Big Data Visualization in Extended Reality

Mahnaz Wahidi Taher Ahmad and Trevina Litchmore

DIVISION OF ERGONOMICS AND AEROSOL TECHNOLOGY
DEPARTMENT OF DESIGN SCIENCES
FACULTY OF ENGINEERING LTH | LUND UNIVERSITY
2023

MASTER THESIS

Big Data Visualization in Extended Reality

Mahnaz Wahidi Taher Ahmad
mahnaz.wahidi@gmail.se

Trevina Litchmore
tlitchmo@gmail.com

June 21, 2023

Master's thesis work carried out at Axis Communications AB.

Supervisors: Günter Alce, gunter.alce@design.lth.se
Jonas Alfredsson, jonas.alfredsson@axis.com
Magnus Bäck, magnus.back@axis.com

Examiner: Mattias Wallergård, mattias.wallergard@design.lth.se

Big Data Visualization in Extended Reality

Published by
Department Design Sciences
Faculty of Engineering LTH, Lund University
P.O Box 118, SE-221 00 Lund, Sweden

Subject: Virtual Reality and Augmented Reality MAMM15
Division: Ergonomics and Aerosol Technology
Supervisor: Günter Alce
Examiner: Mattias Wallergård

Abstract

Effectively utilizing Big Data can be a challenge for many companies. Due to its volume, many aspects of the data processing pipeline such as storage, retrieval and usage may quickly become problematic. For example, it could be difficult identifying the relevant metrics when analyzing a service failure. Visualizing and deriving insights from the data could also prove to be quite onerous. Immersive analytics is one field that aims to solve some of the issues around Big Data by using technologies such as virtual reality, augmented reality and mixed reality.

For our thesis project, we have worked with Axis Communications, a Swedish multinational company within the video surveillance industry to design an immersive analytics application for their R&D Tools department. To define our scope, we employed a co-design approach together with an iterative Double Diamond design process. As a result of academic and user research, concept development and iterative lo-fi prototyping, we decided to implement a hi-fi prototype of an application that would help engineers physically navigate their data. In an extrema identification task, our prototype allowed users to navigate the data more quickly, but in terms of accuracy, users performed roughly the same across both tools. Given that this was a relatively simple task, more research would be needed to investigate the performance of immersive analytics applications in complex data analysis operations.

Keywords: Immersive Analytics, Big Data, Virtual Reality, Co-design, Participatory Design, Data Visualization

Acknowledgements

We would like to thank our superivisors Magnus Bäck, Günter Alce and Jonas Alfredsson for supporting us during our thesis and giving us useful feedback. Also, we greatly appreciate Jonas for helping us settle into the department and supplying us with a copious amount of Post-it notes.

We would also like to thank Emely Jonasson, our manager, for her support in ensuring that we had everything we needed to complete the thesis and navigating the terms of service minefield with us. With that we would also like to thank both Emely Jonasson and Hiromi Blomberg for the speed in which they were able to arrange the approval of the Unreal licenses for our project.

We would like to thank Thomas Ekdahl for providing us with the Meta Quest 2 headset for our lo-fi prototype.

And we would also like to thank Carl-Axel "Cacke" Alm, Rikard Almgren, Stefan Gangefors and David Åkerman for their technical support and advice.

Lastly, a huge thank you goes out to the entire team in R&D Tools, especially those who have participated in our user interviews, focus groups and testing sessions. Without your insightful feedback, we would not have been able to come as far as we did. Also, we appreciate your willingness to include us in your team activities and events as we had quite a bit of fun!

Contents

1 Introduction **5**
- 1.1 Thesis Goals and Research Questions 6
- 1.2 Novelty and Significance 6
- 1.3 Structure of the Report 7

2 Related Work **8**
- 2.1 Big Data Visualization 8
- 2.2 Immersive Analytics 9
- 2.3 Data Representation 9
  - 2.3.1 Situated and Embedded Analytics 10
  - 2.3.2 Graphical Perception 11
- 2.4 Multisensory Presentation 13
- 2.5 Embodied Data Exploration 14
- 2.6 Applications of Immersive Analytics 15
- 2.7 Summary 17

3 Methodology **19**
- 3.1 Design Guidelines and Framework 19
- 3.2 Co-design 21
- 3.3 Prototyping 21
- 3.4 Double Diamond 22
- 3.5 Experimental Design 23
- 3.6 Usability Testing 23
- 3.7 Data Collection Methods 24
  - 3.7.1 System Usability Scale 24
  - 3.7.2 Questionnaires 24
  - 3.7.3 Interviews 24
  - 3.7.4 Focus Groups 24

4 Scope Refinement **25**
4.1 Focus Group Interviews . . . 25
4.2 Affinity Diagram . . . 26
4.2.1 Data and Metrics . . . 26
4.2.2 Data Visualization tools . . . 27
4.2.3 Workspace, Other Services . . . 29
4.3 Brainstorming . . . 30
4.4 Conceptual Designs . . . 30
4.4.1 Tree Concept . . . 30
4.4.2 Mountain Concept . . . 33
4.4.3 Soundscape . . . 35
4.4.4 Smellscape . . . 36
4.4.5 Feedback . . . 36
4.4.6 Summary of Feedback . . . 38

5 Lo-fi Prototyping **40**
5.1 First Version . . . 40
5.2 Second Version . . . 42
5.3 Testing . . . 44
5.3.1 Test Setup . . . 44
5.3.2 Test Procedure . . . 44
5.3.3 Pilot Testing . . . 45
5.3.4 Results . . . 45

6 Hi-fi Prototype **49**
6.1 Data Sources . . . 49
6.2 Building the Environment . . . 49
6.3 The Headset . . . 51
6.4 The Application . . . 52
6.5 Testing Overview . . . 53
6.5.1 Participant Profile . . . 53
6.5.2 Task List . . . 55
6.5.3 Data Collection . . . 56
6.5.4 Test Procedure . . . 56
6.5.5 Pilot Testing . . . 57
6.6 Test Results . . . 57
6.6.1 SUS Questionnaire Results . . . 57
6.6.2 Interview Responses . . . 60
6.6.3 Observations . . . 62
6.7 Summary . . . 63

7 Discussion **65**
7.1 Effectiveness of the Design Process . . . 65
7.2 Prototyping in VR . . . 66
7.3 Legal and Technical Challenges . . . 67
7.4 Limitations . . . 68

8 Conclusion **69**

References **70**

Appendix A **75**

Appendix B **76**

Appendix C **77**

Appendix D **79**

Chapter 1
Introduction

Big Data refers to very large and complex data sets that are difficult to handle and analyze with traditional data analysis applications. It is also referred to as Enormous Information and refers to the sum of information generated by PCs, online media, portable gadgets and other devices or applications. These data sets are characterized using the three Vs: Variety, Volume, and Velocity [1]. Thus, Big Data refers to data that is always increasing in volume with high velocity and variety.

By providing a viewing context such as maps or graphs, data visualization could facilitate the identification of patterns within large data sets [1]. However, the complexity of Big Data can also make it challenging to choose appropriate techniques and methods [2]. Many data visualization tools are developed for use on desktop PCs, which come with a few limitations such as a small display size and difficulties presenting 3D or higher dimensional data in a way that is understandable. Thus, other hardware devices have been utilized to solve this issue such as wall-sized displays [3]. Immersive technologies have also been explored as a potential solution. Given the capabilities of immersive technologies, it could be possible for such devices to lead to the development of more intuitive visualization techniques and methods for Big Data analysis.

Extended Reality (XR) is a general term that covers all immersive technologies such as virtual reality (VR), augmented reality (AR) and mixed reality (MR) [3]. VR allows the user to experience a completely virtual world through a head-mounted display (HMD) or other hardware, while AR overlays virtual information and objects onto the user's physical environment. AR experiences could be facilitated through glasses, tablets, smartphones or other digital displays. Unlike VR and AR, MR combines physical objects with digital content. Within MR experiences, users can manipulate virtual objects which can result in a change to physical objects and vice versa [4].

For our thesis, we worked with Axis Communications. It is a Swedish multinational company in the video surveillance industry that manufactures network cameras, access control systems, intercoms, and audio systems to improve physical security [5]. Within their R&D Tools department, engineers are regularly engaged with examining large internal datasets. The insights gained from their analyses are used within the company to support hardware and software development.

Hundreds of thousands of metrics with trillions of data points are monitored by the team. Terabytes of metrics are collected from cameras and servers located on Axis premises. Given the sheer volume and cardinality of the data, it is difficult to isolate the relevant metrics when troubleshooting an issue or correlating different data sets to identify trends. Thus, XR was explored as a potential solution to some of their current problems with data handling and analysis.

1.1 Thesis Goals and Research Questions

The goal of this thesis was to explore whether or not XR could facilitate monitoring tasks and the derivation of insights for large data sets in real-time. We intended to make this process easier by developing an application that could allow users to efficiently navigate their data and find relevant metrics faster. The R&D Tools team, regularly visualizes, stores and structures their time series and log data with desktop applications such as Icinga, Kibana, Elastic, Grafana and Graphite. Since the team engages in data analysis as part of their everyday work and have experienced some shortcomings with the aforementioned applications, we decided to work with the engineers within R&D Tools as a case study.

To understand their needs, we have adopted a co-design approach to ensure that the application has sufficient usability and may solve some of their current problems in regards to analyzing large data sets. As a result of our process, we would be able to collect the information needed to develop an application that could feasibly be used in their workflow. Thus, this would demonstrate the practical implications of immersive technology.

During the course of our thesis, we have attempted to answer the following research questions:

- How can VR or AR be used to visualize real-time Big Data sets?
- How can we effectively visualize data in a VR/AR environment and make it interactive?
- How can we create such a tool that does not overload the user's cognitive and perceptual capabilities?

1.2 Novelty and Significance

Contribution to the Field: Many data visualization tools for XR technologies employ graphical representations that were originally designed for 2D displays [3]. While there are new methods for such representations tailored specifically for immersive technologies, there is still much work to be done in terms of defining a set of best practices and developing effective data visualization methods. Thus, our thesis project would contribute to the growing body of knowledge in regards to data visualization in XR as we iteratively tested our prototypes with engineers in order to understand the most efficient way to present the data.

Also, there is little research as to how such immersive technologies fare in practical use-cases in the field of data visualization, and our thesis would help to fill this knowledge gap. In our paper we examined the possibilities of using human interaction, cognition, and various modalities to represent data in a way that could be easier to understand and use. This process resulted in a hi-fi prototype that was tested with real data in a realistic, yet extremely limited data analysis task: extrema identification.

UN Global Goals: The UN Global Goals were developed to serve as a blueprint for "peace and prosperity for people and the planet, now and into the future" [6]. Our project fulfills goal eight in that it supports productive employment. The intent of our application is to assist engineers with their work processes and to make the usage of data less onerous. If successful, it could accelerate the troubleshooting process, which would lead to less frustration and more time for the engineers to devote to other tasks.

1.3 Structure of the Report

This report documents the process in which we developed a data visualization application using immersive technologies. In the following chapter, we describe the research that we conducted in the course of our thesis. The third chapter describes the methods and theories underlying our work. In the fourth chapter, we explain how we refined the scope of our project. Our lo-fi prototype, the testing undertaken as well as the test results are described in the fifth chapter. Within the sixth chapter, we describe our hi-fi prototype, the testing procedure and the results. For the seventh chapter, we discuss the findings of the thesis as a whole and reflect on the effectiveness of our design process. In chapter eight, we conclude with our perspective on the feasibility of an immersive data visualization application in the real-world context of a software development tools department.

Chapter 2
Related Work

Developing a data visualization tool that utilizes immersive technologies can be challenging. For general immersive applications, the design guidelines are still evolving as the technology is relatively new and not as ubiquitous as mobile phones or desktop computers. Thus, it is important to understand the underlying design principles for data visualization and immersive analytics.

2.1 Big Data Visualization

Managing and extracting insights from Big Data is often difficult. Depending on the characteristics of the data, the analysis of such data sets can be quite complex. Since working with Big Data requires human interaction, logical thought and judgment, it is important to consider our perceptual and cognitive capabilities. Using traditional visualization techniques such as diagrams, tables, plots and images could result in complex figures that may be difficult to comprehend. Thus, improvement upon current visualization techniques are needed in order to meet those challenges [7].

Typically, feature extraction and geometric modelling are applied to large data sets in order to increase the interpretability of graphical representations [8]. In [7], the authors concluded that one of the most promising methods of enhancing Big Data visualization involves AR and VR technologies. These technologies could accommodate our sensory limitations. Hence, the field of immersive analytics becomes relevant when seeking new ways of displaying Big Data.

2.2 Immersive Analytics

The field of immersive analytics is concerned with the problem of representing and analyzing data using immersive technologies. It is focused on overcoming the "barriers between people, their data and the tools they use for analysis and decision making" [9]. The amount and complexity of data in our current era challenges our ability to make sense of it and to utilize it effectively. Thus, one of the goals of the field is to help make data more understandable by contextualizing the data and leveraging our spatial and sensory capabilities with XR.

Realizing the full potential of immersive analytics requires research into applications that transcend the current desktop-centric paradigm. The emergence of natural interfaces that use speech, gesture and touch provides opportunities to develop applications that support this shift and allow for the flowering of new forms of collocated and remote collaboration [10]. Thus, these interfaces could be a key aspect in the creation of engaging and embodied analysis tools. Beyond natural interfaces and immersive technologies, the field encompasses several different concepts such as data representation and visualization, situated analytics, embodied data exploration, multi-sensory presentation and graphical perception.

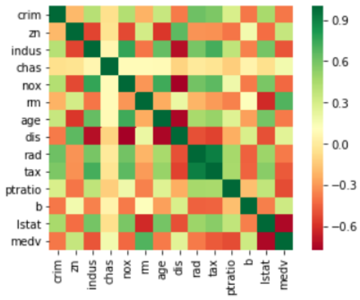
2.3 Data Representation

Data representation is one of the key components of analytics applications. It involves the encoding and abstraction of data into visual variables. This includes both data physicalizations and visualizations. When an object's form encodes a data value, it is referred to as a data physicalization. However, when the data representation is purely visual, it is referred to as a data visualization [11]. Examples of visual variables could include changes to the corresponding object's appearance in relation to temperature such as the height of the central red line common to many analog thermometers (see Fig. 2.1a) or the colors of a heatmap that symbolize the correlation value between different features in a machine learning model (see Fig. 2.1b).

When it comes to immersive analytics, it is important to consider the proximity and the level of integration of the data representation with the relevant object. Situated data representations are located close to the data's physical referent (i.e. the object to which the data refers) or in a relevant physical context (see Fig. 2.2a). Visualizations that are displayed on phones, tablets or laptops can also become situated by placing the device in a relevant space or near a relevant object. These can include ambient, peripheral and AR displays [11, 14]. Embedded data representations build upon the concept of situated representations. Not only are such representations located close to the physical referent, but they are also incorporated into its surroundings. Often these embedded data representations are groups of situated data representations [11] (see Fig. 2.2b).

For some use cases, it is not always possible to have the data depicted in proximity to the physical object. This could be due to many factors such as safety or the sheer size of the referent. In this situation, a virtual facsimile may be used that approximates the relevant

(a) Source: [12]



(b) Source: [13]

Figure 2.1: Examples of data representation. 2.1a is a physical representation, while 2.1b is purely visual.

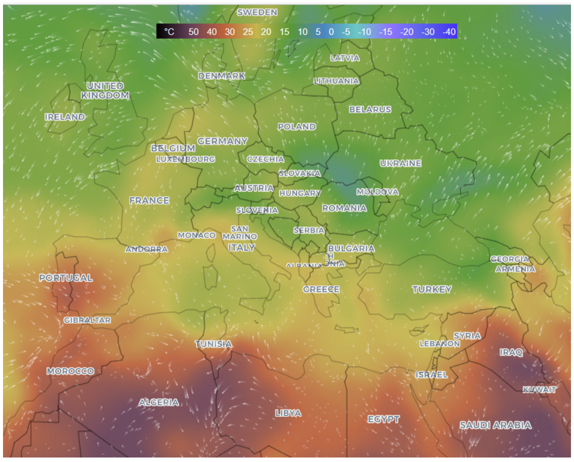
object in some form (see Fig. 2.2b for an example). While this may introduce some indirectness in regards to the spatial relationship between the data representation and the object, the concepts of situated and embedded data representation remain valid in this case [11].

2.3.1 Situated and Embedded Analytics

Analytics utilizes data representations, but also involve the display of information. Situated analytics is similar to situated data representations in the sense that situated analytics have a close proximal relationship to the physical or virtual referent [11]. An example of this concept in practice could be seen in Fig. 2.3. While cooking, the user is able to see the temperature of the food and the amount of time a pot should be left on the stove. Within the context of immersive analytics, situated analytics is often found in AR or MR applications. Video pass-through and digital overlays facilitate the mapping of relevant analytical information onto the physical world. This concept could also be applied in VR applications. However, in that case the referent would be virtual as the environment is wholly artificial.

In Fig. 2.3, we can see three examples of situated analytics (cooking time, food temperature and a reminder). When we consider the presentation of these data points and information as a group, we transition from situated to embedded analytics. Instead of viewing the data and information in isolation, embedded analytics facilitates a more global interpretation [11]. Thus, from a collective perspective, we can discern that the user of this application

(a) Source: [15]



(b) Source: [16]

Figure 2.2: 2.2ais an example of a situated data representation. The thermometer is within the meat and displaying its temperature. 2.2b is an example of embedded data representation with a virtual facsimile (the map). Its coloring is indicative of air temperatures at the locations shown.

is in the midst of cooking.

2.3.2 Graphical Perception

When it comes to viewing graphs in XR displays, the characteristics of the graphs and the hardware determine the effectiveness of the visualization. Certain data representations and analytical tasks are easier to conduct in XR compared to desktop and vice versa. 3D representations are better for showing the overall structure for multidimensional data and networks when viewed with immersive technologies [18, 19]. This benefit was also apparent when viewing complex scenes that involved spatial manipulation [18]. Hardware such as HMDs with head-tracking, provide depth cues that improve the utility of 3D visualizations [18]–[20]. For more precise manipulation, accurate data value measurement or comparison, 2D representations are preferable [18].

Figure 2.3: A screenshot from a video demonstrating a prototype for a cooking application. It shows the temperature of the food, the time left for a dish to finish cooking as well as a reminder.
Source (of image): Adapted from [17].

With AR applications, there is an extra challenge when it comes to displaying graphs: color perception. In [19], a study was conducted where user experience and performance was compared across five different visual channels for the HTC Vive (a VR HMD), the HTC Vive with the Zed mini (an AR HMD) and a desktop monitor. Participants with the AR display had significantly higher error in terms of color perception than with desktop or VR. A similar result was also found in [9]. Color ramping and other strategies such as edge enhancement or recoloring were suggested to increase the viability of color encoding in AR [19]. It was also suggested that visual channels other than color should be used for differentiation within AR applications [9].

