Why Visualization?

Task Abstraction for Analysis and Design

by

Matthew Michael Brehmer

B.Cmp. Cognitive Science, Queen's University, 2009

M.Sc. Computer Science (Human-Computer Interaction),

The University of British Columbia, 2011

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Computer Science)

The University of British Columbia
(Vancouver)

April 2016

Abstract

Why do people visualize data?

People visualize data either to consume or produce information relevant to a domain-specific problem or interest. Visualization design and evaluation involves a mapping between domain problems or interests and appropriate visual encoding and interaction design choices. This mapping translates a domain-specific situation into *abstract visualization tasks*, which allows for succinct descriptions of tasks and task sequences in terms of *why* data is visualized, *what* dependencies a task might have in terms of *input* and *output*, and *how* the task is supported in terms of visual encoding and interaction design choices. Describing tasks in this way facilitates the comparison and cross-pollination of visualization design choices across application domains; the mapping also applies in reverse, whenever visualization researchers aim to contextualize novel visualization techniques.

In this dissertation, we present multiple instances of visualization task abstraction, each integrating our proposed typology of abstract visualization tasks. We apply this typology as an analysis tool in an interview study of individuals who visualize dimensionally reduced data in different application domains, in a post-deployment field study evaluation of a visual analysis tool in the domain of investigative journalism, and in a visualization design study in the domain of energy management.

In the interview study, we draw upon and demonstrate the descriptive power of our typology to classify five task sequences relating to visualizing dimensionally reduced data. This classification is intended to inform the design of new tools and techniques for visualizing this form of data.

In the field study, we draw upon and demonstrate the descriptive and evaluative power of our typology to evaluate *Overview*, a visualization tool for investigating large text document collections. After analyzing its adoption by investigative journalists, we characterize two abstract tasks relating to document mining and present seven lessons relating to the design of visualization tools for document data.

In the design study, we demonstrate the descriptive, evaluative, and generative power of our typology and identify matches and mismatches between visualization design choices and three abstract tasks relating to time series data.

Finally, we reflect upon the impact of our task typology.

Preface

Parts of this dissertation have been previously published with various co-authors:

A version of **Chapter 2** has been published as *A Multi-Level Typology of Abstract Visualization Tasks* by **Matthew Brehmer** and Tamara Munzner; in IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis 2013), 19(12), p. 2376–2385 [33][1]. I conducted the literature review. Tamara and I both contributed to the meta-analysis of the literature and writing. A modified version of the task typology proposed in this chapter appears in *Visualization Analysis and Design* by Tamara Munzner (AK Peters Visualization Series, CRC Press, 2014) [219].

A version of **Chapter 3** has been published as *Visualizing Dimensionally Reduced Data: Interviews with Analysts and a Characterization of Task Sequences* by **Matthew Brehmer**, Michael Sedlmair, Stephen Ingram, and Tamara Munzner; in Proceedings of the ACM Workshop on Beyond Time and Errors: Novel Evaluation Methods For Information Visualization (BELIV 2014), p.1-8 [36][2]. This publication was preceded by a technical report entitled *Dimensionality Reduction in the Wild: Gaps and Guidance* by Michael Sedlmair, **Matthew Brehmer**, Stephen Ingram and Tamara Munzner (UBC CS TR-2012-03) [283][3], and by an unpublished manuscript entitled *Dimensionality Reduction in the Wild* by Michael Sedlmair, **Matthew Brehmer**, Stephen Ingram and Tamara Munzner (2013)[4]. Michael con-

[1]`http://dx.doi.org/10.1109/TVCG.2013.124`

[2]`http://dx.doi.org/10.1145/2669557.2669559`

[3]`http://cs.ubc.ca/cgi-bin/tr/2012/TR-2012-03`

[4]Included in the appendices as Section B.6.

ducted the majority of the interviews with analysts between 2010 and 2012. All authors contributed to the initial analysis of the collected data. For the BELIV paper [36], I re-analyzed this data using the task typology described in Chapter 2. I performed the majority of the writing for the BELIV 2014 submission (which omitted much of the material from the earlier technical report and manuscript); Tamara and Michael contributed to the editing process.

A version of **Chapter 4** has been published as *Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists* by **Matthew Brehmer**, Stephen Ingram, Jonathan Stray, and Tamara Munzner; in IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis 2014), 20(12), p. 2271–2280 [35][5]. *Overview* was developed by Jonathan Stray with contributions from Jonas Karlsson, Adam Hooper, and Stephen Ingram. Stephen's algorithmic contributions are documented in greater detail in an earlier technical report [151] and in his PhD dissertation [146]. I conducted a post-deployment evaluation of *Overview* and its use by investigative journalists. Jonathan and I interviewed the TULSA, RYAN, and DALLAS journalists; I interviewed the GUNS journalist, while Jonathan interviewed the NEWYORK journalist. Jonathan conducted the think-aloud evaluation with journalists. I performed the analysis of the interview data (including transcripts and screen captures), as well as the *Overview* log data. I performed the majority of the writing for the InfoVis 2014 submission, while Tamara and Jonathan contributed to the editing process.

A version of **Chapter 5** has been published as *Matches, Mismatches, and Methods: Multiple-View Workflows for Energy Portfolio Analysis* by **Matthew Brehmer**, Jocelyn Ng, Kevin Tate, and Tamara Munzner; in IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis 2015), 22(1), p. 449–458 [37][6]. I conducted the work domain analysis, sandbox prototyping, and the analysis of feedback on prototype designs from energy analysts. Jocelyn and I both contributed to the workflow design.

[5] `http://dx.doi.org/10.1109/TVCG.2014.2346431`
[6] `http://dx.doi.org/10.1109/TVCG.2015.2466971`

Kevin initiated the project and provided feedback on my process during my internship at EnerNOC (then Pulse Energy); he also provided introductions to energy analysts. EnerNOC's Energy Manager development team, led by Cailie Crane and Reetu Mutti, implemented some of our prototype designs into a new commercial version of Energy Manager. I performed the majority of the writing for the InfoVis 2015 submission, while Tamara and Jocelyn contributed to the editing process.

All images in Chapter 2, Chapter 4, and Chapter 5 are reprinted with the permission of the IEEE. Figure 4.3 is a detail from Figure C.1, an image produced by Jonathan Stray [305]. All images in Chapter 3 are reprinted with the permission of the ACM, with the exception of Figure 3.3, which appears in Tenenbaum et al. [313] and is reprinted with the permission of the AAAS. This dissertation includes several illustrations that were originally created by Eamonn Maguire for *Visualization Analysis and Design* by Tamara Munzner [219], including Figure 1.1, Figure 1.8, Figure 6.1, Figure 6.2, Figure 6.3, and Figure 6.4; these illustrations are available for use under the Creative Commons Attribution 4.0 International license (🅭 BY 4.0).

The studies described in this dissertation were conducted with the approval of the UBC Behavioural Research Ethics Board (BREB): certificate number H10-03336.

Table of Contents

Abstract . . . . **ii**

Preface . . . . **iv**

Table of Contents . . . . **vii**

List of Tables . . . . **xiii**

List of Figures . . . . **xv**

Glossary . . . . **xxi**

Acknowledgments . . . . **xxiii**

1 Introduction **1**

1.1 Research Trajectory 2

1.2 Motivation 4

1.3 Thesis Contributions 5

1.3.1 A Typology of Abstract Visualization Tasks 5

1.3.2 Use of the Typology in an Interview Study 7

1.3.3 Use of the Typology in a Field Study 10

1.3.4 Use of the Typology in a Design Study 12

1.3.5 Summary of Contributions 15

1.4 Extension and Impact of the Typology 16

1.5 A Note on Chronology 17

2 A Typology of Abstract Visualization Tasks **20**

2.1 Motivation 21
2.2 Background Context 23
2.3 A Typology of Tasks 27
2.3.1 Why is Data Being Visualized? 28
2.3.2 How Does the Visualization Technique or Tool Support the Task? 31
2.3.3 What are the Inputs and Outputs of the Task? 34
2.3.4 Concise Task Descriptions 35
2.4 Example: A Sequence of Interdependent Tasks 36
2.5 Connections to Previous Work 39
2.5.1 Existing Classifications 40
2.5.2 Theoretical Foundations 45
2.6 Discussion 47
2.6.1 Using the Typology to Describe 47
2.6.2 Using the Typology to Generate 48
2.6.3 Using the Typology to Evaluate 49
2.7 Summary 50

3 Interview Study: Visualizing Dimensionally Reduced Data: Interviews with Analysts and a Classification of Task Sequences 51
3.1 Motivation 51
3.2 Related Work 53
3.3 Research Process 54
3.4 Task Sequences 56
3.4.1 Dimension-Oriented Task Sequences 57
3.4.2 Cluster-Oriented Task Sequences 61
3.5 A Task Typology Revisited 64
3.6 Discussion 67
3.6.1 Implications for Evaluation 67
3.6.2 Limitations 70
3.7 Summary 71

4 Field Study:
Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists . . . 73
4.1 Motivation . . . 74
4.2 Related Work . . . 76
4.3 Initial Use Case . . . 79
4.4 Design of Overview . . . 80
4.5 Observations of Real World Usage . . . 84
4.5.1 Case Studies . . . 86
4.5.2 Think-Aloud Evaluation . . . 91
4.6 Analysis . . . 93
4.6.1 Task Abstractions Reconsidered . . . 93
4.6.2 Design Rationale . . . 95
4.7 Discussion . . . 101
4.8 Summary . . . 103
4.9 Addendum . . . 104

5 Design Study:
Matches, Mismatches, and Methods: Multiple-View Workflows for Energy Portfolio Analysis . . . 106
5.1 Motivation . . . 107
5.2 Methodology . . . 108
5.3 Abstractions . . . 110
5.3.1 Data Abstraction . . . 110
5.3.2 Task Abstraction . . . 113
5.4 The Previous Version of Energy Manager . . . 114
5.5 Related Work . . . 119
5.6 Prototyping Environments . . . 121
5.7 Visual Encoding Matches and Mismatches . . . 123
5.7.1 Faceted Views for Overview and Drill Down . . . 124
5.7.2 Rank-Based Overviews . . . 126
5.7.3 Matrix-Based Overviews . . . 127
5.7.4 Map-Based Overviews . . . 128

5.7.5 Stack-Based Roll Up Encodings . . . 129
5.8 Workflow Design with Multiple Views . . . 130
5.9 Results . . . 133
5.10 Discussion . . . 135
5.10.1 Guidelines: Familiarity and Trust . . . 135
5.10.2 Methodological Reflection . . . 137
5.11 Summary . . . 138
5.12 Addendum . . . 139

6 Reflection and Conclusion . . . 142
6.1 Reflecting on the Task Typology . . . 142
6.1.1 An Extended Task Typology . . . 142
6.1.2 Comparisons to Roth (2013), Schulz et al. (2013) . . . 149
6.1.3 Impact of the Task Typology . . . 153
6.2 Reflecting on the Interview Study . . . 157
6.3 Reflecting on the Field Study . . . 159
6.4 Reflecting on the Design Study . . . 160
6.5 Concluding Thoughts . . . 162

Bibliography . . . 165

A Appendix: Task Typology (A Chronology) . . . 213
A.1 Preliminary Influences . . . 214
A.2 Dedicated Literature Review . . . 227
A.3 Meta-Analysis of Existing Classifications . . . 242
A.4 Our Initial Classifications of Tasks . . . 246
A.5 Mid-Level Visualization Tasks . . . 271
A.6 Our Proposed Taxonomy of Tasks . . . 275
A.7 Revisions: From Taxonomy to Typology . . . 276
A.8 Presenting our Typology . . . 282
A.9 Subsequent Evolution of our Typology . . . 283

B Appendix: Interview Study . . . 285
B.1 Complete List of Interviews . . . 285

B.2 Interview Foci and Questions 285
B.3 Data Collection, Analysis, and Abstraction 288
B.4 Data Analysis Artefact Examples 289
B.5 Previous Interpretations of Findings 289
B.6 Dimensionality Reduction in the Wild 296
B.6.1 Introduction 297
B.6.2 Related Work 301
B.6.3 Methodology 305
B.6.4 Taxonomy 309
B.6.5 Challenges 324
B.6.6 Benchmarks 327
B.6.7 Discussion 329
B.6.8 Conclusions 331

C Appendix: Field Study 332
C.1 Initial Use Case 332
C.2 Overview v1 332
C.3 Field Study Proposal 334
C.3.1 Research Questions 335
C.3.2 Research Context 336
C.3.3 Methodology 336
C.3.4 Outcomes and Follow-on Work 341
C.4 Interview Protocol 341
C.5 Preliminary Field Study Results 350
C.5.1 Introduction 350
C.5.2 The Overview Project 352
C.5.3 Methodology 354
C.5.4 Findings 359
C.5.5 Discussion 363
C.5.6 Conclusion and Future Work 367
C.6 Overview v3 367

D Appendix: Design Study 369

D.1 Design Study Proposal 369
D.1.1 Domain Background 370
D.1.2 Information Visualization Background 371
D.1.3 Objectives 372
D.1.4 Methodology 373
D.1.5 Desired Outcome 375
D.1.6 Milestones 376
D.2 Example Research Artefacts 378

E Appendix: Consent Forms 391

Index 394

List of Tables

Table 2.1 Nodes in the *why* part of our typology and their relation to the vocabulary used in previous work. 41
Table 2.2 Nodes in the *how* part of our typology and their relation to the vocabulary used in previous work. 42

Table 3.1 A summary of task sequences performed by the ten analysts that we interviewed and found in papers discussing dimensionality reduction (DR) and visualization. 58

Table 4.1 A summary of the six case studies. 92

Table 5.1 Data abstraction summary. 111
Table 5.2 A summary of the *matches* and *mismatches* between abstract tasks and visual encoding design choices. 124

Table A.1 Additional dimensions of previous classifications. 245

Table B.1 The complete set of twenty-four interviews with nineteen analysts. 286
Table B.2 Establishing conceptual relationships across summaries. . 292
Table B.3 Establishing conceptual relationships across summaries (continued). 292
Table B.4 Establishing conceptual relationships across summaries (continued). 296
Table B.5 Establishing conceptual relationships across summaries (continued). 296

Table B.6 Usage examples described using our preliminary classification of people who use DR, DR techniques, and tasks. . . 298
Table B.7 Usage examples described using our classification of tasks relating to DR. 309

Table D.1 Characterizing energy analysts activities as abstract tasks. 380

List of Figures

Figure 1.1 Munzner's nested model of visualization design. 6
Figure 1.2 Our multi-level typology of abstract visualization tasks. . 7
Figure 1.3 Concise task descriptions using elements of the typology. 8
Figure 1.4 An example sequence of tasks using the structure and vocabulary of our typology. 8
Figure 1.5 Task sequences that involve visualizing dimensionally reduced data. 10
Figure 1.6 *Overview*, a multiple-view application intended for use during an investigation of a large collection of text documents. 11
Figure 1.7 A sandbox environment for creating visualization data sketches pertaining to energy portfolio analysis. 15
Figure 1.8 A modified version of our typology appearing in Munzner [219]. 18
Figure 1.9 Timelines of the projects described in this dissertation. . 19

Figure 2.1 Our multi-level typology of abstract visualization tasks: *why*, *how*, and *what*. 27
Figure 2.2 Task descriptions for Example #1 (choropleth map) and Example #2 (large trees). 36
Figure 2.3 An example sequence of tasks as using the structure and vocabulary of our typology. 37

Figure 3.1 A task sequence involving dimensionally reduced data. . 52

Figure 3.2 Five task sequences that involve visualizing dimensionally reduced data. 56
Figure 3.3 A visual encoding of dimensionally reduced data, in which three synthesized dimensions have been identified. 60
Figure 3.4 Scatterplots of dimensionally reduced data illustrating tasks related to item clusters. 62
Figure 3.5 Six tasks related to dimensionally reduced data, characterized using our abstract task typology. 65
Figure 3.6 A refinement to the *why* part of our abstract task typology. 66

Figure 4.1 *Overview* is a multiple-view application intended for the systematic search, summarization, annotation, and reading of a large collection of text documents. 75
Figure 4.2 A timeline of *Overview*'s development, deployment, and adoption phases. 76
Figure 4.3 Detail from *"A full-text visualization of the Iraq War Logs"*. 80
Figure 4.4 *Overview v2*, a desktop application released in Winter 2012. 81
Figure 4.5 *Overview v4*, a web-based application released in Summer 2013. 82
Figure 4.6 The human-centred design process development cycle. . 103
Figure 4.7 The two tasks characterized in Section 4.6.1. 105
Figure 4.8 The task of `annotating` documents and clusters. 105

Figure 5.1 The previous version of Energy Manager, our collaborators' energy analysis tool. 116
Figure 5.2 The previous version of Energy Manager, our collaborators' energy analysis tool (continued). 117
Figure 5.3 A sandbox design environment for visualizing energy data from a portfolio of buildings. 122

Figure 5.4 Faceted boxplots that encode aggregate area-normalized energy *demand* distributions. 125
Figure 5.5 A *bar + bump plot* of *energy intensity*. 127
Figure 5.6 A time series calendar matrix of *energy intensity savings*. 129
Figure 5.7 An interactive auxiliary boxplot prototype. 131
Figure 5.8 A stacked area graph of *energy demand*. 132
Figure 5.9 The redesigned Energy Manager that incorporates many aspects of our prototype designs. 133
Figure 5.10 The three tasks identified in Section 5.3.2. 141

Figure 6.1 The *why* part of our typology, slightly extended and recast as a set of *actions*. 144
Figure 6.2 A set of `targets`, to be used in conjunction with a specification of `actions`. 145
Figure 6.3 Munzner's classifcation of *what* [219]. 146
Figure 6.4 Munzner's extension [219] to the *how* part of our typology. 147
Figure 6.5 Our typology is used to describe Shneiderman's *visual information seeking mantra*: *overview first, zoom and filter, details-on-demand*. 153

Figure A.1 Early brainstorming on the topic of task classification. . 229
Figure A.2 Dimensions of high-level, mid-level, and low-level tasks. . 243
Figure A.3 Mid-level abstract tasks along the axes of domain specificity and interface specificity. 244
Figure A.4 Our first classification. 248
Figure A.5 Our second classification. 250
Figure A.6 Whiteboard and post-it diagramming prior to our third classification 251
Figure A.7 Our third classification. 252
Figure A.8 Whiteboard and post-it diagramming prior to our fourth classification. 253
Figure A.9 Our fourth classification. 259

Figure A.10 Cross-cutting dimensions of previous classifications or frameworks. . . . 264
Figure A.11 Our fifth classification. . . . 269
Figure A.12 Our sixth classification. . . . 274
Figure A.13 Our seventh and proposed classification. . . . 276
Figure A.14 Our multi-level typology of abstract visualization tasks as of October 2013. . . . 283
Figure A.15 Previous classifications sorted from low to high level of abstraction. . . . 284

Figure B.1 An example of an interview transcript. . . . 290
Figure B.2 An example of raw inter notes with post-hoc annotations. 291
Figure B.3 An example document sent to us by an analyst. . . . 293
Figure B.4 An early version of an analyst summary. . . . 294
Figure B.5 A later version of an analyst summary. . . . 295
Figure B.6 Our preliminary classification of people who use DR, DR techniques, and tasks. . . . 297
Figure B.7 Our classification of dimensionality reduction algorithms. 303
Figure B.8 Our data collection, analysis, and abstraction methodology. . . . 305
Figure B.9 Our classification of tasks relating to DR. . . . 309
Figure B.10 Example scatterplots of dimensionally reduced data illustrating potential tasks related to checking cluster separability. . . . 318

Figure C.1 *"A full-text visualization of the Iraq War Logs"*. . . . 333
Figure C.2 *Overview v1*, a prototype desktop application completed in Fall 2011. . . . 334
Figure C.3 *Overview v2*, displaying TULSA's email corpus. . . . 353
Figure C.4 An excerpt of *Overview*'s log file, listing timestamped interaction events. . . . 359
Figure C.5 The median time spent viewing a single document. . . . 362

Figure C.6 *Overview v3*, a web-based application released in Fall 2012. 368

Figure D.1 Eleven slide decks created between Nov 2013 and February 2014. 378
Figure D.2 Partial summary of findings from initial interviews with energy analysts. 379
Figure D.3 Partial summary of findings from initial interviews with energy analysts (continued). 379
Figure D.4 Partial characterization of data abstractions relevant to energy analysts' activities. 380
Figure D.5 Partial characterization of data abstractions relevant to energy analysts' activities (continued). 381
Figure D.6 Verifying the task and data abstractions with power user energy analysts. 381
Figure D.7 Verifying the task and data abstractions with power user energy analysts (continued). 382
Figure D.8 Initial data sketches produced within the sandbox environment. 382
Figure D.9 Initial data sketches produced within the sandbox environment (continued). 383
Figure D.10 Following-up with the power user energy analysts with designs from our sandbox design. 383
Figure D.11 Following-up with the power user energy analysts with designs from our sandbox design (continued). 384
Figure D.12 Another iteration of data sketches produced using the sandbox environment. 384
Figure D.13 Another iteration of data sketches produced using the sandbox environment (continued). 385
Figure D.14 Early view coordination design depicting a matrix with auxiliary boxplots. 385
Figure D.15 Proposed workflow design involving multiple views based on consolidated feedback from energy analysts. 386

Figure D.16 Storyboards using sandbox screenshots based on power user workflows. 386
Figure D.17 Storyboards using sandbox screenshots based on power user workflows (continued). 387
Figure D.18 Color stock charts with juxtaposed line charts as alternative to matrix with juxtaposed boxplots. 388
Figure D.19 Values from the brushed time period are highlighted on the juxtaposed boxplots. 389
Figure D.20 Boxplot for the brushed time period (red) is shown alongside the boxplot for the entire time series. 389
Figure D.21 An example of how feedback was documented. 390

Glossary

API application programming interface

BELIV Beyond Time and Errors: Novel Evaluation Methods for Visualization, a bi-annual ACM workshop focusing on the challenges of evaluation in visualization

CCA canonical correlation analysis

CSV comma-separated values

DR dimensionality reduction

FOIA Freedom of Information Act, allows individuals such as journalists to request documents and data from public institutions

HCI human-computer interaction

HDD heating degree day, one of several approaches to normalizing energy performance using weather data; a full discussion of them is beyond the scope of this thesis

JSON JavaScript object notation

kW kilowatts

kWh kilowatt-hours

MDS multi-dimensional scaling

NLP natural language processing

OCR optical character recognition

PCA principal component analysis

PDF portable document format

SFS sequential forward selection

SPLOM scatterplot matrix

t-SNE t-distributed stochastic neighbor embedding

TF-IDF term frequency-inverse document frequency

UI user interface

WIMP windows-icons-menus-pointer

Acknowledgments

Acknowledgements by chapter:

Chapter 2: First, I thank my co-author Tamara Munzner, who first conceived this project during her PhD and during early discussions with François Guimbretière. I acknowledge Ron Rensink for the original idea of classifying lookup, browse, locate, and explore according to target identity and location. I also thank Jessica Dawson, Joel Ferstay, Stephen Ingram, Joanna McGrenere, Miriah Meyer, Michael Sedlmair, and Colin Ware for their feedback on the paper. We received financial support for this project from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Chapter 3: I thank my co-authors: Michael Sedlmair, Stephen Ingram, and Tamara Munzner; Michael was the first author of the technical report [283] that was the starting point for our 2014 Beyond Time and Errors: Novel Evaluation Methods for Visualization (BELIV) paper [36] and this chapter. I also thank the data analysts who participated in the original interview study for their time and energy: Kerem Altun, Ryan Brinkman, Jennifer Büttgen, Anamaria Crişan, Klaus Dress, Des Higgins, Carrie Holt, Heidi Lam, Kevin Leyton-Brown, Cindy Marven, Greg Mori, Sareh Nabi-Abdolyousefi, Cydney Nielsen, Ahmed Saad, Jonathan Stray, Sid Thakur, John Westbrook, James Wright, and Hong Yi. Finally, I thank Steven Bergner, Jessica Dawson, Joel Ferstay, Miriah Meyer, Torsten Möller, Tom Torsney-Weir, Melanie Tory, and Hamidreza Younesy for assisting with interviews and/or feedback on paper drafts. We received financial support for this project from NSERC.

Chapter 4: I thank my co-authors: Stephen Ingram, Jonathan Stray, and Tamara Munzner. I thank Jonas Karlsson and Adam Hooper, who contributed to the development *Overview* v3-v4. I also thank the case study journalists: Jack Gillum, Ian James, Michael Keller, Adam Playford, Jonathan Stray, Jarrel Wade, and the DALLAS journalist who requested to remain anonymous. Finally, I thank Jessica Dawson, Joel Ferstay, Heidi Lam, Joanna McGrenere, Ron Rensink, and Michael Sedlmair for their comments on the project and paper. We received financial support for this project from the Knight News Challenge and NSERC.

Chapter 5: I thank my co-authors: Jocelyn Ng, Kevin Tate, and Tamara Munzner. I also thank our collaborators at EnerNOC (formerly Pulse Energy): Bruce Cullen, Ben Gready, David Helliwell, Bruce Herzer, Steve Jones, Jamie King, Sarah Laird, Fritz Lapastora, Ari Lesniak, Jordana Mah, Harish Raisinghani, Maria Serbenescu, Paul Teehan, and especially James Christopherson. I also especially thank Cailie Crane, Reetu Mutti, and the Energy Manager development team. I also thank the energy analysts that we interviewed: Marc Etienne Brunet, Andy Constant, Bill Edbrooke, Chris Goodchild, Marc Tabet, Sean Terry, Natalie Vadeboncoeur, Lillian Zaremba, and especially Jerome Conraud and Kevin Ng. Finally, I thank Michelle Borkin, Anamaria Crişan, Jessica Dawson, Johanna Fulda, Enamul Hoque, Sung-Hee Kim, Narges Mahyar, and Joanna McGrenere for their feedback on the paper. We received financial support for this project from NSERC and Mitacs.

For each of the four research chapters, I thank the anonymous reviewers who reviewed the associated research paper, and in particular I thank the reviewer of our typology paper [33] who introduced me to *Cognitive Work Analysis* by Vicente [337], which proved to be an excellent resource.

Personal acknowledgements:

First and foremost, I thank Tamara Munzner, my primary thesis advisor, who first reached out to me during the 2009 graduate recruitment season, taught me the principles of visualization analysis and design in her graduate course later that year, and took me on as a PhD student in 2011. I thank her for all of our discussions, her attention to detail, and the career mentoring that she has provided over the last several years.

I thank Joanna McGrenere, my co-advisor. Joanna encouraged me to pursue a PhD, and has since provided thoughtful mentoring and perspectives on HCI research.

I thank Ron Rensink, the third member of my supervisory committee. I am particularly thankful for Ron's graduate seminar in visual display design, which provided dual perspectives from perceptual psychology and design, as well as opportunities to practice succinct writing and peer review.

I thank my external examiner Jason Dykes for his very thorough report and the thought-provoking questions that he raised.

I thank my university examiners Giuseppe Carenini and Alfred Hermida, as well as my thesis defense chair Luanne Freund.

I thank Giuseppe Carenini, chair of my thesis proposal defense meeting in 2014, and Ron Garcia, chair of my research proficiency evaluation meeting in 2012.

I thank the members of the Tamara Munzner's InfoVis group between 2011 and 2016: Michelle Borkin, Anamaria Crişan, Jessica Dawson, Kimberly Dextras-Romagnino, Wenqiang (Dylan) Dong, Joel Ferstay, Johanna Fulda, Stephen Ingram, Zipeng Liu, and Michael Sedlmair. I also thank regular group meeting attendees Giuseppe Carenini, Enamul Hoque, Sung-Hee Kim, and Narges Mahyar, as well as group alumni Heidi Lam, Miriah Meyer, and Melanie Tory who joined in person or remotely for occasional meetings.

I thank the members of Joanna McGrenere's LUNCH (Lab for Universal usability, persoNalization, Cscw, and Hci) research group between 2011 and 2016: Kamyar Ardekany, Jessica Dawson, Shathel Haddad, Mona Haraty, Sung-Hee Kim, Juliette Link, Matei Negulescu, Antoine Ponsard, Diane Tam, Charlotte Tang, and Kailun Zhang.

I thank the members of the UBC MUX (Multimodal User eXperience) lab between 2011 and 2016 for the feedback and support they provided during research update presentations and practice talks. In addition to all of those already mentioned above from the InfoVis and LUNCH groups, I thank faculty members Kellogg Booth and Karon MacLean, as well as all the post-docs, visiting researchers, and students who have contributed to the culture of the MUX lab.

I thank the InfoVis subcommittee of the 2014 IEEE Doctoral Colloquium for their feedback on my thesis research program: Christopher Collins, Petra Isenberg, and Chris Weaver. I also thank my fellow Doctoral Colloquium participants: Sriram Karthik Badam, Richard Brath, Samuel Gratzl, Julia Jürgens, and Jorge Poco.

I also thank those that I collaborated and co-authored with on other projects between 2011 and 2016: Benjamin Bach, Johanna Fulda, Nathalie Henry Riche, Claudia Jacova, Joanna McGrenere, Charlotte Tang, Melanie Tory, Sheelagh Carpendale, Donghao Ren, and especially Bongshin Lee.

I thank Eamonn Maguire for the excellent illustrations that he created for Tamara's book [219] and for making it possible for me to use them in this dissertation.

I thank Sandra Mathison, instructor of EPSE 595 (a qualitative research methods course offered in Winter 2012), who introduced me to a number of epistemological and theoretical perspectives, methodologies, and methods.

I thank Laura Selander, group assistant for the UBC Imager lab, and Joyce Poon, UBC Department of Computer Science graduate program assistant, for their logistics support.

I thank Joanna McGrenere and Claudia Jacova, my M.Sc advisors, as well as T. C. Nicholas Graham, my undergraduate honours thesis advisor; my decision to pursue a PhD was in part motivated by my early research experiences with them and by their encouragement.

I thank my family for all of their love and support over the years: my parents Leslie and Michael Brehmer, my brother Nicholas Brehmer, my grandparents, and my uncle Thomas Brehmer, who introduced me to the work of Edward Tufte at an impressionable age.

Lastly, I thank my partner, Anamaria Crişan: for enduring, for understanding, and for joining me on this path.

Chapter 1

Introduction

Why do people visualize data?

Ultimately, visualizing data allows people to consume and produce information in order to solve domain-specific problems or to communicate an understanding about phenomena relevant to a particular domain. Visualization is often associated with data analysis and communication processes, however it is important to stress that not all data analysis and communication tasks are addressed via visualization. This dissertation describes an approach that researchers and practitioners can use to systematically classify and abstract visualization tasks, whether they occur in a data analysis or communication context, and whether they involve the consumption or the production of information. This abstraction of tasks is necessary and important because this abstraction facilitates visualization analysis and design: it can be used to communicate and transfer lessons learned from studying visualization tasks in specific application domains or with specific datatypes, it can be used to understand the implications of findings from controlled experiments, and it can be used to contextualize novel visualization techniques.

In this dissertation, we[1] present a typology of abstract visualization tasks

[1]With the exception of personal anecdotes in Section 1.1 and in the conclusion chapter, I will use the pronoun *we* throughout this dissertation to reflect the collaborative nature of the work reported therein.

and document its application in three studies: in an interview study of individuals who visualize dimensionally reduced data in ten different application domains, in a post-deployment field study evaluation of a visual analysis tool in the domain of investigative journalism, and in a visualization design study in the domain of energy management. We also survey how our approach to task analysis has been adopted and extended by others visualization community.

1.1 Research Trajectory

Before discussing the contents and contributions of this dissertation in more detail, I will tell the story of how I came to study why people visualize data.

When I entered my PhD program in late 2011, I posed the following question: *How do we evaluate visualization techniques and tools in an application domain context, particularly if these techniques and tools are used for data analysis?* At this time, I read an early manuscript of a survey of visualization evaluation by Lam et al. [183], in which the authors read and coded over eight hundred recent visualization research papers that report an evaluation component. While many of these papers discuss human perceptual performance or visualization usability, relatively few of these papers document an attempt to evaluate the use of visualization tools or techniques in settings other than in a controlled experiment, and fewer still comment on *adoption*: whether a deployed visualization tool was incorporated into the recurring data analysis workflows of individuals working in a specific application domain. The findings of this survey prompted me to ask: *Why is the study of real-world usage and adoption of visualization techniques or tools reported so infrequently?* and *If this research is difficult to conduct, what makes it so difficult?*

Initially, I focused my study on the corpus of research papers emanating from the ACM Beyond Time and Errors: Novel Evaluation Methods for Visualization (BELIV) workshop series, where a number of methodologies and methods for evaluating visualization tools or techniques "in the wild" have been proposed. At the 2012 BELIV workshop, there was substantial discussion pertaining to a need for a better shared understanding of the *tasks* of people who visualize data or use visualization tools, and that the effective application of visualization evaluation methodologies depends upon this understanding.

My interest in evaluating visualization techniques and tools in real-world settings prompted me to undertake field study and a design study; both cases involved a visualization tool that was developed to address domain-specific problems. In a design study, it is critical that the researchers have correctly abstracted the domain problems or use cases and mapped these to appropriate visual encoding and interaction design choices, perhaps incorporating design choices originally applied to other domains. However, this abstraction and mapping is seldom straightforward: Sedlmair et al. [284] describe how initial designs often fail to address the tasks that people are expected to perform, how inappropriate evaluation methods are chosen, or how prematurely deployed visualization tools fail to be adopted by the people for which they were designed.

By early 2013, my thinking had coalesced into a thesis statement expressed as follows: *visualization design and evaluation is difficult because mapping a person's tasks to visual encoding and interaction design choices requires multiple levels of abstraction. Researchers and practitioners would benefit from a domain-agnostic, consistent, and validated approach for assisting in this mapping.*

What is a "task"? A *task* is an ill-defined concept, and as the reader will discover in Chapter 2, there is currently little agreement in the visualization literature as to the appropriate granularity for describing a task. For instance, *finding an extreme value* [8] is less abstract than *exploring* [366] or *integrating insights* [301], while *comparing sequence variants in a human genome* is quite domain-specific. This confusion is the result of a conflation of two axes on which we might characterize a task: level of abstraction and applicability, in which the latter refers to the specificity of a task with respect to a particular application domain or datatype. Relating these task descriptions is difficult, though not impossible; however, visualization practitioners hardly have a shared lexicon when describing these relations between levels of abstraction and application areas.

Researchers and practitioners should strive to go beyond merely describing the use of visualization tools or techniques in a specific application domain; rather they should *abstract* these domain-specific tasks in order to realize an appropriate visualization design space. This abstraction also lets practitioners contribute back to the visualization research community, transferring their findings beyond a single domain.

My dissertation examines visualization task abstraction from multiple perspectives. Chapter 2 documents the synthesis of related work classifying tasks, interactions, and visualization design choices. The result of this synthesis was a new approach to task analysis, a typology for classifying visualization tasks at multiple levels of abstraction. However, proposing this typology was not enough. We had to validate this typology as a pragmatic tool [16]; to do so, we used the typology to *describe* existing interactions between people and visualization techniques or tools, to *generate* new designs, and to *evaluate* these designs. The three forms of validation are intertwined, as the ability to *generate* or *evaluate* implies the ability to *describe*; in this dissertation, we address all three types of validation.

The remainder of this dissertation serves to validate this typology in applied settings spanning multiple domains. In Chapter 3, we used our typology to *analyze* findings from an interview study spanning several different application domains, focusing on individuals who visualize dimensionally reduced data. In Chapter 4, we used our typology in a field study to *evaluate Overview*, a visual analysis tool that was adopted by investigative journalists. Finally, in Chapter 5, we used our typology to *design* and evaluate visualization designs in the domain of energy management.

1.2 Motivation

Task analysis is essential for visualization analysis, evaluation, and design. While there are many approaches to visualization task analysis, they vary in terms of level of abstraction, applicability across domains and datatypes, as well as in terms of the vocabulary that they use. The visualization community requires a synthesis of existing task analysis approaches and theoretical foundations, one that spans multiple levels of abstraction, spans across domains and datatypes, and introduces a common and consistent task lexicon[2].

This dissertation demonstrates that such a synthesis is possible. Furthermore, we demonstrate our approach to task analysis in visualization analysis, evaluation, and design projects. We also report on the adoption of our approach by others in the community.

1.3 Thesis Contributions

There are four research chapters contained in this dissertation, each offering contributions to the visualization research community. A succinct summary of all contributions appears at the end of this section.

1.3.1 A Typology of Abstract Visualization Tasks

The visualization design and evaluation process is characterized by multiple levels [217, 219], as characterized by Munzner's nested model, shown in see Figure 1.1. These levels include the domain problem, data and task abstractions, visual encoding and interaction design choices, and ultimately the algorithms that drive them. In this dissertation, our focus is on task abstractions, how they map upward to domain problems, and how they map downward to visualization design choices.

As indicated in Section 1.1, my personal motivation for developing an approach for classifying and abstracting tasks was pragmatic: we had amassed observational data of the use of visualization tools and techniques in the interview study and field study projects, described in Chapter 3 and Chapter 4 respectively, where we struggled to describe and compare tasks performed by different people, tasks performed with different visualization techniques or tools, as well as tasks associated with different application domains. We required a systematic approach for analyzing tasks abstractly, allowing us to describe and evaluate visualization design choices that address these tasks.

[2]We elaborate on this motivation in Chapter 2

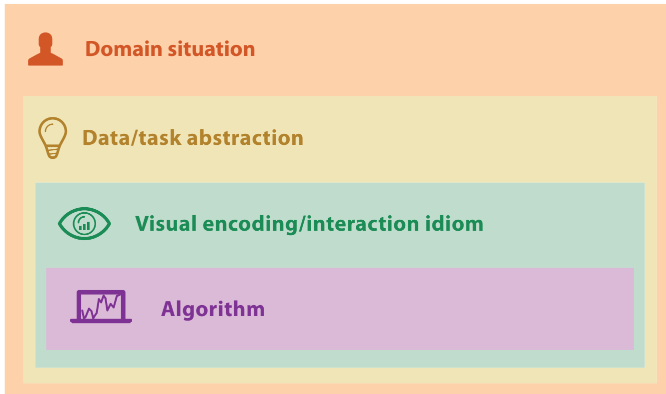


Figure 1.1: Munzner's nested model of visualization design [217, 219]; the arrows indicate the cascading effects of design decisions made at higher levels. Illustration: 🅭 E. Maguire (2014).

Methodology: In Chapter 2, we describe our comprehensive review of previous work that classified tasks, interactions, activities, and visualization design choices. This review included over two dozen previous classification systems and theoretical frameworks from the literatures of visualization, human-computer interaction (HCI), information retrieval, communications, and cartography. We examined the vocabulary and definitions used in this body of previous work, and after multiple rounds of coding, we had grouped similar terms, determined representative terms for each group, and arranged these representative terms into multiple levels of abstraction[3]. We reasoned about how tasks could be described using this arrangement of terms, either in isolation, or as a sequence of interdependent tasks.

Contributions: The result of our synthesis was a typology of abstract visualization tasks, illustrated in Figure 1.2. This typology allows for succinct descriptions of tasks, in which a task description is comprised of *why* data is visualized (at multiple levels of abstraction), *what* dependencies a task might have in terms of *input* and *output*, and *how* the task is or can be supported in terms of visual encoding and interaction design choices; given

[3] Appendix A documents the evolution of our typology.

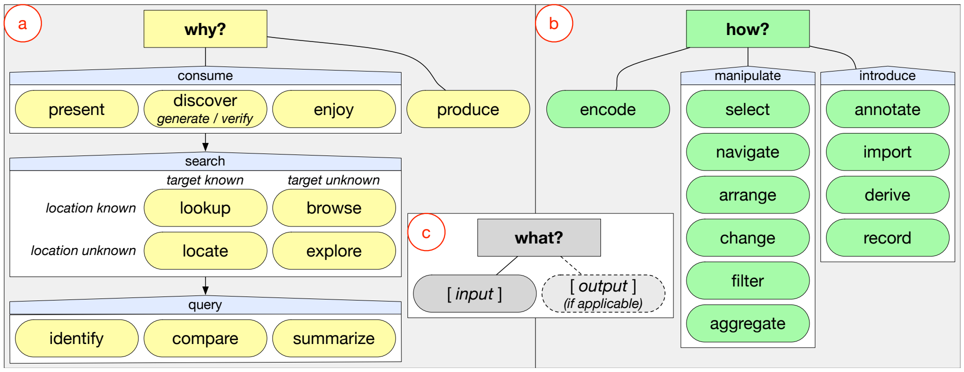


Figure 1.2: Our multi-level typology of abstract visualization tasks, which classifies (a) *why* data is visualized, (b) *how* the task is supported in terms of visual encoding and interaction design choices, and (c) *what* dependencies a task might have. Note that the colors used for *why* and *how* correspond to the abstraction and technique design levels of Munzner's nested model [217], shown in Figure 1.1.

this structure, it is possible to describe sequences of interdependent tasks, as illustrated in Figure 1.3 and in the example of Figure 1.4. Our typology has since proven to be useful in our subsequent interview study (Chapter 3), field study (Chapter 4), and design study (Chapter 5) projects, as well as in recent work by others; we reflect upon the adaptation and use of our typology by others in Chapter 6.

1.3.2 Use of the Typology in an Interview Study

In Chapter 3, we used our typology to *analyze* data analysis and the use of visualization techniques and tools "in the wild" by way of an interview study. In particular, we used the typology to examine the data analysis tasks of individuals working in several different domains, and specifically tasks related to the analysis of high-dimensional data; we sought to better understand this data, the dimensionality reduction (DR) transformations applied to it, as well as *why* and *how* visualization techniques and tools are used throughout analysts' domain-specific workflows.

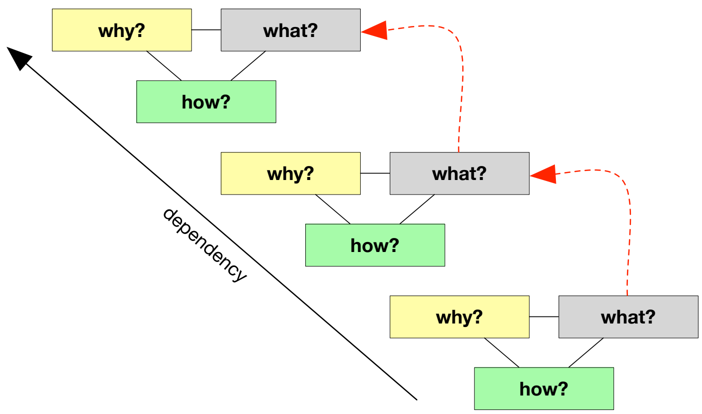


Figure 1.3: Concise task descriptions are constructed using elements from each part of the typology. In specifying the input and output of tasks, we can describe sequences of interdependent tasks.

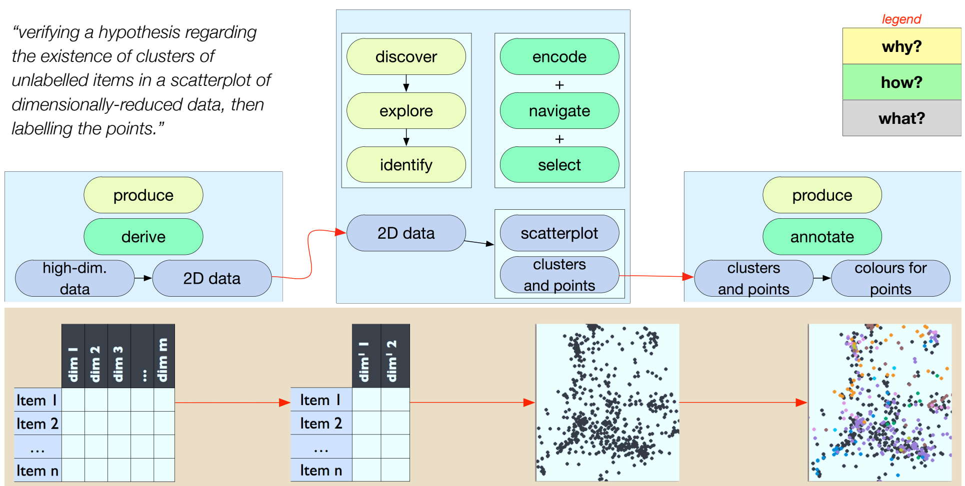


Figure 1.4: An example sequence of tasks, described as a sentence (top left), as well as using the structure and vocabulary of our typology (top); the bottom depicts a series of transformations corresponding to the inputs and outputs of each task. This particular series of abstract tasks is relevant to both the interview study (Chapter 3) and field study (Chapter 4) projects, as they both involve high-dimensional data, dimensionality reduction, and the visualization of dimensionally reduced data.

Methodology: The focus of this research was to classify the tasks associated with the visualization of dimensionally reduced data, such as in the example of Figure 1.4. Our data collection and analysis methodology included twenty-four interviews with researchers and a literature survey spanning several application domains, including HCI, chemistry, bioinformatics, computer science, and policy analysis. Our approach was similar to an interview study by Kandel et al. [163], one classifying data analysis and visualization among enterprise data analysts; we view our work as being complementary to their findings, given that both projects addressed data analysis and visualization "in the wild" for a broad group of domains. We collected a large amount of data: diagrams, screen shots, interview notes, recordings, and transcripts, as well as interviewees' research papers, their data, and other research artifacts.

Using a qualitative coding approach, we developed a classification of task sequences relating to visualizing dimensionally reduced data, which are illustrated in Figure 1.5, where each sequence is comprised of tasks, and each task can be defined using the vocabulary of our task typology. We distinguished between tasks relating to learning about the synthetic dimensions resulting from DR and those relating to learning about clusters of items in the dimensionally reduced data.

Contributions: With the advent of the task typology proposed in Chapter 2, we had a theoretical lens and vocabulary with which to approach the considerable amount of data that we collected from our interview study. Using the vocabulary of our typology, we were able to classify *why* these techniques are applied in sequential workflows, as well as *what* the *inputs* and *outputs* of these tasks are.

We contribute a *datatype-specific* classification of tasks grounded in observations of real-world analyst behaviour. We encourage the further classification of tasks specific to datatype, as these are complementary to our datatype-agnostic typology that we introduce in Chapter 2; examples in the literature include the often-cited task by datatype taxonomy by Shneiderman [291], classifications of graph-specific tasks [186, 272], tabu-

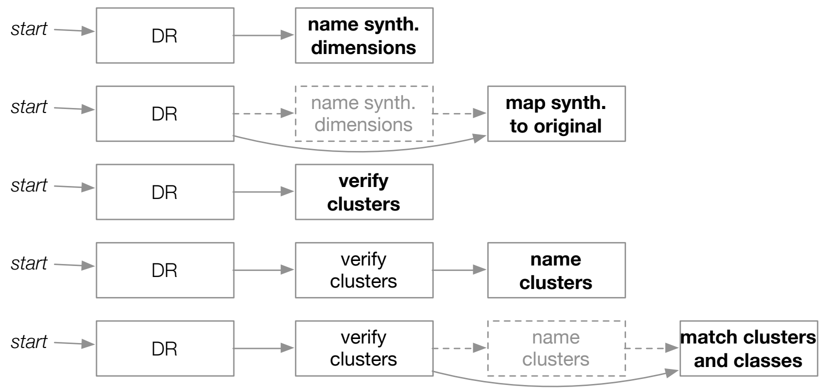


Figure 1.5: A classification of five task sequences that involve visualizing dimensionally reduced data, based upon findings from an interview study, documented in Chapter 3.

lar data [133], and time-oriented data [185]. When the paper about our interview study was first published, there was no prior classification of tasks relating to visualizing dimensionally reduced data[4]. The findings of our interview study and our classification of tasks is further contextualized with references to specific visual encoding design choices; as a result, our classification of tasks can serve to validate and inform visualization technique research, a challenge that we identified in previous work [150].

1.3.3 Use of the Typology in a Field Study

In 2010, our research group began collaborating with a professional journalist who was developing a visualization tool intended for the exploration of large text document collections. Since this time, the tool has been deployed as *Overview*[5] (shown in Figure 1.6), a web-based application for investigative journalists who report on large document collections attained from Freedom of Information Act (FOIA) requests or from whistleblower organizations, collections ranging in size from hundreds to tens of thousands of documents. Between 2012 and 2014, we conducted a post-deployment field study evalu-

[4]A 2015 task taxonomy by Etemadpour et al. [94] is discussed in Section 6.2.
[5]`https://www.overviewdocs.com/`

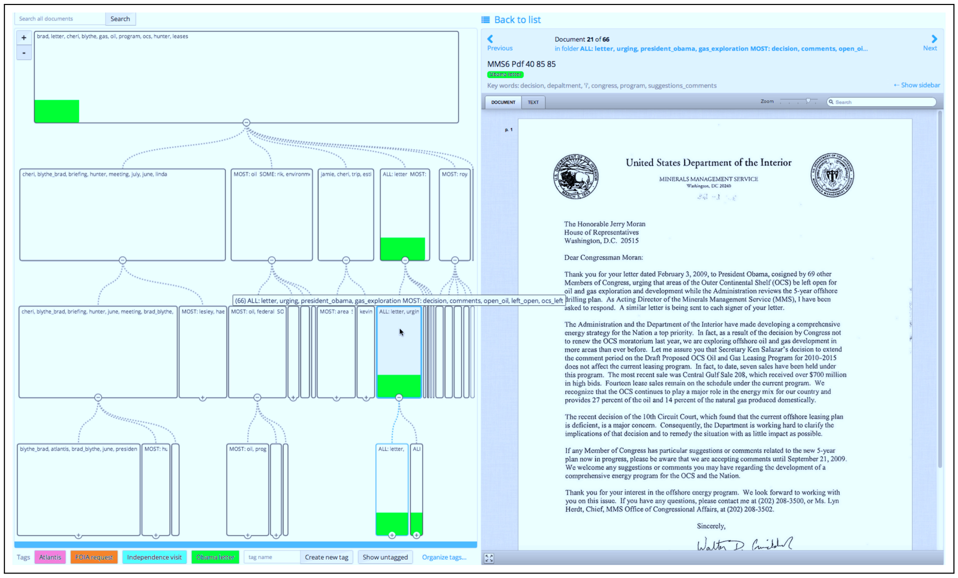


Figure 1.6: *Overview* is a multiple-view application intended for the systematic search, annotation, and reading of a large collection of text documents, which visualizes hierarchical clusters of documents as a tree (left).

ation of *Overview*, in which we analyzed its adoption and self-initiated use by investigative journalists. Chapter 4 documents this field study.

Methodology: We conducted case studies of six journalists who used *Overview* to conduct investigations involving large document collections; in five of these cases, the investigation resulted in a published story, and one of these stories [236] was a finalist for the 2014 Pulitzer Prize in journalism[6]. A critical difference between our approach and other post-deployment field studies that focus on the usage of visualization tools or techniques [195, 277, 292] is that our case study participants were not solicited by the researchers: they freely chose to use *Overview* and they did not inform preceding phases of design. We also engaged a different set of people at each stage of design, rather than the same set of people. This difference reflects *Overview*'s context of use: repeat usage cannot be predicted and *Overview* is only ap-

[6]`http://www.pulitzer.org/2014_public_service_finalist`

propriate for some investigations; we have yet to encounter a journalist who specializes in investigations pertaining to large document collections.

We interviewed these six journalists about the form and provenance of their documents, the objectives of their investigation, and their use of *Overview*; we also collected their logged interaction data and their annotated document collections. We used our task typology, which we introduce in Chapter 2, to better understand *why Overview* was adopted by these journalists to perform their investigations.

Results: The analysis of journalists' use of *Overview* revealed that our initial understanding of their task was insufficient: the task of *"exploring"* a document collection, a term that appears often in previous work on visualizing document data, is both too vague and too narrow to capture how journalists actually used *Overview*. Instead, we identified two different tasks using the vocabulary and structure of our typology: one of *generating* hypotheses and *summarizing* the contents of a document collection, and another of *locating* and *identifying* specific evidence in order to *verify* or *refute* prior hypotheses.

Contributions: Given our more precise understanding of journalists' tasks, we were able to rigorously analyze the rationale for *Overview*'s visual encoding and interaction design choices. This analysis is transferable beyond the domain of journalism and speaks to the design of visualization techniques and tools addressing document data and to some extent any data that can be hierarchically structured. Finally, we reflect upon *Overview*'s design and evaluation process, comparing our approach to previous human-centred visualization design processes [153, 195]; we also discuss the value, logistics, and limitations of studying the adoption of visualization techniques or tools.

1.3.4 Use of the Typology in a Design Study

In 2013, we initiated a visualization design study project that provided an opportunity to validate the *generative* potential of our typology. This project was a collaboration with a company that develops energy usage reporting software for multi-building organizations such as universities, school boards, or hotel chains. Many of these client organizations have designated energy analysts who oversee large portfolios of buildings; these analysts are responsible for identifying cost saving opportunities, diagnosing erratic energy usage behaviour, and attempting to understand the role of fluctuating external factors such as weather, occupancy, operating hours, and equipment usage within buildings. Tools and techniques for addressing these tasks with respect to single buildings already exist, however they do not scale to portfolios of dozens or hundreds of buildings. We conjectured that an interactive application integrating visualization while considering these issues of scale could address the tasks of these analysts. Chapter 5 documents this design study.

Methodology: We began by analyzing the energy domain and interviewing energy analysts from commercial client organizations who had previously used our industry partner's software, asking them about their roles and responsibilities, their technical background, their portfolio of buildings, and the limitations of current tools. We also presented our interview findings and sought additional feedback from members of our industry partner's client services team, who have expertise with the current software and act as liaisons to client organizations.

Once again, we used our task typology, to identify and abstract the tasks of these analysts. We narrowed our scope to tasks that recurred often among the analysts and those that were consistent with the mandate of our industry partner's product development team to support analysis of energy consumption in building portfolios.

Over the course of four months, we designed and implemented over a dozen interactive visualization *data sketches* [195] to address the tasks of these analysts, following a process of rapid iteration in which functional sketches featuring the analysts' data were used to further refine our understanding of their tasks and context of use. These data sketches were produced within the interactive interactive sandbox environment shown in Figure 1.7. Our task abstractions informed the process of mapping these tasks to a set of appropriate visual encoding and interaction design choices.

The design choices that we considered included those for performing multiple comparisons between aggregate and individual items over time [2], for identifying cyclic and acyclic events using meaningful temporal granularities [335], and for identifying differences in multiple lists of ranked items while simultaneously identifying the cause of rank changes [119].

In early 2014, we conducted *chauffeured demos* [195] of these interactive data sketches with four groups of analysts; the energy usage data used in these demos was collected from analysts' own building portfolios. By integrating the feedback we received on our sketches and our understanding of energy analysts' tasks, we then envisioned ways to juxtapose and sequence discrete *views* of the data in order to support workflows, and we continued to elicit feedback from analysts and our collaborators' client services team.

Results: Our collaborators have since adopted a number of our designs into a new version of their commercial energy analysis software tool. They assigned over ten full-time software developers to the project since mid-2014 and the tool has since been released to some client organizations in a small pilot deployment; the tool will soon[7] be deployed to thousands of other clients.

Contributions: As a result of abstracting the data and tasks relating to the energy management domain, visualization practitioners working in other domains might benefit from our classification of matches and mismatches between abstract tasks and visualization design choices, particularly for domains that involve comparing many concurrent time series.

We also confronted issues of domain convention in this project; in the energy sector, some visual encodings carry very specific meanings. We considered how to introduce unfamiliar visual encodings and how to get people working in this domain to trust them.

Finally, we contribute some methodological guidance for visualization design studies, including our approach to work domain analysis, a systematic task analysis and abstraction, our sandbox prototyping and workflow design, as well as how to effectively present visualization design documentation.

[7] Relative to November 2015.

Figure 1.7: A sandbox environment for creating visualization data sketches pertaining to the analysis of energy usage in large building portfolios. In this visualization data sketch, a calendar-based time series matrix is juxtaposed with summary boxplots, where each row is a group of buildings. Both the matrix and the boxplots encode the difference between average energy demand in 2012 and 2013.

1.3.5 Summary of Contributions

The contributions of this dissertation can be summarized as follows:

- A typology of abstract visualization tasks, which allows for succinct descriptions of tasks and task sequences in terms of *why* data is visualized, *what* dependencies a task might have in terms of *input* and *output*, and *how* the task is supported in terms of visual encoding and interaction design choices (Chapter 2).
- A synthesis of the literature relating to visualization tasks (Chapter 2).
- A datatype-specific classification of five task sequences relating to vi-

sualizing dimensionally reduced data, one based on findings from our interview study with data analysts spanning several application domains. This classification draws upon and demonstrates the descriptive power of our typology of tasks and is intended to inform the design of new tools and techniques for visualizing dimensionally reduced data (Chapter 3).

- A field study evaluation of *Overview*, a visualization tool for investigating large text document collections. We draw upon and demonstrate the descriptive and evaluative power of our typology of tasks and characterized two abstract tasks relating to document mining (Chapter 4).
- Seven lessons relating to the design of interactive visualization tools for hierarchical data and document data in particular. These lessons are based on an analysis of successive deployed versions of *Overview* and its adoption by self-initiated journalists (Chapter 4).
- A methodological reflection on the study of visualization adoption (Chapter 4).
- A demonstration of the descriptive, evaluative, and generative power of our typology of tasks in a visualization design study within the energy domain (Chapter 5).
- An identification of matches and mismatches between visualization design choices and three abstract tasks for concurrent time series data (Chapter 5).
- Two lessons pertaining to familiarity with visual encodings, two lessons pertaining to the trust of data aggregation design choices, and three lessons pertaining to visualization design methodology (Chapter 5).

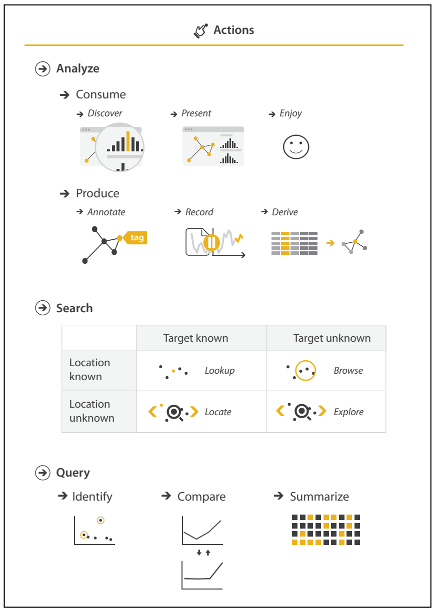
1.4 Extension and Impact of the Typology

In Chapter 6, we comment on how our task typology was subsequently extended by Munzner in her 2014 book *Visualization Analysis and Design* [219]; this modified typology is shown in Figure 1.8. Munzner moved *introduce* nodes from the *how* part of the typology to become forms of *produce* in the *why* part of the typology; she also added *targets* to the *why* part of the typology, referring to the original *why* part of our typology as *actions*; finally, she reorganized the *how* part of the typology and elaborated on forms of *encode*. We explicitly make reference to and use Munzner's modifications to the typology in our interview study (Chapter 3) and in our design study (Chapter 5). Chapter 6 also contains commentary on the origin, the benefits, and the potential drawbacks of Munzner's modifications.

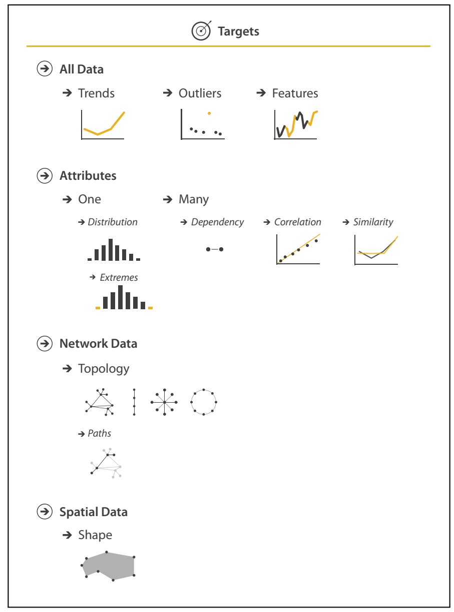
We also present a survey of how our task typology and our approach to systematically analyzing and abstracting tasks has been used and/or extended by others in the visualization community, including how our typology may integrate with novel theoretical frameworks. This survey includes the use of our typology to analyze domain-specific usage of visualization techniques or tools, from bioinformatics to malware analysis, as well as datatype-specific visualization usage, from geospatial data to multiplex networks. This survey also includes the use of our typology to specify and contextualize tasks in experimental studies, as well as the use of our typology to motivate the design of novel visualization techniques and tools.

1.5 A Note on Chronology

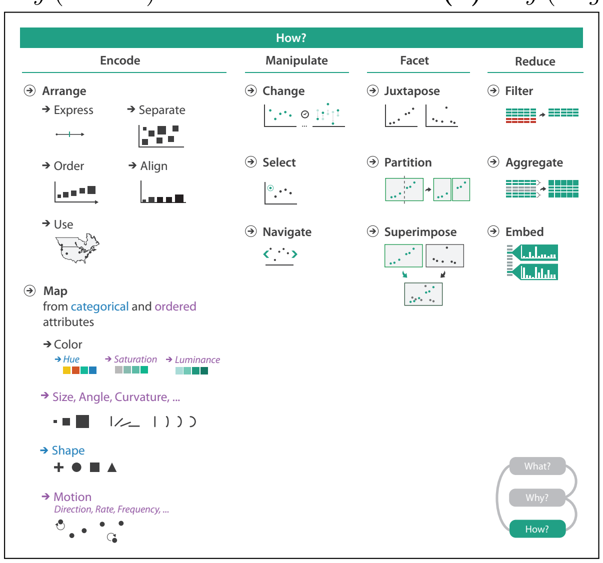
The duration of the projects described in this dissertation extended long periods of time. As a result, periods of focused research on these projects were interleaved or overlapping. Figure 1.9 illustrates the chronological history of these projects, indicating the core focus periods of projects, important milestones, as well as periods of part-time focus. Figure 1.9 also indicates other milestones in my PhD, research projects not included in this dissertation [34, 106], and internships.



(a) *why* (*actions*).



(b) *why* (*targets*).



(c) *how* (design choices).

Figure 1.8: A modified version of our typology appearing in Munzner [219] (c.f. Figure 1.2): (a) forms of *produce* were moved from *how*; (b) a new classification of *targets*; (c) a reorganization of *how*. Illustrations: 🅭 E. Maguire (2014).

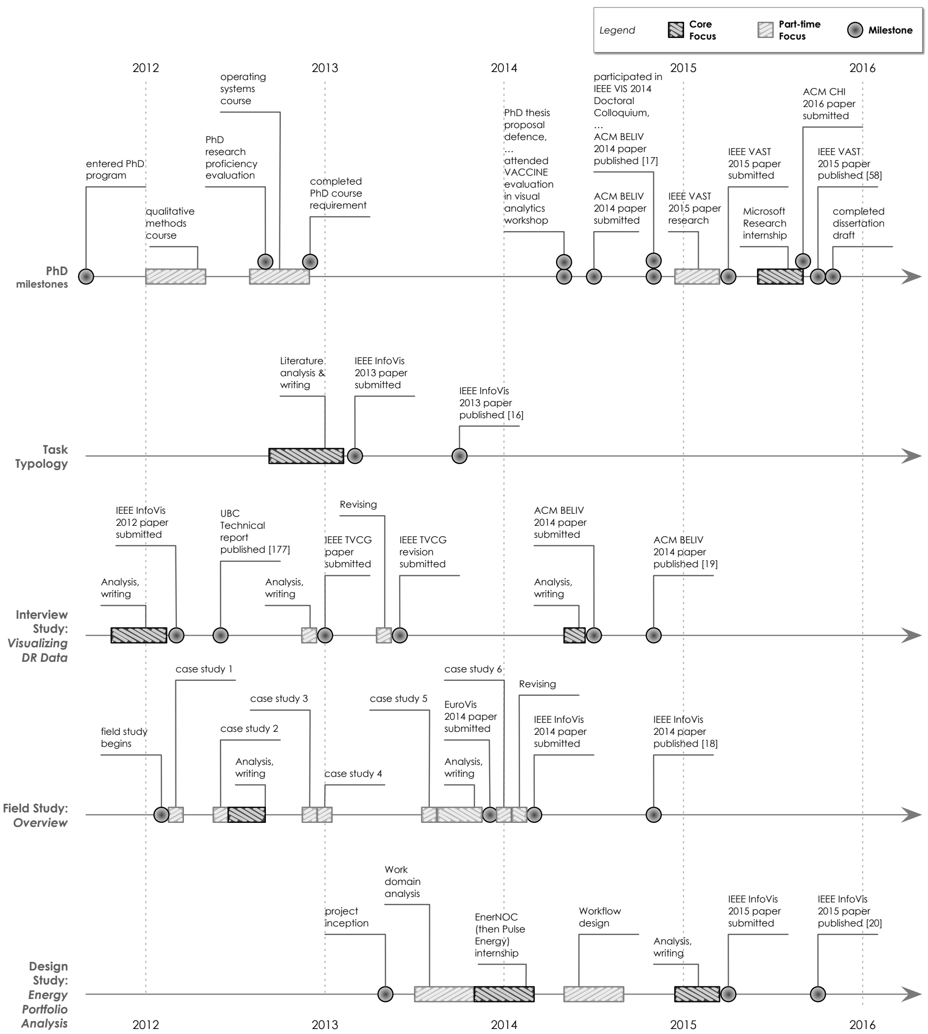


Figure 1.9: Timelines of the projects described in this dissertation. Note that the projects described in this dissertation overlapped in time. While the order of the chapters in this dissertation reflect the order in which the projects were completed, they do not reflect the order in which they were initiated.

Chapter 2

A Typology of Abstract Visualization Tasks

"By thinking about visualization as a process instead of an outcome, we arm ourselves with an incredibly powerful thinking tool." — Jer Thorp in "Visualization as process, not output" [316] (*Harvard Business Review*, April 3, 2013)

[1]The considerable previous work characterizing visualization processes has focused on low-level tasks or interactions and high-level tasks, leaving a gap between them that is not addressed[2]. This gap leads to a lack of distinction between the ends and means of a task, limiting the potential for rigorous analysis. We contribute a multi-level typology of visualization tasks to address this gap, distinguishing *why* and *how* a visualization task is performed, as well as *what* the task inputs and outputs are. Our typology allows complex tasks to be expressed as sequences of interdependent tasks, resulting in concise and flexible descriptions for tasks of varying complexity

[1]This chapter is a slightly modified version of our paper *A Multi-Level Typology of Abstract Visualization Tasks* by Matthew Brehmer and Tamara Munzner; in IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis 2013), 19(12), p. 2376–2385 [33]. `http://dx.doi.org/10.1109/TVCG.2013.124`.

[2]Referring to the examples cited in Section 1.1, *finding an extreme value* [8] is an example of a low level of abstraction while *exploring* [366] or *integrating insights* [301] are examples of a higher level of abstraction.

and scope. It provides abstract rather than domain-specific descriptions of tasks, so that useful comparisons can be made between visualization techniques or tools targeted at different application domains. This descriptive power supports a level of analysis required for the generation of new designs, by guiding the translation of domain-specific problems into abstract tasks, and for the qualitative evaluation of visualization tools or techniques. We demonstrate the benefits of our approach in a detailed example, comparing task descriptions from our typology to those derived from related work. We also discuss the similarities and differences between our typology and over two dozen existing classifications and theoretical frameworks from several research communities, including visualization, HCI, information retrieval, communications, and cartography.

2.1 Motivation

Consider a person who encounters a choropleth map while reading a blog post in the aftermath of an American presidential election. This particular map is static and visually encodes two attributes, candidate and margin of victory, encoded for each state using a bivariate colour mapping. This person decides to compare the election results of Texas to those of California, motivated not by an explicit need to generate or verify some hypothesis, nor by a need to present information to an audience, but rather by a casual interest in American politics and its two most populous states. How might we describe this person's *task* in an abstract rather than domain-specific way?

According to Munzner's nested model for visualization design and validation [217], abstract tasks are domain- and interface-agnostic operations that people perform. Disappointingly, there is little agreement as to the appropriate granularity of an abstract task among the many existing classifications in the visualization, HCI, cartography, and information retrieval literature [7, 8, 12, 42, 48, 51, 58, 60, 75, 117, 130, 166, 175, 186, 193, 216, 239, 242, 252, 260, 262, 263, 291, 298, 301, 329, 330, 343, 349, 366, 370]. One of the more frequently cited of these [8] would classify the above example as being a series of value retrieval tasks. This low-level characterization does not describe the person's context or motivation; nor does take into account prior experience and background knowledge. For instance, a description of this task might differ if the person was unfamiliar with American geography: the person must locate and identify these states before comparing their values. Conversely, high-level descriptions of exploratory data analysis and presentation emanating from the sensemaking literature [7, 48, 175, 242] cannot aptly describe this person's task.

The gap between low-level and high-level classification leaves us unable to abstractly describe tasks in a useful way, even for the simple static choropleth map in the above example. This gap widens when interactive visualization techniques are considered, and the complexity of its usage is compounded over time. We must move beyond describing a single task in isolation, to a description that designates when one task ends and another begins. To close this gap, visualization tasks must be describable in an abstract way across multiple levels.

The primary contribution of this chapter is a multi-level typology of abstract visualization tasks that unites the previously disconnected scopes of low-level and high-level classifications by proposing multiple levels of linkage between them. Our typology provides a powerful and flexible way to describe complex tasks as a sequence of interdependent simpler ones. While this typology is very much informed by previous work, it is also the result of new thinking and has many points of divergence with existing models. Central to the organization of our typology are three questions that serve to disambiguate the means and ends of a task: *why* data is being visualized, *how* the visualization technique or tool supports the task, and *what* are the task's inputs and outputs. We have found that no prior characterization of tasks satisfactorily answers all of these questions simultaneously at multiple levels of abstraction. Typically, low-level classifications provide a sense of *how* a task is performed, but not *why*; high-level classifications are the converse. One major advantage of our typology over prior work is in providing linkage between these two questions. Another advantage is the ability to link sequences of tasks, made possible by the consideration of *what* tasks operate on.

Our typology provides a consistent lexicon for description that supports making precise comparisons of tasks between different visualization tools and across application domains. Succinct and abstract descriptions of tasks are crucial for analysis of people using visualization tools and techniques. This analysis is an essential precursor to the effective design and evaluation of visualization tools, particularly in the context of problem-driven design studies [284]. In these studies, visualization practitioners work with people from specific application domains to determine *why* and *what*, subsequently drawing from their specialized knowledge of visual encoding and interaction design choices as well as known human capabilities with respect to perception [62, 254] and interaction to envision *how* that task is to be supported. A need for task analysis also arises in visualization evaluation [183], particularly in observational studies of people using visualization tools and techniques. Our typology provides a code set for qualitatively describing the behaviour of participants in such studies.

2.2 Background Context

As we expect some readers to be unfamiliar with the context that motivated this work, we begin with a brief discussion of our current inability to succinctly describe and analyze visualization tasks. The primary limiting factor in using existing classifications as tools for analysis is that we cannot easily distinguish between the ends and means of tasks. Making this distinction is a central problem for practitioners during the *abstraction* phase of design studies [284] and during the analysis phase of qualitative studies of people using visualization tools or techniques [183].

For instance, a number of existing classifications mention the word *derive* [8, 60, 130, 186, 239, 301]. Is *derive* a task, or the means by which another task is performed? A person may derive data items as an end in itself, for example to reduce the number of dimensions in a dataset, or as a means towards another end, such as to verify a hypothesis regarding the existence of clusters in a derived low-dimensional space. The ends-means ambiguity exists for many terms found in existing classifications: consider *filter* [8, 48, 117, 130, 166, 175, 186, 216, 239, 242, 260, 262, 291, 366], *navigate* [130, 298, 343], or *record* [130, 216, 301]. The first step towards distinguishing ends from means involves asking *why* someone would visualize data separately from *how* the visualization tool or technique supports the task, a question that is central to the organization of our typology.

The separation of *why* and *how* does not in itself resolve all confusion. Consider *sort*, another term appearing in existing classifications [8, 117, 130, 186, 239]. Sorting has an input and an output; in some cases, it is items of data within a single view [251]; in others, views themselves may be sorted [17]. In both cases, the sorted output can serve as input to subsequent tasks. The next step in distinguishing ends from means is thus characterizing *what* the task's inputs and outputs are, allowing us to describe sequences of interdependent tasks.

To illustrate how the ends-means ambiguity arises during the course of analysis, we will now attempt to use representative existing classification to describe two example tasks:

Example #1: recall the example stated above in Section 2.1, that of a casual encounter with an electoral map in which a person compares two regions; election results for each state are encoded as a choropleth map based on two attributes, candidate and margin of victory. Furthermore, we know that this person is familiar with American geography and its regions; this prior knowledge dictates the type of search.

Using the typology of Andrienko and Andrienko [12], we might describe this example as an *elementary direct comparison task*. While richer than a series of *retrieve value* tasks [8], this description tells us little about *why* and *how* this comparison was performed. Low-level descriptions derived from a number of other classifications are similarly impoverished [51, 117, 263, 330, 349, 366, 370].

We might enrich our description of this task using a recent *taxonomy of cartographic interaction primitives* by Roth [260, 262], a much more comprehensive approach that distinguishes between *goals*, *objectives*, *operators*, and *operands*. Using his taxonomy, this task would be described as follows:

- **goals**: *procure*
- **objectives**: *compare*
- **operators**: *retrieve* and *calculate*
- **operands**: *attribute–in–space* (search target); *general* (search level)

While the dimensions of this description are similar to the questions of *why*, *how*, and *what*, the description is incomplete, particularly in its classification of *goals* and *objectives*. Roth's taxonomy provides us only with a partial sense of *how* the comparison is performed: *retrieve* does not tells us about whether the person knows the spatial location of the regions to be compared a priori. The goal, *procure*, does not provide us with any higher-level context or motivation for *why* the person is *procuring*; specifically, the person's casual interest in these two regions is lost. Finally, Roth's taxonomy imposes a spatial constraint on *operands*, leaving us unable to fully articulate *what* is being compared.

Example #2: in evaluation studies [183], it is sometimes necessary to perform a comparative analysis of a task being performed using different visualization tools or techniques. Consider a person using a tree visualization tool whose interest relates to two nodes in a large tree, and her intent is to present the path between these nodes to her colleagues. SpaceTree [122] and TreeJuxtaposer [220] are two tree visualization tools that allow people to locate paths between nodes by means of different focus + context techniques. Both tools allow for path selection, in which the encoding of selected paths differs from that of non-selected paths. The tools differ in *how* the elements that have been visualized are manipulated: TreeJuxtaposer allows a person to arrange areas of the tree to ensure visibility for areas of interest, while SpaceTree couples the act of selection by aggregating and filtering unselected items.

As in the previous example, task descriptions from existing classifications seldom answer all three questions: *why*, *how*, and *what*. Using the *taxonomy of interactive dynamics for visual analysis* by Heer and Shneiderman [130], we might describe this task as being an instance of *data and view specification* (*visualize* and *filter*) as well as *view manipulation* (*navigate* and *select*). This description tells us *how*, but it doesn't specify *why* the data is being visualized.

We might complement Heer and Shneiderman's description with one based on a taxonomy of graph visualization tasks by Lee et al. [186], in which this task would be classified as a *topology task*, namely one of *determining connectivity* and subsequently *finding the shortest path*. As the scope of Lee et al.'s taxonomy is specialized, we are provided with a clear indication of *what* the person's interest is, this being a *path*. Unfortunately, this description provides only a partial account of *why* data is being visualized; we are not provided with a high-level motivation beyond *determining* and *finding*.

Both descriptions do not relate the person's actions to the high-level goal of *presenting* information to others. Second, and more importantly, these descriptions fail to distinguish *how* this task is performed using SpaceTree from *how* it is performed using TreeJuxtaposer.

Summary: these examples demonstrate our inability to comprehensively analyze tasks using existing classifications of behaviour of people who use visualization tools or techniques. Note that we are not directly criticizing these classifications; we acknowledge that their scope is often deliberately constrained, with some focusing on low-level tasks, interactions, or operations [8, 12, 42, 51, 58, 60, 75, 117, 166, 186, 263, 291, 329, 330, 343, 349, 366, 370], while others focus on high-level tasks or goals [7, 48, 175, 193, 242], or on the behaviour of people who work in specific domains or contexts [186, 260, 262, 299]. We lack guidance on how to integrate these disjoint bodies of work, to compose task descriptions that draw from all of them. This integration is the aim of our typology, which will allow practitioners to describe tasks that address critical questions posed during visualization design and evaluation, namely *why*, *how*, and *what*.

It could be argued that a classification of *tasks* should focus solely on

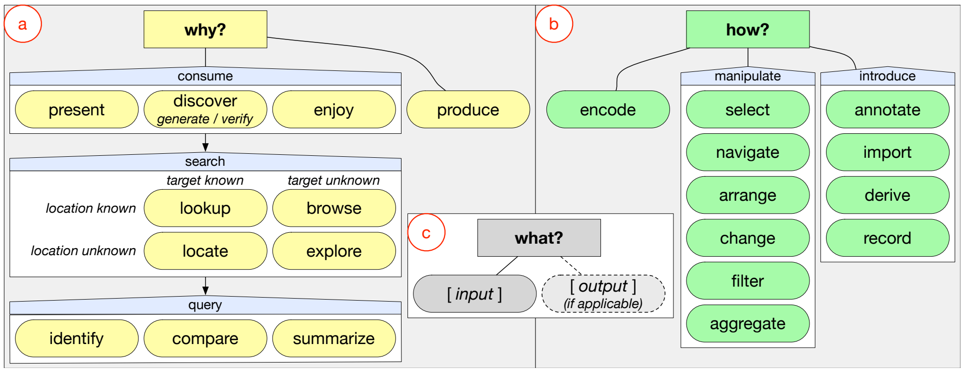


Figure 2.1: Our multi-level typology of abstract visualization tasks. The typology spans *why*, *how*, and *what*; task descriptions are formed by nodes from each part: (a) *why* data is visualized, from high-level (`consume` vs. `produce`) to mid-level (`search`) to low-level (`query`); (b) *how* a visualization tool or technique supports the task in terms of visual encoding and interaction design choices; (c) *what* the task `inputs` and `outputs` are.

the goal of the person who uses the visualization tool or technique, or *why* data is visualized; people are often not immediately concerned with *how* a task is performed, as long as their task can be accomplished. We argue that by classifying tasks according to *how* they are performed, *in addition* to *why* they are performed and *what* they pertain to, we can improve communication between visualization practitioners working in different domains, facilitating tool-independent comparisons, the analysis of diverging usage strategies for executing tasks [337, 371], and improved reasoning about design alternatives.

2.3 A Typology of Tasks

Our multi-level typology[3] of abstract visualization tasks, represented in Figure 2.1, is encapsulated by three questions: *why* the data is being visualized,

[3] We denote this work as a *typology*, rather than a *taxonomy*, as the former is appropriate for classifying abstract concepts, while the latter is appropriate for classifying empirically observable events [13]. For instance, one could construct a taxonomy of the observable ways in which a person could interact with a particular visualization tool, or a taxonomy of existing visual encoding design choices for tree-based data [279].

how the visualization tool or technique supports the task, and *what* does the task pertain to (Figure 2.1a-c). Complete task descriptions, such as those for Examples #1-2 (represented in Figure 2.2), must include nodes from all three parts of this typology. In the remainder of this dissertation, we use a `fixed-width font` to highlight vocabulary from this typology; this vocabulary is also indexed separately at the end of this dissertation.

This structure, while unusual relative to many existing classifications, mirrors the analytical thinking process undertaken in design studies [284]. *Why*, *what*, and *how* are also used in *Cognitive Work Analysis* by Vicente [337], particularly for relating abstractions within a work domain, as well as in an analysis of techniques and tools for visualizing time-oriented data by Aigner et al. [2], which asks *what is presented?*, *why is it presented?*, and *how is it presented?*.

We will introduce *why* before *how*, as this order reflects the translation of empirically observable domain problems into abstract tasks and subsequently into visual encoding and interaction design choices: practitioners first determine *why* data is visualized, and then must decide upon *how* the visualization tool or technique will support the task. We then discuss the *what* part of our typology, which considers the `input` and `output` of tasks. Our typology supports the description of a complex domain-specific visualization workflow as a sequence of interdependent tasks, where the `output` of a prior task may serve as the `input` to a subsequent task, as we demonstrate in the example featured in Section 2.4.

For clarity, we first present our typology in its entirety with minimal discussion of the previous work that informed its organization, and then focus on these connections in Section 2.5 and in Table 2.1 and Table 2.2.

2.3.1 Why is Data Being Visualized?

The *why* part of our typology, shown in Figure 2.1a, allows us to describe *why* the data is being visualized, and includes multiple levels of abstraction, a narrowing of scope from high-level (`consume` vs. `produce`) to mid-level (`search`) to low-level (`query`).

Consume: People visualize data in order to `consume` information in many domain contexts. In most cases, this consumption is driven either by a need to present information or to discover and analyze new information [334]. However, there are many other contexts in which the information being visualized is simply enjoyed [78, 247, 299], where people indulge their casual interests in a topic.

`Present` refers to the visualization of data for the succinct communication of information, for telling a story with data, guiding an audience through a series of cognitive operations. Presentation using a visualization technique or tool may take place within the context of decision making, planning, forecasting, and instructional processes [102, 199, 260, 262]. Presentation brings to mind collaborative and pedagogical contexts, and the way in which a presentation is given may vary according to the size of the audience, whether the presentation is live or pre-recorded, and whether the audience is co-located with the presenter [177].

`Discover` is about the generation and verification of hypotheses and is associated with modes of scientific inquiry [239]. Scientific investigation may be motivated by existing theories, models, and hypotheses, or by the serendipitous observation of unexpected phenomena [9].

`Enjoy` refers to casual encounters with visualized data [247, 299]. In these contexts, a person is not driven by a need to verify or generate a hypothesis; novelty stimulates curiosity and thereby exploration [77, 299, 303, 318]. This motivation is notably absent from existing classifications, as shown in Table 2.1 and Table 2.2. Casual encounters with visualized data can be fleeting, such as in the earlier example of encountering a static choropleth electoral map while reading a blog post. Conversely, these encounters might be immersive and time-consuming experiences, such as in museum settings [247].

Produce: we use `produce` in reference to tasks in which the intent is to generate new information. This information includes but is not limited to: transformed or derived data, annotations, recorded interactions, or screenshots of static visualizations. Examples of `produce` in previous work include the production of graphical annotations and explanatory notes to describe features of line graphs of time series data [355], or the production of *graphical histories* in Tableau intended to document the analytical provenance of a person using this tool [131]. Additional examples of `produce` involving derived data and annotations are featured in the example of Section 2.4.

It is important to note that the products of a `produce` task may be used in some subsequent task that may or may not involve a visualization tool or technique. For example, some visualization tools for analyzing high-dimensional data allow people to `produce` new categorical attributes for labelling clustered data points in a dimensionally reduced coordinate space; these attributes might be used later for constructing a predictive model.

Search: Regardless of whether the intent is to `present`, `discover`, or merely `enjoy`, a person will `search` for aspects of interest in the visualized data. While terms relating to `search` and exploration are often conflated [199, 318], we have imposed a characterization of `search` that depends on *what* is being sought. We classify them according to whether the identity or location of the search target is known a priori. Whether the identity of the search target is known recalls the concept of *references* and *characteristics* introduced by Andrienko and Andrienko [12]: `searching` for known reference targets entails `lookup` or `locate`, while `searching` for targets matching particular characteristics entails `browse` or `explore`. Consider our earlier example of a person who is familiar with American geography and is `searching` for California on an choropleth map; we would describe this as an instance of `lookup`. However, a person who is *unfamiliar* with American geography must `locate` California.

In contrast, the identity of a search target might be unknown a priori; a person may be `searching` for characteristics rather than references [12]; these characteristics might include particular values, extremum, anomalies, trends, or ranges [8]. For instance, if a person using a tree-based visual encoding is `searching` within a particular subtree for leaf nodes having few siblings, we would describe this as an instance of `browse` because the location is known a priori. Finally, `explore` entails `searching` for characteristics without regard to their location; many visualization tools provide an overview of the data, which is often the starting point for exploration. Examples include `searching` for outliers in a scatterplot, for anomalous spikes or periodic patterns in a line graph of time series data, or for unanticipated spatially-dependent patterns in a choropleth map.

Query: Once a target or set of targets has been found, a person will `identify`, `compare`, or `summarize` these targets. If a search returns known or *reference* targets [12], either by `lookup` or `locate`, `identify` returns their *characteristics*. For example, someone who uses a choropleth map representing election results can `identify` the winning candidate and margin of victory for the state of California. Conversely, if a search returns targets matching particular *characteristics*, either by `browse` or `explore`, `identify` returns *references*. For instance, our election map enthusiast can `identify` the state having the highest margin of victory.

The progression from `identify` to `compare` to `summarize` corresponds to an increase in the amount of search targets under consideration [12, 42, 329], in that `identify` refers to a single target, `compare` refers to multiple subsets of targets, and `summarize` refers to a whole set of targets. As with `explore`, `summarize` is also often associated with overviews of the data [186]. Continuing with the choropleth map example, the person `identifies` the election results for one state, `compares` the election results of one state to another, or `summarizes` the election results across all states, determining how many favoured one candidate or the other, or the overall distribution of margin of victory values.

2.3.2 How Does the Visualization Technique or Tool Support the Task?

We now turn our consideration to the *how* part of our typology, which contains *idioms*, defined as families of related visual encoding and interaction design choices. This part of our typology, shown in Figure 2.1b, is likely to be most familiar to readers, as it contains a number of *idioms* associated with interaction design choices that are well-represented by several existing classifications [117, 216, 260, 262, 366]. We distinguish between three classes of idioms: those for `encoding` data, those for `manipulating` previously encoded elements, and those for `introducing` new elements.

Encode: The majority of visualization tasks rely on *how* data is initially encoded as a visual representation[4]. A full enumeration of visual encoding design choices for various datatypes beyond the scope of this chapter and appears in Munzner's book [219].

Manipulate: The following idioms affect previously encoded elements, modifying them to some extent. These idioms represent families of interrelated design choices incorporating both interaction and visual encoding. We consider visual encoding and interaction design choices in a unified way because many idioms incorporate aspects of both [211, 217], such as focus + context techniques [122, 220].

`Select` refers to the demarcation of one or more encoded elements, differentiating selected from unselected elements [252]. Examples range from directly clicking or lassoing elements in a scatterplot to brushing design choices used to highlight elements in visualization tools incorporating multiple linked views [348].

`Navigate` refers to instances where the person using a visualization tool or technique alters their viewpoint, such as zooming, panning, and rotating. Other navigation instances include the triggering of details-on-demand views, combining `navigate` and `select` [291].

`Arrange` refers to the process of organizing encoded elements spatially. This includes arranging representations of data [193, 216, 353], such as reordering the axes in a parallel coordinates plot or the rows and columns of a scatterplot matrix (SPLOM). Other forms of arrangement allow people to coordinate the spatial layout of views [130, 348].

`Change` pertains to alterations in visual encoding. Simple examples include altering the size and transparency of points in a scatterplot or edges in a node-link graph, altering a colour-scale or texture mapping, or trans-

[4]Some tasks do not depend on *how* the data is visually encoded, or take place before the data is encoded; consider, for instance, `produce` tasks that involve `deriving` new data or `recording` states of a visual analysis process or presentation for downstream consumption.

forming the scales of axes. Other alterations have more pronounced effects, changing the visual encoding, such as transitioning between grouped and stacked bar charts, or between linear and radial layouts for line graphs of time series data. Pronounced changes in visual encoding such as these are often facilitated by smoothly animated transitions, which reduce their disruptive effects [129].

`Filter` refers to adjustments to the exclusion and inclusion criteria for encoded elements. Some forms of filtering allow for elements to be temporarily hidden from view and later restored, while other forms are synonymous with outright deletion. As an example of temporary `filtering`, consider a person examining an age histogram based on population census data. First, she decides to exclude males, then further adjusts her filter criteria to focus solely on unemployed females. Finally, she revises the gender criteria to focus on unemployed males.

A common example of permanent `filtering`, or deletion, is that of manually `selecting` and removing outliers resulting from errors in data entry. Alternatively, consider a scatterplot in which some data points are labelled with manually generated categorical tags. Deleting a tag would remove this categorical label from all data points having that tag.

`Aggregate` concerns changes in the granularity of encoded elements; we also consider its converse, segregate, as being associated with this family of design choices. For example, a person may adjust the granularity of a continuous time scale in a line graph, aggregating daily values into monthly values, or segregating annual values into quarterly values. Alternatively, a person may aggregate a clique within a node-link graph into a representative glyph, or segregate clique glyphs into their component nodes.

Introduce: While `manipulate` idioms alter previously encoded elements, `introduce` idioms add new elements.

`Annotate` refers to the addition of graphical or textual annotations associated with one or more encoded elements. When an annotation is associated with data elements, an annotation could be thought of as a new attribute for these elements. The earlier example of manually tagging points in a scatterplot with categorical labels is one such instance of annotating data.

`Import` pertains to the addition of new data elements. In some environments, these new data elements might be loaded from external sources, while others might be manually generated.

`Derive` refers to the computation of new data elements given existing data elements. Aggregating data often implies deriving data, however this may not always be true: we further specify that derived data must be persistent, while aggregated data need not be. For instance, a person might `derive` new attributes for tabular data using a multi-dimensional scaling (MDS) algorithm.

Finally, `record` refers to the saving or capturing of elements as persistent artefacts. As a consequence, `record` is often associated with `produce`. These artefacts include screen shots, annotations, lists of bookmarked elements or locations, parameter settings, or interaction logs [293]. An interesting example of `record` is that of assembling a *graphical history* [131], in which the `output` of each task includes a static snapshot of the state of the visualization tool, and as these snapshots accumulate they are encoded as a branching tree. `Recording` and retaining artefacts such as these are often desirable for maintaining a sense of analytical provenance, allowing people who use the tool to revisit earlier states or parameter settings.

2.3.3 What are the Inputs and Outputs of the Task?

Previous work has reached no agreement on the question of *what* is visualized. Many classifications do not address it at all; others discuss *what* implicitly, as indicated by the parenthetical terms in Table 2.1 and Table 2.2. Of those that classify *what*, some focus on the level of the entire dataset, such as *tables* composed of *values* and *attributes* or *networks* composed of *nodes* and *links* [291]. Others allow more precise specification of data-attribute semantics, such as *categorical*, *ordinal*, and *quantitative* [48]. A few classifications include not only *data* but also *views* as first-class citizens [58, 60, 130, 343]. Specific examples of *what* as classified in previous work include:

- Values, extremum, ranges, distributions, anomalies, clusters, correlations [8].

- *Graph-specific objects* [186]: nodes, links, paths, graphs, connected components, clusters, groups.

- *Time-oriented primitives* [2]: points, intervals, spans, temporal patterns, rates of change, sequences, synchronization.

- *Interaction operands* [343]: pixels, data [values, structures], attributes, geometric [objects, surfaces], visualization structures.

In this typology, we have chosen a flexible and agnostic representation of *what* that accommodates all of these modes of thinking: in short, we have a "bring your own *what*" mentality. The only absolute requirement is to explicitly distinguish a task's `input` and `output` constraints when describing sequences of interdependent tasks [329]. An extensive discussion of *what* that dovetails well with this typology appears in Munzner's book [219][5].

2.3.4 Concise Task Descriptions

Our multi-level typology can be used to concisely describe visualization tasks. Each task is defined by *why* data is being visualized, *how* the visualization tool or technique supports the task, and by *what* are the `inputs` and `outputs` of the task. Single tasks may involve multiple nodes from each part of the typology, as shown in Figure 2.2.

We have chosen to present these descriptions using a simple and flexible visual notation, rather than with a formal grammar [12, 185, 280, 329, 353]; in doing so, creating and iterating on task descriptions can be easily integrated into existing collaborative design and ideation activities, making use of materials such as coloured sticky notes and whiteboards. A crucial aspect of these descriptions is that sequences of interdependent tasks can be chained together, such that the `output` from earlier tasks forms the `input`

[5]Munzner [219] provides a structured classification of data as well as a classification of *targets*; both can be used in the analysis of `inputs` and `outputs`. The classification of *targets* is represented in Figure 6.2.

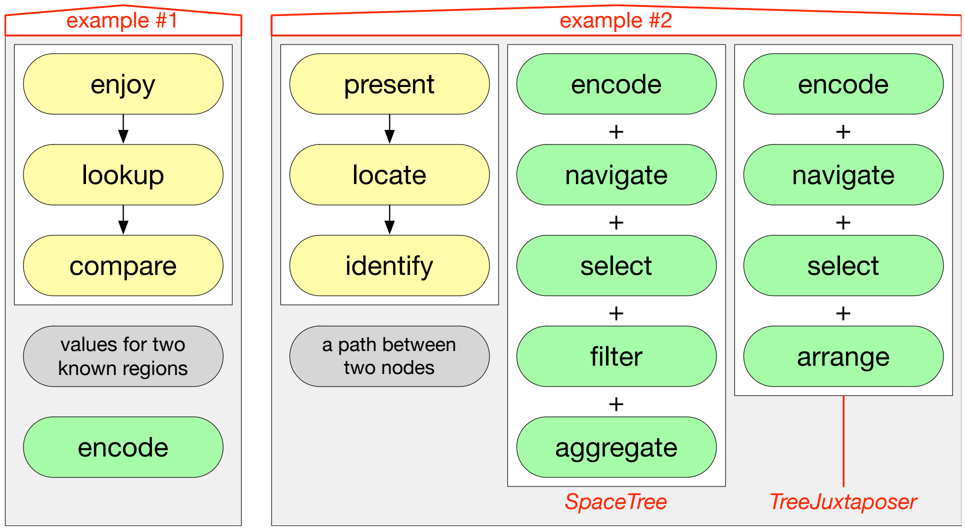


Figure 2.2: Task descriptions for Example #1 (left): casually encountering an choropleth electoral map and comparing election results for two regions; and Example #2 (right): presenting a path between two nodes in a large tree using SpaceTree [122] and TreeJuxtaposer [220].

to later tasks, as discussed in the following example and as represented in Figure 2.3.

2.4 Example: A Sequence of Interdependent Tasks

Visualization tasks are seldom executed in isolation, and the `output` of one task may serve as `input` to a subsequent task. To illustrate this type of dependency, we present an example in which our typology is used to describe a sequence of interdependent tasks.

Consider the case of labelling clusters of related items in a dataset with many dimensions, where a label is a new categorical attribute value for the item. Labels are assigned to clusters by means of `annotation`. However, one must first `explore` the visualized dataset and `identify` clusters of interest. Here a person uses a visualization technique in which items in the dataset

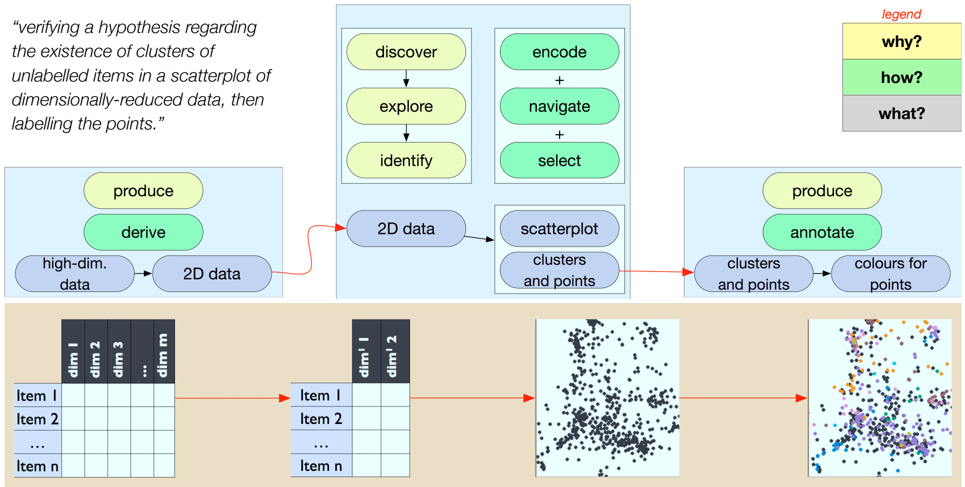


Figure 2.3: An example sequence of tasks, described as a sentence (top left), as well as using the structure and vocabulary of our typology (top); the bottom depicts a series of transformations corresponding to the `inputs` and `outputs` of each task.

are `encoded` as points in an interactive scatterplot. `Identifying` clusters is facilitated by `navigating` and `selecting` items in this scatterplot; upon `selection`, additional attribute values for the item are shown in details-on-demand secondary displays or in tooltips. This task too has a dependency on the result of an earlier task. Before the the data is `encoded` in the scatterplot, a set of two-dimensional distances between data points must be `produced`: they are `derived` from the original set of dimensional attributes using DR.

Using our typology, we can express dependencies in which the `output` of one task serves as the `input` of another, such as the relationship between how data is `derived` and the choice of visual encoding. Such dependencies are represented in Figure 2.3.

As in the examples of Section 2.2, we can compare our description to those generated by other classifications. Consider the classification of basic visualization tasks by Chuah and Roth [60], which distinguishes between three categories of *operations*. Using this classification, this sequence of tasks could be described as having *data operations* (*derived attributes*), *graphical operations* (*encode data*), and *set operations* (*create set, express membership*). This description does specify *how* and *what*, but it does not express the interdependencies within a sequence of tasks, nor does it tell us *why* the data is being visualized. Neither can we easily distinguish when sets (or clusters) are created, as this might occur before the data is encoded, or it might occur via interactive `selection` of items in the scatterplot.

While the description based on Chuah and Roth [60] classification is atemporal, the *operator interaction framework* by Chi and Riedl [58] defines stage- and transformation- based operators occurring along the visualization pipeline. Their framework does not contain a comprehensive list of operators, so we draw from the example *operators* cited in their paper to describe this sequence of tasks as follows:

1. *visualization transformation operators*:
   dimension reduction (DR)

2. *visual mapping transformation operators*:
   scatterplot

3. *view stage operators*:
   zoom, focus, details-on-demand, pick

This description does capture the interdependencies for this sequence of tasks, though it mischaracterizes the processes of dimension reduction as transformations to visualized elements, rather than transformations on data, a distinction central to our definition of `derive`. While this description captures up until the second task in our description, it does not capture the final task of producing cluster labels by means of annotation.

