University of Zurich[UZH]

# Visual Complexity of Bike Maps

GEO 511 Master's Thesis

Author
Donat Büchel
16-736-852

Supervised by
Dr. Tumasch Reichenbacher
Prof. Dr. Sara Irina Fabrikant

Faculty representative
Prof. Dr. Sara Irina Fabrikant

23.09.2022
Department of Geography, University of Zurich

# Summary

More and more cities try to encourage residents to cycle more. Therefore, governments are developing comprehensive bike maps to facilitate bicycle trip planning and, as a result, increase the popularity of cycling in general (Pucher and Buehler, 2008). However, research on the topic of bike maps is rare and the versatility of possible features shown on a bike map makes these visually more complex than others. It is critical to understand how maps are perceived and understood to improve their overall design and efficiency (Castner and Eastman, 1984).

The purpose of this thesis is understand how base maps and the display of various cycling related features affect the visual complexity of bike maps. Different metrics (GMLMT, Subband Entropy, Edge Density, Feature Congestion, and Distinct Object-Type Counts) are applied on bike maps to measure visual map complexity. Following that, an eye-tracking experiment with 35 participants is carried out. Five different everyday tasks have to be solved on bike maps with four complexity levels. The experiment aims to find out how base maps and cycling related features influence the effectiveness of a map.

The findings suggest that adding more detail to base maps and displaying more cycling related features on a map resulted in a visually more complex bike map. Size, shape, and color were found to have the biggest influence on the applied metrics. The eye-tracking study discovered that the display of cycling related features can affect the time needed for successful task completion. To deepen the gained understanding, further research should in more detail investigate how base maps influence bike maps efficiency. To gain maximal learning from such studies, large and representative test groups should be examined in a fully randomized manner.

# Contents

# 1 Introduction

"Zurich invites you to cycle" (City of Zurich, 2013). More and more cities try to promote bicycling. There are various advantages to encouraging individuals to cycle more. Cycling causes no noise or pollution and utilizes significantly fewer resources than any other mode of transportation currently available. The energy required for cycling is provided by the traveler, resulting in additional health benefits for the traveler. Cycling also takes up a fraction of the road space required by cars and is significantly less expensive than a private car or public transportation. To summarize, cycling is the most environmentally, socially, and economically sustainable means of transportation (Pucher and Buehler, 2008). As a result, it is not surprising that cities want to promote cycling. Zurich, for instance, aims to highly increase the number of cyclists across all sections of society while also ensuring that they feel and are safe (City of Zurich, 2013). Different aspects can help promote bicycling, such as direct routes, the presence of bicycle facilities, road safety, and others (Rybarczyk, 2014). Zurich wants to invest in infrastructure. For more experienced cyclists, a rapid main cycling network should be established. Easy routes will be constructed for less experienced cyclists that are largely separated from the rest of the traffic (City of Zurich, 2013). Additionally to infrastructure projects, cities publish bike maps to promote cycling. Features that may be important to cyclists are highlighted on these cycling maps. The amount of features that may be seen on a map varies greatly. In some cases, all streets are documented, while in others, only the major routes are represented.

Cyclists are not a homogeneous group. They have differences in terms of abilities, destinations, and consequently different needs. While some use their bicycles to work, others prefer to ride in their leisure time (Wessel and Widener, 2015). Unlike automobile drivers, not all routes are suitable for all types of cyclists. Dill and McNeil (2016) therefore divided cyclists into categories, ranging from experienced to inexperienced cyclists. Different types of cyclists may be interested in different aspects of a bike map. Aspects of features of a bike map might be the availability of bicycle lanes, paved or unpaved streets, one-way streets, dangerous crossings, height differences, bicycle stations, or pumping stations. This versatility of possible features makes bike maps more visually complex than other maps.

It may be possible to improve the overall design and effectiveness of bike maps if it is possible to understand how they are perceived and understood. This train of thought was already described by Castner and Eastman (1984) and applies to the overall scope of map complexity. Since the early 1970s, the field of map complexity has gotten more attention. Nonetheless, research on this topic is still ongoing because no definitive answer has yet been found (Castner and Eastman, 1984). This thesis examines the visual complexity of bike maps, specifically how they are perceived and how they can be improved. This is accomplished by examining base maps and the features displayed on bike maps.

## 1.1 Objective of the Thesis

The goal of this thesis is to contribute to the field of visual map complexity. Various methods, such as eye-tracking or pixel-based quantification tools, have been adapted and tested up to this point. However, comparing real-life data from map readers with pixel-based quantification tools has only happened a few times. Furthermore, almost no research has been conducted in the field of bike maps. As bike maps have not been designed according to any particular scheme, they have a wide range of appearances. This is undoubtedly an area where progress can be made. Aside from that, all research on map complexity contributes to making this field more understandable and, as a result, provides more user-friendly and efficient maps.

## 1.2 Research Questions

In this thesis, the following two research questions will be addressed. These research questions will be answered with the experiment that will be conducted.

**1a)** How visually complex are different bicycle base maps?

**1b)** How does visual complexity of base maps affect efficiency of bike maps?

**2a)** How does the display of different cycling related features affect the visual complexity of bike maps?

**2b)** How does the display of different cycling related features affect efficiency of bike maps?

Research questions 1a and 1b focus on the visual complexity of base maps, whereas 2a and 2b are concerned with the cycling related features provided on bike maps. Research questions 1a and 2a, as well as 1b and 2b are related. 1a and 2a focus on how visually complex different components of bike maps are. To answer those two questions the base maps used in the eye-tracking experiment will be quantified using various prevalent metrics. 1b and 2b focus on the aspect of efficiency. In this context, efficiency refers to the speed with which a task is completed successfully (Çöltekin et al., 2017). The data from the eye-tracking experiment will be used to answer these two research objectives.

For every research question a hypothesis has been formulated:

**1a)** More detailed base maps are visually more complex.

**1b)** Bike maps with visually complex base maps are less efficient.

**2a)** More displayed cycling related features are visually more complex.

**2b)** Bike maps with more displayed cycling related features are less efficient.

# 2 Background

## 2.1 Bike Maps

### 2.1.1 Biking as Mode of Transport

In many countries, biking is not a viable alternative when moving to a given destination because it is neither safe nor appealing. In many western countries bicycling is a marginal mode of transport, utilized rather as leisure activity than for daily travel needs. Furthermore, the socioeconomic distribution of cyclists is uneven; men and young people cycle more frequently. However, this is not the case in every country. In the Netherlands, Germany, and Denmark cycling is ten times more popular than in the UK and the USA, and there is no gender difference when it comes to cycling. People were served in those countries by making their cities people-friendly rather than car-friendly (Pucher and Buehler, 2008).

When cycling is encouraged, there are numerous advantages. Cyclists produce no noise or pollution, and they utilize significantly fewer nonrenewable resources than any other means of transportation. The energy required is provided by the cyclists themselves. This has a positive impact on one's health. Cycling and health have a positive correlation, according to studies. As cardiorespiratory fitness improves, all-cause morbidity and disease risks decrease (Oja et al., 2011). When compared to cars, cycling requires far less room for use and parking. Overall, riding is less expensive than driving a car or taking public transportation. They are less expensive for both the individual and the public infrastructure. According to Pucher and Buehler (2008), "it is hard to beat cycling when it comes to environmental, social and economic sustainability".

Because of the aforementioned advantages, research was conducted to see what elements contribute to increased cycling. Some elements, such as trip distance, topography, and warmer weather, are impossible or difficult to adjust. Nevertheless, elements such as configured roads, and the presence of bicycle facilities can be improved (Rybarczyk, 2014). Cities in the Netherlands, Denmark, and Germany are attempting to promote cycling on a variety of levels, including access to bicycles, cycle trip planning, public awareness campaigns, and citizen participation. On the level of bicycle trip planning, most cities have created comprehensive bike maps (Pucher and Buehler, 2008).

### 2.1.2 Research on Bike Maps

Overall, research on the topic of bike maps is still limited. Nonetheless, there are certain research papers in this section that are worth discussing.

Wessel and Widener (2015) criticize that bike maps produced by city governments to promote cycling, rely on an ideal "typical cyclist". However, there is no such thing as a typical cyclist. Cycling is done for a variety of purposes. People cycle in their leisure time, but some cyclists also want to go somewhere in their daily life. There is a four-category typology for cyclists provided by Dill and McNeil (2016): *The Strong and Fearless*, *The Enthused and Confident*, *The Interested but Concerned*, and *The No Way No How. Strong and Fearless* take part of their identification from riding and would ride whatever the roadway conditions are. *The Enthused and Confident* do ride on roads with cars, but they would prefer their own, separated facilities and are thus keen on improved infrastructure. *Interested but Concerned* people would like to ride but hesitate to do so, as they are afraid. The *No Way No How*, do not cycle for various reasons including topography, inability, or a lack of interest. For the population of Portland the inhabitants were categorized based on this topology. Less than 1% are in the *Strong and Fearless* group. 7% of the people are *Enthused and Confident*, 59% are *Interested but Concerned*. The last group, the *Now Way No How* are about 33% (Dill and McNeil, 2016). Such a categorization is not available for other cities. However, this classification approach backs up the claim of Wessel and Widener (2015) that cyclists are not a homogeneous group.

Wessel and Widener (2015) attempted to design a map for the city of Cincinnati from a cyclist's perspective. They called the situation in which cars overtake cyclists a "fearful friction". This friction is affected by several factors, including speed, elevation, the width of a street, availability of bicycle lanes, and car-free roads. When cars drive rapidly the "fearful friction" gets higher for the cyclists, especially when the speed difference is large. When the road is hilly, this is especially true, as cars drive faster than cyclists. The width of a street is a variable, as more than one lane is available, which generates space between the passing and the passed. For cyclists, on the one hand bicycle lanes can create a sense of security. On the other hand, they can result in faster and closer passing. Car-free routes are the best way to anticipate this friction.

The implementation of the cyclist's perspective can be seen on the Cinncinati Bike Map (Figure 1). The roadways should be presented in such a way that cyclists can see the possibility of "friction". Included is the speed indicated by colors, the width of streets, bike lanes, and elevation. Also included is additional useful information such as traffic signals, water, and bicycle shops. This approach on a bike map is new, although in practice some issues arose, as Wessel and Widener (2015) stated. Many map readers did not examine the legend and just guessed the meaning of the streets' colors. As a result, because the map is not intuitive, the interpretation was frequently inaccurate. For example, some thought that the color scheme reflects a good or bad scale (Wessel and Widener, 2015).

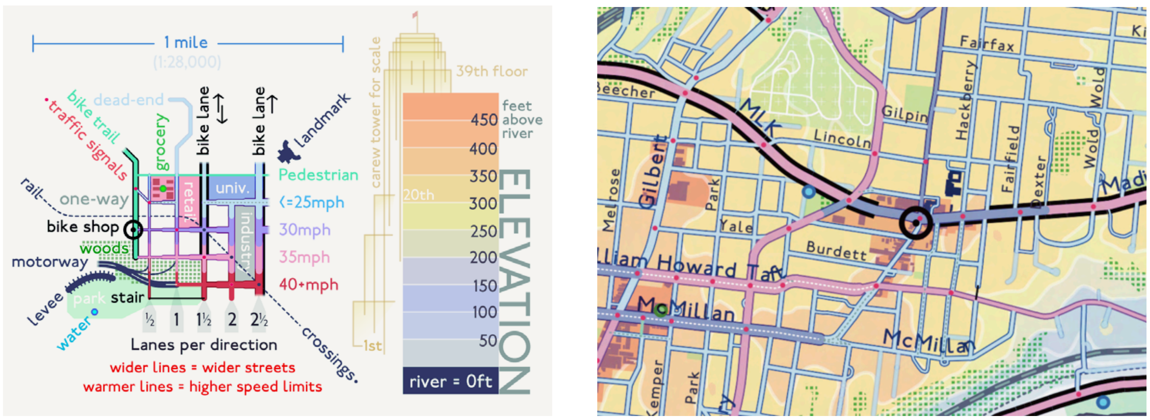



Figure 1: Legend and Map Extract of the Cincinnati Bike Map (Wessel and Widener, 2015)

### 2.1.3 Existing Bike Maps

Because there has been little research on bike maps to date, there are no clear rules on how to design bike maps. Nonetheless, different bike maps are available all over the world. The Copenhagenize Index is the only index that has been designed to analyze the bicycle-friendliness of cities throughout the world. The index contains 13 parameters, including cycling facilities, bicycle-sharing programs, and gender split. Every parameter is ranked from zero to four points, and each city can receive up to 12 bonus points, for a total of 64 points (Zayed, 2016). In 2019 the top five of the index were Copenhagen, Amsterdam, Utrecht, Antwerp, and Strasbourg (Copenhagenize Design Company, 2019). This index will be utilized later in this thesis to conduct a systematic search for existing bike maps.

## 2.2 Map Complexity

Cartographers are interested in the process of map-reading since the early 1970s (Montello, 2002). Their ultimate hope was to learn how their maps were perceived and understood to improve the overall design and effectiveness of maps (Castner and Eastman, 1984). MacEachren (1982) assumed, that map complexity and map effectiveness are negatively correlated, i.e. if a map is more complex, the reader needs more skills to read the map (MacEachren, 1982). This assumption has also been supported by more recent studies. Harrie and Stigmar (2007), for example, claimed that map complexity can affect readability. As there is agreement that map complexity influences the effectiveness of maps, there has not been a conclusive answer to how maps are perceived and understood. As a result, research on the topic of map complexity continues. Even the term “complexity” itself can be defined in a variety of ways, as academics from different fields use the term differently (Schnur et al., 2018).

In Cartography and GIScience map complexity has been researched from different perspectives. All in all, two categories have emerged: visual (or graphic) and intellectual complexity. According to Ciolkosz-Styk and Styk (2011), the two complexity aspects visual complexity and intellectual complexity, correspond to two fundamental aspects of a map: syntactic and semantic. Visual complexity is concerned with the display’s content, what we perceive, and how the visualized information is processed. The two types of map complexity will be described in the following sections.

### 2.2.1 Intellectual Map Complexity

Intellectual complexity refers to the subject (or phenomena) represented in the map (Fairbairn, 2006). Intellectual complexity is determined by the amount of presented information, the type of its presentation, processing level, classification method, and the number of classes (Ciolkosz-Styk and Styk, 2011). Because different map users have varying abilities and knowledge with which they decode map language information, measuring intellectual complexity is exceedingly challenging (Ciolkosz-Styk and Styk, 2011).

### 2.2.2 Visual Map Complexity

Even if the map images are adequately chosen and the objects are legible, users may still struggle to understand the map’s content if the amount of information shown on the map is excessive (Li and Huang, 2002). Visual complexity is determined by the degree of extensiveness, generalization, and visual variable order (Ciolkosz-Styk and

Styk, 2011). Cartographers have more control over the visual than the intellectual component of the map, as the map is easier to influence than map readers.

Barvir and Vozenilek (2020) define visual complexity of a map as the fullness of a map. The density of labels, map symbols and their properties (e.g., form, size, fill), and spatial distribution all influence the fullness (Barvir and Vit, 2021).

Since measuring intellectual complexity is difficult, different studies have developed criteria to determine the graphic map load. The different ways of quantifying visual map complexity will be examined later in this thesis. Alongside the development of metrics, user experiments using eye-tracking became an experimental approach to estimating map complexity (Barvir and Vit, 2021).

## 2.3 Visual Perception

In the following section the process behind vision and the human eye will be explained. This is necessary to understand and draw conclusion from the collected eye-tracking data.

### 2.3.1 Eye Movements

Ware (2019) compares the human eye to a camera. The eye "contains a variable focus lens, an aperture (the pupil), and a sensor array (the retina)" (Ware, 2019). The lens focuses an inverted image onto the photoreceptors of the retina. Two kinds of photoreceptors are found in the retina: cones and rods. Humans have up to 6.5 million cones and up to 125 million rods. Cones can respond to both chromatic and achromatic light, but rods can only respond to achromatic light. Cones and rods are not distributed evenly across the retina. The fovea is dominated by cones, while the periphery is dominated by rods. The fovea is the part of the eye that has the highest amount of detail, which means that vision is sharpest and best for fine-detail vision here. Only in this section visual attractions can be identified by their outline and color (Mangold, 2013 & Hubel, 1995).

The eye needs its eye muscles to position themselves, to ensure that the visual attraction gets into the fovea. If necessary, humans rotate their heads or body to focus on an object. The eyes are in constant movement. The eyes reposition on a new extract in the field of view three to five times every second. Movement is required not only to concentrate on information but also to preserve visual perception. In a study always the same extract has been presented to people. After a while, people's perceptions deteriorated, and they turned into a grey region (Mangold, 2013). According to research, there are eight main types of eye movements, which are classified into movements for stabilization, search, and micromovements (Joos et al., 2017). Here, the four most important movements will be covered: saccades, smooth pursuit movements, vergence movements, and vestibulo-ocular movements (Purves et al., 2001).

Saccades are rapid movements of the eyes that change the point of fixation abruptly. They range in their amplitude from small to big. Examples of small movements can be when someone reads a book. Big saccades are done when a person gazes around a landscape. When a person focuses on a target for a saccade, it takes about 200 ms before the eye movement begins. This delay is because the position of the target with respect to the fovea has to be computed and is called saddic suppression. During this time the eye muscles have to work so that the eyes can be moved at the correct distance and appropriate direction (Purves et al., 2001 & Young and Sheena, 1975).

When the eyes try to maintain a moving stimulus on the fovea, they make smooth pursuit movement. This movement is under voluntary control, as the observer can choose to follow a moving object or not. It is nearly impossible to make this movement smoothly when there is no moving target. Instead, most people do a saccade (Purves et al., 2001).

When each eye has an object at a different distance from the observer, vergence motions are made. Other eye movements have the two eyes move in the same direction (conjugate eye movements). Vergence movements are disconjugate eye movements. When people look at a closeup target that is near to them, their eyes are drawn together. When they look at something in the distance the eyes diverge (Purves et al., 2001).

Vestibulo-ocular movements are made to stabilize the eyes in relation to the external world. An example would be if a person moves their head. The vestibulo-ocular movements prevent the visual images from "slipping". When we look at something and move our head from side to side, our eyes begin to adjust to the movement of our head. As a result, the fixated object stays in the same place in the retina (Purves et al., 2001).

Although it seems contradicting fixations are also classified as eye movements. Fixation takes place when the eye movement stabilizes the retina on a stationary object.

During the fixation event, the eyes do not remain entirely motionless; instead, there are few eye movements. Fixation takes between 200 and 400 milliseconds, and information might be received throughout this time. Fixations and saccades are used to evaluate maps. It is possible to determine where map readers glance at the map and whether they can locate useful information (Ware, 2019). Eye-tracking is a technique that allows researchers to see where a person's eyes are at any given time, as well as the sequences in which their eyes move from one spot to another. Eye-tracking can thus aid in the understanding of visual and display-based information processing, as well as the aspects that influence map usability and readability (Jacob and Karn, 2003).

### 2.3.2 Visual Attention

There has not been a precise definition of what attention is until now. However, there is evidence that the visual system can focus on a sensory attribute while ignoring the others as noise when it comes to visual attention. On this subject, a large number of studies have already been done. The so-called guiding attributes should be highlighted in this section. Wolfe and Horowitz (2004) divided attributes into five categories: undoubted attributes, probable attributes, possible attributes, doubtful cases, and probable non-attributes. Table 1 shows the undoubted and probable guiding attributes. Undoubted attributes are backed up by large amounts of convincing data. Probable attributes would need some more data to clear still existing ambiguities.

| Undoubted Attributes | Probable Attributes |
|---|---|
| Color | Luminance onset |
| Motion | Luminance polarity |
| Orientation | Vernier offset |
| Size | Stereoscopic depth and tilt |
| | Pictorial depth cues |
| | Shape |
| | Line termination |
| | Closure |
| | Topological status |
| | Curvature |

Table 1: Undoubted and Probable Attribute that Might Guide the Deployment of Attention (Wolfe and Horowitz, 2004)

## 2.4 Eye-Tracking

In the following section, different existing eye-tracking techniques will be shown. Following that, other metrics that can be monitored using eye-tracking will be introduced. This section concludes with different studies on map complexity with eye-tracking that already have been conducted.

### 2.4.1 Techniques

Eye-Tracking technology was pioneered over a century ago (Jacob and Karn, 2003). Since many different approaches have been used. The electro-oculographic technique was one of the initial attempts. To assess potential variations, skin electrodes were placed around the eye (Young and Sheena, 1975). A more detailed description and other techniques were discussed by Young and Sheena (1975), Barea et al. (2002), and Jacob and Karn (2003).

In the research for this thesis, the so-called corneal-reflection technique is used. The trackers in this method are made up of a regular desktop PC with an infrared camera positioned beneath the monitor. The participants are shown so-called stimuli on the screen. Infrared determines the features of the eyes needed to track them in conjunction with image processing software. The camera's infrared light is focused on the eye to create a bright reflection that makes them simpler to track. The light entering the retina makes the pupil appear as s bright disc and also the corneal reflection appears as a small glint (Jacob and Karn, 2003). The larger white circle in Figure 2 represents the pupil, while the smaller white circle represents the corneal reflection. The software needs to be able to define the center of the pupil and the location of the corneal reflection. Then, using trigonometric calculations, the point-of-regard can be computed.

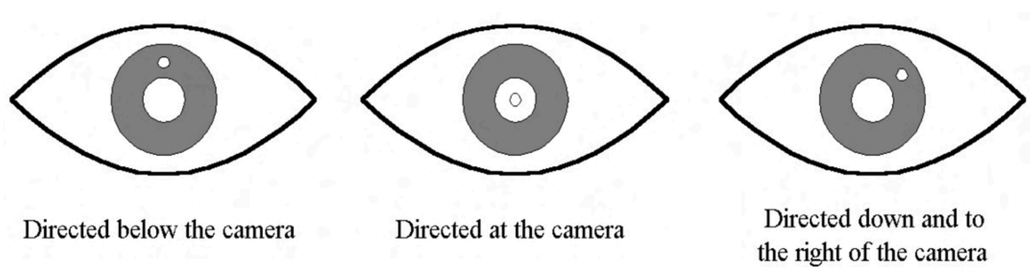



Figure 2: Corneal Reflection Position According to Point of Regard (Jacob 2003)

### 2.4.2 Metrics

After having discussed the general information about eye movements and different measurement techniques, the most frequent eye-tracking metrics will be summarized. The most used eye-tracking measurements are fixation, saccades, and scan paths. Different metrics can deviate from these measurements. In the appendix, a list of fixation-derived metrics is attached. For completeness have a look at Jacob and Karn (2003).

Fixation occurs when the retina of the eyes is stabilized on a stationary movement. Fixations and saccades are therefore readily distinguished. Fixation time is sometimes referred to as saccadic reaction time or intersaccadic interval. Fixation duration is the time between the beginning and the end of a saccade. Although fixation normally lasts from 200 ms to 400 ms, some research has found that short fixations of 50 and 100 milliseconds exist. However, information can not be processed by the viewer, when the fixation duration is too short (Joos et al., 2017). In general fixation duration tend to be much longer, for reading around 225 milliseconds, 275 milliseconds for visual search activities, and 300 milliseconds for picture perception. Furthermore, there appears to be a correlation between the fixation period and the task's difficulty, or required precision (Joos et al., 2017). Eye movement metrics that can be deviated from fixations are listed in the appendix.

During saccades no encoding takes place. Nevertheless, backtracking eye movements can act as a measure of processing difficulty (Jacob and Karn, 2003). In the appendix, metrics derived from saccades are listed.

Scan paths describe a complete sequence of saccade-fixation-saccade. When looking at a scan path of a search path, the optimal scan path would be a straight line to the desired target, with a short fixation on the target (Goldberg and Kotval, 1999 & Jacob and Karn, 2003). Again, the table for scan path-derived metrics can be found in the appendix.

Furthermore, blink rate and pupil size can be considered metrics, as they correspond to cognitive activity. A lower blink rate can suggest a higher workload, while a higher blink rate could indicate exhaustion. Larger pupils could also be a result of a higher cognitive workload. Blink rate and pupil size, on the other hand, could be caused by a variety of different things. As a result, they are employed less frequently in eye-tracking studies (Jacob and Karn, 2003).

### 2.4.3 Visual Map Complexity and Eye-Tracking

Eye-tracking has been used in a variety of fields of study. One method for estimating map complexity that arose was user studies using eye-tracking (Barvir and Vit, 2021). Mapmakers were curious as to when and where specific symbols, elements, or zones of a map are fixated. They aimed to find solutions to map design challenges (Castner and Eastman, 1984). In the beginning, it was difficult to interpret the data coming from the eye-movement experiment. The distinction between whether responses are due to the test-map design or the respondents' ability is particularly difficult. However, it became evident that eye-gazing is a highly selective and purposeful activity. It was discovered that a majority of eye fixation occurs in areas where the information load is high, or it contains unpredictable or unusual details (Mackworth and Morandi, 1967).

Castner and Eastman (1984) distinguish between spontaneous looking and task-specific viewing. When there is no task, spontaneous looking, also known as free examination, occurs. When the eyes must solve a visual challenge, task-specific viewing occurs. When seeing a task-specific display, people tend to be significantly more focused on certain features of the display. Also, more peripheral vision takes place, as possible fixation sites are spotted and quickly assessed to see if it matches the viewing goal (Castner and Eastman, 1984).

Eye-Tracking has been employed in reference and thematic maps. Three studies will be shown in the following sections to provide an overview of how eye-tracking and its metrics might be employed in a geographic setting.

Çöltekin et al. (2017) compared two legends of a soil-landscape map. The first legend listed the categories in alphabetical order, while the second legend listed them by perceptual grouping. Twenty people took part in the testing, and 90% of them said they were comfortable with maps. The transitions between the legend and the map, areas of interest (AOI), and a qualitative examination of the motions were all used to study eye movements. According to the findings, persons with strong map interpretation skills made fewer transitions between the legend and the map. In addition, Çöltekin et al., 2017 used density heat maps to conduct a qualitative examination between the high- and no-ability groups.

Keil et al. (2020) investigated the impact of visual map complexity on the attentional processing of landmarks using reference maps. The study included 57 students of the Ruhr University Bochum. Fixation counts and overall fixation length were calculated using AOIs placed on several landmark pictograms. They sought to see if landmark representations near the depicted path were more likely to be fixed.

Liao et al. (2019) measured the influence of map label density on the perceived visual complexity of maps. Their study included 40 people conducting a visual search. A target name was shown on the screen and after five seconds it disappeared and the participants had to find the target. The shown stimuli had varying label densities. The eye-tracking experiment showed a positive correlation between the label-density and response time.

## 2.5 Approaches to Quantify Map Complexity

There have been several approaches to finding a suitable computation for visual map complexity up until now. In this thesis, complexity quantification will be used to estimate the visual complexity of the stimuli shown to the participants. The used metrics are Graphic Map Load Measurement, Subband Entropy, Edge Density, Feature Congestion, and distinct object-type counts. This chapter will provide the study's measures as well as some additional approaches.

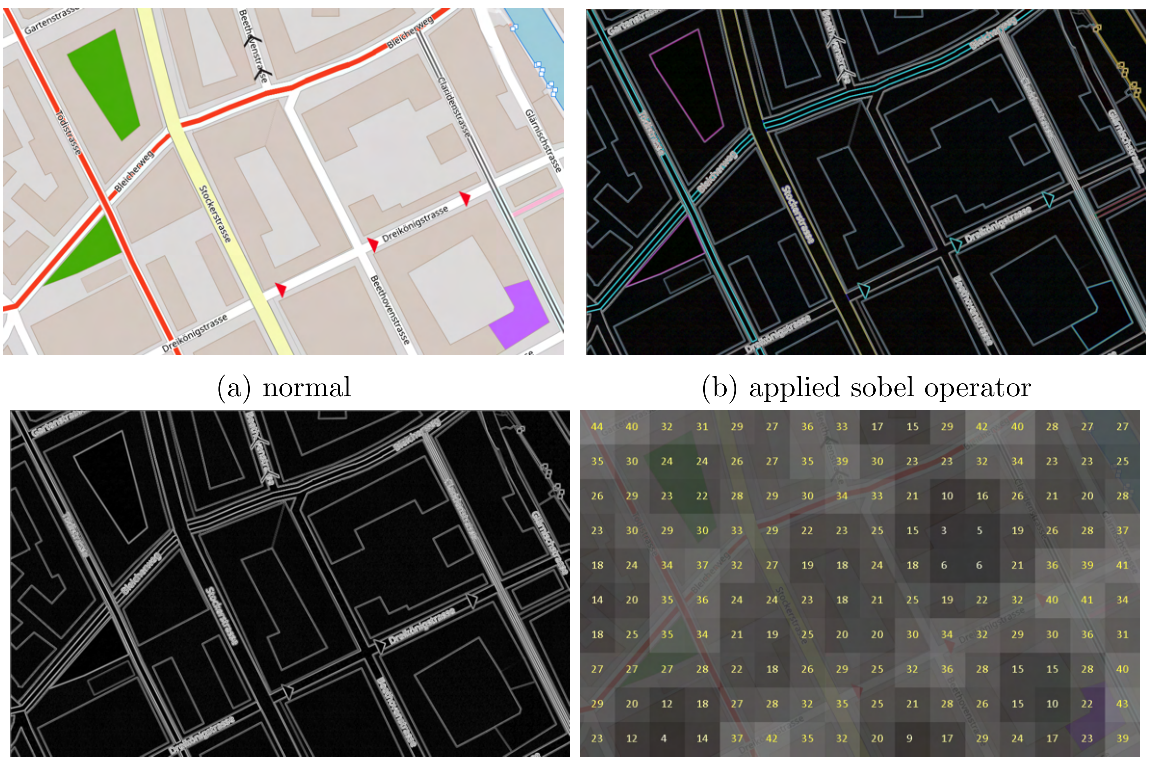
### 2.5.1 Pixel-Based Quantification

#### Graphic Map Load Measurement Tool (GMLMT)

There have been numerous approaches for calculating a map-load value to assume the complexity of a map. None of these, however, have been commonly used. Therefore, Barvir and Vit (2021) wanted to develop a simple and freely available tool. From this intention, the Graphic Map Load Measurement Tool (GMLMT) has been created.

The goal was to build the GMLMT around an edge detection approach, as prior research had shown that this method had a lot of promise. They chose the open-source software The GIMP Development Team (2019) (GNU Image Manipulation Program) as their application. Various edge detection filters are included in this software. They implemented all of the filters in their code and tested them on 26 distinct map extracts, both reference, and thematic maps. As potential filters, they used Sobel, Prewitt, Gradient, Differential, Laplace, and Neon. As a result, for the GMLMT the Sobel operator was chosen, as this filter outperformed the others. The Sobel operator's principle of operation will not be discussed in this context. For a better understanding see Vairalkar and Nimbhorkar (2012).

The GMLMT was created to function with any picture, but it is essential to ensure that the image has a resolution of 100 DPI in order to compare the images on an equal footing. Figure 3 shows the different steps the GMLMT performs. First, the Sobel operator is applied to the map extract. The image is then transformed to monochrome. In the next step of the GMLMT script, the histogram is utilized to calculate the average pixel value of the monochromatic image. For easier comprehension, the range of 0 to 255 is converted into percentage values between 0% and 100%. The map-load level is then calculated as the average of the map's current structures. The computed gets saved into a text file, in addition, the image is given a grid indicating which parts of the image have a higher load (bright tones) and which have a lower load (dark tones) for visualization purposes. The map-load for this example was 12.8%, but without comparing this value is worthless unless it is compared to other maps.




