Hermann Villanger

Exploring temporal map animations in VR

Master's thesis in Engineering and ICT
Supervisor: Terje Midtbø
June 2020

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Civil and Environmental Engineering

NTNU
Norwegian University of Science and Technology

Hermann Villanger

Exploring temporal map animations in VR

Master's thesis in Engineering and ICT
Supervisor: Terje Midtbø
June 2020

Norwegian University of Science and Technology
Faculty of Engineering
Department of Civil and Environmental Engineering

NTNU
Norwegian University of
Science and Technology

Master thesis

(TBA4925 - Geomatics, Master thesis)

Spring 2020

for

Hermann Villanger

Exploring temporal map animations in VR

BACKGROUND

Newer and improved VR equipment allows for more immersive VR experiences, with fewer of the drawbacks from earlier VR equipment.
This equipment enables the virtual environment to more naturally respond to human movements, without breaking the illusion of a real world.
This opens for studying how these advances in VR technology can be used for interacting with maps and 3D models in a virtual environment.

TASK DESCRIPTION

Explore animated maps in VR and carry out an experiment based on prototypes developed during the master thesis.

Specific tasks:

- Study related literature and give an insight in relevant technology
- Develop prototypes for interaction with maps in VR
- Design and carry out an experiment to test interaction with these prototypes

ADMINISTRATIVE/GUIDANCE

The work on the Master Thesis starts on January 15th, 2020

The thesis report as described above shall be submitted digitally in INSPERA at the latest at June 6th, 2020.

External supervisor:

Supervisors at NTNU and professor in charge:
Terje Midtbø

Trondheim, February, 2020

NTNU
Faculty of Engineering
Department of Civil and Environmental Engineering

Date	Our Reference
14.5.2020	
Your date	Your ref

Extraordinary circumstances due to the Corona pandemic

This Master thesis is based on work that was accomplished in the spring semester 2020. In this period the Corona pandemic was active and influenced the work of several master students. The grading of the thesis must take the pandemic situation into consideration.

If this Master thesis is affected by the Corona pandemic, the student will point out the influenced elements in the beginning of the report. More details about this may also be explained later in the thesis.

Terje Midtbø

Terje Midtbø
Professor

Abstract

Recent improvements in Virtual Reality (VR) technology have led to increased interest in and availability of the technology. This have opened up new possibilities for creating interactive visualizations, and has led to increased research into the usage of VR.

This increased interest has not led to a similar increase into research related to how maps can be used together with VR, despite the potential. This thesis aims at adding to this field of study by using VR as a platform for conducting an experiment around animated maps.

A test environment was created in VR, and an experiment was designed to compare three different temporal legends using data drawn from different types of phenomena. The goal was to investigate whether each temporal legend and form of data interact in some way to produce animations that are easier to interpret.

The Covid-19 pandemic significantly reduced the scale of the experiment, which meant that only a pilot experiment with two participants and a small scale experiment with three participants could be conducted.

The results of the small scale experiment indicates that there might be interaction effects present between temporal legends and the phenomena data sets are drawn from. The results are promising, but the small scale of the experiment makes it difficult to draw definitive conclusions from them. A full scale experiment is therefore needed to properly investigate this topic.

Sammendrag

Nylige forbedringer innen Virtual Reality (VR) har økt interessen for og tilgjengeligheten av teknologien. Dette har gitt nye muligheter for å lage interaktive visualiseringer, og har ført til økt forsking på bruken av VR.

Denne økte interessen har ikke ført til en tilsvarende økning i hvordan kart kan brukes sammen med VR, til tross for potensialet. Hensikten med denne oppgaven er å bidra til dette fagfeltet ved å anvende VR som plattform for å gjennomføre et eksperiment med animerte kart.

Et testmiljø ble laget i VR, og et eksperiment ble designet for å sammenligne tre forskjellige temporale tegnforklaringer ved hjelp av data hentet fra forskjellige typer fenomener. Målet var å undersøke om hver temporale tegnforklaring og datatype interagerer på et eller annet vis for å gjøre animasjoner lettere å tolke.

Covid-19 pandemien førte til en betydelig redusering av skalaen til eksperimentet, hvilket førte til at det kun var mulig å gjennomføre et piloteksperiment med to deltakere og et eksperiment i redusert skala med tre deltagere.

Resultatene fra dette reduserte eksperimentet indikerer at kan være interaksjon mellom temporale tegnforklaringer og fenomenene datasett er hentet fra. Resultatene er lovende, men den reduserte skalaen av eksperimentet gjør det vanskelig å trekke endelige konklusjoner. Et fullskala eksperiment er derfor nødvendig for å undersøke dette emnet grundigere.

Preface

This Master's Thesis is submitted to the Norwegian University of Science and Technology (NTNU), and completes my Master's Degree in Geomatics at the Department of Engineering and ICT (MTING). The thesis is written under the supervision of Terje Midtbø. The work has been carried out during the spring semester 2020, and has offcial subject title TBA4925 - Geomatics, Master's Thesis. I am grateful for the help and encouragement Terje Midtbø has provided during the last year.

Hermann Villanger
June 25nd, 2020

Contents

List of Figures

List of Tables

Abbreviations

VR	=	Virtual Reality
HMD	=	Head-Mounted Display
UI	=	User Interface
DoF	=	Degree of Freedom
ANOVA	=	Analysis of Variance

1 Introduction

In recent years, with the gradual maturing of Virtual Reality technology, there has been an increased interest in Virtual Reality, with Virtual Reality devices rapidly becoming more available to the general public.

Virtual Reality used to be a niche technology, only used in a few areas of industry, or by selected groups of enthusiasts. Today, with the current VR revolution, it's rapidly becoming mainstream, and is finding new uses.

For a long time VR has been used on the subject of visualizing complex 3D models, which VR is uniquely suited for, and new visualization opportunities are being explored.

Cartography is an area in which VR has great potential. Maps might not be the first thing that comes to mind when one thinks about VR, but the new technology is opening up new opportunities for visualizing and interacting with maps, which might improve some uses of maps today. Current digital maps are usually viewed through 2D computer screens, which has issues with occlusion of data, and does not allow for true 3D visualizations. With VR, true 3D experiences can be created, which brings with it several challenges and opportunities. Maps show the spatial position of data, and visualize both place and value. With VR, and its ability to show true 3D environments, an additional dimension can be added to the traditional 2D map. This gives us more tools when visualizing complex 3D data, which could allow us to create better visualizations of complex data.

Mapping in VR has great potential uses, but there has so far been limited exploration of these applications. One goal of this thesis is to explore the opportunities which are opened by VR. Among the plethora of maps used today, one in particular sticks out as interesting to look at in the context of VR. This is the animated map, whose popularity has greatly increased with the rise of the internet. These maps show both a spatial position and a temporal position. When the temporal aspect of the data is important for the understanding of the animated maps, these kinds of maps may also be called temporal maps. In these cases, it is therefore important that the temporal aspect of these maps are properly conveyed to and understood by map readers. This can be done through the use of temporal legends, which represents the time-stamp of the data. Several types of temporal legends are used in animated maps, and there is ongoing research into the effect each legends have when viewing an animated map.

This thesis attempts to provide a supplement to this research, and explore it in a different environment. It was decided to focus on temporal maps, with an emphasis on temporal legends, and to use the platform VR to explore this topic. Further, it was chosen to conduct an experiment in which a group of users is asked to compare the various temporal legends in a VR environment. The game engine Unity was used for this purpose. Several VR prototypes were developed and tests were conducted using the VR platform Oculus Quest. There main question that this thesis aims to cast light over is the following:

- Are visualizations that use certain temporal legends more easily understood than others?
- How does different types of data affect the legibility of each temporal legend?

This thesis is organized as follows: In chapter 2, relevant theory of VR and cartography is presented. This includes topics such as virtual environments, user interface design in VR and temporal maps. In chapter 3, the development of the test environment will be discussed. Due to Covid-19 pandemic and in order to reduce the risk of transmitting the disease, a full experiment could not be conducted in this thesis. It was instead decided to perform a small scale experiment to obtain some preliminary data.

In chapter 4, the details regarding the framework and conduction of this pilot experiment will be explained. Finally, chapter 5 will show and discuss the results of the present work.

2 Background

The main purpose of this work is to investigate how users experience the interaction between the geographic information that is displayed and the legend(here a time indicator) in a Virtual Realty environment. This chapter will present some background information necessary to understand the technology used and choices made during this thesis. This will focus on Virtual Reality technology and the design of user interfaces in VR, and information regarding temporal maps.

2.1 Virtual Reality

Virtual Reality (VR) refers to technologies that can create simulated environments, and allows a person to interact with and become immersed in those environments (Mazuryk and Gervautz, 1999 [1]).

VR has a long history, and both large government agencies like NASA, and private entities like Oculus have played a part in the development of VR technologies. Despite VR's long history, it is just recently that it became available for general use. This event was marked by the release of the Oculus Rift in 2016, which led to renewed interest in the use and development of VR technology. Today, several large technology companies are developing VR technology, and its availability for the general public is increasing. This section will explain various important aspects within VR, some fields of applications, and how user interfaces may be adapted to fit VR.

2.1.1 Virtual Environments

Virtual Environment is the term that is often used for the digital surroundings created and shown in VR. It consists of all the virtual elements a user can interact with and move around. Virtual environments are usually developed by using 3D game engines like Unity, Unreal Engine and CryEngine, as 3D environments can be easily translated to VR. Most VR platforms, such as Oculus, supports translation of the developed 3D environments to virtual environments that can be used on the respective platforms.

In order to view and interact with the virtual environments, Head-Mounted Displays (HMD) are commonly used. These are display devices worn on the head that give the user a stereoscopic image of the virtual environment and gives the appearance of being inside the virtual environment. An image of the HMD device Oculus Quest, which is the HMD used in this thesis, can be seen in figure 1.

Figure 1: The Head-Mounted Display platform Oculus Quest [2].

In order to achieve immersion, HMDs use various tracking sensors to keep track of the user position, and change their position within the virtual environment to reflect this. Earlier HMDs, such as the Oculus CV1, needed externally placed sensors to accurately track the position of the users. This would be cumbersome to set up, and reduced the mobility of the device. The latest VR devices uses internal tracking, where the tracking sensors are built into the device itself. This makes the device easier to set up and use, and can help to improve the users perception of VR.

The system which keeps track of the users movements and calculates their position is called Degree of Freedom (DoF). HMDs with fewer or more primitive sensors, such as earlier models, were only able to track three of these parameters, namely rotation around the X-, Y- and Z-axis, and hence called 3 DoF. Newer HMDs are able to track movement in the X-, Y- and Z-axis, in addition to rotation. Consequently they are called 6 DoF, as all six parameters are tracked, which allows for natural movement through the Virtual Environments [3]. Figure 2 shows the rotations and movements that are tracked with 3 DoF and 6 DoF.

Currently, the most common method for interacting with virtual environment is through hand held controllers. They differ from traditional gaming controllers in a few ways, with the most important distinction being that the controllers also need to be tracked, in order to show the users hands correctly in VR, and that the user is able to see and use them intuitively.

2.1.2 Applications of VR

Throughout its history, VR has been used in a multitude of areas, and as the technology is developed further, its applications are expanding. Historically VR have only been used within larger companies or government agencies, and it is only recently that VR has become a consumer product.

Today, VR is used within various industries. These include manufacturing (Nee and Ong, 2013 [4]), where it's used to visualize prototypes, aid in product design and allow for intuitive interaction with virtual models. It's easier and cheaper to make and evaluate mock-ups when they do not need to be physically made. VR simulators have been adopted for use within medicine and the military, where they offer a way to train

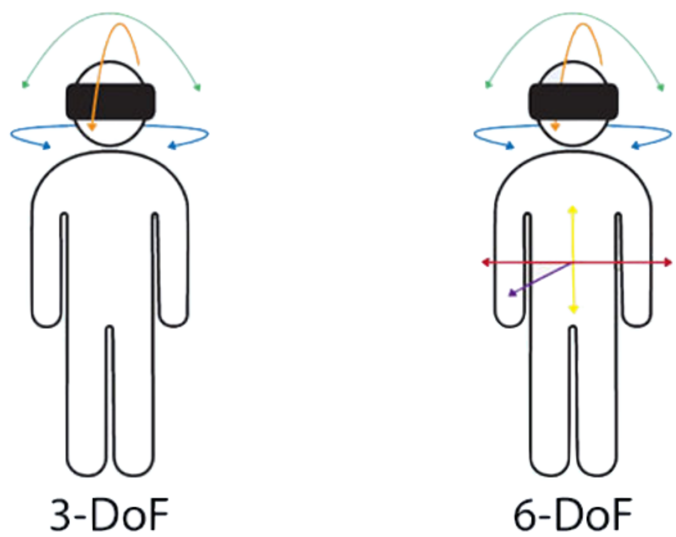


Figure 2: Comparison of 3 Degree of Freedom (3-DoF) to 6 Degree of Freedom (6-DoF) movement (Dom Barnard 2019 [3]).

personnel in realistic scenarios at a much lower cost (Mazuryk and Gervautz, 1999 [1]). An agency that has a long history with both the use and development of VR technology is NASA. NASA has been working with VR simulators since 1991, and use VR for both immersive training and for remote controlling robots. They also have a long history of using VR simulators to train astronauts before sending them on real flights [5].

An area in which VR is being increasingly used is entertainment, particularly gaming. Better HMDs have led to an explosion in software development for VR, and several prominent VR games, such as Half Life Alyx and Beat Saber, have seen wide success on the consumer market. Some VR games, such as Skyrim VR, are ports of already existing games, while others, such as Half Life Alyx, are games designed around the use of VR.

