TASK PERFORMANCE WITH SPACE-TIME CUBE VISUALIZATIONS: DIFFERENCES BETWEEN HOLOLENS AND DESKTOP USERS

by

Michael A Saenz

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Computer Graphics Technology

West Lafayette, Indiana

December 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Paul Parsons, Chair

Department of Computer Graphics Technology

Dr. Brandon J. Pitts

School of Industrial Engineering

Dr. Austin Toombs

Department of Computer Graphics Technology

Approved by:

Dr. Colin M. Gray

Graduate Program Co-Chair

Prof. Nicoletta Adamo

Graduate Program Co-Chair

De colores

ACKNOWLEDGMENTS

I would like to acknowledge my Advisor, Dr. Paul Parsons. Thank you for your understanding and patience while working with me through this research journey. Furthermore, I would like to acknowledge my committee members Dr. Brandon J. Pitts and Dr. Austin Toombs for their advice and support. Additionally, I would like to acknowledge my family and friends for their words of encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Saenz, Michael, A. MS
Institution: Purdue University
Degree Received: December 2018
Title: Task Performance with Space-Time Cube Visualizations: Differences Between HoloLens and Desktop Users
Committee Chair: Paul Parsons

The researcher's intent in this study was to understand users' performance, specifically in terms of time, error and workload, in different display conditions while manipulating a space-time cube visualization. A convergent mixed-method design was applied to allow the researcher to better understand the research problems. In the study, time, error and perceived workload were investigated to test performance to detect if a display condition had a positive or negative influence on users' abilities to perform a task. The qualitative data explored the differences in users' experiences with the HoloLens and desktop.

CHAPTER 1. INTRODUCTION

This chapter provides an overview of the problem and background of the research. It also explains the scope, significance, foundation. Additionally, it provides an outline of limitations and assumptions of the study.

1.1 Problem Statement

The visualization community has debated the value of three-dimensional (3D) visualization since the early days of data visualization. While 3D visualizations produce an extra dimension compared to two-dimensional (2D) visualizations, they often have visual issues with occlusion and depth perception when displayed on a computer screen. Consequently, 3D visualizations work better than 2D visualizations for tasks requiring overviews and holistic reasoning (Wickens, Merwin, & Lin, 1994) but not for detailed analysis and precise reasoning (Smallman, St. John, Oonk, & Cowen, 2001). While other studies have found mixed results for 3D visualizations, sometimes the mixed results can be traced to the display conditions and not to the 3D visualization (Keehner, Hegarty, Cohen, Khooshabeh, & Montello, 2008). In addition, physical models perform better than 3D visualizations in a computer screen for information retrieval activities. For example, Jansen, Dragicevic, and Fekete found that users perform better in information retrieval tasks like range, order and compare in physical models (2013). Furthermore, physical models allow users to use their physical body to measure distance and are coupled to the environment. However, software 3D visualizations have an advantage over static physical models in that they are normally interactive and support dynamic datasets. Virtual reality (VR) has been used to evaluate 3D visualizations with some success. VR is interactive and support dynamic datasets. However, VR is known to be inherently transportive which decouples the users with the physical environment; when users put on a VR system, they believe they are in a new environment (Turner, Turner, & Burrows, 2013) thus making it difficult to use the coupling effect found in physical models. The decoupling effect may be beneficial for scientific 3D visualizations but not data visualizations (Shaw, Green, Liang, & Sun, 1993). This decoupling may make it difficult for users to evaluate 3D data visualizations because users may lose their spatial presence, social presence, self-presence, body awareness and environmental awareness if designed incorrectly (Büschel et al., 2018). This could create an overall detrimental experience as without a physical reference point users may end up colliding with objects unintentionally. Therefore, 3D data visualizations must be evaluated in an interactive immersive display that does not decouple or dissociate the user from the physical environment.

I, the researcher believes that augmented reality (AR) could be used to evaluate 3D visualizations, and the results of my scientific and informatic visualization projects from the past five years suggest AR head mounted display (HMD) would best suited to the task. As an undergrad, I evaluated a desktop computer, tablet and Epson smart glasses for a scientific visualization (Saenz, Strunk, Maset, Malone, & Seo, 2015). From my observations of the way each device affected users' experience, I concluded that the range of possible interactions strongly depends on the quality of the display. At the time of this study, I decided that the HoloLens, a mixed-reality head-mounted device developed by Microsoft, had the best visual display and range of interactions among AR devices on the market.

Ergo, the researcher proposes using a HoloLens as a testbed to examine how users perform 3D visualization tasks. The justification for using the HoloLens as the testbed is that the HoloLens allows digital objects to be superimposed in the real world, providing a more natural interface for interacting with digital objects with which users are unlikely to lose their body environmental awareness. Furthermore, the HoloLens is one of the most advanced devices available for hands-free augmented reality. For example, while walking around and using gestures to manipulate digital objects, users would be able to use their surroundings as a cognitive aid to assist in their understanding of the objects due to the superimposed nature of the holograms thus not decoupling the user from the real world. Additionally, binocular stereoscopic displays, the technology used in the HoloLens, are known to improve users' ability of spatial judgment (McIntire, Havig, & Geiselman, 2012) and 3D object recognition which may be beneficial when interacting with 3D data visualizations. The researcher believes that these benefits justify the use of the HoloLens as a testbed to evaluate 3D data visualizations.

The researcher evaluated users' interactions with 3D data visualizations using a mixed-method qualitative and quantitative approach. This is the approach used since in previous studies in data visualization has mostly focused on performance evaluations of data visualization using quantitative methods. In the education space Gorard contends that quantitative approaches cannot neglect the qualitative factors that can affect the interpretation of the results (2004).

Moreover, in the Immersive Analytics book the authors argue that “task performance is not the only measure of effectiveness” (pg.13) that other element can play a factor like engagement and recall (2018). Furthermore, a significant benefit of qualitative research is “seeing the big picture,” the circumstances in which a 3D visualization might be supported. The HoloLens is a new technology, and knowing participants' motives, misgivings and feelings can show areas were the 3D data visualizations excels. This can efficiently help the development of new 3D visualizations and inform the design of future applications. Thus, the researcher wants to compare performances of HoloLens users compared to desktop users. The purpose of the research is to see if one performs better in displaying 3D data visualizations statistically and the researcher also wants to compare the experiential differences of using the HoloLens and the desktop. To see if there’s a difference in perceived experience and what are the experiences in general of using these displays.

In addition, it can be argued that visualizations that people want to explore and interact with can be more important than just performance metrics. For example, the users can perform better in all task in HoloLens, but they find it difficult. In this thesis we are trying to understand performance in a holistic way by incorporating users preferences and experiences. This type of knowledge is essential in the initial stages of defining what types of 3D data visualizations may be of worth. Therefore, a mixed-method approach was determined to be the best solution for permitting the quantitative approach common in the data visualization research community while at the same time allowing data to be contextualized qualitatively.

1.2 Scope

To narrow down the scope of the study, the researcher focused on three aspects of performance: time, error and workload. The reason for using these three aspects is that data visualization researchers commonly evaluate only and error when testing software; with this information the researcher was able to see how quickly and successfully users completed each task. Moreover, evaluating workload informed the researcher of the cognitive cost to accomplish a task (Hart, & Staveland, 1988), allowing the researcher to see how taxing an activity was to the user.

The researcher also limited the display environments to a desktop and a HoloLens. A desktop was chosen because 3D visualizations are commonly displayed on computer screens and the majority of people are comfortable interacting with this type of display. As stated in the previous section, there are many reasons to use the HoloLens. Another reason the researcher used the HoloLens rather than another head-mounted AR display device was that the HoloLens, at the time of the study, was the most advanced technology on the market and the researcher had access to one.

This study utilized one type of visualization known as a space-time cube, a 3D data visualization that displays spatiotemporal data using the X and Y dimensions to represent longitude and latitude and the Z dimension to represent time. The rationale behind this was that the data used in a space-time cube is intrinsically three-dimensional, with two spatial dimensions and one temporal dimension. Additionally, it has the natural spatial mappings that may be beneficial in augmented reality. This study focused on evaluating users' interactions with a space-time cube while performing low-level components of analytic activity with metrics of time, error and workload, using a HoloLens and desktop computer as the display environments.

1.3 Significance

This study follows the call of action paper Immersive Analytics, which defines immersive analytics as "an emerging research thrust investigating how new interaction and display technologies can be used to support analytical reasoning and decision making" (Chandler et al., 2015, p. 1). The researcher used HoloLens to evaluate a space-time cube. Research focusing on evaluating a HoloLens to display a space-time cube visualization is limited. Moreover, few studies (Beitzel, Dykstra, Toliver, & Youzwak, 2018 ; Baumeister et al., 2017) use the HoloLens to measure perceived workload in 3D data visualization because of the novelty of the device. The researcher also examined the experiential differences between using the HoloLens and a desktop computer. Thus, this work adds to the body of knowledge in data visualization, specifically regarding how users perform using the HoloLens in terms of time, error and workload when doing tasks involving a space time cube. This study can also be used as a stepping stone in understanding if a 3D visualization displayed in an immersive environment such as a HoloLens display performs differently than in a desktop environment. This type of work may benefit the field of augmented reality and head-mounted display industry as the result will give the industry some insight into the capabilities of the HoloLens. In a general sense this study tells us which device is better suited to specific tasks and what factors may influence users' performance.

1.4 Purpose Statement

The researcher's intent in this study was to understand users' performance, specifically in terms of time, error and perceived workload, in different display conditions while manipulating a space-time cube. A convergent mixed-method design was applied to allow the researcher to better understand the research problems. Convergent mixed-method design is a research design in that qualitative and quantitative data are gathered separately, analyzed independently and then combined (Creswell, 2014). In the study, time, error and perceived workload were used to test performance to detect if a display condition had a positive or negative influence on users' abilities to perform a task. The qualitative data explored the differences in users' experiences with the HoloLens and desktop. This was done by observing the users during the study and debriefing them after the study. The rationale for using both quantitative and qualitative data was to achieve a broader understanding of the differences in user experience between the displays by incorporating the perspectives of the users and analyzing them in terms of the data to explain the experiment results.

1.5 Philosophical Foundation

To understand the method used for this study one must understand the researcher's philosophical foundation, which is a pragmatic worldview. Works within the pragmatic research space focus on the research questions and use any existing methods to comprehend the research. (Rossman & Willson, 1985). A pragmatic worldview does not align with any reality (Creswell, 2014). Morgon states the pragmatic worldview is the foundation of mixed-method research (2007). Of the commonality of pragmatic and mixed-method research, Creswell has stated "pragmatists do not see the world as an absolute unity. In a similar way, mix-method researchers look to many approaches for collecting and analyzing data rather than subscribing to one way (e.g., qualitative or quantitative)" (2014). Due to the researcher falling into the pragmatic mindset, the researcher based the study on the premise that collecting multiple types of data would provide the most comprehensive understanding of the research problems.

1.6 Research Questions

The study was intended to answer three research questions. The first question analyzed the users' performance and compared performance between the HoloLens and desktop devices. The second question investigated differences in user experiences between HoloLens and a desktop. The last used mixed-method analysis to gauge how the qualitative themes could help the researcher understand the meaning of the quantitative responses.

What are the differences in task performance on a space-time cube between HoloLens users and desktop users?

What are the experiential differences using the HoloLens compared to the desktop as reported by users?

What aspects of users' experiences potentially influence task performance?

1.6.1 The Hypotheses

There is a significant difference in time for tasks performed in a HoloLens and a computer screen.

There is a significant difference in error for tasks performed in a HoloLens and a computer screen.

There is a significant difference in perceived workload for tasks performed in a HoloLens and a computer screen.

1.6.2 Null Hypotheses

There is no significant difference in time for tasks performed in a HoloLens and a computer screen.

There is no significant difference in error for tasks performed in a HoloLens and a computer screen.

There is no significant difference in perceived workload for tasks performed in a HoloLens and a computer screen.

1.7 Definitions

Workload "is the cost of accomplishing mission requirements for the human operator" (Hart, 2006, p. 904).

Effectiveness (Error) "is the accuracy and completeness with which users achieve certain goals. Indicators of effectiveness include quality of solution and error rates" (Frøkjær, 2000, p. 345).

Efficiency (Time) "is the relation between (1) the accuracy and completeness with which users achieve certain goals and (2) the resources expended in achieving them" (Frøkjær, 2000, p. 345).

1.8 Assumptions

The researcher made the following assumptions when designing the study:

- The participants will complete all the tasks to the best of their ability's.
- The participants were not in a rush and good-naturedly went through the study.

1.9 Limitations

The subsequent limitations were reflected when designing the study:

- The study uses a questionnaire to determine workload and not an eye tracker or electroencephalography.
- The study was narrow to students at Purdue University.
- The study uses one type of augmented reality head-mounted display and cannot be generalized to all augmented reality displays
- The study uses one type of 3D visualization and cannot be generalized to all 3D visualization.
- The study used one six type of task and cannot be generalized to all task.

1.10 Summary

This chapter provided an overview of the proposed study, including problem statement, purpose statement, philosophical foundation, significance, scope, definitions, assumptions and limitations. The next chapter will provide an overview of related research.

CHAPTER 2. REVIEW OF LITERATURE

The chapter will provide a summary of 3D visualizations, how they are displayed, how people interact with them and how they are measured.

2.1 3D Visualization

3D visualizations display abstract and non-abstract information in 3D space. Abstract visualizations can display representative or numerical data, which cannot be referred to by size. Abstract 3D data visualization focuses on translating abstract information into 3D imagery. An example is seen in the work of Carpendale, Cowperthwaite, and Fracchia, who visualized 2D data as a 3D representation (1996). Another example is Liang and Sedig's work in mathematical visualization (2010). Non-abstract visualization, also called scientific visualization, focuses on visualizing objects that exist in the real world, such as a human heart or a volcano. Non-abstract 3D visualizations are believed to be useful for spatial data. Examples of this type of visualization can be found in chemistry visualization, geography visualization, health visualization, engineering visualization and more. One of the differences between scientific visualization and data visualization is that in scientific visualization the data gives the spatial representation and in data visualization, the designer picks the spatial representation. Thus, one of the advantages of scientific visualization is that the user of the visualization can reference their personal experiences to assist them in mapping the 3D space (Teyseyre & Campo, 2009).

In conclusion, non-abstract and abstract 3D visualizations are comprehensive (R. Brath, 2014) works including space-time cubes (Bach, Dragicevic, Archambault, Hurter, & Carpendale, 2017) and 3D scatterplots (Elmqvist, Dragicevic, & Fekete, 2008), among other types of representations. Every type of data visualization, abstract or non-abstract, has its own advantages and disadvantages. In the next subsection, the researcher will focus on one type of data visualization, a space-time cube, and in the section after that will summarize the advantages and disadvantages of this type of 3D data visualization.

2.2 Space Time Cube

A space-time cube is a 3D data visualization that displays spatiotemporal data using the X and Y dimensions to represent longitude and latitude and the Z dimension to represent time. Figure 1 shows an example of a space-time cube. The concept of a space-time model, originally introduced by Torsten Hagerstrand, includes space-time prisms, space-time paths and space-time cubes (1975). Kraak believed the space-time model to be the beginning of time-geography studies (2003). Within the data visualization community, the space-time cube is used as a tool to visualize spatiotemporal data (Gatalsky, Andrienko, & Andrienko, 2004; Kapler & Wright, 2004). Recently, space-time cubes have been used to visualize crime rates (Nakaya & Yano, 2010), sport (Moore, Whigham, Holt, Aldridge, & Hodge, 2003) and other types of spatiotemporal data (Bach, Dragicevic, Archambault, Hurter & Carpendale, 2014). One advantage of the space-time cube relevant to this study is that it is three-dimensional in nature. Kristensson et al. found that novice users are more likely to have performance errors, while more experienced users are likely to find complex space-time patterns faster (2009). The space-time cube was chosen as the testbed for this study because of its inherent 3D nature and use of spatiotemporal data. This is important because previous researchers who studied the differences between a HoloLens and a desktop for 3D visualizations all used spatial temporal data (Bach, Sicat, Beyer, Cordeil & Pfister, 2018; Nguyen., Ketchell, Engelke, Thomas &, De Souza, 2017). Thus, the researcher wants to evaluate another form of spatial temporal data that was not evaluated in previous studies so as a research community we can have a more holistic understanding of how these 3D visualizations are displayed in HoloLens and desktop.

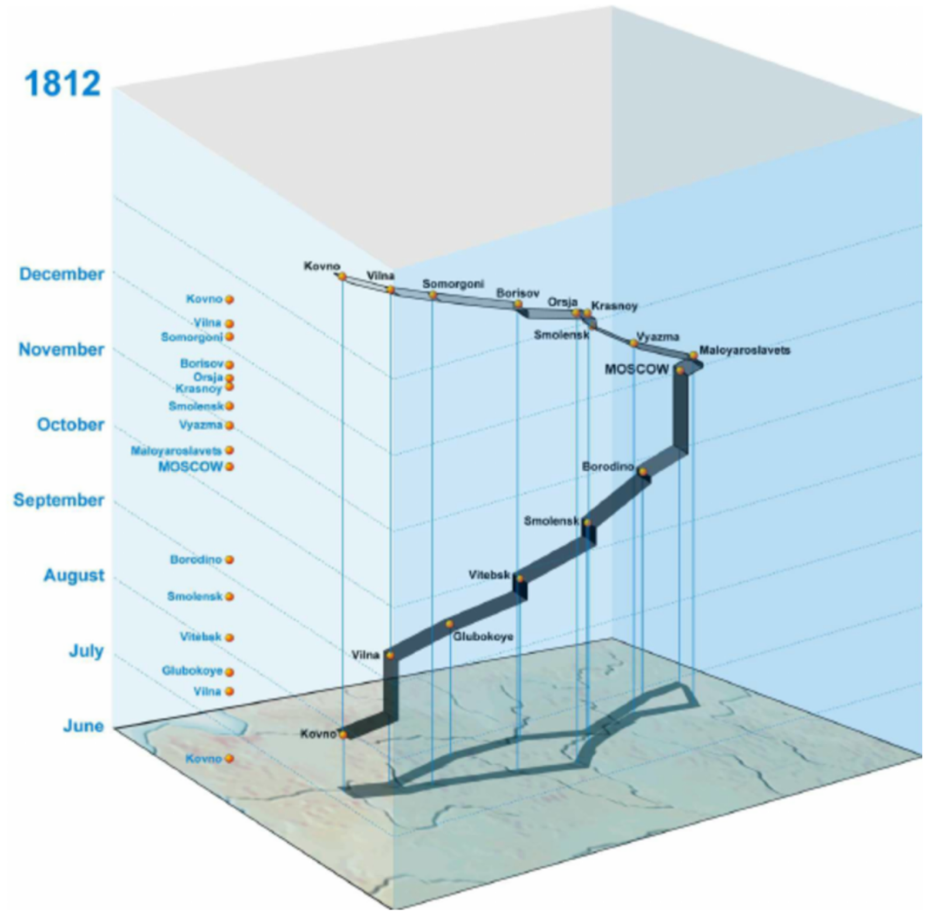


Figure 1: Space-time cube; data visualization of napoleon's 1812 march to Russia (M.Kraak, 2003)

2.3 The Advantages and Disadvantages of 3D Visualization

The research community has debated the advantages and disadvantages of 3D visualization and identified several issues (Brath, 2014; Shneiderman, 2014; McIntire & Liggett, 2014; Smallman, John, Oonk & Cowen, 2001). In Brath's (2014) position paper, "3D InfoVis is here to stay; deal with it," he lists six issues with 3D visualization. His concerns are navigation, occlusion, manipulation, non-anchored points, perspective perception, and 3D fonts.

