Exploratory Visual Analysis in Large High-Resolution Display Environments

BY

MHD KHAIRI REDA
B.S., University of Damascus, Syria, 2005
M.S., University of Illinois at Chicago, 2009

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Computer Science
in the Graduate College of the
University of Illinois at Chicago, 2014

Chicago, Illinois

Defense Committee:

Jason Leigh, Chair and Advisor
Andrew E. Johnson
Thomas Moher
Stellan Ohlsson, Department of Psychology
Michael E. Papka, Argonne National Laboratory

I dedicate this to my family:

Ouseima, Nader, and Eyad.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. Jason Leigh, who gave me inspiration, motivation, and guidance throughout my career as a graduate student. I could not have asked for a better mentor. I am deeply indebted to Dr. Andy Johnson, who was never too busy despite his extremely busy schedule. He was always available to provide feedback, offer advice, and share his remarkable experience. Many thanks go to Dr. Michael Papka, who has always believed in the potential of my research, and was instrumental in connecting me with collaborators.

I would also like to thank the rest of my dissertation committee for their insightful feedback, critique, and encouragement throughout this project. Dr. Tom Moher provided thought-provoking ideas and seasoned advice, which were instrumental in keeping this research on track. Dr. Stellan Ohlsson's seminar in Cognitive Modeling has also been a great inspiration.

Special thanks go to Lance Long, Alan Verlo, and Dr. Luc Renambot, for they were never too busy to respond to my questions, share their experience, and provide invaluable support during my projects. Without them, this work would not have been possible. Many thanks to Maxine Brown for her amazing help in editing research papers and grant proposals, and for her invaluable career advice. Thanks also to Dana Plepys for helping me navigate UIC's administrative maze of paperwork. Lastly, thanks to everyone at the Electronic Visualization Lab. While I may be graduating, I know I will always remain a member of the EVL family.

–KR

CONTRIBUTION OF AUTHORS

Chapter 1 introduces the central thesis in this research and presents the main research questions addressed by this dissertation. Some of these research questions were framed in a published extended abstract (Reda et al., 2014) for which I was the primary author and major driver of the research. Chapter 2 contributes the theoretical foundation for this research and surveys the state of the art. Chapter 3 represents a design framework for scaling up visualizations to large high-resolution display environments. Chapter 4 describes a case study exploring the use of large displays in tasks involving the visual analysis of ensemble data. This chapter is partly based on two published manuscripts for which I was the primary author and major driver of the research. Sections 4.1, 4.2, 4.5 and Figure 14 are based on material appearing in (Reda et al., 2012). Copyright © 2012 IEEE. Sections 4.4, 4.5, and 4.6 have appeared in (Reda et al., 2014). Copyright © 2014 MHD Khairi Reda. Victor Mateevitsi and Catherine Offord performed field experiments which ultimately generated the biological data used in the evaluation of the visualization environment. My research mentors, Dr. Andrew Johnson and Dr. Jason Leigh, contributed ideas and feedback during the development of the visualization environment discussed in the chapter. Chapter 5 represents an experimental study that investigates the effects of increasing the physical size and resolution of the visualization interface on user behavior and insight acquisition. Chapter 6 concludes the dissertation and summarizes its primary contributions.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

FIGURE **PAGE**

FIGURE **PAGE**

LIST OF FIGURES (Continued)

FIGURE **PAGE**

LIST OF ABBREVIATIONS

CAVE2	CAVE Automatic Virtual Environment 2
CMV	Coordinated Multiple Views
DPI	Dots Per Inch
GPS	Global Positioning System
HCI	Human-Computer Interaction
LCD	Liquid Crystal Display
LHD	Large High-resolution Display
LOI	Level Of Insight
RQ	Research Question
VTDP	Visual Thinking Design Pattern

SUMMARY

As society continues to generate escalating quantities of data, there is an increasing demand for perceptual and cognitive aids to help us make sense of the available troves of digital information. Visualization represents one of the most effective ways for exploring large datasets by leveraging our visual-perceptual capacities. Unfortunately, the scalability of visualizations has been limited by the prevailing display technology. Conventional desktop and laptop displays provide too few pixels to visualize today's datasets, forcing users to contextually switch between different views in order to see alternate projections of the information space. This limitation has been known to reduce user performance and potentially hinder exploration by inducing a 'tunnel vision' phenomenon where the analysis is focused on and limited to isolated subsets of the information space. However, thanks to advances in display technology, it is becoming increasingly feasible and affordable to surpass the limitations of conventional displays by building and acquiring Large High-resolution Displays. These displays are rapidly proliferating, providing researchers and scientists with more scalable platforms for the visual analysis and exploration of large and complex datasets.

This dissertation investigates the impact of adopting large high-resolution displays on user strategy and insight acquisition during exploratory visual analysis. First, we present a theoretical account of the cognitive costs involved in visual exploration, and highlight the implicit role of the visualization interface in modulating these costs. Second, we propose design patterns for constructing coordinated, multi-view-based visualizations for large high-resolution displays.

SUMMARY (Continued)

We then empirically investigate the effects of increasing the physical size and resolution of the visualization interface through observational and experimental studies. Our findings indicate improvement in discovery and insight acquisition, when users are provided with physically larger displays and more pixels. This effect manifests in a significant increase in the number of observations reported during visual exploration as well as the acquisition of higher-level, more integrative insights. These results suggest a role for large high-resolution displays in fostering discovery in data-intensive visual exploration scenarios.

CHAPTER 1

INTRODUCTION

> “ Nothing –not the careful logic of mathematics, not statistical models and theories, not the awesome arithmetic power of modern computers– nothing can substitute here for the flexibility of the informed human mind.
>
> *(Tukey and Wilk, 1966)* ”

Researchers and scientists in numerous disciplines are increasingly adopting a *data-driven* approach to scientific discovery. In this paradigm, the data is first generated from simulations or collected using scientific instruments (such as telescopes and high-throughput genome sequencing methods) before specific hypotheses are developed. Researchers then broadly *explore* this data cache to observe patterns of interest, identify promising leads, and formulate hypotheses and narratives in an attempt to extract knowledge from the data. While the ultimate goal is often to use the data to provide quantitative evidence in support of a theory or a model, the immediate goal of exploratory analysis is to look at the data from a flexible point of view, and see relationships that are beyond what existing models and theories would suggest (Tukey, 1977). By leveraging *human perception, insight, and intuition*, exploratory analysis favors a broad inquiry, encouraging one to ask plenty of meaningful questions and formulate multiple hypotheses, which could subsequently be quantitatively verified using confirmatory techniques (Tukey, 1980).

1.1 Exploratory visual analysis

Exploratory data analysis paradigms emphasize the use of visualization tools to empower the human mind to directly see and interact with data. Visualization is one of the most effective ways for communicating large quantities of information to our cognitive centers (Ware, 2012). Encoding data visually allows one to quickly recognize trends, patterns, and outliers (Card et al., 1999), which, when combined with one's domain knowledge, could lead to new insights. Interactive visualizations take this further by affording interactions that enable users to navigate large datasets and explore relationships between different components of the information space. As the user evolves his/her exploratory goals and mental model, the visualization environment transforms itself to present more relevant information (Heer and Shneiderman, 2012; Pirolli and Card, 2005), respond to user queries (Shneiderman, 1994), and re-organize the visual layout so that it reflects the user's mental model (Endert et al., 2012a; Andrews et al., 2010). Hence, the emerging field of *visual analytics* seeks to harness the power of interactive visualizations to facilitate human analytical reasoning (Thomas and Cook, 2005).

1.2 The data deluge

The increasing adoption of the data-driven discovery model presents many opportunities, but also poses new challenges created by the need to deal with increasingly larger, more complex, multi-dimensional, and multi-faceted datasets. This data deluge, wrought by the rise of the supercomputer and the proliferation of high-resolution, high-throughput data collection instruments, is limiting the opportunity for human-guided exploration and causing scientists to rely excessively on automated techniques (Johnson et al., 2007). Automated analysis techniques, while fulfilling an important role in data-intensive science workflows, are less suitable for exploratory analysis when scientists have little a priori knowledge about what is important or interesting in the data. Our limited capacity to explore 'big data' sources and find connections between disparate pieces of information may lead to dire consequences. For example, failure to foresee the 9/11 attacks is partially attributed to a failure in 'connecting the dots' between otherwise readily available pieces of intelligence (Kean, 2011). Arias-Hernandez et al. observe that such catastrophic failure in intelligence analysis can be explained in terms of "the human and technological inability to cope with the information overload produced by enormous amounts of constantly generated intelligence-related data" (Arias-Hernandez et al., 2012).

As the scale and complexity of data continue to escalate, these types of problems are likely to become more common in the future, threatening the viability of human-guided, exploratory analysis paradigms altogether. Therefore, new perceptual and cognitive aids are needed to ensure the ability of scientists, researchers, and data analysts to *see*, *explore*, and *make sense* of their data so that they can continue to make discoveries.

1.3 Visual scalability

Even with the best of designs, the scale and complexity of today's datasets can easily overwhelm visualization interfaces. Visualizations frequently suffer from visual clutter, which decreases recognition accuracy (Rosenholtz et al., 2007), increases the cognitive workload, and potentially interferes with the user's reasoning processes. To cope with larger datasets and reduce visual clutter, visualization designers employ *virtual navigation* techniques, such as panning and zooming, to allow users to navigate the information space and selectively show or hide different aspects of the data. However, since the majority of users still interact with visualizations through conventional desktop and laptop displays, most visualization interfaces are primarily designed to either show a small fraction of the data at an adequate level of detail or visualize a highly abstracted overview of an entire dataset at a single time. Either scenario makes the presented information far less useful for exploratory analysis, which is contingent on the ability to identify and study perceived patterns and outliers across time and space, from the micro to the macro. This technologically-imposed dichotomy between seeing detail without context or gaining an overview without detail could ultimately lead to a *tunnel vision* phenomenon (Schaffer et al., 1996), and bias users to focus their analyses on increasingly smaller information fragments that are examined in isolation. Furthermore, virtual navigation consumes precious cognitive cycles by forcing users to consciously perform operations that are not essential to the task at hand. It also subjects users to unnecessary optical flow (as experienced when panning a map to compare different regions, for instance), making it difficult for them to maintain a 'mental map' of the visualization (Purchase et al., 2007). Ultimately, these shortcomings serve only to increase the mental effort needed to operate a visualization and distract users from the analytical task at hand (Ball and North, 2005; Rønne Jakobsen and Hornbæk, 2011).

1.4 Large high-resolution displays

It is important to note that –at least from a perceptual point of view– the scalability of contemporary digital visualizations is constrained by limitations in the prevailing display technology, as opposed to sensory or perceptual limitations (Yost and North, 2006). Cockburn et al. observe that our "visual field spans approximately 200x120 degrees, whereas typical displays extend only 50 degrees horizontally and vertically, when viewed from normal distance." (Cockburn et al., 2008). The human retina is also capable of resolving approximately 500 dots-per-inch (DPI) in the fovea from a focal distance of 7cm, easily exceeding the resolution of most computer displays which are typically limited to 110 DPI (Woodson and Conover, 1964; Ware, 2012).

To overcome the limitations of conventional desktop and laptop screens, researchers and data scientists are increasingly adopting large high-resolution displays (LHDs) as a platform for the visual analysis of large-scale data (Leigh et al., 2012; Reda et al., 2013b). Constructed by tiling multiple LCD monitors to form contiguous display surfaces, these environments often span entire walls, making for high-resolution display surfaces on which a variety of data artifacts can be juxtaposed for analysis and correlation. In essence, LHDs "afford users the opportunity to trade *virtual navigation* for *physical navigation* (turning, leaning, moving around) thus allowing the user to exploit embodied human abilities such as spatial awareness, proprioception, and spatial memory" (Andrews et al., 2011).

Research has shown that LHDs are not mere 'pixel reservoirs', but rather serve to fundamentally alter the user's perception of the technology by consolidating the virtual information world they beget with our physical reality (Swaminathan and Sato, 1997; Tan, 2004). These displays can also improve user productivity and satisfaction, particularly when one is engaged in cognitively demanding tasks (Czerwinski et al., 2003). Prior studies have also identified numerous perceptual and productivity benefits associated with the use of LHDs, compared to conventional desktop and laptop-based interfaces (Tan et al., 2003; Ball et al., 2007; Yost et al., 2007; Shupp et al., 2009). However, much of these benefits pertain to basic perceptual tasks, such as target acquisition and map navigation. On the other hand, studies that have looked at the cognitive affordances of large displays have primarily focused on the analysis of text documents as a model task (Andrews et al., 2010; Endert et al., 2012b). In contrast, this dissertation investigates the potential role of LHDs in creating *digital information lenses* that foster the *visual exploration* of large and complex datasets. In particular, we focus on the cognitive processes involved in exploratory visual analysis in order to characterize and quantify the impact of increasing the physical size and resolution of visualization interface on insight acquisition in realistic, open-ended discovery scenarios.

1.5 Research questions

The central thesis in this research is that the adoption of LHDs as a visualization instrument would impact the *quantity and quality of insight* in data-driven, exploratory analysis scenarios.

The effects of being able to see and interact with more information at once could have important consequences for exploratory visual analysis. As information becomes instantaneously available on a LHD, visual search sets in as the primary mechanism for information foraging. Attending to different pieces of information becomes be as easy as moving ones eyes across the screen or turning one's head, a far less costly alternative to virtual navigation. This increased utility of visual search serves to reduce the cognitive cost associated with frequent tasks such as visual comparison and correlation (Plumlee and Ware, 2006). With appropriate designs, these changes in the cost structure of low-level visual operators can be channeled to encourage the breadth of the exploration, and ultimately improve the outcome of the process.

Our goal in this research is to understand the effects of having instantaneous visual access to orders of magnitude more information on user behavior in the context of exploratory data analysis. Furthermore, our desire is to incorporate this knowledge into a set of *design patterns* to inform the design of future visual analytic interfaces for emerging high-resolution displays. We shall distill this goal into the following research questions:

- RQ_1: What is the effect of increasing the size and resolution of the visualization interface on user behavior during exploratory visual analysis?
- RQ_2: Compared to conventional displays, how does the ability to simultaneously see and interact with orders of magnitude more information on LHDs affect insight acquisition?
- RQ_3: Are there new design patterns for scaling up multi-view-based visualization interfaces to LHD environments?

1.6 Organization

The rest of this dissertation is organized as follows: In chapter 2, we contribute a theoretical foundation for this research and survey the state of the art. In chapter 3, we present a design framework for scaling up existing visualization designs to LHD environments. Chapter 4 describes a real-world case study exploring the use of LHD environments in tasks involving the visual analysis of large ensemble datasets. In chapter 5, we describe an experimental study which investigates the effects of increasing the physical size and resolution of the visualization interface on user behavior and insight acquisition during visual exploration. We conclude in chapter 6, summarizing our contributions and outlining future research directions.

CHAPTER 2

FOUNDATION AND RELATED WORK

> "Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data… To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection.
>
> *(Cleveland, 1985)*"

Interactive visualizations constitute digital lenses through which we peer at large and complex information worlds. Yet, these lenses are delivering increasingly smaller and isolated fragments of information upon which we base our decision making. Unless we want to leave correct decision making to serendipity, we must understand the role of the visualization interface in shaping user behavior and decision making (Amar and Stasko, 2005). In this dissertation, we focus on understanding the effects of scaling up of the physical size and resolution of the visualization interface on user behavior and insight formation during exploratory visual analysis.

We begin our discussion with the theoretical foundations of distributed cognition, and present a conceptual model of the components and processes involved in thinking with the aid of interactive visualizations. We then turn our attention to exploratory visual analysis, highlighting the limitations of conventional visualization interfaces and their unintended consequences. Lastly, we motivating the need for more scalable visualization instruments and describe prior research on the use of large high-resolution displays in information visualization and visual analytics.

2.1 Distributed cognition

Humans, among many other organisms, have evolved the ability to process the signal conveyed by the visible light, and use it to perceive their surrounding environment, navigate landscapes, find food resources and mates, and solve spatial problems. Clearly, these capacities are not unique to humans. In fact the basic visual perceptual architecture is shared among all mammals, albeit with some important differences. However, what is unique among humans is the ability to apply these perceptual mechanisms to solve abstract problems that are inherently non-spatial. This ability to recognize and manipulate external objects in our visual scene as if they are artifacts of thought is highly important to our capacity to solve complex problems. One may be tempted at first to downplay the role of external representations in facilitating higher-order cognition. After all, abstract thought can occur in our heads with our eyes closed. However, it is generally undisputed that a "person working with the aid of thinking tools is much more cognitively powerful than that person alone with his or her thoughts" (Ware, 2012). The line between the human mind and the external world may, as a matter of functional interpretation, be more blurry than the "boundaries of skin and skull" (Clark and Chalmers, 1998).

Cognition can be thought of as an emergent phenomenon that results from the interaction between an individual and their surrounding environment, including external tools and artifacts as well as other individuals who participate in the cognitive activity (Hutchins, 1995a).

In this view, none of the agents or artifacts involved hold complete knowledge or control over the process. Rather, the cognitive state of the system and its processes are distributed across the different components. Accordingly, any cognitive output that system manages to produce should be attributed to the system as a whole, with the interaction between its various components serving as a key mediator of the cognitive function.

This theoretical framework, known as *distributed cognition*, is relevant to the study of virtually all information processing systems, both in prehistoric and modern contexts. For instance, ancient maritime navigation relied extensively on celestial objects, which served as a persistent form of memory and used to compute an appropriate heading to reach a particular destination. Today, modern GPS systems largely replaced the need to consult celestial objects, which cannot be reliably accessed at all times. However, it is important to note that GPS navigation relies on a complex network of satellite to provide accurate location service. Navigation in modern cars, vessels, and aircrafts relies on memory deposited in the GPS network, information processing provided by local GPS units, as well as the knowledge of crew members, all actively participating in the transformation of memory and sensory readings into a series of decisions and steering commands (Hutchins, 1995b).

In an era of 'big data' and abundant computing systems, problem-solving and decision-making almost always take place with the aid of digital, external representations of information, which increasingly take the form of interactive visualizations. Examples range from an epidemiologist trying to understand an ongoing epidemic situation by looking at a real-time map showing reported disease incidence, to a financial analyst attempting to forecast market trends by analyzing and comparing charts depicting historical stock prices. Interactive visualization systems provide scientists, analysts, and engineers with adaptable artifacts of thought that are not limited to the passive feeding of information to human decision-makers, but rather serve to scaffold their thinking processes (Pike et al., 2009), while providing essential computational processing in the background. Modern *visual analytic* tools thus form an integral part of our extended cognitive system (Hollan et al., 2000; Liu et al., 2008), providing us with high-resolution digital lenses while profoundly influencing our decision-making processes.

2.1.1 Thinking analytically with interactive visualizations

Visual analytic systems join cognizing human agents with highly-interactive information processing and communication tools. In this role, the visualization tool harnesses the computational power of modern computers to transform large amounts of data into visual representations. The human user, on the other side, contributes a "highly flexible pattern finder coupled with an adaptive decision-making mechanism" (Ware, 2012). This coupling of humans and computers provides a more scalable and effective cognitive function, compared to what each component can provide by itself. Visual analytic systems can thus be conceptualized at a high-level as an interaction between the human user and the visualization tool. However, it is helpful to unravel this black box so that we can begin to understand the flow of information and the distribution of processing within the system. Sedig et al. presents a conceptual model that breaks down a visual analytic system into five distinct components: the mental space, the interaction space, the representation space, the computing space, and the information space (Sedig et al., 2012). Figure 1 illustrates this structure.

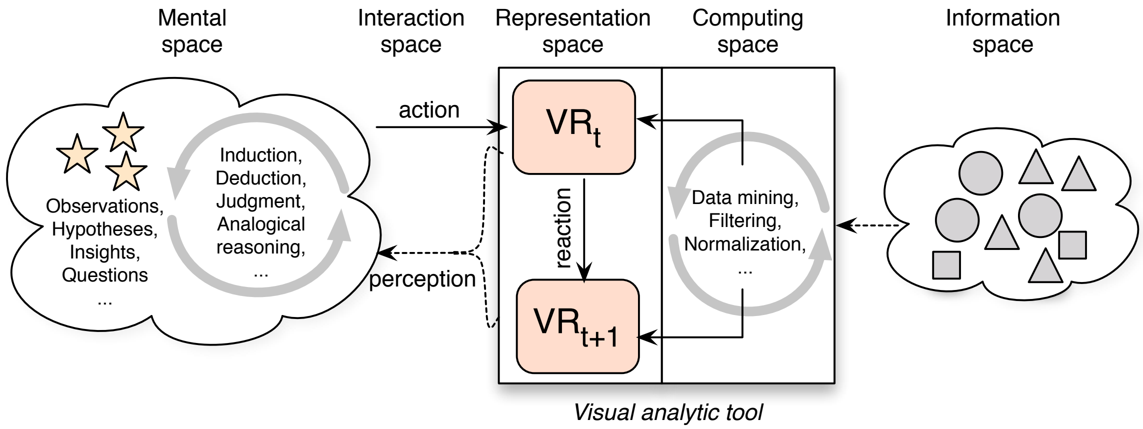


Figure 1. A visual analytic cognitive system results from the interaction of five distinct spaces. VR denotes Visual Representation. Adapted from (Sedig et al., 2012).

Information is usually construed as significant regularities in data, which in turn refers to discernible energy measurements within natural of simulated phenomena (e.g., climates, molecular systems, markets, web browsing histories, and health records) (Bates, 2005; Sedig et al., 2012). A visual analytic tool uses *computational* resources to transform, sample, normalize, and filter the data, mapping it to *visual representations* (denoted as VR). Encoding data visually enables users to utilize their visual perceptual system to search for and discover regularities in the data. Transforming such regularities into information and acquiring new knowledge, however, is a complex *mental* activity that involves the assimilation of newly acquired information, higher-order mental computation (e.g., inductive and analogical reasoning), and potentially conceptual change (Dunbar, 1993; Ohlsson, 2009).

2.1.2 Distribution of processing in visual analytic systems

Information processing in a visual analytic cognitive system can be distributed among the different components of Figure 1 (Parsons and Sedig, 2014), although, the optimal distribution would depend on the particular task the system is intended to support. For instance, in monitoring and situational awareness scenarios (as in operation control centers, for instance), a significant portion of information processing would typically fall on the computational component of the system, so as to reduce the cognitive workload on the human operator and allow for faster response. In learning applications, on the other hand, the learner usually benefits from being required to mentally process the presented information so that he/she can develop his/her mental model (Hullman et al., 2011; Parsons and Sedig, 2014). Some where in between these two ends, exploratory visualization interfaces could benefit from a hybrid approach, offloading tedious operations onto the computational space, while allowing the user to concentrate on developing insights, hypotheses, and beliefs. A related aspect is the *locus of control* over the state of the system, which has to be balanced between the human and the visual analytic tool.

Since our goal in this dissertation is to understand the effects of scaling up the visualization interface on user behavior in exploratory analysis scenarios, we shall limit our analysis to three spaces of the model in Figure 1: the mental space, the interaction space, and the representation space. In particular, we focus on the *interaction* space, as it represents the link between the human and the visual analytic tool, serving to coordinate the flow of information between the two subsystems. Furthermore, we shall limit our discussion to human-guided systems where control over the system and the exploratory process resides within the human agent, as this constitute the most common paradigm in visual analytics.

2.2 Exploratory visual analysis

Generally speaking, visual exploration is a process of broadly surveying the data from a flexible point of view in order to identify patterns of interests in the data, formulate meaningful questions and hypotheses, and construct plausible narratives to account for observations. In contrast to confirmatory analysis, which is typically aimed at obtaining an answer to a specific hypothesis or question (Tukey, 1980), the goal of visual exploration is to obtain a broad overview of the data, while allowing one to make unexpected discoveries (Krestel et al., 2011).

Visual exploration can be driven by the observations of patterns, outliers, and salient visual features in the visualization (Treisman, 1986), or it can be guided by prior hypotheses, intuition, and existing knowledge brought by the analyst (Privitera, 2006; Liu and Stasko, 2010). These bottom-up and top-down processes often interact in complex ways (Healey and Enns, 2012). This makes visual exploration a highly emergent activity, during which analysts constantly adjust their goals and frequently shift their focus to different parts of the information space, reframing existing beliefs to incorporate new findings (Klein et al., 2006). This fluidity is crucial to fostering a meaningful analysis during the early stages of the inquiry, as it enables analysts to evolve their mental models and formulate new theories and beliefs that make them see their data from new perspectives (Heuer, 1999).

The ultimate goal of visual exploration is to aid humans in developing insight. The notion of insight in the visual analytics community, however, is often used in a nebulous way to refer to a range of cognitive and neural states, from profound and spontaneous *'Eureka!'* moments to the gradual acquisition of "individual observations about the data" (Chang et al., 2009). In the context of exploratory analysis, the notion of insight could be expanded to include the formulation of new hypotheses and questions, as these constitute an important outcome of the scientific reasoning process (Klahr, 2002; Zimmerman, 2005).

2.2.1 Exploratory visual interfaces

One of the most useful paradigms for creating effective exploratory visual interfaces is to distribute information among multiple views. The main premise of this technique is that "users understand their data better if they interact with the presented information and view it through different representations." (Roberts, 2007). This paradigm, known as *Coordinated Multiple Views* (CMV), has been proven effective in a variety of applications and has become a standard technique in visual analytics. Figure 2 illustrate two example visualizations that employ CMVs to facilitate exploration.

CMVs allow designers to simplify visualizations by distributing information across multiple views, making the visualization as a whole easier to understand. However, the scalability of CMV-based visualizations has traditionally been limited by display technology, since adding an extra view requires the sacrifice of precious screen estate (Wang Baldonado et al., 2000). Consequently, visualizations usually incorporate a limited number of views (typically 2–6). When the amount of information is too large to fit, views are presented in sequence using *virtual navigation*, such as panning and zooming, *overview + detail*, and window switching (Cockburn et al., 2008). Virtual navigation thus involves the *temporal* separation of views, al-

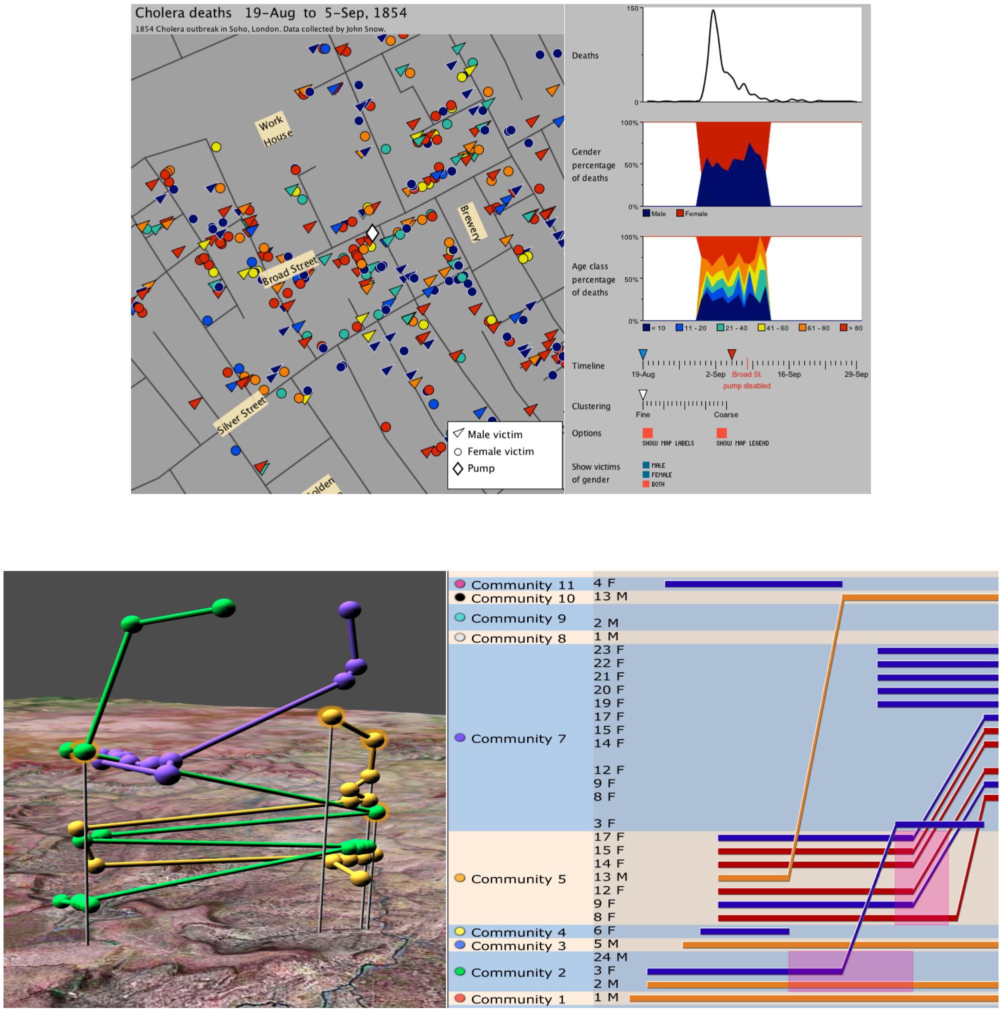


Figure 2. Two example visualization that illustrate the use of *coordinate multiple views* to facilitate visual exploration. The top picture illustrates a recreation of the landmark visualization of the 1854 Cholera outbreak in London by John Snow, augmented with additional timeline charts on right side to show statistical information about the number victims as well as their (fictitious) age and gender. The bottom pictures illustrates a visualization of the social network in a zebra population, with the left view depicting the movement of zebra communities in space and time, and the right part depicting the evolution of their social structure (Reda et al., 2011).

lowing the user to sequentially move between different projections of the data. Conversely, *spatial* separation entails the rendering of multiple views side-by-side, enabling a non-sequential form of access to information. Furthermore, spatial separation typically allows for *coordination* between views, which is typically achieved via brushing-and-linking (Becker and Cleveland, 1987).