Since it is only recently that VR has become widely available, there have been few applications within cartography, though some exists. These includes a project by Henriksen and Midtbø (2015 [6]), who used VR to study human wayfinding with maps by using a virtual environment. And a project by Herman, Kvarda and Stachoň (2018 [7]) who studied how users of several low-cost VR devices could interpret maps in VR, and how well suited the devices were for this task.

These articles show that VR provides new opportunities for conducting research within cartography. It makes it easier to design experiments to focus on the aspects that are wanted, while removing unwanted aspects from the virtual environment. Henriksen and Midtbø could remove factors such as the sun and shadows, leaving only the map for navigation.

2.1.3 User interfaces in VR

User Interfaces (UIs) represent the main way in which humans interact with computers. The design and implementation of UI has a long history within 2D screen based devices, but additional challenges appear when developing UI for VR. One of these challenges is how the user interfaces should be presented. There is no screen in VR where the user interface can be projected, so another method has to be found for presenting the UI.

To solve this, we can look at how UIs for First-person video games are designed. First-person video games share several similarities with VR, for instance that the player sees through the eyes of the player character, and are viewing the environment from within the scene itself. To use First-person games as a basis for creating UIs in VR, we first need to know the different types of First-person game UIs, and the ways these can be translated to VR. There are several types of UI elements commonly found in First-person video games, which can principally be divided into four main categories: Diegetic, Meta, Spatial and Non-diegetic interface elements (Fagerholt and Lorentzon, 2009 [8]). The distinctions are based on how connected the elements are to the game's narrative, and whether they exist as elements in the game world itself.

Diegetic

Diegetic UI elements exist both in the game's narrative and the environment of the game (Norrman, 2020 [9]). As they are elements within the narrative, the player character can interact directly with these elements without breaking immersion. Examples include manually checking the magazine of a gun to see the remaining ammunition and using a wristwatch to see remaining time in a timed challenge.

Non-Diegetic

Non-diegetic are elements that do not exist within the game's narrative or environment (Norrman, 2020 [9]), and can therefore only be affected by the player and not the character they are controlling. Examples include menus which the player can use to access their items, or overlays that gives the player information about their character, like health, or remaining ammunition.

Meta

Meta elements are part of the game's narrative, but does not exist within the game world itself(Norrman, 2020 [9]). Instead, they are shown on a 2D viewing plane and convey elements within the narrative that cannot easily be conveyed to the player through other means. Examples include showing blood splatter on the screen to signify that the player character is damaged, or showing their remaining ammunition with a number on the screen.

Spatial

Spatial elements are not part of the game's narrative, but exist within the 3D geometry of the game (Norrman, 2020 [9]). They are placed within the game world to provide information to the player that would already be known by the player character. Examples include trail markers that shows the path to different objectives, or highlighting interactable objects in the world.

UI elements which are placed within the game world, such as Diegetic and Spatial elements, are particularly suited for VR environments, and can often be easily translated from First-person video games to VR. Figure 3 shows the menu on the Oculus Quest. This is an example of a Spatial UI in VR, where the UI elements exists as objects in the world, and the player can move around them and interact with them.

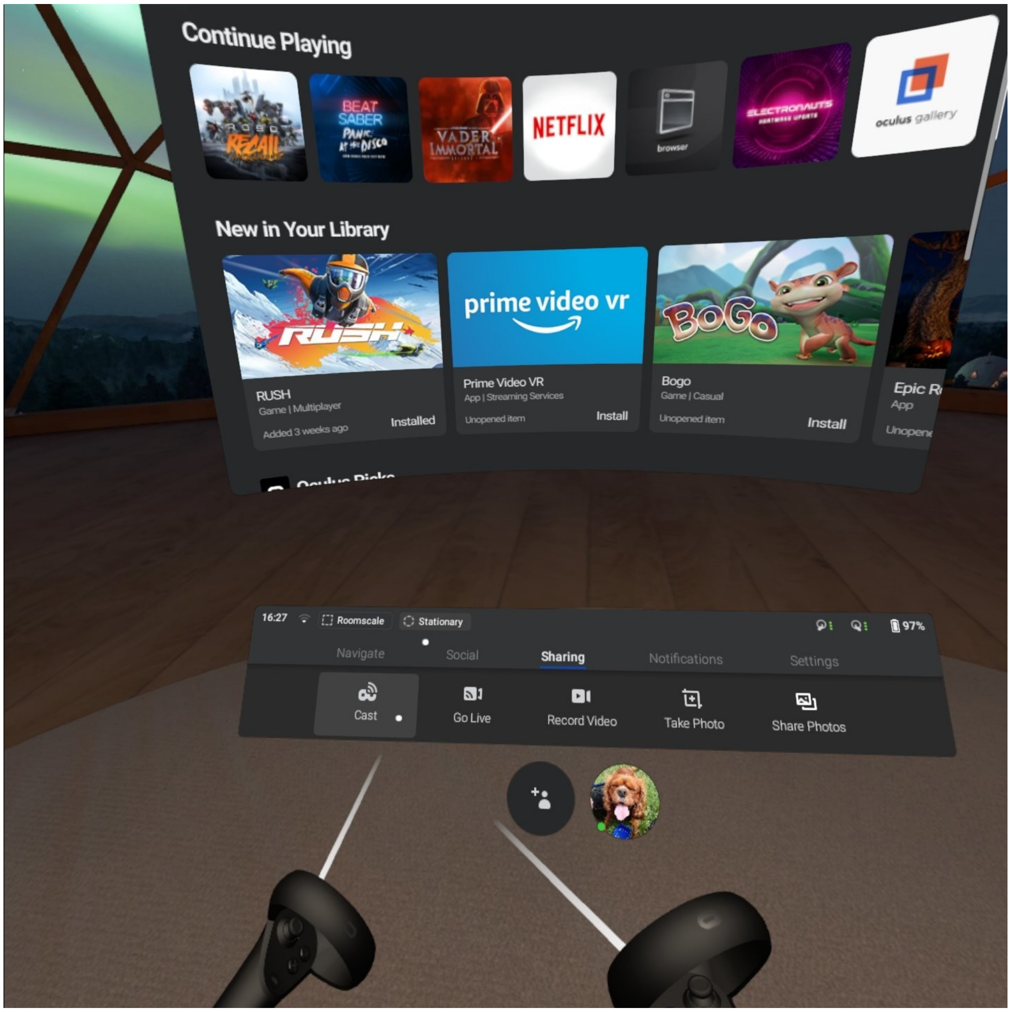


Figure 3: The menu screen in the Oculus Quest. (Sean Endicott, 2019 [10])

Meta and Non-Diegetic elements can be more difficult to translate, since we will need to create some VR analogue to the 2D screen in which they are usually projected. This can be solved by creating an 2D screen-like object within the game which the player can more around and interact with. Figure 4 shows UI from the VR mode of the game No Man's Sky. In this case, the information is projected on an invisible screen some distance away from the player, and it follows and changes according to the players movements. This can be seen as a implementation of Meta or Non-Diegetic UI elements in VR.

Figure 4: User Interface from the Virtual Reality mode of the game No Man's Sky. (Ben Plays VR, 2019 [11]).

As we have seen, there are methods for implementing both spatial and non-spatial based UI elements in VR, and both have been done in VR products that exists today. In addition, there are other aspects of the UI that have to be considered when designing UI for VR experiences. Due to the UI being placed in the virtual world, it is important to consider how far away the UI should be placed from the user. If an UI element is placed too close, it can lead to Vergence-Accommodation mismatch (LaValle, 2019 [12, Chap. 5.4]). This is the phenomena where both eyes try to converge and focus on an object close to the eyes. As the object is virtual, and only appears to be placed close to the eyes, the eyes do not need to focus on it in the same way they do with real world objects. The resulting stimulus provided by VR is inconsistent with experiences from the real world, which can lead to the user feeling strain or fatigue.

Several guidelines for developing UI in VR has been created to avoid problems such as this. The UI should be centered and narrow, in order to minimize the eye and head movement required to use it. It may be more comfortable to interact with the UI if it appears as objects in the virtual environment (LaValle, 2019 [12, Chap. 12.2]). Figure 5 shows Oculus's recommendations for maximum dimensions of critical VR elements. As Oculus recommends to keep one meter between the user and UI elements [13] , these recommendations means that a UI should be less than 1.15m wide and 1.04m tall.

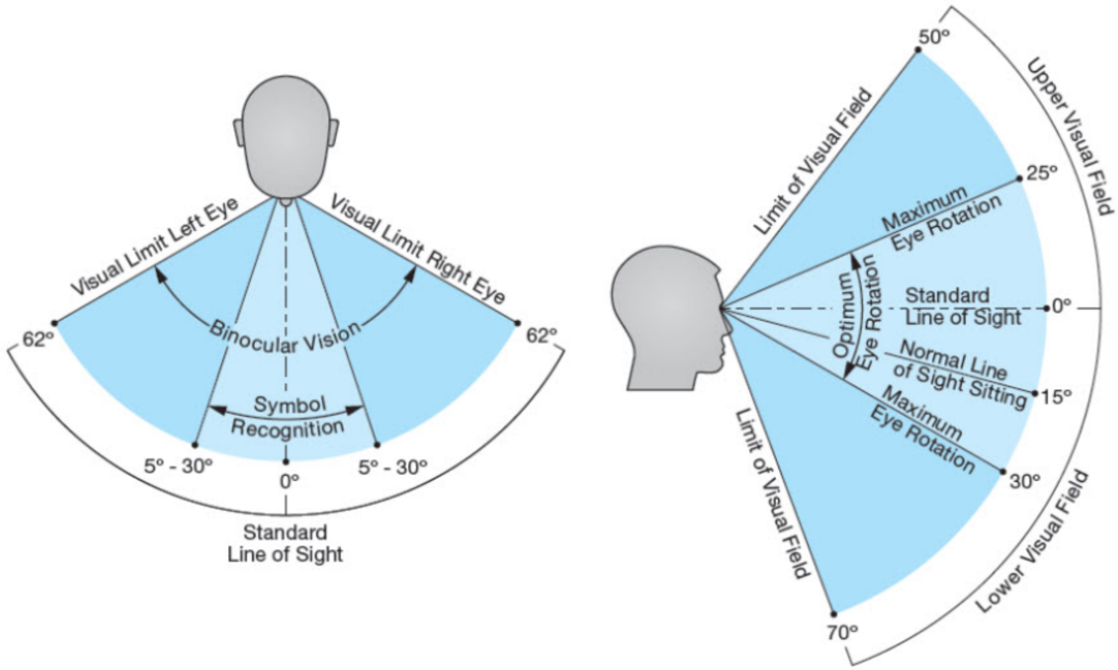


Figure 5: The immediate line of sight of a user in Virtual Reality, which is the area where critical elements should be placed (Oculus [13]).

2.1.4 Oculus Quest

The Oculus Line of VR headsets were the first VR headsets that achieved widespread popularity, and ushered increased interest in VR. Oculus began with a Kickstarter campaign in 2012 to develop the headsets, and the first Oculus Rift headsets were shipped to the public in 2016 [14]. The first headsets required external cameras and sensors in order to track the users movements and achieve 6 DoF tracking. This made the system cumbersome to set up and move. The first Oculus headsets used Xbox controllers for interaction with the virtual environment, but this was quickly replaced with the Oculus Touch controllers [15]. These controllers act as the users hands, and are tracked by the same sensors that tracks the position of the headset.

As VR technology was developed further, there was a need for an upgrade to Oculus VR headsets. This came with the next generation of Oculus headsets, the Oculus Quest and Rift S, which were released in May 2019 [16]. This generation of headsets does not rely on external sensors, and instead use internal cameras and sensors to achieve 6 DoF tracking. As the Oculus Quest is the headset used in this thesis, it will be the focus for the rest of this section.

The main difference between the Quest and the Rift S, is that while the Rift S needs to be powered by a computer, the Quest is an entirely standalone headset. This allows for more freedom of movement for the user, with the drawback of having less processing power than a Rift combined with a powerful PC. As the Quest has a much shorter history than the Rift, it offers a limited selection of released games, though this is expected to change as more games are developed for both headsets.

Other than the Quest being standalone, and the Rift being tethered, there are few major differences. It was chosen to use the Quest for this project as it was thought that the increased freedom of movement would be useful during user experiments, and that the loss of processing power would not be that important for the test environments developed for the experiment.

A new feature of the Quest is the Oculus Link [17], which enables it to operate as an Oculus Rift. This allows the Quest to access games developed for the Rift, potentially simplifying the development and implementation of new applications.

2.1.5 Virtual Reality sickness

Virtual Reality sickness, also called cybersickness, is a phenomenon that can occur during the use of VR devices. It is similar to motion sickness, with symptoms such as nausea, dizziness and headache (LaValle, 2019 [12, Chap. 12.3]), among others. These symptoms can last for hours or even days after using a VR device, though most people will have recovered within a shorter period. With the increased use of VR equipment by the general public, this phenomenon has gotten renewed interest among researchers.

The exact causes of VR sickness is unknown but some theories have been developed, with the most dominant one being the sensory-mismatch theory (LaValle, 2019 [12, Chap. 12.3]). This theory states that VR sickness is caused by mismatch between visual and vestibular senses. Most commonly, the body senses that it is static while the eyes sees in VR that it is moving. This leads to the mismatch between these senses (Gavgani et al, 2018 [18]). As this is usually caused by movement outside the control of the user, reducing this type of movement and letting the user control the movement and rotation of the camera as much as possible is one of the easiest way of reducing VR sickness. Other methods include using better VR hardware that can provide more realistic experiences, or having a higher frame rate. This may necessitate creating simpler virtual environments to reduce the processing strain.

