THE EFFECT OF VISUAL REPRESENTATIONS OF LARGE QUALITATIVE DATA SETS ON DECISION-MAKING PROCESSES

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Meena Natarajan

May 2009

 ABSTRACT

This study investigated the influence of visual representation of qualitative data at progressively greater levels of abstraction on decision-making processes in order to address a gap in research that currently focuses predominantly on the final choice phase of decision-making and representation of quantitative information. Specifically, this research investigated how four forms of data representation, varying progressively in their use of visualization and data abstraction, compare in the effort required to arrive at a decision, the ease with which themes are identified, satisfaction with the level of detail obtained, the confidence in decisions made, and the intuitiveness of representations.

The experimental design closely simulated a real world decision-making scenario with a decision-making task developed in consultation with industry experts and a large qualitative dataset obtained from a survey on workplace environmental design conducted by a large global company. Qualitative data can be open to

interpretation and final decisions or conclusions can be difficult to evaluate for accuracy. Therefore, decisions were compared across the four conditions to see how varying the level of visual abstraction of data representation encouraged participants to focus on certain themes versus others and to arrive at their decisions.

The results indicate that in a matter of an hour, most participants identified key themes in the data regardless of the level of abstraction of the visual representation. However, there were significant differences in operational effort required, the intuitiveness of representations, and a marginaly significant difference in the ease with which themes were identified. Participants' reported satisfaction with the level of detail reached and ratings of confidence in decisions-made were low or neutral and at odds with their objective performance. Insights into the effect of the visual representations and the subjective experience of the decision-maker are discussed.

BIOGRAPHICAL SKETCH

Meena Natarajan was born in south India. She was raised in various places in India and the UK. The exposure to diverse cultures led to a deep interest in the role of the enviroment in development, and a strong inclination towards making connections between different disciplines. She completed her Bachelors degree in Commerce and Master's in Organizational Psychology from the University of Madras in 2002 and 2005 respectively. The graduate program in the Department of Design and Environmental Analysis taught her the value of conducting research that is grounded in real world problems and made accessible to all. This project is a first step towards her goal of contributing to the development of tools that make research universally accessible and influential in building validating environments.

To Amma, Appa and Shankar with love

ACKNOWLEDGMENTS

I am grateful for the generosity of many people in this most memorable graduate school journey. Greatest thanks to my advisor, Professor Alan Hedge for caring about each of my interests, even when they were numerous and contradictory, and for creating opportunities for me to discover meaninful ways to bring them together. Your expertise and passion for research has been a true inspiration. Special thanks to Professor Kathleen Gibson, my minor committee member, for her invaluable guidance. Your kind encouragement was an important source of motivation. Thank you to Nancy Wells for her cheerful support and sense of humor. Thank you to Gary Evans for his kindness and belief in me. DEA was a most special place because of all of you.

I am indebted to Mike O'Neill and Patricia Bergquist for an incredible internship experience and for continuing their support by making this project happen. Special thanks to Pierre Kergall, Greg Theobald, David Lantrip, JC and Florio Arguillas for their generosity and expertise. I greatly value the time you took to share your technology and knowledge. I learned a lot from all of you.

Thank you Anton, Vincent, Anshu, Sarah and Jeanne for being such good friends and always making me laugh. Special thanks to Robert Rodgers and A.J Rubineau for their care. There were many hurdles in the last few years and I am grateful to all of you for your support. Thank you most of all to my family for love and encouragement.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH……...……iii

DEDICATION……..……iv

ACKNOWLEDGEMENTS…….……..……v

TABLE OF CONTENTS……....……vi

LIST OF FIGURES……...…....x

LIST OF TABLES..……...……xiii

CHAPTER 1. INTRODUCTION.....1

1.1 Background.....3

1.2 Sensory and Arbitrary Codes.....3

1.3 Visual Representations as External Cognition Aids.....4

1.4 Textual and Graphical Representations..... 5

1.5 Classification of Visual Representations.....6

1.5.1 Scientific Visualization.....7

1.5.2 Information Visualization.....8

1.5.3 Knowledge and Collaborative Visualization.....11

1.6 Classification of Visualization Systems.....13

1.7 Data.....13

1.7.1 Data Size.....14

1.7.2 Data Dimensions.....14

1.7.3 Task by Data Type Taxonomy.....14

1.7.3.1 Data Types.....14

1.7.3.1.1 Tree/Hierarchical.....15

1.7.3.1.2 Network.....16

1.7.3.1.3 Temporal.....17

1.7.3.2 Tasks.....18

1.8 Macro/Micro View of Data.....19

1.9 Continuum of Visualizations.....21
1.10 Decision-Making Theory.....23
1.10.1 Sequential Models of Decision-Making.....24
1.10.2 Non-Sequential Models of Decision-Making.....25
1.11 Information Search Behavior.....26
1.12 Cognitive Incentives Systems.....27
1.13 Research Questions and Hypotheses.....28
1.13.1 Objective.....29
1.13.2.1 Themes/Patterns and Levels of Detail.....29
1.13.2.2 Confidence Levels.....30
1.13.2.3 Understanding Representations.....30
1.14 Conclusion.....30
CHAPTER 2. METHODOLOGY.....32
2.1 Participants.....32
2.2 Apparatus.....32
2.2.1 Decision-Making Scenario.....33
2.2.2 Information Representations.....33
2.2.2.1 Table Visualization.....35
2.2.2.2 Table Visualization, Link Nodes and Bar Charts.....36
2.2.2.3 Spatial Graphics and Pivot Tables.....38
2.2.2.4 Abstract Visualization.....39
2.3.3 Interactions with the Data.....41
2.2.5 Data Set used in Decision-Making Task.....42
2.3 Procedure.....43
2.4 Experiment Setting.....44
2.5 Measures.....44
2.5.1 Post Experiment Survey.....44

2.5.2 Post Experiment Interview....45
2.6 Independent and Dependent Variables....45
2.6.1 Percieved Ease with which Themes were Identified....45
2.6.2 Operational Requests....45
2.6.3 Satisfaction with Level of Detail Obtained....46
2.6.4 Levels of Confidence....46
2.6.5 Number of Doubts in Understanding Representations....46
2.7 Other Variables Measures....46
2.7.1 Percieved Effectiveness of Representations....46
2.7.2 Operational Doubts....47
2.7.3 Data Doubts....47
2.7.4 Final Decisions....47
2.7.5 Comments on Decision-Making Process....47
2.8 Data Analysis....47
2.8.1 Chat Transcripts....48
2.8.2 Survey Responses....48
2.8.3 Interviews....48
2.8.4 Response Sheet....48
CHAPTER 3. RESULTS....49
3.1 Process Analysis....49
3.1.1 Operations on the Data....49
3.1.1.1 Operational Requests....49
3.1.1.2 Operational Doubts....50
3.1.2 Understanding Representations....51
3.1.3 Questions about the Data....52
3.2 Perception of Decision-Making Process and Information Representations..53
3.2.1 Level of Analysis....53

3.2.2 Level of Confidence in Decisions.....54
3.2.3 Perceived Effectiveness of Representations for Decision-Making.....56
3.2.4 Condition Specific Statements.....57
3.2.4.1 TABLE.....58
3.2.4.2.TEXTA.....58
3.2.4.3 IRD.....58
3.2.4.4 TIM.....59
3.3 Interview.....59
3.3.1 TABLE.....59
3.3.2 TEXTA.....61
3.3.3 IRD.....62
3.3.4 TIM.....64
3.4 Analysis of Decisions.....65
3.4.1 Themes Identified.....65
3.4.2 Number of Themes Identified.....67
CHAPTER 4: DISCUSSION.....70
4.1 Operational Effort.....71
4.2 Ease in Identifying Themes.....72
4.3 Satisfaction with Level of Detail.....73
4.4 Level of Confidence.....74
4.5 Understanding Representations.....75
4.6 Strategies, Credibility and Quantitative Analysis.....76
4.7 Limitations.....79
4.8 Future research.....80
4.9 Conclusion.....81
APPENDIX A. Management Decision Making Scenario.....83
APPENDIX B. Response Sheet Results.....85

APPENDIX C. Transcript Analysis Results.....86
APPENDIX D. Survey.....87
APPENDIX E. Survey Results.....90
APPENDIX F. Sample Chat Transcripts.....92
REFERENCES.....122
IMAGE REFERENCES.....129

LIST OF FIGURES

Figure 1.1 Fire Spread Simulation…….……7
Figure 1.2 NameVoyager graphs names starting with ‘kr’……..…..10
Figure 1.3 NameVoyager graphs names starting with ‘w’……...10
Figure 1.4 Visualization of the article ‘Chocolate’ on Wikipedia based on version history……....11
Figure 1.5 Figure 1.5 A concept map of concept maps from CmapTools……12
Figure 1.6 Multiple users in VR-VIBE, a virtual reality information retrieval environment……..12
Figure 1.7 A node-link representation of hierarchical data……..15
Figure 1.8 Map of the Market……...16
Figure 1.9 Collaboration at Ars Electronica visualized using network……17
Figure 1.10 Tufte’s sparklines illustrating trends in medical information……18
Figure 1.11 Fish-eye display of linear data from the University of Maryland…….20
Figure 1.12 Continuum of data visualization……22
Figure 2.1 Table Visualization.....35
Figure 2.2 TEXTA Interface.....36
Figure 2.3 TEXTA Category Bar.....37
Figure 2.4 TEXTA Category Web.....37
Figure 2.5 IRD Overview Map.....38
Figure 2.6 IRD Pivot table.....39

Figure 2.7 TIM Interface.....40

Figure 3.1 Mean Operational Requests.....49

Figure 3.2 Mean Operational Doubts.....50

Figure 3.3 Mean Interface Doubts.....51

Figure 3.4 Mean Data Doubts.....52

Figure 3.5 Mean scores on survey for level of analysis.....53

Figure 3.6 Mean scores on survey for confidence in decisions.....55

Figure 3.7 Mean scores on survey for effectiveness of representations.....56

Figure 3.8 Percentage of subjects identifying themes.....66

Figure 3.9 Proportion of subjects in each condition ranking meeting as top priority...66

Figure 3.10 Mean number of global themes identified.....67

Figure 3.11 Mean number of UK themes identified.....67

Figure 3.12 Mean number of Philippines themes identified.....68

Figure 3.13 Mean number of Australia themes identified.....69

LIST OF TABLES

Table 2.1 Features of each condition based on Schneiderman's information seeking tasks.....34

1. INTRODUCTION

The form in which information is presented to decision-makers has a powerful influence on real world policy decisions (Kuo, 2002). Qualitative research, with techniques such as focus groups, observations and interviews, produces data that is rich in insight but typically time consuming and difficult to analyze and present to others, even with small sample sizes. Technologies, such as the internet, data mining tools and information retrieval systems, provide an opportunity to collect and analyze data on a larger scale, with participants from around the world, and recent computer software advances now allow for the systematic analysis of qualitative research. The end users of such data include social scientists, managers and professionals from a wide range of disciplines, and consequently the user interface and data representations need to facilitate efficient analysis and decision making (Speier & Morris, 2003) in an intuitive manner.

> "...in an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it" (Simon, 1971)

Herbert Simon (1996) pointed out that because information systems were originally designed to address a problem of information scarcity, designers had to find ways to present larger amounts of information at a time, whereas today these systems need to address an attention scarcity and should instead be built to filter out irrelevant information and quickly focus attention on important aspects. This need has led to a growing number of studies on information representation in areas such as data mining, statistics, learning, pattern recognition, and computer graphics (Kreuseler &

Schumann, 2002). One of the most powerful techniques that recently has arisen out of all this research is visual data mining (Wong, 1999). Data mining is the process of uncovering "patterns or models from observed data, usually as part of a more general process of extracting high-level, potentially useful knowledge, from low-level data" (Oliveira & Lewkovitz, 2003). Visual data mining or visual data exploration brings together information visualization and the analytical process into a single tool (Badjio & Poulet, 2005). This model of data exploration capitalizes on strong human spatial and visual abilities, which allow for retrieval and comprehension of information using visual cues such as motion, color and shape (Kosslyn, 1994). The process of exploring unknown data for patterns and hypothesis generation is called exploratory data analysis and was introduced by John Tukey (1977) as a "counterbalance to the traditional statistical techniques of checking a priori selected hypotheses." (Tukey, 1977 as cited in Adrienko & Adrienko, 2003). Visual data exploration is the analytical mining of data in which results are presented in graphical form (Badjio & Poulet, 2005).

According to Keim (2002) visual data exploration provides a more intuitive interface for exploration, as the user does not need to know complex mathematical or statistical algorithms or parameters, necessary in other data mining interfaces. Keim also suggests that since the users are directly exploring the data they can shift and adjust their goals as they discern patterns and trends (Keim, 2002). Therefore, visualization is beneficial for the first stage of exploration – pattern recognition. This stage is also similar to the first step in the decision-making process: "situation analysis and problem recognition" (Simon, 1960 as cited in Hansson, 1994).

Research thus far; however, has focused mainly on developing various mining techniques and visualization algorithms or paradigms (Au et al. 2000). Anecdotal evidence suggests that visual mining tools can lead to an increase in decision quality and confidence levels (Borzo 2004; Esfahani 2005; Miller 2004, as cited in Lurie & Mason, 2007). However, there is a general lack of systematic human factors research

on user guidance (Au et al. 2000, Badjio & Poulet, 2005; Lurie & Mason, 2007) or the effect these representations have on decision-making processes, when raw textual data is represented in increasingly graphical and abstract forms. This study addresses this research gap by investigating the influence of information representation on decisionmaking processes using four information representation forms that fall on a continuum of visual abstraction, from raw data in a tabular form to a highly abstract graphical representation, a real-world, large qualitative workplace environmental design dataset and a decision-making scenario. The following sections present related research that describe visualization's role in aiding decision making.

1.1 Background

Visualization refers to the "Binding or mapping of data to representations that can be perceived. The types of binding could be visual, auditory, tactile, or a combination of these" (Foley & Ribarsky, 1994).

1.2 Sensory and Arbitrary codes

All visualizations consist of smaller visual components. These components can be classified as sensory and arbitrary (Ware, 2004). According to Ware, the sensory aspect relates to components of a representation that can be perceived without any training on the part of the viewer. Instead, sensory representations rely on the perceptual processing power of the brain to convey information. This information is perceived in an immediate manner that is difficult to change and is valid across cultures. Arbitrary aspects of a representation, however, are those that require the perceiver to learn certain conventions in order to understand the representation. These may be cultural or domain specific conventions. Written language and electronic circuit diagrams are two examples. According to Ware (2004), most representations are a combination of sensory and arbitrary aspects. Depending on the kind of information that needs to be conveyed, a representation may make use of sensory and arbitrary aspects to varying degrees. Both types of visual components play a role in amplifying cognition by reducing the load on memory.

1.3 Visual Representations as External Cognition Aids

One's ability to hold a representation of the external world in memory is limited (Ware, 2005). Studies have shown that individuals retain about three objects, in visual working memory per second (Vogel et al., 2001) and people are oblivious to more than 99% of what is in their visual field at any moment (Simons & Levin, 1998). Ware (2005) suggests that since the external world is there to perceive if needed, we do not have to hold it all in our minds. Depending on what our task is, he suggests that we make "visual queries" to the world – our visual system focuses on the external environment to gather information based on what is required. This happens with such frequency and speed that we do not notice this process consciously. Visual representations act as external cognition aids, providing a faster and more efficient way to access information than queries to our memory (Hutchins, 1995). Scaife and Rogers (1996) refer to this interweaving of external and internal representations, and processing as external cognition.

Graphical representations belong to one class of external aids that amplify cognition. They not only serve to communicate an idea but also to discover and think about the idea itself (Card et al., 1999). According to Ware (2005), the "power of a visualization comes from the fact that it is possible to have a far more complex concept structure represented externally in a visual display than can be held in visual and verbal working memories."

External representations such as diagrams also allow users to gather computational information at a glance instead of having to compute information from purely text-based representations (Scaife & Rogers,1996). They call this phenomenon "computational offloading". Our visual system is also best at the pattern recognition that visual displays facilitate (Ware, 2005). Information visualization allows the perceptual system to absorb the cognitive load (Lohse 1997; Zhang & Whinston 1995) and expand problem-solving abilities by facilitating easier analysis of large volumes of data (Tegarden 1999).

However, as will be discussed, the ease with which graphical representations are processed can sometimes also lead to cognitive bias when a combination of textual and graphical forms are used. It is also important to consider the task involved as the effectiveness of graphical forms may depend on the level of task complexity.

1.4 Textual and Graphical Representations

Research suggests that when both text and graphical information is presented, the viewer tends to focus and give more importance to the graphical information than the text (Stone, Yates, and Parker 1997). As described above, visual representations use sensory codes (Ware, 2004) of color, size and orientation, features of graphics that are processed with minimal effort by our perceptual system (Ware, 2004, Bederson & Shneiderman, 2003). Representations that use sensory codes are more salient or vivid (Nisbett & Ross 1980, as cited in Lurie & Mason, 2007). This may result in biased conclusions as users may ignore less salient or text based representations in the decision-making process (Raghubir & Krishna, 1999).

According to Vessey, the usefulness of textual or graphical representations depends on whether the task involved is spatial or symbolic (Vessey, 1991). According to Vessey, spatial information is best analyzed using graphical representations as it is easier to spot trends and patterns, and tabular representations are more appropriate for symbolic information as they aid easy identification of required data values (Vessey, 1991). In a study on query interfaces, Speier and Morris (2003) found that decision quality was high when the text-based interface was used for tasks with low complexity. However, for highly complex tasks, the visual interface enabled higher levels of decision quality. They suggest that visual interfaces are useful for analyzing general patterns and may be an effective way to make large data sets easier to process. However, their effectiveness is compromised when details need to be analyzed or when there are only a few data points as visual interfaces may require more operations in order analyze on a micro level (Speier & Morris, 2003).

Decision-making tasks different to those used in empirical studies may be less likely to fit into a specific task category. Many decision-making scenarios may involve a combination of subtasks of varying levels of complexity. Decision-makers may need to discern patterns or trends in a dataset to gain a high level overview and also understand the details in order to arrive at a final decision. Some parts of a dataset may be richer in terms of data points than others. But which type of representation or combination of representation forms would facilitate this type of a decision-making scenario? One goal behind studying the influence of visual abstraction of that data is to explore whether there is an optimal point or a specific combination of representations that facilitate an effective decision-making process.

In order to better understand research in visualization, information seeking and decision-making, which are relevant to this study, the following sections present visualization, data and task classifications, information seeking research, decisionmaking theory and finally, the cognitive incentive system by which individuals decide on a specific strategy for a decision-making task. The hypotheses derived from the review of related research are then presented.

1.5 Classification of visual representations

Visual representations are broadly classified as scientific or information visualizations based on the kind of data they represent. When the intention behind visualization is to communicate expert knowledge or share information in groups, they are classified as knowledge and collaborative visualization. These categories are not mutually exclusive but overlap and offer broad ways to think about representations. Scientific visualization pertains to the visual representation of physical phenomena (Oliveira & Levkowitz, 2003) whereas information visualization refers to the visual representation of abstract data that has no obvious physical form.

1.5.1. Scientific Visualization

A 1987 report by the National Science Foundation's Advisory Panel on Graphics, Image Processing, and Workstations led to the birth of scientific

visualization as a discipline initially known as “visualization in scientific computing”. In scientific visualization, data pertaining to objects or concepts in the physical world are graphically represented (Oliveira & Levkowitz, 2003) to facilitate deeper understanding of those physical phenomena. For example, Figure 1.1 is one image from a simulation rendered from temperature data of a 10-m heptane pool fire. This simulation allows researchers to observe the roll-up of vortices and better understand how large-scale fires spread.

Figure 1.1 Fire Spread Simulation, Phil Smith, Rajesh Rawat, James Bigler, Center for the Simulation of Accidental Fires, Explosions (C-SAFE).

This kind of rendering acts as an external cognition aid by allowing researchers to view a phenomenon that would otherwise be difficult to observe or imagine. Other examples of scientific visualization including medical information and weather patterns. The development of scientific visualization is closely related to progress in many scientific fields (Card et al., 1999). As scientists placed greater value on empirical approaches to scientific investigation versus a theoretical approach, the use of graphical representations has increased in order to represent and facilitate the interpretation of data (Wainer & Velleman, 2001). In the mid eighties and early nineties, satellites sent back extremely large sets of data and the artificial intelligence and computer graphics fields focused on the automatic design of graphical

representations of these kinds of data. Visualization was found to be useful in analyzing and understanding this data (Card et al., 1999).

Following these initial developments in scientific visualization, the user interface field saw opportunities for new kinds of user interfaces using visualization. The task that this was most pertinent to was that of user navigation of large multivariate databases. The emphasis of researchers in this area was not so much on the quality of the graphics but on the animation and interactivity that amplified cognition (Card et al., 1999). Card et al suggest that the interaction of the different fields and communities that were interested in visualization as a tool to enhance insight generation in their respective fields, led to the field of information visualization.

1.5.2 Information Visualization

Cognitive psychologists use "information visualization" as a broad term to include different kinds of visualization or a representational mode that describes specific information in a "visual spatial manner" (Tergan & Keller, 2005). However, in computer science and in the context of this study, the term is used specifically for "computer supported interactive, visual representations of abstract non-physically based data to amplify cognition" (Card, Mackinlay & Shneiderman, 1999). According to Voigt (2002), “abstract data has no inherent mapping to space”. Results of a survey, responses to open-ended questions, database of staff information such as academic background, unstructured text and interview transcripts are some examples of abstract data.

Information visualization as understood in computer science is especially pertinent when analyzing data sets that are extremely large (Carr, 1999). Figure 1.2 and 1.3 are screenshots from an online visualization tool of baby names called ‘Name Voyager’ (Wattenberg, 2005). The interactive tool allows users to type in letters and displays a graph showing the popularity of names starting with those letters over the last century. This tool intended initially for parents soon became a popular social

analysis tool based on its display of social trends. Figure 1.2 shows that names starting with the letters 'KR' peaked in popularity during the 1970s and Figure 1.3 shows that there has been a steady decline in the popularity of names starting with the letter 'W' in general. However, the name 'William' ranked 8th in 2007 continues to be popular.

Another example of information visualization is the IBM's 'History Flow'. This visualization tool allows users to view the changes made to a Wikipedia article over a period of time or over different versions. Figure 1.4 shows the different versions of the article 'Chocolate' (Viegas, Wattenberg, & Dave 2004). The zigzag pattern reveals an edit-war over the existence of a surrealist sculpture related to this topic (Viegas, Wattenberg, & Dave 2004). Visualization helps easily identify vandalism and anonymous edits common to Wikipedia. This tool also allows users to see all the authors of the page and the article in its current state.

