

**VISUALIZATION OF CONSTRUCTION MANAGEMENT DATA**

by

Chao-Ying Chiu

B.Sc., National Chiao Tung University, Taiwan, 1994

M.Sc., National Cheng Kung University, Taiwan, 1996

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

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES  
(Civil Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA  
(Vancouver)

October 2011

© Chao-Ying Chiu, 2011

## **Abstract**

To date, the research and development effort as reported in the literature for presenting input/output data in support of human judgment for conducting construction management (CM) functions and associated tasks has been relatively limited. In practice, CM practitioners often find it difficult to digest and interpret input/output information because of the sheer volume and high dimensionality of data. One way to address this need is to improve the data reporting capability of a construction management information system, which traditionally focuses mainly on using tabular/textual reports. Data visualization is a promising technology to enhance current reporting by creating a CM data visualization environment integrated within a CM information system.

Findings from a literature review combined with a deep understanding of the CM domain were used to identify design guidelines for CM data visualization. A top-down design approach was utilized to analyze general requirements of a CM data visualization environment (e.g. common visualization features) that effect visual CM analytics for a broad range of CM functions/tasks. A bottom-up design process integrated with design guidelines and the top-down design process was then employed to implement individual visualizations in support of specific CM analytics and to acquire lessons learned for enriching the design guidelines and common visualization features. Taken together, these three components provide a potent approach for developing a data visualization tool tailored to supporting CM analytics.

A research prototype CM data visualization environment that has an organization of thematic visualizations categorized by construction conditions and performance measures under multiple views of a project was created. Features of images generated from the foregoing visualizations can be characterized by different themes, types, contents, and/or formats. The visualization environment provides interaction features for changing/setting options that characterize images and enhancing readability of images as well as a mechanism for coordinating interaction features to increase efficiency of use. Case studies conducted using this environment provide the means for comparing its use with

current (traditional) data reporting for CM functions related to time, quality, and change management. It is demonstrated that visual analytics enhances CM analytics capabilities applicable to a broad range of CM functions/tasks.

## Preface

The research reported in this thesis consists of identification of research problems and questions, formulation of research methodologies in pursuing answers to research questions, comprehensive and critical literature review, analysis of design guidelines and development methods of a CM data visualization environment, analysis, design, and implementation of a prototype CM data visualization environment, collecting and organizing construction management data from actual projects, and conducting case studies for assessing and seeking lessons learned from the application of data visualization technology and the prototype visualization environment developed. The topic of dissertation was proposed by the author's Ph.D. program supervisor Dr. Alan D. Russell. The data collection part of this research was an ongoing collaboration work by the author, three former master students (Miss Tanaya Korde, Mr. Ali Mehrdana, and Mr. Jehan Zeb), and a former co-op student (Mr. Phiradej Luechachandej). The programming work for implementing a prototype CM data visualization environment was done by Mr. William Wang, a senior programmer, in the Department of Civil Engineering, UBC. The author was solely responsible for the rest of other components of this research with the guidance from Dr. Alan D. Russell. The thesis includes three manuscripts:

- A version of Chapter 2 has been published. Russell, Alan. D., Chiu, Chao-Ying., Korde, Tanaya. (2009). "Visual Representation of Construction Management Data." *Automation in Construction*, 18(8), 1045-1062.
- A version of Chapter 3 has been published. Chiu, Chao-Ying., and Russell, Alan D. (2011). "Design of a Construction Management Data Visualization Environment: a Top-Down Approach." *Automation in Construction*, 20 (4), 399-417.
- A version of Chapter 4 has been submitted for publication. Chiu, Chao-Ying, and Russell, Alan D. "Design of a Construction Management Data Visualization Environment: a Bottom-Up Approach".

# Table of Contents

<b>Abstract.....</b>	<b>ii</b>
<b>Preface.....</b>	<b>iv</b>
<b>Table of Contents .....</b>	<b>v</b>
<b>List of Tables .....</b>	<b>x</b>
<b>List of Figures.....</b>	<b>xii</b>
<b>Acknowledgements.....</b>	<b>xxvii</b>
<b>Chapter 1 Introduction.....</b>	<b>1</b>
1.1 Problem statements and proposed solutions .....	1
1.2 Terminology definitions .....	12
1.3 Research questions, objectives, and hypothesis .....	16
1.4 Research scope and assumptions.....	17
1.4.1 Scope or focus .....	17
1.4.2 Assumptions .....	20
1.5 Research methodologies .....	21
1.6 State-of-the-art data visualization and its application to CM.....	24
1.6.1 Shortcomings of current applications of data visualization in CM .....	24
1.6.2 How state-of-the-art data visualization can be adopted/adapted .....	26
1.7 Research contributions .....	29
1.8 Structure of the thesis .....	46
<b>Chapter 2 Visual Representation of Construction Management Data .....</b>	<b>49</b>
2.1 Introduction.....	49
2.2 Motivation for use of visualization.....	53
2.3 Data visualization in construction .....	54
2.4 General principles of visual analytics design processes .....	58
2.4.1 Understanding the purposes of analytical reasoning .....	58
2.4.2 Organizing data representations and data transformations .....	59
2.4.3 Designing visual representations and interaction features .....	61
2.4.4 Design evaluation .....	63

2.5 Design of visual representations of change order data .....	64
2.5.1 Change order management.....	65
2.5.2 Visual representations for project 1 and design 1.....	67
2.5.2.1 Purposes of analytical reasoning for project 1 and design 1 .....	68
2.5.2.2 Choice of data representations and transformations – project 1 and design 1 .....	69
2.5.2.3 Choice of visual representations – Figures 2.1 and 2.2, project 1 and design 1 ...	70
2.5.2.4 Evaluation – project 1 .....	71
2.5.2.5 Lessons learned – project 1 .....	73
2.5.3 Visual representations for project 2 .....	74
2.5.3.1 Visual representations for project 2 and design 1, Figures 2.3 - 2.5.....	75
2.5.3.2 Visual representations for project 2 – design 2, Figure 2.6.....	84
2.5.3.3 Visual representations for project 2 and design 3, Figure 2.7 .....	86
2.6 Some general observations .....	87
2.6.1 Applying identified principles for design process .....	87
2.6.2 Evaluation and feedback .....	88
2.6.3 Organizing lessons learned for development of a general CM visual analytics model .....	89
2.6.4 Issue of CM data management .....	91
2.6.5 Data exploration flexibility .....	92
2.7 Conclusions.....	93

**Chapter 3 Design of a Construction Management Data Visualization Environment:  
a Top-Down Approach..... 95**

3.1 Introduction.....	95
3.2 Approach and structure of chapter .....	98
3.3 Data visualization in construction management.....	100
3.4 Concepts of analytics and relation to development of a construction management data visualization environment .....	102
3.4.1 Analytics for construction management .....	103
3.4.2 Visual CM analytics supported by a data visualization environment .....	107
3.4.3 Analytics for time performance management .....	109
3.4.3.1 Visual CM analytics for planning/predicting time .....	110
3.4.3.2 Visual CM analytics for monitoring/diagnosing/controlling time .....	112

3.4.3.3 Visualization requirements deduced from time performance management analytic needs .....	114
3.5 Case study of CM analytics using a data visualization environment .....	115
3.5.1 Case study overview .....	115
3.5.2 Visualization for CM analytics for planning/predicting time.....	117
3.5.3 Visualization for CM analytics for monitoring/diagnosing/controlling time .....	120
3.6 Evaluation of and extensions to the current data visualization environment .....	125
3.7 Conclusions.....	135

**Chapter 4 Design of a Construction Management Data Visualization Environment:  
a Bottom-Up Approach..... 137**

4.1 Introduction.....	137
4.2 Visualization design 1--time performance measure variance visualization.....	141
4.2.1 Visualization requirements.....	141
4.2.1.1 CM variables involved.....	142
4.2.1.2 How characteristics of time performance variance measures can be observed for identifying potential causes of time performance as a function of project context dimensions .....	142
4.2.2 Visualization design specifications.....	145
4.3 Visualization design 2--PCBS attributes visualization.....	150
4.3.1 Visualization requirements.....	151
4.3.1.1 CM variables involved.....	152
4.3.1.2 How characteristics of PCBS attributes can be observed for identifying potential impacts on context dimensions.....	152
4.3.2 Visualization design specifications.....	155
4.4 Visualization design 3--time performance cause-effect visualization .....	157
4.4.1 Visualization requirements.....	159
4.4.1.1 CM variables involved.....	159
4.4.1.2 How characteristics of construction conditions related to a certain activity can be observed for identifying abnormalities and their timing.....	161
4.4.2 Visualization design specifications.....	162
4.5 Use of a state-of-the art CM data visualization environment.....	165
4.5.1 Description of projects and project data .....	165
4.5.2 Demonstration cases .....	166

4.6 General observations .....	177
4.7 Conclusions.....	189
<b>Chapter 5 Conclusion-Summary, Answering the Research Questions, Contributions, Future Work.....</b>	<b>193</b>
5.1 Overview of conclusions .....	193
5.2 Research summaries .....	193
5.3 Demonstrating and analyzing the merit of using a CM data visualization environment ...	198
5.3.1 Demonstration cases- project 1 .....	200
5.3.1.1 Project 1- demonstration case 1.....	205
5.3.1.2 Project 1 - demonstration case 2.....	210
5.3.2 Demonstration cases- project 2 .....	217
5.3.2.1 Project 2- demonstration case 1.....	222
5.3.2.2 Project 2- demonstration case 2.....	229
5.3.3 Demonstration cases- project 3 .....	234
5.3.3.1 Project 3- demonstration case 1.....	238
5.3.3.2 Project 3- demonstration case 2.....	245
5.3.4 Analysis of demonstration results.....	250
5.3.5 Conclusions from the demonstration and analysis .....	255
5.4 Summary of research contributions .....	256
5.5 Future work.....	257
<b>Bibliography .....</b>	<b>259</b>
<b>Appendices.....</b>	<b>279</b>
Appendix A Data Visualization in Construction Management .....	279
A.1 Visual representations of planned/baseline construction conditions .....	279
A.2 Visual representations of predicted/ baseline construction performance.....	281
A.3 Visual representations of how changes of planned/baseline conditions affect predicted/baseline performance-optimizing construction plans .....	285
A.4 Visual representations of how changes of planned/baseline conditions affect predicted/baseline performance- identifying and analyzing construction risks .	291
A.5 Visual representations of actual or actual vs. planned/baseline construction conditions .....	298
A.6 Visual representations of actual or actual vs. predicted/baseline construction performance .....	300

A.7	Visual representations of dependency/cause-effect between conditions and performance .....	304
A.8	Interacting with computerized visual representations of CM data .....	306
Appendix B	Overview of State-of-the-Art Data Visualization .....	313
B.1	Introduction to data visualization .....	313
B.2	Toward better visual representations for analytics .....	316
B.3	Interacting with computerized visual representations.....	320
B.4	Designing and developing data visualization tools.....	322
B.5	State-of-the-art data visualization tools .....	327

## List of Tables

Table 1.1 Correspondence between research questions and research methodologies for answering them .....	22
Table 1.2 Primary common CM visualization features: presenting construction conditions or performance measures mapped against primary project context dimensions .....	37
Table 1.3 Secondary common CM visualization features: secondary feature consideration after addressing the primary visualization features .....	42
Table 1.4 Design guidelines in addition to those presented in Chapter 2 .....	43
Table 2.1 Change order properties of interest.....	66
Table 2.2 Summary of visual representation evaluations for Figures 2.3 and 2.4, project 2 .....	80
Table 3.1 Data visualization environment requirements for visual CM analytics and conformance/non-conformance of current environment ( <i>monitoring/diagnosing for time performance without the use of explicit explanatory CM models</i> ) ...	127
Table 3.2 Visualization features that are available in or in development process for the current CM data visualization environment .....	130
Table 4.1 Summary of data representations/transformations for time performance variance measure visualization .....	146
Table 4.2 Summary of data representations/transformations for quantitative PCBS attribute visualization .....	155
Table 4.3 Summary of data representations/transformations for visualizing the distribution of construction conditions versus time .....	163
Table 5.1 Descriptions of the exploration- answer process using the CM data visualization environment and current (traditional) data reporting functionalities: project 1-demonstration case 1 .....	206
Table 5.2 Descriptions of the exploration- answer process for both the use of a CM data visualization environment and the current data reporting functionalities in project 1-demonstration case 2 .....	211

Table 5.3 Descriptions of the exploration-answer process for both the use of a CM data visualization environment and current data reporting functionalities: project 2-demonstration case 1 .....	223
Table 5.4 Descriptions of the exploration-answer process for both the use of a CM data visualization environment and the current data reporting functionalities: project 2-demonstration case 2.....	230
Table 5.5 Descriptions of the exploration-answer process for both the use of a CM data visualization environment and current data reporting functionalities: project 3-demonstration case 1 .....	239
Table 5.6 Descriptions of the exploration-answer process for both the use of a CM data visualization environment and the current data reporting functionalities in project 3-demonstration case 2 .....	246
Table 5.7 Summarized comparison between the use of a CM data visualization environment and current (traditional) data reporting features for the six demonstration cases .....	251
Table B.1 Representative state-of-the-art data visualization tools and their general functionality .....	328

## List of Figures

Figure 1.1 Data flows for the task "develop schedule" as suggested by PMBOK (2008) ..	3
Figure 1.2 Data flows for the task "monitor and control project work" as suggested by PMBOK (2008) .....	4
Figure 1.3 The relationship between several current state-of-the-art CM_IS, computer assistance, human judgment, and conducting stand alone CM tasks.....	5
Figure 1.4 One of many pages of CM_IS generated tabular reports of planned/actual schedule data and problems encountered during execution .....	7
Figure 1.5 One of many pages of Microsoft Word documents of deficiency lists .....	8
Figure 1.6 One of many pages of Microsoft Excel spreadsheets for change order data .....	9
Figure 1.7 A proposed solution in terms of enhancing data reporting functionalities in the relationship between a current state-of-the-art CM_IS, computer assistance, human judgment, and conducting stand alone CM tasks.....	10
Figure 1.8 A two way structured top-down and bottom-up CM data visualization environment development process.....	34
Figure 1.9 A user interface for image selection by construction conditions that are grouped under project views (i.e. the tab items such as "process", "as-built"). This user interface only allows users to choose images representing distribution of values of construction conditions in certain project context dimensions. This figure showcases the primary common visualization feature of “image theme by construction conditions or performance measures” described in the 1st row of Table 1.2.....	39
Figure 1.10 Two record distribution images visualizing number of deficiencies distributed in different project context dimensions: (a) product and location dimensions, (b) project participant and location dimensions. This figure showcases the primary common visualization feature of “image type by context dimensions” described in the 2nd row of Table 1.2.....	39
Figure 1.11 Records distribution images visualizing the number of deficiencies distributed in the location dimension but at different levels of granularity: (a) location set (stories), (b) location (rooms). This figure showcases the primary	

common visualization feature of “image contents by granularity of context dimensions” described in the 3rd row of Table 1.2 .....	40
Figure 1.12 Records distribution images visualizing the number of deficiencies distributed in the location dimension but of different data value ranges: (a) “all” deficiencies, (b) only “long lead time” deficiencies. This figure showcases the primary common visualization feature of “image contents by items selection of project context dimensions” described in the 4th row of Table 1.2 .....	40
Figure 1.13 The user interface (the left bottom combo box) for adjusting data status (i.e. planned, actual, and planned vs. actual work area received percentage). This figure showcases the primary common visualization feature of “image contents by data status states” described in the 5th row of Table 1.2 .....	41
Figure 2.1 Project 1 CO history in terms of ID & Location, timing and value of work ...	68
Figure 2.2 Project 1 History of COs by location, time, responsibility and number .....	69
Figure 2.3 Project 2 Number and reasons for change orders.....	76
Figure 2.4 Project 2 Distribution of value and reasons for change orders .....	79
Figure 2.5 Project 2 Stacked graphs for number, values and reasons for change order....	83
Figure 2.6 Project 2 Distribution of change orders by physical system.....	85
Figure 2.7 Project 2 Causal model reasoning – number of COs and corresponding schedule update dates and projected completion dates .....	87
Figure 3.1 Differentiating between current state-of-art CM systems and potential of systems with a formal visualization environment applicable to a wide range of functions .....	97
Figure 3.2 CM analytics flow charts .....	105
Figure 3.3 Product view – (a) project locations; (b) location attributes; and (c) photo of components.....	116
Figure 3.4 LP and Bar Chart schedule representations .....	118
Figure 3.5 Tiled LP chart views of: (a) as-built schedule for all activities, and (b) planned vs. actual work trajectory for lead activity, excavate & mud slab.....	121
Figure 3.6 LP chart (a) and problem status charts- problem status by problem code vs. time (b); problem status by problem code vs. location (c).....	123

Figure 3.7 Parallel comparisons of operation process of using a CM data visualization environment and the thought process of CM analytics it supports reflected in Figure 3.6.....	126
Figure 3.8 Generate and juxtapose (a) production rate chart, (b) activity status charts, and (c) environment condition charts .....	134
Figure 4.1 Illustration of the definitions of time performance and time performance variances .....	143
Figure 4.2 (a) A specific CM analytical reasoning task/question regarding time performance variance measures, (b) corresponding data dimensions and visual encodings, and (c) the generated visual representation.....	148
Figure 4.3 (a) A specific CM question related to product attributes, (b) corresponding data dimensions and visual encodings, (c) the generated visual representation for the entire data space, and (d) the generated visual representation for an area of interest in (c).....	156
Figure 4.4 Illustration of concept of cause-effect reasoning for identifying potential reasons for time performance. ....	158
Figure 4.5 (a) A specific CM question pertaining to the construction conditions of activity execution status and problem status, (b) corresponding data dimensions and visual encodings, and (c) the generated visual representation.....	164
Figure 4.6 (a) A screenshot of a 3D version of schedule variance graphics presenting the duration variance values of the activity sets related to constructing substructures for project 1 for all locations, (b) a screenshot of a 3D version of schedule variance graphics zooming into regions of interest seen in Figure 4.6(a) (i.e. the phase 5 locations, from locations F533L to F 458).....	168
Figure 4.7 Three 2D versions of PCBS attribute graphics presenting the distribution of planned vs. actual values for location attributes: (a) percentage work area received, (b) underground (UG) utilities relocation by others, (c) overhead (OH) utilities relocation by others. ....	171
Figure 4.8 A 3D version of schedule variance graphics presenting the activity time variance values of the sets of trade activities related to construction work at the parkade location for project 2 .....	173

Figure 4.9 A multiple view image representing selected construction conditions associated with the activity "shotcrete shoring at parkade level": (a) comparison schedule (progress date of 31 December 03 vs. planned project early start date of 20 October 03), (b) activity status, (c) problems encountered, (d) temperature, (e) ground conditions, (f) daily and cumulative precipitation. ....	175
Figure 4.10 A multiple view image representing selected construction conditions associated with the activity "bulk excavate substructure at parkade level": (a) problems encountered, (b) daily and cumulative precipitation, (c) comparison schedule (progress date of 31 December 03 vs. planned project early start date of 20 October 03), (d) activity status, (e) Equipment (truck) planned resource usage, (f) Equipment (hydraulic excavator) planned resource usage (for (e) and (f) the early and late plots are identical because the activity is a critical one) . ....	176
Figure 4.11 An illustration of the hierarchical relationship between various time performance variance measures. ....	184
Figure 4.12 In contrast to Figure 4.8, different bar shapes are used to represent different levels in the hierarchical relationship between various time performance variance measures. ....	186
Figure 4.13 An improvement on Figure 4.8 by using a mock-up of stacked-bar charts to represent the hierarchical relationship about summing together quantities of various time performance variance measures. ....	187
Figure 4.14 An improvement on Figure 4.7(c) by using two more distinctive colors for representing planned and actual data status and by reversing the ordering of labeling on the Z axis. ....	188
Figure 4.15 A mock-up 2D version image for the 3D graphics shown in Figure 4.8. ....	190
Figure 5.1 The PCBS (product) view of project 1. The dialogue box shows how users can define attributes (e.g. concrete quantity) for product items. ....	201
Figure 5.2 The PCBS view (physical work locations) of project 1 .....	202
Figure 5.3 The process view for project 1 (correspond to original modeling of the project). ....	203

Figure 5.4 A photo of actual columns of project 1..... 204

Figure 5.5 The organization view of project 1 ..... 204

Figure 5.6 A traditional bar graph of as-planned (blue bars) vs. as-built (green bars) schedule representing the "excavate and mud slab" activity executed in the location range F 737 to F 654..... 208

Figure 5.7 A non-traditional as-planned (blue lines) vs. as-built (green lines) schedule generated from a CM data visualization environment representing the "excavate and mud slab" activity executed in the location range F 737 to F 654 ..... 209

Figure 5.8 First page of the 14 page tabular report of planned values of product attributes (e.g. concrete quantity, formwork area) for the foundations and columns at all work locations..... 213

Figure 5.9 (a&b) Number and lengths of piles by location, (c&d) number and lengths of rock anchors by location..... 214

Figure 5.10 (a) Planned formwork areas, (b) Planned concrete quantities, (c) Planned reinforcing bar lengths required by the foundations and columns at all locations..... 215

Figure 5.11 (a) Planned schedule for "pour footing" (left connecting lines) and "pour column" (right connecting lines) activities executed at the first 54 locations, (b) Planned concrete quantities required by the foundations and columns at all locations..... 216

Figure 5.12 The PCBS view (physical locations) of project 2 ..... 218

Figure 5.13 The PCBS view (products) of project 2..... 219

Figure 5.14 The organizational view (project participant) of project 2 ..... 220

Figure 5.15 The As-built view (deficiency records) of project 2. The dialogue box shows how users can associate a deficiency item with items from other views such as the project participant view and PCBS view ..... 221

Figure 5.16 The first page of the 304 page tabular report of deficiency records that include information about project participants who are responsible for the deficiencies, deficient products, and locations of the products. .... 225

Figure 5.17 Number of deficiencies distributed in the location dimensions of two levels of location dimension granularity ("location set" and "location" levels) and the project participant dimension at the level of granularity of individual "project participant". .....	226
Figure 5.18 Number of deficiencies (excluding the ones of the Painter and Cleaning trades) distributed in two levels of location dimension granularity ("location set" and "location" levels) and the project participant dimension at the level of granularity of individual "project participant" .....	227
Figure 5.19 Number of (a) Painter trade deficiencies, and (b) Cleaning trade deficiencies distributed in the product dimensions of three different levels of granularity ("System", "Subsystem", and "Element" levels) .....	228
Figure 5.20 The first page of a 16 page tabular report of deficiencies that require a longer time to fix. It includes information about project participants who are responsible for the deficiencies, deficient products, locations of the products, and types of deficient work. ....	232
Figure 5.21 Number of long lead time deficiencies (i.e. deficiencies that need a longer time to correct) distributed in: (a) the location dimensions at two levels of granularity ("location set" and "location" levels) and the project participant dimension at the level of granularity of "project participant", (b) the keyword (i.e. deficiency type) dimension at the level of granularity of "2nd level of deficiency definitions" and the the project participant dimension at the level of granularity of "project participant". ....	233
Figure 5.22 The PCBS view (both product and location) of project 3.....	235
Figure 5.23 The organizational view (project participant) of project 3 .....	236
Figure 5.24 As-built view (change order records) for project 3. The dialogue box shows how users can associate a change order item with items from other views such as the project participant view and PCBS view .....	237
Figure 5.25 The first page of the 43 page tabular report of change orders that include information about project participants who will execute them, products and locations involved, and types of change order. ....	241

Figure 5.26 Number of (a) scope change orders, (b) design change orders, (c) site condition change orders, (d) owner change orders, distributed in the product dimension at the level of granularity of "system" and the location dimension at the level of granularity of "location"..... 242

Figure 5.27 Number of (a) scope change orders related to the building site work , (b) design change orders related to the interior and service systems, distributed in the product dimensions of three different levels of granularity ("System", "Subsystem", and "Element" levels) and the location dimensions of two levels of granularity ("location" and "sub-location" levels"). ..... 243

Figure 5.28 Number of change orders distributed in the project participant dimension at the level of granularity of "project participant" ..... 244

Figure 5.29 The first page of the 24 page tabular report of change orders that includes information about project participants who will execute the change orders, products and locations involved with a change order, and change order issue date. .... 248

Figure 5.30 (a) Number of all change orders distributed in the time dimension at the level of granularity of "month" and the project participant dimension at the level of granularity of "project participant". Number of change orders issued between mid February and June 2005 that are associated with the (b) Broadway trade, and (c) Celtic trade, distributed in the time dimension at the level of granularity of "month" and the location dimension at the level of granularity of "location". ..... 249

Figure A.1 has been removed due to copyright restrictions. It was a graphics visualizing planned resource allocation. Original source: O'Brien, J. J. (1965). CPM in Construction Management 1st Ed: Fig. 15.1., pp 186 ..... 280

Figure A.2 has been removed due to copyright restrictions. It was a graphics visualizing activity sequencing. Original source: de Leon, G. P. (2008). Project Planning using Logic Diagramming Method. AACE International Transactions: Figure 1, pp. PS.S05.03 ..... 280

Figure A.3 Visualizing temporal and spatial distribution of activities. Source: (The National Building Agency 1968) ..... 281

Figure A.4 A construction schedule in bar graph format seen as early as 1917. Source: (Brinton 1939) .....	282
Figure A.5 Network diagram--activities on arrows. Source: (Fondahl 1962) .....	283
Figure A.6 Network diagram--activities on nodes. Source: (Fondahl 1962) .....	283
Figure A.7 has been removed due to copyright restrictions. It was a graphics visualizing S curve. Original source: O'Brien, J. J. (1965). CPM in Construction Management 1st Ed: Fig. 14.5., pp 168 .....	284
Figure A.8 Cash flow diagram. Source: (Cooke and Williams 2004) .....	285
Figure A.9 Changing planned construction conditions (resources) vs. changed forecast performance (time) in multiple views. Source: (Russell et al. 2009) .....	287
Figure A.10 Changing planned construction conditions (profit margin desired) vs. changed forecast performance (cash flow) in a single view. (© 2000 IEEE. Reprinted, with permission, from Khosrowshahi, Information Visualization in aid of Construction Project Cash Flow Management, Proceedings of the International Conference on Information Visualisation,2000).....	288
Figure A.11 Changing planned construction conditions (resources) vs. changed forecast performance (unit cost/productivity) in a single view. T = team; S = saw; O = trucking old panels, N = trucking new panels. Source: (Zhang et al. 2008)	289
Figure A.12 A nomograph that encodes a mathematical model predicting required crew size. The model considers factors such as CPM duration, project deadline, number of sites for a repetitive activity. Source: (Elhakeem and Hegazy 2005) .....	290
Figure A.13 A four variable influence diagram representing uncertain relationships between them. Source: (Diekmann 1992) .....	292
Figure A.14 has been removed due to copyright restrictions. It was a graphics visualizing distribution in time and space of risk from project participants. Original source: Korde, T., Wang, Y., Russell, A. D. (2005). Visualization of Construction Data. Proceedings of 6th Construction Specialty Conference: Figure 3, pp. CT-148-6.....	292
Figure A.15 has been removed due to copyright restrictions. It was a Tornado Plot visualizing how negative/positive 10% change in independent variables	

<p>affects the net present value. Original source: Vrijland, M. S. A. (2003).  Visual Display of Sensitivity and Risk. AACE International Transactions:  Figure 2, pp. RISK 11.4 .....</p>	293
<p>Figure A.16 has been removed due to copyright restrictions. It was a Spider Plot  visualizing how negative and positive percentage changes in various  independent variables affect net present value. Original source: Vrijland, M.  S. A. (2003). Visual Display of Sensitivity and Risk. AACE International  Transactions: Figure 7, pp. RISK 11.7.....</p>	293
<p>Figure A.17 has been removed due to copyright restrictions. It was a Radar plot showing  sensitivity scores for the independent variables. Original source: Vrijland, M.  S. A. (2003). Visual Display of Sensitivity and Risk. AACE International  Transactions: Figure 8, pp. RISK 11.8.....</p>	294
<p>Figure A.18 Probability distribution (includes cumulative probability distribution) for  right of way cost and construction cost. Source: (Washington State  Department of Transportation 2010).....</p>	295
<p>Figure A.19 Probability impact matrix showing likelihood and degree of schedule/cost  consequence of risk events associated with a highway project. Source:  (Washington State Department of Transportation 2010) .....</p>	296
<p>Figure A.20 Tabularized risk register of a highway project. Source: (Washington State  Department of Transportation 2010).....</p>	297
<p>Figure A.21 Representing as-built data of problems and site conditions. Source: (Russell  and Udaipurwala 2002a) (with Permission from ASCE).....</p>	298
<p>Figure A.22 has been removed due to copyright restrictions. It was a graphics visualizing  actual crew size and planned vs. actual crew size. Original source: Pinnell, S.  S. (1998). How to Get Paid for Construction Changes: Figures 10.38 and  10.39, pp. 330 .....</p>	299
<p>Figure A.23 "Causes and who" responsible for change order costs. Source: (Cox et al.  1999) .....</p>	300
<p>Figure A.24 Actual daily status of activities. Source: (Hegazy et al. 2005).....</p>	301

Figure A.25 Actual time performance of activities is encoded in colors and projected onto product components corresponding to those activities. Source: (Song et al. 2005) ..... 301

Figure A.26 Actual activity status at operational level of detail. (© 2008 Reprinted, with permission, from Vrotsou, K., Ynnerman, A., Cooper, M., Seeing Beyond Statistics: Visual Exploration of Productivity on A Construction Site, Proceedings of 2008 International Conference in Visualization, 2008) ..... 301

Figure A.27 has been removed due to copyright restrictions. It was a graphics visualization of EVM index. Original source: Anbari, F. T. (2003). Earned Value Project Management Method and Extensions. Project Management Journal 34(4): Figures 4, 5, and 9, pp. 14 and 16. .... 302

Figure A.28 Actual vs. planned cost distribution in activities at different levels of detail. Source: (Nie et al. 2007)..... 303

Figure A.29 has been removed due to copyright restrictions. It was a Tree-map representations for cost index (i.e. actual cost of work performed/budgeted cost of work performed) of pay items of different levels of detail - the color scale is used for representing cost index values. Original source: Songer, A. D., Hays, B., North, C. (2004). Multidimensional Visualization of Project Control Data. Construction Innovation: Information, Process, Management 4(3): Figure 4, pp. 185..... 303

Figure A.30 has been removed due to copyright restrictions. It was a graphics juxtaposing weather conditions with activity status for validating/invalidating cause-effect relationship between the two. Original source: Zeb, J., Chiu, C., Russell, A. (2008). Designing a Construction Data Visualization Environment. Proceedings of the 1st Forum on Construction Innovation: Figure 4, pp. 7 . 304

Figure A.31 Visual representation of output data of an explanatory model using generic C4.5 decision-tree classification rules for explaining reasons for delays in pipeline laying activities. Source: (Soibelman and Kim 2002) ..... 305

Figure A.32 Visual representations of output data of an explanatory model using generic relevance partitioning/significance testing rules to explain reasons for the increase of budgeted cost. Source: (Roth and Hendrickson 1991) ..... 306

Figure A.33 A multiple view created by the Vico Control commercial software including images of: (a) schedule in flow line format, (b) resource usage, (c) cash flow, and (d) activity status. .... 309

Figure A.34 A system that can generate computer graphics in both: (a) flow line, (b) bar graph, and (c) network formats for visualizing part of the schedule of a transit guideway project. Source: (Russell and Udaipurwala 2002b) ..... 310

Figure A.35 A system incorporating the flow line representation with an simulation model for both computationally and visually optimizing a project's schedule. Source:(Hegazy and Kamarah 2008) ..... 311

Figure A.36 Visualization of EVM indices for any selectable combination of location and product. Source: (Zhang et al. 2009)..... 312

Figure B.1 A graphic (Fig. 5) presenting data of "Expansion of Air" (vertical coordinate) and "Height of Mercury" (horizontal coordinate). Source: (Halley 1686).... 314

Figure B.2 A graphic comparing wages of a "good mechanic" (line) with wheat prices (bars) from the year 1565 to 1821. Source: (Playfair 1821); Image from: (Friendly 2008) ..... 314

Figure B.3 A specimen of a time line chart presenting the names, lengths of lives, birth and death dates, and occupations of the most distinguished persons from BC 1200 to AD 1800. Source:(Priestley 1744) ..... 315

Figure B.4 has been removed due to copyright restrictions. It was a table of guidelines of using multiple views in terms of rules and their corresponding positive and negative impacts on the utility of information visualization. Original source: Wang Baldonado, M. Q., Woodruff, A., Kuchinsky, A. (2000). Guidelines for Using Multiple Views in Information Visualization. Proceedings of the Working Conference on Advanced Visual Interfaces: Table 1, pp. 118..... 320

Figure B.5 A classification of functions of interaction features in information visualization. (© 1996 IEEE. Reprinted, with permission, from Chuah, M. C., and Roth, S. F., On the Semantics of Interactive Visualizations, Proceedings of the 1996 IEEE Symposium on Information Visualization (INFOVIS '96), 1996) ..... 321

Figure B.6 has been removed due to copyright restrictions. It was a graphics showing the use of sliders to specify attribute values of chemical elements. In the Periodic Table, elements that do not meet the query specifications are dimmed in real time while the user moves the buttons of sliders. Original source: Ahlberg, C., Williamson, C., Shneiderman, B. (1992). Dynamic Queries for Information Exploration: an Implementation and Evaluation. CHI '92 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Figure 2, pp. 621..... 330

Figure B.7 A data visualization system with intuitive interfaces for users to select data dimensions, filter data ranges, sort and group data, and specify how data dimensions to be mapped to visual variables. (© IEEE 2002. Reprinted, with permission, from Stolte, C., Tang, D., Hanrahan, P., Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases, Transactions on Visualization and Computer Graphics 8(1), 2002)..... 330

Figure B.8 Selecting data points in a scatter plot by a rectangular brush and data points in other scatter plots associated with the selected data get highlighted. Source: (Cleveland and Becker 1987) (Reprinted with permission from Technometrics Copyright 1987 by the American Statistical Association) ..... 331

Figure B.9 Using a "mouse lasso" to replace with, intersect with, add, subtract, or toggle the previous rectangle-select data points. (© IEEE 1996. Reprinted, with permission, from Wills, G., Selection: 524,288 Ways to Say "This Is Interesting", Proceeding of IEEE Symposium on Information Visualization, 1996) ..... 331

Figure B.10 has been removed due to copyright restrictions. It was graphics showing (a) A magnification function by which the scale of an image is determined by distance from the focal point, (b) application of the magnification function to the horizontal dimension, (c) application of the magnification function to both horizontal and vertical dimensions. Original source: Leung, Y. K., and Apperley, M. D. (1994). A Review and Taxonomy of Distortion-Oriented Presentation Techniques. ACM Trans. Comput. -Hum. Interact.: Figures 5(b)~5(d), pp. 133 ..... 332

Figure B.11 has been removed due to copyright restrictions. It was two graphics showing interaction features that allow users to specify sizes of 3D bars thereby reducing occlusion problems. The blue and red bars in (b) were adjusted to be thinner than the ones in (a) so that green bars can be seen more clearly. Original source: Chuah, M. C., Roth, S. F., Mattis, J., Kolojejchick, J. (1995). SDM: Selective Dynamic Manipulation of Visualizations. Proceedings of UIST' 95 Symposium on User Interface Software and Technology: Figures 8 and 11, pp. 65..... 332

Figure B.12 A interactivity mechanism by which data density changes along with the zooming in/out of a map. (© IEEE 2003. Reprinted, with permission, from Stolte, C., Tang, D., Hanrahan, P., Multiscale Visualization Using Data Cubes, IEEE Trans. Visual. Comput. Graphics 9(2), 2003)..... 333

Figure B.13 An anomalous pattern at the "Morris" barley site can be identified in the left "Trellis view" but not in the right view while: in (a) dot charts are sorted by median yields of barley sites (from bottom to top) and by median yield of year; data points in each dot chart are sorted by median yields of barley varieties, in (b) dot charts and data points in a dot chart are simply sorted by alphanumeric ordering. Source: (Becker et al. 1996) (Reprinted with permission from Journal of Computational and Graphical Statistics Copyright 1996 by the American Statistical Association)..... 334

Figure B.14 has been removed due to copyright restrictions. It was a flow chart of a multiple view coordination model. In this model, an event is initiated by users in one of the two views (V1, V2). This event may include basic visualization processes such as filtering data (i.e. enhance), mapping data to visual variables (i.e. map), rendering an images (i.e. render), and rotating view points (i.e. transform), which consist of a coordination space that is applicable to both views. Source: (Boukhelifa et al. 2003) (© IEEE 2003. Reprinted, with permission, from Boukhelifa, N., Roberts, J. C., Roberts, P. J., Rodgers, P. J., A Coordination Model for Exploratory Multi-view Visualization, Proceedings of the Conference on Coordinated and Multiple Views in Exploratory Visualization, 2003) ..... 335

- Figure B.15 has been removed due to copyright restrictions. It was a multiple view of many 2D vertical bar charts. Original source: Pirolli, P., and Rao, R. (1996). Table Lens as a Tool for Making Sense of Data. Proceedings of the Workshop on Advanced Visual Interfaces: Figure 1, pp. 69 ..... 336
- Figure B.16 A multiple view of a choropleth map, a parallel coordinates chart, a scatter plot matrix, and a scatter plot for users to explore insights from different angles in a demographic dataset. (© IEEE 2005. Reprinted, with permission, from Feldt, N., Pettersson, H., Johansson, J., Jern, M., Tailor-made Exploratory Visualization for Statistics Sweden, Proceedings of the Coordinated and Multiple Views in Exploratory Visualization, 2005) ..... 336
- Figure B.17 Parallel coordinate representations for visualizing multi-dimensional fatal accident data at different levels of detail. (© IEEE 1999. Reprinted, with permission, from Fua, Y., Ward, M. O., Rundensteiner, E. A., Hierarchical Parallel Coordinates for Exploration of Large Datasets, Proceedings of the Conference on Visualization '99, 1999) ..... 337
- Figure B.18 has been removed due to copyright restrictions. It was a "same bin size Mosaic Plot" that represents the data about answers to the question of "how you heard about survey". The degree of grey shades represents counts of responses to certain answer options. Original source: Hofmann, H. (2006). Multivariate Categorical Data-Mosaic Plots. Graphics of Large Datasets- Visualizing a Million: Figure 5.13, pp. 120 ..... 337
- Figure B.19 A multiple view of three images that shows the biological data of stickleback and pufferfish with regard to synteny relationship at the : (a) genome, (b)chromosome, and (c)block level. (© IEEE 2009. Reprinted, with permission, from Meyer, M., Munzner, T., Pfister, H., MizBee: A Multiscale Synteny Browser, IEEE Trans. Visual. Comput. Graphics 15(6), 2009)..... 338
- Figure B.20 has been removed due to copyright restrictions. It was a visualization called "Table Lens" that allows users to view detailed values (i.e. focus) and overall patterns (i.e. context) simultaneously. Original source: Rao, R., and Card, S. K. (1994). The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus + context Visualization for Tabular Information.

Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Color Plate 1, pp. 481 .....	338
Figure B.21 A tree map representation that shows planned values (rectangle sizes) and cost variances (colors) of both individual projects and project groups (grouped by departments). (© IEEE 2004. Reprinted, with permission, from Chintalapani, G., Plaisant, C., Shneiderman, B., Extending the Utility of Treemaps with Flexible Hierarchy, Proceedings of Eighth International Conference on IV, 2004).....	339
Figure B.22 has been removed due to copyright restrictions. It was a graphics showing a non-Euclidian geometry providing a smoothly varying focus+ context for visualizing hierarchical data. Original source: Lamping, J., and Rao, R. (1996). Visualizing Large Trees Using the Hyperbolic Browser. CHI '96: Proceedings of Conference Companion on Human Factors in Computing Systems: Figure 1, pp. 388 .....	339
Figure B.23 has been removed due to copyright restrictions. It was a graphics showing the use of 3D virtual space to house 2D and 3D graphics for visualizing data from Statistics Canada. Original source: Brath, R. (2003). Paper Landscapes: A Visualization Design Methodology. Proceedings of Conference on Visualization and Data Analysis (VDA 2003): Figure 4(right), pp. 131. ....	340

## **Acknowledgements**

Many people contributed to the conception, growth, and blossoming of this challenging research undertaking. Of many helps I received during this journey, my deepest gratefulness is to my research advisor Professor Alan Russell. His prophetic and realistic vision, wisdom, and knowledge provide invaluable guidance throughout this research program. He also provides me with firm support morally and financially so that I can persist and persevere with the research work.

Secondly, I would like to express my thanks to my supervisory committee members, Professor Scott Dunbar, Professor Sheryl Staub-French, and Professor Tamara Munzner for their insightful advices and guidance throughout my research work. Special thanks are to Professor Tamara Munzner for her provision of exceptional expertise in information visualization that is core to this research.

The programming work by Mr. William Wong helped to implement my research ideas and is particularly appreciated. Without a workable prototype, this research work could be very challenging. Special thankfulness goes out to former and current fellow students who were involved in the data collection work; Tanaya Korde, Ali Mehrdana, Jehan Zeb, and Phiradej Luechachandej, thank you all. Concert Properties Ltd. and Scott Management and Group are also greatly appreciated for their generosity in providing their project data.

Lastly, I am very grateful for my family's full support of my pursuing this Ph.D. research. My father Hui-Chin Chiu and mother Su-Chen Chiang play the most important role in encouraging me to hold on to my dream. Dedication must go to my wife Hsiu-Wen Tsai for her sacrifice of discontinuing her job to accompany me and care for our two lovely, healthy, and considerate daughters, Yan-Lin and Yan-Chi, who make this journey more joyful.

## **Chapter 1 Introduction**

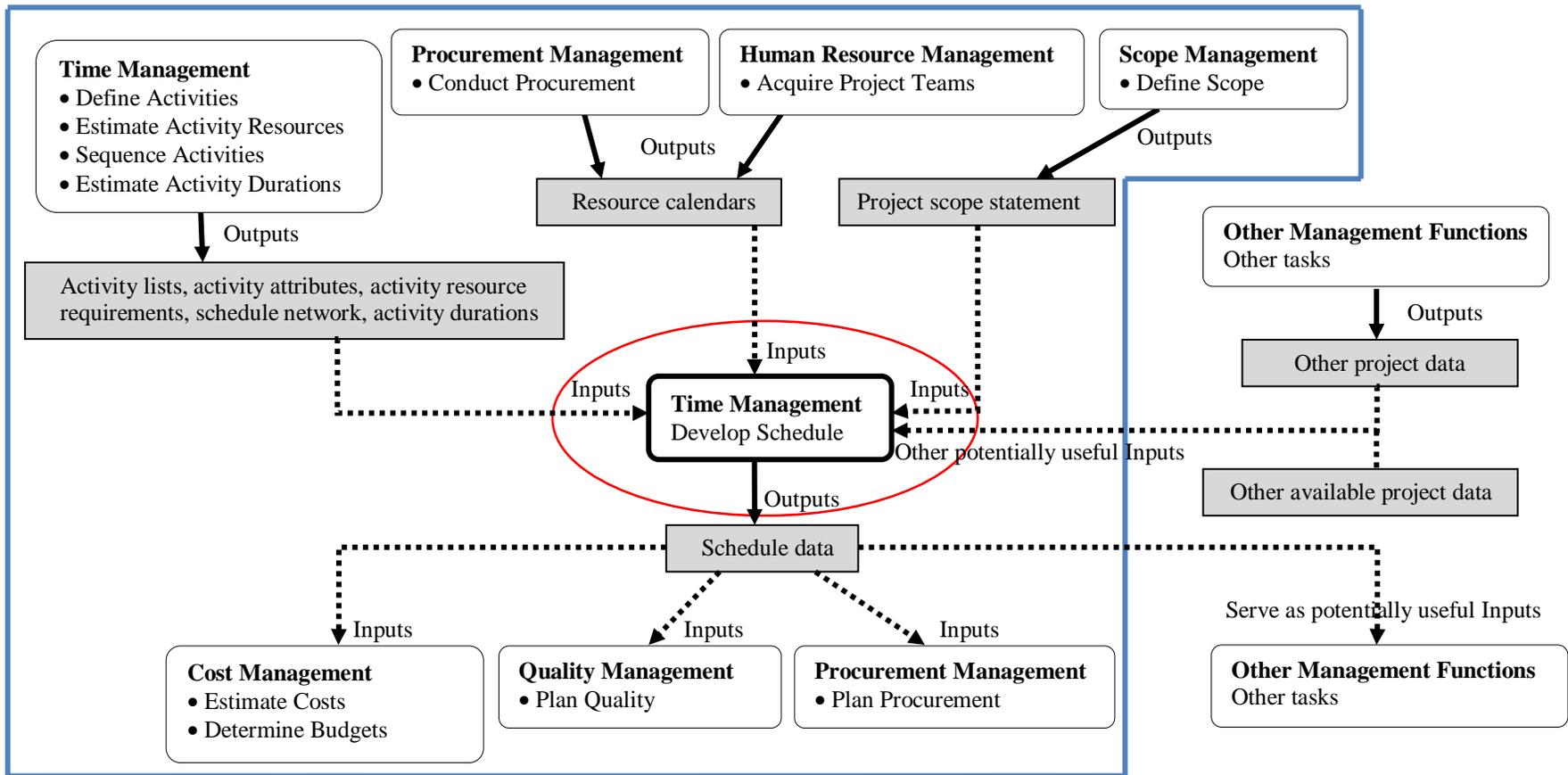
This thesis is a manuscript-based document describing the research topic of applying data visualization technology to construction management (CM). The research centers on seeking answers to three research questions. Answers to the first two research questions: "*How should a CM data visualization environment be developed?*" and "*What are the key features of a CM data visualization environment that best reflect the functions expected for it?*" are described in three manuscripts. Two have been published (Chapters 2 and 3) and the third (Chapter 4) has been submitted for review. In terms of the overall structure of the thesis, Chapter 1 describes the background, goals, methodologies, literature review (presented in the form of appendices), and contributions of the research. Chapters 2 through 4 describe the findings of phases 1 through 3 of the research work respectively (see section 1.5 Research Methodologies). Chapter 5 concludes this thesis by reporting on the fourth and final phase of the research work in terms of answering the last research question: "*How does the use of a CM data visualization environment help conduct CM analytics that cannot be done or which are difficult to do with current data reporting practices?*". It summarizes the research work and contributions, and suggests future work for tackling challenges/limitations encountered in this research endeavor. The literature review that underpins the research work is organized in two major Appendices due to the number and richness of the images associated with the text for describing the results of the literature review. Appendix A treats a full literature review of the use of data visualization in construction management (including the data visualization capabilities of current commercial CM information systems), while Appendix B treats an overview of state-of-the-art data visualization technologies.

### **1.1 Problem statements and proposed solutions**

Construction management processes serve several functions including time, cost, scope, and quality management as well as their integration. Each function requires CM personnel to conduct one or more tasks, which, in general terms, involves referencing "input" information and knowledge and utilizing techniques to produce "output" information and knowledge useful for other tasks. Two examples of tasks are "develop schedules" (time management function; planning) and "monitor and control project work"

(integration management function; monitoring and controlling), as elaborated upon in the Project Management Body of Knowledge (PMBOK; (Institute 2008)). Data flows of inputs/outputs and their originating/destination management functions/tasks as suggested by PMBOK are illustrated in Figures 1.1 and 1.2. In Figure 1.1, the task of interest, "develop schedule", is enveloped in a red oval. For this task, users need to consider the impact of referencing input data which in turn are outputs from various management functions/tasks as shown in the text boxes above the red oval in Figure 1.1. The outputs of this task, the schedule data, are in turn reference information for other management tasks (i.e. have impacts on other management tasks) as shown in the text boxes below the red oval in Figure 1.1. The items enveloped by the blue lines in Figure 1.1 correspond to the scope of good practice suggested by PMBOK; contents outside of the blue envelope represent other inputs that may be relevant to the "develop schedule" task and other management functions/tasks that may be impacted by its outputs (i.e. schedule data). The conventions used in Figure 1.2 are the same as the ones for Figure 1.1, but here the task of interest is "monitor and control project work". Two layers of data flow for generating inputs for this task are depicted.

Explicit CM knowledge developed over many years has been codified into good practices/guidelines/rules/procedures as to the tasks involved in each management function, and the relevant inputs/outputs/tools for executing the task. A complete CM process encompasses many tasks, which can be grouped by CM functions (e.g. management of time, cost, scope, quality), project phases (pre-planning, execution, post-execution) and/or CM purposes (planning/predicting, monitoring/diagnosing/controlling). Usually inputs for one task are outputs from other tasks. A CM information system (shortened as CM\_IS hereafter) can be developed for providing a computerized environment that implements good practices/guidelines/rules/procedures to support CM tasks. Automation in support of the execution of CM tasks to generate outputs constitutes "computer assistance" by which the CM\_IS produces outputs, given the requisite inputs. The manual process of executing CM tasks and generating outputs relies partly on the judgment of CM personnel by referencing relevant input/output information and their own CM knowledge. This human judgment includes detecting and considering patterns



Contents enveloped by blue lines represent good practices suggested by the PMBOK in terms of one-layer sources of inputs and usefulness of outputs for the "develop schedule" task

Management function/management task      Inputs for or outputs from a management task

Figure 1.1 Data flows for the task "develop schedule" as suggested by PMBOK (2008)

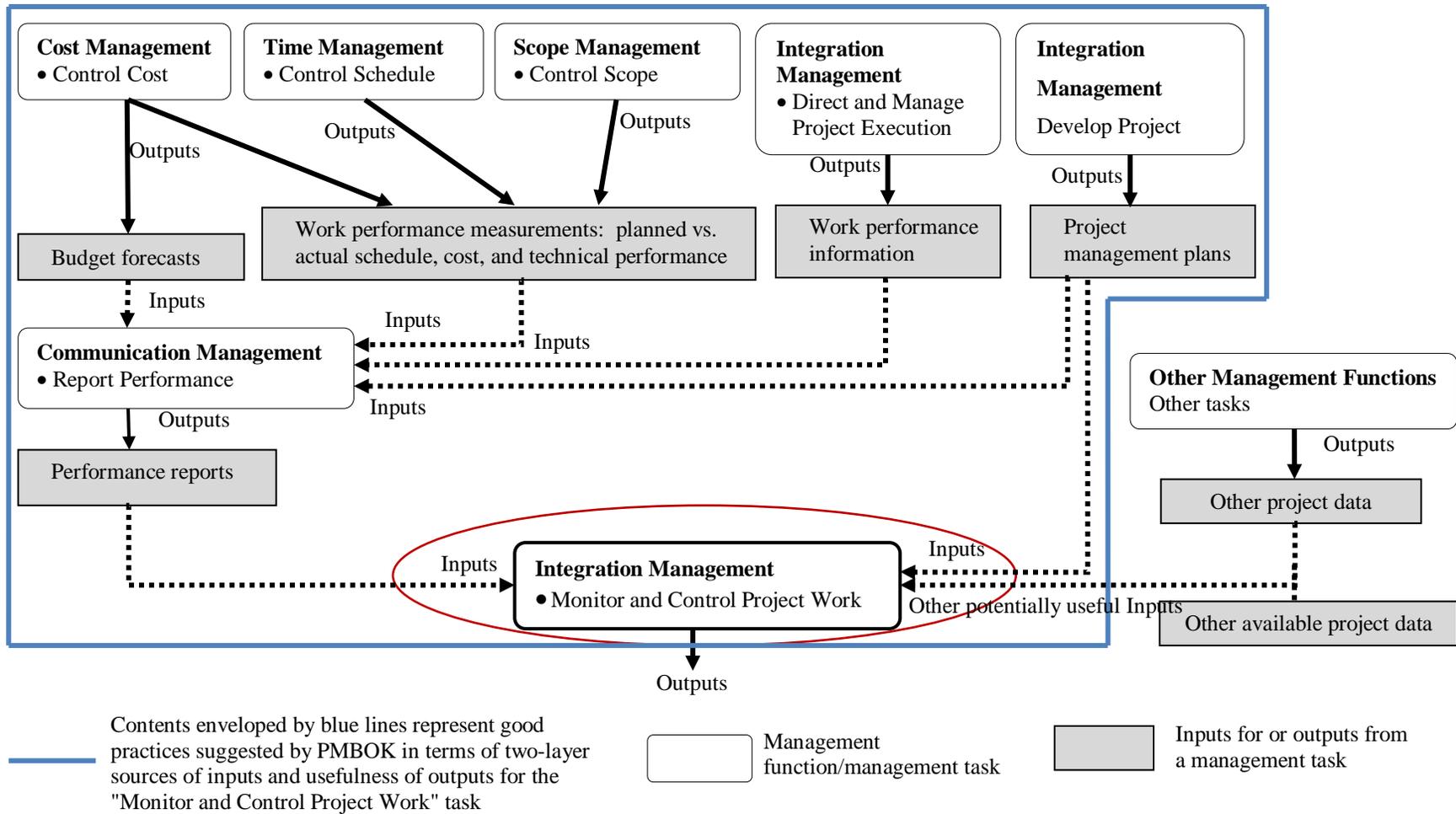


Figure 1.2 Data flows for the task "monitor and control project work" as suggested by PMBOK (2008)

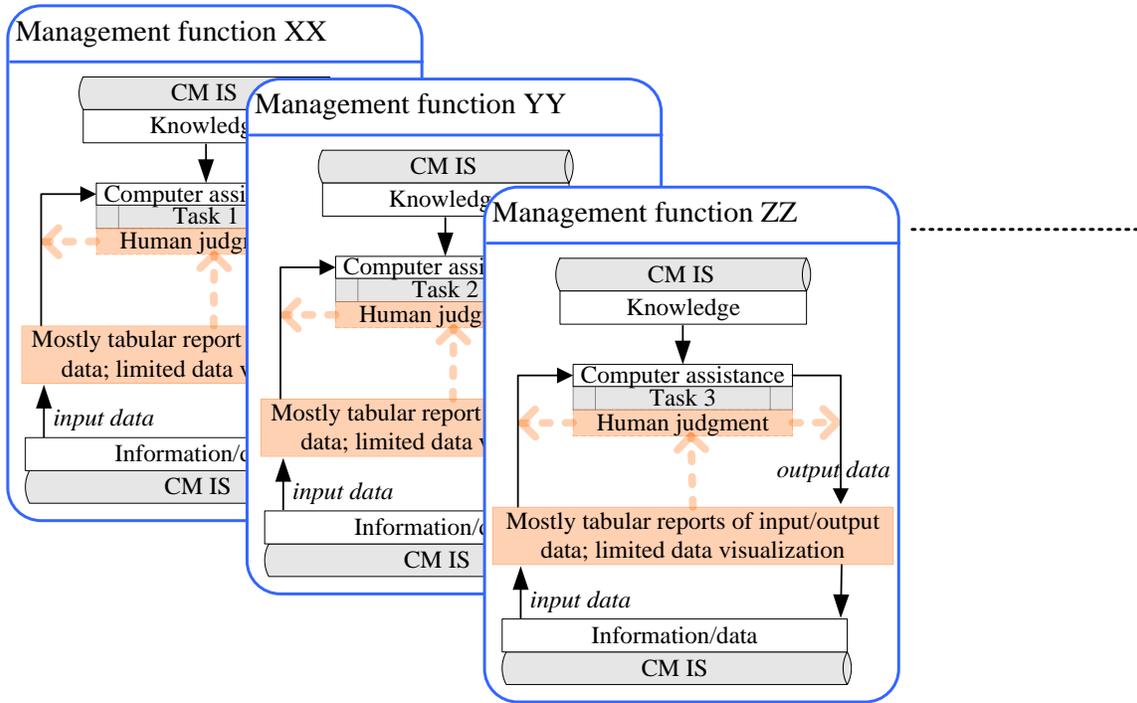


Figure 1.3 The relationship between several current state-of-the-art CM\_IS, computer assistance, human judgment, and conducting stand alone CM tasks

hidden in input and output data that indicate potential effects (impacts) on and/or causes of CM variables. These identified potential effects and/or causes may be relevant to the CM tasks at hand and therefore actions to address the effects (i.e. correction) and/or causes (i.e. prevention) can be taken for conducting/refining the execution of CM tasks. Examples of the different kinds of input/output data and associated management tasks can be observed in the shaded rectangles and round-edged rectangles in Figures 1.1 and 1.2. The relationship between several individual current state-of-the-art CM\_ISs, computer assistance, human judgment, and conducting CM tasks is illustrated in Figure 1.3. Both computer assistance and human judgment complement each other by addressing weaknesses in its counterpart. For example, fast generation of derived data that requires human inspection can be leveraged by computer assistance. Inspecting the characteristics of CM variables that are relevant to a CM task can remedy shortcomings of predefined knowledge that of necessity oversimplifies the complexity that accompanies reality by considering only a limited number of variables.

In the past, as compared to the focus on developing a CM\_IS that enhance its degree of computerization and its role in automatically generating outputs, the development effort for presenting input/output data in support of human judgment for conducting individual CM tasks and complete CM functions has been relatively limited. As a result, CM users often find it difficult to digest and leverage input/output information because of the sheer volume and high dimensionality of data. Two main issues in current CM\_IS that create this difficulty are elaborated upon below.

1. *Issues of report format:* Currently, data reports, specifically referred to here as reports of metadata along with some descriptive remarks describing project models or contents of source documents that can be generated from the state-of-the-art CM\_IS are mostly tabular or textual, with very few of them being graphical. Figure 1.4 represents one out of almost 60 pages of a CM\_IS-generated data report that encompasses input/output data obtained from CM tasks carried out during the planning and execution phases of an actual project. Data treated include planned/actual schedules and records of problems encountered during execution. In fact, many CM staff simply use spreadsheets or even word documents to generate and store input/output data relevant to CM tasks. Seen in Figures 1.5 and 1.6 are actual data reports of deficiencies and change orders collected using either Word documents or Excel spreadsheets. Such tabular reports can provide details of individual raw data for executing CM tasks (e.g. referencing items in a deficiency list to inspect the rectified work). However, they do not provide overall insights that can be deduced from examination of all or part of a data set, which in turn provides leads as to potential effects (impacts) on and/or causes of CM variables. Therefore, a preliminary research question emerges as:

***Rather than using only tabular/textual reports, how should input/output data sets be presented so that CM users can reason about potential effects and/or causes that matter to the CM tasks at hand?***

(A) AS AGREED CONTRACT SCHEDULE							(B) AS - BUILT							(C) ANALYSIS OF DELAYS CAUSED BY	
NO	ITEM	LEVEL	EARLY		LATE		SCHEDULED		ACTUAL		VARIATION	DURATION	ITEMS OF DELAY & EXTRA COSTS		
			START	FINISH	START	FINISH	START	FINISH	START	FINISH					
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
18200	Exc./Mid slab	21	FC 646	1 28NOV83 29NOV83	28NOV83 29NOV83	28NOV83 29NOV83	28NOV83 29NOV83	28NOV83 29NOV83	28NOV83 29NOV83	28NOV83 29NOV83	-15	2	1	- ACCESS TO (ROW, ROAD CLOSURE, 15TH AVE) WAS NEVER MADE AVAILABLE AS PER APPROVED TRAFFIC PLAN ACCESS TO (ROW, PROPERTY ACQUISITION) BLDG. REMOVAL BY OTHERS NOT COMPLETED UNTIL DEC. 15/83	
18200	Exc./Mid slab	22	EC 677	1 28NOV83 29NOV83	28NOV83 29NOV83	28NOV83 29NOV83	28NOV83 29NOV83	28NOV83 29NOV83	28NOV83 29NOV83	28NOV83 29NOV83	-1	2	3	- UTILITY CONFLICT (10% UTILITY) GAS LINE WAS NOT REMOVED BY OTHERS -- (10% UTILITY) WATER LINE WAS BROKEN	
18200	Exc./Mid slab	23	FC 674	1 28NOV83 1 DEC83	1 DEC83 1 DEC83	1 DEC83 1 DEC83	1 DEC83 1 DEC83	1 DEC83 1 DEC83	1 DEC83 1 DEC83	1 DEC83 1 DEC83	3	2	1	- WEEKEND WORK SAT., OCT. 29/83 UTILITY CONFLICT (10% UTILITY) GAS LINE WAS NOT REMOVED BY OTHERS	
18200	Exc./Mid slab	24	EC 671	1 30NOV83 1 DEC83	30NOV83 1 DEC83	30NOV83 1 DEC83	30NOV83 1 DEC83	19DEC83 19DEC83	19DEC83 19DEC83	19DEC83 19DEC83	-12	2	1	- ACCESS TO (ROW, ROAD CLOSURE, 14TH AVE) WAS NEVER MADE AVAILABLE AS PER APPROVED TRAFFIC PLAN	
18200	Exc./Mid slab	25	FC 668	1 2 DEC83 5 DEC83	2 DEC83 5 DEC83	2 DEC83 5 DEC83	2 DEC83 5 DEC83	31JAN84 31JAN84	31JAN84 31JAN84	31JAN84 31JAN84	-20	2	1	- ACCESS TO (ROW, WORK AREA) WAS REDUCED TO ALLOW PLUMBING BUSINESS ACCESS TO THEIR SHOP -- RESEQUENCING OF WORK UTILITY CONFLICT (10% UTILITY) GAS LINE WAS NOT REMOVED BY OTHERS	
18200	Exc./Mid slab	26	FC 664	1 2 DEC83 5 DEC83	2 DEC83 5 DEC83	2 DEC83 5 DEC83	2 DEC83 5 DEC83	1 DEC83 1 DEC83	1 DEC83 1 DEC83	1 DEC83 1 DEC83	7	2	1	- ACCESS TO (ROW, WORK AREA) WAS REDUCED TO ALLOW PLUMBING BUSINESS ACCESS TO THEIR SHOP -- RESEQUENCING OF WORK	
18200	Exc./Mid slab	27	EC 661	1 6 DEC83 7 DEC83	6 DEC83 7 DEC83	6 DEC83 7 DEC83	6 DEC83 7 DEC83	19DEC83 19DEC83	19DEC83 19DEC83	19DEC83 19DEC83	-8	2	1	- ACCESS TO (ROW, ROAD CLOSURE, 13TH AVE) WAS NEVER MADE AVAILABLE AS PER APPROVED TRAFFIC PLAN	
18200	Exc./Mid slab	28	FC 659	1 6 DEC83 7 DEC83	6 DEC83 7 DEC83	6 DEC83 7 DEC83	6 DEC83 7 DEC83	6 DEC83 7 DEC83	6 DEC83 7 DEC83	6 DEC83 7 DEC83	4	2	2	-	
18200	Exc./Mid slab	29	EC 656	1 8 DEC83 9 DEC83	10DEC83 10DEC83	8 DEC83 9 DEC83	8 DEC83 9 DEC83	2 DEC83 5 DEC83	2 DEC83 5 DEC83	2 DEC83 5 DEC83	6	2	2	-	
18200	Exc./Mid slab	30	FC 654	1 8 DEC83 9 DEC83	10DEC83 10DEC83	8 DEC83 9 DEC83	8 DEC83 9 DEC83	7 FEB84 8 FEB84	8 FEB84 8 FEB84	8 FEB84 8 FEB84	-20	2	2	- SURFACE CHANGE (OF FOUNDATION) REQUIRED DUE TO ROCK EXCAVATION -- CHECKS WORKING ON VICTORIA DIVERSION - REMOVAL RESEQUENCED UTILITY CONFLICT (10% UTILITY) 4" GAS LINE NOT REMOVED BY OTHERS	
18200	F Figs./R FAC	1	FC 737	1 3 NOV83 8 NOV83	3 NOV83 8 NOV83	3 NOV83 8 NOV83	3 NOV83 8 NOV83	31JAN84 31JAN84	31JAN84 31JAN84	31JAN84 31JAN84	-51	4	1	- ACCESS TO (ROW, VICTORIA DIVERSION CLOSURE) WAS NEVER MADE AVAILABLE UTILITY CONFLICT (0.8, WIRES # GARDENWAY) WERE NEVER REMOVED BY OTHERS	
18200	F Figs./R FAC	2	EC 734	1 3 NOV83 8 NOV83	3 NOV83 8 NOV83	3 NOV83 8 NOV83	3 NOV83 8 NOV83	18JAN84 18JAN84	18JAN84 18JAN84	18JAN84 18JAN84	-58	4	1	- ACCESS TO (ROW, VICTORIA DIVERSION CLOSURE) WAS NEVER MADE AVAILABLE UTILITY CONFLICT (0.8, WIRES # GARDENWAY) WERE NEVER REMOVED BY OTHERS	
18200	F Figs./R FAC	3	FC 732	1 3 NOV83 8 NOV83	3 NOV83 8 NOV83	3 NOV83 8 NOV83	3 NOV83 8 NOV83	7 FEB84 7 FEB84	7 FEB84 7 FEB84	7 FEB84 7 FEB84	-58	4	1	- ACCESS TO (ROW, VICTORIA DIVERSION CLOSURE) WAS NEVER MADE AVAILABLE UTILITY CONFLICT (0.8, WIRES # GARDENWAY) WERE NEVER REMOVED BY OTHERS	
18200	F Figs./R FAC	4	FC 729	1 3 NOV83 8 NOV83	3 NOV83 8 NOV83	3 NOV83 8 NOV83	3 NOV83 8 NOV83	11JAN84 11JAN84	11JAN84 11JAN84	11JAN84 11JAN84	-29	4	1	- ACCESS TO (ROW, VICTORIA DIVERSION CLOSURE) WAS NEVER MADE AVAILABLE UTILITY CONFLICT (0.8, WIRES # GARDENWAY) WERE NEVER REMOVED BY OTHERS	
18200	F Figs./R FAC	5	EC 727	1 9 NOV83 15NOV83	9 NOV83 15NOV83	9 NOV83 15NOV83	9 NOV83 15NOV83	12JAN84 12JAN84	12JAN84 12JAN84	12JAN84 12JAN84	-36	4	1	- ACCESS TO (ROW, VICTORIA DIVERSION CLOSURE) WAS NEVER MADE AVAILABLE UTILITY CONFLICT (0.8, WIRES # GARDENWAY) WERE NEVER REMOVED BY OTHERS	
18200	F Figs./R FAC	6	FC 724	1 9 NOV83 15NOV83	9 NOV83 15NOV83	9 NOV83 15NOV83	9 NOV83 15NOV83	11JAN84 11JAN84	11JAN84 11JAN84	11JAN84 11JAN84	-37	4	1	- ACCESS TO (ROW, VICTORIA DIVERSION CLOSURE) WAS NEVER MADE AVAILABLE UTILITY CONFLICT (0.8, WIRES # GARDENWAY) WERE NEVER REMOVED BY OTHERS	
18200	F Figs./R FAC	7	EC 721	1 9 NOV83 15NOV83	9 NOV83 15NOV83	9 NOV83 15NOV83	9 NOV83 15NOV83	17JAN84 17JAN84	17JAN84 17JAN84	17JAN84 17JAN84	-29	4	1	- ACCESS TO (ROW, VICTORIA DIVERSION CLOSURE) WAS NEVER MADE AVAILABLE UTILITY CONFLICT (0.8, WIRES # GARDENWAY) WERE NEVER REMOVED BY OTHERS	
18200	F Figs./R FAC	8	FC 718	1 9 NOV83 15NOV83	9 NOV83 15NOV83	9 NOV83 15NOV83	9 NOV83 15NOV83	18JAN84 18JAN84	18JAN84 18JAN84	18JAN84 18JAN84	-48	4	1	- ACCESS TO (ROW, VICTORIA DIVERSION CLOSURE) WAS NEVER MADE AVAILABLE UTILITY CONFLICT (0.8, WIRES # GARDENWAY) WERE NEVER REMOVED BY OTHERS	
18200	F Figs./R FAC	9	EC 715	1 16NOV83 21NOV83	16NOV83 21NOV83	16NOV83 21NOV83	16NOV83 21NOV83	15JAN84 15JAN84	15JAN84 15JAN84	15JAN84 15JAN84	-27	4	1	- ACCESS TO (ROW, VICTORIA DIVERSION CLOSURE) WAS NEVER MADE AVAILABLE UTILITY CONFLICT (0.8, WIRES # GARDENWAY) WERE NEVER REMOVED BY OTHERS	
18200	F Figs./R FAC	14	FC 712	1 16NOV83 21NOV83	16NOV83 21NOV83	16NOV83 21NOV83	16NOV83 21NOV83	21JAN84 21JAN84	21JAN84 21JAN84	21JAN84 21JAN84	-48	4	1	- ACCESS TO (ROW, VICTORIA DIVERSION CLOSURE) WAS NEVER MADE AVAILABLE UTILITY CONFLICT (0.8, WIRES # GARDENWAY) WERE NEVER REMOVED BY OTHERS	
18200	F Figs./R FAC	11	EC 709	1 16NOV83 21NOV83	16NOV83 21NOV83	16NOV83 21NOV83	16NOV83 21NOV83	25NOV83 25NOV83	25NOV83 25NOV83	25NOV83 25NOV83	-5	4	2	- ACCESS TO (ROW, VICTORIA DIVERSION CLOSURE) WAS NEVER MADE AVAILABLE -- WEEKEND WORK SAT JAN 04/84 UTILITY CONFLICT (0.8, WIRES # GARDENWAY) WERE NEVER REMOVED BY OTHERS	
18200	F Figs./R FAC	12	FC 707	1 16NOV83 21NOV83	16NOV83 21NOV83	16NOV83 21NOV83	16NOV83 21NOV83	18NOV83 18NOV83	18NOV83 18NOV83	18NOV83 18NOV83	5	4	2	- ACCESS TO (ROW, VICTORIA DIVERSION CLOSURE) WAS NEVER MADE AVAILABLE -- DESIGN CHANGE (OF FOUNDATION) REQUIRED DUE TO POOR SOILS.	
18200	F Figs./R FAC	13	EC 705	1 22NOV83 29NOV83	22NOV83 29NOV83	22NOV83 29NOV83	22NOV83 29NOV83	14NOV83 14NOV83	14NOV83 14NOV83	14NOV83 14NOV83	6	4	4	- ACCESS TO (ROW, VICTORIA DIVERSION CLOSURE) WAS NEVER MADE AVAILABLE UTILITY CONFLICT (0.8, WIRES # GARDENWAY) WERE NEVER REMOVED BY OTHERS	
18200	F Figs./R FAC	14	FC 702	1 22NOV83 29NOV83	22NOV83 29NOV83	22NOV83 29NOV83	22NOV83 29NOV83	14NOV83 14NOV83	14NOV83 14NOV83	14NOV83 14NOV83	6	4	4	- DESIGN CHANGE (OF FOUNDATION) REQUIRED DUE TO POOR SOILS -- +/-2 METRES DEEPER EXCAVATION REQUIRED	
18200	F Figs./R FAC	15	EC 699	1 22NOV83 29NOV83	22NOV83 29NOV83	22NOV83 29NOV83	22NOV83 29NOV83	16NOV83 16NOV83	16NOV83 16NOV83	16NOV83 16NOV83	5	4	3	-	
18200	F Figs./R FAC	16	FC 696	1 22NOV83 29NOV83	22NOV83 29NOV83	22NOV83 29NOV83	22NOV83 29NOV83	21NOV83 21NOV83	21NOV83 21NOV83	21NOV83 21NOV83	2	4	3	- DESIGN CHANGE (OF FOUNDATION) REQUIRED DUE TO POOR SOILS.	
18200	F Figs./R FAC	17	EC 693	1 29NOV83 1 DEC83	29NOV83 1 DEC83	29NOV83 1 DEC83	29NOV83 1 DEC83	29NOV83 29NOV83	29NOV83 29NOV83	29NOV83 29NOV83	5	4	3	-	
18200	F Figs./R FAC	18	FC 690	1 29NOV83 1 DEC83	29NOV83 1 DEC83	29NOV83 1 DEC83	29NOV83 1 DEC83	24NOV83 24NOV83	24NOV83 24NOV83	24NOV83 24NOV83	4	4	2	- ACCESS TO (ROW, ROAD CLOSURE, 16TH AVE) WAS NEVER MADE AVAILABLE AS PER APPROVED TRAFFIC PLAN	
18200	F Figs./R FAC	19	EC 687	1 29NOV83 1 DEC83	29NOV83 1 DEC83	29NOV83 1 DEC83	29NOV83 1 DEC83	4 JAN84 5 JAN84	5 JAN84 5 JAN84	5 JAN84 5 JAN84	-19	4	2	- ACCESS TO (ROW, ROAD CLOSURE, 16TH AVE) WAS NEVER MADE AVAILABLE AS PER APPROVED TRAFFIC PLAN ACCESS TO (ROW, WORK AREA) WAS USED FOR REBAR FABRICATION DUE TO VICTORIA DIVERSION ACCESS PROBLEMS	

Figure 1.4 One of many pages of CM-IS generated tabular reports of planned/actual schedule data and problems encountered during execution

Building: [REDACTED]  
 Suite: 103

Inspection Report

Inspection Dates	
1st	April 30, 2009
2nd	June 12, 2009
3rd	June 17, 2009
4th	June 25, 2009
5th	July 14, 2009
6th	

Location	Description	Trade	✓
Inspection Date	June 12, 2009		
Inspection Date	June 17, 2009		
Inspection Date	June 25, 2009		
Inspection Date	July 14, 2009		
<b>Completed</b>			
2nd Bedroom	window frame - chipped - front edge bottom L	♦W	✓
	door hinge - has paint on it - top mid bottom at closet		✓
	carpet - damaged - at L closet jamb - possible screw in under lay		✓
	wall corner - fill sand paint - mid R of closet		✓
	carpet - loose - at door		✓
	small space by casing - fill sand paint - mid L inside		✓
	baseboard corner - fill sand paint - R of closet		✓
	closet door - chipped - L side inside bottom L		✓
	jamb - fill sand paint - top R		✓
	label all switched outlets		✓
	wall - fill sand paint - below tel		✓
	baseboard - sand and paint - top edge L of tel		✓
	baseboard - has paint on it - south mid		✓
	baseboard heater - has glue on it		✓
	window sill - redo caulking and paint - at L side wall corner		✓
	window - has paint drips - bottom R		✓
	wall - fill sand paint - at cable		✓
	carpet - damaged - at corner R of closet		—
Bathroom	wall tiles - need cleaning at top course to ceiling		✓
	ceiling at wall - improve cut line - over sink & toilet		✓
	wall edge - paint caulking - top R of door casing inside		✓
	casing - scratched - mid R outside		✓
	casing - has paint drips - bottom R outside		✓
	casing - redo caulking and paint - top L + R at jamb		✓
	jamb - sand and paint - bottom L		✓
	jamb - fill sand paint - bottom L		✓
	grout - missing - near floor south in shower		✓
	drawer front - incomplete - bottom mid top		✓
	sprinkler - has paint on it		✓
	ceiling - fill sand paint - at pot light in shower		✓
	baseboard heater - not working		✓
	bathroom exhaust fan - too loud		—
Den	French doors - has paint drips - at glass all over		✓
	French doors - paint touch-up - at mid hinge west door inside & bottom hinge inside east door		✓
	casing - fill sand paint - mid L outside		✓
	French doors - chipped - L side at bottom R inside		✓
	casing - fill sand paint - L below bottom hinge		✓
	jamb - redo caulking and paint - L side over mid hinge		✓
	door hinge - scratched - top L		✓
Entry	jamb - paint - at mid hinge		✓
	small space by casing - paint - mid R		✓
	entry door closer - needs adjusting		✓
	wall - paint touch-up - mid L of den casing		✓

[REDACTED] Inspection - Suite 103.doc

Figure 1.5 One of many pages of Microsoft Word documents of deficiency lists

Base Building Change Order Registry - Division 17							
C.O. No.	Reference No.		Date Issued	Projected Cost	Date Approved	Approved Amount	Trade Change Orders Issued
17001	-	Extra asbestos abatement in west wing and tower not identified in PHH's report	11-Feb-04	132,890.00	Feb 5/04	132,890.00	
17002	ASI #2 ?	Pin coping stones	29-Mar-04	4,900.00		4,900.00	
17003	-	Sawcut & remove concrete window sills	07-May-04	17,400.00		17,400.00	
17004	-	Removal of level 3 mezzanine - east wing	31-May-04	24,800.00		24,800.00	
17005	-	Remove additional concrete walls and granite block - 5th floor	25-Mar-04	7,700.00	Aug 13/04	7,700.00	Abba, Rev02, \$7,700.00
17006	-	Remove & dispose of 2 chimney stacks - west wall	25-Mar-04	1,800.00	Aug 13/04	1,800.00	Abba, Rev03, \$1,800.00
17007	-	Remove & dispose of concrete stairs in basement - not on drawings	25-Mar-04	250.00	Aug 13/04	250.00	Abba, Rev04, \$250.00
17008	-	Remove & dispose of chimney stack on main, 2nd, 3rd & 4th floors - GL's Q & R	25-Mar-04	750.00	Aug 13/04	750.00	Abba, Rev05, \$750.00
17009	-	Remove & dispose of additional materials	15-Jun-04	5,941.00	Aug 13/04	5,941.00	Abba, Rev02, \$5,941.00
17010	-	Additional demolition work - Backcharge McLeod Masonry	15-Jun-04	3,810.00	Aug 13/04	3,810.00	Abba, Rev03, \$3,810.00
17011	-	Extra shoring on main, 2nd & 3rd floors - JKK fix	15-Jun-04	360.00	Aug 13/04	360.00	Abba, Rev04, \$360.00
17012	ASI #2	Lower 2 window sills at belfry	13-Apr-04	20,800.00	Aug 13/04	13,000.00	McLeod, Rev09, \$13,000.00
17013	-	Remove & catalogue stone entrance at new link	21-Jun-04	3,890.00	Aug 13/04	3,890.00	McLeod, Rev10, \$3,890.00
17014	-	Revised Structural Steel from " Post Tender Addenda " drawings, dated January 16, 2004, to " IFC ", dated March 12, 2004	21-Jun-04	32,396.00	Dec 7/04	32,396.00	George, Rev02, \$32,396.00
17015	SSK-1	Remedial steel channels for cut joist on 2nd floor west wing	21-Jun-04	1,740.00	Dec 7/04	1,740.00	George, Rev03, \$1,740.00
17016	SSK-10, RFI # 26	Change direction of stair # 5 and redetail - West wing	21-Jun-04	871.00	Dec 7/04	871.00	George, Rev04, \$871.00
17017	SI # 84R	West Roof Replacement & SSI # 8 & MSI # 26	Nov 19/04	270,000.00	Dec 7/04	200,000.00	Lower Mainland, Rev02, \$6,303.50/Western, Rev02, \$18,440.34/Marine, Rev01, \$67,565.00/Western, Rev03, \$25,957.81/Western, Rev05, \$1,084.00/Shanahan's, Rev25, \$642.00
17018	-	Storm Connection		97,661.00	Dec 7/04	33,392.35	
17019	-	Sanitary Connection		23,913.00	Dec 7/04	23,913.00	
17020	-	New embeds. Revisions for seismic tubes for Tower windows		1,948.00	Dec 7/04	1,948.00	George, Rev06, \$1,948.00
17021	-	Remedial steel plates & anchors West wing		3,024.00	Dec 7/04	3,024.00	George, Rev07, \$3,024.00
17022	ASI #8	Stair 10 struct revision	May 31/04	15,385.00	Dec 7/04	15,385.00	George, Rev08, \$15,385.00
17023	ASI #13	Struct SI #2	Jul 6/04	43,282.00	Dec 7/04	43,282.00	George, Rev09, \$43,282.00

Figure 1.6 One of many pages of Microsoft Excel spreadsheets for change order data

2. *Issues of the need to continuously refine the search of several reports before generating the "right" reports:* Another issue relates to the need of CM\_IS users to go through many data reports, which may come from:

- Inputs prepared for and outputs generated from computer assisted automation or any relevant required inputs for the task dictated by explicit knowledge (e.g. shaded rectangles enveloped by the blue lines in Figures 1.1 and 1.2), or
- Outputs from other tasks that are deemed by the implicit knowledge of CM\_IS users, but not expressed in terms of explicit knowledge, as relevant inputs to the tasks at hand (e.g. shaded rectangles outside the blue envelopes in Figures 1.1 and 1.2).

The processes involved in going through many reports on demand helps CM\_IS users to pose and answer questions continuously until they are satisfied with their analytical reasoning. These processes may include iterations of: 1) looking for and selecting relevant reports and data contents, and 2) adjusting presentation formats of the reports to ensure readability. However, current CM\_ISs provide limited interactive features with which users can readily conduct the foregoing processes. Also, while some prototypes of integrated CM\_IS have been developed, most commercial CM\_ISs only support specific CM tasks (e.g. Microsoft Project mainly supports CM tasks of time management functions both in the planning and monitoring phases). This creates a

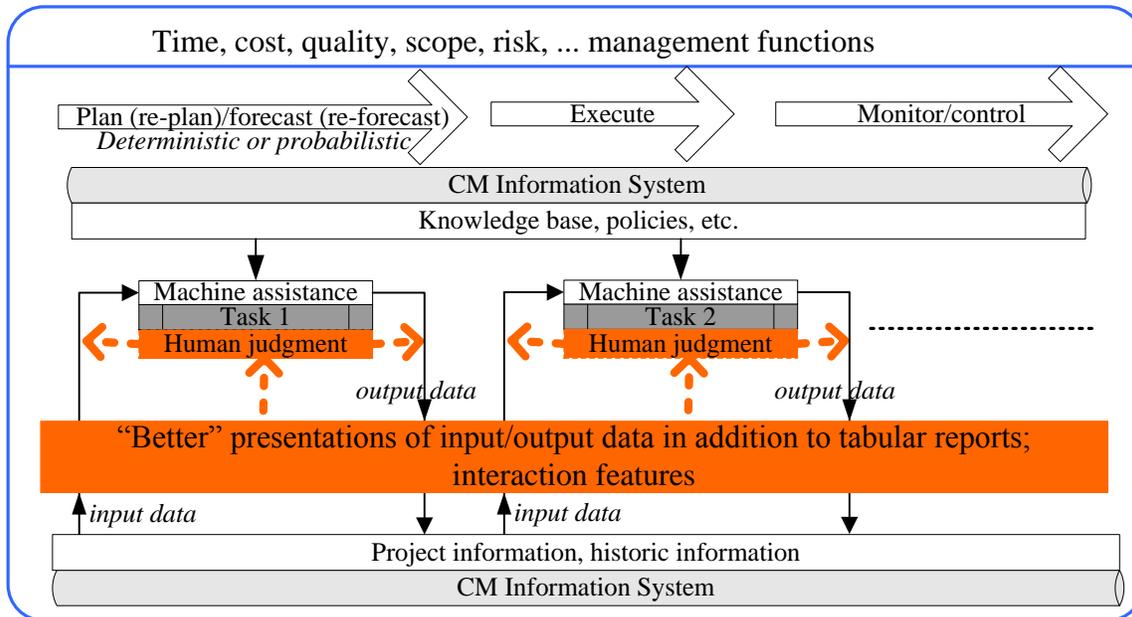


Figure 1.7 A proposed solution in terms of enhancing data reporting functionalities in the relationship between a current state-of-the-art CM\_IS, computer assistance, human judgment, and conducting stand alone CM tasks

difficulty for CM users to seamlessly seek and view a variety of relevant data reports.

Thus, another preliminary research question is:

***How can CM users explore project data sets in order to generate the "right" data reports in visual form for assisting their analytical reasoning processes on an ongoing basis?***

To address the foregoing issues, my preliminary idea for a solution is to improve the data reporting capability of a CM\_IS system which focuses mainly on using tabular/textual reports. The proposed enhancement is conceptualized in Figure 1.7. As seen in this figure, the layer (the orange rectangle) between an integrated CM\_IS and CM processes represents an enhanced "data reporting center" that adds the following two features (an enhanced data reporting center will be referred to hereinafter solely to mean a data reporting center that has these two features but excluding tabular data reporting) that are lacking in current CM\_ISs and their data reporting features as depicted in Figure 1.3:

1. In addition to tabular data reports, input/output data from a diverse range of CM tasks can also be viewed using "new" data presentation formats that can deliver messages

hidden in datasets that are not readily discernable using traditional tabular/textual reports; and

2. These "new" presentations of input/output data should be available "through a single shop or interface" with sufficient interactive features to allow users to flexibly select and adjust the presentations.

The goal of this thesis is to demonstrate how the foregoing can be achieved through the use of data visualization. Data visualization is defined as "the use of computer-based, interactive visual representations of data to amplify cognition" (Card et al. 1999). Issues related to it have been research topics in other knowledge domains including computer science, statistics, and psychology. Suitable visual representations of data can be designed and created by transforming original raw datasets into desired structured data representations and then representing data dimensions by effective visual variables (e.g. spatial positions, colors). For at least three reasons, visual representations of data help derive overall information/insights that tabular/textual/audio/video raw data cannot otherwise provide:

1. For unstructured raw data, important topics and their contexts are parsed into discrete data dimensions (i.e. variables) and values corresponding to them, which in turn provide a focused and structured CM variable data space for investigation. Thus, unstructured raw data is turned into tabular metadata. For example, one of the important topics of construction meeting minutes relates to locations of troubled activities and therefore specific locations and activities mentioned in each meeting are documented. Thus, the minutes can be structured into at least a two dimensional data table with the dimensions being "location" and "activity" and values of the dimension being codes or names of locations/activities mentioned in the meeting minutes. For the foregoing transformed tabular metadata, new data dimensions such as "counts of meeting minutes" can be derived, leading to information as to the number of meeting minutes related to a certain activity and/or location. It is the aforementioned data transformation of original raw data and the derivation of data tables that provides users with a focus for examining certain variables that are most pertinent to their CM analytic tasks. The "data" in the "visual representation of data" usually refers to the

foregoing transformed tabular metadata along with the additional data derived from the metadata.

2. Features such as spatial positions or colors provide low similarity amongst different features as compared to texts or numbers which is one of the key reasons why human beings can be visually attentive to certain symbols (Duncan and Humphreys 1989) and identify visual patterns. These patterns indicate important messages/insights hidden in a collection of data, prior to conscious attention (Ware 2004).
3. Large amounts of visual/diagrammatic information can be processed by the human visual perception system in parallel as opposed to the serial processing required for textual or numeric information (Larkin and Simon 1987; Ware 2004).

In addition to the foregoing merits of visual representations of data, data visualization also enables interaction with visual representations of data thereby allowing users to continuously formulate the data content and format of the visual representations during their insight generation exploration processes. This facility provides users with a continuously updated information platform and aids the decision making process. Another concept is visual analytics, which is a new data analysis paradigm. It is defined as "the use of visual representations of data and interactions to accelerate rapid insights into data" (Thomas and Cook 2005). The purpose of visual analytics is similar to the needs of the CM community in terms of exercising human judgment when conducting CM tasks through interacting with an enhanced "data reporting center".

Use is made in this thesis of terminologies employed by both the visualization and construction management communities combined with the concepts presented in the first part of this chapter. To ensure a clear understanding of the vocabulary used, definitions of terms used are presented in the next section and adhered to throughout this document.

## **1.2 Terminology definitions**

1. ***CM data***: the input/output data generated in support of CM tasks.
2. ***Visual representation of CM data (or images of CM data or simply images)***: presentations of CM data that are either in their natural form (if the CM data

represents tangible objects such as a building) or visual form that map data dimensions to visual variables (if the CM data represents non-tangible abstractions). Since this research focuses on abstract CM data, the latter definition is adopted herein.

3. ***Pre-coded image***: an image whose default specifications (e.g. how data dimensions are mapped to visual variables, levels of detail of data to be presented) have been pre-defined in a data visualization system. The system has user interfaces for showing selections of pre-coded images, which can be in the form of menu items, check boxes, or small icons. If a user chooses one of them, the visualization system will generate the image selected based on the pre-defined specification of that image.
4. ***Thematic images***: images can be categorized by certain themes. For example, various images that portray values of different attributes of a product can be categorized as product thematic images.
5. ***Interaction features***: content and display controls that allow users to interact with images such as querying data ranges and then updating the images according to the query.
6. ***Interactivity***: depending on the context in which this term is used, "interactivity" is loosely used to describe a feature in particular or a capability in general of interacting with images. It is a term commonly used in the computer science literature.
7. ***CM data visualization environment***: an "enhanced CM data reporting center" that is created based on data visualization concepts and technologies.
8. ***Construction conditions***: an umbrella concept covering construction strategies imposed, construction requirements dictated in contracts, construction constraints encountered, and so forth.
9. ***Construction performance measures***: time, cost, scope, quality, safety, and risk.
10. ***CM analytics***: CM user focused analytical reasoning for identifying potential effects and/or causes from the characteristics of CM variables related to construction conditions and construction performance measures that pertain to the CM task/function at hand.
11. ***Visual CM analytics***: the use of a CM data visualization environment for conducting CM analytics.
12. ***Analytic reasoning artifacts***: this term was defined in (Thomas and Cook 2005) as

"tangible pieces of information that contribute to reaching defensible judgments" about a question. These artifacts can be elemental ones such as relevant information and evidence, patterns, high-order knowledge constructs, and complex reasoning constructs.

13. ***Top-down development approach:*** A "top-down approach" involves a design process that is focused on identifying CM analytics applicable to one or more CM functions/tasks and visualization requirements responsive to them. The analysis of visualization requirements includes scoping what CM variables are involved in the CM analytics and identifying general rules of how their characteristics can be observed, on demand, in support of the CM analytics. Determination of the relevant CM variables provides the scope/direction of the visualization development; the analysis of how characteristics of these variables can be observed in general (i.e. in support of CM analytics common to various CM functions/tasks) results in required common visualization features. For example, from an overall CM perspective, the common CM analytics applicable to most CM functions/tasks is to explore potential causes and/or effects amongst construction conditions/performance measures. Using the multiple view modeling of a project, these construction conditions/performance measures can be attributes of the project context dimensions of product, process, project participant, environment, etc. This recognition helps form the scope/direction of CM data visualization environment development, i.e., CM variables in terms of construction conditions/performance measures to be visualized. Also, it is recognized that the characteristics of the aforementioned CM variables in general can be observed by project context dimensions, different levels of detail, and data status. Thus, the common visualization features should support presenting the distribution of values of construction conditions/performance measures in terms of different data status, different context dimensions, and at different levels of detail.

14. ***Bottom-up development approach:*** A "bottom-up approach" deals with the detailing needed in order to create an actual visualization. These detailing processes include:

- Analysis of more specific CM analytics (i.e. CM questions) in relation to CM functions/tasks;
- Identifying specific visualization requirements in terms of what CM variables are

involved for specific CM analytics and how their characteristics can be observed, on user demand, in support of the CM analytics. Analysis of the latter is conducted by first analyzing specifics of the top-down common visualization features that fit the nature of particular CM analytics and/or CM variables. Additional new visualization features may also be sought. Inclusion of interaction features for changing default settings of a visual representation that are as flexible as possible is key to allowing users to decide how to observe the characteristics of CM variables on demand. In the bottom-up design, the focus is on "what cannot" be changed if these changes do not add value to the usability and utility of the visualization;

- Specifying required data representations/transformations, visual encodings and non-visual encoding attributes of a visual representation, and interaction features according to specific visualization requirements;
- Implementing the specifications;
- Evaluating the implemented visualization. Both designers and/or end users can utilize the "inspection evaluation method" (Amar and Stasko 2005; Ardito et al. 2006; Zuk et al. 2006) to identify any deficiency in terms of non-conformance to the requirements/specifications and new requirements/specifications that better help answer the CM questions posed.

The test of operating an implemented visualization should make use of sizeable sets of actual as well as synthetic project data that is representative and reflective of real world projects in order to ensure the visualization tool can handle the realities of construction project data. The function of the bottom-up development process, from the perspective of developing a CM data visualization environment, is that through focusing on details and using the implemented visualization on actual and representative data sets, lessons can be learned for contributing to refining design guidelines and/or top-down common visualization features.

15. ***Environment architecture***: an organization of thematic visualizations, mainly hierarchical from abstract to specific (e.g. time performance visualization vs. milestone finish date monitoring visualization), categorized by construction conditions and performance measures under multiple views of a project. Each

visualization is developed in a consistent way by following design guidelines and addressing common visualization features required for supporting general CM analytics. These include the use of consistent user interfaces and mechanisms for users to adjust them. A direct benefit is enhanced environment learnability.

### **1.3 Research questions, objectives, and hypothesis**

Based on the observed problems, proposed solutions, and terminology described previously, several research questions have been formulated along with attendant research objectives which seek answers to these questions. The questions and research objectives are as follows:

1. How should a CM data visualization environment be developed?
  - What are the status and shortcomings of current data visualizations for effecting visual CM analytics both in commercial CM\_IS software and academic research?
  - Are there concepts, theories, processes, and technologies from state-of-the art data/information visualization that can be adopted or adapted to developing a CM data visualization environment?
  - What methodology should be used to adopt or adapt state-of-the-art data visualization in order to develop/enhance a CM data visualization environment, which addresses shortcomings of the current use of data visualization in CM?
  
2. What are the key features of a CM data visualization environment that best reflect the functions expected for it?
  - What are the key features of images for assisting CM analytics?
  - What are the key features of interactivity and environment architecture that allow users to flexibly explore a variety of images useful for CM analytics?
  
3. How does the use of a CM data visualization environment help conduct CM analytics that cannot be done or which are difficult to do with current data reporting practices?

The formulation of and seeking answers to the third question along with research assumption 3 (see next section) form the research hypothesis as:

*The use of an appropriately developed CM data visualization environment helps CM personnel interpret CM data for assessing, learning, and communicating causes and/or effects amongst a wide range of CM variables of a construction project thereby improving the quality of CM processes and decision-making.*

#### **1.4 Research scope and assumptions**

In order to explore meaningful answers to the previously formulated research questions, it is essential to have a clear and focused scope of work along with supporting assumptions that are appropriate for a Ph.D. thesis. This is essential in order to understand the context (assumptions/conditions) for which the answers sought are valid. Research assumptions and scope of work that underlie the thesis are as follows:

##### **1.4.1 Scope or focus**

1. The focus of this research is on understanding how to visualize a collection of abstract CM data visualizations for identifying potential causes by and/or effects on CM variables. This "abstract CM data" refers to structured metadata describing important topics/aspects of textual descriptions, project documentation (e.g. meeting minutes, videos, photos), and project models (e.g. built products, processes). Visualization of their actual contents (e.g. actual text in documents, appearance and geometry of products) is excluded from this research.
2. To the greatest extent possible, limited yet sufficiently diverse and sizeable sets of CM data originating from different CM tasks/functions (as-built records of change order data, deficiency data, schedule data, and product/location attributes data) are used to deduce general knowledge/insights surrounding the topic of utilizing CM data visualization to apply CM analytics to CM processes.
3. The use of CM data visualization is a complement to CM data analysis that uses a computational approach. In fact, both the data visualization approach and computational data analysis approach can leverage each other to achieve greater utility. Given the focus of my work on data visualization, a computational data analysis approach is not used as a comparison benchmark – i.e. the research question (hypothesis) of interest does not relate to the superiority or inferiority of computational

data analysis versus data visualization, but to the question of what and when benefits can be derived through visual analytics.

4. Inquiring into industry end user performance differences arising from the deployment of data visualization tools and accompanying visual analytics paradigm in a day-to-day work environment is excluded from this research. The purpose of developing a prototype of a CM data visualization environment in this research is to provide a test bench for generating and analyzing insights with regard to the design and use of such an environment. This does pose some challenges however if third party evaluation is sought of the ideas developed, as the focus can and most often move to the full set of features and robustness one finds in commercial software rather than the concepts involved. It is believed that the insights generated through implementation of a prototype will contribute to future research work directed at extending the visualization and usability capabilities of the current prototype as well as those developed by other researchers to facilitate conducting full scale observations on applying data visualization and visual analytics to CM functions and eventually to the development of commercial software that has the features and robustness required to effect adoption by industry.
5. The main focus of this research in response to the research question of "How should a CM data visualization environment be developed" is on how to develop a **CM analytics-oriented one** that helps identify potential causes and/or effects from a collection of data representing a variety of CM variables. Limited visualization capabilities found in the literature and commercial software, along with commonly used tabular/textual data reports are mostly task-oriented for supporting execution of narrowly focused CM tasks (e.g. read project scope statements and create work breakdown structure according to statements; read activities in the upcoming days in a schedule for executing construction work, etc). The kind of visualization features my research is exploring have largely been unaddressed heretofore. Therefore, when evaluating an implemented new visualization the focus is on whether it serves the CM analytics for which it is implemented. In most cases, alternative visualizations in support of the same analytic task do not exist, precluding comparisons and evaluations as to superiority.

6. The author recognizes that in pursuing answers to the research questions, the front end part of the development processes involves analysis of CM analytic needs and the corresponding visualization requirements (identification of CM analytic themes and important CM variables or data dimensions associated with these themes). This front end part is knowledge domain specific. A later stage of the development process involves turning requirements into visualization specifications (e.g. map multi-data dimensions to visual variables considering choices of interaction features) and implementing the specifications, which is not knowledge domain specific and remains as a difficult design problem in the general data visualization domain. It is a difficult general data visualization problem because there could be too many combinations of visual encodings leveraging on supporting interaction features for mapping the same data. In other words, there could be many alternative specifications that may all meet the visualization requirement (e.g. the need to map ten data dimensions important to a certain analytic need into visual forms) but which have different levels of usability (admittedly usability may have an impact on utility to some extent). Therefore, another focus of this research with regard to answering research question 2 "What are the key features of a developed CM data visualization environment...." is on the front end analysis of CM analytic needs and visualization requirements. "Satisficing" specifications for some visualizations are identified and implemented, but they are only used to demonstrate the completeness of the development process and attention to state-of-the-art data visualization, and provide a test bench for answering research question 3. Identification of the best specification amongst several specification alternatives is believed by the author to be a generic data visualization problem and is excluded as a focus of this research.
7. As part of this research work, various visualizations and supporting infrastructure were implemented to demonstrate the usefulness of these features. However, the focus of this implementation effort was not on developing novel programming approaches in support of the implementation work or the use of new programming languages. Accordingly, the programming work associated with implementation is not claimed as a research contribution. It was assisted by a dedicated programmer who had extensive knowledge of the system in which the visual images are embedded.

### 1.4.2 Assumptions

1. As mentioned in the beginning of this chapter, a CM\_IS is developed to implement explicit knowledge of CM processes. There could be different ways of organizing and recognizing CM knowledge and hence there will be variations in different CM\_IS in terms of system architectures, data structures, and terminologies used. One of the structured ways of organizing CM knowledge is the concept that a construction project can be described and modeled by multiple views (e.g. process view, cost view, quality view) (Russell and Udaipurwala 2004), and each view has its unique data structures, data definitions, and data computation routines for implementing required inputs/outputs and techniques associated with the CM functions/tasks supported. Multiple views are tightly associated by sharing common project context dimensions so that data from multiple views can be more readily used as inputs/outputs for specific CM tasks/functions. The exploration of how to develop a CM data visualization environment in this research makes use of this "multiple views of a project" concept and a research CM\_IS (REPCON (Russell 1985; Russell and Udaipurwala 2004)) that implements this concept. This research CM\_IS provided the platform for exploring the data visualization concepts described in this thesis. It provided important information system infrastructure for allowing full scale implementation and testing of concepts. However, it also provided some constraints with respect to the visualization tool kit that could be used because the visualization application programming interface (API) used needs to be compatible with the architecture of the research CM\_IS. The visualization API used is ChartFX 6.2 Client Server (Software FX Inc.), and the visualization features are somewhat limited to the visualization components this API can offer (e.g. have no control over attributes of the third coordinate).
2. CM data / information needs to be abstracted into structured and electronic data formats that are compatible with the CM information system(s) with which the CM data visualization environment is integrated. While at least in theory data abstraction can be done by computer through data mining techniques, this research will rely on manually pre-processing messy and unstructured CM data into structured data formats. For example, for the case study that focused on deficiency data, considerable effort

was expended by the author to review as-built data that was documented in plain sentences. Sentence contents were parsed into dimensions (e.g. participant, product, process, schedule) that match the data fields of the CM information system, and then the corresponding data was entered into the system.

3. The cause-effect relationship between using a CM data visualization environment and improving CM processes is an indirect two layer cause-effect relationship: using a CM data visualization environment enhances the quality of CM analytics through strengthening human CM analytics ability, which in turn improves the quality of conducting CM functions/tasks and consequently CM processes. The latter relationship is an assumption used in this research, and hence the focus on validating/invalidating the hypothesis that a CM data visualization environment can help enhance CM analytics capabilities.

### **1.5 Research methodologies**

Correspondence between research questions for which answers are sought and research methodologies for searching for them can be found in Table 1.1. The research techniques adopted are observational ones that include literature/state-of-the-art review, structured analysis, and prototype development through case studies. These research techniques were applied in the following sequential research phases:

1. *Phase 1:* To develop guidelines and principles for designing a CM data visualization environment based on experience gained from operating state-of-the-art data visualization software and structured analysis of past literature regarding visual analytics, data visualization, and construction management. A case study using change order (CO) data analysis application and CO data of a complex rehabilitation project was utilized to:
  - Demonstrate the merits of presenting data in visual form;
  - Test the use of identified design guidelines/principles for analyzing visualization requirements and specifications for the change order analysis application; and,
  - Identify lessons that can be learned regarding how to develop a CM data visualization environment and key features required of images and interaction features.

Table 1.1 Correspondence between research questions and research methodologies for answering them

Research Questions (objectives)	Research methodologies
<ul style="list-style-type: none"> <li>• What are the status and shortcomings of the current use of data visualization for effecting visual CM analytics both in commercial CM_IS software and academic research?</li> <li>• Are there concepts, theories, processes, and technologies from state-of-the art data/information visualization that can be adopted or adapted for developing a CM data visualization environment?</li> <li>• What methodology should be used to adopt or adapt state-of-the-art data visualization in order to develop/enhance a CM data visualization environment, which can address shortcomings of the current use of data visualization in CM?</li> </ul>	<ul style="list-style-type: none"> <li>• Literature reviews which focus on the use of either visual representations of data or data visualization to present CM data; review of the state-of-the-art of data visualization for CM_IS with the focus on mainstream commercial representations and an academic one available to me</li> <li>• Literature reviews with the focus on fundamentals of how to develop data visualization tools and prevailing data visualization features/technologies</li> <li>• Analysis and case studies (both application cases and project data cases) for executing the prototype development process in order to obtain lessons learned in terms of how to develop a CM data visualization environment</li> </ul>
<p>What are the key features of a CM data visualization environment that best reflect the functions expected for it?</p>	<p>Analysis and case studies (both application cases and project data cases) for executing the prototype development process in order to obtain lessons learned in terms of deducing key features of the images implemented and their interaction features</p>
<p>How does the use of a CM data visualization environment help conduct CM analytics that cannot be done or which are difficult to do with current data reporting practices?</p>	<p>Case studies, structured comparison evaluation, and analysis for comparing analytical reasoning capability between the use of a CM data visualization environment and past data reporting functionalities</p>

2. *Phase 2:* To develop a CM data visualization environment using a top-down approach, based on the guidance of design principles/guidelines and structured analysis of relevant data visualization and CM literature. The focus of the design is to identify requirements from the conceptual level for an overall CM data visualization environment that can serve CM analytics in general, to the more detailed level for the subpart of an overall CM data visualization environment that is specific to time performance management. An existing CM data visualization environment, which was designed and developed in the past using a bottom-up approach, was compared against prescribed requirements. A case study using schedule/as-built data from an actual project and an existing CM data visualization environment were utilized for time performance management applications to:

- Demonstrate the merits of utilizing interaction features to explore and formulate

visual representations of CM data,

- Test the use of a top-down approach for analyzing visualization requirements that relate to common/general CM analytics for CM functions/tasks, and
- Identify lessons that can be learned regarding how to develop a CM data visualization environment and key features of images and interaction features.

3. *Phase 3*: To develop a CM data visualization environment using a bottom-up approach in accordance with the design principles/guidelines and visualization requirements identified in phase 2. The focus of the development is to identify, implement, and evaluate new visualization features of a CM data visualization environment in support of specific CM analytics including schedule variance analysis, product/location attribute analysis, and reasons for time performance analysis within the CM function of time performance control. The inspection kind of evaluation method (Amar and Stasko 2005; Ardito et al. 2006; Zuk et al. 2006) is used to evaluate implemented features against prescribed specifications, requirements, and analytic needs in order to identify the need for refinements. Case studies using product/schedule/location/as-built data from actual projects and the newly enhanced CM data visualization environment were utilized for time performance control applications to:

- Demonstrate the merits of the environment architecture proposed in this thesis,
- Test the use of the bottom-up approach for developing visualization features that are responsive to both common and specific CM analytics for CM functions/tasks, and
- Identify lessons that can be learned regarding how to develop a CM data visualization environment and key features required of images and interaction features.

4. *Phase 4*: Demonstrate how the key concepts used to define a CM data visualization environment may enhance CM analytics capability. The demonstration and analysis are done by comparing a CM information system that is integrated with the developed prototype CM data visualization environment against current data reporting practices that mainly use textual/tabular reports and a few graphics (e.g. bar graph schedule). Specifically, case studies, conventional data reporting functionality, use of the CM data visualization environment developed as part of this research work, and a

structured comparison process were employed to demonstrate and analyze how the use of a CM data visualization environment may enhance CM analytics capability.

## **1.6 State-of-the-art data visualization and its application to CM**

A full literature review of the use of data visualization in construction management (including the data visualization capabilities of current commercial CM information systems) and an overview of state-of-the-art data visualization technologies were carried out in order to understand the current state of applying data visualization technologies to facilitate CM visual analytics in support of CM processes. Due to the number and richness of the images associated with the text for describing the results of the literature review, more detailed descriptions of them are presented in Appendices A and B (Appendix A: Data visualization in construction management; Appendix B: Overview of state-of-the-art data visualization). A distillation of the findings from the literature review documented in Appendices A and B contributes to answering the first two questions under the question heading of "How should a CM data visualization environment be developed?" (see page 16), which includes the identification of shortcomings of current applications that should be addressed and how state-of-the-art data visualization can be adopted/adapted to developing a CM data visualization environment. The findings of this review are summarized below.