There are also challenges when displaying spatial and spatio-temporal visualizations in VR. Despite VR's advantage over desktop in regards to the comprehension of multidimensional data, not all visualizations are suited for every task. In [21], the authors compare 2D and 3D representations of the Earth in a VR head-tracked binocular HMD. The researchers compared the following visualizations: an exocentric globe placed in front of the viewer, an egocentric globe located around the viewer, a flat map rendered onto a plane in VR and a curved map created by projecting the map onto a section of a sphere that was positioned around the user. When using the exocentric globe, users were more accurate than the egocentric globe and the flat map for distance comparison and direction estimation. With direction estimation, users were also faster with the exocentric globe than all other visual presentation methods tested. However, when comparing areas, more time was needed when presenting the data on exocentric and egocentric globes than on flat and curved maps. In almost all cases the egocentric globe was the least effective visualization mode.

Thus, when designing an immersive analytics application, care must be taken in regards to the design of the visualization, and the design must be tailored to each platform. With current AR tools, color differentiation can prove to be unreliable because the background would influence how the digital overlay would be perceived. Additionally, with VR, the way in which data is displayed can affect how quickly and accurately a user can make a judgement about a set of data points. In some cases a 3D representation is the best solution. In others, a 2D representation would suffice. However, XR encompasses more than just the visual modality. By incorporating other human senses, the utility of these representations could be improved.

2.4 Multisensory Presentation

In the previous section, we discussed purely visual immersive applications. However, other senses can be utilized such as hearing, touch, smell and taste. When mapping data to different sensory channels, it should be noted that our senses differ in regards to fidelity and channels of perception. For the visual modality (in order of decreasing effectiveness) size, color, motion and shape can be used to show categorical attributes, while linear position, length, angle, area, depth, color, curvature and volume can be used to show ordinal attributes [22].

With sound, the three most prominent characteristics are loudness, pitch and timbre, and it is best perceived between one and four kHz, which are the characteristic frequencies of human speech. Timbre is often referred to as tone color, and it describes the frequency characteristics of a specific sound. It is also possible for people to recognize rhythmic patterns and to locate the source of a sound in space [22]. All of these characteristics can be used to create data representations.

Additionally, audial data representations can consist of multiple individual sources and varying overtones, which together can produce a unique and distinguishable "sound image" [22]. However, care needs to be taken when incorporating sound in an immersive analytics application as constant low-level sounds can rapidly fade from conscious awareness while loud fluctuating sounds, which are effective at attracting attention, can cause fatigue or distress after long-term exposure [22].

The sense of touch also provides various channels in which data and other information could be communicated. Humans are able to detect indentations and vibrations on the skin through mechanoreceptors, which allow us to feel different sensations such as pressure, vibrations and texture. Our most sensitive mechanoreceptors are found within the top layers of the skin and on non-hairy parts of the body such as the lips, tongue, palms or the soles of our feet. Thermorecptors allow us to perceive the temperature of objects on our skin and are found all over the body. Proprioceptors help us understand the position of our body in space and are found in muscles, tendons and joints. They are also integrated with the vestibular system. Nociceptors allow us to feel pain by detecting an excess of mechanical, thermal and chemical stimuli, which help us to protect our body from harm [22].

While it is possible to induce many different tactile stimulations with the use of haptic technology, it should also be noted that people may struggle to identify arbitrary objects.

However, haptics can be quite useful when trying to identify familiar objects [22]. Thus, it could be possible for people to learn associations between objects and their corresponding tangible properties. Force, position, vibration, contact geometry, slippage and temperature are potential candidates for variables that could be employed in haptic data representations. Complex structures could also be represented as one unit or a haptic glyph. Braille is one such example, but other tactile systems could also be developed [22].

Our gustatory and olfactory senses offer many opportunities when it comes to presenting data and information, and they are also tightly integrated. When we taste something, we are capable of distinguishing five different qualities, while the nose can discern 100s of different substances [22]. However, the dimensions of smell perception are not well understood. There have been attempts to create smell maps in order to assist with the classification and creation of smells. One such smell map was proposed in [22]. This map consisted of four main segments: chemical smells, sweet and floral smells, earthy smells and acrid smells.

When it comes to immersive applications, there are some that employ synthetic smells. Smell systems typically have arrays of pre-mixed chemicals that are dispersed when needed. Unfortunately, these technologies are limited in the number of smells that they can produce, and the most sophisticated systems can be prohibitively expensive for personal use. The most advanced systems are often utilized in cinemas or amusement parks.

Our senses operate synergistically in order for us to perceive the world. This is due in part to how our physical world is constructed. For example, when we see birds, not only are we able to hear them, but we can touch them as well. Thus, in multisensorial analytics applications, redundant encoding across multiple senses could be employed. It would be possible to not only represent data with different visual channels, but also other senses as well. For example, a haptic glyph could be paired with audio or visualizations with smells.

The potential of multisensory data representation is also accompained with drawbacks. One example is sensory crosstalk, which can interfere with a user's judgement. In [23], users misjudged the frequency of a visual signal because of a poorly timed audio signal. Thus, it is possible for multisensorial systems to lead to the falsification of signals. However, the conditions in which this occurs requires further research.

Given that different modalities can be utilized to encode data and convey information, this could also lead to sensory overload. Sensory overload arises when there are too many signals in the environment arriving from different modalities, which may reduce a user's ability to decipher the information [22]. Thus, it is important that a multisensory system supports the user's flow state and minimizes distractions.

2.5 Embodied Data Exploration

While immersive applications could be developed to utilize the full sensory capabilities of the human body, the body itself should also be taken into consideration. It is thought that since we inhabit a physical world that "we are [the] most engaged and effective when working physically" [24]. The same could also be true for virtual environments as well. Allowing the user to interact with the virtual environment as if they were in a physical environment allows for a greater range of interaction and navigation modalities such as embodied data exploration. Also, when navigating virtual spaces, it has been shown that physical navigation (i.e. rotational movements of the eyes, head and body) have been shown to be advantageous over virtual navigation (i.e. teleportation or joystick control) [24, 25].

Distributed cognition can also play a role in designing immersive analytics applications. This concept involves utilizing the affordances of the physical space to support cognitive activities [25]. An example of distributed cognition would be writing down a set of tasks to be done on a whiteboard. In this case, the individual would not have to memorize the list. The person could simply glance at the board to be reminded of what must be done next. Thus, space can play an important role in supporting the tasks of data analysts. It would allow them to organize and structure the collected information and data presentations in a way that would facilitate their workflow and thought processes.

However, one must also be aware of the potential pitfalls that may occur when implementing embodied data exploration such as the "gorilla-arm" syndrome. This is a phenomenon that occurs when a person holds their hands in the air for extended periods of time which leads to fatigue [26]. Additionally, there may be individuals who may not have the ability to move whether through medical illness or another cause. Thus, it is important that such applications are flexible enough to accommodate individuals with a wide range of physical capabilities.

2.6 Applications of Immersive Analytics

Even though immersive analytics is a relatively new field, there are quite a few potential use cases in literature. Many of these applications explore ways in which data can be recontextualized with immersive technologies. Others explore alternative sensory modalities such as sound and touch as a way to communicate visual information.

There are a few AR applications developed with the Microsoft HoloLens. In [9], a wall-mounted display provided a frame of reference for AR content and allowed users the ability to interact with the application using pen and touch. The AR HMD would help guide awareness, show remote parts of the screen and allow for a more personalized viewing experience. Within this study, the researchers also explored different features that could be used in AR such as hinged visualizations, curved screens, brushing and linking and extended axis views (see Fig. 2.4). From their findings, they recommended embedded AR visualizations as a means to avoid focus switches, enable the support for multiple data levels and to allow for the incorporation of additional data if needed.

Another application developed with the HoloLens was described in [27], which presented a different design approach than in [9]. Instead of utilizing additional displays, the authors only used the HoloLens. Within their paper, they demonstrated a use case for a security application that would help users investigate abnormal events in heterogeneous, time series

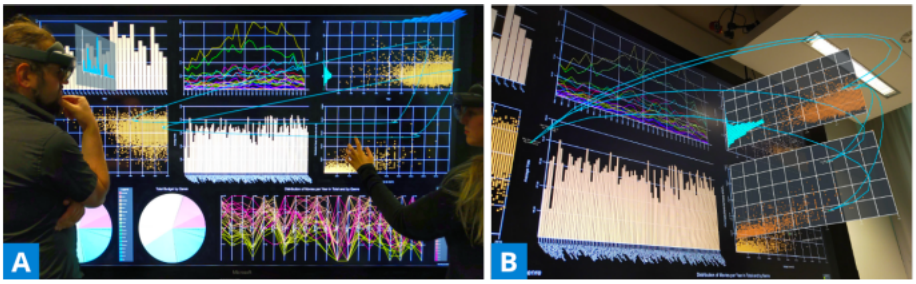
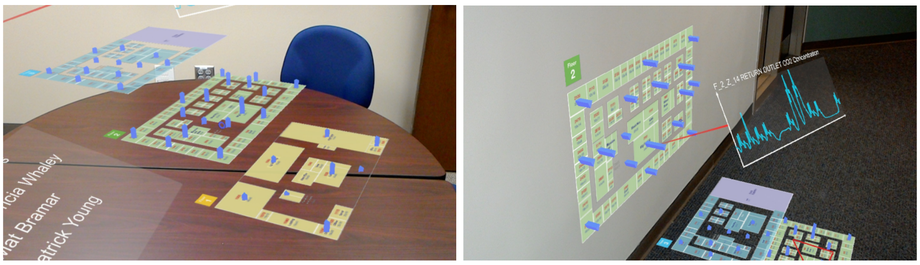


Figure 2.4: Two images of the prototype developed in [9]. Image A shows two analysts interacting with the application. Image B shows the hinged visualization and linking features.

data collected in real-time from various on-site sensors. Instead of utilizing various visualization modes as depicted in [9], the representations were mainly in the form of 3D floor plans and linked temporal line charts (see Fig. 2.5). However, it did offer different interaction modalities than the application in [9] such as voice, gestures and gaze and was also designed for collaborative activity. Similar to the AR application in [9], it also mitigated the need for focus switching and allowed for collaboration.



(a) **(b)**

Figure 2.5: An image of the AR security application described in [27]. Visualizations can be placed on physical surfaces such as a wall, table or the ground.

In terms of VR, there are applications designed to handle multidimensional data. One such application used VR to provide real-time data analytics relating to the COVID-19 pandemic for 50 US states [28]. With both the Oculus Rift S (a VR HMD) and a smartphone, they created two versions of the application. They mentioned that the VR HMD led to a better understanding of the data compared to the mobile experience. However, there was also a concern as to whether VR analytics applications would be easy to use with complex datasets, especially for those without VR experience.

In [29], the researchers addressed the problem of information overload with the GeneVR application. It was designed to make dealing with multidimensional and highly interconnected data easier for biomedical researchers who were new to VR. The application provided an interactive visualization of gene networks in 3D space on an Oculus headset. At the end of their study, they validated the suitability of GeneVR for novice users. Thus, proving that it is possible to create VR analytics tools suitable for those without VR experience.

There are also XR applications that utilize the auditory sense to convey analytical information. Two examples are Audiofeeds and Cheddar. Audiofeeds is a Big Data application that monitors user's social network data in real-time and creates a sonified overview of events that occur on their social media feed. This application was designed to be run on a smartphone. It also had a GUI which allowed users to view more detailed information about these events. With over 86% accuracy individuals were able to identify peaks in activity and were able to identify the platform with over 99% accuracy. Though it should be noted that only three social media platforms were integrated into the device [30].

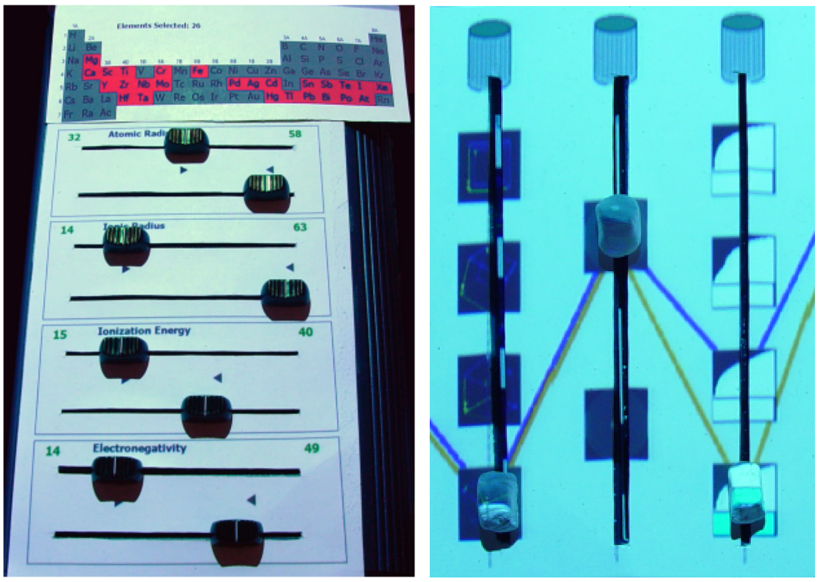
While Cheddar was also designed to handle large discrete data sets, it was not designed as a mobile application. MIDI (Musical Instrument Digital Interface) sliders, a computer keyboard and a mouse were needed in order to operate the application. It also sonified data, but unlike Audiofeeds, the application allowed for the customization of temporal, spatial and sonic parameters in real-time. However, there were limitations regarding the number of variables per data point as well as how parameters were assigned. Despite these restrictions, it was employed within scientific use cases in geology and gestural behavior [31].

Haptics have also been used for data visualization tasks. In [32], researchers used a force-feedback device, the SensAble technologies PHANTOM haptic device, and audio to provide quick overviews of bar charts. A keyboard was also used to change the mode of the software from browsing to building and vice versa. It was based upon an existing application called Graph Builder. Bars were represented as "V"-shaped grooves cut into the background while the axes were represented as cylinders. When moving the bars up and down notes would be played. A sound bar was also incorporated that would play a note with a pitch proportional to the height of the bar in the graph in order to give a quick overview of the graph shape. With this application, users could construct graphs correctly 90% of the time.

MR combined with haptics also offer interesting possibilities for data exploration. In [33], researchers developed a MR data visualization application composed of a haptic mixing-board and projected video. Part of the visualization output and dynamic labels were projected onto the mixing board (see Fig. 2.6). With various forms of data representations, the device was used to accomplish various data exploration tasks such as finding certain values/areas of interest and identifying abnormalities. This study was undertaken not to compare this application to existing tools, but rather to determine its potential.

2.7 Summary

Immersive technologies offer a plethora of alternatives to the traditional desktop-based analytics solutions. By also leveraging the potential of our physical bodies as well as our sensory capabilities, such applications can offer novel ways of interacting and exploring data. However, we must also take into account human physical and sensory limitations in order to ensure that we do not overwhelm the user. Additionally, these new methods of interaction while more natural, may not always be more intuitive. With many of the applications listed



(a) (b)

Figure 2.6: Images of different configurations of the MR application developed in [33]. Fig. 2.6a shows an exploration task centered around the periodic table while in Fig 2.6b, the device has been prepared for exploring a 3D model.

in the previous section, training was also required in order to learn how to use these tools effectively. In order for these tools to have practical use, it must also be possible for users to learn how to use these tools independently.

Chapter 3
Methodology

For this thesis, we have adopted design guidelines specifically for immersive analytics applications. In the analytics field, there is a common understanding as to what an effective analytics application should be able to accomplish. Those conventions provided a foundation for the functionality of our prototype. Also, given our unfamiliarity with the way in which the R&D Tools department prefers to work with data, we decided to adopt a co-design approach. This ensured that we were meeting the needs of the team and developing an application that could feasibly be part of their work processes.

3.1 Design Guidelines and Framework

In [24], a set of design guidelines were suggested for immersive analytics applications, which are described in this section. It was also acknowledged in the same paper that further research into how best to design such applications was needed. Nevertheless, it provided us with a design direction. The facilitation of fluid interactions was a desired capability for our application. Thus, we considered the following suggestions:

- Avoid abrupt switches between different modes by using smooth, animated transitions between different states.
- Provide real-time visual feedback for every interaction including low-level interactions such as key presses or cursor movement.
- Whenever possible, direct manipulation should be utilized. Interactions should be incorporated into the visual representation instead of being separated into control panels.
- To keep users engaged in the application, it is important to indicate to the user when and how interactions can be performed.

- Ensure that continued exploration is always possible.
- Support the user in developing a clear conceptual model. It should always be possible to reverse operations, and the user should be able to have a clear idea of the system's state at all times.
- Where possible, operations should be integrated and accessible in the same mode.

We also considered the tasks that interactive visualization systems are expected to support:

- **Encoding/Visualization**: Choosing an appropriate visualization or visual representation for the data.
- **Filtering**: Data records are excluded based on specific criteria.
- **Sorting**: The ordering of data items.
- **Derivation**: The process in which the analyst elicits data from the primary input data. Though this task is more related to visual analytics than visualization.
- **Reconfiguration**: Changing the graphical display for specific data records as needed or desired.
- **Adjusting**: Allowing for the analyst to adjust parameter settings for the purpose of analysis.
- **Selecting**: The act of choosing different data items.
- **Navigating/Exploring**: How the analyst would move through the interface and view data.
- **Coordinating/Connecting**: Enabling analysts to relate views or individual graphical entities to each other.
- **Organizing**: Allowing the arrangement and grouping of different views or tabs that supports the analysis process.

The following tasks were also considered to be important for immersive analytics applications even though they are not specific to analytics. However, they are needed in order to support analysts in their work as they are related to process and provenance. They included the following functions:

- **Recording**: The software should keep a record of the analysts' operations throughout their process. This would allow them to understand how they came across their findings.
- **Annotating**: Allowing analysts to add labels, highlights or other notes to visualizations could help them keep track of the context in which they discovered their insights by supporting their working memory.
- **Sharing**: Visualization can contribute greatly to collaborative tasks in addition to any notes made and the history of the analysts' process.

- **Guiding**: Supporting the development of visualizations that are static yet self-explanatory. This can be accomplished through animations or stepped interactions.

In addition to the guidelines, we had also employed the 5 Question Design Framework for IA (immersive analytics). It was based on Brehmer and Munzer's "What-Why-How" data visualization framework. For the purpose of IA, the questions Who and Where were added. Thus, the five questions within this framework became "Where-What-Who-Why-How" [34].