The first issue, navigation, affects users interacting with visualizations using a mouse and keyboard, making it more difficult for these users to understand the visualizations (Brath, 2014). The second issue, occlusion, can occur in all types of visualizations instead of just 3D visualizations. Occlusion occurs when visual points overlap each other, preventing users from seeing the point in the dataset (Brath, 2014).

The third issue of manipulation, which afflicts 3D visualizations, is a consequence of navigation and occlusion and occurs when users have difficulty selecting and moving the correct data point (Brath, 2014).

The fourth issue of non-anchored points refers to a visualization having 3D points floating in space with no background or grounding image (Brath, 2014). Users will have no ground to use as in reference to the anchor points, viewing them instead as a 3D scatterplot. This is an issue for virtual reality and desktop environments.

The fifth issue with data visualization is perspective perception (Brath, 2014). This occurs when an object in the foreground looks bigger than an object in the background, making it hard to interpret the objects' size correctly. This is a prominent issue in non-interactive 3D visualizations.

The sixth issue with 3D visualizations is that 3D font is difficult to read in 3D space. This is a result of technological limitations in displaying font, such as low resolution. Brath suggests that new technologies may remove this obstacle (2014).

The design of the visualization used in this study used Brath's issues of 3D visualization twofold. First, the concerns were used as guidelines and rules that the researcher had to follow to implement a testable 3D visualization. Second, the list of issues was used as design rules when choosing criteria for this study such as display technologies, interaction and data visualization. Despite the drawbacks, 3D data visualization researchers suggested 3D visualization has several advantages (Brath,, 2014, ; McIntire, & Liggett, 2014). An advantage of 3D visualization is that it uses the extra dimension to encode more data and helps the understanding of relationships between objects by integrating local views with a worldview (Brath, 2014). The use of the extra dimension is seen in spatiotemporal data; the x- and y-axes are used to anchor users in a geographical location and the z-axis in time. Furthermore, Brath, Peters and Senior believe that the aesthetic elements of animation and 3D graphics enhance the design appeal to the user, thus increasing the intuitiveness and memorability of the visualizations (2005). In Teysseyre and Campo's (2009) survey paper, they state that the strength of 3D visualization is that it holds information density (2009). Furthermore, Brath found that lighting is an advantage for 3D visualizations, allowing designers to use light elements to find data points in the visualization (2014). While keeping in mind the advantages and disadvantages of 3D visualization, the researcher also investigated cognitive and perceptual considerations for working with 3D data visualizations. The next section will focus on these cognitive and perceptual considerations.

2.4 Cognitive and Perceptual Considerations for 3D Data Visualization

When viewing objects, we view them in three dimensions (i.e. having length, width and depth). Nevertheless, when images are projected on the retina, we see them as flat (2D), causing depth information to be distorted. This can cause issues such as occlusion, relative size, and linear perspective (Marriott et al., 2018). The depth information lost is assumed to be present when a 3D figure is flattened as on a printout or computer screen (Neubauer, Bergner & Schatz, 2010). The effects of perception have been studied within the augmented reality community, specifically augmented objects in the real word (Kruijff, Swan & Feiner, 2010; Renner, Velichkovsky & Helmert, 2013). In the past, researchers have used spatial cues when encoding data, position, size and motion in data visualizations (Lum, Stompel & Ma, 2002; Mackinlay, 1986). While potential solutions for addressing perception problems have been, for example, using cast shadows, grids, and focus work still need to be done (Luboschik, Berger, & Staadt, 2016).

In addition, specific concepts and structures that are inherently 3D (e.g. organs and atoms) are often depicted with 2D diagrams. This may lead to confusion, as students often form incorrect or incomplete mental models. It is argued that 3D visualizations may help correct and construct mental models (Chen, Hsiao & She, 2015). However, even though 3D visualizations may be beneficial for developing mental models, research suggests that not all students will benefit from working with them.

2.5 Visualization Displays

In this section, the researcher will discuss how visualizations are displayed and how the display affects the performance of the visualizations. 3D visualizations are evaluated on different display technologies such as 2D (monoscopic) displays, large 2D displays (Andrews, Endert, Yost & North, 2011), 3D (stereo) displays (Alper, Hollerer, Kuchera-Morin & Forbes, 2011), augmented reality head-mounted displays (Mojica et al., 2017), augmented reality on tablets (Saenz, Strunk, Maset, Malone & Seo, 2015), data physicalizations (Jansen, Dragicevic &

Fekete, 2013), and immersive virtual reality environments (Donalek et al., 2015). As discussed in the previous sections, when a desktop is used to show 3D data issues of distortion, occlusion and multi-dimensional information loss have occurred. Numerous techniques have been suggested to assist with the problem of occlusion for all displays when working with simple data sets (Luboschik, Berger & Staadt, 2016). One such remedy is the use of a 3D stereoscopic display. McIntire, Havig and Geiselman found that 3D stereoscopic displays perform better than 2D screens for shape understanding, classification tasks and interaction (2012). One immersive technology researchers have been using to try to solve the issue of occlusion is virtual reality. When the data surrounds the user, increasing the field of view has been found helpful (Barrie, Cassell & Cooper, 2005) but also questionable (McMahan, Gorton, Gresock, McConnell & Bowman, 2006). The issue with using virtual reality to display data is that it decouples the user from the real world. This decoupling effect does not occur with augmented reality because the user can see the world around them, and head-mounted display augmented reality displays such as the HoloLens provide stereoscopic views without the decoupling. Furthermore, a study by Baumeister et al. showed that the use of augmented reality resulted in increased performance and reduced cognitive load (2017). Thus, the researcher believes that if one is testing with immersive technology at this point in time, it may be best to use a HoloLens.

2.6 HoloLens

This subsection will contain examples of work that use the HoloLens as the display setting for 3D visualizations. The HoloLens is a mixed-reality head-mounted device developed by Microsoft. Mixed-reality visualizations place 3D visualizations in a natural environment requiring the use of a desktop mouse and keyboard.

Mahfoud, Wegba, Li, Han and Lu studied immersive visualization using the HoloLens within an aspect of a security application that shows irregular events in multi-source and time-series data. (2018) (see Figure 2). They present their work as a case study of an analyst using the system on site. There focuses is on the design and algorithms of their visualizations. They suggest that the HoloLens can improve spatial cognition of users for a more effective analysis and connection with data. They imply that 3D visualization will have advantages displayed on the HoloLens over desktop systems. Their study does not have any user quantitative components attached only system response time for their visualizations, but they do ask users to perform tasks within the software. It addresses similar issues to this thesis such as how a HoloLens performs for displaying spatiotemporal data compared to a desktop but the authors present their study of the application being using real settings and discuss limitations that are found in the application in the world while my study focuses on understanding the hardware in a lab setting.

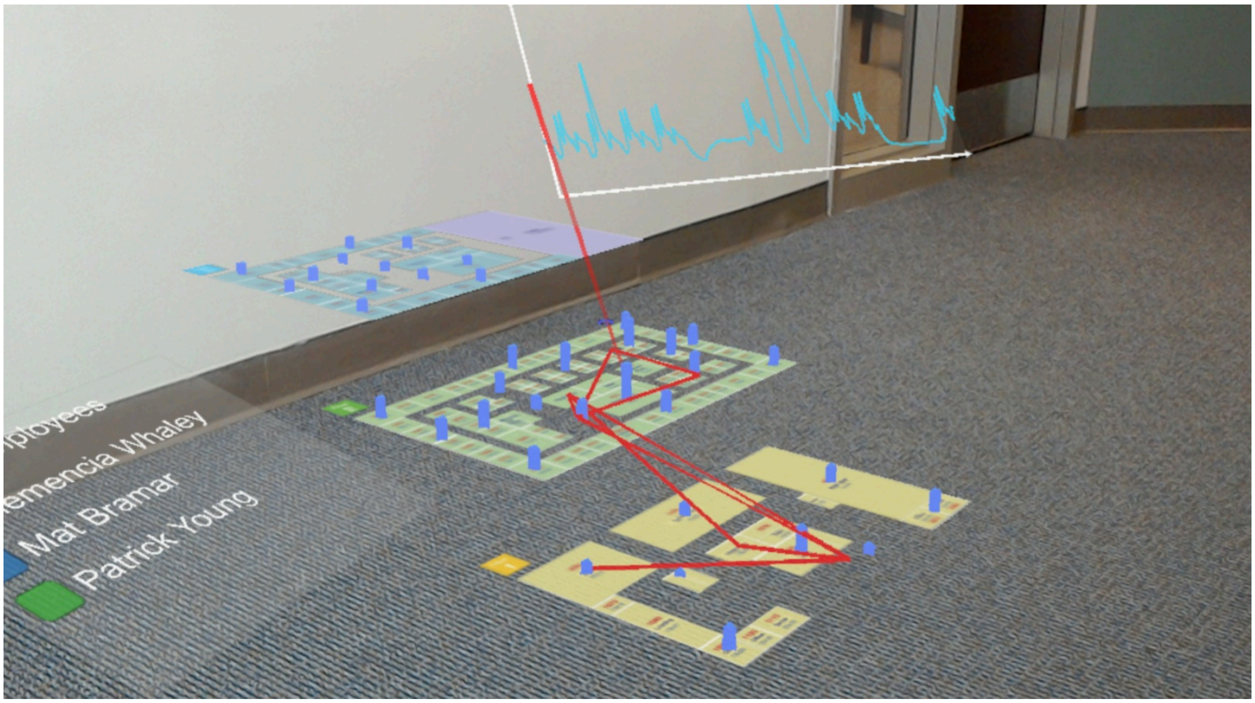


Figure 2: HoloLens security application (Mahfoud, Wegba, Li, Han & Lu, 2018)

Mojica et al. (2017) describe a HoloLens application for planning MRI-guided neurosurgeries. They compared a 3D visualization on a HoloLens against a 3D visualization on desktop and 2D MRI slices (see Figure 3). They found that the HoloLens' visualization worked better for critical structure detection (e.g. vessels and lesions) and assisted in demonstrating spatial relationships for both normal and pathological conditions. While this is a scientific visualization, this shows that the HoloLens might be useful for understanding spatial relationships and finding points, which are common for space-time cubes. Their study uses self-reported data to find their findings while in this thesis also uses quantitative metrics.

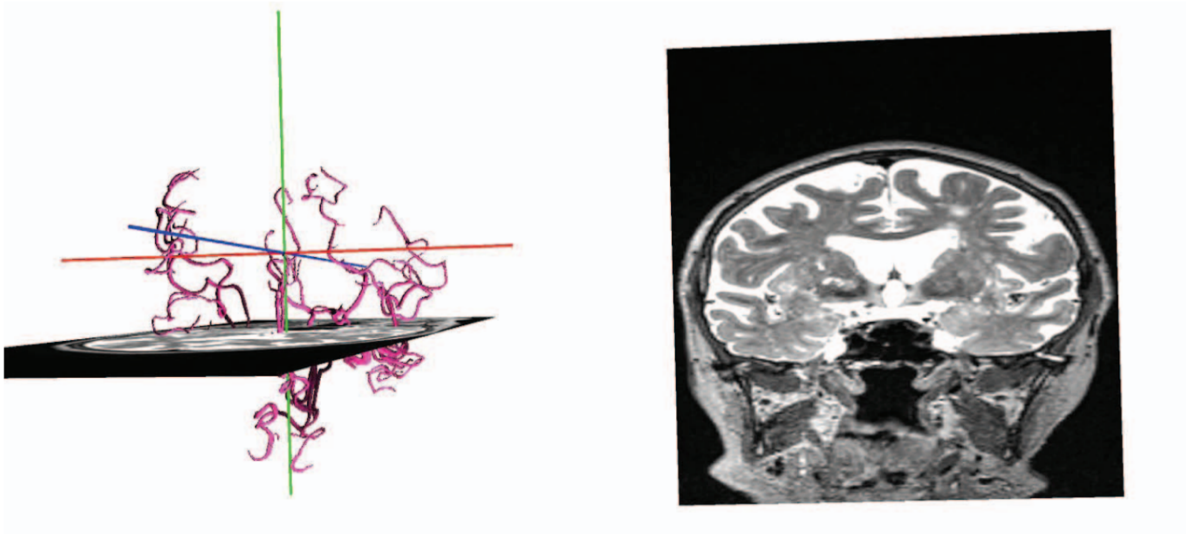


Figure 3: HoloLens MRI 3D visualization (Mojica et al., 2017)

Kuhlemann, Kleemann, Jauer, Schweikard and Ernst (2017) developed a navigation framework which projects a 3D visualization of the vascular system. (see Figure 4). Their study allowed people to interact with the software, finding that users are acceptive of the HoloLens in clinical applications. While not the same as a data visualization application, it shows that a HoloLens is accepted as a tool in other settings.

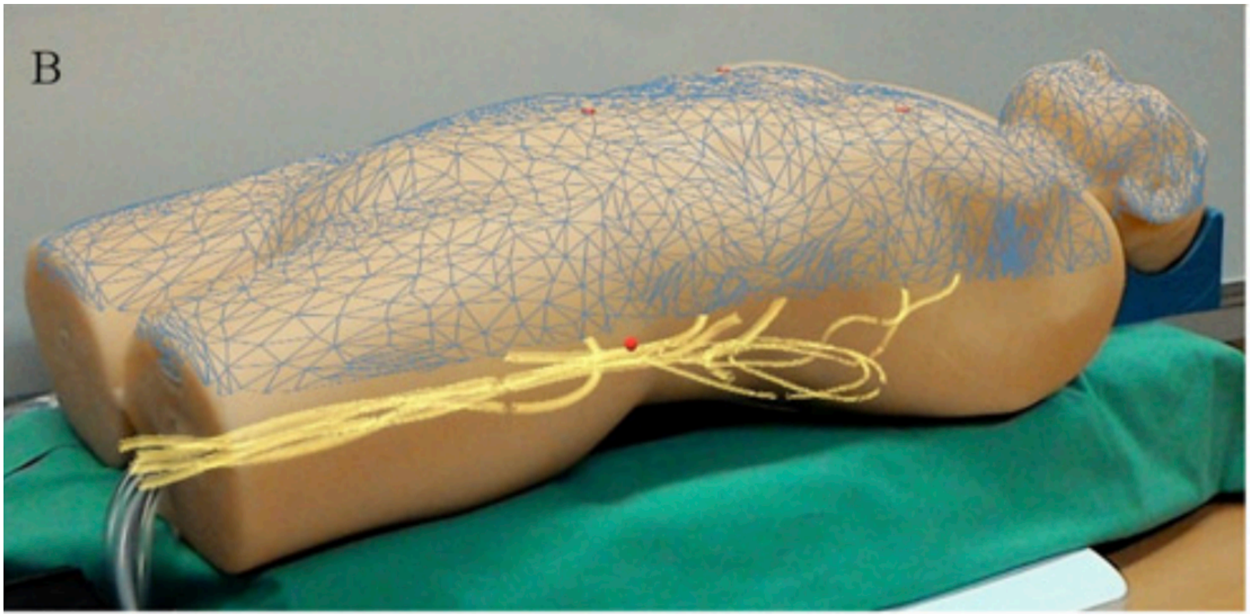


Figure 4: Vascular system using the HoloLens (Kuhlemann, Kleemann, Jauer, Schweikard & Ernst, 2017)

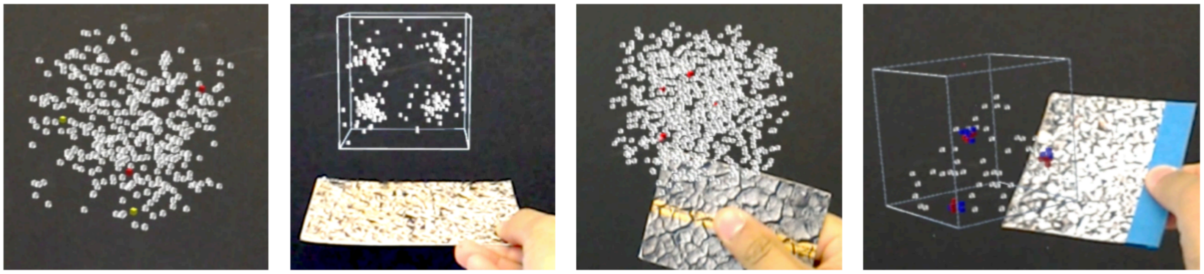
2.7 Interaction Styles of 3D Visualization

Within this section the researcher will discuss how users interact with 3D visualizations, covering traditional interaction such as a mouse and keyboard along with emergent interactions. The most common form of interacting with a data visualization is using a mouse and keyboard. In order to interact, a user must use the keyboard and the mouse simultaneously; for example, if a user wants to do a 90-degree rotation they would have to press a key on the keyboard to activate the rotation command and use the mouse to move the object 90 degrees. Another approach to interacting with a 3D visualization is using natural gesture interaction. In this approach, users execute a gesture in midair to interact with the visualization; for example, the user can swipe their hand from left to right to rotate an object left to right. In Nancel, Wagner, Pietriga, Chapuis and Mackay's work, they found that natural gesture interaction is less beneficial for highly precise interactions than a keyboard but can be valuable for high-level interaction (2011). Another method of interacting with 3D visualization is using tangible user interfaces. Among this variety of interaction, users use wedges for a command - for example, by using a tangible saw to cut the visualization in half. Bach, Sicat, Beyer, Cordeil and Pfister (2018) found that tangible interaction coupled with an augmented reality head-mounted display is beneficial for highly interactive tasks requiring precise manipulation. Each form of interaction has its strengths and weaknesses, and the best type to use depends on the context that the researcher wants to study.

2.8 Immersive Analytics

Commonly, visual analysis focuses on a desktop environment, but with the advance and availability of immersive displays like HTC Vive and HoloLens a new research thrust has emerged known as Immersive analytics. Immersive analytics "investigates how new interaction and display technologies can be used to support analytical reasoning and decision making" (Candler et al., 2015, p. 1). Immersive analytics comes from technologies such as interactive large interactive displays, VR and AR environments such as the HCT Vive and HoloLens, natural user interface devices and other technologies (Candler et al., 2015). The questions asked within the research thrust center on how displays and interface affect visual analysis (Candler et al., 2015). Many different works focus on immersive analytics; thus, to limit the scope the researcher focused on the works that use the HoloLens. The difference between this subsection and the previous HoloLens section is that these types of work investigate the use of the HoloLens for visual analytics tasks.

Bach, Sicat, Beyer, Cordeil and Pfister (2018) did a study comparing three display environments - desktop, HoloLens and tablet - for completing interactive exploration tasks using 3D visualizations. They found that direct interaction with 3D visualizations using tangible user interface improved time on task and error for tasks that required synchronization between user perception and interaction but had less error then the desktop. They also noticed all users stood when interacting with the holograms but were not fatigued even after 40 minutes of the study. They believe the HoloLens may increase user engagement, while the desktop may be easiest to use in terms of interaction, perception and fatigue. However, while this study has similarities to the researcher's thesis, it does not investigate workload.



(a) Distance task (b) Cluster task (c) Selection task (d) Cutting plane task

Figure 5: 3D visualizations using tangible markers for the HoloLens (Bach, Sicat, Beyer, Cordeil & Pfister, 2018)

Nguyen, Ketchell, Engelke, Thomas and De Souza's (2017) work HoloBees is an analytic system that tracks data of the movement of bees. In the study, they compared the desktop against the HoloLens for performing analytical reasoning tasks, such as asking users to answer, "How many bees moved from Hive 3 in each month of the whole period" (p. 4)? They used time, error, numbers of interaction and user experience questionnaires as the metrics to compare the HoloLens and the desktop. The researchers found that the HoloLens and desktop had the same performance score in time and error, but the HoloLens required significantly fewer interactions than the desktop. This is interesting because fewer interactions could possibly mean a lower workload.