2.2 Temporal maps

There exists two major forms of animated maps: Temporal and non-temporal maps (Łucjan, 2017 [19]). Temporal maps depicts changes in spatial patterns over time (Kraak et al, 1997 [20]), and have become widely popular on the internet. These maps usually presents dynamic events sorted in chronological order together with actual passage of time in the world (Harrower and Fabrikant, 2008 [21]). While temporal maps are a popular method for visualizing animated geographic data, there is ongoing research into the effectiveness of these forms of maps, with no definitive conclusions. Midtbø and Larsen (2005 [22]) found that those who viewed animations performed better than those who viewed static maps under some circumstances. Griffin et al (2006 [23]) found that animated map users were better at identifying clusters, and could answer questions faster than static map users. The use of temporal maps in VR is a focal point of this thesis. This section will talk about important components of these types of maps, with a focus on temporal map legends.

2.2.1 Temporal map legends

Legends in maps are used to explain the visual symbols displayed on a map. These legends are generally static, as symbols or classes do not change over time, only which areas display them. Legends for temporal maps differ, since the legend now represents the passage of time instead of static classes. As a result it will generally be animated, and observers can with little difficulty understand the relationship between data points and moments in time. The animated map covers a specific time span. The role of the temporal legend is to place the current state of the animation in time, and to show where it took place in relation to the greater time span (Harrower and Fabrikant, 2008 [21]). For time spans that can be divided into smaller cycles, like a week being divided into days, it may also show where it took place within these cycles.

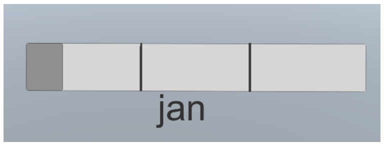
Legends are based on different models for how humans perceive time, and the most commonly used models are time as a linear and cyclic phenomenon (Harrower and Fabrikant, 2008 [21]). In linear time, time is expressed as a chronological order of events, with a single overarching time span. In cyclic time, time is expressed with repeating patterns, usually taken from the rotation of the earth and the movement of celestial bodies (Frank, 1998[24]). Time spans such as a day, month and year are commonly used in cyclic time. Within cyclic time lies the idea that events can be repeated or reoccur in a regular or irregular manner (Harrower and Fabrikant, 2008 [21]). Kraak et al (1997 [20]) thought that different legends may be easier to interpret, depending on the data shown in the animation.

With these two models in mind, we can see them both applied to a weather forecast. The path a raging storm is taking can be seen as linear. The storm is not a repeating event, but has a defined beginning and end. On the other hand, temperature ranges can be seen as a cyclic event. The temperature from day to day will follow similar patterns, with colder temperatures during the night and increasing temperatures during the morning.

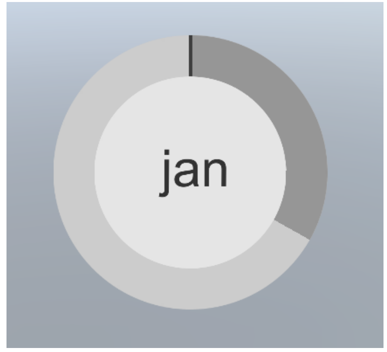
It seems intuitive that some type of temporal legend would be better suited to show a cyclical or linear phenomena, and this assumption has been the subject of several experiments into map understanding, such as described by Edsall et al (1997 [25]) and Midtbø, Clarke and Fabrikant (2007 [26]). In these experiments, map animations were designed in order to make use of linear-, cyclic and text based temporal legends, and the participants had to answer questions regarding the point in time certain events took place. Despite having expected a variance in how good the different animations were interpreted, the authors found no significant differences between the legends.

Midtbø, Clarke and Fabrikant (2007 [26]) did find a significant difference in which legend the participants preferred for various tasks, and the results were in line with what would be expected. For clear cyclic time spans, the cyclic legend was preferred, and for clear non-cyclic time spans, the linear legend was preferred. The much used textual legend was the least preferred legend in all cases, which seem to suggest that people desire some animated representation of time, rather than just a simple text representation.

The three temporal legends most commonly used in map animations, and which were the focus of Edsall et al (1997 [25]), are the Timeline, the Clock and Text representation. The first two are based on the two models of time, linear and cyclic, discussed above. The third, text based representation, does not belong to a specific model of time, but is derived from the way digital time often is expressed on computers and phones today. Of these legends, the Timeline and Text legends are the most common legends found in maps today, with the Text legend being most popular (Cybulski, 2016 [27]). The three temporal legends will be explained in more detail below. See figure 6 for an example of the three temporal legends.



(a) Timeline legend.



(b) Clock legend.

(c) Text legend.

Figure 6: The timeline, clock and text legends created in the test environment used in this thesis.

Timeline

This legend is based on the understanding of time as a linear phenomenon, and usually consists of a bar which gradually fills with color, and some text to show milestones between the start and end. An example of this legend can be seen in figure 6a. With the timeline it is easy to see the temporal position of the animation within the wider time span (Edsall et al, 1997 [25]). On the other hand, having the whole time span represented by the time line can make it more challenging to distinguish the smaller cyclical time periods that may make up the greater time span. Midtbø, Clarke and Fabrikant (2007 [26]) found that this type of legend was the preferred legend in tests that made use of linear time spans. They also found that Timeline legend was preferred to the same extent as Clock for one year time spans, which can be thought of as cyclic.

Clock

This legend is based on the understanding of time as a cyclic/repeating phenomenon. It is represented using a clock shape, which bring up associations of analogue clocks where the hands move around the center (Edsall et al, 1997 [25]). An example of this legend can be seen in figure 6b. One full cycle around the clock translates to one cycle of time. This could represent a day, a month, a year, or other time periods. After this, the clock repeats cycles until the time span is finished. Clocks clearly show position in time within a cycle, but do not show position within the bigger time span (Kraak et al, 1997 [20]). This can make it difficult to understand the broader context for the current temporal position. Midtbø, Clarke and Fabrikant (2007 [26]) found that the Clock was the preferred legend when the time span was 24 hours, which can be thought of as a cyclic time span. They also found that it was equally preferred to Timeline when the time span was one year.

Text

This legend show time as a text string. While the Clock legend has similarities to analogue clocks, the Text legend has similarities to digital clocks, where only the current time is shown (Edsall et al, 1997 [25]). An example of this legend can be seen in figure 6c. The current time is shown precisely to the observer, but the position in relation to the full time span is not shown (Kraak et al, 1997 [20]). This type of legend also requires the user to take their eyes away from the animation to read it, while much of the meaning inherent in the other types of legends can be interpreted with peripheral vision. Midtbø, Clarke and Fabrikant (2007 [26]) found that this type of legend was the least preferred for all time spans, even less preferred than "Circular" for "four decades", despite there being no significant difference in performance between the legends.

2.2.2 Split attention

An unresolved problem with temporal map legends is splitting the users attention between watching both the map and the legend. As both the animated map and the legends change constantly, the user must change their focus between the map and legend and may therefore miss important information (Harrower and Fabrikant, 2008 [21]). The more the users changes their attention, the greater the possibility that they miss out on important information. There are several suggestions for how to design map legends in order to limit this effect.

Kraak et al (1997 [20]) suggest several methods for embedding the legend within the map. These can be using the background as a temporal legend, by dimming it at night and brightening it at day. Other suggestions are using sounds to represent time, but this can be difficult to control. Midtbø, Clarke and Fabrikant (2007 [26]) also came with several suggestions in the form of different animations. These include the use of background shadow and embedding a clock behind the map. Most of the suggestions are focused on placing the legend behind, or around the map in some form.

Translating these recommendations to a virtual environment, where the observer has complete freedom of movement, quickly becomes problematic. It is important that it looks the same if user moves to the other side of the map, which means legend rotates under the map while the map itself stays static. This introduction of movement could become distracting for an observer. A legend placed in a fixed location within the virtual environment can lead to the legend not always being visible to the observer. This would cause the observer to regularly change focus between the map and the legend. Slocum et al (2004 [28]) suggested that a legend spread out along the width of a map might be easier to interpret that a legend occupying a small portion of the display window. However, as there is no traditional display window in VR, the legend would have to be held static next to the map in order for the appearance of its width to be the same. This leads to the problem static legends brings to VR as outlined above.

3 Methods

In order to perform a user experiment based on maps in Virtual Reality, test environments had to be developed. Prototype environments created during the preceding project [29] served as a basis for the test environments created during this thesis. This chapter will describe the development of the test environment, and important choices made during the development.

3.1 Test environment design

As mentioned, the goal of this thesis is exploring the use of maps in VR, and focus was placed on the usage of temporal maps in VR. Several virtual environments were developed in order to test this. These environments consist of a central map with some animated data, and a UI that follows the user is used to control the animation on the map. Figure 7 shows an overview of the virtual environment with these basic elements. These test environments were developed in the game engine Unity [30] and are build on the work done during a preceding project (Villanger, 2019 [29]). The Unity engine is mainly used for creating 3D games, but has a lot of support for VR development. It was therefore to use this platform for developing the test environments. A more thorough overview of both the Unity engine and the development of the initial test environments can be found in the report from the preceding project (Villanger, 2019 [29]).

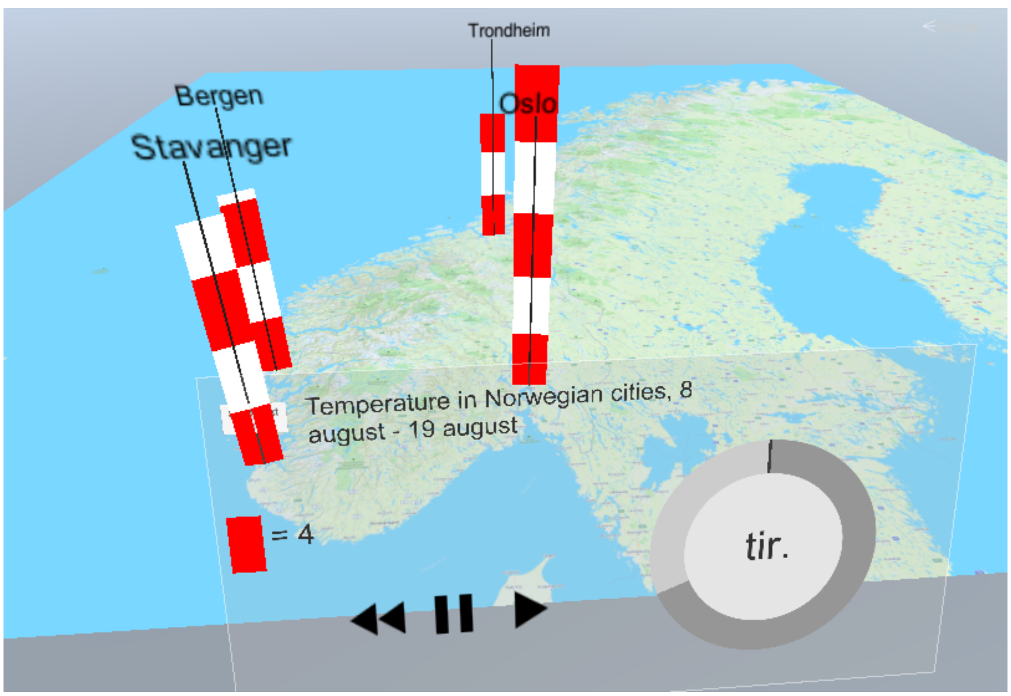


Figure 7: Overview of the virtual environment developed during this thesis. It contains a map with several data point and a user interface which follows the user.

Some measures were taken to reduce the possibility of VR sickness during the development of the environment. These are based on the measures presented in section 2.1.5. The VR environment as a whole is relatively simple, with few moving parts that could make the users disoriented. There are no movements or rotations that are not initiated by the users. This helps keep the stimulus the user gets from VR consistent with their own senses, which is important for preventing VR sickness.

Two important aspects of the test environments are showing the temporal legends to the user, and presenting the geographic data in an easily understandable way. The three temporal legends described in section 2.2.1 were implemented and placed within the UI. Figure 6 shows these implementations in the UI. The "Timeline" and "Clock" legends also had black lines to help distinguishing between adjacent months or days. This section will discuss the overall design of the UI, and the choices made when implementing a UI for VR. After this, the visualization of geographic data in VR will be discussed.

3.1.1 User Interface design

Traditional digital maps and VR are very different mediums, which means methods for interacting with digital maps cannot be directly translated to VR. A map placed in a VR environment will not always be seen from the same side, as the position of the user can change. This makes it difficult to find a location to place the map legend, and rules out static legends placed in the environment itself. The legend can't be embedded into the map either, as the legend must be readable from several sides and should not take too much attention away from the map Temporal legends in particular needs to be visible to the user at all times, in order to show the temporal aspect of the data.

Section 2.1.3 describes some basic forms of UI elements, and shows some examples of these implemented in VR games. These should be considered and compared when choosing how to implement a UI with temporal legends in VR. Problems arise when we look at Spatial UI elements. These are UI elements that are placed in fixed positions in 3D space, such as those shown in the Oculus Menu screen in figure 3. If the temporal legends were implemented with this design, it would be possible for the user to move to a position where they cannot see both the legends and the map animations at the same time. This would not fit the circumstances of this experiment, since a main focus of this experiment is interpreting temporal legends together with map animations.

Due to the problems with Spatial UI elements, it was chosen to implement a UI based on Non-Diegetic UI elements instead, as seen in the game No Man's Sky in figure 4. This type of UI follows the user's movements, and reacts to their rotation, and would therefore always be in the user's field of view. This means that they can view both the map and the temporal legends at the same time, which is crucial for the experiment. A downside to using these UI elements, is that it might get distracting and can lead to split attention, as described in section 2.2.2. If the user were to constantly look at the UI, they might miss important details of the animation. Still, with a UI that requires minimal interaction, only showing the passage of time, this ensures that the user is always able to identify the relationship between the data and the current time.