(a) normal (b) applied sobel operator

(c) monochromatic (d) Output

Figure 3: Application of the GMLMT on a Map Extract

### Edge Density

The Edge Density quantification, like the GMLMT, is based on edge detection. Oliva et al. (2004) was the first to introduce this method. To obtain the Edge Density measure for visual clutter, Rosenholtz et al. (2007) applied MATLAB (2019)'s Canny edge detector. Again, the edge detector's technological functionality will not be examined in detail. See Ding and Goshtasby (2000) for a better understanding.

There are various parameters that can be changed in this detector. The low and high thresholds, which can be modified manually, are the most significant. Weak edges are only kept by the Canny edge detector if they are related to strong edges. Figure 4 shows how the Canny edge detector, which is similar to the Sobel operator, is used.

Figure 4: Canny's Edge Detector on a Map Extract

A pixel density measurement is required to produce the Edge Density metric. The histogram in the Gimp software (The GIMP Development Team (2019)) was employed in this specific scenario. For the example map extract shown in Figure 4, 5.3% of the pixels are edge pixels.

### Subband Entropy

Subband Entropy is based on the notion that as a picture becomes more crowded, the number of bits required for subband image coding will increase. Rather than clutter density, this metric assesses spatial uniformity. In the first stage, wavelets are employed to approximate the image content. To approximate a complex image, a higher diversity of wavelet coefficients is required. Thus, Subband Entropy measures how hard it is to encode the information that is present in the image. Various so-called subbands of the image, such as brightness, chrominance, color, and edge orientation, are taken into account to do so Speed et al., 2017. For the calculation of this entropy, Rosenholtz et al. (2007) provides a MATLAB (2019) code. The value of the Subband Entropy for the map extract shown in Figures 3 and 4 was *3.0489*.

### Feature Congestion

Feature Congestion by Rosenholtz et al. (2007) is a measure for display clutter. Clutter is defined as "...the state in which excess items, or their representation or organization, lead to a degradation of performance at some task" (Rosenholtz et al., 2005). When more and more items are added to a map, there is less space to add new items. This state can be described as Feature Congestion. There are already too many colors, sizes, and shapes that make up a crowded area. The implementation of this measure is done in four steps. First, three scales of local feature (co)variance are computed. Second, this is done across the scale. Third, the clutter is combined across different feature types. As a final step, this is pooled over the space, so that there is one measure output for the entire image. Every step is explained in detail in the according paper by Rosenholtz et al. (2005). Alongside a feature congestion value, a visual image is given as output when running the corresponding MATLAB (2019) code (Figure 5). The feature congestion value for the map in Figure 5 was *4.4513*.

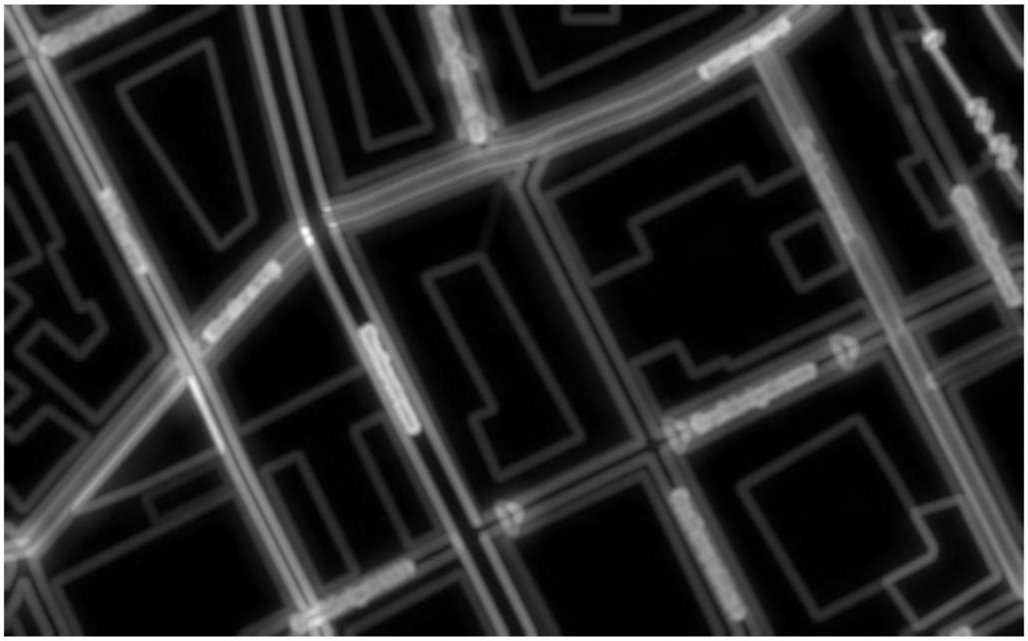


Figure 5: Feature Congestion Clutter on a Map Extract

**Further Pixel-Based Quantification Approaches**

Fairbairn (2006) developed graphical metrics using various measures. The majority of the measurements were made with Fragstats, a program that allows the calculation of various landscape characteristics. Ecologists use Fragstats to study environmental patterns and variety in general. This software can calculate patch-size standard deviation, contrast weighted edge density, double log fractal dimension, landscape shape index, Simpson's and Shannon's Index, and others. Fairbairn (2006) also used other general variables for map quantification like file size or compression.

### 2.5.2 Non-Pixel-Based Quantification

In this section, approaches to quantify map complexity that are not pixel-based will be assessed. In the following, different object-type count metrics will be discussed. Some argue that object-type counts are rather for intellectual than visual map complexity. Nevertheless, they are taken into account, as the boundary between visual and intellectual map complexity is fluent.

### Different Object-Type Count Metrics

Some approaches for quantifying map complexity aim to count the number of objects on the map. However, determining what an object is, is not straight forward. I.e., it is unclear if a road as a whole or certain segment of it count as an object. Harrie and Stigmar (2007) describe different measurements for evaluating map complexities, such as number of objects, number of points in the objects, object line length, object length, the spatial distribution of objects, and spatial distribution of points. These measures were implemented by using Java and further open-source packages (Harrie and Stigmar (2007)). As they only performed these metrics on different generalization levels of buildings, the findings of this study will not be further examined.

### Distinct Object-Type Counts

The distinct object-type count was introduced by Schnur et al. (2010). This calculation is based on the notion that complexity might be affected not only by the total number of individual items on a map but also by the number of categories on the map. The expectation is that an increasing number of distinct map symbols increases human working memory more than the number of times a symbol is used (Schnur et al., 2018). This hypothesis relies on research stating that visual working memory can only store three objects at a time and objects with multiple features are stored as single combined objects (Schnur et al., 2010). When applying the object-type counts every object on a map is counted manually. Objects must be distinguishable based on three of Bertin's visual variables shape, size, and color. Object counts can be divided into several categories. A completed object count with its according categorization can be seen in Figure 6.


Figure 6: Distinct Object Types for a Google Maps Extract (Schnur et al., 2018)

# 3 Methods

## 3.1 Experimental Design

### 3.1.1 Participants

As there is no such thing as a typical cyclist, bike maps should be useful for everyone, from *Strong and Fearless* cyclists to those with *No Way No How*. Thus, there are no restrictions on who can take part in the experiment. The maps were examined with Color Oracle and the participants were asked if they are color-blind in the questionnaire (for Color Oracle visit Jenny, 2020). For the recruitment of the participants, people from the personal surroundings were invited. All in all, data from 35 participants was collected.

### 3.1.2 Variables

Research questions 1a and 1b relate to base maps, whereas 2a and 2b focus on cycling related features. Before conducting the empirical research the included variables and the experimental design are clearly defined.

In the experimental method, statements are made about what circumstances cause a change in behavior. Variables can be used to describe the manipulation and measurements. In the next subsections, the thesis' defined variables are listed.

#### Independent Variable

Independent variables are not influenced by the behavior of participants. This variable is controlled by the experimenter (Martin, 2008). The visual map complexity of the base maps and the cycling related attributes are the independent variables in this study. There will be two levels of complexity for both (see Figure 7 for an example). Furthermore, the various tasks might be viewed as independent variables.

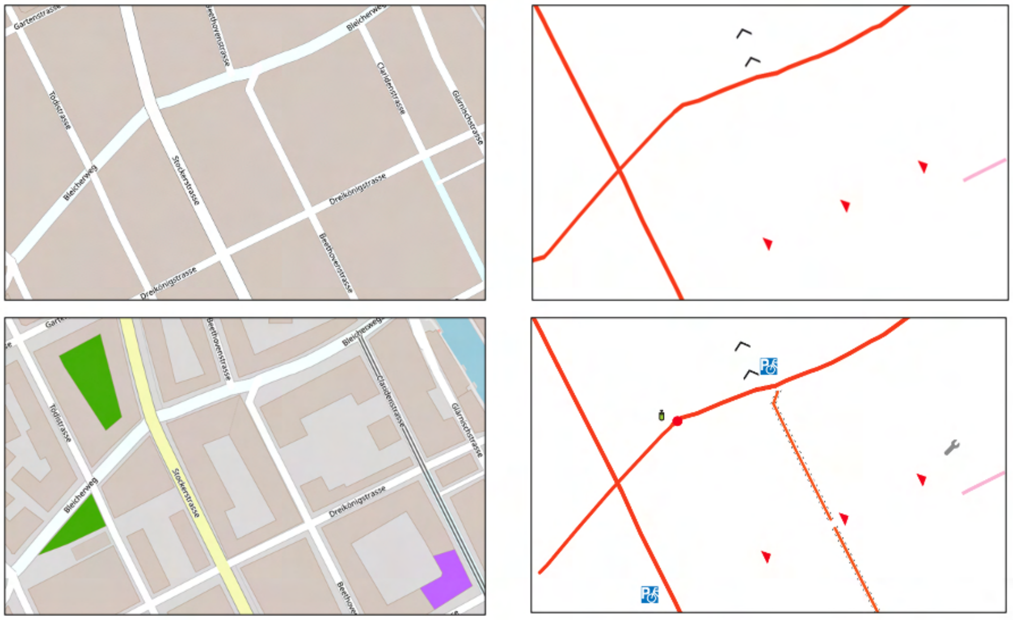


Figure 7: Two Levels of Complexity for Base Maps (left) and Cycling Related Features (right)

### Dependent Variable

When an experiment is conducted, the participants' behavior is measured concerning the various independent factors. The measured behavior is called the dependent variable, as this variable is dependent on what the participant does. In this case, the dependent variable is eye-tracking data and its different metrics that can be derived. In this thesis, time to first fixation, fixation count, heat maps, and task length are the dependent variables.

Hypotheses are formulated before conducting an experiment. Hypotheses are statements concerning the expected relationship between the dependent and independent variables (Martin, 2008).

### Control Variable

Independent variables are changed on purpose and dependent variables are measured, however, also other factors influence the experiment. Controlling them is one method to deal with them. Then they become control variables. Examples of control variables are characteristics of the room the experiment is conducted in (Martin, 2008). Thus, the setup in the room and its condition (for example temperature, lightning condition) will stay the same for all participants during this experiment. Later in this chapter, the specifics of the setup will be introduced. Age, gender, and prior familiarity with bike maps, and maps in general are also control variables in this investigation. This data will be obtained using a questionnaire.

The questionnaire contains socio-demographic questions, such as age, gender, and field of work. Questions about (bike) map familiarity, different map providers, and how often maps are used. Also, a binary question of whether the participant is attested color-blind is included.

### Factorial Design

The number of variables plays a role in deciding which experimental design to utilize. For this study, a factorial design was adopted. The design for this thesis includes two variables, each with two levels. This corresponds to a two-by-two factorial design with four cells (Martin, 2008). Figure 8 shows a combination of the complexities shown in Figure 7. The four cells will be abbreviated with *BM1CRF1*, *BM1CRF2*, *BM2CRF1*, and *BM2CRF2*. BM relates to the base map, while CRF relates to cycling related feature. The numerals 1 and 2 represent the degree of difficulty. For example, BM1CRF2 represents a map with low complexity in terms of the base map (BM = 1) and high complexity in terms of cycling related features (CRF = 2).

The factorial experiment allows the study of the interaction of different variables. Interactions can appear when the participant's behavior on one independent variable is dependent on the level of a second variable. This method is appropriate when the base map and features on the map are viewed as two separate map contents (Martin, 2008).

One drawback of the factorial design is that the investigated factors can quickly add up, resulting in a large matrix with numerous cells. As a result, the experiment will take more time and effort to complete. This is not a problem in this thesis because the matrix only has four cells. When it comes to analyzing the results, there is a second disadvantage or possibility. Often an analysis of variance (ANOVA) is used for the statistical procedure. A normal distribution is required for the analysis of variance to work. If the obtained data is not normally distributed, this analysis

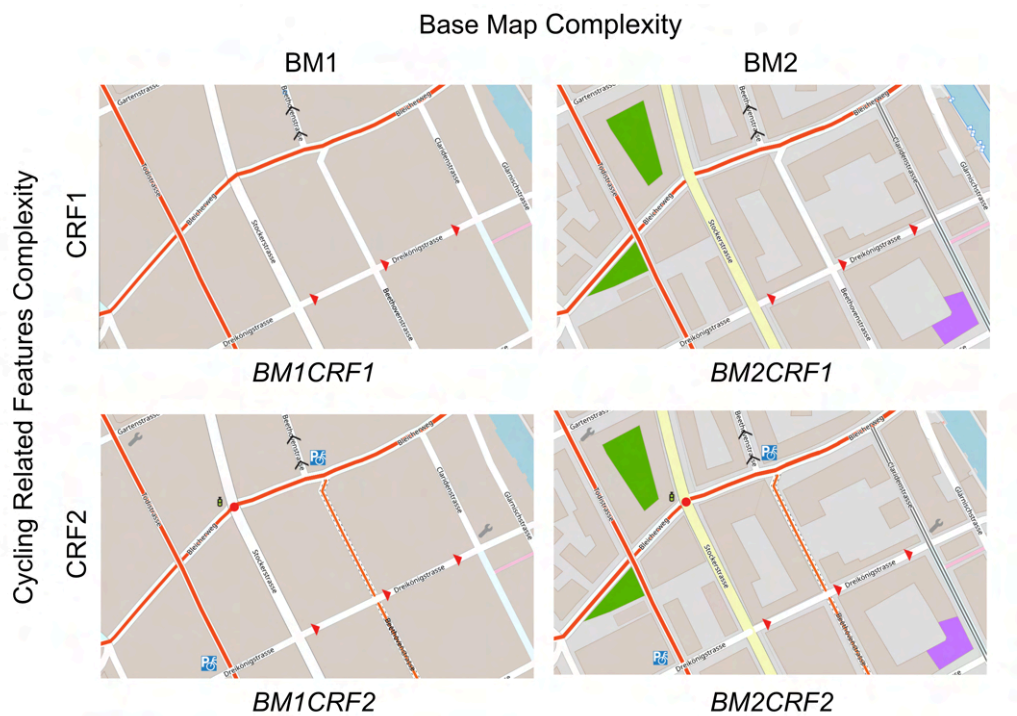



Figure 8: Example of a 2x2 Factorial Design

is irrelevant (Martin, 2008). More details of the statistical method for analyzing the eye-tracking data will follow later in this chapter.

## 3.2 Materials

There is no general agreement on the best way to design bike maps. Therefore, creating the stimuli that are shown in the experiment is not trivial. The Copenhagenize Index was utilized to gain an idea of how bike maps are most commonly designed. The top twenty cities in the world in terms of bicycle friendliness are listed in this index (Copenhagenize Design Company, 2019). For the stimuli design, bike maps for those cities are searched manually on the internet. Twelve of the top twenty cities offer bike maps that are suitable for this thesis. Some cities are relying on interactive online maps, which are certainly useful but cannot be included in this analysis. The next stage was to look at what is displayed and how it is displayed for base maps and cycling related elements. The findings of this search will be taken into account when the stimuli are designed. It is crucial to note, however, that not all observable elements can be included in the map because the map must be coherent.

### Quantification Measurements

When creating the stimulus, the quantification metrics were crucial, as the measured complexity of the four stimuli should differ. The best-case scenario would be if BM1CRF1 was less complex than BM2CRF2 and BM1CRF2 and BM2CRF1 were in between having the same complexity. Therefore, the complexity of the stimuli was frequently measured during the design process to identify a balanced set of stimuli that fit this aim.

The GMLMT was mostly utilized to calculate complexity during the design phase because it was established most recently and considers other, older metrics (Barvir and Vit, 2021). Nevertheless, also Feature Congestion and Subband Entropy was measured from time to time. These three, as well as other metrics, will be applied to the stimulus in the result section.

### 3.2.1 Stimuli

Rather than coming up with a fictional map, a real city is picked so the map has a context. Various requirements must be met before choosing a city. The city should have enough bicycle infrastructure. Because maps of various complexities will be constructed, there must be sufficient features that can be depicted on the map. Additionally, the city should be unknown to the participants. Participants who are familiar with the city depicted in the experiment may have an advantage. Following considerable research, the decision was made to decide on Nashville, which meets all of the previously specified criteria. Most of the attendees are probably unfamiliar with Nashville. The city of Nashville has enough infrastructure allowing the creation of more complicated maps. There is also a cycling map that can be taken and customized for its different purposes (see Figure 9).

The Nashville bike map is imported into Affinity Designer (see Serif Ltd, 2022), where the stimuli set is developed. The map extract has to be reduced, as the participants have to be able to see everything clearly on the computer screen without having to move closer, which would interfere with the eye-tracking device. The design of the base map, as well as cycling related characteristics, will be covered in the following sections.

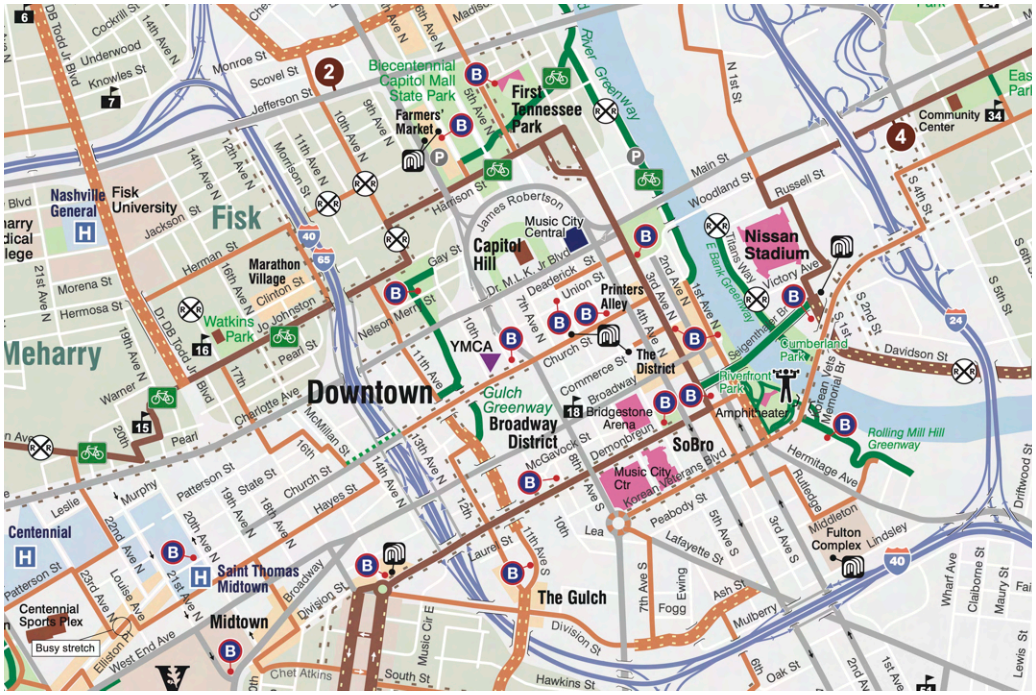
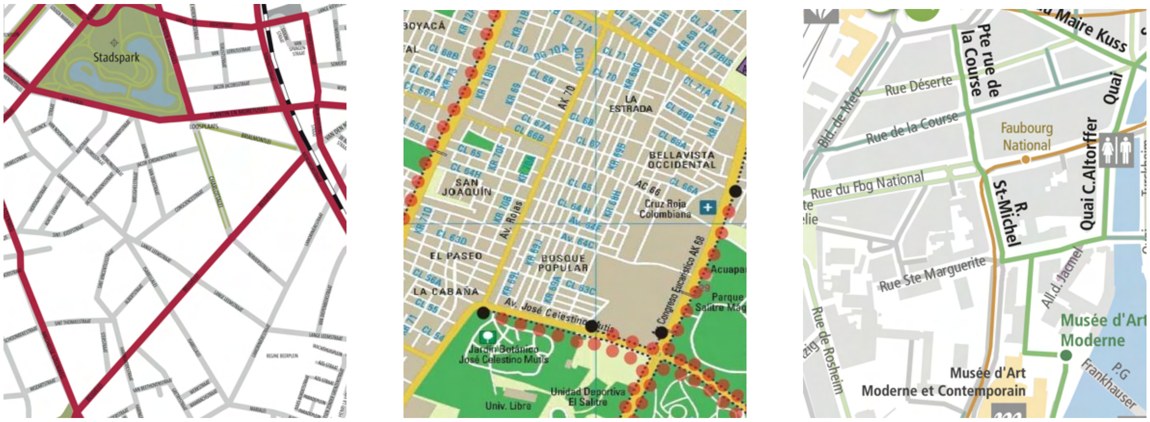



Figure 9: Extract of the Nashville Bike Map (Informing Design, Inc, 2015)

### Base Map

The gathered maps show a wide range of different base maps. Every map includes some elements, such as water bodies, streets, and parks. Other features are not always depicted on maps. Roughly, the base maps can be classified into three complexities (see Figure 10 for examples). The least complex (see Figure 10a) only contains streets, railways, parks, water bodies, and the corresponding labels. Figure 10b depicts a higher complexity, with additional information and varied colors. In the example of Bogota (see Figure 10c), the bike infrastructure seems to have been drawn over an existing map, which is a common practice. In the group with the highest complexity, even the structure of the buildings is represented.




(a) Antwerpen (Slim naar Antwerpen, 2019)

(b) Bogota (Instituto Distrital de Turismo, 2015)

(c) Strasbourg (Vélohop, 2020)

Figure 10: Example of Maps with Different Base Maps Complexities

For the implementation in the experiment, a low-complexity base map and a high-complexity base map were chosen. The following features were chosen for the low-complexity base map (BM1):

- street
- highway
- park
- water body
- street label
- park label

A normal, intuitive depiction is chosen for the style of the features. Water bodies are blue, whereas parks are green. Highways are also shown in blue to differentiate them from the bicycle infrastructure. Because street and park labels are represented in different colors, they are mentioned individually. Figure 11 shows BM1.

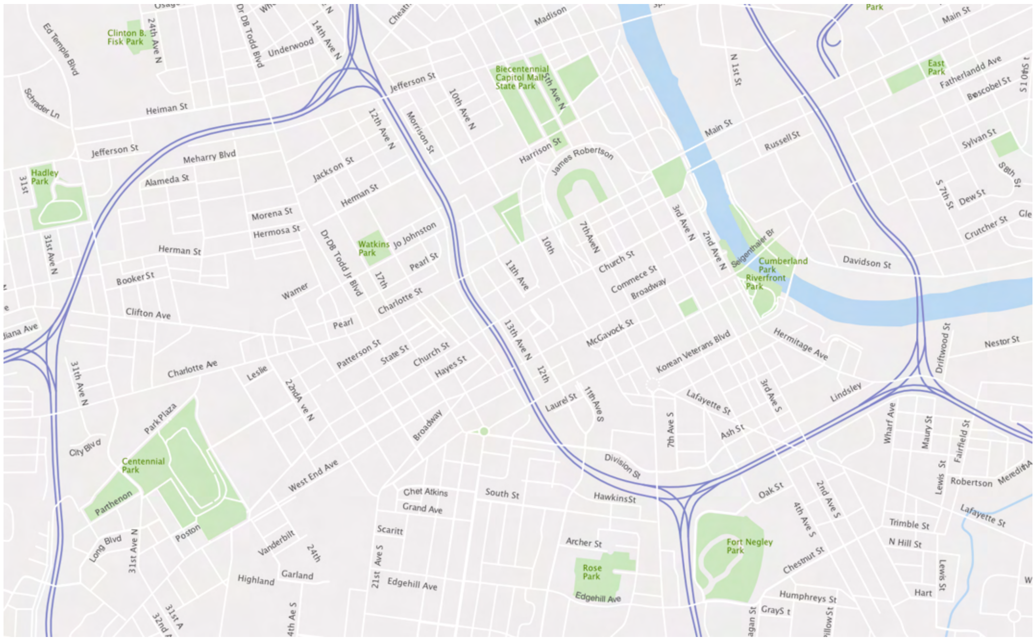



Figure 11: Base Map with Low Complexity (BM1)

The higher base map complexity displays the same features as the low complexity map, but the following features have been added as well.

- Hospital
- Building

Note that alongside the two new features, the number of labels has been raised. The hospitals are blueish, whereas the buildings are grizzly. For the buildings data from OpenStreetMap was added to the map (OpenStreetMap contributors, 2022). Again, it is important to mention that features that are relevant for cyclists should be represented, but also a big enough increase in visual complexity has to be provided. The base map with the higher complexity is visible in Figure 12.


Figure 12: Base Map with High Complexity (BM2)

## Cycling Related Features

Looking at cycling related features on different maps, it is striking that there are many distinct features that may be displayed on a map. The most common were undoubtedly cycle tracks, bicycle lanes, shared lanes, bicycle stations, pedestrian areas, and repair shops. Some maps included bicycle paths that were either proposed or under development. The analysis shows that circumstances change from city to city, particularly when looking at bike maps from multiple continents. The majority of bicycle tracks, lanes, and roadways are colored. Bicycle lanes are often depicted dashed on bike maps, as they are frequently found in reality. Bicycle stations are often represented by a bicycle beneath a roof, and repair shops by pliers.

The decision on which features to display and what term to use is based on the Nashville bike map. The selection of the exact terms are negligible for the tasks. Here are the features that were used for the lower complexity of cycling related features:

- Bicycle Rack
- Bicycle Route
- Physically Protected Bicycle Lane
- Bicycle Lane
- Non-cycling Road

The bicycle racks are a monochrome symbol. All bicycle lanes are red, however, dots represent bicycle lanes and smaller lines represent physically protected bicycle lanes. Non-cycling roads are grey. Figure 13 depicts the low complexity of the cycling related features layer (CRF1).

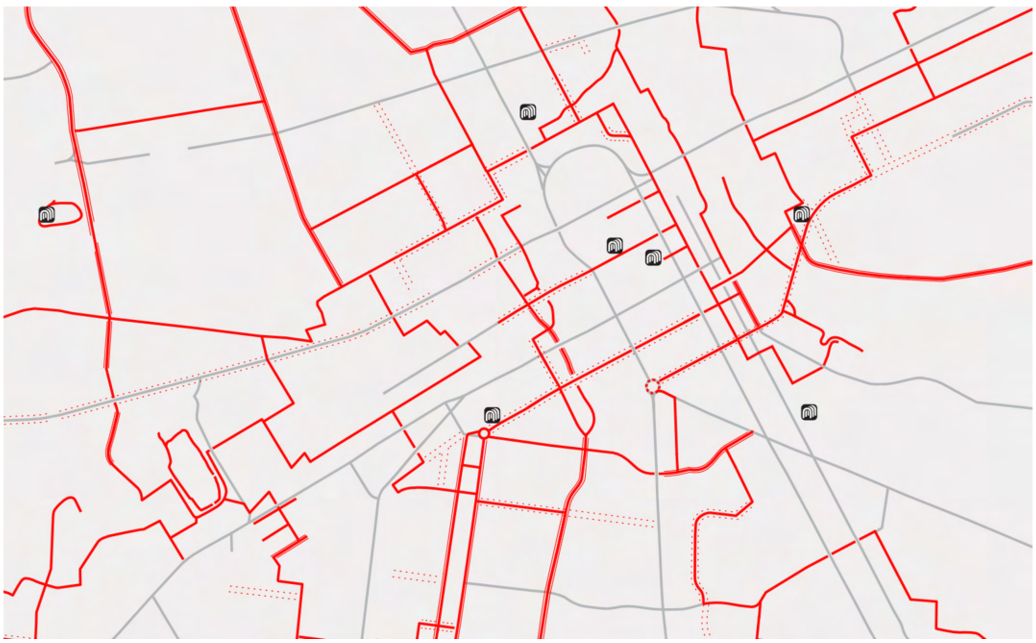


Figure 13: Cycling Related Features with Low Complexity (CRF1)

The second complexity level of cycling related features comprises more features. To begin with, there are more symbols on the map overall. Second, the bicycle paths have been further separated into two categories. Third, there are designated "easy-riding zones" (see Figure 14). For the increased complexity, below is the whole list of cycling related features. The features from the lower complexity are present in the higher complexity level as well:

- Pumping Station
- Railway Crossing
- Bicycle Sign
- Bicycle Rental
- Main Bicycle Route
- Off-Street Bicycle Route
- Easy-Riding Zone

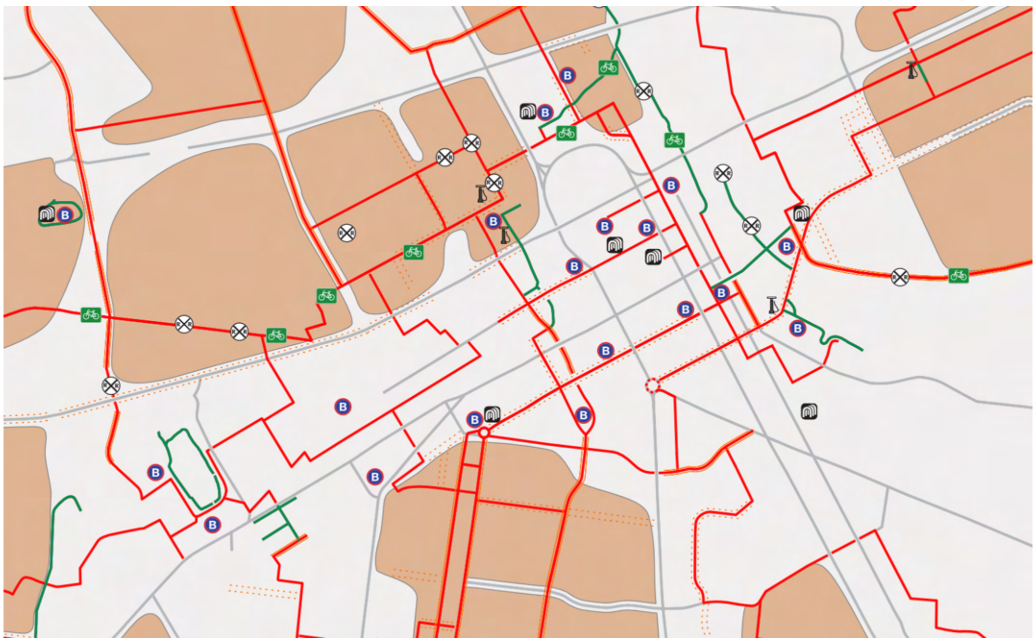


Figure 14: Cycling Related Features with High Complexity (CRF2)

## Combination of the Layers

The combinations of the layers to the four stimuli used in the experiment are shown in Figure 15 to 18.


