Supporting Collaborative Exploratory Visual Data Analysis in Multi-device Environments

by

ABEER ALSAIARI
BSc., Taibah University, Madina, Saudi Arabia, 2005
MSc., King Saud University, Riyadh, Saudi Arabia, 2011

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, 2021

Chicago, Illinois

Defense Committee:
Andrew E. Johnson, Chair and Advisor
Joseph Michaelis
Wei Tang
Steve Jones, Department of Communication
Khairi Reda, Indiana University-Purdue University Indianapolis

"Dedicated to my father Abdullah and my mother Muslaha.".

ACKNOWLEDGMENT

Despite all the challenges along the way in writing this dissertation, writing this acknowledgment is the most difficult, in a sense. Many people influenced me to reach the end of this journey, the start of a brighter new chapter. First, I would like to thank my advisor professor Andrew Johnson for his continuous support and valuable feedback that helped to better shape my research. His continuous appreciation of whatever small effort I make kept me motivated. From him that I learned how to explore the path as a researcher and see things in new and unexpected ways. I could not have asked for a better advisor. I would also like to thank my committee, professors Steve Jones, Joseph Michaelis, Wei Tang, and Khairi Reda for their time in reading my dissertation and providing valuable critique and feedback.

Special thanks go to my friends and colleagues at the Electronic Visualization Lab. I was particularly fortunate to be surrounded by wonderful people like you, who are always ready to provide help, advice and inspiration. Many thanks to Arthur. He never hesitated to help and even to find solutions for my technical problems. Thanks to Juan for all the good times in the hard days. Thanks to Jillian, Sai, Lance, Luc, Dana and everyone at EVL who were always willing to help. Keep the good and creative work and I will always remain a member of the EVL family.

Thanks to Azmi, for his uncompromising support, love, and patience in the last two years, for understanding stressful times, and for always managing to encourage me.

ACKNOWLEDGMENT (Continued)

Lastly, my heartfelt thanks go to my family, my first supporters since the beginning. With my papa, it is all started that afternoon when we left to Riyadh. I know I cannot pay him back for all of the sacrifices he has made for his family. Thank you mama for keeping me in your prayers, my source of tranquility. It's time to spend more time with you my old good friend after being living overseas for more than eight years. Thank you my brothers, Naif, Saleh, and Abdulrahman (may your soul rest in peace) for your endless love and support. I will always remember how you were believing in me and motivating me to take the challenge. I always see in your eyes that I can do it. Thanks to Abdulmjjed for sharing this journey with me. Seeing your success as well nicely seasoned the accomplishments. With love, I dedicate this to you.

AA

CONTRIBUTIONS OF AUTHORS

Chapter 1 introduces the thesis problem and presents the main research questions addressed by this dissertation. Some of these research questions were framed in a poster (Alsaiari and Johnson., 2019) for which I was the primary author. Chapter 2 presents background, related work, and design principles for this research. Some of these related work and design principles appeared in a published paper (Alsaiari and Johnson., 2019) for which I was the primary author. Copyright © 2019 IEEE. Chapter 3 represents a designed framework for multi-device visual data analysis. This chapter appeared in a published paper (Alsaiari and Johnson., 2019) for which I was the primary author. Copyright © 2019 IEEE. Chapter 4 describes a user study to understand collaborative visual data analysis in multi-device environments. Addressed research question and used methodologies appeared in a poster (Alsaiari and Johnson., 2019) for which I was the primary author. Used software and data sets were previously appeared in a published paper (Alsaiari and Johnson., 2019) for which I was the primary author. Parts of this chapter appeared in a published paper (Alsaiari et al., 2020) for which I was the primary author. Data sets were provided courtesy of http://service.iris.edu/ and http://www.occeweb.com/. Chapter 5 presents the design of visualization tool to visualize dimensions search space. Chapter 6 presents and evaluation study to evaluate the effects of visualizing the dimensions search space on exploratory visual data analysis. Krishna Bharadwaj and Arthur Nishimoto helped in video coding. Chapter 7 concludes the dissertation and summarizes the main contributions of this research.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

FIGURE **PAGE**

LIST OF FIGURES (Continued)

FIGURE **PAGE**

SUMMARY

There is a steadily growing interest in leveraging ecosystems of digital devices that go beyond a single desktop for collaborative visual data analysis and exploration. This new thrust of multi-device interfaces supports new models for complex collaboration scenarios, and have great potential to support analysts in their data analysis by utilizing each device's capabilities. However, there are some challenges inherently associated with visual data analysis in multi-device environments (MDE). This dissertation investigates how the analytical process occurs in multi-user multi-device environments to provide a theoretical understanding of collaborative exploratory visual data analysis and better inform the design of visualization tools. First, I touched on the challenges of designing cross-device visualization tools by introducing the design and implementation of a multi-device system for collaborative visual data analysis that enables cross-device visualization sharing and simultaneous interaction. Then, through an exploratory user study, I evaluated strategies of exploratory visual data analysis in a collaborative multi-user multi-device environment. I synthesized a two-level characterization of the analysis structure from observed analysis behaviors. I observed that subjects navigate the data space in three identified exploration patterns and the analysis was primarily depth-oriented. In addition, the cost of deciding what to explore next "Gulf of Goal Formation" is higher in collaborative settings due to short-term memory and the recency effect. I hypothesized that visualizing the dimensions search space would increase the breadth of the analysis and reduce the decision cost. Using a between-groups study, I evaluated the effect of revealing information about what dimension's data space coverage(s) were investigated and what were left. The results indicate that visualizing dimensions search space increases the breadth of the analysis and reduces the decision cost by positively affecting the rate of views generation.

CHAPTER 1

INTRODUCTION

Parts of of this chapter were previously published as: Alsaiari, A. and Johnson, A. (2019). "Towards Understanding Collaborative Visual Data Analysis in Multi-Device Environments". In *2019 IEEE VIS.*

Visual Analytics, as defined by Cook and Thomas in their Research and Development Agenda, is "the science of analytical reasoning facilitated by interactive visual interfaces" (5). The analytical reasoning is an iterative process that involves cycles of visualization creation, interaction and refinement. Therefore, Visual Analytics tools facilitate the human reasoning process by the means of technological support and analytical techniques.

With the increased amount of data that comes from different sources and domains, visual analytics became rarely a solitary activity. Analysts from different backgrounds need to work together to contribute their contextual knowledge and create a better understanding of their data. The integration of visualization and collaboration into a new direction of research imposed new challenges for designers and generated new prospects for researchers to expand the state of art design and evaluation of visualization tools. Designing for collaborative visual data analysis requires special considerations to fully support the sense making process (6).

As shown in Figure 1, collaborative visualization can occur in four different scenarios classified over time and space. Co-located collaborative systems involve a shared workspace such

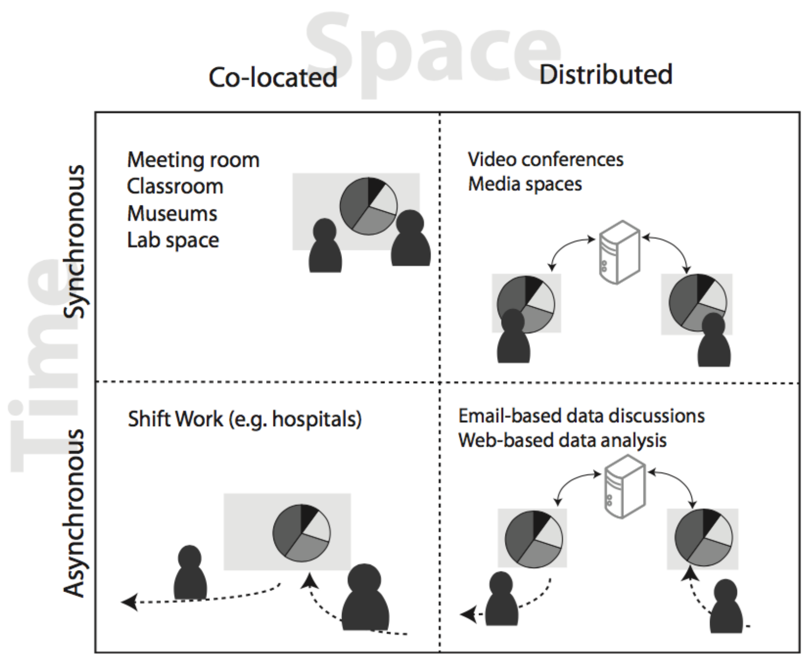


Figure 1: Different scenarios of collaborative visualization classified over time and space (1)

as large displays or tabletops, while distributed systems involve a shared virtual workspaces for remote collaboration.

Visual data analysis tools should support suitable social interactions according to the type of the collaborative environment. A rich body of research investigated the design of visualization tools for co-located (7) (8) (9) and distributed collaboration (10), each of which required unique design principles derived from the visualization and computer-supported cooperative work (CSCW) communities. Besides the technical aspects, a vast majority of past research has focused on more human-centered questions to address issues regarding work coordination, sharing, and groups' awareness in co-located and distributed settings.

With the popularity and availability of various types of devices with different input and output modalities, a new thrust of research has emerged to explore the potential of these display technologies in supporting analytical reasoning and sense making. Multi-device environments (MDE) have great potential to support analysts in their data analysis by utilizing each device's capabilities. However, little is known on how to design visualization tools for multi device environments to support efficient visual data analysis. This research investigates the design of visualization tools for collaborative exploratory visual data analysis in multi device environments.

1.1 Motivation

Recently, there has been an increased interest in leveraging ecosystems of multiple devices for collaborative visual data analysis (11) (12) (13) (14) (15) (16) (17) (18), imposing the need to rethinking the design and evaluation of visualization tools for these environments.

This new thrust of leveraging multi-device environments for visual data analysis supports new models for complex collaboration scenarios and provides the means for users to immerse themselves in their data by creating flexible and mobile exploration territories. However, there are some challenges inherently associated with visual data analysis in multi-device environments.

First, as the analytical process is underway, many visualizations become scattered among different devices and displays. Building a mental model of the analysis flow can render the analytical process more challenging as it would be difficult to track many visualizations. Therefore, it's difficult for analysts to keep track of all the prior analyses and the cost of deciding what to explore next can be even higher. Analysts in collaborative settings need to understand

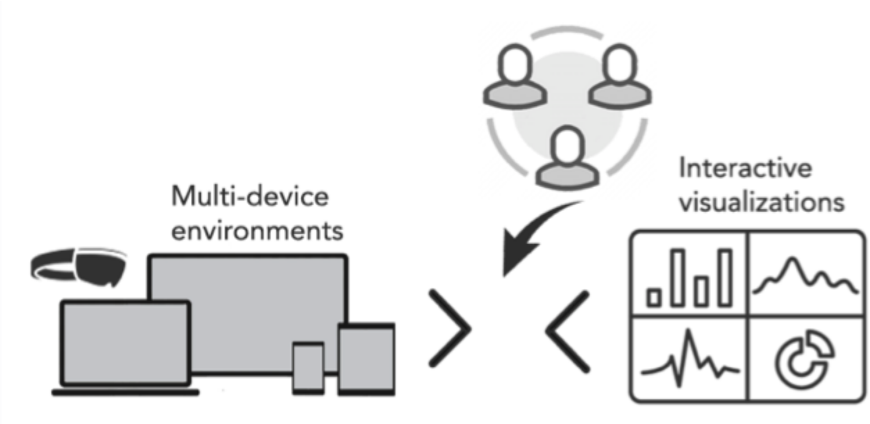


Figure 2: This research investigates the design of visualization tools for collaborative exploratory visual data analysis in multi device environments.

what courses of analysis were investigated by team members and what were left. Therefor, supporting exploratory visual data analysis is essential especially when multiple analysts work together in a setting beyond the single desktop.

In addition, with factors such as the recency effect and the short-term memory, they tend to remember the most recent exploration. This can affect the breadth of the analysis. The tendency to recall the most recent items encourage a depth-oriented exploration. Therefor, visualization tools should promote a breadth-first analysis.

1.2 Thesis Problem

Motivated by the above-mentioned challenges, this research explores the collaborative visual data analysis in multi device environments. The goal is to investigate **how the analytical process occurs** in this multi-user multi-device environment **to provide a theoretical un-**

derstanding of collaborative exploratory visual data analysis **and better inform the design** of visualization tools.

This dissertation investigates the following research questions:

RQ1: In the first phase of this research, I started with an exploratory study to address the question of **what is the complex picture of users' experience during a collaborative visual data analysis in a multi-user multi-device environment?**

Activities in multi-device environments can be complex. It is important to understand aspects around tools, users, and tasks, and how these aspects shape the analysis process. Understanding the analytical strategies and their associated challenges will help us to identify important design considerations and requirements that support some of these challenges.

Earlier studies that aimed to understand the collaborative process of visual data analysis focused on a few elements (display use, processes, work styles, etc.) as they address group's work around a single display. To capture the complexity of collaborative visual data analysis in a multi-user multi-device environment, I presented an Activity-Centered approach that identifies the network of actors that make the activity takes place in this environment: users, tools, and task. As presented in Chapter 4, these activity actors were identified based on the visualization reference models, and used to apply appropriate empirical methods in terms of each aspect for analysis. I believe that these three aspects (users, tools, and tasks) shape the complex picture of user experience in this environment. In Chapter 4, I discuss the study and the application of the hybrid analysis approach. I synthesized an overall understanding of the process and identified a set of observed challenges.

RQ2: The second important aspect is to derive an understanding of **what is the characterization of the analysis process in this environment**. After multiple passes of qualitative coding, I found that the analysis proceeds at two levels. Within each level, I further observed a set of exploration patterns. At the higher level, participants were taken along a set of exploration paths, and along each analysis path, I observed a set of view-to-view exploration patterns that occur within the larger cycles of the analysis. In Chapter 4, I present a structural categorization of the analytical process in such a complex environment and discuss how this categorization corresponds to the current structural assumptions of exploratory visual data analysis. I also discuss research implications and touch on the potential value of augmenting visualization tools with supportive mechanism for efficient exploration.

The presented characterization of the analysis flow revealed patterns of how participants searched the space of data. I identified three patterns of navigating the dimensions data space as will be discussed in Chapter 4. However, participants relied on their mental model on navigating the data space. They were blocked form how much, and what, of the data space they have covered. As the analysis proceeds, it was hard to keep track of prior analyses due to many visualizations. Therefor, the exploration was oriented towards limited dimensions' space coverage. We need an approach that increases the awareness of the group and individual exploration of the data search space, and that can facilitate the exploration process. This led me to explore how to augment the design with a visualization of dimensions search space and what are the effects on the analysis.

RQ3: **What are the effects of visualizing the coverage of the dimensions search space on exploratory visual data analysis?** I hypothesized that explicit visualization of dimensions search space would improve the performance of the exploratory task. In chapter 5, I presented the design and the implementation of visualizing the dimensions search space. In chapter 6, I evaluated the design in a between-groups study. I tested three hypotheses in the presented study. I hypothesized that visualizing the dimensions search space will reduces the decision cost, increases the breadth of the analysis, and increases formed questions and observations. The results from the study supported the first and the second hypotheses. The results showed that the visualization of the dimensions search space reduces the cost of decision and increases the breadth of the analysis.

1.3 Research Scope

This research focuses on supporting visual data analysis in co-located settings where a small group of collaborators work together using multiple devices to make sense of their data. The focus on this direction was motivated by the benefits that co-located collaboration offers in many disciplines. Co-located collaboration reduces communication barriers that appear in other settings, as collaborators communicate directly at the same time and in the same place. With direct interaction, collaborators can easily assess their team's need and adjust the team's work. Especially in the case of using multiple devices, collaborators can easily switch roles and change the analysis strategy without the need for the cumbersome installation of additional instruments.

Furthermore, the exploratory nature of visual data analysis requires social interactions for discussion, ideation, etc. which can be handy in co-located settings.

In addition, in co-located settings, collaborators can employ multiple devices at the same time taking advantageous of offered opportunities. As I will discuss in the next chapter, multi device settings have great potential for collaborative visual data analysis. In this research, I employ a workspace with a large display integrated with portable devices. Specifically, the large display is integrated with tablets, laptops, and an AR headset (See Chapter 3 for more details). Earlier studies on large displays have shown that their physical affordances result in the emergence of different kinds of collaboration that have been used in many domains. The presence of portable devices would allow for different collaboration styles. As reported by Isenberg et al.(2011b), collaborators tend to branch from the group work which emphasizes the importance of supporting different work styles in groupware applications (8). In this research, I investigate the geoscience application domain. A few reasons motivated the selection of this domain. First, geoscience domain data typically has spatial and non-spatial features. The large scale exploration of these types of data benefits from the emergence of solutions that go beyond a single desktop. For example, large displays offer a large-scale exploration of spatial 2D representation, while AR headsets offer a spatial 3D representation. In addition, the analysis of heterogeneous spatiotemporal data has been emerging recently.

1.4 Methodological Approach

Empirical study approaches have been widely adopted by visualization research for visualization evaluation, and for understanding the behavior of individuals using the visualization tools.

Tory (19) provided a categorization of user study methods applied to visualization research based on their goals of conducting. She stated that user studies are performed in visualization not only for "*evaluation*" but also for "*understanding*" the context of use. After specifying the study goals, researchers should delineate their research questions and objectives and identify appropriate empirical methods. This categorization helps the study designers to articulate their goals and narrow down their choices of appropriate empirical methods. Empirical approaches common in visualization research include the quantitative experiment, the qualitative observational study, and the usability study. Qualitative methods are widely used to answer exploratory questions using collected qualitative data. However, it has become common to use a mixed method to offset the shortcomings of each empirical method.

The study conducted in chapter 4 falls into the category of user studies for "*understanding*". The goal is to understand the context of use to enhance the tool design. More specifically, my goal was to observe the collaborative process of visual data analysis to inform the design space. Therefore, I designed the user study by paying attention to methods appropriate for this goal of empirical studies. As I discuss in Chapter 4, I used the exploratory user study method by applying a mixed methods employing both qualitative and quantitative analysis. The study conducted in chapter 6 falls into the category of user studies for "*evaluation*" to evaluate the context of use.

CHAPTER 2

BACKGROUND AND RELATED WORK

Parts of this chapter were previously published as: Alsaiari, A., Johnson, A., Nishimoto, A., "PolyVis: Cross-Device Framework for Collaborative Visual Data Analysis", In the Proceedings of *2019 IEEE International Conference on Systems, Man, and Cybernetics* (IEEE SMC 2019), October 6-9, 2019, Bari, Italy.

2.1 Collaborative Visualization

Collaborative visualization as defined by Isenberg et al. (20) is "the shared use of computer-supported, (interactive,) visual representations of data by more than one person with the common goal of contribution to joint information processing activities". It lies at the intersection of two areas, visualization and computer-supported cooperative work (CSCW). Each of these areas has a long history of research, and specific challenges and requirements. Therefore, collaborative visualization brings its unique challenges to the intersection of these areas.

During the last twenty-years, many frameworks were proposed to support collaborative visualization for small groups to internet scale users. For example, Lark (9) is a visualization tool that support co-located collaboration for small groups around tabletops. In contrast, Many Eyes (10) is a web-based framework proposed to support a large-scale data visualization and asynchronous collaboration at the internet-scale.

Hence, collaborative visualization is classified into different scenarios based on the setup and the style of collaboration. According to the space-time matrix shown in Figure 1.1., there are four scenarios of collaborative visualization. Each setting requires specific design considerations and requirements.

Both synchronous and asynchronous visual analytics need special considerations due to the unique requirements for each setting. Work partitioning across space and time in asynchronous collaborative settings provides scalability yet introduces new challenges. Heer and Agrawala (6) defined a set of design considerations that identify important aspects for achieving effective collaboration in visual analytics settings. Those aspects with regards to asynchronous collaboration are important to increase the collaboration awareness and work engagement during asynchronous visual analytics. However, asynchronous visual analytics is out the scope of this research, therefore, I focus here on some design principles for co-located synchronous visual data analysis. Other efforts have been made to identify the requirements and design considerations for specific settings such as collaboration around tabletops (21) and collaboration in multi-display environments (22).

Petra et al. (20) presented an overview of collaborative visualization scenarios and their associated challenges. They pointed out that designing for each of these settings should handle specific technical and social challenges. The technical challenges arise from the designing and the implementation of the physical and the digital environment. It should address and differentiate appropriate aspects of group work. The physical environment brings addition challenges unique to the type of the environment, either a large-display, tabletop, or multi-device environment.

2.2 Interactive surfaces for Information Visualization

Analyzing data that comes from different sources and domain requires multiple analysts from different background to work together in order to understand data and derive an insight. Therefor, there has been an increased interest in developing frameworks that go beyond a single desktop for visual data exploration and analysis. Petra et al. (23) , in their research agenda on visualization and interactive surfaces, stated the advantages and opportunities that multi-device environments offer for visualization. These include:

- Analysts have larger space than what one device can offer, to visualize and work on more data.
- It allows the distribution of the data to the appropriate device for visualization.
- It allows different collaboration styles by enabling individual and group work.

2.2.1 Literature Themes

A rich body of research investigated different aspects of designing visualization tools for multi-device environments. In the beginning of this research, I surveyed the recent research articles about visualization tools in multi-device/interactive surfaces. After a closer look at these publications, I found that they fell into few categories. The majority of these publications address the development/comparison of interaction techniques for interactive surfaces. The second research focus is the development of specific physical setups for visualization tools or applications for specific domains. Few publications addressed aspects of users' collaboration around interactive surfaces.

2.2.1.1 Interaction beyond mouse and keyboard for InfoVis

Different interaction mechanisms were proposed to facilitate exploration of visualization on interactive surfaces. Instead of using traditional mouse and keyboards, natural and direct interactions were used. Chegini et al. (24) presented a set of touch-based interactions for collaborative exploration of scatter plots on large displays. It enables multiple people to interact with visualization at the same time using different techniques for manipulation.

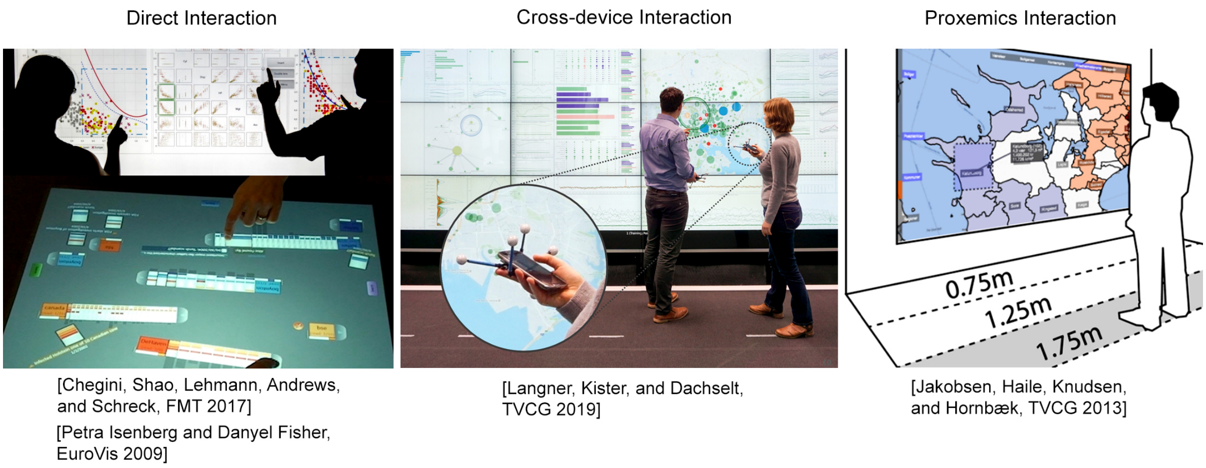


Figure 3: Themes of Interaction for information visualization.

Cross-device interaction examines the design and the development of interaction techniques that leverage the combination of different devices. Due to their popularity and portability, tablets, mobile phones and smart watches have been integrated with large displays and tabletops to steer the interaction and the visual exploration. By leveraging each device's display and input modalities, they provide fluid interplay between them to support the visual data analysis tasks (16). Langner et al. (25) examined interaction techniques for multiple-coordinated views on large displays. They found that interaction from distance using mobile devices offer flexible movements, which is essential for collaboration and perception of many visualization at the large display.

Due to the physical nature of large displays and other devices, many approaches considered the space in front and around interactive surfaces for interaction. An interesting possibility for that is the use of proxemics. Jakobsen at al. (26) studies the possibility of using body movements to drive interaction with visualizations. They developed proxemics-based interaction techniques as input for visualization manipulation. In their approach, they used the spatial relations among people and visualization as an input for visualization exploration. In VisTiles presented by Langner et al. (27), they instead used the spatial relations among devices to steer interaction. VisTiles utilized the portability and dynamics of mobile devices to enable flexible layout and distribution of coordinated multiple views. Therefore, it aids a user-friendly interface. The coordinated multiple views can adapt to the spatial arrangement of devices enabling new visualization composition and exploration of multivariate data.

2.2.1.2 Setups development beyond a single desktop for InfoVis

Other frameworks investigated the composition of multi display environments to utilize the capabilities of heterogeneous devices, and extend the visual space for visual data exploration. Towards this goal, Badam et al. (11) presented the software of Munin that was developed to unify the composition of multi device environments through a service-based model. It envisions the anytime and anywhere visual data analysis. Through the service-based model, a user can specify the physical setup, input, output, and visualization services for the assembled devices.

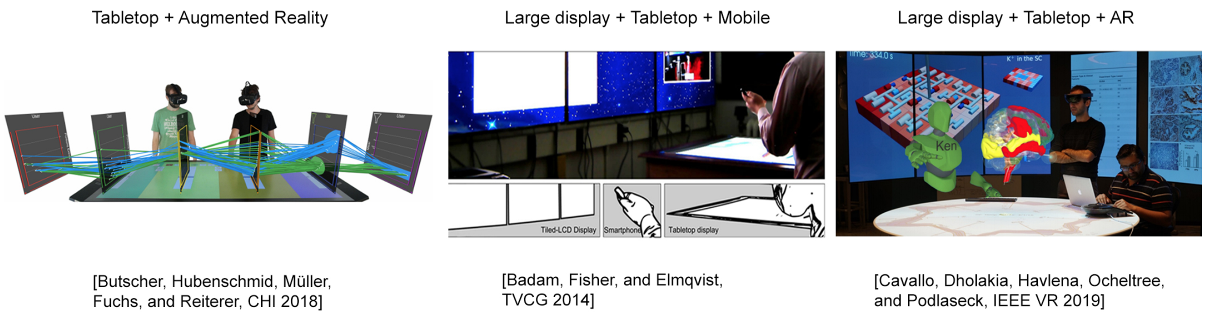


Figure 4: Examples of setups beyond a single desktop for information visualization.

Other physical setups were presented in the literature for the visual exploration of multidimensional data. These systems investigated the opportunities that the new technologies offer for visualization. Butscher et al. (17) and Cavallo et al. (28) presented new design spaces for visual data analysis through the use of immersive technologies. These systems enable an immersive collaborative analysis of multidimensional data where users can immerse themselves into the data.

2.2.1.3 Collaboration beyond a single device for InfoVis

The last theme of work addressed aspects of users' collaboration such as collaboration styles, territoriality, processes, and coupling and decoupling of work. I conducted a search-based survey on publications that their main research focus was around collaboration. I chose the major visualization and interactive surfaces venues such as IEEE VIS, EuroVis, ACM CHI, ACM ITS, CoVis, and IV.

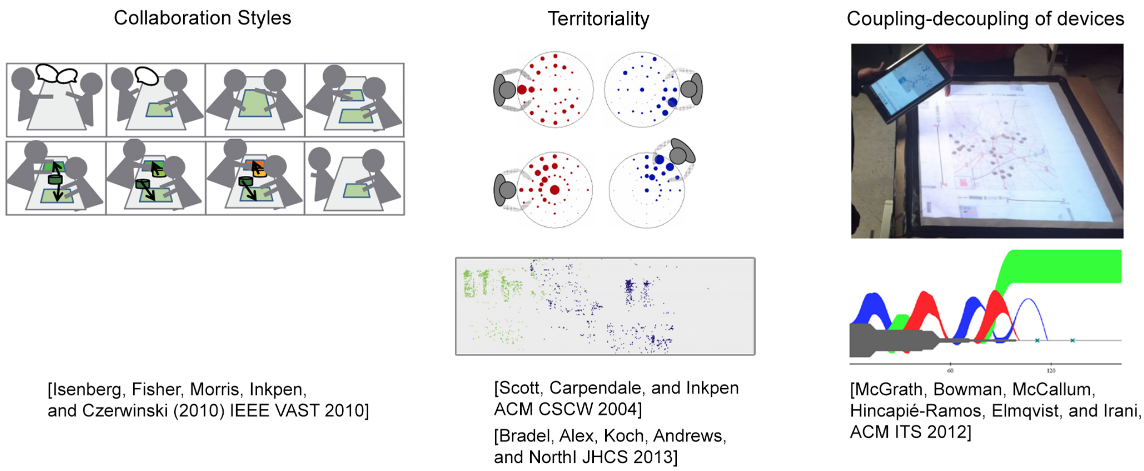


Figure 5: Studying collaboration beyond desktop environments.

Figure 6 lists the surveyed publications, their setup, and research focus. This theme of work is the most related to the work presented in this research, although the goal of this research is not only to understand the collaboration in multi-device environments but also to provide tools support that enhance this collaboration. The work presented here differs from the work presented in the literature in two aspects. First, we address the flow of the analysis process from the dimension of analytical flow and structure with regards to the use and formation around multiple devices. Second, we study this problem in a dynamic environment of multiple devices including large display, laptop, tablets, and AR headset.

			Physical setup					Visualization type				
Author	Year	Group's members	Mobile	Tablet	Laptop	Tabletop	Large display	Map	Charts	Graphs	Other	Research Focus
Mahyar et al.	2009	3				■	■		■			Note taking
Mahyar et al.	2010	3				■	■		■			Processes and note taking
Isenberg et al.	2010	2				■					■	Collaboration styles
McGrath et al.	2012	3		■		■		■				Pattern of coupling and decoupling
Wallace et al.	2013	4		■		■			■			Breadth vs. depth sensemaking
Bradel et al.	2013	2					■				■	Territoriality
Chung et al.	2014	3		■		■	■			■		Collaboration and cross-device sharing
Mahyar and Tory	2014	3			■					■		Communication and coordination
Mahyar et al.	2016	4-5		■		■	■	■				Public engagement
Langner et al.	2018	2	■				■	■	■			Movements and distant interaction

Figure 6: Publications that their main research focus is collaboration around visualization tools and interactive surfaces.

Mahyar et al. (29) (30) studied the collaborative visual data analysis around large displays with the focus on the record-keeping activity. They identified how and when users keep notes and charts during the analysis. They analyzed the use and the contents of those saved items. Then, they classified the analysis activities into five categories. The presented framework of the analysis activities encompasses the record-keeping as a main activity that takes place along with all other activities. As their work focused merely on the analysis task, this research studies the analysis task and its structural characteristics that was shaped by the dynamic social and digital interaction. This structural definition informs the further support of the analysis process.

Isenberg et al. (31) performed an exploratory study to observe collaboration styles of pairs around a single large display. They identified eight collaboration styles of team's work around the tabletop. These styles classified as close and loose collaboration. Their work investigated the styles of collaboration around single device where multiple devices bring more dynamic collaboration as we observed in this work.

McGrath et al. (15) proposed Branch-Explore-Merge protocol to support the coupled and decoupled visual data exploration in an environment of tabletop and tablets. Portable devices, i.e. tablets, allowed for private exploration and merging of results onto the shared space, and hence, the branching and merging protocol facilitates flexible levels of exploration territories. Chung et al. (18) studied the sharing and organization of information entities across devices through gestural interaction. Their work addressed users' collaboration around devices and the organization of information entities while the work presented in this research addresses the task flow from the dimension of analytical flow and structure with respect to the use and formation around devices.

Understanding territoriality on interactive surfaces is essential for collaboration and communication. Some studies focused on the collaborative use of devices and territoriality on large displays (32) and tabletops (33). Bradel et al. (32) explored how pairs of participants used the large display during a collaborative analysis task of textual data. They observed that the large space offered by the large display allowed participants to construct different spatial schemas. Users used areas on large display to non-verbally communicate space ownership. They created territories for private and shared use, in addition to territories for storage. This highlights the importance of supporting the individual and group work on collaborative interactive surfaces. Lngner et al. (25) tracked the movements pattern and interaction from distance with the large display to understand territoriality of individual in the physical space. The territoriality in this research requires different evaluation metrics as it is scattered across physical and digital spaces. However, I save the investigation of physical and digital boundaries for future work.

Wallace et al. (34) investigated different displays configurations and how each setup affected the sensemaking and equity of participation. All three setups designed around a single tabletop which implies different considerations for the task and social interaction.

Other aspects of collaborative visualization for multi-device were investigated. Mahyar and Tory (35) studied the effect of linking individuals work in a virtual space using an office setup with personal desktop computers. Related to collaboration, Mahyar et al. (36) studied the iterative design of multi-device urban planning environment that engage a broad range of stakeholders. Sarvghad et al. (37) presented the notion of dimension search space visualization.

Briefly, this research differs from previous work by addressing the analysis process from the dimension of analytical flow and structure with regards to the use and formation around multiple devices. The resulted characterization of the analysis process informed the design of the proposed approach of visualizing the dimension search space. The novelty of the proposed approach is identified by two features. First, the differentiation of individual and group search space. Second, the guiding of the analysis process through the realization of the analysis paths patterns.

2.2.2 Roles of Devices

The goal of utilizing heterogeneous devices for visual data analysis is to leverage their different capabilities and strengths during analytical activity. Here I review their potential roles in supporting visual data analysis.

2.2.2.1 Portable Devices

Portable devices such as phones, tablets, and smartwatches are small personal devices mostly used privately by their users. Portable devices have been adapted to serve as a secondary displays for different needs. They were used as a controller for interaction in front of large displays (25) allowing multiple users to interact with the large display form distance. In addition, they can take the role of a private display in collaborative settings. Users can branch from the public work to due their analysis and merge finding later (15). Smartwatches are special type of portable devices and they are in fact lightweight and wearable devices. They are non-intrusive devices allowing users to be hands free and able to interact with other devices. Due to their limited display capabilities, they require special considerations for visualization design (38).

2.2.2.2 Personal Computers

Personal computers have been widely used in offices and homes. Their input capabilities using mouse and keyboard are familiar by the majority of users. In fact, traditional visualization tools were mainly designed for the desktop setup, supporting its input and output capabilities. Personal computers are beneficial for personal visual exploration of data.

2.2.2.3 Large Displays

The large space offered by large displays serves as a canvas to visualize multiple visualizations and juxtapose them for analysis. They have been demonstrated to positively affect the visual data exploration by providing a large "space to think" (39). In addition, the large space enables multiple analysts to interact with them at the same time. They serve as a shared display that is accessible to everyone. Large displays also support natural an intuitive interaction metaphors such as touch and speech interactions, increasing their interactivity and support for multi user interaction.

2.2.2.4 Immersive Displays

The notion of immersive displays are more broader than Augmented Reality (AR) and Virtual Reality (VR) displays. For example, very large displays with large field of view qualify as immersive displays. Nevertheless, for simplicity, here I refer by immersive displays to AR and VR displays. Big Data characteristics required non-traditional means to support the limited human ability to extract information and gain knowledge from the data. AR and VR are one of promising techniques to support the challenges of big data. They are suitable for the limited perception capabilities of the human brain. VR displays showed better exploration of data that holds spatial dimensions. They have been used as an interactive and collaborative platforms for scientific visualization (40) and visual data exploration (41) moving from traditional visualization of 3D data on 2D screens.