When designing the UI, the recommendations for VR development that were presented in section 2.1.3 were followed. The UI was placed one meter in front of the user, and is one meter wide and 40cm tall. This is much shorter than the maximum recommended height, as the tests did not require a taller UI. This gives a distance to the participants which is not too short, and makes sure that the participants can interact with the whole UI without moving their head. Figure 8 shows the implemented UI with a linear temporal legend. It contains the temporal legend, buttons for controlling the animation, a button for progressing through the tests and some information about the animated map. This includes a title to give some context for the data, and a legend which shows how much of a given unit each of the vertical blocks of the animation corresponds to.

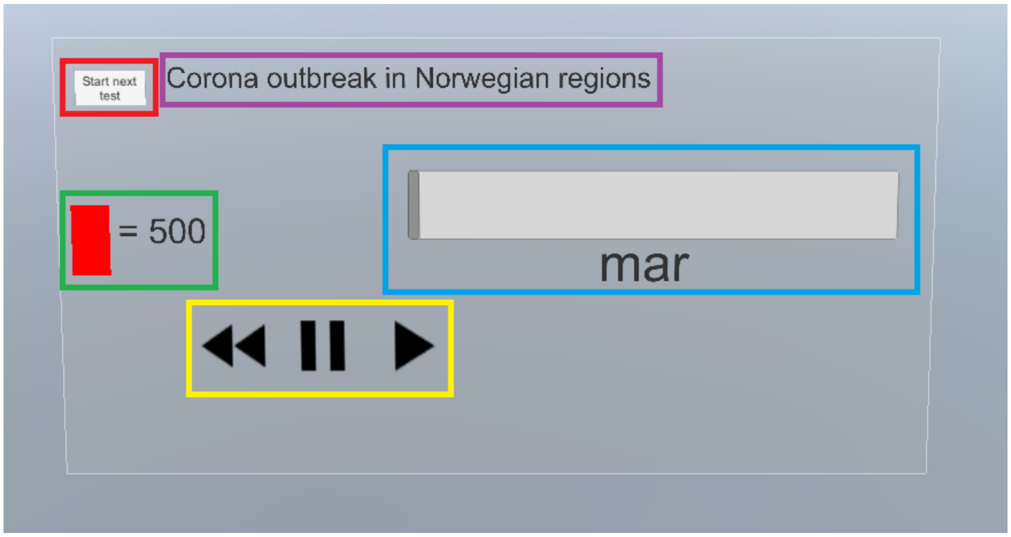


Figure 8: The user interface created for the test environment. It contains temporal legend(blue), time control buttons(yellow), spatial legend(green), title(purple) and button for starting new tests(red).

3.1.2 Geographic data

An important part of the experiment was to show some sort of geographic data on the map. The frameworks used during the development contained some basic functions for placing objects on the map, but most of the implementation had to be done from scratch. When creating this solution, differences between 2D and VR had to be taken into account. Since the map is placed within a 3D environment, the data points will have positions with varying distances from the user. This leads to one point placed further away from the user appearing smaller than a point placed closer, even though both have the same height. When using data points where their height indicated their value, this problem must be taken into account. During the development of the test environment, two methods for displaying different types of data on the map were implemented, namely column and line representation.

Column Representation

The first method displays data as a vertical column, where the height of the column represents the value of the data point. This technique is used for showing changes in data over time for a given location. This procedure can for instance show Covid-19 cases for a city over time or hourly temperature for a region.

Two methods for directly comparing the height of different columns were evaluated. One promising method was using a plane parallel to the map, which cuts through each of the columns at a given height. This makes it easy to quickly see which columns has reached a specific height. This works great when only focusing on one height, but it quickly gets more complicated when more parallel planes are added. Several planes makes it difficult to see which plane corresponds to which value, as they are stacked on top of each other. There were also several issues when attempting to implement this solution within the test environment, as it proved difficult to see the intersection between plane and column. This method for comparing the height of the columns was therefore not used.

The other method was dividing the columns into clearly marked vertical sections, where each section corresponds to some value. Figure 9 shown an implementation of Column Representation using this method.

This scheme makes it much easier to compare different columns, regardless of different distances to the observer. The observer simply need to count how many sections each column consists of. It is also easier to get an overview of which columns are highest and lowest, as this is evident from the relative number of sections. A drawback to this method is that it is difficult to see when a column reaches a specific value, unless this value corresponds to the transition point between sections. This method proved to be much easier to implement than the plane method, and it was chosen to use the section method for column representation of data.

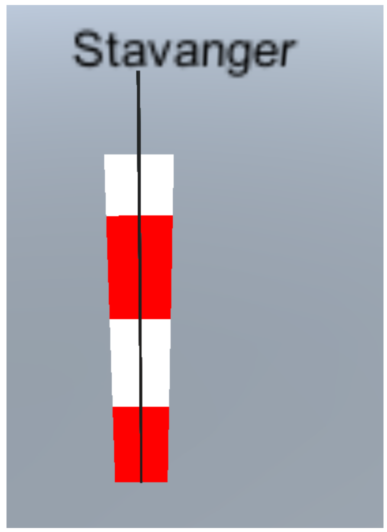


Figure 9: Column used to display point data that changes value over time.

Line representation

The Line Representation method displays data as a series of line segments, where the coordinates of the endpoint represents the position of a data point and the line segments represents the path taken. This method is intended to show the path taken by an event which changes location over time. This method can for instance show the path of a train across a country or a ship sailing along a coast. Figure 10 shows an example of an event which changes location over time. This method was investigated in order to find more alternatives when looking for data to use in the experiment, but it was ultimately not used in the current experiment. Nevertheless, it allows for other types of data to be explored or opens for different tests to be created in the test environment.

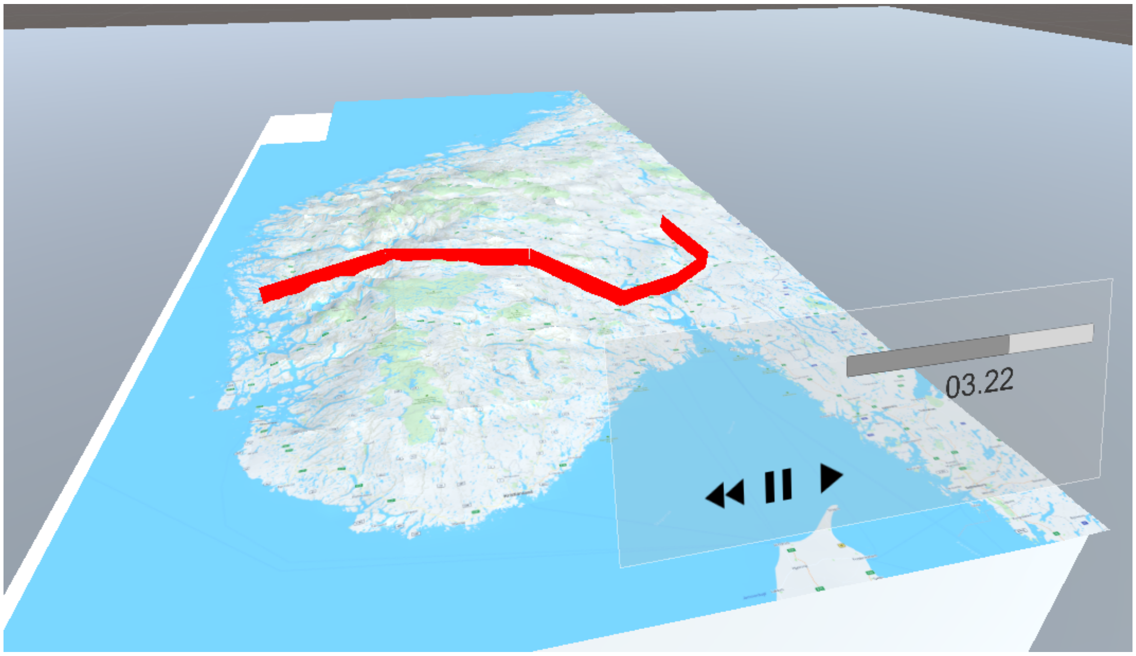


Figure 10: Line used to display position data that changes coordinates over time.

4 Experimental setup

The experiment is set up as an observer trial. Volunteers will use the VR environment to view a visualization of a data set evolving in time, together with a visible time indicator. They are asked to pay specific attention to some details, which they will be queried about when the visualization is finished. This chapter will explain the statistical framework of the experiment, and go through the conduction of a small scale pilot experiment.

4.1 Framework

There are several approaches to take when carrying out an experiment. Previous experiments into temporal legends, such as Edsall et al (1997 [25]) and Midtbø, Clarke and Fabrikant (2007 [26]) compared various temporal legends in order to find significant differences. In this experiment it was instead chosen to focus on how the different temporal legends interacted with various forms of data sources. This would allow us to find out if some of these data sources leads to better results from the users when used together with some temporal legends.

It seems reasonable that a Clock legend, which is a cyclic temporal legend, would be better for showing temperature changes over a day. This assumption forms the basis for the research questions this thesis aims to cast light over, which are:

- Are visualizations that use certain temporal legends more easily understood than others?
- How does different types of data affect the legibility of each temporal legend?

This section will talk about the framework of the experiment, which factors are involved, and how it should be analysed.

4.1.1 Factorial experiment

Factorial experiments are experiments which study the effects and interactions of several factors. "Factors" are a general term used to denote any feature of the experiment that varies between trials. The actual values the factors hold in each trial are called "Levels" (Walpole et al, 2016 [31, Chap. 14]). When referring to these types of experiments, one writes the number of levels each factor holds. For example, an experiment with two factors where each have two levels, is called a 2x2 factorial experiment.

"Full factorial experiments" are conducted by performing trials with all possible combinations of factors and levels. This allows us to efficiently analyze the interaction of several factors and levels within the same experiment, without needing to perform additional experiments with other factors.

A drawback with factorial experiments is that since each combination of factors needs a separate trial, the experiments can easily get too complex. Each additional factor increases the complexity exponentially, which quickly makes it not feasible to carry out the experiment.

An alternative to full factorial experiments is "fractional factorial experiments" (Walpole et al, 2016 [31, Chap. 15]). In these types of experiments only a fraction of the full experiment plan is carried out, such as $\frac{1}{2}$ or $\frac{1}{4}$ of the experiment. This makes the experiment easier and cheaper to carry out, but does not give us the opportunity evaluate higher order interactions between individual factors.

In this thesis is was chosen to conduct a full factorial experiment, but with only a handful of factors to reduce the complexity of the experiment. The statistical software Minitab 19 [32] was used to set up and analyse the experiment. The randomized order of tests for each participants was generated by Minitab.

4.1.2 Factors

Table 1 summarizes the three main factors of the experiment, with associated levels. Since the experimental design consists of one factor with three levels and two factors with two levels, this is a 3x2x2 factorial experiment.

Main factor	Level 1	Level 2	Level 3
Temporal Legend	Timeline	Clock	Text
Data source	Cyclic phenomenon	Linear phenomenon	-
Time scale	Cyclic	Linear	-

Table 1: Table over the main factors and their levels used in the experiment.

Temporal legend

The first main factor is the Temporal Legend. This factor consists of the three temporal legends described in section 2.2.1, which are the Timeline, Clock and Text. This factor allows us to compare each of the temporal legends and study how they interact with the other factors.

The next two main factors are related to the data that is being visualized. As described in section 2.2.1, the different temporal legends are based upon specific models of time. A prevalent theory regarding temporal legends is that these legends are better at visualizing a phenomenon if that phenomenon follows the same model of time. It was therefore chosen to create two factors which describes several aspects of real phenomenona that are related to these models.

Data Source

The first of these is the Data Source, which describes whether the data is drawn from a cyclic or linear phenomenon. In this context, cyclic phenomenona are phenomenona which reoccur at regular intervals and have large similarities between cycles. This can for instance include temperature between days or certain weather patterns.

Linear phenomenona are events which occur in chronological order and do not have cyclic characteristics. Examples include one-off events, such as pandemic outbreaks, or events connected to human development like population growth.

Time Scale

The other factor is the Time Scale, which describes whether the time interval is cyclic or linear. This factor is directly related to the time models the temporal legends are based on, as described in section 2.2.1. Time scales such as a full year, a month or a day can be considered cyclic time scales, while time scales without these cyclic characteristics are considered linear.

These two factors makes a separation of the phenomenon data is drawn from, and the time interval the phenomenon occurs within. This allows us to analyse how each of these factors individually affect temporal legends, while combining them might mask this effect. The experiment was limited to these factors due to the nature of factorial experiments. The complexity of experiments like this rises exponentially with each added factor, making the experiment much more difficult to perform. Choosing a 3x2x2 factorial design leads to a total of 12 tests, which gives the experiment some complexity, while keeping it manageable.

Other factors were considered for the experiment, but it was ultimately chosen to not include them. These additional factors include the placement of the temporal legend within the UI, and how the height value of a data point would be visualized in a VR environment. These factors would both increase the complexity of the experiment and make it less focused. Though the current experiment does not take supplementary factors into account, it may be interesting to evaluate these factors in future explorations of maps in VR.

4.1.3 Tests

12 tests were created for this experiment, one for each combination of the factors and levels discussed in section 4.1.2. The tests were made on the basis of the game engine Unity, and by using the environment described in section 3. These tests were then populated with data sets that could represent the various combinations of the factors "Data Source" and "Time Scale". These data sets needed to cover different types of phenomenona and time scales. Table 2 shows the combinations of the factors "Temporal Legend", "Data Source" and "Time Scale" that make up each test, along with the data set used for the specific tests.