Each question within the framework concerned a different design aspect for IA applications. *Where* is concerned with how the data is presented to the user, the type of interaction the user will have with the data, how much the system would need to know about its physical location and the environment in which the application would be used. *What* describes the data that would be presented, if they are static or dynamic and the algorithms needed to represent the data. *Who* pertains to the users of the data and the type of collaboration that would need to be supported. *Why* relates to the different tasks that users can perform while in the application. This category is split into high-level tasks (e.g. presenting data) and medium-level tasks (e.g. browsing datasets). Lastly, the *How* category relates to the implementation of the application such as the graphics fidelity required, aggregating and filtering data, mapping the data to various sensory channels and spatial variables, arranging, combining and placing views as well as how the application would facilitate collaboration [34].

The aforementioned guidelines and the 5 Question Design Framework for IA, provided us with a foundation on which we could develop an effective analytics tool. However, these guidelines alone were not sufficient for our use case. To design an application for actual users, we would also need their input.

3.2 Co-design

With co-design or participatory design, the idea is to work together with users to identify a problem and develop solutions [35]. Given that the way of working can vary between companies, departments and individuals, we felt that co-design would allow us to reach a more feasible solution for the R&D Tools team that can complement their existing workflows. Co-designing can take many forms. This process can involve the use of focus groups, brainstorming sessions, informal testing, conceptual design and storyboarding. These all assist us in developing a suitable application for our user group.

3.3 Prototyping

After deciding on a general design, the prototyping phase commences. Prototyping is an experimental stage in product development where design ideas are tested. Designers create prototypes with varying fidelity to present concepts and evaluate the content, functions and interaction modalities with the user group [36]. At this stage, designers can iteratively refine and validate the design, thereby allowing them to rectify potentially expensive flaws before production. Prototypes usually have two levels of fidelity: low and high.

Low-Fidelity prototyping (aka lo-fi prototyping) is a very quick and easy way to transfer high-level design concepts into testable artifacts. Such prototypes can be constructed with software applications, paper or other materials. At this level of prototyping, the goal is to inspect and validate the operation of the proposed application or product. The prototype itself would have a fairly simple visual layout with basic shapes representing the main elements and content [36].

Paper prototypes and clickable wireframes are two common low-fidelity prototyping techniques. Wireframes can also be used to display the functionality and interactivity of the design with software applications such as Figma or AdobeXD. These prototypes allow for testing early in the design process. They help designers identify and remedy problems with the product before moving further into development where changes may not be as easy to implement. Furthermore, they are very low cost and easy to build and augment, which provides opportunities for more iterations. Additionally, such prototypes allows stakeholders and end users to develop a better understanding of the potential functionality of the product.

When designers have a clear direction and are confident about their solution, they would start to create the high-fidelity prototype (aka hi-fi prototype). This prototype is intended to be quite similar to the final product as it would have a more realistic and detailed design that has almost the same content and interaction methods as the final product [36]. Thus, users would interact with it as if it were the final application.

This type of prototype can be very beneficial for testing as feedback would be related to the finer details of the application such as its appearance, user experience and overall functionality. In contrast to lo-fi, hi-fi prototypes are more expensive and time consuming to make due to their complexity. There are also several digital platforms where designers can create complex prototypes with many interactive elements and animations.

3.4 Double Diamond

While we have used a few different methods in our thesis such as co-design and design guidelines for IA, our overarching design process follows the Double Diamond approach. Fig. 3.1 is a visual representation of the process. This model has four iterative phases:

- **Discover**: It covers the research phase in which the designer begins to understand the problem. Often it involves conducting user research with target groups.
- **Define**: In this phase, the designer will analyze the information obtained in the first phase in order to formulate a problem definition.
- **Develop**: The designer brainstorms several solutions to the defined problem.
- **Deliver**: For the last phase, the designer will test the different solutions and choose the best one. The selected design will become the final product.

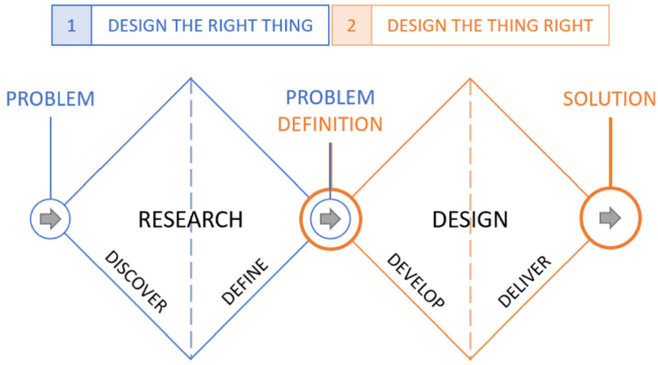


Figure 3.1: Double Diamond Design Process
Source: [37]

3.5 Experimental Design

At the end of the design process, the hi-fi prototype is often evaluated with some form of user testing. For this thesis, we have decided to go with a within-subjects experimental design. It describes a testing setup where different groups are exposed to the same test conditions [38]. In the context of this project, the design protocol would involve two different user groups that are given the same task list to accomplish within our application and within one of the applications that the R&D Tools team are currently using. Complete counterbalancing would also be used (i.e. an equal number of participants who start with our application would also start with the current programs used in the Tools department) [39].

We chose this experimental setup as it would negate the affect of familiarity with existing software programs and reduce order effects (e.g. level of fatigue, learning transfer, etc.). Also, it would give us some insight as to the learnability of our application compared to the existing analytics applications in the Tools department. While learnability is not integral to our research question, this information would be useful in evaluating the practicality of IA applications.

3.6 Usability Testing

Usability refers to how well a user in a specific context can use a product to perform specific tasks in an effective, efficient and satisfactory manner. According to ISO 9241-11, a product or service should be effective, efficient and satisfying to be usable [40]. Usability testing is the process of evaluating a product with participants representing the target users. This testing is undertaken in order to evaluate the degree to which the product meets certain attributes such as effectiveness, efficiency and satisfaction. It can also assist designers in determining the strengths and weaknesses of the product or service. This type of testing consists of many techniques such as giving tasks to users to perform in certain conditions and collecting data from the test sessions by using different data collection methods such as observations and interviews.

3.7 Data Collection Methods

3.7.1 System Usability Scale

The System Usability Scale (SUS) is a questionnaire created by John Brooke used to measure the usability of different products such as software, hardware, websites and other applications [41]. It has ten questions which provide five options for responses that range from strongly disagree to strongly agree (see Appendix A). This questionnaire typically can be used to evaluate a participant's experience with a product.

3.7.2 Questionnaires

A questionnaire is a simple and cheap research tool that consists of a list of questions to collect data from users about a particular subject. Usually there are a mix of closed and open-ended questions. It is common practice to use a screening questionnaire prior to user testing to gain insight about the user's background.

3.7.3 Interviews

Interviews are a qualitative user research method used to gather user feedback. With interviews, designers can collect a large quantity of information from small groups of users. Usually they are conducted in a one-to-one format. When using this method, it is also important that the interviewer does not influence the interviewee's responses by guiding them towards a desired answer. To gain the maximum benefit from this method, it is also important to ensure that the conversation remains relevant to the topic.

3.7.4 Focus Groups

A focus group is a type of interview that involves smaller groups of users with comparable demographics and backgrounds. It is typically an open or guided discussion related to a particular study topic. This method allows designers to examine how potential customers may react to a new product or service through interactive dialogue, which can lead to incredibly rich information. The participants should be carefully selected as they are intended to be representative of the broader user group, and their responses can provide insight into the reception of a product or service by a specific community.

Chapter 4
Scope Refinement

After completing our review of the literature on IA, we began our user research phase within the framework of the co-design process. (This was a continuation of the Discovery phase of the Double Diamond approach.) To gain a deeper understanding of the R&D Tools team, their work processes, how they currently visualize data and their pain points, we conducted eight focus group interviews. From their responses, we created an Affinity Diagram, which helped us define our design problem.

4.1 Focus Group Interviews

Each focus group consisted of individuals from the same subteam within R&D Tools. For each service or collection of services within the department, a subteam is responsible for providing maintenance, monitoring and support. Collectively, the services they maintain support a CI/CD (continuous integration and continuous development) pipeline. These services include source code management systems, log collection and data visualization tool stacks as well as autonomous testing and integration. The interviews were scheduled to last for 30~minutes each in order to allow for us to get a broad overview of the different work processes at R&D Tools.

For each focus group, we had the same set of questions prepared (see Appendix B). It was a list of 14 questions that covered the following topics: their working environment (on-site, remote or hybrid), their typical work day, whether they often switched subteams, the type of data and data sources they worked with, as well as how they utilized the data. To help us gather ideas for our application, we also asked them about what they would like to see in a program. These questions allowed us to investigate the broader problem space of our project.

4.2 Affinity Diagram

To analyze the answers from the focus group interviews, we applied the affinity diagramming process. It is a method of qualitative data analysis. The aim of this method is to identify common themes amongst the participants' responses. Sometimes quantitative data can also be included in the affinity diagram as well, but it is primarily a way to organize qualitative data [42].

From their responses, we were able to identify four major themes: data and metrics, their experience with different data visualization tools (Data Visualization Tools), the space in which they conducted their work (Workspace) and the other systems and services that they interacted with in the course of their work (Other Systems/Services). (Fig. 4.1 shows the resulting diagram.) For the purposes of our project, we were most interested in their responses for the categories of Data/Metrics, Data Visualization Tools and Workspace. These three categories were key in determining the type of XR analytics application that would be suitable for the team. Immersive technologies have a wide range of capabilities and limitations that are heavily task and space dependent. Thus, their responses in these categories were critical in determining our design direction.

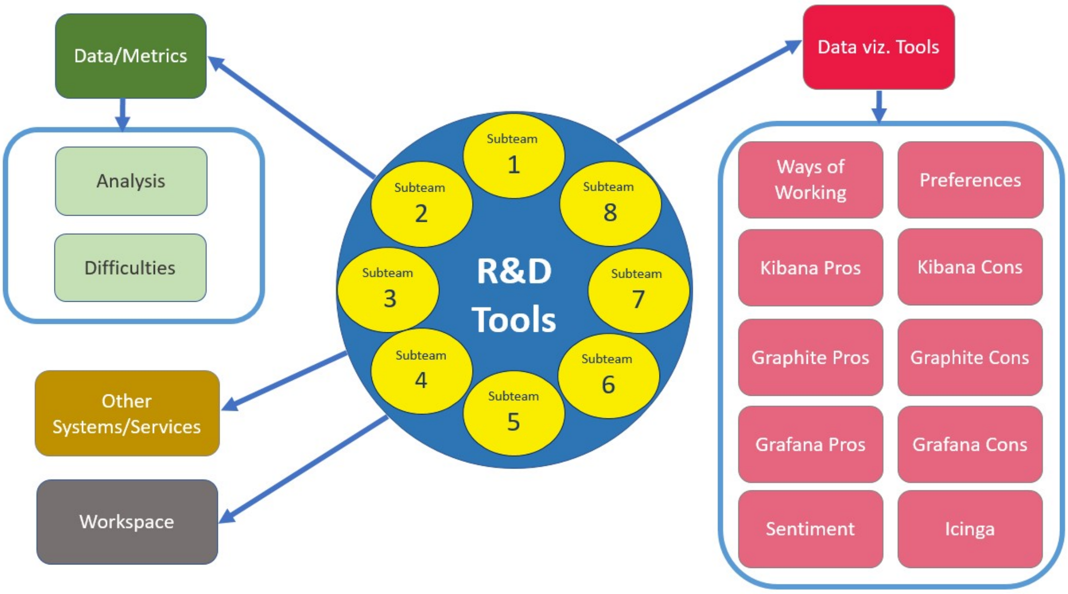


Figure 4.1: The affinity diagram from the focus group interviews. It highlights the four major themes and relevant subthemes from the responses.

4.2.1 Data and Metrics

Each service used within the department has their own logs. As part of their normal function, some services require logs from other services. Most subteams work with time series data while others work with document-based data in plain text or JSON format. How they present the data varies: line graphs, bar graphs, tables or histograms are often utilized.

- **Data Analysis:** Some of the subteams conduct some sort of analysis, which mostly involves correlation, aggregation and filtering. Sometimes visualizations of the data are used for the purpose of exploration. With numerical data, additional data transforms are often employed such as moving average and derivatives.

- **Challenges:** (All subteams except one were included in this category.) Generally, they are overwhelmed with data. It is often difficult to determine which metrics are relevant for some of the tasks they carry out during the course of their work. In the same vein, some engineers have mentioned that it can be hard to visualize and interpret the data. (Part of this problem could be related to the difficulty of merging data.) Also, for new engineers it can be particularly hard to navigate and identify pertinent datasets. One of subteams faces additional hurdles relating to unstructured data such as reduced searchability.

4.2.2 Data Visualization tools

The R&D Tools department uses a few different applications for data visualization. Elastic displays event-based data derived from the logs of the various services that they manage. Graphite is used to visualize time series data, while Grafana allows them to combine both time series and event-based data. All three tools allow users to create dashboards, but their capabilities are quite varied.

- **Ways of Working:** Some of the subteams share dashboards as many of them use the same metrics in their work processes. These dashboards are the product of accumulated knowledge. However, the most advanced users of these applications prefer to explore the data and create their own visualizations as it helps them to understand the data and identify abnormalities.

  Kibana is often used to inspect the logs and to search for errors. However, each subteam uses it differently. For example, one of the subteams regularly use Kibana to filter through logs while another only uses it for elaborate queries. The other subteams primarily use it as a searching tool to various degrees. Only one of them uses the application for periodic reporting.

  Within the department, Grafana and Graphite are the most widely used visualization tools. The general consensus is that they are easier to use than Kibana. However, some individuals use python for more elaborate visualizations and for larger data sets since Grafana can only import a limited amount of data (less than 500 data points).

- **Kibana:**
  - **Positive Aspects:** It is an advanced tool with powerful search capabilities. In addition to its own query language, it is also possible to use the Lucene query language. Some engineers have found it to be easier to use Lucene when looking

for specific patterns in the log data.

Users can create graphs in Kibana (e.g. histograms). Also, if the data is structured properly, it is easy to retrieve information about individual devices within the program. Additionally, the auto-complete feature allows users to find a metric even if its specific name is unknown.

- **Negative Aspects:** Despite the powerful query language, many people find it difficult to use. The language itself is difficult to learn and remember. Also, it is not the optimal choice for certain use cases. For example, if you want to look for a URL you have to escape all backslashes, which can be inconvenient if it is particularly long. Thus, some engineers use other query languages like Lucene. However, the general sentiment is that with regular use, one can eventually learn how to use it effectively.

  Browsing the data can also prove to be difficult given the manner in which the results are structured in the UI. In order to inspect each entry in its entirety, the user must click on it. If the engineer suspects that a metric from a different data source is related to an abnormality found in Kibana, it is not possible to merge the data. Thus, correlation is quite difficult to perform within this platform. Also, the program is generally not intuitive.

- **Graphite:**
  - **Positive Aspects:** Many found it to be easy to use and quite useful. The way in which the UI was structured was intuitive given that the metrics were arranged in a folder tree. Thus, even though a query language exists for Graphite, many do not use it given the efficiency of the folder tree structure.

    Data transformations such as moving averages, derivatives and time shifts could also be utilized within the program. To compare different time series, the user would select the desired metrics in order to add it to a common graph. Additionally, any data transformations performed on the metrics in the graph would be applied to all. Thus, the program makes it easy to identify correlations between different time series.

  - **Negative Aspects:** While simple to use, the user interface is extremely dated. Furthermore, it does not allow users to import event-based data. Even though the folder tree structure is helpful, some have found that the program is not intuitive. Additionally, its dashboarding capabilities are very limited, and the program is hard to customize.

- **Grafana:**
  - **Positive Aspects:** In general, the engineers are quite satisfied with it. It is possible to import multiple data sources into Grafana. Thus, they are able to plot both event-based data and time series data on the same graph within the program.

  Also, the engineers feel that is has a nice UI for dashboarding and constructing graphs.
  - **Negative Aspect:** The amount of joining is limited, and it can only extract 500 data points from Elastic.

- **Icinga:** Icinga is a reporting and monitoring service. Within this program, the subteams set a threshold for certain metrics that when crossed would trigger an alert. Sometimes these alerts form the basis of an emerging or ongoing issue with their services that the subteam would investigate.

- **Preferences and Sentiment:** Within the department there is a desire for a more integrated data visualization tool. They would like to see a tool that would allow them to easily access, query and correlate their data. In regards to data visualization, some would like to see a solution that would have a larger diversity of data representations as well as making it easy to switch between different views.

  When it came to the general functionality of the program, we received other recommendations from the team. Some would like a service that could suggest other metrics of interest based upon their previous queries. One user proposed implementing voice control for selection and filtering tasks as it is currently quite onerous to do with the existing software. Another user would like to see an application that would implement the best aspects of the current software: an application that would allow for the easy generation of aesthetic visualizations with the powerful search capabilities of Kibana.

- **Datadog:** While this tool was not currently being used by the department when we conducted the interviews, it was mentioned quite often. It provides more integration between datasets than what is currently possible with Graphite, Grafana and Kibana. The advantage with using this application is that everything is in the same interface, and it has more sophisticated forecasting and anomaly detection than Icinga. Also, the choice of visualization options is comparable to Grafana. Querying within Datadog remains label-based (similar to Kibana), but Datadog offers the possibility of navigating by host and allowing users to find metrics in that manner.

4.2.3 Workspace, Other Services

In terms of workspace, the department is fairly hybrid. About half of the team work within a hybrid arrangement, while the other half are primarily on-site. Also, most employees do not switch subteams frequently, which means that a majority of the team work consistently with the same services.

Even though the category, other systems and services lies outside of the realm of data visualization, it is essential in understanding the larger picture in terms of the role and importance of visualization within their work processes. Many teams use Jira and Confluence for documentation and issue tracking as well as other services such as Zoekt. These other services and systems are also key to their work as a software tools department. Thus, data visualization for most subteams is not their primary task, but is used to support their main goal which is to develop and maintain the services needed for software development at Axis.

4.3 Brainstorming

As a way of organizing the information that we acquired through our literature search and user research, we used the 5 Question Design Framework for IA. This framework served as the underlying structure of our brainstorming process. The responses from the focus groups enabled us to develop different solutions to the categories of *Where*, *Who*, *What* and *Why*. From this brainstorming session, we decided to focus on the problem of how to encode data in a way that makes it easier to understand and navigate with the current capabilities of immersive technologies. Consequently, we decided to conduct another round of literature review focusing on this problem as well as the *How* aspect of our approach.

4.4 Conceptual Designs

After we completed our second round of literature review and the subsequent brainstorming session, we developed four different concepts. Two of which were applications designed with the intent of assisting the engineers with data exploration and identifying relevant metrics. The other two concepts were notification applications. They used smell and sound to notify users of abnormalities.