2.2.3 SAGE2

SAGE2 (2), the successor of SAGE (42), is a middleware developed using web-browser technologies to take multiple displays and unify them as one high-resolution workspace. It enable users to collaboratively share and display their contents on the large display (Figure 2.5).

Figure 7: User collaborating during a SAGE2 session where they share digital contents (i.e. PDFs, images, etc.) on the large display. (2)

Display clients provide information of the corresponding viewport in the workspace via their URLs. Any number of displays on different systems can be joined to form a unified view of the SAGE2 workspace. SAGE2 native applications are written in JavaScript using SAGE2 API. Applications open simultaneously on the large workspace enabling users to collaboratively interact with them. Users interact with the workspace through UI clients running on their devices using a SAGE2 pointer, which is an html element that collects the native mouse events and propagates them to the corresponding display client for handling. Due to its distributed application and event model, all users input events are passed to the head node server which in turn distributes them to display clients for handling. Each display client has its own instance of running applications and receives events to handle them consistently. In this system, I integrate the SAGE2 large display with portable devices of different modalities like tablets and AR headsets to create additional visual exploration territories. Coupling and coordinating with different devices requires middle modules for data sharing, translation and synchronization due to different platforms inter-dependency. To tackle this issue, I developed the PolyVis framework based on declarative visualization design and operation transformation (OT) for seamless migration of visualizations and their interactivity between devices.

2.3 First Set of Design Principles for MDE

2.3.1 (D1) Device agnostic visualization sharing

Generally, there are two ways to develop visualizations. One is a native development for a specific platform, and the other is a web-based development. Unlike native applications, web-based applications can be deployed to any device using web technology. Many frameworks and toolkits were developed based on web technology like D3 (43) and JavaScript InfoVis Toolkit (44) to support information visualization applications. PolyChrome (12), Vistrates (45) and Visfer (46) are all web-based frameworks developed to support the collaborative visual analysis. However, sometimes, going natively cannot be avoided when working with devices like AR/VR headsets. In addition, native applications are essential to take the full advantage and support of the target device. Going with one way is not enough to support all applications and user requirements. To close this gap, solutions for cross-platform infrastructures are essential (47). Grammar-based representation of visualizations has been introduced in many works with various levels of abstraction. They provide a mechanism to define visualization interdependently from rendering platforms. Examples of these declarative languages include Vega (48), Vega-Lite (49), ggplot2 (50), and ggvis (51). In addition, PolyChrome (12) adapts a centralized server to maintain concurrent web-based visualization exploration by pushing DOM events between browsers. DOM Events are wrapped into a global space and inverted on the target display to support different display sizes and configurations. This mechanism is called Operation Transformation (OT) and it is originally developed to maintain concurrent use and consistency in text editing tools (52).

2.3.2 (D2) Support of parallel and joint activities

The style of collaboration between participants is affected by the display setup, the problem under investigation and the analysis metaphors. Studies showed that collaboration around interactive surfaces for information visualization in co-located settings takes the forms of completely independent, partially independent and joint (coupled) work (9)(53). Other studies by Isenberg et al. (54)(31) identified the styles of collaboration as a spectrum that varies from loosely coupled to tightly coupled. These findings emphasize the importance of supporting individual and group work, and efficient transitions between styles. Another aspect that is related to the style of collaboration around interactive surfaces, is the use of the space. Territoriality, which is the spatial coordination of collaborative work, also takes three forms as identified by Scott et al. (33). Users use the space for personal work, group work and for storage.

2.3.3 (D3) Fluid cross-device interaction

Spreading visualizations and the analysis tasks to multiple devices requires intuitive cross-device interactions. Information sharing and management should not distract users from the actual analysis. Embodied interactions (55) leverage the proximity of devices to develop interactions that carry out these operations. Badam and Elmqvist (46) presented a cross-device interaction technique for data sharing in ubiquitous environments based on a design elicitation study. The interaction technique leverages the physicality of the devices, to effortlessly share visualizations across devices using a built-in camera and embodied QR codes. In VisPorter (18), gestural interaction was utilized to transfer information across displays in an intuitive and direct way. Their approach was based on the concept of physical references of shared information, rather than using symbolic references such as IDs and URLs.

2.3.4 (D4) Exploiting the physical space

Utilizing physical space is essential in scalable visual data analysis. Andrews et al. (39) showed that analysts exploit the spatial affordances of large displays to serve as an external memory and as a semantic layer for spatial data layout and organization. In collaborative settings around tabletops, users frequently move and organize information to approach their analysis tasks (54). Multi device ecologies enable users to carry information and form dynamic exploration territories across displays that populate the physical space. The view and the analysis process can be extended to span multiple exploration sites across the physical space.

The affordances of the physical space enable the flexible configuration and coordination of devices to approach the task. In addition, physical space is essential to embody information and immerse users in their data.

2.4 Modeling of Visual Data Analysis

Neumann et al. (56) presented an information visualization framework describing collaborative activities in information visualizations context. The presented framework is derived from an exploratory study that was designed to understand the process of collaborative visual data analysis around tabletop display. They contributed an evolving understanding of this process and informing earlier models of information visualization. Brehmer and Munzner (57) reviewed the literature on visualization tasks and derived a multi-level typology of visual analysis tasks. The typology comprises why and how a task is performed, and what are the input and output to complete it. It helps to express high-level tasks as sequences of low-level tasks. Lam et al. (58) presented a framework based on a review of 20 design study papers to describe the high-level analysis goals and how they can be achieved with low-level tasks identified from the review. As in (57), the high level context of analysis goals helps to interpret the low level actions.

Several studies modeled the behavior of users during exploratory visual analysis as a set of states. Reda et al. (59) used a Markov chain to model the transition between different cognitive and computational processes. The weighted transition between processes states help to understand user analytical behavior and predict future interaction. Sarvghad et al. (60) defined the analysis states during the analysis session as newly created visualization with new set of attributes. The same definition of analysis state is used in Voyager (61) and Voyager2

(62). Unlike other categorization of the analysis, they defined the analysis as a navigation of the dimension space where the analysis state changes by the change in the data dimension space.

The previous representations of the analysis as sets of states captures the complex flow of the analysis but it doesn't provide a full understanding of the analysis structure. Battle and Heer (63) reviewed the literature on exploratory visual data analysis and identified a set of assumptions regarding analysis performance, goals, and structure. After evaluating those assumptions through analytic provenance in Tableau, they synthesized a definition of exploratory visual data analysis contributing an understanding of it structure. We further add to this definition by presenting a two level categorization of analysis structure synthesized from observations of collaborative visual data analysis sessions.

CHAPTER 3

POLYVIS: DESIGNING FOR VISUAL DATA ANALYSIS IN MULTI-DEVICE ENVIRONMENTS

Parts of this chapter were previously published as: Alsaiari, A., Johnson, A., Nishimoto, A., "PolyVis: Cross-Device Framework for Collaborative Visual Data Analysis", In the Proceedings of *2019 IEEE International Conference on Systems, Man, and Cybernetics* (IEEE SMC 2019), October 6-9, 2019, Bari, Italy.

3.1 Introduction

Visual analytics encompasses a large amount of data that comes form different sources and domains. Therefore, collaborative visual data analysis has wide application across domains to enable multiple users (often called analysts) to work together to collaboratively contribute their contextual knowledge and deepen their understanding of the data. The heterogeneity of datasets and the need for multiple analysts to work together demanded solutions that go beyond the single desktop (64) (65). There has been a shift to big and multi-surface interfaces for visual data analysis. Tiled wall displays have been shown to increase the performance of visualization tasks (66) and the productivity of exploratory visual analysis (67). In recent years, spreading to multi-device settings for co-located collaborative visual data analysis has emerged to leverage different devices capabilities (17) (27).

Designing multi-display interfaces that combine multiple devices for collaborative visual data analysis faces multiple challenges. They should support key principles for effective collaboration. First, the ability to share visualization between collaborators and devices is important to support different collaboration styles. Collaborators should be able to share visualization between different devices. However, cross-device visualization sharing requires the development of flexible visual representations that can seamlessly migrate between devices regardless of the rendering platform. In addition, cross-device collaborative systems should allow simultaneous interaction with visualization. Cross-device simultaneous interaction can be a grand challenge due to platforms disparity. An interaction (e.g. touch) on a specific device should be interpreted in other synced platform to execute the same action (e.g. click).

This chapter addresses the above mentioned challenges by introducing the design and implementation of a multi-device system for collaborative visual data analysis that enables cross-device visualization sharing and simultaneous interaction between devices. I integrate SAGE2 large display with portable devices (laptop, tablets and augmented reality headset) for co-located visual data analysis. The system implements a front-end multi-display user interface and a networked communication and coordination protocols for visualization sharing and simultaneous interaction between clients devices. Each device plays specific roles according to its display modality as described in section 3.2.2. Back-end protocols for communication, sharing, interaction synchronization are described in sections 3.2.3 and 3.2.4.

3.2 PolyVis System

Below, I discuss the primary features of the framework. I refer to the design principles discussed in the last chapter (D1-D4) in the description of the framework and how the choice is made to meet these principles.

3.2.1 Overview

The proposed framework is specifically designed to seamlessly support collaborative visual data analysis that can span multiple devices of different modalities. The framework is built on top of SAGE2 middleware that drives tiled wall displays and unifies them as one high-resolution display. PolyVis integrates portable devices with SAGE2 display to compose a heterogeneous visual data analysis environment enhanced with further exploration capabilities.

Earlier studies of collaborative visualization emphasized on the importance of supporting individual and group work for different collaboration styles. While the large display is a primary display, tablets enable different exploration styles. They allow users to branch from the main analysis to conduct a local exploration or to conduct a coupled exploration with other collaborators (D2). PolyVis users can join the analysis session using their tablets or phones to pull and push visualizations from the large display and do further analysis activities as will be described below.

The integration of AR/VR devices enables a different display modality. However, PolyVis only support the HoloLens AR headset for 3D immersive visualization. This choice can be justified by the fact that VR display modality requires additional considerations to be integrated effectively. Unlike VR, AR headsets can enable collaborators to exploit the physical space as an additional exploration territory without blocking them from their surrounding environment.

The system is a server/client based system where SAGE2 server handles the communication with and between several clients applications for the laptop, tablet, SAGE2 app, and HoloLens. Figure 8 shows an overview of each device capabilities. Users can iteratively filter data, specify the visual encoding, and create visualization. They can pull/push visualizations and change their visual representations using portable devices.

As users can pull/push visualization, they also can sync interaction on a view between the wall and the tablet displays. Any changes are made on the tablet will update the synced view on the wall and vise versa. The simultaneous interaction approach is described in 3.2.3. To share visualizations between devices (pull/push), PolyVis adapts the visualization declarative design as described in section 3.2.4.

Developing visualizations can be a tedious process for users with no programming skills, such as data analysts. Therefore, visualization authoring systems and toolkits have been widely adapted in recent years. The presented framework enables the rapid construction of visualizations by a visualization authoring UI following the flow of the information visualization reference model (68) in which users filter the data, specify the visual encoding, and create the visualization. Here, users play a major role in the visual mapping task that maps each data attribute onto a single visual channel.

As will be described in the next chapters, groups of three participants used the system to perform exploratory visual data analysis of geoscience data sets.

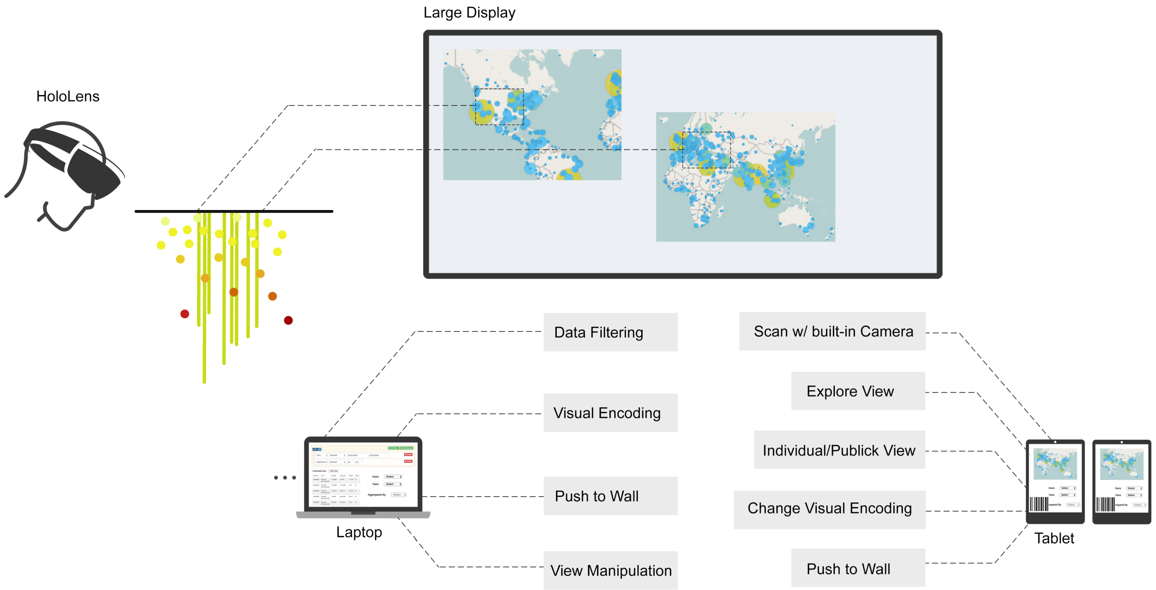


Figure 8: An overview of each device capabilities.

3.2.2 Physical Environment

Different devices like smart-watches, phones, tablets, laptops, large displays, AR and VR headsets became common display metaphors for information visualization. Some of these devices like smart-watches and VR headset require unique design considerations due to their field of regard either a very small or a very big. Therefore, I limited my scope to support the integration of portable devices that vary in between like tablets and the HoloLens AR headsets. Any number of mobile devices with a built-in camera and web browsers (i.e. tablets and phones) can be joined to pull and push visualizations from and to other devices (D3). The HoloLens client device extends the exploration into the third space. Each distinct HoloLens client should run on a separate machine. While theoretically the system can support a larger number of devices clients, I used one laptop, two tablets and one HoloLens in the conducted studies.

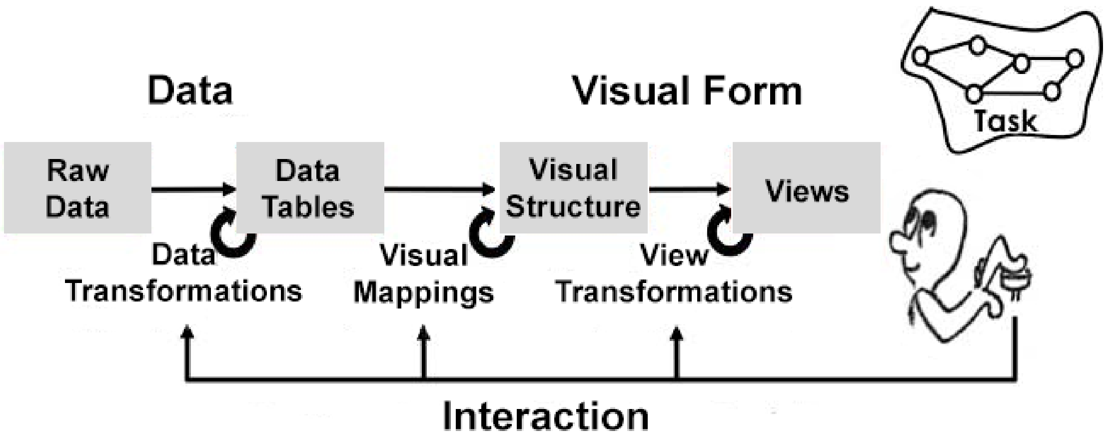


Figure 9: Visualization reference model (3).

3.2.3 Cross-device Visualization Coordination Approach

PolyVis can enable coupled exploration style by coordinating views between two devices (D2). That is, users can link a local view on the tablet with a global version on the wall and interact with them simultaneously. To allow coupled exploration styles between devices, the views on both devices should be synchronized. When it is needed to maintain visualization coordination, the system should synchronizes visualization state between client devices, regardless of rendering platform, to ensure that collaborators see the same data.

In this section, I discuss the choice of the selected coordination approach. To understand how to support cross-device visualization synchronization, we need to analyze the structure of visualization applications in order to specify the possible places for synchronization in different platforms.

According to the visualization reference model (Figure 9) , the "*view transformation*" converts the ***"visual structure"*** into a ***"view"***. To maintain the same view across all clients, one possible way is to synchronize the view itself. However, it is difficult to sync the view itself (and its rendering components like SVGs) due to different rendering environments across platforms.

Another way to coordinate views across all clients is through interactions. As shown in Figure 9, the user can control the parameters of the views through *interactions*. User interaction with the view triggers an action (program logic) that updates the view. Therefore, interactions will trigger the view transformation cycle that updates the view. So any interaction triggers an action that updates the visualization should be triggered in all coordinated views. Synchronizing interaction is more generic than synchronizing the view itself.

Nonetheless, synchronizing interaction is not straightforward. Low-level interaction events can be variant in different platforms. Click event for example corresponds to pinch event in HoloLens. While they differ in their representation, they posses a similar activity semantic. Gotz and Zhou (69) characterized user's visual analytic interactions into a multi-tier activity model based on their semantic richness. The bottom tier of their 4-level model is the low-level interaction events which have little meaning without context such as click, mousemove, etc. The second tier is the actions tier. Actions are meaningful units that can be achieved by one or more low-level events such as brush, filter, inspect, etc. While they posses a richer semantic than low-level events, they are generic in visualization tools (69).

3.2.3.1 Interaction Synchronization

The interaction synchronization approach presented in this chapter is based on the actions tier. All interactions are wrapped into predefined rich semantic visualization actions and shared with peer clients for synchronization. The interaction synchronization API interprets and executes (triggers) the action in the target device for coordination.

The representation of rich semantic action captures its properties and parameters. Similar to (69), a rich semantic actions are defined as:

$$\mathsf{Action} =< \mathsf{Id}, \mathsf{Type}, \mathsf{Parameters}[\mathsf{value}, \mathsf{valueType}, \mathsf{deviceId}, \mathsf{timestamp}] >$$

Where the type represent the type of the action and the parameters hold the values to execute this action.

Similar to PolyChrome (12), the server is used to maintain the global state between all clients. Low-level events are wrapped into a global space (actions) and inverted on the target display. The framework encapsulates coarser interaction operations instead of low-level events, so they can be shared and inverted by the target device. There are four types of actions that are supported for synchronized interaction. These are inspect (details-on-demand' for a visual object), brush (highlighting a subset of visual objects), pan (scrolling a visualization), and zoom (scaling a visualization). Although the tablet client is written in JavaScript similar to

SAGE2 applications, the coordination layer is necessary due to the difference in interactivity handling between SAGE2 applications and other JavaScript-based applications.

3.2.3.2 Visualization Persist State

The second challenge that PolyVis addresses is sharing visualization between devices as will be described in the next sub-section. However, it is especially important in collaborative settings to share visualizations in their current state. Therefor, PolyVis maintains visualization state after interaction.

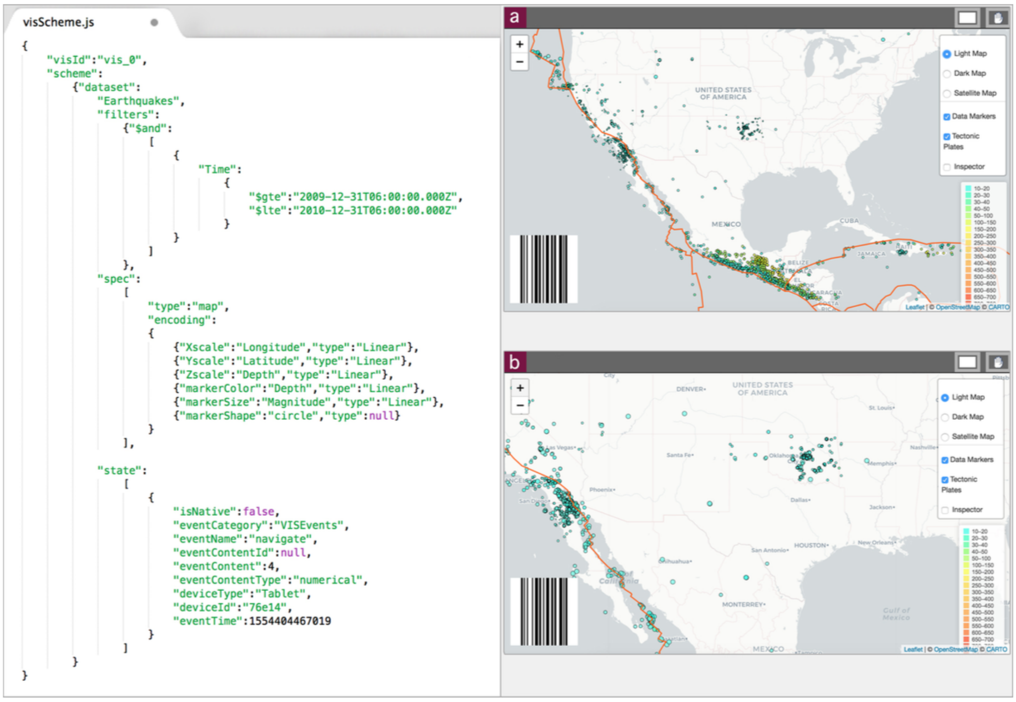


Figure 10: An example of a visualization scheme structure. (a) visualization at initial state. A new visualization state is pushed to the scheme after an exploration event occurred in (b).

Most visualization frameworks lack the ability to capture the visual exploration state and the path that led to it. The most challenging aspect is how to capture the visualization state. From the visualization task perspective, interactions in visualization can include a set of low-level events, such as brushing interaction which is composed of the events: mouse-down, mouse move and mouse up. Do we consider the visualization state after each low-level event or after a richer semantic interaction that is composed of a set of low- level events?

Figure 11: The visualization declaration scheme can span different devices for rendering: (a) large display, (b) tablet, and (c) HoloLens.

The state definition needs to be identified first before any effort to capture it is made. As discussed in the last section, I define operations as interaction-centric operations. To enable consistency between different platforms, I chose to define the visualization state based on semantic rich interactions. I enable client side maintenance of a persist state. The state is recorded as the user interacts with the visualization. I defined an intermediate layer to record and push the state to the visualization scheme. When the visualization is shared, the state is recovered according to the device-dependent interactivity and visual channels encoding.

3.2.4 Visualization Sharing

As I discussed in the previous chapter, visualization development can be either native to a specific platform like AR/VR headsets or it is web-based application that can be deployed to any device using web technology. Grammar-based representation of visualizations (i.e Vega-Lite(49)) has been introduced in many works with various levels of abstraction. They provide a mechanism to define visualization interdependently from rendering platforms. For visualization sharing between different devices, I treat visualizations as user-configurable semantic units (D1). I use a grammar-based representation of visualization to represent the visualization semantic.

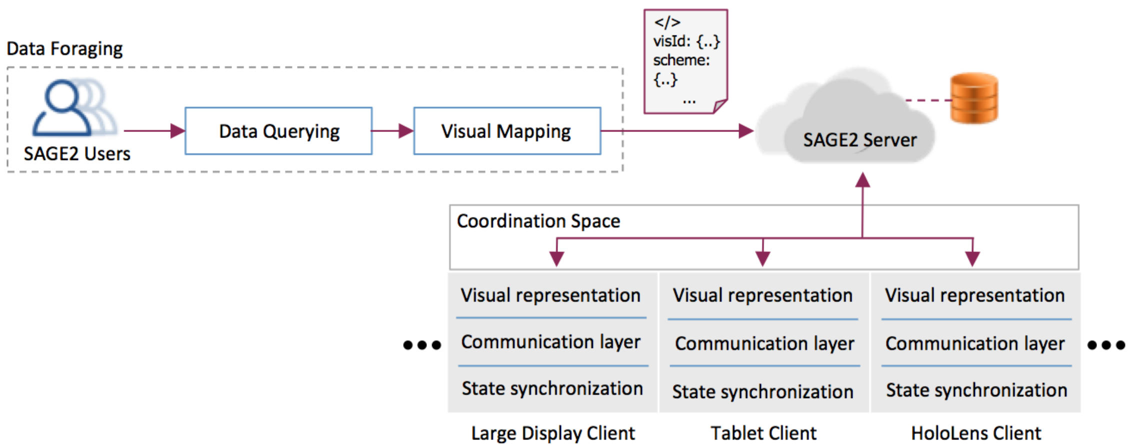


Figure 12: An overview of the system components. A visualization scheme is defined by the user through a set of filtering and visual encoding specifications. The server coordinate the spanning of the scheme to the target device and coordinate the event wrapping and sharing between devices.

Unlike other grammar-based applications, I assume a dynamic visualization scheme that gets updated with user interactivity with the visualization. I employ an all-in-one JSON format to declare three main components of the visualization in our framework.

These components are: query specification for data retrieval, visual encoding channels, and interactivity state. I capture those components during user composition of visualization. The interactivity state is captured automatically using our persist state mechanism and update the scheme accordingly. Figure 10 shows an example structure of these components. Decoupling the visualization semantic from its view transformation process enabled a seamless migration of visualizations across devices (D1).

Using this approach of declarative visualization design, visualization can be shared between devices regardless of the rendering platform. To share a visualization, the application shares the visualization scheme with peer client. The view transformation is delegated to the target device for rendering (Figure 11).

3.3 Evaluation

To evaluate the use of the prototype system for the visual data analysis of real world datasets, I conducted a collaborative session with two visualization researchers. Here, I outline the data analysis scenario and discuss feedback from experts.

3.3.1 Collaborative Scenario

Two researchers with a background in visualization, one has additional experience using immersive technologies, conducted a visual data analysis of two geosciences datasets. For reference, I will refer to the users as U1 and U2. The users performed a visual analysis task to ascertain the relationship between injection volume, the pressure of fracking wells and the frequency of earthquakes in Oklahoma State. The first dataset contained information about earthquake incidents in Oklahoma and California from the years 2000 to 2010 (Appendix C). The Wells dataset contained information about the fracking activities in Oklahoma and California also from the years 2000 to 2010 (Appendix C). The earthquake dataset consisted of 24555 records and 12 attributes while the Wells dataset consisted of 5138 records and 9 attributes. These datasets have attributes with similar meaning such as the location, the time, and the depth. The earthquake dataset was provided courtesy of http://service.iris.edu/ and the Wells injection dataset was provided courtesy of http://www.occeweb.com/.

The users started with the question: Is there any correlation between the injection volume of wells and earthquake events? U1 began by mining the data for all earthquake events during 2010 and then he visualized them on a large display map. He also created a map of the locations of active well during 2010. U2 captured the barcode attached to the map of earthquakes by using the camera of the handheld device to pull the map visualization and performed analysis of the mapped data. He created a line chart to plot the frequency of earthquake events over the year and pushed the chart to the wall. They observed an increase in the number of earthquake incidents during the month of December.

To investigate the temporal relationship with injection volume, he moved to the map of wells and captured the attached barcode. Then, he created another line chart of total volume injection per month. A pattern is observed, so he pushed the chart to the wall and started to discuss with U1. They observed an increase of volume injection during the month of November, which

Figure 13: In a collaborative session, the user on the left is examining data in 3D using a HoloLens device. Data points (Wells) within the blue rectangle on the left map are viewed in 3D via HoloLens. The other user on the right is using a tablet (with linked visualization) to inspect specific areas on the right map.

has no temporal relation with the increase in earthquake events, but they made a hypothesis: can a high volume injection cause an increase in earthquake frequency for the next month? U2 used the HoloLens to examine the relative depth of the wells compared to the depth of the earthquakes. They concluded that an additional investigation of the observed pattern is needed for different years and probably for different states to test their hypothesis.

3.3.2 Expert Feedback

I collected feedback from the experts regarding the usage of the system for visual data analysis and the benefits of integrating different devices into the process of visual data analysis. U1 mentioned that the use of the tablet gave more freedom of movement, obtain the data they want, process it and push it back. He also believes that this will allow different people to focus on different things of the analysis process. Because of the affordance of portability, both users mentioned that it would be beneficial to use the portable devices as a controlling metaphor to control visualization on other devices (i.e. tablet to control a visualization on large display or on the HoloLens). Controlling here is different than coordinating or linking visualizations. In this context, it means moving visuals around, minimize or maximize them, etc. U2 mentioned that it is useful to view datasets in 2D on the large wall and in 3D on the HoloLens, but the hardest part is to determine what the HoloLens user is seeing. As U1 used the HoloLens to view the data in 3D, he added that it also needs a kind of representation on the large display or any mechanism that would increase the awareness. Experts gave good feedback on how the devices are complementary to each other.

3.4 Conclusion

In this chapter, I presented the PolyVis framework for the building and promoting of visualizations in multi device environments. It supports visual data exploration by utilizing multiple devices of different modalities. The primary goal was to maintain consistent sharing and interaction with visualizations across different platforms. To achieve this, I relied on the declarative visualization design and the operation transformation paradigms. I treat visualizations as semantic units (in the form of grammar) to migrate to and render by different devices. SAGE2 users assume a major role in the composition of visualization grammar without any need for programming skills. The interactivity with the visualization is captured and stored in a global space for consistent representation. Therefore, the state of the visualization will be maintained as the data analysis proceeds regardless of the processing device. There are a few areas that

I plan to improve in the future. First, the visualization layers at each device only support few visualization types. I plan to extend that to support more advanced types of visualization such as multi lines, stacked bars, parallel coordinate, node-link, etc. I plan also to support the 3D version of these types on the HoloLens client. In addition, as suggested by experts, I would like to implement a mechanism for cross-device multi-coordinated views. With multiple visualizations at a time, it would be beneficial for the visual exploration to connect data points across scattered views.

CHAPTER 4

UNDERSTANDING COLLABORATIVE VISUAL DATA ANALYSIS IN MULTI-DEVICE ENVIRONMENTS

Parts of this chapter were previously published as: Alsaiari, A., Johnson, A., Nishimoto, A., "PolyVis: Cross-Device Framework for Collaborative Visual Data Analysis", In the Proceedings of *2019 IEEE International Conference on Systems, Man, and Cybernetics* (IEEE SMC 2019), October 6-9, 2019, Bari, Italy. AND as: Alsaiari, A. and Johnson, A. (2019). "Towards Understanding Collaborative Visual Data Analysis in Multi- Device Environments". In *2019 IEEE VIS*.

4.1 Introduction

The work of this chapter addresses the questions: **what is the complex picture of users' experience during a collaborative visual data analysis in a multi-user multi-device environment?** and **what is the characterization of the analysis process?**

Collaborative visual data analysis is a complex process. There are several factors add to this complexity. As I discussed in Chapter 2, users and tools influence one another in system-user interaction. Therefore, the complexity of the analysis process is not influenced only by the respective task, but also by users and tools. The goal of this chapter is to understand this process based on involved factors, and identify challenges that shed light on requirements to improve the design. To achieve this goal, I conducted an exploratory study to observe how users approach the analysis task in a multi-device environment, and how this differs from one-display settings. In the first phase of the study analysis, I decomposed the problem into three dimensions. I performed a hybrid analysis approach with mixed methods to analyze these dimensions. Specifically, I analyzed the usage of the tools, the analytical activities, and the strategies of collaboration around devices. In the second phase of the analysis, I performed an in-depth qualitative analysis and provided a structural categorization of the visual data analysis process. This categorization was influenced, on abstract level, by the three dimensions of the analysis environment. The findings and observed challenges highlighted the importance of a supportive tool that unites the scattered effort. I envision this through a hybrid model of a visualization recommendation tool that cast the analysis process as an assignment problem based on the different aspects of the environment: the tasks, the tools, and the users. This motivated the future work of this research. In Chapter 5, I provided a set of design guidelines to enhance visualization tools with recommendation system that steer the analysis process in flexible and efficient way. In Chapter 5, I discussed the current designs of visualization recommendation systems and their building metrics. In contrast to previous work, my proposed model frames its recommendation metrics based on the different aspects of the environment. The design, implementation, and evaluation of this model is the core of the next phase of this research and will be investigated in more detail.

4.2 Exploratory Study

The study presented in this chapter falls into the category of user studies for "*understanding*". The goal of this type of user studies is to build a rich understanding of user experience and context of use by observing the user-system interaction. Then, this understanding can help informing the design space of such systems. Within a multi-device environment, collaborators employ different tools to perform different kinds of activities in approaching the respective task. The main question is: what are the flow patterns that collaborators follow in approaching the task? I believe that the flow of the data analysis is shaped by the following sub-questions: How do users use the tools? What types of activities do they perform? And how do they collaborate? Each question corresponds to one aspect: tools, tasks, and users. I aim to synthesize the flow patterns of the analysis process by quantitatively and qualitatively analyzing the three aspects. Subsequently, I target the categorization of how the analysis process unfolds. Through this analysis, I can define further requirements on how to provide tool support for collaborative use and coordination of analytical components. This chapter contributes a better understanding of the visual data analysis process and provides the directions for further development of information visualization tools around interactive surfaces.

4.2.1 Participants

I recruited 18 subjects, 6 groups of 3, from a pool of undergraduate and graduate students at the electronic visualization laboratory and computer science department of the University of Illinois at Chicago. Participants comprised of 13 male and 5 female students between ages of 18 and 34. They participated in the study for the duration of 45min-1.5hrs. Participants had varied experience in visual data analysis, ranging from moderate to advanced. EVL affiliated students had advanced background in visualization while participants who were not affiliated with EVL, had taken at least one EVL course and had moderate experience with visualization.

4.2.2 Software

For these studies, I used a cross-device framework that I developed for collaborative visual data analysis, PolyVis (70). In PolyVis, I integrated SAGE2 (2), a large display collaborative application, with portable devices for co-located, multi-device visual data analysis. Portable displays can vary from smartwatches to the fully immersive VR headsets. Due to the unique requirements of integrating devices from these categories for information visualization, I limited my scope to support the integration of portable devices like tablets, laptops, and HoloLens AR headsets. Specifically, PolyVis integrates SAGE2 large display with laptops, tablets and the HoloLens AR headset. It provides users with an environment for visualization compositions and sharing across displays. PolyVis also offers the capability to utilize each device for specific tasks. Some of these tasks include data filtering, visual mapping, visual representation, visualization construction and sharing. This environment allowed analysis across different devices and many visualizations with the ability to move and share them.

PolyVis usage scenario: Using a laptop or a tablet, users can start by mining the data for all earthquake events during 2010, and then specify their visual representation (i.e. map) to visualize them on the large display. Any user with a tablet can capture the barcode attached to the map of earthquakes using the camera of the device to pull the map visualization to the portable device. Analysis charts like scatterplots, line or bar charts can be created for the pulled map for analysis and then they can be pushed back to the wall. Using the laptop, the user can select a specific area on the map to view data points in 3D using HoloLens. PolyVis was developed based on a declarative visualization design like Vega-lite (49) and the paradigm of operation transformation for seamless migration of visualizations and their interactivity between devices.

4.2.3 Datasets and Tasks

Each group performed visual analysis tasks using two geoscience datasets. The first dataset contained information about earthquake incidents in Oklahoma and California from the years 2000 to 2010. The Wells dataset contained information about the fracking activities in Oklahoma and California also from the years 2000 to 2010. I collected, cleaned, preprocessed and stored datasets in NoSQL database using MongoDB. The earthquake dataset was provided courtesy of http://service.iris.edu/ and the Wells injection dataset was provided courtesy of http://www.occeweb.com/. The earthquake dataset consisted of 24555 records and 12 attributes while the Wells dataset consisted of 5138 records and 9 attributes. These datasets have attributes with similar meaning such as the location, the time, and the depth. Earthquakes dataset has other attributes like magnitude while Wells dataset has other attributes like well status, well type, injection volume, and injection pressure. Tasks were designed to ascertain the relationship between injection volumes, the pressure of fracking wells and the frequency of earthquakes in Oklahoma and California states.