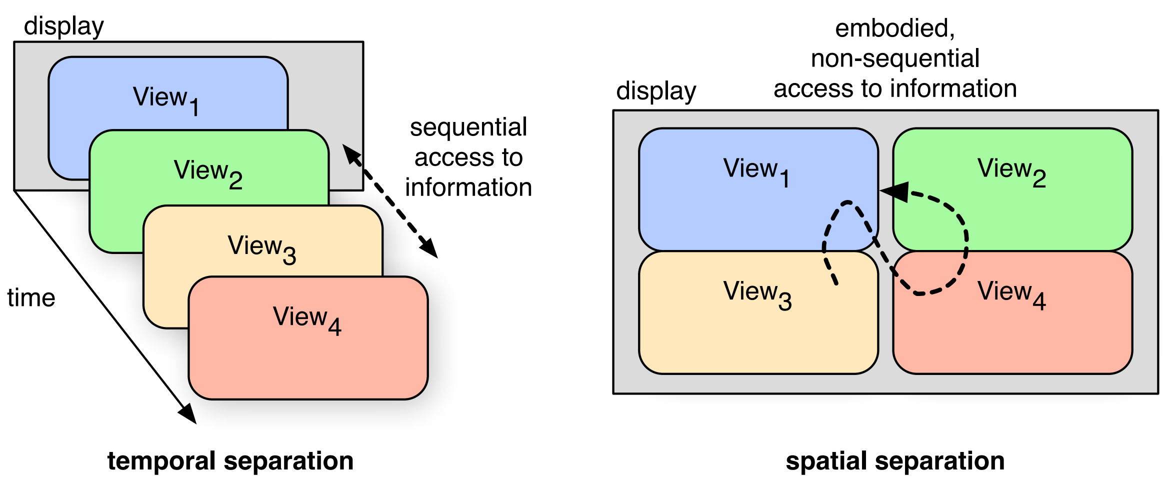


Figure 3. Temporal separation of views compels users to sequentially switch between views, causing them to rely excessively on their working memory to retain and integrate information across views. In contrast, spatially separated views make the relevant information simultaneously available, affording non-sequential access to information and thus reducing the cost associated with integrative operations, such as comparison and correlation.

Naturally, conventional desktop and laptop displays skew visualizations towards the *temporal separation* of views, which tends to increase the cognitive workload on users as they resort to retaining increasingly complex features and patterns in their limited visual working memory (Plumlee and Ware, 2006). On the other hand, the cost of switching between simultaneously visible views appears to be insignificant (Convertino et al., 2003). This favors the spatial separation of views in visual exploration so as to reduce the cognitive costs associated with low-level analytic operations, including comparison, correlation, and identification of outliers. Figure 3 contrasts these two alternatives.

At the macro-level, the effort required to navigate, examine, and integrate information across a large number of temporally-separated views could become too prohibitive, particularly in scenarios involving the exploration of large-scale datasets. These elevated cognitive costs –which are subconsciously perceived by users– may act as 'soft constraints' (Gray and Fu, 2004), prompting users to reduce the frequency of view switching and virtual navigation. Such accommodating behavior may save precious time and reduce the cognitive workload. Unfortunately, it may also contribute to a 'tunnel vision' phenomenon by inducing users to narrow in onto increasingly isolated parts of the information space. Furthermore, the ability of users to spontaneously recognize unexpected relationships across temporally separated views is likely to diminish, causing them to inadvertently miss important connections.

The above two factors can be understood in terms of inhibitory costs that impede exploration (i.e., top-down costs) as well as missed opportunity costs for making unexpected inferences (i.e., bottom-up costs). Their side effects have been observed, to some extent, in studies that address how intelligence analysts work with large-scale corpora of text documents (e.g., (Patterson et al., 2001)). However, these unintended consequences are likely to be more pronounced in exploratory visual analysis scenarios, where the exploratory goals are less clear to begin with, and the constraints on visual working memory are sharper. We discuss both of these facets in the following sections, and then consider how *large high-resolution displays* can provide technological intervention to mitigate these unintended consequence, potentially improving the outcome of the exploratory process.

2.2.2 Missed opportunities: bottom-up costs in visual exploration

The main bottleneck in visual exploration is our visual working memory, which has a limited capacity of approximately 3–5 items (Vogel et al., 2001). With conventional visualizations, one needs to frequently switch between a large number of projections in order to compare and correlate features during visual exploration. Visual context-switching is not only time consuming but often requires the explicit manipulation of a graphical user interface, potentially causing interference with existing stored patterns by 'flushing' the content of the visual working memory (Maxcey-Richard and Hollingworth, 2013). This reduces the chance of retaining potentially important visual features that are essential to the inferential task.

It is also informative to contrast this form of visual context-switching experienced by visualization users as they are attempting to move through a large information space, by looking at one projection at a time, with more familiar activities, such as solving a jigsaw puzzle. In the latter, one tends to move individual pieces and position them side-by-side, in different arrangements, until the connections between them is clear. Juxtaposition and rearrangement of pieces is thus essential to solving the puzzle in a reasonable amount of time. The ability to utilize space to rearrange "physical tokens of statement" allows the player to "change the cost structure of the *inferential landscape*", and is thus essential to his/her ability to logically move downstream towards a more complete solution (Kirsh, 2013). After all, solving the same jigsaw puzzle looking at once piece at a time is a daunting proposition. Similarly, the sequential switching between view (representing different data projections) during exploratory visual analysis could ultimately reduce the probability of making inferences about how these temporally-separated views relate to each other. For example, a user looking at geospatial visualization may fail to recognize similar clusters of chemical contamination sites across a large map when such clusters occur in disparate locations that cannot be viewed simultaneously.

2.2.3 Top-down costs in visual exploration

From a top-down perspective, interacting with a visualization incurs significant cognitive costs that could impeded exploration. While such costs are acknowledged by the visualization community, they are often understood as factors that hinder the perception of glyphs in visual representations or decrease user performance on specific tasks. Here, we argue that the costs associated with the use of visualizations should also be thought of as important factors that could influence the formation of exploratory goals, the selection of strategy, and ultimately the outcome of the visual exploration process. These costs are determined in part by the visual representations employed in a visualization and the set of interaction techniques incorporated into the visualization interface. However, additional factors, such as the physical size and resolution of the visualization interface, implicitly affect these costs by modulating the utility of the visualization with respect to the analytic operators available to users.

2.2.3.1 Seven stages of action

To understand potential interface-related factors affecting user strategy and goals in visual exploration, we need to properly distinguish between the different costs associated with the use of interactive visualizations. For that we turn to one of the longest standing usability models in the human-computer interaction community, the *seven stages of action*, developed by Donald Norman in his seminal work *The Design of Everyday Things* (Norman, 2002). Norman postulates a series of stages a user goes through when interacting with a piece of technology in order to accomplish a specific goal. These seven stages are in the following order:

1. Forming the goal
2. Forming the intention
3. Specifying an action
4. Executing the action
5. Perceiving the state of the world
6. Interpreting the state of the world
7. Evaluating the outcome

The first step in interacting with the world (or with a piece of technology) is to formulate a goal and specify a state to be achieved. After the goal is formulated, it must be translated into an intention–a desire to perform some action. This action must be unpacked and mapped onto the set of available operators. The action is then executed. The next stage is to evaluate whether the performed action satisfied the original goal, which begins with perceiving the new state of the world. The perceived state must be interpreted. Lastly, the user compare the interpreted state and determined whether the goal has been met.

In addition to the above seven stages, Norman proposes two types of gulfs, which can be construed as costs or difficulties one has to overcome. The *gulf of execution* represent difference between the user's intention and the set of operators available to the user. Similarly, the *gulf of evaluation* reflects the effort needed to interpret the state of the world to determine whether the goals have been met.

2.2.3.2 Interaction costs in visualizations

At its heart, the *seven stages of action* is a framework that abstracts the intricacies of human interaction with their world, with the ultimate goal of improving the design of everyday technologies from door knobs to computer interfaces. While the framework is intended to ease the complexity of technologies that serve a well-defined goal, the framework is also adaptable to human-guided knowledge discovery systems (e.g., interactive visualizations) where the goals are more fuzzy. Lam's framework of *interaction costs with information visualization* represents one such adaption (Lam, 2008). Lam's framework is similar in spirit to the seven stages of actions, with a minor difference in that the costs implied by Norman's two gulfs are more explicitly articulated. Lam describes the following types of costs associated with the use of interactive visualizations, which are also illustrated in Figure 4:

1. *Decision costs to form goals:* When a visualization depicts a large dataset, users need to decide on which subsets of the data to explore and what questions to pursue.

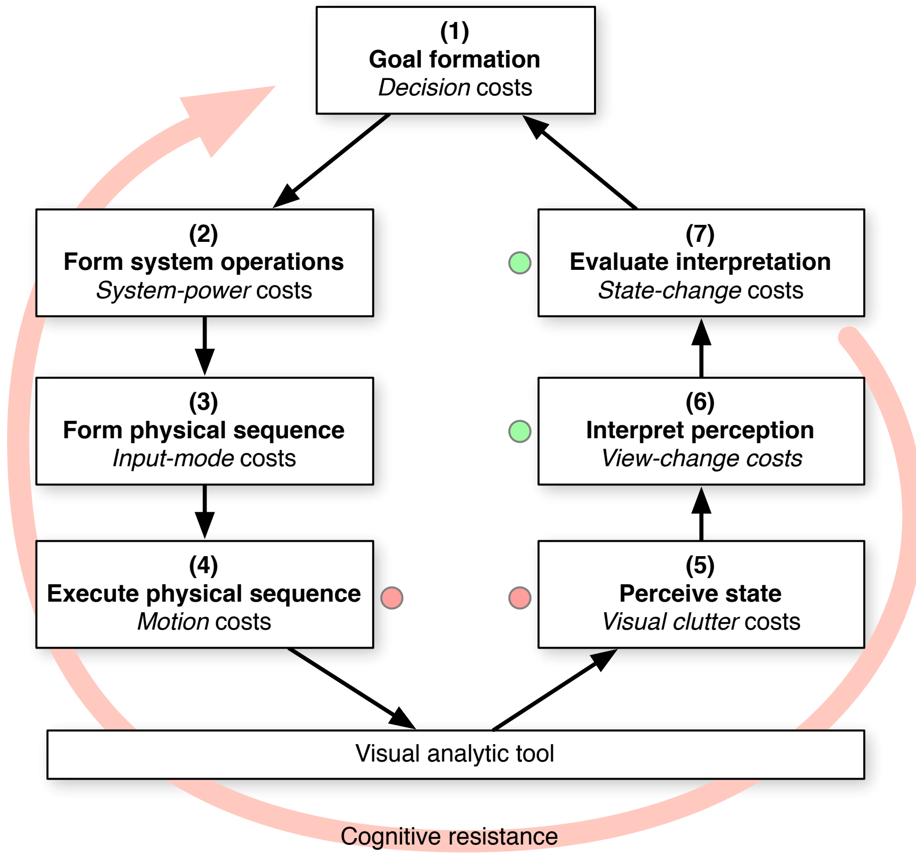


Figure 4. The sequence of steps a user takes when interacting with a visualization (in **bold**) along with the associated interaction costs (in *italics*). Adapted from (Lam, 2008). In addition to the costs, we postulate a 'cognitive resistance' feedback force, which could dissuade users from pursuing integrative questions and prevent them from forming new or lateral exploratory goals. The circles represent the hypothesized effect of adding additional views to the interface to reduce temporal-separation, with red circles indicating a hypothesized increase in the costs and green circles indicating a decrease.

2. *System-power costs to form system operations:* Once the goal is formulated, it has to be translated into operations that can be performed using the provided set of interaction techniques. For example, the user may choose to navigate a map using pan and zoom opera-

tions to focus on a particular region, starting from an existing view instead of spawning a new one.

3. *Input mode costs to form physical sequences*: The operation must then be translated into a series of input commands. For instance, a series of drag-and-drop and mouse wheel scroll commands could be used to perform a navigation operation. A possible confounding factor here is the need to choose between multiple input modes when the controls are overloaded.

4. *Physical-motion costs to execute sequences:* The input must be executed using motor actions (e.g., moving the mouse and turning one's eyes and head). This physical effort could become strenuous if repeated over an extended period of time.

5. *Visual-cluttering costs to perceive state:* Interactions may increase visual clutter (due to an increase in the number of data points being rendered, for instance) or introduce occlusion (e.g. overlapping views), impeding perception.

6. *View-change costs to interpret perception:* Interactions can result in significant changes to the visualization state, which must be interpreted and related to the previous state.

7. *State-change costs to evaluate interpretation:* To gain insight and update one's mental model, the user need to perform mental computation to cognitively integrate information extracted form multiple views and/or analysis states so that one can begin to understand the effects of their inquiry.

2.2.3.3 Soft constraints on visual exploration

The costs articulated in Lam's framework are usually regarded as a series of processes or barriers users need to overcome to accomplish a particular exploratory goal. However, the cost framework also implies 'cognitive resistance' (the red feedback arrow in Figure 4), which manifests as soft constraints that prevent users from forming new or lateral exploratory goals. Users, as cognitive agents, tend to minimize the cost of acquiring information, just as animals attempt to minimize the cost of foraging for food in the wild (Pirolli, 2007). Furthermore, users are likely to alter their exploratory strategy based on the cost of operators available (O'Hara and Payne, 1998). For instance, visual comparison, a commonly used operator in visual analytics (Amar et al., 2005), is likely to incur a relatively high cost when users are forced to compare temporally-separated objects (Plumlee and Ware, 2006). We would therefore expect users to perform less visual comparisons even though this operation might be integral to forming insight. This bias to the *path of least resistance* is particularly relevant in exploratory scenarios where the goals are vaguely defined and the *information scent* is rather weak (Pirolli et al., 2003). In such cases, users may unknowingly steer their exploratory strategy to exploit isolated subsets of data, in their attempt to minimize the overhead of visual context-switching induced by the chronic temporal-separation of views in conventional interfaces.

Adding additional views to the visualization to reduce the frequency of view-switching could affect interaction costs in a number of important ways (changes are indicted by green and red circles in Figure 4). More specifically, we would expect a decrease in *view-change* costs (step 6) as their would be less frequent disruption to the visual layout wrought by virtual navigation (Ball and North, 2005), and users would thus have to rebuild their 'mental map' less frequently. Additionally, we would expect a reduction in the costs of integrating information found in multiple views (step 7) as users would be able to 'physically' move between the different views with embodied actions (Andrews et al., 2010), integrating information in a non-sequential fashion. On the other hand, the addition of views also results in some overhead, which manifests in rising *physical motion* costs (step 4) as a result of having to execute more strenuous physical actions (e.g., head turns, eye movements, and walking up to the display) to look at the views. Additionally, we would expect an increase in *visual clutter* costs (step 5) induced by having to 'read' a larger number of visual glyphs, and from being required to manage one's attention in a larger spatial environment.

2.2.4 Designing around the limits of working memory

The most fundamental tasks in visual exploration revolves around the integration of meaningful visual patterns encountered during the analytic activity. Yet, our ability to integrate information from multiple sources will always be bound by innate cognitive limits. Chief among those is the capacity of our working memory (both visual and verbal). When designing visual analytic interfaces, it is therefore essential to minimize the burden on working memory. As Ware observes, designs that "work around the limits of working memory capacity can, in many cases, result in impressive gains in efficiency" (Ware, 2012).

There is an inherent trade-off between the temporal separation of views and the increase in interpretation and integration costs on one hand, and the spatial separation of views and the increase in the physical and perceptual/attentional effort required to 'read' a more complex visualization. Therefore, a key design issue in visualization is to balance this trade-off. The question is, then, do current visualization interfaces effectively balance these costs? Or do they tip the balance towards either sides of the equation? Naturally, the prevailing display technology favors the temporal separation of views which could potentially tip the balance, thus increasing the cost of visual exploration. Fortunately, however, we are witnessing a disruptive trend in display technology marked by rapid increase in the resolution of LCD panels accompanied by a sharp drop in their cost. It is thus becoming increasingly affordable for visualization users to acquire and/or build larger displays with more pixels from commodity LCD panels. *Large high-resolution displays* are opening the possibility for a new set of visualization designs with a focus on relieving the burden on human cognitive resources, as opposed to conserving pixels.

2.3 Large high-resolution displays

There is a large body of research that demonstrate advantages to using LHDs in knowledge-based activities, compared to conventional displays (Czerwinski et al., 2006; Ni et al., 2006). These benefits are derived from a range of perceptual and cognitive affordances that LHD environments appear to provide or enhance. We can group these affordances under two categories: *visual scalability* and the provisioning of *space to support sense making*.

2.3.1 Visual scalability

Visual scalability is typically defined as the "capability of visualization tools effectively to display large data sets, in terms of either the number or the dimension of individual data elements" (Eick and Karr, 2002). Eick and Karr's definition, however, deliberately avoids referring to human performance, claiming that these aspects cannot be adequately quantified. In some respect, this observation is correct, as it is difficult to quantify the value and type of insight one can gain from looking at and interacting with visualization (North, 2006; Chang et al., 2009). It is quite possible, however, to measure low-level aspects of human performance, such as the time to complete a particular task and the accuracy of the outcome, so that we can begin to understand the potential for visualizations to scale to larger datasets.

There are various ways to increasing the scalability of visualizations. One may use more scalable visual metaphors and data aggregation methods (Shneiderman, 2008), utilize interaction and multi-scale navigation techniques to selectively show and hide information on demand (Shneiderman, 1996; Keim and Schneidewind, 2005), break the data into multiple composite views (Javed and Elmqvist, 2012), or use larger displays with more pixels so as to show more data points (Reda et al., 2013b). Of course, one may also combine these approaches.

As the technology behind displays improve and their cost continues to decline, Large High-Resolution displays (LHDs) are becoming an increasingly attractive option. Numerous studies have explored the relationship between display size, resolution and human performance with basic visual analysis tasks, including visual search, map navigation, and pattern perception. Ball et al. explored the tradeoff between virtual navigation and physical navigation afforded by LHDs. Their results demonstrate that users take significantly less time to perform map navigation and visual search tasks, as the size and resolution of the display increases (Ball et al., 2007). Furthermore, users preferred physical navigation (by walking up to and in front of the display to access information) over virtual virtual navigation (by panning the map, in this study).

These results suggest that virtual navigation has a detrimental effect on performance time in basic visualization tasks. Furthermore, they suggest that one is able to simply add more displays (possibly up to a certain extent) to improve the scalability of visualizations. However, the precise reason for this improvement could not be identified. In a followup study, Ball and North attempted to identify the reason for improved performance in LHD environments by isolating two factors: the field of view, and the ability of users to acquire more information using physical navigating (Ball and North, 2008). Their results indicated that the opportunity for acquiring more information by physically navigating the display was more crucial to improving performance, compared to the wide field of view afforded by LHDs. This study suggests the need for both an adequate size and resolution in displays, so that users can physically navigate the information space and acquire information with embodied actions as opposed to virtual navigation. In that respect, *context + focus* instruments (see Figure 5), which are designed to provide a high-resolution focal area embedded in a large but low-resolution display (Baudisch et al., 2002), may be less useful in information visualization scenarios.

The question of what is the optimal display size and resolution remains unanswered (Simmons, 2001). Ware has argued that the optimal display resolution is 4,000 x 4,000, as this number provides a correspondence between display pixels and 'brain pixels', with the latter being determined by the density of photoreceptors in the retina (Ware, 2000). However, this argument does not consider the effects of physical navigation. Yost et al. conducted a pair of studies

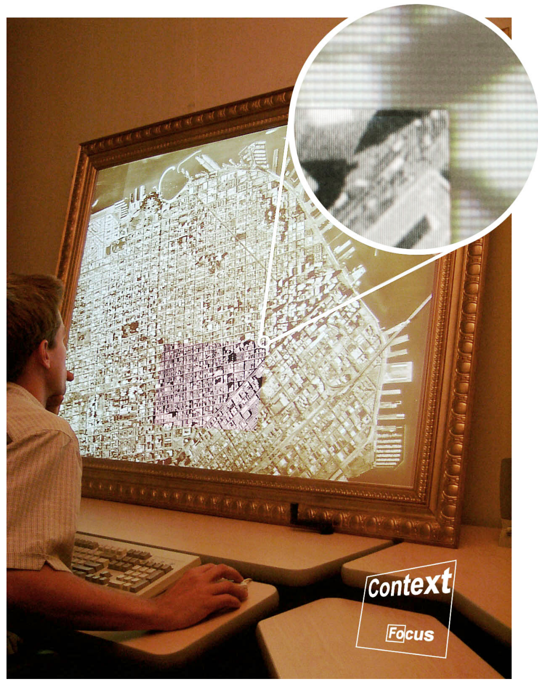


Figure 5. A *context + focus* instrument provides a high-resolution focal area that is embedded in a large but low-resolution display. Such displays provide a less costly alternative to large high-resolution displays, but could negatively impact performance in situations that require frequent, high-fidelity access to data in the peripheral area, as this would essentially compel users to perform repetitive virtual navigation. Picture copyright by Patrick Baudisch. Used in accordance with the Creative Commons Attribution-ShareAlike 3.0 license.

to determine the effects of increasing the display's resolution beyond the visual acuity of the human eye, by using a large enough display that require users to walk up to the display in order to see the full detail. (Yost and North, 2006; Yost et al., 2007). They found out that –again for basic visualization tasks– the time needed to complete the task increases at a slower rate compared to the increase in the size of the dataset, as long as the resolution of the display is increased proportionally. They also observed continued performance gains, even when the display's resolution exceeded the visual acuity of the human eye. Similar studies by Shupp et al. demonstrated additional benefits to curving a LHD around the user (Shupp et al., 2009).

2.3.2 Space to support sense making

One of the main advantages of LHD environments is their ability to engage our spatial cognition and proprioceptive capacities (Tan, 2004; Andrews et al., 2011). Humans cognition is a fundamentally embodied phenomena that emerges from our interaction with the world. One of the main resources we have evolved to leverage is space. Space provides us with a flexible memory system that we can use to externalize our internal limited memory. Kirsh classifies the myriad ways in which humans intelligently use space under three main categories: using space to simplify choice by highlight the set of possible actions one can perform on external objects, using space to simplify perception primarily by means of clustering, and using space to save expensive internal computations by offloading these computations onto spatial operations (Kirsh, 1995). One of the fundamental features that make space such an efficient cognitive resource is the ease with which we can access and manipulate external representations. We can easily reach out to our immediate surroundings, access physical objects using eye movements and head turns, and manipulate these external objects with our fine motor abilities. All of these operations can be performed efficiently with little thought and effort, enabling us to adapt and structure our spatial environments to help us accomplish complex tasks. In fact, as Kirsh and Maglio point out, experts rely extensively on space, frequently recruiting spatial resources throughout their mental activities. For instance, observations of Tetris players demonstrated that expert players tend to rotate falling pieces more frequently compared to novices (Kirsh and Maglio, 1994). These non-pragmatic spatial operations saved expert players from exerting mental effort to rotate the pieces, relying instead of space as a computational resource, ultimately improving the overall performance in the game.

Andrews et al. observed how both professional and novice intelligence analysts were able to leverage the space and resolution afforded by a LHD in a scenario that involved the analysis of large collections of text documents (Andrews et al., 2010). The study revealed qualitative differences in strategy between participants who performed the task on a traditional desktop display and those who utilized a LHD to conduct their analyses. Participants in the latter group frequently opened a large number of documents and used the extra screen space, enabling them to efficiently switch between documents with simple embodied actions (e.g., eye movements and head turns). Andrews et al. also observed how participants spontaneously leveraged the expansive display surface to organize information spatially into clusters. For instance, related documents were clustered on different parts of the display, with some participants creating more subtle arrangements (e.g., timelines reflecting chronological ordering of events in the documents). Theses LHD-based arrangements can be considered *schemas*– high-level representations that summarize the relevant information combine them with the analyst's mental model, so as to tie them together into a coherent narrative (Pirolli and Card, 2005). Andrews et al. postulate that the main factor in enabling this form of space-based sense making lies in the ability to access display pixels physically using embodied interactions, as opposed to using virtual navigation techniques. In a follow up study, users who were given the same amount of *virtual space* while performing the same task on a conventional desktop display created fewer structures compared to their counterparts who utilized a LHD (Andrews and North, 2013).

The above studies demonstrate the benefits of having an abundance of digital screen estate, highlighting the fact that people are able to think of LHDs as inherently spatial environments that afford similar cognitive resources compared to real-world physical environments (e.g. war rooms (Teasley et al., 2002)). Furthermore, analysis of user-created clusters revealed that these clusters are organized semantically, indicating that users were able to project their mental model onto the display. (Endert et al., 2012b). Building on these affordances, Endert et al. introduced *semantic interaction*, a technique that leverages user-created spatial clusters to steer and progressively adapt the underlying data model to the analyst's mental model throughout the visual analysis process (Endert et al., 2012a). By recording and analyzing the user's interaction with documents on a LHD, Endert et al. were able to infer some of the cognitive operations users implied and use this knowledge to steer the computational processing of data, effectively unifying information foraging and synthesis (Endert et al., 2011). Combining these concepts into an integrated tool, Fiaux et al. describe a visual analytics environment for exploring large document corpora (Fiaux et al., 2013). The tool creates spatially-arranged document clusters which are chained based on semantic similarity, allowing users to quickly jump between documents based on topics.

2.4 Summary

The above line of research provides powerful examples and presents a compelling treatment of how one could leverage the space and resolution afforded by LHD environments in complex analytical activities involving large quantities of digital information. However, the above studies focused primarily on the analysis of *text documents* in scenarios that emphasized the *synthesis* of information into one cohesive narrative, so as to model the task of intelligence analyst. This research, on the other hand, focuses on *exploratory visual analysis* where the goal is generally aimed at attaining breadth and diversity. We are particularly interested in the effects of increasing the visual fidelity of visualizations on the articulation of hypotheses and insights. Although it is reasonable to expect favorable results in exploratory visual analysis when using larger displays with more pixels, the perceptual and cognitive mechanisms employed as well as corresponding user strategies are likely to be very distinct (Larkin and Simon, 1987), compared to text analytic scenarios. Consequently, the appropriate *design principles and guidelines* for LHD-based visualization interfaces are likely to be distinctively unique. Such design principles, however, are largely absent from the literature (Andrews et al., 2011). This lack of knowledge has often led to designs that employ a 'giant desktop' metaphor, despite the numerous noted usability issues (Czerwinski et al., 2003; Hutchings et al., 2004). Indeed, there is evidence from prior studies to suggest that existing visualization and interaction techniques will not simply scale up, when the size and resolution of the display is increased (Swaminathan and Sato, 1997; Rønne Jakobsen and Hornbæk, 2011; Jakobsen and Hornbæk, 2013). While some of these studies also provide cautionary tales on the danger of flooding users with too much information, we argue that it is possible to overcome some of these challenges by providing appropriate *cueing* mechanisms that help users manage their attention and enable them to focus on the relevant information without loosing context

We believe that the *coordinated multiple views* (CMV) paradigm provides a good starting point to scale up visualization interfaces from the desktop world to LHD environments. However, we need to re-examine some of the assumptions in this model and adapt it to this new platform. In the next chapter, we reconsider the design space of CMV-based visualizations and present visualization design patterns aimed at taking advantage of the expansive screen and exquisite resolution afforded by LHDs while reducing the potential of inducing *information overload*.

CHAPTER 3

DESIGN PATTERNS FOR LARGE HIGH-RESOLUTION DISPLAYS

Modern visualization practice is predicated on decades-old infrastructure. Small, low-resolution computer monitors continue to be the principle technological components on which we base our visualization design principles. However, in the last several years, we have seen great advances in visualization infrastructure marked by the proliferation of new technologies, including novel sensing technologies (e.g., Microsoft Kinect), multi-touch devices, and large high-resolution display (LHD) environments. Unfortunately, current generation of visualization tools have not been designed to take advantage of such new infrastructure. In fact, many existing visualization and interaction models were specifically conceived to work around technological limitations. For instance, visualizations rely extensively on virtual navigation – as epitomized by the *overview first, zoom and filter, then details-on-demand* mantra (Shneiderman, 1996)– to compensate for limitations in display resolution. Virtual navigation, however, can be detrimental to user performance (Czerwinski et al., 2003; Ball and North, 2005). One alternative is to use *coordinated multiple views* (CMVs), which are commonly employed, but often not leveraged to their full potential as such interfaces tend to quickly use up the limited screen estate available in conventional displays.

This chapter examines some of the commonly used visualization design patterns, and analyzes their expected performance and adaptability to LHDs. Our goal is to improve the scalability of these designs, by leveraging the expanded screen estate and resolution afforded by

LHDs, while providing additional perceptual cues to reduce the potential of inducing *information overload*. We focus on the task of comparative visual analysis, as it represents a commonly recurring theme in visual analytics. We base our discussion on Ware's *Visual Thinking Design Patterns* (VTDP), which represents a collection of best-in-class designs that have been identified by the visualization community after years of research (Ware et al., 2013).

3.1 Scope

The design space of interactive visualizations is enormous. Covering the entire design space would not be feasible in a single dissertation. Therefore, this dissertation will focus on one task subset, namely *comparative visual analysis*.

3.1.1 Comparative visual analysis

Comparison is one of the primary elemental operations that can benefit from visual representations (Amar et al., 2005; Kehrer et al., 2013). Many complex analytical tasks are built around comparison, including the analysis of ensemble, spatio-temporal, and genomic data. Moreover, visual comparison represents a cognitively demanding task that entails the recognition and conceptual reconstruction of relationships that implicitly exist among a large number of data items. As the size and complexity of datasets grow, the number of possible patterns and relationships grows at an even faster rate. While data mining and machine learning techniques can be used to automatically classify large collections of objects into separate groups, users often find it difficult to interpret and understand the visual and conceptual basis for such classifications.