To express these concepts, we made storyboards and presented them to two focus groups. One group consisted of expert and/or regular users of Grafana, Kibana and Graphite. The other group were made of individuals who used these tools less often. Some individuals in the second group were included because we were unsure as to how proficient they were with the aforementioned tools. Group dynamics were also taken into account as we wanted to ensure that everyone would feel comfortable contributing to the discussion. In the end, we managed to meet with at least one individual from each of the subteams.

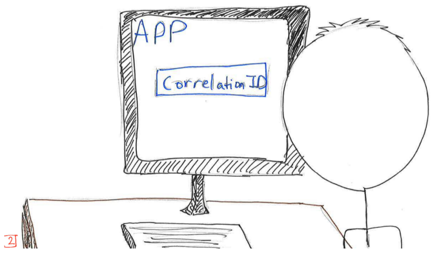
4.4.1 Tree Concept

This was one of the concepts that was designed to facilitate the discovery of relevant metrics (see Fig. 4.2). When an engineer receives an issue from Jira, they could enter the relevant correlation ID or build URL in our application (see Fig. 4.2a). (The correlation ID and build URL function as trackers that could be used to follow a particular process as it travels between the different services managed within the department.) Afterwards, the application would collect all the logs associated with this ID or URL (see Fig. 4.2b). From the logs that were collected, it would parse the fields and create filters based on the fields (see Fig. 4.2c). Examples of different fields would be time, client, host, application, container, etc.

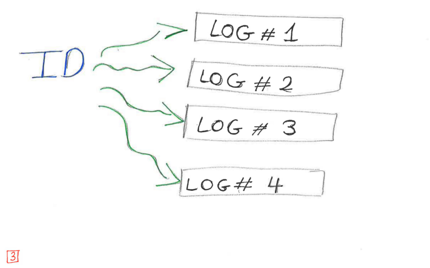
The data would be classified into three main categories: services, containers and applications (see Fig. 4.2d). From this classification, a tree structure would be created. The first layer would be the contain the services, the containers would be placed in the second layer, and the third layer would contain the applications. In other words, the deeper the user delves into the tree, the more specific the data becomes (see Fig. 4.2e).

When a user selects one of the nodes of the tree, a selection menu would appear. From this menu, the user would chose between time series data or log data (see Fig. 4.2f). (For the sake of differentiation, the log data was depicted as a scatterplot graph.) If the user chooses to look at the log data, that metric could be visualized as a scatterplot within a glass box that the user could pick like a fruit (see Fig. 4.2g). If the user wishes to explore other branches in the tree or look at other trees, the metrics could be collected in a fruit basket that would be accessible at any point during the experience. Thus, the user would be able to compare data across teams and services whenever they desired.

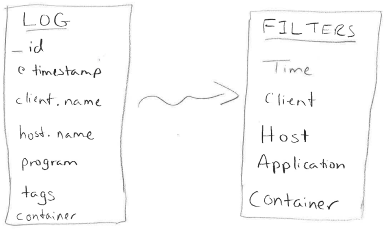
Time series data would be depicted as line graphs (see Fig. 4.2h). The user could take one of these visualizations, adjust the time frame and compare it with a graph of the same metric for a different time interval. They would be able to layer the graphs in 3D space, creating a 2.5D visualization in which the user could determine if the metrics are correlated or not (see Fig. 4.2i). It would also be possible to compare metrics across services in a similar way with the log data. Once the user has a visualization that they are satisfied with, it could be exported to their computer for further use.



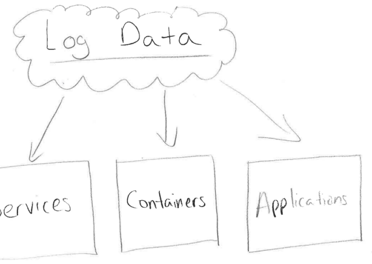
(a) The engineer enters the correlation ID or build URL into the application.



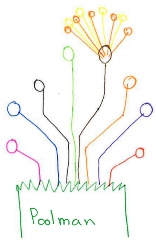
(b) Correlation ID or build URL is used to collect relevant logs and metrics.



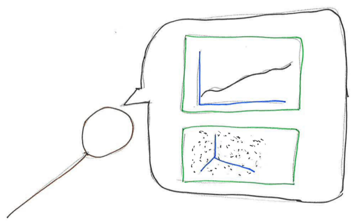
(c) Application creates fields from the log data.



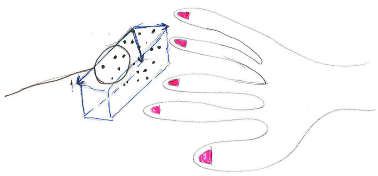
(d) Separation of log data and metrics into application, service and container categories.



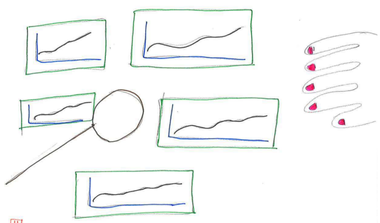
(e) Once the data has been processed, a hierarchical tree structure is created.



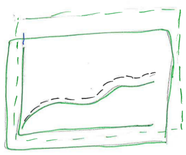
(f) Data selection menu. The user can choose between time series or log data.



(g) Visualization of log data.



(h) Visualization of time series data.



(i) 2.5D comparison of time series data.

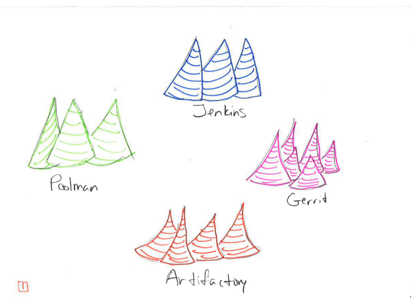
Figure 4.2: Storyboard of the tree concept idea.

4.4.2 Mountain Concept

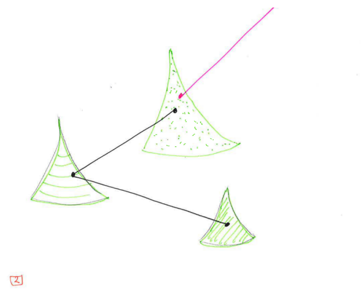
The mountain concept is the second data exploration concept (see Fig. 4.3). It is similar to the tree concept except that the main way of visualizing data is in the form of mountain ranges. Each team is depicted as a mountain range where each mountain represents a different service (see Fig. 4.3a). When looking at a team, it is possible to see the connections between the services within that team as well as connections to services outside of that team (see Fig. 4.3b).

If the user wishes to get a closer look at the metrics, they could hit one of the mountains which would trigger the transition to the cylindrical view (see Fig. 4.3c). In this view, the user is in an egocentric cylinder that would allow them to see the metrics. These metrics would be divided into three categories: application metrics, server metrics and container metrics. For each panel that the viewer can see within the cylinder, there are trees representing the hierarchy of metrics within each application. The first node would contain aggregate metrics for each application and the child nodes would contain metrics that describe the functioning of the application on an increasingly granular level (see Fig. 4.3d). When the user selects one of the nodes, a visualization of the different metrics would appear. In the example depicted, the user would see both time series and log data (see Fig. 4.3e).

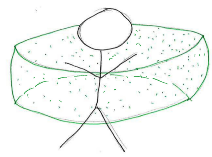
It would also be possible to view the general network of each of the services and how the different applications, servers and containers are connected with each other. From the cylindrical view, the user would push the cylinder surrounding them forwards, which would cause it to shrink to the size where it could be easily manipulated by the individual. In this mode, the user would be able to see a network graph that illustrates these connections (see Fig. 4.3f).



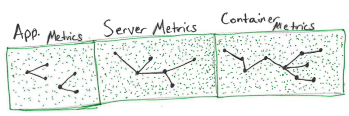
(a) Each team is visualized as a mountain range.



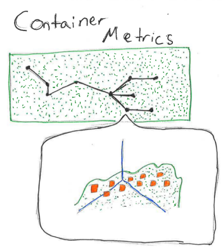
(b) Connections can be visualized between services.



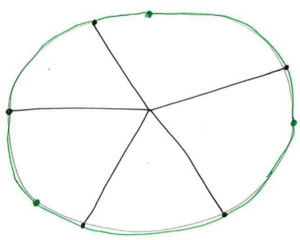
(c) User would have a cylindrical view of the metrics.



(d) Application, sever and container data are presented within the cylinder.



(e) Visualization of graph with both log and time series data.



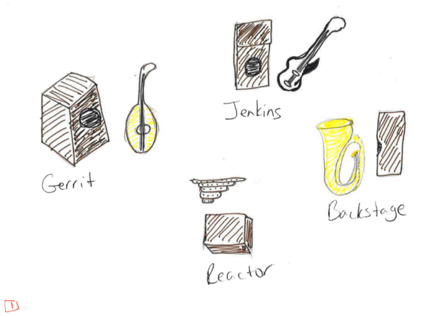
(f) Cylindrical network graph. Connections are shown between applications, containers and servers within the service.

Figure 4.3: Storyboard of the mountain concept idea.

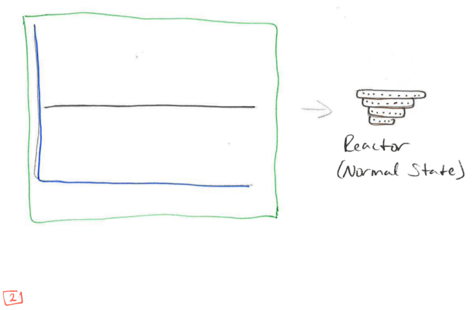
4.4.3 Soundscape

The intent behind this concept was to find a method of notifying engineers of abnormalities in their metrics without relying on the visual modality. For this concept, we assigned distinct soundscapes to each team (see Fig. 4.4). In our sketch, we assigned one genre to each team: traditional Persian music, traditional Native American pan flute music, rock music or brass band music (see Fig. 4.4a). When the metrics are within normal limits, the user would hear sounds that are congruent with their subteam (see Fig. 4.4b), but when there is an abnormality the team would begin to hear sounds outside of their assigned genre (see Fig. 4.4c).

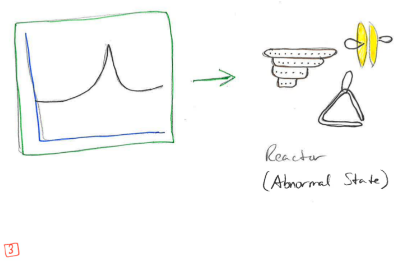
We also presented an alternate idea where if there was an abnormality caused by another subteam, then the engineer would begin to hear sounds coherent with that subteam's soundscape. For example, if the Reactor team normally heard Native American folk music, but then began to hear an electric guitar riff, the engineer could recognize that malfunctioning services within the Jenkins team are affecting their services.



(a) Each team is assigned a unique soundscape.



(b) When the metrics are within a normal range, the assigned genre would play.

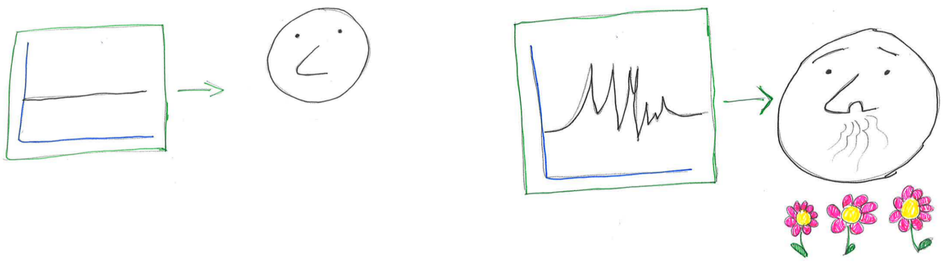


(c) When abnormalities are present, sounds outside of the team's soundscape would be heard.

Figure 4.4: Soundscape Concept

4.4.4 Smellscape

We have also introduced another concept which uses different scents to notify the engineers (see Fig. 4.5). When the metrics are in a normal range, the user would not be exposed to any smells (see Fig.4.5a). However, once an abnormality appears in the metric, for example a sudden sudden peak, a scent would be released (see Fig. 4.5b). This scent would indicate that a service may be malfunctioning and should be investigated.



(a) No smells are released when the metrics are in a normal range.

(b) Once an abnormality is detected, a scent is released.

Figure 4.5: Smellscape Concept

4.4.5 Feedback

The concepts were presented to two focus groups: one of frequent users of their current visualization tools and one of infrequent users. Each group had four participants from different subteams. (Two of the individuals within the user group of infrequent users had a competence level that was unknown at the time of their selection.) Within each session we asked for feedback on the concepts, but also asked the participants to contribute their own ideas. These sessions also served as a way to confirm that our understanding of how R&D Tools worked with and handled data was correct.

General feedback

For all participants, the tree concept was more intuitive than the data mountain concept. However, there were some aspects of the data mountain concept that they liked such as its overview capabilities. Additionally, the exploratory capabilities of the tree concept were also positively received. Thus, it was suggested to combine these features from the tree and mountain concepts.

Other ideas were also discussed during these sessions. One of the participants mentioned that when an error is discovered they are interested in the origin of the issue. For this individual, a graph of system nodes and errors would be more useful than the concepts that we had presented. From our discussions, we also concluded that the application should be flexible enough to be used in different contexts by different teams. Not every team was only concerned about data from their services, but from other parts of the CI/CD pipeline that were not covered by our designs.

Concerns over sensory overload were also mentioned not only in relation to our concepts, but also in regards to their current worklife. For some engineers they often found their existing setup to be quite visually cluttered and mentioned how they have also become accustomed to tuning out certain sounds. Thus when it came to the notification concepts, our participants generally preferred the smellscape over the soundscape.

Tree Concept feedback

The tree concept was not very useful for one of the subteams as they were one of the few subteams that did not manage a service, but focused mainly on monitoring a system. However, they mentioned that it could be a great tool for onboarding new employees. It had the potential to help them understand the different services and networks that the team manages as a whole.

For the other subteams, the tree concept was more useful. Within the application, the user could see where they needed to look in order to find relevant metrics. Also, the concept helped to narrow the search space. Many participants viewed these features as beneficial.

While discussing this concept, we had realized that where we had envisioned the services to be placed in the data source hierarchy was incorrect. Instead of them being a parent of the containers, they should instead be a child of the hosts. Additionally, it was mentioned that containers and applications have a one-to-one relationship and that services and applications are essentially the same within the department. Thus, with this information, we revised the way in which we structured the data in our proposed application.

Mountain Concept feedback

Overall it was perceived to be too abstract and nonintuitive. However, our participants were happier with its aesthetics compared to the tree concept. To improve upon the idea, some of the subteams mentioned that it would be useful if the small mountains could report the status of different parts of the system. Another suggestion was that each mountain could represent a server cluster instead of a service.

Some participants also thought that mountains could be used to visualize network or service activity as it is currently difficult to describe how everything is connected. It was mentioned that this application could also be useful when they experience issues related to certain processes or server outages as it would assist the engineers in determining what is relevant to the current problem. When a server or a tool is misbehaving, the mountains could flash or Icinga alerts could be incorporated in some manner. Additionally, the prospect of modelling the absence of activity was also discussed.

During our conversations, it was mentioned that the subteams prefer to have one application which contains all of the features they need instead of three. Being able to customize the representation of the metrics was desirable, but the biggest problem they faced was related to finding metrics related to the issue they were troubleshooting and simply knowing which metrics were stored in their systems. However, one of the subteams doubted this solution would be beneficial for them as their team was quite experienced and already knew how to find the metrics they needed.

Sound Glyph Feedback

For this concept, sensory overload was a concern as it could easily become overwhelming if several errors were detected. To help avoid this, it was suggested that a specific sound could be mapped to each issue. It was thought that this mapping could facilitate collaboration as everyone would know the issues currently being handled based upon the sounds being played. Also, many thought that sound notifications could be better than visuals, especially if the sounds were differentiated in some form as they normally received the same notification sound from Outlook regardless of message type. However, it was also noted that some individuals would prefer a silent work environment.

Sound design was also considered. AI-generated sounds were brought up as an option as people could become accustomed to the same sound. In support of this viewpoint, one of the participants mentioned that they would not want to listen to the same sound all time. Implementing an ambient soundscape was considered a better option as it would allow the engineers to listen to the streams of the other teams. Other suggestions included linking Outlook notifications to the system and integrating Spotify in some manner.

Smell Glyph Feedback

We discussed the feasibility and preferences of the participants in relation to this concept. Some suggested that a burning scent could be released when an abnormality was discovered within the metrics. It also proved to be quite an interesting concept for one individual as they noted that they often ignore audial and visual cues. However, such an application could be problematic for those with smell sensitivity. Individuals could become nauseated and dizzy, which would not be conducive to a safe working environmment. Additionally, the risk of smell contamination was considered.

4.4.6 Summary of Feedback

With the feedback we received on our various concepts, we were able to decide on the features that would be included in our application. Given that one of the main issues was finding relevant metrics when troubleshooting, one of the features we included was allowing the user to input a key linked to a specific issue in order to assist them in finding relevant logs, metrics and data sources. Another frequent comment was that the data mountain concept could provide a good overview of the health of the infrastructure and servers. This would also allow it to be utilized as a type of living diagram of the services and servers that the R&D Tools department maintains. Consequently, we decided to develop an application that was a combination of the tree and the data mountain concept.

Given the amount of data that we would have to display and the very limited field of view provided by current AR hardware, we felt that it would be best to develop our application for a VR HMD. Placing the application in this setting would allow for the user to be able to efficiently navigate through our data landscape and avoid issues related to color perception and environmental interference. Taking into account sensory overload, the VR environment would allow us to reduce the number of conflicting signals in the user's surroundings which would help them focus on the task at hand.

Chapter 5
Lo-fi Prototyping

Our prototype was created in ShapesXR, which is a VR prototyping tool exclusively for Meta Quest headsets. It allows for simultaneous collaboration and quick ideation within a VR space. Even though it lacked user interaction at the time, it provided a decent platform for spatial design, which allowed the engineers to get a feel for the virtual space.

Given that it was not possible to create a fluid user flow in the prototype, it was presented over a series of static scenes. Despite its limitations, the prototype generated plenty of actionable feedback during the testing phase. Some of these recommendations led to changes in its design.

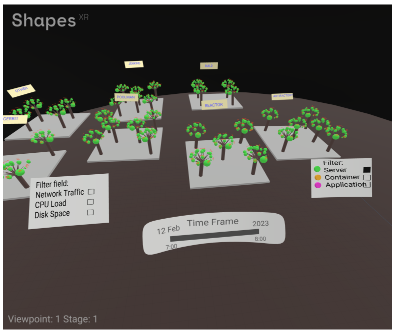
5.1 First Version

In the first scene, we presented the overview mode (see Fig. 5.1a). Here, it was possible to see all of the teams, their services and servers. At the center of user's field of view was a time slider where the user could adjust the time frame they wished to view. On the left was a filter menu where the user could select the metric they wished to view, and on the right a filter based on the data source. The data source could be either the servers, containers or applications. (In this prototype, the server data source was selected.) In front of the user were different tables. Each table represented a subteam, and the trees on the each table were the services managed by that subteam.