### **1.6.1 Shortcomings of current applications of data visualization in CM**

1. *Data for limited CM tasks/functions is visualized:* Exclusive of the visualization of the geometric aspects of a project (e.g. built products, equipment) and how the geometric aspects progress in time (e.g. 4D model, equipment operation-topics that have been extensively investigated in the literature, and which are not covered in my literature review), the use of visualizations of abstract CM data in the past focuses mainly on construction process planning, scheduling, and estimating. These visualizations usually present values of the independent and dependent variables that are explicitly defined in a prediction model (e.g. Figure A.9 in Appendix A) and are usually limited to "showing a plan" in terms of schedules, cash flows, and resource allocation. However, many CM variables that relate to planning

assumptions/conditions are not either explicitly recorded or visualized. These variables include planned values of project environment attributes in support of risk management (e.g. political environment, economic environment, climatic environment), product attributes in support of scope, cost, and time management (e.g. product quantity, quality requirement/grades, design features), location attributes in support of risk management and planning (e.g. work area size, utility distribution), and organization/contractual attributes for contract management (e.g. organization performance appraisal).

It is only recently that researchers have started to explore the potential of visualizing as-built data to support the control function of CM. However, the focus is mainly on monitoring deviations between planned and actual cost (e.g. Figure A.27, Figure A.29 in Appendix A). Little was found about using visualization of CM data to monitor other performance measures and to seek reasons for performance from data representing actual values of planning conditions/assumptions including visualization of metadata used to describe documents that record a wide range of as-built events/conditions (e.g. problems encountered, change directives, meeting minutes discussing project/construction issues, site reports, etc).

2. *No domain wide CM analytics-oriented visualization is supported:* Most current CM data visual representations (e.g. bar graph schedule, earned value index graphics) are developed as part of data-oriented and task-oriented approaches and their use tends to be restricted to a single function- e.g. schedule data for planning and scheduling (e.g. Figure A.4 in Appendix A), earned value data for cost control (e.g. Figure A.36 in Appendix A), probability-impact register for risk analysis (Figure A.19 in Appendix A). They are not developed and used in conjunction with the need to know potential causes of and/or effects on CM variables that in turn could be helpful in conducting CM tasks (e.g. identify a schedule's potential impacts on cost implications of project participants and refine the schedule to eliminate the impacts). Understanding potential causes and/or effects amongst CM variables relevant to a CM task can add significant value to all kinds of CM tasks/functions.

Currently neither academia developed nor commercial CM information systems have provided a visualization environment that allows users to generate and interact with a wide variety of visual representations of CM data that support common and unique analytic tasks associated with the diversity of tasks/functions that comprise CM. For example, current commercial CM information systems are usually modularized into different software components or packages that support specific CM tasks/functions (e.g. Oracle-Primavera Risk Analysis software is for risk management and Oracle-Primavera P6 software is mainly for scheduling and schedule tracking). Hence the data and data visualization are specific to the tasks/functions supported and can only be accessed in the individual software component. To achieve the integrated use of data visualization in different commercial software packages, issues that have to be addressed include: 1) data integration, 2) image accessibility and composition across software packages, and 3) whether images designed for one package support common CM analytics shared by other CM tasks.

### **1.6.2 How state-of-the-art data visualization can be adopted/adapted**

1. *Limitations of state-of-the-art data visualization and focus of development:* Although computer technology development and research accomplishments in data visualization have helped advance the use of data visualization from paper-based, static, and passively-presented graphics to computer-based, dynamic, and actively-exploring visual interfaces, the development of data visualization tools still faces two fundamental problems. The first one is the technical problem of limited resources, e.g., limited size of visualization space, number of visual variables, and human visual memory. The second problem is the complex interactions amongst variables including characteristics of visual display, task demands, data complexity, human factors, and user characteristics (e.g. users knowledge about the designed graph, context or information the graph represents). State-of-the-art data visualization can be creatively applied to overcome some of these technical limitations (e.g. use of interaction features to more quickly browse through many images representing different data dimensions/contents). To date, however, fundamental research in data visualization at best has answered a small part of the questions related to the aforementioned complex

relationships (e.g. simplistic relationship between visual encoding, data characteristics such as measurement scales of data, and effectiveness in judging data values). From the perspective of applying existing state-of-the-art data visualization principles and tools as opposed to adding to them, the challenge becomes one of utilizing the state-of-the-art to overcome technical issues (e.g. accommodate more images in one view with clarity for quicker scanning utilizing commonly used graphics).

2. *The need to use domain wide analytics-oriented data visualization development methodologies:* For many state-of-the-art information/data visualization tools, they tend to be one or more of technology-oriented (develop innovative technologies), data-oriented (solve issues of visualizing massive data; how to portray the same data in different ways), and/or specific task-oriented (solve specific analysis or visual analysis problems). In general their focus is not on the breadth of analytical reasoning associated with a complete knowledge domain, nor is it oriented to generalists as opposed to specialists within a domain. For example, task/specialist-oriented visualization development may focus on solving clutter issues associated with presenting activity sequences in a traditional bar graph schedule (i.e. how can one see sequences between activities more effectively and efficiently). An analytic-oriented development that covers the entire domain of CM on the other hand will first figure out the implications to CM of activity sequencing as a design/development guidance (i.e. the conclusion may be that there is little or no benefit to show sequences in bar graph schedules from the CM analytics perspective). Non-domain wide analytics-oriented visualization development usually goes through a structured development process by eliciting requirements in terms of specific task needs (or task problems, visualization problems) from a small number of end users, then design and implementation, followed by conducting user studies to quantitatively or qualitatively appraise the usability and/or utility of the developed tools. However, the goal of a CM data visualization environment is to support CM domain wide analytics. Thus, the development approach proposed by (Amar and Stasko 2005) appears to be more suitable for pursuit of this objective. This approach involves the idea of developing an

analytics-centered visualization system, and it tries to identify heuristic rules<sup>1</sup> in terms of how a visualization system should support certain high-level analytics (e.g. a visualization system should assist in creating/acquiring/transferring knowledge about important domain parameters) as a reference for designing the system and inspecting whether the developed system meets these analytics needs. Therefore, development methodologies that are more in line with this idea should be adapted to formalizing structured processes for developing a CM data visualization environment, which includes an analytics oriented, heuristic inspection kind of evaluation.

3. *Understand the underlying fundamental functionalities of novel data visualization techniques/tools:* Table B.1 in Appendix B demonstrates how computer technologies and two areas of data visualization fundamentals (i.e. visual encodings and the use of interaction features) were applied to creating novel/advanced data visualization tools. While they help provide tangible examples of innovative visualization techniques, designers should understand the underlying functionalities (i.e. items in the left column of Table B.1 in Appendix B such as good practices of mapping data to visual variables, data query, adjusting visual attributes such as scales of images) supported by the special techniques of these novel tools and contemplate their underlying usefulness to CM analytics rather than being tempted to incorporate these new visualization techniques in a CM data visualization environment because of their novelty (i.e. applying their use irrespective of the nature of real analytics needs and data structures). For example, Hyperbolic Tree (Figure B.5.17 in Appendix B) visualization that is built on an innovative visual representation of data may enhance the easiness of navigating and searching textual information organized in a hierarchical structure. This understanding of the fundamental functionalities of hyperbolic tree visualization (i.e. goal of enhancing usability; for navigating information; applicable to hierarchical categorical data) enables designers to assess its fit of use in support of the analytical tasks at hand (e.g. it is not suitable for identifying cause-effect relationship between CM variables). Designers should refrain from enthusiastically trying/pursuing the use

---

<sup>1</sup> Heuristic rules are a set of pre-defined criteria, rules, or guidelines that need to be considered when developing a software interface or visualization (Carpendale 2008) and are for looking for problems in the developed products in terms of their compliance to the heuristic rules (Mankoff et al. 2003).

of new visualization techniques simply because of their novelty or popularity.

4. *Desired degree of generality/flexibility of a data visualization tool*: current state-of-the-art data visualization tools at one end of the spectrum are single function tools either supporting narrowly defined information search/data analysis tasks or dealing with particular data characteristics (e.g. a Treemap specifically for visualizing hierarchical data, Parallel Coordinate for visualizing multidimensional quantitative data) and at the other end of the spectrum are generic visualization systems (e.g. (Eick 2000; Stolte et al. 2002)) for general public use (i.e. require no programming to access state-of-the-art data visualization features). A specific tool may provide only one or two images and a set of interaction features to interact with these images for supporting a few particular analysis tasks. A generic system is flexible enough for users to choose whatever data dimensions from databases and specify the mapping between data dimensions and visual variables. In other words, many images could be created, if not an unlimited number (e.g. unlimited combinations of data dimensions to include in a visualization compounded by unlimited choices of visual encodings). The level of generality (or uniqueness) of a CM data visualization environment should be in between the two extremes of specific tools and generic systems, and there should be a limited number of thematic images that are particularly useful for CM analytics. However, a certain degree of flexibility can be still be maintained through the use of essential interaction features to query background data for viewing different data contents associated with the same image theme, to adjust visual attributes to enhance readability, and to coordinate the use of interaction features for efficiency.

## **1.7 Research contributions**

In this section, the first two of three major research contributions and related sub-contributions made are described and explained. The validation for each of contribution 1 and contribution 2 is done by describing how the research results answer the research questions in this section. While the third contribution is overviewed here, the assessment of the contribution is done through case studies, structured comparison evaluation, and results analysis, as documented in the conclusion chapter.

**Contribution 1<sup>2</sup>: Answering the research question: of "What methodology should be used to adopt or adapt state-of-the-art data visualization in order to develop/enhance a CM data visualization environment, which address shortcomings of the current use of data visualization in CM?"**

*Contribution 1.1: Formulation of a set of design guidelines/ principles in terms of how to apply state-of-the-art data visualization (or visual analytics) techniques to the CM domain data.* These guidelines contribute to the application of structured data visualization design/development processes tailored to CM needs (e.g. all visualization designs start with analyzing analytical reasoning needs) and extracting knowledge fundamental to data visualization and CM data that pertains to the design/development of general CM data visualizations. The design guidelines address in a comprehensive manner the following issues that were not covered in previous methods proposed for developing CM information visualization (Lee and Rojas 2009; Shaaban et al. 2001; Songer et al. 2004) :

- Recognizing that CM analytic needs govern visualization design (analytics-oriented design) and provide evaluation benchmarks.

This recognition is a contribution because "CM analytics-oriented" design differs from traditional "task oriented", "data oriented", or "technology oriented" design approaches (Lee and Rojas 2009; Shaaban et al. 2001; Songer et al. 2004). Consider, for example, visualizing schedule data. The focus of CM analytics-oriented visualization is on "showing the rationales behind a construction plan (e.g. potential effects (impacts) by or reasons for such a plan)" as opposed to "showing a plan", "visualizing available schedule data ", or "visualizing something that available visualization technologies can offer".

- Emphasizing the need to organize CM data representations and transformations for portraying information and knowledge responsive to CM analytics needs, which leads

---

<sup>2</sup> Contribution 1 deals with the third question under the question heading of "How should a CM data visualization environment be developed?" (see page 16). The first two questions under this heading are addressed in Appendices A & B and section 1.6 which contribute to an in depth understanding of the shortcomings of current applications of data visualization in CM and how state-of-the-art data visualization can be adopted/adapted to develop a CM data visualization environment.

to the recognition of project context/performance data dimensions and the consideration of data granularity and data derivation of those data dimensions.

This recognition is a contribution because data reports are usually in textual formats and concepts of project context/performance data dimensions provide a direction for transforming fuzzy messages/insights contained within the data into focused and important CM variables in the form of structured data dimensions, which in turn can be mapped to visual variables thereby creating images that address CM analytic needs.

- Incorporating ideas from state-of-the-art data visualization and characteristics of CM users for designing visual representations of CM data, which include collectively considering conventions/good practices of visually encoding data, use of 2D/3D space, and leverage of interaction features.

The recognition and use of these visualization design rules is a contribution because it unburdens designers in terms of allowing them to conduct "satisficing" visual encoding designs so that they can focus on analyzing CM analytics and visualization requirements.

- Including an evaluation step to ensure users are enabled to glean insights from the implemented visualizations that help answer CM questions thereby leading to the understanding necessary to take appropriate management actions as required. This evaluation step can be integrated into the design process to assist in refining design of the visual images and their interaction features. A qualitative kind of self valuation combined with contrasting the design against those proposed by others (if any exists) is recommended. (A methodology that was considered but excluded is qualitative evaluation by construction management practitioners. This user study type of evaluation is suitable for inquiring into performance differences of industry end users and degree of adoption by industry, a research topic that is beyond the scope of my research work. For assessing basic concepts or changed paradigms of thinking, it is fraught with difficulties given the action orientation of most CM personnel).

This identification is a contribution because it not only addresses the needs of evaluation and evolving design/development, but also sets out guidance as to how to conduct CM data visualization evaluation that fits the reality of the CM domain (both in industry and academia).

*Contribution 1.2: Introduction of a top-down design approach for identifying common CM analytic needs and the corresponding visualization requirements.* Applying this approach results in the identification of the general/common CM analytic reasoning required by the overall CM function/task and time management function/task domains, and visualization requirements responsive to those analytics needs. Items identified are as follows:

- Overall CM function/task domains
  - Concept of understanding characteristics of construction conditions, construction performance, construction dependency relations; CM analytics for predicting/planning purposes; CM analytics for monitoring/diagnosing/controlling purposes.
  - General visualization requirements that respond to the foregoing CM analytics such as representing different status value states for condition/performance couplets.
- Time management function/task domains
  - CM analytics for planning/predicting time (mainly for appraising quality of a schedule); CM analytics for diagnosing/monitoring/controlling time.
  - General visualization requirements responding to the foregoing CM analytic reasoning needs such as associating construction conditions that can be represented visually with individual activities.

The top-down design methodology is a contribution because it provides a structured way of identifying the scope and direction of visualization development and formalizing common and essential features of images and associated interaction features required of a CM data visualization environment (e.g. images to present distribution of values of construction conditions/performance measures in terms of primary project context dimensions). Through a top-down approach, an environment architecture for a CM data visualization tool can be established. In this environment architecture, an organization of thematic CM data visualizations, mainly hierarchical from the abstract (e.g. time performance visualization) to the specific (e.g. duration variances displayed by location and activity), can be developed in a consistent way. The major benefits of employing a

top-down design approach to design a visualization environment architecture are two fold: 1) the environment developed possesses visualization features for supporting CM analytics that are common to a wide range of CM functions/tasks and hence can be utilized by a variety of CM functions/tasks, 2) the learnability of using the visualization environment is enhanced because different visualizations have consistent image features/attributes, allowing users to quickly apply the learning from using a few visualization to operating others. The first benefit addresses a crucial property of visualization environment architecture, i.e., the sharing of common design features amongst multiple CM functions/tasks, as shown in Figure 1.7.

*Contribution 1.3: Employment of a structured development process that combines a bottom-up design process integrated with design guidelines and a top-down design process tailored to the needs of the CM discipline.* The bottom-up development method treats details for creating an actual visualization that supports a specific CM analytics task. The process embraces CM analytics, specific visualization requirements, and visualization specification analysis, implementation, and evaluation. Its integration with design guidelines ensures that state-of-the-art data visualization is applied. Its integration with a top-down approach in terms of ensuring that common visualization features are addressed helps to preserve the environment architecture described in Contribution 1.2. Most importantly, its integration with the design guidelines and the top-down approach means lessons learned or new visualization features identified that may be common to other visual analytic reasoning needs can be added to the design guidelines or checklists of common visualization features. The concept of the foregoing integrated development process is depicted in the flow chart seen in Figure 1.8.

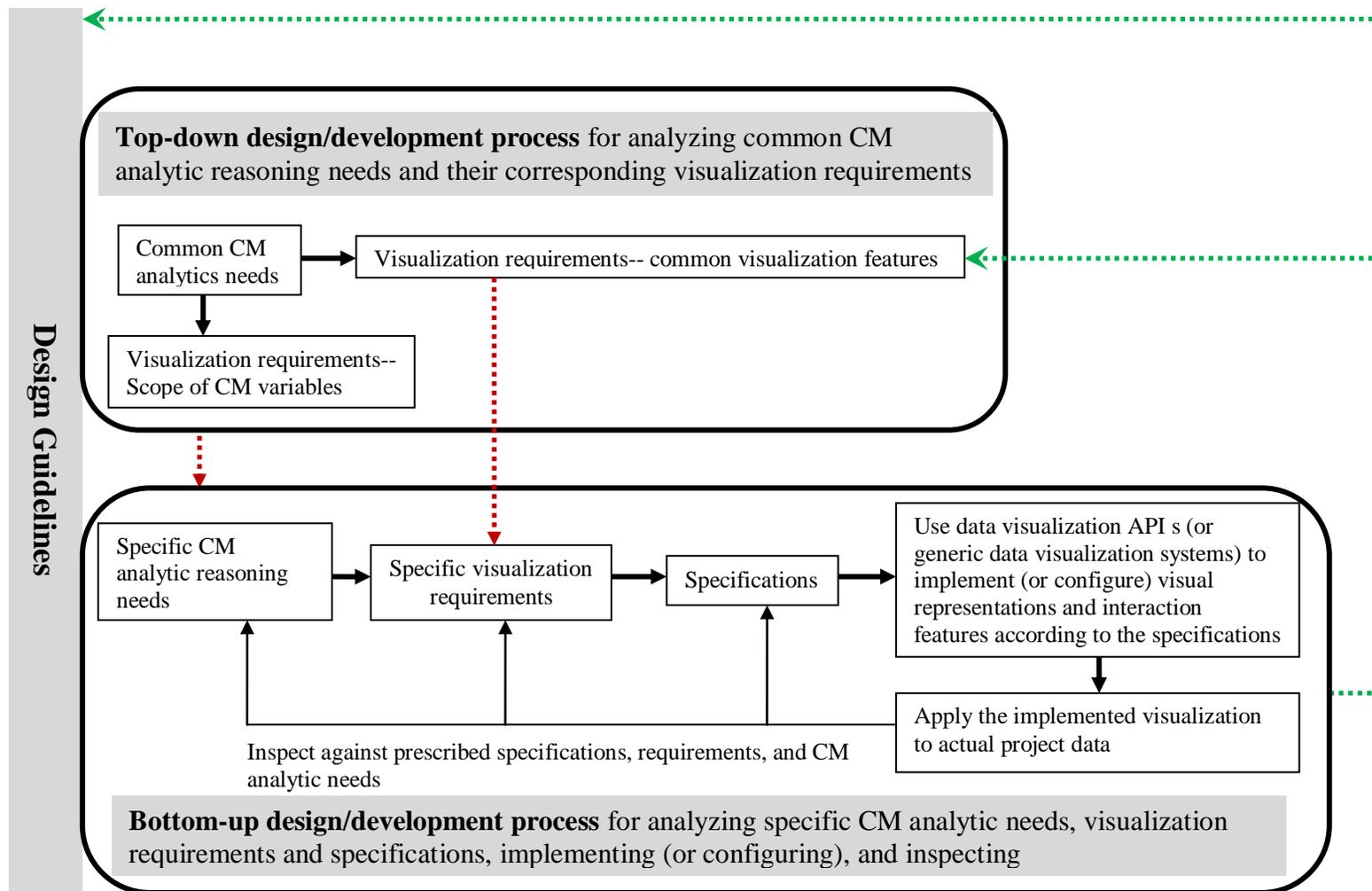


Figure 1.8 A two way structured top-down and bottom-up CM data visualization environment development process

**Contribution 2: answers the research questions of "What are the key features of a CM data visualization environment that best reflect the functions expected for it?"**

Lessons learned from conducting this research contribute to understanding key features of a CM data visualization environment that pin point *the needs of assisting CM analytics and allowing users to flexibly explore a variety of images useful for CM analytics*. These key features can be described as follows:

1. An organization of thematic visualizations, mainly hierarchical from abstract to specific, that are categorized by construction conditions and performance measures under multiple views of a project.
2. Each visualization has common primary and secondary features as summarized in Table 1.2 and Table 1.3 and is designed in accordance with the design guidelines described in Chapter 2 plus the additions presented in Table 1.4. In Table 1.2 and Table 1.3, the first column contains the common image features that need to be addressed when designing a new visualization; the second column contains settings for common image features that can be chosen to fit the uniqueness of a specific visualization designed to serve particular CM analytics. The last column presents the kind of interaction features corresponding to users selecting or changing the settings listed in the second column. The interaction features shown in the third column and elaborated upon in Table 1.4 in terms of the ability to change visual encodings are the "image specific" interaction features. During the bottom-up design process for specific visualizations, the focus of designing "image specific" interaction features is on "what settings of common image features and/or visual encoding **cannot** be changed". Principles of general interactivity for enhancing readability that deal with attributes of scales, orientations, and positions that are independent from visual encoding data values are treated as part of the design guidelines and can be observed in Table 1.4.
3. Different visualizations have commonalities in terms of shared visualization features and adherence to design guidelines. They are also unique because they respond to different specific CM analytics (e.g. different CM variables) and hence require different settings of the common visualization features or even new visualization features.

The primary common image features are the most fundamental ones and have been fully incorporated into the implementation of a prototype of a CM data visualization environment. Figures 1.9 to 1.13 illustrate these primary common image features as described in Table 1.2.

The identification of the common visualization features and refined design guidelines is a contribution because a CM data visualization environment can be developed in a structured way and tailored to CM user needs to explore visual representations of CM data at different levels of detail in order to identify visual patterns that represent the following insights that could not be readily deduced through other means:

1. Potential impacts of planned values of construction conditions on project context dimensions (e.g. activities, locations, project participants, time windows, pay items) and in turn inference of impacts on future construction performance (i.e. in support of evaluating quality of plans and identifying risk).
2. Potential impacts of planned and/or actual values of construction conditions on project context dimensions (e.g. activities, locations, project participants, time windows, pay items) and in turn inference of whether the planned or actual values of construction conditions have impacts on actual construction performance (i.e. in support of diagnosing reasons for actual performance).
3. Potential reasons for planned and/or actual values of performance measures by project context dimensions (e.g. activities, locations, project participants, time windows, pay items) and in turn inference that variances of performance measures are dependent on certain activities, locations, project participants, etc. (i.e. in support of diagnosing reasons for actual performance).
4. Potential cause-effect relationship amongst construction conditions and performance measures by comparing time stamped planned and/or actual values of them on the time dimension (i.e. in support of diagnosing reasons for actual performance).

Table 1.2 Primary common CM visualization features: presenting construction conditions or performance measures mapped against primary project context dimensions

Image features	Setting options for image features	Interaction features for changing option settings
<p>Image themes by construction conditions or performance measures (scope, time, cost, safety, quality, etc), which are treated as measurement dimensions</p> <p>Examples of choices of image themes are shown as selectable items in Figure 1.9</p>	<p><i>Kinds of CM data dimensions treated as construction conditions or performance measures</i></p> <ul style="list-style-type: none"> <li>• Attributes characterizing project context dimensions (e.g. activity-attributes, product-attributes, location-attributes, organization-attributes, environment-attributes)</li> <li>• Derived attributes (e.g. productivity)</li> <li>• Counts of data records related to values of context dimensions (i.e. number of data topics that are related to certain project contexts)</li> <li>• Other attributes of context dimensions (e.g. daily weather conditions, daily site conditions, daily problems encountered)</li> </ul>	<p>Select visualization</p>
<p>Image types by project context dimensions against which construction conditions or performance measures are mapped</p> <p>Examples of different image types (distribution <b>by product and location dimensions vs. distribution by project participant and location dimensions</b>) of the same image theme (as-built record distribution) are shown in Figure 1.10.</p>	<p><i>Kinds of CM data dimensions treated as project context dimensions</i></p> <ul style="list-style-type: none"> <li>• Definition of construction conditions or performance measures</li> <li>• Activity</li> <li>• Location</li> <li>• Product</li> <li>• Project participant</li> <li>• Occurrence time (data version dates can be used as occurrence time if the occurrence time is difficult to know or no "occurring behaviour")</li> <li>• Others (e.g. pay item, environment)</li> </ul>	<p>Select visualization</p>
<p>Image contents by granularity of context dimensions</p> <p>Examples of different image contents (record distribution in the location dimension that is at the level of granularity of <b>“location set” vs. “location”</b>) of the same image theme/type (record distribution; record distribution by the location dimension) are shown in Figure 1.11.</p>	<p>Inclusion of all, singular, or multiple levels of granularity of the project context dimensions</p>	<p>Select data granularity</p>
<p>Image contents by items selection of</p>	<p>Inclusion of all, singular, or multiple (group or range) values of the project context</p>	<p>Select data range</p>

<b>Image features</b>	<b>Setting options for image features</b>	<b>Interaction features for changing option settings</b>
<p>project context dimensions</p> <p>Examples of different image contents (<b>“all deficiencies” vs. “long head time deficiencies”</b> record distribution in the location dimension) of the same image theme/type (record distribution; record distribution by the location dimension) are shown in Figure 1.12.</p>	<p>dimensions. In addition to direct selection of items, they can be selected by values of attributes characterizing the context dimensions.</p>	
<p>Image contents by data status states</p> <p>Examples of choices of different image contents by different status states can be selected by user interface such as a combo box shown in Figure 1.13</p>	<ul style="list-style-type: none"> <li>• Non-variance case: planned, actual, and/or planned vs. actual</li> <li>• Variance case: actual (current) - actual (target), actual (current) - planned (target), and/or planned (current) - planned (target)</li> </ul>	<p>Select data status</p>

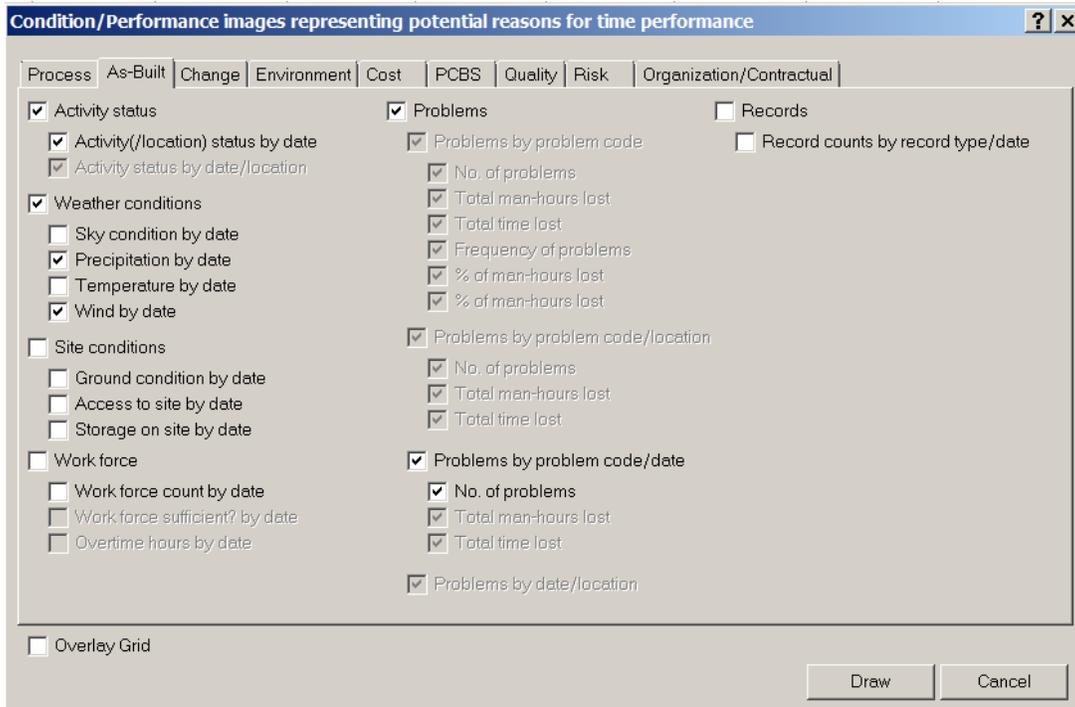


Figure 1.9 A user interface for image selection by construction conditions that are grouped under project views (i.e. the tab items such as "process", "as-built"). This user interface only allows users to choose images representing distribution of values of construction conditions in certain project context dimensions. This figure showcases the primary common visualization feature of “image theme by construction conditions or performance measures” described in the 1st row of Table 1.2

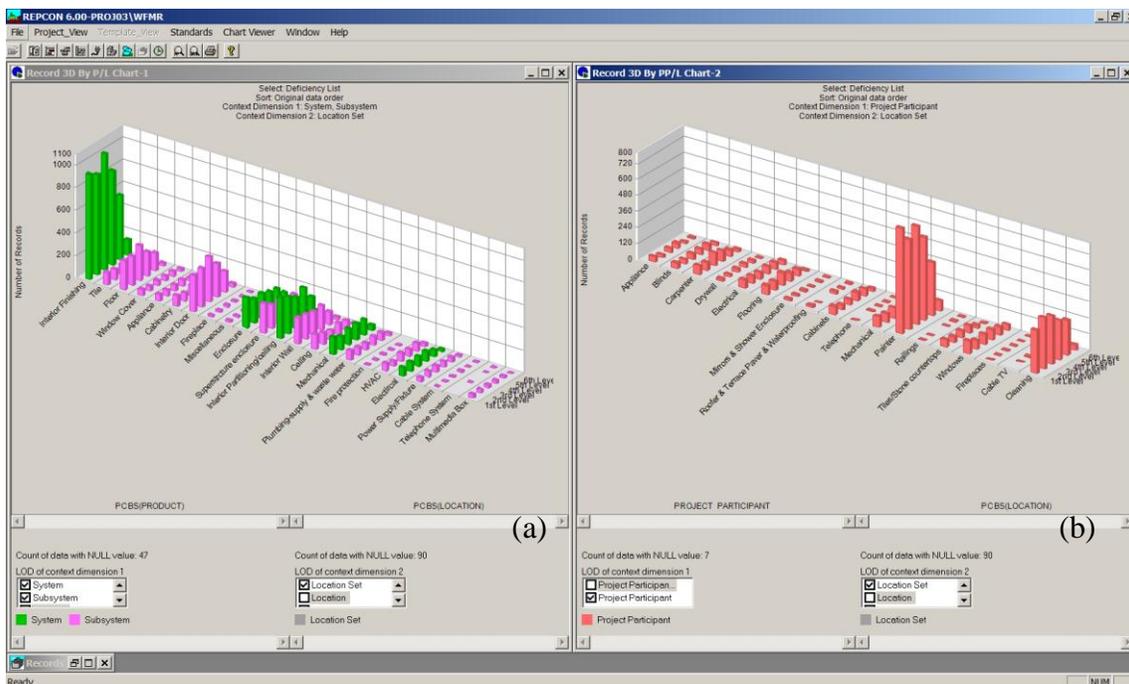


Figure 1.10 Two record distribution images visualizing number of deficiencies distributed in different project context dimensions: (a) product and location dimensions, (b) project participant and location dimensions. This figure showcases the primary common visualization feature of “image type by context dimensions” described in the 2nd row of Table 1.2

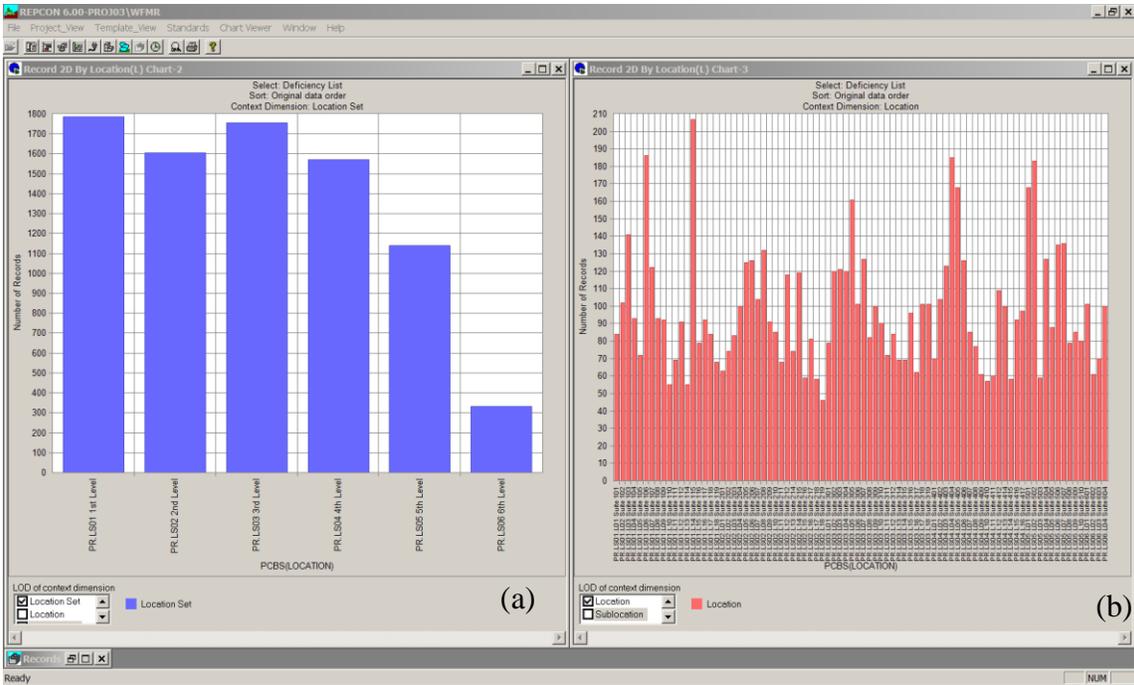


Figure 1.11 Records distribution images visualizing the number of deficiencies distributed in the location dimension but at different levels of granularity: (a) location set (stories), (b) location (rooms). This figure showcases the primary common visualization feature of “image contents by granularity of context dimensions” described in the 3rd row of Table 1.2

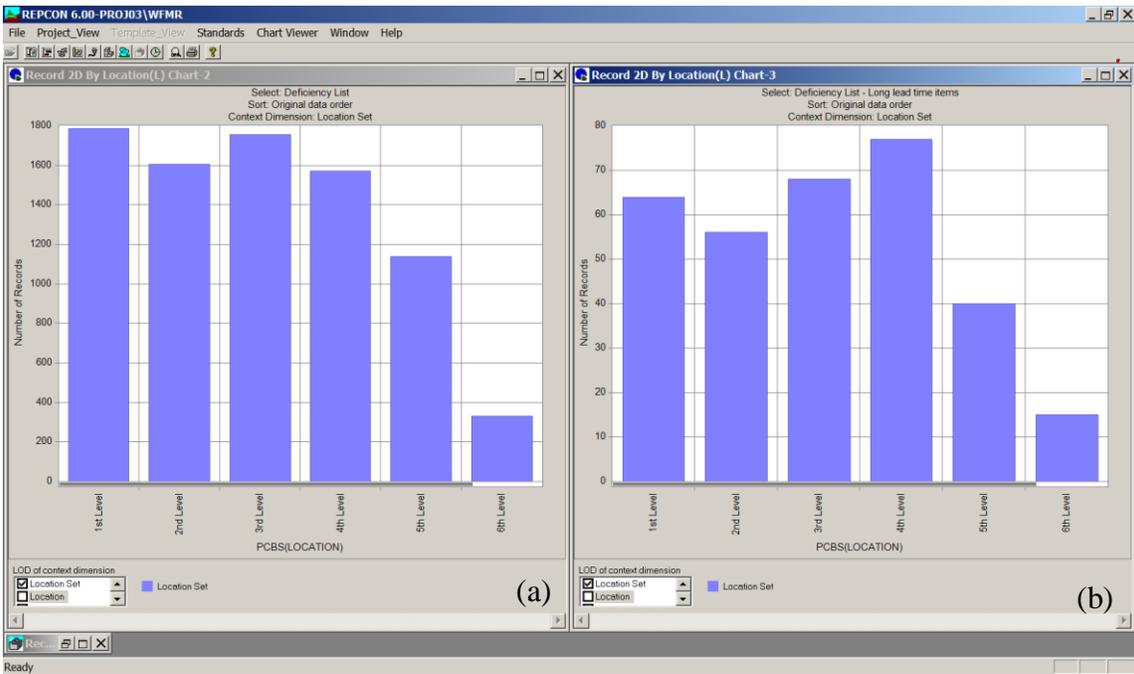


Figure 1.12 Records distribution images visualizing the number of deficiencies distributed in the location dimension but of different data value ranges: (a) "all" deficiencies, (b) only "long lead time" deficiencies. This figure showcases the primary common visualization feature of “image contents by items selection of project context dimensions” described in the 4th row of Table 1.2

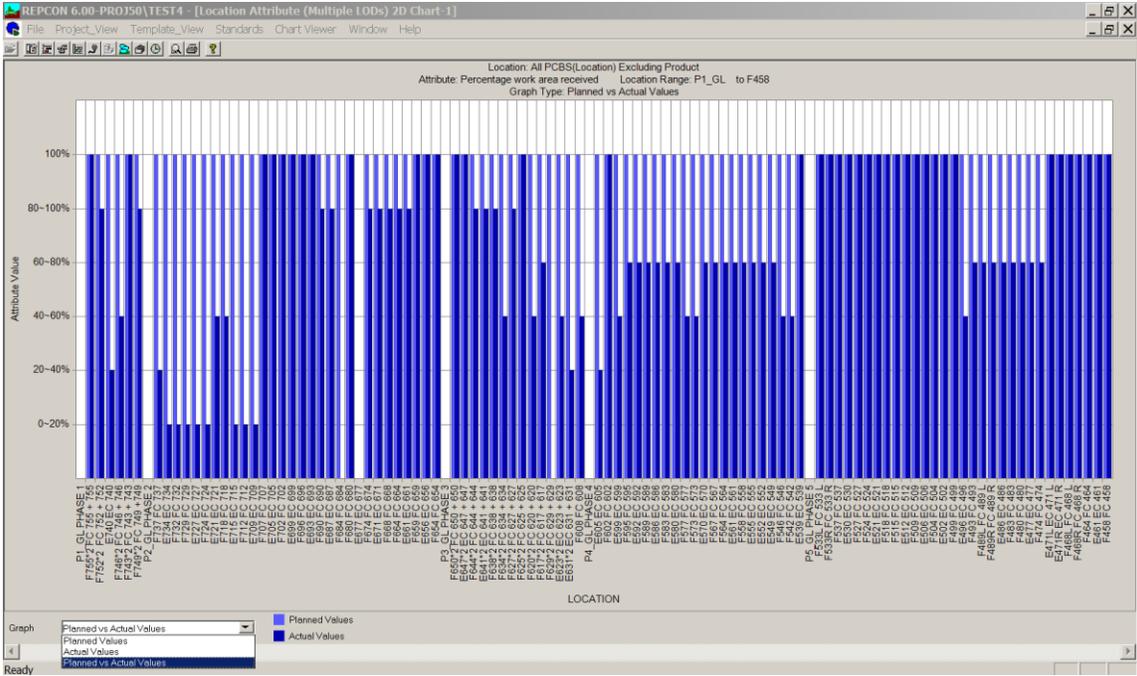


Figure 1.13 The user interface (the left bottom combo box) for adjusting data status (i.e. planned, actual, and planned vs. actual work area received percentage). This figure showcases the primary common visualization feature of “image contents by data status states” described in the 5th row of Table 1.2

Table 1.3 Secondary common CM visualization features: secondary feature consideration after addressing the primary visualization features

Image features	Settings options for image features	Interaction features for changing option settings
At least two basic image formats	<ul style="list-style-type: none"> <li>• Distribution of values of construction conditions or performance measures in project context dimensions. Examples of this image format are discussed in the design cases 1 and 2 of Chapter 4.</li> <li>• Distribution of values of several construction conditions or performance measures in the occurrence time dimension and definition dimension (i.e. definitions of construction conditions or performance measures ). Examples of this image format are discussed in the design case 3 of Chapter 4.</li> </ul>	Select visualization
Image types by non-variance or variance values	<ul style="list-style-type: none"> <li>• Visualizing non-variance values</li> <li>• Visualizing variance values</li> </ul>	Select visualization
Image contents by data versions (dates)	<p>Non-variance case: original data version (date), and/or several data update versions (dates)</p> <p>Variance case: one or several paired data versions (dates)</p>	Select data versions (dates)
Image contents by how to aggregate measurements	<ul style="list-style-type: none"> <li>• Aggregation not allowed</li> <li>• Sum, Max, Min, Median</li> <li>• Others</li> </ul>	Select data aggregation method
Image contents by how to compute variances between the planned and actual values of measurements	<ul style="list-style-type: none"> <li>• No variance computation</li> <li>• Variance computation method</li> </ul>	Select variance computation method
Encode "holiday" "non-working day" if the time dimension is involved	<ul style="list-style-type: none"> <li>• Encode "non-working day" and by what working calendar</li> <li>• Do not to encode "non-working day"</li> </ul>	Select display options
How to include values of context dimensions	<ul style="list-style-type: none"> <li>• Include all values of a context dimension (e.g. all activities modeled, continuous time points)</li> <li>• Only include values of a project context dimension that have corresponding values of construction conditions or performance measures (e.g. only activities that have computed variance values)</li> </ul>	Select display options
Enhancing the visual grouping of certain items of context dimensions that are not visually encoded by spatial position	<ul style="list-style-type: none"> <li>• Use connecting lines (e.g. using lines to connect bars representing activities of the same project participant)</li> <li>• Use normal visual encoding such as coloring (e.g. bars representing activities of the same project participant have the same color)</li> </ul>	Select display options

Table 1.4 Design guidelines in addition to those presented in Chapter 2

Guidelines subjects	Guidelines
<b>Choice of data representation/transformations</b>	<ol style="list-style-type: none"> <li>1. The time dimension can be treated as the context dimension (i.e. occurrence time dimension) or time performance measures</li> <li>2. To create a null item as one value of a project context dimension because values of a construction condition may be associated with items of one context dimension but not associated with another context dimension (therefore associate with Null)</li> </ol>
<b>Choice of visual representations</b>	<ol style="list-style-type: none"> <li>1. Choice of visual marks for different ways of value assignments for measurements dimensions <ul style="list-style-type: none"> <li>• Singular valued--bars (i.e. lines), points</li> <li>• Multiple valued--points</li> <li>• Value range--Gantt bar</li> </ul> </li> <li>2. Suggested visual encodings <ul style="list-style-type: none"> <li>• If the visualizations involve visually encoding more than one data status state, two colors can be used to encode planned and actual status; three other colors can be used for three variance status states.</li> <li>• <i>For image format 1 described in the first row of Table 1.3:</i> Map context dimensions and the data version (date) dimension to positions on three orthogonal coordinates; map construction conditions/performance measures to Z axis (i.e. the coordinate that goes upward). How to encode other data dimensions (e.g. differentiating levels of granularity, data status states) will be treated on a case by case basis. <i>For image format 2 described in the first row of Table 1.3:</i> Map construction conditions/performance measures to Z axis (i.e. the coordinate that goes upward), map the time dimension (e.g. events occurring dates, data version dates) to positions on the axis perpendicular to Z axis. How to encode other data dimensions (e.g. context dimension other than the time dimension, differentiating levels of granularity, data status states) will be treated on a case by case basis.</li> <li>• When multiple project context dimensions are mapped to spatial positions, visual marks representing all items of these dimensions at different levels of granularity will be positioned. The ordering of visual marks follows the rules of: 1) first ordered by different project context dimensions (the ordering is to be specified by users), 2) then ordered by dimensions of different levels of granularity (from the coarsest to the finest), 3) then ordered by data values of context dimensions (the ordering is to be specified by users if the type of value scale is categorical). Visually differentiating project context dimensions of multi-levels of granularity and different project context dimensions by applying visual variables to their corresponding chart gridlines, labels, and/or strips</li> </ul> </li> </ol>
<b>Choice of interaction features</b>	<p><i>Image specific interaction features</i> Provide interfaces for users to change the settings of common visualization features and/or visual encodings. During the bottom-up design processes, designers can determine what settings and/or visual encoding cannot be changed.</p> <p><i>General interactivity</i></p> <ul style="list-style-type: none"> <li>• Global 3D virtual space and local optimum image format</li> </ul>

Guidelines subjects	Guidelines
	<ul style="list-style-type: none"> <li>○ Allow the users to globally change the position, orientation, and scale of the 3D virtual space</li> <li>○ Allow locally change positions, orientations, and scales of individual visual representations</li> <li>○ Optimum orientations and scales are fixed for elements (e.g. labels, titles) in a visual representation</li> </ul> <ul style="list-style-type: none"> <li>● <i>Coordinating image specific interaction features</i></li> </ul> <p>It is suggested that the coordination be specified by users. Interfaces and mechanisms should allow users to specify what visual representations to include for interaction coordination and what kind of shared image specific interaction features should be coordinated.</p>
<b>Evaluation</b>	<p><i>Method and focus of evaluation 1: evaluation that is done by designer/developer and needs shorter time to perform after implementation</i></p> <ul style="list-style-type: none"> <li>● Conformance to specifications and requirements</li> <li>● Identify features required by the CM analytics not addressed in the requirements/specifications analysis</li> <li>● Improvement on usability, mostly for ease of use of interaction features and good visual effects (e.g. visual encoding does help provide visual effects of data grouping)</li> <li>● Identify new lessons learned or new visualization features common to the visualization environment for updating design guidelines and checklists of common visualization features</li> </ul> <p><i>Method and focus of evaluation 2: evaluation that is done by users and need longer time to perform after deployment</i></p> <ul style="list-style-type: none"> <li>● Identify which visualization in the organization of thematic visualization are frequently used</li> <li>● Identify new visualizations or new visualization features needed</li> </ul>

**Contribution 3: Answers the research question of "How does the use of a CM data visualization environment help conduct CM analytics that cannot be done or which are difficult to do with current data reporting practices?"**

*Contribution 3.1: Demonstrated and analyzed, through case studies, how the ability to carry out CM analytics useful for dealing with the CM tasks at hand may be enhanced (i.e. human judgment enhancement from the "light color" in Figure 1.3 to the "dark color" in Figure 1.7) using data reports in visual form that are responsive to these analytics tasks.*

*Contribution 3.2: Demonstrated and analyzed, through case studies, how CM analytics ability may be enhanced using interaction features and an environment architecture that allows users to flexibly explore CM data collected/computed for a range of CM tasks/functions presented in visual form.*

Through the demonstration cases described in section 5.3 (sections 5.3.1~5.3.3) of the conclusion chapter, it is shown that a CM data visualization environment allows users to identify more analytic reasoning artifacts and more quickly than examining data in its original textual or tabular form. Furthermore, through the analysis conducted on the demonstration cases, it was found that it is certain features of a CM data visualization environment, which are related to “reports in visual forms”, “use of interaction features”, and “an environment architecture” and which are lacking in the current tabular data reporting mechanism, are the ones that contribute to enhanced analytical reasoning performance. How these features differ from the ones that accompany tabular data reporting and traditional data images to contribute to enhanced CM analytics was analyzed and explained in a concrete way through referring to the demonstration cases and their corresponding images. A detailed discussion can be found in section 5.3.4 of the conclusion chapter. How the main features of a CM data visualization environment that contribute to the enhancement of existing CM analytics capabilities can be summarized as follows:

1. By presenting data in visual form, salient visual patterns representing values and patterns of behaviour of construction conditions or performance measures can be

instantly observed.

2. Images that depict the distribution of values of either construction conditions or performance measures in various context dimensions provide helpful insights as to whether or not their causes can be inferred or the impacts they generate can be inferred.
3. A thematic visualization representing CM data, which is collected/computed for a certain CM function/task, can also be utilized for CM analytics supporting other CM functions/tasks. Many such thematic visualizations have been organized for CM\_IS users to access and conduct flexible analytical reasoning.
4. Users can adjust/select visual formats of data presentations of a specific thematic visualization on demand to meet their CM analytics needs.
5. Users can adjust/select data contents in terms of granularity of data sets of the visualization chosen on demand to meet their CM analytics needs. This involves choice of levels of granularity of context dimensions and aggregations of values over context dimensions of different levels of granularity.

## **1.8 Structure of the thesis**

**Chapter 1 Introduction:** This chapter describes the background (problems and proposed solutions), goals (questions, scope or focus, assumptions), methodologies, literature review (presented in the form of appendices), and contributions of the research.

The "contributions of the research" part of the chapter focuses on describing and justifying the first two main contributions that piece together research findings identified in Chapters 2 through 4 for providing a complete view of answers to the first two research questions. The third contribution related to answering the third research question is also overviewed. The assessment of that contribution is done through case studies, structured comparison evaluation, and results analysis, as documented in the conclusion chapter.

**Chapter 2 Visual Representation of Construction management Data** (*a version of a published paper*): This chapter describes the phase 1 research work related to developing guidelines and principles for designing a data visualization tool tailored to CM use.

Research findings provide partial answers to the first and second research questions.

**Chapter 3 Design of a Construction Management Data Visualization Environment: a Top-Down Approach** (*a version of a published paper*): This chapter describes the phase 2 research work related to developing a CM data visualization environment using a top-down approach in order to identify visualization requirements for an overall CM data visualization environment that can serve *CM analytics in general*. Research findings provide partial answers to the first and second research questions.

**Chapter 4 Design of a Construction Management Data Visualization Environment: a Bottom-Up Approach** (*has been submitted for publication*): This chapter describes the phase 3 research work related to developing a CM data visualization environment using a bottom-up approach in order to identify, implement, and evaluate new visualization features of a CM data visualization environment in support of *specific CM analytics* including schedule variance analysis, product/location attribute analysis, and reasons for time performance analysis within the CM function of time performance control. Research findings provide partial answers to the first and second research questions.

**Chapter 5 Conclusion- Summary, Answering the Research Questions, Contributions, Future Work:** This chapter concludes the research, and incorporates reporting on the fourth and final phase of the research work in terms of answering the last research question. Six demonstration and analysis cases are used to: 1) show and assess how the use of a CM data visualization environment helps conduct CM analytics that cannot be done or which are difficult to do with current data reporting practices, and, 2) provide an indicator of degree of validity and generality of the research findings in answering the research questions set forth in Chapter 1.

**Appendix A Data Visualization in Construction Management:** This appendix provides a full literature review of the use of data visualization in construction management (including the data visualization capabilities of current commercial CM information

systems).

**Appendix B Overview of State-of-the-Art Data Visualization:** This appendix provides an overview of state-of-the-art data visualization technologies in order to understand the current state of applying data visualization technologies to facilitate CM visual analytics in support of CM functions.

## **Chapter 2 Visual Representation of Construction Management Data<sup>3</sup>**

### **2.1 Introduction**

Construction project participants are confronted with the need to make high quality and timely decisions based on the information content that can be deduced from the very large data sets required to represent the various facets of a project through its development life cycle. How best to extract information from such data sets is a question that preoccupies researchers and practitioners alike across a number of disciplines, including construction. One approach to reasoning about data is visual analytics, the science of analytical reasoning facilitated by interactive visual interfaces (Thomas and Cook 2005). We believe it has special appeal to the construction industry because of its visual orientation, and because visual analytics has the potential to be directly usable by construction practitioners without a requirement for specialist knowledge or assistance.

Visual analytic models provide the building blocks for development of an interactive visualization environment which is tailored to the special needs of a particular industry and user audience profile in order to help personnel glean insights from large and complex data sets. In this chapter, we use the term visual analytics environment to refer to a computerized information system which treats pre-coded scenes that consist of one or more visual representations and accompanying interaction features, and a user interface that assists users to interact with data and pre-coded images for analytic reasoning purposes. Further, we use the term visual analytics model to refer to the specifications of requirements of components for implementing a visual analytics environment. Also with respect to terminology, a distinction is drawn between the terms, visual analytics, and data visualization, with the latter referring to the use of computer-based, interactive visual representations of data to amplify cognition (Card et al. 1999). In effect, data visualization corresponds to one of the components of visual analytics.

---

<sup>3</sup> A version of this chapter has been published. Russell, Alan. D., Chiu, Chao-Ying., Korde, Tanaya. (2009). "Visual Representation of Construction Management Data." *Automation in Construction*, 18(8), 1045-1062.

The design of effective visual analytics models is built on four pillars: (i) the purpose(s) of the analytical reasoning; (ii) the choices of data representations and transformations; (iii) the choices of visual representations and interaction technologies (i.e. data visualization); and, (iv) the production, presentation, and dissemination of the visual analytics findings. Of these four dimensions, data representations and transformations constitute the foundation on which visual analytics is built, while the use of visual representations and interactions to accelerate rapid insights into complex data is what distinguishes visual analytics software from other types of analytics tools (Thomas and Cook 2005). A data representation is a structured form, which is generated from the original raw data and which retains the information and knowledge content within the original data to the greatest degree possible (Thomas and Cook 2005). A data transformation deals with transforming data into varying levels of abstraction or deriving additional data that has new semantic meaning. Visual representations translate data into a visible form that helps the analyst perceive salient aspects of the data quickly. Interaction technologies support dialogue between the analyst and data (Thomas and Cook 2005). Thus, the second and third pillars identified previously constitute the core of a visual analytics model.

Described in this chapter are aspects of our work related to the application of visual analytics to the domain of construction management. Our particular focus is on how visual representations and interaction technologies, in concert with a nine-view data representation of a construction project (e.g. physical, , process, cost, as-built, quality, change, organization/contractual, environmental, and risk views) (Russell and Udaipurwala 2004) that supports a range of construction management (CM) functions, can improve the construction management process. Two research hypotheses guide our work: (i) the application of visual analytics to CM functions (e.g. change order management, quality management, drawing control, schedule analysis, etc.) improves the CM process through enhanced understanding of project status and reasons for it, improves communication amongst project participants, assists with the detection of potential causal relationships, and improves decision making; and, (ii) a visual analytics environment can be developed which is sufficiently general to serve the needs of a broad range of CM functions. Our

concern is with the practical application of visual analytics, with practical meaning the use of visualization technologies that are compatible with the constraints associated with the construction industry – e.g. a heterogeneous user audience with highly variable education backgrounds, a focus on action and results as opposed to exploration, and ease of use without the requirement for specialist assistance. Implications of the foregoing statement are at least two-fold: (i) the user audience is comprised of generalists as opposed to specialists; and (ii) usability by practitioners of the visual representations developed depends on an implementation strategy that encodes the representations in a ready-to-use manner along with the ability to interact with them to extract the greatest meaning possible. To illustrate the concepts and principles presented, attention is focused on change management and its corresponding change-data view, which also interacts with one or more of the other eight project-data views. Nevertheless, the concepts and principles described are broadly applicable to other CM management functions.

Designing good visual representations involves several challenges: fit of a visual representation with the characteristics of users within a given industry (e.g. visual perception and cognitive abilities); scalability of a visual representation; and, the degree to which a visual representation is useful. With the aid of interaction techniques and careful arrangement of data representations and transformations, these challenges can be addressed. In the construction world, there can be multiple target audiences, and the type of visual representations used may vary from one audience to another depending on their comfort with 2D, 3D, and more complex ones. The scalability issue of visual representations also exists due to the volume of data generated for a large scale, complex, construction project. Thus while construction data representations have many dimensions which need to be translated into visual representations, at most one must work with a three dimensional visualization space. Therefore, one is confronted with the need to consider various combinations of dimensions to develop the understanding required for analytical reasoning. Considering the foregoing issues, substantive research and design challenges must be addressed in formulating data representations and transformations that reflect knowledge and information in the context of the analytic reasoning tasks of interest and

crafting them into visual representations with relevant interaction techniques so that expert to novice users can extract important information hidden in the data. Success in meeting these challenges can be measured in terms of the breadth of analytical reasoning supported, including the number and total domain value of insights gleaned by industry users (i.e. the total sum of the significance of the insights generated) (Saraiya et al. 2004), and the ease of use of the visualizations supported.

Of several questions that need to be addressed in pursuit of proving the research hypotheses described previously, three are examined in this chapter, with emphasis being placed on the first two.

- Q1: What principles and guidelines should be used for designing a visual analytics environment for a broad-based treatment of construction management functions, in terms of analytical reasoning tasks supported, data representations and transformations supported, and corresponding visual representations and interaction features?
- Q2: Can the usefulness of individual visual representations designed in response to specific analytic reasoning tasks be perceived by users, thus lending support to the hypothesis that the application of visual analytics to construction management functions improves the CM process?
- Q3: How should a visual analytics environment be designed and implemented so that it is responsive to the realities of the construction industry and satisfies the criterion or test of practicality?

The remainder of the chapter is structured as follows. A brief overview of the general motivation that drives research on visualization is provided. This is succeeded by a short description of recent construction data visualization work. Important principles or guidelines related to visual representations are then discussed, with emphasis on the viewing dimensions of data and thus how it may be portrayed, features desired of a visualization environment, and how the usefulness of various visual representations may be evaluated and validated. Our interest lies with examining collections of entities as

opposed to individual entities (e.g. a register of change orders as opposed to an individual change order). Then, visual representations of change order data are explored in detail in order to examine issues associated with the questions posed and application of the principles/guidelines set out in the previous section. The chapter concludes with a discussion of findings from work performed to date, and their extension to other construction management functions and data types.

## **2.2 Motivation for use of visualization**

Representing data in a visual format helps amplify cognitive ability or reduce complex cognitive work (Card et al. 1999; Keim et al. 2006). Humans can derive overview information from data better and faster if it is presented in a suitable visual format other than textual/numerical scripts or tables. This is because features such as spatial positions or colors provide low similarity amongst different features than do texts or numbers, which is one of the key reasons why human beings can be visually attentive to certain symbols (Duncan and Humphreys 1989) and identify visual patterns prior to conscious attention (Ware 2004). Another explanation states that this is because large amounts of visual/diagrammatic information can be processed by the human visual perception system in parallel as opposed to the serial processing required for textual or numeric information (Larkin and Simon 1987; Ware 2004). Based on these theories, various attributes of the data of interest are mapped against certain features in the visual representation like color, size, shape, location or position thereby reducing the need for explicit selection, sorting and scanning operations within the data (Shneiderman 1994; Tufte 1990). These techniques thus tailor the data to be retrieved, such that the large arrays of neurons in the eyes can rapidly extract features of visual representations and distinguish salient visual patterns (Ware 2004) that correspond to patterns hidden in the data. This helps the target audience achieve insights faster and better as to the information content of a data set that may otherwise be concealed or not easy to comprehend from its representation in tabular or text form.

For the current state-of-the-art of computerized visualization techniques, data

representation is often coupled with real time interactive tools like zooming and filtering, details-on-demand windows and setting dynamic query fields, which allow users to browse through and study the represented data. Emphasis is also placed on the rapid filtering of data to reduce the result sets (Ahlberg and Shneiderman 1994). This is called visual data exploration. Thus, visualization can be described as a two-fold process of data presentation and data exploration. Effective visual representation schema assist the efficient scanning of different parts of an organization's or project's database, allowing users to instantly "identify the trends, jumps or gaps, outliers, maxima and minima, boundaries, clusters and structures in the data" (Brautigam 1996). Exploration tools allow continuous interaction between users and the graphic displays by offering scope for "constant reformulation" of search goals and parameters as new insights into the data are gained (Ahlberg and Shneiderman 1994). They provide a continuously updated information platform to users, thereby aiding the decision making process.

Of the several references reviewed that describe frameworks for classifying visualization techniques, we mention one in particular because of its potential applicability to the construction domain. Specifically, to reduce the 'complexity inherent in choosing a visualization technique for a particular application context', Lengler and Eppler (Lengler and Eppler 2007) compiled a pre-selected group of a hundred visualization techniques thought to be applicable to management functions in the form of a periodic table analogous to the Dmitri Periodic table of elements. Through this structure, the authors highlighted the fact that for a given requirement there need not be just one appropriate visualization method. Rather, there is a potential of employing a combination of different methods to enhance understanding. Such an approach may be particularly appropriate for the construction domain, as it has the potential to enhance practicality and ease of use of visual representations for different management functions.

### **2.3 Data visualization in construction**

In carrying out the literature review on visualization techniques, we also undertook to identify the extent to which they have been applied to the field of construction. The

majority of the work described in the literature has focused on visualizing the spatial and temporal aspects of construction project data, with very limited emphasis being placed on the visualization of abstract, non-spatial data. A rich literature has developed over many years dealing with 2D, 3D, and 4D and even nD visualization of the physical artefact to be constructed (Collier and Fischer 1995; Heesom and Mahdjoubi 2004; McKinney and Fischer 1998; U.S. General Service Administration 2008). For example, there is a growing use of 3D and 4D models to minimize the potential for design and construction errors in the construction product, to identify critical space and time during construction (Dawood et al. 2005), to determine the most suitable construction methods and sequence, and to monitor construction progress (e.g. (Sriprasert and Dawood 2003; Staub and Fischer 1998)). For visualizing some aspects of a project's data, Song et al. (Song et al. 2005) proposed a 3D model-based project management control system where the visual platform (i.e. the 3D building model) itself serves as a construction information delivery platform. The system enables the user 'to show a holistic picture of a project by applying the multiple project data sets to the geometric attributes (such as shapes, faces, and edges) of the 3D building model components through color- tone variation and motion. The proposed control system uses a Project Dashboard as the user control interface allowing the user to freely choose the sets of data to apply to the different visual attributes of the 3D model. Although this approach makes it relatively easy to visually associate project control data with components of the physical product, how best to generate insights from abstract construction data that require representations of salient spatial and one or more of temporal, or organizational patterns (e.g. clusters, trends, and anomalies) is not obvious because spatial positions have been dedicated to represent only the geometric data of the built product.

In contrast to visualizing the physical artefact to be built for purposes of constructability reasoning or workability of the methods selected for its construction, or even accessing relevant project information through the mechanism of a 3D model, our primary focus is on the visualization of abstract construction management data in support of exploratory data analysis and the application of project participant tacit knowledge. Specifically, our

interest is with collections of entities (e.g. change orders, drawings, RFIs (Request for Information), etc.) and their association with other entity collections, with the definition of a collection being determined by the choice of values for one or more properties of an entity associated with a management function. Somewhat surprisingly, there is very little literature that addresses visualization of construction data (Rojas and Lee 2007), particularly with respect to how data visualization can play an important role in aiding analytic reasoning for a range of CM functions. This observation results in both an opportunity and a challenge for researchers exploring the use of data visualization and visual analytics for construction. The opportunity is that it is a relatively virgin field of inquiry. The challenge is that when positing ideas either for the design of visual images themselves or complete visualization environments, in terms of validating ideas there is very little with which to compare and contrast. Consequently, it is important to set out some basic principles and guidelines against which one can assess the usefulness of image and visualization environment designs proposed.

A project's database is voluminous, containing data that varies from textual form such as drawing specifications and contractual clauses, to quantitative data like number of change orders and related properties (e.g. value, timing, number of participants), RFIs issued and turn around times, SIs (site instructions), correspondence, photos, drawing control data, planned and actual schedule data, weather conditions on site, and cost breakdowns. The data is generally time and location variant and originates from or affects multiple project participants. The sheer volume and nature of the data poses significant management challenges. Further complicating these challenges is the observation that construction data is often poorly organized because it lacks proper grouping and sub grouping which can lead to missed opportunities to associate related data or facts, and more often than not it is incomplete. For effective management of a project, efficient handling, monitoring and control of all project data is essential. Buried within this data are important messages which relate to the reasons for performance to date, but extracting this information from any database, especially a poorly organized one can be very difficult (even if a database is well organized, linkages amongst different data items may not be obvious – data

visualization may in fact help one forge relevant links). As a consequence, explaining different aspects of construction project performance often qualifies as a classic case of “data rich - information poor” problems (Songer et al. 2004). Thus, the massive amount of data available to management personnel results in information overload (Songer et al. 2004) unless it is accompanied by a high level of organization and accompanying reporting mechanisms.

Songer et al. (Songer et al. 2004) explored the use of Treemaps and other visual aids like scatter plots and histograms for assessing cost performance. They described an iterative process of structure-filter-communicate while considering level of detail, density, and efficiency of data representation. Vrotsou et al. (Vrotsou et al. 2008) applied Time Geographical methods to visualize work sampling data to allow analysts to understand better the distribution of activities and the interdependencies amongst them. For assessing schedule quality and aiding communication (e.g. feasibility, matching production rates, avoiding trade stacking, achieving work continuity, making clear the location sequence of work, etc.), especially for projects characterized by repetitive work, Russell and Udaipurwala (Russell and Udaipurwala 2000a; Russell and Udaipurwala 2000b; Russell and Udaipurwala 2002) and Zeb et al. (Zeb et al. 2008) demonstrated the value of using linear planning charts. Combined with ancillary images pertaining to the distribution of resource usage in time and space, additional insights on the quality of a schedule can be gleaned. Zeb et al. (Zeb et al. 2008) also explored the visualization of as-built data in terms of job site conditions encountered, problems associated with individual activities, and the juxtaposition of site condition parameters with daily activity status in support of causal reasoning. Zhang et al. (Zhang et al. 2009) used an integrated building information system and digital images captured on site to semi-automate the calculation of progress measurements (e.g. cost and schedule variance) for items of a work breakdown structure and then facilitated their visualization using data filtering techniques (i.e. single work package selection) and a composite of images to represent various progress measurements.

A limitation of work to date on abstract construction data visualization as opposed to

physical product visualization is that it is mainly exploratory rather than systematic in nature, with limited breadth in terms of the type of data and information entities examined, management functions examined, and guiding principles for designing relevant visual images. Thus while such individual explorations are useful, there has been a lack of an extensive program of research directed at determining what roles visualization can play across multiple functions using a common framework of principles, and what properties should be present in a visual analytics environment tailored to visualizing construction data.

## **2.4 General principles of visual analytics design processes**

The beneficial application of visual analytics begins with understanding the purposes of the analytical reasoning involved in conducting the management functions of interest. This understanding in turn provides guidance as to what data representations and transformations are desired. Then, based on the structure of these representations and transformations, data can be collected or derived. Lastly, with structured data at hand, strategies of mapping the data onto visual representations can be explored while considering the limitations on visualization space, differences in end user cognitive and visual perception abilities, and interaction techniques available. The associated design process is an iterative one that integrates an evaluation process which captures and incorporates feedback from the intended user audience. In the following subsections, principles of conducting the aforementioned steps in the visual analytics design process as applied to construction management functions are explained. They have been gleaned from an extensive review of the literature and from hands on design and exploration of visual images for various analytic reasoning tasks for a range of management functions. Later in the chapter they are applied in the context of change order management; nevertheless they are broadly applicable to a range of functions.

### **2.4.1 Understanding the purposes of analytical reasoning**

Different project managers have different thinking styles, experiences, and knowledge (Tullett 1996). Therefore, it is difficult to predict the steps a person takes to explore, acquire, organize, and use information to assist analytical reasoning (Stolte et al. 2002;

Tullett 1996). However, in general, the analytical reasoning involved in construction management is about gaining understanding from the perspectives of different project context dimensions, the characteristics of construction conditions (e.g. constraints, requirements, environment) and construction performance dimensions (e.g. time, cost, quality), and then confirming/exploring how construction conditions and construction performance are interrelated – i.e. identifying potential causal relationships amongst the two. In essence, the focus of analytic reasoning is on *assessing* conditions and performance and *communicating* findings to different audiences in forms that facilitate interpretation. Condition and performance characteristics can be described at three different levels: overall characteristic (i.e.: overall qualitative pattern); local characteristic (i.e.: local qualitative pattern); and, individual characteristic (i.e.: single value.) This somewhat oversimplified categorization represents a generalization from the authors’ distillation of problem solving intentions observed in past CM researches and project management principles. This distillation analysis along with a more detailed taxonomy of analytical reasoning involved in construction management functions across different construction phases are left for extended discussion in chapters 3 and 4. Nevertheless, suffice it to say that identification of the primary purposes (assessing and communicating) to be served by analytical reasoning is very important as it provides guidance and focus for the design of the components of a visual analytics model (e.g. a collection of useful pre-coded images presenting various construction conditions, construction performance dimensions, and possible causal relationship amongst them), and serves as a benchmark for evaluating the efficacy of a design (Amar and Stasko 2005).

#### **2.4.2 Organizing data representations and data transformations**

In response to the purposes of analytical reasoning outlined in the foregoing, we identified several context dimensions (e.g. time (when), space (where), responsibility (who), physical system/component (what), work environment (natural & man-made work conditions)) for representing a project’s context. Each context dimension can be characterized by a number of quantitative and qualitative attributes to describe planned versus actual construction conditions. Further, one instance of a context dimension can be interrelated with another

instance of the same dimension through the sharing of the same value for other dimensions (e.g. two activities can share the same time, space, responsibility and work environment). Complementing context dimensions are performance dimensions which include measures such as time (how long), cost (how much), quality (e.g. number of deficiencies), safety (e.g. days and man hours lost to accidents), and scope (e.g. value of change orders). The notion of characterizing a construction project in terms of both quantitative and non-quantitative context dimensions and quantitative performance dimensions provides a cornerstone for forming structured data representations that reflect the information and knowledge needed by construction personnel.

Data transformations are directed at qualitative abstractions and quantitative aggregations and dis-aggregations of both context and performance dimensions to reflect different levels of granularity, and at deriving new semantically meaningful dimensions. As an example of a data transformation dealing with level of granularity, the space dimension can be expressed at different levels of detail, such as a sub-location (e.g. east wing), an individual location (e.g. 2<sup>nd</sup> floor), or a group of locations (e.g. all superstructure locations). An example of a data transformation dealing with data derivation is the computation of a site or work location congestion index which is defined as work space area divided by resource usage rate.

The aforementioned conceptual principles regarding organizing data representations and data transformations are based on the nature of CM analytical reasoning requirements and the characteristics of structured data (e.g. measurement scales of data values, data items being relationally and hierarchically related, etc.). Therefore, they are independent of construction project information models proposed by several researchers (e.g. (Abudayyeh and Al-Battaineh 2003; Froese 1996; Karim and Adeli 1999; Kim and Liu 2007)). However, because the essence of exploratory CM analytical reasoning is to be able to examine construction conditions, construction performance, and causal relationships amongst them for various project context dimensions, an integrated construction management information model (e.g. (Russell et al. 2004)) is essential to support analytical

reasoning based construction data representations and data transformations.

### **2.4.3 Designing visual representations and interaction features**

When trying to design visual representations and their interaction features, four major constraints need to be taken into account: (i) purposes of analytical reasoning; (ii) characteristics of the construction domain (e.g. the rather broad spectrum of user cognitive and visual perception abilities encountered in the construction industry, and limited resources such as time and cost for conducting the analysis and communicating the results); (iii) space limitations on the visual display and the multidimensional data representations that need to be presented; and (iv) the extent to which the user can interact with data and its visual representation. With respect to the last constraint, its removal by maximizing interaction capabilities helps to cope with the other constraints, thereby facilitating the design and use of more flexible visual representations that best meet the purposes of analytical reasoning.

To date we have identified three main general rules of thumb for initiating draft designs of visual representations. We observe that such rules have not been systematically and integrally discussed in the CM literature:

1. **Follow conventions and good practices:** Use effective visual encoding principles (i.e.: choice of encodings depends on measurement scales of data values (Cleveland and Robert 1984; Mackinlay 1986)); use conventions and good practice of graphing data (Bertin 1983 (originally published in French in 1967); Brath 1999; Cleveland 1985; Schmid 1983; Tufte 1986; Unwin 2008; Wainer 1997; Wilkinson 1999); and, use conventional graphics elements (e.g. orthogonal coordinate layout, points, lines, bars, pies) because natural standards and organized standards of graphics (Schmid 1978) have formed people's basic graphics literacy over the years, which is one of the factors explaining how effectively people interpret visual representations of data (Shah 2005).
2. **Possible use of virtual 3D space:** Provided with the tool of interactive computer graphics which is far advanced from only the pen and paper used by Playfair to create the first static 2D bar chart (Beniger and Robyn 1978) 200 years ago, researchers have

started exploring the opportunity of utilizing and enhancing its power in order to creatively generate dynamic 3D visualizations to assist data analysis (Cook 2009) and information search (Card et al. 1991). This opportunity should not be overlooked for designing visual representations of CM data especially for purposes of exploratory data analysis. Recently, an innovative design methodology has been proposed (Brath 2003) in which 3D virtual space “houses” several 1D, 2D, and 3D statistical graphics representing data of several dimensions in order to treat one or more analytical functions in one image, for promoting image aesthetics, and for adapting to the evolving desire for and comfort with 3D images by users. We speculate that addressing more than one analytical function in a single image as opposed to in several images may reduce the time needed to analyze multidimensional data. This is because the aesthetic appeal of and human preference for 3D scenes may prolong users’ patience (Cawthon and Moere 2007) and thus keep management staff more attentive/engaged. As well, communication may be enhanced (Brath et al. 2005). Therefore, this design approach should be explored for its potential use in CM data analysis applications similar to what has been done for advancing the use of visualization of geometric data of 3D product models. However, the 2D version of the 3D image should also be produced in order to accommodate users who are more inclined to use 2D visualization.

3. **Use interaction features to solve graphics problems encountered in the design of visual representations:** Interaction features that allow users to interact with data and its visualization (e.g. data query, view navigation, image editing, etc.) can deal with major issues encountered in the use of static graphics such as: image readability problems (e.g. occlusions, illegible labels); the inability to present large and multidimensional data in just one image making it difficult to thoroughly examine datasets from different perspectives and level of detail; and, the inability to change visual encoding to accommodate user visual perception preferences. Thus with the leverage provided by interaction features, flexible designs are possible and rosters of designs can be supported so as to not be constrained by these issues. For example, given space limitations of the display medium, it is difficult to have visual representations of a large data set in which users can observe both overall patterns and

detailed data simultaneously. With the use of an interaction feature for coordinated multiple views, one can design a visual representation to have two or more images, with one showing overview patterns and one or more showing details of data that users select in the overview image in order to know their exact values (Ball and Eick 1996). On the other hand, if using the interaction feature of differential scaling an image on demand of the user, one can design a visual representation requiring only one image in which a focus (showing details of the data of interest) plus context (showing overall pattern of the data) effect can be observed (Rao and Card 1994). Classifications of the generic functionality of interaction features can be found in (Chuah and Roth 1996; Unwin 2006; Yi et al. 2007).

#### **2.4.4 Design evaluation**

The final product of the iterative design process is the set of implementation requirements for the components of the visual analytics environment for the management functions of interest. Two important aspects for validating the final product are usability and usefulness (Grinstein et al. 2003; Plaisant 2004; Scholtz 2006). Usability refers to the ease of use while usefulness examines whether or not the models serve the intended purposes of analytical reasoning. Although usability plays a part in achieving usefulness, the requirements to achieve it are more technology dependent while the requirements for usefulness are much more dependent on the fundamental concepts for designing visual analytics models. Our interest here is to validate the usefulness of the application of visual analytics. We believe that usability issues can be addressed by leveraging the capabilities of cutting edge technology.

In terms of usefulness of a visual analytics model to serve the analytical reasoning purposes identified, one must demonstrate that users are enabled to glean insights and to apply their tacit CM knowledge through viewing the salient patterns shown in the visual representations of data while interacting with the data. These insights must then lead to the understanding necessary to take appropriate management actions as required. We suggest that the process of evaluation be integrated into the design process and in the form of a

qualitative type of method that includes heuristic inspection, collecting opinions, and/or contextual interview (Carpendale 2008). Such an approach allows for the capture of the perceptions of CM experts as to the usefulness of visual analytics in assisting with analytical reasoning for complex CM data analysis tasks, and the identification of features that heighten reasoning capabilities. Contrast this approach with quantitative methods such as controlled experiments which demand significant sample sizes and domain expert time, which in our experience is very difficult to obtain for the CM domain. Therefore, it is recommended that the evaluation process be basically one of self evaluation (the self evaluators themselves are domain experts) combined with comparison and contrasting against visual representation designs proposed by others (which, based on a thorough review of the literature, tend to be very modest in number). Most importantly, the designs can be evaluated by construction personnel to test for the ability of the visual analytics model to provide the analytical reasoning capabilities sought at the outset and to obtain feedback to allow further refinement and analytical reasoning. These steps in the evaluation process are applied to the design images presented in the next section of the chapter.

## **2.5 Design of visual representations of change order data**

For the remainder of this chapter, using data from two retrofit / rehabilitation projects (denoted as Project 1 and Project 2 herein), we focus on the design and evaluation of visual representations for change order data in order to demonstrate application of the thought processes and principles described in the previous section of the chapter. The representations developed can be readily adapted to the exploration of other management functions and data types. They illustrate how visual analytics can facilitate analytic reasoning by providing insights into reasons for performance to date, identifying potential cause-effect relations (e.g. an implicit causal model is that the impact of change orders on time performance is likely to be highest if they are clustered simultaneously in one or more of time, space, by project participant, or physical system), and improving communication amongst project participants. For both projects, our perspective is mainly that of the general contractor (GC) or construction manager (CM) in terms of the change order

management function and the possible impacts of changes on project performance. The examples given here are illustrative of the kinds of situations often encountered on capital projects, and which can be missed because of a preoccupation with individual items as opposed to the collection of many items and related patterns of occurrence – i.e. there can be a failure to see the big picture. This in turn can lead to several undesirable situations, including an underestimation of consequences, failure to initiate corrective action in a timely way, delays, management burnout, loss of entitlement, and loss of reputation, to name a few.

### **2.5.1 Change order management**

A change order (CO) (also referred to as an extra herein) corresponds to an instance of one of the sub-dimensions that comprise the process/information context dimension. COs are tracked at the instance level whether in an integrated information management system or simply by spreadsheet. From a system design perspective, it is useful to treat CO properties in a separate data view (e.g. change view), which is the perspective adopted herein. Properties of interest include CO\_ID (change order identification), date of initiation, date of approval, reason(s) for the change order, project participants affected, estimated vs. approved vs. actual cost, and associations with components used to define other project data views. Other properties derived from associations with other project data views include start and finish dates of the work and hence actual duration (As-built view), physical components affected, where and related drawings (Physical view), and required procurement activities (Process view). Some of these properties are specified by system users while others are derived by the system based on information provided (e.g. durations). A list of change order properties of interest herein, their distribution across different project data views, data type and source are provided in Table 2.1. Typically for projects, a roster of change orders is maintained (e.g. a spreadsheet), and depending on the type of project and procurement mode used, this roster can become very lengthy. As illustrated later, visual analytics provides one approach for extracting and communicating the information content in such a roster.

Table 2.1 Change order properties of interest

<b>Change Order (CO) Property</b>	<b>View*</b>	<b>Data type</b>	<b>Source</b>
CO ID (identity)	CO Mgmt	alphanumeric	User
Date CO process initiated	CO Mgmt	date	User
Date CO approved (cancelled)	CO Mgmt	date	User
Duration of CO initiation/approval process	CO Mgmt	number	Derived
Reason for CO (client initiated, design error/omission,)	CO Mgmt	alphanumeric	User
Date CO work started	As-built	date	User
Date CO work completed	As-built	date	User
Duration of executing CO work	As-built	number	Derived
Number of consultants involved with CO	CO Mgmt	number	Derived
Identity of consultants involved (e.g. architect, structural engineer, .....	CO Mgmt	alphanumeric	User
Number of trades involved with CO	CO Mgmt	number	Derived
Identity of trades involved (e.g. GC, mechanical, electrical, ....)	CO Mgmt	alphanumeric	User
Basis for payment (lump sum, unit price, time & materials, ..)	CO Mgmt	alphanumeric	User
Base cost of CO and cost breakdown, exclusive of impact costs	CO Mgmt	numbers	User
Estimate of impact costs of CO if applicable	CO Mgmt	number	User
Physical component(s) of project affected by CO and locations	Physical	alphanumeric	User
Long lead time procurement items associated with CO	Physical	alphanumeric	User
Procurement item procurement sequence	Process	alphanumeric	User
Association with existing schedule activities	Process	alphanumeric	User
Number of existing activities affected	Process	number	Derived
Association with new activities as a consequence of CO	Process	alphanumeric	User
Number of new activities as a consequence of CO	Process	number	Derived
As-built problems associated with CO	As-built	alphanumeric	User
Identity of existing drawings revised due to CO	Physical	alphanumeric	User
Identity of new drawings due to CO	Physical	alphanumeric	User
Number of RFI's associated with CO	As-built	number	Derived
Identity of RFI's associated with CO	As-built	alphanumeric	User

\* Use is made by the authors of a nine-view data representation of a project: product (physical), process, organizational/contractual, cost, quality, as-built, change (CO Management), environmental and risk (Russell and Udaipurwala 2004)

Changes and change orders are an inevitable part of any construction project. They can have a significant effect on a project and its participants in terms of productivity, and overall project performance. Further, they can give rise to contentious disputes because of their cumulative impact on the efficient execution of other work, and the additional load placed

on management staff. Various researchers (e.g. (Hanna et al. 2004; Moselhi et al. 2005; Thomas and Napolitan 1995)) in the past have tried to quantify these impacts as well as the properties of change orders that have the most adverse consequences for performance.

In terms of analytic reasoning from the perspective of GC/CM or the client with respect to change orders, example questions of interest include the following:

- ***Assessing***
  - To date, what is the distribution of change orders in terms of the context dimensions of time, space, physical system/component, project participant, etc., and what are the potential consequences of this distribution?
  - To date, what is the distribution of change order cost (a performance dimension) in terms of time, space, physical system/component, project participant, etc.?
  - What is the distribution of reasons for change orders, and are they limited to a specific facet of the project or a small subset of project participants?
  - What causal relations appear to exist between the distribution of change orders and project performance as measured in terms of productivity and schedule?
- ***Communicating***
  - How can the change order history to date be communicated in as factual and objective a manner as possible to key participants (e.g. client, architect)?

### **2.5.2 Visual representations for project 1 and design 1**

As indicated previously, rather than focus on the properties of an individual change order, here we show how visual representations can provide a ‘big picture’ of what is happening to a project in the way of changes during its construction phase. In presenting the images in Figures 2.1 and 2.2 for Project 1 which we refer to as Design 1, use has been made of a 122 change order data set including information related to value, timing, location and responsibility of the work. The impact of change orders on labour productivity and project duration became a contentious issue for this project. One approach applied to assess the impact of the value and number of change orders involved use of the kind of analysis offered by Moselhi et al. (Moselhi et al. 1991). But such an analysis ignores the timing

and location of the work, and implicitly contains a retroactivity principle (i.e. future change orders impact work already done). By visualizing the distribution of CO's using relevant meta-data (in this case timing, location and responsibility for the work), a more accurate assessment of potential impact of COs on productivity can be made and other assessment and communication issues addressed. In what follows, the properties of Figures 2.1 and 2.2 are analyzed in terms of the principles presented previously.

### 2.5.2.1 Purposes of analytical reasoning for project 1 and design 1

The analytical reasoning purpose of Design 1 is to examine the as-built change order history to identify trends of change orders versus time, the clustering of change orders in time, space, and by participant to examine possible site congestion issues which could impact productivity or schedule performance, or overwhelm management's capabilities to process and coordinate change orders to minimize the impacts on project performance.

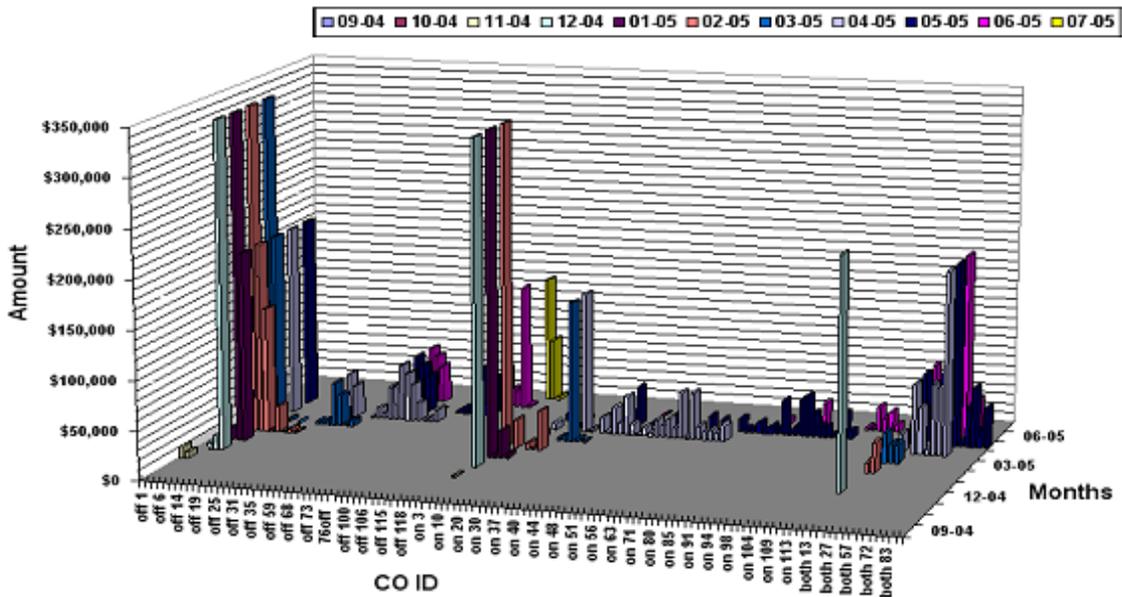


Figure 2.1 Project 1 CO history in terms of ID & Location, timing and value of work

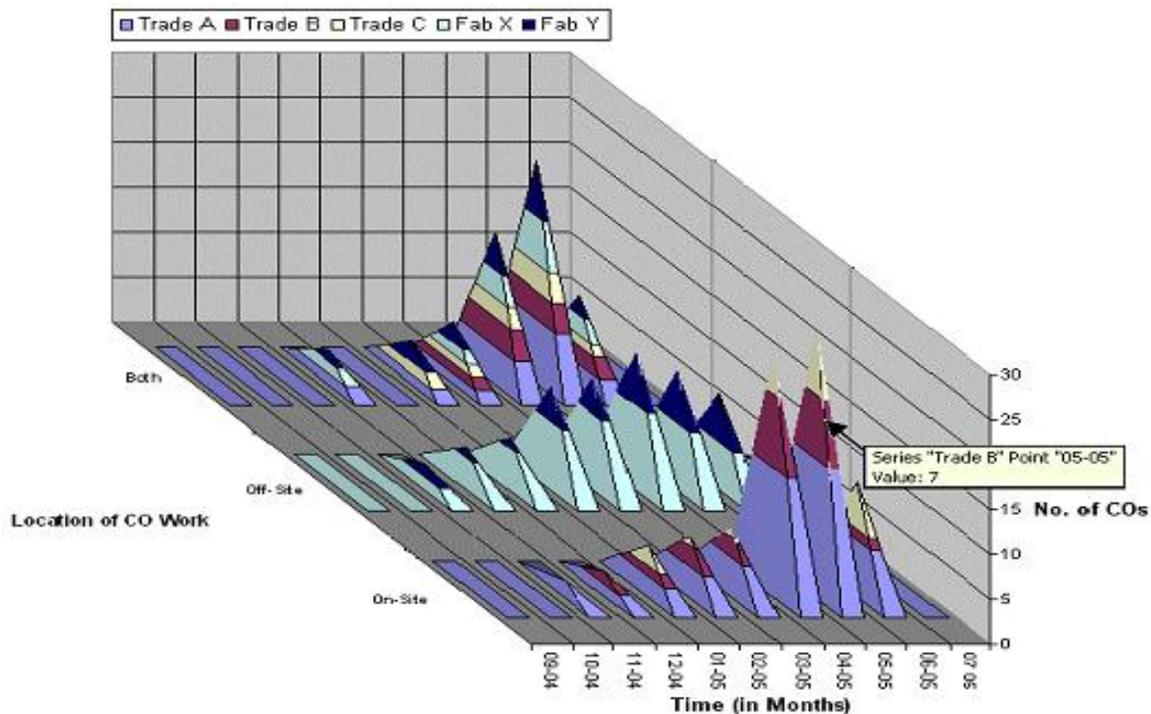


Figure 2.2 Project 1 History of COs by location, time, responsibility and number

### 2.5.2.2 Choice of data representations and transformations – project 1 and design 1

*Data representations:* Given the purposes set for the analytical reasoning, the relevant context dimensions include the process entity of change order (CO) in terms of identification (i.e. CO\_ID), time window of execution and where executed, and performance dimensions of cost and number of change orders. In terms of the original data, time was measured in days and months, and the location dimension was highly aggregated into three values – on-site, off-site, and both off and on site, with the reasoning being that offsite CO's would make little or no contribution to productivity loss or congestion on site. As a general observation, we note that it is important to support different granularities in the definition of time (e.g. day, week, month), location (e.g. individual, group, class), project participants (individual, group, class) and physical components (e.g. individual, group, system).

*Data transformations:* To enhance clarity of the visual representations, we have

transformed the original data by using a more coarse definition of time in terms of months. In response to this transformation, a CO is counted once for each month it is active, and its dollar value is distributed uniformly over its duration. In order to reduce the original four dimensions describing a CO to three to facilitate 3D visual representation, a new CO data dimension was derived by concatenating space and identification number.

### **2.5.2.3 Choice of visual representations – Figures 2.1 and 2.2, project 1 and design 1**

Three dimensions of change order data need to be translated into visual representations. We chose to use positions on three visual space dimensions to encode them in Figure 2.1 because some researchers have found promise in this approach (Robertson et al. 1998). In addition, all COs executed in a given month are mapped against one colour to add clarity to the visual representation. Along the X axis, individual COs are not serially ordered according to their IDs but are sorted by their location. Thus as evident from the figure, the bars grouped at the left end are ‘off-site’ COs, the one in the central area are ‘on-site COs. And COs classified in the ‘both’ category are found at the right end. Thus, from this figure, for a given time instance, one can deduce the total number of COs generated, total base costs associated with the COs, and their concentration in space in terms of an aggregated location descriptor.

Figure 2.2 provides a deeper insight into the project’s set of COs and perhaps tells a more compelling story than Figure 2.1. In this visual representation, each project participant is mapped onto its own colour (as observed later for designs of Project 2, the use of colour to identify participants can become problematic when a large number are involved). The participants are stacked over one another in a predefined order. In this case we have dealt with five participants in total, three on-site trades, Trade A, Trade B and Trade C, and two fabricators, namely Fab X and Fab Y. The vertical performance dimension axis represents the number of COs active for a specific participant in a given month (a dollar value could also have been used). The COs have also been sorted according to their location along the X-axis. This makes the available information easier to assimilate. A single cell in the horizontal plane of the graph yields the project participants involved, the number of COs

active per participant, the active month and the location of the COs. For instance, the arrow in the figure indicates that in the month May-05, Trade B had 7 active ‘On-site’ COs. An interesting observation made from this representation is that Trade A and Fab X have been affected by more change orders in terms of number than any other project participant. This figure also reiterates the message delivered by Figure 2.1 that most of the change orders generated were towards the end of the project time line.

Figure 2.2 also highlights one of the challenges involved in designing visual representations to maximize the clarity and visibility of the data represented, especially for communication purposes with external parties when static or hard copy representations must be used. For larger data sets, if vertical columns had been used, the taller columns in the front of the image would obstruct the view of the bars in behind, thereby hiding much of the content of the image. (In an interactive environment, this problem is lessened as users can experiment with different view angles.) To avoid this problem, we experimented with the use of cones and pyramids, and found the latter provided the most pleasing and useful image. However, perception problems can arise from such a representation. While only height of the pyramid is important, in looking at the image, most individuals implicitly use volume or surface area as the quantification metric, thereby underestimating (or overestimating) the level of effort of specific participants (e.g. Fab Y). Hence, for the representations produced for Project 2, only cylinders are used in order not to bias or distort the insights provided to the user.

#### **2.5.2.4 Evaluation – project 1**

*Overall level* – The two visual representations shown demonstrate that most of the change orders are clustered in the latter stages of the project, although a significant share of the total value of CO work was performed earlier and was associated with just a few on-site and off-site COs. Thus, from an analytical reasoning perspective regarding a potential causal relationship between number and value of COs and reduced productivity and schedule difficulties, one could argue that the clustering of the number of change orders in the latter stage of the project could have impacted productivity, schedule performance, and

management's ability to coordinate effectively all of the changes. In terms of explaining or reasoning about relative performance of project participants, it is clear that Trade A and Fabricator X were affected most by the COs, which could explain why their productivity and schedule performance suffered more than for other project participants. However, missing from the visual representations, but addressed for Project 2 is the link between schedule performance and change order occurrence, information that is crucial to strengthening the argument about CO impact.

In summary, the two visual representations provided insights about how change orders were distributed in time, space and by project participant, which in turn could assist (and did) the client, contractor, and those adjudicating the dispute resolve differences of opinion about the impact of change orders on project performance. The same benefits were not derived from examining the spreadsheet of change order data no matter how sorted by those assisting the project's contractor. The individuals involved did not attempt to forge visual images of the contents of the roster of change orders in order to comprehend how they were clustered in terms of one or more of the project's context dimensions. Instead, they simply relied on presenting a listing of change orders. We have witnessed first hand similar approaches in practice, and in fact encountered such for Project 2, with these practices being an impediment to telling the construction story in a readily comprehensible manner. With respect to Design 1, Project 1, unfortunately, we are unable to compare and contrast the design of our visual representations with those proposed by others due to the lack of alternative designs being documented in the literature. However, our own critical evaluation of the images led to improvements in the design of the visual representations for Project 2.

*Components level* – Evaluation at this level is done by checking the choices of data representations and transformations, visual representations, and interaction features. In the design of Figures 2.1 and 2.2, change orders were represented by the context dimensions of *time* when change orders were active, *trades responsible* for executing change orders, and *locations* where the change orders were executed along with the performance measurement

dimensions of *number* and *dollar values* of change orders. This data representation consists of the information and knowledge fundamental to identify trends of change orders versus time and the clustering of change orders in time, space, and by participant. The original data were then transformed by abstracting locations into three categories (on-site, off-site, and both on-site and off-site) and representing time by months, a more aggregated level of detail than by individual days. The former one is essential for management to identify site congestion issues if change orders were executed on-site in clusters; the latter one is essential to add clarity of visual representations and echo industry's practice of processing and monitoring change orders in a longer time interval. Other data transformation such as deriving performance dimensions of change order percentage (cost of one change order/cost of all change orders or cost of one change order/cost of original related work) could provide more insights and should be considered. As to the choices of visual representations, Design 1 (Figures 2.1 and 2.2) utilized three dimensional visualization space in order to maximize the use of spatial positions to encode multi-dimensional data. Lastly, because the primary purpose of Design 1 is to provide management with an overview of the entire distribution of change orders, interaction features supporting further data exploration were not considered essential to the analytical reasoning purpose of this design. However, interaction features for enhancing data readability like "details on demand" and "navigating visual representations" could be helpful for comprehensively examining both the details and profiles contained in Figures 2.1 and 2.2.

#### **2.5.2.5 Lessons learned – project 1**

Users may have preferences for adopting different visual representations of basically the same format. For example, instead of concatenating CO\_ID and location together as was done in Figure 2.1, one could also concatenate time and location together. It is left to the user to determine which visual representation best suits their cognitive abilities, but the main message is that the design of a visual analytics environment must allow the user to experiment with different representations. From the practical perspective of construction users, what this means is that a relatively large range of representations needs to be pre-coded, along with some guidance as to the advantages of each for analytic reasoning.

Further, considerable care must be taken in choosing the shape of visual objects used to represent context entities or performance dimensions in order not to create misleading or false insights on the part of the user.

### **2.5.3 Visual representations for project 2**

Having experimented with different visual formats to represent aspects of the change order data set for Project 1, we explored a broader range of images for a more extensive change order dataset for a complex rehabilitation project, Project 2. The sheer volume of the extra work orders generated (531) during the first 2/3 of the project duration and their occurrence frequency made change order management on this project a challenging task (the construction manager providing the data used the words extras, extra work order and change order as synonyms). These 531 change orders correspond to 560 subtrade involvements – i.e. if three subtrades are involved in a single CO the actual contribution to 560 is three. For the total project, slightly more than 750 change orders were generated. The issue confronting both the construction manager and client on this project was one of communication between the two as to the reasons for the large number and attendant cost of change orders. Printouts of the construction manager's change order spreadsheet provided to the client did not resolve the communication problem. A variation of the visual representation presented in Figure 2.3, developed as part of our interaction with the CM firm, assisted in clarifying the change order story of the project, especially with respect to the origins of the change orders.

A reality of current industry practice is that data sets for a number of functions are invariably incomplete, either because only a subset of the properties defined for the item of interest are recorded, and/or an incomplete set of properties have been defined. Since a primary focus of project management staff is to maintain momentum on the job, keeping and updating records in a comprehensive manner often takes a backseat. Thus, data records for many of the change orders generated on this project were found to have certain missing properties in terms of trades affected, issue date and/or date of approval, when the work was actually completed, dollar consequences for each of the affected trades, etc.

Though our work is focused mainly on visualization of datasets, the usefulness of visualization is dependent on the completeness of the data set. Hence considerable effort was expended in trying to obtain as complete a data set as possible. To do so, we made use of relevant and associated documents like the contract register, site instruction (SI) and request for information (RFI) lists, we reviewed individual SIs and RFIs, and through discussions with on and off-site management personnel, we tried to track the missing links in the data. This allowed us to cluster data items using different attributes such as location of the work, physical system affected, trades involved, and turnaround times, thereby yielding more insightful visual representations, which proved to be beneficial to the CM when communicating with the client. We were able to accomplish this because of the direct access provided to the site, site records and management staff. Moreover, the staff members were enthusiastic in offering their comments and providing us with prompt additional information as and when required, and finally, senior management was motivated to use findings of the work as appropriate to enhance communication with the client.

A total of 3 different visual representation designs were generated, corresponding to Figures 2.3 through 2.7. In the discussion that follows, observations are made about the specific features of these designs. A detailed critique of Figures 2.3 and 2.4 is summarized in Table 2.2 to show the kind of evaluation procedure that should be conducted as an integral part of the design process.

#### **2.5.3.1 Visual representations for project 2 and design 1, Figures 2.3 - 2.5**

Figure 2.3 represents the distribution and reasons for the change orders, and is particularly useful for communicating with the client while also providing valuable insights on how the project is evolving. This figure conveys the distribution of changes over time, trades affected, and primary reason for the change. To enable the user to identify trends in the datasets and thus obtain additional valuable insights, cumulative totals for all change orders versus primary reason for change integrated over time and for all change orders versus time integrated over reason for change are presented as an option on the side and

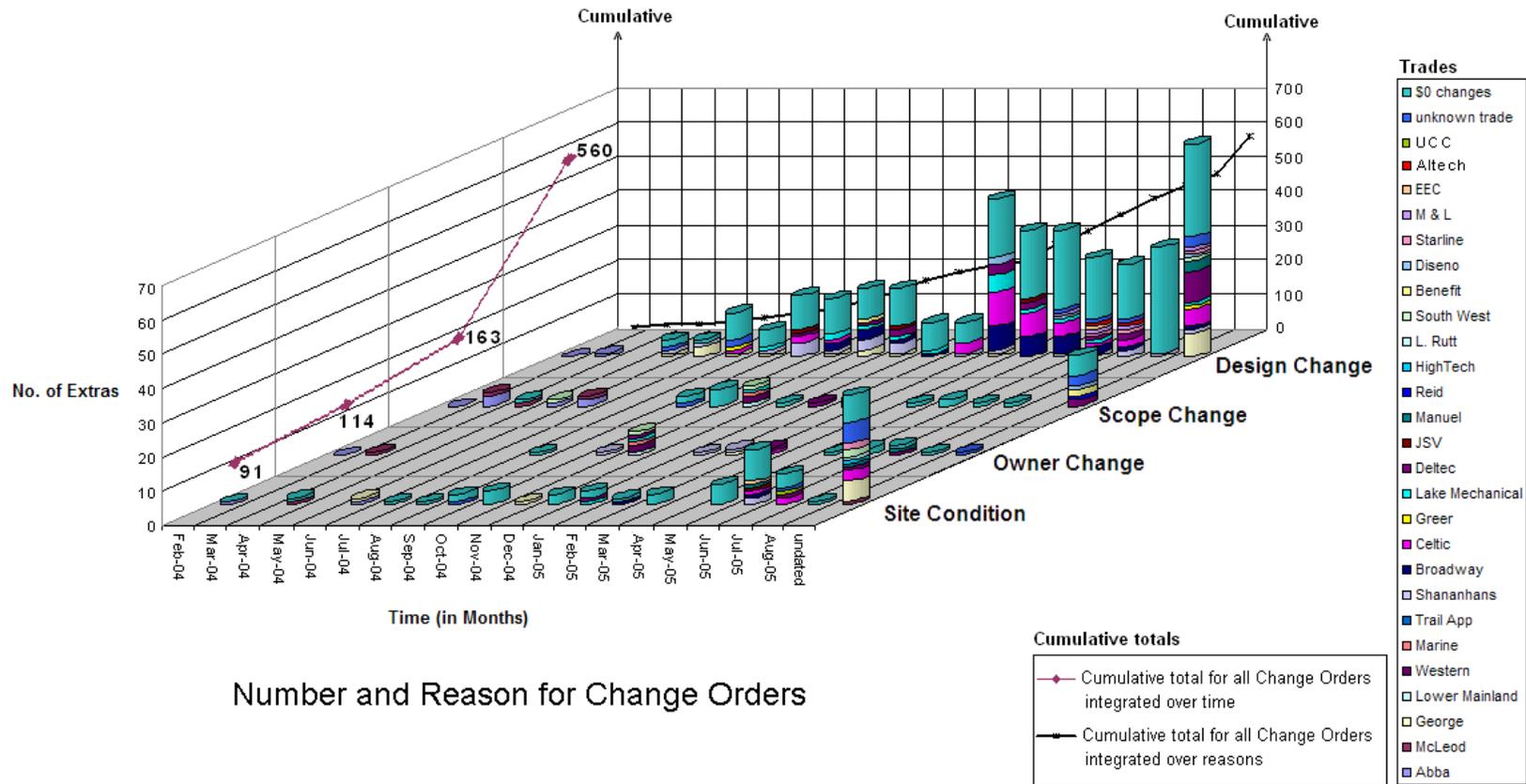


Figure 2.3 Project 2 Number and reasons for change orders

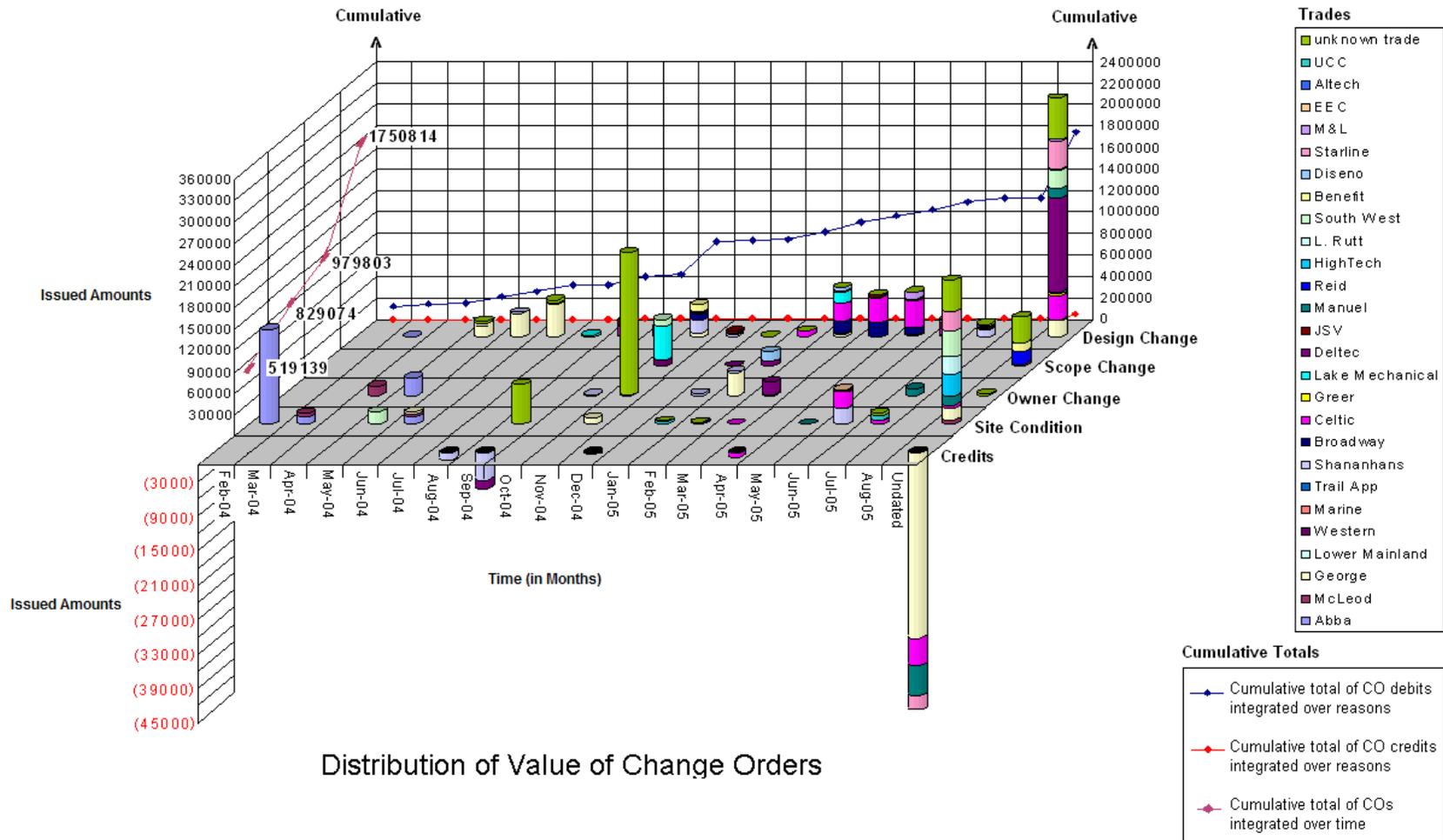
back panels of the chart, respectively (an example of how additional information can be incorporated into the visualization space through an interaction feature). The X-axis represents time in months when a change order was issued. A more fine-grained representation of time did not add value. The right most section on this axis flags time as ‘undated’. The COs included in this section are the ones for which the issue date could not be identified. As noted previously, datasets are invariably incomplete, and thus mechanisms to treat incomplete data have to be incorporated into the design of visual images. How best to do this is not always clear and hence more exploration on this issue and related ones (e.g. zero value COs and COs involving multiple trades, see below) is needed. The Y-axis divides the entire graph into 4 separate zones depending upon the reasons for the issued COs. As described, later, every change order in the datasheet was eventually allocated to a single primary reason for issuance. The vertical axis (Z-axis) which corresponds to the performance measure or variable of interest, represents the total number of COs affecting different trades issued in that particular month as in the case of the previous image. The majority of change orders involved the work of a single trade. Nevertheless, for some COs, two or more trades were involved. In such cases, for accurate representation in Figure 2.3 when the breakdown by trade is also treated, a CO will be ‘double or triple counted’ for the month in which it was issued (hence the 560 count in Figure 2.3). We observe that if the facility to generate an image like that shown in Figure 2.3 was to be incorporated into construction management software, then the option to include a breakdown by trade should be included, and a footnote automatically included in regard to the counting issue. On the other hand, if the breakdown by trade was not chosen as an option, then the correct count of change orders would be shown on the figure.