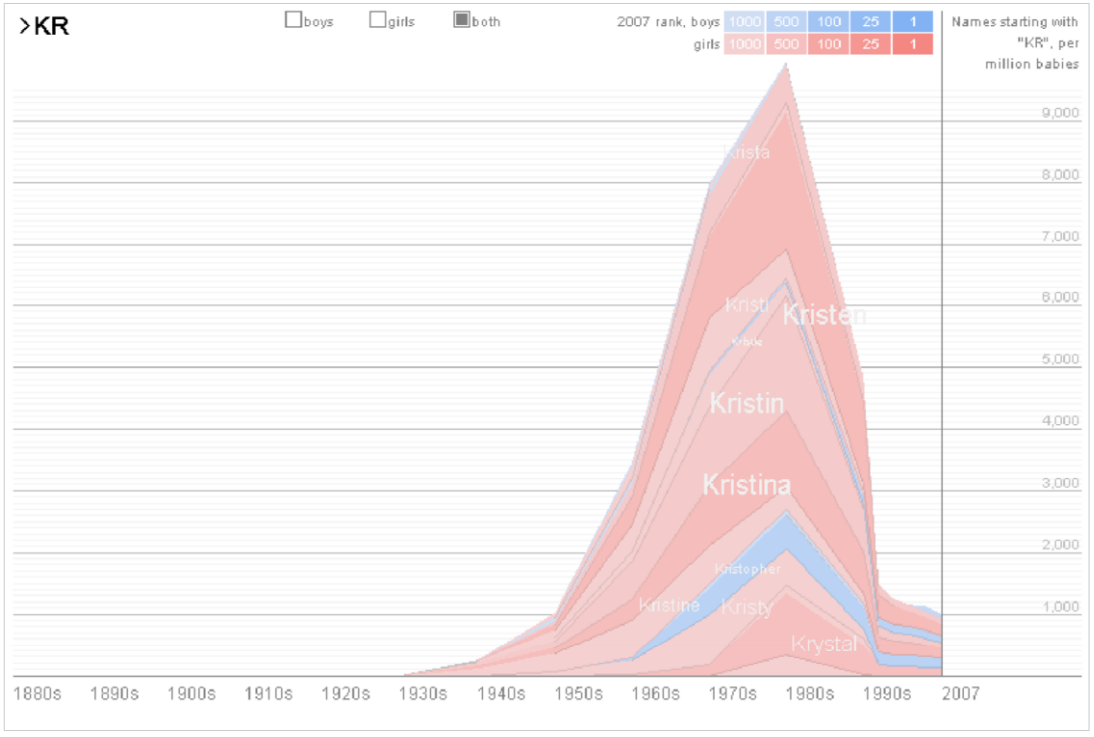


Figure 1.2 NameVoyager (Wattenberg, 2005) graphs names starting with 'kr'

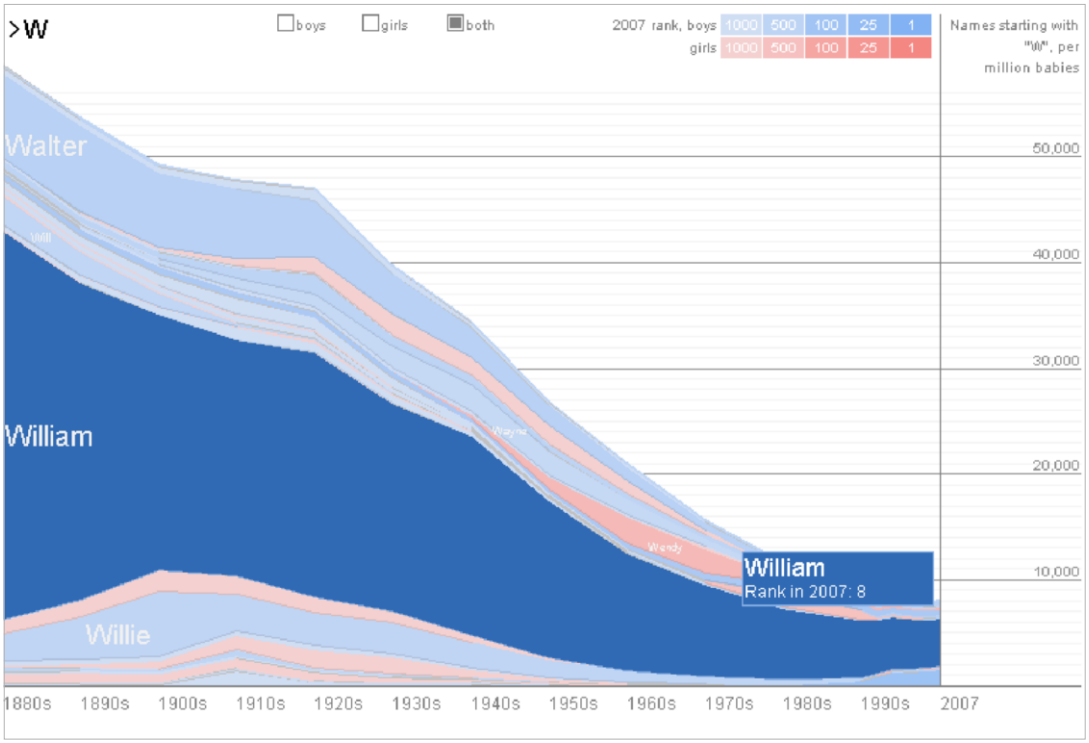


Figure 1.3 NameVoyager (Wattenberg,2005) graphs names starting with 'w'

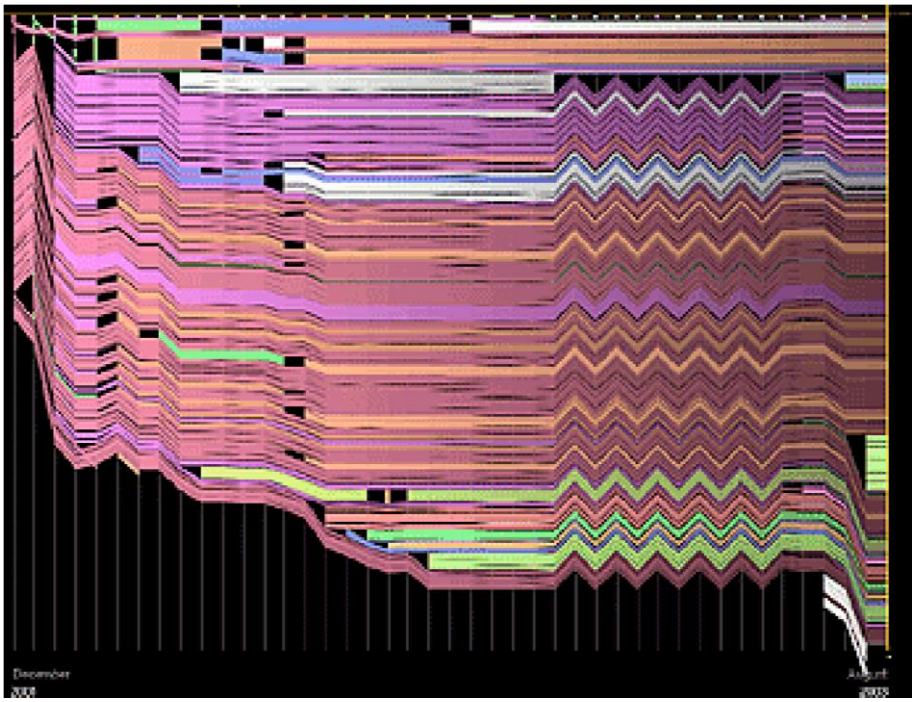


Figure 1.4 Visualization of the article 'Chocolate' on Wikipedia based on version history. The zigzag pattern reveals an edit-war.

1.5.3. Knowledge and Collaborative Visualization

Knowledge visualization is used when the knowledge structure of an expert is represented through visualization to students, intended to serve as a tool for selfassessment of knowledge and to facilitate comprehension and navigation (Dansereau, 2005). Here emphasis is placed not just on the transfer of facts but also on the communication of insights, experiences, and predictions. Examples of knowledge visualization include mind maps and concept maps (Jonassen et al., 1993). Mind maps refer to the use of keywords and images as a spatial strategy to aid note taking and structuring of ideas (Buzan, 1995). Concept visualizations or maps allow users to map, manage and manipulate concepts (Canas, Leake & Wilson, 1999). Collaborative knowledge visualizations are based on the belief that sharing of knowledge between different communities contributes to innovation (Novak & Wurst, 2005). Figure 1.5 is a concept map of concept mapping from CmapTools. Figure 1.6 is an image of a virtual reality information retrieval environment in which users collaboratively explore a document describing virtual reality research (Benford et al., 1995).

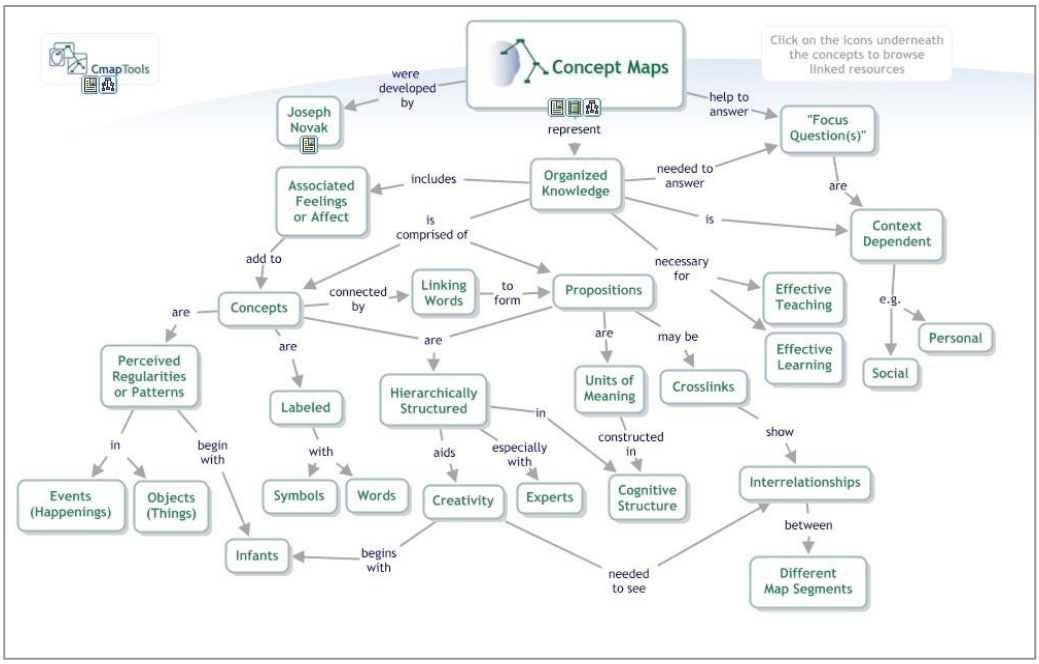


Figure 1.5 A concept map of concept maps from CmapTools

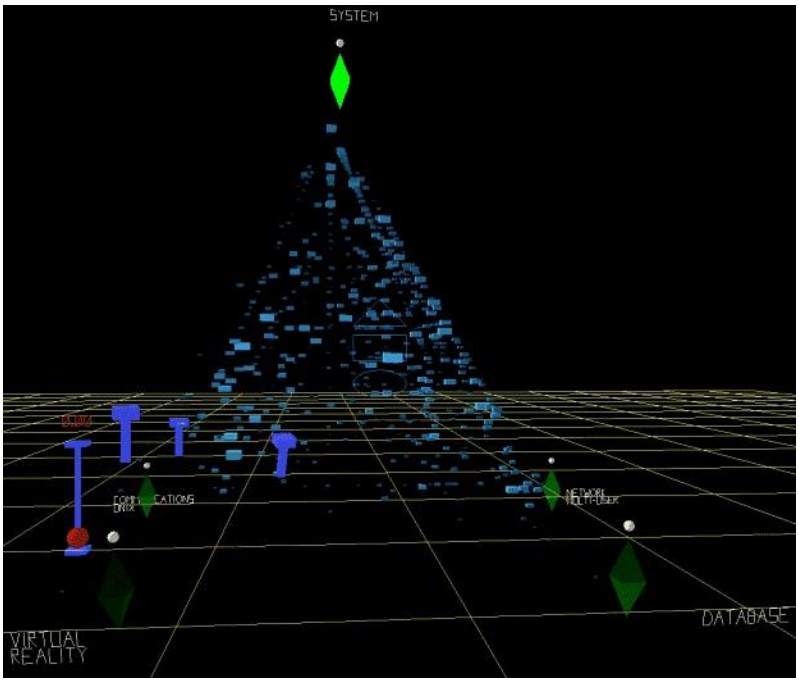


Figure 1.6 Multiple users in VR-VIBE, a virtual reality information retrieval environment

1.6. Classification of Visualization Systems

Card et al. (1999) classify visualization systems in the following four categories based on levels. In the first level are visualizations that enable a user to access information outside of their workspace through the internet or online databases. On the next level are visualizations that allow more efficient use of an information workspace. On the third level are visualizations which allow for data to be represented with a set of interactive tools for pattern finding and discovering relationships in the data. This level relates to the focus of most data mining tools. At the fourth level are visualizations that represent a physical object in order to better understand their structure or workings. This also corresponds to the scientific visualization classification. The focus of this research project is on the visual representation of qualitative data or information visualization using interactive tools that Card et al (1999) classify as the third level of visualization systems. The abstract data represented in information visualization can be further categorized based on type and size. The following section describes the different classification of data.

1.7 Data

Data refers to facts or symbols that are not interpreted or analyzed. By itself, data is raw and without meaning (Tergan & Keller, 2005). When data undergoes some analysis or interpretation it becomes Information. This may be through relating it to other variables or placing it in a specific context (Tergan & Keller, 2005). All visualization software process raw data in some form before the representation is delivered (Lurie & Mason, 2007). Data can be classified according to size, number of dimensions, and type. In data-mining and information visualization, the following classifications of data and the tasks involved in their analysis are most pertinent.

1.7.1 Data Size

In data-mining contexts a dataset is considered large if it contains around 100,000 items. Around a million data items make up a very large dataset and any data set with more than a million classified as massive. Most data mining software focus on representing massive data sets (Oliveira & Lewkovitz, 2003).

1.7.2. Data Dimensions

Oliveira & Lewkovitz (2003) categorize data arbitrarily as low dimension when it contains up to 4 attributes, medium when there is anywhere between 5 to 9 attributes and high for data sets with ten or more. They suggest that a representation of more than 4 dimensions can be overwhelming and too complex for our perceptual abilities. For most people there may be no difference between analyzing a representation of a 5 dimension or 50 dimension dataset (Oliveira & Lewkovitz, 2003).

1.7.3 Task by Data Type Taxonomy

In order to classify existing prototypes and guide future research in information visualization, Shneiderman (1996) developed the widely cited “task by data type taxonomy” for information visualization. He delineated seven data types and seven

tasks involved in the analysis of representation of all data types. These are described below.

1.7.3.1 Data Types

These first four classifications are dimension based. The first type is linear or one dimensional datasets. Typical examples include a list of names, programming code and textual documents. The second type is two-dimensional data. The third type is volumetric or three dimensional data which refer to objects in the real world and the data about their physical form. Visualization of this kind of data is classified as scientific visualization as previously discussed. The fourth data type is multidimensional as discussed in the Oliveira & Lewkovitz classification above.

1.7.3.1.1 Tree/Hierarchical

The fifth type, tree or hierarchical data refers to data that has a hierarchical structure. Each item has a parent and a child. The last level of children or the roots will not have children of their own. Common tasks with hierarchical or tree structured data are identifying the number of levels, the children under a specific parent and the number of children (Schneiderman, 1996). Visualization techniques for hierarchical data include node-link diagrams as illustrated in figure 1.7, and tree maps.

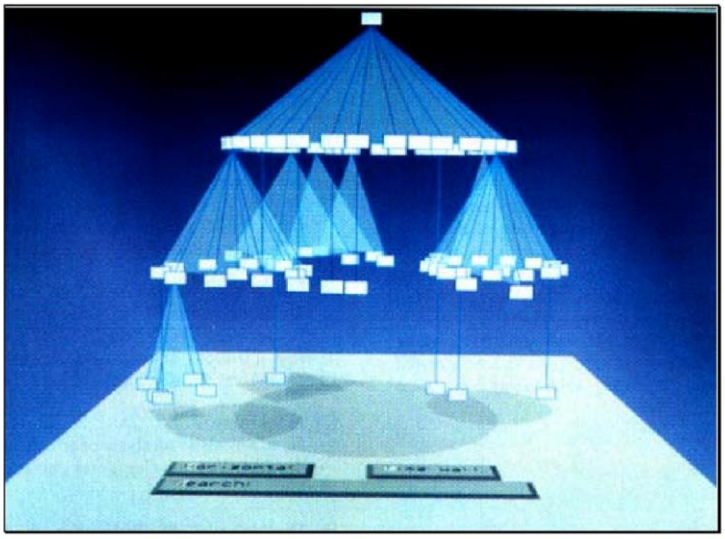


Figure 1.7 A node-link representation of hierarchical data (Robertson et al., 1991)

In order to optimize the use of space, hierarchical data is often represented using

tree map or mosaic representations as developed by Shneiderman (1992). Figure 1.8 is a modified version of the tree map also called a “change tree map” from smartmoney.com. The ‘Map of the Market’ was designed by Wattenberg (1999) to show viewers the state of the market. Each rectangle represents a company. The size of the rectangles reflects the market capitalization of the company. The color indicates the changes in stock prices with increase in value shown in green and a decrease in red. All companies fall under market categories such as capital goods, technology, utilities, energy etc (Wattenberg, 1999). The Map of the Market was found to be difficult for new users to learn (Bederson & Shneiderman 2003) and viewers typically take around ten to twenty minutes to process complex treemaps (Schneiderman, 1996). Treemaps remain a popular representation form and the Map of the Market set the standard for many financial representations (Wattenberg, 1999).

Figure 1.8 Map of the Market (Wattenberg,1999)

1.7.3.1.2. Network

Network data refers to data where each item may be linked to a number of other items. Social network data is an example of this type of data. Figure 1.9 is a social network visualization of the collaboration between artists and scientists at the Ars Electronica conference from its inception to 2004. Each dot denotes an individual. The colors indicate the year. The visualization shows that there has been more intense collaboration in recent years.

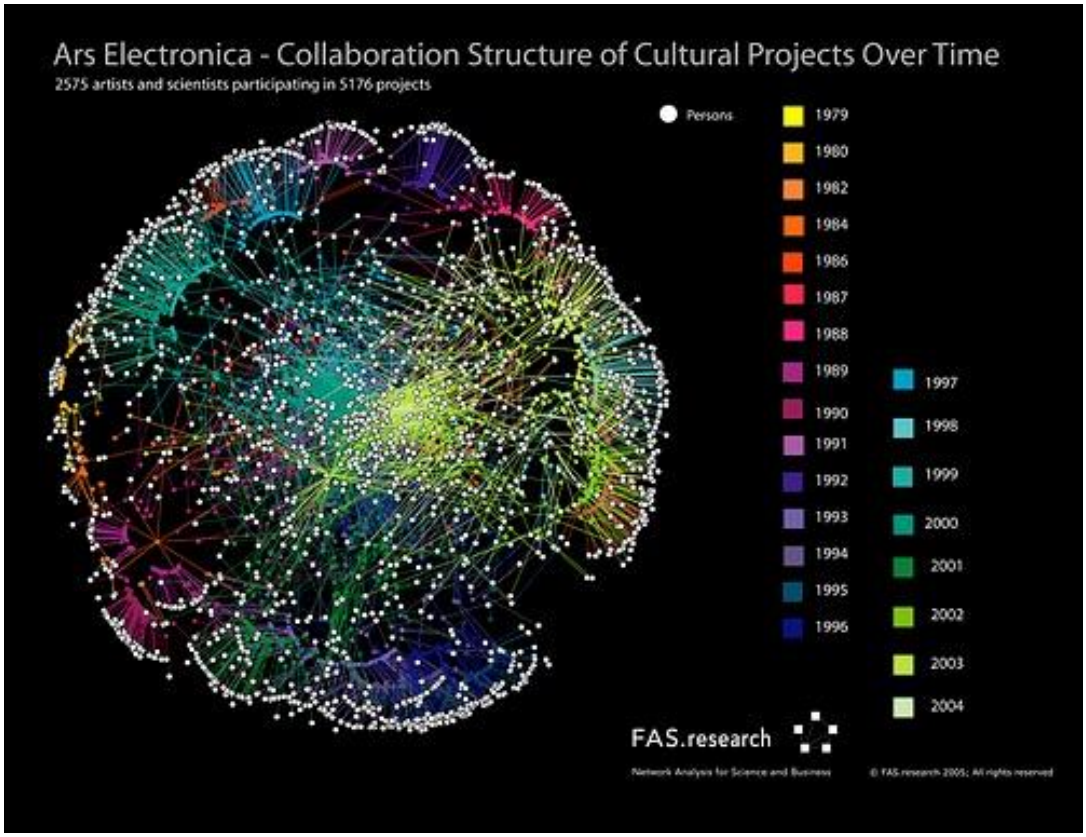


Figure 1.9 Collaboration at Ars Electronica visualized using network.

1.7.3.1.3. Temporal

Finally, temporal data refers to timelines or items that have a start and finish time. Temporal data is important in medicine, history and finance (Shneiderman, 1996). Figure 1.10 illustrates changes in medical information for a patient using sparklines. Sparklines are "data-intense, design-simple, word-sized graphics" (Tufte,

2006).The grey bar shows the normal range for each measure. The red dot at the end of the spark line is the most recent data point.

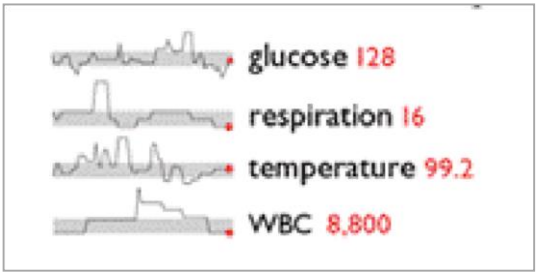


Figure 1.10 Tufte's sparklines illustrating trends in medical information

1.7.3.2 Tasks

The tasks that Schneiderman (1996) suggests as important in analysis of all types of data are described below.

- *Overview* – With an overview, the user obtains a high level understanding of a dataset using a visual representation of the entire collection of items in the dataset.
- *Zoom* – By zooming, the user views a subset or an individual item of the dataset
- *Filter* - With filtering, data that is irrelevant or unnecessary to the overall task is removed.
- *Details-on-demand* - Here, the user needs to obtain detailed information about a subset of the data
- *Relate* – The relationship between different subsets and/or the subset in relation to the whole dataset may need to be analyzed.
- *History* - Information about the interactions with the dataset may need to be accessed in order to undo or replay.

- *Extract* - This refers to extracting relevant information or subsets for further detailed analysis.

The overview, zoom, details-on-demand and relate tasks all require representations that provide a macro as well as micro view of the data. Whether this is done simultaneously on one screen or separately has implications for cognitive load. This is an important aspect of visualization and is described in greater detail below.

1.8 Macro/Micro View of Data

In order to gain a high level understanding of the data, users need to view a macro level representation or overview. This is also important when trying to analyze the relationships between subsets or a subset's relationship to the whole dataset. According to Tufte, often, the visual task is to contrast, compare and make a choice (Tufte, 1990). In these cases Tufte suggests that it is a combination of both macro level overview and micro level details that make for effective information design. Since our capacity to hold information in our visual memory is limited (Ware, 2005), "the more relevant information there is within eye span, the better" as the display will allow for viewers to rely less on their visual memory to compare and contrast information in order to make a decision or choice (Tufte, 1990). The micro/macro level displays allow for easy context switching between global and local comparison, which is essential for good reasoning about information (Tufte, 1990). Card et al (1999) refer to representations that facilitate this as providing a "focus +context". "Focus+Context start from three premises: First, the user needs both overview (context) and detail information (focus) simultaneously. Second, information needed in the overview may be different from that needed in detail. Third, these two types of information can be combined within a single (dynamic) display, much as in human vision" (Card et al., 1999)

The following figure illustrates a visualization technique called 'fish-eye view' (Sarkar & Brown, 1994) used to provide a focus and context to the viewer while working with linear data. As the cursor move over an area in the list, it expands

showing the details of the information contained in that area. The fisheye menu allows the user to gain an understanding of details while maintaining awareness of the entire list and its contents. Hornbæk and Frøkjær (2001) found that reading speed increases with the use of fish-eye views.

Figure 1.11 Fish-eye display of linear data from the University of Maryland

Similar to Fish-eye views are bifocal views in which information that is central is enlarged and information that is peripheral is reduced in size (Robertson &

Mackinlay 1993). Another strategy is to present information in two windows, with one providing an overview and the other, detailed information (Beard & Walker, 1990). Research suggests that satisfaction increases when users have access to a visual overview (Hornbæk, Bederson, & Plaisant, 2002) and data is better understood when both overview and detailed views are provided (Hornbæk & Frøkjær, 2001). However, it was also found that the need to process two views simultaneously is taxing and slows down navigation (Hornbæk et al., 2002) and that the speed with which questions are answered is faster when only a detailed view is provided (Hornbæk & Frøkjær, 2001).

1.9 Continuum of Visualizations

The visualizations discussed all use different combinations of sensory and arbitrary components to represent data in varying levels of abstraction (Ware, 2004). We can think of information representations in the qualitative data mining context as falling on a continuum of visualizations, from raw data on the one extreme to highly abstract visualizations of that data on the other (Figure 1.12). It is important to research how different points or levels in this continuum will influence the decisionmaking process, whether there may be an optimal point for a decision-making scenario that involves tasks of varying levels of complexity or whether different representations may be more appropriate for each subtask. This study compares two extremes and two intermediate levels of visual abstraction, from tabular representation with some graphic such as bar charts, or network diagrams, and pivot tables to abstract spatial representations. These levels or points are illustrated in figure 1.12 using four software packages: Microsoft Excel (TABLE), SPSS Text Analysis (TEXTA), Intelligent Results Discover (IRD) and finally Eaagle Information Mapper (TIM).

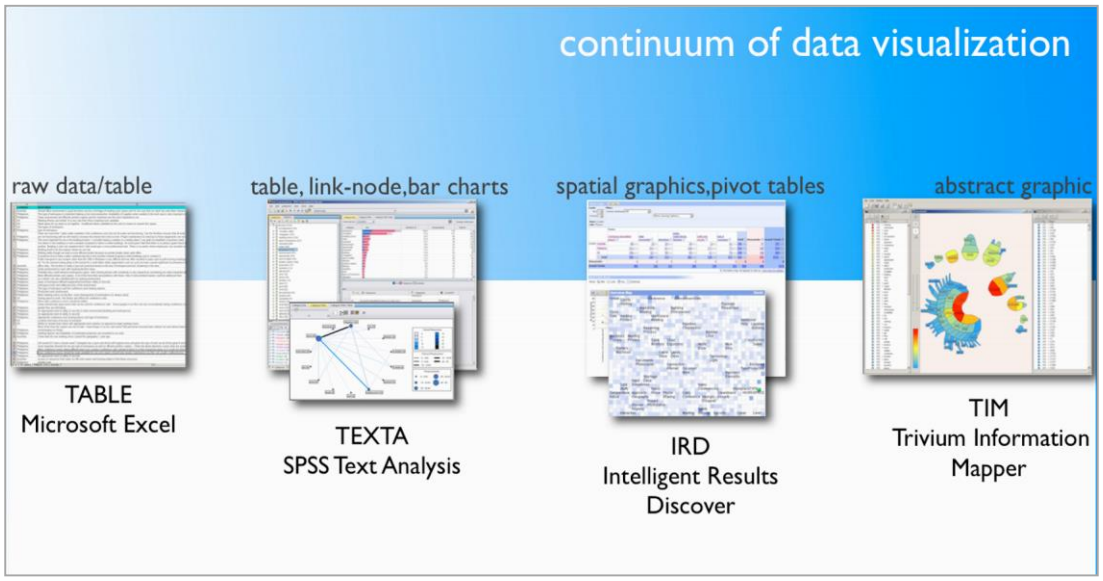


Figure 1.12 Continuum of data visualization

Raw textual data presented in a table falls on one end of the visual abstraction continuum as it uses only written language or arbitrary visualization components of text. This requires prior learning and takes longer to process as the viewer needs to decode the language in order to understand the content (Ware, 2004). Most data mining software give a frequency analysis of the raw text. Although, this still requires a viewer to decode the numerical information, it makes processing less overwhelming as some analysis is already done.

Supplementing tables with conventional graphics, such as bar charts, node-link or network diagrams, represent the next level on the continuum.

On the next level of the continuum, we find representations which make use of increasingly abstract data visualizations. For example, a spatial data abstraction to present an overview along with a representation of frequency and relationships, which uses more sophisticated representations such as pivot tables versus tabular numerical representations.

Finally, the other extreme of the continuum involves displaying highly abstract data visualizations that make use of sensory codes such as shape, color and proximity primarily to represent qualitative data.

In order to understand the decision-making processes which these representations are intended to facilitate, the following sections review decisionmaking theory, information seeking behavior and cognitive incentive systems which suggest which strategy a decision-maker will adopt based on an analysis of the representations available to him or her.

An important characteristic of a majority of research in visualization and decision-making is their focus on the speed and accuracy of the final decision versus the process by which the decision is made. Decision-making research in general has focused on errors and biases and in developing tools to help the "error-prone decisionmaker" (Klienmuntz, 1990). These tools usually involve representations of numerical data. According to Kleinmuntz (1990) although training and decision aids improves decision-making, users are resistant to depending on them. In order to better understand the decision-making process and hypothesize the specific influence of representation on processes, the following section reviews decision-making theory on processes.

1.10 Decision-Making Theory

Decision theory focuses on "goal directed behavior in the presence of options" (Hansson, 1994). The goal in this behavioral process is most often to choose the best alternative from a given set of choices. This decision making process is affected by multiple factors such as the characteristics of the decision maker, the task and the environment in which the decision needs to be made. Decision maker characteristics include task domain knowledge, work experience and decision making authority. Task characteristics include task realism and task structure and finally environmental characteristics include distractions, time constraints and the importance of the task in a larger context (Borgman, 1996).