Figure 6: Augmented Reality Based Bee Drift Analysis: A User Study (Nguyen, Ketchell, Engelke, Thomas & De Souza's, 2017)

Zhang, Chen, Dong and El Saddik tested city data visualization to see how users performed in tasks such as finding and editing base task using a HoloLens and desktop. The study used the following metrics to compare the HoloLens and desktop: interaction, visual, combination with haptics, time and flexibility. The study found that HoloLens performed better in all metrics than a desktop except for response time. Additionally, the researchers asked each user to choose one environment; eight out of the 12 users chose the HoloLens, demonstrating that more users prefer the HoloLens over the desktop.

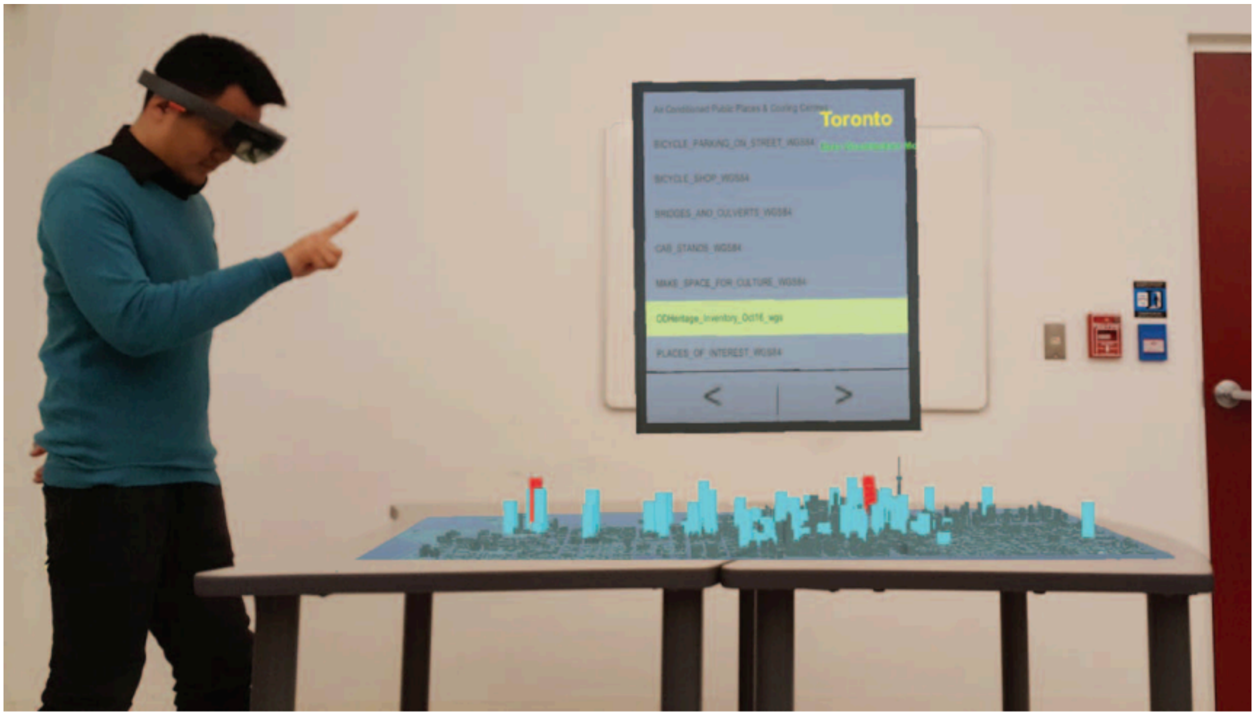


Figure 7: Data City visualization (Zhang, Chen, Dong & El Saddik, 2018)

2.9 Evaluating Data Visualizations

This section contains information on how to evaluate visualizations. A common method for evaluating data visualizations is a user study, which can use a range of techniques such as survey, field study, field experiment, laboratory experiment, experimental simulation, judgment study and computer simulation (Carpendale, 2008). When choosing a technique, researchers must consider the tradeoffs of each one and the desired goal of the study. For example, Carpendale states that there are three goals when evaluating data visualizations: generalizability, precision and realism. Carpendale used the work of McGrath (1995) to define these terms as “the extent to which it can apply to other people (than those directly in the study) and perhaps even extend to other situations” (p. 22), “the degree to which one can be definite about the measurements that were taken and about the control of the factors that were not intended to be studied” (p. 22) and “the extent to which the context in which it was studied is like the context in which it will be used” (p. 22), respectively.

This thesis collected data on three measures: time, error and perceived workload. Efficiency (time) is the measure of how quickly a person can finish a task and is commonly used when testing software. Efficiency "is the relation between (1) the accuracy and completeness with which users achieve certain goals and (2) the resources expended in achieving them" (Frøkjær, 2000, p. 345). This type of information can inform the researcher on how quickly a user can finish a task. Another common usability metric, effectiveness (error), "is the accuracy and completeness with which users achieve certain goals. Indicators of effectiveness include quality of solution and error rates" (Frøkjær, 2000, p. 345). While time and error are commonly used to test visualizations, workload is not as widely used.

2.10 Perceived Workload

In this subsection, the researcher will discuss the metric of workload. Workload is "the cost of accomplishing mission requirements for the human operator" (Hart, 2005, p. 904). The goal of data visualization is to convert data into illustrations to assist the cognitive burden of understanding of information, and it is known that visualization aids people in understanding data and provides cognitive support (Tory & Moller, 2004). Most visualizations are external memory aids for understanding data, an example of which is a space-time cube. A space-time cube can be used to assist with the understanding of spatiotemporal data by providing an external memory aid in cube form.

Time and error are the most common metrics for visualizations, but there are situations when they alone cannot distinguish performance differences between visualizations. If the goal of visualization is to reduce cognitive burden, then a metric to understand the cognition or workload input when using a visualization might assist in determining what type of visualization would be most beneficial. Huanga, Eades and Hongb argue that understanding workload can help designers make visualization to be more accurate since a direct correlation has been found between workload and error rate (2006).

In this study, the author uses the NASA-TLX survey to understand perceived workload (Hart & Staveland, 1988). Over 7,000 studies have cited the NASA-TLX survey, which was created by Sandra G. Hart and Lowell E.Staveland (1988) to measure perceived workload through six components: mental demand, physical demand, temporal demand, effort, performance and frustration level. Each factor is gauged by a question that the user can use as a guide to fill out the survey, as seen in Figure 8. The survey was originally designed to study stress endured by flight crews but has recently been used in other domains such as computer software, automobile, and medicine (Hart, 2006).

Fischer et al. (2008) studied workload using NASA-TLX survey in examining how animation speed impacts the understanding of dynamic visualizations. Yost and North (2006) similarly measured workload utilizing an altered type of NASA-TLX to understand the workload of different display sizes for visualization. Fjeld et al. (2007) used NASA TLX to compare the workload of using a chemistry visualization in augmented reality against the same visualization displayed with a physical model. Another study found that there was about the same amount of workload in augmented reality than in a desktop (Beitzel, Dykstra, Toliver,& Youzwak, 2018). Wiegmann, Overbye, Hoppe, Essenberg and Sun (2006) used NASA-TLX to compare a 3D visualization of power system information with a 2D diagram. They found no significant differences in workload but noted tasks were completed more quickly using the 3D visualization. Workload is also examined in other studies such as 3D anatomical visualization (Foo et al., 2013), multivariate visualization (Livingston, Decker & Ai, 2012) and ontology visualization (Fu, Noy & Storey, 2013). Overall, workload is a metric that can inform the researcher how much cost there is to accomplish a task, and while it is not as common a metric for evaluating visualizations as time and error there is still precedent for using it.

Figure 8: NASA-TLX RATING SCALE DEFINITIONS

Title	Endpoints	Descriptions
MENTAL DEMAND	*Low /High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	*Low /High*	How much physical activity was required (e.g.. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous. restful or laborious?
TEMPORAL DEMAND	*Low/ High*	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	*good/poor*	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	*Low/High*	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL*i*	*Low /High*	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Figure 8: NASA TLX survey (Hart & Staveland, 1988).

2.11 Summary

This chapter covered related literature, including 3D visualization, displays, cognitive and perceptual considerations for 3D data visualization, interaction styles of 3D visualization and evaluating data visualizations. The next chapter presents the methods implemented in this study.

CHAPTER 3. METHODS

This chapter will include the methods the researcher used to answer the research questions. This chapter consists of the research type, experiment procedure, testbed, tasks, data collection, sampling, data analysis, validation and perspective.

3.1 Research Type

To understand users efficiency, effectiveness and workload in performing tasks within a space-time cube visualization with a HoloLens and a computer screen, a convergent parallel mixed-method design approach was applied. When finding a definition for mixed-method studies Johnson, Onwuegbuzie and Turner (2007) analyzed 19 definitions of mixed-method research to develop a more holistic definition. They stated that,

> *"Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration" (p.123).*

The researcher also used convergent parallel design in this study. Creswell, a researcher in mixed-method design, defines convergent parallel design as an "approach, a researcher collects both quantitative and qualitative data, analyzes them separately, and then compares the results to see if the findings confirm or disconfirm each other" (Creswell 2003 p.219). The combination of qualitative and quantitative methods allowed a holistic understanding of students' performances by providing multiple types of information. An entirely quantitative research approach might be suitable for measuring time, error and workload and would provide insight on performances, but would not produce the meaning from data such as why users had high physical demand scores. Furthermore, it would not offer an understanding into peoples' opinions on why they believe a task is challenging or simple. The opposite is true about a qualitative research approach. It might provide potential insights into peoples' thoughts and emotions, for example individual preferences and characteristics of interaction, regarding a task but not statistical analysis and generalizations about time, error and workload. Hence, a mixed-method strategy can combine these techniques by allowing the researcher to collect more comprehensive data and use all tools available for both exploration and analysis in the same study. To qualify this, a mixed-method approach can be risky because it is very time-consuming, and sometimes the data doesn't converge. Nevertheless, the researcher argues that this study should be done in mixed-method because it can provide comprehensive analysis of users' performance.

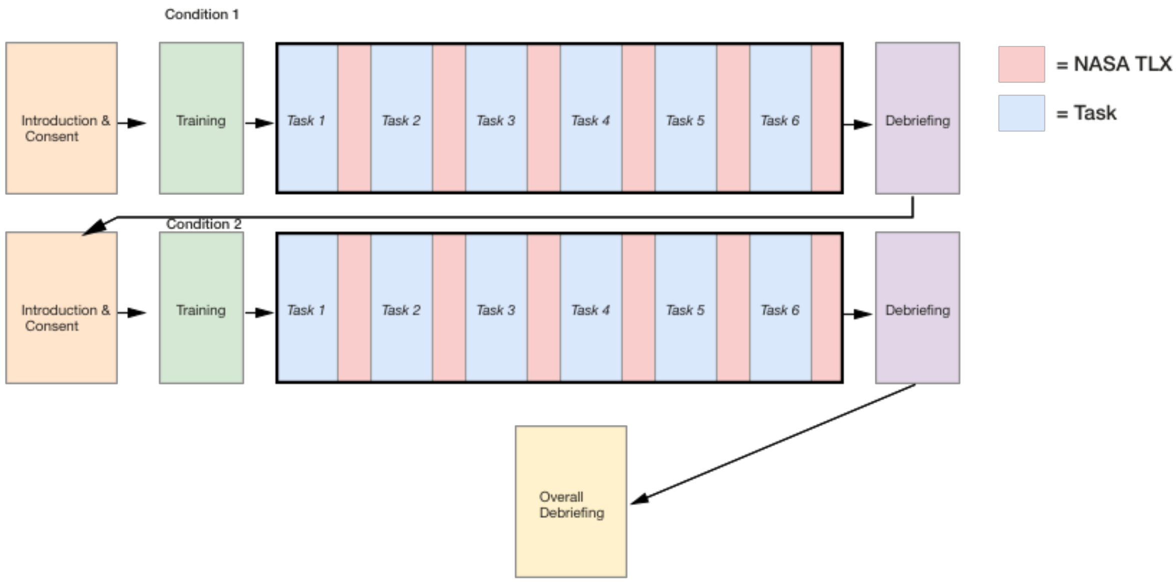


Figure 9: An example of a user's progression through the study

3.2 Experiment Procedure

A within-subject study design was used for this project because this approach allows each user to use both devices (HoloLens and desktop) and permitted a smaller sample size. Each participant was randomly distributed into two different groups. The study took place inside a lab at Purdue university's' Wang hall (see Appendix D for lab diagram) A visual representation of the study can be seen in Figure 9. The study started with an introduction and signing of consent. The user then underwent a short training session where they learned how to work with a space-time cube, how to interact with the device and how to solve each task in the study without being explicitly told the nature of the tasks. After the training the users were requires to tell the researcher what each axis represent in the space time cube, where is the north, south, east and west in the space time cube and were required to do a full rotation around the space time cube, pan around the space time cube and zoom in the space time cube. The training continued until the user felt comfortable with the device, task and verbally assented to their understanding. After the training the researcher turned on and started the video camera and the screen recorder. Group A did all tasks on a HoloLens, followed by the same tasks on a computer with a different dataset. The study had six tasks: retrieve value, filter data, find extremum, determine range, sort and cluster, pattern finding (see Appendix A or Table 1 for more information on the task). The researcher used Amar, Eagan, and Stasko taxonomy (2005) as a guide for the tasks. The rationale for using these taxes because they are simple in nature and an are commonly done while working with any data visualization. Moreover, these tasks helped the researcher understand user's performance of low-level components of analytic activity. Group B did all tasks on a computer first and then on a HoloLens. Users were instructed to think aloud while performing the tasks. The task order was the same for HoloLens and the desktop. The user was allowed to take as long as they wanted to complete a task, performing the same task for both groups and both display conditions but with different dataset. (see Testbed section for information) After each task, the user filled out the NASA TLX survey to rate their perceived workload (see Appendix B for NASA TLX survey). The user had the options to fill out the NASA TLX survey verbally or written. The researcher took notes on the user's behavior and comments for the duration of the study, which were used to ask questions in the debriefing and for the data analysis. After completing all the tasks for both display conditions, the user was debriefed about their experience (see Appendix C for debriefing protocol).

Table 1: Task

Task Number	Task	Question
Task 1	Retrieve Value	How long (time) did the cat travel?
Task 2	Filter Data	In what location (latitude and longitude) did the cat spend the most time?
Task 3	Find Extremum	How far east did the cat go? Identify the specific point. How far south did the cat go? Identify the specific point.
Task 4	Determine Range	What is the range of latitude/longitude in which the cat traveled?
Task 5	Sort and Cluster	Are there clusters in time? Can you rank the clusters the from biggest to smallest?
Task 6	Pattern Finding	Is there a clear pattern in the visualization? Can you provide a summary of content within the visualization?

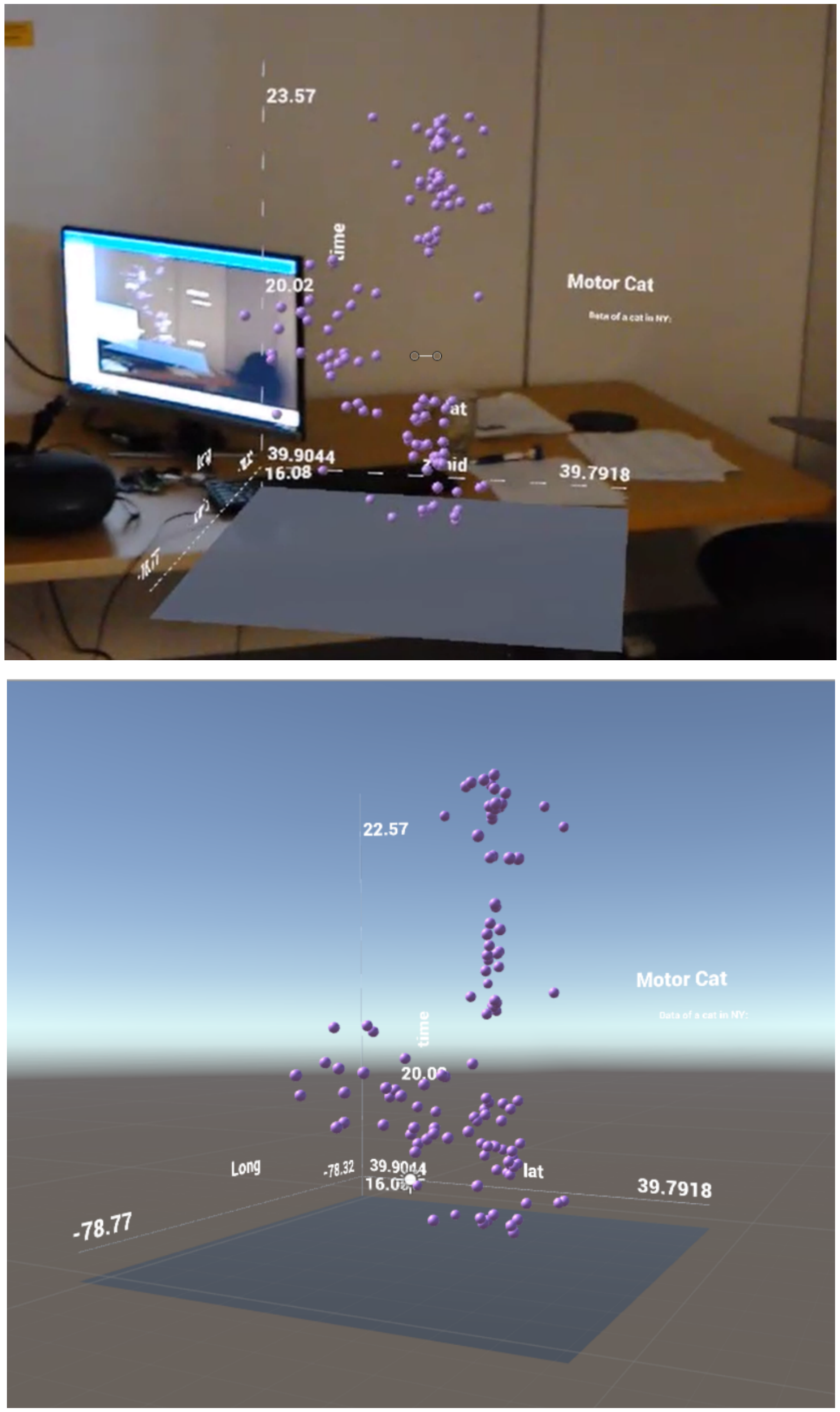


Figure 10: Space time cube in the computer and HoloLens

3.3 Testbed

Two data visualizations, one for the HoloLens and one for the desktop, were designed and developed for the purpose of testing the research questions for this study. Both visualizations are space-time cubes of the same cat traveling in New York but on two different days. The dataset has 81000 data points but only 246 data points were used. The dataset came from movebank.org. and was called Pet Cats (Kays, 2016) The data visualization was developed in the Unity game engine due to the researcher's familiarity with the program. Figure 10 shows both data visualizations.

3.4 Sampling

In this study, a purposeful sample was used. A purposive sample is a sampling technique that allows the researcher to select users base on their characteristics. (Palinkas et al, 2015). This was necessary for the researcher to be able to recruit only users who were not experts with a space-time cube or a HoloLens. The reason inexperienced users were selected is because the researcher is interested in what non-experts' initial opinions of using the HoloLens to analyze 3D visualizations may be. To find the sample size for the quantitative section that would meet statistical requirements, the researcher used the ideal power, effect size and significance level to calculate the sample size. Sample size for a two-group T-test with a power of .80 and effect size of .5 and significance level of .05 was calculated to be 17 per group. However, given that this study used a within-subject design, a sample size of 17 was determined to be too low to meet the researcher's face validity. Thus, the researcher investigated a local standard to increase the sample size. Local standards are strategies based on similar studies that have been published (Caine 2016). The "Local Standards for Sample Size at CHI" for an in-lab mixed-method study is a mean of 25 users (Caine 2016), thus the researcher increased the sample size to a minimum of 25. In addition, the local standards for CHI interview studies is 15 users (Caine, 2016). Therefore, debriefing for this study used a minimum sample size of 15. While a sample of 15 was propose during the data analysis process, the researcher decided to analyze all debriefing from the study. In total, 32 participants were recruited for this study.