Test	Temporal legend	Data source	Time scale	Data set
1	Timeline	Cyclic phenomenon	Cyclic	Hours of sunlight
2	Timeline	Cyclic phenomenon	Linear	Hourly temperature
3	Timeline	Linear phenomenon	Cyclic	Covid-19, March
4	Timeline	Linear phenomenon	Linear	Covid-19
5	Clock	Cyclic phenomenon	Cyclic	Hours of sunlight
6	Clock	Cyclic phenomenon	Linear	Hourly temperature
7	Clock	Linear phenomenon	Cyclic	Covid-19, March
8	Clock	Linear phenomenon	Linear	Covid-19
9	Text	Cyclic phenomenon	Cyclic	Hours of sunlight
10	Text	Cyclic phenomenon	Linear	Hourly temperature
11	Text	Linear phenomenon	Cyclic	Covid-19, March
12	Text	Linear phenomenon	Linear	Covid-19

Table 2: The 12 tests developed for the experiment, showing which combination of factors they test, along with the data set used in each test.

The data sets used for the experiment needed to have the properties described in section 4.1.2. It was chosen to use data for major cities and regions within Norway. Participants to the experiment would have some familiarity with these areas, and would therefore be an easier task to get oriented with the map. It would also be easier to find data for Norway than for other regions.

Some difficulties were encountered while searching for data sets related to weather phenomena. Many potential data sets had large gaps in them, and few data sets were complete. Some editing was done on the data sets used for the tests, but these only lacked data for short periods, such as a few days each year. Some data sets that were considered briefly but discarded as they had significantly larger gaps of several months or even years.

Another issue was the availability of data sets for different locations. In addition to gaps in the data sets, many locations did not measure specific phenomena at all, or only did so for a brief period of time. This made it difficult to find data for enough locations to use in the tests.

Two weather based data sets without serious drawbacks were found which, along with data over the Covid-19 outbreak in Norway, were used for the tests. These data sets are described below, along with the changes made to them to better fit the tests.

Temperature

The first data set shows the hourly temperature in a selection of Norwegian cities during one week of august 2019. This data was obtained from Norsk Klimaservicesenter [33]. As this data set shows hourly temperature it represents a cyclic phenomenon, as hourly temperature follows a similar pattern for each day. Seven days of data was used since this can be seen as a linear time scale. The original data set was very choppy and had few easily discernible extreme values, which made it difficult to design tests around it. The data set was first smoothed out to make it easier to distinguish days from each other. Then some of the extreme values were made more pronounced, so it would be easier for the participants to distinguish them during the tests.

Sunlight

The second data set shows the weekly hours of sunlight for a selection of Norwegian cities during one year. This data was obtained from Norsk Klimaservicesenter [33]. As this data set shows hours of sunlight it can be seen as a cyclic phenomenon, as hours of sunlight follow a similar pattern from year to year. The time scale is also cyclic, as the whole one-year cycle was shown. and hours of sunlight follow similar patterns between years. The data set originally showed hours of sunlight for each day of the year, which was is far more data points than any other data set. The values for each day varied considerably, which made the data set choppy and difficult to follow. To reduce the number of values, the hours of sunlight for each day of the week was averaged to produce one value for each week. The resulting data set was still choppy but was smoothed out, and the extreme values were made more pronounced to make it easier for the participants to complete the tests.

Covid-19 outbreak

The third data set shows the cumulative number of confirmed Covid-19 cases in Norwegian regions during the 2020 Covid-19 outbreak. This data was obtained from Folkehelseinstituttet [34]. As this data set depicts the outbreak of a disease, it represents a linear phenomenon. Due to the size of this data set, it was chosen to use it for both the linear and cyclic time scales. For tests requiring linear time scales, the whole data set, ranging from 26 February to 22 April, was used. For tests requiring cyclic time scales, the data set was shortened to only contain data for the month of March, which is a cyclic time scale. No region in the data set was the subject of more than one question, to prevent learning effects between tests. Unlike the other data sets, no changes were made to this data set before using it in the tests, as there were no missing values of regions.

In order to ensure that participants can easily identify the correct point to focus on during the tests, these locations are clearly marked on the map with both name and a marker as is shown in figure 7.

4.1.4 Response variable

After the participants have viewed the animation part of a test, they are asked a question about it. The questions are about when a specific event took place, and the participants answer with the point in time they believe this occurred. Due to the fact that the different tests cover various lengths of time and have different sample lengths, we cannot directly compare these results. In order to obtain the response variable of the test, we compare the answer given by the participants to the true answer for the test, and calculate the temporal difference between them. This difference is the response variable of the experiment. When calculating the response variable, we first look at the temporal resolutions of the data set.

- For the Temperature data set, this is one hour.
- For the Sunlight data set, this is one week.
- For the Covid-19 data set, this is one day.

We then count how many of these time units the participants answer was away from the true answer. This means that smaller response values are closer to the true point in time, and larger are further away. Since this response variable takes into account the differences in temporal resolution between each data set, they are comparable.

4.2 Pilot experiment

In order to evaluate and improve the experimental setup, a small pilot experiment was conducted. This was carried out with two participants without backgrounds in cartography or geography. As a start, the participants were introduced to the experiment and informed in detail about how it would be organized. Next they were given a presentation of the Oculus Quest headset, which they would wear during the experiment. In order to be familiarized with the test environments, a special learning environment was shown first. This contained all the various aspects they would interact with, shown in a similar way to the proper tests. After they had become comfortable with moving around in and interacting with the virtual environments, they started the proper tests.

In order to reduce biases, the order of the tests were randomized. As described in section 4.1.3, each test were designed to contain data for a diversity of places within Norway. Before each test, the participant was introduced to the specific test and informed about what the data represented and which time scale it took place over. They were then asked to keep their attention to one specific location, and given some broad information about which aspects of the data they should focus on. They were also informed that they should focus on both the animated map, and the temporal legend, and make a note of the time certain events occurred. When the animation had finished, they were asked a specific question related to the location they had observed, and they should answer with the point in time they believed the event occurred. This was written down, along with any comments or thoughts they had relating to each test. Finally, they were also asked a few questions related to the whole experiment.

Improvements

Based on their comments and answers, some changes were made to improve the experiment. During the pilot experiment, there had been a short introduction before the tests and the experiment as a whole. The introduction was extended to provide the participants with more information about both the experiment and the tests. Such as describing the particular data used for each test, which time period the data is taken from, and better explains what they should be focusing on in the data. There were some feedback given by the participants. In the tests covering a period of one year, the participants found it difficult to properly interpret the black lines showing the transition points between months. Additional text was added to some of these lines to make it easier to understand the legends, which can be seen for the Timeline legend in figure 11.

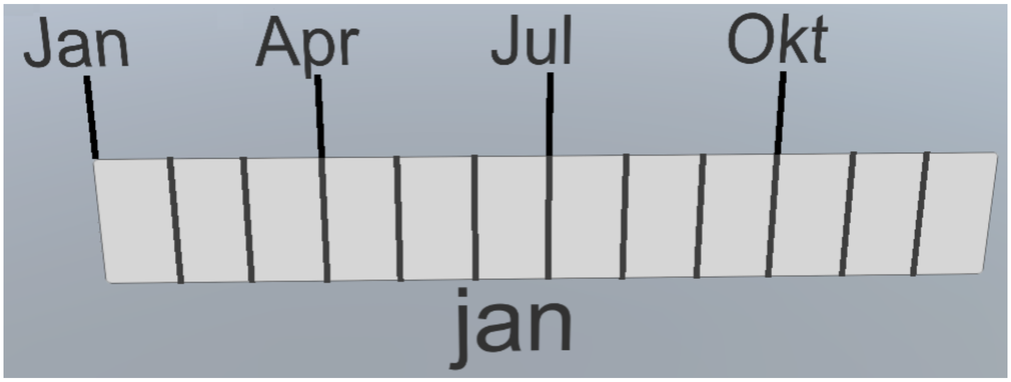


Figure 11: The updated Timeline legend after the pilot experiment.

The participants thought that the data in a few of the tests changed too fast, which made it difficult to focus on. The results also made it apparent that it was much easier to find the correct answer in some tests compared to others. These problems stemmed from the different lengths of the data sets, where each animation was the same length despite some data sets being shorter or longer than the others. To fix the mismatch in difficulty, the animation time was changed to one second for each data value. This made the easier tests shorter and the more difficult tests longer, which should help reduce this problem.

The participants also thought some of the data points were placed too close together. These criticisms were directed towards the placement of "Oslo" and "Viken" in the tests based on the Covid-19 data set. In response to this, the coordinates of "Viken" was placed further east to make it easier to distinguish the two, while still keeping "Viken" within its correct geographic region. The final placements of all regions used for the Covid-19 data set tests can be seen in figure 12.

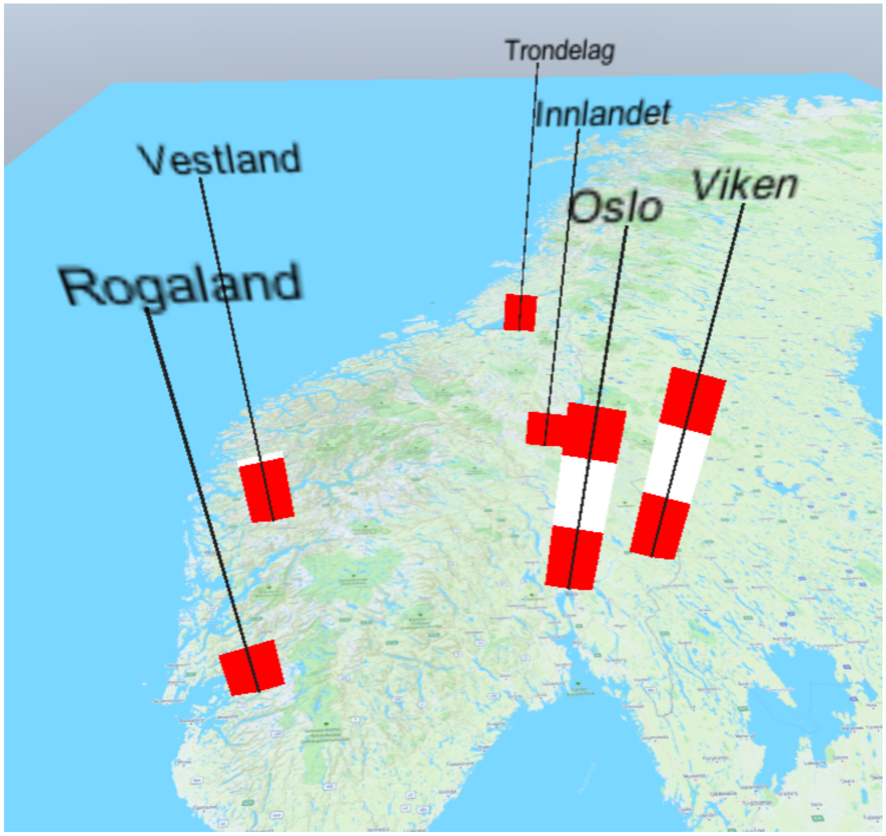


Figure 12: The placement of the regions used in the tests based on the Covid-19 data set.

Finally, some visual issues were encountered when the UI overlapped with the data columns. When this occurred, the column would be rendered in front of the UI, when it should have been rendered several meters behind. This was a bug with the implementation in Unity, and could not be resolved in time. It could be avoided if the participants positioned themselves so that the UI and data column did not overlap. Since this was easy to do, this was added to the information the participants received before the experiment.

4.3 Small scale experiment

With improvements from the pilot experiment included in the experimental setup, a small scale experiment was conducted. It was only possible to get three volunteers to participate in this experiment. These volunteers did not have a background in cartography or VR. The experiment was conducted similarly to that which is described about the pilot experiment in section 4.2, which is the intended way the planned full scale experiment would be conducted.

The order of tests was again randomized for each person by Minitab to prevent bias in regard to test order. The experiment itself was conducted with one participant at a time and with no additional people in the room to avoid distractions. An area of about 2x3 meters was used as the test area which the participants could adjust their position if needed. This could be necessary for some tests so the participants could see the correct point of interest from a better angle for the given test. As large changes in the participants positions were not needed, the area did not need to be expanded for the tests.

The whole experiment was conducted with the participant wearing the HMD. The HMD only showed the animations themselves, and the questions were asked and answered orally. The answers were written down and converted to useful response variables after all the tests were finished. The participants had no major problems with finishing the experiment. The duration of each session was between 20 and 30 minutes, and the participants did not show any signs of VR sickness. This could indicate that VR sickness is less of a concern with the newer HMD models, or that the static nature of the tests aided in preventing the phenomenon.

5 Analysis and Results

This chapter will analyse and discuss the results obtained from the small scale experiment. Due to the low number of participants, these results have a large degree of uncertainty associated with them, and they should mainly be regarded as preliminary findings. This chapter's role is primarily to evaluate the experiment and method of analysis, and provide an example of how the results from a full scale experiment could be analyzed.

5.1 Data Analysis

The statistical software Minitab [32] was used to setup and analyse the experiment, as this software has support for creating and analysing factorial experiments. The procedure is as follows; Minitab first performs a regression analysis to find the best fit for the various factors and interactions to the response variable "Score". The difference between each observation and this regression model is called the "Residual" for that observation. This regression analysis determines the effect of each factor and interaction, and a two-way Analysis of Variance (ANOVA) is performed to determine the significance of each of each factor and interaction. Two-way ANOVA is an analysis method that compares the response variable at different factor levels to determine whether they are significantly different [35]. In addition to assessing the main effects, it also assesses whether there are any interaction effects present.