Figure 15: BM1CRF1

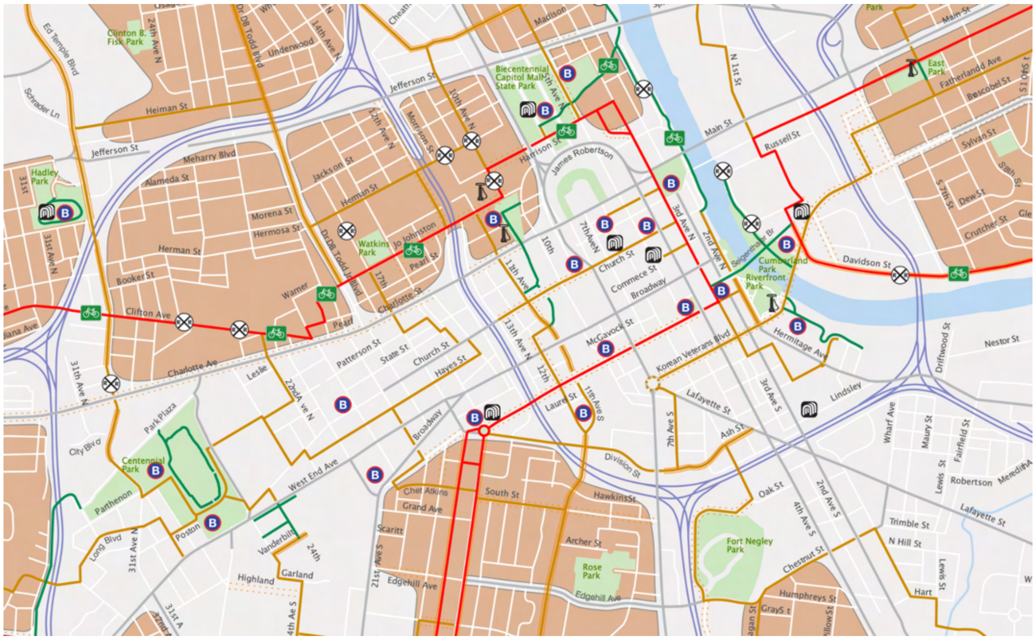


Figure 16: BM1CRF2

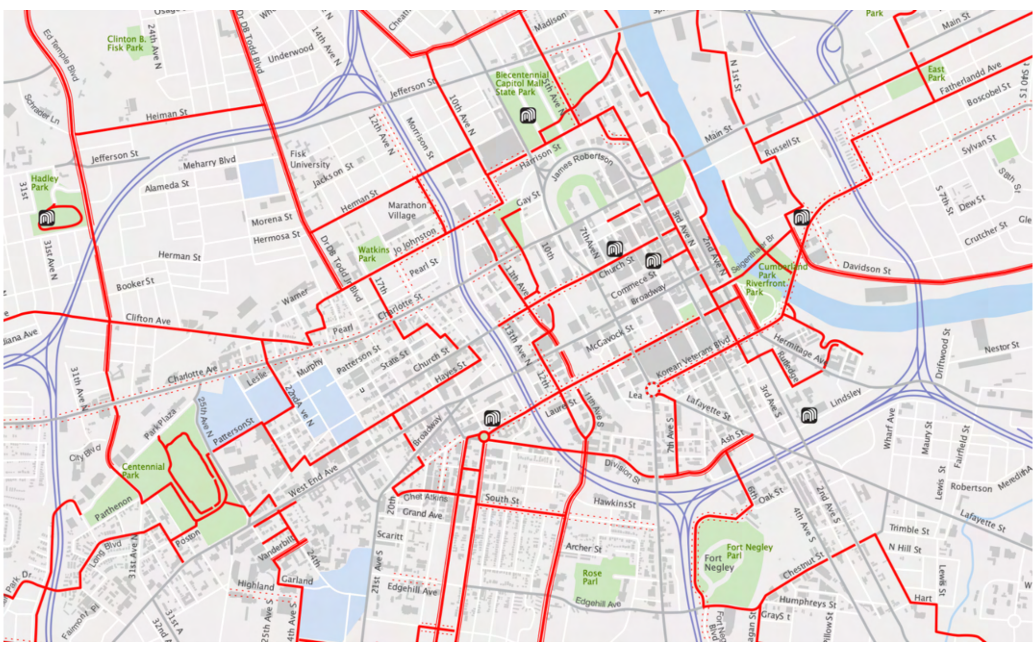


Figure 17: BM2CRF1

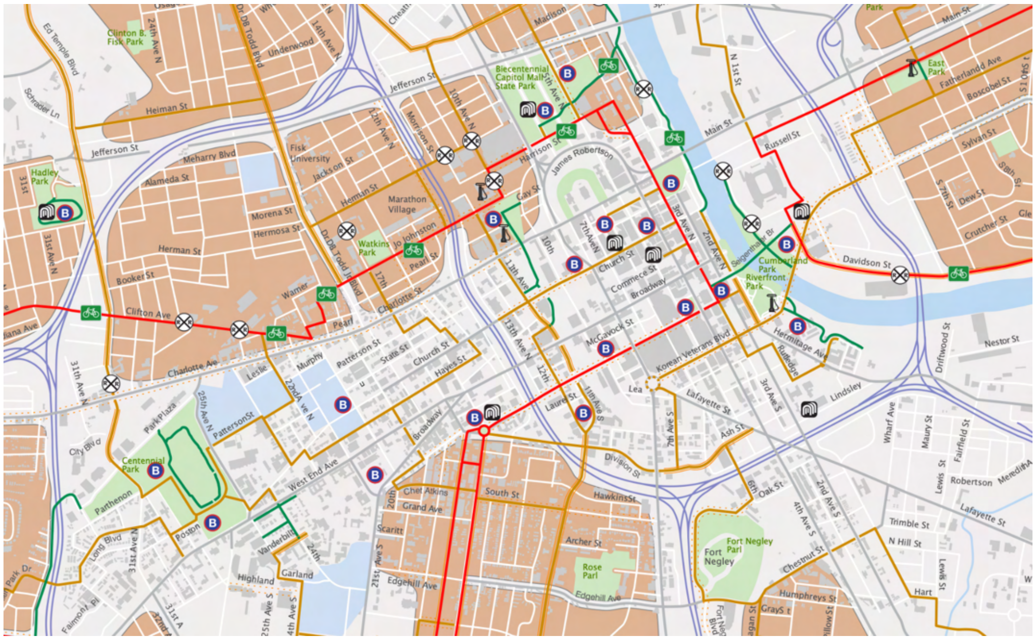



Figure 18: BM2CRF2

### 3.2.2 Task Design

Five tasks must be completed in the experiment. Tasks 1 to 4 are entirely within-subject, meaning that all the participants see the same stimuli. Between-subject design is used in Task 5. Meaning that a variable is manipulated among the participants. In this particular case, participants see BM1CRF1 or BM2CRF2. All of the tasks are related to biking. A white slide with a cross in the middle is displayed after each stimulus shown in the experiment. Before the following activity begins, the participants should gaze at that cross to ensure that their eyes are aligned neutrally. All tasks will be covered in the following sections.

#### Task 1

The first assignment involves the you-are-here (YAH) symbol, which is one of the most crucial aspects of a reference map. Those YAH maps are traditionally in situ since they represent the area in which they are located (Montello, 2010). Nowadays, YAH maps are also available on digital devices, such as cell phones with GPS (Klippel et al., 2006). Eight maps are shown in the first challenge, two for each factorial design cell. Participants must look for the symbol and then click on it when they have found it. The symbols are strewn around the map at random. But, two conditions are taken into account. First, the symbols should not be placed too close to the edge of the map. In reality, if the YAH symbol is right on the edge of a map, the extract would be made bigger. As a result, it would have been counterintuitive to the participants. Second, symbols that are in the river are moved for obvious reasons. A grid has been placed over the map for the random distribution. The x-axis ranges from 0 to 90, and the y-axis ranges from 0 to 50. Random numbers between 3 and 87 and 3 and 57 were picked to account for the edge problem. Random.org was used to generate these random values (Haahr, n.d.).

### Task 2

The second assignment requires participants to look for and count bicycle racks. When they have finished counting, they can proceed by clicking anywhere on the screen, where a scale appears. This task resembles the act of looking for a parking spot for the bicycle. The distribution of bicycle racks follows the same procedure as for the YAH symbols in task one. In total, four maps are displayed with bicycle rack locations constantly changing. The original plan was to always show seven bicycle racks per map. However, the pilot run has shown, that always displaying the same number is too conspicuous. Therefore, two times six and two times seven bicycle racks are displayed in this task. Due to the need for randomization, the bicycle racks are not distributed as they are in reality. Furthermore, depending on the distribution, the visual map complexity changes. This aspect was tested before utilizing the stimulus.

### Task 3

In task one and two, the emphasis was on symbols. The third assignment, which serves as a complement to the first two, focuses on areas and labels. This is particularly interesting since not only does the symbol density change inside the different cells of the factorial design, but so does the amount of areas. In the third task, participants have to search for a specific park. The name of this particular park appears at the top of each map. After finding the park, the participant can proceed by clicking on it. For this task there are four maps that are shown to the participant. However, this task's design is a little more difficult. The park names must also be randomized; otherwise, there may be a learning effect if the park names remain the same. As a result, the names of ten labelled parks always changed. Furthermore, variation was introduced, which means that four distinct maps were created for each cell, but only one was shown to the participant. This time-consuming method ensured that the task would be random.

**Task 4**

A starting and ending point is displayed on the map in task 4. A legend detailing the various bicycle infrastructure displayed on the map is also available. Participants must find the quickest route between the two spots. They should follow the fastest route by hovering over it with the mouse once they have found it. The routes picked were not the most important aspect of this endeavor. Rather, the eye-movements are of interest for later investigation. This task is also performed on the four different cells, thus on four maps with varying degrees of complexity. The starting and endpoint were not assigned randomly. Rather, two routes with almost equal distance and complexity were chosen. The starting and ending points for both routes were altered. In other words, point A became point B and vice versa. This has not been realized in either of the pilot runs. BM1CRF1 and BM2CRF2 had the same but introverted route, as well as BM1CRF2 and BM2CRF1.

**Task 5**

The last task is the only one designed as a between-subject experiment. Some participants are shown a BM1CRF1 map, while others are confronted with a BM2CRF2 map. This task does not include the remaining cells in the matrix (BM1CRF2 & BM2CRF1). The sample size of the experiment must be larger if four maps are shown in the between-subject design. As a result, the two images with the greatest visual complexity difference are displayed.

The objective of this task is to estimate the percentage of roadways with cycling infrastructure. Again, the legend with the bicycle infrastructure is available for the participants. When the participants are ready to enter their estimation, they can proceed by clicking the button. Then, a scale from 0 to 100% is displayed, where the estimated value can be clicked.

GIMP is used to calculate the ratio between streets with and without bicycle infrastructure. To count the pixels of bicycle streets and normal streets, all streets were shrunk to the same width. The histogram is then used to calculate the ratio.

**Dry Run**

A dry run was conducted with each participant to ensure that everything was clear to them and that they were familiar with the calibration process. The same five tasks that were just introduced must be completed in the dry run. In the dry run, participants are simply given one task per map to complete. Furthermore, the map shown in the dry run was only a small extract of a fictional map.

## 3.3 Procedure

All the sessions followed the same procedure. To ensure that all participants had the same circumstances, a protocol was created and the experiment was always led by the same person. All to ensure a consistent procedure. The protocol listed the several phases, as well as what to say, do, and an estimation of how long each step takes.

The protocol for the experiment is given in this section (see Figure 19). The entire experiment took roughly 20 minutes to complete. Half-hour intervals were set up to ensure that there was enough time for preparation and post-processing, as well as some buffer.

In order to conduct an experiment, some preparations had to be made. This included sanitizing all surfaces, printing the consent form, and turning on the computer. Then creating a new participant in Tobii Studio and opening the questionnaire in Google Chrome.

In the next phase, the participant entered the room. A do not disturb sign in front of the lab ensured that no one entered the room during the experiment.

The participant got seated to sign the consent form. To expedite the procedure, each participant received a consent form in advance. There was adequate time for the participant to read the permission form in-person if necessary (see appendix for the consent form).

In the next phase, the introduction to the study and the eye-tracking took place. The participant was informed that she/he would see different bike maps on the computer screen and was required to complete various tasks. The eye-movements made while doing these tasks were recorded. Also mentioned were the slides with the white crosses on them. These were displayed after each map to make sure the eyes remained focused on the computer's center. Task 3 was already mentioned in advance because the pilot runs revealed that hovering over the path with the mouse can be deceiving. Instead of hovering, the pilots clicked on the screen. Finally, the participant was informed that a dry run and a main run were held.

Then, the calibration was explained. The calibration for the dry run was slightly shorter than for the main run. It was crucial to make sure the participant was properly positioned in front of the screen. She/He was also told to ensure not to move too much, as the eye-tracker does not handle movement well. Overall, they must be sitting comfortably, as they may sink lower into the seat during the experiment. After explaining how the calibration works, the calibration could be started. Depending on the accuracy, the participant could proceed or had to repeat the calibration. For this experiment, a maximal acceptable data loss of 10% was chosen. This threshold was found to work well in the pilot run and other previous runs.

It was important to keep an eye on the participant throughout the dry run to ensure that everything worked properly. Following the dry run, the participant got the opportunity to ask any questions she/he may have had. They got told that the procedure of the main run is the same. The experiment began with a calibration, followed by the experiment. The tasks remained the same, but the amount of maps displayed and the number of subtasks changed.

Only the questionnaire remained after the participant completed the tasks. LimeSurvey was set up in a browser tab so that the participant could jump right to the questions.

The participant was thanked for participating in the study and given a chocolate bar. The data was backed up and the consent form was scanned and submitted once the participant had left the room. The procedure then repeated itself, beginning with the preparation phase for the next participant.

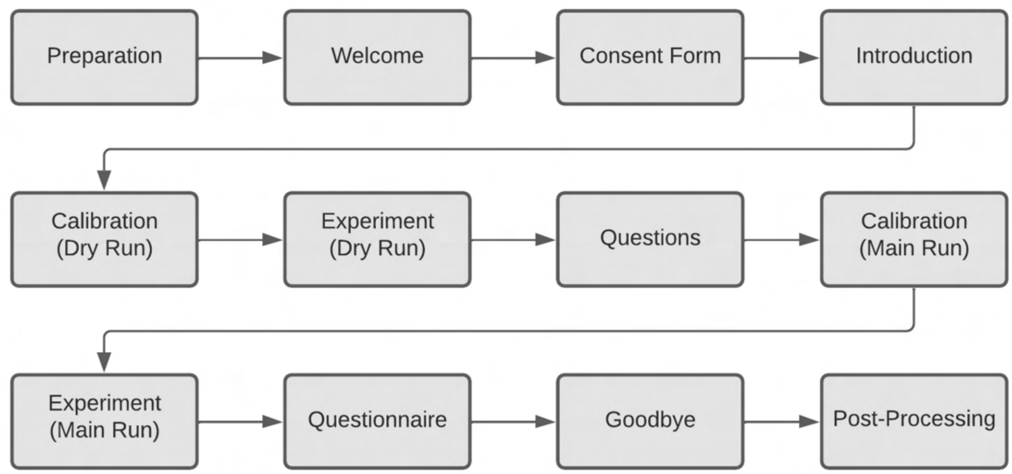



Figure 19: Procedure of the Experiment

**Setup**

The experiment took place at the University of Zurich's Irchel Campus in the Geographic Information Visualization and Analysis (GIVA) group's Eye-Movement Recording Lab (EML). This lab has no windows to maintain consistent lighting conditions, so the pupil size of the participants is constant.

The test was carried out on a Dalco PC running Windows 7 Enterprise (SP 1). This computer is powered by an Intel Core i5 760 processor (2.80 GHz, 8 MB cache, 16GB RAM). A 23-inch Estecom screen displays the stimuli. This display has a 1920 x 1080 resolution, a 5 millisecond reaction time, and a 16:9 image aspect ratio. To be able to display the stimuli on this display, the resolution must be reduced accordingly. The participant's workspace is similar to a typical office desk (see Figure 20). The setup included a wheel mouse, a regular QWERTZ keyboard, speakers, and a webcam. To ensure that the participants stay in place, the chair provided is not movable.

The binocular Tobii TX300 eye tracker was employed, which has a data rate of 300Hz and a 0.4 degree precision. The required software is Tobii Studio running on version 1.181.

The same Dalco PC is utilized for the questionnaire, using Google Chrome running a Lime Survey questionnaire.

Figure 20: Setup of the Participant's Workplace

# 4 Results

This chapter will give an overview of the study's findings. Starting with the created stimuli shown in the experiment and their corresponding complexity quantifications. Then, the participant's background and knowledge are presented. In the next step, the results of every task of the experiment will be presented.

## 4.1 Measured Map Complexity

During the design process, the quantification measurements were of importance. As the four stimuli should be a well-balanced set of complexities. The measured complexity of the stimuli will be displayed in this section. It was necessary to modify the maps for some tasks. The modified maps were also quantified to see if they have an impact on the complexity.

### 4.1.1 Stimuli

**Pixel-Based Quantification**

The calculated GMLMT, Features Congestion, Subband Entropy, and Edge Density for all four layers are shown in Table 2. It can be observed that for every measure BM1 < BM2, while CRF1 < CRF2. Furthermore, base map layers are more complex than cycling related feature layers (BM1 > CRF1, and BM2 > CRF2). The initial goal was to have similar BM1 and CRF1, as well as BM2 and CRF2 values. However, having unbalanced sets was inevitable, as base maps have a per se higher complexity. To match the base maps' complexities, a high amount of cycling related features would have to be added to the map, which would lead to an unrealistic result.

| Layer | GMLMT | Feature Congestion | Subband Entropy | Edge Density |
|---|---|---|---|---|
| BM1 | 10.7% | 3.95 | 3.86 | 3.8% |
| BM2 | 15.5% | 4.98 | 4.29 | 6.1% |
| CRF1 | 8.4% | 3.98 | 3.38 | 3.1% |
| CRF2 | 11.0% | 4.680 | 3.53 | 4.0% |

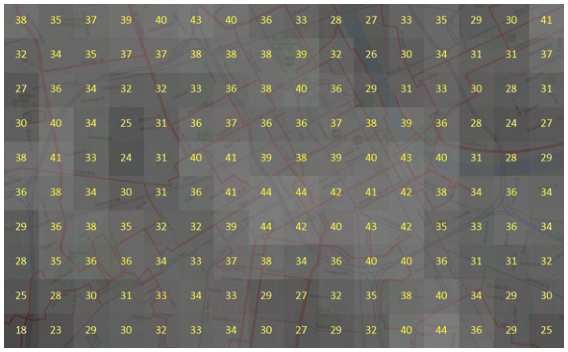
Table 2: Quantification of Layers

The quantification for all four stimuli is shown in Table 3. For every measure, BM1CRF1 > BM1CRF2 & BM2CRF1 > BM2CRF2. As a result, the overall goal of balancing was achieved. BM1CRF1 has always the smallest complexity, BM2CRF2 the highest, and the others are somewhere in between. The goal was for BM1CRF2 and BM2CRF1 to be as similar as possible. Again, this was not possible because base maps have a higher complexity than cycling related features. BM1CRF2 is more complex than BM2CRF1 according to GMLMT, Feature Congestion, and Edge Density. Only for Subband Entropy, BM1CRF2 was smaller than BM2CRF1. It is no surprise that Subband Entropy has distinct values, given this metric takes a different methodology than the other. It is the only measure that is not based on edge detection, instead, it considers the difficulty of encoding an image or a map. Another essential aspect was to make sure that the spacing between BM1CRF1 and BM1CRF2 & BM2CRF1 is more or less equal to BM1CRF2 & BM2CRF1 and BM2CRF2. This aspect was likewise fulfilled. GMLMT, for example, measures a spacing of 4.2% and 3.8%. Again, the spacing has some acceptable variation. Acceptable is not meant scientifically, rather, it is objectively acceptable after having attempted various approaches to get a balanced set.

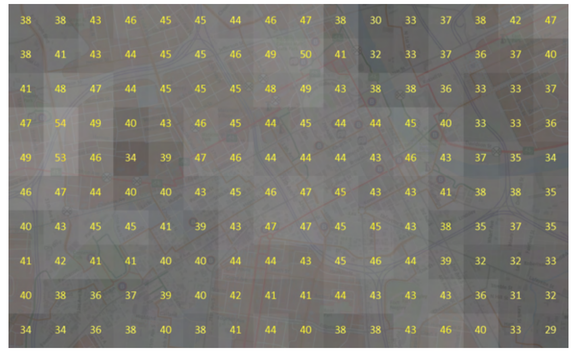
| Stimuli | GMLMT | Feature Congestion | Subband Entropy | Edge Density |
|---|---|---|---|---|
| BM1CRF1 | 16.5% | 5.71 | 4.31 | 6.7% |
| BM1CRF2 | 20.7% | 6.66 | 4.44 | 8.8% |
| BM2CRF1 | 20.1% | 6.25 | 4.55 | 7.1% |
| BM2CRF2 | 23.9% | 7.06 | 4.61 | 9.2% |

Table 3: Quantification of Stimuli

BM1CRF1 and BM2CRF2 are shown in Figure 21. As previously stated, bright cells and high values indicate a high level of complexity. It can be observed that BM2CRF2 is more complex, as it has a brighter grid than BM1CRF1. The distribution of the brightness of the grids is comparable between the two stimuli. The center of both stimuli is the most complex, while the upper-right section is the least complicated.




(a) BM1CRF1




(b) BM2CRF2

Figure 21: Outputs of the GMLMT

**Not Pixel-Based Quantification**

The distinct object-type count was also applied for the produced stimuli. This method is based on the notion that the number of various categories on a map influences complexity as well as the overall number of individual elements on the map. Figure 22 shows the distinct object-count per complexity, subdivided into labels, roads, land use, and hydrology. In all complexities, the 12 distinct objects of BM1CRF1, are visible. BM1CRF2 has 19 objects. BM2CRF1 contains 14 objects, while BM2CRF2 has the highest object-count with 21. This set is uneven from the standpoint of object-count. BM1CRF1 is still the least complex and BM2CRF2 is the most complex map. But, BM1CRF2 and BM2CRF1 have big differences. To achieve a balanced set of measures, an imbalanced object-count metric was unavoidable. The only way around this would be to create a fictitious map.


Figure 22: Distinct Object-Count for every Stimuli

### 4.1.2 Modified Stimuli

Some of the stimuli have to be altered for the tasks. The task-specific stimuli were also quantified to ensure that the different metrics did not change too significantly. Table 4 shows the measures split by task. The numbers .1, .2, and .3 refer to the stimuli used in the experiment. In Task 1, for example, eight different stimuli were shown.

Only the placement of the YAH symbol changes in Task 1, therefore the measures nearly show no change. For GMLMT the subtasks with the same stimuli always have the same value (Task1.1 = Task1.2, Task1.3 = Task1.4, and so on). There are very few changes in Feature Congestion and Subband Entropy. The stimulus set is still balanced on all measures.

Bicycle racks are displayed in Task 2. Six (in Task2.2, Task2.3) or seven (in Task2.1, Task2.4) bicycle racks change their location on the map. As a result, the GMLMT varies from the original value. In BM1CRF1, BM1CRF2, and BM2CRF1 the complexity decreased slightly, as for BM2CRF2 the complexity increased. The bicycle

| Complexity | Task | GMLMT | Feature Congestion | Subband Entropy |
|---|---|---|---|---|
| BM1CRF1 | Task1.1 | 16.6% | 5.71 | 4.29 |
| BM1CRF1 | Task1.2 | 16.6% | 5.72 | 4.28 |
| BM1CRF2 | Task1.3 | 20.7% | 6.67 | 4.43 |
| BM1CRF2 | Task1.4 | 20.7% | 6.67 | 4.433 |
| BM2CRF1 | Task1.5 | 20.1% | 6.25 | 4.52 |
| BM2CRF1 | Task1.6 | 20.1% | 6.26 | 4.52 |
| BM2CRF2 | Task1.7 | 20.7% | 7.07 | 4.61 |
| BM2CRF2 | Task1.8 | 20.7% | 7.07 | 4.61 |
| BM1CRF1 | Task2.1 | 16.0% | 5.62 | 4.29 |
| BM1CRF2 | Task2.2 | 20.6% | 6.69 | 4.46 |
| BM2CRF1 | Task2.3 | 20.0% | 6.37 | 4.56 |
| BM2CRF2 | Task2.4 | 24.6% | 7.24 | 4.59 |
| BM1CRF1 | Task3.1 | 16.2% | 5.64 | 4.34 |
| BM1CRF2 | Task3.5 | 20.3% | 6.59 | 4.47 |
| BM2CRF1 | Task3.10 | 19.7% | 6.17 | 4.58 |
| BM2CRF2 | Task3.13 | 23.5% | 7.00 | 4.65 |
| BM1CRF1 | Task4.1 | 16.7% | 5.99 | 4.24 |
| BM1CRF2 | Task4.2 | 20.8% | 6.86 | 4.42 |
| BM2CRF1 | Task4.3 | 20.3% | 6.51 | 4.48 |
| BM2CRF2 | Task4.4 | 24.0% | 7.23 | 4.59 |
| BM1CRF1 | Task5.1 | 17.8% | 6.19 | 4.24 |
| BM2CRF2 | Task5.2 | 24.9% | 7.50 | 4.56 |

Table 4: Quantification of Task Stimuli

racks were placed on the map at random. As a result, the numbers fluctuate depending on where the bicycle racks are placed. However, the stimuli are still well-balanced.

Task 3 includes 16 different stimuli, however, only four were shown to each participant. Table 4 contains four examples, one for each level of complexity. In this task, the names of the parks were assigned at random. The names vary in length and complexity, but this has little effect on the total complexity score.

Task 4 requires the participant to find the quickest path. A and B points were included in the map, which has an impact on the map's complexity. However, as only two symbols change, the difficulty has only changed marginally.

In Task 5, the stimuli shown to the participant contains the task description. As a result, the measurements for task 5 are slightly higher.

## 4.2 Participants' Background

Initially, 44 people signed up to take part in the eye-tracking experiment. While attempting to calibrate the system in four experiments, issues arose. Tobii Pro Lab was unable to detect the eyes. Only now and then do the eyes appear for a short amount of time. The project had to be stopped after several attempts. No data were obtained for these four participants, hence no data is included in the results section. After completing 35 successful experiments, the eye-tracking device failed, and no more experiments were possible. As a result, this study includes data from 35 people.

### 4.2.1 Socio-Economic Background

The 35 participants had an average age of 25 years, with a standard deviation of 2.3 years. The age range was 21 to 34 years old. 11 participants were female and 24 were male. None of these participants indicated color-blindness.

The majority of the participants are geography students (67%), as they were individually invited to participate in the experiment. Only a small percentage of the participants had other backgrounds (8% teaching, 25% other).

### 4.2.2 Familiarity with Maps and Bicycles Usage

To find out the prerequisite the participants had with maps and cycling, they were asked about their familiarity in the questionnaire. Figure 23 - 25 show the plotted likert scale, for the familiarity questions. Figure 23 and 24 have the same levels (maps, bike maps, and bicycles).

First, the participants were asked about their familiarity with maps in general, bike maps, and cycling, then they were asked how often they use them. It is no surprise they are almost congruent. The participants have a high level of familiarity with maps and frequently utilize them. The same appears for cycling. The participants are familiar with cycling and also use them often. Compared to the inhabitants of Portland, it can be said that the familiarity of the present sample with cycling is higher. The categorization of the inhabitants by Dill and McNeil (2016) was created with different criteria. However, the sample might use bicycles more often than the average population. This could be due to the relatively young age of the participants. The intersection between maps and cycling are bike maps. However, it is remarkable that this criterion has the lowest familiarity and utilization.

The familiarity with various types of maps can be seen in Figure 25. The familiarity of the sample with Google Maps is by far the highest. Everyone is at least somewhat familiar, while most are even moderately or extremely familiar. The familiarity with paper maps is lower than with Google Maps, but still, there are not many people being not familiar with paper maps. OpenStreetMap has the third-highest familiarity. This might be explained by the participants' background, as many of them work or study in the field of geography. The sample is least familiar with Bing Maps and Apple Maps.

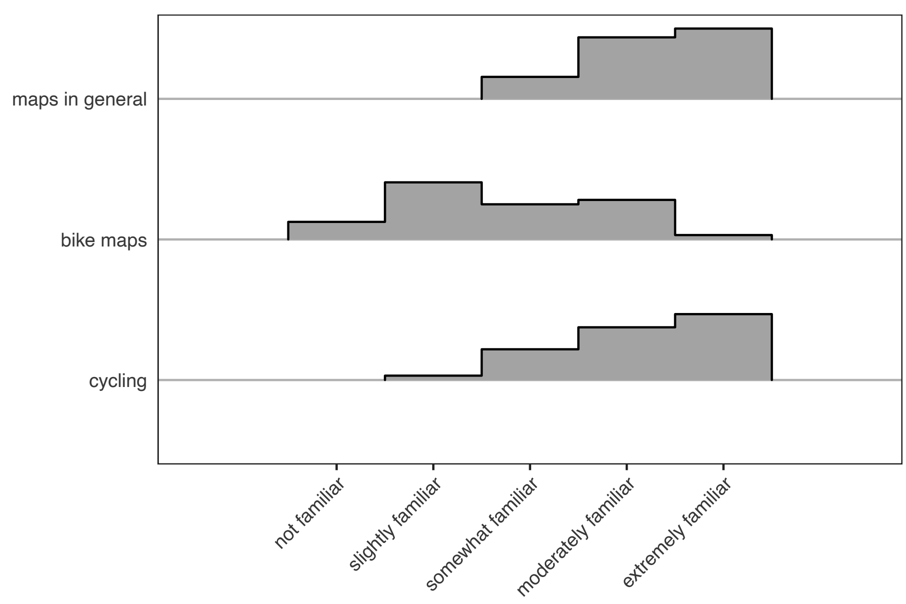



Figure 23: Distribution of participants' familiarity with maps in general, bike maps, and cycling

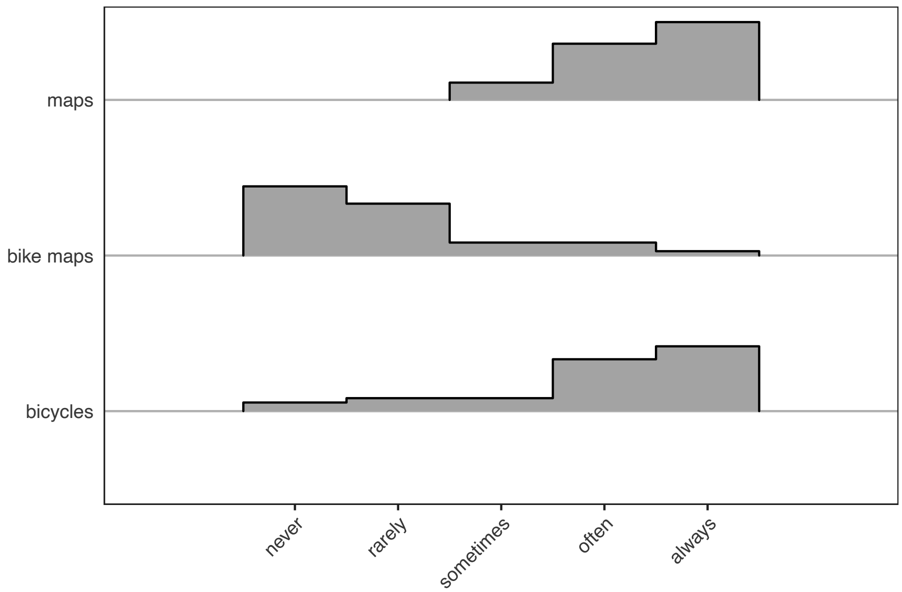



Figure 24: Distribution of how often participants use maps, bike maps, and bicycles

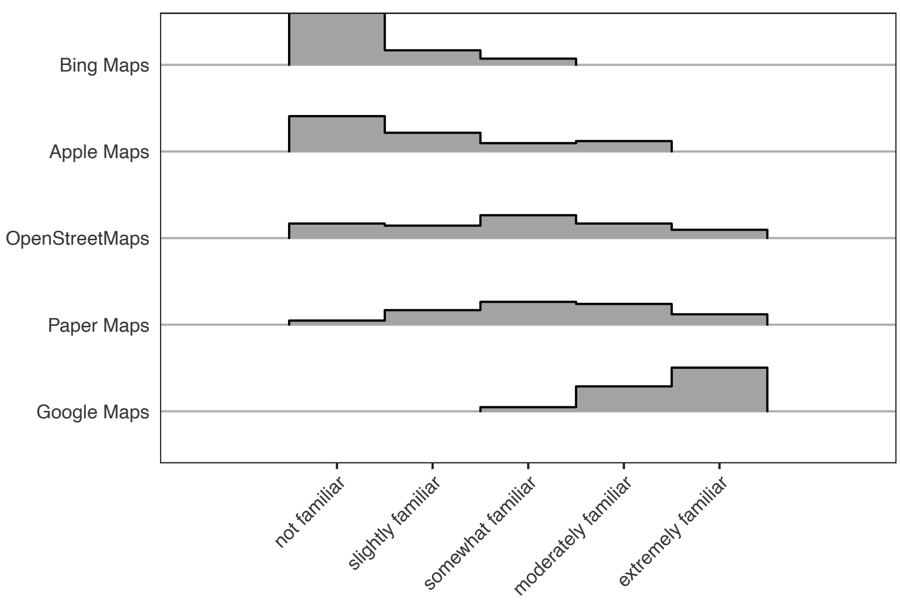



Figure 25: Distribution of participants' familiarity with different sorts maps

## 4.3 Eye-Tracking

In this section the results of the eye-tracking experiment will be discussed. Before, the essential eye-tracking features and needed statistics will be shortly introduced.