Of the total number of change orders generated on this project, a significant number were issued as a result of design changes. A large fraction of these were found to be zero dollar changes i.e. change orders having no dollar consequences. In generating this figure, \$0 change orders have been coloured as though they belong to a trade, in this case \$0 trade (see color legend in Figure 2.3). From a work monitoring perspective, such changes would still have to be tracked on a trade-by-trade basis, but for keeping count of all COs issued, it

was deemed acceptable to treat under a \$0 trade designation. This particular case is mentioned as it highlights the kind of situations often encountered when attempting to represent data in a visual format. The need exists, however, to explore other ways of treating such situations in order to present as objective a view of data as possible.

Figure 2.3 was developed based on refinements to the CM's dataset. In the original dataset, the construction manager used a suite of six reasons and allowed for a many to one relationship – i.e. many reasons to one extra. Some of these reasons overlapped to a certain extent, creating considerable ambiguity in interpreting the data and communication challenges with the client. Upon seeing a first draft of the figure, management personnel realized they needed to adopt a less ambiguous set of reasons, which led to the use of the 4 reasons shown and a one-to-one relationship between a change order and the primary reason for it. The CM revised the dataset, which provided the basis for Figure 2.3. The foregoing observations speak to the challenges of having data accurately, unambiguously and completely collected while it is current, a non-trivial task given the preoccupation of management to maintain momentum on the job.

Figure 2.4 looks at the distribution of the value of change orders, and assists with client communications while providing useful insights on budget matters. This image is very similar to Figure 2.3, the only difference being that the Z-axis now corresponds to dollar amount instead of number of change orders. The cumulative total of the dollar amount of COs integrated over time for each primary reason for change orders is shown on the side panel while the back panel has two separate line graphs for cumulative total of CO debits and CO credits vs. time integrated over reasons for change. It is observed that the number of COs is not necessarily proportionate to the dollar consequences of change orders. There can be situations where a large number of COs generated in a month totals to an insignificant amount whereas in other cases a single CO may cost a very significant amount (as discussed later, such observations provide powerful motivation for being able to create and navigate scenes comprised of multiple visual representations). Management staff therefore faces a two fold challenge of managing the flow of change orders and



Distribution of Value of Change Orders

Figure 2.4 Project 2 Distribution of value and reasons for change orders

Table 2.2 Summary of visual representation evaluations for Figures 2.3 and 2.4, project 2

EVALUATION ITEMS	CRITIQUE ANALYSIS FOR EVALUATION
<p align="center"><b>Overall Evaluation</b></p>	<ul style="list-style-type: none"> <li>• <b>Strength</b> The visual representations in Figures 2.3 &amp; 2.4 provide clear visualizations for:               <ol style="list-style-type: none"> <li>1. Showing the existence of data patterns that may help to identify potential root causes or impacts of change orders.</li> <li>2. Showing an overview of characteristics of change orders for monitoring a project from a CO management perspective.</li> <li>3. Communicating with the client.</li> </ol> </li> <li>• <b>Weakness</b> The visual representations don't convey the full range of insights possible. For example, both Figures 2.3 &amp; 2.4 do not provide information of how change orders are distributed by sub-trades, e.g., identifying ranking of trades by number or dollar value of CO. Another example is that this design is not able to present insights that can only be gleaned from a subset of the data such as dollar value exceeding a certain threshold.</li> </ul>
<p align="center"><b>Components Evaluation (Data Representations and Data Transformation)</b></p>	<ul style="list-style-type: none"> <li>• <b>Strength</b> <ol style="list-style-type: none"> <li>1. Representing COs by the contexts of reasons for change, time, &amp; responsibility and the performance measurements of # and \$ values of COs provides the information and knowledge essential to i) identify root causes or impacts of COs and ii) understand selected properties of COs.</li> <li>2. Transforming CO data to different levels of abstraction or granularity can increase the clarity of a visual representation, which is essential to ensuring usefulness to the intended industry audience.</li> <li>3. Transforming CO data by aggregating COs in counts or dollar values based on various data query conditions, which is essential for observing the distribution of COs in various context dimensions. This method matches the current state-of-art concept of OLAP and data cube.</li> </ol> </li> <li>• <b>Weakness</b> The aggregation is not exhaustive, and thus some insights may be missing – e.g. the design did not aggregate number of COs by COs that are of the same sub-trade. Another example is that this design did not aggregate number of COs if we query a subset of CO data with the filtering condition of dollar value being over a certain amount of money. However, this can be remedied by providing interaction features for users to choose level of aggregation on demand.</li> </ul>

EVALUATION ITEMS	CRITIQUE ANALYSIS FOR EVALUATION
<p style="text-align: center;"><b>Components Evaluation (Visual Representations and Interaction Features)</b></p>	<ul style="list-style-type: none"> <li>• <b>Strength</b> Compact as much information as possible into fewer and clear images for quick scanning by users.</li> <li>• <b>Weakness</b> <ol style="list-style-type: none"> <li>1. The visual encodings used would be undesirable if interaction features are not supported (e.g. use color hue to represent many sub-trades and bar length to represent breakdown of COs by trade).</li> <li>2. Lacks interaction features for:               <ul style="list-style-type: none"> <li>○ Enhancing image readability- use of 3D visualization space requires view navigation to find an optimum scale and angle of 3D chart so that occlusion is minimized. Brushing technique also can be used to alleviate issues of occlusions and ineffective color coding.</li> <li>○ Querying data- visual analytics in essence is querying data that is presented in visual forms. Therefore, basic data query abilities such as filtering data value ranges, sorting/grouping data values, and simple data transformation (e.g. data aggregation) are a must.</li> <li>○ Choosing visual representations- different users have different visual perception preferences or cognitive styles. This difference could be a factor affecting the effectiveness of analytical reasoning. The interaction feature of changing visual representations on users' demand should be supported – e.g. users should be able to change from 3D charts in Figures 2.3 or 2.4 to 2D charts (e.g. Figure 2.5).</li> <li>○ Coordinating views- when observing Figure 2.4, users may become interested in COs having dollar values that are over a certain threshold, and want to know whether this subset of COs cluster in time and/or reasons for change, which could be observed in Figure 2.3. This can be done by directly selecting visual marks in Figure 2.4 as an instruction of filtering data, and then Figure 2.3 would highlight visual marks representing data that are only related to the data selection in Figure 2.4.</li> </ul> </li> </ol> </li> </ul>

observing the cost of change orders as they affect the overall project cost. Thus Figure 2.3 helps management assess the effect of distribution of changes by number as they affect the targeted project completion time while Figure 2.4 helps assess the effect of cost of change orders by value of work on the overall budget. Since for the latter case one is dealing with the dollar consequences of COs on different trades, the issue of double counting of COs does not exist. Another observation is that some of the change orders actually generate credits. In order to identify these credits with greater ease they have been allotted a separate zone at the forefront in the image. Again, the need exists to explore other alternatives of displaying such information. One important message from both Figures 2.3 and 2.4 is that incomplete information can result in the inability to derive completely accurate insights. Without being able to properly distribute the number and value of change orders in time (the undated missing data problem), especially when the numbers involved are significant, the potential impact of the cumulative effect of COs may not be properly gauged. By portraying the data in the way we have chosen, this problem is highlighted, and could provide the incentive needed to search out the data required and/or being more diligent in recording essential data.

Figure 2.5 is a 2D stacked graph presenting information similar to the content of Figures 2.3 and 2.4. As noted earlier, different users have different preferences and capabilities for visualizing data, especially when it comes to 3-D representations. Hence it becomes necessary to develop alternate formats for the same data. Figure 2.5 represents all of the information from Figures 2.3 and 2.4 in a single representation consisting of stacked graphs with time as a common context dimension. This figure can be read in two parts. The top part of the graph is a scatter plot representing the total number of COs issued each month over the project execution phase. The pie charts in the graph are comprised of an inner circle that corresponds to the reasons for initiating these COs while the outer ring depicts the fraction of the number of COs affecting individual trades. For this figure, the X-axis indicates the time when change orders were issued and the vertical Y-axis indicates the total number of change orders issued. One important advantage of this graph is that COs associated with multiple trades are not double counted, as is the case in Figure 2.3. In

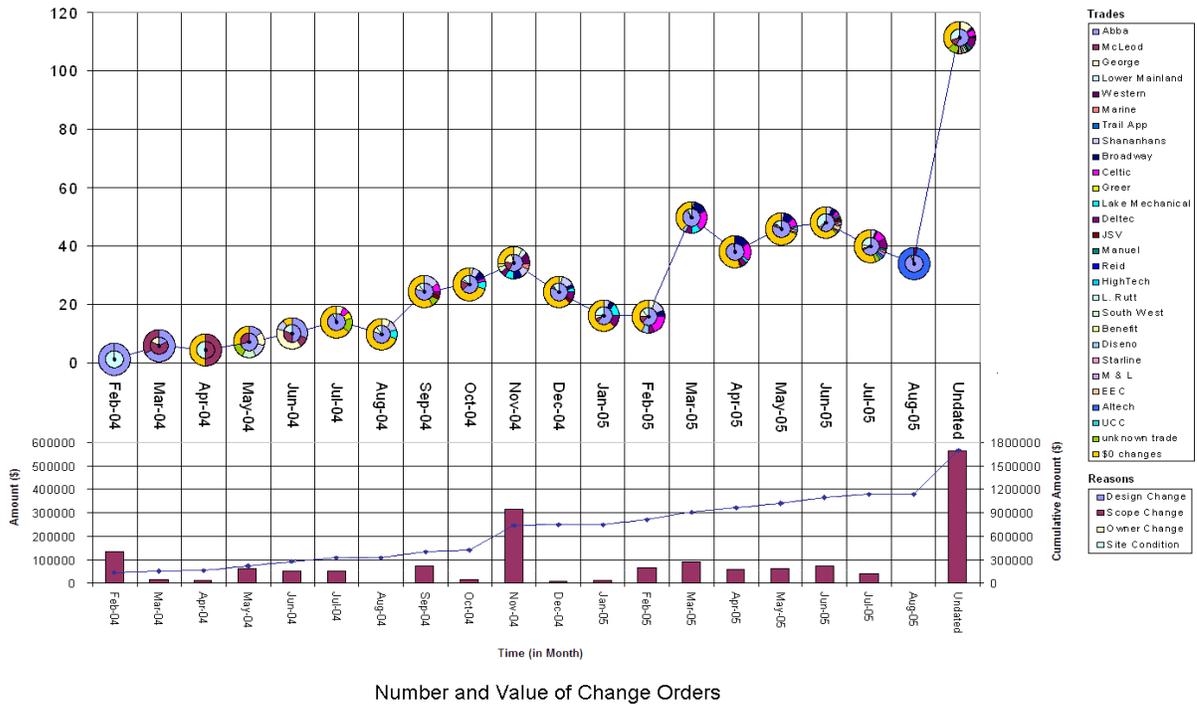


Figure 2.5 Project 2 Stacked graphs for number, values and reasons for change order

the bottom half of Figure 2.5, the total dollar amount of the COs issued each month is shown. Also shown on this graph is the cumulative dollar amount of COs issued to date. Figure 2.5 thus enables the user to determine the number of change orders generated, corresponding trades affected and the subsequent dollar amount in one go. However this graph does not show the division of dollar amount by trade as per Figure 2.4. This could be achieved, however, in the bottom half of Figure 2.5. From our experience in dealing with construction personnel, we venture the opinion that Figure 2.5 is probably preferred to the images shown in Figures 2.3 and 2.4 simply because of their greater familiarity with 2D project representations (e.g. drawings, sketches, etc.). As 3D representations start to permeate the industry with the adoption of Building Information Modeling (BIM), 3D representations of construction management data are likely to receive greater acceptance.

Figures 2.3 through 2.5 also highlight the challenges involved in trying to represent as much information as possible or too much information on the same image. For example, for Figures 2.3 and 2.4, by including a breakdown of number and value of COs by trade,

accurate counts for each are difficult to discern, especially when small numbers are involved. Further when many organizations are involved (for the case at hand 28 trades, including the \$0 ‘trade’), the use of colour to distinguish between organizations breaks down – one simply runs out of a sufficient number of distinct and easily identifiable colours. This problem would only be exacerbated for much larger projects, when many more organizations are involved. Thus there are practical limits on how much information can be depicted on one image, even when supported by an array of user interaction features. Such challenges provide in part the motivation for examining data through coordinated data views, in which overview data can be portrayed along with supporting details (e.g. number of COs in each month, and then breakdown by trade and reason for change).

### **2.5.3.2 Visual representations for project 2 – design 2, Figure 2.6**

Figure 2.6 examines the distribution of change orders by physical system and time, and thus helps identify clustering of work and potentially speaks to the quality of design documents issued by the various professional disciplines involved. In this case the vertical Z-axis represents the total number of change orders generated, the X-axis indicates the time in months when the change orders were issued and the other horizontal Y-axis represents the physical systems affected. These physical systems are further grouped under different ‘Major elements’ (e.g. Substructure, Shells, Interiors, Services) along the Y-axis. A single cell in the graph represents the number of change orders issued in a particular month affecting a particular physical system. For example, a total of 8 change orders were generated in the month of May-05 affecting the Exterior closure which forms a part of the group ‘Shells’. In some cases a single change order is found to affect multiple physical systems. In such cases the CO gets ‘double or triple counted’ for that month and that Main Element group. Thus the number of change orders affecting different physical systems of a group do not necessarily add up to the total number of change orders affecting that group. In the form shown, the use of colour does not add value. However, if it was desired to show additional information like the reasons for change, the use of color coding would be beneficial.



### **2.5.3.3 Visual representations for project 2 and design 3, Figure 2.7**

The analytical reasoning purpose behind the visual representation shown in Figure 2.7 is to explore the potential existence of a causal relationship between number and timing of change orders and schedule performance. In generating this representation, use has been made of the first 402 change orders encountered. This 3D representation deals with the trajectory of forecast project completion time versus the cumulative effect of number of change orders with time (with the underlying causal model being that the greater the number of changes, the more the potential for an extended project duration). Note that number of changes, the Z axis, is used as the surrogate measure here, not value. For quick reference, the cumulative total of the change orders considered is also displayed on the back panel of the graph. To generate this representation, use was made of the sequence of project schedules generated by the CM (ready access to this data in the form of update date and projected completion date speaks to the advantage of having an integrated, multi-view data representation of a project, which was not a feature of the CM's data). Across the horizontal axis (X-axis) is time, which serves two purposes: (i) to indicate the months when change or extra work was identified; and, (ii) to represent the dates of schedule update, starting with the original schedule before work started all the way to the last update observed by the research team. On the other horizontal axis (Y-axis) are listed the months when the project was forecast to be completed, with the dates of project completion reflecting the update version on the X-axis. The change order work is stretched out over these months of completion, to indicate how many more changes have occurred since the last update and projected completion date. The red line reflects the trajectory of movement of the forecast completion date. This visual representation portrays that a relation appears to exist between the number of changes occurring over time and the change in the projected completion date. However it would not be fair to state that all the movement in the projected completion date is solely due to number of changes (or for that matter value of changes if used instead of number) since there might be several other factors impacting the completion date (e.g. weather, labour shortages, etc.). We have, however, limited our scope to assessing the impact of changes on the project performance outcome as measured by project duration. While not straightforward to generate, this visual representation is a

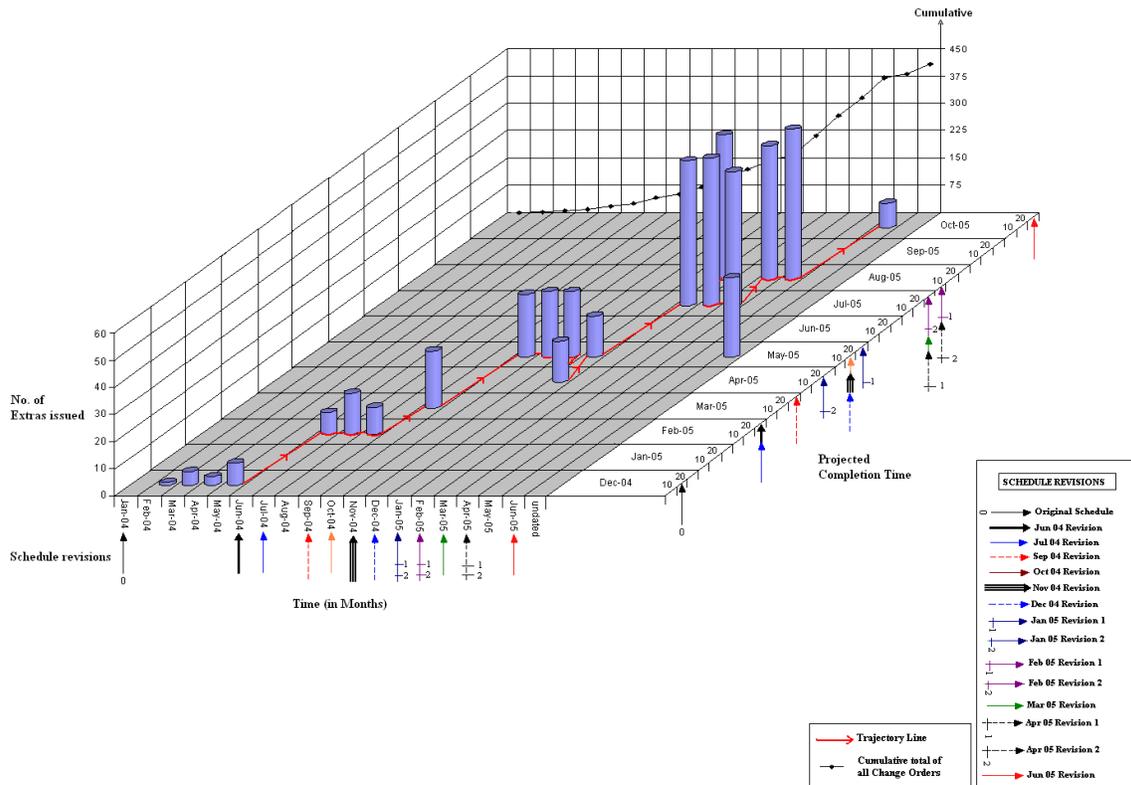


Figure 2.7 Project 2 Causal model reasoning – number of COs and corresponding schedule update dates and projected completion dates

reasonably compelling one, and not only helps with identifying cause-effect relationships, but assists greatly in communication with the client. The main point here, however, is that carefully designed visual images that juxtapose data from two or more project data views can offer assistance in exploring potential causal relationships between project context dimensions and project performance dimensions.

## 2.6 Some general observations

In this section we discuss a number of issues relevant to the development of a general visual analytics model for CM functions.

### 2.6.1 Applying identified principles for design process

As a prelude to pursuing the application of visual analytics to a CM function, the first step is to determine the analytical reasoning tasks that could benefit from the visual

representation of associated data. Project context dimensions and performance dimensions involved in these reasoning tasks then need to be identified. Effective visual encodings of X, Y positions and colors can be utilized to map non-quantitative context dimension while Z positions can be used to map quantitative performance dimensions (i.e.: # or \$ values of COs). The use of conventional graphics elements such as orthogonal coordinates, bars, and lines are generally sufficient to accommodate CM users from a broad range of educational and experience backgrounds. Trying to represent multi-dimensional data sets in a 2D or 3D space is difficult while maximizing image information content. Compactness is viewed as a virtue so that management can develop a holistic view (overview and details) as quickly as possible, without the requirement to navigate through multiple images. Thus, the design of visual images involves a great deal of iterative design and self evaluation using both hand drawn sketches and a variety of software tools in order to formulate visual representations in terms of their dimensionality (two-dimensional or three-dimensional), scale, viewing angles and colour-coding in order to maximize both the information content of each image and the insights that can be extracted. Issues like occlusion for 3D graphics and too many colours for effective color coding might be alleviated to some extent through interaction features (e.g. linking brushing, view navigation).

### **2.6.2 Evaluation and feedback**

Included as part of the design and evaluation process for Project 2 was exposing earlier versions of the images to a group of construction personnel including senior management, and incorporating feedback received (interestingly, personnel normally worked with large spreadsheets or other tabulations of data, and had not explored on their own how data visualization could assist them in their management tasks). One non-definitive observation of the reaction by construction personnel was that the notion of image compactness can lead to information overload, and the use of multiple images as opposed to a single image to convey the insights involved may be a better choice. However, the overwhelming reaction and positive feedback by management staff that Figure 2.3 would go a long way to having the client understand the change order story for this difficult project, a

preoccupation of management at that time, outweighed any downside of too much information on a single image.

Apart from the industry evaluation, our self evaluation also identified a number of merits of the images designed by following the principles of the visual analytics design process described previously. A consensus of the evaluation results is that an overall qualitative understanding of change order characteristics and impacts on project performance dimensions (e.g. the majority of change orders are design changes as seen in Figure 2.3, and changes orders related to contract issues increased as the project progressed as seen in Figure 2.6) can be perceived by glancing at those images for only a few seconds per image, particularly when Figures 2.3, 2.4, and 2.6 are placed closely together. We believe that such a quick and rich understanding could further trigger the tacit knowledge of project participants as to the impact of change orders on different project performance dimensions, thus leading to deeper insights.

### **2.6.3 Organizing lessons learned for development of a general CM visual analytics model**

Based on the design/evaluation work we have done to date, we have identified three general structures of visual representation designs and a suite of interaction features that are tailored to CM use and that can be readily extended to other CM functions. However, some design details still need to be tailored to the unique analytical reasoning needs associated with specific CM functions.

Lesson 1: General structure for visual representations of CM data

- Scene structure 1 - visualizing characteristics of construction conditions:  
A 3D scene of several charts could be generated to present the characteristics of construction conditions observed from different project context dimensions. Each chart uses an X axis and/or Y axis, and/or color coding to represent three non-quantitative context dimensions (e.g. process, product, organization, etc.) and the Z-axis to represent a quantitative attribute dimension representing a construction condition (e.g. product

quantity, resource usage, problems encountered, etc.). The side panel and back panel design are used to visualize aggregated data values of the condition similar to the use in Figure 2.3. Different charts represent conditions observed from different combinations of context dimensions if the investigated condition associates with more than three project context dimensions (e.g. time vs. space vs. trade, time vs. trades vs. activity, etc.).

- Scene structure 2- visualizing characteristics of construction performance:

A scene structure that is similar to the one for visualizing characteristics of a construction condition can be used for performance dimensions (i.e. the Z-axis is used to represent performance dimensions such as number of deficiencies, time variances, cost variances, etc.).

- Scene structure 3: visualizing potential cause-effect amongst construction conditions and performance dimensions:

Scenes juxtaposing or overlaying charts of construction conditions with charts of construction performance (similar to Figure 2.7) can be very useful for exploring hypotheses as to reasons for performance. Also, one should be able to juxtapose or overlay charts of construction conditions with construction conditions or construction performance with construction performance. For this type of scene design, the “floor/wall” of the virtual 3D space could be flexibly used to position charts of construction performance/ conditions.

## Lesson 2: A suite of interaction features

In the near future, the visual representations developed should be coupled with interactive features like ‘zooming and filtering’, ‘details-on-demand windows’ or setting ‘dynamic query fields’, thus greatly enhancing the potential for analytic reasoning. For example, a simple click on a particular CO in Figure 2.2 would pop up a ‘detail-on-demand window’ listing all the required details of the specific data item, in this case CO properties (trade name, the month of interest and the Number of COs associated with the trade) selected from the list in Table 2.1 and contained in a user defined content profile. Further, by introducing filtering techniques, users would have the flexibility to view only data of current interest. For example, if a user prefers to obtain the distribution of extras only by

number and trade, with the use of appropriate filter options one should be able to generate the required image which would represent a subset of the content in Figures 2.3 through 2.5. Such selection and filtering capabilities would help management absorb the content of images faster and improve the quality of insights obtained, allow users to adjust image content to reflect their own cognitive style, help pinpoint specific issues and assist with decision making directed at resolving existing or emerging problems. The range of interactions features that should be incorporated are identified in the critique of Figures 2.3 and 2.4 contained in Table 2.2.

#### **2.6.4 Issue of CM data management**

With respect to the data itself, during the process of designing a visual analytics model for change order diagnosis, we identified two major issues regarding current industry practice of data management: missing data values, and incomplete and dissociated data representations. Both of these issues speak to the importance of good data management for generating useful visual representations of data for assessing and communicating performance and related issues.

The problem of missing data values was observed and described in the previous section with respect to one or more properties of individual change orders. Although we addressed this problem through a combination of searching through project records, discussing items with management personnel and in some cases assigning default values to some properties (e.g. assigning 'undated' status to COs with missing date values), unless accurate recording of properties is achieved the actual patterns of data in practice could be quite different from what would be visualized using incomplete data. For example, as stated previously, the patterns shown in Figures 2.3 and 2.4 would no doubt be changed somewhat if all of the undated COs were positioned when they actually occurred.

The problem of incomplete and dissociated data representations was also encountered with industry practice, either as reflected in the commercial software applications used or internally generated spreadsheets. As a result, data fields and data association simply do

not exist with which to record several properties, including the association of a CO with the context dimensions of space and physical systems/components, which would help provide useful insights on potential causal relationships between context dimensions and performance dimensions. The reality is that management personnel are focused on maintaining project momentum and are often stretched to capacity, leading to only partially populating the predefined properties of different project records (e.g. COs, drawings, RFIs, etc.), with little consideration given to properties not explicitly defined. Fundamental to persuading personnel to collect additional information is the ability to demonstrate that the benefits significantly outweigh the costs, a proposition that in most cases is not easy to prove.

#### **2.6.5 Data exploration flexibility**

Currently, many commercial data visualization systems are available for supporting generic visual data analysis. At a first glance, it seems that they are sufficient with which to explore integrated CM databases. However, based on our experience using these generic tools, we found that even with the ease of use facilitated by their state-of-art interaction and data query capabilities, users could still spend much of their time examining what data is available, deciding which data items can be useful for being visualized, and determining how best to visualize them. Although these visualization environments provided very flexible interactive features (e.g. iteratively changing data query conditions and visual representations on user demand) thereby increasing the potential to detect interesting visual patterns representing unexpected phenomena hidden in data, the significant amount of time required for exploring data in this type of environment should not be underestimated. However, on the other hand, if the visual analytics environment imposes a strict analytic scenario and forces users to follow steps of viewing only certain images, the rigidity may limit the usefulness of visual analytics. How to strike a balance between these two extremes and optimize the level of data exploration flexibility when designing a CM visual analytics environment is a topic that needs further work.

## **2.7 Conclusions**

Visual analytics, the science of analytical reasoning facilitated by interactive visual interfaces, has the potential to improve the construction management process through the enhanced understanding of project status and reasons for it, better informed decision making, and improved communication amongst project participants. To date, while some useful exploratory work on data visualization has been carried out by a few researchers, no significant body of work exists on the application of visual analytics to the discipline of construction, despite successes in other disciplines. An approach for developing such a body of work for construction has been outlined in this paper. Of the four pillars of visual analytics, namely the purpose(s) of the analytical reasoning, the choices of data representations and transformations, the choices of visual representations and interaction technologies, and the production, presentation and dissemination of visual analytics findings, the focus herein has been on choices of visual representations. General principles to guide the design of visual representations useful for construction management processes have been identified, with emphasis on the two primary purposes served by analytical reasoning – i.e. assessing and communicating. In terms of assessing performance, visual analytics can assist with predicting the future based on lessons learned to date, examining the past in order to better understand the as-built situation, comparing performance, and identifying potential causal relationships. Other advantages offered by visual analytics include the ability to work on a more factual as opposed to perception (feeling) driven diagnosis of reasons for performance to date, and the quickness, versatility and relative ease with which data can be represented and interpreted without the need for specialist assistance. As part of the general principles identified, the notion of context dimensions vs. performance dimensions was introduced, which is of direct assistance in formulating visual representation designs. To demonstrate the application of the concepts presented, an in-depth examination of how visual analytics can assist with change order management was described. Data sets from two different projects were used to demonstrate the design and practical data collection challenges involved in formulating visual representations that are useful for analytical reasoning. A detailed assessment of two of the images was presented, both in terms of strengths and weaknesses, and interaction features desired were

highlighted. While somewhat obvious, it is important to use large scale datasets when designing and testing visual representations, as significant challenges exist with respect to scale in terms of the context dimensions of time, space, responsibility, physical components, process entities and work environment. It is believed that the lessons learned are readily extendable to other construction management functions, including the need to examine the use of coordinated data views as opposed to maximizing the compactness of an image in terms of providing both overview and detailed information in a single image, despite the desirability of doing so for a construction audience that is action driven.

In the near term, our focus will be on exploring visual analytics models for quality and risk management to demonstrate broad applicability of the approach and supporting principles. As part of this work, including previous work on change order management, comparisons of the utility of compact visual representations vs. coordinated data views will be made. Attention will also be directed on the design of visual representations for assisting in formulating and determining the validity of hypotheses for explaining construction performance (e.g. productivity, delays) – i.e. visual causal model reasoning. The most promising of the foregoing visual representations and accompanying interaction features will be implemented using state-of-the-art visualization tools and field-tested on actual projects. Feedback from such tests is essential in order to ensure the usefulness of the representations and their responsiveness to the practicalities and constraints of the industry. Our ultimate goal is to contribute to the design of a visual analytics environment that is attuned to the needs and attributes of construction managers. We believe that this environment should include a palette of pre-coded images and related interaction features.

## **Chapter 3 Design of a Construction Management Data Visualization Environment: a Top-Down Approach<sup>4</sup>**

### **3.1 Introduction**

The primary focus of this chapter is on the use of a top-down approach coupled with the extension of thought processes, principles and guidelines previously described (Russell et al. 2009b) for the design and development of a data visualization environment for construction management. For most projects, the large volume of data generated while executing a diverse set of CM functions poses significant challenges to their constructors, particularly in regard to generating insights and deducing cause-effect relations in a timely manner in support of decision making. Data visualization can play a pivotal role in addressing these challenges and improving project performance in terms of “cost and profit, time, scope, quality, safety and regulatory compliance” (Russell and Udaipurwala 2004). It deals with the effective portrayal of construction data to generate insights about the data and to unveil the undiscovered useful information embedded in it (Keim 1996).

Visualization of construction data offers several benefits. These include:

- identifying and communicating interdependent relationships across various data items thus enhancing the ability of the construction team to interpret data and improve decision making (Liston et al. 2000);
- amplifying cognition of quantitative data;
- improving and verifying the completeness and accuracy of data;
- reducing the time spent in comprehending and explaining information;
- providing managers with information rich overviews about the status of various project components;
- avoiding misconceptions due to inadequacies in data sets;
- explaining the divergence and disparity between the planned and as-built stories (Pilgrim et al. 2000; Shaaban et al. 2001; Songer et al. 2004); and,

---

<sup>4</sup> A version of Chapter 3 has been published. Chiu, Chao-Ying., and Russell, Alan D. (2011). "Design of a Construction Management Data Visualization Environment: A top-down Approach." *Automation in Construction*, 20 (4), 399-417.

- assessing the quality of a construction schedule (Russell and Udaipurwala 2000a).

The term visualization is defined by Card et al. (Card et al. 1999) as “the use of computer supported, interactive, visual representations of data to amplify cognition”, and as “the act or process of interpreting in visual terms or of putting into visible form” (Nielsen and Erdogan 2007). Card et al. (Card et al. 1999) has identified the need to support three types of interaction for information visualization in order to be able to modify: (a) data transformation; (b) data mapping between the data and its visual representation; and, (c) view transformation (navigation) for navigating through the visual representation.

In using the term *data visualization environment*, which we consider herein as being synonymous with a visual analytics environment, we mean a computerized information system which enables users to create their own scenes composed of one or more pre-coded visual representations and an interface that assists users to interact (filter, sort, zoom, highlight & coordinate, etc.) with data and a palette of pre-coded images designed to facilitate analytic reasoning. Distinguishing features of a holistic data visualization environment from that of data visualization capabilities of current state-of-the-art CM systems revolves in large part around two issues. First is support for dynamic versus static images, specifically the degree to which the user can interact in real time with an image or collection of images across the temporal, spatial, organizational, product and process dimensions of a project. And, second is the breadth of functions treated and associated analytical reasoning tasks supported, within the confines of individual tasks for a function, across multiple tasks within a function and across multiple functions.

Addressed in Figure 3.1 are several aspects of the foregoing issues for a subset of CM functions and associated tasks. Depicted is an approximate subjective assessment of the current state-of-the-art in terms of functionality versus what could be achieved in a holistic data visualization environment (the triangular / trapezoidal shapes reflect intensity (height) and breadth (length) of support offered). Significant opportunities exist for enhancing the insights that can be generated from the speedy exploration of large information spaces in

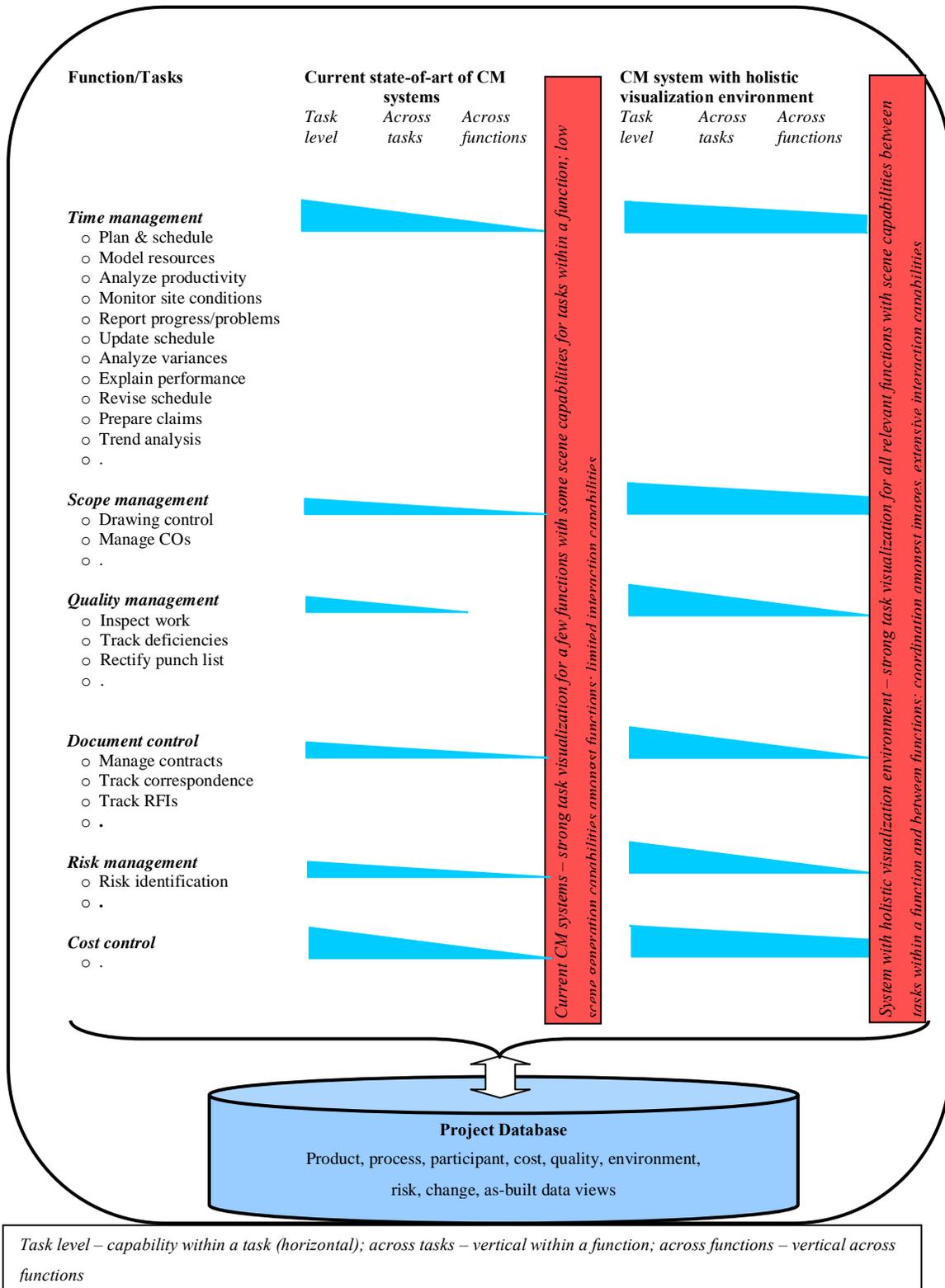


Figure 3.1 Differentiating between current state-of-art CM systems and potential of systems with a formal visualization environment applicable to a wide range of functions

terms of improving and broadening analytic reasoning at the individual task level, and amongst tasks and functions in terms of scene generation. Also, data visualization has the potential to help overcome some of the integration challenges created by the need to discretize the job of construction management into somewhat independent functions in order to make it manageable. In summary, data visualization can assist with at least three significant management needs: analytical reasoning, communication, and learning (Pich et al. 2002; Puddicombe 2006). Ideally, data in support of all functions would reside in a single system facilitating its ready access, but the reality is that multiple systems may be involved. An underlying assumption of our work is that however the data is stored (single system or multiple systems), all of it can be accessed for use in a visualization environment.

### **3.2 Approach and structure of chapter**

In Chapter 2, three questions posed and pursued, with emphasis on the first two, dealt with the principles and guidelines for design of a visual analytics environment, the ability of users to perceive the usefulness of individual visual representations, and how to design and implement an analytics environment which is responsive to the realities of the construction industry. With respect to the principles treated, topics addressed included understanding the purposes of analytical reasoning, organizing data representations and data transformations, designing visual representations and interaction features, and the evaluation of visual representation designs. In this chapter, using primarily a top-down design approach which is described later, we extend our past treatment of principles and guidelines by setting out some concepts of analytics as they apply to CM in general. We then elaborate on these concepts in the context of time management, with the focus being on planning, monitoring, diagnosing and controlling time. This allows us to address more fully the third question posed previously, namely “How should a visual analytics environment be designed and implemented so that it is responsive to the realities of the construction industry and satisfies the criterion or test of practicality?” (Russell et al. 2009b).

The visualization environment design task is aided by a top-down approach because it helps one:

- identify the functions to be treated as well as the linkages amongst functions and related tasks in terms of the use of shared data either directly or through transformations;
- identify the analytical reasoning processes about project performance and conditions to be served (i.e. the suite of questions for which answers / insights are sought);
- establish requirements for consistency in the formulation of images and interaction features developed (e.g. filtering, sorting, highlighting, coordination, and navigation capabilities);
- describe the degree of flexibility that users should have especially in terms of being able to create scenes to assist in establishing likely casual relationships;
- determine how the environment should be evaluated to assess the degree of conformance with established requirements; and,
- identify opportunities for data visualization to extend existing management capabilities in terms of reasoning, communication and learning by critically assessing strengths and weaknesses of current management practices (a particular benefit of a top-down approach).

A bottom-up approach on the other hand assists one to identify the properties required of individual images that could be helpful in addressing the analytic reasoning associated with individual tasks within a function and at the function level itself. It is at this detailed level where one applies many of the findings of researchers on image design and image specific interaction features (encodings, use of colour, etc.) (Meyer et al. 2009). In carrying out this approach, opportunities for improving insight-generation capabilities can be identified through the design of novel images that juxtapose information in a manner that aids reasoning (e.g. clustering of change orders in time versus the trajectory of forecast project completion date – see (Russell et al. 2009b)). In reality, one iterates between a top-down and bottom-up approach as the design of a visualization environment is evolutionary in nature – development of one set of capabilities generally leads to ideas for additional capabilities – from the detailed level up to the overall environment level and vice versa.

The remainder of the chapter is structured as follows. A brief overview of recent construction data visualization work is first provided. Then, as part of the top-down approach, we introduce concepts and useful terminology related to a structured way of thinking about analytical reasoning and visual analytics, and their relationship with construction management functions. The focus of the latter then shifts to how a construction data visualization environment can support project participants' analytical reasoning needs for the management of time, specifically planning/predicting and monitoring/diagnosing/controlling construction conditions/time performance. A case study of aspects of an actual project examined using the construction data visualization environment developed to date is then presented. Purposes served include demonstrating the breadth of support that can be offered for reasoning by such an environment and thus its usefulness for conducting CM analytic tasks, and providing a test case for demonstrating the kind of evaluation process one should engage in to assess how well an environment conforms to the requirements set out for it. Time management functions treated for this case study include assessing quality of a baseline schedule, assessing actual vs. planned construction conditions/time performance, and assessing reasons for deviations. An evaluation of the current environment is then made to assess conformance / non-conformance with the requirements established for it and to identify worthwhile extensions to it. This evaluation involves both a top-down and bottom-up approach. The chapter concludes with a discussion of lessons learned from work performed to date, and their application to create a more comprehensive visualization environment that supports key tasks within a CM function and multiple functions, as depicted in Figure 3.1.

### **3.3 Data visualization in construction management**

Significant opportunities exist for the integration of advanced interactive tools and techniques along with visual analytic tools in support of a diverse range of CM functions. To date, however, only a modest level of effort has been expended by the construction academic community on this topic. Despite the fact that the field of visualization has been instrumental in representing how physical artefacts are to be built from constructability reasoning and construction method workability perspectives (e.g. (McKinney and Fischer

1998; Sriprasert and Dawood 2003; Staub and Fischer 1998)), the literature reveals very little about the visualization of heterogeneous multi-source, multi-dimensional, and time varying data in the context of construction management.

With regards to visualizing structured-abstract construction data, Russell and Udaipurwala (Russell and Udaipurwala 2000b; Russell and Udaipurwala 2002) demonstrated the value of using linear planning charts which, when combined with ancillary images pertaining to the distribution of resource usage in time and space, permitted additional insights on the quality of a schedule to be gleaned. Extensions to this work to include 4D CAD further enhanced the ability to generate insights on quality of schedule and strategies to improve schedule performance (Russell et al. 2009a). Songer et al. (Songer et al. 2004) developed and evaluated four visual representations including scatterplot, linked histogram, hierarchal tree, and treemap to represent structured cost control data. Song et al. (Song et al. 2005) proposed a 3D model-based project management control system where the visual platform (i.e. the 3D building model) itself serves as a construction information delivery platform. Vrotsou et al. (Vrotsou et al. 2008) applied Time Geographical methods to visualize work sampling data to allow analysts to understand better the distribution of activities and the interdependencies amongst them. Zhang et al. (Zhang et al. 2009) used an integrated building information system and digital images captured on site to semi-automate the calculation of progress measurements (e.g. cost and schedule variance) for items of a work breakdown structure. Lee and Rojas (Lee and Rojas 2009) recommended principles for designing effective visual representation for actual construction performance data as did the author in Chapter 2.

Based on the literature review of past research on construction management data visualization, it is observed that “a limitation of work to date on abstract construction data visualization as opposed to physical product visualization is that it is mainly exploratory in nature, with limited breadth in terms of the type of data and information entities examined, management functions examined, and guiding principles for designing relevant visual images” (Russell et al. 2009b). In particular, relatively little research seems to have been

conducted that addresses a detailed and systematic process for designing either specific images or a comprehensive CM data visualization environment reflective of clearly identified analytical reasoning, decision making, communication, and learning needs. Such a process is required in order to develop a detailed specification / set of requirements for the design of a construction data visualization environment comprehensive enough to support the full spectrum of CM functions.

### **3.4 Concepts of analytics and relation to development of a construction management data visualization environment**

As part of our top-down approach, we seek a structured way of thinking that will help in formulating visual images of data and supporting interaction features to assist with the analytical reasoning tasks associated with various CM functions, singly or in combination, and related performance measures. Introduced in this section is important terminology related to analytical reasoning and visual analytics, and its adaptation to the context of construction management. We make extensive use of the term ‘model’ in reference to explicit and implicit (tacit) knowledge models and predictive versus explanatory (diagnostic) models. The term explicit model is used to describe a quantitative relationship between input and output variables, whether based on fundamental principles (e.g. Critical Path Method (CPM) network model) or empirically derived (e.g. regression equation, factor based mathematical relationship, neural net). The model structure may or may not be transparent to the user. Implicit models are viewed as residing in the minds of experienced construction management personnel, and in general cannot be readily documented. A predictive model refers to the application of either an explicit or implicit model to forecast or predict likely outcomes in the future based on some assumed fact pattern or set of input variables. In contrast, an explanatory model involves explaining actual outcome values, and basically involves working an explicit or implicit model in reverse and as measured against some originally assumed set of conditions or objectives.

### 3.4.1 Analytics for construction management

Here we present several analytic concepts as they relate to construction management in general and which assist in determining CM data visualization environment requirements. Construction project management involves an integrated process of managing construction conditions to achieve required levels of performance, dimensions of which include time, cost, scope, quality, safety and risk, and values of which depend on how the construction conditions imposed and/or encountered are managed. The terminology of “construction condition” is an umbrella concept covering construction strategies imposed, construction requirements dictated in contracts, construction constraints encountered, and so forth. Dependency relationships exist amongst construction conditions and performance (e.g. a dependency relationship between productivity and activity duration). Therefore, the terminology of “construction dependency relations” is used for referring to such relationships.

*Construction conditions, performance, and dependency relations can be understood by their characteristics, which can be viewed from three perspectives. Firstly, they have a temporal status – one of planned, in progress, or actual. Secondly, conditions, performance and dependency relations can be characterized at different levels of detail ranging from overall characteristic (i.e. overall qualitative pattern), local characteristic (i.e. local qualitative pattern), and individual characteristic (i.e. single value). For example, the characterization of time performance can be described as stable overall (i.e. the entire project remains on schedule), the rate of production is accelerating in the middle of the project, or simply the duration of an individual activity. Also, the level of detail can be defined by data range (e.g. entire data set vs. a data subset) and/or data granularity (e.g. all footings vs. spread footings only). Lastly, conditions, performance and dependency relations can be examined from various perspectives such as time, space, product, process, and participant. In general, the primary purpose of the analytical reasoning (i.e. analytics) involved in construction management processes is to gain an understanding of the characteristics of construction conditions, performance, and dependency relations. This “analytics for construction management” is referred to as CM analytics herein.*

Most of the CM processes associated with different project phases involve the iterative application of CM analytics for the purposes of planning/predicting, monitoring/diagnosing/controlling/, and re-planning/ re-predicting construction conditions, performance, and dependency relations. For example, construction time management during the planning phase involves assuming/imposing (i.e. planning) certain construction conditions, applying explicit or implicit dependency relations or models to obtain a forecast (prediction) of time performance and then seeing whether the forecast time performance satisfies the contractual requirements. Purposes served, tasks, and workflows of CM analytics applicable to construction management functions are depicted as CM analytics flow charts in Figure 3.2 and elaborated upon as follows:

1. *CM analytics for predicting/planning purposes, Figure 3.2(a)*: Steps involved include the specification of inputs for the explicit CM prediction model of interest (e.g. (Babu and Suresh 1996; Chao and Skibniewski 1998; Chong et al. 2005; Motawa et al. 2007; Staub-French et al. 2003)), running the relevant model, and obtaining model outputs. The models can be mathematical-function based or artificial intelligence based, and deterministic or probabilistic. Having obtained model outputs, one then examines both inputs (i.e. construction conditions/performance assumed) and outputs (i.e. construction performance predicted) to gain insights (i.e. the purposes served) into:

- characteristics of the inputs and outputs (e.g. crewing levels vs. milestone dates, undesirable construction conditions vs. performance impacts, etc.) so that project participants can identify additional construction dependency relationships possibly unique to the project at hand or limitations on the models used. Other purposes served include inspecting quality of data entries, and assessing validity of the models used; and
- how the change of inputs affect outputs, with the goal being to identify the best plan possible.

2. *CM analytics for monitoring/diagnosing/controlling purposes*: The primary purposes served here deal with examining performance to date, explaining (diagnosing) reasons for it, and then determining the most relevant actions to take. Use is made of, and thus support is required for, both formal (explicit), diagnostic (explanatory) models as well as

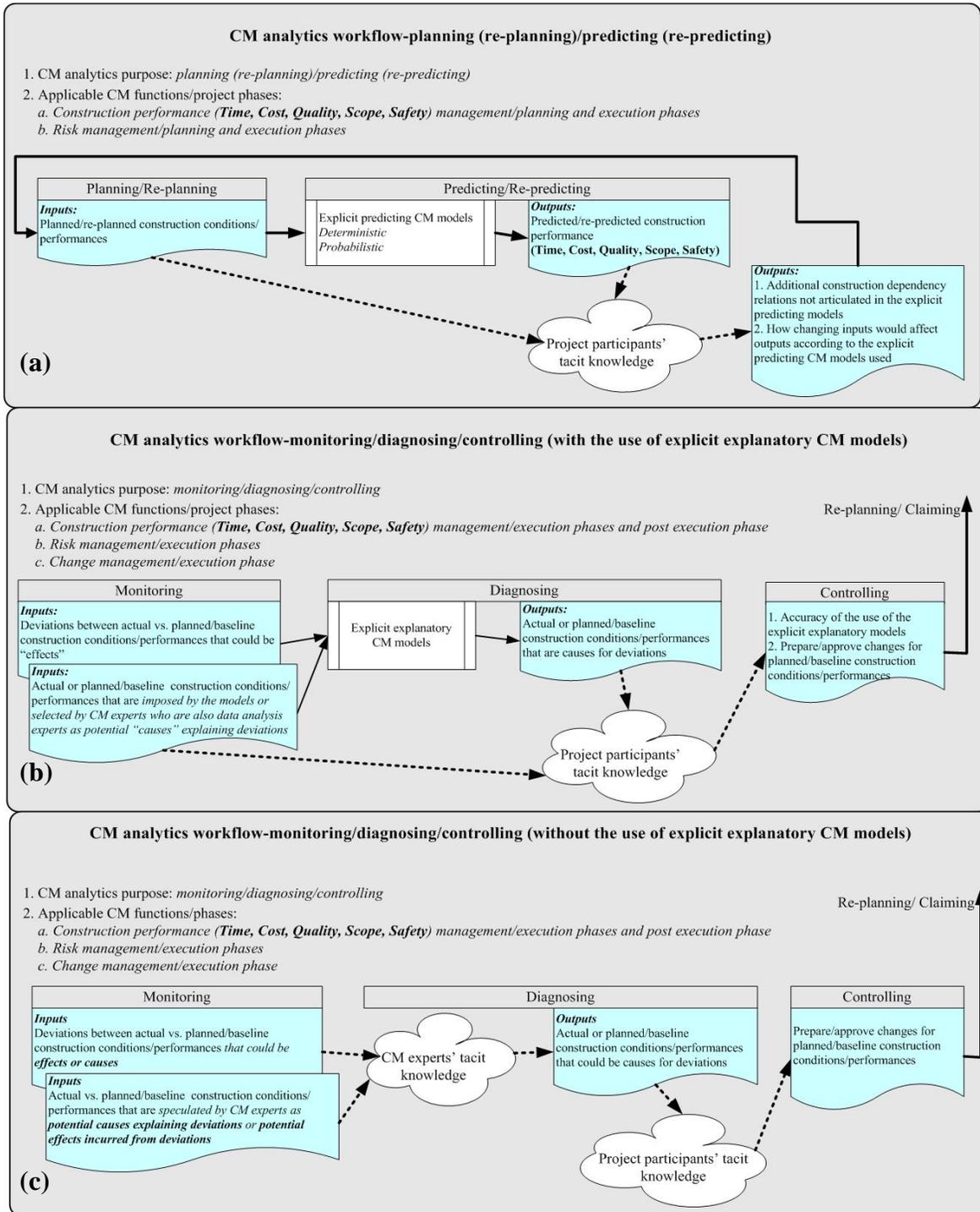


Figure 3.2 CM analytics flow charts

the implicit ones which reside in the minds of project participants. Specifically, diagnosis can be made:

- with the use of explicit CM explanatory models (see Figure 3.2(b)) by preparing inputs for an explicit CM explanatory model (e.g. Battikha 2008; Moselhi et al. 1991; Russell and Fayek 1994; Soibelman and Kim 2002), running the model, and obtaining outputs. The models can be mathematical-function based or artificial intelligence based. By examining the inputs (i.e. deviation of actual vs. planned/baseline construction conditions/performance) and outputs (i.e. actual or planned/baseline construction conditions/performance that explain the deviations) of the explanatory models, insights into their characteristics can be gained, reasonableness of model results assessed, and changes made to planned/baseline construction conditions/performance as appropriate; and,
- without the use of explicit CM explanatory models (see Figure 3.2(c)). Steps involved include examining actual vs. planned/baseline construction conditions/performance to gain insights into their characteristics so that likely reasons for any deviations between the actual and the planned/baseline can be inferred, using human-based reasoning processes. Once these reasons are identified, changes to planned/baseline construction conditions/performance may be pursued as part of the corrective/preventive actions.

The solid flow lines connecting the text boxes in the CM analytics flow charts (Figure 3.2) represent the CM analytics workflows that utilize explicit, machine-based CM models (referred to later as machine-based CM analytics) while the dashed flow lines connecting text boxes and mind maps represent CM analytics workflows that utilize human tacit CM knowledge (referred to later as human-based CM analytics). Different project managers have different thinking styles, experiences, and knowledge (Tullett 1996) and it is difficult to predict the steps a person takes to explore, acquire, organize, and use information to assist analytical reasoning (Stolte et al. 2002; Tullett 1996). Nevertheless, the CM analytics processes depicted in Figure 3.2 account for most CM analytical reasoning scenarios that project participants use in carrying out construction management functions.

### **3.4.2 Visual CM analytics supported by a data visualization environment**

For project participants to conduct the CM analytics tasks just described, an essential complementary task is to search out and analyze data relevant to construction conditions, performance, and dependency relations. The current state-of-the-art of CM data analysis is dominantly computational in nature and machine-based. However, computational data analysis requires pre-defined algorithms that capture explicit CM knowledge either in transparent or non-transparent form. While useful, these algorithms on the one hand can be too simplistic because they leave out complex and comprehensive real world phenomena, or on the other hand, require data analysis experts to operate them (Diekmann 1992), leading to reservations on the part of industry to use them. Computational data analysis for construction management constitutes the “machine-based CM analytics” mentioned in the previous sub-section.

As an alternative to the foregoing, a relatively recent data search/analysis paradigm of “the use of visual representations of data and interactions to accelerate rapid insights into complex data” and coined “visual analytics” (Thomas and Cook 2005) is advocated herein. It transcends the use of computerized statistical graphics as an essential data analysis tool which began in the 60s~70s (Friendly 2008; Schmid 1983) and the development of visualization for information search emerging in the 80s~90s (e.g. (Card et al. 1991; Shneiderman 1994)). The paradigm of visual analytics has the potential of addressing some of the shortcomings of computational data analysis. Equally if not more importantly, it is also a particular fit for CM use because a construction project involves all types of data ranging from structured abstract data (our primary focus herein), geometric data, and unstructured data such as pictures and textual documents. With the advanced capabilities of graphics automation, interactivity, and computation provided by computer technology, it is possible to develop a computerized construction data visualization environment that allows project participants to go quickly through complex data presented in easily-understood/natural visual forms. This facility allows them to gain important insights by applying their experience and knowledge to interpreting what they see in the images. Therefore, a terminology of “visual CM analytics” can be defined as “conducting CM

analytics by the visual analytics approach”. This kind of visual data analysis is mainly applied to support the “human-based CM analytics” mentioned in the previous sub-section and focused on hereafter.

A data visualization environment should allow users to conduct visual CM analytics in support of both human-based and machine-based CM analytics. Such an environment is meant to be an interactive one built as an integral part of a computerized construction information system. It should allow users to access and navigate galleries of visual representations on demand. These representations depict the complex data stored in the information system and illustrate salient characteristics of construction conditions, construction performance, and construction dependency relations hidden in this data. The shaded text boxes of the CM analytics flow charts seen in Figure 3.2 illustrate construction data that should be turned into visual representations or images that are pre-coded in the construction data visualization environment for project participants to select and view. This gallery of images has to be capable of presenting the characteristics of a spectrum of construction conditions and performance reflective of different time statuses, at different levels of detail, and as observed from the perspectives of different project views (e.g. product, process, participant, as-built, change, quality, etc.). The dashed arrowed lines going from shaded boxes to mind maps in the same figures represent the possible CM analytics workflows that project participants may take by interacting with this visualization environment. This interaction takes the form of iteratively accessing and navigationally viewing the pre-coded images individually or in the form of image scenes, depending on participant cognitive styles and purposes of the CM analytics. Inclusion of a comprehensive set of interaction features is core to the successful development of a responsive visualization environment.

To support the types of analytics described in the foregoing, three essential requirements must be addressed in a CM data visualization environment, as follows:

1. An extendable gallery of ready to use (pre-coded) visual representations of CM data should be supported. These ready to use representations encapsulate a range of

construction conditions/performance mapped against the primary CM data dimensions of time, space, product, process, and participant in support of analytic reasoning. Pre-definition of these representations in as flexible and easy to use manner as possible is seen as being essential for them to be employed for everyday use by construction practitioners.

2. When representing specific construction conditions/performance couplets, the ability to represent different status states must be supported (e.g. planned versus actual). This includes the ability to show the differences (deviations) in status values of performance measures as a function of the differences of status values of condition parameters.
3. A rich set of interaction features should be present to allow users to interact with the environment in support of the following essential tasks: (a) choose the visual representation to be generated; (b) adjust granularity and value range of data dimensions for the visual representation of interest; and, (c) view several visual representations which portray different aspects of construction conditions/performance either sequentially or simultaneously (i.e. multiple views (Wang Baldonado et al. 2000). Being able to display multiple views is particularly useful because not only does it reduce the burden on human visual memory (Plumlee and Ware 2002), it also supports (i) scanning overviews and then viewing details, (ii) comparing differences (Roberts 2007), (iii) viewing a range of attributes (Wang Baldonado et al. 2000) related to the user's chosen CM strategy, and (iv) comparing conditions with performance measures to observe likely cause-effect relationships.

### **3.4.3 Analytics for time performance management**

Described in this section are the analytics reasoning processes associated with planning and then monitoring, diagnosing and controlling time performance. In the case study section of the chapter, we demonstrate an implementation of these processes in the form of a research visualization environment.

### 3.4.3.1 Visual CM analytics for planning/predicting time

An important and representative CM analytics task for planning/predicting analytic reasoning purposes involved in construction time management is appraising the quality of a planned schedule (i.e. inputs and outputs of the scheduling model). This task encompasses many aspects of planned construction conditions/performance (e.g. crewing and sequencing conditions, early/late finish dates, etc.). A complementary reasoning task that involves the same data and which could benefit from the application of visual analytics is judging the quality of data entry, which includes identifying erroneous data values (e.g. crewing data, activity duration data, etc.).

Reasoning about a schedule and associated data in terms of both formulation and evaluation from the contractor's perspective involves a number of considerations which are also applicable to the reasoning associated with monitoring/diagnosing/controlling. They include:

- **Ordering of construction processes:** The effective ordering or sequencing of construction processes, requires consideration of at least three project dimensions, these being time, space, and project participant. The time dimension speaks to the logic or order in which work should be performed, and hence its placement in time. Intersecting with this ordering is consideration of the sequence in which trades (project participant) should perform their work to the extent that it is discretionary, so as to minimize deficiencies plus access and possibly congestion issues. For large scale linear projects in particular, consideration also needs to be given to the order in which work flows through locations – i.e. the location sequence for multi-location activities. For example, it is desired to execute such activities in order of physical location adjacency so as to reduce mobilization efforts to a minimum. Time ordering considerations other than technological ones and preferred trade sequencing deal with balancing production rates by adjusting resource levels to reflect different location and trade work scopes and achieving work continuity to minimize interruptions between work at the locations of a multi-location activity.
- **Distribution of construction processes:** With respect to the distribution of

construction processes, consideration should be given to the project dimensions of time, space, project participant and physical component. The issue of granularity with which these dimensions is expressed also comes into play, as discussed below. Schedule properties of interest include: (a) how much work is/can be packed into a specific time window and over how many locations; (b) how much work is being executed during this time window at the same work location with emphasis on simultaneous or overlapping activities and what is the distribution of this work in terms of sub-locations, if any; (c) what is the footprint of the work being performed for a specific time window (i.e. number of locations in progress simultaneously; (d) how many work faces are there and their distribution over the project site; (e) what spans of control for project participants in terms of active locations are implied; and (f) how much time is required to complete an individual physical component or collection of similar components at a work location. All of these properties reflect on the workability and quality of a schedule.

- **Granularity:** Granularity refers to the level of aggregation or decomposition that is useful for communicating information to project participants and for extracting meaning from the schedule. Dimensions involved include time, space, project participant, physical component (product), and activity (process). Ideally, it would be desirable to be able to work at different levels of detail for each dimension. In terms of time, from hours to days to weeks to months to years; for space, from collections of locations to individual locations to sub-locations of a location; for participants, from individual trades to collections of participant types; for physical components, from individual constituents of a component to a complete component to collections of a specific component, and for activities from individual activities to collections of activities for shared properties of interest – i.e. belong to the same physical component or same project participant.
- **Compliance:** Compliance refers to meeting requirements specified by contractual language (e.g. milestone dates, including project completion and client specified time windows for specific aspects of the work), applicable codes, regulations and agreements (e.g. permits required, allowable working hours and days, etc.), and constraints imposed

by prevailing natural and man-made conditions (e.g. black out times for certain types of work, resource availability limitations, restrictions on work face access, etc.).

#### **3.4.3.2 Visual CM analytics for monitoring/diagnosing/controlling time**

During the execution phase, construction time management is mainly an iterative three-step process of monitoring, diagnosing, and controlling construction conditions/time performance. As an overview of the primary roles that CM analytics can play, for the monitoring function, it can assist with identifying deviations between actual and planned/baseline characteristics of construction conditions/time performance. For the diagnostic function, it can help in identifying actual or planned/baseline characteristics of construction conditions/time performance that explain the deviations. And for control, it can help to identify changes to planned/baseline construction conditions/time performance that are reflective of findings from the diagnostic function and that may aid in mitigating problems encountered to date. An additional function served by CM analytics is to assist in determining the completeness and accuracy of conditions data by providing big picture views of it in order to determine data gaps and anomalous values. Incomplete and inaccurate data can make difficult the task of diagnosing the true causes(s) of deviations encountered and can result in conflicting interpretations of reasons for performance to date.

In essence, in order to design a visualization environment for the monitoring, diagnosing and controlling functions in the context of time management, one seeks answers to the following question groups.

Question group (1) – Fact representation of time performance

- (i) What are the key performance indices for construction time performance?
- (ii) What are their values and patterns of behaviour and how can they provide useful insights into time performance versus various project dimensions and assist in positing cause-effect hypotheses?

Addressing these questions allows one to examine time performance as a function of one or more project dimensions (i.e. space, time, participant, product, process) and determine

regions of inadequate performance – i.e. one seeks to determine the facts. This might then be followed by reflecting on possible reasons for it, and hence conditions that should be explored. For example, if activities (process dimension) only experienced delays in the later stages of a project (the time dimension), for one project participant (e.g. subtrade), then construction personnel could focus their attention on conditions pertaining to this stage of construction in order to identify potential causes of delay.

Question group (2) – Fact representation of conditions and other performance measures

- (i) What construction conditions or other construction performance measures (i.e. measures not directly time related such as scope, safety) may have contributed to the time performance detected in addressing question group 1?
- (ii) What are their values and patterns of behaviour and how can they provide useful insights into time performance versus various project dimensions and assist in positing cause-effect hypotheses?

The primary value in seeking answers to these questions is to determine if the factors contributing to unsatisfactory time performance are more broadly based than ones for which management personnel believe there is a direct causal relationship with time performance. This is reflective of the intricate web of interactions amongst the condition variables and performance dimensions that accompany construction projects. For example, it has been demonstrated that construction data collected for one management function has the potential to be used in CM analytics for more than one management function. This was illustrated by Lu and Anson (Lu and Anson 2004) who used quality control data to analyze the actual productivity of placing concrete. Thus, the facility must be present to allow users to explore a wide store of data, not just data directly related to one performance measure.

Question group (3) – Evidence in support of participant causal models

- (i) What is the specific evidence in terms of comparisons between planned and actual conditions and between conditions and performance that support project participant causal models to explain the time performance of a particular collection of activities related to one or more project dimensions?

Here one seeks to demonstrate compelling causal links between conditions and performance determined by addressing question groups 1 and 2 as a function of known and generally accepted quantitative models, participant tacit knowledge, or by direct learning from the unique project context at hand. In so doing, one seeks to pinpoint conditions/performance evidence that can objectively explain the time performance of interest. This can be a challenging task, given that the analytical reasoning involved may not be limited to simply identifying one layer of cause-effect relationship between construction conditions and construction performance.

### **3.4.3.3 Visualization requirements deduced from time performance management analytic needs**

Based on the foregoing discussion of the time management function, the requirements for a generalized CM data visualization environment identified earlier in the chapter should be augmented by the following ones:

1. One focus should be on visualizing individual activity time performance, for which the time units of date and duration are used as performance metrics. The ability to represent visually contractual requirements with regard to planned and actual time performance of activities should also be supported in order to assess compliance with contractual requirements, both for the planning and execution phases of a project.
2. The ability to associate construction conditions that can be represented visually with individual activities should be supported. Thus, for activities that experience unsatisfactory time performance, conditions directly relevant to those activities can be readily displayed. The association can be explicitly specified by users or inferred because of their proximity in time and space, and possibly with participant and product.
3. The ability to display sequences of activities and associated conditions should be supported. Such sequences can relate to one or more paths in a network model, the sequence of work of an individual trade, or the sequence of work at a specific work location.

### **3.5 Case study of CM analytics using a data visualization environment**

A case study that treats construction time management during the planning, execution, and post execution phases is used for a two fold purpose: (a) to demonstrate how a construction data visualization environment can assist with important CM analytic tasks; and (b) to elaborate on the process for developing a CM data visualization environment, with emphasis on evaluation to assess compliance with requirements developed as part of a top-down and bottom-up design approach. Analytic purposes served include: planning/predicting (e.g. assessing planned construction condition/time performance such as quality of schedule) and monitoring/diagnosing (e.g. assess actual vs. planned/baseline construction conditions/time performance). Such case studies also help to identify the benefits that could be derived from additional features or requirements of a visualization environment. The current implementation of the visualization environment forms part of the REPCON research system (Russell and Udaipurwala 2004). It employs custom designed schedule graphics routines and pre-coded as-built graphics and associated features using CHARTFX 6.2 Client Server (Software FX Inc.). During the original development of the system when creation of a visualization environment was not the primary focus, a bottom-up approach was used for visual representations, with the focus being on the design of individual images. This led to recognition of the potential for a comprehensive data visualization environment for integrating the use of these images and hence pursuit of the current top-down approach. As a result, there has been a significant enrichment of the palette of images and user interface features supported.

#### **3.5.1 Case study overview**

The case study data comes from a 3 km segment of the original Advanced Light Rapid Transit Project (ALRT) in Vancouver, British Columbia built some 24 years ago. It reflects the actual as-planned and as-built schedule data and the problems encountered, as seen through the eyes of the contractor. The scope of the work consisted of building 103 foundations and piers in support of a pre-cast beam elevated guideway, with installation of the beams being performed by others. Use is made here of three project views: (i) the product view which contains data related to what is to be built and the site context; (ii) the

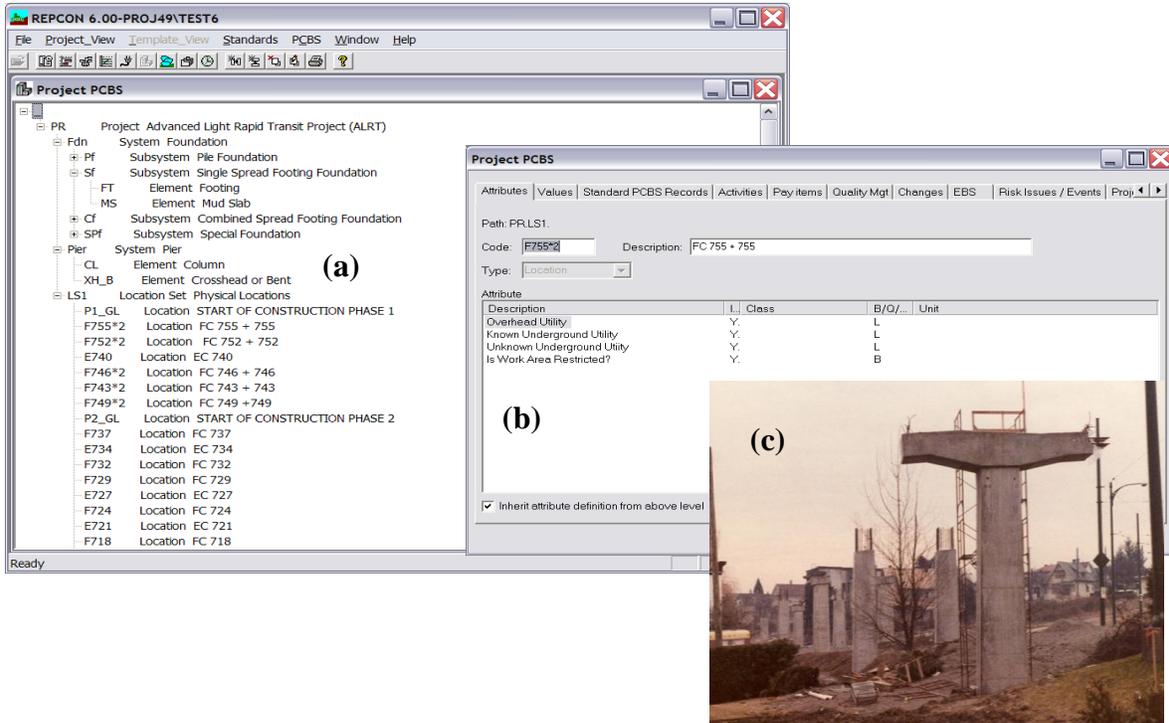


Figure 3.3 Product view – (a) project locations; (b) location attributes; and (c) photo of components

process view which contains data pertaining to how, when, where and by whom a project is being built; and, (iii) the as-built view which captures data that describes what happened, why and actions taken. The product view (Physical Component Breakdown Structure (PCBS)) consists of a simplified list of project components and a listing of all work locations in planned location sequence, as partially shown in Figure 3.3(a). Location attributes related to problems encountered are depicted in Figure 3.3(b), and a photo of actual columns is shown in Figure 3.3(c). The project plan and schedule (process view) was recreated using the original assumptions and constraints related to work location sequence, work continuity constraints, number of crews, contract milestone constraints, and decomposition into phases because of the then capacity constraints of the scheduling system used. (A much more elegant modeling of the project could now be done using current technology which would enhance the clarity of the schedule images.) As can be observed in Figures 3.4(b) and (c), respectively, the bold red lines in the linear planning (LP) chart and filled bars in the bar chart schedule representations show the critical activities. Because of the way the schedule was originally defined with start milestones to

link the phases and no finished milestones with late date constraints, no critical path is shown from beginning to end of the project. Scoping of activities includes: (i) survey & layout; (ii) excavate & mud slab; (iii) form & reinforce footings; (iv) pour footings; (v) form columns; (vi) pour columns; (vii) cure & strip columns; (viii) backfill & grade; and, (ix) cleanup. Other activities not present at all locations include piling at locations of weak soil conditions and construction of column crossheads (form, reinforce, pour, cure & strip, and post tension). The as-built view contains data pertaining to the daily status of site conditions as well as activity status from the available records maintained by the contractor. The activity status includes “start”, “ongoing”, “finish”, “same day activity start and finish”, “idle” (work started but later interrupted), and “postponed” (start of activity delayed), respectively. Daily site environmental data for the project’s time frame was retrieved from the Environment Canada web site as we did not have access to weather data recorded by the contractor. Three parameters, (i) daily high and low temperatures; (ii) precipitation; and, (iii) wind speed were coded due to their significance for construction operations and performance measures. Problems encountered during project execution, as compiled and documented by the contractor, were categorized, coded, and associated with the daily status of each activity in the as-built view, as appropriate.

### **3.5.2 Visualization for CM analytics for planning/predicting time**

The current status of the data visualization environment includes the capability to generate visual representations of planned schedule data in various formats ranging from linear planning graphics, bar charts, and network diagrams. Interaction features such as querying data (e.g. filter data, sort data) and view navigation (e.g. zoom in/out) are supported for users to enhance image readability and to cope with project scale. Treated here is the use of linear planning graphics (Figures 3.4(a) and (b)) and their associated interaction features for assisting CM analytics relevant to assessing the quality of a schedule. It is noted that Figure 3.4(a) was derived by zooming in on the schedule both in time and space and 3.4(b) was derived by zooming out on the schedule in time and space to provide a global view of the schedule (i.e. the ability to overview and view detail simultaneously).

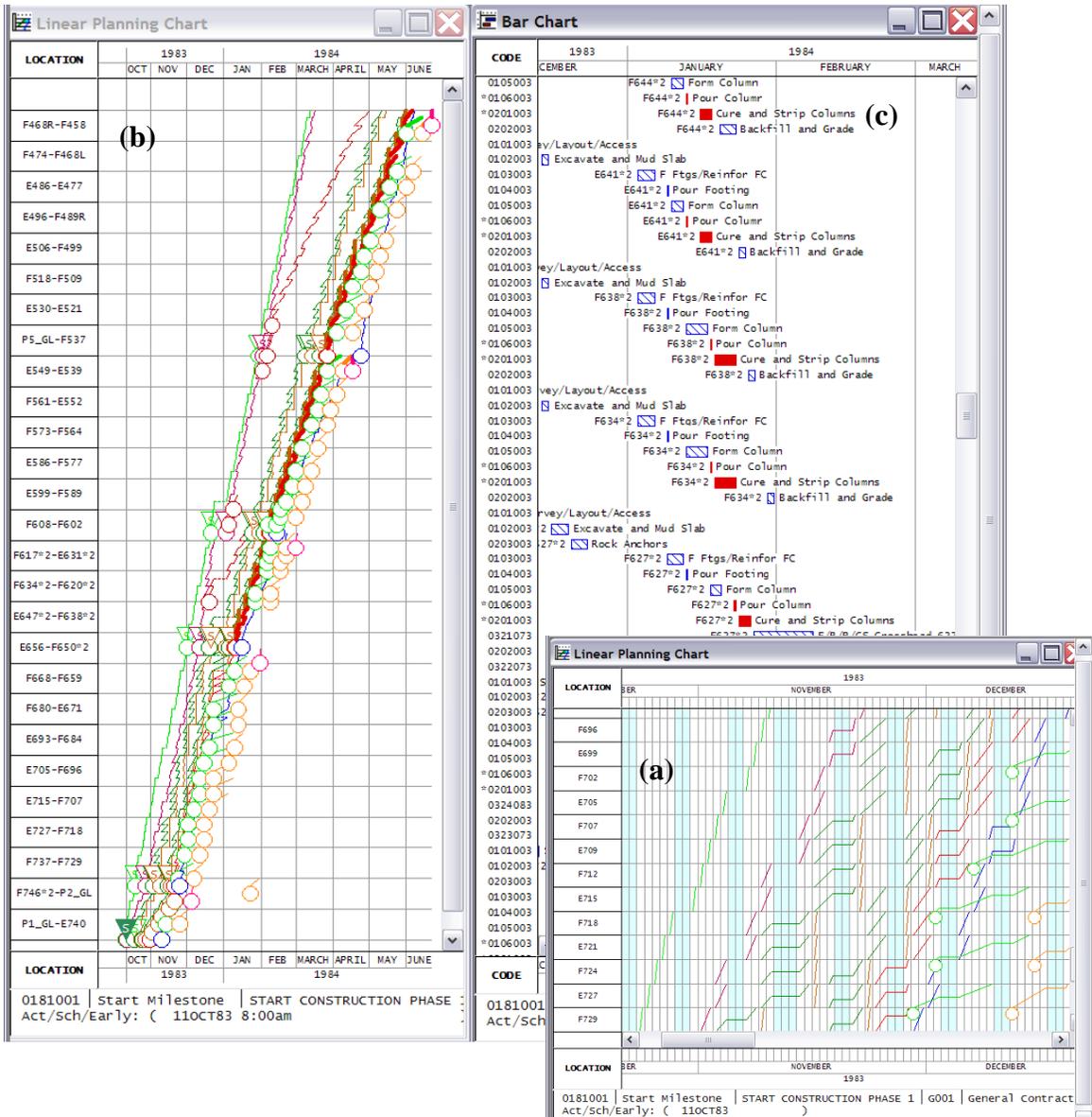


Figure 3.4 LP and Bar Chart schedule representations

In terms of the attributes required of a good “quality” schedule, the linear planning charts shown in Figures 3.4(a) and (b) reflect some of the considerations previously described that should be taken into account in order to ensure development of an effective construction strategy, especially for work in an urban environment. First, as observed from Figures 3.4(a) and (b), work is executed in an ordered work location sequence. Next, as evident from the figures, production rates have been balanced by matching them through

one or more of scope of work, resource allocation level, or the use of multiple crews (meaning in this context the simultaneous execution of the same work at two or more work locations).

Further, as shown, a good strategy allows for work continuity at the activity level (as soon as work is completed at one location, it commences at the next location in the ordered sequence of locations) which ensures efficient utilization of onsite resources as well as offsite fabrication and handling (e.g. pre-casting) of components, if relevant. In terms of the distribution of work, it is sufficiently separated in terms of space and time to avoid work site congestion and thus detrimental impacts on productivity. In addition, minimizing the length of the construction foot-print on the work corridor throughout the project duration also contributes to the quality of the schedule, as it helps minimize third party stakeholder concerns.

While bar chart representations (see Figure 3.4(c) in which activities are grouped by work location) employ visual analytic capabilities relating to comparisons and interactive tools of filtering and zooming, for large-scale linear projects, they cannot provide a holistic overview of the schedule, nor do they make easy the task of assessing schedule quality in terms of the ordering or distribution of work. As observed by Russell and Udaipurwala (Russell and Udaipurwala 2002), the inherent limitations of bar chart representations stem “from the fact that bar charts essentially provide a local view of the project, i.e. locally within a particular time window, making it extremely difficult to anticipate effects on the global project level for changes made to local activities. Furthermore, for sizeable projects, bar charts can lead to vast, difficult to navigate information spaces in which one can easily lose their bearings”. Nevertheless, the strength of the bar chart lies in its long-standing and pervasive use in the construction industry, and the ease with which it communicates work that has to be done and when. Thus, one seeks to complement its strengths, not replace it.

As seen from Figures 3.4(b) and (c), current interaction features allow for juxtaposing schedule views, horizontally or vertically. Note the use of different granularity in terms of

the time axis – for this representation, the LP (linear planning) chart provides a global view of the project schedule, while the bar chart provides a more refined level of granularity for time. In terms of the current implementation, the juxtaposed views are not linked in that as you navigate one, the corresponding activity is not highlighted in the other.

In summary, and within the context of large scale linear projects in particular, the visual representation of a schedule in linear planning form along with a rich set of interaction features facilitates the speedy yet in-depth consideration of the ordering and distribution of work in time and space at different levels of granularity as well as consideration of compliance with natural and man-made constraints. The insights provided by this format enhance the analytical reasoning abilities of users, and when fully integrated with a scheduling algorithm to form part of a visual CM analytics environment, users are empowered to iteratively design a high quality schedule.

### **3.5.3 Visualization for CM analytics for monitoring/diagnosing/controlling time**

In addition to assisting with the planning function, the LP chart schedule representation can be very useful in telling the as-built story for monitoring/diagnosing /controlling time performance. Depicted in Figure 3.5(a) is a schedule representation using actual as opposed to planned dates. As observed from this figure, a significant amount of work was done out of location sequence. While the image does not convey reasons for this out of sequence work or delay in completing the project, it does suggest likely productivity losses for the contractor because of the need to shift resources, including crews, formwork, and other equipment, seemingly at times on a random basis from location to location. The task of the contractor, having portrayed what happened, then becomes one of explaining reasons and responsibilities for the actual as opposed to planned pattern of work. Complementing Figure 3.5(a) is Figure 3.5(b) which contrasts the as-planned and as-built schedules for the lead activity excavate & mud slab, over the entire location range and time window to show the time trajectory of location work flow. The monitoring of time performance for this lead activity at individual locations can be done through further examination of Figure 3.5(b). In this figure, the slopes of the solid lines and their different colour codings represent the “actual” (green color lines) and “planned” (blue color lines)

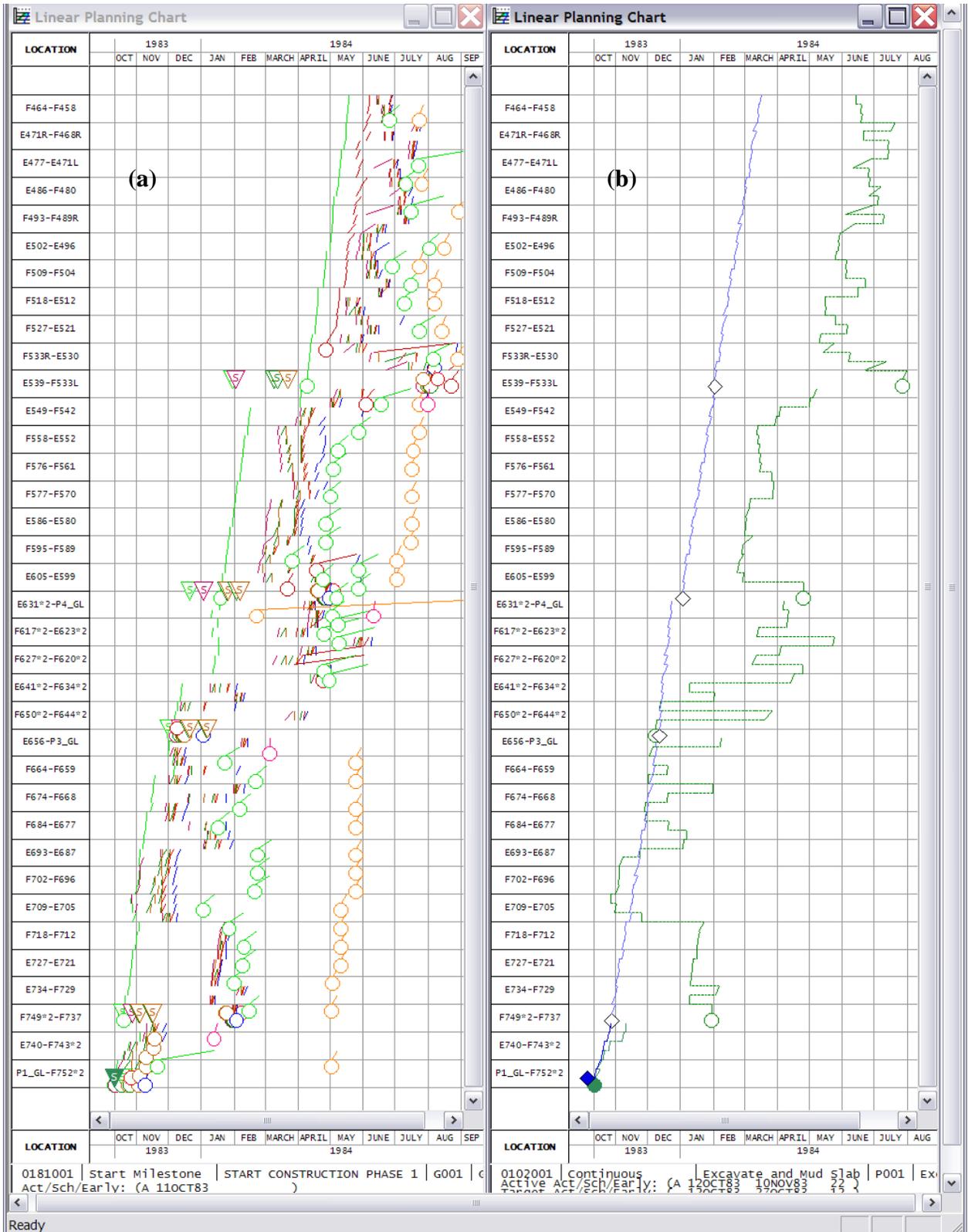


Figure 3.5 Tiled LP chart views of: (a) as-built schedule for all activities, and (b) planned vs. actual work trajectory for lead activity, excavate & mud slab.

production rates at each location. Also, the slopes of both lines at each location do not vary very much at all, which means the contractor maintained an actual production rate performance similar to the planned one. In contrast to this finding is the observation that this lead activity suffered severe start date delays at many locations, which likely was the main contributory factor to the overall time performance deviation (i.e. late finish) for this activity as well as for the entire project. This detrimental start date slippage began early in phase 2 (locations F737 to F654). As a response, the contractor worked out of the planned location sequence in an attempt to maintain the planned pace of construction. As seen from Figures 3.5(a) and (b), the actual footprint of the work was larger than planned, and in addition, continuity of work was not maintained. In summary, the LP aspect of the visualization environment helps one grasp on a holistic basis what actually happened versus what was planned, likely implications, and where to focus attention in order to diagnose reasons for performance. In terms of enhancing the LP aspect of the visualization environment, a complementary feature to the existing one of being able to show the time trajectory of contiguous work locations is the recently added option of being able to show a location trajectory of work. This feature assists in assessing the movement of resources, both planned and actual. It came about from application of the top-down approach described in this chapter.

For diagnosing the phase 2 time performance deviations identified in the previous paragraph, the visualization environment can be used to continuously and on demand present and view images of structured daily site data representing actual construction conditions (e.g. weather conditions) or construction condition deviations (e.g. execution problems encountered). This “on demand” capability of the visualization environment allows users to formulate and/or validate cause-effect hypotheses in terms of how construction condition deviations might cause time performance deviations based on user knowledge, experience, and cognitive style. Figures 3.6(a) through 3.6(c) depict images of construction data sequentially generated through the authors’ interactions with the visualization environment. The sequence of presenting and using images is attuned to the authors’ diagnostic strategy for seeking reasons for time performance deviations (delayed

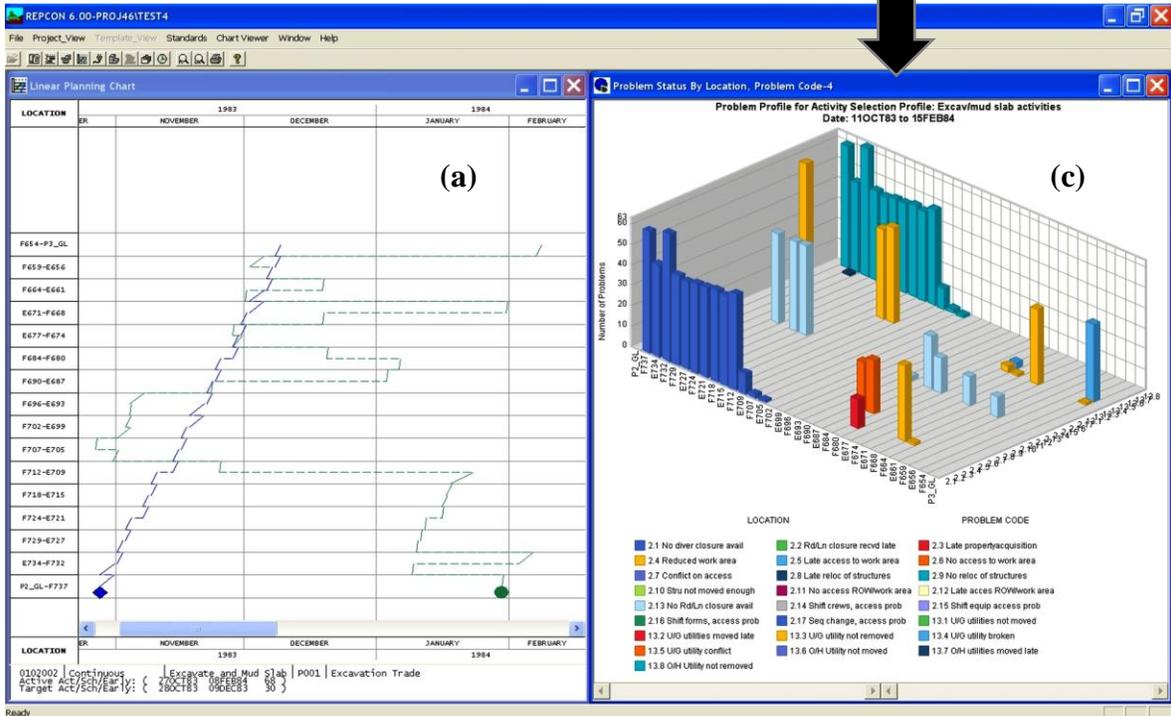
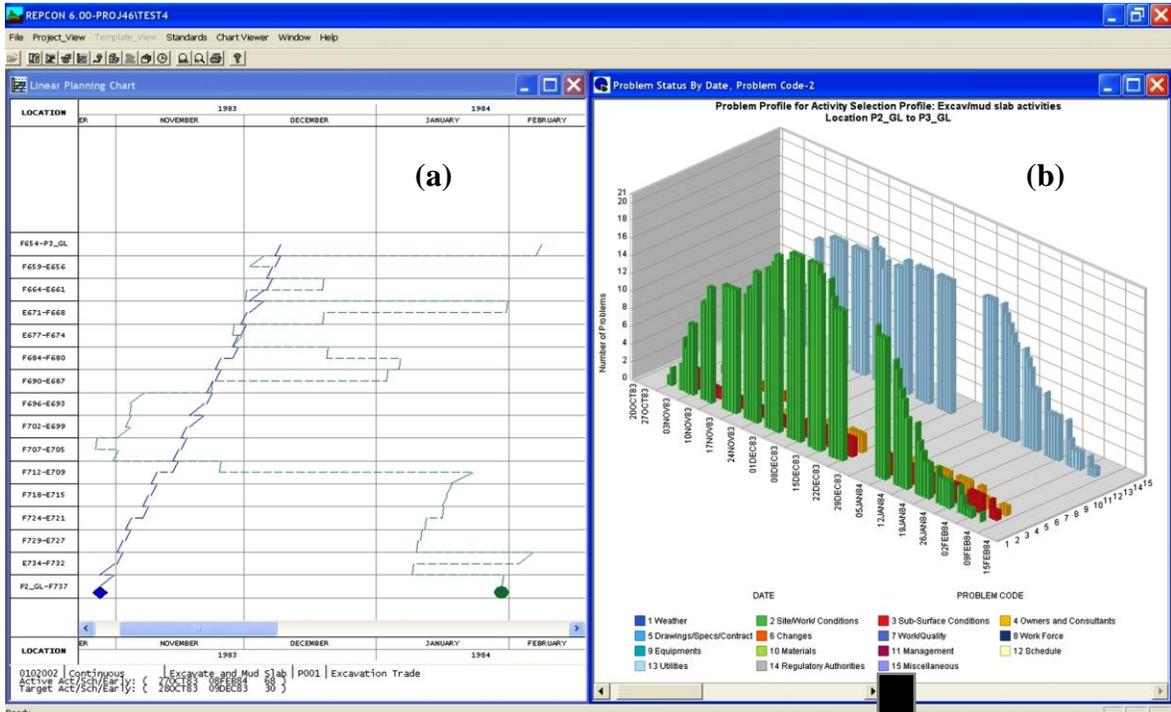


Figure 3.6 LP chart (a) and problem status charts- problem status by problem code vs. time (b); problem status by problem code vs. location (c)

starts at individual work locations). Other users may pursue their own strategy by taking different routes of image selection and viewing for the same diagnostic purpose.

Figure 3.6 reflects one of the authors' cause-effect hypothesis generation scenarios, the goal being to assess which, if any, problems (i.e. deviations in or unanticipated conditions) were encountered that contributed to the delayed commencement of the excavate & mud slab activity at various locations during the construction of phase 2. Specifically, we sought an answer to the question *“Did a clustering of problems coincide with the clustering of start date delays and altered work location sequence?”*. The authors started the hypothesis formulation process by generating an image of problems recorded in daily site reports that correspond to the excavate & mud slab activity's execution for the phase 2 time window and location range, as shown in Figure 3.6(b). The authors also re-focused Figure 3.5(b) to show only the phase 2 part of the comparison LP chart, as shown in Figure 3.6(a). In Figure 3.6(b), the challenges encountered are observed from the perspectives of “time” and “problem code category”. The category level constitutes an aggregation of all instances of problems recorded against user-defined problems in that category in terms of one of number of problem instances encountered, time lost, or man-hours lost – see Figure 3.6(c) as an example. By having the ability to place Figures 3.6(a) and 3.6(b) adjacent to each other and compare them, it can be seen that during phase 2, problems encountered by the excavate & mud slab activity were clustered in two problem code categories, namely Site/work conditions (in green) and Utilities (in blue). These two categories of problems are widely recognized (tacit knowledge) by construction personnel as being able to seriously impact a lead activity like excavation both in terms of timing and work location sequence. Hence, based on initial supporting evidence at the category level, confidence is built in the hypothesis that a clustering of problems for one or more problem categories could explain all or a significant part of the schedule behaviour shown in Figure 3.6(a).

The authors then sought further validation of the hypothesis by generating Figure 3.6(c) which portrays the specific problem types encountered under the two dominant problem categories identified in Figure 3.6(b), as observed from the perspectives of “location” and

“individual problem code”. By juxtaposing and comparing Figure 3.6(c) with the linear planning chart in Figure 3.6(a), it can be observed that locations where the start of the excavate & mud slab activity was severely postponed (seen in Figure 3.6(a)) matched the ones where the specific problems of “no divert closure available” and “overhead utilities not removed” were clustered. Therefore, through examination of the images of Figures 3.6(a) to 3.6(c), knowledgeable construction personnel could be reasonably confident that the issue of readiness of working areas was the main culprit for late work commencement in certain locations, leading to work being performed in a non-contiguous work location sequence.

Figure 3.7 depicts in detail the formal thought process followed in exploring question groups 1 through 3 set out as part of the discussion on analytic reasoning in the previous section of the chapter. The right hand side of the figure demonstrates the thought process of interest, while the left hand side indicates how this thought process is executed in the visualization environment. The detailed documentation in this manner of such reasoning processes and the features required to support them constitute an important component of the data visualization environment design process.

### **3.6 Evaluation of and extensions to the current data visualization environment**

Presented in Table 3.1 is an evaluation of the current data visualization environment with respect to its conformance/non-conformance to the functional requirements established in sections 3.4.2 and 3.4.3.3 for a CM data visualization environment that assists users in monitoring/diagnosing construction conditions, time performance, and dependency relations without the use of explicit CM explanatory models. Presented in Table 3.2 is a list of images and interaction features currently supported or under development. Taken together, these two tables illustrate both the breadth and depth of issues that must be addressed for the design of a data visualization environment to be responsive to the analytical reasoning tasks targeted for support. From Table 3.1, five main non-conformance issues are: (i) vague definitions of time performance metrics; (ii) insufficient visual representations with which to depict the range of construction conditions associated

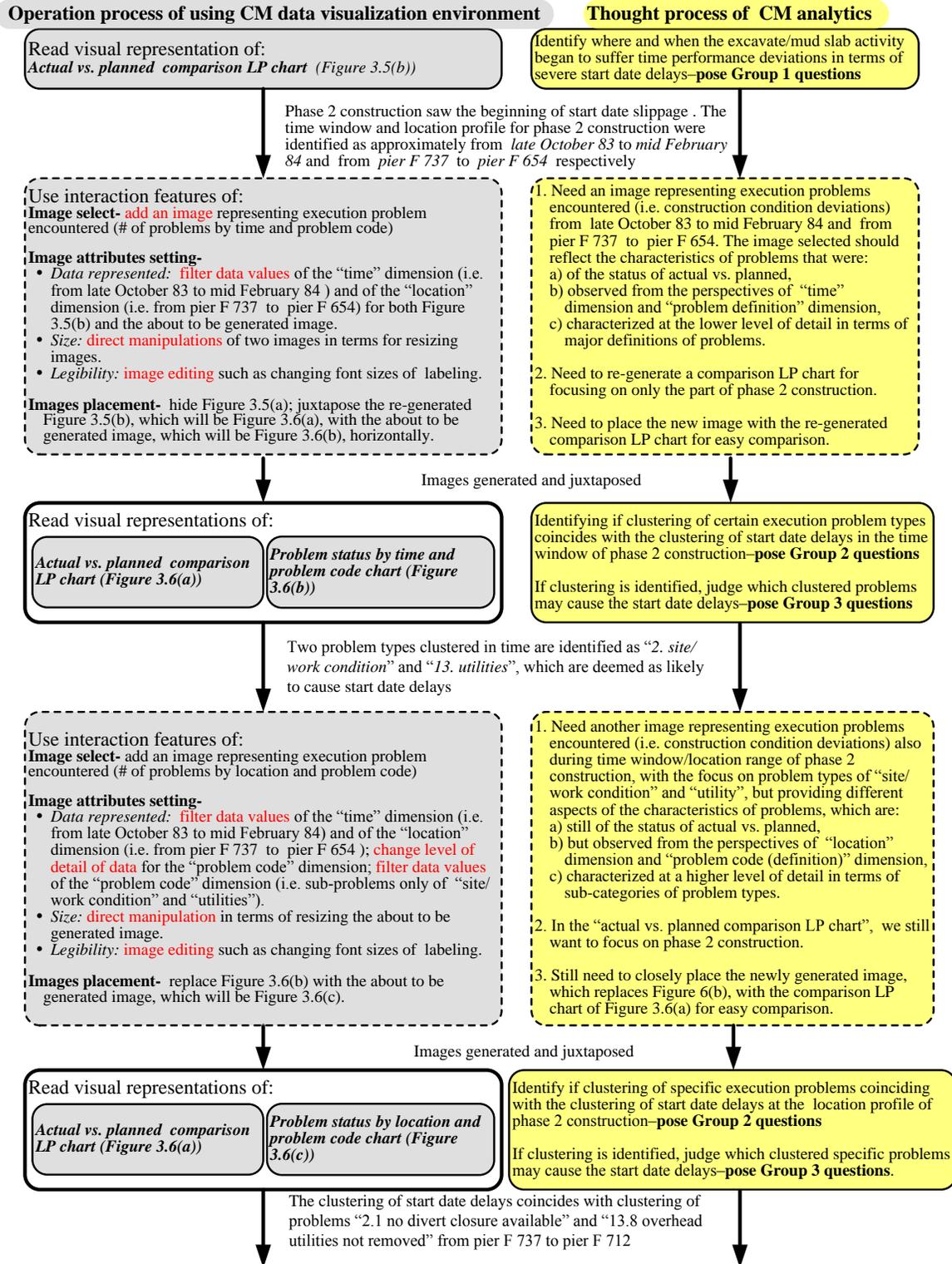


Figure 3.7 Parallel comparisons of operation process of using a CM data visualization environment and the thought process of CM analytics it supports reflected in Figure 3.6

Table 3.1 Data visualization environment requirements for visual CM analytics and conformance/non-conformance of current environment  
(*monitoring/diagnosing for time performance without the use of explicit explanatory CM models*)

	<b>Descriptions of requirements</b>	<b>Area of conformance</b>	<b>Area of non-conformance</b>
<b>General requirements for overall CM data visualization environment</b>	An extendable gallery of pre-coded visual representations of CM data encapsulating a range of construction conditions / performance mapped against the primary CM data dimensions of time, space, product, process, and participant in support of analytic reasoning.	Visual representations available for viewing can be found in Table 3.2.	Some potentially useful visual representations of construction conditions/performance measures that are related to analytics for monitoring/diagnosing time performance are not yet available (e.g. change view, environment view). Some candidate visual representations are also noted in Table 3.2.
	To represent different status states (e.g. planned versus actual) and specific construction differences (deviations) in status values of conditions/performance couplets.	All available visual representations such as the LP chart and the productivity chart can show data values of various possible status states.	Most of the available visual representations do not consider computing and visualizing value changes between status states.
	To provide interaction features for choosing the visual representation to be generated.	All available visual representations can be accessed through graphical user interface.	None
	To provide interaction features for adjusting granularity and value range of data dimensions for the visual representation of interest.	Some visual representations such as the LP chart partially allow adjusting data granularity in time and location dimension.  Some visual representations such as LP charts allow flexibly for changing data ranges (e.g. show only critical activities )	Some visual representations do not allow changing data granularity. For example, the project quantity image does not allow changing levels of detail for products.  Some visual representations do not allow changing data value range extensively. For example, the image visualizing production rates of activities do not allow filtering activities by project participants (e.g. trades).
	To provide interaction features for viewing several visual representations which portray different aspects of construction conditions / performance either sequentially or	All available visual representations can be viewed sequentially or simultaneously	None

	<b>Descriptions of requirements</b>	<b>Area of conformance</b>	<b>Area of non-conformance</b>
	simultaneously.		
<b>Specific requirements for CM visualization environment supporting time performance management</b>	<p>To visualize individual activity time performance, for which the time units of date and duration are used as performance metrics.</p> <p>To represent visually contractual requirements with regard to planned and actual time performance of activities.</p>	Visual representations such as LP charts allows users to see activity time performance metrics such as start dates, finish dates, and execution durations.	<p>The time performance metrics such as start dates, finish dates, and execution durations are not explicitly indicated in the visual representations such as the LP chart.</p> <p>The contractual requirements are not shown.</p>
	To associate construction conditions that can be represented visually with individual activities. The association can be explicitly specified by users or inferred because of their proximity in time and space, and possibly with participant and product.	Some visual representations such as "problem status" images can be filtered by activities, or the time window and the location range in which the activities were executed.	<p>Some potentially useful visual representations of construction conditions that are closely associated with activity execution are not yet available (e.g. location attributes such as condition of underground/overhead utility). Some candidate visual representations are noted in Table 3.2.</p> <p>Some visual representations such as "product quantity" can not be easily filtered by activities with which the products are associated.</p>
	To display sequences of activities and associated conditions. Such sequences can relate to one or more paths in a network model, the sequence of work of an individual trade, or the sequence of work at a specific work location.	The network diagram allows users to track the sequence of activities, activity by activity.	<p>Either LP chart or Gantt chart can not show both complex sequencing relationship and path in a easily understood form.</p> <p>Showing change of sequencing between various statuses (e.g. from the planned to the actual or from the planned to the re-planned) are not yet available.</p>

with activities; (iii) lack of visual representations depicting value changes between various status states of construction conditions and time performance; (iv) data query; and, (v) visualizing activity sequencing. In addition to checking conformance with the requirements determined previously, the process of applying the current CM data visualization environment to the case study helps to identify additional requirements not addressed in the original requirements analysis. Most of these additional requirements deal with usability of the environment in terms of: (i) readability of visual representations, and (ii) ease and quickness of use of the interaction features provided in support of the visualization requirements (e.g. adjusting data range, improving readability). The requirements of readability and ease and quickness of use are further elaborated upon in Table 3.2. One example of usability issues is the legibility of long labeling, which can be enhanced through being able to rotate the labels. Combining the check of conformance with observing issues encountered during use of the current CM data visualization environment, the two most important lessons learned from employing a top-down approach for enhancing the current status of the visualization environment relate to:

*1. Developing visual representations and associated image specific interaction features in support of a broad range of management functions:*

While commercial CM information systems in general have rich databases for models of projects and logging as-built data that reflect planned and actual performance and conditions, most of this data is displayed in textual form or tabular formats. Such displays can be very ineffective as visual representations, especially when large volumes of data are involved.