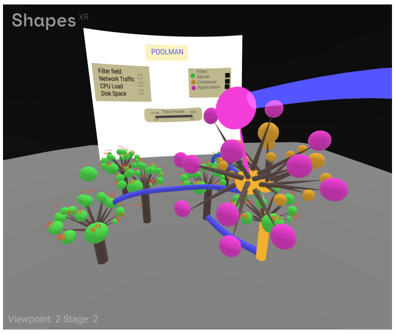
The second scene depicted our concept of the troubleshooting mode (see Fig. 5.1b). This mode would allow the user to input an identifier connected to the issue that they were investigating on the home screen. (The home screen was not implemented in this prototype.) When troubleshooting, the relevant nodes and connections to other programs would be highlighted (i.e. the contrast and saturation of the nodes and the blue connections would indicate the servers, containers and applications pertinent to the current issue). The filters in the overview mode (5.1a) could also be applied in this context as well. For this scene, we used one of the subteams (Poolman) to illustrate the proposed user experience.

The third and fourth scenes demonstrated how the user could traverse the tree. In the third scene, the user would select the node and a menu would appear that allowed the user to chose to view the metrics, aggregate the metrics or to go deeper into the tree (see Fig. 5.1c). The fourth scene depicted the user's viewpoint when viewing the metrics of that node (see Fig. 5.1d). At this stage, a menu would appear where the user could select log or time series data. (The time series data was represented as a line graph, and the log data was represented as a 3D scatter plot.) Once those selections were made, another series of menus would appear that would allow them to select the particular log and metric data that they would like to see. As the user cycled through the options, the graphs would display the selected metric.

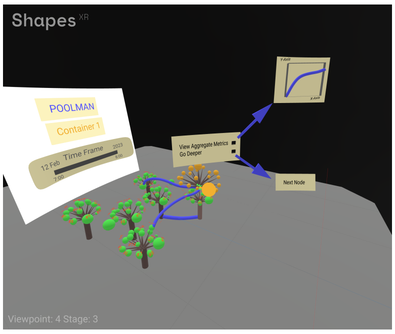
For the final scene of the prototype, the data space was depicted (see Fig. 5.1e). It was a space in which the user would be able to explore the data and manipulate the metrics that they had chosen in the previous mode. (This space could also be accessed if the user chose the option "view all" in the previous scene.) Within this workspace, the engineers would be able to identify correlations and metrics of interest as well as create visualizations to share with others. It would be possible to do visual comparisons of graphs by overlaying them and apply mathematical transformations to the data displayed on the graphs. Additionally, it would be possible to scale and transform the graphs in order to support embodied data exploration.



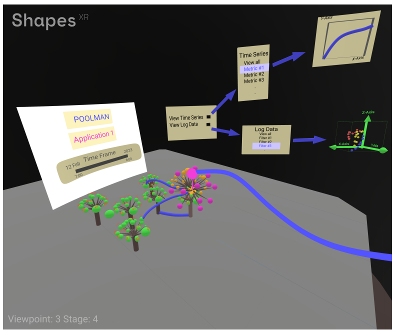
(a) The Overview Mode



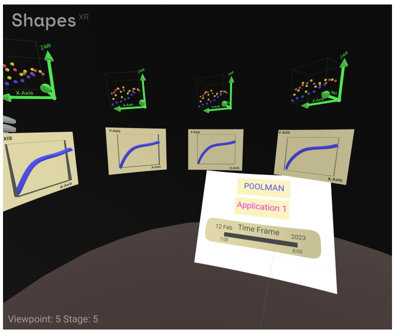
(b) Troubleshooting Mode



(c) Tree Traversal Menu



(d) Selecting Metrics



(e) The Data Space

Figure 5.1: The First Version of our Prototype

5.2 Second Version

In the first version of the prototype, the menus and filters were billboards that were present in the space (see Fig. 5.1). However, we felt that if the menus and filters were controller-based, they would be easier for users to manage. This change would also reduce the amount of visual clutter in the prototype.

For the second iteration of our prototype, the first scene showed the menu options that could be expected when starting the application (see Fig. 5.2a). On the far left of the menu was a button that would transport the user to the data space. The middle button would be used to open the filter menu, and the camera button on the right would allow the user to take a screenshot of their visualizations.

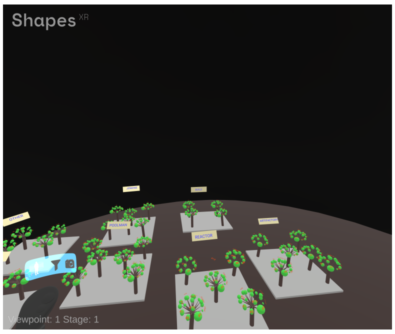
The second scene depicted the filters that would be available (see Fig. 5.2b). They were the same as the filters in the first prototype. (It should also be noted that all menus in the prototype were intended to be reduced for the user when they were in the experience.)

In the third scene, the troubleshooting mode was displayed (see Fig. 5.2c). (The Poolman subteam was also used in this example.) When troubleshooting, the relevant nodes for the issue would be highlighted and connections to other programs would be visualized in the same manner as was mentioned in the first version. The filters could also be applied in this context as well.

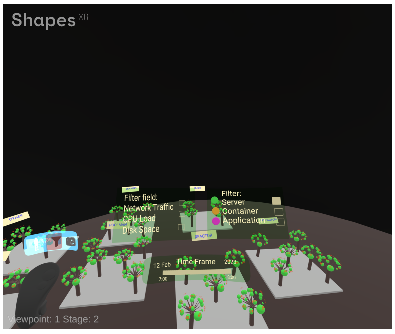
For all views except for the general view, there would be a tree map to the left of the user's field of view that indicated the user's position in the tree (see Fig. 5.2c). Since the troubleshooting mode was presented in this scene, the branches and nodes that were not relevant were grayed out, while those that were relevant retained their full color. When outside of troubleshooting mode (It can be toggled on or off by pressing the wrench button on the far right of the controller menu.), the entire tree map would have a normal coloration.

The fourth scene presented the view the user would encounter when they wished to access the metrics (see Fig. 5.2d). To select this menu, the user would press a button when clicking on a node. Visualizations of both log and time series data would appear, and the user would be able to select which metric they would like to view. It would be possible to chose one metric at a time or all of them. To see all of the metrics, the user would select the view all option and would be transported to a different virtual environment, the data space.

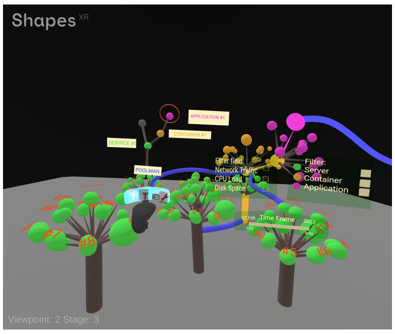
Within the fifth scene lies the data space (see Fig. 5.2e). This would be a space where the user could bring in the metrics they have selected, which is similar to the data space concept in the first version of the prototype. It would allow them to combine and compare the metrics. Additionally, users would be free to place their data visualizations anywhere in the virtual space.



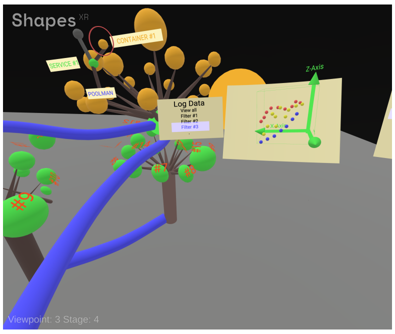
(a) The Overview Mode



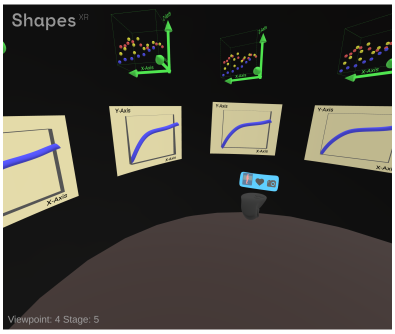
(b) Overview Mode: Filtering Options



(c) Troubleshooting Mode



(d) Selecting Metrics



(e) The Data Space

Figure 5.2: The Second Version of our Prototype

5.3 Testing

We tested our lo-fi prototype with a subset of the team at R&D Tools to not only validate our concept, but also as a way to glean additional user requirements that we may have missed in previous stages. While the participants were in the experience, we encouraged them to explore. A semi-structured interview was incorporated into the testing session and mainly related to the spatial design of the application, the different features we have included and whether or not their workflow would be complemented by the software.

5.3.1 Test Setup

The prototype was developed in ShapesXR with Meta Quest headsets, and it was demonstrated on a Meta Quest 2. If one of our participants could not use the headset for any reason, we also had a desktop version as a backup.

In total we had scheduled seven sessions with one participant per session. Each participant represented a specific user profile or subteam. In this way, we would be able to ensure that it would be useful for as many user groups as possible.

The prototype consisted of five scenes and four viewpoints. Thus, some scenes shared viewpoints that the participant needed to navigate through. In order to provide guidance to the participants, we used an Oppo A72 mobile phone with the Oculus app installed. This setup allowed us to cast into the headset and see the participant's view of the application. We were then able to lead the participant to the appropriate viewpoint and scene.

The testing area was our two-person office. One portion of the office was devoted for testing the prototype. The testing area was 118cm x 156cm, and the Guardian boundary was within the space.

5.3.2 Test Procedure

Each participant was seated in an office chair and was instructed as to how to move through the different scenes. One person would help the participant put on the headset and ensured that the participant was able to successfully enter the application. We explained each scene to the user. After each scene, the participant was asked a set of questions about how the data was represented, the effectiveness of the representation, the proposed workflow and whether or not the participants could envision themselves using this application within their work process (see Appendix C). At the end of the testing session, the user was asked about whether the data was represented in a satisfactory manner or if they saw any immediate problems with the application.

5.3.3 Pilot Testing

For our pilot test, we selected one of the engineers in the R&D Tools team that we did not plan to include in this round of testing. This test was conducted on a Meta Quest 2 headset, and we were able to cast to an Oppo A72 mobile phone in order to see what the participant was viewing at the time. Since we wanted to ensure that our procedure was sound and estimate how much time we should allocate for each testing session, we subjected our participant to the full testing protocol.

While the pilot test did run smoothly, we changed some of the questions included in the interview portion of the test. This was due primarily to repetition. We realized that some of the questions we had originally included covered the same subjects. Thus, we removed them for the sake of efficiency. [1]

5.3.4 Results

The testing of our lo-fi prototype was conducted over the course of three days. After each test session, we collected the feedback and discussed whether or not it could be incorporated into the lo-fi prototype based on the feasibility of the change, whether it would lead to an improvement in the user experience and the number of people who have requested the feature. Therefore, the prototype was revised twice as a result of our user testing process.

Revisions Undertaken during Testing

- **Day One:** We received feedback that the graphs within the data space were too far away from the user. Also, looking at the aforementioned graphs induced neck strain. Thus, we changed organization of data space, moving the graphs closer to the user and then adjusted their height to remedy the issue.
- **Day Two:** For two out of three of our participants, the data representations within the prototype itself proved to be highly distracting as they felt they were too unrealistic. Even though lo-fi prototypes are not intended to be realistic, we felt that too much time was spent discussing the accuracy of the sizes of the nodes in the application. Our intention with these sessions was to evaluate the high-level concepts such as feasibility and suitability instead of its aesthetics. Thus, we slightly randomized the size of the nodes in scenes one and two to make it more realistic.

  The scatterplots in scenes four and five were changed from a 3D representation to a 2D representation as it faced a similar problem to the node representation. A majority of our participants (two out of three) were distracted by the presentation of log data as a 3D scatterplot instead of a 2D scatterplot. Despite explaining to our participants that we were not evaluating the visual aspect of the application, an inordinate amount of time was spent discussing this topic. To avoid derailing the interviews in further testing sessions, we changed the existing scatterplots and added another scene with

[1] The final version of the interview questions are in Appendix C.

3D scatterplots to show the visualization possibilities.

From these sessions, we also decided to add a feature to our design. One of our participants mentioned that it would be nice to allow the user to select multiple metrics in the overview scene that could be viewed simultaneously. As we had not considered this idea, we began brainstorming different possible implementations of this feature.

- **Day Three:** No changes to the design of the application were needed after the testing session. However, in addition to the user tests, we also showed the preliminary tree concepts that would allow users to view multiple metrics to our participant. In this last test, we ran into technical difficulties casting to the phone and had to cast to the computer, but it did not affect the user's experience of the prototype.

General Findings

From our interview questions, we were able to derive a few insights. Flexibility is important both in terms of data representation and navigation. Not every subteam needs the same architecture or data for their trees. Some would prefer trees to be instances of ongoing processes instead of servers and applications. In terms of navigation, some would like to move around the virtual space, while others would rather like to manipulate the space itself. Similar preferences appear when our participants describe how they want to manipulate the visualizations in the data space. Some would like to move the visualizations themselves while others would prefer to physically explore the space and become immersed in the graphs.

While the content of the application was generally understandable, there were some aspects of the application that proved to be confusing. Within the overview mode most people understood what was presented, but some were confused. The same was said of the network graph in another scene. To reduce confusion, it was suggested to add labels in order to clarify what was present scene. In regards to the network graph, it was recommended that we could add labels to the endpoints of each of the connections and indicate the direction of the data flows.

When it came to the organization of the data, it was noted that the container and application level could be compressed into one node as both levels would contain the same data. Another suggestion was to compare different metrics and time frames in the overview mode. To visualize multiple data points, it was recommended to cluster them according to the time frame and allowing the user to zoom into a particular cluster. The size of the cluster could reflect a value of some sort.

In regards to the manner of data representation many felt that it was suitable. One suggestion was that it could also be used to investigate the frequency of errors. For example, the user could use to see the nodes animated over a particular time frame in order to observe the change over time. It was also brought to our attention that some of the metrics that they record are rates, which would require our program to aggregate them. However, it was mentioned that the visual nature of the application could lead to visual overload and reduce the learnability of the application. Perhaps the use of labels for summary statistics such as how many servers are running on this service, would make it easier to interpret, which was proposed by some participants. Others suggested that when the user is in troubleshooting mode, removing unrelated services could also reduce visual clutter. Despite this concern, some participants were open to more visual elements. Instead of text labels for the different subteams, some would like to see the use of logos.

We also received some suggestions in regards to the data represented. In its current form, it was mentioned that it would be useful for investigating the frequency of errors that occur. Some of the participants wanted to see the incorporation of alarms from Icinga and other monitoring software. Others said that the application could be used to monitor Kubernetes clusters. Also in line with our research, it was mentioned by some participants that VR would be best for 3D data instead of 2D. Some participants were unsure if the visual fidelity of a VR environment would allow for them to read the logs. (For all of the subteams, logs are one of the main sources of data that they peruse when troubleshooting.)

There were some concerns regarding the menus that arose during this round of user testing. The titling of the filters were ambiguous to some. One of the suggestions offered was to label one of the filters data sources and the other with a more appropriate name for its content. For the base level of the controller-centric menu, there was some confusion as to the meaning of the heart. A suggestion we received was to use a star icon instead as the participants noted that the star had a stronger association with favorites than the heart.

The data space (a space for the engineers to explore data and create visualizations) also garned some feedback. It was clear during our sessions that it could be useful for some teams, but not for everyone. Some users suggested that it should be possible to construct a graph with a dual y-axis where one can compare the same metric, but on different scales. Another suggested representation was a timeline of events where the y-axis could be related to specific jobs being processed by their service. We also received some feedback about the proposed organization of the graphs in the data space. Some users said that instead of displaying the graphs around the user, the metrics could be arranged in the form of forward-facing rows.

Summary

Overall, the participants saw the potential of the application, but were skeptical in regards to the practicality of the implementation and if it would be a benefit to their workflow. The weight of the headset was also mentioned as a drawback.

In regards to the application itself, we incorporated some of the feedback we received into the hi-fi application. Given that the application was meant for independent use and the confusion that could arise due to the data representations, we have identified a need for a help feature. Also, the wording on the menus could be clarified and its appearance to be made less abstract so that it is easy for the user to map colors to the particular data source. Incorporating more labels into our application could also reduce confusion, but we must also be aware of the potential of such labels to increase visual clutter. However, we identified opportunities to utilize the existing user interfaces in a better way such as using the tree map to display summary statistics of the service (e.g. the number of servers currently running).

Chapter 6
Hi-fi Prototype

The hi-fi prototype was created with Unreal Engine 5.1.1. Unreal Engine is an advanced game engine with powerful lighting and rendering capabilities. This software is not only used in game development, but also in the film, TV, automotive and simulation industries [43]. This prototype was created with the Linux version of the engine. We had used the VR template included with the software as the basis for our project, which was only available as a Blueprint project on Linux. (Blueprint is the scripting language for Unreal Engine). Since our application needed to display data in real-time and send RestAPI calls every second, we created C++ classes to provide this function within the application.

6.1 Data Sources

In order to retrieve the data needed to build the trees, we used RestAPIs to query two applications: Icinga and Graphite. With Icinga we were able to build the tree itself as it contained data relating to the corresponding subteam and service for each server we wished to display in the application. Graphite provided us with server data.

6.2 Building the Environment

To construct the environment, we used a pre-made, free Unreal assets package from the Unreal Market Place: "Art, Fashion, Automotive Galleries and Showcases" [44]. This package provided us with many assets that we used to design the environment. We also tried to minimize the number of assets in the scene to avoid user distraction. The environment itself was a singular, large room. In order to create the room, we used some walls with lights from this package to create an aesthetic and bright indoor environment (see Fig. 6.1).