3.5 Data Collection Method

A core concept of convergent parallel mixed-method design is that the same topics must be addressed in the qualitative and quantitative data. The three topics were measured in the study, time, error and workload, were addressed in terms of both qualitative and quantitative data, at least at a high level. To measure time, error and workload through both the methods, the researcher gathered video data and conducted surveys while performing, debriefings and observations during the study.

3.5.1 Time and Error

The researcher collected various quantitative measures such as time, error and workload index score. *Time* is how long a person takes to finish a task, which was used to measure efficiency. The timer started when the researcher had finished explaining the task and the timer stopped when the participant stated their answer. *Error* is whether the user completed the tasks correctly or incorrectly, which was used to measure effectiveness. To find errors, the researcher referred to the video recordings of the tasks and marked the result correct or incorrect; the researcher then gave a binary error score, zero or one, per task. For example, if the task is to find the easternmost point in the visualization and the user identified the wrong point, their binary score for that task would be a zero.

3.5.2 NASA TLX Survey Responses

The *workload index score* is the workload on the user while doing a task. To find perceived workload, the researcher used a hybridized version of the NASA TLX survey (Hart, & Staveland, 1988). NASA TLX is a survey used in human factor, human-computer interaction and other fields to understand users' mental workload (Hart, 2006). For this study the researcher used the raw NASA TLX that removes the weightings between paired dimensions, which is easier to use and shows results similar to the weighted-score version (Byers, Bittner, & Hill, 1989). The hybrid survey replaces the NASA TLX's twenty-one-point scale with a ten-point scale, the reason for the modification being that users did not understand twenty-one-point scale in the pilot study. A ten-point NASA TLX is seen in Schoeffmann's work (2014). Finally, the researcher asked the user which display method was easier to use and which one they prefer.

3.5.3 Debriefing

For the qualitative method, debriefing and observations were used in this study. According to Berg and Lune (2012), interviews can be used to understand users' thoughts, and can get the story behind their experiences. Therefore, the researcher used semi-structured interviews as a method of collecting data regarding users' perceptions of using each device. According to Berg and Lune (2012), a semi-structured interview is a type of interview that is not entirely standardized but not fully unstandardized. This kind of interview permitted the researcher to adjust to each user-specific situation; for example, it allowed the researcher to question any habits observed in the study. While Berg and Lune paper focuses on interviews the researcher used it as a guide for the debriefing. The goal of the debriefing is to contribute a broader perspective of the experiential differences of using the HoloLens and the desktop. The researcher made observations, by taking notes in his computer, during the study on notable patterns or methods of interaction the user did.

3.6 Data Analysis

To understand the data collection methods and data analysis methods a visual model, Figure 11, is used to help the reader comprehend the mixed method approaches being used in the study. Its argued that in mixed method research a visual model can assist readers to understand the study (Morse 1991; Tashakkori and Teddlie 1998; Creswell et al. 2003; Creswell 2005). The visual model in the study follows the ten guidelines for drawing visual models for mixed method model made by Ivankova, Creswell and Stick. (2006). The visual model is a summary of the data collection method and data analysis sections.

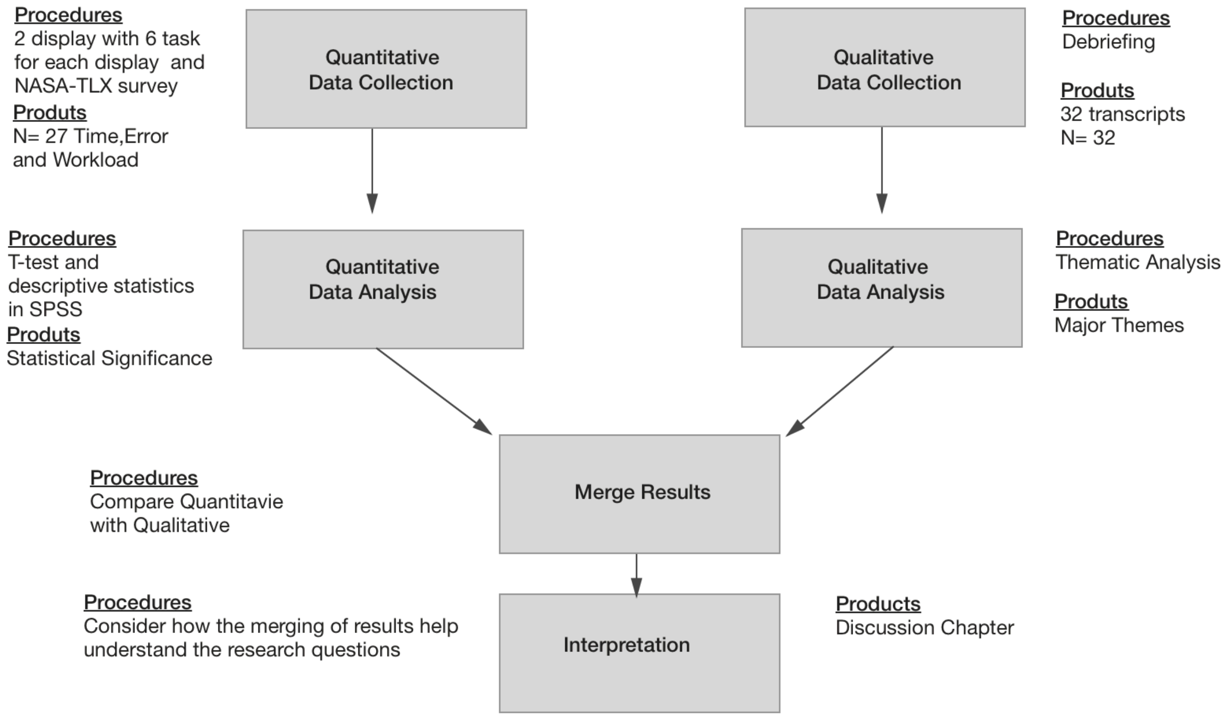


Figure 11: A visual model of the methods and analysis

3.6.1 Quantitative Analysis

For the quantitative analysis, descriptive statistics are reported, e.g. means. The independent variable was the display type and the dependent variables were time, error and workload in performing the tasks. The quantitative data analysis was used to determine any statistical differences between the task performed and the display environment, allowing the researcher to reject the null hypotheses. To examine the hypothesis, the researcher used a Student's T-test or Wilcoxon tests to see if there was a statistically significant difference between the means when Alpha = 0.05.

After the conclusion of the experimental data collection, recorded video data was transferred into Camtasia and VLC for analysis. The researcher identified time-on-task by examining each video and adding time stamps for when each user started and ended each task. These time stamps were recorded in a spreadsheet. During this process the researcher cleaned the data by discounting any time that the user stopped performing a task—for example, when a user stopped to ask a question. Out of 324 tasks analyzed in this study only 21 tasks had to have time remove because of user help. Out of 21 tasks a majority of users needed help with understanding their directions in relationship to the space time cube this was seen in Task 3 and Task 4. Few users asked the researcher to repeat the task question again. Moreover, two users need a reminder on what are cluster and how to find them? The researcher also discovered that video or audio was not properly recorded for one or two tasks for 5 participants and was thus unable to identify time-on-task for these participants. Data for these participants was removed from the quantitative data analysis due to the possibility that such incomplete data could affect the accuracy of the quantitative results, with data on 14 Group A users and 13 Group B users remaining. All 32 participants were considered in the qualitative analysis. To compute error rates, the researcher repeated the process of analyzing each video and marking for error. After both error and time on task were identified for all participants, data was transferred to SPSS along with the paper-based surveys. Next, variables were created using display type, task number and workload category. Afterward, the data was checked for normal distribution via Shapiro Wilk test. If the data was normally distributed a pair T test was used. However, if the data was not normally distributed Wilcoxon signed-rank test was used to find significant differences on the display.

3.6.2 Qualitative Analysis

For the qualitative data analysis, a thematic analysis was used. According to Braun and Clarke (2006), "thematic analysis is a method for identifying, analyzing and reporting patterns (themes) within data. It minimally organizes and describes (the) data set in (rich) detail" (p. 79). To utilize a thematic analysis, the researcher followed the method provided in the article by Braun and Clarke (2006). First, the researcher compiled all the debriefing videos to more easily view and transcribe them. Then the researcher viewed the videos multiple times in order to determine the quantitative data results. The researcher then transcribed the audio recordings into text using Temi, a speech-to-text transcription service, and subsequently reviewed and corrected the transcriptions. After reviewing the data multiple times, the researcher generated initial codes by highlighting and commenting on the transcription in a word editor. He then printed all the initial codes on note cards, search for theme and sub themes for themes, the researcher was able to find and define 24 themes. By grouping note cards with similar meanings, the researcher was able to condense this number to 15 themes. By using the codes to further generalize the themes, the researcher ultimately determined and defined five themes. Consequently, the researcher changed his second research question to "What are the experiential differences using the

HoloLens vs. the desktop?" from "What is the user's perception of performance in tasks on a space-time cube in a HoloLens and the desktop environment?" The rationale for the change was that the original themes did not answer the initial question. Furthermore, this change allowed the researcher to focus more on the users' perceived experience other than just performances. Considering this change, the researcher reanalyzed the data based on the 5 themes and 15 codes used in the previous code cycle. The researcher ultimately found 11 themes belonging to 5 categories, which he reviewed, defined and reported (see Chapter 4 for the report).

3.6.3 Mixed Analysis

Both quantitative and qualitative analyses were used to evaluate the space time cube visualization. After analyzing each, a side-by-side comparison was performed, with the researcher integrating each set of findings during the discussion of the study. Types of results that can come from this comparison is a discussion where the data converges. As stated in Chapter 1, the researcher asked both quantitative and qualitative research questions to better understand users' holistic performance and communicate the findings of the study to appeal to a larger audience. Following this process, the researcher used Student's T-test and descriptive statistics to answer research question number one, thematic analysis to answer research question number two, and both thematic analysis and Student's T-test to answer research question number three.

3.7 Validation & Trustworthiness

An issue with internal validity is mortality. People may drop out of the study. To combat this the researcher recruited 32 of participants and get a sub-set from them. To address trustworthiness and rigor in the qualitative data, procedures for ensuring trustworthiness and rigor of qualitative data were employed. Using validity techniques found in Creswell (2014), work. The methods are data triangulation; several different sources were used to collect data; a detailed description of the findings were provided, and any negative or contradicting findings were reported. Another method uses is confirming interpretations with the participants during the debriefings the researcher asked the participants if he was interpreting what they were saying correctly while this isn't a validity technique it did help the researcher interpret the debriefing.

3.8 Perspective

The researcher perspective is presented in first person here: I come from a background of human computer interaction, ubiquitous computing and visualization. I have five years of experience developing and designing for AR and VR. I have no monetization connections with HoloLens or Microsoft. I am a graduate student, positive or negative results do not affect my chances to graduate. I am approaching this study as a user researcher interested in how users perceive their performance.

3.9 Summary

This chapter provided an overview of the methods used to solve the research goal time, error, workload and debriefing are the data collection method. The analysis methods used are Thematic analysis and T-test. The deliverables of this research study will be the statistical report and themes generated from the data.

CHAPTER 4. FINDINGS

This chapter presents the findings of the study, such as demographic data including user-reported experience levels; descriptive statistics and T-test results; and the themes that emerged during the thematic analysis. The themes show insight on the experiential differences between using the HoloLens and desktop computer. Each theme will be described and accompanied by an example of how they were seen in the study, a direct quote from the debriefing, or users' spoken thoughts (see chapter 5 for a discussion on the theme). The chapter will conclude with a summary. A discussion on how the data converges is shown in the next chapter.

4.1 Demographics

For this study 32 participants were recruited (20 male and 12 female). The participants were recruited via a mailing list. Each participant was randomly assigned to one group; Group A or Group B. The highest level of education attained is as follows: 1 doctorate degree; 11 bachelor's degrees; 2 associate degrees; and 18 high school degrees. The participants were also asked to self-report their prior experience with the following categories: data visualizations, head-mounted displays, 3D software in general and space-time cube visualizations. Users self-reported their experience levels as none, beginner, intermediate and advanced. A self-report of none means no experience, a self-report of beginner means little experience and only understand the basic of the subject, a self-report of intermediate means they have some experience and understand a good amount of the subject and a self-report of advanced means they have a lot of experience and understand most of the subject. The results are as followed for data visualizations experience: 2 advanced, 3 intermediate, 18 beginner, 9 none. The results are as followed for head-mounted displays experience: 5 advanced, 3 intermediate, 17 beginner, 7 none. The results are as followed for 3D software in experience: 2 advanced, 3 intermediate, 19 beginner, 8 none. The results are as followed for space-time cube visualizations experience: 6 intermediate and 26 none. The demographic data of the participants is reported in Table 1.

Table 2: Demographic data of the participants. The highlighted rows are the user that were removed in the quantitative data analysis.

Age	Gender	Education	Computer Experience	Mouse or touchpad	HMD Experience	3D Software Experience	Data Visualization Experience	Space Time Cube Experience
18	Female	high school	advanced	touchpad	none	advanced	beginner	none
22	Female	high school	advanced	both	beginner	none	none	none
26	Female	bachelor	advanced	touchpad	beginner	advanced	intermediate	none
25	Female	bachelor	advanced	both	beginner	advanced	beginner	none
23	Female	bachelor	intermediate	both	beginner	none	beginner	none
21	Female	bachelor	advanced	touchpad	beginner	advanced	intermediate	intermediate
20	Male	high school	advanced	both	none	advanced	intermediate	none
23	Male	high school	advanced	mouse	none	beg	beginner	none
21	Male	bachelor	advanced	both	advanced	none	beginner	none
20	Male	high school	advanced	mouse	intermediate	intermediate	none	none
18	Male	high school	advanced	both	intermediate	intermediate	beginner	none
25	Male	high school	advanced	mouse	beginner	beginner	beginner	none
18	Male	bachelor	advanced	mouse	beginner	beginner	beginner	none
21	Male	associate	advanced	both	beginner	intermediate	beginner	none
25	Male	associate	advanced	mouse	none	advanced	beginner	none
20	Female	bachelor	advanced	touchpad	beginner	beginner	beginner	intermediate
21	Male	high school	advanced	touchpad	intermediate	beginner	beginner	none
28	Male	bachelor	intermediate	mouse	advanced	beginner	beginner	none
19	Male	high school	advanced	mouse	advanced	intermediate	beginner	none
24	Female	high school	advanced	mouse	advanced	beginner	advanced	none
21	Male	bachelor	advanced	both	beginner	beginner	beginner	intermediate
21	Male	high school	advanced	both	beginner	beginner	beginner	none
18	Male	high school	advanced	touchpad	beginner	beginner	beginner	none
21	Male	high school	advanced	mouse	beginner	intermeddle	none	none
25	Female	high school	advanced	mouse	beginner	none	advanced	intermediate

Table 2 continued

19	Male	bachelor	advanced	both	none	intermediate	none	none
22	Male	high school	advanced	mouse	none	intermediate	none	none
21	Male	high school	advanced	mouse	beginner	beginner	beginner	none
18	Female	high school	advanced	mouse	beginner	none	none	none
21	Female	high school	advanced	both	beginner	none	none	intermediate
34	Female	PhD	intermediate	mouse	none	none	none	none
31	Female	bachelor	intermediate	both	advanced	none	none	intermediate

4.2 User Performance

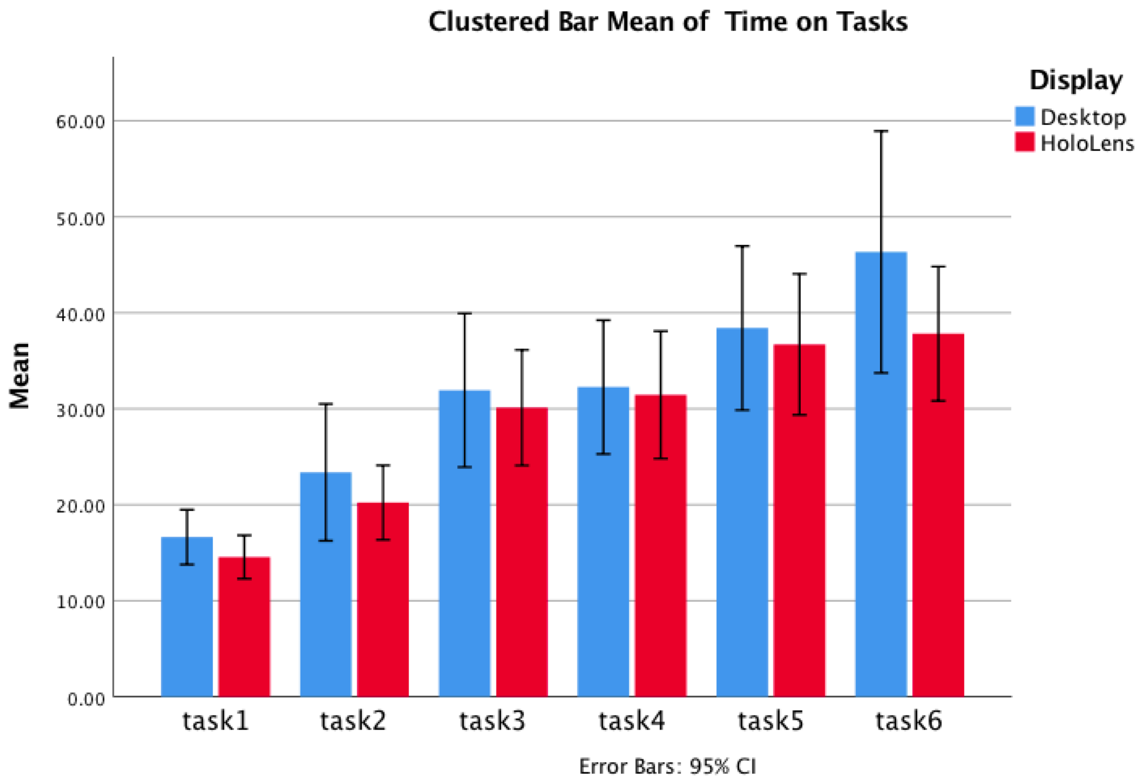
In this section, will be testing our null hypotheses and reporting the data using bar charts and error bars to show standard deviation. To check if there was a statistically significant difference a paired T-test and nonparametric test was conducted. The following are the null hypotheses used for testing users' performances.

There is no significant difference in time for tasks performed in a HoloLens and a computer screen.

There is no significant difference in error for tasks performed in a HoloLens and a computer screen.

There is no significant difference in perceived workload for tasks performed in a HoloLens and a computer screen.

.