Numerous visualization techniques have been designed to facilitate the comparison of specific types of information, including graphs (Alper et al., 2013), rankings (Behrisch et al., 2013), and genomic sequences (Meyer et al., 2009). However, in this research we shall focus on generic visualization designs that can be used to support comparison in a data-type agnostic manner. According to Gleicher et al., there are three primary methods for supporting comparisons in visual representations: *juxtaposition*, *superimposition*, and *explicit encoding of differences* (Gleicher et al., 2011). We illustrate these three basic approaches in Figure 6.

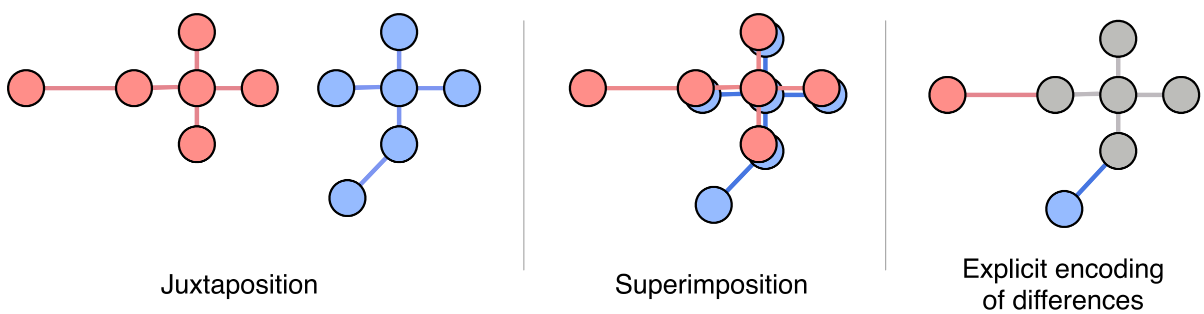


Figure 6. Two node-link diagrams can be compared using one of three primary visual comparison methods:*juxtaposition*, *superimposition*, and *explicit encoding of differences*. Adapted from (Gleicher et al., 2011).

Explicit encoding of differences provides a powerful way of highlighting similarities and differences between a set of objects. However, this method is often difficult to use in practice and can potentially require the use of new visual metaphors altogether. Superimposition of is often quite effective at highlighting visual differences, but could on the other hand lead to visual clutter, particularity when the objects of comparison are already complex. Among the three, juxtaposition is arguably the most straightforward method involving the placement of two or more views side-by-side, intuitively mimicking the way we compare physical artifacts. Moreover, juxtaposition provides a scalable way of leveraging the expansive screen estate and pixel density in LHD environments. One drawback of juxtaposition is that it compels the user to manually resolve the differences and similarities by visually scanning the objects. However, in many domains, such as astronomy and ecology, users frequently express their desire to see the objects of comparison side-by-side and visually 'eye ball' the variations instead of relying on more simplified representations.

Rather than rely on domain-specific techniques, we shall focus our analysis on designs that involve the juxtaposition of *high-resolution coordinated views*, as such designs are likely to be applicable to a wide range of domains and problems. We couple these designs with interactions that provide users with *visual cues* to help them manage their attention. In doing so, we would be scaling up the long standing model of *coordinated multiple views* to LHD environments in a manner that avoids inducing *information overload*.

3.2 Visual Thinking Design Patterns

Pattern language originated as an architectural concept developed by Christopher Alexander in 1977 (Alexander et al., 1977). It centers on capturing architectural design ideas as archetypal and reusable descriptions to aid the design of cities and buildings. This concept inspired the notion of *design patterns* developed in the early 90s to codify commonly occurring software design problems and approaches to their solution (Gamma et al., 1993). Today, design patterns are used in the development of all major software systems and have greatly influenced the design of programming languages (such as Java) and programming toolkits (such as Qt).

Visualization experts have also began to identify a number of recurring designs that are applicable in different domains and analysis scenarios. For instance, *Visual Thinking Design Patterns* (VTDP) is a concept developed by Colin Ware as a design framework that categorizes best-in-class in visualization designs (Ware et al., 2013). In addition to prescribing a set of effective designs, VTDP also articulate the cognitive and perpetual operations a viewer employs to reason about data depicted in a visualization.

The advantage of basing our analysis on VTDPs is that we are starting with known designs that are thought to exemplify best practices, which have been identified by the visualization community after years of research. The downside is that we could potentially miss entirely new design patterns that are unique to LHD platforms. Nevertheless, given the relative novelty of LHDs and our limited understanding of their affordances, it is reasonable to start with the existing design space and see how it can be adapted to this new platform.

3.2.1 Components of a VTDP

Each VTDP encompasses one or more cognitive tasks. It also describes a visual form that can help accomplish the said tasks. Some also include a possible set of interactions that represent *epistemic actions* (Kirsh and Maglio, 1994), such as clicking on an object to see more detail. Epistemic actions represent actions that a user takes to change the state of the visualization with the goal of acquiring additional information.

Taken together, the components of a VTDP describe a common design for solving recurring visual analytic problems in a domain-independent fashion. This description is decoupled from the technology so as to be as generic as possible. However, it is possible to ask the following question: given a particular VTDP, what benefits and drawbacks can we expect if we were to apply the prescribed design to render more data elements using a larger display with a higher number of pixels? We shall therefore look at some of the VTDP identified by Ware et al. and analyze the potential gains and costs that can be expected when the visual layout is 'stretched' to a big display. By stretch we imply rendering the visualization to take advantage of the extra resolution afforded by a LHD, while still being able to visually resolve individual glyphs (by a human eye with 20/20 vision). Furthermore, we shall propose a *query-by-example* interaction technique that can be used to selectively highlight features of interest in a perceptually salient manner.

3.2.2 VTDPs for comparative visual analysis

Table I lists three of Ware et al's VTPDs along with a brief description of each patterns. The order of listing starts from low-level patterns which rely mostly on perceptual processing to high-level design patterns that apply to cognitive tasks. We discuss these VTDPs in the rest of this section.

3.2.2.1 Visual query

The benefits of using LHDs to improve performance in basic visualization tasks has been established in numerous studies (Ball et al., 2007; Yost et al., 2007; Shupp et al., 2009). Compared to a traditional desktop display, a larger number of data points can be simultaneously

TABLE I

A subset of Ware's Visual Thinking Design Patterns encompassing design patterns that are applicable in comparative visual analysis scenarios.

VTDP	Description
Visual query	Visual queries represent the most basic operation one can perform with a visualization, by scanning the display for a particular visual pattern. For instance, a viewer scans a weather chart looking of lightening symbols to identify locations of thunderstorms.
Pattern comparison in large information spaces	A common problem in visualization is to compare two or more visual objects of varying complexity. The user executes an epistemic action to locate the first objects (by panning and zooming, for instance), retaining a subset of the object in his/her visual working memory. The user then executes a second epistemic action to locate a second object, comparing the contents of the working memory to it. The sequence is repeated until the comparison is complete.
Seed and grow	This design pattern can be employed when the information space is too large or complex that a meaningful overview cannot be obtained from one view. Starting with a particular piece of information, the user expands laterally, exploring related elements by following a series of links or analysis states. For instance, exploring a social network by following the chain of links starting from a known contact.

visualized. However, the time needed to execute a visual query is dependent on the saliency of the target pattern. When the target pattern is pre-attentively distinguishable based on at least one basic visual feature such as color (Ware, 2012), the search can proceed in parallel and the search time is significantly reduced. Although we would expect a serial search to take a significantly longer time on a LHD due to the cost of executing physical movements, finding perceptually salient targets can be generally sped by up to 30% (Ruddle et al., 2013). For example, finding the locations marked by red triangles across the map in Figure 7 is generally faster when a LHD is used in place of a conventional display with a pan and zoom interface.

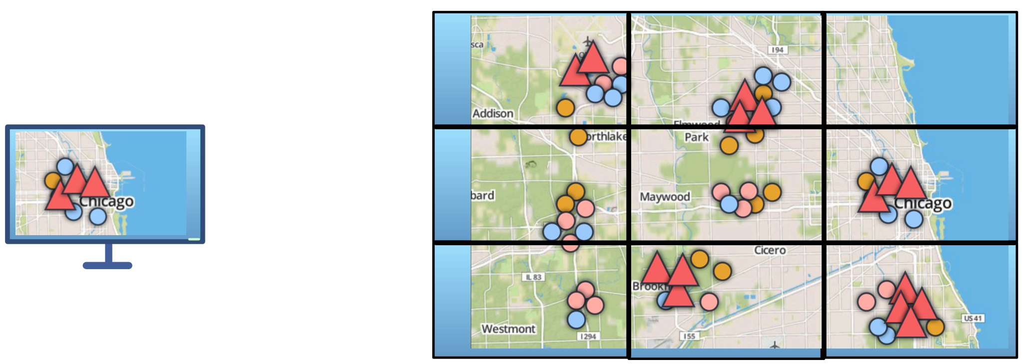


Figure 7. Visually querying for the red triangles can be up to 30% faster on a LHD (right) compared to a conventional desktop display with a pan and zoom interface (left). The improved performance is conferred by the perceptual saliency of the target and the ability to utilize physical navigation in place of virtual navigation.

3.2.2.2 Pattern comparison in large information spaces

A LHD enables more views to be juxtaposed side-by-side, allowing a larger set of objects to be compared. In this case, users can move their attention across different objects, with epistemic actions reduced to simple embodied actions, such as eye movements and head turns. This eliminates the need to temporally switch between views and reduces cost of comparing complex objects that cannot be easily held in the visual working memory (Plumlee and Ware, 2006). Figure 8 illustrates an application of this design pattern to visually compare brain scans from multiple patients.

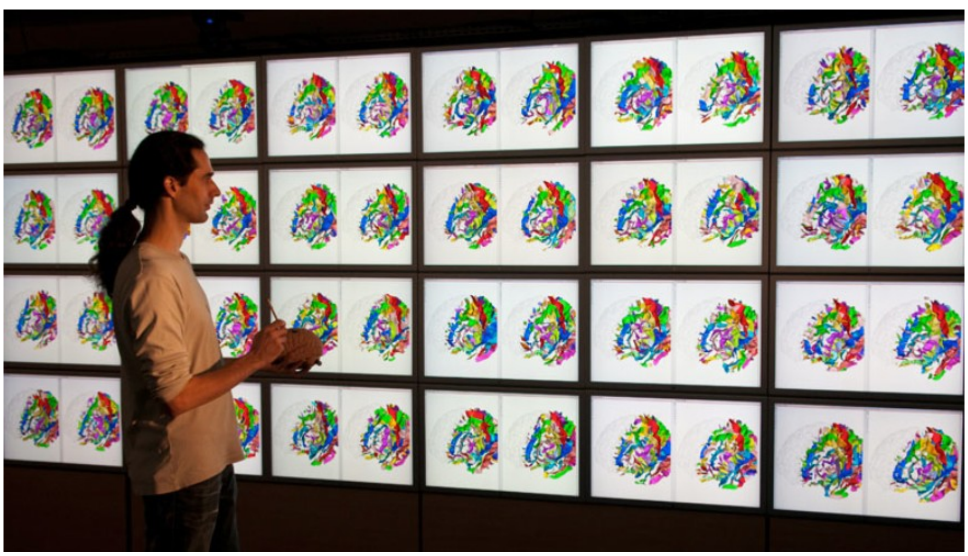


Figure 8. Comparative visual analysis of brain scans from multiple patients. The expansive screen estate and resolution afforded by the LHD allows for the juxtaposition of a large number of high-resolution scans. Consequently, comparison between the scans can be done in-place without disrupting the layout. A brain prop allows the user to control the orientation of the scan. Courtesy of the *WILD* project (Beaudouin-Lafon et al., 2012). Used with permission.

There is potential to overload users with too many views, making visual comparison more difficult. To reduce the chance of inducing an information overload phenomenon, the visualization should make it easier for users to efficiently move their attention among the objects. This can be done by opting for structured layout that can be navigated predictably, providing visual landmarks, and/or adding interactions that can serve as visual cuing mechanisms by enabling users to selectively and interactively highlight patterns of interest. For instance, a visual *query-by-example* paintbrush can be used to brush a pattern of interest in one view, and automatically see instances of that pattern highlighted in other views using a perceptually salient encode (see Figure 9 for an example). The user can then efficiently attend to these features using the *visual query* design pattern.

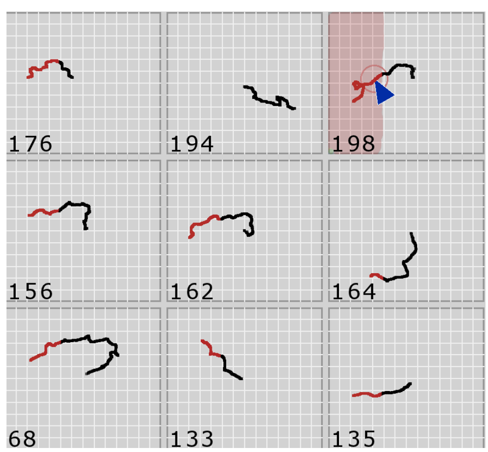


Figure 9. A visual *query-by-example* technique can be used to simplify the comparison of complex objects. A paintbrush tool enables the user to select a visual pattern of interest in one view (top right corner), causing the visualization tool to automatically extract instances of similar patterns in other views and highlight them with a perceptually salient encode.

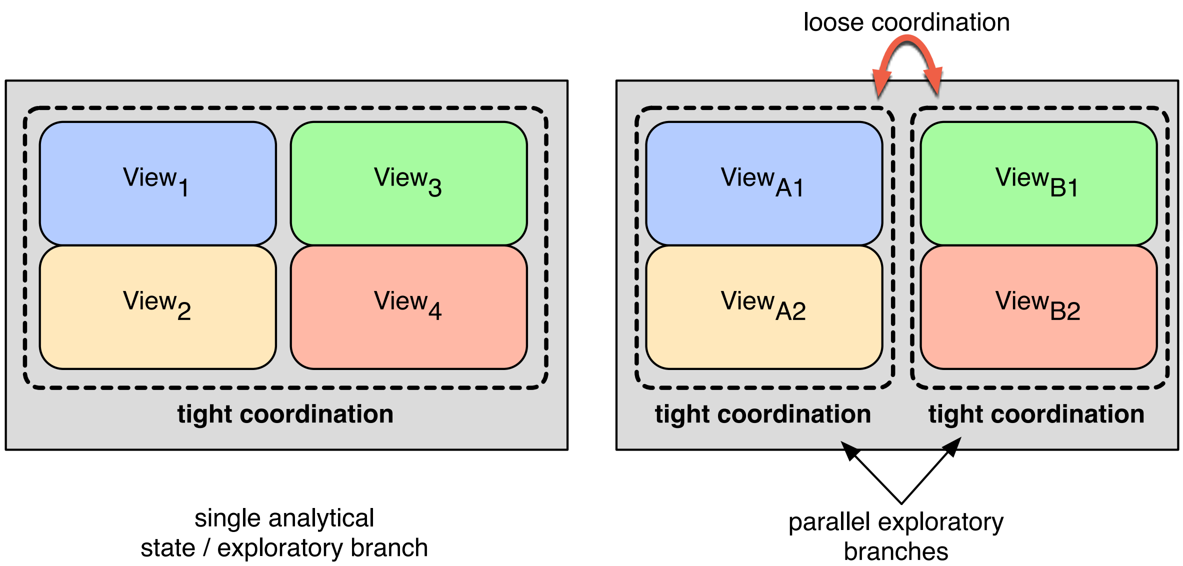


Figure 10. Traditionally, a visualization environment maintains a single analytical state representing one exploratory branch with all views tightly coordinated (left). However, it is beneficial to allow users to maintain separate analytical states and interact with them in parallel. One way to achieve this is to provide multiple groups of *loosely coordinated* views (right). This paradigm could in turn enable users to develop and grow parallel exploratory branches, benefiting the breadth of the exploratory activity.

3.2.2.3 Seed and grow

Visual exploration can be seen as an instance of the *seed and grow* design pattern. The user starts with a particular view and progressively manipulates the visualization tool to affect the visual representations and navigate through different projections of the data. At any point in time, the visualization tool encodes not only a particular projection of the data, but also encapsulate the current analysis state (Jankun-Kelly et al., 2007). On conventional displays, the exploratory activity typically assumes a sequential nature as there is usually enough screen space to show a single analytical state at a time, which forces a sequentialization of the exploration activity, restricting it to a single trajectory. With an LHD environment on the other hand, a visualization tool can potentially show multiple parallel analytical states at a time, by providing screen estate to support this activity. This is illustrated in Figure 10.

3.3 Summary

This chapter addresses RQ_3 with a set of LHD-based Visual Thinking Design Patterns (VTDP). The three VTDPs described above were selected to provide a *vertically integrated* framework for designing exploratory visualizations aimed at supporting tasks involving the comparative analysis of large, homogeneous information spaces. Our adaptations of these three VTDPs should be thought of as broad principles that can guide the design towards perceptually and cognitively efficient visual interfaces for LHDs. As Ware contends, these VTDPs do not represents reusable modules as any modularization would necessarily restrict their application to a particular domain or problem.

In the following chapter, we present the first of our two studies that aim to evaluate the effects of using LHD environments within the context of an exploratory visual analysis scenario.

CHAPTER 4

VISUAL ANALYSIS OF ENSEMBLES WITH LARGE HIGH-RESOLUTION DISPLAYS: A CASE STUDY

Large high-resolution display (LHD) environments have the potential to provide data scientists and researchers with long-needed technological support to enable them to visually explore and make sense of large-scale datasets. Based on theories of distributed cognition and an analysis of the cognitive costs involved in interacting with visualizations, we argued that LHDs can potentially affect user analytic behavior and improve the outcome of the analytic process.

This chapter presents an exploratory study of an LHD-based visual analysis environment designed to support the comparative analysis of large ensemble datasets. We describe and present the results of a user study we conducted to understand how a domain expert employs the environment to explore a complex real-world dataset in the domain of ecology and behavioral biology. This case study contributes a qualitative understanding of the effects of increasing the physical size and resolution of the visualization interface on user analytic behavior. It also provides preliminary evaluation of two design patterns presented in the previous chapter.

We motivate the use case, describe the design of the visualization environment, outline the methodology of this study, and lastly present and discuss our observations.

4.1 Scenario

To understand the navigational strategy and decision-making processes of animals, ecologists track and analyze their movement patterns. However, biological organisms typically exhibit stochastic, locally scoped behavioral responses that are difficult to characterize on a case-by-case basis. Therefore, ecologists resort to collecting large sample sets of animal movement trajectories under a variety of conditions in an attempt to tease out general behavioral responses. Due to the large number of plausible hypotheses that might explicate an observed behavioral pattern, ecologists need a scalable and efficient way of exploring these hypotheses and generalizing spatio-temporal movement patterns to theories. However, the sheer number of trajectories collected during experimentation makes the analysis difficult and time-consuming on conventional displays. This scenario thus provides a good use case to study the affordances of LHD environments and their impact on user behavior and insight formation within the context of an exploratory visual analysis task.

4.1.1 Dataset

Our dataset comprised approximately 500 trajectories, which represent the movement patterns of Seed harvester ants (*Messor cephalotes*) under a variety of experimental conditions. The trajectories were obtained by tracking the movement of ants in the field at approximately 3mm spatial resolution and a temporal resolution of 30 fixations per second. Each trajectory represents the movement of a single ant, which has been captured, taken away from its colony, and placed on the center of an experimental arena to record its behavioral response in a novel landscape (see Figure 11). Trajectories range in duration from 10 seconds to approximately 3 minutes.

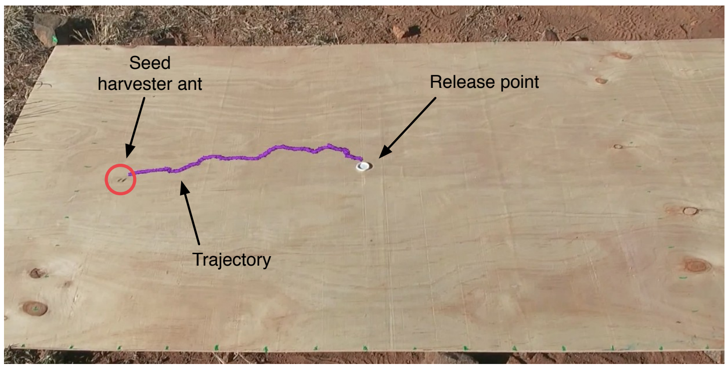


Figure 11. A video freeze-frame from one of the experiments used to generate the Seed harvester ants trajectory dataset. Ants were captured, one by one, and released in the middle of an experimental arena (a rectangular plywood sheet). This posed a navigational problem to the ant in a visually novel area, prompting it to attempt to return to its colony using the available environmental cues. The movement of the ant was analyzed and its trajectory was extracted using a semi-automated computer vision algorithm. The extracted trajectory for this particular instance is superimposed in purple for illustration.

The goal of the behavioral experiments was to understand the navigational strategy employed by the ants, and investigate the possibility an adaptive decision-making mechanism employed by ants to select an appropriate strategy depending on context (Offord et al., 2013). Trajectories were categorized based on the state of the ant at the time of capture. Some of the variables included: position of the ant at time of capture relative to the colony's trail network, initial heading, journey direction (heading away from or returning to the colony's nest), and whether the ant was carrying a seed (i.e., food item).

4.2 Visual analysis environment

The design of the visualization environments employed two of the design patterns presented in chapter 3, namely, *visual query* and *information comparison in large information spaces* patterns. We describe the LHD apparatus employed in the study, describe the visual encoding of individual trajectories, and then elaborate on the visual layout and the interactive features designed into the visualization environment.

4.2.1 Apparatus

We used the *Cyber-Commons* environment to visualize the Seed harvester ants dataset. The Cyber-Commons consisted of thin-bezel, tiled display wall consisting of 18 stereo-capable LCD panels that are arranged in 6 columns and 3 rows, at a physical size of 7 x 3 meters (see Figure 12). The display is also capable of showing stereoscopic 3D content using a micro-polarization technology, which polarizes each pixel row in an alternative direction. The Cyber-Commons display provided a resolution of 8,192 x 2,304 (approximately 19 Megapixels). However, due to technical limitations in the graphics architecture at the time, only two thirds of the total display surface was utilized in order to maintain an interactive frame-rate, at a total resolution of 8,192 x 1,536 (about 12.5 Megapixels).

4.2.2 Visual encoding

We utilized the 3D capability of the Cyber-Commons display to encode the temporality of trajectories. Each trajectory was rendered in stereoscopic 3D, with the XY plane (the display

Figure 12. *Cyber-Commons* is a thin-bezel, tiled, high-resolution LCD wall providing a total resolution of approximately 19 Megapixels. The panels are equipped with a micro-polarization layer, enabling 3D stereoscopic content to be viewed with polarized 3D glasses.

surface) encoding the movement of ant on the ground, and the Z+ axis (away from display) encoding time. The extents of the view was mapped to the borders of the experimental arena. To generate a stereo-pair, we applied a sheer transformation along the X-axis (negative sheer for the left eye image, positive for the right), which produced in an orthographic 3D projection that does not suffer from perspective distortion. Figure 13 illustrates this concept. The trajectories appear as cylinders sprouting from the display surface and extending out to ‘float’ in front of the display.

The use of stereoscopic depth cues as an additional perceptual channel made the temporality of ant movement more evident. This was of paramount importance to our participant who was interested in understanding the insects' decision-making process at the micro-level. Since our data had relatively high temporal resolution with each trajectory containing 30 fixations per second, we were able to utilize the expansive horizontal space of the Cyber-Commons display to depict trajectories at a high temporal resolution. This was achieved by exaggerating the amount of stereo disparity and increasing the shear coefficient, which caused an increased horizontal divergence in the stereo pair. The increased divergence resulted in added depth perception, which helped reveal additional temporal details in the trajectory. To avoid excessive eyestrain,however, we provided a slider to allow the user to control the maximum stereo disparity (see section 4.2.5)

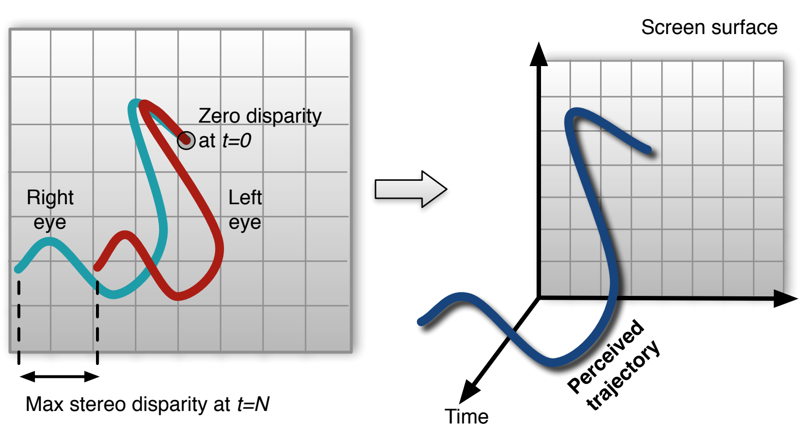


Figure 13. Visual encoding of an ant trajectory, with stereoscopic depth cues to convey time.

4.2.3 View layout

We visualize ant trajectories in a grid-based, small-multiples layout (Tufte, 2001). This layout enabled the participant to divide the screen estate into configurable 'bins' that can be used to group related trajectories. For example, one bin can show trajectories of ants captured *east* of the colony's main foraging trail, while a second bin might contain ants captured on the trail while carrying a seed. Figure 14 shows the overall visualization running on the CyberCommons display. The contents of the groups can also be changed independently, allowing the user to bring in additional information without disrupting the entire layout.

The number of trajectories in the grid can also be varied interactively. The user can switch between a number of configurations by pressing a shortcut number on the keypad, such as 1 for a 15 x 4 layout, 2 for a 24 x 6 layout, and 3 for a 36 x 12 layout. These configurations were chosen to avoid trajectory-bezel overlap, as we wanted to make use of bezels as natural dividers.

4.2.4 Interactive features

To facilitate the comparative analysis task, we provided two interactive features to enable the user to explore hypotheses about ant behavior, and quickly determine whether those hypotheses are supported by the data. First, a *query-by-example brush* allows the user to brush the background of a single trajectory, causing the visualization to highlight motion segments in all other trajectories when the insect moves over the brushed area. Second, a *temporal filter* lets the user specify a filtering time window, causing the visualization to display trajectory seg-

Figure 14. Visualization of ant trajectories on the emphCyber-Commons wall in a 2D layout with stereoscopic depth cues. Trajectories can be grouped into 'bins' which are given distinct background colors to make them visually discernible. In this figure, five distinct trajectory groups are juxtaposed corresponding to ants that were captured on the colony's main trail (blue background), west (red), east (yellow), north (gray), and south (green) of the trail. Copyright © 2012 IEEE. Used with permission.

ments corresponding to insect motion within the specified duration only. We describe these two features in more detail.

4.2.4.1 Query-by-example brush

The query-by-example brush allowed the user to highlight a particular spatial pattern of interest in one view, and have the visualization automatically highlight similar instances in the rest of the layout with a perpetually salient visual encode. In our application, the user can brush the background of a single trajectory using a paintbrush tool. This causes segments in all currently displayed trajectories to be highlighted when the insect moved over a brushed area. For example, the researcher could brush the west (left) side of a trajectory to highlight all instances where the ant exited the experimental arena from the west. Figure 15 illustrates this feature.

When we designed this feature, we envisioned that it would be helpful when looking at pair-wise similarities between a small collection of trajectories. However, the study showed that this simple feature played a far more important role: it enabled the participant to visually cue the trajectory data and perform fast *visual queries* on the entire layout, owing to the perceptually silent visual encode.

4.2.4.2 Temporal filter

The temporal filter allowed the user to interactively specify a time window of interest, causing the visualization to filter out movement segments that occurred outside the specified window. For example, the user can choose to display the first 30 seconds of the experiment. The time window can be selected using a range slider (see Figure 15).

4.2.5 Ergonomics

Prolonged viewing of stereoscopic images has been known to cause discomfort for some viewers, mainly due to excessive binocular parallax and accommodation-convergence conflict (Lambooij et al., 2007). To reduce the chance of inducing fatigue, we included a set of controls in the visualization to modify the 3D view to allow for comfortable, prolonged viewing. A slider enables the user to push trajectories so that they lie in front of the display surface,

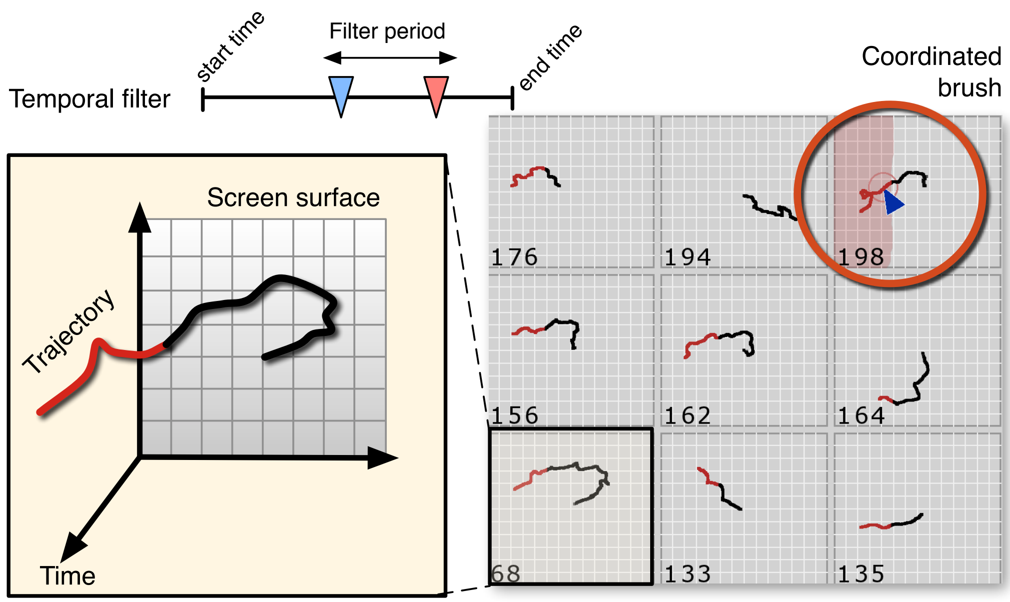


Figure 15. The coordinated query-by-example brush (top right corner) along with the temporal filter (top left) can be used to visually specify and test for spatio-temporal patterns of interest across a large group of trajectories. In this example, the brush is being used to query for trajectories bearing east-side exits using a red highlight. A highlight in the majority of trajectories indicate somewhat uniform exit points suggesting a consistent navigational strategy employed by ants in this particular group.

behind the display surface, or somewhere in between. Additionally, the time-scale can be de-exaggerated using a second slider. Using these two sliders, the user can control the maximum amount of binocular parallax and keep it within a comfortable range while maintaining sufficient depth cues.