#### Replay Recordings

Tobii Pro Lab allows you to playback any recording. The stimuli, as well as the place where the eyes are focused, can be re-watched in real-time. The replay speed can be changed, and a timeline zoom range can be used to browse specific points of interest (Tobi Pro, 2021). This tool helps comprehend what happened during the experiment. For instance, if the experiment's executor wishes to know if particular AOIs were missed or if the AOI box was just too small. This tool came in handy for determining the size of the AOIs.

#### Heat Maps

Heat maps are usually generated on top of a stimulus. Heat maps do aggregate the duration or numbers of all fixations at the same location. Heat maps are visualized using several colors, the most frequent of which are red for high or extended fixations and green for low fixations, with varying degrees in between. Heat maps can be made for a single person, but they can also be used to visually summarize the eye movements of multiple participants intuitively (Tobi Pro, 2021).

#### Areas of Interest

Based on a region of interest in the stimuli, AOIs can provide statistical information for additional studies. Drawing the AOI is a crucial step. The participants may miss the AOIs if they are drawn too small. If they are drawn too large, they may be struck without the participant even realizing it. Furthermore, there is no set size for AOIs because it is dependent on the size of the object of interest. Before conducting the experiment the size of the AOIs was modified several times to find the appropriate size. For example, to complete the first task of the experiment, each participant must be able to locate a symbol, which necessitates a hit of the AOI. When drawing the AOIs, it was assured that when the items of interest are the same size, so are the AOIs.

### Statistics

The exported metrics from Tobii Pro Lab are loaded into R for data processing and statistical analysis. For data management, statistical tests, and plot creation and styling, the packages agricolae, ggrdiges psych, tidyverse, and yarr are utilized.

The goal is to utilize analysis of variance to analyze the variance in the gathered data. Within-subject (repeated measures) and between-subjects ANOVA tests are available (separate groups). Before conducting ANOVA, tests for normal distribution and homoscedasticity have to be conducted.

To check for normal distribution, the Shapiro-Wilk test is utilized (Royston, 1995 & Martin, 2008). Data is additionally examined on a log-normal distribution as time is not normally distributed. Another criterion for performing ANOVA is that the groups' variances are equal. To verify the data for homoscedasticity, Levene's test is executed. If ANOVA reveals significant differences, Tukey's honestly significant difference (HSD) test is used to determine where differences within the groups exist (Hervé and Williams, 2010). When the data is not (log-)normally distributed, a non-parametric test was used. The Mann-Whitney U test, also known as the Wilcoxon rank sum test, is used for this purpose (Weiner and Craighead, 2009). The data used in this test is not log-transformed.

### 4.3.1 Task 1: Locate the YAH Symbol

Task 1 required participants to locate a YAH symbol on the map. All in all, eight maps were shown to the participants. For Task 1 the time to first fixation (TTFF) per subtask and per stimuli are measured.

### Time to First Fixation per Subtask

The TTFF for each subtask is shown in Figure 26. It is possible to observe differences not only between but also within the stimuli. The biggest difference within the stimuli can be seen in tasks 1.3 and 1.4. Possible reasons will be discussed in the discussion chapter. Furthermore, it is striking that the stimuli BM2CRF1 (Task 1.7 and Task 1.8) have the biggest boxes and whiskers with the overall highest time values.
In the next step, the gathered data will be statistically analyzed. As the intention was to perform an analysis of variance, the data must be normally distributed and fulfill homoscedasticity. The data was neither normally distributed nor log-normally distributed (see appendix for test results). As the prerequisite of normal was not

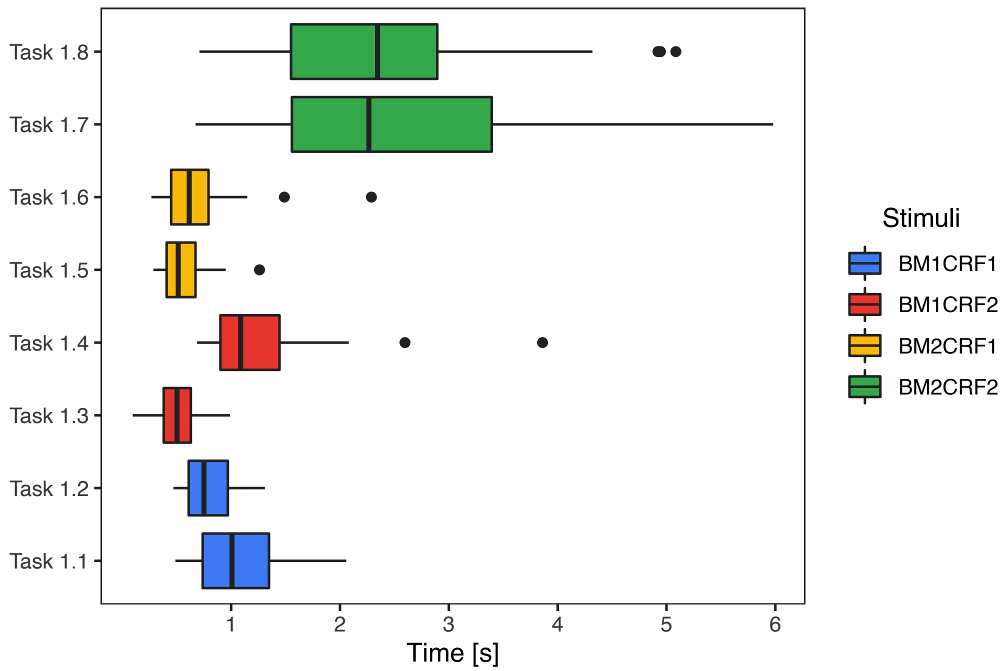



Figure 26: Time to First Fixation per Subtask

met, the Mann-Whitney U test was used. Figure 27 shows a matrix for the pairwise test. Many paired tests are significant. Stimuli BM2CRF1 and BM2CRF2 are not significantly different within them. Furthermore, Task 1.1 and Task 1.4 are also not significant.

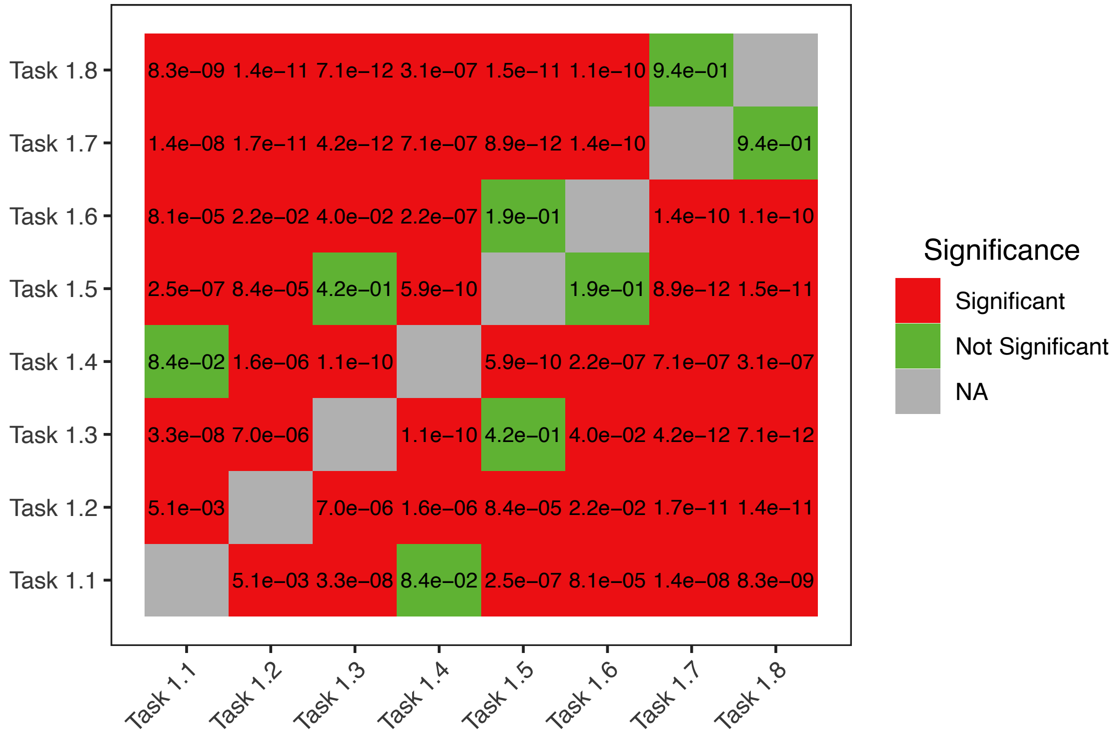



Figure 27: Matrix for Mann-Whitney U Test per Subtask

### Time to First Fixation per Stimuli

The data is aggregated to check between the different stimuli after having checked the individual tasks. Figure 28 shows the TTTFF per stimuli. In the aggregated boxplot, BM1CRF1 and BM1CRF2 have fairly similar boxes and medians, however, BM2CRF1 has lower time values and BM2CRF2 has higher values.

Again, the requirement of normal-distribution for ANOVA was not met, and thus the Mann-Whitney U test was performed. The matrix for the paired test is shown in Figure 29. The boxplot shows that BM1CRF1 and BM1CRF2 are quite similar and that they are the only two stimuli that are not significant. It is crucial to note, however, that BM2CRF1 is significantly smaller than BM1CRF1, which was unexpected.

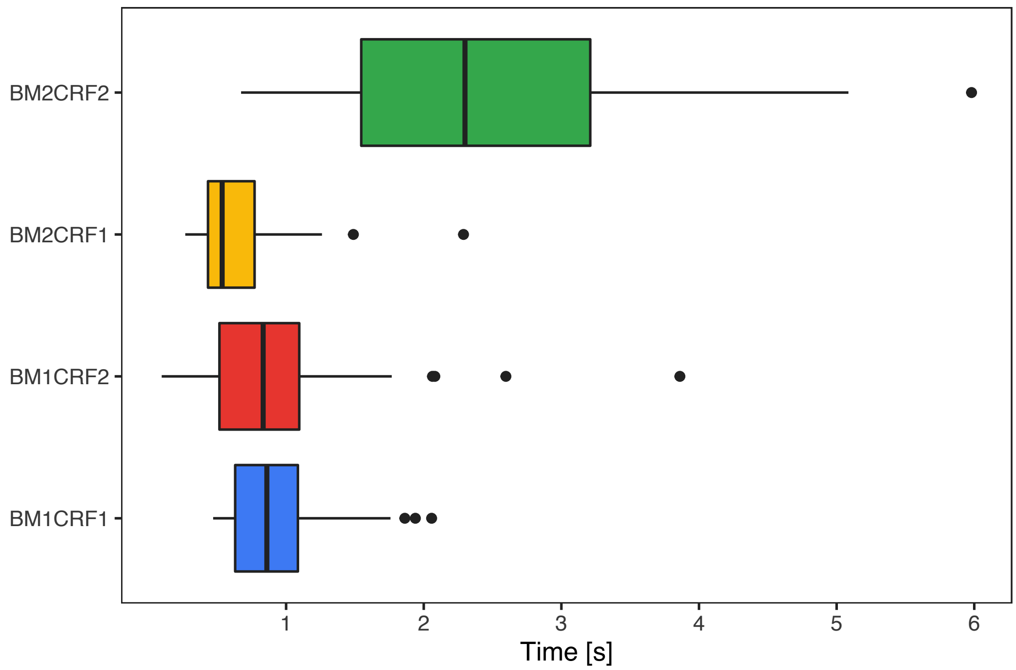



Figure 28: Time to First Fixation per Stimuli

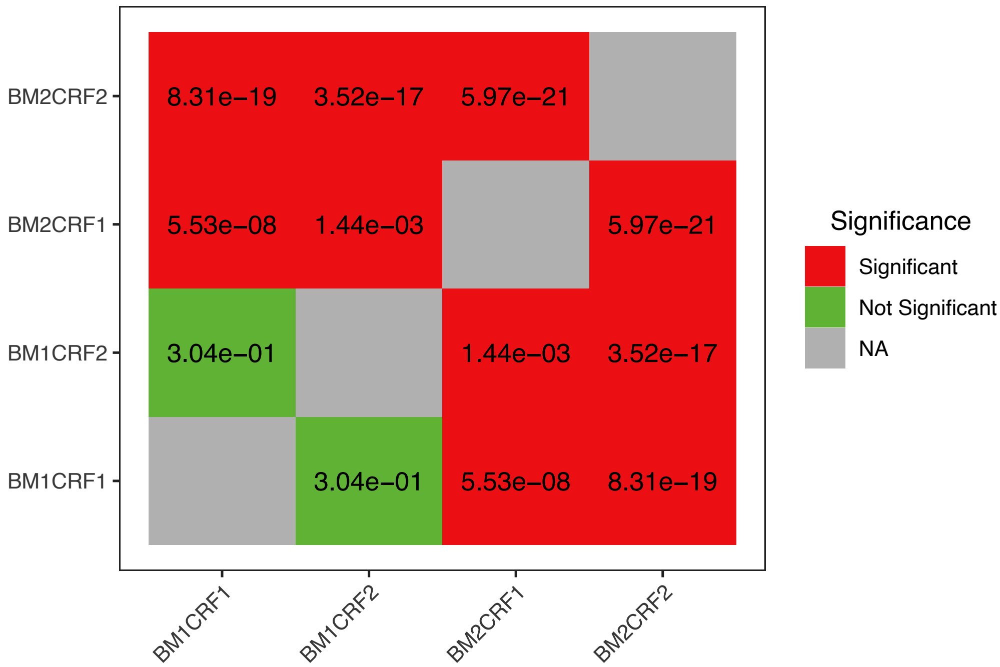



Figure 29: Matrix for Mann-Whitney U Test per Stimuli

#### 4.3.2 Task 2: Count Bicycle Racks

Task 2 required participants to count all bicycle racks on the map. There were four maps in total, one for each stimulus. Seven bicycle racks are depicted in BM1CRF1 and BM2CRF2, whereas six bicycle racks are depicted in BM2CRF1 and BM1CRF2. Average time to new fixation, duration of task completion, fixation count, and accuracy will be statistically examined in the following sections.

##### Average Time to New Fixation

The average time to first fixation is shown in Figure 30. For example, the first bicycle rack in BM1CRF1 gets fixated the earliest on average, whereas participants took the longest to fixate on the first bicycle rack in stimuli BM1CRF1. There is no data for bicycle rack seven for stimuli BM1CRF2 and BM2CRF1, as only six were displayed. This is why the lines stop at bicycle rack number six. All in all, it can be stated that for BM1CRF1 it took participants the shortest to fixate on the bicycle racks, followed by BM2CRF1, BM2CRF2, and BM1CRF2. However, for the fixation on the fifth and sixth bicycle racks, it took the participants a little less time for BM2CRF1 than BM1CRF1.

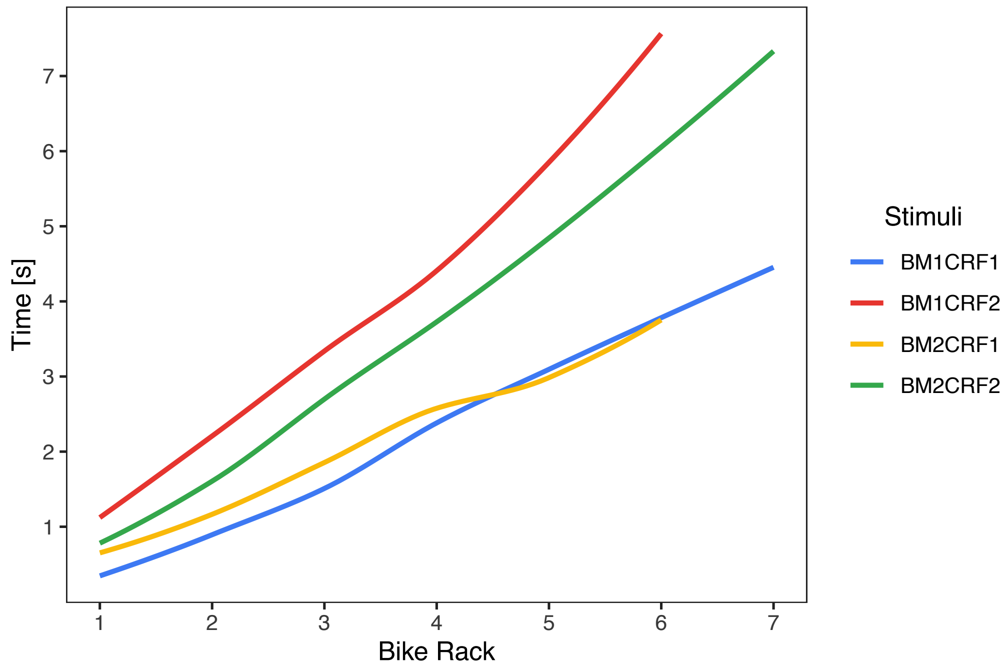



Figure 30: Average Time to First Fixation per Bicycle Rack

Figure 31 shows the average time needed to a new fixation on a bicycle rack. As a result, the previously described trend of Figure 30 remains unchanged. For BM1CRF1 participants took the least amount of time to find a new bicycle rack, while for BM1CRF2 they took the most time to find a new one.

An analysis of variance was again the goal. The data is not normally distributed, although it does follow a log-normal pattern (see appendix for Shapiro-Wilk test results). Therefore, Levene's Test for Equality of Variances can be performed. The test reveals a p-value of 0.77, indicating that the homogeneity of variance assumption is met. ANOVA reveals that there are significant differences between some stimuli (p-value = <2e-16). The Tukey HSD shows that BM1CRF1 and BM2CRF1 are not significantly different, just like BM1CRF2 and BM2CRF2. All of the other combinations are significant. The appendix contains all values and results (for Shapiro-Wilk, Levene's Test, ANOVA, and Tukey HSD).

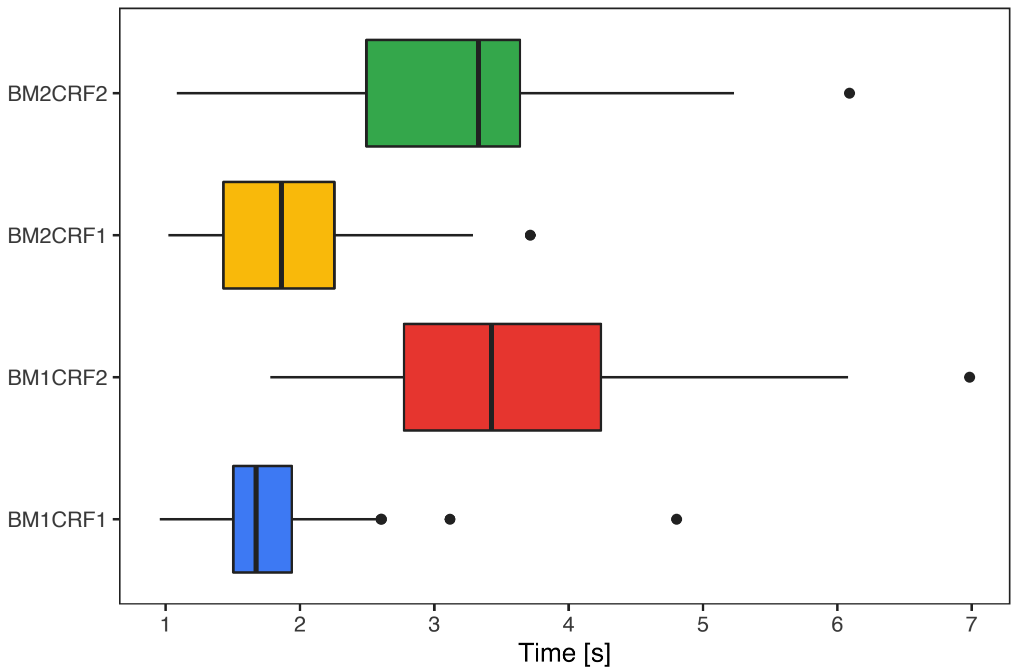



Figure 31: Average Time to New Fixation

### Duration of Task Completion

After examining the time to first fixation for task 2, the duration of task completion is investigated. Figure 32 shows the boxplot for the time needed for task completion. The plot seems similar to Figure 31. This seems understandable, as people finish tasks faster when they are faster at spotting the symbols.

As both Shapiro-Wilk tests reveal a lack of normal distribution, the Mann-Whitney U test is used. The matrix for this test can be found in the appendix. BM1CRF1 and BM2CRF1 were significantly different from BM1CRF2 and BM2CRF2. BM1CRF1 and BM2CRF1, like BM1CRF2 and BM2CRF2, were not significant.

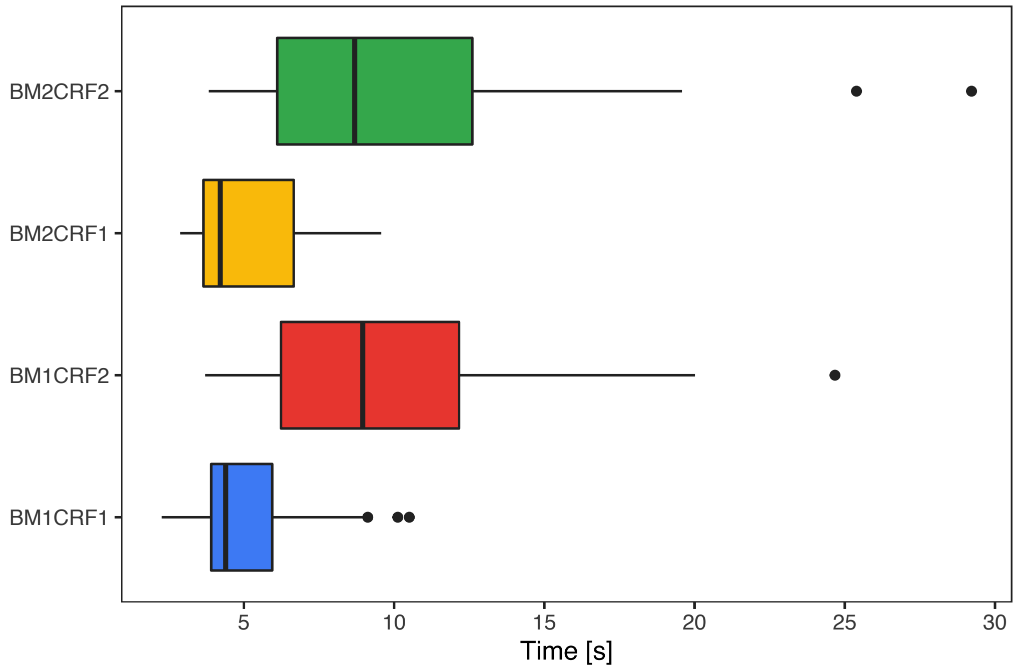



Figure 32: Time Needed for Task Completion

### Fixation Count

Next, the number of count fixation will be analyzed. Meaning, how often do the participants look on average on a bicycle rack, before moving on. Figure 33 shows the mean fixation count for the four stimuli.

The tests for normal and log-normal distributions are in the appendix, as neither test show normal distribution, the Mann-Whitney U test is applied. The results of the test were the same as the duration for task completion (see appendix for the matrix).

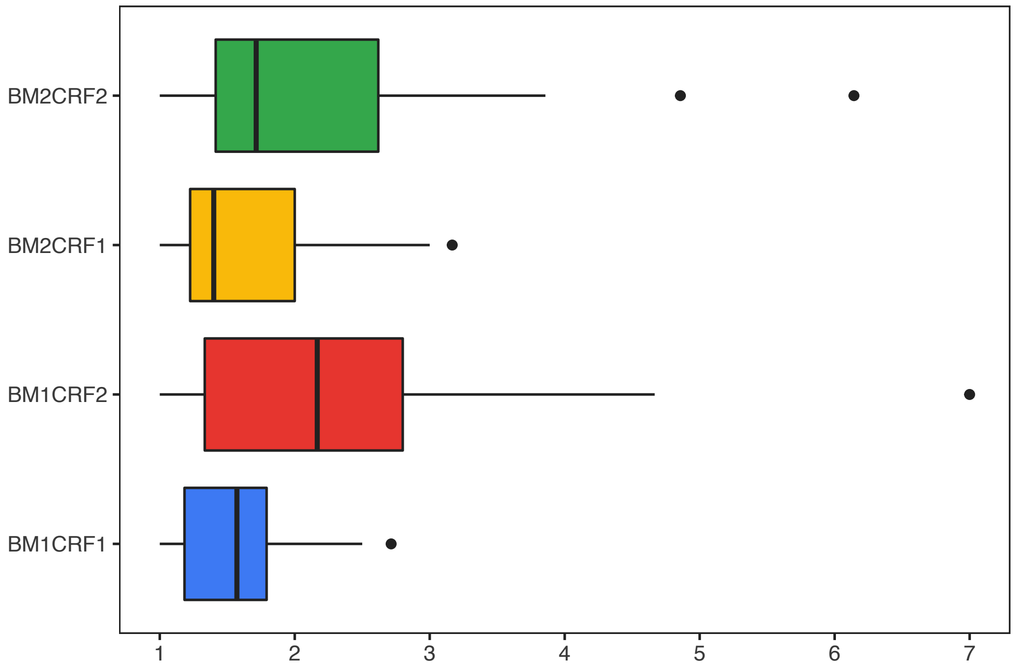



Figure 33: Fixation Count

**Accuracy**

Initially, the plan for task 2 was to set a time limit of 15 seconds to complete this assignment to provide some pressure. Then, the idea altered to not impose a time limit because there is none in reality. Rather, all participants should be able to find all bicycle racks. Nonetheless, not all participants in the experiment were able to identify all bicycle racks, necessitating this section on accuracy.

Figure 34 shows the distribution of counted objects per stimuli. The correct answer for BM1CRF1 was seven bicycle racks, while the correct answers for BM1CRF2 and BM2CRF1 were six bicycle racks. For every stimulus, the majority counted the objects correctly. For BM1CRF1 all the participants counted gave the correct answer (100%). The second greatest accuracy has BM2CRF1 (94%), followed by BM1CRF2 (91%) and the lowest accuracy of counted bicycle racks has BM2CRF2 (74%).

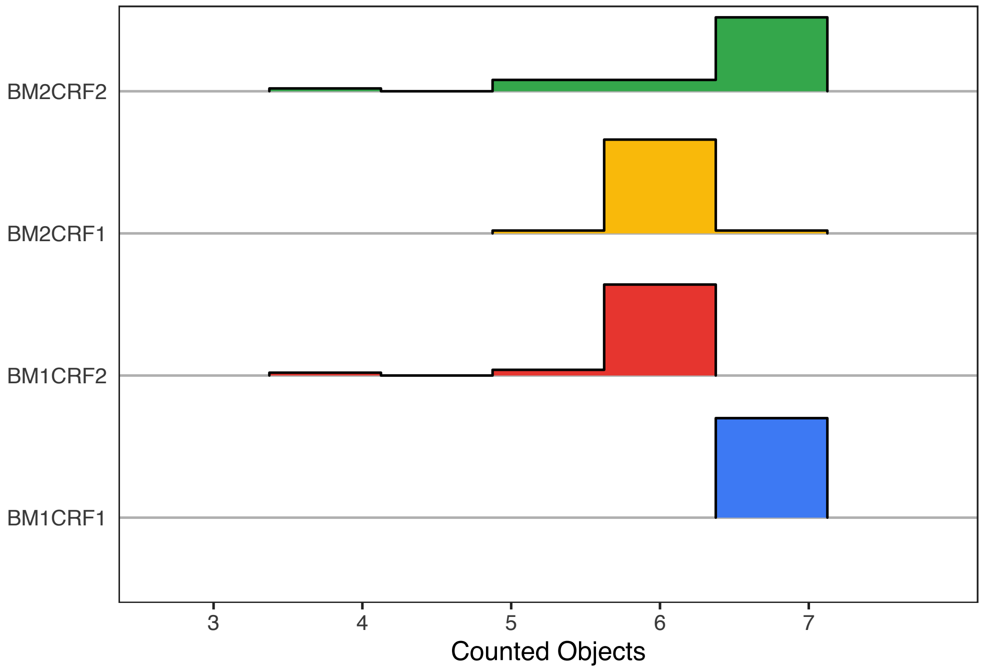



Figure 34: Counted Objects per Stimuli

### 4.3.3 Task 3: Locate a Park

The goal of task 3 was to locate a park marked on the map. In Task 3 the aim was to search for a park represented on the map. For randomization the names of the parks changed for every subtask, however, the location of the parks did not change. A learning effect can not be excluded. The outcomes of this task should be handled with caution. The amount of parks that have been focused on in Task 3 is dependent on when the searched park was detected. In Figure 35 a gaze plot for Task 3 can be seen. The park searched park was detected after looking at two other parks. The average time to new fixation and duration of task completion will be examined in the following sections.