Each group completed two tasks, with focused and open questions. In the first task, the subjects were given focus questions that can be answered by creating one or two visualizations. The focus question is designed in a way that helps subjects learn how to use the system and be familiar with the datasets. For example, "how does the injection volume on Oklahoma compare to California in 2010?" Subjects were then asked an open exploratory question to determine the correlation between earthquake events and wells' volume injection of two states, California and Oklahoma, from the years 2000 to 2010. They were asked to create as many visualizations as they needed with no restrictions regarding using devices or moving in the space.

4.2.4 Setup and Data Capture

The study was conducted in a room space of approximately 10.61 by 5.59 meters, equipped with a high-resolution large display. Overall display size is approximately 7.3 by 2.05 meters at a resolution of 11,520 by 3,240 pixels. Other portable devices were placed on a table in the middle for use during the study: one MacBook Pro (macOS Sierra, 2.4 GHz Intel Core i5), one 8" Samsung - Galaxy Tab A (32GB, Android 9 (Pie)), one 10" Samsung - Galaxy Tab A (64GB, Android 9 (Pie)), and one Microsoft HoloLens 1 (Windows Mixed Reality OS, Intel 32-bit (1GHz) CPU, 2 GB RAM). There were no chairs provided in the working area as shown in Figure 14. Each of the portable devices was attached with Mocap markers for tracking. In addition, three caps with attached Mocap markers were provided to the users.

The whole room was tracked using the OptiTrack Mocap system. The position and orientation of devices and users were streamed from the OptiTrack Mocap system to a Unity application depicting a 3D model of the physical space. We sampled the captured data at a rate of one frame per second. The unity application was running on a separate machine in the corner of the room. Systems usage logs were collected from all deployed devices. I wrote a script to capture all interaction events with the system. Each log included the device id and type, the action type, and the timestamp. System logs will be used in my quantitative analysis of the device usage. The study was video and audio recorded using two cameras, one showing

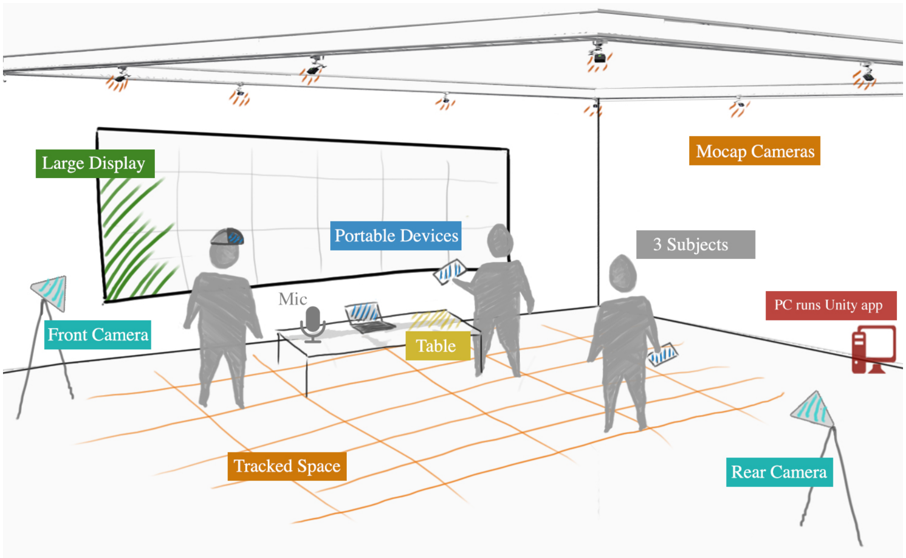


Figure 14: Illustration of the study setup.

the full room from behind and one showing the subjects' interaction with the large display from the front. The setup is pictured in Figure 14.

4.2.5 Procedure

First, participants were greeted and provided with consent and media forms. Then, they spent 2-5 minutes to read and fill out forms. Once participants finished completing forms, I started with a 5-minute introduction to give the users an overview of the software and tools. Next, participants were given the first task of multiple focus questions that could be solved by creating one or two visualizations to familiarize them with the software and tools. I opted for this approach as a practical tutorial on how to use the system. They were told to feel free to ask for a clarification or instruction at any time during this task. They spent 20-30 minutes on this task. Next, they started the main task of an open exploratory question to find the correlation

Figure 15: Subjects examining a set of created visualizations while using different devices. Position and orientation of subjects and devices are streamed from OptiTrack Mocap system to a Unity application depicting a 3D model of the physical space.

between earthquake events and well's volume injection in two states, Oklahoma and California. Participants were told to work on the task in any way they preferred and to create as many visualizations as they wanted. I left it to the participants to find their own way for completing the task. This task was exploratory in nature and took between 25 to 70 minutes to complete. Upon completing the main task, participants spent 2-5 minutes to fill out surveys about their experience in the study.

4.2.6 Coding and Data Analysis

I collected data in the form of recorded videos, system logs, tracking data, and questionnaires. About 420 minutes of videos were collected (an average of 70 minutes per session, 41 minutes for the main task). I divided the analysis into two parts and in each part I performed multiple coding passes. In the first part, I focused on analyzing tools usage, analytical activities, and user collaboration styles around devices while in the second part the focus was on the flow of the analysis process. The main goal of this study is to identify how we can provide supporting tools to facilitate the collaborative analysis process; therefore, it is crucial to understand how the analysis flowed. In addition, revealing aspects of the tools usage, the performed activities, and the users' collaboration styles shed light on requirements that should be taken into consideration when designing a tool.

For the first part of the analysis, I created an excel sheet with 5-second intervals. For each time interval, I coded from the videos the users' formation styles around devices, the tools used, and the type of use. Then, each formation style of user collaboration was classified as close, moderate, or loos. By this, I created 5-second interval logs of the collaboration styles, the tools usage, and performed activities. I converted these log files into timeline visualizations as shown in Figure 17 and Figure 18.

I started coding the analysis flow by creating a second excel sheet and then for every created visualization, I documented how it was created, its relation to previous visualizations, and why it was created. Then, I drew a flow diagram of the created visualizations in chronological order with arrows indicating the first set of visualization relationship codes. After multiple coding passes, I identified a categorization of the analysis flow as will be discussed below.

4.3 Findings

In this section, I present findings from the two major analyses conducted on the collected data. In the first part, I present results from analyzing three aspects of the analysis sessions to provide an understanding of users' activities and collaboration. In the second part, I present a categorization of analysis strategies observed in the study.

4.3.1 Understanding User Activity: A Hybrid Approach

Previous work examined three factors (users, tools, and tasks) independently. In this study, I focus on the synthesis of these factors. To do this, I adopted a hybrid analysis approach that focused on three different aspects: users, tools, and tasks. We believe these findings will help us identify associated challenges and better inform design goals in developing multi-device tools for visual data analysis.

In their book "*Acting with technology*" (71), Kaptelinin and Nardi discussed the development of Activity Theory as an approach and theoretical foundation for research in the fields of HCI and CSCW. They discussed its use in "interaction design" to understanding the human interaction with technology and using this understanding to better design technological systems. Interaction design was defined by Winograd (72) as "the design of spaces for human communication and interaction".

They ultimately framed the contribution of activity theory to HCI field as: "Activity theory provides a coordinated description of the use of technology at several hierarchical levels at the same time, and thus opens up a possibility to combine, or at least coordinate, analyses of different aspects of the use of technology, such as physical interaction, conceptual interaction, and social "contextual" interaction". The adapted approach presented in this paper was mainly motivated by their framing of activity theory in HCI field.

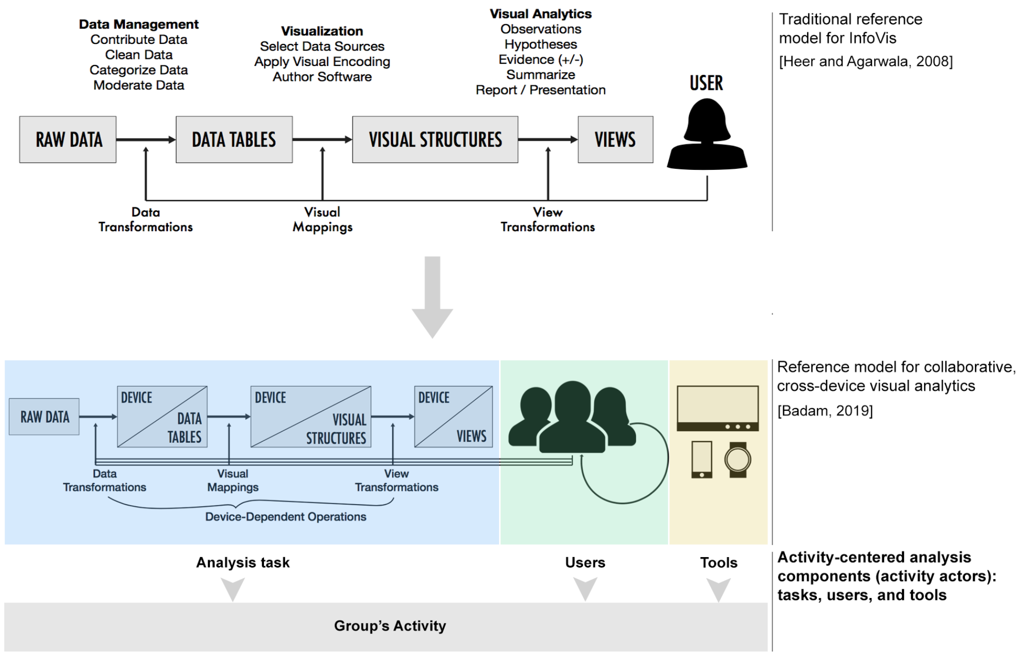


Figure 16: The traditional reference model for InfoVis (above) is redefined for multi-user multi-device environments (below). The activity-centered approach maps this definition into activity actors: tasks, users, and tools. For each actor, we apply appropriate empirical methods to present a structural analysis of group's activity in this complex environment.

The "activity system" triangle proposed by Engestrom (73) is one of the main model that was developed based on the concept of activity theory. It has been considered by several researchers as a theoretical framework for the analysis and evaluation of system-user interaction. This model describes actions through six elements: the objective of the activities, subject engaged in the activities, social context, tools or the artifacts, division of labor or roles, rules or guidelines regulating activities. Therefore, Kaptelinin and Nardi argued that the most leverage from this model are complex systems with multiple subjects and objects, as the focus here changes from single user-system interaction into collaborative uses of technology with mixed virtual and physical settings and set of activities. The activity system model has been applied to many HCI research to inform the design and analysis of technologically mediated activities. We framed an approach in terms of the components of this framework that facilitates the selection and application of appropriate analysis methods.

As defined by the activity system, the activity relies on network of actors to make it possible. Actors are the people, tools, rules, social context, etc, that interact to make the activity happens. Therefore, we need to define the network of actors. As shown in Figure 16 (above), the traditional visualization reference model illustrates the iterative process of the visual analysis task. Badam (74) redefined this model (Figure 16 , below) for multi-user multi-device environments where multiple users utilize multiple devices to perform the iterative process of visual data analysis. This new definition included two additional actors to the earlier definition: group of users and set of devices. So beside the process of the analysis task, users and devices compose the network of actors that make the activity takes place in this environment. These three working actors map onto the top components of the activity system triangle. The approach uses these dimensions to apply appropriate empirical methods within each one. Therefore, this structural analysis provides an overall understanding of group's activity in complex visual analytics environments.

4.3.1.1 Tools Usage

Each device has specific capabilities to serve users in the analysis course. We were interested in capturing the usage frequency of each device and how they contributed to the analysis flow.

Some devices like large displays naturally offer a public usage to multiple users at the same time while tablets and laptops tend to be privately used. However, participants were not restricted to use devices in specific ways. Rather, I left it to the users to decide how and when to use the tools. I aimed to capture the dynamic of using devices in parallel, in conjunction, etc. and the patterns of how users utilized them. As shown in Figure 17, I coded the use of devices from videos and system logs at 5-second intervals. I considered a device as under use if one or more participants are interacting with it. That includes direct interaction using touch, click, etc., or indirect interaction like looking at and discussing information. The view exploration task as discussed below encompasses a direct and indirect interaction with devices. In the next subsections, I discuss the types of activities and the styles of user formations around a device. A complete list of tools usage timelines for all trials is presented in Appendix A.

Utilization of devices affordances. Participants used the large display to share and arrange the many visualizations they created through the analysis session. This came naturally from the bigger area offered by the large display. Although users were able to create many visualizations privately on portable devices, they shared publicly what they thought was important to their analysis.

The large space was also used to lay out visualizations. The layout was important to indicate implicit relationships between visualizations. As I discuss in the next section, participants were taken different analysis paths and in some cases layout was used to differentiate paths. Tablets along with the large display offered different analysis styles as suggested by (15). This was important to allow participants to try different analyses on their own and then share with

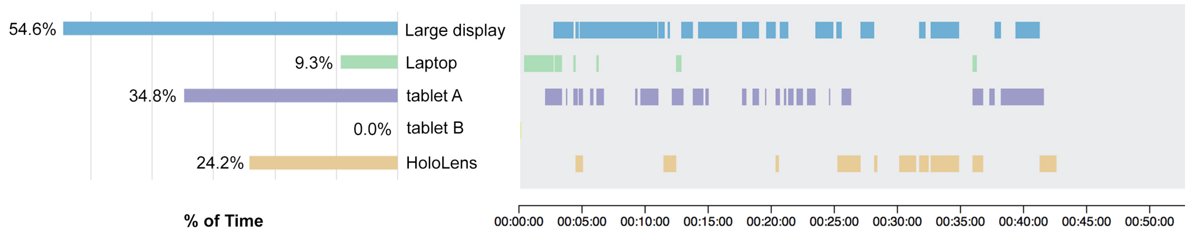


Figure 17: Logging of tools usage during the analysis session. The large display was the most used device.

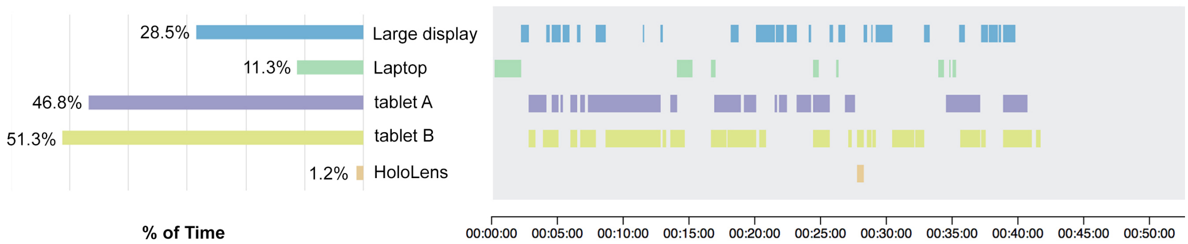


Figure 18: In session 2, participants loosely collaborated by engaging in individual analyses using tablets. See Figure 24.

others. Unlike stationary devices like laptops, tablets also offered some mobility to users by enabling them to move within the space while working on their analysis. In three out of six sessions, participants used the Hololens to compare the depth of wells and earthquakes in 3D. The 3D view offered a quick comparison of depth. In the other three sessions, participants used 2D charts to find a depth relationship, which required them to create many visualizations.

Frequency of use. In five out of six sessions, the large display was the most used device. This can be explained by the fact that users use the large display as a public canvas to plot and arrange all created visualizations and conduct their discussion and analysis around it. In addition, the large display offers a better space for collaboration than other devices, which align with the current assumptions in the literature (75). In an exceptional session, users used tablets most of their time analyzing visualizations individually, which led to loose collaboration for around 34% of their analysis time. This highlights the importance of supporting communication and work coordination in these systems to enhance collaboration. We need to develop guidelines to better design visualization tools that can utilize the affordances of multiple devices and overcome the communication and coordination challenges.

Joint and parallel use. I observed a frequent presence of coupling styles between devices. In the case of coupling, participants mostly used a large display along with one or two portable devices. While the large display served as a canvas to place visualizations, other devices were used to further analyze these visualizations. Participants were engaged in cycles of creating visualizations and analyzing them (i.e. filter, change representation, aggregate data). Although participants had two tablets, in most sessions they used one of them more than the other one. There were some cases where participants divided tasks and worked on both in parallel. Here, I stress on the importance of a guidance mechanism that can help users to better utilize devices of same capabilities to divide work.

4.3.1.2 Analysis Activities

In this section, I report the common high-level activities I observed among all groups. First, I coded all analytical processes performed by users; then, I classified them into four high-level

TABLE I: Percentage of time spent in using each device.

	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Avg.
Task time	43 min	42 min	66 min	28 min	22 min	48 min	42 min
Large display	55%	29%	41%	37%	34%	60%	44%
Laptop	9%	11%	19%	15%	15%	11%	13%
tablet A	35%	47%	12%	18%	14%	47%	29%
tablet B	0%	51%	4%	28%	15%	13%	17%
HoloLens	24%	1%	23%	11%	0%	13%	14%

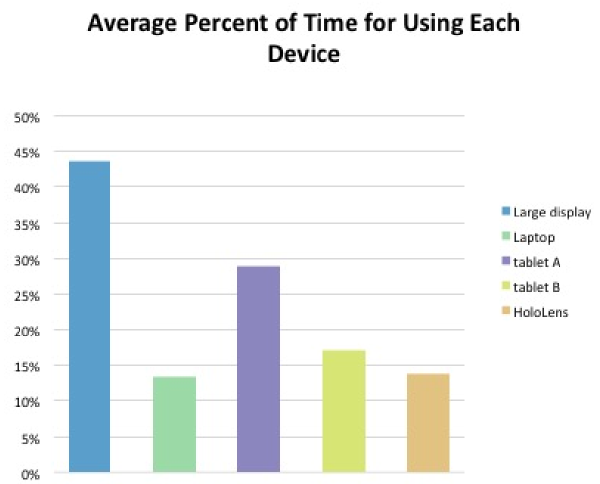


Figure 19: Average percent of time spent in using each device.

activities as shown in Table II. These high-level activities are: creating new views, exploring views, manipulating views, and analyzing views. Table II shows all the micro processes under each of these high-level categories. This categorization has similarity to some visual analysis processes described in (76) (54) which suggests that micro processes within the characterization can occur in settings outside the context of this study. Previous works (76) (54) characterized users' activities from the perspective of the visual analysis task. In the case of multi-device settings, the affordances of the devices influence the analytical task activities. For example, changing a visual representation into an immersive view is a cross-device analytical activity. Therefore, the characterization of groups' activities is based on our focus on the problem from the different dimensions of the environment. I coded these activities from video recordings and system logs. In this categorization, I focused on "actionable processes". Actionable processes are the type of processes when users directly or indirectly interact with analytical components (visualizations, filters, data subset, etc.) using one or more digital media (devices). In addition to actionable processes, I identified two other non-actionable processes: discuss observations and form hypothesis.

Creating new view. This activity comprises processes such as browsing data attributes, applying filters to datasets, and mapping attributes to visual channels. As the goal of these processes is to produce a new visualization, I identified them as an activity of creating a new view. The final product of these processes is a new visualization that is not derived from an existing one. I classified deriving visualization from an existing one as an activity of analyzing a view. Created new visualizations have no direct ancestors but they mostly have implicit relationships with other views along the historical paths of the analysis. I will discuss the explicit and implicit relationships of views in a later section.

TABLE II: The four activities I observed in the study. Micro processes are performed for the goal of the corresponding high-level activity

High-level activities	Micro processes	
Creating new view	Brows attributes Filter data Map data to visual channels	
Exploring views	Share visualization Reading visualization Direct interaction Indirect interaction	Cross-device analytical activities
Manipulating views	Layout views Resize views	
Analyzing views	Change view representation Calculate aggregation	

Exploring views. The exploring views activity captures all the processes where participants aim to derive information individually or collaboratively from visualizations. I noticed participants most of the time share private visualizations on the public wall display to explore information with others. In other cases, participants share their portable devices with others to explore visualizations. Participants explore by reading the information about the visualizations. This process is usually followed by a discussion or forming a hypothesis activity. Participants also interact directly or indirectly with visualizations to explore information. Direct interactions are the direct zooming, navigation, selection, etc., on the visualization while indirect interaction is through another device (i.e. tablet to wall, wall to HoloLens) or merely through the looking at and discussing information.

Manipulating views. I noticed a few processes where participants arrange views on the large display for different purposes. In some cases, participants position views in specific layouts for comparison. In other cases, they resize and position views to create clusters of views.

Analyzing views. Through their analysis, participants derived many views from existing ones. The goal is to render a further analysis of the current subset of data. Further analyzing the view comprises the changing of the visual representation. For example, changing a 2d map into a 3d representation, or into a scatter plot to correlate the distribution of specific attributes. Analyzing view activity also comprises the aggregation of data points using an aggregation functions.

Non-actionable activities. There were some activities where participants do not directly interact with the system. I identified those activities as common non-actionable. Under this definition, I considered when participants discuss observations they found and formulate hypotheses.

Non-linear Temporal Order . I noticed that these activities have no linear temporal order as participants switched between them frequently. To reveal unseen patterns and temporal relationships, I coded the time interval of each activity for all groups. These activities can temporally overlap when performed simultaneously by multiple users. Therefor, I charted each activity in a separate timeline as shown in Figure 20 and Figure 21.

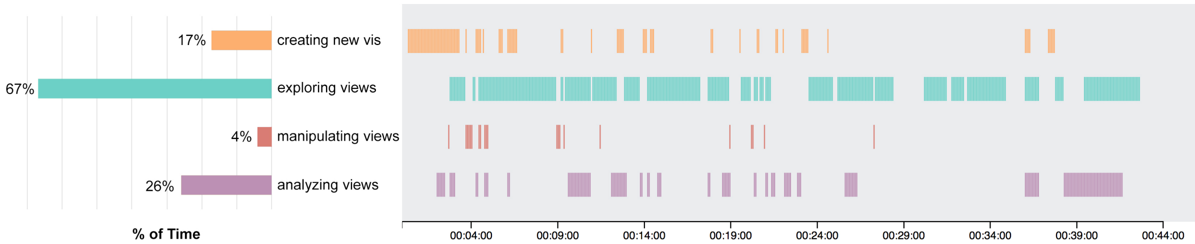


Figure 20: Logging of activities during the analysis session.

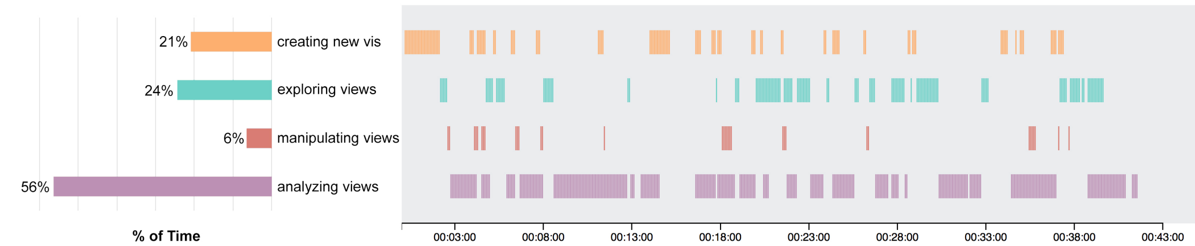


Figure 21: Logging of activities during the analysis session.

Figure 20 and Figure 21 show timeline visualizations of performed activities during two sessions. A complete list of timeline visualizations for all trials is presented in Appendix A.

The Analysis Outcome. The task was exploratory in nature to infer any correlation between the provided datasets. There were no correct or wrong answer. In all sessions, participants came to one or two observations. They wrote down those observations, mostly at the end, on the task paper that was given. I didn't observe that they wrote down any other information during the analysis either regarding what dimension space they are working on or other related

TABLE III: Percentage of time spent in performing each activity.

Session	1	2	3	4	5	6	Avg.
Task time	43 min	42 min	66 min	28 min	22 min	48 min	42 min
creating new vis	17%	21%	19%	15%	25%	8%	17%
exploring views	67%	24%	58%	34%	21%	66%	49%
manipulating views	4%	6%	4%	19%	9%	7%	7%
analyzing views	26%	56%	12%	23%	14%	43%	30%

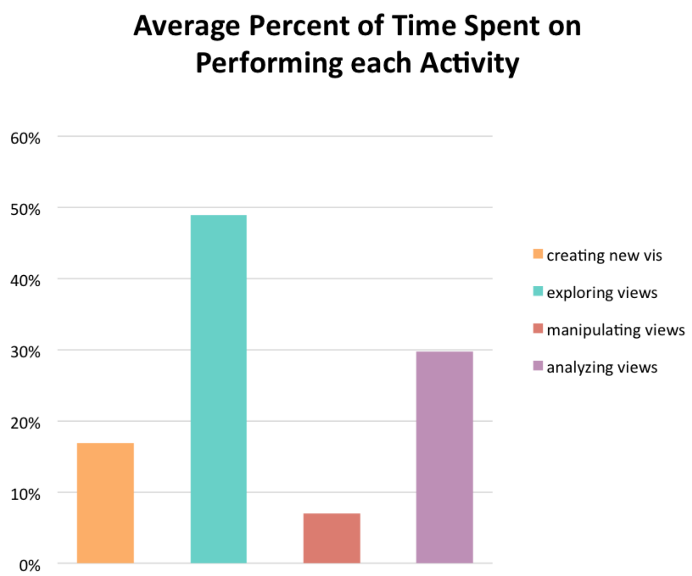


Figure 22: Average percent of time spent in performing each activity.

information. In most sessions, participants started with a dimension space of the data specified by time, location, depth, etc., and then they went broader or deeper in the exploration. At this time, I saved the analysis of the dimension space coverage to be investigated in the future work. Analyzing how they explored the dimension space helps in designing recommendations.

4.3.1.3 Collaboration Styles

Unlike the case of a single display, the presence of multiple devices arises the question of how closely or loosely participants have worked. This also requires us to redefine the close and loose coupling of work. Petra et. al. (31) discussed the close coupling around tabletop display. Several aspects were used to define the range of close to loose coupling. From the most to the least influencing factor, these were the engagement in the discussion, the working on the same information, and the working on the same view.

Having multiple devices to complete the analysis task, I was interested to observe how they would work together. So I envisioned the collaboration as how they would engage in discussing the task, working on the task, as a group or individually, on one device, or using multiple devices. Similar to Petra et. al. (31), I used a few aspects to rate the collaboration as close or loose collaboration. These are the engagement in discussing the task, working on same information or same view, either using one or multiple devices. So in this environment, I consider the collaboration is close in cases where all participants are engaged in a discussion about the task and looking into one or multiple views using one or multiple devices. As shown in Figure 23, there are a few cases were observed in the study in which they can be considered as close collaboration. In all four cases, all participants were engaged in the task. In cases when one participant detaches himself from the group to work individually or to not engage in the work, these are considered to be medium collaboration as the two other participants will still be working as a group in different ways.


Close collaboration
Moderate
all are working on same view, same device
all are working on multi views, same device
all are working on multi views, two devices
one is working on one device, two are engaged in the discussion
two are working on one device, one is working individually
two are working on two devices, one is working individually
one is working on device, one is engaged with him, one is working individually
two are working on one device, one is not engaging
one is working on devices, one is engaged with him, one is not engaging

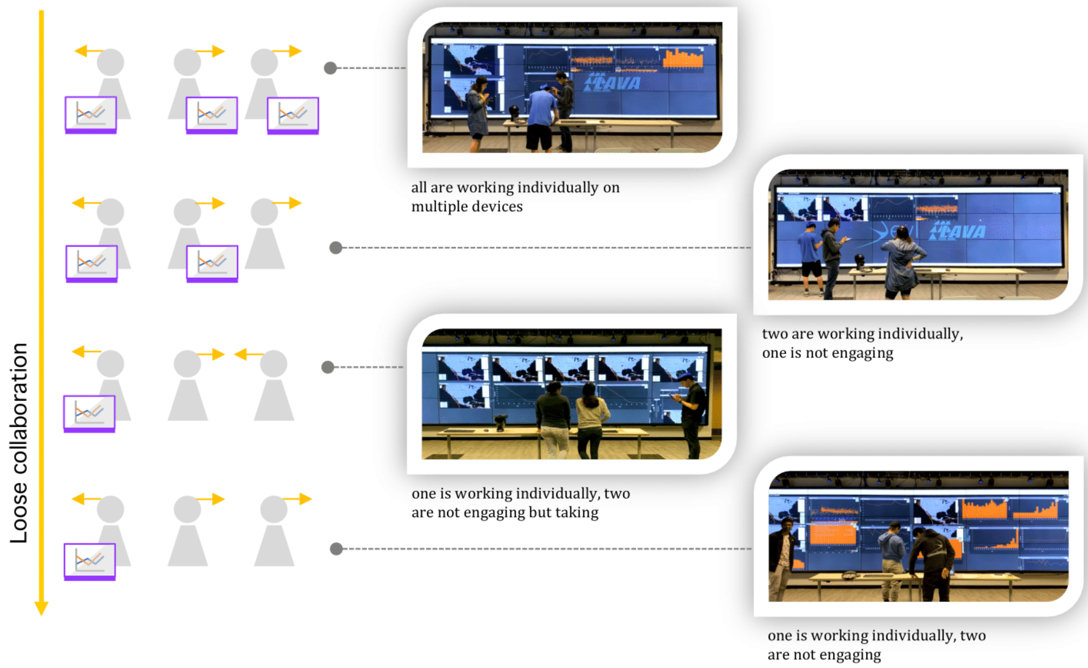


Figure 23: Several cases of users' formation rated as close, medium, or loose collaboration.

Lastly, when the members of the group start working individually or not engaging in the work, I considered these cases to be loose collaboration. Although these cases of users' formation will have a collaboration rate as a close, medium, or loose, as shown in Figure 23, they take position on the collaboration scale. For example, the first case of *working on the same view on the same device* has higher closeness than the case of *working on multi views on two devices.*

The set of users' formation cases is what I observed from this study. Therefore, when having larger groups, new cases of users' formations can emerge. However, the concept of using a scale of collaboration degrees can be used to define and rate different cases of users' formations in other settings. In addition, this definition of collaboration is related to the context of multi-user multi-device systems. Although I assumed co-located collaboration, it can be considered in other settings such as distributed collaboration, assuming users can use the same device or number of devices with the proper technology.

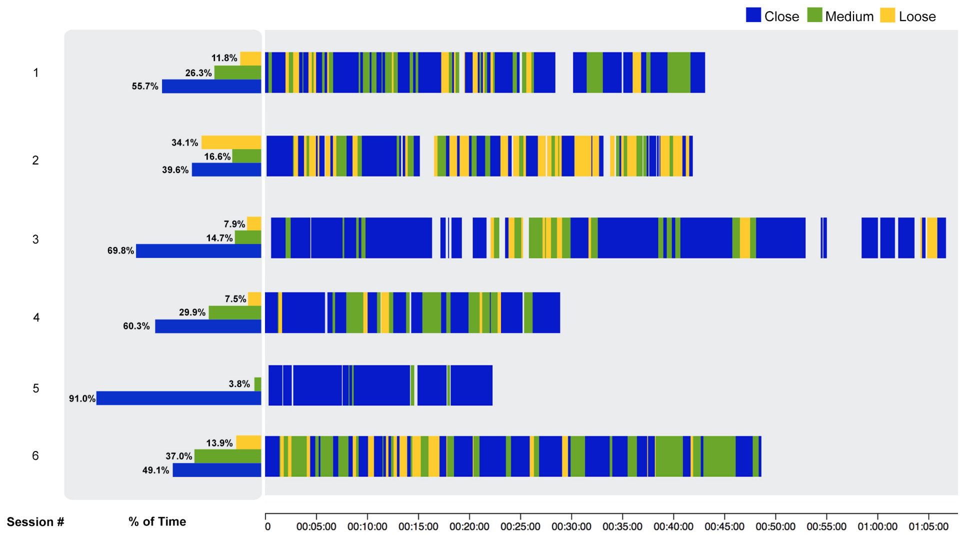


Figure 24: Users' collaboration logged as Close, Medium, and Loose collaboration.

During the qualitative analysis, I coded several cases of users' formation around devices. Then, for each case, I rated their collaboration as close, medium, or loose collaboration depending on the observed engagement in the discussion, task, and the device.

TABLE IV: Percentage of time spent in each collaboration mode.

Session	1	2	3	4	5	6	Avg.
Task time	43 min	42 min	66 min	28 min	22 min	48 min	42 min
close collaboration	56%	40%	70%	60%	91%	49%	59%
medium collaboration	26%	17%	15%	30%	4%	37%	22%
loose collaboration	12%	34%	8%	8%	0%	14%	14%

As shown in Figure 24, participants closely collaborated for more than 50% of their time, except in one session where participants spent almost only 40% of their time in close collaboration. In that case, participants were less engaged in the group work and they spent 34% of their time on individual analyses.

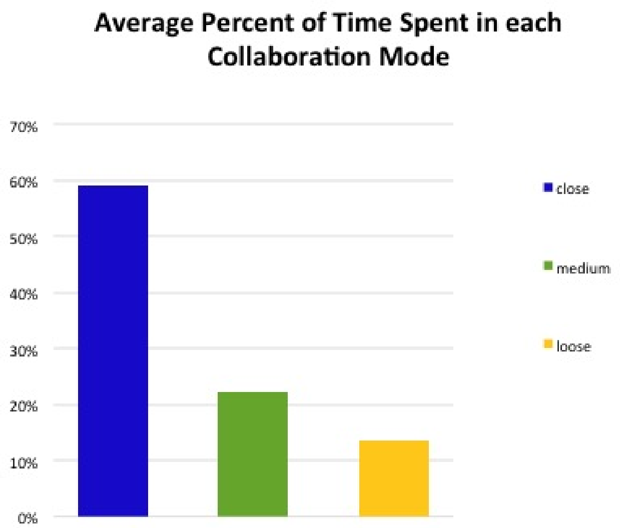


Figure 25: Average percent of time spent in each collaboration mode.