4.3 Method

We had two general goals in this exploratory study: first, we wanted to get a sense of how a user might utilize a visualization tool built around the design patterns presented in the previous chapter to explore a large and complex dataset. Secondly, we wanted to understand how the user adapts his/her analytic strategy to take advantage of an LHD-based visualization environment. Rather than attempting a full comparative experiment at this stage, we decided that we could learn more by closely observing one expert user who had a real-world dataset and a curiosity to explore it.

4.3.1 Participant

Our participant for this study was a doctoral student in behavioral ecology and evolutionary biology, with research interests in the collective behavior of insects. The participant was the principle investigator of the Seed harvester ant projector (described below), and had extensive background on the behavior of socially foraging insects. For the reasons mentioned in section 4.1, the analysis of insect behavior is a particularly challenging problem that can benefit from new exploratory analysis tools.

4.3.2 Procedures

The participant sat approximately 3 meters away from the display and interacted with the visualization using a mouse and a keyboard. At the beginning of the study, the participant was given a brief training on how to use the visualization environment. After that, the participant was given approximately an hour to freely explore and analyze the Seed harvester ants dataset. The subject wore a pair of lightweight polarized 3D glasses (identical to the typical theatrical 3D movie glasses) during the entire duration of the study to resolve the 3D stereoscopic encoding employed in the visualization (see section 4.2.2). We instructed the participant to think aloud during the activity and video and audio-recorded the session.

4.3.3 Analysis

We used a two-pass coding scheme to analyze the data with emphasis on the verbal protocol. The first pass focused on capturing the participant's low-level interaction with the visualization environment as well as tagging any insights, questions, or hypotheses uttered by her. The second pass focused on characterizing the participant's higher-level strategy and the flow of the exploratory activity.

4.4 Results

We describe our general observations on how the participant utilized with the environment and the type of interactions performed during the study. We then examine the flow of the analytic activity and analyze the participant's strategy.

4.4.1 Interaction pattern

The participant started her analysis by first attempting to juxtapose and organize trajectory groups on the large display in a meaningful way. Often, the display was divided into 2–5 groups with each group displaying trajectories of ants captured under different but related experimental conditions. Once the trajectory group were configured, the participant spent most of her time investigating potential relationships between the visible groups before moving on to investigate different groups.

To get a sense of the distribution and types of interaction initiated by the participant during the study, we broadly categorize them into two categories: *layout-preserving* interactions, which comprise actions that do not cause significant layout changes. This category covers two interactive features in our visualization: the query-by-example brush and the temporal filter.

TABLE II

Breakdown of the time spent interacting with the visualization environment.

Type of interaction	Time spent interacting
Query-by-example brush and temporal filter (*layout-preserving*)	57 min (74%)
Workspace management (*layout-changing*)	20 min (26%)

The second category comprises *layout-changing* interactions which cause significant changes to the visual layout, potentially requiring the user to rebuild a mental map of the visualization. In our visualization, this category refers *workspace management* actions, such as organizing the trajectory groups on the display, by creating and bringing in additional groups of trajectories, removing existing ones, or adjusting the number of trajectories displayed in the small-multiples layout.

Table II shows the time spent by the participant interacting with the visualization environment broken-down by the two categorizes defined above. From this table, can see that the majority of interactions comprised layout-preserving actions (i.e., query-by-example brushing and temporal-filtering). This represents a somewhat expected interaction pattern, as we would anticipate a reduction in the frequency of virtual navigation, when a LHD is used in place of a conventional display (Ball and North, 2005). The pattern in Table II also serves to confirms our general observation that the bulk of the activity was done using layout-preserving actions without the need to perform non-essential actions to explicitly manage the information contents of the display (by frequently switching between different trajectory groups,), as is typically the case with a conventional low-resolution display.

Overall, the participant was able to cover 416 out of 496 trajectories in the original dataset (approximately 84% of the data) during the exploratory activity. These trajectories were visible at least once during the study, though we cannot tell if the participant looked closely at each one. We note, however, that a quick glance over trajectories is sufficient when attempting to identify correlations or variations between multiple trajectory groups, if the query-by-example brush is used to highlight patterns of interest.

4.4.2 Transition diagram

To get a high-level understanding of the participant's exploratory strategy, we created a state transition diagram to characterize the flow between the various components of the analytic activity. Crucially, this diagram captured key *mental states*, which were identified from the verbal protocol during the second coding pass, as well as transitions between these states. This technique was inspired by the work of Ratwani et al. who applied it to study the cognitive processes involved in qualitative reasoning with the aid of graphs (Ratwani et al., 2008). However, in addition to identifying relevant cognitive states, we also included *epistemic states*. Recall that epistemic actions represent actions that a user takes to change the state of the visualization with the goal of acquiring additional information (Ware et al., 2013).

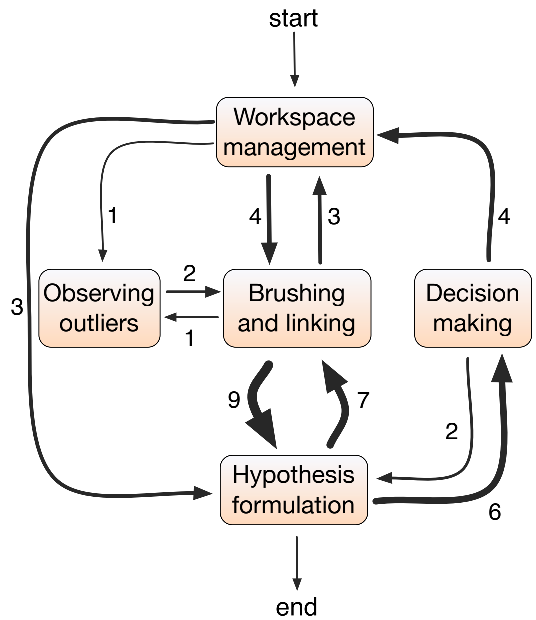


Figure 16. A state transition diagram illustrating key states in the participant's analytic activity. Three main mental states were identified from the verbal protocol: *Observing outliers*, *Hypothesis formulation*, and *Decision making*. Additionally two epistemic states were included: *Workspace management* and *Brushing and linking*. The weights of the arrows indicate how many times a transition has occurred between two states during the activity.

We identified three principle mental states from the verbal protocol: *Observing outliers*, *Hypothesis formulation*, and *Decision making*, which represented instances when the participant made a clear decision to accept or reject a particular hypothesis. In addition to the three cognitive states, two epistemic states were included: *Workspace management* reflected layout-changing actions induced when the participant would bring in new trajectory groups or rearrange existing trajectories, and *Brushing-and-linking* which reflected layout-preserving actions invoked using the query-by-example brush and the temporal filter.Figure 16 illustrate the resulting transition diagram.

The most salient features in Figure 16 is the entwinement of *Brushing and linking* and Hypothesis formulation, which is reflected by frequent transitions between these two states. In comparison, we see fewer transition to and from the *Workspace management* state. Generally, the transition diagram suggests that the activity was primarily driven by hypothesis formulation, reflecting a top-down approach to the exploratory task (i.e., from mental model to observations).

4.4.3 User strategy

The main effect we observed for the use of the Cyber-Commons display manifests in the participant's reliance on layout-preserving interactions to carry out the bulk of the exploratory process. The most salient aspect of this process was the use of the query-by-example brush and the temporal filter, in concert, as a mechanisms to visually query a large collection of trajectories. Often, such queries were formulated and structured as visual tests to evaluate a particular hypothesis. For example, to verify the whether *off-trail* ants employed the sun as a celestial compass to orient themselves back to the trail, the participant isolated trajectories of ants captured *east* of the trail in one group and proceeded to brush the *left* side of one of the trajectories to query for *west* exit points. Upon seeing a red highlight in the majority of trajectories in that group, the participant concluded that her hypotheses is supported (see Figure 15). Moreover, the same hypothesis was reinforced upon applying the same visual query to *on-trail* ants, and noticing that this group exhibited entirely different exit points that are non-consistent with their *off-trail* counterparts.

One key advantage to using a LHD manifests in how visual queries were operationalized during the exploratory activity. Rather than making pair-wise comparisons between individual trajectories, the participant appeared to visually compare sets of trajectories at a time. For instance, *off-trail* trajectories were compared against trajectories of *on-trail* ants. This form of comparison did not necessarily relate to any particular item, but rather seem to be driven by collective difference between the two sets (though, in few instance, individual outliers were noted and were investigated individually). Importantly, this form of set-level comparison was possible as a consequence of two factors: the ability to simultaneously see two sets of trajectories and the ability to perceptually highlight the relevant visual attributes (using the query-by-example brush) so that the two sets can be compared with relative ease.

Another observation we made during the study relates to how new hypotheses were derived by refining an existing hypothesis. Reframing of existing hypotheses often occurred when a particular visual query could not account for an observed spatio-temporal pattern (hence partially invalidating the current hypothesis). For instance, upon seeing that a collection of *on-trail* ants did not demonstrate a preferential exit point exhibited by *off-trail* ants, the participant hypothesized that the former group would be confused by the sudden disappearance of pheromone cues normally found on the trail, and thus would take longer time before deciding on their next strategy. She then proceeded to brush the center of a trajectory (corresponding to the point of release) with a green color, expecting "*off-trail* ants [to] start green and turn black faster [than their *on-trail* counterparts] because they know where they're going." This strategy of refining existing hypotheses, based on how different trajectory groups reacted to a visual query, appears to be an important factor in driving the exploratory activity and expanding the search towards new hypotheses. Once a relationship is recognized and noted, the participant seemed to be eager to test for a new visual pattern that offered a somewhat complimentary narrative.

A further point to note here is the crucial role of stereoscopic depth cues in revealing complex spatio-temporal behavioral patterns. One particularly interesting example came out when the participant was attempting to investigate whether ants that have dropped their seed during the capture process spent more time in the center searching for the seed. To test for this complex pattern, the participant brushed the center of the experimental arena with green and set the temporal filter to show the beginning of the experiment. Her intuition became stronger upon seeing green segments that were roughly perpendicular to the display surface, which indicated little movement. This ability to see temporal pattern unraveled in space appeared to be crucial in this scenario. The participant later commented that, owing to the 3D stereoscopic encoding, she was able to perceive the periodicity of ant behavior not only on a single-trajectory basis, but also on a larger scale.

4.5 Discussion

Overall, the experience of the participant was overwhelmingly positive. Originally, the participant relied on *Matlab* as her primary analysis platform, which allowed her to look at few trajectories at a time. Our LHD-based visualization environment, on the other hand, appeared to be more helpful to the participant in carrying our a large-scale comparative analysis, making it "easier to think visually than in Matlab". Results of the study seem to support the usefulness of the design patterns described in chapter 3, particularly the *comparison in large information spaces* pattern and the *query-by-example brush*. Yet, there is one point that is worthy of further discussion.

4.5.1 Exploring the hypothesis space

The impact of being able to perform large-scale visual comparisons in a cognitively efficient manner can be understood from looking at the cycle between *Brushing and linking* and *Hypothesis formulation* in Figure 16. The frequent transitions between these two states suggest a close association between the ability to visually query for set-level relationships and the conceptualization of new hypotheses. Interestingly, these set-level visual queries directly expressed domain questions, enabling the participant to evaluate hypotheses with primarily perceptual operations. Importantly, this form of hypothesis validation did not involve cycling back-and-forth between a series of temporally-separated views, as is typical on a conventional display, but rather involved quick glancing over trajectories that are already visible on the Cyber-Commons display. This cognitive efficiency appeared to encourage the participant to formulate and explore multiple hypotheses in succession, within a short time. In fact, the participant spent most of the time contemplating a variety of theories and scenarios and evaluating them with quick visual queries. Although such visual queries may not be enough to quantitatively substantiate a particular hypothesis, they nevertheless provided a high-fidelity, low-cost data assessment scheme, enabling the participant to explore a larger *hypothesis space* (Klahr and Dunbar, 1988).

The utility of visual queries in this application lies in design of the query-by-example tool, which provided a flexible visual mechanisms for specifying and highlighting complex spatio-temporal patterns with a perceptually salient encode. However, this utility is also derived from the exquisite resolution provided by the Cyber-Commons display, which made it feasible to visually highlight arbitrary patterns across a large number of trajectories, effectively turning them into visually-cued information sets that can be efficiently compared using a combination of visual scans, eye movements, and head turns. On conventional displays, performing this form of set-level comparison would typically require a significant amount view switching to see different objects and/or different parts of the information. In contrast, with a LHD such as the Cyber-Commons, comparing complex objects can be mostly performed in-place, using layout-preserving interactions and embodied physical actions.

The above distinction has important implications, as it heralds fundamental changes to the cost structure of visual exploration. Preservation of the visual layout helps users maintain their *mental map* (Purchase et al., 2007; Alper et al., 2014), effectively reducing (or perhaps even eliminating) the *view-change* costs (see section 2.2.3.2). This reduction in cognitive costs appears to have encouraged our participant to explore a wide range of relationships and follow up on multiple hypotheses. Although we did not find evidence that the participant was actively considering competing theories, our observations indicate that she was able to formulate and explore hypotheses that offered complementary accounts, which served to refine her beliefs during the exploratory process.

Looking at the comments of the participant after the study, we also find evidence that our LHD-based visualization seems to have broadened her understanding of the problem. For instance, the participant recounted that the LHD-based visualization represented “the only way we can appreciate on a big scale –we are not taking about individual trajectories here– things like the periodicity of ant behavior, the patches of color [resulting from the application of query-by-example brush], and effects [of spatial context] on navigation and decision making”.

4.6 Conclusions and study limitations

This case study directly addresses RQ_1. It provides qualitative evidence that the physical properties of the visualization interface play an indirect role in modulating the visual exploration activity, potentially affecting the analytic behavior of users in important ways. A larger visualization interface with more pixels allows one to see a larger portion of the information space at once, and may thus improve the diversity of relationships one can visually observe. These low-level affordances may in turn encourage users to explore their data more broadly, and invest more time in formulating and testing complimentary hypotheses and narratives during the analysis.

This case study also addresses RQ_3, providing preliminary evaluation of the *visual query* and *comparison in large information spaces* design patterns. In particular, the study provides a compelling example that demonstrate the effectiveness of the *query-by-example brush* in facilitating the comparative analysis of large-scale ensemble datasets in LHD environments.

There are two main limitations in this study. The first obvious limitation is that our analysis is based on data collected from a single participant. Since our participant was a domain expert acting as the principle investigator for the Seed harvester ants project, it was practically difficult to find and recruit other potential participants with similar background and experience. This limits our ability to draw broad conclusions from this study. On the one hand, we are able to provide evidence on the usefulness of the visualization environment presented here for tasks involving the comparative analysis of ensemble data. On the other hand, it is difficult to generalize the results of the study beyond this particular application domain. The second limitation lies in the lack of a baseline condition that would have allowed us to measure the relative effects of increasing the resolution and size of the display on the exploratory process. Although our participant had used Matlab extensively (on a conventional laptop display) to visualize and compare trajectories prior to our study, we do not consider the two conditions to be comparable, as they provide entirely different sets of interactions.

Following up on results from this exploratory case study, the next chapter presents an experimental study that addresses theses limitations, within the context of a different application domain.

CHAPTER 5

EFFECTS OF INCREASING THE DISPLAY SIZE AND RESOLUTION ON USER BEHAVIOR AND INSIGHT ACQUISITION

In this chapter, we investigate the impact of increasing the physical size and resolution of the visualization interface on user behavior and insight generation during exploratory visual analysis. We employ a custom-designed visual analytic environment based on the design patterns articulated in chapter 3, and study users' analytic behavior and performance under two display configurations: a *small* and a *large* display. As in the previous study, we employ an open-ended exploratory task situated within the context of a real-world scenario and dataset.

We present the design of an LHD-based visualization environment for exploring spatio-temporal crime patterns in large metropolitan areas. We then present and discuss the results of the user study. But first, we motivate the scenario and describe the dataset, which together provide the analytical task framework for this study.

5.1 Scenario

As cities and government embrace data transparency initiatives, the number of publicly-available municipal datasets is rapidly increasing. With the availability of detailed police reports on crime incidence, the analysis of crime patterns in is becoming a 'big data' problem. This provides opportunity for government officials, civic groups, and community organizations to understand evolving crime patterns in urban centers and target affected communities with appropriate social support programs at the micro-level, in an attempt to reduce crime incidence and prevent outbreaks before they occur.

We based this study on a scenario that revolved around the analysis of crime activity and patterns in the city of Chicago. The city has made available a database containing the majority of crimes that happened between 2001 and present, with detailed information containing accurate location (down to a city block), date and time, crime type, arrest records, among other information. To make the exploratory task more manageable during the limited experiment time, we used a subset of the dataset comprising crimes committed between 2006 and 2012, inclusive. The total number of crimes in this dataset totaled approximately 2.8 million crime reports, which were broken down into 8 different crime types depending on the nature of offense.

The above scenario furnishes a suitable framework to study the analytical behavior of users within the context of a visual exploratory task for the following reasons. First, the scenario provides a semantically rich task for our participants to try to find temporal patterns that govern crime activity in different parts of the city, identify spaito-temporal correlations between distinct types of crime, and explore cause-and-effect narratives in an attempt to explicate the observed patterns and correlations. Additionally, the complexity and subtlety of criminal activity in Chicago, which varies greatly across different neighborhoods and yet appears remarkably consistent over time, provides a compelling use case of the use of LHD environments to visually explore the dataset in an attempt to discover 'hidden' patterns and outliers within the larger picture. Lastly, while the analysis of crime activity is by no means an easy task and certainly requires an appropriate level of domain expertise, the task can be considered to be semantically appropriate for a participant pool drawn from the general population. Although the correct interpretation and forecasting of crime trends require skill and experience, common sense judgement and everyday knowledge maybe sufficient to to bootstrap participants and engage them with the analytical task so that they can begin to generate insights. Since our goal is to study the process of exploratory visual analysis, we do not require participants to rigorously verify their observations.

5.2 Visual analysis environment

We describe the LHD apparatus employed in this study. We then proceed to describe the design of the visualization environments, the visual representations employed, and the set of interactions provided to participants.

5.2.1 Apparatus

We used the *CAVE2* environment as our LHD display apparatus for this study. CAVE2 is a cylindrical system measuring 7.3 meters in diameter and 2.4 meters in height. The environment consists of 72 thin-bezel stereoscopic LCD panels arranged in 18 columns and 4 rows, creating an approximately 320-degree panoramic environment at a total resolution of 74 Megapixels (Reda et al., 2013b). Figure 17 illustrates the environment.

The choice of CAVE2 was motivated by the fact that the cylindrical configuration of the environment makes it possible to view the displays at similar distances and visual angles form the center, which reduces the amount of geometric distortion in the periphery compared to flat wall displays (Bezerianos and Isenberg, 2012).

Figure 17. The interior of the *CAVE2* environment consists of 72 thin-bezel LCD panels cylindrically arranged in 18 columns and 4 rows. The inset illustrates a schematic bird's-eye view of the environment.

5.2.2 Visualization

The visualization employs two main types of visual representations: heatmaps and timeline charts. A heatmap is juxtaposed over a geographic map to show crime density in a particular area in the city over a single year. The timeline charts illustrate aggregate crime trends over three time scales: monthly to show crime patterns over an entire year, daily to illustrate crime pattern in a typical week, and hourly to illustrate trends in crime activity during a typically day. The two representations are combined in one compound view, with the three timeline charts showing aggregate number of crimes occurring within the extents of the geographic map. Each view shows crime density of a particular type of crime (e.g. robbery, narcotics violation, serious crimes, or all crimes combined) over a single year in one particular area. Panning the geographic heatmap can be paned using familiar drag-and-drop operations to move the map and show different parts of the city. Moving the map causes the charts to instantly update to reflect crime trends in the new area. To simplify navigation, we also designed an *overview + detail* navigation method by providing an overview map showing the entire city, with a rectangular selection box providing an alternate way to pan the map by dragging the rectangle. Figure 18 illustrates the compound view and the overview + detail navigation method.

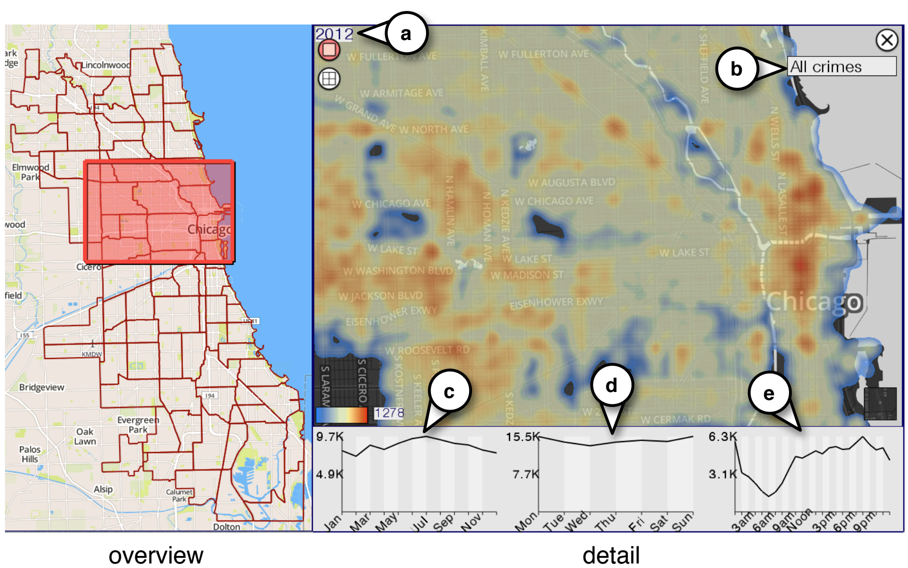


Figure 18. The overview map (left) and the compound view which combines a geographic heatmap with three timeline charts (right). The compound view shows one year's worth of criminal activity (a) for a particular type of crime (b). The heatmap shows the spatial distribution of criminal activity during the entire year. The three timeline charts show crime trends occurring within the map at three time scales: monthly (c), daily (d), and hourly (e).

One important design decision we took early on was to limit the visualization to a single scale. That is, there was no way to zoom in onto a particular area in the heatmap. This decision was primarily done to eliminate the confounding factor introduced when participants allowed to 'read' the visualization at multiple scales. To compensate for this restriction, the overview map provided an alternative way of navigating the city. Additionally, we also provided participants with a slider to adjust the 'coarseness' of the heatmap; a coarse-grained heatmap coalesce crimes into larger blobs, whereas a fine-grained map allows one to almost see the individual crime incidents as distinct points on the map. This features was important as the spatial distribution patterns for different types of crimes require different aggregation levels in the heatmap.

5.2.3 View coordination

To facilitate the comparative analysis of crime patterns between different years and/or crime types, additional views can created in two ways: from the overview map, or by extending an existing view. Views created directly from the overview map are not coordinated, whereas views extended from existing ones are 'chained' together and coordinated to collectively respond to user actions. *Chained views* share the same geographic extent and the same brushing-and-linking state, but can be set to show different crime types or different years. Panning the map in one view causes a corresponding movement in all chained views. In addition to being geographically coordinated, chained views can also be synchronized in one of two ways: temporally by showing crime activity for the same year but across different crimes categories (e.g., narcotic crimes, burglaries, and thefts in 2012), or categorically by showing the same crime category but over multiple years (e.g. vehicle thefts in 2006, 2009, and 2012). Figure 19 illustrates this concept. This features was added to simplify the layout of views and provide a semi-automated way of juxtaposing and organizing a large number of views.

This design was motivated by two aspects we want to provide support for. First, we wanted to give participants a way to drill down into the data while being able to leverage the size and resolution of the display environment to maintain contextual awareness. The geographic heatmap provided a detailed in picture of crime density down the block-level. Secondly, the flexibility afforded by the chained views meant that they could be used to view different parts of the city, different years, or different types of crimes, enabling participants to visually look for correlations in criminal activity across these dimensions. For instance, time synchronization provide a way to look at temporal patterns in crime activity. Alternatively, categorical synchronizations of views based supports the comparison of crime patterns for different types of criminal activities, enabling one to find possible relationships between narcotics-related crimes and homocides, for instance.

Ultimately, chained views represent a set of *tightly coordinated* views that can be manipulated and navigated collectively. However, multiple autonomous chains can be created in parallel and manipulated independently of each others (see Figure 19). This features represents an instantiation of the *seed and grow* design pattern (see section 3.2.2.3), providing a mechanisms for users to instantiate multiple exploratory branches that can be manipulated in parallel and followed up upon independently. Such independent branches can be used to explore crime patterns in different parts of the city, or can potentially provide scaffolds for users to explore and compare two or more alternative narratives or hypotheses, for instance.

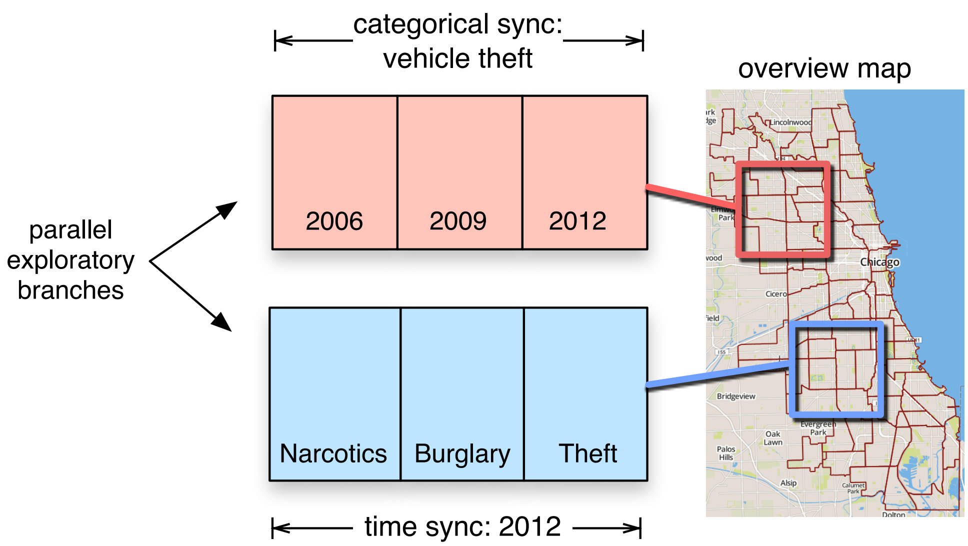


Figure 19. Two sets of *chained views*. The red chain represents a categorically synchronized set of views illustrating vehicle theft trends in 2006, 2009, and 2012 in the northwest part of the city. The blue chain is temporally synchronized, illustrating narcotics, burglary, and theft crimes in 2012 in the south-central part of the city. Chains are visually linked to their geographic location in the overview map. Views in each chain are tightly linked to respond collectively to brushing-and-linking actions and map movements. However, the two chains can be manipulated independently, providing two parallel exploratory branches.

5.2.4 Automated view layout

To help users organize the screen space, we provided two layout modes:

- A freeform layout enabled users to create their own view-layout and position views freely in the environment using a familiar window-based metaphor.
- A structured layout spread the chained views around the center of the display, aligning them with the bezels. This layout mode further provided users with two options, allowing them to choose between 2 or 4 independent view chains

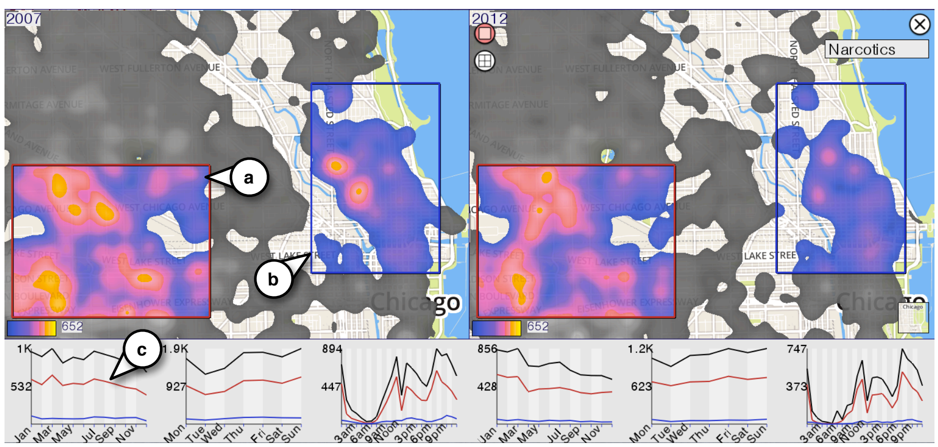


Figure 20. Two chained views comparing narcotics-related crimes in 2007 (left) and 2012 (right). Two bounding boxes highlight crime patterns in the west-central part of the city (a) as well as downtown and the near-north side (b). Bounding boxes created in one view are automatically propagated to all chained views. In addition to highlighting the selected subregions in the heatmap by greying out non-bounded areas, bounding boxes cause additional trend lines to be visualized, allowing one to examine detailed crime trends in subregions of interest within the context of a broader trends (c).