A palette of images can be organized using two classification schemes: construction conditions vs. construction performance; and process, product, environment, participant, and as-built views of a project. As demonstrated in Table 3.2, the current status of the CM data visualization environment including work in progress encompasses several images, mainly focused on visualizing the process and as-built aspects of project data with emphasis on certain tasks within the time management function. Nevertheless, the limited number of images designed and implemented to date makes cross referencing data for cer-

Table 3.2 Visualization features that are available in or in development process for the current CM data visualization environment

Data Views for Representing Construction	Images for Representing Data Views In Terms of Construction Conditions/Performance Measures			Interaction Features for Adjusting Level of Detail <sup>2</sup> of Data View in Primary Context Dimensions <sup>1</sup>		General Interactivity Capability Applicable to All Images	
	Conditions/performance measures visualized	Observed from primary context dimensions <sup>1</sup>	Inclusion of data status	Adjust data granularity <sup>2</sup>	Adjust data value range <sup>2</sup>		
Process View	Productivity	L	p, a, p/a, v	NA	P, D (Resource Type ), A	<ul style="list-style-type: none"> <li>Adjust visual attributes for visual representations, including elements of the visual representations. The fundamental visual attribute adjustments should include, but are not limited to:                             <ul style="list-style-type: none"> <li>Adjust scales in three physical dimensions (e.g. zoom in/out in an image, resize images).</li> <li>Adjust orientations in three physical dimensions (e.g. rotate a chart, labels, etc.).</li> <li>Adjust position in three physical dimensions (e.g. sort visual marks, move images for creating scenes).</li> <li>Adjust view port in three physical dimension (e.g. pan in a image).</li> <li>Adjust color coding/ saturation /transparency (e.g. highlight or dim a visual mark).</li> </ul> </li> <li>Intuitive user interfaces for using interaction features, including automating the execution of the concurrent interaction instructions applied to single or multiple views (i.e. coordination of interaction). The three main interaction features that should be supported by intuitive</li> </ul>	
	Production Rate	L	p, a, p/a, v	NA	A		
	Product Quantity	L	p, a, p/a, v	P	P, L,A		
	Schedule@(Bar Chart)	A & T	p, a, p/a, v	T	P, L, T, A		
	Schedule@(LP Chart)	A & T & L	p, a, p/a, v	T, L	P, L, T, A		
	Schedule@(Network Diagram)	D (Sequencing Logics)	p, a, p/a, v	NA	A		
	Resource Usage	T, L, T&L	p, a, p/a, v	T	D (Resource Type ), A, L		
	Site Congestion Index	L, T&L	p, a, p/a, v	T	D (Resource Type ), A, L		
	<i>Schedule Variance</i>	In progress					
Physical (Product/Location) View	<i>Location Attributes</i>	In progress					
	<i>Product Attributes</i>	In progress					
As-Built View	Problem Status	D (Problem code), T&L, D (Problem code) &L, D (Problem code) &T	a	D (Problem code)	T, L, A, PP, D (Problem code)		
	Workforce status	T	a	NA	T, PP		
	Activity Status	T, T&L	a	NA	T, A		
	Working conditions ( <i>weather, site</i> )	T	a	NA	T		

Data Views for Representing Construction	Images for Representing Data Views In Terms of Construction Conditions/Performance Measures			Interaction Features for Adjusting Level of Detail <sup>2</sup> of Data View in Primary Context Dimensions <sup>1</sup>		General Interactivity Capability Applicable to All Images
	Conditions/performance measures visualized	Observed from primary context dimensions <sup>1</sup>	Inclusion of data status	Adjust data granularity <sup>2</sup>	Adjust data value range <sup>2</sup>	
	<i>condition, etc)</i>					<p>GUIs should include, but are not limited to:</p> <ul style="list-style-type: none"> <li>▪ select and generate visual representations,</li> <li>▪ adjust data granularity/ range,</li> <li>▪ adjust visual attributes.</li> </ul> <p>The often used interaction features requiring coordination should include, but are not limited to:</p> <ul style="list-style-type: none"> <li>▪ select data then highlight visual elements (i.e. change color coding) representing the selected data,</li> <li>▪ apply the same data selection condition to multiple views,</li> <li>▪ navigate along the same data dimension (e.g. time dimension) in multiple views.</li> </ul>
	<i>Records</i>	In progress				
<b>Quality View</b>	NA					
<b>Change View</b>	NA					
<b>Environmental View</b>	NA					
<b>Organizational/ Contractual View</b>	NA					
<b>Cost View</b>	NA					
<b>Risk View</b>	NA					

• Keys for abbreviating primary context dimensions: **L**-location; **A**-activity; **T**-time; **PP**-project participant; **P**-product; **D**-condition/performance measure definition.  
 • Keys for abbreviating data status: **a** - actual; **p** - planned; **p/a** - planned versus actual; **v**- variance between the planned and the actual.  
 • Texts in italics or "NA" (except for contents of "general interactivity capability") represent features that are potentially useful but not currently available and need future work. The current CM data visualization environment has partial features regarding general interactivity capability.  
 • Footnotes:  
 1. Primary context dimensions refer to time (when), location (where), product (what), activity (how), project participant (who is responsible), and definitions of conditions/ performance measures (e.g. resource, problem, time variance).  
 2. Level of detail of data can be characterized by data range and/or data granularity. Data range relates to windows on a particular data set, with the dimensions of one or more ranges corresponding to time, location, participant, etc. Data granularity relates to the level of aggregation or disaggregation sought from a data set – e.g. only categories of problems vs. individual problem types within a category.

tain management functions difficult, thereby impeding the ability of users to conduct thorough CM analytics that stretch across key tasks within a certain management function and across management functions. For example, consider the physical components from the product view that are associated with an activity in the process view. They can be described in terms of a number of user specified attributes with both planned and actual values. Of particular interest here are scope attribute values (e.g. quantities). In addition, both products and activities map onto locations, which in turn can be described in terms of user specified attributes again with both planned and actual values. Of particular interest are location conditions attributes that could impact work – e.g. planned (unrestricted) versus actual (restricted) work face access, or planned versus actual presence of underground utilities. This data can serve several useful purposes. In the planning phase, it can be used for estimating activity duration as a function of scope and work conditions and in assessing risks and likely mitigation measures. In the execution phase, it can be used for managing scope. But it can also be used to help explain performance deviations during both the execution and post-project analysis phases of a project. Thus it would be helpful if this data could be translated into effective images so that it could be referenced quickly as part of the data exploration effort in search of likely causes for performance deviations.

When designing these images and determining how they can be used in scenes, data representations/transformations should be carefully reviewed in order to characterize and/or quantify the essential aspects of various construction performance measures and conditions in terms of their status (i.e. predicted, actual, and predicted versus actual) and value changes between status states. Also to be reviewed are the level of detail required, and the associations with context dimensions that would make this characterization useful. For example, for identifying causes of schedule variances, the definition of variances needs to include an activity's start date variance and its explicit predecessors' start date or finish date variances along with implicit predecessor condition states (planned versus actual) because in this way users can tell whether an activity's start date delay is completely or partially caused by a delay in one or more predecessors. Thus the design of these additional images and associated interaction features should be focused on supporting a clearly articulated CM analytics task. This could lead to the need for unique interaction features for operating on individual images, which involve mostly the operations of selecting visual representations and data content, in addition to the general

interaction features necessary to support the ability of users to browse quickly and clearly scenes of images in the visualization environment. This is elaborated upon as part of the second important lesson below. The detailed design of the individual images and ability to meld them into scenes along with both general and image specific interaction features corresponds to the bottom-up design approach described earlier.

*2. Incorporating general interactivity capability to enable users to quickly and clearly browse scenes of visual representations.*

Several interaction features are required to enable users to browse quickly and clearly visual representations central to CM analytics. First, as many capabilities as possible should be provided for users to control visual attributes in order to enhance readability and the composition of multiple representations. Examples of adjusting attributes include sorting visual marks in a chart, changing scale of axes, labels or visual marks in the chart, resize or move the chart, etc. And, the GUIs for effecting these interaction features and image specific ones need to be easy and intuitive to use.

One issue that can arise from the requirement to view multiple visual representations simultaneously (i.e. a scene) relates to the complexity of the interaction features required in terms of the ability to apply them to all or a subset of the representations in the scene. The quickness of viewing/comparing multiple representations will be lessened considerably if users need to interact with each image one by one. For example, using the current status of the CM data visualization environment, suppose users want to view the five charts shown in Figure 3.8 (one production rate chart, two activity status charts, and two weather condition charts) all within the same date range. During this time interval, activity duration difficulties were encountered for a few work locations. Hence the desire to compare various construction conditions with planned versus actual activity duration status in order to explore reasons for the duration performance experienced for selected pier locations, specifically for F702 and F712. Then, users would need to conduct the interaction action for zooming into or filtering to that date range separately for each chart. This would take five times longer than specifying the date range once and then automatically applying this data filtering specification to all five charts. Therefore in addition to including the interaction features mentioned previously, a mechanism for automatically and sim-

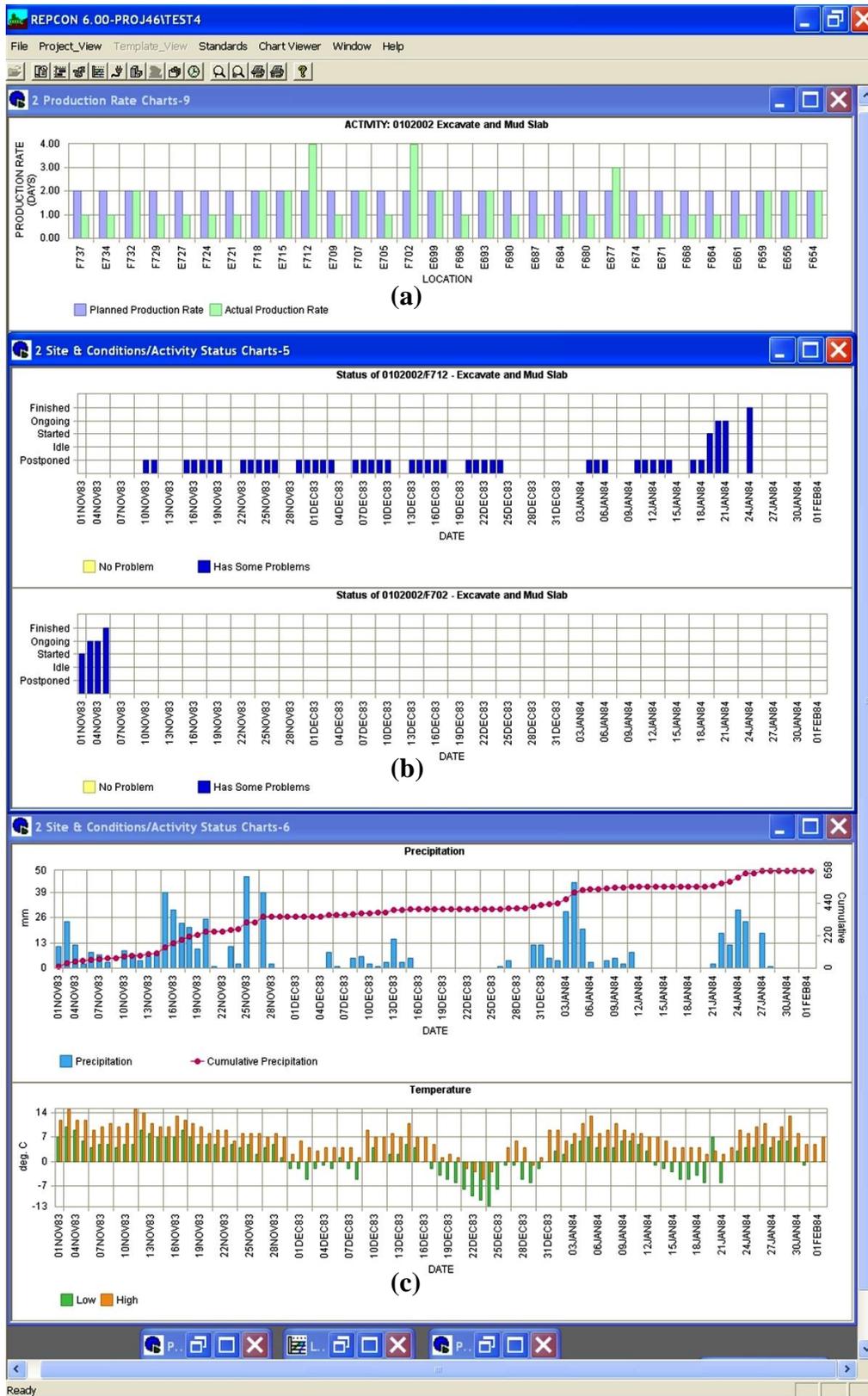


Figure 3.8 Generate and juxtapose (a) production rate chart, (b) activity status charts, and (c) environment condition charts

ultaneously coordinating shared interaction actions such as filtering data ranges of a data dimension to several images all representing that data dimension should be included.

### **3.7 Conclusions**

The primary purpose of CM analytics in conducting a specific task within a CM function, tasks across a CM function, and tasks across CM functions throughout the different phases of a construction project is mainly to understand the characteristics of construction conditions, performance, and dependency relations from three different perspectives: status, level of detail, and construction project dimensions. CM analytics requires and enables project participants to apply their implicit or tacit knowledge modeling capabilities as well as quantitative, explicit knowledge models. Visual CM analytics, as an adjunct to CM analytics, which is conducted through interacting with visual representations of construction management data, addresses the need of project participants to apply their own implicit CM models. A CM data visualization environment integrated with a CM information system can provide a CM analytics information technology infrastructure that enables project participants, from experts to novices, to journey quickly through complex CM data presented in easily-understood/natural visual forms. This facility allows them to gain important insights by applying their experience and knowledge to the interpretation of what they see in the images.

Emphasized in this chapter is the use of a top-down design process for identifying requirements for a comprehensive data visualization environment that treats a broad range of CM functions, tasks within each function, and relationships amongst functions. This process permits development of detailed specifications of individual visual representations and accompanying interaction features that can be pursued using a bottom-up design approach incorporating state-of-the-art design principles (Munzner 2009; Russell et al. 2009b) and overall environment design requirements gleaned from a top-down approach. We observe that the design/development lifecycle is an ongoing evolutionary one, because over time the visualization environment needs to be adapted to reflect lessons and insights learned through use, changing CM practices and the visual CM analytics demanded in support of them, and enhancements to the state-of-the-art of visualization utilities and technologies. The design process described in this chapter which adopts the concept of visual analytics and a top-down design perspective to revisit past

individual visual representations originally developed with a bottom-up approach to forge a comprehensive data visualization environment speaks to this observation.

As part of the goal of demonstrating the versatility and exploring the full potential of CM visual analytics, our ongoing research work is focused on how the functionality of a CM data visualization environment can be expanded to support a broader range of CM functions and processes with emphasis on the visual CM analytics that should be associated with them. The analytic concepts elucidated in this chapter, how they were applied to the case study presented, use of state-of-the-art visualization technologies and utilities, the testing of ideas developed through case-studies of full scale projects, and interaction on an ongoing basis with industry professionals are providing the foundation for this continuing work. A particular focus of current work is the identification and pre-coding of more visual representations of data (images and scenes of images) that are of potential importance to planning/predicting and monitoring/diagnosing/controlling construction conditions (e.g. variance analysis, cause-effect analysis), performance (e.g. punch list), and dependency relations (e.g. risk register). Part of this work involves incorporating interaction features into the environment so that users can on demand 1) choose and set contents of visual representations based on their own analytics strategies, 2) ensure visual representations are readable and understandable, 3) make contents of visual representations relevant to a specific visual CM analytics task easy to cross reference, and 4) make visual representations quickly accessible. Over the longer term, the research effort needs to be as fully inclusive as possible of findings by other research communities dealing with cognitive science, statistics, and computer science in order to fully understand the variables and relationships amongst them in order to maximize the benefits that can be derived from data visualization.

## **Chapter 4 Design of a Construction Management Data Visualization Environment: a Bottom-Up Approach<sup>5</sup>**

### **4.1 Introduction**

In this chapter, work related to the third phase of an ongoing research program on the visualization of construction management (CM) data is described. The specific focus is the role of a bottom-up design methodology as an essential part of a CM data visualization environment development process which also makes use of design guidelines (Russell et al. 2009) and a top-down design approach (Chiu and Russell 2011). The environment is meant to provide ready-to-use visualizations that help users answer specific CM questions relevant to a certain CM function/task while at the same time creating an environment architecture so that visualizations developed for particular CM purposes also support general CM analytics involved in a range of other CM functions/tasks. As used herein a visualization is comprised of one or more images of data along with their respective user interfaces. By "CM analytics" we mean CM user focused analytical reasoning with regard to identifying potential causes and/or effects from characteristics of CM variables (i.e. measurement dimensions of construction conditions and performance measures) related to the CM tasks/functions at hand. "Environment architecture" means an organization of thematic visualizations developed in a consistent way for addressing common visualization features required for supporting general CM analytics (e.g. the distribution of values of construction conditions and performance measures with respect to context dimensions). As a consequence, each visualization can be useful for multiple CM functions/tasks. Three visualization development case studies using location/schedule/as-built data from one actual project and one synthetic project derived from several actual projects for the CM function of time management/time performance control were used to help illustrate lessons learned about the value of a bottom-up design approach.

The bottom-up visualization designs described in this chapter adhere to the principles of the design process (abbreviated as design guidelines hereafter) identified in (Russell et al. 2009) and incorporate the top-down visualization requirements described in (Chiu and Russell 2011). The design guidelines set out the highest level principles in terms of how to apply state-of-the-art

---

<sup>5</sup> A version of Chapter 4 has been submitted for publication. Chiu, Chao-Ying, and Russell, Alan D. "Design of a Construction Management Data Visualization Environment: a Bottom-Up Approach"

data visualization (or visual analytics) techniques to CM domain data. The design guidelines contribute to formulating a structured CM data visualization design/development process that starts with analyzing analytical reasoning needs followed by applying fundamental knowledge about data visualization and CM data essential to the design of general CM data visualizations. For example, knowledge of effective visual variables can be applied when specifying visual encoding for visual representations during the bottom-up design process. Further, knowledge about CM data in terms of the concepts of data dimensions of project context, attributes of project context, and project performance dimensions helps articulate in a structured way the relationship between data dimensions and CM analytics when analyzing visualization requirements during both top-down and bottom-up design approaches.

The top-down design approach involves a design process that is focused on identifying CM analytics common to a certain scope of CM functions/tasks and hence visualization requirements that are responsive to them (e.g. CM analytics common to multiple CM functions as well as CM analytics applicable to an individual CM function). The analysis of visualization requirements includes scoping the CM variables involved in CM analytics and identifying general rules of how their characteristics can be observed on user demand in support of CM analytics. The scoping of relevant CM variables provides the breadth of sub visualizations that need to be developed in support of a CM function; analysis of how their characteristics should be observed in general results in identifying the common visualization features required. For example, from an overall CM perspective, the common CM analytics applicable to most CM functions/tasks involves exploring potential causes and/or effects amongst construction conditions and project performance measures. Using a multiple view modeling of a project, these construction conditions and performance measures can be attributes of the context dimensions of products, processes, project participants, etc. This recognition helps form both the scope and direction of CM data visualization environment development. Also, the characteristics of the aforementioned CM variables can be observed by project context dimensions, different levels of detail (abbreviated as LOD hereafter), and data status. Therefore, the common visualization features should support presenting values and distributions of construction conditions/performance measures of different data status in different context dimensions and at different levels of granularity. The "level of granularity" is more specific than level of detail and is specifically

used to represent the hierarchical levels of context dimensions. For example, the product context dimension has different levels of granularity of system, subsystem, element, etc. The level of granularity is abbreviated as LOG hereafter.

A "bottom-up approach" deals with the detailing processes needed to create actual visualizations. These processes include:

1. Analysis of specific CM analytics (i.e. CM questions) in relation to CM functions/tasks;
2. Identifying specific visualization requirements in terms of the CM variables involved in specific CM analytics and how their characteristics can be observed, on user demand, in support of these analytics. The analysis of how variable characteristics can be observed is conducted by first analyzing properties of the common top-down visualization features that fit the nature of particular CM analytics and/or CM variables. Additional new visualization features may also be sought. Making the interaction features for changing default settings of a visual representation as flexible as possible is key to enabling users to decide how best to observe the characteristics of CM variables on demand. In bottom-up design, the focus is on "what cannot" be changed if these changes do not add value to the usability and utility of the visualization;
3. Specifying required data representations and transformations, and visual and non-visual encoding attributes of a visual representation, plus essential interaction features reflective of the specific visualization requirements;
4. Implementing the specifications;
5. Evaluating the implemented visualization. Both designers and end users can utilize the "inspection evaluation method" (Amar and Stasko 2005; Ardito et al. 2006; Zuk et al. 2006) to identify any deficiencies in terms of non-conformance to the requirements/specifications and new requirements/specifications that help answer better the CM questions posed.

Testing the implemented visualization can make use of actual project data and well scaled synthetic project data reflective of actual projects in order to ensure the visualization tool can handle the realities associated with construction project data. The function of the bottom-up development process, from the perspective of developing a CM data visualization environment, is that through working on details and intensive exploration of the visualizations implemented,

lessons can be learned for refining design guidelines and/or top-down common visualization features as well as providing visualizations that are very responsive to the needs of CM personnel.

In the following sections of the chapter, the first three sections describe the analysis for three visualization designs for CM analytics applicable to the time performance control task. The first design treats the visualization of time variance performance measures. It provides a new visualization that is complementary to traditional schedule visualizations for supporting CM analytics directed at answering questions related to *"What are the values and patterns of behaviour of key performance indices for construction time performance and how can they provide useful insights into time performance vs. various project dimensions and assist in positing cause-effect hypotheses?"* Under the umbrella of the foregoing questions, particular emphasis is placed on *"understanding characteristics of time performance variance measures"*. The second design relates to a project's product view defined in terms of a Physical Component Breakdown Structure (PCBS) and visualization of the attributes of PCBS components. It provides a new visualization useful for supporting CM analytics with regard to answering questions related to *"What are the values and patterns of behaviour of construction conditions or other construction performance measures (referred to hereafter as construction condition/performance) and how can they provide useful insights into time performance vs. various project dimensions and assist in positing cause-effect hypotheses?"* Of particular interest is *"understanding characteristics of PCBS component attributes"*. The third design deals with time performance cause-effect visualization design. It provides a new visualization for supporting CM analytics that address questions related to *"What are the comparisons between planned and actual conditions and between conditions and performance?"* The specific goal is one of *"validating/invalidating cause-effect hypotheses between the time performance of individual activities and potential causes by comparing the two"*. Examined in the section that follows the sections that treat the visualization design analysis process are the case studies used to illustrate implementation of the designs that flow from the analysis process. Following the case study section, observations regarding the roles of the bottom-up design methodology in the CM data visualization environment development process and how it combines with the design guidelines and top-down design approach to form a visualization environment architecture that

supports CM analytics applicable to a broad range of CM functions are described. The chapter concludes with lessons learned during this phase of our data visualization research work along with suggestions for future work.

## **4.2 Visualization design 1--time performance measure variance visualization**

The CM analytics questions relevant to time performance control application framed as: "*What are the values and patterns of behaviour of key performance indices for construction time performance and how can they provide useful insights into time performance vs. various project dimensions and assist in positing cause-effect hypotheses?*" relate to CM analytics in terms of identifying potential causes of time performance deviation as a function of various data project dimensions. While many dimensions may affect time performance, the first preliminary step is a more focused and limited analytics one in terms of understanding whether reasons for time performance are a function of primary context dimensions (e.g. which activities and the participants responsible for them). Traditionally, a construction schedule in either traditional bar-chart format or linear planning chart format is used to visualize activity time performance measures such as start dates, finish dates, or durations. Superimposing or juxtaposing two or more versions of schedules allows users to make comparisons between planned/baseline and actual/updated baseline time performance. However, visualizations that address variances between planned and actual time performance have not been well developed and adopted for use. This has served as motivation for developing a case study focused on the design of a visualization to assist users in examining the quantitative variances of any two of several versions of a schedule in order to help identify whether the variances are a function of certain context dimensions. The case study also helps demonstrate application of a "bottom-up" design approach and to garner lessons from its use.

### **4.2.1 Visualization requirements**

For this case study, the specific CM analytics focus is on "understanding characteristics of variance values of time performance measures for identifying possible reasons for time performance deviations as a function of project context dimensions". The analysis of visualization requirements deals with identifying what CM variables are involved, particulars of the common visualization features derived from a top-down approach and additional new

visualization features that respond to the specific CM analytics involved. The analyzed visualization requirements are elaborated upon below.

#### **4.2.1.1 CM variables involved**

Metrics that are quantitatively measured in time units for characterizing time performance measures and hence time performance variance measures of a construction project have yet to be definitely defined. For example, although the Acumen Fuse project analysis system (Acumen PM 2010) organizes several industry standard metrics that are related to scheduling into several categories such as characteristics, duration, logic, lags, and status, not all of them are expressed in time units and explicitly defined as time performance measures. Another example is the Earned value management technique which uses a Schedule Performance Index (SPI) for representing whether an activity's execution is on schedule. However it is an unitless index derived from cost data and its time implication is not obvious. Use of the term "time performance" in this chapter is meant to suggest that time performance should be monitored and understood in the context of time units (e.g. durations, dates). Figure 4.1 provides an illustration for defining time performance measures including variance measures. The primary interest is at a finer LOD in terms of measuring time performance by dates and calendar days/working days ("the day being the most common basic unit for scheduling in construction"(Johnston 1981)), with emphasis on measurement at the activity level. By associating the activity dimension with other context dimensions such as project participants, images can be developed to characterize time performance as a function of these dimensions (e.g. time performance of project participants).

#### **4.2.1.2 How characteristics of time performance variance measures can be observed for identifying potential causes of time performance as a function of project context dimensions**

##### **1. Specifics for the top-down common visualization features**

- *Representing different status states of time performance variance measures:* For time performance measures, the status states include planned and actual. These states may change from planned to actual over different data versions (i.e. different schedule data updates). For time performance variance measures, the status states include "planned (cur-

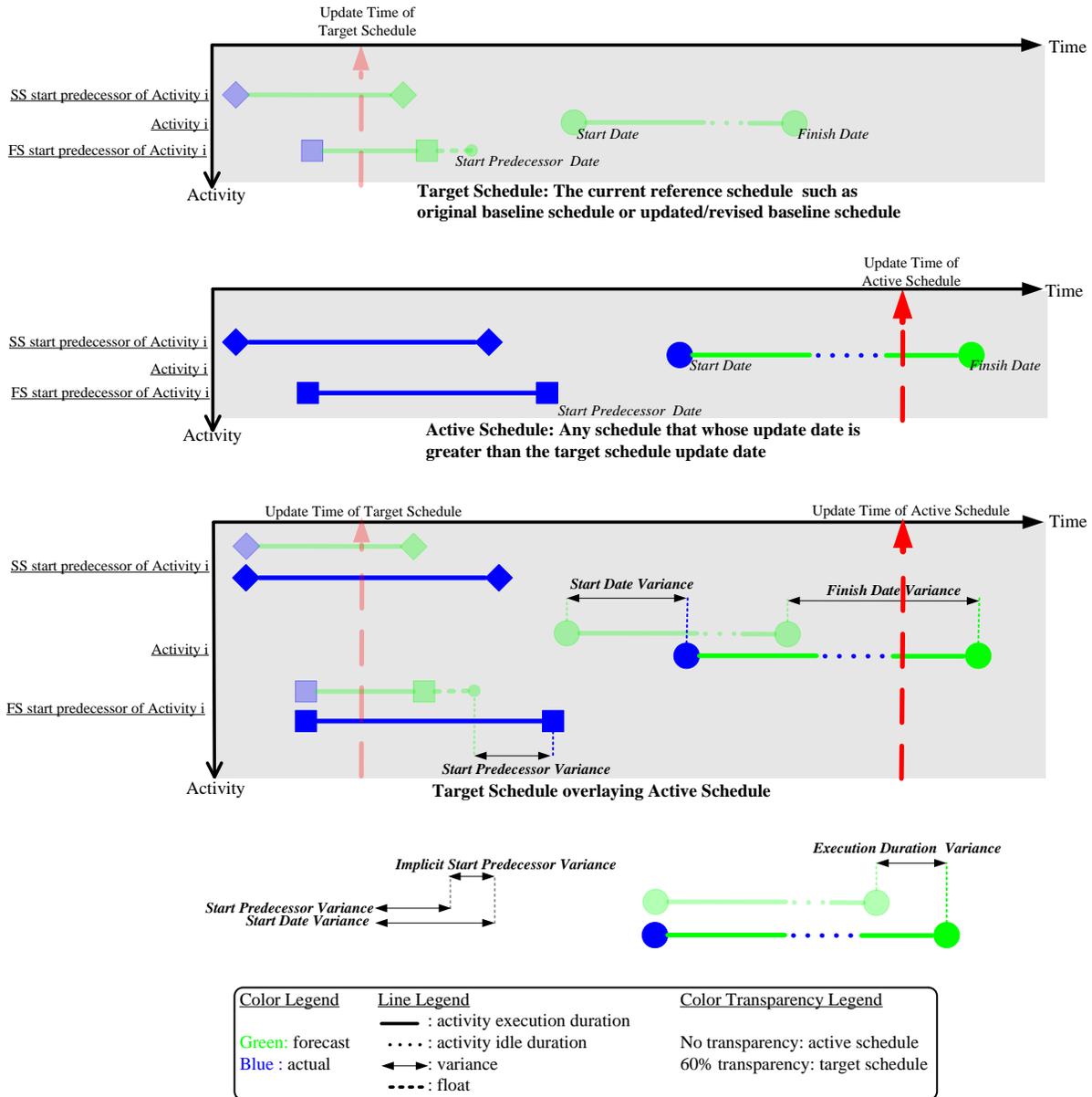


Figure 4.1 Illustration of the definitions of time performance and time performance variances

rent data version) minus planned (target data version)", "actual (current data version) minus planned (target data version)", and "actual (current data version) minus actual (target data version)". Planned status can be further categorized into "planned- early schedule ", "planned- late schedule", and "planned- scheduled dates".

- *Time performance variance measures mapped against data dimensions:* By seeing interesting patterns (e.g. anomalies, clustering, trends) of the distribution of values of time

performance variance measures in the context dimensions of definition of time performance variance measures, activity (process), product, location, project participant, and/or others (e.g. pay item, environment), one can infer likely causes of time performance deviations as a function of these context dimensions. For this kind of observation to be possible, however, a crucial requirement is that all relevant context dimensions need to be associated with the activity dimension.

- *Data granularity and data aggregation:* In this design case, values of time performance variance measures are computed according to the critical path method algorithm, or location-based hierarchical process modeling, and the definitions of variance measures are as illustrated in Figure 4.1. Therefore, each variance value computed corresponds to a certain activity, a certain location, and a certain variance type. Aggregating these computed variance values using common methods (i.e. SUM, MIN, MAX) such as summing variance values of a group of activities as the variance value for this activity group is not semantically meaningful. Meaningful "aggregation" for this case requires special computation algorithms that involve either true hierarchical scheduling, the use of hammock activities or summary activity structures, in order to obtain accurate variance values for collections or groups of activities that may not be completely contiguous in terms of their logic in the network. Because this is not our primary focus, it is not discussed further herein. Leaving aside the treatment of the foregoing special aggregation computation, in a time performance variance measures graphic, the computed variance values as well as their corresponding activities, locations, and variance types all need to be visually encoded. For example, if locations are not visually encoded, the variance values shown must be ones that aggregate variance values of all individual locations, which is not allowed in this design case. If users want to know and view the variance value of an activity group, no variance value aggregation should be computed and visualized for the group of activities; instead, all data points representing variance values of activities in this group should be shown. When generating the visualization of variance attributes, as a default time performance variance measures are mapped against the context dimensions of all LOG.
- *Data selection:* The time performance variance measure dataset can be selected by the values of variance, values of the context dimensions, and status states of time performance

measures. As a default, all values are included for visualization when generating the first image.

## 2. Features in addition to common visualization features

- *Consideration of representing variances for repetitive (multi-location) activities:* An activity that is repetitively executed through several locations is common to many construction projects. Visualizing a large number of activities executed repetitively through many of the same locations in one axis would take up a very large space along that axis. Therefore, activity and location dimension, if both are being visualized, should be mapped to orthogonal axes in order to avoid encoding an extraordinarily large dataset along one coordinate.
- *Considering the dimension of schedule update versions:* In theory, an activity's time performance needs to be updated regularly (e.g. schedules updated weekly or monthly). Therefore, an additional dimension of "schedule version" is useful for characterizing time performance variances. For example, showing several variance values computed by comparing a current schedule against several previous target schedules can show a trend in variance values, which in turn is an indirect indication of the effectiveness of management's actions for controlling time performance deviations.

### 4.2.2 Visualization design specifications

Visualization design specifications are concrete details about the choices of data representations/transformations, visual representations, and interaction features that can satisfy the visualization design requirements described in the foregoing subsection.

*Data representations and transformations:* Choices of data representations and transformations determine what data dimensions should be visually encoded, the LOD (data granularity and data range) of data, and what data can be derived through means such as data aggregation and computation of variances. Here, details of the definitions of time performance measures and time performance variance measures are first explained, followed by a summary in Table 4.1 of specifications with regard to data representations and transformations based on the analyzed visualization requirements.

Table 4.1 Summary of data representations/transformations for time performance variance measure visualization

<b>Measurement dimensions</b>	
<b>Construction conditions or performance measures</b>	Time performance variance measures
<b>Variance computation method</b>	There are several ways of computing variances that give different values: 1) variance values in working days or calendar days, 2) actual minus early planned or late planned, or 3) current minus target or target minus current
<b>Aggregation method</b>	No aggregation
<b>Context dimensions</b>	
<b>Primary project context dimensions</b>	Definition of time performance variance measures, activity, location, product, project participant, others
<b>Data granularity</b>	All LOG for applicable context dimensions as the default setting
<b>Data selection</b>	All data items of the context dimensions as the default setting
<b>Data status states</b>	
<b>Non variance status</b>	Planned and/or actual
<b>Variance status</b>	Planned (current schedule data version) - planned (target schedule data version) Actual (current schedule data version) - planned (target schedule data version) Actual (current schedule data version) - actual (target schedule data version)
<ul style="list-style-type: none"> <li>• <b>Data version (date)</b></li> <li>• <b>Paired data versions (dates)-- for variance</b></li> </ul>	Several combinations of paired schedule data versions (dates)

1. Measurement dimensions for representing time performance metrics of an activity

- **Start/finish date**
- **Governing start/finish predecessor date:** Latest date by which all start/finish predecessor requirements are fulfilled.
- **Duration:** This includes idle time (assumed to be zero in planned schedule and idle time duration recorded in the daily site report) and non-idle time.

2. Measurement dimensions for representing various time performance variance measures (i.e. variance type) of an activity between any two versions of a schedule

- **Start/finish date variance:** start/finish date given in the active schedule minus start/finish date given in a specified target schedule.
- **Start/finish predecessor variance:** start/finish predecessor date given in the active schedule minus start/finish predecessor date given in a target schedule.
- **Implicit start/finish predecessor variance:** start/finish date variance minus start/finish predecessor variance. Computed values have the following implications: 1) If the value is zero, the start/finish date variance is solely inherited from its predecessors, 2) if the value is greater/less than zero, it is likely that one or more unscheduled events are the cause of further start/finish date variance in addition to any variance caused by the explicit start/finish

predecessors.

- **Duration variance:** duration variance of an activity equals the active schedule duration minus the duration in a target schedule. This duration variance can be further broken down into: 1) idle time variance: idle time given in the active schedule minus idle time given in a target schedule, 2) non-idle time variance (i.e. extended working time): non-idle time given in the active schedule minus non-idle time given in the target schedule.

*Visual representations:* Determining possible choices of visual representations deals largely with how data dimensions should be mapped to visual variables (attributes) of selected visual marks (points, lines), labels, and gridlines. This mapping is termed "visual encoding". Other aspects of specifying visual representations involves labeling, providing keys (legends), providing titles, use of gridlines, and specifying default visual attributes of an image that are independent from visual encoding (e.g. scale of axis, which is not used as a visual variable to encode data values). The latter aspects are more related to usability while the former one of visual encoding that deals directly with presenting data values is central to the utility of a representation. As a consequence, only visual encoding is explained in more detail here.

In accordance with design guidelines that treat good visual encoding practice (e.g. choosing effective visual variables based on measurement scale) and the use of 3D virtual space, the line connecting scheme shown in Figure 4.2(b) illustrates how the applicable choices of visual encoding can be determined by mapping data dimensions on the left and right to visual variables on the middle. These are then applied to visual marks (the visual mark chosen in this design case is the bar, as shown in Figure 4.2(c)), labels, and gridlines except for the visual variable called "singular selection". If one context dimension such as the location dimension is mapped to "singular selection", the image can only visualize the dataset corresponding to the one item, e.g., a location, selected in "singular selection". By examining Figure 4.2(b), it should be clear that there could be many mapping alternatives. Each alternative creates different images and several of them may assist in answering different CM questions related to the umbrella CM analytics question of: "*What are the time performance variances as a function of variance type, project activities, project participants, locations, pay items, and/or other project context dimensions?*".

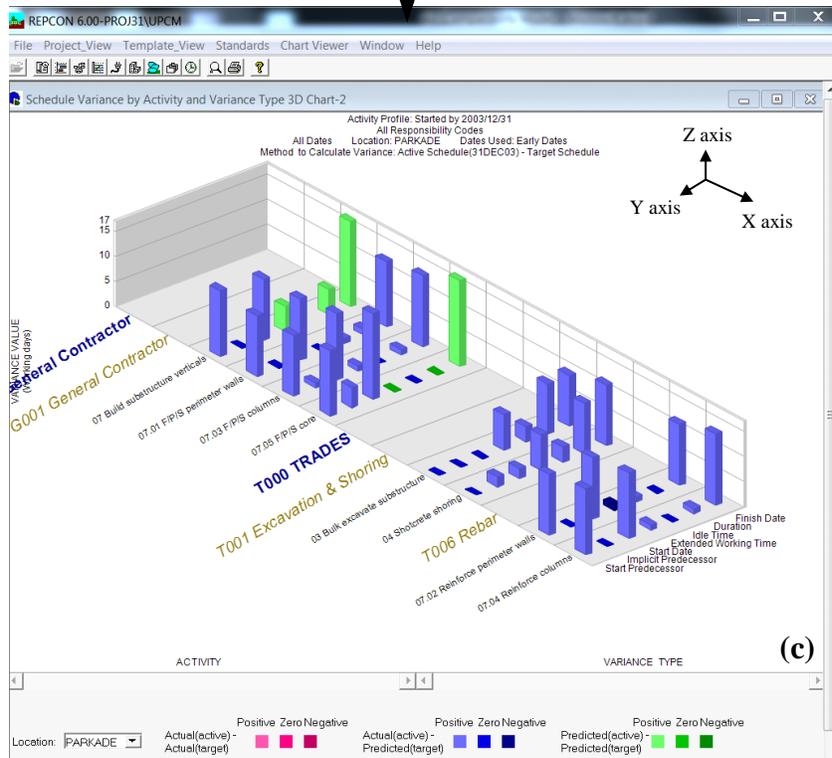
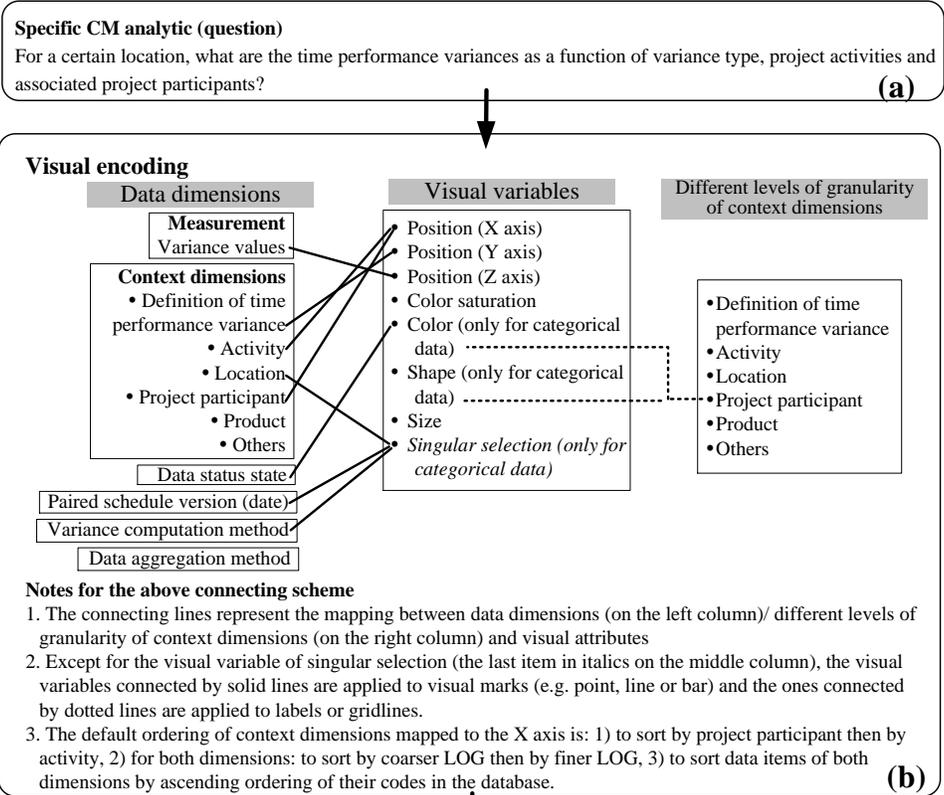


Figure 4.2 (a) A specific CM analytical reasoning task/question regarding time performance variance measures, (b) corresponding data dimensions and visual encodings, and (c) the generated visual representation

For users to continuously formulate and seek answers to various specific questions contained within the foregoing CM analytics question, interaction features should accompany a visualization to allow users to specify and change the default visual encoding in order to generate images using different visual encodings that may provide insights with respect to specific questions. For the purpose of distinguishing between the concept of "sub visualization" described below, a visualization that allows users to generate many images using different visual encodings is termed a "general visualization". An alternative or complementary design approach involves the analysis of certain specific questions contained within the umbrella CM analytics question and development of "sub-visualizations" for answering them. Usually more specific questions involve fewer data dimensions and therefore visual encoding alternatives for each sub visualization are greatly reduced. Then, when applied in the practice of CM, if users pose similar specific questions, they can directly choose to use these sub visualizations that involve only one or two images using different visual encodings instead of using a general visualization to explore many visual encoding alternatives by themselves. This way of identifying more specific CM analytics and designing "sub visualizations" in support of them constitutes part of the process of top-down design. To illustrate the foregoing, a more specific question with respect to time performance variance is framed as: "For a certain location, what are the time performance variances as a function of variance type, selected project activities, and/or associated project participants?". The visual encoding (i.e. mapping between data dimensions and visual variables) of a visual representation that responds to this analytical reasoning task is illustrated by the connecting line scheme seen in Figure 4.2(b) and the resulting image is shown in Figure 4.2(c).

*Interaction features:* A user interface (i.e. interaction features) is provided to users to allow them to change several of the default settings of a visualization, as described in the foregoing choices of data representations/transformations and visual representations. The ability to change these default settings helps users to continuously reformulate questions and helps to ensure readability of the visual representations. For example, users can change the visual encoding setting illustrated in Figure 4.2(b) by mapping the context dimensions of location, activities, and definitions of time performance variance measures to the visual variables of position (X axis), position (Y axis), and singular selection respectively for answering another specific CM question framed as : "For a certain variance type, what are the time performance variances as a function of

project activities and/or locations?". However, if a sub visualization designed to answer a very specific question (which usually involves fewer data dimensions) proves to be effective, including a user interface to allow users to change the default settings (e.g. settings for visual encodings and/or the context dimensions to be included for visualization) is not necessary. For example, using color coding to represent three variance status states (see the connection between "data status state" and "color" in Figure 4.2(b)) is deemed to be very effective because temporary human visual working memory can readily retain information color coded up to four colors (Luck and Vogel 1997); it is much less effective to use colors to encode tens to hundreds of items of location or activity items. Thus, it is unnecessary to allow users to change this default visual encoding. Furthermore, the effectiveness of the foregoing visual encoding can be maintained as long as three different colors are used to represent three different variance status states irrespective of what the colors are, thus it is not necessary to allow users to change the three default colors. The general rules of primary specification items that can be changed include:

1. Settings of measurement dimension, which include how to compute variances, how to aggregate data values, and selecting datasets by measurement values;
2. Settings of context dimensions, which treat the context dimensions to be included for visualization and their LOD (both data granularity and data selection);
3. Settings of data status state, which treat the status states and data versions (or paired data versions) to be included for visualization; and,
4. Settings of visual representations, which include the visual encodings to be used (e.g. change the ordering of activities in the image) and the visual attributes of the representations that are independent of the visual encodings (e.g. change the orientation of labels; change the scale of the X axis).

### **4.3 Visualization design 2--PCBS attributes visualization**

The second high level CM Question group relevant to the time performance control application is framed as: "*What are the values and patterns of behaviour of construction conditions or other construction performance measures (shortened as construction conditions) and how can they provide useful insights into time performance vs. various project dimensions and assist in positing cause-effect hypotheses?*" This question group relates to CM analytics in terms of

identifying potential impacts on time performance deviations as a function of various construction conditions. To prove or disprove a cause-effect relationship is a rather difficult task. Therefore, the initial analytical reasoning is more focused and limited in terms of identifying whether these construction conditions potentially impact the primary context dimensions and the attributes that characterize them. For example, relatively large concrete quantities required for certain products implies potential impacts on the product dimension (e.g. the quality control attributes of the concrete products), location dimension (e.g. the attribute of site access to the locations of the products), and activity dimension (e.g. the duration attribute of the concrete pour activities associated with the products). All of the foregoing potential impacts may directly or indirectly cause time performance deviations. While there are potentially many construction conditions in a construction project, to date only a few have been presented in visual forms by other researchers or software developers. Amongst many largely unexplored CM variables (e.g. skill level of various trades, communication performance, activity sequencing logic variance) are the attributes of products and/or the locations where the products are to be built. The term "product" as used here refers to the built components or products of construction processes. With the increasing use and adoption of built-product information modeling such as building product models (Eastman 1999) (i.e. building information model or BIM) or bridge information model (BrIM; (Shirole et al. 2008)), research efforts dealing with visualization of these rich information models are focused mostly on presenting their geometric aspects. Designing visualizations to assist users to conduct building/facility/infrastructure information analytics in support of CM by quickly going through product data or location attribute data is yet to be extensively investigated. This observation has provided the impetus for developing a case study focused on the design of a visualization to assist users in examining many aspects of product/location information, which also helps demonstrate application of a "bottom-up" design approach and derive lessons from its use. In this chapter the products and locations where the products are constructed are collectively referred to as a Physical Component Breakdown Structure (PCBS; (Russell and Chevallier 1998)). Essentially it is a product model tailored to the needs of construction management.

#### **4.3.1 Visualization requirements**

For this case study of visualization design, the specific CM analytics or analytical reasoning is focused on "understanding the characteristics of values of PCBS attributes for identifying

potential impacts on project context dimensions". The analysis of visualization requirements deals with identifying the CM variables involved and the specifics of the top-down common visualization features including the need for additional new visualization features that respond to the CM analytics at hand.

#### **4.3.1.1 CM variables involved**

The PCBS attributes are user-defined to characterize products and locations. For product attributes, their values correspond directly to a certain product at a certain location; for the location attributes, their values correspond directly to a certain location. Depending on how one models a project, product and/or location dimensions are closely associated with other context dimensions such as activity (i.e. process), pay item, and project participant. Therefore, product and location attributes and their values indirectly correspond to these other context dimensions. Product and location attribute values have two states - planned and actual. Computing variances between planned and actual product and location attribute values in some cases is tricky. One such case is when attributes do not have quantitative values. Therefore, there is a need to implement various computational routines to support data transformations for computing variances between actual and planned attribute values. For example, one way of computing variance values for non-quantitative attributes (e.g. the difference between "Yes" and "No") is to derive a new data definition such as "existence of variance/change", and its counts, count percentage, or level of variance (e.g. for the ordinal data of "good", "medium", and "high", the level of variance between "good" and "high" is 2 because they are two intervals apart) can be used as a measurement to quantify the differences. However, due to the complexity of this requirement, only computing variances for quantitative and singular value assignment attributes is considered herein.

#### **4.3.1.2 How characteristics of PCBS attributes can be observed for identifying potential impacts on context dimensions**

##### **1. Specifics for common top-down visualization features**

- *To represent different status states of PCBS attributes:* As stated previously, PCBS attributes can have both planned and actual values. These values are separate and somewhat independent of the time status values (planned (early, late, scheduled) and

actual) of the activities associated with the PCBS components of interest. Normally, the status state of a PCBS attribute variance would be the same as time performance variance measure (i.e. three status states). However, for the case that PCBS data is not updated regularly or only intermittently (i.e. only have original planned values and final actual values), the data dimension representing the status state of PCBS attribute variance can be neglected for visual encoding because there is only one status state (i.e. actual (final data version) minus planned (original data version)).

- *Mapping PCBS attributes against other data dimensions:* By seeing interesting patterns (e.g. anomalies, clustering, trends, similarities/differences) in the distribution of PCBS attribute values in the context dimensions of definition of PCBS attributes, activities, products, locations, time, project participants, and/or others (e.g. pay items, quality criteria), one is better positioned to infer how the PCBS attributes may have impacted these dimensions. For this kind of observation to be possible, however, a crucial requirement is that associations between product and location dimensions need to be forged with the other context dimensions.
- *Data granularity and data aggregation:* For PCBS attributes, the finest LOD of data is the one users entered into the CM information system by inputting and associating attribute values with an item in the product and/or location dimensions of a certain LOG (e.g. LOG for the product dimension, from coarser to finer, includes system level (e.g. structural), subsystem level (e.g. verticals), element level (e.g. columns, walls)). When generating PCBS attribute visualizations, as a default the PCBS attributes are mapped against the context dimensions of all LOG. If the attribute values are quantitative and singular in their assignment, they can be aggregated by the common method of SUM. For example, concrete quantities required by all products on the 2nd floor of a building can be aggregated to become concrete quantities for the 2nd floor. Whether to and how to aggregate non-quantitative attribute values requires further analysis and is not discussed in this chapter.
- *Data ranges:* A dataset of PCBS attributes can be selected by the values of attributes, values of the context dimensions, and status states of PCBS attributes. As a default, all values are included for visualization when generating the first image. For large scale datasets, the resulting image needs to be filtered in order to enhance readability (see

Figures 4.3(c) and 4.3(d).

## 2. Features in addition to common visualization features

- For the definitions of attributes to more accurately capture the rich information/knowledge that accompanies most PCBS components, there could be additional ways of assigning attribute values. For example, the “equal to” value assignment is for attributes such as product quantities; the “within the range of” value assignment is for attributes such as the time window during which work can take place at a location; there are instances where a multiple value assignment for location attributes is required - e.g. "what are the different utilities encountered at a location?" Therefore, the design of the visualization needs to accommodate visualizing attribute values in a variety of ways (later on it is concluded that such a capability should be a common feature applicable to other visualizations in a CM visualization environment).
- The visual representations need to reflect the structure of the common knowledge models about a built product, which include the hierarchical modeling of a PCBS (both product components and locations) and relational relationships between product components and locations. While tree-like images (Herman et al. 2000) are common for showing hierarchical data and matrix-like images (e.g. Design Structure Matrix (Keller et al. 2005)) are intuitive for presenting one-to-one relational data relationships, the visual representations of PCBS attributes should be capable of simultaneously showing both hierarchical and relational relationships amongst locations/products. Therefore, product and location dimensions are mapped to two orthogonal coordinates (similar to the image format of a design structure matrix) and product/location dimensions of all LOG are simultaneously mapped to these two coordinates and ordered from coarser to finer in order to mimic a hierarchical tree structure (see the X axis in Figure 4.3(d) where built products are labeled in order from the coarsest granularity of foundation related products of "foundation", then "pile foundation", then the finest granularity of "piling"; each of the foregoing labels is repeated twice because one corresponds to the planned values of "piling length" while the other corresponds to the actual values (the empty columns seen in Figure 4.3(d) reflect the fact that actual values have not yet been entered).

### 4.3.2 Visualization design specifications

*Data representations and transformations:* Choices of data representations and transformations determine the data dimensions that should be visually encoded, LOD of data (data granularity and data range), and how data should be derived such as data aggregation and computation of variances. Specifications with regard to the applicable data representations and transformations are summarized in Table 4.2.

Table 4.2 Summary of data representations/transformations for quantitative PCBS attribute visualization

<b>Measurement dimensions</b>	
<b>Construction conditions or performance measures</b>	Quantitative and singular valued PCBS attributes
<b>Variance computation method</b>	Actual values minus planned values
<b>Aggregation method</b>	Sum
<b>Context dimensions</b>	
<b>Primary project context dimensions</b>	Definition of PCBS attributes, activities, locations, products, project participants, others
<b>Data granularity</b>	All LOG for applicable context dimensions as the default setting
<b>Data selection</b>	All data items of the context dimensions as the default setting
<b>Data status states</b>	
<b>Non variance status</b>	Planned and/or actual
<b>Variance status</b>	The same as the ones for time performance variance measures unless the PCBS data is not updated regularly or only intermittently
<ul style="list-style-type: none"> <li>• <b>Data version (date)</b></li> <li>• <b>Paired data versions (dates)-- for variance</b></li> </ul>	The same as the ones for time performance variance measures unless the PCBS data is not updated regularly or only intermittently

*Visual representations and interaction features:* In this design case study, the same problem of having many visual encoding alternatives similar to what was encountered in the time performance variance measures visualization design case is also faced by the designer. Sub visualizations related to PCBS attributes can be developed for ready use (i.e. no need for users to explore visual encoding alternatives) in order to answer more specific CM questions surrounding PCBS attributes. For example, a more specific question into PCBS (product) attributes is "For a certain product attribute, what is its distribution of planned and/or actual values in the product and location dimensions so that one can identify potential impacts on products and/or locations?" The corresponding visual encoding of a visual representation that responds to this analytical reasoning task is illustrated by the line connecting scheme seen in Figure 4.3(b). The resulting image is shown in Figure 4.3(c). Figure 4.3(c) illustrates an example where the data space is large and hence not easy to read. However, it reflects the design principle of showing the full

**Specific CM analytic (question)**  
 For a certain product attribute, what is the distribution of its planned and/or actual values in the product and location dimensions so that one can identify potential impacts on products and/or locations? **(a)**

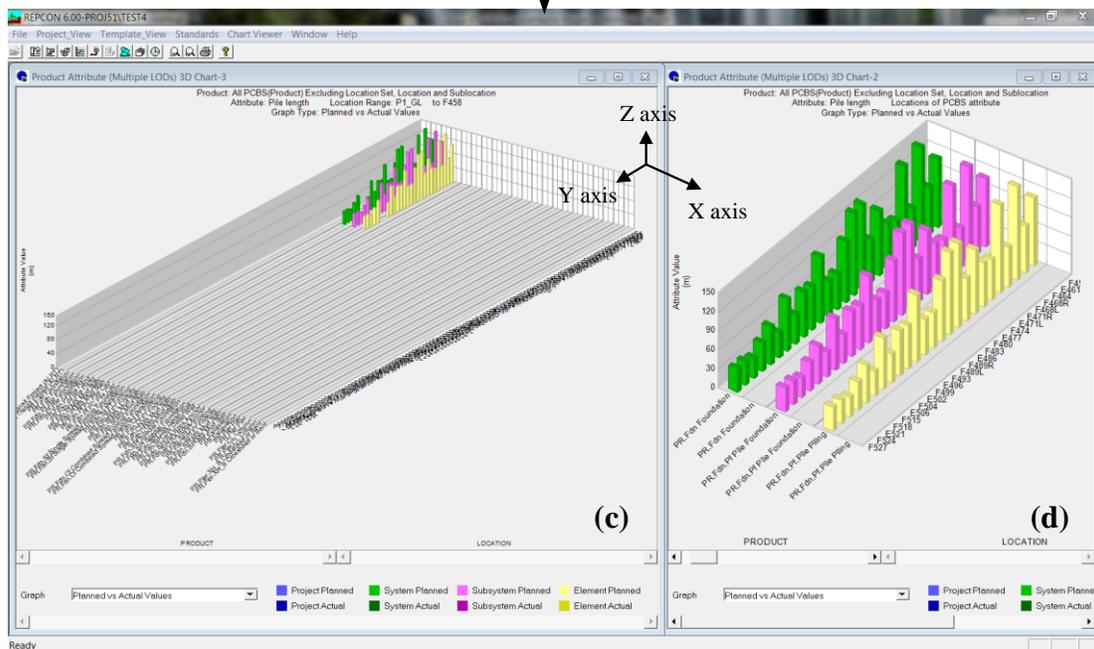
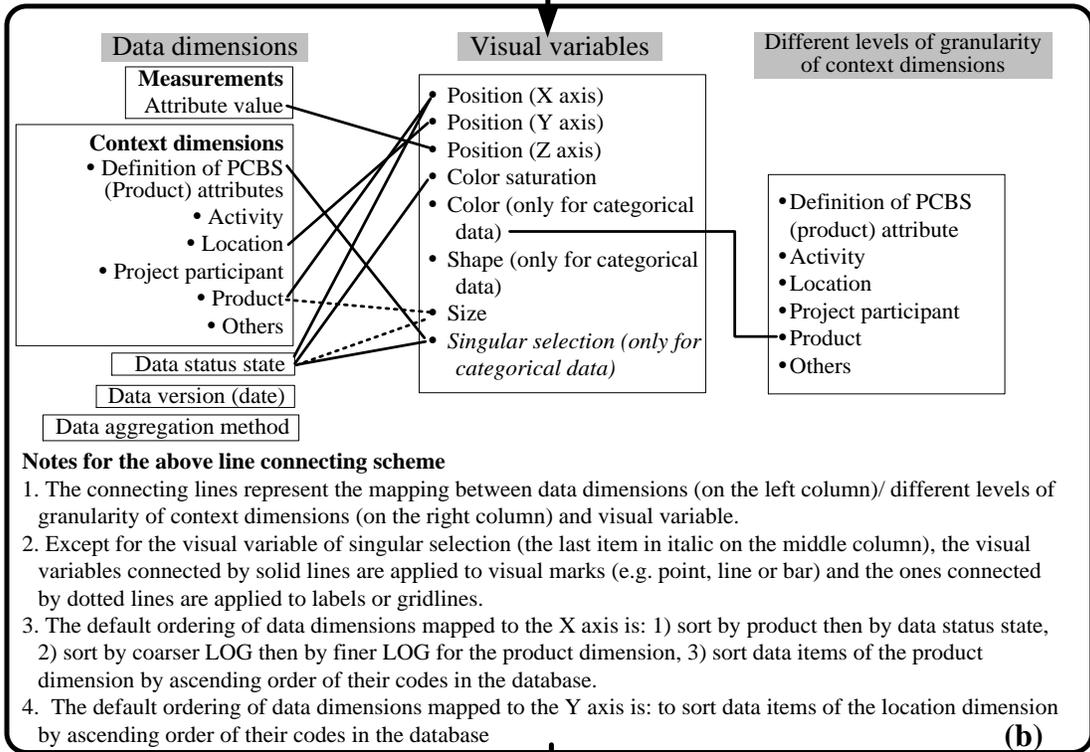


Figure 4.3 (a) A specific CM question related to product attributes, (b) corresponding data dimensions and visual encodings, (c) the generated visual representation for the entire data space, and (d) the generated visual representation for an area of interest in (c).

range of the data as a default setting and then allowing the user to pinpoint areas of interest by applying relevant filters to create an image that is more readable as shown in Figure 4.3(d). As to the interaction features that should be available to users, the principle in terms of providing user interfaces so they can change certain default settings of data representations/transformations and visual representations is the same as the one described in the previous design case.

#### **4.4 Visualization design 3--time performance cause-effect visualization**

The third high level CM Question group relevant to the time performance control application is framed as: "*What are the comparisons between planned and actual conditions and between conditions and performance that support project participant causal models to explain the time performance of a particular collection of activities related to one or more project dimensions?*"

This question group addresses the analytical reasoning involved in checking whether certain construction conditions are potential reasons for time performance deviations. The end users' cause-effect hypotheses in terms of what construction conditions might cause time performance deviations may be formed based on their own causal models or from investigating project data as part of group 1 and/or group 2 questions (e.g. identify time performance variances associated with certain trades and then determine if one or more attributes characterizing these trades may be the root cause of the variances). When an hypothesis is formed, users need to perform rigorous checks to validate or invalidate it. This checking can be done by comparing temporal values (planned, actual, planned vs. actual, or actual minus planned values) of activity time performance measures with those of construction conditions. The purpose is to see whether there are abnormalities in the temporal values of construction conditions, which occurred before or at the same time when the time performance deviations developed. The abnormalities may arise from abnormal planned values/patterns, abnormal actual values/patterns, or undesired values/patterns changes between planned and actual. This kind of comparison can provide evidence that better satisfies stricter causation requirements in terms of whether there are effects (e.g. time performance deviations of an activity as an effect) and whether the effects, if they exist, developed at the same time or after the causal conditions occurred (e.g. undesired changes between planned values and actual values of construction conditions as causes) (Li 2009). The conceptualization of the foregoing reasoning scheme is illustrated in Figure 4.4. In this chapter we have articulated this analytic strategy and formalized the design of CM data visualizations in

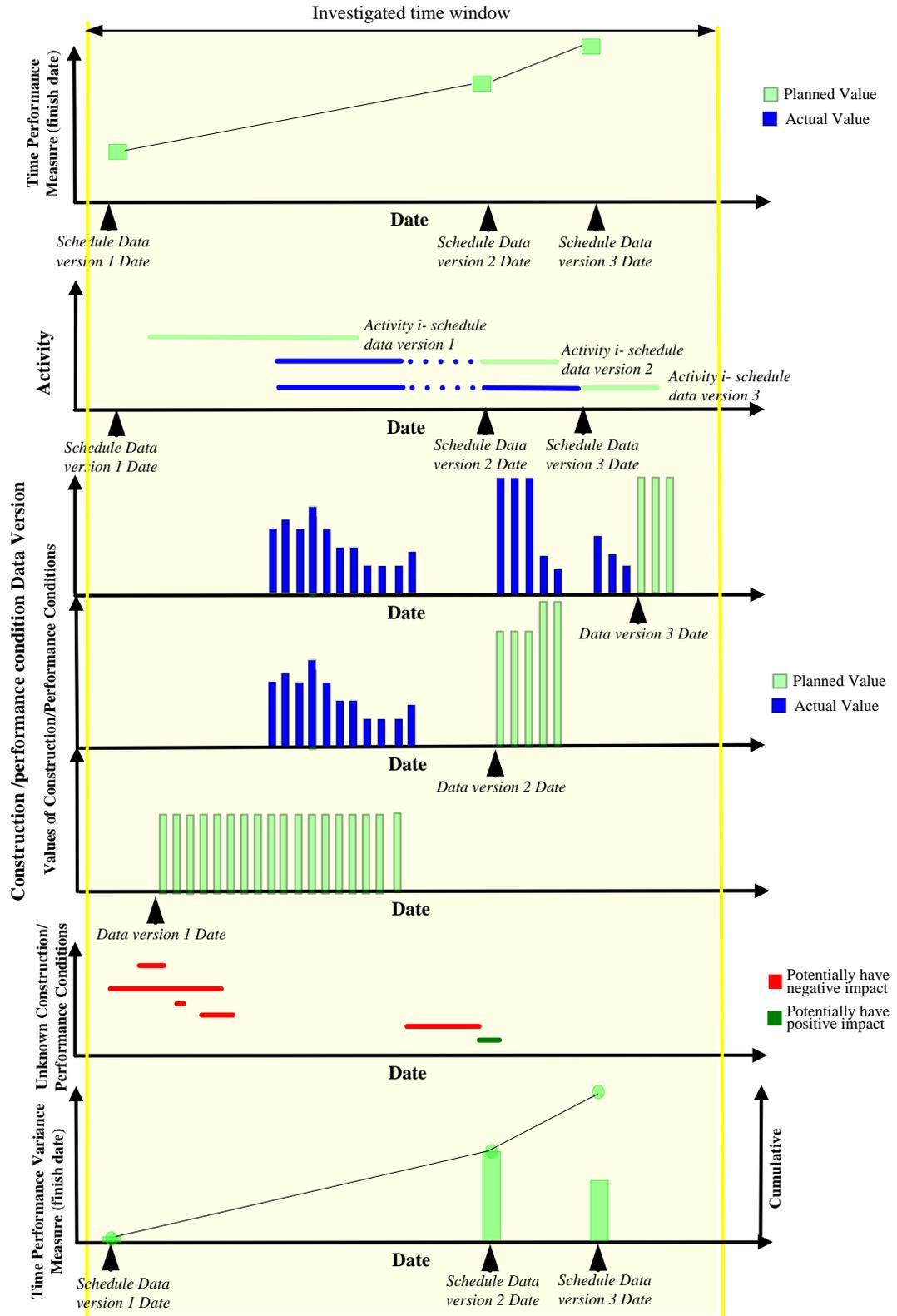


Figure 4.4 Illustration of concept of cause-effect reasoning for identifying potential reasons for time performance.

support of cause-effect hypothesis validation in the CM context. Thus, a case study focused on designing visualizations that allow users to compare temporal values of construction conditions with temporal values of individual activity time performance measures by "day" is used to demonstrate and learn from the application of a "bottom-up" design process. Comparing on a "weekly" or "monthly" basis may be more suitable for finding reasons for time performance at the activity group or project level because their execution usually spans weeks, months, or years.

#### **4.4.1 Visualization requirements**

For this case study of visualization design, the specific CM analytics focus on: 1) understanding characteristics of activity time performance measures for identifying deviations and their timing, 2) understanding characteristics of construction conditions which are relevant to the activity for identifying abnormalities and their timing, and 3) understanding whether the abnormalities occur before or at the same time as the time performance deviations thereby providing evidence that may not be definitive by itself for satisfying causation requirements. The analysis of visualization requirements deals with identifying the CM variables involved and the specifics of the top-down common visualization features or additional new visualization features needed to respond to the underlying CM analytics. The visualization design explained in the following subsection is focused on the individual visualizations presenting the construction conditions that support the second type of CM analytics (i.e. understanding characteristics of construction conditions which are relevant to the activity for identifying abnormalities and their timing). Similar design features also apply to single image visualizations presenting the time performance measures in support of the first type of CM analytics (i.e. understanding characteristics of activity time performance measures for identifying deviations and their timing). Another aspect of visualization requirements is that single image visualizations that support the first and second types of CM analytics can be juxtaposed and coordinated to create a multi-image visualization in support of the third type of specific CM analytics described in this section. The analyzed visualization requirements are itemized and explained in what follows.

##### **4.4.1.1 CM variables involved**

Construction performance/condition attributes are mainly ones that characterize important project context dimensions (e.g. activity attributes, PCBS component attributes, project

participant attributes, environment attributes). Defining them, including the computation of variance values, needs a separate analysis for each attribute type (quantitative, linguistic, Boolean, Date) and is not discussed here. However, for creating visualizations to support the analytical reasoning described in the beginning of this section, the following common features are required:

1. *Temporal values of time performance measures and construction conditions and their time stamps*: The temporal values of time performance measures and construction conditions, planned or actual, can be time stamped by their date of occurrence (i.e. occurring in time). For the ones that do not have the characteristic of "occurring in time" such as the planned vs. actual quantity of a certain product or the planned vs. actual duration of an activity) or the date of occurrence is difficult to know, data version/recording date can be treated as their date of occurrence. The definition of time stamps (i.e. date of occurrence) is important in this design case because if one wants to compare values of several construction conditions with values of time performance measures along the time dimension, the values have to meaningfully correspond to certain points in time, and the unit of time is at the LOG of day.
2. Construction conditions in relation to the individual activity under investigation: When modeling a construction project, several project context dimensions can be associated with one another by users (e.g. associate products with activities, quality requirements, etc). An important reason that users establish the associations amongst modeling components is because there are known or potential cause-effect relationships amongst the attributes that characterize project context dimensions, and hence they can help to predict or explain construction performance. Therefore, if items in the various project context dimensions are associated with the relevant activities, their corresponding attributes (including those of an activity itself) are construction conditions that are potential reasons for the time performance of an activity. For ease of explanation in the following discussion, the foregoing construction conditions can be referred to as "having direct associations with the activity dimension". As to other construction conditions that do not have direct associations with the activity dimension, they still can be useful for exploring reasons for time performance and can be indirectly related to activities through time windows/locations of the activities or a user's understanding of how they are related to activities.

#### 4.4.1.2 How characteristics of construction conditions related to a certain activity can be observed for identifying abnormalities and their timing

##### 1. Specifics for the top-down common visualization features

- *To represent different status states of construction conditions:* Depending on the kind of construction conditions, some have both planned and actual status states (product quantities, activity sequences) while others have only an actual status state (e.g. problems encountered, change orders). Some construction conditions such as environment conditions (e.g. weather related data such as temperature and precipitation) theoretically can have both forecast (e.g. planning assumptions) and actual status, however in practice usually only the actual states are recorded. For construction conditions that have both planned and actual status states, the status state changes over different data versions (dates). Thus, the data version dimension needs to be considered when representing status states. For construction conditions that only have an actual status state, since this state will not change there is no need to visually encode it and data version (date) can be neglected.
- *Construction conditions mapped against context dimensions:* By seeing anomalies, clustering, trends, and similarities/differences in the distribution of construction condition values in the context dimensions of the construction conditions definition, activity, location, and occurrence time, one can identify abnormalities in temporal values of the construction conditions and their timing that correspond to a certain activity/location combination. The occurrence time context dimension refers to the "time stamp" dimension discussed previously.
- *Data granularity and data aggregation:* The finest LOD of data and whether and/or how to aggregate data depends on the particular construction condition. For construction conditions "having direct associations with the activity dimension", a common principle is that data at its finest LOD should have its temporal values correspond to a day and an activity/location combination in support of a design case that is focused on finding reasons for the time performance of individual activities. For the case that construction conditions are not directly associated with the activity dimension, users need to specify an appropriate LOD that is equivalent to that of an activity/location combination. When generating a visualization, as a default construction conditions are mapped against the context dimensions of all LOG except for the occurrence time dimension for which LOG equals

"day".

- *Data ranges*: As seen in Fig. 4.4, for conducting cause-effect analysis in search of reasons for time performance, users need to first specify the activity/location combination of interest and the time window they want to examine. If construction conditions are directly associated with an activity/location combination, as a default the data range is determined by filtering the construction conditions dataset by the activity/location combination and the "time stamp" that is within the time window of interest. Otherwise, in addition to the time window specified, users may need to specify a data range based on their knowledge of the project at hand and past experience in terms of how the construction conditions encountered may be related to the selected activity/location combination.

#### **4.4.2 Visualization design specifications**

The analysis method for determining the specifications of the visualization design is the same as those used in the previous two design cases. The focus is on individual visualizations presenting the distribution of values of a certain construction condition or performance measure in the time domain. For any individual visualization, the required data representations/transformations are summarized in Table 4.3. Visual encoding specifications for a visual representation that responds to an exemplary CM question such as "*Are there abnormalities in the temporal values of actual execution status and problem status (i.e. counts of occurrence) that associate with a certain activity/location combination, and what is their timing?*" along with an example image are shown in Figure 4.5. Images representing the distribution of values of other construction conditions and time performance measures in the time domain have formats similar to the image seen in Figure 4.5. Two or more of these images can be juxtaposed for comparison. As to the kind of interaction features that should be made available to users, the same principle as the one described in the schedule visualization design case (e.g. provide interfaces to allow users to change certain settings of data representations/transformations and visual representations) is applied here. However, because many images could be generated and interacted with by users, changing the same settings for all images should be coordinated (i.e. select the changes once and then apply to all images) to enhance efficiency. For example, if a user wants to focus on a narrower investigation time window, the time filter for one image should be automatically applied to others.

Table 4.3 Summary of data representations/transformations for visualizing the distribution of construction conditions versus time

<b>Measurement dimensions</b>	
<b>Construction conditions or performance measures</b>	Mostly the attributes characterizing context dimensions such as activity, product, location, project participant.
<b>Variance computation method</b>	Depending on respective construction condition/performance
<b>Aggregation method</b>	Depending on respective construction condition/performance
<b>Context dimensions</b>	
<b>Primary project context dimension</b>	Definition of construction condition/performance, activity, location, occurrence time
<b>Data granularity</b>	All LOG for the activity dimension; "location" level for the location dimension; "day" level for the occurrence time dimension
<b>Data selection</b>	User specified time window and activity/location combination based on user knowledge of the project at hand and past experience in terms of how the construction conditions encountered may be related to the selected activity/location combination
<b>Data status state</b>	
<b>Non variance status</b>	Planned and/or actual
<b>Variance status</b>	The same as the ones for time performance variance measures unless the construction condition data is not updated regularly or only intermittently or has only an actual status
<b>Data version (date) Paired data versions (dates)-- for variance</b>	The same as the ones for time performance variance measures unless the construction condition data is not updated regularly or only intermittently or has only an actual status

**Specific CM analytic (question)**  
 Are there abnormalities in the temporal values of actual execution status and problem status (i.e. count of occurrence) that associate with to a certain activity/location combination, and what is their timing? (a)

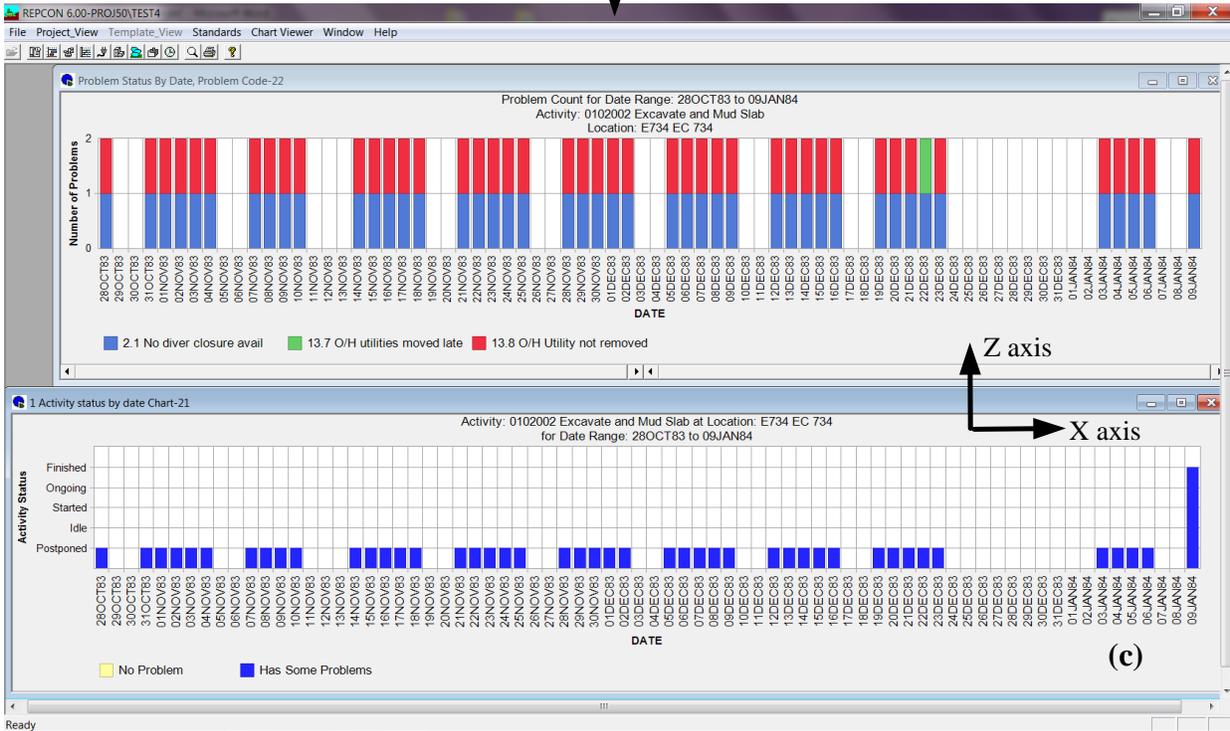
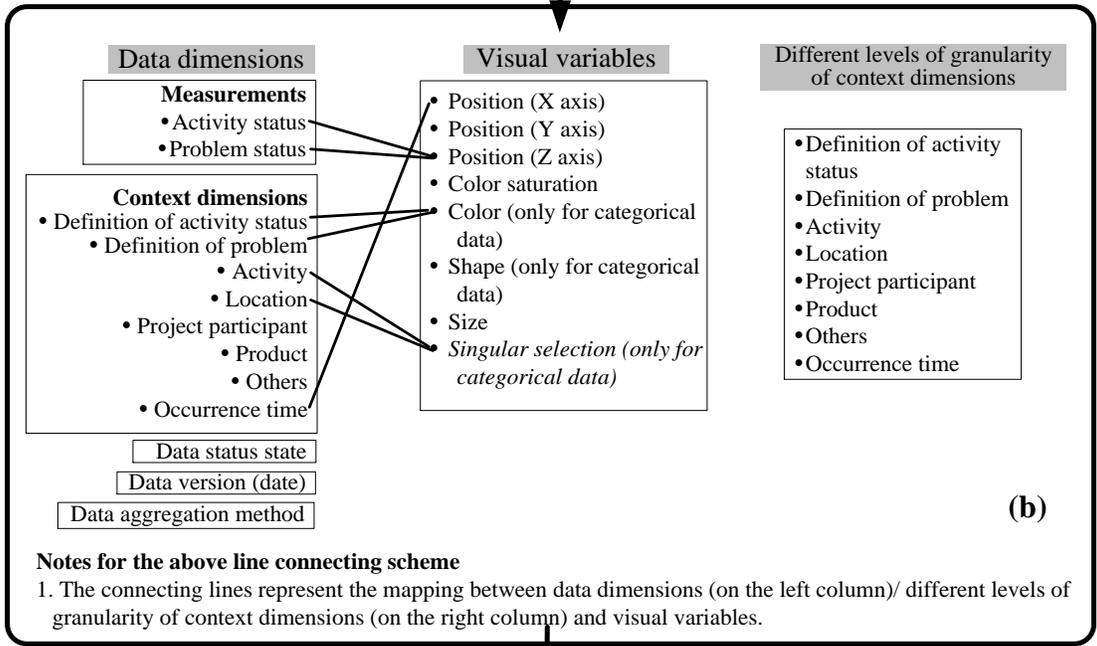


Figure 4.5 (a) A specific CM question pertaining to the construction conditions of activity execution status and problem status, (b) corresponding data dimensions and visual encodings, and (c) the generated visual representation.

## **4.5 Use of a state-of-the art CM data visualization environment**

Following the design of the three foregoing visualizations and their implementation in a research prototype CM data visualization environment, case studies using two sets of project data were carried out. The primary purpose of the case studies is to demonstrate the merits of an environment architecture (of a CM data visualization tool) created using a development methodology that systematically follows design guidelines and both top-down and bottom-up development processes. Benefits sought should be that the visualizations and accompanying features developed are responsive to common CM analytics (i.e. identifying potential causes and/or effects from characteristics of CM variables) as well as specific ones (i.e. identifying potential causes and/or effects from characteristics of particular CM variables related to specific CM analytics and their supported CM functions/tasks). The secondary purpose of the case studies is to showcase how insights into additional desired visualization features can be obtained. These additional features are targeted at both enhancing usability and utility (i.e. better support for CM analytics) and ideally they could become common features applicable to other visualizations. Our experience is that they can be identified only when a real visualization is implemented and applied to real world project data. The reality is, that the design of a software environment evolves over time as experience is gained in its utilization by a diverse set of users and on a range of real world applications. In the following subsections, project descriptions and project data, and its usage in the evaluation cases are first outlined. Then three evaluation cases used to assess the three newly developed visualizations are demonstrated.

### **4.5.1 Description of projects and project data**

Two projects are used to test the visualizations developed. The first project (referred to as project 1 hereafter) is a 3 km segment of the original Advanced Light Rapid Transit Project (ALRT) in Vancouver, British Columbia built some 26 years ago. The project was planned to start on October 11th 1983 and finish on June 22nd 1984. However, due to numerous reasons, project completion was delayed until September 24th 1984. Thus, the planned and as-built data pertaining to the products, locations, and the processes were modeled and/or collected from the original contract documents (drawings) and post-project analysis documents. These data were visualized using the time performance variance measures visualization and PCBS attributes visualizations for understanding their usefulness in answering CM analytic reasoning questions

that fall under question groups 1 and 2 and for seeking design refinements to enhance usefulness. This project was selected because it is representative of linear projects that involve many repetitive activities executed through many locations. Usage of the ALRT data is described in demonstration cases 1 and 2.

The second project used is a 6 story residential building project (Upper Crust Manor project, referred to as project 2 hereafter), which was created based on actual experience with a number of high-rise building projects. The planned start is October 20th 2003 and projected finish is August 31st 2004. The project data used includes planned schedule data, two as-built schedules as of progress dates November 28th 2003 and December 31st 2003, planned resource usage data, and as-built data such as daily construction problems, workforce conditions, and records (e.g. meta data for photos, letters). These data are used to evaluate visualizations of time performance variance measures and cause-effect visualization in order to assess their usefulness in answering CM analytic reasoning questions that fall under question groups 1 and 3, and to identify possible design refinements. The as-built data, drawn from actual projects, was nevertheless constructed so that the true answers regarding reasons for time performance could be framed into data sets complete enough to eliminate the data quality factor. We recognize that most real world project data sets are not as complete as the ones used here, which complicates further the observation that cause-effect relationships between project variables are usually fuzzy and complex in practice. However, the purpose in this chapter is to show the value of the foregoing visualizations to assist in identifying cause-effect relationships. The data set for the Upper Crust Manor project was used for demonstration case 3.

#### **4.5.2 Demonstration cases**

Three demonstration cases are described. The first case tests the use of the time performance variance measures visualization for planned and post-project as-built schedule data for project 1. Case 2 demonstrates the visualization of PCBS location attributes data for project 1. The last case deals with time performance cause-effect visualization of planned and as-built data for project 2. For each case, answers are sought to the specific CM questions posed under the three high level question groups for the design cases examined in the previous sections. For each case, inferred implications relevant to the answers sought and the context of project are also explained.

Desired enhancements to the visualizations developed and which were identified when assessing their responsiveness to the CM questions posed for the demonstration cases examined are also described.

*Case 1:* In this case, visualizing variance values for several activity sets that were involved in constructing the foundations and verticals for project 1 is tested. The activities involved can be grouped into different activity sets using different grouping schemes. For example, they can be grouped by trades, by locations, by location sets, and/or by groupings of repetitive activities. A specific analytical reasoning question is formulated as "*What are the values and patterns of behaviour of duration variance and how can they provide useful insights into time performance vs. activity, project participant, and/or location dimensions?*" For answering this question, a 3D version of the schedule variance graphics presenting the duration-variance values of the foundation and verticals activity sets is shown in Figure 4.6 (a) (all trades, activities and locations are shown, leading to a relatively large data space). Salient visual patterns and areas of attention for viewing these patterns are boxed in the figure. These patterns can be translated into values and patterns of behaviour of duration variance. An arrow is used to point to a bar that is a standout anomaly; it corresponds to a high duration variance incident for the form trade activity, "form column" activity, at location "F537". The foregoing observation can be interpreted as a very large duration variance being, a function of the project participant, activity, and location dimensions. Certain attributes of the form trade, of the "form column" activity, and of the "F537" location may have a combinatory effect that contributed to the variance. Nevertheless, it can also be simply a data quality issue (e.g. erroneous data entry or recording), which in fact was not the case for this project.

By comparing the overall visual patterns of the five "boxed" groups of bars, in Figure 4.6(a) the insight gained is that activities of the form trade and the cure/strip trade mostly had *negative* duration-variance values (i.e. actual performance better than planned), activities of the concrete pour trade mostly had *zero* duration-variance values, and excavation trade and backfill/grade trade activities experienced *positive* duration-variance values from time to time. The foregoing visual patterns mean that positive duration variances are clustered in the excavation trade and the

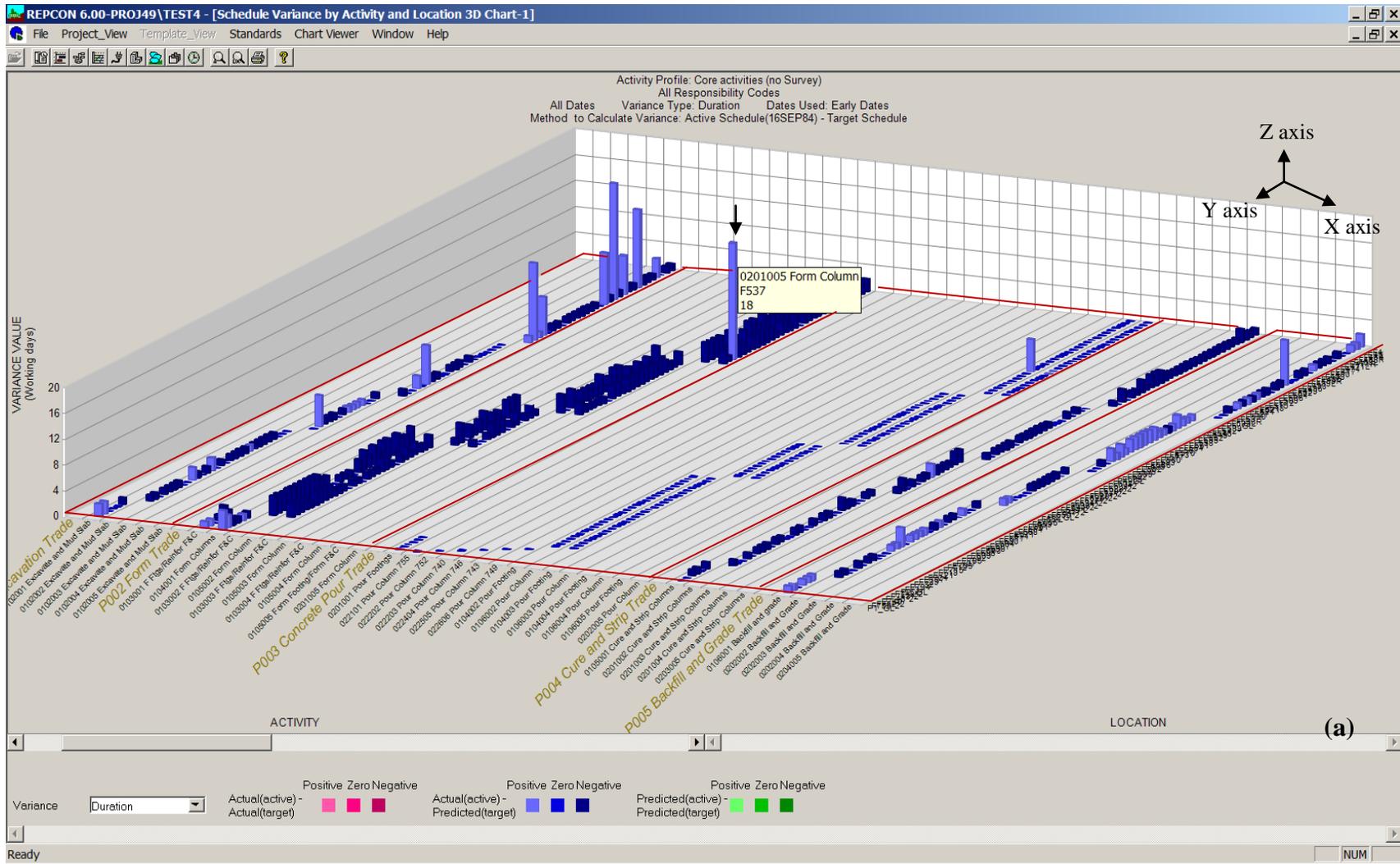
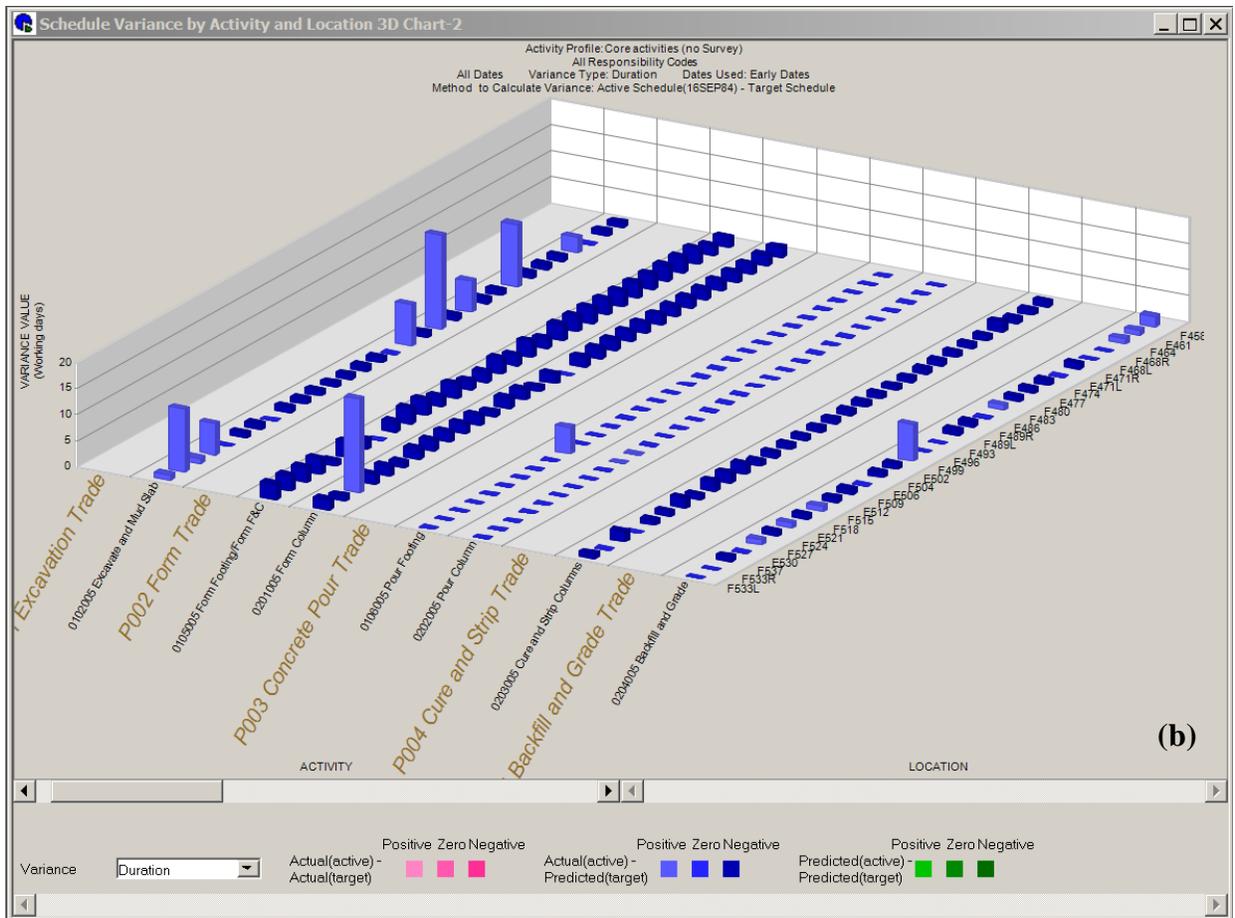


Figure 4.6 (a) A screenshot of a 3D version of schedule variance graphics presenting the duration variance values of the activity sets related to constructing substructures for project 1 for all locations, (b) a screenshot of a 3D version of schedule variance graphics zooming into regions of interest seen in Figure 4.6(a) (i.e. the phase 5 locations, from locations F533L to F458)



backfill/grade trade. Since the two trades correspond to the "excavate and mud slab" and "backfill and grade" activities respectively, one possible interpretation is that the positive duration variances are caused by attributes of one or more of the activity and project participant dimensions. One attribute value common to these activities and their trades is that work type involved is earthworks (e.g. cut, fill, store, borrow, and haul earth), which could be a significant causative factor for the positive duration variances.

An example of a useful enhancement to the visualization requirements or specifications is that the visual representation should emphasize the "grouping" visual effect in the same way as "red boxing" the data points by trade as seen in Figure 4.6(a). A user interface could be provided to facilitate the choices of different visual encoding strategies such as using thicker gridlines or wider spacing to separate different activity sets (e.g. the way the authors manually boxed the image into five zones by trades). Once areas of particular interest are pinpointed by viewing the

entire data space as shown in Figure 4.6(a), then one's attention can be narrowed to regions of interest (e.g. excavation and backfill work within a particular location region) by zooming in or filtering Figure 4.6(a) to create Figure 4.6(b) for viewing details such as labeling for data values.

*Case 2:* The built products of project 1 were constructed in an urban environment. Characterizing the conditions of the construction locations corresponding to pier stations involves a rather rich data set. The location attributes that matter to construction include soil conditions, working area accessibility/availability, maintaining street traffic, encroachment to property lines, underground and overhead utilities, etc. Using planned vs. actual values of the three attributes of "percentage work area received", "overhead (OH) utilities relocation by others", and "underground (UG) utilities relocation by others" as a test case, three analytics questions specific to *"What are the values and patterns of behaviour of construction conditions"* were formulated as: *"What are the values and patterns of behaviour of " percentage work area received ", "overhead (OH) utilities relocation by others", and "underground (UG) utilities relocation by others", and how can they provide useful insights into time performance vs. the location dimension?"* Three 2D images of PCBS attributes representing the distribution by location of the foregoing planned vs. actual attribute values are shown in Figure 4.7. In Figure 4.7(a), the circled visual pattern indicates actual values that were relatively lower than planned values (shortened as high undesired change in values) for the "percentage work area received" attribute and which were clustered in between locations FC 737 and FC 709. This undesired change in values of "percentage work area received " may have impacts on the location dimension and in turn on its associated context dimensions such as the activity dimension, depending on the type of work involved. Thus, activities executed at those locations may have their attributes affected - e.g. actual start date may be different than planned, durations may be different than planned, etc. In Figure 4.7(c), the circled visual pattern points to the fact that undesired overhead utility relocation conditions (overhead utilities were never moved by utility companies as opposed to being relocated on time) were encountered and there was significant clustering in the location range of FC 737 to FC 705. Again, depending on the type of work being performed, activity performance could be impacted significantly in these locations. Lastly, by comparing the two visual patterns in Figures 4.7(a) and 4.7(c), it is observed that the undesired change in values of both attributes "percentage work

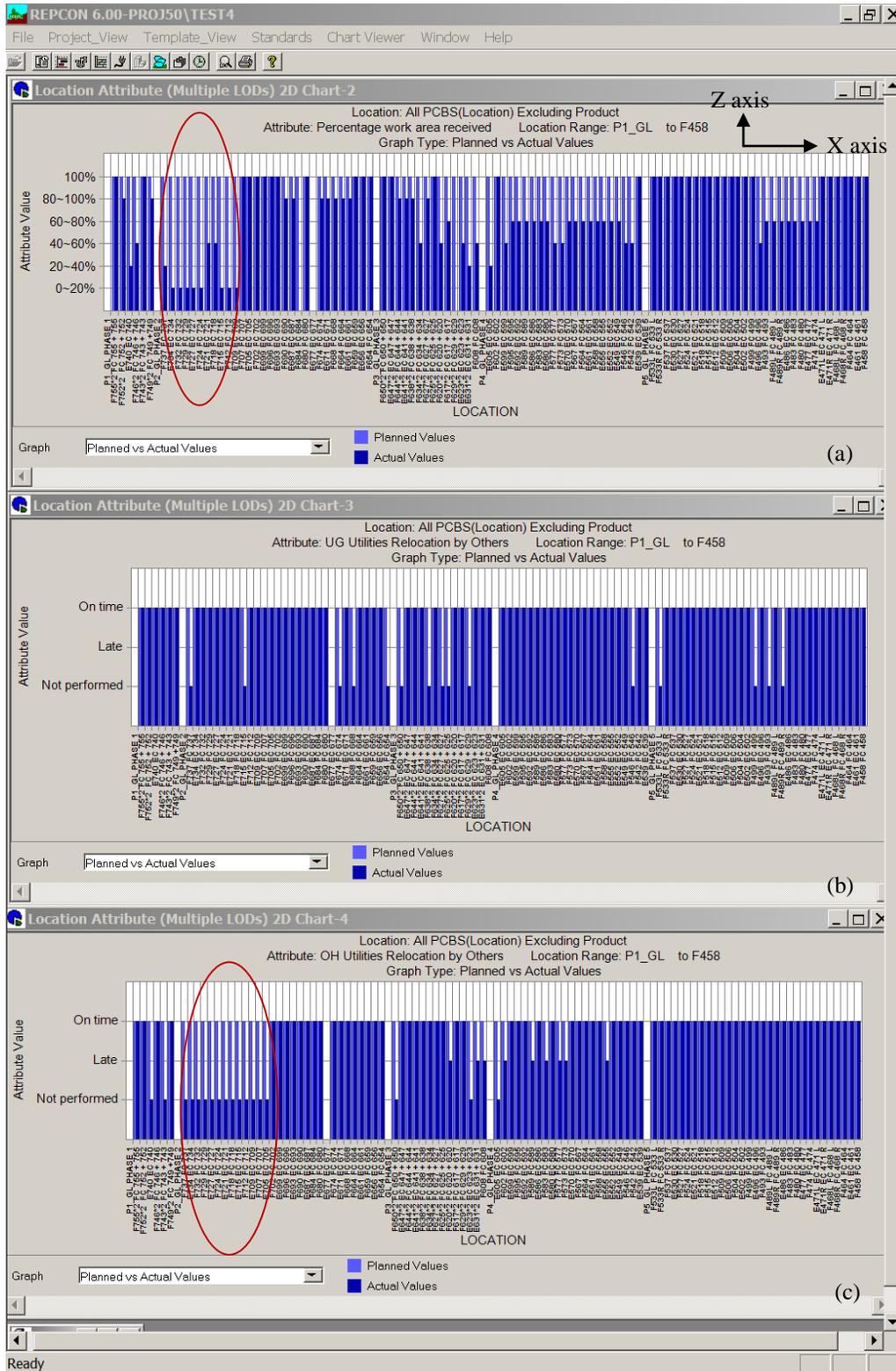


Figure 4.7 Three 2D versions of PCBS attribute graphics presenting the distribution of planned vs. actual values for location attributes: (a) percentage work area received, (b) underground (UG) utilities relocation by others, (c) overhead (OH) utilities relocation by others.

area received" and "OH utility relocation by others" clustered in the same location range may singly or in combination impact negatively the time performance of activities executed at these locations. Therefore, in this location range there could be start delays for the lead activity "excavate and mud slab" at each location due to not being able to receive the work areas as planned, and further, start delays for the activities associated with crosshead construction because the unrelocated overhead utilities may interfere with work in the air (cranes in close proximity to active lines).

*Case 3:* Project 2 and its synthetic project data set are used here for evaluating and demonstrating how time performance cause-effect visualization enhances the ability to reason about potential causes of activity time performance variances. By way of an example, the analytical reasoning process starts with identifying that the duration variances of the excavation and shoring trade's activities impacted other activities at the parkade location. This is inferred by interpreting the similarity/difference in variance patterns (the Y axis treats variance types) between this trade's activities and those of other trades as seen in Figure 4.8 (i.e. the excavation and shoring trade's activities have larger duration variances but small start predecessor/start variances; activities of other trades have the opposite variance pattern indicating that their delays are caused mainly by predecessor activities, which are the excavation and shoring trade's activities). In order to find the possible reasons for the duration variances of the excavation and shoring trade's two activities, two questions can be formulated as:

1. *For the approximate time window October 28th 2003 to December 5th 2003 for execution of the activity "shotcrete shoring at parkade level", are there abnormalities which occur before or at the same time when the finish date variance develops (between November 19th 2003 and December 4th 2003) in the temporal values of the following construction conditions:*
  - *The daily execution status of the activity;*
  - *Temperature and precipitation environmental conditions;*
  - *Ground condition; and*
  - *Problems encountered that are associated with the activity?*
2. *For the approximate time window October 26th 2003 to November 28th 2003 for execution of the activity "bulk excavate substructure at the parkade level", are there abnormalities which*

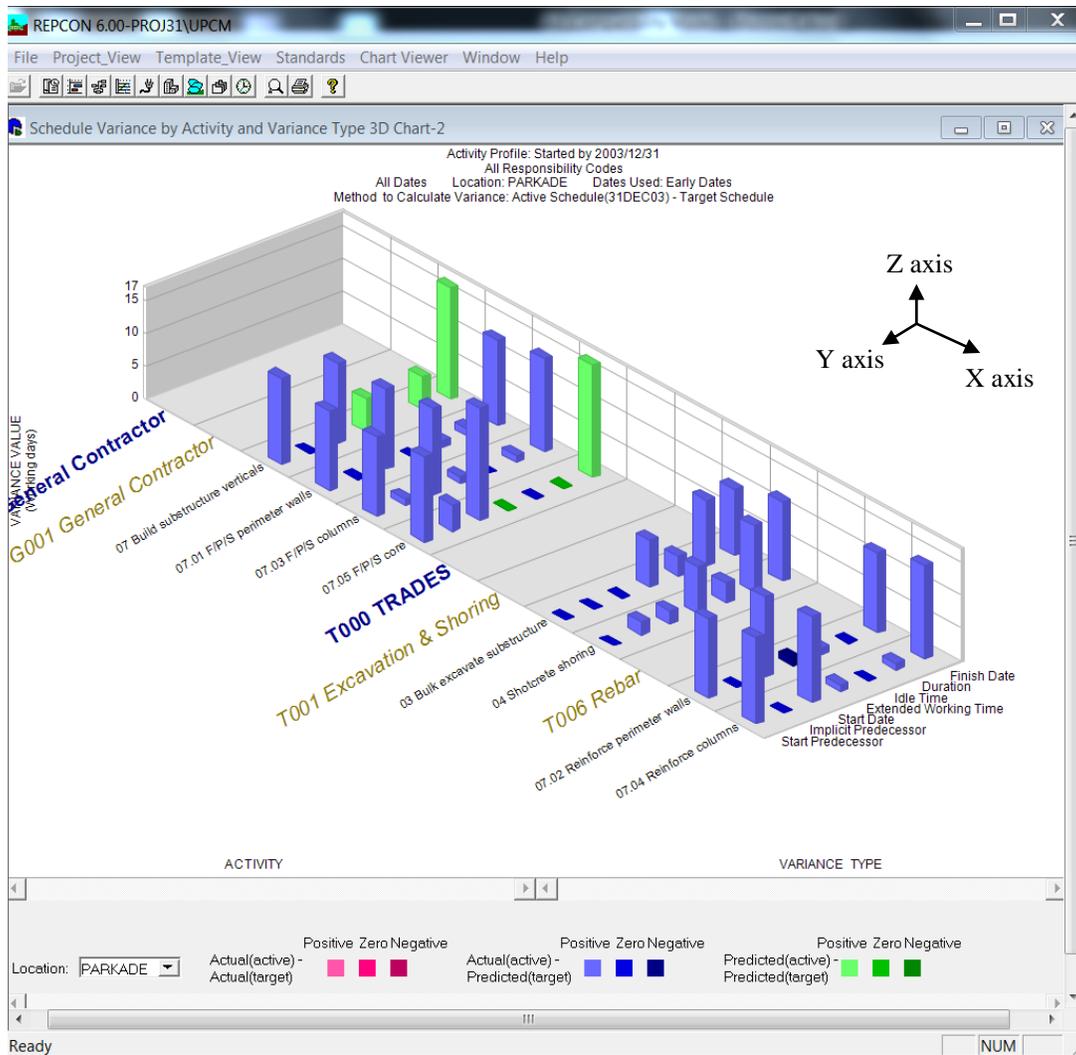


Figure 4.8 A 3D version of schedule variance graphics presenting the activity time variance values of the sets of trade activities related to construction work at the parkade location for project 2

*occur before or at the same time when the finish date deviation develops (between November 14th 2003 and November 28th 2003) in the temporal values of the following construction conditions:*

- *The daily execution status of the activity;*
- *Precipitation as an environmental condition;*
- *Problems encountered that are associated with the activity; and*
- *Resource (excavator, truck) usage?*

Depending on the availability of data and visualizations to represent construction conditions mapped onto the time dimension, users can query images of CM variables that reflect the causal models of CM users as potential reasons for the time performance deviations that treat the

conditions listed previously as well as other potential explanatory conditions.

Having designed and implemented such cause-effect visualization features, two multiple view images representing time performance measures and construction conditions that are mapped against the time dimension are shown in Figures 4.9 and 4.10. By quickly scanning Figure 4.9 from top to bottom, it can be seen that during the time period of executing the "shotcrete shoring" activity at the parkade level, potentially abnormal construction conditions such as continuous rain, poor ground conditions, and soil contamination were encountered. Also by scanning the left side of Figure 4.10 vertically, it is observed that during the time period of executing "bulk excavate substructure at parkade level", potentially abnormal conditions encountered included rain and insufficient equipment which may have impacted time performance. When searching for further evidence regarding insufficient equipment problems, only the planned daily equipment usage images are currently available as shown in the right hand side of Figure 4.10. If the visualization of actual daily equipment usage was also available (assuming that such data was collected which is likely for excavation work), then it would be possible to see whether there were usage reductions from the planned one to the actual one, thereby strengthening the reasoning that a problem of insufficient equipment did occur and may have caused the finish date delay.

During the test of using the cause-effect visualizations developed, it was observed that a key feature of this type of visualization is many individual images are juxtaposed to support the analytics of validating/invalidating a cause-effect hypothesis and they can be coordinated to enhance efficiency. For example, Figures 4.9(b) and 4.10(d), Figures 4.9(c) and 4.10(a), and Figures 4.9(f) and 4.10(b) respectively are of the same image types with different contents by two different activity/location combinations and/or time windows. Currently, this change of content between the three dual images only requires one operation of specifying a different activity/location combination and time window which can be applied to the three paired images automatically. In addition to the currently supported common features and interaction features coordination just described, a future enhancement for analytics would be to encode non-working day information so that the gaps seen in the current images do not mislead viewers to think that they are abnormalities.