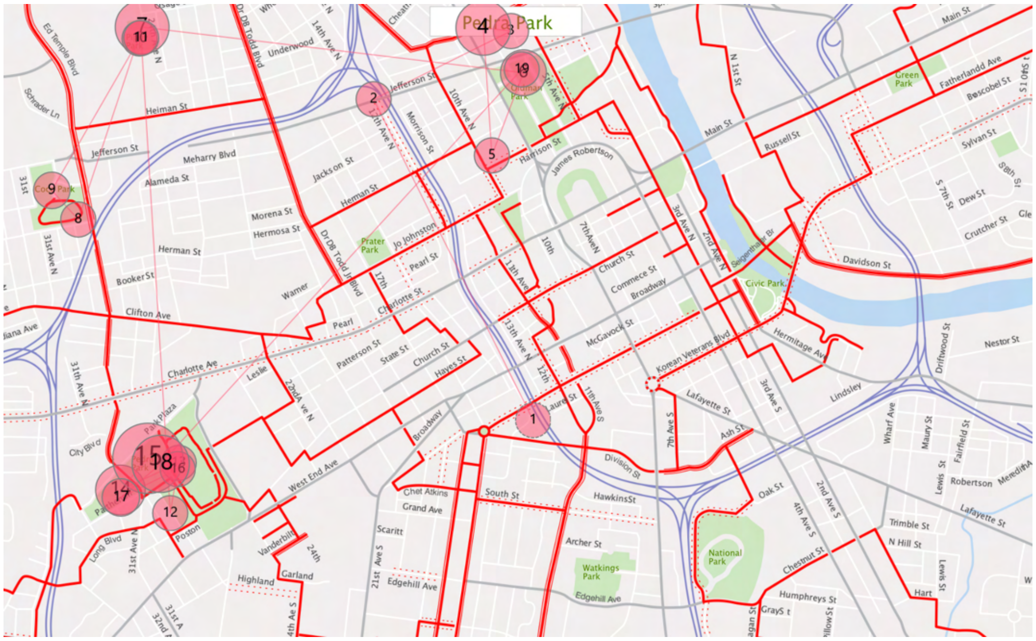



Figure 35: Gaze Plot Example

**Average Time to New Fixation**

As previously stated, the participants could complete the task without looking at every park. Figure 36 shows the average time to fixation for the first, second, third, and subsequent parks. Only five parks are shown on the x-axis because higher values have fewer samples. There is no clear result in the graph. The average time appears to be the longest for BM1CRF1. For the other stimuli, it is difficult to see a pattern, as they intersect before four parks.

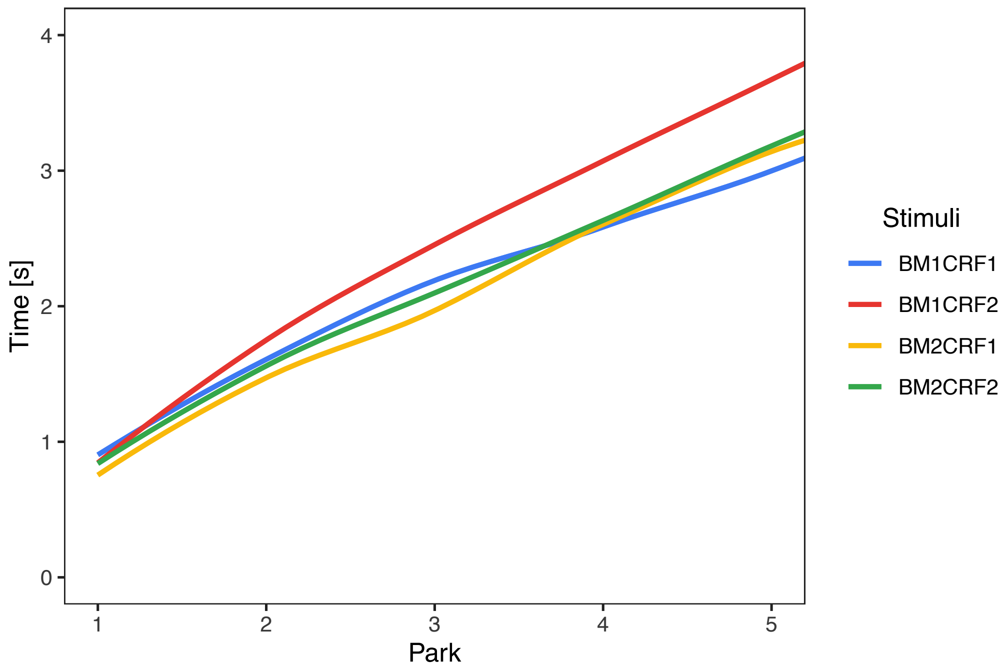



Figure 36: Average Time to First Fixation

In Figure 37 the average time to new fixation is shown. Again, for BM1CRF2 participants had the longest to focus on new objects. This is the same observation as can be drawn from Figure 36. It is important to keep in mind that with rising park numbers, fewer observations are taken. As a result, parks one, two, and three have more entries in Figure 37, than parks six, seven, eight, and nine.

Shapiro-Wilk reveals a normal distribution for all averaged time to new fixation. Levene's test has a p-value of 0.313, allowing an ANOVA to be performed. The analysis of variance depicts no significant differences within the four stimuli.

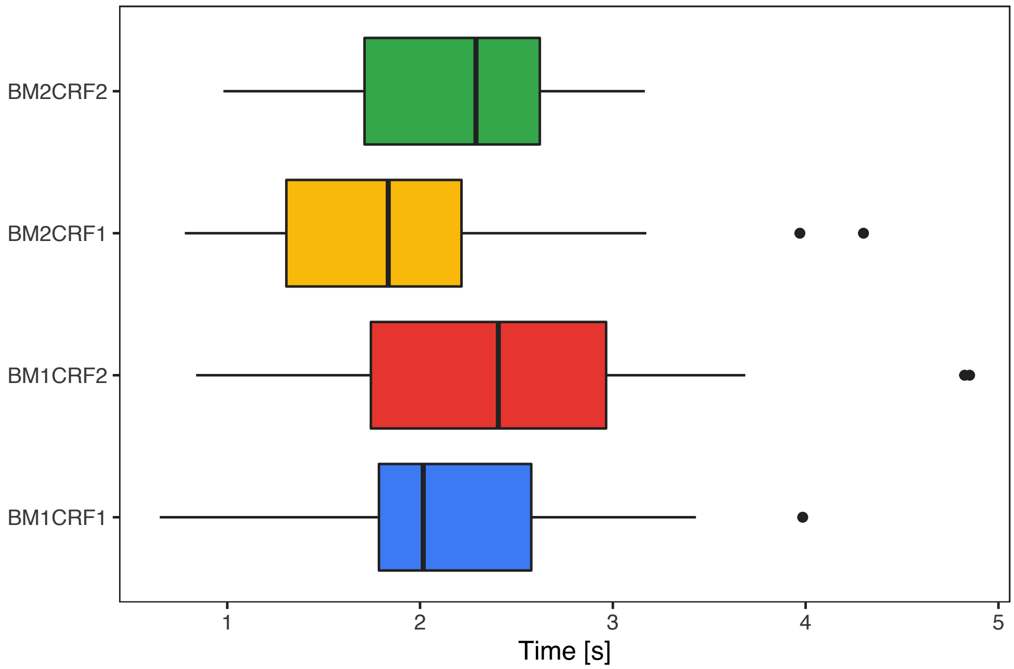



Figure 37: Average Time to New Fixation

### Duration of Task Completion

Figure 38 shows the time needed for the completion of the task. It can be observed that all stimuli have similar values. This observation is also highlighted when applying statistics. The Mann-Whitney U test was used since the data was not (log-)normally distributed. The test reveals that there are no significant differences in the time needed for task completion. The appendix contains all the tests that were undertaken.

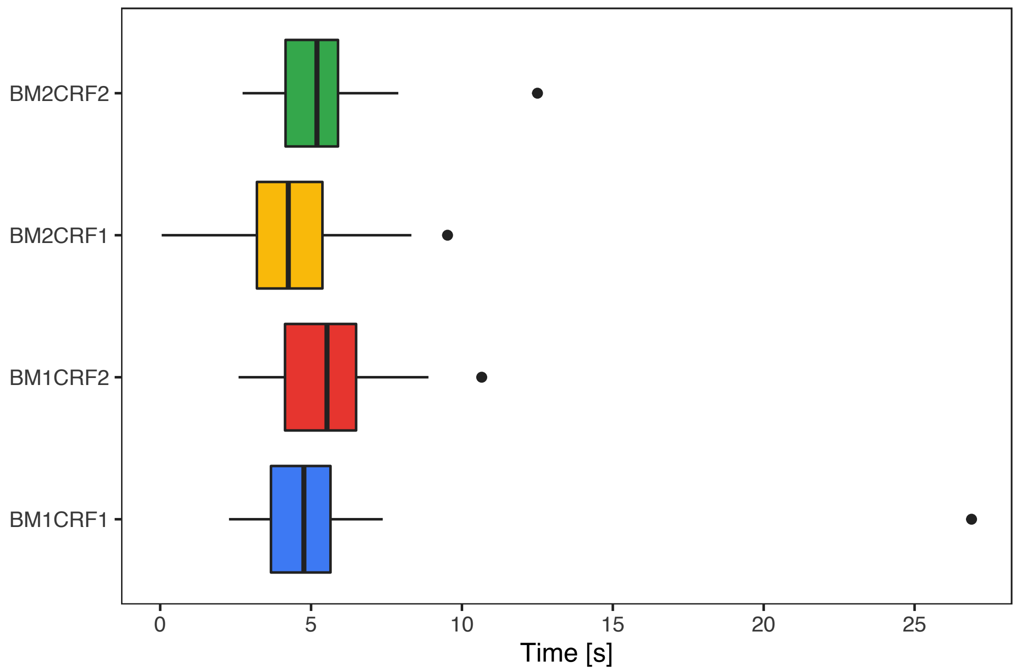



Figure 38: Time Needed for Task Completion

### 4.3.4 Task 4: Search for a Route

In task 4, participants must find the quickest route between points A and B. For stimuli BM1CRF1 and BM2CRF2 A and B had the same position but were inverted. BM1CRF2 and BM2CRF1 were treated in the same way. The duration of task completion and the time to the first fixation will be analyzed and shown in the following heat maps.

#### Heat Maps

Figure 39 and 40 show the absolute count heat maps for Task 4. In all stimuli, the legend was fixated often. The frequency with which the tale was fixed will be compared later in this chapter. Some differences are illustrated in Figure 39. In stimuli BM1CRF1 more participants looked at a route that followed the euclidian shortest path, whereas in BM2CRF2 more people decided on a route following a street in the east. In Figure 40 stimuli BM2CRF1, more participants focused on a route in the south, whereas in BM1CRF2, a section of the most fixated route was further north.

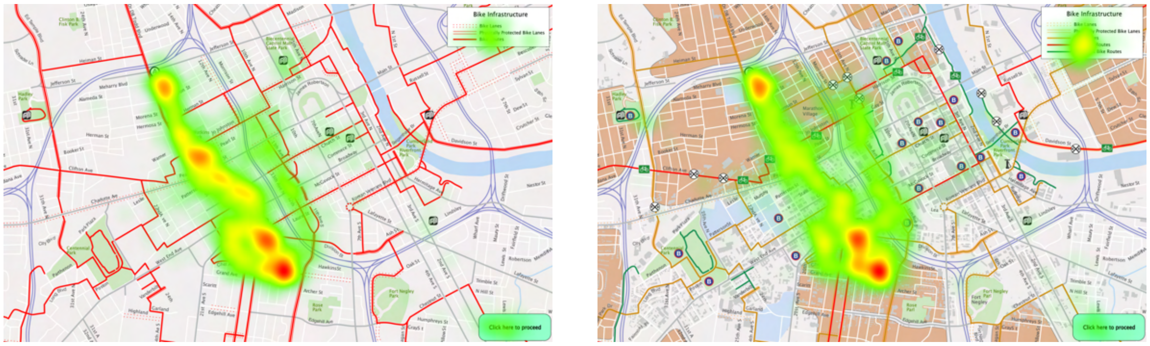


Figure 39: Heat Maps for BM1CRF1 (left) and BM2CRF2 (right)

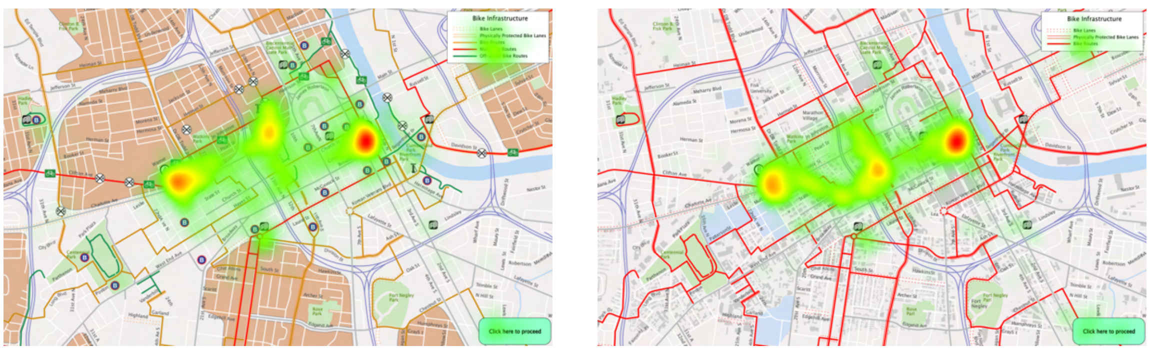


Figure 40: Heat Maps for BM1CRF2 (left) and BM2CRF1 (right)

**Time to First Fixation**

Figure 41 shows the time to first fixation per symbol. It is important to mention, that it is random on which symbol the participants fixate initially on. Therefore, looking first at point B was interpreted as a fixation on point A and vice versa. This procedure ensured that symbol A has a lower value than symbol B. It can be said that the time to the first fixation on point A is quite similar within the stimuli. It seems that like in task 1 BM1CRF1 and BM2CRF1, the maps with a low amount of symbols, have lower values than BM1CRF2 and BM2CRF2. The same appears to be true for symbol B. This was also statistically verified.

First fixations on symbol A and B show no (log-)normal distribution. Therefore, the Mann-Whitney U test is applied to both datasets. For the start symbol, no significance was found. For the second symbol, significance could be found. Table 5 shows the calculated values for the Mann-Whitney U test.

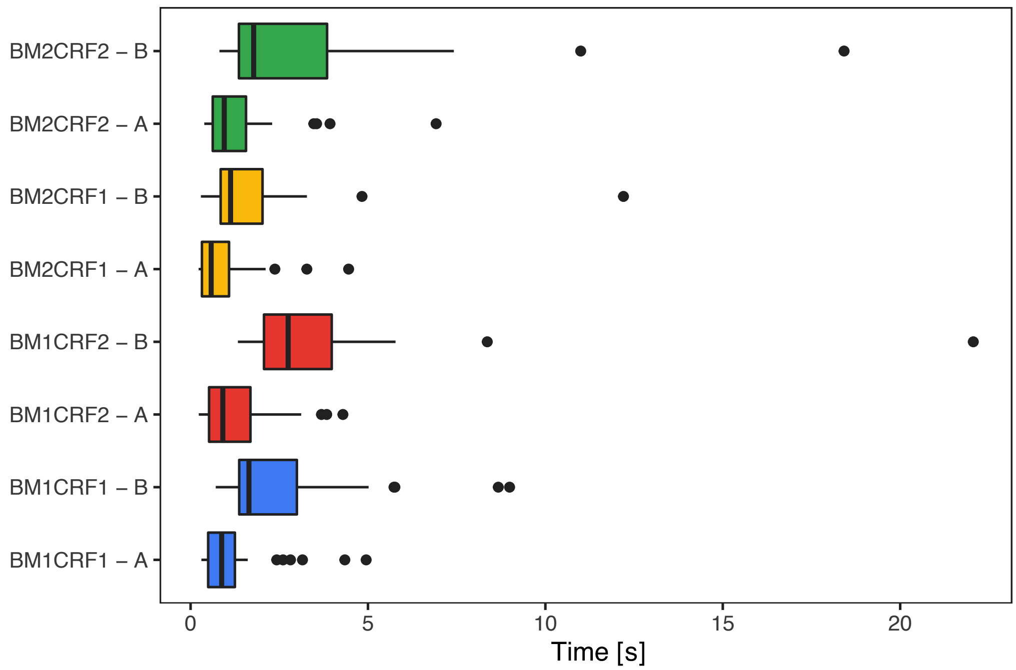



Figure 41: Time to First Fixation per Symbol

| | BM1CRF1 | BM1CRF2 | BM2CRF1 | BM2CRF2 |
|---|---|---|---|---|
| BM1CRF1 | - | - | - | - |
| BM1CRF2 | 0.0087 | - | - | - |
| BM2CRF1 | 0.0018 | 2e-07 | - | - |
| BM2CRF2 | 0.8549 | 0.0251 | 0.0006 | - |

Table 5: Mann-Whitney U Test for TTFF for the Second Symbol

### Duration of Task Completion

As the final analysis step for task 4, the duration for task completion was investigated. Figure 42 shows the task duration per stimuli. It was expected that the participants would take the least amount of time to complete BM1CRF1, but instead, BM2CRF1 took them the least amount of time followed by BM1CRF2.

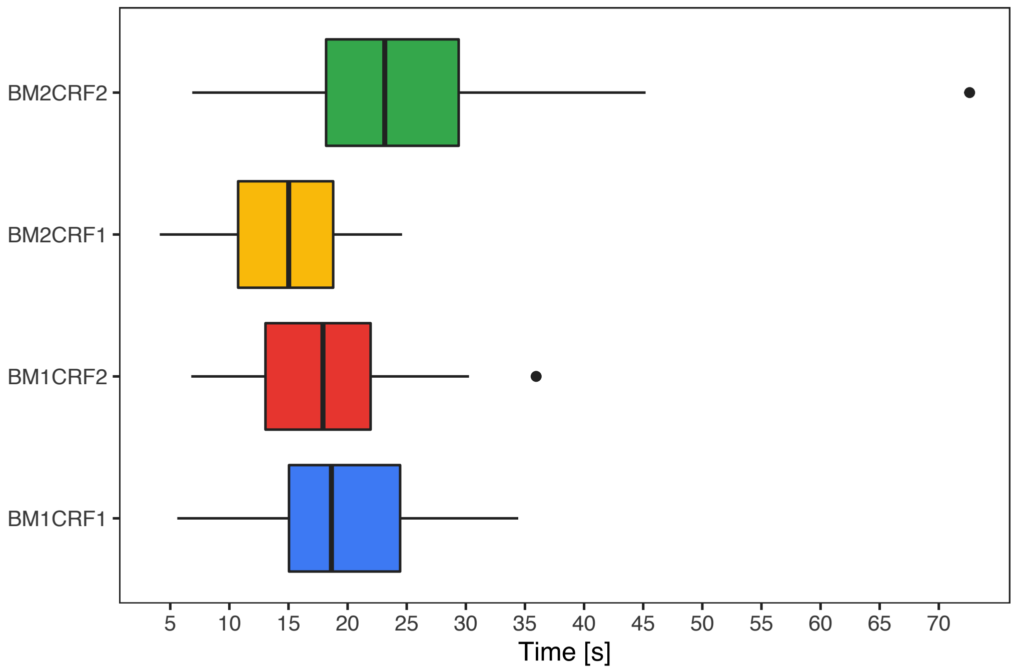



Figure 42: Duration for Task Completion

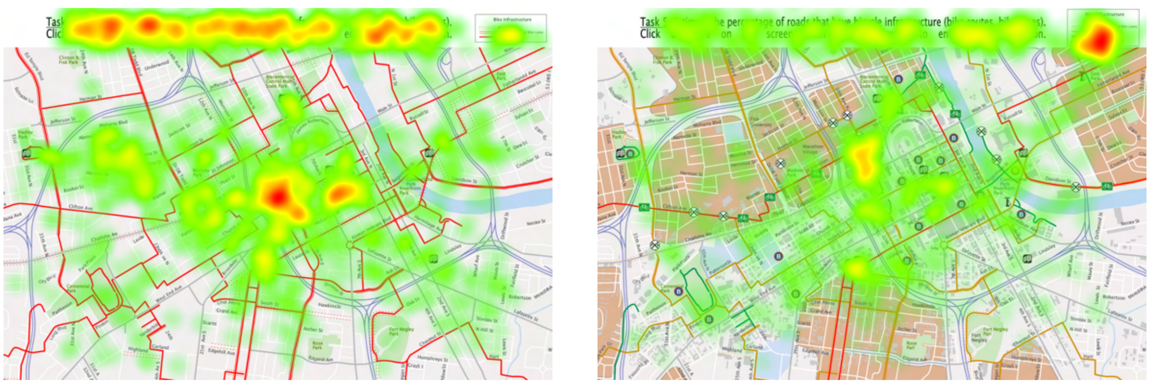
Table 6 shows the Mann-Whitney U test, as the obtained data is not (log-) normally distributed. BM1CRF1 is significantly different from BM2CRF1 and BM2CRF2. BM2CRF2 differs from the other three stimuli.

| | BM1CRF1 | BM1CRF2 | BM2CRF1 | BM2CRF2 |
|---|---|---|---|---|
| BM1CRF1 | - | - | - | - |
| BM1CRF2 | 0.530 | - | - | - |
| BM2CRF1 | 0.0258 | 0.0622 | - | - |
| BM2CRF2 | 0.0258 | 0.0056 | 8.8e-06 | - |

Table 6: Mann-Whitney U Test for Duration of Task Completion

### 4.3.5 Task 5: Estimate Bicycle Infrastructure

Task 5 was designed as a between-subject design. Half of the sample had to assess BM1CRF1, while the other half were asked to give an estimation on BM2CRF2. The stimulus shown to the individual was randomized to achieve this. Due to technical difficulties, always the identical stimulus was shown to the participants. Luckily, this issue was discovered after conducting the experiment with 19 participants. After that, only the second stimulus was shown to the participants. In total, 19 participants have seen BM1CRF1 and 16 looked at BM2CRF2. Figure 43 shows the aggregated heat maps for absolute fixation count on both stimuli. The task description was on the top of both stimuli, hence this location had a high fixation count. In the following, the duration of task completion and accuracy of the estimations will be investigated.



(a) BM2CRF2 (b) BM2CRF2

Figure 43: Heat Maps of Fixation Count

#### Duration of Task Completion

Figure 44 displays the duration for the task completion, meaning in this case, how long did the participants look at the shown map, before entering their estimation. It is visible that for BM1CRF1 the participants needed a shorter amount of time. This perception can be validated after doing statistical tests, as they differ significantly (see appendix for results).

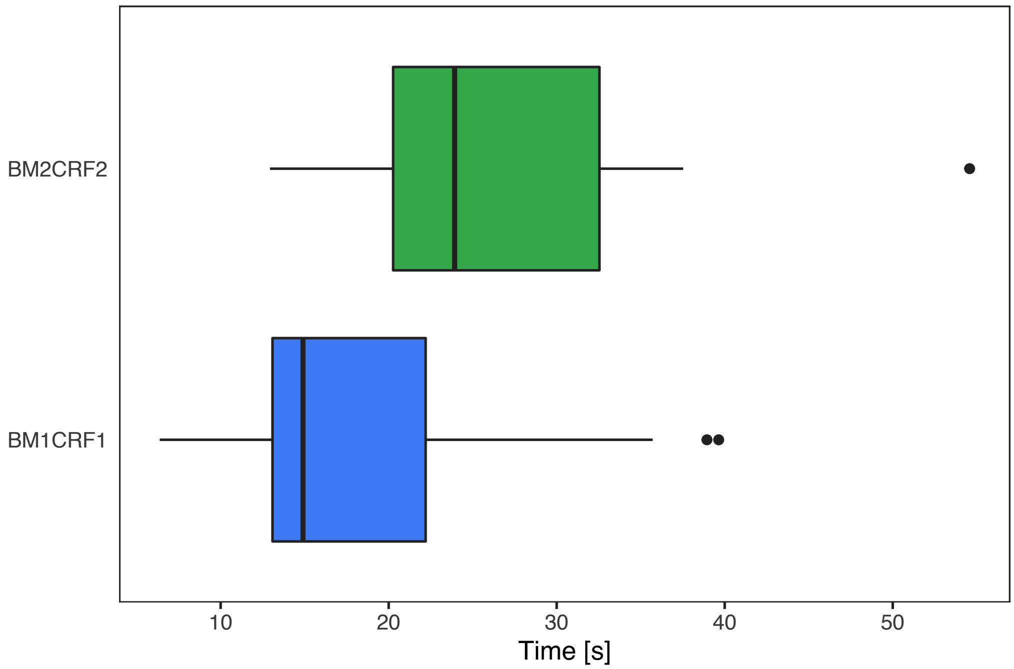



Figure 44: Duration of Task Completion

### Accuracy of the Estimations

Figure 45 shows the estimation of task 5. The red line represents the correct value of 27.6%. Participants' estimates for BM1CRF1 are generally greater, with a wider range. Aside from the plot, the values were evaluated using the accurate value of 27.6%, as a parameter to formulate the null hypothesis. The Wilcoxon-Test was applied to accomplish this (see Table 7). The estimation based on BM1CRF1 differs significantly from the correct answer (p-value = 0.03). However, the estimation based on BM2CRF2 do not differ (p-value = 0.45). It is interesting to see, that when the participants see BM1CRF1, they make decisions faster but estimate worse.

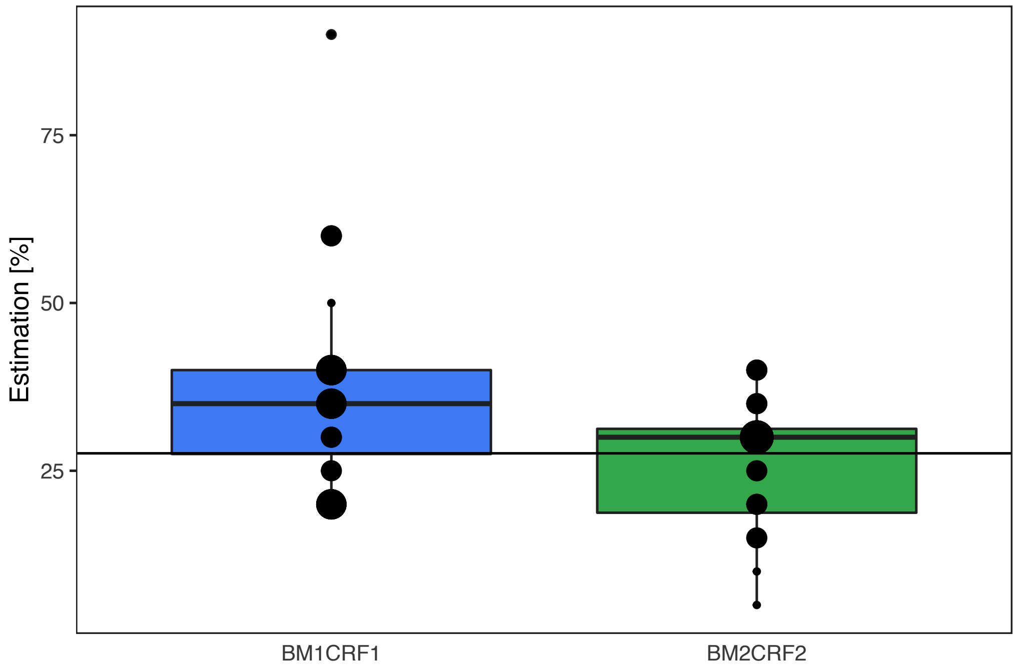



Figure 45: Estimations on Bicycle Infrastructure

| | BM1CRF1 | BM2CRF2 |
|---|---|---|
| V | 149 | 0.03078 |
| p-value | 53 | 0.4516 |

Table 7: Wilcoxon Test

# 5 Discussion

## 5.1 Interpretation of the Results

This section interprets the findings to answer the research questions and test the hypotheses that have been proposed. The first subsection will address research questions 1a and 2a about the visual complexity of bike maps. The effects of the visual complexity on bike map efficiency will be examined in the second subsection. This corresponds to research questions 1b and 2b.

### 5.1.1 Visual Complexity of Bike Maps

**Base Maps**

The following research question and hypothesis address the visual complexity of base maps:

**1a)** How visually complex are different bicycle base maps?

**H1a)** More detailed base maps are visually more complex.

After a quantitative analysis of the experiment's base maps, it can be concluded that the detail of base maps and visual complexity is positively correlated. For all measurements, GMLMT, Feature Congestion, and Edge Density, the measured complexity was higher for the more detailed base map.

This conclusion is intriguing because only two items are added to the more detailed base map: hospitals and buildings. It is reasonable to conclude that size of the added element is an important factor. The measurements are highly sensitive when elements have a lot of edges. Buildings, for example, have a lot of edges and are distributed all over the map, which adds to the visual complexity.

Another factor to consider is contrast. When an element's color is drastically different from its surroundings, it creates a greater edge. This can be observed when the layers are combined. The complexity measure of BM1CRF1 is lower than the sum of BM1 and CRF1. BM1 and CRF1 are depicted on a white background, resulting in sharp edges. Some edges get weaker when they are merged to form BM1CRF1, as the transitions become less sharp. This was also detected during the development of BM1. The complexity of the building's layer decreases as transparency is increased.

Detail may be interpreted in two different ways. First, when more categories are depicted on a map, it becomes more detailed. Second, the level of detail (LOD) has also an effect. In this context, examples are the number of streets or labels that are displayed. The level of detail also has an impact on visual complexity. It is, however, dependent on the size and color of the inserted elements. Adding a few extra labels or small streets, for example, has a minor impact on the estimated map complexity.

The findings of the complexity measurements contradict the distinct-object count in certain ways. As the metrics emphasize major differences, BM2 includes only two additional categories (hospitals and buildings) than BM1. Comparing distinct-object count to quantification measures could be argued to be fairly pointless. The display of one or more objects from a category does not affect the distinct-object count, but it does affect the measurements. This is correct in this case. Distinct-object counts make greater sense when the efficiency of map users is taken into account.

Overall, the hypothesis for 1a) can be accepted. A more detailed base map leads to a visually more complicated cycling map for any quantification method, whether pixel-based or not.

### Cycling Related Features

The following research question and hypothesis address visual complexity of cycling related features:

**2a)** How does the display of different cycling related features affect the visual complexity of bike maps?

**H2a)** More displayed cycling related features are visually more complex.

The quantitative analysis of cycling-related elements reveals that showing more cycling-related features increases visual complexity. Likewise, for base maps, all applied quantification measurements reveal a higher level of complexity. For Sub-band Entropy, the slightest difference can be seen. Compared to GMLMT, and Edge Density, this metric has a different methodology, as it takes the difficulty of encoding a map into account.

Compared to base maps, more categories are added to CRF2. Pumping stations, railway crossings, bicycle signs, bicycle rentals, and easy riding zones are depicted for the higher complexity of cycling related features. Also, bicycle routes have been further dived into main bicycle routes and off-street bicycle routes. In total, there are six more different objects on CRF2 than on CRF1. In this context, adding more cycling-related elements has a smaller effect than adding basic map categories, according to the complexity assessments. This can not be said for all stimuli, but it is the case for the ones used in the experiment.

The easy-riding zone is the feature that covers the most ground. Easy-riding zones are not always depicted on bike maps. However, for example, pedestrian areas are frequently displayed. Pedestrian areas and easy-riding zones both have the characteristic of being widely distributed, resulting in a bigger impact on complexity than other features. Zones have a smaller impact than buildings regarding visual complexity, as zones have fewer edges. In some cases, the depiction of zones can minimize complexity, when it is represented with a color leading to less contrast and thus less strong edges on the maps. But, poor contrast caused map users to perceive this zone worse. It would be interesting to conduct further research and experimentation to see how different colors affect complexity. Likewise for buildings, picking the color can influence measured map complexity.

The hypothesis for 2a) can be accepted. The more cycling-related features that are displayed, the more visually complicated they are. When the cycling-related characteristics are integrated with the base map, however, it's vital to remember that zones might minimize visual complexity.