There are several outputs from performing this analysis:

- Residual Plots from the observations, which are used to decide whether transformation of the observations are necessary.
- ANOVA results, which describes the probability-values(P-values) for all factors and interactions in the experiment.
- Pareto Chart, which shows showing importance and significance of factors.
- Main Effects Plot, which shows the effect of each factor.
- Interaction Plot, which shows the interaction effects between factors.

This section will present and explain these outputs.

Residual Plots

Figure 13 shows the Residual Plots from the experiment. The plots are used to verify various assumptions about the observations [36] from the experiment. If the assumptions relating to a normal distribution, such as constant variance, independence of variables, etc do not hold, the data may need to be transformed to remedy this. The four residual plots are explained below.

The Normal Probability plot shows the residuals plotted against their expected values. This is used to verify whether the residuals are normally distributed. Normally distributed residuals should approximately follow a straight line in the plot. Most of the residuals shown in this plot do seem to follow a straight line, but there are a some that deviate from this. A clear outlier can be seen in the top right of the plot, and to the lower left, there are several residuals that do not follow the line. These are signs that the residuals might not be normally distributed.

The Versus Fits plot shows the residuals against the fitted values. This is used to verify whether the residuals are randomly distributed and have constant variance. If they are, they should be randomly distributed on both sides of zero, and should not follow any clear patterns. The residuals shown in the plot seem to follow a fan-like pattern, with greater fitted values having larger residuals. This can be a sign of nonconstant variance among the observations. There also seems to be an outlier among the residuals, as one fitted value has far greater residuals than the others. This possible outlier can be seen in the top right of the plot.

The Histogram plot shows the distribution of the residuals for all observations. This is used to determine whether the data are skewed or includes outliers. There does seem to be an outlier, as one residual is far larger than any others, and the residuals appear skewed towards negative values.

The Versus Order plot shows the residuals in the order they were collected. This is used to verify whether the residuals are independent from each other. Independent residuals should result in no trends or patterns in the plot. The residuals seem to be randomly distributed, and there seems to be no clear patterns in the observations.

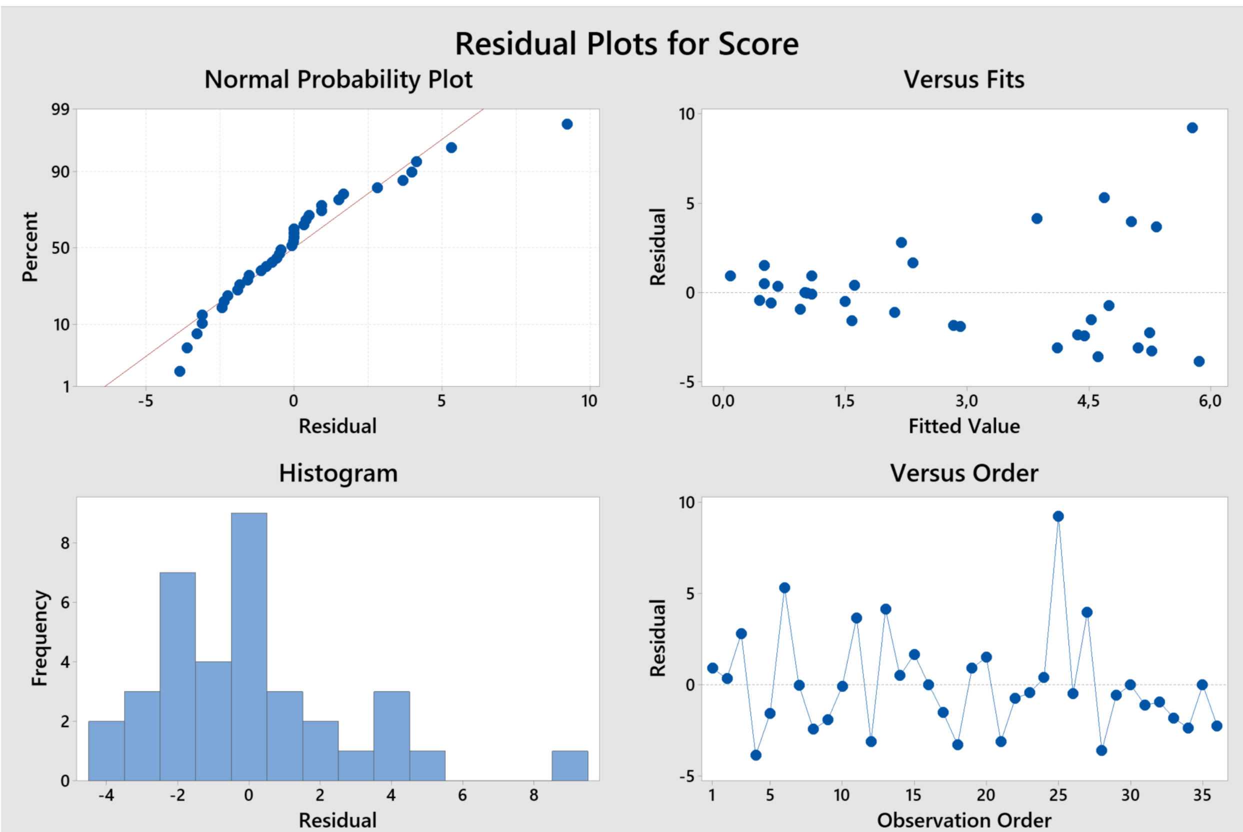


Figure 13: Residual plots of the observations from the experiment.

We observe that the assumptions about the observations may not be entirely correct. This might be caused by the low number of participants in the experiment, and few observations. Normality of observations is a important assumption for this statistical analysis and Minitab recommends transforming the response variable using a Box-Cox transformation. The Box-Cox transformation transforms non-normal dependent variables into a normal shape [37], which should make the observations be more in line with our assumptions. Figure 14 shows the residuals chart after this transformation. We see that the assumptions about normal distribution of observations, constant variance, lack of outliers and non-skewness hold better with the transformed observations. Therefore, the transformed observations are used in the rest of the results.

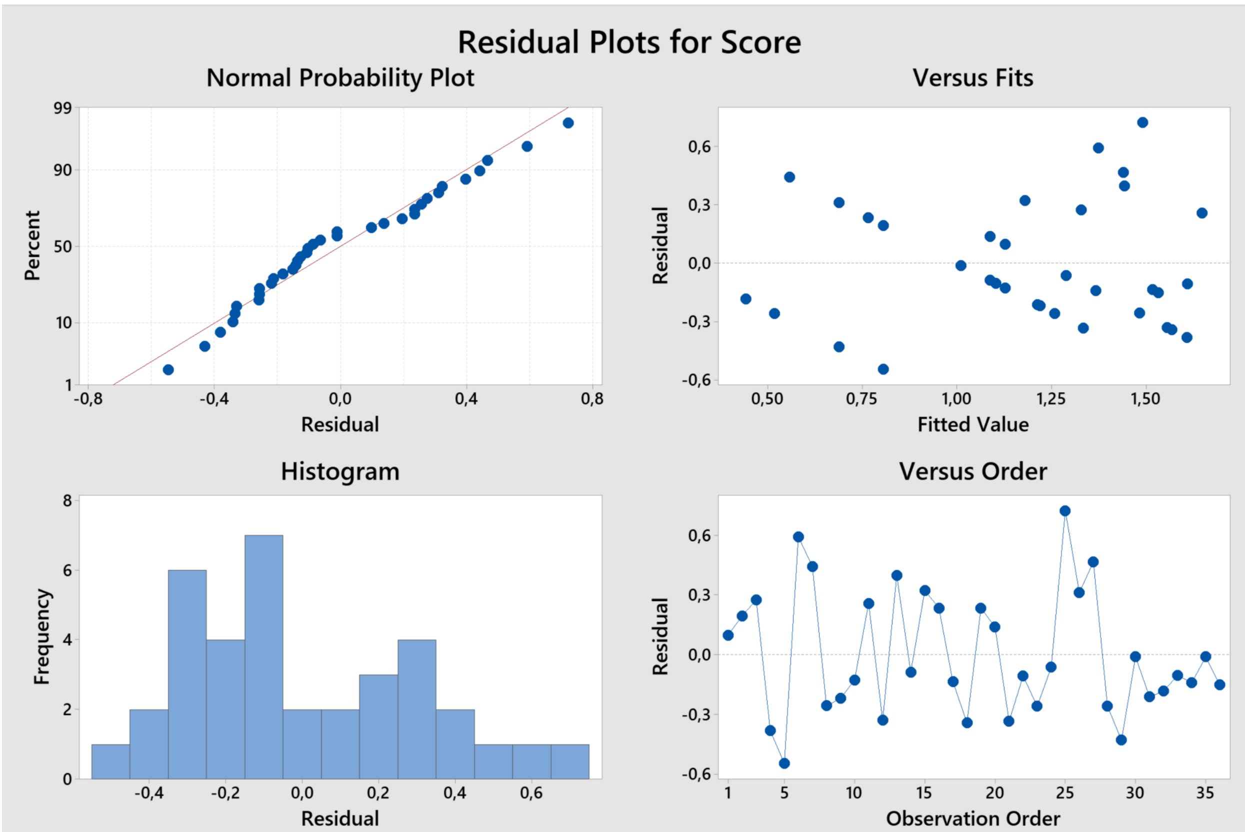


Figure 14: Residual plots of the transformed observations from the experiment.

ANOVA

Table 3 shows the results from the two-way Analysis of Variance (ANOVA). This table shows, among other things, the probability values (P-values) for each of the main factors and interactions [35]. This table can be used to determine which factors have a significant effect on the response variable. The significance level of the experiment is $\alpha = 0.05$. We observe that only Time Scale and the interaction between Temporal Legend and Data Source have p-values greater than 0.05, meaning that only these factors are significant.

Pareto Chart

Figure 15 shows the Pareto Chart from the experiment. This chart shows the standardized effect of each factor, which are t-statistics that test whether the effect is significantly different from zero [38]. The factors are ordered from most to least significant. The red dotted line give the value at which the effects are significant at $\alpha = 0.05$. As can be seen in the figure, the effect of Time Scale (C) and the effect of the interaction between Temporal Legend and Data Source (AB) were found to be significant, whereas all the other factors

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	13	4,137	0,32	2,1	0,063
Blocks	2	0,084	0,04	0,3	0,764
Linear	4	1,960	0,49	3,2	0,033
Temporal legend	2	0,647	0,32	2,1	0,145
Data source	1	0,014	0,01	0,1	0,766
Time scale	1	1,299	1,30	8,5	0,008
2-Way Interactions	5	1,480	0,30	1,9	0,129
Temporal legend*Data source	2	1,129	0,56	3,7	0,042
Temporal legend*Time scale	2	0,006	0,00	0,0	0,981
Data source*Time scale	1	0,345	0,35	2,3	0,148
3-Way Interactions	2	0,614	0,31	2,0	0,159
Temporal legend*Data source*Time scale	2	0,61	0,31	2,0	0,159
Error	22	3,369	0,15		
Total	35	7,506			

Table 3: Results from the two-way Analysis of Variance from the experiment.

and interactions are not significant.

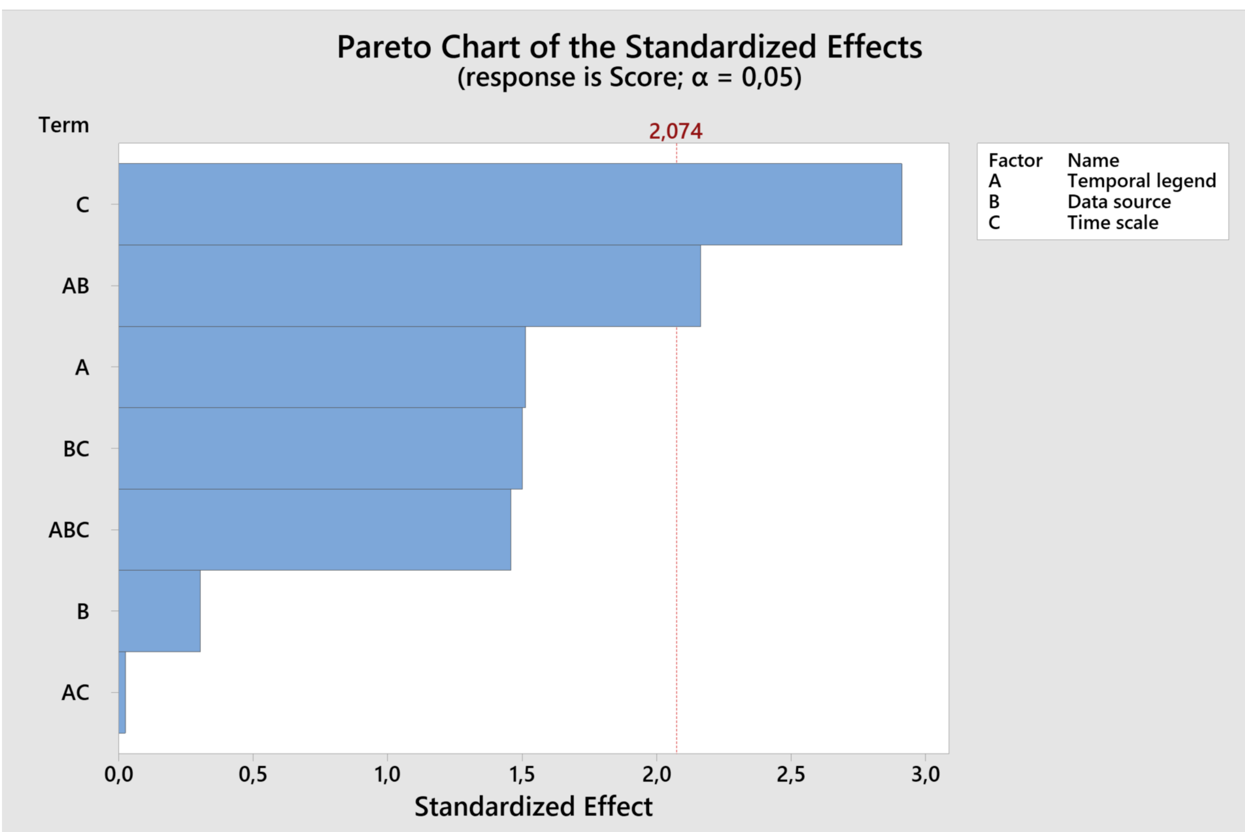


Figure 15: Pareto chart of all factors and interaction in the experiment.