5.2.5 Brushing-and-linking

Brushing-and-linking in one view propagates to all chained views, and can be done in one of two ways. First, rectangular bounding boxes can be created and moved within the heatmap, which are updated accordingly across all chained views. Bounding boxes cause additional trend lines to be visualized in the chart area, showing crime pattern corresponding to the selected regions. Figure 20 illustrates how bounding boxes work across two chained views. A second way to achieve linking between the heatmap and the timeline charts is to brush one of the charts to select a time range, which causing the heatmap to update to reflect crimes that happened within the brushed time range only.

The bounding boxes serve two main proposes. First, they allow one to visualize crime trends within a subregion in the heatmap, and compare those trends to the overall trend. Secondly, they allow a user to mark and highlight important features in the heatmap for comparison. This is also facilitated by the fact that the bounding boxes are automatically synchronized across all chained views, allowing a user to compare crime density in a particular subregion region across multiple years without disrupting the layout, for instance.

5.3 Methodology

Our first goal in this study is to understand the effects of increasing the physical size and resolution of the display on the quantity and quality of insights and hypotheses formulated by participants during the exploratory activity. Our second goal is to understand potential variations in user behavior induced as a result of increasing the size and resolution of the visualization interface.

In this study, we employ two experimental conditions in this study: a *small* and a *large* display. The two conditions provide identical visual representations coupled with the same set of interactions. However, the two displays afford distinct experiences; the larger display promotes the juxtaposition of views at the potential risk of overloading the interface with too much information, while the small display provides a potentially simpler visualization interface that promotes the temporal-separation of views at the risk of providing a narrower view onto the information space. The key different between the two conditions therefore lies in the amount of information that can be simultaneously viewed, which serve to implicitly modulate the utility of the visual analytic operators afforded (hypothesized to be higher with the large display) as well as the effort needed to manage one's attention (hypothesized to be lower with the small display) during the exploratory activity.

5.3.1 Hypotheses

Our hypotheses for this experiment were:

- H_1: We hypothesize that the cumulative number of observations would increase with the large display, driven by the ability to see more views side-by-side. We also expect the rate of observations (defined as the number of observations reported during a minute of analysis) to increase with the large display as participants would be able to access information more efficiently using embodied actions as opposed to virtual navigation, and thus would be able to devote more time to the actual analytic task.

- H_2: Extrapolating from the case study in chapter 4, we expect an increase in number of hypotheses formulated by participants given the large display. Similarly, we also expect an increase in the rate of hypotheses.
- H_3: Lastly, we expect participants to acquire higher-level, more integrative insights with the large display, given the ability to see a larger portion of the information space at once. One way to measure this effect is to quantify the "broadness" (or "narrowness") of observations and hypotheses formulated by participants. We define criteria to measure this quality in section 5.3.6.

The above predictions are based on results from earlier studies which indicate improved performance time in basic visualization tasks (Ball and North, 2005; Yost et al., 2007). We believe that that such low-level performance gains induced by having access to a larger display would in turn translate to higher-level analytical gains in exploratory visual analysis scenarios. A counter argument, however, is that the surge in the amount of information visible on the large display would dramatically increase the complexity of visualization interface, hamper participant's exploration efforts, and ultimately result in a poorer analytic outcome compared to the small display. Alternatively, participants may perceive the additional screen estate and resolution afforded by the large display to be futile, thus resulting in a similar outcome and user behavior across the two conditions.

5.3.2 Participants

Ten volunteer participants (4 female) were recruited from our lab. The majority of participants were computer science graduate students, with one being a staff member. Participants ranged from 25 to 54 years (*mean* = 31.3, *SD* = 8.3). All participants were members of the *Electronic Visualization Laboratory*, and thus were familiar with LHD environments and had prior experience in using the CAVE2 environment in some capacity. However, none of the participants had seen or used the particular visualization environment reported here prior to the study. Participants reported their familiarity with interactive visualizations on a 5-points Likert scale, with 1 being "not at all familiar" and 5 being "very familiar" (*mean* = 4.3, *SD* = 0.8). Participants also reported how often they perform data analysis tasks as part of their normal work activities on a 5-points Likert scale, with 1 being "never" and 5 being "a great deal" (*mean* = 3.2, *SD* = 1.0). All participants resided in the Chicago Metropolitan area, and thus were familiar with most neighborhoods in the city, which made it easier for them to engage in the task. Although it is likely that participants brought their intuitions and existing perceptions of crime prevalence throughout the city, we consider this to be an acceptable compromise as our goal is not to measure the accuracy of participants' findings, but rather to understand their exploratory strategy and the type of discoveries they end up making. We refer to participants as S1–S5 and L1–L5 depending on whether they undertook the experiment with the small or the large display, respectively.

5.3.3 Procedure

Participants sat in the center of the environment on a swivel, height-adjustable chair at a distance of approximately 4 meters from the displays and interacted with the visualization using a standard mouse and keyboard setup. Participants were also provided with a notepad and pen to use during the activity, if desired. Figure 21 illustrates the experimental setup. While the study was open-ended, we provided participants with a written description of their ‘task’ which outlined the scenario and provided a general description of the dataset. Additionally, the scenario document contained a list of questions for use as starting points. Examples of questions provided to participants include:

- What year has the most crime occurred in?
- Identify crime ‘hotspots’ throughout the city. Are there distinct hotspots for specific types of crime?
- How does crime in the city vary with time of day, day of the week, and season?
- When do most crimes occur (time of day, day of week, and season)?
- Are there correlations between different types of crimes?

However, participants were instructed to use the provided list of questions for general guidance, and were told that they are otherwise free to explore the dataset in any way they desire and followup on questions that are of interest to them. The scenario document is included in appendix A.

The experiment began with a 15 minutes training session with the experimenter explaining the task and demonstrating the visualization environment and its various interactive features. Participants were then given 150 minutes (2.5 hours) to explore the Chicago crime dataset, and instructed to think aloud during the activity and report interesting observations, salient patterns and outliers, correlations, trends, as well as hypotheses that explicate their observations. The session was video and audio recorded. Additionally, we collected activity logs which

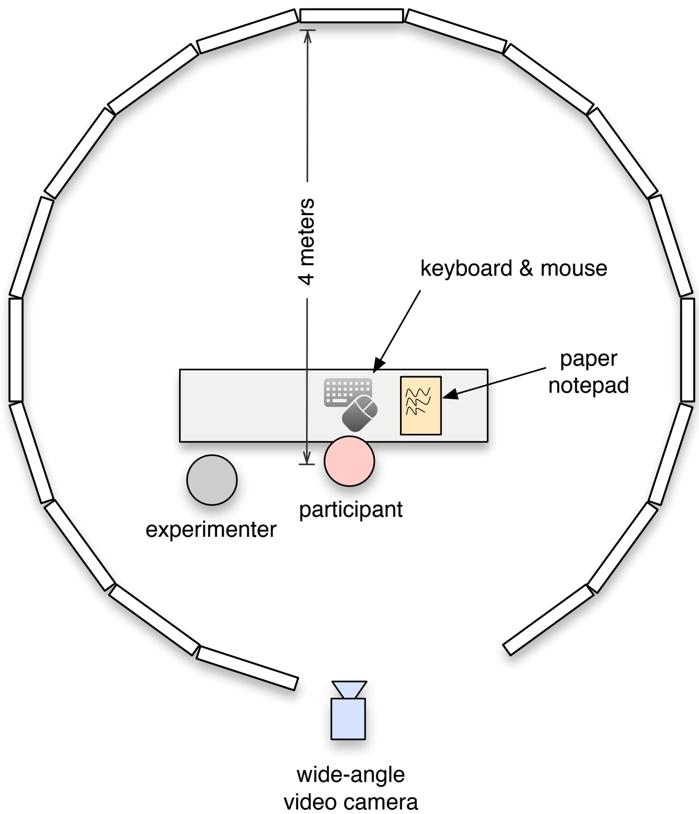


Figure 21. A top down view of the experimental setup in *CAVE2*.

recorded all interactions with the visualization, such as creating new views and bushing the time charts. Participants were informed that they could end the experiment when they felt that they have exhaustively explored the dataset. Otherwise, the experiment was terminated when the 150 minutes of allotted exploration time has elapsed. Participants were allowed to take breaks during the experiment, if desired. At the conclusion of the study, participants underwent a short semi-structured interview to obtain subjective feedback and comments about their experience.

5.3.4 Study design

We used a between-subjects design with a single dependent variable: The display size (Small, Large). Half of the participants undertook the study using a *small display* while the other half experienced the visualization environment on a *large display*. The small display condition utilized 3 of the 18 columns available in the CAVE2 environment, giving participants a resolution of 4,098 x 3,072 (12.5 Megapixels) and approximately a 40-degree field-of-view. The large display condition, on the other hand, utilized 13 columns giving participants a resolution of 17,758 x 3,072 (54.5 Megapixels) and approximately a 190-degree field-of-view. Figure 22 illustrates these two conditions.

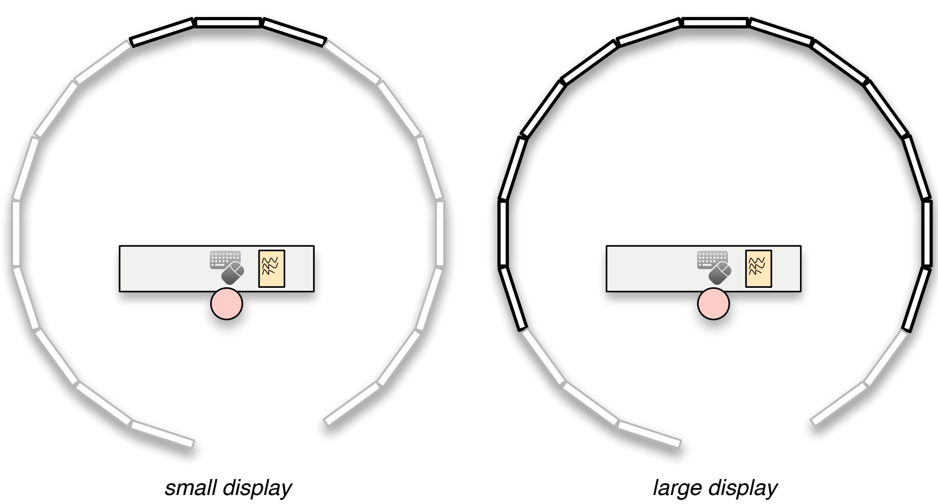


Figure 22. A comparison between the two experimental condition: The *small display* condition utilized 3 of the 18 columns in the *CAVE2* environments, giving participants a total resolution of 12.5 Megapixels providing an a 40-degree field of view. The *large display* condition utilized 13 columns with a total resolution of 54.5 Megapixels and a field-of-view of 190 degrees.

The two experimental conditions represented by the small and the large display provide a substantially similar visualization interface comprising identical visual representations and analytic operators. The only difference between the two conditions lies in the number of views that can fit on the display, which is influenced by the display size. This factor serves to implicitly modulates the amount of information a participant can simultaneously see and interact with during the exploratory activity.

The difference between the two experimental conditions, in terms of raw pixel count, is quite significant, with the larger display providing approximately 4 times the resolution and screen estate. However it is worthy to note that our small display provides 12.5 Megapixels, outweighing the average resolution of high-end desktop monitors, which are typically limited to 4 Megapixels. Thus, while we do anticipate an advantage to the large display, we do not believe the small display condition to be inferior.

5.3.5 Verbal protocol encoding

Our main focus in this analysis was on the video and audio data, which contained a record of the participant's verbal protocol as well as the state of the visualization environment at the time. Due to the close correspondence between the verbal protocol and the state of the visualization, we decided to encode the verbalizations directly from the video. While Ericsson and Simon's original methodology includes transcription and segmentation of the verbal data as two pre-processing steps (Ericsson and Simon, 1993), omitting these two steps is not unusual (Chi, 1997), particularly when protocol segmentation can be done more effectively from the video stream.

To encode the verbal protocol, we developed a coding scheme inspired from insight-based evaluation methodologies (Saraiya et al., 2005; North, 2006). As our goal is to understand the effects of increasing the display size on insight acquisition and user exploratory behavior, we were interested in two main themes: *insights* and *exploratory goals*. Insights comprise units of knowledge acquired by participants while interacting with the visualization. They comprise *observations* as well as *hypotheses* (Saraiya et al., 2005), both of which are relatively straightforward to recognize from the verbal protocol. Exploratory goals, on the other hand, are verbalizations that reflect objectives participants set for themselves during visual exploration. Because goal formation is the first step in Norman's seven stages of actions as well as Lam's framework of interaction costs (see section 2.2.3.2), coding for goals would enable us to measure the impact of display size on top-down exploratory behavior (e.g., hypothesis-driven inquiries). Conversely, measuring insights would also allow us to quantify the impact of display size on bottom-up analysis processes (i.e., from visual features to insights).

The coding scheme was refined over three coding cycles and finally consisted of the following set of codes (see Table III for example coded verbal statements):

- *Observation:* A unit of knowledge acquired from looking at and interacting with the visualization.
- *Hypothesis:* A conjuncture made by the participant, usually as a result of making a series of observations.
- *Question:* A statement reflecting an inquiry about the data.

- *Goal:* A statement reflecting the formation of an exploratory objective. Questions and goals are functionally similar in the sense that they both serve to orient the exploration process. However, we found that objectives, when articulated, were more indicative of forthcoming exploratory actions, and thus where coded separately from questions.
- *Physical navigation:* Represents an instance when the participant stood and walked up to and/or moved in front of the display.
- *Comment:* Represents a generic verbal statement made by the participant, usually about some aspect of the visualization environment or the user interface.

5.3.6 Level-of-insight metric

In addition to the above coding scheme, we assigned a *level-of-insight* (LOI) score to each observation or hypothesis based on its "broadness" (or "narrowness"), on a scale of 1 to 5. The goal of this score, which was inspired by the work of Shupp et al. (Shupp et al., 2009), is to measure the effects of using a larger display on the *quality of insights* formed, and allow us to categorically differentiate between potentially distinct types of insight induced by our two experimental conditions. In this scheme, higher LOI scores are indicative of higher-level, integrative insights which tie together multiple aspects and/or dimensions of the information space.

For instance, level-1 observations refer to isolated features found by the participant in one view (and hence are associated with a single year, for one type of crime, in a particular region or neighborhood in the city). Level-2, on the other hand, is ascribed to insights that are derived

TABLE III

Examples of coded statements from the verbal protocol.

Code	Example statement
Observation	"In the Loop, the crime pattern is high on Monday, high on Sunday, low through the week. This is a... There's a dip from Tuesday, Wednesday, Thursday, Friday, Saturday, but it's high on Sunday and Monday."
Hypothesis	"Maybe there was a drop in supply [in reference to a noticeable dip in narcotics-related crimes in 2012]. Because... I mean there's a lot of things that go into the narcotics trade; it's not just a local phenomenon, it's like a global thing."
Question	"Why is the peak crime time at night has changed [to an earlier hour] across all crimes?"
Goal	"So I'm gonna go and investigate, why is it that in the earlier years, there were more crimes at one point in time happening in the afternoon."
Comment	"One thing I'd comment on is that it is hard for me to compare the trends in this whole region in the city to the trends in the, um... the little blocks that I've perched out, because it's so small... It's such a small region. I don't know if you'd need two different scales, or that would be confusing."

from comparing two distinct level-1 features that may or may not be in the same view. For instance, a participant may find higher incidence of robberies in the near-north side of the city compared to the westside, in 2010. Level-3 scores refer to insights comprising temporal trends (over several years) or spatial patterns (involving multiple neighborhoods or regions). Level-4 scores take level-3 insights by implicating a second dimension. For instance, one may find vehicle thefts to be fairly constant over the years in the north-side of the city, but detect a decrease in the south-side during the same period. Lastly, level-5 insights integrate spatio-temporal patterns with observed variations or correlations between different types of crimes. For instance, based on an analysis of several years of crime trends, a participant may observe that homocides are correlated with narcotics-related crimes in the west-side, whereas the south-side exhibits a more pronounced correlation between weapon violations and homocides. Table IV details the criteria we used to assign the LOI scores and provides example statements from the verbal protocol for each level.

5.4 Qualitative observations

This section describes our general observations on how participants utilized the provided environment to explore the Chicago crime dataset. We describe the high-level analytic behavior of participants, highlighting similarities and variations in strategy between the two display conditions. Quantitative analysis results are presented in section 5.5.

5.4.1 Gaining an overview

At the beginning of the study, participants exhibited remarkable similarity in how they approached the task. All participants started by attempting to gain an "overview" of the dataset,

TABLE IV

Criteria for assigning a *level-of-insight* (LOI) score to each observation and hypothesis reported by participants along with exemplifying verbal statements. Insights are scored on a 5-points scale, with higher scores implying broader and more integrative insights.

LOI score	Criterion	Example statement
1	Refers to isolated features in a single year, for one type of crime, in a particular neighborhood/region.	"I can see that there are a lot of non-serious crimes, which is kind of expected, for downtown Chicago."
2	Refers to comparative insights (variation or correlation) between two level-1 features.	"In this time period [2012], [there are] more weapons violations near the lake [compared to inland areas]."
3	Refers to insights about spatial patterns (across neighborhoods/regions) or temporal trends (across years).	"Across all years it's the same seasonal trends [in narcotics-related crimes], I think, of peaking in the summer and then dropping off in the fall."
4	Refers to level-3 insights combining one additional dimension (e.g., temporal trend with observed variations/correlations in two or more neighborhoods/regions).	"The spikes [in burglaries], they seem to be consistent [across the years], form what I see. Especially... It looked like there's a pattern. Look here, this, this, and this. It's exactly the same. If you compare this one [burglaries in the northside], with this one [burglaries in the southside], same thing."
5	Refers to insights combining spatio-temporal patterns with observed variations/correlations in two or more crime types.	"A lot of people in this region [in the near northside] are buying drugs, but they're not fighting. There's no gang activity. There are much fewer deaths resulting from the narcotics trade. Where as disproportionally in the southern regions there are more narcotics related homocides."

which was typically done with the aid of one or two views created in the freeform layout. Due to the peculiar geographical aspect ratio of Chicago (approximately 1:2), a single view does not cover the entire height of the city, making in necessary to pan the view up and down. Most participants chose to use a single view, setting it to visualize all crimes in the most recent year (2012 in our case). However, few participants utilized two views at this stage to compare crime patterns in two years (e.g., 2012 and 2009), or in two separate regions (e.g., north and south). As an example, Figure 23 illustrates the dual-view setup created by participant L1.

Figure 23. Participant L1 utilized two disjoint views to get an initial overview of the Chicago crime dataset. The two views were set to visualize crime in 2006 and 2012, enabling him to make initial assessment of changes in crime patterns. To cover the entire geographical extent of the city, the participant panned the two views independently.

During the overview phase, which typically lasted for about 10–20 minutes, participants frequently extracted salient 'anchor' features from the heatmap (which were typically scored as level-1 or level-2 insights). For instance, participants often reported obvious hotspots indicating high concentration of crimes in the downtown area, the west side, and the far northeast side. This, the overview phase can be tied to the formation of a *qualitative mental model* which would later be used by participants to orient the exploratory activity (Gregory Trafton et al., 2000).

5.4.2 From overview to exploration

After the overview phase, participants proceeded to explore the dataset in more detail. For the large display participants, the end of the overview phase was typically marked by a transition from the freeform to the automated layout mode, with the instantiation of one or two view chains that typically wrapped around the entire environment. Figure 24 shows the view setup created by participant L4 immediately after she had completed the initial assessment of the dataset. Interestingly, this abrupt transition from overviewing the data to what appeared to be a goal-directed exploratory phase was not observed with the small display. Compared to the large display participants who quickly adopted the automated layout, most of the small display participants continued using their initial overview layout with some 'tweaking'. For instance, Figure 25 illustrates the layout created by participant S1, with views slightly extended to make use of the additional screen space, causing some overlap with the overview map.

This preference for the manual layout is justifiable considering that it gave the small display participants more flexibility in utilizing the available screen estate "down to the last pixel", albeit often at the cost of introducing visual clutter in the interface. Alternatively, participants may have seen little value in an automated layout on the small display. Participants with

Figure 24. The post-overview layout created by participant L4 comprised two temporally-synched view chains that show density for multiple types of crimes in the central-southern region of the city in 2012. While the two chains spanned 13 display columns, the participant limited her attention to the 7 central columns of CAVE2, focusing the two chains on the same area.

Figure 25. Participant S1 adopted this dual-view layout in the overview as well as the exploration phase.

the large display, on the other hand, quickly learned to utilize both layouts, switching to the freeform mode when they felt the need to see an overview of the dataset, and switching back to the automated mode to conduct goal-directed exploration.

5.4.3 Exploration phase

The exploration phase showed great diversity in strategies between individual participants. As one would expect, the activity comprised undirected search for 'interesting' variations, correlations, or outliers in the heatmap and/or the timeline charts as well as goal-driven exploration. Despite the wide variations in strategy and exploration styles, we observed a number of consistent differences between the two display conditions.

5.4.3.1 Organizational strategies

Generally speaking, the layout of the workspace in the large display reflected an organization that was suitable for a particular exploratory task. For instance, Figure 26 illustrate the setup created by participant L5 to explore the distribution of narcotics-related crimes relative to the overall crime patterns. This organization scheme reflected L5's desire to investigate a sudden drop in narcotics-related crimes in 2012 compared to other years, based on the hypothesis that there was a drop in the supply of illicit drugs. Overall, the large display participants seem to extensively rely on the automated layout to line up views in a manner that would facilitate search and comparison, making use of the bezels as natural dividers. Often, a change in the exploratory goal was accompanied by a significant change in the layout.

Participants on the small display also adopted specific organizational layout in response to specific exploratory goals they set for themselves. For example, participant S2 utilized 4

Figure 26. Participant L5 utilized four view chains to survey trends in narcotics-related crimes (top view chains) compared to the overall crime pattern (bottom chains) between 2007 and 2012. This view setup also covered two regions comprising near-north (right) and near-south (left) neighborhoods.

independent views to observe changes in crimes patterns in the west side between 2009 and 2012 (see Figure 27). However, such goal-specific organizational strategies were observed less frequently among the small display participants. Instead, we see a tendency in this group to opportunistically search for 'interesting' features in the heatmap and temporal charts, based on the current layout scheme. Often, participants resorted to 'recycling' the current layout instead of rearranging views or creating new ones. This bias to leverage existing resources is understandable, given the limited screen space available on the small display. However, it may have inadvertently 'locked' participants into existing visualization and goal states, preventing them form formulating new exploratory goals.

Figure 27. Participant S2 utilizing 4 independent views to explore changes in crime trends in the west side between 2009 and 2012.

5.4.3.2 Cognitive integration

A key difference between the two groups lies in how much time they allocated to explore each visualization state. Participants with the large display seemed to allocate a considerable amount of time to explore the current state of the visualization before changing it. During this time, participants seemed to iteratively generate insight by relating newly observed patterns to previously discovered ones (which would have still been visible on the display). As additional information is gathered, observations were incrementally refined and generalized. To help with integration across a large number of views, participants often utilized the bounding boxes to highlight patterns of interest (see Figure 28 for an example), seemingly using those to help them manage their attention in the environment. In contrast, we observed considerable 'paging' behavior on the smaller display with participants rapidly cycling through the years and/or the different crime types, leading to rapid changes in the state of the visualization (i.e., temporal view separation).

Figure 28. Participant L5 analyzing the evolution of crime trends in one region in north-side of the city (consisting of portions of Wrigleyville and Uptown). Four view chains illustrate changes in vehicle thefts, serious, non-serious, and all crimes between 2007 and 2012. Note the use of the bounding boxes to highlight relevant features in the heatmap. To see the 'big picture', the participant stepped away from the desk and stood at the entrance of CAVE2, taking few minutes to observe the entire layout. During this time, the participant was able to develop a number of level-4 and level-5 insights on how the distribution of crimes is affected by proximity to the lake.

To deal with this constant paging, participants developed two general strategies. The first strategy was to reduce the frequency of paging by simply ignoring one dimension in the information space. For instance, participant S3 focused his analysis primarily on the most recent year (2012 in our dataset). In the few instances where he switched the views to a different year, he seemed to be motivated to learn about a specific area.

The second strategy involved summarizing observations into verbal descriptions and noting them down on the paper notepad. For example, to understand changes in crime patterns in the south side, participant S2 resorted to performing pairwise comparison two between views (showing distribution of weapon violations and homocides) while writing down his observations on the notepad before switching to a different year. After cycling through the all years between 2006 and 2012, he went back to his summary and concluded that homocides are becoming more diffused whereas weapon violations are increasing. This strategy aided participants in integrating patterns observed in multiple temporally-separated visualization states. Yet, it also appeared to result in significant overhead as some participants seem to eventually abandon it, and is thus likely to be less effective compared to the ability to see multiple visualization states side-by-side afforded by the large display. Additionally, it may be difficult to distill down complex pattern into concise verbal descriptions (though we did not analyze the contents of the written notes generated by participants in this study).

5.5 Quantitative results

We first look at the length of the exploratory activity, and then discuss the quantity and rate of insights reported for observations and hypotheses, separately. Following that, we describe variations in the strategy and behavior of participants by analyzing the interaction patterns under the two display conditions.

5.5.1 Length of exploratory activity

The length of the exploratory activity comprises the time period spent by the participants actively exploring the dataset. Recall that the maximum time for exploration was 150 minutes (2.5 hours). However, participants were free to end this activity earlier, with most participants choosing to terminate the activity before the 150 minutes of allotted time had elapsed. Figure 29 illustrates average exploration times under the two display conditions.

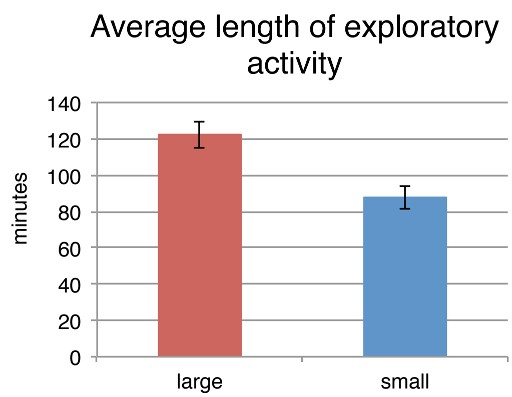


Figure 29. Average length of the exploratory activity under two display conditions (Large, Small). Error bars represent the standard error.

We found a significant effect for display size on exploration time ($two - samplet(8) = 3.551, p < .01$), with participants choosing to spend an additional 35 minutes on average with the large display (*mean* = 122.5, *SD* = 17.9) compared to the small display (*mean* = 87.3, *SD* = 15.7).

5.5.2 Reported observations

To determine the effect of increasing the display size on the generation of insights, we analyzed the number of observations reported by participants during the exploratory activity. We also looked at the distribution of observations according to our LOI metric. Figure 30 illustrates the average number of reported observations for each LOI score as well as the combined total.

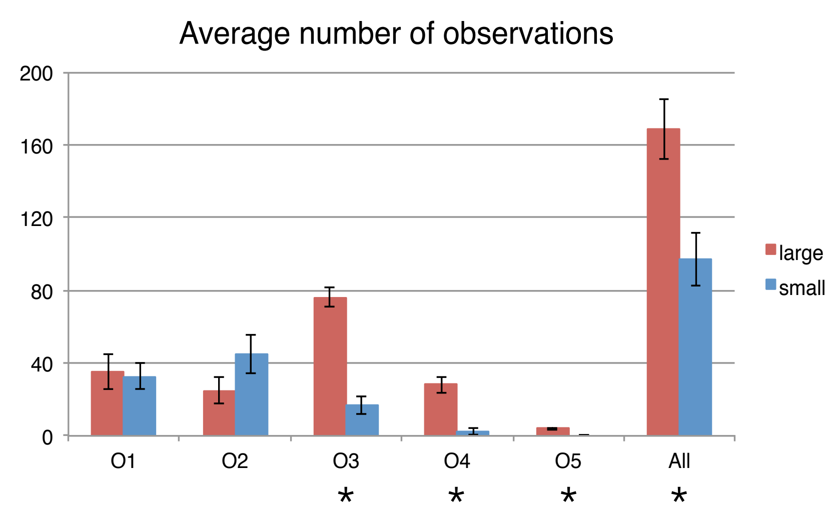


Figure 30. Average number of observations reported by participants under two display conditions (Large, Small). In addition to the combined number of observations (first column from the right), we show the distribution of observations under our LOI metric (*O1* through *O5*). Error bars represent the standard error. An asterisk denotes a significant difference between the two display conditions ($p < .05$, Bonferroni corrected for 5 tests).

There was a significant effect for display size on the compounded number of observations reported by participants ($t(8) = 3.23, p < .05$), with the large display eliciting approximately 74% more observations on average. We performed chi-squared test with Yates' correction to examine the relation between the display size and the distribution of observations. The display size had a significant effect on the distribution of observations ($\chi^2(4, N = 1327) = 263.37, p < .001$), with the large display inducing higher LOI scores in general.

We also performed post-hoc analysis to further investigate the effects of display size on observations for each LOI individually. Bonferroni correction was applied to maintain a significance level of $p < .05$ to each of the 5 comparisons tests. The was also a significant effect for display size on level-3 observations, which comprise insights about temporal and/or spatial trends ($t(8) = 8.161, p < .001$), with the large display eliciting approximately 3 times more observations on average. Similarly, there were significantly more level-4 and level-5 observations reported with the large display ($t(8) = 5.475, p < .001$ and $t(8) = 6.324, p < .001$, respectively). The numbers of level-1 and level-2 observations reported, on the other hand, did not seem to be affected by the display size ($t(8) = .217, p = .83$ and $t(8) = 1.552, p = .15$, respectively). Interestingly, none of the participants on the small display condition reported level-5 observations. We take these results as evidence that the large display supports the generation of more complex, more integrative insights compared to the small display.