### 5.1.2 Impacts of Visual Complexity on Bike Map Efficiency

Accepting hypotheses for 1a and 2a allows the usage of the prepared stimuli in the experiment. The eye-tracking experiment was done to learn more about the impact of visual complexity on bike map efficiency. The following are the impact-related study questions and hypotheses:

**1b)** How does visual complexity of base maps affect efficiency of bike maps?

**H1b)** Bike maps with visually complex base maps are less efficient.

**2b)** How does the display of different cycling related features affect efficiency of bike maps?

**H2b)** Bike maps with more displayed cycling related features are less efficient.

The base maps and cycling-related features were always depicted simultaneously in the experiment to create a realistic map. As a result, analyzing BM and CRF separately can be difficult. When the four stimuli are shown together, the effects of BM and CRF on efficiency can still be seen, as each complexity layer appears twice. For example, BM1CRF1 and BM1CRF2 have the same base map complexity. Comparing those two layers, it is still possible to conclude about the influence of base map complexity on efficiency.

### Task 1: Locate the YAH Symbol

For task 1, the participants were expected to find the YAH sign in the stimuli BM1CRF1 first, then BM2CRF1, BM1CRF2, and BM2CRF2. According to the hypotheses, more complex bike maps are less efficient. Furthermore, CRF1 contains fewer symbols than BM1CRF2, thus the participant seeking the symbol may be less distracted.

The position of the symbols had an impact on the outcomes. In some cases, two subtasks performed on the same stimuli show significant differences. This is true for two stimuli (Task 1.1 with Task 1.2, as well as Task 1.3 and Task 1.4). When looking at the stimuli (all of which are included in the appendix) it is striking that symbols towards the middle of the maps are noticed faster than those nearer to the maps' borders. This is independent of the map's complexity. To avoid this problem, complete randomization of the symbol's placement would be needed.

In the next step, the subtasks were summed up to enable comparison between the stimuli. It is interesting to see that overall participants took the least amount of time to locate the symbol BM2CRF1, followed by BM1CRF2, BM1CRF1, and BM2CRF2. The position of the symbols could potentially be a reason why BM2CRF1 has lower values than BM1CRF1. For BM2CRF1, the symbols were again closer located to the center. In both BM1CRF1 and BM2CRF2, the symbols were about the same distance from the center. The symbols in BM2CRF2 are located in areas where map information is rather low. According to Mackworth and Morandi (1967) there is fewer fixation in places where the information load is low. As a result, the high values for stimuli BM2CRF2 can also be explained by the position of the symbols.

As previously stated, complete randomization of this task is required to generate non-biased results. The location appears to have more of an impact on the outcome, than the map complexity. This is also something map designers should bear in mind while creating maps. Selecting an extract with the YAH symbol in the center of the map can help users avoid searching for a long time.

#### Task 2: Count Bicycle Racks

The expectations for the second task were the same as the first. Given the continued emphasis on symbols, less efficient performance for the second level of cycling-related features complexity was expected.

For this task, the outcomes are more in line with the expectations. Maps with fewer cycling related feature complexity have lower values and are thus more efficient than those with more cycling related features. This holds true for both the average time to new fixation and the duration it takes to complete a task. These results are bolstered by the statistical tests performed. In terms of average time to first fixation per bicycle rack it can be observed that for bicycle racks five and six, BM2CRF1 has lower values than BM1CRF. This seems to be rather random and no explanation was found. It is also worth noting that BM2CRF2 was more efficient than BM1CRF2. The distribution of bicycle racks could be again the reason, although nothing unusual can be found.

Another aspect to consider for this task is the correctness of the values entered by the participants. The accuracy matches predictions, with BM1CRF1 having the highest accuracy, followed by BM2CRF1, BM1CRF2, and BM2CRF2. One probable explanation is that as the map becomes more complicated and more symbols are depicted on the map, the map reader becomes more confused and struggles to locate all the bicycle racks. The fact that BM2CRF2 has by far the lowest accuracy rate is intriguing. This aspect demonstrates that also the design of the base map might have an impact on efficiency, as the participants were only partially successful in completing the task.

#### Task 3: Locate a Park

Task 3 was designed to have an antipole of tasks 1 and 2. In those tasks, CRF and symbols were of interest. As a result, task 3 should take into account the various levels of base map complexities. This task was not well-constructed enough. The participants recognized where the parks are on the map after the first subtask, and there was a learning effect as only the names of the parks, not their locations, changed. Rather, the participants' ability to read park names quickly was the focus in the end.

This effect is mirrored in the task's outcomes. There were no significant changes in the average time to new fixation and the time required to complete the activity. Looking at the created boxplot some trends can be seen, however, discussing them in detail would be rather inappropriate.

### Task 4: Search for a Route

In previous assignments, a symbol or a park had to be found. In task 4 the combination of base maps and cycling related features should be both addressed. The assumption for this task is that symbols of less complex maps are detected faster and the time needed for task completion is shorter.

The heat maps reveal that the stimuli's legends are more or less fixated equally often. It would be feasible to discuss the various routes taken by the participants based on the heat maps. When comparing BM1CRF1 with BM2CRF2, map complexity appears to have an effect, as other routes have been partially fixed. Because the impact of complexity on route choice is not considered part of efficiency, it is not covered here.

Time to first fixation of symbols was already investigated in task 1. Many significant differences could be detected in this task. In the instance of task 4, this was not the case. The fixation of the first symbol showed no significant differences. However, a pattern can be seen. The values of BM1CRF1 are lower than those of BM2CRF2, while the values of BM2CRF1 are lower than those of BM1CRF2. Task 1 revealed that finding a symbol takes longer when the amount of cycling related features displayed is greater. For the second symbol, significant differences could be found. Only BM1CRF1 and BM2CRF2 showed no significant differences. It is hard to pinpoint a plausible cause for this case; perhaps further testing should be run on this task to determine if this is an outlier or not. All in all, the same pattern could be observed for the detection of the first symbol. There is a trend that stimuli containing CRF1 symbols are identified faster than stimuli containing the second level of cycling related features.

The previous trend can be noticed again for the duration of task completion. When providing stimuli with a low level of visual complexity, it took less time to decide on a route. However, the location of the stimuli may have influenced the time needed. For this task, the symbols have been inverted. For stimuli BM1CRF1 and BM1CRF2, the route is from top to bottom and from left to right, respectively. It is plausible that if the route is from the bottom to the upper part of the map (BM2CRF2) or from left to right (BM2CRF1) the participants require more time, as this is less intuitive. Hence, the intuition of the participants may have increased the trend. More tests are needed to evaluate the impact of the route's direction.

Overall, task 4 has shown a trend that presenting more cycling related features is less efficient. However, the design of the task also influences the outcomes. More testing with different starting and ending positions on the map is required.

#### Task 5: Estimate Bicycle Infrastructure

The assumption for this work was that the time required to complete the task would be less for the less complex stimulus BM1CRF1 because estimating is easier when only a few elements are depicted on the map.

This assumption proved to be correct. The participants required much less time to give an estimate for BM1CRF1. When the map is less complex, it takes less time to understand and acquire an overview.

Surprisingly, the estimation accuracy for BM2CRF2 was higher. The following could be a hypothesis for this case. People tend to overestimate the proportion of streets with bicycle facilities to other streets. For BM2CRF2, the contrasts of cycling related features to the base map are weaker and the bicycle infrastructure is, therefore, harder to detect. As a result, the bicycle infrastructure is detected worse for more complex stimuli, and the estimation is lower. For verification, this hypothesis would need to be tested further.

This task has two possible constraints. First, the task was rather challenging, it is probable that not all participants truly comprehended it. Second, the sample size is quite limited, with 19 participants looking at BM1CRF1 and 16 looking at BM2CRF2.

It can be concluded that giving an estimate for a less complex map takes less time. However, complexity did not have the expected effect on estimating accuracy. Rather, the design could have an impact on precision.

**Summary**

For research questions 1b) and 2b), the formulated hypotheses cannot be accepted. An influence of map complexity on efficiency could only be found partially, but often just as an observed trend. The experiments on base maps do not reveal any noticeable tendencies. This may be because the experiments lacked tasks that were appropriate for this aim. Task 3 was planned to assess the impact of base map complexity. However, the data reveals that this task was not well built. the learning effect was too large to yield useful results. Hence, no differences between BM1 and BM2 were found. The complexity of cycling related features had a greater influence on efficiency. For the search of the YAH symbol (Task 1) and the counting of the bicycle racks (Task 2) significance could be found. However, the task design also had an impact on the outcomes. The significance of CRF1 being more efficient than CRF2 could not be confirmed in tasks 4 and 5. For both tasks, participants were faster when the cycling related features are displayed as less complex, however, better accuracy was not achieved, and therefore, and so better efficiency was not met. Hence, hypothesis 2b) could not be accepted.

## 5.2 Potential Biases and Limitations of the Study

Potential biases and limitations of the study can be found in the experimental design, the measuring instrument, the participants, and the sample size.

### 5.2.1 Experimental Design

The experiment is primarily set up as a within-subject design. A between-subject design is used in only one out of five tasks. When utilizing a within-subject design, the most significant concern is the participants' learning effect (Martin, 2008). Participants may be able to learn from the maps, and as a consequence, it is feasible that they perform better after seeing the stimuli numerous times. To reduce this effect the stimuli within a task are shown in random order. It would have been even better to randomize the stimuli even across tasks. However, this would lead to major confusion for the participants, having to solve alternating tasks. Overall, it does not appear that the learning effect is a substantial issue. In tasks 1 and 2, the symbols always change their location. In task 3 the names of the parks change, but the park areas remain the same. In this case, the learning effect is certainly an issue. In task 4 the symbols A and B appear on the map for the first time, thus, there is also no learning effect. It could be that the participants already know some streets that are displayed on the map, but the streets were not of importance until this point, as the previous tasks focused on different aspects. Therefore, it is rather unlikely that participants can remember the alignment of the streets.

The fact that the symbols are always put on the same area for all participants in tasks 1, 2, and 4 is a study restriction. Some positions on the map are more likely to get focused on early. A YAH symbol in the center of the stimuli, for example, is likely to be recognized faster than one on the map's edge. This is especially true because a calibration cross is displayed between tasks, causing participants to fixate on the center of the screen. Another randomization should take place here, with each participant's symbols being placed in a distinct location. This would have taken too long for this thesis because Tobii Pro does not allow you to move symbols arbitrarily, and importing various stimuli with different symbol locations takes too long. Furthermore, there are some locations on a map where a YAH symbol is meaningless. If the symbol is put on a body of water, for example.

### 5.2.2 Measuring Instrument

The Tobii TX300 eye-tracker delivers reliable results, measuring the dependent variables of the experiment. The replay recording function described in *3.4.1 Eye-Tracking Features* can be used to verify this. 65cm is the ideal distance between the eyes and the tracking equipment. Participants do not keep this exact distance during the whole experiment. This leads to some inaccuracies. To counterbalance this, the AOIs are drawn slightly larger. It is critical to find a balance while designing AOIs, as too small AOIs may not be hit, and too large AOIs may be overly sensitive, measuring strikes that were not on the actual AOI.

The experiment was carried out in an eye laboratory. The setup resembles a workplace setup. This is probably not the most common place, where people look at bike maps. However, if the experiment is conducted outdoors, more variables can not be controlled, such as sunlight, temperature, and other factors.

### 5.2.3 Participants and Sample Size

The sample of the experiment is rather unbalanced. A minority of the participants are female (11 female and 24 male). They were between the ages of 21 and 34. Because the majority of the participants had a geography background, they declared themselves to be moderately to extremely familiar with maps in general. In addition, the familiarity and frequency with which participants ride bicycles appear to be higher compared to the data of the city of Portland (Dill and McNeil, 2016). Overall, the participants had a different background than the broader public. Furthermore, the sample size of N = 35 results in data that was often not normally distributed. Instead of ANOVA, other non-parametric tests had to be conducted.

Sample size and background might seem unfavorable, however, the sample size was above average when compared to other studies (Jacob and Karn, 2003). Experiments involving people with a background in geography are also rather popular for practical reasons (see Çöltekin et al., 2017, Keil et al., 2020, and Liao et al., 2019).

## 5.3 Possible Enhancement and Further Need for Research

It would be fascinating to do further testing and experimentation to measure the visual complexity of bike maps. Especially, how different colors can influence the measured complexity. Combining the measurements then again with eye-tracking to find out how the map reader's perception is influenced would make up for interesting further research.

Another facet of future study that should be considered is the background of the subjects. In this experiment, but also in other studies, having students or other people with a geographical background is the norm. This can influence the outcome of a study. In this particular case, it is unclear what advantages a background in geography could provide when searching for symbols, parks, choosing routes, and estimating bicycle infrastructure. Experimenting with a more unbiased sample size should be implemented in future research.

The number of static maps has decreased in recent years. Interactive maps are becoming increasingly popular. In this study, a static map has been used for practical reasons, as implementing an interactive map in an experiment is tricky. The experiment has also been conducted in a closed room, to minimize uncontrolled variables. Bike maps used outdoors and maybe on a smaller smartphone screen are more realistic. Thus, putting interactive maps to the test outside would be beneficial.

# 6 Conclusion

This paper combines the topic of visual map complexity with bike maps. In the past, different pixel-based complexity metrics have arisen, and also eye-tracking has become an important tool for assessing map complexity (MacEachren, 1982, Barvir and Vit, 2021). Meanwhile, more and more bike maps are being created to make cycling more appealing to the general public. As there are no design guidelines on how to design a bike map, very different approaches have resulted (Pucher and Buehler, 2008). Wessel and Widener (2015) came up with an attempt to design a perfect bike map suitable for all different kinds of cyclists.

To find out how the design of a bike map influences its complexity, four different bike maps for the city of Cincinnati were created and analyzed. In a factorial design, two layers of base maps and two levels of cycling-related features were combined. GMLMT, Feature Congestion, Subband Entropy, Edge Density, and distinct object counts were used to measure the bike maps' complexity. The measurements showed that more detail in base maps and the depiction of more cycling related features have a positive correlation with map complexity. The size, shape, and color of the elements are considered to have the most impact on complexity measurements. When the element is large and has various boundaries, it adds more edges to the map, increasing its complexity. The color contrast with the surroundings is of importance, as it can contribute to the creation of strong or weak edges. In this particular case, cycling related features had a smaller impact on visual complexity, as the symbols were small in size and number.

The effects of visual complexity on efficiency were explored in addition to measuring the visual complexity of bike maps on a technical level. For this purpose, an eye-tracking experiment with 35 participants was done. Five tasks that were similar to those faced by cyclists on a daily basis had to be completed. The stimuli assessed in the previous step were shown to the participants for this purpose.

No effect of map complexity on efficiency could be found for the base maps. The task constructed to investigate the base map did not work as expected, as there was a big learning effect. In task 2, count bicycle racks, an influence of the base maps on the estimation accuracy could be found. All in all, a definitive answer to this research question can not be given.

For the complexity of cycling related features, influence on efficiency was discovered. Significant variations between the two levels of cycling related features could be noticed in two tasks. When the participants searched for the YAH symbol on the map and when they had to count depicted bicycle racks. However, those two tasks were not fully randomized, thus, the location of the symbols is likely to also have an influence on the participants' performance. For the other tasks none of the stimuli showed a significantly better efficiency. Rather, recurring trends could be observed that complexity has an impact on the map's efficiency.

Not only Zurich, but more and more cities invite their inhabitants to cycle more. To do so, an increasing amount of bike maps are being published. However, the map design of the maps varies greatly. This thesis shows that a wisely designed bike map can influence efficiency and potentially bring even more people onto their bicycles.

# References

Barea, R., Boquete, L., Mazo, M., & López, E. (2002). System for assisted mobility using eye movements based on electrooculography. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *10*(4), 209–218. https://doi.org/10.1109/TNSRE.2002.806829

Barvir, R., & Vit, V. (2021). Graphic Map Load Measuring Tool–development and verification. *International Journal of Cartography*, *7*(3), 285–303. https://doi.org/10.1080/23729333.2021.1972907

Barvir, R., & Vozenilek, V. (2020). Developing versatile graphic map load metrics. *ISPRS International Journal of Geo-Information*, *9*(12). https://doi.org/10.3390/ijgi9120705

Castner, H. W., & Eastman, R. J. (1984). Eye-movement parameters and perceived map complexity—i. *American Cartographer*, *11*(2), 107–117. https://doi.org/10.1559/152304084783914768

Ciolkosz-Styk, A., & Styk, A. (2011). Measuring maps graphical density via digital image processing method on the example of city maps. *Geoionformation Issues*, *3*(1), 61–76. https://doi.org/https://doi.org/10.34867/gi.2011.5

City of Zurich. (2013). *Zurich invites you to cycle*.

Çöltekin, A., Brychtová, A., Griffin, A. L., Robinson, A. C., Imhof, M., & Pettit, C. (2017). Perceptual complexity of soil-landscape maps: a user evaluation of color organization in legend designs using eye tracking. *International Journal of Digital Earth*, *10*(6), 560–581. https://doi.org/10.1080/17538947.2016.1234007

Copenhagenize Design Company. (2019). *Copenhagenize 2019 - the index*. https://copenhagenizeindex.eu/the-index

Dill, J., & McNeil, N. (2016). Revisiting the four types of cyclists: Findings from a national survey. *Transportation Research Record*, *2587*(1), 90–99. https://doi.org/10.3141/2587-11

Ding, L., & Goshtasby, A. (2000). Computer Science and Engineering Department Wright State University. *Pattern Recognition*, *34*(34), 721–725.

Fairbairn, D. (2006). Measuring map complexity. *Cartographic Journal*, *43*(3), 224–238. https://doi.org/10.1179/000870406X169883

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, *24*(6), 631–645. https://doi.org/10.1016/S0169-8141(98)00068-7

Haahr, M. (n.d.). *RANDOM.ORG: true random number service*. Retrieved January 8, 2022, from https://www.random.org

Harrie, L., & Stigmar, H. (2007). An evaluation of measures for quantifying map information. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(3), 266–274. https://doi.org/10.1016/j.isprsjprs.2009.05.004

Hervé, A., & Williams, L. J. (2010). Honestly Significant Difference (HSD) Test. *Encyclopedia of Research Design*, 1–5. https://doi.org/10.4135/9781412961288.n181

Hubel, D. (1995). *Eye, Brain, and Vision.* Scientific American Library.

Informing Design, Inc. (2015). *The nashville groove bike map.* https://www.nashville.gov/sites/default/files/2021-06/NashvilleGroove.pdf?ct=1623768924

Instituto Distrital de Turismo. (2015). *n/A.*

Jacob, R. J., & Karn, K. S. (2003). Eye Tracking in Human-Computer Interaction and Usability Research. *The Mind's Eye*, 573–605. https://doi.org/10.1016/b978-044451020-4/50031-1

Jenny, B. (2020). *Color oracle. design for the color impaired.*

Joos, M., Rötting, M., & Velichkovsky, B. M. (2017). Spezielle Verfahren I: Bewegungen des menschlichen Auges: Fakten, Methoden und innovative Anwendungen. *Psycholinguistik*, (1), 142–168. https://doi.org/10.1515/9783110114249.2.142

Keil, J., Edler, D., Kuchinke, L., & Dickmann, F. (2020). Effects of visual map complexity on the attentional processing of landmarks. *PLoS ONE*, *15*(3), 1–20. https://doi.org/10.1371/journal.pone.0229575

Klippel, A., Freksa, C., & Winter, S. (2006). You-are-here maps in emergencies –the danger of getting lost. *Journal of Spatial Science*, *51*(1), 117–131. https://doi.org/10.1080/14498596.2006.9635068

Li, Z., & Huang, P. (2002). Quantitative measures for spatial information of maps. *International Journal of Geographical Information Science*, *16*(7), 699–709. https://doi.org/10.1080/13658810210149416

Liao, H., Wang, X., Dong, W., & Meng, L. (2019). Measuring the influence of map label density on perceived complexity: a user study using eye tracking. *Cartography and Geographic Information Science*, *46*(3), 210–227. https://doi.org/10.1080/15230406.2018.1434016

MacEachren, A. M. (1982). Map complexity: Comparison and measurement. *American Cartographer*, *9*(1), 31–46. https://doi.org/10.1559/152304082783948286

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception Psychophysics*, *2*(11), 547–552. https://doi.org/10.3758/BF03210264

Mangold, R. (2013). *Informationspsychologie - Wahrnehmen und Gestalten in der Medienwelt.* Springer.

Martin, M., David. (2008). *Doing psychology experiments.* Thomson Wadsworth.

MATLAB. (2019). *Version 9.6.0 (r2019a)*. The MathWorks Inc.
Montello, D. R. (2002). Cognitive map-design research in the twentieth century: Theoretical and empirical approaches. *Cartography and Geographic Information Science*, *29*(3), 283–304. https://doi.org/10.1559/152304002782008503
Montello, D. R. (2010). You are where? The function and frustration of you-are-here (YAH) maps. *Spatial Cognition and Computation*, *10*(2-3), 94–104. https://doi.org/10.1080/13875860903585323
Oja, P., Titze, S., Bauman, A., de Geus, B., Krenn, P., Reger-Nash, B., & Kohlberger, T. (2011). Health benefits of cycling: A systematic review. *Scandinavian Journal of Medicine and Science in Sports*, *21*(4), 496–509. https://doi.org/10.1111/j.1600-0838.2011.01299.x
Oliva, A., Mack, M. L., Shrestha, M., & Peeper, A. (2004). Identifying the perceptual dimensions of visual complexity of scenes. *Proc. of the 26th Annual Meeting of the Cognitive Science Society*, 1041–1046.
OpenStreetMap contributors. (2022). *Planet dump retrieved from https://planet.osm.org*.
Pucher, J., & Buehler, R. (2008). Making cycling irresistible: Lessons from the Netherlands, Denmark and Germany. *Transport Reviews*, *28*(4), 495–528. https://doi.org/10.1080/01441640701806612
Purves, D., Augustine, George, J., Fitzpatrick, D., Katz, Lawrence, C., LaMantia, A.-S., McNamara, James, O., & Williams, Mark, S. (2001). No Types of Eye Movement and Their Functions. *Neuroscience*, *2*.
Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005). Feature congestion: a measure of display clutter. *Proceedings of the SIGCHI ...*, 761–770. http://dl.acm.org/citation.cfm?id=1055078
Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2), 1–22. https://doi.org/10.1167/7.2.17
Royston, P. (1995). Remark AS R94: A Remark on Algorithm AS 181: The W-test for Normality. *Journal of the Royal Statistical Society*, *44*(4), 547–551.
Rybarczyk, G. (2014). Simulating bicycle wayfinding mechanisms in an urban environment. *Urban, Planning and Transport Research*, *2*(1), 89–104. https://doi.org/10.1080/21650020.2014.906909
Schnur, S., Bektas, K., Salahi, M., Çöltekin, A., Bekta, K., Salahi, M., & Çöltekin, A. (2010). A Comparison of Measured and Perceived Visual Complexity for Dynamic Web Maps. *GIScience 2010*, (March 2015). https://doi.org/10.5167/uzh-38771
Schnur, S., Bektaş, K., & Çöltekin, A. (2018). Measured and perceived visual complexity: a comparative study among three online map providers. *Cartography*

*and Geographic Information Science*, *45*(3), 238–254. https://doi.org/10.1080/15230406.2017.1323676

Serif Ltd. (2022). *Affinity designer*.

Slim naar Antwerpen. (2019). *Fietsplan Antwerpen en omgeving*.

Speed, A., Stracuzzi, D., Lee, J., & Hund, L. (2017). Applying Image Clutter Metrics to Domain-Specific Expert Visual Search. *Sandia National Laboratories*.

The GIMP Development Team. (2019, June 12). *Gimp*. https://www.gimp.org

Tobi Pro. (2021). Pro Lab User Manual.

Vairalkar, M., & Nimbhorkar, S. U. (2012). Edge Detection of Images Using Sobel Operator. *International Journal of Emerging Technology and Advanced Engineering*, *2*(1), 291–293. www.ijetae.com

Vélohop. (2020). *Plan vélo vélostras*.

Ware, C. (2019). *Information visualization: Perception for design*. Morgan Kaufmann.

Weiner, I. B., & Craighead, E. W. (2009). Mann-Whitney U Test. *The Corsini Encyclopedia ofPsychology*. https://doi.org/10.3109/9780203450307-28

Wessel, N., & Widener, M. (2015). Rethinking the urban bike map for the 21st century. *Cartographic Perspectives*, (81), 6–22. https://doi.org/10.14714/CP81.1243

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501. https://doi.org/10.1038/nrn1411

Young, L. R., & Sheena, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods  Instrumentation*, *7*(5), 397–429. https://doi.org/10.3758/BF03201553

Zayed, M. A. (2016). Towards an index of city readiness for cycling. *International Journal of Transportation Science and Technology*, *5*(3), 210–225. https://doi.org/10.1016/j.ijtst.2017.01.002

# Glossary

| Notation | Description |
|---|---|
| ANOVA | Analysis of Variance |
| AOI | Area of Interest |
| BM | Base Map |
| CRF | Cycling Related Features |
| DPI | Dots per Inch |
| EML | Eyemovement Recording Lab |
| GIVA | Geographic Information Visualization and Analysis |
| GMLMT | Graphic Map Load Measurement Tool |
| HSD | Honestly Significant Difference |
| LOD | Level of Detail |
| TTFF | Time to First Fixation |
| YAH | You Are Here |

# List of Figures

# List of Tables

# Appendix

## A Metrics

| Eye-Movement Metric | What it Measures |
|---|---|
| Number of fixation overall | More overall fixations indicate less efficient search (perhaps due to sub-optimal layout of the interface). |
| Fixations per area of interest | More fixations on a particular area indicate that it is more noticeable, or more important, to the viewer than other areas. |
| Fixation duration | A longer fixation duration indicates difficulty in extracting information, or it means that the object is more engaging in some way. |
| Gaze | Gaze is usually the sum of all fixation durations within a prescribed area. It is best used to compare attention distributed between targets. It can also be used as a measure of anticipation in situation awareness if longer gazes fall on an area of interest before a possible event occurring. |
| Fixation spatial density | Fixations concentrated in a small area indicate focussed and efficient searching. Evenly spread fixations reflect widespread and inefficient search. |
| Repeat fixations | Higher numbers of fixations off-target after the target has been fixated indicate that it lacks meaningfulness or visibility. |
| Time to first fixation on-target | If a low proportion of participants is fixating an area that is important to the task, it may need to be highlighted or moved. |
| Percentage of participants fixating an area of interest | If a low proportion of participants is fixation an area that is important to the task, it may need to be highleted or moved. |
| On-target (all target fixations) | Fixations on-target divided by total number of fixations. A lower ratio indicates lower search efficiency. |

Fixation-derived metrics and possible interpretation (Jacob & Karn 2003)

| Eye-Movement Metric | What it Measures |
|---|---|
| Number of saccades | More saccades indicate more searching. |
| Saccade amplitude | Larger saccades indicate more meaningful cues, as attention is drawn from a distance. |
| Saccades revealing marked directional shifts | Any accade larger than 90 degrees from the saccade that preceded it shows a rapid change in direction. This could mean that the user's have changed or the interface layout does not match the user's expectations. |

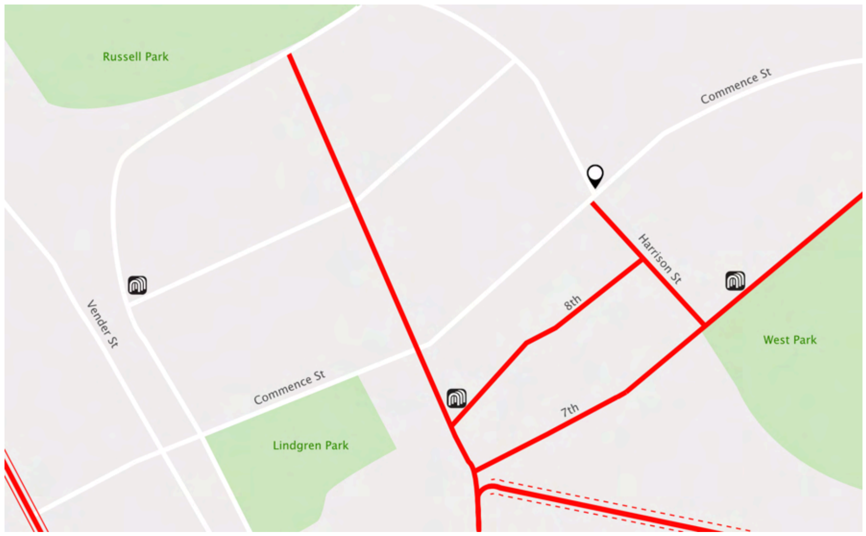
Saccades-derived metrics and possible interpretation (Jacob and Karn, 2003)

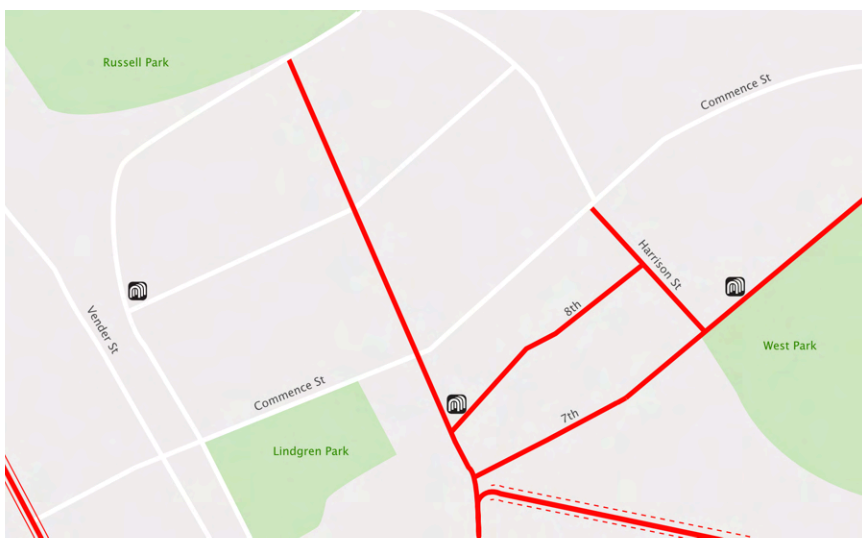
| Eye-Movement Metric | What it Measures |
|---|---|
| Scanpath duration | A longer-lasting scanpath indicates less efficient scanning. |
| Scanpath length | A longer scanphat indicates less efficient searching (perhaps due to a sub-optimal layout). |
| Spatial density | Smaller spatial density indicates more direct search. |
| Scanpath direction | This can determine a participant's search strategy. |
| Saccade/fixation ratio | This compares time spent searching (saccades) to time spent processing (fixating). A higher ratio indicates more processing or less searching. |

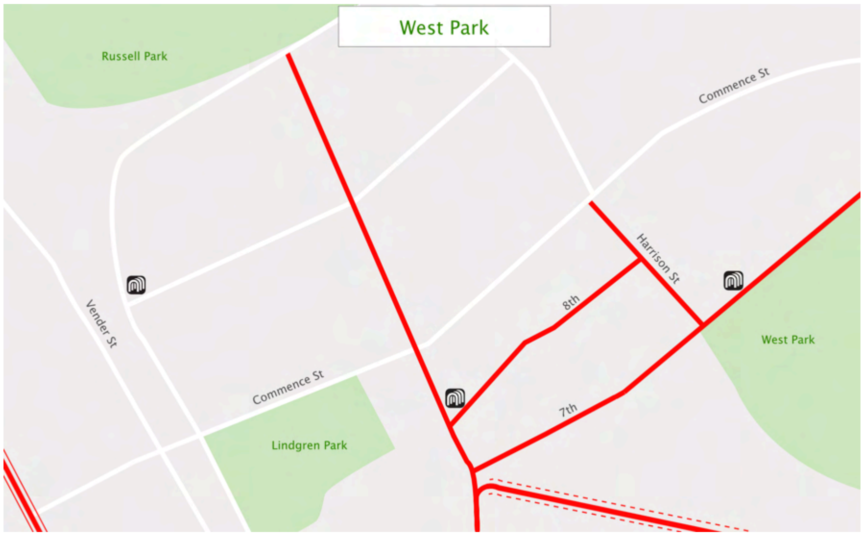
Scanpath-derived metrics and possible interpretation (Jacob & Karn 2003)