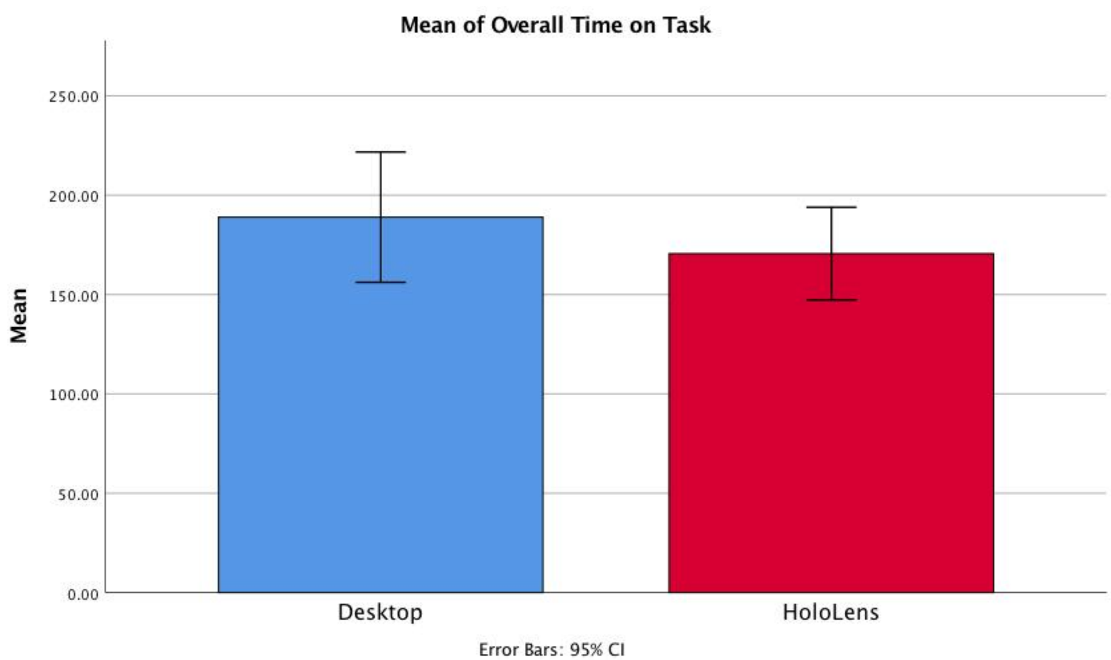


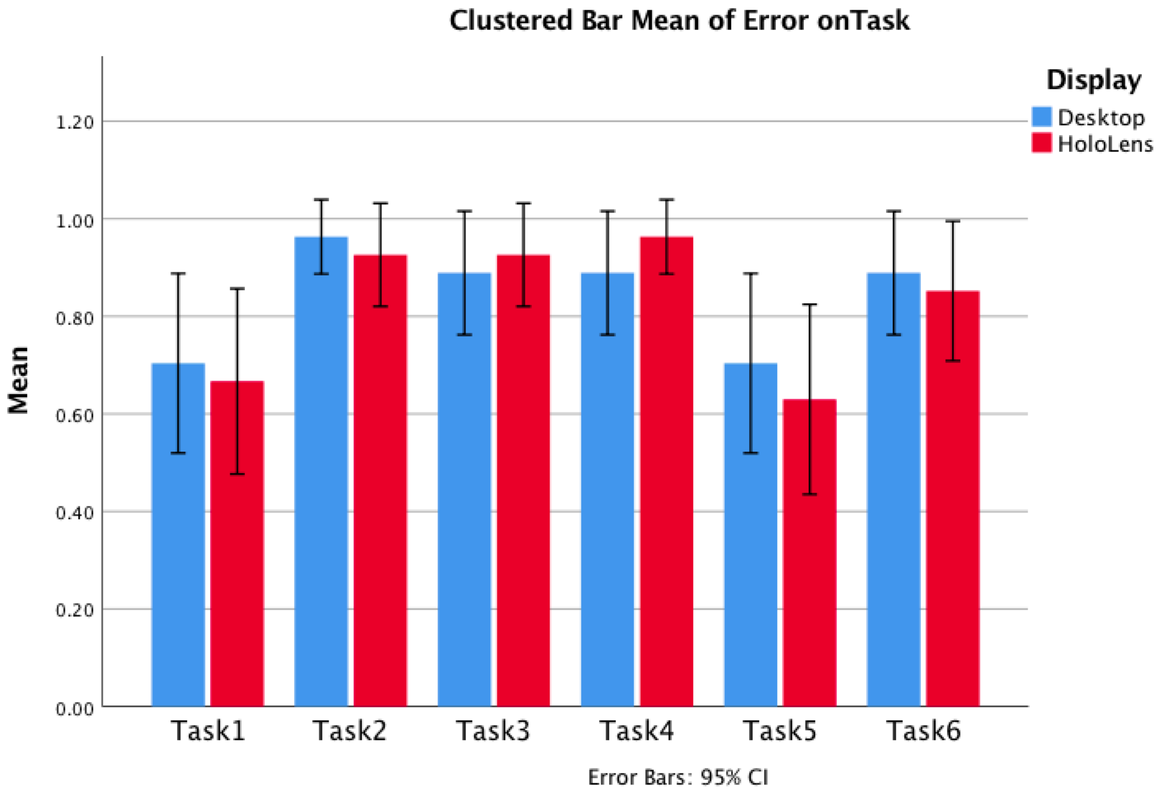
Figure 12: Demonstrates the means on time on task, with time given in seconds.

Table 3: Demonstrates the results for the T-test and Wilcoxon test on time on task, with time given in seconds.

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Overall
HoloLens	14.555	20.222	30.111	31.444	36.703	37.814	170.592
Desktop	16.629	23.370	31.925	32.259	38.407	46.333	188.925
P-Value	0.215	0.727	0.709	0.666	0.516	0.075	.156

First, the researcher checked the time on task data for normal distribution using the Shapiro Wilk test. It was found that task 1 was normally distributed but task 2,3,4,5 and 6 was not. Therefore, the researcher ran a T-test for task 1 and a Wilcoxon test for task 2,3,4,5 and 6.

Figure 12 exhibits the mean of time on task with a HoloLens compared to tasks performed with a desktop. Table 2 indicates that the p- value for all tasks is greater than 0.5. Thus, we fail to reject the null hypothesis for our first null hypothesis. Which means there is no statistically significant difference in time for all tasks. Figure 13 illustrates the mean of error on tasks performed with a HoloLens rivalled to tasks performed with a desktop.



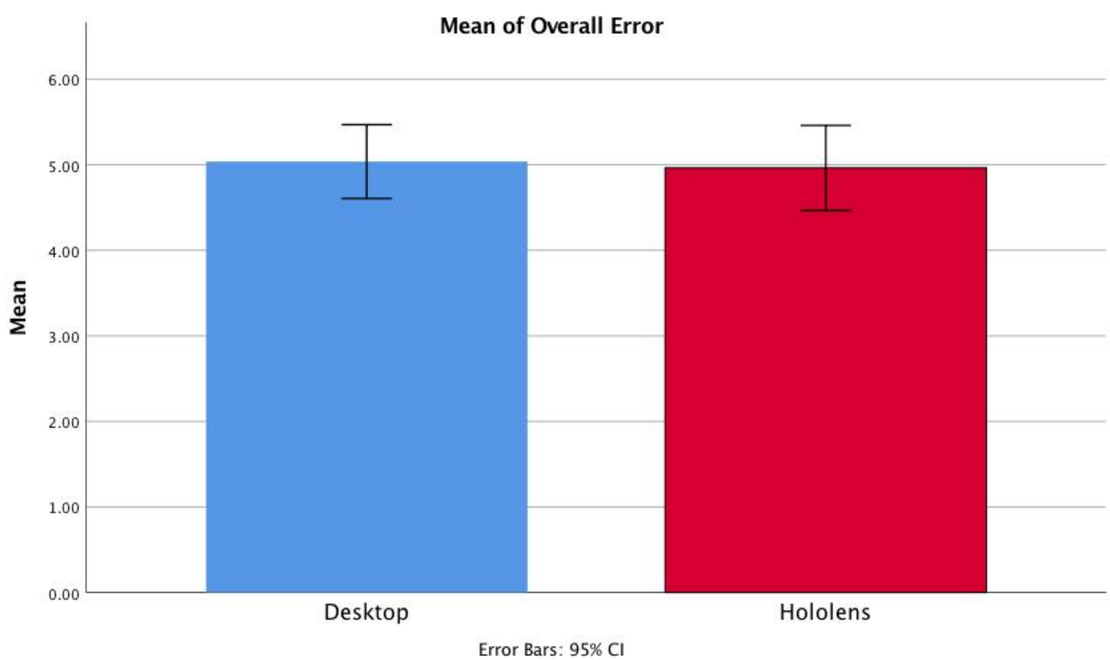


Figure 13: Demonstrates the means on error on task.

Table 4: Summarizes the results of the Wilcoxon test for error.

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Overall
HoloLens	0.666	0.925	0.925	0.962	0.629	0.851	5.037
Desktop	0.703	0.962	0.888	0.888	0.703	0.888	4.962
P-Value	.655	.564	.312	.157	.480	.317	.652

Likewise, the researcher checked the error data for normal distribution using the Shapiro Wilk test. It was found that no data was normally distributed. So, the researcher ran a Wilcoxon test for all tasks.

Table 3 signifies that the p-value for all tasks is larger than 0.5. Thus, we fail to reject the null hypothesis. Figure 14 shows the mean of workload on tasks performed with a HoloLens compared to tasks performed with a desktop.

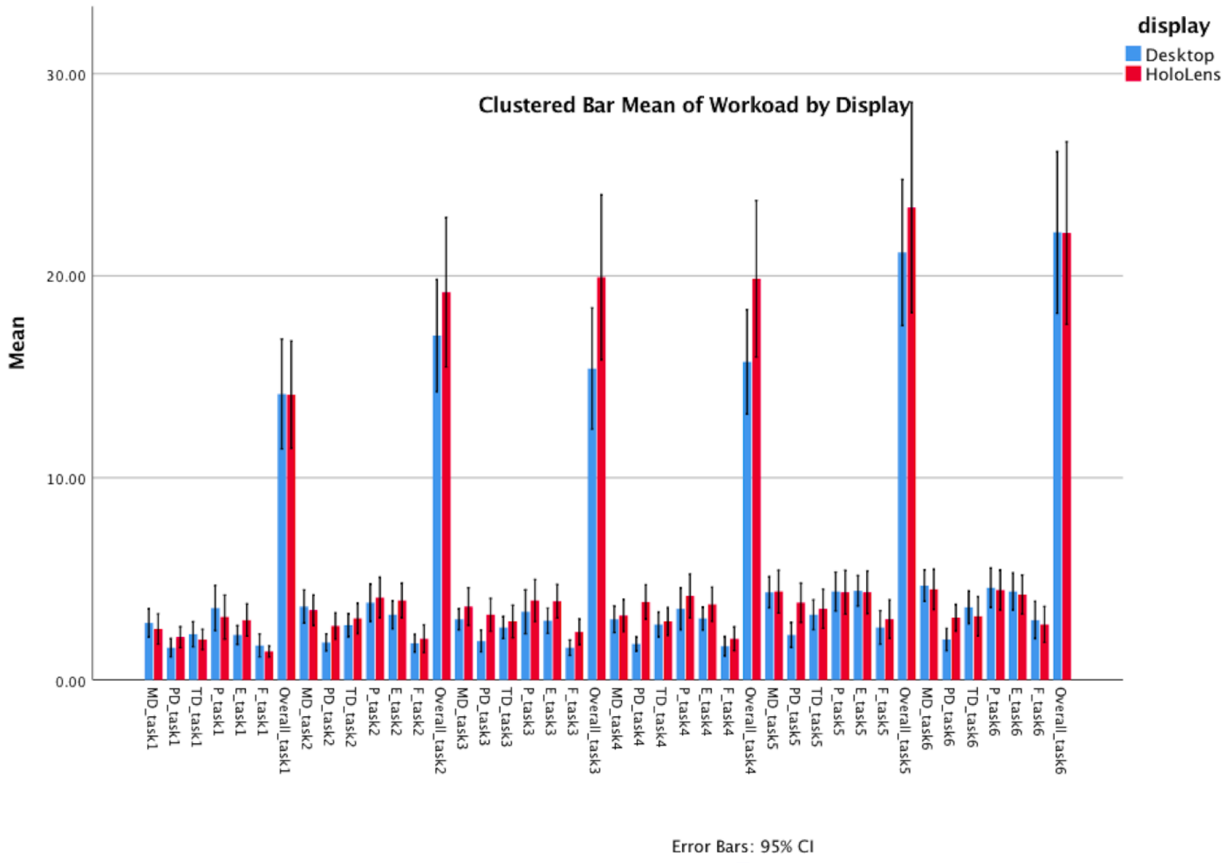


Figure 14: Demonstrates the means on Workload on task.

Table 5: Summarizes the results of the T-test and Wilcoxon test for Workload.

Task 1	Mental Demand (MD)	Physical Demand (PD)	Temporal Demand (TD)	Performance(P)	Effort(E)	Frustration(F)	Overall
HoloLens	2.518	2.111	2.000	3.111	2.963	1.403	14.111
Desktop	2.814	1.592	2.259	3.555	2.222	1.703	14.148
P-Value	.691	.019	.205	.087	.041	.399	.852

Task 2	Mental Demand (MD)	Physical Demand (PD)	Temporal Demand (TD)	Performance(P)	Effort(E)/	Frustration(F)	Overall
HoloLens	3.444	2.666	3.037	4.074	3.9256	2.037	19.185
Desktop	3.629	1.851	2.703	3.814	3.222	1.814	17.185
P-Value	.791	.015	.150	.829	.092	.641	.079

Task 3	Mental Demand (MD)	Physical Demand (PD)	Temporal Demand (TD)	Performance(P)	Effort(E)	Frustration(F)	Overall
HoloLens	3.629	3.222	2.888	3.925	3.888	2.370	19.925
Desktop	3.000	1.925	2.592	3.370	2.925	1.592	15.407
P-Value	.166	.003	.453	.226	.020	.023	.012

Table 5 continued

Task 4	Mental Demand (MD)	Physical Demand (PD)	Temporal Demand (TD)	Performance(P)	Effort(E)	Frustration(F)	Overall
HoloLens	3.185	3.851	2.888	4.148	3.740	2.037	19.851
Desktop	3.000000	1.777	2.740	3.518	3.037	1.666	15.740
P-Value	.672	.00001	.543	.097	.111	.226	.010

Task 5	Mental Demand (MD)	Physical Demand (PD)	Temporal Demand (TD)	Performance(P)	Effort(E)	Frustration(F)	Overall
HoloLens	4.370	3.814	3.518	4.333	4.333	3.000	23.370
Desktop	4.333	2.222	3.222	4.370	4.407	2.592	21.148
P-Value	.894	.003	.626	.893	.791	.295	.285

Task 6	Mental Demand (MD)	Physical Demand (PD)	Temporal Demand (TD)	Performance(P)	Effort(E)	Frustration(F)	Overall
HoloLens	4.481	3.074	3.148	4.444	4.222	2.740	22.1111
Desktop	4.666	2.000	3.592	4.555	4.370	2.963	22.1481
P-Value	.513	.002	.205	.578	.785	.659	.871

Moreover, the researcher checked the workload data for normal distribution using the Shapiro Wilk test. It was found that for the category overall data was normally distributed for task 2,3 and 4. Also, performance for task 6 was normally distributed for the other data it was not. Hence, the researcher ran a Wilcoxon test for all task that was not normally distributed and a T-test when the data were normally distributed. Table 4 demonstrates that the p-value for physical demand for task 1(Cohen's d=.419), tasks 2 (Cohen's d=0.589), 3 (Cohen's d=0.742), 4

(Cohen's D= 1.25), 5 (Cohen's D= 0.772) and 6 (Cohen's D= 0.696) is less than 0.5. Moreover, Table 4 determines that the p-value for effort for task,1(Cohen's D= 0.440) tasks 3 (Cohen's d= 0.516) is less than 0.5. Additionally, Table 4 reveals that the p-value for frustration for task,3. (Cohen's D= 0.582) is less than 0.5. Likewise, Table 4 demonstrates that the p-value for overall workload for task, 3 (Cohen's D= 0.498) tasks 4 (Cohen's S= 0.489) is less than 0.5. Thus, we reject the null hypothesis for the third null hypothesis. Which mean there is a statistically significant difference in perceived workload between the HoloLens and desktop.

4.3 Cross Analysis

For cross analysis we first check for normal data distribution using the Shapiro Wilk test if the data was normally distributed the researcher ran a two-sample t-test. However, if the data was not normally distributed the researcher ran a Mann-Whitney test. The reason for a two-sample t test is because we are comparing two groups to see if they are equal value and the statistical test to do so is a two sample tests for normal data distribution and a Mann-Whitney test for non-normal data distribution. Within this section we are only reporting the significant difference found within the test to see a full report see Appendix E. No statistically significant difference was found in time. A statistically significant difference was found in error. Group A (HoloLens first) did statistically significant better for HoloLens task 5(p- value .038) and overall HoloLens task (p- value .012). For workload no statistically significant difference was found.

4.4 Thematic Analysis

Within this sub-section the researcher is reporting themes and sub-themes of the thematic analysis. The themes are place into five categories table 5 show the theme found in the data analysis.

The researcher used all 32 users for the qualitative data analysis. 31 of the deferring videos were transcribed using Temi, a speech-to-text transcription service. The researcher then watched each video and fixed any errors in the transcription from Temi. One video was transcribed by hand rather than with Temi due to noise in the background of the video. The data was then analyzed using an inductive thematic analysis (see Chapter 3 for more detail on thematic analysis). Furthermore, when presenting the theme, the researcher included the amount of times the theme was seen in the data set a higher number does not show more significance of the theme, but it shows proof of data saturation.

Table 6: The themes and sub- themes found in this study

Categories	HoloLens (Themes)	Desktop (Themes)
Understanding	Object Mimicry	3D Reconstruction
Users Experience	Playful Exploring, Engaging	Controlled, Familiar
Perception	Depth	Flat
Interaction	Embodied (Natural), Familiar	Controller, Familiar
Hardware	How hardware affects the user experience.	

4.4.1 Understanding

The first category is how users understands the visualization. In this category, the sub-themes of object mimicry and 3D reconstruction encompass the users' strategy for trying to understand the presence of the 3D model. During the data analysis this theme was seen

4.4.1.1 Object Mimicry

The first theme is object mimicry. The users believed that the 3D visualization - in this case the space-time cube - was a "real" physical object. This caused the users to interact and experience the model physical object. During the data analysis this theme was seen 17 times.

> *"this is kind of sounding philosophical but, you can look at any 3D(model)in the HoloLens, when you put it on. The image .... it looks as if it's, (pause)it looks like this pen or kind of it just, it's an actual physical object that's in the room".* -P 13

4.4.1.2 3D Reconstruction

The theme of 3D reconstruction was seen in the desktop condition. Users had to examine each side of the visualization separately to help them recreate the visualization in 3D space. This was seen when users need to compare points in different planes. Users accomplished this in one of two ways: either by looking at every single plane of the space-time cube and combining each plane to recreate the 3D visualization, or users would create the cube in their hands for example, a user would place their left-hand flat and then count the number of clusters they can see in one plane of the cube then repeat to the other side of the cube. During the data analysis this theme showed up nine times.

> "*(The user is talking about desktop) I can't see this in 3D. I look at it, I have to like, almost like put the piece of pictures together, ., like each plate like actually is like I've seen this[gesture to the computer] and then I have to overlay that image with an understanding of what it looks like from the top down and it looks like from the other side."* -P22

4.4.2 Users Experience

The second category, user experience, describes how users felt and perceptions of system when performing tasks on the space-time cube. With the HoloLens users' experiences were playful exploring and engaging, while with the desktop they were controlled and familiar.

4.4.2.1 Playful Exploring

The HoloLens' superimposition of the visualization over the real world created a physical presence that contributed a playful quality to users' experience, similar to the feeling of a child making a discovery or exploring a new toy. During the data analysis this theme found 13 times.

> *" this was much more of almost like childish exploration, like actually like going out and exploring something because you get to like move. Whereas [in the desktop] the only time you are moving is your eyes,"* -P 5

4.4.2.2 Engaging

With the HoloLens, users reported and showed a greater sense of engagement. The physical act of walking caused them to care more about the data, and the superimposition of the visualization over the real world made them focus more on the data itself. During the data analysis this theme was seen 17 times.