5.5.3 Observation rates

While the quantity of observations appears to be higher with the large display, the above results do not take into account the fact that participants spent longer times exploring the dataset on the large display. Thus, we calculated a rate of observations, which normalizes the number of observations for each participant by the time it took him/her to complete the activity. Figure 31 illustrates these rates (in units of *observation per minute of analysis*).

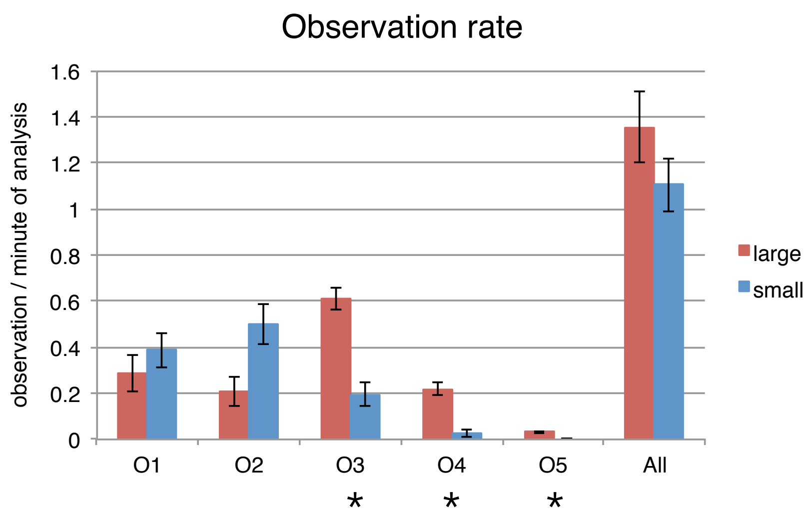


Figure 31. Observation rates under two display size conditions. Error bars represent the standard error. Asterisks denote significant differences between the two conditions ($p < .05$, Bonferroni corrected for 5 tests).

After normalization, we do not find significant difference in the combined rate of observations between the two displays ($t(8) = 1.294, p = .23$). However, post-hoc analysis with Bonferroni correction finds significant differences when separately comparing the rate of individual LOIs. The rates of level-3 and level-4 insights were significantly higher with the large display ($t(8) = 5.96, p < .001$ and $t(8) = 5.597, p < .001$, respectively). The average rate of level-5 was a mere $.03$ observations per minute with the large display, compared to *nil* with the small display ($t(8) = 7.466, p < .0001$).

We summarize the overall results on the effects of *increasing* the display size on the quantity and rate of observations in Table V.

TABLE V

Summary of results on the effects of *increasing* the display size on the numbers and rates of observations.

Measure	LOI	Difference in means ($\overline{X}_{Large} - \overline{X}_{Small}$)	Effect size (*Cohen's d*)	t-statistic	*p*-value
Number of observations reported	1	02.60	0.15	0.21	$p = .83$
	2	-19.80	-0.96	1.55	$p = .15$
	3	59.40	1.89	8.16	$p < .0001$
	4	25.60	1.77	5.47	$p < .001$
	5	04.00	1.82	6.32	$p < .001$
	All	71.80	1.50	3.23	$p < .05$
Rate of observation per minute of analysis	1	-00.09	-0.60	0.90	$p = .39$
	2	-00.29	-1.38	2.71	$p < .05$
	3	00.41	1.80	5.96	$p < .001$
	4	00.19	1.78	5.59	$p < .001$
	5	00.03	1.87	7.46	$p < .0001$
	All	00.25	0.83	1.29	$p = .23$

5.5.4 Hypothesis formulation

Next, we looked at the numbers and rates of hypotheses formulated during the activity. Figure 32 illustrate the average distribution of hypotheses formulated under our 5-points LOI scale, in addition to the combined average.

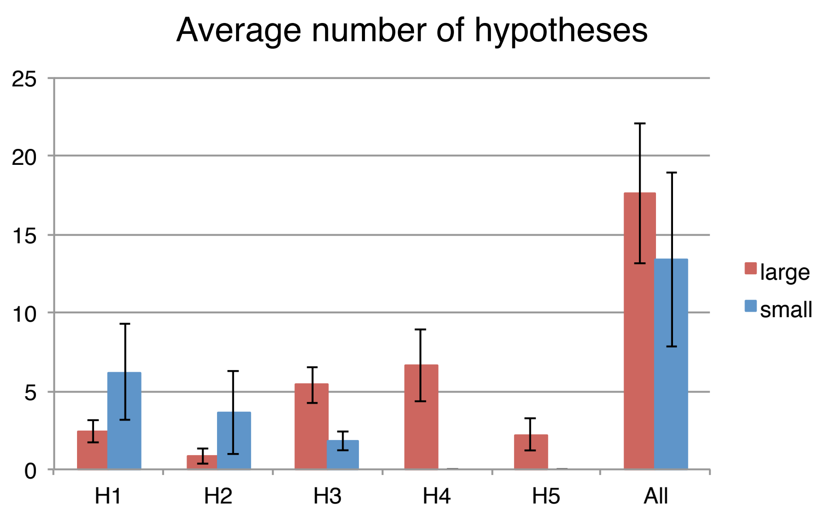


Figure 32. Average number of hypotheses formulated by participants under two display size conditions. In addition to the combined number of hypotheses (first column from the right), we show the distribution of hypotheses under our LOI metric (*H1* through *H5*). Error bars represent the standard error.

In contrast to reported observations, we do not find a significant effect for display size on the combined number of hypotheses ($t(8) = .587, p = .57$). A chi-squared test with Yates' correction, however, indicated a significant effect on the distribution of LOI scores ($\chi^2(4, N =$

$145) = 67.38, p < .001$). The large display appeared to skew the distribution towards higher LOI scores, with the distribution seemingly centered around level-4. In the small display, on the other hand, hypotheses seem to be distributed between level-1 and level-3. Interestingly, none of the small display participants formulated level-4 or level-5 hypotheses. Post-hoc analysis with Bonferroni correction finds no further significant differences between the two display conditions.

It is worthy to point out that the between-subjects variance in the quantity of reported hypotheses was quite large: ($\overline{X}_{hyp} = 17.6, SD_{hyp} = 10.0$) for the large display and ($\overline{X}_{hyp} = 13.4, SD_{hyp} = 12.4$) for the small display. In comparison, there was less between-subject variation in the number of reported observations: ($\overline{X}_{obs} = 168.6, SD_{obs} = 37.6$) for the large display and ($\overline{X}_{obs} = 96.8, SD_{obs} = 32.4$) for the small display.

5.5.5 Hypothesis formulation rates

We also calculate the rates of hypothesis formulation, normalizing the number hypotheses by the duration of the exploratory activity for each participant. Figure 33 illustrates the these rates (in units of *hypothesis per minute of analysis*). The combined rate of hypotheses formulation was not affected by the display size ($t(8) = .215, p = .83$). Post-hoc analysis finds no significant effect for the display size on the individual rates.

We summarize results on the effects of *increasing* the size of the display on the numbers and rates of hypotheses in Table VI.

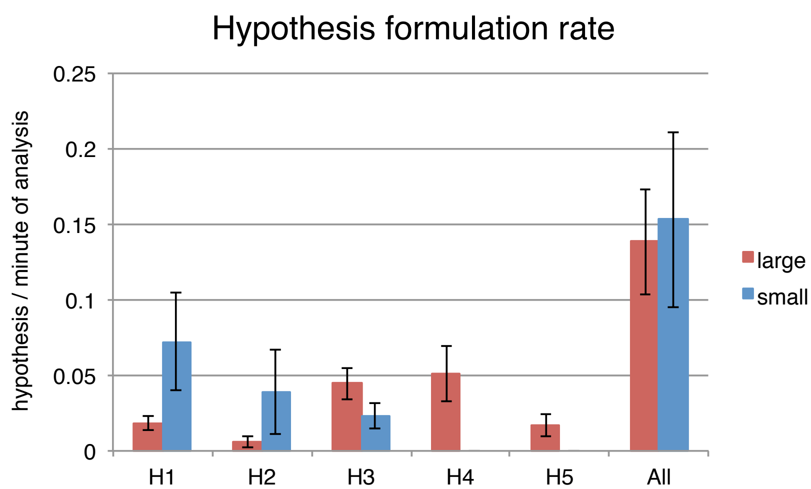


Figure 33. Hypothesis formulation rates under two display size conditions. Error bars represent the standard error.

5.5.6 Interaction pattern and analytic behavior

To understand the relation between the display size and participants' analytic behavior and interaction pattern, we constructed transition diagrams to characterize and quantify the flow between mental states identified from the verbal protocol and epistemic actions. Recall that epistemic actions comprises actions that a user takes to change the state of the visualization with the goal of acquiring additional information (Ware et al., 2013).

The transition diagrams comprised two epistemic states: *Brush, link, pan map*, which represent layout-preserving actions, and *Modify layout*, which represent layout-disruptive actions resulting in major change to the visualization state, potentially requiring the participant to rebuild his/her mental map. Examples of layout-disruptive actions include switching between

TABLE VI

Summary of results on the effects of *increasing* the display size on the numbers and rates of hypotheses formulated during the exploratory activity.

Measure	LOI	Difference in means ($\overline{X}_{Large} - \overline{X}_{Small}$)	Effect size (*Cohen's d*)	t-statistic	p-value
Number of hypotheses formulated	1	-3.80	-0.79	1.220	$p = .25$
	2	-2.80	-0.70	1.051	$p = .32$
	3	3.60	1.39	2.761	$p < .05$
	4	6.60	1.42	2.877	$p < .05$
	5	2.20	1.21	2.157	$p = .06$
	All	4.20	0.40	0.587	$p = .57$
Rate of hypothesis formulation per minute of analysis	1	-0.05	-1.00	1.639	$p = .13$
	2	-0.03	-0.76	1.172	$p = .27$
	3	0.02	0.96	1.558	$p = .15$
	4	0.05	1.40	2.787	$p < .05$
	5	0.01	1.23	2.212	$p = .05$
	All	-0.01	-0.15	0.215	$p = .83$

views and changing the contents of one or more view. Additionally, we included three primary mental states in the transition diagrams: *Make observation*, *Form goal*, and *Formulate hypothesis*, which are collectively responsible for insight acquisition. We constructed a transition diagram for each participant.

Formally, the transition diagrams model the exploratory analysis activity as a *Markov chain* process (Norris, 1998) involving probabilistic transitions between mental states (i.e., mental processing) and interaction states (i.e., instances when processing is offloaded onto the visualization tool). Aside from their formal properties, these diagrammatic representations provide

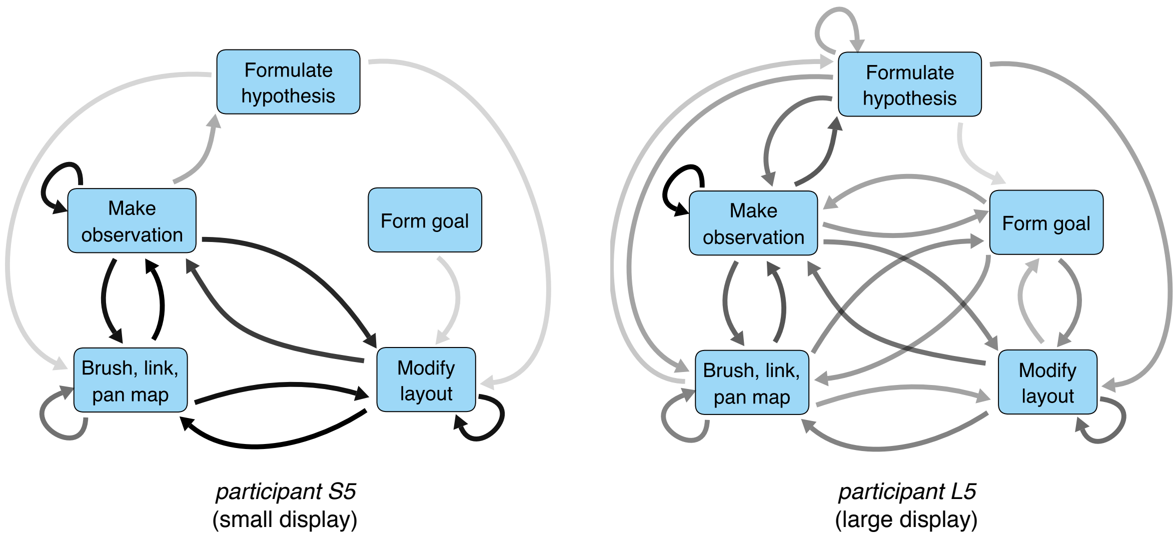


Figure 34. Two state transition diagrams illustrating differences in strategy between *participant S5* who used the small display to undertake the exploratory task (left) and *participant L5*, who utilized the large display (right). The weight of edges represent transition probability between two states (with log transformation applied). Thus, darker arrows represent more likely transitions.

us with a convenient way of capturing variations in the analytic behavior of participants. First, they visually illustrate micro-level differences in strategy between participants. Second, they enable us to characterize the impact of display size on cognition at a relatively fine-grained temporal scale by quantitatively analyzing the probability of moving between the different mental states and epistemic actions during the exploratory activity.

By way of example, Figure 34 illustrates the state transition diagrams for participants S5 and L5. We can see that S5 had to perform an extensive amount of layout-disruptive operations on the small display compared to L5. Moreover, we can see more transitions to the goal and hypothesis formulation states in L5's diagram, which suggests that the participant was able to devise and follow up on a larger number of exploratory goals.

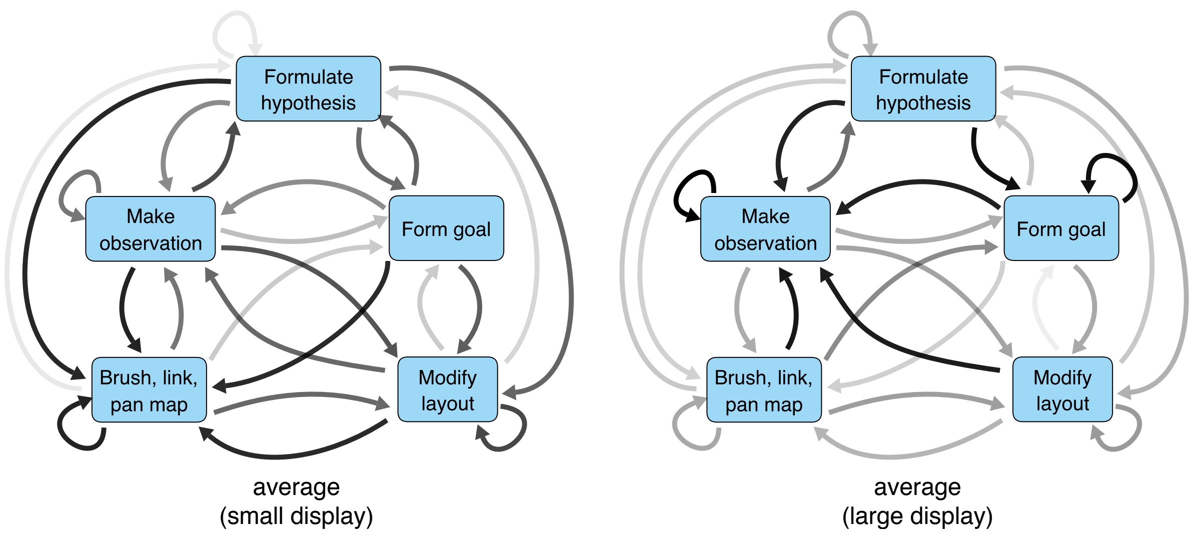
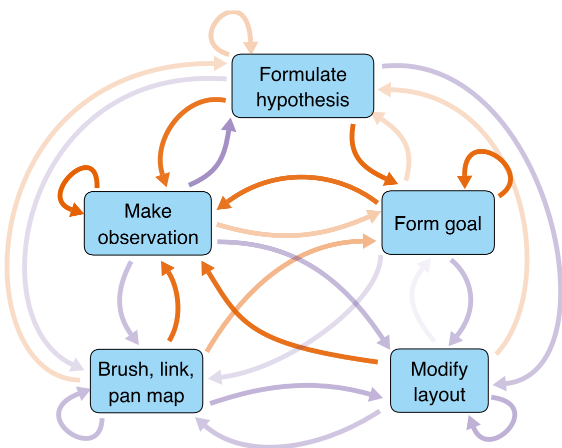


Figure 35. Two transition diagrams illustrating 'average' behavior of participants under the small (left) and large (right) display conditions.

We also 'averaged' the individual diagrams for participants under the same condition, giving us two average diagrams corresponding to the small and large display. To guarantee equal contributions from each participant to the average diagrams, we normalized transition frequencies by the time it took a participant to complete the activity. Figure 35 shows the two average transition diagrams side-by-side, illustrating important differences in the overall behavior of participants under the two display conditions. Figure 36 highlights these variations



	Modify layout (1)	Brush, link, pan map (2)	Form goal (3)	Make observation (4)	Formulate hypothesis (5)	
Modify layout	-51%	-70%	-27%	48%	12%	1
Brush, link, pan map	-37%	-68%	43%	56%	25%	2
Form goal	-48%	-76%	100%	61%	73%	3
Make observation	-50%	-69%	13%	**64%**	-26%	4
Formulate hypothesis	-48%	-77%	54%	60%	36%	5
	1	2	3	4	5	

Figure 36. A transition diagram illustrating the effect of *increasing* the display size on the behavior of participants (left). The weight of an edge represent difference in the transition probability between the large and the small display ($P(a,b) = P_{large}(a,b) - P_{small}(a,b)$). Edges that have been strengthened are color coded with orange whereas weakened edges are color coded with purple. The adjacency matrix on the right shows percentage changes in transition probabilities, with outlined cells indicating significant differences between the two displays conditions ($p < .05$, Bonferroni corrected for 25 tests).

with a 'difference' diagram and transition matrix, showing the relative changes in transition probabilities as a result of *increasing* the display size.

The large display diagram is marked by a decrease in the transition to the *Modify layout* state (column 1 of the transition matrix in Figure 36), indicating that participants were less likely to initiate layout-disruptive epistemic actions on the large display. We also see decreased transition probability to the *Brush, link, pan map* state, indicating that participants were also less likely to initiate brushing-and-linking and map panning operations (column 2 of the transition matrix). However, generally, we see an increased tendency for participants to transition from epistemic actions to insight-generating mental states with the large display (columns 3, 4, and 5). Furthermore, we see an increased likelihood to stay in these states on the large display (represented by elements (3,3), (4,4), and (5,5) in the adjacency matrix). Indeed, post-hoc analysis indicates a significant increase in the probability of remaining in the *Make observation* state ($t(8) = 4.995, p < .001$). Overall, these results suggest that the large display was more effective at eliciting insights and keeping participants in the 'cognitive zone' (Green et al., 2009), where they are likely to continue to generate additional insights.

5.6 Discussion

The results partially confirm our first hypothesis, which predicts an increase in the quantity and rate of observations reported by participants with the large display. As for our second hypothesis, which predicts a similar increase in the quantity and rate of hypotheses with the large display, we do not find evidence to support it. Lastly, our results indicate a significant tendency for participants to acquire higher-level, more integrative insights with the large display, confirming our third hypothesis. We discuss these results and then reexamine the analysis of interaction patterns and user behavior.

5.6.1 Quantity and rate of insight

Our results show a significant increase in the quantity of observations reported by participants during visual exploration, given a larger display with more pixels. Participants generated 74% more observations on average with the large display, partially confirming our first hypothesis. Given the disparity in the time spent by participants on the activity under the two display conditions, there are two potential explanations for this increase.

The first possible explanation is that the larger display afforded participants the ability to see more ‘interesting’ features, by enabling them to perform the bulk of their analysis using mental map-preserving interactions (e.g. brushing-and-linking) and physical embodied actions (eye movements and head turns). The large display participants were also able to juxtapose more views (representing crime patterns in different years, in multiple areas of the city, and/or distinct types of crime), which may have helped them in perceiving and interpreting complex crime patterns. Participants who used the smaller display, on the other hand, had to perform a greater amount of layout-disruptive actions, including cycling between multiple years or crimes types and rearranging the workspace to keep important views on top. The efficiency of being able to retrieve information with embodied actions may have induced the large display participants to revisit information more often, ultimately leading to more inquiries and insights. This explanation is supported by our qualitative observations which suggest a tendency for participants to thoroughly investigate the current state of the visualization and integrate information from multiple views on the large display (see section 5.4.3.2). It is also supported by previous studies which suggest a tendency for users to revisit information more frequently in LHD environments (Andrews and North, 2013).

The second possibility is that both display conditions afforded an equal opportunity for participants to generate meaningful insights into crime patterns, but participants on the large display benefited from a longer exploration time. Recall that participants spent an additional 35 minutes on average performing the exploratory task with the large display. Indeed, our analysis of observation *rates* does not find a significant difference between the two conditions.

Although we see a slight trend suggesting a possible increase in the rate of observation with the large display (see Figure 31), additional data is needed to quantitatively verify this trend. Hence, we cannot rule out the possibility that observed differences in the cumulative number of observations were simply due to participants spending a longer amount of time exploring the Chicago crime dataset with the larger compared to the small display.

As for our second hypothesis, which predicts an increase in the number of hypotheses formulated by participants when given a larger display, we do not find sufficient evidence to support it. Although, again, we see a slight trend with an increase in the cumulative number of hypotheses formulated by the large display subjects (see Figure 32), these differences are not significant. Furthermore, the aggregate rate of hypothesis formulation is quite similar for all participants, regardless of which condition they were assigned to. Thus we do not find evidence for an effect for display size on hypothesis formulation.

The above conclusion stands in contrast with findings from the case study in chapter 4, which suggest a positive impact for LHDs on hypothesis formulation. One possible explanation lies in individual differences in problem solving style and analytic ability between participants, which could have weakened differences between the two conditions. Considering that hypothesis formulation is generally indicative of higher-order thinking (Bloom, 1956), this explanation remains plausible even when taking into account the large differences in the reported observations between the two conditions. A second plausible explanation is that our participants simply lacked the domain expertise necessary to interpret the data to a level that would have allowed them to formulate a sufficient number of hypotheses, thus preventing us from detecting a measurable difference between the two conditions. Compared to the case study in chapter 4, where the subject was a domain expert having had strong motivation and curiosity to explore her dataset, this study was carried out with a participant pool drawn from a population of primarily graduate students who had no prior experience in the analysis of crime patterns.

5.6.2 Quality of insight

Another key difference between the large and small display lies in distribution of insights according to our LOI metric (see section 5.3.6 for a definition). Figure 30 shows that the distribution of insights is seemingly centered around level-3. Conversely, the majority of insights reported on the small display were scored at level-1 or level-2. This provides evidence that the large display affords the acquisition of more complex and integrative insights. We see a similar trend in the distribution of hypotheses (see Figure 32), though these differences were not statistically significant.

This result can be explained by taking into account changes to the cost structure of visual exploration wrought by the increase in the display size and resolution (see section 2.2.3 for a theoretical discussion). An important cognitive component in exploratory analysis is sense making, which refers to the integration of disparate pieces of information into a coherent narrative (Pirolli and Card, 2005). Sense making benefits significantly from the *spatial distribution* of information artifacts, with people naturally taking advantage of space as resource to facilitate the assimilation of fragmentary information (Bradel et al., 2013). Similarly, the large display, by emphasizing the *spatial separation* of views, affords the integration of patterns found in different views, thus allowing people to make more complex inferences. We see evidence of this form of integration in our analysis of participants' behavior (see section 5.5.6). In particular, Figure 36 shows a tendency for participants to remain in insight-generating mental states (particularly, the *Make observation* state), where they are likely to integrate their observations and develop more complex insights about crime patterns without having to 'leave' this state (Green et al., 2009). Our qualitative observations also support this conjecture, with participants seemingly allocating more time to integrate information form multiple views, before changing the state of the visualization in the large display condition. Participants with the small displays, on the other hand, were more likely to cycle between views. This may have ultimately o decrease the probability of arriving at integrative insights, inducing participants to report their isolated observations, which represent the 'low hanging fruits'.

Conversely, from a top-down perspective, participants with the large display may be more likely to form ambitious exploratory goals, which could be perceived as too costly on the small display. We see evidence of this in Figure 36, which shows an increase in transitions to the *goal formation* state with the large display. We also see anecdotal evidence of this from interviewing participants who undertook the experiment on the small display. For instance, when participant S3 was asked why he focused his exploration on crime activity in 2012 and did not explore variations across different years, the participant responded: "I only bothered to look at the years when I knew something about an area– like Cabrini Green and the Taylor area". This suggests that the cost of exploring temporal trends in this case was inhibitory that the participant generally avoided it unless he knew a priori what attributes to focus on. In other words, in circumstances of weak information scent (Pirolli, 2007), the small display seem to skew participants' exploratory goals towards the acquisition of low-level, isolated insights. This behavior resembles 'narrowing in' behavior observed by Patterson et al. in their study of how intelligence analyst cope with information overload (Patterson et al., 2001), which caused some analysts to inadvertently miss highly-profitable pieces of information that lied beyond their search scope. From this perspective, a large display affords users a wider 'spotlight', enabling them to incorporate a larger variety of views in their analysis.

5.6.3 Cognitive engagement

Our study also raises the question of why participants 'choose' to devote significantly more time to the task, when given a larger display with more pixels. One possible explanation here is that participants were simply more 'engaged' in the analytical task, with the large display serving to sustain their attention for a longer amount of time.

The notion of engagement is widely cited in the HCI literature (usually in a nebulous manner) as being a positive quality that is often correlated with effective designs. However, engagement is seldom discussed in the circles of visual analytics, except when referring to *persuasive* and *casual* visualizations, which aim to provide rhetorically effective visual designs intended to induce personal reflection and/or some form of behavioral change (Fogg, 2002; Pousman et al., 2007). In the realm of visual analytics, we can define *cognitive engagement* as the "tendency towards investing or being concerned with a stimulus" (Peters et al., 2009). From this perspective, a larger display with more pixels may provide a more compelling visual stimulus that invites more interest from users. Alternatively, engagement can be conceptualized of as a process, where gaining insight is likely to drive one's interest in seeking further discoveries. With LHD environments inducing cognitive engagement, they are likely to improve the analytic performance of users in open-ended analysis tasks.

5.6.4 The cognitive costs of physically large displays

Although some participants were able to utilize the full size of the environment, by stepping way from the display for instance (see Figure 28), most participants in the large display condition appeared to limit their use of the CAVE2 environment to 6 or 7 columns at a time. Often, participants focused their attention on either the left or the right side of the environment. When asked about this, most participants indicated that it was difficult to compare pieces of information when located at different sides of the display. Participant L3, for instance, maintained that by the time he had turned his head to the opposite side and located the necessary information, he would have had 'forgotten' the pattern he was trying to remember. Conversely, participants with the small display appeared to have no problem in utilizing the entire size of the display.

This phenomenon suggests an increase in *physical motion costs* needed to integrate information located across spatially-separated views (see section 2.2.3), particularly if the views are separated by a relatively large visual angle. Our observations also suggests an upper limit on the amount of rotational motion users are willing to tolerate, which we estimate to be approximately 87–102 degrees of visual angle (corresponding to the field-of-view afforded by 6–7 columns in the CAVE2 environment, from its center).

There are also instances when the large display participants appeared to suffer from *information overload*. Participant L5 commented that it was often "so hard to compare so many things" at once. This may indicate an increase in the effort needed to visually resolve the visualization and manage one's attention, given a physically large display. This, in turn, could have reduced the utility of the large display in some cases, and may have ultimately diminished the differences between the two experimental conditions for some measures, such as the hypothesis formulation rates. Furthermore, this phenomenon could partly explain why the large display participants devoted more time to the activity, by virtue of the task being more physically and mentally demanding. It is worthy to note, however, that this additional investment (both in time and effort) translated into considerable advantage with respect to the analytical outcome.

These observations call for future research into interactions and designs that can further reduce the potential for information overload in LHD-based visualization environments. While some of the interactions provided by our visualization, such as the heatmap bounding boxes, do seem to help (as all participants utilized them at some point during the activity), they appear to be insufficient. This motivates the need for more effective perceptual techniques to help users manage their attention, and efficiently locate and highlight the relevant information in spatial visualization environments.

5.7 Conclusions and study limitations

This study addresses RQ_1. It demonstrates that increasing the physical size and resolution of the visualization interface positively affects user behavior during exploratory visual analysis. In particular, we see a tendency for users to allocate more time to exploring the visualization and integrating information contained in *spatially-separated views*, given a large display with more pixels. Moreover, from a micro-cognition perspective, we see an increased tendency for users to transition from epistemic actions (i.e., interactions) to insight-generating mental states. We attribute these key behavioral differences to changes in the cost structure of visual exploration wrought by the spatial separation of views promoted by LHDs, compared to the chronic *temporal-separation* of visualization states induced by conventional displays.

This study also addresses RQ_2, demonstrating that the above behavioral adaptations to LHDs result in a significantly improved analytical outcome, reflected by an increase in the quantity of insights reported. Furthermore, we see a significant tendency for users to develop higher-level, more integrative insights, when given a larger display with more pixels.

Lastly, this study addresses RQ_3 by providing a preliminary evaluation of the *seed and grow* design pattern. Our observations suggest that users were able to leverage the design concept of *view chains* to devise effective organizational strategies and collectively control views on a LHD in support of exploratory goals.

The main limitation in this study is that it relies on a novice participant pool drawn from the general population to undertake a complex analytical task. Participants had no prior experience in the analytical scenario posed by the study, which revolved around the analysis of crime patterns in a major metropolitan area. This may have impacted participants' level of motivation, selection of strategy, as well as their understanding of phenomenon depicted in the visualization. Therefore, we expect some of the effects measured in this study to change in real-world visual analytic applications. One important consideration is that a domain expert might be able to override some of the heuristic biases faced by out participants in the small display condition, and thus diminish the analytical advantage conferred by the large display.