The description based on our typology retains the separability of these tasks, ensuring the distinction between interim `inputs` and `outputs`. Another problem with descriptions generated by existing classifications was that of coverage; the *how* part of our typology includes both `derive` and `annotate`, while descriptions generated by other classifications could not account for the latter [8, 58], or both [239, 291, 330, 349, 366, 370]. Finally, our description also accounts for both *why* data is `derived` and *why* clusters are `annotated` with tags, whereas descriptions generated using existing classifications mention *how* a task is performed in relation only to *when* it is performed [58] or to *what* it is performed on [60]. We maintain that a task description requires *why*, *how*, and *what*; the question of *when* for a sequence of interdependent tasks is best served by denoting task `input` and `output`.

2.5 Connections to Previous Work

Our typology was informed in part by related work, including existing classifications and established theoretical models, and in part by new thinking with many points of divergence from previous work. We surveyed work relating to tasks spanning the research literature in visualization, visual analytics, HCI, cartography, and information retrieval. We focus on two subsets that informed the configuration of our typology: thirty works that explicitly contribute a taxonomy, typology, characterization, framework, or model of tasks, goals, objectives, intentions, activities, or interactions [7, 8, 12, 42, 48, 51, 58, 60, 75, 117, 130, 166, 175, 186, 193, 216, 239, 242, 252, 260, 262, 263, 291, 298, 301, 329, 330, 343, 349, 366, 370], along with twenty other references that make compelling or noteworthy assertions about the behaviour of people who use visualization tools or techniques [2, 9, 77, 78, 102, 163, 199, 218, 244, 247, 261, 293, 299, 303, 318, 321, 327, 334, 345, 353][6]. The similarities between the individual nodes of our typology and those of existing classifications and other related work are presented in detail in Table 2.1 and Table 2.2[7].

Table 2.1 and Table 2.2 serve three purposes: they document our choices of terms for the purpose of reproducibility, they illustrate the influence of previous work on our thinking, and they indicate overrepresented and underrepresented areas in the literature, such as `consume` and `enjoy` in the *why* part of our typology. Note that non-leaf nodes in the *how* part of our

[6]Appendix A includes additional meta-analysis of this literature and documents the evolution of our typology.

[7]Yalçın [364] has visualized the vocabulary from previous work represented in these tables here: `http://keshif.me/demo/vis_tasks.html`.

typology are poorly represented, serving to indicate the gap between low and high levels of abstraction.

2.5.1 Existing Classifications

The scope of existing classifications can be categorized in three different ways: level of abstraction, temporality, and applicability.

Level of abstraction: Much of previous work can be divided into those having a low or high level of abstraction, with very little falling in between. Relying solely on either type of classification leads to the aforementioned ends-means confusion, thereby limiting the potential for rigorous analysis. Low-level use of a visualization technique or tool is well represented in related work [8, 12, 42, 51, 58, 60, 75, 117, 166, 186, 263, 291, 329, 330, 343, 349, 366, 370]. Elements common to many of these classifications include *select*, *filter*, and *navigate*. Following Lee et al. [187] and Roth [262], we note that low-level classifications of tasks are often conflated with those of interaction design choices. At the high level, abstract tasks can be found in the context of theoretical models, but without explicit connections to low-level visualization tasks [7, 48, 175, 193, 242]. Examples of these include *confirm hypotheses*, *present*, and *explore*. Low-level classifications often provide a sense of *how* a task is performed, but not *why*; high-level models are the converse. Our focus on multi-level descriptions of visualization tasks is intended to close this gap and resolve the ends-means confusion.

Temporality: Most classifications are atemporal in that they do not have any way to express sequences or dependencies between different stages. A few classifications are explicitly temporal in that they divide the behavior of people who use visualization tools or techniques into larger stages that occur in specific sequences or cycles. Examples include pipeline models for visualization construction [58] or data analysis [163], or cyclic models such as knowledge crystallization [48] or information foraging and sensemaking [242]. However, empirical observations of the use of visualization tools and techniques have indicated a mismatch between the specific cyclical or sequential patterns proposed by these models and the actual behaviour of people [152].

why?	
`consume`	–
→ `present`	*present*, [293, 334], *author, compose* [48]*, *build (case), tell (story)* [242]*, *depict* [239]*, *express (ideas), describe* [301]*, *guide, share* [130]* *inform, elaborate* [370]*, *report* [163],
→ `discover (generate, verify hypotheses)`	*discover* [199], *explore* [370]*, [334], *verify* [51]*, [199], *synthesize* [216]*, [199], *investigate, integration (of insight)* [301]*, [199], *frame operations: construct, elaborate, question, reframe* [175]*, *assimilate, assess, understand* [239]*, *infer* [330]*, *analyze* [216, 239]*, [199], *support, reevaluate (hypotheses)* [242]*, *monitoring* [345], *confirm (hypotheses), expose (uncertainty), formulate (cause and effect), concretize (relationships), learn (domain parameters), multivariate explanation* [7]*, *evaluate, learn, investigate* [199], *open-ended exploration, diagnosis* [244], *abduction, deduction, induction* [239], *generate, confirm (hypotheses)* [9, 102], *integrate, interpret* [102], *exploratory and confirmatory data analysis* [327]
→ `enjoy`	(using visualized data in casual contexts) [247, 299], *strolling* [78]
`produce`	*export* [260, 262]*, *store* [216]*, *save* [193, 260, 262]*, *extract* [48, 291]*, *generating (images)* [244], (a *classification*) [48, 301]*, (a *categorization*) [216, 349, 370]*, (a *record* of one's history / process) [130, 291, 301]*
`search`	*search* [48, 242, 370]*, *acquire* [216]*, *visual queries* [345]
→ `lookup`	*lookup* [51]* [199], *identify: lookup (value)* [349]*, *(value) lookup* [263]*, *retrieve (value)* [8, 186, 239, 260, 262]*[321], *procure* [260, 262]*
→ `browse`	*browse* [48, 216, 239, 298]*, [78, 318], *search* [260, 262]*, *finding (gestalt)* [42]*, *browsing tasks: follow (path)* [186]*
→ `locate`	*locate* [193, 216, 330, 349, 370]*, [102], *search* [51]*[78], *search (for known item)* [199], *seek* [298]*, *pathfinding* [345]
→ `explore`	*explore* [193, 239, 366]*, [345, 353], *forage* [48, 193, 242]*, *finding (gestalt)* [42]*, *(overview) tasks* [186]*, *find (clusters, correlations, extremum, anomalies)* [8, 186, 239]*, *determine (correlations)* [263]*, *determine (clusters)* [349]*
`query`	*query* [252]*, *posing queries* [42]*, *elementary and synoptic tasks* [12]*, *levels of questions* [329]*, *question answering* [199]
→ `identify`	*identify* [186, 216, 239, 260, 262, 330, 349, 370]*, [260, 262], *reading (the data)* [102], *read (fact, pattern)* [48]*, *lookup* [12]*, *examine* [301]*, *determine (range)* [8, 186, 239]*, *determine / characterize (distribution)* [8, 186, 239, 349]*, *recognize* [175]*
→ `compare`	*compare* [12, 175, 216, 239, 260, 262, 301, 330, 370]*, [199], *compare (within a relation vs. across / between relations)* [263, 349]*, *relation seeking* [12]*, *read comparison* [48]*, *making comparisons* [42]*, [345], *discriminate* [216]*, *associate* [260, 262]*
→ `summarize`	*summarize* [370]*, *summarize (set), enumerate (set objects)* [60]*, *overview* [48, 75, 291]*, *(overview) tasks* [186]*, *scan* [186, 216]*, *connectional tasks* [12]*, *count* [186, 291]*, *visualization* [78], *review* [293]

Table 2.1: Nodes in the *why* part of our typology of abstract visualization tasks and their relation to the vocabulary used in previous work. Underlining is used where a term used in our typology appears in related work. Terms in parentheses are encompassed by the *what* part of our typology. Previous work that explicitly contributes a classification system is denoted by *; other sources incidentally make compelling or noteworthy assertions about the use of visualization tools or techniques.

	how?
encode	*encode* [60, 239, 366, 370]*, *create mapping* [60]*, *visualize* [130, 330]*, *generate* [301]*, *transform (visual mapping)* [58]*
manipulate	*manipulate* [353], *(object) manipulation* [216]*, *modify* [252]*, *(data) manipulation loop* [345]
→ select	*select* [130, 216, 239, 252, 343, 366]*, *brush* [117, 166, 239]*, [58, 345, 353], *distinguish* [349, 370]*, *emphasize* [370]*, *differentiate* [239]*, *highlight* [75, 130, 252]*, [345], *identify: portray, individualize, profile* [370]*, *indicate* [216, 252]*, *mark* [216, 366]*, *reference* [216]*, *outline (clusters)* [370]*, *promote* [48]*, *track* [366]*, *pick* [216]*[58], *express (set membership)* [60]* *connect* [239, 366]*
→ navigate	*navigate* [130, 298, 343]*, [199, 218, 244, 345, 353], *focus* [42, 75]*, [58], *details-on-demand* [48, 291]*, [58], *flip through* [58], *zoom* [42, 48, 75, 117, 166, 216, 239, 260, 262, 291, 366]*, [58, 218, 353], *pan* [42, 117, 216, 239, 260, 262, 366]*, [353], *elaborate* [239, 366]*, *abstract* [239, 366]*, *change (range)* [117]*, *drill down* [75]*, *maneuver / navigate* [301]*, *rotate* [58, 353] *revisit* [117, 186]*
→ arrange	*arrange* [42, 260, 262]*, *sort* [8, 117, 130, 186, 239]*, [218], *rank* [260, 262, 349, 370]*, *coordinate* [130]*, *delineate, sequence* [260, 262]*, *index* [263]*, *move* [216, 252]*, *edit* [216]*, *organize* [130]*, [293], *orient, permute, position, translate* [58], *reorder* [48, 353], *configure* [330]*, *reconfigure* [239, 366]*, *restructure* [193]*
→ change	*change (parameters)* [75]*, [58], *change (metaphor)* [117]*, *change (representation)* [75]*, *change (vis. encoding)* [218], *transform* [252]*, [199, 353], *transform (mapping), shift, scale, set (graphical value)* [60]*, *rotate, scale* [58], *configure* [330]*, *animate* [58, 353], *distort* [166, 343]* [58], *orient / transform* [301]*, *(object) manipulation: transform, stretch, shape* [216]*, *re-express, re-symbolize, re-project* [260, 262]*, *edit* [216, 260, 262]*, *activate* [252]*
→ filter	*filter* [8, 48, 117, 130, 166, 175, 186, 216, 239, 242, 260, 262, 291, 366]*, [218, 321, 353], *subsetting, (value) filtering, (view) filtering* [58], *exclude* [199, 321], *screen: filter, suppress, conceal* [216]*, *maneuver: (data) management / culling* [301]*, *configure* [330]*, *delete (objects, sets, graphical objects)* [60]*, *delete* [48, 117, 252]*, *overlay* [260, 262]*, *restore* [117, 216]*
→ aggregate	*aggregate* [216]*, [58, 218], *cluster* [48]*, [58], *associate* [216, 349, 370]*, *simplify* [58], *link* [42, 75, 166, 216]*, [293, 353], *merge* [117]*, *generalize / merge* [370]*, *assemble* [216]*, *create (set)* [60]*, *split* [117]*, *disassemble* [216]*, *disassociate* [216]*, *reveal: itemize, separate* [370]*, *segregate: ungroup, unlink* [216]*, *withdraw, overlay* [216]*
introduce	*introduce* [216]*
→ annotate	*annotate* [117, 130, 260, 262]*, *add placemark* [366], *create (anchors)* [193]*, *create / copy (graphical objects)* [60]*, *create / modify (note)* [117]*, *externalize (analysis artefacts)* [293], *give a meaningful name to (groups / clusters)* [186]*
→ import	*import* [260, 262]*, *add (objects)* [60]*, *create* [48, 216]*, *generate* [252]*, *(data) entry* [216]*, *load* [193]
→ derive	*derive* [130]*, *derived (attributes)* [60]*, *derive (new conditions)* [301]*, *compute (derived value)* [8, 186, 239]*, *copy* [252]*, *compute* [370]*, *calculate* [216, 260, 262, 301]*, *configure, determine* [330]*, *average* [48]*, *computation operators* [51]*, *transform (data)* [58]*, *estimate, generate (statistics)* [301]*, *extrapolate* [216], *[102], *interpolate* [216], *[102]
→ record	*record* [130, 216, 301]*, *bookmark* [117]*, *history* [291]*, *redo, undo* [117, 366]*

Table 2.2: Nodes in the *how* part of our typology and their relation to the vocabulary used in previous work. Typographic conventions follow those used in Table 2.1.

Vicente argues that sequence-based approaches to task analysis are overly rigid and thus inappropriate for describing such open-ended tasks [337], and that *constraint*-based approaches to task analysis allow for more flexibility in terms of *how* a task is performed. Descriptions based on our typology do not force any strict *global* temporal orderings, as imposed by sequence- or cycle-based models; instead, they accommodate *local* interdependencies within sequences of tasks by way of *constraints* on task `input` and `output`.

Applicability: Many classifications represented in our survey are applicable across domains and datatypes, though specifically-targeted classifications and models do exist. Examples include Lee et al.'s task taxonomy for graph visualization [186] and Lammarsch et al.'s task framework for time-oriented data [185]. Our typology encompasses and complements these specific classifications, and we encourage further development of more like these.

We are also aware of five domain- and datatype-agnostic classifications that span low-level and high-level tasks. These classifications had the highest contributions to the organization of our own typology:

Springmeyer et al. [301] (1992): This classification of scientific data analysis covers both *how* and *why*, but these aspects are not clearly distinguished within its hierarchical structure, that which begins with a high-level distinction between *investigation* and *integration of insight*.

Mullins and Treu [216] (1993): This extensive taxonomy contains over 150 items: an exhaustive list of high-level *mediation* and *coordination* tasks, which overlaps with our classification of *why* and *how*, as well as many low-level object-oriented interactions relating to physical interface `input` and `output`. We do not attempt to specify tasks at this lowest level, though we have adopted a consideration of `input` and `output` in the *what* part of our typology.

Pike et al. [239] (2009): Their characterization of *analytic discourse* draws from earlier work, distinguishing high-level modes of inquiry [7] as *goals*, from low-level tasks [8] and interactions [366]. These are in turn distinguished from the separable *intents* of representation and interaction design choices. Bringing these formerly disjoint classifications together is laudable, though the integration of this information for the purpose of analyzing tasks was not the focus of Pike et al.'s article. The aim of our typology of abstract tasks is to make this integration explicit, relating these intents and design choices (*how*) to modes of inquiry, goals, and tasks (*why*).

Heer and Shneiderman [130] (2012): Their *taxonomy of interaction dynamics* provides a top-level distinction between *data*-, *view*-, and *process*-centric tasks. The focus of their taxonomy is on interactive elements and operations; ten of the twelve task types they characterize are encompassed by the *how* part of our typology. The two remaining *process and provenance* tasks, *share* and *guide*, are captured by the definition of `present` in the *why* part of our typology.

Roth [262] (2012): Roth's taxonomy, based on Norman's *Stages of Action* model [226], classifies *cartographic interaction primitives* as *objectives*, *operators*, and *operands*. Norman's model describes a series of translations between a person's goal, an immediate intention (or *objective*), and a series of actions (*operators*) performed on an environment (of *operands*). Roth's classification is closely aligned with our notions of *why*, *how*, and *what*, and thus has a high-level structure similar to that of our own typology. However, Roth's taxonomy imposes a spatial constraint on where *operands* are located in space, as discussed in Section 2.2. In contrast, we restrict our classification of *what* to that of `input` and `output`; the location of *operands* is represented by the `search` node in the *why* part of our typology.

What these five classifications have in common is that they are atemporal, and most span our characterization of *why* and *how*. Our typology integrates and extends this work, adding a specification of *what*, the `input` and `output` of tasks[8]. As a result, our typology can be used to describe sequences of tasks, in that the `output` of one task may serve as the `input` of another.

[8] As indicated in Section 2.3.3, we chose a flexible and agnostic representation of *what* the `inputs` and `outputs` can be. Munzner [219] has since provided a more structured classification of *targets* that can be used in the analysis of the `outputs` of tasks; for more detail about this extension to the typology, see Section 6.1.

2.5.2 Theoretical Foundations

Our typology was also informed by four theoretical frameworks:

Distributed cognition: The distributed cognition literature offers us a useful distinction between *pragmatic* and *epistemic* actions [173, 194][9]. Pragmatic actions are explicitly and consciously goal-directed, while epistemic actions serve to coordinate actors' internal mental models with external representations of information [194], where an external representation could be an image or interface associated with a visualization tool or technique. Given this distinction, epistemic actions are often performed in support of pragmatic actions. This distinction is lost in low-level classifications; in isolation from higher-level goals we are unable to discern between pragmatic and epistemic actions. Our typology accommodates this distinction. External representations are the graphical and interface elements displayed to or created by a person. Pragmatic actions correspond to the *why* part of our typology, while epistemic actions are captured by the *how* part of our typology. The set of `manipulate` idioms are particularly well-suited for the purpose of describing epistemic actions and their role in coordinating between internal and external representations.

Stages of Action: Norman's *Stages of Action* model [226] and its influence on Roth's *objective-operand-operator* meta-analysis [261] of previous classifications helped shape the *why-what-how* organization of our typology. In the process of evaluating visualization tools, we can discuss Norman's *gulf of execution* with respect to the *how* part of the typology, in which we describe the *means* by which a person can execute the task with a visualization tool. Also central to Norman's model is the *gulf of evaluation*, useful for reasoning about whether the `output` of a task matches a person's expectation. However, this gulf is more applicable when reasoning about specific interaction design choices, which are not directly addressed by our typology. More recently, Lam [180] extended the model with a *gulf of goal formation*, relevant

[9]Distributed cognition theory is fundamental to the study of collaboration, however our typology does not at present explicitly address collaborative visualization tasks; here we focus solely on other aspects of distributed cognition, namely the distinction between pragmatic and epistemic actions.

whenever a person articulates their own questions pertaining to visualized data, thereby specifying the *ends* of a task. This gulf corresponds to the *why* part of our typology, which allows us to abstractly describe these questions.

Sensemaking: The *why* part of our typology overlaps with and bridges to high-level processes of decision making and prediction described in theories of information foraging and sensemaking, both temporal stage-based models [48, 242] and atemporal data-frame models [175]. In particular, sensemaking models connect at the levels of `discover`, denoting hypothesis generation and formation, `present`, and the types of `search`: `lookup`, `locate`, `browse`, and `explore`.

Play theory: Casual interactions with visualized data pose another set of problems for many existing classifications, in that task specifications for these contexts are not easy to motivate by a need to `present`, `discover`, or `produce` [299]. We included `enjoy` in the *why* part of the typology to encompass casual consumption of information, curiosity-driven tasks without expectations or predicted outcomes [77, 318]. The choropleth map example used in Section 2.1 is an instance of this type of task. As visualized data becomes increasingly pervasive in casual contexts, we may turn to theories of casual information seeking and newsreading behaviour, such as Stephenson's *Play theory* [303] to motivate visualization tasks in these contexts. This theory accounts for media consumption activities that bring no material gain, serving no "work" functions, but instead induce moments of absorption and self-enchantment. Casual media consumption relies upon serendipitous apperception, a readiness to interact with information relating to existing interests. Studies of newsreading behaviour indicated that people read most avidly what they already know about [303], a seemingly irrational activity that cannot be described as an explicit *need* to *discover* new information. We posit that this behaviour is also true of some consumption of visualized data, particularly in non-work contexts [247, 299].

2.6 Discussion

Our motivation to develop a multi-level classification of abstract tasks grew in part from our own needs. We have specifically noted that our ability to rigorously analyze tasks has been constrained in the context of the design and evaluation of visualization tools or techniques in general [211, 219] and of design studies in particular [284]. We offer this new typology as a next step in the ongoing discussion in the literature, rather than as a final answer. Our efforts also serve the broader purpose of strengthening the science of analytical reasoning [315] by further uniting the frameworks and methodologies of the cognitive sciences with those in the field of visualization [246]. This work also calls for a wider range of evaluation methods centred around task analysis, with a feedback loop in which tasks observed in field settings can inform subsequent design and evaluation. Our multi-level task typology will serve to expedite this translation and analysis.

We now discuss the capabilities and potential usage of our typology in terms of its descriptive, evaluative, and generative power [16, 18].

2.6.1 Using the Typology to Describe

The typology's descriptive power is in its provision of a consistent lexicon for tasks in terms of *why*, *how*, and *what* in a way supports precise comparisons across different visualization tools and application domains. This lexicon can be used to describe and compare tasks as they occur in situ, of particular use to those analyzing current work practices and use of visualization tools "in the wild". This form of inquiry is often performed within a single domain, such as enterprise data analysis [163] or intelligence analysis [164], wherein tasks are described in a domain-specific way. An interface- and domain-independent vocabulary for multi-level tasks allows practitioners to perform comparative analyses of tasks involving different visualization tools occurring in different disciplines[10].

Only the descriptive aspect of the typology has been directly validated

[10] The interview study described in Chapter 3 is an example of using the typology to classify and compare tasks spanning multiple domains.

in this chapter; we used our typology to describe several empirical cases including single tasks and a sequence of interdependent tasks. We also demonstrated its ability to facilitate the comparison of tasks as they are performed using different visualization tools. Future work includes a further examination of its descriptive power by analyzing whether it covers the full set of abstract tasks described in previously published design studies. In addition, we acknowledge that the typology does not at present explicitly address collaborative use of visualization tools, although we did consider some of the issues involved during its development [153]. Future work will verify if the typology can sufficiently describe collaborative tasks, or if extensions are needed[11].

2.6.2 Using the Typology to Generate

The typology's generative power stems from its ability to prescribe and inform design[12]. In particular, the typology is well-suited to support task analysis occurring throughout the formative *discover* and *design* stages of Sedlmair et al.'s nine-stage design study framework [284]. In the *discover* stage, the practitioner must transform a domain problem into an abstract task description; the typology provides an explicit set of choices for *why* data is being visualized, possibly making this difficult aspect of design studies more tractable. In the *design* stage, the practitioner then chooses *how* the task will be supported, calling upon the existing repertoire of encoding and interaction design choices or inventing new ones. During both the *discover* and *design* stages, the practitioner must consider *what* comprises the `inputs` and `outputs` of these tasks, remaining aware that these tasks may have interdependencies. Once a set of candidate design choices have been identified, the designer must consider additional constraints beyond interdependencies, including human capabilities with respect to perception and interaction, domain conventions, and display medium. Regarding visual perception in particular, seminal research by Cleveland and McGill [62] iden-

[11]This future work may involve revisiting the distributed cognition literature and its discussion of collaboration, as indicated in footnote 9.

[12]We use the typology to inform design in Chapter 5.

tified the perceptual constraints and limitations with respect to elementary perceptual tasks along different visual channels (e.g., comparison of position, length, area, shading, angle, direction, etc.); different visual encoding choices will involve different combinations of visual channels, so knowledge of these constraints and limitations allows us to rank these choices in terms of expected effectiveness [197]. Taking all of these constraints and limitations into consideration, the designer can then make informed decisions about candidate design choices intended to support the task or sequence of tasks.

2.6.3 Using the Typology to Evaluate

The typology is intended to facilitate the evaluation of the experience of using a visualization tool or technique, which includes field studies such as in Chapter 4. We can validate the typology's evaluative power by using it as a set of codes for labelling human behaviour, a common practice in open-ended observational studies of people using visualization tools or techniques; these include longitudinal insight-based studies [229] and multidimensional in-depth long-term case studies (MILCs) [292]. A MILC study of SocialAction [237], a social network visualization tool, incorporated a categorization of interaction design choices by Yi et al. [366] into the analysis of *how* people performed the tasks; we intend that our task typology be used in a similar manner, in which the scope of analysis is expanded to include *why* and *what*. Mixed-method qualitative evaluation studies allow practitioners to determine *how* a task is performed along with its `inputs` and `outputs` via interaction logs and observational analysis; we can also determine *why* data is being visualized via interviews, think-aloud protocols, and artefact analysis.

Task descriptions generated by our typology can also be used to better understand individuals' analytical strategies and the context-dependent variability with regards to *how* a task is performed [337, 371]. Understanding individual problem solving strategies in terms of mental model formation and coordination is also an ongoing goal of distributed cognition research [137, 193]. Our typology and its accommodation of pragmatic and epistemic actions may serve to further this research in the study of people using visualization tools and techniques.

Finally, while our typology may not provide the low level of specification required for defining the procedures of empirical experiments aimed at evaluating the performance of human subjects with respect to specific interaction and visual encoding design choices [183], it may be use to connect and contextualize these low-level experimental tasks to high-level tasks and domain-specific activities.

2.7 Summary

The primary contribution of this chapter is a multi-level task typology that relates both *why* and *how* a task is performed to *what* the task pertains to in terms of `inputs` and `outputs`. The typology allows for the precise description of complex tasks as sequences of simpler tasks, with their interdependencies made explicit. One major advance of the new typology is that it bridges the gap between the low-level and high-level tasks of previous work by providing linkages between them, distinguishing the ends and means of a task. Our typology integrates new thinking with existing classifications of tasks, and with previously established theoretical frameworks spanning multiple literatures. The multi-level task typology presented here is another step towards a systematic theoretical framework for visualization, helping us to describe existing visualization experiences, evaluate them, and generate new ones.

Chapter 3

Interview Study:

Visualizing Dimensionally Reduced Data: Interviews with Analysts and a Classification of Task Sequences

[1]We characterize five task sequences related to visualizing dimensionally reduced data, drawing from data collected from interviews with ten data analysts from different application domains, and from our understanding of the technique literature. Our classification of visualization task sequences for dimensionally reduced data fills a gap created by the abundance of proposed techniques and tools that combine high-dimensional data analysis, dimensionality reduction (DR), and visualization, and is intended to be used in the design and evaluation of future techniques and tools. We discuss implications for the evaluation of existing work practices, for the design of controlled experiments, and for the analysis of post-deployment field observations.

3.1 Motivation

DR is the process of reducing a dataset with many dimensions to a lower-dimensional representation that retains most of its important structure. It

[1]This chapter is a slightly modified version of our paper *Visualizing Dimensionally Reduced Data: Interviews with Analysts and a Characterization of Task Sequences* by Matthew Brehmer, Michael Sedlmair, Stephen Ingram, and Tamara Munzner; in Proceedings of the ACM Workshop on Beyond Time and Errors: Novel Evaluation Methods For Information Visualization (BELIV 2014), p.1-8 [36]. `http://dx.doi.org/10.1145/2669557.2669559`.

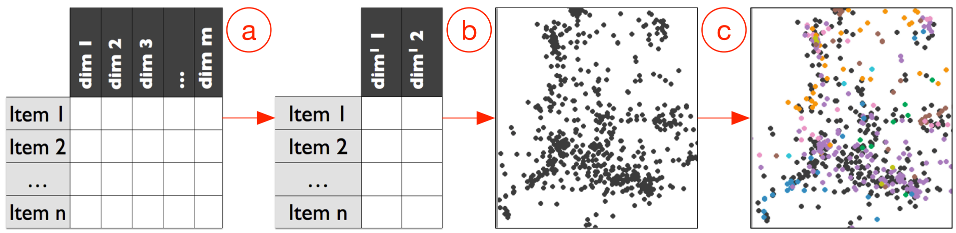


Figure 3.1: A task sequence involving dimensionally reduced data. (a) Data is reduced to two dimensions; (b) `encoded` in a scatterplot to `verify` visible clusters; and (c) colour-coded according to preexisting class labels to *match* clusters and classes.

has been an active research area throughout several decades and across many domains, from its origins in psychology [319, 368] through statistics [41] to machine learning [155, 313, 331] and visualization [149, 150, 159, 365].

While many techniques and tools combining DR with visualization have been proposed, there is still no perfect automated solution that will generate the most effective visual encoding for every situation. Analysts are faced with complex choices between alternative DR techniques and between different visualization techniques for analyzing the resulting data. These choices are strongly dictated by the analysts' data and tasks [312]. The statistics and machine learning communities have provided extensive classifications of DR techniques based on data and technique characteristics [73, 101, 124, 155, 332, 360]. In contrast, there is very little that is explicitly stated about the characteristics of tasks that analysts engage in when visually analyzing dimensionally reduced data. To guide designers, analysts, and those who conduct evaluations of techniques and tools, a better understanding of these tasks is essential.

The contribution of this chapter is a classification of five task sequences related to the visualization of dimensionally reduced data: *naming synthesized dimensions*, *mapping a synthesized dimension to original dimensions*, *verifying clusters*, *naming clusters*, and *matching clusters and classes*. In the last of these sequences, illustrated in Figure 3.1, an analyst uses DR and scatterplots to verify clusters, and then match them with existing classes.

Our classification is based on an in-depth analysis of ten interviews with analysts who use DR for visualizing their data, as well as on a literature review of papers that apply DR for the purpose of data visualization. Our analysis framework is our typology of abstract tasks proposed in Chapter 2. Our typology allows practitioners to characterize task sequences based on observed work practices, occurring in requirements gathering activities and in field evaluations of deployed tools.

3.2 Related Work

Classifying tasks: The systematic analysis of worker activities and tasks is a critical process in the design and evaluation of technology, and task analysis frameworks appear in many different fields, including human factors and ergonomics [337], HCI [216], and visualization, including the typology of tasks proposed in Chapter 2.

While many classifications of visualization tasks are agnostic to datatype, some address specific types of data [291], such as network data [186], time-oriented data [185], and tabular data [133]. As we discussed in Section 2.5.1 and in Meyer et al. [211], datatype-specific task classifications consider a specific set of *data abstractions*, facilitating a mapping to appropriate visual encoding and interaction design choices. A datatype-specific task classification of tasks is also critical for evaluation, such as when specifying tasks to be performed by participants in controlled experiments. In this chapter, we propose a datatype-specific classification of task sequences for *dimensionally reduced* data.

Classifications of tasks are often based on their authors' own experience in conjunction with a thorough consideration of the literature [7, 291], while others are based on observations of human behaviour in controlled laboratory settings [8]. In contrast, our classification of task sequences is primarily based on an interview study with analysts working with their own data [203], allowing us to ground our findings in real data analysis practices.

Mapping tasks to design choices for high-dimensional data analysis: There are many approaches that combine analysis of high-dimensional data, DR, and visualization, including some developed by our research group [149, 150, 357]. While there are existing classifications of high-dimensional data analysis techniques [24] and of dimensionally reduced data [285], the mapping between data, tasks, and appropriate design choices remains unclear [312]. This problem is particularly apparent when designing to accommodate *workflows*, or instantiations of task sequences within software tools for high-dimensional data analysis [150, 159].

One task for dimensionally reduced data is that of matching clusters and categorical classes given with the data, discussed below in Section 3.4.2. Based on findings from an empirical data study, we previously identified effective visual encoding design choices that support this task [286], and we called for similar work to be done for other tasks relating to dimensionally reduced data. Our classification of task sequences moves us closer to this goal.

Expert judgments and dimensionally reduced data: We are aware of one other study involving expert analysts' interpretations of visualized dimensionally reduced data, though they do not share our explicit examination of analysts' domain problems and tasks: Lewis et al. [188] asked expert and novice analysts in a controlled lab setting to subjectively rate the value of two-dimensional scatterplots of seven dimensionally reduced datasets, generated using nine different DR techniques. Their findings showed that experts were more consistent than novices in their positive and negative ratings. Judging the value or quality of a visual encoding of dimensionally reduced data should occur regardless of task, and analysts can additionally leverage automated quality metrics based on human perception [5, 24]. In our study, the domain experts we interviewed varied in terms of their perceived understanding of DR; furthermore, we sought to characterize experts' tasks and activities in naturalistic settings, rather than in a controlled lab study.

3.3 Research Process

Our methodological choice was motivated by a vibrant thread of work in the visualization community using qualitative methods in general [49, 153, 322], and interview studies in particular [163, 164][2].

Data collection: Between 2010 and 2012, we interviewed nineteen data analysts working in academic and industry settings, representing over a dozen domains, spanning the natural sciences, computer science, policy analysis, and investigative journalism. These analysts were recruited from our extended personal and professional networks via snowball sampling, and they were known to work with high-dimensional data. These interviews were semi-structured, lasting in duration from one to four hours[3]; some of these interviews were more akin to contextual inquiries [139], occurring at the analyst's workplace, while others were performed in our department or via teleconference.

We discussed the analysts' domain context, their data analysis goals, and their data; we also asked more specific questions about how they transformed their data and their use of DR and visualization techniques[4]. We also collected artifacts from these analysts, including their published papers and theses, their unpublished manuscripts, screenshots of their visualized data, and in some cases, even their data.

Data analysis: We alternated between data collection and analysis, progressing from *initial* to *focused* coding of the data [55][5].

In this chapter we concentrate our attention on the ten analysts who (a) specifically used dimensional synthesis DR algorithms in analyzing their high-dimensional data, and who (b) also visualized their dimensionally reduced data.

To analyze the data that we collected from these ten interviews, we used our typology of abstract visualization tasks, proposed in Chapter 2. Our typology distinguishes *why* data is being visualized at multiple levels of abstraction, *what* `inputs` and `outputs` a task may have, as well as *how* a task is supported by visual encoding and interaction design choices. This

[2] We elaborate on the evolution of our methodology and the foci of our analysis in Section B.3.

[3] In Section B.1, we indicate that we conducted twenty-four interviews in total, as five analysts were interviewed twice; see Table B.1.

[4] The interview foci and questions can be found in Section B.2

[5] Example artefacts from this data analysis process can be found in Section B.4.

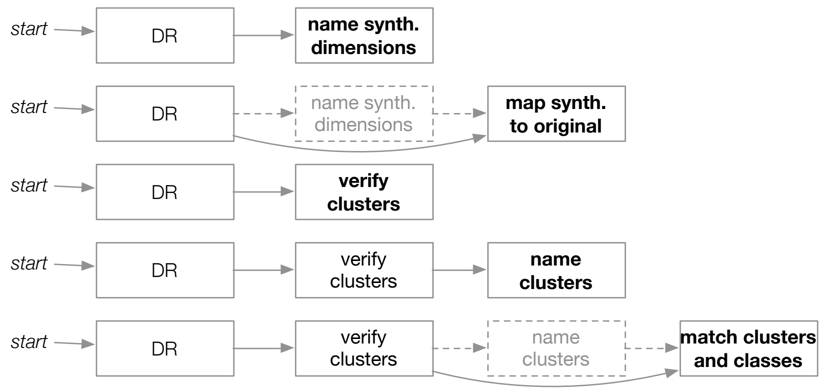


Figure 3.2:
Five task sequences that involve visualizing dimensionally reduced data. Individual tasks are described using our typology in Figure 3.5.

lens allowed us to focus on a subset of our findings from the standpoint of visualization design and evaluation[6], culminating in the task sequences presented in Section 3.4. In Section 3.5, we revisit the typology and illustrate how it can describe our five task sequences.

Finally, we enriched our analysis with further examples from the literature. We specifically sought papers that report on applications where DR and visualization were performed in conjunction for analysis, and we consider these applications with respect to the task sequences we characterized.

3.4 Task Sequences

We have identified five task sequences related to dimensionally reduced data. In this section, we describe each task sequence and illustrate the sequence in Figure 3.2. Each is named after the terminal task appearing in the sequence. We also comment on how these task sequences arose in our interviews, and

[6]The previous interpretations of our findings are documented in Section B.5. We initially focused on characterizing DR techniques, people who use them, their tasks, and their challenges [283]. Later, we narrowed our focus to that of DR techniques and related tasks (see Section B.6). In this chapter, our focus is even narrower, on tasks relating to the visualization of dimensionally reduced data following the use of dimensional synthesis DR techniques.

which visualization techniques were used to address these sequences. These task sequences are not exclusive: some analysts performed multiple task sequences in the course of their work. This descriptive survey of analysts' data, task sequences, and visualization is summarized in Table 3.1. The dataset sizes being investigated by these analysts ranged from dozens to over a million dimensions, and from hundreds to hundreds of thousands of items.

Dimensionality reduction: All the task sequences we characterized begin with DR. In our context, we define DR as a means of dimensional synthesis: a set of m synthesized dimensions is `derived` from n original dimensions, where $m < n$. Dimensional synthesis techniques are commonly differentiated between *linear* and *non-linear* [155]. Linear techniques such as principal component analysis (PCA) [161] or classical MDS [319, 368] produce synthetic dimensions from linear projections of the original data. However, many datasets have an intrinsic structure that can only be revealed using non-linear techniques, such as Isomap [313], t-distributed stochastic neighbor embedding (t-SNE) [331], or Glimmer MDS [149]. Further distinction between linear and non-linear dimensional synthesis is outside of the scope of this chapter, though we note that some techniques are more appropriate for verifying the existence of local cluster structure while others are more appropriate for identifying global intrinsic dimensions (or *manifolds*) [188]. In Table 3.1, we note who used linear and non-linear DR.

It is not our intent to catalog and differentiate the large body of DR techniques; we will concentrate our analysis on their `output`, asking *why* do analysts visualize these synthesized dimensions.

3.4.1 Dimension-Oriented Task Sequences

We describe two task sequences that specifically relate to *synthesized* dimensions as generated by dimensional synthesis DR techniques: *naming synthesized dimensions* and *mapping synthesized to original dimensions.* The verbs *name* and *map* were deliberately chosen and are defined using the vocabulary of our typology in the following two subsections.

Case	Data		DR		Task Sequence					Visualization Techniques						
ID	**Description**	**# Dims x Items**	**Linear**	**Non-Linear**	*Name Dimensions*	*Map Dimensions*	*Verify Clusters*	*Name Clusters*	*Match Clusters*	**2D Scatterplots**	**3D Scatterplots**	**SPLOMs**	**Scree plots**	**Graph / Tree**	**Correl. matrix**	**Heat maps**
A1	usage logs from online music service	48 x 310	✔		✔	✔	✔	✔	✔	✔		✔				
A2	aggregated search engine metrics	12–31 x 1,463	✔	✔	✔	✔	✔	✔	✔	✔						
A3	recreational boating survey data	39 x 543	✔	✔		✔	✔	✔		✔	✔	✔		✔		
A4	protein region data	160 x 10–100K	✔	✔		✔	✔	✔				✔				✔
A5	polymer molecule feature vectors	1K x 10K	✔	✔			✔	✔		✔					✔	✔
A6	bibliometric co-occurrence matrix	20K x 20K	✔	✔			✔	✔		✔		✔	✔		✔	
A7	human motions from multiple sensors	1,170 x 9,120	✔				✔		✔	✔	✔					
A8	genomic, clinical data from patients	1.4M x 600	✔	✔			✔		✔	✔	✔					
A9	distance matrix of genome sequences	100K x 100K	✔	✔			✔	✔	✔	✔	✔	✔		✔		
A10	distance matrix of text documents	10K x 10K		✔			✔	✔	✔	✔				✔		
Ref.																
[41]	distance matrix of Morse codes	36 x 36		✔	✔		✔	✔		✔						
[201]	BRDF reflectance model	4.36M x 104	✔	✔	✔					✔			✔			
[256]	quadruped skeleton models	348–406 x 9	✔		✔								✔			
[313]	64 x 64 px images	4,096 x 698–1K		✔	✔					✔			✔			

Table 3.1: *Top*: A summary of task sequences performed by the ten analysts that we interviewed, along with the visualization techniques(s) used to perform these tasks sequences. *Bottom*: examples of task sequences in papers discussing DR and visualization.

Name synthesized dimensions: Given a set of synthesized dimensions, an analyst may want to `discover` what these dimensions mean, to `generate hypotheses` about the semantics of these synthesized dimensions. An analyst will `browse` the set of synthesized dimensions, and for each dimension of interest, she will `browse` items and their corresponding values; as a result, she may be able to `identify` the name of a synthesized dimension.

This task sequence was attempted by two of the analysts we interviewed (A1 and A2 in Table 3.1). Both worked in the field of HCI and attempted to `identify` the intrinsic dimensions related to usage data collected about online search behaviour and music listening behaviour, respectively.

A common approach, employed by both analysts, is to inspect data points plotted according to two synthesized dimensions in a two-dimensional scatterplot, in which the analyst may be able to discern an interesting semantic relationship along the axes. In some cases, these scatterplots are augmented with text labels containing categorical information, such as item name, annotated adjacent to a subset of the plotted points [41, 201, 313] or available through interaction. Tenenbaum et al.'s paper describing the Isomap algorithm [313] contains a particularly compelling example (reproduced in Figure 3.3), in which each data point in a scatterplot corresponds to an image of a face; a random sample of these images are displayed directly in the scatterplot as thumbnails adjacent to their corresponding points. Given this display, it is possible to discern names for the three synthesized dimensions resulting from dimensional synthesis.

Map synthesized to original dimensions: Regardless of whether an analyst is interested in naming synthesized dimensions, another possible task sequence involves mapping synthesized dimensions back to original dimensions. In the context of PCA [161], this mapping is often referred to as the *loading* of the synthesized dimensions by the original dimensions. Given a synthesized dimension, an analyst may want to `discover` this mapping. More specifically, the analyst may either `verify` a hypothesis that this mapping exists, or `generate` a new hypothesis about it. The analyst will `browse` items and their values along this synthesized dimension and `compare` these

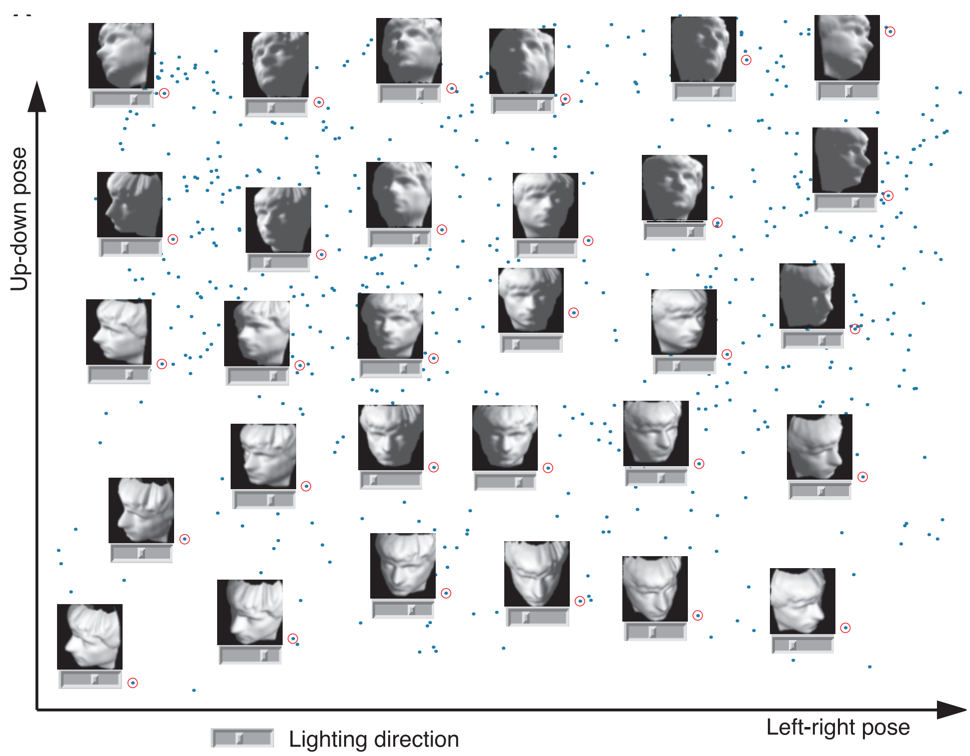


Figure 3.3: A visual encoding of dimensionally reduced data, in which three synthesized dimensions have been identified: *up-down pose* along the y-axis, *left-right pose* along the x-axis, and *lighting direction* indicated below each image. Figure from Tenenbaum et al. [313] (© [2000] AAAS).

values to those along the set of original dimensions, looking for similarities and correlations. This mapping could allow analysts to `identify` groups of correlated original dimensions.

Four of the analysts we interviewed attempted to perform this sequence of tasks; two of these analysts had previously attempted to name some of their synthesized dimensions. A1 mapped her synthesized dimensions to a set of original dimensions in aggregated usage logs from an online music streaming service, while A2 attempted the same task sequence with aggregate search engine metrics but was unable to confidently map any of her synthesized dimensions to her original dimensions. Both used two-dimensional scatterplots to carry out this task sequence. The other two analysts were explicitly interested in grouping original dimensions based on this mapping: a policy analyst (A3) investigating survey data pertaining to recreational boating practices used two-dimensional scatterplots to `compare` synthesized dimensions and original dimensions, while a bioinformatician (A4) investigating protein regions used a SPLOM, heat maps, and density plots.

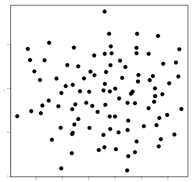
3.4.2 Cluster-Oriented Task Sequences

There exists another set of task sequences where the semantics of the synthesized dimensions are not a central interest; instead, analysts are interested in clusters of items that might be revealed in the dimensionally reduced data. We characterize three task sequences: *verify clusters*, *name clusters*, and *match clusters and classes*. As with the dimension-oriented task sequences, the verbs *verify*, *name*, and *match* were deliberately chosen and are defined using the vocabulary of our typology in the following three subsections.

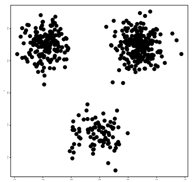
Verify clusters: Analysts might seek to `verify` the hypothesis that clusters of items will be revealed in the dimensionally reduced data, or to `verify` hypotheses about specific conjectured clusters. In order to `discover` clusters, analysts must `locate` and `identify` item clusters in the low-dimensional representation of the data; in the example of Figure 3.4b, we can `identify` three clusters.

All ten of the analysts we spoke to were interested in verifying that clusters exist in their data. This task sequence is also captured by a discussion by Buja and Swayne [41] about visualizing data following multidimensional scaling. The analysts we interviewed used a variety of visualization techniques when performing this task sequence, including two-dimensional monochrome scatterplots, such as those depicted in Figure 3.4a-b, as well as three-dimensional scatterplots, SPLOMs, dendrograms, heat maps, and density plots.

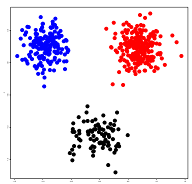
Name clusters: Once the existence of clusters has been verified, such as in the example of Figure 3.4b, the next task is often one of `generating` hypotheses regarding the meaning of these clusters in the form of a name.



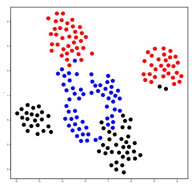
(a) No discernible clusters.



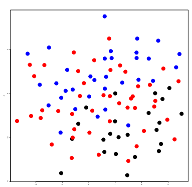
(b) Three discernible clusters.



(c) A match between clusters and class labels.



(d) A partial match between clusters and class labels.



(e) No discernible class separation.

Figure 3.4: Example scatterplots of dimensionally reduced data illustrating tasks related to item clusters: Verifying the existence of clusters, naming clusters, and matching clusters and classes.

In this `discover` task, an analyst will `browse` items within a cluster and attempt to `summarize` the cluster with a meaningful name. In some cases, this name is made explicit, as the analyst will `annotate` the cluster, thereby using the visual encoding to `produce` new information about their data.

Eight of the analysts who had previously *verified clusters* also attempted to *name clusters* in the course of their work, using the same visualization techniques. For instance, A6 examined bibliometric data from a corpus of life sciences research literature, who attempted to `identify` and name clusters of related research concepts, such as *“cancer”* or *“RNA”*.

Match clusters and classes: The final task sequence we characterize is matching clusters with classes. The `input` to this *match* task is not only a set of item clusters, `identified` in the earlier *verify clusters* task, but also a set of categorical class labels. These classes might come directly with the data, be assigned using a clustering algorithm run by the analyst, or be the result of manual labeling. The analyst must `verify` a hypothesis that a cluster of items matches the class for those items. To `discover` a match, the analyst performs a `lookup` for the class and cluster membership of an item in order to `compare` them, resulting in a match (as in Figure 3.4c), otherwise referred to as a *true positive*, or a mismatch (as in Figure 3.4d-e), which could either be a *true negative* or a *false negative*. This task was examined in our recent paper [286], a paper that offered guidance for choosing appropriate visualization techniques for dimensionally reduced data.

Naming the clusters is not a pre-requisite for this *match* task, though we did encounter four analysts who reported performing both tasks in succession (A1, A2, A9, A10); two other analysts performed this task without previously naming the clusters they identified (A7, A8). Typically, this task was performed using two-dimensional scatterplots, wherein the points were coloured using the class labels; SPLOMs, interactive and non-interactive three-dimensional scatterplots, and node-link graphs were also used. Note that the visual separability of colour-coded clusters differs perceptually from the separability of monochrome clusters, as described in our recent taxonomy of cluster separation factors [285]. These perceptual differences should be taken into account particularly when determining which experimental stimuli for use in controlled experiments.

A possible outcome of this task sequence is a partial match between classes and clusters: there may be more clusters than classes, or vice versa. In cases where there are more clusters than class labels, illustrated in Figure 3.4d, this outcome suggests that the class labels may not capture a finer-grained cluster structure in the data, as was the case for the investigative journalist that we interviewed (A10). In cases where there are more classes than clusters, illustrated in Figure 3.4e, this result may either be a *true negative*, in which perfect class separation is not possible, or a *false positive* [286]. If this mismatch is suspected to be a *false negative*, Sedlmair et al. recommend `selecting` other dimensions to visualize, using other design choices such as a SPLOM, or revisiting the choice of DR technique.

3.5 A Task Typology Revisited

The analysts that we interviewed hailed from very different domains, each using a different terminology to describe their work processes. For instance, we needed a way to compare how *diagnosing cancer patients based on their genomic data* (A8) was like *classifying types of human motion through the use of sensors attached to the body* (A7). We required an abstract vocabulary for describing and comparing the work processes of these analysts.

For this reason, we used our typology of abstract visualization tasks, introduced in Chapter 2, which provided a domain-agnostic vocabulary and framework for describing visualization tasks in terms of *why*, *what*, and *how*. By describing a task in this manner, we can link `outputs` and `inputs` to describe sequences of interdependent tasks, which Norman would refer to as *activities* [227]. We use it here to describe task sequences relating to visualizing dimensionally reduced data across multiple domains.

Our analysis concentrated on the *why* and *what* aspects of the tasks pertaining to dimensionally reduced data, as summarized in Figure 3.5. We chose not to be prescriptive about *how* these task sequences should best be supported by visualization techniques; instead, we described the variety of techniques used by the analysts that we interviewed for each task sequence, as summarized in Table 3.1.

The analysts we interviewed were all interested in `discovery`, which involves the `generation` and `verification` of hypotheses. Figure 3.5b-f show which tasks relate to `hypothesis generation` and which relate to `hypothesis verification`. The graphical depiction also shows which task can be associated with pure `consumption` of information and which task can additionally lead to the `production` of new information. When consuming information, an analyst will `search` for targets within a visual encoding. Whether the location and identity of these targets is known a priori will determine the type of `search`. In tasks related to visualizing dimensionally reduced data, we found that `search` strategies used by analysts were either

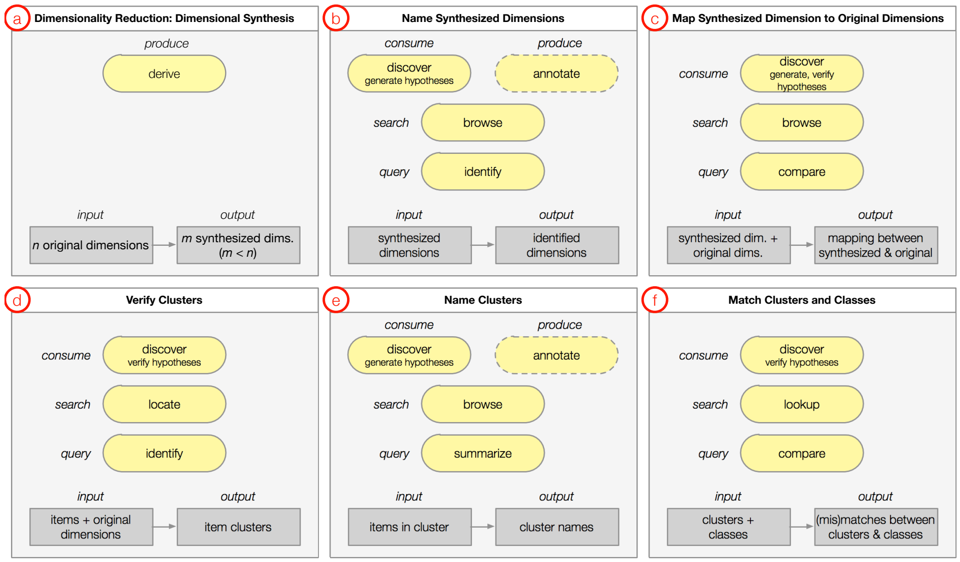


Figure 3.5: Six tasks related to dimensionally reduced data, characterized using our abstract task typology introduced in Chapter 2, which describes *why* the data is being visualized at multiple levels of abstraction (yellow) and *what* `inputs` and `outputs` a task has (grey). These tasks are combined to form the task sequences described in Section 3.4.

`browse`, `locate`, or `lookup`, as indicated in Figure 3.5b-f. Once targets are found, an analyst will execute some form of `query`: they might `identify` a single target, such as an item cluster, `compare` multiple targets, such as values along a synthesized dimension to values along an original dimension, or `summarize` all the targets, such as when *naming a cluster*.

Dependencies: The task sequences described in Section 3.4 contain dependencies. For example, in order to *match clusters and classes*, an analyst must first *verify* that clusters exist. Each of the sequences also depend on the `output` of DR techniques, the `derived` synthetic dimensions. The application of DR to a set of original dimensions is itself a task, as shown in Figure 3.5a. However, unlike the other tasks described in this chapter, it is about neither hypothesis generation nor verification, but rather about

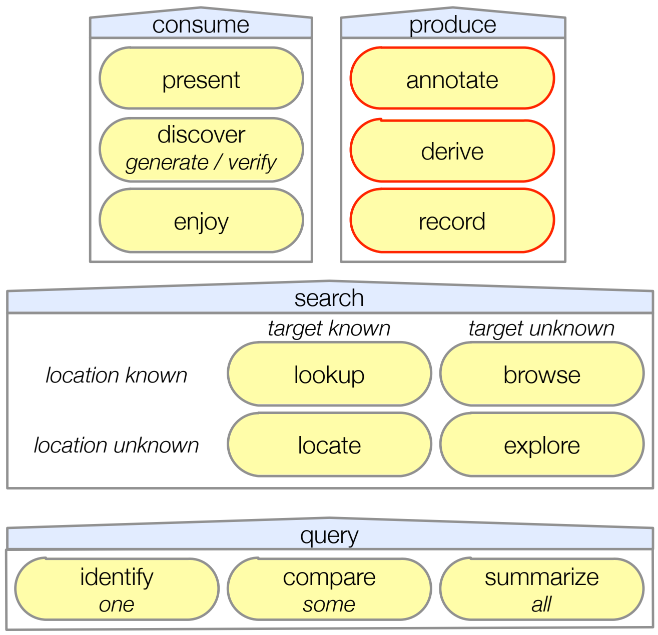


Figure 3.6: The *why* part of our abstract task typology from Chapter 2, with the refinement (emphasized in red) that the actions of `annotate`, `record`, and `derive` are forms of `produce` [219].

`producing` new information intended to support subsequent tasks.

While the distinctions between these tasks and task sequences may seem obvious in hindsight, we initially struggled to find a vocabulary and framework that would allow us to distinguish between these task sequences and their interdependencies. Our task typology, introduced in Chapter 2, allows us to describe these task sequences explicitly, whereas they were implicit in previous work combining DR and visualization.

Extended typology: Figure 3.6 reproduces the *why* part of an extended task typology [219][7].

The changes relevant to our analysis in this chapter pertain to three actions: an analyst may `annotate` information, `derive` new information from existing, or `record` their use of a visualization tool so as to provide ana-

[7] We comment further on the the extensions to our typology in Section 6.1.1.

lytical provenance or to facilitate subsequent presentations of the visualized data. The terms `annotate`, `derive`, and `record` were previously attributed to families of interaction design choices in the *how* part of our typology; the extended typology classifies them as ends rather than means and thus situates them as forms of `produce`. Both versions of the typology distinguish whether a person will visualize data either to `consume` or `produce` information. The remaining aspects of the typology describing lower levels of abstraction are unchanged.

3.6 Discussion

We discuss the utility of our classification of task sequences with regard to several visualization evaluation scenarios, the limitations of our current findings, and our planned future work.

3.6.1 Implications for Evaluation

Task analysis and evaluation are closely linked. An understanding of visualization tasks informs how an evaluation is conducted, from the justification of experimental procedures to the collection and analysis of field observations.

Our current work adds to previous task classifications proposed in the visualization evaluation literature [133, 186, 330]. As evaluation takes on many forms, we frame our discussion around four of Lam et al.'s scenarios for empirical studies [183].

Understanding work practices: Work practice evaluation or work domain analysis can provide a richer understanding of the perspective of people who might benefit from visualizing their data, reflecting real work practices and activities. While we have outlined their immense importance several times [34, 211, 217], only a few dedicated examples exist in the visualization literature [163, 164, 322].

More commonly, however, such work practice evaluations occur in design studies, an increasingly popular form of problem-driven visualization research. In particular, a design study's early *discover stage* [284] involves the analysis of work practices within a very specific usage context in a particular domain. These concrete work practices are then translated into abstract visualization tasks and design requirements.

Our current work goes beyond task classification in design studies by conducting interviews with analysts across different application domains. We then cast our findings as task sequences or activities [227] that abstract away domain-specific language. In doing so, we intend to support researchers when conducting and analyzing future *work practice* evaluations, specifically when DR techniques are to be used. We encourage practitioners to adopt our classification of task sequences into a lexicon for coding observations of work practices and for translating domain-specific descriptions of these practices. We believe that using our task sequences will make the analysis process more efficient and, furthermore, will allow for transferability between design studies from different application domains [284].

Evaluating human performance: Our classification of task sequences can inform the design of experimental procedures and participant instructions in controlled laboratory studies, where the aim might be to quantitatively assess human performance on a newly proposed visualization technique. Many previous classifications of tasks have informed experimental design, such as the adoption of a task classification by Zhou and Feiner [370] in a laboratory evaluation of an information retrieval tool [214]. We expect that our classification of task sequences will play a similar role in the evaluation of techniques or tools that visualize dimensionally reduced data. For instance, an experiment might compare multiple visualization techniques for *verifying clusters* and subsequently *matching clusters and classes*, where performance might be measured in terms of speed and accuracy.

Munzner [217] refers to such studies as a form of downstream validation, in which a design has been implemented for its investigation in a study. In contrast, upstream validation in this case refers to the justification of visual encoding and interaction design choices before its implementation. We deem our task sequences to be similarly helpful for such upstream evaluations. Researchers presenting new visual encoding or interaction design choices can refer to our task sequences to concisely state assumptions about which abstract tasks are supported, rather than leaving this description implicit in a way that places a burden on a potential adopter of the design choice.

Evaluating the experience of using a visualization tool or technique: In either lab or field settings, a researcher can evaluate the experience of using a tool or technique by dictating the tasks without specifying *how* to execute them, asking study participants to verbalize their actions while they attempt to execute a sequence of tasks. Such a think-aloud protocol might allow the researcher to understand if features of the tool are learnable, useful, or in need of further usability improvements. Questionnaires and interview questions relating to the experience of using a visualization tool or technique could also be framed around our classification of task sequences.

We note that expertise has many facets; the distinction between novices and experts is a particularly nuanced question for studies considering DR. Several of the high-dimensional data analysts that we interviewed might be described as *middle-ground users* [150]: they had significant domain expertise but only a partial understanding of the available DR tools and of the mathematics underlying these techniques. This characteristic is important to keep in mind when recruiting participants for evaluations of performance or experience, as some evidence exists that participants with an understanding of DR will interpret visual encodings of dimensionally reduced data differently than those who do not have this understanding [188].

Evaluating visual data analysis and reasoning: While a researcher must dictate the tasks in a controlled laboratory experiment, another scenario is the observation of tasks in an open-ended qualitative evaluation of a visualization tool or technique. Here, the researcher must recognize when these task sequences appear in naturalistic settings, in order to better understand how visual data analysis and reasoning are supported following the introduction of a new visualization tool. This form of evaluation is typical in design studies [284, 292], particularly after a tool is deployed.

As with evaluations of work practices, our classification of task sequences could become part of a lexicon for coding observed behaviour after a tool is deployed. In cases where direct observation of tool use is not possible, our classification of task sequences might be used to analyze interaction log files, or used as a basis for diary or interview questions, suggesting a consistent vocabulary for coding participant responses. Precedents for the use of task classification in evaluation of deployed tools include the adoption of a classification by Yi et al. [366] in a longitudinal field study of a social network analysis tool [237], or how we used our task typology introduced in Chapter 2 to evaluate why and how journalists used *Overview*, a tool for analyzing large document collections [35][8].

Finally, if we consider the task sequences *name synthesized dimensions* and *name clusters* in particular, one conceivable evaluation of visual data analysis and reasoning would involve collecting participant annotations and explanations of synthesized dimensions or clusters in visual encodings of dimensionally reduced data. Such a study might adopt a protocol similar to one used by Willett et al. [355] to elicit participant annotations and explanations of visualized time series data in an application deployed online. This evaluation could help to identify the features of a visualization tool that facilitate or inhibit visual data analysis and reasoning.

3.6.2 Limitations

Our interview findings are certainly not exhaustive, and despite conducting interviews with nineteen analysts, only ten of these analysts contributed to our classification of task sequences. This selection was based on our goal of studying task sequences relating to visualizing data reduced with *dimensional synthesis* techniques. There are many other interesting areas of high-dimensional data analysis that we did not address. Specifically, we found that many of our excluded interviewees used *dimensional filtering* techniques, in which a subset of the original dimensions are retained [159, 365]. Alternatively, other analysts applied DR to their data without visually analyzing it. In these cases, DR was used to reduce the data for algorithmic

[8] This use of the task typology is documented in Chapter 4.

input, such as for classification and other machine learning applications.

We consider our findings to be existence proofs of the task sequences as performed by analysts as part of their ongoing work. We do not make claims about the prevalence of these task sequences in high-dimensional data analysis, nor do we make claims about completeness: our classification of task sequences might be incomplete due to sampling or observer bias.

3.7 Summary

In this chapter, we presented a classification of five task sequences related to visualizing dimensionally reduced data:

- Name synthesized dimensions: `discover` meaning of these dimensions, `generate hypotheses` about their semantics, `browse` these dimensions and their corresponding values, and ideally `identify` their names.
- Map synthesized to original: `discover` this mapping, `verify` a hypothesis that this mapping exists, or `generate` a new hypothesis about this mapping; for a synthesized dimension, `browse` items and their values and `compare` these values to those from the original dimensions and ideally `identify` groups of correlated original dimensions.
- Verify clusters: `verify` a hypothesis that clusters of items exist, or `verify` a hypotheses about specific conjectured clusters, `locate` clusters.
- Name clusters: `generate` hypotheses regarding the meaning of these clusters, `browse` items within a cluster, `summarize` the cluster with a meaningful name; in some cases, `annotate` the cluster (`produce` new information about the data).
- Match clusters and classes: `verify` a hypothesis that a cluster of items matches the class for those items; to `discover` a match, `lookup` the class and cluster membership of an item in order to `compare` them.

Our abstract classification of these task sequences fills a gap between the large body of technique-driven literature and analysts' domain problems in this area. We encourage other researchers to consider these task abstractions in the evaluation of existing work practices, in the *discover* phase of future design studies involving high-dimensional data and DR, in the design of controlled experiments, and in field evaluations of deployed visualization tools.

Chapter 4

Field Study:

Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists

"The Street finds its own uses for things - uses the manufacturers never imagined." — William Gibson in "Rocket Radio" (*Rolling Stone*, June 15, 1989)

[1]For an investigative journalist, a large collection of documents obtained from a Freedom of Information Act (FOIA) request or a leak is both a blessing and a curse: such material may contain multiple newsworthy stories, but it can be difficult and time consuming to find relevant documents. Standard text search is useful, but even if the search target is known it may not be possible to formulate an effective keyword search term. In addition, summarization is an important non-search action. We present *Overview*[2], an application for the systematic analysis of large document col-

[1]This chapter is a slightly modified version of our paper *Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists* by Matthew Brehmer, Stephen Ingram, Jonathan Stray, and Tamara Munzner; in IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis 2014), 20(12), p. 2271–2280 [35]. `http://dx.doi.org/10.1109/TVCG.2014.2346431`. Section 4.9 is a new addendum section that is unique to this dissertation. High-resolution versions of the figures in this chapter are available here: `http://cs.ubc.ca/labs/imager/tr/2014/Overview/`.

[2]Throughout this chapter, *Overview* is italicized to distinguish it from "overview", an overloaded term in the visualization literature.

lections based on document clustering, visualization, and tagging. This work contributes to the small set of studies which evaluate a visualization tool "in the wild", and we report on six case studies where *Overview* was voluntarily used by self-initiated journalists to produce published stories. We find that the frequently-used language of "exploring" a document collection is both too vague and too narrow to capture how journalists actually used our application. Our iterative process, including multiple rounds of deployment and observations of real world usage, led to a much more specific classification of tasks. We analyze and justify the visual encoding and interaction design choices used in *Overview*'s design with respect to our final task abstractions, and propose transferable lessons for visualization design methodology.

4.1 Motivation

FOIA requests, leaks, government transparency initiatives, or other disclosures can result in thousands or millions of pages of potentially newsworthy material. Investigative journalists must find the stories lurking in these massive document collections, but it is frequently impossible to read every document. Standard text search can be used to `locate` documents containing particular terms, but not all information retrieval problems can be expressed as word search queries, especially if the relevant information is unexpected or novel. Journalists may also be interested in patterns of text across many documents, which can reveal significant trends, categories, or themes. We conjectured that this *document mining* problem could be solved by a visualization tool built around clustering and tagging documents. The path from this hypothesis to a tool that working journalists would voluntarily use was a long one; we needed to refine both our understanding of the problem and the ways in which journalists might want to solve it.

This chapter reports on the design, adoption, and analysis of *Overview*[3], an application developed by co-author Jonathan Stray in collaboration with our research group over several years. *Overview*, shown in Figure 4.1, visualizes a document collection as a tree where nodes represent clusters of

[3] `https://www.overviewdocs.com/`

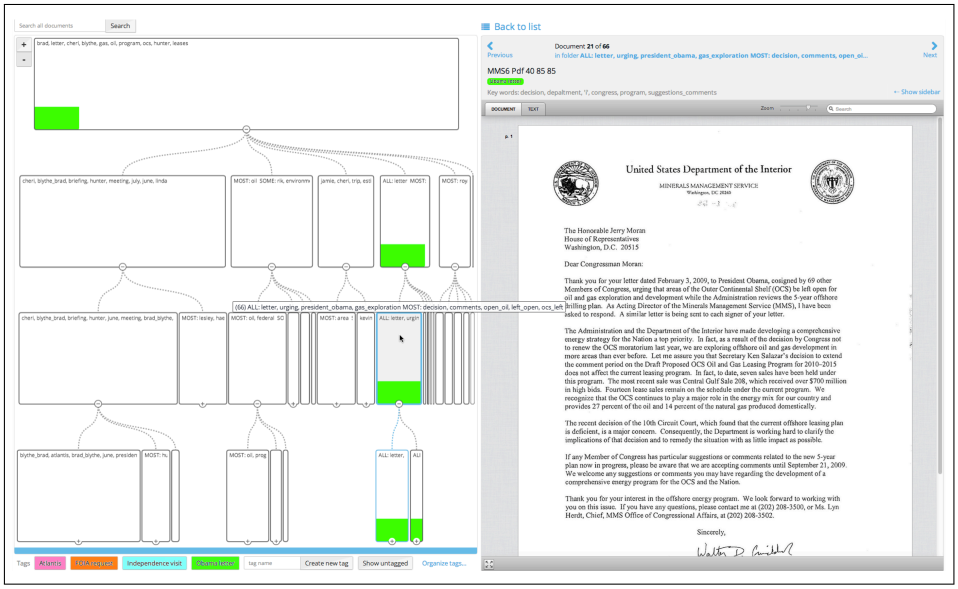


Figure 4.1: *Overview* is a multiple-view application intended for the systematic `search`, `summarization`, `annotation`, and reading of a large collection of text documents, hierarchically clustered based on content similarity and visualized as a tree (left). Pictured: a collection of White House email messages concerning drilling in the Gulf of Mexico prior to the 2010 Deepwater Horizon oil spill.

similar documents; a person can `navigate` this tree, `identify` clusters, read individual documents, and `annotate` documents with meaningful tags. A timeline illustrating *Overview*'s development, deployment, and adoption phases is shown in Figure 4.2. Beginning with an initial use case, we developed a research prototype (*v1*), a publicly available cross-platform desktop application (*v2*), and finally a web-based application (*v3-v4*). Ultimately, we succeeded in building a useful tool for journalists: we report on multiple case studies where *Overview* was adopted for real investigations. Analysis of these cases revealed that journalists often used the application in ways we did not anticipate, and we found that the often-used concept of "exploring" a document collection fails to capture the tasks that journalists actually perform.

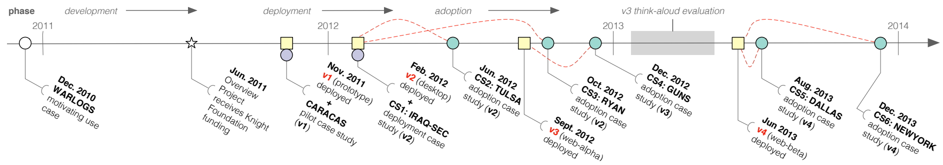


Figure 4.2: A timeline of *Overview*'s development, deployment, and adoption phases: deployments are represented as yellow squares; deployment-phase case studies are represented as ● (purple circles), while adoption-phase case studies are represented as ● (turquoise circles). The dotted red lines indicate which version of *Overview* was used in each case study.

We frame this work as a visualization field study, one that took place during and after a process of iterative design addressing a particular domain problem, involving collaborators and people from that domain. The contributions of this chapter include our classification of data and task abstractions, a description of its usage in real investigations spanning four deployments and six case studies, and a detailed analysis of the mapping from these abstractions to visual encoding and interaction design choices. This analysis led to important design revisions, based on a better understanding of *why* and *how* journalists use *Overview*. From this experience we propose transferable lessons for visualization design methodology.

4.2 Related Work

There have been a number of approaches and tools to support the analysis of document collections, spanning a range of data transformations and visual encodings. We also review how these tools were evaluated.

Topic model visualization: One common approach to visualizing a document collection uses probabilistic topic models inferred from the collection. These define topics as distributions of words and assign a distribution of topics per document. Both distributions are visualized directly in recent work by Chaney and Blei [52], while other tools or techniques focus on the number of documents in each topic [70, 80, 191], or use the topic assignments to compute the similarity between documents [57, 86]. *Overview* does not use distribution-based topic models but directly creates a hard hierarchical clustering, which is visually encoded as a tree.

Documents as points: Many visualization tools, including the first two versions of *Overview*, encode individual documents as points in a scatterplot. InfoSky [118] places points according to a pre-existing hierarchical arrangement of documents; in contrast, *Overview* is intended for document collections that do not have a pre-existing hierarchical structure. Other approaches begin with an unstructured document collection and place points based on document similarity metrics and DR techniques, such as Leaksplorer [38], PEx [232], and EV [57]. *Overview v1-v2* included a similar scatterplot which placed points by DR through MDS. Finally, ForceSPIRE [91] and TopicViz [86] incorporate a scatterplot where points corresponding to documents can be interactively placed according to one's own semantics or mental model, adaptively adjusting the underlying similarity metric used between document pairs. In Section 4.6.2, we discuss in greater detail why a scatterplot was omitted from later versions of *Overview*, and how tagging documents and clusters is an effective alternative to interactive placement.

Documents as landscapes or clouds: Document collections have also been encoded as landscapes, three-dimensional visual encodings of two- dimensional scatterplots where height represents density, as in In-Spire [135] and recent work by Österling et al. [231]. However, empirical studies have shown that spatial landscapes are not well suited for encoding inherently non-spatial data, and exhibit poor visual memory performance in comparison to two-dimensional scatterplots [324].

It is also possible to visualize a document collection by encoding clusters of documents as interactive tag clouds, as in Newdle [192]. Once again, previous research has documented the perceptual drawbacks of tag clouds [128]. By encoding a document collection as a tree, *Overview* circumvents these issues.

Documents as networks of entities: Jigsaw's approach [115, 165] to document collection analysis differs from *Overview* in that it emphasizes the extraction of entities from documents, linking names, places, events, and dates, visualizing these relationships. The emphasis on entities is reflective of the domains in which Jigsaw is used, which include intelligence analysis, law enforcement, and academic research [165]. Journalists frequently start with barely-legible scanned documents which must first be converted to text through optical character recognition (OCR), greatly reducing the accuracy of standard entity extraction techniques. As a flexible multiple-view application, Jigsaw also has a significant learning curve, and people have reported investing many months into learning how to use it [165]. The journalists we spoke to are accustomed to short deadlines and may only intermittently be working on a story involving a large document collection, so simplicity is a crucial requirement.

Documents as trees and rivers: Like *Overview*, HierarchicalTopics [80] features a tree of document clusters, initially arranged by similar keywords. It allows people to re-arrange the tree according to their own semantics, similar to how a person who uses ForceSPIRE can rearrange documents in a scatterplot [91]. HierarchicalTopics [80] additionally allows people to track topic prevalence over time with a stacked area graph visual encoding in the style of a ThemeRiver [127]. However, this approach requires temporal metadata that would be difficult to extract from the diverse document sources supported by *Overview*.

Evaluating visual document mining tools: Several of the aforementioned tools have been evaluated via controlled experiments and case studies. Controlled experiments, such as those used to evaluate Newdle [192] or HierarchicalTopics [80], often involve non-specialists conducting domain-agnostic tasks specified by the researchers, who conjecture that they match with real world usage. Moreover, the documents used in these controlled experiments were collections of online news articles which are not appropriate test data for *Overview*, as professionally produced news articles are clean and homogeneous, unlike the diverse and messy documents obtained by our case study journalists, which often contain little or no metadata; news articles are the output of the journalistic document mining process, not the input.

Most similar to our approach is a series of case studies of academic researchers, intelligence analysts, and law enforcement personnel who had adopted Jigsaw [165]. These case studies resulted in a better understanding of Jigsaw's utility in relation to the tasks of people working in a specific application domain; like us, they identified similar barriers to adoption and their results suggested new directions for design [115, 116].

4.3 Initial Use Case

The *Overview* project began in December 2010, when journalist and co-author Jonathan Stray visualized a subset (11,616 of 391,832) of the WikiLeaks Iraq War Logs [305]. Journalists had previously examined these documents by using text search to retrieve specific records and by visualizing the structured data fields such as time and location, but they had not attempted an analysis of the unstructured text of the reports. In this initial use case, which we will refer to as WARLOGS, documents were visualized as points placed according to a measure of similarity between documents and coloured according to pre-existing categorical labels, such as *"friendly action"* and *"criminal incident."* As shown in Figure 4.3, this design revealed meaningful cluster structure that cross-cuts the colourings, showing that the pre-existing coarse categorization does not capture the whole story[4].

The WARLOGS scatterplot had serious limitations: it was not possible to interactively and systematically examine the contents of clusters of documents. However, it demonstrated that visual cluster analysis could illuminate previously unknown and meaningful structure in a real world document collection, a conjecture that Stray had synthesized from his previous experience reporting on this collection of documents. On the basis of this promising result, Stray collaborated with us to design an interactive visualization tool for document mining.

[4]The entire image is shown in Section C.1.

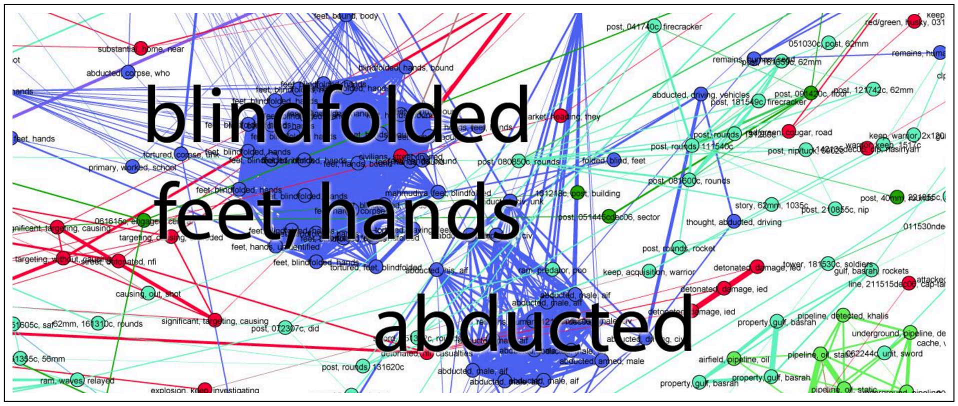


Figure 4.3: Detail from *"A full-text visualization of the Iraq War Logs"* (WARLOGS) [305], in which distinct clusters of documents are visible; these documents pertain to "criminal incidents" during the Iraqi civil war involving abductions and blindfolding.

4.4 Design of Overview

We now describe our initial task abstraction, *Overview*'s underlying data abstractions, and the elements of its interface.

Initial task abstraction: During the development of *Overview v1-v2*, our task abstraction was based on the WARLOGS use case: journalists would be motivated by the hypothesis that their document collection contained a semantically interesting cluster structure, and would require a means for `exploring` that structure, drilling down into these clusters to examine the contained documents. During this exploration, they would need a way to keep track of what they had discovered, allowing them to revisit previously examined clusters and documents.

Data abstractions: Although *Overview*'s design has evolved over the course of four deployed versions, it continues to reflect several underlying data abstractions. *Overview* does not incorporate any novel text analysis techniques; following a practice common in that domain, we convert each document to a vector of words weighted by the term frequency-inverse document frequency (TF-IDF) formula, and compute similarity between documents using the cosine distance metric [275]. We generate our document

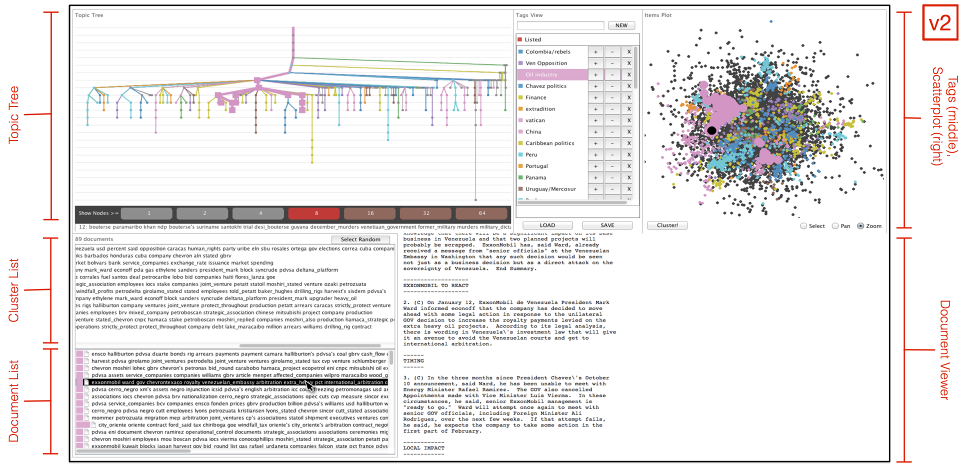


Figure 4.4: *Overview v2*, a desktop application released in Winter 2012. Shown here is 6,849 of the U.S. State Department diplomatic cables released by WikiLeaks, those pertaining to Venezuela. The *"Oil industry"* tag is `selected`; clusters containing documents having this tag are emphasized in pink in the *Topic Tree* and are shown in the *Cluster List* as a set of keywords. Individual documents having the *"Oil industry"* tag are emphasized in the scatterplot and shown in the *Document List* as a set of keywords. The fifth document is `selected`; its contents are displayed in the *Document Viewer* and it is marked as a larger black dot in the scatterplot.

clusters by hierarchically clustering these distances and encoding the result as a tree [146, 151]. Clusters are labeled with keywords extracted via TF-IDF scores.

Multiple meaningful clusterings may exist for any collection of documents [121]; our particular distance metric and hierarchical clustering algorithm is but one possible choice. Human-generated clusterings that leverage domain knowledge can complement automatic clusterings [80, 91]. For these reasons, *Overview* allows for an arbitrary number of human-generated *tags* on each document, which can be assigned to individual documents or at the cluster level. Tags allow people to keep track of what they have found and where they have looked so far.

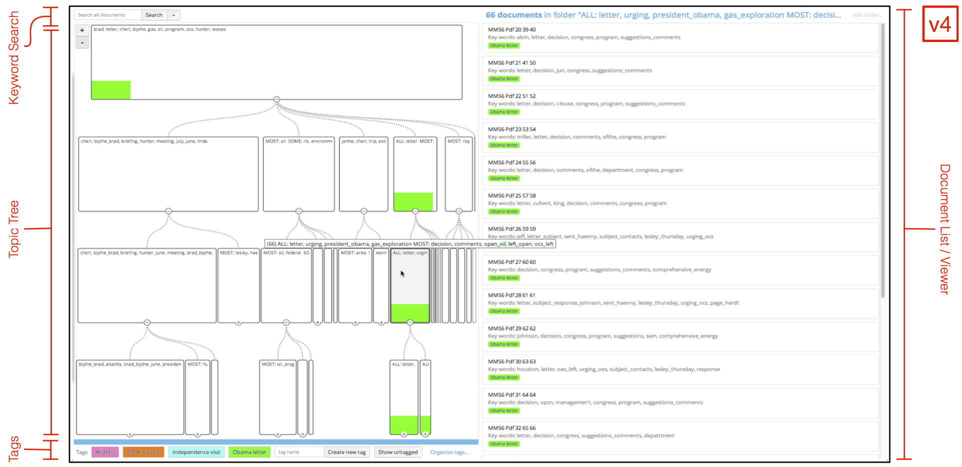


Figure 4.5: *Overview v4*, a web-based application released in Summer 2013. Shown here is 625 White House email messages concerning drilling in the Gulf of Mexico prior to the 2010 Deepwater Horizon oil spill. The *"Obama letter"* tag is `selected`; clusters containing documents having this tag are highlighted in green in the *Topic Tree*. One of these clusters is `selected` and its keywords are displayed in a tooltip; the 66 documents in this cluster are listed in the *Document List*. `Selecting` a document from this list reveals the *Document Viewer* (cf. Figure 4.1).

Interface: With each deployment came changes to the interface, though we will focus on the differences between *Overview v2* and *v4*, shown in Figure 4.4 and 4.5[5], respectively. The visualization design of *v1* and *v2* are quite similar to each other, as are *v3* and *v4*[6].

Common to all deployed versions of *Overview* is the *Topic Tree* visual encoding, representing a hierarchical clustering of similar documents, the *Document List*, showing currently `selected` documents, the *Document Viewer*, and the ability to create and assign custom categorical tags to clusters or individual documents; tags are `encoded` as coloured labels on documents and clusters. `Selections` of documents are propagated and highlighted across

[5] A video demonstration of *Overview v4* is available here: `http://vimeo.com/71483614`.

[6] Screenshots of *v1* and *v3* can be found in Appendix C as Figure C.2 and Figure C.6, respectively.

views.

The *Topic Tree* underwent some of the most significant changes. It was redesigned to emphasize nodes, and to `encode` the number of documents in each node, instead of focusing on the edges between identically-sized nodes. In *v1-v2*, the *Topic Tree* could be pruned based on a threshold cluster size, controlled using a set of coloured radio buttons below; in *v3*, we replaced threshold pruning with an open/close interface that allows a person to show or hide the children of any node. Pan and zoom controls were also added, including an auto-zoom feature that automatically zooms and pans to a selected node.

Another prominent change was the removal of the interactive scatterplot, in which individual documents were `encoded` by points and their placement corresponded to a two-dimensional projection of the original high-dimensional TF-IDF vector space, generated via MDS; pairs of documents appearing closer together were deemed to be more similar than pairs of documents that were farther apart. The scatterplot had panning and zooming controls, and document-points could be `selected` via clicking or lassoing.

We also removed the *Cluster List* and consolidated the *Document Viewer* with the *Document List* (c.f. Figure 4.1). The *Document List* now displays the document title, extracted keywords, and coloured labels indicating which tags have been applied to each document. We added full-text keyword search in *v4*; documents matching a search term are highlighted with colour labels in the *Topic Tree*, and these results can be saved as a persistent tag. Finally, we added a *"Show Untagged"* button in *v4*, which highlights documents and clusters where no tags have been applied, a crucial feature for the (initially unexpected) task of exhaustively reviewing a document collection.

This section described the design without providing any rationale for its evolution. Our decisions were based on observations of real world usage; we provide concrete examples of *why* and *how Overview* was used by journalists in Section 4.5. Then, in Section 4.6, we present our final task abstraction, the outcome of analyzing these observations, and justify our design choices with respect to these revisited tasks.

4.5 Observations of Real World Usage

We conducted six case studies where we analyzed the use of *Overview* by investigative journalists. We distinguish between a *case study* and a *usage scenario* [284], in which the former involves a person from the target application domain who uses a tool to examine their own data, having goals related to their ongoing work; in contrast, the latter reports usage of a tool by its designers with curated data and conjectured tasks.

Pilot case study: The first person who used *Overview* was the Associated Press Caracas bureau chief, whom we asked in November 2011 to use the *v1* prototype to examine 6,849 of the 251,287 U.S. State Department diplomatic cables released by WikiLeaks, those pertaining to Venezuela; this document collection is featured in Figure 4.4. Although he found the tool to be interesting, his analysis did not lead to a published story. This informal pilot case study revealed basic usability problems and the experience prompted us to formalize the case study process and determine foci of interest, such as utility, usability, learnability, and journalists' tasks in context.

Metrics: In addition to the qualitative analysis of journalists' tasks, we also focus on the metric of adoption defined as *self-initiated* use: did a journalist freely chose to use the tool for their own investigation, rather than trying out the tool in response to direct solicitation by the researchers? According to this distinction, adoption occurred in five of the six case studies we report, as indicated by the turquoise circles in Figure 4.2; the journalist in the remaining case study (IRAQ-SEC) was co-author Stray. We were also interested in the outcome of a journalist's investigation: did they complete their investigation to satisfaction as a result of using *Overview*, either by choosing to publish a story or by deciding that their findings did not merit a story? Or did they abandon *Overview* because the tool did not help further their investigation?

Recruitment: Since the *v2* deployment, Stray has promoted *Overview* within the data journalism community. Several hundred journalists have created accounts on the public server, and they have collectively uploaded more than nine million documents; *Overview* is used by approximately two hundred unique people each month[7]. As of April 7, 2016, we are aware of twenty published stories where *Overview* played a part in the investigative process[8], five of which are discussed as case studies below. The self-initiated journalists featured in case studies 2–6 were recruited to participate in our study after they contacted Stray with technical questions, which often pertained to workflow difficulties such as wrangling their document collection into a format that *Overview* could ingest.

Methods: Our case study findings are the result of triangulating between multiple data collection and analysis methods[9]. Our primary data collection method was that of a semi-structured interview[10]. We conducted interviews via Skype or Google^{+} Hangout, as our journalists were geographically remote; both services include a screen sharing feature, allowing journalists to demonstrate aspects of their investigative process. We recorded these interviews and demonstrations using a screen capture application and later transcribed them. The deadline-driven nature of journalism precluded multiple interviews during an ongoing investigation, so we chose to interview each journalist after their investigation was complete, despite the known limitations of retrospective introspection [93]. Journalists were encouraged but not expected to keep a diary relating to their ongoing use of *Overview*. Five of our case study journalists wrote or contributed to retrospective blog posts about their process [168, 307–309, 338], and one of them (TULSA) also sent us his personal notes.

We also collected usage logs for each journalist, consisting of timestamped interactions with *Overview*, which included `selecting`, viewing, and `annotating` documents and clusters with tags. Log file analysis allowed us to partially reconstruct a journalist's analysis process, complementing information divulged to us in their retrospective interview. Finally, each

[7] As of March 2014.

[8] Links to these stories can be found here: `https://github.com/overview/overview-server/wiki/News-stories`. A blog post that summarizes several of these stories is available here: `https://blog.overviewdocs.com/completed-stories/`

[9] Section C.3 provides additional detail regarding our data collection and analysis methodology.

[10] Our interview protocol is provided in Section C.4.

journalist provided us with their tagged document collection, which helped to establish a shared context.

4.5.1 Case Studies

The six case studies we present, summarized in Table 4.1, took place between February 2012 and December 2013, as indicated in Figure 4.2.

CS1: IRAQ-SEQ [306]: Our first case study took place in February 2012, when journalist and co-author Stray used *Overview v2* to analyze recently declassified documents from the Iraq war concerning the behavior of private security contractors. In particular, he wanted to categorize and count types of documented incidents involving these contractors; aside from the high-profile incidents that made headlines, he wanted to determine the prevalence of other incidents that these contractors were involved in during the Iraq war.

The document collection was the result of a FOIA request to the U.S. State Department, comprised of 666 incident reports over 4,500 pages, which were scanned using OCR. After the documents were loaded in *Overview*, Stray examined document clusters over the course of five days: he `navigated` the *Topic Tree*, `selected` clusters and their documents, `aggregated` clusters using the tree pruning controls, and `annotated` approximately 48% of the documents with 28 unique tags. After a lengthy "orientation" phase to determine incident categories of interest, he sampled the documents using the "Select Random" button (above the *Cluster List* in Figure 4.4), which would `select` a document from the *Document List* to be shown in the *Document Viewer*. With this approach, he read and tagged 50 of the 666 reports, which allowed him to develop hypotheses regarding the prevalence of certain incident types. Afterward, he followed up with U.S. State Department representatives, who provided additional context and a timeline for these incidents. His published story [306] combines his categorical summarization with the context of the war.

CS2: TULSA [339]: The first case of self-initiated adoption by a journalist took place in June 2012[11], revealing a different motivation for using

[11] Additional analysis of this case study is provided in Section C.3.

Overview. In this case, the journalist wanted to `locate` and `identify` evidence, documents that would support or refute a pre-existing hypothesis: he was following-up on an anonymous tip regarding municipal government mismanagement and potential conflicts of interest between city hall, municipal police, and police equipment vendors. He filed a FOIA request with the City Hall of Tulsa, Oklahoma for email messages between these organizations, and then used *Overview v2* to examine 5,996 of these email messages.

His search for corroborating evidence spanned multiple sessions over 18 days, beginning with an exhaustive and systematic left-to-right `navigation` of the *Topic Tree*, `aggregating` clusters using the tree pruning controls, and `selecting` clusters to view their contained documents. He viewed roughly 70% of the documents in the *Document Viewer* at least once, `annotating` 92% of them with 22 unique tags. We observed that he undertook multiple iterations of tagging: he began by tagging entire clusters using terms appearing in cluster keywords, but later tagged individual documents throughout the tree with tags such as *"important"*, *"weird"*, and *"follow-up."* As a result of this thorough tagging, the journalist was able to `lookup` and `browse` previously identified clusters or documents of interest, focus on documents `annotated` by multiple tags, or `locate` documents that remained untagged; the latter was accomplished by `selecting` uncoloured points in the scatterplot. These tags also provided a starting point for the further `annotation` of 129 *"important"* documents with notes relating to his hypothesis; these notes eventually became integral parts of his published story [339].

CS3: RYAN [110]: In October 2012, *Overview v2* was used yet again to `locate` evidence in support of a hypothesis, though there are several differences as compared to the TULSA case study. In this case, a journalist wanted to follow-up on an earlier story and on accusations made by Vice President Joe Biden that vice-presidential nominee Paul Ryan's campaign statements were hypocritical. In order to support or refute this hypothesis, the journalist sought to `compare` Ryan's campaign statements regarding wasteful government programs to his correspondence with various federal agencies concerning these same programs. After filing over 200 FOIA requests to these agencies, the journalist received 8,680 pages of correspondence. These physical documents arrived in several batches, and were scanned using OCR.

The journalist wanted to find genuine correspondence signed by Ryan; however, prevalent OCR errors prevented him from `locating` these documents using keyword search. *Overview* was able to cluster documents effectively on the remaining intact text, and most of the documents in this collection were quickly found to be irrelevant to his hypothesis. Over the course of half a day, he `navigated` the *Topic Tree* to `locate` and `identify` a small subset of clusters containing one hundred and seventy-six pages of genuine correspondence containing Ryan's signature; the remainder could be safely ignored, comprised of attachments and other irrelevant correspondence. Unlike the TULSA journalist, the RYAN journalist `annotated` a mere 8% percent of the document collection with 12 unique tags. As with TULSA, the RYAN journalist used tags as a starting point for the further `annotation` of his source documents with notes; his published story [110] compares these findings to Ryan's campaign statements.

CS4: GUNS [167]: The first documented adoption of *Overview*'s web application deployment (*v3*) took place in December 2012. Shortly after the Newtown school shooting, the journalist asked *Daily Beast* readers to self-identify as gun owners or non-owners, to report where they lived, and to post their opinion on the debate over gun ownership on a discussion board. He collected 1,278 comments: 757 from gun owners and 521 from non-owners. He aimed to determine what the debate on gun ownership is about: do gun owners and non-owners raise the same issues? He was also curious about geographical differences.

He uploaded the responses from gun owners and non-owners into two separate instances of *Overview*. Like the IRAQ-SEC case study, the GUNS journalist was interested in `summarizing` a document collection, though the form of this `summarization` was different. In IRAQ-SEC, the journalist wanted to categorize and count types of documented incidents; in contrast, the GUNS journalist sought to `identify` documents that were representative of their clusters, the sensational and polarizing speaking points from both sides of the debate over gun ownership; he was less interested in a fine-grained classification or quantification. For both sets of documents, he `navigated` and `selected` clusters and their contained documents, `compared` related clusters between the *gun owner* and *non-owner* instances, and later `browsed` previously identified clusters to `identify` representative quotes from people on both sides. Ultimately, he read nearly all the discussion board comments over the course of a day. Unlike the previous case studies, he did not use *Overview*'s tagging functionality, instead opting to copy quotes into an Excel spreadsheet, where he integrated geographical metadata and iteratively arranged quotes to construct a narrative for his story [167].

CS5: DALLAS: In August 2013, a journalist used *Overview v4* in a similar fashion to that of the TULSA journalist, though the outcome of their investigations differed. In the DALLAS case study, the journalist had recently reported on a collection of 4,653 email messages resulting from a FOIA request regarding the state government's response to an emergency incident. The journalist believed that some remaining evidence was left to be `located`, beyond what had already been reported in the earlier story. Despite having already read all the documents in the collection (unassisted by *Overview*), the journalist used *Overview* to verify that nothing was overlooked and sought to gather material for a follow-up story. She subsequently used *Overview* to examine four additional collections of messages, analyzed individually, ranging in size between 1,858 and 3,564 email messages.

The keyword search feature introduced in *Overview v4* was found to be particularly useful: the journalist alternated between `identifying` clusters by `navigating`, `aggregating`, and `selecting` nodes in the *Topic Tree*, and `locating` documents via keyword search, then `identifying` related documents. As her analysis progressed, we observed that the journalist relied more upon keyword search to highlight clusters of interest within the *Topic Tree*. She applied tags to each of the five document collections: the number of tags ranged between three and seven, and between 7% and 52% of documents were `annotated` with at least one tag; in total, 14 out of 31 tags were created from keyword search results.

In this case, *Overview* was used to make the decision *not* to publish: after 12 hours of *Overview* usage spanning several weeks, the journalist was sufficiently confident that nothing significant had been overlooked in the previous investigation, ultimately deciding not to write a follow-up story. This journalist estimated that it would have taken "more than a week" to reach this conclusion without *Overview*, and is "definitely planning on using it again for large document sets".

CS6: NY [236]: The final case study we report took place in December 2013, in which a journalist used *Overview v4* to confirm that a document collection *did not* contain evidence that would refute his hypothesis. In the NY case study, the journalist had gathered material to investigate the state of New York's process for handling and responding to police misconduct cases, including 1,680 proposed and passed bills retrieved from the State Senate Open Legislation application programming interface (API). He hypothesized that the state legislature had failed to pass any bills addressing this misconduct by increasing oversight.

A considerable amount of data wrangling was required before this journalists could use *Overview*. The State Senate API provided the bills in JavaScript object notation (JSON) format; to address this, the journalist wrote a script to `import` these documents into a database, which was in turn used to export a comma-separated values (CSV) file that *Overview* could ingest.

Following data ingestion, the journalist used Overview for about four hours over the course of three days to read *all* the document titles and keywords in a systematic fashion: starting with the smaller nodes, he would `select` a node in the *Topic Tree* and scan the document titles and keywords appearing in the *Document List*; the titles tended to be verbose and descriptive, and any that were deemed interesting were read in the *Document Viewer* or tagged as *"review"* . He eventually examined the largest node, which contained 732 documents with similar titles and keywords, their contents mostly comprised of boilerplate text; the journalist tagged the entire node as *"no unless"*, meaning that any document contained by the node was not significant unless there was another tag on it. He later returned to documents tagged with *"review"*, replacing this tag with one of five descriptive tags. Though the tag highlighting used in *Overview*'s *Topic Tree* allowed the journalist to quickly `locate` tagged documents, he suggested that the tree could alternatively hide all documents *not* marked with a particular tag, such as his *"not of interest"* tag.

His approach was similar to TULSA and DALLAS, in that they all sought to `locate` and `identify` clusters containing potential evidence. However, the TULSA and DALLAS journalists could have stopped their search once this evidence was found, as it is unlikely that any additional evidence would invalidate their previous findings. In contrast, the NY journalist sought to prove the *non-existence* of evidence, which required review of every document, as any evidence that went overlooked would have invalidated a claim of non-existence.

As a result of his analysis, the journalist was confident that no bills had been passed to address police misconduct, though several relevant bills had been proposed multiple times; conveniently, multiple versions of proposed bills were clustered together in *Overview*'s *Topic Tree*. While this finding is reported in a only a single paragraph of his published story [236], it played a key role in his argument that the state of New York is facing a police oversight problem; this story received considerable acclaim from the journalism community and was a finalist for the 2014 Pulitzer Prize[12].