1.10.1 Sequential Models of Decision-Making

Models that describe this process of decision-making can be categorized as sequential or non-sequential models. Sequential models organize the decision-making

process into phases or stages that always occur in a specific order. Herbert Simon (1960) developed a sequential model of decision-making for organizational contexts that involves three primary phases. The first phase termed "intelligence" relates to identifying opportunities or situations for decision-making. The second phase called "design", involves developing different courses of action that are possible. In the final phase, "choice", the most suitable option from those identified in the design phase, is chosen.

In another important sequential model the decision-making process is presented as a six-step process - "identification of problem, obtaining necessary information, production of possible solutions, evaluation of the solutions, choosing a strategy for performance and finally, implementation of the decision" (Brim et al., 1962 as cited in Hansson, 1994).

According to Kleinmuntz and Schkade (1993) complex decision-making tasks involve set of subtasks that are carried out in a sequence and the overall strategy is a series of sub-strategies designed to complete each of the subtasks. They suggest and that these sub-strategies may involve stages such as information acquisition or evaluation, simple subtasks such as pattern recognition or summarizing information and basic operations such as reading, mathematical calculations and comparing (Kleinmuntz & Schkade, 1993)

Empirical research by Witte (1972 as cited in Hansson, 1994) challenges the sequential nature of decision-making that these models propose. Witte's findings suggest that as an individual collects information, alternatives are developed simultaneously. These alternatives are analyzed as soon as they are generated and in doing so a decision is made. Witte suggests that these actions are a "package of operations and the succession of these operations over time, constitutes the total decision-making process". Non-sequential models take a similar perspective to decision-making.

1.10.2 Non-Sequential Models of Decision-Making

An important non-sequential model of decision-making is that of Mintzberg, Raisinghani and Theoret (1976). This model includes three phases similar to those of Herbert Simon's intelligence, design and choice. In this model these phases are called "Identification, Development and Selection". An important distinction from that of Simon's model is that these three phases do not occur in a specific order or even in a simple sequence. Each phase also contains sub-phases or "routines".

There are two routines within the identification phase. In the first routine is decision recognition. This usually happens by reviewing "streams of ambiguous, largely verbal data that decision makers receive" (Mintzberg et al., 1976). In the next routine called diagnosis, decision-makers review existing information and look for new information in order to understand the issues more clearly.

In the development phase, existing courses of action or options are uncovered through the search routine and new options are developed in the design routine. Existing solutions may also be refined in the design routine. There are three routines leading up a final decision in the selection phase. These are screen, evaluation-choice, and finally authorization. In the screening routine, available options that are clearly below required standards are eliminated. This is only done if it is felt that the number of existing options or courses of action is too large that they cannot all be evaluated in a thorough manner. Once the screen routine leaves a few good solutions to be evaluated, using techniques such as "intuitive judgment, bargaining and analysis" (Hansson, 1994), a final choice is made in the routine called evaluation-choice. The final choice then needs to be approved by a higher authority in the final routine called authorization.

The phases and routines delineated above do not always occur in a specific sequence. According to Johnson et al., decision-making is a cyclical process of "hypothesis formation, active, expectancy guided information acquisition, and information processing" (Johnson et al., 1992 as cited in Borgman, 1996). Often a

decision-maker will cycle through a few routines or phases before moving onto another (Hansson, 1994).

Although managers spend a large proportion of their time in the problem identification phase and expend even more effort in development of possible solutions (Mintzberg et al., 1976) than in the choice phase, much of decision-making research has focused on analysis of the end result of the process or the final choice (Beach, 1993; Hansson, 1994). This study's primary focus is on the influence of visual representation on the information search and screening processes that lead to a final decision.

1.11 Information Search Behavior

The main components of Information acquisition or search behavior are depth, content and sequence. Depth relates to the amount of information that is acquired. Content relates to which information is retrieved and sequence is the order in which the different information is searched for and retrieved (Jacoby et al, 1987).

Borgman (1996) investigated manager's pre-decisional information search behavior using a similar focus on process as this study. Using a process tracing technique Borgman studied the characteristics that distinguish successful information system search behavior from unsuccessful approaches. Participants were asked to play the role of a 'senior business analyst' of a fictional company, 'Multilevel'. As senior business analysts the participants analyzed financial, marketing and sales data about a failed daughter company of Multilevel and put together a brief containing analysis of the situation that this company was in. The data was presented in the form of graphs with a total of 960 graphs that included all combination of variables. He found that successful search behavior was characterized by an initial scanning of information using a general but structured checklist, followed by more focused and detailed analysis. Participants "plunging in" or immediately focusing on a detailed analysis of an issue based on preconceived ideas characterized unsuccessful search behavior (Borgman, 1996).

Interestingly, this study found that participant's perceived performance did not significantly correlate with objective score of performance ($p>0.05$). Retrieval of more important information actually resulted in decreased perceived performance. CatsBaril and Huber (1987) also found that higher degree of information retrieval leads to decreased perceived performance but improves actual performance. They suggest that this may be due to the "lack of closure" that an individual feels when the amount of information retrieved extends beyond a certain level and more nuances are discovered in a poorly defined problem situation (Cats-Baril & Huber, 1987).

1.12 Cognitive Incentive Systems

The strategy that a decision-maker adopts is influenced by the representation forms available. Different strategies are used for numerical, textual and graphic forms (Stone & Schkade. 1991) as representations differ in the amount of effort needed to use them and the resulting level of accuracy. A cognitive incentive system is set in place in any decision-making task and decision-makers will adopt different strategies based on a cost-benefit analysis of the representation forms (Kleinmuntz & Schkade, 1993).

When a decision-maker adopts a specific strategy, feedback about the amount of effort it requires is obtained quickly and easily understood. However, feedback about accuracy can be delayed and vague resulting in decision-makers being better able to estimate effort than accuracy (Einhorn & Hogarth, 1978 as cited in Kleinmuntz & Schkade, 1993). This is further complicated by the fact that it is difficult to learn the level of accuracy a strategy affords as each decision-making scenario may have different criteria for determining accuracy. However, perception of effort expended does not vary as much from one scenario to another (Kleinmuntz & Schkade, 1993).

1.13 Research Questions and Hypotheses

This section relates the literature reviewed to the goals of this study. Research in information visualization has focused primarily on developing algorithms for new representations and less on how these representations influence the decision-making

process. Decision-making theory has been driven by a model of the decision-maker as error prone and biased, and a focus on the final choice stage of decision-making (Beach, 1993; Hansson, 1994). Consequently, a majority of research on information representations designed to aid decision-making do not address earlier stages of the decision-making process. Representation of qualitative data in data-mining and decision-making is a relatively new field of research. The analysis of qualitative information may be less likely to have a definite outcome as it is more open to interpretation than quantitative data. Investigating the process by which users strategize, analyze and identify themes in qualitative data is a required step towards designing interfaces that support the phases where decision-makers expend the most time and effort (Mintzberg et al., 1976). The following section describes the objective of this study and hypotheses from the research reviewed.

1.13.1 Objective

To investigate how four different forms of information representation (tabular; text and conventional graphics; pivot table and spatial visualization; complete data visualization) affect how users strategize, analyze and identify themes in a decisionmaking scenario for a qualitative dataset.

1.13.2 Hypotheses

1.13.2.1 Themes/patterns and levels of detail

Visualizations are useful for identifying patterns in a dataset. However, it does not afford detailed analysis (Speier & Morris, 2003). This suggests that as the level of data visualization increases, the ease with which decision-makers identify themes or patterns will also increase. It also suggests that the difficulty of getting at the details in a highly abstract graphical representation may lead to increased effort on the part of the decision-maker. Will the number of operations a decision-maker performs on a dataset increase as the level of data visualization increases? Consequently, will

satisfaction with the level of detail reached in a given time frame decline as the level of data visualization increases? These questions lead to the following hypotheses:

- H1 - Participants will find it easier to identify themes as the level of data visualization increases
- H2 - The number of operations on the data required to arrive at a decision will increase as a function of increasing data visualization
- H3 - Participants will be less satisfied with the level of detail reached with increasing levels of data visualization.

1.13.2.2 Confidence levels

Studies have also shown that when the decision-maker is exposed to a high level of detail or sees nuance in a dataset, the less confident they are likely to be about their decisions (Cats-Baril & Huber, 1987). How will confidence levels change with increasing levels of abstract visualization?

- H4 – Participants will rate their confidence in decisions lower as the level of data visualization increases

1.13.2.3 Understanding representations

The more elements there are in a representation space, the longer it takes to process (Hornbæk et al., 2002). Also, representations which use a greater number of arbitrary visual codes versus sensory codes may require more time and learning on the part of the viewer as arbitrary codes need to be learnt and decoded in order to be processed (Ware, 2004). Will decision-makers ask fewer questions in order to understand a representation as the level of data visualization increases to extreme of the continuum, which mainly uses sensory codes that are processed immediately?

- H5 – Participants will ask fewer questions related to understanding representations as the level of data visualization increases

1.14 Conclusion

Since a majority of research on decision-making and visualization has focused on the final choice phase, the efficacy of representations have been judged by the accuracy and speed of the decisions made. Most of these studies focused on problems of risk, financial information and other quantitative tasks, which are more amenable to judgments of accuracy. Qualitative information is more open to interpretation and analyzing the accuracy of final decisions against pre-determined criteria may not be appropriate or possible in this context. Instead, comparisons of decisions across representations may help clarify whether one representation form is likely to sway a decision maker in one direction versus another.

In order to investigate the questions and test the hypotheses described above, this study uses the combination of quantitative and qualitative methods required for easy comparison of decisions made and to gain insight into subjective decisionmaking process. The following chapter describes this methodology in detail.

2. METHODOLOGY

2.1 Participants

Thirty-two participants, 16 male and 16 female, participated in this betweenparticipants experiment with four conditions of 8 participants each, but gender was not balanced because no effect of gender was anticipated. All participants were either graduate students or staff members at Cornell University (28 graduate students, 4 staff members). The average number of years of past work experience was 5.63 years (SD = 7.73). Average age was 30.16 years (SD =8.98). None of the participants had prior experience with large qualitative data sets or data mining software. The participants were recruited using employee and graduate student list serves, flyers, and course email lists. Participants were asked if they were color blind at the time of scheduling to screen for color blindness. None of the participants were color blind. They were compensated $20 for their participation in the 1.5 hours experiment session in which they were trained in the software and given a decision-making task to perform.

2.2 Apparatus

The experiment used a managerial decision making scenario and four software packages to represent a multidimensional qualitative dataset of 1574 responses to an

open-ended survey. The data was presented to each participant using one of the four software packages on a 17-inch Dell AS501, 1280 x 1024 pixels 48-cm TFT LCD screen. The participants communicated requests to the researcher to manipulate the data using an Instant Messaging client on a Dell Latitude D620 Laptop computer. The Instant Messaging client was used to document all interactions between the researcher and participant. The participants filled in a response sheet during the exploration process to document their findings and to finally present their decisions.

2.2.1 Decision Making Scenario

The participants were presented with a managerial decision making scenario based upon the real world uses of the workplace survey responses by managers. In this scenario, the participant was instructed to play the role of a manager responsible for workplace design policies for a fictional company with operations in three countries. This company had initiated a company wide survey with open-ended questions to get feedback from its employees on their physical work environment. As the manager, the participants were asked to examine the data collected to identify what important areas employees voice concerns about, what the specific problems may be and what they suggest as improvements. They then needed to prioritize and rank areas where they thought action to be taken by the company, globally as well as for each country. This managerial decision making task was designed based on consultation with managers about the kind of analysis and decisions that are made with similar data in their companies. (See Appendix A)

2.2.2 Information Representations

The 1,574 responses mentioned above were obtained from a real internal workplace survey conducted by a major furniture company. The anonymous responses then were inputted into each of four software packages to form the information representations that were tested in the four conditions of this study. Two conditions represented the dataset using one dominant representation form: textual representation

or abstract visualization. The other two conditions different combinations of graphical, text and quantitative representations. Summaries of the four information representations based on the information seeking tasks delineated by Schneiderman (1996) and described in detail in Chapter 1 are illustrated in the table 2.1.

None of the software conditions allowed users to view a history of their interactions and extracting to an external software or tool for further analysis was not in the scope of the experiment task.

Table 2.1 Features of each condition based on Schneiderman's information seeking tasks

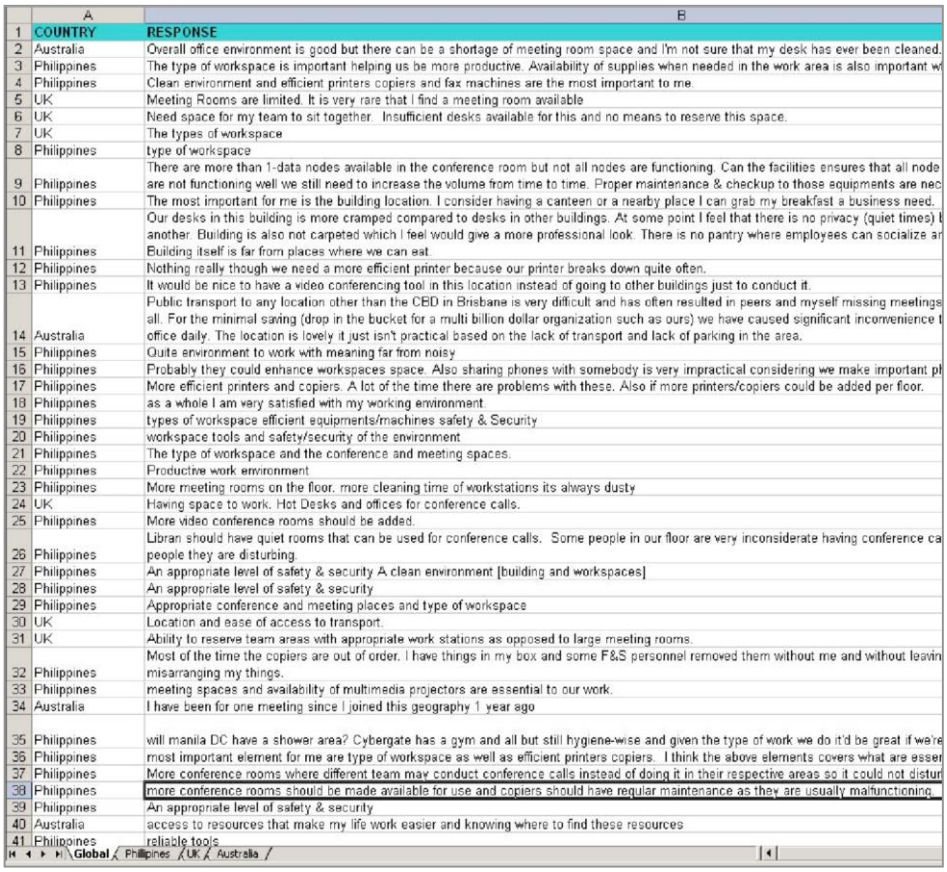
Condition	Representation	Overview	Zoom	Detailsondemand	Relate	Filter
TABLE (Excel)	Text or Table visualization	No	No	Yes	No	No
TEXTA (SPSS Text Analysis)	Word/concept frequency and line/network diagrams	Yes	No	Yes	Yes	Yes
IRD (Intelligent Results Discover)	Spatial graphics and pivot tables	Yes	No	Yes	Yes	Yes

TIM (Eaagle Information Mapper)	Abstract visualization	Yes	No	Yes	Yes	Yes

2.2.2.1. Table Visualization

Table visualization or a data table is a form of representation where text data is listed in table form. This representation comprised a list of all responses copied verbatim in a spreadsheet (Microsoft Excel 2003). The spreadsheet contained 4 worksheets, one for each country and one worksheet containing responses from all three countries. Each worksheet contained two columns, one for the country variable and one for the response.

The participants were allowed to organize the data in any way they needed arrive at a decision as long as it was possible to execute in this software. Typical manipulation included alphabetizing responses, searching for keywords, color coding, deleting responses, and identifying frequency of occurrence for specific keywords.



	A	B
1	COUNTRY	RESPONSE
2	Australia	Overall office environment is good but there can be a shortage of meeting room space and I'm not sure that my desk has ever been cleaned.
3	Philippines	The type of workspace is important helping us be more productive. Availability of supplies when needed in the work area is also important w
4	Philippines	Clean environment and efficient printers copiers and fax machines are the most important to me.
5	UK	Meeting Rooms are limited. It is very rare that I find a meeting room available
6	UK	Need space for my team to sit together. Insufficient desks available for this and no means to reserve this space.
7	UK	The types of workspace
8	Philippines	type of workspace
9	Philippines	There are more than 1-data nodes available in the conference room but not all nodes are functioning. Can the facilities ensures that all node are not functioning well we still need to increase the volume from time to time. Proper maintenance & checkup to those equipments are nec
10	Philippines	The most important for me is the building location. I consider having a canteen or a nearby place I can grab my breakfast a business need.
11	Philippines	Our desks in this building is more cramped compared to desks in other buildings. At some point I feel that there is no privacy (quiet times) b another. Building is also not carpeted which I feel would give a more professional look. There is no pantry where employees can socialize ar Building itself is far from places where we can eat.
12	Philippines	Nothing really though we need a more efficient printer because our printer breaks down quite often.
13	Philippines	It would be nice to have a video conferencing tool in this location instead of going to other buildings just to conduct it.
14	Australia	Public transport to any location other than the CBD in Brisbane is very difficult and has often resulted in peers and myself missing meetings all. For the minimal saving (drop in the bucket for a multi billion dollar organization such as ours) we have caused significant inconvenience t office daily. The location is lovely it just isn't practical based on the lack of transport and lack of parking in the area.
15	Philippines	Quite environment to work with meaning far from noisy
16	Philippines	Probably they could enhance workspaces space. Also sharing phones with somebody is very impractical considering we make important ph
17	Philippines	More efficient printers and copiers. A lot of the time there are problems with these. Also if more printers/copiers could be added per floor.
18	Philippines	as a whole I am very satisfied with my working environment
19	Philippines	types of workspace efficient equipments/machines safety & Security
20	Philippines	workspace tools and safety/security of the environment
21	Philippines	The type of workspace and the conference and meeting spaces.
22	Philippines	Productive work environment
23	Philippines	More meeting rooms on the floor. more cleaning time of workstations its always dusty
24	UK	Having space to work. Hot Desks and offices for conference calls.
25	Philippines	More video conference rooms should be added.
26	Philippines	Libran should have quiet rooms that can be used for conference calls. Some people in our floor are very inconsiderate having conference ca people they are disturbing.
27	Philippines	An appropriate level of safety & security A clean environment [building and workspaces]
28	Philippines	An appropriate level of safety & security
29	Philippines	Appropriate conference and meeting places and type of workspace
30	UK	Location and ease of access to transport.
31	UK	Ability to reserve team areas with appropriate work stations as opposed to large meeting rooms.
32	Philippines	Most of the time the copiers are out of order. I have things in my box and some F&S personnel removed them without me and without leavin misarranging my things.
33	Philippines	meeting spaces and availability of multimedia projectors are essential to our work.
34	Australia	I have been for one meeting since I joined this geography 1 year ago
35	Philippines	will manila DC have a shower area? Cybergate has a gym and all but still hygiene-wise and given the type of work we do it'd be great if we're
36	Philippines	most important element for me are type of workspace as well as efficient printers copiers. I think the above elements covers what are esser
37	Philippines	More conference rooms where different team may conduct conference calls instead of doing it in their respective areas so it could not distur
38	Philippines	more conference rooms should be made available for use and copiers should have regular maintenance as they are usually malfunctioning.
39	Philippines	An appropriate level of safety & security
40	Australia	access to resources that make my life work easier and knowing where to find these resources
41	Philippines	reliable tools

Global / Philippines / UK / Australia /

Figure. 2.1 Table Visualization

2.2.2.2 Table Visualization, Link-node and Bar charts

This second condition used the SPSS Text Analysis (TEXTA) software to present a combination of text and graphical representations of the dataset. The interface for this software consisted of four panes.

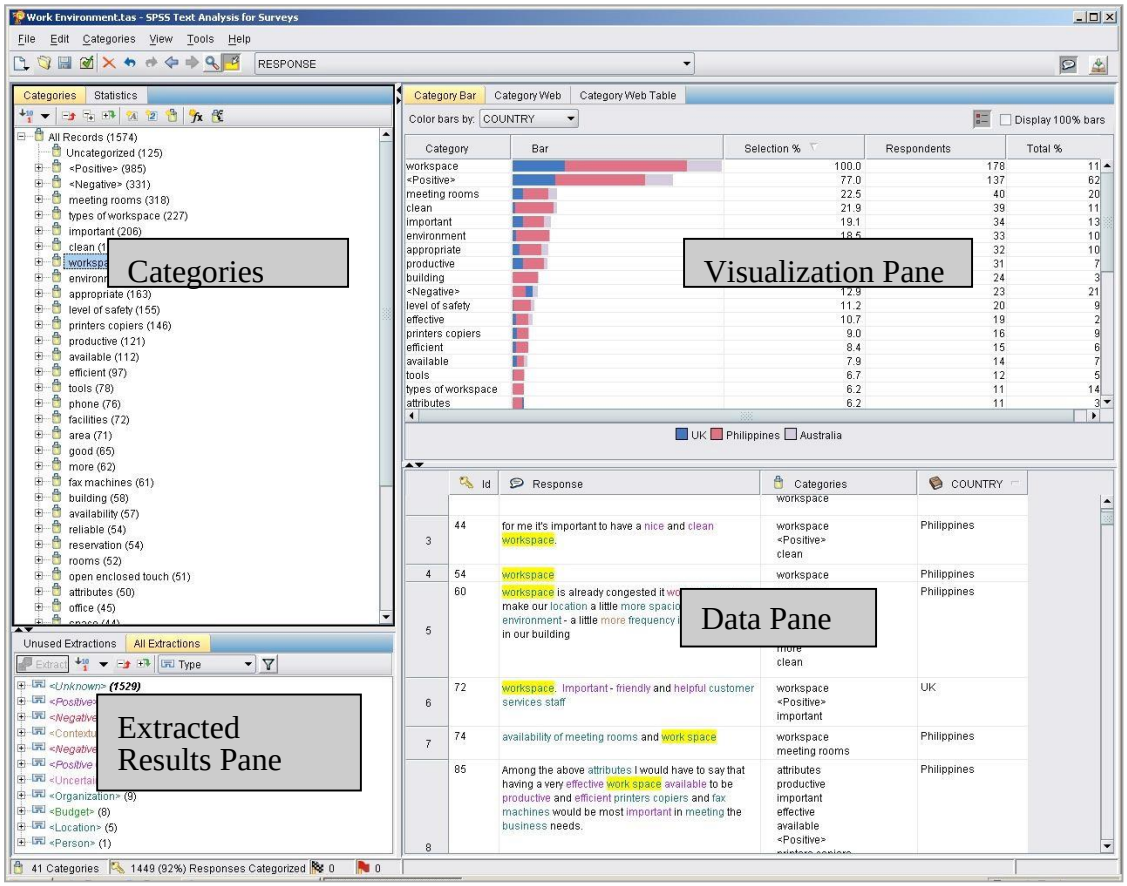
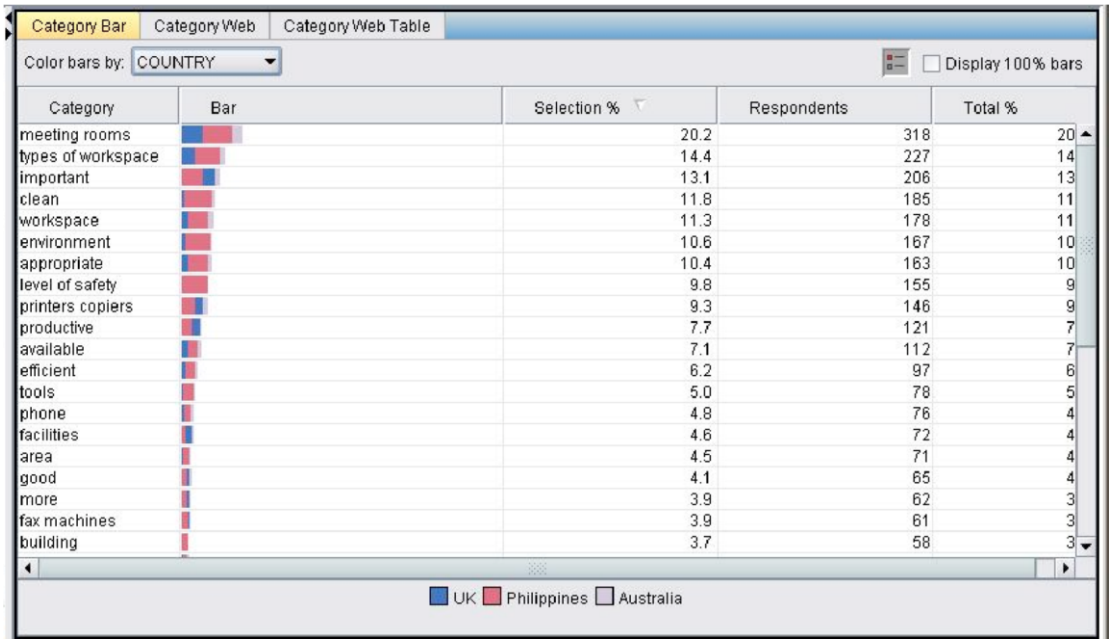


Figure. 2.2 TEXTA Interface

The Data pane displays in table form the response ID, the actual text response, categories in which it falls and any other structured variable such as ‘country’. Displayed responses are filtered according to what is selected in the Extracted Results or Categories pane. The Extracted Results pane lists key concepts or terms extracted from the responses. These are used to create the categories that are listed in the Categories pane. The fourth pane contains the visualization of the data in three forms – category bar, category table or a link-node visualization called the category web.

Based on selections in the category pane these visualizations show how responses fit into categories and how different categories share responses.