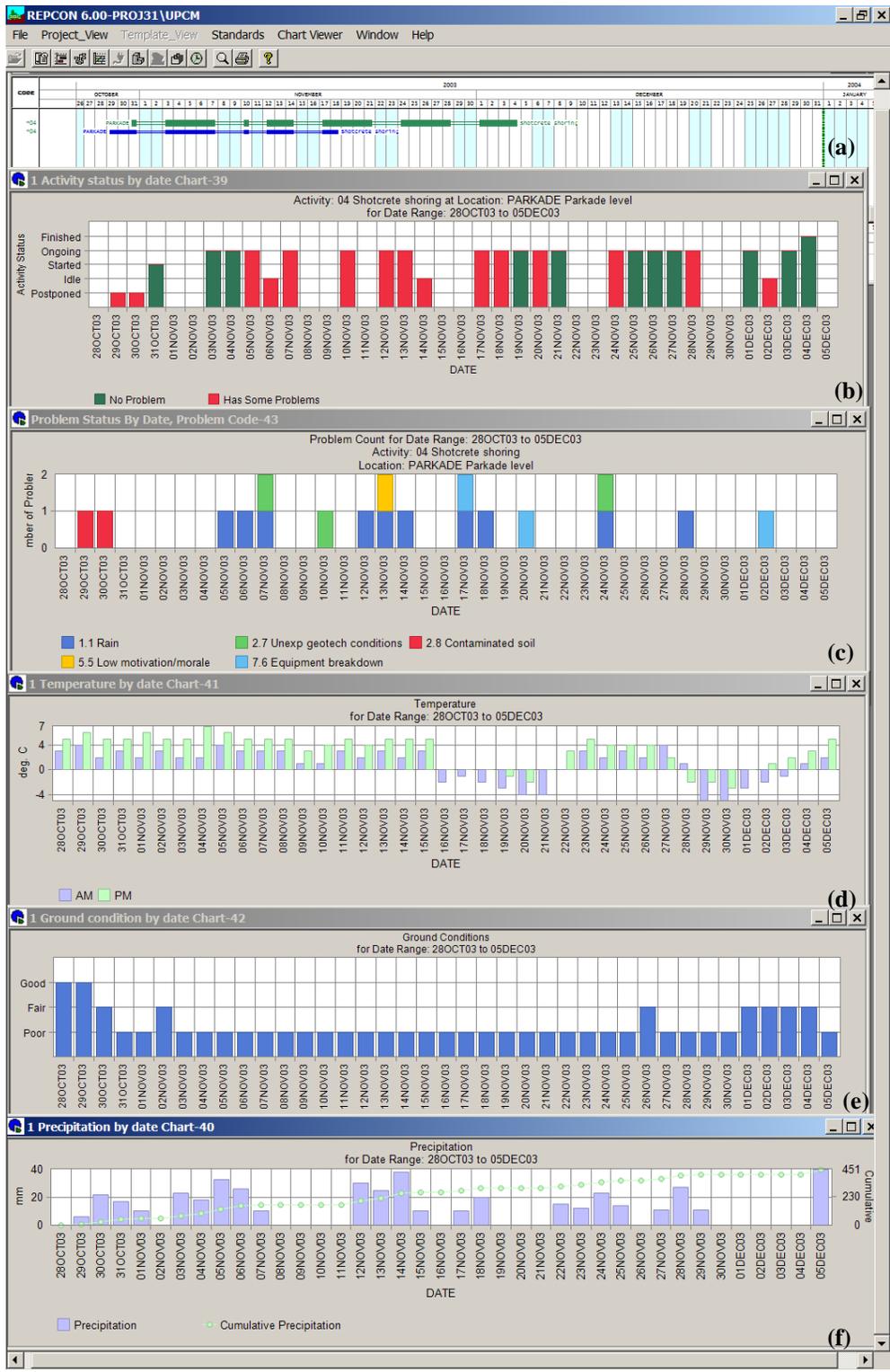


Figure 4.9 A multiple view image representing selected construction conditions associated with the activity "shotcrete shoring at parkade level": (a) comparison schedule (progress date of 31 December 03 vs. planned project early start date of 20 October 03), (b) activity status, (c) problems encountered, (d) temperature, (e) ground conditions, (f) daily and cumulative precipitation.

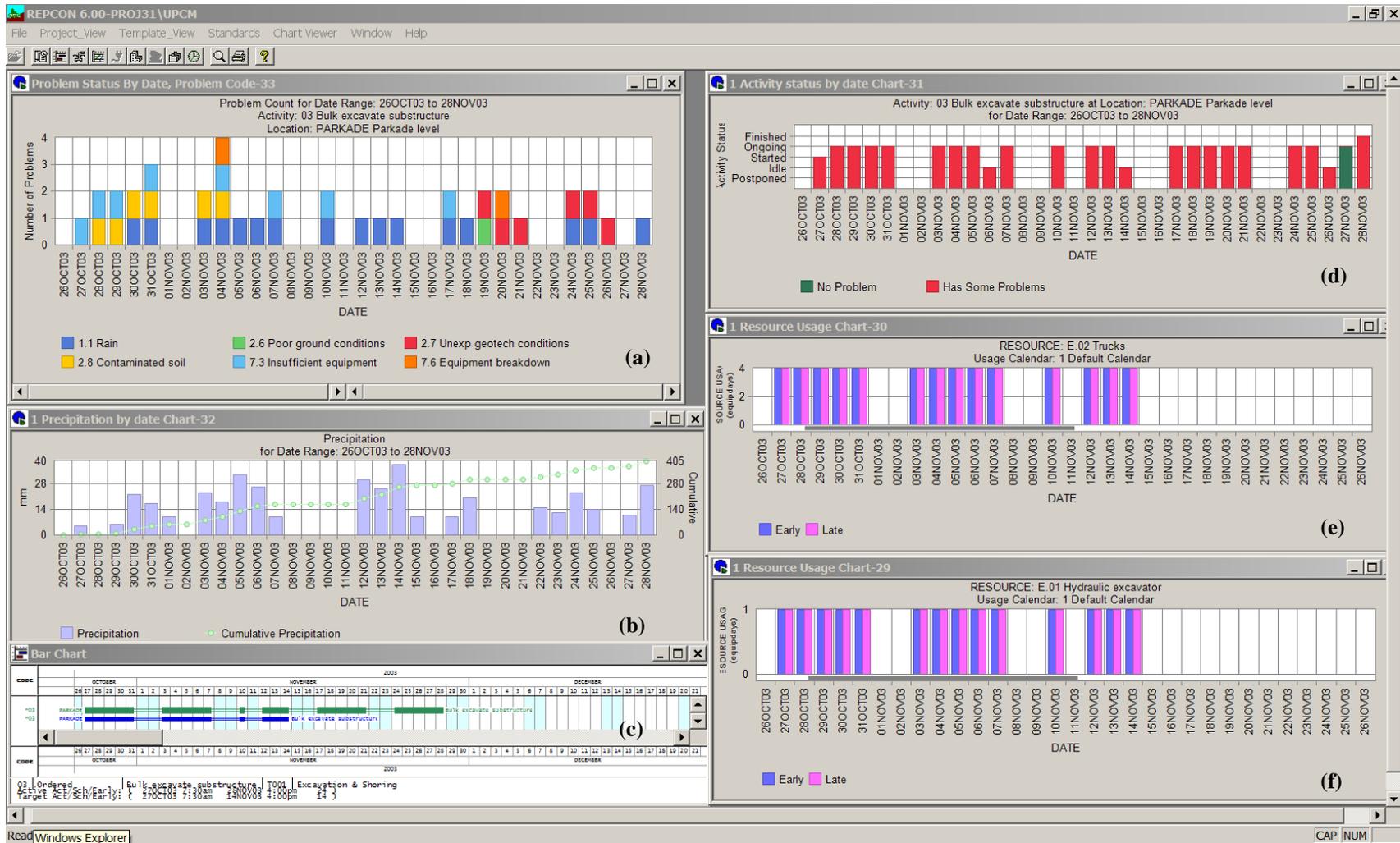


Figure 4.10 A multiple view image representing selected construction conditions associated with the activity "bulk excavate substructure at parkade level": (a) problems encountered, (b) daily and cumulative precipitation, (c) comparison schedule (progress date of 31 December 03 vs. planned project early start date of 20 October 03), (d) activity status, (e) Equipment (truck) planned resource usage, (f) Equipment (hydraulic excavator) planned resource usage (for (e) and (f) the early and late plots are identical because the activity is a critical one).

## 4.6 General observations

The three new visualizations presented in this chapter, developed using a structured bottom-up design process integrated with design guidelines and a top-down design approach, showcase the kind of environment architecture of a CM data visualization tool required for supporting CM analytics common to a range of CM functions/tasks. In this environment architecture, an organization of thematic CM data visualizations, hierarchical from the abstract (e.g. time performance visualization) to the specific (e.g. duration variances distributed by location and activity), can be developed in a consistent way. The common image features shown by these visualizations and the CM analytics supported are as follows:

1. Images created can be of different themes by construction conditions (i.e. attributes characterizing project context dimensions of activity, location, product, etc.) or performance measures (i.e. scope, time, cost, safety, quality, etc.). Images of the same theme can be further categorized into different types based on: a) project context dimensions against which construction conditions or performance measures are mapped, b) showing non-variance or variance values, and/or c) visual encodings. Images of various themes represent visual depictions of important CM variables; images of various types for a certain theme represent values or value changes (i.e. types by showing non-variance or variance values) of a particular CM variable in relation to project context dimensions (i.e. types by project context dimensions against which construction conditions or performance measures are mapped), and how this relation be observed (i.e. types by visual encodings). Figures 4.6(b) and 4.8 are examples of images that are of the same theme (showing time performance measures) but of different types by visual encodings (in Figure 4.6(b) the location and variance type dimensions are mapped to the Y axis and "singular selection" respectively while in Figure 4.8 the mapping is the other way around). The contents of images of any type can be further changed for observing values or value changes of a particular CM variable in relation to certain project context dimensions by different levels of detail and ways of computing the values or value changes. Image contents can be changed by: a) granularity of project context dimensions, b) item selection for project context dimensions, c) data status states, c) data versions (dates), d) how to aggregate measurements, and, e) how to compute variances between the planned and actual values of measurements. For example, in going from Figure 4.6(a) to Figure 4.6(b) the image contents are changed by "items selection for project context

dimensions" which in this case is a change of content from all locations to only phase 5 locations.

2. In addition to the classification of images just described, images can also be categorized by at least two formats: a) distribution of values of construction conditions or performance measures in project context dimensions, and b) distribution of values of construction conditions or performance measures in the occurrence time dimension and condition /performance definition dimension. The first format supports the common analytics of identifying whether reasons for or impacts of the behaviour of CM variables are a function of these project context dimensions. The images generated by the time performance variance visualization (e.g. Figure 4.6(a)) and the PCBS attribute visualization (e.g. Figure 4.3(c), Figure 4.7) described in this chapter correspond to this format. The second format is used for the common CM analytics for checking whether construction conditions have a possible cause-effect relationship with other construction conditions so that root causes of performance might be traced. Multiple view images generated by the time performance cause-effect visualizations such as shown in Figures 4.9 and 4.10 belong to this format.

Although visualizations supported by an environment architecture have common features, each visualization is unique because of the nature of the specific CM analytics involved and because related CM variables demand different settings of common visualization features and in some cases require special features in addition to the common ones. Recognition of this can be observed from differences in the analysis of visualization requirements for the three design cases described in this chapter and is accompanied by the insight that each new visualization under development always requires separate analysis. For example, the time performance variance visualizations do not allow the common aggregation of variance values of individual activities because network model based computation algorithms are required for obtaining semantically meaningful and accurate variance values of grouped activities. In contrast, PCBS attribute visualizations may permit the common aggregation of quantitative attribute values from components at a finer level of granularity to ones at coarser level of granularity because a simple summation operation is applicable. Thus, different choices of visual encodings are needed to address such differences.

The design cases demonstrated in this chapter are meant to support CM analytics related to time performance control. However, visualizations built within this environment architecture can be readily applicable to other CM functions/tasks. A design analysis process similar to the one used for design cases 1 and 2 can be applied for developing visualizations for presenting the distribution of values of CM variables (i.e. performance measures and construction conditions) pertaining to other CM functions such as quality management, safety management, and cost management against relevant context dimensions. This will assist in identifying whether reasons for or impacts by the behaviour of these variables are a function of these context dimensions. For example, a product design feature (e.g. whether walls are straight or curved), of a specific PCBS component attribute, might be a CM variable that affects cost (Staub-French et al. 2003). Thus, values of product design feature can be visually mapped against associated product and pay item dimensions for understanding their potential impacts on certain products and pay items. A design analysis similar to the one used for design case 3 for time performance cause-effect visualization can be applied for developing visualizations presenting the distribution of construction condition values, which may be related to performance measures relevant to other management functions (e.g. quality, safety, cost) versus the occurrence time dimension. Thus, for example, a cost performance cause-effect visualization could be developed.

The methodology demonstrated in this chapter for evaluating an implemented CM data visualization is cost effective (Ardito et al. 2006) for evaluating newly developed visualizations. The evaluation process involves the assessment of the implemented products against prescribed analytic needs, requirements, and specifications. Non-conformances to specifications are seen as deficiencies or bugs; non-conformances to requirements leads to a review of specifications for enhancements; and non-conformances in terms of serving analytics needs require the review of both requirements and/or specifications. In addition to identifying any items of non-conformance, enhancements to visualization features that can improve usability and/or utility may also be observed. These new visualization features along with ones identified from the visualization requirement analysis may be applicable to other visualization designs and hence rolled up to become new design guideline items or check list items of top-down common visualization features. Several new visualization features along with other lessons learned are:

1. **Enhancing the visual effects of groupings:** Construction conditions/performance measures

grouped by different context dimensions (e.g. when, where, and/or who is responsible) help identify potential effects on or causes by those individual dimensions or some combination of them. This insight has been demonstrated by "boxing" activities of the same trade in Figure 4.6 (a). Several visual encoding techniques can be used to emphasize grouping. If the items of context dimensions are encoded by spatial positions, applying different colors and/or sizes to their gridlines, labels, or chart sections can be used for strengthening visual effects of groupings. Using coloring of visual marks is not recommended because it may be better used for encoding data status states (i.e. planned vs. actual). If they are not encoded by spatial positions (e.g. activities in a linear planning chart), connecting lines or the coloring of visual marks become essential to enhance the visual effects of groupings.

2. **Encoding different ways of data value assignments:** Many current data visualization tools and systems only deal with visualizing singular data values. In the construction management context, it is not uncommon that a project model is characterized by dimensions of various data types and data value assignments. In addition to common knowledge regarding the correspondence between data scale types and effective visual variables (e.g. positions, color), a knowledge base in terms of what kinds of data value assignments can be appropriately encoded by certain visual marks (a visual mark (e.g. point/line/area), is different from a visual variable (e.g., spatial position/color/orientation), in that visual variables are attributes of visual marks) can also be established. For example, if the data value assignment is "within" a value range, the visual marks of lines (e.g. Gantt bar) can be used to present value ranges; if the data value assignment is "greater than" or "less than" a value threshold, singular open ended Gantt bars or singular arrow-headed lines can be used to visualize these data range types.
3. **The time dimension as a context dimension or as a time performance measure:** The temporal dimension can be seen as a context dimension or as a time performance measure. In the construction management context, the time dimension as a context dimension usually refers to time stamps of event occurrences or data recording. On the other hand, the time dimension portrayed in a bar graph or linear planning schedule represents time performance measures (start dates, finish dates, and durations as the ranges between start dates and finish dates).
4. **Issues of visualizing multiple conditions/performance measures and/or their distribution in multiple context dimensions:** When the number of data dimensions that need to be

visually encoded grows, the number of visual encoding alternatives grows exponentially. Different ways of encoding the same data dimensions may provide insights in support of different CM analytics tasks. In developing a CM data visualization environment, this issue can be addressed either by: 1) providing users with full control over changing visual mappings, or 2) continuing with a top-down design and breaking a visualization down into several sub visualizations in support of various specific CM analytics. For example, in our first case study of designing visualizations for time performance variance measures, the first approach is to directly conduct the bottom-up development for a single visualization by specifying default visual mappings for all relevant dimensions and provide interaction features for users to explore and change the visual mappings. This approach is of limited practical use because most CM functions/tasks do not permit longitudinal exploratory data analysis. Therefore, a more practical approach is to break a visualization down into sub visualizations that have fixed visual encoding for focused (but limited) analytics tasks that may be more frequently used by end users. One of the exemplary sub visualizations and the specific CM analytics it supports related to time performance variance measures is shown in Figure 4.2. In addition to this practical approach, three fixed visual encoding rules are recommended to further reduce the burden on both designers and users of choosing visual encodings:

- Always map measurement dimensions to the Z axis and map primary project context dimensions to the X, Y, or Z axes because values of these dimensions tend to be many and visual variables other than spatial positions are less effective in differentiating many values;
- Visual variables are scarce resources. Therefore, it is best not to encode data with more than one visual variable. For example, in the visualization of time performance variance measures, the variance values are "double" encoded because in addition to spatial position on the Z axis, color saturations are also used to differentiate positive, zero, and negative values. While the intention is to reduce potential occlusion problems of 3D bars, the benefit of doing so is not easy to justify especially when many important data dimensions needed to be encoded;
- If two context dimensions have many to many relationship (e.g. activity vs. location) and they are to be mapped to spatial positions in the visualization space, they should be

mapped to orthogonal coordinates for equal use of visualization space in two coordinates. An even more general rule for more than two context dimensions that have a many to many relationship is that the space use of three coordinates should be equivalently long (i.e. each spatial coordinate encodes a similar number of instances of context dimensions). This permits a more optimum mapping for purposes of visualization.

5. **Issues of enhancing image readability:** The use of a CM data visualization environment may involve juxtaposing many images for instant scanning. One example of this use is the viewing of multiple images (Figures 4.9 and 4.10) generated by the time performance cause-effect visualization. While interaction features can be provided for adjusting scale, orientation, and position settings of visual elements in images (e.g. visual marks, legends, labels) in order to enhance readability, applying input device operations (e.g. mouse clicks, selecting from combo boxes, etc.) for changing the settings for visual elements and images one by one is simply too inefficient. Therefore, it is recommended to use optimum scales and orientations of visual elements for implementing individual images that are most readable (e.g. use font size of 12 for labeling). These images are then placed in a virtual 3D space and the only user interactions required for enhancing readability are to change the scale, orientation, or position of the individual images themselves in three orthogonal coordinates. Users can also globally change the position, orientation, and scale of the 3D virtual space thereby applying the changes to all images in the space (i.e. interaction coordination).
6. **Method and focus of evaluation:** Construction management personnel are generalists as opposed to being data analysis specialists. Project data analysis is not their primary work and their incentive to use new data analysis tools is limited unless they see direct benefit to their CM functions/tasks. Therefore, it is difficult to obtain immediate feedback from practitioners about the usefulness of newly developed visualizations especially if there is not a direct relationship to their focus at work. Thus, for evaluating newly developed visualizations, two evaluation phases involving different evaluation methods and focus are recommended. The first evaluation phase is immediate after a visualization is implemented. The method and focus in this phase is the same as the inspection type of evaluation described in this chapter. This evaluation is done by the designer/developer and involves a shorter time to perform after implementation with the evaluation being focused on: a) checking conformance to specifications and requirements, b) identifying features required by the CM analytics not

addressed in the requirements/specifications analysis, c) improving usability, mostly for ease of use of interaction features and good visual effects (e.g. visual encoding does help provide the visual effect of data groupings), and e) identifying new lessons learned or new visualization features common to the visualization environment for updating design guidelines and checklists of common visualization features. The main purpose of evaluation in this phase is to ensure a newly developed visualization is potentially useful, usable, and robust for use. The second evaluation phase is during the use of a CM data visualization environment on actual projects for supporting day to day CM functions/tasks. The evaluation is done by CM information system users and needs much more time to perform (i.e. months to years that span life cycles of several construction projects) with the evaluation focus being on: a) identifying which visualizations in the organization of thematic visualizations are used frequently, and b) identifying new visualizations or new visualization features needed.

7. **Choices of visual encodings:** The focus in this thesis of applying a bottom-up development process integrated with design guidelines and a top-down approach is on the creation of new CM data visualizations to assist with CM analytics that cannot be done or which are difficult to do with current modes of data reporting. The case studies described in section 4.5.2 demonstrate promising results from use of these newly created visualizations. However, as already pointed out in item 6 of section 1.4 and section 4.2.2 of the thesis, alternative specifications to the ones used for these visualizations, especially with respect to the visual encodings adopted are possible and in some cases may be preferred. Here we address briefly how different choices of visual encodings may affect usability or utility of visualization, and do this by examining several of the images presented in section 4.5.2. We treat the following four characteristics of an image: 1) whether or not to visually encode implicit information; 2) choice of colors; 3) ordering of labels and visual marks; and, 4) whether use of spatial position on the third coordinate may affect the CM analytics performance of users.

- *Whether or not to visually encode implicit information:* The variable of "time performance variance measures" itself is a variable derived from the time performance metrics illustrated in Figure 4.1 and explained in 4.2.2. There are seven categorical values for the "time performance variance measures" variable as seen by the seven Y-axis categories shown in Figure 4.8. This transformation allows users to see deviations of various time performance measures, from planned to actual, of an activity by the quantity of calendar or

working days rather than simply by a comparison of planned versus actual dates. By showing an activity's variance values corresponding to the seven categories of the "time performance variance measures" variable, CM domain experts can quickly understand the amount of finish date delay of an activity (see the bars that correspond to the Y-axis category of "Finish Date") and whether this delay comes from one or more of:

- Start delay (variance) (see the bars that correspond to the Y-axis category of "Start Date"): this delay can be further attributed to whether it is caused by delays of predecessor activities (see the bars that correspond to the Y-axis category of "Start Predecessor") and/or issues resulting in start delay beyond those associated with explicit predecessors (see the bars that correspond to the Y-axis category of "Implicit Predecessor").
- Duration variance (see the bars that correspond to the Y-axis category of "Duration"): this delay can be further attributed to whether it is caused by idle work (see the bars that correspond to the Y-axis category of "idle time") and/or slower work resulting in extended working time (see the bars that correspond to the Y-axis category of "Extended Working time").

In addition to the aforementioned explicit meanings of each of seven categories of the derived "time performance variance measures" variable, there is an implicit relationship amongst them. For example, start data variance is the sum of implicit predecessor and start predecessor variance, as shown in the three-level hierarchical relationship illustrated in Figure 4.11.

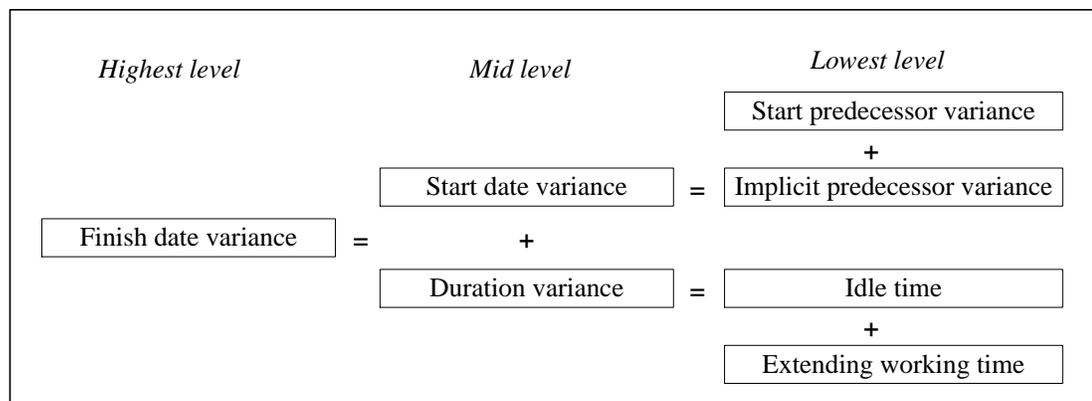


Figure 4.11 An illustration of the hierarchical relationship between various time performance variance measures.

In formulating Figure 4.8, we relied on the intuitive understanding or mental model (formalized in Figure 4.11) of construction personnel of how an activity's finish date is impacted by predecessor performance (both explicit (logic) and implicit (other understood but required conditions)) in terms of establishing actual start time, and work performance including unscheduled idle time in terms of establishing actual elapsed duration. This information in turn provides the basis for computing finish date variance, the time metric of greatest interest for an activity. But without making the visual representation of the underlying mental model more explicit, the messages conveyed in Figure 4.8 may not be transparent and hence the image may not provide the insights intended. Specifically, not all variance types shown in Figure 4.8 are of equal status as a functional relationship exists amongst them. And thus there is a problem. The question becomes – how can the intent of this image be best realized? We use this example as a way of providing some additional insights on how individual images may be enhanced or reformulated, especially in terms of the effective use of established and proven visual encoding principles. We do this by way of a 2 step process. The first step addresses a modification to Figure 4.8 while retaining its 3D form. The second step involves abandoning a 3D image and using a 2D stacked bar chart one which makes the relationships shown in Figure 4.11 much more explicit and resolves a problem of how to treat negative variances.

In our first step, we choose to use three different bar encodings to imply a relationship between the components of a higher level variance and its constituents. Specifically, as shown in the mock-up in Figure 4.12, we used a visual encoding of cone bars to demonstrate: (i) the explicit and implicit predecessor variances that combine to determine the start date variance (encoded as a square bar) of an activity (note that all three can be positive or negative; and, (ii) the working time (positive or negative) and unscheduled idle time (can only be positive) variances to determine the duration variance of an activity (encoded as a square bar). The start date variance and duration variance then combine to determine the finish date variance, which can be positive, zero or negative, and which is shown as a cylindrical bar. Missing from the figure is a key that differentiates between the meanings of cone, square and cylindrical bars. Such a key could be readily added and would provide assistance in interpreting the relationships amongst the seven variance components shown. However, adding more visual marks provides greater complexity, and

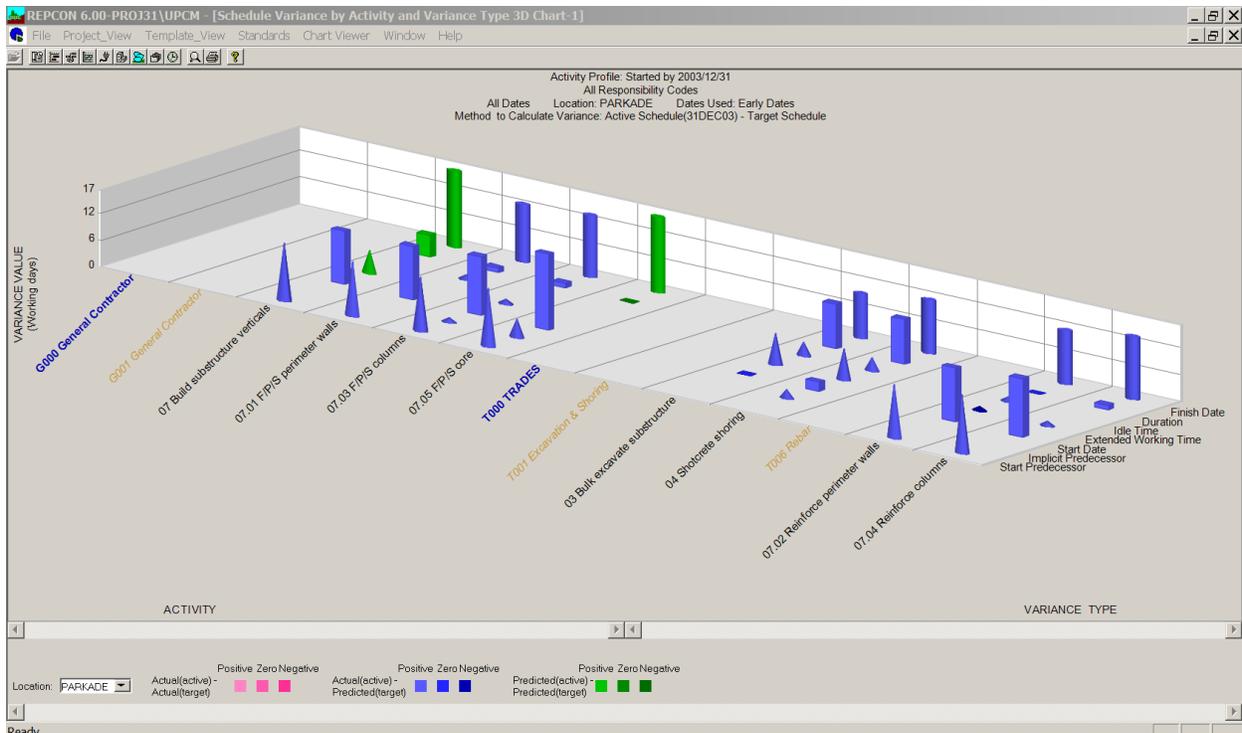


Figure 4.12 In contrast to Figure 4.8, different bar shapes are used to represent different levels in the hierarchical relationship between various time performance variance measures.

the challenges associated with 3D representations remain. Hence, an alternative image concept was sought.

As our second step, we applied a visual encoding strategy that makes use of stacked bar charts to encode the hierarchical relationship about summing together quantities. We believe that this helps users to understand better the semantics of what the variance type labels on the Y-axis in both Figures 4.8 and 4.12 stand for. As shown in Figure 4.13, it is easier to recognize that the height of the triangle visual marks is the sum of the heights of two stacked bars in each bar chart. Also in each bar chart, colors of the two stacked bars are in the same color family, which make it easier to tell that each bar chart represents a certain "variance measures group" that has its own "variance measures summing" relationship amongst group members (Figure 4.13 is a mock-up image only for the purpose of illustrating the aforementioned visual encoding idea; it does not visually encode the information about data status as done in Figure 4.8). One other benefit of this format is that negative variances are easily accommodated without having to use colour coding to differentiate between negative, zero and positive values, as was required in the 3D representation.

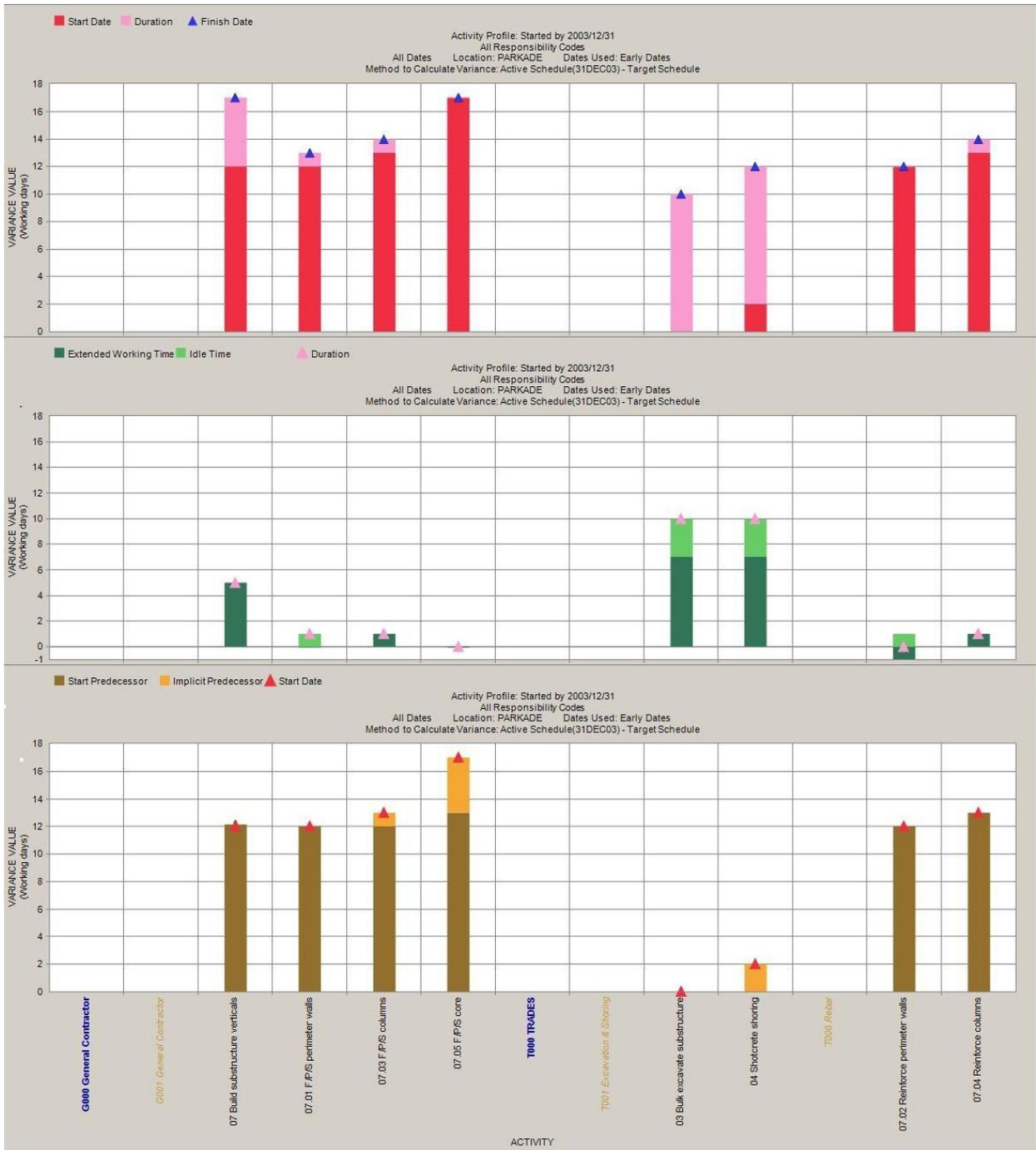


Figure 4.13 An improvement on Figure 4.8 by using a mock-up of stacked-bar charts to represent the hierarchical relationship about summing together quantities of various time performance variance measures.

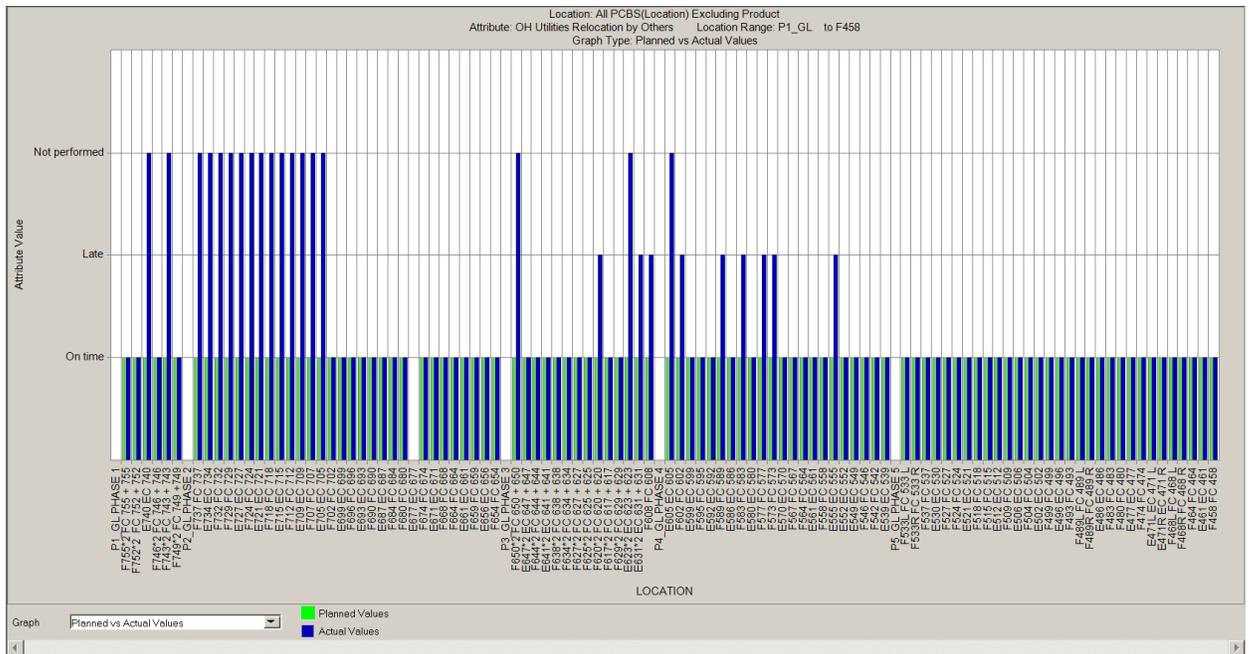


Figure 4.14 An improvement on Figure 4.7(c) by using two more distinctive colors for representing planned and actual data status and by reversing the ordering of labeling on the Z axis.

- Choice of color:* When using colors to represent linguistic values (i.e. categorical data values), the colors used need to be as distinctive from each other as possible. For example, in Figure 4.6 the idea of using three distinctive color hues for representing three data status is a good one. However, the differences amongst the three color saturations actually used within each respective color hue for representing negative, zero, or positive values are somewhat subtle, which makes differentiating bars representing negative values from bars representing positive values challenging. The same issue exists in Figure 4.7 in which two colors that are in proximity of the color saturation spectrum are used to differentiate planned from actual data status. In Figure 4.14, two colors that are more distinctive from each other have been used, which provides a better distinction between bars representing planned values and bars representing actual values in comparison to Figure 4.7(c).
- Ordering of labels and visual marks:* Careful ordering of labels and visual marks can help important information stand out in images. For example, in Figure 4.7 (c), the information that is of greatest concern to CM personnel is whether overhead utility relocations were not performed because if not, they will result in the contractor having to work around these utility issues, thereby likely causing delay and extra cost to the contractor. By reversing the ordering of the Z-axis labels as shown in Figure 4.14, bars representing overhead utility

relocations that were not performed become the tallest and make this important information stand out.

- *The use of spatial position on the third coordinate:* Although using spatial position on the third coordinate (i.e. Y axis) to represent a data dimension seems to make images more compact, it creates several issues such as occlusion and perspective distortion which in turn lead to difficulties in accurately reading off individual data values, as observed in Figure 4.8. An alternative to the use of spatial positions on all three coordinates for representing three data dimensions is that spatial positions on one coordinate can be repeatedly used to create a stacked 2D graphic as per the mock-up image shown in Figure 4.15. As seen in this figure, spatial positions (or position ranges) on the Z axis are used to represent: 1) various time performance variance measures (see the labeling of various variance types on the right hand side of Figure 4.15), and 2) values of each variance measure (see the labeling of numbers on the left hand side of Figure 4.15), thus eliminating the previously identified visualization problems. Another strategy of making use of stacked 2D graphics has been shown in Figure 4.13 in which color, shape, and position range on the Z axis are collectively used to encode various time performance variance measures.

In summary, considerable refinement through improved or alternative visual encodings for the visualizations presented in this thesis is possible. However, as stated previously, our primary focus has been on identifying and responding to the analytical reasoning needs associated with various CM functions in order to demonstrate how data visualization can enhance insights into conditions encountered and performance achieved and contribute to improved decision making by CM personnel.

## **4.7 Conclusions**

From the perspective of developing a CM data visualization environment, design guidelines contribute to setting out the highest level principles in terms of how to apply state-of-the-art data visualization (or visual analytics) techniques to the CM domain. The "top-down" approach contributes to identifying the CM analytics common to various CM functions/tasks and therefore scoping the kinds of visualizations to be developed and the general visualization features required to support them. The "bottom-up" approach involves, in combination with the guidance

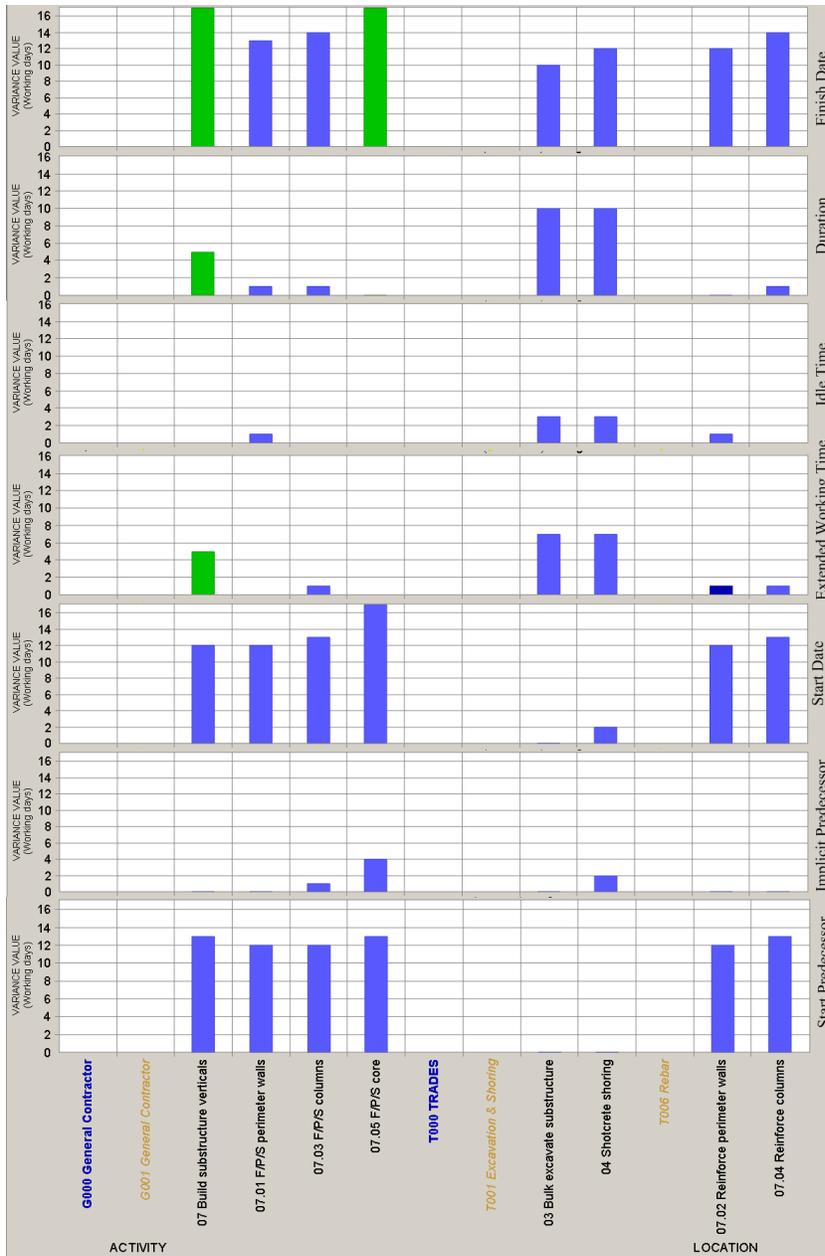


Figure 4.15 A mock-up 2D version image for the 3D graphics shown in Figure 4.8.

of design guidelines and check lists of top-down common visualization features, working out and implementing details in creating visualizations in support of specific CM analytics, and evaluating their use. Lessons learned and additional visualization features identified that are potentially common to other "sibling" or even "ancestor" visualizations can be rolled up to become added design guidelines or check list items of common visualization features for future development of visualizations.

In this chapter, a bottom-up development process integrated with design guidelines and a top-down development approach were applied to develop three new CM data visualizations. Their focus is on helping answer CM questions encountered in the time management/time performance control application. All three designs started with analyzing specific CM analytics useful for this application. The CM analytics treated are: 1) understanding characteristics of variance values of time performance measures for identifying reasons for time performance deviations potentially as a function of project context dimensions, 2) understanding the characteristics of values of product/location attributes for identifying potential impacts on project context dimensions, 3) understanding characteristics of activity time performance measures for identifying deviations and their timing, understanding characteristics of construction conditions which are relevant to the activity for identifying abnormalities and their timing, and understanding whether the abnormalities occur before or at the same time as the time performance deviations.

Steps involved in applying a bottom-up approach include the following. First, the relevant CM variables and their nature in relation to the specific CM analytics involved are identified. Specifics of the top-down common visualization features that fit these CM analytics and/or CM variables along with the need for additional visualization features are then determined. The common visualization features considered are to ensure that the visualizations developed support the general concept of CM analytics, i.e., identifying potential causes and/or effects from the characteristics of the CM variables that matter to the time management/time performance control. The common visualization features considered include: 1) representation of different status states of selected CM variables, 2) the mapping of selected CM variables against context dimensions, and 3) data granularity, aggregation, and selection of CM variables. Default settings of the visualization features can be applied when users generate the first image. Some of the settings are configured as fixed and no user interface is provided for adjusting those settings (e.g. a CM variable is mapped to certain context dimensions; activity and location dimensions must be mapped to orthogonal coordinates in the visualization space). For the next step in the bottom-up approach, the foregoing analysis of specific visualization requirements is then turned into a specification that details choices of data representations/transformations and visualization representation/interaction features for implementation. The inspection kind of evaluation method is then applied to appraise the use of the implemented visualization on actual project data for

identifying any non-conformance with the specifications/requirements and new visualization features that can enhance usability and utility. Finally, lessons learned and new data visualization features identified that are applicable to other visualization designs can be added to existing design guidelines or checklists of top-down common visualization features.

For future research work related to the topic of developing a comprehensive CM data visualization environment, more visualization design cases will be conducted using the structured development process described herein in order to: 1) construct a more complete organization of thematic visualizations, and 2) enrich and refine the design guidelines and checklists of common visualization features for developing other visualizations, both in support of analytical reasoning for a broader range of functions and tasks. A design case that has high priority for future development are visualizations dealing with the function of risk management because of the involvement of uncertainties in the values of CM variables. Another future research topic deals with a CM data visualization "development environment". Such a development environment would incorporate the knowledge of design guidelines, workflow of the top-down and the bottom-up design approaches, and checklists of common visualization features with respect to user interfaces to specify requirements and specifications of new or enhanced visualizations. Included would be mechanisms to automatically generate visualizations in compliance with these specifications. With this development environment, new visualizations could be quickly prototyped and appraised to assess their potential usefulness and hence whether or not to include them in the organization of thematic visualizations for use in practice.

## **Chapter 5 Conclusion-Summary, Answering the Research Questions, Contributions, Future Work**

### **5.1 Overview of conclusions**

Presented in this chapter is :

1. A brief summary of the research background, research development, and research findings (section 5.2);
2. Case studies that demonstrate and analyze how key concepts behind a CM data visualization environment may enhance CM analytics (section 5.3). The demonstration and analysis answer the third research question and provide an indicator of degree of validity and generality of the research findings in answering the research questions set forth in Chapter 1;
3. Summary of research contributions (section 5.4); and,
4. Suggestions for future work (section 5.5).

### **5.2 Research summaries**

To date, the research and development effort as reported in the literature for presenting input/output data in support of human judgment for conducting CM functions and associated tasks has been relatively limited. In practice, CM practitioners often find it difficult to digest and leverage input/output information because of the sheer volume and high dimensionality of data. One way to address this need is to improve the data reporting capability of a CM\_IS system, which traditionally focuses mainly on using tabular/textual reports. Data visualization is a promising technology to enhance current reporting by creating a CM data visualization environment integrated with a CM\_IS. Use of such a visualization environment enhances the CM analytics capabilities of CM personnel such that they can interpret, learn, and communicate causes and/or effects amongst a wide range of CM variables of a construction project thereby improving the quality of CM processes and decision-making. Three research questions about the application of data visualization to CM use were posed to guide the research: 1) *How should a CM data visualization environment be developed;* 2) *What are the key features of a CM data visualization environment that best reflect the functions expected for it;* and, 3) *How does the use of a CM data visualization environment help conduct CM analytics that cannot be done or which are difficult to do with current data reporting practices?*

In search for the answers to the three research questions, an extensive literature review of the use of data visualization in construction management, which included the examination of the data visualization capabilities of current commercial CM information systems and an overview of state-of-the-art data visualization technologies, was carried out. This led to an in-depth understanding of the current state of applying data visualization technologies to facilitate CM visual analytics in support of CM functions. A distillation of the findings from the literature reviewed helped to identify how state-of-the-art data visualization can be adopted/adapted to developing a CM data visualization environment as well as how best to address current reporting shortcomings. These shortcomings include: 1) data for limited CM tasks/functions is visualized; and 2) no domain wide CM analytics-oriented visualization is supported. In terms of how state-of-the-art data visualization can be adopted/adapted, observations made relate to: 1) limitations of state-of-the-art data visualization and focus of development; 2) the need to use domain wide analytics-oriented data visualization development methodologies; 3) understanding the underlying fundamental functionalities of novel data visualization techniques/tools; and, 4) the desired degree of generality/flexibility of a data visualization tool.

By incorporating findings from the literature review and considering the generalist nature of CM practitioners as opposed to a specialist one, it was determined that any visual representations developed need to be ready-to-use, i.e., the target audience is results oriented. The first research phase focused on identifying design guidelines for the design and development of a CM data visualization environment. Such an environment should enhance CM analytics capabilities and therefore: 1) improve the CM process through enhanced understanding of a project's status and reasons for it; 2) improve communication amongst project participants; 3) assist with detection of potential causal relationships; 4) improve decision making; and 5) serve the needs of a broad range of CM functions. The design guidelines identified include the design processes and design components that are applicable to conceptualizing an overall CM data visualization environment and designing details of individual visualizations. **Detailed descriptions of these guidelines were discussed in Chapter 2.** A preliminary test of the concept of visual analytics and the design guidelines formulated were applied to a case study of designing and using visual representations of change order data. Lessons learned through this case study revealed several key points for developing a full scale CM data visualization tool, and included the recognition of

three potential "standard formats" (described in section 2.6.3) of visual representations incorporating interactive features for enhancing image readability, querying data, choosing the format of visual representations, and coordinating multiple visual representations.

With design guidelines and lessons learned from the first research phase in hand, the research effort was then directed to the design of a CM data visualization environment with the focus being on identifying its scope and generality. A top-down design approach was adopted for analyzing the relationship between CM functions, CM analytics associated with those functions, visual CM analytics, and finally the role of a CM data visualization environment in support of visual CM analytics. The identification of this relationship led to the formulation of high level visualization requirements for the environment. The same top-down design analysis was then applied to the narrower CM function of time management. This allowed visualization requirements, which augmented the foregoing high level requirements, for designing visualizations in support of time management functions/tasks to be identified. These visualization requirements provided: 1) scope and direction; and, 2) the common visualization features that need to be addressed for developing new visualizations. **Details about this analysis process are described in Chapter 3.** Through conducting a case study, a research prototype data visualization environment developed as part of the research effort which reflects several of these requirements was used to demonstrate the breadth of support that can be offered for reasoning in support of time performance planning and time performance monitoring/controlling. The prototype environment was also checked against the high level requirements established in this phase in order to identify any deficiencies. Lessons learned through this case study assisted in identifying several key points for designing individual visualizations in support of the time management function, as follows:

1. A CM data visualization environment should have an organization of thematic visualizations, mainly hierarchical from abstract to specific, categorized by construction conditions and performance measures under multiple data views of a project. These multiple data views include process, physical, as-built, quality, change, environment, organizational/contractual, cost, and risk views. The data dimensions of each view define it with components at different levels of granularity along with a characterization of each component in terms of attributes.
2. When designing/developing an individual thematic visualization in support of particular CM

analytics, the following fundamental visualization requirements need to be addressed: 1) scope and definition of CM variables in terms of construction conditions or performance measures that are the central theme of the visualization; and, 2) common visualization features with regard to status states of CM variables, the project context dimensions the variables are to be mapped against, and levels of detail of data required.

3. A CM data visualization environment should provide interaction features for users to interact with visual representations of data. These interaction features can be categorized into two groups with the first being image-specific that deals with images and data content selection, and the second treating general interactions useful for quickly and clearly browsing scenes of visual representations. For example, interacting with an image in terms of adjusting levels of detail of *data* in the image is an image-specific interaction feature which is a function of the nature of the data dimensions the image portrays. On the other hand, the ability to zoom in and out on an image is an image-independent interaction feature which should be the same for all visualizations.

Based on the design guidelines, scope and direction of visualization development, and common visualization features of a CM data visualization environment, the research focus became one of investigating how to design and develop visualizations in support of the CM analytics core to time performance monitoring and controlling. Three new visualizations (time performance measure variance visualization, PCBS attributes visualization, time performance cause-effect visualization) were designed and developed by a bottom-up design process integrated with design guidelines and a top-down design process. All three designs started with analyzing the specific CM analytics associated with the time management/time performance control application. These CM analytics involve : 1) understanding characteristics of variance values of time performance measures for identifying potential causes of time performance as a function of project context dimensions; 2) understanding characteristics of PCBS attributes for identifying potential impacts on important project context dimensions; and, 3) understanding characteristics of time performance measures of an activity and construction performance/conditions associated with the activity to observe abnormalities, if any, and their timing. CM variables involved in the specific CM analytics of interest were first identified. Details of the top-down common visualization features that fit the nature of the specific CM analytics and/or CM variables were

analyzed. Features in addition to those shared with other CM analytics were also identified. The common visualization features that require detailed analysis include: 1) representing different status states of selected CM variables; 2) mapping selected CM variables against project context dimensions; and, 3) data granularity and aggregation, and their selection for CM variables. Default settings of these common visualization features can be applied when users generate the first image. Some of the settings are configured as fixed and no user interface is provided for adjusting them (e.g. a CM variable is mapped to certain context dimensions; activity and location dimensions must be mapped to orthogonal coordinates in the visualization space).

The foregoing analysis of specific visualization requirements was turned into specifications in terms of choices of data representations/transformations and visualization representation/interaction features for implementation. The inspection kind of evaluation method was applied to appraise the use of the implemented visualizations on actual project data in order to determine any non-conformance with specifications/requirements as well as new visualization features that could enhance usability and utility. Lessons learned and new data visualization features identified through the bottom-up design processes applicable to other visualization designs were added to the design guidelines and checklists of top-down common visualization features. **Details about these design/development processes, and applying the new visualizations developed to actual project data for time performance monitoring and control, along with lessons learned are discussed in Chapter 4.**

In addition to the specific visualization designs described in Chapter 4, the bottom-up development process integrated with design guidelines and a top-down design approach were extended to the development of new images that visualize count distribution of as-built records where counts of records are treated as actual construction conditions (e.g. count of change orders is viewed as the construction condition of encountering unexpected changes). While the design processes and lessons learned for developing these images are not elaborated upon in this thesis, the implemented work, along with other thematic visualizations, are used to demonstrate and analyze the merit of using a CM data visualization environment later in this chapter. Lessons learned have contributed to the design guidelines/checklists of top-down common visualization features.

### 5.3 Demonstrating and analyzing the merit of using a CM data visualization environment

Described in this section are the findings from phase 4 of the research. The goal of this phase was to systematically demonstrate and analyze how the key concepts of a CM data visualization environment can enhance CM analytics. The demonstration and analysis was done by comparing a research prototype CM data visualization environment integrated within a research CM information system called REPCON (Russell 1985; Russell and Udaipurwala 2004) against current data reporting practices that mainly use textual/tabular reports and a few graphics (e.g. bar graph schedule). The *demonstration* involves comparing the use of the visualization environment and current data reporting methods to answer questions relevant to various CM functions/tasks for 3 actual projects. Data from these projects was entered into REPCON and both visual representations and tabular reports of CM data are generated from the same REPCON system. The subjects of comparison include the kinds of data reports generated, interaction features utilized, messages useful for answering CM questions received from viewing the reports/images, and time spent on viewing the reports/images. The *analysis* involves analyzing how two key concepts of a CM data visualization environment may lead to differences between using it and using current data reporting practices. The two concepts are: 1) the ability to carry out CM analytics useful for dealing with the CM tasks at hand can be enhanced using data reports in visual forms that are responsive to the intended analytics tasks; and 2) CM analytics capabilities can be enhanced using interaction features and an environment architecture that allows users to flexibly explore data collected/computed for a range of CM functions/tasks presented in visual form.

For each of the 3 project data sets used there are two corresponding demonstration cases for comparing the use of a CM data visualization environment versus current data reporting practices. The first case for each project focuses on comparing how visual CM analytic capabilities in support of a certain CM function/task are performed when interpreting CM data collected specifically for that CM function/task. For demonstration case 1 of project 1, as-planned and as-built schedule data (i.e. time management related CM data) are used for comparing how CM analytics in support of time management/time performance control are performed. For demonstration case 1 of project 2, deficiency data (i.e. quality management related CM data) are used for comparing how CM analytics in support of quality

management/quality performance control are performed. For demonstration case 1 of project 3, change order data (i.e. change management related CM data) are used for comparing how CM analytics in support of change management/change order control are performed.

The second demonstration case for each project focuses on comparing how visual CM analytics capabilities in support of a certain CM function/task are performed when interpreting CM data generated/collected for other CM functions/tasks. For demonstration case 2 of project 1, product quantity data (i.e. scope management related CM data) along with planned schedule data are used for comparing how visual CM analytics in support of time management with emphasis on assessing the quality of duration estimates are performed. For demonstration case 2 of project 2, deficiency data (i.e. quality management related CM data) are used for comparing how visual CM analytics in support of time performance control are performed. For demonstration case 2 of project 3, change order data (i.e. change management related CM data) are used for comparing how visual CM analytic in support of time performance control are performed.

Brief descriptions of the projects and project data used are first introduced and then succeeded by the demonstration cases. For each demonstration case, the following information is included in aid of the assessment process:

1. CM analytics tasks ( i.e. CM questions dealing with CM functions/tasks);
2. Supporting CM functions/tasks;
3. Project data shown in images for answering CM questions;
  - Type of data (e.g. planned/actual schedule data; as-built records);
  - Approximate scope of data (size and/or contents of the data); and
  - Original CM use of the data collected and/or data analysis (e.g. scheduling, cost estimating, document control)
4. Current data reporting practices for the data;
5. Visualization approach; and
6. Descriptions of the exploration-answer process using both a CM data visualization environment and current traditional tabular data reporting functionalities:
  - Data reports (tabular reports and images respectively) generated;
  - Types of primary interactive operations involved reflecting exploration processes for

finding answers to CM questions;

- Messages/answers obtained from the data reports (tabular reports and images respectively); and
- Time spent on viewing data reports (tabular reports and images respectively) before the messages or insights and answers can be obtained (abbreviated as "viewing time" hereafter). Due to the approximations involved in measuring viewing time, time was measured in intervals of 5 seconds. It is observed that the author's viewing time may be shorter than readers and other users because of his familiarity with the formats of both the tabular reports and images.

After completing all six demonstration cases, the findings extracted from these demonstration cases are collectively analyzed to identify insights as to how key concepts of a CM data visualization environment may lead to the differences between using it and current data reporting features.

### **5.3.1 Demonstration cases- project 1**

Project 1 corresponds to a 3 km segment of the original Advanced Light Rapid Transit Project (ALRT) in Vancouver, British Columbia built some 26 years ago. The scope of work consisted of building 103 foundations and piers in support of a pre-cast beam elevated guideway, with installation of the beams being performed by others. The PCBS view of this project consists of a simplified list of project components and a listing of all work locations in planned location sequence, as partially shown in Figures 5.1 and 5.2 (a photo of actual columns is shown in Figure 5.4). The process view of the project, as seen in Figure 5.3, is defined by the main activities of: (i) survey & layout; (ii) excavate & mud slab; (iii) form & reinforce footings; (iv) pour footings; (v) form columns; (vi) pour columns; (vii) cure & strip columns; (viii) backfill & grade; and, (ix) cleanup. Other activities not present at all locations include piling/rock anchor at locations of weak soil conditions and construction of column crossheads (form, reinforce, pour, cure & strip, and post tension). The organizational view of this project, i.e., project participants of the project, is shown in Figure 5.5.

The project was planned to start on October 11th 1983 and finish on June 22nd 1984. However,

due to numerous reasons, project completion was delayed until September 24th 1984. Thus, the planned and as-built data pertaining to the product view and the process view of this project were assembled and examined for the purpose of exploring reasons for time performance that may arise from the aspects of quality of the original planning and the conditions encountered during execution in an urban environment.

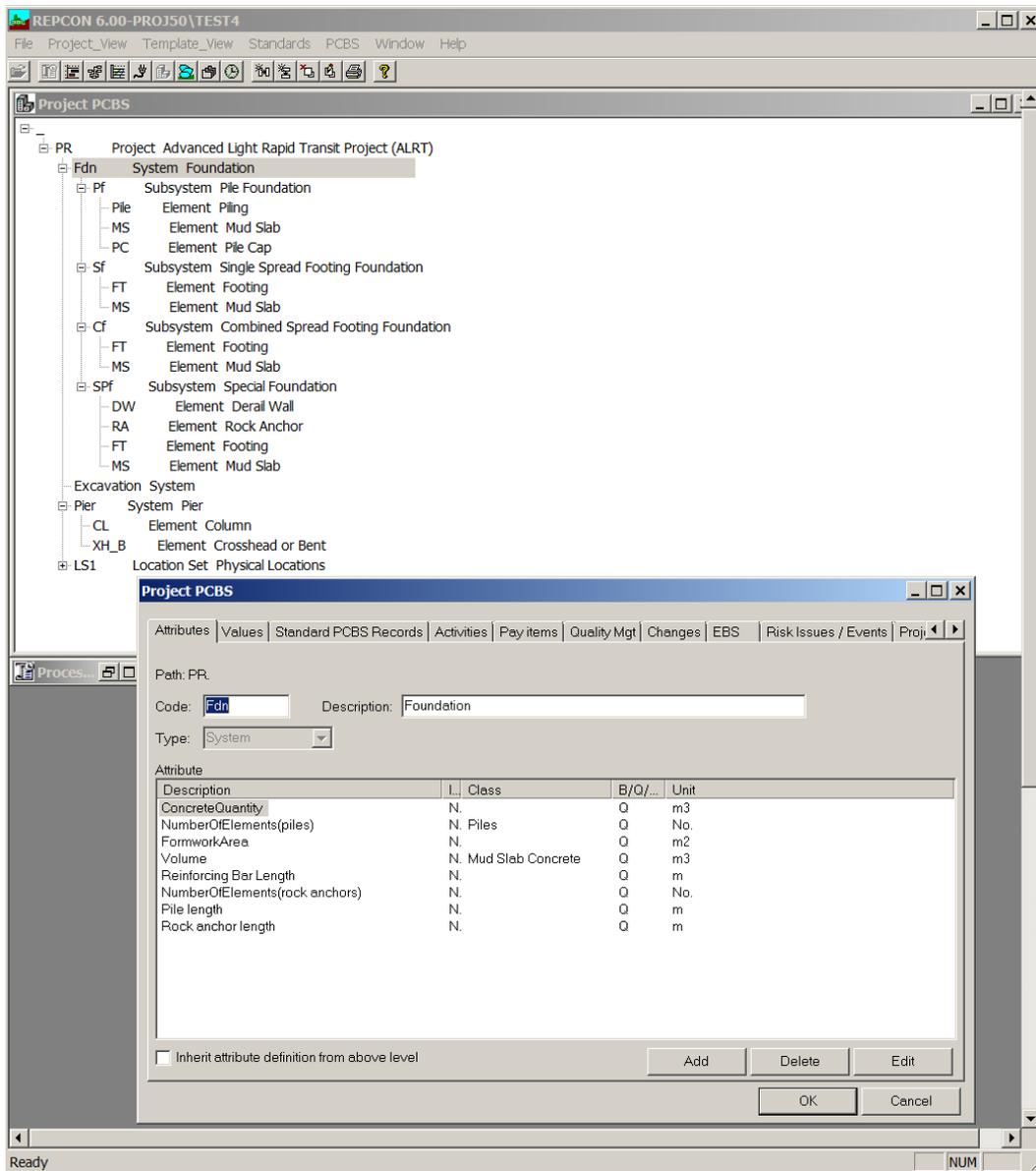


Figure 5.1 The PCBS (product) view of project 1. The dialogue box shows how users can define attributes (e.g. concrete quantity) for product items.

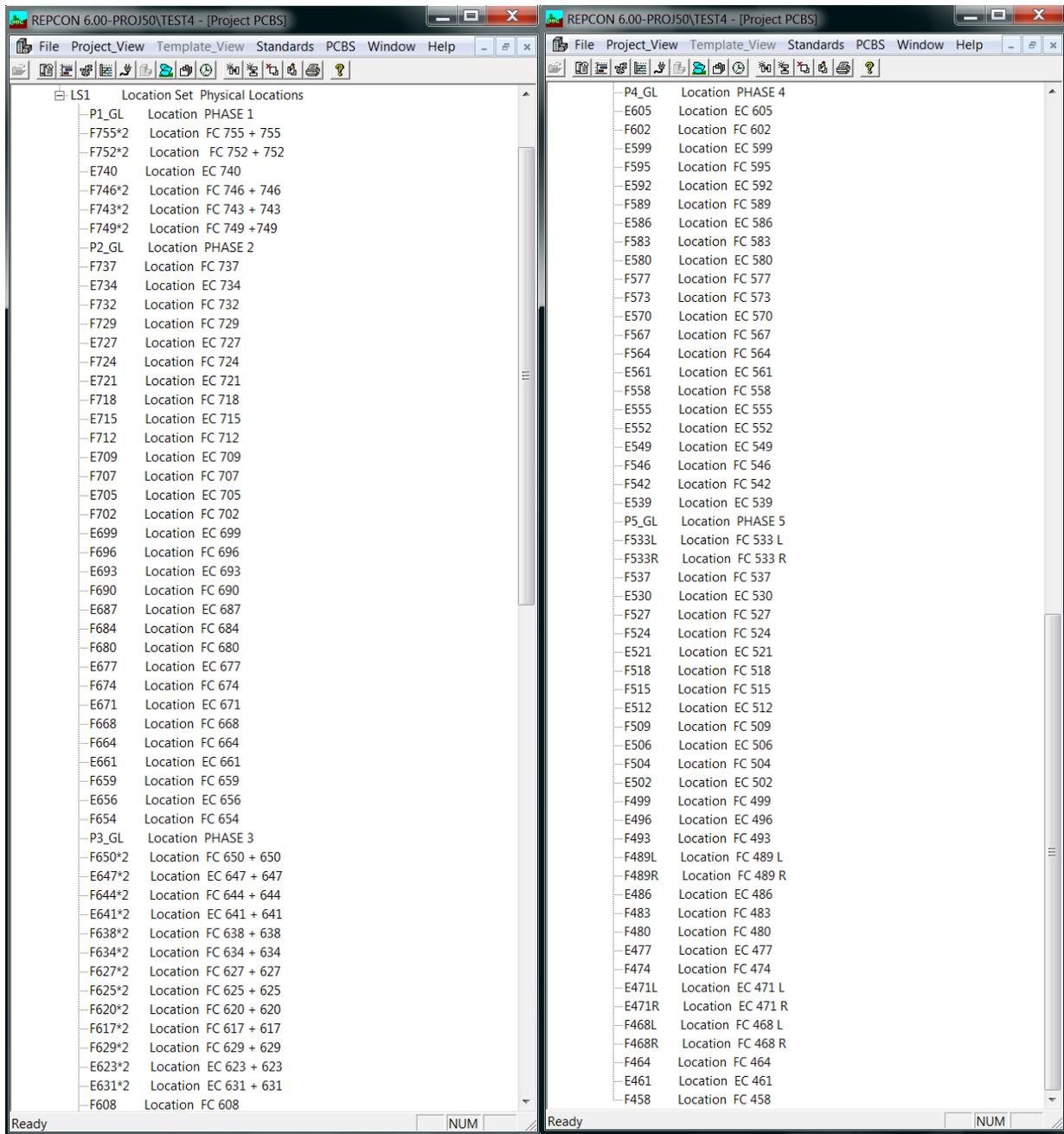


Figure 5.2 The PCBS view (physical work locations) of project 1





Figure 5.4 A photo of actual columns of project 1

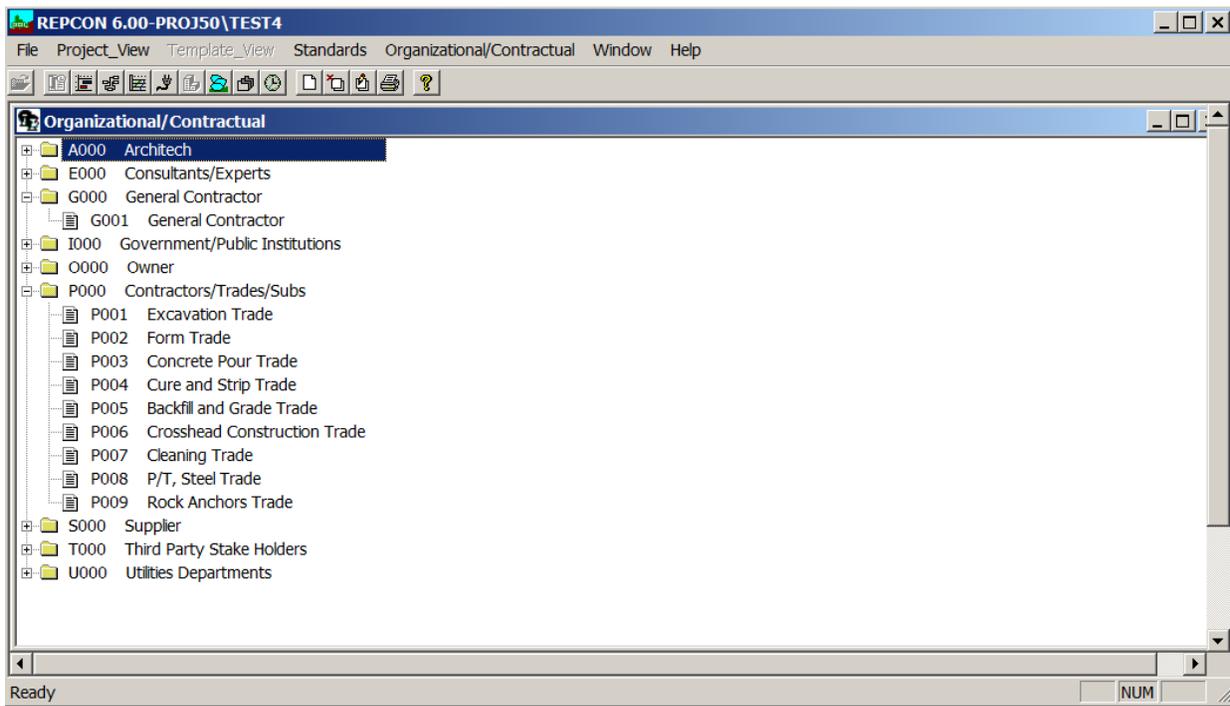


Figure 5.5 The organization view of project 1

### 5.3.1.1 Project 1- demonstration case 1

#### 1. CM analytics tasks:

- To understand the values and patterns of behaviour of execution conditions in terms of the temporal and spatial distribution of the "excavate and mud slab" activity.
- Based on the foregoing understanding, try to identify potential causes of execution conditions as a function of time and/or location and identify their potential impacts.

#### 2. Supporting CM functions/tasks: Time management/time performance control

#### 3. Project data shown in tabular reports/images for answering CM questions:

- Type of data: As-planned and as-built schedule data.
- Approximate scope of data: Schedule data for one activity repetitively executed through 30 locations (locations of the second of five construction phases).
- Original CM application of data: Time management/time performance control.

#### 4. Current data reporting practice: Bar graph of as-planned and as-built schedules (Figure 5.6).

#### 5. Visualization approach: A non-traditional as-planned vs. as-built schedule (Figure 5.7).

#### 6. Descriptions of the exploration-answer process using both a CM data visualization environment and current traditional tabular data reporting functionalities: see Table 5.1.

Table 5.1 Descriptions of the exploration- answer process using the CM data visualization environment and current (traditional) data reporting functionalities: project 1-demonstration case 1

<b>CM analytics tasks (i.e. CM questions):</b>			
<ul style="list-style-type: none"> <li>• To understand the values and patterns of behaviour of execution conditions in terms of the temporal and spatial distribution of the "excavate and mud slab" activity.</li> <li>• Based on the foregoing understanding, try to identify potential causes of execution conditions as a function of time and/or location and identify their potential impacts.</li> </ul>			
<b>Data reports</b>	<b>Primary interactive operations</b>	<b>Insightful messages received and/or answers identified</b>	<b>Viewing time</b>
<p>A traditional bar graph type of as-planned vs. as-built schedule: <i>Figure 5.6</i></p>	<p><i>Figure 5.6</i></p> <ul style="list-style-type: none"> <li>• <b>Visualization selection:</b> Select to generate the "bar graph schedule" visualization.</li> <li>• <b>Data selection:</b> Select the "excavate and mud slab" activity and the location range between F 737 and F 654 (i.e. the locations of phase 2 construction). Also select two schedule data versions for comparison. One is the final actual version as the current schedule and the other is the originally planned one as the target schedule.</li> <li>• <b>Visual encoding selection:</b> The visual encoding has been fixed as seen in the figure except for the following options: <ul style="list-style-type: none"> <li>○ The vertical chart strips representing dates can be highlighted to denote whether the dates are non-working days. Here non-working days are highlighted in the light blue color.</li> </ul> </li> <li>• <b>Data granularity selection:</b> The time dimension is at the level of granularity of "day" and the location dimension is at the level of granularity of "location".</li> </ul>	<p><i>Figure 5.6</i></p> <ol style="list-style-type: none"> <li>1. The work of the "excavate and mud slab" activity in the locations ranging from F 737 to F 654 was planned to be executed continuously in the time window between mid October 1983 and early December 1983.</li> <li>2. The work of the "excavate and mud slab" activity in the locations ranging from F 737 to F654 was actually executed non-continuously and spanned a longer period of time from mid October 1983 to early February 1984. There were four time windows when work was discontinued with the first two ones being the more serious ones (spanning approximately two weeks and one week respectively). Therefore, it is suspected that reasons for the finish delay of the "excavate and mud slab" activity executed between F 737 and F 654 are a function of the time dimension, and during the time period between mid November 1983 and early December 1983 some unexpected events may have occurred thereby preventing the activity from being executed in a timely manner.</li> </ol>	<p><i>Figure 5.6</i> <b>10 seconds</b></p>
<p>A non-traditional as-planned vs. as-built schedule generated from</p>	<p><i>Figure 5.7</i></p> <ul style="list-style-type: none"> <li>• <b>Visualization selection:</b> Select to generate the "linear planning chart" visualization.</li> <li>• <b>Data selection:</b> the same as for Figure</li> </ul>	<p><i>Figure 5.7</i></p> <ol style="list-style-type: none"> <li>1. In addition to the 1st message obtained in Figure 5.6, it is also observed that the activity was planned to be executed sequentially by the spatial ordering of physical locations in the location range F 737 to F654.</li> </ol>	<p><i>Figure 5.7</i> <b>10 seconds</b></p>

<b>CM analytics tasks (i.e. CM questions):</b>			
<ul style="list-style-type: none"> <li>• To understand the values and patterns of behaviour of execution conditions in terms of the temporal and spatial distribution of the "excavate and mud slab" activity.</li> <li>• Based on the foregoing understanding, try to identify potential causes of execution conditions as a function of time and/or location and identify their potential impacts.</li> </ul>			
<b>Data reports</b>	<b>Primary interactive operations</b>	<b>Insightful messages received and/or answers identified</b>	<b>Viewing time</b>
a CM data visualization environment: <i>Figure 5.7</i>	5.6 <ul style="list-style-type: none"> <li>• <b>Visual encoding selection:</b> The visual encoding has been fixed as seen in the figure except for the following options:               <ul style="list-style-type: none"> <li>○ The dotted lines can be selected to be shown or not shown. If the dotted lines are chosen for display, users can further choose to use the dotted lines for connecting solid lines by the spatial ordering of physical locations or the ordering of start dates of the activity (there are different start dates for the activity executed at different locations). Here dotted lines are shown and connect the solid activity lines at each location by the ordering of start dates.                   <ul style="list-style-type: none"> <li>○ Non-working days highlighted</li> </ul> </li> <li>• <b>Data granularity selection:</b> The same as for Figure 5.6</li> </ul> </li> </ul>	2. In addition to the 2nd message obtained in Figure 5.6, the following insights are also observed: <ol style="list-style-type: none"> <li>1) The planned sequencing of the activity was disrupted during actual execution. The activity advanced through locations randomly, a prominent phenomenon differentiating between what was planned and what was actual. The "difference" pattern in terms of a comparison between planned and actual activity sequencing by the spatial ordering of physical locations indicates that reasons for the displaced temporal and spatial distribution of the "excavate and mud slab" activity are potentially a function of one or both of the location or time dimensions. However, due to the fact that spatial displacements consistently occurred throughout the actual execution time window, it is suspected that the reasons may be more of a function of the location dimension (i.e. location dependent issues)</li> <li>2) The displaced spatial distribution of the activity also indicates a potential impact on the location dimension (e.g. neighborhood and vehicular roads nearby the work locations) because sections of work locations cannot be swiftly closed off and returned to normal. Also impacted is the project participant dimension (i.e. the trade executing this activity) due to its need to constantly move equipment for long distances from location to location.</li> </ol>	

Advanced Light Rapid Transit Project (ALRT)

Base File Used: PROJ50\TEST4  
 Target File Used: PROJ44\TEST4

Progress Date: 17SEP84 12:00am

Date Selection: Act/Sch/Early  
 Schedule Window  
 Time: 15OCT83 8:00am To 15FEB84 5:00pm  
 Locations: F737 To F654  
 Non-working time according to calendar: 1

Nomenclature: P = Planned IP/C = in-progress/completed C = Critical NC = Noncritical S = Start F = Finish [ ] = estimated  
 Current Target Current Target Current Target Current Target  
 S P Milestone ▼ ▼ Derived P-NC [ ] [ ] Activity P-NC [ ] [ ] Positive Float [ ] [ ]  
 S IP/C Milestone ▼ ▼ Derived P-C [ ] [ ] Activity P-C [ ] [ ] Negative Float [ ] [ ]  
 F P-NC Milestone ▲ ▲ [% Installed] [% Installed] [% Installed] [% Installed] Progress Date [ ] [ ]  
 F P-C Milestone ▲ ▲ Derived IP/C-NC [ ] [ ] Activity IP/C-NC [ ] [ ]  
 F IP/C-NC Milestone ▲ ▲ [% Installed] [% Installed] [% Installed] [% Installed]  
 F IP/C-C Milestone ▲ ▲ Derived IP/C-C [ ] [ ] Activity IP/C-C [ ] [ ]  
 Hammock [ ] [ ]

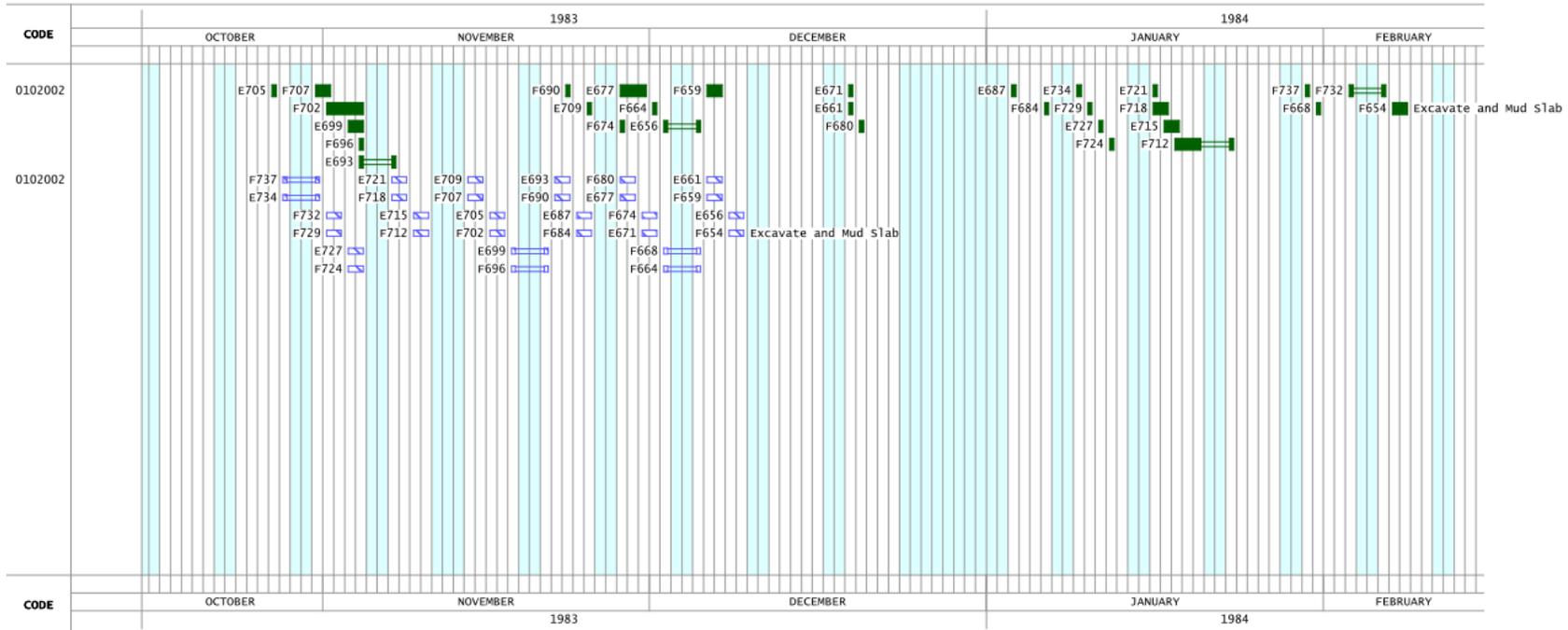


Figure 5.6 A traditional bar graph of as-planned (blue bars) vs. as-built (green bars) schedule representing the "excavate and mud slab" activity executed in the location range F 737 to F 654

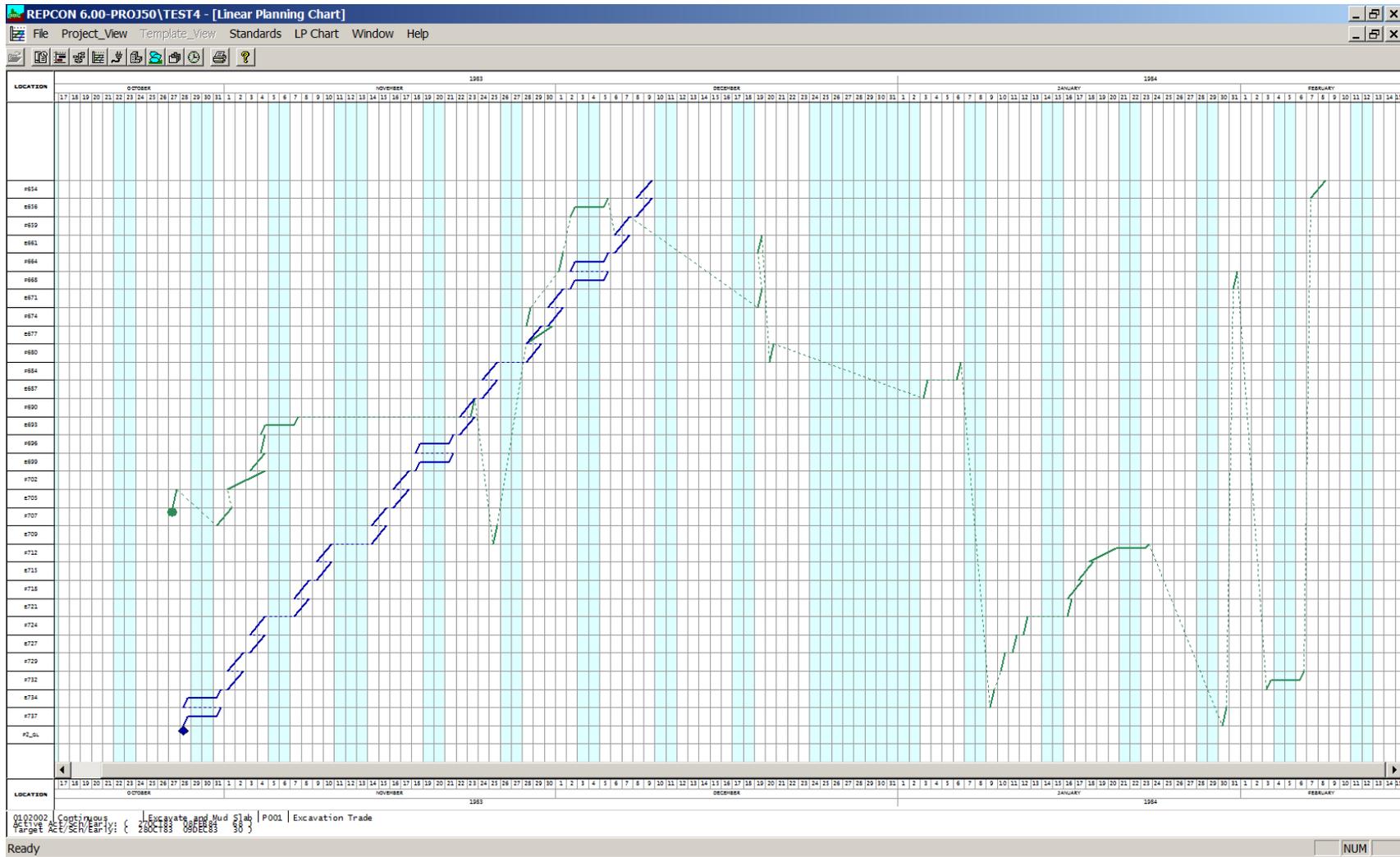


Figure 5.7 A non-traditional as-planned (blue lines) vs. as-built (green lines) schedule generated from a CM data visualization environment representing the "excavate and mud slab" activity executed in the location range F 737 to F 654

### 5.3.1.2 Project 1 - demonstration case 2

#### 1. CM analytics tasks:

- To understand the values and patterns of behaviour of planned conditions in terms of relevant product quantities for foundations and columns.
- Based on the foregoing understanding, try to identify potential impacts of these conditions on planned time performance in terms of durations of individual activities.

#### 2. Supporting CM functions/tasks: Time management/assess quality of duration estimates

#### 3. Project data used:

- Type of data:
  - Planned quantities of concrete, formwork, reinforcing bars, piling, and rock anchors for the vertical structures of a rapid transit guideway system
  - Planned schedule data
- Approximate scope of data: Planned schedule data of two activities and quantitative data for attributes of two products at 102 locations.
- Original CM application of data: Scope management; cost estimate; scheduling

#### 4. Current data reporting practice (for product attributes data): tabular reports (Figure 5.8).

#### 5. Visualization approach: product attribute graphics (Figures 5.9 and 5.10, Figure 5.11(b) ), linear planning chart (Figure 5.11 (a)).

#### 6. Descriptions of the exploration-answer process using both a CM data visualization environment and current traditional tabular data reporting functionalities: see Table 5.2.

Table 5.2 Descriptions of the exploration- answer process for both the use of a CM data visualization environment and the current data reporting functionalities in project 1-demonstration case 2

<b>CM analytics tasks (i.e. CM questions):</b>			
<ul style="list-style-type: none"> <li>• To understand the values and patterns of behaviour of planned conditions in terms of relevant product quantities for foundations and columns.</li> <li>• Based on the foregoing understanding, try to identify potential impacts of these conditions on planned time performance in terms of durations of individual activities.</li> </ul>			
<b>Data reports</b>	<b>Primary interactive operations</b>	<b>Insightful messages received and/or answers identified</b>	<b>Viewing time</b>
14 pages of the kind of tabular report seen in <i>Figure 5.8</i>	<p><i>Figure 5.8</i></p> <ul style="list-style-type: none"> <li>• <b>Data report selection:</b> Select to generate "PCBS attribute" reports</li> <li>• <b>Data selection:</b> Select PCBS attributes datasets that are related to the "foundation" and the "column" and include only planned attribute values</li> </ul>	<p><i>Figure 5.8</i></p> <ol style="list-style-type: none"> <li>1. There are 14 locations requiring rock anchors for resisting the movement of foundations; and there are 25 locations needing deep pile foundations to improve bearing capacity of soils.</li> <li>2. Number and lengths of piles and rock anchors are indirect indicators for the soil conditions at individual locations, i.e., the more and longer piles/rock anchors a foundation needs, the more problematic the soil conditions at a location are. Thus, a pattern of problematic soil conditions that cluster in certain locations indicates potential impacts on the location dimension and its associated activity dimension. Activities executed in these locations may be prone to prolonged durations, design changes for stabilizing soils, and consequently start delays for subsequent activities.</li> </ol>	<p><i>Figure 5.8</i> <b>150 seconds</b></p>
Three images generated from a CM data visualization environment: <i>Figure 5.9</i> <i>Figure 5.10</i> <i>Figure 5.11</i>	<p><i>Figure 5.9</i></p> <ul style="list-style-type: none"> <li>• <b>Visualization selection:</b> Select to generate the "product attribute" visualization.</li> <li>• <b>Data selection:</b> Select product attribute datasets that are of the "number of elements (piles)" attribute, associated with all locations, and associated with the "foundation system" and "column element".</li> <li>• <b>Visual encoding selection:</b> Specify that the location dimension be mapped to the X axis and the product dimension be mapped to the singular selection combo box.</li> <li>• <b>Data granularity selection:</b> The level of granularity of location dimension has been fixed at the "location" level, and the level of granularity of product dimension is dependent on singular product selection (e.g. if users select product attribute datasets that are associated with the "foundation system", the level of granularity chosen is</li> </ul>	<p><i>Figure 5.9</i> The same messages received from viewing Figure 5.8</p> <p><i>Figure 5.10</i></p> <ol style="list-style-type: none"> <li>1. Overall, the required formwork areas for foundations and columns are similar (Figure 5.10(a)); the required concrete quantities for foundations are much more than the ones for columns (Figure 5.10 (b)); the required reinforcing bar lengths for columns are slightly more than the ones for foundations and there is one location demanding many more steel bars for the column (Figure 5.10 (c)). The visual perception of the stark difference in terms of required concrete quantities between foundations and columns (see Figure 5.10(b)) prompts the reasoning of potential impacts on the product dimension and its associated activity dimension. The impact effect on activities is that the planned execution durations of "pour columns" activities may take much less</li> </ol>	<p><i>Figure 5.9</i> <b>5 seconds</b></p> <p><i>Figure 5.10</i> <b>5 seconds</b></p>

<b>CM analytics tasks (i.e. CM questions):</b>			
<ul style="list-style-type: none"> <li>• To understand the values and patterns of behaviour of planned conditions in terms of relevant product quantities for foundations and columns.</li> <li>• Based on the foregoing understanding, try to identify potential impacts of these conditions on planned time performance in terms of durations of individual activities.</li> </ul>			
<b>Data reports</b>	<b>Primary interactive operations</b>	<b>Insightful messages received and/or answers identified</b>	<b>Viewing time</b>
	<p>automatically the "system" level). Repeat the foregoing steps three more times except that the data selection step is to select product attribute datasets that are of the "pile length", "number of elements (rock anchors)", and "rock anchor length" attributes respectively.</p> <p><i>Figure 5.10</i></p> <ul style="list-style-type: none"> <li>• <b>Visualization selection:</b> Select to generate the "product attribute" visualization.</li> <li>• <b>Data selection:</b> Select product attribute datasets that are of the "concrete quantity", associated with all locations, and associated with the "foundation system" and "column element".</li> <li>• <b>Visual encoding selection:</b> Specify that the location dimension be mapped to the X axis and the product dimension be mapped to the Y axis.</li> <li>• <b>Data granularity selection:</b> The same as the one for Figure 5.9</li> </ul> <p>Repeat the foregoing steps two more times except for the data selection step (select product attribute datasets for the "formwork area" and "reinforcing bar length").</p> <p><i>Figure 5.11(a)</i></p> <ul style="list-style-type: none"> <li>• <b>Visualization selection:</b> Select to generate the "linear planning schedule" visualization.</li> <li>• <b>Data selection:</b> Select the "pour footing" and "pour column" activities.</li> <li>• <b>Visual encoding selection:</b> The visual encoding has been fixed as seen in the figure. However, the dotted lines can be selected to be on or off.</li> <li>• <b>Data granularity selection:</b> The time dimensions is at the level of granularity of "day" and the location dimension is at the level of granularity of "location".</li> </ul>	<p>time than the ones of "pour foundations (footings)" activities.</p> <p><i>Figure 5.11</i></p> <p>1. Fig 5.1.11(a) is a linear planning chart showing planned schedules for the "pour footing" and "pour column" activities executed in the first 54 locations. While in the figure labels of lines for informing what activities these lines represent are not available (labelling lines in a linear planning chart is an image design problem), it is known that the connecting lines to the right represent the "pour footing" activity and the other connecting lines represent the "pour column" activity. A steady pattern found is that the production rate (locations/day) of "pour footing" activity is always better than the one of "pour column" activity (i.e. always complete one more location in a day).</p> <p>By comparing Figure 5.11(a) and Figure 5.11(b), a "similarity" pattern is observed (i.e. required concrete quantities for foundations are always <b>much more than</b> the ones for columns (see Figure 5.11(a)), and the production rate (locations/day) of "pour footing" activity is also <b>more than</b> its counterpart). However, this identification of the pattern is counterintuitive to the instinct that more quantities require longer time to produce. For the case at hand, the task of placing concrete in column differs considerably from the mass pour that can be used for footings may alert one to review and ensure correct values of parameters such as work method, crew size, and standard work output rate have been used in arriving at the duration estimates.</p>	<p><i>Figure 5.11</i> <b>10 seconds</b></p>

Advanced Light Rapid Transit Project (ALRT)

Planned Attribute Values

File Used: PROJ50\TEST4

Report Date: 2011-03-03  
 Report Time: 10:44:44am  
 Revision Number: 0  
 Progress Date: 16SEP04

Select: Selectively

PHYSICAL COMPONENT BREAKDOWN STRUCTURE		PLANNED ATTRIBUTE VALUE		CLASS	TYPE	UNIT	AGGREGATED
PCBS PATH	DESCRIPTION	INHERITED DESCRIPTION	REL COND VALUE 1				
PR.Fdn	Foundation	N ConcreteQuantity				Quantitative m3	Y
		P1_GL	No value entered at below component.				
		F755*2	EQ	87.00			
		F752*2	EQ	65.30			
		E740	EQ	56.30			
		F746*2	EQ	87.00			
		F743*2	EQ	87.00			
		F749*2	EQ	87.00			
		P2_GL	No value entered at below component.				
		F737	EQ	63.80			
		E734	EQ	37.70			
		F732	EQ	50.80			
		F729	EQ	50.80			
		E727	EQ	37.70			
		F724	EQ	63.80			
		E721	EQ	37.70			
		F718	EQ	63.80			
		E715	EQ	56.30			
		F712	EQ	120.00			
		E709	EQ	72.50			
		F707	EQ	72.50			
		E705	EQ	56.30			
		F702	EQ	85.40			
		E699	EQ	72.50			
		F696	EQ	90.20			
		E693	EQ	72.50			
		F690	EQ	85.40			
		E687	EQ	50.80			
		F684	EQ	74.40			
		F680	EQ	74.40			
		E677	EQ	50.80			
		F674	EQ	74.40			
		E671	EQ	50.80			
		F668	EQ	74.40			
		F664	EQ	74.40			
		E661	EQ	50.80			
		F659	EQ	50.80			
		E656	EQ	37.70			
		F654	EQ	58.20			
		P3_GL	No value entered at below component.				
		F650*2	EQ	111.70			
		E647*2	EQ	65.30			
		F644*2	EQ	87.80			
		E641*2	EQ	83.20			
		F638*2	EQ	137.10			
		F634*2	EQ	137.10			
		F627*2	EQ	83.70			
		F625*2	EQ	83.70			
		F620*2	EQ	111.40			
		F617*2	EQ	156.00			
		F629*2	EQ	83.70			
		E623*2	EQ	83.70			
		E631*2	EQ	83.70			
		F608	EQ	181.20			
		P4_GL	No value entered at below component.				
		E605	EQ	37.70			
		F602	EQ	129.20			
		E599	EQ	37.70			
		F595	EQ	94.70			
		E592	EQ	56.30			
		F589	EQ	85.40			
		E586	EQ	56.30			
		F583	EQ	85.40			
		E580	EQ	37.70			
		F577	EQ	63.80			
		F573	EQ	63.80			
		E570	EQ	37.70			
		F567	EQ	63.80			

Figure 5.8 First page of the 14 page tabular report of planned values of product attributes (e.g. concrete quantity, formwork area) for the foundations and columns at all work locations

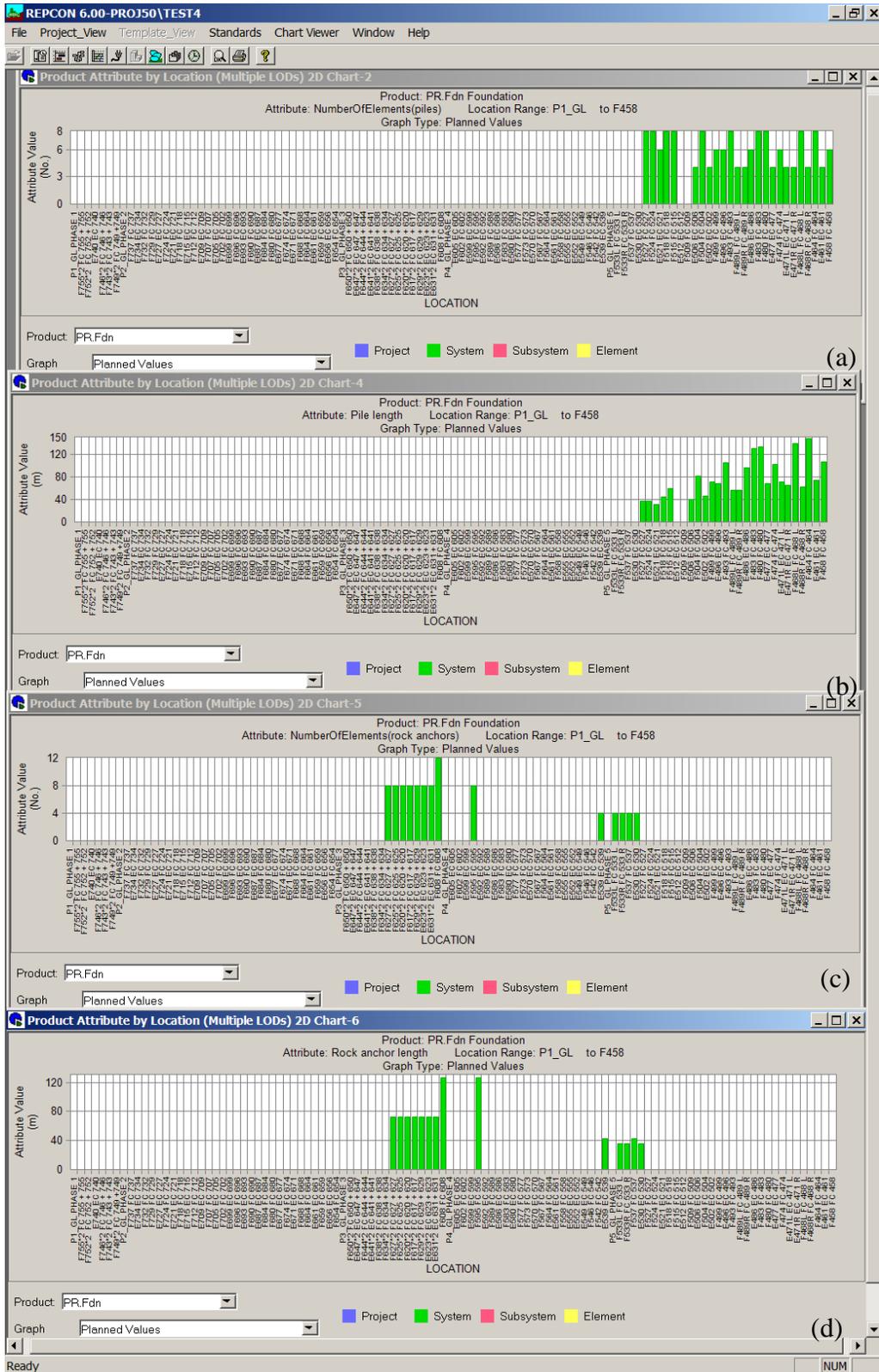


Figure 5.9 (a&b) Number and lengths of piles by location, (c&d) number and lengths of rock anchors by location.

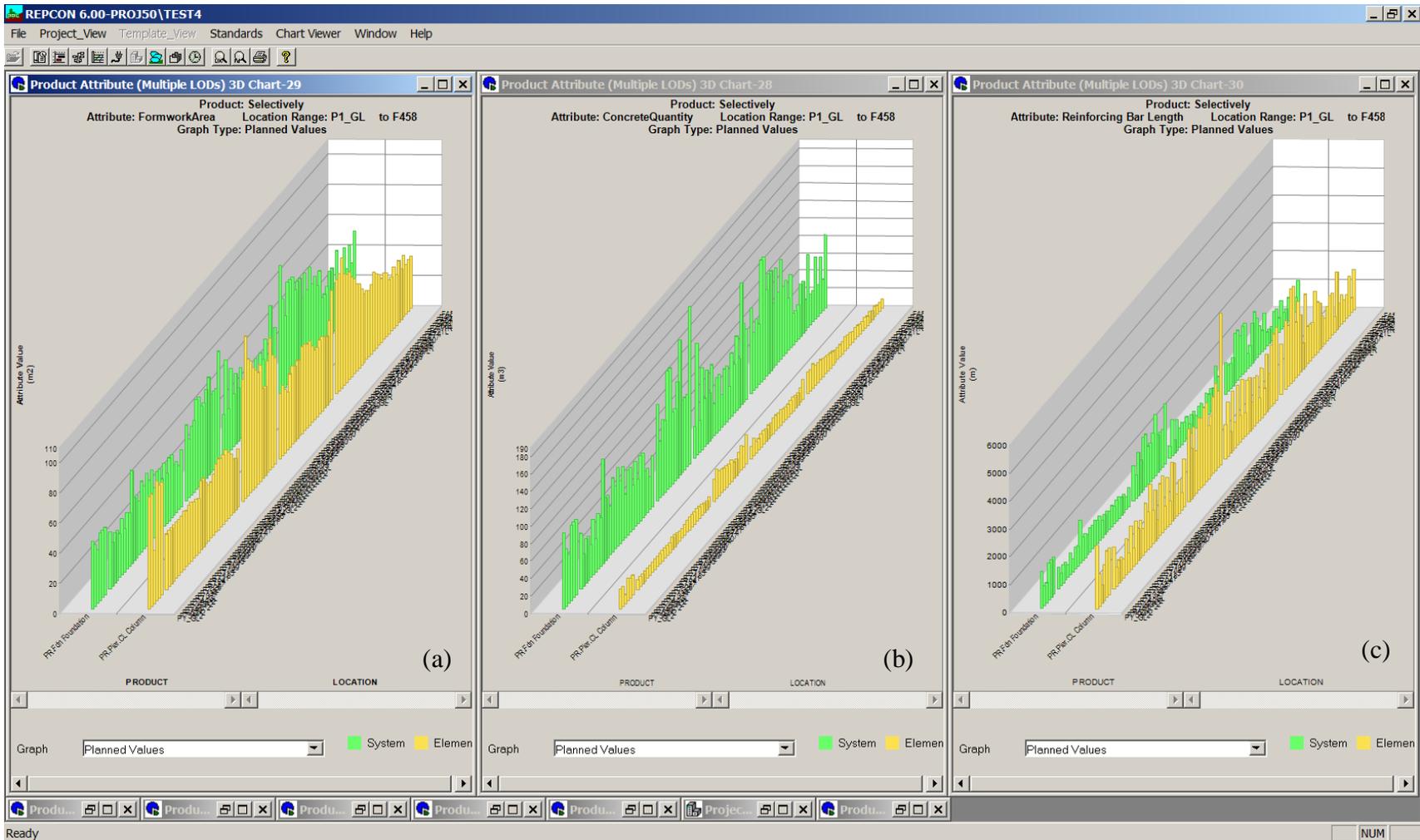


Figure 5.10 (a) Planned formwork areas, (b) Planned concrete quantities, (c) Planned reinforcing bar lengths required by the foundations and columns at all locations

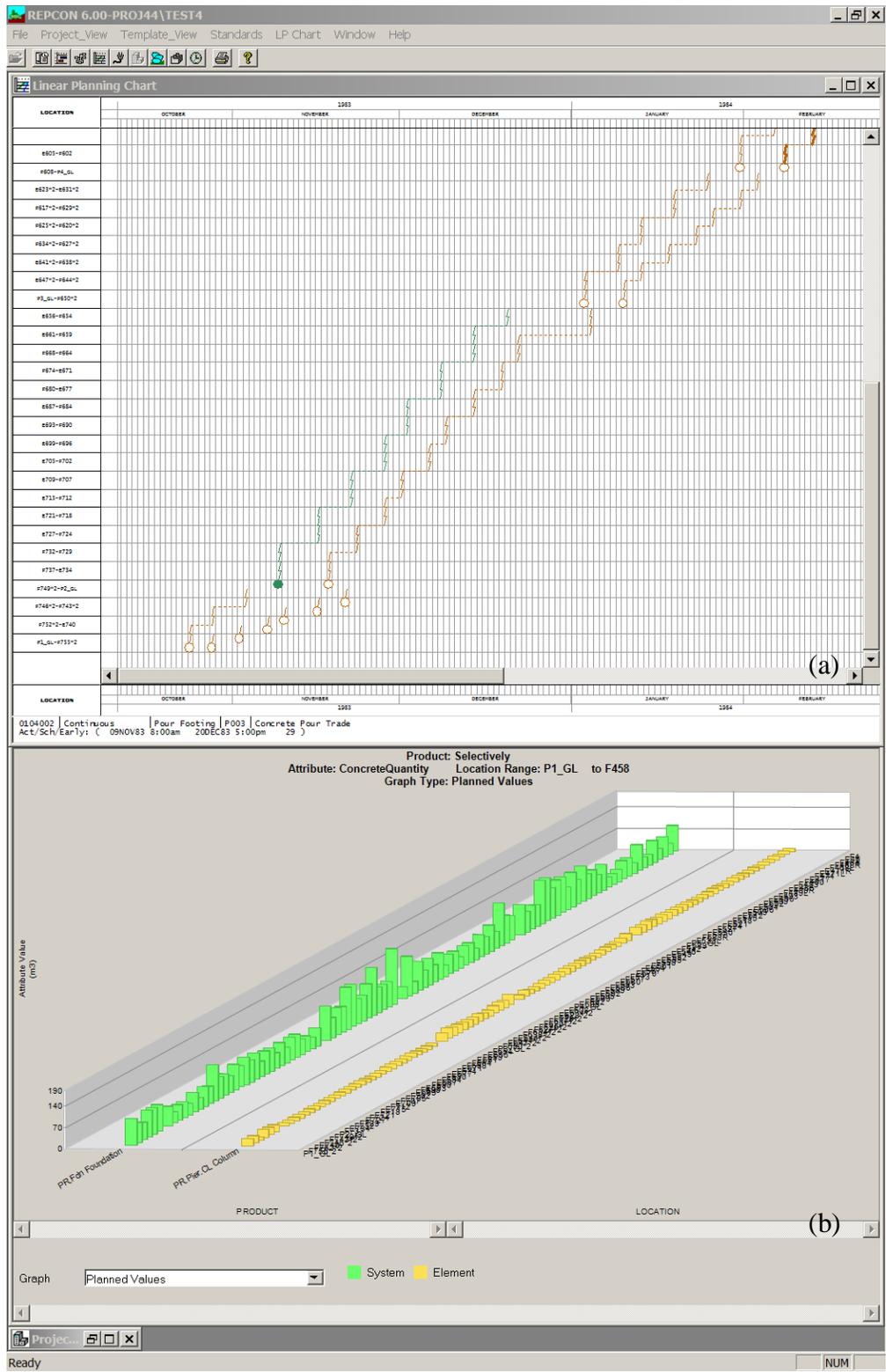


Figure 5.11 (a) Planned schedule for "pour footing" (left connecting lines) and "pour column" (right connecting lines) activities executed at the first 54 locations, (b) Planned concrete quantities required by the foundations and columns at all locations

### **5.3.2 Demonstration cases- project 2**

The project 2 used here is a 6 story residential building project. The process view of the project is not modeled because this project was used mainly for collecting and visualizing quality control related data. Hence, the focus of project modeling and project data collection is more on products and the trades responsible for constructing the products. The PCBS view of this project consists of rather detailed lists of project components and all work locations, as shown in Figures 5.12 and 5.13. The organizational view of this project, i.e., project participants of the project, is depicted in Figure 5.14. The as-built view in terms of deficiency records and their association with the foregoing views can be found in Figure 5.15.