4.5.2 Think-Aloud Evaluation

To complement our case study observations, we also solicited feedback from other journalists. After the deployment of the web-based *Overview v3*, which included usage tracking, we observed that *Overview* and its individual features were not being used to the extent that we had hoped. We suspected usability problems so we embarked on a discount usability testing program inspired by the work of Nielsen [224]: five naïve journalists were independently presented with an example document collection, such as the collection

[12] `http://www.pulitzer.org/finalists/5325`

Case Study	1: IRAQ-SEC [306]	2: TULSA [339]	3: RYAN [110]	4: GUNS [167]	5: DALLAS	6: NY [236]
Date	Feb. 2012 ○	Jun. 2012 ○	Oct. 2012 ○	Dec. 2012 ○	Aug. 2013 ○	Dec. 2013 ○
Version	*v2* / desktop	*v2* / desktop	*v2* / desktop	*v3* / web	*v4* / web	*v4* / web
Document Collection	666 reports / 4,500 pages from FOIA (scanned using OCR).	5,996 email messages from FOIA.	8,680 pages of correspondence from multiple FOIAs (scanned using OCR).	2 collections of online discussion board comments (757 in the first, 521 in the second).	5 collections of email messages from FOIAs, ranging from 1,858 to 4,653 messages.	1,680 proposed and passed bills retrieved with NY Senate Open Legislation API.
Task	**T1**: `generate hypotheses` → `explore` → `summarize`	**T2**: `verify hypotheses` → `locate` → `identify`	**T2**: `verify hypotheses` → `locate` → `identify`	**T1**: `generate hypotheses` → `explore` → `summarize`	**T2**: `verify hypotheses` → `locate` → `identify`	**T2**: `verify hypotheses` → `locate` → `identify`
Outcome	Summarized prevalence of document categories.	Located evidence supporting hypothesis.	Located a small subset of document clusters relevant to hypothesis.	Summarized using exemplar documents.	Could not locate evidence to support hypothesis.	Proved non-existence of evidence.

Table 4.1: A summary of the six case studies; deployment-phase case studies are represented as ○ (purple circles),while adoption-phase case studies are represented as ○ (turquoise circles).

featured in Figure 4.4, and asked to narrate their actions as they interacted with *Overview* using a think-aloud protocol[13], resulting in a qualitative understanding of usability problems.

All who participated in these think-aloud sessions found *Overview* to be confusing; much of this confusion was due to the visual complexity of its multiple-view interface, as well as a lack of affordances for common and critical interactions, such as `selecting` a document to read. We suspect that many previous document set visualization tools would face similar usability problems in real workflows, either by lacking a robust document `import` feature, or by not providing a means to read individual documents [70].

[13]Think-aloud protocols are limited in that they do not capture automatic, non-conscious reactions to stimuli. The retrospective introspection of our case study interviews is similarly limited. Interaction logs and screen captures may provide some indication of non-conscious reactions to stimuli. Eye-tracking equipment may provide additional indication of these reactions. However, these think-aloud sessions were performed opportunistically in non-laboratory settings, and we were unable to gather additional data.

An exception is Jigsaw, whose developers have noted and overcome similar problems [116]. In the next section, we discuss how the design of *v4* resolved these usability problems.

4.6 Analysis

Given our observations or real world usage, we now revisit our initial task abstraction and discuss the rationale for *Overview*'s design.

4.6.1 Task Abstractions Reconsidered

After the GUNS case study, we struggled to distinguish between journalists' goals, approaches, and outcomes. Specifically, the TULSA and RYAN journalists used *Overview* in more directed and systematic ways that we did not anticipate, in that they sought to `locate` specific evidence or a subset of clusters that were relevant to a pre-existing hypothesis, forcing us to reconsider our initial task abstraction of *"exploring"* a document cluster structure, which was based on the WARLOGS use case. A number of previous tools aim to help people to *"explore"* a document collection (e.g., [52, 70, 80, 91]), though few of these tools have been evaluated with people who work in a specific target domain who bring their own data, making us suspect that this imprecise term often masks a lack of understanding of the tasks that people perform.

In48 Chapter 2, we proposed a typology of abstract visualization tasks, the purpose of which was precisely to articulate such differences in the use of visualization tools at multiple levels of abstraction. According to this typology, a task description is broken down into *why* data is visualized, *what* dependencies a task might have, and *how* the task is supported. *How* is somewhat orthogonal to *why*, as exemplified by the differences in usage reported in the previous section. We applied this typology to the coding of our observational data, characterizing two different tasks, **T1** and **T2**, that replace and improve upon our initial task abstraction. In this section, we will use the vocabulary and notation of this typology to focus on *why* and *what*; in Section 4.6.2, we analyze *how Overview* supports these tasks.

T1: `generate hypotheses` → `explore` → `summarize`: When approaching a collection of leaked documents or a corpus of social media content, a journalist may have little prior knowledge regarding the collection's content, eliciting a need to `generate hypotheses` and to ask *"what's in this collection?"*. To support the generation of hypotheses, a journalist must be able to `explore` a document collection and `summarize` clusters of documents. The term `explore` is defined more precisely in our typology as a form of `search` in which neither the identity nor the location of a search target are known a priori. In the context of a document collection, a search target may be content within a document, an individual document itself, a cluster of related documents, or an arbitrary set of documents and clusters. `Exploring` is distinguished from `browsing`, in which the location of a search target is known but its identity is not, `locating`, in which the converse is true, and `lookup`, in which both the location of the target and its identity are known. The result of `summarizing` is a compressed representation of the full contents of the document collection, such as the categories and counts produced in the IRAQ-SEC case study, or the exemplar documents that the journalist ultimately `selected` in GUNS case study.

T2: `verify hypotheses` → `locate` → `identify`: In contrast, a journalist who asks for documents via FOIA request typically has some pre-existing hypotheses, and their aim is to `verify`, `refute`, or `refine` these hypotheses by `locating` evidence. In these cases, a journalist likely has a sense of what the documents are about, but they may not be able to specify the evidence they seek in terms of a standard keyword search (for example, "corruption" would not suffice), and there may also be unexpected but valuable material waiting to be discovered. In the language of our typology, the aim is to `locate` and `identify` clusters containing potential evidence, beginning with those labelled by interesting keyword terms; alternatively, a journalist will `locate` documents containing specific search terms and subsequently `browse` and `identify` related documents found in the same cluster. **T2** describes the use of *Overview* in the TULSA, RYAN, DALLAS, and NY case studies.

Throughout both **T1** and **T2**, a journalist will often `produce` notes or annotations for documents as they `generate`, `verify`, or `refine` their hypotheses, perhaps `comparing` documents to each other or to secondary sources outside of the collection. **T1** and **T2**, along with the subsequent annotation task, are illustrated using the visual notation of our typology in Figure 4.7 and Figure 4.8, respectively, in the addendum (Section 4.9) at the end of this chapter.

4.6.2 Design Rationale

With a more precise understanding of journalists' tasks, we now analyze the rationale for our visual encoding and interaction design choices in the hope that our choices may transfer to other domain problems involving similar data and task abstractions.

Why show a tree? *Trees afford structured and systematic exploration.* When a document collection contains separable clusters of similar documents, a tree-based visual encoding affords a systematic and, if desired, exhaustive traversal of these clusters. The TULSA case study is an example where the journalist based his choice of which documents to read based on the structure of the tree, sweeping from left to right. The *Topic Tree* also includes a visual encoding of applied tags, which makes it possible for a person to `identify` the documents they have and have not already tagged, and how tags correspond to clusters.

How to show a tree? *Emphasize interior nodes (not edges or leaves); instill trust in the underlying algorithm.* The *Topic Tree* in the first two versions of *Overview* (Figure 4.4) rendered all clusters as identical nodes. While tree-based visual encodings are often associated with the task of path tracing and determining connectivity [186], people who use *Overview* are primarily interested in the properties of nodes corresponding to document clusters, such as the number of documents contained by a cluster, or the key terms that describe these documents.

The *Topic Tree* of *v3-4* directly `encodes` cluster size as node width; this design choice allows a person to `compare` cluster sizes directly, or work systematically from larger to smaller topics, of particular use when trying to `summarize` a document collection to some desired degree of detail (**T1**). In enlarging the width of nodes, it also became possible to `encode` the number of documents tagged within the cluster as a colour label having a width proportional to the size of the node, as shown in Figure 4.5. We opted not to use a space-filling treemap visual encoding of hierarchical document clusters because this approach would place too much emphasis on leaf nodes; when `summarizing` a collection (**T1**) or when `locating` a subset of documents (**T2**), the mid-level interior nodes in the tree are typically the most informative. While less space efficient, a tree with variable-width nodes provides more flexibility, especially given the differences between **T1** and **T2**.

With larger nodes, we were able to display cluster keyword terms directly in the node itself, rather than in a separate *Cluster List* view, as in *v1-2*, which displayed keywords only for the selected cluster and its descendants. Displaying keywords within nodes allows a person to `compare` keywords at a glance, both between and within clusters.

The design of the *Topic Tree* required a balance between usability and cluster fidelity: in *v3*, there was no limit on the number of children allowed for each node, a situation reported to be overwhelming by journalists who participated in the think-aloud evaluation. When a node can have many children, tree exploration reduces to linear search; for this reason, the maximum number of child nodes was limited in *v4* by switching to a recursive adaptive K-means algorithm [238] with an upper limit of five children per node.

We added explicit cluster fidelity labels to the *Topic Tree* nodes in *v4* to help people interpret the content of a cluster: the labels *"Some"*, *"Most"*, and *"All"* show how many documents in a cluster contain each keyword and thus signal the consistency of topics found in that node, as shown in Figure 4.5. These labels help a person to decide whether to treat the node as a conceptual unit that might be tagged as a whole, or expand it to examine its children individually. They help a person assess cluster consistency and separability and serve to build trust in the clustering algorithm [61]. Previously, people had to judge the topical consistency of a cluster by examining the individual documents within it, or by referring to the scatterplot in a way that Sedlmair et al. [286] demonstrated to be difficult for dimensionally reduced data.

How to interact with a tree? *Selective pruning and informative tooltips.* In *v1-v2*, people were able to clarify the *Topic Tree* by pruning (`aggregating`) small nodes, according to a threshold `selected` from a set of seven coloured radio buttons below the *Topic Tree*. Our case studies revealed that many people never understood that the variable tree-pruning threshold used in *v1-v2* was hiding nodes from them, a problem especially for those intent on `locating` evidence or proving the non-existence of evidence (**T2**). We replaced threshold-based node pruning with a selective expand/-collapse option on each node; when combined with panning and zooming, these interactions provide people with a fine-grained control over focus and context.

Upon `selecting` a node in *v1-v2*, keywords for the selected cluster and its descendants were shown in a status bar between the *Topic Tree* and the *Cluster List*, however these were spatially removed from a person's point of focus. To resolve this, we added tooltips in *v3* that show cluster keywords when the cursor hovers over a node.

Why no scatterplot? *Unstructured exploration is redundant for* **T1** *and* **T2**. Using scatterplots to visualize document collections is an approach common to previous work [38, 57, 91, 118, 232]. We thought that a scatterplot would allow people to judge cluster size, quality, and separability [286]. However, scatterplots do not directly show cluster content, such as document keywords, unless tooltips or point aggregation is used. We did not pursue the use of these design choices because we discovered that the scatterplot was seldom used in the case studies of journalists who adopted *Overview*. The TULSA journalist was an exception in that he used the scatterplot to `locate` untagged documents containing potential evidence after extensive use of the *Topic Tree* (**T2**). This task would have been better served by providing a direct way to show how many untagged documents a node contains; the *"Show Untagged"* button introduced in *v4* accomplishes this.

Ultimately, we realized that a scatterplot does not help people overcome the burden of choice overabundance when determining which cluster to investigate next [281], whereas the tree-based hierarchical clustering used in the *Topic Tree* affords a form of structured `navigation`. In addition, the cluster fidelity labels introduced in *v4* help people to assess cluster consistency and separability, thereby eliminating any further need for a scatterplot.

Why tags? *Tags provide simple annotation, progress tracking, and human-defined semantics.* Tagging was used extensively in five of the six case studies. Some tags aligned with cluster boundaries (IRAQ-SEC, RYAN, and the first set of tags created in TULSA), while other tags appeared throughout the tree (DALLAS, NY, and the second set of tags created in TULSA). Tags are a simple and flexible form of `annotation` that help people track where they have been and what they have learned. They can also be used to impose a context-specific organization scheme on a document collection. No single clustering will meet all analysis needs, since any high-dimensional dataset is likely to have multiple cross-cutting semantically interesting clusterings [121]. The "best" clustering will depend on the documents and the story. *Overview* does not support manual re-arrangement of the *TopicTree* hierarchy, as in tools such as HierarchicalTopics [80]. Instead, we support manual tagging as a simple and flexible way for people to impose their own semantics on a document collection, where the cluster structure can be leveraged as a useful scaffold when it matches a person's mental model, but ignored when it does not. The most recent feature added to *Overview*, developed after the case studies in this chapter, supports the creation of multiple trees, giving different views of same document collection. A person can control the clustering by entering words to ignore, which prevents *Overview* from clustering based on document letterhead or boilerplate text, and by entering especially important words which are weighted higher when constructing document vectors. It is also possible to create a tree containing only a subset of the documents, specified by `selecting` an existing tag.

Multiple views: how many and how to coordinate? *Less is more; provide obvious affordances.* The evolving design of *Overview* recalls the challenges and considerations for designing multiple view tools, which we have discussed elsewhere [181]. These considerations include determining the appropriate number of discrete views, the arrangement of these views, and how these views should be coordinated. A consensus on these questions has not yet been reached, as multiple-view visualization tools range from dual-view to over twenty, with a similar range in view coordination patterns [348].

The CARACAS pilot case study with *Overview v1* revealed that the *Document Viewer* was too small and the `selection` of documents and clusters across the different views was poorly coordinated. Despite improvements to view coordination in *v2* and *v3*, the views were not always well understood; for example, those who participated in the think-aloud evaluation did not initially realize that the *Document List*, displayed as a line of keywords for each document, was in fact a list of `selectable` documents. In *v4*, *Overview*'s interface was streamlined into three views coordinated with linked `selection` and highlighting: the *Topic Tree*, a consolidated *Document Viewer/List* featuring document titles and list `navigation` controls, and a list of *Tags*; as described above, we removed the scatterplot and *Cluster List*, both made redundant by the redesigned *Topic Tree*. While we might have made *Overview* a single-view tool, we instead reasoned that having the *Topic Tree* visible provides helpful context when reading a document and deciding what to read next.

How to support workflows? *Simplify for infrequent use; reduce data wrangling.* Our findings show that simplicity and learnability are critical for journalists, because any one journalist only deals with large document collections intermittently.

After *Overview v2* was deployed and promoted within the journalism community, it became clear that many people had great difficulty downloading, installing, and configuring it. Additionally, *v2* could only `import` documents in a CSV file; we quickly learned that journalists receive document collections in every conceivable format, from stacks of paper to database dumps. We confirmed that the need to wrangle data into compatible formats is a considerable barrier to adopting a visualization tool into an analysis workflow, as discussed in a recent research agenda [162]. We should not expect journalists to write custom data wrangling scripts, as the NY journalist had to do. To minimize the amount of configuration and wrangling required, the web-based *Overview v3-v4* required no installation and supported `import` from a folder of portable document format (PDF) documents or from DocumentCloud[14], a document hosting service used by journalists which can itself ingest a wide variety of formats. Without this integration, we suspect that the DALLAS journalist would have been unable to make use of *Overview*. The DocumentCloud interface is integrated into *Overview*'s *Document Viewer*, which includes a function for `annotating` documents with notes.

We also added full-text keyword search in *v4*, as members of the journalism community and our case study journalists had expressed a desire to flexibly alternate between `locating` clusters in the *Topic Tree* and a directed search for `locating` documents of interest, without having to use a search tool such as grep or DocumentCloud's search interface. We expect that the TULSA and RYAN journalists, both performing **T2** with *v2*, would have benefited from this keyword search feature, and that the absence of this feature would have been a deterrent for the DALLAS and NY journalists, who also performed **T2**.

We note that the use of *Overview* forms part of a larger investigation and reporting workflow: each case study journalist combined its use with many computer-assisted and non-computer-assisted methods for data collection, data transformation, and eventual story presentation.

Summary: This section presented seven lessons or guidelines that can be applied to the design of visualization tools for document data in particular, though we expect that these lessons may also be transferable to tools that address other forms of hierarchically-structured data:

- Why show a tree? *Trees afford structured and systematic exploration.*
- How to show a tree? *Emphasize interior nodes (not edges or leaves); instill trust in the underlying algorithm.*

[14] `http://documentcloud.org/`

- How to interact with a tree? *Selective pruning and informative tooltips.*
- Why no scatterplot? *Unstructured exploration is redundant.*
- Why tags? *Tags provide simple annotation, progress tracking, and human-defined semantics.*
- Multiple views: how many and how to coordinate? *Less is more; provide obvious affordances.*
- How to support workflows? *Simplify for infrequent use; reduce data wrangling.*

4.7 Discussion

In this section, we discuss the value and logistics of conducting field studies involving multiple deployments and the analysis of adoption, as well as the limitations of this type of research.

Why study adoption? As with any iterative human-centred design process, it is difficult to know when to declare success; we consider adoption defined as repeated instances of self-initiated use to be a form of success. Adoption is particularly interesting in the domain of journalism because tool use is a decision made separately by each journalist for each story, rather than dictated by a central authority.

Though the visualization design process is cyclical and may include multiple deployments, there are surprisingly few papers that comment on the adoption of a visualization tool without the prompting of designers: in a recent survey of eight hundred visualization papers containing an evaluation component, only five commented on adoption [183]. Of these, two provide a thorough description of who adopted their visualization tool, how it had been used, whether it was still in use, and what problems people reported [170, 205]. Our work adds to this short list, as does the recent study of Jigsaw's adoption [165].

Many visualization design studies report on deployment to a target group of individuals and evaluate their reaction to the tools during a period of intense study, but our own experience and personal communication with other practitioners leads us to believe that visualization tool use typically drops off after a paper is submitted. Gonzalez and Kobsa [113] provide a rare example of explicitly checking back after this time period; they found that their target population did not adopt the visualization tool despite the promising initial results reported in their original paper, and conjecture that a misunderstanding of workflow was the primary factor behind this lack of adoption. We conjecture that this situation might be the common case, and that longer-term adoption rates may be very low for research prototypes. Sustained follow-up by researchers until adoption is achieved provides a way to disambiguate whether the barrier to adoption is truly only a workflow issue, or an indication that the tool failed to address the true needs of the people for whom the tool was built.

The logistics of studying adoption: Before *Overview* was deployed, we could not have fully predicted *why* and *how* journalists would approach large document collections; we were unable to verify the correctness of our task and data abstractions. *Overview* is sufficiently novel that its value could not be assessed without adequately functional prototypes. We argue that this situation is common in visualization because of the complexity of the data and tasks at play, recalling the argument of Lloyd and Dykes [195] about the need for *data sketches*: functional example designs for establishing context of use and eliciting requirements[15]. They contrasted their design-first approach, illustrated by the red trajectories in Figure 4.6, to a design-then-evaluate approach (the green trajectory) and to an approach grounded in a person's context [153] (the blue trajectory). The purple trajectory illustrates that close collaboration with domain experts from the very start of a project means they bring expertise about requirements to the table, so it is not a design-first endeavour. The multiple loops in the purple trajectory emphasize the importance of deploying a visualization tool as a precursor

[15]Unlike typical HCI approaches such as paper prototyping, functional example designs in a visualization context entails the development of interactive software prototypes.

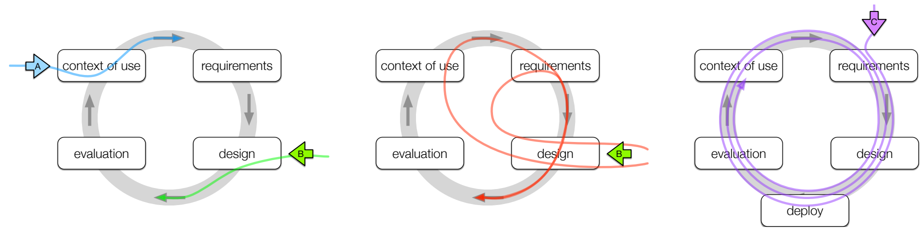


Figure 4.6: The human-centred design process development cycle, in which Lloyd and Dykes [195] discern between alternative entry points (A,B) and between design-then-evaluate (green), grounded (blue), and their own approach in which example designs are used to establish context of use and elicit requirements (red). In contrast, we begin with some requirements at point C, and only after multiple deployments do we arrive at a clear understanding of context of use (purple). Figure adapted and extended from Lloyd and Dykes [195].

to evaluation, and we note that after these loops the trajectory ends at context of use: it took several deployments and case studies of self-initiated journalists who adopted *Overview* before we attained a clear understanding of journalists' tasks and their broader analysis workflows.

Our use of case studies to study adoption is methodologically similar to qualitative longitudinal evaluation studies described in previous work [195, 277, 292]. Our approach differs from these in that we engaged a different set of people at each stage of design, rather than the same set of people. This difference reflects *Overview*'s context of use: repeat usage cannot be predicted and *Overview* is only appropriate for some investigations; we have yet to encounter a journalist who specializes in investigations pertaining to large document collections.

4.8 Summary

We presented a field study of *Overview*, an application for the systematic analysis of large document collections. *Overview* has proven to be useful, having been used in investigations leading to at least nine published stories. Using the typology of visualization tasks proposed in Chapter 2, we characterized two task abstractions based on findings from six case studies. Given our data and task abstractions, we rigorously analyzed the effectiveness of *Overview*'s visual encoding and interaction design. Our analysis may transfer beyond the domain of journalism, and speaks to the design and evaluation of other visualization tools for supporting the analysis of document collections and clustered dimensionally reduced data. Finally, this work adds to the small number of studies found in the visualization literature that include observations of adoption[16]; in observing real world usage by self-initiated journalists, we presented detailed evidence to support the contention that that several iterations of design and deployment are required before fully understanding *why* and *how* a visualization tool will be used in practice.

4.9 Addendum

Figure 4.7[17] illustrates the two tasks that we characterized in Section 4.6.1; colours correspond to the *why-what-how* structure of our task typology (see Figure 2.1)[18]. Figure 4.7 indicates the *how* and *what* aspects of these two tasks explicitly, as this information was discussed implicitly throughout this chapter. In **T1**, the `input` is a document collection and the `output` is evidence to either refute or verify a hypothesis. In **T2**, the `input` is once again a document collection, however the `output` is a set of exemplar documents or a set of clusters that can be used to `summarize` the entire collection. *Overview* supports both of these tasks via `encoding` the document collection as a tree, by allowing people to interactively `navigate` this tree structure, to `select` clusters and individual documents, to `aggregate` (or disaggregate) clusters of documents using the interactive expand and collapse controls on each node of the tree.

Figure 4.8 depicts a task that follows either **T1** or **T2**, in which the person who uses *Overview* will `annotate` clusters or individual documents

[16] While studies of adoption and appropriation may be common in the HCI community, they are rare in the visualization literature.

[17] Figure 4.7 and Figure 4.8 did not appear in our paper about *Overview* [35].

[18] We use the revised typology described in Section 3.5.

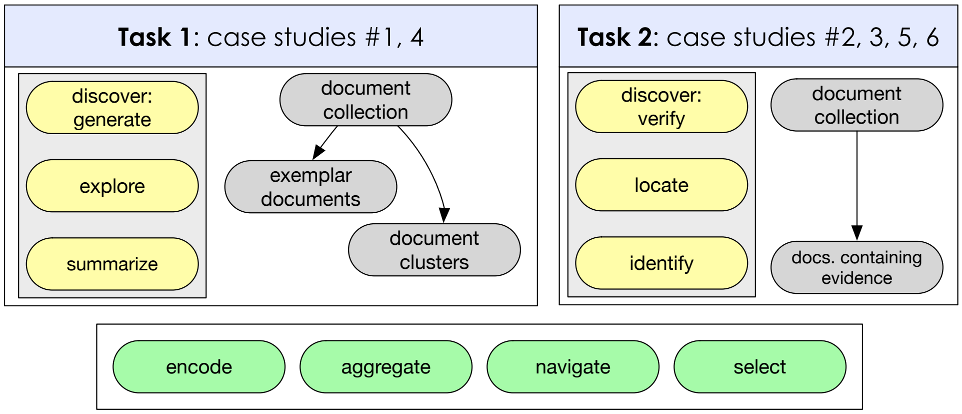


Figure 4.7: The two tasks characterized in Section 4.6.1; colours correspond to the *why-what-how* structure of our task typology (see Figure 2.1), where yellow is *why*, green is *how*, and grey is *what* defined in terms of `inputs` and `outputs`.

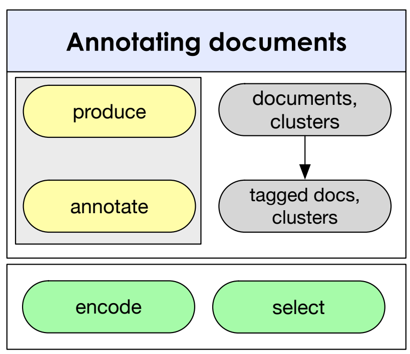


Figure 4.8: The task of `annotating` documents and clusters (which follows **T1** or **T2**); colour conventions follow that of Figure 4.7.

with meaningful tags (a `produce` action)[19]. To `annotate` documents or clusters with tags, a person `selects` documents or clusters, `selects` a tag, and the tag is `encoded` as a coloured label for the document or cluster.

[19] As per the discussion in Section 3.5, we now view `annotate` as a form of `produce`, as indicated in Figure 4.8.

Chapter 5

Design Study:

Matches, Mismatches, and Methods: Multiple-View Workflows for Energy Portfolio Analysis

[1]The energy performance of large building portfolios is challenging to analyze and monitor, as current analysis tools are not scalable or they present derived and aggregated data at too coarse of a level. We conducted a visualization design study, beginning with a thorough work domain analysis and a classification of data and task abstractions. We describe visual encoding design choices for time-oriented data framed in terms of *matches* and *mismatches*, as well as considerations for workflow design. Our designs address several research questions pertaining to scalability, view coordination, and the inappropriateness of line graphs for derived and aggregated data. We also present guidelines relating to *familiarity* and *trust*, as well as methodological considerations for visualization design studies. Our designs were adopted by our collaborators and incorporated into the design of an energy analysis software application that will be deployed to their large client base.

[1]This chapter is a slightly modified version of our paper *Matches, Mismatches, and Methods: Multiple-View Workflows for Energy Portfolio Analysis* by Matthew Brehmer, Jocelyn Ng, Kevin Tate, and Tamara Munzner; in IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis 2015), 22(1), p. 449–458 [37]. `http://dx.doi.org/10.1109/TVCG.2015.2466971`. Section 5.12 is a new addendum section that is unique to this dissertation.

5.1 Motivation

Consider a university campus containing about a hundred buildings. For a university operations worker looking for opportunities to conserve energy, visualization can be helpful when analyzing and monitoring the energy performance of this large portfolio of buildings.

In this design study, we collaborated with a team of people at EnerNOC, a company that develops energy analysis and reporting software for organizations such as commercial business chains, universities, and utility companies. Our collaborators' goal was to deploy a redesigned version of Energy Manager, their energy analysis software tool; in doing so, they hoped to retain their existing client base encompassing thousands of organizations, attract new clients, and increase engagement with their software. Meanwhile, our goal as researchers was to successfully integrate our research process into our collaborators' software development practice. We designed and evaluated potential visual encodings and interactions with a variety of stakeholders in an industry setting, which included our collaborators' colleagues as well as their clients. This chapter documents a success story, where our collaborators committed software development resources and adopted our designs that resulted from our research.

Visualization researchers and practitioners working in domains unrelated to energy analysis will find several transferable aspects of this chapter, beginning with our classification of data and task abstractions. This project required that we design and promote sophisticated visual analysis by individuals accustomed to unsophisticated visual encoding design choices. We needed scalable visual encoding and interaction design choices that can accommodate dozens to hundreds of concurrent time series. To complicate matters further, we could not rely upon the visual encoding design choice of a line graph due to a combination of data semantics and domain convention. Once we found appropriate mappings between visual encoding design choice and individual tasks, we next addressed the question of accommodating task sequences: determining which visual encodings ought to be juxtaposed in the same display as multiple views on the data, and which ought to be presented sequentially. Coordinating these views also posed a challenge; specifically, we sought to reduce the amount of manual navigation between views, as this was an issue with the previous version of Energy Manager.

The primary contribution of this chapter is a set of design choices and guidelines framed in terms of *matches* and *mismatches* between abstractions and visual encoding design choice for time-oriented data, guidelines that transfer beyond the energy management domain. We also present guidelines relating to the themes of *familiarity* and *trust*. The former refers to the spectrum between ubiquitous visual encodings and those that a person may have never seen before, while the latter refers to the appropriate display of derived and aggregated data, as well as giving people control over these data transformations. Finally, we contribute methodological advice for visualization design projects. This includes considerations for designing workflows that incorporate multiple views; while prototyping the visual encoding design of a single view has received considerable attention in the literature [195], workflow prototyping has received far less.

5.2 Methodology

In this section, we focus on how we conducted our research; we reflect upon our methodological decisions and offer advice in Section 5.10.2.

Analyzing the work domain: Over the course of five months in 2013, we conducted nine in-depth interviews with people who have used Energy Manager. Eight of those interviewed were employed by client organizations, which included three North American universities, an engineering consulting firm, and a school board. The final interviewee was a colleague of our collaborators who regularly consulted with new clients. Although we use the term *energy analyst* to describe these individuals, we encountered a wide range of roles, job titles, skill sets, and levels of training. The use of energy analysis tools such as Energy Manager also varied in terms of context and frequency of use. Despite these differences, we did find several common goals and activities relating to the energy analysis of building portfolios, which we classify in terms of data and task abstractions in Section 5.3.

Validating the abstractions: We validated our abstractions by checking back with a subset of the people that we had previously spoken to during the work domain analysis phase: our primary collaborators and the two energy analysts who provided the most feedback during the work domain analysis phase. To do so, we consolidated our thoughts and findings into a slide deck that contained screenshots, examples, mockups, and notes. These slides were a living research artefact: we used them to present previous findings to collaborators and energy analysts, and we recorded their feedback as further annotations.

Eliciting feedback on visual encoding designs: We developed an interactive sandbox prototyping environment that allowed us to rapidly explore the design space of visual encodings, which we describe in Section 5.6. This phase of the project lasted about four months. In addition to weekly design feedback sessions with our collaborators, we conducted two 60 to 90 minute design feedback sessions with the two energy analysts that had participated in the prior two phases, as well as two feedback sessions with energy analysts that we had not previously spoken to.

We continued with the method of slide decks as research artefacts. For each session with an external energy analyst, we created a personalized slide deck that included screenshots from our sandbox environment along with explanations; these slides were sent to energy analysts in advance. Recognizing the importance of showing project stakeholders their own data [195], these slides featured data from the energy analyst's own portfolio of buildings. Since many of the energy analysts that we consulted with are located in other cities or countries, many methods for participatory design and evaluation methods [114, 204] that depend on the researchers and stakeholders being co-located could not be implemented, and we have previously commented on the methodological implications of this situation [34].

During these sessions, we shared our screen and conducted *chauffeured demos* [195] using our visualization sandbox: we would present the energy analyst with data from their own portfolio and ask them to step through their energy analysis workflow with our alternative visual encoding designs.

These sessions were recorded for further analysis and their feedback was later transcribed as annotations on the session's slide deck. The result of this phase was the identification of a set of *matches* and *mismatches* between visual encoding design choice and individual tasks, which we discuss in Section 5.7.

Prototyping workflows: Based on feedback collected during the chauffeured demos with energy analysts, we prepared storyboards of these workflows using sandbox screenshots and mockups. We then continued our design process by considering how to best juxtapose, link, and sequence multiple views on the data, all while continuing to consult with our collaborators and energy analysts. This phase of the project resulted in the workflow described in Section 5.8 and realized in the redesigned Energy Manager described in Section 5.9.

Example artefacts: We generated 11 slide decks containing a total of 302 slides over the course of this project. We include example slides in Section D.2[2], along with other research artefacts.

5.3 Abstractions

Following the work domain analysis phase, we recast activities from domain-specific terminology to data and task abstractions.

5.3.1 Data Abstraction

The energy analysts to whom we spoke oversee the energy performance of dozens to hundreds of buildings, which are referred to as *portfolios*. We now abstract a portfolio of buildings and its associated time-oriented energy data, which we summarize in Table 5.1.

Building metadata: Consider a university, where buildings vary by area, age, and number of occupants. They can also be differentiated using any number of categorical *tags*, such as the type of building (*"lecture hall"*,

[2] Additional supplemental materials for this chapter, including high-resolution versions of the figures and a video, are available here: `http://cs.ubc.ca/labs/imager/tr/2015/MatchesMismatchesMethods/`

Term	Abstraction	Example
	Building metadata	
Building ID	unique categorical	#123
Building area	quantitative	450 m^2
Building age	quantitative	20 years
# occupants	quantitative	50 people
Location	spatial	49.26° N, 123.25° W
Tag	categorical	*"restaurant"*
	Temporal data for each building	
Energy demand	quantitative	200 kilowatts (kW)
Outdoor temperature	quantitative	18° C
Open / closed	categorical	Open Mon–Fri, 08-18h
	Derived temporal data for each building	
Consumption	quantitative	800 kilowatt-hours (kWh)
Energy intensity	normalized quantitative	1.78 kWh / m^2
Weather-independent performance†	normalized quantitative	50 kWh / heating degree day (HDD)
Predicted perform.†	quantitative	190 kW
% Savings	normalized quantitative	40%
Rank	ordinal	1st, 2nd, 3rd

Table 5.1: Data abstraction summary. †*Performance* could be assessed in terms of *demand*, *consumption*, or *intensity*.

"laboratory"), its campus or department (*"chemistry"*, *"physics"*), or the name of its building operations manager.

Building groups: Given all of this building metadata, we can have groups of buildings with shared attribute values or ranges. A portfolio may have many building groups, and they may overlap.

Multiple time series per building: The energy performance of each building in a portfolio is monitored over time, and each building has multiple time series associated with it. A building may consume several forms of energy, such as electricity, natural gas, or steam. Many non-residential buildings are equipped to report raw energy demand values at the granularity of minutes, along with outdoor air temperature. Finally, building opening and closing hours are also recorded.

Derived data: Raw continuous energy *demand* data is typically examined during detailed investigations of single buildings at fine granularities of time; for instance, a building might exhibit an unexpected spike in demand on a Sunday morning. However, for an energy analyst overseeing a large portfolio of buildings, analysis tasks typically revolve around *derived* and *aggregated* data, rather than raw continuous data. These may include *averages*, *minimums*, and *maximums* for different temporal granularities of interest, such as the average weekday electricity demand in January. *Consumption* is the most common of these derived attributes, which is an accumulated amount of energy. *Intensity* is consumption normalized by a building's area, which allows the energy analyst to directly `compare` the energy performance of buildings of different size. Similarly, it is possible to normalize energy performance values by considering outdoor temperature, which allows the energy analyst to `compare` buildings at different times of the year, or buildings in locations with different climates. *Predicted energy performance* based on statistical models is also considered, though predicted values are problematic for reasons we describe in Section 5.4. *Relative* and *absolute differences* between the observed energy performance and the predicted or historical performance are also considered; for example, the energy analyst can determine how a building is performing this year relative to its performance last year. Finally, given any of these derived values, the energy analyst can `compare` multiple *rankings* of buildings, allowing them to `identify`, for instance, the buildings most in need of an energy conservation measure.

In summary, we have a multidimensional table of buildings and building attributes, along with quantitative energy-related attributes with values changing over time.

Domain convention: The energy analysts to whom we spoke are accustomed to interpreting any encoding that incorporates a continuous line graph as raw energy *demand*, and thus it is inappropriate and potentially misleading to use such encodings to display derived and aggregated time series values such as *energy consumption* or *intensity*. In the previous version of Energy Manager, `derived` and `aggregated` values are `encoded` in bar charts or listed in tables; we will discuss in Section 5.4 how these encodings do not scale, and in Section 5.7, we examine alternative encodings.

5.3.2 Task Abstraction

Energy analysts need to balance reducing costs and conserving energy while ensuring the comfort and safety of people who use buildings in their portfolio. To achieve these goals, the energy analysts need to: (i) assess the performance of buildings in their portfolio following the implementation of an energy conservation measure, such as installing new windows or lighting; (ii) determine which buildings in their portfolio require energy conservation measures; and (iii) find and diagnose anomalous energy performance such as spikes, outages, surges, or otherwise erratic and inconsistent behaviour.

We classify these activities as abstract visualization tasks according to the typology introduced in Chapter 2. We further distinguish between `action` and <u>`target`</u> terms[3], a distinction introduced in Munzner's extension to our typology [219], where a <u>`target`</u> is the `output` of the task; it can also be helpful to think of `actions` as verbs and <u>`targets`</u> as nouns[4]. Each of the tasks can be described by a need to `discover`: to generate and verify hypotheses. While energy analysts also occasionally `present` their findings to colleagues and other stakeholders, our current focus is predominantly on the discovery process.

T1 / Overview: An individual will `lookup` and `summarize` time-oriented data from all the items spanning a coarse period of time. Of interest are <u>`trends`</u>, <u>`outliers`</u>, <u>`distributions`</u>, <u>`extreme values`</u>, and <u>`similarities`</u>. This abstract task corresponds to the domain activities of assessing overall energy performance and determining which buildings in a portfolio require energy conservation measures. A concrete example of this task would be an energy analyst who asks: *"How did the entire university perform this past year? Do any buildings stand out?"*

T2 / Drill Down: The second task is the result of drilling down from the portfolio to a group of buildings and to a finer temporal resolution, examining energy performance in detail: a person `locates` and `compares`

[3]We underline <u>`target`</u> terms to distinguish them from terms that appeared in our original typology.

[4]Munzner's distinction between `actions` and <u>`targets`</u> [219] is summarized in Figure 6.1 and Figure 6.2, respectively.

`trends`, `outliers`, and `features` in time-oriented data for items in the group. This abstract task corresponds to the domain activities of assessing building performance following an energy conservation measure, or finding and diagnosing anomalous energy performance, wherein a `feature` of interest could be a spike or sudden outage. One concrete example is *"Are my restaurants in Vancouver performing better this January than they did last January?"*

T3 / Roll Up: The third task is the converse of drilling down from the portfolio to a group and from the group to individual buildings. This task is one in which a person `explores` and `locates` `trends`, `outliers`, and `features` in time-oriented data to `identify` the proportion of individual behaviour relative to the group's behaviour, or the `dependency` between the aggregate amount and the individual contribution. Concrete examples of this task include *"What proportion of a university's energy consumption is consumed by its computer science building over time?"* or *"Which science faculty building consumes the largest proportion of energy?"*

Task sequences: These three tasks are not performed in isolation, but in sequential energy analysis workflows, and some task sequences occur more often than others. We know from our interviews that the frequency of this work varies, so the Overview (**T1**) task may be more prevalent among energy analysts who infrequently analyze their portfolio's energy performance. Others may skip the Overview task because they have a specific question about a group of buildings, proceeding directly to the Drill Down (**T2**) or Roll Up (**T3**) tasks. An energy analyst will alternate between the Drill Down and Roll Up tasks, as the completion of one task may prompt new questions.

T1, **T2**, and **T3** are illustrated using the visual notation of our typology in Figure 5.10 in the addendum (Section 5.12) at the end of this chapter.

5.4 The Previous Version of Energy Manager

We now analyze the previous version of Energy Manager and determine the extent to which the tasks are supported.

The previous version of Energy Manager, shown in Figure 5.1 and Figure 5.2, is a multiple-page web-based application, one that uses a small number of visual encoding design choice adhering to the domain convention mentioned in Section 5.3: bar charts and tables are used for derived aggregate values such as *consumption*, while line graphs are used for continuous values such as *energy demand*. The main page presents a summary dashboard for a portfolio of buildings (Figure 5.1). At the bottom of this dashboard is a sortable table listing *consumption*, *intensity*, as well as aggregate *savings* values for the currently selected time period.

Line graphs and bar charts such as those in Figure 5.2 (top) and Figure 5.2 (bottom) are indexed on the other pages, similar to how a Microsoft Excel workbook has sheets, which can in turn contain multiple charts. Unlike Excel, Energy Manager's charts provide some interactivity: an energy analyst can zoom, pan, and reveal values upon mouseover. However, none of the charts are directly linked to one another or to the dashboard, so the energy analyst must `navigate` between them manually.

Task support: Energy Manager only partially supports the Overview (**T1**) task: with the portfolio dashboard (Figure 5.1), the energy analyst can observe the aggregate *consumption* for a portfolio over time, or alternatively she can observe single aggregate values for individual buildings in the sortable table, but she will not be able to directly `compare` how individual buildings vary over time. As a result, the energy analysts that we interviewed essentially ignored this dashboard, and none of them had found the sortable table to be useful.

The Drill Down (**T2**) task is supported but the current approach used by Energy Manager does not scale. The examples in Figure 5.2 (top) and Figure 5.2 (bottom) respectively display *energy demand* and *consumption* performance for three restaurants. Since superimposed line graphs and grouped bar charts are limited by the number of discriminable colours, these charts are inappropriate for the energy analyst who needs to consider more than a handful of buildings.

The Roll Up (**T3**) task is not explicitly supported. Though it is possible

Figure 5.1: The previous version of Energy Manager, our collaborators' energy analysis tool: a dashboard for a portfolio of buildings.

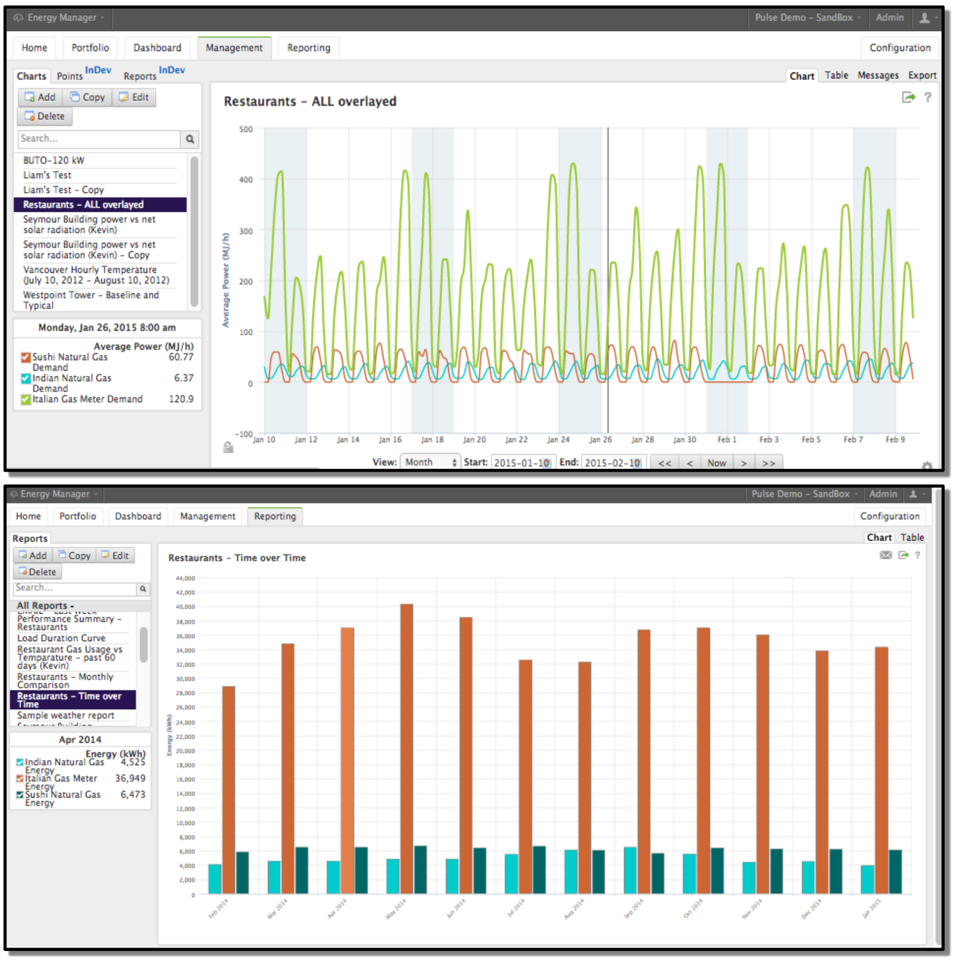


Figure 5.2: The previous version of Energy Manager, our collaborators' energy analysis tool (continued). A superimposed line graph of *energy demand* (top) and a grouped bar chart of *energy consumption* (bottom) for a group of three restaurant buildings within this portfolio.

to estimate the proportion of a building's energy performance relative to its group with bars and lines, this process is error-prone.

Task sequences not supported: Because the line graphs, bar charts, and tables in Energy Manager are not coordinated or linked in any way, it is difficult and tedious to alternate from Overview (**T1**) to Drill Down (**T2**) and Roll Up (**T3**) tasks. As in Excel, the energy analyst will have to `locate` an existing chart or specify a new chart using a wizard dialog; if the bar or line graph for a set of buildings does not already exist when the energy analyst needs it, she has to create it. By the time she has created it, she may have forgotten her objective[5].

Limited filtering and aggregation: The bar charts and line graphs in Energy Manager allow an energy analyst to hide or show individual buildings. However, the energy analyst cannot `filter` buildings with shared attributes, such as `filtering` a portfolio of buildings to only show restaurants. Similarly, it is impossible to aggregate buildings together when they share attributes: for instance, the energy analyst cannot `compare` the aggregate energy performance of restaurants in one city to those in another city.

No faceting: Aside from the seldom-used portfolio dashboard shown in Figure 5.1, there is no faceting or juxtaposition of charts in Energy Manager: the energy analysts' workarounds included opening multiple browser windows, adjusting the line or bar charts to display the same scale and ranges, and tiling these windows manually across their monitor. Similarly, one energy analyst that we spoke to printed and taped charts together to accomplish the same result.

Summary: Due to these limitations, energy analysts can only observe narrow slices of their portfolio data, or they are presented with aggregate data that is too coarse to be useful. In addition, they did not trust Energy Manager's derived predicted values based on statistical models, and would have preferred to `compare` observed energy performance to historical values. In

[5]Task sequences such as these are likely to be poorly supported in most interactive system that involves wizard dialogs.

other words, the derived and aggregated values currently shown may hide information such as extreme values and unusual distributions. As a result of this loss of detail, energy analysts routinely export tabular data from Energy Manager and `import` it into Excel, with which they would conduct a time-consuming custom analysis. Finally, many of the energy analysts to whom we spoke remarked that energy analysis software tools[6] built by our collaborators' competitors have limitations similar to those of Energy Manager. Altogether, these problems may help to explain why the number of energy analysts who actively use Energy Manager is substantially less than the number of client accounts.

5.5 Related Work

We now review relevant previous work, beginning with work in the energy domain. We also discuss the visualization of time-oriented data in other domains, as well as evaluation studies that assess the effectiveness of visualization design choices for similar data and task abstractions.

Visualization in the energy domain: Technology that allows for the continuous measurement of a building's energy demand is becoming increasingly available, and several techniques to monitor and present this data have recently been proposed, especially for residential buildings [87, 92, 114, 259]. Erickson et al. [92] developed a web-based residential energy dashboard for homeowners, allowing them to `compare` against their neighbours with familiar bar charts and line graphs. However, such a dashboard would be unsuitable for the work of an energy analyst who oversees a portfolio of many buildings.

Though bar chart and line graph depictions of energy data are most common, other visual encodings have also been employed, from abstract and artistic ambient visual encodings [259] to a compelling calendar-based visual encoding [335], in which calendar dates with similar energy behaviour are visually associated using a common colour. We also explore visual encodings

[6] e.g., Energent EMIS: `http://energent.com/emis/`, Northwrite Energy Expert: `http://northwrite.com/energyexpert.asp`, Schneider Electric PowerLogic Ion EEM: `http://goo.gl/Nh7mJv`

beyond bar charts and line graphs, and in Section 5.7.3, we consider how to `encode` data from multiple buildings using calendars. Another approach to `summarizing` the energy behaviour of multiple buildings is to use map-based visual encoding, though we discuss the limitations of this approach in Section 5.7.4.

More closely related to our work is that of Goodwin et al. [114], who visualized modelled residential energy use across thousands of households at the scale of individual household appliances, resulting in four prototype data sketches. The domain activities they address overlap partially with those performed by the energy analysts we spoke to, such as the need to find anomalous energy performance across many buildings; another activity they address, in which energy modelers perform energy load-shifting simulations to estimate potential savings, is not an activity performed by our energy analysts. Their designs incorporated several visual encodings not typically seen in the energy management domain: horizon graphs [132], boxplots [352], and matrix-based encodings. However, the focus and main contribution of their paper was on creative methods for visualization requirements analysis, rather than on a thorough analysis of their visual encoding and interaction design choices. In our work, we reexamine some of these visual encodings, among others, and evaluate their effectiveness in the context of our data and task abstractions.

Visualizing multiple time series: Many techniques for visualizing time-oriented data have been proposed, and a survey of these techniques by Aigner et al. [2] has provided us with a structured way to think about this design space. In their terminology, our designs incorporate *linear* and *cyclic* encodings of time, depicting *abstract multivariate interval* data.

Another axis on which we can analyze existing techniques is the number of time series being considered. At the low end of this continuum, superimposed line graphs or grouped bar charts are appropriate for a small number of time series. In the middle of this continuum, faceting techniques such as faceted line graphs, horizon graphs [132] and matrix-based encodings [126, 290] are appropriate. At the high end of this continuum, dense multi-form faceting techniques and those that aggregate time series together are appropriate when dealing with thousands of time series, such as in LiveRAC [206] or Line Graph Explorer [182][7]. Since we are addressing portfolios of dozens to hundreds of buildings, we position our designs toward the middle of this continuum, and we evaluate faceting and matrix-based approaches in the following section.

Evaluation of visualization techniques for time-oriented data: We also situate our work with regards to experimental studies [4, 66, 105, 158] that have examined the viability of alternative visual encodings for abstract tasks similar to those that we classified: `identifying` and `comparing` averages, trends, extreme values, and outliers. Some studies address the viability of encodings for a single time series [4, 66], while others consider multiple concurrent time series; one study considered up to sixteen time series [158], while another considered forty-eight [105]. We too assessed the viability of different visual encodings for multiple concurrent time series; however, our approach involves a qualitative evaluation (see Section 5.2), as opposed to a controlled experiment.

Moreover, while these studies considered continuous time series data, we must consider alternative scalable encodings, since in our case domain conventions dictate that continuous line graph encodings would be misleading for the display of derived and aggregated time series values, as discussed in Section 5.3.1.

5.6 Prototyping Environments

Shiny sandbox: We developed an interactive browser-based visualization design sandbox[8] to produce *data sketches* [195], shown in Figure 5.3. The sandbox allowed us to rapidly prototype different visual encoding designs and conduct chauffeured demos with energy analysts, a process that we described in Section 5.2. All of the designs discussed in Section 5.7 were

[7]LiveRAC and LGE were both developed by members of our group.
[8]`http://mattbrehmer.shinyapps.io/PortfolioSandbox`

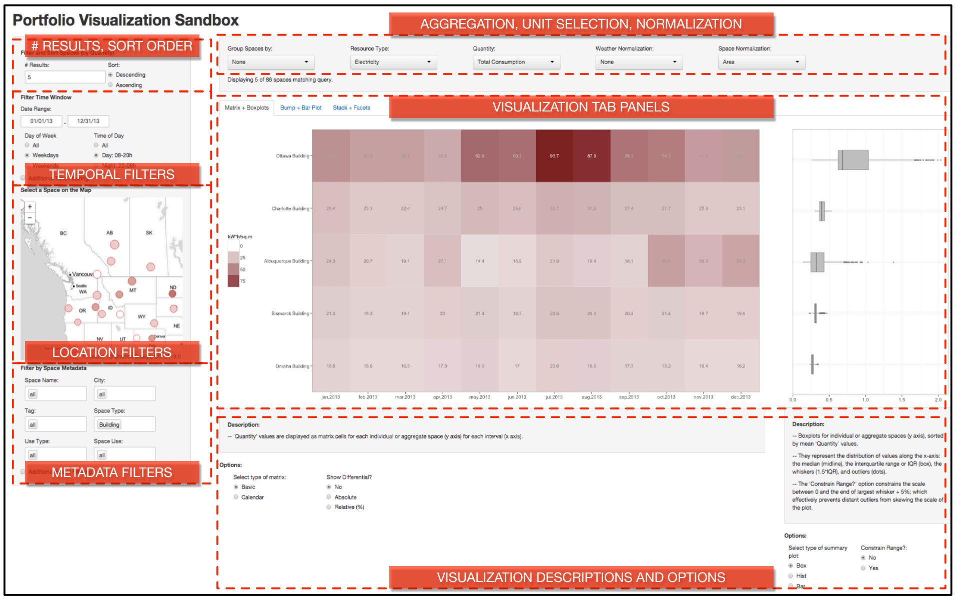


Figure 5.3: A sandbox design environment for visualizing energy data from a portfolio of buildings. A matrix of aggregate *energy intensity* values with auxiliary boxplots is shown for 5 (of 86) buildings, those with the highest *intensity*.

produced within this environment, which was developed[9] using the Shiny web application framework[10].

This sandbox has interactive controls for sorting (`arranging`), `filtering`, and `aggregating` buildings, controls for `selecting` units of interest such as *demand* or *consumption*, as well as controls for area and weather normalization; recall that there were no such controls in the Energy Manager interface. Whenever these controls are adjusted, we sort the filtered set of buildings according to the currently selected energy unit of interest and time span. The sandbox operator can `select` the number of results to show and sort (`arrange`) them. For instance, Figure 5.3 shows

[9]`http://github.com/mattbrehmer/PortfolioSandbox`

[10]Shiny is a framework for integrating statistical analysis and visualization techniques within an interactive web application; Shiny web applications are implemented using the R statistical programming language. More information is available here: `http://shiny.rstudio.com/`

a view described in Section 5.7.3, and in it we show 5 buildings from a geographically anonymized portfolio of 86 buildings, those with the highest *energy intensity* in 2013.

D3 interactive prototypes: Since our Shiny-based sandbox implementation did not allow us to directly experiment with interactions involving coordinated `selection` across juxtaposed views, we developed several interactive prototypes using D3 [30] that specifically address this coordination; these prototypes are discussed in Section 5.8 and one of them[11] is shown in Figure 5.7.

5.7 Visual Encoding Matches and Mismatches

Meyer et al.'s nested blocks and guidelines model [211], which extends Munzner's nested model [217, 219], describes a need for guidelines that relate the domain, abstraction, idiom (visual encoding and interaction design choices), and algorithm levels of visualization design. In Section 5.3, we described the relationship between domain activities and the data and task abstractions. In this section, we consider the space of visual encoding design choices and present guidelines for matching design choices to abstractions. Since the space of possible visual encoding design choices for time-oriented data is large [2], we undertook a typical design study approach [284]: considering several design choices, implementing a subset of them, and selecting only the few good matches.

We identified five *matches* {✔} between visual encoding design choices and the combination of data and task abstractions, based on evidence resulting from our process described in Section 5.2. We also identified four *mismatches* {✖} and two *potential matches* {?}. These matches and mismatches, indicated in Table 5.2, serve as guidelines that are transferable beyond the energy management domain, especially when we consider the similarity between our abstract tasks to those addressed in domain-agnostic evaluation studies [4, 158]. Furthermore, these matches and mismatches fill a gap with regards to identifying suitable visual encoding design choices for

[11] `http://bl.ocks.org/mattbrehmer/287e44c9a12151967874`

Task	Design Choice	Match?
T1: Overview	Faceted bar chart	✖
	Bump plot	✖
	Bar + bump plot	?
	(Calendar) matrix	?
	Map	✖
	Juxtaposed matrix and boxplots	✔
T2: Drill Down	Faceted bar chart	✔
	Faceted boxplot	✖
	Faceted line graph	✔
T3: Roll Up	Stacked area graph	✔
	Stacked bar chart	✔

Table 5.2: A summary of the *matches* and *mismatches* between abstract tasks and visual encoding design choices.

multiple time series in which values are not continuous, but derived and aggregated values that should *not* be `encoded` as line graphs.

5.7.1 Faceted Views for Overview and Drill Down

We initially thought that faceted "small multiple" views would be a good match for *both* the Overview (**T1**) and Drill Down (**T2**) tasks, in that they provide a scalable alternative to grouped bar charts and superimposed line graphs.

Faceted bar charts: *a mismatch* {✖} *for the* Overview (**T1**) *task, yet a match* {✔} *for the* Drill Down (**T2**) *task.* Faceted bar charts were among the first designs that we considered, especially after one energy analyst provided us with his own mockup of such a design. However, if an energy analyst has dozens or hundreds of buildings in their portfolio, faceting is unlikely to scale [158]. We determined that it was a poor match for the Overview (**T1**) task, though a match for the Drill Down (**T2**) task, provided that the energy analyst has already `filtered` down to a smaller group of buildings, such as `filtering` a university portfolio to show only the *"laboratory"* buildings. In addition, one of the energy analysts lamented that bar charts only show coarse aggregate values, typically an average or a sum, and as a result of this loss of detail, they do not show other aggregate values of interest, such

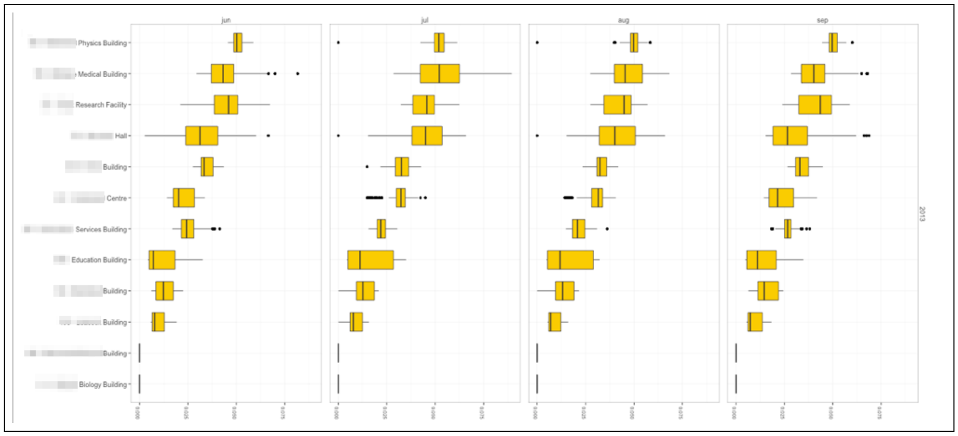


Figure 5.4: Faceted boxplots that `encode` aggregate area-normalized energy *demand* distributions for twelve buildings across four months, sorted in descending order according to the average *demand* value for this four month period. Faceted boxplots are a mismatch {✖} for the Drill Down task (T2). Building names are blurred to sanitize real client portfolio data.

as ranges or extreme values within each corresponding time interval.

Faceted boxplots: *a mismatch* {✖} *for the* Drill Down (**T2**) *task.* We expected that faceted boxplots would allow energy analysts to `compare` ranges, distributions, and extreme values for multiple buildings at different points in time, such as in Figure 5.4. However, despite the long history of boxplots [352] and support from influential visualization practitioners [97], we found that most energy analysts are not familiar with boxplots, except for a minority who had taken a post-secondary statistics course. Furthermore, comparisons in faceted boxplots are more difficult than in faceted bar charts, where positions are aligned to each facet's baseline; with faceted boxplots, the observer must `compare` multiple positions and widths across separate facets. Our design was therefore a daunting introduction to boxplots for those unfamiliar with them and a poor match for the Drill Down task.

Faceted line graphs: *a match* {✔} *for the* Drill Down (**T2**) *task.* Faceted line graphs are a good match when observing non-derived continuous quantitative time series values such as *energy demand*; an example is shown in

Figure 5.8 (bottom). They are a scalable alternative to superimposed line graphs [158] and the line graphs encoding is already very familiar to energy analysts. As mentioned above in Section 5.3, line graphs are *not* appropriate for derived and aggregated values such as *energy consumption* or *intensity*.

5.7.2 Rank-Based Overviews

As faceting seemed unlikely to be effective for the Overview (**T1**) task, we considered non-faceted designs and the encoding of aggregate values. Recall how the sortable table in Energy Manager's portfolio dashboard (Figure 5.1, bottom) was never used for the Overview task; it contained only coarse aggregate values for each item, providing little detail about temporal variation. We therefore experimented with encodings for displaying rank as well as rank change over time.

Bump plots: *a mismatch* {✖} *for the* Overview (**T1**) *task*. Bump plots `encode` rank and rank change; they incorporate a familiar line encoding across equally-spaced temporal intervals [326]. However, as with superimposed line graphs, it becomes difficult to distinguish individual items using colour. One possible solution is to highlight items that vary in rank, rather than requiring the observer to `locate` these items. Another problem is that bump plots only show relative rank and rank change, whereas the absolute values that produce these ranks are not shown. Due to this loss of detail, the bump plot is also a poor match for the Overview task.

Bump + bar plots: *a potential match* {?} *for the* Overview (**T1**) *task*. We next considered an encoding that incorporates relative rank, rank change, and absolute value, by adding bars to each series in the bump plot, as shown in Figure 5.5. This approach is similar to two recently proposed techniques that `encode` both relative rank and absolute value [119, 142]. With this design, we still face the scalability problem associated with colour discriminability. A combination of interaction and highlighting rank variation may facilitate this discriminability; in Figure 5.5, rank variation is encoded using the alpha channel, so the pink series that varies considerably over time is most salient.

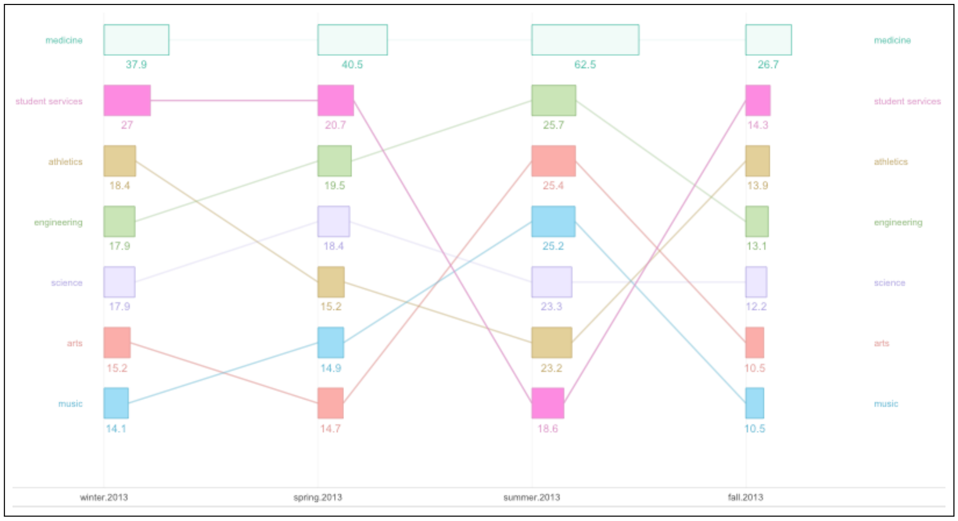


Figure 5.5: A *bar + bump plot* of *energy intensity*, encoding rank change for the top seven building groups (buildings aggregated by tag) across four seasons. The alpha channel `encodes` rank variation to highlight inconsistent buildings; in this instance, so the pink series (beginning in the top left) varies considerably over time and is most salient. The *bar + bump plot* is a potential match {?} for the Overview task (T1).

Energy analysts responded positively to this visual encoding, as it is comprised of familiar bar and line encodings. However, despite this positive response, we discovered that *ranks* as derived values are actually infrequently considered during energy analysis, and that they tend to be more appropriate for annual planning and presentation activities, such as determining how to prioritize energy conservation projects, and less so for recurring analysis and monitoring activities. Thus, the hunt for a match for the Overview (**T1**) task continued.

5.7.3 Matrix-Based Overviews

Time series matrix: *a potential match* {?} *for the* Overview (**T1**) *task*. Matrix encodings are scalable and space-efficient [114, 126], as can be seen in the center of Figure 5.3. Matrix encodings allow us to display observed as well as differential values, allowing an energy analyst to review *energy savings* relative to predicted or historical values; a matrix displaying differential energy data is shown in Figure 5.6. Most of the energy analysts that we interviewed were unfamiliar with this form of encoding, except one who had made use of a similar visual encoding in Excel. As a result, it took more effort to convince our collaborators of the value of these matrix-based encodings for the Overview task.

We also learned that energy analysts found matrices with diverging colour scales easier to interpret than than those with unidirectional colour scales. Finally, we found that while red is appropriate for use in diverging colour scales, as it has a negative connotation, it is inappropriate for unidirectional colour scales in this context. As a result of this mixed response to matrix-based encodings, we realized that more work needed to be done.

Calendar matrix: *a potential match* {**?**} *for the* Overview (**T1**) *task.* We altered our matrix encoding by partitioning the cells corresponding to months into calendars (Figure 5.6), a design decision inspired by previous work [184, 335]. Energy analysts responded positively to this encoding, which helped to resolve the unfamiliarity of the more generic matrix encoding. However, months and days are not the only time granularities of interest, so this encoding may not be appropriate for all time ranges.

5.7.4 Map-Based Overviews

Maps: *a mismatch* {✖} *for the* Overview (**T1**) *task.* We explored the use of maps based on their popularity in the energy domain[12]. We conjectured that maps may be appropriate for buildings in a shared neighbourhood, such as a university campus, even though that encoding may be less appropriate for building portfolios spanning large geographic areas. However, after interviews with energy analysts, we realized that maps are better suited for `presenting` coarse aggregate summary values of energy performance to a casual observer, and they are less appropriate for recurring analysis work;

[12]e.g., HEAT (Heat Energy Assessment Technologies) Map: `http://saveheat.co/heat-scores.php`, the McGill Energy Project's Energy Map: `http://mcgillenergyproject.com/map.php`

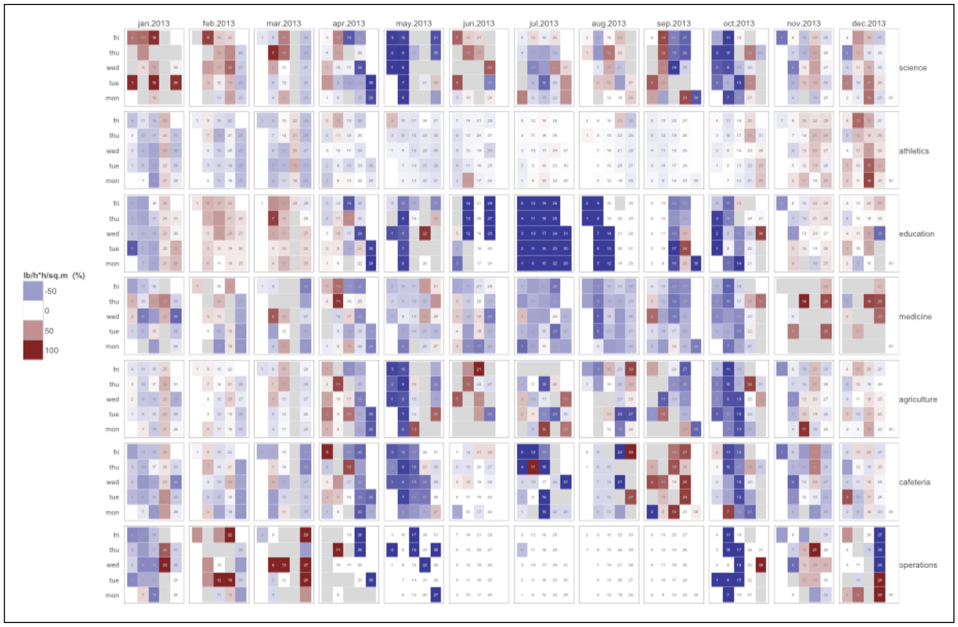


Figure 5.6: A time series calendar matrix of *energy intensity savings* for seven building groups (buildings aggregated by shared categorical tag), relative to historical values (blue = *savings*, red = higher than historical *intensity*). The time series calendar matrix is a potential match {?} for the Overview task (T1).

to view energy behaviour varying over time, animating or faceting the map would be necessary. Furthermore, an energy analyst overseeing a portfolio of buildings is already likely to be familiar with the locations of buildings in her portfolio, and their relative locations are not informative. While using a map to `encode` energy data was found to be inappropriate for the tasks that we classified, an interactive map may be an effective means to `filter` a portfolio by building location, which we considered in our sandbox environment shown in Figure 5.3.

5.7.5 Stack-Based Roll Up Encodings

Stacked area graphs & stacked bar charts: *matches* {✔} *for the* Roll Up (**T3**) *task.* The obvious visual encodings of stacked area [46] and stacked bar charts were indeed matches; an example of the former is shown in Figure 5.8 (top). Stacked area graphs are appropriate when considering *energy demand* values, while stacked bar charts are appropriate for derived and aggregated values such as *energy consumption* or *intensity*. For both of these encodings, differentiating individual time series can be accomplished by interactive highlighting [346], as using hue to differentiate stack elements will not scale.

5.8 Workflow Design with Multiple Views

Our design discussion up to this point has focused on visual encoding choices for single views; we also want to stress the importance of interaction and workflow design, which involves juxtaposing and linking multiple views.

Juxtaposed matrix and boxplots: *a match* {✔} *for the* Overview (**T1**) *task*. One reason to juxtapose views is to support a single task with complementary data. None of the encodings discussed thus far are a clear match for the Overview task, although the time series matrix designs described in Section 5.7.3 showed promise. A problem with the matrix encodings is that they only display coarse aggregate values, such as averages or sums for each matrix cell. Recall from Section 5.7.1 that an energy analyst made the same remark about faceted bar charts. Partitioning a matrix into calendars is one way to show a finer resolution in the same amount of space, however this encoding will not always be appropriate: for instance, an energy analyst may be interested in a time span shorter than a month, or longer than several years. The alternative that we developed involves complementing and reinforcing the aggregate values in the original matrix design by juxtaposing single boxplots that `encode` ranges and distributions for each time series, as shown in Figure 5.3. Though boxplots remain unfamiliar to energy analysts, these auxiliary boxplots are easier to interpret than faceted boxplots (c.f. Figure 5.4), as they require no comparison of length or width across separate facets. We reflect further upon the balance between familiarity and the use of auxiliary charts to combat information loss in Section 5.10.1.

We then sought a better way to coordinate and link the matrix and

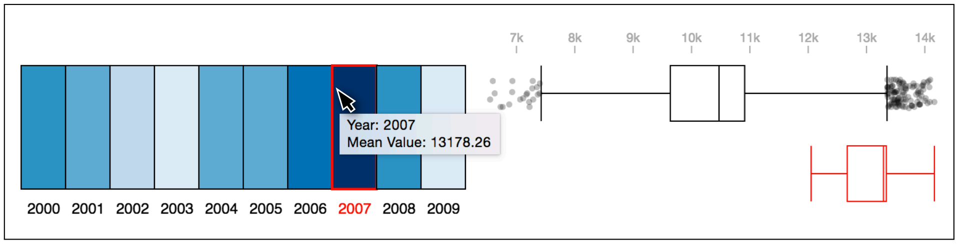


Figure 5.7: An interactive auxiliary boxplot prototype: boxplots corresponding to the brushed time period are shown alongside the boxplot for the entire time series.

its juxtaposed auxiliary boxplots. We created several prototypes, and one is shown in Figure 5.7. The interactive linked highlighting in this prototype served both to promote engagement with these juxtaposed views and to facilitate the learning of these visual encodings, which were previously unfamiliar to energy analysts[13].

Juxtaposed stack and facets: Another reason to juxtapose views is to support fast alternation between tasks. Recall that the Drill Down (**T2**) and Roll Up (**T3**) tasks are often performed in alternation, and we were concerned about the loss of context when switching between stacked bar charts or stacked area graphs and faceted bar charts or line graphs. To prevent this loss of context, we juxtaposed the stacked chart with the faceted charts, and provide linked highlighting between elements in the stack and those in the facets, as shown in Figure 5.8; as a result, both the Drill Down and Roll Up tasks are supported in a single display. Currently, four facets are shown in a row, with additional facets wrapping to subsequent rows, sorted in the same order as the elements in the stack.

Sequenced view navigation: Recall that the Drill Down (**T2**) and Roll Up (**T3**) tasks involve fewer buildings and finer temporal resolutions than the Overview (**T1**) task, which has a broader scope; thus, it would be inappropriate to juxtapose the Overview with the Drill Down and Roll Up views in a single display. Instead, we considered how an energy analyst

[13]This interactive prototype is available here: `http://bl.ocks.org/mattbrehmer/287e44c9a12151967874`

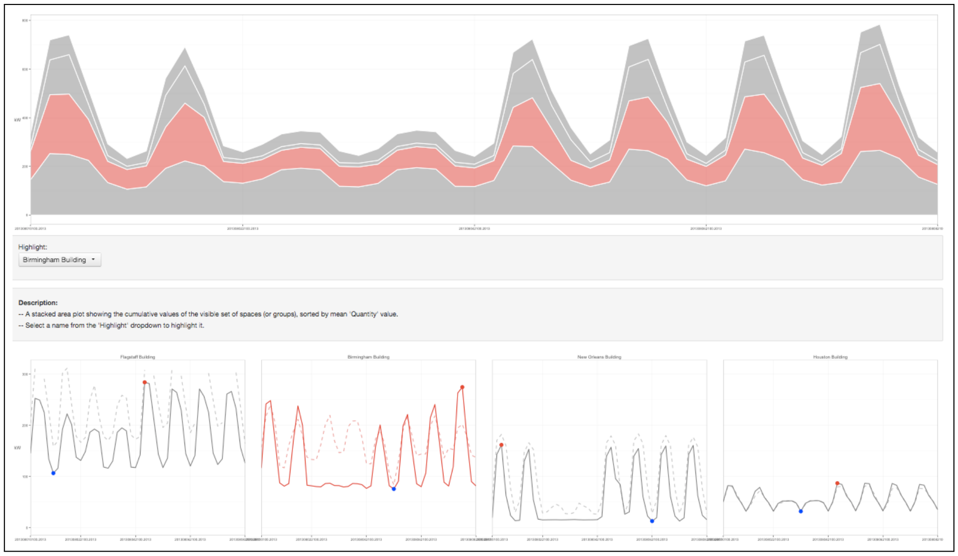


Figure 5.8: A stacked area graph of *energy demand* data for four library buildings, juxtaposed alongside faceted line graphs of the same data. The same building is highlighted in red in both the stack and the facets.

would `navigate` between these views shown on separate displays. Beginning with the matrix and auxiliary boxplots, the energy analyst can perform the Overview task, `select` units of interest, and `filter` or `aggregate` buildings in the portfolio. If the currently selected unit of interest is *consumption* or *intensity*, `selecting` a column of the matrix directs the energy analyst to juxtaposed faceted and stacked bar charts that include every building from the matrix and spanning the time period corresponding to the selected column. For *demand* data, the energy analyst is directed to faceted and stacked line graphs. At this point, the energy analyst can perform the Drill Down and Roll Up tasks in alternation. We demonstrate this multiple-view workflow in a supplemental video[14].

Finally, we also envisioned drilling down further to individual buildings. If the energy analyst `selects` a cell or row of the matrix, she will `navigate`

[14] This video is available here: `http://cs.ubc.ca/labs/imager/tr/2015/MatchesMismatchesMethods/`

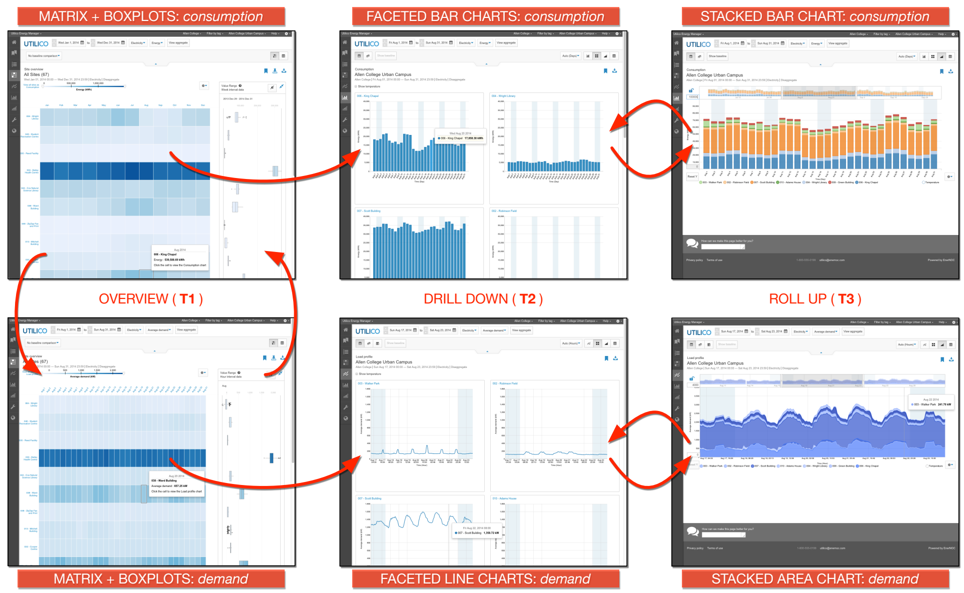


Figure 5.9: The redesigned Energy Manager that incorporates many aspects of our prototype designs. On the left, the *Site Overview* (a time series matrix) is juxtaposed with coordinated *Value Range* (boxplot) views. An energy analyst can easily switch between units such as *energy consumption* or *energy demand* and `filter` or `aggregate` the set of buildings to those that share a common categorical tag; by `selecting` a column of the matrix, she can drill down to faceted or stacked visual encodings of *consumption* (top middle, top right) or *demand* (bottom middle, bottom right).

to a single bar or line graphs for the corresponding building and time period.

5.9 Results

We are pleased to report that our collaborators have adopted a number of our designs into a new version of Energy Manager, shown in Figure 5.9, which will soon be deployed to their large client base.

As in our sandbox environment, options for `filtering` and `aggregating` buildings according to shared categorical tags are now prominently and persistently shown in the menu at the top of the interface.

The interface also provides the ability to `select` units of interest and `compare` observed energy values against trusted historical values, as an alternative to comparing observed values to less trusted predicted values, overcoming a limitation of the original Energy Manager.

Coordinated matrix and boxplots: a variant of our design described in Section 5.8 has been incorporated into the redesigned Energy Manager. The number of buildings shown depends on the window size, and more buildings appear as the energy analyst scrolls. Our collaborators did consider the alternative calendar-based encoding, but ruled it out based on a requirement that arose late in the design process: that the redesigned Energy Manager be accessible on a tablet device. Partitioning the matrix cells into calendars may result in calendar dates too small to be `selectable` without zooming, which may incur a high implementation cost. Meanwhile, the boxplots update when the energy analyst brushes the matrix by hovering over a cell in the corresponding matrix row, similar to the behaviour of the prototype we described in Section 5.8. Unlike our earlier prototype, a single boxplot is shown instead of showing one boxplot for the entire range and another boxplot for the brushed time period; when no time period in the matrix is brushed, the boxplot for the entire time series is shown.

Interactive workflows realized: Another critical improvement over the original Energy Manager is the ability for an energy analyst to drill down from a row, column, or cell of the matrix to stacked, faceted, superimposed, or individual line graphs and bar charts, as shown in Figure 5.9. The selected energy unit is retained across these transitions, so faceted line graphs and stacked area graphs are used for *demand*, while faceted bar charts and stacked bar charts are used for *consumption*. In faceted views, individual facets can be manually resorted via drag and drop. Stacked and faceted views are currently shown separately; our collaborators are considering juxtaposing stacked and faceted views, such as in the design described in Section 5.8, which would allow for an uninterrupted alternation between the Drill Down (**T2**) and Roll Up (**T3**) tasks in the same display.

5.10 Discussion

We now step back from specific aspects of visualization design for time-oriented data to discuss higher-level guidelines, to reflect upon on our methodology, and to indicate possible future work.

5.10.1 Guidelines: Familiarity and Trust

In addition to the specific guidelines regarding matches and mismatches between design choices and abstractions described in Section 5.7, we also propose guidelines relating to the themes of *familiarity* and *trust*.

Familiarity: As with professionals in many other domains, energy analysts are accustomed to working predominantly with familiar visual encodings, namely bars and lines. When we introduced them to unfamiliar visual encodings, we learned several things:

Persevere despite unfamiliarity: Though counter-intuitive, we learned that the juxtaposition of a matrix and a boxplot, two unfamiliar encodings, together with coordinated interaction and highlighting, received more positive feedback than either of these encodings in isolation. The issue of unfamiliarity with the time series matrix was also partially resolved when we partitioned the cells into calendars. We similarly persevered with the unfamiliar bump plot: by superimposing a layer of familiar bars on top of the bump plot, energy analysts were able to more easily interpret this rank-based visual encoding.

Beware assuming familiarity: Introducing visual encoding design choices using names that are well-known in the visualization literature can be misleading when they allude to familiar concepts in a way that is unfamiliar to the target audience, as we found with energy analysts and the term "*heat map*" [98, 354]. We initially referred to the time series matrices as "heat maps", but this visualization term led to considerable confusion because of conflicting domain conventions with the energy-related meaning of *heat* and expectations raised by the word *map*: this encoding does not show energy solely used for heating, nor does it show the geographic location of buildings[15]. In the redesigned Energy Manager, the time series matrix is referred to as a *Site Overview*.

We also had a difficult experience gathering feedback about boxplots because the term itself was unfamiliar. In the redesigned Energy Manager, the boxplots are referred to as the *Value Range* chart, a term that appears to be understood. In hindsight, we could have explicitly solicited names for these views from energy analysts early on in the process based on their own descriptions [209].

Trust: When a visualization technique or tool is used as part of the hypothesis generation and verification process, trust is imperative, especially for derived and aggregated values. Previous work has investigated the trustworthiness of visual encodings for text-based data [61], and we now discuss the topic of trust motivated by our design of visual encodings for time-oriented data.

Auxiliary charts to combat information loss: When the number of concurrent time series grows large, it is difficult and overwhelming to visualize individual values from each time series; instead, a common approach is to visualize derived aggregate values [206]. This loss of detail is apparent in the original Energy Manager's portfolio dashboard (Figure 5.1) as well as in the cells of our time series matrix (Figure 5.3). Whenever there is a loss of detail, there is a loss of trust: one of the energy analysts remarked that these derived aggregate values hide information such as extreme values. In juxtaposing the time series matrix with auxiliary boxplots that update whenever a matrix cell containing an aggregate value is brushed, we are not only restoring lost information: we are also restoring trust.

Promote agency over derived values: In our sandbox environment and in the redesigned Energy Manager, we provided explicit and obvious interactive controls for `filtering` and `aggregation`, as well as controls for unit `selection` and normalization, controls that were missing in the original Energy Manager. With these controls, we provide energy analysts agency over the creation of derived values and these values become more trusted.

[15]This confusion is not unique to the energy domain [98, 354].

Similarly, the redesigned Energy Manager includes the option to `compare` observed energy performance to selected historical values, as an alternative to `comparing` against predicted values generated by a "black box" statistical model; until there is some visual indication as to how the underlying model algorithms generate these values [215], there will be little trust, and it is therefore preferable to provide the option to `compare` against trusted historical values.

5.10.2 Methodological Reflection

Though our overall methodological approach is similar to many other visualization design studies [204, 284], there are some specific aspects of our methodology that are unique to projects executed in company settings [282]: we negotiated access to clients and to their portfolio data at the very beginning of our collaboration, and we encourage researchers considering similar collaborations to do the same. In addition, we also engaged primarily with remote energy analysts, and our methodological decisions described in Section 5.2 reflect this logistical difficulty [34]. We now take the opportunity to reflect on three other aspects of our methodology:

Work domain analysis: *Worth it, and don't be daunted.* Conducting a rigorous and systematic work domain analysis can be time consuming and logistically challenging. However, it is helpful to realize that authorities on work domain analysis [337] established their methodologies in the design of high-risk, high-cost systems, such as nuclear power plant control rooms. Work domain analysis and requirements analysis methodologies for many visualization design projects can be more flexible [208] and creative [114] relative to those used for high-risk, high-cost systems. A thorough work domain analysis need not take a year to complete, and subsequent phases of abstraction and iterative design can be carried out while continuing to develop an understanding of the work domain.

Workflow prototyping: *A successful tool is more than a collection of views.* In addition to using our sandbox environment to identify appropriate visual encodings for individual tasks, we also used it as a tool to generate possible *workflows* that support a sequence of tasks. Some visualization research projects stop before this step, but we argue for its importance. We thought that confronting energy analysts with a combinatorial explosion of possibilities from a large set of visual encodings and view parametrization options would fall short of a full solution to the problem at hand.

Grounding design decisions: *Document everything, strive to be consistent and systematic.* One collaborator remarked that our approach often confirmed some earlier suspicions rather than introduced major surprises, where the novelty lay in a clear path to design decision-making that was missing before: *"we performed an analysis of* [*Energy Manager's*] *flaws in a systematic way, put a name on them, and then tested with users"*. The exhaustive collecting and analyzing of qualitative data before and during design allowed us to justify the design decisions described in Section 5.7 and Section 5.8. Presenting our consolidated findings and design justifications in concise and consistent annotated slide decks was highly appreciated by our collaborators. Given this presentation of evidence, our collaborators adopted our designs with confidence, much in the same way that the results of a controlled quantitative experiment can convince stakeholders [282].

5.11 Summary

We conducted a design study in the energy management domain, one in which we collaborated with an energy analysis software company and their clients to develop a tool for analyzing the energy performance of large building portfolios. We described visualization design choices framed in terms of *matches* and *mismatches* between abstractions and visual encoding design choices that are transferable beyond the energy analysis domain, to other domains involve `summarizing` and `comparing` many concurrent time series. The matches include:

- Juxtaposed matrix and boxplots for an Overview task: `lookup` and `summarize` time-oriented quantitative data from all the items spanning a coarse period of time.

- Faceted bar charts and faceted line graphs for a Drill Down task: `locate` and `compare` time-oriented quantitative data for a group of items.

- Stacked area graph and stacked bar charts for a Roll Up task: `explore`, `locate`, and `identify` the proportion of a quantity associated with one item relative to the quantity associated with the item's group over time.

We also contributed seven lessons or guidelines pertaining to the themes of *familiarity* and *trust*, along with methodological guidance for visualization design studies:

- Familiarity: *Persevere despite unfamiliarity.*

- Familiarity: *Beware assuming familiarity.*

- Trust: *Auxiliary charts to combat information loss.*

- Trust: *Promote agency over derived values.*

- Work domain analysis: *Worth it, and don't be daunted.*

- Workflow prototyping: *A successful tool is more than a collection of views.*

- Grounding design decisions: *Document everything, strive to be consistent and systematic.*

As a result of our research and design process, our visualization designs were adopted by our collaborator into their development of a redesigned commercial energy analysis application that will be deployed to client organizations.

5.12 Addendum

Figure 5.10[16] illustrates the three tasks that we identified in Section 5.3.2.

[16]Figure 5.10 did not appear in our design study paper [37].

In Section 5.3.2, we explicitly listed the `actions` and `targets` involved in these three tasks, and in Figure 5.10, `targets` are depicted as the `output` of each task. Figure 5.10 also captures the sequential relationships between these tasks, as well as the *how* perspective, indicating *how* the redesigned Energy Manager (shown in Figure 5.9) supports these three tasks in terms of design choices[17].

The Overview task (**T1**) is supported by two views: the data is `encoded` as a time series matrix and as a series of boxplots, which are juxtaposed (`arranged`) in a single display. The energy analyst can `aggregate` all the buildings to view their combined consumption or demand, `navigate` the list of buildings, `navigate` the temporal range of the data, and finally `select` data points in the time series, a brushing operation that updates the juxtaposed boxplots.

The Drill Down task (**T2**) is supported as follows: the energy analyst begins with the time series matrix and boxplots, whereby she `selects` or `filters` the set of buildings. She is then directed to a new display, in which the data is `encoded` as a either a set of bar charts or line graphs (depending on whether the data in question is consumption, intensity, or demand), which are juxtaposed (`arranged`) in a faceted small multiple layout. She can then `navigate` this set of buildings and `navigate` the temporal range of the data.

Finally, the Roll Up task (**T3**) is supported by a single view in which the data for a group of buildings is `encoded` as stacked bar charts or stacked area graphs (again depending on whether the data in question is consumption, intensity, or demand). With either of these charts, the energy analyst can `navigate` the temporal range of the data and `select` time points for individual buildings to view the specific values.

[17] This classification of *how* follows an extension to the typology by Munzner [219], as illustrated in Figure 6.4, rather than our original set of *how* design choices included in Figure 2.1.

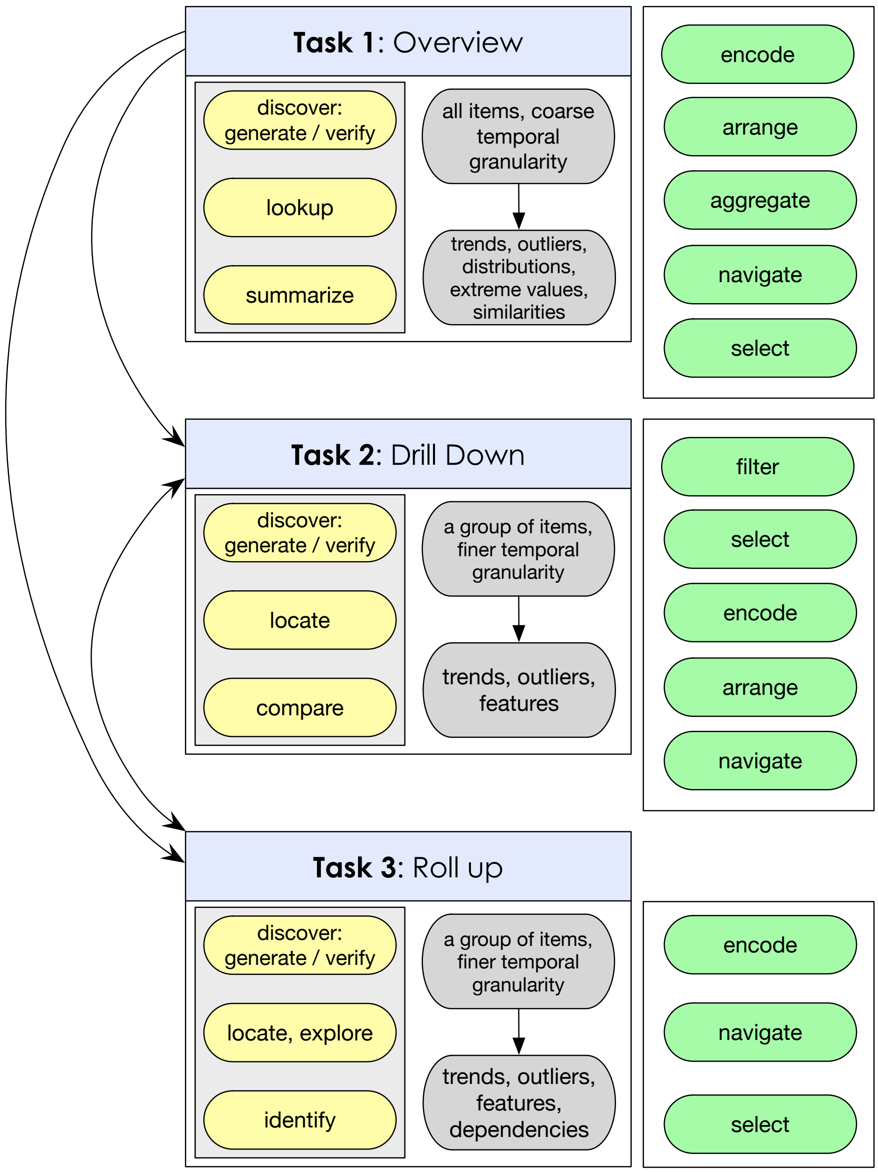


Figure 5.10: The three tasks identified in Section 5.3.2; colours correspond to the *why-what-how* structure of our task typology (see Figure 2.1), where yellow is *why*, green is *how*, and grey is *what* defined in terms of `inputs` and `outputs`, where the `outputs` are framed as `targets`. The classification of *how* corresponds to the redesigned Energy Manager as shown in Figure 5.9.

Chapter 6

Reflection and Conclusion

"Because everything connects in the end, or only seems to, or seems to only because it does." — Don DeLillo in *Underworld* (Scribner, 1997)

Prior to concluding this dissertation, we revisit each of the four projects described in the chapters above, to reflect upon their impact in the research community, to point out their limitations, and to indicate open research questions and ideas for future work.

6.1 Reflecting on the Task Typology

In this section, we discuss Munzner's extension to our task typology, new classifications of tasks appearing since the initial publication of our typology, and the impact of our typology in the visualization research community and beyond.

6.1.1 An Extended Task Typology

Since the initial publication of our typology paper [33], Munzner has expanded upon the *why*, *how*, and *what* aspects of the typology in her recent book [219].

We used Munzner's extension to the typology in our analysis of task sequences involving dimensionally reduced data in Chapter 3. Section 3.5

summarized this extension, in which `annotate`, `record`, and `derive`, previously attributed to families of interaction design choices in the *how* part of our typology, are reclassified as ends rather than means, as subtypes of `produce`.

Munzner [219] further distinguishes between `actions` and `targets` in the *why* part of the typology. The set of *actions*, shown in Figure 6.1, are identical to the extended *why* part of typology described in the previous paragraph and in Section 3.5. The set of `targets`, shown in Figure 6.2, are are meant to be used in conjunction with *actions* as a way to characterize *why* data is visualized and *what* the `output` of the task should be. As mentioned in Section 5.3.2, it can be helpful to think of `actions` as verbs and `targets` as nouns. In the energy management design study, described in Chapter 5, our initial task abstraction followed that of our typology paper [33]; however, when preparing the manuscript that would eventually become our IEEE InfoVis 2015 paper [37], we integrated Munzner's distinction between `actions` and `targets` into our task abstraction in Section 5.3.2.

Munzner [219] has also introduced a classification of *what*, shown in Figure 6.3, a remarkable difference from our "bring your own *what*" mentality discussed in Section 2.3.3, where we suggested that a simple specification `inputs` and `outputs` was sufficient for describing tasks and task sequences. Munzner's classification of *what* refers to abstract datatypes, attribute types, and dataset availability, among other criteria.

Finally, Munzner [219] has proposed a revision to the *how* part of the typology, as indicated by Figure 6.4. Relative to our original formulation of the typology (c.f. Figure 2.1b), Munzner expands upon subtypes of `encode` and further distinguishes between `manipulate` (retaining subtypes `navigate`, `select`, and `change`), `facet` (formerly `arrange`, now with subtypes `juxtapose`, `superimpose`, and `partition`), and `reduce` (encompassing `filter` and `aggregate`, along with the previously absent `embed`).

Commentary on these extensions: Our typology of visualization tasks is an evolving entity, a proposal that the visualization community might use to classify tasks, and Munzner's extensions are part of this evolution, along

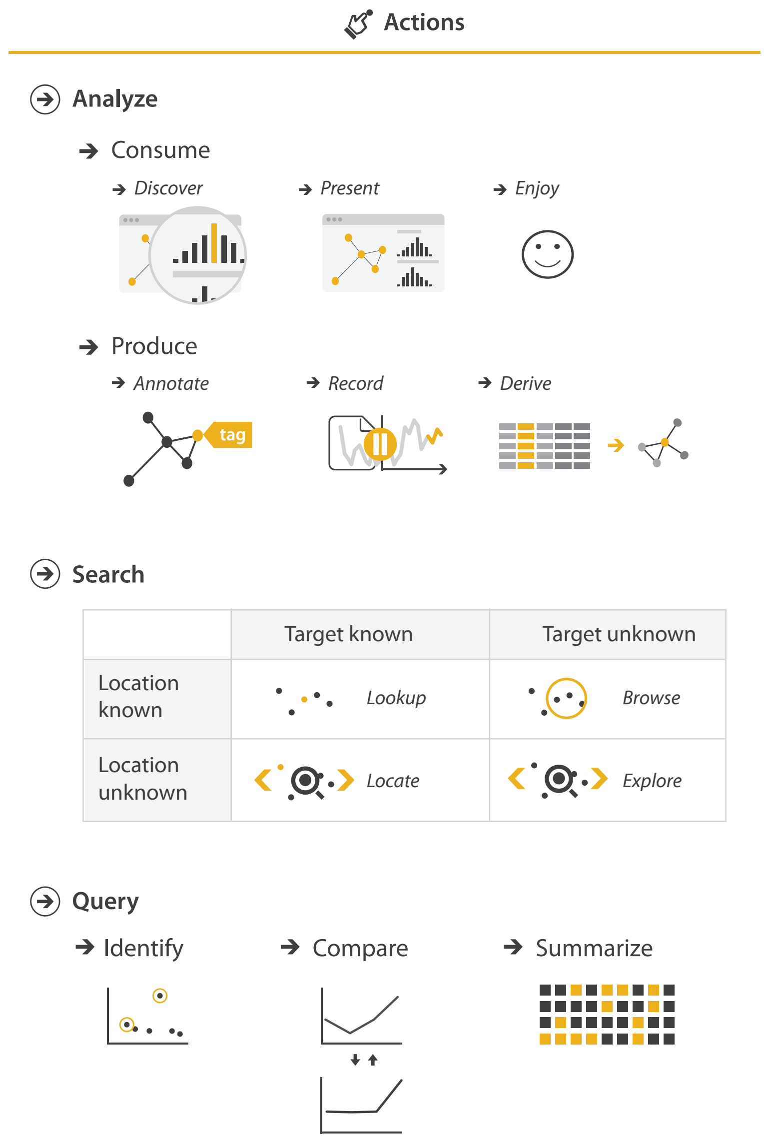


Figure 6.1: The *why* part of our typology, slightly extended as per Section 3.5 and recast as a set of `actions` [219]; c.f. Figure 2.1a. Illustration: 🅭 E. Maguire (2014).