Category	Bar	Selection %	Respondents	Total %
meeting rooms		20.2	318	20
types of workspace		14.4	227	14
important		13.1	206	13
clean		11.8	185	11
workspace		11.3	178	11
environment		10.6	167	10
appropriate		10.4	163	10
level of safety		9.8	155	9
printers copiers		9.3	146	9
productive		7.7	121	7
available		7.1	112	7
efficient		6.2	97	6
tools		5.0	78	5
phone		4.8	76	4
facilities		4.6	72	4
area		4.5	71	4
good		4.1	65	4
more		3.9	62	3
fax machines		3.9	61	3
building		3.7	58	3

Figure 2.3 TEXTA Category Bar

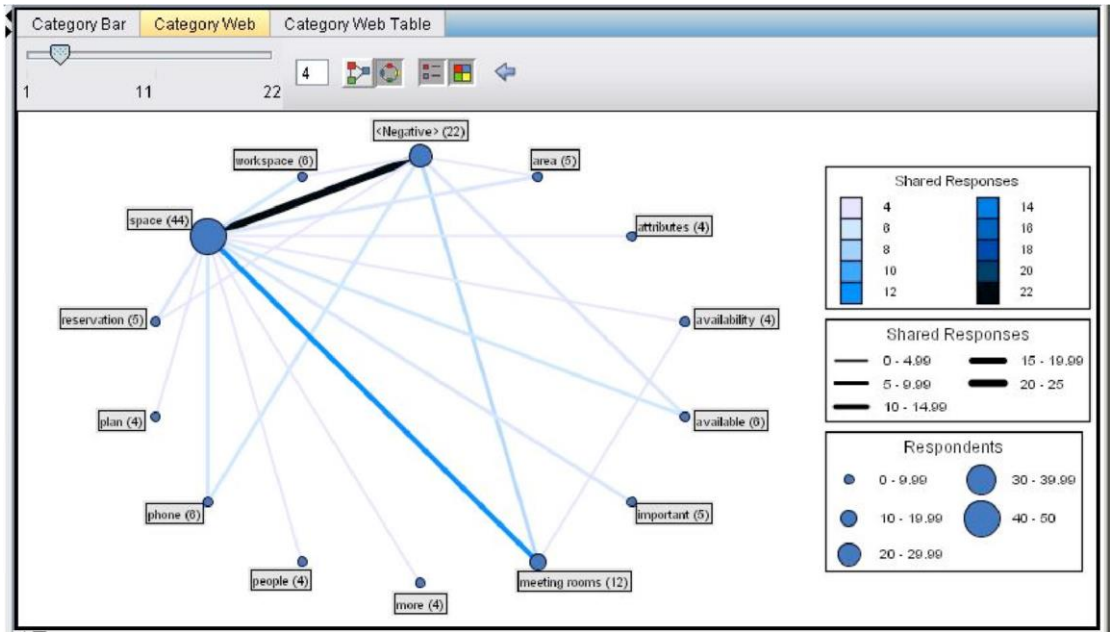


Figure 2.4 TEXTA Category Web

2.2.2.3 Spatial Graphics and Pivot Tables

A data mining software called Intelligent Results – Discover (IR) formed the third condition where the dataset was represented using an overview map, a list of dimensions and a set of pivot tables.

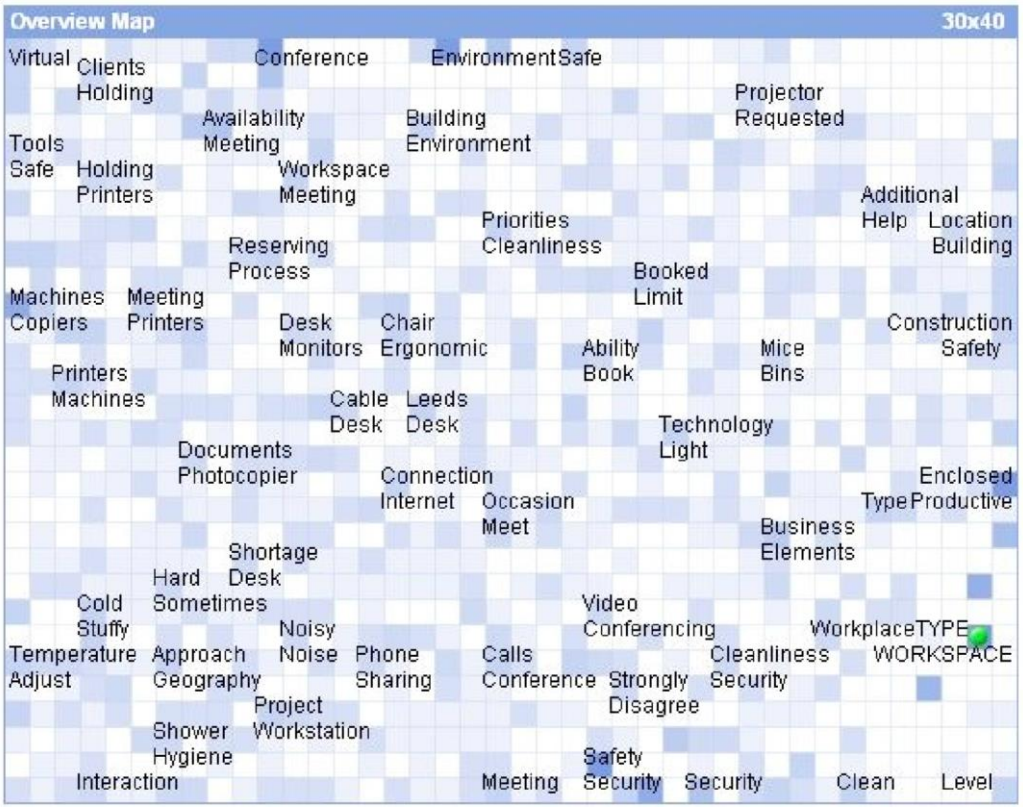


Figure 2.5 IRD Overview Map

Categories that have been identified in the data by the researcher are listed as dimensions on the left hand side of the interface. Selecting a dimension results in a listing of responses associated with it in the main pane and also an illustration of how these responses are spread in the entire data set through the overview map. The software also presents an analysis of the data in the form of a pivot table. This representation allows for a comparison of responses by country, sentiment and dimensions using quantitative information such as percentages of responses and number of responses in one category. Under the pivot tables are also visualization options such as bar chart, line chart, pie chart or a grid like visualization as in the overview map.

Figure 2.6 IRD Pivot Table

2.2.2.4 Abstract Visualization

A content analysis software called Eaagle Information Mapper (TIM) formed the fourth condition. This condition used abstract visualization of the dataset as the primary representation through which the data is explored. This condition was most representative of the visual data mining method. The software interface contained three columns – the thesaurus, graphic and corpus.

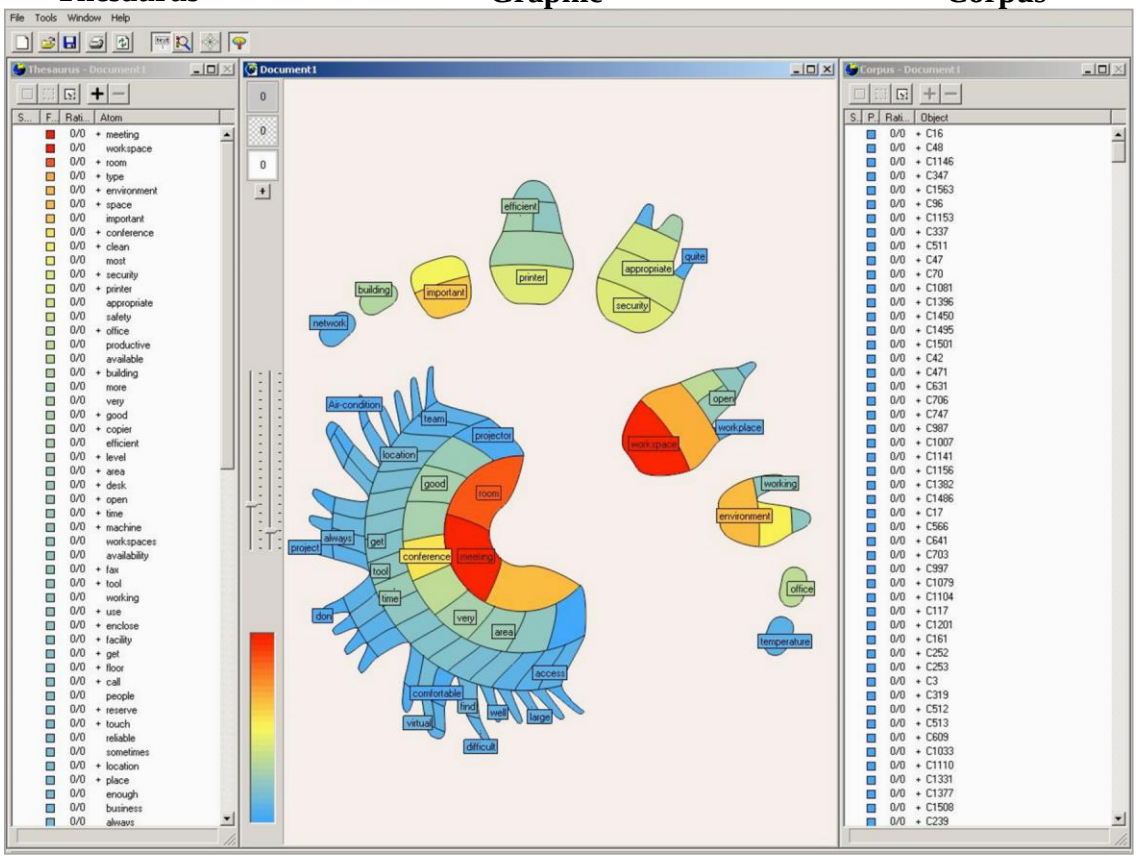


Figure 2.7 TIM Interface

The thesaurus lists the hundred most frequent words found in all the responses. In default mode, this list is arranged by frequency. It can also be arranged alphabetically. Colors denote frequency, with red being the most frequent and blue, the least frequent.

The graphic panel contains the graphic representation of the words listed in the thesaurus panel. This graphic is formed based on the following criteria - the frequency of a word across all responses in the data, the number of times it occurs in an individual response and the proximity of different words in the responses.

Color, size and proximity are most important in reading the graphic. Shapes in red most significantly meet the criteria listed above and shapes in blue meet the criteria to a lesser degree. The larger the shape, the more frequently the word occurs. Proximity of shapes denotes the proximity of the words associated with the shapes in the data set. Apart from size, the form of the shapes has no meaning.

This graphic has a zoom function, which allows one to view the data in different levels of detail. At zero percent zoom, one sees an overview of the data. As the zoom level increases, one starts to see more details in how the words cluster together. Separate clusters of words form islands that represent different themes in the data.

By selecting individual shapes or clusters in the graphic, one can read the specific responses associated with the selection from a third panel called the corpus. Once a selection is made in the graphic panel, the corpus panel arranges the list of all responses in the data set according to their relevance to the selection. Responses which contain all the words in the selection are arranged first, coded in red and as one goes down the list, the responses contain a fewer number of words from the selection. The least relevant responses to the selection are coded in blue. Selection of a response in the corpus column opens a new window with the full response.

2.2.3 Interaction with the data

The software used in the four conditions differed in the amount of training required to manipulate and explore the data. To minimize the effect of varying levels of learning abilities on the results and to more closely simulate how a manager might interact with the data in an organization, direct participant interaction with the software was curtailed. Instead, the representation of the dataset was presented on a screen and the participants communicated requests to manipulate this representation through the Trillium Instant Messaging client using a Dell Latitude D620 Laptop computer to the researcher. The researcher, trained in all software packages then controlled and navigated through the software based on the directions of the participant, who was able to observe the interactions. All communication between participant and researcher was recorded and time stamped by the Instant Messaging client.

2.2.4 Response Sheet

As they explored the dataset, the participants filled in an electronic response sheet consisted of five columns: Theme/ issue, Description, Working rating of importance, Comments and Rank Order in a spreadsheet (Microsoft Excel). Of the five fields, identification of theme/issue and the rank order of importance of the themes identified were required fields. The others were meant to help the participants describe and categorize the data during the exploration process, for their own reference. The participants filled in this information on a global level and for each individual country.

2.2.5 Data set used in decision making task

The set of comments that were explored as part of the decision making task was extracted from a larger data set obtained from an international furniture company. The original data set comprised employee responses to a set of eleven open-ended questions on various aspects of the work environment. All responses that were not in English were eliminated from the test set, resulting in a remainder of approximately 14,000 responses. Responses ranged from one to five lines in length. These responses were contained in an Excel spreadsheet and this was spell checked using the spelling and grammar tool in Excel 2003. The company name and all employee identifiers were stripped from the responses. The data set was then further narrowed down to responses to two open ended questions (See Appendix A).

Responses of employees from eleven countries were then fed into two of the software programs to examine the visualization and representation formed in order to further narrow down the number of countries. Countries with less than 200 responses were removed. Since many of the responses were one to two lines in length, a data set of less than 200 responses did not contain enough information for themes to emerge in the visualization and the resulting graphic showed scattered responses. Countries with more than 200 responses but with data that showed no clear clustering into themes were also removed. Four countries were then left - India (approx 1,100 responses),

Philippines (913 responses), UK (474 responses) and Australia (237 responses). Comparison of the responses from India and the Philippines revealed that they were very similar in the themes that emerged in the dataset. India also had more responses than the Philippines. Inclusion of the large dataset from India would have made the table condition of the list of comments in the Excel spreadsheet unwieldy, therefore, this data set was removed. The final three countries shared some common themes, but differed in the presence of certain others and in the level of significance of each theme. This selection of 1,574 responses formed the final data set that the participants explored during the decision making task. It should be noted that this data was used only to create the different information representations for the four conditions. The focus of this study was not on the analysis of this data per se but on the effect that each representation had on the exploration and decision making process for the same data set.

2.3 Procedure

The experiment was conducted in the Human Factors and Ergonomics Laboratory in the Department of Design and Environmental Analysis at Cornell University. The participants were first briefed about the study and how long it would take. They were then given an informed consent form, which explained the purpose of the study, risks and benefits and compensation for participation in the study. This was followed by a training session of approximately twenty minutes on the software condition they would use. The training consisted of what was possible to do in the software and how the researcher and participant would communicate different requests to manipulate data. The Excel condition had a shorter training period of 5 – 10 minutes as all participants were familiar with the software. There was also a 2 -3 minute practice session with the chat client. The participants were then presented with the managerial decision making scenario. After any doubts were cleared about the task, one hour was allotted to explore the data and fill in the response sheet. An optional five minutes was given after the process to complete the response sheet.

Participants then filled in a questionnaire to elicit feedback on the process and were asked some open-ended questions of how they perceived the process in an unstructured interview.

2.4 Experiment Setting

Participants were tested individually. Each participant was seated at a computer workstation in front of a computer screen that presented the representation of the data that they needed to explore. The participant did not directly interact with this data. Instead they used the Instant Messaging client on a Dell Latitude D620 Laptop computer to interact with the researcher. The final response sheet was also completed using the laptop. The lack of interactivity and the presence of the researcher more closely simulated corporate conditions where the results of the survey are presented to managers rather than managers undertaking their own data analysis.

2.5 Measures

In addition to the chat log and the response sheet described in the apparatus, the other measures in this study included a survey and an interview.

2.5.1 Post Experiment Survey

This survey consisted of two parts. The first section included basic demographic questions. The second section was a five-point Likert scale that was developed to obtain feedback on the decision making process and confidence levels. The questionnaire included eight common questions for the four conditions. These questions related to four main aspects – the ease with which themes could be identified, the level of analysis that the representation allowed, confidence in decisions made, and the perceived effectiveness of the representations. The survey also had two to three condition specific questions that related to unique aspects of the representations in that condition. (see Appendix D).

2.5.2 Post Experiment Interview

Participants were asked open-ended questions in a short post-experiment interview about whether they had any comments on their experience and what they found interesting or difficult during the process.

2.6 Independent and Dependent Variables

This study tests the effect of four levels of data visualization (tabular; text and conventional graphics; pivot table and spatial visualization; complete abstract data visualization), on the following dependent variables: percieved ease with which themes are identified, operational requests, satisfaction with level of detail obtained, level of confidence, and the number of doubts in understanding representations. These main dependent variables are described in detail below, followed by other variables measured to support the analysis.

2.6.1 Percieved Ease with which Themes were Identified

Participant rating of agreement or disagreement to the statement: "It was easy to identify themes in the data", was used to measure percieved ease with which themes were identified. This related to hypothesis 1: Participants will find it easier to identify themes as the level of data visualization increases.

2.6.2 Operational Requests

The extent to which the data needed to be manipulated to arrive at a decision was analyzed for each participant. Requests made to the researcher to alter the data in some way was coded as "operational requests" and counted for each participant. This was measured to test hypothesis 2: The number of operations on the data required to arrive at a decision will increase as a function of increasing data visualization.

2.6.3 Satisfaction with Level of Detail Obtained

Participant rating of agreement or disagreement to the survey statement: "I am satisfied with the level of detail I could reach in the analysis of the data.", was used to measure the satisfaction with the level of detail reached in the analysis. This related to

hypothesis 3: Participants will be less satisfied with the level of detail reached with increasing levels of data visualization.

2.6.4 Level of Confidence

Participants' rating of agreement or disagreement to the survey statements: "I am confident with the decisions I made on global priorities" and "I am unsure of my analysis of differences in priorities for each country" were used to measure confidence in decisions made. This related to hypothesis 4: Participants will rate their confidence in decisions lower as the level of data visualization increases.

2.6.5 Number of Doubts in Understanding Representations

Doubts about how to read the information representation were counted for each participant from the chat transcript. This variable was measured to test hypothesis 5: Participants will ask fewer questions related to understanding representations as the level of data visualization increases.

2.7 Other Variables Measured

Apart from the variables relating to the hypotheses of this study, the following were also measured in order to provide greater insight and support the analysis of the primary dependent variables desribed above.

2.7.1 Percieved Effectiveness of Representations

Participants' rating of agreement or disagreement to the statements: " I would not use this form of data representation to analyze this type of data and make decisions in the future" and " I think this representation helped me arrive at good decisions quickly and easily", were used to measure the percieved effectiveness of the representations in facilitating decision-making.

2.7.2 Operational Doubts

In addition to the operational requests, the number of doubt clarifications that were needed between the researcher and the participant in order to manipulate the data were counted. This was measured to be able to distinguish between the possible

ambiguity of the information representation and any issues that arise because of miscommunication between the participant and researcher in terms of navigation.

2.7.3 Data Doubts

Any doubts about the data such as what kind of company it was from or which industry it related to were counted for each participant.

2.7.4 Final Decisions

The number of themes that were identified in each condition on a global and individual country level were analyzed using descriptive statistics. Major themes were also identified by the researcher and compared with the themes identified by the participants.

2.7.5 Comments on Decision-Making Process

Participant feedback on the decision-making process was also obtained through the post experiment interview. Responses to open ended questions on which aspects of the process they found difficult and how they felt about the decision-making process were analyzed.

2.8 Data Analysis

The following section describes the data analysis for chat transcripts, survey responses, interviews and response sheets.

2.8.1 Chat Transcripts

The interactions between the researcher and participant, documented in the chat transcripts, were analyzed for the following interactions and dependent variables: operational requests, operational doubts, representation doubts and data doubts. Requests from the participants that fell in one of the four interaction categories mentioned above were coded as such and analyzed using analysis of variance statistics.

2.8.2 Survey Responses

The questionnaire was scored to analyze ease with which themes were identified, satisfaction with level of detail reached, confidence levels with decisions made, percieved effectiveness of representations and other representation specific questions. Participant ratings on a five point likert scale, with 1 being strongly disagree to 5 being strongly agree, were analyzed using descriptive statistics.

2.8.3 Interviews

The participants were asked for any comments they had on their experience of the task and the information representations they used. These responses were transcribed and analyzed in order to identify recurrent themes within a condition.

2.8.4 Response Sheet

Using descriptive statistics, the number of themes identified globally and locally were analyzed. The percentage of respondents who ranked a specific theme as top priority was calculated and compared across conditions.

3. RESULTS

3.1 Process Analysis

This section describes the results derived from analysis of the chat transcripts.

3.1.1 Operations on the Data

The extent to which the data needed to be manipulated to arrive at a decision was analyzed. This was based on the operational requests made to the researcher to alter the information representation in some way or view individual responses. The number of clarifications that were needed between the researcher and the participant in order to manipulate the representation and data was also analyzed.

3.1.1.1 Operational Requests

The analysis of variance ($F_{(3,31)} = 9.576$) indicated a significant difference between conditions in the number of operational requests ($p=0.0002$).

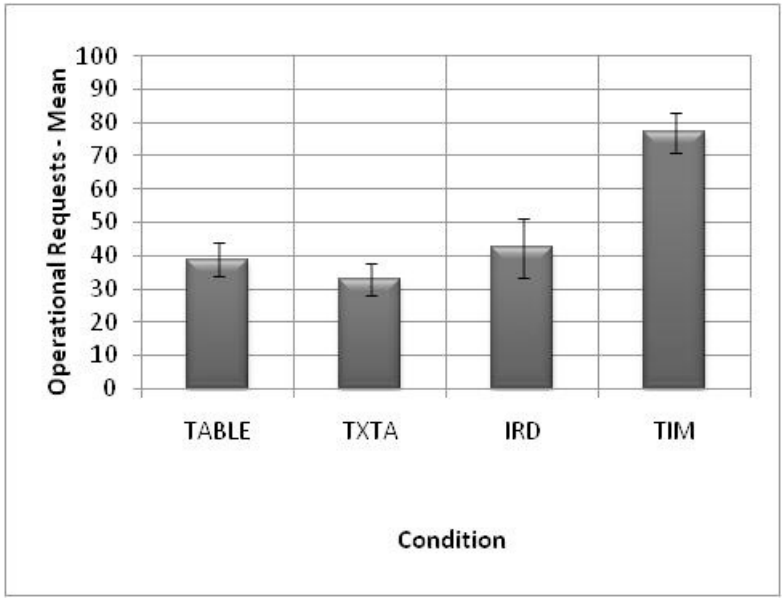


Figure 3.1 Mean Operational Requests

TIM had the highest number of operational requests with a mean of 77.125 (SD =14.827). The other three conditions had relatively lower means. IRD followed with a mean of 42.5 (SD= 25.196).

The high standard deviation suggests wide variation in operational requests for this condition. The TABLE condition had the next highest mean of 38.875 (SD = 16.89) and TEXTA had the lowest average number of operational requests with a mean of 32.75 (SD = 13.792). TIM was significantly different from each of the other three conditions ($p = 0.0002$). Although TIM had the highest mean, followed by IR, the results also indicate that TABLE scored higher than TEXTA, which suggests that hypothesis 2, "the number of operations on the data required to arrive at a decision will increase as a function of increasing data visualization" is not supported.

3.1.1.2 Operational Doubts

The analysis of variance ($F_{(3,31)} = 16.867$) shows a significant difference between conditions in the number of operational doubts ($p<0.0001$).

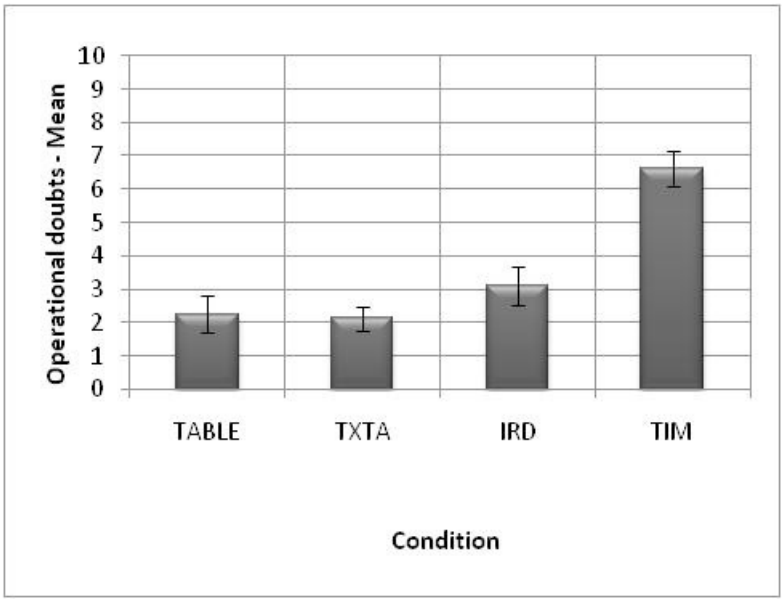


Figure 3.2 Mean Operational Doubts

TIM had the highest average number of operational doubts with a mean of 6.625 (SD= 1.506). IRD followed with a mean of 3.125 (SD=1.642). The TABLE condition followed with a mean of 2.25 (SD= 1.581) and TEXTA had the least number of operational doubts with a mean value of 2.125 (SD= 1.584). TIM was significantly different from each of the other three conditions ($p<0.0001$).

3.1.2 Understanding Representations

The number of doubts about the way the information was represented in the interface and how to read the data were analyzed for each subject and condition. The difference in interface/representation doubts between conditions ($F_{(3,31)} = 4.371$) was statistically significant ($p=0.0121$).

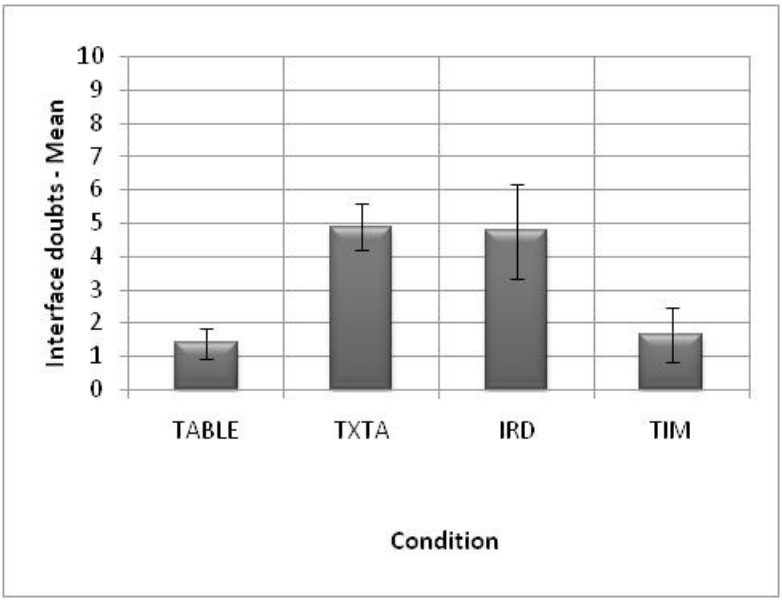


Figure 3.3 Mean Interface Doubts

TEXTA had the highest average number of doubts regarding the interface with a mean value of 4.875 (SD=1.959). IRD followed with a mean of 4.75 (SD=3.991). TIM and the TABLE condition had fewer number of interface doubts with a mean of 1.625 (SD=2.326) and 1.375 (SD=1.303) respectively. Relatively high standard deviations for IR, TIM and the TABLE condition suggest a wide variation in the number of interface doubts for these conditions. The hypothesis that participants will ask fewer questions related to understanding representations as the level of data visualization increases is rejected. Interestingly, the two lowest scores for this variable fell on the conditions that lie on the two extremes of the continuum.