The project started in 2008 and completed in 2009. Near the completion of the project, the project owner (the developer) conducted very rigorous and thorough inspections on the completed work (It is observed that the development company is a large seasoned one with operations in a number of venues in Canada). The inspection process started in March 2009 and continued until August 2009. The tracking of deficiencies was by sections in terms of individual suites (84 suites in total), common areas on each floor (1st floor to 6th floor, underground parkade, and roof), and exterior areas. Developer personnel generated 93 Word documents, one for each section, for recording deficiencies identified and re-inspected in these sections. For each inspection section, there were approximately three to six cycles of inspection and re-inspection until all deficiencies were fixed. On average there were 100~150 deficient items identified for each section. In each Word document, dates of inspecting/re-inspecting the section were recorded, and each deficient item recorded includes the information of location, description, and whether or not a long time to fix is required (a special concern where lengthy procurement times are involved). The description of a deficiency contains information about the deficient product, local location of the product (e.g. upper right corner of the wall) where quality issues were observed, and the quality deficiency of the product. Information as to the trade responsible for fixing the deficiency (the deficiency may or may not have been caused by the trade) was not recorded- the responsibility for correcting the deficiency was generally assigned by the job site superintendent. Further, detailed date data tended not be collected. The missing information about trades responsible for and inspection/re-inspection dates of each deficient item would be very useful for understanding quality performance of trades. Trades information was eventually

made available through requesting additional information from the developer, but since date data was not recorded, this data was not forthcoming. In support of this phase of the research, the semi-structured deficiency data of 84 suites in the original Word documents was parsed into structured data and entered into REPCON.

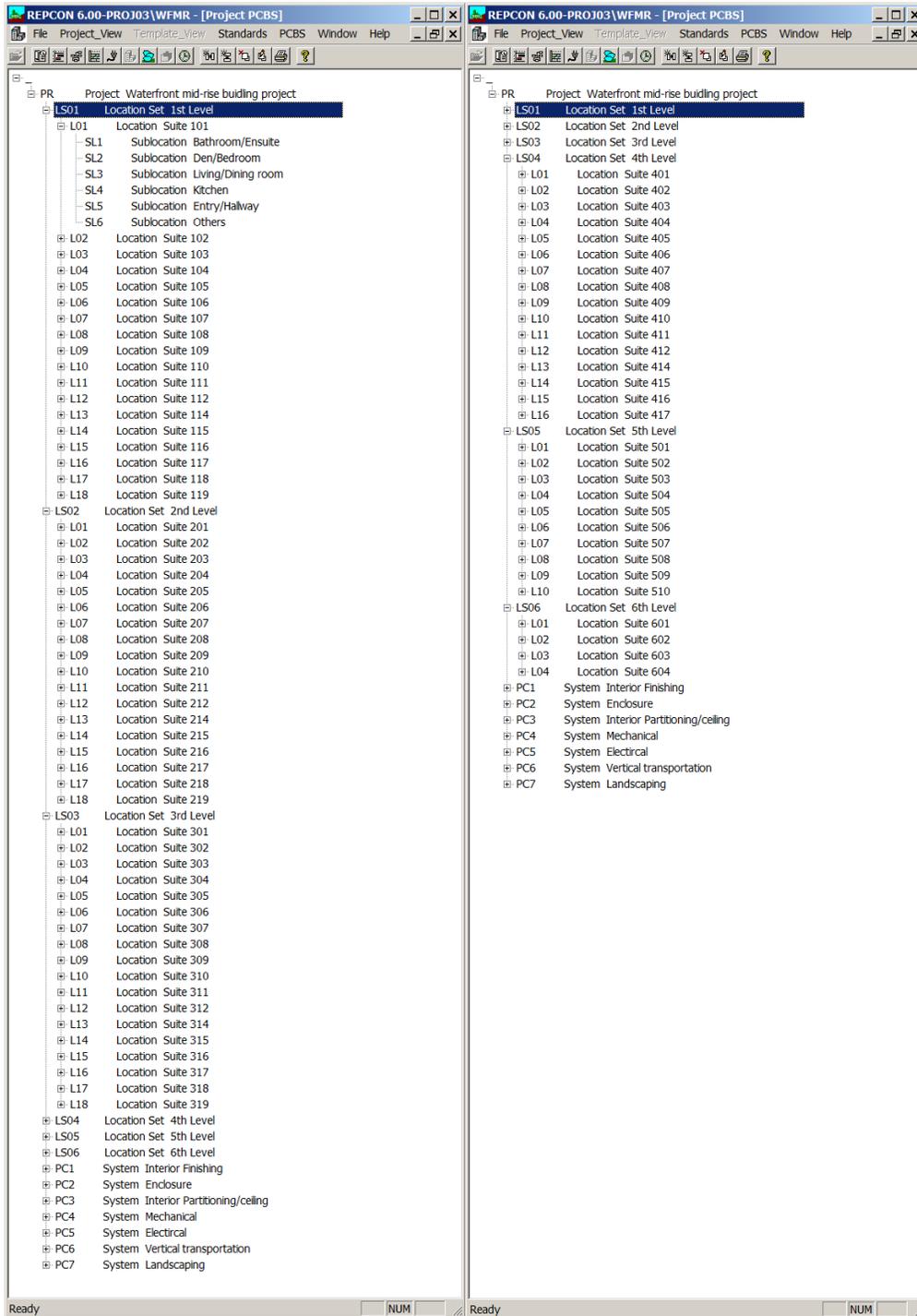


Figure 5.12 The PCBS view (physical locations) of project 2

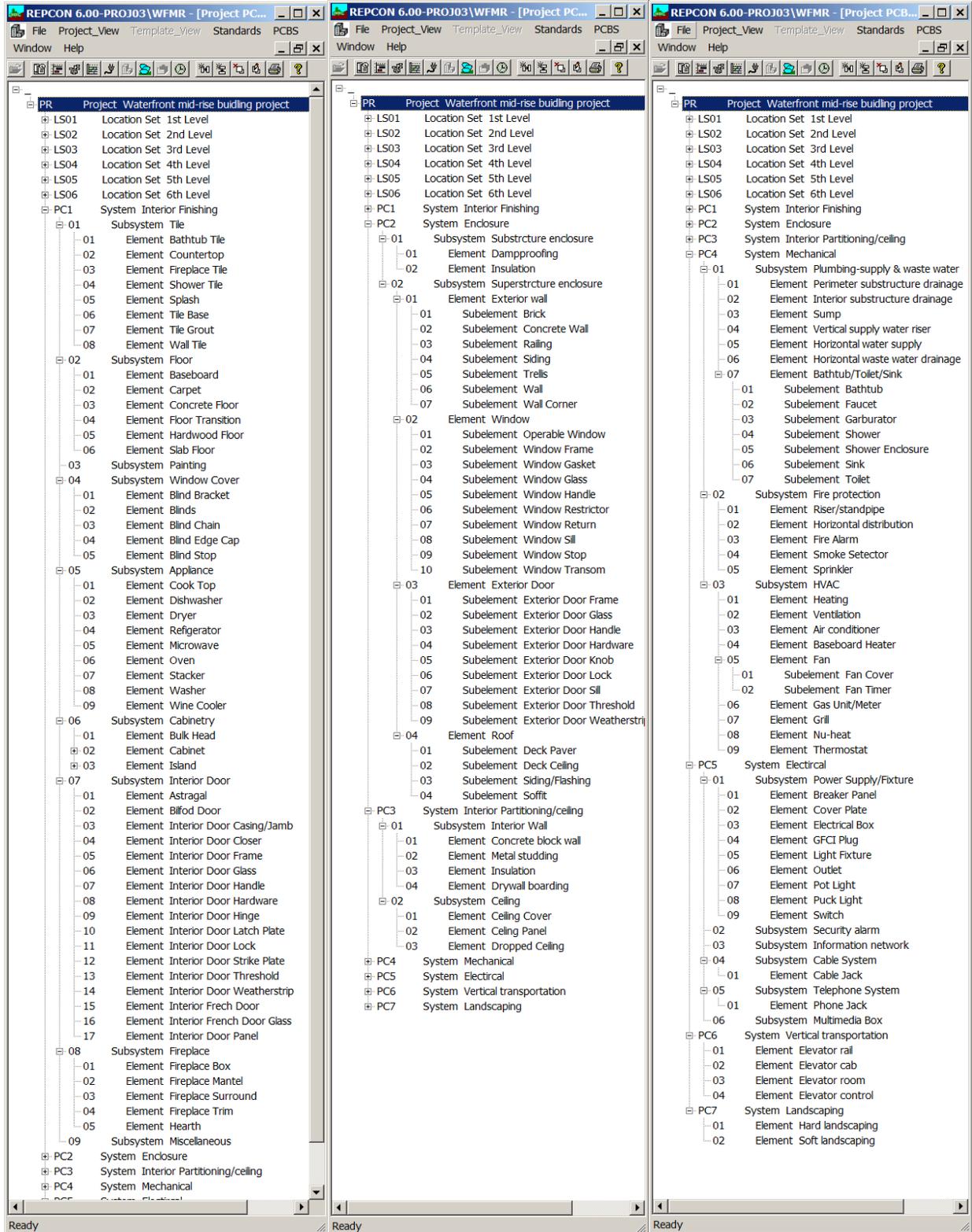


Figure 5.13 The PCBS view (products) of project 2

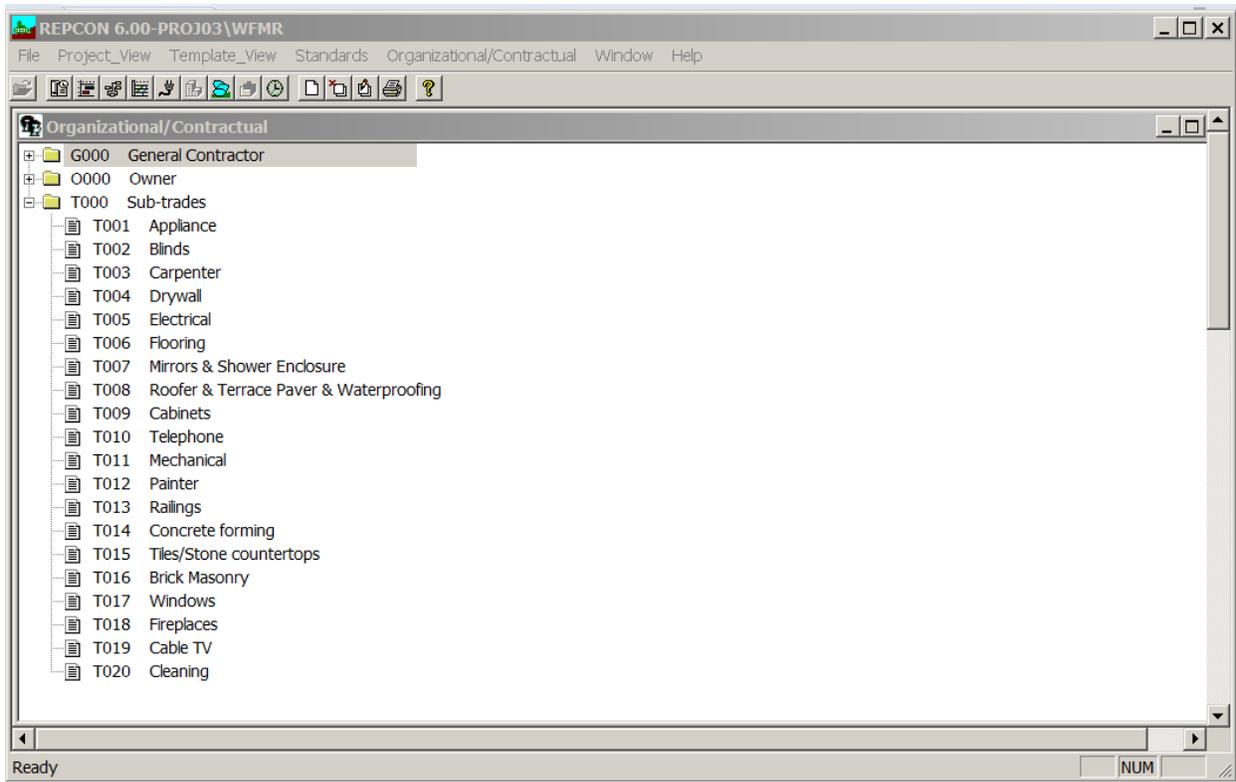


Figure 5.14 The organizational view (project participant) of project 2

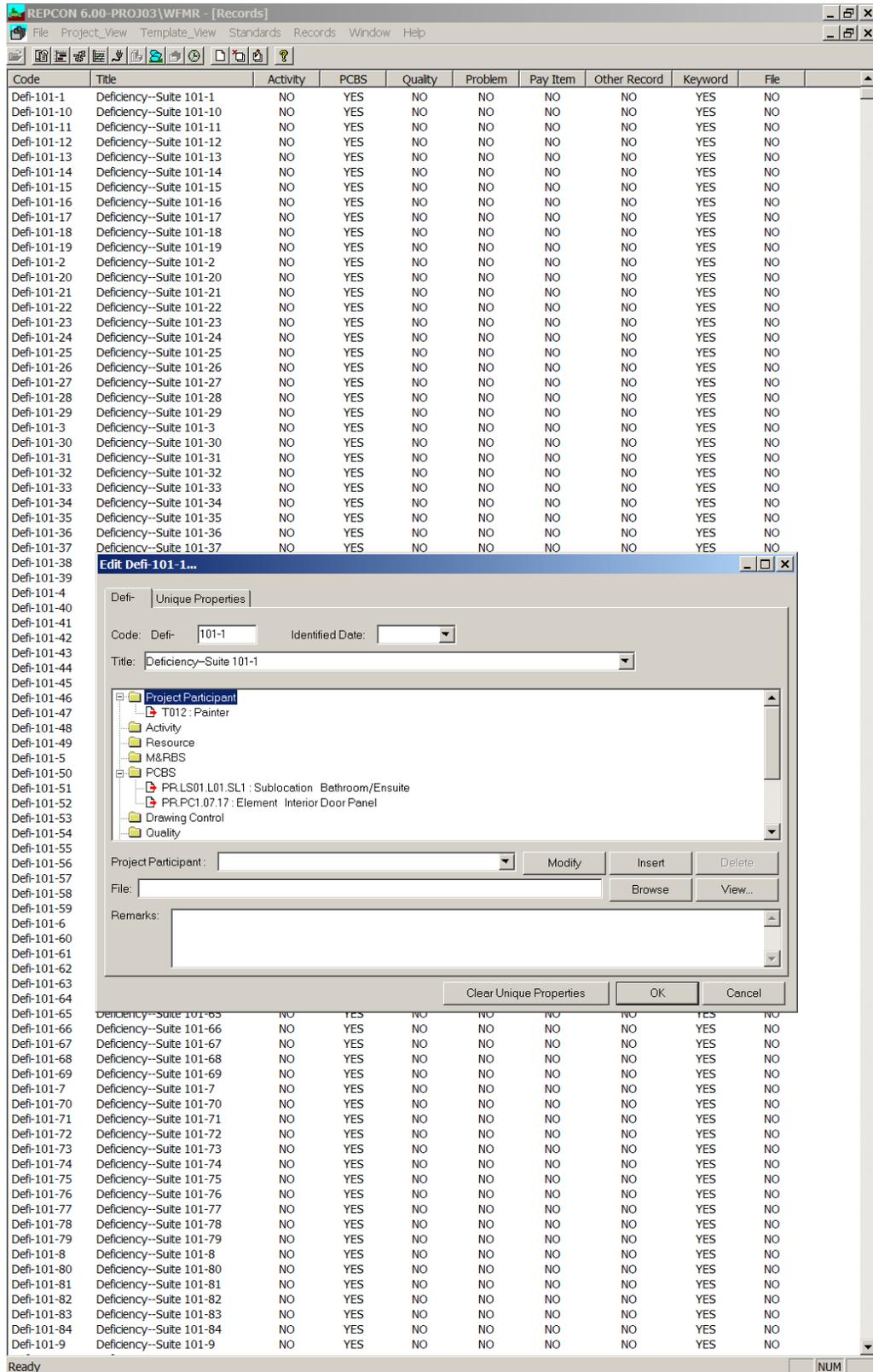


Figure 5.15 The As-built view (deficiency records) of project 2. The dialogue box shows how users can associate a deficiency item with items from other views such as the project participant view and PCBS view

### 5.3.2.1 Project 2- demonstration case 1

#### 1. **CM analytics tasks:**

- To understand the values and patterns of behaviour of quality performance in terms of number of deficiencies identified during inspection of interior construction.
- Based on the foregoing understanding, try to identify potential reasons for quality performance possibly as a function of trade, product, and/or location.

#### 2. **Supporting CM functions/tasks:** quality management/ quality performance control.

#### 3. **Project data used:**

- Type of data: records of deficiency data.
- Approximate scope of data: 8200 deficiency data entries.
- Original CM application of data: quality management/ quality performance control.

#### 4. **Current data reporting practice:** tabular reports (Figure 5.16).

#### 5. **Visualization approach:** as-built record (deficiency list) distribution graphics (Figures 5.17 ~ 5.19).

#### 6. **Descriptions of the exploration-answer process using both a CM data visualization environment and current traditional tabular data reporting functionalities:** see Table 5.3.

Table 5.3 Descriptions of the exploration-answer process for both the use of a CM data visualization environment and current data reporting functionalities: project 2-demonstration case 1

<b>CM analytics tasks (i.e. CM questions):</b>			
<ul style="list-style-type: none"> <li>• To understand the values and patterns of behaviour of quality performance in terms of number of deficiencies identified during inspection of interior construction.</li> <li>• Based on the foregoing understanding, try to identify potential reasons for quality performance possibly as a function of trade, product, and/or location.</li> </ul>			
<b>Data reports</b>	<b>Primary interactive operations</b>	<b>Insightful messages received and/or answers identified</b>	<b>Viewing time</b>
304 pages of the kind of tabular report seen in <i>Figure 5.16</i>	<p><i>Figure 5.16</i></p> <ul style="list-style-type: none"> <li>• <b>Data report selection:</b> Select to generate the "as-built record" reports.</li> <li>• <b>Data selection:</b> Select as-built record datasets that are of the "deficiency list" type. Include data fields in terms of products, locations, and project participants that are associated with deficiencies.</li> </ul>	<p><i>Figure 5.16</i></p> <p>The data examination task was abandoned after spending approximately five minutes scrolling and viewing the first 40 pages.</p>	<p><i>Figure 5.16</i> <b>&gt; 300 seconds</b></p>
Three images generated from a CM data visualization environment: <i>Figure 5.17</i> <i>Figure 5.18</i> <i>Figure 5.19</i>	<p><i>Figure 5.17</i></p> <ul style="list-style-type: none"> <li>• <b>Visualization selection:</b> Select to generate the "counts of as-built record" visualization.</li> <li>• <b>Data selection:</b> Select as-built record datasets that are of the "deficiency list" type.</li> <li>• <b>Visual encoding selection:</b> Specify that the location dimension be mapped to the X axis and the project participant dimension be mapped to the Y axis.</li> <li>• <b>Data granularity selection:</b> Select that the location dimension is at the levels of granularity of "location set" and "location" while the project participant dimension is at the level of "project participant".</li> </ul> <p><i>Figure 5.18</i></p> <p>Similar to the interaction features applied to generating Figure 5.17 except for:</p> <ul style="list-style-type: none"> <li>• <b>Data selection:</b> Select as-built record datasets that are of the "deficiency list" type AND excluding ones that are associated with the Painter trade and Cleaning trades.</li> </ul>	<p><i>Figure 5.17</i></p> <p>1. The deficiencies identified are from the 1st level to 6th level of the building. The 1st level to 3rd level have 18 suites, the 4th level has 16 suites, the 5th level has 10 suites, and the 6th level has 4 suites. The deficiencies identified involve 20 trades. A relatively much larger number of deficiencies is clustered in two trades (the Painter trade and the Cleaning trade). This clustering pattern is consistent throughout all locations and sub-locations (i.e. items of the location dimensions of different levels of granularity) thereby providing the insight that the deficiencies are a function of the project participant dimension.</p> <p><i>Figure 5.18</i></p> <p>1. Excluding the Painter and the Cleaning trade, deficiencies of other trades do not have the clustering patterns exhibited by the Painter and Cleaning trades. In other word, datasets excluding the deficiency data of the Painter and the Cleaning trades do not indicate reasons for deficiencies as a function of the location and/or project participant dimensions.</p> <p><i>Figure 5.19(a)</i></p> <p>1. For the Painter trade, deficiencies cluster in the "interior finishing" and "interior partitioning/ceiling" systems, and the intensities of clustering in both systems are similar. The pattern of clustering in the "interior</p>	<p><i>Figure 5.17</i> <b>5 seconds</b></p> <p><i>Figure 5.18</i> <b>5 seconds</b></p> <p><i>Figure 5.19(a)</i> <b>10 seconds</b></p>

<b>CM analytics tasks (i.e. CM questions):</b>			
<ul style="list-style-type: none"> <li>• To understand the values and patterns of behaviour of quality performance in terms of number of deficiencies identified during inspection of interior construction.</li> <li>• Based on the foregoing understanding, try to identify potential reasons for quality performance possibly as a function of trade, product, and/or location.</li> </ul>			
<b>Data reports</b>	<b>Primary interactive operations</b>	<b>Insightful messages received and/or answers identified</b>	<b>Viewing time</b>
	<p><i>Figure 5.19(a)</i></p> <ul style="list-style-type: none"> <li>• <b>Visualization selection:</b> Select to generate the "counts of as-built record" visualization.</li> <li>• <b>Data selection:</b> Select as-built record datasets that are of the "deficiency list" type AND associated with the Painter trade.</li> <li>• <b>Visual encoding selection:</b> Specify that the product dimension be mapped to the X axis. Specify that values of the product dimension are sorted by ascending order of their corresponding deficiency counts.</li> <li>• <b>Data granularity selection:</b> Select that the product dimension is at the levels of granularity of "system", "subsystem", and "element".</li> <li>• <b>General interactivity:</b> Rotate the chart 90 degree clockwise.</li> </ul> <p><i>Figure 5.19(b)</i> Similar to the interaction features applied to generating Figure 5.19(a) except for: <b>Data selection:</b> Select as-built record datasets that are of the "deficiency list" type AND associated with the Cleaning trade.</p>	<p>partitioning/ceiling" system is not surprising because the major part of the Painter trade's work is "fill/sand/paint walls and ceilings". However, the pattern of clustering in the "interior finishing", which involves painting doors and baseboards, is somewhat concerning because the scope of work of painting with regard to the interior finishing is relatively smaller than the one related to interior partitioning/ceiling.</p> <p>2. When drilling down to see the number of deficiencies related to the subsystem of the "interior finishing" and "interior partitioning/ceiling" systems respectively, it is unexpected to see many painting deficiencies cluster in the "door" subsystem and the number of painting deficiencies for the "door" subsystem is even more the ones for the "interior wall" despite the fact that painting areas for doors are relatively much smaller than the ones for interior walls. Therefore in addition to the issue of general capability of the Painter trade, the reasons for the Painter's deficiencies is possibly also a function of the product dimension, i.e., it has particular problems in dealing with the door painting work.</p> <p><i>Figure 5.19(b)</i></p> <p>1. For the cleaning trade, the deficiencies cluster in the "interior finishing". This clustering pattern is not surprising because the major part of the cleaning trade's work is on tidying up and cleaning after other interior finishing. This indicates that cleaning trade deficiencies have no apparent pattern in terms of reasons for deficiencies as a function of the product dimension.</p>	<p><i>Figure 5.19(b)</i> <b>5 seconds</b></p>

FLATIRON CONSTRUCTION CORP		waterfront mid-rise building project		REPCON™	
		RECORDS SELECTION REPORT		Page 1 of 304	
File Used: PROJ03\WFMR				Report Date: 2011-02-09 Report Time: 3:30:13pm Revision Number: 0 Progress Date:	
RECORD CODE	DESCRIPTION	PROJECT PARTICIPANT CODE TRADE	ASSOCIATE PCBS PATH	DESCRIPTION	
Defi-101-1	Deficiency--Suite 101-1	T012 Painter	PR.LS01.L01.SL1 PR.PCL.07.17	Bathroom/Ensuite Interior Door Panel	
Defi-101-10	Deficiency--Suite 101-10	T020 Cleaning	PR.LS01.L01.SL1 PR.PCL.02.04	Bathroom/Ensuite Floor Transition	
Defi-101-11	Deficiency--Suite 101-11	T005 Electrical	PR.LS01.L01.SL1 PR.PC4.03.08	Bathroom/Ensuite Nu-heat	
Defi-101-12	Deficiency--Suite 101-12	T005 Electrical	PR.LS01.L01.SL1 PR.PC5.01.07	Bathroom/Ensuite Pot Light	
Defi-101-13	Deficiency--Suite 101-13	T007 Mirrors & Shower Enclosure	PR.LS01.L01.SL1 PR.PC4.01.07.05	Bathroom/Ensuite Shower Enclosure	
Defi-101-14	Deficiency--Suite 101-14	T011 Mechanical	PR.LS01.L01.SL1 PR.PC4.01.07.06	Bathroom/Ensuite Sink	
Defi-101-15	Deficiency--Suite 101-15	T015 Tiles/Stone countertops	PR.LS01.L01.SL1 PR.PCL.01.06	Bathroom/Ensuite Tile Base	
Defi-101-16	Deficiency--Suite 101-16	T015 Tiles/Stone countertops	PR.LS01.L01.SL1 PR.PCL.01.05	Bathroom/Ensuite Splash	
Defi-101-17	Deficiency--Suite 101-17	T011 Mechanical	PR.LS01.L01.SL1 PR.PC4.01.07.07	Bathroom/Ensuite Toilet	
Defi-101-18	Deficiency--Suite 101-18	T012 Painter	PR.LS01.L01.SL2 PR.PCL.02.01	Den/Bedroom Baseboard	
Defi-101-19	Deficiency--Suite 101-19	T005 Electrical	PR.LS01.L01.SL2 PR.PC5.01.02	Den/Bedroom Cover Plate	
Defi-101-2	Deficiency--Suite 101-2	T012 Painter	PR.LS01.L01.SL1 PR.PCL.07.17	Bathroom/Ensuite Interior Door Panel	
Defi-101-20	Deficiency--Suite 101-20	T011 Mechanical	PR.LS01.L01.SL2 PR.PC4.03.05	Den/Bedroom Fan	
Defi-101-21	Deficiency--Suite 101-21	T012 Painter	PR.LS01.L01.SL2 PR.PCL.07.16	Den/Bedroom Interior French Door Glass	
Defi-101-22	Deficiency--Suite 101-22	T012 Painter	PR.LS01.L01.SL2 PR.PCL.07.16	Den/Bedroom Interior French Door Glass	
Defi-101-23	Deficiency--Suite 101-23	T012 Painter	PR.LS01.L01.SL2 PR.PCL.07.15	Den/Bedroom Interior French Door	
Defi-101-24	Deficiency--Suite 101-24	T012 Painter	PR.LS01.L01.SL5 PR.PCL.02.01	Entry/Hallway Baseboard	
Defi-101-25	Deficiency--Suite 101-25	T003 Carpenter	PR.LS01.L01.SL5 PR.PCL.07.17	Entry/Hallway Interior Door Panel	
Defi-101-26	Deficiency--Suite 101-26	T020 Cleaning	PR.LS01.L01.SL5 PR.PC5.06	Entry/Hallway Multimedia Box	
Defi-101-27	Deficiency--Suite 101-27	T004 Drywall	PR.LS01.L01.SL5 PR.PC5.06	Entry/Hallway Multimedia Box	
Defi-101-28	Deficiency--Suite 101-28	T012 Painter	PR.LS01.L01.SL5 PR.PC5.06	Entry/Hallway Multimedia Box	
Defi-101-29	Deficiency--Suite 101-29	T010 Telephone	PR.LS01.L01.SL5 PR.PC5.06	Entry/Hallway Multimedia Box	
Defi-101-3	Deficiency--Suite 101-3	T011 Mechanical	PR.LS01.L01.SL1 PR.PC4.01.07.01	Bathroom/Ensuite Bathtub	

Figure 5.16 The first page of the 304 page tabular report of deficiency records that include information about project participants who are responsible for the deficiencies, deficient products, and locations of the products.

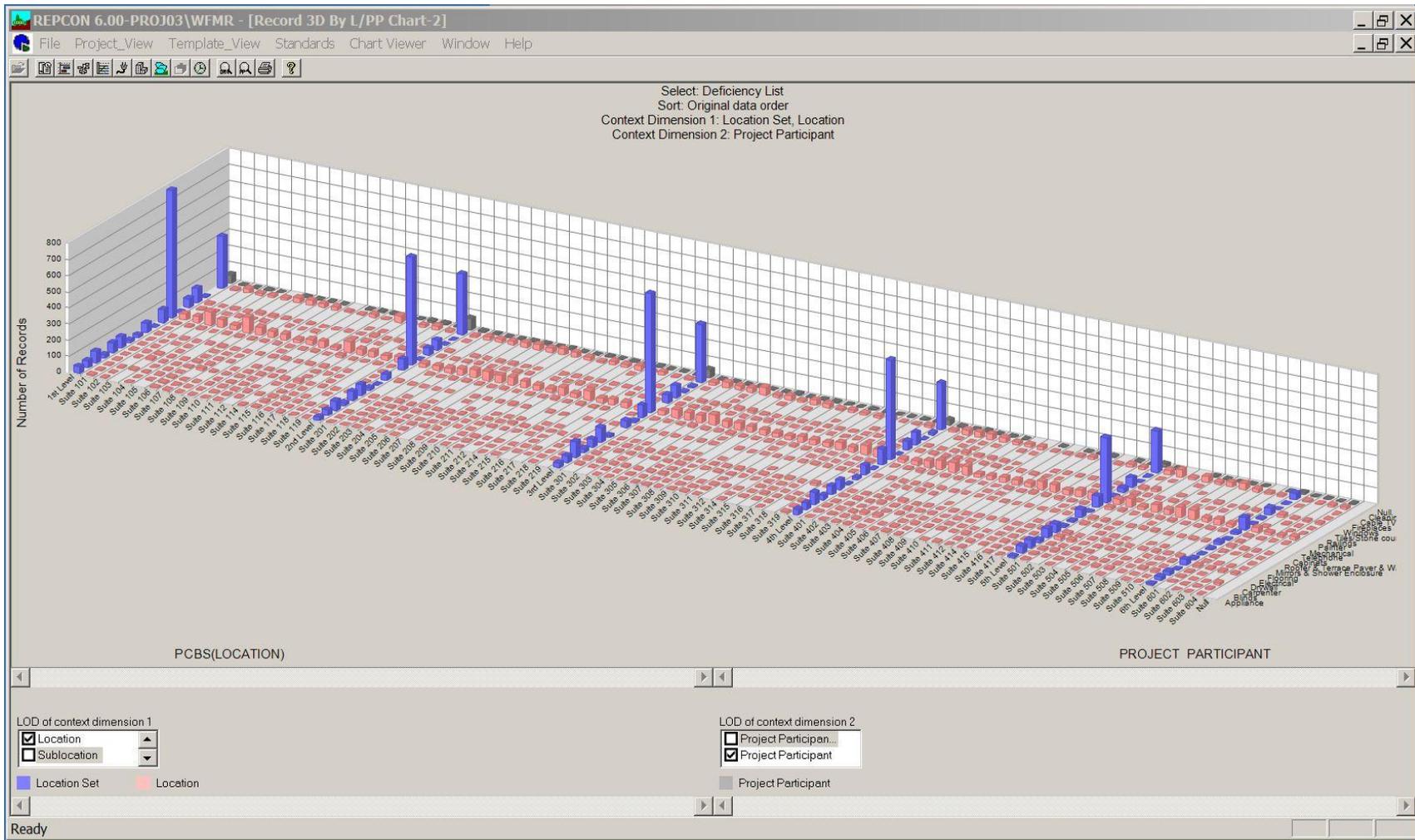


Figure 5.17 Number of deficiencies distributed in the location dimensions of two levels of location dimension granularity ("location set" and "location" levels) and the project participant dimension at the level of granularity of individual "project participant".

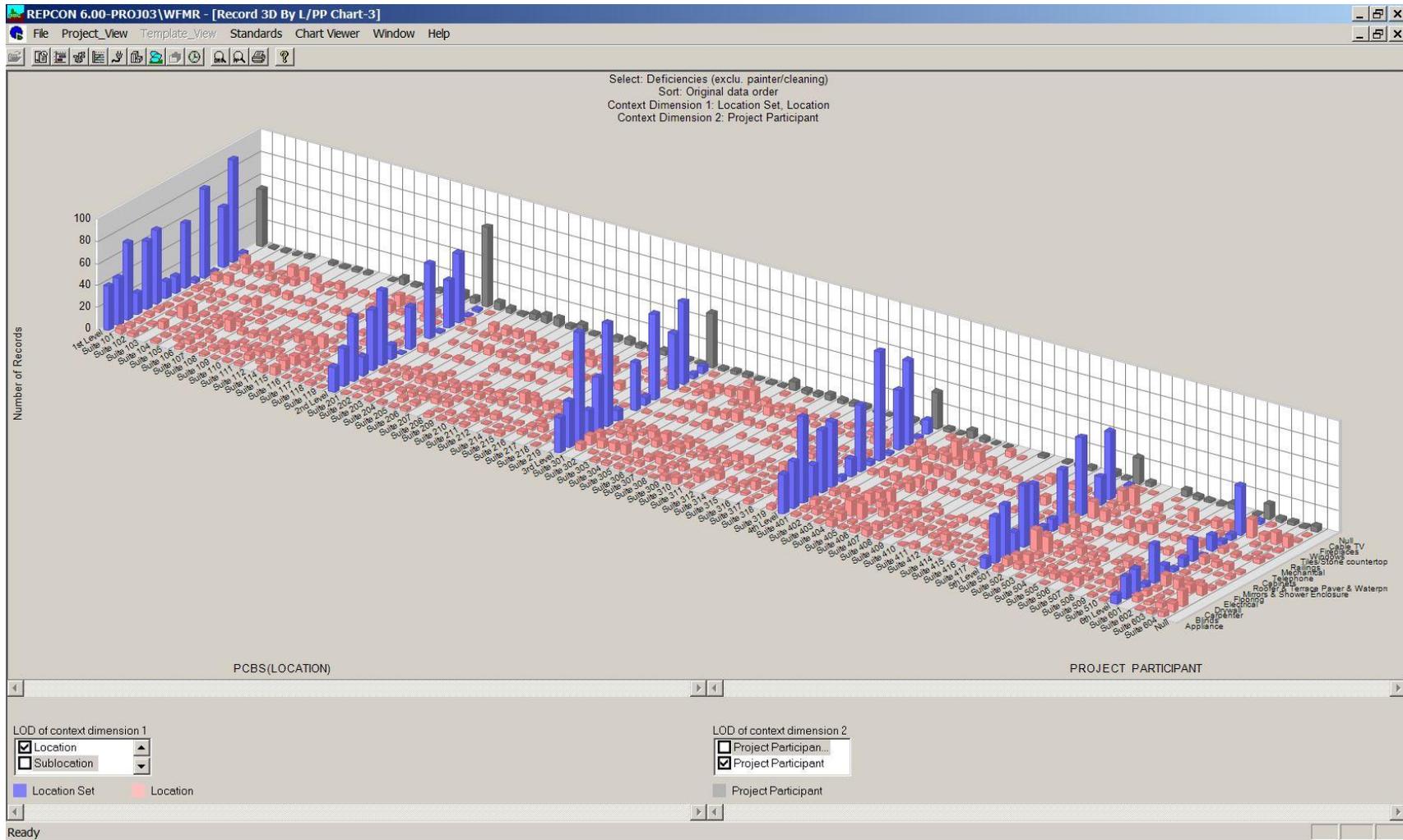


Figure 5.18 Number of deficiencies (excluding the ones of the Painter and Cleaning trades) distributed in two levels of location dimension granularity ("location set" and "location" levels) and the project participant dimension at the level of granularity of individual "project participant".

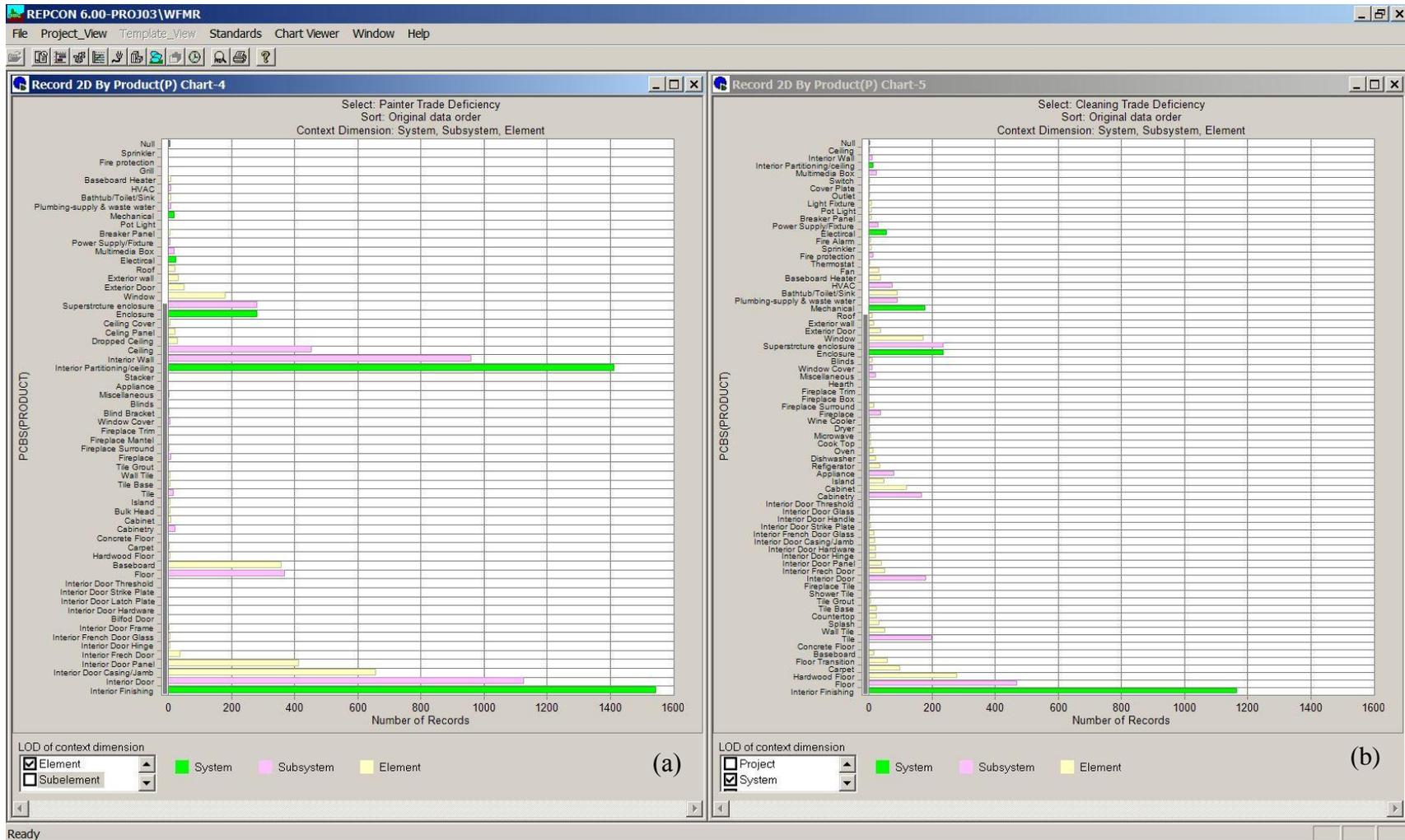


Figure 5.19 Number of (a) Painter trade deficiencies, and (b) Cleaning trade deficiencies distributed in the product dimensions of three different levels of granularity ("System", "Subsystem", and "Element" levels)

### 5.3.2.2 Project 2- demonstration case 2

#### 1. **CM analytics tasks:**

- To understand the values and patterns of behaviour of execution conditions in terms of number of deficient work items identified during interior construction.
- Based on the foregoing understanding, try to identify potential impacts on time performance.

#### 2. **Supporting CM functions/tasks:** time management/time performance control.

#### 3. **Project data used:**

- Type of data: records of deficiencies data.
- Approximate scope of data: approximately 8200 deficiency data entries.
- Original CM application of data: quality control application.

#### 4. **Current data reporting practice:** tabular report (Figure 5.20).

#### 5. **Visualization approach:** as-built record (deficiency list) distribution graphics (Figure 5.21 ).

#### 6. **Descriptions of the exploration-answer process using both a CM data visualization environment and current traditional tabular data reporting functionalities:** see Table 5.4.

Table 5.4 Descriptions of the exploration-answer process for both the use of a CM data visualization environment and the current data reporting functionalities: project 2-demonstration case 2

<b>CM analytics tasks (i.e. CM questions):</b>			
<ul style="list-style-type: none"> <li>• To understand the values and patterns of behaviour of execution conditions in terms of number of deficient work items identified during interior construction.</li> <li>• Based on the foregoing understanding, try to identify potential impacts on time performance.</li> </ul>			
<b>Data reports</b>	<b>Primary interactive operations</b>	<b>Insightful messages received and/or answers identified</b>	<b>Viewing time</b>
16 pages of the kind of tabular report seen in: <i>Figure 5.20</i>	<p><i>Figure 5.20</i></p> <ul style="list-style-type: none"> <li>• <b>Data report selection:</b> Select to generate the "as-built record" reports.</li> <li>• <b>Data selection:</b> Select as-built record datasets that are of the "deficiency list" type and have the record keyword of "long lead time item". Include data fields in terms of keywords (i.e. types of deficiencies), products, locations, and project participants that are associated with the deficiencies (it was intended to only include locations but the system must generate locations and products at the same time).</li> </ul>	<p><i>Figure 5.20</i></p> <ol style="list-style-type: none"> <li>1. It seems that most of the deficiencies that potentially require longer time to fix relate to the Window trade. This pattern indicates that the deficiencies have potential impacts on the project participant dimension (i.e. impacting the Window trade), and the "substantial completion date" attribute for the Window trade may be affected, which may in turn affect the substantial completion date for the whole project.</li> <li>2. Also it seems that many of the deficiencies requiring longer time to fix are due to the "damage/defect" type of quality issue.</li> </ol>	<p><i>Figure 5.20</i> <b>150 seconds</b></p>
Images generated from a CM data visualization environment: <i>Figure 5.21</i>	<p><i>Figure 5.21 (a)</i></p> <ul style="list-style-type: none"> <li>• <b>Visualization selection:</b> Select to generate the "counts of as-built record" visualization.</li> <li>• <b>Data selection:</b> Select as-built record datasets that are of the "deficiency list" type and have the record keyword of "long lead time item".</li> <li>• <b>Visual encoding selection:</b> Specify that the location dimension be mapped to the X axis and the project participant dimension be mapped to the Y axis.</li> <li>• <b>Data granularity selection:</b> Select that the location dimension is at the levels of granularity of "location set" and "location" while the project participant dimension is at the level of granularity "project participant".</li> </ul> <p><i>Figure 5.21 (b)</i></p> <ul style="list-style-type: none"> <li>• <b>Visualization selection:</b> Select to generate the "counts of as-built record" visualization.</li> <li>• <b>Data selection:</b> Select as-built record datasets that are of the "deficiency list" type and have the record keyword of</li> </ul>	<p><i>Figure 5.21 (a)</i></p> <ol style="list-style-type: none"> <li>1. In addition to the 1st message received in Figure 5.20, it is also observed that long lead time deficiencies cluster in the Appliance trade. This pattern indicates that the deficiencies have potential impacts on the project participant dimension (i.e. impacting the Appliance trade), and the substantial completion date attribute for the Appliance trade.</li> </ol> <p><i>Figure 5.21 (b)</i></p> <ol style="list-style-type: none"> <li>1. For the Window trade, most of their deficient work is due to the "damage/defect" type of quality issues.</li> <li>2. For the Appliance trade their deficient work comes from both "damage/defect" and "unsatisfactory functions" issues.</li> <li>3. Therefore, in order to prevent the foregoing two trades from repeatedly producing the same deficient work thereby requiring rework and consequently delaying</li> </ol>	<p><i>Figure 5.21(a)</i> <b>5 seconds</b></p> <p><i>Figure 5.21(b)</i> <b>5 seconds</b></p>

**CM analytics tasks (i.e. CM questions):**

- To understand the values and patterns of behaviour of execution conditions in terms of number of deficient work items identified during interior construction.
- Based on the foregoing understanding, try to identify potential impacts on time performance.

<b>Data reports</b>	<b>Primary interactive operations</b>	<b>Insightful messages received and/or answers identified</b>	<b>Viewing time</b>
	<p>"long lead time item".</p> <ul style="list-style-type: none"><li>• <b>Visual encoding selection:</b> Specify that the keyword (i.e. deficiency type) dimension to be mapped to the X axis and the project participant dimension to be mapped to the Y axis.</li><li>• <b>Data granularity selection:</b> Select that the keyword (i.e. deficiency type) dimension is at the level of granularity of "2nd level of deficiency definitions" (an example of 1st level of deficiency definition is "interior deficiency classification" and 2nd level of definition are the ones seen in the figure such as "unsatisfactory installation"). The project participant dimension is at the level of "project participant".</li></ul>	<p>substantial completion of the project, the project owner or general contractor can issue special notices to these two sub-trades in terms of their focus on fixing deficient work.</p>	

FLATIRON CONSTRUCTION CORP		Waterfront mid-rise building project		REPCON™
File Used: PROJ03\WFMR		RECORDS SELECTION REPORT		Page 1 of 16
RECORD CODE	DESCRIPTION	PROJECT PARTICIPANT CODE TRADE	ASSOCIATE PCBS PATH DESCRIPTION	ASSOCIATED KEYWORDS
Defi-101-42	Deficiency--Suite 101-42	T001 Appliance	PR.LS01.L01.SL4 Kitchen PR.PC1.05.04 Refrigerator	Damages/Defects Long ordering time
Defi-101-44	Deficiency--Suite 101-44	T015 Tiles/Stone countertops	PR.LS01.L01.SL4 Kitchen PR.PC1.01.05 Splash	Damages/Defects Long ordering time
Defi-101-65	Deficiency--Suite 101-65	T017 Windows	PR.LS01.L01.SL3 Living/Dining room PR.PC2.02.02 Window Frame	Damages/Defects Long ordering time
Defi-101-80	Deficiency--Suite 101-80	T017 Windows	PR.LS01.L01.SL2 Den/Bedroom PR.PC2.02.02 Window	Damages/Defects Long ordering time
Defi-101-81	Deficiency--Suite 101-81	T017 windows	PR.LS01.L01.SL2 Den/Bedroom PR.PC2.02.02 Window	Damages/Defects Long ordering time
Defi-102-48	Deficiency--Suite 102-48	T001 Appliance	PR.LS01.L02.SL4 Kitchen PR.PC1.05.01 Cook Top	Damages/Defects Long ordering time
Defi-102-60	Deficiency--Suite 102-60	T017 Windows	PR.LS01.L02.SL3 Living/Dining room PR.PC2.02.03.02 Exterior Door Glass	Damages/Defects Long ordering time
Defi-102-98	Deficiency--Suite 102-98	T017 Windows	PR.LS01.L02.SL2 Den/Bedroom PR.PC2.02.02 Window	Damages/Defects Long ordering time
Defi-103-91	Deficiency--Suite 103-91	T017 windows	PR.LS01.L03.SL3 Living/Dining room PR.PC2.02.03.02 Exterior Door Glass	Damages/Defects Long ordering time
Defi-103-98	Deficiency--Suite 103-98	T017 windows	PR.LS01.L03.SL3 Living/Dining room PR.PC2.02.02 Window	Damages/Defects Long ordering time
Defi-103-99	Deficiency--Suite 103-99	T017 windows	PR.LS01.L03.SL3 Living/Dining room PR.PC2.02.02 Window	Damages/Defects Long ordering time
Defi-104-43	Deficiency--Suite 104-43	T001 Appliance	PR.LS01.L04.SL4 Kitchen PR.PC1.05.01 cook Top	Damages/Defects Long ordering time
Defi-104-45	Deficiency--Suite 104-45	T001 Appliance	PR.LS01.L04.SL4 kitchen PR.PC1.05.02 Dishwasher	Unsatisfactory Installation Long ordering time
Defi-104-70	Deficiency--Suite 104-70	T017 Windows	PR.LS01.L04.SL3 Living/Dining room PR.PC2.02.02 Window	Damages/Defects Long ordering time
Defi-105-11	Deficiency--Suite 105-11	T017 windows	PR.LS01.L05.SL2 Den/Bedroom PR.PC2.02.02 Window	Damages/Defects Long ordering time
Defi-105-31	Deficiency--Suite 105-31	T015 Tiles/Stone countertops	PR.LS01.L05.SL4 Kitchen PR.PC1.01.02 Countertop	Damages/Defects Long ordering time
Defi-105-49	Deficiency--Suite 105-49	T017 Windows	PR.LS01.L05.SL3 Living/Dining room PR.PC2.02.02 window	Damages/Defects Long ordering time

Figure 5.20 The first page of a 16 page tabular report of deficiencies that require a longer time to fix. It includes information about project participants who are responsible for the deficiencies, deficient products, locations of the products, and types of deficient work.

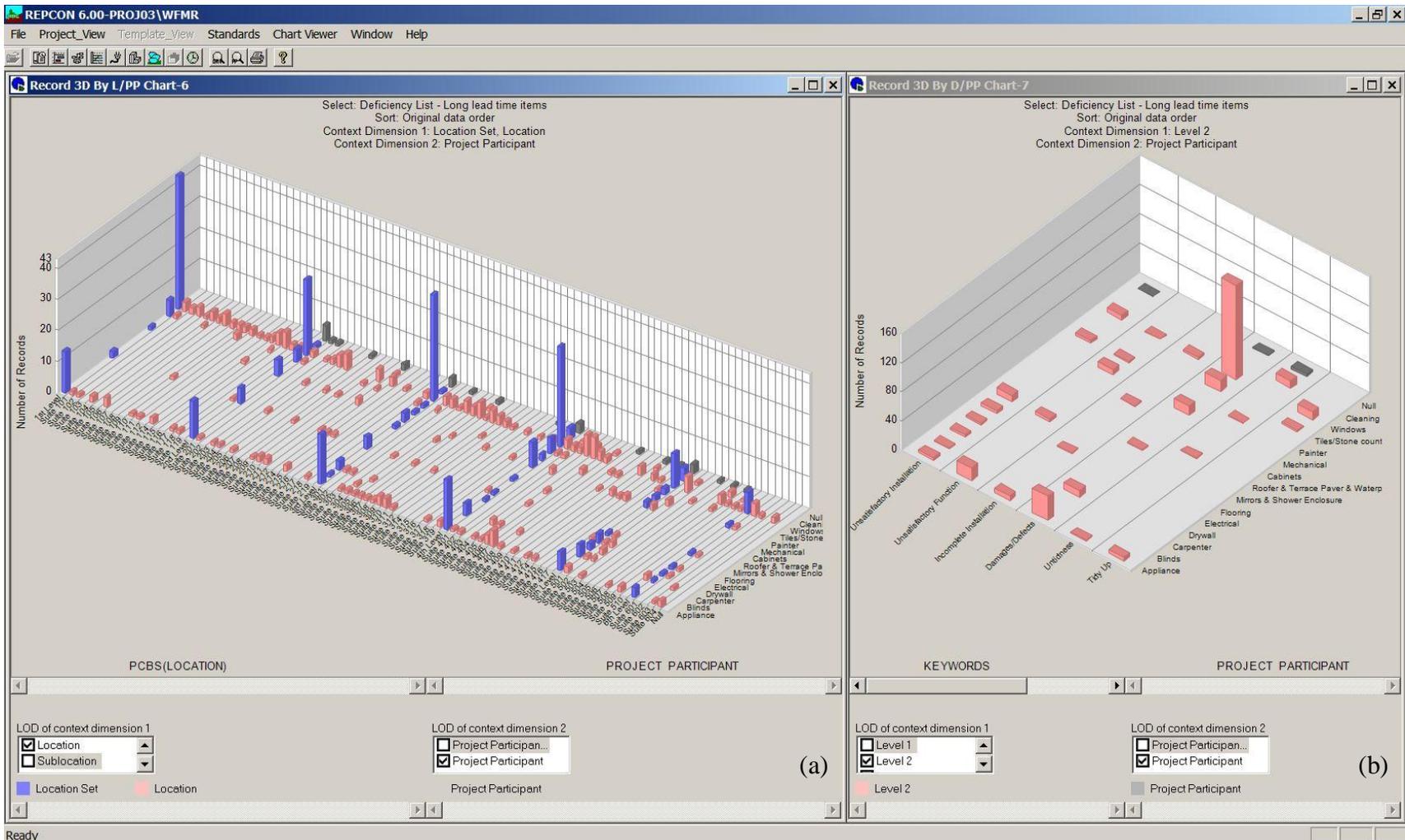


Figure 5.21 Number of long lead time deficiencies (i.e. deficiencies that need a longer time to correct) distributed in: (a) the location dimensions at two levels of granularity ("location set" and "location" levels) and the project participant dimension at the level of granularity of "project participant", (b) the keyword (i.e. deficiency type) dimension at the level of granularity of "2nd level of deficiency definitions" and the the project participant dimension at the level of granularity of "project participant".

### 5.3.3 Demonstration cases- project 3

Project 3 corresponds to a complex building rehabilitation project. The process view of the project is not modeled because this project was used mainly for collecting and visualizing change management related data. Hence, the focus of project modeling and project data collection is more on products and trades who are responsible for constructing the products. The PCBS view of this project consists of a rather detailed list of project components and all work locations in the planned location sequence (except for the sub locations), as shown in Figure 5.22 The organizational view of this project, i.e., project participants of the project, is depicted in Figure 5.23. The as-built view in terms of change order records and how it can be associated with the foregoing views can be found in Figure 5.24.

The sheer volume of the extra work orders generated during the first 2/3 of the project duration and their occurrence frequency made change order management on this project a challenging task (the construction manager who provided the data used the words extras, extra work order and change order as synonyms). In the original change order spreadsheet organized by a former master's student (Korde 2005), 448 change orders were recorded. Information recorded with each change order included change order initiating sources (e.g. request for information, site instruction), description, issue date (from general contractor to owner), general contractor proposed cost, approval date, approved cost, proposed cost breakdown for sub-trades, and type of change order (e.g. change due to design omissions/errors, change due to owner's changing product requirements). Each item in the original change order spreadsheet does not have an ideal one to one information relationship such that a change order relates to a certain product, a certain location, a certain trade, and cost. Therefore, each original change order was broken down to several "sub-change orders" so that a one to one information relationship could be created. As a result, the original 448 change order items became approximately 850 change order data entries. However in practice and also during the foregoing "data disaggregation" process, it proved difficult to break the proposed or approved cost down to such a level of detail. Therefore, for demonstration purposes only, change order information excluding cost information are parsed into structured data and entered into REPCON.

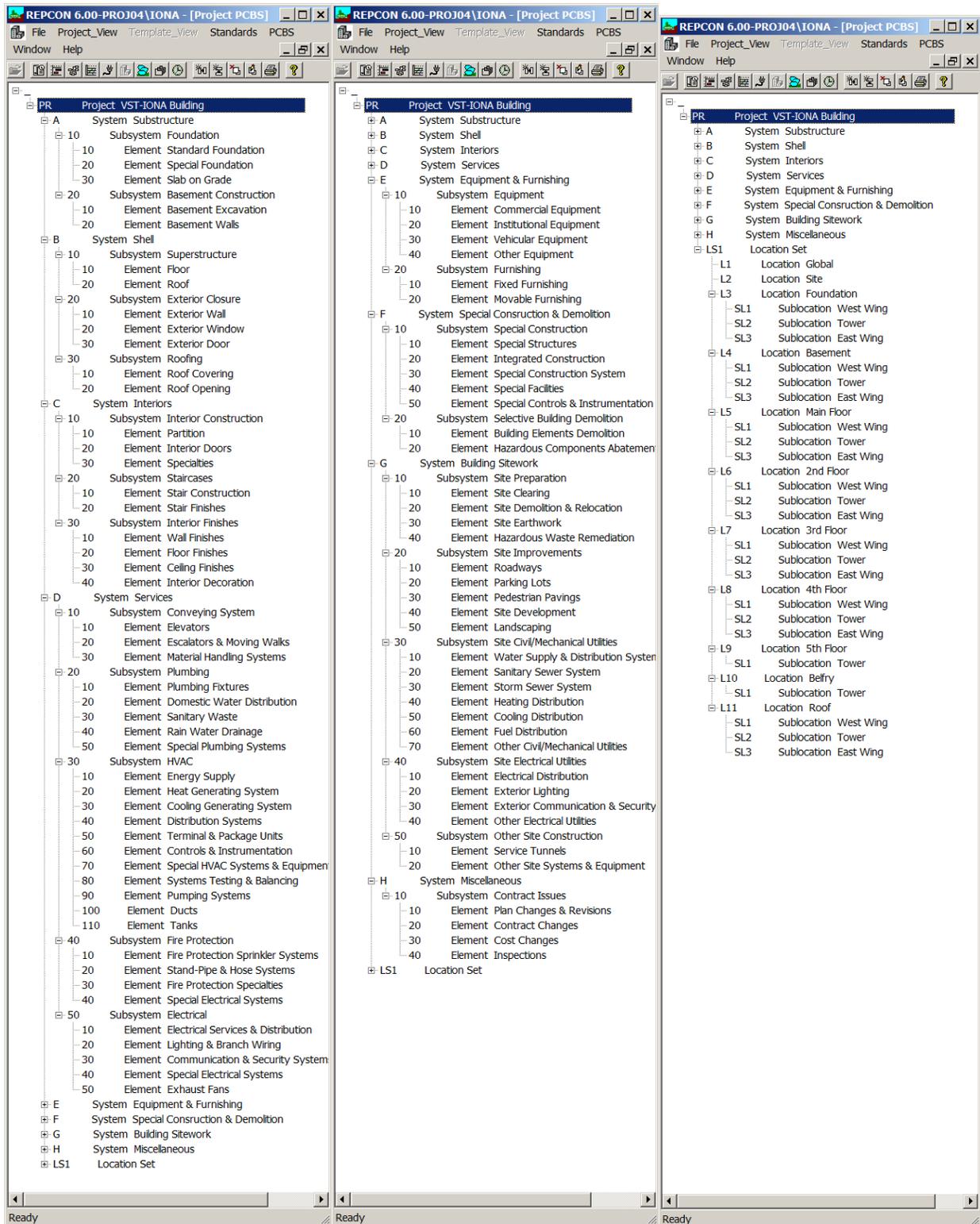


Figure 5.22 The PCBS view (both product and location) of project 3

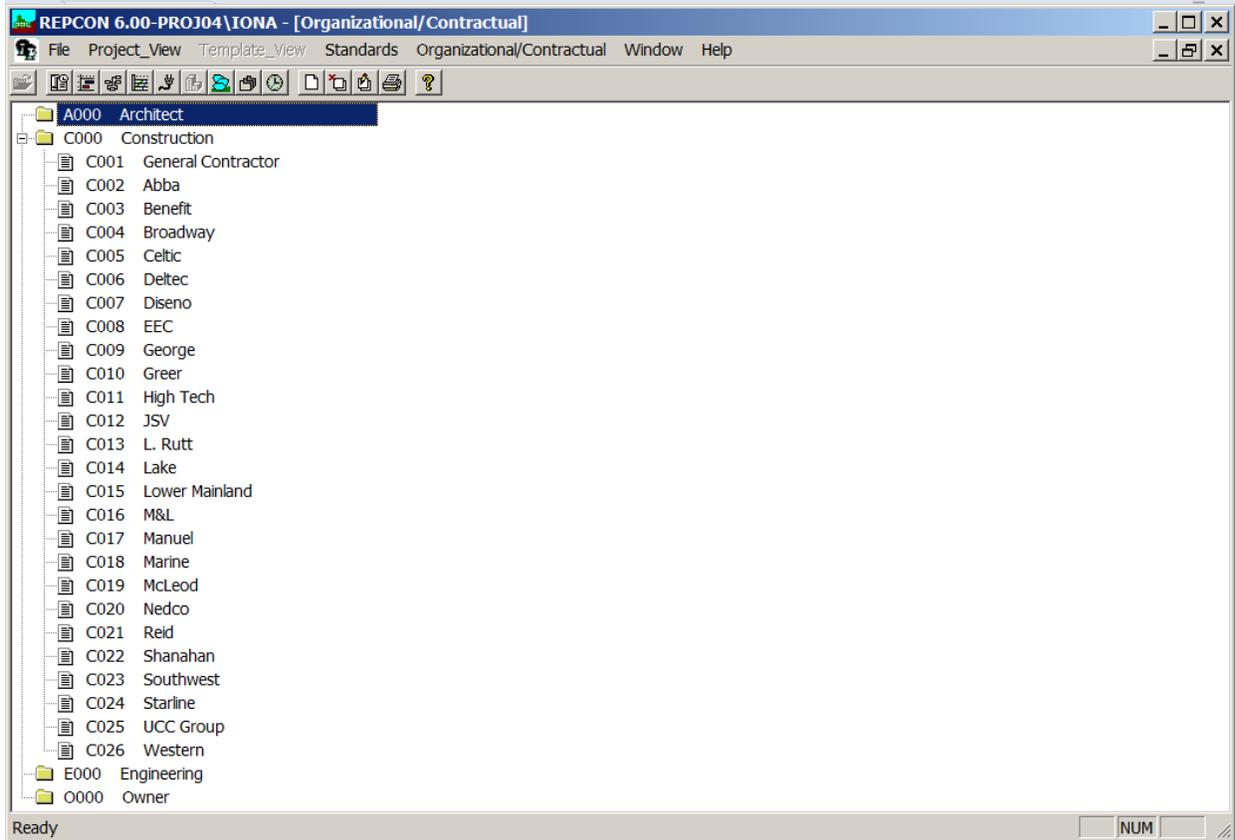


Figure 5.23 The organizational view (project participant) of project 3

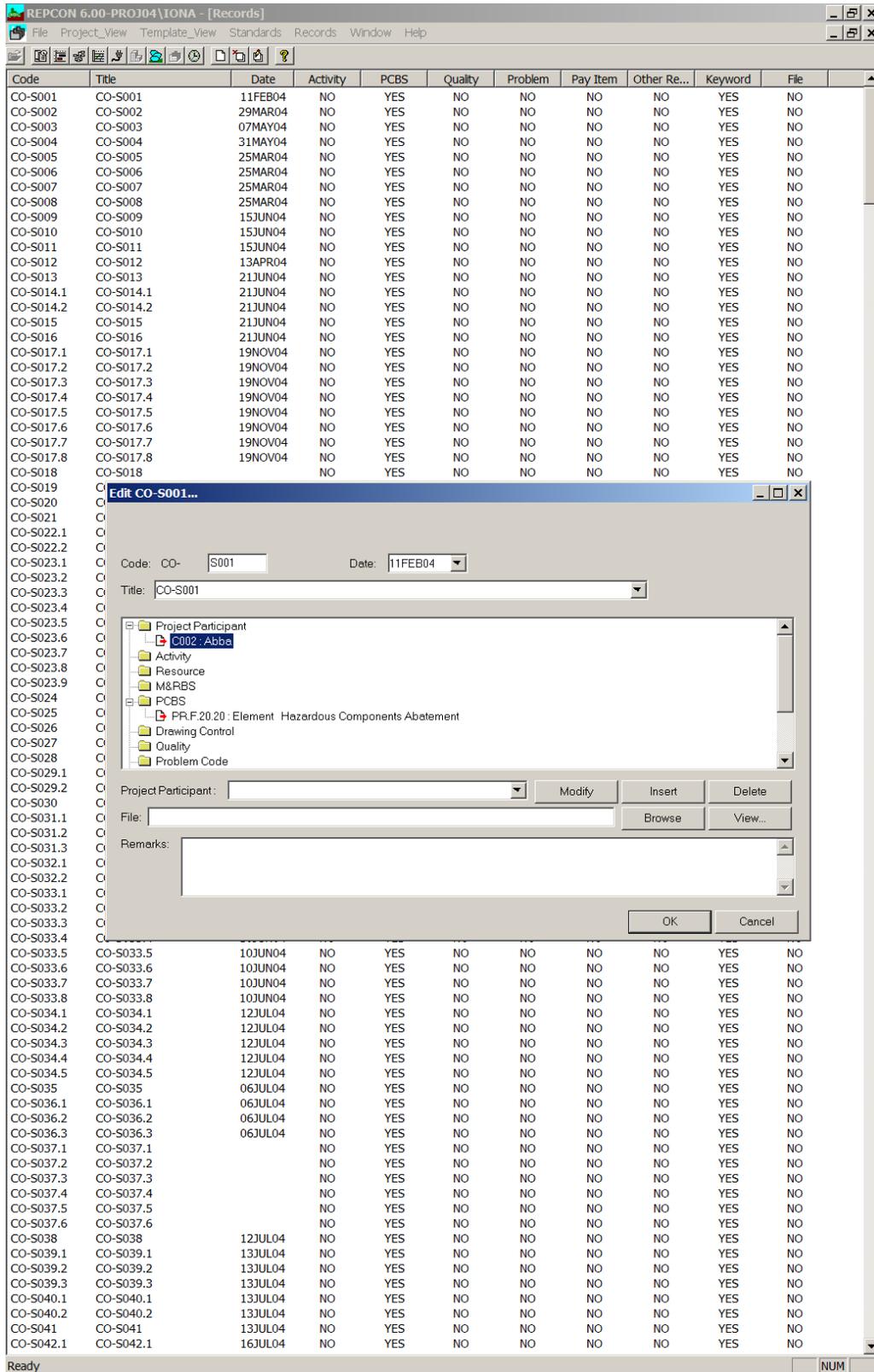


Figure 5.24 As-built view (change order records) for project 3. The dialogue box shows how users can associate a change order item with items from other views such as the project participant view and PCBS view

### 5.3.3.1 Project 3- demonstration case 1

#### 1. CM analytics tasks:

- To understand the values and patterns of behaviour of execution conditions in terms of number of change orders generated during the first 2/3 of the construction phase duration.
- Based on the foregoing understanding, try to identify potential causes of or impacts by these conditions possibly as a function of the change order type, product, location, and/or trade dimensions.

#### 2. Supporting CM functions/tasks: change management/change order control.

#### 3. Project data used:

- Type of data: records of change order data.
- Approximate scope of data: 850 change order entries.
- Original CM application of data: change management/change orders control and preparation.

#### 4. Current data reporting practice: tabular reports (Figure 5.25).

#### 5. Visualization approach: as-built record (change orders) distribution graphics (Figures 5.26 ~ 5.28).

#### 6. Descriptions of the exploration-answer process using both a CM data visualization environment and current traditional tabular data reporting functionalities: see Table 5.5.

Table 5.5 Descriptions of the exploration-answer process for both the use of a CM data visualization environment and current data reporting functionalities: project 3-demonstration case 1

<b>CM analytics tasks (i.e. CM questions):</b>			
<ul style="list-style-type: none"> <li>• To understand the values and patterns of behaviour of execution conditions in terms of number of change orders generated during the first 2/3 of the construction phase duration.</li> <li>• Based on the foregoing understanding, try to identify potential causes of or impacts by these conditions possibly as a function of the change order type, product, location, and/or trade dimensions.</li> </ul>			
<b>Data reports</b>	<b>Primary interactive operations</b>	<b>Insightful messages received and/or answers identified</b>	<b>Viewing time</b>
43 pages of the kind of tabular report seen in <i>Figure 5.25</i>	<ul style="list-style-type: none"> <li>• <b>Data report selection:</b> Select to generate "as-built record" reports.</li> <li>• <b>Data selection:</b> Select as-built record datasets that are of the "change order" type. Include data fields in terms of products, locations, project participants, and record keywords that are associated with the change orders.</li> </ul>	<p><i>Figure 5.25</i> It seems that most of the change orders relate to "design change".</p>	<p><i>Figure 5.25</i> <b>250 seconds</b></p>
Three images generated from a CM data visualization environment: <i>Figure 5.26</i> <i>Figure 5.27</i> <i>Figure 5.28</i>	<p><i>Figure 5.26</i></p> <ul style="list-style-type: none"> <li>• <b>Visualization selection:</b> Select to generate the "counts of as-built record" visualization</li> <li>• <b>Data selection:</b> Select as-built record datasets that are of the "change order type" and have the record keyword of "scope change".</li> <li>• <b>Visual encoding selection:</b> Specify that the product dimension be mapped to the X axis and the location dimension be mapped to the Y axis.</li> <li>• <b>Data granularity selection:</b> Select the location dimension at the level of granularity of "location set" while the product dimension is at the level of "system".</li> </ul> <p>Repeat the foregoing steps three more times except for as-built record datasets that are of the "change order type" and have the record keyword of "site condition", "design change", and "owner change" respectively.</p> <p><i>Figure 5.27(a)</i></p> <ul style="list-style-type: none"> <li>• <b>Visualization selection:</b> Select to generate "counts of as-built record" visualization.</li> <li>• <b>Data selection:</b> Select as-built record datasets</li> </ul>	<p><i>Figure 5.26(a)~(d)</i></p> <ol style="list-style-type: none"> <li>1. The change orders identified relate to the locations of: <i>global, site, foundation, basement, main floor, 1st floor, 2nd floor, 3rd floor, 4th floor, 5th floor, belfry, and roof</i>. They also relate to the products of: <i>substructure system, shell system, interiors system, services system, equipment &amp; furnishing system, special construction &amp; demolition, and building site work</i>. The change orders identified are defined into four types: <i>design change, scope change, owner change, site condition change</i>.</li> <li>2. It is apparent that most of the change orders come from the "design change", and relatively higher numbers of design change orders cluster in the "interior system" and "service system". It is also noted that the relatively larger numbers of "scope change" change orders cluster in the product of "building site work" and the location of "main floor".</li> </ol> <p><i>Figure 5.27(a)</i></p> <ol style="list-style-type: none"> <li>1. For the "scope change" type of change order associated with the "building site work", further drilling down to see their associations with products of finer levels of granularity leads to the observation that these change orders are clustered in the "electrical utilities subsystem" and "exterior lighting element" and at the "main floor" location. This pattern implies that reasons for the large number of scope change orders of building site work are a function of the product and location</li> </ol>	<p><i>Figure 5.26(a)~(d)</i> <b>5 seconds</b></p> <p><i>Figure 5.27(a)</i> <b>5 seconds</b></p>

<b>CM analytics tasks (i.e. CM questions):</b>			
<ul style="list-style-type: none"> <li>• To understand the values and patterns of behaviour of execution conditions in terms of number of change orders generated during the first 2/3 of the construction phase duration.</li> <li>• Based on the foregoing understanding, try to identify potential causes of or impacts by these conditions possibly as a function of the change order type, product, location, and/or trade dimensions.</li> </ul>			
<b>Data reports</b>	<b>Primary interactive operations</b>	<b>Insightful messages received and/or answers identified</b>	<b>Viewing time</b>
	<p>that are of the "change order" type, have the record keyword of "scope change", and are associated with the " building site work" including its "descendent" components.</p> <ul style="list-style-type: none"> <li>• <b>Visual encoding selection:</b> Specify that the location dimension be mapped to the Y axis and the product dimension be mapped to the X axis.</li> <li>• <b>Data granularity selection:</b> Select the location dimension at the level of granularity of "location set" while the product dimension is at the levels of "system", "subsystem", and "element".</li> </ul> <p><i>Figure 5.27(b)</i> Similar to the interaction features applied to generating Figure 5.27(a) except for as-built record datasets that are of the "change order" type, have the record keyword of "design change", and are associated with the "interiors system" and "services system" including their "descendent" components.</p> <p><i>Figure 5.28</i></p> <ul style="list-style-type: none"> <li>• <b>Visualization selection:</b> Select to generate the "counts of as-built record" visualization.</li> <li>• <b>Data selection:</b> Select as-built record datasets that are of the "change order" type.</li> <li>• <b>Visual encoding selection:</b> Specify that the project participant dimension be mapped to the X axis.</li> <li>• <b>Data granularity selection:</b> Select that the project participant dimension is at the level of granularity of "project participant".</li> </ul>	<p>dimensions, and they mainly arise from the exterior lighting required at the main floor.</p> <p><i>Figure 5.27(b)</i></p> <ol style="list-style-type: none"> <li>1. For the "design change" type of change order associated with the "interiors system" and "services system", further drilling down to see their associations with products at finer levels of granularity leads to the observation that these change orders do not exhibit apparent clustering patterns, although in the location range of basement to 4th floor a slightly higher number of change orders are encountered. This pattern combined with the one identified in Figure 5.26 imply that reasons for large numbers of design change orders mostly come from the "interior system" and "service system" in general, which were designed by the architect and mechanical engineer respectively. Therefore, their deficient design work may be the source of design changes. It can be further interpreted as design changes are a function of the product as well as project participant dimensions.</li> </ol> <p><i>Figure 5.28</i></p> <ol style="list-style-type: none"> <li>1. The change orders identified relate to 26 sub-trades and the general contractor. The observation that the general contractor does not have any change orders but a high number of change orders are not associated with any responsibility code (see the right most bar in the figure) shows a possible deficiency in data collection. It is suspected that the lack of referring to general contractor in the data is because the change orders are prepared by the general contractor soliciting change proposals from its sub-trades. Therefore the change order data are recorded mainly according to information provided by sub-trades (and owners regarding the change orders approval).</li> <li>2. It is also noted that high numbers of change orders are clustered in six sub-trades, an indication that the current change order situation may impact the project participant dimension, i.e., the six sub-trades.</li> </ol>	<p><i>Figure 5.27(b)</i> <b>10 seconds</b></p> <p><i>Figure 5.28</i> <b>5 seconds</b></p>

VST-IONA Building  
**RECORDS SELECTION REPORT**

File Used: PROJ04\IONA

Report Date: 2011-03-03  
 Report Time: 10:27:06am  
 Revision Number: 0  
 Progress Date:

RECORD CODE	DESCRIPTION	PROJECT PARTICIPANT CODE TRADE	ASSOCIATE PATH	PCBS DESCRIPTION	ASSOCIATED KEYWORDS
CO-S001	CO-S001	C002 Abba	PR.F.20.20	Hazardous Components Abatement	Site condition
CO-S002	CO-S002	C019 McLeod	PR.F.20.10 PR.LS1.L10.SL1	Building Elements Demolition Tower	Site condition
CO-S003	CO-S003	C023 Southwest	PR.F.20.10 PR.LS1.L1	Building Elements Demolition Global	Site condition
CO-S004	CO-S004	C002 Abba	PR.F.20.10 PR.LS1.L7.SL3	Building Elements Demolition East Wing	Owner Change
CO-S005	CO-S005	C002 Abba	PR.F.20.10	Building Elements Demolition	Site condition
CO-S006	CO-S006	C002 Abba	PR.F.20.10	Building Elements Demolition	Site condition
CO-S007	CO-S007	C002 Abba	PR.F.20.10	Building Elements Demolition	Design Change
CO-S008	CO-S008	C002 Abba	PR.F.20.10	Building Elements Demolition	Site condition
CO-S009	CO-S009	C002 Abba	PR.F.20.10	Building Elements Demolition	Site condition
CO-S010	CO-S010	C002 Abba	PR.F.20.10	Building Elements Demolition	Site condition
CO-S011	CO-S011	C002 Abba	PR.B.10.10	Floor	Site condition
CO-S012	CO-S012	C019 McLeod	PR.F.20.10 PR.LS1.L10.SL1	Building Elements Demolition Tower	Owner Change
CO-S013	CO-S013	C019 McLeod	PR.F.20.10	Building Elements Demolition	Site condition
CO-S014.1	CO-S014.1	C009 George	PR.H.10.10 PR.LS1.L1	Plan Changes & Revisions Global	Site condition
CO-S014.2	CO-S014.2	C009 George	PR.H.10.10 PR.LS1.L11.SL1	Plan Changes & Revisions West Wing	Site condition
CO-S015	CO-S015	C009 George	PR.LS1.L6.SL1	West Wing	Others
CO-S016	CO-S016	C009 George	PR.H.10.10	Plan Changes & Revisions	Site condition
CO-S017.1	CO-S017.1	C015 Lower Mainland	PR.B.10.20 PR.LS1.L11.SL1	Roof West Wing	Site condition
CO-S017.2	CO-S017.2	C018 Marine	PR.B.10.20 PR.LS1.L11.SL1	Roof West Wing	Site condition
CO-S017.3	CO-S017.3	C022 Shanahan	PR.B.10.20 PR.LS1.L11.SL1	Roof West Wing	Site condition
CO-S017.4	CO-S017.4	C026 Western	PR.B.10.20	Roof	Site condition

Figure 5.25 The first page of the 43 page tabular report of change orders that include information about project participants who will execute them, products and locations involved, and types of change order.

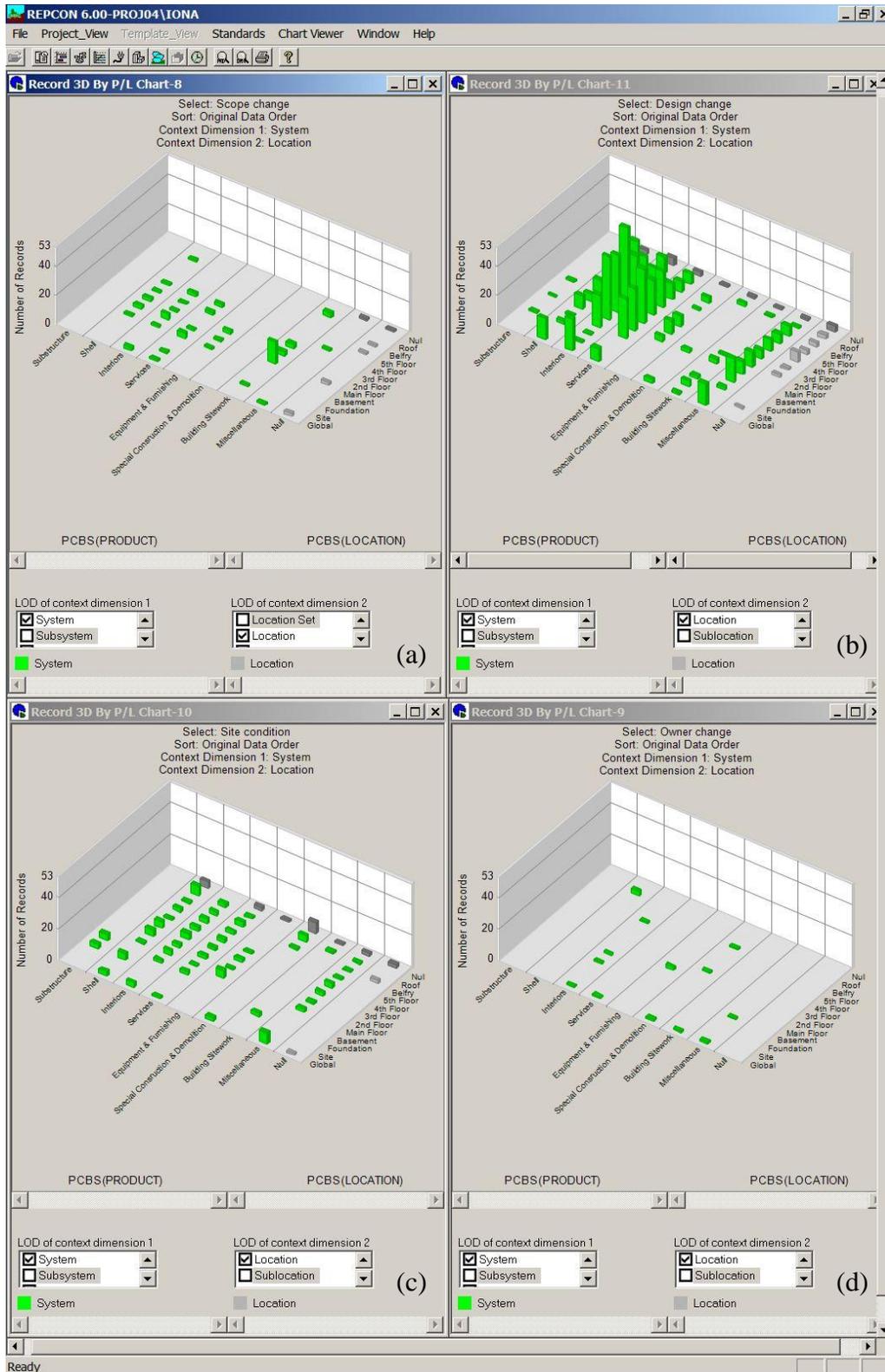


Figure 5.26 Number of (a) scope change orders, (b) design change orders, (c) site condition change orders, (d) owner change orders, distributed in the product dimension at the level of granularity of "system" and the location dimension at the level of granularity of "location".

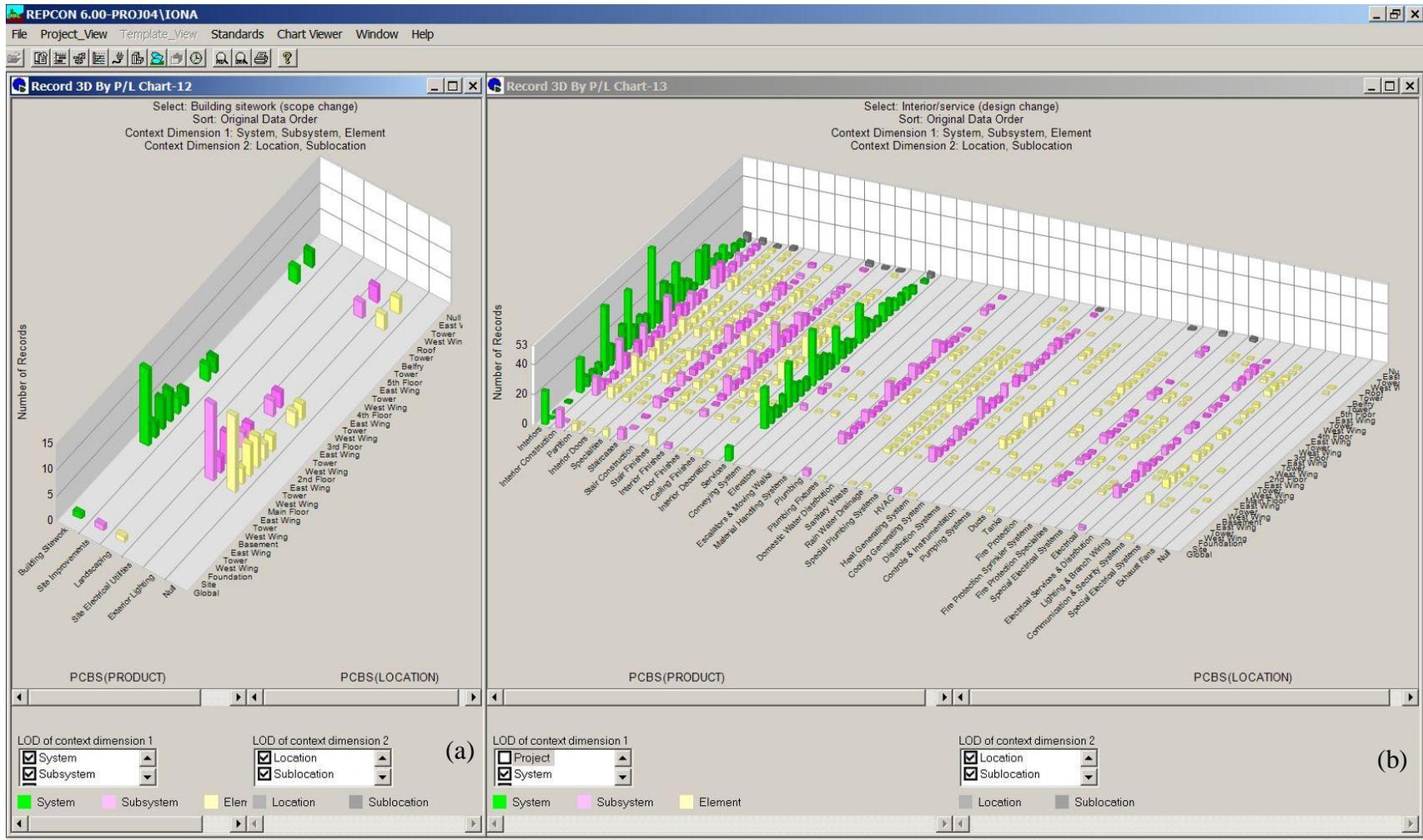


Figure 5.27 Number of (a) scope change orders related to the building site work , (b) design change orders related to the interior and service systems, distributed in the product dimensions of three different levels of granularity ("System", "Subsystem", and "Element" levels) and the location dimensions of two levels of granularity ("location" and "sub-location" levels).

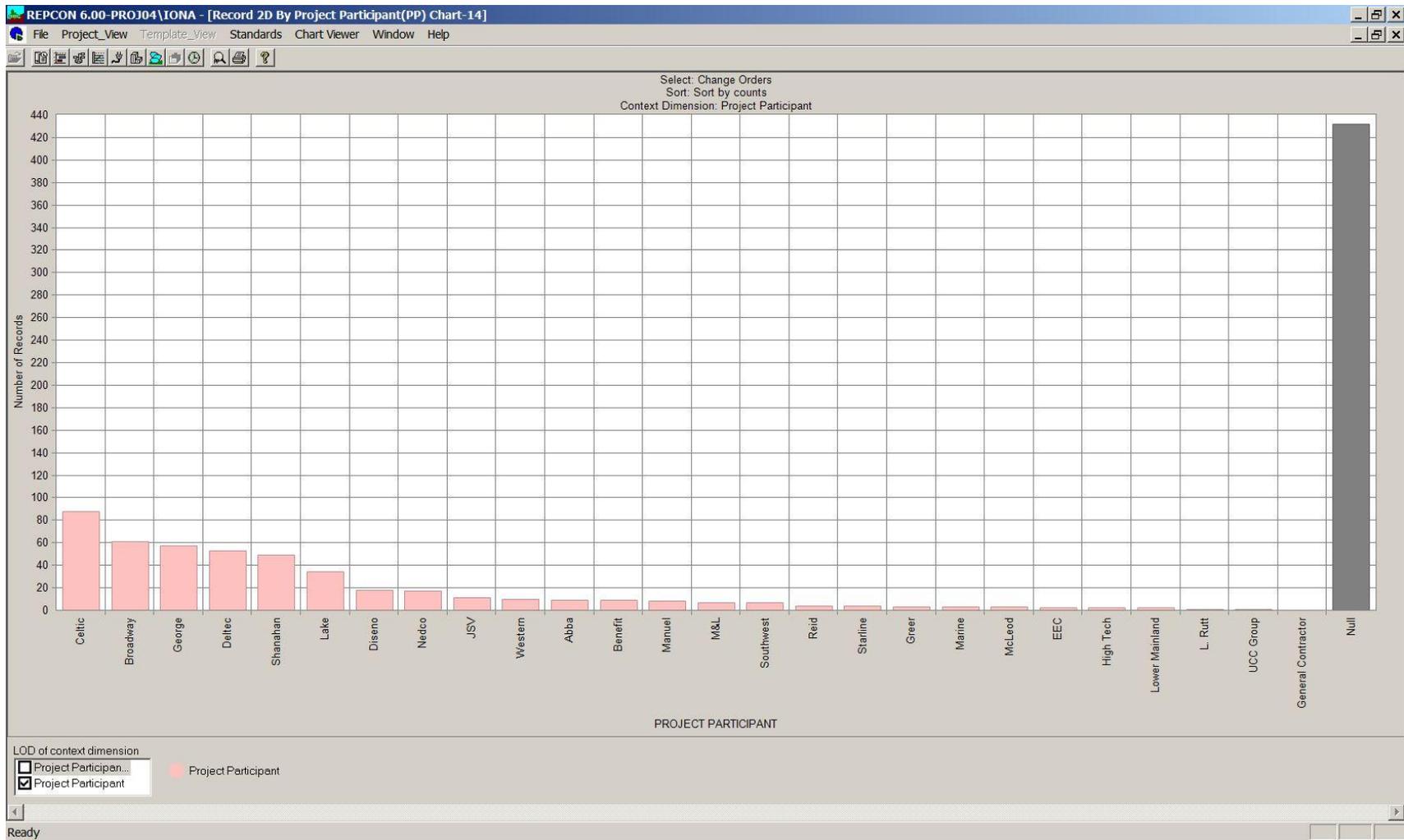


Figure 5.28 Number of change orders distributed in the project participant dimension at the level of granularity of "project participant".

### 5.3.3.2 Project 3- demonstration case 2

#### 1. **CM analytics tasks:**

- To understand the values and patterns of behaviour of execution condition in terms of numbers of change orders generated during the first 2/3 of the construction phase duration.
- Based on the foregoing understanding, try to identify potential impacts on time performance.

#### 2. **Supporting CM functions/tasks:** time management/time performance control.

#### 3. **Project data used:**

- Type of data: records of change order data.
- Approximate scope of data: 850 change order entries.
- Original CM application of data: change order applications/ change order control and preparation

#### 4. **Current data reporting practice:** tabular report (Figure 5.29).

#### 5. **Visualization approach:** as-built record (deficiency list) distribution graphics (Figure 5.30).

#### 6. **Descriptions of the exploration-answer process using both a CM data visualization environment and current traditional tabular data reporting functionalities:** see Table 5.6.



<b>CM analytics tasks (i.e. CM questions):</b>			
<ul style="list-style-type: none"> <li>• To understand the values and patterns of behaviour of execution condition in terms of numbers of change orders generated during the first 2/3 of the construction phase duration.</li> <li>• Based on the foregoing understanding, try to identify potential impacts on time performance.</li> </ul>			
<b>Data reports</b>	<b>Primary interactive operations</b>	<b>Insightful messages received and/or answers identified</b>	<b>Viewing time</b>
	<p><i>Figures 5.30(b) and 5.30(c)</i></p> <p>Similar to the interaction features applied to generating Figure 5.30(a) except select as-built record datasets that are of the "change order" type, record dates between February 15th 2005 and June 30th 2005, and are associated with the Broadway trade and the Celtic trade respectively.</p>	<p>there is not a high level of concern requiring focused management actions for the Broadway trade.</p> <p>2. However, in Figure 5.30 (c) it is noted that for change orders that were generated at the same time period but are associated with the Celtic trade, high numbers of change orders cluster in February ~April 2005 and in the 2nd floor to the 4th floor. Therefore, the Celtic trade's change orders may have impact on the time and location dimensions and their associated activity dimension. That is, time performance of the Celtic trade's activities scheduled between February and April 2005 and on the 2nd to the 4th floor may be impacted more significantly.</p>	

FLATIRON CONSTRUCTION CORP		VST-IONA Building		REPCON™	
		RECORDS SELECTION REPORT		Page 1 of 24	
File used: PR0704100A				Report Date: 2011-02-11 Report Time: 11:43:03am Revision Number: 0 Progress Date:	
RECORD CODE	DESCRIPTION	DATE	PROJECT PARTICIPANT CODE TRADE	ASSOCIATE PCSB PATH	DESCRIPTION
CO-S001	CO-S001	11FEB04	C002 Trade Abba	PR.F.20.20	Hazardous Components Abatement
CO-S002	CO-S002	29MAR04	C019 Trade McLeod	PR.F.20.10 PR.LSI.L10.SL1 Tower	Building Elements Demolition
CO-S003	CO-S003	07MAY04	C023 Trade Southwest	PR.F.20.10 PR.LSI.L1	Building Elements Demolition Global
CO-S004	CO-S004	31MAY04	C002 Trade Abba	PR.F.20.20 PR.LSI.L7.SL3	Building Elements Demolition East Wing
CO-S005	CO-S005	25MAR04	C002 Trade Abba	PR.F.20.10	Building Elements Demolition
CO-S006	CO-S006	25MAR04	C002 Trade Abba	PR.F.20.10	Building Elements Demolition
CO-S007	CO-S007	25MAR04	C002 Trade Abba	PR.F.20.10	Building Elements Demolition
CO-S008	CO-S008	25MAR04	C002 Trade Abba	PR.F.20.10	Building Elements Demolition
CO-S009	CO-S009	15JUN04	C002 Trade Abba	PR.F.20.10	Building Elements Demolition
CO-S010	CO-S010	15JUN04	C002 Trade Abba	PR.F.20.10	Building Elements Demolition
CO-S011	CO-S011	15JUN04	C002 Trade Abba	PR.B.10.10	Floor
CO-S012	CO-S012	13APR04	C019 Trade McLeod	PR.F.20.10 PR.LSI.L10.SL1 Tower	Building Elements Demolition
CO-S013	CO-S013	21JUN04	C019 Trade McLeod	PR.F.20.10	Building Elements Demolition
CO-S014.1	CO-S014.1	21JUN04	C009 Trade George	PR.H.10.10 PR.LSI.L1	Plan Changes & Revisions Global
CO-S014.2	CO-S014.2	21JUN04	C009 Trade George	PR.H.10.10 PR.LSI.L11.SL1 West Wing	Plan Changes & Revisions West Wing
CO-S015	CO-S015	21JUN04	C009 Trade George	PR.LSI.L6.SL1	West Wing
CO-S016	CO-S016	21JUN04	C009 Trade George	PR.H.10.10	Plan Changes & Revisions
CO-S017.1	CO-S017.1	19NOV04	C015 Trade Lower Mainland	PR.B.10.20 PR.LSI.L11.SL1 West Wing	Roof West Wing
CO-S017.2	CO-S017.2	19NOV04	C018 Trade Marine	PR.B.10.20 PR.LSI.L11.SL1 West Wing	Roof West Wing
CO-S017.3	CO-S017.3	19NOV04	C022 Trade Shanahan	PR.B.10.20 PR.LSI.L11.SL1 West Wing	Roof West Wing
CO-S017.4	CO-S017.4	19NOV04	C026 Trade Western	PR.B.10.20 PR.LSI.L11.SL1 West Wing	Roof West Wing
CO-S017.5	CO-S017.5	19NOV04	C015 Trade Lower Mainland	PR.F.20.10 PR.LSI.L11.SL1 West Wing	Building Elements Demolition West Wing
CO-S017.6	CO-S017.6	19NOV04	C018 Trade Marine	PR.F.20.10 PR.LSI.L11.SL1 West Wing	Building Elements Demolition West Wing
CO-S017.7	CO-S017.7	19NOV04	C022 Trade Shanahan	PR.F.20.10 PR.LSI.L11.SL1 West Wing	Building Elements Demolition West Wing
CO-S017.8	CO-S017.8	19NOV04	C026 Trade Western	PR.F.20.10 PR.LSI.L11.SL1 West Wing	Building Elements Demolition West Wing
CO-S018	CO-S018			PR.G.30.30	Storm Sewer System
CO-S019	CO-S019			PR.G.30.20	Sanitary Sewer System
CO-S020	CO-S020		C009 Trade George	PR.H.10.10 PR.LSI.L1	Plan Changes & Revisions Global
CO-S021	CO-S021		C009 Trade George		
CO-S022.1	CO-S022.1	31MAY04	C009 Trade George	PR.C.20.10 PR.LSI.L7.SL3	Stair Construction East Wing
CO-S022.2	CO-S022.2	31MAY04	C009 Trade George	PR.C.20.10 PR.LSI.L8.SL3	Stair Construction East Wing
CO-S023.1	CO-S023.1	06JUL04	C009 Trade George	PR.B.10.10 PR.LSI.L7.SL3	Floor East Wing
CO-S023.2	CO-S023.2	06JUL04	C009 Trade George	PR.B.10.10 PR.LSI.L9.SL1	Floor Tower
CO-S023.3	CO-S023.3	06JUL04	C009 Trade George	PR.B.10.10 PR.LSI.L11.SL2	Floor Tower
CO-S023.4	CO-S023.4	06JUL04	C009 Trade George	PR.B.10.20 PR.LSI.L7.SL3	Roof East Wing
CO-S023.5	CO-S023.5	06JUL04	C009 Trade George	PR.B.10.20 PR.LSI.L9.SL1	Roof Tower
CO-S023.6	CO-S023.6	06JUL04	C009 Trade George	PR.B.10.20 PR.LSI.L11.SL2	Roof Tower
CO-S023.7	CO-S023.7	06JUL04	C009 Trade George	PR.C.20.10 PR.LSI.L7.SL3	Stair Construction East Wing
CO-S023.8	CO-S023.8	06JUL04	C009 Trade George	PR.C.20.10	Stair Construction

Figure 5.29 The first page of the 24 page tabular report of change orders that includes information about project participants who will execute the change orders, products and locations involved with a change order, and change order issue date.

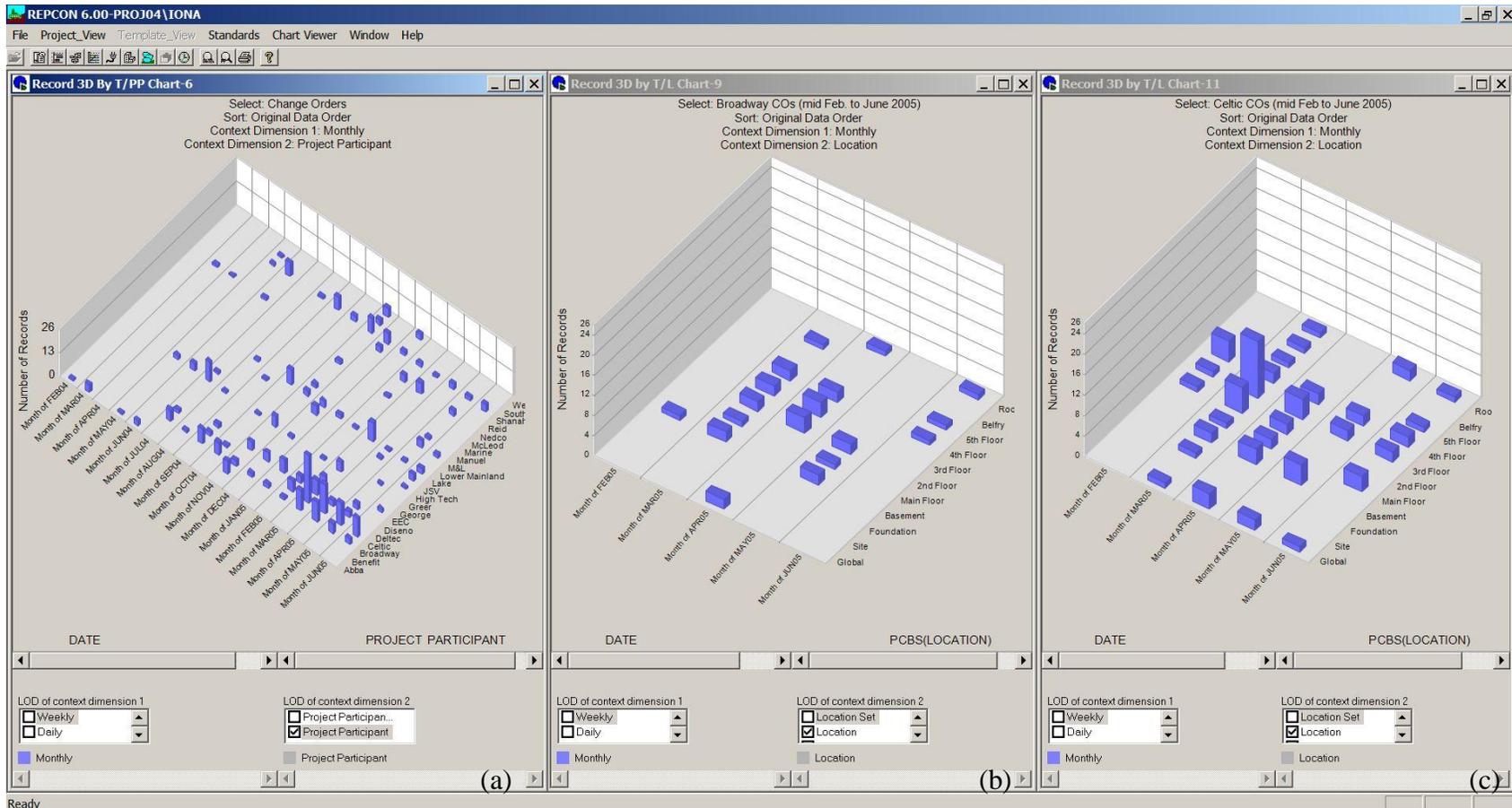


Figure 5.30 (a) Number of all change orders distributed in the time dimension at the level of granularity of "month" and the project participant dimension at the level of granularity of "project participant". Number of change orders issued between mid February and June 2005 that are associated with the (b) Broadway trade, and (c) Celtic trade, distributed in the time dimension at the level of granularity of "month" and the location dimension at the level of granularity of "location".

### 5.3.4 Analysis of demonstration results

Summarized in Table 5.7 is a comparison between the use of a CM data visualization environment and current traditional data reporting features in the six demonstration cases. From this table it is observed that the use of a CM data visualization environment allows users to identify more analytic reasoning artifacts and more quickly than examining data in its original textual or tabular form, which conforms with the popular belief that utilizing data visualization helps one interpret data better than reading data in its original textual or tabular form. How the concepts of "data presented in visual forms", "environment architecture", and "use of interaction features" central to a CM data visualization environment contribute to enhanced CM analytics is summarized in Table 5.7. In this table, the lettered items represent key features that correspond to the foregoing concepts. As seen in the rightmost column of Table 5.7, the differences between features possessed by the CM data visualization environment and current data reporting capabilities indicate that the following features play the most important roles in enhancing CM analytics capabilities:

1. By presenting data in visual form, salient visual patterns representing values and patterns of behaviour of construction conditions or performance measures can be instantly observed (represented by "A" in Table 5.7).
2. Images that depict the distribution of values of either construction conditions or performance measures in various context dimensions provide helpful insights as to whether or not their causes can be inferred or the impacts they generate can be inferred (represented by "B1" in Table 5.7).
3. A thematic visualization representing CM data, which is collected/computed for a certain CM function/task, can also be utilized for CM analytics supporting other CM functions/tasks. Many such thematic visualizations have been organized for CM\_IS users to access and conduct flexible analytical reasoning (represented by "B2" in Table 5.7).
4. Users can adjust/select visual formats of data presentations of a specific thematic visualization on demand to meet their CM analytics needs (represented by "C2" in Table 5.7).
5. Users can adjust/select data contents in terms of granularity of data sets of the visualization chosen on demand to meet their CM analytics needs (represented by "C4" in Table 5.7). This involves choice of levels of granularity of context dimensions and aggregations of values over context dimensions of different levels of granularity.