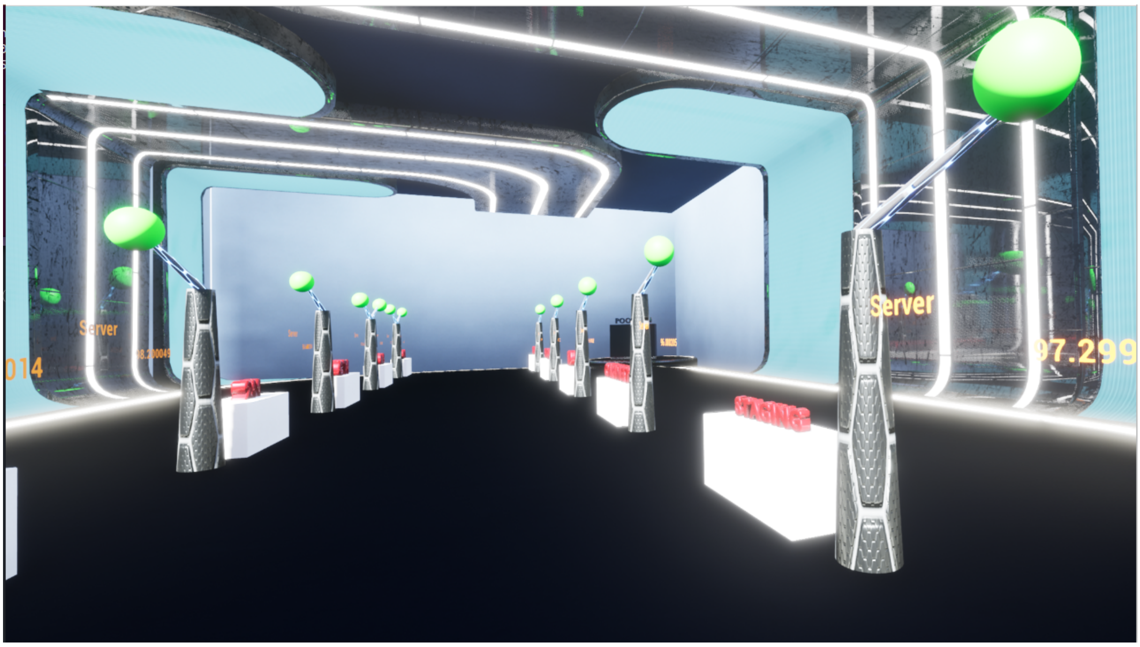


Figure 6.1: A screenshot of our hi-fi prototype.

The trees in our environment were created with Blender, an open source 3D creation software application, in which we modeled the tree trunks, branches and nodes. The site labels, which denoted the location of the servers, were also made in Blender as 3D Text. Within the VR environment, these labels were placed on top of one of the stand meshes from the package, "Art, Fashion, Automotive Galleries and Showcases."

From Icinga, we were able to access the host groups. These host groups provided the data used to build the trees and spawn the appropriate site name mesh for each tree. Each host group contained data such as the site at which the servers were located, the service that was running on them as well as the subteam that maintained those servers. Within the application, we only queried the API for the hostgroups relevant to the Poolman subteam. The site name was assigned to the trunks and the server names were assigned to the branches and nodes. When the trees were spawned within the application, the trunks were connected to the relevant branches and nodes based on the hierarchy from the Icinga application. The stands with the site names as 3D text were also located next to the corresponding trunks.

Host names from the servers were used to query the Graphite API for related server metrics. Due to the time constraint, we limited the query to a single metric: the average percentage of the CPU that was idle. Once a second, the application would query the API for the value of this metric. The size of the nodes were proportional to the amount of CPU that was idle and would update once a change in value was registered. (The more CPU available, the larger the node.) 3D text labels were created in Unreal Engine that displayed both the host name and the value of the metric under each of the nodes in our application (see Fig. 6.2).

Figure 6.2: A closer look at one of the trees in our hi-fi prototype.

6.3 The Headset

We used the Valve Index headset for our application. We ordered the Valve Index VR Kit, which came with two base stations, hand controllers with grip detection and other needed accessories such as cables and a lens cleaning kit. In order to ensure that it is accessible to as many people as possible, we also ensured that it came with a headset face gasket as well as a headset cradle adapter to make the device suitable for people with small heads.

Setting it up in Linux was a bit more difficult than expected as the system did not recognize the Valve Index as a HMD. In order for the system to recognize the device and run the display at a reasonable frequency, it required us to find the correct system settings and GPU configurations. Our hi-fi prototype was developed with the Ubuntu 22.04 OS and the X.Org Window System. Thus, we used the xrandr configuration utility for X.Org to find the headset and explicitly set it as a non-desktop display. This setting allowed the system to recognize the headset. Afterwards, it was necessary to set it as a joint display and set the refresh rate to 144.0Hz. This refresh rate also had to be set using xrandr to ensure that the headset ran at that frequency. Since our GPU was from Nvidia, we used the Nvidia X Server Settings application to ensure that the GPU would run with a preference for maximum performance.

After adjusting the settings, it was able to run at a reasonable frequency, but there was still some minor lag when the user turned their head. Unfortunately, we were not able to optimize the GPU configurations to our satisfaction as the options for the NVIDIA X Server settings were quite limited. Ideally, we would have wanted to change the number of frames that were held in the GPU before rendering to one, but were unable to do so.

Within Unreal Engine, there were also other adjustments that could have been made to ensure a smoother experience. However, we did not have the time to find the optimal settings for the parameters recommended by Unreal for XR experiences. Thus, we decided to use the default settings provided by Unreal Engine in the VR Template.

6.4 The Application

It was designed as a seated experience due to the limited space that we had in our office and lack of a proper guardian boundary. (The guardian boundary was not implmented as it required additional setup on our Linux system that we did not have time to implement.) To enter the experience, the user would put on the Valve Index headset. Afterwards, they could navigate the environment with their hand controllers by rotating or teleporting (see Fig. 6.3). The participants could also turn their heads as well.

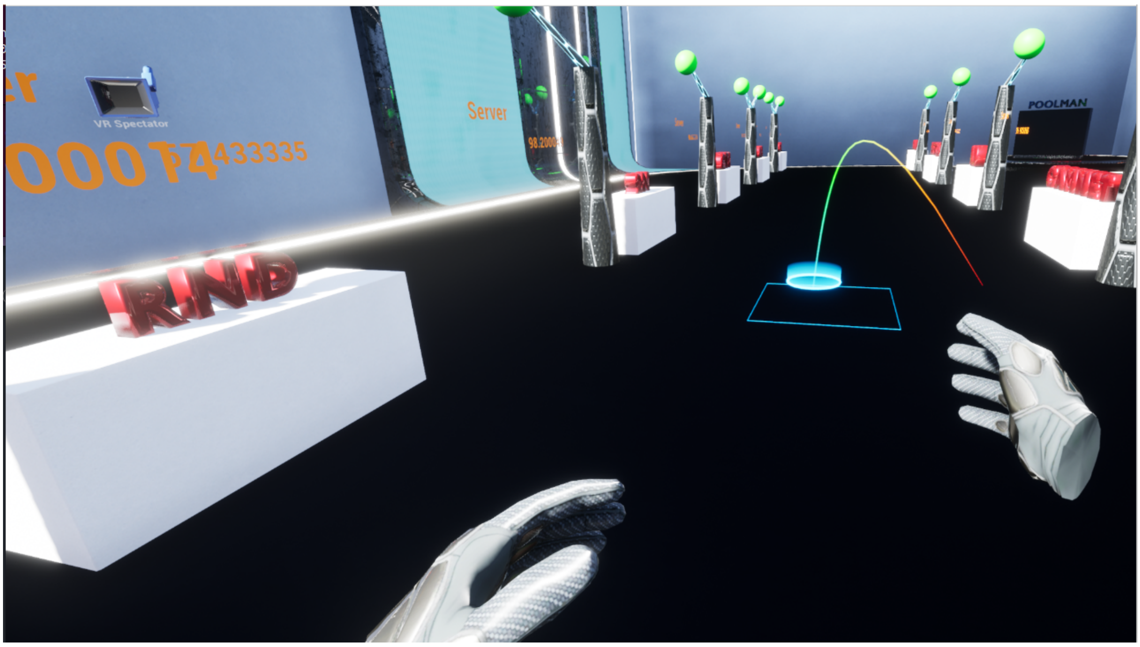


Figure 6.3: Teleporting within our hi-fi prototype.

In the environment were a set of trees and stands that represented the different sites and the servers that the Poolman team monitored as well as the amount of CPU available for each of the servers. The environment was relatively bare except for those elements and a sign that stated Poolman, which was intended to inform the user as to which subteam the servers and sites referred. Users would be able to see the nodes expand and the values of the metrics change in relation to the value of the metric over time. It was unfortunately not real-time due to the practical constraints of the data pipeline within the department, but every ten seconds, the values would change.

6.5 Testing Overview

To test our prototype, we chose six R&D Tools engineers as our participants. Given the limitations of the prototype, we selected individuals who were part of the Poolman team or were familiar with Graphite. A screening questionnaire was sent to all participants for the purpose of creating a participant profile (see Appendix D). This questionnaire was also sent with the intent of filtering out individuals who were sensitive to motion sickness as the experience itself would lag.

The tasks given to the participants consisted of finding the servers under the Poolman team with the highest and lowest CPU load both in our VR application and in Graphite. Half of the users started in our VR application and the other half started in Graphite. For the Graphite portion of our test, we provided all participants with a list of Poolman servers.

After the tasks were completed for each application a SUS questionnaire was provided. Additional questions were asked after the use of the VR application that addressed perceived effectiveness, user distraction and sensory overload. Also, concluding questions were asked about which application the participants would prefer to use for this task and if they had other comments about the tasks or applications. The goal of our user test was to measure the following:

- Satisfaction with the VR application vs. Graphite within the context of this task.
- Efficiency: How long does it take to complete the task? And are they able to complete the task correctly?
- Learnability: How long does it take for users to be able to effectively use the VR application?

6.5.1 Participant Profile

Since we aimed to provide a safe testing environment, we aimed to exclude users who were or potentially could be sensitive to motion sickness. This was due to the fact that our application had relatively high latency due to poor tuning. Unfortunately nearly all potential test subjects either had motion sickness or were unsure if they suffered from motion sickness. Thus, we decided to allow them to participate in the test as the participants had expressed a desire to try the VR application, but we re-iterated that if they felt sick in the VR application to let us know and we would stop that segment of the test.

When it came to demographics, there was a reasonable spread in terms of age, but not in terms of gender. All of our participants were male, but their ages spanned a range from 18 to 69. Two were between 18 and 29, three were within the 30-49 range and one was in 50-69 range (see Fig: 6.4).

We had also asked them about their background with VR applications, IA tools and the three main data visualization tools that they use in their department (Graphite, Kibana/Elastic and Grafana). Two have used VR applications before, but not IA applications. It is shown

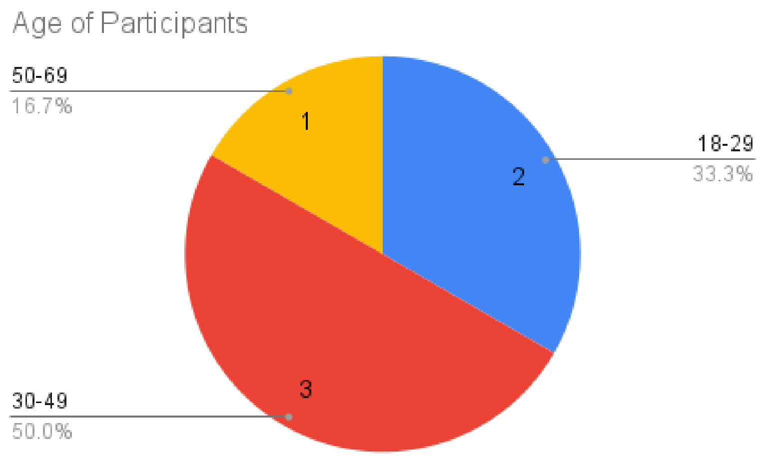


Figure 6.4: Participants' age distribution.

in the questionnaire responses that all of the participants have experience with Graphite, Kibana, Elastic and Grafana. However, the frequency of usage varies (see Fig. 6.6). Participants two and five never used it. Participants one and six used it once a week, while participant three used it two to three times per week. Participant four used it four to five times per week.

Additionally, we asked them about their height. This question was asked to determine if their perception of the environment and the trees was height dependent. Based on the results, our users were between 171 and 189 cm tall, and the average height was 182.3 cm (see Fig 6.5).

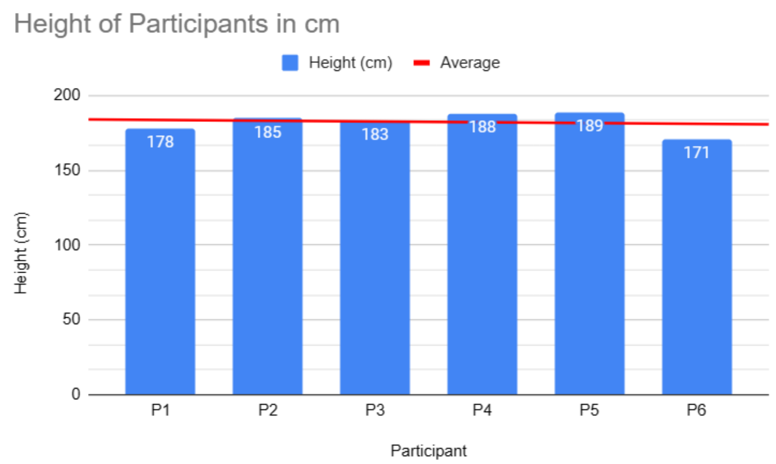


Figure 6.5: Participants' heights. The average height of 182.3 is depicted with a red line.

Since we aimed to compare how efficiently the tasks could be completed with both Graphite and our VR application, we categorized them based on their experience with Graphite.

From the questionnaire results, two of our participants used it regularly, two used it once a week and the other two were not using it at all (see Fig. 6.6). (**Note:** We asked about all three programs to prevent our participants from anticipating the tasks we would ask them to complete.)

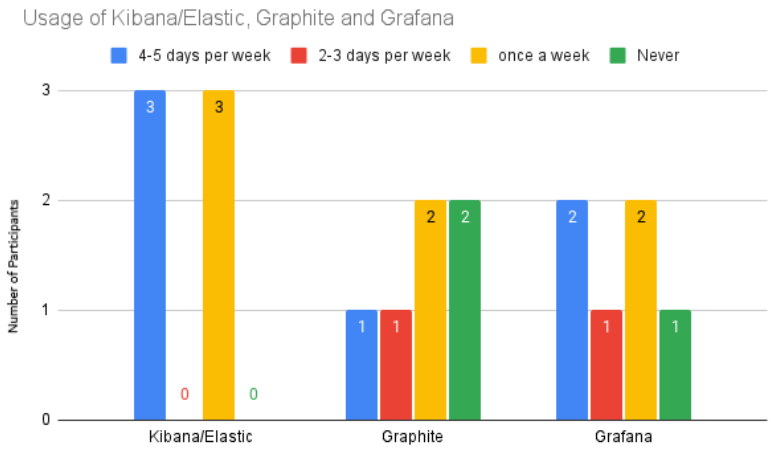


Figure 6.6: Participants usage of tools.

Other Participant Characteristics

- In regards to their backgrounds, four of our users were working within Poolman subteam at the time: participants one, two, three and five. Additionally, four of our users have been working in their current subteam for zero to two years (participants two, three, four and six), while the other two have been present in their subteam for three to five years (participants one and five). However, only participant six was not familiar with Poolman specific metrics.
- Out of six participants, two wore glasses: participants two and six.

6.5.2 Task List

For user testing, the participants were asked to complete the following tasks with our VR application and with Graphite.

1. Identify the servers with the least and the most average CPU load with Graphite at a specific time.
2. Identify the servers with the least and the most average CPU load in the VR application at a specific time.

To keep the comparisons fair for each task, the server with the highest CPU load was always the outlier, while the server with the lowest CPU load was harder to discern. The times were chosen so that this generalization would remain true across both applications. Part of the reason for this choice was due to the data available during testing. When we conducted our user tests, the server with the highest CPU load was always the outlier while the servers with the lowest CPU loads were not always easy to distinguish from each other.

For each task, the server for the lowest CPU load was different between Graphite and the VR application. Unfortunately, due to data constraints, since we had used real-time data for our tests, the server with the highest CPU load was always the same regardless of the time and thus was the same for both applications. The difference in answers between the applications was implemented to prevent participants from gaining an unfair advantage in the second part of the test.

6.5.3 Data Collection

For the purposes of evaluating the performance of our VR application with Graphite, we captured screen recordings of both the Graphite and VR application as well as video recordings of the participants themselves. The screen recordings were used to evaluate how well the participants could navigate both the VR environment and the Graphite application and their focus on the tasks. Video recordings were used to evaluate the participants physical and mental state and to capture anything that was missed in the observations.

The SUS questionnaire and the focused interview questions for the VR application were used to compare the usability of our application to their existing tool, Graphite. They provided both qualitative and quantitative data. The focused interview questions were in service to our overall thesis questions. These questions focused on the perceived effectiveness of VR in terms of visualizing data and if such an application could be designed in a way to prevent sensory overload.

To evaluate efficiency, we recorded the time for completion of each of the tasks and defined task success as arriving at the correct answer. A mobile phone was used as a time keeping device during the sessions, and we also recorded the participants' answers.

6.5.4 Test Procedure

For conducting the tests, we developed the following protocol [1]:

1. Send the screening questionnaire to the participant.
2. Welcome the participant into our office and have them sign the consent form.
3. Explain the tasks to the user and remind them that it is okay to ask questions.
4. For the VR application, explain the controllers and assist them into the headset and adjust it for comfort.
5. When the user is in the VR environment, ensure that the participant is able to move around the space.

[1] For participants who began with Graphite, the procedure was reversed. Thus, steps eight to eleven were carried out after step two, and steps three to seven were carried out afterwards.

6. Start the timer and the camera once the user is ready. They will be stopped once the task has been completed.
7. At the end of the experience give them the SUS questionnaire to fill out and have them answer the VR follow-up interview questions.
8. For the Graphite application, show the user the list of Poolman servers and ensure the user knows where to find the desired data within the folder tree.
9. Explain the tasks to the participant and remind them that it is okay to ask questions.
10. Start the timer and the camera. They will be stopped once the task has been completed.
11. Administer the SUS questionnaire once the participant has completed the tasks.

6.5.5 Pilot Testing

We carried out the pilot test with one of the members of the Poolman team. As a result of the test, we found that our procedure was sound and that the questions and the SUS questionnaire were suitable evaluation tools. The only change that we made was setting up the screen recordings as we had discovered that we had only recorded part of the screen. In subsequent tests, we adjusted the recording area to correct this issue. Despite the screen recording issues, we included the pilot test in our test results as it was considered to be valid.

6.6 Test Results

Since the aim was to compare participants' performance in Graphite and in our VR application, we did not specify a maximum time completion threshold. Additionally, there is no time limit when it comes to identifying abnormalities within metrics in the normal course of their duities. When the metrics show irregularities, it could indicate a system-wide issue, a user-specific issue or a false alarm. If abnormal metrics indicate a system-wide or user-specific issue, the team would solve it as quickly as possible. Sometimes the issue would be identified immediately, in other cases it could take hours, days or even longer to find the root cause of the problem. Thus, this condition was also reflected in how we managed the timing of our participants. Despite our lenience in regards to time, not all participants successfully completed the tasks in both applications (see Tables 6.1 and 6.2). [2] However, it was clear that most of the participants completed the task faster or in a similar time frame in the VR application in comparison to their performance in Graphite (see Fig 6.7).