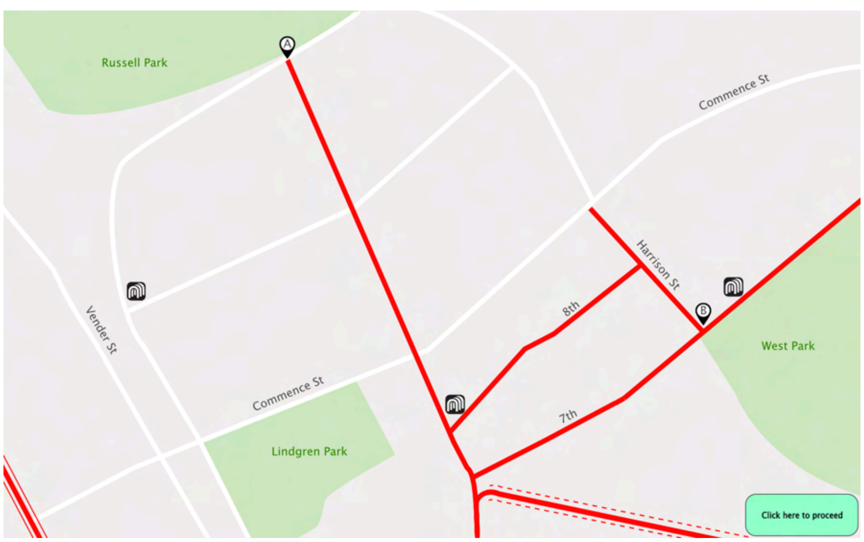
# B Stimuli

## Dry Run




West Park
Russell Park
Commence St
Harrison St
8th
7th
West Park
Vender St
Commence St
Lindgren Park


A
Russell Park
Commence St
Harrison St
B
8th
7th
West Park
Vender St
Commence St
Lindgren Park
Click here to proceed

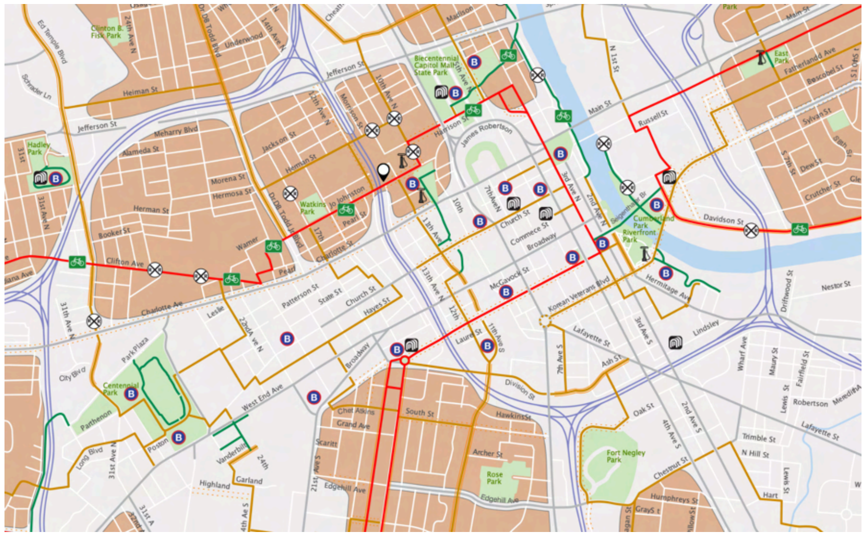

Task 5: Estimate the percentage of roads that have bicycle infrastructure (bike routes, bike lanes).
Click anywhere on the screen when you are ready to enter your estimation.
Bike Infrastructure
Bike Lanes
Physically Protected Bike Lanes
Bike Routes
Commence St
Harrison St
8th
Vender St
West Park
Commence St
7th
Lindgren Park