Lastly, it is worthy to point out again that our participants had prior experience in using LHD environments within the context of general knowledge-based activities, such as conducting collaborative brain storming sessions and meetings. Consequently, the results of the experiment may have been slightly biased in the favor of this more experienced participant pool. It is therefore reasonable to expect a learning curve with users who are novice with the technology before they are able to achieve comparable results.

CHAPTER 6

CONCLUSION

As we enter a ‘big data’ era, our society is being confronted with staggering quantities of data that are being generated at ever escalating rates. Consequently, there is an increasing need for effective perceptual and cognitive aids to help us make sense of the available troves of digital information. Visualization represents one of the most effective ways for exploring and communicating large quantities of data. However, the scalability of visualizations has often been limited by the prevailing display technology. Thanks to advances in display technology, Large High-resolution Displays (LHDs) are becoming increasingly common, and beginning to take on an important role in providing data scientists and researchers with a visualization instrument for the exploration and analysis of large-scale datasets. Compared to conventional desktop and laptop displays, LHDs enable greater quantities of information be visualized at once, thus potentially alleviating the ‘tunnel vision’ phenomenon often experienced by people working with large data sources.

This dissertation has sough to provide an understanding of the analytical affordances of LHDs and their effects on scientific discovery in scenarios that involve the exploration of large and complex datasets. We have demonstrated, through exploratory and experimental studies, that increasing the physical size and resolution of the visualization interface can fundamentally impact user behavior and insight acquisition during exploratory visual analysis. Our results provide evidence to suggest a positive effect for the use of LHDs which manifests in a broader exploratory behavior along with the acquisition of higher-level, more integrative insights.

This chapter concludes the dissertation by summarizing our contributions and outlining future research directions suggested by this research.

6.1 Contributions

This dissertation was motivated by three research questions:

- RQ_1: What is the effect of increasing the size and resolution of the visualization interface on user behavior during exploratory visual analysis?
- RQ_2: Compared to conventional displays, how does the ability to simultaneously see and interact with orders of magnitude more information on LHDs affect insight acquisition?
- RQ_3: Are there new design patterns for scaling up multi-view-based visualization interfaces to LHD environments?

Following is a summary of our primary contributions along with a discussion of how they address the above research questions.

6.1.1 Theory of how interaction costs affect exploratory behavior

A primary concern of visual analytics is the design of effective visual interfaces that facilitate human *exploration* and *analytical reasoning* in circumstances involving large quantities of information. In this sense, a visual analytic interfaces does not passively supply information but rather serves to scaffold user activity, shaping the entire analytical cycle while providing key cognitive resources (e.g., memory). It is therefore crucial to understand how variations in the design of the interface affect user strategy and behavior. Based on the work of Lam (Lam, 2008) and Norman (Norman, 2002), we contributed a theory of how interaction costs affect user behavior in exploratory scenarios where user goals not defined a priori and the information scent is weak. Our theory postulates a 'feedback' effect associated with costly interactions which propagates back to influence decision making and goal formation processes, causing users to 'narrow in' and focus on the exploitation of isolated subsets of the information space. This theory, which provides us with a conceptual foundation to begin to address RQ_1 and RQ_2 empirically, also predicts that the spatial separation of views afforded by LHDs would alleviate some of these costs, thus encouraging a broader exploratory behavior and possibly the acquisition of more insights.

6.1.2 Visualization design patterns for LHD environments

LHDs presents visualization designers with opportunities to design more scalable visual analytic environments, enabling users to see and interact with orders of magnitude more information. However, existing design paradigms need to be readapted to this new environment. In chapter 3, we developed an extension to Ware's Visual Thinking Design Patterns (VTDP), extending four of his VTDPs to LHD environments. Our adaptations provide vertically integrated framework (based on empirically and theoretically motivated design heuristics) for constructing multi-view-based visualization interfaces for LHDs. This subset of VTDP is aimed at supporting tasks involving the comparative analysis of large, homogeneous information spaces. Additionally, our design patterns contribute two conceptually novel design ideas:

- *Scalable query-by-example technique:* Involves an extension of the traditional brushing-and-linking technique in conjunction with a small-multiples layout. It can be used to facilitate the comparative analysis of large ensemble datasets. Using a familiar brush metaphor, the user can specify a pattern of interest in one views and cause the system to automatically highlight similar patterns in other views (based on a domain-dependent similarity function), using a perceptually salient encode.
- *Loose coordination model:* An extension to the Coordinated Multiple Views (CMV) model, allowing for multiple levels of coordination between views. This extension allows for better utilization of the screen estate in LHD, giving users more control over the state of the visualization to support parallel exploration threads as well as the efficient storage and retrieval of previous exploratory outcomes.

Although the presented design patterns are by no means complete, they contribute design solutions to RQ_3, and provide practitioners and researchers with a starting point to begin to systematically investigate how LHDs reshape the design space of multi-view-based visualizations.

6.1.3 Impact of LHDs on user analytic behavior

A key contribution of this dissertation is to demonstrate important differences in user behavior wrought by a possible reduction in visual exploration costs as a result of increasing the size and resolution of the visualization interface. This difference is marked by a significant tendency for users to invest more time in the exploratory process. Furthermore, we have observed slight tendency for participants to form and pursue more exploratory goals, possibly increasing user engagement with the task and leading to a broader exploratory strategy. Lastly, our analysis of verbal protocols suggest a significant increase in users' tendency to transition to and 'remain' in *insight generating* mental state, when given a larger display with more pixels. These results suggest that the ability to utilize visual search as the primary mechanism for information foraging is likely to drive the exploratory process towards the formation of more ambitious exploratory goals, while keeping users in the 'cognitive zone' by avoiding unnecessary context-switching (Green et al., 2009).

The above results directly address RQ_2. While some of these effects are based on qualitative observations rather than significant quantitative differences, the gamut of behaviors exhibited by participants are consistent with our theoretical account. We believe that additional data would serve to quantitatively confirm these trends.

6.1.4 Impact of LHDs on insight acquisition and scientific discovery

Lastly, this dissertation has demonstrated that the use of LHDs as an interactive visualization medium can fundamentally impact the outcome of the analytical process, and the nature of insights acquired during visual exploration. As the study described in chapter 5 illustrates, increasing the physical size and resolution of the visualization interface is correlated with a significant increase in the quantity of observations reported by users. Furthermore, we see a significant tendency for users to develop broader, more integrative insights, when given access to a larger display with more pixels. These observed effects demonstrate a potential role for LHDs in supporting the cognitive processes implicated in scientific discovery, particularly in scenarios that involve the visual analysis of large datasets.

These results address RQ_3, but also raise questions about the common wisdom of championing minimalistic visualization interfaces so as to avoid distracting users and overloading them with too much information. While such long held wisdom reflects valid concerns, we would argue that it is often invoked prematurely.

6.2 Future research directions

This research has investigated the impact of using LHDs on user behavior and insight acquisition in exploratory visual analysis. While we do contribute design patterns that are suitable for constructing LHD-based visual interfaces, this work makes no attempt at defining or even constraining the design space. Having identified significant benefits to reducing the temporal-separation of views by increasing the size and resolution of the visualization interface, a natural next step is to begin to systematically investigate this design space. There have been some earlier attempts at sketching out some of the broad outlines for a design space for multi-view-based visualizations. For instance, Yost et al. propose that *view layout*, which characterize how views are physically located within the display, as one of the principle axes of such a design space (Yost et al., 2007). Some of the distinct design instances on this axis include *attribute-centric* layouts, which are commonly referred to as small-multiples, and *space-centric* layouts, which rely on embedding abstract information views in a larger spatial context (e.g., a geographical map) at a corresponding location.

Although the above definition of the design space addresses the issue of information layout within a large, spatial visualization environment, it does not make an attempt at constraining the possible range of interactions that can be realized with multi-view, LHDs-based visualizations. Other researchers attempted to address interaction design by leveraging the affordances of freeform layouts as a way to communicate semantics between the user and the visualization tool, by dynamically clustering related information artifacts (Andrews et al., 2010; Endert et al., 2012a).

Our *seed and grow* design pattern, which allows for loose coordination between views, can be construed as one point on an orthogonal axis, which we shall refer to as the *coordination model*. Tight coordination between views have always assumed to be essential in conventional multi-view interfaces. However, there is an inherent benefit to loosening coordination in LHD environments and restricting it to a subset of views. For instance, as our second study shows, uncoordinated views could function as independent 'lenses', allowing one to project and juxtapose two disparate parts of the information space side-by-side for comparison. They can also be used as non-volatile storage of earlier exploratory outcomes. Disjoint views could also provide scaffolds to enable users to explore and follow up on multiple narratives in parallel. We also found from informal comments of participants in the second study that they want to have some control over the coordination model. That is, users want to be able to dynamically specify which views are coordinated and which are not, so that they can adapt the visualization environment to their analytic process. The challenge here is to support dynamic arrangement and coordination while retaining interface usability.

Articulation of a coherent interaction design space for multi-view, LHD-based visualizations is still an open question (Andrews et al., 2011). The 'overview first, zoom and filter, then details-on-demand' have been the dominant interaction model for information visualizations on conventional displays (Shneiderman, 1996). However, the studies in this dissertation suggest that zooming and panning would subject users to excessive visual flow and temporal-view separation. Are there new interaction models that can better support visual exploration of large information spaces on LHD environments? And how should interactions be operationalized in a larger spatial environments with orders of magnitude more information? These are some of the open questions that warrant future research.

Another aspect that was not investigated in this research is collaboration between multiple users. One of the main reasons why people acquire LHD environments is to provide a platform that foster collaborative problem solving between collocated teams (Leigh et al., 2012; Reda et al., 2013b). While this dissertation has limited itself to studying the cognitive processes of a single person, it would be of interest to look at the exploratory visual analysis process when multiple collocated individual participate in the activity. Would we expect the analytical performance of a group of researchers or scientists to improve, when given access to a larger display with more pixels?

A collaborative LHD-based visualization interfaces would require distinctive set of interactions to support a wide range of synchronous and asynchronous processes within the activity. Previous research by Jagodic et al. have resulted in the design of an multi-modal input and interaction framework intended for use within the context of knowledge-based activities that fall under everyday usage scenarios (Jagodic et al., 2011). However, it is likely that there are unique requirements and solutions for data-intensive visualization applications (Isenberg et al., 2011). Collaborative visual analytics with LHD environments remains a widely-open field that is ripe with many opportunities and challenges.

6.3 Final remarks

The results of this research is a testament of human perceptual and cognitive abilities to deal with scale and complexity, when given appropriate technological support. We have argued that large high-resolution displays can fundamentally impact the analytic behavior of users and ultimately improve the outcome of the visual analytic process. However, much research lies ahead to fully understand the unique affordances of this platforms, and much work remains to be done to apply this research towards the development of scalable visual analytic environments to help us stay abreast with the rising tide of data.

APPENDICES

Appendix A

USER STUDY DOCUMENTS

UNIVERSITY OF ILLINOIS
AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Exemption Granted

February 17, 2014

Mhd Reda, MS
Computer Science
851 S. Morgan, Rm. 1120 SEO
M/C 152
Chicago, IL 60607
Phone: (312) 996-3002 / Fax: (312) 413-7585

RE: Research Protocol # 2014-0140
"Evaluating the Use of Large, High-Resolution Displays in Visual Data Analysis"

Sponsor(s): None

Dear Mhd Reda:

Your Claim of Exemption was reviewed on February 16, 2014 and it was determined that your research protocol meets the criteria for exemption as defined in the U. S. Department of Health and Human Services Regulations for the Protection of Human Subjects [(45 CFR 46.101(b)]. You may now begin your research.

Exemption Period:	**February 16, 2014 – February 16, 2017**
Performance Site:	UIC
Subject Population:	Adult (18+ years) subjects only
Number of Subjects:	30

The specific exemption category under 45 CFR 46.101(b) is:
(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

You are reminded that investigators whose research involving human subjects is determined to be exempt from the federal regulations for the protection of human subjects still have responsibilities for the ethical conduct of the research under state law and UIC policy. Please be aware of the following UIC policies and responsibilities for investigators:

1. Amendments You are responsible for reporting any amendments to your research protocol that may affect the determination of the exemption and may result in your research no longer being eligible for the exemption that has been granted.

2. Record Keeping You are responsible for maintaining a copy all research related records in a secure location in the event future verification is necessary, at a minimum these documents include: the research protocol, the claim of exemption application, all questionnaires, survey instruments, interview questions and/or data collection instruments associated with this research protocol, recruiting or advertising materials, any consent forms or information sheets given to subjects, or any other pertinent documents.

3. Final Report When you have completed work on your research protocol, you should submit a final report to the Office for Protection of Research Subjects (OPRS).

4. Information for Human Subjects UIC Policy requires investigators to provide information about the research protocol to subjects and to obtain their permission prior to their participating in the research. The information about the research protocol should be presented to subjects in writing or orally from a written script. When appropriate, the following information must be provided to all research subjects participating in exempt studies:
   a. The researchers affiliation; UIC, JBVMAC or other institutions,
   b. The purpose of the research,
   c. The extent of the subject's involvement and an explanation of the procedures to be followed,
   d. Whether the information being collected will be used for any purposes other than the proposed research,
   e. A description of the procedures to protect the privacy of subjects and the confidentiality of the research information and data,
   f. Description of any reasonable foreseeable risks,
   g. Description of anticipated benefit,
   h. A statement that participation is voluntary and subjects can refuse to participate or can stop at any time,
   i. A statement that the researcher is available to answer any questions that the subject may have and which includes the name and phone number of the investigator(s).
   j. A statement that the UIC IRB/OPRS or JBVMAC Patient Advocate Office is available if there are questions about subject's rights, which includes the appropriate phone numbers.

Please be sure to:
→Use your research protocol number (listed above) on any documents or correspondence with the IRB concerning your research protocol.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact me at (312) 355-2908 or the OPRS office at (312) 996-1711. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Charles W. Hoehne
Assistant Director
Office for the Protection of Research Subjects

cc: Peter C. Nelson, Computer Science, M/C 152
Andrew Johnson, Computer Science, M/C 154

Leave box empty - For office use only

University of Illinois at Chicago
Consent for Participation in Research
"Evaluating the Use of Large, High-Resolution Displays in Visual Data Analysis"

Why am I being asked?

You are being asked to be a subject in a research study about Large, High-Resolution displays conducted by Computer Science PhD candidate *MHD Khairi Reda* at the University of Illinois at Chicago. You have been asked to participate in the research because you are not a minor and may be eligible to participate. We ask that you read this form and ask any questions you may have before agreeing to be in the research.

Your participation in this research is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University or your grade in any UIC courses. If you decide to participate, you are free to withdraw at any time without affecting that relationship.

Why is this research being done?

Large, High-Resolution displays are becoming increasingly common. Scientists and researchers often use these displays to analyze large amounts of data, which would otherwise be difficult to do on traditional computer screens. It is important to develop techniques for creating effective visualization tools for such environments. The Electronic Visualization Laboratory at UIC has been a pioneer in the study of visualization interfaces for large displays. The development of better visualization tools has the potential to increase productivity and accelerate scientific discovery.

What is the purpose of this research?

The purpose of this research study is to understand the benefits and drawbacks associating with using Large, High-Resolution displays in the context of visual data analysis. Results form this study will allow us to devise guidelines for creating effective visualization tools in the future.

What procedures are involved?

If you agree to be in this research, we will ask you to do the following:

- You will undertake a simulated data analysis scenario, and use an interactive visualization tool to explore a dataset that will be given to you. The description of the scenario will be provided to you along with a list of questions about the data. Your task is to try to answer those questions. There are no 'right' and 'wrong' answers here. Rather, your answer should reflect your assessment of the data.

- Follow a short training on how to use the visualization tool.

- Use the visualization tool to explore the provided dataset and answer the questions. You may also follow up on any additional questions you may come up with during your analysis. The maximum durarion for this part of the study will be 2.5 hours.

- Undertake a short interview after you finish the task.

- During the study, you will be video and audio taped. We also ask that you try to keep talking aloud while thinking. Try to say anything that goes through your head.

- If you have questions during the study, please feel free to ask the researcher at any time. You may also take breaks at any time during the study.

Approximately 30 subjects may be involved in this research at the University of Illinois at Chicago.

What are the potential risks and discomforts?

The research has no foreseeable risks to you as a participant. Your information will remain confidential, and is not linked to your performance with the University of Illinois at Chicago.

Are there benefits to taking part in the research?

You will not receive any direct benefit for participating in this study. By participating in this research, however, you will also be contributing to the field of data visualization.

Will I be reimbursed for any of my expenses or paid for my participation in this research?

You will not be reimbursed for any expenses you incur by participating in this research nor will you be compensated in any way for your participation.

What are the costs for participating in this research?

There is no cost to you for participating in this study.

What about privacy and confidentiality?

The only people who will know that you are a research subject are members of the research team. No information about you, or provided by you during the research, will be disclosed to others without your written permission, except:

- when necessary to protect your rights or welfare (for example, if you are injured and need emergency care or when the UIC Institutional Review Board monitors the research or consent process); or

- when required by law.

When the results of the research are published or discussed in conferences, no information will be included that would reveal your identity. If photographs, videos, or audiotape recordings of you will be used for educational purposes, your identity will be protected or disguised. Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law.

- All identifying data, including images, recordings, and questionairs will be kept under lock and key at the Electronic Visualization Laboratory for the duration of the study. Access to this data will be restricted to the prinicipal investigator, *MHD Khairi Reda*, and his co-investigators.

- All other information that might identify you, such as response times, will be labeled with a numerical identifier to maintain your anonymity. The index of study participant names and numbers will be kept under lock in the Elecronic Visualization Laboratory.

- We will reguest additional consent from you if we desire to use identifying images or recordings of you in a publication or for public presentation.

- All identifying images, recordings, and questionnaire results will be destroyed once the data has been fully analyzed or within one year, whichever comes first.

What if I am injured as a result of my participation?

In the event of injury related to this research study, treatment will be made available through the University of Illinois at Chicago Hospital. However, you or your third party payer, if any, will be responsible for payment of this treatment. There is no compensation and/or payment for such medical treatment from the University of Illinois at Chicago for such injury, except as may be required of the University by law. If you feel you have been injured, you may contact the researcher, *MHD Khairi Reda* at the Electronic Visualization Laboratory, 312-996-3002.

Can I withdraw or be removed from the study?

You can choose whether to be in this study or not. If you volunteer to be in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you don't want to answer and still remain in the study. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

Who should I contact if I have questions?

The researcher conducting this study is *MHD Khairi Reda*. You may ask any questions you have now. If you have questions later, you may contact him at:
Phone: 312-996-3002 Email: mreda2@uic.edu

The faculty sponsor of this research is Associate Professor *Andrew E. Johnson*. You may contact him at:
Phone: 312-996-3002 Email: ajohnson@uic.edu

What are my rights as a research subject?

If you have any questions about your rights as a research subject, you may call the Office for Protection of Research Subjects at 312-996-1711.

What if I am a UIC student?

You may choose not to participate or to stop your participation in this research at any time. This will not affect your class standing or grades at UIC. The investigator may also end your participation in the research. If this happens, you class standing or grades will not be affected. You will not be offered or receive any special consideration if you participate in this research.

What if I am a UIC employee?

Your participation in this research is in no way a part of your university duties, and your refusal to participate will not in any way affect your employment with the university, or the benefits, privileges, or opportunities associated with your employment at UIC. You will not be offered or receive any special consideration if you participate in this research.

Remember: Your participation in this research is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University. If you decide to participate, you are free to withdraw at any time without affecting that relationship.
You will be given a copy of this form for your information and to keep for your records.

Signature of Subject or Legally Authorized Representative

I have read (or someone has read to me) the above information. I have been given an opportunity to ask questions and my questions have been answered to my satisfaction. I agree to participate in this research. I have been given a copy of this form.

Signature   Date

Printed Name

E-mail address

Signature of Researcher   Date (must be same as subject's)

University of Illinois at Chicago
Consent to Use Identifying Media from

"Evaluating the Use of Large, High-Resolution Displays in Visual Data Analysis"

Why am I being asked?

We would like to use images, video, or audio recordings of your participation in the study for publication or presentation. We seek your consent to use this media in unaltered form that may allow others to identify you. We ask that you read this form and ask any questions you may have before giving consent.

Your decision to give this consent is voluntary. Your decision whether or not to give consent will not affect your current or future relations with the University or your grade in any UIC courses.

What will this media be used for?

We wish to use images, video, or audio recordings that include your likeness in a publication about the results of our study. This may also lead to opportunities to present the results of the study in a conference setting. The media selected with your likeness or voice has not been altered to prevent others from identifying you. The media will only be used to support arguments regarding the hypothesis of our study within the publication or at the presentation. The media will not be used to convey any personal information about your individual mannerisms, personality, or behavior traits.

Can I review or edit the media before they are used?

You have the opportunity to review the images, video recordings, and audio material bearing your likeness at this time. You may decline to give your consent for individual media items if you so desire. Once you have reviewed the media items and signed this consent agreement, you will not have another opportunity to edit or review the images, video, or audio content before publication or conference presentation.

Who should I contact if I have questions?

The researcher conducting this study is *MHD Khairi Reda*. You may ask any questions you have now. If you have questions later, you may contact him at:
Phone: 312-996-3002, Email: mreda2@uic.edu

The faculty sponsor of this research is Associate Professor *Andrew E. Johnson*. You may contact him at:
Phone: 312-996-3002, Email: ajohnson@uic.edu

What are my rights as a research subject?

If you have any questions about your rights as a research subject, you may call the Office for Protection of Research Subjects at 312-996-1711.

Remember: Your participation in this research is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University. If you decide to participate, you are free to withdraw at any time without affecting that relationship.
You will be given a copy of this form for your information and to keep for your records.

Signature of Subject or Legally Authorized Representative

I have read (or someone has read to me) the above information. I have been given an opportunity to ask questions and my questions have been answered to my satisfaction. I agree to participate in this research. I have been given a copy of this form.

Signature Date

Printed Name

Signature of Researcher Date (must be same as subject's)

Evaluating the Use of Large, High-Resolution Displays in Visual Data Analysis

Hypothetical scenario: Exploratory analysis of crime patterns in Chicago

The City of Chicago has asked for your help in analyzing crime patterns in the city. The city has been witnessing a slow but steady decline in crime rate over the last decade. However, in the last two years, there has been a resurgence of crime in some areas. City officials want your expertise to help characterize crime patterns across the different neighborhoods. The city is also asking for your advise on how to deploy additional law enforcement resources to combat and deter crimes.

The city has given you access to its database of crimes from the year 2006 to the present. This dataset lists the majority of crimes that happened in the last 7 years. Crimes are identified by location, time, and type of crime (e.g., robbery, violation of liquor law, vehicle theft, homicide, etc...). The city has given you access to an interactive visualization tool, which shows crime data on city maps. You will be using this tool to explore the crime database, and try to identify and characterize crime patterns in the different neighborhoods as well as out of the ordinary crime activity.

To help you with this task, the city has provided a list of questions for you to use as a starting points. The city is also interested in other questions you can come up with and any additional insights you can glean from the data.

List of questions:

- What year has the most crime occurred in?
- What neighborhoods are inflected by crimes most?
- Identify crime 'hotspots' throughout the city. Are there distinct hotspots for specific types of crime?
- How does crime in the city vary with time of day, day of the week, and season?
- When do most crimes occur (time of day, day of week, and season)?
- Are there correlations between different types of crimes?
- The city wants to deploy 50 additional full-time police officers. Which areas should the police officers by deployed to? And during what times of the day?

Appendix B

PERMISSION TO USE THIRD PARTY MATERIAL

Permission to use Figure 5

Size of this preview: 474 × 599 pixels. Other resolutions: 190 × 240 pixels | 380 × 480 pixels | 475 × 600 pixels | 608 × 768 pixels | 811 × 1,024 pixels | 1,284 × 1,622 pixels.

Original file (1,284 × 1,622 pixels, file size: 550 KB, MIME type: image/jpeg)

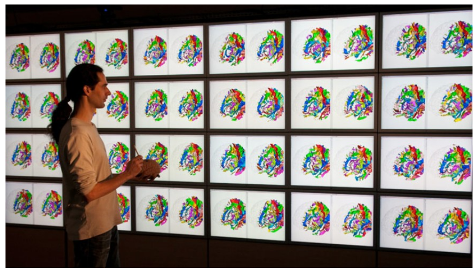
Expand view

Summary [edit]

Author Patrick Baudisch; Image created by the author based on a photo of the prototype

Licensing [edit]

Permission to use Figure 8



Permission to use picture

Emmanuel Pietriga <[illegible]@inria.fr> Sun, May 4, 2014 at 3:52 PM
To: Khairi Reda <mreda2@uic.edu>

Sure, go ahead. Just give the proper credits.

--
Emmanuel Pietriga
INRIA & INRIA Chile
http://pages.saclay.inria.fr/emmanuel.pietriga

On 4 mai 2014, at 22:30, Khairi Reda <mreda2@uic.edu> wrote:

> Dr. Pietriga,
>
> I'm writing my PhD dissertation on the use of Large High-Resolution displays in exploratory visual analysis. I came across your WILD project (http://insitu.lri.fr/Projects/WILD), and found the examples to be very compelling.
>
> I was wondering if it would be OK to use one of your pictures on the comparison of brain scans in my dissertation document. Of course, I would add citation and credit.
>
> Thank you so much!

Permission to use materials from previously published manuscript in chapter 4 (including Figure 14).

Copyright Clearance Center RightsLink®

Home | Account Info | Help

IEEE Requesting permission to reuse content from an IEEE publication

Title: Scalable Visual Queries for Data Exploration on Large, High-Resolution 3D Displays

Conference Proceedings: High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:

Author: Reda, K.; Johnson, A.; Mateevitsi, V.; Offord, C.; Leigh, J.

Publisher: IEEE

Date: 10-16 Nov. 2012

Logged in as:
Mhd Reda

LOGOUT

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

Expanding the Porthole: Leveraging Large, High-Resolution Displays in Exploratory Visual Analysis

Khairi Reda
Computer Science Department
University of Illinois at Chicago
Chicago, IL 60607 USA
mreda2@uic.edu

Jason Leigh
Computer Science Department
University of Illinois at Chicago
Chicago, IL 60607 USA
spiff@uic.edu

Catherine Offord
Department of Ecology &
Evolutionary Biology
Princeton University
Princeton, NJ 08544 USA
cofford@princeton.edu

Andrew E. Johnson
Computer Science Department
University of Illinois at Chicago
Chicago, IL 60607 USA
aej@evl.uic.edu

CHI 2014, April 26–May 1, 2014, Toronto, Ontario, Canada.
ACM 978-1-4503-2474-8/14/04.
http://dx.doi.org/10.1145/2559206.2581202

Abstract
The scale and complexity of today's datasets frequently overwhelm conventional visualization interfaces, which could negatively impact the quality of the visual analytic activity. In this paper, we investigate the use of Large, High-Resolution displays in exploratory visual analysis scenarios. We argue that the ability to see and interact with more information at once fundamentally affects users' analytic behavior, prompting them to explore their data more broadly. This positive effect may also enhance the diversity of questions and hypotheses conceived and explored by users during their analysis.

Author Keywords
Information visualization; visual exploration; visual analytics; large high-resolution displays.

ACM Classification Keywords
H.5.2. Information intH.5.2erfaces and presentation (e.g., HCI): Graphical user interfaces (GUI).

Introduction
Exploratory data analysis paradigms emphasize the use of interactive visualization tools to explore and make sense of large amounts of information. This approach favors a broad inquiry with the goal of generating

CITED LITERATURE

[Alexander et al. , 1977] Alexander, C., Ishikawa, S., and Silverstein, M.: Pattern languages. Center for Environmental Structure, 2, 1977.

[Alper et al. , 2013] Alper, B., Bach, B., Henry Riche, N., Isenberg, T., and Fekete, J.-D.: Weighted graph comparison techniques for brain connectivity analysis. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 483–492. ACM, 2013.

[Alper et al. , 2014] Alper, B. E., Henry Riche, N., and Hollerer, T.: Structuring the space: a study on enriching node-link diagrams with visual references. In Proceedings of the 32nd annual ACM conference on Human factors in computing systems, pages 1825–1834. ACM, 2014.

[Amar et al. , 2005] Amar, R., Eagan, J., and Stasko, J.: Low-level components of analytic activity in information visualization. In Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on, pages 111–117. IEEE, 2005.

[Amar and Stasko, 2005] Amar, R. A. and Stasko, J. T.: Knowledge precepts for design and evaluation of information visualizations. Visualization and Computer Graphics, IEEE Transactions on, 11(4):432–442, 2005.

[Andrews et al. , 2010] Andrews, C., Endert, A., and North, C.: Space to think: large high-resolution displays for sensemaking. In Proc. of CHI'10, pages 55–64. ACM, 2010.

[Andrews et al. , 2011] Andrews, C., Endert, A., Yost, B., and North, C.: Information visualization on large, high-resolution displays: Issues, challenges, and opportunities. Information Visualization, 10(4):341–355, 2011.

[Andrews and North, 2013] Andrews, C. and North, C.: The impact of physical navigation on spatial organization for sensemaking. Visualization and Computer Graphics, IEEE Transactions on, 19(12):2207–2216, 2013.

[Arias-Hernandez et al. , 2012] Arias-Hernandez, R., Green, T. M., and Fisher, B.: From cognitive amplifiers to cognitive prostheses: Understandings of the material basis of cognition in visual analytics. Interdisciplinary science reviews, 37(1):4–18, 2012.

[Ball and North, 2005] Ball, R. and North, C.: Analysis of user behavior on high-resolution tiled displays. Human-Computer Interaction-INTERACT 2005, pages 350–363, 2005.

[Ball et al. , 2007] Ball, R., North, C., and Bowman, D.: Move to improve: promoting physical navigation to increase user performance with large displays. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 191–200. ACM, 2007.