3.1.3 Questions about the Data

Additional questions about the data such as where the data was from, what kind of employees responded and clarifications of any terms in the data that the subject was not familiar with were analyzed. The difference in means between conditions ($F_{(3,31)}$ = 1.305) was not statistically significant (p =0.2923).

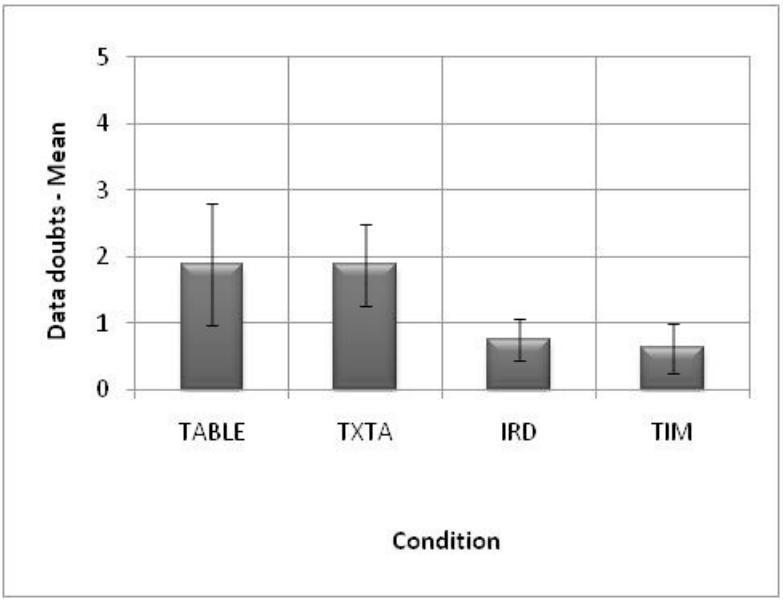


Figure 3.4 Mean Data Doubts

TEXTA and the TABLE condition had equal means at 1.875 with high standard deviations of 1.727 and 2.588 respectively. IRD followed with a mean of 0.75 and a standard deviation of 0.886. TIM had the lowest number of data doubts with a mean of 0.625 and a high standard deviation of 1.061. The results indicate a high level of variation in doubts about the data within each condition and no significant difference between conditions.

3.2 Perception of Decision-Making Process and Information Representations

A survey and short interview was used to evaluate subjects' perception of the decision-making process and information representation. Subjects indicated the extent to which they agreed or disagreed with eight statements common to all conditions and a few condition specific statements in the survey. The results of the survey follow.

3.2.1 Level of Analysis

Three questions in the survey related to the perception of the levels of analysis that each condition facilitated. The results are described below.

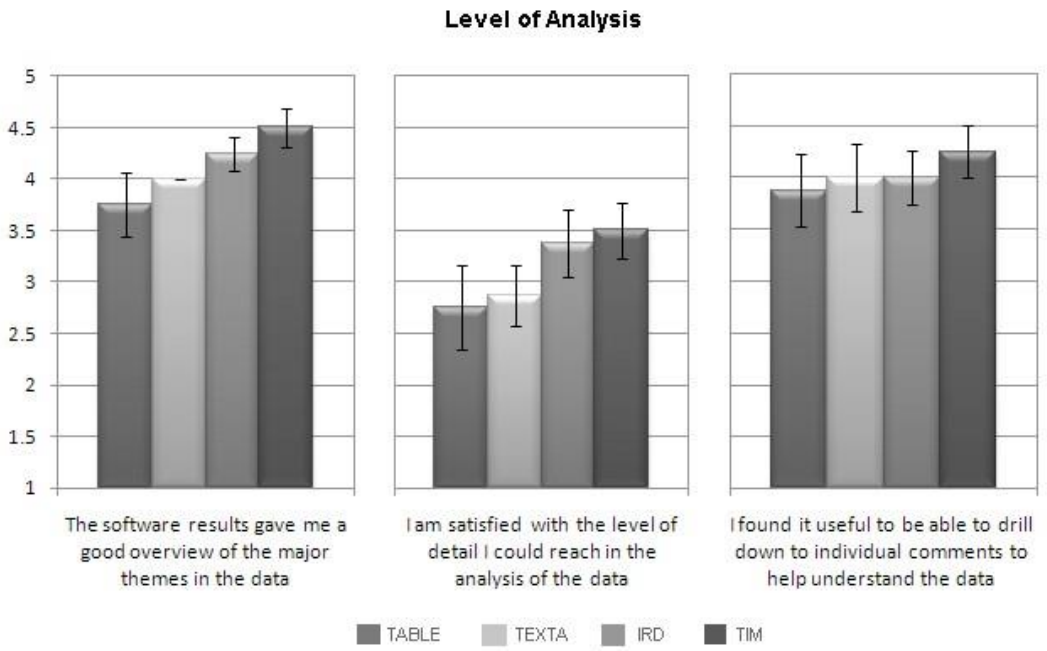


Figure 3.5 Mean scores on survey for level of analysis

3.2.1.1 The software results gave me a good overview of the major themes in the data. Subjects in the TIM condition agreed with this statement to a higher degree than the other conditions. TIM scored a mean of 4.5 (SD= 0.535 and SE=0.189). IRD followed TIM with a mean of 4.25 (SD= 0.463 and SE=0.163). TEXTA and the TABLE condition scored lower with mean scores of 4 (SD = 0, SE=0) and 3.75 (SD=0.886 and SE=0.31) respectively. The difference in mean scores was marginally statistically significant (p = 0.0725).

3.2.1.2 I am satisfied with the level of detail I could reach in the analysis of the data. The average score for this statement did not reach "agree" for any of the conditions. TIM scored slightly higher than IRD with a mean of 3.5 (SD = 0.756, SE = 0.267). IR had a mean of 3.375 (SD = 0.916, SE = 0.324). TEXTA scored 2.875 (SD=0.835, SE=0.295) and the TABLE condition scored 2.75 (SD = 1.165, SE =0.412). Although, the differences in means were not statistically significant (p = 0.3102). The direction of the result tends towards rejection of the hypothesis that participants will be less satisfied with the level of detail reached with increasing levels of data visualization.

3.2.1.3 I found it useful to be able to drill down to individual comments to help understand the data.

Subjects in the TIM condition agreed with this statement to a slightly higher degree than other conditions with a mean score of 4.25 (SD=0.707 and SE =0.25). Subjects in both the IRD and TEXTA condition agreed with this statement. Mean score was 4 for both (SD = 0.756, SE = 0.267 and SD= 0.925 and SE = 0.327 respectively). The TABLE condition scored the lowest with a mean of 3.875 (SD= 0.991 and SE= 0.350). The differences in mean scores were not statistically significant (p=0.845).

3.2.2. Level of Confidence in Decisions

Two statements related to the extent to which subjects were confident of the decisions they arrived at on a global and local level.

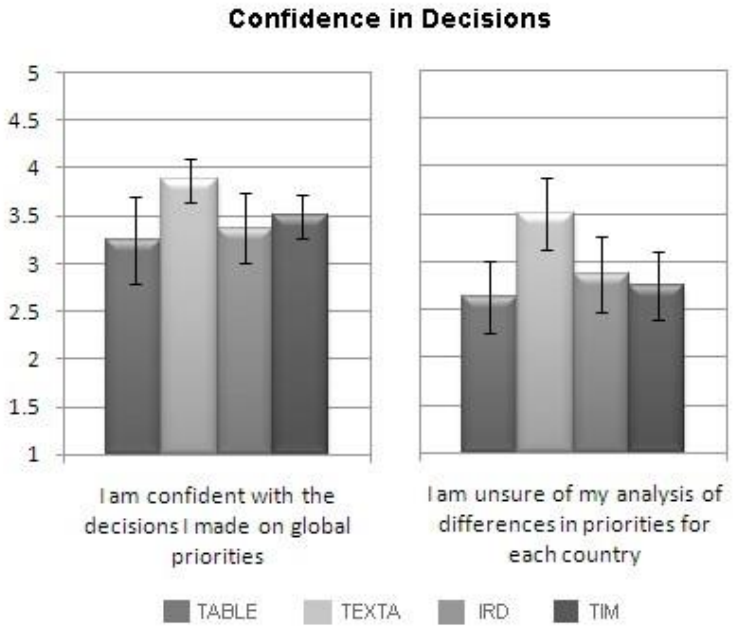


Figure 3.6 Mean scores on survey for confidence in decisions

3.2.2.1 I am confident with the decisions I made on global priorities.

TEXTA scored the highest with a mean value of 3.875 (SD =0.641, SE=0.227). TIM and IRD scored around the same point with means of 3.5 and 3.375 (SD =0.926, SE=0.227 and SD =1.061 and SE =0.375). TABLE condition scored the lowest with mean of 3.25 (SD = 1.282 and SE =0.45316). Differences were not statistically significant (p=0.634)

3.2.2.2 I am unsure of my analysis of differences in priorities for each country.

Subjects in the TEXTA condition tended to agree with this statement to a higher degree than the other conditions; however the mean score was not at the 'Agree' point. The mean score for this condition was 3.5 (SD= 1.07, SE=0.378). IRD scored lower with a mean of 2.875 (SD= 1.126, SE 0.398) and TIM scored 2.75 (SD=1.035 SE =0.366). The TABLE condition scored 2.625 (SD = 1.061, SE = 0.375). Differences were not statistically significant (p=0.3863).

The difference in means were not significant for either of the two statements, which evaluated confidence, and the results tends towards the rejection of the hypothesis that participants will rate their confidence in decisions lower as the level of data visualization increases.

3.2.3 Perceived effectiveness of representations for decision-making

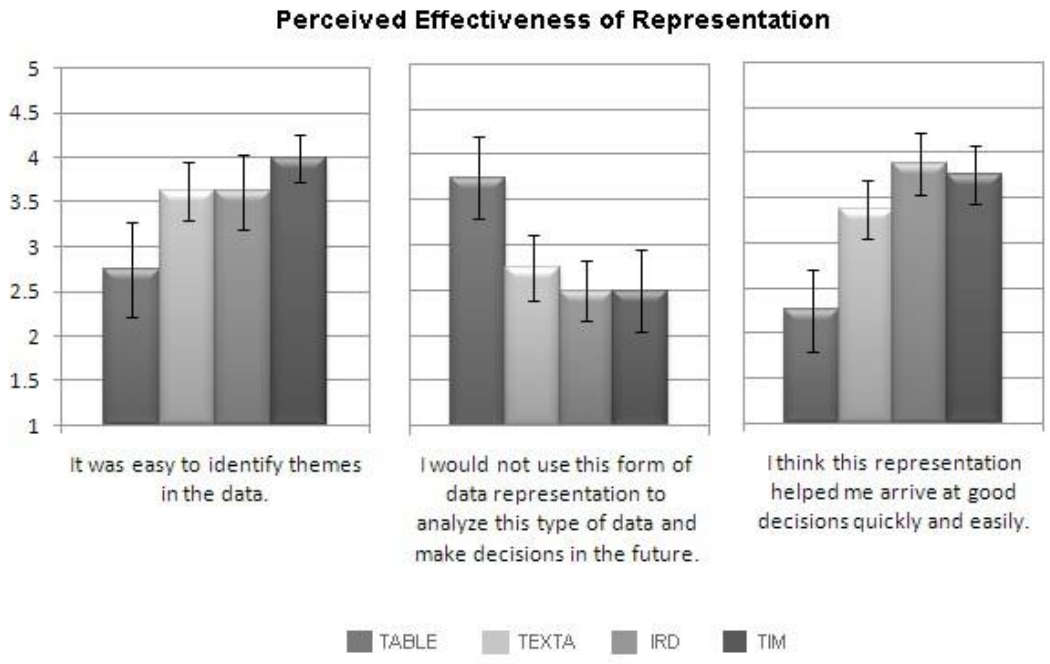


Figure 3.7 Mean scores on survey for effectiveness of representations

3.2.3.1 It was easy to identify themes in the data

Subjects in the TIM condition tended to agree with this statement. TIM scored the highest with a mean of 4 (SD= 0.756, SE= 0.267). IRD and TEXTA fell on the same point at 3.625 (SD = 1.187 and SE = 0.42, SD = 0.916, SE =0.324). The TABLE condition scored the lowest with 2.75 (SD=1.49 and SE=0.526). Although the results are in the direction of the hypotheses that participants will find it easier to identify themes as the level of data visualization increases, the difference between means are not statistically significant (p=0.173) to accept the hypothesis.

3.2.3.2 I would not use this form of data representation to analyze this type of data and make decisions in the future.

Subjects in the TABLE condition tended to agree most to this statement with a mean score of 3.75 (SD=1.282, SE =0.453). TEXTA followed with a mean of 2.75 and standard deviation of 1.035. IRD and TIM both scored 2.5, falling between disagree and neutral with standard deviations of 0.9258 and 1.309 (SE=0.327 and 0.463 respectively). The high SD for TIM indicates greater variability. The differences in mean scores were not statistically significant (p=0.1171)

3.2.3.3.I think this representation helped me arrive at good decisions quickly and easily

IRD scored the highest with a mean value of 3.875 (SD = 0.991, SE =0.350). TIM scored a mean of 3.75 (SD=0.916 SE=0.324). TEXTA scored 3.375 (SD= 0.916, SE=0.3234) and subjects in the TABLE condition tended to disagree with this statement with the lowest mean score at 2.25 (SD=1.282 SE=0.453). The difference between conditions was statistically significant (p=0.0199).

3.2.4. Condition Specific Statements

In addition to the statements mentioned in the previous section of this chapter, the survey also included a few condition specific statements. These are described below.

3.2.4.1 TABLE

3.2.4.1.1 I found the number of responses overwhelming

The mean score for this statement was 3.375 with a high standard deviation at 1.188. This suggests high variability in responses to this statement. The mean suggests that scores fell closer to the neutral point.

3.2.4.2.TEXTA

3.2.4.2.1 I found the bar chart feature helpful for making country specific decisions.

A mean of 4 and standard deviation of 0.534 suggests that subjects tended towards agreeing or strongly agreeing to this statement.

3.2.4.2.2 The colors used to represent the "types" were easy to remember while exploring the data.

A mean of 2.75 suggests that subjects fell in the disagreement side of the scale. However a standard deviation of 1.165 suggests that there is high variability in scores.

3.2.4.2.3 The category web/bar chart feature helped me understand easily how each category was related to other categories.

Subjects tended towards a neutral position with a mean score for this statement at 3.375. The standard deviation was 0.916.

3.2.4.3 IRD

3.2.4.3.1 I found the visualization or overview map confusing

The mean score of 3.75 for this statement suggests that subjects in this condition tended towards agreeing with this statement. However there are indications of high variability with a standard deviation of 1.165

3.2.4.3.2 I found it useful to be able to see quantitative information such as frequencies and percentages

The mean score of 4.25 and a low standard deviation (SD = 0.463) suggests that subjects in this condition tended to agree that quantitative information was useful.

3.2.4.4 TIM

3.2.4.4.1 I found the graphic images confusing

The mean score of 1.75 (SD = 1.035) suggests that subjects tended to fall between strongly disagree and disagree for this statement, suggesting that they did not find the graphic confusing

3.2.4.4.2 .I liked the use of size, shape, and color in the graphic images The mean score of 4.25 suggests that subjects agreed with this statement. However, the standard deviation was high at 1.389.

3.3 Interview

The interview consisted of open-ended questions on which aspects of the process they found difficult and how they felt about the decision-making experience.

3.3.1 TABLE

Three subjects (n=8) in this condition said they would have liked to have more time. Four subjects expressed how difficult and overwhelming the process was.

- "It was tiring and monotonous. After a while they seemed like they were saying the same thing. I was not getting much after 15minutes. After reading 300 responses in the Philippines, everything looked the same."
- "It was overwhelming, especially with such a large population."
- "Really confusing, there was so much info, I didn't know how to organize it"
- "Terrible. Hard to identify anything."

Three subjects mentioned that it is valuable to read all the comments

- "Reading the comments was good as I would put my own interpretation on a graphical representation."

- "I did not think outside of the given attributes and If I had more time I would scan down line by line to get an idea of what might not be included in the attributes but may be important to look for."

Two subjects mentioned using preconceived notions about what might be important and two subjects mentioned assuming there was a pattern and trying to generate hypothesis based on a few responses in order to guide their decision-making process.

- "I started by thinking about what was important for my work and in relation to that I went from there. I had a personal idea that safety might be a concern in the Philippines and checked it out."

- "I had my own estimation about which issues might exist in different countries...almost as if it is given in the UK and Australia that it would be clean and safe".

- One subject read a few comments to generate hypothesis to test by searching the data with keywords. However the subject didn't see a common theme to test hypothesis. "Everyone wants a productive environment"

- "Went through each response for around ten responses. If they were similar, I assumed there was a pattern."

Subjects said that the lack of organization of the data was difficult in the process and mentioned how they tried to organize the data.

- "Tried to use color highlighting to see what might be redundant. Used numbers especially for the Philippines."

- "I am used to quantifying and it might not be the best method but it usually helps me refine my index. I tried to figure out common keywords, give them a number and then sort according to keywords and then take out commonality to look at specific comments left"

- "The responses that were copy pasted from the question were junk. There was no organization to the comments. I tried to sort alphabetically"

One subject brought up the issue of not being able to see sentiment associated with responses in this condition also.

- "I am not sure of the results." " As a manager I would not trust this. Randomly searching for words -you can make out a little from the numbers but you do not know what their comments were...whether it was good or bad" Suggestions for better representations included "visual representation of where the important stuff is", "graphical picture would have been good" and "maybe a grading system." One subject also mentioned that the weaknesses of the questionnaire should also be addressed and another mentioned that she is used to more consistency in terms that found in the data.

3.3.2 TEXTA

Three subjects (n=8) mentioned that they felt they needed more time to explore the data. One subject also mentioned that they felt like they were missing out on information because they could not go through everything. Another subject said, " I wanted to read everything" and "felt bad not reading individual comments". Two subjects mentioned the visualizations. One subject mentioned that the ability to select a particular issue and see how other themes overlap or relate to the selection was a good feature. Another subject commented on the use of frequency and bar chart together.

- "The frequency of comments can tell you which issue is of more concern". You can look at the bar chart for a single category and then see for each country. The combination of frequency information and the bar chart lets you see how much of a concern it is."

One person mentioned that "The type of survey in itself is difficult to analyze" and another mentioned that there was a lot of overlap and not a great level of specificity.

- "Clean and environment always went together and so should not have been made two categories"

One person suggested having percentage information for countries and another suggested randomizing the order in which comments are listed for any selection in order to reduce the risk of relying on first impressions.

- " When you look at comments for a certain issue, there are a lot and it may be natural for people to read from the beginning and form ideas based on what you see first. I think random ordering of comments every time you choose the category may be better."

One subject also mentioned that she did not take categories with less than 50 comments seriously and felt this representation works for identifying larger themes. One subject also said that the interface was user friendly

- "I am not good with this kind of stuff but this was really user friendly and I was kind of impressed by how easy it was to use and to be able to sort things."

Regarding positive or negative sentiment in responses, one subject mentioned that the positive category was not helpful.

- "I used the negative filter in the beginning and found that I had to keep going back to it. The positive was not helpful."

3.3.3 IRD

Four subjects (n=8) mentioned that they would have liked more direct interactivity with the software package. Three subjects mentioned that the overlap in themes represented and the fact that multiple themes were mentioned in individual responses resulted in a redundancy that made the process difficult. Three subjects said that it was easier once they got the hang of it and developed a strategy. However, trying to figure out a strategy was the difficult part.

- "I got more efficient half way through when I actually got down to using it and realized how it works."

Two subjects said they needed more time to complete the task. Two subjects mentioned that they were very confused during the process with how to go about the task, with the nature of the comments and inconsistencies between representations of the data.

- "At first it is amazing the amount of information you can get about the data but then you don't know where to start with the limited amount of time." "Comments had so much more than just the category I was exploring it in so there was a depth to the comments that was difficult to see from percentages."

- " I did not know where I was or how it would come together." "All representations did not match up. The visualization below the pivot table did not totally correlate with the table. Sometimes it was difficult to tell from the overview map where people said more. For example, a cell with only 19 responses was dark blue. Why is that important? The nature of the comments with a lot of cut and pasting from the original question made it difficult also."

One subject mentioned however that it seemed like a good way to mine data and seemed useful overall. It was pretty easy to pick up on basic functions and know how to go about looking for something.

Regarding the negative or positive connotations associated with categories, one subject said that the categories by themselves seemed neutral and so had to read the responses to confirm. Another subject said that they trusted the sentiment analysis and dimensions of the software.

- "Made the assumption that the software used some reasonable logic in the way it classified negative and positive and assumed no overlap."

  One person mentioned that they felt the line chart was the more useful than the table and would have liked to combine search queries and mix and match results.

- " Line graphs were much more useful than percentages. It didn't strike me that clean was way higher in data table but it was obvious from chart. Would choose chart

if I had to"

3.3.4 TIM

Seven subjects (n=8) mentioned that the graphic was easy to understand and they found the representation engaging to work with. Three people mentioned that one difficult aspect of this representation was that given the time constraints, it was difficult to tell just by looking at the graphic, the sentiment associated with a cluster or theme.

- "Major themes were easy to identify but it was hard to be certain whether it was positive or negative".

- "I got paranoid about what they actually might be saying. Just because something occurs frequently does not necessarily mean the employees were saying it was good or bad"

It was also mentioned that since the representations are based on frequency, it would be easy to generalize that a theme has a negative or positive connotation based on what was seen in the country viewed initially.

- " Once you figure that cleanliness is an issue for one country and you see in the map that it appears for another, you could easily assume that it is also an issue but it may have been employees saying it was great."

Five people said that in the beginning, the hardest part was trying to find a strategy to explore the data and make decisions.

- "It was hard to figure out how to attack the data. There was a lot going on"

Subjects also mentioned the strategies they used to arrive at a decision

- "First I picked up common themes at 25% zoom and then compared it with the same zoom level for individual countries. I prioritized by color and the number of segments. The more segments there were, the more complex the theme seemed as it had higher number of aspects associated it with it and hence more important."

- "I started with the frequency of words in the beginning and used the graphic to finally make eliminations."

- "I had ideas of words that were important and looked at graphic to see if the words we selected in the list were connected in the data."

- One subject mentioned that the zoom function was useful to test the strength of association between words. "If words remained connected as the zoom increases, I took that as a sign"

One person also mentioned that it felt strange to analyze qualitative data in such a cursory manner since qualitative analysis is traditionally associated with in depth analysis of a response.

3.4 Analysis of Decisions

3.4.1 Themes Identified

The average number of themes identified by participants on a global level ranged from 5.75 to 6 for the four conditions. Six major themes were identified based

on the researcher's analysis of the data using the four software packages. The following figure illustrates the percentage of subjects in each condition who identified these themes.

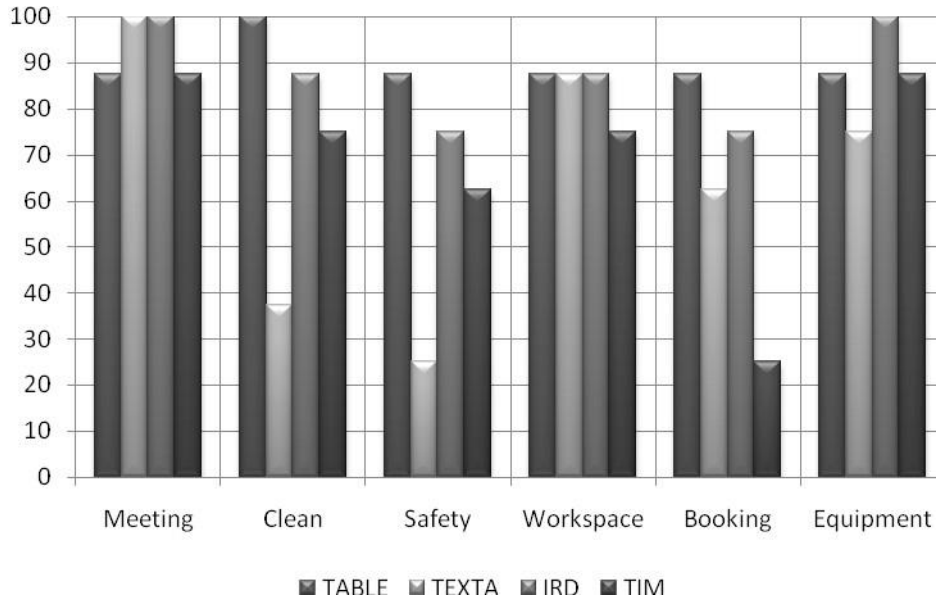


Figure 3.8 Percentage of subjects identifying themes

The themes 'meeting', 'workspace' and 'equipment' were most consistently identified by all four conditions. Meeting was identified as the highest priority most often with 20 of the 32 subjects ranking it as number one on a global level.

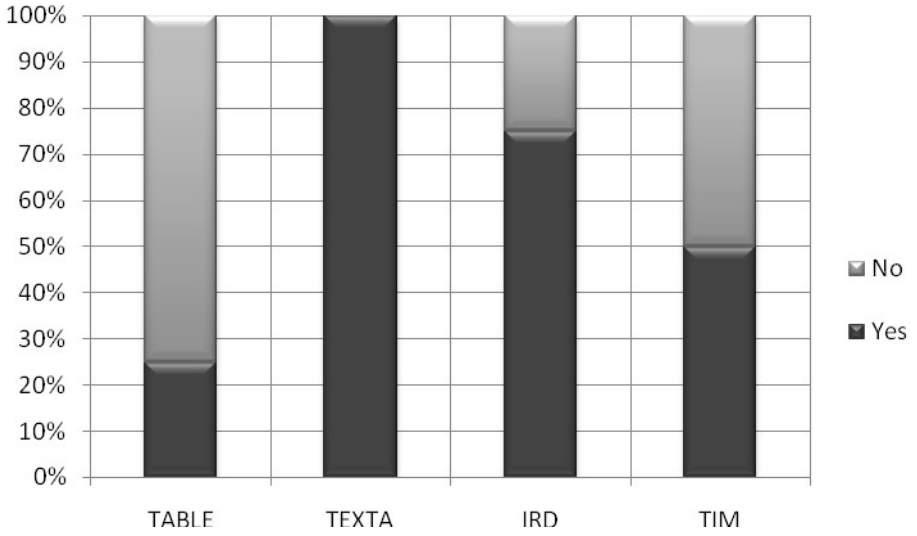


Figure 3.9 Proportion of subjects in each condition ranking meeting as top priority

All subjects rated 'meeting' as number one in TEXTA. In the IRD condition, 75% of the subjects ranked this theme as the highest priority. TIM had 50% of its subjects do so and Excel had only 20% of the subjects rank this theme as number one.