Table 5.7 Summarized comparison between the use of a CM data visualization environment and current (traditional) data reporting features for the six demonstration cases

Demonstration Case	Data reporting methods	Number of analytics	Viewing time	Primary features of a CM data visualization environment and the current data reporting functionalities
Project 1- demonstration case 1	Current data reporting practice	2	10 sec	Partly A <sup>6</sup> , B (B1), C (C1, C2, C3, C4)
	Use of a CM data visualization environment	5	10 sec	A, B (B1), C (C1, C2, C3, C4)
Project 1- demonstration case 2	Current data reporting practice	2	150 sec	C (C1, C3)
	Use of a CM data visualization environment	4	20 sec	A, B (B1, B2), C (C1, C2, C3, C4)
Project 2- demonstration case 1	Current data reporting practice	0	>300 sec	C (C1, C3)
	Use of a CM data visualization environment	5	25 sec	A, B(B1, B2), C (C1, C2, C3, C4)
Project 2- demonstration case 2	Current data reporting practice	2	150 sec	C (C1, C3)
	Use of a CM data visualization environment	5	10 sec	A, B (B1, B2), C (C1, C2, C3, C4)
Project 3- demonstration case 1	Current data reporting practice	1	250 sec	C (C1, C3)
	Use of a CM data visualization environment	6	25 sec	A, B (B1, B2), C (C1, C2, C3, C4)
Project 3- demonstration case 2	Current data reporting practice	0	240 sec	C (C1, C3)
	Use of a CM data visualization environment	4	20 sec	A, B (B1, B2), C (C1, C2, C3, C4)
<p><b>Alphanumeric numbering representing primary features of a CM data visualization environment and the current data reporting functionalities</b></p> <p>A: By presenting data in visual form, salient visual patterns representing values and patterns of behaviour of construction conditions or performance measures can be instantly observed.</p> <p>B: Environment architecture:            B1: Images that depict the distribution of values of either construction conditions or performance measures in various context dimensions provide helpful insights as to whether or not their causes can be inferred or the impacts they generate can be inferred.            B2: A thematic visualization representing CM data, which is collected/computed for a certain CM function/task, can also be utilized for CM analytics supporting other CM functions/tasks. Many such thematic visualizations have been organized for CM_IS users to access and conduct flexible analytical reasoning.</p> <p>C: Use of interaction features:            C1: Users can select a suitable thematic visualization (or thematic tabular report) from the organization of visualizations (or tabular reports) representing CM data collected/computed for a range of CM functions/tasks on demand to meet their CM analytics needs.            C2. Users can adjust/select visual formats of data presentations of a specific thematic visualization (or thematic tabular report) on demand to meet their CM analytics needs.            C3. Users can adjust/select data contents in terms of range of data sets of the visualization (or tabular reports) chosen on demand to meet their CM analytics needs            C4. Users can adjust/select data contents in terms of granularity of data sets of the visualization (or tabular reports) chosen on demand to meet their CM analytics needs</p>				

<sup>6</sup> Because the location information is presented in textual forms; the information about both spatial and temporal ordering of activities is also not visually encoded

The significance of statement "A" in Table 5.7: "*By presenting data in visual form, salient visual patterns representing values and patterns of behaviour of construction conditions or performance measures can be instantly observed*" is amplified and demonstrated by the *project 1-demonstration case 1*. At first glance, both Figures 5.6 and 5.7 are visual representations of data. However, in Figure 5.6 the important information regarding the locations of activities is in fact in textual form and the information about both spatial and temporal ordering of activities is also not visually encoded. The performance differences in terms of numbers of analytic reasoning artifacts obtained can be solely attributed to whether or not data are presented in visual forms. Another important aspect of the value of presenting data in visual forms is that the compelling visual patterns instantly perceived help provide more insights and direction for further investigation of the data. For example, in *project 2-demonstration case 2*, although through the use of the 16 page tabular report (Figure 5.20) likely patterns of long lead time deficiencies being mostly either related to the Window trade or of the "damage/defect" type of issue were perceived, one's confidence in this perception is much weaker than the clarity and hence greater confidence afforded by the visual patterns in Figure 5.21. As a consequence of being somewhat overwhelmed when viewing the 16 page tabular report, the initial observation of the repeated occurrence of "Window trade" or "Damage/defect" texts may cause the user to miss the high number of occurrences of "Appliance trade" texts and the relationship between which trades have what kinds of deficiencies. As a result, users may find it difficult to formulate more specific follow up CM questions with which to continue the data analysis process in order to arrive at more concrete conclusions. In comparison, the deficiency data presented in visual form as shown in Figure 5.21 project strong visual patterns to the viewer and therefore messages hidden in the data are hard to miss and are immediately received. This in turn prompts an investigation into the Appliance and Window trades and ultimately helps the user to realize the main causes of the long lead time deficiencies of these two trades and their implications for time performance.

An example signifying the importance of statement "B1" in Table 5.7: "*Images that depict the distribution of values of either construction conditions or performance measures in various context dimensions provide helpful insights as to whether or not their causes can be inferred or the impacts they generate can be inferred*" can be observed in the *Project 3-demonstration case*

*1* and *Project 3-demonstration case 2*. In Figure 5.26, the number of change orders distributed by the change order type, product, and location context dimensions is shown for reasoning about whether change orders mainly originate from a certain change order type, product, and/or location so that management actions can be taken to control to the extent possible the proliferation of change orders in the later phase of the project. For the same change order data, in Figure 5.28 the number of change orders distributed in the project participant context dimension help the user to identify potential impacts of change orders on an individual participant and its associated context dimensions such as pay items and activities. In Figure 5.30 the distribution of change orders in the time (change order issued dates), project participant, and location context dimensions also helps in identifying potential impacts on individual trades, work time windows, and/or locations and their associated activities. As a comparison, the tabular report seen in Figure 5.25 has a fixed format that lists data dimensions horizontally at the top and vertically lists values corresponding to these dimensions in a long linear way. This makes observation of the distribution of construction conditions and performance measure values in different combinations of context dimensions very difficult, if not impossible. Further, opportunities for grasping potential causes of or impacts by change orders as a function of different combinations of context dimensions are lost.

An example demonstrating the significance of statement "B2" in Table 5.7: "*A thematic visualization representing CM data, which is collected/computed for a certain CM function/task, can also be utilized for CM analytics supporting other CM functions/tasks. Many such thematic visualizations have been organized for CM-IS users to access and conduct flexible analytical reasoning*" can be observed in the *Project 1-demonstration case 2*. The image of Figure 5.10 shows how planned values of the "concrete quantity", "formwork area", and "reinforcing bar length" attributes are distributed in the product and location dimensions. Understanding this distribution helps in reasoning about how these distributions may impact the product and location dimensions and their associated activity dimension, thereby prompting an investigation of activity related data. Therefore, although the image presenting data regarding products/locations and product quantities (Figure 5.11(b)) is originally intended for scope management use, it can also be utilized alongside other images (e.g. a LP chart as seen in Figure 5.11(a)) to assist users in conducting analytical reasoning for time management related tasks (e.g.

assessing the quality of duration estimates as demonstrated in this case). As a comparison, although the current data reporting functionalities also allow users to select from a variety of tabular reports and a limited number of graphics, the difficulties in obtaining insights from individual tabular reports hinder the opportunity for achieving new or additional insights from a combination of reports. In other words, their use is mainly limited to the purposes for which they were originally designed.

One example showcasing the usefulness of statement "C2" in Table 5.7: "*Users can adjust/select visual formats of data presentations of a specific thematic visualization on demand to meet their CM analytics needs*" is seen in the *Project 2-demonstration case 1*. Each image in Figure 5.17 is configured to show how number of deficiencies as a performance measure distributed in the project participant and location dimensions can help the user to reason how the sources of these deficiencies may be a function of these two context dimensions. The same thematic visualization in terms of deficiency distribution graphics can also be configured to present their distribution in the product dimension at different levels of detail, as seen in Figure 5.19. This assists users in testing their hypothesis as to whether or not other context dimensions may also be a source of or contribute to the deficiency issues. As a comparison, the tabular report seen in Figure 5.16 has a fixed format that lists data dimensions horizontally at the top and vertically lists values corresponding to these dimensions. This makes observation of the distribution of data values in different combinations of context dimensions very difficult, if not impossible. (The manner in which this issue is addressed in most current commercial systems is through a proliferation of predefined tabular reports plus the inclusion of an ever more general report writer. However, one is still left with a tabular report, now potentially many more of them, and the required insights may still be hard to come by.)

*Project 2-demonstration case 1* provides an example that shows the necessity of statement "C4" in Table 5.7: "*Users can adjust/select data contents in terms of granularity of data sets of the visualization chosen on demand to meet their CM analytics needs*". "Data in visual forms" as seen in Figure 5.17 is not just simply the visual encoding of values of the almost 8200 rows in the 304 page tabular report. It also involves data transformations in terms of aggregating counts of deficiencies corresponding to items of context dimensions at different levels of granularity.

Thus, a 304 page report recording 8200 deficiencies can be compacted into an single image showing the distribution of deficiencies by floor and project participant as well as by individual suites floor by floor. With this single compact image one can reason as to whether deficiencies are dependent on floors, suites, and/or project participants. In the same vein, one can observe in Fig 5.19 as to whether a Painter's deficient work relates to particular systems, subsystem, and/or elements. As a comparison, it is almost impossible to obtain the foregoing rich insights by perusing the 304 page report seen in Figure 5.16.

### **5.3.5 Conclusions from the demonstration and analysis**

The demonstration examples and the post-demonstration analysis present a strong case for the merits of using a CM data visualization environment and how the key concepts behind it contribute to enhancing CM analytics abilities. The demonstration example and associated analytic reasoning artifacts present research findings that respond to or answer research question 3 (how the use of a CM data visualization environment may help conduct CM analytics that cannot or are difficult to be done with current data reporting practices), as set forth at the beginning of this research. The highly structured comparison methodology employed, both in the demonstration and post-demonstration analysis, speaks to the validity of this research findings.

The structured way of developing a CM data visualization environment that combines a bottom-up development approach (see Chapter 4) integrated with the top-down design approach (see Chapter 3) and design guidelines (see Chapter 2) is neither CM function dependent nor CM data type dependent and therefore can be readily applied to developing other thematic visualizations in support of other yet to be explored CM functions/tasks. For example, as part of the last stage of research work, this methodology was easily extended to developing the thematic visualization of "counts of as-built records" whereas counts of various types of records (e.g. meeting minutes, request for information (RFIs), change orders) are treated as actual execution conditions or performance measures (e.g. number of deficiencies identified represents the status of quality performance). Many of the images generated in the demonstration cases (Figure 5.17, Figure 5.18, Figure 5.19, Figure 5.21, Figure 5.26, Figure 5.27, Figure 5.28, Figure 5.30) are images of different types and/or contents derived from this thematic visualization. Furthermore, it was demonstrated that the visualizations as developed can be utilized across functions, and not for

just a single purpose. Hopefully CM users will recognize the multi-use potential of individual images in support of CM analytics applicable to multiple functions/tasks. This generality of the research findings in terms of development methods and application of a CM data visualization environment is seen as an important contribution.

#### **5.4 Summary of research contributions**

Three main research contributions were made in this Ph.D. research, as outlined in Chapter 1, along with some elaboration of the first two. In addition to their description in Chapter 1, the third contribution was demonstrated to readers in section 5.3 of this chapter.

The first contribution relates to the identification of methodologies for developing a CM visualization environment, which includes: 1) formulation of a set of design guidelines/principles in terms of how to apply state-of-the-art data visualization (or visual analytics) techniques to the CM domain data, 2) introduction of a top-down design approach for identifying common CM analytic needs and the corresponding visualization requirements, and 3) employment of a structured development process that combines a bottom-up design process integrated with design guidelines and a top-down design process.

The second contribution treats the key features required of a CM data visualization environment. The key features can be found in Tables 1.2~1.4 of Chapter 1 and are summarized as:

1. An organization of thematic visualizations, mainly hierarchical from abstract to specific, that are categorized by construction conditions and performance measures under multiple views of a project.
2. Features of images generated from the foregoing visualizations categorized as:
  - Image themes by construction conditions or performance measures (scope, time, cost, safety, quality, etc), which are treated as measurement dimensions.
  - Image types by: a) project context dimensions against which construction conditions or performance measures are mapped, b) non-variance or variance values, c) visual encodings.
  - Image contents by: a) granularity of context dimensions, b) items selection of project context dimensions, c) data status states, c) data versions (dates), d) how to aggregate

measurements, e) how to compute variances between the planned and actual values of measurements.

- Image formats by: a) distribution of values of construction conditions or performance measures in project context dimensions, b) distribution of values of several construction conditions or performance measures in the occurrence time dimension and definition dimension (i.e. definitions of construction conditions or performance measures ).
- Image display options by: a) encode "holiday" "non-working day" if the time dimension is involved, b) how to include values of context dimensions, c) enhancing the visual grouping of certain items of context dimensions that are not visually encoded by spatial position.

3. Provide image specific interaction features to allow users to change and/or set options that characterize image themes, image types, image contents, image formats, and image display options. Also provide general interactivity such as image navigation for enhancing readability of images as well as a mechanism for coordinating interaction features to increase the efficiency of operating a CM visualization environment.

The third contribution relates to demonstrating and analyzing, through case studies, how the ability to carry out CM analytics useful for dealing with the CM tasks at hand may be enhanced using data reports in visual form that are responsive to these analytics tasks and by using interaction features and an environment architecture that allows users to flexibly explore CM data collected/computed for a range of CM tasks/functions presented in visual form.

## **5.5 Future work**

For future research work related to the topic of developing a comprehensive CM data visualization environment, more visualization design cases will be conducted using the structured development process described in this thesis in order to: 1) construct a more complete organization of thematic visualizations, and 2) enrich and refine the design guidelines and checklists of common visualization features for developing other visualizations, both in support of analytical reasoning for a broader range of functions and tasks. A design case that has high priority for future development are visualizations dealing with the function of risk management because of the involvement of uncertainties in the values of CM variables. Another future research topic deals with a CM data visualization "development environment". Such a

development environment would incorporate the knowledge of design guidelines, workflow of the top-down and the bottom-up design approaches, and checklists of common visualization features with respect to user interfaces to specify requirements and specifications of new or enhanced visualizations. Included would be mechanisms to automatically generate visualizations in compliance with these specifications. With this development environment, new visualizations could be quickly prototyped and appraised to assess their potential usefulness and hence whether or not to include them in the organization of thematic visualizations for use in practice. Followed up work to this would involve deployment of both a CM data visualization environment and a CM data visualization "development environment" in CM organizations and their use in day to day project data analysis work in support of various CM functions/tasks. Through longitudinal field observations, an in depth understanding into the thematic visualizations and their corresponding image types, formats, contents, and visual encoding most often utilized by professional practitioners could be elicited.

## Bibliography

Abudayyeh, O., and Al-Battaineh, H. T. (2003). "As-Built Information Model for Bridge Maintenance." *J.Comp.in Civ.Engrg.*, 17(2), 105-112.

Acumen PM, L. (2010). "Acumen project confidence-included metrics." <http://www.projectacumen.com/fuse/overview/metrics/> (11/11, 2010).

Ahlberg, C., Williamson, C., Shneiderman, B. (1992). "Dynamic queries for information exploration: An implementation and evaluation." *Proc., CHI '92: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Monterey, California, United States, 619-626.

Ahlberg, C., and Shneiderman, B. (1994). "Visual information seeking: Tight coupling of dynamic query filters with starfield displays." *Proc., Human Factors in Computing Systems*, ACM Press, New York, NY, USA, 313-317.

Alonso, D. L., Rose, A., Plaisant, C., Norman, K. L. (1997). "Viewing Personal History Records: A Comparison of Tabular Format and Graphical Presentation using LifeLines." *Behav. Inf. Technol.*, 17(5), 249-262.

Amar, R., Eagan, J., Stasko, J. (2005). "Low-level components of analytic activity in information visualization." *Proc., INFOVIS '05: Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, IEEE Computer Society, Minneapolis, Minnesota, USA, 111-117.

Amar, R. A., and Stasko, J. T. (2005). "Knowledge Precepts for Design and Evaluation of Information Visualizations." *IEEE Trans. Visual. Comput. Graphics*, 11(4), 432-442.

Anbari, F. T. (2003). "Earned Value Project Management Method and Extensions." *Project Management Journal*, 34(4), 12-23.

Ardito, C., Buono, P., Costabile, M. F., Lanzilotti, R. (2006). "Systematic inspection of information visualization systems." *Proc., Proceedings of the 2006 AVI Workshop on BEyond*

*Time and Errors: Novel Evaluation Methods for Information Visualization*, ACM, Venice, Italy, 1-4.

Babu, A. J. G., and Suresh, N. (1996). "Project Management with Time, Cost, and Quality Considerations." *European Journal of Operational Research*, 88(2), 320-327.

Battikha, M. G. (2008). "Reasoning Mechanism for Construction Nonconformance Root-Cause Analysis." *Journal of Construction Engineering and Management*, 134(4), 280-288.

Ball, T., and Eick, S. G. (1996). "Software Visualization in the Large." *Computer*, 29(4), 33-43.

Becker, R. A., Cleveland, W. S., Wilks, A. R. (1987). "Dynamics Graphics for Data Analysis." *Statistical Science*, 2(4), 355-396.

Becker, R. A., Cleveland, W. S., Shyu, M. (1996). "The Visual Design and Control of Trellis Display." *Journal of Computational and Graphical Statistics*, 5(2), 123-155.

Beniger, J. R., and Robyn, D. L. (1978). "Quantitative Graphics in Statistics: A Brief History." *The American Statistician*, 32(1), 1-11.

Bertin, J. (1983 (originally published in French in 1967)). *Semiology of Graphics*, University of Wisconsin Press, Milwaukee.

Boukhelifa, N., Roberts, J. C., Roberts, P. J., Rodgers, P. J. (2003). "A coordination model for exploratory multi-view visualization." *Proc., CMV '03: Proceedings of the Conference on Coordinated and Multiple Views in Exploratory Visualization*, IEEE Computer Society, London, England, 76-85.

Brath, R. K. (1999). "Effective Information Visualization, Guidelines and Metrics for 3D Interactive Representations of Business Data." *Master's Thesis, University of Toronto, Canada*.

Brath, R. (2003). "Paper landscapes: A visualization design methodology." *Proc., Conference on Visualization and Data Analysis (VDA 2003)*, Society of Photo-Optical Instrumentation Engineers, Santa Clara, CA, USA, 125-132.

Brath, R., Peters, M., Senior, R. (2005). "Visualization for Communication: The Importance of Aesthetic Sizzle." *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, Washington, DC, USA, 724-729.

Brautigam, M. (1996). "Applying Information Visualization Techniques to Web Navigation " *Thesis Proposal: UC Santa Cruz, CA, USA.*,

Brinton, W. C. (1939). *Graphic Presentation*, 1st Ed., Brinton associates, New York city, USA.

Card, S. K., Mackinlay, J., Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*, 1st Ed., Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Card, S. K., Robertson, G. G., Mackinlay, J. D. (1991). "The information visualizer, an information workspace." *Proc., CHI '91: Conference on Human Factors in Computing Systems*, ACM, New Orleans, Louisiana, United States, 181-186.

Carpendale, S. (2008). "Evaluating Information Visualization." *Information Visualization*, Springer, U.S.A., 21-49.

Casner, S. M. (1991). "Task-Analytic Approach to the Automated Design of Graphic Presentations." *ACM Trans.Graph.*, 10(2), 111-151.

Cawthon, N., and Moere, A. V. (2007). "The Effect of Aesthetic on the Usability of Data Visualization." *Information Visualization, 2007. IV '07. 11th International Conference*, IEEE Computer Society, Zurich, Switzerland, 637-648.

Chao, L., and Skibniewski, M. J. (1998). "Fuzzy Logic for Evaluating Alternative Construction Technology." *Journal of Construction Engineering and Management*, 124(4), 297-304.

Chen, C., and Yu, Y. (2000). "Empirical Studies of Information Visualization: A Meta-Analysis." *Int. J. Hum. -Comput. Stud.*, 53(5), 851-866.

Chintalapani, G., Plaisant, C., Shneiderman, B. (2004). "Extending the utility of treemaps with flexible hierarchy." *Proc., Information Visualisation, 2004. IV 2004. Proceedings. Eighth*

*International Conference on*, IEEE Computer Society, London, England, 335-344.

Chiu, C., and Russell, A. D. (2011). "Design of a Construction Management Data Visualization Environment: A top-down Approach." *Autom. Constr.*, 20 (4), 399-417.

Chong, W. K., O'Connor, J. T., Chou, J., Lee, S. (2005). "Predicting the production rates of foundation construction using factor and regression analysis." *Proc., Construction Research Congress 2005*, ASCE, San Diego, CA, USA, 56.

Chuah, M. C., Roth, S. F., Mattis, J., Kolojejchick, J. (1995). "SDM: Selective dynamic manipulation of visualizations." *Proc., Proceedings UIST' 95 Symposium on User Interface Software and Technology*, ACM, Pittsburgh, PA, USA, 61-70.

Chuah, M. C., and Roth, S. F. (1996). "On the semantics of interactive visualizations." *Proc., INFOVIS '96: Proceedings of the 1996 IEEE Symposium on Information Visualization (INFOVIS '96)*, IEEE Computer Society, Washington DC, USA, 29.

Cleveland, W. S. (1985). *The Elements of Graphing Data*, Wadsworth Advanced Book Program, USA.

Cleveland, W. S., and McGill, R. (1984). "Graphical Perception: Theory, Experimentation and Application to the Development of Graphical Methods." *Journal of the American Statistical Association*, 79(387), 531-554.

Cleveland, W. S., and Becker, R. A. (1987). "Brushing Scatterplots." *Technometrics*, 29(2), 127-142.

Cockburn, A., Karlson, A., Bederson, B. B. (2008). "A Review of overview+detail, Zooming, and focus+context Interfaces." *ACM Comput.Surv.*, 41(1), 1-31.

Collier, E., and Fischer, M. (1995). "Four-Dimensional Modeling in Design and Construction", *CIFE Technical Report #101*, CIFE, Stanford University, USA.

Convertino, G., Chen, J., Yost, B., Ryu, Y., North, C. (2003). "Exploring context switching and

cognition in dual-view coordinated visualizations." *Proc., CMV '03: Proceedings of the Conference on Coordinated and Multiple Views in Exploratory Visualization*, IEEE Computer Society, London, England, 55-62.

Cooke, B., and Williams, P. (2004). *Construction Planning, Programming & Control*, Blackwell Publishing, UK.

Cox, I. D., Morris, J. P., Rogerson, J. H., Jared, G. E. (1999). "A Quantitative Study of Post Contract Award Design Changes in Construction." *Construction Management & Economics*, 17(4), 427-439.

Dawood, N., Scott, D., Sriprasert, E., Mallasi, Z. (2005). "The Virtual Construction Site (VIRCON) Tools: An Industrial Evaluation." *ITcon*, 10(Special (From 3D to nD modelling )), 43-54.

de Leon, G. P. (2008). "Project Planning using Logic Diagramming Method." *AACE International Transactions*, 2008, 1-6.

Demian, P., and Fruchter, R. (2006). "Finding and Understanding Reusable Designs from Large Hierarchical Repositories." *Information Visualization*, 5(1), 28-46.

Diekmann, J. E. (1992). "Risk Analysis: Lessons from Artificial Intelligence." *International Journal of Project Management*, 10(2), 75-80.

Duncan, J., and Humphreys, G. W. (1989). "Visual Search and Stimulus Similarity." *Psychol. Rev.*, 96(3), 433-458.

Eastman, C. M. (1999). *Building Product Models: Computer Environments Supporting Design and Construction*, CRC Press, Inc, Boca Raton, FL, USA.

Eick, S. G. (2000). "Visual Discovery and Analysis." *IEEE Transactions on Visualization and Computer Graphics*, 6(1), 44-58.

Elhakeem, A., and Hegazy, T. (2005). "Graphical Approach for Manpower Planning in

Infrastructure Networks." *J. Constr. Engrg. and Mgmt.*, 131(2), 168-175.

Feldt, N., Pettersson, H., Johansson, J., Jern, M. (2005). "Tailor-made exploratory visualization for statistics Sweden." *Proc., CMV '05: Proceedings of the Coordinated and Multiple Views in Exploratory Visualization*, IEEE Computer Society, London, England, 133-142.

Fondahl, J. W. (1962). *A Non-Computer Approach to the Critical Path Method for the Construction Industry*, 2nd Ed., Department of Civil Engineering, Stanford University, Stanford, California, USA.

Forsell, C., and Johansson, J. (2010). "An heuristic set for evaluation in information visualization." *Proc., Proceedings of the International Conference on Advanced Visual Interfaces*, ACM, Roma, Italy, 199-206.

Freitas, C. M. D. S., Luzzardi, P. R. G., Cava, R. A., Winckler, M. A. A., Pimenta, M. S., Nedel, L. P. (2002). "Evaluating usability of information visualization techniques." *Proc., 5th Symposium on Human Factors in Computer Systems*, Brazilian Computer Soc. Press, Fortaleza, Ceara, Brazil, 40-51.

Friel, S. N., Curcio, F. R., Bright, G. W. (2001). "Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications." *Journal for Research in Mathematics Education*, 32(2), 124-158.

Friendly, M. (2008). "A Brief History of Data Visualization." *Handbook of Data Visualization*, Springer, USA, 57-78.

Froese, T. (1996). "Models of Construction Process Information." *J.Comp.in Civ.Engrg.*, 10(3), 183-193.

Fua, Y., Ward, M. O., Rundensteiner, E. A. (1999). "Hierarchical parallel coordinates for exploration of large datasets." *Proc., VIS '99: Proceedings of the Conference on Visualization '99*, IEEE Computer Society Press, San Francisco, California, United States, 43-50.

Garner, W. R. (1974). *The Processing of Information and Structure*, Lawrence Erlbaum

Associates, Potomac, Maryland, USA.

Gonzalez, V., and Kobsa, A. (2003). "Benefits of information visualization systems for administrative data analysts." *Proc., IV '03: Proceedings of the Seventh International Conference on Information Visualization*, IEEE Computer Society, London, England, 331-336.

Graham, M., Kennedy, J., Beyond, D. (2000). "Towards a Methodology for Developing Visualizations." *International Journal of Human-Computer Studies*, 53(5), 789-807.

Grinstein, G., Kobsa, A., Plaisant, C., Shneiderman, B., Stasko, J. (2003). "Which comes first, usability or utility?" *Proc., VIS '03: Proceedings of the 14th IEEE Visualization 2003 (VIS'03)*, IEEE Computer Society, Washington DC, USA, 605-606.

Halley, E. (1686). "On the Height of the Mercury in the Barometer at Different Elevations Above the Surface of the Earth, and on the Rising and Falling of the Mercury on the Change of Weather." *Philosophical Transaction*, 16(January 1), 104-116.

Hanna, A. S., Camlic, R., Peterson, P. A., Lee, M. (2004). "Cumulative Effect of Project Changes for Electrical and Mechanical Construction." *J. Constr. Engrg. and Mgmt.*, 130(6), 762-771.

Heesom, D., and Mahdjoubi, L. (2004). "Trends of 4D CAD Applications for Construction Planning." *Construction Management and Economics*, 22(2), 171-182.

Hegazy, T., Elbeltagi, E., Zhang, K. (2005). "Keeping Better Site Records using Intelligent Bar Charts." *J. Constr. Engrg. and Mgmt.*, 131(5), 513-521.

Hegazy, T., and Kamarah, E. (2008). "Efficient Repetitive Scheduling for High-Rise Construction." *J. Constr. Engrg. and Mgmt.*, 134(4), 253-264.

Herman, I., Melancon, G., Marshall, M. S. (2000). "Graph Visualization and Navigation in Information Visualization: A Survey." *Visualization and Computer Graphics, IEEE Transactions on*, 6(1), 24-43.

- Hofmann, H. (2006). "Multivariate Categorical Data-Mosaic Plots." *Graphics of Large Datasets: Visualizing a Million*, Springer, USA, 73-101.
- Ibbs, C. W. (1997). "Quantitative Impacts of Project Change: Size Issues." *J. Constr. Engrg. and Mgmt.*, 123(3), 308-311.
- Institute, P. M. (2008). *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)*, 4th Ed., Project Management Institute, Inc, Newtown Square, Pa., USA.
- Johnston, D. W. (1981). "Linear Scheduling Method for Highway Construction." *J. Constr. Div.*, 107(2), 247-261.
- Karim, A., and Adeli, H. (1999). "CONSCOM: An OO Construction Scheduling and Change Management System." *J.Constr.Engrg.and Mgmt.*, 125(5), 368-376.
- Keim, D. A., Mansmann, F., Schneidewind, J., Ziegler, H. (2006). "Challenges in visual data analysis." *Proc., IV '06: Proceedings of the Conference on Information Visualization*, IEEE Computer Society, London, England, 9-16.
- Keller, R., Eckert, C. M., Clarkson, P. J. (2005). "Multiple Views to Support Engineering Change Management for Complex Products." *Coordinated and Multiple Views in Exploratory Visualization, 2005. (CMV 2005). Proceedings. Third International Conference on*, IEEE Computer Society, London, England, 33-41.
- Khosrowshahi, F. (2000). "Information visualization in aid of construction project cash flow management." *Proc., IV '00: Proceedings of the International Conference on Information Visualisation*, IEEE Computer Society, London, England, 583-588.
- Kim, C. S., and Liu, L. Y. (2007). "Cost Information Model for Managing Multiple Projects." *J.Constr.Engrg.and Mgmt.*, 133(12), 966-974.
- Kobsa, A. (2001). "An empirical comparison of three commercial information visualization systems." *Proc., Information Visualization, 2001. INFOVIS 2001. IEEE Symposium on*, IEEE Computer Society, San Diego, CA, USA, 123-130.

Korde, T. (2005). "Visualization of Construction Data." *Master's Thesis, University of British Columbia, Canada.*

Korde, T., Wang, Y., Russell, A. D. (2005). "Visualization of construction data." *Proc., 6th Construction Specialty Conference, Canadian Society of Civil Engineers, Toronto, Ontario, Canada, CT-148-1-CT-148-11.*

Lamping, J., Rao, R., Pirolli, P. (1995). "A focus+context technique based on hyperbolic geometry for visualizing large hierarchies." *Proc., Proceedings of the Conference on Human Factors in Computing Systems, ACM Press/Addison-Wesley Publishing Co., Denver, CO, USA, 401-408.*

Lamping, J., and Rao, R. (1996). "Visualizing large trees using the hyperbolic browser." *Proc., CHI '96: Conference Companion on Human Factors in Computing Systems, ACM, Vancouver, British Columbia, Canada, 388-389.*

Larkin, J. H., and Simon, H. A. (1987). "Why a Diagram is (Sometimes) Worth Ten Thousand Words." *Cognitive Science, 11(1), 65-100.*

Lee, N., and Rojas, E. M. (2009). "Developing effective visual representations to monitor project performance." *Proc., 2009 Construction Research Congress, ASCE, Seattle, WA, USA, 826-835.*

Lengler, R., and Eppler, M. (2007). "Towards a periodic table of visualization methods for management." *Proc., IASTED Proceedings of the Conference on Graphics and Visualization in Engineering (GVE 2007), ACTA Press, Clearwater, Florida, USA, six pages.*

Leung, Y. K., and Apperley, M. D. (1994). "A Review and Taxonomy of Distortion-Oriented Presentation Techniques." *ACM Trans. Comput. -Hum. Interact., 1(2), 126-160.*

Levy, E., Zacks, J., Tversky, B., Schiano, D. (1996). "Gratuitous graphics? putting preferences in perspective." *Proc., CHI '96: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Vancouver, British Columbia, Canada, 42-49.*

Li, M. (2009). "Diagnosing Construction Performance by using Causal Models." *Ph.D. Thesis,*

*University of British Columbia, Vancouver, B.C., Canada.*

Liston, K., Fischer, M., Kunz, J. (2000). "Designing and evaluating visualization techniques for construction planning." *Proc., Eighth International Conference on Computing in Civil and Building Engineering*, ASCE, CA, USA, 1293-1300.

Luck, S. J., and Vogel, E. K. (1997). "The Capacity of Visual Working Memory for Features and Conjunctions." *Nature*, 390, 279-281.

Lu, M., and Anson, M. (2004). "Establish Concrete Placing Rates using Quality Control Records from Hong Kong Building Construction Projects." *Journal of Construction Engineering and Management*, 130(2), 216-224.

Mackinlay, J. (1986). "Automating the Design of Graphical Presentations of Relational Information." *Transactions on Graphics*, 5(2), 110-141.

Mankoff, J., Dey, A. K., Hsieh, G., Kientz, J., Lederer, S., Ames, M. (2003). "Heuristic evaluation of ambient displays." *Proc., Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Ft. Lauderdale, Florida, USA, 169-176.

McKinney, K., and Fischer, M. (1998). "Generating, Evaluating and Visualizing Construction Schedules with CAD Tools." *Automation in Construction*, 7(6), 433-447.

Merwin, D. H., Vincow, M. A., Wickens, C. D. (1994). "Visual Analysis of Scientific Data: Comparison of 3D-Topographic, Color and Gray Scale Displays in a Feature Detection Task." *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 38, 240-244.

Meyer, M., Wong, B., Styczynski, M., Munzner, T., Pfister, H. (2010). "Pathline: A Tool for Comparative Functional Genomics." *Comput. Graphics Forum*, 29 (3), 1043-1052.

Meyer, M., Munzner, T., Pfister, H. (2009). "MizBee: A Multiscale Synteny Browser." *IEEE Trans. Visual. Comput. Graphics*, 15(6), 897-904.

Morrison, A., Ross, G., Chalmers, M. (2003). "Fast Multidimensional Scaling through Sampling,

Springs and Interpolation." *Information Visualization*, 2(1), 68-77.

Morton, S., Cook, D., Stuetzle, W., Buja, A. (1995). "Computer graphics in statistics: The last 30 years in brief." *ASA Statistical Graphics Video Lending Library*, USA.

Moselhi, O., Leonard, C., Fazio, P. (1991). "Impact of Change Orders on Construction Productivity." *Canadian Journal of Civil Engineering*, 18(3), 484-492.

Moselhi, O., Assem, I., El-Rayes, K. (2005). "Change Orders Impact on Labor Productivity." *J. Constr. Engrg. and Mgmt.*, 131(3), 354-359.

Motawa, I. A., Anumba, C. J., Lee, S., Peña-Mora, F. (2007). "An Integrated System for Change Management in Construction." *Automation in Construction*, 16(3), 368-377.

Munzner, T. (2009). "A Nested Process Model for Visualization Design and Validation." *IEEE Trans. Visual. Comput. Graphics*, 15(6), 921-928.

Nie, H., Staub-French, S., Froese, T. (2007). "OLAP-Integrated Project Cost Control and Manpower Analysis." *J. Comput. Civ. Eng.*, 21(3), 164-174.

Nielson, Y., and Erdogan, B. (2007). "Level of Visualization Support for Project Communication in the Turkish Construction Industry: A Quality Function Deployment Approach." *Canadian Journal of Civil Engineering*, 34(1), 19-36.

North, C. L. (2000). "A User Interface for Coordinating Visualization Based on Relational Schemata: Snap-Together Visualization." *Ph.D. Thesis, University of Maryland, College Park, Maryland, USA*.

North, C., and Shneiderman, B. (2000). "Snap-Together Visualization: Can Users Construct and Operate Coordinated Visualization?" *Int. J. Human-Computer Studies*, 53, 715-739.

O'Brien, J. J. (1965). *CPM in Construction Management*, 1st Ed., McGraw-Hill, Inc, USA.

Pich, M. T., Loch, C. H., Meyer, A. D. (2002). "On Uncertainty, Ambiguity, and Complexity in Project Management." *Manage.Sci.*, 48(8), 1008-1023.

- Pilgrim, M., Bouchlaghem, D., Loveday, D., Holmes, M. (2000). "Abstract data visualization in the built environment." *Proc., IV '00: International Conference on Information Visualisation*, IEEE Computer Society, London, England, 126-134.
- Pinker, S. (1990). "A Theory of Graph Comprehension." *Artificial Intelligence and the Future of Testing*, Lawrence Erlbaum Associate, Inc., Hillsdale, NJ, USA, 73-126.
- Pinnell, S. S. (1998). *How to Get Paid for Construction Changes*, McGraw-Hill, USA.
- Pirolli, P., and Rao, R. (1996). "Table lens as a tool for making sense of data." *Proc., AVI '96: Proceedings of the Workshop on Advanced Visual Interfaces*, ACM, Gubbio, Italy, 67-80.
- Plaisant, C., Grosjean, J., Bederson, B. B. (2002). "SpaceTree: Supporting exploration in large node link tree, design evolution and empirical evaluation." *Proc., Information Visualization, 2002.INFOVIS 2002.IEEE Symposium on*, IEEE Computer Society, Boston, Massachusetts, 57-64.
- Plaisant, C. (2004). "The challenge of information visualization evaluation." *Proc., AVI '04: Proceedings of the Working Conference on Advanced Visual Interfaces*, ACM, Gallipoli, Italy, 109-116.
- Plaisant, C., Grinstein, G., Scholtz, J., Whiting, M., O'Connell, T., Laskowski, S., Chien, L., Tat, A., Wright, W., Görg, C., Liu, Z., Parekh, N., Singhal, K., Stasko, J. (2008). "Evaluating Visual Analytics at the 2007 VAST Symposium Contest." *IEEE Comput. Graphics Appl.*, 28(2), 12-21.
- Playfair, W. (1821). *A Letter on our Agricultural Distresses, their Causes and Remedies*, 1st Ed., William Sams, London, England.
- Plumlee, M. D., and Ware, C. (2006). "Zooming Versus Multiple Window Interfaces: Cognitive Costs of Visual Comparisons." *ACM Trans.Comput.-Hum.Interact.*, 13(2), 179-209.
- Priestley, J. (1744). *A Description of a Chart of Biography*, Warrington, England.
- Puddicombe, M. S. (2006). "The Limitations of Planning: The Importance of Learning." *J.*

*Constr. Engrg. and Mgmt.*, 132(9), 949-955.

Rao, R., and Card, S. K. (1994). "The table lens: Merging graphical and symbolic representations in an interactive focus + context visualization for tabular information." *Proc., CHI '94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Boston, Massachusetts, USA, 318-322.

Roberts, J. C. (2007). "State of the art: Coordinated & multiple views in exploratory visualization." *Proc., CMV '07: Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization*, IEEE Computer Society, Zurich, Switzerland, 61-71.

Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D., van Dantzich, M. (1998). "Data mountain: using spatial memory for document management" *Proceedings of the 11th annual ACM symposium on User interface software and technology*, ACM, San Francisco, California, United States, 153-162.

Rojas, E. M., and Lee, N. (2007). "Visualization of project control data: A research agenda." *Proc., 2007 ASCE International Workshop on Computing in Civil Engineering*, ASCE, Pittsburgh, Pennsylvania, USA, 37-42.

Roth, S., and Hendrickson, C. (1991). "Computer-Generated Explanations in Project Management Systems." *J.Comp.in Civ.Engrg.*, 5(2), 231-244.

Russell, A. D. (1985). "Microcomputers, Management, and High-Rise Construction: The Next Step" *Can. J. Civ. Eng.*, 12(2), 396-414.

Russell, A. D., and Fayek, A. (1994). "Automated Corrective Action Selection Assistant." *J. Constr. Engrg. and Mgmt.*, 120(1), 11-33.

Russell, A., and Chevallier, N. (1998). "Representing a Project's Physical View in Support of Project Management Functions." *Can. J. Civ. Eng.*, 25(4), 705-717.

Russell, A. D., and Udaipurwala, A. (2000a). "Assessing the quality of a construction schedule."

*Proc., Construction Congress VI: Building Together for a Better Tomorrow in an Increasingly Complex World*, ASCE, Orlando, FL, USA, 928-937.

Russell, A. D., and Udaipurwala, A. (2000b). "Visual representation of project planning and control data." *Proc., Eighth International Conference on Computing in Civil and Building Engineering (2000)*, ASCE, Stanford, CA, USA, 542-549.

Russell, A. D., and Udaipurwala, A. (2002a). "Advantages of a multi-view representation of project execution." *Proc., Proceedings of Specialty Conference on Fully Integrated and Automated Project Processes*, Virginia Polytechnic and State University, Blacksburg, Virginia, USA, 364-375.

Russell, A. D., and Udaipurwala, A. (2002b). "Construction schedule visualization." *Proc., International Workshop on Information Technology in Civil Engineering 2002*, ASCE, Washington, D.C., USA, 167-178.

Russell, A. D., Udaipurwala, A., Robinson-Fayek, A., Pedrycz, W. (2004). "Interpreting data derived from a holistic representation of construction projects." *Proc., Proceedings, 1st International Conference, World of Construction Project Management*, CSCE, , 455-469.

Russell, A. D., and Udaipurwala, A. (2004). "Using multiple views to model construction." *Proc., CIB World Building Congress*, Toronto, 12 pages.

Russell, A., Staub-French, S., Tran, N., Wong, W. (2009a). "Visualizing High-Rise Building Construction Strategies using Linear Scheduling and 4D CAD." *Autom. Constr.*, 18(2), 219-236.

Russell, A. D., Chiu, C., Korde, T. (2009b). "Visual Representation of Construction Management Data." *Autom. Constr.*, 18(8), 1045-1062.

Saraiya, P., North, C., Duca, K. (2004). "An evaluation of microarray visualization tools for biological insight." *Proc., INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization*, IEEE Computer Society, Austin, TX, USA, 1-8.

Schmid, C. F. (1978). "The role of standards in graphic presentation." *Proc., 136th Annual*

*Meeting of the American Statistical Association*, Bureau of the Census, U.S. Department of Commerce, Boston, Massachusetts, USA, 69-78.

Schmid, C. F. (1983). *Statistical Graphics: Design Principles and Practices*, 1st Ed., John Wiley & Sons, USA.

Scholtz, J. (2006). "Beyond usability: Evaluation aspects of visual analytic environments." *Proc., Visual Analytics Science and Technology, 2006 IEEE Symposium on*, IEEE, Baltimore, MD, USA, 145-150.

Shaaban, S., Lockley, S., Elkadi, H. (2001). "Information visualisation for the architectural practice." *Proc., Fifth International Conference on Information Visualisation (2001)*, IEEE Computer Society, London, UK, 43-50.

Shah, P. (2001). "Graph Comprehension: The Role of Format, Content, and Individual Differences." *Diagrammatic Representation and Reasoning*, Springer-Verlag New York, Inc, Secaucus, NJ, USA, 173-186.

Shah, P., and Hoeffner, J. (2002). "Review of Graph Comprehension Research: Implications for Instruction." *Educational Psychology Review*, 14, 47-69(23).

Shah, P. (2005). "The Comprehension of Quantitative Information in Graphical Displays." *The Cambridge Handbook of Visuospatial Thinking*, Cambridge University Press, New York, NY, USA, 426-476.

Shirole, A. M., Chen, S. S., Pucket, J. A. (2008). "Bridge information modeling for the life cycle progress and challenges." *Proc., 10th International Conference on Bridge and Structure Management*, Transportation Research Board, Buffalo, New York, USA, 313-323.

Shneiderman, B. (1994). "Dynamic Queries for Visual Information Seeking." *IEEE Software*, 11(6), 70-77.

Shneiderman, B. (1996). "The eyes have it: A task by data type taxonomy for information visualizations." *Proc., 1996 IEEE Symposium on Visual Languages*, IEEE Computer Society,

Boulder, CO, USA, 336-343.

Simkin, D., and Hastie, R. (1987). "An Information-Processing Analysis of Graph Perception." *Journal of the American Statistical Association*, 82(398), 454-465.

Smallman, H. S., John, M. S., Oonk, H. M., Cowen, M. B. (2001). "Information Availability in 2D and 3D Displays." *IEEE Comput. Graph. Appl.*, 21(5), 51-57.

Software FX Inc., . "Chart FX for COM." <http://www.softwarefx.com/sfxComProducts/cfxForCOM/> (August/23, 2009).

Soibelman, L., and Kim, H. (2002). "Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases." *J.Comp.in Civ.Engrg.*, 16(1), 39-48.

Song, K., Pollalis, S. N., Pena-Mora, F. (2005). "Project dashboard: Concurrent visual representation method of project metrics on 3D building models." *Proc., 2005 ASCE International Conference on Computing in Civil Engineering*, ASCE, Cancun, Mexico, 12 pages.

Songer, A. D., Hays, B., North, C. (2004). "Multidimensional Visualization of Project Control Data." *Construction Innovation: Information, Process, Management*, 4(3), 173-190.

Sriprasert, E., and Dawood, N. (2003). "Multi-Constraint Information Management and Visualisation for Collaborative Planning and Control in Construction." *ITCon.*, 8(Special (eWork and eBusiness)), 341-366.

Staub, S., and Fischer, M. (1998). "Constructability reasoning based on a 4D facility model." *Proc., Structural Engineering World Congress (Structural Engineering World Wide)*, Elsevier Science Ltd., San Francisco, CA., USA.

Staub-French, S., Fischer, M., Kunz, J., Paulson, B. (2003). "A Generic Feature-Driven Activity-Based Cost Estimation Process." *Advanced Engineering Informatics*, 17(1), 23-39.

Stolte, C., Tang, D., Hanrahan, P. (2002). "Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases." *Transactions on Visualization and*

*Computer Graphics*, 8(1), 52-65.

Stolte, C., Tang, D., Hanrahan, P. (2003). "Multiscale Visualization using Data Cubes." *IEEE Trans. Visual. Comput. Graphics*, 9(2), 176-187.

The National Building Agency. (1968). *Programming House Building by Line of Balance*, 1st Ed., The National Building Agency, London, England.

Theus, M., and Urbanek, S. (2008). *Interactive Graphics for Data Analysis: Principles and Examples*, 1st Ed., Chapman & Hall, USA.

Thomas, H. R., and Napolitan, C. L. (1995). "Quantitative Effects of Construction Changes on Labor Productivity." *J. Constr. Engrg. and Mgmt.*, 121(3), 290-296.

Thomas, J. J., and Cook, K. A. (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, 1st Ed., IEEE Computer Society Press, Los Alamitos, CA, USA.

Trafton, J. G., Kirschenbaum, S. S., Tsu, T. L., Miyamoto, R. T., Ballas, J. A., Raymond, P. D. (2000). "Turning Pictures into Numbers: Extracting and Generating Information from Complex Visualizations." *Int. J. Hum. -Comput. Stud.*, 53(5), 827-850.

Tufte, E. R. (1986). *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT, USA.

Tufte, E. (1990). *Envisioning Information*, Graphics Press, Cheshire, CT, USA.

Tullett, A. D. (1996). "The Thinking Style of the Managers of Multiple Projects: Implications for Problem Solving when Managing Change." *International Journal of Project Management*, 14(5), 281-287.

U.S. General Service Administration. (2008). "3D-4D building Information Modeling." <<http://www.gsa.gov/Portal/gsa/ep/channelView.do?pageTypeId=8195&channelPage=%2Fep%2Fchannel%2FgsaOverview.jsp&channelId=-18161>> (August/5, 2008).

Unwin, A. (2006). "Interacting with Graphics." *Graphics of Large Datasets: Visualizing a*

Million, Springer, USA, 73-101.

Unwin, A. (2008). "Good Graphics?" *Handbook of Data Visualization*, Springer, USA, 57-78.

Vico Software, I. (2010). "Vico control 2009." <<http://www.vicosoftware.com/products/Vico-Control/tabid/84573/Default.aspx>> (September/27, 2010).

Vrijland, M. S. A. (2003). "Visual Display of Sensitivity and Risk." *AACE International Transactions*, 2003, RISK.11.1-RISK.11.8.

Vrotsou, K., Ynnerman, A., Cooper, M. (2008). "Seeing beyond statistics: Visual exploration of productivity on a construction site." *Proc., 2008 International Conference in Visualization*, IEEE Computer Society, London, UK, 37-42.

Wainer, H. (1997). *Visual Revelations*, Copernicus Springer-Verlag, USA.

Wang Baldonado, M. Q., Woodruff, A., Kuchinsky, A. (2000). "Guidelines for using multiple views in information visualization." *Proc., AVI '00: Proceedings of the Working Conference on Advanced Visual Interfaces*, ACM, Palermo, Italy, 110-119.

Ware, C. (2004). *Information Visualization: Perception for Design*, 2nd Ed., Morgan Kaufman, San Francisco.

Washington State Department of Transportation. (2010). "Cost risk assessment -sample report." <<http://www.wsdot.wa.gov/Projects/ProjectMgmt/RiskAssessment/>> (September/27, 2010).

Weaver, C., Fyfe, D., Robinson, A., Holdsworth, D., Peuquet, D., MacEachren, A. M. (2006). "Visual analysis of historic hotel visitation patterns." *Proc., Visual Analytics Science and Technology, 2006 IEEE Symposium on*, Baltimore, MD, USA, 35-42.

Wickens, C. D., and Carswell, C. M. (1995). "The Proximity Compatibility Principle: Its Psychological Foundation and Relevance to Display Design." *Human Factors*, 37(3), 473-494.

Wijk, J. J. V., and Selow, E. R. V. (1999). "Cluster and calendar based visualization of time series data." *Proc., IEEE Symposium on Information Visualization (INFOVIS'99)*, IEEE

Computer Society, San Francisco, CA, USA, 4-9.

Wilkinson, L. (1999). *The Grammar of Graphics*, Springer-Verlag New York, Inc, New York, NY, USA.

Wills, G. (1996). "Selection: 524,288 ways to say "this is interesting"." *Proc., Proceeding of IEEE Symposium on Information Visualization*, IEEE Computer Society, San Francisco, CA, USA, 54-63.

Yi, J. S., Kang, Y. A., Stasko, J., Jacko, J. (2007). "Toward a Deeper Understanding of the Role of Interaction in Information Visualization." *IEEE Trans. Visual. Comput. Graphics*, 13(6), 1224-1231.

Zacks, J., and Tversky, B. (1999). "Bars and Lines: A Study of Graphic Communication." *Memory and Cognition*, 27(6), 1073-1079.

Zeb, J., Chiu, C., Russell, A. (2008). "Designing a construction data visualization environment." *Proc., Proceedings of the 1st Forum on Construction Innovation*, Canadian Society for Civil Engineering, Quebec City, Quebec, Canada, 11 pages.

Zhang, C., Zayed, T., Hammad, A. (2008). "Resource Management of Bridge Deck Rehabilitation: Jacques Cartier Bridge Case Study." *J. Constr. Engrg. and Mgmt.*, 134(5), 311-319.

Zhang, J. (1996). "A Representational Analysis of Relational Information Displays." *Int. J. Hum.-Comput. Stud.*, 45(1), 59-74.

Zhang, X., Bakis, N., Lukins, T. C., Ibrahim, Y. M., Wu, S., Kagioglou, M., Aouad, G., Kaka, A. P., Trucco, E. (2009). "Automating Progress Measurement of Construction Projects." *Autom. Constr.*, 18(3), 294-301.

Zuk, T., Schlesier, L., Neumann, P., Hancock, M. S., Carpendale, S. (2006). "Heuristics for information visualization evaluation." *Proc., Proceedings of the 2006 AVI Workshop on BEyond*

*Time and Errors: Novel Evaluation Methods for Information Visualization*, ACM, Venice, Italy,  
1-6.

## **Appendices**

In Appendix A, a full literature review of the use of data visualization in construction management (including the data visualization capabilities of current commercial CM information systems) is described. In Appendix B, an overview of state-of-the-art data visualization technologies is provided in order to understand the current state of applying data visualization technologies to facilitate CM visual analytics in support of CM functions. Where appropriate, earlier work by authors is cited in order to emphasize the longevity of some visual images.

### **Appendix A Data Visualization in Construction Management**

#### **A.1 Visual representations of planned/baseline construction conditions**

The commonly used visual representations of CM data depicting planned/baseline construction conditions correspond to graphics for visualizing important inputs to scheduling models (e.g. CPM) such as:

- Resource allocation (e.g. Fig. A.1),
- Activity sequencing (e.g. Fig. A.2)
- Temporal and spatial distribution of activities (e.g. Fig. A.3).

Another major type of condition relates to product information model. Currently, the geometric aspects of product information are encoded mainly in visual forms as virtual 2D/3D models. Other than these conditions, many construction conditions are implicitly assumed by schedulers or cost estimators based on their experience, their referencing historical data, or even their gut feelings when they work on construction plans. These conditions include environment conditions, organizational capability conditions (e.g.: productivity, skill, experience), etc. Since these conditions are rarely recorded explicitly as CM data, rich visual representations presenting them for other project participants to understand planning/predicting assumptions are almost non-existent.

Figure A.1 has been removed due to copyright restrictions. It was a graphics visualizing planned resource allocation.  
Original source: O'Brien, J. J. (1965). CPM in Construction Management 1st Ed: Fig. 15.1., pp 186

Figure A.2 has been removed due to copyright restrictions. It was a graphics visualizing activity sequencing.  
Original source: de Leon, G. P. (2008). Project Planning using Logic Diagramming Method. AACE International Transactions: Figure 1, pp. PS.S05.03

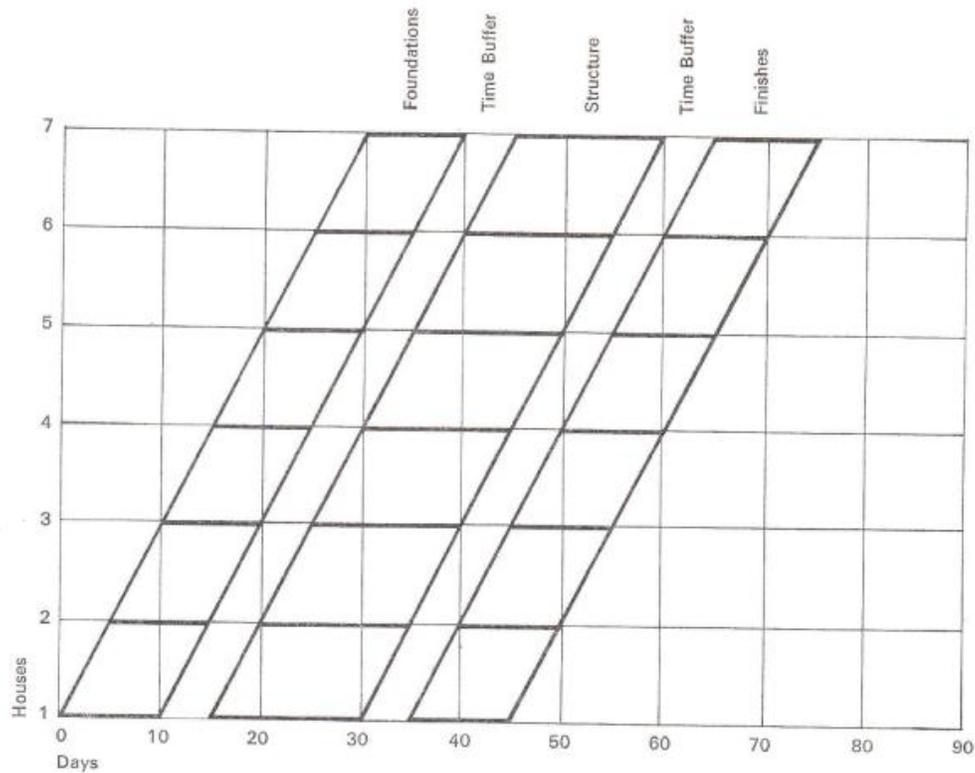


Figure A.3 Visualizing temporal and spatial distribution of activities. Source: (The National Building Agency 1968)

## A.2 Visual representations of predicted/ baseline construction performance

*Time performance:* The most commonly used visual representations of CM data that describe construction time performance is in the form of "schedule" graphics. In the past, a variety of visual forms of construction schedules were proposed. Some of them include:

- Bar graph (e.g. Fig. A.4),
- Network diagram--activities on arrows (e.g. Fig. A.5),
- Network diagram--activities on nodes (e.g. Fig. A.6).

All of these visual representations allows users to examine at least three construction time performance dimensions in terms of activity start dates, activity finish dates, and activity duration. Users are able to examine these performance dimensions from the project aspects of locations, activities (or work breakdown structures), and time. Several researchers integrated the aforementioned visual representations with 3D product information models and their corresponding 3D CAD so that users can further examine construction time performance from the product aspect of the project (4D model).

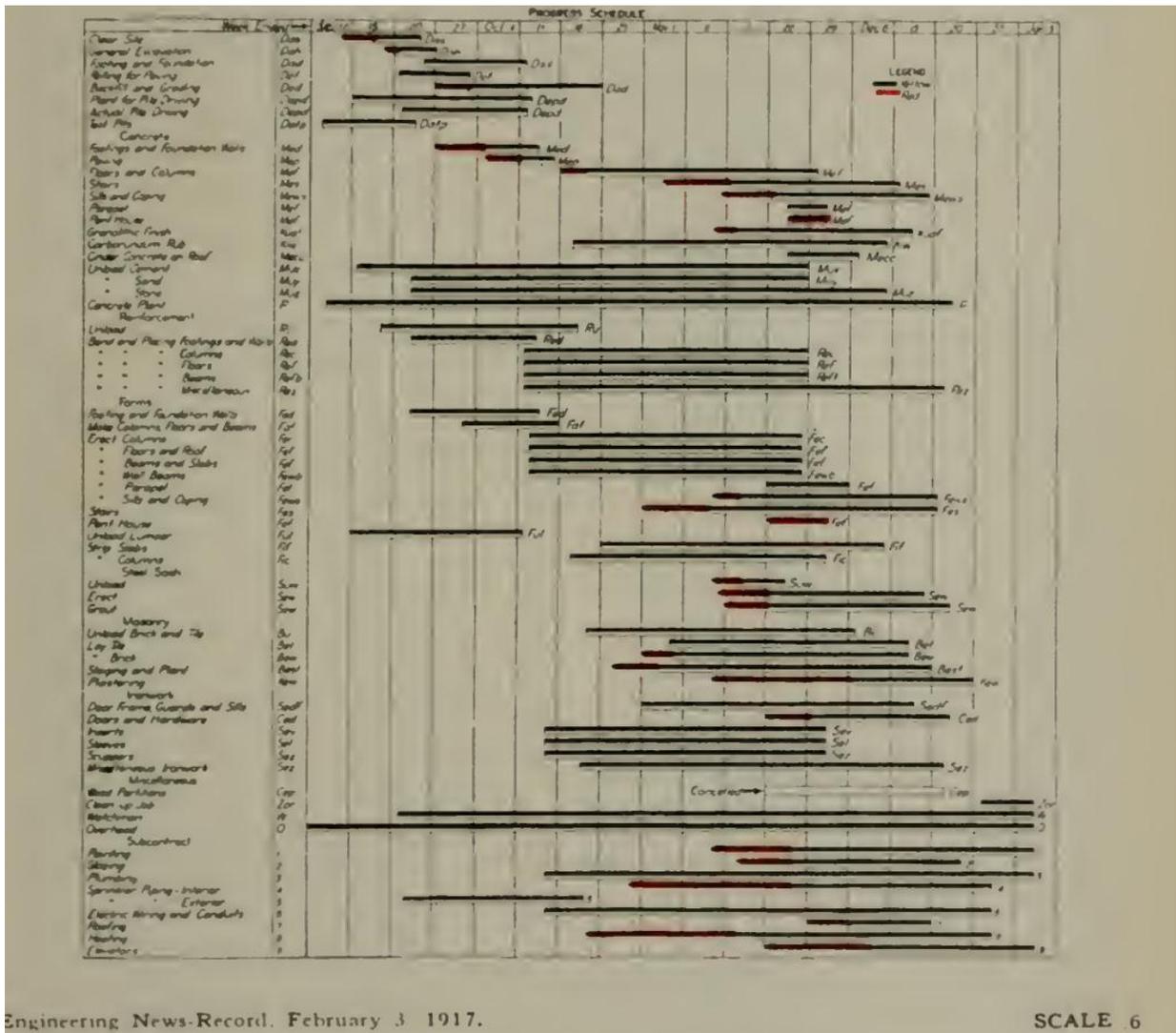


Figure A.4 A construction schedule in bar graph format seen as early as 1917. Source: (Brinton 1939)



*Cost performance:* The commonly used visual representations of CM data to describe construction cost performance are the S curve (e.g. Fig. A.7) and the cash flow diagram (e.g. Fig. A.8). They are usually used for tracking the actual cost performance by comparing the planned and actual S curve/cash flow diagram.

Figure A.7 has been removed due to copyright restrictions. It was a graphics visualizing S curve. Original source: O'Brien, J. J. (1965). CPM in Construction Management 1st Ed: Fig. 14.5., pp 168

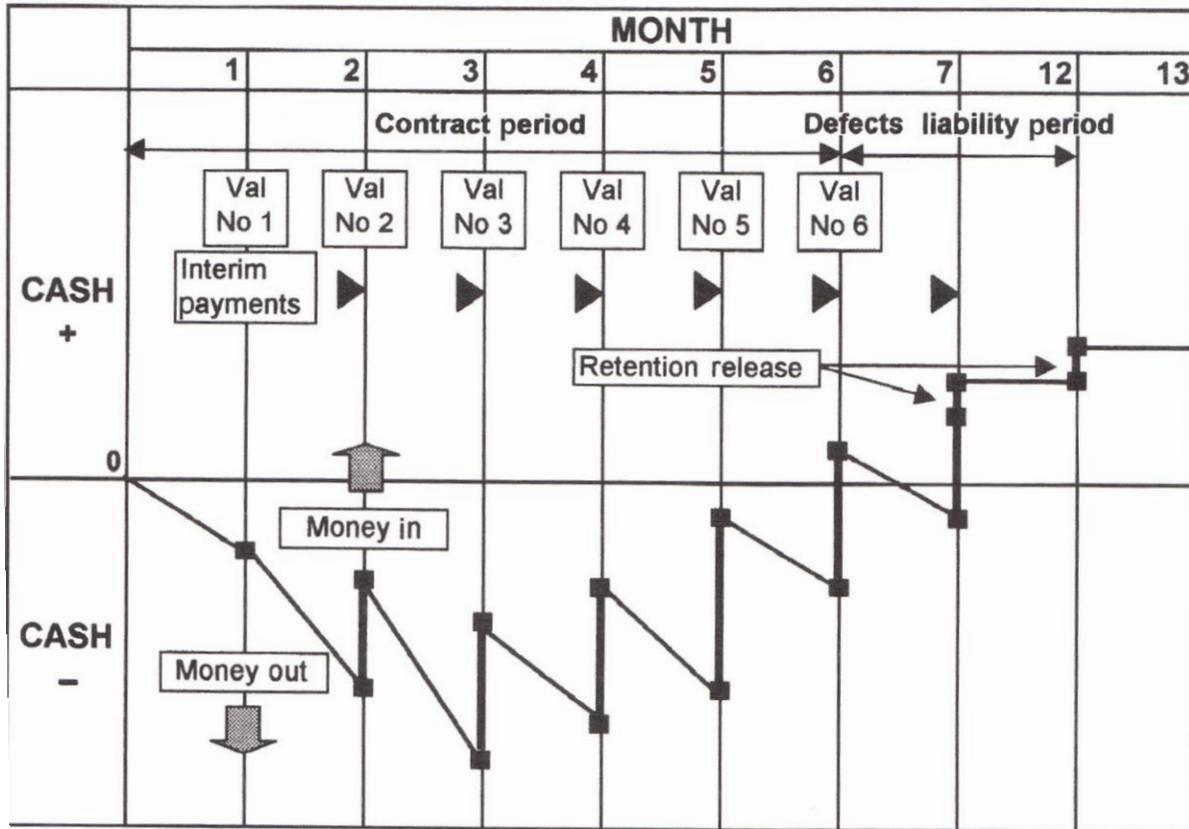


Figure A.8 Cash flow diagram. Source: (Cooke and Williams 2004)

*Quality performance:* Commonly used visual representations of CM data for depicting quality performance are almost non-existent. Planned/baseline quality performance is part of the quality assurance specifications that are usually presented in textual formats describing the product quality requirements (i.e. planned quality performance), construction method requirements for achieving the product quality requirements, and inspection requirements for assuring compliance.

### **A.3 Visual representations of how changes of planned/baseline conditions affect predicted/baseline performance-optimizing construction plans**

The application of exploring the consequences of different values of construction conditions (i.e.: simulating changes of conditions) on a performance prediction model in order to assess the change in construction performances is known as the optimization or simulation technique. The main purpose of doing so is to identify a deterministic construction plan that can achieve optimal/better construction performance. There was no unique image for visualizing how

changes of conditions affect performance. One way is to juxtapose or superimpose images depicting different construction conditions and/or images presenting related construction performances in response to the various construction conditions for easy comparison. For example, in Figure A.9, Figure A.9(A) and Figure A.9(B) at the left hand side represent two resource allocation scenarios (scenario a and scenario b), and Figure A.9(A') and Figure A.9(B') at the right hand side represent the schedules resulting from those two scenarios. By viewing Figure A.9, project participants quickly realize that scenario B requires resources earlier in the middle of construction than scenario A if the project owner requests an earlier finish date. Another way to show how change of performance may be a function of change in conditions is to use a single image. As seen in Figure A.10, each curved line in the line chart represents the cash flow corresponding to the contractors' desired marginal profit which ranges between 0% to 15%. Another example is seen in Figure A.11 where each data point represents the correspondence between the planned resource combination (T: team, S: saw, O: truck transporting old bridge panels, N: truck transporting new bridge panels) and the resultant unit cost (Figure A.11(a)) and productivity (Figure A.11(b)) in a paneled bridge rehabilitation project. A particular way of visualization is to devise a chart similar to a Psychrometric Chart for representing the prediction model itself so users can directly look up the crew size required given a certain combination of conditions such as CPM duration, project deadline, number of sites, etc. (Fig. A.12).

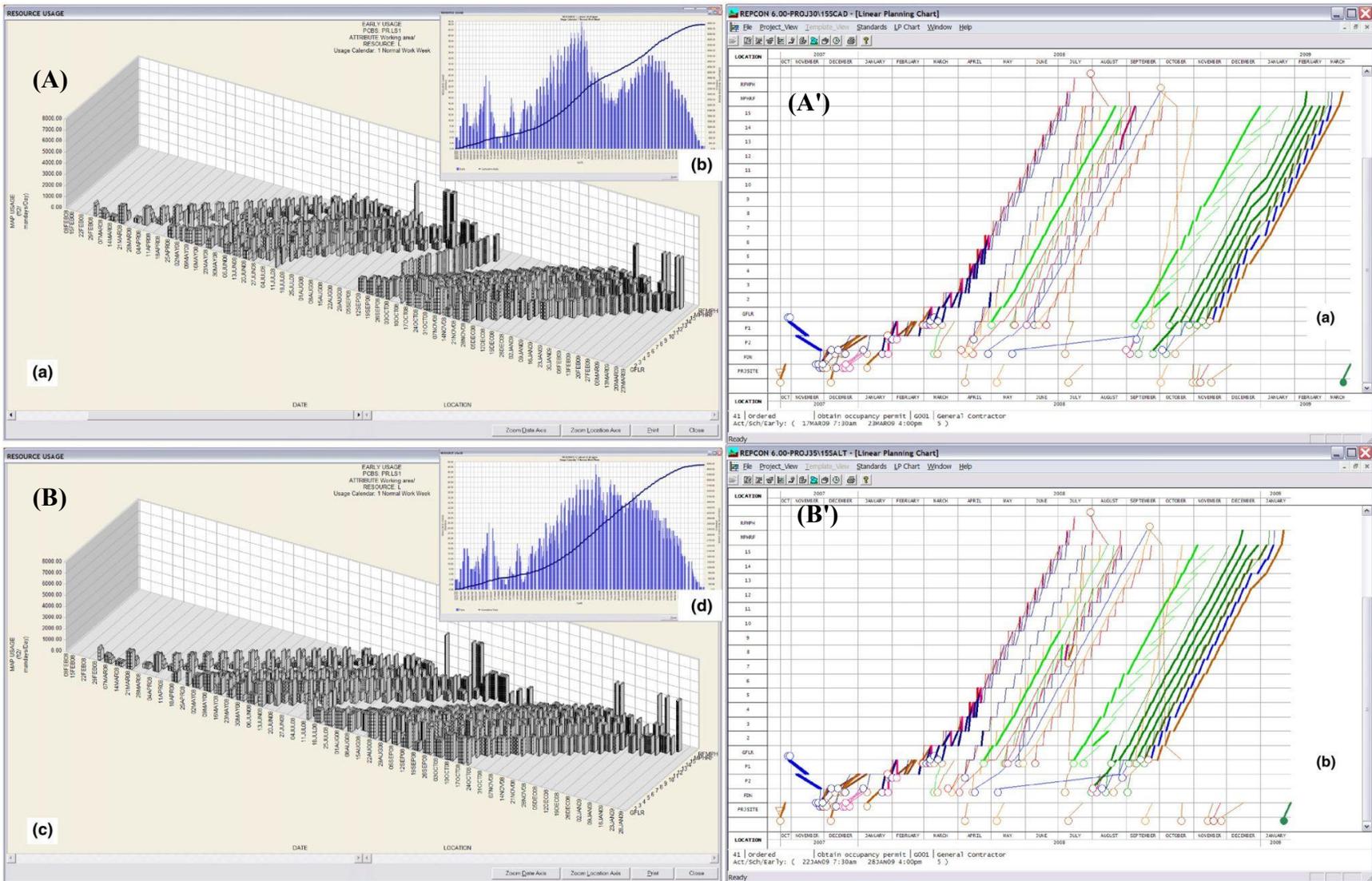


Figure A.9 Changing planned construction conditions (*resources*) vs. changed forecast performance (*time*) in multiple views. Source: (Russell et al. 2009)

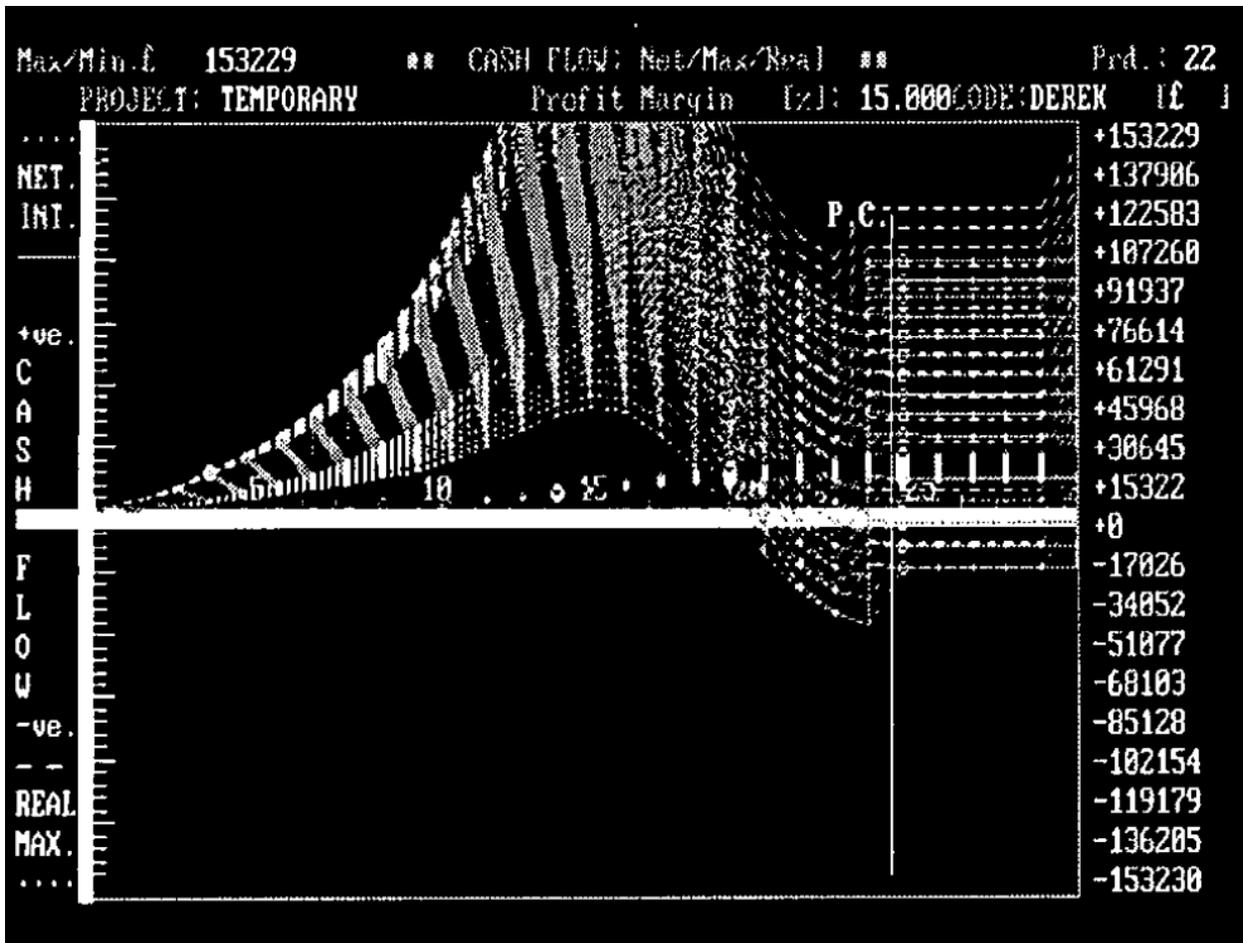


Figure A.10 Changing planned construction conditions (*profit margin desired*) vs. changed forecast performance (*cash flow*) in a single view. (© 2000 IEEE. Reprinted, with permission, from Khosrowshahi, Information Visualization in aid of Construction Project Cash Flow Management, Proceedings of the International Conference on Information Visualisation, 2000)

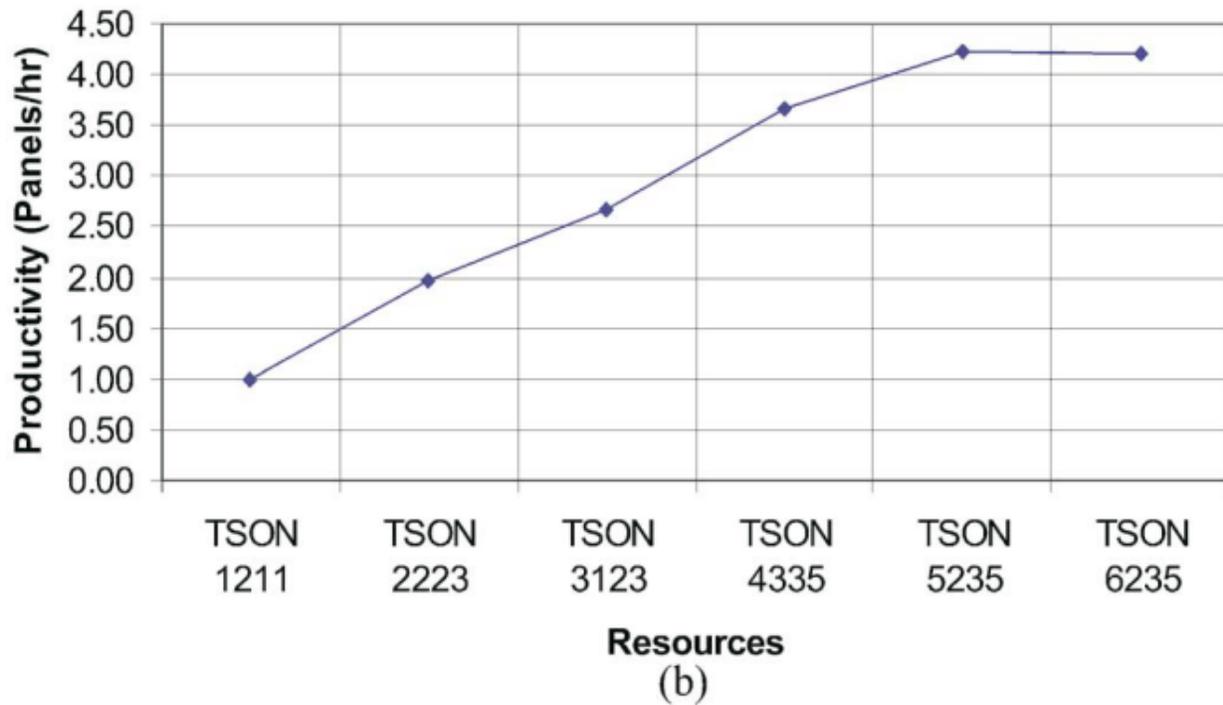
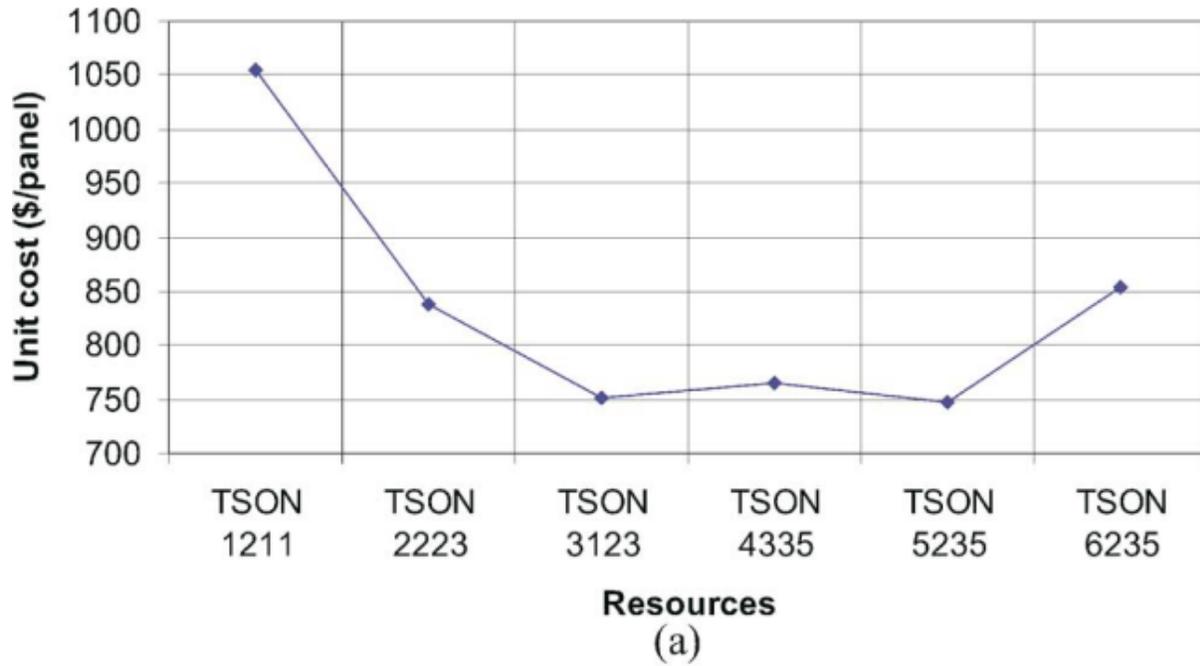


Figure A.11 Changing planned construction conditions (*resources*) vs. changed forecast performance (*unit cost/productivity*) in a single view. T = team; S = saw; O = trucking old panels, N = trucking new panels. Source: (Zhang et al. 2008)

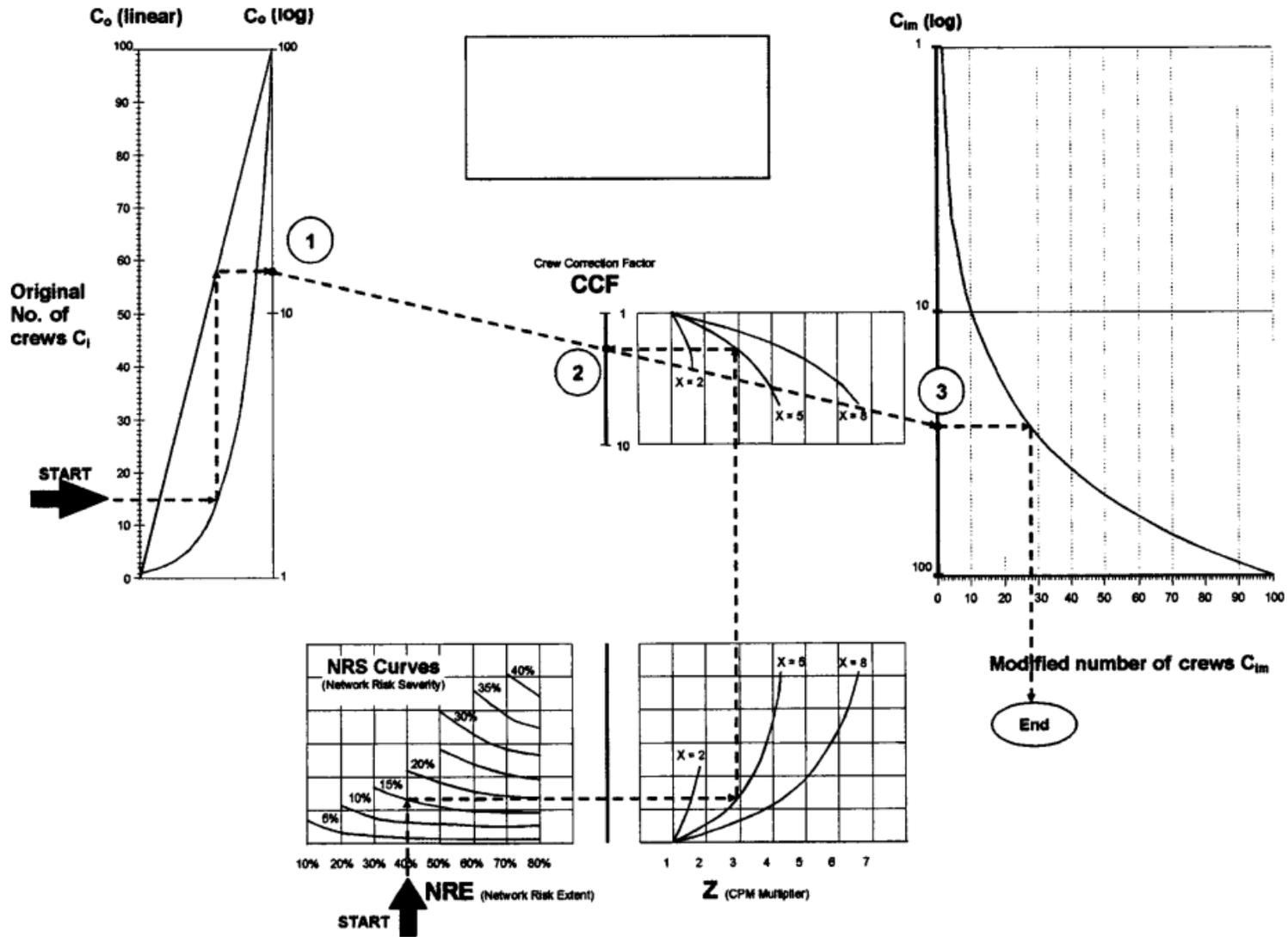


Figure A.12 A nomograph that encodes a mathematical model predicting required crew size. The model considers factors such as CPM duration, project deadline, number of sites for a repetitive activity. Source: (Elhakeem and Hegazy 2005)

#### **A.4 Visual representations of how changes of planned/baseline conditions affect predicted/baseline performance- identifying and analyzing construction risks**

Although varying as many inputs of a prediction model (i.e. construction condition items and the values thereof) as possible to see which ones lead to optimal construction performance can help identify an optimal construction plan, it does not take uncertainty and complexity into consideration. There is a need to identify construction condition items that would possibly have undesirable values thereby possibly impacting “normal” or “optimized” construction performance. A re-planning process may follow for eliminating, avoiding, mitigating, or accepting those identified undesirable conditions and attendant performance. The investigation of potential undesirable construction condition items and associated values, and their impact on “normal” construction performance with probabilistic treatments is in fact the management function of risk identification and analysis, a key component of the risk management process.

One application of visualization in risk management is for identifying risk items (i.e. risk identification). Image types such as a cause & effect diagram and influence diagram (Figure A.13) visualize details of risk entities in terms of how undesirable construction conditions (i.e. risk items) may affect other construction conditions and ultimately impact construction performance. Risk distributions graphics such as the one shown in Figure A.14 visualize collections of risk entities and help identify where and when the risk items come from.

With the identification of risk items, further quantitative analysis of them can be conducted if there is a mathematical model that links the independent variables with the dependent ones. The commonly used visual representations of CM data with respect to showing results of risk analysis are those visualizing how value changes of independent variables (e.g. construction conditions) impact values of dependent variables (e.g. construction performance such as net present value, finish dates). This type of risk analysis is also termed sensitivity analysis. Exemplary visual representations include the tornado diagram (Figure A.15), the spider plot (Figure A.16), and the radar plot (Figure A.17). Other frequently used images focus on visualizing probability and values of dependent variables (probability distribution diagram; Figure A.18) or probability/impact of independent variables (probability/impact matrix; Figure A.19).

Tabularizing qualitative and quantitative properties of risks item by item is a common way of presenting the risk identified in a risk register (Figure A.20).

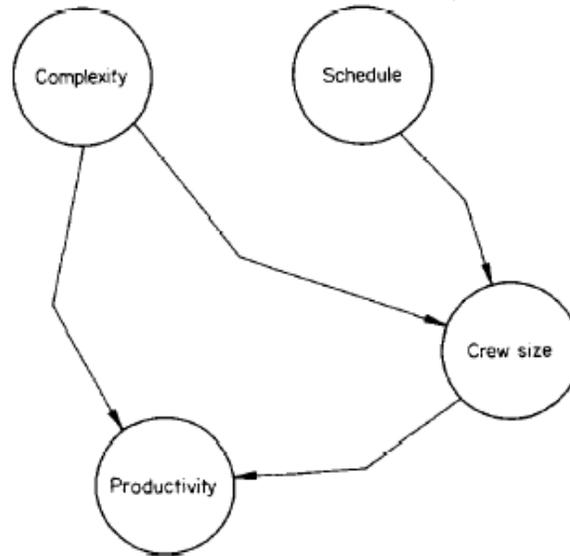


Figure A.13 A four variable influence diagram representing uncertain relationships between them. Source: (Diekmann 1992)

Figure A.14 has been removed due to copyright restrictions. It was a graphics visualizing distribution in time and space of risk from project participants. Original source: Korde, T., Wang, Y., Russell, A. D. (2005). Visualization of Construction Data. Proceedings of 6th Construction Specialty Conference: Figure 3, pp. CT-148-6

Figure A.15 has been removed due to copyright restrictions. It was a Tornado Plot visualizing how negative/positive 10% change in independent variables affects the net present value. Original source: Vrijland, M. S. A. (2003). Visual Display of Sensitivity and Risk. AACE International Transactions: Figure 2, pp. RISK 11.4

Figure A.16 has been removed due to copyright restrictions. It was a Spider Plot visualizing how negative and positive percentage changes in various independent variables affect net present value. Original source: Vrijland, M. S. A. (2003). Visual Display of Sensitivity and Risk. AACE International Transactions: Figure 7, pp. RISK 11.7

Figure A.17 has been removed due to copyright restrictions. It was a Radar plot showing sensitivity scores for the independent variables. Original source: Vrijland, M. S. A. (2003). Visual Display of Sensitivity and Risk. AACE International Transactions: Figure 8, pp. RISK 11.8

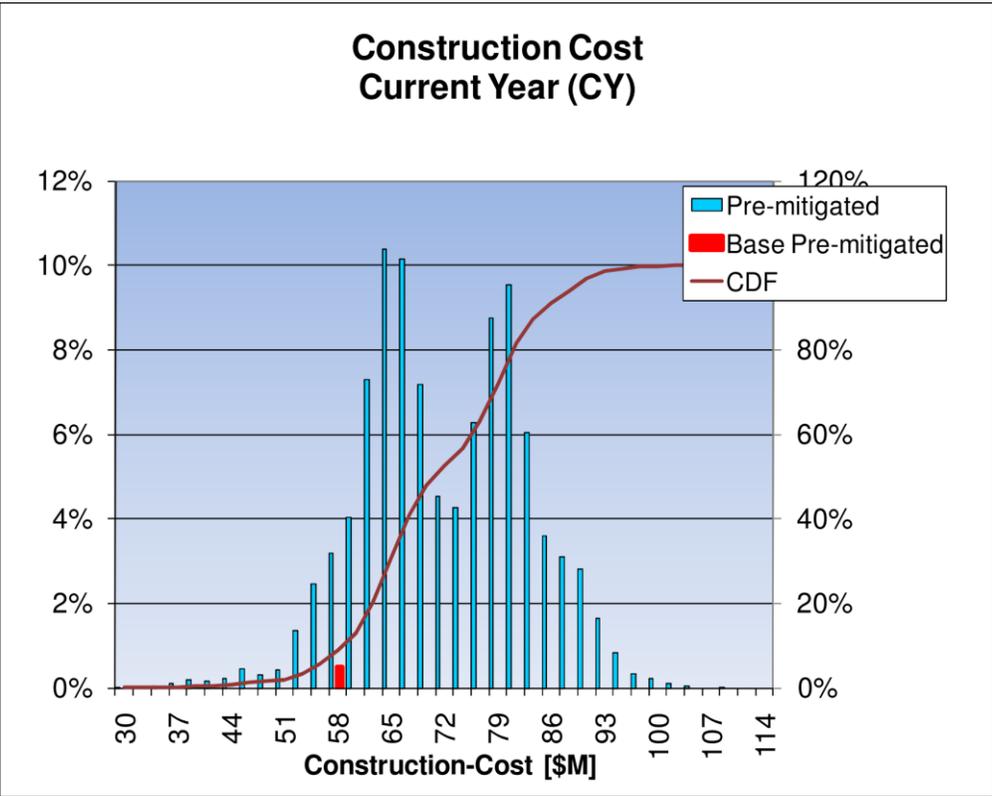
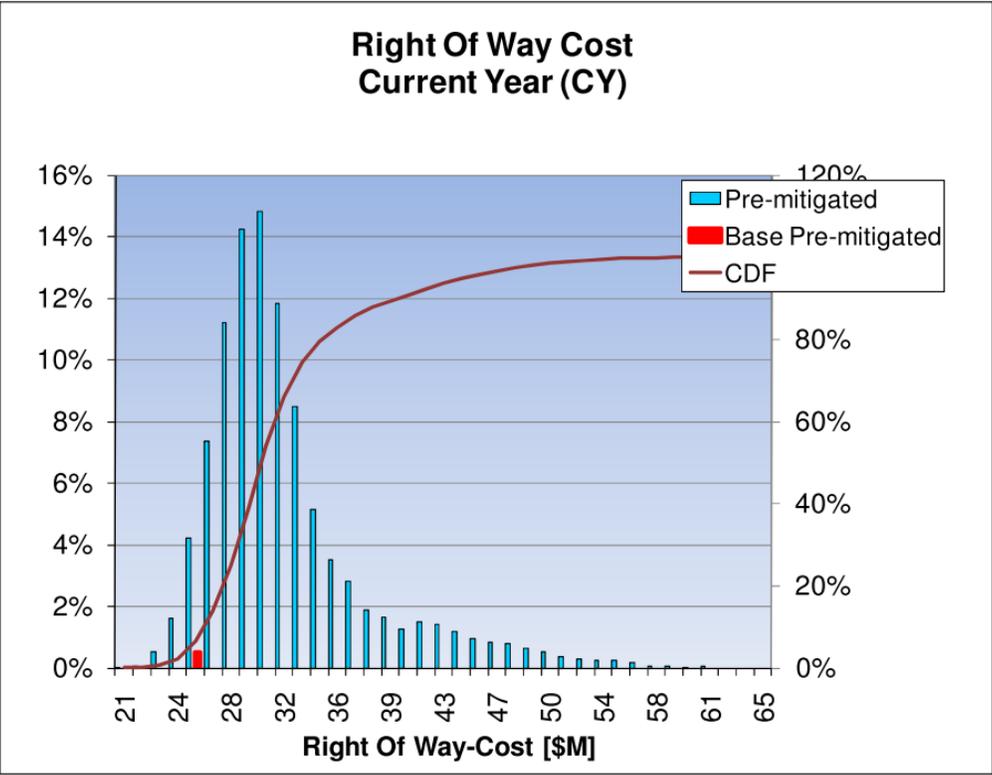


Figure A.18 Probability distribution (includes cumulative probability distribution) for right of way cost and construction cost. Source: (Washington State Department of Transportation 2010)

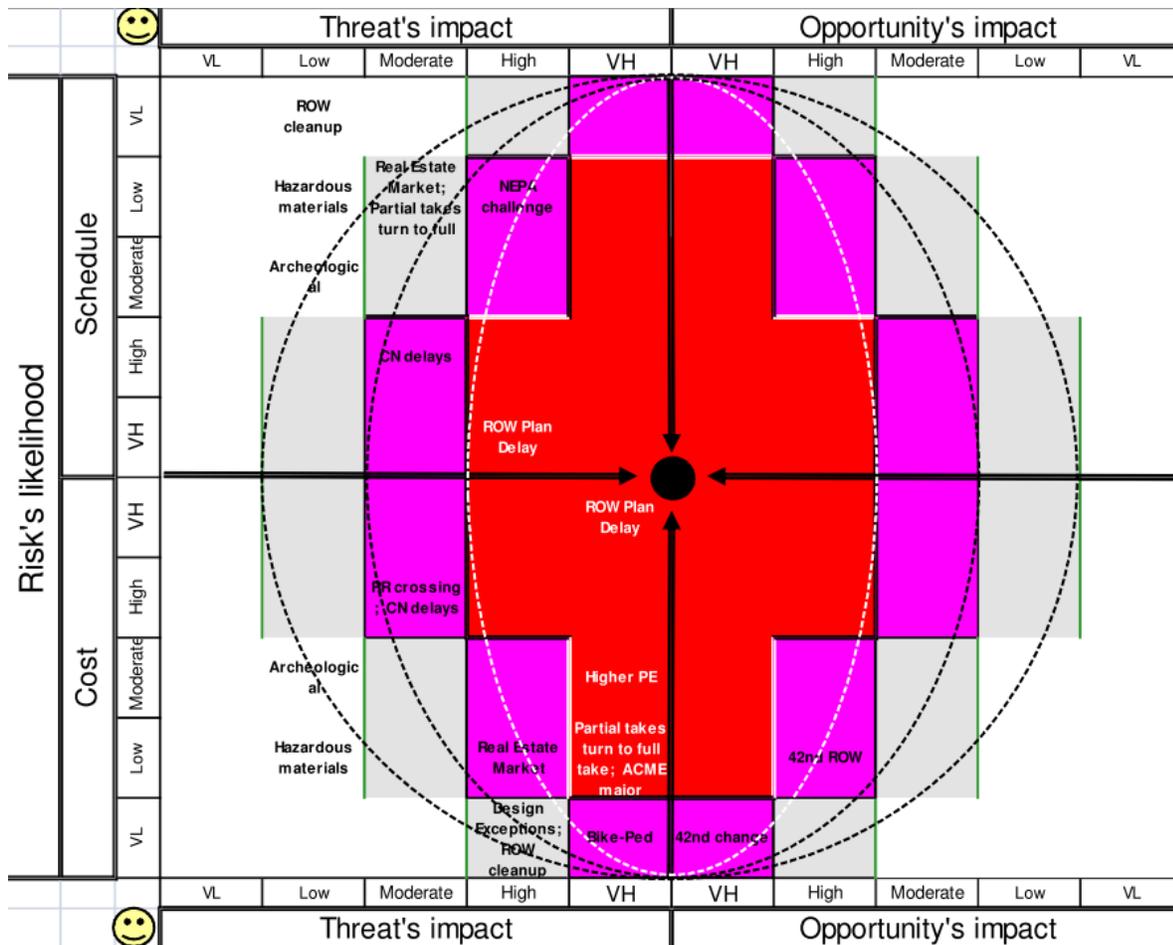


Figure A.19 Probability impact matrix showing likelihood and degree of schedule/cost consequence of risk events associated with a highway project. Source: (Washington State Department of Transportation 2010)

Risk Identification					Quantitative Analysis					Qualitative Display of the Best Guess Impact					
1	Active	ROW	Threat	Scheduling is a major risk as Right of Way and Utility relocations are dependent on other units to complete their design work prior to right of way beginning. This can be measured by the slippage in the date R/W and utility design/accommodation can begin their process.	Des Doc Pkg	Cost	95%	MIN	0.00\$M	3.17\$M	Very High	Very High	Probability	VH	
			MAX					8.00\$M	H						
			Most Likely					3.00\$M	M						
			Schedule			0	Master Duration Risk		11.4Mo	High		L			
							MIN	0.0Mo				VL			
							MAX	24.0Mo				L			
Most Likely	12.0Mo	M													
Impact															
2	Active	DEP-INCL Construction	Threat	Design exception required for reduced spirals, horizontal curves, and super transitions, shoulder width on the new mainline at the west transition to existing structure. There is a risk that the Design Exceptions will not be approved.	DE is reviewed and denied by Region and Tech Services	Cost	10%	MIN	0.50\$M	0.21\$M	Very Low	High	Probability	VH	
			MAX					4.00\$M	H						
			Most Likely					2.00\$M	M						
			Schedule			0	MIN		0.0Mo	NO RISK		L			
							MAX					VL			
							Most Likely					L			
Impact															

Figure A.20 Tabularized risk register of a highway project. Source: (Washington State Department of Transportation 2010)

## A.5 Visual representations of actual or actual vs. planned/baseline construction conditions

Using visual representations of CM data to make comparisons of planned/baseline and actual construction conditions is not common. Some sporadic research efforts used different visualization strategies to present examples of such comparisons, which include the use of conventional 2D or 3D charts to show construction conditions such as problems encountered (Fig. A.21), resources allocations (e.g. Fig. A.22), change orders (e.g. Fig. A.23) and superimposition of as-built construction conditions causing performance deviations onto products' 3D representations of the project itself., etc.

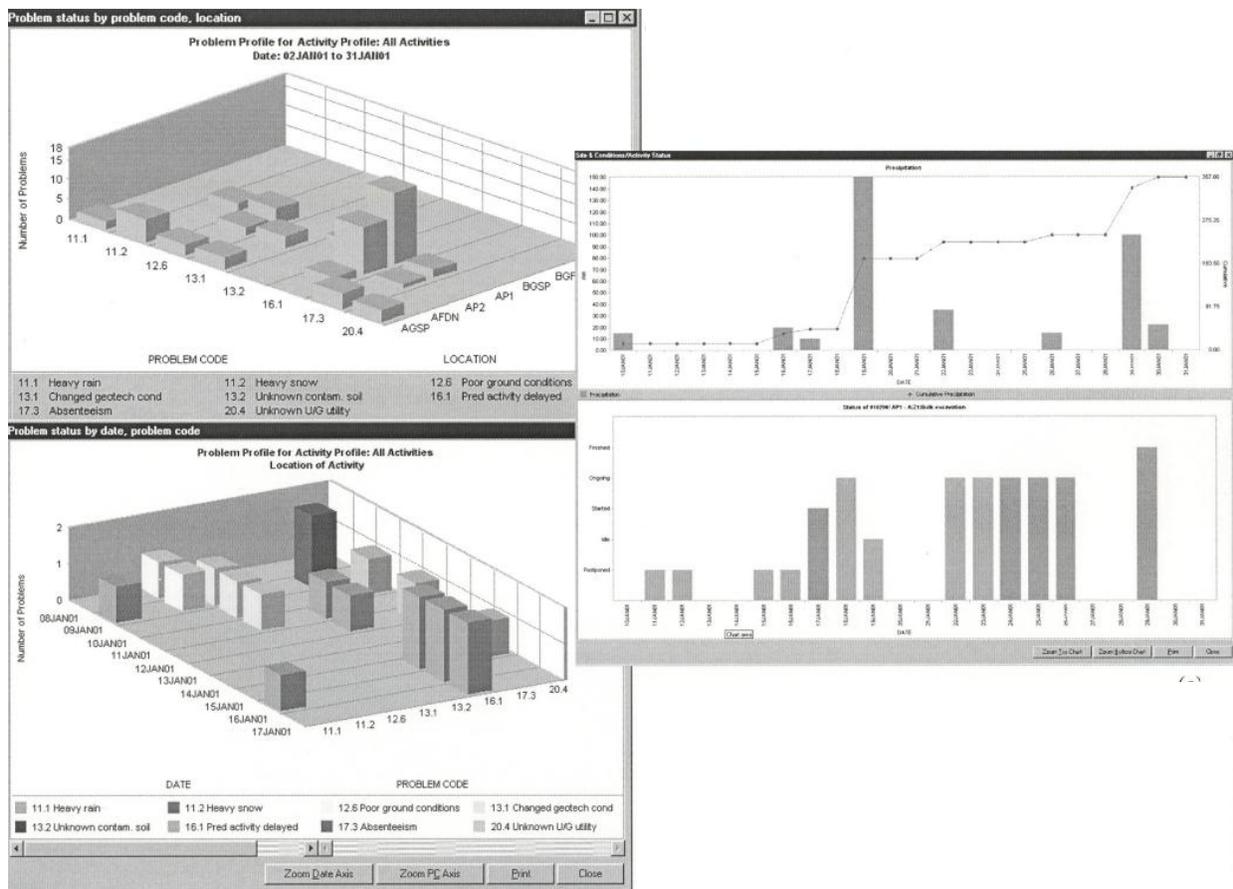


Figure A.21 Representing as-built data of problems and site conditions. Source: (Russell and Udaipurwala 2002a) (with Permission from ASCE)

Figure A.22 has been removed due to copyright restrictions. It was a graphics visualizing actual crew size and planned vs. actual crew size. Original source: Pinnell, S. S. (1998). *How to Get Paid for Construction Changes: Figures 10.38 and 10.39*, pp. 330

Label	Meaning
A	Improvement resulting from subcontract design
B	Cost saving measures
C	New information on existing site conditions
D	Employer has changed his requirements
E	Forced upon project from shop drawing coordination
F	Programme advantage or assurance
G	Statutory body requirement came to light since placing of trade contract
H	Public utility requirement came to light since placing of trade contract
I	Designers omission in tender documents
J	Coordination defects in tender documents
K	Management contractor omission from packaging
DD	Design development
L	Other
Empty	Not recorded

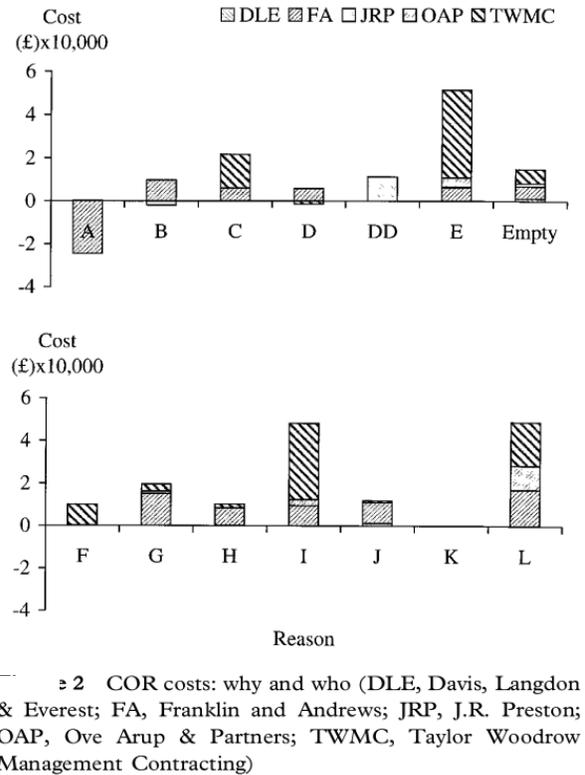


Figure A.23 "Causes and who" responsible for change order costs. Source: (Cox et al. 1999)

## A.6 Visual representations of actual or actual vs. predicted/baseline construction performance

*Time performance:* Frequently seen visual representations of CM data depicting a comparison between planned/baseline and actual construction time performance involve the superimposition of a planned/baseline schedule with an actual schedule, portrayed in the form of bar charts or linear planning charts. Other formats include the visual representations used to facilitate various schedule delay analysis techniques (e.g. Fig. A.24). Another visual representation for monitoring time performance can be found in the cost-time line charts used in earned values management (EVM) techniques (e.g. Fig. A.27). Recently, research has been directed at showing actual time performance on 3D representations of the project itself (Fig. A.25). Graphics used to analyze operational productivity of labour is also a visual representation of long standing for time performance data (e.g. Fig. A.26).