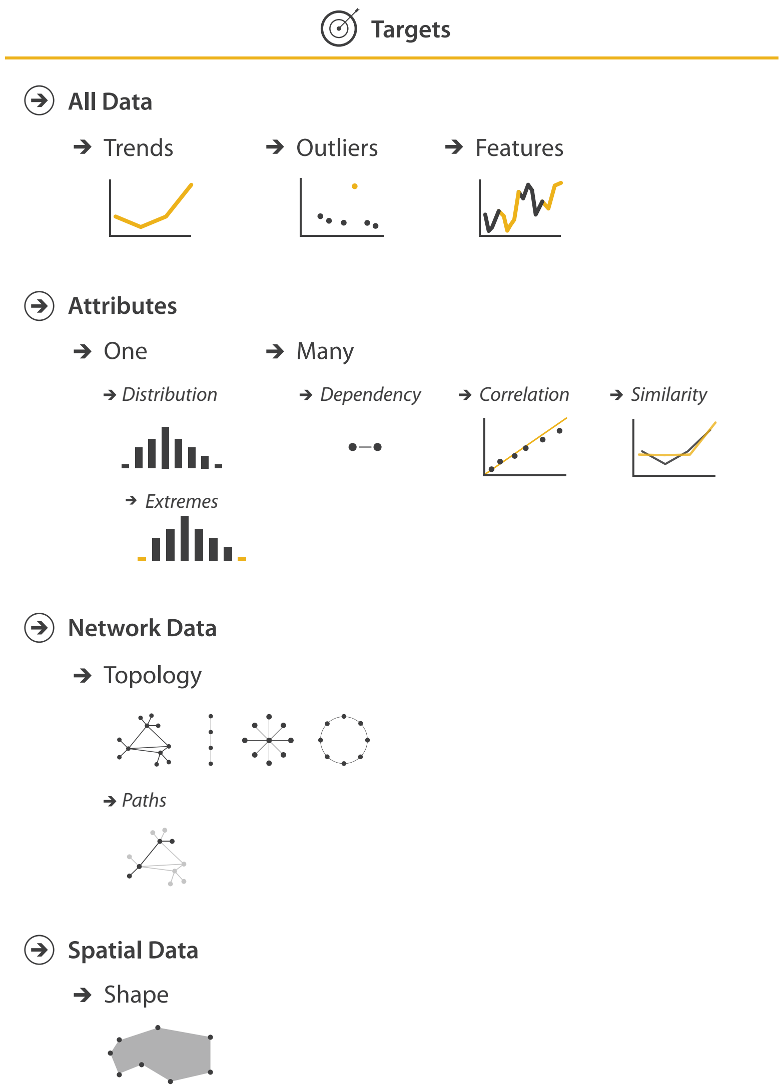


Figure 6.2: A set of `targets`, to be used in conjunction with a specification of `actions` drawn from the set in Figure 6.1 [219]. Illustration: ⓒⓒ E. Maguire (2014).

Figure 6.3: Munzner's classifcation of *what* [219]. Illustration: 🅭 E. Maguire (2014).

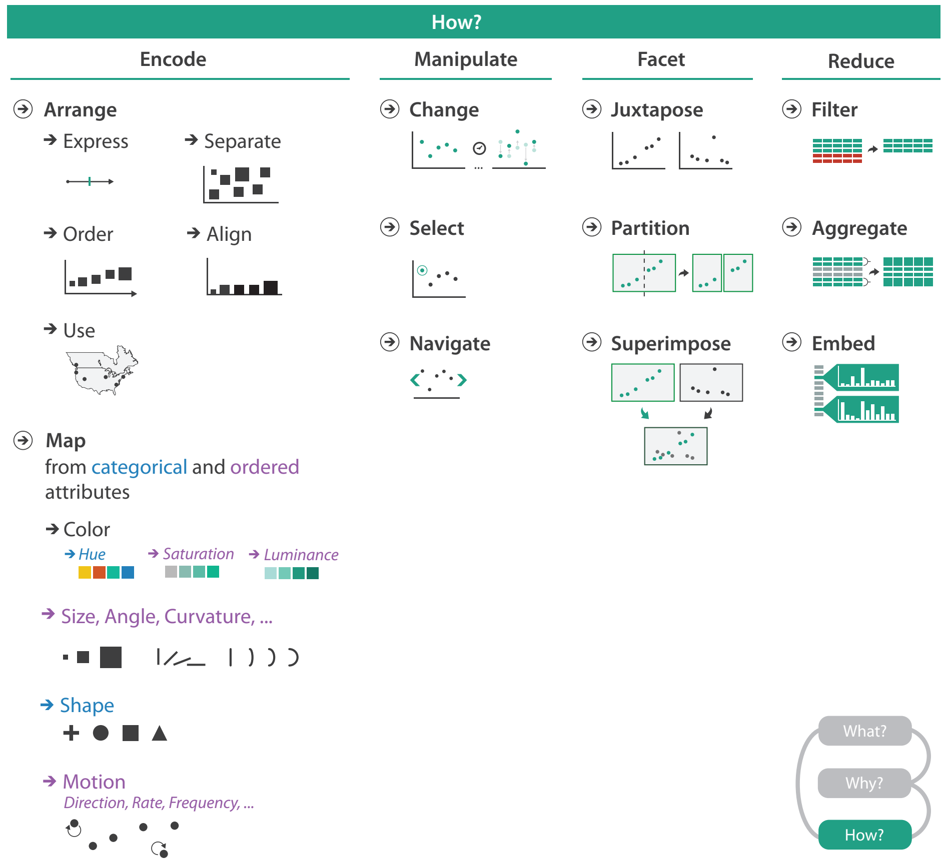


Figure 6.4: Munzner's extension [219] to the *how* part of our typology; c.f. Figure 2.1b. Illustration: 🅭 E. Maguire (2014).

with the efforts of other researchers who have extended our typology to speak about tasks involving specific datatypes or domains, which we discuss below in Section 6.1.3.

Is this variability and evolution a problem? Not necessarily. One the one hand, Munzner's proposed modifications allow us to create more specific descriptions of *how* a visualization technique or tool can support a task, as well as *what* these tasks pertain to. We surely benefited from Munzner's modifications in our interview study (Chapter 3) and in our design study (Chapter 5), in which we acknowledged the `introduce` nodes being reclassified as forms of `produce` and the addition of `targets`, respectively. On the other hand, an evolving classification is more difficult to learn, and may complicate communication between researchers, practitioners, and students, which surely reduces the benefit of having a common lexicon for describing tasks. Furthermore, one could argue that with the increase in specificity that Munzner's modifications bring, there is a reduction in flexibility. For instance, by leaving `input`, `output`, and `encode` as open-ended placeholder terms in our original typology, people were free to consider inputs and outputs specific to their situation, or whichever visual encoding they could imagine.

Munzner's decision to move `introduce` nodes `annotate`, `record`, and `derive` from the *how* part of our typology to the *why* part of our typology as subtypes of `produce` reflects an uncertainty about these terms that appeared throughout the typology's history (see Appendix A): at one point, `derive` was classified under the question of *which* (see Figure A.9); earlier, `annotate` and `record` were classified as *provenance tasks* along with `present` (see Figure A.5). This uncertainty with regards to whether `annotate`, `record`, and `derive` fit best under *why* or *how* seems to imply that they could be seen as either ends or means, subject to context and their position within a sequence of tasks. For now, the advantage of thinking of `annotate`, `record`, and `derive` as forms of `produce` in the *why* part of our typology is that it forces us to consider *how* the annotation, recording, or derivation are supported by a visualization tool. For instance, consider `annotate`: annotation may be supported with `selection` of visualized data elements via direct manipulation, or an automatic annotation of data elements may occur in response to interactive `navigation` or `aggregation`.

Munzner's addition of `targets` and a thorough classification of *what* is also not surprising given the history of our typology. In Appendix A, we document how earlier drafts of what would be become our typology included a classification of *what* beyond `input` and `output`[1]. We realized that we could not present a satisfyingly comprehensive classification of *why*, *how*, and *what* in a single 10-page paper submission to IEEE InfoVis, and

[1] See Figures A.8 – A.12.

thus we opted to focus on *why* and *how*. Nevertheless, I personally prefer the flexibility that comes with thinking about *what* simply in terms of `input` and `output`, as this line of thinking also primes one to think in terms of task sequences and interdependencies, while there is no explicit sequence-based language in Munzner's `targets` or in her classification of *what*. Furthermore, I view Munzner's `targets` and her classification of *what* as suggestions, like any term in our original typology: if a person has more specific terms in mind, she should be encouraged to use them in place of our abstract conceptual terms, especially if it helps her to think. This reflects how others have used our typology, which we describe below in Section 6.1.3: they retain the *why-what-how* structure of our typology and perhaps some of our vocabulary, while introducing their own terms to create new a datatype or domain specific taxonomy. In summary, out typology is part of an ongoing conversation about tasks, and any incarnation of it should not be interpreted as a rigid system.

6.1.2 Comparisons to Roth (2013), Schulz et al. (2013)

When I presented our typology paper [33] at IEEE InfoVis 2013, my presentation was preceded by a classification of interaction primitives for cartographic data by Roth [262] (an extension of his 2012 *GIScience* workshop paper [260], which we included in the literature review that grounded our typology). My presentation was also preceded by "*A Design Space of Visualization Tasks*" by Schulz et al. [280], which independently arrived at some of the same defining questions forming the structure our typology, namely *why* a task pursued (the *goal* of the task), *how* a task is carried out (the *means* of the task), and *what* a task seeks (the data *characteristics*); they also pose the questions of *where* does a task operate in the data (the *target* and *cardinality* of data entities within that target), *when* a task is performed (for specifying the order of tasks), and *who* is executing the task (for specifying the type of user). Schulz et al. [280] further introduce a formal notation for describing tasks, and they realize their design space with an implementation of a tool for recommending visualization techniques in relation to climate impact data.

There are certainly ideas that are common between our typology and the classifications of Roth [262] and Schulz et al. [280], however there are also notable differences: there are aspects of tasks that we account for that these other classifications do not (and vice versa), as well as aspects that we organize in a different way. Moreover, if we consider the external influences on these three classifications in terms of their bibliographies, only seven references are shared by all three papers[2].

At the highest level, the three questions of our typology suggests a correspondence with the dimensions of Schulz and colleagues' design space [280] and with the four parts of Roth's taxonomy [262].

One might expect that our question of *why* a task is undertaken to overlap with Roth and Schulz et al.'s characterization of *goals*, however we also observed a partial overlap with the design space's dimensions of *means* and *cardinality*, as well as with Roth's *objectives* and *operators*. Schulz and colleagues have analogues to `present` and `discover`, and our definition of `discover` similarly refers to the *generating*, *refining*, and *verifying* of hypotheses. Meanwhile, Roth's *procure* bears a similarity to our definition of `consume`; while his *predict* and *prescribe* suggest an even higher-level motivation for `consuming` information, as *predict* and *prescribe* are often associated with decision making processes. We are unique in that we include the term `produce`, the converse of `consume`, which allows us to account for tasks that `produce` new information. Moving further down the *why* part of our typology, we claim a unique characterization of `search` that depends on whether the identify and location of the search target are known a priori. Our characterization of `search` overlaps somewhat with Schulz and colleagues' *goals* pertaining to *directed* and *undirected search*; it also relates to their definition of *navigation*, listed as a *means* in their design space. Our characterization of `search` also differs from that of Roth's taxonomy, which includes several

[2]Amar et al. [8], Chuah and Roth [60], Pirolli and Card [242], Shneiderman [291], Wehrend and Lewis [349], Yi et al. [366], Zhou and Feiner [370]. We also have 12 references in common with both Roth [262] and Schulz et al. [280], who in turn share 8 common references that we do not cite.

related `search` terms but defined as *operators*. With regards to our `query` types: `identify`, `compare`, and `summarize`, there is an obvious mapping to the *cardinality* dimension of Schulz et al.'s design space, which refers to *single instances*, *multiple instances*, and *all instances* of available data; meanwhile, our `query` types would be *objectives* according to Roth, which increase in sophistication along with *cardinality*.

Our question of *how* a task is supported has a clearer correspondence to both Schulz et al.'s *means* and Roth's *operators*; in particular, how we distinguish `manipulate` and `introduce`[3] partially corresponds with Roth's distinction between *enabling operators* and *work operators*. Nevertheless, there are still a number of subtle differences worth pointing out. For one, we include `encode`, which relates directly to how the data is visually encoded. Based on the absence or presence of other methods in a task description, we can discern between static and interactive visualization artefacts. Another term unique to our typology is `change`, a verb which might appear vague without a qualifying noun; this term is found in many previous classifications, perhaps pertaining to a change in visual encoding parameters, scale, or an animated transformation. When used in the context of a full task description that includes explicitly defined `inputs` and `outputs`, the meaning of `change` is no longer unclear.

Another prominent difference from our typology and the classifications of Roth [262] and Schulz et al. [280] is in our handling of *what*. Our characterization of *what*, comprising of a task's `inputs` and `outputs`, overlaps partially with several dimensions of the Schulz et al.'s design space, as well as with Roth's *operands*. In early drafts, we tried to classify *what* comprises a visualization in more detail (documented throughout Appendix A), but our classification became far too complicated, and we decided to talk about *what* in greater detail in Munzner [219][4]. We simplified to the agnostic "bring your own *what*" mentality discussed in Section 2.3.3, as we

[3]`introduce` nodes became forms of `produce` in Munzner's extension to the typology [219].

[4]Summarized in Figure 6.3, above; it is also worth noting that Munzner [219] also uses the term `targets`, a term used by Schulz et al. [280].

realized that as long as one specifies the `inputs` and `outputs` of a task, any classification of visualization elements can be used, including the *operand* classification by Roth [262] or the *characteristics* and *target* classifications by Schulz et al. [280].

Yet another difference is that Roth [262] and Schulz et al. [280] do not explicitly accommodate the casual use of visualization artefacts, this being usage motivated not by a need to `present` or `discover` information, whereas we use the term `enjoy` to describe such tasks.

A final difference to illustrate, particularly between our typology and the classification by Schulz et al. [280], is how these classifications can be used to describe task sequences. Schulz et al. [280] provide a textual notation for describing what they refer to as workflows, and they offer an example in their paper where they use their (*goals, means, characteristics, target, cardinality*) notation to describe the Shneiderman's well-known *visual information seeking mantra* [291]: *overview first, zoom & filter, details-on-demand*:

$$(\textit{exploratory}, \textit{summarize}, *, *, \textit{all}) \Rightarrow$$
$$(\textit{exploratory}, \textit{elaborate} \mid \textit{filter}, *, *, \textit{multiple}) \Rightarrow$$
$$(\textit{exploratory} \mid \textit{confirmatory}, \textit{gather} \mid \textit{look-up}, *, \textit{single})$$

Our corresponding visual notation, shown in Figure 6.5, captures the same three steps while explicitly indicating the links between `inputs` and `outputs` of subsequent tasks in the sequence. In the *overview first* task, a person `explores` and `summarizes` an *overview of the data*, which is supported by the visual `encoding`. In the *zoom & filter* step, she `browses` the *overview* to `identify` a *subset of items* that interests her, supported by `navigation` and `filtering`. Finally, in the *details-on-demand* step, she `browses` that *subset* and `identifies` a particular item that interests her, `navigating` and `selecting` it to learn more. Also note that each part of the *overview first, zoom & filter, details-on-demand* mantra is about `consuming` information, and it could apply equally to `present`, `discover`, or `enjoy` contexts.

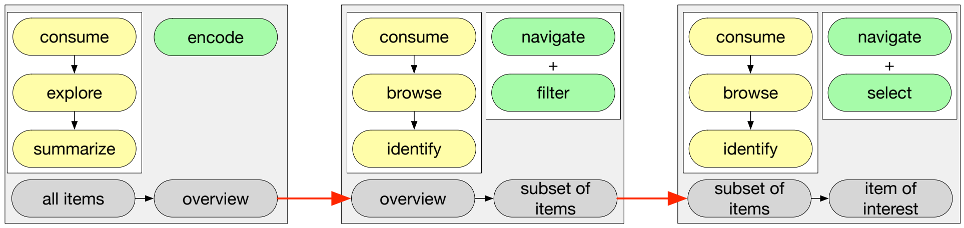


Figure 6.5: Our typology is used to describe Shneiderman's *visual information seeking mantra*: *overview first, zoom & filter, details-on-demand*.

6.1.3 Impact of the Task Typology

In addition to playing an important role in the three other projects described in Chapter 3, 4, and 5, our task typology has had considerable impact within the visualization community. Our task typology paper [33] has been cited over 60 times[5], and it is one of the most-cited papers from IEEE *Transactions on Visualization and Computer Graphics* published since Fall 2013 (Volume 19, Number 12, Proceedings of InfoVis 2013).

Papers written by other researchers which cite our typology paper include eight IEEE *Transactions on Visualization and Computer Graphics* journal papers [28, 169, 171, 269, 272, 278, 287, 369][6], an IEEE *Symposium on Visual Analytics Science and Technology (VAST)* paper [112], three *Computer Graphics Forum (Proceedings of EuroVis)* journal papers [65, 274, 342], four *EuroVis Short Papers* [213, 230, 271, 273], two *EuroVis State of the Art Reports* [27, 340], three *Information Visualization* journal articles [258, 295, 359], and an ACM *Human Factors in Computing Systems (CHI)* conference paper [311]. Our typology paper has also been cited in papers presented at various other conferences, symposia, and workshops, as well as in a number of journal articles, technical reports, and theses. In addition to informing the work of other researchers, we have also referred to our task typology in other subsequent papers appearing in IEEE *Transactions on Visualization and Computer Graphics* [106], the *Information Visualiza-*

[5]According to Google Scholar, excluding self-citations, as of April 7, 2016.

[6]According to Google Scholar, IEEE *TVCG* papers have the second highest impact factor in the field of computer graphics, following ACM *Transactions on Graphics*.

tion journal [74, 211], and at the ACM BELIV workshop [34].

Other researchers have discussed the typology in the context of new models or frameworks. In the *knowledge generation model for visual analytics* by Sacha et al. [269], our typology is invoked in relation to their discussion of a *exploration loop* and a *verification loop*, wherein lower-level actions are associated with the former and higher-level actions are associated with the latter. Meanwhile, in the *algebraic process for visualization design* by Kindlmann and Scheidegger [171], the concept of *data symmetries*, or transformations in data space, are associated with the `manipulate` nodes in the *how* part of our typology. Rind et al. [258] discuss our typology (among many other classifications of tasks) in the context of their meta-analysis of task classifications, defining three dimensions for describing the space of task classifications: *abstraction* (concrete vs. abstract), *composition* (high-level vs. low-level), and *perspective* (*why* vs. *how*). Finally, in a conceptual model for characterizing domain problems by Winters et al. [359], they describe different types of people who use visualization tools, and how these people are typically either those that `produce` information or those that `consume` information.

There are several new task classifications specific to particular datatypes that adopt the vocabulary and/or the *why-what-how* structure of our typology, including those relating to node-link-group data [271, 272], cartograms [230], multiplex networks [253], and multidimensional data projections [94]. Other novel datatype-specific classifications that refer to our typology include classifications of tasks for multivariate network analysis [249] and temporal graphs [169]; while these classifications do not adopt our typology's vocabulary or structure, their authors reiterate our arguments about the purpose of classifying visualization tasks, and Kerracher et al. [169] they explicitly motivate their work in response to our call for more datatype-specific classification of tasks (see Section 2.5.1).

There are also several application-domain specific classifications of tasks that adopt or extend our typology, again using our vocabulary and/or the *why-what-how* structure, including those specific to bioinformatics [212], computational chemistry [333], telemedicine [314], and malware analysis [340]. One particularly unusual and fascinating adoption of our typology is a classification of tasks related to artistic human plant interfaces [350].

Applying our typology to a specific domain or datatype is not always a straightforward endeavour, as indicated by Sedlmair et al. [287] in their classification of visual parameter space analysis tasks:

> *"While we found the general categories of why and how helpful in guiding our analysis, we could not directly match our framework into this typology. Our work addresses analysis tasks specific to visual parameter space analysis that have not been discussed in their typology."* [287, p. 2167]

Sedlmair et al. [287] reiterate our call for additional datatype-specific and domain-specific classifications of tasks, and they ultimately characterize six tasks: *optimization*, *partitioning*, *fitting*, *outliers*, *uncertainty*, and *sensitivity*. It would be an interesting exercise to express these six tasks using the vocabulary and structure of our typology, tasks which are highly intertwined with other data analysis and simulation activities beyond those that directly involve a visual encoding.

Others have used our typology in specifying the tasks addressed by novel visualization techniques or tools, such as in a technique for analyzing traffic flows [278], in the design of a multi-touch tangible user interface for biological data visualization [3], in ColorCAT, a colour map creation tool [213], or in VA^2, a tool for evaluating the eye movements and interactions of people while using visual analytics tools [28]. Our typology has also been used to describe the capabilities of a family of off-screen visualization techniques [154]. We also note the use of our typology to characterize the tasks addressed by VisExpress, a recent design study project in the bioinformatics domain [295].

Finally, Saket et al. [274] have used our typology to specify tasks in an experiment comparing map-based and node-link graph layouts.

Altogether, we are very pleased with how our typology has been adopted. In Section 2.6, we proposed ways in which our typology could be used to *describe* (or *analyze*) the use of visualization tools or techniques, *evaluate* these tools or techniques, and *generate* new designs; this brief survey has revealed that the typology has accomplished all of these goals: it has described the use of visualization tools and techniques in the context of specific domains or datatypes, it has been used in the specification of tasks for experimental evaluation studies, and it has grounded the design of new visualization tools and techniques.

We also note the variability with respect to how aspects of our typology have been used. While some adopt the *why-what-how* framing, others use specific vocabulary from parts of our typology, while some have even adopted the visual style of our task diagrams, such as those used in Chapter 2 (e.g., [154, 333]). This variability is not surprising; we did not imagine or intend that our typology would be used to develop a rigid specification of tasks; we reiterate that it our typology is a proposal of how to classify tasks: it is not the final answer, and we invite others to extend it and propose alternatives. Our own use of the typology has also varied and evolved between Chapter 3, Chapter 4, and Chapter 5, especially since the advent of Munzner's extension to the typology [219], and we expect that our use of the typology will continue to evolve in our future projects.

Other uses of the typology: In addition to its adoption and extension by the visualization research community, the typology has also appeared in other contexts.

For instance, our typology was featured in a 2014 IEEE VIS conference tutorial by McNamara et al. [208] on practical methods for design studies, where our typology was referred to as a viable tool for task abstraction following the work domain analysis phase of a design study project.

Another interesting use of the typology was in an interactive faceted browser by Yalçın [364], in which he visualized the vocabulary and bibliography of our typology paper[7].

Our typology paper [33] has also been included on the syllabi of several graduate courses in visualization[8]; many visualization courses are also making use of Munzner's book [219], and thus are reading about our typology

[7] `http://keshif.me/demo/vis_tasks.html`

[8] These include courses taught at the University of Chicago (`http://goo.gl/htBFju`), the University of Utah (`http://goo.gl/cLsJmY`), the University of Wisconsin-Madison (`http://goo.gl/lsRAHC`), Tufts University, and at the University of British Columbia (`http://goo.gl/Xt7ry2`)

indirectly[9]. Our paper is also listed on the computer science doctoral qualifying resources list for students pursuing research in information visualization at Georgia Tech[10]. These institutions are well-known for visualization research, and it is encouraging to see that our typology is reaching the next generation of researchers and practitioners.

6.2 Reflecting on the Interview Study

The study reported in Chapter 3 was preceded by a technical report entitled *Dimensionality Reduction in the Wild: Gaps and Guidance* [283]; in addition to informing our own subsequent work [125, 146, 148, 286, 287], the paper and technical report have garnered over fifteen citations by others in the visualization community[11].

Like our original task typology in Chapter 2, our classification of task sequences relating to dimensionally reduced data can also be discussed in relation to subsequent work. One of the papers mentioned above that extends and refers to the *why-what-how* structure of our original typology is a datatype specific classification of tasks by Etemadpour et al. [94], one specific to multidimensional data projections (in other words, visual encodings of dimensionally reduced data). Etemadpour et al. [94] consolidate all of the tasks that are explicitly or implicitly addressed in previous DR technique papers. They arrive at a classification of four groups of tasks (35 tasks in total): seven *pattern identification* tasks, fourteen *relation-seeking* tasks, ten *behaviour comparison* tasks, and four *membership disambiguation* tasks. The individual tasks are quite specific and are worded much like instructions in a controlled experiment, such as *estimate the number of observed clusters* (a *pattern identification* task), or *find the cluster with the most/least number of points or size* (a *behaviour comparison* task). Though they do not define each of the 35 tasks in detail, they do provide two examples in which they decompose the task using our typology's vocabulary and *why-what-how* structure.

[9] A list of these courses can be found here: `http://cs.ubc.ca/~tmm/vadbook/`

[10] `http://goo.gl/GY3hQw`

[11] According to Google Scholar (as of April 7, 2016)

Recall that in Chapter 3, we characterized five task sequences relating to the visualization of dimensionally reduced data: two focused on dimensions (*name synthetic dimensions* and *map synthetic dimensions to original dimensions*), while three focused on clusters (*verifying that clusters exist*, *naming clusters*, and *matching clusters and classes*). In contrast, the list of 35 tasks by Etemadpour et al. [94] pertains almost exclusively to clusters and their spatial or quantitative properties (30 tasks), while the remaining five tasks focus on individual points or outliers; there are no tasks pertaining to *naming the projected synthetic dimensions* or *understanding the relation between synthetic dimensions and original dimensions*, nor are there tasks about *naming clusters* or *matching clusters and classes*. There is nevertheless some overlap between our work and that of Etemadpour et al. [94]: the task sequence that we refer to as *verifying that clusters exist* could correspond with their task of *estimating the number of observed clusters*. Arguably, the tasks by Etemadpour et al. [94] are stated at a very low level of abstraction, and that when executed in sequence, they might be described using our cluster-related task sequences. The disconnect between our work and that of Etemadpour et al. [94] is interesting, particularly when the provenance of our respective task classifications are considered: ours was largely informed by an interview study, while theirs largely informed from a survey of the DR technique literature, suggesting that the applied use of various DR techniques may be different from what their developers had anticipated.

Our classification of task sequences for dimensionally reduced data has informed the design of a recent tool for *probing projections* by Stahnke et al. [302], a tool that allows analysts to `encode` the correspondence between the original set of dimensions and a projection of the data along two synthetic dimensions that were generated by way of MDS. The same tool also allows analysts to `select` clusters of data points in the two-dimensional scatterplot projection, and as a result, histograms of this selection are overlaid on histograms of the original set of dimensions. Stahnke et al. [302] explicitly ground their design around our set of task sequences, and their tool is perhaps the first to provide data analysts with an elegant interactive way to map synthetic dimensions onto the set of original dimensions, as well as a means to name clusters using interactive linked views.

Future work: We intend to use our classification of task sequences in the design and evaluation of software tools to support them in the context of domain-specific workflows. Finally, we hope to expand upon this set of task sequences to characterize additional high-dimensional data tasks, including those related to the use of *dimensional filtering* techniques.

6.3 Reflecting on the Field Study

A limitation of adoption-phase research is that a set of individuals from a specific application domain cannot be identified in advance, in contrast to a typical design study chronology [284]. As a result, there is an inherent selection bias in our case studies of journalists who used *Overview*, because they largely represent successful cases; a similar observation was made by McKeon [205], who interviewed people who used his deployed visualization tool prolifically. In future work, we would like to know more about cases in which *Overview* was used briefly and then abandoned as being unsuitable for the problem at hand.

To broaden our understanding of how *Overview* is used, we hope to investigate the use of *Overview* in other domains where large document collections are prevalent, such as intelligence analysis [165], law [123], medicine, and digital humanities research. Consequently, we expect that our set of task abstractions may continue to expand.

Meanwhile, we are continuing to monitor and learn from new cases of adoption by journalists. In addition to the case studies featured in Chapter 4 [110, 167, 236, 306, 339], recent stories in which *Overview* was used as part of an investigation include those about veteran medical benefits [99], problems with an online food stamp registration service [81][12], American Senator John McCain's memos [347], the fine print of credit card agreements [356], and an amusing analysis of comedian Louis C. K.'s emails [179][13].

[12] Journalist Tyler Dukes [82] also wrote a blog post about how *Overview* was used during his investigation.

[13] As of April 7, 2016, we are aware of twenty stories where *Overview* was as used

Our 2014 paper about the *Overview* field study [35] has been cited over a dozen times by other researchers [14], and we are optimistic that an interest in visualization for journalism and the digital humanities will continue to grow in the research community.

Another recent development is the *Overview* API, which allows developers to use alternative visual encodings to represent their document collections and their tags, such as tag clouds or a DocuBurst radial-hierarchical encoding [64]. We have recently began to use the *Overview* API to integrate TimeLineCurator [106], our tool that uses natural language processing (NLP) to automatically extract temporal references from unstructured text documents in order to generate visual timelines. Previously, TimeLineCurator was a standalone tool; its integration with *Overview* will allow for visual timelines to be generated directly from *Overview* document collections.

Finally, in the time that has passed since the publication of our *Overview* field study paper [35], we have come to realize that the two abstract tasks that *Overview* supports, (i) generating hypotheses by summarizing themes and (ii) verifying hypotheses by locating evidence, reflect the views and values of journalists [134]. For any investigation, whether it involves large document collections or other types of information, journalists tend to gravitate toward one of these two tasks. We arrived at this characterization of tasks via an analysis of the processes of six journalists who used *Overview*; investigative journalists already understand this characterization well as a result of their training and on-the-job experience. In other words, though our journeys differed, we arrived at a definition that is already well understood by journalists.

6.4 Reflecting on the Design Study

In Chapter 5, we focused on the design and evaluation of a visual analysis tool for energy management, as we remarked upon how our designs were

during an investigation leading to a published story: `https://github.com/overview/overview-server/wiki/News-stories`

[14] According to Google Scholar (as of April 7, 2016).

adopted into our collaborators' production timeline. Since mid-2014, our collaborators have committed over ten full-time developers to the project. Energy Manager is currently being piloted with client organizations[15]; in the future, we would like to assess the adoption of the redesigned Energy Manager following its wide-scale deployment. We will track usage over an extended period of time and we hope to speak to more energy analysts via interviews and focus groups.

Though it is still too early to gauge the impact of this work on the visualization research community, we hope that our methodological considerations inspire future visualization design studies in various application domains.

Open questions: The results of our design study have motivated a couple of interesting new directions for future research.

First, let us consider the role of domain convention: how and when should visualization designers adhere to convention, and when should they attempt to break it? When a visualization design choice that breaks domain convention proves to be successful, where success might be measured in terms of adoption of the visual encoding or interaction design choice within the domain, what are the factors that contribute to its success? A recent survey of graphical conventions in the visualization and infographic design communities [45] suggests that more inquiry is needed into domain-specific graphical conventions, and that curated lists of domain-specific conventions would be useful resources for visualization practitioners who engage with multiple domains.

Related to domain convention is the concept of familiarity and the question of how to introduce unfamiliar visual encodings to those familiar with only a small set of ubiquitous visualization design choices, such as bar charts, line graphs, scatterplots, and pie charts, or to those familiar only with visualization conventions used within their own domain. In the energy management domain, many of the energy analysts that we spoke to worked primarily with bar charts and line graphs, and yet our eventual design involved the juxtaposition and linking of two unfamiliar visual encodings: a

[15] As of November 2015.

time series matrix with auxiliary boxplots, a design that was positively received and green-lighted for production by our collaborators. We now need to return to laboratory settings to better understand when and how the juxtaposition of unfamiliar visual encodings is effective: to test juxtapositions of unfamiliar visual encodings, juxtapositions or familiar and unfamiliar, and also to examine the role of the number of distinct views in a single display, as well as the role of interactive linking and brushing across these views. Considerations for the design of multiple-view visualization tools are well-documented, both by us [181] and by Weaver [348], and include questions such as *how many discrete views are appropriate*, *how should views be arranged or sequenced* and *how should views be coordinated*, such as with linking and brushing techniques; we now need to examine the factor of familiarity in the design of these multiple-view visualization tools. Ruchikachorn and Mueller [264] have recently examined the role of smoothly animating between familiar and unfamiliar visual encodings, such as between a data table to a parallel coordinates plot, or between a pie chart and a treemap. We might now ask: in which circumstances is it preferable to toggle an animated transition between a familiar to an unfamiliar visual encoding, and in which circumstances is it preferable to juxtapose and link the familiar and unfamiliar?

6.5 Concluding Thoughts

We return now to the question stated in Chapter 1: *why do people visualize data?* Now we have a systematic approach to answering this question, one that allows us to describe the use of visualization tools and techniques with the vocabulary and structure of our abstract task typology.

Ultimately, people visualize data to `consume` or `produce` information. When one `consumes` information, they may be doing so to `discover`, to `generate` or `verify` hypotheses. Or they may `present` information to others so that they too may consume the information that is to be communicated. Or perhaps a person might `consume` information merely to `enjoy` or play with data relevant to a casual interest. On the other hand, one may visualize data as a means to `produce` new information: to `derive` new data, to `annotate` the data with additional contextual information or personal insights, or to `record` an analysis process to ensure reproducibility or credibility. At a lower level of abstraction, one may visualize data in order to `search` for information, to `identify` items in the data, to `compare` items, or to `summarize` all of them. All of the aforementioned terms introduced in Chapter 2 serve to describe *why* people visualize data, and when combined with a description of *how* a visualization tool or technique supports the task and *what* the task's `inputs` and `outputs` are, we have an even more complete picture, and we can combine these task descriptions together to form task sequences that reflect real domain-specific workflows.

We applied our typology of abstract visualization tasks in three projects, allowing us to better understand *why people visualize data* in specific contexts.

In Chapter 3, we discovered *why* a data analyst would visualize dimensionally reduced data. They do so to better understand the synthetic dimensions resulting from the use of DR: to name these synthetic dimensions and to understand the mapping between them and the original set of dimensions. Or they might visualize this data in order to verify the existence of clusters of items, to name these clusters, and to match these clusters with preexisting class labels.

In Chapter 4, we discovered *why* a journalist would use a visualization tool called *Overview* to investigate large collections of text documents. Initially, we thought that *Overview* would be used to `generate` hypotheses, and to `explore` and `summarize` a document collection. In a post-deployment field study of *Overview*'s self-initiated adoption by investigative journalists, we discovered, to our surprise, that *Overview* was also used to `verify` hypotheses, to `locate` and `identify` specific pieces of evidence within a document collection.

In Chapter 5, we considered *why* an energy analyst who oversees organizational energy usage in large building portfolios would visualize data in order to `lookup` and `summarize` the energy performance of buildings in their portfolio, to `locate` and `compare` trends and outliers in subsets of their portfolio, and to `identify` the proportion of energy used by a single building relative to its building group or to the entire portfolio. This understanding of tasks and the sequential relationships between them led us to design a series of prototypes, and through an iterative feedback process with energy analysts, some of our designs have been incorporated into a forthcoming release of a commercial energy analysis software tool.

Along the way, our efforts have provided us and the visualization community with implications for visualization design that may transfer to situations involving similar tasks and datatypes. In the case of our interview study about visualizing dimensionally reduced data, the community has responded with the design of a new technique [302]. From the *Overview* field study, we provided an approach to studying the adoption of visualization tools, as well as a discussion on the merits of multiple-view tools for analyzing document collections. Finally, from the energy management design study, we provided a discussion on the themes of familiarity, trust, and domain convention, as well as methodological guidance for visualization design studies.

The response by the visualization community and its adoption of our task typology has been tremendous. We hope that our approach to analyzing visualizations tasks will continue to resonate with visualization researchers, students, and practitioners in the years to come. We also anticipate that our typology will continue to evolve, that it will be applied to new domains and datatypes, that it will facilitate the specification of tasks in experimental studies, and that it will motivate and contextualize new designs.

Bibliography

[1] S. N. Abdolyousefi. *Equilibria of a Nonlocal Model for Biological Aggregations: Linear Stability and Bifurcation Studies.* MSc thesis, Simon Fraser University, 2011. `http://summit.sfu.ca/item/11771`. → pages 286

[2] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data.* Springer, 2011. → pages 14, 28, 35, 39, 120, 123, 276, 278, 372, 375

[3] S. Al-Megren and R. A. Ruddle. Comparing tangible and multi-touch interaction for interactive data visualization tasks. In *To appear in the proceedings of the ACM International Conference on Tangible, Embedded and Embodied Interaction (TEI)*, 2016. → pages 155

[4] D. Albers, M. Correll, and M. Gleicher. Task-driven evaluation of aggregation in time series visualization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 551–560, 2014. `http://dx.doi.org/10.1145/2556288.2557200`. → pages 121, 123, 388

[5] G. Albuquerque, M. Eisemann, D. J. Lehmann, H. Theisel, and M. Magnor. Improving the visual analysis of high-dimensional datasets using quality measures. *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 19–26, 2010. `http://dx.doi.org/10.1109/VAST.2010.5652433`. → pages 54

[6] K. Altun, B. Barshan, and O. Tunçel. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognition*, 43(10):3605–3620, 2010. `http://dx.doi.org/10.1016/j.patcog.2010.04.019`. → pages 286

[7] R. Amar and J. T. Stasko. A knowledge task-based framework for design and evaluation of information visualizations. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 143–150, 2004. http://dx.doi.org/10.1109/INFVIS.2004.10. → pages 21, 22, 26, 39, 40, 41, 43, 53, 214, 222, 228, 260, 299, 302, 335, 340, 351

[8] R. Amar, J. Eagan, and J. T. Stasko. Low-level components of analytic activity in information visualization. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 111–117, 2005. http://dx.doi.org/10.1109/INFVIS.2005.1532136. → pages 3, 20, 21, 23, 24, 26, 30, 35, 38, 39, 40, 41, 42, 43, 53, 150, 226, 228, 230, 238, 247, 248, 260, 261, 302, 310, 335, 352, 356, 358

[9] P. André, J. Teevan, and S. T. Dumais. Discovery is never by chance: Designing for (un)serendipity. In *Proceedings of the ACM Conference on Creativity & Cognition*, pages 305–314, 2009. http://dx.doi.org/10.1145/1640233.1640279. → pages 29, 39, 41, 220, 222, 356

[10] K. Andrews. Evaluation comes in many guises. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2008. http://goo.gl/eya0cc. → pages 350

[11] S. Andrews, G. Hamarneh, and A. Saad. Fast random walker with priors using precomputation for interactive medical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 9–16, 2010. http://dx.doi.org/10.1007/978-3-642-15711-0_2. → pages 286

[12] G. Andrienko and N. Andrienko. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer-Verlag, 2006. → pages 21, 24, 26, 30, 31, 35, 39, 40, 41, 277, 278

[13] K. D. Bailey. *Typologies and Taxonomies: An Introduction to Classification*. Sage Publications, 1994. → pages 27, 280

[14] M. Balasubramanian and E. Schwartz. The Isomap algorithm and topological stability. *Science*, 295:7, 2002. http://dx.doi.org/10.1126/science.295.5552.7a. → pages 329

[15] D. Baur, J. Büttgen, and A. Butz. Listening factors: A large-scale principal components analysis of long-term music listening histories. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1273–1276, 2012. http://dx.doi.org/10.1145/2207676.2208581. → pages 286

[16] M. Beaudouin-Lafon. Designing interaction, not interfaces. In *Proceedings of the ACM Conference on Advanced Visual Interfaces (AVI)*, pages 15–22, 2004. http://dx.doi.org/10.1145/989863.989865. → pages 4, 47, 226, 247, 263, 264, 352, 365

[17] R. A. Becker, W. S. Cleveland, and M.-J. Shyu. The visual design and control of trellis display. *Computational and Statistical Graphics*, 5(2):123–155, 1996. http://dx.doi.org/10.1080/10618600.1996.10474701. → pages 24, 271

[18] B. B. Bederson and B. Shneiderman. Theories for understanding information visualization. In *The Craft of Information Visualization: Readings and Reflections*, pages 349–351. Morgan-Kaufmann, 2003. → pages 47, 268

[19] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 585–591, 2001. http://goo.gl/F9PJ2R. → pages 298, 304, 327

[20] R. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961. → pages 322

[21] S. Bergner, M. Sedlmair, T. Möller, S. N. Abdolyousefi, and A. Saad. ParaGlide: Interactive parameter space partitioning for computer simulations. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 19(9):1499–1512, 2013. http://dx.doi.org/10.1109/TVCG.2013.61. → pages 314

[22] J. Bertin. *Semiology of graphics: Diagrams, networks, maps*. Esri Press, 2011. Translated by W. J. Berg. → pages 272, 277

[23] E. Bertini, A. Perer, C. Plaisant, and G. Santucci. Beyond time and errors: Novel evaluation methods for information visualization. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel*

evaLuation methods for Information Visualization (BELIV), 2008. http://dx.doi.org/10.1145/1168149.1168152. → pages 351

[24] E. Bertini, A. Tatu, and D. A. Keim. Quality metrics in high-dimensional data visualization: an overview and systematization. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 17(12):2203–2212, 2011. http://dx.doi.org/10.1109/TVCG.2011.229. → pages 54, 302

[25] E. Bingham and H. Mannila. Random projection in dimensionality reduction: applications to image and text data. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pages 245–250, 2001. http://dx.doi.org/10.1145/502512.502546. → pages 310

[26] G. Blackshields, F. Sievers, W. Shi, A. Wilm, and D. G. Higgins. Sequence embedding for fast construction of guide trees for multiple sequence alignment. *Algorithms for Molecular Biology*, 5(21):1–11, 2010. http://dx.doi.org/10.1186/1748-7188-5-21. → pages 286

[27] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. State-of-the-art of visualization for eye tracking data. In *Proceedings of the Eurographics Conference on Visualization (EuroVis) State of the Art Reports*, 2014. http://dx.doi.org/10.2312/eurovisstar.20141173. → pages 153

[28] T. Blascheck, M. John, K. Kurzhals, S. Koch, and T. Ertl. VA^2: A visual analytics approach for evaluating visual analytics applications. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VAST 2015)*, 22(1):61–70, 2016. http://dx.doi.org/10.1109/TVCG.2015.2467871. → pages 153, 155

[29] M. Booshehrian, T. Möller, R. M. Peterman, and T. Munzner. Vismon: Facilitating analysis of trade-offs, uncertainty, and sensitivity in fisheries management decision making. *Computer Graphics Forum (Proceedings of EuroVis)*, 31(3):1235–1244, 2012. http://dx.doi.org/10.1111/j.1467-8659.2012.03116.x. → pages 286, 301, 314

[30] M. Bostock, V. Ogievetsky, and J. Heer. D^3: Data-driven documents. *IEEE Transactions on Visualization and Computer*

Graphics (Proceedings of InfoVis), 17(12):2301–2309, 2011. http://dx.doi.org/10.1109/TVCG.2011.185. → pages 123, 378

[31] I. Boyandin, E. Bertini, and D. Lalanne. A qualitative study on the exploration of temporal changes in flow maps with animation and small-multiples. *Computer Graphics Forum (Proceedings of EuroVis)*, 31(3pt2):1005–1014, 2012. http://dx.doi.org/10.1111/j.1467-8659.2012.03093.x. → pages 300

[32] U. Brandes and C. Pich. Eigensolver methods for progressive multidimensional scaling of large data. In *Proceedings of the International Symposium on Graph Drawing*, pages 42–53, 2007. http://dx.doi.org/10.1007/978-3-540-70904-6_6. → pages 327

[33] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 19(12):2376–2385, 2013. http://dx.doi.org/10.1109/TVCG.2013.124. → pages iv, xxiv, 20, 142, 143, 149, 153, 156, 213, 216, 217, 218, 219, 221, 223, 232, 234, 240, 241, 243, 246, 255, 256, 257, 261, 265, 266, 273, 280

[34] M. Brehmer, S. Carpendale, B. Lee, and M. Tory. Pre-design empiricism for information visualization: Scenarios, methods, and challenges. In *Proceedings of the ACM Workshop on Beyond Time And Errors: Novel Evaluation Methods For Visualization (BELIV)*, pages 147–151, 2014. http://dx.doi.org/10.1145/2669557.2669564. → pages 17, 67, 109, 137, 154

[35] M. Brehmer, S. Ingram, J. Stray, and T. Munzner. Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 20(12):2271 – 2280, 2014. http://dx.doi.org/10.1109/TVCG.2014.2346431. → pages v, 70, 73, 104, 160, 286

[36] M. Brehmer, M. Sedlmair, S. Ingram, and T. Munzner. Visualizing dimensionally-reduced data: Interviews with analysts and a characterization of task sequences. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, pages 1–8, 2014.

http://dx.doi.org/10.1145/2669557.2669559. → pages iv, v, xxiii, 51

[37] M. Brehmer, J. Ng, K. Tate, and T. Munzner. Matches, mismatches, and methods: Multiple-view workflows for energy portfolio analysis. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis 2015)*, 22(1):449–458, 2016. http://dx.doi.org/10.1109/TVCG.2015.2466971. → pages v, 106, 139, 143

[38] H. Bretschneider. Leaksplorer Beta, 2012. https://twitter.com/leaksplorer. → pages 77, 97

[39] A. Bronstein, M. Bronstein, A. Bruckstein, and R. Kimmel. Matching two-dimensional articulated shapes using generalized multidimensional scaling. In *Proceedings of the Conference on Articulated Motion and Deformable Objects (AMDO)*, pages 48–57, 2006. http://dx.doi.org/10.1007/11789239_6. → pages 298, 308, 309

[40] J. Buhl, D. J. T. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller, and S. J. Simpson. From disorder to order in marching locusts. *Science*, 312(5778):1402–1406, 2006. http://dx.doi.org/10.1126/science.1125142. → pages 286

[41] A. Buja and D. Swayne. Visualization methodology for multidimensional scaling. *Classification*, 19:7–43, 2002. http://dx.doi.org/10.1007/s00357-001-0031-0. → pages 52, 58, 59, 61, 298, 308, 309, 314

[42] A. Buja, D. Cook, and D. F. Swayne. Interactive high-dimensional data visualization. *Computational and Graphical Statistics*, 5(1): 78–99, 1996. http://dx.doi.org/10.1080/10618600.1996.10474696. → pages 21, 26, 31, 39, 40, 41, 42, 238, 271

[43] J. Büttgen. *What's in a History? A Large-Scale Statistical Analysis of Last.fm Data*. MSc thesis, Ludwig-Maximilians Universitat Munchen, 2010. . → pages 286

[44] W. Buxton. Chunking and phrasing and the design of human-computer dialogues. In *Proceedings of the IFIP World*

Computer Congress, pages 475–480, 1986.
http://www.billbuxton.com/chunking.pdf. → pages 265

[45] L. Byrne, D. Angus, and J. Wiles. Acquired codes of meaning in data visualization and infographics: Beyond perceptual primitives. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis 2015)*, 22(1):509–518, 2016. http://dx.doi.org/10.1109/TVCG.2015.2467321. → pages 161

[46] L. Byron and M. Wattenberg. Stacked graphs – geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 14(6):1245–1252, 2008. http://dx.doi.org/10.1109/TVCG.2008.166. → pages 129

[47] S. K. Card and J. D. Mackinlay. The structure of the information visualization design space. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 92–99, 1997. http://dx.doi.org/10.1109/INFVIS.1997.636792. → pages 265

[48] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, 1999. → pages 21, 22, 24, 26, 34, 39, 40, 41, 42, 46, 216, 234, 272

[49] S. Carpendale. Evaluating information visualizations. In A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, chapter 2, pages 19–45. Springer, 2008. http://dx.doi.org/10.1007/978-3-540-70956-5_2. → pages 54, 288, 300, 305, 351

[50] D. O. Case. *Looking for Information: A Survey of Research on Information Seeking, Needs and Behavior*. Emerald Group Publishing, 2nd edition, 2008. → pages 256, 257, 258, 270

[51] S. M. Casner. A task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics*, 10(2): 111–151, 1991. http://dx.doi.org/10.1145/108360.108361. → pages 21, 24, 26, 39, 40, 41, 42, 228, 235

[52] A. J.-B. Chaney and D. M. Blei. Visualizing topic models. In *Proceedings of the International AAAI Conference on Weblogs and*

Social Media (ICWSM), pages 419–422, 2012.
http://goo.gl/nqmKPY. → pages 76, 93

[53] R. Chang, C. Ziemkiewicz, T. M. Green, and W. Ribarsky. Defining insight for visual analytics. *IEEE Computer Graphics and Applications (CG&A)*, 29(2):14–17, 2009.
http://dx.doi.org/10.1109/MCG.2009.22. → pages 222, 356

[54] R. Chang, C. Ziemkiewicz, R. Pyzh, J. Kielman, and W. Ribarsky. Learning-based evaluation of visual analytic systems. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2010.
http://dx.doi.org/10.1145/2110192.2110197. → pages 351, 357, 367

[55] K. Charmaz. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis.* Sage Publications, 2006. → pages 55, 247, 288, 300, 305, 307, 330, 336, 339, 340, 355, 357, 358

[56] L. Chen and A. Buja. Stress functions for nonlinear dimension reduction, proximity analysis, and graph drawing. *Machine Learning Research*, 14:1145–1173, 2013. http://goo.gl/EmJtIf. → pages 305

[57] Y. Chen, L. Wang, M. Dong, and J. Hua. Exemplar-based visualization of large document corpus. *IEEE Trans. Visualization and Computer Graphics (Proceedings of InfoVis)*, 15(6):1161–1168, 2009. http://dx.doi.org/10.1109/TVCG.2009.140. → pages 77, 97

[58] E. H. Chi and J. T. Riedl. An operator interaction framework for visualization systems. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 63–70, 1998.
http://dx.doi.org/10.1109/INFVIS.1998.729560. → pages 21, 26, 34, 38, 39, 40, 42, 231, 234, 238, 263, 264

[59] D. Child. *The Essentials of Factor Analysis.* Continuum, 2006. → pages 304

[60] M. C. Chuah and S. F. Roth. On the semantics of interactive visualizations. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 29–36, 1996.
http://dx.doi.org/10.1109/INFVIS.1996.559213. → pages 21,

23, 26, 34, 37, 38, 39, 40, 41, 42, 150, 234, 237, 238, 243, 245, 247, 248, 263, 264

[61] J. Chuang, D. Ramage, C. D. Manning, and J. Heer. Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 443–452, 2012. http://dx.doi.org/10.1145/2207676.2207738. → pages 96, 136

[62] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387): 531– 554, 1984. http://dx.doi.org/10.1080/01621459.1984.10478080. → pages 23, 48

[63] R. R. Coifman. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences*, 102(21), 2005. http://dx.doi.org/10.1073/pnas.0500334102. → pages 304

[64] C. Collins, S. Carpendale, and G. Penn. DocuBurst: Visualizing document content using language structure. In *Proceedings of the Eurographics / IEEE VGTC Symposium on Visualization (EuroVis)*, pages 1039–1046, 2009. http://dx.doi.org/10.1111/j.1467-8659.2009.01439.x. → pages 160

[65] C. Conati, G. Carenini, E. Hoque, B. Steichen, and D. Toker. Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. *Computer Graphics Forum (Proceedings of EuroVis)*, 33(3):371–380, 2014. http://dx.doi.org/10.1111/cgf.12393. → pages 153

[66] M. Correll, D. Albers, M. Gleicher, and S. Franconeri. Comparing averages in time series data. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1095–1104, 2012. http://dx.doi.org/10.1145/2207676.2208556. → pages 121

[67] J. A. Cottam, A. Lumsdaine, and C. Weaver. Watch this: A taxonomy for dynamic data visualization. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*

(VAST), pages 193–202, 2012.
http://dx.doi.org/10.1109/VAST.2012.6400552. → pages 241, 243

[68] M. Crotty. *The Foundations of Social Research*. Sage Publications, 1998. → pages 337

[69] R. J. Crouser and R. Chang. An affordance-based framework for human computation and human-computer collaboration. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VAST)*, 18(12):2859–2868, 2012.
http://dx.doi.org/10.1109/TVCG.2012.195. → pages 241, 244

[70] W. Cui, S. Liu, L. Tan, S. C, Y. Song, Z. J. Gao, X. Tong, and H. Qu. Textflow: Towards better understanding of evolving topics in text. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 17(12):2412–2421, 2011.
http://dx.doi.org/10.1109/TVCG.2011.239. → pages 76, 92, 93

[71] A. C. Culhane, G. Perriere, and D. G. Higgins. Cross-platform comparison and visualization of gene expression data using co-inertia analysis. *BMC Bioinformatics*, 4(59), 2003.
http://dx.doi.org/10.1186/1471-2105-4-59. → pages 286

[72] A. C. Culhane, J. Thioulouse, G. Perriere, and D. G. Higgins. MADE4: An R package for multivariate analysis of gene expression data. *Bioinformatics*, 21(11):2789–2790, 2005.
http://dx.doi.org/10.1093/bioinformatics/bti394. → pages 286

[73] P. Cunningham. Dimension reduction. In *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*, pages 91–112. Springer, 2008.
http://dx.doi.org/10.1007/978-3-540-75171-7_4. → pages 52, 298, 302, 303, 327

[74] J. Q. Dawson, T. Munzner, and J. McGrenere. A search-set model of path tracing in graphs. *Information Visualization*, 14(4):308–338, 2015. http://dx.doi.org/10.1177/1473871614550536. → pages 154

[75] A. Dix and G. Ellis. Starting simple: Adding value to static visualization through simple interaction. In *Proceedings of the ACM*

Conference on Advanced Visual Interfaces (AVI), pages 124–134, 1998. http://dx.doi.org/10.1145/948496.948514. → pages 21, 26, 39, 40, 41, 42, 238, 272

[76] A. Dix, J. Finlay, G. D. Abowd, and R. Beale. *Human-Computer Interaction*. Pearson Education, 3rd edition, 2004. → pages 374

[77] M. Dörk, S. Carpendale, and C. Williamson. The information flaneur: A fresh look at information seeking. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1215–1224, 2011. http://dx.doi.org/10.1145/1978942.1979124. → pages 29, 39, 46, 279

[78] M. Dörk, N. Henry Riche, G. Ramos, and S. Dumais. PivotPaths: Strolling through faceted information spaces. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 18 (12):2709–2718, 2012. http://dx.doi.org/10.1109/TVCG.2012.252. → pages 29, 39, 41, 279

[79] W. Dou, D. H. Jeong, F. Stukes, W. Ribarsky, H. R. Lipford, and R. Chang. Recovering reasoning processes from user interactions. *IEEE Computer Graphics & Applications (CG&A)*, 29(3):52–61, 2009. http://dx.doi.org/10.1109/MCG.2009.49. → pages 240

[80] W. Dou, L. Yu, X. Wang, Z. Ma, and W. Ribarsky. HierarchicalTopics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VAST)*, 19(12):2002–2011, 2013. http://dx.doi.org/10.1109/TVCG.2013.162. → pages 76, 78, 81, 93, 98

[81] T. Dukes. DHHS downplayed food stamp issues. *WRAL.com*, Dec. 9, 2013. http://goo.gl/otfiu8. → pages 159

[82] T. Dukes. Human-assisted reporting gets the story, Feb. 27, 2014. https://goo.gl/Jlkise. → pages 159

[83] R. Eftimie. *Modeling Group Formation and Activity Patterns in Self-Organizing Communities of Organisms*. PhD dissertation, University of Alberta, 2008. . → pages 286

[84] R. Eftimie. Hyperbolic and kinetic models for self-organized biological aggregations and movement: A brief review. *Mathematical Biology*, 2011. `http://dx.doi.org/10.1007/s00285-011-0452-2`. → pages 286

[85] R. Eftimie, G. de Vries, M. A. Lewis, and F. Lutscher. Modeling group formation and activity patterns in self-organizing collectives of individuals. *Bulletin of Mathematical Biology*, 69(5):1537–1565, 2007. `http://dx.doi.org/10.1007/s11538-006-9175-8`. → pages 286

[86] J. Eisenstein, D. H. P. Chau, A. Kittur, and E. P. Xing. TopicViz: Interactive topic exploration in document collections. In *Extended Abstract Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 2177–2182, 2012. `http://dx.doi.org/10.1145/2212776.2223772`. → pages 77

[87] K. Ellegård and J. Palm. Visualizing energy consumption activities as a tool for making everyday life more sustainable. *Applied Energy*, 88(5):1920–1926, 2011. `http://dx.doi.org/10.1016/j.apenergy.2010.11.019`. → pages 119

[88] G. Ellis and A. Dix. An explorative analysis of user evaluation studies in information visualisation. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2006. `10.1145/1168149.1168152`. → pages 350

[89] A. Endert, C. Han, D. Maiti, L. House, S. Leman, and C. North. Observation-level interaction with statistical models for visual analytics. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 121–130, 2011. `http://dx.doi.org/10.1109/VAST.2011.6102449`. → pages 301

[90] A. Endert, P. Fiaux, and C. North. Semantic interaction for visual text analytics. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 473–482, 2012. `http://dx.doi.org/10.1145/2207676.2207741`. → pages 301

[91] A. Endert, P. Fiaux, and C. North. Semantic interaction for sensemaking: Inferring analytical reasoning for model steering. *IEEE Transactions on Visualization and Computer Graphics (Proceedings*

of VAST), 18(12):2879–2888, 2012.
http://dx.doi.org/10.1109/TVCG.2012.260. → pages 77, 78, 81, 93, 97

[92] T. Erickson, M. Li, Y. Kim, A. Deshpande, S. Sahu, T. Chao, P. Sukaviriya, and M. Naphade. The Dubuque electricity portal: Evaluation of a city-scale residential electricity consumption feedback system. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1203–1212, 2013.
http://dx.doi.org/10.1145/2470654.2466155. → pages 119, 370

[93] K. A. Ericsson and H. A. Simon. Verbal reports as data. *Psychological Review*, 87(3):215–251, 1980.
http://psycnet.apa.org/doi/10.1037/0033-295X.87.3.215. → pages 85

[94] R. Etemadpour, L. Linsen, C. Crick, and A. Forbes. A user-centric taxonomy for multidimensional data projection tasks. In *Proceedings of the International Conference on Information Visualization Theory and Applications (IVAPP)*, pages 51–62, 2015.
http://goo.gl/hIE9nk. → pages 10, 154, 157, 158

[95] A. Fagan, A. C. Culhane, and D. G. Higgins. A multivariate analysis approach to the integration of proteomic and gene expression data. *Proteomics*, 7(13):2162–2171, 2007.
http://dx.doi.org/10.1002/pmic.200600898. → pages 286

[96] R. C. Fetecau and R. Eftimie. An investigation of a nonlocal hyperbolic model for self-organization of biological groups. *Mathematical Biology*, 61(4):545–579, 2010.
http://dx.doi.org/10.1007/s00285-009-0311-6. → pages 286

[97] S. Few. Distribution displays, conventional and potential. *Visual Business Intelligence Newsletter*, 2014. http://goo.gl/zy8uAo. → pages 125

[98] K. Field. When is a heat map not a heat map, Feb. 17 2015. Blog post. http://goo.gl/WnOPXo. → pages 135, 136

[99] A. Figueroa. A losing battle: How the army denies veterans justice without anyone knowing. *Fusion Media Network*, 2014.
http://goo.gl/96Kiky. → pages 159

[100] A. Fontana and J. Frey. Interviewing: The art of science. In *The Handbook of Qualitative Research*, pages 361–376. Sage Publications, 3rd edition, 1994. → pages 339, 357

[101] S. France and J. Carroll. Two-way multidimensional scaling: A review. *IEEE Transactions on Systems, Man, and Cybernetics*, 41 (5):644 –661, 2011. http://dx.doi.org/10.1109/TSMCC.2010.2078502. → pages 52, 298, 302, 303

[102] S. N. Friel, F. R. Curcio, and G. W. Bright. Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Research in Mathematics Education*, 32(2):124–158, 2001. http://dx.doi.org/10.2307/749671. → pages 29, 39, 41, 42, 217, 242, 266

[103] M. Friendly. *Visualizing Categorical Data*. SAS Press, 2000. → pages 327

[104] J. Froehlich, L. Findlater, and J. Landay. The design of eco-feedback technology. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1999–2008, 2010. http://dx.doi.org/10.1145/1753326.1753629. → pages 370

[105] J. Fuchs, F. Fischer, F. Mansmann, E. Bertini, and P. Isenberg. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 3237–3246, 2013. http://dx.doi.org/10.1145/2470654.2466443. → pages 121

[106] J. Fulda, M. Brehmer, and T. Munzner. TimeLineCurator: Interactive authoring of visual timelines from unstructured text. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VAST 2015)*, 22(1):300–309, 2016. http://dx.doi.org/10.1109/TVCG.2015.2467531. → pages 17, 153, 160

[107] D. Furniss, A. Blandford, and P. Curzon. Confessions from a grounded theory PhD: Experiences and lessons learnt. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 113–122, 2011. http://dx.doi.org/10.1145/1978942.1978960. → pages 288

[108] M. J. Gabanyi, P. D. Adams, K. Arnold, L. Bordoli, L. G. Carter, J. Flippen-Andersen, L. Gifford, J. Haas, A. Kouranov, W. A. McLaughlin, D. I. Micallef, W. Minor, R. Shah, T. Schwede, Y.-P. Tao, J. D. Westbrook, M. Zimmerman, and H. M. Berman. The structural biology knowledgebase: A portal to protein structures, sequences, functions, and methods. *Structural and Functional Genomics*, 12(2):45–54, 2011. `http://dx.doi.org/10.1007/s10969-011-9106-2`. → pages 286

[109] S. Gerber, T. Tasdizen, and R. Whitaker. Dimensionality reduction and principal surfaces via kernel map manifolds. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 529–536, 2009. `http://dx.doi.org/10.1109/ICCV.2009.5459193`. → pages 325

[110] J. Gillum. Ryan asked for federal help as he championed cuts. *Associated Press*, Oct. 12, 2012. `http://goo.gl/RM7uk`. → pages 87, 88, 92, 159

[111] M. Gleicher. Stop the evaluation arms race! A call to evaluate visualization evaluation. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2012. → pages 241, 247, 275

[112] S. R. Gomez, H. Guo, C. Ziemkiewicz, and D. H. Laidlaw. An insight- and task-based methodology for evaluating spatiotemporal visual analytics. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 63–72, 2014. `http://dx.doi.org/10.1109/VAST.2014.7042482`. → pages 153

[113] V. M. Gonzalez and A. Kobsa. A workplace study of the adoption of information visualization systems. In *Proceedings of the International Conference on Knowledge Management*, pages 92–102, 2003. `http://goo.gl/WZ7GHr`. → pages 102, 350

[114] S. Goodwin, J. Dykes, S. Jones, I. Dillingham, G. Dove, A. Duffy, A. Kachkaev, A. Slingsby, and J. Wood. Creative user-centered visualization design for energy analysts and modelers. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 19(12):2516–2525, 2013. `http://dx.doi.org/10.1109/TVCG.2013.145`. → pages 109, 119, 120, 127, 137, 370

[115] C. Görg, Z. Liu, J. Kihm, J. Choo, H. Park, and J. T. Stasko. Combining computational analyses and interactive visualization for document exploration and sensemaking in Jigsaw. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 19 (10):1646–1663, 2013. http://dx.doi.org/10.1109/TVCG.2012.324. → pages 77, 79

[116] C. Görg, Z. Liu, and J. T. Stasko. Reflections on the evolution of the Jigsaw visual analytics system. *Information Visualization*, 13(4), 2014. http://dx.doi.org/10.1177/1473871613495674. → pages 79, 93

[117] D. Gotz and M. X. Zhou. Characterizing users' visual analytic activity for insight provenance. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 123–130, 2008. http://dx.doi.org/10.1057/ivs.2008.31. → pages 21, 24, 26, 32, 39, 40, 42, 234, 247, 248, 302

[118] M. Granitzer, W. Kienreich, V. Sabol, K. Andrews, and W. Klieber. Evaluating a system for interactive exploration of large, hierarchically structured document repositories. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 127–134, 2004. http://dx.doi.org/10.1109/INFVIS.2004.19. → pages 77, 97

[119] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. LineUp: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 19 (12):2277–2286, 2013. http://dx.doi.org/10.1109/TVCG.2013.173. → pages 14, 126

[120] M. Greenacre and J. Blasius. *Multiple correspondence analysis and related methods*. Chapman & Hall/CRC, 2006. → pages 327

[121] J. Grimmer and G. King. General purpose computer-assisted clustering and conceptualization. *Proceedings of the National Academy of Sciences (PNAS)*, 108(7):2643–2650, 2011. http://dx.doi.org/10.1073/pnas.1018067108. → pages 81, 98

[122] J. Grosjean, C. Plaisant, and B. B. Bederson. SpaceTree: supporting exploration in large node link tree, design evolution and empirical evaluation. In *Proceedings of the IEEE Symposium on Information

Visualization (InfoVis), pages 57–64, 2002.
http://dx.doi.org/10.1109/INFVIS.2002.1173148. → pages 25, 32, 36, 271

[123] M. R. Grossman and G. V. Cormack. Technology-assisted review in e-discovery can be more effective and more efficient than exhaustive manual review. *Richmond Journal of Law and Technology*, XVII(3): 1–33, 2011. http://goo.gl/I8cIxQ. → pages 159

[124] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Machine Learning Research*, 3:1157–1182, 2003.
http://psycnet.apa.org/doi/10.1162/153244303322753616. → pages 52, 298, 303

[125] B. Hammer, S. Ingram, S. Kaski, E. Parviainen, S. Peltonen, J. Yang, and L. Zhang. Emerging tasks at the crossing of machine learning and information visualisation. *Dagstuhl Reports: Bridging Information Visualization with Machine Learning*, 5, 2015.
http://dx.doi.org/10.4230/DagRep.5.3.1. → pages 157

[126] M. Hao, U. Dayal, D. A. Keim, and T. Schreck. Multi-resolution techniques for visual exploration of large time-series data. In *Proceedings of the Eurographics / IEEE VGTC Symposium on Visualization (EuroVis)*, pages 27–34, 2007.
http://dx.doi.org/10.2312/VisSym/EuroVis07/027-034. → pages 120, 127

[127] S. Havre, E. Hetzler, P. Whitney, and L. Nowell. ThemeRiver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 8(1):9–20, 2002.
http://dx.doi.org/10.1109/2945.981848. → pages 78

[128] M. A. Hearst and D. Rosner. Tag clouds: Data analysis tool or social signaller? In *Proceedings of the IEEE Hawaii International Conference on System Sciences (HICSS)*, 2008.
http://dx.doi.org/10.1109/HICSS.2008.422. → pages 77

[129] J. Heer and G. G. Robertson. Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 13(6):1240–1247, 2007.
http://dx.doi.org/10.1109/TVCG.2007.70539. → pages 33

[130] J. Heer and B. Shneiderman. Interactive dynamics for visual analysis: A taxonomy of tools that support the fluent and flexible use of visualizations. *Communications of the ACM*, pages 1–26, 2012. `http://dx.doi.org/10.1145/2133416.2146416`. → pages 21, 23, 24, 26, 32, 34, 39, 41, 42, 44, 228, 247, 248

[131] J. Heer, J. D. Mackinlay, C. Stolte, and M. Agrawala. Graphical histories for visualization: supporting analysis, communication, and evaluation. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 14(6):1189–1196, 2008. `http://dx.doi.org/10.1109/TVCG.2008.137`. → pages 30, 34

[132] J. Heer, N. Kong, and M. Agrawala. Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1303–1312, 2009. `http://dx.doi.org/10.1145/1518701.1518897`. → pages 120

[133] N. Henry and J.-D. Fekete. Evaluating visual table data understanding. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2006. `http://dx.doi.org/10.1145/1168149.1168154`. → pages 10, 53, 67

[134] A. Hermida. Personal communication, 2016. → pages 160

[135] E. Hetzler and A. Turner. Analysis experiences using information visualization. *IEEE Computer Graphics & Applications (CG&A)*, 24 (5):22–26, 2004. `http://dx.doi.org/10.1109/MCG.2004.22`. → pages 77

[136] D. G. Higgins. Sequence ordinations: a multivariate analysis approach to analysing large sequence data sets. *Computer Applications in the Biosciences (CABIOS)*, 8(1):15–22, 1992. `http://dx.doi.org/10.1093/bioinformatics/8.1.15`. → pages 286

[137] J. Hollan, E. Hutchins, and D. Kirsh. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(2): 174–196, 2000. `http://dx.doi.org/10.1145/353485.353487`. → pages 50, 241, 254, 255, 305

[138] C. Holt and M. Bradford. Evaluating benchmarks of population status for Pacific Salmon. *North American Journal of Fisheries Management*, 31(2):363–378, 2011. http://dx.doi.org/10.1080/02755947.2011.578525. → pages 286

[139] K. Holtzblatt and S. Jones. Contextual inquiry: A participatory technique for system design. In D. Schuler and A. Namioka, editors, *Participatory Design: Principles and Practices*, chapter 9, pages 177–210. Lawrence Erlbaum Associates, 1993. → pages 55, 374

[140] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28 (3/4):321–377, 1936. http://dx.doi.org/10.2307/2333955. → pages 326

[141] A. Howard. Profiling data journalists. *O'Reilly Strata: Making Data Work*, Mar. 2012. http://goo.gl/64Yh2. → pages 352

[142] I. Hur, S.-H. Kim, A. Samak, and J. S. Yi. A comparative study of three sorting techniques in performing cognitive tasks on a tabular representation. *International J. Human-Computer Interaction (IJHCI)*, 29(6):379–390, 2013. http://dx.doi.org/10.1080/10447318.2012.713802. → pages 126

[143] F. Hutter, T. Bartz-Beielstein, H. H. Hoos, K. Leyton-Brown, and K. P. Murphy. Sequential model-based parameter optimization: An experimental investigation of automated and interactive approaches. In T. Bartz-Beielstein, M. Chiarandini, L. Paquete, and M. Preuss, editors, *Experimental Methods for the Analysis of Optimization Algorithms*, pages 363–414. Springer-Verlag, 2010. http://dx.doi.org/10.1007/978-3-642-02538-9_15. → pages 286

[144] F. Hutter, H. Hoos, and K. Leyton-Brown. Tradeoffs in the empirical evaluation of competing algorithm designs. *Annals of Mathematics and Artificial Intelligence*, 60:65–89, 2010. http://dx.doi.org/10.1007/s10472-010-9191-0. → pages

[145] F. Hutter, H. Hoos, K. Leyton-Brown, and K. Murphy. Time-bounded sequential parameter optimization. In C. Blum and R. Battiti, editors, *Learning and Intelligent Optimization*, volume 6073 of *Lecture Notes in Computer Science*, pages 281–298. Springer, 2010. http://dx.doi.org/10.1007/978-3-642-13800-3_30. → pages 286

[146] S. Ingram. *Practical Guidance for Dimensionality Reduction: User Guidance, Costly Distances, and Document Data*. PhD dissertation, University of British Columbia, 2013. http://hdl.handle.net/2429/45175. → pages v, 81, 157

[147] S. Ingram and T. Munzner. Glint: An MDS framework for costly distance functions. In *Proceedings of SIGRAD*, pages 29–38, 2012. http://goo.gl/W3Jnnn. → pages 330

[148] S. Ingram and T. Munzner. Dimensionality reduction for documents with nearest neighbor queries. *Neurocomputing*, 150:557–569, 2015. http://dx.doi.org/10.1016/j.neucom.2014.07.073. → pages 157

[149] S. Ingram, T. Munzner, and M. Olano. Glimmer: Multilevel MDS on the GPU. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 15(2):249–261, 2009. http://dx.doi.org/10.1109/TVCG.2008.85. → pages 52, 54, 57, 297, 298, 303, 304, 318, 321, 327, 330

[150] S. Ingram, T. Munzner, V. Irvine, M. Tory, S. Bergner, and T. Möller. DimStiller: Workflows for dimensional analysis and reduction. In *Proceedings of the IEEE Symposium on Visual Analytics Software and Technologies (VAST)*, pages 3–10, 2010. http://dx.doi.org/10.1109/VAST.2010.5652392. → pages 10, 52, 54, 69, 301, 327

[151] S. Ingram, T. Munzner, and J. Stray. Hierarchical clustering and tagging of mostly disconnected data. Technical report, Dept. Computer Science, University of British Columbia, 2012. http://cs.ubc.ca/cgi-bin/tr/2012/TR-2012-01. → pages v, 81, 330, 353, 355

[152] P. Isenberg, A. Tang, and S. Carpendale. An exploratory study of visual information analysis. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1217–1226, 2008. http://dx.doi.org/10.1145/1357054.1357245. → pages 40, 216

[153] P. Isenberg, T. Zuk, C. Collins, and S. Carpendale. Grounded evaluation of information visualizations. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2008.

http://dx.doi.org/10.1145/1377966.1377974. → pages 12, 48, 54, 102, 288, 300, 305, 338, 351, 356

[154] D. Jäckle, B. C. Kwon, and D. A. Keim. Off-screen visualization perspectives: Tasks and challenges. In *Proceedings of the Symposium on Visualization in Data Science (VDS)*, 2015. http://goo.gl/m3mqwL. → pages 155, 156

[155] A. Jain, R. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000. http://dx.doi.org/10.1109/34.824819. → pages 52, 57, 297, 298, 302, 303, 304, 319

[156] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31:651–666, 2010. http://dx.doi.org/10.1016/j.patrec.2009.09.011. → pages 330

[157] T. J. Jankun-Kelly, K. L. Ma, and M. Gertz. A model and framework for visualization exploration. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 13(2):357–69, 2007. http://dx.doi.org/10.1109/TVCG.2007.28. → pages 234

[158] W. Javed, B. McDonnel, and N. Elmqvist. Graphical perception of multiple time series. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 16(6):927–934, 2010. http://dx.doi.org/10.1109/TVCG.2010.162. → pages 121, 123, 124, 126

[159] S. Johansson and J. Johansson. Interactive dimensionality reduction through user-defined combinations of quality metrics. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 15(6):993–1000, 2009. http://dx.doi.org/10.1109/TVCG.2009.153. → pages 52, 54, 70, 289, 298, 304

[160] P. Joia, F. V. Paulovich, D. Coimbra, J. A. Cuminato, and L. G. Nonato. Local affine multidimensional projection. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 17(12):2563–2571, 2011. http://dx.doi.org/10.1109/TVCG.2011.220. → pages 301, 304

[161] I. T. Jolliffe. *Principal Component Analysis.* Springer, 2nd edition, 2002. → pages 57, 59, 297, 303, 304, 315

[162] S. Kandel, J. Heer, C. Plaisant, J. Kennedy, F. van Ham, N. Henry Riche, C. Weaver, B. Lee, D. Brodbeck, and P. Buono. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, 10(4):271–288, 2011. `http://dx.doi.org/10.1177/1473871611415994`. → pages 100

[163] S. Kandel, A. Paepcke, J. M. Hellerstein, and J. Heer. Enterprise data analysis and visualization: An interview study. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18 (12):2917–2926, 2012. `http://dx.doi.org/10.1109/TVCG.2012.219`. → pages 9, 39, 40, 41, 47, 55, 67, 225, 298, 300, 302, 305

[164] Y.-A. Kang and J. T. Stasko. Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 19–28, 2011. `http://dx.doi.org/10.1109/VAST.2011.6102438`. → pages 47, 55, 67, 223, 298, 300, 302, 305, 338, 351, 356

[165] Y.-A. Kang and J. T. Stasko. Examining the use of a visual analytics system for sensemaking tasks: Case studies with domain experts. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VAST)*, 18(12):2869–2878, 2012. `http://dx.doi.org/10.1109/TVCG.2012.224`. → pages 77, 78, 79, 101, 159

[166] D. A. Keim. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 8 (1):1–8, 2002. `http://dx.doi.org/10.1109/2945.981847`. → pages 21, 24, 26, 39, 40, 42, 238, 273

[167] M. Keller. Own a gun? Tell us why. *The Daily Beast*, Dec. 22, 2012. `http://goo.gl/2yRgk`. → pages 88, 89, 92, 159

[168] M. Keller. The other month, we asked readers why they did or did not own guns. *NewsBeast Labs*, Feb. 12, 2013. `http://goo.gl/jKBdeH`. → pages 85

[169] N. Kerracher, J. Kennedy, and K. Chalmers. A task txonomy for temporal graph visualisation. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 21(10):1160–1172, 2015. http://dx.doi.org/10.1109/TVCG.2015.2424889. → pages 153, 154

[170] R. Kincaid, A. Ben-Dor, and Z. Yakhini. Exploratory visualization of array-based comparative genomic hybridization. *Information Visualization*, 4(3):176–190, 2005. http://dx.doi.org/10.1057/palgrave.ivs.9500095. → pages 101

[171] G. Kindlmann and C. Scheidegger. An algebraic process for visualization design. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 20(12):2181–2190, 2014. http://dx.doi.org/10.1109/TVCG.2014.2346325. → pages 153, 154, 272

[172] D. Kirsh. Distributed cognition: A methodological note. *Pragmatics & Cognition*, 14(2):249–262, 2006. http://dx.doi.org/10.1075/pc.14.2.06kir. → pages 254

[173] D. Kirsh and P. Maglio. On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18(4):513–549, 1994. http://dx.doi.org/10.1207/s15516709cog1804_1. → pages 45, 241, 254, 255, 263, 264

[174] D. Klahr and H. A. Simon. Studies of scientific creativity: Complimentary approaches and convergent findings. *Psychological Bulletin*, 125(5):524–543, 1999. http://psycnet.apa.org/doi/10.1037/0033-2909.125.5.524. → pages 221, 222

[175] G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems*, 21(5):88–92, 2006. http://dx.doi.org/10.1109/MIS.2006.100. → pages 21, 22, 24, 26, 39, 40, 41, 42, 46, 241, 261, 273

[176] G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems*, 21(4):70–73, 2006. http://dx.doi.org/10.1109/MIS.2006.75. → pages 261

[177] R. Kosara and J. D. Mackinlay. Storytelling: The next step for visualization. *IEEE Computer*, 46(5):44–50, 2013.

http://doi.ieeecomputersociety.org/10.1109/MC.2013.36. → pages 29, 265

[178] T. S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, 4th edition, 1962. → pages 255, 256

[179] A. LaFrance. The brilliance of Louis C.K.'s emails: He writes like a politician. *The Atlantic*, Jul. 16, 2014. http://goo.gl/VJeZDu. → pages 159

[180] H. Lam. A framework of interaction costs in information visualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 14(6):1149–1156, 2008. http://dx.doi.org/10.1109/TVCG.2008.109. → pages 45, 233, 259

[181] H. Lam and T. Munzner. A guide to visual multi-level interface design from synthesis of empirical study evidence. *Synthesis Lectures on Visualization*, 1(1):1–117, 2010. http://dx.doi.org/10.2200/S00308ED1V01Y201011VIS001. → pages 99, 162

[182] H. Lam, T. Munzner, and R. Kincaid. Overview use in multiple visual information resolution interfaces. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 13 (6):1278–1285, 2007. http://dx.doi.org/10.1109/TVCG.2007.70583. → pages 121

[183] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18 (9):1520–1536, 2012. http://dx.doi.org/10.1109/TVCG.2011.279. → pages 2, 23, 25, 50, 67, 101, 214, 305, 350, 372, 375

[184] T. Lammarsch, W. Aigner, A. Bertone, J. Gärtner, E. Mayr, S. Miksch, and M. Smuc. Hierarchical temporal patterns and interactive aggregated views for pixel-based visualizations. In *Proceedings of the IEEE International Conference on Information Visualization*, pages 44–50, 2009. http://dx.doi.org/10.1109/IV.2009.52. → pages 128

[185] T. Lammarsch, A. Rind, W. Aigner, and S. Miksch. Developing an extended task framework for exploratory data analysis along the structure of time. In *Proceedings of the Eurographics International*

Workshop on Visual Analytics (EuroVA), 2012.
http://dx.doi.org/10.2312/PE/EuroVAST/EuroVA12/031-035. → pages 10, 35, 43, 53, 278

[186] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry. Task taxonomy for graph visualization. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2006.
http://dx.doi.org/10.1145/1168149.1168168. → pages 9, 21, 23, 24, 26, 31, 35, 39, 40, 41, 42, 43, 53, 67, 95, 228, 233

[187] B. Lee, P. Isenberg, N. Henry Riche, and S. Carpendale. Beyond mouse and keyboard: Expanding design considerations for information visualization interactions. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 18 (12):2689–2698, 2012.
http://dx.doi.org/10.1109/TVCG.2012.204. → pages 40, 241, 262

[188] J. M. Lewis, L. van der Maaten, and V. R. de Sa. A behavioral investigation of dimensionality reduction. In *Proceedings of the Cognitive Science Society*, pages 671–676, 2012.
https://goo.gl/0wP4hH. → pages 54, 57, 69, 302

[189] K. Leyton-Brown, E. Nudelman, and Y. Shoham. Empirical hardness models: Methodology and a case study on combinatorial auctions. *Journal of the ACM*, 56(4):22:1–22:52, 2009.
http://dx.doi.org/10.1145/1538902.1538906. → pages 286

[190] T. Lin and H. Zha. Riemannian manifold learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5): 796–809, 2008. http://dx.doi.org/10.1109/TPAMI.2007.70735. → pages 304

[191] S. Liu, M. X. Zhou, S. Pan, Y. Song, W. Qian, W. Cai, and X. Lian. TIARA: Interactive, topic-based visual text summarization and analysis. *ACM Transactions on Intelligent Systems and Technology*, 3(2), 2012. http://dx.doi.org/10.1145/2089094.2089101. → pages 76

[192] Y. Liu, S. Barlowe, Y. Feng, J. Yang, and M. Jiang. Evaluating exploratory visualization systems: A user study on how clustering-based visualization systems support information seeking

from large document collections. *Information Visualization*, 12(1): 25–43, 2013. http://dx.doi.org/10.1177/1473871612459995. → pages 77, 78

[193] Z. Liu and J. T. Stasko. Mental models, visual reasoning and interaction in information visualization: a top-down perspective. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 16(6):999–1008, 2010. http://dx.doi.org/10.1109/TVCG.2010.177. → pages 21, 26, 32, 39, 40, 41, 42, 50, 270

[194] Z. Liu, N. Nersessian, and J. T. Stasko. Distributed cognition as a theoretical framework for information visualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 14(6):1173–1180, 2008. http://dx.doi.org/10.1109/TVCG.2008.121. → pages 45, 241, 249, 270

[195] D. Lloyd and J. Dykes. Human-centered approaches in geovisualization design: investigating multiple methods through a long-term case study. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 17(12):2498–2507, 2011. http://dx.doi.org/10.1109/TVCG.2011.209. → pages 11, 12, 13, 14, 102, 103, 108, 109, 121, 351, 374, 375

[196] A. M. MacEachren, R. E. Roth, J. O'Brien, N. Li, D. Swingley, and M. Gahegan. Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 18(12):2496–2505, 2012. http://dx.doi.org/10.1109/TVCG.2012.279. → pages 372

[197] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2): 110–141, 1986. http://dx.doi.org/10.1145/22949.22950. → pages 49

[198] S. F. Madden, S. B. Carpenter, I. B. Jeffery, H. Bjorkbacka, K. A. Fitzgerald, L. A. O'Neill, and D. G. Higgins. Detecting microRNA activity from gene expression data. *BMC Bioinformatics*, 11(257), 2010. http://dx.doi.org/10.1186/1471-2105-11-257. → pages 286

[199] G. Marchionini. Exploratory search: From finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006. http://dx.doi.org/10.1145/1121949.1121979. → pages 29, 30, 39, 41, 42, 225, 351

[200] S. Mathison. Why triangulate? *Educational Researcher*, 17(2):13, 1988. http://dx.doi.org/10.3102/0013189X017002013. → pages 340

[201] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Transactions on Graphics*, 22(3):759–769, 2003. http://dx.doi.org/10.1145/882262.882343. → pages 58, 59, 298, 308, 309, 328

[202] E. Mayr, M. Smuc, and H. Risku. Many roads lead to Rome: Mapping users' problem-solving strategies. *Information Visualization*, 10(3):232–247, 2011. http://goo.gl/x6CMbu. → pages 216, 351, 356

[203] J. E. McGrath. Methodology matters: Doing research in the behavioral and social sciences. In R. M. Baecker, J. Grudin, W. Buxton, and S. Greenberg, editors, *Readings in Human-Computer Interaction: Towards the Year 2000*, pages 152–169. Morgan Kaufmann, 2nd edition, 1995. → pages 53, 300, 305

[204] S. McKenna, D. Mazur, J. Agutter, and M. Meyer. Design activity framework for visualization design. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 20 (12):2191–2200, 2014. http://dx.doi.org/10.1109/TVCG.2014.2346331. → pages 109, 137

[205] M. McKeon. Harnessing the web information ecosystem with wiki-based visualization dashboards. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 15 (6):1081–1088, 2009. http://dx.doi.org/10.1109/TVCG.2009.148. → pages 101, 159

[206] P. McLachlan, T. Munzner, E. Koutsofios, and S. North. LiveRAC: interactive visual exploration of system management time-series data. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1483–1492, 2008. http://dx.doi.org/10.1145/1357054.1357286. → pages 121, 136

[207] L. A. McNamara and N. Orlando-Gay. Reading, sorting, marking, shuffling: Mental model formation through information foraging. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2012. → pages 241, 246

[208] L. A. McNamara, K. Cole, and S. Stevens-Adams. "Where Do I Start?" Practical Methods for Design Studies in InfoVis and Visual Analysis, 2014. IEEE VIS Tutorial. `http://goo.gl/7wziik`. → pages 137, 156

[209] R. Metoyer, B. Lee, N. Henry Riche, and M. Czerwinski. Understanding the verbal language and structure of end-user descriptions of data visualizations. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1659–1662, 2012. `http://dx.doi.org/10.1145/2207676.2208292`. → pages 136

[210] M. Meyer, M. Sedlmair, and T. Munzner. The four-level nested model revisited: Blocks and guidelines. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, pages 1–6, 2012. `http://doi.acm.org/10.1145/2442576.2442587`. → pages 271

[211] M. Meyer, M. Sedlmair, P. S. Quinan, and T. Munzner. The nested blocks and guidelines model. *Information Visualization*, 14(3): 234–249, 2015. `http://dx.doi.org/10.1177/1473871613510429`. → pages 32, 47, 53, 67, 123, 154, 271, 300, 372

[212] B. Mirel and C. Görg. Scientists' sense making when hypothesizing about disease mechanisms from expression data and their needs for visualization support. *BMC Bioinformatics*, 15(1):117–128, 2014. `http://dx.doi.org/10.1186/1471-2105-15-117`. → pages 154

[213] S. Mittelstädt, D. Jäckle, F. Stoffel, and D. A. Keim. ColorCAT: Guided design of colormaps for combined analysis tasks. In *Short Paper Proceedings of the Eurographics Conference on Visualization (EuroVis)*, pages 115–119, 2015. `http://dx.doi.org/10.2312/eurovisshort.20151135`. → pages 153, 155

[214] E. Morse, M. Lewis, and K. A. Olsen. Evaluating visualizations: Using a taxonomic guide. *International J. Human-Computer Studies*,

53(5):637–662, 2000.
http://dx.doi.org/10.1006/ijhc.2000.0412. → pages 68, 232

[215] T. Muhlbacher, H. Piringer, S. Gratzl, M. Sedlmair, and M. Streit. Opening the black box: Strategies for increased user involvement in existing algorithm implementations. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VAST)*, 20 (12):1643–1652, 2014.
http://dx.doi.org/10.1109/TVCG.2014.2346578. → pages 137

[216] P. M. Mullins and S. Treu. A task-based cognitive model for user-network interaction: Defining a task taxonomy to guide the interface designer. *Interacting with Computers*, 5(2):139–166, 1993.
http://dx.doi.org/10.1016/0953-5438(93)90016-M. → pages 21, 24, 32, 39, 41, 42, 43, 53, 235

[217] T. Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 15(6):921–928, 2009.
http://dx.doi.org/10.1109/TVCG.2009.111. → pages 5, 6, 7, 21, 32, 67, 68, 123, 255, 265, 269, 271, 274, 275, 298, 300, 306, 308, 371

[218] T. Munzner. Visualization. In P. Shirley and S. Marschner, editors, *Fundamentals of Graphics*, chapter 27, pages 675–707. AK Peters, 3rd edition, 2009. → pages 39, 42

[219] T. Munzner. *Visualization Analysis and Design*. AK Peters Visualization Series, CRC Press, 2014. → pages iv, vi, xv, xvii, xxvi, 5, 6, 16, 18, 32, 35, 44, 47, 66, 113, 123, 140, 142, 143, 144, 145, 146, 147, 151, 156, 228, 271, 283

[220] T. Munzner, F. Guimbretière, S. Tasiran, L. Zhang, and Y. Zhou. TreeJuxtaposer: scalable tree comparison using focus + context with guaranteed visibility. *ACM Transactions on Graphics*, 22(3): 462–471, 2003. http://dx.doi.org/10.1145/882262.882291. → pages 25, 32, 36, 271

[221] T. Munzner, A. Barsky, and M. Williams. Reflections on QuestVis: A visualization system for an environmental sustainability model. In H. Hagen, editor, *Scientific Visualization: Interactions, Features, Metaphors. Dagstuhl Follow-Ups*, volume 2, chapter 17, pages 240–259. Dagstuhl Publishing, 2011.

http://dx.doi.org/10.4230/DFU.Vol2.SciViz.2011.240. → pages 350

[222] K. P. Murphy. *Machine Learning: A Probabilistic Perspective.* MIT Press, 2012. → pages 299, 310, 323, 330

[223] C. B. Nielsen, S. D. Jackman, I. Birol, and S. J. M. Jones. ABySS-Explorer: Visualizing genome sequence assemblies. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 15(6):881–888, 2009. http://dx.doi.org/10.1109/TVCG.2009.116. → pages 286

[224] J. Nielsen. Why you only need to test with 5 users, March 19, 2000. http://goo.gl/6Huppn. → pages 91

[225] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 249–256, 1990. http://dx.doi.org/10.1145/97243.97281. → pages 375

[226] D. A. Norman. *The Psychology of Everyday Things.* Basic Books, 1988. → pages 44, 45, 233, 238, 249, 251, 258, 263, 264, 266

[227] D. A. Norman. Human-centered design considered harmful. *ACM interactions*, 12(4):14, 2005. http://dx.doi.org/10.1145/1070960.1070976. → pages 64, 68

[228] C. North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications (CG&A)*, 26(3):6–9, 2006. http://dx.doi.org/10.1109/MCG.2006.70. → pages 220, 356

[229] C. North, P. Saraiya, and K. Duca. A comparison of benchmark task and insight evaluation methods for information visualization. *Information Visualization*, 10(3):162–181, 2011. http://dx.doi.org/10.1177/1473871611415989. → pages 49, 224, 375

[230] S. Nusrat and S. G. Kobourov. Task taxonomy for cartograms. In *Short Paper Proceedings of the Eurographics Conference on Visualization (EuroVis)*, pages 61–65, 2015. http://dx.doi.org/10.2312/eurovisshort.20151126. → pages 153, 154

[231] P. Österling, G. Scheuermann, S. Teresniak, G. Heyer, S. Koch, T. Ertl, and G. H. Weber. Two-stage framework for a topology-based projection and visualization of classified document collections. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 91–98, 2010. http://dx.doi.org/10.1109/VAST.2010.5652940. → pages 77

[232] F. V. Paulovich, M. C. F. Oliveira, and R. Minghim. The projection explorer: A flexible tool for projection-based multidimensional visualization. In *Proceedings of the IEEE Brazilian Symposium on Computer Graphics and Image Processing*, pages 27–36, 2007. http://dx.doi.org/10.1109/SIBGRAPI.2007.21. → pages 77, 97

[233] F. V. Paulovich, L. G. Nonato, R. Minghim, and H. Levkowitz. Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. *IEEE Trans. on Visualization and Computer Graphics (TVCG)*, 14(3):564–575, 2008. http://dx.doi.org/10.1109/TVCG.2007.70443. → pages 304

[234] F. V. Paulovich, C. T. Silva, and L. G. Nonato. Two-phase mapping for projecting massive data sets. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of SciVis)*, 16(6):1281–1290, 2010. http://dx.doi.org/10.1109/TVCG.2010.207. → pages 304

[235] F. V. Paulovich, D. M. Eler, J. Poco, C. P. Botha, R. Minghim, and L. G. Nonato. Piece wise Laplacian-based projection for interactive data exploration and organization. *Computer Graphics Forum (Proceedings of EuroVis)*, 30(3):1091–1100, 2011. http://dx.doi.org/10.1111/j.1467-8659.2011.01958.x. → pages 301

[236] S. Peddie and A. Playford. For their eyes only: Police misconduct hidden from public by secrecy law, weak oversight. *Newsday*, Dec. 28, 2013. http://goo.gl/KBT7CA. → pages 11, 90, 91, 92, 159

[237] A. Perer and B. Shneiderman. Integrating statistics and visualization for exploratory power: from long-term case studies to design guidelines. *IEEE Computer Graphics and Applications (CG&A)*, 29 (3):39–51, 2009. http://doi.ieeecomputersociety.org/10.1109/MCG.2009.44. → pages 49, 70, 217, 232

[238] D. Pham, S. Dimov, and C. Nguyen. Selection of k in k-means clustering. *Mechanical Engineering Science (Proceedings of the IMechE)*, 219:103–119, 2005. http://dx.doi.org/10.1243/095440605X8298. → pages 96

[239] W. A. Pike, J. T. Stasko, R. Chang, and T. A. O'Connell. The science of interaction. *Information Visualization*, 8(4):263–274, 2009. http://dx.doi.org/10.1057/ivs.2009.22. → pages 21, 23, 24, 29, 38, 39, 41, 42, 43, 222, 260

[240] H. Piringer, W. Berger, and J. Krasser. HyperMoVal: Interactive visual validation of regression models for real-time simulation. *Computer Graphics Forum (Proceedings of EuroVis)*, 29(3):983–992, 2010. http://dx.doi.org/10.1111/j.1467-8659.2009.01684.x. → pages 301, 314

[241] P. Pirolli. *Information Foraging Theory: Adaptive Interaction with Information*. Oxford University Press, 2009. → pages 218, 222, 241, 335, 340, 351, 356, 358

[242] P. Pirolli and S. K. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of the International Conference on Intelligence Analysis*, 2005. http://dx.doi.org/10.1057/ivs.2009.22. → pages 21, 22, 24, 26, 39, 40, 41, 42, 46, 150, 215, 216, 218, 223, 236, 241, 272, 273

[243] C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the ACM Conference on Advanced Visual Interfaces (AVI)*, pages 109–116, 2004. http://dx.doi.org/10.1145/989863.989880. → pages 305, 351

[244] C. Plaisant, D. Carr, and B. Shneiderman. Image-browser taxonomy and guidelines for designers. *IEEE Software*, 12(2):21–32, 1995. http://dx.doi.org/10.1109/52.368260. → pages 39, 41, 42, 238

[245] M. Pohl, S. Wiltner, and S. Miksch. Exploring information visualization: Describing different interaction patterns. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2010. http://dx.doi.org/10.1145/2110192.2110195. → pages 227, 359, 375

[246] M. Pohl, M. Smuc, and E. Mayr. The user puzzle: Explaining the interaction with visual analytics systems. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VAST)*, 18(12): 2908–2916, 2012. http://dx.doi.org/10.1109/TVCG.2012.273. → pages 47, 241, 253, 254, 267

[247] Z. Pousman, J. T. Stasko, and M. Mateas. Casual information visualization: depictions of data in everyday life. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 13(6):1145–1152, 2007. http://dx.doi.org/10.1109/TVCG.2007.70541. → pages 29, 39, 41, 46, 280

[248] A. J. Pretorius, M. A. P. Bray, and A. E. Carpenter. Visualization of parameter space for image analysis. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 17 (12):2402 –2411, 2011. http://dx.doi.org/10.1109/TVCG.2011.253. → pages 301, 314

[249] A. J. Pretorius, H. C. Purchase, and J. T. Stasko. Tasks for multivariate network analysis. In A. Kerren, H. C. Purchase, and M. O. Ward, editors, *Multivariate Network Visualization*, pages 77–95. Springer, 2014. http://dx.doi.org/10.1007/978-3-319-06793-3_5. → pages 154

[250] J. D. Prown. Mind in matter: An introduction to material culture theory and method. *Winterthur Portfolio*, 17(1):1–19, 1982. http://jstor.org/stable/1180761. → pages 341

[251] R. Rao and S. K. Card. The Table Lens: Merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 318–322, 1994. http://dx.doi.org/10.1145/191666.191776. → pages 24, 271

[252] J. Raskin. *The Humane Interface: New Directions for Designing Interactive Systems*. Addison-Wesley Longman, 2000. → pages 21, 32, 39, 41, 42, 278

[253] B. Renoust. *Analysis and Visualisation of Edge Entanglement in Multiplex Networks*. PhD dissertation, University of Bordeaux, 2013. http://goo.gl/mrP5eY. → pages 154

[254] R. A. Rensink. On the prospects for a science of visualization. In W. Huang, editor, *Handbook of Human Centric Visualization*, pages 147–175. Springer, 2014. http://dx.doi.org/10.1007/978-1-4614-7485-2_6. → pages 23, 240

[255] M. Rester, M. Pohl, S. Wiltner, K. Hinum, S. Miksch, C. Popow, and S. Ohmann. Evaluating an InfoVis technique using insight reports. In *Proceedings of the IEEE International Conference on Information Visualization*, pages 693–700, 2007. http://dx.doi.org/10.1109/IV.2007.44. → pages 358

[256] L. Reveret, L. Favreau, C. Depraz, and M.-P. Cani. Morphable model of quadrupeds skeletons for animating 3D animals. In *Proceedings of the Eurographics Symposium on on Computer Animation (SCA)*, pages 135–142, 2005. http://dx.doi.org/10.1145/1073368.1073386. → pages 58, 308, 309

[257] W. Ribarsky, B. Fisher, and W. M. Pottenger. Science of analytical reasoning. *Information Visualization*, 8(4):254–262, 2009. http://dx.doi.org/10.1057/ivs.2009.28. → pages 273

[258] A. Rind, W. Aigner, M. Wagner, S. Miksch, and T. Lammarsch. Task cube: A three-dimensional conceptual space of user tasks in visualization design and evaluation. *Information Visualization*, (in press). http://dx.doi.org/10.1177/1473871615621602. → pages 153, 154

[259] J. Rodgers and L. Bartram. Exploring ambient and artistic visualization for residential energy use feedback. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2489–2497, 2011. http://dx.doi.org/10.1109/TVCG.2011.196. → pages 119, 370

[260] R. E. Roth. An empirically-derived taxonomy of cartographic interaction primitives. In *Proceedings of the GIScience*, 2012. http://goo.gl/7HgB2N. → pages 21, 24, 26, 29, 32, 39, 41, 42, 149, 251, 259, 263, 264, 283

[261] R. E. Roth. Cartographic interaction primitives: Framework and synthesis. *Cartographic Journal*, 49(4):376 – 395, 2012. http://dx.doi.org/10.1179/1743277412Y.0000000019. → pages 39, 45, 238, 243, 245, 249, 258

[262] R. E. Roth. An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 19(12):2356–2365, 2013. http://dx.doi.org/10.1109/TVCG.2013.130. → pages 21, 24, 26, 29, 32, 39, 40, 41, 42, 44, 149, 150, 151, 152, 260, 283, 302

[263] S. F. Roth and J. Mattis. Data characterization for intelligent graphics presentation. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 193–200, 1990. http://dx.doi.org/10.1145/97243.97273. → pages 21, 24, 26, 39, 40, 41, 42, 231