3.4.2 Number of Themes Identified

There was no significant difference between groups on the number of themes identified on a global or individual country level. There was no significant difference in the number of themes identified on a global level (p=0.996).

IRD had the highest mean of 6 (SD= 1.512). TIM and TEXTA both had a mean number of global themes of 5.875 with standard deviations of 1.356 for TIM and a high standard deviation of 3.182 for TEXTA. The TABLE condition had the lowest mean of 5.75 (SD=1.389).

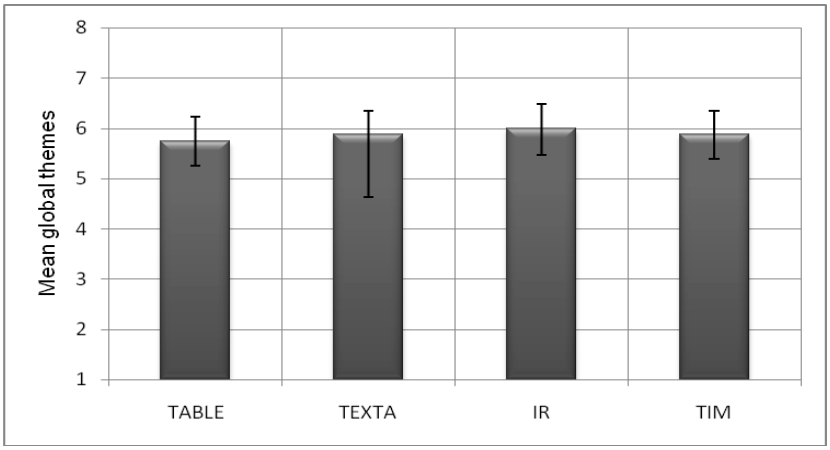


Figure 3.10 Mean number of global themes identified

There was no significant difference in the number of themes identified for the UK (p=0.199), Philippines (p=0.459) or Australia (p=0.338)

For the UK, IRD had the highest mean of 5.38. Subjects in IR were also the most consistent with a relatively lower standard deviation than other conditions at 1.061. TIM followed IRD with a mean of 5.25 and a high standard deviation of 2.712.

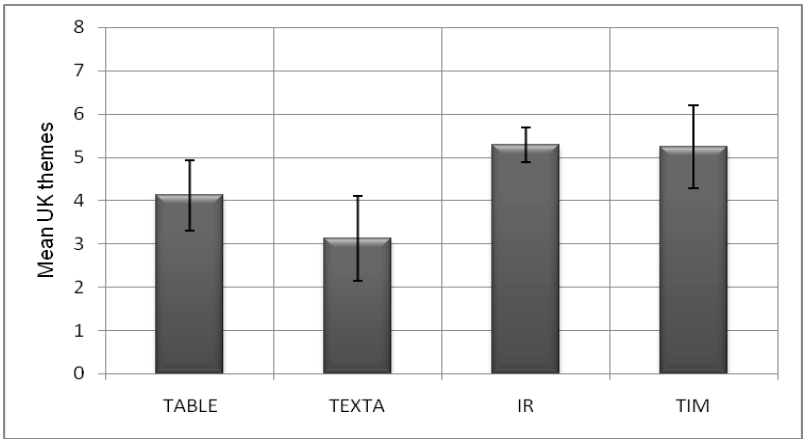


Figure 3.11 Mean number of UK themes identified

The TABLE condition's mean number of themes identified was 4.13 (SD= 2.295). TEXTA had the lowest mean of 3.13 and a high standard deviation of 2.80.

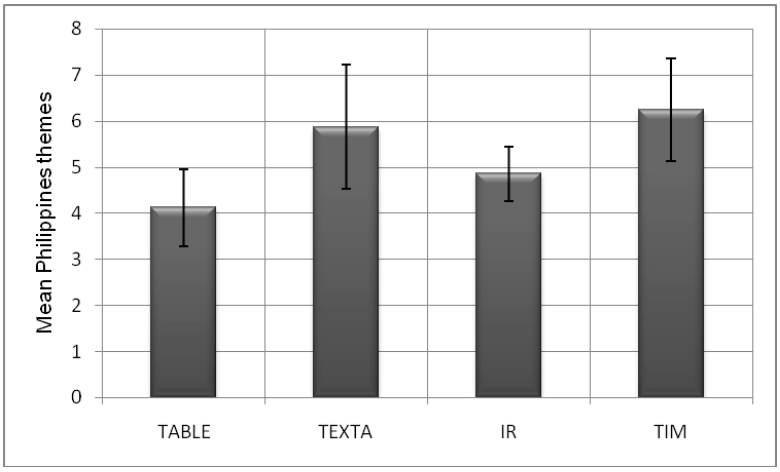


Figure 3.12 Mean number of Philippines themes identified

The average number of themes identified for the Philippines was the highest for TIM with a mean of 6.25. Standard deviation was high at 3.151. TEXTA had the next highest mean of 5.88 with standard deviation of 3.796. IR had a mean of 5 with a standard deviation of 1.604. Finally, the average number of themes identified in the TABLE condition was 4.13 with a standard deviation of 2.357. Again, subjects in the IRD condition were the most consistent with a relatively low standard deviation.

Subjects in the TIM condition identified the highest number of the themes for Australia with a mean of 5.5. TIM also had a high standard deviation of 3.117 indicating wide variation in the number of themes identified. IRD had a mean of 5.13 and the lowest standard deviation of 1.356 amongst the four conditions.

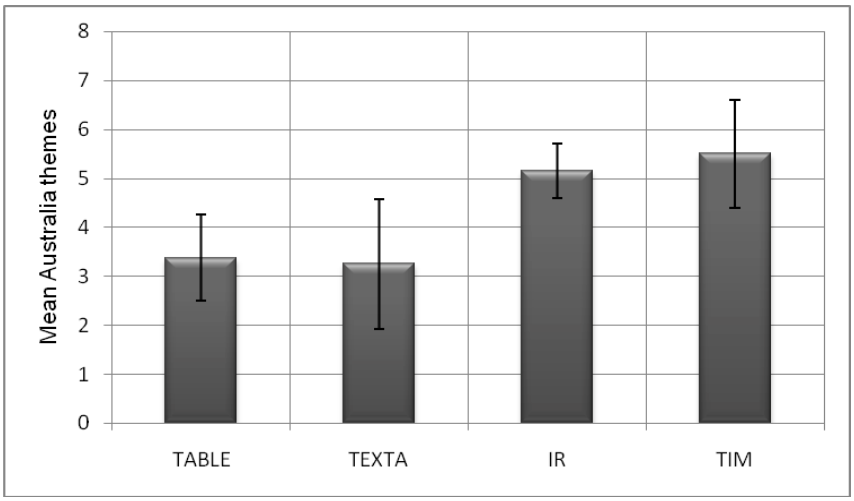


Figure 3.13 Mean number of Australia themes identified

TEXTA followed with a mean of 3.75 and the highest standard deviation of 3.732. The average number of themes identified for Australia in the TABLE condition was 3.25 with a standard deviation of 2.493.

4. DISCUSSION

This study investigated the influence of visual representation of qualitative data at progressively greater levels of abstraction on decision-making processes in order to address a gap in research that currently focuses predominantly on the final choice phase of decision-making and representation of quantitative information. Specifically, this research investigated how four forms of data representation, varying progressively in their use of visualization and data abstraction, compare in the effort required to arrive at a decision, the ease with which themes are identified, the level of detail obtained, the confidence in decisions made, and the intuitiveness of representations. Qualitative data can be open to interpretation and final decisions or conclusions can be difficult to evaluate for accuracy. Therefore, decisions were compared across the four conditions to see how varying the level of visual abstraction of data representation encouraged participants to focus on certain themes versus others and to arrive at their decisions.

The results obtained showed that the degree of visualization and abstraction did not seem to significantly influence the themes extracted from the qualitative data. The average number of themes identified by participants was not significantly different across conditions on a global or local level. The dataset was the result of a survey of workplace attitudes and issues. The themes ‘meeting’, ‘workspace’ and ‘equipment’ were consistently identified in all four conditions. ‘Meeting’ was identified as the highest priority by a majority of participants on a global level. Only the TABLE condition had less than a quarter of its users rate this as the number one priority, though this was not a statistically significant difference compared to the other conditions. Over half of all participants in all conditions identified the six themes that

had also been identified by the researcher (meeting, cleanliness, safety, workspace, booking and equipment), except for 'safety' and 'cleanliness' by participants in TEXTA, and 'booking' by TIM. Overall, in a period of an hour, a majority of participants in all conditions identified and shared at least three issues or themes in the data. Although the results suggests that the degree of visualization and abstraction did not have a significant influence on the final decisions made, the process by which these decisions were arrived at differed considerably between conditions in the following ways.

4.1 Operational effort

The operational effort required was analyzed for each representational condition. The results indicate that the highly abstract graphical representation required the most number of operations in order to arrive at a decision. This condition was significantly different from all the others. Interestingly, the TABLE condition had higher operational requests than either TEXTA or IRD. However, TABLE condition was not significantly different from TEXTA or IRD, The results suggest that the hypotheses, "The number of operations on the data required to arrive at a decision will increase as a function of increasing data visualization" is only partially supported. When users were allowed to view information in multiple forms on one screen (TEXTA) fewer operations were required in order to arrive at a decision. Although IR gives multiple representations of the data, they are not accessible on the screen at the same point in time. As Tufte suggests when users need to compare and make a choice, it is important to present the data in a macro and micro level on the same representational space for easy analysis (Tufte, 1990). However, only the number of operations on the representation was evaluated and not the mental effort required to comprehend representations. The results suggests that the number of operations on the data required to arrive at a decision will be higher for representations that make use of a high level of abstract graphics (TIM) versus representations that provide multiple representations of text and graphics (TEXTA & IRD) or present qualitative data in raw

form. However, it is important to note that this relates only to operations on the data and is not indicative of overall effort required to analyze the data across conditions.

4.2 Ease in identifying themes

In general the first phase of the decision-making process is problem identification (Mintzberg et al., 1976) or situational analysis and problem recognition (Simon, 1960). In a data mining context, this would involve identifying the themes or issues in the dataset. Research suggests that for a task with high complexity such as this one, visual representations of the data best facilitates the recognition of patterns and results in higher decision quality and not detailed analysis (Speier & Morris, 2003). The reported ease with which participants were able to identify themes was highest for TIM followed by IRD, TEXTA and finally the TABLE condition. This result is congruent with research which found that highly graphical representations facilitate easy identification of patterns and themes (Speier & Morris, 2003). However, the participants were asked to rate two statements relating to identifying themes – 'the software results gave them a good overview of the major themes in the data' and 'It was easy to identify themes in the data'. For both statements, the highly graphical representation scored the highest or participants tended to agree to a higher degree and the TABLE condition scored the lowest. However, the difference between conditions was marginally statistically significant for the first statement and not significant for the second. It is possible that participants understood the first statement as an evaluation of the software's ability to analyze data and the second as relating to the ease with which the analysis could be understood. In both cases, although the differences are not statistically different, the trend is congruent with the research and hypotheses that participants will find it easier to identify themes as the level of data visualization increases.

4.3 Satisfaction with level of detail

The reported satisfaction with the level of detail afforded did not reach the "agree" point for any of the conditions. Interestingly, participants using TIM came

closest to satisfied of the four conditions with the TABLE condition coming last. Given, that the TABLE representation presents the most detail and IR gives detailed analysis of themes in the pivot table and other representations, it is surprising that the highly graphical representation scored the highest. However, the differences were not statistically different and did not reach the agree level. According to Speier and Morris (2003) the effectiveness of visual representations is compromised when details need to be analyzed as visual interfaces may require more operations in order to analyze on a micro level (Speier & Morris, 2003). The fact that TIM had the highest number of operations on the data suggests that participants were attempting to retrieve the details. Of the four conditions, participants in TIM tended to agree most that it was useful to be able to drill down to individual comments although this was only marginally different from other conditions. Interviews with participants revealed that they felt the representation was easy to understand and themes easy to identify. They also reported in the survey that they did not find the images used confusing and liked the use of shape, color and size in the representation. However, they also stated that it was difficult to tell just by looking at the representation whether a theme had negative or positive connotations. This relates to Tufte's observation (1997) that there is an optimal level of detail that is required for data to be meaningful and simplicity in representations is not an effective communication technique as users are often suspicious of information that seems opaque. Also, the themes 'booking' and 'meeting' identified in the data are related but are two different issues as respondents to the survey complain of a lack of meeting rooms and an inefficient process for booking the ones that are available. It is interesting that this was not identified as two separate issues by participants in TIM and suggests that it inhibited the detailed analysis required to distinguish the two themes. However, the fact that TIM scored the highest in satisfaction with detail of the four conditions suggests that the hypotheses that participants will be less satisfied with the level of detail reached with increasing levels of data visualization is not supported in this study. However, the interactivity

present in these representations and the ability of participants to dig deeper if they chose makes it inappropriate to generalize that participants will be satisfied with details in a highly graphical representation that is static.

4.4 Level of Confidence

Ratings of confidence did not reach the agree point for any of the conditions, all falling around the neutral point instead. Ratings of confidence levels indicate that participants in the TEXTA condition tended to be more confident than the other conditions in their global decision. The hypotheses that participants will rate their confidence in decisions lower as the level of data visualization increases, was not supported. TEXTA provides multiple representations of the data on one screen and according to research, users gain better understanding of data when overview and detailed information is presented simultaneously (Hornbæk & Frøkjær, 2001). Participants in the TABLE condition rated their confidence the lowest and tended to agree that they would not use this representation to analyze and make decisions with this kind of data in the future. However, the participants in the TABLE condition did identify three of the top themes and had similar number of overall themes identified as the other conditions in the same period of time. It is possible that their perceived suboptimal performance may be due to the "lack of closure" they felt as nuances in the data were revealed in reading a large amount of individual comments (Cats-Baril & Huber, 1987). Although, TEXTA scored the highest in confidence on a global level, on a local level, participants tended to fall between neutral and agree points for the statement, "I am unsure of my analysis of differences in priorities for each country", whereas the other conditions fell between the disagree and neutral points. This differences were not significant. Overall, participants in the IR condition tended to agree with the statement – "I think this representation helped me arrive at good decisions quickly and easily". IRD also presents a quantitative analysis of the data in a pivot table which all participants in this condition felt was useful for decision making. TIM came second with marginally lower rating and participants in the TABLE

condition tended to fall around the disagree point and significantly lower than the other conditions.

4.5 Understanding representations

The perceived ease with which a representation is understood was measured by the number of questions or doubts regarding the representations and what they meant. Interestingly, TEXTA which had the lowest number of operations also had the highest number of doubts regarding the interface. IRD which represents information in multiple forms also was slightly lower than doubts about the interface than TEXTA. TIM and the TABLE condition had the lowest number of doubts about the interface. The hypotheses, "participants will ask fewer questions related to understanding representations as the level of data visualization increases", is not supported. There seems to be a trade-off between ease of understanding and the number of operations needed to arrive at a decision. This result suggests that representations using multiple forms of representations (TEXTA and IRD) will be more difficult to learn than those that use fewer representation forms (TIM and TABLE). Research by Hornbæk et al. (2002) suggests that the need to process two views in a macro and micro level representation increases cognitive load and slows down the navigation process. The multiple representations involved in TEXTA and IRD may have a similar effect. The lower number of doubts about the TIM interface suggests that the graphical elements were easy to understand. This may be due to the use of sensory codes (Ware, 2004) of color, size and proximity, features which are processed with minimal effort by our perceptual system (Ware, 2004, Bederson & Shneiderman, 2003) and due to a predominant representation form in this condition. The TABLE condition was also cited as easier to use than the TEXTA or IRD. This may be due the relative familiarity of this representation to all users.

4.6 Strategies, credibility and quantitative analysis

According to Stone and Schkade (1991), decision makers adopt different strategies depending how the data is represented. From open interviews a majority of

participants in the TIM condition stated that it was difficult to figure out a strategy and there was wide variation amongst participants. Participants mentioned that they used the graphic to test their hypotheses about whether issues would be similar for individual countries and to see how strong the association of words would be at different levels of zoom. Participants voiced concerns that the graphic may result in biased conclusions as information about negative or positive sentiment is not obvious. Although, TIM was mentioned as easy to understand and is similar to other software in its use of frequency and proximity of words for analysis and subsequent representation, the fact that the graphic lacked depth in what it conveyed was an issue for some participants. However, TIM did allow users to view actual responses associated with a graphical selection. It was also observed by the researcher that participants would only read a few of the responses for each selection and relied more on the graphic and frequency analysis to make their decisions. This is congruent with research that suggests that when text and graphical information is available, users tend to pay more attention to the graphic. However, given the large number of data points, it is difficult to tell whether this was due to a preference for graphical information in general or a preference resulting from the overwhelming number of responses in this dataset. The general feeling of missing out on some aspect of the data was expressed by participants in TEXTA as well. The concern of bias expressed by participants in TIM was also stated in this condition by one participant who suggested that the order in which responses are listed should be randomized each time the list is selected in order to prevent the first few responses in the list swaying the user in one direction.

Participants used the frequency of words to decide which categories they would explore further and also in combination with bar charts in order to compare issues in different countries. Participants in IR mentioned that they found in difficult to develop a strategy. This was the only condition that stated that they needed more direct interactivity with the software. Although IR represents information in multiple forms, participants said that the data did not seem consistent across representations. For

example, the pivot table information did not match with the graphical representations below it. However, one participant mentioned that the line chart was more useful than the pivot table information as it was easier to pick up on issues. The greater dependence on the charts under the pivot table was noted in observations by the researcher also. However, participants seemed to like access to multiple representations.

The lack of organization of information in the TABLE condition was stated as a difficult aspect of working with the data here. Although, participants identified similar themes as all the other conditions and had a similar number of themes identified, they felt this representation was monotonous, tiring and overwhelming. One participant stated that it was good to be able to read all the comments as if it were represented in graphical form, there would be a tendency to put one's own interpretation on the data. However, it was seen that in trying to navigate through the data in this condition, some participants did use their own ideas of what might be important in each country to decide what to look for. According to Borgman (1996) participants focusing on a specific issue based on preconceived ideas was a sign of unsuccessful search behavior. However, in the same condition, there were participants who scanned the entire dataset before focusing on specific themes. This kind of analysis according to Borgman characterizes successful information search. Therefore, the strategies employed could also be a result of personal differences. Participants also voiced concern over the credibility of analysis based on frequency and that as a manager they would not trust this. This suggests that there is a chance of bias here also.

According to research, decision-makers adopt a specific strategy based on the amount of perceived effort required and resulting levels of accuracy (Kleinmuntz & Schkade, 1993). However, feedback about effort is more readily available than accuracy (Einhorn & Hogarth, 1978 as cited in Kleinmuntz & Schkade, 1993). In the time allotted for the task, participants had to develop a strategy to arrive a decision in

conditions that were unfamiliar. The TABLE condition though presented in software that all participants had used, nearly all of them had never used it to analyze qualitative data or dealt with the number of responses present in this dataset. The themes identified by all participants were similar to each other and to those identified by the researcher. However, the relative experience of arriving at these decisions, confidence and perceived effectiveness varied based on the representation used and sometimes within a representation. Although the final analysis was similar, the future actions that a manager would take would depend on their perceived performance. Results from this study suggest the specific influences of different representation forms described thus far on the decision-making process. However, it is important to view these results in light of the limitations of this study.

4.7 Limitations

The dataset used to create representations was from a survey that had a quantitative rating section with the list of possible issues and then followed by the open-ended question. A number of respondents copy-pasted from the list in the rating section for the answer to the open-ended question. Therefore, sometimes when participants investigated individual responses, they received a set of responses that lacked depth. This may have led to a reliance on the graphic or a dependency on frequency analysis in many conditions. As one participant who read hundreds of individual comments in the TABLE condition said "After reading 300 responses in the Philippines, everything looked the same". However, it is reasonable to expect a large amount of repetition given that the respondents were employees who worked in the same company and issues are likely to be common. Also, data points in real world data sets typically vary in their level of depth. Consequently, it is important to consider that the sense of missing out on nuances and the feeling that data points are redundant, are likely to arise in other data sets, and these concerns may increase as the size of the data set increases.

Although the representations were the focus of the study, there was additional meaning that a participant received based on the unique features of the software used in the condition. For example, IR and TEXTA had a sentiment analysis feature which was lacking in TIM and the TABLE condition. However, in order to gain insight into a specific form of representation, it was felt to be necessary to eliminate the information needs of users that we already knew about from experiments in the other two conditions. This was an intentional compromise in the level of control used in the study, in order to gain insight into a valuable nuance in qualitative data.

The scenario and decision-making task asked participants to identify and rank themes on a global and local level. However, the data from local decisions could not be analyzed as some participants wrote down only themes that were unique to a country while others stated all themes relevant to that country and subsequent rankings that were difficult to analyze as it was not known what their exact interpretation of the task for the local level was. Therefore, it is difficult to compare local decisions across participants and only their perceived confidence and effectiveness of representations could be analyzed.

Participants had limited time to learn the software and perform a complex decision-making task with a large complex dataset. In order to focus the study on representation and not features of the software and to address the different learning curves of the various participants, direct interactivity was curtailed. The training session focused just on how to read the representation. This may have limited the participants' ability to develop a strategy in a more efficient manner as direct experimentation with the tool may have facilitated a faster trial and error process. However, the priority in this experiment was to investigate the kind of strategies these representations would help develop in a specific time frame for novice users.

4.8 Future research

The results of this study suggests that representation of large qualitative data sets may have a stronger influence on the decision-making process and the

decisionmaker's subjective experience than on the final decisions made. In a real world scenario, whether a decision is followed through with action depends on how the decision-maker feels about his or her analytical experience. The effort required, the decision-maker's satisfaction with level of detail reached, and confidence in decisions made, are likely to play an important role in determining the real world impact of large scale qualitative research. It is important that future resesarch in data mining and visualization takes into account the interpretative nature of qualitative research, the importance of nuance and sentiment, and features of software design that lend credibility to results presented, from the qualitative researcher's perspective.

4.9 Conclusion

Current research in visual data mining and decision-making is characterized by a focus on the final choice phase of decision-making, the evaluation of simple tasks of either low or high complexity, and a proliferation of visualization and data mining techniques rather than systematic human factors research on user guidance. This study provides insights into the initial stages of decision-making process, problem recognition and situational analysis, for a real world data set, and a decision-making scenario consisting of tasks of varying levels of complexity.

In an hour and a half, participants learned new software features, strategized on how to analyze the data at hand and made decisions similar to that of the researcher who had months of experience with the data set. The results do not indicate that there is a specific representation that is appropriate for a decision-making scenario that consists of subtasks of varying levels of complexity. Instead, the results reveal that information representations of qualitative data, when highly graphical results in more operational effort but is easier to understand. Participants also expressed how engaging they found this form of representation. However, the nuances inherent in qualitative data make it necessary for representations of large qualitative datasets, to also feature graphical elements that can reveal sentiment or other nuance relevant to the dataset, as decision-makers tend to feel like they are missing out on important

information otherwise, and also tend to pay less attention to this information when it is text based.

Representations that are a composite of different forms require less operational effort but may take longer to understand and process. Successful search behavior is characterized by an initial scanning and overview of the data before focusing on specific themes. In order to encourage this behavior, representations that provide an overview and detailed information will allow users to form a more holistic understanding of the data and possibly reduce biases before they focus on the details.

Unless, a highly detailed representation of a dataset is also an organized representation, users will not be satisfied with the level of detail and may be overwhelmed.

Although users may identify similar themes from various kinds of representations, it is important for future research to identify elements that add credibility and enhance confidence as subsequent decisions depend on the user's confidence levels and satisfaction with their analysis, however accurate their objective performance may be. It is important to make apparent in layman terms how the software processes the data to lend it credibility.

Finally, as data mining software is adopted to analyze qualitative data in various fields, it is important to also design the data collection tools carefully, keeping in mind the logic that the software uses to analyze and represent the data, in order to avoid possible issues of data cleansing, redundancy and possibly, misleading frequency analysis.

APPENDIX A

Management Decision Making Scenario

You are to play the role of the VP of Workplace Futures, responsible for workplace design policies for company H, Inc. The company is an international operation and it initiates a companywide survey to get feedback from employees on their work environment. Employees are asked open-ended questions relating to a set of work environment attributes (see below). As the manager, you need to examine the data collected to see what the important areas are that employees voice concerns about, what the specific problems may be and what do they suggest as improvements? You then need to prioritize areas where you think action needs to be taken by the company both globally as well as prioritize issues for each country. You will use the help of a data analyst to navigate through the information for this task. You will drive the exploration through a chat client and the analyst will follow your directions on what you want to see and focus on. Please include in your communication with the analyst, questions you are looking to answer to aid your task, your rationale, and please tell the analyst as you identify themes and issues, what they are. You will have a chance to revise these themes after the exploration process if you choose to do so.

Workplace and Work Environment Survey Open Ended Questions

Q1 - Which of the attributes or elements of the work environment listed below are most important in meeting your business needs? In addition to what is listed above, are there other elements of the work environment that are also or more important in meeting your business needs.