Activity ID	Description	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
1	Excavation	50.0%	0	0	0	50.0%																				
2	Foundation						50.0%	50.0%																		
3	Joining Wall											100.0%														
4	House Walls									25.0%	25.0%	0+C	0+C	C	25.0%	25.0%										
5	House Roof																33.0%	33.0%	34.0%							
6	Select Finishes												0	0	0	0	0	0	100.0%							
7	Interior Finishes																		33.0%	33.0%	34.0%	C	C			
8	Clean Up																								100.0%	
9	Fab. Garage Doors												C	C	C	16.0%	16.0%	17.0%	17.0%	17.0%	17.0%					
10	Garage Walls												33.0%	33.0%	C	C	C	34.0%	C							
11	Garage Walls																		50.0%	50.0%						
12	Garage Roof																			0	0	0	0	50.0%	50.0%	

Figure A.24 Actual daily status of activities. Source: (Hegazy et al. 2005)

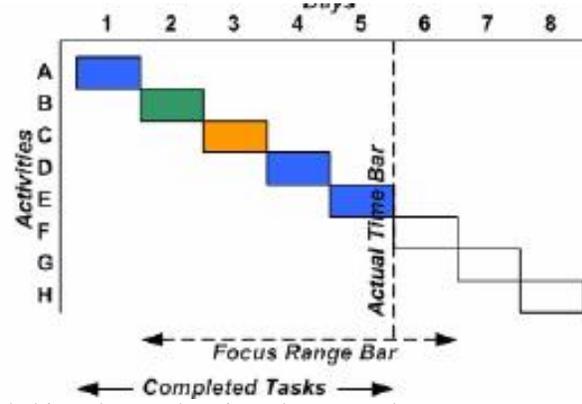
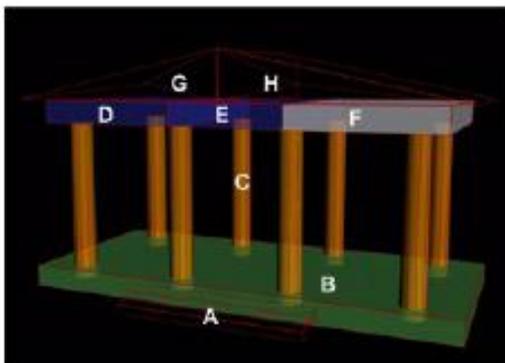
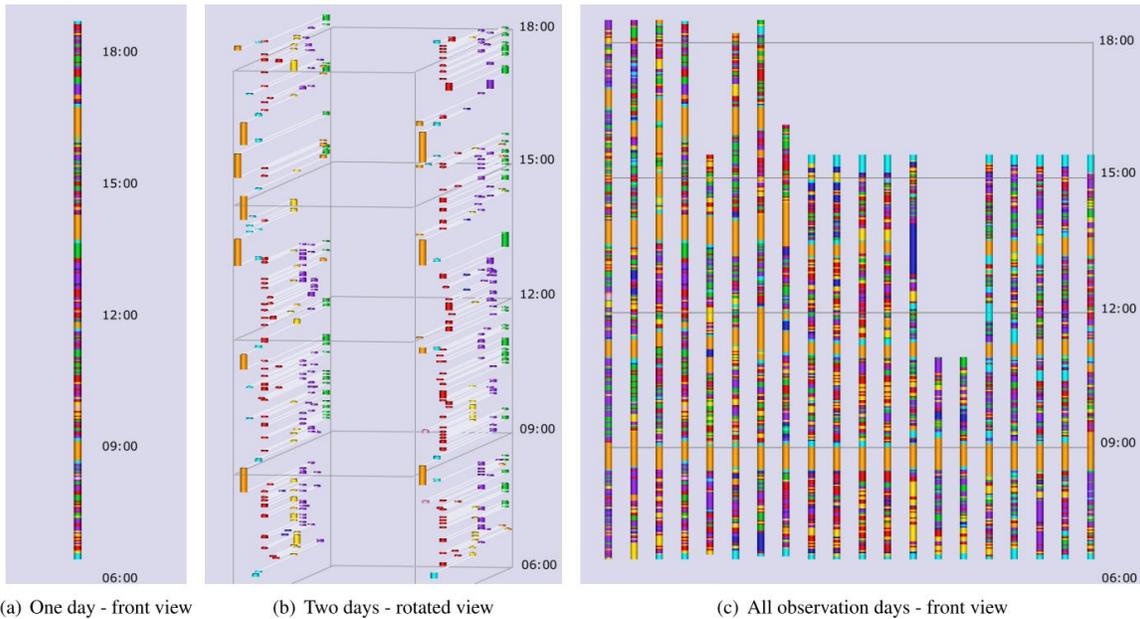


Figure A.25 Actual time performance of activities is encoded in colors and projected onto product components corresponding to those activities. Source: (Song et al. 2005)



(a) One day - front view (b) Two days - rotated view (c) All observation days - front view  
 Figure A.26 Actual activity status at operational level of detail. (© 2008 Reprinted, with permission, from Vrotsou, K., Ynnerman, A., Cooper, M., Seeing Beyond Statistics: Visual Exploration of Productivity on A Construction Site, Proceedings of 2008 International Conference in Visualization, 2008)

*Cost performance:* Using conventional charts to represent how several cost performance indices fluctuate with time during a construction process is a relatively common practice in industry and a core visualization method used in EVM techniques (e.g. Fig. A.27). Several visual representations different from earned values charts were developed to add the aspect of the hierarchical structure of cost items in addition to cost performance (Fig. A.28, Fig. A.29)

Figure A.27 has been removed due to copyright restrictions. It was a graphics visualization of EVM index. Original source: Anbari, F. T. (2003). Earned Value Project Management Method and Extensions. Project Management Journal 34(4): Figures 4, 5, and 9, pp. 14 and 16.

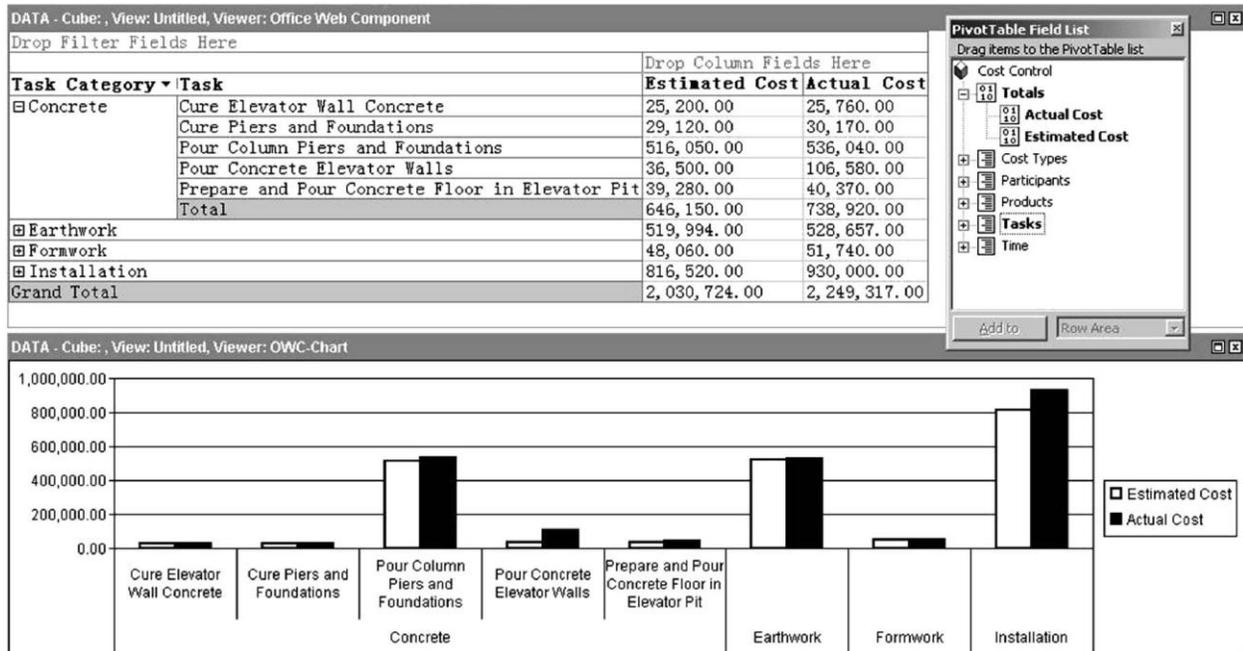


Figure A.28 Actual vs. planned cost distribution in activities at different levels of detail. Source: (Nie et al. 2007)

Figure A.29 has been removed due to copyright restrictions. It was a Tree-map representations for cost index (i.e. actual cost of work performed/budgeted cost of work performed) of pay items of different levels of detail - the color scale is used for representing cost index values. Original source: Songer, A. D., Hays, B., North, C. (2004). Multidimensional Visualization of Project Control Data. Construction Innovation: Information, Process, Management 4(3): Figure 4, pp. 185.

## **A.7 Visual representations of dependency/cause-effect between conditions and performance**

*Without the use of explanatory models:* A few visual representations have been developed in academia or in practice for visualizing the dependency relationships between actual/deviated construction conditions and actual/deviated performance without the assistance of explanatory models. Zeb et al. (Zeb et al. 2008) juxtaposed images of actual weather conditions with images of actual activity performance (i.e. activity status in terms of postponed, started, ongoing, finished, etc; Fig. A.30) for users to view and judge whether weather conditions may have impacted the execution of one or more activities.

Figure A.30 has been removed due to copyright restrictions. It was a graphics juxtaposing weather conditions with activity status for validating/invalidating cause-effect relationship between the two. Original source: Zeb, J., Chiu, C., Russell, A. (2008). Designing a Construction Data Visualization Environment. Proceedings of the 1st Forum on Construction Innovation: Figure 4, pp. 7

*With the use of explanatory models:* With the use of explanatory models implemented within a computerized information system, deviations in construction performance and the potential construction conditions causing them can be automatically identified. The purpose of trying to

visualize them is to simply report items identified and their structure in graphical form. A decision tree format (Fig. A.31) and hierarchical bar format (Fig. A.32) have been used to present the most likely reasons identified for explaining performance.

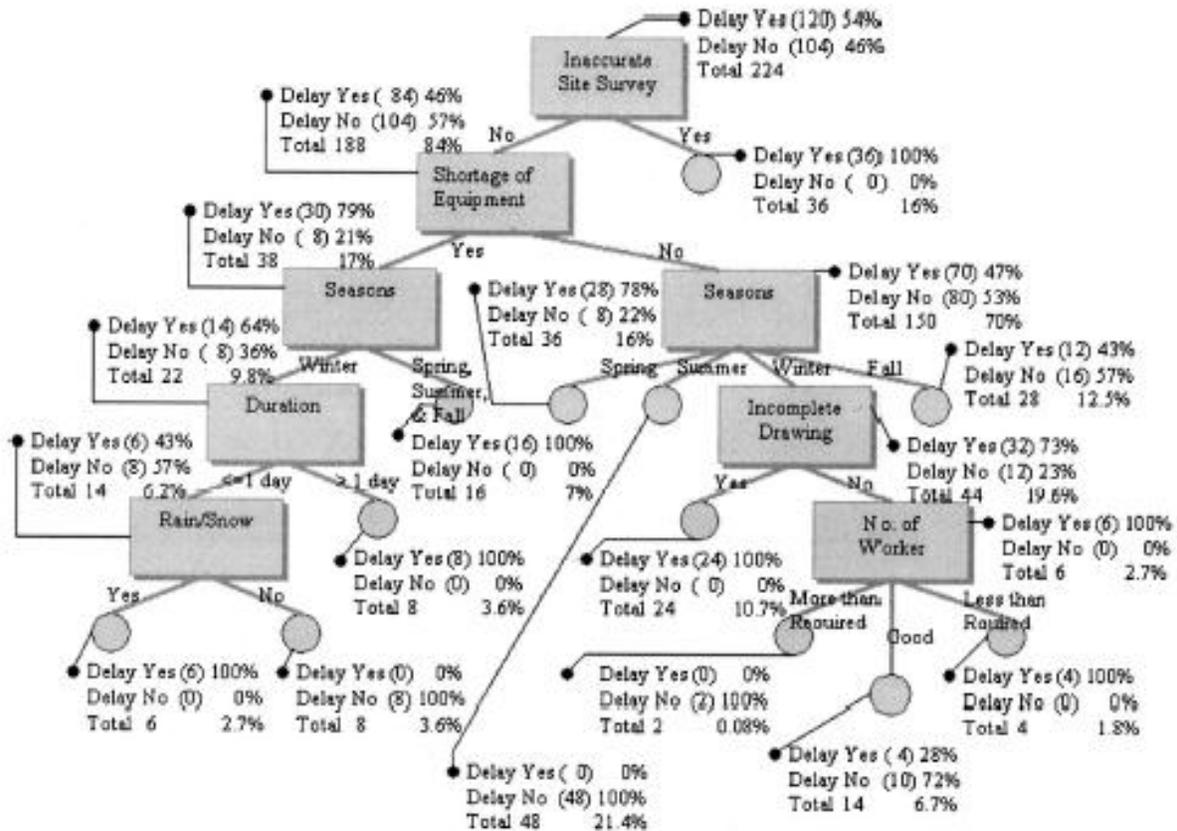


Figure A.31 Visual representation of output data of an explanatory model using generic C4.5 decision-tree classification rules for explaining reasons for delays in pipeline laying activities. Source: (Soibelman and Kim 2002)

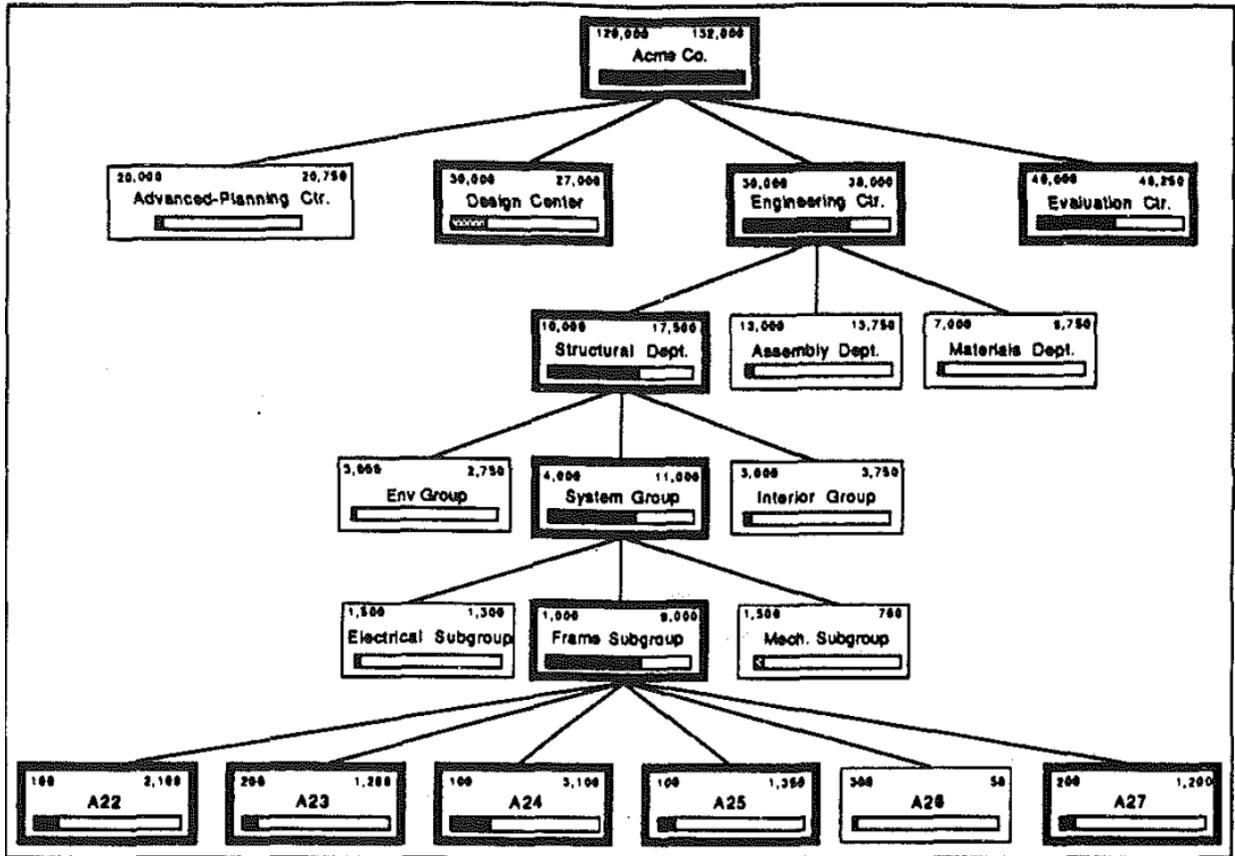


Figure A.32 Visual representations of output data of an explanatory model using generic relevance partitioning/significance testing rules to explain reasons for the increase of budgeted cost. Source: (Roth and Hendrickson 1991)

### A.8 Interacting with computerized visual representations of CM data

With the fast advancement of and easy access to computer technology, some of the visual representations of CM data presented in the previous section have been implemented as built-in computer graphics for users to access, view, and interact with in project/construction information management system software. Most of the visual representations of CM data incorporated into management software are the ones related to *time management*, *cost management*, and *risk management*, with a focus on the planning function. The kinds of popular visual representations that were computerized in main stream commercial project/construction information management software include:

1. Planned (baseline)/actual schedule:
  - Bar graph (MS project; Oracle-Primavera P6; Vico Control)
  - Time-space flow line (Vico Control)

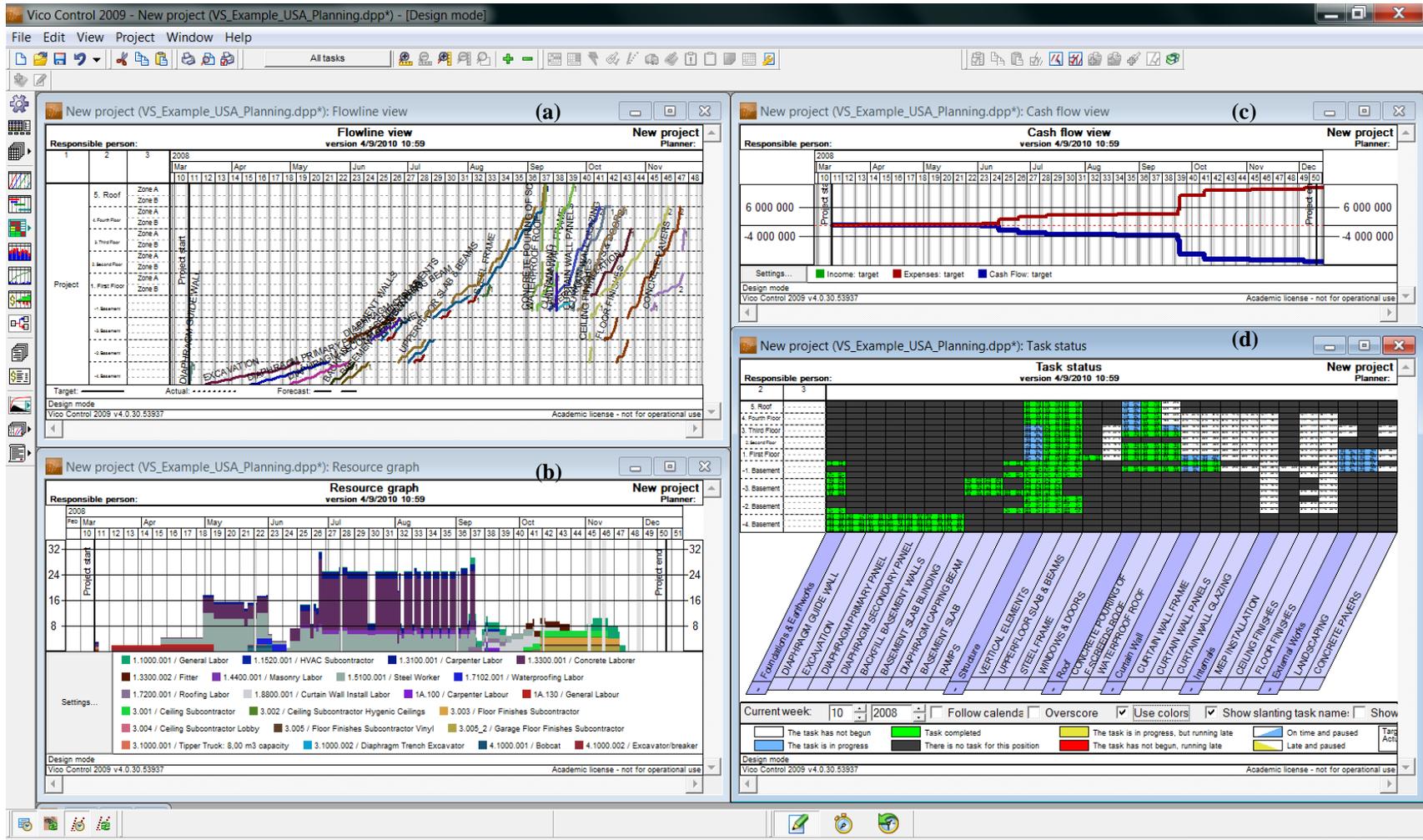
- Network (MS Project; Primavera P6; Vico Control)
2. Planned (baseline)/actual resource allocation bar chart (MS project; Primavera P6; Vico Control)
  3. Planned (baseline)/actual construction cost/budget:
    - Cash flow (Vico Control)
    - S curve (Oracle-Primavera EVM; Vico 5D Presenter)
    - EVM metrics graphics (Primavera EVM)
  4. Risk identification/analysis (Primavera Risk Analysis):
    - Risk register table
    - Probability distribution of performance metrics (Primavera Risk Analysis; Vico Control)
    - Sensitivity analysis graphics such as Tornado graph

All of the aforementioned visual representations as implemented in various software systems have user interfaces that allow users to conduct one or more of the following tasks:

1. Select which visual representations to show/hide;
2. Juxtapose visual representations;
3. Query data, i.e., filter, sort, and group data;
4. Run explicit prediction models and conduct simulations by adjusting values of input variables; and,
5. Adjust viewing attributes such as scales and visual formats of visual representations.

Since the commercial software identified in the foregoing have more or less similar functionality, the Vico Control system (Vico Software 2010), a commercial software supporting construction planning, scheduling, estimating, risk predicting, and performance monitoring, is used to show the latest visualization capabilities of CM information systems. In Figure A.33, four images generated in Vico Control shows the visual representation of data representing forecasts of performance (schedule in flow line format, cash flow, and activity status) and conditions (resource usage). Also seen in the images are several GUIs (e.g. the small graphics symbols at the left hand side) that are for users to choose the visual representations to be generated, display options, activity selection, etc.

Over the years, a number of academic researchers have developed information systems for implementing research ideas relevant to construction planning and control. Some of these systems have a visualization component, i.e., visual representations of data similar to several of the figures in the previous section along with accompanying interaction features primarily for the purpose of illustrating work related to construction planning or construction performance diagnosis. For example, the REPCON system (Russell and Udaipurwala 2002b) leveraged different schedule formats (flow line charts, bar graph, network) to present the planning/scheduling paradigm that integrates characteristics of repetitive construction processes and critical path method based on a generalization of CPM to include linear scheduling (Figure A.34). Accompanying these visual representations is an interaction capability that allows users to juxtapose the three schedule formats, select/sort activities, change scales of images, and adjust display options (e.g. whether to append activity information to bars). Hegazy and Kamarah (Hegazy and Kamarah 2008) developed a system incorporating the flow line representation with a simulation model for both computationally and visually optimizing a project's scheduling that involves many repetitive activities (Figure A.35). Zhang et al. (Zhang et al. 2009) implemented a system for users to select and view graphics of earned value indices for product components at different levels of detail (Figure A.36). Since the use of visualization in the academia-developed systems is mainly for illustrating the underlying knowledge models, not much thought about analytical reasoning using the visual representations was put into their design and development. Issues of clarity/scalability of the visual representations and required interactivity capability were also seldom addressed.



Command succeeded: Operation mode change

Figure A.33 A multiple view created by the Vico Control commercial software including images of: (a) schedule in flow line format, (b) resource usage, (c) cash flow, and (d) activity status.

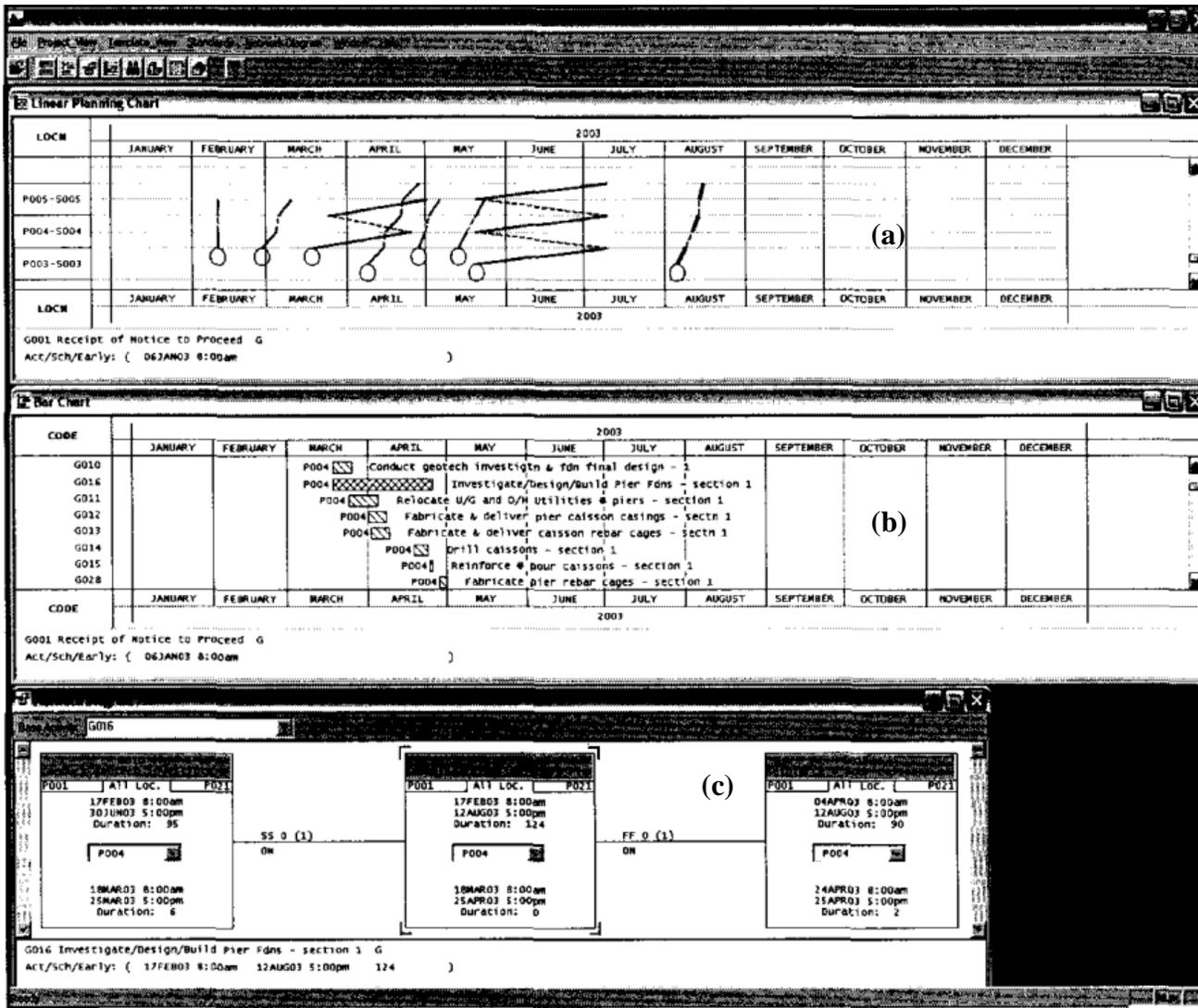


Figure A.34 A system that can generate computer graphics in both: (a) flow line, (b) bar graph, and (c) network formats for visualizing part of the schedule of a transit guideway project. Source: (Russell and Udaipurwala 2002b)

Project (c) Dr. Tarek Hegazy - E. Mail: tarek@uwaterloo.ca

**Activities**

**Conditional Methods**

**Structural Core**

**Vertical Constraint**

**Other Constraints**

Maximum Repetitions: 13 units

Deadline Duration: 220 days

Indirect Cost: 5000 \$/day

Liquidated Damages: 100000 \$/day

Incentive: 10000 \$/day

You may change project START DATE from MS Project (Project - Project Info...) menu option.

Reset

**Productivity Factors:**

Jan.:	70 %	April:	85 %	July:	100 %	Oct.:	85 %
Feb.:	70 %	May:	85 %	Aug.:	100 %	Nov.:	85 %
March:	70 %	June:	100 %	Sept.:	100 %	Dec.:	90 %

Quit Save & Proceed

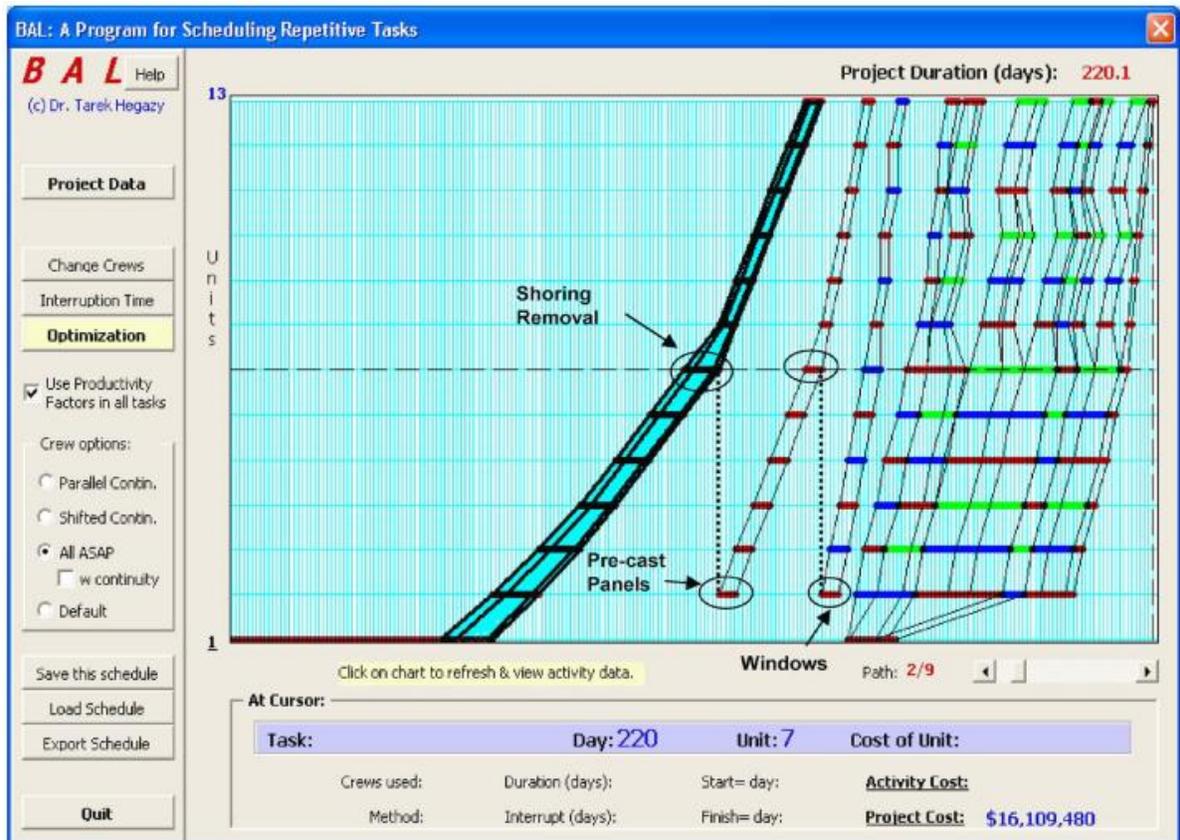


Figure A.35 A system incorporating the flow line representation with an simulation model for both computationally and visually optimizing a project's schedule. Source:(Hegazy and Kamarah 2008)

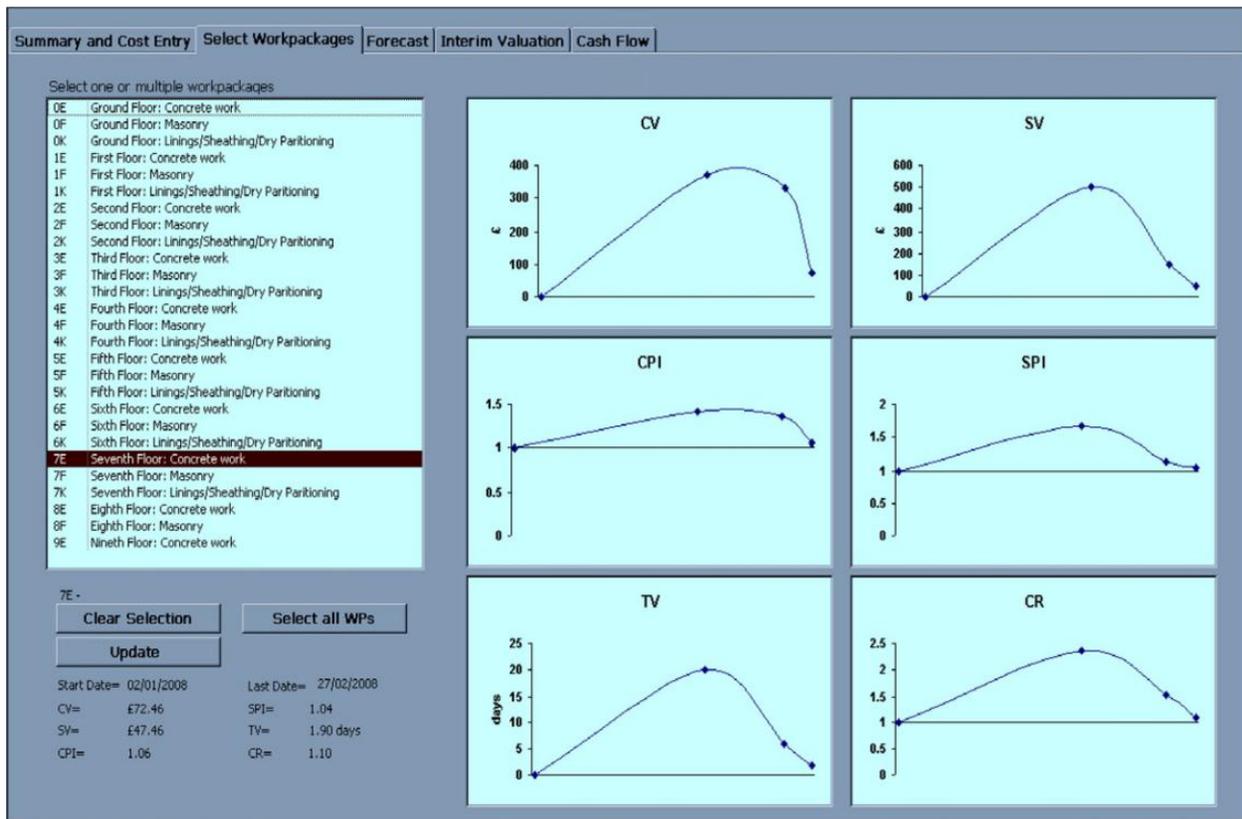


Figure A.36 Visualization of EVM indices for any selectable combination of location and product. Source: (Zhang et al. 2009)

## **Appendix B Overview of State-of-the-Art Data Visualization**

### **B.1 Introduction to data visualization**

Data visualization is the use of computer-based, interactive visual representations of data to amplify cognition (Card et al. 1999). The visual representations can be graphics (e.g. bar charts, histograms), symbols, maps, or natural forms of real world objects. Visual representations of data can:

1. Reduce the complex cognitive work needed to perform certain tasks (Keim et al. 2006);
2. Amplify cognition by: 1) increasing the memory and processing resources available to the users, 2) reducing the search for information, 3) enhancing the detection of patterns, and 4) enabling perceptual inference operations (Card et al. 1999);
3. Help users comprehend huge amounts of data, perceive emergent properties that were not anticipated, enable problems with the data itself to become immediately apparent, facilitates understanding of both large-scale and small-scale features of the data and facilitate hypothesis formation (Ware 2004); and,
4. Help the human brain to process and analyze complex information more effectively and much faster compared to sequential textual or verbal forms.

Due to the fact that visual representations of data are effective visual aids for human cognition, their use has a long history. The earliest use of visual representations of information trace back to the earliest map making and visual depiction, and later into thematic cartography, statistics and statistical graphics (Beniger and Robyn 1978; Friendly 2008). For example, the visual encoding strategies seen in popularly used bar/pie/line charts nowadays have been applied to depict observational data (e.g. Edmund Halley's Bivariate plot of barometric readings; Figure B.1) or statistical data (e.g. Playfair's line/bar chart of wages and wheat prices; Figure B.2) at least 200~300 years ago. Another example is that the visual encoding ideas in depicting construction schedules in a time line form can be seen as early as 1744 in Joseph Priestley's Chart of Biography (Figure B.3). Today, graphics are a vital part of statistical data analysis and communication in science and technology, business, education, and mass media (Cleveland and McGill 1984). In fact, graphics constitute a pervasive species of cognitive artifact, used both to reason about data and to communicate them (Zacks and Tversky 1999).

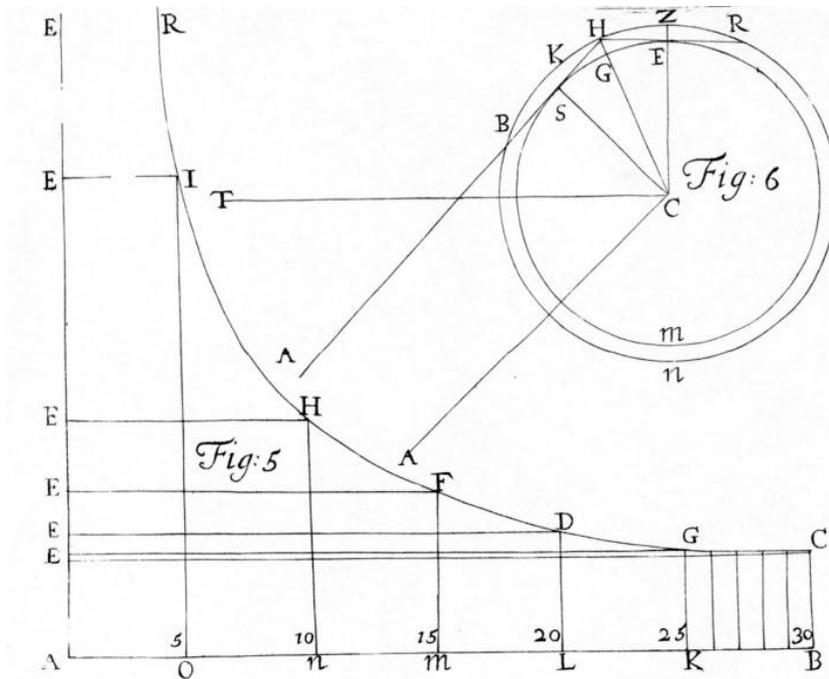


Figure B.1 A graphic (*Fig. 5*) presenting data of "Expansion of Air" (vertical coordinate) and "Height of Mercury" (horizontal coordinate). Source: (Halley 1686)

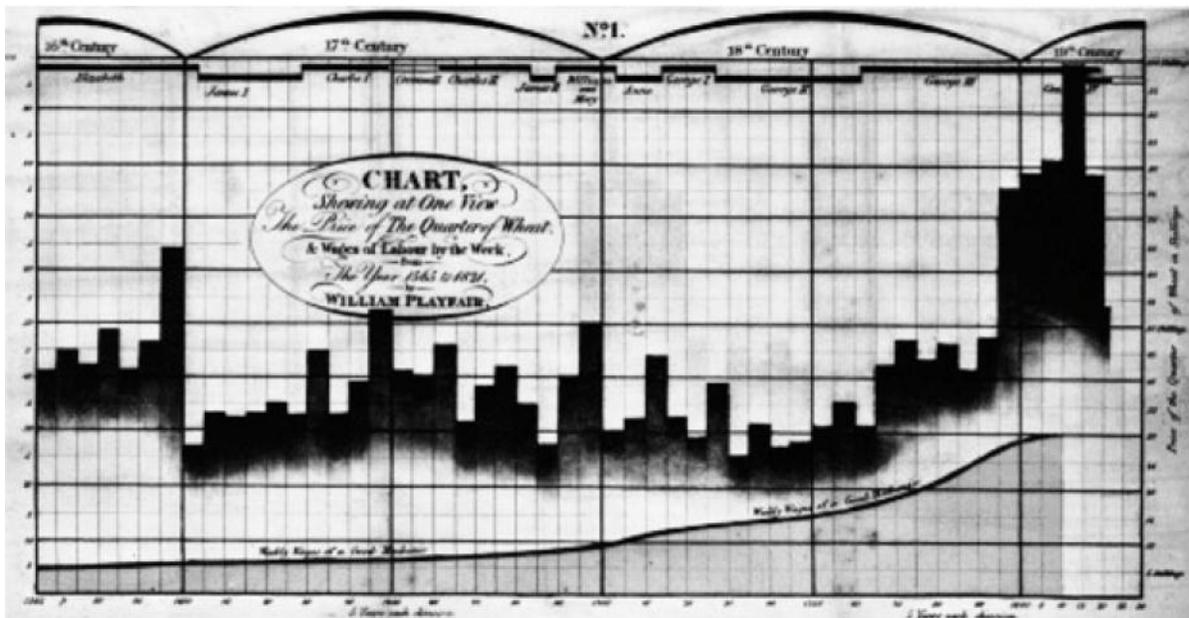


Figure B.2 A graphic comparing wages of a "good mechanic" (line) with wheat prices (bars) from the year 1565 to 1821. Source: (Playfair 1821); Image from: (Friendly 2008)

# A Specimen of a Chart of Biography

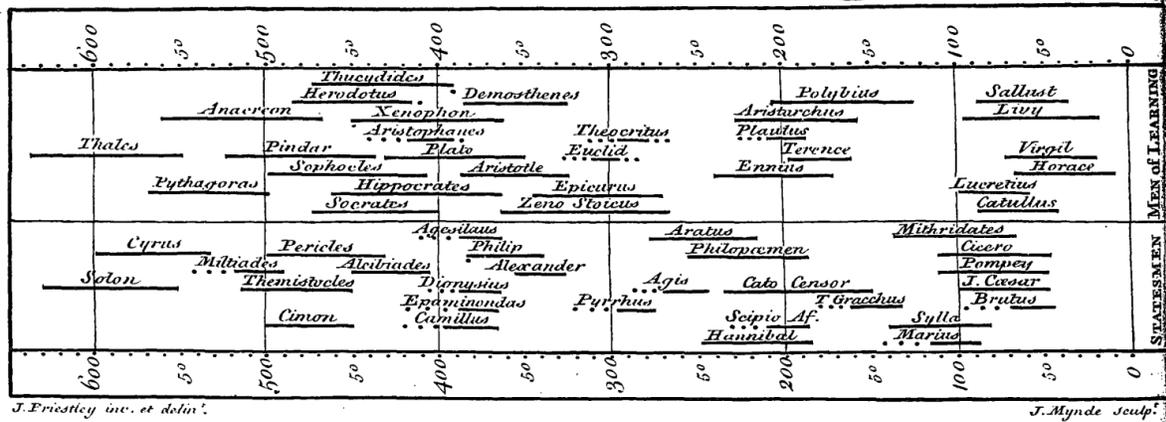


Figure B.3 A specimen of a time line chart presenting the names, lengths of lives, birth and death dates, and occupations of the most distinguished persons from BC 1200 to AD 1800. Source:(Priestley 1744)

Over the last three decades, the advancement of computer technology has brought the use of visual representations of data to a new level. Researchers started to capitalize on computer graphics, user-computer interaction, and database technology to turn static visual representations into dynamic and interactive ones. In the statistics community, several forms of “dynamic graphics” that allow users to interact with computerized data and graphics were developed to solve the difficulties of visually exploring statistical data patterns in multi-dimensional variables (Morton et al. 2009). In the user interface research community, interaction techniques were also utilized to query, browse, and navigate large numbers of visual metaphors of information spaces more quickly (Card et al. 1991).

Interactivity of data visualizations are capable of displaying more information as needed, disappearing when not needed, and accepting user commands to help with the thinking process (Ware 2004). Interaction is particularly required for 3D visualization to enable users to explore the spatial environment of the visual display. The application of interactivity to data visualization can also assist users in transforming data, mapping data to visual structures, and generating different views of the visual structures by specifying graphical parameters such as view port position, scaling, etc. (Card et al. 1999). From the perspective of the analytics experience of users, interaction lets them iteratively gain an overview of the entire collection of data, zoom in on items of interest, filter out uninteresting items, select an item or group and get details when

needed, and view relationships amongst items. Therefore, interaction helps present information rapidly and allows for rapid user-controlled exploration (Shneiderman 1996). To sum up, interaction is core to conducting visual analytics through data visualization because it "supports a true human-information discourse in which the mechanics of interaction vanish into seamless problem solving" (Thomas and Cook 2005).

## **B.2 Toward better visual representations for analytics**

Because abstract data does not have natural visual forms, researchers have investigated how to translate the data into visual representations that better fit user visual perception abilities and cognitive styles, either sensory or arbitrary (Ware 2004). This better fit may be key to how viewers can effectively and efficiently obtain and digest messages that data presents through seeing salient visual stimuli and patterns effortlessly. Three main topics related to "better visual representations for analytics" are:

1. *Formalization of mapping data to visual variables*: Bertin (Bertin 1983 (originally published in French in 1967)) seems to be the first one to formalize the process of information analysis, properties of a graphics system, and the rules of using a graphics system to represent information. Under the limitation of a flat sheet of white papers and the condition of normal lights, he proposed a planar graphic system using marks and eight visual variables describing those marks. This sign system emphasizes that the design of visual representations of data mainly depends on the measurement scale of data. At the same time, the U.S. Census Bureau also initiated research to identify best practice/standardization for statistics graphics (Schmid 1978), which was followed by several research efforts to identify elements of how to make a good graphic (Cleveland 1985; Tufte 1986; Wainer 1997; Wilkinson 1999, Unwin 2008;).

In the cognitive science field, researchers also tried to figure out puzzles regarding the graphics comprehension ability of humans so that designers can produce graphics that assist in better communication and learning. Many researchers tried to identify attributes of a visual representation that are insensitive to human differences. Some research findings concluded a correlation between characteristics of visual representations, data characteristics (e.g. measurement scales of data), and effectiveness in judging data values (a finding that is similar to Bertin's effectiveness ranking of visual variables). However, this ranking only considers

effectiveness in differentiating data values, and it did not take other issues into considerations such as space limitation of graph, use of 3D space, and support for higher level analytic tasks (e.g. finding trends or causal effects instead of only observing which value is quantitatively larger). Research findings of particular interest include:

- In judging the ratio between two different quantitative values, position judgments are more accurate than length judgments by factors varying from 1.4 to 2.5; they are also more accurate than angle judgments by factors of around 1.96 (Cleveland and McGill 1984);
- Spatial dimensions are the best dimensions for representing (perceiving) quantity (Shah 2005);
- X-Y spatial dimensions are indispensable (Pinker 1990);
- A ranking of perception effectiveness of visual variables (Mackinlay 1986); and
- Relational information display (RID) is perceived efficiently when its external representation conveys exactly the same scale property as the relational information it represents (Zhang 1996).

Other research investigated correlations between characteristics of visual representations, low-level analytic tasks, and effectiveness of conducting those tasks. Useful findings include:

- A pie chart (i.e. use of angle) is better than a stacked bar chart (use of length) in terms of allow one to discern quantity proportions (Simkin and Hastie 1987);
- Applying the Gestalt Laws to using visuals of spatial proximity, similarity in color/shape, continuity, symmetry, and closure for enhancing the perception of groupings (Pinker 1990; Ware 2004). Connectedness, relative size, and figure & ground are also strong patterns for representing (perceiving) groupings (Ware 2004);
- Viewers are more likely to describe x-y trends when viewing line graphs and are more accurate in retrieving x-y trend information from line graphs than bar graphs (Shah and J. 2002; Zacks and Tversky 1999); and,
- Proximity compatibility principle: if various information channels should be processed together, these information should be placed close together or be coded by the same visual variables (Wickens and Carswell 1995).
- Theory of integral and separate dimensions: visual variables may be perceived integrally or separately depending on the compositions of visual variables. For example, shape width

and shape height are tended to be perceived as a whole while position and color are perceived separately (Garner 1974; Ware 2004).

Although the foregoing research findings were supported by scientific evidences such as results from controlled subjects experiments, they did not address the complex relationships between effectiveness of real world problems solving and characteristics of visual representations. The main variables involved in this complex relationship include characteristics of visual displays, tasks demand, characteristics of readers (e.g. readers' knowledge about the designed graph, contexts or information the graph represents), and data complexity (Friel et al. 2001; Shah 2005).

2. *Use of 3D virtual space*: One particular topic discussed at length in the literature is the use of 2D or 3D graphics and the use of 2D or 3D virtual space to present graphics. There were mixed and somewhat opposite conclusions regarding this discussion subject. Shah (Shah 2005) identified that three-dimensional displays were better than two-dimensional ones when the answers to questions require integrating information across all three dimensions. Levy et al. (Levy et al. 1996) also found that 3-D graphs were preferred "more for depicting details than trends, more for memorability than immediate use, and more for showing others than oneself". However, some researchers expressed reservations about using 3D graphics or 3D space. Merwin et al. (Merwin et al. 1994) argued that the use of three-dimensional linear perspective drawings can degrade or occlude information; Shah (Shah 2001) found that viewers were inaccurate on tasks that requires reading individual data points in bar graphs and also had trouble identifying general trends; Smallman et al. (Smallman et al. 2001) observed that in 2D displays line of sight (LoS) ambiguity makes the altitude of an aircraft entirely ambiguous; the oblique LoS of the 3D perspective viewing angle spreads ambiguity across all three space dimensions (i.e. x, y, z coordinate), making each of them somewhat uncertain. Due to the conflict in the opinions of researchers, Shah (Shah 2005) pointed out that there has been relatively little researches on the psychological validation of the superiority between using three-dimensional or two-dimensional graphics. Thus, Brath (Brath 1999) is the first one to try and provide good practice guidelines for developing computerized and interactive visual representations in virtual 3D spaces, which may remedy shortcomings but retain the strengths of 3D visualization. Some of these rules include: 1) provide a simple 3D navigation

model (3D must be easy to use, or users will not use 3D) , 2) don't use a 3D visualization if the number of data points is low (e.g. less than 500), 3) redundant mapping of data to multiple visual attributes helps the user discriminate graphical objects in the scene and aids learning by reinforcing one another, 4) do not rely on interaction (the 3D scene should be minimally comprehensible with no user intervention), 5) complete occlusion is undesirable.

3. *How to present many visual representations of multi-dimensional data:* Abstract data representing phenomena of the world is multi-dimensional, relationally or hierarchically, but there is only limited number of visual variables (e.g. position in 3D space, color) to encode them. It is inevitable to have visual representations that consist of several separate images. How to organize many images for viewing thereby effectively providing both individual and collective insights also interests researchers. Ways of organizing them can be placing them in the time dimension (i.e. present images sequentially) and/or be placing them in the space dimension (i.e. present image concurrently). Images presented concurrently in the space dimension can be called multiple views (Wang Baldonado et al. 2000). Some research has been done to investigate whether viewing images sequentially or viewing multiple views is more effective. For example, Plumlee and Ware (Plumlee and Ware 2006) found that "extra windows are needed when visual comparisons must be made involving patterns of a greater complexity than can be held in visual working memory"; Cockburn et al. (Cockburn et al. 2008) recognized that the temporal separation of views can easily create substantial cognitive load for users in assimilating the relationship between pre- and post-zoom states—zooming is easy to do badly, as indicated by many studies in which it has performed poorly. Therefore, Wang Baldonado et al.(Wang Baldonado et al. 2000) proposed guidelines in terms of when and how to use multiple views in information visualization (Figure B.4). Although the controlled experiment type of research such as the one for validating that the context switch issues that may be encountered in using multiple views (Convertino et al. 2003) can be pursued to turn guidelines into theories, currently a sound and comprehensive theoretic system for designing multiple views is not available. In addition to leveraging multiple view design to deal with visualizing large amounts of information, Card et al.(Card et al. 1991) also recommended that for the purpose of visualization, visual representations of data should be organized hierarchically, in clustered localities that are of different information themes, and from abstraction to details.

Figure B.4 has been removed due to copyright restrictions. It was a table of guidelines of using multiple views in terms of rules and their corresponding positive and negative impacts on the utility of information visualization. Original source: Wang Baldonado, M. Q., Woodruff, A., Kuchinsky, A. (2000). Guidelines for Using Multiple Views in Information Visualization. Proceedings of the Working Conference on Advanced Visual Interfaces: Table 1, pp. 118

### **B.3 Interacting with computerized visual representations**

Human computer interaction (HCI) capability is an essential part of computerized machines that enables human beings to operate them. For users to utilize efficiently and effectively computerized visual representations of data, interactivity plays an important role. Functions fundamental to interacting with computerized visual representations have been categorized by researchers in several ways. Yi et al.(Yi et al. 2007) categorized interaction features based on high level analytic tasks as follows:

- Select: mark something as interesting;
- Explore: show me something else;
- Reconfigure: show me a different arrangement;
- Encode: show me a different representation;

- Abstract/Elaborate: show me more or less detail;
- Filter: show me something conditionally; and
- Connect: show me related items.

Chuah and Roth (Chuah and Roth 1996) on the other hand recognized that interaction features can be differentiated by low level operational tasks (Figure B.5) including the three main tasks of graphical operations, set operations, and data operations. Theus and Urbanek (Theus and Urbanek 2008) organized interaction features into the four main categories of querying, selection and linked highlighting, sorting, and zooming. Thomas and Cook (Thomas and Cook 2005) identified that for effecting visual analytics processes, interaction features should be supported for users to modify data transformations (i.e. filtering), to modify visual mapping, to modify view transformations (i.e. navigation), and to discourse with information.

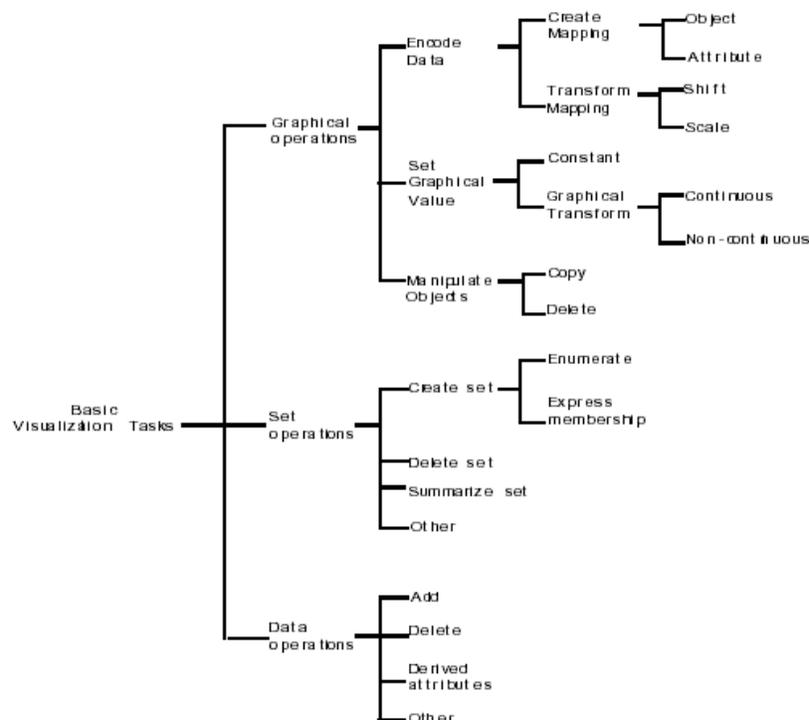


Figure B.5 A classification of functions of interaction features in information visualization. (© 1996 IEEE. Reprinted, with permission, from Chuah, M. C., and Roth, S. F., On the Semantics of Interactive Visualizations, Proceedings of the 1996 IEEE Symposium on Information Visualization (INFOVIS '96), 1996)

Because data/information that users need to explore is many and there is limited visualization space to present several visual representations of data at the same time, it is inevitable that users need to sequentially interact with computer systems to obtain and view the images they desire to

see. It was found that there is a wait-time tolerance, which can be referred to as "time constant", for users to complete a certain type of human-computer interactions (Thomas and Cook 2005) before they would abandon interactions or the effectiveness of interactivity would be significantly degraded. The types of human-computer interactions and their corresponding time constants can be loosely categorized as:

1. Time to perceive an expected immediate response (e.g. brush to highlight): from milliseconds to a second;
2. Time to respond to user actions (e.g. click "close window", click a "generate images" button): from a second to 10 seconds; and,
3. Time for conducting analytical reasoning and human-information discourse: usually minutes to even hours.

Therefore, while there could be many ways of providing a graphical user interface for users to interact with a information system, GUI designs and data processing algorithms that respond to users instructions should meet the minimum requirements regarding the users tolerance to system response time.

#### **B.4 Designing and developing data visualization tools**

1. *Design and development process*: The foregoing research efforts directed at identifying good visualization design practices were, for the most part, focused on deducing general principles that can be used in broader applications. However, most of them were tested on generic low-level analytics tasks such as retrieving values, finding extremes, clusters, etc (Amar et al. 2005). Indiscriminately applying those principles to simply encoding the data available into visual representations may not address complex and unique data analysis needs for different knowledge domains. Therefore, many researchers started to formalize the entire visualization design process that starts from understanding domain problems, representing the problems, designing visualizations with reference to general principles, and conducting design evaluation, including possible iterations of these steps in order to refine the design until users perceive the usefulness of the final design in helping solve their domain problems. Brath (Brath 2003) used an actual visualization development project to demonstrate the entire development process of collecting requirements, proposing possible designs, testing the

designs, and collecting feedback. Graham et al.(Graham et al. 2000) analyzed that a complete visualization tool development process requires the steps of gathering initial requirements, prototyping, conducting qualitative testing of functionality for refinement, conducting qualitative testing of usability and scalability for refinement, and finally conducting quantitative testing of the entire design for validation. Amar and Stasko (Amar and Stasko 2005) established heuristic rules as a reference for designing and evaluating a visualization system that support high-level analytics tasks such as decision-making and learning. These heuristic rules state that a visualization system should assist in: a) knowledge domain specific analytics such as creating/acquiring/transferring knowledge about important domain parameters, b) discovering useful correlative models, and c) clarifying possible sources of causation. Munzner (Munzner 2009) structured a nested model for designing and validating visualization. The model encompasses four nested layers related to tasks of domain problem characterization, operation/data abstraction design, encoding/interaction technique design, and algorithm design. Each layer of design tasks has potential pitfalls that can be identified and remedied by conducting both validation steps immediately after each task and the downstream validation steps after a visualization system is implemented. For example, for the task of domain problem characterization, the potential threat is formulating the wrong problem. This threat can be eliminated by observing/interviewing target users to refine or rectify the problem characterization immediately after the problem definitions is formed. The threat can be confirmed as no longer present by the downstream validation process of observing the adoption rate of implemented visualization products.

2. *Evaluation guidelines*: Evaluation is a crucial step to either validate the quality of visualization methods/systems developed or research into data visualization. Ware (Ware 2004) categorized that there are several research-validation goals for conducting an evaluation:

- Uncover fundamental truths, i.e. theory, for applying data visualization;
- Discover the nature of the world - a feeling of the range of phenomena of the world that are related to applying data visualization;
- Ascertain if an existing theory generalizes to practices, i.e., to see if well known laboratory results generalize visualization problems;
- Objectively compare between two or more display methods;

- Objectively compare between two or more display systems;
- Measure task performance of utilizing data visualization; and,
- Ascertain user preferences for different display methods. Sometimes the "cool appearance of a particular interface can be decisive in its adoption".

He also recommended several evaluation methods depending on the kind of visualization research is being validated. For identifying 'truth' about human behaviour involved in applying data visualization, research-validation techniques used by psychophysics and cognitive psychology fields along with a statistical exploration processes are more appropriate. For understanding a wide range of issues surrounding data visualization (e.g. discover the nature of the world, objectively compare between two or more display systems), structural analysis that mainly involves interviewing subjects and eliciting their appraisal opinions (e.g. judging by the use of certain rating scales) would be more feasible and suitable.

There were also many reports on developing new data visualizations in which how evaluations of the new visualizations were conducted was also documented. It is clear that a "one size fits all" methodology for evaluating data visualizations or validating data visualization researches does not exist. Based on these reports, evaluation approaches can be categorized by:

- Purpose of evaluation, which include:
  - *Comparing visualization component(s)/systems with or amongst existing ones--* Kobsa (Kobsa 2001) compared three commercial visualization systems in terms of speed and accuracy in completing analytic tasks; Demian and Fruchter (Demian and Fruchter 2006) compared innovative visual representations with other tree-structure visual representations.
  - *Comparing visualization component(s)/systems amongst designed alternatives--* Plaisant et al. (Plaisant et al. 2008) determined superiority of several innovative visual analytic designs by predetermined scoring criteria.
  - *Comparing the use of data visualization with other analytic tools or current analytic practices--*Gonzalez and Kobsa (Gonzalez and Kobsa 2003) conducted such a comparison to determine if an information visualization systems can be complementary

to data analysts' current data analysis practices.

- *Demonstrating usefulness of visualization research results*: depending on the usefulness of visualization research results in terms of generalizability, precision, and/or the realism sought, Carpendale (Carpendale 2008) analyzed and categorized several evaluation approaches as to which approach may better demonstrate one of the forgoing result factors, but possibly at the expense of others. These approaches can also be coarsely grouped as quantitative ones or qualitative ones.
- Methodologies of evaluation, which include:
  - *Examination by checking against pre-defined metrics, guidelines, or by experts*-- Freitas et al. (Freitas et al. 2002) self checked a new tree visualization against pre-defined criteria of usability.
  - *Controlled subject experiments*-- Many controlled experiments conducted for evaluating information visualization designs can be found in a review work conducted by Chen and Yu (Chen and Yu 2000).
  - *Feedback elicitation*-- In (Weaver et al. 2006), user feedback for refining and validating a new visualization solution were iteratively sought.
  - *Longitudinal and field study*-- In (Trafton et al. 2000), a real world information seeking situation for weather forecasting was simulated in order to observe how users utilize the visualization.
- Components under evaluation, which include:
  - *Data representation & transformation*-- Weaver et al (Weaver et al. 2006) included the "what more data needed?" as one of the criteria for refining and evaluating a new visualization design.
  - *Computational algorithm*-- Morrison et al (Morrison et al. 2003) developed a new algorithm to map high dimensional variables to low-dimensional spaces (i.e. multidimensional scaling) and compared its running speed against older algorithms as an evaluation approach.
  - *Visual representations*-- Alonso et al. (Alonso et al. 1997) evaluated that visual representation in the form of a Gantt chart can help users to perform tasks of interval comparison faster than using tabular representations.
  - *Interaction features*-- Plaisant et al. (Plaisant et al. 2002) evaluated whether a new

interaction design of rescaling of tree branches, optimized camera movement, and preview icons integrated with search and filter functions can improve the task of large tree browsing.

- Metrics of evaluation, which include:
  - *Usability* (e.g. operability, efficiency, satisfaction, learn-ability)-- North and Shneiderman (North and Shneiderman 2000) observed how well users can learn and operate a new interaction feature to construct a link amongst views of relational data and then use the coordinated views beneficially (i.e. the usability metric of operability & learn-ability); Songer et al. (Songer et al. 2004) compared several visual representations of cost data by the metrics of number of questions correctly answered and subjective preferences (i.e. the usability metric of satisfaction); Pirolli and Rao (Pirolli and Rao 1996) compared several information visualizations by the metric of time taken to finish the same task (the usability metric of efficiency).
  - *Usefulness* (e.g. whether one can generate insights or meet task purposes)-- Saraiya et al (Saraiya et al. 2004) evaluated different visual analytic systems based on scores on number and importance of insights obtained (i.e. the usefulness metric of whether one can generate insight); Demian and Fruchter (Demian and Fruchter 2006) compared the quality of conducting a pre-defined task (the usefulness metric of whether one can meet task purposes) by using different visual representations.
  - *Heuristics*: Through meta-analysis, Zuk et al. (Zuk et al. 2006) organized three sets of heuristics identified in past research that dealt with perceptual and cognitive heuristics (e.g. considering people with color blindness, considering Gestalt Laws), interactivity heuristics (e.g. provisions of interaction features such as details on demand and zoom and filter (Shneiderman 1996)), and analytics heuristics (e.g. whether one can formulate cause and effect, confirm hypothesis, etc (Amar and Stasko 2005)) that can be used in evaluating information visualizations. Forsell and Johansson (Forsell and Johansson 2010) also tried to identify the 10 most useful heuristics for evaluating usability problems in an information visualization.
- Measurements of evaluation, which include:
  - *Quantitative measurement*: Number of accurate answers (Songer et al. 2004), time to complete tasks (North and Shneiderman 2000).

- *Qualitative or subjective measurement*: Likert scale score, feedback, checklist (Gonzalez and Kobsa 2003).
- *Conformance to heuristics*: Ardito et al.(Ardito et al. 2006) and Amar and Stasko (Amar and Stasko 2005) used descriptive reporting to document whether the information visualization tool under evaluation *can* or *cannot* support user's ability to conduct tasks prescribed in the heuristics (e.g. "users are enabled to focus on a specific interval and try to zoom in/out it" (Ardito et al. 2006), "users are enabled to expose cause and effect" (Amar and Stasko 2005)).

### **B.5 State-of-the-art data visualization tools**

By fusing an understanding of human behaviour in comprehending visual stimuli and interacting with computer systems with the latest information technology such as computer graphics, databases, and artificial intelligence, state-of-the-art data visualization tools have been developed to enhance user capabilities of conducting information search and analytics tasks. Although these tools can range from a new application improving on usability of a certain interaction feature to a visualization system incorporating several state-of-the-art visualization technologies for general use, all in all they automate tediously manual steps of iteratively producing graphics one desires to see, either using pen and paper or by computer system, so that users can focus on viewing images and conduct analytical reasoning quickly. In Table B.1, representative state-of-the-art data visualization tools developed are introduced.

Table B.1 Representative state-of-the-art data visualization tools and their general functionality

General functionality	Descriptions of Specific Techniques
<b>Automating generation of effective images</b>	A visualization tool that can intelligently choose the most effective way of visually encoding data according to data properties such as measurement scales or according to analytics tasks needs. With this kind of tool, users can focus on choosing the data/tasks for a visualization and leave the work of formulating/generating effective images to the computer (Mackinlay 1986; Casner 1991).
<b>Enhancing interactivity-data selection</b>	<ul style="list-style-type: none"> <li>• <i>Dynamic query</i>: The image of data can be seamlessly refreshed to reflect user constantly changing data selection conditions in real time with the use of sliders (Figure B.6).</li> <li>• <i>Visual query</i>: Users can randomly select data dimensions and specify which visual variables to represent what data dimensions. The system then can automatically generate visual representations based on user specifications (Figure B.7).</li> <li>• <i>Brushing/Visual selection</i>: Users can formulate data selection conditions by directly brushing/selecting data points in the visual representations (Figure B.8, Figure B.9).</li> </ul>
<b>Enhancing interactivity-manipulating visual representations</b>	<ul style="list-style-type: none"> <li>• <i>Manipulating scales</i>: The scales of images are differentiated based on user "point of interest" in the images. The part of images that users are interested in viewing more closely get larger scales. The determination of "point of interest" is based on user specified data selection conditions, which can be done by brushing/selecting on the images or specifying data query conditions using dialogue boxes. Examples of this type of scaling manipulation include the techniques of focus + context (Figure B.10) and selective dynamic manipulation (Figure B.11). The manipulation of scales is not limited to scale the sizes of components in an image. It can also be used to change data density. For example, when zooming in a visual representation, visual marks representing more detailed information can emerge (Figure B.12).</li> <li>• <i>Manipulating ordering of visual marks</i>: Being able to sort visual marks by various sorting conditions (e.g. sorting by data values, manual sorting) was found to be useful to reveal intrinsic visual patterns (Figure B.13), which may indicate insights hidden in data.</li> </ul>
<b>Enhancing interactivity-interaction coordination</b>	Applying the same interaction feature to several images can be time consuming. Therefore, the technique of interaction coordination (e.g. (Becker et al. 1987; North 2000; Boukhelifa et al. 2003)) was developed so that users only need to give interaction instructions once (e.g. select and highlight data of interest in one image) and the system can automatically apply that instruction to all other presented images. The flow chart of this coordination mechanism is shown in Figure B.14.
<b>Innovative visualization (combined use of computer graphics and interaction features)</b>	<ul style="list-style-type: none"> <li>• <i>Visualizing multi-dimensional data</i>: Multiple views of individual graphics are juxtaposed and integrated as another image mainly for revealing correlations between variables. Examples of this type of innovative visualizations include the visual representations of Tablelens (Figure B.15), Scatter Plot Matrix (Figure B.8), Trellis Graphics (Figure B.13), and a dashboard of visual representations of arbitrary data types (Figure B.16). Interaction features such as sorting and linking brushing are incorporated into these types of visualizations so that users can explore correlations. A single view is also possible for presenting multi-dimensional data. Examples are Parallel Coordinates plot (Figure B.17) and Mozaic Plot (Figure B.18). Sorting interaction features are essential to make use of this type of single-view multi-dimensional visualizations.</li> </ul>

General functionality	Descriptions of Specific Techniques
	<ul style="list-style-type: none"> <li>• <i>Visualizing various levels of detail of data:</i> multiple views of individual graphics that represent various levels of detail of data are juxtaposed or superimposed so that users can cross reference information that is at different levels of detail. Techniques for achieving this kind of visualization include overview + detail (Figure B.19). Coordinating data selection between an "overview" image and "detail" images is essential to effect the "overview + detail" type of visualization; The Focus + Context (Figure B.20) visualization is able to present different levels of detail (e.g. overall patterns vs. detailed individual values) in a single view. Being able to specify differentiated image scales is essential to use the "focus + context" kind of visualization. A single view of a tree-structure image such as Tree Map (Figure B.21) or Hyperbolic Tree (Figure B.22) is also a common way of presenting data of various levels of detail. The interactivity of traversing levels of a tree (e.g. zoom in/out a tree map) is crucial to make good use of this visualization.</li> <li>• <i>3D scene:</i> Several visual representations of data can be placed in a 3D virtual space (Figure B.23) so that more images can be compacted into a scene for users to have a quick scan. The interactivities of navigation and zooming in/out are essential to maximize the usefulness of this type of visualization.</li> <li>• <i>Visualizations for assisting in particular analysis tasks:</i> Several specialized visualization tools were developed to assist in or improve particular analysis tasks such as analyzing number of employees vs. power demand by multiple time scales simultaneously (Wijk and Selow 1999), hotel visitation patterns (Weaver et al. 2006), the levels of gene activity and metabolites by pathways, time, and species (Meyer et al. 2010), etc.</li> </ul>
<p><b>Visualization systems for general use</b></p>	<p>Visualization systems fuse as many general functionalities of state-of-the-art data visualizations into an environment so that allow users to have high flexibilities in specifying which data dimensions in a database to include for visualization, how to map selected data dimensions to visual variables, how to query data ranges/granularities, and how to adjust visual attributes of images without the need of programming.</p> <p>For example, the visualization system described in (Eick 2000) allows users to select from 13 predefined effective graphics formats, which include traditional ones (e.g. bar charts, scatter plots, line charts ) and innovative ones (e.g. Time Tables, ParaBoxes that combine the format of Parallel Coordinate chart and bubble plot), and then assign which data dimensions to be visualized by the selected graphics format. However, there is a restriction imposed in that data dimensions of certain types of measurement scale can only be visualized by the selected format in order to ensure the effectiveness of images. This restriction is based on the knowledge about the effectiveness ranking of visual encodings. The system developed by Stolte et al.(Stolte et al. 2002) provides even more flexibility. It allows users to specify which data dimensions should be visually encoded and by what visual variables, which can be only selected from a group of candidate effective visual variables. Both aforementioned systems have intuitive and easy graphical user interfaces for users to connect to mainstream databases, and specify the data dimensions that will be included for visualization. To some extent, the charting capability of Microsoft Excel is similar to this kind of visualization system.</p>

Figure B.6 has been removed due to copyright restrictions. It was a graphics showing the use of sliders to specify attribute values of chemical elements. In the Periodic Table, elements that do not meet the query specifications are dimmed in real time while the user moves the buttons of sliders. Original source: Ahlberg, C., Williamson, C., Shneiderman, B. (1992). Dynamic Queries for Information Exploration: an Implementation and Evaluation. CHI '92 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Figure 2, pp. 621

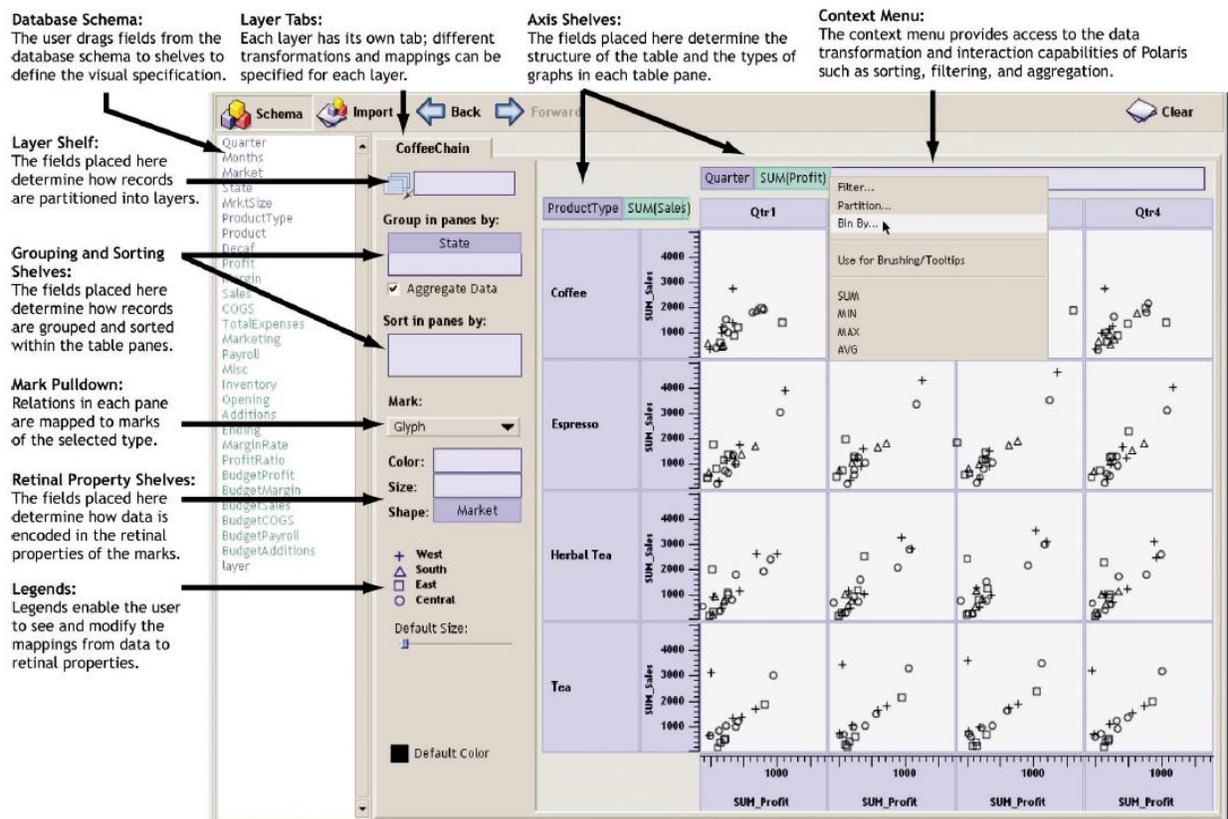


Figure B.7 A data visualization system with intuitive interfaces for users to select data dimensions, filter data ranges, sort and group data, and specify how data dimensions to be mapped to visual variables. (© IEEE 2002. Reprinted, with permission, from Stolte, C., Tang, D., Hanrahan, P., Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases, Transactions on Visualization and Computer Graphics 8(1), 2002)

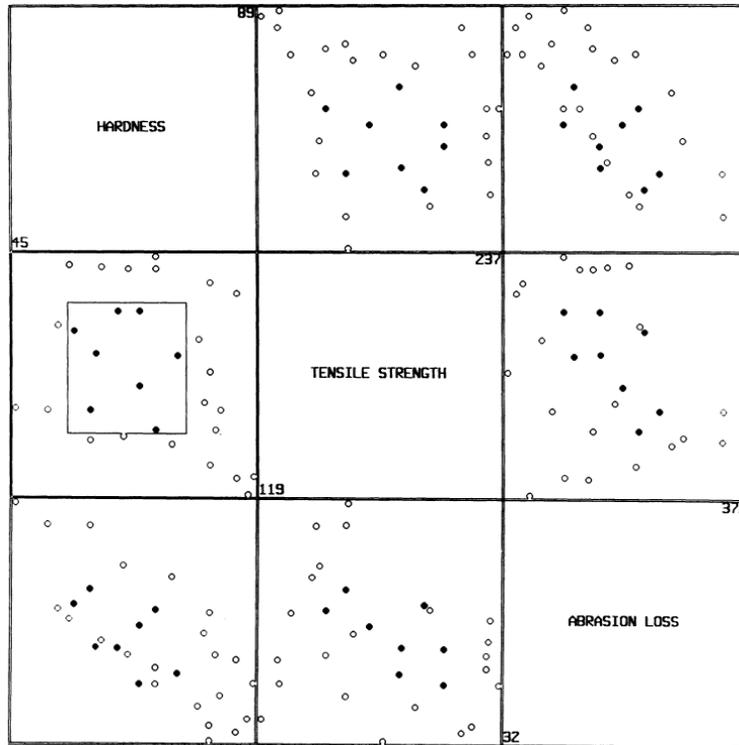


Figure B.8 Selecting data points in a scatter plot by a rectangular brush and data points in other scatter plots associated with the selected data get highlighted. Source: (Cleveland and Becker 1987) (Reprinted with permission from Technometrics Copyright 1987 by the American Statistical Association)

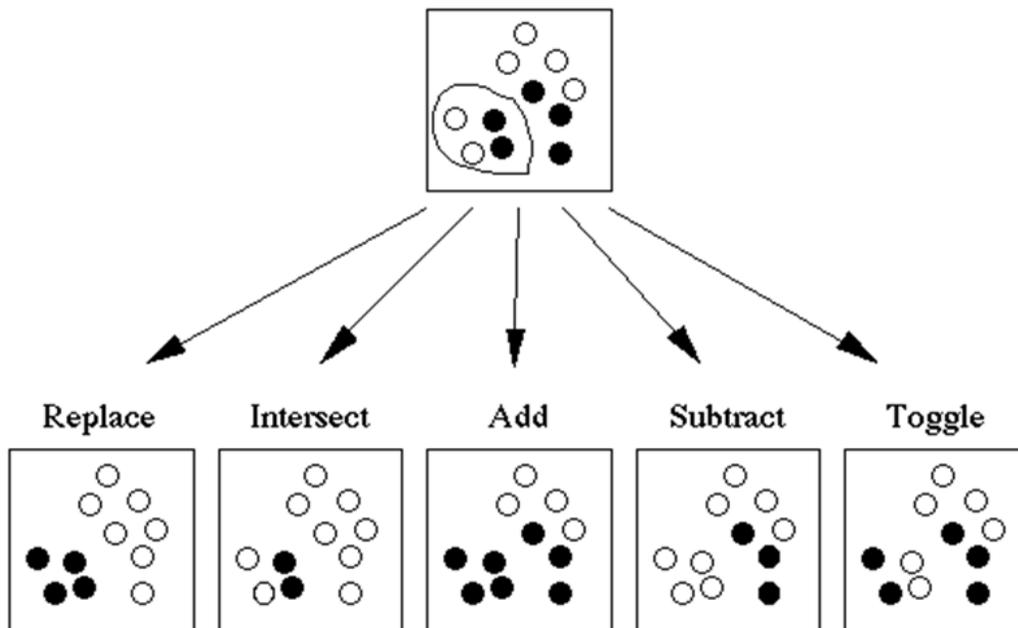


Figure B.9 Using a "mouse lasso" to replace with, intersect with, add, subtract, or toggle the previous rectangle-select data points. (© IEEE 1996. Reprinted, with permission, from Wills, G., Selection: 524,288 Ways to Say "This Is Interesting", Proceeding of IEEE Symposium on Information Visualization, 1996)

Figure B.10 has been removed due to copyright restrictions. It was graphics showing (a) A magnification function by which the scale of an image is determined by distance from the focal point, (b) application of the magnification function to the horizontal dimension, (c) application of the magnification function to both horizontal and vertical dimensions. Original source: Leung, Y. K., and Apperley, M. D. (1994). A Review and Taxonomy of Distortion-Oriented Presentation Techniques. *ACM Trans. Comput. -Hum. Interact.*: Figures 5(b)~5(d), pp. 133

Figure B.11 has been removed due to copyright restrictions. It was two graphics showing interaction features that allow users to specify sizes of 3D bars thereby reducing occlusion problems. The blue and red bars in (b) were adjusted to be thinner than the ones in (a) so that green bars can be seen more clearly. Original source: Chuah, M. C., Roth, S. F., Mattis, J., Kolojejchick, J. (1995). SDM: Selective Dynamic Manipulation of Visualizations. *Proceedings of UIST' 95 Symposium on User Interface Software and Technology*: Figures 8 and 11, pp. 65.

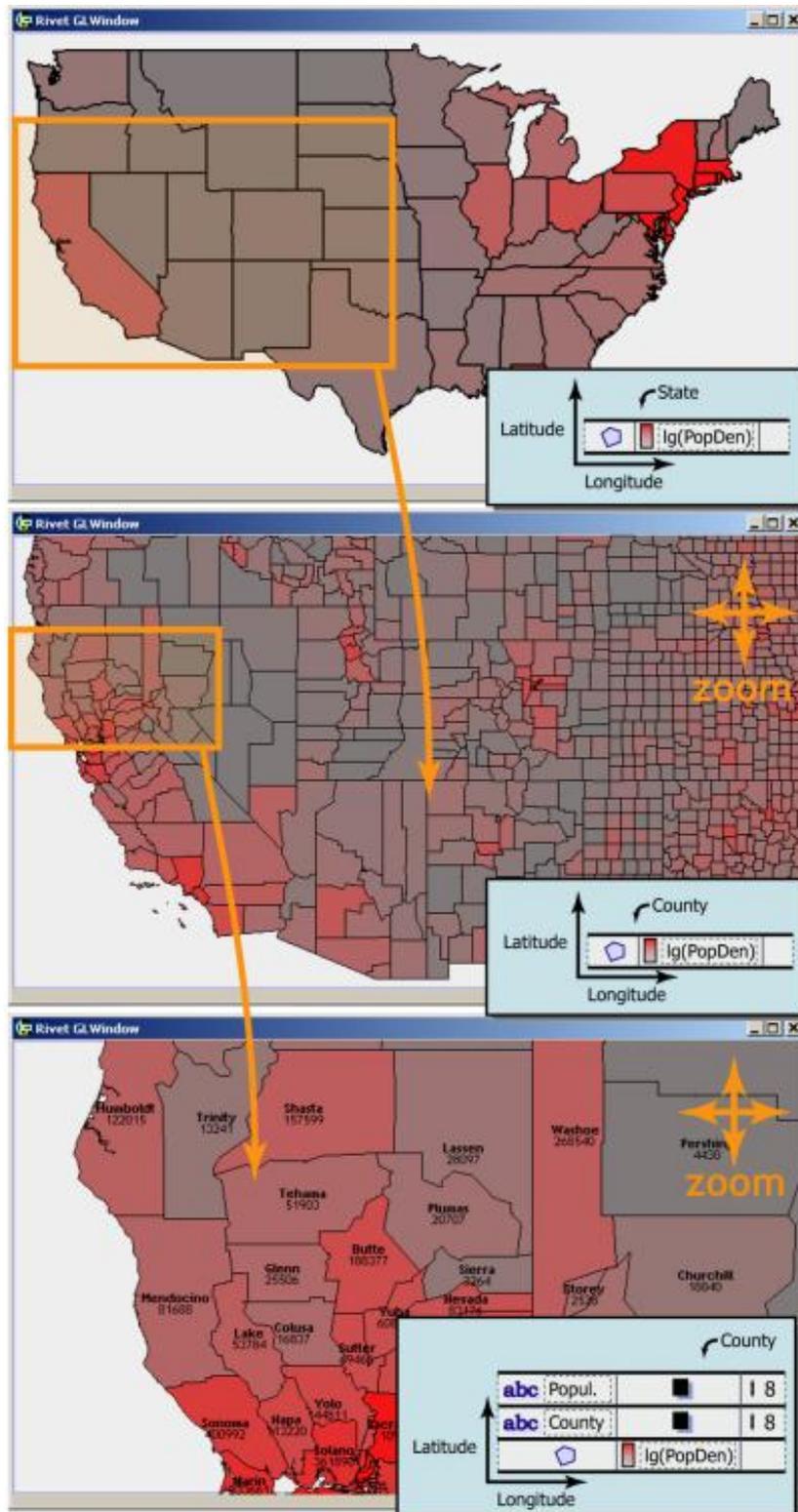


Figure B.12 A interactivity mechanism by which data density changes along with the zooming in/out of a map. (© IEEE 2003. Reprinted, with permission, from Stolte, C., Tang, D., Hanrahan, P., Multiscale Visualization Using Data Cubes, IEEE Trans. Visual. Comput. Graphics 9(2), 2003)

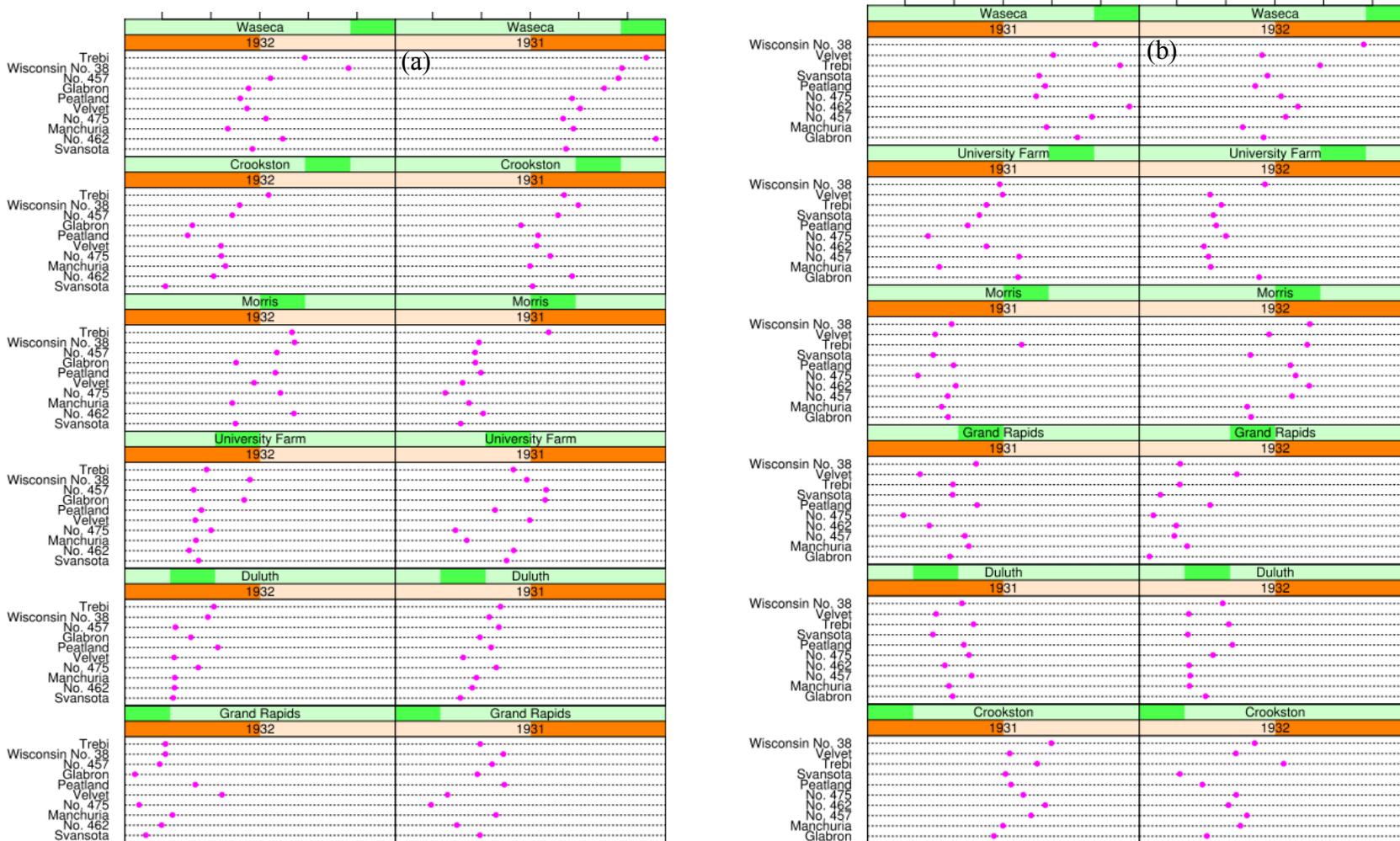


Figure B.13 An anomalous pattern at the "Morris" barley site can be identified in the left "Trellis view" but not in the right view while: in (a) dot charts are sorted by median yields of barley sites (from bottom to top) and by median yield of year; data points in each dot chart are sorted by median yields of barley varieties, in (b) dot charts and data points in a dot chart are simply sorted by alphanumeric ordering. Source: (Becker et al. 1996) (Reprinted with permission from Journal of Computational and Graphical Statistics Copyright 1996 by the American Statistical Association)

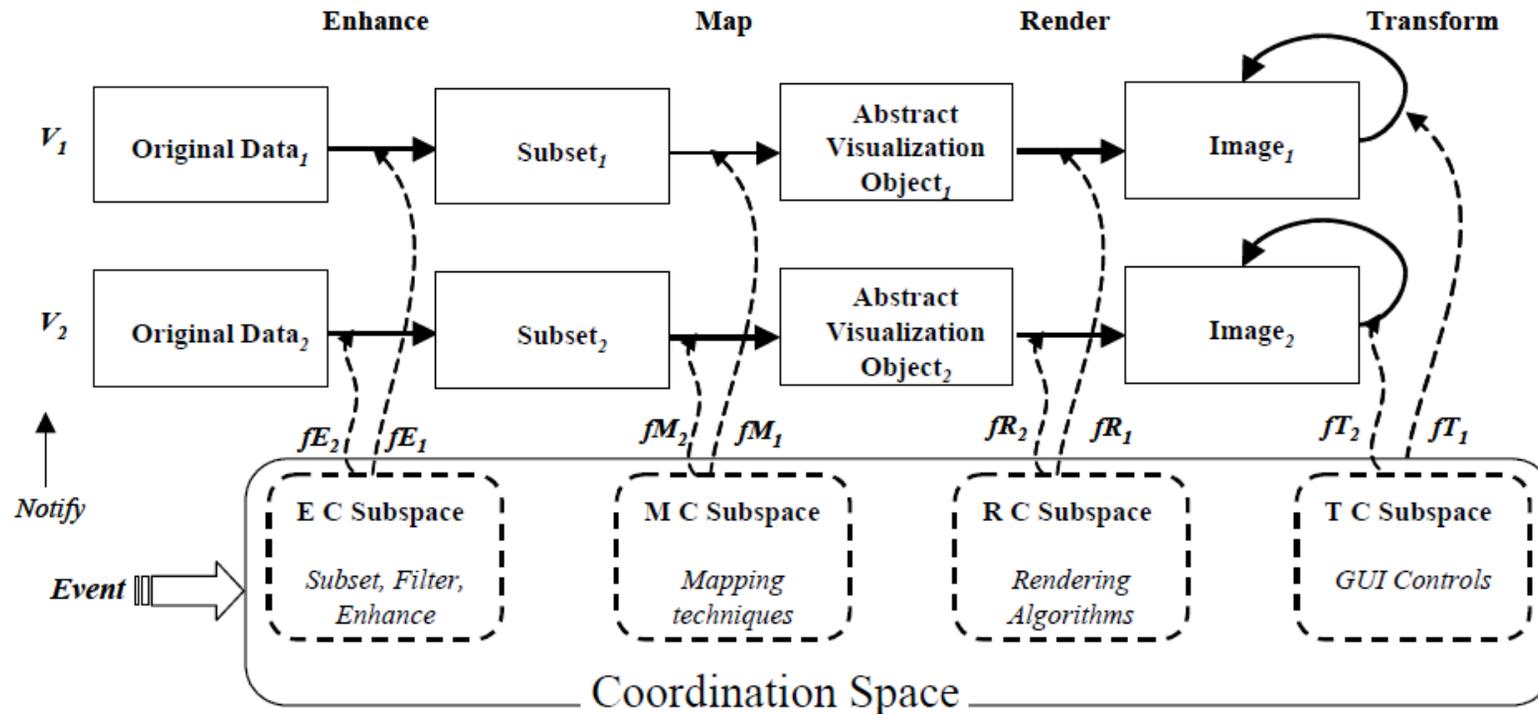


Figure B.14 has been removed due to copyright restrictions. It was a flow chart of a multiple view coordination model. In this model, an event is initiated by users in one of the two views ( $V_1, V_2$ ). This event may include basic visualization processes such as filtering data (i.e. enhance), mapping data to visual variables (i.e. map), rendering an images (i.e. render), and rotating view points (i.e. transform), which consist of a coordination space that is applicable to both views. Source: (Boukhelifa et al. 2003) © IEEE 2003. Reprinted, with permission, from Boukhelifa, N., Roberts, J. C., Roberts, P. J., Rodgers, P. J., A Coordination Model for Exploratory Multi-view Visualization, Proceedings of the Conference on Coordinated and Multiple Views in Exploratory Visualization, 2003)

Figure B.15 has been removed due to copyright restrictions. It was a multiple view of many 2D vertical bar charts. Original source: Pirolli, P., and Rao, R. (1996). Table Lens as a Tool for Making Sense of Data. Proceedings of the Workshop on Advanced Visual Interfaces: Figure 1, pp. 69

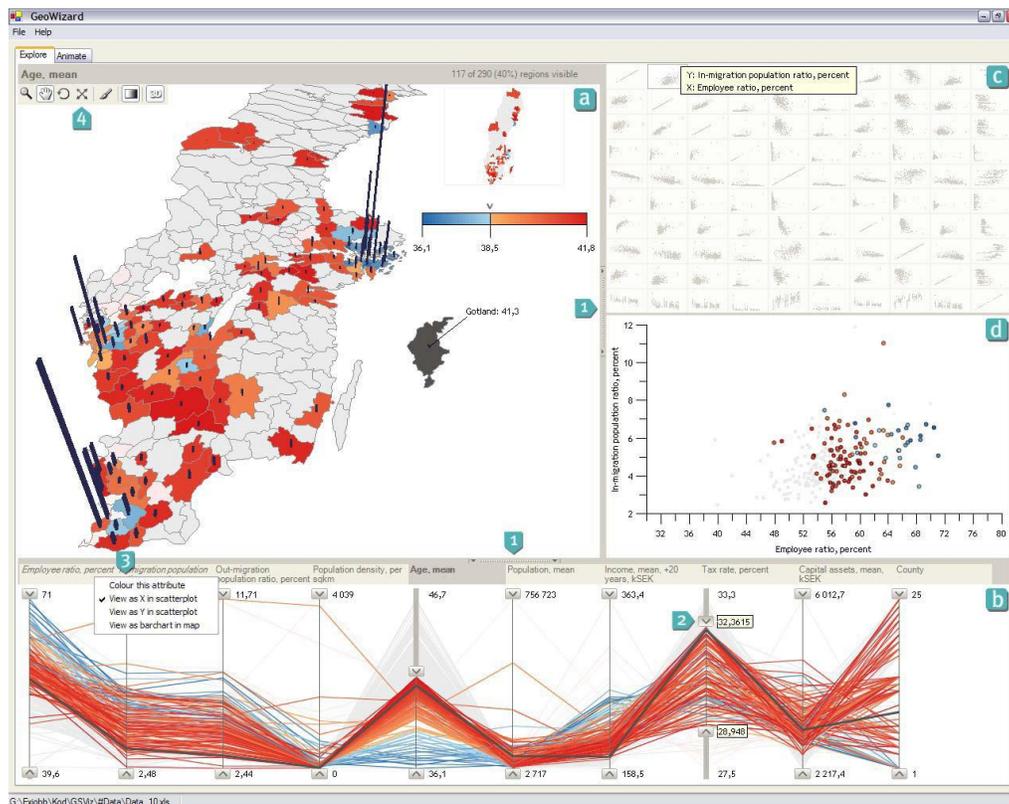


Figure B.16 A multiple view of a choropleth map, a parallel coordinates chart, a scatter plot matrix, and a scatter plot for users to explore insights from different angles in a demographic dataset. (© IEEE 2005. Reprinted, with permission, from Feldt, N., Pettersson, H., Johansson, J., Jern, M., Tailor-made Exploratory Visualization for Statistics Sweden, Proceedings of the Coordinated and Multiple Views in Exploratory Visualization, 2005)

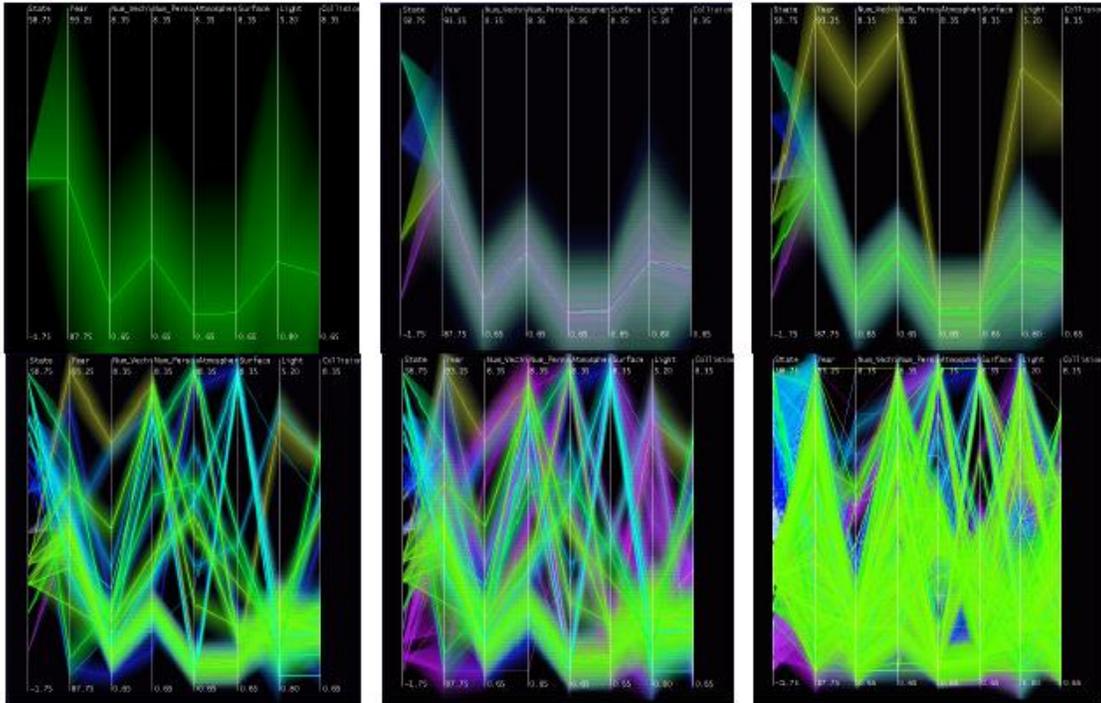


Figure B.17 Parallel coordinate representations for visualizing multi-dimensional fatal accident data at different levels of detail. (© IEEE 1999. Reprinted, with permission, from Fua, Y., Ward, M. O., Rundensteiner, E. A., Hierarchical Parallel Coordinates for Exploration of Large Datasets, Proceedings of the Conference on Visualization '99, 1999)

Figure B.18 has been removed due to copyright restrictions. It was a "same bin size Mosaic Plot" that represents the data about answers to the question of "how you heard about survey". The degree of grey shades represents counts of responses to certain answer options. Original source: Hofmann, H. (2006). Multivariate Categorical Data-Mosaic Plots. Graphics of Large Datasets- Visualizing a Million: Figure 5.13, pp. 120

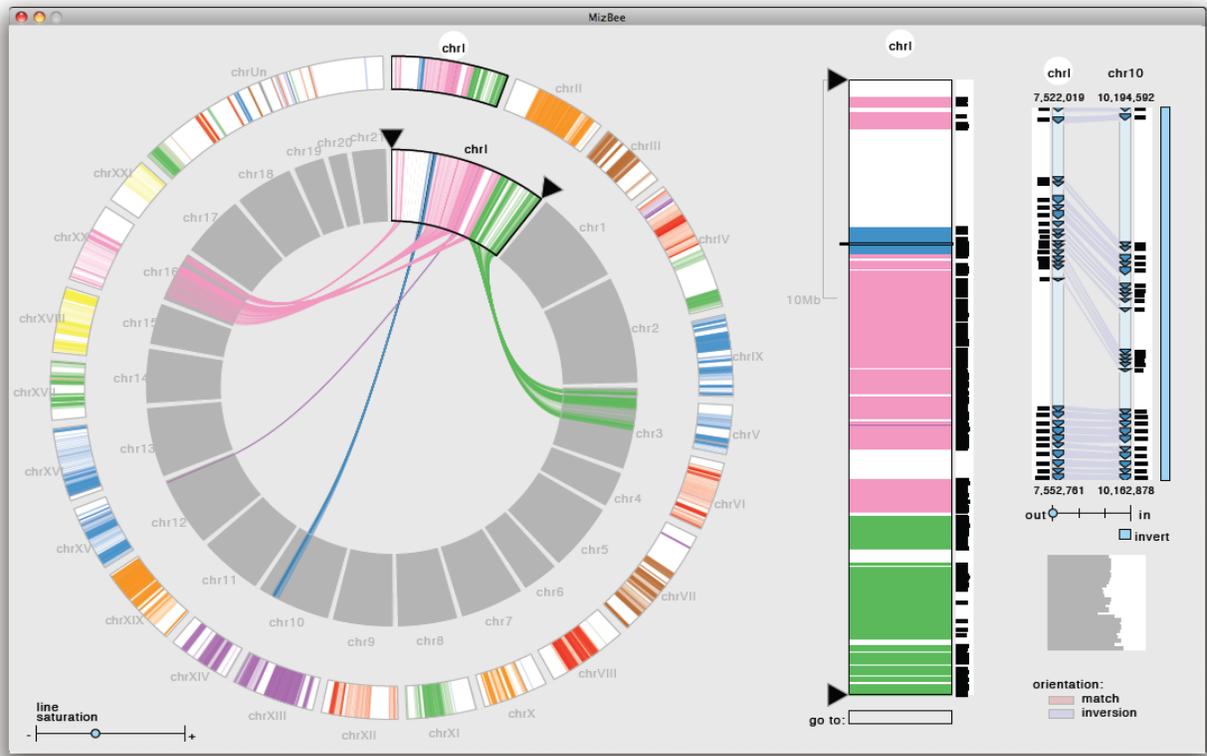


Figure B.19 A multiple view of three images that shows the biological data of stickleback and pufferfish with regard to synteny relationship at the : (a) genome, (b) chromosome, and (c) block level. (© IEEE 2009. Reprinted, with permission, from Meyer, M., Munzner, T., Pfister, H., MizBee: A Multiscale Synteny Browser, IEEE Trans. Visual. Comput. Graphics 15(6), 2009)

Figure B.20 has been removed due to copyright restrictions. It was a visualization called "Table Lens" that allows users to view detailed values (i.e. focus) and overall patterns (i.e. context) simultaneously. Original source: Rao, R., and Card, S. K. (1994). The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus + context Visualization for Tabular Information. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Color Plate 1, pp. 481

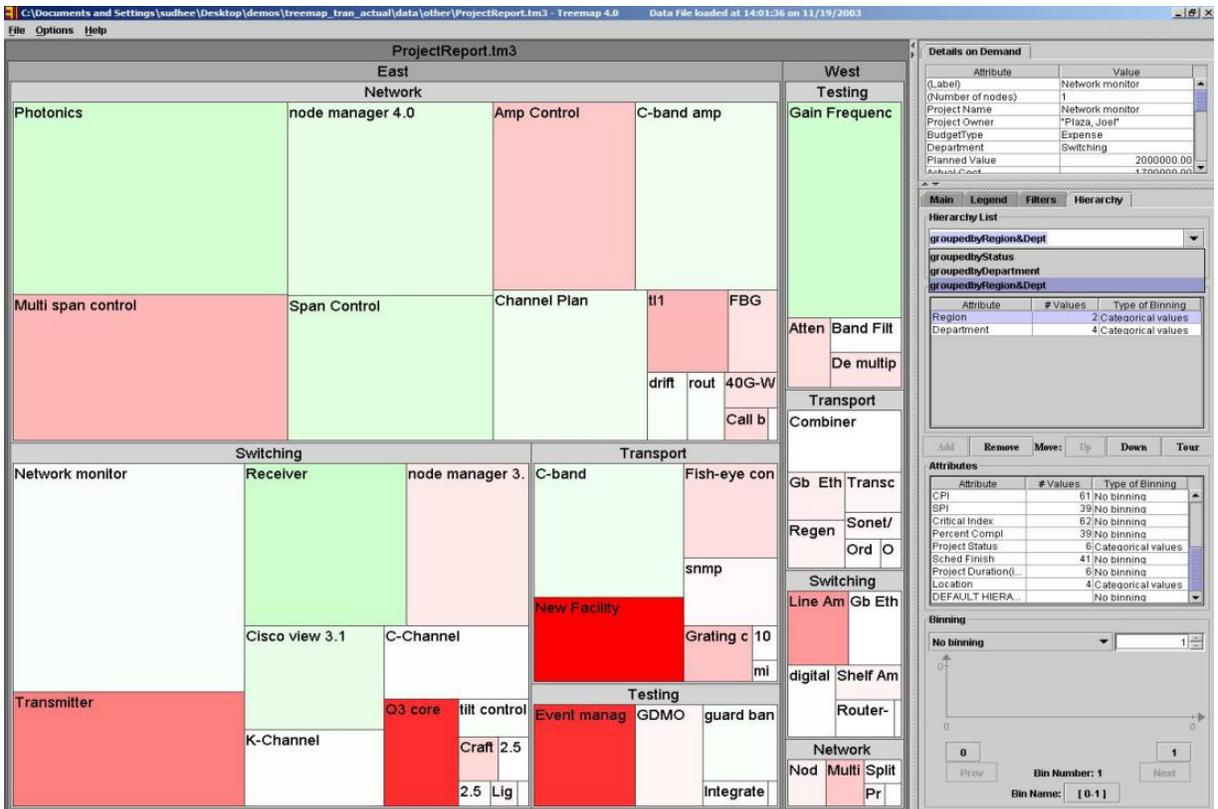


Figure B.21 A tree map representation that shows planned values (rectangle sizes) and cost variances (colors) of both individual projects and project groups (grouped by departments). (© IEEE 2004. Reprinted, with permission, from Chintalapani, G., Plaisant, C., Shneiderman, B., Extending the Utility of Treemaps with Flexible Hierarchy, Proceedings of Eighth International Conference on IV, 2004)

Figure B.22 has been removed due to copyright restrictions. It was a graphics showing a non-Euclidian geometry providing a smoothly varying focus+ context for visualizing hierarchical data. Original source: Lamping, J., and Rao, R. (1996). Visualizing Large Trees Using the Hyperbolic Browser. CHI '96: Proceedings of Conference Companion on Human Factors in Computing Systems: Figure 1, pp. 388

Figure B.23 has been removed due to copyright restrictions. It was a graphics showing the use of 3D virtual space to house 2D and 3D graphics for visualizing data from Statistics Canada. Original source: Brath, R. (2003). Paper Landscapes: A Visualization Design Methodology. Proceedings of Conference on Visualization and Data Analysis (VDA 2003): Figure 4(right), pp. 131.