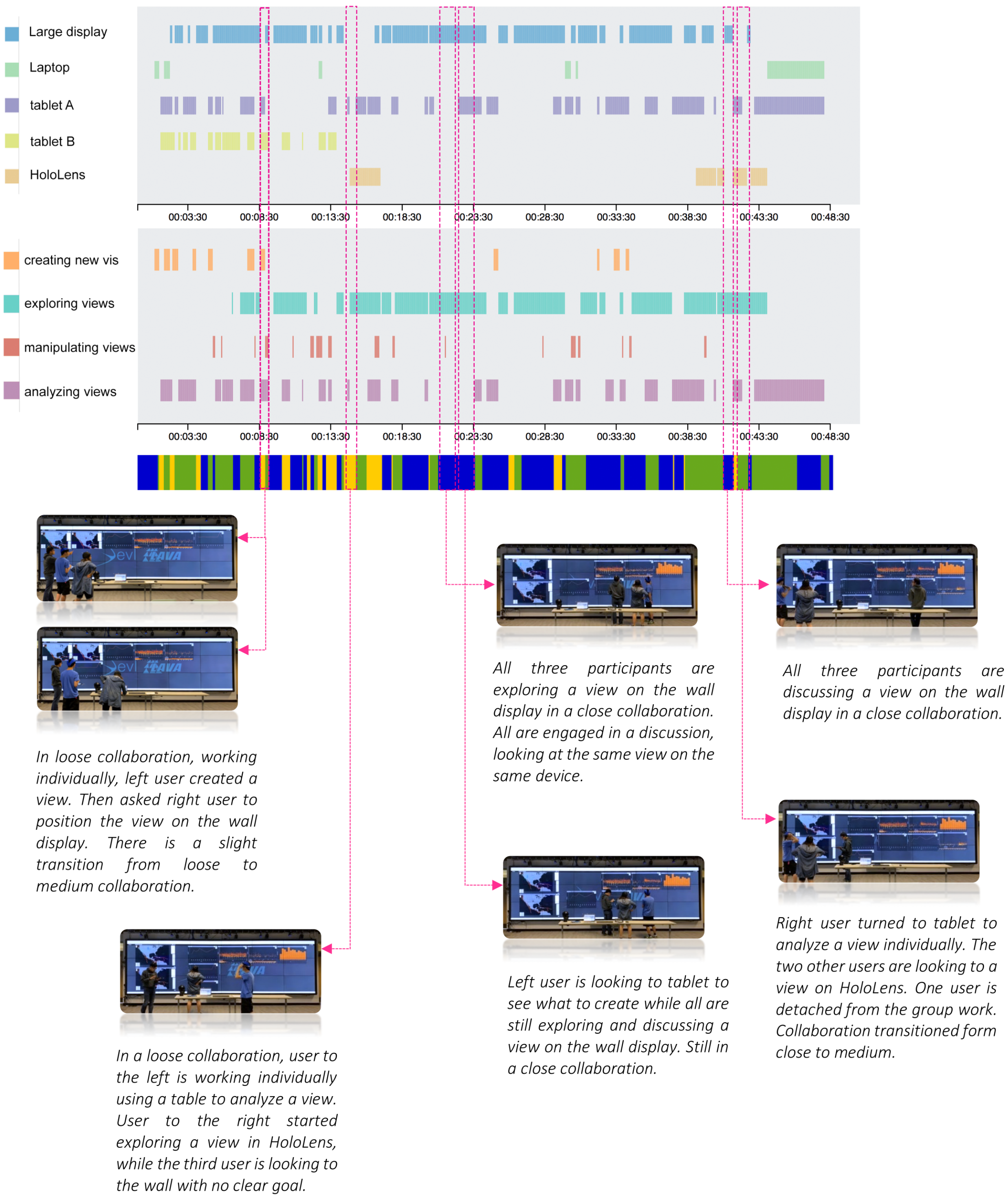


Figure 26: Snippets of activities and collaboration transitioning.

4.3.2 Strategies of Collaborative Visual Data Analysis

I started the second part of the qualitative analysis by documenting all created visualizations with their relationship to other visualizations. Then, I plotted a flow diagram of the created visualizations during each session in chronological order with arrows indicating the first set of codes. In a second pass of coding with a secondary coder, we developed a theme of the analysis structure by grouping and refining the set of codes. In this section, I report the observed exploration patterns and how they compose a structure of the analysis flow as: (1) a high-level analysis paths initiation and (2) a low-level visualizations exploration that occur along paths.

4.3.2.1 Two-level Structure of Analysis Sessions

I observed that the course of the analysis happens at two levels. Within each level, I observed a set of exploration patterns. At the higher level, participants were taken a set of exploration paths. I observed three path emerging patterns as below. Along each analysis path, I also observed a set of view-to-view generation patterns. The latter are the low-level exploration patterns that occur within the larger cycles of the analysis. I discus below these observations and shed a light on how this structure corresponds to the current definition of exploratory visual data analysis.

A) Analysis Cycles: Initiation of Analysis Paths

I identified the higher level of the analysis structure by analysis paths that were taken by participants. I coded times when participants individually or collaboratively specify a new subset of data points and start analyzing this subset through a set of visualizations to find patterns. All created visualizations from this subset compose analysis states along this path. I observed that participants start with one analysis path and subsequently start another analysis path to work along both paths in parallel or sequentially. Except in one case when one group initiated four analysis paths in parallel. I observed three patterns of path initiation in the study as follow.

Parallel analysis paths: This type of analysis path was the most common among all groups since the open-ended task was exploring two datasets and make observations of possible correlation. Participants fetch subsets of data points from different datasets using same attributes value (*i.e. year = 2009, or location = "California"*). Then, they analyze these subsets of data using different visualizations and measures. The analysis of visualizations along one path was highly affected by the flow of the other path. As shown in the example in Figure 30 (a), participant P1 requested to see the count of earthquakes over time after participant P2 merged the average volume over time of wells injection. Participants in general did organize visualizations from the same analysis path in a specific layout. However, there were no specific patterns on how groups work on parallel paths. They alternated frequently between working on both at the same time and focusing on one for some time and then switching to the other one. I hypothesize that the layout serves as a base for visualizations clustering but participants build a mental model of visualizations interconnections, which can be hard to track with many visualizations. Therefore, solutions such as meta-visualization approaches need to be investigated further.

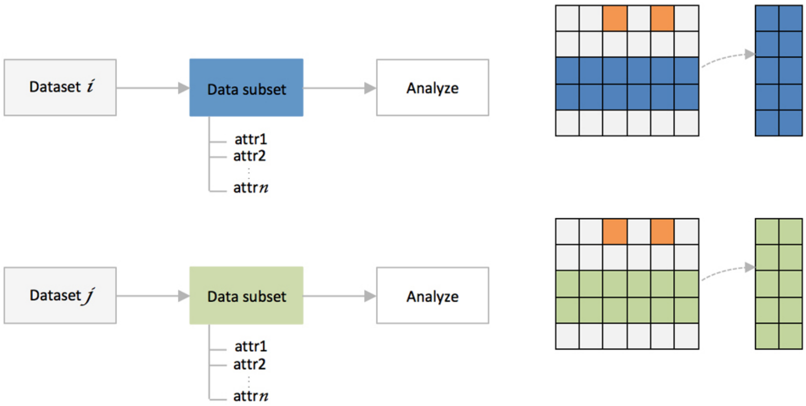


Figure 27: Participants frequently subset data points by same attributes from different datasets to work on both sets in parallel.

Sequential analysis paths: This type of analysis path takes place later in the analysis session. The Sequential term here refers to the temporal sequence and sequential course of analysis, and not the sequential spatiality. Participants initiated this type of analysis path to re-examine formed hypotheses or observations by looking into different dimensions of the datasets. As shown in Figure 30 (b), participant P1 and P2 found no spatial correlation between earthquakes events and wells locations on the year 2009. P2 suggested: " *We can try another year...*" in order to test their hypothesis. With the dynamic nature of the analysis, sequential paths soon become parallel path of another one. Although taking this direction in the analysis session is less frequent, I believe that it is very important for validating hypotheses especially when working on large datasets. This can be envisioned by an automatic regeneration of analysis paths on different dimensions of the dataset for rapid validation of hypotheses, with leaving the control over the analysis to the user.

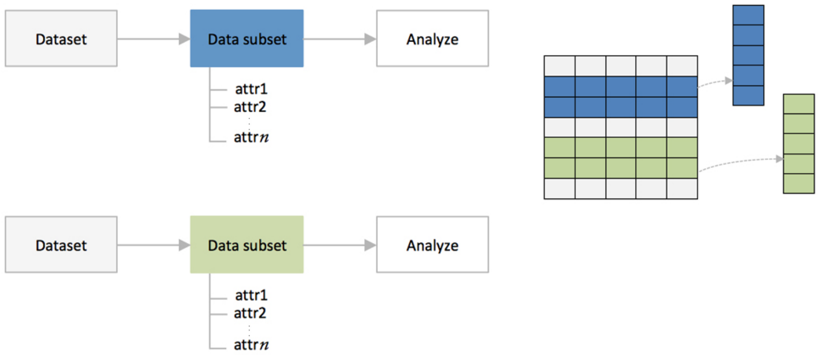


Figure 28: Later in the analysis session, participants initiate a subsequent analysis on a different dimension of the data to validate hypotheses.

Focused analysis paths: Taking this direction in the analysis happens when participants start with a larger scope of data subset and then drill down into the dimension space. This is understandable when hypotheses cannot be generalized over a larger dimension space. For example, as shown in Figure 30 (c), participants started an analysis on a larger dimension space (location = "California" AND "Oklahoma") and after forming some hypotheses; they drilled down to verify that the hypothesis is still correct per state. Focus analysis can be considered as special validation task similar to subsequent analysis. I observed that participants were also focused on their analysis along the last two types of paths as they focused only on attributes of interest from which they formed their hypotheses. I believe that this is because of the gained knowledge from the exploratory analysis on previous dimensions space.

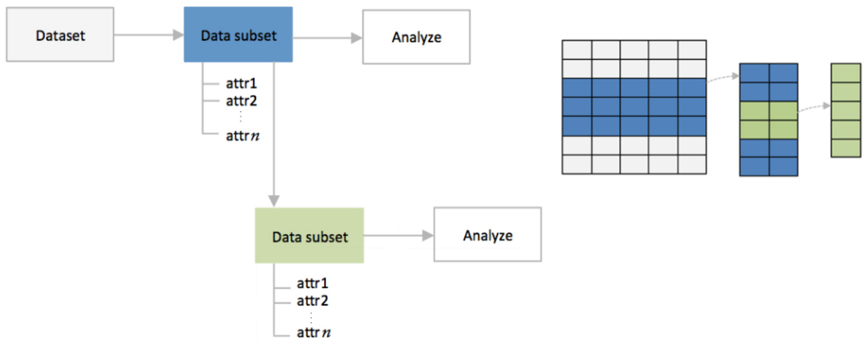


Figure 29: Participants later in the analysis session initiate a focused analysis path by drilling down into a dimension space.

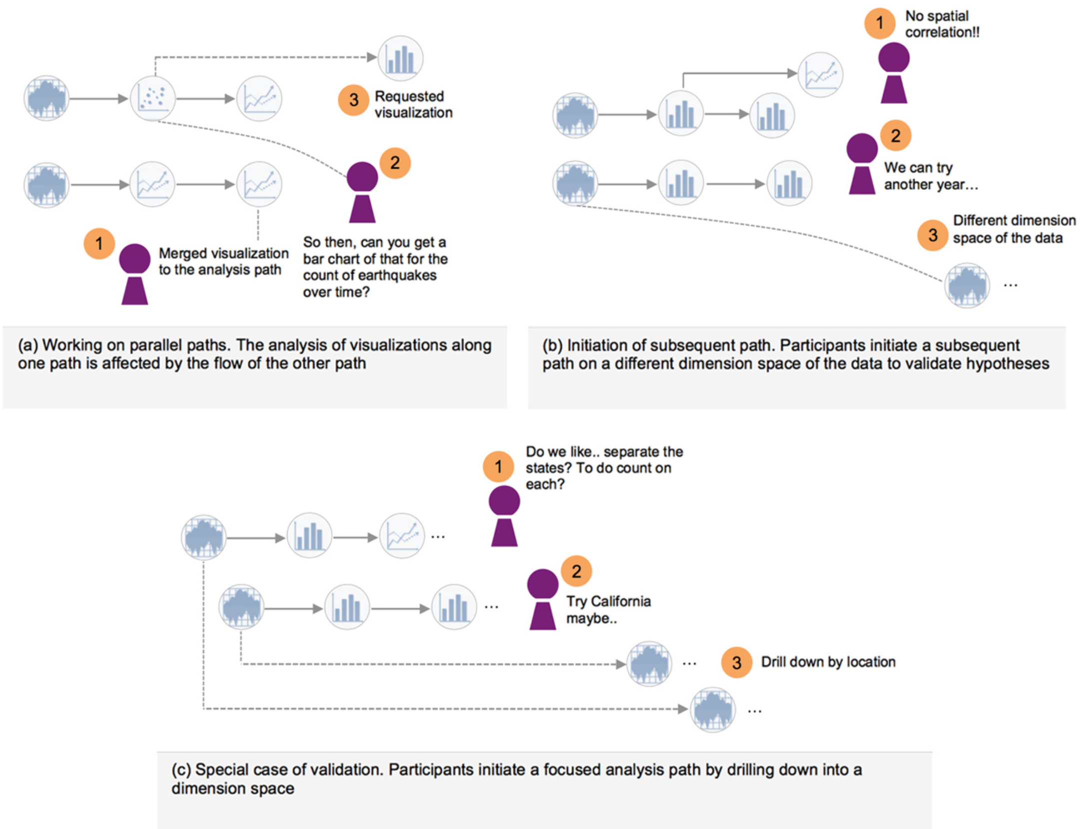


Figure 30: Examples of analysis paths initiation.

B) View-to-view Generation along Analysis Paths

Participants generated and explored sets of visualizations that compose the evolution of analysis paths. They specify visual encodings, group, or aggregate data for all produced visualization resulting as a new analysis state. The production of visualization can have an implicit or explicit relationship to another visualization in the analysis session. In other words, participants produced some visualization by explicitly referring to another visualization either to directly compare, correlate, etc. Some other visualizations are produced to be explored solely without explicit relationship to another visualization (implicit to the context).

I categorized those exploration patterns based on the goal of exploring every visualization. Those categories I present below are not a one-to-one relationship. That is, they can work in hierarchy as shown in Figure 32 where items in a low-level exploration compose a counterpart for a higher exploration task. For simplicity, I decomposed those exploration patterns and explained below.

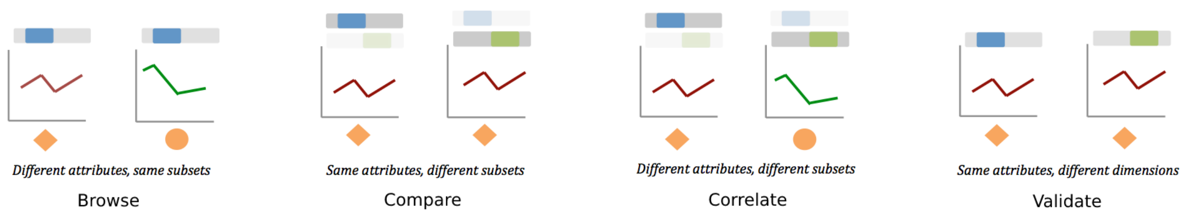


Figure 31: Views-to-views exploration pattern.

Browse: I observed that browsing attributes usually takes place mostly at the beginning of the analysis session. At this point participants are aiming to navigate through the data for interesting attributes that may involve interesting trends. After browsing, participants proceed to form different analysis tasks. When participants browse set of attributes, there is no obvious relationship between them, but they form the take away points for further analysis.

Compare: It takes place when participants create visualizations that share an attribute and visual representation. It's a common exploration task to find similarities and differences between data attribute. Compare can take place between attributes of data subsets from same or different population. For example, comparing the average depth of earthquakes in different years output an understanding of attribute trends over subsets from same data population. On the other hand, comparing the average depth of earthquakes and wells in a specific location outputs similarities and differences between an attribute trends over subsets from different data populations.

Correlate: As compare, correlate was also a common exploration task to infer the dependency of one attribute on another. Participants created many visualizations, mostly, of the same representation but different attributes to infer possible correlation. I observed that correlating different attributes form different subsets was the most common. In very few cases, it involved the same attributes (i.e. depth) from different data subsets. As mentioned earlier, the presented exploration patterns are not one-to-one. In many cases, participants correlated more than one visualization to infer attributes relationship.

Validate: It takes place within the larger cycle of the validation process. It involves the creation of visualization of similar attributes and representation but different dimension space of data points. Participants aim from this task to validate observations made on an earlier visualization from a different dimension space or data subset. It was less frequent than other patterns as the validation process was limited due to the hard tracking of the analysis flow.

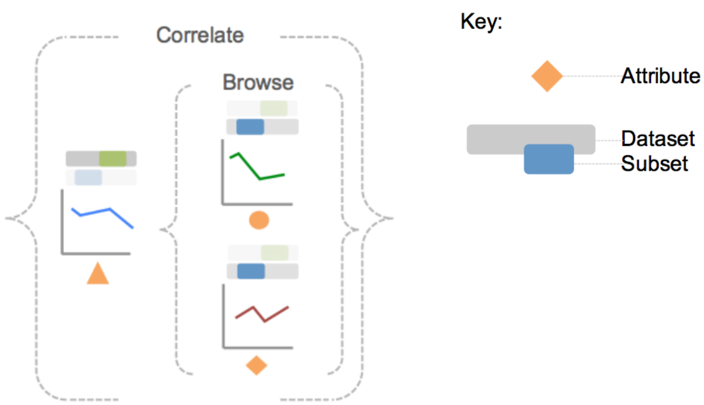


Figure 32: Exploration patterns can occur at different levels. Components of Browse task are used to find if they correlate to another attribute.

4.3.2.2 Temporal Relationships of Observed Patterns

At the higher level of analysis paths initiations, some paths are initiated naturally after another. For example, a focused analysis path is a drill-down of an earlier analysis path, and it was started later in the analysis session. Sequential paths also naturally come after an earlier analysis path. However, I observed that one team initiated a sequential path, for validation, along the current analysis path. This is because the team planned its analysis strategy at the beginning of the analysis session. Therefore, I believe it is important to provide flexibility in designing visualization tools to support different analysis styles and strategies. Working on both focused and sequential path is independent of their ancestor paths. So by default, participants can work on them independently later in the session or along with their ancestors. On the other hand, the goal of initiating a parallel path is to be working on them in parallel with another analysis path. Therefor, they come sometime after the initiation of an analysis path. When participants were working on more than one analysis path, there were no specific patterns on how they alternated between them. At the lower level of exploring different visualizations, the temporal relationships of creating and working on visualization were more dynamic. Participants switched frequently between creating visualizations for browsing, comparing, correlating and validating. They worked forward and backward along the analysis paths.

4.4 Discussion

4.4.1 The Current Assumptions of Analysis Structure

In this section, I discuss efforts made by other researchers to define and understand the structure of exploratory visual data analysis. It is crucial to understand how the process of exploratory visual data analysis is conceptualized to better support the design of visualization tools. In this work, I presented a categorization of analysis behavior graphs exhibited in the study. Therefore, it is important to shed light on how analysis behaviors are conceptualized in the literature.

A recent review by Battle and Heer (63) indicated that the current literature of exploratory visual data analysis provides different assumptions about how it is defined and performed, and there is no consistent definition about how it is structured. It is unclear yet whether exploratory visual data analysis follows certain structural patterns or, due to behavioral differences, the pattern is hard to predict. They highlighted different assumptions about the structure of the analysis from the literature:

- "Overview first, zoom and filter, details on demand"(77):

  The first theme of work follows Shneiderman's mantra: "Overview first, zoom and filter, details on demand" for exploratory visual data analysis. It is assumed that the visual data analysis is an iterative process that alternates between these phases.

- View creation, exploration, and refinement:

  Another theme of work assumes three phases of exploratory visual data analysis that the user alternates between. It is assumed that user alternates between view creation, exploration, and refinement.

- Discovering insights and recording insights:

  A different theme of work assumes that the analysis process alternates between discovering insights and recording insights.

- Open-ended exploration and targeted question answering:

  Another theme of work assumes that exploratory visual analysis is an alternation between focused and open-ended exploration.

The common theme among these assumptions, that the analysis process is an iterative process that alternate between different phases. As shown above and as stated by Battle and Heer (63), “the literature is inconsistent in defining exactly what the different phases of EVA are.”

The presented characterization aligns with the existing assumptions that the analysis process is an iterative process that alternate between different phases. I identified that users involve in cycles of views creation, in different identified patterns, which align with the exploration cycles in the themes above. However, I identified that these views creation cycles take place in large cycles of exploration which are the analysis paths initiation. Therefore, the presented categorization involves two levels, analysis path initiations and views creation along each path.

So this characterization has similarity to existing themes in term of alternating between different phases. It assumes the analysis process is an alternation between analysis paths in the larger cycle of the exploration.

However, this characterization contributes more fine-grained patterns within the different phases. At the higher level of exploration which is the analysis path initiations, I presented patterns of how users alternate between different paths. Along each analysis path, I presented patterns of how users alternate between views creation. The categorization presented here contributes a new definition of exploratory visual data analysis as an effort to motivate the community to better support visual exploration tasks.

4.4.2 Overall Interpretation of the Results

Due to the complexity of performing an integrated analysis of user’s experience in this environment, the presented approach provides a structural analysis on different dimensions:

tools, tasks, and users. Above, we presented a detailed analysis on each dimension. Here, we provide an overall reading of the results. The integration of the analyses we performed on the three dimensions can help us interpret the overall experience of the users.

From the results of the tools dimension, we found that the large display was the most used device in 5 out of 6 sessions for an average of 45% of the time. We found from the results of the task dimension that the exploration task was the most performed task in 4 out of 6 sessions and mostly took place on the large display. By integrating results from the users dimension, we found that performing the exploration task using the large display is accomplished in close collaboration for the most of the time. This yields that the large display encourages the close collaboration for exploration and discussion of visualizations, which inline with earlier findings that large displays facilitate users collaboration (8) and analysis of data (78).

Finding I: In this dynamic environment, users performed cycles of activities during the data analysis, in which they transitioned between tasks and devices in collaboration styles that vary from close to loose collaboration. The most common activity that took place in the study was the ***close collaborative exploration task*** of visualizations on the ***large display***.

The earlier studies of data analysis using large displays either target a single display or a single user (8)(78). The finding is in the context of MDEs where transitioning between devices and tasks is more dynamic. The finding above is important for designing visualization tools that target multi-device environments. This finding highlights the importance of promoting collaborative jobs such as collaborative interaction, workspace organization, and user engagement during the exploration task on the large display, for more effective collaboration.

The other tasks that took place using the other small devices were less frequent in the analysis cycle except in one session. Users were transitioning to other devices like tablets, laptop, and HoloLens to further analyze existing visualizations (e.g. changing filters, visual representation, and aggregations) or to create new ones. It did happen that they performed exploration task on these small devices, but it didn't take long until they switched to explore the visualizations on the large display. Working on the tasks using the smaller devices was generally in medium collaboration (two are engaged in the work on the device) to loose collaboration (one user is working individually). This inline with the findings of Wallace et al. (34) where they found a negative correlation between the use of personal devices and equity of member participation.

Finding II: ***collaboration varies based on the used devices and the type of the tasks***. The collaboration was less during the task of analyzing existing views or creating new views on the small devices than the task of exploring views on the large display.

This finding highlights the importance of supporting awareness and work coordination during specific periods of the analysis cycle. When working on small devices, users rely on verbal communication which is not effective to fully communicate their work and results. Therefore, an awareness techniques of the analysis work that takes place on small devices will improve users' awareness and help in work coordination. The above findings help us understand the design space of visualization tools for MDEs.

CHAPTER 5

VISUALIZATION OF DIMENSIONS SEARCH SPACE IN EXPLORATORY VISUAL DATA ANALYSIS

5.1 Introduction

One of the challenges that motivated this work, is the biases to the depth first analysis in this environment. When analysts explore a dataset that they haven't seen before, they typically start with a broad overview to familiarize themselves with the data. Then, they might focus on answering specific question about field A and B which can spark another questions about them (anchoring bias) leading to a focused or depth-oriented exploration. In addition, factors such as recency effect and short term memory (availability bias) support depth-oriented exploration.

As described by Law and Basole (79), depth-oriented exploration can cause a biased exploration path due to constrained coverage of the dataset, and hence confirming hypotheses inaccurately. They illustrated how insights can be missed with depth-oriented exploration. They argued that breadth-oriented exploration can alleviate biases caused by recency effect and anchoring and called for systems that support breadth-oriented analysis.

Therefore, the main goal of the presented approach is to support the breadth of the analysis in term of covered dimensions space. Visualizing the coverage of the dimensions space provide information of what has not been explored so analysts can orient and steer their analysis towards unexplored data.

In Chapter 4, I investigated exploratory visual data analysis in collaborative multi-device settings. I presented an evaluation of exploration strategies from observed analysis behaviors. Three exploration patterns were identified which involved frequent exploration of different subsets in the attributes data space. Since the exploration was performed by the team members in a scattered fashion, it was hard to keep track of the analysis flow because of the limited short-term memory. The dependency on short-term memory led to focusing on the most recent exploration (recency effect) which in turn led to a more depth-oriented analysis.

Supporting the breadth of the analysis is still a challenging problem. Battle and Heer (63) found that exploratory visual data analysis is primarily depth-oriented using Tableau and called for a greater exploration breadth. Voyager (61) promotes a breadth oriented exploration by providing users with useful visualization recommendations to increase the coverage of unique data attributes. They evaluated how visualization recommendations increase the breadth of the analysis by evaluating the number of unique attributes combination explored by users.

The focus of this chapter is to design a supportive visualization that help users navigate the dimensions' data space and increase the breadth of the analysis. Analysts especially in collaborative setting beyond the single desktop need to understand what courses of analysis were done by the team and what were left. By visualizing the dimensions search space, the visualization communicates to the team what have been investigated in the attributes data space and what were left.

In this chapter, I present the design and implementation of visualizing the dimensions search space that will serve as a supportive tool for users during exploratory visual data analysis.

Based on previous observations, I hypothesize that visualizing the dimensions search space will positively affect the analysis process. In the next chapter, I present results from an evaluative study of providing the users with visualization of dimensions search space during exploratory visual data analysis.

Throughout this chapter, the reader will come across few terms that I define here for clarity:

- **Dimensions' data space coverage** (or shortly data space coverage): is making a coverage selection based on multiple dimensions.
- **Dimensions search space**: is the process of selecting different data space coverage. It is what the visualization tool mainly visualize.
- **Dimension's data space coverage(s)**: is the selected coverage on this dimension. It can have multiple of them.

5.2 Guidance for Visual Data Analysis

Prior research has already investigated several methods to assist and guide users during visual data analysis. These methods provide guidance in diverse ways and for different purposes. To facilitate a comprehensive reasoning about designing guidance in the context of visual data analysis, Ceneda et al. (80) proposed a general conceptual model of guidance describe it's different components and characteristics. They used existing guidance approaches to illustrate it's components and extended van Wijk's model of visualization with these fundamental guidance components. According to this model, a guidance in the context of visual data analysis "can be characterized in terms of three main aspects: knowledge gap, input and output, and guidance degree" (80):

Knowledge Gap. Before designing any guidance, it is important to identify what a particular problem is facing the user. The knowledge gap pertains to the problems that the user needs help with to make an efficient progress in the analysis. According to this model, there are two types of knowledge gap. The first problem is that the target is unknown, which means that the user doesn't know the result of a specific analysis activity. The second type of knowledge gap is the unknown path. The unknown path means that the user knows the result to be achieved but doesn't know how to achieve it.

In this research, the supportive visualization assists users in their exploration. As defined by Munzner (81), '*exploration*' is a searching task where the target and it's location are unknown. Although the goal is not to assist users finding the target, the visualization helps users broad their exploration to find possible targets. The other aspect of the knowledge gap pertains to where the guidance is to be provided. There are many places where we can provide a guidance to the user. The guidance in the *Data* provides assistance in terms of data subsets or features. The guidance can be also provided for *Task* structuring and for *VA methods*. It can be also provided in terms of who should do a specific task (*Users*), and what *Infrastructure* to use.

Input and Output. To generate a guidance, there should be a process that takes some kind of inputs to produce a guidance to the user. Based on which guidance to be generated, the input sources can be the data, the visualization image, the history in terms of interaction or provenance, and domain or user knowledge. The guidance generation process produces the guidance as direct or indirect answer and uses an appropriate means to present the answer to the user. In the context of visual data analysis, visuals are the typical means to convey the guidance to the user. The designed guidance in this research takes the history of exploration and provides a visual representation of it.

Guidance Degree. Once the guidance is provided, there are different scenarios on how the user proceeds with a guidance. The guidance that is provided through visual cues, as used in this research, can orient the user's analysis behavior which in turn can build or maintain the user's mental model. The second degree of guidance provides the user with different options that differ in terms of quality or costs that lead to specific results. The user has a freedom to follow or not to follow these options. Prescribing guidance approaches provide a higher degree of guidance through automated process that operates on their own to generate desired results.

5.3 Visualization of Dimensions Search Space

In this chapter, I provide the design and implementation of a visualization tool as a guidance metaphor during exploratory visual data analysis. It visualizes the dimensions search space in terms of what dimensions (and their data space coverage) have been investigated and in what combination.

5.3.1 Related Work

There are few visualization approaches in the literature that were used to visualize the dimensions search space. These approaches include Scented widgets, Circos plot, and Treemap.

Scented widgets. They were used by Willet et al. (82) to enhance visualization controls with embedded visual cues to enable users understand the navigation of the data space. These visual cues guide and orient users in their analysis. Willet et al. (82) enhanced HomeFinder, a home searching tool, with scented controls to help users understand the prior searches made by other users and which data dimensions were investigated more than the others. They proved in a controlled experiment that users made more discoveries using scented widgets than baseline tool, and it helped them understand the navigation of the information space.

Similarly, Sarvghad et al. (37) used the notion of scented widgets and embedded information revealing each dimension's frequency of investigation and co-mapping with other attributes. The list of data dimensions were attached with bar charts to their left revealing the number of investigation of each dimension. Selecting a dimension (in orange) highlights in blue the attributes that this dimension was investigated with (appeared together in analysis chart). This information is embedded within the interface of the visualization tool and evaluated on how they can positively assist data analysis.

Circos plot. Sarvghad and Tory (60) used Circos plot to visualize the co-mapping of tabular data attributes in exploratory visual data analysis. Co-mapping (in other word co-investigation) of a pair or a set of attributes means that they appeared together in a chart for analysis. The Circos plots each attribute as a curved segment labeled with the attribute's name and the number of charts it appeared within. The longer the segment, the larger the number of charts this attribute was included in. Selecting an attribute segment (e.g. Profit) shows co-mapping arrows to the set of attributes this segment appeared with in a chart. The width of the arrow indicates how frequent the pair of the attributes appeared together. They used the Circos to study how it positively helps analysts understand past analysis done by other.

Treemap. Due to chart clutter that Circos can encounter with a higher number of dimensions and higher-order of co-mapping relationships, they changed the design to Treemap for visualizing the dimensions as a set of rectangles (60). Each dimension is represented by a cell labeled with the dimension's name. The white cells represent the un-investigated attributes while the grey ones are the investigated attributes. The size of the cell represents the number of investigation of that attribute. Larger rectangles indicate frequent investigation of these attributes in prior analysis. User can reveal information of co-mapping through interaction by clicking on the cells. The selected cell is highlighted in orange and all other dimensions that appeared in chart with the selected cell are highlighted in blue.

These approaches have been evaluated with tabular data. Willet et al. (82) evaluated their approach for information foraging tasks while other approaches (37) (60) have been evaluated for open-ended exploratory tasks. The domain of the presented approach is open-ended exploratory tasks of tabular data.

5.3.2 Design Goals

The above described approaches focus on visualizing the frequency that dimensions were investigated. While these are important features, they lack revealing information about what dimensions' data space coverage were investigated when analysts pivot the analysis between different subsets of the data space. There are design goals that I want the selected visual representation to satisfy. These design goals were identified based on observations from the earlier study. In addition to frequency of investigation and co-mapping, I hypothesize based on observations that revealing information of the dimensions' data space coverage would improve and increase the breadth of the analysis.

I observed in the earlier study that participants evolved in their analysis by investigating selected subsets from the data space, and the analyses were mostly depth oriented in these data subsets. In the last chapter, I identified three patterns of moving from one subsets to another on one or more dimensions. These patterns are the analysis path initiation where participants specify new subsets of data points for analysis. These analysis paths were parallel, sequential, and focused. Each of which represent a pattern of selecting a subset from the data space for analysis. The reader can refer to chapter 4 for more details about these exploration patterns.

I focus on revealing information about what dimension's data space coverage(s) were investigated and what were left to encourage a broader analysis. I selected the parallel set as the visual representation for visualizing the dimensions search space. This selection was based on a few design goals I wanted to achieve in a visual representation. These design goals are the ability to:

- Visualize the current dimensions' data space coverage that the analysis is pivoted to (**G1**)
- Visualize what parts of each dimension has been investigated so far and what were left (**G2**)
- Reveal information about each dimension's frequency of investigation and the co-mapping information of it's data space coverage(s) (**G3**)

In the next section, I discuss the visual representation design and how it can meet the above mentioned goals.

5.3.3 Paper-based Visual Representation Prototyping

The above mentioned approaches do not directly support G1 and G2 although they can be modified to satisfy these goals. For example, Treemap can nest each rectangle to show different values of the dimension and what ranges were investigated and what were left. The same can be achieved in Circos by dividing curved segments into sub-segments representing the different values of the dimension. However, that can cause visual clutter especially with the changing sizes of the rectangles and the segments in the Treemap and the Circos respectively. Therefor, after reviewing other visual approaches for categorical data, Parallel Sets were the simplest and the more intuitive approach and were selected to be used.

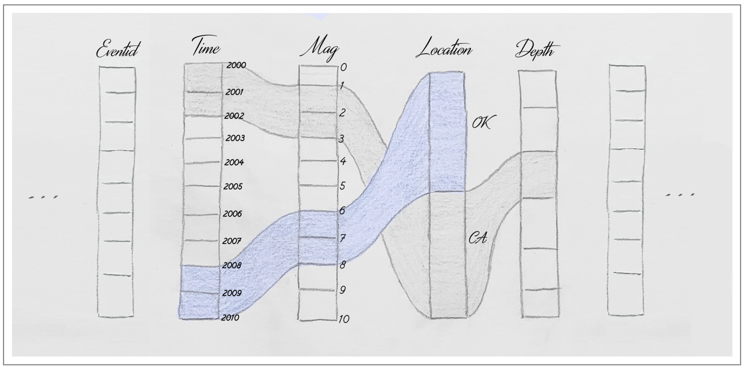


Figure 33: The basic idea of using parallel set to show dimensions search space. The analysis is pivoted to specific data space coverage on the dimensions time, magnitude, and location. All past data space coverage will be combined into one category (grey) and send to the back.

Parallel Sets is a visualization approach to visualize categorical data which can be though of as the parallel coordinate with the addition of the "proportional" component. They show distributions over categories. Although we only have two categories of information to visualize, which are here the current and past dimensions' data space coverage, they can provide a simple and intuitive way to visualize such data. I used a version of parallel set to show the coverage of investigated data spaces from the dimensions. Each line-set corresponds to one dimension and it's divided into values/categories of that dimension.

I used data from the study at chapter 4 and a paper-based version of the visual representation to test it's ability to visualize the dimensions' data space coverage of a real analysis session. Two types of information were extracted and visualized. One is the dimensions' data space coverage. Second is the frequency of investigation of each dimension. The complete paper sketching can be found at Appendix B. The main reason of this preliminary evaluation is to put the visual representation in a real use and recognize if there are any changes or improvements have to be made. There were no changes regarding the visual representation but there were other design questions have to be addressed. One early design choice is to use different tabs of parallel set for the group and individuals analysis. During this evaluation, I didn't differentiate individuals analysis from the group analysis since the goal was to test it's functionality. I evaluated other design questions regarding the application prototyping using a scenario-based interview with UI design experts as following.

5.3.3.1 Scenario-based Interview using Paper Prototype

I did a paper-based interview with two computer science students who I considered as UI design experts. They have a good background in the user interface design as both of them took the User Interface Design and Programming course and worked at EVL. In addition, both participated in the earlier study of chapter 4 and are familiar with the system and can provide feedback on the design development. The paper-based questionnaire is in Appendix B.

One designer didn't support using tabs to show individual's work as this can be redundant data and suggested to possibly highlight on the main view what dimensions individual are working with.

In addition, at this point I haven't had a design choice regarding the co-investigation and one early thought was to use click and highlight on the main view. In this regard, one of the designer suggested to show information of how each data dimension was investigated. As illustrated in Figure 34, clicking on a dimension highlight it's investigated data space coverage in orange and connect it to other data space coverage(s) it appeared with in a chart.

During the implementation phase later, I improved this to show all investigated data space coverage(s) on a dimension (Figure 35 (d) in yellow) and the most recent one will be active and connected to co-mapping peers. Users can click to show co-mapping for other data spaces (Figure 35 (d) in light yellow).