Q2 – Please provide any additional comments or suggestions you have regarding the work environment

Attributes

- Engaging workplace (appropriate image of company H, welcoming environment, collaborative workplace)
- The types of workspace (open, enclosed, touch down, etc.) I need to be productive.
- Appropriate conference and meeting spaces available for my needs.
- A simple and effective process for reserving workspace/meeting rooms.
- Efficient printers, copiers and fax machines.
- Reliable tools for holding virtual meetings (conference calls, video conferencing).
- An appropriate level of safety and security.
- A clean environment (building and workspaces)

APPENDIX B

Response Sheet Results

Number of themes identified by participants in each condition:

	Number of Themes Identified

Condition	Participant	Global	UK	Philippines	Australia
TABLE	1	5	3	5	2
	2	4	4	4	4
	3	7	8	8	8
	4	5	5	5	3
	5	5	3	2	3
	6	8	0	0	0
	7	5	5	5	5
	8	7	5	4	1
TEXTA	9	4	1	4	3
	10	3	3	6	3
	11	12	2	6	2
	12	8	7	14	11
	13	4	1	3	1
	14	8	8	8	8
	15	3	1	4	1
	16	5	2	2	1
IRD	17	7	6	6	6
	18	9	5	3	3
	19	6	6	6	6
	20	5	6	5	6
	21	5	5	6	5
	22	6	6	6	6
	23	4	3	2	3
	24	6	6	6	6
TIM	25	6	5	6	4
	26	7	10	10	9
	27	8	5	6	5
	28	7	8	12	11
	29	5	5	5	5
	30	5	5	4	5
	31	5	2	3	1
	32	4	2	4	4

Percentage of participants identifying most commonly mentioned themes in each condition:

	THEMES					
CONDITION	Meeting	Clean	Safety	Workspace	Booking	Equipment
TABLE	87.5	100	87.5	87.5	87.5	87.5
TEXTA	100	37.5	25	87.5	62.5	75
IRD	100	87.5	75	87.5	75	100

TIM	87.5	75	62.5	75	25	87.5

APPENDIX C

Transcript Analysis Results

Condition	Participant	Operational Requests	Operational doubts	Interface doubts	Data doubts
TABLE	1	24	4	0	1
	2	43	3	1	0
	3	27	2	1	0
	4	65	4	4	1
	5	19	2	0	0
	6	53	3	2	6
	7	27	0	1	6
	8	53	0	2	1
TEXTA	9	23	1	5	0
	10	14	2	7	5
	11	49	2	2	1
	12	50	3	4	3
	13	33	2	3	3
	14	19	1	4	0
	15	30	4	7	1
	16	44	2	7	2
IRD	17	25	1	0	2
	18	50	5	10	0
	19	9	1	2	1
	20	35	5	2	0
	21	89	4	2	0
	22	53	4	8	1
	23	57	3	4	2
	24	22	2	10	0
TIM	25	87	7	7	0
	26	61	5	0	0
	27	56	7	2	0
	28	76	9	2	0
	29	104	8	0	1
	30	79	7	1	3
	31	75	5	1	0
	32	79	5	0	1

APPENDIX D

Survey

Basic Information Participant code:

No. of years of work experience:

Qualitative data analysis experience:

Qualitative Software used if any:

Common Questions

Please indicate the degree of agreement/disagreement with the following statements based on your exploration and decision making process.

1. The software results gave me a good overview of the major themes in the data

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

2. It was easy to identify themes in the data

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

3. I am satisfied with the level of detail I could reach in the analysis of the data

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

4. I am confident with the decisions I made on global priorities

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

5. I am unsure of my analysis of differences in priorities for each country

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

6. I would not use this form of data representation to analyze this type of data and make decisions in the future if I had a choice

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

7. I found it useful to be able to "drill down" to individual comments to help understand the data

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

8. I think this representation helped me arrive at good decisions quickly and easily.

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

Condition specific questions

TABLE:

1. I found the number of responses overwhelming

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

TEXTA:

1. I found the bar chart feature helpful for making country specific decisions.

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

2. The colors used to represent the "types" were easy to remember while exploring the data.

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

3. The category web/bar chart feature helped me understand easily how each category was related to other categories.

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

IRD:

1. I found the visualization or overview map confusing

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

2. I found it useful to be able to see quantitative information such as frequencies and percentages

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

TIM:

1. I found the graphic images confusing.

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

2. I liked the use of size, shape, and color in the graphic images.

Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

APPENDIX E

Survey Results

Scoring: 1= Strongly disagree, 2= disagree, 3=neither agree or disagree, 4=agree, 5=strongly agree

TABLE

Statement	P1	P2	P3	P4	P5	P6	P7	P8	Mean	SD
1	4	4	4	3	4	4	5	2	3.75	0.88640526
2	4	3	2	1	4	2	5	1	2.75	1.488047618
3	4	3	2	1	4	2	4	2	2.75	1.164964745
4	3	4	1	2	4	4	5	3	3.25	1.281739889
5	2	2	3	4	3	1	2	4	2.625	1.060660172
6	2	4	5	5	4	3	2	5	3.75	1.281739889
7	5	2	3	4	4	4	5	4	3.875	0.991031209
8	4	3	1	1	2	2	4	1	2.25	1.281739889

9*	2	3	5	3	4	3	2	5	3.375	1.187734939

* TABLE specific statement:

9 - “I found the number of responses overwhelming”

TEXTA

Statement	P9	P10	P11	P12	P13	P14	P15	P16	Mean	SD
1	4	4	4	4	4	4	4	4	4	0
2	2	4	5	3	4	4	3	4	3.625	0.916125381
3	3	3	4	3	4	2	2	2	2.875	0.83452296
4	4	4	4	3	4	4	3	5	3.875	0.640869944
5	4	4	3	4	4	4	4	1	3.5	1.069044968
6	2	4	2	4	3	3	3	1	2.75	1.035098339
7	5	4	4	2	4	4	4	5	4	0.9258201
8	2	4	4	4	3	4	2	4	3.375	0.916125381
9*	4	4	4	4	4	3	4	5	4	0.534522484
10*	4	3	4	2	2	4	2	1	2.75	1.164964745
11*	4	3	4	4	4	4	2	2	3.375	0.916125381

* TEXTA specific statements:

9 - I found the bar chart feature helpful for making country specific decisions. 10 - The colors used to represent the "types" were easy to remember while exploring the data.

11 - The category web/bar chart feature helped me understand easily how each category was related to other categories.

IRD

Statement	P17	P18	P19	P20	P21	P22	P23	P24	Mean	SD
1	4	5	4	4	5	4	4	4	4.25	0.46291005
2	5	4	4	2	5	2	4	3	3.625	1.187734939
3	4	4	4	2	4	2	3	4	3.375	0.916125381
4	4	3	3	1	4	4	4	4	3.375	1.060660172
5	2	4	3	5	2	2	3	2	2.875	1.125991626
6	4	2	2	4	2	2	2	2	2.5	0.9258201
7	4	5	4	3	5	4	4	3	4	0.755928946
8	4	3	4	2	5	5	4	4	3.875	0.991031209
9*	4	4	4	1	4	4	5	4	3.75	1.164964745
10*	4	4	4	4	4	5	4	5	4.25	0.46291005

* IRD specific statements:

9 - I found the visualization or overview map confusing.

10 - I found it useful to be able to see quantitative information such as frequencies and percentages.

TIM

Statement	P25	P26	P27	P28	P29	P30	P31	P32	Mean	SD
1	5	5	4	4	5	5	4	4	4.5	0.534522484
2	4	5	4	3	3	5	4	4	4	0.755928946
3	4	4	4	2	4	4	3	2	3.375	0.916125381
4	2	4	4	4	2	4	4	4	3.5	0.9258201
5	4	2	4	2	4	2	2	2	2.75	1.035098339
6	2	1	4	5	2	2	2	2	2.5	1.309307341
7	4	5	4	4	5	5	4	3	4.25	0.707106781
8	2	5	3	4	5	4	4	3	3.75	1.035098339
9*	4	1	2	2	1	1	2	1	1.75	1.035098339
10*	4	5	4	1	5	5	5	5	4.25	1.38873015

* TIM specific statements:

9 - I found the graphic images confusing.

10 - I liked the use of size, shape, and color in the graphic images.

APPENDIX F

Sample chat transcripts

TABLE

[17:19] analystHFE: session start

[17:20] analystHFE: you are viewing the global data set which is a combination of three countries.

[17:20] analystHFE: you have one hour

[17:20] infohfe: Okay, thanks

[17:20] infohfe: Let

[17:20] infohfe: us begin from the beginning

[17:21] infohfe: Give me a description of the type of people surveyed.. for e.g. managerial, staff, all of them, etc.

[17:21] analystHFE: they cover a number of levels. You do not have access to that information for this task

[17:21] analystHFE: they do cover senior levels to staff levels

[17:21] infohfe: okay

[17:22] infohfe: So we only have access to the answers to the open ended questions given in the sheet

[17:22] analystHFE: yes and which country it is from

[17:23] infohfe: Is there any way we can q2uantify answers to Q 1

[17:23] infohfe: LEt me be specific..

[17:23] analystHFE: ok

[17:24] infohfe: First off, do you have answers to q1 and q2 listed separately

[17:24] analystHFE: no. they are merged

[17:26] infohfe: I'm first going try to figure out which of the listed attributes ranks important to most people.. we will first look at it globally. I will think of how to obtain this info.

[17:26] analystHFE: ok

[17:26] infohfe: Simple try: Find all response that have the word "engaging"

[17:27] infohfe: how many entries is that?

[17:27] analystHFE: two responses

[17:27] infohfe: ok

[17:27] infohfe: that's not very promising.. let's try efficient and words related to efficient

[17:27] infohfe: efficien*

[17:28] analystHFE: there are 103 responses with the word 'efficient' in them

[17:28] analystHFE: what would you like to do

[17:28] infohfe: I'm going quickly scan the entries..

[17:28] analystHFE: should i check for other efficien*

[17:28] analystHFE: ok

[17:28] infohfe: yes

[17:28] infohfe: check for effiecn*

[17:28] infohfe: (correct sp)

[17:29] analystHFE: 5 for efficiency

[17:29] infohfe: okay is there a way to look at the country of response when you look for a search item?

[17:30] analystHFE: sure. would you like me to go back to the original request to view the 103 responses related to the word "efficient" before we do this?

[17:30] infohfe: yes please

[17:30] analystHFE: i will go through each one, and they will be highlighted in the body of Excel and you can see the corresponding country. There are also country wise spread sheets

[17:30] infohfe: got it, when I say next, can go to the next entry

[17:31] infohfe: n

[17:31] infohfe: n

[17:31] infohfe: n

[17:31] infohfe: n

[17:31] infohfe: n

[17:32] infohfe: n

[17:32] infohfe: n

[17:32] infohfe: n

[17:32] infohfe: total how many entries again?

[17:32] analystHFE: 103 for efficient

[17:32] infohfe: not this search..

[17:32] analystHFE: out of 1574

[17:32] infohfe: total # responsed

[17:33] analystHFE: globally 1574

[17:33] infohfe: I want to look at all entries with the word important

[17:34] analystHFE: 231 . should i do the same as before.

[17:34] infohfe: yes

[17:34] infohfe: actually.. wait at each entry till I say "n"

[17:34] infohfe: for next!

[17:34] analystHFE: okay

[17:35] infohfe: what is type of workspace?

[17:35] analystHFE: type of workspace is whether is it open plan, enclosed etc

[17:35] infohfe: for e.g. what does touch down mean

[17:35] analystHFE: a desk that you can just use when you come in for a short time

[17:35] analystHFE: to the office

[17:35] infohfe: ok.

[17:36] infohfe: I need help with some context.. I want to first figure out what type of industry this company is..

[17:37] analystHFE: they are in consulting and include analysts. software engineers, managers and all departments usually found in a large global company

[17:37] analystHFE: this survey covers all employees in all departments

[17:40] infohfe: okay.. back to keyword search "important"

[17:40] analystHFE: il lwait for you to say n

[17:40] infohfe: n

[17:41] infohfe: n

[17:41] infohfe: n

[17:41] infohfe: n

[17:41] infohfe: n

[17:42] analystHFE: you have forty minutes left

[17:42] infohfe: ok

[17:42] infohfe: n

[17:42] infohfe: n

[17:43] infohfe: n

[17:44] infohfe: It looks to me that a lot of people are interested in better hardware...

[17:44] infohfe: let's search for printer

[17:44] analystHFE: okay

[17:44] analystHFE: 166 cells with the word printer

[17:45] analystHFE: would you like to read them like last time?

[17:45] infohfe: yes..

[17:45] analystHFE: can u tap on the n loudly so i dont need to see the chat window to know you want to proceed?

[17:45] infohfe: sure..

[17:45] infohfe: NEXT

[17:46] infohfe: let's now sort the data by country

[17:46] infohfe: Aus

[17:46] analystHFE: ok. which country would you like to see first.

[17:46] analystHFE: aus

[17:46] infohfe: look for printer in this tab

[17:47] infohfe: oka

[17:47] infohfe: okay

[17:47] infohfe: okay.. have some structure now..

[17:48] analystHFE: ok

[17:48] infohfe: Can you tell me the rough total entries per country?

[17:48] analystHFE: sure - australia - 237

[17:48] analystHFE: phillipines - 903

[17:48] analystHFE: UK - 423

[17:49] infohfe: here's what I want to do..

[17:50] analystHFE: you have thirty minutes

[17:50] infohfe: First look at global responses.. find words the keywords that I will mention and highlight them in colors I say..

[17:50] analystHFE: okay.

[17:51] infohfe: let's look for "engaging"

[17:51] infohfe: yellow.

[17:51] infohfe: before we go on..

[17:51] infohfe: can we sort data by cell color?

[17:51] analystHFE: no

[17:51] infohfe: okay..

[17:51] infohfe: can you insert a column?

[17:52] analystHFE: sure

[17:52] infohfe: let's type in 1 in cells that have engaging.. don't know if thjat's easy

[17:52] analystHFE: done

[17:53] infohfe: well.. it's going to be tough wehn we get to 100's of entries.. thinking of a bettwe way to do this..

[17:53] analystHFE: okay

[17:56] infohfe: let's look for "type of workspace"

[17:57] analystHFE: 93

[17:57] infohfe: also look for "types of workspace"

[17:57] analystHFE: 139

[17:58] infohfe: look for conference

[17:58] analystHFE: 208

[17:59] infohfe: look for "reserv"

[17:59] analystHFE: 86

[17:59] infohfe: printer

[17:59] analystHFE: 166

[18:00] infohfe: safety

[18:01] infohfe: look for that word

[18:01] infohfe: safety

[18:01] analystHFE: 148

[18:01] analystHFE: you have twenty minutes left

[18:01] infohfe: can you look for "clean"

[18:01] analystHFE: 233

[18:02] infohfe: So okay.. I'm now going to total these..

[18:02] analystHFE: okay.

[18:03] infohfe: so there are 1075 entries in the keyeords I searched.. There are 500 ish entries without these keywords.. I am interested in looking at these.. any suggestions?

[18:04] analystHFE: I can only help you perform functions in excel

[18:04] infohfe: Ok, let me think a bit..

[18:04] analystHFE: ok

[18:06] analystHFE: you have 15 minutes left

[18:09] infohfe: okay.. when you seach an entry and find it.. can you delete that entry from the file (my way of looking at the rest of the 500 entries I want to look at)

[18:09] analystHFE: sure

[18:09] analystHFE: where would you like tos tart

[18:09] infohfe: let's do that..

[18:09] analystHFE: okay

[18:09] infohfe: "effiecien"

[18:10] infohfe: hold on

[18:10] infohfe: one sec

[18:10] analystHFE: some entries have more than just one attribute listed

[18:10] infohfe: right.. just thought of that too!

[18:11] infohfe: that's okay.. I will lose some data.. overlap is not a major concern to me now.

[18:11] analystHFE: okay

[18:11] infohfe: let's delete all thse entries

[18:11] infohfe: let's find "type of", and let me skim quickly

[18:12] analystHFE: is this view ok?

[18:12] infohfe: yes

[18:12] infohfe: let's delete all these

[18:12] infohfe: actually.. hikld on

[18:12] infohfe: hold on

[18:13] infohfe: let's go back and undo the delete of efficien

[18:13] analystHFE: you have 8 minutes

[18:13] infohfe: sorry.;.

[18:13] analystHFE: the delete of efficien did not work anyway

[18:13] infohfe: okay.. too bad..

[18:13] infohfe: okay.. I'm going to jot stuff now.

[18:13] analystHFE: okay. you can have 5 min after the exploration process to complete your response sheet also. However I will have to close the window at the end of seven minutes starting now

[18:14] infohfe: ok..

[18:15] infohfe: let's look for "recreation"

[18:15] infohfe: "music"

[18:16] infohfe: gym

[18:16] infohfe: facility

[18:16] infohfe: parking

[18:17] infohfe: transport

[18:17] infohfe: out of time?

[18:17] analystHFE: three minutes

[18:17] infohfe: let's look for transport

[18:17] infohfe: look for internet

[18:18] infohfe: okay..

[18:18] infohfe: think I'm done here.

[18:18] analystHFE: ok. do you need time (5 min) to fill in your excel sheet?

[18:18] infohfe: yes.. very much.

[18:18] analystHFE: okay

TEXTA:

[16:58] analystHFE: okay. we are going to start now. you have one hour to explore and you may take 5 min after to complete your excel sheet if you would like.

[16:58] infohfe: ok

[16:58] infohfe: need to see thye negative qualifiers

[16:59] infohfe: yup

[16:59] analystHFE: would you like to see the list of words that come under it

[16:59] infohfe: ok, we'll look at the categories now

[16:59] analystHFE: ok let me nkow if you need to scroll down in that pane

[17:00] infohfe: i need to see meeting rooms and clean in the category screen

[17:00] infohfe: the connection

[17:00] infohfe: thx

[17:00] analystHFE: would you like to see the web also?

[17:01] infohfe: yup

[17:01] infohfe: oops

[17:01] infohfe: :)

[17:01] infohfe: could i see the connection between the meeting rooms and the negative things?

[17:02] infohfe: is there a way?

[17:02] analystHFE: yes. one sec

[17:02] infohfe: ok

[17:03] infohfe: so negative comments about meeting rooms- 25%

[17:03] analystHFE: yes. let me know if you would like to read those

[17:03] infohfe: can we see thje web?

[17:03] infohfe: could i see the complaints for the MR, under negative?

[17:03] analystHFE: sure

[17:04] infohfe: like, what they did not like

[17:04] analystHFE: the data pane lists the responses that have meeting room and negative

[17:07] infohfe: you can go faster

[17:08] infohfe: ok, id like to see the same thing on copiers

[17:08] analystHFE: ok

[17:09] analystHFE: let me know when u have read enough

[17:09] analystHFE: this is -ve and printers copiers

[17:09] infohfe: yup

[17:10] infohfe: 331 negative !!!!

[17:10] infohfe: comments?

[17:10] analystHFE: 40 for printer copiers and negative together

[17:11] infohfe: there were 80 for the last category right?

[17:11] infohfe: ah, ok

[17:11] analystHFE: let me check . one sec

[17:11] analystHFE: yes 80

[17:11] infohfe: ok ty

[17:11] infohfe: ok, restrooms

[17:12] infohfe: no complaints on the restrooms

[17:13] analystHFE: these three have the exact word "Restroom"

[17:13] infohfe: so i should look at lavatory, toilet or others?

[17:13] analystHFE: you could if you would like. should i look for them?

[17:13] infohfe: its an important detail, right?

[17:13] analystHFE: it is up to you to decide

[17:13] infohfe: well, i think clean would be more interesting

[17:14] infohfe: actually, clean + negative

[17:14] analystHFE: okay. should i look for clean and negative or just clean

[17:14] analystHFE: ok

[17:14] infohfe: good, seems that there are no complaints on lcleanliness of the workspace

[17:15] infohfe: need computers

[17:15] analystHFE: ok

[17:15] infohfe: + negative

[17:16] analystHFE: remember that because people talk about multiple things in one comment, if any of those other aspects have a negative connotation it will be catagorized as negative. and so they may not neccarily be talking about negative and computer but negative and something else in that comment and computer just happens to be someting else they commented on.

[17:16] infohfe: thats true

[17:17] infohfe: ok

[17:17] infohfe: need to see comments on space

[17:17] analystHFE: negative and space or just space

[17:17] infohfe: just space

[17:18] infohfe: and then negative +

[17:18] analystHFE: let me know when you would like negative

[17:18] infohfe: lets see the grid

[17:18] analystHFE: the web?

[17:18] infohfe: yup[

[17:18] infohfe: ok, negatives
[17:19] infohfe: ok, enough

[17:19] infohfe: how about tools

[17:19] analystHFE: ok

[17:20] analystHFE: viewing just tools

[17:20] infohfe: the web

[17:20] infohfe: phone

[17:20] infohfe: and virtual conferences

[17:21] infohfe: ok

[17:21] infohfe: any negatives about these?

[17:23] infohfe: ok

[17:23] infohfe: need to see productive

[17:23] infohfe: the web

[17:24] infohfe: workspace and negative

[17:26] infohfe: ok, environment and safety

[17:27] infohfe: each of them, and negatives

[17:27] infohfe: ok

[17:27] infohfe: web

[17:28] infohfe: ok

[17:28] infohfe: level of safgety

[17:28] analystHFE: you have 30 minutes left

[17:28] infohfe: need to see the negatives

[17:29] infohfe: ok

[17:30] infohfe: need to see chairs
[17:33] infohfe: need to see some of them again to watch for specific countries problems

[17:33] analystHFE: you can also look at the bar charts for country info.

[17:33] analystHFE: UK - Blue, phill - pink etc

[17:34] infohfe: yeah, thats what i mean

[17:34] analystHFE: clicking on those will give you country specific. what would you like to see

[17:34] infohfe: to look back at some we have covered and see the country info

[17:34] analystHFE: ok

[17:34] infohfe: ahhhhh

[17:34] infohfe: meeting rooms

[17:34] infohfe: ok

[17:35] infohfe: meeting rooms in uk negatives

[17:36] infohfe: ok, phillipines

[17:37] infohfe: phones and philipines and negatives

[17:37] infohfe: i saw smthg

[17:37] analystHFE: ok

[17:39] infohfe: environment and uk

[17:39] analystHFE: negative and uk or just uk?

[17:39] infohfe: lets see why philipines has so many comments on environment

[17:40] infohfe: ok, general considerations

[17:40] analystHFE: ok

[17:41] infohfe: many in the Ph say they need a clean env
[17:41] analystHFE: yes

[17:41] infohfe: does that mean they dont have it?

[17:41] analystHFE: i dont know

[17:41] infohfe: could be

[17:41] infohfe: ok

[17:41] analystHFE: ok

[17:41] infohfe: what could uk be low on?!!!!

[17:42] infohfe: lets see

[17:42] infohfe: please click on each category

[17:42] infohfe: ok

[17:42] infohfe: great

[17:42] analystHFE: ok.

[17:43] infohfe: can you tell me again how many responses come from each country?

[17:43] analystHFE: 913 phillipines

[17:43] analystHFE: 424 approx UK

[17:43] infohfe: thats why most are ph

[17:43] analystHFE: 232 approx Aust

[17:43] infohfe: aaaaaaaaaaa

[17:44] infohfe: i should have considered this all the way

[17:44] infohfe: :(

[17:44] infohfe: ok

[17:44] infohfe: uk is not happy wt space

[17:44] analystHFE: ok
[17:44] infohfe: seems like

[17:44] analystHFE: ok

[17:44] analystHFE: you have 15 minute sleft

[17:45] infohfe: can you minimize the chat window

[17:45] analystHFE: you can choose to look at the number of responses for each country for each category also if youw ould like

[17:46] infohfe: can i see the respondents as percentages?

[17:46] infohfe: percent of uk

[17:46] infohfe: like on space, and negative

[17:47] analystHFE: ok. let me try

[17:47] analystHFE: 23 respondents out of 232approx

[17:47] analystHFE: out of 424 approx sorry

[17:47] infohfe: ok

[17:48] infohfe: it would be helpful to have a percentage function for that

[17:48] analystHFE: there is no percentage function here

[17:48] infohfe: i know, but as an interface thing :)

[17:48] analystHFE: yes it would

[17:48] infohfe: ok

[17:49] infohfe: i guess ill put them in order now

[17:49] analystHFE: ok.

[17:49] analystHFE: you have 9 minutes left to explore the data if you choose to at any point

[17:49] infohfe: ok
[17:49] analystHFE: will close the window at 5:58

[17:50] infohfe: actually

[17:50] infohfe: lets l;ook at bfax machines

[17:50] analystHFE: okay

[17:50] analystHFE: let me know if you want the non 100% bars

[17:50] infohfe: web view

[17:50] infohfe: pls

[17:51] infohfe: can i see a breakdown of the negatives

[17:52] infohfe: and the countries

[17:53] infohfe: australia negative offices

[17:53] infohfe: 10 comments

[17:54] infohfe: ok im done

[17:54] analystHFE: ok thank you!

IRD:

[11:03] analystHFE: we are starting now. you are viewing the overview map of the global data. you have one hour to explore the data and you may have 5 min after to complete your excel sheet

[11:04] infohfe: OK, Lets go!

[11:04] analystHFE: ok!

[11:04] infohfe: can you show me the darkest blue square by environmentws

[11:06] infohfe: can I see under themes conferenfce and meeting spaces

[11:06] analystHFE: it has four subthemes

[11:07] infohfe: lets look at availability

[11:07] infohfe: subtheme
[11:07] analystHFE: let me knwo if you want to see the full comment for any of thses

[11:08] infohfe: can you scroll back up to the image map

[11:08] infohfe: please scroll through comments

[11:09] infohfe: can I see terms

[11:10] infohfe: close this please

[11:11] infohfe: can we go back to the main image map

[11:12] infohfe: I would like to explore issues with machines and copiers

[11:13] analystHFE: you have the following options to explore that - through clickin on the overview map, through selecting "printer, copier and fax machine in the themes or by looking at any of the exhibits in the bottom left hand corner

[11:13] infohfe: lets first click on overview map

[11:14] infohfe: can I see exhibits

[11:14] analystHFE: which one would you like to see

[11:14] infohfe: themes by senitment

[11:15] infohfe: when it searches for "negative sentiment" what kinds of terms is it using to filter the data?

[11:15] analystHFE: i can show you. one sec

[11:16] infohfe: ok, thanks let's look at the charts again

[11:16] analystHFE: let me knwo if you wnat me to explain any of the numbers.

[11:17] infohfe: can I see country by sentiment exhibit

[11:17] analystHFE: sure

[11:18] infohfe: is this just a general one... is there anyway to view the country by sentiment in relation to certain issues such as conference, machines

[11:19] analystHFE: yes. you can filter this table for themes or you could go to the country by themes and filter for sentiment
[11:19] infohfe: lets do the second option

[11:19] infohfe: can I see a bar chart

[11:20] analystHFE: let me know when you want to filter this by sentiment if you do

[11:20] infohfe: if I want to be taking notes I do that on the spread sheet right?