[264] P. Ruchikachorn and K. Mueller. Learning visualizations by analogy: Promoting visual literacy through visualization morphing. *IEEE Transactions on on Visualization and Computer Graphics (TVCG)*, 21(9):1028–1044, 2015. http://dx.doi.org/10.1109/TVCG.2015.2413786. → pages 162

[265] G. W. Ryan and H. R. Bernard. Techniques to identify themes. *Field Methods*, 15(1):85–109, 2003. http://dx.doi.org/10.1177/1525822X02239569. → pages 341

[266] A. Saad, G. Hamarneh, T. Möller, and B. Smith. Kinetic modeling based probabilistic segmentation for molecular images. In D. Metaxas, L. Axel, G. Fichtinger, and G. Székely, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2008)*, volume 5241 of *Lecture Notes in Computer Science*, pages 244–252. Springer, 2008. http://dx.doi.org/10.1007/978-3-540-85988-8_30. → pages 286

[267] A. Saad, G. Hamarneh, and T. Möller. Exploration and visualization of segmentation uncertainty using shape and appearance prior information. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of SciVis)*, 16(6):1366–1375, 2010. http://dx.doi.org/10.1109/TVCG.2010.152. → pages 286

[268] A. Saad, T. Möller, and G. Hamarneh. ProbExplorer: Uncertainty-guided exploration and editing of probabilistic medical image segmentation. *Computer Graphics Forum (Proceedings of EuroVis)*, 29(3):1113–1122, 2010.

http://dx.doi.org/10.1111/j.1467-8659.2009.01691.x. → pages 286

[269] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VAST)*, 20(12):1604–1613, 2014. http://dx.doi.org/10.1109/TVCG.2014.2346481. → pages 153, 154

[270] Y. Saeys, I. Inza, and P. Larrañaga. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19):2507–2517, 2007. http://dx.doi.org/10.1093/bioinformatics/btm344. → pages 286

[271] B. Saket, P. Simonetto, and S. Kobourov. Group-level graph visualization taxonomy. In *Short Paper Proceedings of the Eurographics Conference on Visualization (EuroVis)*, 2014. http://dx.doi.org/10.2312/eurovisshort.20141162. → pages 153, 154

[272] B. Saket, P. Simonetto, S. Kobourov, and K. Börner. Node, node-link, and node-link-group diagrams: An evaluation. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 20(12):2231–2240, 2014. http://dx.doi.org/10.1109/TVCG.2014.2346422. → pages 9, 153, 154

[273] B. Saket, C. Scheidegger, and S. G. Kobourov. Towards understanding enjoyment and flow in information visualization. In *Short Paper Proceedings of the Eurographics Conference on Visualization (EuroVis)*, pages 109–113, 2015. http://dx.doi.org/10.2312/eurovisshort.20151134. → pages 153

[274] B. Saket, C. Scheidegger, S. G. Kobourov, and K. Börner. Map-based visualizations increase recall accuracy of data. *Computer Graphics Forum (Proceedings of EuroVis)*, 34(3):441–450, 2015. http://dx.doi.org/10.1111/cgf.12656. → pages 153, 155

[275] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):

513–523, 1988.
http://dx.doi.org/10.1016/0306-4573(88)90021-0. → pages 80

[276] P. Saraiya, C. North, and K. Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 11(4):443–456, 2005. http://dx.doi.org/10.1109/TVCG.2005.53. → pages 351, 367

[277] P. Saraiya, C. North, V. Lam, and K. Duca. An insight-based longitudinal study of visual analytics. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 12(6):1511–1522, 2006. http://dx.doi.org/10.1109/TVCG.2006.85. → pages 11, 103, 351, 355, 358

[278] R. Scheepens, C. Hurter, H. van de Wetering, and J. van Wijk. Visualization, selection, and analysis of traffic flows. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis 2015)*, 22(1):379–388, 2016. http://dx.doi.org/10.1109/TVCG.2015.2467112. → pages 153, 155

[279] H. J. Schulz. Treevis.net: A tree visualization reference. *IEEE Computer Graphics and Applications (CG&A)*, 31(6):11–15, 2011. http://dx.doi.org/10.1109/MCG.2011.103. → pages 27

[280] H. J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 19(12):2366–2375, 2013. http://dx.doi.org/10.1109/TVCG.2013.120. → pages 35, 149, 150, 151, 152, 283

[281] B. Schwartz. *The Paradox of Choice: Why More Is Less*. Ecco, 2003. → pages 98

[282] M. Sedlmair, P. Isenberg, D. Baur, and A. Butz. Information visualization evaluation in large companies: Challenges, experiences and recommendations. *Information Visualization*, 10(3):248–266, 2011. http://dx.doi.org/10.1177/1473871611413099. → pages 137, 138, 306, 350, 364, 375

[283] M. Sedlmair, M. Brehmer, S. Ingram, and T. Munzner. Dimensionality reduction in the wild: Gaps and guidance. Technical report, Dept. Computer Science, University of British Columbia,

2012. `http://cs.ubc.ca/cgi-bin/tr/2012/TR-2012-03`. → pages iv, xxiii, 56, 157, 227, 288, 289, 297, 298

[284] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 18(12):2431–2440, 2012. `http://dx.doi.org/10.1109/TVCG.2012.213`. → pages 3, 23, 28, 47, 48, 67, 68, 69, 84, 123, 137, 159, 271, 299, 308, 373, 374, 376

[285] M. Sedlmair, A. Tatu, M. Tory, and T. Munzner. A taxonomy of visual cluster separation factors. *Computer Graphics Forum (Proceedings of EuroVis)*, 31(3):1335–1344, 2012. `http://dx.doi.org/10.1111/j.1467-8659.2012.03125.x`. → pages 54, 63, 300, 302, 317, 327, 330

[286] M. Sedlmair, T. Munzner, and M. Tory. Empirical guidance on scatterplot and dimension reduction technique choices. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 19(12):2634–2643, 2013. `http://dx.doi.org/10.1109/TVCG.2013.153`. → pages 54, 63, 97, 157, 301

[287] M. Sedlmair, C. Heinzl, S. Bruckner, H. Piringer, and T. Möller. Visual parameter space analysis: A conceptual framework. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 20(12):2161–2170, 2014. `http://dx.doi.org/10.1109/TVCG.2014.2346321`. → pages 153, 155, 157

[288] F. Sha and L. K. Saul. Analysis and extension of spectral methods for nonlinear dimensionality reduction. In *Proceedings of the ACM International Conference on Machine Learning (ICML)*, pages 784–791, 2005. `http://dx.doi.org/10.1145/1102351.1102450`. → pages 304

[289] W. Shi, F. Lei, C. Zhu, F. Sievers, and D. G. Higgins. A complete analysis of HA and NA genes of influenza A viruses. *PloS ONE*, 5 (12):e14454, 2010. `http://dx.doi.org/10.1371/journal.pone.0014454`. → pages 286

[290] M. H. Shimabukuro, E. F. Flores, M. C. F. de Oliveira, and H. Levkowitz. Coordinated views to assist exploration of

spatio-temporal data: A case study. In *Proceedings of the IEEE International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV)*, pages 107–117, 2004. `http://dx.doi.org/10.1109/CMV.2004.1319531`. → pages 120

[291] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, 1996. `http://dx.doi.org/10.1109/VL.1996.545307`. → pages 9, 21, 24, 26, 32, 34, 38, 39, 40, 41, 42, 53, 150, 152, 228, 232, 234, 238, 272, 302, 310, 335, 352

[292] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2006. `http://dx.doi.org/10.1145/1168149.1168158`. → pages 11, 49, 69, 103, 215, 217, 351, 355, 375

[293] Y. B. Shrinivasan and J. J. van Wijk. Supporting the analytical reasoning process in information visualization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1237–1246, 2008. `http://dx.doi.org/10.1145/1357054.1357247`. → pages 34, 39, 41, 42, 239, 240

[294] D. Silverman. *A Very Short, Fairly Interesting and Reasonably Cheap Book about Qualitative Research*. Sage Publications, 2007. → pages 340

[295] S. Simon, S. Mittelstädt, B. C. Kwon, A. Stoffel, R. Landstorfer, K. Neuhaus, A. Mühlig, S. Scherer, and D. A. Keim. VisExpress: Visual exploration of differential gene expression data. *Information Visualization*, (in press). `http://dx.doi.org/10.1177/1473871615612883`. → pages 153, 155

[296] M. Sips, B. Neubert, J. Lewis, and P. Hanrahan. Selecting good views of high-dimensional data using class consistency. *Computer Graphics Forum (Proceedings of EuroVis)*, 28(3):831–838, 2009. `http://dx.doi.org/10.1111/j.1467-8659.2009.01467.x`. → pages 304

[297] K. B. Smith. Typologies, taxonomies, and the benefits of policy classification. *Policy Studies Journal*, 30(3):379–395, 2002. http://dx.doi.org/10.1111/j.1541-0072.2002.tb02153.x. → pages 280

[298] R. Spence. *Information Visualization: Design for Interaction*. Prentice Hall, 2nd edition, 2007. → pages 21, 24, 39, 41, 42, 266, 277

[299] D. Sprague and M. Tory. Exploring how and why people use visualizations in casual contexts: Modeling user goals and regulated motivations. *Information Visualization*, 11(2):106–123, 2012. http://dx.doi.org/10.1177/1473871611433710. → pages 26, 29, 39, 41, 46, 223, 224

[300] D. W. Sprague and M. Tory. Motivation and procrastination: Methods for evaluating pragmatic casual information visualizations. *IEEE Computer Graphics and Applications (CG&A)*, 29(4):86–91, 2009. http://dx.doi.org/10.1109/MCG.2009.70. → pages 223

[301] R. R. Springmeyer, M. M. Blattner, and N. L. Max. A characterization of the scientific data analysis process. In *Proceedings of the IEEE Conference on Visualization*, pages 235–242, 1992. http://dx.doi.org/10.1109/VISUAL.1992.235203. → pages 3, 20, 21, 23, 24, 39, 41, 42, 43, 219, 222, 226, 247, 248, 302, 351, 356, 358

[302] J. Stahnke, M. Dörk, B. Muller, and A. Thom. Probing projections: Interaction techniques for interpreting arrangements and errors of dimensionality reductions. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis 2015)*, 22(1):629–638, 2016. http://dx.doi.org/10.1109/TVCG.2015.2467717. → pages 158, 164

[303] W. Stephenson. *The Play Theory of Mass Communication*. University of Chicago Press, 1967. → pages 29, 39, 46, 256, 258, 268

[304] C. Stolte, D. Tang, and P. Hanrahan. Polaris: A system for query, analysis, and visualization of multidimensional relational database. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 51(11):52–65, 2002. http://dx.doi.org/10.1109/2945.981851. → pages 271

[305] J. Stray. A full-text visualization of the Iraq War Logs. *Associated Press*, Dec. 10, 2010. `http://goo.gl/IITFpE`. → pages vi, 79, 80, 332, 333

[306] J. Stray. What did private security contractors do in Iraq? *Associated Press*, Feb. 21, 2012. `http://goo.gl/qqGGh`. → pages 86, 92, 159, 352, 364

[307] J. Stray. Using Overview to analyze 4500 pages of documents on security contractors in Iraq, Feb. 21, 2012. `http://goo.gl/6SmNc`. → pages 85

[308] J. Stray. The document mining Pulitzers, May 1, 2014. `https://goo.gl/iUwGX6`. → pages

[309] J. Stray. Document mining shows Paul Ryan relying on the the programs he criticizes, Nov. 2 2012. `https://goo.gl/xJy2eZ`. → pages 85

[310] A. G. Sutcliffe, M. Ennis, and J. Hu. Evaluating the effectiveness of visual user interfaces for information retrieval. *International Journal of Human-Computer Interaction (IJHCI)*, 53(5):741–763, 2000. `http://dx.doi.org/10.1006/ijhc.2000.0416`. → pages 351

[311] F. Taher, J. Hardy, A. Karnik, C. Weichel, Y. Jansen, K. Hornbæk, and J. Alexander. Exploring interactions with physically dynamic bar charts. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 3237–3246, 2015. `http://dx.doi.org/10.1145/2702123.2702604`. → pages 153

[312] A. Tatu, G. Albuquerque, M. Eisemann, P. Bak, H. Theisel, M. Magnor, and D. A. Keim. Automated analytical methods to support visual exploration of high-dimensional data. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 17 (5):584–597, 2010. `http://dx.doi.org/10.1109/TVCG.2010.242`. → pages 52, 54, 304

[313] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000. `http://dx.doi.org/10.1126/science.290.5500.2319`. → pages vi, 52, 57, 58, 59, 60, 297, 298, 303, 304, 308, 309, 314, 327, 328

[314] S. Theisa, A. Mertensa, and C. M. Schlicka. Preliminary results of an exploratory study towards a general task and data model for telemedical visualizations. In *Proceedings of the Triennial Congress of the International Ergonomics Association (IEA)*, 2015. `http://goo.gl/Y2Qbjx`. → pages 154

[315] J. J. Thomas and K. A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE, 2005. → pages 47, 217, 222

[316] J. Thorp. Visualization as process, not output. *Harvard Business Review*, Apr. 3, 2013. `https://goo.gl/1xgkI6`. → pages 20

[317] E. G. Toms. Serendipitous information retrieval. In *Proceedings of the DELOS Workshop on Information Seeking, Searching and Querying in Digital Libraries*, pages 1–4, 1999. `http://goo.gl/01pwgQ`. → pages 257

[318] E. G. Toms. Understanding and facilitating the browsing of electronic text. *International J. of Human-Computer Studies (IJHCS)*, 52(3):423–452, 2000. `http://dx.doi.org/10.1006/ijhc.1999.0345`. → pages 29, 30, 39, 41, 46, 257, 270

[319] W. S. Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17:401–419, 1952. `http://dx.doi.org/10.1007/BF02288916`. → pages 52, 57, 298, 304

[320] T. Torsney-Weir, A. Saad, T. Möller, B. Weber, H. C. Hege, J. M. Verbavatz, and S. Bergner. Tuner: Principled parameter finding for image segmentation algorithms using visual response surface exploration. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of SciVis)*, 17(12):1892–1901, 2011. `http://dx.doi.org/10.1109/TVCG.2011.248`. → pages 286

[321] M. Tory and T. Möller. Rethinking visualization: A high-level taxonomy. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 151–158, 2004. `http://dx.doi.org/10.1109/INFVIS.2004.59`. → pages 39, 41, 42, 237

[322] M. Tory and S. Staub-French. Qualitative analysis of visualization: A building design field study. In *Proceedings of the ACM Workshop*

on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV), 2008. http://dx.doi.org/10.1145/1377966.1377975. → pages 54, 67, 300, 338, 351, 356

[323] M. Tory, D. Sprague, F. Wu, W. Y. So, and T. Munzner. Spatialization design: comparing points and landscapes. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 13(6):1262–1269, 2007. http://dx.doi.org/10.1109/TVCG.2007.70596. → pages 330

[324] M. Tory, C. Swindells, and R. Dreezer. Comparing dot and landscape spatializations for visual memory differences. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 15(6):1033–1039, 2009. http://dx.doi.org/10.1109/TVCG.2009.127. → pages 77

[325] J. G. Trafton, S. S. Kirschenbaum, T. L. Tsui, R. T. Miyamoto, J. A. Ballas, and P. D. Raymond. Turning pictures into numbers: Extracting and generating information from complex visualizations. *International Journal of Human Computer Studies*, 53(5):827–850, 2000. http://dx.doi.org/10.1006/ijhc.2000.0419. → pages 216, 302

[326] E. Tufte. *Envisioning Information*. Graphics Press, 1990. → pages 126

[327] J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977. → pages 39, 41, 222, 239

[328] O. Tunçel, K. Altun, and B. Barshan. Classifying human leg motions with iniaxial piezoelectric gyroscopes. *Sensors*, 9(11):8508–8546, 2009. http://dx.doi.org/10.3390/s91108508. → pages 286

[329] L. Tweedie. Characterizing interactive externalizations. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 375–382, 1997. http://dx.doi.org/10.1145/258549.258803. → pages 21, 26, 31, 35, 39, 40, 41, 268, 275

[330] E. R. A. Valiati, M. S. Pimenta, and C. M. D. S. Freitas. A taxonomy of tasks for guiding the evaluation of multidimensional visualizations. In *Proceedings of the ACM Workshop on Beyond time*

and Errors: novel evaLuation methods for Information Visualization (BELIV), 2006. http://dx.doi.org/10.1145/1168149.1168169. → pages 21, 24, 26, 38, 39, 40, 41, 42, 67, 217, 351, 352

[331] L. van der Maaten and G. Hinton. Visualizing data using t-SNE. *Machine Learning Research*, 9:2579–2605, 2008. https://goo.gl/wXOQEo. → pages 52, 57, 298, 329

[332] L. van der Maaten, E. O. Postma, and H. J. van den Herik. Dimensionality reduction: A comparative review. Technical report, Tilburg University, TiCC-TR 2009-005, 2009. https://goo.gl/Bx2sS1. → pages 52, 298, 302, 303, 329

[333] J. M. van der Woning. *Interaction Design for Fragment-Based Molecule Parameterisation*. MSc thesis, University of Amsterdam, 2014. http://dare.uva.nl/document/529854. → pages 154, 156

[334] J. J. van Wijk. Views on visualization. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 12(4):421–432, 2006. http://dx.doi.org/10.1109/TVCG.2006.80. → pages 29, 39, 41, 219, 222

[335] J. J. van Wijk and E. R. van Selow. Cluster and calendar based visualization of time series data. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 4–9, 1999. http://dx.doi.org/10.1109/INFVIS.1999.801851. → pages 14, 119, 128

[336] J. Venna and S. Kaski. Nonlinear dimensionality reduction as information retrieval. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2007. http://goo.gl/usGF7F. → pages 304, 330

[337] K. J. Vicente. *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*. CRC Press, 1999. → pages xxiv, 27, 28, 43, 49, 53, 137, 281

[338] J. Wade. How the Tulsa World used AP's Overview to report on 8,000 police department emails. *PBS MediaShift Idea Lab*, Jun. 26, 2012. http://goo.gl/QJ64R. → pages 85, 360

[339] J. Wade. TPD working through flawed mobile system. *Tulsa World*, Jun. 3, 2012. http://goo.gl/yE98e2. → pages 86, 87, 92, 159, 359

[340] M. Wagner, F. Fischer, R. Luh, A. Haberson, A. Rind, D. A. Keim, and W. Aigner. A survey of visualization systems for malware analysis. In *Proceedings of the Eurographics Conference on Visualization (EuroVis) State of the Art Reports*, pages 105–125, 2015. http://dx.doi.org/10.2312/eurovisstar.20151114. → pages 153, 154

[341] I. M. Wallace and D. G. Higgins. Supervised multivariate analysis of sequence groups to identify specificity determining residues. *BMC Bioinformatics*, 8(135), 2007. http://dx.doi.org/10.1186/1471-2105-8-135. → pages 286

[342] J. Walny, S. Huron, and S. Carpendale. An exploratory study of data sketching for visual representation. *Computer Graphics Forum (Proceedings of EuroVis)*, 34(3):231–240, 2015. http://dx.doi.org/10.1111/cgf.12635. → pages 153

[343] M. Ward and J. Yang. Interaction spaces in data and information visualization. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 137–145, 2004. http://dx.doi.org/10.2312/VisSym/VisSym04/137-146. → pages 21, 24, 26, 34, 35, 39, 40, 42, 238, 273

[344] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 2nd edition, 2004. → pages 236, 277

[345] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 3rd edition, 2012. → pages 39, 41, 42, 277

[346] M. Wattenberg. Baby names, visualization, and social data analysis. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 1–7, 2005. http://dx.doi.org/10.1109/INFVIS.1999.801851. → pages 130

[347] N. Watzman. Private memo reveals winding tale involving John McCain, the NRA, and... condors. *Sunlight Foundation*, Sep. 18, 2014. https://goo.gl/mctTUJ. → pages 159

[348] C. Weaver. Patterns of coordination in Improvise visualizations. In *Proceedings of the SPIE–IS&T Conference on Visualization and Data Analysis (VDA)*, 2007. http://dx.doi.org/10.1117/12.704570. → pages 32, 99, 162, 271

[349] S. Wehrend and C. Lewis. A problem-oriented classification of visualization techniques. In *Proceedings of the IEEE Conference on Visualization*, pages 139–143, 1990. `http://dx.doi.org/10.1109/VISUAL.1990.146375`. → pages 21, 24, 26, 38, 39, 40, 41, 42, 150, 219, 228, 231, 234, 238

[350] F. Weil. *Artistic Human Plant Interfaces*. MA thesis, Kunstuniversität Linz, 2014. `http://blog.derhess.de/human-plant-interfaces/`. → pages 155

[351] M. West. Bayesian factor regression models in the "large p, small n" paradigm. *Bayesian statistics*, 7:733–742, 2003. `http://goo.gl/cunVdl`. → pages 327

[352] H. Wickham and L. Stryjewski. 40 years of boxplots. Technical report, had.co.nz, 2011. `http://vita.had.co.nz/papers/boxplots.html`. → pages 120, 125

[353] L. Wilkinson. *The Grammar of Graphics*. Springer, 2nd edition, 2005. → pages 32, 35, 39, 41, 42, 267

[354] L. Wilkinson and M. Friendly. The history of the cluster heat map. *The American Statistician*, 63:179–184, 2009. `http://dx.doi.org/10.1198/tas.2009.0033`. → pages 135, 136

[355] W. Willett, J. Heer, and M. Agrawala. Strategies for crowdsourcing social data analysis. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 227–236, 2012. `http://dx.doi.org/10.1145/2207676.2207709`. → pages 30, 70, 271

[356] F. O. Williams. Surprise! Many credit card agreements allow repossession. *creditcards.com*, Feb. 18, 2014. `http://goo.gl/Yeuhf2`. → pages 159

[357] M. Williams and T. Munzner. Steerable, progressive multidimensional scaling. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 57–64, 2004. `http://dx.doi.org/10.1109/INFVIS.2004.60`. → pages 54, 330

[358] M. A. Winckler, P. Palanque, and C. M. D. S. Freitas. Tasks and scenario-based evaluation of information visualization techniques. In

Proceedings of the ACM Conference on Task Models and Diagrams (TAMODIA), pages 165–172, 2004.
http://dx.doi.org/10.1145/1045446.1045475. → pages 219, 222, 352

[359] K. M. Winters, D. Lach, and J. B. Cushing. A conceptual model for characterizing the problem domain. *Information Visualization*, (in press). http://dx.doi.org/10.1177/1473871615608902. → pages 153, 154

[360] I. Witten, E. Frank, and M. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2011. → pages 52, 298, 302, 303, 304

[361] J. Wood, P. Isenberg, T. Isenberg, J. Dykes, N. Boukhelifa, and A. Slingsby. Sketchy rendering for information visualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 18(12):2749–2758, 2012.
http://dx.doi.org/10.1109/TVCG.2012.262. → pages 375

[362] J. Wright and K. Leyton-Brown. Beyond equilibrium: Predicting human behavior in normal-form games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 901–907, 2010.
http://goo.gl/WVZOSl. → pages 286

[363] R. Xu and D. Wunsch. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678, 2005.
http://dx.doi.org/10.1109/TNN.2005.845141. → pages 330

[364] M. A. Yalçın. References in "a multi-level typology of abstract visualization tasks", 2014.
http://keshif.me/demo/vis_tasks.html. → pages 39, 156

[365] J. Yang, W. Peng, M. O. Ward, and E. A. Rundensteiner. Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 105–112, 2003. http://dx.doi.org/10.1109/INFVIS.2003.1249015. → pages 52, 70, 289, 297, 298, 303, 304

[366] J. S. Yi, Y.-A. Kang, J. T. Stasko, and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*

(TVCG), 13(6):1224–1231, 2007.
http://dx.doi.org/10.1109/TVCG.2007.70515. → pages 3, 20, 21, 24, 26, 32, 38, 39, 40, 41, 42, 43, 49, 70, 150, 217, 228, 230, 238, 260, 261, 277, 299, 302, 310

[367] J. S. Yi, Y.-A. Kang, J. T. Stasko, and J. A. Jacko. Understanding and characterizing insights: How do people gain insights using information visualization? In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2008.
http://dx.doi.org/10.1145/1377966.1377971. → pages 218, 356

[368] G. Young and A. Householder. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3:19–22, 1938.
http://dx.doi.org/10.1007/BF02287916. → pages 52, 57, 298, 304

[369] L. Zhou and C. Hansen. A survey of colormaps in visualization. *IEEE Transactions on on Visualization and Computer Graphics (TVCG)*, (in press).
http://dx.doi.org/10.1109/TVCG.2015.2489649. → pages 153

[370] M. X. Zhou and S. K. Feiner. Visual task characterization for automated visual discourse synthesis. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 392–399, 1998. http://dx.doi.org/10.1145/274644.274698. → pages 21, 24, 26, 38, 39, 40, 41, 42, 68, 150, 219, 228, 229, 232, 238

[371] C. Ziemkiewicz, S. Gomez, and D. H. Laidlaw. Analysis within and between graphs: Observed user strategies in immunobiology visualization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1655–1658, 2012.
http://dx.doi.org/10.1145/2207676.2208291. → pages 27, 49, 226

Appendix A

Task Typology Supplemental Material: A Chronology

This appendix supports Chapter 2. It contains an chronological annotated bibliography of references which led to a meta-analysis of existing classifications, as well as a progression of diagrams that represent the evolution of our typology of abstract visualization tasks between late 2012 and early 2013.

In this section, I reconstruct a chronological account of the evolution of our task typology and its bibliography between 2011 and 2013. For each reference consulted, I took notes and recorded the date, and this information was maintained in a wiki document that I shared with Tamara; I reproduce some of this material here, along with comments regarding relevant events and project milestones. References are denoted by the order in which they were consulted with [R-#]. The structure of this annotated bibliography is a combination of commentary, the vocabulary of prior classifications (demarcated with *an italic font*), and references to other works. Most of these references were cited in Chapter 2 and in Brehmer and Munzner [33]; other references listed below may have been cited in an earlier draft or in the original paper submission, and these instances are remarked upon throughout this appendix.

My literature search progressed as follows: for each reference consulted, I reviewed its bibliography as well as the subsequent works that cite it (for the latter, I consulted Google Scholar, IEEE Xplore[1], and the ACM Digital Library[2], depending on where the source was published). This process allowed me to collect more references that propose a classification of tasks, activities, interactions, and the like, or those that discuss the theoretical foundations for such classifications.

A.1 Preliminary Influences

October, 2011: Work on topic of classifying visualization tasks began in earnest in September 2012. However, I had already consulted a great deal of relevant previous work between the beginning of my PhD studies (October 2011) and September 2012, during which time I was concentrating primarily on the interview study (Chapter 3) and field study (Chapter 4) projects, having throughout this period a conscious interest in evaluation methodologies for visualization.

October 7, 2011 [R-1]: H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18(9):1520–1536, 2012. `http://dx.doi.org/10.1109/TVCG.2011.279` (Pre-print version) [183].

> A survey of evaluation scenarios based on a review of over 800 references published at visualization venues. A need for task analysis arises in visualization evaluation, particularly in observational studies of people using visualization tools and techniques.

October 11, 2011 [R-2]: R. Amar and J. T. Stasko. A knowledge task-based framework for design and evaluation of information visualizations. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 143–150, 2004. `http://dx.doi.org/10.1109/INFVIS.2004.10` [7].

[1] `http://ieeexplore.ieee.org/`
[2] `http://dl.acm.org/`

Contributes a high-level classification / model that includes *rationale-based tasks* (*expose uncertainty*, *concretize relationships*, *formulate cause and effect*) and *worldview-based tasks* (*determine domain parameters*, *multivariate explanation*, *confirm hypotheses*).

October 11, 2011 [R-3]: B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2006. `http://dx.doi.org/10.1145/1168149.1168158` [292].

Describes their Multi-dimensional In-depth Long-term Case study methodology (MILC) for evaluating deployed visualization tools and techniques.

October 17, 2011 [R-4]: P. Pirolli and S. K. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of the International Conference on Intelligence Analysis*, 2005. `http://dx.doi.org/10.1057/ivs.2009.22` [242].

Contributes a high-level classification / model: describes the cyclic model of *sensemaking*. A theory of analytical reasoning involving a cycle of processes, including *information foraging or gathering*, *representation of relevant information*, *manipulation of new representations to develop insight*, and *communication of insights generated*.

In the *information foraging* loop, a person must *determine the tradeoff between exploration, enrichment, and exploitation*, *facilitate scanning, recognizing, and selecting items for further attention*, *allow shifting attentional control*, and *allow shifting attentional control*.

In the *sensemaking* loop, a person must *use external working memory for analysts to manage evidence and hypotheses*, *support* *adequate comparison of alternative hypotheses*, and *provide clear confirmation or dis-confirmation of hypotheses*.

October 17, 2011 [R-5]: J. G. Trafton, S. S. Kirschenbaum, T. L. Tsui, R. T. Miyamoto, J. A. Ballas, and P. D. Raymond. Turning pictures into numbers: Extracting and generating information from complex visualizations. *International Journal of Human Computer Studies*, 53(5):827–850, 2000. `http://dx.doi.org/10.1006/ijhc.2000.0419` [325].

Reports findings from a cognitive task analysis and protocol analysis of expert meteorologists' workflows involving visualization.

We did not cite this work in Chapter 2 or Brehmer and Munzner [33], though in hindsight perhaps we should have, given that this is a rare example of an observational study of people using visualization artefacts and an attempt to identify the tasks being performed.

October 18, 2011 [R-6]: P. Isenberg, A. Tang, and S. Carpendale. An exploratory study of visual information analysis. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1217–1226, 2008. `http://dx.doi.org/10.1145/1357054.1357245` [152].

Contributes empirical observations of people using visualization tools and techniques that refute cyclical or sequential patterns proposed by high-level classifications / models such as knowledge crystallization [48] and sensemaking [242].

October 18, 2011 [R-7]: E. Mayr, M. Smuc, and H. Risku. Many roads lead to Rome: Mapping users' problem-solving strategies. *Information Visualization*, 10(3):232–247, 2011. `http://goo.gl/x6CMbu` [202].

Contributes a high-level classification / model, distinguishing between *reading the data* (*locating and extracting the data*), *reading between the data* (*interpolating and identifying relationships*), and *reading beyond the data* (*extrapolating relationships*).

This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33], as we opted to cite Friel et al. [102] instead, as a similar classification appears there.

October 21, 2011 [R-8]: E. R. A. Valiati, M. S. Pimenta, and C. M. D. S. Freitas. A taxonomy of tasks for guiding the evaluation of multidimensional visualizations. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2006. `http://dx.doi.org/10.1145/1168149.1168169` [330].

Contributes a low-level classification, distinguishing between *identify*, *determine*, *visualize*, *compare*, *infer*, *configure*, and *locate*.

October 21–22, 2011 [R-9]: A. Perer and B. Shneiderman. Integrating statistics and visualization for exploratory power: from long-term case studies to design guidelines. *IEEE Computer Graphics and Applications (CG&A)*, 29(3):39–51, 2009. `http://doi.ieeecomputersociety.org/10.1109/MCG.2009.44` [237].

A Multi-dimensional In-depth Long-term Case study (MILC) [292] of SocialAction, a social network visualization tool. They use the low-level classification by Yi et al. [366] in their analysis of how people interacted with this tool.

October 24–25, 2011 [R-10]: J. J. Thomas and K. A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics.* IEEE, 2005 [315].

A call for establishing and strengthening a science of analytical reasoning. Describes visual analytics as serving the purposes of *increasing (cognitive and computational) resources*, *reducing search time*, *enhancing pattern recognition*, *allowing for perceptual inferences*, and *supporting interactivity.*

October 26, 2011 [R-11]: J. S. Yi, Y.-A. Kang, J. T. Stasko, and J. A. Jacko. Understanding and characterizing insights: How do people gain insights using information visualization? In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2008. `http://dx.doi.org/10.1145/1377966.1377971` [367].

> A meta-review and attempt to characterize *insight*; they posit that insight is not a product (as in Pirolli and Card [242]), but a midpoint for more cycling and iterations back and forward towards a product. Insight is gained during several overlapping but distinct processes: when *providing an overview (or big picture)*, when *adjusting the level of abstraction* (i.e., *grouping / filtering*), when *detecting patterns*, and when *the data matches one's mental model*.
>
> This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33], as we decided to refrain from discussing *insight*.

October 31 – Novermber 2, 2011 [R-12]: P. Pirolli. *Information Foraging Theory: Adaptive Interaction with Information.* Oxford University Press, 2009 [241].

> Contributes a high-level classification, a model of *sensemaking*. This work aims to generally describe *information foraging* behaviour of any rational agents, with a heavy emphasis on information on the web. This work is is a response to the immense amount of data constantly being generated and added to the web and the many information retrieval methods and tools at our disposal.
>
> This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33], opting to cite Pirolli and Card [242] instead (see above).

November 2, 2011 [R-13]: M. A. Winckler, P. Palanque, and C. M. D. S. Freitas. Tasks and scenario-based evaluation of information visualization techniques. In *Proceedings of the ACM Conference on Task Models and Diagrams (TAMODIA)*, pages 165–172, 2004. `http://dx.doi.org/10.1145/1045446.1045475` [358].

> A recapitulation of the classifications of Wehrend and Lewis [349] and Zhou and Feiner [370]; contributes a low-level classification, distinguishing between *abstract tasks* (*locate, identify, distinguish, reveal, cluster, emphasize, explore*), *user tasks* (*identify by name, portray, individualize, profile*), *application tasks* (*focus, isolate, reinforce, expose, itemize, specify, separate, outline, individualize, highlight, colour, zoom*), and *interactive tasks* (*select, finish*).
>
> This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33] because we cited Wehrend and Lewis [349] and Zhou and Feiner [370].

November 5, 2011 [R-14]: J. J. van Wijk. Views on visualization. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 12(4):421–432, 2006. `http://dx.doi.org/10.1109/TVCG.2006.80` [334].

> Makes compelling or noteworthy assertions about the behaviour of people who use visualization tools or techniques: visualization is used to *present* information or to *discover* and *explore* new information.

November 16, 2011 [R-15]: R. R. Springmeyer, M. M. Blattner, and N. L. Max. A characterization of the scientific data analysis process. In *Proceedings of the IEEE Conference on Visualization*, pages 235–242, 1992. `http://dx.doi.org/10.1109/VISUAL.1992.235203` [301].

> Contributes a domain- and datatype-agnostic classification that spans low-level and high-level tasks, distinguishing between

investigation (*interacting with representations, applying math, maneuvering*) and *integration of insight* (*maneuvering, expression of ideas*).

November 29, 2011 [R-16]: P. André, J. Teevan, and S. T. Dumais. Discovery is never by chance: Designing for (un)serendipity. In *Proceedings of the ACM Conference on Creativity & Cognition*, pages 305–314, 2009. http://dx.doi.org/10.1145/1640233.1640279 [9].

> Makes compelling or noteworthy assertions about the serendipitous observation of unexpected phenomena. This paper is a meta-review of computational tools built to support serendipity, however the paper argues that these systems only account for one part of serendipity: *the chance discovery of something unexpected*, or something sought after in an unexpected location (the cause). Many systems do not account for the second aspect of serendipity (the effect), *the sagacity or insight to acknowledge an unexpected connection with earlier knowledge and expertise*, and the will to act upon these connections, by reinforcing an existing problem or solution, rejecting or confirming ideas, or starting a new research direction.

December 2, 2011 [R-17]: C. North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications (CG&A)*, 26(3):6–9, 2006. http://dx.doi.org/10.1109/MCG.2006.70 [228].

> A short article addressing the issue of measuring *insight* generation in visualization. Describes the disadvantages of simple benchmark lookup and search tasks and calls for experimental tasks with greater complexity (such as *characterizing distributions, correlations, and patterns, or estimating various statistical metrics*).
>
> North also suggests that we measure insight in a qualitative, unconstrained way with real domain users interacting with their

own data. A rigorous coding protocol would be required for such studies. On the other hand, the experimenter no longer needs to design benchmark tasks.

This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33], as we decided to refrain from discussing *insight*.

December 6, 2011 [R-18]: D. Klahr and H. A. Simon. Studies of scientific creativity: Complimentary approaches and convergent findings. *Psychological Bulletin*, 125(5):524–543, 1999. `http://psycnet.apa.org/doi/10.1037/0033-2909.125.5.524` [174].

Contributes a high-level classification / model, distinguishing between *normal everyday problem solving* and *scientific problem solving*. The latter requires combination of *strong and weak methods*; strong methods incorporate a rich amount of domain expertise, methodology, and background, while weak methods are domain-independent, incorporating trial and error, hill climbing, means-ends analysis, and planning, and bridging between strong and weak via analogy.

Klahr and Simon [174] also discuss the role of surprise and they distinguish between scientific investigations that are theory or hypothesis driven, and those that are driven by observation of an unexpected or surprising phenomena.

Finally, this paper includes a discussion of the role of analogical reasoning for formulating initial hypotheses and the notion of multiple search spaces: a parallel search of a hypothesis space, an experiment space, and a representation space (abstractions, visual representations, notation), and a strategy/instrumentation space.

This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33]; in our discussion of generating and verifying hypotheses, we refer to others, including André et al. [9], Pike et al. [239], and Tukey [327].

December 8, 2011: I prepared a summary of the literature I had reviewed up until this point in a presentation entitled *"The black box... of Sensemaking and Scientific Discovery"*. As the title suggests, I discussed sensemaking and information foraging models of Pirolli [241], and I distinguished between the high-level data analysis taxonomies of Amar and Stasko [7] and Springmeyer et al. [301], the low-level data analysis taxonomy of Winckler et al. [358], and the scientific discovery process described by Klahr and Simon [174]. I attempted to establish commonalities between these models and taxonomies and their relevance to the visualization and visual analytics communities, making references to arguments by Thomas and Cook [315] and van Wijk [334]. I indicated the need for a classification of *mid-level activities*, calling upon constructs such as *insight, discovery, serendipity, learning, creativity*, and *problem solving*. I also pointed out that metrics associated with any of these constructs are difficult to define and that the evaluation of visual analysis and visualization processes will likely involve multiple metrics related to more than one of these constructs. Altogether, my thinking about these constructs in the context of evaluating visualization was rather nebulous and unfocused in retrospect, and it was not yet clear as to what this line of thinking would lead to. Nevertheless, I continued with the literature review in a part-time capacity over the course of the next nine months.

December 14, 2011 [R-19]: R. Chang, C. Ziemkiewicz, T. M. Green, and W. Ribarsky. Defining insight for visual analytics. *IEEE Computer Graphics and Applications (CG&A)*, 29(2):14–17, 2009. `http://dx.doi.org/10.1109/MCG.2009.22` [53].

> Distinguishes between two definitions of *insight* and the implications of these two definitions for the design and evaluation of visualization tools. One view is that that insight is an event; and other view characterizes insight as a quantity, an amount

of knowledge gained that occurs upon integrating and building upon one's existing representations, making associations between disparate concepts. While neither is trivial to track or measure, the authors suggest that in the context of visualization, the two forms of insight support each other and occur in a loop, wherein knowledge-based insight elicits or enables event-based insight.

This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33], as we decided to refrain from discussing *insight*.

December 14, 2011 [R-20]: Y.-A. Kang and J. T. Stasko. Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 19–28, 2011. `http://dx.doi.org/10.1109/VAST.2011.6102438` [164].

Contributes a domain-specific classification of intelligence analysis activities, in which they identify parallel rather than sequential tasks: *constructing and refining a conceptual model*, *data collection*, *analysis*, and *production*.

They emphasize the need for the support of collaboration and sharing, and that the intelligence process is not linear or sequential the model of Pirolli and Card [242] would imply.

April 12, 2012 [R-21]: D. W. Sprague and M. Tory. Motivation and procrastination: Methods for evaluating pragmatic casual information visualizations. *IEEE Computer Graphics and Applications (CG&A)*, 29(4): 86–91, 2009. `http://dx.doi.org/10.1109/MCG.2009.70` [300].

A short methods paper, proposing techniques for evaluating the usability and utility of casual visualization artefacts in situ, with several case studies of casual visualization artefacts.

We did not cite this work in Chapter 2 or Brehmer and Munzner [33], opting to cite the more comprehensive 2012 study by Sprague and Tory [299] instead.

April 12, 2012 [R-22]: D. Sprague and M. Tory. Exploring how and why people use visualizations in casual contexts: Modeling user goals and regulated motivations. *Information Visualization*, 11(2):106–123, 2012. `http://dx.doi.org/10.1177/1473871611433710` [299].

Makes compelling or noteworthy assertions about the behaviour of people who use visualization artefacts in casual settings, describing contexts in which the information being visualized is simply enjoyed, where people indulge their casual interests in a topic, where novelty stimulates curiosity and thereby exploration. These artefacts include information graphics, floor plans, advertisements, and forms of entertainment.

They describe aspects that promote the use of visualization artefacts in casual settings such as personal interest, usefulness, curiosity, data correctness and trust, the cost of misinterpretation, and aesthetics. They also describe aspects that inhibit the use of visualization artefacts in these settings: time constraints and higher priority tasks, learning effort, and insufficient data context.

A person's goals in these contexts can be *extrinsic* (motivated by social pressure or a desire to avoid boredom) and *intrinsic*, which can be distinguished by referring to *learning and understanding* (*curiosity, information acquisition*), *utility* (*instruction, scheduling, task completion, orientation*), and *entertainment* (*humour, self-expression*).

This paper in part motivated our inclusion of `enjoy` in our typology.

April 20, 2012 [R-23]: C. North, P. Saraiya, and K. Duca. A comparison of benchmark task and insight evaluation methods for information visualization. *Information Visualization*, 10(3):162–181, 2011. `http://dx.doi.org/10.1177/1473871611415989` [229].

Expands on the differences between *task-based* and *insight-*

based evaluation methodologies. Insight-based methodologies treats tasks as dependent measures; assessing how a visualization artefacts promotes tasks, rather than how it supports tasks. How a tool supports tasks is a question better suited for a task-based methodology. An insight-based methodology can address higher-level questions regarding task taxonomies, conclusions about visualization artefacts, time spent by participants in a study, and effort spent analyzing the data. Insight-based methodologies are particularly effective when a data analysis process is exploratory in nature.

June 27, 2012 [R-24]: S. Kandel, A. Paepcke, J. M. Hellerstein, and J. Heer. Enterprise data analysis and visualization: An interview study. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18 (12):2917–2926, 2012. `http://dx.doi.org/10.1109/TVCG.2012.219` [163].

Makes compelling or noteworthy assertions about the behaviour of people who use visualization tools or techniques in enterprise contexts. Contributes a temporal / pipeline model of data analysis containing the following terms: *discover*, *wrangle*, *profile*, *model*, and *report*.

July 24, 2012 [R-25]: G. Marchionini. Exploratory search: From finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006. `http://dx.doi.org/10.1145/1121949.1121979` [199].

Makes compelling or noteworthy assertions about the behaviour of people who use search tools or techniques, distinguishing between *lookup* (*fact retrieval, known item search, navigation, transaction, verification, question answering*), and *exploratory search*, which can be further separated into *learning* (*knowledge acquisition, comprehension / interpretation, comparison, aggregation / integration, socialize* and *investigating* (*accretion, analysis, exclusion / negation, synthesis, evaluation, discovery, planning / forecasting, transformation*).

August 14, 2012 [R-26]: C. Ziemkiewicz, S. Gomez, and D. H. Laidlaw. Analysis within and between graphs: Observed user strategies in immunobiology visualization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1655–1658, 2012. `http://dx.doi.org/10.1145/2207676.2208291` [371].

> Describes an observational study of four biologists using a visualization tool called GenePattern, in which the authors coded their observations using the classifications of Springmeyer et al. [301] and Amar et al. [8]. Their findings revealed two interaction strategies despite participants having similar goals and context, a *within-graphs interaction strategy* and a *between-graphs interaction strategy*. Participants used the visualization tool to *increase confidence in results*, to *ensure the validity of the data*, and for *efficiently viewing lots of data*.

August 22, 2012 [R-27]: M. Beaudouin-Lafon. Designing interaction, not interfaces. In *Proceedings of the ACM Conference on Advanced Visual Interfaces (AVI)*, pages 15–22, 2004. `http://dx.doi.org/10.1145/989863.989865` [16].

> A position paper discussing how models or frameworks describing interaction can be evaluated, distinguishing between *descriptive power* (the ability to describe a range of existing interfaces), *evaluative power* (the ability to help assess multiple design alternatives), and *generative power* (the ability to help designers create new designs). He writes: "*High level-models tend to have good descriptive power but poor evaluative and generative power. Low-level models tend to have poor descriptive and evaluative power, but higher generative power. A good interaction model must strike a balance between generality (for descriptive power) concreteness (for evaluative power), and openness (for generative power).*"

August 28, 2012 [R-28]: M. Pohl, S. Wiltner, and S. Miksch. Exploring information visualization: Describing different interaction patterns. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2010. `http://dx.doi.org/10.1145/2110192.2110195` [245].

> Presents findings from a study in which the authors evaluated use of a visualization tool intended for exploratory data analysis by means of qualitative and quantitative log file analysis. They found task-specific usage patterns, trends of macro-tasks (common sequences of interactions). They characterize some of these patterns according to Gestalt definitions of problem solving.

A.2 Dedicated Literature Review

September, 2012: As mentioned at the start of the previous section, a deliberate effort toward a classification of tasks began in September 2012, when we began to consolidate notes from the aforementioned sources and we initiated a dedicated literature review of additional classifications and associated frameworks or theories.

At this time I had arrived at an impasse in both the interview study (Chapter 3) and field study (Chapter 4) projects, which I had been working on in parallel up until this point. Our 2012 IEEE InfoVis submission about the interview study had been rejected (we then published it as a technical report [283]), as reviewers found the connections to visualization to be tenuous. Meanwhile, with regards to the field study, I mentioned an explicit need for a classification of visualization tasks among my preliminary field study findings (see Section C.5):

> *"A valid and comparative evaluation methodology requires a robust mid-level task characterization of exploratory data analysis, one that spans domains and tool interfaces. . . . My long-term goal is to contribute to the construction of such a task characterization."*

As indicated in this quote, my interest in developing a classification of tasks was motivated out of an interest in qualitative evaluation. However, over the next few months, I would come to realize the value of a classification of tasks in other contexts, such as in visualization design and quantitative evaluation.

Meanwhile, Tamara had been thinking about task classification in the context of the book she was writing[3], and she was struggling to reconcile existing classifications by Amar and Stasko [7], Amar et al. [8], Casner [51], Lee et al. [186], Shneiderman [291], Wehrend and Lewis [349], Yi et al. [366], Zhou and Feiner [370] and a new classification by Heer and Shneiderman [130]. She had already sketched out some ideas as to how a classification of tasks could be presented, such as in Figure A.1. It became clear at this point that Tamara and I had similar goals: in my case, I was in need of task classification a data analysis tool to use in my projects, whereas Tamara was in need of task classification to communicate concepts in her book. We then decided to continue our literature review and meta-analysis of existing classifications toward a new task classification.

September 11, 2012 [R-29]: J. Heer and B. Shneiderman. Interactive dynamics for visual analysis: A taxonomy of tools that support the fluent and flexible use of visualizations. *Communications of the ACM*, pages 1–26, 2012. `http://dx.doi.org/10.1145/2133416.2146416` [130].

> Contributes a domain- and datatype-agnostic classification that spans low-level and high-level tasks, distinguishing between *data / view specification* (*visualize, filter, sort, derive*), *view manipulation* (*select, navigate, coordinate, organize*), and *process and provenance* (*record, annotate, share, guide*).
>
> One of the existing classifications that Tamara was attempting to reconcile.

September 13, 2012 [R-30]: M. X. Zhou and S. K. Feiner. Visual task characterization for automated visual discourse synthesis. In *Proceedings*

[3] which would be published over a year later [219].

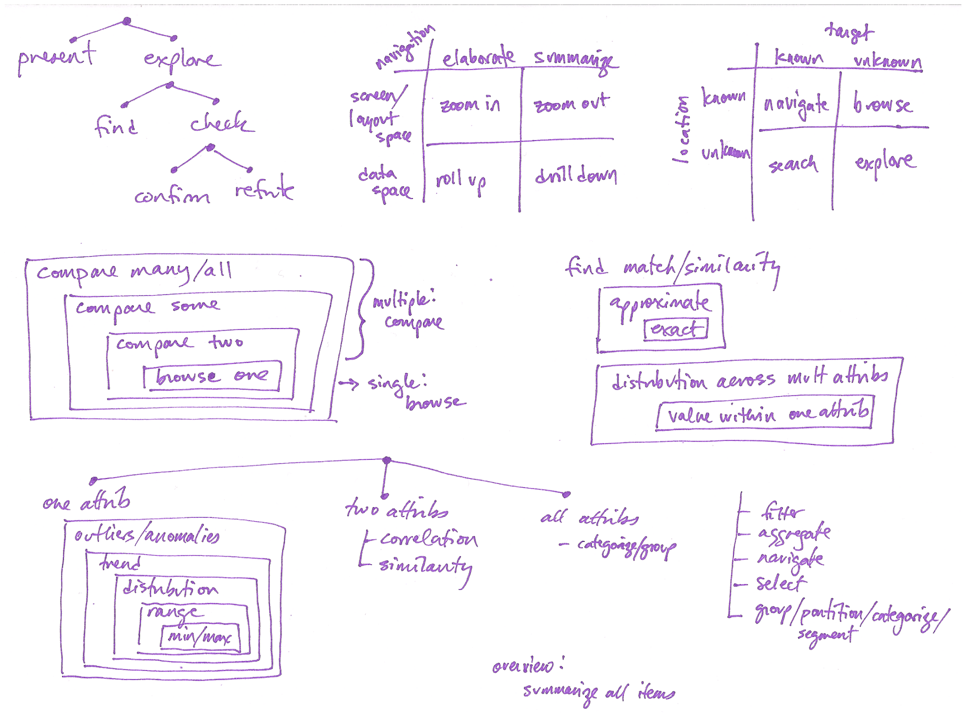


Figure A.1: August 27, 2012: Tamara's early brainstorming on the topic of task classification.

of the ACM Conference on Human Factors in Computing Systems (CHI), pages 392–399, 1998. `http://dx.doi.org/10.1145/274644.274698` [370].

Contributes a low-level classification, distinguishing between *associate (collocate, connect, unite, attach), background, categorize (mark distribution), cluster (outline, individualize), compare (differentiate, intersect), correlate (plot, mark compose), distinguish (mark distribution, isolate), emphasize (focus, isolate, reinforce), generalize (merge), identify (name, portray, individualize, profile), locate (position, situate, pinpoint, outline), rank (time), reveal (expose, itemize, specify, separate), switch*, and *encode (label, symbolize (quantify, iconify), portray, tabulate, plot, structure, trace, map)*.

These terms can be occur in various combinations with respect to *inform* and *enable* intents. The types of *inform* include *elaborate* and *summarize*. The types of enable include *explore* (*search, verify*) and *compute* (*sum, compute*).

One of the existing classifications that Tamara was attempting to reconcile.

September 13, 2012 [R-31]: J. S. Yi, Y.-A. Kang, J. T. Stasko, and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 13(6):1224–1231, 2007. `http://dx.doi.org/10.1109/TVCG.2007.70515` [366].

Contributes a low-level classification that distinguishes between *select, explore, reconfigure, encode, abstract, elaborate, filter, connect* and a category for *other* processes including *undo* and *redo*, *change configuration / layout / settings*.

One of the existing classifications that Tamara was attempting to reconcile.

September 18, 2012 [R-32]: R. Amar, J. Eagan, and J. T. Stasko. Low-level components of analytic activity in information visualization. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 111–117, 2005. `http://dx.doi.org/10.1109/INFVIS.2005.1532136` [8].

Contributes a low-level classification that distinguishes between *retrieve, filter, compute derived value, find extremum, sort, determine range, characterize distribution, find anomalies, cluster*, and *correlate*.

One of the existing classifications that Tamara was attempting to reconcile.

September 18, 2012 [R-33]: S. F. Roth and J. Mattis. Data characterization for intelligent graphics presentation. In *Proceedings of the ACM*

Conference on Human Factors in Computing Systems (CHI), pages 193–200, 1990. `http://dx.doi.org/10.1145/97243.97273` [263].

> Contributes a low-level classification that distinguishes between *value lookup, compare within a relation, compare across or between relations, determine distributions, determine correlations, indexing*, and *sorting*.
>
> One of the existing classifications that Tamara was attempting to reconcile.

September 18, 2012 [R-34]: S. Wehrend and C. Lewis. A problem-oriented classification of visualization techniques. In *Proceedings of the IEEE Conference on Visualization*, pages 139–143, 1990. `http://dx.doi.org/10.1109/VISUAL.1990.146375` [349].

> Contributes a low-level classification that includes *identify (lookup value), locate, distinguish, categorize, cluster (determine), distribution, rank, compare (within and between relations), associate*, and *correlate*.
>
> One of the existing classifications that Tamara was attempting to reconcile.

September 18, 2012 [R-35]: E. H. Chi and J. T. Riedl. An operator interaction framework for visualization systems. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 63–70, 1998. `http://dx.doi.org/10.1109/INFVIS.1998.729560` [58].

> Contributes a low-level classification, a temporal / pipeline model of visualization construction, describing four stages interleaved with four transformations: *the data stage (value filtering, subsetting), data transformation (deriving, computing new attributes, aggregating), the analytical abstraction stage (select subset to visualize), visualization transformation (e.g., cluster, MDS), the visualization abstraction stage (simplify), visual mapping transformation (choose visual encoding technique)*, and *the*

view stage (navigation, orientation, view-filter, dynamic view-filtering, brushing, animating, focus, permute).

September 19, 2012 [R-36]: E. Morse, M. Lewis, and K. A. Olsen. Evaluating visualizations: Using a taxonomic guide. *International J. Human-Computer Studies*, 53(5):637–662, 2000. `http://dx.doi.org/10.1006/ijhc.2000.0412` [214].

An example of an evaluation where the authors use the classification of Zhou and Feiner [370] to evaluate four different visualization tools for information retrieval. They translate a subset of this abstract classification into concrete domain tasks particular to the datasets used in the study.

We cited this work in early drafts but not in Chapter 2 or Brehmer and Munzner [33], as we opted to cite Perer and Shneiderman [237] as an example of an evaluation incorporating a prior classification.

September 20, 2012 [R-37]: B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, 1996. `http://dx.doi.org/10.1109/VL.1996.545307` [291].

Contributes a low-level classification, distinguishing between *overview, zoom, filter, details-on-demand, relate, history*, and *extract*.

Also contributes a classification of tasks by data type for one-dimensional (*counting, filtering, details-on-demand*), two-dimensional (subsumes one dimensional tasks, *containment, compare, relate*), and three-dimensional data (subsumes one and two dimensional data tasks, *adjacency, understanding position and orientation, resolving occlusion*), tasks for temporal or time-oriented data (subsumes one-dimensional data tasks, *determine start / end, find overlap, find events before / after / during*), for multi-dimensional data (*finding patterns of variables, gaps, outliers, resolving disorientation, occlusion*), as well as for tree-based data (subsumes one-dimensional data tasks applied to items and links, *determine how many levels in the tree, how many children does an item have, examine types of objects at different tree depths, examine breadth / depth*) and network data (subsumes tree-based data tasks, *examine shortest/less costly paths, network traversal*).

One of the existing classifications that Tamara was attempting to reconcile.

September 21, 2012 [R-38]: H. Lam. A framework of interaction costs in information visualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 14(6):1149–1156, 2008. `http://dx.doi.org/10.1109/TVCG.2008.109` [180].

Describes a survey of 484 references appearing in visualization venues and analyzed according to the *Seven Stages of Action* framework by Norman [226]. In addition to Norman's *gulf of execution* and *gulf of evaluation*, Lam adds a *gulf of formation* to represent high-level cognitive decision costs related to data analysis and intent formation.

September 21, 2012 [R-39]: B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry. Task taxonomy for graph visualization. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2006. `http://dx.doi.org/10.1145/1168149.1168168` [186].

Contributes a low-level classification specific to graph-based data, distinguishing between *topology tasks* (*determine adjacency, determine accessibility, find common connection, determine connectivity: shortest path, clusters, connected components, bridges, articulation points*), *attribute tasks* (*find node at-*

tributes, find link attributes), *browsing tasks* (*follow path, revisit*), and *overview tasks*.

One of the existing classifications that Tamara was attempting to reconcile.

September 25, 2012 [R-40]: T. J. Jankun-Kelly, K. L. Ma, and M. Gertz. A model and framework for visualization exploration. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 13(2):357–69, 2007. `http://dx.doi.org/10.1109/TVCG.2007.28` [157].

Proposes a formal model to explain exploration with visualization tools, a formal "*P-set*" grammar for describing the model, and contributes a software framework for recording exploration. The model centres around *parameter adjustments* and *transformations* that reference Card et al. [48], including *data filtering, data transformation, visual (primitive) mapping, rendering, and view transformations*. Claims that previous classifications by Chi and Riedl [58], [60], Shneiderman [291], and Wehrend and Lewis [349] "*focus on the goal of the user, not how the visualization was used to achieve those goals*".

This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33], as we also cited Card et al. [48], who in turn inspired the choice of elements in the *P-set* grammar.

September 26, 2012 [R-41]: D. Gotz and M. X. Zhou. Characterizing users' visual analytic activity for insight provenance. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 123–130, 2008. `http://dx.doi.org/10.1057/ivs.2008.31` [117].

Contributes a low-level classification that distinguishes between *exploration*, *insight actions*, and *meta actions*.

Exploration includes *data exploration* (*filter, inspect, query, restore*) and *visual exploration* (*brush, change metaphor, change range, merge, sort, split*).

Insight actions include those related to *visual insight* (*annotate, bookmark*) and *knowledge insight* (*create, modify, remove*).

Meta actions include *delete, edit, redo, revisit*, and *undo*.

September 26, 2012 [R-42]: S. M. Casner. A task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics*, 10(2):111–151, 1991. `http://dx.doi.org/10.1145/108360.108361` [51].

Contributes a low-level classification, distinguishing between *search operators* (*search, lookup, verify*) and *computation operators* (*equal, less than, greater than, plus, difference, times, quotient*).

One of the existing classifications that Tamara was attempting to reconcile.

September 26, 2012 [R-43]: P. M. Mullins and S. Treu. A task-based cognitive model for user-network interaction: Defining a task taxonomy to guide the interface designer. *Interacting with Computers*, 5(2):139–166, 1993. `http://dx.doi.org/10.1016/0953-5438(93)90016-M` [216].

Contributes a domain- and datatype-agnostic classification that spans low-level and high-level tasks, containing over 140 terms: an exhaustive list of high-level *mediation* and *coordination* tasks, as well as many low-level object-oriented interactions relating to physical interface input and output.

Their list of *mediation* and *coordination* terms include: *assess* (*success*, *evaluate*, *test*, *compare*, *verify*), *analyze* (*interpret*, *calculate*, *categorize*, *count*, *itemize*, *tabulate*), *synthesize* (*integrate*, *translate*, *remember*, *prioritize*, *estimate*, *extrapolate*, *interpolate*), *solve problems* (*plan*, *formulate*, *plan*, *program*, *diagnose*, *decide*, *choose*), *learn* (*query*, *tutorial*, *browse*), *undo*, *reset*, and *cross-reference*.

Their list of *object space modeling* terms include: *create* (*associate*, *name*, *group*, *link*, *assemble*, *aggregate*, *paste*, *overlay*, *insert*, *replicate*, *copy*, *instance*, *store*, *introduce*, *data entry*, *restore*), *eliminate* (*remove*, *cut*, *delete*, *purge*, *disassociate*, *rename*, *ungroup*, *unlink*, *disassemble*, *segregate*, *filter*, *suppress*, *withdraw*), *activate* (*execute a process*, *start*, *invoke*, *change status*, *re-start*, *foreground/background switch*, *stop process*, *suspend*, *terminate*, *set-aside*, *quit*), *indicate* (*pick*, *reference*, *mark*), *edit*, and *display*

Their list of *communicate* terms include: *transmit* (*call*, *acknowledge*, *respond*, *suggest*, *direct*, *inform*, *instruct*, *request*, *record*, *transform*, *stretch*, *sketch*, *re-orient*, *shape*, *pan*, *zoom*, *move*, *select*, *position*, *orient*, *quantify*, *text*, *hold*, *push*, *pull*, *control*), *reach through*, and *receive* (*attend*, *monitor*, *notice*, *filter*, *accept*, *acquire*, *observe*, *scan*, *search*, *inspect*, *extract*, *screen*, *detect*, *discriminate*, *recognize*, *identify*, *locate*).

A neglected paper that Tamara pointed me to.

September 27, 2012 [R-44]: C. Ware. *Information Visualization: Perception for Design.* Morgan Kaufmann, 2nd edition, 2004 [344].

Contributes a high-level classification. In Ware's *Interacting with Visualizations* chapter, there are three loops of activity that define interaction with a visualization: *the data manipulation loop, the exploration and navigation loop*, and *the problem solving loop.*

In Ware's *Thinking with Visualizations* chapter, the *problem solving loop* is tangentially addressed, referencing the sensemaking model of Pirolli and Card [242]. However, this chapter mainly deals with memory models and knowledge costs, limits to visual working memory, and eye movements. There is a section on *visual problem solving*, which breaks the process down into the following nested hierarchy: *problem solving, visual query, the* *pattern finding loop, eye movements, and intrasaccadic image-scanning.*

The final section of the chapter addresses *creative problem solving* and *creative thinking* as high-level tasks, which involves stages of *preparation*, *production*, and *judgment*.

September 28, 2012 [R-45]: M. Tory and T. Möller. Rethinking visualization: A high-level taxonomy. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 151–158, 2004. `http://dx.doi.org/10.1109/INFVIS.2004.59` [321].

Makes compelling or noteworthy assertions about the behaviour of people who use visualization tools or techniques. The authors outline the relationships between tasks with different visualization design models. Their classification of tasks breaks down into *whether spatialization is constrained* and whether the design model is assumed to be *continuous* or *discrete*.

For a *continuous design model*, tasks include *find[ing] spatial relationships and spatial regions of interest* for a *given spatialization* and *finding numeric trends* for a *chosen spatialization*.

For a *discrete design model*, tasks can pertain to *values* (*finding patterns such as clusters and outliers*) or *structure* (*analyzing connectivity relationships*); other tasks include *retrieving item details and filtering or excluding items*.

October 1, 2012 [R-46]: M. C. Chuah and S. F. Roth. On the semantics of interactive visualizations. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 29–36, 1996. `http://dx.doi.org/10.1109/INFVIS.1996.559213` [60].

Contributes a low-level classification, which distinguishes between *graphical operations*, *set operations*, and *data operations*.

Their list of *graphical operations* includes *encode data* (*create mapping, transform mapping*), *set graphical value*, *manipulate objects* (*creating, copying, deleting*).

Their list of *set operations* includes *create* (*enumerate, express membership*), *delete*, and *summarize*.

Their list of *data operations* includes *add objects, delete objects*, and *derive attributes*.

October 2, 2012 [R-47]: C. Plaisant, D. Carr, and B. Shneiderman. Image-browser taxonomy and guidelines for designers. *IEEE Software*, 12 (2):21–32, 1995. `http://dx.doi.org/10.1109/52.368260` [244].

Presents a taxonomy of browser interfaces, both single-view and multiple-view variants, including overview and detail displays. Contributes a high-level task classification that justifies the different browser types: *image generation*, *open-ended exploration*, *diagnostic*, *navigation*, and *monitoring*.

October 3, 2012 [R-48]: R. E. Roth. Cartographic interaction primitives: Framework and synthesis. *Cartographic Journal*, 49(4):376 – 395, 2012. `http://dx.doi.org/10.1179/1743277412Y.0000000019` [261].

Contributes a meta-analysis of interaction taxonomies and classifications in the visualization and cartographic literature, distinguishing these prior classifications by referring to Norman [226] and his terms *objective (intents)*, *operator (tools/widgets)*, and *operand (data objects/abstractions)* from his *Stages of Action* model.

Roth's meta-analysis partially overlaps with our own: he includes the classifications of Amar et al. [8], Buja et al. [42], Chi and Riedl [58], Chuah and Roth [60], Dix and Ellis [75], Keim [166], Shneiderman [291], Ward and Yang [343], Wehrend and Lewis [349], Yi et al. [366], and Zhou and Feiner [370]. His meta-analysis also includes 13 other classifications, largely from the geovisualization and cartographic literature.

Roth identifies several concordances and differences across the classifications that he surveyed: the terms *identify* and *compare* are the most common; most *objective*-oriented classifications are at a high-level of abstraction; *objective* and *operator* classifications are often hard to delineate; *brushing* is the most common *operator*, *focusing* is defined in many different ways, there is ambiguity related to *changing or altering the encoding or symbolization*; *viewpoint operators* are related to *distortion, navigation, observer motion, object rotation, panning, re-centring, re-projecting, viewpoint manipulation, and zooming*; *operand* classifications vary between being *type-centric* and *state-centric*, while others vary between *data operands* and *representation operands*.

October 4, 2012 [R-49]: J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977 [327].

Makes compelling or noteworthy assertions about *exploratory data analysis*, described as being more than descriptive statistics, it requires flexibility and an attitude a "*willingness to look for what can be seen*". *Confirmatory data analysis* can be automated, but it depends upon *exploratory data analysis*.

October 4, 2012 [R-50]: Y. B. Shrinivasan and J. J. van Wijk. Supporting the analytical reasoning process in information visualization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1237–1246, 2008. `http://dx.doi.org/10.1145/1357054.1357247` [293].

Contributes a high-level classification that distinguishes between *externalizing evidence, hypotheses, assertions, and causal links, organizing evidence to support/refute a claim, reviewing and revising the exploration process, linking externalized evidence to support claims*, and *presenting findings*.

The focus of this paper is on *visualization history tracking* and *knowledge externalization*. History and knowledge externalization can involve artefacts such as screen shots, annotations, lists of bookmarked elements or locations, parameter settings, or interaction logs.

October 5, 2012 [R-51]: W. Dou, D. H. Jeong, F. Stukes, W. Ribarsky, H. R. Lipford, and R. Chang. Recovering reasoning processes from user interactions. *IEEE Computer Graphics & Applications (CG&A)*, 29(3):52–61, 2009. `http://dx.doi.org/10.1109/MCG.2009.49` [79].

> The authors conducted a study in which a team of four volunteers qualitatively coded interaction logs of a visual analytics tool for financial data, a tool they were familiar with: they interactions they coded were grouped into three categories: *finding*, *strategy*, and *method*.
>
> This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33], as we opted to refer to Shrinivasan and van Wijk [293] in our discussion of interaction logs.

October 10, 2012 [R-52]: R. A. Rensink. On the prospects for a science of visualization. In W. Huang, editor, *Handbook of Human Centric Visualization*, pages 147–175. Springer, 2014. `http://dx.doi.org/10.1007/978-1-4614-7485-2_6` (Pre-print version) [254].

> Rensink discusses prospects for a science of visualization with regards to *low-level visual perception tasks*, borrowing experimental methodologies from vision science but retaining stimuli and visual tasks from the visualization domain, which include: *perception of correlation, pattern detection, cluster detection, outlier detection, grouping, finding convex hull of points, following a curve, finding point of maximum intensity.*
>
> Rensink proposes an *extended vision thesis* and an *optimal reduction thesis*. The former refers to the viewer and visualization system as a single system. The latter refers to reducing a task to low-level operations in the extended systems (low-level visual tasks that do not involve interactivity).

October 14–19, 2012: I attended the 2012 IEEE VIS Conference and the 2012 ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV). Some discussion at BELIV included arguments against task classification: that tasks are nonlinear systems, and that new contexts, new data types, and changes of scale result in new tasks all the time; further complexity is added with collaborative tasks and tasks spread over multiple tools.

I also noted several papers of potential relevance at InfoVis/VAST [67, 69, 187, 246], and at BELIV [111, 207] which I would add to our literature review and meta analysis over the next few months.

October, 2012 [R-53]: M. Gleicher. Stop the evaluation arms race! A call to evaluate visualization evaluation. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2012 [111].

> This position paper asks whether evaluation (of a visualization tool / technique) is necessary. The results of an evaluation should be *actionable* and *persuasive*.
>
> We cited this work in early drafts but not in Chapter 2 or Brehmer and Munzner [33], as our claim that we had developed an *actionable* and *persuasive* classification was disputed by readers of early drafts, as discussed below.

November 13, 2012 [R-54]: M. Pohl, M. Smuc, and E. Mayr. The user puzzle: Explaining the interaction with visual analytics systems. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VAST)*, 18(12):2908–2916, 2012. `http://dx.doi.org/10.1109/TVCG.2012.273` [246].

> Contributes a meta-analysis of cognitive science theories and frameworks that can explain visual analysis using visualization tools and techniques, including *sensemaking* (see our discussion of Pirolli and Card [242], Pirolli [241], and Klein et al. [175]), *Gestalt theory*, *distributed cognition* (see our discussion of Hollan et al. [137], Liu et al. [194], and Kirsh and Maglio [173]), *graph*

comprehension (see our discussion of Friel et al. [102]) and *skill-rule knowledge theories*.

They indicate whether (and the degree to which) these theories or frameworks help to explain *preattentive processing, tool and mapping comprehensibility, visual pattern detection, open data exploration, the solving of ill-defined and well-defined problems, interaction strategies, interpretation, hypothesis generation and testing, insights, sensemaking, decision making, operationalizability, errors*, and *collaboration*.

A.3 Meta-Analysis of Existing Classifications

November 2012 – January 2013: After surveying the sources listed up until this point, Tamara and I began to characterize the dimensions and foci of existing classifications (Figures A.2 and A.3 and Table A.1). These figures are culled from eight slide presentations generated between November 2012 and March 2013[4], as a slide presentation was the primary medium in which we consolidated, framed, refined, and commented on our ideas.

This meta-analysis eventually culminated in the discussion in Section 2.5 and in Table 2.1 and Table 2.2.

Figure A.2 indicates some of the dimensions considered in this meta-analysis of existing classifications, along with examples of high-level, mid-level, and low-level tasks. These dimensions refer to the domains interested in studying the tasks, the domain-specificity of the task, the linearity of the task, whether the tasks can be performed with variation from person to person, how the tasks could be studied, and how these tasks are supported.

Meanwhile, Figure A.3 presents another view of our meta-analysis, in which our notion of "mid-level tasks" are described as being interface- and domain-independent.

Table A.1 summarizes the metadata associated with existing classifications, which included bibliographic information, the depth or number of hierarchical levels in the particular classification, and meta-classifications

[4] 489 slides in total, with an average of 60 slides per presentation.

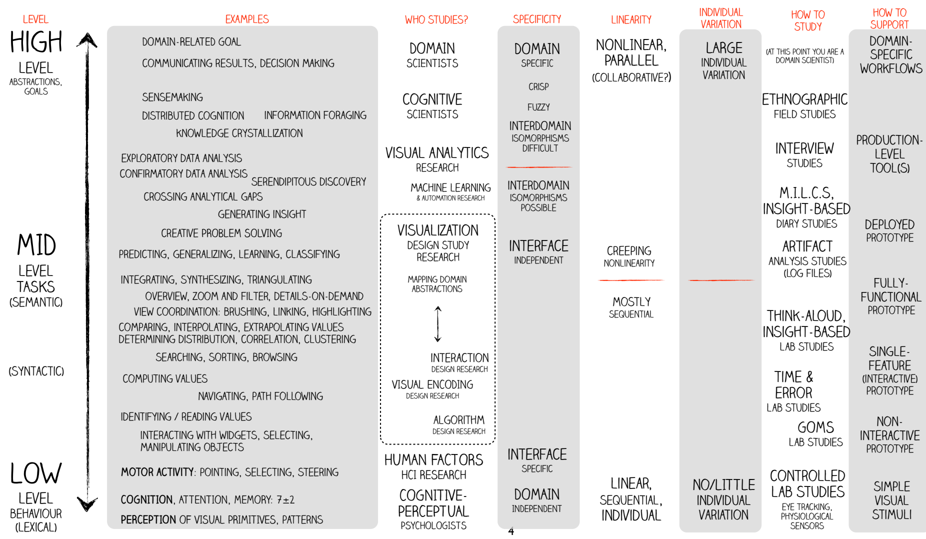


Figure A.2: November 14, 2012: Cross-cutting dimensions of high-level, mid-level, and low-level tasks: who studies them, their specificity, their linearity, individual variation, how they could be studied, and how these tasks are supported.

terms originating with Roth [261] and Chuah and Roth [60], as well as the method or means by which the classification was developed.

I also continued to consult additional literature during this period.

November 21, 2012 [R-55]: J. A. Cottam, A. Lumsdaine, and C. Weaver. Watch this: A taxonomy for dynamic data visualization. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 193–202, 2012. `http://dx.doi.org/10.1109/VAST.2012.6400552` [67].

> Discuses visualization tools and technique that involve dynamic (i.e., changing) data, introduces a vocabulary for quantifying *dynamic change*: *identity preserving transformations, transitional transformations, and immediate transformations*, where the type of transformation depends on the task.
>
> We did not cite this work in Chapter 2 or Brehmer and Munzner [33], as this classification pertains more to the design space

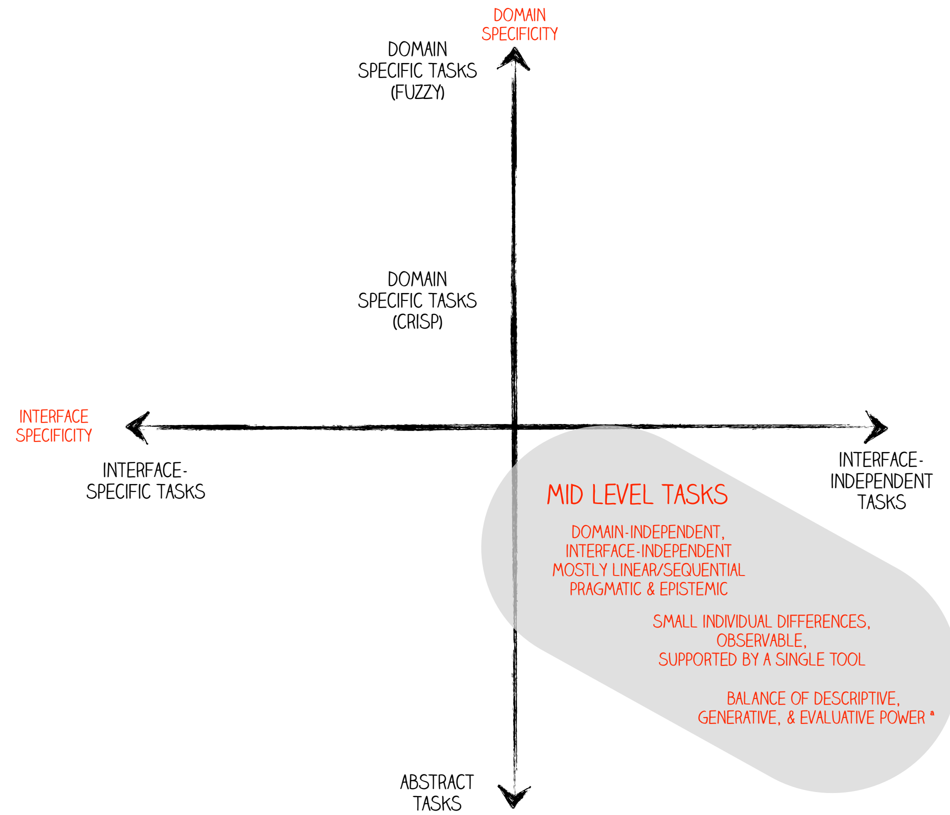


Figure A.3: November 14, 2012: Mid-level abstract tasks along the axes of domain specificity and interface specificity.

of transformation techniques and not to tasks or interactions.

November 21, 2012 [R-56]: R. J. Crouser and R. Chang. An affordance-based framework for human computation and human-computer collaboration. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of VAST)*, 18(12):2859–2868, 2012. `http://dx.doi.org/10.1109/TVCG.2012.195` [69].

Contributes an affordance-based framework for mixed-initiative systems (human-computer collaboration systems), distinguishing human affordances and machine affordances, choos-

PAPER	YEAR	VENUE	CITATIONS (SCHOLAR)	CITATIONS (ACM)	CITATIONS (IEEE)	DOMAIN	LEVEL	DEPTH	FOCUS	METHOD
AMAR & STASKO (2004)	2004	IEEE INFOVIS	104	4	21	INFOVIS	HIGH	2	OBJECTIVE / SEMANTIC	INTROSPECTION, EXAMPLES
AMAR, EAGAN, & STASKO (2005)	2005	IEEE INFOVIS	126		2	INFOVIS	LOW	1	OBJECTIVE* / SYNTACTIC	OBSERVATIONAL STUDY W/ STUDENTS, AFFINITY DIAGRAMMING
CARD, MACKINLAY, SHNEIDERMAN (1999)	1999	READINGS IN INFORMATION VISUALIZATION	3,207			INFOVIS	MID-HIGH	2	OBJECTIVE / SEMANTIC	INTROSPECTION
CASNER (1991)	1991	ACM T. GRAPHICS	323	79		HCI	LOW	2	OPERATOR / SYNTACTIC	DESIGN AUTOMATION LANGUAGE, EXPERIMENTAL STUDY
CHI & RIEDL (1998)	1998	IEEE INFOVIS	142		4	INFOVIS	LOW	3	OPERAND* / TEMPORAL / SEMANTIC	LITERATURE SEARCH
CHUAH & ROTH (1996)	1996	IEEE INFOVIS	78		4	INFOVIS	LOW	2.5	OPERATOR*, OPERAND / SYNTACTIC	INTROSPECTION, EXAMPLE
GOTZ & ZHOU (2008)	2008	VAST	45		1	VA	LOW-MID	3	OPERATOR, OPERAND / SYNTACTIC	OBSERVATIONAL STUDY, NON-EXPERTS
HEER & SHNEIDERMAN (2012)	2012	COMM. ACM	7	0		INFOVIS	LOW-MID	3	OPERATOR / SEMANTIC	REFLECTING ON EXAMPLE SYSTEMS
LEE, & PLAISANT (2006)	2006	BELIV	73	22		INFOVIS	LOW	3	OPERAND / SYNTACTIC	INTROSPECTION
MARCHIONINI (2006)	2006	COMM. ACM	514	150		HCI	HIGH	3	OBJECTIVE / SEMANTIC	INTROSPECTION, EXAMPLES
MAYR, SMUC, RISKU (2010)	2010	BELIV / IV JOURNAL	2	0		INFOVIS	HIGH	2	OBJECTIVE / SEMANTIC	LAB STUDY, NON-EXPERTS, INTERACTION LOGGING
MULLINS & TREU (1993)	1993	INTERACTING W/ COMPUTERS	9			HCI	LOW-MID	4	OBJECTIVE, OPERATOR, OPERAND / LEXICAL, SYNTACTIC, SEMANTIC	INTROSPECTION, VALIDATED BY QUESTIONNAIRE W/ EXPERTS
PIROLLI & CARD (2005)	2005	INTL. CONF. IA	200			VA	HIGH	2	OBJECTIVE, TEMPORAL, SEMANTIC	INTROSPECTION
ROTH & MATTIS (1990)	1990	CHI	224	57		INFOVIS	LOW	1	OPERAND, SYNTACTIC	INTROSPECTION
SHNEIDERMAN (1996)	1996	IEEE VISUAL LANG.	1,977		108	INFOVIS	LOW-MID	2	OPERATOR*, OPERAND* / SEMANTIC, SYNTACTIC	INTROSPECTION
SPRINGMEYER (1992)	1992	IEEE VIS	78		4	VIS	MID-HIGH	3	OBJECTIVE / SEMANTIC	INTERVIEWS W/ SCIENTISTS, OBSERVATIONS
VALIATI & FREITAS (2006)	2006	BELIV	28	6		INFOVIS	LOW-MID	1	OPERAND/ SYNTACTIC	LAB STUDY, NON-EXPERTS, EXPERT REVIEWER, QUESTIONNAIRE
WEHREND & LEWIS (1990)	1990	IEEE VIS	207		8	VIS	LOW	1	OBJECTIVE*, OPERAND* / SYNTACTIC	LITERATURE SEARCH
WINCKLER & FREITAS (2004)	2004	TAMODIA	20	4		INFOVIS	LOW-MID	2	OBJECTIVE, OPERATOR / SYNTACTIC	SCENARIOS
YI, KANG, STASKO, JACKO (2007)	2007	TVCG	169		22	INFOVIS	LOW	2	OBJECTIVE* / SYNTACTIC, LEXICAL	LITERATURE SEARCH, AFFINITY DIAGRAMMING
ZHOU & FEINER (1998)	1998	CHI	137	35		INFOVIS	LOW-MID	3	OBJECTIVE* / SEMANTIC, SYNTACTIC	TASK GRAMMAR, EXAMPLES, USED IN EVAL BY MORSE (2000)

Table A.1: November 14, 2012: Additional dimensions of previous classifications. Citation counts are as of December 2012. Level pertains to the degree of abstraction. Depth pertains to the number of hierarchical levels in the particular classification. Asterisks in the focus column represent a classification by Roth [261]; semantic and syntactic are terms used by Chuah and Roth [60]. Method pertains to how the particular classification was developed.

ing not to take a task-centric or deficiency-centric approach to constructing the framework. In other words, rather than allocate tasks to a human because a computer cannot do it (computationally intractable), their framework takes the view that tasks should be allocated to the human because a human is good at it. The authors recognize that humans adapt and can learn complex tasks. Humans and computers should leverage the affordances of the other for harmonious cooperation and problem-solving.

Human affordances include *visual perception, visuospatial thinking, audiolinguistic ability, sociocultural awareness, creativity*, and *domain expertise*.

Machine affordances include *large-scale data manipulation, collecting and storing large amounts of data, efficient data move-*

ment, bias-free analysis.

The framework can be used for guiding function (i.e., task) allocation: to the human, to the computer, or split between both. The authors state: "*We need to to develop a set of canonical actions that humans can perform with known complexity, but compiling this list is nontrivial*".

We did not cite this work in Chapter 2 or Brehmer and Munzner [33], as we opted to focus on processes (i.e., verbs) rather than the characteristics of the actors performing these processes.

November 21, 2012 [R-57]: L. A. McNamara and N. Orlando-Gay. Reading, sorting, marking, shuffling: Mental model formation through information foraging. In *Proceedings of the ACM Workshop on Beyond time and Errors: novel evaLuation methods for Information Visualization (BELIV)*, 2012 [207].

A position paper distinguishing between general forms of *analytics* and specific forms of *analysis*, informed by interviews with 18 intelligence analysts.

We did not cite this position paper in Chapter 2 or Brehmer and Munzner [33].

A.4 Our Initial Classifications of Tasks

November 2012 – February 2013: At this point, Tamara and I had examined the vocabulary and definitions used in the existing classifications that we had surveyed. Our meta-analysis reflected a top-down perspective, in which we had established the dimensions on which we could distinguish prior classifications. However, this process did not yield our own classification of tasks; to do so, we tried for a bottom-up approach, in which we would begin with a small set of existing classifications and gradually add and combine similar terms and remove redundant or duplicate terms, and we would continue to do so we considered additional classifications and the definitions that their authors had provided for the terms they contained. we relied upon affinity diagramming using tools such as OmniGraffle[5], Keynote[6], as well as post-its and whiteboards. The diagrams and images that follow appeared throughout the eight slide presentations mentioned in Section A.3. We began by consolidating the classifications of Heer and Shneiderman [130], Springmeyer et al. [301], Amar et al. [8], Gotz and Zhou [117], and Chuah and Roth [60], as shown in Figure A.4.

At this point, our notion of an ideal *task taxonomy* was one that describes tasks at a middle level of abstraction, is domain- and interface-independent, focusing on the semantics of tasks (their objective or intent) and their temporal or sequential dependencies. Such as taxonomy would have *descriptive*, *evaluative*, and *generative* power [16], providing *actionable* and *persuasive* guidance [111] for visualization design and evaluation. Our plan was to translate existing classifications into a semantic and objective language (i.e., the active voice, a stance encouraged in qualitative data analysis and grounded theory in particular [55]), identify emergent patterns following this translation, and attempt to map root nodes of low-level classifications to the leaf nodes of high-level classifications.

Through multiple rounds of coding, we continued to group similar terms and consider additional classifications, we selected representative terms for each group, and we arranged these representative terms into multiple levels of abstraction. This process is reflected in Figure A.5, Figure A.6, and Figure A.7. In our second classification, (Figure A.5), we considered an reorganization around `explore` and "provenance tasks" (the latter referring mostly to data and process management, inspired by the *process and provenance* classification by Heer and Shneiderman [130]). In both cases, leaf nodes pointed to operands (aspects of the data or view). This reorganization also reflects a hybrid between the bottom-up approach reflected in Figure A.4 and the high-level top-down thinking reflected in Tamara's early proposal (see Figure A.1).

In our third classification (Figure A.7), we introduced the questions

[5] `https://omnigroup.com/omnigraffle`
[6] `http://apple.com/mac/keynote/`

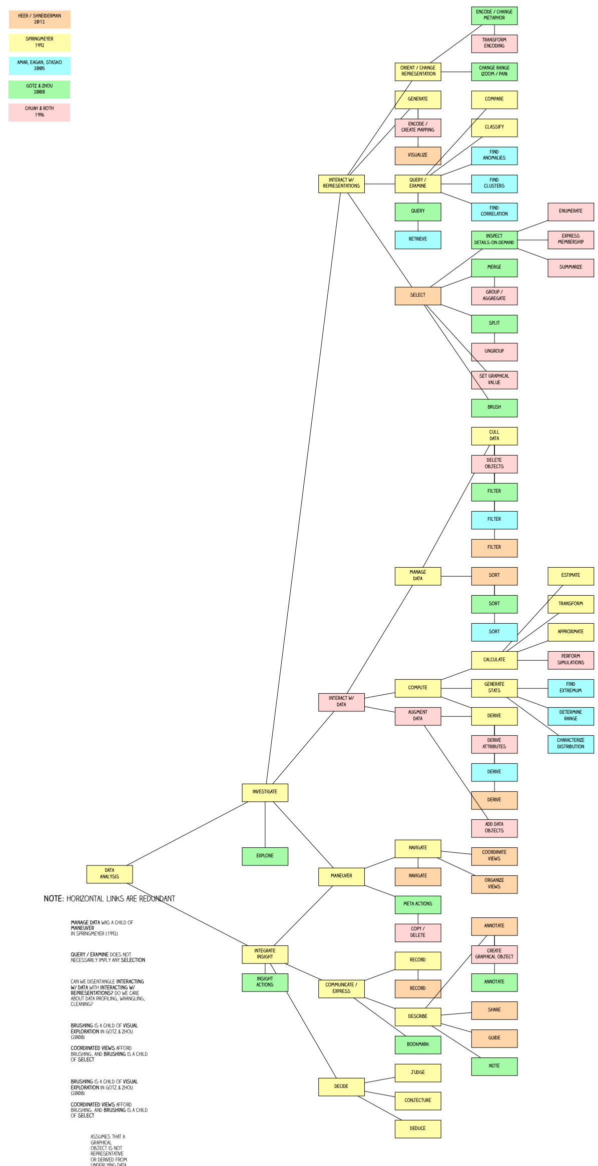


Figure A.4: November 21, 2012: Our first classification, originating from "data analysis" and consolidating existing terms used by Heer and Shneiderman [130] (orange), Springmeyer et al. [301] (yellow), Amar et al. [8] (blue), Gotz and Zhou [117] (green), and Chuah and Roth [60] (pink).

of *why* (ends, objectives, goals), *what* (operands), *which* (also referring to operands), and *how* (means, methods, operators); we also considered the special case of *search*. The objective, operator, and operand language is attributed to Norman [226], which had also inspired a meta-analysis of cartographic interaction classifications by Roth [261].

November 2012 – February 2013: While we refined our classification, I continued to consult additional literature during this period, now venturing further from the core visualization literature to literature from the HCI, information retrieval, communications, and distribution cognition research communities.

December 7, 2012 [R-58]: Z. Liu, N. Nersessian, and J. T. Stasko. Distributed cognition as a theoretical framework for information visualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 14(6):1173–1180, 2008. `http://dx.doi.org/10.1109/TVCG.2008.121` [194].

> A discussion of distributed cognition theory in the context of visualization, distinguishing between *epistemic* and *pragmatic* actions. Pragmatic actions are explicitly and consciously goal-directed, while epistemic actions serve to coordinate actors' internal mental models with external representations of information.
>
> Information is propagated as a series of *representation states*, some are external within or between representations and artefacts, while some are internal to individuals (or shared between individuals). In the context of visualization, distributed cognition can help us to acknowledge both external and internal states of information, to capture the process of coupling, coordination, interaction strategies for sensemaking and analytical reasoning, the creation and evolution of external representations, and the role of interaction in aiding understanding.
>
> The authors state: "*A science of interaction should not be just a taxonomy of interaction techniques or a framework of the*

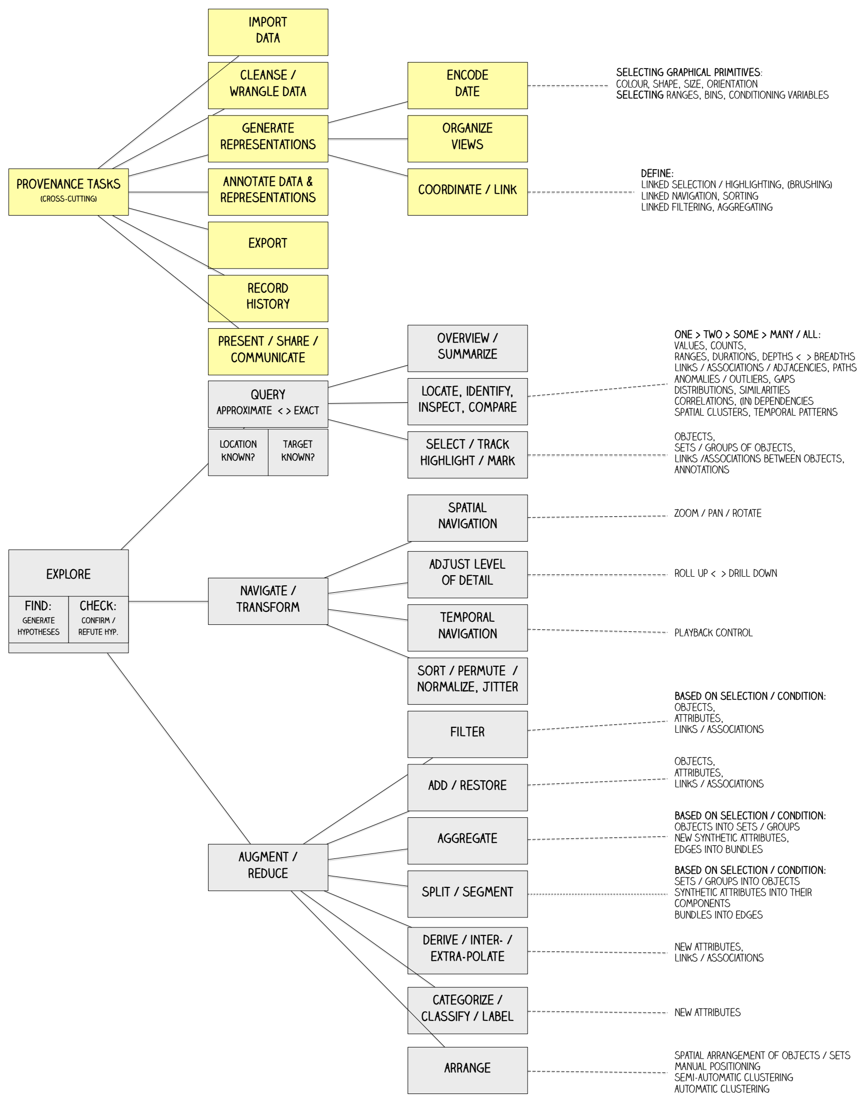


Figure A.5: November 23, 2012: Our second classification, reorganizing terms and distinguishing between "provenance tasks" and `explore`.

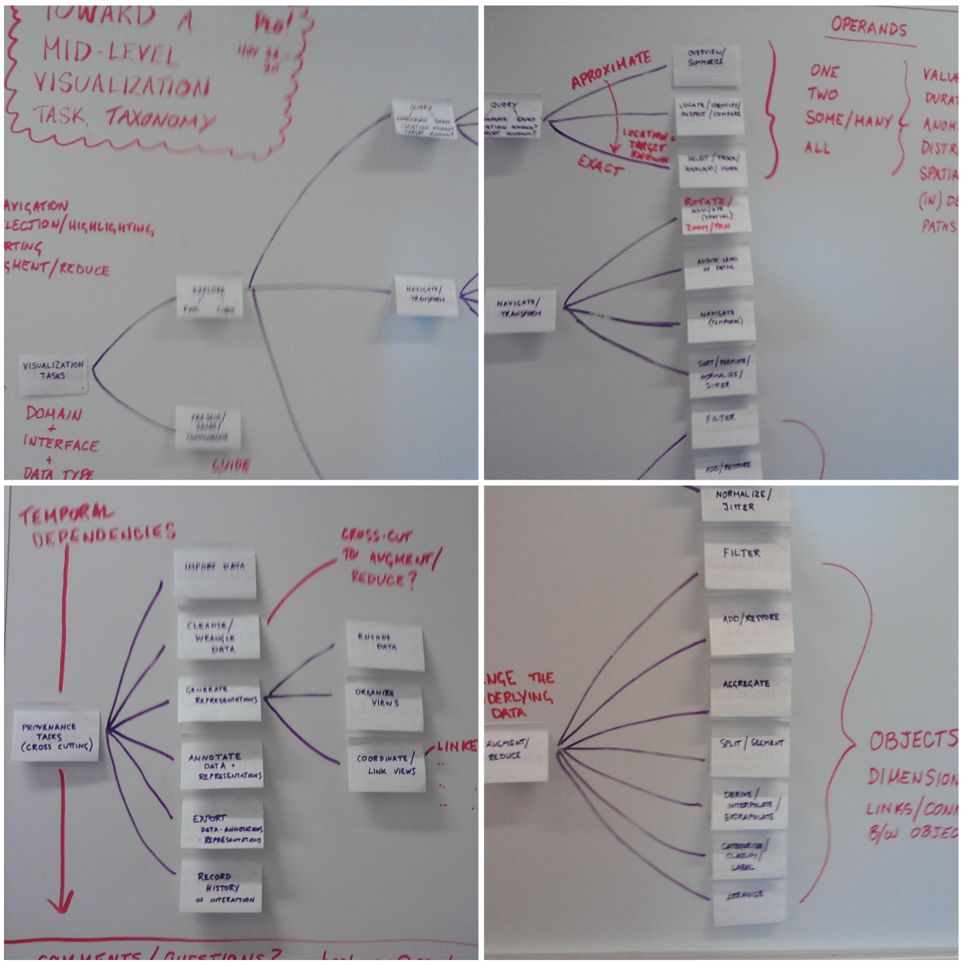


Figure A.6: November 26–28, 2012: Whiteboard and post-it diagramming prior to our third classification reflecting the influence of Roth [260] (and Norman [226], indirectly), reflecting an organization around operands (top right), objectives (bottom right), and temporal dependencies (bottom left).

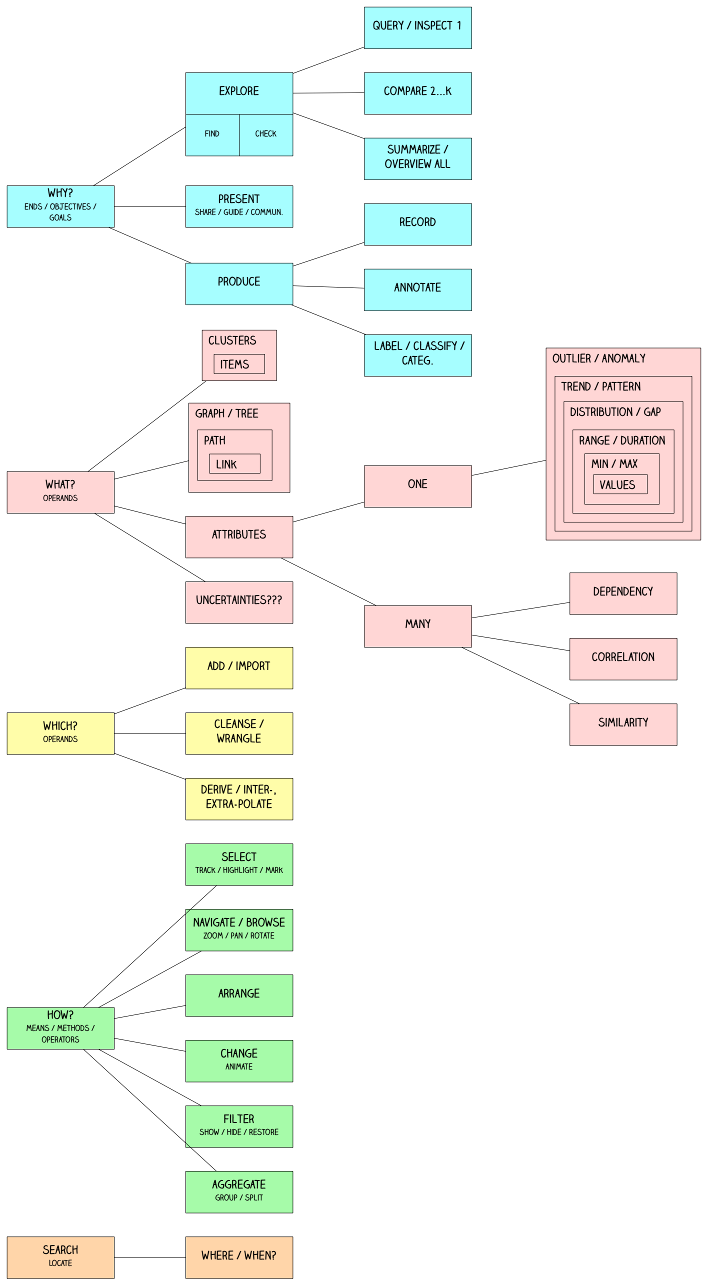


Figure A.7: November 30, 2012: Our third classification, which introduced *why*, *how*, *what*, *which*, as a means to group terms, as well as a special case for `search`.