Main Effects Plot

Figure 16 show the Main Effects plot from the experiment. This plot shows how each level of the main factors affect the response variable [39]. Lower values imply that the level resulted in more accurate observation, while higher values imply that the observations were more inaccurate. A nearly horizontal line, such as the one for Data Source, means that there is almost no main effect. Non horizontal lines means that there are main effects, and steeper lines means greater effects.

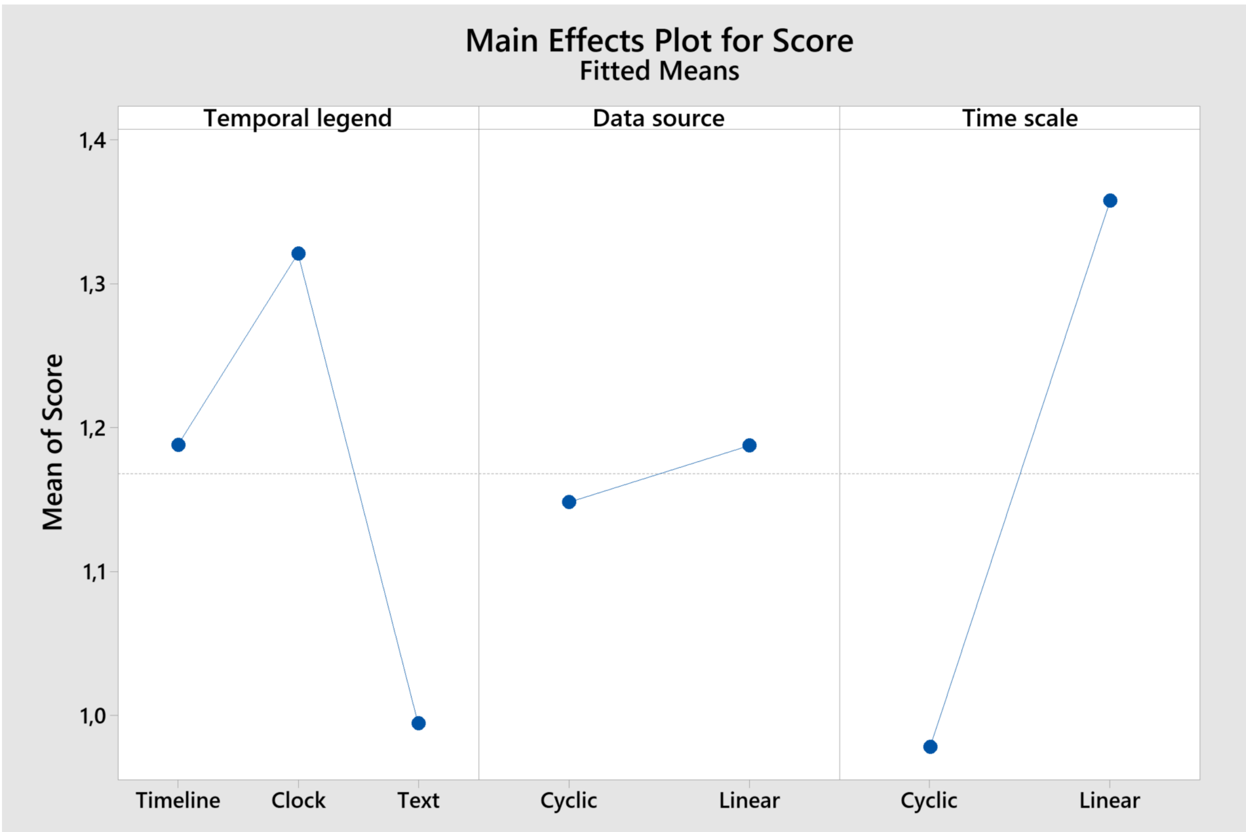


Figure 16: The effects of each main factor in the experiment.

Interaction Plot

The Interaction effects from the experiment can be seen in figure 17. This plot shows how the interaction between main factors affect the response variable [40]. Parallel lines imply that there is no interaction effects between factors, while nonparallel- or differently shaped lines suggest that some interaction occurs. From the plot we observe the following:

- In the interaction plot between Temporal Legend and Time Scale the two lines have the same shape, indicating that there is no interaction between these factors.
- In the interaction plot between Data Source and Time Scale the two lines are not parallel, indicating that there is some interaction between these factors.
- In the interaction plot between Temporal Legend and Data Source the two lines form completely different shapes, indicating that there is some interaction between these factors.

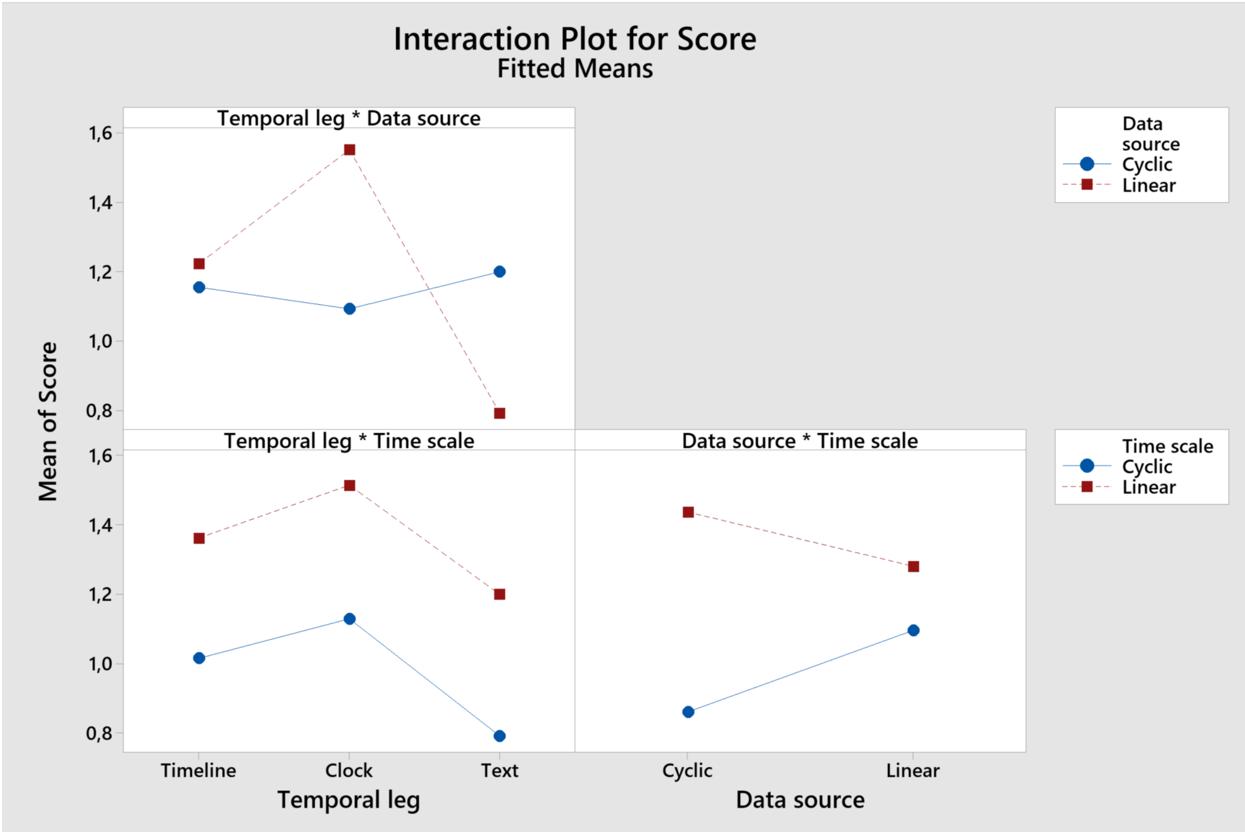


Figure 17: Interaction between factors in the experiment.

5.2 Discussion of analysis results

This section will analyse and discuss the results presented in section 5.1, and discuss the experiment as a whole. Due to the low number of participants in the experiment, it is difficult to draw firm conclusions from the results. This section mainly serves to show how the results can be analyzed, and to provide recommendations for future full scale experiments.

As was explained in section 5.1, a transformation of the variables was needed, as they violated several of the properties necessary for performing the analysis. This might be caused by the low number of participants in the experiment, and an experiment with more participants might not have the same problem. Generally, as the number of participants increases, distributions tend towards normality.

The following key observations can be made from the results: The factor Time Scale, and the interaction between Temporal Legend and Data Source were found to have significant effects. No other factors or interactions had significant effects, though some were relatively close. This is shown in both the ANOVA table in table 3, and the Pareto Chart in figure 15. This section will analyse and discuss the results, beginning with the two key observations.

The first main result we will examine is the significance of the Time Scale factor. This signifies that the different time scales has a significant effect on the response variable from each test observation. Figure 16 shows the way the response variable changes based on each level of Time Scale. As we can see, the line for

Time Scale shows a clear difference between the levels Cyclic and Linear. This means that tests which used a Cyclic Time Scale resulted in better accuracy from the participants than the tests which used a Linear Time Scale.

There are possible reasons for this. As discussed in section 4.2, one lesson from the pilot experiment was that some tests were easier to answer correctly than others. Several of these tests used cyclic time scales, and were shorter than many tests with linear time scales. They were given much less animation time than other tests to make up for this, but this might not have been enough to remove this problem. These tests might still be easier to complete than other tests, leading to the difference in effect between Cyclic and Linear Time Scales. This seems like a plausible explanation for this result, but further investigation is necessary to determine the true cause.

The second main result we will examine is the significance of the interaction between Temporal Legend and Data Source. This signifies that the various combinations of temporal legends and data sources have a significant effect on the response variable from each test observation. A goal of this thesis is to investigate these types of interactions, and to understand whether certain types of data would be better visualized using certain types of temporal legends. Figure 17 shows how the response variable changes based on the combinations of each level from Temporal Legend and Data Source. When we compare this to figure 16, which shows the effect of each level in Temporal Legend, we can make the following observation. The interaction between Temporal Legend and linear Data Source follows the same shape as for just Temporal Legend. This indicates that this level does not interact greatly with Temporal Legend. The level Cycle, on the other hand, has a completely different interaction effect with Temporal Legend. The plot shows that while Text is associated with more accurate observations overall, Clock is better when paired with Cyclic Data Source. In fact, Text has worse results than the two other legend types when paired with a Cyclic data source, despite being the legend type associated with most accurate results.

The results indicate that there might be interactions between these factors, which is worth to investigate further in a full scale experiment. The low number of participants means that the results found here are far from certain, and should only serve to indicate the potential for future research.

There were some comments from the participants regarding the Text legend. During the tests with the Text legend, they could quickly glance at the legend to check the time, and then quickly look back without spending too much time looking at the legend. With the other legend types, this was more difficult to do, as more mental calculations were needed to understand the current time. Figure 16 seems to reflect this, as Text leads to a better Score than the other legends. While the effects of Temporal Legend are not significant, they are in line with the comments from the experiment participants. These comments might indicate that the tests themselves were created in a way which inadvertently suited the Text legend better than the other types. The events the participants looked for were fairly obvious, and the participants did not need to reach a broad understanding about the animation to answer the questions. This should be kept in mind in future experiments within this topic.

While the temporal legend Text led to more accurate observations, the experiment did not find any significant differences between the temporal legends. This is in line with previous research into this topic, as is described in section 2.2.1. Here it is noted that experiments performed by Edsall et al [25] and Midtbø et al [26] did not find significant differences between the three temporal legend types. This highlights a strength with factorial experiments, as significant interaction effects can be found while their main factor components are not significant. The interaction effects can be hidden in experiments that focuses on comparing the legends directly. This is particularly interesting in light of Data Source having the least significant effect of all the main factors in the experiment, while being part of the most significant interaction effect.

There are other results from the experiment which, while not significant, are worth discussing.

We will first examine the interaction effect between Data Source and Time Scale, which can be seen in figure 17. This effect describes how the various aspects of the data sets interact with each other. Based on the interaction plot, we can see a trend which indicates that using a cyclic time scale for cyclic data and linear time scale for linear data could lead to more accurate observations. This is in line with our predictions, as we would expect cyclic phenomena visualized with a cyclic time scale to complement each other. Still, these results are not significant, but could be investigated further to properly evaluate the effect.

From the Pareto chart in figure 15, we can observe that the main factor Data Source (Term B) is the least significant of all the main factors. This factor describes what kind of phenomenon the data for the test was drawn from. It is expected that whether a phenomenon is cyclic or linear should not significantly impact the observers ability to read the data, which is what we observe. While Data Source is not significant, the interaction effects it is associated with are much greater than its main effect. This indicates that this factor should be taken into account in future experiments, as it does have an impact on several other factors.