5.3.3.2 The Final Visual Representation Design

As illustrated in Figure 35 (a), each dimension is represented by a line-set rectangle. The line-sets are divided into several blocks representing a discrete number of categories for cate-

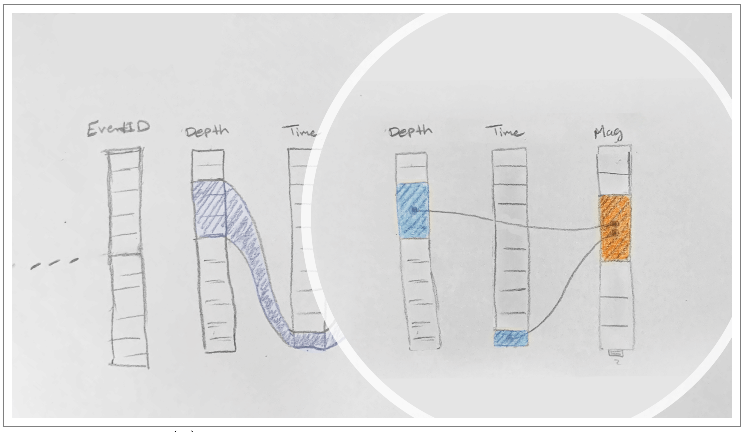


Figure 34: Data space coverage(s) co-investigation: clicking on the dimension *magnitude* shows covered data space (orange) with connections to other data space coverage(s) (in blue) it appeared with in a chart.

gorical dimensions or a distribution of values for numerical dimensions. As shown in sub-figure (b), each dimension is attached with a bar representing the number of charts this dimension appeared within. When users subset a new data subset based on filtering some dimensions, the graph gets updated to reflect the new state of the dimensions' data space coverage as shown in Figure 35 (c). The dimensions of the new data space coverage will be stacked to the left showing the current coverage of the data space (in blue).

5.3.4 Implementation

The visualization is implemented using JavaScript and D3 (83) to visualize the dimensions search space. It runs as a separate application and communicates with PolyVis using SAGE2 server. SAGE2 server handles all the communications required for the application to run, without interfering with PolyVis functionalities.

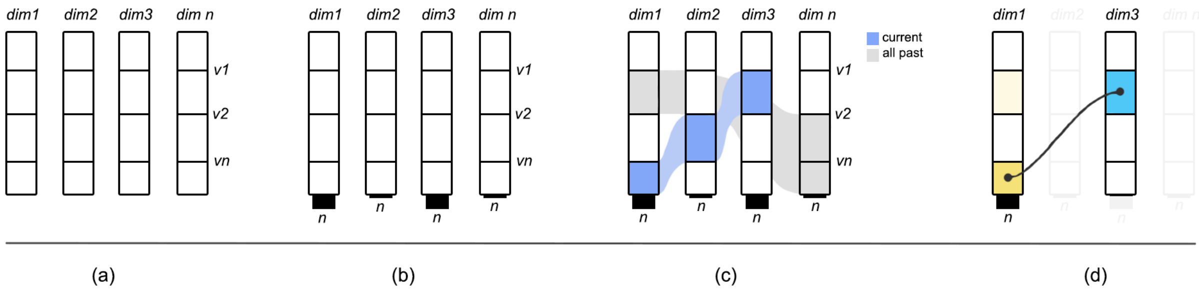


Figure 35: The final visual representation design: (a) Each dimension is represented by a line-set divided into several blocks representing its values distribution (categorical or numerical). (b) Each dimension is attached with a bar representing the frequency of investigation of this dimension. (c) The current dimensions that the analysis is pivoted to will be stacked to the left showing the current coverage of the data space (in blue) and all past data space coverage will be combined into one category (in grey) and send to the back. (d) Clicking on a dimension shows co-investigation information (appearing together in a chart) of its data space coverage(s).

Upon running, the application initializes the data graph as will be described below and start communicating with SAGE2 server to interactively visualize the dimensions search space. As described above, users can interact with the application and freely move it to any position on the big wall display. I used WindowTop (84), a freeware utility, to pin the window of the application and keep it always on top.

5.3.4.1 Pre-processing

The main functionality of the application is to visualize the current and the past dimensions' data space coverage(s) and what covered on each dimension. First, I did a one time pre-processing to extract all the dimension keys from the DB and create their values distribution. The first step is to extract all the dimension keys from the DB and classify them to numerical or categorical keys. For categorical keys, all distinct values were extracted to represent the different categories of the key. For numerical keys, the min, max, and median were used to create a distribution of numerical ranges for the key. These value ranges for all the dimension keys will be used upon running the application to initialize the visualization graph as will be described next.

5.3.4.2 Visualization Graph Initialization

Before drawing the visualization, a graph will be created to form the structure of the visualization. In other words, the different components of the visualization that interact with each other when updating the visualization. The graph of the visualization is composed of three different components: nodes, blocks, and links (Figure 36). A node represent a dimension and has features like name, state, and investigation number of this dimension. All dimensions are represented as nodes. Each node has a number of blocks. These blocks represent the different data space coverage(s) on the dimension (ranges being selected for investigation at some point). Each block has features like it's node source, starting and ending values, state, list of co-investigation nodes, and list of co-investigation blocks.

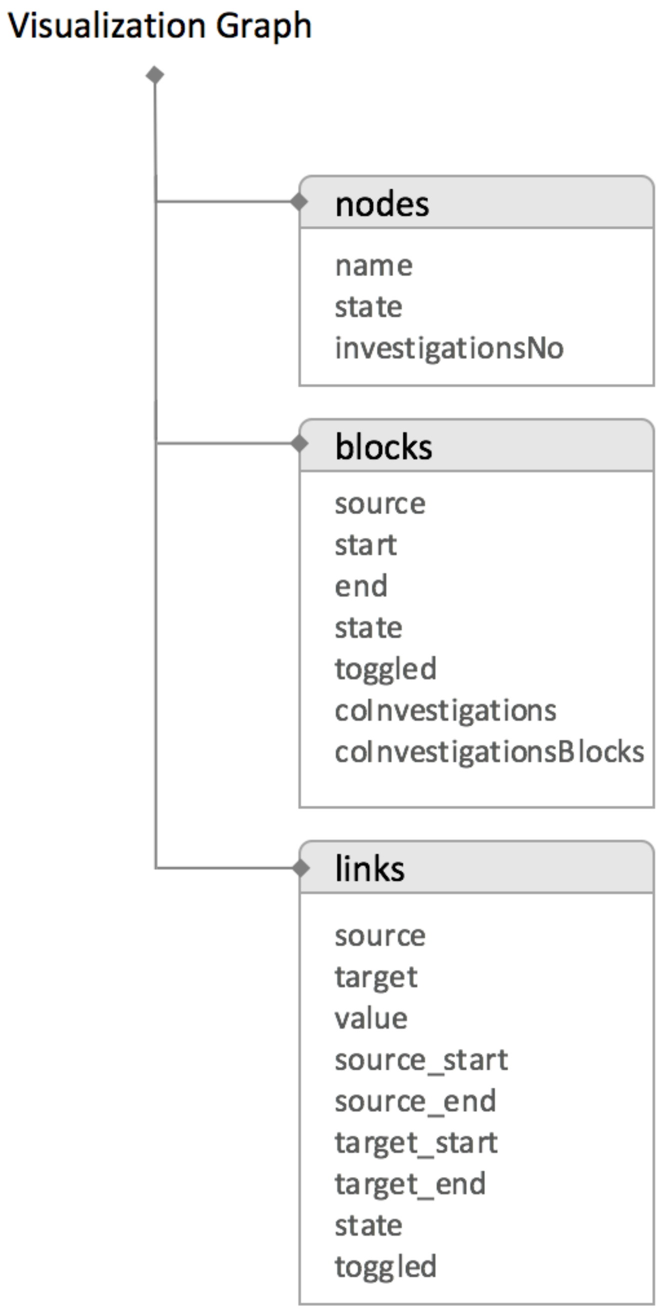


Figure 36: The visualization graph consists of nodes, blocks, and links.

A link shows proportions of data selected on two nodes: source and target. These proportions are defined as blocks on the two dimensions. Links do not have a flow and they represent the coverage over multiple nodes. A links has features like source and target nodes, starting and ending values on the source and on the target nodes.

At the start of the application, a visualization graph will be initialized using the list of nodes and their computed value distribution. Then, an initial visualization will be drawn using this graph. The graph will be updated during the analysis by adding blocks and links representing the state of the search space. The visualization will be re-drawn each time the graph is updated. This graph structure facilitates the heavy interaction between the different components when adding and sorting the blocks and links.

5.3.4.3 Update Dimensions' Data Space Coverage

When a new visualization is created using any of the devices, the graph gets updated to reflect the new state of the dimensions' data space coverage. The server sends the new visualization schema and all required information is extracted to update the graph.

The visualization schema contains information about the visualization including the mongoDB query to the database. As shown in the example in Figure 37, the dimensions names and the new coverage values for each dimension will be extracted. The node's state that correspond to each dimension will be activated if it's not active yet. New blocks and links will be created for each dimension based on the new coverage values.

```
[
    {"$match":{"$and":[
        {"Time":{"$gte":"2002-12-31T06:00:00.000Z","$lte":"2003-12-31T06:00:00.000Z"}},
        {"Magnitude":{"$gte":2,"$lte":6}},
        {"Location":"California Area"}]}
    },
    {"$group":{"_id":{"t":{"$dateFromParts":{"year":{"$year":"$Time"},"month":{"$month":"$Time"},"day":1,"hour":12}}},
    "total":{"$sum":"$Magnitude"},"count":{"$sum":1},"avg":{"$avg":"$Magnitude"},"min":{"$min":"$Magnitude"},"max":{"$max":"$Magnitude"}}},
    {"$project":{"_id":"$_id.t","total":"$total","count":"$count","avg":"$avg","min":"$min","max":"$max"}},
    {"$sort":{"_id":1}}
]
```

Figure 37: mongoDB query is used to extract dimensions and their new coverage values.

5.3.4.4 Update Dimension's Co-investigations

In addition to the coverage information of the data space, the graph tracks the co-investigations of each dimension. As shown above, when the coverage of each dimension is computed, it's co-investigation information with other dimensions will be recorded as well. As shown in Figure 35 (d), clicking on a dimension node shows co-investigations for its different data space coverage(s) where each data space coverage is used in one or more visualizations.

CHAPTER 6

EVALUATING THE EFFECTS OF VISUALIZING DIMENSIONS SEARCH SPACE ON EXPLORATORY VISUAL DATA ANALYSIS

6.1 Introduction

The focus of this chapter is to evaluate **the effects of visualizing the coverage of the dimensions search space on exploratory visual data analysis.**

The study presented in this chapter tests three hypotheses. These hypotheses were motivated by observations from the earlier study in chapter 4 and results of prior research in related contexts. In the study of chapter 4, I observed that participants evolved in their analysis by investigating selected subsets from the data spaces, and the analyses were mostly depth oriented in these data subsets. They typically rely on memory to remember what already explored and what to explore next. Earlier studies of visualizing the frequency of investigation of dimensions (60)(37) proved a positive effect on exploratory visual analysis. In addition to frequency of investigation, I also reveal information about what dimensions' data space coverage were investigated when analysts pivot the analysis between different subsets of the data space, and I hypothesize that this information will support the following hypotheses:

H1: Reducing the decision cost. The cost of deciding what to explore next "Gulf of Goal Formation" bears the cost of "finding a data subset to explore" and "choosing amongst interface options" (85). The visualization of the dimensions search space communicate the dimensions and their data space coverage which I hypothesize will help participants build their interest of investigation and select the next course of analysis, and hence, reducing the decision cost.

H2: Increasing the breadth of the analysis. Supporting the breadth of exploratory visual analysis has been a central goal in the visualization community (63)(61). By providing the visualization of the dimensions search space, I aim to guide participants to broad their analysis.

H3: Increasing formed questions and observations. Analysts constantly form questions and observations during exploratory visual data analysis. I hypothesize if the above hypotheses were supported, that participants will ask more questions and find more observations.

I hypothesized that visualizing the coverage of the dimensions search space will reduces the decision cost, increases the breadth of the analysis, and increases formed questions and observations. The results from the study support the first and the second hypotheses as will be described below.

6.2 Evaluation Study

I conducted a between-groups study to evaluate the effect of visualizing dimensions' data space coverage and co-investigation on exploratory visual analysis. The study contained two conditions: baseline and full versions where half of the groups used a baseline visual analysis tool and the other half of the groups used a full version of the tool enhanced with a visualization of the dimensions search space. The main reason for choosing a between-groups study is the difficulty of conducting the study with two conditions per session as one condition can last for about 70 minutes. In addition, with between-groups study we can eliminate the transferred knowledge about the domain after completing one condition. The independent variables are the full and baseline tools. The dependent variables are the total number of created visualization, the views generation rate, the number of unique attributes combination, the number of formed questions and observations.

6.2.1 Participants

I recruited 33 participants as 11 groups of 3. I did a pilot study with one group and completed 10 sessions for the formal study. The recruitment pool was UIC College of Engineering undergraduate and graduate students. Participants were 27 male and 6 female students between the ages of 21 and 35 years old. They participated in the study for the duration of 45min-2hrs. Participants self-reported that they have basic background in basic charts.

6.2.2 Software

For this study, I used two versions of the tool. The baseline version is the PolyVis system described in chapter 3. The full version is the PolyVis tool enhanced with an interactive visualization of the dimensions search space as described in chapter 5. The visualization of the dimensions search space is integrated as an independent application running on the big display.

6.2.3 Datasets and Tasks

Each group performed visual analysis tasks using two geoscience datasets. The first dataset contained information about earthquake incidents in Oklahoma and California from the years 2000 to 2010. The second dataset contained information about the fracking activities in Oklahoma and California also from the years 2000 to 2010. I collected, cleaned, preprocessed and stored datasets in NoSQL database using MongoDB. The earthquake dataset was provided courtesy of http://service.iris.edu/ and the Wells injection dataset was provided courtesy of http://www.occeweb.com/. The earthquake dataset consisted of 24555 records and 12 attributes while the Wells dataset consisted of 5138 records and 9 attributes. These datasets have attributes with similar meaning such as the location, the time, and the depth. Earthquakes dataset has other attributes like magnitude while Wells dataset has other attributes like well status, well type, injection volume, and injection pressure.

Each group completed two tasks, with focused and open questions. In the first task, the subjects were given focus questions that can be answered by creating one or two visualizations. I opted for the focus questions approach to be a practical tutorial on how to use the system. Subjects were then asked to explore the earthquakes events and wells injection activities and identify trends/observations in the data.

6.2.4 Setup and Data Capture

The study was conducted in the Continuum room at EVL. The room is approximately 10.61 by 5.59 meters, equipped with a high-resolution large display. Overall display size is approximately 7.3 by 2.05 meters at a resolution of 11,520 by 3,240 pixels. Other portable devices were placed on a table in the middle for use during the study: one MacBook Pro (macOS Sierra, 2.4 GHz Intel Core i5), one 8” Samsung - Galaxy Tab A (32GB, Android 9 (Pie)), one 10” Samsung - Galaxy Tab A (64GB, Android 9 (Pie)), and one Microsoft HoloLens 1 (Windows Mixed Reality OS, Intel 32-bit (1GHz) CPU, 2 GB RAM).

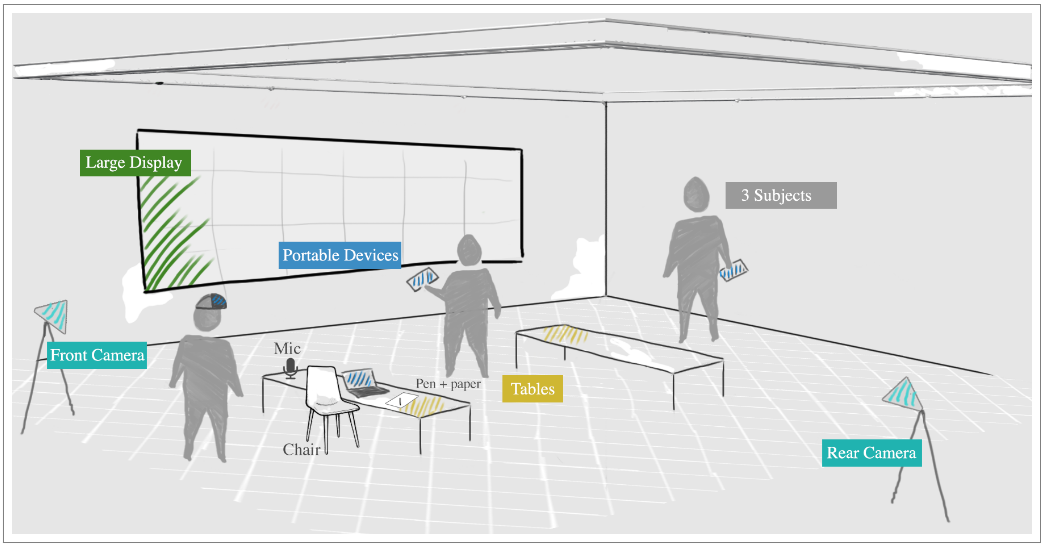


Figure 38: Illustration of the study setup.

For safety protocols, two tables were placed in the middle as shown in Figure 38 and Figure 39 to compose three movement areas. Although users were free to move around, these spaces were created to maintain social distancing. Users can walk back and forth to the big display but they should be standing at 6' distancing. The tables help them keep an appropriate distance between them.

In addition to maintaining social distancing, users were instructed to not share devices between them. When a user pick a device to work with, no one is allowed to use that device. Therefor, the group divide roles on using devices at the beginning of the task. Since only one user will be using the laptop all the time, a chair was provided for the laptop user.

Systems usage logs were collected from all deployed devices. I wrote a script to capture all interaction events with the system. Each log included the device id and type, the action type, and the timestamp. In addition, all created visualizations and selected attributes were collected. The study was video recorded using two cameras, one showing the full room from behind and one showing the subjects' interaction with the large display from the front. In addition, a microphone was placed at the table for audio recording. The setup is illustrated in Figure 38.

Figure 39: Participants examining a set of created visualizations on the large display. The middle window shows the visualization of dimensions search space.

6.2.5 Procedure

Prior to the study start, three disinfected disks and chairs in three apart locations in the room were prepared for the participants. Participants were greeted and seated in the specified locations. They were provided with consent and media forms to read and sign. They spent 2-5 minutes to read and fill out forms. Once participants finished completing forms, I started with a 5-minutes introduction to give the users an overview of the tool and the task.

Next, participants were given the first task of a single question that could be solved by creating one or two visualizations. I opted for this approach as a practical tutorial on how to use the system and to familiarize themselves with the tool. Participants took turns in completing this task. All devices were disinfected in between. They were told to feel free to ask for a clarification or instruction at any time during this task. Each individual took from 5 to 10 minutes to complete this task. In total, they spent 20-30 minutes on this part.

Next, they started the main task of an open-ended exploration of the earthquake events and wells' activities. They were asked to explore both data sets and write down any observations or findings. There was no time limit and they can stop at any time. Participants were instructed to work on this task together but to maintain social distancing and to avoid sharing of devices. They can discuss before starting the task who would be responsible for each of the devices. I left it to the subjects to make this decision rather than make a determinate assignment of devices. They can use the laptop and tablets to create visualizations and push them onto the big display. The visualization of the dimensions search space is placed in the middle of the large display. Subjects can interact with the application and freely move it to any position on the big wall display but it will always be on top of other views using WindowTop (84) utility. This task was exploratory in nature and took between 25 to 70 minutes to complete.

6.2.6 Coding and Data Analysis

I collected data in the form of recorded videos, audio, and system logs. I wrote a script to capture all created visualizations, their visual encoding, selected attributes, and timestamps. About 506 minutes of videos were collected (an average session time of 54.4 minutes for full version and 46.8 minutes for baseline version).

I did two passes of coding. In the first pass, I used a video coding tool called ChronoVis (4) to code the study videos using annotation timelines (Figure 40. As shown in Figure 40, I created annotation timelines with distinct colored codes. These timelines are used to code the created charts, the indirect questions, the direct questions, and the observations.

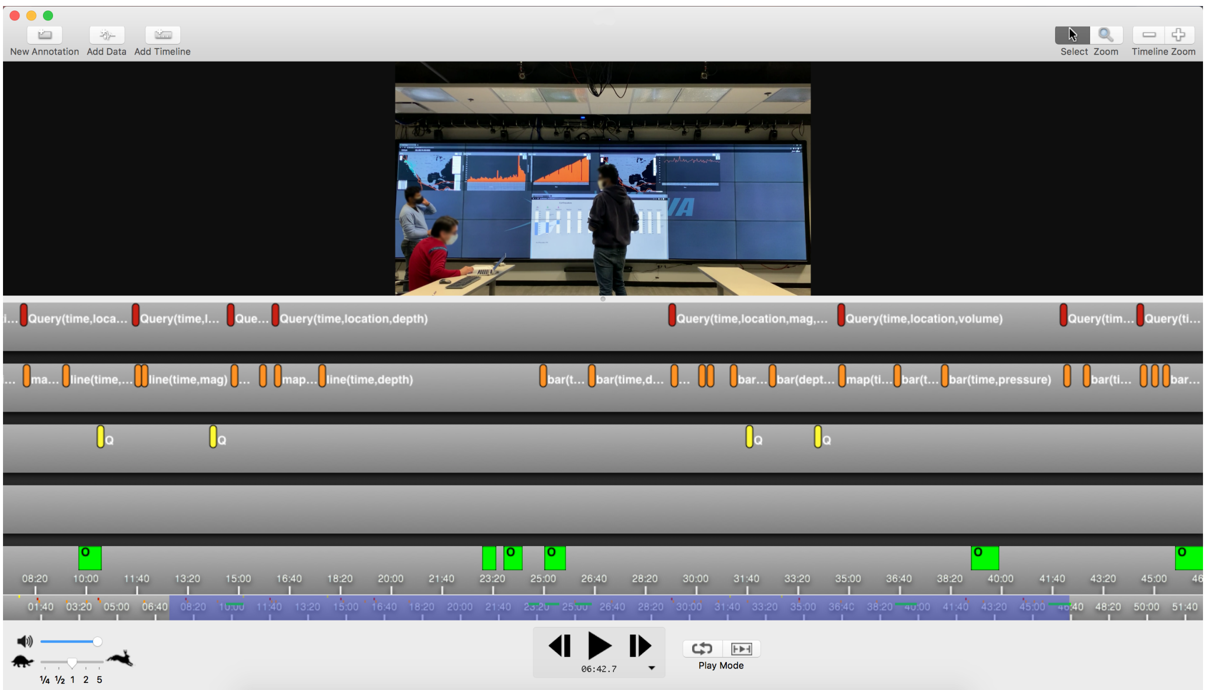


Figure 40: ChronoVis (4), a video coding tool was used to code sessions' videos.

Then, annotations were extracted and quantified for each session. Two coders [1] completed a second pass of coding of four randomly selected videos. For time limitation, they worked on the videos from the first pass and edited the timelines for any changes they want to make. With this approach, the inter-coder reliability agreement was 95% and 98% with both coders respectively.

Next, I did another pass of qualitative coding of the analysis workflow to create search tree graphs for the sessions. The purpose of this is to use the search trees aspect ratio as a second metric of measuring the breadth and depth of the analysis. As shown in Figure 46 (a), the width of the tree represent the depth of the analysis while the height represent the breadth. When participants start an analysis path by analyzing a new subset of the data, a new branch will be created increasing the height (breadth) of the tree. When creating a new chart of unique attributes along this branch, a forward edge will be add increasing the width (depth) of the tree.

---

[1] Krishna Bharadwaj and Arthur Nishimoto.

6.3 Findings

In the following subsections, I present the results of providing the visualization of the dimensions search space on the decision cost, the breadth of the analysis, and the number of questions and observations.

6.3.1 Gulf of Goal Formation

The visualization of the dimensions search space played a central role in data selection and attributes co-investigation. The ability to see values distribution of each dimension and their coverage, assisted participants to easily know what data ranges each dimension has and build their interest of investigation around that.

Selecting what to explore next, also known as the "Gulf of Goal Formation" (85), is a major component of interaction costs in information visualization. Lam defined this cost as the cost of "finding a data subset to explore" and "choosing amongst interface options" (85). When a team of analysts collaborates using multiple devices to work on an analysis task, the decision cost can be higher due to short-term memory and the recency effect. Analysts in collaborative settings need to understand what was investigated by the team and what was left. Visualizing the dimensions search space can communicate to the team what dimensions have been investigated (and in what combination) and what were left. I hypothesized that the visualization of dimensions search space would reduce the decision cost (***H1***).

Based on the definition of the decision cost, one way to measure it is to calculate for every view the time required to decide on the data, choosing amongst interface options, and creating the view. In addition, other factors can contribute to this cost such as findings and hypothesis formation, as they can affect the nature and the number of next created views.

I used an approximation approach to measure the decision cost based on the assumption that *generating a larger number of views in the same time period indicates that the time required to select a data and create a view is lesser*. I used an approximation approach to measure the decision cost because there is no consistent way to measure the decision cost in the literature, and hence, a new approach is needed to be developed and tested through a qualitative analysis. Therefore, by following the assumption above, I estimated an approximate of the decision cost measure for the whole session.

For the approximation approach, I calculated the rate of producing views for each condition. Initial statistics showed that full version groups created more charts than baseline groups. To eliminate the effect of sessions' time, I chose the views' generating rate as the decision cost measure. The higher the rate of views generation, the less the cost of goal formation.

First, I counted the number of created visualizations by each group. Full version groups created an average of 43.6 views (SD = 18.15), versus 20.4 (SD = 13.07) for baseline version groups. A tow-tail independent t-test showed that full version groups generated more views than baseline groups (*t = 2.3198, df = 8, p = 0.0489 at p ¡ .05*).

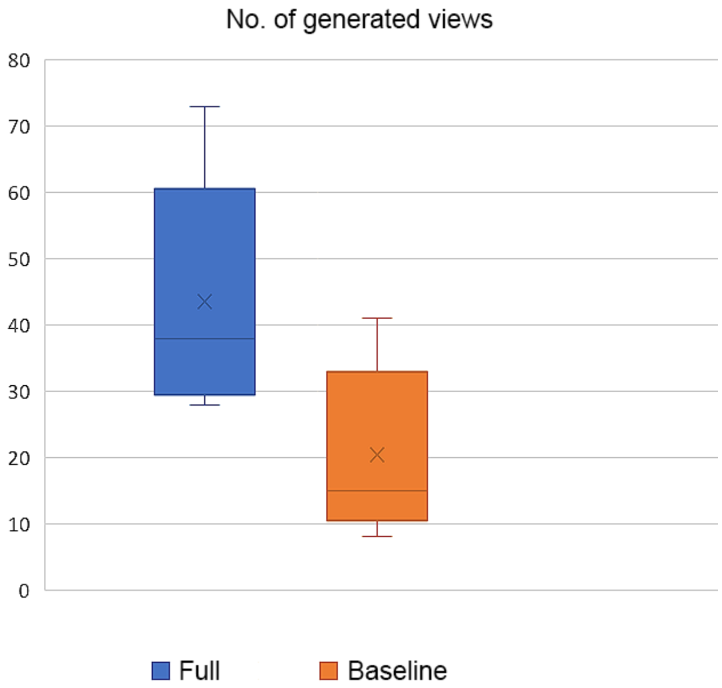


Figure 41: Total count of generated views in each condition.

Then, I calculated the views generation rate by dividing the number of created views by the session's time. Full version groups created an average of 0.7880 views per minute (SD = 0.1987), versus 0.4360 views per minute (SD = 0.1781) for baseline version groups. A tow-tail independent t-test showed that full version groups generated views at a higher rate than baseline groups (*t = 2.9498, df = 8, p = 0.0184 at p ¡ .05*). I can conclude that higher rate of views generation indicates a reduction in decision cost.

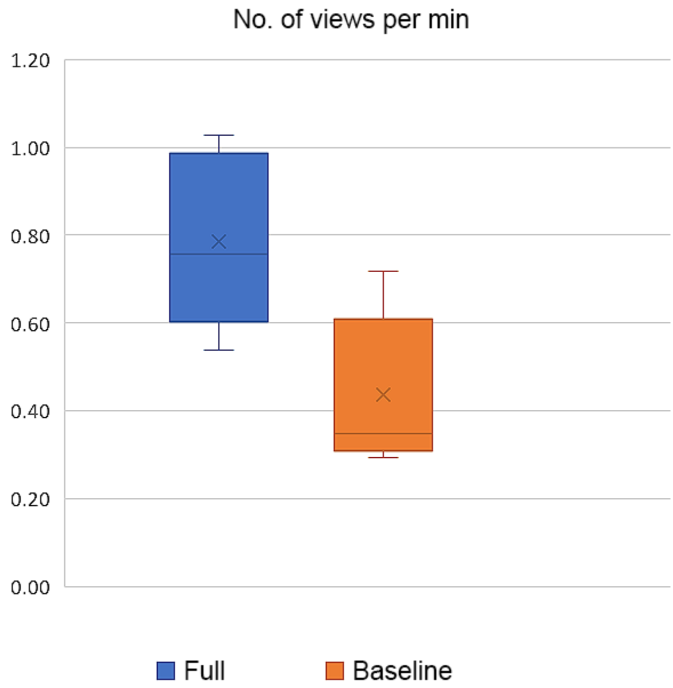


Figure 42: Rate of generated views per minute in each condition.

I can also speculate that baseline groups generated views at a lower rate because of short term memory and the recency effect. They typically rely on memory to remember what already explored and what to explore next. Due to the recency effect and the short-term memory, they tend to remember the most recent exploration. These factors also affected the breadth of the analysis as will be described in the next subsection. The tendency to recall the most recent items encouraged a depth-oriented exploration.

6.3.2 Breadth vs Depth Oriented Analysis

Supporting the breadth of exploratory visual analysis has been a central goal in the visualization community. Battle and Heer (63) called for a greater exploration breadth after they studied exploratory visual data analysis using Tableau and found that the analysis is primarily depth-oriented. Voyager (61) promotes a breadth oriented exploration by providing users with useful visualization recommendations to increase the coverage of unique data attributes. They evaluated how visualization recommendations increase the breadth of the analysis by evaluating the number of unique attribute combinations explored by users.

To evaluate the breadth of the analysis (***H2***), I used two metrics as will be described below. The first metric is counting the unique attributes combination similar to the approach of Wongsuphasawat et al. (61) and Sarvghad et al. (37) . The second metric is using the aspect ratio of analysis search trees as used by Battle and Heer (63).

6.3.2.1 No. of Unique Attributes Combination

Similar to the approach of Wongsuphasawat et al. (61) and Sarvghad et al. (37) to measure the breadth of the analysis, I counted the number of created views of unique attributes combination.

I considered the attributes combination is unique if this set of attributes visualizes a new subset from the data space. For example, the average magnitudes over time for the year 2000 is unique from the average magnitudes over time for the year 2010. I considered this approach as it represents a broader exploration in the data space. In the cases which I considered the attributes combination is not unique, participants reproduced the same attributes combination to change

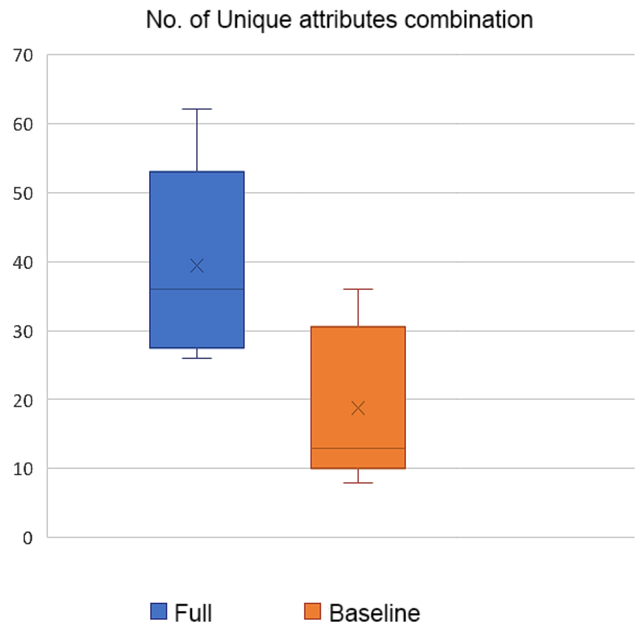


Figure 43: Total count of unique attributes combination in each condition.

the visual representation (e.g. scatter to line chart) or to change the aggregation function (e.g. avg. to max). The number of unique attributes combination was mostly proportional to the number of created visualization.

Full version groups explored an average of 39.40 unique attributes combination (SD = 14.42), versus 18.80 (SD = 11.52) for baseline groups. A two-tail independent t-test showed that full version groups explored more unique attributes combination than baseline groups (t = *2.4963, df* = *8,* p = *0.0372 at* p ¡ *.05*).

6.3.2.2 Aspect Ratio of Search Trees

As a second metric, I used aspect ratios of search trees similar to the approach of Battle and Heer (63) to measure if search trees tend to be breadth or depth oriented. When participants start an analysis path by analyzing a new subset of the data, a new branch will be created increasing the height (breadth) of the tree. As shown in Figure 44, they started the analysis with a data subset based on the attributes time and location. Then, they pivoted the analysis into a new data subset (second row) based on the attributes time, location, and magnitude. When creating a new chart of *unique* attributes along each branch, a forward edge will be added increasing the width (depth) of the tree. Figure 44 shows that they created three charts in the first analysis path. Hence, the width of the tree represent the depth of the search while the height of the tree represent the breadth.

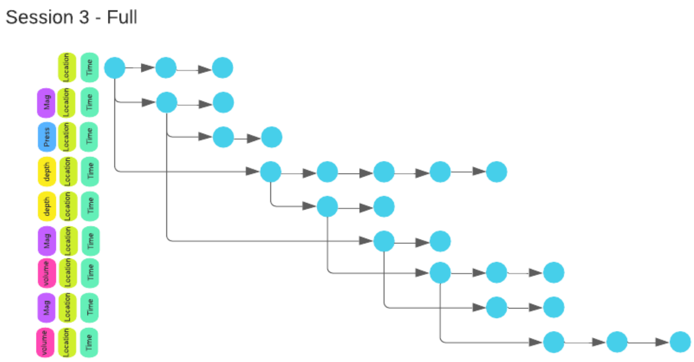


Figure 44: Search tree S3

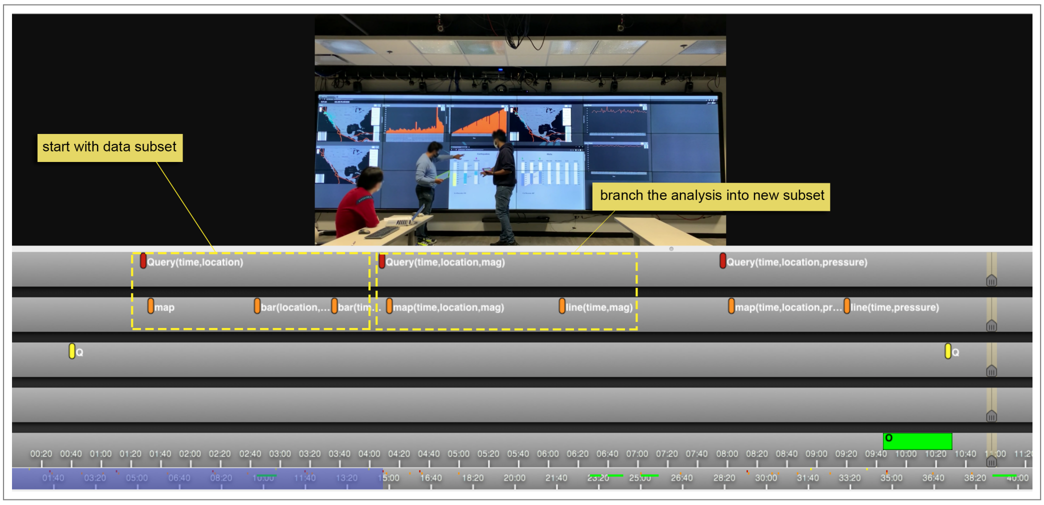


Figure 45: Coding search tree branching

Figure 46 shows the distribution of aspect ratios of all sessions compared to the total average. Full version search trees have larger aspect ratios showing greater breadth than baseline version. Statistically, full version search trees have an average of 0.80 aspect ratio (SD = 0.07) versus 0.68 (SD = 0.11) for baseline version. A one-tail independent t-test showed that full version search trees have larger aspect ratio than baseline groups (*t = 2.0322, df = 8, p = 0.0383 at p ¡ .05*).