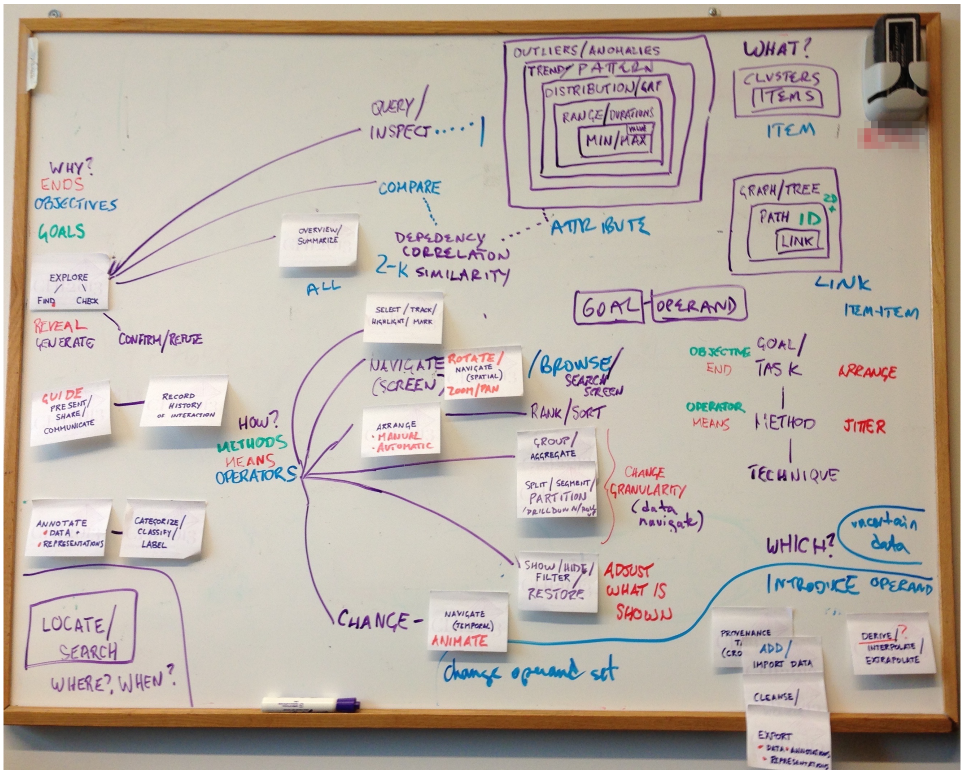


Figure A.8: December 9, 2012: Whiteboard and post-it diagramming prior to our fourth classification; we continued to iterate on a classification revolving around *why*, *how*, *what*, *which*, and `search`.

abstracted task procedures; it should be a scientific approach to understand how cognition emerges as a property of interaction between internal and external representations."

Referred to in the meta-analysis of cognitive science theories by Pohl et al. [246] (see above).

December 9, 2012: Prior to our fourth classification, we performed more brainstorming and diagramming such as in Figure A.8.

December 17, 2012 [R-59]: J. Hollan, E. Hutchins, and D. Kirsh. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7 (2):174–196, 2000. `http://dx.doi.org/10.1145/353485.353487` [137].

> An introduction to distributed cognition theory and the implications for HCI research. Their suggested research framework involves studying *how people establish and coordinate structure within an environment*, *how coordination is maintained over time*, *how cognitive effort is offloaded to the environment when it is practical to do so*, achieving a better *conceptualization of what is going on and what ought to be done*, and *when and how cognitive load-balancing is achieved between human actors, the environment and its artefacts*.
>
> Referred to in the meta-analysis of cognitive science theories by Pohl et al. [246] (see above).

December 17, 2012 [R-60]: D. Kirsh and P. Maglio. On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18(4):513–549, 1994. `http://dx.doi.org/10.1207/s15516709cog1804_1` [173].

> A distributed cognition paper that distinguishes *epistemic actions* (gaining information about the problem at hand) from *pragmatic actions* (which support users toward their goals). *Epistemic actions* are those that *make mental computation easier, faster, more reliable, less error prone, mediating intermediate results, reducing cognitive load, simplifying one's problem solving task, reducing both space and time complexity. Epistemic tasks* are *memory- and time-saving actions, information gathering and exploratory actions*. They are intended as *external checks or verifications to reduce the uncertainty of judgments*.
>
> Referred to in the meta-analysis of cognitive science theories by Pohl et al. [246] (see above).

December 18, 2012 [R-61]: D. Kirsh. Distributed cognition: A methodological note. *Pragmatics & Cognition*, 14(2):249–262, 2006. `http://dx.doi.org/10.1075/pc.14.2.06kir` [172].

A brief article that discusses the methodological and design implications of a distributed cognition framework for HCI.

Kirsh states: "*Formalism-based abstractions on task environments or task procedures are the major approaches for studying how people interact with artefacts and other people, yet these are known to be flawed in making unreasonable assumptions about the predictability of human behaviours and characterizing the environment as a fixes set of choice points.*"

This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33], as we opted for more comprehensive references to the distributed cognition literature, such as Hollan et al. [137] and Kirsh and Maglio [173].

December 19, 2012: As we continued to iterate on our meta-analysis and classification efforts, we considered how to explicitly frame our project using the language of Tamara's Nested model [217][7]. We also added a distributed cognition framing to our meta-analysis of existing classifications and specifically the distinction between *pragmatic* and *epistemic* actions. We also began to group terms around nodes in our classifications, which would eventually lead to Table 2.1 and Table 2.2.

December 15 2012 – January 3, 2013 [R-62]: T. S. Kuhn. *The Structure of Scientific Revolutions.* University of Chicago Press, 4th edition, 1962 [178].

Discusses how *normal science* is a puzzle-solving process governed by *paradigms* (abstractions shared by those within a specific discipline, describing what the puzzle looks like and what parts of the puzzle remain to be solved), which in turn influence a set of rules, by which the course of normal scientific experimentation, measurement, analysis, and theorizing proceed.

[7] This framing was documented in our second slide presentation, as explained in Section A.3.

Anomaly, *invention*, and *discovery* are violations to the order of *normal science* and the expectations that a paradigm has provided. Exploration of the anomaly leads to a *paradigm shift*, and during this shift, disagreement about these discoveries leads to a state of crisis in which new theories emerge.

Recommended by committee member Ron Rensink; we did not cite this work in Chapter 2 or Brehmer and Munzner [33], as Kuhn is referring to a much larger scope of processes encompassing in entire scientific discipline, whereas we focused on processed of individuals or groups as they use a certain class of tools and techniques.

January 3–4, 2013 [R-63]: D. O. Case. *Looking for Information: A Survey of Research on Information Seeking, Needs and Behavior*. Emerald Group Publishing, 2nd edition, 2008 [50].

In a Chapter called "*Models of information behaviour*", Case surveys five models of *information seeking* intended to generalize over a wide population, considering various needs, individual differences, forms of media. Case places an emphasis on casual information seeking and non-work-related motivators.

In a Chapter called "*Theories, perspectives, paradigms*", Case discusses the paradigms (with nods to Kuhn [178]) and theoretical perspectives relating to information behaviour. Case reviews five major theories and their sources relating to information seeking behaviour, emanating from psychology, sociology, communications, management and business, consumer research, economics, and linguistics. Included among these theories is *sensemaking* and *play theory* (see discussion of Stephenson [303] below).

I became aware of this work via UBC InfoVis group member Jessica Dawson, who in early 2012 had completed a UBC Library and Information Studies graduate seminar course (LIBR

553 2011-12: *Understanding information users in diverse environments*); Case [50] was required reading for this course. This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33], opting to cite original sources such as Toms [318] instead.

January 7, 2013 [R-64]: E. G. Toms. Serendipitous information retrieval. In *Proceedings of the DELOS Workshop on Information Seeking, Searching and Querying in Digital Libraries*, pages 1–4, 1999. `http://goo.gl/01pwgQ` [317].

Discusses *serendipity* and the motivation for open-ended exploration: "*significant evidence exists to support the concept that people also acquire information that was never sought and about which the individual may have had no predisposition.*"

Helps to account for casual information retrieval: "*"There was no need, no anomalous state of knowledge and no knowledge gap evident. This was simply an information gathering experience without expectations or predicted outcome novelty stimulated curiosity (and thus exploration).*" This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33], opting to cite the more comprehensive Toms [318] paper instead.

January 8, 2013 [R-65]: E. G. Toms. Understanding and facilitating the browsing of electronic text. *International J. of Human-Computer Studies (IJHCS)*, 52(3):423–452, 2000. `http://dx.doi.org/10.1006/ijhc.1999.0345` [318].

Makes compelling or noteworthy assertions about the browsing behaviour of people who use information retrieval tools, reporting on an experiment in which two groups of participants read related news articles, where one group was given a learning objective and another was not; the two groups' interactions and usage of the articles varied.

Toms remarks that the terms *browsing, navigating, exploring, way-finding, travelling, orienteering, foraging, grazing, wandering, surfing*, and *skimming* are often conflated. *Browsing* relies on juxtapositions of content in time and space, presented within a context that stimulates a person; "*browsing may be more like news reading, more in line with the play theory of Stephenson [303]. The success of browsing may be in the experience itself and not in the outcome.*"

This paper was also part of the syllabus for LIBR 553 2011-12: *Understanding information users in diverse environments*, along with Case [50].

January 9, 2013: Our fourth classification is represented in Figure A.9. By this point, we consolidated *search* and *why* and we established a stronger associated between *which* and *what*. We also began to use the terms *epistemic* and *pragmatic*, which we borrowed from the distributed cognition literature[8].

January 11, 2013 [R-66]: D. A. Norman. *The Psychology of Everyday Things*. Basic Books, 1988 [226].

Norman's *Stages of Action* model has seven stages and two gulfs: *stage 1: forming the goal*, the *gulf of execution*, *stage 2: forming the intention*, *3: specifying the action*, *4: executing the action*, the *gulf of evaluation*, *stage 5: perceiving the state of the system*, *6: interpreting the state of the system*, and *7: evaluating the outcome*.

Norman's model and its influence on Roth's *objective-operand-operator* meta-analysis [261] of previous classifications helped shape the *why-what-how* organization of our typology.

[8]We documented these developments in our third slide presentation, as explained in Section A.3.

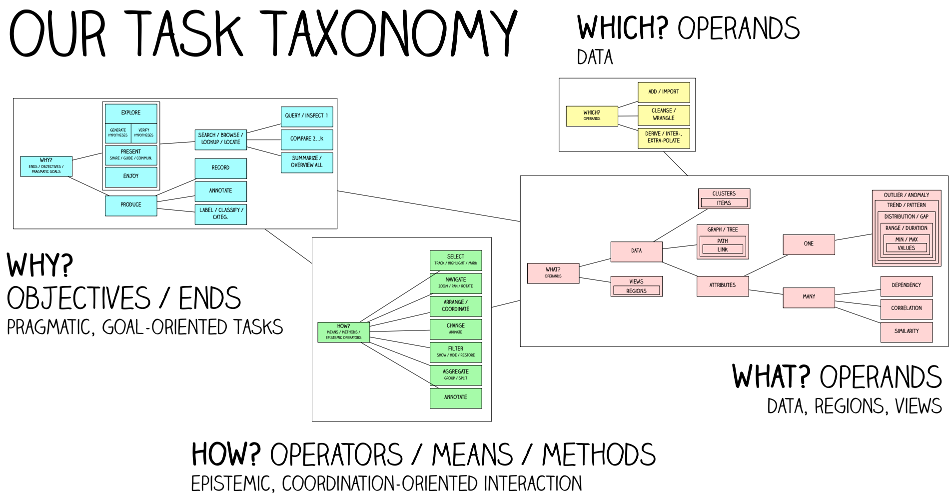


Figure A.9: January 9, 2013: Our fourth classification, where *why* corresponds to objectives / ends, *how* corresponds to operators / means / methods, while *what* and *which* correspond to operands. Note that we begin to use the terms *epistemic* and *pragmatic*, which we borrow from the distributed cognition literature.

Lam [180] (discussed above) extended Norman's model in the context of visualization with a *gulf of goal formation*, relevant whenever a person articulates their own questions pertaining to visualized data.

January 14, 2013 [R-67]: R. E. Roth. An empirically-derived taxonomy of cartographic interaction primitives. In *Proceedings of the GIScience*, 2012. `http://goo.gl/7HgB2N` [260].

Contributes a datatype specific classification of interaction primitives relating to cartographic data based on Norman's *Stages of Action* model, distinguishing between *goals* (or *meta-objectives*: *procure, predict, prescribe*), *objectives* (*identify, compare, rank, associate, delineate*), *operators*, and *operands*.

Among Roth's list of *operators* are *work operators* (*re-express, arrange, sequence, re-symbolize, overlay, re-project, pan, zoom,* *filter, search, retrieve, calculate*) and *enabling operators* (*import, export, save, edit, annotate*).

With respect to *operands*, Roth distinguishes between the *search target* (*in space alone, in space-in-time, an attribute-in-space*) and *search level* (*elementary, general*).

Roth arrived at this classification via a series of card-sorting exercises on *objectives* and *operators* and 15 cartographic interface designer/developer participants. His rationale for this method: "*One limitation of extant taxonomies contributing to their lack of general adoption is that the majority of these taxonomies are not empirically derived, instead relying on secondary sources or personal experience.*" Roth noted a high degree of variability in responses across the *objectives*, as there appeared to be some confusion stemming from different *operands*.

A longer version of this short workshop paper appeared as Roth [262] at IEEE InfoVis 2013 (see below).

January 14, 2013 [R-68]: W. A. Pike, J. T. Stasko, R. Chang, and T. A. O'Connell. The science of interaction. *Information Visualization*, 8(4):263–274, 2009. `http://dx.doi.org/10.1057/ivs.2009.22` [239].

A survey and research agenda highlighting challenges in interaction within the visual analytics community. Prior to addressing the research agenda, they provide a domain- and datatype-agnostic classification that spans low-level and high-level tasks, consolidating prior work from the same group by Amar and Stasko [7], Amar et al. [8], and Yi et al. [366].

The authors state that despite many prior classifications, they see a need to understand a relationship between task/interaction components and modes of inquiry: *abduction* (constructing hypotheses), *deduction* (refuting prior hypotheses), *induction* (verification, ranking alternative hypotheses, identifying the best explanation).

Among *user goals* or *tasks*, they distinguish between *high-level* (*explore, analyze, browse, assimilate, triage, show understand, compare*) and *low-level* (from Amar et al. [8]: *retrieve value, filter, sort, compute derived value, find extremum, correlate, determine range, cluster, characterize distribution, find anomalies*).

The authors distinguish between *techniques* and *intents*, in that the former should never be considered as an ends in itself, but "*a means to support the user's information understanding*".

When characterizing *interactive visualization*, the authors distinguish between *high-level representation intents* (*depict, differentiate, identify, show outliers, compare*), *high-level interaction intents* (from Yi et al. [366]: *select, explore, reconfigure, encode, abstract, elaborate, filter, connect*), *low-level representation techniques* (*charts, graphs, networks, treemaps, parallel coordinates*) and *low-level interaction techniques* (*selection, brushing, dynamic query, pan, zoom*).

January 14, 2013 [R-69]: G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems*, 21(4):70–73, 2006. `http://dx.doi.org/10.1109/MIS.2006.75` [176].

Discusses *sensemaking*, misconceptions about *sensemaking*, and its distinguishing features. *Sensemaking* is about *understanding connections, anticipating trajectories, making predictions*, and *acting effectively (i.e., making decisions)*; it is highly iterative and has no clear start and end points.

We did not cite this work in Chapter 2 or Brehmer and Munzner [33], as we opted to cite Klein et al. [175] instead (discussed below), which contributes a classification.

January 14, 2013 [R-70]: G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems*, 21(5):88–92, 2006. `http://dx.doi.org/10.1109/MIS.2006.100` [175].

Contributes a high-level classification / model: the *data-frame* model of *sensemaking*, which involves a *elaboration cycle* and a *reframing cycle*.

The *elaboration cycle* involves *recognizing and constructing a frame*, *managing attention within a frame, defining, connecting*, and *filtering data*, *elaborating a frame* (*adding and filling slots, seeking and inferring new data and relationships, discarding data*), *questioning a frame* (*tracking anomalies, detecting inconsistencies, judging plausibility, gauging data quality*), and *preserving a frame*.

The *reframing cycle* involves *comparing* and *seeking new frames*.

The relationship between a data and a *frame* is one of *assimilating* data into a frame and *accommodating* frames via *questioning, rejecting*, and *comparing* of alternative frames. *Sensemaking* is a process by which multiple *frames* or hypotheses are considered.

January 15, 2013 [R-71]: B. Lee, P. Isenberg, N. Henry Riche, and S. Carpendale. Beyond mouse and keyboard: Expanding design considerations for information visualization interactions. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 18(12):2689–2698, 2012. `http://dx.doi.org/10.1109/TVCG.2012.204` [187].

A discussion of post-windows-icons-menus-pointer (WIMP) interfaces and the prospects for visualization tools and techniques. With respect to interaction in general, the authors characterize how an interaction begins with an *intent*, followed by an *action*, which triggers some *feedback* from the system and in return a *reaction* to that feedback.

They argue: "*InfoVis interaction taxonomies consider the system reaction or the resulting functionality part of the interaction timeline largely discuss interaction from the viewpoint of* *interaction tasks to be performed with and on the data, or from interaction techniques that can be performed on the data or its visual representation. We take a human-centric approach instead and do not specify data or tasks.*"

The authors correctly point out that many classifications of interaction techniques for data exploration "*are often mixed and discussed interchangeably with those related to an analyst's tasks with a visualization.*"

January 16, 2013: We continued to refine our meta-analysis of existing classifications, as represented in Figure A.10. These cross-cutting dimensions refer to terms defined by Chi and Riedl [58] (*temporal* vs. *atemporal*), Roth [260] (*objective*, *operator*, and *operand*, which is in turn influenced by Norman [226]'s *Stages of Action* model), Chuah and Roth [60] (*semantic*, *syntactic*, and *lexical*, Beaudouin-Lafon [16] (*descriptive*, *evaluative*, and *generative*), and Kirsh and Maglio [173] (*pragmatic* and *epistemic*).

January 16, 2013: I gave a pre-paper talk[9] about this project to the UBC InfoVis group, entitled *"Mid-level Task Abstractions for Visualization Design and Evaluation"*[10]. This talk summarized our meta-analysis and our proposed classification shown in Figure A.9.

The group found our work to be compelling, however there was some confusion regarding the vocabulary that we used (e.g., "`produce`" vs. "`present`" and "`explore`" vs. "`enjoy`"), and whether a consideration of *who* was warranted alongside *why*, *what*, *how*, and *which*. They asked whether our scope accounted for even higher levels of abstract tasks, such as *understanding the past*, *decision making in the present*, and *predicting the future*. They asked for more concrete examples to explain aspects of our classification. There was also a question as to whether *which* was necessary and if it could be subsumed into the questions of *how* and/or *what*. Some were

[9]For most papers written by members of the UBC InfoVis group, the lead author will present the paper to the group before it is written in the style of a conference presentation as a means to get feedback on the framing and contributions of the work; for more information, see `http://people.cs.ubc.ca/~tmm/policy.txt`.

[10]This was pre-paper talk was our fourth slide presentation, as explained in Section A.3.

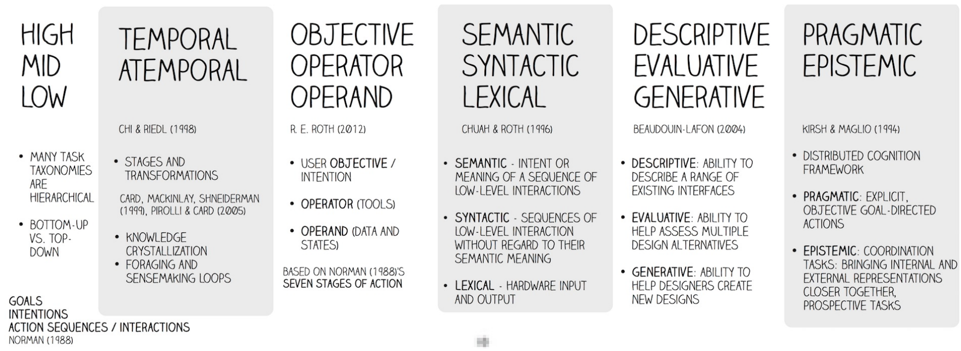


Figure A.10: January 16, 2013: Cross-cutting dimensions of previous classifications with dimensions using terms defined by Chi and Riedl [58], Roth [260] (which is in turn influenced by Norman [226]'s *Stages of Action* model), Chuah and Roth [60], Beaudouin-Lafon [16], and Kirsh and Maglio [173].

confused about of the dimensions of our meta-analysis, particularly the *semantic*, *syntactic*, and *lexical* dimensions inspired by Chuah and Roth [60]. The question of supporting evidence for our classification and our methodology also arose, indicating that we would need to thoroughly explain these aspects in subsequent presentations and in our eventual paper. Finally, there was the question of how to validate our classification, which was not something we had planned to address in the paper; we planned to treat our classification as a proposal, and that validation would follow in subsequent projects over time.

January 16 – February 2, 2013: We consolidated the feedback we received in response to our pre-paper talk and refined our ideas and intended contributions over the next two weeks, and in this time we generated three additional iterations of slide presentations, as explained in Section A.3. We also used this period to clarify our scope: we were not proposing an interaction taxonomy, nor were we proposing a temporal pipeline model, and we certainly were not planning to enumerate or classify visual encoding choices. With respect to *what*, we realized that we needed to distinguish *data* and *views*. Finally, we realized that too much of our framing was contingent upon a deep familiarity with previous work from our group, namely

Tamara's Nested Model [217].

January 25, 2013 [R-72]: W. Buxton. Chunking and phrasing and the design of human-computer dialogues. In *Proceedings of the IFIP World Computer Congress*, pages 475–480, 1986. `http://www.billbuxton.com/chunking.pdf` [44].

> Discusses interaction techniques designed around chunking single actions into compound actions as a means to facilitate the development of expert skills.
>
> Recommended by co-advisor Joanna McGrenere in response to our pre-paper talk and our discussion of "nested tasks". We did not cite this work in Chapter 2 or Brehmer and Munzner [33], as we recast nested tasks as sequences of interdependent tasks.

January 28, 2013 [R-73]: R. Kosara and J. D. Mackinlay. Storytelling: The next step for visualization. *IEEE Computer*, 46(5):44–50, 2013. `http://doi.ieeecomputersociety.org/10.1109/MC.2013.36` [177].

> Discusses the role and prospects of visualization in presentation and storytelling. The authors discuss presentation in collaborative and pedagogical contexts, and the way in which a presentation is given may vary according to the size of the audience, whether the presentation is live or pre-recorded, and whether the audience is co-located with the presenter.

February 1, 2013 [R-74]: S. K. Card and J. D. Mackinlay. The structure of the information visualization design space. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 92–99, 1997. `http://dx.doi.org/10.1109/INFVIS.1997.636792` [47].

> Presents a grammar or language for describing a mapping between abstract data types and common (visual encodings), distinguishing between data and *derived* or *transformed* data.

Transformations can include *filtering, sorting, MDS, interactive input*, and other functions.

This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33], as the authors do not explicitly discuss tasks but focus on the visual encoding design space, which we did not address.

February 1, 2013 [R-75]: R. Spence. *Information Visualization: Design for Interaction.* Prentice Hall, 2nd edition, 2007 [298].

Contributes a classification that distinguished between *interaction modes* (*continuous, stepped, passive*, and *composite interaction*) and *interaction spaces* (*continuous, discrete*).

Regarding intention, Spence says that a person could be *learning* (*exploring*), *seeking* (*finding*), or *interacting in an opportunistic or involuntary manner. Browsing* (*perusal*) can occur in any of these situations, and Spence refers to the perception and interpretation of content, including navigational cues.

Regarding interaction, Spence invokes Norman [226] and the *Stages of Action* model. He distinguishes some interaction as being related to *navigation*, *sensitivity*, *making dynamic queries*, and *evaluating residue or information scent*.

February 1, 2013 [R-76]: S. N. Friel, F. R. Curcio, and G. W. Bright. Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Research in Mathematics Education*, 32(2):124–158, 2001. `http://dx.doi.org/10.2307/749671` [102].

Makes compelling or noteworthy assertions about the behaviour of people who use graphs, distinguishing two high-level uses of graphs: *translation* (*locating*) and *interpretation / integrating* (involving *re-arranging, sorting, filtering*, and its *generative extensions interpolation* and *extrapolation*).

Discusses analogous concepts in the comprehension of text, discerning between *inference, application, synthesis*, and *evaluation*, as well as *identifying gaps, contradictions, incongruities, anomalies*, and *ambiguities*.

Included in the meta-analysis of graph comprehension theories by Pohl et al. [246], discussed above.

February 1, 2013 [R-77]: L. Wilkinson. *The Grammar of Graphics.* Springer, 2nd edition, 2005 [353].

Makes compelling or noteworthy assertions about the behaviour of people who use visualization tools or techniques, distinguishing between *building graphics* and *exploring interactive graphics*; the latter comprising of *filtering navigating* (*zooming, panning, lens*), *manipulating* (*node dragging, categorical reordering*), *brushing and linking*, *animating*, *rotating*, and *transforming.*

Wilkinson also comments on the value of classification: "*Classification for its own sake, however, is as unproductive in design as it is in science. In design, objects are only as useful as the system they support. And the test of a design is its ability to handle scenarios that include surprises, exceptions, and strategic reversals. [Classifications] may be useful for developing interfaces but contributes nothing to a deeper understanding of graphs. Customary usage and standards can blind us to the diversity of the graphics domain; a formal system can liberate us from conventional restrictions.*"

Presents a formal grammar for specifying graphics. This grammar is object-oriented, comprised of *sets, relations, functions, graphs, compositions, transformations, algebras, variables, varsets, frames.* It describes how to *create variables, apply algebra, apply scales, compute statistics, construct geometries, apply coordinates*, and *compute aesthetics.*

February 7, 2013 [R-78]: B. B. Bederson and B. Shneiderman. Theories for understanding information visualization. In *The Craft of Information Visualization: Readings and Reflections*, pages 349–351. Morgan-Kaufmann, 2003 [18].

> Discusses the purpose of theories for guiding visualization design and the *descriptive, explanatory, predictive, prescriptive*, and *generative* power of theories.

February 7, 2013 [R-79]: L. Tweedie. Characterizing interactive externalizations. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 375–382, 1997. `http://dx.doi.org/10.1145/258549.258803` [329].

> Contributes a low-level classification of *interaction types* (*manual, mechanized, instructable, steerable, and automatic*) as well as the *directness* of interaction (*direct* and *indirect manipulation*).
>
> Describes three different levels of questions regarding data (*objects* and *attributes*): a single item, a set of items, or the whole set.
>
> Tweedie highlights *input* and *output* relations, as "*this provides an externalization of the current state of the interaction*". which is important if a person is to engage in a dialogue with a tool.

February 7, 2013: Our fifth classification is represented in Figure A.11. Based on the discussion that followed the pre-paper talk, we consolidated *which* with *how* and established a mid-level classification within *how*, distinguishing between `encoding` data elements, `manipulating` elements, and `introducing` elements. The terms previously associated with *which* largely referred to `introducing` elements, and thus were placed here.

February 14, 2013 [R-80]: W. Stephenson. *The Play Theory of Mass Communication*. University of Chicago Press, 1967 [303].

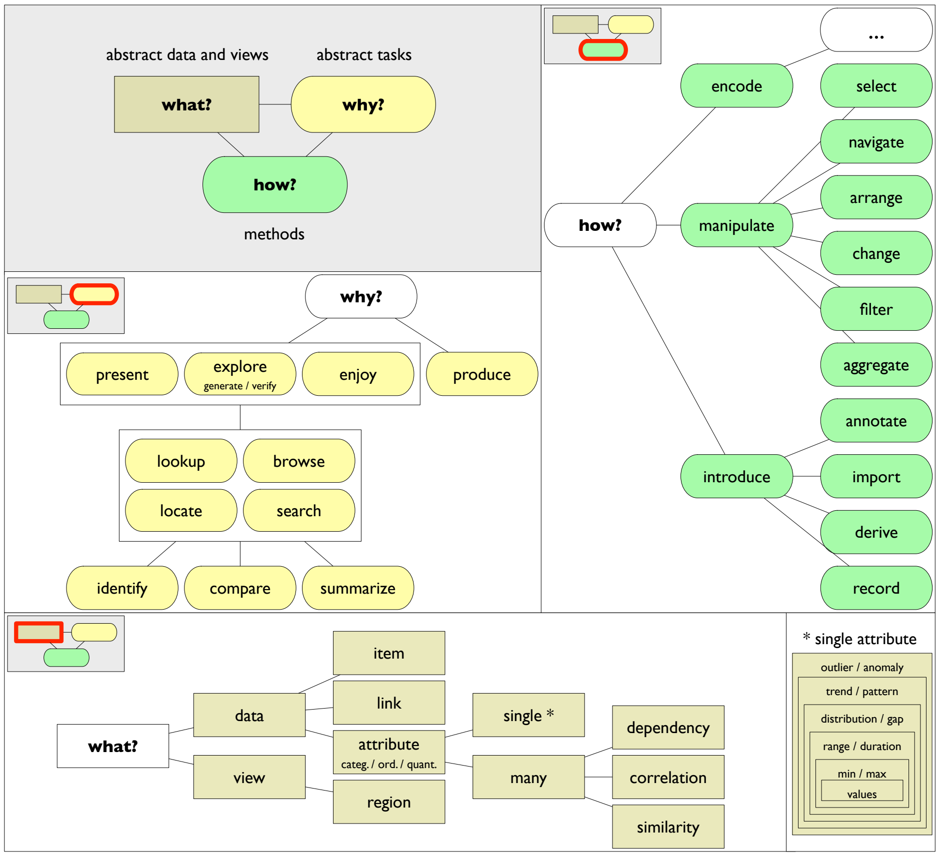


Figure A.11: February 13, 2013: Our fifth classification, where *why* corresponds to abstract tasks, *how* corresponds to methods, and *what* corresponds to abstract data and views; the colours now correspond to Munzner's nested model [217].

Makes compelling or noteworthy assertions about the behaviour of people who use communication artefacts (such as newspapers).

Describes *Play theory*[11], which accounts for media consump-

[11] Stephenson also discusses the word *play*, and how the English language conflates the many forms of play: *agon* (antagonistic play; e.g., football), *alea* (games of chance; e.g., lotteries), *mimicry* (acting, pretending), and *ilinx* (producing dizziness; e.g., swings and carousels). He also indicates the *ways of playing*: *paideia* (primitive, carefree gaiety and fantasy), *ludus* (formal play with rules and conventions, involving patience and development of skill), and *wan* (the quietly sensual Chinese way of playing; e.g., polishing jade).

tion activities that bring no material gain, serving no "work" functions, but instead induce moments of absorption and self-enchantment. Casual media consumption relies upon serendipitous *apperception*, a readiness to interact with information relating to existing interests. Stephenson argues that forms of mass media serve the purpose of mutual socialization, to give people "something to talk about", and that maximizing social interaction in the process of digesting mass media can be enjoyed for its own sake.

Stephenson points out that studies of newsreading behaviour indicated that people read most avidly what they already know about, a seemingly irrational activity that cannot be described as an explicit need to discover new information.

Cited by Case [50] and Toms [318], both discussed above.

February 14, 2013 [R-81]: Z. Liu and J. T. Stasko. Mental models, visual reasoning and interaction in information visualization: a top-down perspective. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 16(6):999–1008, 2010. `http://dx.doi.org/10.1109/TVCG.2010.177` [193].

A sequel to Liu et al. [194], considering distributed cognition, mental models, and the interplay between internal and external visualization. Contributes a high-level classification / model, describing how people *internalize* visualizations and engage in *mental simulation* with *mental models*. The dynamics between mental models and external visualizations include the *internalization of functional models*, *processing*, *augmentation*, and *creation* via *discovery* and *innovation*. People *construct* and *manipulate* mental models in working memory for *reasoning, anticipation*, and *planning*. Meanwhile, people interact with external visualization

Finally, play is distinguishable from work and represents a voluntary interlude directed by *convergent selectivity*, which concerns fads, manners, fashions, taste, etc.

artefacts for several reasons, including *external anchoring* (*projecting, locating*), *information foraging*, *restructuring* (*reconfiguring, encoding*), *exploring*, and *cognitive offloading* (*highlighting, arranging, creating, saving, loading*).

The authors indicate a need for a "*taxonomy of mental simulation*": the activities that people do internally in relation to external visualization.

A.5 Mid-Level Visualization Tasks

February 25, 2013: We added the following set of references added during the course of paper writing to support our arguments throughout Chapter 2; the exact dates corresponding to when these papers were read were not recorded. In addition, we added several references to visualization tools and techniques to illustrate aspects of the typology; these include the trellis display by Becker et al. [17], SpaceTree by Grosjean et al. [122], TreeJuxtaposer by Munzner et al. [220], the Table Lens by Rao and Card [251], Polaris (later Tableau) by Stolte et al. [304], Improvise by Weaver [348], and techniques for crowdsourcing data analysis by Willett et al. [355]. Finally, we also made reference to work by our group [210, 211, 217–219, 284].

February, 2013 [R-82]: A. Buja, D. Cook, and D. F. Swayne. Interactive high-dimensional data visualization. *Computational and Graphical Statistics*, 5(1):78–99, 1996. `http://dx.doi.org/10.1080/10618600.1996.10474696` [42].

Contributes a low-level classification, distinguishing between *focusing* (*choosing the projection or aspect ratio, zooming, panning, ordering, scaling, animation, rotation*), *linking* (*brushing as conditioning, sectioning, querying*), and *arranging views* (e.g., scatterplot matrix, conditional plot).

The authors also distinguish between *finding Gestalt*, *posing queries*, and *making comparisons*.

February, 2013 [R-83]: S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, 1999 [48].

> Contributes a high-level classification / model: describes the cyclic model of *knowledge crystallization*, a high-level hierarchy of tasks that occurs in a cycle. These tasks involve *ill-defined problem solving*, a *need to communicate or act upon results*, large amounts of *heterogeneous information*, a *well-defined goal* (e.g., *communication, decision-making*) requiring *insight*. The stages are as follows, with information visualization supporting the subtasks: *information foraging* (*browsing, querying*; see Pirolli and Card [242]; also involves the *overview, zoom and filter, details-on-demand* mantra of Shneiderman [291]), *search for or generate a schema, representation, abstraction* (*reorder, cluster, class, average, derive new data, promote, detect pattern*), *instantiate schema with data* (*reduce residue not fitting schema, improve schema*), *problem-solve to trade-off features* (*manipulate, create, delete*, and the *read fact, read comparison, read pattern* classification by Bertin [22]), *search for new schema that reduces the problem*, and *package the patterns found in some output*.

February, 2013 [R-84]: A. Dix and G. Ellis. Starting simple: Adding value to static visualization through simple interaction. In *Proceedings of the ACM Conference on Advanced Visual Interfaces (AVI)*, pages 124–134, 1998. `http://dx.doi.org/10.1145/948496.948514` [75].

> Contributes a low-level classification, distinguishing between *highlight & focus*, *accessing extra information*, *overview & context*, and *linking representations*. Also discusses instances involving the same representation but changing parameters, or instances involving the same data but a changing representation[12].

[12] A topic recently revisited by Kindlmann and Scheidegger [171] in their *Algebraic process for visualization design*.

February, 2013 [R-85]: D. A. Keim. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 8(1):1–8, 2002. `http://dx.doi.org/10.1109/2945.981847` [166].

> Contributes a low-level classification, distinguishing between *dynamic projection, filtering, zooming, distortion*, and *linking & brushing*.

February, 2013 [R-86]: W. Ribarsky, B. Fisher, and W. M. Pottenger. Science of analytical reasoning. *Information Visualization*, 8(4):254–262, 2009. `http://dx.doi.org/10.1057/ivs.2009.28` [257].

> Discusses the science of analytical reasoning and *sensemaking*, as well as three primary classes of objects: *stages*, *artefacts*, and *data tasks*. Proposes a hierarchical relationship with *sensemaking tasks* at the highest layer of abstraction, *artefacts* at the next lower layer, and *data tasks* at the lowest layer. The authors also discuss the possibility of inferring tasks from recorded interface interactions for the purpose of analytical provenance.
>
> This work was cited in our original submission but not in the Chapter 2 or the published version of Brehmer and Munzner [33], as we opted to cite earlier sensemaking work by Pirolli and Card [242] and Klein et al. [175].

February, 2013 [R-87]: M. Ward and J. Yang. Interaction spaces in data and information visualization. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 137–145, 2004. `http://dx.doi.org/10.2312/VisSym/VisSym04/137-146` [343].

> Contributes a low-level classification, distinguishing between *navigation, selection*, and *distortion*, as well as *interaction spaces* (*screen-space, data value-spaces, data structure-space, attribute-space, object-space, and visualization structure-space*), and *interaction parameters* (*focus, extents, transformation, blender*).

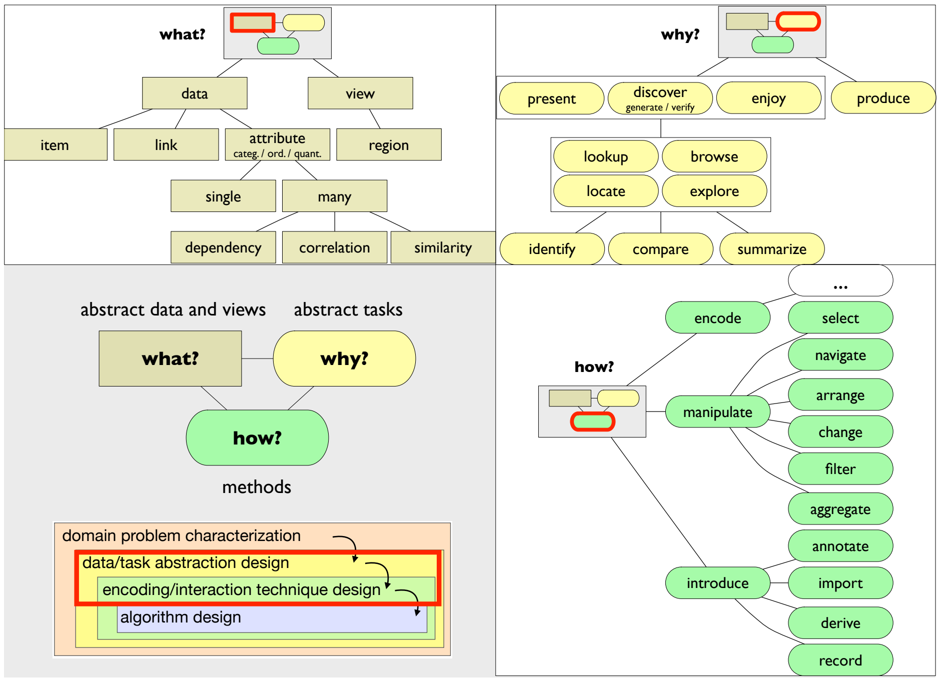


Figure A.12: February 25, 2013: Our sixth classification, which simplifies our characterization of single attributes and makes the connection to Munzner's nested model [217] more explicit.

February 25, 2013: Our sixth classification is represented in Figure A.12; this classification was included in a paper draft entitled *"Mid-Level Tasks for Visualization Design and Evaluation"*, which we circulated to members of the UBC InfoVis group to invite discussion and criticism. This version simplified our classification of *what* and made the connection to Munzner's nested model [217] more explicit.

March 4, 2013: The UBC InfoVis group along with visitor Colin Ware met to discuss our draft: *"Mid-Level Tasks for Visualization Design and Evaluation"*.

The group's critique included: (i) they disputed the term *mid-level* since we had multiple levels in our classification, and though the *why-what-how* division was appreciated (with the strengths being *why* and *how*), our classification of *what* was disputed and was thought to be underdeveloped; (ii) they acknowledged our abundance of motivation but suggested that many readers may not appreciate this motivation without additional context (beyond the Nested Model [217]), running examples, and use cases; (iii) our claims of *evaluative* and *generative* power were disputed, as were our claims regarding that our work was *actionable* and *persuasive*; (iv) some of our task description examples were expressed with a pseudocode-like notation, which was deemed to be confusing, and it was similarly it was unclear as to why we presented the terminology of related work in a table (referring to Table 2.1 and Table 2.2); (v) finally, they asked us to be more explicit with regards to what was gained with the advent of our classification and why ours was superior to all prior classifications.

A.6 Our Proposed Taxonomy of Tasks

March 13, 2013: Our seventh classification is shown in Figure A.13; this classification is identical to the typology as it is presented in Chapter 2 (reproduced in Figure A.14 below) in all aspects save its visual layout, which includes an illustration of a sequence of interdependent tasks. In response to feedback from UBC InfoVis group readers and to our recent reading of Tweedie [329], we drastically simplified our classification of *what* to an even greater extent, reducing it to `input` and `output`. We also started to refer to our classification as a *multi-level taxonomy* as opposed to one classifying *mid-level tasks*. We emphasized our taxonomy as resulting from a combination of literature review and new thinking. We cut back on our discussion of *evaluative* and *generative* potential, emphasizing the *descriptive* power of our classification. We also dropped the discussion of whether our classification was *actionable* or *persuasive* [111].

March 31, 2013: Submission of *"A Multi-Level Taxonomy of Abstract Visualization Tasks"* to IEEE Information Visualization.

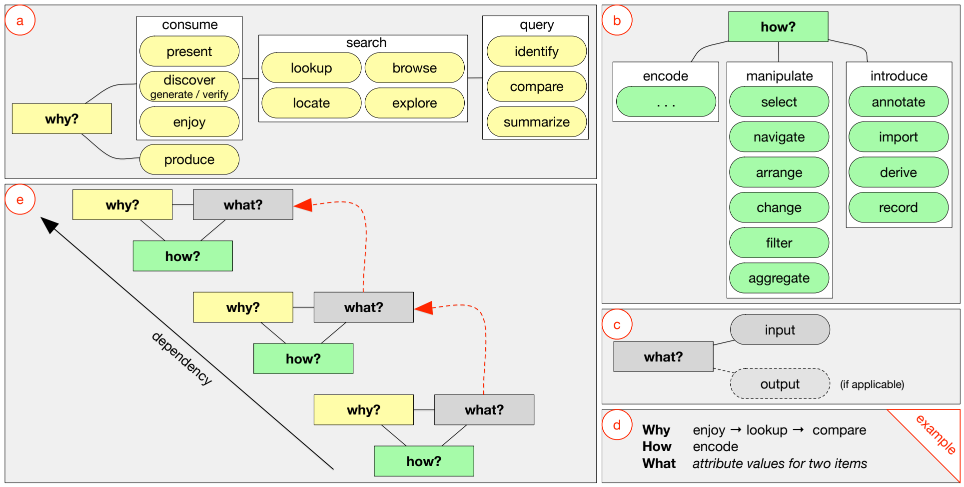


Figure A.13: March 13, 2013: Our seventh and proposed classification, including a simplification of *what*: (a) *why* a task is performed; (b) *how* the task is supported; (c) *what* are the `inputs` and `outputs` of the task; (d) an example task description; (e) a sequence of interdependent tasks.

A.7 Revisions: From Taxonomy to Typology

June 6, 2013: *"A Multi-Level Taxonomy of Abstract Visualization Tasks"* was conditionally accepted to IEEE Information Visualization, and we revisited our literature review in response to the anonymous reviews.

June 10, 2013 [R-88]: W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, 2011 [2].

> Recommended by an anonymous reviewer and makes compelling or noteworthy assertions about the behaviour of people who use time-oriented visualization tools or techniques.
>
> Includes an analysis framework centred around *why*, *what*, and *how*, asking *what is presented?*, *why is it presented?*, and *how is it presented?* (implying a focus on visual encodings).
>
> Also includes *time-oriented primitives* that we cite in reference to *what*, which includes *points, intervals, spans, temporal patterns, rates of change, sequences*, and *synchronization*.

They also refer to classifications by Andrienko and Andrienko [12], Spence [298], and Yi et al. [366]. The authors to higher-level analysis tasks to which visualization of time-oriented data contributes to, which includes *classification, clustering, search, retrieval, pattern discovery*, and *prediction*.

June 10, 2013 [R-89]: C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 3rd edition, 2012 [345].

In addition to the content carried over from the 2nd edition (Ware [344], see Section A.2), Ware contributes a high-level classification, adding ten *visual thinking algorithms* in the 3rd edition, which include *visual queries*, *pathfinding on a map or diagram*, *reasoning with a hybrid of a visual display and mental imagery*, *design sketching*, *brushing*, *small pattern comparisons in a large information space*, *degree-of-relevance highlighting*, *generalized fisheye views*, *multidimensional dynamic queries with scatterplot*, and *visual monitoring strategies*.

June 10, 2013 [R-90]: G. Andrienko and N. Andrienko. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer-Verlag, 2006 [12].

Recommended by an anonymous reviewer and contributes a low-level classification that distinguishes between *elementary* and *synoptic* tasks (inspired by the *elementary, intermediate*, and *overall levels of reading* described by Bertin [22]), as well as *references* and *characteristics*.

Elementary tasks include *direct lookup* (*identification*), *inverse lookup* (*localization*), *direct comparison* (*identification and interrelation of characteristics*), *inverse comparison* (*localization and interrelation of references*), and *relation seeking*.

Synoptic tasks include *descriptive tasks* (*direct lookup / pattern definition and identification, inverse lookup / pattern search* *and localization, direct pattern comparison, inverse pattern comparison, relation seeking*) and *connectional tasks* (*identify heterogeneous behaviour, identify homogeneous behaviour*).

Includes a formal quasi-algebraic notation for specifying these tasks in terms of *referents* (*points*) and *characteristics* (*attributes*).

The authors define a task as having two parts: a *target* (*what information needs to be obtained*) and *constraints* (*what conditions this information needs to fulfil*). Targets for synoptic tasks are types of patterns or configurations of *characteristics* for a *reference set*; they include: *association, differentiation, arrangement*, and *distribution summary*.

June 10, 2013 [R-91]: T. Lammarsch, A. Rind, W. Aigner, and S. Miksch. Developing an extended task framework for exploratory data analysis along the structure of time. In *Proceedings of the Eurographics International Workshop on Visual Analytics (EuroVA)*, 2012. `http://dx.doi.org/10.2312/PE/EuroVAST/EuroVA12/031-035` [185].

Recommended by an anonymous reviewer for connecting the *what* to the *how*; contributes a datatype specific classification of tasks relating to time-oriented data, extending the top-down task classification for spatial and temporal data by Andrienko and Andrienko [12], introducing a formal notation for specifying these tasks and incorporating the time-oriented primitives introduced by Aigner et al. [2]: *scale, scope, arrangement, viewpoints* (*ordered* vs. *branching time*), *granularities*, and *determinancy*.

June 10, 2013 [R-92]: J. Raskin. *The Humane Interface: New Directions for Designing Interactive Systems*. Addison-Wesley Longman, 2000 [252].

Recommended by an anonymous reviewer in response to the `manipulate` nodes in our *how* classification, and contributes a low-level classification, a *taxonomy of elementary actions* that might be performed with a mouse and keyboard. Raskin's classification is addressed at an HCI readership, though his classification applies to many visualization tools.

Rasking distinguishes between *indicating* (*pointing*), *selecting, activating*, and *modifying* (*or using: generating, deleting, moving, transforming, copying*).

June 13, 2013 [R-93]: M. Dörk, S. Carpendale, and C. Williamson. The information flaneur: A fresh look at information seeking. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1215–1224, 2011. `http://dx.doi.org/10.1145/1978942.1979124` [77].

Makes compelling or noteworthy assertions about the behaviour of people who use visualization artefacts in casual contexts, including curiosity-driven tasks without expectations or predicted outcomes, where novelty stimulates curiosity and thereby exploration.

Added in response to an anonymous reviewer who indicated that our term `enjoy` was somewhat vague.

June 13, 2013 [R-94]: M. Dörk, N. Henry Riche, G. Ramos, and S. Dumais. PivotPaths: Strolling through faceted information spaces. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 18(12):2709–2718, 2012. `http://dx.doi.org/10.1109/TVCG.2012.252` [78].

Makes compelling or noteworthy assertions about the behaviour of people who use visualization artefacts in casual contexts in which the information being visualized is simply enjoyed; introduces the concept of *strolling* through information spaces, and discusses the characteristics of *browsing* and *searching*.

Added in response to an anonymous reviewer who indicated that our term `enjoy` was somewhat vague.

June, 2013 [R-95]: Z. Pousman, J. T. Stasko, and M. Mateas. Casual information visualization: depictions of data in everyday life. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis)*, 13(6):1145–1152, 2007. `http://dx.doi.org/10.1109/TVCG.2007.70541` [247].

> Makes compelling or noteworthy assertions about the behaviour of people who use visualization artefacts in casual contexts in which the information being visualized is simply enjoyed, including immersive and time-consuming experiences, such as in museum settings.
>
> Added in response to an anonymous reviewer who indicated that our term `enjoy` was somewhat vague.

June 18, 2013 [R-96]: K. B. Smith. Typologies, taxonomies, and the benefits of policy classification. *Policy Studies Journal*, 30(3):379–395, 2002. `http://dx.doi.org/10.1111/j.1541-0072.2002.tb02153.x` [297].

> Smith indicates that "*the key characteristic of a typology is that its dimensions represent concepts rather than empirical cases [and that] typologies create useful heuristics and provide a systematic basis for comparison ... Their central drawbacks are categories that are neither exhaustive nor mutually exclusive.*" Taxonomies, on the other hand, "*classify items on the basis of empirically observable and measurable characteristics.*"
>
> Considered in response an anonymous reviewer who questioned whether we had truly proposed a *taxonomy*, implying completeness, indicating that others have use the less strict *typology* in this context; we did not cite this work in Chapter 2 or Brehmer and Munzner [33], opting to cite the original source Bailey [13] instead.

June 18, 2013 [R-97]: K. D. Bailey. *Typologies and Taxonomies: An Introduction to Classification*. Sage Publications, 1994 [13].

Bailey indicates that a *typology* is appropriate for classifying abstract concepts, while a *taxonomy* is appropriate for classifying empirically observable events and are often but not always hierarchical.

Added in response an anonymous reviewer who questioned whether we had proposed a *taxonomy*, or a *typology*.

June 19–27, 2013 [R-98]: K. J. Vicente. *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work.* CRC Press, 1999 [337].

Recommended by an anonymous reviewer who stated that "*The authors will enjoy it*"[13].

Describes *cognitive work analysis*, a framework that combines *constraint-based task analysis* with *work domain analysis* as the "*foundation for a holistic and socio-technical account of a person's work involving information technology*", which involves five phases: *work domain analysis, control task analysis, strategies analysis, social organization and cooperation analysis*, and *worker competencies analysis.*

Vicente contrasts *normative work analysis* (how work *should* be done), *descriptive work analysis* (how work *is* done, including workarounds and strategies outside of what is prescribed by the normative account of the work), and *cognitive work analysis* as a form of *formative work analysis* (how work *could* be done, subject to known constraints, such as *inputs* and *outputs*, social organization, and worker competencies).

The *strategies analysis* phase involves describing *how* a task is carried out independent of who does it, still in a device-agnostic fashion, examining the different possible strategies by which the task could be executed.

Like us, Vicente appreciates the power and simplicity of asking *why, what*, and *how* for elucidating *means* and *ends*, though

[13] I did enjoy it; thanks R3!

he is referring to the *structural means-ends* relations in a work domain's *abstraction hierarchy*, while we are concerned with *action means-ends relations* of tasks; in other words, relations between nouns vs. relations between verbs. Finally, Vicente argues that sequence-based approaches to task analysis are overly rigid and thus inappropriate for describing such open-ended tasks; *constraint*-based approaches to task analysis allow for more flexibility in terms of *how* a task is performed.

June 27, 2013: Submission of a revision to IEEE Information Visualization, which we renamed *"A Multi-Level Typology of Abstract Visualization Tasks"*. In addition to opting to use the term *typology* and including the recommended references, we extended the definition for each node of our typology, now including one or more succinct examples, and we also made an effort to disambiguate the terminology flagged by reviewers. We added a section providing background context (Section 2.2) to outline the limitations faced by practitioners conducting task analysis, that previous classifications cannot be used to generate concise and abstract task descriptions needed for design and evaluation. Finally, we extended our discussion of *what* to include several illustrative example classifications of *what* as proposed by previous work (see Section 2.3.3).

A.8 Presenting our Typology

July 11, 2013: *"A Multi-Level Typology of Abstract Visualization Tasks"* was accepted to IEEE Information Visualization.

August 1, 2013: We submitted a camera-ready version of *"A Multi-Level Typology of Abstract Visualization Tasks"* to IEEE Information Visualization. In this revision, we responded to the meta-reviewer's comments regarding the presentation of the typology in Figure A.14, clarifying the differences between `discover` types *generate hypotheses* and *verify hypotheses*, the *means* and *ends* of `produce`, the scope of `change`, and the difference between `annotate` (a verb: *how*) and an annotation (a noun: *what*). We

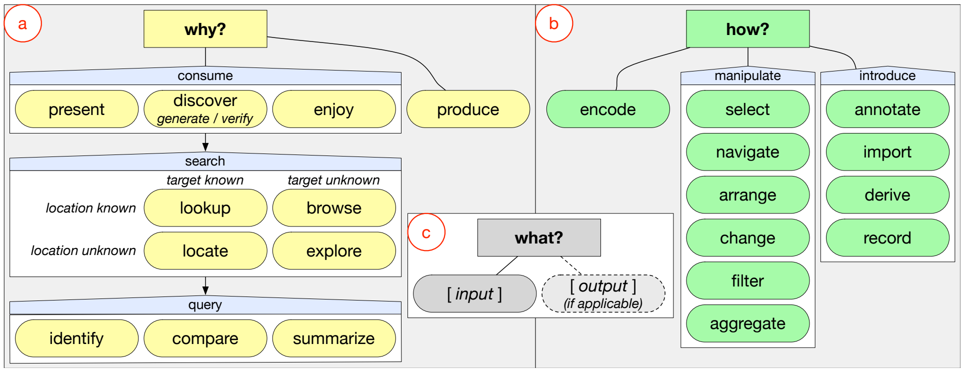


Figure A.14: October 13–18, 2013: Our multi-level typology of abstract visualization tasks as it was presented in Chapter 2 and at IEEE VIS 2013.

also highlighted terms from previous work relating to *what* in Table 2.1 and Table 2.2 using parentheses.

October 13–18, 2013: I attended the 2013 IEEE VIS Conference and presented *"A Multi-Level Typology of Abstract Visualization Tasks"* in the "Defining the design space" InfoVis session, which also included task classification papers by Schulz et al. [280] and Roth [262]. Roth's paper was a longer journal version of his 2012 *GIScience* workshop paper discussed above [260]; his classification remain unchanged. I comment upon the differences between our typology and "*A Design Space of Visualization Tasks*" by Schulz et al. [280] in Section 6.1.2. I presented the typology as it is shown in Chapter 2 and in Figure A.14; we omitted the illustration of a sequence of interdependent tasks due to space constraints. Finally, Figure A.15 is one aspect of our meta-analysis of existing classifications that I presented at IEEE VIS 2013.

A.9 Subsequent Evolution of our Typology

Summer 2014: Tamara continued to iterate on the typology as she completed her book [219]. The version of the typology that appears in her book is reproduced in Section 6.1, where I also indicate and reflect upon the dif-

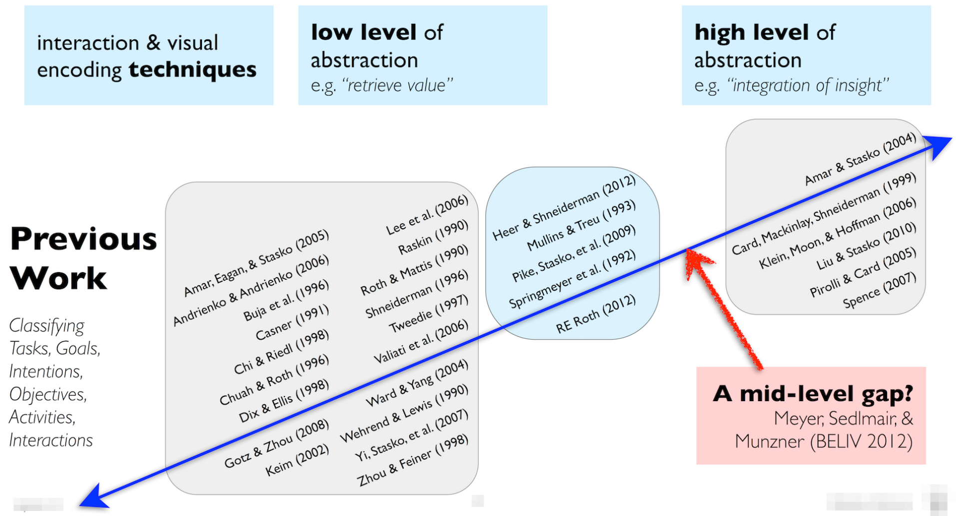


Figure A.15: October 13–18, 2013: Previous classifications sorted from low to high level of abstraction, highlighting a mid-level gap; one aspect of our meta-analysis of existing classifications that we presented at IEEE VIS 2013.

ferences between the version of the typology presented at IEEE InfoVis 2013 and the version in her book.

Appendix B

Interview Study Supplemental Material

This appendix supports Chapter 3. It contains a summary of all twenty-four interviews with nineteen analysts (nine of whom were excluded from the final analysis presented in Chapter 3), our interview questions, our methodological choices, example artefacts from our data analysis process, and an unpublished manuscript that preceded the re-interpretation of our findings presented in Chapter 3.

B.1 Complete List of Interviews

Table B.1 summarizes the twenty-four interviews that we conducted with nineteen analysts. Analysts 11-19 (below the midline in Table B.1) were excluded from the analysis presented in Chapter 3 because they did not use dimensional synthesis techniques and/or did not visualize their data as part of their analysis.

B.2 Interview Foci and Questions

Data (and data analysis):

- What does your data look like?

Analyst	Domain	Date / Duration / Location	Interviewers[1]	Artefacts[2]
A1 (Music)	HCI	2011-01-22 / 0.5h / phone	ms	p [15], m, m, t [43]
A2 (Search)	HCI	2011-01-26 / 1.5h / phone	ms,tm	m, e
A3 (BoatAct)	policy analysis	2011-01-20 / 1.5h / phone	ms,si	m, d, e, v
A4 (EpiGen)	bioinformatics	2010-11-16 / 2h / remote 2010-12-21 / 2h / remote	ms,hy,tmö ms,hy,tmö	p [223], s, s
A5 (Polymers)	chemistry	2010-11-18 / 1h / phone	ms,si,tm	e
A6 (Concept)	social networks	2010-11-18 / 1h / phone	ms,si,tm	v, e
A7 (MoCap)	HCI	2011-11-23 / 1h / ubc	ms,mb	p [6, 328], s
A8 (ProstCan)	bioinformatics	2011-04-04 / 1.5h / remote 2011-04-29 / 4h / remote	ms ms,si,tm	p [270], m
A9 (SeqAln)	bioinformatics	2010-04-20 / 1.5h / phone	ms,si,tm	p [26, 71, 72] [95, 136, 198] [289, 341], e
A10 (TxtDocs)	journalism	2012-03-05 / 3h / ubc	ms,mb,tm	[35], m, w
A11 (FishPop)	fisheries sciences	2010-09-17 / 2h / remote	ms,si	p [29, 138], m
A12 (CompVis)	computer vision	2010-10-06 / 1h / remote	ms,hy,tmö	
A13 (ChemRel)	chemistry	2010-10-07 / 1h / phone	ms,si,tm	
A14 (CompBio)	bioinformatics	2010-10-13 / 1h / remote	ms,si	
A15 (NPAlgo)	machine learning	2010-12-01 / 1h / phone	ms,tm	p [143–145, 189]
A16 (GamMdl)	machine learning	2010-12-07 / 2h / ubc 2011-04-07 / 0.5h / ubc	ms ms	p [362], m
A17 (FlockSim)	mathematics	2011-03-18 / 1.5h / remote 2011-04-05 / 1h / remote	ms,sb ms,sb,tmö	p [40, 84, 85, 96], t [1, 83]
A18 (MedImg)	computer vision	2011-04-05 / 1.5h / remote 2011-04-15 / 1.5h / remote	ms,sb ms,sb	p [11, 266, 267], m p [268, 320]
A19 (StrucGen)	bioinformatics	2011-04-20 / 1h / phone	ms,si,tm	p [108], v

Table B.1: The complete set of twenty-four interviews with nineteen analysts, including those excluded from our subsequent analysis (below the midline). [1]Michael Sedlmair (ms), Matthew Brehmer (mb), Stephen Ingram (si), Tamara Munzner (tm), Hamidreza Younesy (hy), Steven Bergner (sb), and Torsten Möller (tmö). [2]Artefacts from analysts: articles published or referred to us by interviewees (p), unpublished manuscript (m), data (d), presentation slides (s), thesis / dissertation (t), screenshots (v), additional email correspondence (e), web site / blog (w).

- One dataset, more datasets?
- What are the major problems, challenges in the data analysis?
- Which information in the data is important for you / what do you read from the data?
- What else do you want to read from the data?

Tasks and goals:

- What are you doing?
- What are you working on?
- What are your goals?
- What is the ultimate goal?
- What data analysis tasks are involved in your work?
- How important is data analysis in your daily work?
- What other tasks apart from data analysis?
- Collaboration or alone?
- What are the questions/hypotheses you try to answer by analyzing your data?

Current practices and tools; problems and challenges:

- What are the current tools you use for data analysis?
- What visualization techniques are you currently using?
- How is your procedure in analyzing the data with these tools (hypotheses)?
- Good things / bad things about these tools?
- What are you missing with these tools? What is the a perfect analysis tool?

Dimensionality reduction:

- Do you use DR in your work?
- If not yet, why do you think it is important for you?
- What are your expectations?

Patterns of Interest:

- Clusters
- Outliers
- Correlation between dimensions (between axis, should be rare after DR)
- Finding meaningful low-dimensional axes

B.3 Data Collection, Analysis, and Abstraction

In Chapter 3, we applied our task typology to the analysis of our interview findings. However, this was not the first attempt to analyze this data[1]; we initially analyzed our findings using a bottom-up approach in the spirit of grounded theory [55], in which we did not impose a previous task classification.

Our data analysis involved iteratively *coding* our collected data. For an extended discussion of the method of iterative coding, see both the textbook by Charmaz [55] and the paper on using grounded theory methods within an HCI context by Furniss et al. [107]. An extended discussion on the use of grounded theory methods within a visualization context is provided by Isenberg et al. [153], and the subject is also discussed by Carpendale [49]. Initial coding is a process for identifying themes and concepts; the code set is constructed iteratively. It is followed by focused coding, the grouping of codes with conceptual relationships, guided by our background knowledge and insights accumulated from previous data analysis. Our first classification of tasks [283] was the result of an iterative process of focused coding, a process of arranging our initial codes based on their importance and representativeness, into hierarchical relationships.

Another outcome of our data analysis was a set of *usage examples*, which refer to specific combinations of analysts' tasks with DR and visualization techniques, documented by rows in Table 3.1, Table B.6, and Table B.7.

[1] Our first attempt is documented by Sedlmair et al. [283], our second attempt is documented in Section B.6.

The classification of task sequences presented in Chapter 3 is the result of a new task abstraction analysis, one in which we used our task typology introduced in Chapter 2 as a code set. Our earlier attempts to classify tasks related to visualizing dimensionally reduced data are represented in Figure B.6 and Figure B.9. Unlike the analysis presented in Chapter 3, these previous classifications did not focus solely on the union of visualization and dimensionally reduced data; these classifications are summarized in Sedlmair et al. [283], as well as below in Section B.5 and in Section B.6.

B.4 Data Analysis Artefact Examples

We provide examples from our process of data collection and from our initial data analysis (reported in the original technical report [283]) The data that we gathered included interview transcripts (e.g., Figure B.1), notes (e.g., Figure B.2), publications, and other documents from our interviewees (e.g., Figure B.3), which we initially coded to identify concepts. We then organized the concepts that we identified around the properties and dimensions represented in summaries produced for each analyst (e.g., Figure B.4, Figure B.5). Tables B.2, B.3, B.4, and B.5 represent the result of an iterative process of focused coding to establish conceptual relationships. This process led to the development of our initial task classification, as described in our technical report [283].

B.5 Previous Interpretations of Findings

The initial interpretation of our findings [283] (summarized in Figure B.6) classifies people who use DR, DR techniques, tasks, and the challenges faced by these people. Unlike the analysis presented in Chapter 3, we did not exclude cases in which visualization did not occur, nor did we exclude cases in which *dimensional filtering* techniques (e.g., [159, 365]) were used[2]. Our findings are summarized according to this classification in Table B.6. A

[2]In Chapter 3, we concentrate on tasks relating to visualizing the results of *dimensional synthesis* techniques, such as PCA or MDS.

sequences, then additional information like fields that can superimpose. ultimate challenge is to learn how the structure relates to the genomes.

me: functional

functional information. all the information you measure or compute about experimental result. 3D structur eitself, various ways that can be represented in less atomistic forms. properties taht can compute or associate with surface.

binding cavity or cleft, for which small molecule or another part of the proetin could be substrate or site of information. those are things where once we have a 3D structure we can create a quant model. 3D structure becomes a quant model.

then question of creating biological hypotheses - given that kind of shape, we want to look at how many observations do we have of that shape or similar shapes that have epxerimentally exhibited a particular biological function.

maybe protein undergoes or facilities bio function of a particular type. if you want to start exploring and eo associations between function and shape, fo rinstance, or properties that may help to explain function that are mapped onto shape, like elctrostatic potential, then that's the kinds of simple view of a particular problem.

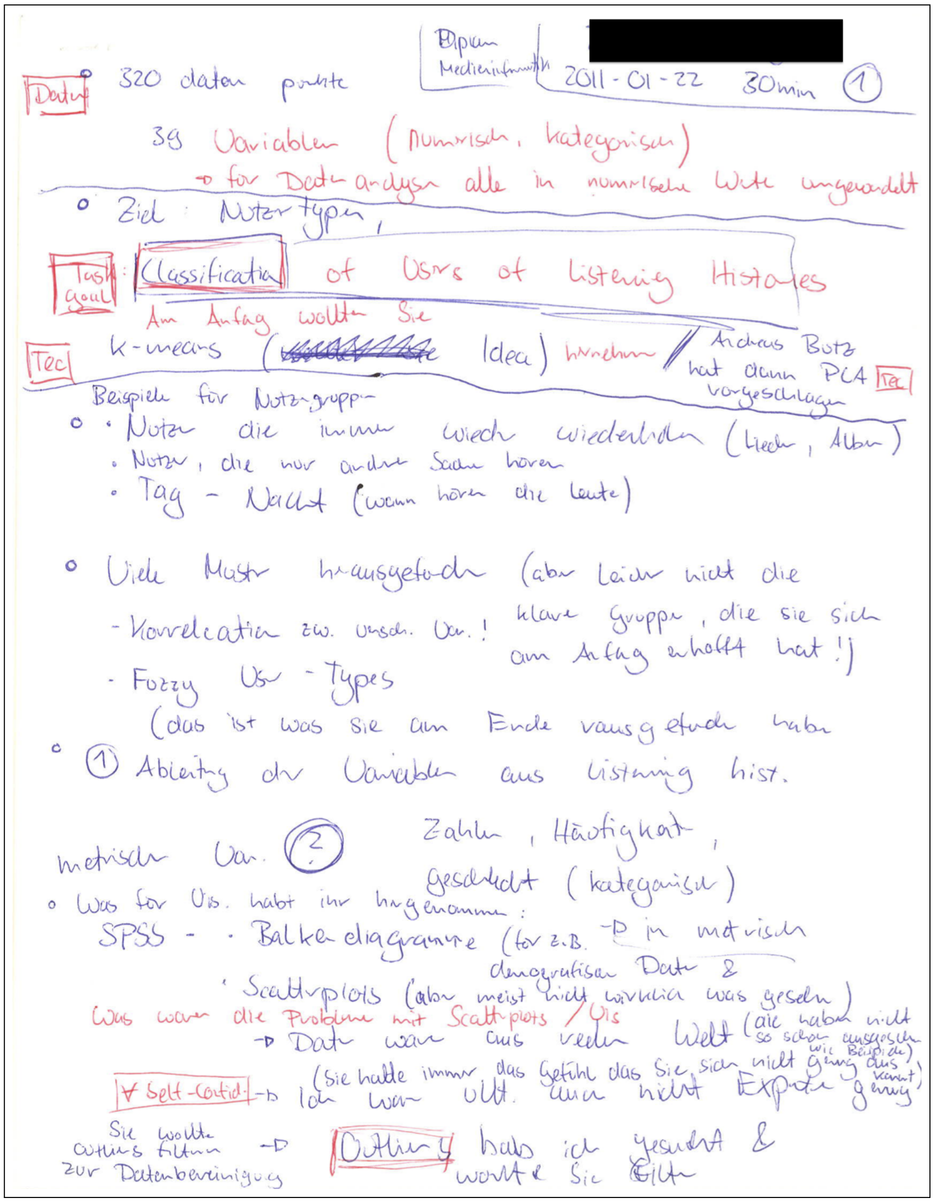
to give practical example, very large project protein structure initiative set out to solve 3D structures at a genomic scale. the way the human genome project tried to sequence the human genome. in a relatively uncharacteristic way for structural biology, they solved structures where did not know bi9ochemical function of them. normally it's hypothesis driven, where people try to understand function by looking at 3D structure dat.a but in this case idea was to solve structures that cover lots of sequence space, then to back and analyze late.r so have lots of structures where don't know the function.

computational task to find something you can see or compare about structure or one of its reduced forms where might have a clue about bio context. by looking at thingss wh

functional stuff. is it just out of GO ontology? or do you have your own?

GO describes biochemical functions. normally this particular protein is involved in a pathway that performs something related to cell

Figure B.1: An example of an interview transcript (Analyst A19).



Diplom Medieninformatik [redacted] 2011-01-22 30min 1

Daten 320 daten punkte

39 Variablen (numerisch, kategorisch)

-> für Datenanalyse alle in numerische Werte umgewandelt

o Ziel: Nutzertypen,

Task goal: Classification of Users of Listening Histories

Am Anfang wollten Sie

Tec k-means (~~...~~ Idea) heranziehen / Andreas Butz hat dann PCA vorgeschlagen Tec

Beispiele für Nutzergruppen

o • Nutzer die immer wieder wiederholen (Lieder, Alben)

• Nutzer, die nur andere Sachen hören

• Tag – Nacht (wann hören die Leute)

o Viele Muster herausgefunden (aber leider nicht die klare Gruppe, die sie sich am Anfang erhofft hat!)

- Korrelation zw. unsch. Var.!

- Fuzzy User-Types

(das ist was sie am Ende rausgefunden haben

o 1 Ableitung der Variablen aus Listening hist.

metrische Var. 2 Zahlen, Häufigkeit, Geschlecht (kategorisch)

o Was für Vis. habt ihr hergenommen:

SPSS - • Balkendiagramme (für z.B. -> in metrisch demographische Daten &

• Scatterplots (aber meist nicht wirklich was gesehen)

Was waren die Probleme mit Scatterplots / Vis

-> Daten waren aus vielen Welt (sie haben nicht so schön ausgesehen wie Beispiele)

∀ Self-Confid -> Ich war vllt. auch nicht Experte genug (sie hatte immer das Gefühl das Sie sich nicht genug aus kannt)

Sie wollte outliers filtern zur Datenbereinigung -> Outliers habe ich gesucht & wollte sie filtern

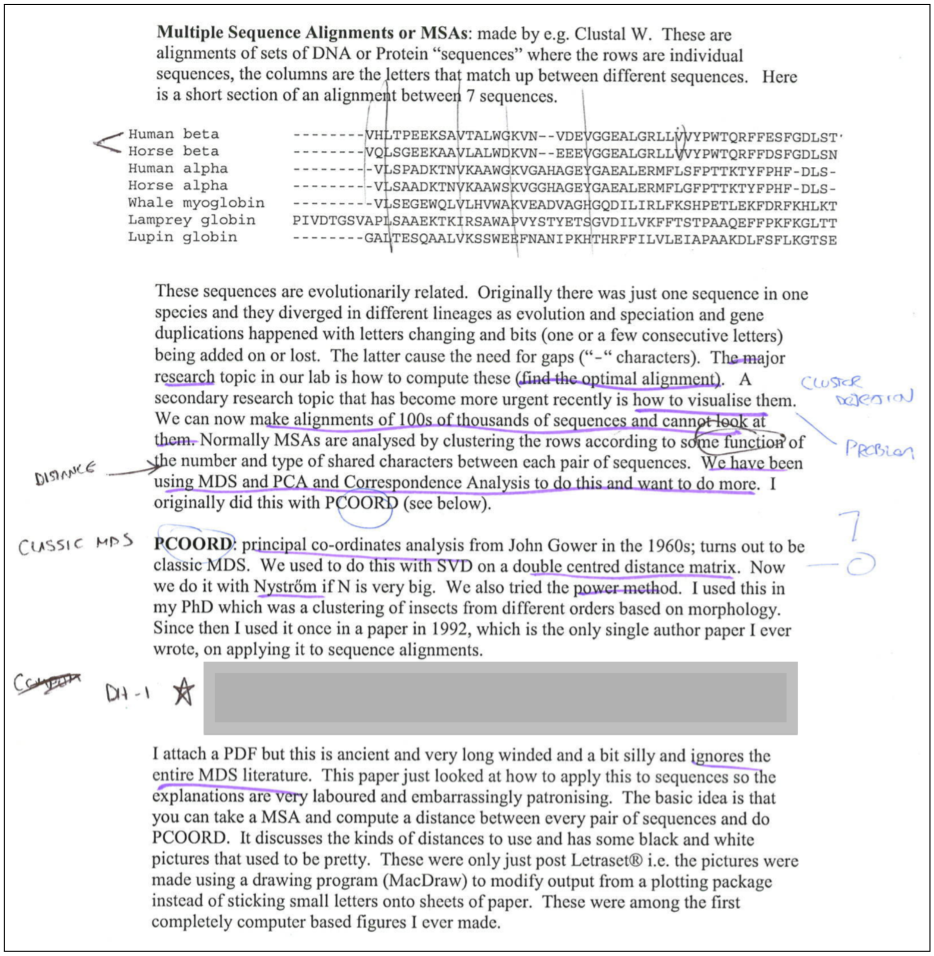
Figure B.2: Raw interview notes (blue ink) for Analyst A1, with post-hoc open coding (red ink) to identify concepts.

Short description of current project	Domain	Experience with DR	Type of user	Data	N	Dim	Task	Pattern of interest	Metrics	Current Analysis Techniques	Current Problems / Challenges
Evaluating different fishing simulation models in terms of harvest control rule	Biology - Fisheries	No	Analyst	Models tested with different parameterization	?	5 to 10	- Evaluating HCR - Finding best HCR	- Correlations between dimensions - Correlations in general	- Insight: Level of extirpation	- R - Line Graphs	- Missing Overview - Tedious interaction
Helping SpinPro Designers to reuse their images.	Vision	Yes	Researcher / Tool designer	Image database (design galleries)	?	25	- Developing new search algorithms - Debugging algorithms	- Clusters / Groupings	- Insight: What is the right algorithm - Insight: Model improvement/ optimization	- Matlab	
?? No explicit examples	Computational chemistry	Yes	Analyst	e.g. compounds with features represented in fingerprints	?	?	- Analyzing data for somebody else - Communicating findings	?	- Finding cause/effect relationships - Communicating cause/effect relationships	- MDS, correlation matrices - Pipeline Plot / Knime - Spotfire / Tableau (rare)	- Sparse data - Missing guidance - Is information really true? - How to trace back data analysis to build better compounds...
Finding better clustering methods for biologists	Biology - Cancer / HIV	Yes	Analysts / Algo designers	Cancer / HIV / FlowCap	20.000	5 to 20	- Comparing clustering algorithms (differences/uniques) - Compare automatic clustering vs manuel results - Developing automatic clustering techniques	- Clusters / Classifications	- Better Analysis - Faster clustering - More accurate clustering	- R - C for speed - Heatmaps - Scatterplots - By biologists: FlowJo tool	- Scalability Communication with biologists - Misslabels (dirty data) - PCA removes important information - PCA hard to interpret for biologists - Interactive clustering with visualization would be great
Recreational boating activity	coastal and ocean resources	little / no	researcher	survey data (ordinal and ratio) from recreational boaters	543 survey respondents	39 survey questions (mix or ordinal (Likert scale) and ratio (i.e. boat length)	- find meaningful clusters of types of recreational boaters - characterize boaters	clusters correlations b/w dims factors that characterize boaters (meaningful new dims)	- insight into clusters	- PCA - MDS - SPLOMs - k-means clustering - focal point clustering	- missing guidance - stability of clusters - the "blob" - dirty data
Epidemics. Providing a better overview over genomes.	Biology - Genomes	Little	Researcher / Tool designer	Heatmaps of genome parts	10.000 - 100.000	160 (8x20)	- Understanding problem - Implementing new tool	- Clusters	- Insight: Getting Overview		- The manual sorting of the rows in the heatmap
Visualizing research concepts in life science	Life science	Little	Researcher / Tool designer	Research concepts	?	20. 000	- Implementing a VisTool	- Clusters	- Insight into Clusters	- co-occurrence matrix of research concepts - force-directed layout: 2d scatterplots	- Scalability - Blob - Are clusters real
Visualizing bio polymer data	Biology - polymer bio.	Yes	Researcher / Tool designer	feature vectors about about polymer molecules	10.000 - 100.000	up to 1.000	- Implementing a VisTool	- Clusters	- Insights	- MDS - force-directed layouts - correlation matrices - mdsi, c, mavis	- Are my clusters real - See clusters in higher D
Different projects with simulating run time of algorithms	- Computational Intelligence -	Yes	Researcher / Algo designer	Features about algos (eg parameters), run time, sampled	100.000 - millions	100-200	- Generation of knowledge about algorithm parameterization, etc. - Come up with new algos, models, etc	- correlation between dimensions - clusters of algorithms (UC3)	- Insight	- Scatterplots and co - mathematical calculations	- usual DR techniques do not work because of the dependencies between his variables (run-time!)
Meta-evaluation of behavioural game models	Behavioural game theory	No	Researcher / Algo designer	Sampling models with different parameters	100.000 - millions	5 to 8	- Getting insights into the problem	- Clusters of dims		- mathematical calculations - Line Graphs - Basically computational meth - Line Graphs, Scatterplots	- Data sampling / generating a data set that includes all maximas
Taxonomy for search tasks	Web search machines	Yes	Researcher	data from a big survey where people kept a diary of their search tasks	300 Subjects	lots of (aspects that people used to describe their tasks)	- Come up with (lowd) axes for the taxonomy - understanding how tasks fit in predefined buckets	- low dimensional axes - clusters in this low dimensional space	- Correct Taxonomy: can other people put their task task in correctly		- what are dims/what are
Classification of mutations in Prostate tumour cells	bioinformatics	Yes	Researcher / Algo designer	genomic and clinical data from cancer patients	600 subjects	1.4 M (genomic exons), 20 clinical variables	- classify potential cancer patients based on their genomic data into categories: aggressive, non-aggressive, none - communicating the results to physicians	- clusters - clusters at each stage of the analysis pipeline	- comparisons against physicians' noisy ground truth data - more accurate classification	- R - filtering, feature selection, PCA, MDS - scatterplots used at end of the pipeline	- is the analysis pipeline valid? trust - batch effect: noisy, perhaps mostly indifferentiated data - need insight at interim stages of the analysis pipeline - noisy ground truth - dimensions highly intercorrelated - runtime complexity
Multiple Sequence Analysis	bioinformatics	Yes	Researcher / Tool designer	protein sequences, each 10-30K letters long	- 100K sequences (use case 1,2) - 5 - 50 sequences (use case 3)	- UC 1,2: dimensions not meaningful (N^2 distance matrix) - UC 3: several hundred to 50K	- UC1: verify alignment of 100K sequences, locate clusters of sequences, drill down - UC2: use DR avoid computing large distance matrix, as scaffolding to sequence alignment algorithm - UC3: analysis of a virus strain where D >> N	- previously unknown clusters - cluster neighbourhood - clusters that facilitate sequence alignment (UC2)	- UC1: cluster detection, hypothesis generation, cluster insight - UC2: scalable cluster alignment	- many DR techniques and custom algorithms (particularly for UC2): MDS, PCA, CA, PCOORD, mBED, FastMap, SeedMap, BGA, MADE4, CIA - phylogenetic trees, reticulate networks	- trees do not scale past 1K - communication of results in publications (no consensus in the research community) - UC3: D >> N
Models of biological aggregate (flock) behaviour	biological mathematics	unknown	Researcher	2D spatial + temporal biological aggregate motion produced by a model and ground truth data from recorded aggregate motion	# of samples = # of aggregates (exact N unknown) (not individuals within an aggregate)	14 model parameters	- given a set of model parameters, compare the predicted aggregate movement with recorded ground truth data - reduce / refine the number of model parameters, perform sensitivity analysis - generate and validate hypotheses	- clusters - original parameters contributing to the model - groups of dimensions	- comparison with ground truth (recorded biological aggregate motion) - model stability, computation speed	- Matlab to filter model parameters - linear and weakly linear analytical methods, numerical simulations, bifurcation analysis, - Paraglide, multidimensional visualization tool	- stability of parameterization: how quickly is the model changing (need vis for sensitivity analysis) - computational processing time - sparse statistics - visualizations difficult to interpret
Models of medical image segmentation parameters	computer graphics	Yes	Researcher / Algo designer	2D / 3D / 4D medical images (SPECT, dynamic PET), model with input and output parameters, ground truth data (manual image segmentations)	50 - 250 samples, very noisy, less if working in 4D (10-15 samples)	6 to 8 input parameters, 6 to 10 output parameters	- model parameterization, refining, reducing # parameters - sensitivity analysis - human guidance and verification of parameter choice and settings - comparison against ground truth	- clusters as means (areas of parameter stability - shapes of clusters - groups of dimensions - cluster neighbourhood	- comparisons against physicians' ground truth data - more accurate classification - model stability, computation speed	- Paraglide, multidimensional visualization tool - feature selection / manual DR	- needs visualization support - overfitting the model - stability of the model - doesn't trust DR algorithms / understand meaning of new dimensions: results in blob problem
Constructing a structural genomics database for proteins	structural genomics / bioinformatics	unknown	Researcher / Tool designer	protein data base including: - 3D structural data - sequence data - functional data (gene ontology)	range from 70K to 15M protein models	huge space, 3 groups of dimensions	- model building - visualization for communication - hypothesis validation and generation: cluster structure space, correlate with known clusters in functional and sequence space	- clusters as means and ends - groups of dimensions	- smallest number of degrees of freedom needed in structural models - insight into clusters	- dendrograms - custom visualization tools	- need for intuitive visualization - combining categorical and quantitative data, establishing distance metrics
Classifying music listener user types on last.fm	media informatics	Little	researcher / analyst	listening histories and demographic data from Last.fm	310 users (sampled from 5000, sparse data)	49-48 dimensions in 7 categories, mix of categorical and numerical	- hypothesis validation: validate existence of hypothesized user groups	- clusters - semantics of reduced	- separable user groups - validated hypotheses	- k-means clustering - PCA	- no clear user groups after performing initial PCA (blob

Table B.2: Establishing conceptual relationships across summaries.

Task Taxonomy										Other stuff						Problems				Data Taxonomy							Notes	
Pure Data Analysis				Make/ Refine Model			Reduce for Algo Input	Curve Fitting	Manifold Finding	Hypotheses		... xxx		Need Vis		The blob (false negative?)	Communicate to others	Trust the results	other	Type	Scale		Struct. of dims		Smooth Manifold	Indifferentiated		done
Clusters		Dims		Generate	Reduce	Parameterize				Generation	Validation	Explore	Compare	Yes	No						points	dims	semantics of dims	groups of dims				
mean	end	mean	end																									
X				X		X				X	X							X	local maxima, finding the right k	Images - Feature vectors	?	25		X			no case study: not interesting enough?	x
	X									X	X						X		use PCA, but removes too much important information for them	HIV data	20k	~20					no case study: not interesting enough?	x
no			X			X				X	X		X				X	X		sim fishery data	?	05 to 10	X				no case study: vismon	x
unclear										no	X			X			X	X	sparse, varying aged and inconsistent data	different	?	?		X			no case study: very top level interview (not enough info)	x
	X									X		X		X		X	X	X		research concepts	20k	(similarity matrix)					possible case study: descriptive	x
	X									X				X				X	are my clusters real?	feature vectors of polymer molecules	10k-100k	up to 1k					possible case study: descriptive	x
	X									X		X		X			X		clustering of a subgroup of dimensions (rows of heatmaps), order of dims is interesting	genomes	10-100k	160					possible case study: descriptive	x
X		X		X	X	X	X			X	X		X						PCA, MDS, etc do not work for grouped data	sampled in/outs from algos	100k-1M	100-200		X			no case study: no DR	x
		X				X				X			X				X		local maxima	sampled in/outs from game models	100k-1M	5 - 8		X			no case study: no DR	x
	X		X							X						X		X	missing values in the data, clustering is unstable/changes	traffic data from survey	543	39 vars					possible case study: descriptive	x
X				X			X			X			X	X			X	X		Classification of mutations in Prostate tumour cells	~600	1.4M	no	x		maybe	possible case study: predictive	x
	X		X								X	X	X	?	?				what are dims/what are points	diary survey about search tasks	300	a lot	x				possible case study: descriptive	x
X	X						X			X		X		X			X		scalability, #dims >> #points	multiple sequence analysis	100K	(similarity matrix)					possible case-study: predictive	x
X		X		X		X	X					X	X	X		X	X	X	Stability of parameterization, speed	medical image segmentation	250 (s)	6i/6o		X			no case-study - paraglide	x
		X			X	X				X	X		X	X					Stability of parameterization, speed	simulated animal group patterns	?	14i/pattern out	X	X			no case-study - paraglide	x
X	X		X	X	X					X	X		X	X			X		mix of categorical and quantitative parameters, no distance metric	structural genomics / protein structures	70K- 15M	a lot		X			no case study: not enough info	x
	X		X							X	X	X	X	X		X		X		last.fm listener data	310 - 5K	39 - 48	X	X			possible case study: descriptive	x
X							X						X	X			X	X	speed,	classifying human motion	9120	1170					possible case-study: predictive	x
	X									X	X	X	X	X		X	X	X	indifferentiated data, word order not preserved by bag-of-words model	Investigating the Iraqi war logs, wikileaks info dumps	10^3	(similarity matrix)				X	no case study - MoDisco	

Table B.3: Establishing conceptual relationships across summaries (continued).



Multiple Sequence Alignments or MSAs: made by e.g. Clustal W. These are alignments of sets of DNA or Protein "sequences" where the rows are individual sequences, the columns are the letters that match up between different sequences. Here is a short section of an alignment between 7 sequences.

```
Human beta         --------VHLTPEEKSAVTALWGKVN--VDEVGGEALGRLLVVYPWTQRFFESFGDLST·
Horse beta         --------VQLSGEEKAAVLALWDKVN--EEEVGGEALGRLLVVYPWTQRFFDSFGDLSN
Human alpha        --------VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-
Horse alpha        --------VLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHF-DLS-
Whale myoglobin    ---------VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKT
Lamprey globin     PIVDTGSVAPLSAAEKTKIRSAWAPVYSTYETSGVDILVKFFTSTPAAQEFFPKFKGLTT
Lupin globin       --------GALTESQAALVKSSWEEFNANIPKHTHRFFILVLEIAPAAKDLFSFLKGTSE
```

These sequences are evolutionarily related. Originally there was just one sequence in one species and they diverged in different lineages as evolution and speciation and gene duplications happened with letters changing and bits (one or a few consecutive letters) being added on or lost. The latter cause the need for gaps ("-" characters). The major research topic in our lab is how to compute these (find the optimal alignment). A secondary research topic that has become more urgent recently is how to visualise them. We can now make alignments of 100s of thousands of sequences and cannot look at them. Normally MSAs are analysed by clustering the rows according to some function of the number and type of shared characters between each pair of sequences. We have been using MDS and PCA and Correspondence Analysis to do this and want to do more. I originally did this with PCOORD (see below).

CLUSTER DETECTION

PROBLEM

DISTANCE

CLASSIC MDS

PCOORD: principal co-ordinates analysis from John Gower in the 1960s; turns out to be classic MDS. We used to do this with SVD on a double centred distance matrix. Now we do it with Nyström if N is very big. We also tried the power method. I used this in my PhD which was a clustering of insects from different orders based on morphology. Since then I used it once in a paper in 1992, which is the only single author paper I ever wrote, on applying it to sequence alignments.

DH-1 [redacted]

I attach a PDF but this is ancient and very long winded and a bit silly and ignores the entire MDS literature. This paper just looked at how to apply this to sequences so the explanations are very laboured and embarrassingly patronising. The basic idea is that you can take a MSA and compute a distance between every pair of sequences and do PCOORD. It discusses the kinds of distances to use and has some black and white pictures that used to be pretty. These were only just post Letraset® i.e. the pictures were made using a drawing program (MacDraw) to modify output from a plotting package instead of sticking small letters onto sheets of paper. These were among the first completely computer based figures I ever made.

Figure B.3: Document sent to us by Analyst A9, with post-hoc coding to identify concepts.

revised interpretation of our findings[3] focuses primarily on tasks relating to DR (Figure B.9), with less emphasis on challenges. Our findings are summarized according to this classification in Table B.7.

[3]Reproduced below in Section B.6.

Summary: Interview, [redacted] 2010-09-17

------- personal ------------

Biologist with statistics background.

Domain: Biology - Fisheries.
Type of user: Data analyst for fisheries data.

Work place: DOF.
Overall User Goal: Give recommendations about which harvest control rule to use. Publish papers.

Experience with DR techniques:
- No

---------- *project* -----------

Short project description:
Comparison of mathematical models simulating the behavior of fish populations. All models take a set of parameters such as carrying capacity and productivity. Each model is run with a variety of parameter combinations and [redacted] checks if the fish population has died out for this model/parameterization.

Data:
- Models tested with different parameterization
- simulation data (input and output)
- (n/dim unknown so far)

Tasks:
- Evaluating different harvest control rules (HCRs)
- Finding best HCR (i.e. Most robust against a variety of assumptions about fish populations)

Pattern of interest:
- correlations

Goals/Metrics:
- Level of extirpation

Current analysis techniques (Visualization, Stats, DR):
- R
- Matrices of line plots

Current problems / challenges:

Figure B.4: An early version for an analyst summary (Analyst A11), where the italicized headings reflect the result of focused coding to organizing concepts, properties, and dimensions.

Summary: [redacted]

Based on:

- 1.5h Skype interview (11-01-26 MS + TM Notes)
- CHI '11 submission
- Email communication

Type of user	Researcher, wants to better understand search browser intent
User goal	Conduct research, publish papers, build better search UI/support for targeted ads
DR User	Non-expert, some background (alumni UBC InfoVis group), recruited machine learning expert to help with classification
Experience with DR techniques	Used correspondence analysis to view structure of query tasks. Exposed to DimStiller, tried to use it but it didn't work out: couldn't understand results.

Brief research description:

A better understanding of search browser query intent is needed to improve search browser user interfaces and support targeted advertising that better matches queries. They seek to build, refine, and validate a search query intent taxonomy, and then classify queries with their validated taxonomy.

Several hundred [redacted] users were recruited to maintain a diary of their search query tasks over the course of several months. Queries were annotated by these users. Search metrics and topical content related to these queries were also collected. The researchers constructed a taxonomy based on this dataset, based on two intrinsic dimensions and 12 query intent clusters. This taxonomy will be used to classify future search queries based on intent.

Use case Instance:

Search query logs decomposed into individual search tasks from potentially hundreds-thousands of [redacted] users need to be mapped to a low-dimensional space. The researchers seek to validate the intrinsic dimensionality of their taxonomy and its search query intent clusters with this data.

Data

- N = 1290 - 1463 search tasks (from 36 users); subsequent field study with 300 people (N much higher) (small by [redacted] standards)
- D = many: (dirty data (sparse, incomplete), 12-13 diary/questionnaire fields, search metrics (clicks, refinement events, abandonment, query length), topical content, mix of categorical and numerical dimensions

Figure B.5: A later version of an analyst summary (Analyst A2), where the bold headings reflect our focused coding: concepts, properties, and dimensions.

Use case instance	Primary Person / Paper	Add. persons / papers	Task / Data: clust new	clust old	clssification	dim new	dim old	dim groups	dim semantics	Use Case Classes: PDA	MM	MR	MP	AD	RAI	SA	CF	NOP	DR Technique (done): none	filter	linear	nonlin	Comments
Evaluating different fishing simulations							x	x	x	x			x						x	x			
Create Image search algorithm				x				x			x		x					(x)	x		x	x	
More accurate clustering for HIV data / flow cytometry			x	x						x				x*					x	x	may		*Refining Clustering Algos
Genome Overview			x				x	x*		x											x	x	*rows and cols of the matrix
Visualizing research concepts in life science			x							x											(x)	x	similarity matrix, no dims
Visualizing bio polymer data			x							x											x	x	
Characterizing dynamic travel patterns of recreational boaters			x				x		x	x	(x)								x		x	x	
Run-time prediction of algorithms for NP-hard problems				x				x		x	x	x								x			PCA not applicable, dims too intercorrelated
Parameters in Algorithms													x						x				
Local Search Algorithm				x						x									x				tree-based distance metric
Meta-evaluation of behavioural game models							x	x				x	x						x				
Classification of mutations in Prostate tumour cells			x		x					x	x				x					x*	x	x	*and feature selection
Taxonomy for search tasks			x		x*	x	x		x	x											x	x	*cluster into given taxonomy
Multiple Sequence Analysis			x*							x					x						x	x	*cluster neighbourhood
Bird moving patterns		a		x?			x	x	x			(x)	x*			x			x	x			*understand parameterization
New Model for medical image separation					x			x				x	x	x	x	x				x			
Structural patterns in proteins, matching with functional, sequence patterns			x					x		x	x								x*				*unknown
PCs of Music listening histories			x		x	x*	x	x	x	x											x		*clusters of dimensions
Classification of human motion with wearable sensors				x	x										x					x	x		
Computational Chemistry					x			x		x												x	cause/effect relationships
Insights and Com of large document sets (Journalism)			x							x												x	

Table B.4: Establishing conceptual relationships across summaries (continued).

Use case instance	Primary Person / Paper	Problems: blob	Communicate	trust	dirty data	other	Other stuff: Need Vis	hypo generate	hypo validate	compare	indiff data	n	d	other comments
Evaluating different fishing simulations			x	x			x			x*		?	5 - 10	* compare 4-5 different models
Create Image search algorithm				x		local maxima, finding the right k		x (vis)	x			?	25	sparse infos from interviews
More accurate clustering for HIV data / flow cytometry			x	x			x			x*		20K	~20	* compare different clustering algorithms
Genome Overview			x			order of dims, clustering subgroups of dims	x		x			100K	160	
Visualizing research concepts in life science		x	x	x	x		x	x				20K		similarity matrix, no dims
Visualizing bio polymer data				x	x		x	x				100K	1K	
Characterizing dynamic travel patterns of recreational boaters		x		x	x	missing guidance which techniques to use,	x	x		x*		534	39	*compare diff clustering techniques for cross-validation (trust)
Run-time prediction of algorithms for NP-hard problems						run time. highly intercorrelated dimensions				x		1M	100	
Parameters in Algorithms						local maxima, intercorrelated dimensions				x		1M	100o, 63i	
Local Search Algorithm						intercorrelated dimensions				x		11	lots	
Meta-evaluation of behavioural game models			x			local maxima		x		x		1M	5 - 8	
Classification of mutations in Prostate tumour cells			x	x	x	#dims >> #points	x			x	x*	600	1.4M	* possibly indifferentiated
Taxonomy for search tasks					x	what are dims/what are points	?		x	x		300	a lot	
Multiple Sequence Analysis			x			scalability: #dims >> #points	x	x				100K		similarity matrix, no dims
Bird moving patterns						stability of parameterization	x	x	x	x*		?	14i/ 1o	* compare model to ground truth
New Model for medical image separation		x	x	x		stability of parameterization	x			x		250	6i,6o	
Structural patterns in proteins, matching with functional, sequence patterns			x			mix of categorical and quantitative parameters, no distance metric	x	x	x	x		70K-1M	a lot	
PCs of Music listening histories		x		x		missing guidance which techniques to use,	x	x	x	x		300-5K	39-48	
Classification of human motion with wearable sensors			x	x		speed	x			x		9K	1.1K	
Computational Chemistry			x	x	x	sparse data, varying aged and inconsistent data, missing guidance	x		x			?	?	
Insights and Com of large document sets (Journalism)		x	x		x	indifferentiated data, runtime complexity	x				x	10^3		similarity matrix, no dims, indifferentiated

Table B.5: Establishing conceptual relationships across summaries (continued).

B.6 Dimensionality Reduction in the Wild

[4] In this manuscript, we contribute the first systematic and broad analysis of dimensionality reduction (DR) usage conducted "in the wild", by observing the usage patterns of real data analysts, along with their needs and problems. We present the results of a two-year qualitative research endeavor, in which we iteratively collected and analyzed a rich corpus of data. We interviewed nineteen data analysts from ten different domains and selected five papers describing data analysis using DR, deriving twenty-seven real-world usage examples in total. Grounded in this data, the main contribution of this paper

[4] This section is a slightly modified version of our unpublished manuscript *Dimensionality Reduction in the Wild* by Michael Sedlmair, Matthew Brehmer, Stephen Ingram, and Tamara Munzner (2013).