Another observation that can be made from the Pareto chart in figure 15, is the interaction effect between Temporal Legend and Time Scale (Term AC). This is the least significant effect of all the effects present in the experiment. These factors are related in some ways, as each temporal legend is used to visualize each time scale. It could be reasonable to believe that a clock legend would be more suited for showing a cyclic time span, or that a timeline would be more suited for showing a linear time span, but no such interactions were found between the factors. In fact, given the very low significance of the effect, it is more likely that these factors do not interact.

Having covered the results, the rest of this section will comprise the experiment itself. Here we will discuss various issues that were noticed during the experiment and lessons learned which should be incorporated into future experiments.

Regarding the virtual environment itself, the participants commented that they found it easy to understand and interact with. It was not overly complex, which probably aided this. The participants did not need to shift their attention between too many points of interest during the tests, which also assisted in simplifying the VR experience. Future experiments using VR should ensure that the participants do not get overwhelmed by needing to focus on too many virtual elements.

The column representation of data also got a positive response from the participants, as they found it easy to understand. Since maps in VR have an additional dimension that can be used to display the data, it is important to take advantage of the new possibilities opened by this third dimension. Given the positive feedback, the column representation could form a basis for similar visualizations in VR. It is important to remember that it is only possible to know the values of the transitions between each section in the column. Observers cannot be expected to accurately estimate values between section transitions. Section 3.1.2 explains this further.

The feedback regarding the UI was positive, and the participants found it easy to read and interact with. Due to the Covid-19 outbreak it was not possible to test parts of the experimental framework with volunteers during development. This could have uncovered several of the problems present in implementation, and lead to a better test environment. Future experiments using VR should incorporate testing throughout the development of the test environments to weed out these issues.

One issue that was not discovered until it was too late to solve was the overlap issue between the UI and data columns mentioned in section 4.2. This was a results of the implementation of the data column, and was discovered during the pilot experiment. It did not cause significant problems during the experiment itself, as it could be bypassed relatively easily by careful positioning within the virtual environment. Nevertheless, it does represent the issues that can occur when the evaluation of a test environment is severely limited. Other than this visual issue, no other major issues with the experimental framework were encountered, and the flaws were generally a product of the implementation, and not the underlying design.

Having covered the experimental environment, we will discuss the tests themselves. A few issues were noticed during the experiment and while analyzing the results.

Some issues are related to the data sets used for the tests. The data sets with cyclic time spans were shorter than those with linear time spans, which could have led to these tests being easier. Future experiments should use a wider variety of time spans for the data sets, and try to prevent some test being made easier than others. More extensive testing could have discovered this issue earlier. The current results imply that tests using cyclic time spans are easier to answer correctly, while in reality this could simply be that shorter tests are easier to answer correctly.

Some data sets, in particular the data set describing hours of sunlight, might have been too simple. All participants found that this data set was easy to answer, and the observations for this data set were closer to the true value than any other data set. A more complex data set could be used, but it is important to not make the tests so difficult that they do not give useful results. It is also important that the data sets still have the properties of cyclic phenomena, and if these properties are not clearly present in the data set the analysis will be incomplete. There are limits to the data sets that are available, and it could be difficult to find one that fulfills all criteria.

6 Conclusion

The objective of this thesis is to investigate the effect of different temporal legends in animated maps. In particular, this involves investigating whether data from different phenomena have different effects on each temporal legend.

An experiment was set up to evaluate this issue and VR was chosen as the experiment platform. This experiment was set up as an observer trial. Randomly chosen volunteers would use a developed VR environment to view a visualization of data sets evolving in time(animations), together with visible time indicators. Observers were asked to pay specific attention to several details, which they would be queried about. Due to the fact that the Covid-19 pandemic significantly limited the number of volunteers that could participate in the experiment, the scope of the experiment was reduced. Because of the low number of participants in the experiment, the results should be regarded as preliminary and mainly serves to showcase the potential of a full scale experiment.

A simple test environment was developed in VR for the experiment. This contained a map with animated data and a UI to present the relevant temporal legends to the participants. It was not possible to evaluate and improve the design by performing tests with volunteers during the development, which should be done in future experiments. This is particularly important when developing for new and unfamiliar platforms such as VR. Two volunteers participated in a pilot experiment to evaluate the setup and improve it before the final experiment. Several possible improvements were identified and implemented into the test environment. Finally, a small scale experiment with three volunteers was conducted to obtain preliminary data to answer the research questions.

The results of the experiment indicate that there could be interaction effects between different temporal legends and the phenomena data sets are drawn from. If this is confirmed, it could have implications in the way legends for animated maps are designed. The low number of participants makes this result uncertain, and future full scale experiments are needed to investigate this further. Despite this, the results are promising and showcase the potential of using this method for investigating temporal legends.

There are several issues that should be addressed in future experiments exploring this topic. During the development phase of the experiment, more testing should be performed with volunteers to improve the design of the tests. The experiment itself should be conducted with more volunteers to obtain a result that is more statistically robust. Data sets should have more diverse lengths to avoid interactions between certain factors in the experiment and test difficulty. This could lead to certain patterns emerging where there should not be any. This seemed to occur during this experiment, where tests with cyclic time spans where shorter than tests with linear time spans, which might have made these tests easier than others.

The overall design of the test environments was well received by the test participants, and could serve as a basis for future experiments. The method of visualizing data points with columns can fit several kinds of data, but is unsuited for questions requiring precise data values.

References

[1] Tomasz Mazuryk and Michael Gervautz. “Virtual Reality - History, Applications, Technology and Future”. In: (Dec. 1999).

[2] *Oculus Quest Back In Stock At Oculus.* Accessed: 2020-04-05. URL: https://www.gamespot.com/articles/oculus-quest-back-in-stock-at-oculus/1100-6476020/.

[3] Dom Barnard. *Degrees of Freedom (DoF): 3-DoF vs 6-DoF for VR Headset Selection.* Accessed: 2020-04-05. 2019. URL: https://virtualspeech.com/blog/degrees-of-freedom-vr.

[4] A. Y. C. Nee and S. K. Ong. “Virtual and Augmented Reality Applications in Manufacturing”. In: *IFAC Proceedings Volumes* 46.9 (2013), pp. 15–26. ISSN: 1474-6670 9783902823359. DOI: 10.3182/20130619-3-RU-3018.00637.

[5] Eddie Paddock. *Overview of NASA EVA Virtual Reality Training, Past, Present and Future.* https://nvite.jsc.nasa.gov/presentations/b3/D2_3_Virtual_Reality_Panel_Paddock.pdf/. Accessed: 2020-04-20.

[6] S. P. Henriksen and T. Midtbø. “Investigation of map orientation by the use of low-cost virtual reality equipment”. In: (2015), pp. 75–88. ISSN: 18632351 9783319177373. DOI: 10.1007/978-3-319-17738-0_6.

[7] L. Herman, O. Kvarda, and Z. Stachoň. “Cheap and immersive virtual reality: application in cartography”. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-4.4 (2018), pp. 261–266. ISSN: 1682-1750. DOI: 10.5194/isprs-archives-XLII-4-261-2018.

[8] Erik Fagerholt and Magnus Lorentzon. “Beyond the HUD - User Interfaces for Increased Player Immersion in FPS Games”. In: (Jan. 2009).

[9] Alex Norrman. *User Interface’s Impact on Player’s Immersion.* eng. Umeå universitet, Teknisk- naturvetenskapliga fakulteten, Institutionen för tillämpad fysik och elektronik, 2020.

[10] Sean Endicott. *How to share Oculus Quest games on the TV.* https://www.androidcentral.com/how-share-oculus-quest-games-tv-friends-and-family-see/. Accessed: 2020-05-01. 2019.

[11] Ben Plays VR. *No Man’s Sky VR - Part 2 - Snow Planet.* https://youtu.be/6iXFYKIMaCs?t=697/. Accessed: 2020-05-01. 2019.

[12] Steven M. LaValle. *Virtual Reality.* eng. Cambridge University Press, 2017. URL: http://vr.cs.uiuc.edu/vrbook.pdf.

[13] *Oculus. VR Guidance and Best Practices.* https://developer.oculus.com/design/bp-vision/. Accessed: 2020-03-24.

[14] Alex. *Oculus Rift history - How it all started.* https://riftinfo.com/oculus-rift-history-how-it-all-started/. Accessed: 2020-03-25. 2018.

[15] *The Oculus Rift, Oculus Touch and VR Games at E3.* https://www.oculus.com/blog/the-oculus-rift-oculus-touch-and-vr-games-at-e3/. Accessed: 2020-03-26. 2015.

[16] *Oculus - Introducing Oculus Quest.* https://www.oculus.com/blog/introducing-oculus-quest-our-first-6dof-all-in-one-vr-system-launching-spring-2019/. Accessed: 2020-03-26. 2018.

[17] *Oculus - Play Rift Content on Quest with Oculus Link*. https://www.oculus.com/blog/play-rift-content-on-quest-with-oculus-link-available-now-in-beta/. Accessed: 2020-04-30. 2019.

[18] Alireza Mazloumi Gavgani et al. "A comparative study of cybersickness during exposure to virtual reality and "classic" motion sickness: are they different?" In: *Journal of Applied Physiology* 125.6 (2018), pp. 1670–1680. DOI: 10.1152/japplphysiol.00338.2018. eprint: https://doi.org/10.1152/japplphysiol.00338.2018. URL: https://doi.org/10.1152/japplphysiol.00338.2018.

[19] Kamila Łucjan. "Perception of the contents of animated maps". In: *Polish Cartographical Review* 48.4 (2016), pp. 149–160. DOI: 10.1515/pcr-2016-0015.

[20] Menno-Jan Kraak et al. "Cartographic animation and legends for temporal maps: Exploration and or interaction". In: 23 (1997).

[21] M Harrower and Sara Fabrikant. "The Role of Map Animation for Geographic Visualization". In: *Geographic Visualization: Concepts, Tools and Applications* (June 2008). DOI: 10.5167/uzh-8979.

[22] Terje Midtbø and Elina Larsen. "Map animations versus static maps - when is one of them better?" In: (Jan. 2005).

[23] Amy L. Griffin et al. "A Comparison of Animated Maps with Static Small-Multiple Maps for Visually Identifying Space-Time Clusters". In: *Annals of the Association of American Geographers* 96.4 (2006), pp. 740–753. ISSN: 0004-5608. DOI: 10.1111/j.1467-8306.2006.00514.x.

[24] Andrew Frank. "Different Types of "Times" in GIS". In: (Jan. 1998).

[25] R. Edsall, Alan MacEachren, and Donna Peuquet. "Assessing the effectiveness of temporal legends in environmental visualization". In: (1997).

[26] Terje Midtbø, Keith C Clarke, and Sara I Fabrikant. "Human Interaction with Animated Maps: The portrayal of the passage of time". In: *11th Scandinavian Research Conference on Geographical Information Science*, pp. 45–60. ISBN: 978-82-7636-020-2.

[27] Paweł Cybulski. "Design rules and practices for animated maps online". In: *Journal of Spatial Science* 61.2 (2016), pp. 461–471. ISSN: 1449-8596. DOI: 10.1080/14498596.2016.1147394.

[28] Terry Slocum et al. "A qualitative evaluation of MapTime, a program for exploring spatiotemporal point data". In: *Cartographica* 39.3 (2004), pp. 43–68. ISSN: 03177173. DOI: 10.3138/92T3-T928-8105-88X7.

[29] Hermann Villanger. *Visualisering av temporale kart i Virtual Reality*. Prosjektoppgave. NTNU, Trondheim. 2019-12-20.

[30] *Unity Real-Time Development Platform*. Accessed: 2020-05-20. URL: https://unity.com/.

[31] Ronald E. Walpole et al. *Probability & statistics for engineers & scientists*. 9th edition, Global Edition. Boston: Prentice Hall, 2016. ISBN: 9781292161365.

[32] *Minitab*. https://www.minitab.com/en-us/. Accessed: 2020-03-17.

[33] Norsk Klimaservicesenter. *Observasjoner og værstatistikk*. https://seklima.met.no/observations/. Accessed: 2020-04-27.

[34] Folkehelseinstituttet. *Dags- og ukerapporter om covid-19*. https://www.fhi.no/sv/smittsomme-sykdommer/corona/dags--og-ukerapporter/dags--og-ukerapporter-om-koronavirus/. Accessed: 2020-04-22. 2020.

[35] Minitab. *Analysis of variance table for Analyze Factorial Design.* https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/doe/how-to/factorial/analyze-factorial-design/interpret-the-results/all-statistics-and-graphs/analysis-of-variance-table/. Accessed: 2020-05-30.

[36] Minitab. *Residual plots for Analyse Factorial Design.* https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/doe/how-to/factorial/analyze-factorial-design/interpret-the-results/all-statistics-and-graphs/residual-plots/. Accessed: 2020-05-30.

[37] Stephania Glen. *Box Cox Transformation.* https://www.statisticshowto.com/box-cox-transformation/. Accessed: 2020-05-30.

[38] Minitab. *Effects plots for Analyse Factorial Design.* https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/doe/how-to/factorial/analyze-factorial-design/interpret-the-results/all-statistics-and-graphs/effects-plots/. Accessed: 2020-05-30.

[39] Minitab. *Effects plots for Analyse Factorial Design.* https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/anova/supporting-topics/basics/what-is-a-main-effects-plot/. Accessed: 2020-05-30.

[40] Minitab. *Effects plots for Analyse Factorial Design.* https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/anova/how-to/interaction-plot/interpret-the-results/key-results/. Accessed: 2020-05-30.

NTNU

Norwegian University of
Science and Technology