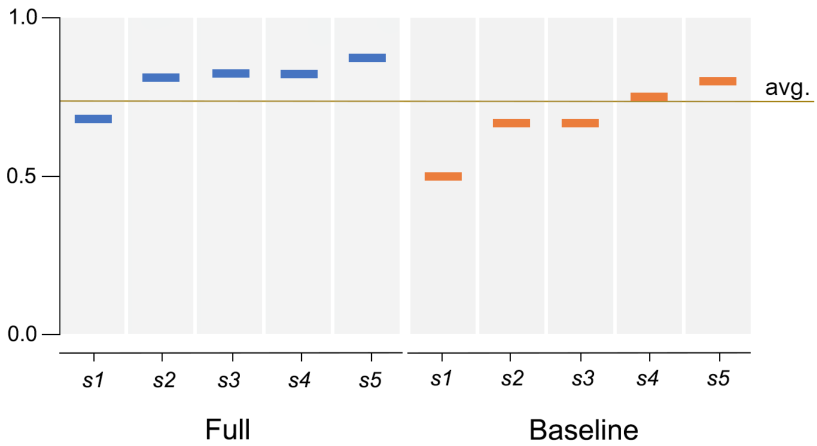


Figure 46: Distribution of aspect ratios of all sessions (breadth/depth).

As shown in the quantitative analysis above, the tool helped participants to explore more dimensions and coverage of the data. I didn't perform a qualitative analysis on how the tool helped participants to build their interest of exploring the data. However, the search trees show that baseline groups pivoted their analysis around the *time* and *location* attributes in 4 out of 5 sessions, while full groups pivoted their analysis around a larger and divers sets of attributes in 4 out of 5 sessions. We can infer that they were able to build their interest in a more broader way. As shown in the example figure below, they started the analysis around some attributes (mag, pressure, and depth) in the same location and time period (open-ended exploration). Later, they planned the analysis to explore high magnitude and large volume for the same time period in California. Then, the same were explored in Oklahoma.

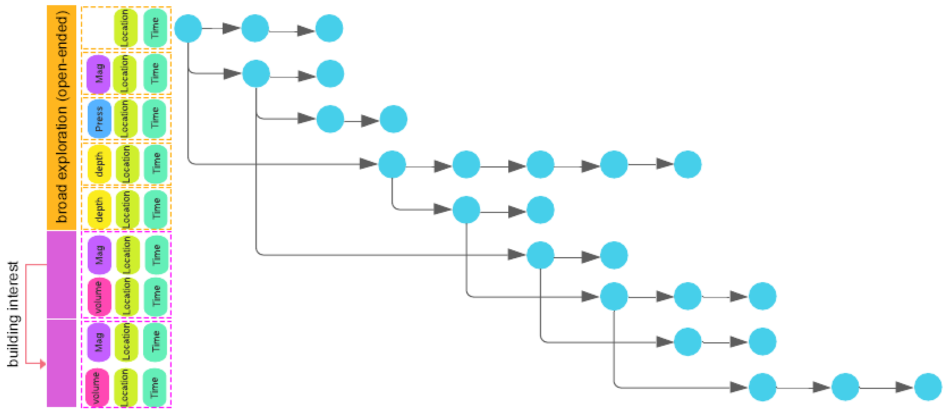


Figure 47: Building an exploration interest.

A thorough qualitative analysis can help understanding how they build or progress the analysis and if there are some patterns. In addition, it can help understanding what types of knowledge, questions, or findings have an effect in progressing the analysis and in what degree.

6.3.3 Questions and Observation

When working on an analysis task, participants ask questions and come up with observations about the data. Liu and Heer defined a question as "*an indication of desire to examine certain aspects of the data*"(86). In both studies of chapter 4 and 6, the groups engaged in discussion and ask questions either as direct questions that end with a question mark or by indicating the interest of exploring a certain aspect of the data. I considered both cases to be valid questions although direct questions that end with a question mark were not common.

In addition to questions, I coded observations made about the data. An observation is a piece of information they derive from a single or multiple visualizations.

The results didn't fully support the hypothesis of increased questions and observations (***H3***). A two-tail independent t-test showed that the number of questions was marginally different between full and baseline condition (*t = 2.1871, df = 8, p = 0.0602 at p ¡ .05*) and the standard deviations were statistically different with a low variance value (*.107*). In addition, a two-tail independent t-test showed that there is no statistically significant difference between the number of observations of both conditions (*t = 1.43813, df = 8, p = 0.188341 at p ¡ .05*).

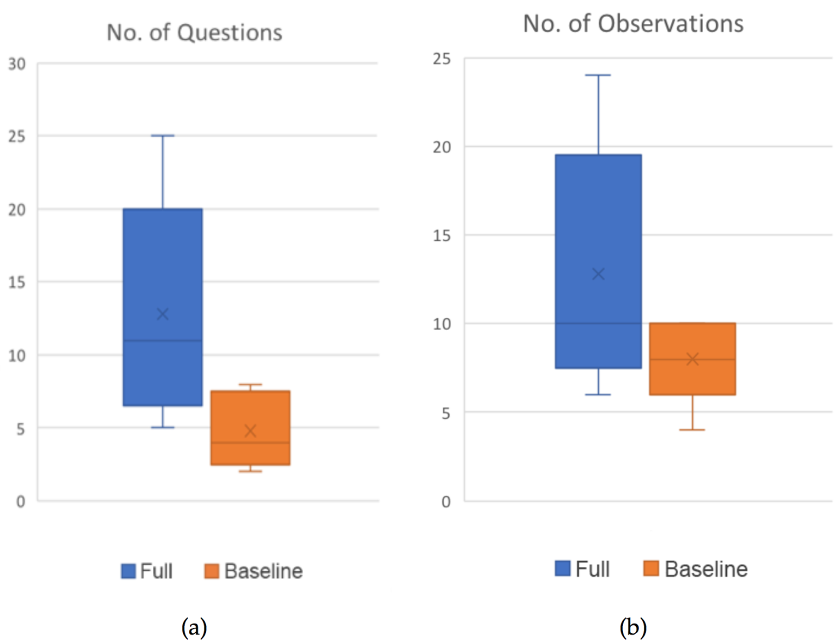


Figure 48: No. of questions and observations in both conditions.

6.4 Conclusion

In this chapter, I evaluated the effect of visualizing dimensions search space on exploratory visual data analysis. The results showed a positive correlation in terms of generating more views and investigating more dimensions.

During the design phase of the visualization tool, we concluded to have the tool in a shared public medium, the large display. Therefore, it is crucial to consider where and how to implement the tool in different contexts. It should be incorporated in a way that doesn't hinder the team or the individual's work and doesn't cause any cognitive load due to browsing activities or additional interactions. In this environment, the large display offered easy access to the tool without hindering the analysis work. Although the mechanism can be used in different contexts such as distributed collaboration, the design and the context of use should be considered properly.

Since the analysis in this environment is performed in a scattered fashion, an individual can create a view while another member is involved in individual analysis and is not aware of what was just created. As shown in Figure 45, the middle user is checking what the right user just created. Therefore, it can be more beneficial in situations where the nature of work is scattered either in co-located or distributed settings. Although it can be adapted in different contexts, it suited this environment, first, due to the nature of work being scattered. Secondly, having a shared public medium for incorporating it effectively into the environment.

6.5 COVID-19 Safety Protocols

6.5.1 Research Study Restart

This study was conducted during a global pandemic!

In response to the COVID-19 pandemic in March 2020, all educational activities were moved online and all on-site research studies were halted. To ensure the continuation of research, the Office of the Vice Chancellor for Research at UIC established a COVID-19 Research Recovery Committee to help investigators and Deans prioritize IRB approved studies for reopening. All Study Principal Investigators (PIs) who have protocols they are prepared to restart were asked to apply for an assessment of their studies ability to re-open. I applied for UIC COVID-19 Human Subjects Research Restart to review the reopening of my approved IRB protocol of this study (Appendix I). I amended my protocol with safety measurements that abide to the campus/college/laboratory requirements as described below. The research restart application was approved by OVCR for restart on or after September 28, 2020, assuming that Illinois and Chicago stay in Phase 4 of Restore Illinois and Protecting Chicago COVID-19 recovery plans.

6.5.2 UIC/EVL Safety Measurements

UIC and EVL established a set of safety protocols regarding entering and working inside buildings/labs. I took the safety measurements that abide to these requirements. According to campus wise safety protocols, all students, including PIs and study participants, must follow the following safety measures:

- Submit temperature self-reporting form one hour prior to entering the campus

- Take the saliva testing once on campus
- Wear a mask
- Maintain 6 feet distancing

In addition to these safety protocols, I maintained the following EVL safety requirements before, during, and after the research studies:

- Schedule the study on the EVL calendar (with no. of people)
- Get a lab entry approval from Prof. Andrew Johnson for all participants who will enter the EVL lab
- All people should sanitize when entering/exiting the lab rooms
- Disinfect the equipment before the study
- Submit EVL exit form after completing the study

The study involved three participants at a time and the total time took about 70 minutes on average to complete. The study took place in the Continuum room at EVL. The room is large enough to maintain 6' distancing. Two tables were used to divide the room into three large spaces to ensure safe movements of the participants without getting closer to each others.

Subjects were asked to use the devices to complete the analysis task. Prior to the task start, subjects divided roles among them to use each of the devices. They were instructed to maintain 6' distancing during the study and to not share devices between them.

Figure 49: Hand sanitizers, masks, gloves, and disinfection supplies were provided around the lab rooms.

CHAPTER 7

CONCLUSION AND FUTURE WORK

Parts of this chapter were previously published as: Alsaiari, A. and Johnson, A. (2019). "Towards Understanding Collaborative Visual Data Analysis in Multi- Device Environments". In *2019 IEEE VIS*.

In the era of big data, visual analytics became rarely a solitary activity. Analysts from different backgrounds need to work together to contribute their contextual knowledge and create a better understanding of their data. The growing interest in leveraging ecosystems of digital devices that go beyond a single desktop for collaborative visual data analysis have great potential in supporting collaboration. First, analysts have larger space than what one device can o er, to visualize and work on more data. Also, it allows the distribution of the data to the appropriate device for visualization. In addition, it allows different collaboration styles by enabling individual and group work. Therefore, visual data analysis is moving from single user/single device into multiple users/multiple devices.

This dissertation explored visual data analysis in multi-device environments in an attempt to provide an understanding of the visual analysis process and how we can support it for efficient exploration. Using an activity-centered approach, I presented an analysis of users' experience and characterization of the analysis behavior. I have demonstrated through an evaluative study that visualizing the dimensions search space can positively affect the analysis by reducing the decision cost and increasing the breadth of the analysis.

I conclude the dissertation by outlining the main contributions and areas of future research.

7.1 Contributions

This dissertation investigates the following research questions:

- RQ1: What is the complex picture of users' experience during a collaborative visual data analysis in a multi-user multi-device environment?
- RQ2: What is the characterization of the analysis process in this environment?
- RQ3: What are the effects of visualizing the coverage of the dimensions search space on exploratory visual data analysis?

In order to answer these questions I conducted two user studies. The study presented in Chapter 4, addresses the first and second research questions. The third research question is addressed through an evaluative study presented in Chapter 6. The following outlines the contributions of this research and a discussion of how they address the above-mentioned research questions.

C1. A hybrid analysis approach to understanding the collaborative visual data analysis in a multi-user multi-device environment

The HCI community gradually realized that focusing on the task only for the design and evaluation of modern systems is not sufficient anymore. The aspects of why and how the task is performed is now a central focus in HCI community. Essentially, it is important to develop a systematic analysis approaches for the analysis of complex activity in multi-device ecosystems that relate the different key components of the big picture. Ghaoui (87) described the impact of activity-theory on HCI as "it served as an analytical framework for design and evaluation of concrete interactive systems, and offered a set of concepts for capturing the context of use". It better suits our goal for understanding the context of group activity in multi-user multi-device ecosystems. Therefore, I based the presented approach on the principles of activity-theory and visualizations reference models. The result is a structural analysis based on the identified activity actors. It frames the complex environment as a network of actors: users, tools, and tasks. In Chapter 4, I demonstrated the approach by analyzing group activity in the presented study. The work based on this approach was submitted to CSCW 2022 titled as "*An Activity-Centered Approach to Understand Collaborative Visual Data Analysis in Multi-Device Environments*".

C2. A two-level characterization of the analysis structure

Supporting exploratory visual data analysis is essential when multiple analysts collaborate using multiple devices. Yet, we still have no full understanding of how the iterative process of analysis unfolds in complex settings. Through exploratory study, I found that the course of the analysis happens at two levels. Within each level, I observed a set of exploration patterns. In Chapter 4, I presented a categorization of the analysis structure in such a complex environment and discussed the implications of device affordances on this categorization. I also discussed this categorization in relation to the current structural assumptions of exploratory visual analysis.

The results of this work appeared in EuroVis2020 titled as "*Evaluating Strategies of Exploratory Visual Data Analysis in Multi Device Environments*".

C3. The Effects of Visualizing Dimensions Search Space on Exploratory Visual Data Analysis

A key contribution of this dissertation is demonstrating the positive effect of visualizing dimensions search space on exploratory visual data analysis. When multiple analysts work together in a setting beyond the single desktop, it can be difficult for analysts to keep track of all the prior analyses as they rely on building a mental model of the analysis flow which can render more challenges due to many visualizations. In addition, factors such as recency effect and the short-term memory lead to focusing on the most recent items which promote a depth first analysis. In Chapter 6, I evaluated these effects in a between-groups study and the results showed that visualizing the dimensions search space reduced the decision cost and increased the breadth of the analysis. The initial result of this study was submitted to VIS2021 and full results will be submitted to CHI2022 or VIS2022 titled as "*Evaluating the Effects of Visualizing Dimensions Search Space on Exploratory Visual Data Analysis*". .

7.2 Future Work

This research is an initial step towards understanding exploratory visual data analysis in settings beyond the single desktop. There are remain open areas for future research. In chapter 4, I presented a synthesized characterization of the analysis structure from observed analysis behaviors. Although it can help us to build an understanding of exploratory visual analysis, this characterization is grounded to the study setup. Analysis behaviors can be evaluated in different collaborative settings to find commonalities and differences and derive a comprehensive definition of exploratory visual data analysis. In addition, the presented categorization of the analysis structure and strategies is what I observed using two datasets with a few sets of attributes. Therefore, we cannot generalize the observations to other types of datasets, which may contain dozens of attributes. Exploration strategies may differ with large datasets of many attributes.

Additionally, in the first study I collected the position and orientation of subjects and devices. However, this data has never been analyzed as the spatial usage was not the focus of this research. It would provide insights about the use of the space and the complex F-formation of users and devices. One research direction is to study and analyze the use of the space and the complex formations of users and devices to derive a new model that describe the user-user, user-device, and device-device F-formation in light of the work by Lee et al. (88).

The presented approach of visualizing dimensions search space has limitations. One limitation is that it combines all the past analysis into one category. Therefore, it visualizes the past analysis coverage as a whole and doesn't provide a complete information about the history of it. One possible solution is to use history mechanisms such as history trees to enable the navigation in the past coverage of the dimensions. Therefor, there are rooms for improvements for the visual representation design. In addition, most of the study results were quantitative-based. For future work, a qualitative analysis can be performed to analyze how subjects interacted and often looked at the dimensions search space visualization. Additionally, how communication and discussion may differ with and without the search space visualization. Another interesting aspect of the qualitative analysis is to analyze the quality of the exploration to see if they created more useful visualizations or explored more interesting aspects of the data.

The presented approach falls into the category "*Orienting*" in the scale of the guidance degree according to Ceneda's model. A direction of research here is to explore different "*Directing*" or "*Prescribing*" guidance mechanisms. Directing mechanisms provide the user with different options that differ in terms of quality or costs and the user has a freedom to follow or not to follow these options. Using visualization recommendations is one type of Directing mechanisms. Prescribing guidance approaches provide a higher degree of guidance through automated process that operates on their own to generate desired results. Although these approaches have been evaluated in prior research, their design space for multi-user multi-device settings has not been explored yet.

Lastly, in identifying the performed analysis activities, I focused on analysis tasks that provoke interactions with the visual analysis system (Table II). These include creating new views, exploring views, manipulating views, and analyzing views. These types of actionable activities falls into the exploration loop according to the knowledge generation model by Sacha et al. (89). Therefore, a related aspect here that can be investigated in future work is to further study and analyze the additional reasoning processes of knowledge generation. These include, in addition to exploration processes, verification and knowledge generation loops (89). While exploration loop includes interactions with the visual analysis systems, verification loop includes cognitive process of developing insights and verifying hypotheses. It is a central loop in knowledge generation process as it develops insights and combine them into the problem knowledge. The knowledge generation loop is challenging because it depends on the user and other factors such as the domain knowledge, but it would help to study how to identify the generation and externalization of the knowledge.

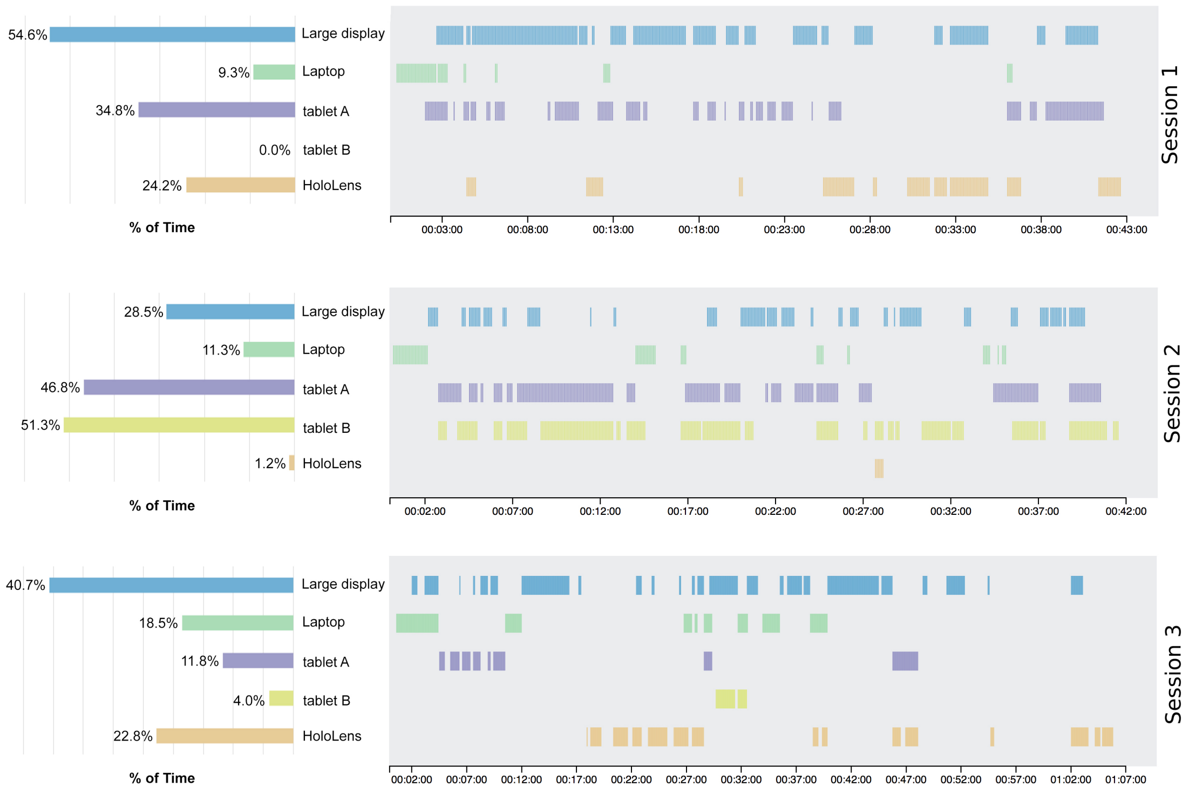
7.3 Final Remarks

This dissertation investigated how the analytical process occurs in the multi-user multi-device environments to provide a theoretical understanding of collaborative exploratory visual data analysis and better inform the design of visualization tools. The results of this research argued that providing guidance mechanisms can fundamentally improve the outcome of the visual analytic process. However, this research contributes a little portion of knowledge to fully support exploratory visual analysis and much work has to be done to apply this research towards the development of visualization tools for multi-device environments.

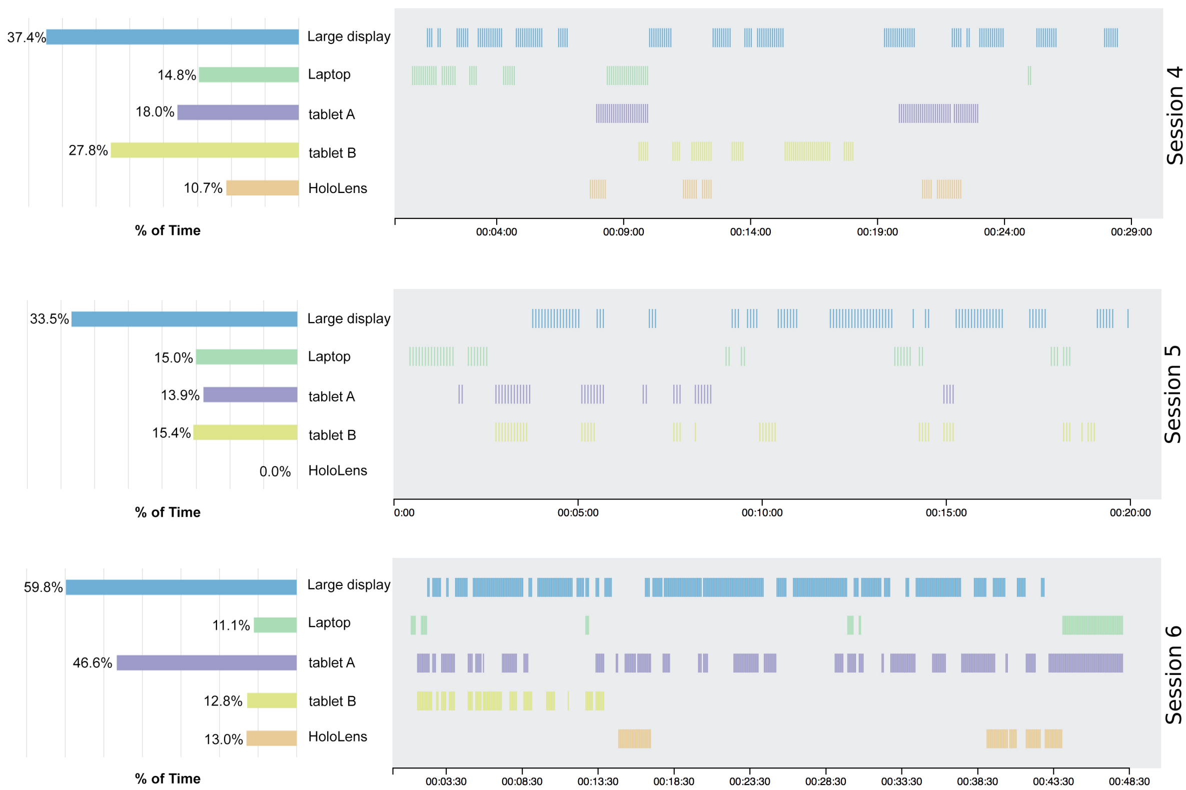
APPENDICES

Appendix A

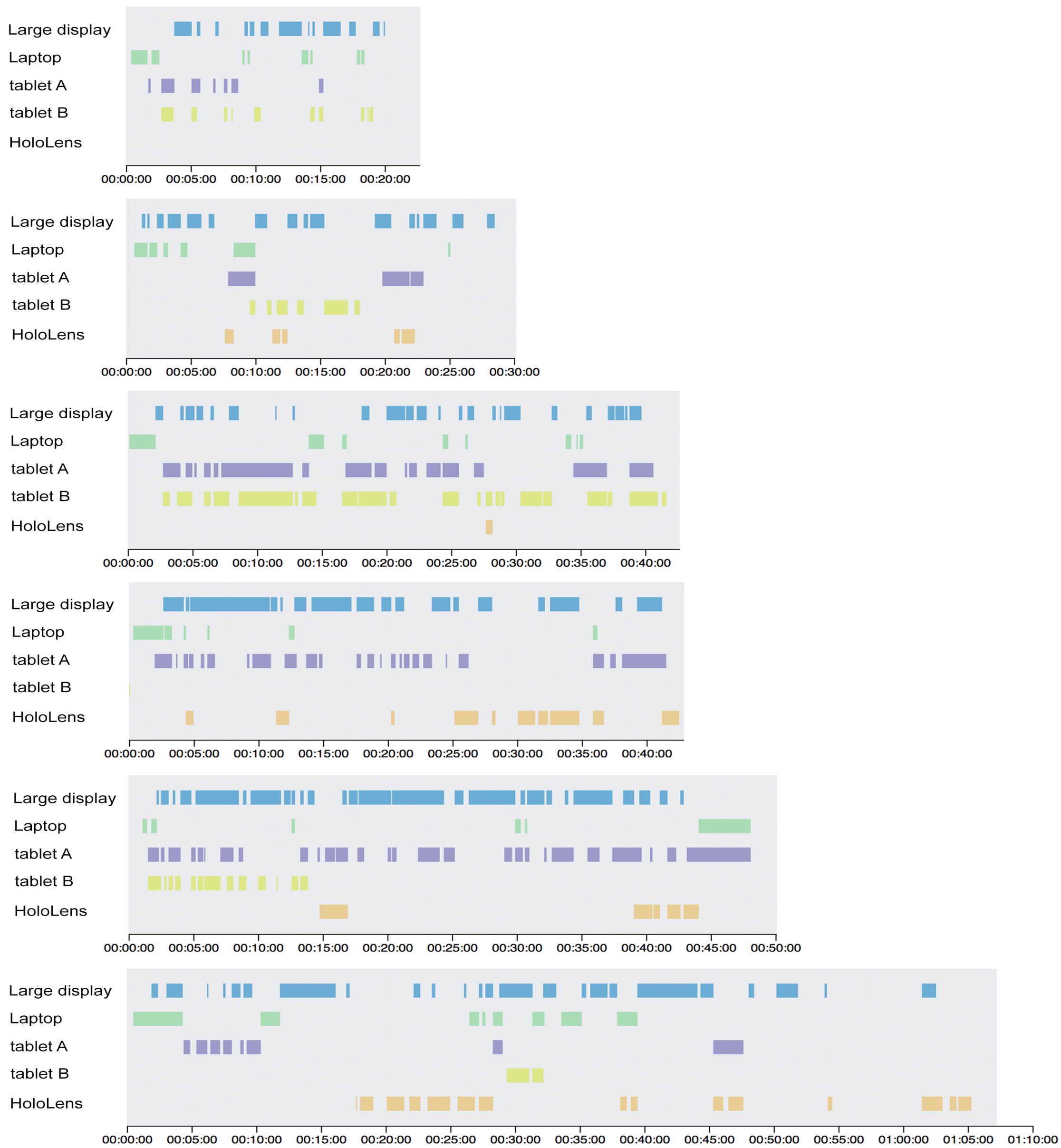
Timeline visualizations of tools usage for each session in the user study, along with their statistics of use.



Appendix A (Continued)



Appendix A (Continued)



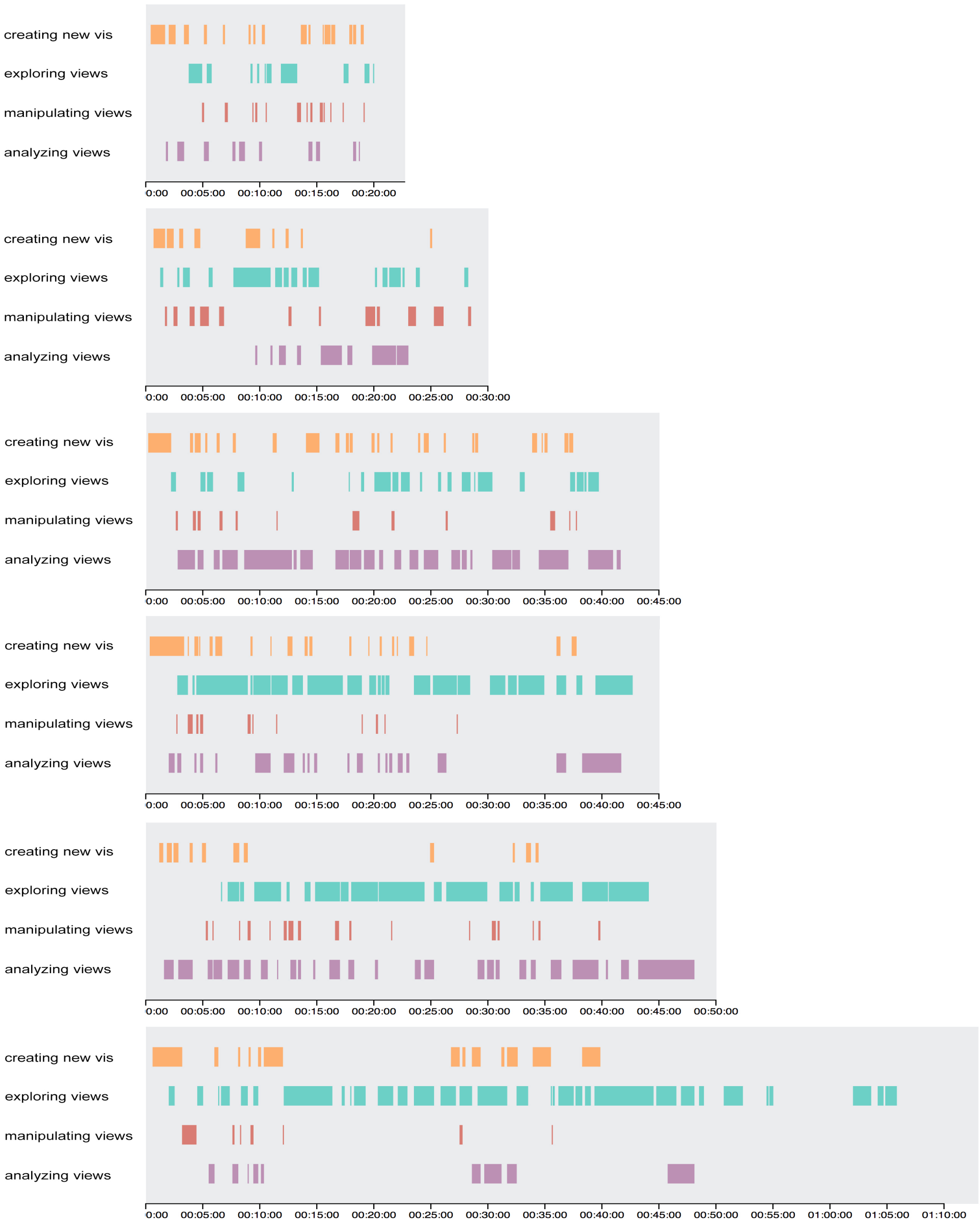
Visualizations in absolute timelines for comparison, ordered from shortest to longest session.

Appendix A (Continued)

Timeline visualizations of performed activities for each session in the user study, along with their statistics.

Appendix A (Continued)

Appendix A (Continued)

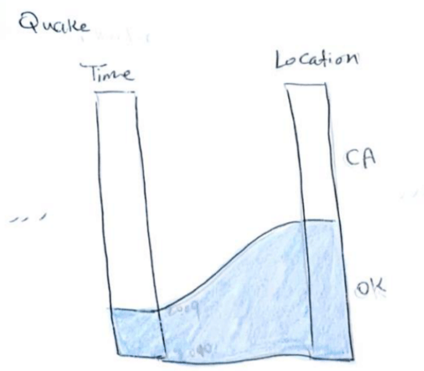


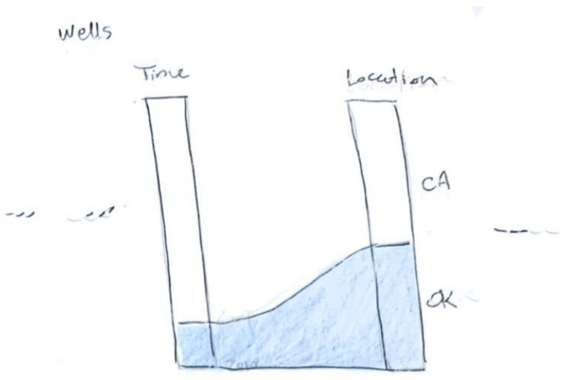
Visualizations in absolute timelines for comparison, ordered from shortest to longest session.