6.6.1 SUS Questionnaire Results

SUS scores lie within a range of 0 to 100. This number represents the composite measure of the overall usability of a product. Based on the SUS questionnaire results, the SUS scores for

[2] Participant five had given the correct answer, but it was incorrect in the context of the task as it was a new server in the midst of being integrated into their systems at the time of testing. Thus, it was not considered to be a Poolman server in the context of the experiment resulting in an incorrect response.

Table 6.1: Participant Responses for Identifying Servers with the Highest and Lowest Loads in Graphite

Participant	Highest Load	Lowest Load
1	Correct	Correct
2	Correct	Incorrect
3	Correct	Incorrect
4	Correct	Incorrect
5	Incorrect	Incorrect*
6	Correct	Correct

Table 6.2: Participant Responses for Identifying Servers with the Highest and Lowest Loads in our VR Application

Participant	Highest Load	Lowest Load
1	Correct	Correct
2	Correct	Correct
3	Correct	Incorrect
4	Correct	Correct
5	Correct	Incorrect
6	Correct	Incorrect

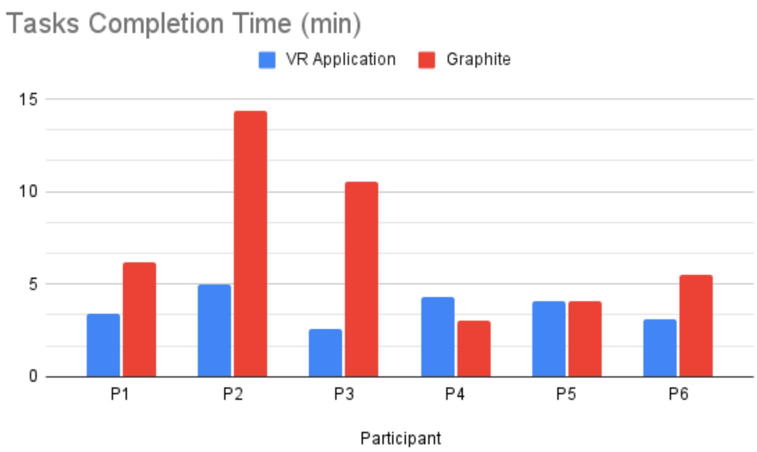


Figure 6.7: Tasks completion time for Graphite and VR application
.

the VR application ranged between 62.5 and 80 (see Fig. 6.8) with a mean of approximately 73.3 (see Fig. 6.10). While the SUS score for Graphite, was between 40 and 65 (see Fig. 6.9) with a mean of approximately 51.7 (see Fig. 6.10). Thus, the SUS score for the VR application was higher than the SUS score for Graphite.

According to the ranges displayed in Fig. 6.11, we can say that our VR application was "Acceptable" with an adjective rating of "Good." On the school grade scale, it earned a "C". While Graphite would be considered "Low" in terms of acceptability and earned a grade of

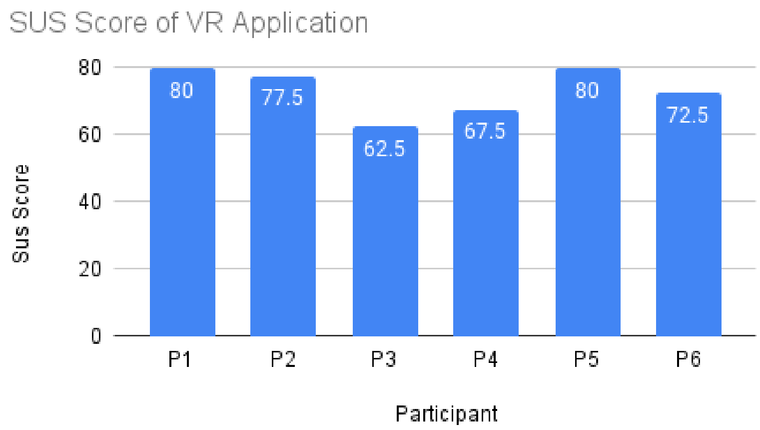


Figure 6.8: SUS scores for our VR application.

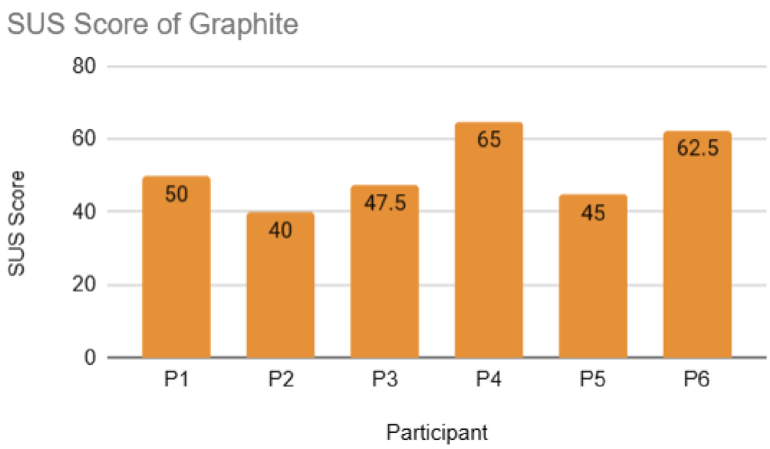


Figure 6.9: SUS scores for Graphite.

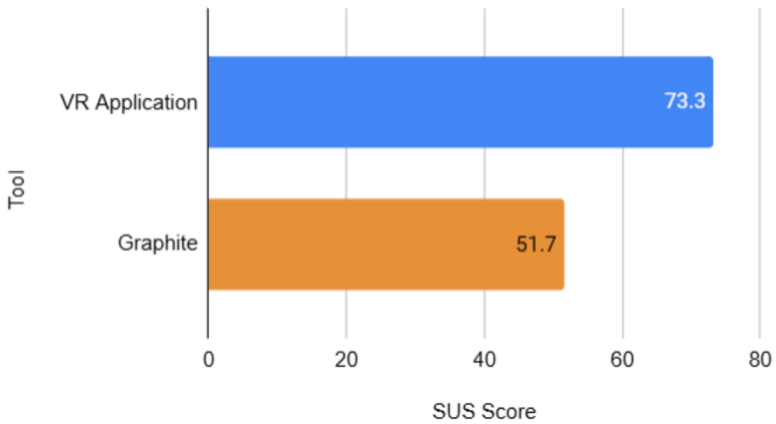


Figure 6.10: The average SUS scores for our VR application and Graphite.

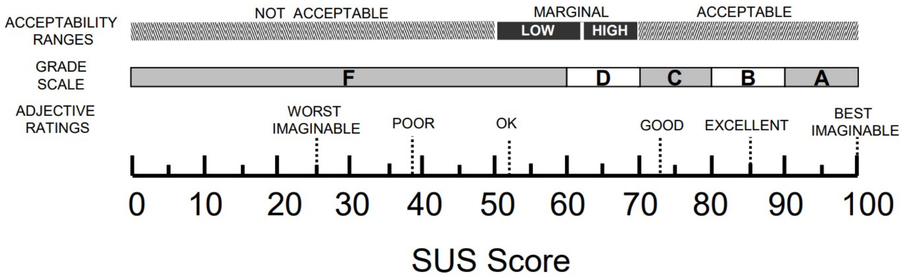


Figure 6.11: A comparison of the adjective ratings, acceptability ranges, and school grading scales, in relation to the average SUS score. Source: [45]

"F" with an adjective rating of "OK".

6.6.2 Interview Responses

Perceived Effectiveness: From the interviews, a majority of participants (five out of six) felt that the VR application allowed them to effectively accomplish the tasks. One of the participants had a mixed view on the application in this context: It was easy to determine the server with the highest CPU load, but it was difficult to determine which server had the lowest CPU load as the nodes representing the servers were of a similar size. A suggestion from this user to improve the application would be to create an outer shell for the nodes that would have provided an additional visual cue to indicate which node was larger and which was smaller.

For two of the participants, while stating the application did support them in completing the tasks, there was some confusion as to the meaning of the numbers. In both cases, it was thought that the numbers represented the CPU load and not the amount of CPU available for each server. Another user who also agreed that the VR application helped them to accomplish their goals also indicated that the application lacked capabilities such as sorting and the lack of a saving capability that would reduce or prevent the user from needing to navigate the space to find the data. Also, a different participant mentioned that it did take them some time to orient themselves in the space as they had not been in this VR environment before. Nevertheless, they still felt that they could accomplish the tasks given.

Distraction: Four out of six of our participants did not feel they were distracted from their task while in the VR environment. Participant 2 noted that there was some mild discomfort from the headset itself due to the pressure on their glasses, which did bother them to some degree. Otherwise, for this participant, the experience itself did not possess any distracting elements. Another participant felt that the way in which the experience was organized felt quite natural to them as it was aligned with how they envisioned the relationships between the sites and servers. However, this participant also noted that it was difficult to see the orange text when it was superimposed on the sky blue background.

For the two participants who were distracted (participants one and four), the aspects of the experience that they were distracted by were different. Participant one saw the name of the subteam in the environment and thought that it was a feature and did not realize that it was meant to indicate that the environment itself was related to the Poolman subteam. Additionally, the spaceship aesthetic of the environment led the participant to think that perhaps there was a place to sit where the name was located as it looked like there could be such a spot. Though the participant was quick to note that if there were other rooms with the names of other subteams that they would not have been confused. The hands were also mildly distracting as they were pleasantly surprised to see them in the environment. Despite these elements, they felt they were able to find their way and knew what to focus on in order to accomplish the tasks.

Participant four found themselves focusing on the wrong aspect of the tree when they were accomplishing the task. Thus, it did take them a while to notice that there were numbers attached to the nodes that displayed the percentage of the CPU that was idle. Instead of looking at the nodes, they looked at the branches instead.

Sensory Overload: For this question, none of the participants felt overwhelmed or experienced sensory overload while in the VR environment. One of the participants said that the environment was nicely designed. They went on to mention that the different colors made it easy to discern the different aspects of the experience.

Preferred Platform: For four out of six of our participants, there was a clear preference for the VR application in the context of the tasks. One of the participants had a mixed view of the VR system. They noted that it was nausea-inducing, which did not make for a comfortable experience. However, they did feel that they were more confident in their responses for the VR application because the numbers were there. The second participant that did not prefer the VR application stated that they had more control and certainty in Graphite. For them, if they had to look for a single node, it would be easy to use, but when looking at 20 or 30 nodes, they felt that it could be difficult to figure out their size at a distance and to feel confident in the answer. However, in Graphite, it is possible to sort the data by size or time and they are more certain that they have the correct answer.

For those who did prefer the VR application over Graphite, it was noted by one participant that the VR application would be comparable to the dashboards that they have in Grafana. The relevant sources are already selected and presented to the user in the VR space unlike in Graphite where the user needs to know in which folder the data resides in order to find the correct time series. Also, a Graphite dashboard could become hard to read as more data is added to it, which makes it difficult to discern the individual lines. The other users in this group also held similar opinions in that it was easier to find the data that they were looking for in the VR application than in Graphite. It was noted by other participants that Grafana, Kibana or Datadog would be a much better tool to compare with our application, if those applications contained the data.

Other comments: At the end of the testing session, some suggestions were made regarding further development of the VR application. It was recommended by some participants that we should assign color gradients that change in relation to the value of the metric that was depicted. For example, a gradient from green to red could be assigned where a node would be green if it was 100% idle and would progressively become more red as the percentage approached 0%. In general, they would have liked to have seen the data values represented not only with the node size, but color as well as it would be more attention grabbing. An increase in the size of the nodes was also suggested. Another user also had difficulty discerning size differences between the nodes in the VR experience and just relied on the values underneath the nodes for the task.

Three of our users also mentioned there was some difficulty in understanding what the nodes represented. It was suggested by one of the users that there should be a VR-based tutorial that onboards the user and explains everything in the environment. Using the hands to explain what the nodes signified would also have been helpful and would reduce the onboarding burden, according to this participant. Another participant mentioned that the prescence of menus similar to those in the second lo-fi concept (see Fig. 5.2) would also have been useful in this context.

6.6.3 Observations

While contemporaneous notes were taken that indicated notable reactions or comments from our participants, the screen recordings and video recordings provided some additional context or were used as a support in case something was missed that was deemed relevant after the testing was completed. Unfortunately due to operator error, there were no screen recordings for the VR application for participants five and six. Also, there was no video recording of participant three when they were in the VR environment. Despite some of the missing data, most of the notable observations were recorded during the sessions. Thus, we did not lose a lot of information due to the missing recordings.

Five out of six of our participants were quite talkative in the VR environment. From those five, three engaged in quite a bit of self-talk while one was more conversational and spoke to us as they were going through the test. The other individual (participant three) did not have a video recording, thus it was not possible to determine if their talkativeness was directed towards us or if they engaged in self-talk. Four of those participants were somewhat talkative during the Graphite segment of the test, but were notably less so. However, one of the participants, participant four maintained the same level of talkativeness in both Graphite and VR. Participant five was not very talkative compared to the rest of the participants. For both conditions, they talked about the same amount, which was quite infrequently.

For most of our participants (four out of six), there was a clear unwillingness to leave the experience despite being reminded by us that they should move on to the next task. (**Note:** The experience did not have any sound.) However, from the screen recordings, it was clear that the participants were having fun in the environment. They were exploring the rest of the room and interacting with the trees and other objects. One participant was ready to leave due to motion sickness, while the other participant was willing to leave when asked as they realized there was nothing left to do in the environment.

In terms of distracting elements, it was clear from the recordings that participants one and four were distracted by the environment. Both participants in the course of the task either explored other parts of the room that were not relevant to the task (the Poolman sign) or were interested in interacting with the nodes and tried to touch them. It should also be noted that participant four did not notice the values under the nodes until later in the experience. The other four participants were able to remain focused when in VR and if they wished to further explore the environment did so once the task was over.

Even though it was a seated VR experience, there was some difference in movement between the participants. Most slightly turned their chair or moved their heads. However, there were two participants who were much more animated (participants one and six). They were spinning around in the chair quite a bit, while the others mainly moved using the controllers. It was interesting that one of our active participants noted that after the spinning and turning with the controllers that they had lost their sense of location in the physical room.

In terms of the differences between node sizes, despite the participants being in the same conditions, some did not see a clear difference in node size while others did notice a clear difference in node size even for the task to find the server with the highest load. For that task, it involved identifying an outlier that was much smaller than the other nodes. However, participants six, four and two noted that it was difficult to see any differences. While the other participants did not mention any difficulties at all except for the task where they had to find the node with the least CPU load. In that condition, most of the nodes were indeed of similar size and difficulties would be expected.

In comparison to the Graphite task, the VR task required quite a bit of explanation. All participants faced minimal difficulty using the controls and were able to use them effectively throughout the task. However, the task and the metric displayed often had to be re-explained for the participants. Additionally, the tree structure and its meaning was also a point of confusion for some users initially, but they were quickly able to understand once another explanation was provided.

6.7 Summary

Overall, the participants were quite satisfied with the VR applicaiton in comparison to Graphite for this particular task. The way in which the data was presented was generally well appreciated. However, some participants did experience some mild discomfort while using the application.

In terms of efficiency, while the participants were able to complete the task faster in the VR application, they had marginally better accuracy than with Graphite. Thus, it cannot be said in this task that the participants were noticeably more efficient. Also, when comparing their backgrounds in terms of familiarity with the Poolman metrics, the subteam in which they worked, the length of time they worked within their subteam and their weekly usage of Graphite, there were no clear differences in performance, which could be due to the small sample size.

When evaluating our VR application in terms of learnability, it was relatively fast to onboard individuals, and they were quick to grasp the controls. However, it was clear that there was some confusion as to what was represented in the environment. This issue could have been resolved with a better user interface or a menu that would help orient the user. Additionally, the differences in node size was difficult for many of our users to discern irregardless of height. Thus, improvements are clearly needed for this application.

Chapter 7
Discussion

In terms of whether or not XR can be used to visualize real-time Big Data sets, the question remains open. Due to many setbacks in the development of our hi-fi prototype, we were unable to adequately answer that question as we were only able to represent a subset of the data that we had intended to display in our prototype. However, it has become clear that the limitations for displaying real-time data does not only depend on the immersive application, but also on the limitations of the existing IT infrastructure. Perhaps if our application were able to be directly connected to the servers and services that generate the logs and data points that are stored in their databases, this could be realized.

As to the effectiveness of visualizing data in an XR environment, it was clear in the testing of our hi-fi prototype that it offered a marginal improvement over the existing Graphite software application in terms of time savings, but the responses generally had the same level of accuracy. Despite the minimal improvement in efficiency, we proved that it is possible to construct an application that does not lead to sensory overload. However, given that the scope of the task in the user test was extremely limited, it remains to be seen if the tool would retain this characteristic when employed in a more complex task.

7.1 Effectiveness of the Design Process

The design framework and guidelines we applied in the course of our thesis helped to guide the development of our application relatively well. Both the general design guidelines for analytics applications as well as the 5 Question Design Framework for IA provided us with a foundation for our design. Thus, we focused more on features that could enhance user experience. Additionally, the Double Diamond approach and the co-design methodology dovetailed nicely with each other. We were able to combine our knowledge and skills with those of the engineers within the R&D Tools department to identify their needs in terms of data visualization. This hybrid approach also lead to the identification of other problems such as visual overload and clutter as well as signal noise from notification tools.

Following our process has led to the generation of interesting design ideas that we had not considered such as using a small electric shock to notify users of abnormal metrics instead of a scent. Also, over time, this process had also allowed us to discover the reason for the skepticism some had concerning the utility of XR applications in the scope of their workflow. Not only were ideas generated, but our misunderstandings were corrected through the process. For example, we considered applications and containers to be separate entities, when in the scope of this department they were essentially the same in many cases.

However, there are some aspects of the process that could have been improved. The focus group interviews was one of the most time consuming stages in our design process. We interviewed nearly everyone and followed up with some individuals, which gave us great insight into their work processes and how they use data. In hindsight, it may have been a bit too ambitious to speak with everyone and should have limited the number of people whom we interviewed given our timeline. Despite the large time investment, we were more than satisfied with the amount of information we managed to glean from this part of the process.

7.2 Prototyping in VR

For our project, we decided to build both our lo-fi and hi-fi prototypes as VR experiences. The lo-fi prototype was built using Shapes XR, which is a prototyping platform for VR experiences. While it lacked the capability of incorporating user interaction at the time, it provided a decent platform for spatial design. Thus, the engineers were able to experience the space. Unfortunately, it was a bit difficult to work with the application in the beginning due to the fact that it required a high amount of bandwidth to start and download the application in a reasonable time. However, we were able to solve this issue by switching to a more powerful wi-fi network.