[Ball and North, 2008] Ball, R. and North, C.: The effects of peripheral vision and physical navigation on large scale visualization. In Proceedings of graphics interface 2008, pages 9–16. Canadian Information Processing Society, 2008.

[Bates, 2005] Bates, M. J.: Information and knowledge: An evolutionary framework for information science. Information Research, 10(4):10–4, 2005.

[Baudisch et al. , 2002] Baudisch, P., Good, N., Bellotti, V., and Schraedley, P.: Keeping things in context: a comparative evaluation of focus plus context screens, overviews, and zooming. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 259–266. ACM, 2002.

[Beaudouin-Lafon et al. , 2012] Beaudouin-Lafon, M., Huot, S., Nancel, M., Mackay, W., Pietriga, E., Primet, R., Wagner, J., Chapuis, O., Pillias, C., Eagan, J. R., et al.: Multisurface interaction in the wild room. Computer, 45(4):48–56, 2012.

[Becker and Cleveland, 1987] Becker, R. A. and Cleveland, W. S.: Brushing scatterplots. Technometrics, 29(2):127–142, 1987.

[Behrisch et al. , 2013] Behrisch, M., Davey, J., Simon, S., Schreck, T., Keim, D., and Kohlhammer, J.: Visual comparison of orderings and rankings. In EuroVis Workshop on Visual Analytics, 2013.

[Bezerianos and Isenberg, 2012] Bezerianos, A. and Isenberg, P.: Perception of visual variables on tiled wall-sized displays for information visualization applications. Visualization and Computer Graphics, IEEE Transactions on, 18(12):2516–2525, 2012.

[Bloom, 1956] Bloom, B. S.: Taxonomy of Educational Objectives: The Classification of Educational Goals. Longmans, 1956.

[Bradel et al. , 2013] Bradel, L., Self, J. Z., Endert, A., Hossain, M. S., North, C., and Ramakrishnan, N.: How analysts cognitively "connect the dots". In Intelligence and Security Informatics (ISI), 2013 IEEE International Conference on, pages 24–26. IEEE, 2013.

[Card et al. , 1999] Card, S., Mackinlay, J., and Shneiderman, B.: Readings in information visualization: using vision to think. Morgan Kaufmann, 1999.

[Chang et al. , 2009] Chang, R., Ziemkiewicz, C., Green, T. M., and Ribarsky, W.: Defining insight for visual analytics. Computer Graphics and Applications, IEEE, 29(2):14–17, 2009.

[Chi, 1997] Chi, M. T.: Quantifying qualitative analyses of verbal data: A practical guide. The journal of the learning sciences, 6(3):271–315, 1997.

[Clark and Chalmers, 1998] Clark, A. and Chalmers, D.: The extended mind. Analysis, 58(1):7–19, 1998.

[Cleveland, 1985] Cleveland, W. S.: The Elements of Graphing Data. Wadsworth Advanced Books and Software, 1985.

[Cockburn et al. , 2008] Cockburn, A., Karlson, A., and Bederson, B. B.: A review of overview+ detail, zooming, and focus+ context interfaces. ACM Computing Surveys (CSUR), 41(1):2, 2008.

[Convertino et al. , 2003] Convertino, G., Chen, J., Yost, B., Ryu, Y.-S., and North, C.: Exploring context switching and cognition in dual-view coordinated visualizations. In Coordinated and Multiple Views in Exploratory Visualization, 2003. Proceedings. International Conference on, pages 55–62. IEEE, 2003.

[Czerwinski et al. , 2006] Czerwinski, M., Robertson, G., Meyers, B., Smith, G., Robbins, D., and Tan, D.: Large display research overview. In Proc. of CHI'06, pages 69–74. ACM, 2006.

[Czerwinski et al. , 2003] Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, G., and Starkweather, G.: Toward characterizing the productivity benefits of very large displays. In Proc. Interact 2003, volume 3, pages 9–16, 2003.

[Dunbar, 1993] Dunbar, K.: Concept discovery in a scientific domain. Cognitive Science, 17(3):397–434, 1993.

[Eick and Karr, 2002] Eick, S. G. and Karr, A. F.: Visual scalability. Journal of Computational and Graphical Statistics, 11(1):22–43, 2002.

[Endert et al. , 2011] Endert, A., Fiaux, P., and North, C.: Unifying the sensemaking loop with semantic interaction. In IEEE Workshop on Interactive Visual Text Analytics for Decision Making at VisWeek 2011, 2011.

[Endert et al. , 2012a] Endert, A., Fiaux, P., and North, C.: Semantic interaction for visual text analytics. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 473–482. ACM, 2012.

[Endert et al. , 2012b] Endert, A., Fox, S., Maiti, D., and North, C.: The semantics of clustering: analysis of user-generated spatializations of text documents. In Proceedings of the International Working Conference on Advanced Visual Interfaces, pages 555–562. ACM, 2012.

[Ericsson and Simon, 1993] Ericsson, K. and Simon, H.: Protocol analysis: Verbal reports as data. MIT Press, revised edition, 1993.

[Fiaux et al. , 2013] Fiaux, P., Sun, M., Bradel, L., North, C., Ramakrishnan, N., and Endert, A.: Bixplorer: Visual analytics with biclusters. Computer, 46(8):90–94, 2013.

[Fogg, 2002] Fogg, B. J.: Persuasive technology: using computers to change what we think and do. Ubiquity, 2002(December), 2002.

[Gamma et al. , 1993] Gamma, E., Helm, R., Johnson, R., and Vlissides, J.: Design patterns: Abstraction and reuse of object-oriented design. Springer, 1993.

[Gleicher et al. , 2011] Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., and Roberts, J. C.: Visual comparison for information visualization. Information Visualization, 10(4):289–309, 2011.

[Gray and Fu, 2004] Gray, W. D. and Fu, W.-T.: Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. Cognitive Science, 28(3):359–382, 2004.

[Green et al. , 2009] Green, T., Ribarsky, W., and Fisher, B.: Building and applying a human cognition model for visual analytics. Information Visualization, 8(1):1–13, 2009.

[Gregory Trafton et al. , 2000] Gregory Trafton, J., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., and Raymond, P. D.: Turning pictures into numbers: extracting and generating information from complex visualizations. International Journal of Human-Computer Studies, 53(5):827–850, 2000.

[Healey and Enns, 2012] Healey, C. G. and Enns, J. T.: Attention and visual memory in visualization and computer graphics. Visualization and Computer Graphics, IEEE Transactions on, 18(7):1170–1188, 2012.

[Heer and Shneiderman, 2012] Heer, J. and Shneiderman, B.: Interactive dynamics for visual analysis. Queue, 10(2):30, 2012.

[Heuer, 1999] Heuer, R.: Psychology of intelligence analysis. US Government Printing Office, 1999.

[Hollan et al. , 2000] Hollan, J., Hutchins, E., and Kirsh, D.: Distributed cognition: toward a new foundation for human-computer interaction research. ACM Transactions on Computer-Human Interaction (TOCHI), 7(2):174–196, 2000.

[Hullman et al. , 2011] Hullman, J., Adar, E., and Shah, P.: Benefitting infovis with visual difficulties. Visualization and Computer Graphics, IEEE Transactions on, 17(12):2213–2222, 2011.

[Hutchings et al. , 2004] Hutchings, D. R., Smith, G., Meyers, B., Czerwinski, M., and Robertson, G.: Display space usage and window management operation comparisons between single monitor and multiple monitor users. In Proceedings of the working conference on Advanced Visual Interfaces, pages 32–39. ACM, 2004.

[Hutchins, 1995a] Hutchins, E.: Cognition in the Wild. MIT press, Cambridge, MA, 1995.

[Hutchins, 1995b] Hutchins, E.: How a cockpit remembers its speeds. Cognitive Science, 19(3):265–288, 1995.

[Isenberg et al. , 2011] Isenberg, P., Elmqvist, N., Scholtz, J., Cernea, D., Ma, K.-L., and Hagen, H.: Collaborative visualization: definition, challenges, and research agenda. Information Visualization, 10(4):310–326, 2011.

[Jagodic et al. , 2011] Jagodic, R., Renambot, L., Johnson, A., Leigh, J., and Deshpande, S.: Enabling multi-user interaction in large high-resolution distributed environments. Future Generation Computer Systems, 27(7):914–923, 2011.

[Jakobsen and Hornbæk, 2013] Jakobsen, M. R. and Hornbæk, K.: Interactive visualizations on large and small displays: the interrelation of display size, information space, and scale. Visualization and Computer Graphics, IEEE Transactions on, 19(12):2336–2345, 2013.

[Jankun-Kelly et al. , 2007] Jankun-Kelly, T., Ma, K.-L., and Gertz, M.: A model and framework for visualization exploration. Visualization and Computer Graphics, IEEE Transactions on, 13(2):357–369, 2007.

[Javed and Elmqvist, 2012] Javed, W. and Elmqvist, N.: Exploring the design space of composite visualization. In Proceedings of the 2012 IEEE Pacific Visualization Symposium (PacificVis'12), pages 1–8. IEEE, 2012.

[Johnson et al. , 2007] Johnson, C., Ross, R., Ahern, S., Ahrens, J., Bethel, W., Ma, K., Papka, M., Rosendale, J., Shen, H., and Thomas, J.: Visualization and knowledge discovery: Report from the doe/ascr workshop on visual analysis and data exploration at extreme scale. Salt Lake City October, 2007.

[Kean, 2011] Kean, T.: The 9/11 commission report: Final report of the national commission on terrorist attacks upon the United States. US Government Printing Office, 2011.

[Kehrer et al. , 2013] Kehrer, J., Piringer, H., Berger, W., and Groller, M. E.: A model for structure-based comparison of many categories in small-multiple displays. Visualization and Computer Graphics, IEEE Transactions on, 19(12):2287–2296, 2013.

[Keim and Schneidewind, 2005] Keim, D. A. and Schneidewind, J.: Scalable visual data exploration of large data sets via multiresolution. Journal of Universal Computer Science, 11(11):1766–1779, 2005.

[Kirsh, 1995] Kirsh, D.: The intelligent use of space. Artificial Intelligence, 73(1):31–68, 1995.

[Kirsh, 2013] Kirsh, D.: Thinking with external representations. In Cognition Beyond the Brain, pages 171–194. Springer, 2013.

[Kirsh and Maglio, 1994] Kirsh, D. and Maglio, P.: On distinguishing epistemic from pragmatic action. Cognitive Science, 18(4):513–549, 1994.

[Klahr, 2002] Klahr, D.: Exploring science: The cognition and development of discovery processes. MIT press, 2002.

[Klahr and Dunbar, 1988] Klahr, D. and Dunbar, K.: Dual space search during scientific reasoning. Cognitive Science, 12(1):1–48, 1988.

[Klein et al. , 2006] Klein, G., Moon, B., and Hoffman, R. R.: Making sense of sensemaking 2: A macrocognitive model. Intelligent Systems, IEEE, 21(5):88–92, 2006.

[Krestel et al. , 2011] Krestel, R., Demartini, G., and Herder, E.: Visual interfaces for stimulating exploratory search. In Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries, pages 393–394. ACM, 2011.

[Lam, 2008] Lam, H.: A framework of interaction costs in information visualization. Visualization and Computer Graphics, IEEE Transactions on, 14(6):1149–1156, 2008.

[Lambooij et al. , 2007] Lambooij, M., IJsselsteijn, W., and Heynderickx, I.: Visual discomfort in stereoscopic displays: a review. Stereoscopic Displays and Virtual Reality Systems XIV, 6490(1), 2007.

[Larkin and Simon, 1987] Larkin, J. H. and Simon, H. A.: Why a diagram is (sometimes) worth ten thousand words. Cognitive Science, 11(1):65–100, 1987.

[Leigh et al. , 2012] Leigh, J., Johnson, A., Renambot, L., Peterka, T., Jeong, B., Sandin, D., Talandis, J., Jagodic, R., Nam, S., Hur, H., et al.: Scalable resolution display walls. Proceedings of the IEEE, 101(1):115–129, 2012.

[Liu et al. , 2008] Liu, Z., Nersessian, N. J., and Stasko, J. T.: Distributed cognition as a theoretical framework for information visualization. Visualization and Computer Graphics, IEEE Transactions on, 14(6):1173–1180, 2008.

[Liu and Stasko, 2010] Liu, Z. and Stasko, J. T.: Mental models, visual reasoning and interaction in information visualization: A top-down perspective. Visualization and Computer Graphics, IEEE Transactions on, 16(6):999–1008, 2010.

[Maxcey-Richard and Hollingworth, 2013] Maxcey-Richard, A. M. and Hollingworth, A.: The strategic retention of task-relevant objects in visual working memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 39(3):760, 2013.

[Meyer et al. , 2009] Meyer, M., Munzner, T., and Pfister, H.: Mizbee: a multiscale synteny browser. Visualization and Computer Graphics, IEEE Transactions on, 15(6):897–904, 2009.

[Ni et al. , 2006] Ni, T., Schmidt, G., Staadt, O., Livingston, M., Ball, R., and May, R.: A survey of large high-resolution display technologies, techniques, and applications. In Virtual Reality Conference, 2006, pages 223–236. IEEE, 2006.

[Norman, 2002] Norman, D. A.: The Design of Everyday Things. Basic Books, 2002.

[Norris, 1998] Norris, J. R.: Markov Chains. Cambridge University Press, 1998.

[North, 2006] North, C.: Toward measuring visualization insight. Computer Graphics and Applications, IEEE, 26(3):6–9, 2006.

[Offord et al. , 2013] Offord, C., Reda, K., and Mateevitsi, V.: Context-dependent navigation in a collectively foraging species of ant, messor cephalotes. Insectes Sociaux, 60(3):361–368, 2013.

[O'Hara and Payne, 1998] O'Hara, K. P. and Payne, S. J.: The effects of operator implementation cost on planfulness of problem solving and learning. Cognitive Psychology, 35(1):34–70, 1998.

[Ohlsson, 2009] Ohlsson, S.: Resubsumption: A possible mechanism for conceptual change and belief revision. Educational Psychologist, 44(1):20–40, 2009.

[Parsons and Sedig, 2014] Parsons, P. and Sedig, K.: Distribution of information processing while performing complex cognitive activities with visualization tools. In Handbook of Human Centric Visualization, pages 693–715. Springer, 2014.

[Patterson et al. , 2001] Patterson, E. S., Roth, E. M., and Woods, D. D.: Predicting vulnerabilities in computer-supported inferential analysis under data overload. Cognition, Technology & Work, 3(4):224–237, 2001.

[Peters et al. , 2009] Peters, C., Castellano, G., and de Freitas, S.: An exploration of user engagement in hci. In Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots, number 9, pages 1–3. ACM, 2009.

[Pike et al. , 2009] Pike, W. A., Stasko, J., Chang, R., and O'Connell, T. A.: The science of interaction. Information Visualization, 8(4):263–274, 2009.

[Pirolli and Card, 2005] Pirolli, P. and Card, S.: The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In Proceedings of International Conference on Intelligence Analysis, volume 5, 2005.

[Pirolli et al. , 2003] Pirolli, P., Card, S. K., and Van Der Wege, M. M.: The effects of information scent on visual search in the hyperbolic tree browser. ACM Transactions on Computer-Human Interaction (TOCHI), 10(1):20–53, 2003.

[Pirolli, 2007] Pirolli, P. L.: Information foraging theory: Adaptive interaction with information. Oxford University Press, 2007.

[Plumlee and Ware, 2006] Plumlee, M. D. and Ware, C.: Zooming versus multiple window interfaces: Cognitive costs of visual comparisons. ACM Transactions on Computer-Human Interaction (TOCHI), 13(2):179–209, 2006.

[Pousman et al. , 2007] Pousman, Z., Stasko, J. T., and Mateas, M.: Casual information visualization: Depictions of data in everyday life. Visualization and Computer Graphics, IEEE Transactions on, 13(6):1145–1152, 2007.

[Privitera, 2006] Privitera, C. M.: The scanpath theory: its definitions and later developments. In Electronic Imaging 2006, pages 60570A–60570A. International Society for Optics and Photonics, 2006.

[Purchase et al. , 2007] Purchase, H. C., Hoggan, E., and Görg, C.: How important is the "mental map"?–an empirical investigation of a dynamic graph layout algorithm. In Graph Drawing, pages 184–195. Springer, 2007.

[Ratwani et al. , 2008] Ratwani, R. M., Trafton, J. G., and Boehm-Davis, D. A.: Thinking graphically: Connecting vision and cognition during graph comprehension. Journal of Experimental Psychology: Applied, 14(1):36–49, 2008.

[Reda et al. , 2013a] Reda, K., Aurisano, J., Febretti, A., Leigh, J., and Johnson, A.: Visualization design patterns for ultra-resolution display environments. In Proceedings of the workshop on Visualization Technology and Systems Infrastructure (VISTech '13), 2013.

[Reda et al. , 2013b] Reda, K., Febretti, A., Knoll, A., Aurisano, J., Leigh, J., Johnson, A., Papka, M., and Hereld, M.: Visualizing large, heterogeneous data in hybrid-reality environments. Computer Graphics and Applications, 33(4):38–48, 2013.

[Reda et al. , 2012] Reda, K., Johnson, A., Mateevitsi, V., Offord, C., and Leigh, J.: Scalable visual queries for data exploration on large, high-resolution 3d displays. In SC Companion: High Performance Computing, Networking, Storage and Analysis, pages 196–205. IEEE, 2012.

[Reda et al. , 2013] Reda, K., Mateevitsi, V., and Offord, C.: A human-computer collaborative workflow for the acquisition and analysis of terrestrial insect movement in behavioral field studies. EURASIP Journal on Image and Video Processing, 2013(1):1–17, 2013.

[Reda et al. , 2014] Reda, K., Offord, C., Johnson, A., and Leigh, J.: Expanding the porthole: leveraging large, high-resolution displays in exploratory visual analysis. In CHI'14 EA on Human Factors in Computing Systems, pages 2047–2052. ACM, 2014.

[Reda et al. , 2011] Reda, K., Tantipathananandh, C., Johnson, A., Leigh, J., and Berger-Wolf, T.: Visualizing the evolution of community structures in dynamic social networks. In Computer Graphics Forum, volume 30, pages 1061–1070. Eurographics Association, 2011.

[Roberts, 2007] Roberts, J. C.: State of the art: Coordinated & multiple views in exploratory visualization. In Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on, pages 61–71. IEEE, 2007.

[Rønne Jakobsen and Hornbæk, 2011] Rønne Jakobsen, M. and Hornbæk, K.: Sizing up visualizations: effects of display size in focus+ context, overview+ detail, and zooming interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 1451–1460. ACM, 2011.

[Rosenholtz et al. , 2007] Rosenholtz, R., Li, Y., and Nakano, L.: Measuring visual clutter. Journal of Vision, 7(2):17, 2007.

[Ruddle et al. , 2013] Ruddle, R. A., Thomas, R. G., Randell, R. S., Quirke, P., and Treanor, D.: Performance and interaction behaviour during visual search on large, high-resolution displays. Information Visualization, 2013.

[Saraiya et al. , 2005] Saraiya, P., North, C., and Duca, K.: An insight-based methodology for evaluating bioinformatics visualizations. Visualization and Computer Graphics, IEEE Transactions on, 11(4):443–456, 2005.

[Schaffer et al. , 1996] Schaffer, D., Zuo, Z., Greenberg, S., Bartram, L., Dill, J., Dubs, S., and Roseman, M.: Navigating hierarchically clustered networks through fisheye and full-zoom methods. ACM Transactions on Computer-Human Interaction (TOCHI), 3(2):162–188, 1996.

[Sedig et al. , 2012] Sedig, K., Parsons, P., and Babanski, A.: Towards a characterization of interactivity in visual analytics. Journal of Multimedia Processing and Technologies, 3(1):12–28, 2012.

[Shneiderman, 1994] Shneiderman, B.: Dynamic queries for visual information seeking. Software, IEEE, 11(6):70–77, 1994.

[Shneiderman, 1996] Shneiderman, B.: The eyes have it: A task by data type taxonomy for information visualizations. In Visual Languages, 1996. Proceedings., IEEE Symposium on, pages 336–343. IEEE, 1996.

[Shneiderman, 2008] Shneiderman, B.: Extreme visualization: squeezing a billion records into a million pixels. In Proceedings of the 2008 ACM SIGMOD international conference on Management of data, pages 3–12. ACM, 2008.

[Shupp et al. , 2009] Shupp, L., Andrews, C., Dickey-Kurdziolek, M., Yost, B., and North, C.: Shaping the display of the future: The effects of display size and curvature on user performance and insights. Human–Computer Interaction, 24(1-2):230–272, 2009.

[Simmons, 2001] Simmons, T.: What's the optimum computer display size? Ergonomics in Design: The Quarterly of Human Factors Applications, 9(4):19–25, 2001.

[Swaminathan and Sato, 1997] Swaminathan, K. and Sato, S.: Interaction design for large displays. Interactions, 4(1):15–24, 1997.

[Tan, 2004] Tan, D. S.: Exploiting the cognitive and social benefits of physically large displays. Doctoral dissertation, Carnegie Mellon University, 2004.

[Tan et al. , 2003] Tan, D., Gergle, D., Scupelli, P., and Pausch, R.: With similar visual angles, larger displays improve spatial performance. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 217–224. ACM, 2003.

[Teasley et al. , 2002] Teasley, S. D., Covi, L. A., Krishnan, M. S., and Olson, J. S.: Rapid software development through team collocation. Software Engineering, IEEE Transactions on, 28(7):671–683, 2002.

[Thomas and Cook, 2005] eds. J. Thomas and K. Cook Illuminating the path: The research and development agenda for visual analytics. IEEE Computer Society, 2005.

[Treisman, 1986] Treisman, A.: Features and objects in visual processing. Scientific American, 255(5):114–125, 1986.

[Tufte, 2001] Tufte, E. R.: The visual display of quantitative information. Graphics Press, Cheshire, CT, 2nd edition, 2001.

[Tukey, 1977] Tukey, J. W.: Exploratory Data Analysis. Addison-Wesley, 1977.

[Tukey, 1980] Tukey, J. W.: We need both exploratory and confirmatory. The American Statistician, 34(1):23–25, 1980.

[Tukey and Wilk, 1966] Tukey, J. W. and Wilk, M.: Data analysis and statistics: an expository overview. In Proceedings of the November 7-10, 1966, fall joint computer conference, pages 695–709. ACM, 1966.

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

[Wang Baldonado et al. , 2000] Wang Baldonado, M. Q., Woodruff, A., and Kuchinsky, A.: Guidelines for using multiple views in information visualization. In Proceedings of the working conference on Advanced visual interfaces, pages 110–119. ACM, 2000.

[Ware, 2000] Ware, C.: Information visualization: perception for design. Morgan Kaufmann, 2nd edition, 2000.

[Ware, 2012] Ware, C.: Information visualization: perception for design. Morgan Kaufmann, 3rd edition, 2012.

[Ware et al. , 2013] Ware, C., Wright, W., and Pioch, N. J.: Visual thinking design patterns. In Proceedings of the 19th International Conference on Distributed Multimedia Systems (DMS'13), 2013.

[Woodson and Conover, 1964] Woodson, W. E. and Conover, D. W.: Human engineering guide for equipment designers. University of California Press, 1964.

[Yost et al. , 2007] Yost, B., Haciahmetoglu, Y., and North, C.: Beyond visual acuity: the perceptual scalability of information visualizations for large displays. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 101–110. ACM, 2007.

[Yost and North, 2006] Yost, B. and North, C.: The perceptual scalability of visualization. Visualization and Computer Graphics, IEEE Transactions on, 12(5):837–844, 2006.

[Zimmerman, 2005] Zimmerman, C.: The development of scientific reasoning skills: What psychologists contribute to an understanding of elementary science learning. Report to the National Research Council, Committee on Science Learning Kindergarten through 8th Grade, August 2005.

VITA

MHD Khairi Reda

Education	Ph.D., Computer Science, *University of Illinois at Chicago*, 2014 M.S., Computer Science, *University of Illinois at Chicago*, 2009 B.S., Computer Science, *University of Damascus*, 2005
Academic Experience	Graduate Research Assistant, *University of Illinois at Chicago* 2008 – present Research Intern, *Argonne National Laboratory* 2012 – present
Professional Experience	Intern Software Engineer, *Midway Games* (now *NetherRealm Studios*) Summer 2007 Software Engineer, *Platinum, Inc.* 2005 – 2006
Professional Membership	ACM Special Interest Group on Human-Computer Interaction (SIGCHI), member Association for Computer Machinery, student member
Awards	Image of Research (2nd place), University of Illinois at Chicago, 2013 Graduate College Presenter Award, 2009 Graduate Student Council Travel Award, 2009 Honor Graduate, University of Damascus, 2005

Publications

Knoll, A., Wald, I., Navratil, P., Bowen, A., Reda, K., Papka, M., Gaither, K. RBF Volume Ray Casting on Multicore and Manycore CPUs. Computer Graphics Forum, 33(3):71–80. Eurographics Association, 2014

Reda, K., Chau, D., Mostafa, Y., Nagrajan, S., Leigh, J., Nishimoto, A., Kahler, E., Demeter, J. Design Guidelines for Multiplayer Video Games on Multi-touch Displays. Computers in Entertainment, 11(1):1–17. ACM, 2014

Reda, K., Offord, C., Johnson, A., Leigh, J. Expanding the Porthole: Leveraging Large, High-Resolution Displays in Exploratory Visual Analysis. In CHI'14 Extended Abstracts on Human Factors in Computing Systems, pages 2047–2052. ACM, 2014

Reda, K., Knoll, A., Nomura, K., Papka, M., Johnson, A., Leigh, J. Visualizing Large-Scale Atomistic Simulations in Ultra-Resolution Immersive Environments. In Proceedings of the IEEE Symposium on Large-Scale Data Analysis and Visualization, pages 59–65. IEEE, 2013

Reda, K., Febretti, A., Knoll, A., Aurisano, J., Leigh, J., Johnson, A., Papka, M., Hereld, M. Visualizing Large, Heterogenous Data in Hybrid-Reality Environments. Computer Graphics and Applications, 33(4):38–48. IEEE, 2013

Reda, K., Mateevitsi, V., Offord, C. A Human-Computer Collaborative Workflow for the Acquisition and Analysis of Terrestrial Insect Movement in Behavioral Field Studies. EURASIP Journal on Image and Video Processing, 2013:48. Springer, 2013

Offord, C., Reda, K., Mateevitsi, V. Context-Dependent Navigation in a Collectively Foraging Species of Ants, Messor cephalotes. Insectes Sociaux, 60(3):361–368. Springer, 2013

Reda, K., Johnson, A., Mateevitsi, A., Offord, C., Leigh, J. Scalable Visual Queries for Data Exploration on Large, High-Resolution 3D Displays. In Proceedings of the SC'12 Companion: High Performance Computing, Networking Storage and Analysis, pages 196–205. IEEE, 2012

Reda, K., Tantipathananandh, C., Johnson, A., Leigh, J., Berger-Wolf, T. Visualizing the Evolution of Community Structures in Dynamic Social Networks. Computer Graphics Forum, 30(3):1061–1070. Eurographics Association, 2011

Reda, M. K., SocioScape—Spatio-temporal Visual Analysis of Group Dynamics in Social Networks, M.S. thesis. University of Illinois at Chicago, 2009

Aurisano, J., Reda, K., Johnson, A., Leigh, J. Bacterial Gene Neighborhood Investigation Environment: A Large-Scale Genome Visualization for Big Displays *(under review)*

Nam, S., Reda, K., Johnson, A., Leigh, J. Multiuser-Centered Resource Scheduling for Collaborative Display Wall Environments *(under review)*

Abstracts

Aurisano, J., Reda, K., Johnson, A., Leigh, J. Bacterial Gene Neighborhood Investigation Environment: A Large-Scale Genome Visualization for Big Displays. Poster at the 4th Symposium on Biological Data Visualization (BioVis'14), 2014

Mateevitsi, V., Reda, K., Leigh, J., Johnson, A. The Health Bar: A Persuasive Ambient Display to Improve the Office Worker's Well Being. In Proceedings of 5th Augmented Human International Conference. ACM, 2014

Reda, K., Tantipathananandh, C., Berger-Wolf, T., Leigh, J., Johnson, A. SocioScape—a Tool for Interactive Exploration of Spatio-Temporal Group Dynamics in Social Networks. In Proceedings of the IEEE Information Visualization Conference. IEEE, 2009

Exhibits & Demonstrations

SIGGRAPH'11, Aug 2011, Vancouver, BC, Canada
Demonstrated interactive applications for Large, High-Resolution Displays including the Scalable Adaptive Graphics Environment and a 20-foot virtual paint canvas.

Chicago Field Museum, Royal Ontario Museum, National Museum of Australia, and others, 2009–2011
RainTable, an interactive museum exhibit I co-developed, was featured as part of the *H_2O = Life* exhibit, traveling to several prominent museums in North America and Australia.

American Association for the Advancement of Science, Feb 2009, Chicago, IL
Demonstrated TacTile, a custom-built multi-touch display for collaborative information exploration. This exhibit was part of a showcase for National Science Foundation's Major Research Instrumentation program.

American Geophysical Union, Dec 2009, San Francisco, CA
Demonstrated *RainTable* and *TacTile*.

Supercomputing'08, Nov 2008, Austin, TX
Demonstrated interactive multi-touch visualizations.

Invited Talks

Future of Software Frameworks for Advanced Visualization Instrumentation, panelist at the workshop on Visualization Technology and Systems Infrastructure, SC'13, Nov, 2013

Advanced Visualization Environments for Scientific Discovery and Collaboration, University of Damascus, Dec, 2010

Visualizing Dynamic Social Networks, and Collaborating with Domain Scientists, guest lecturer on Visualization & Visual Analytics, University of Illinois at Chicago, 2010

Understanding Animal Behavior with Social Network Visualization and Analysis, guest lecturer on Information Aesthetics, University of Illinois at Chicago, Spring, 2008