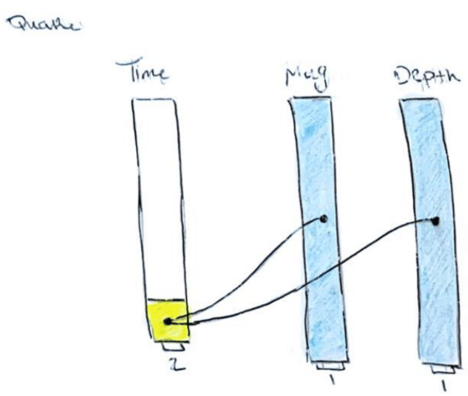
Appendix B

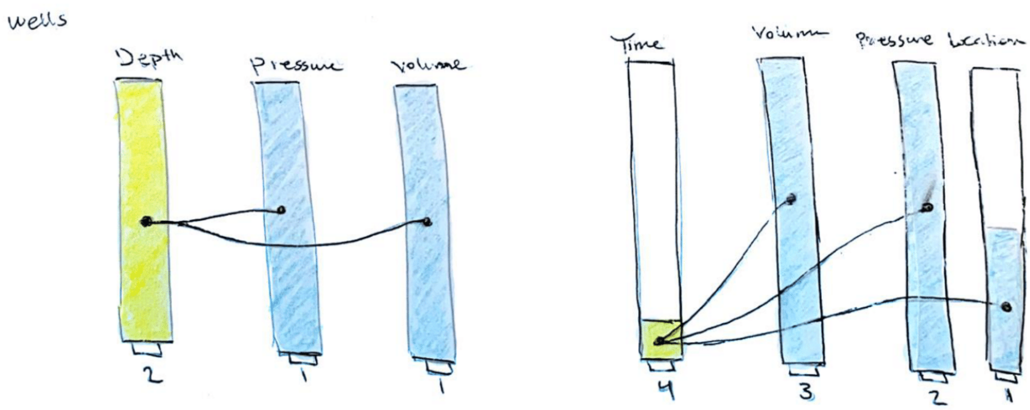
Paper-based sketches of the search space of a session from study 1

Appendix B (Continued)

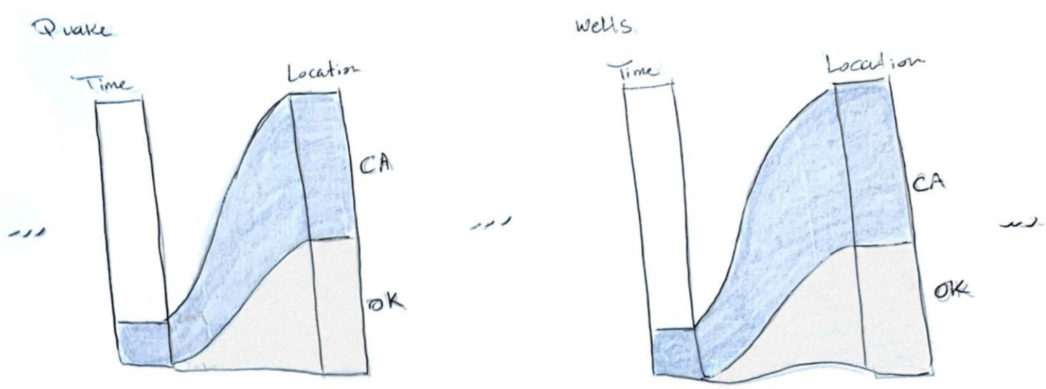


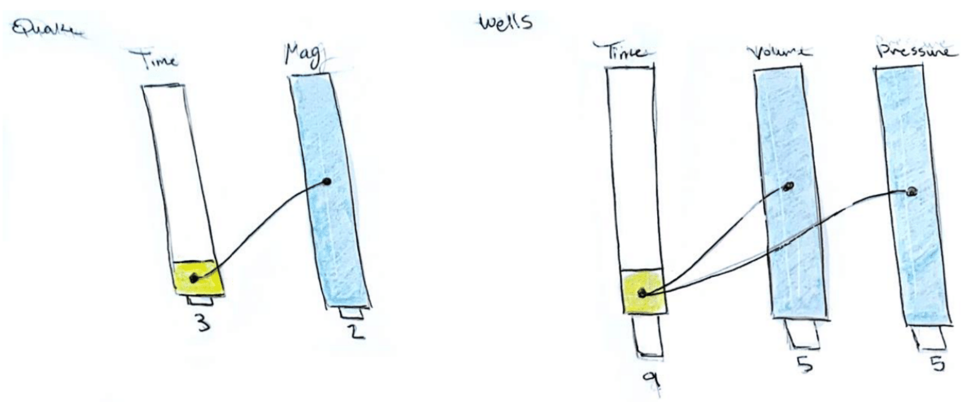




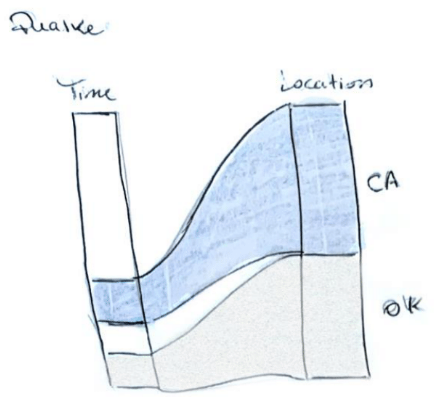


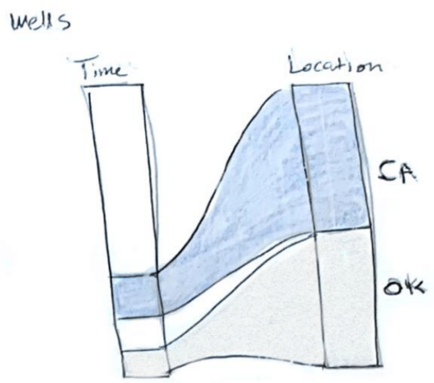
Appendix B (Continued)

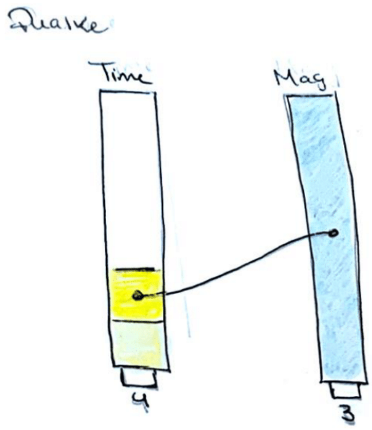


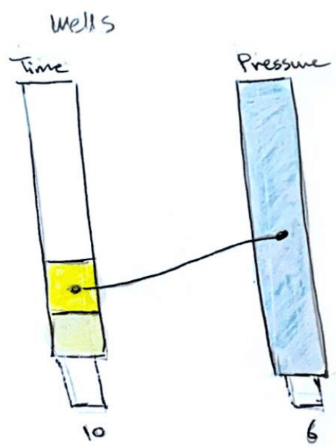


Appendix B (Continued)









Appendix B (Continued)

Paper-based Interview

Appendix B (Continued)

First Name:

Major/degree:

List any taken Visualization/UI related courses:

Introduction

You have been asked to answer the below questions because you participated in the data analysis user study last year. Suppose you do the same user study in the same environment to analyze the earthquake dataset by collaboratively creating and visualizing analysis charts about the data using the big display and multiple portable devices.

This time, suppose that you are provided with the below simple visualization of the investigated attributes and their data space coverage. It visualizes each attribute with regard to how much of that attribute has been investigated. It also communicate information about the group and individual investigations.

The goal of this paper-based interview is to get your feedback to help us make decisions about the design choices of a supportive "visualization of investigated attributes" during an exploratory visual data analysis.

Read the scenarios below and answer the followed questions about the design choices.

Scenario 1: Visualization of the attributes space

You and your two colleagues started the analysis and created few charts based on some attributes. At a time *t* during the analysis, you want to know what attributes have been investigated (used in charts) so far, what data space (range) of these attributes were covered, and how frequent (in how many charts) each attribute has been investigated.

The visualization in Fig. 1 shows this information to you. It lists investigated attributes and highlights in blue the covered data space of each one. The numbered bars at the bottom of each attribute shows the number of investigations (the number of charts that this attribute was used in). The visualization gets updated each time you create a new chart.

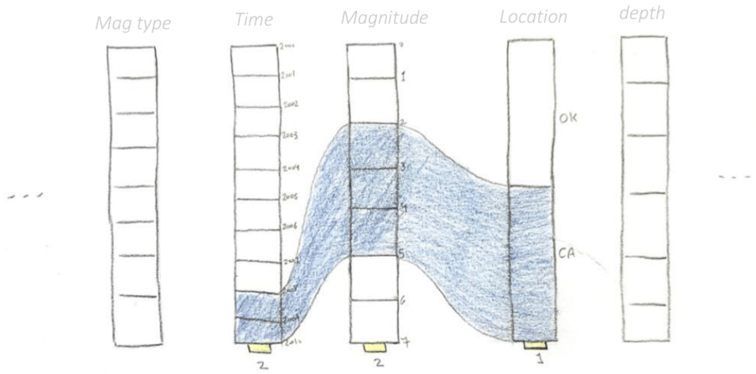


Figure 1. Investigated attributes so far are:
- Time [with data space: 2008-2010] and used in two charts
- Magnitude [data space: 2-5] in two charts
- Location [data space: CA] in one chart

Appendix B (Continued)

Q1: What do you think of using "Parallel Set" as the visual representation of this information? Does it show the information that you want to see (investigated attributes, their coverage, and their frequency of investigation)?

Q2: Do you want this visualization to list all the attributes in the dataset or only the investigated one?

Q3: Where do you want this visualization to be deployed for your team? On the wall display? On the personal tablets? On both? And why?

Q4: Yellow bars that show the frequency of an attribute investigation: What do you think of this design choice? How you would improve it to show the attribute investigation?

Appendix B (Continued)

<u>Scenario 2: Communicating the group and individual investigations</u>

The visualization in Fig. 1 shows the overall investigation by your team. Giving that each of your team members can do some individual investigation using portable devices, you want to know what attributes and data space an individual is working on. The individual tabs shown in Fig. 2 communicate this information to you. Clicking on an individual tab, shows individual investigation: attributes and data space investigated so far by this user. Highlighted dimensions indicate the current attributes under work.

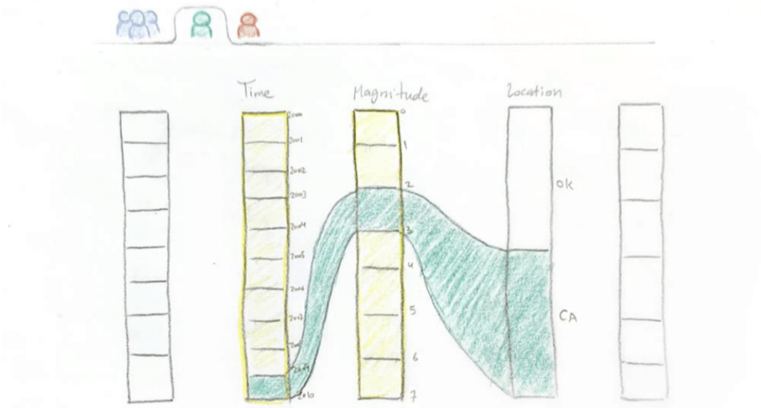


Figure 2. The green user's tab shows the data space that this user is investigating on his device

Q5: Do you want the visualization of individuals investigation to be at a separate tab or to be combined with the group visualization?

Q6: What do you think of this design choice? How you would improve it or simplify it?

Q7: Any other suggestions/questions you want to add?

Appendix C

Database record examples of earthquakes and wells data sets


```
_id: ObjectId("5f52a8cde275392ae77ec150")
EventID: 2198951
Time: 2006-03-05 07:27:41.000
Latitude: 32.4
Longitude: -115.108
Depth: 5
Author: "ECX"
Catalog: "ISC"
Contributor: "ISC"
ContributorID: "7827598"
MagType: "ML"
Magnitude: 3.9
Felt: 2
MagAuthor: "ECX"
Location: "California Area"
__v: 0
```


Figure 50: The earthquake dataset was provided courtesy of http://service.iris.edu/

Appendix C (Continued)


```
_id: ObjectId("6011cb2b4901a9ab5f62c14e")
API_number: 21020
Latitude: 33.9412946
Longitude: -97.7747261
Depth: 891
Well_Type: "Gas"
Well_Status: "Unknown"
Location: "Oklahoma Area"
⌄ Events: Array
  ⌄ 0: Object
      Year: 2009
      _id: ObjectId("6011cb2b4901a9ab5f62c14f")
    ⌄ Months: Array
      ⌄ 0: Object
          Month: 1
          Pressure: 0
          Volume: 12400
          _id: ObjectId("6011cb2b4901a9ab5f62c15b")
      › 1: Object
      › 2: Object
      › 3: Object
      › 4: Object
      › 5: Object
      › 6: Object
      › 7: Object
      › 8: Object
      › 9: Object
      › 10: Object
      › 11: Object
  ⌄ 1: Object
      Year: 2010
      _id: ObjectId("6011cb2b4901a9ab5f62c15c")
    ⌄ Months: Array
      ⌄ 0: Object
          Month: 1
          Pressure: 120
          Volume: 8500
          _id: ObjectId("6011cb2b4901a9ab5f62c168")
      › 1: Object
      › 2: Object
      › 3: Object
      › 4: Object
      › 5: Object
      › 6: Object
      › 7: Object
      › 8: Object
      › 9: Object
      › 10: Object
      › 11: Object
__v: 0
```

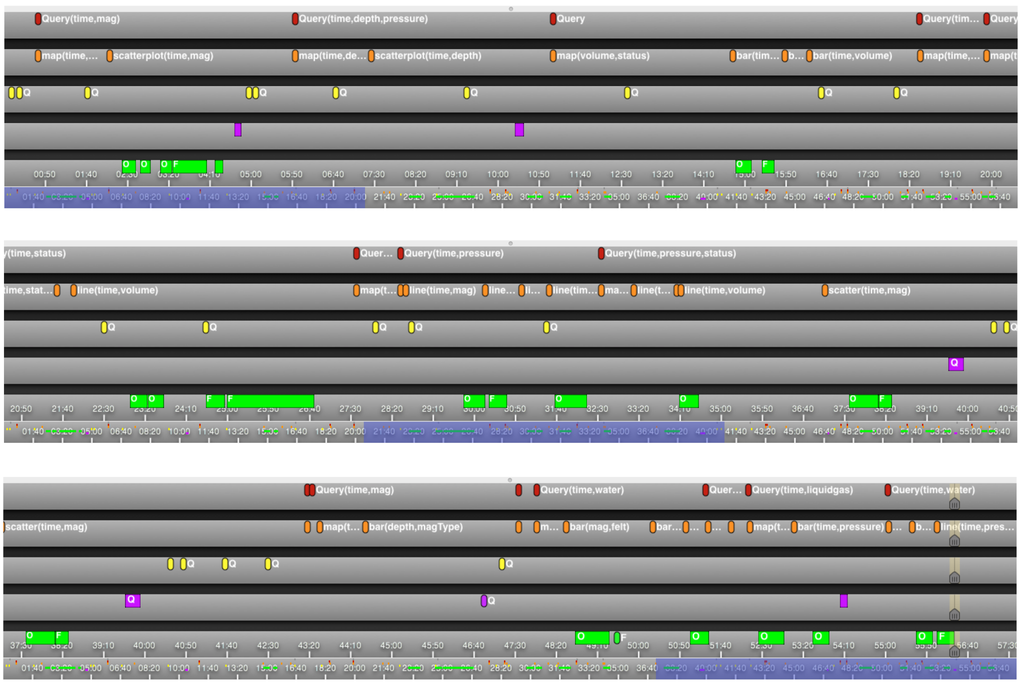

Figure 51: The Wells injection dataset was provided courtesy of http://www.occeweb.com/.

Appendix D

Timelines of video coding for Full and Baseline sessions

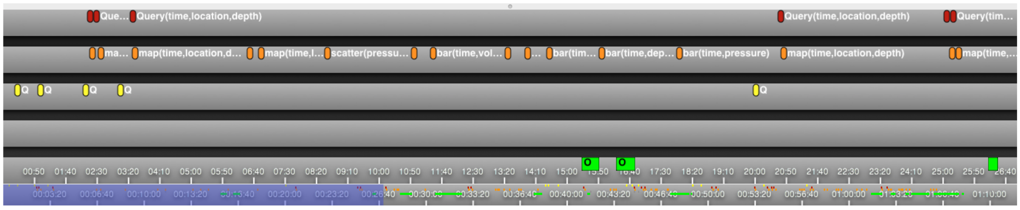
The first set of Full version sessions:

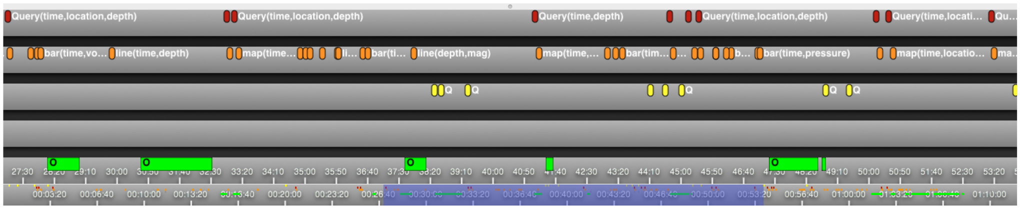
Session 1

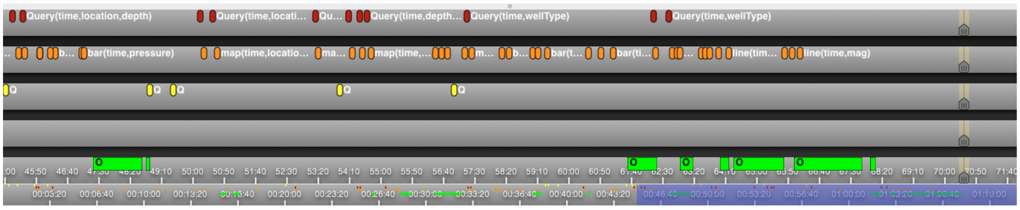


Appendix D (Continued)

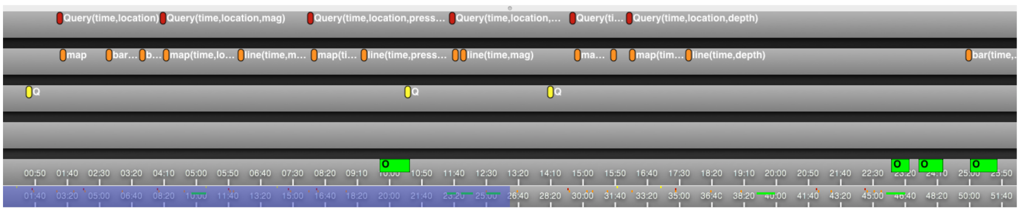
Session 2

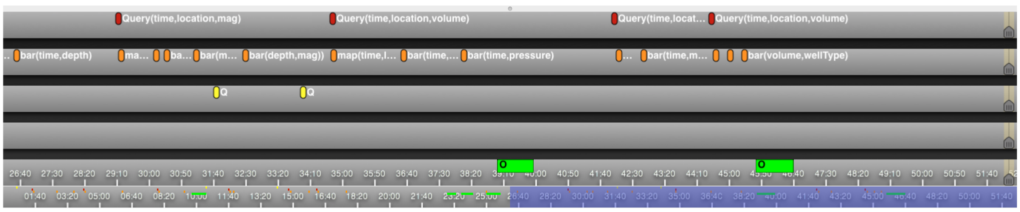






Session 3





Appendix D (Continued)

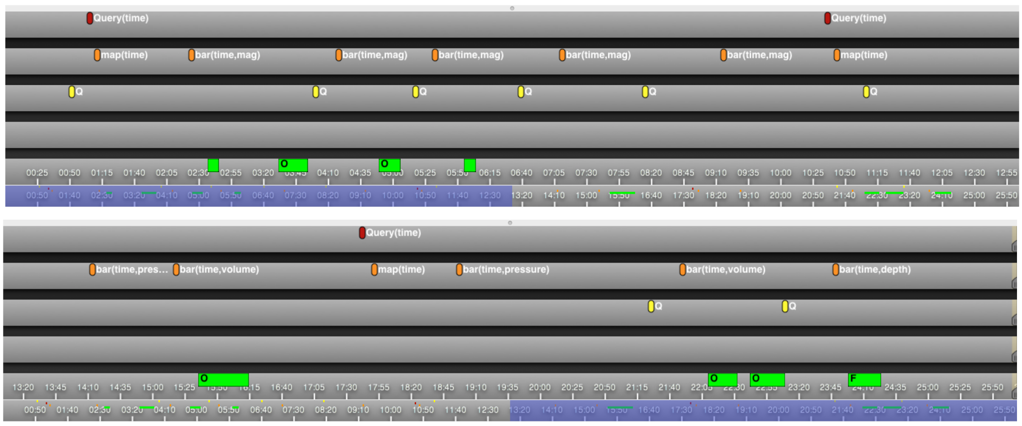
Session 4

Session 5

Appendix D (Continued)

The second set of Baseline version sessions:

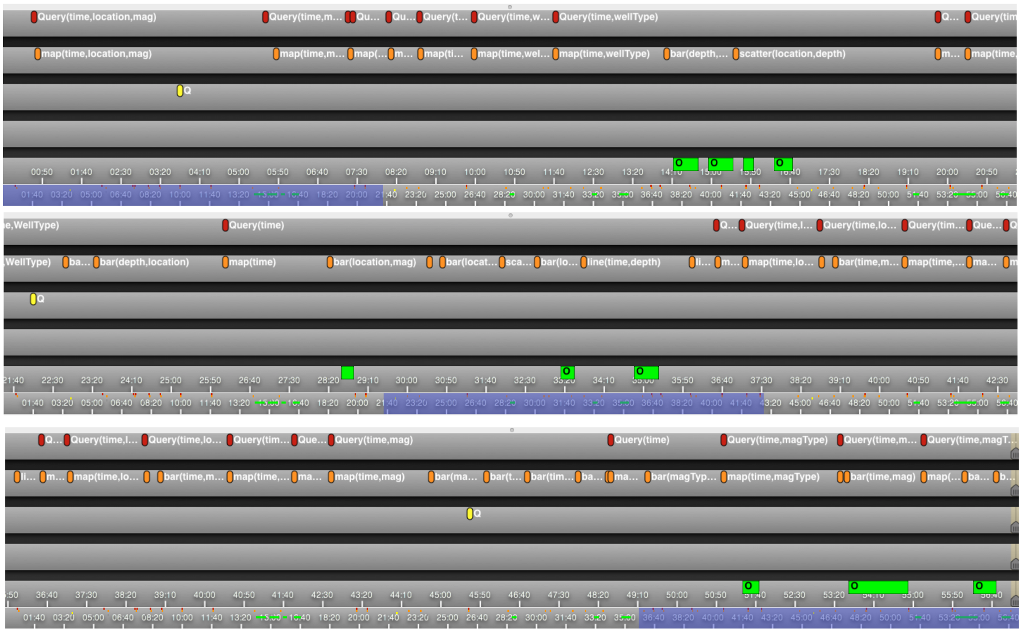
Session 1



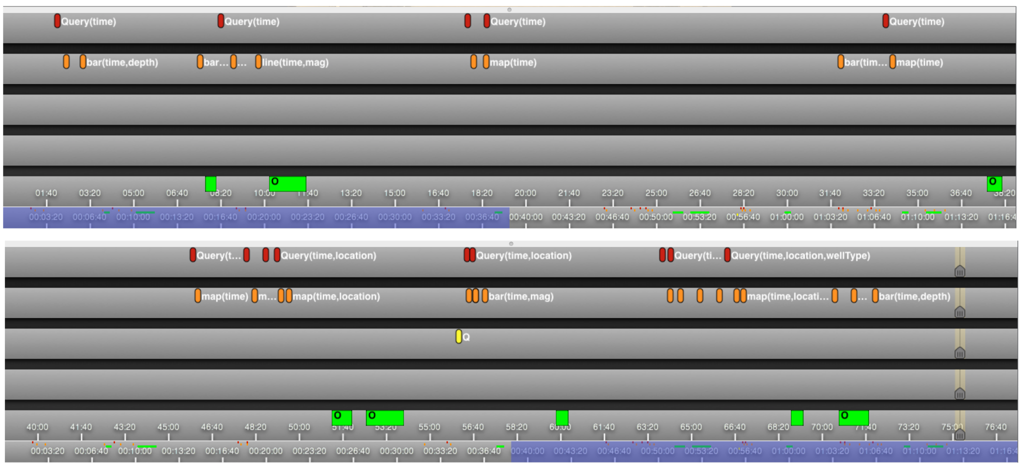
Session 2

Appendix D (Continued)

Session 3

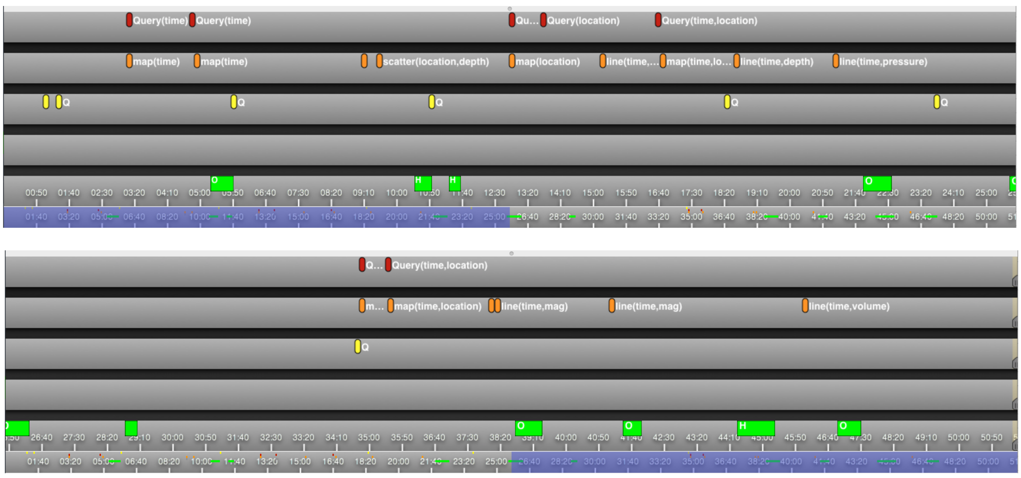


Session 4



Appendix D (Continued)

Session 5



Appendix E

Statistical Analysis of study 2

T-Test

```
[No._of_generated_views]
```

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
No._of_views	Full	5	43.6000	18.14663	8.11542
	Baseline	5	20.4000	13.06905	5.84466

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
		F	Sig.	t	df	Significance One-Sided p	Significance Two-Sided p
No._of_views	Equal variances assumed	.409	.540	2.320	8	.024	.049
	Equal variances not assumed			2.320	7.270	.026	.052

		t-test for Equality of Means			
		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
No._of_views	Equal variances assumed	23.20000	10.00100	.13765	46.26235
	Equal variances not assumed	23.20000	10.00100	-.27187	46.67187

Appendix E (Continued)

T-Test

[No._of_generated_views_per_minute]

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
No._of_views_per_min	Full	5	.7880	.19867	.08885
	Baseline	5	.4360	.17813	.07966

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
		F	Sig.	t	df	Significance One-Sided p	Significance Two-Sided p
No._of_views_per_min	Equal variances assumed	.106	.753	2.950	8	.009	.018
	Equal variances not assumed			2.950	7.907	.009	.019

Independent Samples Test

		t-test for Equality of Means			
		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
No._of_views_per_min	Equal variances assumed	.35200	.11933	.07682	.62718
	Equal variances not assumed	.35200	.11933	.07625	.62775

Appendix E (Continued)

T-Test

```
[No._of_unique_attributes_combination]
```

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
No._of_unique_attributes_combination	Full	5	39.4000	14.41527	6.44670
	Baseline	5	18.8000	11.51955	5.15170

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
		F	Sig.	t	df	Significance One-Sided p	Significance Two-Sided p
No._of_unique_attributes_combination	Equal variances assumed	.139	.719	2.496	8	.019	.037
	Equal variances not assumed			2.496	7.629	.019	.039

Independent Samples Test

		t-test for Equality of Means			
		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
No._of_unique_attributes_combination	Equal variances assumed	20.60000	8.25227	1.57023	39.62977
	Equal variances not assumed	20.60000	8.25227	1.40788	39.79212

Appendix E (Continued)

T-Test

[Search_tree_aspect_ratio]

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
Aspect_ratio	Full	5	.8000	.07106	.03178
	Baseline	5	.6780	.11389	.05093

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
						Significance	
		F	Sig.	t	df	One-Sided p	Two-Sided p
Aspect_ratio	Equal variances assumed	.575	.470	2.032	8	.038	.077
	Equal variances not assumed			2.032	6.705	.042	.083

Independent Samples Test

		t-test for Equality of Means			
		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
				Lower	Upper
Aspect_ratio	Equal variances assumed	.12200	.06003	-.01644	.26044
	Equal variances not assumed	.12200	.06003	-.02123	.26523

Appendix E (Continued)

T-Test

[No._of_questions]

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
No._of_questions	Full	5	12.8000	7.75887	3.46987
	Baseline	5	4.8000	2.58844	1.15758

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
						Significance	
		F	Sig.	t	df	One-Sided p	Two-Sided p
No._of_questions	Equal variances assumed	3.307	.107	2.187	8	.030	.060
	Equal variances not assumed			2.187	4.879	.041	.082

Independent Samples Test

		t-test for Equality of Means			
				95% Confidence Interval of the Difference	
		Mean Difference	Std. Error Difference	Lower	Upper
No._of_questions	Equal variances assumed	8.00000	3.65787	-.43506	16.43506
	Equal variances not assumed	8.00000	3.65787	-1.47314	17.47314

Appendix E (Continued)

T-Test

[No._of_observations]

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
No._of_observations	Full	5	12.8000	7.04982	3.15278
	Baseline	5	8.0000	2.44949	1.09545

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
						Significance	
		F	Sig.	t	df	One-Sided p	Two-Sided p
No._of_observations	Equal variances assumed	4.261	.073	1.438	8	.094	.188
	Equal variances not assumed			1.438	4.952	.105	.210

Independent Samples Test

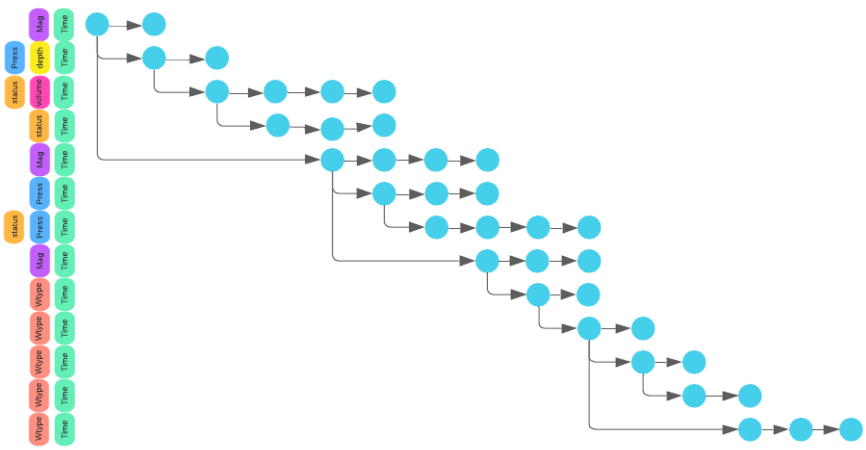
		t-test for Equality of Means			
				95% Confidence Interval of the Difference	
		Mean Difference	Std. Error Difference	Lower	Upper
No._of_observations	Equal variances assumed	4.80000	3.33766	-2.89667	12.49667
	Equal variances not assumed	4.80000	3.33766	-3.80485	13.40485

Appendix F

Search tree graphs of the analysis states from study 2

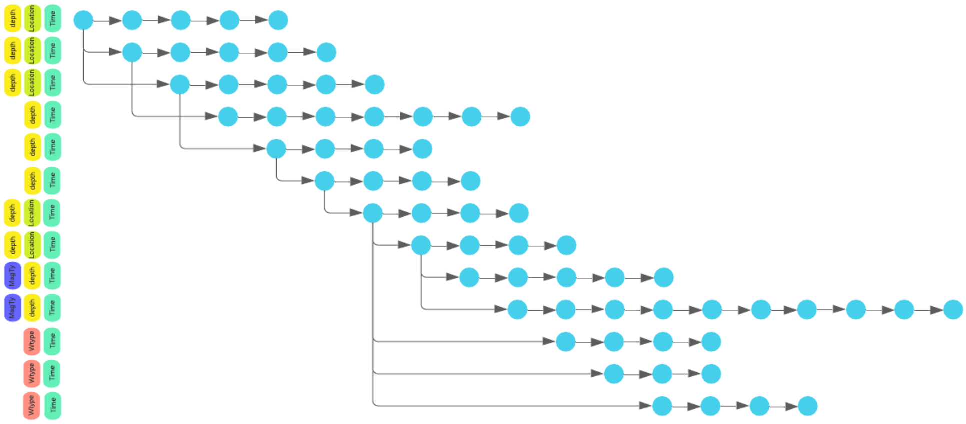
Full version set:

Session 1 - Full

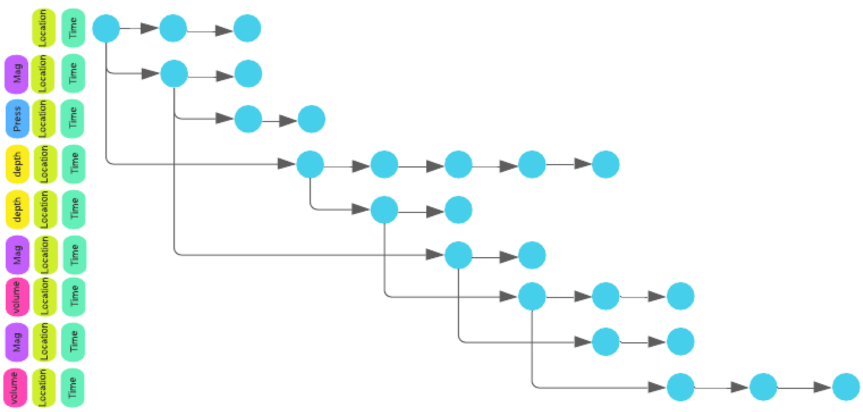


Appendix F (Continued)

Session 2 - Full

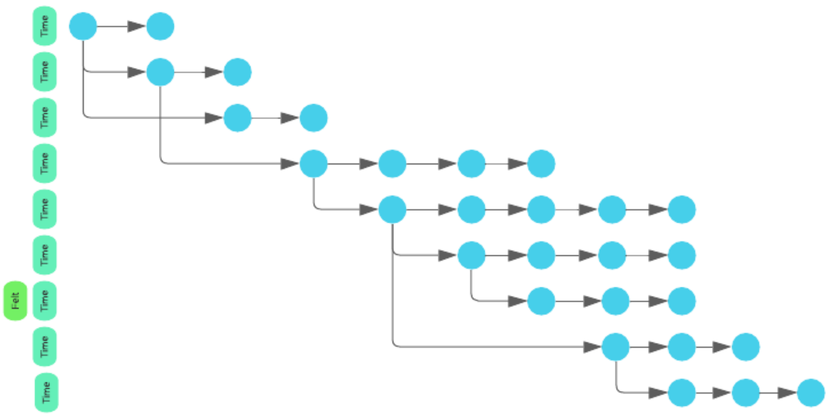


Session 3 - Full

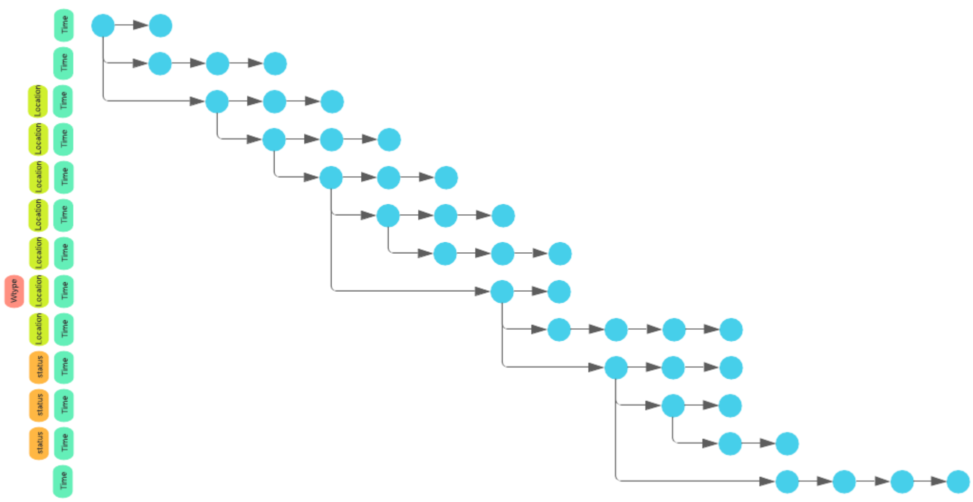


Appendix F (Continued)

Session 4 - Full

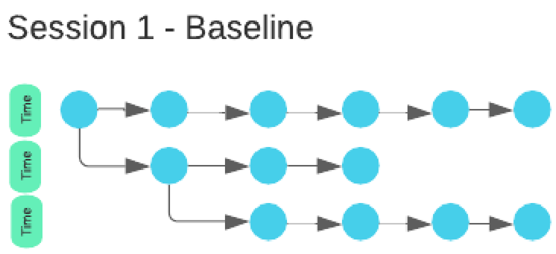


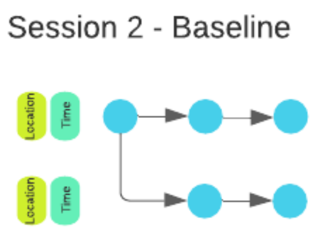
Session 5 - Full



Appendix F (Continued)

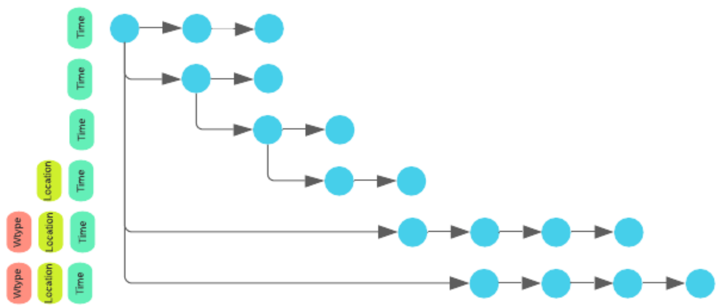
Baseline version set:



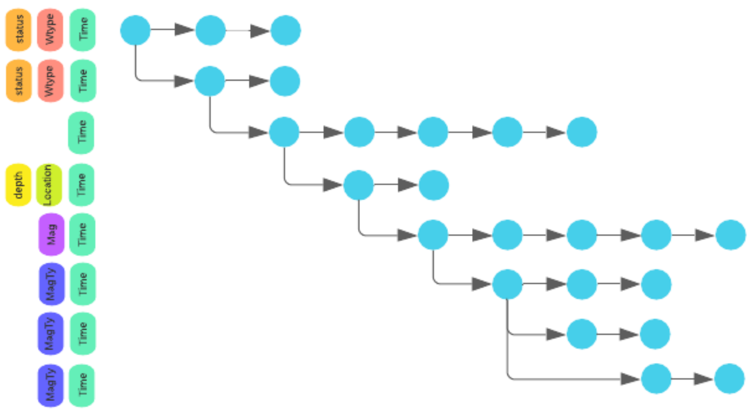


Appendix F (Continued)

Session 3 - Baseline

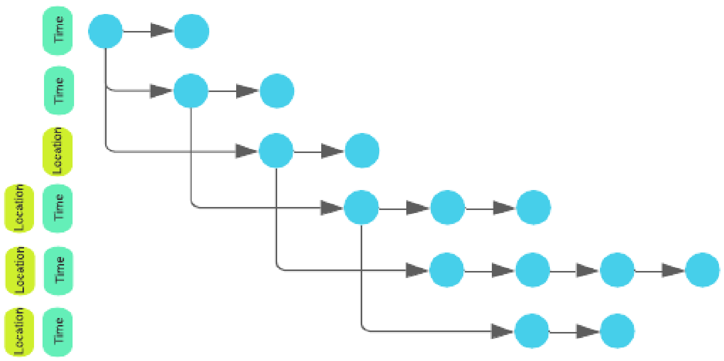


Session 4 - Baseline



Appendix F (Continued)

Session 5 - Baseline



Appendix G

Institutional Review Board (IRP) approval letter from the University of Illinois at Chicago for the user study.