[11:20] analystHFE: that would be good. if you wnat to use paper that is fine too but i need to have the info we talked about filled in the excel sheet in the end

[11:21] infohfe: ok

[11:21] analystHFE: you can take random notes on paper if you want too

[11:21] infohfe: can you filter by senitment now?

[11:21] analystHFE: negative sentiment. let me know if yo uwant positive

[11:22] analystHFE: you can also view any of the comments contained in this table if you would like

[11:22] analystHFE: note - conf and meeting spaces has subthemes

[11:24] infohfe: can I see comments for Australia distrations

[11:27] infohfe: ok I'm done with this

[11:27] infohfe: can we go back to the main chart

[11:27] infohfe: this one

[11:27] analystHFE: this one

[11:27] analystHFE: it is still filtered for negative sentiment -note

[11:28] infohfe: can I see phillipines printer, copier fax

[11:29] infohfe: ok go back

[11:30] infohfe: can I see UK conf. space

[11:30] analystHFE: would you like to see overall or for the subthemes

[11:30] infohfe: sub theme of availability

[11:31] infohfe: is it specific to UKI

[11:31] infohfe: ?

[11:31] analystHFE: this selection yes..there are availability for other countries also

[11:33] infohfe: ok return to main graph

[11:34] analystHFE: do you mean the previous one with the otehr themes also or here

[11:34] infohfe: no this is good

[11:35] infohfe: lets go back to the main chart... no sub themes

[11:36] analystHFE: you have 30 minutes left

[11:36] infohfe: ok

[11:36] infohfe: lets go to the overall main map again

[11:36] analystHFE: the grid one?

[11:36] infohfe: can I see themes by sentiment

[11:36] analystHFE: sure

[11:37] analystHFE: note - you can also filter this by country if you want

[11:38] infohfe: wouldn't it be the same as the last chart?

[11:38] analystHFE: yes. just in case

[11:38] infohfe: sure lets see

[11:38] analystHFE: which country?

[11:38] infohfe: UIK

[11:38] infohfe: can I see the remainder

[11:39] analystHFE: remainder for negative sentiment or overall?

[11:39] infohfe: negative
[11:39] infohfe: ok

[11:39] infohfe: go back

[11:40] infohfe: now can I see australia

[11:41] infohfe: can I see printer copier fax

[11:42] infohfe: ok go back

[11:43] infohfe: lets go to phillipines

[11:43] infohfe: can I see printer

[11:44] infohfe: ok

[11:44] infohfe: go back

[11:45] infohfe: can I see conf.

[11:45] infohfe: actually can you bring up subthemes

[11:45] analystHFE: would you like to read specific sub theme comments or overall?

[11:46] infohfe: I just wanted to look at the chart for now

[11:47] infohfe: ok can I see comments on availability

[11:48] infohfe: lets go back to the very first screen we saw, the overall map

[11:50] infohfe: ok... go back to sentiment chart (overall not by country)

[11:51] analystHFE: this is themes by sentiment

[11:52] infohfe: can I see remainder comments for negativew

[11:54] infohfe: do all the countries use the same type of officing strategy... like "hot desking"

[11:54] analystHFE: yes

[11:54] infohfe: ok

[11:54] infohfe: and they all have open plans?
[11:54] analystHFE: yes

[11:55] analystHFE: you can look at the full comment to any of these to see which country it is from also

[11:56] infohfe: go back

[11:56] infohfe: yes here

[11:57] infohfe: is there any way that the program can filter out when people just copy and paste responses, like you described... from looking at this information you would believe that there is an issue with printers copiers and fax... but when you actually look at most of the comments, they have just copied and pasted "effiicient printers, copiers, fax"

[11:58] analystHFE: i am not sure if there is a way to do that. i could search for those terms but it would just select those for you instead of excluding them

[11:59] analystHFE: or if you choose a certain theme, i can search for the exact phrase and it would show you how many of the documents under that theme have the exact phrase

[11:59] analystHFE: you have 6 minutes left

[11:59] analystHFE: to look at the data

[11:59] infohfe: hm... I'm not sure what to look at next

[12:00] analystHFE: it is up to you. you can choose to work on your excel sheet or look at anything else. you do not have to take the entire one hour

[12:00] infohfe: I would like to filter these again by country

[12:00] infohfe: one last time

[12:00] analystHFE: okay.

[12:00] analystHFE: which country

[12:00] infohfe: UK first

[12:01] infohfe: can I see distractionls

[12:03] infohfe: ok
[12:03] infohfe: go back

[12:03] infohfe: lets see australia

[12:03] infohfe: conf room

[12:04] infohfe: ok

[12:05] infohfe: philipines

[12:05] analystHFE: you have one minute left

[12:05] infohfe: distractions

[12:06] infohfe: ok I'm done... I'll work on the spread sheet now

[12:07] analystHFE: ok thank you!

TIM

[14:57] analystHFE: ok

[14:57] infohfe: can you zoom in to 25 percent?

[14:58] infohfe: so can you select the meeting segment and the one above it?

[14:58] analystHFE: done

[14:59] infohfe: maybe also select the one to the right of meeting

[14:59] infohfe: ok, can you now select the type island

[14:59] analystHFE: do you want to combine it with the original meeting selections or just the type island?

[14:59] infohfe: just the type island for now

[15:00] infohfe: ok, can you show me about fifteen of the comments with 6/8 frequency

[15:00] analystHFE: ok

[15:01] infohfe: ok

[15:01] infohfe: can we go back to the map

[15:01] infohfe: can you now select the first two rings of the main island

[15:02] infohfe: can i see the comments down to 5/12

[15:02] infohfe: the end of 5/12

[15:02] analystHFE: sure

[15:03] infohfe: you can go through them a bit faster

[15:04] infohfe: you can stop at 5/12

[15:04] infohfe: the begging of i mean

[15:04] infohfe: beginning

[15:05] infohfe: let's go back to the map

[15:05] infohfe: can you please highlight the security island

[15:06] infohfe: there seems to be a major worldwide problem with meeting space availability, though it may be concentrated in the philipines

[15:06] analystHFE: ok

[15:06] infohfe: ok, can we now begin cycling through the comments on the irght

[15:06] infohfe: right

[15:06] analystHFE: do you want to specify a number of responses or will you tell me when to stop?

[15:06] infohfe: i will tell you when to stop.

[15:06] analystHFE: ok

[15:07] infohfe: ok, that is good

[15:07] analystHFE: ok

[15:07] infohfe: is there a way to sort responses by the frequency of all thesaurus words?

[15:08] analystHFE: the most frequent is at the top. Do you mean in another way?

[15:08] infohfe: yes, in terms of the responses, not the words
[15:08] infohfe: i guess you could select all segments

[15:08] analystHFE: ok

[15:09] infohfe: ok, can we now cycle through the comments please

[15:12] analystHFE: you have forty four minutes left and three maps left. Would you like to go through these responses or move to a specific country

[15:12] infohfe: why don't we move to the UK map

[15:13] analystHFE: ok

[15:13] infohfe: ok

[15:13] infohfe: why don't you start by selecting the meeting ring [15:14] infohfe: can we look at the 6/10 and the 5/10 comments?

[15:15] analystHFE: we are at the last 5/10

[15:15] infohfe: ok, thanks. i have noted a particular problem in th lack of available workspace, and a problem in the types of workspaces available

[15:16] infohfe: also, people appear to be frustrated with their printing options

[15:16] analystHFE: uhuh

[15:16] analystHFE: ok

[15:16] infohfe: ok, let's return to the kmpa

[15:16] infohfe: map

[15:16] infohfe: can you highlight the segment to the left of the room segment

[15:17] infohfe: in the second ring

[15:17] infohfe: yea

[15:17] infohfe: can you select those two segments so I can see some of the responses?

[15:18] infohfe: thanks, that'll do

[15:18] analystHFE: ok

[15:18] infohfe: it appears that the focus in the UK office is on availability of workspaces, as opposed to meeting rooms

[15:18] analystHFE: ok

[15:19] infohfe: just as a last look, can we select the entire map as we did with the first map?

[15:19] infohfe: actuall\y, wait

[15:19] infohfe: can you first select the third ring of the main island

[15:19] infohfe: i want to see the conference responses

[15:21] infohfe: perhaps i need to amend my previous statement--there is a problem with conference availability, but it is a lesser priority in the UK

[15:22] infohfe: they seem to be much more concerned with the availability of workspaces in their open plan

[15:22] analystHFE: ok

[15:22] analystHFE: should i keep going through the responses?

[15:22] infohfe: shall we go on to the next map?

[15:22] analystHFE: ok

[15:23] analystHFE: this is phillipines at 0% zoom with random words highlighted

[15:23] infohfe: why don't we start this time by selecting the entire map and looking at the responses with the most frequent thesaurus words

[15:24] analystHFE: do you want me to t=go thru the responses now?

[15:24] infohfe: yes please

[15:24] analystHFE: ok

[15:26] infohfe: ok, let's zoom in now

[15:26] infohfe: to 25 percent

[15:27] infohfe: a quick scan just of the word frequencies and adjacencies tells me again that the most important factor here is the type of workspaces available to employees

[15:28] infohfe: can we unselect all

[15:28] analystHFE: you have 28 minutes

[15:28] infohfe: question--when all is selected, how come only some retain their colors?

[15:28] analystHFE: oh those words were the ones that were highlighted initially as word landmarks

[15:28] analystHFE: they retain their color because of that

[15:28] infohfe: oh ok, i misunderstood

[15:29] infohfe: can you highlight the segments in the first ring of the main island

[15:29] analystHFE: can u describe how you misunderstood it?

[15:29] infohfe: and also the first ring of the central-left island

[15:30] infohfe: i thought that the colorless ones were infrequent, but that is obviously incorrect

[15:30] infohfe: just a misunderstanding of process

[15:30] analystHFE: no when a segment is selected the word usually loses its color. it is only because we highlighted it that it retains its color

[15:31] analystHFE: how did you view the map earlier when we highlighted everything?

[15:31] infohfe: correctly--i was only looking at the responses then

[15:31] analystHFE: ok. let me know if you need other clarifications. I have now highlighted the central leftisland ring one

[15:32] infohfe: ok, so scanning the map, the central reported areas have to do with types of workspaces, meeting rooms, and the security and safety of the working environment

[15:32] analystHFE: ok

[15:32] infohfe: can you now select the first and second rings of the main island, so we can look at responses?

[15:33] infohfe: sorry can you go back to the map really quickly

[15:33] analystHFE: ok

[15:33] infohfe: ok thanks, back to the responses

[15:35] infohfe: ok, let's return to the map

[15:35] analystHFE: ok

[15:36] infohfe: can you select the workspace island so that we can look at those responses?

[15:36] analystHFE: you have twenty minutes left

[15:36] infohfe: thanks

[15:38] infohfe: one interesting thing that i have noticed is that respondents who are concerned about workspace issues tend to copy and paste, while people who take issue with conference space availability take much more time and put much more thought into their responses

[15:38] analystHFE: i see

[15:38] analystHFE: ok

[15:38] infohfe: it could be an indication of a more serious problem

[15:38] infohfe: anywho

[15:38] infohfe: let's return to the map

[15:38] analystHFE: would you like me to go thru the rest of the responses in this island?

[15:38] analystHFE: ok

[15:39] infohfe: can i see the responses from the entire security island?

[15:39] infohfe: ok

[15:40] infohfe: lastly, i would like to look briefly at the printer island's responses

[15:41] infohfe: ok, that'll do

[15:41] analystHFE: ok

[15:41] infohfe: let's move on to Australia

[15:41] analystHFE: ok

[15:42] analystHFE: this is australia at 0% zoom with random words highlighted

[15:42] infohfe: can you select the first ring on the main island please

[15:42] infohfe: let's look at the response

[15:42] infohfe: s

[15:44] infohfe: there seems to be a disagreement about the availability of meeting rooms

[15:44] infohfe: let's go back to the map

[15:44] analystHFE: ok

[15:44] infohfe: can we zoom to 25 percent?

[15:45] infohfe: ok, now highlight the second ring of the main island

[15:45] infohfe: ok, now please select those two rings and look at responses

[15:45] analystHFE: teh second ring and

[15:45] infohfe: the first and second

[15:45] infohfe: sorry

[15:46] analystHFE: ok

[15:46] infohfe: let's look at responses

[15:47] infohfe: ok, let's go back to the map

[15:47] analystHFE: ok

[15:47] analystHFE: you have nine minutes left

[15:47] infohfe: can you select tehe workspace island?

[15:47] infohfe: let's look at responses there

[15:48] infohfe: ok, let's return to the map

[15:48] analystHFE: ok

[15:48] analystHFE: you can choose to go back to a map if you would liek

[15:49] infohfe: it seems that in australia, the main workspace related problem is also one of reserving, as opposed to the type as discussed earlier

[15:49] analystHFE: ok

[15:49] infohfe: can you highlight the first ring of the availability island

[15:49] infohfe: ok

[15:50] infohfe: ok, now i will take a few mins to rank them

[15:50] analystHFE: ok

[15:51] infohfe: can we return briefly to the global map

[15:52] infohfe: please unselect

[15:52] infohfe: can you highlight the first ring of the type area

[15:52] infohfe: island that is

[15:52] infohfe: ok, i think i am done

[15:53] analystHFE: ok

[15:53] analystHFE: thank you!

REFERENCES

Adrienko, N., & Adrienko, G. (2003) Informed spatial decisions through coordinated views, *Information Visualization, 2,* 270-285.

Au, P., Carey, M., Sewraz, S., Guo, Y., & Ruger, S. M. (2000) New Paradigms in Information Visualization, in *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* E. Yannakoudakis, N. Belkin, M. Leong, and P. Ingwersen (eds.), ACM Press, New York, 307-309

Badjio, E.F., & Poulet, F. (2005) User guidance – From theory to practice, the case of visual data mining, *Proceedings of the 17th IEEE Conference on Tools with Artificial Intelligence,* 709.

Beach, L.R. (1993), "Broadening the definition of decision making: the role of prechoice screening of options", *Psychological Science, 4(4),* 215-20.

Beard, D.V., & Walker, J.Q.(1990), "Navigational Techniques to Improve the Display of Large Two-Dimensional Spaces," *Behaviour and Information Technology, 9 (6),* 451–66.

Bederson, B. B., Shneiderman, B. (Eds.) (2003). *The Craft of Information Visualization: Readings and Reflections.* San Francisco: Morgan Kaufman.

Benford, S., Snowdon, D., Greenhalgh, C., ingrain, R., Knox, I. & Brown, C. (1995) VR-VIBE: A Virtual Environment for Co-operative Information Retrieval, *Proceedings of Eurographics '95*

Borgman, H.P. (1996) "Managers' Information Search Behavior Using Executive Information System," *Proceedings of the 29th Hawaii International Conference on System Sciences (HICSS) Decision Support and Knowledge-Based Systems, 2,* 99-109

Buzan, T. (1995). *The Mind Map book,* 2nd ed. London: BBC Books

Cañas, A., Leake, D.B., & Wilson, D.C. (1999) "Managing, mapping, and manipulating conceptual knowledge," in *Proceedings of the AAAI-99 Workshop on Exploring Synergies of Knowledge Management and Case-Based Reasoning,* AAAI Press: Menlo Park,10-14.

Card, S.K, Mackinlay, J.D., & Shneiderman, B. (1999) *Readings in Information Visualization – Using Vision to Think,* Morgan Kaufmann Publishers, San Francisco, CA

Carr, D. (1999) Guidelines for Designing Information Visualization Applications. *In Proceedings of ECUE'99, Ericsson Conference on Usability Engineering.* Stockholm, Sweden.

Cats-Baril, W.L. & Huber, G.P.(1987). 'Decision Support Systems for ill-structured problems: an empirical study'. *Decision Sciences, 18,* 350-372

Dansereau. D.F., (2005) Node-link Mapping Principles for Visualizing Knowledge and Information, *S.O Tergan & T. Keller (Eds) Knowledge and Information visualization, LNCS 3426,* Springer – Verlag Berlin Heidelberg

Foley, J., & Ribarsky, B. (1994) Next-generation Data Visualization Tools, in *Scientific Visualization, Advances and Challenges,* Academic Press.

Hansson, S.O. (1994) *Decision Theory: A Brief Introduction*, (Accessed June 2007) [Available at: //209.85.173.132/search?q=cache:PZHZJyqfh00J: www.infra.kth.se/~soh/decisiontheory.pdf+hansson+1994+decision+theory&hl =en&ct=clnk&cd=1&gl=us&client=firefox-a]

Hornbæk, K., Bederson, B.B. & Plaisant, C. (2002), "Navigation Patterns and Usability of Zoomable User Interfaces With and Without an Overview," *ACM Transactions on Computer-Human Interaction (TOCHI)*, 362–89.

Hornbæk, K. & Frøkjær, E. (2001), "Reading of Electronic Documents: The Usability of Linear, Fisheye, and Overview+Detail Interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM Press, 293–300.

Hutchins, E. (1995). *Cognition in the wild.* Cambridge, MA: The MIT Press.

Jacoby, J., Jaccard, J., Kuss, A., & Troutman, T. & Mazursky, D (1987). 'New Directions in Behavioral Process Reserach: Implications for Social Psychology'. *Journal of Experimental Social Psychology, 23*, 146- 175.

Johnson, P.E., Grazioli, S. jamal, K. and Zualkernan I.A. (1992). 'Success and Failure in Expert Reasoning'. *Organizational Behavior and Human Decision Processes, 53,* 173-203.

Keim, D.A. (2002) Information visualization and visual data mining, *IEEE transactions on visualization and computer graphics, 8(1)*, 1-7

Kosslyn, S. M. (1994), *Elements of Graph Design*. New York: W.H. Freeman.

K1einmuntz, B. (I990) Why we still use our heads instead of formulas: Toward an intcgralive approach. *Psychological Bulletin, 107*, 296-310.

Kleinmuntz. D.N. (1990) Decomposition and the control of error in decision analytic models. In R.M. Hogarth (Ed.). *Insights in decision making: A tribute to Hillel J. Einhorn*, Chicago: University of Chicago Press, 107-126.

Kleinmuntz, D. N. & Schkade , D.A (1993), Information Displays and Decision Processes, *Psychological Science, 4,* 221–27.

Kreuseler, M. & Schumann, H (2002) A flexible approach for visual data mining, *IEEE transactions on visualization and computer graphics, 8(1),* 39-51.

Kuo, F.E. (2002). Bridging the gap: How scientists can make a difference. In R.B. Bechteland A. Churchman (Eds). *Handbook Of Environmental Psychology,* NY: Wiley, 335 – 346.

Larkin, J.H (1989) Display-based problem solving. In D. Klahr, & K. Kotovsky (Eds), *Complex information processing. The impact of Herbert Simon.* Hillsdale, NJ:Lawrence Erlbaum Associates.

Lohse, Gerald L. (1997), "The Role of Working Memory on Graphical Information Processing," *Behaviour and Information* Technology, 16 (6), 297–308.

Lurie, N.H., & Mason, C.H.,(2007) Visual Representation: Implications for Decision Making, *Journal of Marketing, 71,* 160-177

Mintzberg, H., Raisinghani, D., & Théorêt, A. (1976), The Structure of 'Unstructured' Decision Processes, *Administrative Sciences Quarterly* 21:246-275.

Nisbett, R.E. & Ross, L. (1980), *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice Hall.*Less*. New York: Ecco.

Novak, J., & Wurst, M. (2005) Collaborative Knowledge Visualization for CrossCommunity Learning, *S.O Tergan & T. Keller (Eds) Knowledge and Information visualization, LNCS 3426,* Springer – Verlag Berlin Heidelberg

Oliveira, M.C.F. & Levkowitz, H. (2003) From visual data exploration to visual data mining; A survey, *IEEE Transactions on Visualization and Computer Graphics, 9(3),* 378- 394.

Raghubir, P., & Krishna, A. (1999) Vital Dimensions in Volume Perceptions: Can the Eye Fool the Stomach, *Journal of Marketing Research,* 36, 313–26.

Robertson, G. G. &. Mackinlay, J.D (1993), The Document Lens, in *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*, Scott Hudson, Randy Pausch, Brad Vander Zanden, and James Foley, eds. New York: ACM Press, 101–108.

Sarkar, M., & Brown, M.H. (1994), Graphical Fisheye Views,"*Communications of the ACM*, 37, 73–84.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies , 45,* 185-213.

Shneiderman, B., (1992) Tree visualization with tree-maps: 2-d space-filling approach, *ACM Transactions on Graphics (TOG), 11(1),* 92-99.

Schneiderman, B. (1996) The eyes have it: A task by Data Type Taxonomy for Information Visualizations, *Visual Languages,* 1996

Simons, D.J & Levins, D.T (1998) Failure to detect changes to people during a real world interaction, *Psychonomic Bulletin and Review, Vol 5,* 644-649.

Simon, H. A. (1971), "Designing Organizations for an Information-Rich World", *Computers, Communication, and the Public Interest,* The Johns Hopkins Press

Simon, H. A. (1996) *The Sciences of the Artificial* (3rd ed.), The MIT Press

Speier, C., & Morris, M. G. (2003). The influence of query interface design on decision-making performance. *MIS Quarterly, 27*(3), 397-423.

Stone. D.N.. & Schkitde. D.A. (1991). Numeric and linguistic information representation in multiattribute choice. *Organizationul Behavior and Human Decision Proce.sses. 48.* 42-59.

Stone, Eric R., J. Frank Yates, and Andrew M. Parker (1997), "Effects of Numerical and Graphical Displays on Professed Risk-Taking Behavior," *Journal of Experimental Psychology: Applied*, 3 (4), 243–56.

Tegarden, David P. (1999), "Business Information Visualization," Communications of the Association for Information Systems, 1 (4), (accessed June, 2007), [available at http://cais. isworld.org/articles/default.asp?vol=1&art=4].

Tergan S.O & Keller T.(2005) Visualizing Knowledge and Information: An introduction *(Eds) Knowledge and Information visualization, LNCS 3426*, Springer – Verlag Berlin Heidelberg

Tufte, E.R. (1990) Micro and Macro Readings Envisioning Information, Graphics Press, Cheshire, Connecticut, 37- 52.

Tufte, E.R. (1997) Visual Explanations: Images and Quantities, Evidence and Narrative, Graphics Press, Cheshire, Connecticut

Tufte, E.R (2006) Sparklines: Intense, Simple, Word-sized graphics, *Beautiful Evidence,* Graphics Press, Cheshire, Connecticut., 46-63.

Vessey, I. (1991), "Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature," *Decision* Sciences, 22, 219–40.

Viégas, F. B., Wattenberg, M., & Dave, K. (2004) "Studying Cooperation and Conflict between Authors with history flow Visualizations." *CHI 2004,* 24-29 [Available at: http:// web.media.mit.edu/fviegas/papers/ history_flow.pdf.]

Vogel, E.K, Woodman,G.F., & Luck S.J. (2001) Storage of features, conjunctions and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27(1),* 92-114.

Voigt, V. (2002) , Classification and Definition of Terms, *An Extended Scatterplot Matrix and Case Studies in Information Visualization,* Master's thesis, Hochschule Magdeburg-Stendal, (Accessed June 2007) [Available at: http://old.vrvis.at/via/resources/DARVoigt/masterthesis.html]

Wainer, H. & Velleman, P.F. (2001) STATISTICAL GRAPHICS: Mapping the Pathways of Science, *Annual Review of Psychology, 52,* 305-335

Wattenberg, M. (1999) Visualizing the stock market, *CHI '99 extended abstracts on Human factors in computing systems,* Pittsburgh, Pennsylvania

Wattenberg, M. (2005) Baby Names, Visualization, and Social Data Analysis, *Proceedings of the 2005 IEEE Symposium on Information Visualization*, 1.

Ware, C. (2004) *Information Visualization - Perception for Design*, Morgan Kaufmann, 2nd edition

Ware, C (2005) Visual Queries – the foundation of visual thinking, *S.O Tergan & T. Keller (Eds) Knowledge and Information visualization, LNCS 3426,* Springer – Verlag Berlin Heidelberg

Wong, P.C. (1999) Visual data mining, IEEE Computer Graphics and Applications, 19(5), 20-21.

Zhang, P., & Andrew B. W. (1995), "Business Information Visualization for Decision-Making Support: A Research Strategy," in Proceedings of the First Americas Conference on Information Systems, Manju K. Ahuja, Dennis F. Galletta, and Hugh J. Watson, eds. Pittsburgh: Association for Information Systems.

IMAGE REFERENCES

Bederson, B.B. (May 2000) *Fisheye Display of Linear Data,* Human Computer Interaction Lab, University of Maryland, [Online Image]. Retrieved from: http://www.cs.umd.edu/hcil/fisheyemenu/

Benford, S., Snowdon, D., Greenhalgh, C., ingrain, R., Knox, I. & Brown, C. (1995) *Multiple users in VR-VIBE,* [Online Image]. Retrieved July 2007, from: http://www.crg.cs.nott.ac.uk/research/technologies/visualisation/vrvibe/

CmapTools, Concept Map about Concept Maps, [Online Image]. Retrieved November 2008, from http://cmapskm.ihmc.us/servlet/SBReadResourceServlet?rid =1064009710027_1483270340_27090&partName=htmltext

FAS.research (2005) *Ars Electronica - Collaboration Structure of Cultural Projects* [Online Image] Retrieved January 2008, from http://www.fas research.com/gallery10.shtml

Robertson, G. G. Mackinlay, J. D. Card, S. K. (1991) Cone Trees: animated 3D visualizations of hierarchical information, *Proc. Human factors in computing systems conference,* 189-194.

Smith, P., Rawat, R. & Bigler, J. *Fire Spread Simulation*, Center for the Simulation of Accidental Fires, Explosions (C-SAFE), [Online Image]. Retrieved May 2007, from: http://www.sci.utah.edu/sci_gallery.php?id=325

Tufte, E.R. (2006) *Sparklines*, Edwardtufte.com, [Online Image]. Retrieved January 2008, from: http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001OR

Viégas, F. B., Wattenberg, M., & Dave, K. (2004) *Chocolate page*, Studying Cooperation and Conflict between Authors with history flow Visualizations." *CHI 2004*, 24-29. Retrieved June 2008 from http:// web.media.mit.edu/fviegas/papers/ history_flow.pdf.

Wattenberg, M. *NameVoyager*, Babynamewizard.com, [Online Image]. Retrieved

May 2007, from http://www.babynamewizard.com/voyager

Wattenberg, M. (1998) *Map of the Market,* Smartmoney.com, [Online Image], Retrieved January 2008, from http://www.smartmoney.com/map-of-themarket/