> *"Interviewer: Do you have a preference between either of the displays?*

User: The HoloLens just because you know, it's pretty cool.
Interviewer: What makes its cool?
User: Yeah, because I mean it's, it's a very different way to view the data and it's much more interesting to me to view the data on the HoloLens.
Interviewer: Why is it interesting?
User: Its interesting because it's, I feel like it's more interactive than it is on the computer because it sort of super imposed the three d model in the room so that you have to walk around it to really look at it all the way.
Interviewer: So are you saying the super imposed nature forced you to interact with it?
User: Yes, of course forced it me to interact more than I did on the computer.
Interviewer: Do you find that as a good thing or a bad thing?
User:Uh, well it really depends on the situation I feel in this situation. It was a good. It made me care about the data, so a little bit more."-P 10

"the computer is more boring life and this is something new and exciting to me. So this[HoloLens] is more fun to play with and walk around rather than the computer. I'm just clicking on points or rotating with the mouse. So I'm actually physically involved like I feel more physically involved with this" -P 1

4.4.2.3 Controlled

The theme of control applies most strongly to the desktop. Users reported feeling more empowered and in charge of the data visualization, with no physical constraints preventing them from performing any desired interaction. The users felt like a conductor controlling the visualization. During the data analysis this theme was seen 12 times.

"the computer was actually easier to understand that visualization just because it was, I'm a lot more concrete, like it was sort of like the cameras in a fixed position and I can move around this camera and that gives me an easier view of what I'm looking at"- P 13

"HoloLens wasn't tightened up sometimes to get a little bit shaky, sometimes it gets a bit blurry, which to some extent it interrupting my judgement more brief while, but once it got used to it, it's no big deal. Um, I do feel. I felt more in control when on the desktop" -P 24

"desktop version I think was much easier to control and made me want to search a little bit harder to see if I have the right content to bring up."-P 25

4.4.2.4 Familiar

Users drew on past experiences using similar 3D modeling software or data visualization to assist them in interacting with the devices. This created an experience of familiarity for users interacting with the desktop, particularly for users who self-reported themselves as advanced 3D software users. During the data analysis this theme was seen 20 times.

> *"Computer feels natural, but that's only because I use computer every day., I'm kind of used to the, you know, once you explain this is how you zoom in, you know, how you scroll. This is all that. It's kind of natural [to] pickup because those are gestures, we use every day."* -P 10

4.4.3 Perception

The third category is perception. Different devices cause users to perceive the space-time cube differently, such as the desktop computer's flattening of the visualization to a 2D image.

4.4.3.1 Flat

Users perceived images generated by the desktop computer to be flat. This visual illusion is an inherent consequence of most computer screens' inability to create the perception of depth. During the data analysis this theme was seen 20 times.

> *"think that the desktop, it was like, it wasn't as intuitive, whereas with the HoloLens you put it on and the data gets set up for me and I just look around. Rather than clicking like you physically just look around like moving your head. I think so. It was just, it was just easier to see all the data are presented in a three d space that wasn't on like a one dimensional I guess because like if this is the three dimensional space, you walk around and see that it really is the three dimensions is rather flat on the computer"* -P 7

4.4.3.2 Depth

While using the HoloLens, users perceived the space-time cube as occupying their own 3D space. For example, while performing the task requiring them to find clusters, they were able to see different points with less rotation compared to the desktop. During the data analysis this theme was seen 22 times.

> *"But since that's 3D I can look at it from one side and I can tell if like this that is farther away from that dot because like, you know, if there's actual depth to it. Uh, so I didn't feel like I had to do as much rotating. Like if there was a cluster here[points], here[point], here[points] and like I could, I could see that this*

> *cluster was farther away without having to like to get on the other side of it and things like that."* -P 27

4.4.4 Interaction

This category encompasses the different styles in which users interacted with the space-time cube.

4.4.4.1 Embodiment

Interacting with this visualization came naturally to users of the HoloLens. They considered walking to be a natural way of interacting with the visualization, as it mimics strategies they would take when looking at a real object in physical space. During the data analysis this theme was seen 36 times.

> *"[The user was talk about interacting with the HoloLens] flexibility in terms of like how I can like what position in, in my, um, what's positioned in my field of view and like how I can control that really, really precisely in the HoloLens"* -P 22

> *"So, looking at the differences between the points, it's easier to walk into. Okay. , cause it's, it just feels more natural to be like, oh it's(the cat) here... He moved from here [*the user is tracking the pattern of the cat with his fingers] *to here and um, because you can kind of be like, oh it's like this long [uses he hand to show distance]"* -P 9

4.4.4.1.1 Spatial Awareness and Body Awareness

Spatial awareness and body awareness code is within the embodiment theme. Users were aware of the position of points relative to the physical space inside the room, which aided in recognizing patterns, collecting points and mapping out the points while changing positions. Also, users could treat their physical bodies as tools to assist with spatial perception this was seen in the observations when users would map out points between their fingers. For example, a user grasped a 3D point with their left hand and then touched another point with their right hand, making a line with their both hands to demonstrate the distance between them. This was a common occurrence while users performed tasks with the HoloLens.

> *"I just knew that in this space whether or not deserving or anything else, I knew that this block was here, and it was, it was staying there the entire time.... just special my spatial awareness. Knew that a point was there. It was always going to be there. I can just point points and that's all that counts. That's what I would do is because I was just like, all right, I pointed this point, this point at this point and then we're good to go. Um, so and then like I think the spatial awareness and just understanding like where each of the points fit in 3 d-space. And as I was a little more excited to kind of explore the different trends and stuff in that space as a space where I was just like, all right, I have to like visualize"* -P 22

> *"I was on the HoloLens, I could actually walk around it and I realize like, I guess like in my brain, know where these points actually are in relationship to the room, on the desktop, like as I scrolled out, maybe that like relationship would change because the scale is different now."* -P 26

4.4.4.2 Controller

During the debriefings, users discussed their perceptions of their control over the visualizations. In the desktop environment, they felt that without the limitation of their physical bodies they had greater control over their interactions. This theme came from operating the mouse and keyboard. During the data analysis this theme was seen 22 times.

> *"interact and look around the computer version because I was manually using a camera, I'm a controller to control where the camera was sort of facing"* -P 9

4.4.4.3 Familiarity(Interaction)

This theme contains users' mental model, or how they believe the interaction would behave. For some users it felt foreign; in other words, it did not match their mental models of how the visualization tool would work. For example, users found it foreign that they had to walk to interact with the device due to the fact that interaction with data visualizations is traditionally done with a mouse and keyboard. During the data analysis this theme was seen 20 times.

> *"I use the mouse all the time in my day to day activity, so that wasn't a burden or anything. But actually, moving around is. I'm not going to say it's foreign, but in a line of work with data visualization"* -P 4

Another example was seen with users unfamiliar with using 3D software. They found it foreign to press multiple buttons at once, describing it as a cognitive burden due their unfamiliarity with tools having similar controls.

> *"thinking about, oh, I gotta hit alt and that drag it with whichever button you just take a few steps and looking at all. That's not the [key]. Or is that the[key] ?"*
> -P 7

4.4.5 How Hardware Affects the User Experience.

This theme encompasses the way hardware limitations such as low color saturation or a small field of view could hinder users in completing a task. The weight of the HoloLens affected some users even in a 30-minute study. Users complained about pressure in their nose, sometimes focusing on the physical discomfort instead of the task at hand. Another hardware element that affected the user experiences was the field of view. Because the HoloLens has a small field of view, users felt uneasy if they moved their heads outside the field of view or got too close to the data, causing it to disappear. During the data analysis this theme was seen 29 times.

> *"with the desktop you can see the entire model all there, but through the HoloLens you have a bit of a, a small field of view. It kind of makes it a little more difficult to get an idea of the bigger picture. You have to look up and down and make sure you've seen everything while on the desktop. You just had access to that view."(P 8)*

Also, users with ocular conditions reported that having the screen directly in their face was overwhelming and distracted or hurt their eyes.

> *" Yeah, so I have eyesight problem . . . within a certain, like certain ranges with my eyesight, they aren't great . . . I don't know, like, well I'm thinking with that HoloLens on I have like I don't know why, what it is like having the screen like constantly in my face it like makes me like kind of feel a little overwhelmed with that category."* -P 21

Some users felt that the low saturation of the model produced by the HoloLens broke the illusion of it being a real 3D object and made it more difficult to see.

"That, HoloLens isn't clear but just the coloring is bad ...and its faded. You can read everything clearly but it's just not as vibrant as it." -P 3

"I would probably just use the computer because it's a clear image." -P 23

This chapter discusses the findings of the study. The quantitative data showed no significant difference in the measures of time and error but there was a significant difference in workload. For the qualitative data five category emerged: Understanding, Users Experience, Perception, Interaction, Hardware. The next chapter will cover a discussion of the results

CHAPTER 5. DISCUSSION

The researcher's goal was to compare users' holistic performance using the HoloLens and the desktop computer for 3D data visualizations. The insights from the study could be used to improve decision-making when selecting hardware to analyze a space-time cube visualization. This chapter will cover the discussion of the conclusions that can be made based on the statistical and thematic analyses. It will also discuss important observations made during the study, provide a practitioner takeaway, and then propose future work and limitations of the study. The chapter will end with a conclusion for the study

5.1 What are the Differences in Task Performance on a Space-Time Cube Between HoloLens Users and Desktop Users?

To answer research question number one, the researcher ran 3 two-tail T-tests or Wilcox test with 95% confidence to see if there was a significant difference in the amount of time, error and workload involved in completing tasks for each display condition. In the quantitative data analysis we found significant difference in overall workload for task 3 and task 4 this is interesting because past literature would suggest that the ability to externalize user cognition and use of the environment could possibly lower the cognitive need (Wilson, 2002). Furthermore, if we look at Baumeister et al. (2018) they found that the HoloLens increases user's performance and decreases the cognitive load difference compared to a desktop. The difference between Baumeister et al study and this thesis is within their study design users sat down in a chair while doing all task and in this thesis, it required users to stand and walk around in the HoloLens condition. Thus, there's a possibility if users sat down the study could find different results this is something that the researcher wants to further explore in a later study. Moreover, Merino, Bergel, Nierstrasz, found that 3D city visualizations displayed in the HoloLens are beneficial in finding patterns (2018). Similarly, in this thesis we also found marginal significance in finding patterns task within the HoloLens condition. Does increasing the possibility of knowing if the HoloLens is useful in finding patterns compared to the desktop.For ease of understanding, the remaining section will be broken down into three sections focusing on one hypothesis each.

5.1.1 Hypothesis 1: There is a Significant Difference in Time for Tasks Performed in a HoloLens and a Computer Screen.

The data showed with a 95% confidence interval that there was not a significant difference between the amount of time it takes to complete a task in the HoloLens and desktop, but there was a marginal significance for task 6 (p value .075). Looking at the mean of time, the HoloLens was faster for task 1 (2.07 sec), task 2 (3.14 sec), task 3 (1.81 sec), task 4, (.81 sec) task 5 (1.70 sec) and task 6 (8.85 sec), but the disparity between these values are not significant. It should be noted that in task 5 in the HoloLens condition, some users walked around the object multiple times before making a decision, which could have affected the result of time on task. The HoloLens and desktop take about the same amount of time to finish a task, so it is up users to decide which one to use. In this study, 51.85% self-reported the HoloLens as easier to use and 59.25% said they preferred the HoloLens.

5.1.2 Hypothesis 2: There is a Significant Difference in Error for Tasks Performed in a HoloLens and a Computer Screen.

The statistics showed with a 95% confidence interval that there was no significant difference in the amount of error involved in task completion between the HoloLens and desktop. If we look at the mean of error, the computer was better in task 1 (.037), task 2 (.037) task 5 (.074) and task 6 (.037) but the HoloLens was better in task 3 (.037) and task 4 (074).

5.1.3 Hypothesis 3: There is a Significant Difference in Perceived Workload for Tasks Performed in a HoloLens and a Computer Screen.

The data showed with a 95% confidence interval that there was a significant difference in the amount of perceived workload involved in task completion between the HoloLens and desktop. Overall, the HoloLens had a higher mean for all perceived workload significance. The self-reported physical demand involved that using the HoloLens was significantly greater for task 1, task 2 , task 3, task 4, task 5 and task 6. Moreover, there was a significant difference with 95% confidence for effort in task 2 and 3. There was also a significant difference with 95% confidence for frustration in task 3. Furthermore, there was a significant difference with 95% confidence for overall workload in task 3 and 4.

Thus, we can say that the HoloLens had a higher self-reported physical demand for tasks 1-6 compared to the desktop, the HoloLens along with higher self-reported frustration compared to the desktop for task 3. Also, we can say HoloLens had a higher self- reported effort in task 2 and task 3. Additionally, we can say HoloLens had a higher self- reported overall workload for task 3 and task 4. (see Table 4 for means of all task)

The cross analysis showed a statistically significant difference in error. Group A (HoloLens first) did statistically significant better for HoloLens task 5(p- value .038) and overall HoloLens task (p- value .012). This is interesting because a significant learning affect didn't happen. There's a possibility that users excitement can explain this phenomenon because for many users this was the first time seeing the HoloLens. Also, there's a possibility that user fatigue could explain this phenomenon as well because Group B(desktop first) users were tired after doing all the task in desktop

5.2 What are the Experiential Differences using the HoloLens Compared to the Desktop as Reported by Users?

To answer research question number two, a thematic analysis was performed to find the experiential differences in using the HoloLens vs. the desktop. In the analysis, the researcher found 5 themes for experiential differences along with 6 sub themes for the HoloLens and 4 for the desktop (see Table 4.4 for the themes). The five themes are: How users observe the model, Users' Experience, Perception, Interaction and Hardware. The User experience and Interaction themes are very similar, but User experience pertains to users' feelings and perceptions of system while interacting with the visualization while Interaction pertains to the act of performing of the task. This section will discuss the impact of each of the five themes and how they relate to previous work.

5.2.1 Understanding: *Object Mimicry vs. 3D reconstruction*

These themes include users' strategies for understanding the 3D visualization. Because the HoloLens used object mimicry to make users perceive the space-time cube as a physical model, users adopted strategies they already had for interaction toward physical objects. During the debriefing, the researcher noticed one user acting on this strategy. When asked for the rationale behind their strategy, they responded as follows.

> *"Well, I think that's how I deal with the physical things. So when I'm looking at my phone and my wants to something closer, I'll zoom in like that (moves face closer to the phone)." (P 4)*

The theme of object mimicry aligns with Pham, Vermeulen, Tang, Vermeulen's theme of physicality (Pham, Vermeulen, Tang, & MacDonald, 2018). Their study, which focused specifically on interaction design, also found that users interacted with virtual objects in the same manner as physical objects. This type of theme is also seen in field of virtual reality, where researchers are focusing on presence - specifically spatial presence. Spatial presence is the user's perception of a virtual object in the environment as a physical object, a condition commonly tied to the user's immersion within the virtual environment (Dwyer et al, 2018). The difference between object mimicry and spatial presence is that the latter focuses on the recreation of its virtual object in the real environment.

The second observed theme was 3D reconstruction. This was seen in the way computer users reconstructed the 3D visualization. They did this by looking at every face in the 3D visualization individually similarly to a spatial awareness test. One of the main differences between using a HoloLens and desktop is the required spatial reasoning abilities needed to process the visualization using these display conditions. I would suggest that this be examined in future works by testing users' spatial abilities beforehand and using the results to sort them into groups. Because there's a possibility that users may need less spatial awareness skills to understand a 3D visualization since the 3D visualization mimics a physical object assuming that 3D physical objects require less spatial awareness.

5.2.2 Users Experience: *Playful Exploring and Engaging vs. Controlled and Familiar*

During the testing, interactions with the HoloLens demonstrated the themes of playfulness and engagement. The superimposition of the visualization over the real world created a physical presence that contributed a playful quality to users' experience. This was observed in users standing on chairs, sitting on their knees, laying on their backs, and such. Users took advantage of the physical space around them to explore the data visualization in more detail. This ties into the second observed theme, engagement. The researcher found that users were more observant of the data in the HoloLens condition. These two themes could possibly force the user to walk around to understand the data. Something to keep in mind is that these two user experiences might be connected to the novelty effect. It is possible that users treated the experience like playing with a new toy because they had not previously used augmented reality for 3D visualization. Well there's a possibility that the novelty could have play a factor within these themes it could also be argued that the inherent physical nature of the HoloLens possibly explains these themes as well and not the novelty.

The theme of control was a consequence of users feeling more empowered while using the devices. This was particularly notable in the desktop environment, where users were not limited by any of the physical constraints of the real world or the HoloLens hardware, and reported being able to achieve greater freedom and precision in their interactions. This could be at least partially a result of hardware issues associated with the HoloLens; for example, if a user gets too close to the data, a clipping plane will eclipse the visualization. The desktop may also more closely align with most users' mental model of the required interactions. This could be beneficial to user experience, with users having the ability to draw on their past knowledge as a reference for how to interact with this new software on a familiar device.

The two sets of sub-themes for the devices are beneficial for analyzing data visualization However, it can be surmised that engagement and playfulness would make the HoloLens more suited to understanding the rough outline of 3D visualizations, while the control afforded by the desktop would make it ideal for studying models in closer detail.

5.2.3 Perception: *Depth Vs Flat*

Users of the HoloLens perceived actual depth between the points of the 3D visualization, while desktop computer users perceived the space between the points of the 3D visualization to be flat. These results are expected; from the table of different depth cues (Marriott et al., 2018)AR head-mounted displays have additional depth cues compared to desktop computers. For example, accommodation, which is "how objects that are not focused on appear proportionally blurry in relation to the distance to the current focus depth"(p. 31), binocular disparity and stereopsis which shows "small disparities in the images received by the left and the right eye are processed in the brain to interpret depth and the 3D shape of objects."(p. 30) and subjective motion "related to controlled points of view in that the viewer can change the perspective of the scene";(p.32). However, in subjective motion the transformations are caused by the user's movement and not automatically in the eye. (Marriott et al., 2018) This information complements the user's visual cues to give depth to the space-time cube visualization. Because the computer lacks these depth cues, the images it produced were perceived as flat. For example, participant one said.

"For the computer, I can still get the sense of like where everything's at, but I feel like I can get a better sense of where everything's located looking at it in the 3D model{HoloLens} because on the computer screen I'm looking at flat image over... So like I can look at this point and I can see from this angle what's further away from me and [points in the space time cube] what's not. Looking at say, just like the starting at points on the computer. I see really just dots that I needed to rotate around it to be able to see where everything's actually located"-P1

Other users said the that depth helped them understand the space time cube.

"Researcher: Do you have a preference between either of the displays?
User : HoloLens
Researcher Why?
User : Its a more personal reason I am not good in [understanding} 3D space visual in the 2D visual [display] so it presented to me in a 3D object I can see it clear and all is fine -P10

Both display methods have their merits; depth could help the user perceive volume or find patterns and clusters in 3D space, while a flat perspective could be used to order points along one dimension. Both themes influenced the way users interacted with and understood the 3D visualizations; consideration should be given to the best method of perception when determining tasks for a study and the best device for completing them.

5.2.4 Interaction: *Embodied (Natural) vs. Controller*

The primary interaction styles observed in this study were embodiment for the HoloLens and controller for the desktop. Embodiment was exhibited in the way users physically navigated the space-time cube, for example by walking to the side of the cube they wanted to examine or macro physical navigation, by moving one's head to the side. Physical navigation received mixed reviews from users, who reported it being intuitive but likely difficult to maintain for extended periods of time. Conversely, virtual navigation was the primary method of interaction with the desktop, where users could maintain control while sitting using controllers rather than having to move around.

Again, each method of interaction has both advantages and disadvantages. Embodiment involving physical navigation could enhance the user's sense of engagement with the space- time cube visualization but does place physical demands on the user that may make the device unsuitable for sustained use.