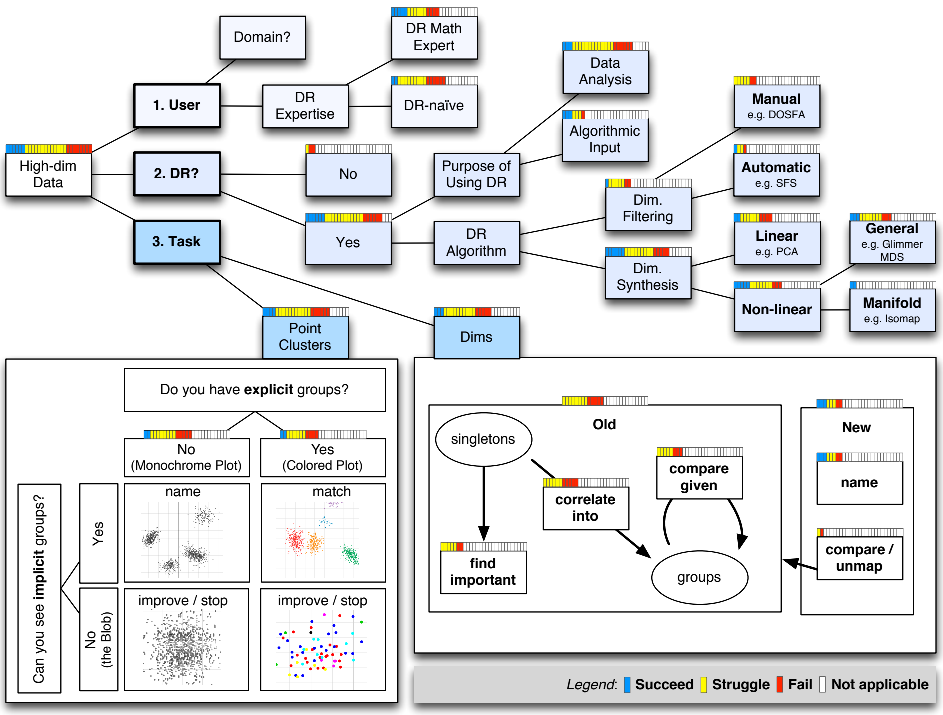


Figure B.6: Our preliminary classification of high-dimensional data analysis in terms of people who use DR, DR techniques, and tasks [283]. References: DOSFA (Dimension Ordering, Spacing and Filtering Approach) [365], SFS [155], PCA [161], Isomap [313], Glimmer MDS [149].

is a classification of tasks that relate to high dimensional data analysis and the use of DR. Its high-level structure differentiates between abstract tasks related to the visual analysis of scatterplots versus of dimensions, in contrast to reducing dimensionality for downstream usage by completely automatic algorithms.

B.6.1 Introduction

DR is the process of reducing a high- dimensional dataset to a low-dimensional visual encoding that retains most of its important structure. It has been an active research area throughout several decades and across many

fail / strug / happy	Usage Scenario	Paper/ Interview	1. User			2. DR							3. Task				Gaps						
			Domain	DR expertise		Purpose of Using DR		DR Algorithm					Point Clusters		Dims		User Gaps: DR-naïve			Task Gaps: Dimensions		Data Gaps: Assumptions	
								Dim. Filtering		Dim. Synthesis													
				DR math expert	DR-naïve	Algo. Input	Data Analysis	Manual	Auto	Lin	Non-lin: Gen.	Non-lin Man.	Verify Explicit	Identify Implicit	Old	New	Concept.	Interpret.	Guidance	Groups	Unmapping	Categorical	Scalability
happy	MoCap A	Interview	machine learning		[X]	[X]			[X]	[X]			[X]						[X]				
fail	MoCap B						[X]	[X]	[X]	[X]			[X]					[X]	[X]				
fail	Music A	Interview	HCI		[X]		[X]			[X]			[X]	[X]	[X]		[X]	[X]	[X]			[X]	
strug	Music B						[X]			[X]					[X]	[X]	[X]	[X]	[X]		[X]	[X]	
strug	Concept	Interview	life sciences		[X]		[X]	[X]		[X]	[X]			[X]		[X]	[X]	[X]	[X]				[X]
fail	NPAlgo	Interview	machine learning	[X]		[X]		[X]							[X]	[X]				[X]			[X]
happy	BRDF	Paper	graphics	[X]			[X]			[X]		[X]				[X]							[X]
fail	FishPop	Interview	fisheries sciences		[X]										[X]				[X]	[X]			
fail	SeqAln A	Interview	bioinformatics	[X]	[X]		[X]			[X]	[X]		[X]	[X]				[X]	[X]				
happy	SeqAln B					[X]					[X]			[X]					[X]				
fail	SeqAln C						[X]				[X]			[X]					[X]				[X]
fail	GamMdl	Interview	machine learning	[X]										[X]	[X]					[X]			[X]
fail	Search	Interview	search optimization		[X]		[X]			[X]	[X]		[X]	[X]	[X]	[X]	[X]		[X]		[X]	[X]	
strug	ProstCan	Interview	bioinformatics		[X]	[X]	[X]	[X]	[X]	[X]	[X]		[X]		[X]		[X]	[X]	[X]	[X]			[X]
strug	EpiGen	Interview	bioinformatics		[X]		[X]			[X]	[X]			[X]	[X]		[X]	[X]	[X]	[X]			
strug	StrucGen	Interview	bioinformatics		[X]								[X]	[X]	[X]		[X]		[X]	[X]		[X]	[X]
strug	FlockSim	Interview	mathematics	[X]			[X]	[X]						[X]	[X]			[X]		[X]			
strug	CompBio	Interview	comp. biology	[X]	[X]	[X]	[X]	[X]	[X]				[X]	[X]				[X]	[X]				[X]
strug	ChemRel	Interview	comp. chemistry	[X]			[X]				[X]		[X]		[X]			[X]		[X]		[X]	
strug	MedImg	Interview	comp. vision	[X]			[X]	[X]						[X]	[X]			[X]		[X]			
strug	TxtDocs	Interview	journalism	[X]	[X]		[X]				[X]		[X]	[X]			[X]	[X]	[X]				[X]
strug	BoatAct	Interview	marine & ocean sci.		[X]		[X]			[X]	[X]		[X]		[X]		[X]	[X]	[X]			[X]	
strug	Polymers	Interview	structural chemistry		[X]		[X]			[X]	[X]			[X]			[X]	[X]	[X]				
strug	Quadrup	Paper	graphics	[X]	[X]	[X]	[X]			[X]						[X]		[X]					
happy	ArtShp	Paper	graphics	[X]		[X]					[X]		[X]										
happy	Faces	Paper	graphics	[X]			[X]					[X]				[X]							
happy	MorseCd	Paper	statistics	[X]			[X]				[X]			[X]		[X]							
Total count				13	14	7	20	7	4	13	13	2	12	15	13	8	10	14	17	9	2	6	9
				57%	61%	26%	74%	26%	15%	48%	48%	7%	44%	56%	48%	30%	37%	52%	63%	33%	7%	22%	33%
			of 18 interviewees and 5 papers			of 27 usage examples																	

Table B.6: Usage examples described using our preliminary classification of people who use DR, DR techniques, tasks, and gaps. For a description of the gaps, see our technical report [283]. Paper references: BRDF [201], ArtShp [39], Faces [313], MorseCd [41]. The usage example abbreviations correspond to analyst ID numbers listed in Table B.1.

domains, from its origins in psychology [319, 368] through statistics [41] to machine learning [19, 124, 155, 313, 331] and visualization [149, 159, 365].

The DR literature is heavily focused on mathematical and algorithmic descriptions of new techniques and their characterization based on algorithmic properties [73, 101, 124, 155, 332, 360]. In recent years, the visualization community has increasingly focused on understanding the analysts' perspective, as part of the quest for visualization design and evaluation that reflects real needs and activities [163, 164, 217]. However, almost no work has considered DR from a such human-centered perspective. Many questions about how DR is actually used *in the wild*, in real world settings with real datasets, remain open: when do those with high-dimensional data use DR techniques? What are their tasks and goals, considered at an abstract level? Which DR techniques do they use and how do they use them? How well do their real datasets match up with common benchmarks? Understanding these questions can guide both technique-driven work such as further algorithmic development, and problem-driven work such as design studies.

With these questions in mind, we engaged in a two-year qualitative research project based on semi-structured interviews with nineteen data analysts from ten application domains, followed by extensive analysis based on iterative coding. We present the results as a classification of abstract tasks that provide a framework for understanding how, when, and why analysts might use DR. Our task classification has three categories at the highest level: *learning about dimensions*, *seeing clusters*, and reducing dimensionality for the purpose of *algorithmic input*. We focus on the former two as visual data exploration tasks are particularly important for the visualization community, rather than the latter usage that is common practice in completely automatic analysis, as in the machine learning community [222].

We also relate our observations of analysts' in-the-wild practices to the technical DR literature, providing a synthesis overview that is accessible to the visualization community. We also provide an analysis of the relationships between the tasks in our classification and benchmark datasets common in DR algorithm development and usage, and discuss challenges encountered by the analysts.

Our larger motivation for this project was a gap in existing work on characterizing tasks in the visualization literature: there was no adequate characterization of abstract tasks that people face when using DR techniques. Design studies are an increasingly popular form of visualization research [284], but they focus on identifying tasks within a very specific usage context in a particular domain. Existing cross-domain classifications of tasks are not specific to high-dimensional data analysis and thus do not provide enough detail about the use of DR in particular [7, 366]. The enormous amount of work on categorizing DR techniques and related methods such as feature selection and factor analysis is difficult to connect with the existing visualization literature on abstract tasks for researchers not already embedded in the area of high-dimensional data analysis.

Our choice of methodology was motivated by a vibrant thread of work in the visualization and HCI communities using qualitative methods in general [31, 49, 153, 285, 322], and in-the-wild field studies in particular [163, 164]. The strength of the methodological approach of iterative coding of qualitative data [55] is realism [203]: we do make existence claims, in that all of the tasks in our classification are grounded in real-world DR usage examples. However, we do not make claims about completeness; our classification might lack some use cases due to sampling or observer bias. We do not claim to have identified new tasks; our choice of naming is intended to be evocative to the visualization community in a way that aligns with existing task classifications such as our task typology from Chapter 2, and we discuss the connection between our categories and usage in other communities throughout this manuscript.

In the language of Munzner's four-level nested model of visualization design and validation [211, 217], this manuscript is focused on the abstraction layer. We do not directly address either of the two layers nested within it, namely visual encoding and interaction design choices, and algorithm design. Task classifications designed for visualization researchers have many uses. Those conducting design studies involving visualization of high-dimensional data can use our abstractions to guide the categorization and abstraction of problems and tasks, including the decision of whether DR is needed at all. Researchers presenting new DR techniques can use the set of abstract tasks to concisely state assumptions about which tasks are supported, rather than leaving this description implicit in a way that places a burden on a potential adopter of the technique. This task classification can also be used in the design of controlled experiments, and can facilitate the analysis of existing systems.

The main contribution of this manuscript is the presentation of results from our qualitative field work, which we summarize as a organized set of abstract tasks (Figure B.9). Each abstract task is illustrated with usage examples from our study, grounding it in the data. Three secondary contributions are the synthesis of many ideas scattered throughout the technical DR literature into a coherent and accessible overview for visualization researchers through the lens of task abstraction, the discussion of the relationship between common benchmark datasets to these tasks, and a discussion of challenges faced by analysts.

B.6.2 Related Work

We discuss three categories of related work: domain-specific design studies for high-dimensional data analysis, human-centred DR research, and classifying DR or tasks.

Design studies with high-dimensional data: Several design studies report on tools created for high-dimensional data analysis within a specific domain context [29, 240, 248]. Design studies typically focus on a single domain, with an emphasis on identifying domain goals and deriving abstract tasks. However, although the targeted data in these field endeavors was high-dimensional, the associated tasks were of parameter optimization and sensitivity analysis; none focused on tasks related to DR as such. From a methodological perspective, our work differs significantly from a design study approach because we operate across many domains.

DR from a human-centred perspective: While there is very little field work on DR available, there is some technical work that focuses on DR from a human-centred perspective. In our own previous work, we presented DimStiller [150], a system for providing people with guidance via specific DR workflows that correspond to specific tasks. It addresses the conjectured needs of "middle-ground users": visualization generalists and application domain experts who do not have a deep understanding of DR mathematics. Endert et al. [89, 90] developed strategies to abstract away some of the complexities of DR by allowing a person to interactively change the underlying mathematical model through directly manipulating a lower-dimensional projection of high-dimensional data. Joia et al. [160] and Paulovich et al. [235] follow a similar approach and proposed DR techniques that allow to integrate interactivity into the reduction process. Our own work on cluster separation for dimensionally reduced data presents guidance on the relationship between visual encoding choices and DR choices [286]. We go beyond this previous work in addressing a much broader scope in terms of both tasks and domains.

One of the rare DR studies involving human subjects is a lab study conducted by Lewis et al. [188]. They compared how experts, novices, and informed novices subjectively judged the value of two-dimensional projections produced with nine different DR techniques. Their findings showed that experts were more consistent in their positive and negative ratings. We aim to answer a different set of human factors questions and take a fundamentally different methodological approach by way of a qualitative interview study.

Classifying DR or tasks: Roth [262] specifies three different perspectives for classifying interaction: by technique, by the data that techniques operator on, and by task. All existing classifications of high-dimensional data analysis use the first two perspectives, classifying DR techniques based on their algorithmic properties [73, 101, 155, 332, 360], or by characterizing the data that DR tools are acting on [24, 285]. We aim to complement this body of work by providing an example of the third perspective, classifying by DR tasks.

Conversely, while the visualization literature contains several examples of abstract task classifications that span multiple domains [7, 8, 117, 291, 301, 366], and a few examples of domain-specific task taxonomies [325], these do not address tasks related to DR.

Moreover, there are methodological differences between our work and this line of previous work. Classifications of tasks are often based on the authors' own experience in conjunction with a thorough consideration of the current state of the art [7, 291, 366], while others are based on observations of human behavior in controlled laboratory settings [8, 117]. In contrast, our work is conducted in the wild by interviewing expert data analysts. It is similar in spirit to understanding of the processes undertaken by intelligence analysts [164] and enterprise data analysts [163], but its scope extends across multiple domains.

DR algorithms: While our work focuses on tasks rather than algorithms, in order to concisely discuss the implications of our classification we need to refer to a classification of DR algorithms. This subject has been the target

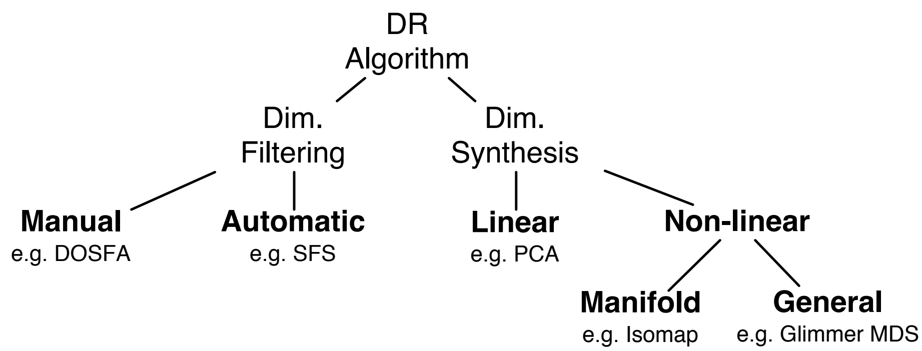


Figure B.7: Our classification of dimensionality reduction algorithms as either dimensional fitlering or dimensional synthesis. References: DOSFA (Dimension Ordering, Spacing and Filtering Approach) [365], SFS [155], PCA [161], Isomap [313], Glimmer MDS [149].

of extensive previous work. Many classifications of DR algorithms that have been proposed are based on different technical distinctions, including feature selection and feature extraction/construction [73, 124, 155, 360], linear and non-linear [155], globally and locally operating techniques [101], or convexity and spectrality [332].

Figure B.7 shows the classification that we use in this manuscript. We illustrate each of these categories with some concrete examples of algorithms, with a preference for well-known, more canonical examples and some examples from the visualization community. Our goal here is not to provide a complete and thorough survey of the latest techniques or an extensive discussion of DR algorithms; for that, we refer the interested reader to the papers above.

This classification reflects distinctions that are well-known in the technical literature, but at a granularity suitable for the discussion of this manuscript that is simpler than many previous classification systems. It also emphasizes the types of algorithms used by the analysts we interviewed. Our choice of terms follows the vocabulary of the visualization community, rather than that of machine learning or other communities in which DR is also used.

At a high level, DR algorithms can be divided into *dimensional filtering*, in which less interesting dimensions are filtered out, and *dimensional synthesis*, in which old dimensions are combined into new synthetic dimensions. In the machine learning literature, these categories are known as feature selection and feature extraction, respectively [360].

Dimensional filtering can be further differentiated between *manual* and *automatic* filtering techniques. Examples of manual filtering techniques are human-defined quality metrics for finding interesting dimensions [159] and DOSFA (Dimensional Ordering, Spacing, and Filtering Approach) [365]. An example of an automatic filtering technique is SFS [155] in which the "best" dimension is selected and others are added iteratively until no further improvement is made, relative to a threshold selected a priori. In the visualization community, various techniques have been proposed that assist people in finding interesting two-dimensional projections of high-dimensional data; in other words finding interesting two-dimensional scatterplots in a SPLOM. Recent advances in this area, for instance, focus on finding two-dimensional projections that separate pre-defined clusters well [296, 312].

Dimensional synthesis is commonly differentiated between *linear* and *non-linear* techniques [155]. Linear techniques such as PCA [161], Factor Analysis [59], or classical MDS [319, 368] produce new dimensions from linear projections of the original data. Many datasets have an intrinsic structure that can only be revealed using non-linear techniques; these are further divided into *general* and *manifold* techniques. We define manifold techniques as the subset of non-linear techniques with the underlying assumption that the data lies on a densely sampled manifold. Of the numerous techniques that have been presented, some relevant examples include Isomap [313], Laplacian Eigenmaps [19], Diffusion Maps [63], Maximum Variance Unfolding [288], NeRV [336], and Riemannian Manifold Learning [190]. Although the machine learning literature often uses the term manifold synonymously with non-linear, we distinguish manifold-specific techniques from general non-linear techniques. Examples of general techniques are the distance scaling MDS approach of Glimmer [149], and the multidimensional projection techniques Least Square Projection [233], PLMP [234] and LAMP [160].

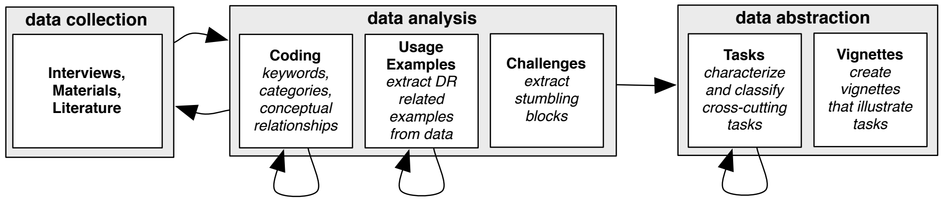


Figure B.8: Our data collection, analysis, and abstraction methodology.

Some nonlinear dimensional synthesis techniques straddle the two classes; for example, Chen and Buja [56] introduce a parametrized family of mapping techniques that can be tuned for either the general or manifold case. The operational distinction between manifold and general will become particularly important in Section B.6.6, where we discuss implications for benchmark datasets and relate them to manifold unrolling techniques.

B.6.3 Methodology

We sought to better understand the tasks, challenges, and context of high-dimensional data analysts working in different domains. To attain this understanding, we adopted a research methodology that alternated between cycles of data collection and analysis, allowing us to gradually identify and refine conceptual relationships grounded in the data [55], as represented in Figure B.8. This methodological approach has been shown to be effective before, particularly in recent work on characterizing the processes undertaken by intelligence analysts [164] and enterprise data analysts [163], and the types of evaluation scenarios found in the visualization literature [183].

"In the wild" rationale: We borrow the phrase "in the wild" from HCI [137], where it differentiates investigation of people with their own work materials performing tasks in real-world settings from studying human behavior in controlled laboratory settings; the goal of in-the-wild studies is maximizing the realism of findings [203]. The importance of better understanding the needs and visualization practices of people "in the wild" has been repeatedly raised in previous work by others [49, 153, 243] as well as in our own previous work [217, 282].

Data collection: We undertook several methods for studying the tasks and challenges of high-dimensional data analysts. Our primary method was semi-structured interviews.

We interviewed nineteen high-dimensional data analysts, recruited from our extended personal and professional networks[5]. The analysts were sampled from a range of ten different domains, which included bioinformatics, machine learning, HCI, policy analysis, and journalism. The majority of our analysts were academic researchers, ranging in years of experience from graduate students to senior professors. The remaining analysts were employed either by a national research lab, or by the private sector.

We conducted one or two interviews per data analyst, ranging in duration from thirty minutes to four hours. Most interviews were individual interviews while others were group interviews with multiple analysts working on similar problems. Whenever possible, we interviewed analysts at their place of work; others were conducted over the phone or using Skype.

We used a set of open-ended questions relating to high-dimensional data and DR, which evolved over the course of data collection and analysis. We asked our interviewees about both projects from their past as well as ongoing work; in each interview, we sought to understand[6]:

- What are the goals, questions, or hypotheses that drive analysts?
- What data analysis tasks do analysts undertake?
- Do these tasks occur often? How long do they take?
- What are the characteristics of analysts' datasets?
- How and when is DR used? What role does visualization play?
- What are the major problems and challenges faced during the data analysis process?

[5]See summary in Table B.1.

[6]See Section B.2 for the complete interview question list.

- What are the patterns of interest? (clusters, outliers, correlation, semantically meaningful dimensions)

One to three interviewers guided the interview sessions. We took extensive notes and audio recordings for later analysis, and we asked follow-up questions via email.

We also gathered materials from the analysts that we interviewed, which included their published papers, unpublished manuscripts, theses, datasets, and static images of their visualized data.

To broaden the corpus of data, we also surveyed the published DR literature. To further enrich our data analysis, we chose to additionally include a small number of papers that featured the application of DR for solving a particular domain problem; we excluded many papers that simply present a new DR technique or algorithm without a detailed discussion of using it to solve a real-world problem.

We concurrently performed data collection and analysis, culminating in the representation of our findings in the set of DR tasks described in Section B.6.4, as well as an understanding of the challenges faced by high-dimensional data analysts, discussed in Section B.6.5. The analysis process yielded five types of results.

Coding: Interview notes and other materials collected were iteratively coded to identify conceptual relationships into higher-level categories [55]. These codes spanned analysts' high-level research goals, their expertise and assumptions, the structure and size of their dataset(s), their current analysis techniques and approaches, the challenges they faced, as well as desired or actual outcomes. This code set provided us with an intermediate result that we used for further analysis.

Usage examples: Using our refined set of codes, we iteratively constructed and revised twenty-seven concise and distinct DR *usage examples*: twenty-two from interviews and five from our literature review; these are summarized in Table B.7. A usage example includes a description of an analyst's task as well as the means by which they perform this task.

While most of the interviews resulted in one usage example, several analysts described multiple yet fundamentally different analysis situations in which they used or wanted to use DR; these became multiple usage examples, allowing for a concise description using our set of abstract tasks. In the case of one interviewee, we did not have enough information to derive a concise usage example, so this data was excluded from subsequent analysis. Additionally, we derived five usage examples from the literature with interesting descriptions of high-dimensional data analysis and DR [39, 41, 201, 256, 313]. These examples were unusual in that they contained sufficient detail about the data analysis process itself to support a usage example; we considered dozens of other technical papers that did not, but instead focused exclusively on algorithmic issues.

Challenges: Our analysts also described problems they faced while performing DR tasks. We report these challenges in Section B.6.5, relate them to the technical literature, and discuss implications for the design of high-dimensional data analysis tools and techniques.

Tasks: We then broke down the usage examples into smaller units of concrete and abstract tasks. We iteratively explored different levels of task abstractions and organization, and evaluated them across the usage examples. The final results are the hierarchical organization shown in Figure B.9.

Vignettes: Finally, we created descriptive *vignettes* based on our usage examples and task classification, to ground our written discussion of these tasks in our collected data. These vignettes, featured in Section B.6.4, provide concise and highly abstract summaries of an analyst's task. In contrast, unprocessed direct quotes from interviews would be prohibitively lengthy and moreover would be framed in the vocabulary of the particular domain, impeding cross-domain abstraction. The importance of such abstractions and its challenges have been discussed in our previous work [217, 284].

Further details about our methodological approach can be found in the supplemental material.

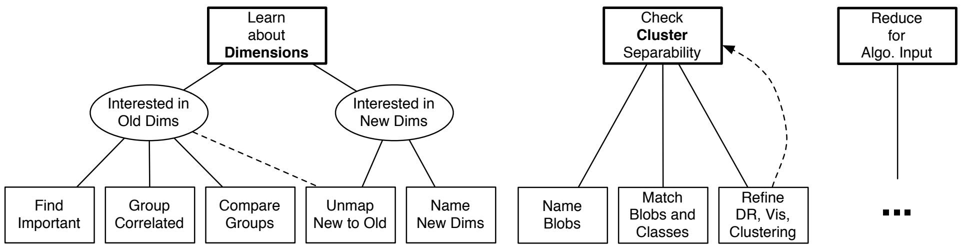


Figure B.9: Our classification of tasks relating to DR.

		Tasks									DR Algorithm				
		Dimensions					Point Clusters			Algorithmic Input	Dim. Filtering		Dim. Synthesis		
		Find important	Group correlated	Compare groups	Name new	Unmap new	Name blobs	Match blobs classes	Refine DR, Vis, clust.		Manual	Auto	Lin	Non-lin: Gen.	Non-lin Man.
BoatAct	policy analysis		X*					X	X				X	X	
Concept	biology				X		X		X*		X		X	X	
FishPop	policy analysis			X*											
FlockSim	mathematics	X*		X			X		X		X				
MoCap A	machine learning							X*	X		X	X	X		
MoCap B	machine learning									X*		X	X		
Music A	HCI						X*	X*	X*				X		
Music B	HCI	X	X		X	X*							X		
NPAlgo	algorithms			X*	X					X	X				
StrucGen	biology			X*			X	X	X		X				
TxtDocs	journalism						X*	X	X					X	
Faces	computer vision				X*										X
ChemRel	chemistry	X		X				X	X					X	
CompBio	biology						X	X	X	X	X	X			
EpiGen	biology	X	X				X		X				X	X	
GamMdl	algorithms			X			X		X						
MedImg	computer vision			X			X		X		X				
Polymers	chemistry						X		X				X	X	
ProstCan	biology	X		X				X	X	X	X	X	X	X	
Search	HCI	X			X	X	X	X	X				X	X	
SeqAln A	biology						X	X	X				X	X	
SeqAln B	biology						X		X	X				X	
SeqAln C	biology						X		X					X	
ArtShp	graphics							X	X	X				X	
MorseCd	statistics				X		X							X	
BRDF	graphics				X								X		X
Quadrup	graphics				X					X			X		
		6	3	8	8	2	15	11	19	7	7	4	13	13	2
		17 of 27 usage examples					20 of 27			7 of 27	25 of 27				

Table B.7: Usage examples described using our classification of tasks relating to DR. Asterisks indicate tasks that we describe in vignettes. Paper references: FACES [313], ARTSHP [39], MORSECD [41], BRDF [201], QUADRUP [256]. The usage examples abbreviations correspond to analyst ID numbers listed in Table B.1.

B.6.4 Taxonomy

Figure B.9 presents our findings in terms of tasks involving DR usage for high-dimensional data analysis. Rectangular nodes represent tasks and oval nodes represent interests that further characterize tasks.

At the top level, we distinguish between three different intents of using DR: using DR to learn about *dimensions*, both the given *old* dimensions as well as synthesized *new* dimensions, using DR to see *point clusters*, and finally using DR for reducing data for *algorithmic input*.

Our analysis focuses predominantly on the former two, dimensions and point clusters, which involve DR usage as part of the data analysis process. Both of these high-level tasks are hierarchically broken further down into mid-level tasks. We specifically abstained from breaking them down into all of their lowest-level components, since low-level interactive visualization tasks are adequately classified in previous work [8, 291, 366]. Instead, we focus on revealing meaningful mid-level tasks.

For tasks relating to dimensions and point clusters, analysts often used DR with the intent of visualizing the dimensionally reduced data, using visual encodings such as two-dimensional scatterplots (twenty usage examples), three-dimensional scatterplots (three usage examples) or a SPLOM (two usage examples). Nevertheless, we also found instances where analysts were using DR for the purpose of data analysis without subsequently visualizing the low-dimensional data, such as the Music B (A1) usage example discussed below.

As a third high-level task, we include using DR for algorithmic input. Here, DR is used as a means of automatic data compression and noise reduction rather than for explorative data analysis. We provide this task as important context, however we do not go into depth as with the other two task categories. The power of using DR for the purpose of algorithmic input has been widely discussed in the machine learning and data mining communities [25, 222].

DR for Learning About Dimensions

We found seventeen usage examples in which analysts were interested in learning about the dimensions of a dataset, reflected in the left-most branch of Figure B.9. We differentiate between interests in the original high-dimensional, or *old*, dimensions, and interests in the *new* dimensions that are the result of dimensional synthesis DR algorithms. Since a single data analyst can be engaged in multiple tasks, having interests in both old and new dimensions within the same usage example is not uncommon.

Find important old dimensions: We identified six usage examples in which the analyst sought to *find important old* dimensions in a dataset, according to some particular metric of interest, for example the variance that a single dimension contributes to the overall variance.

> In the FlockSim usage example, a mathematician (A17) was interested in modeling self-organized aggregate animal behavior; in the case of birds, this is known as flock behavior. Her dataset of recorded flock behavior contained fourteen spatiotemporal dimensions. A desirable model would accurately and precisely predict flock behavior, such as where a flock may land. The model would also need to strike a balance between sensitivity and computation time; a model with two to three spatiotemporal dimensions was therefore thought to be ideal. Additionally, the model parameters should retain the semantics of the original dimensions. Thus the analyst's task is one of *finding the most important dimensions* in the recorded flock data, examining which dimensions contribute heavily to the overall variance. She did that by manually filtering and inspecting all fourteen dimensions.

For identifying important dimensions, analysts usually engage in manual or automatic filtering of dimensions. Some also used linear dimensional synthesis techniques such as PCA, as in the Music B (A1) example discussed below.

Group correlated old dimensions: Analysts may also be interested in grouping old dimensions. Usually, this grouping is done by consolidating singleton dimensions that are correlated together into groups. We identified three usage examples where the analyst engaged in *grouping correlated* dimensions. In these cases linear DR techniques were used as indicators for dimensional correlation.

> In the BoatAct usage example, a geographic information systems (GIS) analyst (A3) was interested in characterizing

> travel patterns of recreational boaters, which carry implications for maritime traffic routing. Using a questionnaire with thirty-nine items, she surveyed five hundred and forty-three recreational boaters about their boating practices and preferences. The analyst wanted to better understand the correlation among the thirty-nine dimensions of her survey and to group correlated dimensions together. To learn about the correlation of dimensions, she used PCA to provide her with a rough understanding of which dimensions she might group together. After manually assigning dimensions to groups, she aggregated the groups. Then she used both the aggregation and the dimensions of a group to describe different recreational boating patterns.

Compare groups of old dimensions: In eight usage examples, datasets were comprised of groups of dimensions. These groups might be known a priori, or might be manually created by an analyst during the analysis process, such as by grouping correlated dimensions as described above. The most typical case is when a dataset was produced by a predictive simulation model, where there exists a group of independent, or *input* dimensions and a group of dependent, or *output* dimensions[7]. When groups of dimensions exist, one common task is to *compare these dimensional groups*, which might or might not coexist with a need for DR.

> In the NPAlgo usage example, a computer scientist (A15) was interested in using empirical methods to construct good prediction models for algorithms solving NP-hard problems, such as the traveling salesman problem. He measured the time required to run an algorithm for a NP-hard problem across many different parameter settings, where the settings are regularly sampled from the available range.

[7]Different communities use different vocabularies for these types of dimensions with independent/dependent, cause/effect, and input/output being the most common ones. We opted for input/output because it was frequently used by our interviewees and is common for describing simulation models.

> This process results in a dataset with between one hundred thousand and one million points, with two groups of dimensions: a group of approximately one hundred input dimensions representing parameter choices, and the measured runtime as a single output dimension. The dataset is then used to train a predictive model for the algorithm at hand. Although the resulting prediction model works well, humans have a hard time understanding this one-hundred-dimensional feature space. The analyst sought a way to reduce the dataset to a lower dimensional set of five-ten input dimensions, with a resulting model that retains nearly the same predictive power as the one that uses the entire feature space. The data analyst knew that the input dimensions are highly correlated to each other, so he wanted to use a dimensional synthesis technique.

Abstractly, the goal of this analyst can be described as getting insight into *groups of dimensions* and to synthesize *new dimensions* without diminishing predictive power. In Section B.6.5, we will further discuss this usage example and see that this combination is not well supported by current state-of-the-art DR algorithms.

In five of the eight usage examples involving groups of dimensions, there existed two groups of dimensions, having an input/output relationship. The remaining three usage examples involved the more general case of n groups of dimensions.

> The StrucGen usage example involved a bioinformatician (A19) with three explicit groups of dimensions in his protein dataset: one group pertaining to proteins' three-dimensional structure as described by high-dimensional feature vectors, one describing proteins within a hierarchical organization of functional properties, and another pertaining to the ordering of amino acid sequences within these proteins. His task was to identify similarities and differences across these groups of dimensions. He independently reduced the dimension of all three

groups by manually filtering them, and then sought to understand the relationships between them.

In general, comparing groups of dimensions can stem from different interests of an analyst. A typical interest is to learn about how groups correlate to each other. Another example is sensitivity analysis, where the goal is to understand whether small changes within one set of dimensions lead to small or large changes in another set of dimensions [29, 240, 248]. Sensitivity analysis is particularly common with predictive simulation models, to understand the sensitivity of output dimensions to changes on input dimensions. A recent example of use cases in which sensitivity analysis is combined with DR can be found in Bergner et al. [21]. While these tasks can co-exist with DR, they can also occur without any need of DR. In Section B.6.5, we will discuss one such usage example.

Name new dimensions: Tasks involving an interest in new dimensions necessarily imply the usage of dimension synthesis techniques. One common task is to *name new* dimensions, which we found in eight usage examples. Here, the analyst attempts to ascertain the semantic meaning of the proposed new dimensions. A common approach is to inspect the original high-dimensional data plotted within the context of the new low-dimensional layout in a scatterplot, wherein the analyst may be able to discern an interesting semantic relationship along the low-dimensional axes: the task of *naming new* dimensions is often used in DR algorithm papers [41, 313], in particular those about non-linear DR techniques:

> In the FACES usage example [313], a dataset of human face images, with each image described by a vector with four thousand and ninety-six, was reduced to three dimensions to showcase the capabilities of the Isomap algorithm. In the low-dimensional space, it became possible to plot the images as thumbnails in a scatterplot, whereupon three meaningful dimensions were identified: up-down pose, left-right pose, and illumination.

While many of the *naming new dimensions* examples are from the literature (four usage examples), we also detected these tasks in our field investigation (four usage examples).

Unmap new dimensions to old dimensions: Another task associated with new dimensions is *unmapping* new dimensions to corresponding old dimensions. Unmapping can be conducted in a straightforward fashion with linear techniques, by inspecting the extent to which any particular old dimension contributed to the synthesis of a new one. In PCA, unmapped old dimensions are often referred to as the "loading" of the (new) dimensions [161].

> The Music analyst (A1) was interested in the listening behavior of digital music consumers. She gathered listening history data and demographic information from approximately three hundred people who use the last.fm music streaming service[8]. The resulting dataset had forty-eight dimensions, which included continuous dimensions, such as the number of tracks streamed per day or per login session, as well as categorical dimensions, which included the gender and geographical location of a person using last.fm.
>
> In the Music B usage example, she engaged in finding important dimensions and consolidating dimensions by correlating them into groups. With a set of grouped dimensions she intended to characterize and describe important factors in music listening behavior. She used PCA and examined the first thirteen principal components, which accounted for seventy-five percent of the variance. Her goal was to account for as much of the variance of the original data as possible with a small number of new dimensions, while maintaining an understandable semantic mapping between old and new dimensions. By closely analyzing this unmapping, she identified proper names and meanings for the new synthetic thirteen dimensions. She did not visualize her data in this analysis.

[8] `http://last.fm/`.

With linear DR techniques, such as PCA, the relation between old and new dimensions is clear. However, most non-linear techniques, while offering a more powerful reduction, do not support unmapping: the mapping that occurs between old and new dimensions is hard to interpret.

The unmapping approach is a top-down approach, in which first a DR synthesis algorithm is used and then the loadings of the new dimensions are inspected. In contrast, a bottom-up approach can lead to similar results. Consider, for instance, the BoatAct usage example, in which A3 first learned about the correlation of the old dimensions, grouped them together accordingly, and then aggregated them into new synthetic ones. Both approaches can lead to similar outcomes.

No correlation among new dimensions: While we found that determining correlation between old dimensions is an important task, we do not include correlation between new dimensions as a task. We did not find any instances of analysts with this interest. We are not surprised by this finding, because many DR techniques produce new synthetic dimensions that have as little correlation between them as possible; the best-known example is PCA. In these cases, searching for correlation between new synthetic dimensions is unlikely to yield meaningful results. Nevertheless, we call out this issue, in particular because we have seen arguments in the visualization literature that confound what we consider to be two very different uses of scatterplots. The use of scatterplots to show correlation in data that is not dimensionally reduced should not be conflated with the use of scatterplots for dimensionally reduced data to name new dimensions or check for cluster separability.

DR for Checking Cluster Separability

DR-related clustering tasks pertain to the identification and verification of separable point clusters in the data (twenty usage examples), reflected in the middle branch of Figure B.9. In terms of visual data analysis, clustering is strongly tied to visualizing dimensionally reduced data in scatterplots: eighteen of the twenty usage examples associated with an interest in clusters involved the use of scatterplots.

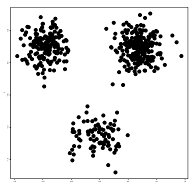
If there is no *class structure* available, these scatterplots are typically shown in monochrome and a common task is to identify and *name separable blobs*, and to see if these blobs match the analyst's mental model of the dataset. By separable blobs, we refer to proximity relationships in the lower-dimensional layout of points, forming a distinguishable structure as shown in Figure B.10a.

On the other hand, a dataset often comes with an associated class structure, which is then typically shown by color coding points according to their classes. These classes might come directly with the data, be assigned using a clustering algorithm run by the analyst, or be the result of manual labeling. Here, we found that the predominant reason that people look at scatterplots is to check if separable blobs match with the color-coded classes as shown in Figure B.10b and Figure B.10c. Note that we previously found [285] that the visual separability of color-coded clusters differs perceptually from the separability of monochrome clusters.

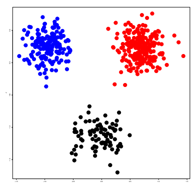
The more common situation however is that the visualization of dimensionally reduced data does not lead to clearly separable blobs. Rather than clearly visible structure, reducing the dimension of the data and plotting it in a scatterplot often leads to an undifferentiated clutter of points, as shown in Figure B.10d and Figure B.10e. In such situations, analysts often engage in iterative *refining of DR, visual encoding, and/or clustering* algorithms to obviate algorithmic artifacts as a reason for the non-visibility of cluster structure.

Name blobs: The *name blobs* task, which we observed in fifteen usage examples, describes a person's intent to identify visually separable blobs of (monochrome) points in a scatterplot of dimensionally reduced data and to assign meaning to these clusters. If blobs are visually separable in a low-dimensional projection, as shown in Figure B.10a, an analyst could assume that these blobs represent meaningful groups.

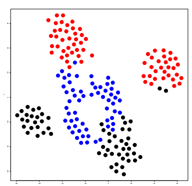
> In the TxtDocs usage example, a journalist (A10) was interested in characterizing patterns of incidents reported by de-



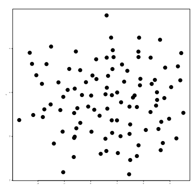
(a) Naming separable blobs of unlabeled data.



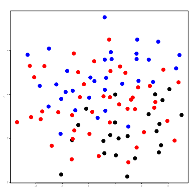
(b) Matching class structure with visible blobs.



(c) A mismatch of class structure with visible blobs.



(d) An undistinguishable clutter of unlabeled data.



(e) An undistinguishable clutter of labeled data.

Figure B.10: Example scatterplots of dimensionally reduced data illustrating potential tasks related to checking cluster separability. Cases (d) and (e) usually lead to refining DR, visual encoding, or clustering algorithm choices and parametrizations.

fense contractors during the Iraq war. Several thousand hard copy incident reports were scanned and subject to optical character recognition, resulting in an unstructured text file for each report. Each report was converted into a weighted vector of words, and the cosine similarity measure was computed between pairs of these vectors. The distances between vectors served as input to Glimmer MDS [149], which generated a monochrome two-dimensional scatterplot visual encoding of the documents.

Overlaid on these points were the top weighted terms from their corresponding vectors.

The journalist observed visible structure in this visual encoding, and was able to identify clusters of incidents and their distinguishing features: the involvement of civilians, official dignitaries, or combatants, the presence and type of casualties, the locations, dates, and times of incidents, and whether the incident involved the use of weapons or vehicles.

Match blobs and classes: If classes are available for color-coding the points and if visually separable blobs are visible as in Figure B.10b and Figure B.10c, a typical task is to evaluate the match between blobs and class structure: that is, checking whether the colors match with the spatial structure in the layout.

The MoCap analyst (A7), a machine learning researcher interested in building predictive motion capture models based on a dataset with given *class structure*. Using forty-five carefully calibrated accelerometers, gyroscopes, and magnetometers, each attached to a part of the body of human subjects, he captured instances of human motion: walking, standing up, lying down, and so on. Instances of the same motion were grouped together in the same class, and represented the ground truth class structure of his motion capture recordings. From these sensors, a large number of time-varying derived variables (approximately 25 per sensor) were recorded, resulting in datasets of approximately one thousand dimensions and approximately ten thousand recorded motions (points).

In usage example MoCap A, A7 was interested in verifying that the classes of motion types formed visibly distinct blobs within the layout of points in a scatterplot. He reduced the data using linear PCA or SFS automatic filtering [155] to either two or three dimensions, then visualized the result in two-dimensional or three-dimensional scatterplots, with points colored according

> to their motion class. However, regardless of the DR technique used, the colored blobs did not unambiguously match the class structure of points.

Another common case is that analysts start out with exploring an unlabeled dataset and then uses a clustering algorithm to suggest a certain class structure. In this case, the analyst starts out with an initial *name blob* task and then engages in *matching* the suggested class labels with the visible point layout (six usage examples). Automatic labeling by way of a clustering algorithm is therefore used to facilitate the detection of clusters that exist in the high-dimensional space.

> Before the MUSIC analyst (A1) shifted her focus to finding and grouping important dimensions as described above (usage example MUSIC B), she was interested in identifying and naming clusters of listeners (usage example MUSIC A). Her particular interest was to cluster people into groups with similar listening behavior. She hypothesized several groups a priori, such as people who listen to the same music repetitively vs. people who listen to new music, or people who listen during the day vs. those who listen at night. She used PCA to reduce the dataset, which she plotted using monochrome scatterplots. Unable to perceive meaningful cluster structure in the monochrome plot, she applied a k-means clustering to generate colored scatterplots of the low-dimensional data. However, even after repeated changes to the size of the k parameter, the number of clusters, she was not able to identify any meaningful cluster structure. She gave up on her clustering endeavors and switched to learning about her dimensions as described above.

Refining DR, visual encoding, and clustering: Low-dimensional projections rarely represent the full details of a high-dimensional dataset and different algorithmic choices and parametrizations lead to different visual encodings of the data. In practice, identifying separable clusters has therefore to be seen as a spectrum between seeing no visible class structure at all, as in Figure B.10d and Figure B.10e, to clearly seeing separable point clusters, as in Figure B.10a and Figure B.10b. A very common task in this spectrum is to iteratively *refine DR, visual encoding, and/or clustering* algorithms by selecting alternative choices and parametrizations, with the goal of seeing clearer separability. In the following usage example an analyst tried different DR algorithms and parametrizations, toward the goal of identifying clusters:

> In the CONCEPT usage example, A6's goal was to visualize research concept clusters, to be used by life science researchers, allowing her to identify higher-level concept clusters and other researchers working in areas related to their own. The dataset stemmed from a database of all life science researchers in which each researcher is represented by a set of ranked research concepts. Overall there are twenty thousand concepts, including terms such as "DNA" or "cancer".
>
> As her eventual goal was to build a helpful visualization tool, her immediate task was to identify groups of concepts. To do so, she computed a distance matrix of concepts based on their co-occurrence in the researcher database. For visualization, this matrix was then used as input first to classical MDS and then to Glimmer MDS [149], the latter with a variety of parametrizations. Yet, no approach revealed any meaningful visible cluster structure, but only an undifferentiated clutter of points. The analyst finally decided that the non-existence of visible separation is not just an algorithmic artifact. Trying different algorithmic choices and parametrizations gave her a certain degree of confidence in her decision.

Judging whether non-visible/visible cluster structure correctly reflects the high-dimensional data or whether it stems from algorithmic artifacts is a major challenge we observed. To address this challenge, all our interviewees concerned with checking point clusters, including the TXTDOCS (A10), MoCap B (A7), and Music A (A1) usage examples described above, engaged to some degree in *refining*.

Abstractly, this central question can be phrased in terms of whether the presence or absence of visible cluster structure is a *false/true negative/positive* result. An undifferentiated clutter of all points is the most common situation, which could be either indicative of a false negative or a true negative. If an analyst suspects a *false negative*, she will probably decide to refine algorithmic choices and/or parametrizations; the analyst might, for instance, decide to use non-linear DR techniques instead of linear techniques, or to use a SPLOM instead of two-dimensional scatterplots. The goal of this refinement is to move incrementally from few or no visible clusters toward more visible structure.

Alternatively, an analyst may consider an undifferentiated clutter of points to be a *true negative*, meaning that there is no cluster structure in the dataset. This consideration often occurs after an analyst has tried many different technique and parameterization choices. The Concept usage example (A6) is an example where the analyst eventually gave up.

When clear clusters are visible, an analyst will often declare victory, as the spatial layout indicates a *true positive*. In contrast, there may be instances when an analyst mistrusts the visible structure, attributing it to an artificial algorithmic artifact (a *false positive*). These situations are probably less common; we observed none in our study.

DR for Algorithmic Input

While all of the above tasks were about using DR for the purpose of data analysis and visualization, another purpose of DR is algorithmic input, a common case in machine learning. We identified seven usage examples in which DR was used for preparing a dataset for algorithmic input, reflected by the right-most branch of Figure B.9. In these usage examples, the goal is to reduce the dataset's dimensions in order to improve the performance of downstream algorithmic processing, and to make a predictive model more reliable, robust and accurate by avoiding the curse of dimensionality [20].

This task is well-described in the machine learning community [222].

Earlier, we described the MoCap A usage example (A7), where the analyst strived to *match blobs with classes* of motion types in two-dimensional or three-dimensional scatterplots. In a larger context, his ambitions to visually analyze the data were one step toward his ultimate goal of building a motion classifier, which involved performing DR for the purpose of algorithmic input:

> In the MoCap B usage example, A7 wanted to use his motion capture dataset to train supervised motion classification algorithms. Once again he used PCA or SFS to reduce the number of dimensions in his dataset. After manually inspecting scree plots, he selected roughly thirty dimensions to train a supervised motion classification algorithm. Despite his failed attempt to visually verify groups in his data (usage example MoCap A), he was sufficiently satisfied by the performance of his classifier trained with the 30-dimensional reduced data that he considered these groups to be a true positive result. His results from MoCap A can therefore be seen as a false negative.

High-dimensional Data Analysis and DR

Although DR was the focus of our investigation, many of the tasks we found might be also conducted without the direct need for the application of DR algorithms. The only tasks we found where DR is inherently necessary for are tasks about new dimensions. All other tasks can be seen as a subset of general high-dimensional data analysis for which we found instances that coexisted with DR usage. Naturally, there will be many other abstract tasks if the scope is broadened to high-dimensional data analysis. We see our work as a first step toward a better understanding of those tasks “in the wild”.

We found that the question whether DR will or will not be a valuable tool for the analysis of a high-dimensional dataset is often not straightforward for an analyst or designer to answer. In particular, we observed two instances in which our interviewees attempted to use DR after our initial interview, but eventually realized that no DR was needed for the purpose of accomplishing the analysis of their high-dimensional data. In the following usage example, sensitivity analysis alone without DR, for instance, was the actual solution:

> The FishPop usage example involves a biologist (A11) whose goal is to provide recommendations about balancing the risks of overfishing with commercial and private fishing interests. She compares and evaluates several different models that simulate the behavior of fish populations. All these models take a set of input parameters, such as carrying capacity and productivity, typically generated via regular sampling in the space of possible parameter configurations. The output of these models is an indication of the probability that a fish population will die out. Her main concern is sensitivity analysis: checking whether small changes in input dimensions lead to small or large changes in output dimensions. Sensitivity is a main aspect of recommending one model over another. Although she experimented with DR algorithms after our initial interview, she ultimately resolved that no existing unmodified DR technique was appropriate or even necessary for her dataset and task at hand, and did not pursue this line of analysis further.

We include this discussion here to illustrate that some analyses may not benefit from DR. While many analysts and designers might be quick to assume a need for DR, it may be inappropriate or even present confounding results. Our suite of DR-related tasks offers more clarity and guidance for those who are facing such decisions at the interplay of high-dimensional data analysis and DR.

B.6.5 Challenges

We observed several challenges encountered by people when using DR for their data analysis endeavors. Their struggles are of interest as stumbling blocks in daily practice that affected multiple people, and are not necessarily a sign that there is no solution in the technical literature. We discuss two of them in detail: the need for DR that explicitly handles different groups of dimensions, and the need for nonlinear techniques that support unmapping new to old dimensions.

DR with groups of dimensions: NPAlgo (A15), StrucGen (A19), and FishPop (A11) were some of the usage examples that involved comparing groups of dimensions. Some analysts, such as in the FishPop usage example, perform this comparison and have no need for DR at all, and sensitivity analysis alone may be the appropriate task for them to perform. However, other analysts have an additional need to reduce the dimensions of their data. This combination does not inherently result in a problem. For instance, the StrucGen analyst (A19) was able to meet his needs with DR in the form of filtering; he reduced the number of dimensions in each of his three groups of dimensions independently from one another, and then compared the reduced groups.

However, in the NPAlgo usage example (A15), which involved a predictive model of NP-hard algorithms, groups of dimensions possessed an inherent and important dependency on one another. In particular, the analyst observed cases where a specific input dimension contributes very little to the overall variance nevertheless had a huge impact on the output dimension run time; converse cases were also observed. In these cases, it is not useful to apply a DR technique solely on the input dimensions, because the crucial information about the intrinsic relation of the input dimensions to the output dimensions is not considered in the DR computation. Neither can a synthetic DR be run simultaneously on both input and output dimensions. The importance of original input variables with regard to their influence on the output variable is not acknowledged by common DR metrics, such as variance as in PCA, or stress as in MDS. In general, this problem may occur in situations with n groups of dimensions and an arbitrary arrangement of dependencies between these groups.

A first approach toward solving this problem was outlined by Gerber et al. [109]. Their focus is on continuous high-dimensional input manifolds that map to one output variable. In brief, the high-dimensional input space is partitioned into monotonic areas that are separately projected to two dimensions using PCA. These two-dimensional input parts can then be plotted together with the one output dimension in three dimensions. This approach allows people to visually explore relations between a group of multiple input dimensions and one output dimension. This approach, however, does not per se provide a DR technique that would handle interdependent dimensional groups appropriately as needed by the NPAlgo analyst (A15). A statistical approach toward that goal is canonical correlation analysis (CCA), a method for identifying highly-correlated linear combinations of both input and output dimensions [140]. As with PCA, CCA is limited to producing linear projections of the input data. Furthermore, the dimensionality of CCA projections can be no greater than the smaller dimensionality of the input and output groups. In case of the NPAlgo (A15), reducing the input space to one linear dimensions would not have been satisfying for the analyst. We see ample opportunities for future technical contributions in this set of challenges.

Non-linear unmapping: Our study emphasized that some analysts are interested in both *old* and *new* dimensions at the same time. Identifying these relationship often takes a top-down approach: a synthetic DR algorithm is used and the analyst tries to *unmap* these new dimensions to their old dimensions, such as in the Music B example (A1). Alternatively, some analysts take a bottom-up approach, in which they manually *group correlated* dimensions together and then aggregate these groups into new dimensions, such as the BoatAct example (A3). When analysts performed this task, they used linear techniques such as PCA because there is no support for unmapping with any nonlinear DR technique.

However, in cases such as usage example Music B, the A1 would have benefited from a more powerful nonlinear DR technique. The analyst wanted to account for as much of the variance of the original data as possible with a small number of new dimensions, while at the same time maintaining an understandable semantic mapping between old and new dimensions. With linear DR techniques, this criterion was only partially satisfied, due to their limited power.

While we are well aware that unmapping non-linear combinations is a difficult undertaking, even a partial solution would improve the state of the art. Interactive visualization may well be a fruitful avenue to pursue, for helping people explore a complex non-linear dimensional mapping space.

Other challenges: Analysts reported several other challenges that have been previously identified in the literature, including DR scalability issues with large datasets [32, 149], extreme ratios between dimensions and points with many more dimensions than points [73, 351], and the existence of categorical dimensions in the data [103, 120].

We also found supporting evidence for our previous hypothesis on the need for guidance [150]. While all of the analysts were experts in their own domain, thirteen analysts indicated a lack of expertise in the mathematical implications of DR techniques, while only eight analysts indicated deep understanding.

B.6.6 Benchmarks

One of our initial motivations for this project related to an apparent gap between benchmark DR datasets used in different communities. In this section, we discuss this gap and show how our work adds clarification by defining tasks.

A canonical DR benchmark often referred to in the machine learning literature is the synthetic *Swiss roll* dataset that is a simple two-dimensional strip rolled into a three-dimensional spiral shape. It shows off the capabilities of the *manifold following* non-linear DR techniques such as Isomap [313] and Laplacian Eigenmaps [19], which assume that the high-dimensional data lies on a densely sampled manifold that can be "unrolled" to a lower-dimensional representation. In contrast, cluster datasets of the type shown in Figure B.10 typify those described by Sedlmair et al. [285]. Interestingly, these cluster datasets appear to completely violate the manifold assumption. We were initially skeptical that methods optimized for manifold datasets would work well on other datasets, and wondered about the contexts in which these datasets were representative of analysts' practices and needs.

Our field work sheds light on this question: both benchmark dataset types are represented by real-world scenarios, but these scenarios have very different underlying tasks and associated assumptions. In particular, the assumption of a dataset lying on a densely sampled, single smooth manifold is strongly tied to the task of *naming new dimensions*.

We note that none of our interviewed analysts chose to apply manifold techniques; the two usage examples where these techniques were applied (BRDF, FACES) were from the literature [201, 313]. In instances of the *name new dimensions* task that emanated from our interviews (such as CONCEPT / A6 or MUSIC B / A1) it was not clear that the assumption of a smooth manifold was met by the data. In all other usage examples, the task itself did not align with manifold assumptions.

We now elaborate further on the relationship between these tasks and DR algorithm assumptions to provide guidance to visualization practitioners.

Densely sampled: Manifold techniques assume densely sampled data. However, analysts may not always be sure if their datasets meet the assumptions of a densely sampled manifold. We derived two criteria for understanding when a dataset is likely to qualify as a densely sampled manifold. First, all dimensions should be numerical and continuous, which negates any datasets containing categorical dimensions; the survey datasets from the CONCEPT, MUSIC (A6), and BOATACT (A3) usage examples, for instance, violate this assumption. Second, the dataset should be generated by a process that has the characteristics of continuous sampling. For example, in the MOCAP usage example (A7), sensors measured body part motion over small time intervals; however, the analyst's tasks did not include *naming new dimensions*, but rather *matching blobs and classes* and reducing the data for *algorithmic input*. Real-world measurements are a common case for manifolds, but manifolds may exist elsewhere, such as in the NPALGO usage example (A15), where the regularly changing values reflected algorithm parameters.

Single vs. multiple manifolds: In addition to densely sampled data, the classical *Swiss Roll* manifold unrolling use case comes with a further assumption: the data resides on a *single manifold*. Formally, the data distribution along all continuous dimensions needs to be homogeneous, rather than heterogeneous or clumpy. The *Swiss Roll* benchmark dataset is the canonical example for a single manifold; here the associated task is one of carefully unrolling the single manifold. Clearly, this assumption matches only with the needs of analysts who perform the *name new dimensions* task. In particular, such assumptions are less likely to be appropriate for tasks involving point clusters. Point clusters may exist due to a non-uniform distribution of samples on a single manifold. Alternatively, clusters may represent samples taken from multiple different manifolds. The data from the MoCap example (A7) is an instance where a multiple manifold structure is likely, with one densely sampled manifold per class of movement type. The machine learning community has noted the instability of older single manifold algorithms [14, 332], and newer techniques such as t-SNE [331] have been proposed for identifying multiple manifolds, thus supporting tasks relating to point clusters. However, the question of whether a dataset reflects the result of dense sampling along continuous dimensions remains.

B.6.7 Discussion

Our work provides a synthesis description of DR-related tasks that is both accessible to a visualization audience and grounded from a systematic study of real-world human behavior. In particular, we emphasize a differentiation between tasks involving *point clusters* and those involving *dimensions*, the idea of *false/true positive/negative* groups of points as a way to think about faithful visual clustering, and of *old* and *new dimensions* for framing dimension-related tasks.

Being a descriptive classification of human behaviour, our main contribution is not a radically new perspective on a problem. Our classification of tasks provides an abstract and structured description and evocative vocabulary for talking and thinking about high-dimensional data analysis and DR tasks from a visualization usage point of view. All our findings are grounded in observations of DR practices "in the wild", adding a new and usage-based perspective to the large body of technique-driven DR literature.

We now discuss some connections between our classification and the terms and ideas in previous work. The relationship between clusters and dimensions has been explored in depth in machine learning, which has existing vocabulary for a related set of tasks. In particular, the supervised context of machine learning is divided between *classification*, which maps input data dimensions to a discrete class variable, and *regression*, which maps input data dimensions to a continuous variable or set of variables [222]. This discrete-versus-continuous division has a rough analogue in the set of tasks described in Section B.6.7. Checking cluster separability involves confirming that there are discrete classes of data and is therefore similar in spirit to classification. Learning about dimensions, especially in naming new dimensions, is about characterizing the continuous output of newly synthesized dimensions, and similar to regression.

Likewise, clustering tasks have been widely discussed [156, 363], where clustering can be considered as unsupervised classification of unlabeled data. Furthermore, our division into old vs. new dimensions echoes the division of feature selection vs feature extraction, again in the vocabulary of machine learning. Venna and Kaski's NeRV method [336] explicitly strives to balance the maximization of true positives with the minimization of false positives in the placement of data points in the reduced set of dimensions. Their work is implicitly focused on a low-level task of examining local point neighborhoods, in contrast to the higher-level tasks in our classification.

A limitation of this work is its restricted scope. We chose to focus on questions about DR techniques in particular, but high-dimensional data analysis includes many related methods that we did not specifically investigate. While we are familiar with a considerable amount of the previous work and have experience in developing DR algorithms and systems [147, 149–151, 357] and evaluating them [285, 323], we do not claim to have all-encompassing knowledge. Our perspective is centered on visualization issues. As is inevitable with qualitative research [55], our previous experience is a lens that influenced which participants we invited to our study, how we questioned them, which papers we selected and read, how we coded the data, and what we decided to report in the manuscript. Our discussion is necessarily limited to the behavior that we encountered in a usage example, given our methodological approach where the findings must emerge from and be grounded in the data. We thus do not claim complete coverage of all possible and relevant usage patterns and tasks involving DR and high-dimensional data analysis.

We see our work as a first step toward a better and more systematic understanding of data analysis in the wild and hope that others will build upon our work, propose alternative methodological approaches, extend our collection of tasks with new findings, and broaden our understanding with new and different perspectives.

B.6.8 Conclusions

We have presented a classification of tasks derived from analyzing DR usage in the wild that provides an abstract understanding of the processes of real-world analysts. The combination of abstract tasks and the usage examples that underlie them serve as a task-centered lens on DR, complementing and connecting to the rich corpus of technique-centered DR literature. We also relate common benchmark datasets to these tasks, and discuss challenges encountered by analysts in practice. We hope that this task-centered approach to high-dimensional data analysis encourages others to continue in this methodological spirit.

Appendix C

Field Study Supplemental Material

This appendix supports Chapter 4. It contains the full scatterplot from the WARLOGS initial use case (Fall 2010, a detail was provided in Figure 4.3), a screen shot of the *Overview v1* prototype (Fall 2011), our original project proposal (Spring 2012), our interview protocol (Spring 2012), preliminary field study results (Fall 2012), and a screen shot of *Overview v3* (Fall 2012).

C.1 Initial Use Case

The initial WARLOGS use case, described in Section 4.3, resulted in the scatterplot shown in Figure C.1 (a detail is also shown in Figure 4.3), which was produced by Jonathan Stray in Fall 2010 [305].

C.2 Overview v1

Overview v1, a prototype desktop application completed in Fall 2011, is shown in Figure C.2.

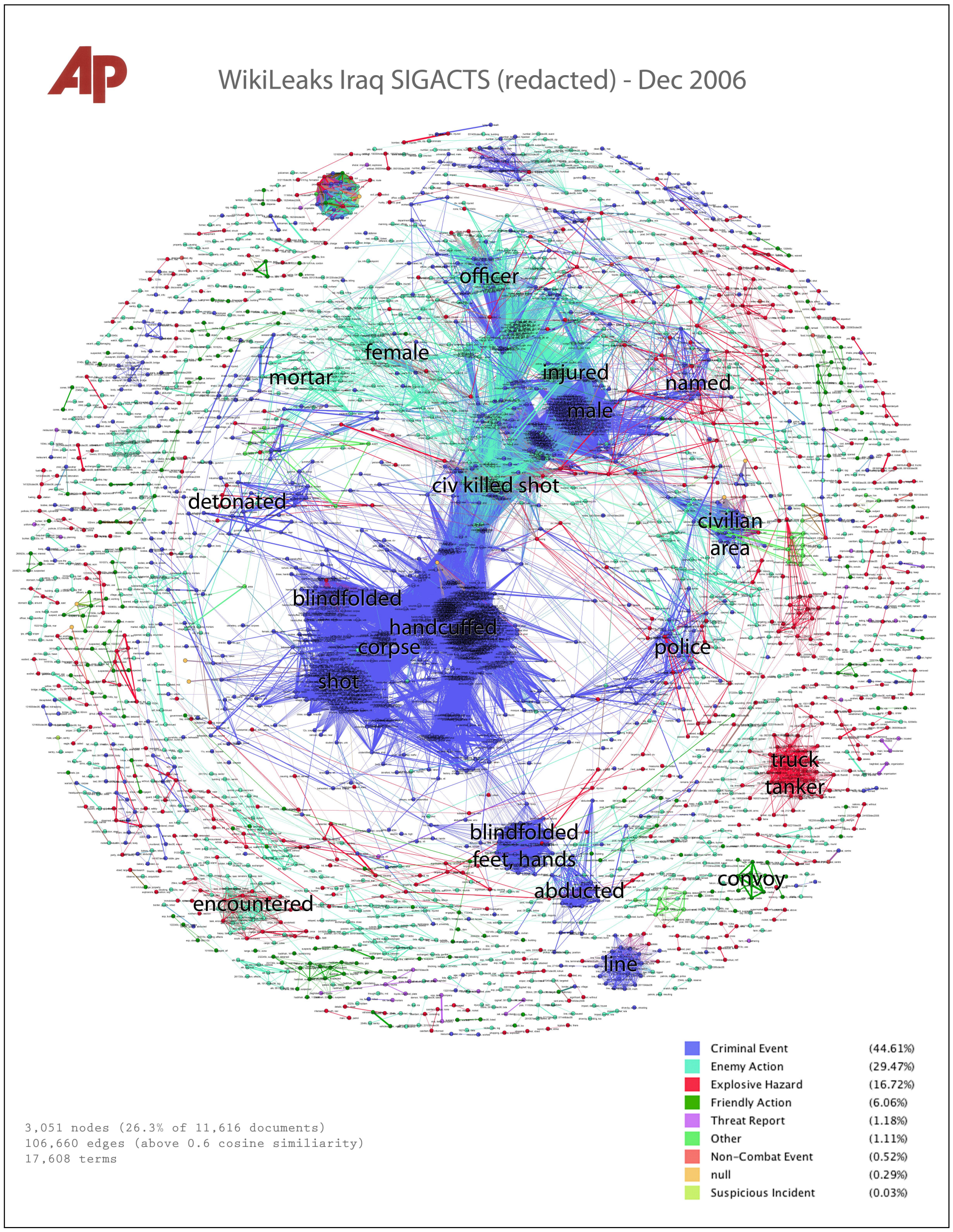


Figure C.1: *"A full-text visualization of the Iraq War Logs"* (the WARLOGS initial use case) by Jonathan Stray [305].

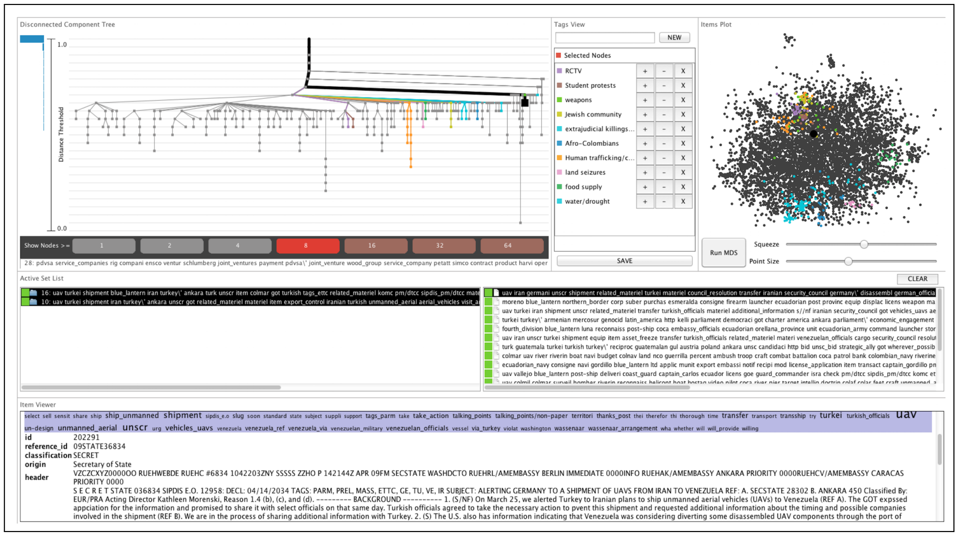


Figure C.2: *Overview v1*, a prototype desktop application completed in Fall 2011.

C.3 Field Study Proposal

[1]In the fall of 2010, WikiLeaks released 391,832 text documents relating to the conduct of American armed forces and independent security contractors during the war in Iraq. Since that time, specialized investigative *data journalists* have reported on what was contained in this vast deposit of documents, which included startling information regarding civilian casualties, friendly fire incidents, and observed breaches of protocol. My proposed research, in brief, asks how data journalists *mine* such deposits, how they seek information to support or refute prior findings, or how they come to discover unanticipated yet newsworthy stories hiding in the data.

The motivation for this research proposal reflects the growing trend of large collections of emails, reports, and other documents being dumped, leaked, released, or declassified by corporations, government agencies, and other organizations, such as WikiLeaks. Fellow journalists and the news-reading public deserve transparency when it comes to the methods of data journalists who investigate these collections. Additionally, developers of

[1]This research proposal for the *Overview* field study was written in April 2012.

data analysis applications require a better understanding of journalists as potential adopters.

Aside from being a consumer of news media, I have no background in journalism. Rather I represent the interests of application developers, and I am sensitive to theoretical concepts relating to the design of tools for supporting data analysis. While this theoretical repertoire is useful for understanding low-level perceptual activity contributing to how individuals interact with data displays [8, 291], or for understanding high-level domain-agnostic abstractions relating to information foraging and sensemaking [7, 241], I am faced with a gap in the middle. How can I characterize the data analysis process within the context of a domain such as journalism? Moreover, how does this process play out with a specific type of data, in this case being collections of text documents? Thus I seek a middle-level theory to explain this process.

C.3.1 Research Questions

My predominant research question asks: What is document mining? That is, how do journalists conduct data analysis when faced with more documents than they could possibly read in a year, let alone in time to meet a deadline?

Several additional questions follow from this: namely, what constraints do these journalists face in the process of their work? What tools do they use and how do they use them? Do they collaborate with other people, and if so, how do they collaborate?

Finally, this work will determine how document mining compares to other analytical processes in data journalism, when the data is comprised of numbers rather than text documents, which could include large financial databases or historical measurements. I will also make comparisons between document mining and other processes of investigative journalism, as well as with processes of data analysis characterized in other domains, such as business intelligence and law enforcement.

C.3.2 Research Context

Journalists engaging in document mining are found in newsrooms around the world. Unfortunately, like many busy professionals, they are often working under tight deadlines, and have little time to participate in academic research. However, I am lucky enough to know someone "on the inside".

Jonathan Stray is a computer-scientist-turned-journalist now employed by the Associated Press, based out of New York City. Working in collaboration with my research group, he has developed *Overview*, a robust data visualization application for document mining, recently made available as a free download[2]. He is currently[3] pitching *Overview* to journalists via conferences, workshops, and social media. Buzz surrounding *Overview* is starting to grow in the data journalism community.

Stray is our gatekeeper to research participants, as his potentially useful application provides an incentive for journalists participate in our research. As a result, we have an opportunity to satisfy two research goals: (1) assess whether *Overview* is usable and useful, as well as how it fits within existing document mining workflows; and (2) characterize the process and context of document mining, with and without this new application. While this proposal focuses on the latter goal, data collection corresponding to both goals will occur simultaneously. Furthermore, it is my intent that in working toward the second goal, my findings will contribute to the future development of *Overview* and other applications like it.

Due to the distributed nature of this research, logistical constraints will keep me from visiting individual journalists in newsrooms. Thus my data collection will occur at a distance, over the phone and online.

C.3.3 Methodology

A need to characterize the process of document mining among journalists necessitates a grounded theory approach [55]. This approach is in turn informed by an interpretivist, symbolic interactionist theoretical perspective

[2] Recall that *Overview v2* was a desktop application (`https://github.com/overview/overview-prototype`), which was later replaced by the *v3-v4* web application.

[3] As of Spring 2012.

and a constructionist epistemology [68]. That is, I intend to focus on the language used by journalists to describe this process, and construct a shared interpretation of this process based on interactions with research participants and the data they generate.

The constant comparative method of grounded theory will allow me to flexibly make comparisons between the process of document mining with other journalistic processes, as well as processes relating to data analysis in other domains. Comparisons will also be made between journalists, between newsrooms, and over periods of time. For instance, I will be comparing the process of document mining both before and after the introduction of *Overview*, the new visualization tool.

A further justification for the use of a grounded theory methodology is that my initial research questions are not theoretically deduced hypotheses. Rather, my questions are informed by sensitizing concepts and assumptions held within my domain. It is these sensitizing concepts that allowed me to frame the data collection methods, particularly a preliminary set of interview questions.

These sensitizing concepts include the notion that data analysis, document mining being an instance of which, occurs in stages. Data analysis may involve stages of hypothesis generation, each necessitating an exploration of the data without a particular set of questions in mind, save perhaps "What's going on here?" Other times there may be stages of hypothesis validation, where the goal is to support or refute prior evidence. These stages necessitate a directed search within a subset of the data, or a comparison between individual items or documents. Individuals may or may not engage in both types of stages during the course of a single investigation.

The products of data analysis are also among my sensitizing concepts. These products include *"Eureka!"* moments of insight, serendipitous discoveries, and both optimal and suboptimal solutions to closed- and open-ended problems. Admittedly, these products of analysis are ill-defined constructs, and it will be necessary to attain our research participants' interpretations of their meaning, as well as their native terminology.

Finally, these sensitizing concepts include the disentanglement of an individual's expertise. By this I mean that an analyst may have expertise within a domain, expertise using specific analytical tools or techniques, and/or expertise regarding the data, its semantics and its provenance.

In my field of research, there exist several precedents for the use of grounded theory, or at least the use of methods inspired by a grounded theory methodology. The methodology has informed prior work which has characterized the data analysis processes of professionals in other domains, including architecture [322] and national security and intelligence [164]. There also exists a grounded evaluation method for determining the effectiveness of visualization software when deployed in target contexts of use [153]. Both uses of grounded theory methods serve as inspiration for my proposed research.

Data collection methods and sources: My primary data collection method will be intensive, open-ended interviews with journalists. These interviews will be teleconferences or group Skype chats. Both Stray (in New York City), and myself (at UBC) will have questions to ask interviewees, with his questions pertaining to the usability and utility of *Overview*. Audio from these interviews will be recorded for later transcription.

Following the methodology of grounded theory, I will not specify the number of interviews that I plan to conduct a priori. The final number of interviews will depend on how much theoretical sampling is required before achieving data saturation, the point where no new categories emerge; I will return to this point below when I discuss data analysis methods. The number of interviews will also depend on how many journalists download and use *Overview*, and among those, how many express a willingness to participate in interviews. Ideally, I would like to perform multiple interviews with each journalist in order to make comparisons over time, as their processes vary or change over time, before, during, and after using the new tool. However, this is an unrealistic plan. As mentioned in Section C.3.2, these journalists will often be conducting their investigation and writing their story under a tight deadline, and will likely only have time to commit to an intensive interview after the story is written. As a result, I will rely on secondary data collection methods, such as follow-up email exchanges, to fill in some of the gaps. These methods are discussed in greater detail below.

Regarding the content of these interviews, I plan to keep the number of initial questions small. I have prepared a list of interview foci with a small set of representative questions for each, composed according to guidelines for open-ended interviews [100] and for interviews conducted in the context of a grounded theory study [55]. These foci correspond with the research questions mentioned above in Section C.3.1. In particular, I will attempt to ground the interview in the interviewee's example of document mining, one drawn from their prior experience. This will invite comparisons with other journalistic processes, as well as comparisons between their processes before and after their initial use of *Overview*.

Many questions are redundant and cross-referential, a deliberate choice, as I have no intention to ask all or even most of them in a single interview. An answer to one open-ended question is expected to answer many of the others; these foci and questions are more so a checklist than a script. I also expect this list of foci and questions to change as I conduct interviews, as a result of theoretical sampling and the possibility that early interviews will illuminate unanticipated themes and concepts.

I plan to complement the interviews by eliciting texts and other information from journalists that I interview. Follow-up questions will be asked via email. I will also request copies of the notes journalists take during the course of their investigation. I expect that in many cases, journalists will be taking notes regardless of whether or not I ask to see them. I will also request information regarding the data, such as how many documents are contained in the document collection being investigated, how they tend to vary from one another, how many were read or skimmed during the course of their investigation, how many were discarded or ignored, as well as why individual documents were read, skimmed, ignored, or discarded. In cases where these documents are publicly available, I will examine the documents as well. Screenshots or pictures of annotated documents, journalist notes, and other analytical artifacts, such as spreadsheets or charts, will also be requested.

Realizing the value of found data [294], I will also collect several extant texts. In particular, I will collect the stories journalists write as a result of their investigations. I will examine the extent to which the document mining process is transparent in their articles, allowing for a comparison with their notes and the remarks they make during interviews. Finally, in cases where these stories are published online, I will also collect the reader perspective, via comments and discussion boards.

Data analysis methods: Data collection and analysis for this project will occur concurrently. Interviews and artifacts collected from journalists will be subject to multiple iterations of coding, each calling upon the constant comparative method, the basis of grounded theory [55]. First, open coding will label the data, at the line or paragraph level, using words or short phrases used in the data. Next, tentative categories of codes will be generated, each with an explanatory rationale based on comparisons between code instances, recorded in memos. This will inform subsequent data collection and focused coding. The process of axial coding follows, a recoding of the data using the categories constructed. At this point, the process of theoretical sampling will direct me to specific data collection, using different interview foci or artifact collection criteria. As categories become refined and theoretical concepts emerge through the process of selective coding and memoing, I will begin to seek theoretical saturation, the point where no unexamined concepts are apparent. At this time, I will begin to construct a mid-level theory of document mining based on the relationships between concepts. This stage will also involve comparisons between my theory and other theories of data analysis, as it occurs in other domains and as it is described at higher levels of abstraction [7, 241].

Triangulation between researchers is highly effective during interpretive analysis [200]. I will share my findings with Stray. While he and I have different foci and research goals, we will both be engaged in participant interviews, and thus can compare notes. Additionally, his own journalistic expertise can also be called upon throughout the stages of my analysis.

I will also triangulate in terms of methods, in that I will take an alternative approach to thematic coding [265]. This will involve an examination of word frequency, word co-occurrence, key words as used in context, linguistic connectors, and metaphors used. Extant texts and artifacts collected will also be analyzed in terms of their descriptive properties, as well as their intellectual and cultural values [250]. I will then compare the codes produced by these alternative methods to the codes and categories generated via the grounded theory methods.

C.3.4 Outcomes and Follow-on Work

I anticipate two audiences to which I will report my findings. The first are readers of peer-reviewed academic publications and/or conference proceedings in the fields of HCI, visualization, and visual analytics. The second audience for my findings are journalists and the news-reading public, so I plan to report my findings online, either via my own website or in collaboration with the Associated Press. I hope to apply my findings in the future development of *Overview* and other applications like it. Finally, I anticipate that an examination of what makes *Overview* ultimately successful or unsuccessful will call for a critical inquiry of existing values and standards in journalism, as well as existing theories of data analysis.

C.4 Interview Protocol

The interview foci and associated questions below can be lumped into two broad categories: understanding the data journalist and understanding their use of *Overview*[4]. In semi-structured interviews, some questions come about naturally as follow-on questions, so this list is by no means a linear interview script; these questions are intended to be asked during / after a screen-sharing walkthrough.

Utility and efficacy:

- What (have / are) you (used / using) *Overview* to do?

[4] This exhaustive list of questions, drafted in Spring 2012, proved to be impractical as a script in actual interviews; despite this, answers to many of these questions arose naturally in conversation.

- Can *Overview* ingest some or all of your data? How long did this take? How much time did you spend cleaning/formatting the data specifically for *Overview*?

- How did you explore your data using *Overview*? How long did this take?

- How did you search your data using *Overview*? How long did this take?

- Did you have hunches about a story a priori? Were you able to verify these hunches?

- Did you develop new hunches about your story while using *Overview*?

- How did you choose which documents to read? Depth vs. breadth? How long did you spend employing these strategies? How many individual documents did you skim / read?

- How did you go about tagging your data in *Overview*? How long did this take? How many tags did you generate? How many did you delete? How many did you consolidate / split? How many of these tags were structural vs. semantic?

- What level(s) of tree pruning[5] were used?

- Context of use: where were you while using *Overview*? What other applications did you have running? Were you taking handwritten or digital notes? Were you using *Overview* alone or in collaboration with others?

- How long did you use *Overview* for? How long was spent reading documents vs. organizing, browsing / sorting / tagging?

- Did this match your time frame expectation prior to using *Overview*, given the size of the dataset?

[5]In *Overview v1-v2*, recall that tree pruning was controlled globally for the entire tree, and not selectively for each interior node; see Section 4.6.2.

- How much time was spent reading a single document? Did this vary?
- What proportion of documents did you read/skim? Are you confident in this proportion? Why or why not?
- What proportion of documents did you tag? Are you confident in this proportion? Why or why not?
- Not enough tags vs. too many tags?
- How did you deal with (unique / weird / important / unimportant / relevant / irrelevant) documents? (these may be overlapping categories).
- How did you dismiss the unimportant? The irrelevant and unimportant?
- Did you ever flag documents for later follow-up? Did this eventually happen?
- Did you ever re-read documents? Was this intentional?
- How, if at all, did you use the sampling functionality[6] to read documents?

Usability:

- On installing *Overview*.
- On pre-processing data / ingesting data.
- Your first use of *Overview*: Orienting yourself within the user interface (UI).

[6]This feature from *Overview v2* randomly chose a document from selected node, highlighted it in each view, and displayed the document in the *Document Viewer*.

- Usability issues with respect to: Linked displays (selecting items); the tree display; the MDS display[7]; UI for sampling; UI for tagging; UI for reading / opening in the browser[8].
- What UI features did you like? Which didn't you like?

Learnability:

- Learning materials: Self-exploration vs. relying on Jonathan Stray's blog posts and instructions[9], other sources? How much time spent on each?
- Developing a conceptual understanding of item layouts in the MDS display and the tree display.

Adoption:

- Has *Overview* become part of your workflow?
- (If not), do you foreseeing it becoming part of your workflow? Will it replace / add to / complement your workflow?
- How often is there a document set where *Overview* could be used?
- Would you recommend use of *Overview* for colleagues? Do you expect them to you use it in their workflows?
- Has *Overview* improved your process? Do you see *Overview* as having the potential to improve your process?
- Problems left unaddressed: What can't *Overview* do for you? What problem remains unaddressed?

---

[7]The scatterplot featured in *Overview v1-v2*.

[8]As *Overview v2* was a desktop application, it included a feature to open a document in a web browser as opposed to displaying its raw text content, provided that a document url was included in input data.

[9]These instructions have evolved since 2012; up-to-date instructional blog posts can be found here: `https://blog.overviewdocs.com/help/`.

- Previously untouched data: Were you able to ingest/analyze data that you couldn't approach with other tools?

- Previously unapproachable tasks: Were you able to perform tasks with *Overview* that you couldn't do (or couldn't do efficiently) with other tools?

- Discoveries with *Overview*: Have you made discoveries (in your data) using *Overview* that you wouldn't have been able to make with other tools?

Personal background:

- Who: What is your background and expertise?

- What brought you to working in this type of journalism?

- Relevant demographic information: Age, number of years working in journalism, number of years working in data journalism, educational background.

- How did you become versed in *data journalism*? Self-educated vs. formally educated, trained, or mentored, mixed (discerning between what was formally taught and what was independently learned).

- Technical skill set: Spreadsheets and table manipulation, data cleansing, internal and external validation; programming / scripting experience.

- Skill set with respect to document mining and unstructured data vs. skill set with respect to structured data.

Context:

- Tell me about your current day-to-day work.

- Career context: Agency / bureau affiliation, past and present.

- Spatial context: Where does the work happen? (Office setting, working remotely, other).
- Task context: What else is going on? Multi-tasking vs. single-task focus?
- Temporal context: Project time frame, deadlines: hours, days, weeks, months, vs. ongoing investigation.

Collaboration:

- To extent is the work collaborative? With respect to data acquisition, data pre-processing, analysis, writing, reliance on technical expertise of colleagues.

Workflows and processes:

- I'd like to hear about a recent story of yours, one from before using *Overview*, involving large collections of documents, representative of this type of work.
- (If there is no precedent) can I hear about a story involving large amounts of data analysis? Is this example representative or unique? (please provide links to articles, or send manuscripts).

Methodology:

- What was your data collection and analysis methodology prior to *Overview*? Personal vs. adopted / taught? Strictly followed protocol vs. ad hoc variations, subject to constraints of data, deadlines?

Tool use:

- What tools/services do you use for data collection (e.g., DocumentCloud[10])?

[10] `http://documentcloud.org/`

- How long does it take to collect data, using these tools/services?
- What tools do you use for pre-processing? Scraping / separating PDFs or semi-structured text? Cleansing? Internal and external validation (e.g., Google Refine[11], Google Fusion[12] tables, spreadsheet manipulation)?
- How long do these processes take, using these tools?
- What analysis tools do you use? (e.g., tools for keyword search, data exploration / orientation, visualization?).

For each tool:

- What did you use the tool to do?
- Were you able to ingest some or all of your data? How long did this take? How much time did you spend cleaning / formatting the data specifically for this tool?
- How did you explore your data using this tool? How long did this take?
- How did you search your data using this tool? How long did this take?
- Did you have hunches about a story a priori? Were you able to verify these hunches?
- Did you develop new hunches about your story while using this tool?
- How did you choose which documents to read? Depth vs. breadth? How long did you spend employing these strategies? How many individual documents did you skim / read?
- Context of use: Where were you while using this tool? What other applications did you have running? Were you taking handwritten or

[11] Now OpenRefine: `http://openrefine.org/`

[12] `https://goo.gl/DPj2pG`

digital notes? Were you using this tool alone or in collaboration with others?

- How long did you use this tool for?

- Usability issues? Installing this tool? Pre-processing data / ingesting data? What was your first use of this tool like? How did you orient yourself within the UI? What UI features did you like? Which didn't you like?

Data:

- When collecting the data, what state was it in?

- Individual documents? Single document?

- When analyzing the data, what state is it in?

- Unstructured text vs. mixed / semi-structured vs. structured? For the latter two, categorical variables, numerical variables, a mixture?

- Are individual points (documents) tagged or untagged? For structured data, do categorical variables exist to tag each point?

- How do you find correlations between structured and unstructured information?

Data provenance:

- Where did the data come from? Who manufactured it? Who collected it or consolidated it?

- Who did the cleansing / pre-processing of the data (if not yourself)?

- (If tags are used), where did the tags come from?

Orientation:

- Regarding orientation, by which I mean getting an initial overview of a dataset of unstructured text documents, how do you orient yourself within a dataset? Does orientation matter?
- How do you get a sense of the high-level structure of a dataset? The size of a dataset?
- To what extent do you rely on metadata? Tags? Word counts? Word co-occurrences? Word clouds? Word frequency? Mapping (when points have a geographical component)?

Exploration vs. targeted search:

- Relative time spent exploring vs. targeted search?
- How much of your work is hypothesis generation / exploration and how much is verification / fact checking?
- What other sources were used to complete the work? Other texts / data, vs. communications / interviews with people.
- When were other sources used? For hypothesis generation? For fact-checking/verification/corroboration? What was the pipeline/workflow for using other sources?

Results:

- What constitutes a strong piece of data journalism involving document mining?
- Reader perspective: How should readers react to document mining in journalism? What do the reader comment boards say?
- Peer perspective: How should document mining and by extension data journalism, its goals and processes, be reviewed and critiqued be peers? How has your data journalism been reviewed / critiqued?

C.5 Preliminary Field Study Results

[13]A post-deployment evaluation of a visualization tool can be difficult to conduct, particularly when evaluation criteria is contingent on determining how domain-specific professionals use the tool in the context of their ongoing work. Remote collaborators and journalists add to the logistical complexity of these evaluation studies. Such is the case with *Overview*, a visualization tool for exploring large document corpuses, built by our collaborator at the Associated Press. In this report I reflect upon the process and findings of an ongoing post-deployment, mixed-method evaluation of *Overview*, which includes an in-depth case study of a journalist who used *Overview* to investigate a large email corpus. I also reflect upon how this work factors into my long-term research goals relating to exploratory data analysis and evaluation in visualization.

C.5.1 Introduction

An evaluative study or experiment helps us to gauge the efficacy of a visualization tool or technique. The prominence of evaluation in information visualization research has grown in recent years. A recent survey [183] of over eight hundred articles published between 1995 and 2010 at four major visualization venues reported an increasing trend of papers containing an evaluation component. The vast majority of these evaluation studies focus on usability or graphical perception issues. However, there has yet to be a similar increase in the number of papers reporting an evaluation of visualization users' context-dependent processes. Evaluating users' processes with and without visualization can be highly informative at both formative and summative stages of visualization tool development [10, 88]. Yet these evaluation studies tend to be time-consuming and costly, posing many logistical challenges, particularly when collaborators and users outside of academia are involved [282]. As a result, post-deployment or summative evaluation studies are rarely published, and often report negative results [113, 221].

[13]This report, written in September 2012, describes preliminary findings from the IRAQ-SEC and TULSA case studies.

Novel evaluation methodologies and methods are emerging to overcome the challenges associated with studying user processes. In 2006, a workshop [23] was established to discuss visualization evaluation *beyond time and error*, the historical metrics of usability and graphical perception studies. These metrics cannot be used to reliably answer our process-centric evaluative questions, nor can we rely upon controlled laboratory settings, prescribed experimental tasks, or research participants who are unrepresentative of an intended user population [49, 243, 310]. Instead, the novel evaluation research community has established alternative metrics, conducting studies in externally valid research settings with representative users. This community has proposed methods for measuring gains in insight [276, 277], for comparing problem-solving strategies [202], and for assessing how visualization tools can facilitate multiple forms of learning [54]. They have demonstrated that evaluation methodologies that incorporate qualitative data collection and analysis methods adopted from the social sciences can be rigorous, inductive, and replicable [153, 164, 322]. Finally, they have shown that valid and reliable evaluation research often requires time, patience, and longitudinal coordination with collaborators and users [195, 277, 292].

The work I discuss in this report, a preliminary evaluation of a visual document mining tool, has been motivated by my interest in the proliferation of emerging evaluation methodologies and methods. I am also motivated by my interest in how people perform exploratory data analysis with visualization tools or techniques: I want to know how visualization techniques or tools is used to solve ill-defined problems, to discover and understand, rather than to lookup and verify [199]. One way I like to describe this process is how someone who visualizes data may not know a priori what they are looking for in a dataset, but they'll know it when they see it. I want to be able to reliably evaluate how a visualization tool supports this process of serendipitous discovery.

A fair question at this point is one of granularity: how do you define the process of exploration and discovery? There exists a body of work that characterizes these processes from an abstract, top-down perspective of human cognition [7, 241, 301, 330], while others have observed and characterized these processes from a bottom-up perspective, at the level of interface events [8, 291, 330, 358]. Task taxonomies constructed either on abstract models of human cognition or on interface-level events can be helpful, but all too often they can be stifling if they are meant to be used as evaluation criteria [16]. This is particularly true when one is required to compare how people interact with different visualization tools across domains. A valid and comparative evaluation methodology requires a robust mid-level task characterization of exploratory data analysis, one that spans domains and tool interfaces.

My long-term goal is to contribute to the construction of such a task characterization. To do this, I will study exploratory data analysis as it occurs across multiple domains, the tools individuals use, and the processes they undertake, while assessing how these tools and processes succeed or fail. I will employ a repertoire of established and emerging data collection and analysis methods, while reflecting upon the efficacy of my methodological choices. This endeavour began in March of this year, studying journalists' use of a visualization tool for exploring large document corpuses.

C.5.2 The Overview Project

In recent years, many large corpuses of emails, reports, and other documents have been "dumped", "leaked", released, or declassified by corporations, government agencies, and other organizations. A well-known example is that of WikiLeaks, an organization that released 391,832 documents relating to the conduct of armed forces and security contractors during the recent war in Iraq. Since that time, journalists have reported on what was contained in this corpus, which included startling patterns of civilian casualties, friendly fire incidents, and observed breaches of protocol [306][14]. My goal has been to better understand how journalists explore or "mine" these corpuses, how they seek information to support or refute prior evidence, or how they come to discover unanticipated newsworthy stories hiding in the data.

Journalism is a field in transition [141]. Areas of specialization and

[14] This is a reference to the IRAQ-SEC case study; see Section 4.5.1.

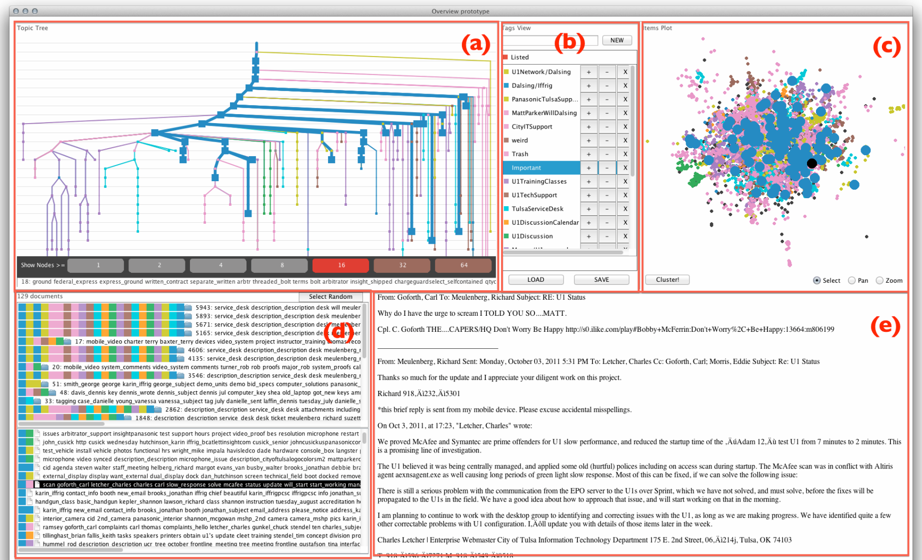


Figure C.3: *Overview v2*, displaying TULSA's email corpus: (a) a topic tree displays hierarchical clusters of related documents, (b) document tagging functions and a list of user-generated tags, (c) a scatterplot of documents in the corpus; nearby documents are similar, (d) a list of documents having the currently selected tag in (b), (e) an embedded document viewer displaying the currently selected document in (d).

experience among journalists are changing, reflecting the shift toward online content presentation and the necessity to address the growing amount of structured and unstructured information at one's disposal. As a result, it is difficult to predict how and when a data visualization tool will be used, who will be using it, and whether it will be an effective part of the process of writing a convincing news story.

These were the motivating questions of Jonathan Stray, our collaborator at the Associated Press, who recently worked with our group to develop *Overview*, a visualization tool to support the process of mining large document corpuses [151]. A prototype of *Overview* was released publicly in early

2012[15]. *Overview*'s user interface, shown in Figure C.3, is comprised of set of linked views for exploring hierarchical clusters of related documents within a corpus, providing means for reading documents, as well as for tagging them with personally meaningful terms or phrases.

Stray believes that the need to mine large document corpuses will increase in the coming months and years, and that current practices made this process impractical. These practices gravitated to either keyword searches or brute-force approaches: reading or skimming all documents in a corpus. The former approach requires one to know a priori what one is looking for, while the latter is too time-consuming, difficult to streamline and manage. In both cases, exploratory analysis is poorly afforded: it is impossible to sample representative documents in a corpus and extract the trends or patterns. *Overview* has been designed to make this exploration possible.

A post-deployment evaluation of *Overview* was an attractive opportunity for both Stray and our research group. The project was also well-situated within my larger research goals relating to evaluation methodologies and the characterization of exploratory data analysis.

In the sections that follow, I describe the methodology of my evaluation, followed by my current findings. Working with a collaborator from outside of academia has been a novel experience; in the final section of this report, I reflect upon the advantages, constraints, and implications of the *collaborator-as-gatekeeper* relationship. I then look back upon my research process thus far, discussing perceived inefficiencies as well as ideas for future improvement. I conclude with thoughts on how this project fits into the larger scheme of my PhD research. I do not yet claim a defined research contribution, as this paper is intended to be a largely reflective account of an ongoing project.

C.5.3 Methodology

My intent of conducting a post-deployment evaluation of *Overview* has been to assess whether or not it meets the exploratory data analysis needs of

[15] `https://overviewdocs.com/`

individuals with large document corpuses and hunches about potential stories contained within them; could *Overview* make writing a story possible in situations where doing so was previously impossible, or at least highly impractical? An evaluation would also serve Stray's need to identify major usability issues and barriers to adoption. Finally, given that *Overview* displays a document corpus using a novel visual encoding design choice [151], our evaluation would also seek to determine if journalists have a conceptual understanding of how this visual encoding represents the structure of semantic relationships in a document corpus.

An ideal research methodology for this work would include in-situ interview and observation sessions, in the spirit of longitudinal insight-based evaluations [277] and multi-dimensional long-term case studies [292]. However, our user base is distributed around the world, and logistical constraints keep me from visiting each individual journalist in their newsrooms. Furthermore, like many busy professionals, journalists are often working under tight deadlines, and have little time to participate in multi-session longitudinal studies. As a result, observing and interviewing these journalists in-situ is infeasible. These constraints limit possible data collection methods, which in turn has dictated my choice of methodology, a less than ideal situation. Relying upon retrospective accounts of journalists' experiences with *Overview*, via in-depth interviews and elicited textual accounts, necessitates an interpretive perspective, a focus on the language journalists use to describe their processes.

A grounded theory methodology [55] appeared to be appropriate, given a need to construct a post-hoc interpretation of journalists' processes of exploratory data analysis. The constant comparative philosophy of grounded theory prompts me to think flexibly, to make comparisons between the process of using *Overview* with other journalistic processes, with processes relating to exploratory data analysis as it occurs in other domains. Comparisons are also made between journalists and over periods of time, before and after the introduction of *Overview* into one's workflow. I initially proposed this methodology in an earlier research proposal, which was submitted as a final project for a course in interpretive and critical research, taken in the winter term of 2011/2012. This proposal is included in the supplemental material.

I am not the first to adopt a grounded theory methodology and its methods in visualization research. The methodology has informed prior work characterizing the use of visualization tools and techniques by professionals in domains such as architecture [322] and intelligence analysis [164], as well as our study of processes of high-dimensional data analysis DR across multiple domains[16]. In evaluation research, a "grounded evaluation" methodology has been used for the in-situ study of a visualization tool's efficacy in a target context of use [153].

A further justification for the use of a grounded theory methodology is that my research questions are not theoretically deduced hypotheses, nor is my objective to explicitly validate or refute prior task characterizations of data analysis [8, 241, 301]. Rather, my eventual goal is to construct a mid-level, cross-domain task characterization. Thus my starting point is not a rigid theoretical framework, but with the personal accounts of *Overview* users. Of course, no research exists in a void, uninfluenced by previous work. As such, my research questions are informed by assumptions and sensitizing concepts held within the visualization community. It is these sensitizing concepts that have allowed me to frame data collection methods, particularly a preliminary set of interview questions.

These sensitizing concepts include the notion that exploratory data analysis occurs in stages [241]. Exploratory data analysis may involve stages of hypothesis generation, without a well-defined set of questions in mind. There are also stages of hypothesis validation, where the goal is to support or refute prior evidence. Individuals may or may not engage in both types of stages during the course of a single investigation.

The products of exploratory data analysis are also among my sensitizing concepts. These products include moments of *insight* [53, 228, 367], serendipitous discoveries [9], and both optimal and suboptimal solutions to closed- and open-ended problems [202]. Admittedly, these products are ill-defined constructs within the language of our research community, and it is necessary for me to attain our participants interpretations of their meanings,

[16] Documented in Chapter 3

as well as their own terminology.

Finally, these sensitizing concepts include the disentanglement of a data analyst's learned expertise [54]. By this I mean that an analyst may have expertise related to their domain, expertise using specific analytical tools or techniques, and expertise regarding the data under examination, its semantics, and its provenance.

Participant recruitment: The recruitment of research participants has been difficult to predict, and dependent on the number of individuals who download, install, and use *Overview* for the purpose of writing a story. As a representative of a reputable news agency and a recipient of a prestigious Knight Foundation grant, Stray's clout meant that *Overview* would be highly visible to the journalism community. Participant recruitment is therefore a matter of waiting for prospective *Overview* users to establish contact with Stray. This means that he acts as a gatekeeper to research participants, an aspect of this project that I reflect upon later in the discussion section.

Data collection: My primary data collection method is that of an in-depth, open-ended interview, recorded for later transcription. Following the methodology of grounded theory [55], I have not specified the number of interviews that I planned to conduct a priori. The final number of interviews will depend upon how many *Overview* users express a willingness to participate in interviews and upon whether and when data saturation occurs, which I discuss in the following section.

Multiple interviews with each participant would be ideal, as processes change over the course of an investigation. However, this is an unrealistic plan; as mentioned above, journalists often conduct their investigation and write articles under tight deadlines, and often only have the spare time required to commit to an intensive interview after their story is written.