Exemption Granted

April 8, 2019

Abeer Alsaiari
Computer Science

RE: **Protocol # 2019-0365**
"Understanding Collaborative Visual Data Analysis Activities In Multi-Device Environments"

Dear Abeer Alsaiari:

Your application was reviewed on **April 8, 2019** and it was determined that your research 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 46.104(d)]. You may now begin your research.

Exemption Granted Date: April 8, 2019
Funding Source/Sponsor: None

The specific exemption categories under 45 CFR 46.104(d) are: 2 and 3

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 remember to:

→ Use your research protocol number (2019-0365) on any documents or correspondence with the IRB concerning your research protocol.
→ Review and comply with the policies of the UIC Human Subjects Protection Program (HSPP) and the guidance ***Investigator Responsibilities***.

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 via OPRS Live.

Sincerely,
Charles W. Hoehne, B.S., C.I.P.
Assistant Director, IRB #7
Office for the Protection of Research Subjects

cc: Robert Sloan
Andrew E. Johnson

Page 1 of 1

UNIVERSITY OF ILLINOIS AT CHICAGO
Office for the Protection of Research Subjects
201 AOB (MC 672)
1737 West Polk Street
Chicago, Illinois 60612
Phone (312) 996-1711

Appendix H

Permission to use previously published work.

IEEE COPYRIGHT AND CONSENT FORM

To ensure uniformity of treatment among all contributors, other forms may not be substituted for this form, nor may any wording of the form be changed. This form is intended for original material submitted to the IEEE and must accompany any such material in order to be published by the IEEE. Please read the form carefully and keep a copy for your files.

PolyVis: Cross-Device Framework for Collaborative Visual Data Analysis
ABEER ALSAIARI, Andrew Johnson, Arthur Nishimoto
2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)

COPYRIGHT TRANSFER

The undersigned hereby assigns to The Institute of Electrical and Electronics Engineers, Incorporated (the "IEEE") all rights under copyright that may exist in and to: (a) the Work, including any revised or expanded derivative works submitted to the IEEE by the undersigned based on the Work; and (b) any associated written or multimedia components or other enhancements accompanying the Work.

GENERAL TERMS

1. The undersigned represents that he/she has the power and authority to make and execute this form.
2. The undersigned agrees to indemnify and hold harmless the IEEE from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.
3. The undersigned agrees that publication with IEEE is subject to the policies and procedures of the IEEE PSPB Operations Manual.
4. In the event the above work is not accepted and published by the IEEE or is withdrawn by the author(s) before acceptance by the IEEE, the foregoing copyright transfer shall be null and void. In this case, IEEE will retain a copy of the manuscript for internal administrative/record-keeping purposes.
5. For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.
6. The author hereby warrants that the Work and Presentation (collectively, the "Materials") are original and that he/she is the author of the Materials. To the extent the Materials incorporate text passages, figures, data or other material from the works of others, the author has obtained any necessary permissions. Where necessary, the author has obtained all third party permissions and consents to grant the license above and has provided copies of such permissions and consents to IEEE

You have indicated that you DO wish to have video/audio recordings made of your conference presentation under terms and conditions set forth in "Consent and Release."

CONSENT AND RELEASE

1. In the event the author makes a presentation based upon the Work at a conference hosted or sponsored in whole or in part by the IEEE, the author, in consideration for his/her participation in the conference, hereby grants the IEEE the unlimited, worldwide, irrevocable permission to use, distribute, publish, license, exhibit, record, digitize, broadcast, reproduce and archive, in any format or medium, whether now known or hereafter developed: (a) his/her presentation and comments at the conference; (b) any written materials or multimedia files used in connection with his/her presentation; and (c) any recorded interviews of him/her (collectively, the "Presentation"). The permission granted includes the transcription and reproduction of the Presentation for inclusion in products sold or distributed by IEEE and live or recorded broadcast of the Presentation during or after the conference.
2. In connection with the permission granted in Section 1, the author hereby grants IEEE the unlimited, worldwide, irrevocable right to use his/her name, picture, likeness, voice and biographical information as part of the advertisement, distribution and sale of products incorporating the Work or Presentation, and releases IEEE from any claim based on right of privacy or publicity.

Appendix H (Continued)

BY TYPING IN YOUR FULL NAME BELOW AND CLICKING THE SUBMIT BUTTON, YOU CERTIFY THAT SUCH ACTION CONSTITUTES YOUR ELECTRONIC SIGNATURE TO THIS FORM IN ACCORDANCE WITH UNITED STATES LAW, WHICH AUTHORIZES ELECTRONIC SIGNATURE BY AUTHENTICATED REQUEST FROM A USER OVER THE INTERNET AS A VALID SUBSTITUTE FOR A WRITTEN SIGNATURE.

ABEER ALSAIARI	17-07-2019
Signature	**Date (dd-mm-yyyy)**

Information for Authors

AUTHOR RESPONSIBILITIES

The IEEE distributes its technical publications throughout the world and wants to ensure that the material submitted to its publications is properly available to the readership of those publications. Authors must ensure that their Work meets the requirements as stated in section 8.2.1 of the IEEE PSPB Operations Manual, including provisions covering originality, authorship, author responsibilities and author misconduct. More information on IEEE's publishing policies may be found at http://www.ieee.org/publications_standards/publications/rights/authorrightsresponsibilities.html Authors are advised especially of IEEE PSPB Operations Manual section 8.2.1.B12: "It is the responsibility of the authors, not the IEEE, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it." Authors are also advised of IEEE PSPB Operations Manual section 8.1.1B: "Statements and opinions given in work published by the IEEE are the expression of the authors."

RETAINED RIGHTS/TERMS AND CONDITIONS

- Authors/employers retain all proprietary rights in any process, procedure, or article of manufacture described in the Work.
- Authors/employers may reproduce or authorize others to reproduce the Work, material extracted verbatim from the Work, or derivative works for the author's personal use or for company use, provided that the source and the IEEE copyright notice are indicated, the copies are not used in any way that implies IEEE endorsement of a product or service of any employer, and the copies themselves are not offered for sale.
- Although authors are permitted to re-use all or portions of the Work in other works, this does not include granting third-party requests for reprinting, republishing, or other types of re-use.The IEEE Intellectual Property Rights office must handle all such third-party requests.
- Authors whose work was performed under a grant from a government funding agency are free to fulfill any deposit mandates from that funding agency.

AUTHOR ONLINE USE

- **Personal Servers**. Authors and/or their employers shall have the right to post the accepted version of IEEE-copyrighted articles on their own personal servers or the servers of their institutions or employers without permission from IEEE, provided that the posted version includes a prominently displayed IEEE copyright notice and, when published, a full citation to the original IEEE publication, including a link to the article abstract in IEEE Xplore. Authors shall not post the final, published versions of their papers.
- **Classroom or Internal Training Use.** An author is expressly permitted to post any portion of the accepted version of his/her own IEEE-copyrighted articles on the author's personal web site or the servers of the author's institution or company in connection with the author's teaching, training, or work responsibilities, provided that the appropriate copyright, credit, and reuse notices appear prominently with the posted material. Examples of permitted uses are lecture materials, course packs, e-reserves, conference presentations, or in-house training courses.
- **Electronic Preprints.** Before submitting an article to an IEEE publication, authors frequently post their manuscripts to their own web site, their employer's site, or to another server that invites constructive comment from colleagues. Upon submission of an article to IEEE, an author is required to transfer copyright in the article to IEEE, and the author must update any previously posted version of the article with a prominently displayed IEEE copyright notice. Upon publication of an article by the IEEE, the author must replace any previously posted electronic versions of the article with either (1) the full citation to the

Appendix H (Continued)

IEEE work with a Digital Object Identifier (DOI) or link to the article abstract in IEEE Xplore, or (2) the accepted version only (not the IEEE-published version), including the IEEE copyright notice and full citation, with a link to the final, published article in IEEE Xplore.

Questions about the submission of the form or manuscript must be sent to the publication's editor.
Please direct all questions about IEEE copyright policy to:
IEEE Intellectual Property Rights Office, copyrights@ieee.org, +1-732-562-3966

Appendix I

Research restart form and approval.

Application Summary

Competition Details

Competition Title:	UIC COVID-19 Human Subjects Research Review Worksheet
Category:	Questionnaire
Award Cycle:	2020
Submission Deadline:	03/31/2021 11:59 PM

Application Information

Submitted By:	Abeer Alsaiari
Appplication ID:	368
Application Title:	Understanding Collaborative Visual Data Analysis In Multi-Device Environments
Date Submitted:	09/14/2020 4:55 PM

Personal Details

Optional: Please explain

Are research procedures conducted during:	Dedicated research visits

Optional: Please specify

How many research subjects are expected to participate per day or per week?:	3 per session

Optional: Please explain

What ages of participants are included?:	19-60 years old
Does this research protocol potentially include participation of subjects with any of the following risk factors?:	None of the above
Are any of the following risks a specific requirement for subject participation in this research protocol?:	None of the above

What procedures does your protocol involve? Please describe your protocol study procedures, e.g., surveys, focus groups, physical exams, blood tests, other laboratory tests, imaging studies.

Appendix I (Continued)

This protocol involves an empirical study that will be conducted at the Electronic Visualization Laboratory (building ERF, room 2068). All safety measurements will be taken according to the campus/college/laboratory requirements. This includes:

- Maintaining 6 feet distancing
- Wearing a mask
- Getting approval by Maxine Brown to enter the EVL lab
- Submitting temperature self-reporting form one hour prior to the study start
- Scheduling the study on the EVL calendar (with no. of people)
- Sanitizing when entering/exiting the study room
- Disinfect the equipment before the study
- Submitting EVL exit form after completing the study

The study will involve three participants at a time and is expected to take about 70 minutes to complete. It will take place in a room equipped with a large display, laptop, 2 tablets, and an augmented reality HMD. The room is large enough to maintain 6' distancing. Subjects will be asked to use the devices to complete a visual analysis task. Prior to the task start, subjects will divide roles among them to use each of the devices.

The subjects will follow the following procedure during the study:

1. Read and sign media and consent forms
2. The subjects will be then given a short tutorial by the PI on how to use the system.
3. Task 1: each subject will be given a focus question to create one or two visualization
4. Task 2: the three subjects will be given a question to perform a visual data analysis task
5. Fill out a brief survey

Do any of these procedures require research subjects and study personnel to be less than 6 ft apart?: No

Please specify

The study will take place in a room that is large enough to maintain 6' distancing. Subjects and study personnel are not required to be less than 6 ft apart.

Do any of these procedures require subjects to remove their masks? If yes, please specify.: No

Please specify.

Do any of these procedures generate aerosols? If yes, please specify.: No

Please specify.

N/A

Appendix I (Continued)

Approximately how long do research procedures take for each subject (hours and minutes)?:	70 minutes

Optional: Please explain

How many research staff will be present with subjects during these study procedures?:	1

Optional: Please explain

First Name:	Abeer
Last Name:	Alsaiari
Email address:	aalsai3@uic.edu
Phone number:	2244257784
Study Name:	Understanding Collaborative Visual Data Analysis In Multi-Device Environments
Primary Organization(s):	Computer Science
IRB Protocol Number:	2019-0365
Do study activities offer potential direct benefits to participants?:	No
Was this protocol modified for non-face-to-face activities?:	No

Optional: Please explain

Where is the research conducted? Please list building(s) and room(s).:	Building ERF, Room 2068

If yes, please explain

Are all study personnel trained in use of personal protective equipment and sanitization protocols needed to minimize the risk of COVID-19 transmission during interactions with human subjects?:	Yes

Optional: Please explain

Has your research program been approved to restart research in BioRAFT?:	Yes

Appendix I (Continued)

UIC **Office of the Vice Chancellor for Research**

Application Management

Dear Abeer Alsaiari,

Thank you for submitting your **UIC COVID-19 Human Subjects Research Restart Worksheet** for your Understanding Collaborative Visual Data Analysis In Multi-Device Environments human subjects research protocol.

Your protocol **is approved for restart on or after September 28, 2020**, assuming that Illinois and Chicago **stay** in Phase 4 of Restore Illinois and Protecting Chicago COVID-19 recovery plans.

To re-start your IRB-approved protocol, your research program must have an approved BioRAFT registration (https://uic.bioraft.com). This is required for ALL research as it rosters research staff and will be used for training and, if needed, exposure contact tracing. In addition, your site or building must be open for faculty and staff activities on site by the Dean of your College. Appropriate disinfection, signage, and physical and temporal spacing of subjects to promote 6 foot physical distancing, masks, and sanitizer/hand-washing must be in place.

We appreciate your plan for maintaining distance between subjects; because your project has multiple subjects at the same time as part of the research we have copied in Environmental Health staff in the VCAs office as they are responsible for the BioRAFT approval and will assess the capacity of the lab.

You must have a plan for screening research subjects for COVID19 symptoms or exposure, for tracking attendance on site by subjects, and for ensuring appropriate personal protective equipment and sanitizer are available for subjects.

We appreciate your patience as the University works to make sure our restart is done safely. **Please note that specific approval is required for each research protocol. If you have other re-opening applications, you will be notified separately about each one.** Please also note that if the course of the pandemic requires a roll-back in activities that some or all studies may need to cease operations again.

A list of approved studies has been shared with the Vice Chancellor for Administration Environmental Health and Safety Office, with college Deans, and with the OVCR Office for the Protection of Research Subjects (OPRS).

If you have any questions please see the COVID-19 guidance on the OVCR website, or email research@uic.edu.

View Application

CITED LITERATURE

1. 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.

2. Marrinan, T., Aurisano, J., Nishimoto, A., Bharadwaj, K., Mateevitsi, V., Renambot, L., Long, L., Johnson, A., and Leigh, J.: Sage2: A new approach for data intensive collaboration using scalable resolution shared displays. In 10th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, pages 177–186. IEEE, 2014.

3. Card, M.: Readings in information visualization: using vision to think. Morgan Kaufmann, 1999.

4. Fouse, A.: Chronoviz visualization of time-based multimodal data., 2011.

5. Cook, K. A. and Thomas, J. J.: Illuminating the path: The research and development agenda for visual analytics. Technical report, Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2005.

6. Heer, J. and Agrawala, M.: Design considerations for collaborative visual analytics. Information visualization, 7(1):49–62, 2008.

7. Isenberg, P. and Fisher, D.: Collaborative brushing and linking for co-located visual analytics of document collections. In Computer Graphics Forum, volume 28, pages 1031–1038. Wiley Online Library, 2009.

8. Isenberg, P., Fisher, D., Paul, S. A., Morris, M. R., Inkpen, K., and Czerwinski, M.: Co-located collaborative visual analytics around a tabletop display. IEEE Transactions on visualization and Computer Graphics, 18(5):689–702, 2011.

9. Tobiasz, M., Isenberg, P., and Carpendale, S.: Lark: Coordinating co-located collaboration with information visualization. IEEE transactions on visualization and computer graphics, 15(6):1065–1072, 2009.

10. Viegas, F. B., Wattenberg, M., Van Ham, F., Kriss, J., and McKeon, M.: Manyeyes: a site for visualization at internet scale. IEEE transactions on visualization and computer graphics, 13(6):1121–1128, 2007.

11. Badam, S. K., Fisher, E., and Elmqvist, N.: Munin: A peer-to-peer middleware for ubiquitous analytics and visualization spaces. IEEE Transactions on Visualization and Computer Graphics, 21(2):215–228, 2014.

12. Badam, S. K. and Elmqvist, N.: Polychrome: A cross-device framework for collaborative web visualization. In Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces, pages 109–118, 2014.

13. Kister, U., Klamka, K., Tominski, C., and Dachselt, R.: Grasp: Combining spatially-aware mobile devices and a display wall for graph visualization and interaction. In Computer Graphics Forum, volume 36, pages 503–514. Wiley Online Library, 2017.

14. Woźniak, P., Lischke, L., Schmidt, B., Zhao, S., and Fjeld, M.: Thaddeus: a dual device interaction space for exploring information visualisation. In Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational, pages 41–50, 2014.

15. McGrath, W., Bowman, B., McCallum, D., Hincapié-Ramos, J. D., Elmqvist, N., and Irani, P.: Branch-explore-merge: facilitating real-time revision control in collaborative visual exploration. In Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces, pages 235–244, 2012.

16. Horak, T., Badam, S. K., Elmqvist, N., and Dachselt, R.: When david meets goliath: Combining smartwatches with a large vertical display for visual data exploration. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pages 1–13, 2018.

17. Butscher, S., Hubenschmid, S., Müller, J., Fuchs, J., and Reiterer, H.: Clusters, trends, and outliers: How immersive technologies can facilitate the collaborative analysis of multidimensional data. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pages 1–12, 2018.

18. Chung, H., North, C., Self, J. Z., Chu, S., and Quek, F.: Visporter: facilitating information sharing for collaborative sensemaking on multiple displays. Personal and Ubiquitous Computing, 18(5):1169–1186, 2014.

19. Tory, M.: User studies in visualization: A reflection on methods. In Handbook of Human Centric Visualization, pages 411–426. Springer, 2014.

20. 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.

21. Wallace, J. R. and Scott, S. D.: Contextual design considerations for co-located, collaborative tables. In 2008 3rd IEEE International Workshop on Horizontal Interactive Human Computer Systems, pages 57–64. IEEE, 2008.

22. Waldner, M., Lex, E., et al.: Design considerations for collaborative information workspaces in multi-display environments. 2009.

23. Isenberg, P., Isenberg, T., Hesselmann, T., Lee, B., Von Zadow, U., and Tang, A.: Data visualization on interactive surfaces: A research agenda. IEEE Computer Graphics and Applications, 33(2):16–24, 2013.

24. Chegini, M., Lin, S., Lehmann, D. J., Andrews, K., and Schreck, T.: Interaction concepts for collaborative visual analysis of scatterplots on large vertically-mounted high-resolution multi-touch displays. In Forum Media Technology, pages 90–96, 2017.

25. Langner, R., Kister, U., and Dachselt, R.: Multiple coordinated views at large displays for multiple users: Empirical findings on user behavior, movements, and distances. IEEE transactions on visualization and computer graphics, 25(1):608–618, 2018.

26. Jakobsen, M. R., Haile, Y. S., Knudsen, S., and Hornbæk, K.: Information visualization and proxemics: Design opportunities and empirical findings. IEEE transactions on visualization and computer graphics, 19(12):2386–2395, 2013.

27. Langner, R., Horak, T., and Dachselt, R.: V is t iles: Coordinating and combining co-located mobile devices for visual data exploration. IEEE Transactions on Visualization and Computer Graphics, 24(1):626–636, 2017.

28. Cavallo, M., Dholakia, M., Havlena, M., Ocheltree, K., and Podlaseck, M.: Dataspace: A reconfigurable hybrid reality environment for collaborative information analysis. In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pages 145–153. IEEE, 2019.

29. Mahyar, N., Sarvghad, A., and Tory, M.: Roles of notes in co-located collaborative visualization. Collaborative Visualization on Interactive Surfaces-CoVIS'09, page 13, 2009.

30. Mahyar, N., Sarvghad, A., and Tory, M.: A closer look at note taking in the co-located collaborative visual analytics process. In 2010 IEEE Symposium on Visual Analytics Science and Technology, pages 171–178. IEEE, 2010.

31. Isenberg, P., Fisher, D., Morris, M. R., Inkpen, K., and Czerwinski, M.: An exploratory study of co-located collaborative visual analytics around a tabletop display. In 2010 IEEE Symposium on Visual Analytics Science and Technology, pages 179–186. IEEE, 2010.

32. Bradel, L., Endert, A., Koch, K., Andrews, C., and North, C.: Large high resolution displays for co-located collaborative sensemaking: Display usage and territoriality. International Journal of Human-Computer Studies, 71(11):1078–1088, 2013.

33. Scott, S. D., Carpendale, M. S. T., and Inkpen, K. M.: Territoriality in collaborative tabletop workspaces. In Proceedings of the 2004 ACM conference on Computer supported cooperative work, pages 294–303, 2004.

34. Wallace, J. R., Scott, S. D., and MacGregor, C. G.: Collaborative sensemaking on a digital tabletop and personal tablets: prioritization, comparisons, and tableaux. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 3345–3354, 2013.

35. Mahyar, N. and Tory, M.: Supporting communication and coordination in collaborative sensemaking. IEEE transactions on visualization and computer graphics, 20(12):1633–1642, 2014.

36. Mahyar, N., Burke, K. J., Xiang, J., Meng, S., Booth, K. S., Girling, C. L., and Kellett, R. W.: Ud co-spaces: A table-centred multi-display environment for public engagement in urban design charrettes. In Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces, pages 109–118, 2016.

37. Sarvghad, A., Tory, M., and Mahyar, N.: Visualizing dimension coverage to support exploratory analysis. IEEE transactions on visualization and computer graphics, 23(1):21–30, 2016.

38. Blascheck, T., Besançon, L., Bezerianos, A., Lee, B., and Isenberg, P.: Glanceable visualization: Studies of data comparison performance on smartwatches. IEEE transactions on visualization and computer graphics, 25(1):630–640, 2018.

39. Andrews, C., Endert, A., and North, C.: Space to think: large high-resolution displays for sensemaking. In Proceedings of the SIGCHI conference on human factors in computing systems, pages 55–64, 2010.

40. Usher, W., Klacansky, P., Federer, F., Bremer, P.-T., Knoll, A., Yarch, J., Angelucci, A., and Pascucci, V.: A virtual reality visualization tool for neuron tracing. IEEE transactions on visualization and computer graphics, 24(1):994–1003, 2017.

41. Donalek, C., Djorgovski, S. G., Cioc, A., Wang, A., Zhang, J., Lawler, E., Yeh, S., Mahabal, A., Graham, M., Drake, A., et al.: Immersive and collaborative data visualization using virtual reality platforms. In 2014 IEEE International Conference on Big Data (Big Data), pages 609–614. IEEE, 2014.

42. Naveen, K., Venkatram, V., Vaidya, C., Nicholas, S., Allan, S., Charles, Z., Gideon, G., Jason, L., and Andrew, J.: Sage: The scalable adaptive graphics environment. 2004.

43. Bostock, M., Ogievetsky, V., and Heer, J.: D^3 data-driven documents. IEEE transactions on visualization and computer graphics, 17(12):2301–2309, 2011.

44. Belmonte, N. G.: Javascript infovis toolkit. Web (Ed.), 2011.

45. Badam, S. K., Mathisen, A., Rädle, R., Klokmose, C. N., and Elmqvist, N.: Vistrates: A component model for ubiquitous analytics. IEEE transactions on visualization and computer graphics, 25(1):586–596, 2018.

46. Badam, S. K. and Elmqvist, N.: Visfer: Camera-based visual data transfer for cross-device visualization. Information Visualization, 18(1):68–93, 2019.

47. Blumenstein, K., Wagner, M., and Aigner, W.: Cross-platform infovis frameworks for multiple users, screens and devices: Requirements and challenges. In Workshop on Data Exploration for Interactive Surfaces DEXIS 2015, page 7, 2015.

48. Satyanarayan, A., Wongsuphasawat, K., and Heer, J.: Declarative interaction design for data visualization. In Proceedings of the 27th annual ACM symposium on User interface software and technology, pages 669–678, 2014.

49. Satyanarayan, A., Moritz, D., Wongsuphasawat, K., and Heer, J.: Vega-lite: A grammar of interactive graphics. IEEE transactions on visualization and computer graphics, 23(1):341–350, 2016.

50. Wickham, H.: ggplot2: elegant graphics for data analysis. Springer, 2016.

51. Chang, W. and Wickham, H.: ggvis: Interactive grammar of graphics, 2015. URL http://CRAN. R-project. org/package= ggvis. R package version 0.4, 1.

52. Ellis, C. A. and Gibbs, S. J.: Concurrency control in groupware systems. In Proceedings of the 1989 ACM SIGMOD international conference on Management of data, pages 399–407, 1989.

53. Vogt, K., Bradel, L., Andrews, C., North, C., Endert, A., and Hutchings, D.: Co-located collaborative sensemaking on a large high-resolution display with multiple input devices. In IFIP Conference on Human-Computer Interaction, pages 589–604. Springer, 2011.

54. Isenberg, P., Tang, A., and Carpendale, S.: An exploratory study of visual information analysis. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 1217–1226, 2008.

55. Dourish, P.: Where the action is: the foundations of embodied interaction. MIT press, 2004.

56. Neumann, P., Tang, A., and Carpendale, S.: A framework for visual information analysis. 2007.

57. Brehmer, M. and Munzner, T.: A multi-level typology of abstract visualization tasks. IEEE transactions on visualization and computer graphics, 19(12):2376–2385, 2013.

58. Lam, H., Tory, M., and Munzner, T.: Bridging from goals to tasks with design study analysis reports. IEEE transactions on visualization and computer graphics, 24(1):435–445, 2017.

59. Reda, K., Johnson, A. E., Papka, M. E., and Leigh, J.: Modeling and evaluating user behavior in exploratory visual analysis. Information Visualization, 15(4):325–339, 2016.

60. Sarvghad, A. and Tory, M.: Exploiting analysis history to support collaborative data analysis. In Proceedings of the 41st Graphics Interface Conference, pages 123–130. Canadian Information Processing Society, 2015.

61. Wongsuphasawat, K., Moritz, D., Anand, A., Mackinlay, J., Howe, B., and Heer, J.: Voyager: Exploratory analysis via faceted browsing of visualization recommendations. IEEE transactions on visualization and computer graphics, 22(1):649–658, 2015.

62. Wongsuphasawat, K., Qu, Z., Moritz, D., Chang, R., Ouk, F., Anand, A., Mackinlay, J., Howe, B., and Heer, J.: Voyager 2: Augmenting visual analysis with partial view specifications. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pages 2648–2659. ACM, 2017.

63. Battle, L. and Heer, J.: Characterizing exploratory visual analysis: A literature review and evaluation of analytic provenance in tableau. In Computer Graphics Forum, volume 38, pages 145–159. Wiley Online Library, 2019.

64. Thomas, J. J. and Cook, K. A.: A visual analytics agenda. IEEE computer graphics and applications, 26(1):10–13, 2006.

65. Lee, B., Isenberg, P., Riche, N. H., and Carpendale, S.: Beyond mouse and keyboard: Expanding design considerations for information visualization interactions. IEEE Transactions on Visualization and Computer Graphics, 18(12):2689–2698, 2012.

66. Ball, R. and North, C.: Effects of tiled high-resolution display on basic visualization and navigation tasks. In CHI'05 extended abstracts on Human factors in computing systems, pages 1196–1199, 2005.

67. Reda, K., Offord, C., Johnson, A. E., and 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. 2014.

68. Card, S. K. and Mackinlay, J.: The structure of the information visualization design space. In Proceedings of VIZ'97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium, pages 92–99. IEEE, 1997.

69. Gotz, D. and Zhou, M. X.: Characterizing users' visual analytic activity for insight provenance. Information Visualization, 8(1):42–55, 2009.

70. Alsaiari, A., Johnson, A., and Nishimoto, A.: Polyvis: Cross-device framework for collaborative visual data analysis. In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), pages 2870–2876. IEEE, 2019.

71. Kaptelinin, V. and Nardi, B. A.: Acting with technology: Activity theory and interaction design. MIT press, 2006.

72. Winograd, T.: Bringing design to software. ACM, 1996.

73. Brown, J. S., Heath, C., Pea, R., et al.: Perspectives on activity theory. Cambridge university press, 1999.

74. Badam, S. K.: Enabling collaborative visual analysis across heterogeneous devices. 2019.

75. Renambot, L., Marrinan, T., Aurisano, J., Nishimoto, A., Mateevitsi, V., Bharadwaj, K., Long, L., Johnson, A., Brown, M., and Leigh, J.: Sage2: A collaboration portal for scalable resolution displays. Future Generation Computer Systems, 54:296–305, 2016.

76. Mahyar, N., Sarvghad, A., and Tory, M.: Note-taking in co-located collaborative visual analytics: Analysis of an observational study. Information Visualization, 11(3):190–204, 2012.

77. Shneiderman, B.: The eyes have it: A task by data type taxonomy for information visualizations: Proc. In 1996 IEEE Conference on Visual Languages, Viewed, volume 17, page 2000, 1996.

78. Andrews, C., Endert, A., and North, C.: Space to think: large high-resolution displays for sensemaking. In Proceedings of the SIGCHI conference on human factors in computing systems, pages 55–64, 2010.

79. Law, P.-M. and Basole, R. C.: Designing breadth-oriented data exploration for mitigating cognitive biases. In Cognitive Biases in Visualizations, pages 149–159. Springer, 2018.

80. Ceneda, D., Gschwandtner, T., May, T., Miksch, S., Schulz, H.-J., Streit, M., and Tominski, C.: Characterizing guidance in visual analytics. IEEE Transactions on Visualization and Computer Graphics, 23(1):111–120, 2016.

81. Munzner, T.: Visualization Analysis and Design.. Print, 2014.

82. Willett, W., Heer, J., and Agrawala, M.: Scented widgets: Improving navigation cues with embedded visualizations. IEEE Transactions on Visualization and Computer Graphics, 13(6):1129–1136, 2007.

83. Bostock, M.: D3.js - data-driven documents, 2012.

84. Eliyahu, G.: Windowtop, 2019.

85. Lam, H.: A framework of interaction costs in information visualization. IEEE transactions on visualization and computer graphics, 14(6):1149–1156, 2008.

86. Liu, Z. and Heer, J.: The effects of interactive latency on exploratory visual analysis. IEEE transactions on visualization and computer graphics, 20(12):2122–2131, 2014.

87. Ghaoui, C.: Encyclopedia of human computer interaction. IGI Global, 2005.

88. Lee, B., Lee, M., Zhang, P., Tessier, A., and Khan, A.: An empirical study of how socio-spatial formations are influenced by interior elements and displays in an office context. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW):1–26, 2019.

89. Sacha, D., Stoffel, A., Stoffel, F., Kwon, B. C., Ellis, G., and Keim, D. A.: Knowledge generation model for visual analytics. IEEE transactions on visualization and computer graphics, 20(12):1604–1613, 2014.

VITA

NAME	ABEER A. ALSAIARI
EDUCATION	PhD., University of Illinois at Chicago, Chicago, IL, 2021
	MSc., King Saud University, Riyadh, Saudi Arabia, 2011
	BSc., Taibah University, Madina, Saudi Arabia, 2005
EXPERIENCE	Graduate Research Assistant, University of Illinois at Chicago, 2017 - 2021
	Graduate Teaching Assistant, University of Illinois at Chicago, 2020 - 2021
PUBLICATIONS	Alsaiari, Abeer., Aurisano, Jillian., Johnson, Andrew. "Evaluating Strategies of Exploratory Visual Data Analysis in Multi Device Environments." In the Proceedings of *EuroViz 2020*, May 25-29, 2020, Norrkoping, Sweden. https://doi.org/10.2312/evs.20201054.
	Jillian Aurisano, Abhinav Kumar, Abeer Alsaiari, Barbara Di Eugenio, and Andrew Johnson. "Many At Once: Capturing Intentions to Create And Use Many Views At Once In Large Display Environments." In *Eurographics Conference on Visualization (EuroVis) 2020, Volume 39 (2020), Number 3.*
	Alsaiari, Abeer and Johnson, Andrew. "Towards Understanding Collaborative Visual Data Analysis in Multi- Device Environments." In *IEEE Transactions on Visualization and Computer Graphics, 2019.*
	Jillian Aurisano, Abhinav Kumar, Abeer Alsaiari, Barbara Di Eugenio, and Andrew Johnson. "Evaluation of Scalable Interactions over Multiple Views in Large Display Environments." In *IEEE Transactions on Visualization and Computer Graphics, 2019.*
	Alsaiari, Abeer, Nishimoto, Arthur, and Johnson, Andrew. "PolyVis: Cross-Device Framework for Collaborative Visual Data Analysis." In the Proceedings of *2019 IEEE International Conference on Systems,*

Man, and Cybernetics (IEEE SMC 2019), October 6-9, 2019, Bari, Italy. https://doi.org/10.1109/SMC.2019.8914209.

Abeer Alsaiari, Ridhi Rustagi, A'aeshah Alhakamy, Manu Mathew Thomas, Angus G. Forbes. "Image Denoising Using A Generative Adversarial Network." In *IEEE 2nd International Conference on Information and Computer Technologies*, 2019.

Kumar, Abhinav, Barbara Di Eugenio, Jillian Aurisano, Andrew Johnson, Abeer Alsaiari, Nigel Flowers, Alberto Gonzalez, and Jason Leigh. "Towards Multimodal Coreference Resolution for Exploratory Data Visualization Dialogue: Context-Based Annotation and Gesture Identification." In *SEMDIAL 2017 SaarDial*.

Abeer Alsaiari and Aqil Azmi. "A calligraphic based scheme to justify Arabic text improving readability and comprehension." In *Computers in Human Behavior* 39. Elsevier (2014). DOI:doi.org/10.1016/j.chb.2014.07.003.

Abeer Alsaiari and Aqil Azmi. "An Algorithm to Justify Arabic Text." In *Egyptian Computer Science Journal Vol. 37 No. 5.*

Abeer Alsaiari and Aqil Azmi. "Arabic Typography: A Survey." In *International Journal of Electrical Computer Sciences* 9.10 (2010): 1.