Similarly, others have found that embodied base interactions can have a higher sense of engagement than a mouse and keyboard interaction (Lindgren et al ,2016). Control using virtual navigation makes it easier for users to control every aspect of the interaction without being limited by their physical capabilities, but with the qualification that interaction sacrifices intuition and requires skills that must be learned.

5.2.5 Hardware: *How Hardware can Affect the Experience*

The limitations of the hardware of the HoloLens had impacts on users' experience. The narrow field of view prevented users from observing the entire 3D landscape at once. Additionally, at certain angles, physical pressure on users' heads and noses affected their ability to think after as little as 20 minutes. Despite the high performance of the hardware, the physical hardships it afflicts on users may mean it would not be ideal for analyzing data over extended periods of time. Beitzel, Dykstra, Toliver and Youzwak found that after 30 minutes the weight of the HoloLens made continued use difficult for some users (2018). This suggests that the device is not a practical solution for extended periods of analysis.

5.2.6 Familiar

Users' interactions with the applications was influenced by their expectations of how it should work based on their past experiences. Some users found the HoloLens familiar, drawing on past experiences interacting with physical 3D models, but others found it foreign because it didn't match their mental models of working with 3D visualizations, which for them lacked a physical representation. The same was true of the desktop; users who had previously worked with Maya or other 3D modeling software or were familiar with the device reported it feeling familiar and tended to perform well on the tasks, while others unused to looking at objects in 3D space or manipulating a viewport using a mouse and keyboard found the interaction foreign. This theme could have been a factor with in the entire study. Currently we captured demographic information that shows users' familiarity with augmented reality-based devices and data visualizations. This information can be used to analyze performance data in future studies.

5.3 What Aspects of Users' Experiences Potentially Influence Task Performance?

In this section, the researcher amalgamated the data to understand the quantitative results in terms of the qualitative themes found. Observations and quotes from transcriptions will be used to explain what factors could have affected the quantitative results. There is a significant difference in perceived workload, specifically physical demand, frustrations, effort and over all workload. These findings seem to relate to the theme of hardware. The users' physical limitations and the design of the HoloLens likely affected the users' experience. For example, the weight of the HoloLens made it quite taxing to use even after only 30 minutes, which could increase the physical demands placed on the user while performing a task. Regarding the fact that the HoloLens required users to physically walk and move their arms to interact with the model, a user compared the physical demand of both devices by stating the following:

> *"I could stay in my chair and literally just moved my mouse for one of them (Desktop), but I had to move my whole body for the HoloLens."* -P 10

Another user stated that the physical demand was greater, but it was more enjoyable to use than the desktop.

> *"physical demand, this was harder physically, but it was more fun so it kind of outweighed itself."* -P 4

Though many users described it as fun, they also took note of the pain caused by the weight of the HoloLens.

> *"it was, it was quite enjoyable, peaceful, and manipulate the decision based upon my own movement and with the body. That was fine. But again, just when it came to like having to move the neck and the head constantly with the weight of the HoloLens [the user made a sad face]"* - P 15

The physical demand could also have affected the error rate if the weight dissuaded users from moving around to examine the visualization.

> *"HoloLens was very heavy in my head and it made it hard to scan the whole visualization. I didn't feel compelled to walk around and see this visualization from other side and then the desktop version I think was much easier to control and made me want to search a little bit harder to see if I have the right content." - P 6*

Users described feelings of frustration at being unable to see the model in their field of view, which likely contributed directly to the high frustration scores reported by HoloLens users. For example, one task required users to determine the easternmost and southernmost points in the data; while trying to compare the locations of points, users reported the points disappearing from their field of view. Baumeister similarly found that a small field of view in the HoloLens can increased cognitive load of a user. (2017)

People who preferred the HoloLens cited embodied interaction and depth as their rationale.

> *"I think so. It was just, it was just easier to see all the data are presented in 3D space that wasn't on like a one dimensional I guess because like if this is the 3D space, you walk around and see that it really is the three dimensions is rather flat on the computer. My scenes like, you could really see the depth"* -P 7

Those that preferred the desktop cited the control they had with a mouse, increased field of view, lack of pressure on their heads and past familiarity with that type of device as their reasons.

> *"Uh, generally, um, when comparing to the HoloLens, like I feel a little bit more comfortable with the desktop, probably more familiar with it. Um, but uh, yeah, that a narrow, like a lens of vision. Um, it, um, just because I felt like I couldn't see all the data points at one time which was. So then when you're asking questions, especially to compare data points are like, once, you know, when there was like a range or like I see you're asking for like a spectrum to see some on this side. And then on this side it was a little bit more labor intensive. Um, uh, yeah. So, it was a little bit more familiar with the desktop. Oh. And then it just, it's more comfortable to see the desktop. Like I felt like without the goggles I was, um, it was kind of straining or my vision a little bit." -P 32*

In this section I tried to converge the date of from the qualitative and quantitative study to see if they could inform us from each other.

5.4 Observations

This section will include observations made during the study the researcher found interesting. During the study, HoloLens users commented on the fact that their interpretations were limited by their physical capabilities. For example, hovering over the model would be easy in the desktop environment but difficult in physical space. While this did not limit the correctness of the task for this specific study, this could have a factor in other studies. The researcher also noticed that some HoloLens users seemed reluctant to move or walk around, staring at the visualization from one corner the entire study. What was interesting, was that they were able to answer every single question correctly, whereas desktop users had to rotate around the model to solve some of the questions. Other HoloLens users made use of the full physical space of the room, getting on chairs or on their backs on the floor to view the model from different angles.

The researcher noted that task 1 had a high rate of error because users did the math incorrectly for both conditions or would look at the middle points and not the lowest point in the graph. Task 3 and task 4 saw an increase in overall workload. This could be because these tasks reintroducing micro navigations forced HoloLens users to move their head to solve them and inspire the HoloLens users to physically walk around to find the range of latitude and longitude and locate points in space. This could explain the higher workload score, as this task is inherently more physically demanding than sitting at a computer. For task 5, users found clustering easier and more enjoyable in the HoloLens than the desktop this could possibly related to the depth theme, but the researcher noted that users were in a hurry which could cause the low error score. Most users correctly completed task 6, which required them to simply describe the pattern of the cat; after completing the other tasks, most users felt more comfortable doing this. It should be noted that in tasks 6, there was no significant difference in time between devices in the 95% range, but the opposite was true for the 90% range. The researcher observed that the HoloLens users would trace the cat's path with their finger before stating their answer. The phenomenon could be caused by the user off-loading their mental demand into the environment, which is one of the views of embodied cognition (Wilson, 2002). The researcher noticed that users found the desktop is easy to navigate and observe micro patterns, while the HoloLens was useful for gathering more detailed information of 3D space and finding macro patterns.

> *"I would say it's easier to, you just want to have the rough detail. It's pretty easy to use a computer monitor, but if I use the HoloLens is more like a, a much more details then using the monitor and I can uh, just move to check which part of the information I need."* -P 28

This observation could possibly be related to dept cues. Users also felt that the HoloLens was intuitive and easy to learn, while using the desktop involved more skill.

> *"so it was a very user friendly for me, but that's because that's the area I'm in. Whereas this [HoloLens], I could see anybody had messing around with it. Like I didn't need a lot of training."* -P 4

5.5 Practitioner Takeaways

This section will state practitioner takeaways for this study. First, users took approximately the same amount of time to finish tasks on either device, so it is up to user preference to dictate which device to use. The amount of error was also comparable between the two devices. Second, the physical demands placed on HoloLens users may be too high for extended use. After the 20 to 30 minutes study, most users commented on the weight of the system. Applications designed for the HoloLens should have an interaction loop no longer than twenty minutes. The users who felt that the HoloLens was good for exploring would describe the HoloLens as a playful and inherently suitable application to help people do exploration. The HoloLens' object mimicry capabilities cause users to perceive objects in the viewfinder as real. This would make the device ideal for previewing a product in virtual space before people actually build it to see how it would look or work in physical space for example physical data visualizations.

5.6 Future Work

Future work for the study should focus more on evaluating the 3D visualization with different metrics like recall and engagement. Additionally, as display hardware continues to improve and reduced in weight, it may be beneficial to study how these advances affect user experience and workload. Furthermore, the researcher can bring in a third display condition such as VR to see if different depth cues affect the tasks performance. Moreover, the researcher wants to investigate if object mimicry within the HoloLens helps users with low spatial awareness understand the 3D visualization. Furthermore, the researcher would like to explore if personality and cognitive styles could affect people perception of the HoloLens?

5.7 Limitations

This study is limited to using only two devices, a desktop computer and a HoloLens and no other immersive displays. Additionally, the data presented was coded by only one person; while methods were used to ensure variability in the data, this limitation should still be considered. Furthermore, self-reported workload was used and not measurable workload. Moreover, all participants came from one location and were not distributed form multiple locations. The user had the options to fill out the NASA TLX survey verbally or written. This could be an issue because they could switch at any time in the study. The researcher feared the user could get dizzy filling out the survey with the HoloLens on. Another limitation observed is that users were allowed to self-report their experience level instead of gauging their experience levels by a test.

5.8 Conclusion

Within the study the researcher investigated performance between a HoloLens and desktop for time, error, and workload. The quantitative data showed no significant difference in the measures of time and error but there was a significant difference in perceived workload. For the qualitative data, 11 themes emerged: object mimicry 3D reconstruction ,playful exploring, engaging, controlled, depth, flat embodied (natural), familiar, controller, and how hardware affects the user experience. There is a possibility that the theme of hardware could have affected the user's performance. The researcher also reported important observations made during the study, provide a practitioner takeaway and limitations of the study.

APPENDIX A. SURVEY

Figure 8: NASA-TLX RATING SCALE DEFINITIONS

Title	Endpoints	Descriptions
MENTAL DEMAND	*Low /High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	*Low /High*	How much physical activity was required (e.g.. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous. restful or laborious?
TEMPORAL DEMAND	*Low/ High*	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	*good/poor*	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	*Low/High*	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL*i*	*Low /High*	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Mental Demand

1 2 3 4 5 6 7 8 9 10

Physical Demand

1 2 3 4 5 6 7 8 9 10

Temporal Demand

1 2 3 4 5 6 7 8 9 10

Performance

1 2 3 4 5 6 7 8 9 10

Effort

1 2 3 4 5 6 7 8 9 10

Frustration level

1 2 3 4 5 6 7 8 9 10

APPENDIX B. TASK

Retrieve Value
How long (time) did the cat travel?

Filter
In what location (latitude and longitude) did the cat spend the most time?

Find Extremum
How far east did the cat go? Identify the specific point.
How far south did the cat go? Identify the specific point.

Determine Range
What is the range of latitude/longitude in which the cat traveled?

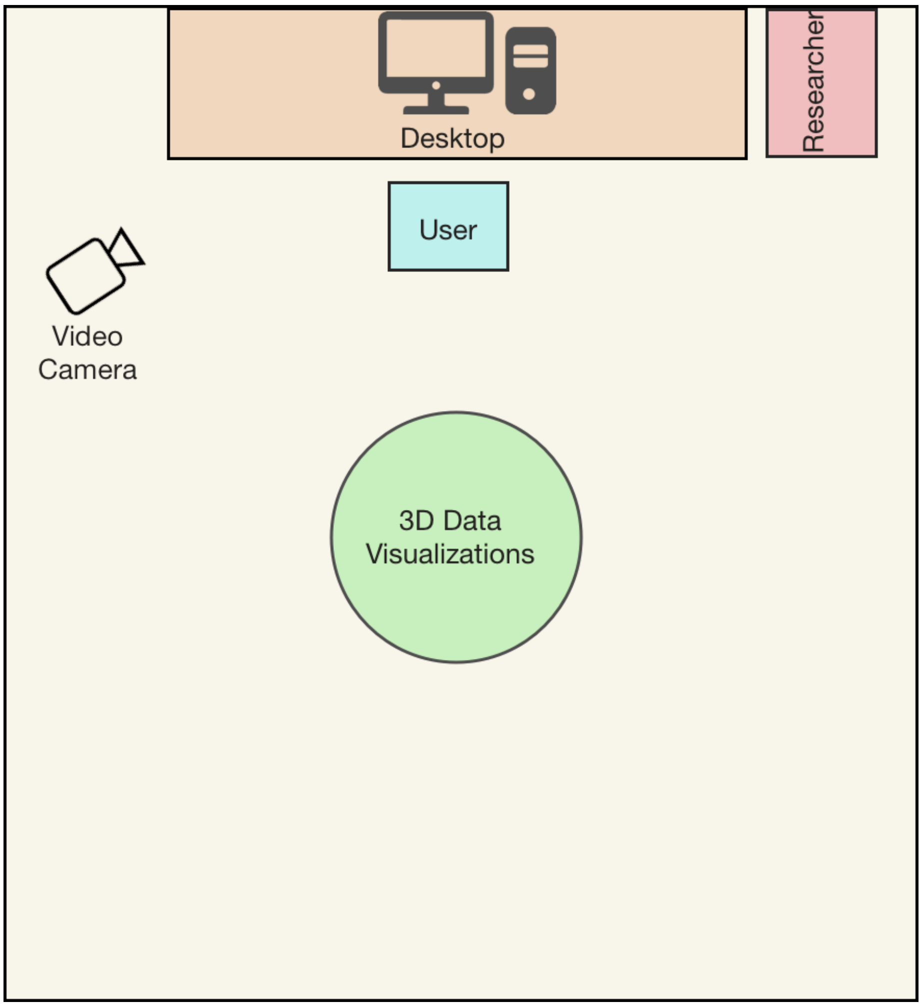
Sort and cluster
Are there clusters in time?
Can you rank the clusters the from biggest to smallest? By number of points.

Pattern
Is there a clear pattern in the visualization?
Can you provide a summary of content within the visualization?

APPENDIX C. DEBRIEFING

How was your overall experience?
I notice blank In A and not in B...
During the think aloud you said about A but not B...
Can you tell me about how you experienced both displays? (then ask these questions only if they are having a hard time with it: _How would you compare your experience of using both displays?_; How did you feel about the interaction for both displays?_)
How did you feel about the 3D model and how you could view it through both displays?
Do you have a preference between either of the displays? Why?
Was one of the displays easier to use? How so?
Can you see yourself using this if you had to analyze this type of data ?

APPENDIX D. LAB DIAGRAM



This is an example on how the lab was set up during the study. Something to note the camera was turned to face the 3D data visualization during the HoloLens conditions

APPENDIX E. CROSS ANALYSIS REPORT

Time	Desktop Task 1	HoloLens Task 1	Desktop Task 2	HoloLens Task 2	Desktop Task 3	HoloLens Task 3
Desktop Frist	18.3846	14.9231	23	21	36.2308	30.3077
Hololens Frist	15	14.2143	23.7143	19.5	27.9286	29.9286
P- value	0.253	0.846	0.319	0.559	0.576	0.884

Time	Desktop Task 4	HoloLens Task 4	Desktop Task 5	HoloLens Task 5	Desktop Task 6	HoloLens Task 6	Desktop OverallTasks	HoloLens OverallTasks
Desktop Frist	33.6923	26.8462	41	36.3846	45.3077	39.5385	169	197.6154
Hololens Frist	30.9286	35.7143	36	37	47.2857	36.2143	172.0714	180.8571
P- value	0.451	0.151	0.297	0.808	0.942	0.56	0.961	0.452

The tables above show the result for the cross analysis for time, desktop first is Group B and HoloLens first is Group A.

Error	Desktop Task 1	HoloLens Task 1	Desktop Task 2	HoloLens Task 2	Desktop Task 3	HoloLens Task 3
Desktop Frist	0.6154	0.5385	1	0.8462	0.8462	0.8462
Hololens Frist	0.7857	0.7857	0.9286	1	0.9286	1
P- value	0.342	0.181	0.335	0.134	0.504	0.134

Error	Desktop Task 4	HoloLens Task 4	Desktop Task 5	HoloLens Task 5	Desktop Task 6	HoloLens Task 6	Desktop OverallTasks	HoloLens OverallTasks
Desktop Frist	0.8462	0.9231	0.6923	0.3846	0.8462	0.7692	4.3077	4.8462
Hololens Frist	0.9286	1	0.7143	0.8571	0.9286	0.9286	5.5714	5.2143
P- value	0.504	0.299	0.902	0.013	0.504	0.253	0.008	0.267

The tables above show the result for the cross analysis for error, desktop first is Group B and HoloLens first is Group A.

	Desktop Task 1 Mental Demand	HoloLens Task1 1Mental Demand	Desktop Task 1 Phyical Demand	HoloLens Task 1 Phyical Demand	Desktop Task 1 Temporal Demand	HoloLens Task 1 Temporal Demand
Desktop Frist	2.6154	2.3077	1.3077	1.6154	2.3077	2.1538
Hololens Frist	3	2.7143	1.8571	2.5714	2.2143	1.8571
P- value	0.96	0.335	0.541	0.113	0.684	0.406

	Desktop Task 1 Performance	HoloLens Task 1 Performance	Desktop Task 1 Effort	HoloLens Task 1 Effort	Desktop Task 1 Frustration	HoloLens Task 1 Frustration	Desktop Task 1 Overall	HoloLens Task 1 Overall
Desktop Frist	3.0769	2.0769	2.3077	2.9231	1.6923	1.2308	13.3077	12.3077
Hololens Frist	4	4.0714	2.1429	3	1.7143	1.5714	14.9286	15.7857
P- value	0.646	0.12	0.959	0.601	0.777	0.251	0.808	0.113

	Desktop Task 2 Mental Demand	HoloLens Task 2 Mental Demand	Desktop Task 2 Phyical Demand	HoloLens Task 2 Phyical Demand	Desktop Task 2 Temporal Demand	HoloLens Task 2 Temporal Demand
Desktop Frist	3.9231	3.1538	1.8462	2.4615	2.9231	3.0769
Hololens Frist	3.3571	3.7143	1.8571	2.8571	2.5	3
P- value	0.291	0.22	0.958	0.861	0.616	1

	Desktop Task 2 Mental Demand	HoloLens Task 2 Mental Demand	Desktop Task 2 Phyical Demand	HoloLens Task 2 Phyical Demand	Desktop Task 2 Temporal Demand	HoloLens Task 2 Temporal Demand	Desktop Task 2 Overall	HoloLens Task 2 Overall
Desktop Frist	3.9231	3.1538	1.8462	2.4615	2.9231	3.0769	18.1538	17.8462
Hololens Frist	3.3571	3.7143	1.8571	2.8571	2.5	3	16	20.4286
P- value	0.291	0.22	0.958	0.861	0.616	1	0.627	0.369

	Desktop Task 3 Mental Demand	HoloLens Task 3 Mental Demand	Desktop Task 3 Phyical Demand	HoloLens Task 3 Phyical Demand	Desktop Task 3 Temporal Demand	HoloLens Task 3 Temporal Demand
Desktop Frist	2.8462	3.8462	1.4615	3.0769	2.3077	2.9231
Hololens Frist	3.1429	3.4286	2.3571	3.3571	2.8571	2.8571
P- value	0.619	0.806	0.126	0.96	0.248	0.94

	Desktop Task 3 Performance	HoloLens Task 3 Performance	Desktop Task 3 Effort	HoloLens Task 3 Effort	Desktop Task 3 Frustration	HoloLens Task 3 Frustration	Desktop Task 3 Overall	HoloLens Task 3 Overall
Desktop Frist	2.4615	3.5385	2.8462	4.1538	1.3846	2.6923	13.3077	20.2308
Hololens Frist	4.2143	4.2857	3	3.6429	1.7857	2.0714	17.3571	19.6429
P- value	0.229	0.518	0.804	0.54	0.523	0.344	0.253	0.942