Regarding the content of these interviews, I began with a short list of interview foci and a small set of representative questions for each, composed according to guidelines for open-ended interviews [100] and for interviews conducted in the context of a grounded theory study [55]. These foci correspond with the sensitizing concepts described above. I ground the interview in the current document corpus under investigation, inviting comparisons from the interviewee's prior body of work, before using *Overview*. This list of foci and questions, included in this report's supplemental material, is flexible and subject to change, owing to the possibility that early interviews illuminate unanticipated concepts.

I compliment interviews by gathering texts and other information from participants, such as *Overview*'s log file of timestamped interface interactions. I also gather information regarding their data, such as the format and number of documents contained in the corpus under investigation. Finally, I request copies of notes participants take during the course of their investigation. Journalists' notes are less intrusive than dedicated progress or insight reports [255, 277]; they are personalized and externally valid, offering yet a another window into their own interpretations of their investigative process.

Data analysis: Data collection and analysis occur concurrently. I subject interview transcripts and textual artifacts collected from journalists to multiple iterations of coding, wherein I call upon the constant comparative method, the basis of grounded theory. First, *initial coding* labels excerpts from transcripts and notes with words from the participant's own vocabulary, using an active voice to emphasize a focus on a process occurring in time [55]. Next, I generate tentative categories of codes, each with an explanatory rationale based on comparisons between code instances, recorded as memos. This process informs subsequent data collection and the process of *focused coding*, wherein I re-code the data using emerging categories.

As categories are refined and theoretical concepts emerge through the process of focused coding, I may reach a point of theoretical saturation, when no major unexamined concepts are expected to appear as a result of future data collection. At this time, a mid-level task characterization of exploratory data analysis in this domain may begin to emerge. This stage will also invite comparisons between my nascent theory and existing theories [8, 241, 301].

Extensive log file analysis is complimentary to the analysis of transcripts

```
"2012-07-25 09:52:54.661","LAUNCH","jarrel-tpd-url-terms.vec"¬
"2012-07-25 09:53:04.318","NEW TAG","Listed"¬
"2012-07-25 09:53:04.318","NEW TAG","Selected"¬
"2012-07-25 09:53:04.378","TREE VIEW PRUNING","8"¬
"2012-07-25 09:53:04.426","MDS PAN TO","0,0"¬
"2012-07-25 09:53:04.426","MDS ZOOM TO","1.0"¬
"2012-07-25 09:53:30.730","TREE VIEW PRUNING","64"¬
"2012-07-25 09:53:32.726","TREE VIEW SELECT NODE",""¬
"2012-07-25 09:53:32.726","SET TAG ITEMS","Listed, 67 items"¬
"2012-07-25 09:53:39.29","VIEW DOC","a78df854762bc0dc2b0efe79cd46ff47"¬
"2012-07-25 09:53:45.884","SELECT RANDOM",""¬
"2012-07-25 09:53:45.885","VIEW DOC","c8a616998272f3b5e790c9ef30b87089"¬
"2012-07-25 09:53:45.890","SET TAG ITEMS","Selected, 1 items"¬
```

Figure C.4: An excerpt of *Overview*'s log file, listing timestamped interaction events, such as tagging a document or tree node, panning or zooming within the scatterplot, or viewing documents.

and textual artifacts, providing me with a partial yet objective account of usage strategy [245]. After parsing, aggregating, and filtering events in the log, I can extract descriptive usage statistics. The log file, as shown in Figure C.4, reveals how many documents were viewed and tagged during the course of a user's investigation, as well as how various user interface features were used over time.

C.5.4 Findings

To date, two professional journalists and a pair of academic researchers have completed an analysis of a large document corpus using *Overview*. I am also aware of several additional journalists and academic researchers who may be currently using it. Finally, I am aware of a journalist, an academic researcher, and business consultant who abandoned use of *Overview*, as it either did not meet their needs or was incompatible with their existing workflow or set of tools.

The discovery of prospective users from fields outside of journalism was unanticipated, indicating that *Overview* may support exploratory data analysis in the digital humanities, communications, and related domains.

Case study: Of the two journalists who completed an analysis of a document corpus using *Overview*, one published his findings [339]. The TULSA journalist was willing to be interviewed, and he provided us with not only his log file, but also his entire dataset. He also kept thorough notes during his investigation, and would later write a blog post intended for prospective *Overview* users [338]. In many ways, he was an ideal research participant, and I should not expect the same amount and depth of information from all future participants.

Beginning with an anonymous tip and hunch relating to a botched, four million dollar police equipment purchase, the TULSA journalist accumulated a document corpus of 5,996 Tulsa Police Department emails via a municipal records request. Using *Overview*, the journalist discovered newsworthy evidence contained in only a handful of emails: several police officials were responsible for the poorly managed purchase, and were caught in a potential conflict of interest with an equipment supplier.

I interviewed the TULSA journalist three days after his story ran. The two-hour interview was conducted via a Google^{+} Hangout video chat, a service that also affords chat participants the ability to share their screen. This feature permitted the TULSA journalist to walk us through his process, both with and without *Overview*. Using a screen capture application, I recorded the TULSA journalist's video feed along with the audio conversation. I later transcribed this interview, whereupon I realized that the ratio of time spent transcribing to interview duration was approximately four to one. I then coded this transcript, alongside the TULSA journalist's notes, according to the initial coding scheme described above. Meanwhile, my analysis of the TULSA journalist's log file provided an objective account of his process, corroborating with his subjective, retrospective description. With over twenty thousand events logged, log analysis was exhaustive yet time-consuming, requiring the better part of a week to complete; I reflect upon the utility and duration of log analysis in the following section.

The TULSA journalist has only been on the Tulsa "cops beat" for a couple of years. While he considers himself to be tech-savvy, he has no formal background in programming or visualization. As such, he required a considerable amount of assistance while installing and configuring *Overview*. This story was, in his words, the biggest story of his early career. The only similar story in his prior body of work was one about the emergency response practices of a local college security force, an investigation that necessitated the examining of a two-foot high stack of emergency call log printouts, while making annotations with highlighters.

What was most fascinating about the TULSA journalist's process was his determination to read and tag most of the documents in the corpus. In 56 hours of use, spread over 15 non-consecutive days, an impressive 70% of the documents in the corpus were selected and viewed in *Overview*'s embedded document viewer for at least one second. Rather than use *Overview* as a means of broadly exploring and sampling documents within a corpus, the TULSA journalist's usage strategy defied expectation:

> *"At the worst I wanted to use (Overview) as a way of organizing me looking through every email, and at best I wanted to look at most of the emails."*

A systematic and efficient process was a recurring theme in the interview: the quick dismissal of uninteresting or irrelevant documents, the worry about overlooking important documents, the prevention of having to re-read documents, and the vigilant scanning for unique or "rogue" documents. Always conscious of deadlines quotas, the TULSA journalist wanted to streamline his process of viewing documents as much as possible, at times referring to his rate of skimming individual documents as a "speed test":

> *"The speed factor, you're talking about just clicking and glancing, it could literally be as fast as three seconds per email. Until I got to one that I needed to know..."*

Figure C.5 shows that during the first several sessions of the TULSA journalist's investigation. Longer median document viewing times reflects the process of getting an initial feel for the documents contained in the corpus. During the longer sessions that followed, shorter median document viewing times reflected his streamlined, exhaustive scouring of the corpus. While finalizing his investigation, fewer documents were read for longer periods of

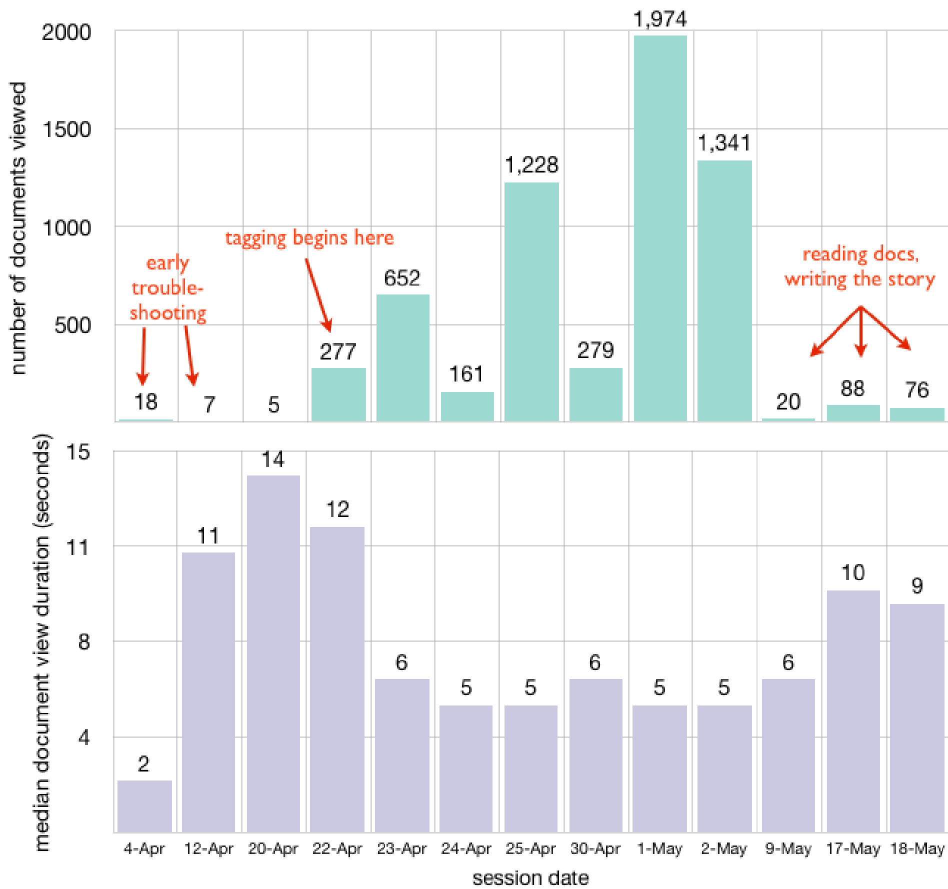


Figure C.5: The median time spent viewing a single document was lower during sessions in which more documents were viewed.

time; returning to a smaller number of significant documents. By this time he was preparing to write his article.

The TULSA journalist's similarly completist use of *Overview*'s document tagging feature may have implications for information management in exploratory data analysis tools. He created twenty-two tags, tagging 92% of the documents in the corpus with at least one of them. He explained that he became "obsessed" with having every document tagged. This came at the expense of having a complicated, cross-cutting, and disorganized set of tags. Regarding his tags, it was clear that eighteen of them reflected the content of the documents: email subject, sender, or recipient. The remaining four tags reflected a document's level of importance (*Important, Trash*) or personal memo for follow-up (*Check on* [this], *Weird*). He explained that he would have additionally liked to have had *unread / read* and *relevant / not relevant* flags for each document. The latter four tags and flags cross-cut the eighteen content-based tags, and suggest that two strategies of information management were being used during his investigation.

All things considered, could the TULSA journalist have carried out his investigation without *Overview*? He admitted the possibility, however it would have taken an estimated four months of full-time dedication to read the entire corpus, while maintaining the same level of organization that *Overview* provided him. By contrast, his investigation using *Overview* took less than two weeks. With news agency deadlines and article quotas to consider, a longer-term project would have been relegated to a part-time assignment; upon its completion, the story would have run the risk of no longer being newsworthy.

C.5.5 Discussion

The post-deployment study of *Overview* is far from over. At present, I have only spoken in depth with a single user, one who may not be representative of all *Overview* users. Finding users with different goals and strategies will take time, patience, and a reliance upon Stray as a collaborator. In the meantime, I can reflect upon what I've learned and refine my processes of data collection and analysis.

Samplers and completists: *Overview* was built for the purpose of broad exploration, sampling documents in a corpus as a means to extract trends or irregularities. It came as a surprise to both Stray and myself that *Overview* would instead be used for performing an exhaustive and systematic search. The TULSA journalist's search criteria was approximate, as opposed to exact, necessitating exploration rather than processes such as keyword search, browsing, or navigating. Broad sampling and exhaustive approximate search are two usage strategies, both being variants of exploratory data analysis.

I presented a comprehensive analysis of the TULSA journalist's use of *Overview* to Stray, whereupon we conjectured that the TULSA journalist's usage strategy reflected his initial records request, which was filtered a priori to emails with specific senders, recipients, and subject lines. His task was to find the small number of "smoking gun" emails, those containing the evidence needed for his story, validating the hypotheses that emanated from the original anonymous tip.

A sample-based usage strategy might arise in cases where the document corpus is "leaked" or "dumped", rather than requested. In the case of document leaks such as the Iraqi war logs [306][17], a journalist may explore the corpus broadly, reading far less than 70% of the corpus in an attempt to attain a gist or summary of its contents, rather than seek the "smoking gun". Instead, the TULSA journalist was a completist, one who used *Overview* as a tool for performing an exhaustive, "I have to read everything" investigation, albeit more systematically and efficiently than that of his earlier story involving the two-foot stack of call log printouts. It also remains to be seen if strategies of information management via tagging used during an exhaustive investigation are also called upon by those using *Overview* for a more sampling-based analysis of a document corpus.

At this point, I would like to interview an individual who uses *Overview* with a sample-based strategy, preferably one working with an unfiltered document corpus emanating from a dump or leak. Should they have an intention to determine the major trends or themes within, I would be curious to compare their process against that of the TULSA journalist's.

The collaborator as gatekeeper: Collaborating on a project with non-academic visualization tool builders has well-known advantages and disadvantages [282]. In my case, a robust, publicly-available tool was completed and deployed before the evaluation project began. As mentioned above, Stray's professional visibility would also attract potential adopters within the journalism community. As a junior visualization researcher, I do not have sufficient clout within the journalism, digital humanities, or related

[17] This is a reference to the IRAQ-SEC case study; see Section 4.5.1.

communities to attract users. It is furthermore inappropriate to evaluate the utility of *Overview* with undergraduate student volunteers or those not working in domains that do not regularly encounter large document corpuses. As a result, our collaborator is the gatekeeper to prospective research participants.

But is reputation enough to attract users? In the six months since *Overview*'s launch, we are aware of only a handful of individuals that have used it, with only one journalist using it to write a story. The time commitment of mining large document corpuses is extensive, not including the time required to install, configure, and learn how to use a new tool such as *Overview*. These problems may be alleviated in a forthcoming release of the *Overview* web application, which will feature a simplified user interface and a reduced feature set. It will also eliminate the need for a desktop installation, requiring less initial configuration. I hope that this upcoming release will attract a larger pool of prospective users, providing me with needed research participants.

Reflections on research process: After many hours spent analyzing the TULSA journalist's interview transcript and log file, I asked myself: what is the value of my analysis efforts? Is the log file a trove of information or a rathole? I admit that I became distracted by the notion that my findings would have an impact on *Overview*'s future design. It was around this time that Stray revealed plans to remove major features and overhaul the user interface in the forthcoming release of the *Overview* web application; I believed that my in-depth analysis could validate or refute these design decisions. I lost focus on the larger goal of understanding how *Overview* was used in the process of exploratory data analysis. This larger understanding of a process is not interface-specific [16]; when we understand the process and its interactions, we can then evaluate specific interface components.

Ultimately, Stray was fascinated by the detail and depth of my findings, but agreed that it was overkill for the purposes of validating design decisions or for identifying usability issues. What mattered most to him was that *Overview* had been used to write a story, and unexpectedly, that it had been used as a tool for streamlining an exhaustive search, rather than for its intended purpose, being a broad, sample-based variant of exploratory data analysis.

I have been rethinking my data collection and analysis methods, my methodology, and how my findings will eventually be presented to the research community. At the level of data collection, I have refined my interview foci to include a deeper examination of the difference between exploratory sampling and approximate search. More questions relating to information management and tag usage will also be added. At the level of methodology, the major question is whether to continue with a grounded theory approach and its constant comparative, bottom-up philosophy, or to instead survey a broad range of *Overview* users and then select specific case studies that appear to be radically different, given the users' goals. I could imagine reporting on cases of differing usage strategy (broad sampling vs. exhaustive approximate search), corpus provenance (records request vs. leak / dump), or user domain (journalism vs. digital humanities research). When presenting my findings, whichever approach taken will need to be consistent and well-justified.

This project in context: A replicable evaluation of a visualization tool that supports the process of exploratory data analysis requires a methodology grounded in a abstract characterization of what this process is and what it is not, an understanding of the form that this process takes across multiple domains. My current project is a small part of this dependency, in that it allowed me to study the use of visualization tools or techniques in a single domain. Over time, I will study exploratory data analysis and the use of visualization tools or techniques in other domains, both personally and via my ongoing comprehensive review of the literature.

I have already adopted multiple data collection and analysis methods; others will surely follow, subject to practical constraints and assessments of expected utility. A useful evaluation methodology and a mid-level task characterization are mutually dependent, and will develop together with further study.

C.5.6 Conclusion and Future Work

I am currently continuing my study and evaluation of *Overview*, a tool built for exploring large text document corpuses. I've still got a long way to go: in my first completed case study, I observed the tool being used for conducting an exhaustive approximate search; not exactly what I was expecting, but an interesting finding nonetheless.

I expect to interview more *Overview* users in the coming weeks and months. There is also likely to be an opportunity to elicit participation from journalism students as they use *Overview* in the context of a course project. Such an opportunity would be logistically simpler than observing professional journalists, affording multi-session in-situ interview and observation methods [54, 276].

Along the way I've learned and reflected upon a great deal: methodological considerations, the process and pitfalls of mixed-method data collection and analysis, and my experience working with an external collaborator. Writing this report has served to get my thinking directed toward my larger goal: the continued study of visualization evaluation and exploratory data analysis.

C.6 Overview v3

Overview v3, a web-based application released in Fall 2012, is shown in Figure C.6.

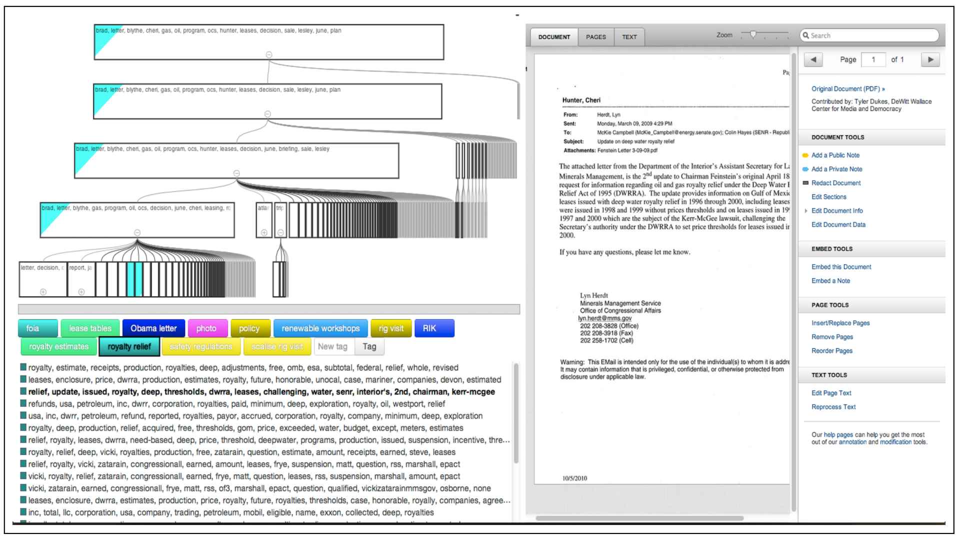


Figure C.6: *Overview v3*, a web-based application released in Fall 2012.

Appendix D

Design Study Supplemental Material

This appendix supports Chapter 5. It contains our original project proposal (Summer 2013) and example research artefacts (Fall 2013 through Summer 2014).

D.1 Design Study Proposal

[1]This project proposal pertains to the domain of large-scale energy management. At a high level, the goal of this project is to improve the process by which professional energy managers and utility company personnel analyze large amounts of data related to energy consumption. This process is often exploratory, meaning that not all analysis questions can be phrased as a directed query or a detailed hypothesis. Open-ended exploration can be supported by visualization, which involves a collection of techniques for representing and interacting with large amounts and varying types of data. Visualization techniques and tools provides analysts with an overview of the data, as well as an ability to drill down into specific subsets of the data as their analysis questions are refined. The application of visualization tech-

[1]This research proposal for the design study project was written in Summer 2013, during the first half of the work domain analysis phase.

niques to the problem of large-scale energy management will involve implementing and adapting an iterative user-centred design and evaluation process. The results of this project are expected to be of value to those working in the domain of energy management as well as visualization practitioners and user experience researchers working in other application domains.

D.1.1 Domain Background

Energy-conscious organizations are seeking ways to better understand energy consumption information beyond their monthly utility invoices; many are interested in identifying ways to conserve energy. However, while access to energy consumption data increases, the scale and complexity of the data is also growing; for example, with the emergence of Smart Meter technology it will soon be possible to collect energy consumption data at the level of rooms or individual appliances. As this deluge of data continues, effective means for presenting and analyzing energy consumption data are needed. The research community has responded with a number of applications and techniques for providing energy consumption feedback, often referred to as eco-feedback [104]. These predominantly address personal or domestic energy consumption, and includes recent work by Goodwin et al. [114], in which visualization is used to assist energy analysts in understanding and modeling domestic energy consumption. Eco-feedback displays are also being designed for direct use by home owners, from ambient visual encodings embedded into household displays [259] to web portals with detailed graphical displays [92]. However, little research to date has addressed energy management for large organizations, such as universities, corporations, governments, or energy providers. Consider a university facilities manager who oversees the energy management of all buildings on a large campus, a corporate facilities manager responsible for multiple properties spanning different geographical regions, or a utility company analyst who has access to energy consumption data for thousands of buildings. Pulse Energy[2] provides a number of commercial energy management solutions, ranging from detailed

[2]Pulse Energy has since been acquired by EnerNOC: `http://www.enernoc.com/`.

summary reports for small business owners to advanced interactive analysis applications for professional energy managers at larger organizations. Some of this feedback already involves visual encodings such as line graphs and bar charts; while useful, these cannot at present support a growing number of analysis questions, nor do they scale the size and heterogeneity of the data collected. Pulse Energy is seeking to improve the analysis workflow of professional energy analysts, which involves the identification of energy consumption patterns, the detection of anomalies, and the prediction of future energy consumption together with measures of confidence. This analysis requires an integration of data having varying scale and type, including raw and aggregated energy consumption, detailed weather information, as well as building size, location, operating hours, and occupancy.

D.1.2 Information Visualization Background

A number of existing visualization techniques may be suitable for representing and interacting with heterogeneous energy consumption data. However, the scope of visualization research involves much more than merely implementing known techniques. According to Munzner's nested model for visualization design and validation [217], visualization research can be considered at four levels. The first level involves the characterization of a domain problem; our domain problem, as we currently understand it, is for energy managers to better understand patterns and variations in energy consumption over time and across multiple buildings, as well as a need for predicting future energy consumption. The second level involves the translating this problem into a set of domain-agnostic abstract data types and tasks. With regards to abstract data types, our current understanding of the problem suggests that we will be faced with largely quantitative multivariate time-oriented data, as well as some spatial data variables. With regards to abstract tasks, our current understanding of the problem suggests high-level `present` and `discovery` tasks, which in turn involve searching for values and patterns, as well as `identifying`, `comparing`, and `summarizing` these results. Further analysis of the problem will be required to definitively identify abstract data types and tasks, their prevalence, and their relative importance. The third level involves the selection of appropriate visual encoding and interaction design choices that map to these abstract data types and tasks; many techniques have been developed for the analysis of time-related data [2]. Techniques for representing uncertainty in data [196] may be of use for the analysis questions pertaining to simulation, anomaly detection, and predictions of future energy consumption. We may also leverage visualization design choices for the `filtering`, `aggregation`, `selection`, `navigation`, `arrangement`, and alteration (`changing`) of data. Finally, the fourth level pertains to the design of algorithms that drive the design choices, which is of particular concern to those developing novel visual encoding or interaction design choices; at this time, it is not apparent that novel design choices will be required for this project. At each level, design choices must be evaluated, to which there exists a growing repertoire of suitable evaluation methods and methodologies [183, 211]. Within the visualization research community, projects traversing these four levels are known as design studies. The contributions of a design study paper include the identification of domain-agnostic task and data abstractions, a well-reasoned mapping of these abstractions to appropriate design choices, and a critical reflection of design and evaluation process, thereby providing guidance to other visualization practitioners working in similar and different application domains.

D.1.3 Objectives

The objectives of this project are as follows:

1. Identify energy management tasks and practices, documenting what works well as well as current unmet analysis goals of collaborating energy managers.

2. Given (1), derive domain-agnostic data and task abstractions and consider appropriate visual encoding and interaction design choices.

3. Engage in an iterative participatory design process with collaborating

energy managers and other project stakeholders.

4. Deploy and evaluate resulting designs as a component of the existing Energy Manager application, determining whether it meets the needs of energy analysts.

5. Document and reflect upon the design process and introduce or improve guidelines for visualization practitioners doing similar work in (a) the energy management domain, (b) in domains with time-oriented data, (c) domains with exploratory analysis tasks, (d) domains with subject-matter experts.

D.1.4 Methodology

This project will progress according to the nine-stage visualization design study methodology of Sedlmair et al. [284]. The first two stages, *learn* and *winnow*, are already underway; these stages involve meeting with potential collaborators, learning about their unsolved problems, determining whether a shared research question exists, establishing rapport, and identifying logistical barriers to continued collaboration, such as access to people who will use the visualization tool or technique and representative data. This research proposal serves as an indication that these preconditions to collaboration have been satisfied. The next stage, *cast*, involves identifying and developing relationships with the people who will use the tool or technique and other project stakeholders, as well as determining their level of involvement in subsequent stages. The project will focus on the core design and evaluation stages: *discover*, *design*, *implement*, and *deploy*. *Discover* refers to the process of identifying abstract data types and tasks, while *design* refers to the process of mapping appropriate visual encoding and interaction design choices to these data and task abstractions, involving iterative and rapid low-fidelity prototyping and evaluation; *implement* refers to the development and refinement of a high-fidelity interactive prototype, and *deploy* releases the finished visualization tool to the people who will use it, in this case as a component of Pulse's Energy Manager web application. The final two stages, *reflect*, and *write*, are expected to take place as part of an ongoing research collaboration with Pulse Energy; in the months after deployment, *reflect* refers to an ongoing critical examination of the design and evaluation process, occurring while longitudinal usage of the visualization tool is recorded. Finally, *write* involves identifying and writing about aspects of our design and evaluation process which may benefit other visualization practitioners working in other application domains, such as our selection of specific visual encoding and interaction design choices, or the efficacy of particular evaluation methods. Sedlmair et al.'s article [284] also contains 32 design study pitfalls during the stages of the methodology; these pitfalls include ignoring existing practices that currently work well (*discover* phase), failing to sufficiently abstract data and tasks (*design* phase), employing non-rapid prototyping methods (*implement* phase), and failing to conduct an externally valid case study, assuming that a hypothesized usage scenario is sufficient (*deploy* phase). These pitfalls are unambiguously worded in that they specify "what not to do" in a visualization design study. Altogether, the converse of these pitfalls form a design study checklist.

A breadth of design and evaluation methods will be adopted at the four core stages of the methodology. Many user-centred design and prototyping methods from the HCI literature [76] are appropriate for the design of visualization tools. For instance, in the *discover* phase, contextual inquiries [139], semi-structured interviews, and artefact analysis are useful for understanding the domain context surrounding existing workflows. In the *design* and *implement* phases, cognitive walkthroughs and think-aloud evaluations [76] are useful for identifying usability problems with designs generated during iterative cycles of rapid prototyping. In the *deploy* phase, follow-up interviews, feature usage statistics extracted from interaction log files, and artefact analysis help to determine the efficacy of the deployed application.

However, some methods must be adapted to suit the unique aspects of visualization design, particularly during the *design* phase. Lloyd and Dykes [195] stress the importance of using real data in early prototypes, thereby necessitating that prototypes address large-scale datasets and interactive transformations of the data. For this reason, traditional paper prototypes are often insufficient for communicating design ideas. This leads to a pitfall in which stakeholders exposed to these higher-fidelity interactive prototypes react in a different manner than they would have with low-fidelity paper prototypes. For this reason, early prototypes should be presented as "sketches" [361]. During the *design* and *implement* stages we will draw from our knowledge of visual encoding and interaction design choices, particularly those used for time-oriented data [2].

Regarding evaluation, we will consider the following methods during the *design* and *implement* stages [183], subject to logistical and timeline constraints [282]. Heuristic evaluation [225], individual and group usability interviews, chauffeured demos [195], and design critiques are among these methods. At the *deploy* stage, we will use longitudinal evaluation methods, often associated with multidimensional in-depth long-term case studies [292], which include interaction log file analysis [245], insight diaries [229], and multiple follow-up interviews with project stakeholders. We do not foresee the use of formal laboratory-based experiments to evaluate our design, unless our design includes novel or untested visual encoding or interaction design choices.

In design studies, it is common practice to work closely with stakeholders, those being experts in their domain and sources of invaluable insight. These people are considered to be collaborators in a research and design process; they are not considered to be research subjects. Nevertheless, should a need arise to elicit feedback from people beyond this core group of collaborators, we may resort to surveys and/or field interviews. These methods have been approved by an existing research ethics application (UBC Behavioural Research Ethics Board certificate number H10-03336: Human Factors in Information Visualization), and we will file a revision if deemed necessary.

D.1.5 Desired Outcome

A desirable outcome of this project is the design of visual encodings and interactions to support the analysis of large-scale energy consumption data, designs which can be embedded into Pulse Energy's Energy Manager application. This application is widely used by a diverse set of energy managers and analysts at universities or large distributed organizations, including utility companies. It is not intended for the analysis of domestic energy consumption. It is our intent that our designs improve energy analysts' understanding of energy consumption, allowing them to make more informed predictions regarding future energy consumption. By appealing to these people, Pulse Energy will gain a competitive advantage and is likely to increase adoption of their services.

D.1.6 Milestones

The following milestones are grouped by design study methodology [284] phases (underlined).

July 2013 – October 2013 / early discover phase:

- Data collection (part-time); site visits and contextual inquiries with energy analyst collaborators.

November 2013 / late discover phase:

- 13.10.28 - 13.11.01: Wrapping up data collection (site visits and contextual inquiries with collaborators and stakeholders, begun in July 2013).
- 13.11.04 - 13.11.08: Contextual inquiry data analysis.
- 13.11.11 - 13.11.15: Participatory ideation and requirements gathering workshop; Remembrance Day holiday (11 / 11).
- 13.11.18 - 13.11.22: Workshop data analysis.

November – December 2013 / design phase:

- 13.11.25 - 13.11.29: Early design mockups, rapid paper prototyping and internal evaluation.

- 13.12.02 - 13.12.06: Rapid *sketchy* interactive prototyping and evaluation.
- 13.12.09 - 13.12.13: Continued rapid sketchy interactive prototyping and evaluation.
- 13.12.16 - 13.12.20: In-house design critiques and heuristic evaluation; evaluation of sketchy prototypes with representative energy analysts.

December – January 2013 / implement phase:

- 13.12.23 – 13.12.27: Commitment to particular design; high-fidelity prototype development; Christmas Day holiday (12 / 25).
- 13.12.30 – 13.01.03: High-fidelity prototype development; New Years Day holiday (01 / 01).
- 14.01.06 - 14.01.10: Continued high-fidelity prototype development and integration.
- 14.01.13 - 14.01.17: Continued high-fidelity prototype development, integration and in- house evaluation.
- 14.01.20 - 14.01.24: Continued high-fidelity prototype development, integration and in-house evaluation.

January – February 2014 / early deploy phase:

- 14.01.27 - 14.01.31: Evaluation of integrated prototype with representative energy analysts, tweaking and preparation for deployment.
- 14.02.03 - 14.02.07: Evaluation of integrated prototype with representative energy analysts, tweaking and preparation for deployment.
- 14.02.10 - 14.02.14: Deployment of integrated prototype and recruitment of energy analysts for longitudinal study.
- 14.02.17 - 14.02.21: Monitoring of usage data, evaluation of deployed prototype begins.

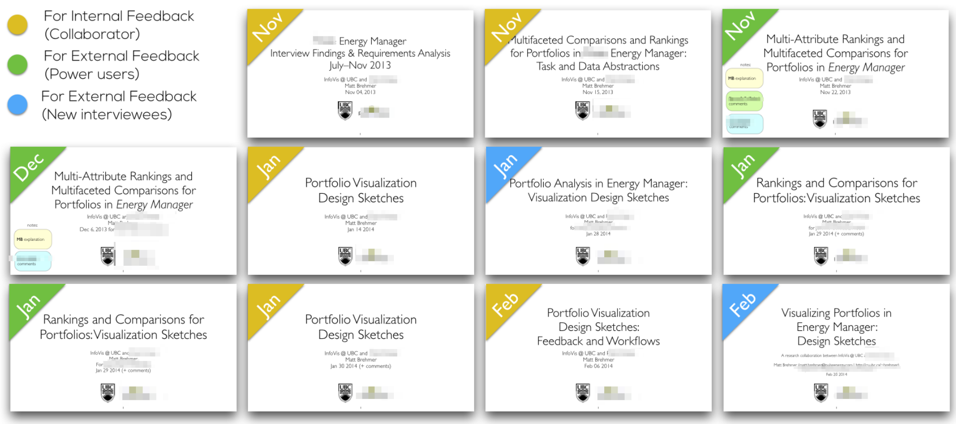


Figure D.1: Eleven slide decks (302 slides in total) created between Nov 2013 and February 2014. Slide decks were iteratively refined research artefacts used to document the research and design process.

- 14.02.24 – 14.02.28: Follow-up interviews and post-deployment usage data analysis.

March 2014 - Spring 2014 / late deploy phase, reflect and write phases:

- Continued longitudinal evaluation and analysis of deployed prototype, critical reflection, and preparation of a manuscript for submission to a conference or journal.

D.2 Example Research Artefacts

The following figures (D.1–D.16) and Table D.1 are example research artefacts from our design process described in Section 5.2, including examples slides from the eleven slide decks (302 slides in total) generated over the course of this project. Slides that attribute individual energy analysts or depict real portfolio data have been sanitized.

Figure D.18, Figure D.19, and Figure D.20 are screenshots of D3.js [30] prototypes developed in Summer 2014 that address view coordination design.

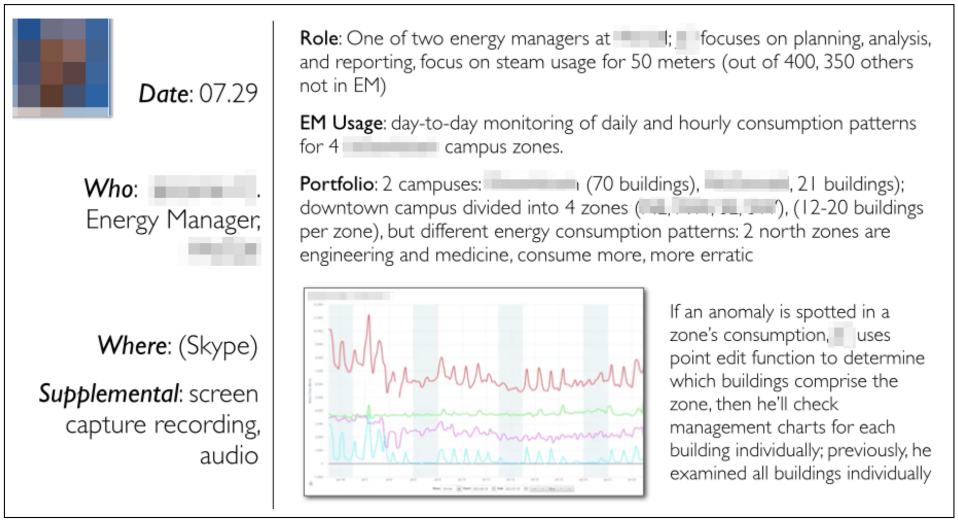


Figure D.2: Partial summary of findings from initial interviews with energy analysts (July – November 2013).

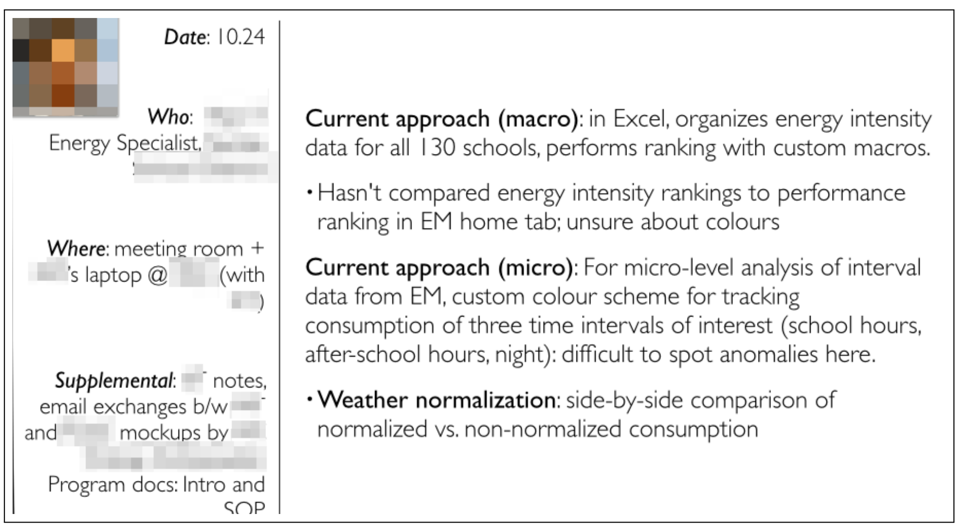


Figure D.3: Partial summary of findings from initial interviews with energy analysts (continued) (July – November 2013).

User	Role	Affiliation	EM Use & Frequency	Port-folio?	Portfolio Size, Organization	Task abstractions: current (not in EM)	Task abstractions: desirable	Task abstractions: possible (does data exist?)	Task abstractions: target
	space automation specialist	[illegible]	meta-user / power-user: frequently setting up charts, baselines for clients	YES	(Client portfolios range in size, hierarchical structure)	• Lookup → Compare: ranked performance (absolute and normalized) • Lookup → Identify: CUSUM of entire portfolio, single space	• Locate → Compare: portfolio performance faceted by any database field (tag, geographical location, primary use, square footage, year constructed, …) • Locate → Identify: space's contribution to portfolio's CUSUM • Lookup → Compare: multivariate ranking of portfolio performance • Locate → Identify: validated savings vs. unvalidated savings • Locate → Identify: end-use disaggregation within a space; • Locate → Identify: contributions of parameters and events baselines (ECMs, weather, outages, holidays, other events) • Locate → Compare: multiple baselines • Produce aggregate baselines • Locate → Identify: noise / confidence / uncertainty in baseline	• Locate → Compare: portfolio performance faceted by any database field (geographical location, primary use, square footage, year constructed, …), faceted by tag • Locate → Identify: space's contribution to portfolio's CUSUM • Lookup → Compare: multivariate ranking of portfolio performance	• Locate → Compare: portfolio performance faceted by space or by space attributes (over time) • Locate → Identify: contribution of individual space performance to aggregate space performance (over time) • Lookup → Compare | Summarize: multivariate ranking of spaces (over time)
	energy analyst	[illegible]	several hours a week, additional analysis in Excel	YES	[illegible]mpus: ~100 spaces (90% concentrated on single campus), subset in EM, departments cross-cuts spaces	• Locate → Compare: consumption of [largest spaces, libraries, mid-size spaces] • Locate → Identify: causes of threshold events in reference to OAT • Lookup → Compare: ranked space performance • Locate → Compare: before & after ECMs • Lookup → Compare: monthly department performance	• Lookup → Compare: department performance at arbitrary time scales • Locate → Identify: contribution of department(s) to space consumption • Lookup → Compare: OAT-demand regression curves before & after ECMs • Locate → Identify: end-use disaggregation within a space • Lookup → Identify: changes in space sensitivity to OAT • Locate → Compare: consumption of [illegible] other universities; • Lookup → Identify: weather predictions, trends	• Lookup → Compare: monthly department performance • Lookup → Compare: departments (arbitrary groups of spaces) performance at arbitrary time scales • Locate → Identify: contribution of department(s) to space consumption (assuming assignment of tags to sq. footage, occupants within a space) • Lookup → Identify: changes in space sensitivity to OAT • Lookup → Identify: weather predictions, trends	• Locate → Compare: portfolio performance faceted by space or by space attributes (over time) • Lookup → Compare | Summarize: multivariate ranking of spaces (over time)
	energy manager	[illegible]	day-to-day monitoring	YES	2 [illegible]mpuses, 4 zones in [illegible]ampus (~70 spaces), [illegible]mpus (~20 spaces); all in EM [illegible] focuses on 50 steam meters	• Locate → Compare | Summarize: combined consumption of two campuses; four groups of spaces for main campus • Browse → Identify: contribution of individual spaces to combined consumption, anomalies (spikes, surges) • Lookup → Identify: threshold events	• Lookup → Identify: contribution of individual spaces to combined consumption, anomalies (spikes, surges) • Locate → Identify: causes of threshold events in wider context • Lookup → Identify: contributions of parameters to PAM baselines (weather, occupancy)	• Lookup → Identify: contribution of individual spaces to combined consumption, anomalies (spikes, surges) • Locate → Identify: causes of threshold events in wider context	• Locate → Compare: portfolio performance faceted by space or by space attributes (over time) • Locate → Identify: contribution of individual space performance to aggregate space performance (over time)
	energy specialist	[illegible]	EM for data export; analysis done in Excel, EM analysis offloaded to student volunteers	YES	~130 schools, 2 accounts, 36 in EM (Electricity, 2 submetered), 4 in EM (Natural Gas)	• Lookup → Compare: ranked performance (absolute and normalized) • Browse → Identify: anomalies (jumps in rankings), trends (consistent rankings) at macro-level between spaces • Locate → Compare: single-space performance across N time periods • Produce annotations to explain aspects of performance	• Lookup → Compare: multivariate ranking of portfolio performance • Locate → Identify | Compare: single space performance, within and between operating hours and between days	• Lookup → Identify: anomalies (jumps in rankings), trends (consistent rankings) at macro-level between spaces • Locate → Compare: single-space performance across N time periods • Produce annotations to explain aspects of performance • Lookup → Compare: multivariate ranking of portfolio performance • Locate → Identify | Compare: single space performance, within and between operating hours and between days	• Lookup → Compare | Summarize: multivariate ranking of individuals (over time) • Locate → Compare: individual performance (over time)
	head maintenance engineer, automation	[illegible]	daily email digest, follow-up in EM ~3-4 hrs / week	YES	[illegible] campus, ~100 spaces and 2 zones in EM, monitors about 10 spaces / week	• Lookup → Compare: ranked space performance • Locate | Explore → Identify: anomalies, causes of threshold events / alerts	• Locate → Identify: end-use disaggregation within a space • Locate → Identify: contributions of parameters to PAM baselines (weather, outages, holidays, other events)		Lookup → Compare | Summarize: multivariate ranking of individuals (over time)
	climate and energy engineer	[illegible]	infrequent (annual, semi-annual reports)	YES	[illegible]mpus, ~100 spaces and 2 zones in EM, [illegible]y interested in handful of [illegible] spaces	• Lookup → Identify: differential between actual and predicted performance • Lookup → Identify: CUSUM • Locate → Compare: actual to baseline performance	• Locate → Identify: cause of long-term trend alerts • Locate → Identify: baseline precisions / uncertainty • Locate → Compare: performance across, arbitrary time periods	Locate → Compare: performance across arbitrary time periods	Locate → Compare: individual performance (over time)
	energy efficiency engineer (consultant)	[illegible]	some exploratory analysis, most analysis done in Excel	NO (small)	(single-space focus or small group of spaces (e.g. 5))	• Explore | Browse → Identify: load profile of space, anomalies; • Lookup | Locate → Compare: within and across spaces: monthly and seasonal differences in consumption / schedule / demand; OAT vs. demand for occupied and unoccupied periods. • Lookup → Summarize: distribution of OAT, demand	Locate → Identify: end-use disaggregation use within a space; Locate → Identify | Compare: effects of simulated ECMs on space performance	• Lookup | Locate → Compare: within and across spaces: monthly and seasonal differences in consumption / schedule / demand; OAT vs. demand for occupied and unoccupied periods. • Lookup → Summarize: distribution of OAT, demand	• Locate → Compare: individual performance (over time) • Lookup → Summarize: distributions of individual's attributes (over time)
	energy efficiency engineer (consultant)	[illegible]	some exploratory annalysis, confirmatory analysis done in Excel	NO	(single-space focus)	Lookup → Compare: month-to month %Δ in consumption, peak demand, actual : baseline	• Locate → Identify: effects of simulated ECMs on a space based on previous success • Locate → Compare: effect of ECMs between spaces	Lookup → Compare: month-to month %Δ in consumption, peak demand, actual : baseline	Locate → Compare: individual performance (over time)

Table D.1: Characterizing energy analyst's activities as abstract tasks according to our typology proposed in Chapter 2 (November 2013).

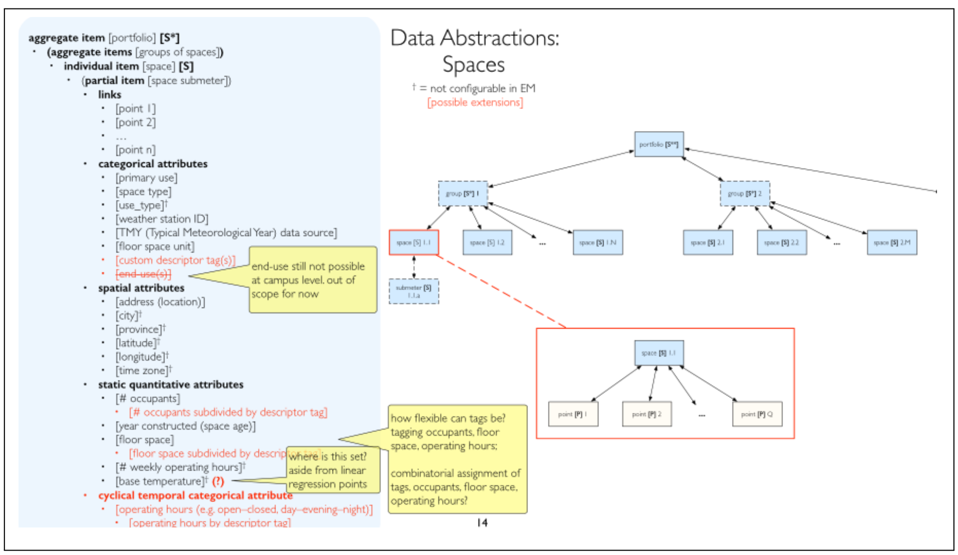


Figure D.4: Partial characterization of data abstractions relevant to energy analysts' activities (November 2013).

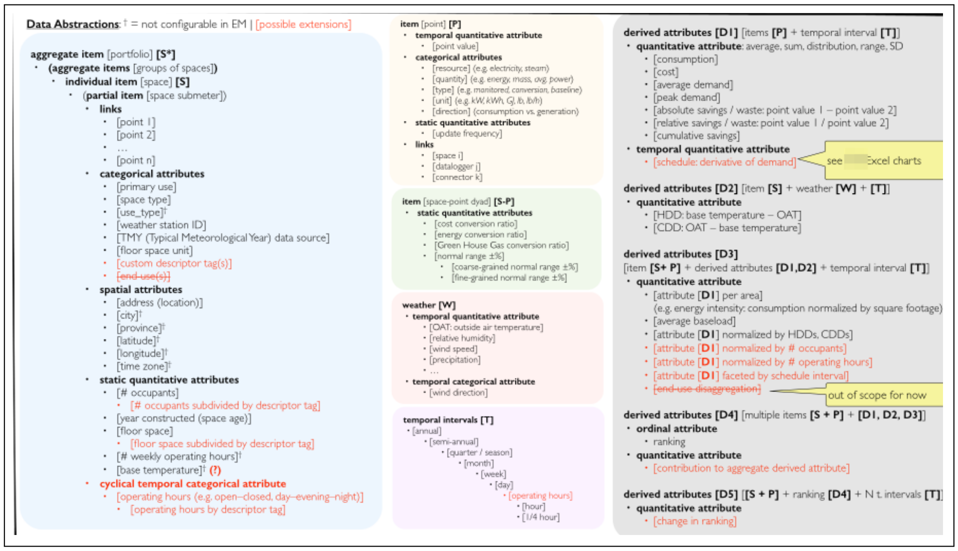


Figure D.5: Partial characterization of data abstractions relevant to energy analysts' activities (continued) (November 2013).

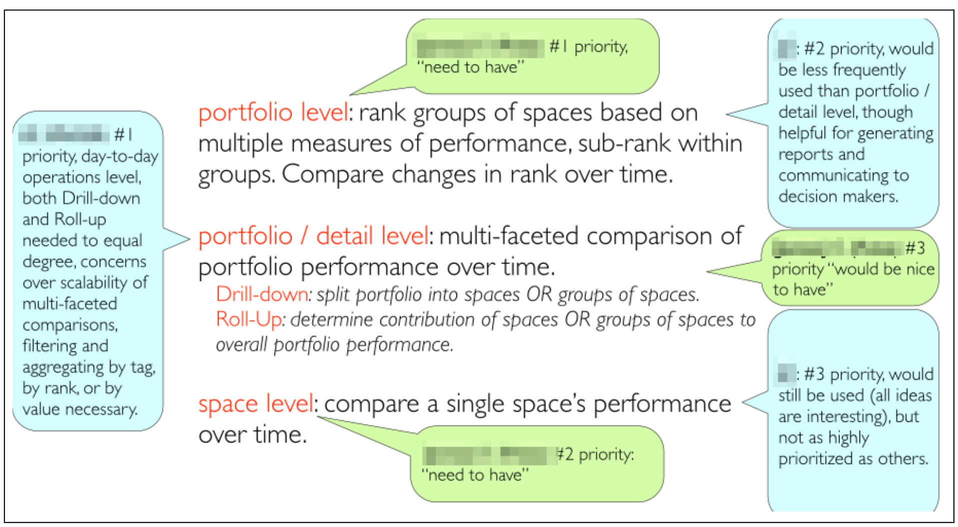


Figure D.6: Verifying the task and data abstractions with power user energy analysts: a summary of tasks. (November – December 2013).

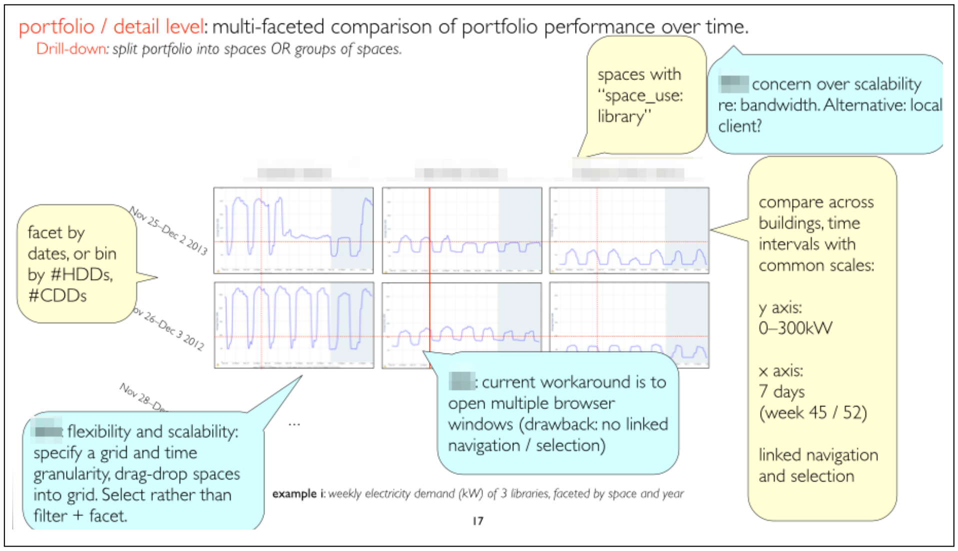


Figure D.7: Verifying the task and data abstractions with power user energy analysts (continued): a mockup of a faceted line graph (November – December 2013).

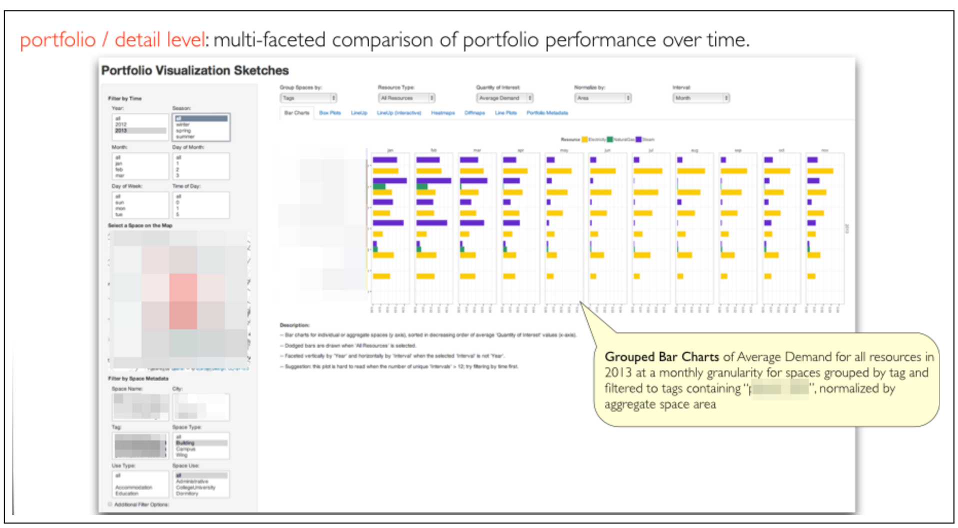


Figure D.8: Initial data sketches produced within the sandbox environment: faceted bar charts (January 2014).

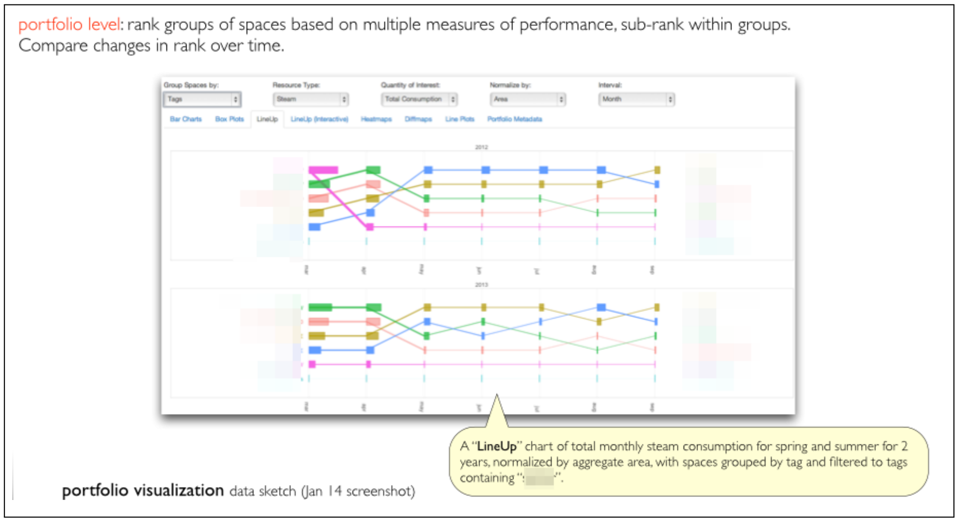


Figure D.9: Initial data sketches produced within the sandbox environment (continued): an early version of the bar + bump plot (January 2014).

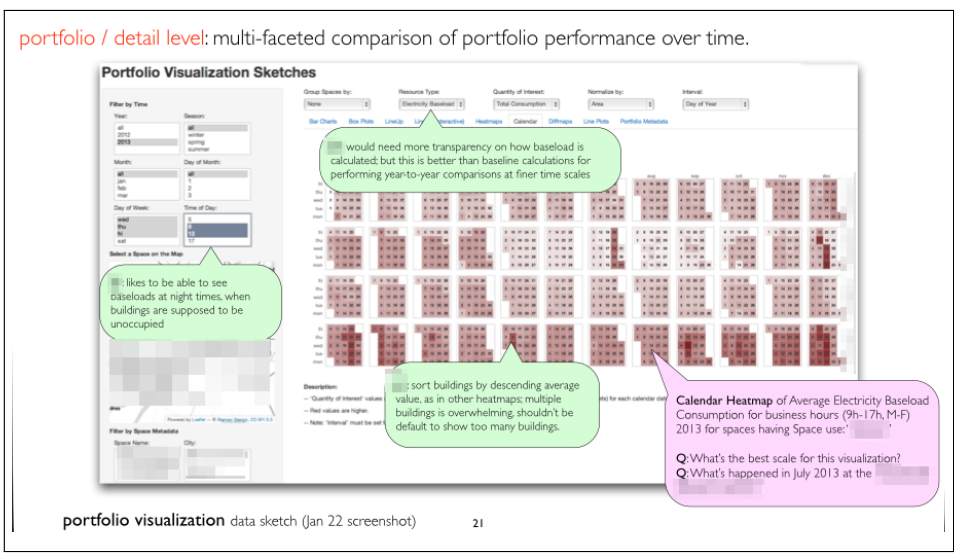


Figure D.10: Following-up with the power user energy analysts with designs from our sandbox design: calendar-partitioned time series matrix (January 2014).

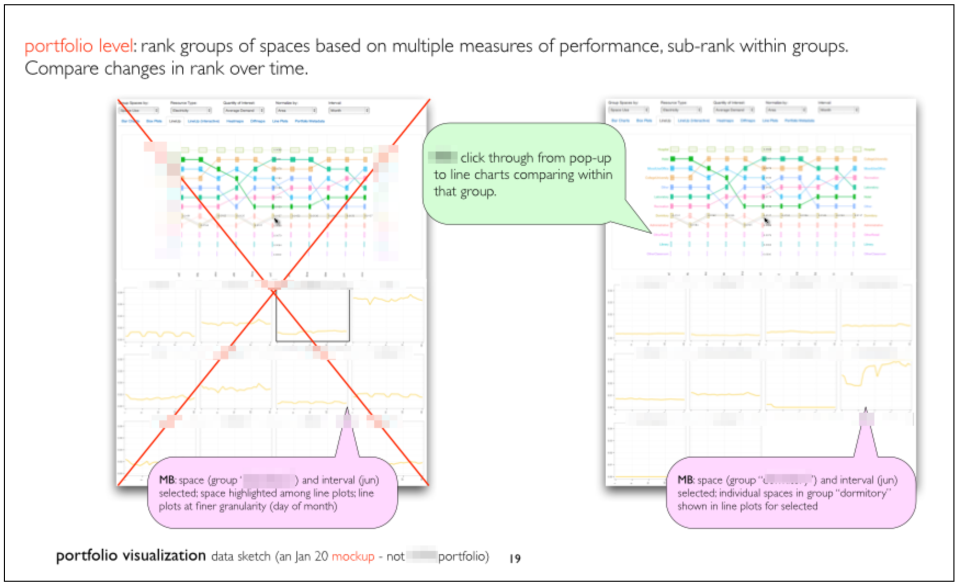


Figure D.11: Following-up with the power user energy analysts with designs from our sandbox design (continued): view coordination mockups (January 2014).

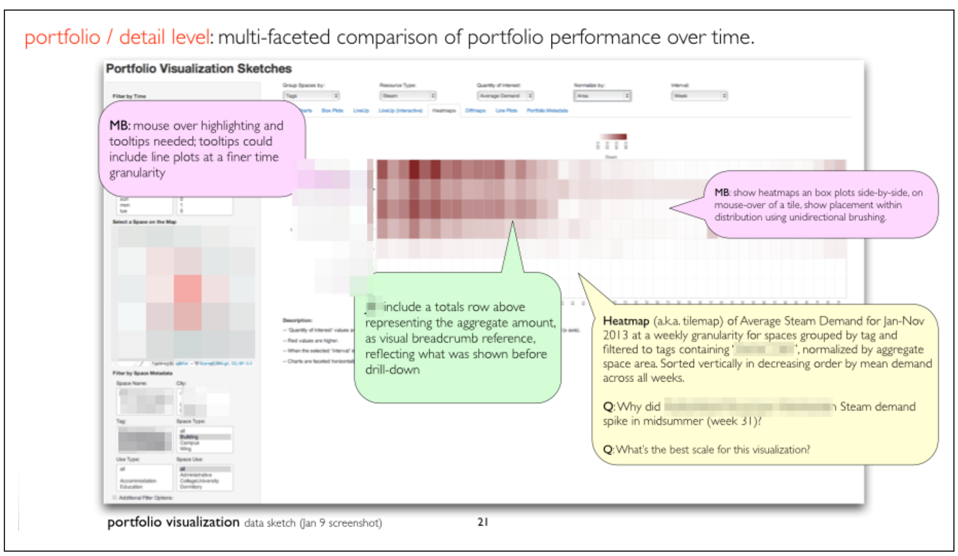


Figure D.12: Another iteration of data sketches produced using the sandbox environment: time series matrix (January 2014).

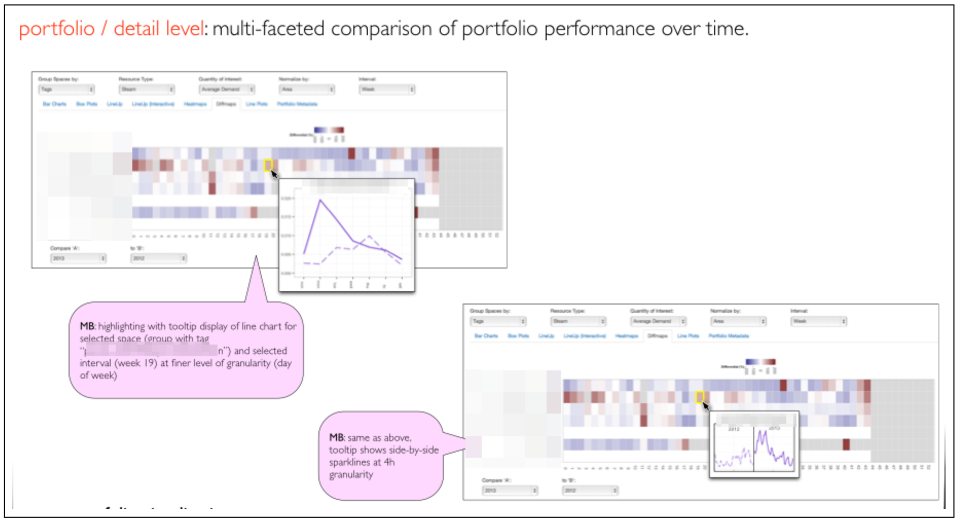


Figure D.13: Another iteration of data sketches produced using the sandbox environment (continued): interactivity mockups (January 2014).

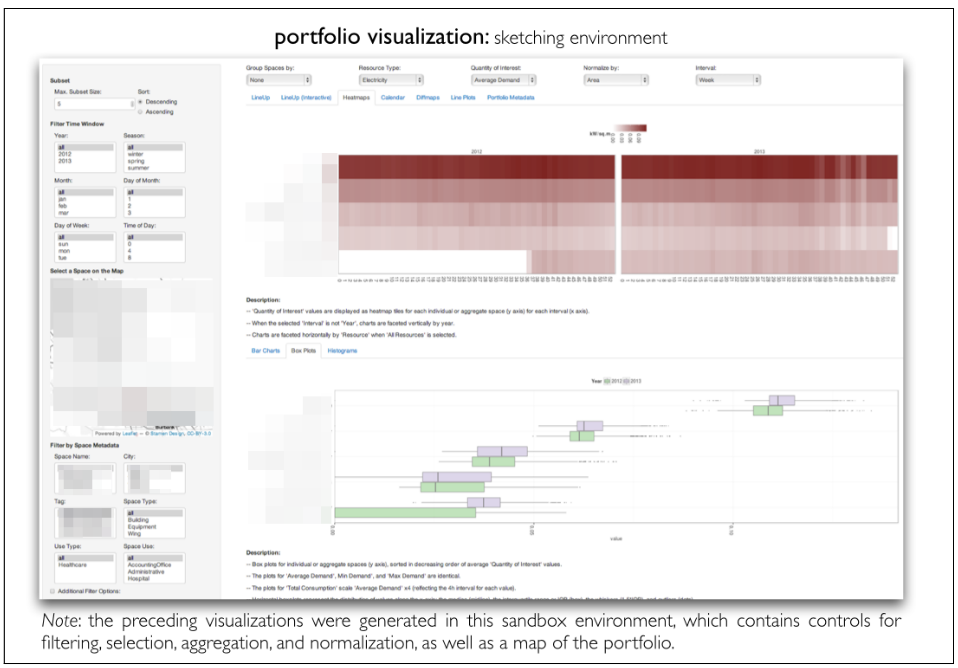


Figure D.14: Early view coordination design depicting a matrix with auxiliary boxplots (February 2014).

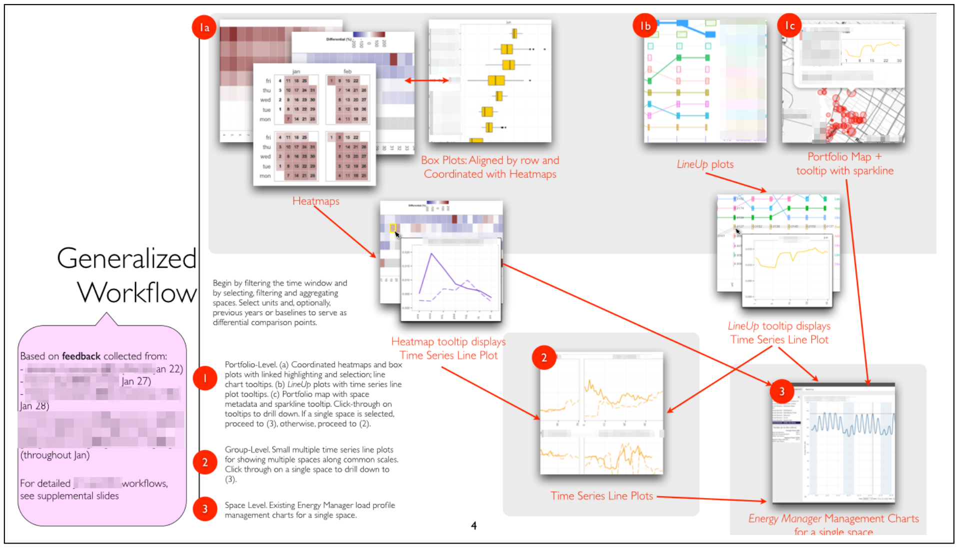


Figure D.15: Proposed workflow design involving multiple views based on consolidated feedback from energy analysts (February 2014).

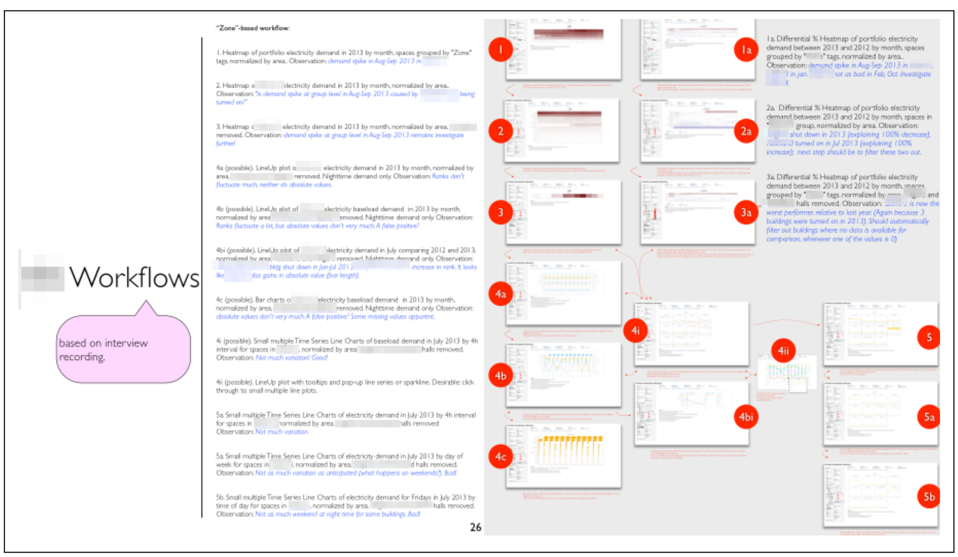


Figure D.16: Storyboards using sandbox screenshots based on power user workflows (February 2014).

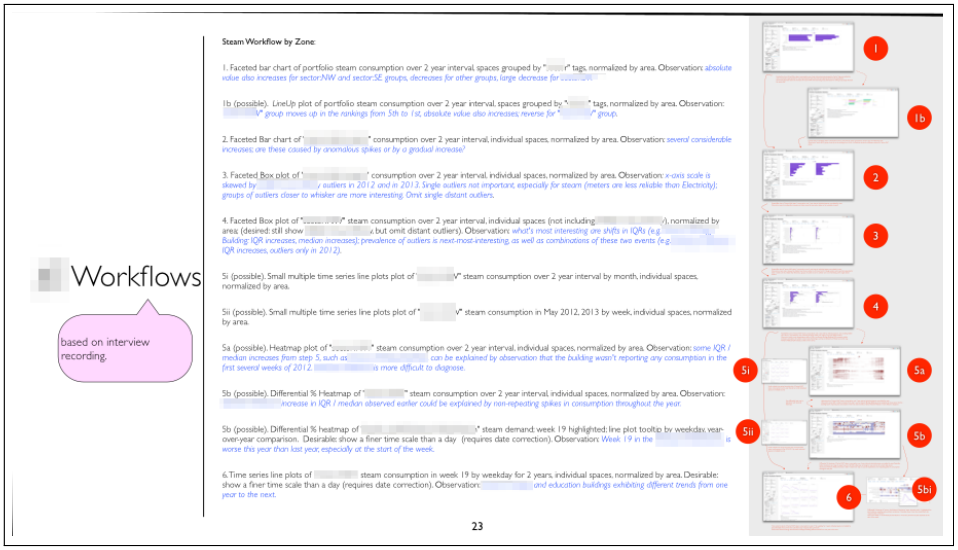


Figure D.17: Storyboards using sandbox screenshots based on power user workflows (continued) (February 2014).

Development on the redesigned Energy Manager continued throughout Summer 2014. During this time, we collected feedback on the new designs from five energy analysts at EnerNOC. Figure D.21 provides an example of how this feedback was documented.

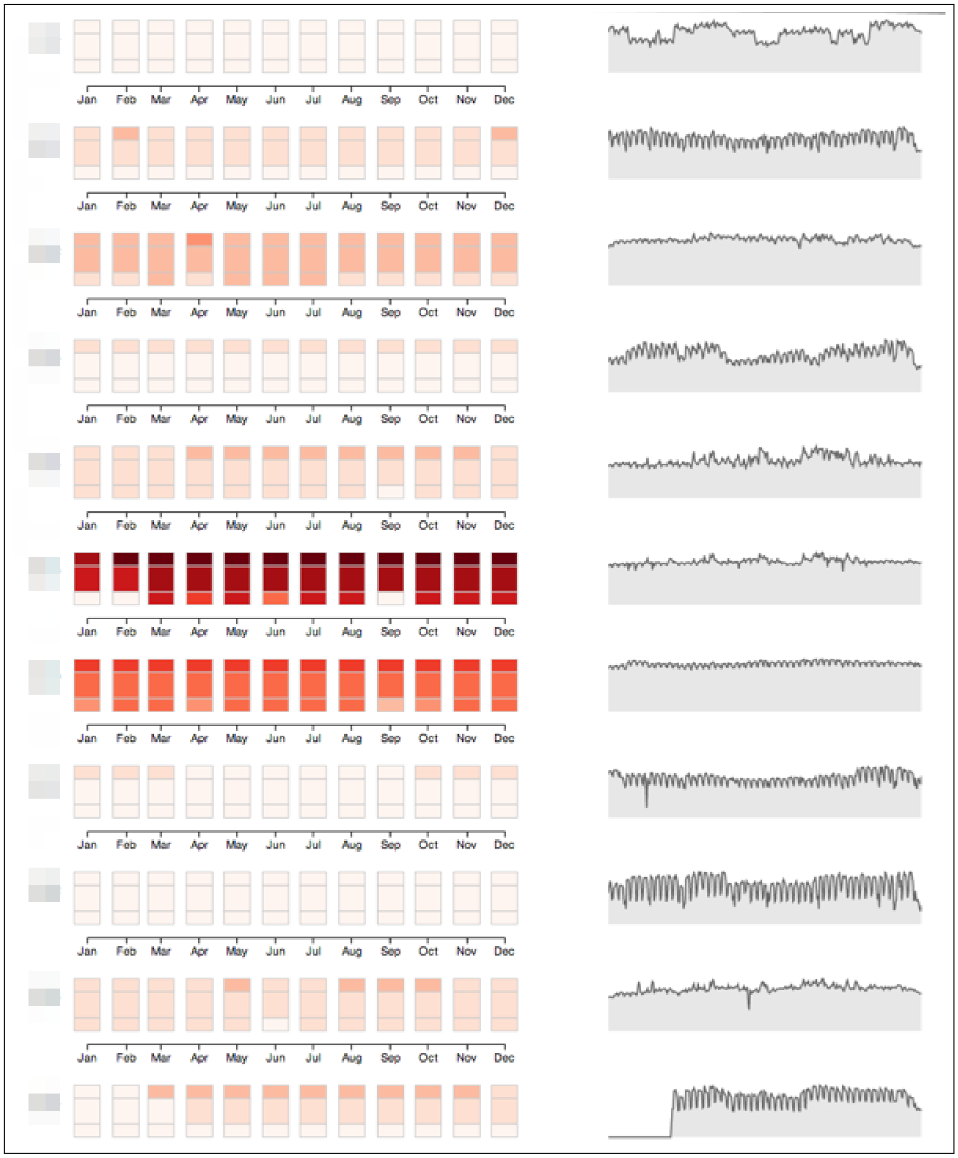


Figure D.18: Color stock charts [4] with juxtaposed line charts as alternative to matrix with juxtaposed boxplots (Summer 2014).

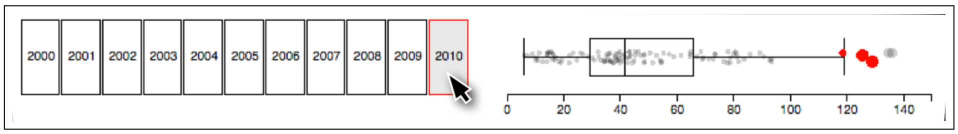


Figure D.19: Values from the brushed time period are highlighted on the juxtaposed boxplots (`http://bl.ocks.org/mattbrehmer/8be29724bdd7a63ff41d`) (Summer 2014).

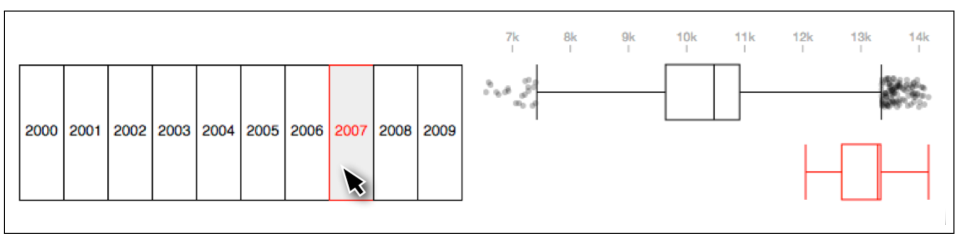


Figure D.20: Boxplot for the brushed time period (red) is shown alongside the boxplot for the entire time series (`http://bl.ocks.org/mattbrehmer/287e44c9a12151967874`) (Summer 2014).

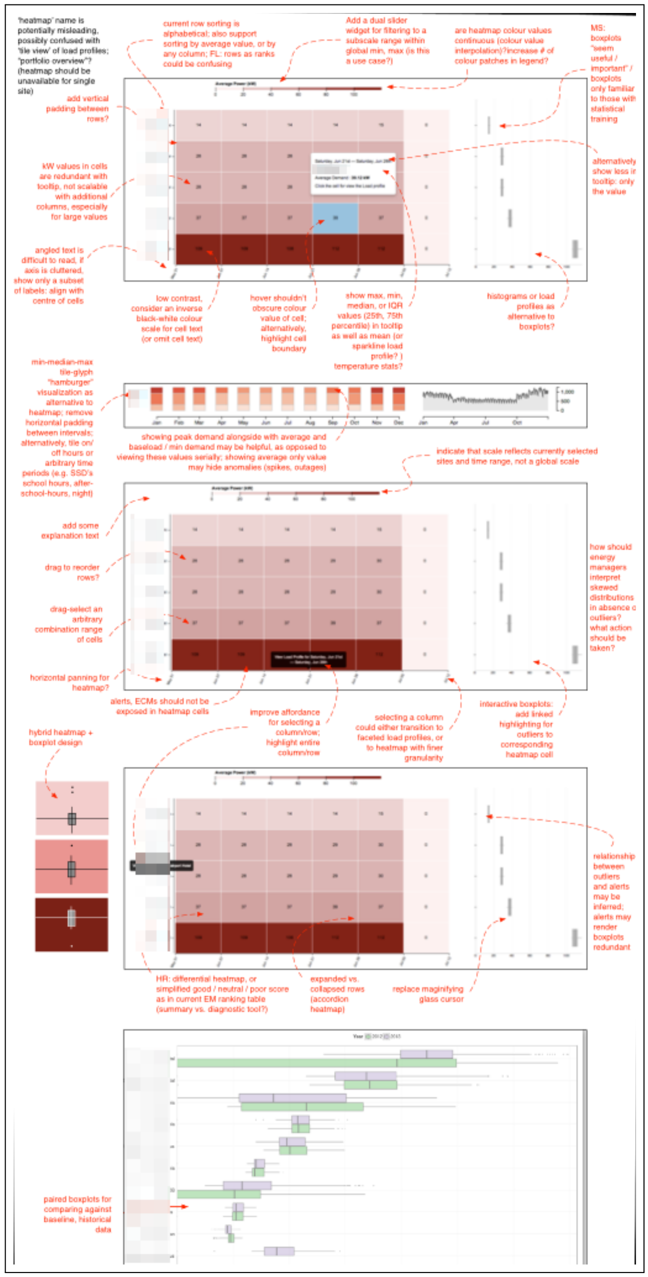


Figure D.21: An example of how this feedback was documented, using a combination of screenshots from the redesigned Energy Manager and earlier mockups (Summer 2014).

Appendix E

Participant Consent Forms

The following consent form was used in the field study (Chapter 4). A consent form with identical wording was used in the interview study (Chapter 3), with the exception of the attribution (the interviews in that study were conducted primarily by Michael Sedlmair).

University of British Columbia
Department of Computer Science

Participant Consent Form

Human Factors in Information Visualization

You are being invited to participate in a study entitled "Human Factors in Information Visualization" that is being conducted by Matthew Brehmer and Tamara Munzner.

Matthew Brehmer is a Doctoral student in the department of Computer Science at the University of British Columbia and you may contact him if you have further questions by email at [redacted] or by phone at [redacted]

This research is being funded by NSERC.

The purpose of this research project is to investigate how people analyze complex data, and how statistical analysis techniques and visual representations are used to make sense of this data. Research of this type is important because it allows us to design better data displays to allow more effective, efficient, and enjoyable analysis of data in a variety of applications.

If you agree to voluntarily participate in this research, your participation may include:

- Participating in a verbal interview.
- Filling out a background questionnaire that asks about your experience with computer technology and data analysis applications, as well as personal characteristics such as age and gender.
- Completing computer-based tasks.
- Filling out a questionnaire about the computer-based tasks and tools you experienced.
- Being audio-taped.
- Being video-taped
- Being watched by live observers.
- Using software with automatic data logging techniques (e.g., screen recording, mouse recording).

This study occurs over multiple sessions. Signing this form implies your consent to participate for the duration of the study. You will be periodically reminded that you can withdraw your participation at any time. Each research session is expected to take approximately 1-2 hours and, depending on the stage of research, will take place at one of the following locations:

- Your place of work.
- UBC Computer Science Department.
- Remotely by phone or a tele/video conference application such as Skype.

There are no known or anticipated risks to you by participating in this research.

Your participation in this research must be completely voluntary. If you do decide to participate, you may withdraw at any time without any consequences or any explanation. If you do withdraw, we will ask whether we may use your data for data analysis. If you decline, your data will be destroyed.

You will be asked to choose whether your participation in this study will be attributed (for example, by an explicit acknowledgement by name for your contributions towards the work in a publication) or confidential. If you choose confidentiality, your confidentiality and the confidentiality of the data will be protected by identifying data only with a participant number rather than your name. All data stored in computer files will be password-protected, and audio/video tapes will be stored in a locked office.

Some interviews will be with focus groups rather than a single individual. In these cases, we cannot fully guarantee confidentiality since other participants in your group may know your identity.

It is anticipated that the results of this study will be shared with others in the following ways:

- Published articles

- Conference presentations
- Theses
- Internet project descriptions

Data from this study will be disposed at the end of this research project. Electronic data will be erased, paper copies will be shredded, and audio tapes will be recorded over or physically destroyed.

In addition to being able to contact the researcher at the above phone numbers, you may verify the ethical approval of this study, or raise any concerns you might have, by contacting the University of British Columbia Office of Research Services (ORS) at [redacted].

Your signature below indicates that you understand the above conditions of participation in this study and that you have had the opportunity to have your questions answered by the researchers.

______________________	______________________	____________
Name of Participant	*Signature*	*Date*

A copy of this consent will be left with you, and a copy will be taken by the researcher.

Index

Typology Terms

actions 113, 140, 143
aggregate 16, 25, 33, 34, 42, 86, 87, 89, 97, 104, 106, 112, 132, 133, 136, 143, 148, 372
annotate 12, 30, 33, 36, 38, 39, 42, 59, 62, 66, 67, 70, 71, 75, 85--89, 95, 98, 100, 101, 104, 105, 143, 148, 163
arrange 25, 32, 42, 77, 78, 89, 98, 99, 122, 140, 143, 162, 372
browse 30, 31, 41, 46, 59, 62, 65, 71, 87, 89, 94, 152
change 32, 42, 143, 151, 282, 372
compare ..3, 14, 21, 22, 24, 25, 31, 41, 59, 61, 63, 65, 71, 89, 95, 96, 112, 113, 115, 118, 119, 121, 125, 130, 134, 137--139, 151, 163, 371
consume 1, 29, 39, 41, 67, 150, 154, 162
derive .23, 29, 34, 37--39, 42, 57, 65--67, 106, 108, 111, 112, 143, 148
discover 12, 21, 23, 29, 30, 41, 46, 48, 52, 59, 61--65, 71, 72, 92, 94, 95, 104, 113, 150, 152, 162, 163, 219, 282, 371
encode 21, 24, 32--34, 37, 38, 42, 77, 81--83, 95, 96, 104, 105, 112, 120, 124, 126, 129, 130, 140, 143, 148, 151, 152
enjoy 29, 30, 39, 41, 46, 152, 162, 224, 263, 279, 280
explore .3, 10, 12, 29--31, 36, 40, 41, 46, 74, 75, 80, 92--97, 100, 101, 114, 139, 152, 163, 263
filter ..24--26, 33, 40, 42, 118, 122, 124, 129, 132, 133, 136, 140, 143, 152, 372
how 6, 7, 15, 20, 22--28, 31, 32, 35, 38--40, 43--45, 47--50, 55, 64, 67, 69, 76, 83, 93, 102, 104, 140, 142, 143, 147--149, 151, 154, 156, 157, 163, 282

identify 12--14, 22, 31, 36, 37, 41, 57, 59--63, 65, 71, 75, 87--89, 91, 92, 94, 95, 112, 114, 121, 139, 151, 152, 157, 163, 164, 371
import 34, 42, 90, 92, 99, 100, 119
input . 6, 9, 15, 20, 22, 24, 28, 35--39, 43, 44, 48--50, 55, 62, 64, 79, 104, 143, 148, 149, 151, 152, 163
introduce 147, 148, 151
locate .. 12, 22, 25, 30, 31, 41, 46, 61, 65, 71, 74, 87--89, 91--94, 96, 97, 100, 113, 114, 118, 126, 139, 163
lookup 30, 31, 41, 46, 63, 65, 71, 87, 94, 113, 138, 163
manipulate 25, 26, 32, 33, 42, 45, 143, 151, 154
navigate . 24, 26, 32, 37, 40, 42, 75, 86--89, 98, 99, 104, 108, 115, 132, 140, 143, 148, 152, 372
output 6, 9, 15, 20, 22, 24, 28, 34--39, 43--45, 48--50, 55, 57, 64, 65, 78, 104, 113, 140, 143, 148, 149, 151, 152, 163
present .. 29, 30, 40, 41, 44, 46, 113, 128, 148, 150, 152, 162, 263, 371
produce 1, 28--30, 34, 37, 41, 46, 57, 62, 66, 67, 71, 94, 105, 143, 148, 150, 154, 162, 263, 282
query 28, 31, 41, 65, 151
record 24, 29, 34, 42, 66, 67, 143, 148, 163
search 24, 28, 30, 31, 41, 44, 46, 64, 94, 150, 151, 163, 371
select .. 25, 26, 32, 33, 37, 38, 40, 42, 63, 82, 83, 85--87, 89, 90, 92, 94, 96--99, 104, 105, 115, 122, 123, 132, 134, 136, 137, 140, 143, 148, 152, 158, 372
summarize 12, 31, 41, 62, 65, 71, 73, 88, 92, 94, 96, 104, 113, 120, 138, 151, 152, 163, 371
targets 113, 140, 143, 148, 149
what .. 6, 9, 15, 20, 22--28, 30, 34, 35, 38, 39, 43, 44, 47--50, 55, 64, 93, 104, 142, 143, 148, 149, 151, 154, 156, 157, 163
why 6, 7, 9, 12, 15, 20, 22--28, 35, 38--40, 43--50, 55, 57, 64, 66, 76, 83, 93, 102, 104, 142, 143, 148--150, 154, 156, 157, 163

A

adoption .. 2--4, 11, 12, 14, 16, 69, 74, 75, 79, 84, 86, 88, 97, 99, 101--104, 106, 107, 133, 138, 139, 159, 161, 163, 164, 300, 355, 376

affinity diagramming 247
algorithms .5, 55, 137, 297--300, 302--304, 307, 308, 312--314, 316, 317, 320--326, 328--330, 372
algorithmic input 71, 299, 310, 322, 323, 328
classification 71, 323
clustering 63, 77, 81, 82, 96, 98, 317, 320, 321, 330
parametrization of 320--322
analytical provenance 30, 34, 44, 228, 247, 273

B

bioinformatics 9, 17, 61, 154, 155, 286, 306, 313

C

canonical correlation analysis (CCA) 326
cartography 6, 21, 24, 39, 44, 149, 238, 249, 259, 260
case study ..11, 49, 74--76, 78, 79, 84--95, 97--100, 103, 104, 159, 350, 352, 355, 364, 366, 367, 374, 375
chauffeured demos 14, 109, 110, 121, 375
chemistry 9, 154, 286
coding (qualitative data analysis) ..6, 9, 55, 68, 70, 93, 247, 288, 289, 299, 300, 340, 341, 358, 360
cognitive work analysis (Vicente) 28, 281
communications 6, 21, 249, 359
computer vision 286

D

data abstraction .53, 76, 80, 95, 102, 106--110, 112, 119, 120, 123, 135, 137, 138, 371--374
data mining 310
data sketch 13, 14, 102, 120, 121, 375
design studies 2, 12, 14, 16, 23, 47, 48, 67--69, 72, 102, 106, 107, 123, 137--139, 143, 155, 156, 159, 161, 164, 299--301, 372--376
digital humanities 159, 160, 359, 364, 366
dimensionality reduction (DR) ..4, 7, 9, 10, 16, 30, 37, 38, 51--57, 61, 63--66, 68--72, 77, 97, 104, 142, 154, 157, 158, 163, 164, 287--289, 293, 296--304, 306--314, 316, 317, 320--331, 356

dimensional filtering 70, 159, 289, 304
dimensional synthesis ...55, 57, 59, 70, 285, 304, 305, 310, 311, 313
multi-dimensional scaling (MDS) ...34, 57, 77, 83, 158, 304, 318, 321, 325, 344
principal component analysis (PCA) ...57, 59, 304, 311, 312, 315, 316, 319, 320, 323, 325, 326
t-distributed stochastic neighbor embedding (t-SNE) 57, 329
distributed cognition 45, 48, 49, 241, 249, 254, 255, 258, 270
epistemic actions 45, 50, 249, 254, 255, 258, 263
pragmatic actions 45, 50, 249, 254, 255, 258, 263
document data 10--12, 16, 73--86, 88--91, 93--100, 102--104, 159, 160, 163, 164, 318, 334, 335, 337, 339, 340, 342, 343, 346, 347, 349, 350, 352--355, 358--365, 367
document mining 16, 74, 78, 79, 335--337, 339, 340, 345, 349, 351
domain convention 14, 48, 107, 112, 115, 121, 135, 161, 164

E

energy management 12, 14, 108, 119, 120, 123, 128, 138, 143, 160, 161, 164, 369--373, 376
Energy Manager 107, 108, 110, 112, 114, 115, 118, 119, 122, 126, 133, 134, 136--138, 140, 161, 373, 375, 387
evaluation 2--5, 21, 23, 25, 26, 45, 47, 49--53, 56, 67--70, 72, 74, 76, 78, 93, 101--104, 107, 109, 119--121, 123, 155, 156, 159, 160, 214, 215, 222, 223, 225--227, 298, 305, 308, 319, 330, 338, 350--352, 354--356, 364--367, 370, 372--378
cognitive walkthrough 374
contextual inquiry 55, 374, 376
heuristic evaluation 375
in the wild ...2, 7, 9, 47, 74, 296, 298, 302, 305, 323, 330, 331
insight-based evaluation 49, 225, 351, 355, 358, 375
think-aloud evaluation 49, 69, 92, 96, 99, 374
Excel (Microsoft) 89, 115, 118, 119, 128

F

familiarity .....14, 16, 106, 108, 119, 125--131, 135, 136, 139, 161, 162, 164 fisheries sciences 286, 324

G

grounded theory 55, 247, 288, 336--341, 355--358, 366

H

hierarchical data 12, 16, 77, 100, 354

high-dimensional data .7, 51, 54, 55, 69--72, 83, 98, 159, 298--302, 304--309, 314, 320, 321, 323, 324, 327, 329--331, 356

human-computer interaction (HCI) 6, 9, 21, 39, 53, 59, 249, 254, 255, 279, 286, 288, 299, 305, 306, 341, 374

I

information foraging 40, 46, 215, 218, 222, 271, 272

information retrieval 6, 21, 39, 68, 74, 80, 218, 232, 249, 257

insight 3, 20, 43, 163, 215, 218, 220--223, 225, 234, 235, 242, 272, 288, 313, 337, 351, 355, 356, 358, 375

intelligence analysis 47, 78, 79, 159, 223, 246

interaction 3--6, 12, 13, 15, 20, 23, 26, 28, 31, 32, 39, 40, 43, 45, 46, 48--50, 53, 55, 59, 67, 68, 74, 76, 92, 95, 97, 104, 107, 120, 123, 126, 130, 135, 143, 149, 155, 161, 214, 226, 235, 238, 240, 242, 244, 249, 255, 258--260, 262--264, 266, 268, 273, 300, 302, 372--375

- interaction logs 34, 49, 70, 85, 239, 240, 358, 374, 375

J

journalism ..10, 12, 16, 55, 63, 70, 73--76, 78--80, 83--91, 93--97, 99--104, 159, 160, 163, 286, 306, 317, 319, 334--341, 345, 349, 350, 352, 353, 355, 357--367

L

law enforcement 78, 79

M

machine learning 52, 71, 286, 298, 299, 303, 304, 306, 310, 319, 322, 323, 327, 329, 330

N

natural language processing (NLP) 160
nested model (Munzner) 5, 21, 123, 255, 265, 274, 275, 300, 371
network data 78, 154, 233, 261

O

Overview (document mining tool) .4, 11, 12, 16, 70, 73--95, 97--100, 102--104, 159, 160, 163, 164, 332, 334, 336--339, 341--346, 350, 353--367

P

perception 23, 48, 240, 245
play theory (Stephenson) 46, 256, 258, 269
policy analysis 9, 55, 61, 286, 306

S

sensemaking 22, 40, 46, 215, 216, 218, 222, 236, 241, 242, 249, 256, 261, 262, 273
sensitivity analysis 155, 301, 311, 314, 324, 325
sequential forward selection (SFS) 304, 319, 323
social networks 49, 70, 217, 286
stages of action (Norman) 44, 45, 258
storytelling 29, 100, 265

T

task 3--7, 9, 12--15, 20--30, 32, 34--40, 43--50, 52--56, 61--70, 74, 75, 78, 79, 83, 84, 93, 95, 97, 102--104, 107, 110, 112--115, 118, 124--132, 134, 137--140, 143, 147, 149, 154--159, 163, 164, 213, 214, 216, 220, 223--225, 227, 228, 232, 235, 237, 240, 241, 243, 246, 247, 253--255, 263, 266, 272, 273, 277, 278, 282, 283, 287--289, 293, 297--302, 305--317, 319--331, 345, 346, 351, 352, 356, 358, 364, 366
- ends 20, 22--24, 40, 46, 50, 67, 143, 148
- high-level tasks 20, 22, 25, 26, 28, 40, 43, 45, 46, 50, 215, 216, 218, 219, 221, 228, 235, 236, 238, 239, 242, 260, 262, 270, 272, 277, 307, 310, 330, 371
- low-level tasks 20, 22, 24, 26, 28, 40, 43, 45, 50, 158, 163, 217, 219, 228--235, 237, 242, 260, 268, 271--273, 277, 278, 310, 330

means ..20, 22--25, 36, 38, 40, 45, 50, 57, 67, 80, 92, 129, 143, 148, 159, 163, 307, 310, 370
mid-level tasks 28, 242, 275, 310
task abstraction ...3--6, 13, 14, 21, 22, 27, 28, 40, 48, 53, 55, 64, 67--69, 72, 74, 76, 80, 83, 93, 95, 102, 104, 106--110, 113, 114, 119--121, 123, 135, 137, 138, 143, 154, 156, 158, 159, 163, 289, 297--301, 308, 323, 329, 331, 371--374
task analysis 2, 4, 5, 14, 23, 26, 43, 47, 48, 53, 67, 164, 214, 216, 282
task applicability 3, 4, 40
task sequence 6, 7, 9, 15, 20, 22, 24, 28, 35--40, 43, 44, 48--54, 56, 57, 59--72, 107, 114, 118, 138, 142, 143, 148, 149, 152, 157--159, 163, 164, 227, 265, 275, 283, 289
task typology 1, 4, 6, 7, 9, 12, 13, 15--17, 20--24, 26--28, 31, 35--37, 47, 49, 50, 53, 55--57, 61, 64, 66, 67, 70, 93--95, 103, 104, 113, 114, 142, 143, 148--157, 162--164, 213, 224, 283, 284, 288, 289
term frequency-inverse document frequency (TF-IDF) 80, 81, 83
time-oriented data 14, 16, 30, 31, 33, 53, 106--108, 110--114, 119--121, 123--125, 130, 134--136, 138--140, 154, 160, 162, 232, 276, 278, 319, 375
topic models 76, 77
trust 14, 16, 95, 96, 100, 164

U

uncertainty 155, 215, 254, 372
usage scenario 84

V

view coordination ..14, 24, 26, 32, 34, 99, 106--108, 110, 130, 137, 139, 140, 159, 162, 378
animated transitions 33, 129, 151, 162, 232, 267, 271
brushing across views 32, 83, 134, 162, 232, 234, 239, 261, 267, 271, 273, 277
details-on-demand 32, 37, 38, 152, 232, 272
faceting (small multiples) 118, 120, 121, 124--126, 129--132, 134, 139, 140, 143
focus + context 25, 32

view juxtaposition 14, 107, 110, 118, 123, 124, 130--132, 134--136, 138, 140, 143, 161, 162
view sequencing 14, 108, 110, 132, 162
visual analytics 39, 154, 155, 217, 222, 240, 260, 341
visual encoding 3, 5, 6, 10, 12--16, 23, 27, 28, 31--33, 37, 49, 50, 52--55, 62, 64, 68--70, 74, 76, 82, 95, 96, 104, 106--110, 115, 119--121, 123, 127--131, 135--138, 148, 151, 152, 155, 157, 160--162, 229--231, 237, 239, 261, 264--266, 268, 271, 276, 297, 300, 301, 310, 317--321, 355, 370--375
bar + bump plot 124
bar chart 112, 115, 118--120, 124, 125, 130, 131, 134, 139, 140, 161, 371
grouped bar chart 33, 115, 120
stacked bar chart 33, 124, 130--132, 134, 139, 140
boxplots 120, 124, 125, 130--132, 134--136, 138, 140, 162
bump plot 124, 126, 135
density plot 61
heat map 58, 61, 135
histogram 33, 158
horizon graph 120
line graph ...30, 31, 33, 106, 107, 112, 115, 118--121, 124--126, 131--134, 139, 140, 161, 371
superimposed line graph 115, 120, 124, 126, 134
map 120, 124, 128, 129
choropleth map 21, 22, 24, 29--31, 46
matrix 120, 121, 124, 127, 128, 130, 132, 134--136, 138, 140, 162
calendar (matrix) 119, 120, 124, 128, 130, 134, 135
correlation matrix 58
node-link graph 32, 33, 58, 155
parallel coordinates plot 32, 162
pie chart 161, 162
scatterplot ..31--34, 37, 38, 52, 54, 58, 59, 61, 63, 77--79, 83, 87, 96--99, 101, 158, 161, 277, 297, 304, 310, 314, 316--320, 322, 323, 332
3D scatterplot 58, 61, 77, 310, 319, 323
scatterplot matrix (SPLOM) ..32, 58, 61, 63, 64, 271, 304, 310,

322
scree plot 58, 323
stacked area graph 78, 124, 130, 131, 134, 139, 140
tag cloud 77, 160
tree 25, 30, 34, 58, 61, 74, 75, 77, 78, 81--83, 86--91, 95--101, 104, 342, 344
treemap 96, 162
visual information seeking mantra (Shneiderman) 152

W

work domain analysis .14, 28, 51, 53, 64, 67--69, 72, 106, 108--110, 137, 139, 156, 281
workflows 7, 9, 14, 28, 54, 85, 92, 100--103, 108--110, 114, 130, 132, 134, 138, 152, 159, 163, 216, 301, 336, 344, 349, 355, 359, 371, 374
workflow design 106, 137, 139