Beyond the issue with network connectivity, it was quite easy to collaborate within the application itself. When building the prototype we used the Oculus Quest, Meta Quest 2 and Meta Quest Pro. The software was compatible with all three headsets, and it was possible to use different headsets simultaneously. However, the Oculus Quest did struggle with the program as it kept freezing unlike the Meta Quest 2 and Meta Quest Pro, but overall it was still usable. Within ShapesXR, there were many assets available from simple geometric shapes to more complex models as well as assets that could be used to construct a UI. Therefore, it was possible for us to quickly construct a prototype.

Since VR also encompasses a wide array of modalities, it is possible to design applications that can utilize not only the visual modalities but others such as sound, smell, taste and touch. However, on this platform the ability to use other modalities besides vision was very limited. Thus, in order to include other modalities, it would require the use of other technologies such as headphones or scent diffusers that would be difficult to incorporate in the prototype itself.

Despite the limitations of ShapesXR, prototyping a VR application while being in the virtual space itself was quite useful. It was much easier to see spatial relationships within the VR environment as we were building the prototype and changing the layout as necessary. This prototyping tool also led to time savings as we did not have to constantly switch between desktop and VR HMDs.

However, we did encounter ergonomic issues when working in the VR environment for long periods of time. The weight of the Oculus Quest, Meta Quest 2 and Meta Quest Pro headsets did cause some physical discomfort. Additionally, the screens within the headset also could cause eye strain with long-term use (more than four hours per day). Thus, while prototyping in VR has benefits, the experience could be further improved with more ergonomic headsets.

During testing, our VR lo-fi prototype was essential in eliciting feedback specific to virtual environments. If we had used other prototyping methods that did not involve an all-encompassing virtual environment, we would not have received such comments. We were able to garner feedback relating to physical discomfort and claustrophobia that we would not be able to easily replicate without an extensive real-world setup.

On the other hand, the fact that our prototype was in VR also became a distraction for the engineers as it elevated their expectations. Despite our explanations that it was a lo-fi prototype that was meant to gather their initial thoughts on our proposed application, the participants expected a more realistic prototype with real data. Using real data would detract from the purpose of a lo-fi prototype as that would require extensive backend work. However, it was clear that we needed a better way to manage expectations with some of our participants.

7.3 Legal and Technical Challenges

Compared to mobile and desktop application development, VR development is not as straightforward. During this thesis, we encountered legal hurdles when it came to our software development platform and the headset itself. The corporate license for Unity was very complex, and the legal department would have needed a substantial amount of time to analyze the agreement. Given the time frame of the thesis, the process would have taken too long to seek approval for a Unity license, hence our choice of Unreal for our development environment for the hi-fi prototype.

The headset itself also had separate terms for corporate use that also needed to be approved by the legal department. Normally, when buying a PC for developing any application, there are no terms of service that dictate what you can do with the programs you have created on the device. For game engines such as Unity and Unreal, there are often terms that apply to what you create with these programs [46], [47]. Given that these programs are widely used for VR development, it is possible that such terms could be one of the reasons that VR technologies have not been widely adapted.

While developing our hi-fi prototype, we had faced some technical issues related to running Unreal, Steam and the Valve Index on Ubuntu. Some of these issues related to how the OS managed HMDs and poor documentation from Unreal and Steam in regards to how to run their programs and devices on Linux. With some googling and visiting several developer boards we were able to get our devices running, but there is a clear need for more support for Linux on VR. Also, it is important to have strong rapport with the IT department when developing in VR as many of the fixes may not be possible on corporate computers due to how devices are configured as well as security needs.

7.4 Limitations

Unlike our original plan we did not engage in rapid iterations of the hi-fi prototype due to the aforementioned technical and legal issues. Thus, we only focused on one subteam for our hi-fi prototype. Consequently, the potential advantages and disadvantages of our IA application could be greater or lesser than what we have seen in the course of this project.

Also, the R&D Tools department is composed primarily of software engineers who use data visualization in the course of their work. While we did chose to focus on one of the many user groups for analytics applications, perhaps a different perspective could be offered by those whose roles are primarily concerned with data such as machine learning engineers, data analysts and data scientists. Their viewpoints could have guided the development of the application in a different direction than we had decided to pursue in this thesis.

Additionally, the hi-fi prototype was not completed as we had intended for this thesis due to technical and legal hurdles. Thus, it was not a complete test of this application. Also, many of the features that our users liked in the lo-fi prototype were not implemented, which could have negatively affected user performance in the testing phase of the hi-fi prototype. If it was completely implemented, then we would have a better basis on which to judge the suitability of an IA application in this context.

Chapter 8
Conclusion

Our application mainly focused on how engineers could potentially navigate their data sets. However, we found that for some use cases, an IA application would not be the best choice given the limitations of the current hardware. Sometimes the best choice was a PC with a monitor, keyboard and mouse if the user wished to look at text data. Within the narrow context in which we tested our prototype, the VR application was faster for most participants. However, the accuracy of the participants' answers was quite similar when comparing our VR application with Graphite. Thus, it remains to be seen if IA applications can truly perform better than existing software solutions.

While our application may not have offered a clear benefit other than faster navigation, it was far from complete. Due to various setbacks that we had faced in the course of the thesis, what was tested was far from the vision we had developed during the course of this thesis project. Thus, this is still an area of research that could yield some interesting findings if a more complete prototype was tested with a larger user group. Therefore, our paper should not be considered as a final statement on the feasibility of an IA tool, but as a stepping stone for further research.

References

[1] Ishika and N. Mittal, "Big data analysis for data visualization a review," in *2021 9th International Conference on Reliabillity, Infocom Technologies and Optimization (Trends and Future Directions)*, ser. ICRITO. Noida, India: IEEE, Sep 2021, pp. 1–6.

[2] E. Olshannikova, A. Ometov, Y. Koucheryavy, and T. Olsson, "Visualizing big data with augmented and virtual reality: challenges and research agenda," *Journal of Big Data*, vol. 2, no. 1, pp. 1–27, Dec 2015.

[3] R. Liu, M. Gao, L. Wang, X. Wang, Y. Xiang, A. Zhang, J. Xia, Y. Chen, and S. Chen, "Interactive extended reality techniques in information visualization," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 6, pp. 1338–1351, Dec 2022.

[4] M. Wallergård, J. Eriksson, and G. Alce, *Virtual Reality, An Introduction*. Lund, Sweden: Studentlitteratur AB, 2022.

[5] Axis Communications, "Axis homepage," https://www.axis.com/about-axis, (accessed: Dec 11, 2022).

[6] United Nations, "The 17 goals," https://sdgs.un.org/goals, (accessed: Mar 16, 2023).

[7] E. Olshannikova, A. Ometov, Y. Koucheryavy, and T. Olsson, "Visualizing big data with augmented and virtual reality: challenges and research agenda," *J. Big Data*, vol. 2, no. 1, Dec. 2015.

[8] D. Thompson, J. A. Levine, J. C. Bennett, P.-T. Bremer, A. Gyulassy, V. Pascucci, and P. P. Pebay, "Analysis of large-scale scalar data using hixels," in *2011 IEEE Symposium on Large Data Analysis and Visualization*. IEEE, Oct. 2011.

[9] P. Reipschläger, T. Flemisch, and R. Dachselt, "Personal augmented reality for information visualization on large interactive displays," *IEEE Transcations on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1182–1192, 2021.

[10] T. Dwyer, K. Marriott, T. Isenberg, K. Klein, N. Riche, F. Schriber, W. Sturezlinger, and B. H. Thomas, “Immersive analytics: An introduction,” in *Immersive Analytics*. Cham, Switzerland: Springer Nature Switzerland AG, pp. 1–23, 2018.

[11] W. Willet, Y. Jansen, and P. Dragicevic, “Embedded data representations,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 461–470, 2017.

[12] “Analog innetermometer,” https://www.billigteknik.se/hem-hushall/32622-analog-innetermometer-7316040001017.html?channable=027077696400706373762d333236323236&msclkid=44296a0efc001a34df3b0e5a36178625&utm_source=bingutm_medium=cpcutm_campaign=Shopping%20everything%20else&utm_term=4579534654452388&utm_content=Ad%20group%20%231, (accessed: 2023-05-05).

[13] WeirdGeek, “Building simple linear regression model using python’s sci-kit library,” https://www.weirdgeek.com/2018/11/linear-regression-model-using-python/, (accessed: 2023-05-05).

[14] B. H. Thomas, G. F. Welch, P. Dragicevic, N. Elmqvist, P. Irani, Y. Jansen, D. Schmalstieg, A. Tabard, N. A. M. ElSayed, R. T. Smith, and W. Willett, “Situated analytics,” in *Immersive Analytics*. Cham, Switzerland: Springer Nature Switzerland AG, pp. 185–220, 2018.

[15] L. White, https://pixabay.com/sv/photos/k%c3%b6tt-steka-n%c3%b6tk%c3%b6tt-mat-laga-mat-4616902/, (accessed: 2023-05-08).

[16] https://map.worldweatheronline.com/temperature?lat=55.52699887874952&lng=96.07474782658285, (accessed: 2023-05-08).

[17] M. Wolf, “Make pie in the metaverse? designer creates ar kitchen assistant with snap spectacles,” https://thespoon.tech/make-pie-in-the-metaverse-designer-creates-ar-kitchen-assistant-with-snap-spectacles/, (accessed: 2023-05-04).

[18] K. Marriott, J. Chen, M. Hlawatsch, T. Itoh, M. A. Nacenta, G. Reina, and W. Stuerzlinger, “Immersive analytics: Time to reconsider the value of 3d for information visualization,” in *Immersive Analytics*. Cham, Switzerland: Springer Nature Switzerland AG, pp. 25–55, 2018.

[19] M. Whitlock, S. Smart, and D. A. Szafir, “Graphical perception for immersive analytics,” in *Proc. of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, ser. IEEE VR. Atlanta, GA: IEEE, 2020, pp. 616–625.

[20] C. Ware and P. Mitchell, “Visualizing graphs in three dimensions,” *ACM Trans. Appl. Percept.*, vol. 5, no. 1, pp. 1–15, Jan. 2008.

[21] Y. Yang, B. Jenny, T. Dwyer, K. Marriott, H. Chen, and M. Cordeil, “Maps and globes in virtual reality,” in *Proc. of the Eurographics Conference on Visualization 2018*, ser. EuroVis. Brno, Czech Republic: The Eurographics Association and John Wiley & Sons, 2018, pp. 427–438.

[22] J. McCormack, J. C. Roberts, B. Bach, C. D. S. Freitas, T. Itoh, C. Hurter, and K. Marriott, "Multisensory immersive analytics," in *Immersive Analytics*. Cham, Switzerland: Springer Nature Switzerland AG, pp. 57–94, 2018.

[23] L. Shams, Y. Kamitani, and S. Shimojo, "Illusions: What you see is what you hear," *Nature*, vol. 408, p. 788, 2000.

[24] W. Büschel, J. Chen, R. Dachselt, S. Drucker, T. Dwyer, C. Görg, T. Isenburg, A. Kerren, C. North, and W. Stuerzlinger, "Interaction for immersive analytics," in *Immersive Analytics*. Cham, Switzerland: Springer Nature Switzerland AG, pp. 95–138, 2018.

[25] W. Steurzlinger, T. Dwyer, S. Drucker, C. Görg, C. North, and G. Scheuermann, "Immersive human-centered computational analytics," in *Immersive Analytics*. Cham, Switzerland: Springer Nature Switzerland AG, pp. 139–163, 2018.

[26] X. Li, F. Han, X. Sun, Y. Liu, Y. Li, and Y. Chen, "Bracelet: arms-down selection for kinect mid-air gesture," *Behavior & Information Technology*, vol. 38, no. 4, pp. 401–409, 2019.

[27] E. Mahfoud, K. Wegba, Y. Li, H. Han, and A. Lu, "Immersive visualization for abnormal detection in heterogeneous data for on-site decision making," in *Proceedings of the 51st Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences, 2018.

[28] S. Sharma, S. T. Bodempudi, and A. Reehl, "Real-time data analytics of COVID pandemic using virtual reality," in *Lecture Notes in Computer Science*, ser. Lecture notes in computer science. Cham: Springer International Publishing, 2021, pp. 106–116.

[29] Á. M. Fernández, L. A. Bongo, and E. Pedersen, "GeneNet VR: Interactive visualization of large-scale biological networks using a standalone headset," Sep. 2021.

[30] T. Dingler and S. Brewster, "Audiofeeds–a mobile auditory application for monitoring online activities," ser. ACMMM. Firenze, Italy: ACM, 2010, pp. 1067–1070.

[31] N. Barrett, "Interactive spatial sonification of multidimensional data for composition and auditory display," *Computer Music Journal*, vol. 40, no. 2, pp. 47–69, 2016.

[32] D. K. McGookin and S. A. Brewster, "Multivis: Improving access to visualizations for visually impaired people," in *Proc. of the Conference on Human Factors in Computing Systems*, ser. CHI. Québec, Canada: CHI, 2006, pp. 267–270.

[33] M. Crider, S. Bergner, T. N. Smyth, T. Möller, M. K. Tory, A. E. Kirkpatrick, and D. Weiskopf, "A mixing board interface for graphics and visualization," in *Proc. of the Graphics Interface Conference*, ser. Graphics Interface. Montréal, Canada: Candian Information Processing Society, 2007, pp. 87–94.

[34] K. Marriott, J. Chen, M. Hlawatsch, T. I. aand Miguel A. Nacenta, G. Reina, and W. Stuerzlinger, "Just 5 questions: Toward a design framework for immersive analytics," in *Immersive Analytics*. Cham, Switzerland: Springer Nature Switzerland AG, pp. 199–288, 2018.

[35] C. Naranjo-Bock, "Creativity-based research: The process of co-designing with users," https://uxmag.com/articles/creativity-based-research-the-process-of-co-designing-with-users, 2012, (accessed 23-05-12).

[36] M. Walker, L. Takayama, and J. A. Landay, "High-fidelity or low-fidelity, paper or computer? choosing attributes when testing web prototypes," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 46, no. 5, pp. 661–665, Sep. 2002.

[37] G. Bardaro, A. Antonini, and E. Motta, "Robots for elderly care in the home: A landscape analysis and co-design toolkit," *Int. J. Soc. Robot.*, vol. 14, no. 3, pp. 657–681, Apr. 2022.

[38] R. Budiu, "Between–subjects vs. within–subjects study design," https://www.nngroup.com/articles/between-within-subjects/, 2018, (accessed 2023-05-11).

[39] P. C. Price, R. Jhangiani, i Chant A. Chiang, D. C. Leighton, and C. Cuttler, *Reearch Methods in Psychology*, 2017.

[40] J. Rubin and D. Chisnell, *Handbook of usability testing*, 2nd ed. Chichester, England: John Wiley & Sons, Apr. 2008.

[41] J. Brooke, "Sus: A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, 11 1995.

[42] R. F. Dam and T. Y. Siang, "Affinity diagrams: How to cluster your ideas and reveal insights," https://www.interaction-design.org/literature/article/affinity-diagrams-learn-how-to-cluster-and-bundle-ideas-and-facts, 2022, (accessed 2023-05-19).

[43] Epic Games, "Unreal engine 5," https://www.unrealengine.com/en-US/unreal-engine-5, (accessed: June 6, 2023).

[44] Dekogon Studios, "Art | fashion | automotive showcases," https://dekogon.notion.site/Art-Fashion-Automotive-Showcases-f5325a5374664be08de55c40981ad2e3, (accessed: May 19, 2023).

[45] A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: Adding an adjective rating scale," *J. Usability Stud.*, vol. 4, pp. 114–123, 04 2009.

[46] Unity, "Unity editor software terms," https://unity.com/legal/editor-terms-of-service/software, (accessed: March 13, 2023).

[47] Epic Games, "Unreal engine end user license agreement," https://www.unrealengine.com/en-US/eula/unreal, (accessed: June 7, 2023).

Appendices

Appendix A

System Usability Questionnaire

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Response Options:

1. Strongly Disagree
2. Disagree
3. Neutral
4. agree
5. Strongly Agree

Appendix B

Focus Group Interview Questions

1. Do you normally work on-site, remote or hybrid?
2. What does your typical day look like? (focus on how they work with data)
3. What are the services that you work with/use normally?
4. What kind of data do you work with?
5. Which data sources do you use/access daily?
6. Are there any data sources that you access less frequently but are relevant to your work?
7. What kind of software do you use for data visualization?
8. Do you do any sort of data analysis? And if so, which algorithms and transformations do you normally use?
9. How is the data normally presented?
10. Do you customize your visualizations? If so, how?
11. Are you satisfied with the software that you use? What do you like about the software? What do you not like about the software?
12. Do you normally share your visualizations or results of your analysis with others?
13. What would the ideal program look like?
14. During your career at R&D Tools do you often switch teams? And if so, how often?

Appendix C

Revised Interview Questions for User Testing (Lo-fi Prototype)

Scenes 1 and 2:

1. Do you think that it provides a quick and clear overview which is easy to understand?
2. Are there any features missing that you would like to see in this overview?
3. Could you differentiate different teams?
4. Would you find filtering useful?

Scene 3:

1. Do you think viewing the data from this perspective is useful?
2. Did you notice the highlighted path?
3. Would you find blue network traffic visualization useful?

Scene 4:

1. Does the workflow make sense to you ?
2. Does the menu selection help you to find the metrics that you want?

Scene 5:

1. How do you feel about the placement of the metrics?
2. Do you think they will support your workflow?/ Do you think this arrangement would support your workflow?

3. How do you feel about the placement of the menu?

General Questions:

1. How do you feel about the way in which the data is represented?

Appendix D

Screening Questionnaire

1. Are you sensitive to motion sickness?
2. Do you consent to video and audio recording of yourself while performing the user test?
3. Do you wear glasses?
4. What is your Axis email address?
5. How old are you?
6. What is your gender?
7. How tall are you?
8. Have you used any of the data visualization tools such as Graphite, Grafana and Elastic/Kibana before?
9. If yes, which one of these data visualization tools have you used?
10. Have you used VR applications before?
11. If yes, which VR application have you used?
12. Have you used any sort of immersive analytics application in the past?
13. If yes, which immersive analytics application have you used?
14. To which subteam are you currently assigned?
15. How long have you been working with your current subteam?

16. Are you familiar with the metrics commonly used by the Poolman team?
17. On average, how many days per week do you use the following tools at least once?
    - 4-5 days per week
    - 2-3 days per week
    - Once a week
    - Never