	Desktop Task 4 Mental Demand	HoloLens Task 4 Mental Demand	Desktop Task 4 Phyical Demand	HoloLens Task 4 Phyical Demand	Desktop Task 4 Temporal Demand	HoloLens Task 4 Temporal Demand
Desktop Frist	3.3077	3.9231	1.6154	4.2308	2.8462	3.1538
Hololens Frist	2.7143	2.5	1.9286	3.5	2.6429	2.6429
P- value	0.345	0.113	0.432	0.432	0.691	0.37

	Desktop Task 4 Performance	HoloLens Task 4 Performance	Desktop Task 4 Effort	HoloLens Task 4 Effort	Desktop Task 4 Frustration	HoloLens Task 4 Frustration	Desktop Task 4 Overall	HoloLens Task 4 Overall
Desktop Frist	3.1538	3.6154	3.4615	4.5385	1.7692	2.4615	16.1538	21.9231
Hololens Frist	3.8571	4.6429	2.6429	3	1.5714	1.6429	15.3571	17.9286
P- value	0.766	0.403	0.14	0.073	0.857	0.377	0.884	0.353

	Desktop Task 5 Mental Demand	HoloLens Task5 Mental Demand	Desktop Task 5 Phyical Demand	HoloLens Task 5 Phyical Demand	Desktop Task 5 Temporal Demand	HoloLens Task 5 Temporal Demand
Desktop Frist	4.5385	4.2308	1.9231	3.9231	2.9231	3.6154
Hololens Frist	4.1429	4.5	2.5	3.7143	3.5	3.4286
P- value	0.334	0.694	0.469	0.961	0.185	0.98

	Desktop Task 5 Performance	HoloLens Task 5 Performance	Desktop Task 5 Effort	HoloLens Task 5 Effort	Desktop Task 5 Frustration	HoloLens Task 5 Frustration	Desktop Task 5 Overall	HoloLens Task 5 Overall
Desktop Frist	4.3077	3.8462	4.8462	5	2.3077	3.0769	20.8462	23.6923
Hololens Frist	4.4286	4.7857	4	3.7143	2.8571	2.9286	21.4286	23.0714
P- value	0.941	0.27	0.209	0.304	0.348	0.86	0.942	1

	Desktop Task 6 Mental Demand	HoloLens Task 6 Mental Demand	Desktop Task 6 Phyical Demand	HoloLens Task 6 Phyical Demand	Desktop Task 6 Temporal Demand	HoloLens Task 6 Temporal Demand
Desktop Frist	5.1538	4.9231	1.8462	2.6923	3.7692	3.3077
Hololens Frist	4.2143	4.0714	2.1429	3.4286	3.4286	3
P- value	0.243	0.522	0.938	0.322	0.691	0.802

	Desktop Task 6 Performance	HoloLens Task 6 Performance	Desktop Task 6 Effort	HoloLens Task 6 Effort	Desktop Task 6 Frustration	HoloLens Task 6 Frustration	Desktop Task 6 Overall	HoloLens Task 6 Overall
Desktop Frist	4.6923	4.6923	4.7692	4.8462	3.2308	2.9231	23.4615	23.3846
Hololens Frist	4.4286	4.2143	4	3.6429	2.7143	2.5714	20.9286	20.9286
P- value	0.844	0.607	0.324	0.27	0.92	0.96	0.395	0.734

The tables above show the result for the cross analysis for perceived workload, desktop first is Group B and HoloLens first is Group A.

REFERENCES

Alper, B., Hollerer, T., Kuchera-Morin, J., & Forbes, A. (2011). Stereoscopic highlighting: 2d graph visualization on stereo displays. *IEEE Transactions on Visualization and ComputerGraphics*, 17(12), 2325-2333.

Amar, R., Eagan, J., & Stasko, J. (2005). Low-level components of analytic activity in information visualization. In *Proceedings - IEEE Symposium on Information Visualization, INFO VIS* (pp. 111–117). https://doi.org/10.1109/INFVIS.2005.1532136

Andrews, C., Endert, A., Yost, B., & North, C. (2011). Information visualization on large, highresolution displays: Issues, challenges, and opportunities. *Information Visualization,10(4),* 341-355.

Bach, B., Dragicevic, P., Archambault, D., Hurter, C., & Carpendale, S. (2017, September). A Descriptive Framework for Temporal Data Visualizations Based on Generalized Space-Time Cubes. *In Computer Graphics Forum* (Vol. 36, No. 6, pp. 36-61*).*

Bach, B., Dragicevic, P., Archambault, D., Hurter, C., & Carpendale, S. (2014, June). A reviewof temporal data visualizations based on space-time cube operations. *In Eurographicsconference on visualization.*

Bach, B., Sicat, R., Beyer, J., Cordeil, M., & Pfister, H. (2018). The Hologram in My Hand: How Effective is Interactive Exploration of 3D Visualizations in Immersive Tangible Augmented Reality*?. IEEE Transactions on Visualization & Computer Graphics.*

Barrie, A., Cassell, B., & Cooper, M. (2005). Benefits of Immersion for Viewing 3D Data.

Baumeister, J., Ssin, S. Y., ElSayed, N. A., Dorrian, J., Webb, D. P., Walsh, J. A., ... & Thomas, B. H. (2017). Cognitive Cost of Using Augmented Reality Displays. *IEEE transactions on visualization and computer graphics*, *23*(11), 2378-2388

Beitzel, S., Dykstra, J., Toliver, P., & Youzwak, J. (2018). Network Anomaly Analysis using the Microsoft HoloLens. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 62, No. 1, pp. 2094-2098).

Berg, B. L., & Lune, H. (2012). Qualitative Research Methods for the Social Sciences. *IEEE VIS International Workshop on* (pp. 25-31).

Brath, R. (2014, November). 3d infovis is here to stay: Deal with it*. In 3DVis (3DVis),* 2014

Brath, R., Peters, M., & Senior, R. (2005). Visualization for communication: The importance of aesthetic sizzle. *In Information Visualisation, 2005. Proceedings. NinthInternational Conference on* (pp. 724-729). IEEE.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101

Büschel, W., Chen, J., Dachselt, R., Drucker, S., Dwyer, T., Görg, C., ... & Stuerzlinger, W. (2018). Interaction for Immersive Analytics. *In Immersive Analytics* (pp. 95-138).

Burdea, G., & Coiffet, P. (2003). Virtual reality technology. In *Presence: Teleoperators & Virtual Environments* (Vol. 12, pp. 663–664). https://doi.org/10.1162/105474603322955950

Caine, K. (2016, May). Local standards for sample size at CHI. *In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 981-992). ACM.

Carpendale, M. S. T., Cowperthwaite, D. J., & Fracchia, F. D. (1996, October). Distortion viewing techniques for 3-dimensional data. *In Information Visualization'96, Proceedings* I*EEE Symposium on* (pp. 46-53). IEEE.

Carpendale, S. (2008). Evaluating information visualizations. *In Information visualization* (pp.19-45). Springer, Berlin, Heidelberg.

Chandler, T., Cordeil, M., Czauderna, T., Dwyer, T., Glowacki, J., Goncu, C., ... & Wilson, E. (2015, September). Immersive analytics. *In Big Data Visual Analytics* (BDVA), 2015(pp. 1-8). IEEE.

Chen, S. C., Hsiao, M. S., & She, H. C. (2015). The effects of static versus dynamic 3Drepresentations on 10th grade students' atomic orbital mental model construction:Evidence from eye movement behaviors. *Computers in Human Behavior,* 53, 169-180.

Choo, J., Lee, C., Kim, H., Lee, H., Liu, Z., Kannan, R., … Park, H. (2015). VisIRR: Visual analytics for information retrieval and recommendation with large-scale document data. In *2014 IEEE Conference on Visual Analytics Science and Technology, VAST 2014 - Proceedings* (pp. 243–244). https://doi.org/10.1109/VAST.2014.7042511

Cobb, S., & Fraser, D. S. (2005). Multimedia Learning in Virtual Reality. In *The Cambridge handbook of multimedia learning* (pp. 525–547).

Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. *Research design Qualitative quantitative and mixed methods approaches*. https://doi.org/10.1007/s13398-014-0173-7.2

Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. *In Perception of space and motion* (pp. 69-117).

Donalek, C., Djorgovski, S. G., Cioc, A., Wang, A., Zhang, J., Lawler, E., ... & Davidoff, S.(2014, October). Immersive and collaborative data visualization using virtual reality platforms. *In Big Data (Big Data), 2014 IEEE International Conference on*(pp. 609-614).IEEE.

Dwyer, T., Marriott, K., Isenberg, T., Klein, K., Riche, N., Schreiber, F., ... & Thomas, B. H. (2018). Immersive Analytics: An Introduction. In *Immersive Analytics* (pp. 1-23). Springer, Cham

Elmqvist, N., Dragicevic, P., & Fekete, J. D. (2008). Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *IEEE transactions on Visualization and Computer Graphics,* 14(6), 1539-1148.

Elmqvist, N., & Tsigas, P. (2008). A taxonomy of 3d occlusion management for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(5), 1095-1109.

Fischer, S., Lowe, R. and Schwan, S. (2008) Effects of presentation speed of a dynamic visualization on the understanding of amechanical system. *Applied Cognitive Psychology* 22: 1126–1141.

Fjeld, M., Fredriksson, J., Ejdestig, M., Duca, F., Bötschi, K., Voegtli, B., & Juchli, P. (2007,April). Tangible user interface for chemistry education: comparative evaluation and re design. *In Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 805-808). ACM.

Foo, J. L., Martinez-Escobar, M., Juhnke, B., Cassidy, K., Hisley, K., Lobe, T., & Winer, E.(2013). Evaluating mental workload of two-dimensional and three-dimensional visualization for anatomical structure localization. *Journal of Laparoendoscopic & Advanced Surgical Techniques,* 23(1), 65-70.

Fu, B., Noy, N. F., & Storey, M. A. (2013, October). Indented tree or graph? A usability study of ontology visualization techniques in the context of class mapping evaluation. *In International Semantic Web Conference* (pp. 117-134). Springer, Berlin, Heidelber

Gatalsky, P., Andrienko, N., & Andrienko, G. (2004, July). Interactive analysis of event data using space-time cube. *In Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on* (pp. 145-152). IEEE.

Gorad,S. Combining Methods in Educational Research. *McGraw-Hill Education, New York,USA, 2004*

Hägerstrand, T. (1975). Space, time and human conditions.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *In Advances in psychology* (Vol. 52, pp. 139-183).

Hart, S. G.: NASA-Task Load Index (NASA-TLX); 20 Years Later. *The Human Factors and Ergonomics Society 50th Annual Meeting,* 904-908 (2006)

Huk, T. (2006). Who benefits from learning with 3D models? The case of spatial ability. *Journal of computer assisted learning,* 22(6), 392-404.

Huang, W., Eades, P., & Hong, S. H. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization,* 8(3), 139-152.

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field methods,* 18(1), 3-20.

Jansen, Y., Dragicevic, P., & Fekete, J. D. (2013, April). Evaluating the efficiency of physical visualizations. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2593-2602). ACM.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research,* 1(2), 112–133.

Kapler, T., & Wright, W. (2005). Geotime information visualization. *Information visualization,* 4(2), 136-146.

Kays, R. (2016, July 25). Pet Cat. Retrieved from https://www.movebank.org/panel_embedded_movebank_webapp

Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P. and Montello, D. (2008). Spatial Reasoning With External Visualizations: What Matters Is What You See, Not Whether You Interact. *Cognitive Science: A Multidisciplinary Journal,* 32(7), pp.1099-1132.

Klein MI, Lio CH, Grant R, Carswell CM, Strup S. A mental workload study on the 2d and 3d viewing conditions of the da Vinci surgical robot. *Proc Hum Factors Ergon Soc 2009*.

Kobsa, A. (2004, October). User experiments with tree visualization systems. *In Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on* (pp. 9-16). IEEE.

Kraak, M. J. (2003, August). The space-time cube revisited from a geovisualization perspective. *In Proc. 21st International Cartographic Conference* (pp. 1988-1996).

Kristensson, P. O., Dahlback, N., Anundi, D., Bjornstad, M., Gillberg, H., Haraldsson, J., ... & Stahl, J. (2009). An evaluation of space time cube representation of spatiotemporal patterns. *IEEE Transactions on Visualization and Computer Graphics,* 15(4), 696-702.

Kruijff, E., Swan, J. E., & Feiner, S. (2010, October). Perceptual issues in augmented reality revisited. In Mixed and Augmented Reality (ISMAR), *2010 9th IEEE International Symposium on* (pp. 3-12). IEEE.

Liang, H. N., & Sedig, K. (2010). Can interactive visualization tools engage and support preuniversity students in exploring non-trivial mathematical concepts?. *Computers &Education,* 54(4), 972-991.

Lio CH, Bailey K, Carswell CM, Seales WB, Clarke D, Payton (2006)GM. Time estimation as a measure of mental workload during the training of laparoscopic skills. *Proc Hum Factors Ergon* Soc 2006;50:1910–1913.

Livingston, M. A., Decker, J. W., & Ai, Z. (2012). Evaluation of multivariate visualization on a multivariate task. *IEEE transactions on visualization and computer graphics,* 18(12),2114-2121

Luboschik, M., Berger, P., & Staadt, O. (2016, November). On spatial perception issues in augmented reality based immersive analytics. *In Proceedings of the 2016 ACM Companion on Interactive Surfaces and Spaces* (pp. 47-53). ACM.

Lum, E. B., Stompel, A., & Ma, K. L. (2002, October). Kinetic visualization: a technique for illustrating 3D shape and structure. *In Proceedings of the conference on Visualization'02* pp. 435-442. IEEE Computer Society.

Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov), 2579-2605

Mahfoud, E., Wegba, K., Li, Y., Han, H., & Lu, A. (2018, January). Immersive Visualization for Abnormal Detection in Heterogeneous Data for On-site Decision Making. *In Proceedingsof the 51st Hawaii International Conference on System Sciences*.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog),* 5(2), 110-141.

Mackinlay, J. D., Robertson, G. G., & Card, S. K. (1991, April). The perspective wall: Detail and context smoothly integrated. *In Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 173-176). ACM.

Marriott, K., Chen, J., Hlawatsch, M., Itoh, T., Nacenta, M. A., Reina, G., & Stuerzlinger, W. (2018). Immersive Analytics: Time to Reconsider the Value of 3D for Information Visualisation. In Immersive Analytics (pp. 25-55). Springer, Cham.

McIntire, J. P., Havig, P. R., & Geiselman, E. E. (2012, May). What is 3D good for? A review of human performance on stereoscopic 3D displays. In Head-and Helmet-Mounted Displays XVII; and Display Technologies and Applications for Defense, Security, and Avionics *VI International Society for Optics and Photonics*.

McIntire, J. P., & Liggett, K. K. (2014, November). The (possible) utility of stereoscopic 3d displays for information visualization: The good, the bad, and the ugly. *In 3DVis (3DVis), 2014 IEEE VIS International Workshop on* (pp. 1-9). IEEE.

McGrath, J. E. (1995). Methodology matters: Doing research in the behavioral and social sciences*. In Readings in Human–Computer Interaction (pp. 152-169).*

McMahan, R. P., Gorton, D., Gresock, J., McConnell, W., & Bowman, D. A. (2006, November). Separating the effects of level of immersion and 3D interaction techniques. *In Proceedings of the ACM symposium on Virtual reality software and technology* (pp. 108-111). ACM.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12), 1321-1329.

Mojica, C. M. M., Navkar, N. V., Tsekos, N. V., Tsagkaris, D., Webb, A., Birbilis, T., &Seimenis, I. (2017, October). Holographic interface for three-dimensional visualization of MRI on HoloLens: A prototype platform for MRI Guided Neurosurgeries. *In Bioinformatics and Bioengineering (BIBE), 2017 IEEE 17th International Conference on* (pp. 21-27). IEEE.

Moore, A. B., Whigham, P., Holt, A., Aldridge, C., & Hodge, K. (2003, September). A time geography approach to the visualisation of sport. *In Proceedings of the Seventh International Conference on Geocomputation.*

Nakaya, T., & Yano, K. (2010). Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics. *Transactions in GIS,* 14(3), 223-239.

Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., & Mackay, W. (2011, May). Mid-air pan-andzoom on wall-sized displays. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (*pp. 177-186). ACM.

Neubauer, A. C., Bergner, S., & Schatz, M. (2010). Two-vs. three-dimensional presentation ofmental rotation tasks: Sex differences and effects of training on performance and brainactivation*. Intelligence,* 38(5), 529-539.

Nguyen, H., Ketchell, S., Engelke, U., Thomas, B. H., & De Souza, P. (2017, November).Augmented Reality Based Bee Drift Analysis: A User Study. *In Big Data Visual Analytics (BDVA), 2017 International Symposium on* (pp. 1-8). IEEE.

Renner, R. S., Velichkovsky, B. M., & Helmert, J. R. (2013). The perception of egocentric distances in virtual environments-a review. *ACM Computing Surveys (CSUR),* 46(2), 23.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research,* 42(5), 533-544.

Pham, T., Vermeulen, J., Tang, A., & MacDonald Vermeulen, L. (2018, June). Scale Impacts Elicited Gestures for Manipulating Holograms: Implications for AR Gesture Design. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018* (pp. 227-240). ACM

Saenz, M., Strunk, J., Maset, K., Malone, E., & Seo, J. H. (2015, August). See the Flex: Investigating Various Display Settings for Different Study Conditions. *In International Conference on Human-Computer Interaction* (pp. 295-300). Springer, Cham.

Schoeffmann, K. (2014, November). The Stack-of-Rings Interface for Large-Scale Image Browsing on Mobile Touch Devices. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 1097-1100). ACM

Shaw, C., Green, M., Liang, J., & Sun, Y. (1993). Decoupled simulation in virtual reality with the MR toolkit. *ACM Transactions on Information Systems (TOIS*), 11(3), 287-317.

Shneiderman, B. (2014). Why not make interfaces better than 3D reality*?. IEEE Computer Graphics and Applications,* 23(6), 12-15.

Smallman, H. S., John, M. S., Oonk, H. M., & Cowen, M. B. (2001). Information availability in 2D and 3D displays. *IEEE Computer Graphics and Applications,* 21(5), 51-57.

St. John, M., Cowen, M. B., Smallman, H. S., & Oonk, H. M. (2001). The use of 2D and 3D displays for shape-understanding versus relative-position tasks. *Human Factors,* 43(1),79-98.

Teyseyre, A. R., & Campo, M. R. (2009). An overview of 3D software visualization. *IEEE transactions on visualization and computer graphics*, 15(1), 87-105.

Tory, M. and Moller, T. (2004) Human factors in visualization research. *IEEE Transactions onVisualization and Computer Graphics* 10(1): 72–84

Turner, P., Turner, S., & Burrows, L. (2013). Creating a sense of place with a deliberately constrained virtual environment. *International Journal of Cognitive Performance Support,* 1(1), 54-68.

Wickens, C., Merwin, D. and Lin, E. (1994). Implications of Graphics Enhancements for the Visualization of Scientific Data: Dimensional Integrality, Stereopsis, Motion, and Mesh. *Human Factors: The Journal of the Human Factors and Ergonomics Society,*36(1), pp.44-61.

Wiegmann, D. A., Overbye, T. J., Hoppe, S. M., Essenberg, G. R., & Sun, Y. (2006, October). Human factors aspects of three-dimensional visualization of power system information. *In Power Engineering Society General Meeting*, 2006. IEEE (pp. 7). IEEE.

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

Wilson, M. (2002). Six views of embodied cognition. Psychonomic bulletin & review, 9(4), 625-636.

Yost, B. and North, C. (2006) The Perceptual scalability of visualization. *IEEE Transactions on Visualization and ComputerGraphics* 12(5): 837–844.

Zhang, L., Chen, S., Dong, H., & El Saddik, A. (2018). Visualizing Toronto City Data with HoloLens: Using Augmented Reality for a City Model. *IEEE Consumer Electronics Magazine,* 7(3), 73-80.