## Main Run

### Task 1

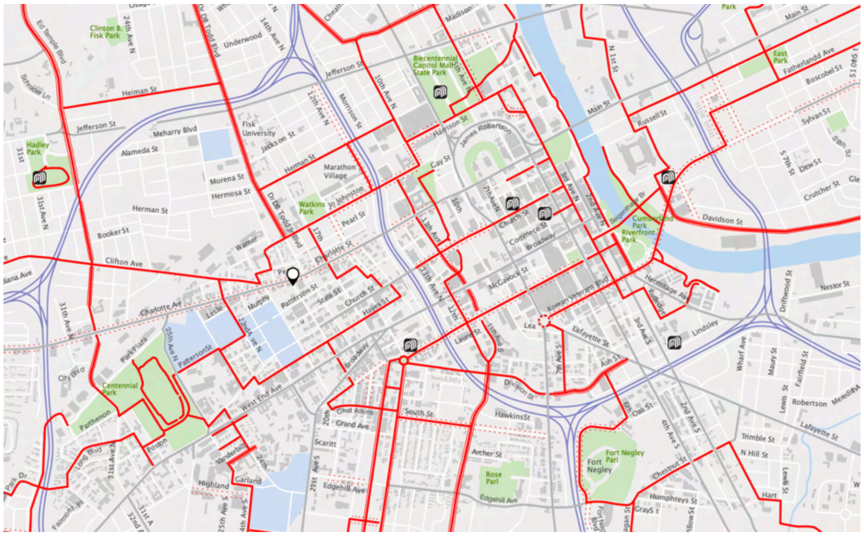

Task 1.1


Task 1.2

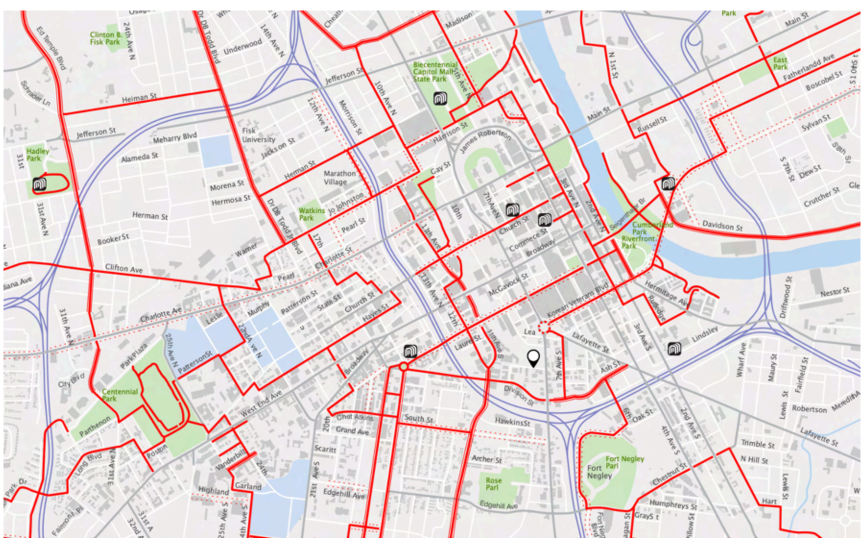



Task 1.3

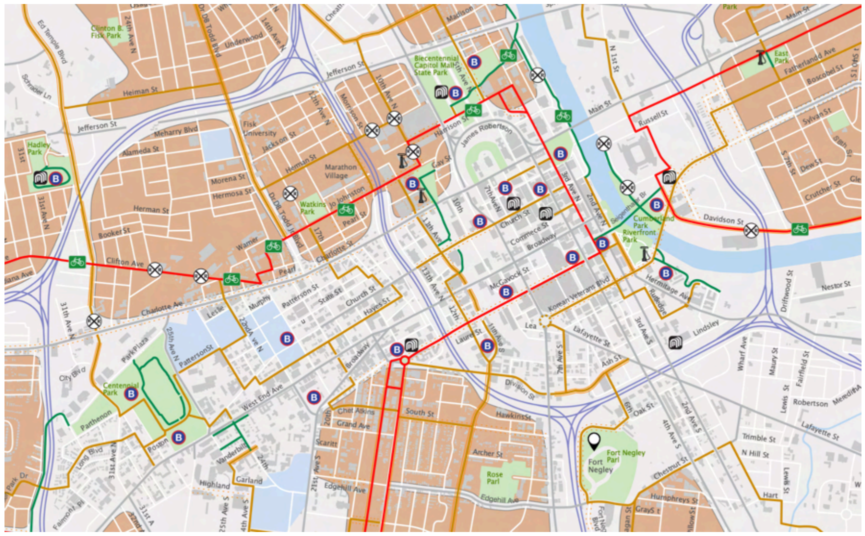

Task 1.4



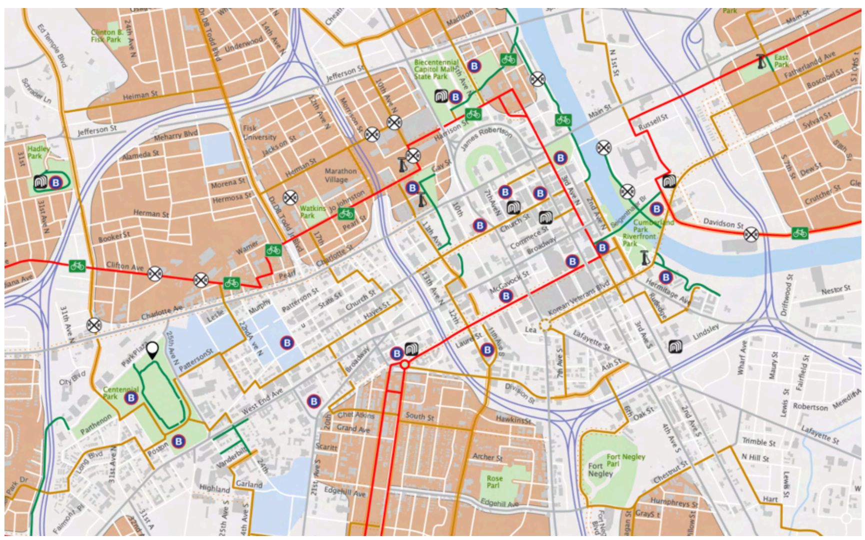
Task 1.5



Task 1.6



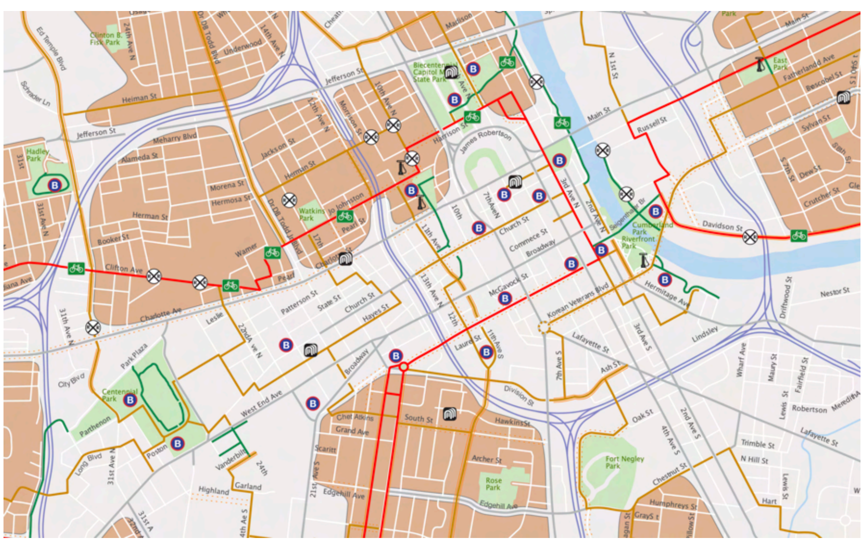

Task 1.7



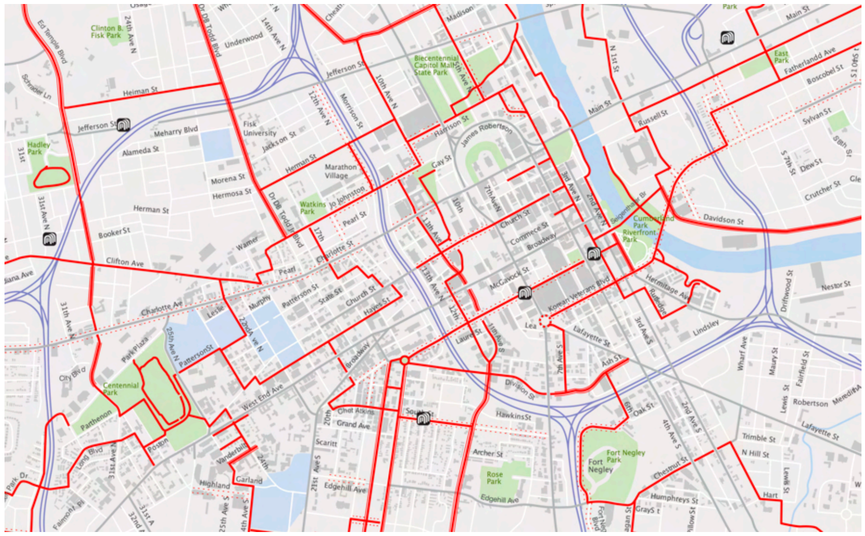

Task 1.8

**Task 2**

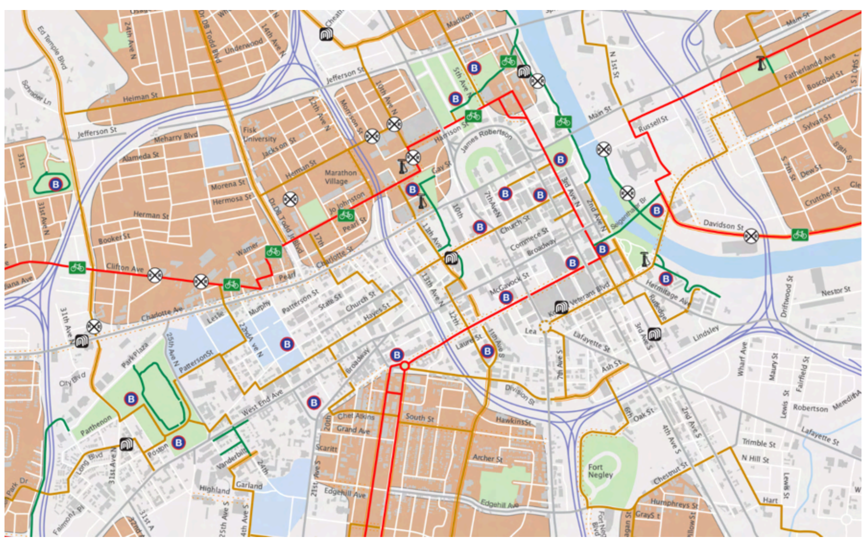
Task 2.1



Task 2.2

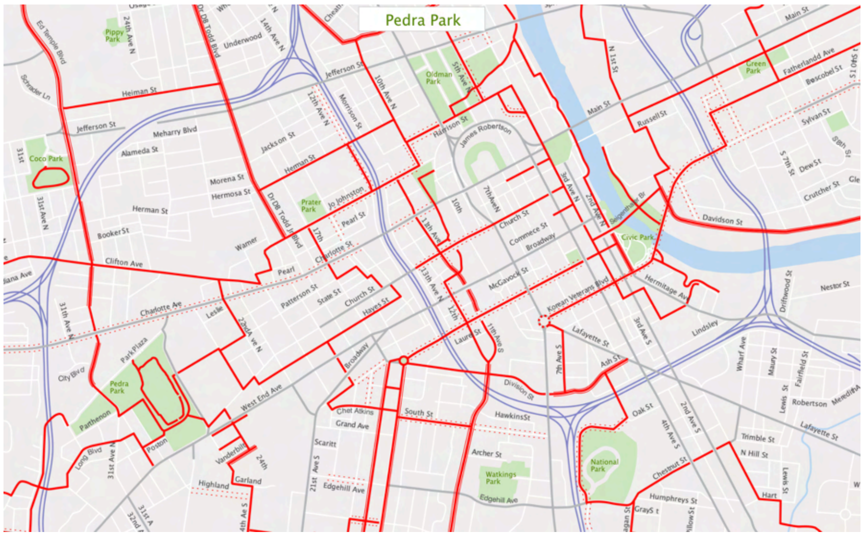


Task 2.3

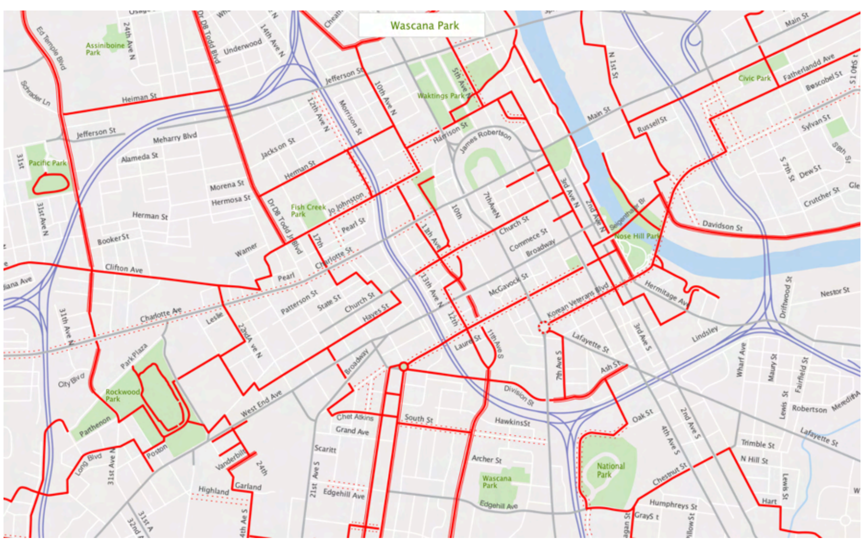


Task 2.4

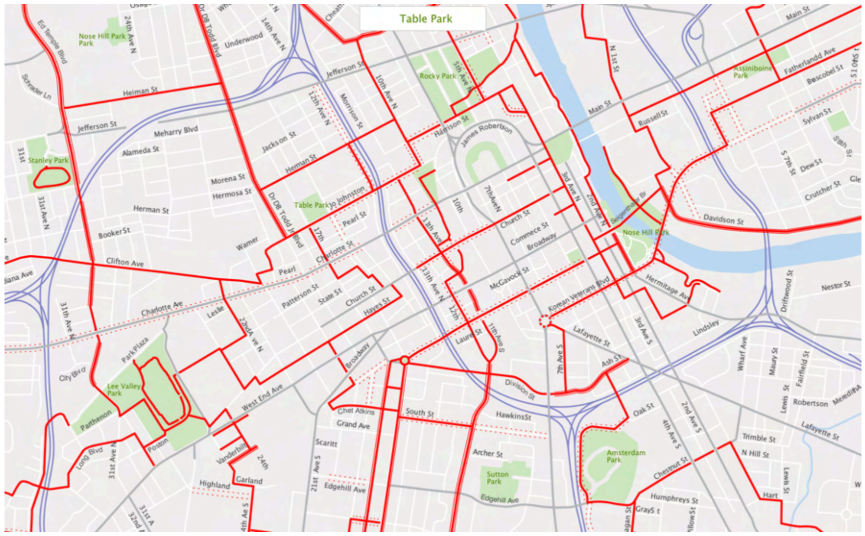
**Task 3**



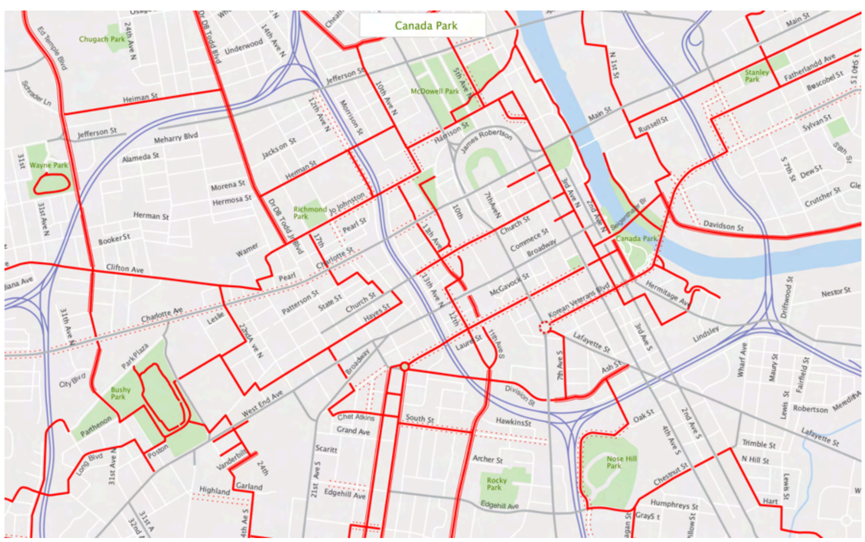

Task 3.1



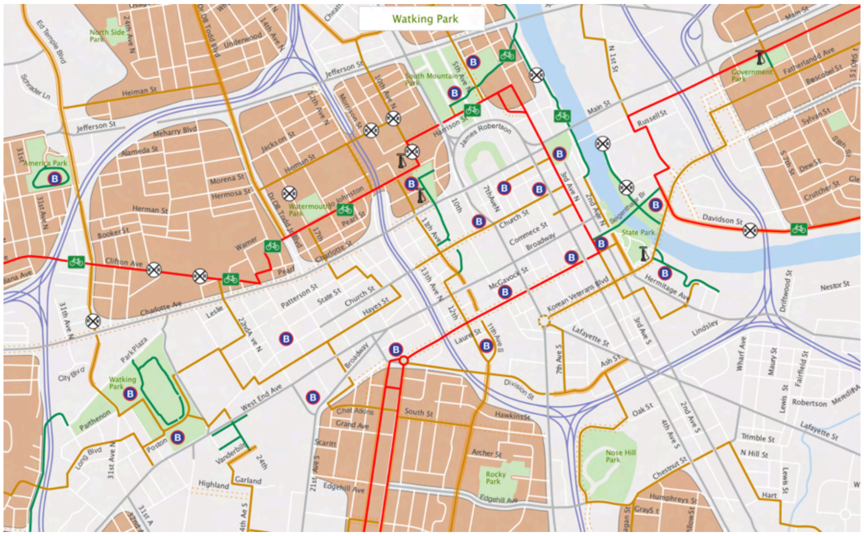

Task 3.2



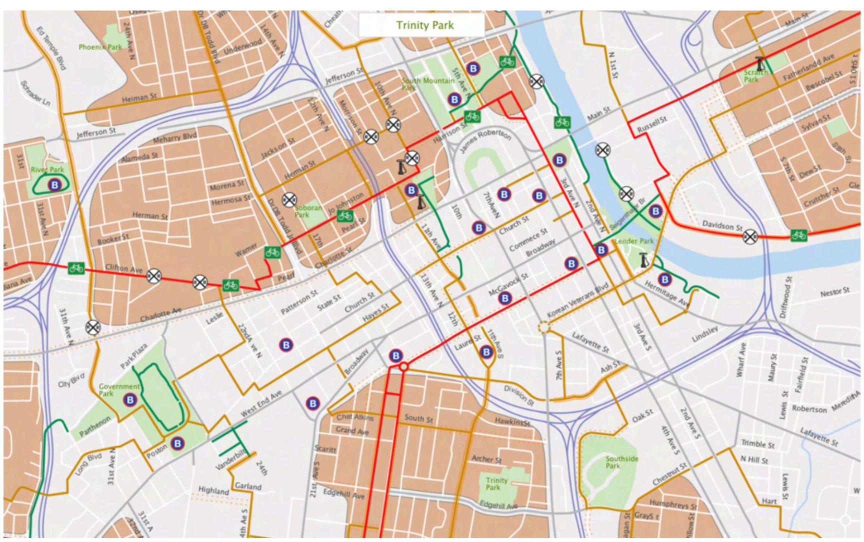

Task 3.3



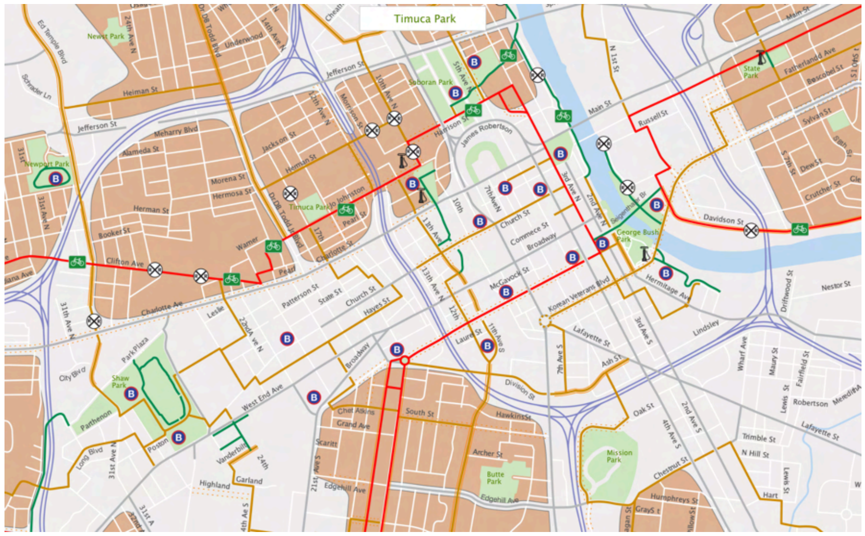

Task 3.4



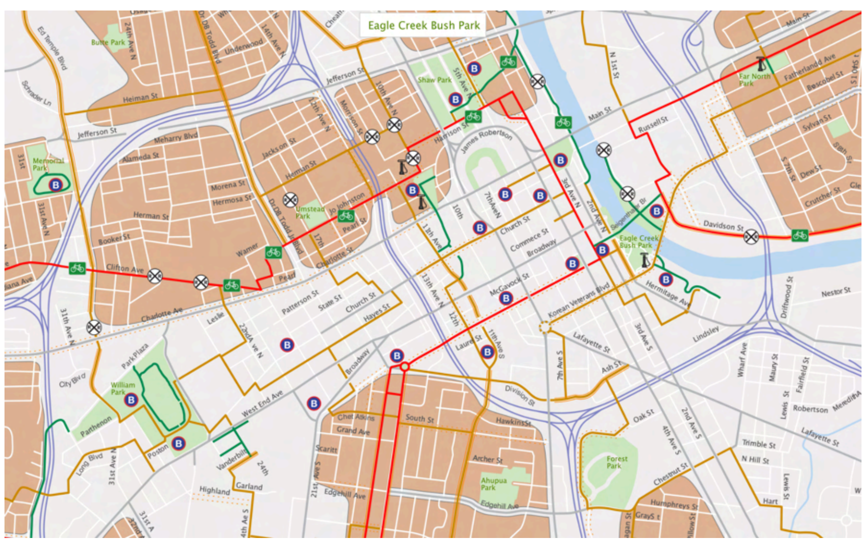

Task 3.5



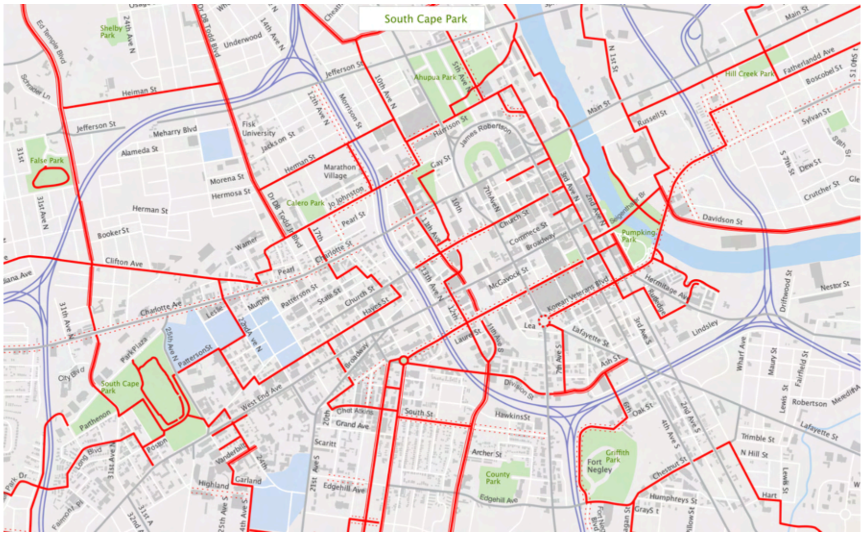

Task 3.6



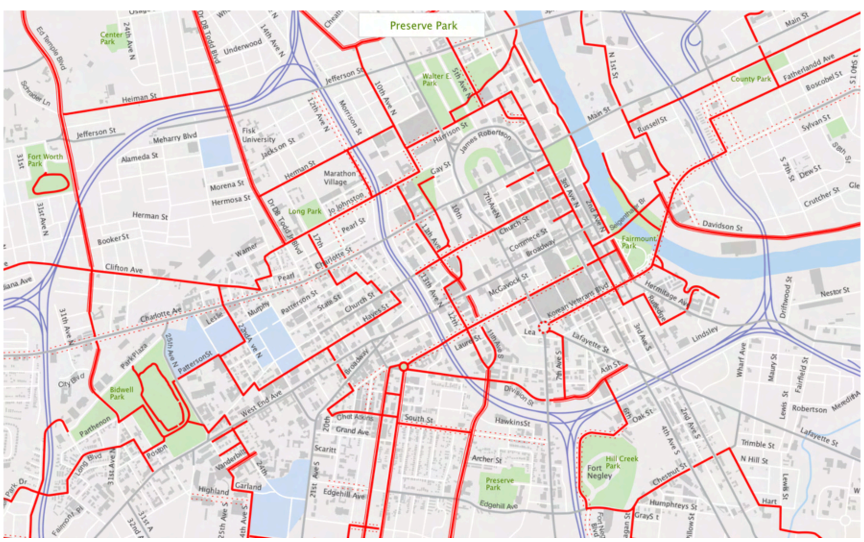

Task 3.7



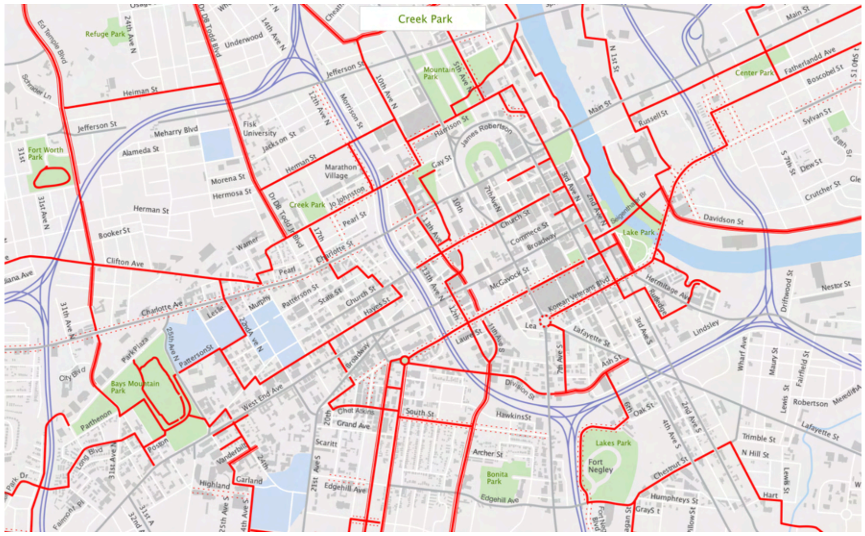

Task 3.8



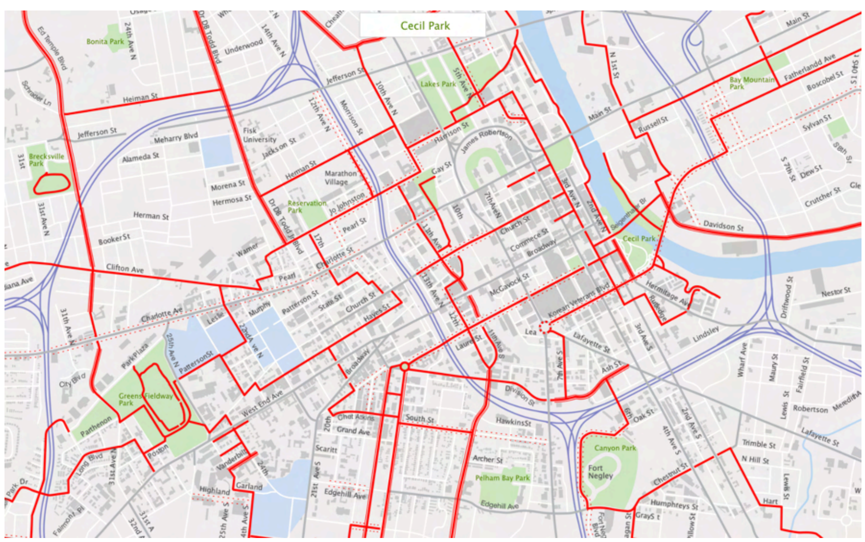

Task 3.9



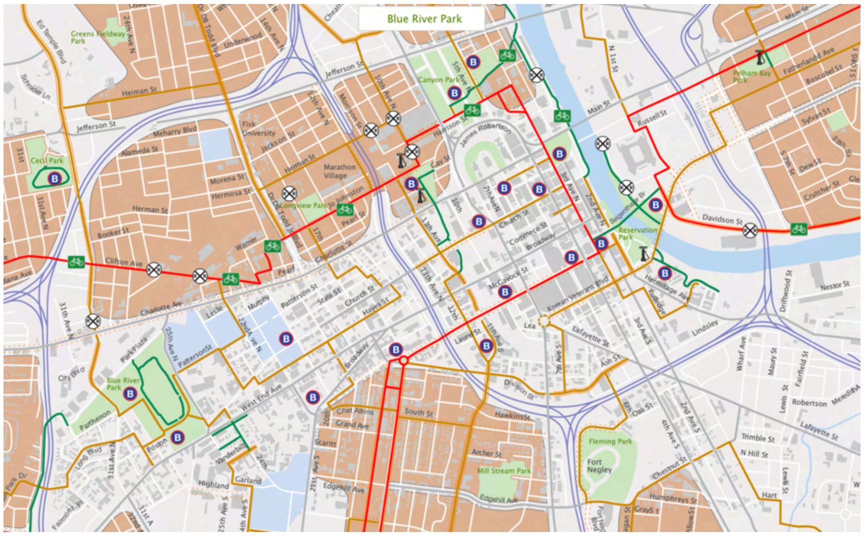

Task 3.10



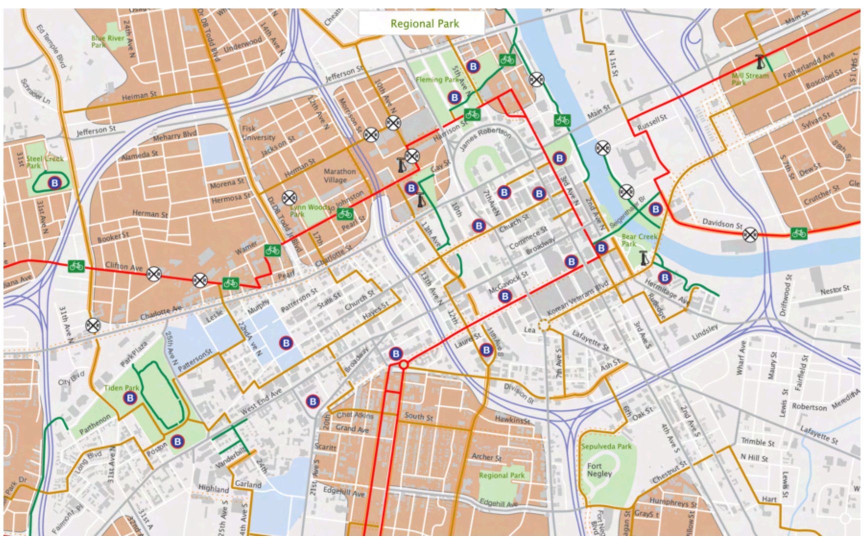

Task 3.11



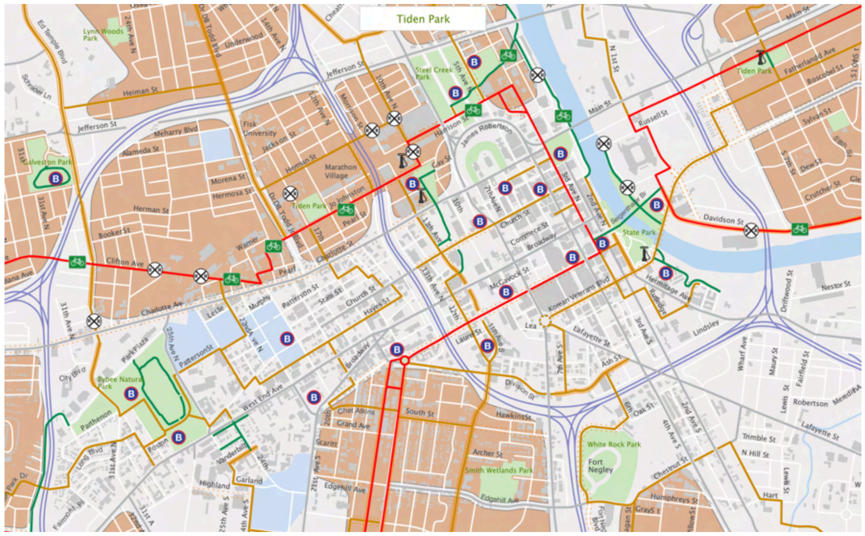

Task 3.12



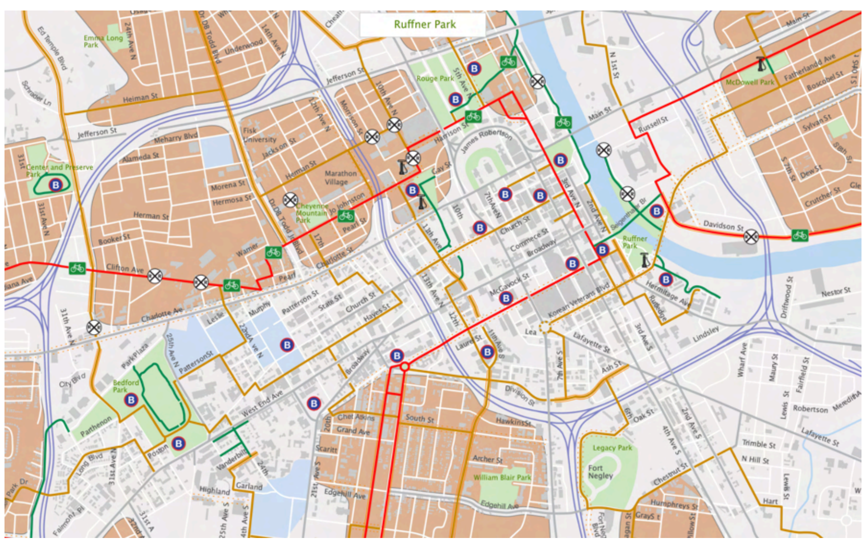

Task 3.13




Task 3.14




Task 3.15

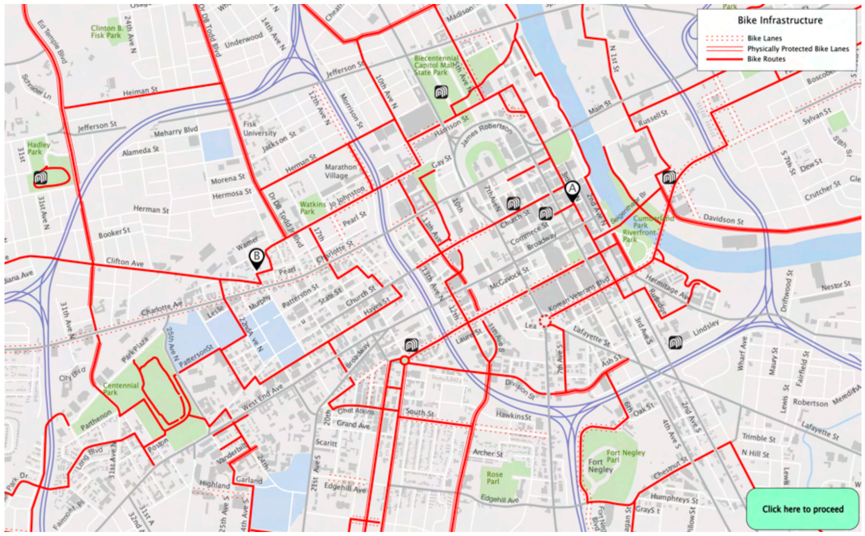



Task 3.16

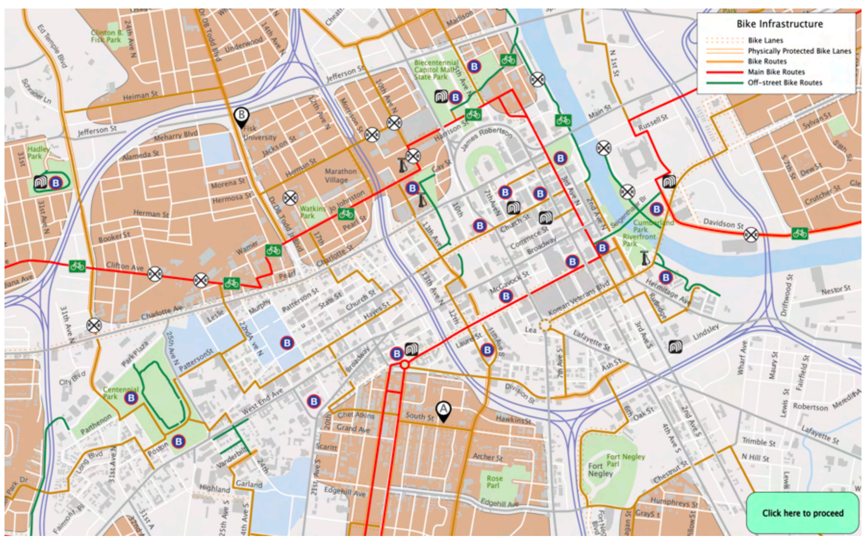
**Task 4**


Task 4.1


Task 4.2

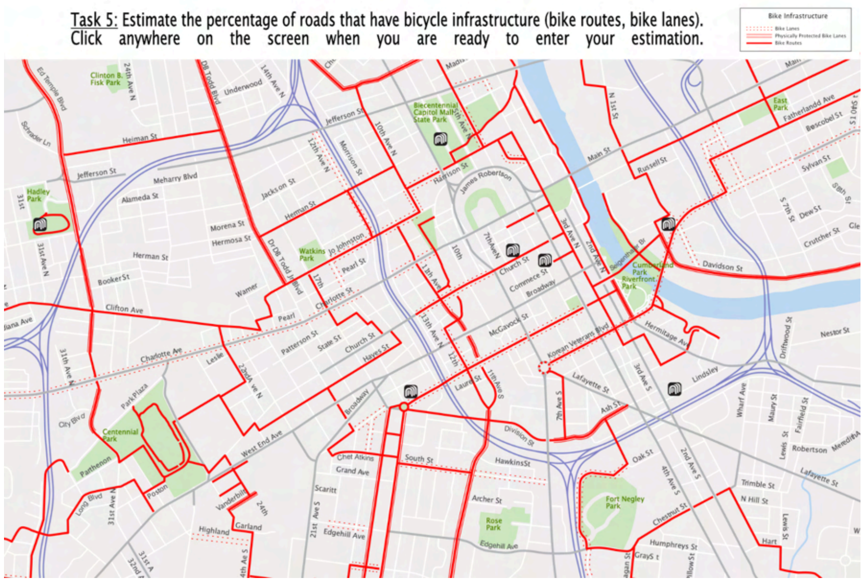



Task 4.3

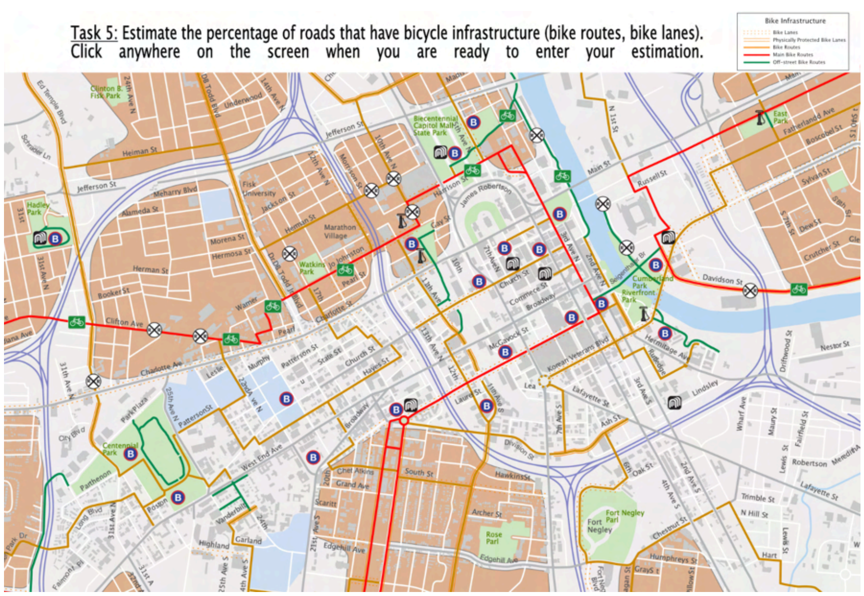



Task 4.4

## Task 5

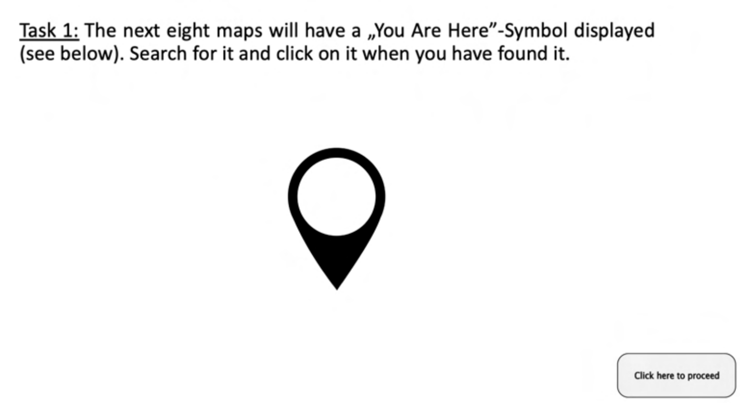


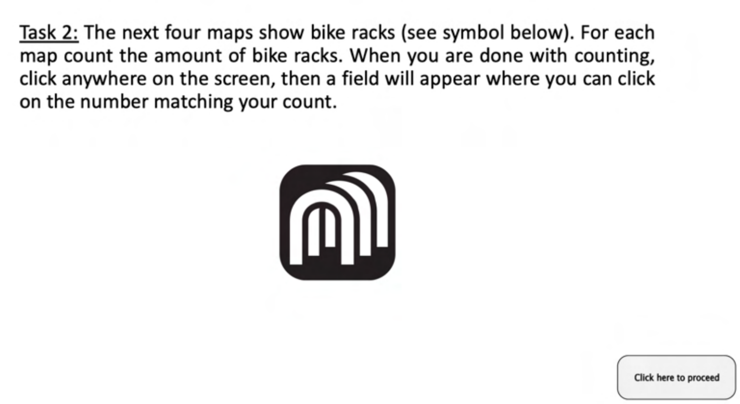

Task 5.1




Task 5.2

# C Task Descriptions



How many bike racks were displayed on the map? Click on the number.

1 2 3 4 5 6 7 8 9 10 11 12

Task 3: On the next four maps, you will have to search for a specific park (see an example for a park below). On top of the map, the name of the park to be searched will be displayed. Once you have found the park, click on it.

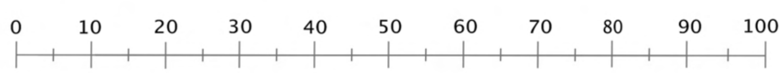

Click here to proceed

Task 4: The next four maps each display a starting point (A) and an end point (B) (see symbols below). Search for the fastest route between A and B. After you have found the fastest route, use the mouse to follow the route by hovering over it.


Click here to proceed

Task 5: Estimate the percentage of roads that have bicycle infrastructure (bike routes, bike lanes). Click anywhere on the screen when you are ready to enter your estimation.


Estimate the percentage of roads that have bicycle infrastructure (bike routes, bike lanes). Click on the scale below on the number that is closest to your estimation.



# D Procedure

Procedure Eye-Tracking MSc Donat Büchel

Procedure Eye-Tracking

| Phase | What to say | What to do | Time |
|---|---|---|---|
| 0. Preparation | - | Sanitize surfaces, print consent form, prepare setup (handle, computer, non-movable chair), chocolate, Tobii (create new participant) & questionnaire (in web) | (5') |
| 1. Welcome | Welcome, thanks for participating in my user study, small talk | Close door, "do not disturb" sign | 1' |
| 3. Consent Form | You received the consent form via mail. Could you already read it? (If needed you can take yourself a few minutes to read it.) Do you have any questions regarding the consent form? Please sign here. I will send you a scan via email. As written in the consent form you can withdraw your participation at any time also after today's experiment. | Sign consent form | 1' |
| 2. Study Introduction | As you have read in the consent form, I am doing research on visual complexity of bike maps. Therefore, I will show you bike maps on this computer screen, and you will complete different tasks. While you perform these tasks your eye movements will be recorded. | - | 1' |

1

| | | | |
|---|---|---|---|
| 4. Eye Tracking Introduction | We are now conducting a dry run of the eye-tracking. Eye-Tracking means that your eye movement will be recorded while you look at the screen. In a next step, we will conduct the dry run. The aim is to already have a look at the tasks to make sure that everything is clear. There is a cross in between some of the maps you should look there. And in one task you hover with the mouse, make sure you don't click but just hover. | | 1' |
| 5. Calibration (Dry Run) | On the screen you can see this square and your eyes. You should move a little bit… (backward, forward, left, right etc.) Make sure that you sit comfortably, as you should try to avoid moving. With one hand you can grab this mouse as you will need it in the experiment. With the other hand you can also grab this handle if you want to. The handle can also help with staying in place. Make sure that you are seated comfortably as you may get lower during the experiment. Now we will calibrate the eye tracking system. Therefore, you will see a small circle around the screen. You should always focus on the middle of the circle. (…) The calibration seems fine / We will do the calibration once again. | Adjust chair, make sure that eyes are in the zone, recalibrate if needed | 2' |





| | | | |
|---|---|---|---|
| 6. Conducting the dry run | We are ready to start the dry run. | Start eye-tracking, have an eye on the participant to make sure that everything is working. | 3' |
| 7. Getting Ready for the real run | Do you have any questions regarding the eye tracker or the tasks? Are you able to understand the tasks? When there is a cross you should look there. If not, we will do the calibration once again, this time for a little bit longer than before. Then, we will start with the actual experiment. | | 2' |
| 8. Calibration | Again, follow the circle. The calibration seems fine / We will do the calibration once again. | Adjust chair, make sure that eyes are in the zone, recalibrate if needed (threshold for recalibration 10%) | 2' |
| 9. Conducting the experiment | I will now start the experiment. | | 7' |
| 10. Questionnaire | Do you need a break or are you all good? I have a small questionnaire for you (about three minutes) about you and your familiarity with maps. | Open LimeSurvey | 2' |
| Goodbye | Thanks for participating in the study! | Hand over chocolate | 2' |
| Post-processing | - | Back up data, scan & send signed consent form to the participant | |

# E Declaration of Consent

## English

Declaration of Consent

Dear study participant

You are invited to participate in a study conducted by Donat Büchel (+41 76 493 76 55, donat.buechel@uzh.ch) as part of his master's thesis "Visual Complexity of Bike Maps" at the Department of Geography, University of Zurich.

**Purpose of the study**
The purpose of this study is to investigate how bicycle maps should be designed so they can be understood better by the map readers. The aims at investigating if and how changes in the level of detail influence the map's effectiveness.

**General Information**
The main study will take place at the Institute of Geography of the University of Zurich (Campus Irchel) in the "Eye Movement Lab" (Y25-L9) and will last about 20 minutes.

**Procedure of the study**
If you decide to take part in the study, you will first be guided through a short introduction. In the next step, you will be asked to complete a series of tasks with maps on the computer. During this part, your eye movements will be recorded. This is completely harmless and painless for you. Finally, you will be asked to fill out a questionnaire with information about yourself. Response times, recorded eye movements, and the questionnaire will all be recorded anonymously.

**Voluntary participation**
Your participation in this study is voluntary. You can withdraw your consent to participate in this study at any time without giving reasons. You can also ask questions about the study at any time.

**Advantages for the study participant**
This study offers no direct benefits to the study participant.

**Confidentiality of data**
This study involves the collection of personal data. All data will be encrypted and anonymized by replacing your name with a code. In addition, your name will not be used in the work. All data collected will be kept encrypted and stored on secure media. Your data can be used in anonymised form for publication in the scientific community.

The personal data provided here will be stored for a period of 10 years due to a legal obligation. A local ethics committee may review the information during this period. All information is stored in a locked archive and on a secure server at the Department of Geography of the University of Zurich.

**Cost for the study participant**
The study does not incur any direct costs for the study participant.

**Compensation**
There is no financial compensation for participating.

**Termination of participation**
Your participation in the study will be discontinued,

- If you are unable to understand / follow instructions from the investigator.
- If you withdraw from participation in the study. If you withdraw your participation, your records will be deleted.

| Place, Date | Signature of the **participant** |
| --- | --- |
| | |

☐ The participant has received the information contained on this form verbally upon request.

**Experimenter's Statement:** I certify that I have explained the study and the use of the study participant's data. I have encouraged the study participant to seek an explanation of the experiment and his/her rights. If there are any changes that affect the study participant during the course of the experiment, I will inform them immediately and ask for their consent. I certify that this study meets all legal obligations and complies with national rules and international guidelines for human experimentation.

| Place, Date | Signature of the **experimenter** |
| --- | --- |
| | |

# German

Einverständniserklärung

Sehr geehrte Studienteilnehmerin, sehr geehrter Studienteilnehmer

Sie sind eingeladen, an einer Studie teilzunehmen, die von Donat Büchel (+41 76 493 76 55, donat.buechel@uzh.ch) im Rahmen seiner Masterarbeit „Visual Complexity of Bike Maps" am Geographischen Institut der Universität Zürich durchgeführt wird.

**Zweck der Studie**
Der Zweck dieser Studie ist es zu untersuchen, wie Velokarten gestaltet werden müssen, damit die Karte von Nutzer*innen besser gelesen werden kann. Dabei wird darauf fokussiert, ob und wie eine Veränderung des Detailgrades der Karte die Effektivität der Karte beeinflusst.

**Allgemeine Information**
Die Hauptstudie findet am Geographischen Institut der Universität Zürich (Campus Irchel) im „Eye Movement Lab" (Y25-L9) statt und wird ca. 20 Minuten dauern.

**Studienablauf**
Wenn Sie sich für eine Teilnahme an der Studie entscheiden, erfolgt als Erstes eine kurze Einführung. Im nächsten Schritt werden Sie gebeten, eine Serie von Aufgaben mit Karten am Computer zu bearbeiten. In diesem Schritt werden Ihre Augenbewegungen aufgezeichnet. Dies ist für Sie völlig ungefährlich und schmerzfrei. Zum Schluss werden Sie gebeten, einen Fragebogen mit Angaben zu Ihrer Person auszufüllen. Die Antwortzeiten, Augenbewegungen und der Fragenbogen werden alle anonymisiert aufgenommen.

**Freiwillige Teilnahme**
Ihre Teilnahme an dieser Studie ist freiwillig. Sie können Ihre Einwilligung zur Teilnahme an dieser Studie jederzeit ohne Angabe von Gründen widerrufen. Sie können auch jederzeit Fragen zur Studie stellen.

**Vorteile für Studienteilnehmende**
Diese Studie bietet keine direkten Vorteile für den Studienteilnehmenden.

**Vertraulichkeit der Daten**
Diese Studie beinhaltet die Erfassung Ihrer persönlichen Daten. Alle Daten werden durch das Ersetzen Ihres Namens mit einem Code verschlüsselt und anonymisiert. Darüber hinaus wird Ihr Name nicht in der Arbeit verwendet. Alle gesammelten Daten werden verschlüsselt aufbewahrt und auf sicheren Datenträgern gespeichert. Ihre Daten können in anonymisierter Form in der wissenschaftlichen Community publiziert werden.

Die erfassten personenbezogenen Daten werden aufgrund einer gesetzlichen Verpflichtung für einen Zeitraum von 10 Jahren gespeichert. Eine lokale Ethikkommission kann die Informationen in diesem Zeitraum prüfen. Alle Informationen werden in einem abgeschlossenen Archivschrank sowie auf einem sicheren Server am Geographischen Institut der Universität Zürich gespeichert.

**Kosten für Studienteilnehmende**
Die Studie verursacht keine direkten Kosten für den Studienteilnehmenden.

**Entschädigung**
Für die Teilnahme gibt es keine finanzielle Entschädigung.

**Abbruch der Teilnahme**
Ihre Teilnahme an der Studie wird abgebrochen,

- wenn Sie nicht in der Lage sind, Anweisungen des Versuchsleiters zu verstehen / zu befolgen.
- wenn Sie die Teilnahme an der Studie widerrufen. Sollten Sie Ihre Teilnahme zurückziehen, werden Ihre Aufzeichnungen gelöscht.

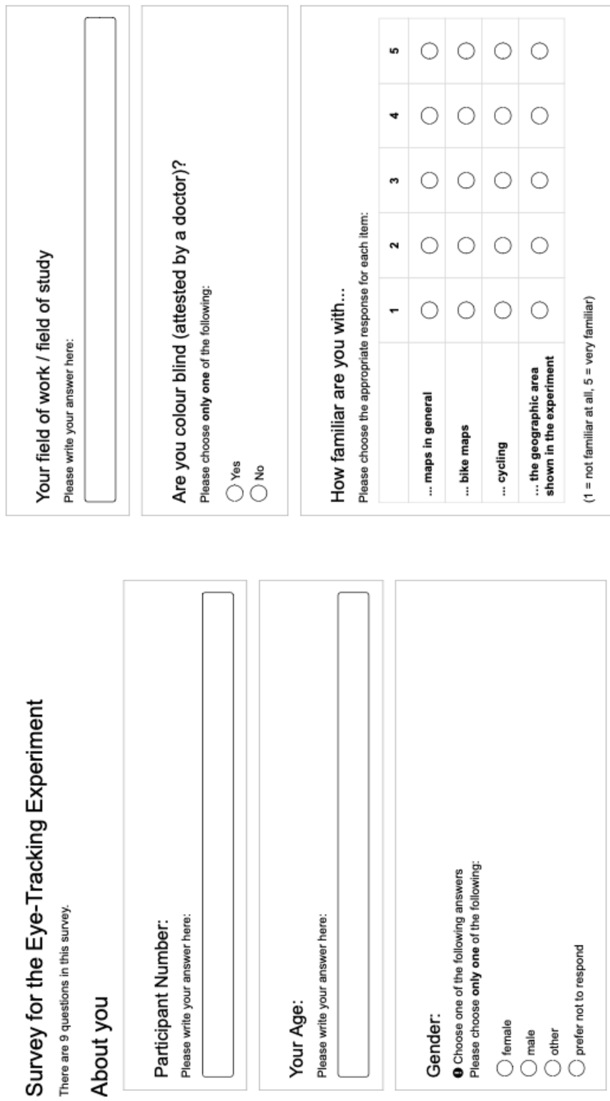
| Ort, Datum | Unterschrift **Studienteilnehmer*in** |
|---|---|
| | |

☐ Die auf diesem Formular enthaltenen Informationen hat der Teilnehmer auf Anfrage mündlich erhalten.

**Erklärung des Versuchsleiters:** Ich bestätige, dass ich die Studie sowie die Verwendung der Daten des Studienteilnehmers erklärt habe. Ich habe den Studienteilnehmer ermutigt, sich um eine Erklärung des Experiments und seiner Rechte zu bemühen. Sollten sich im Laufe des Versuchs Änderungen ergeben, die den Studienteilnehmer betreffen, werde ich ihn unverzüglich informieren und um Zustimmung bitten. Ich bestätige, dass diese Studie alle gesetzlichen Verpflichtungen erfüllt und mit den nationalen Regeln und internationalen Richtlinien für Humanexperimente übereinstimmt.

| Ort, Datum | Unterschrift **Versuchsleiter** |
|---|---|
| | |

# F Questionnaire


Survey for the Eye-Tracking Experiment

There are 9 questions in this survey.

About you

Participant Number:

Please write your answer here:

Your Age:

Please write your answer here:

Gender:

Choose one of the following answers

Please choose **only one** of the following:

- female
- male
- other
- prefer not to respond

Your field of work / field of study

Please write your answer here:

Are you colour blind (attested by a doctor)?

Please choose **only one** of the following:

- Yes
- No

How familiar are you with...

Please choose the appropriate response for each item:

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| ... maps in general | | | | | |
| ... bike maps | | | | | |
| ... cycling | | | | | |
| ... the geographic area shown in the experiment | | | | | |

(1 = not familiar at all, 5 = very familiar)

How familiar are you with...

Please choose the appropriate response for each item:

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **... Google Maps** | ○ | ○ | ○ | ○ | ○ |
| **... Apple Maps** | ○ | ○ | ○ | ○ | ○ |
| **... Bing Maps** | ○ | ○ | ○ | ○ | ○ |
| **... OpenStreetMap** | ○ | ○ | ○ | ○ | ○ |
| **... paper maps** | ○ | ○ | ○ | ○ | ○ |

(1 = not familiar at all, 5 = very familiar)

Do you use some sort of maps that have not been mentioned above?

Please write your answer here:

How often do you use…

Please choose the appropriate response for each item:

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **... maps in general** | ○ | ○ | ○ | ○ | ○ |
| **... bike maps** | ○ | ○ | ○ | ○ | ○ |
| **... bicycles** | ○ | ○ | ○ | ○ | ○ |

(1 = never, 5 = very often)

# G Statistical Tables

## Task 1

| | Task1.1 | Task1.2 | Task1.3 | Task1.4 | Task1.5 | Task1.6 | Task1.7 | Task1.8 |
|---|---|---|---|---|---|---|---|---|
| p-value | 0.05992 | 0.1049 | 0.09026 | 2.975e-06 | 0.01213 | 1.089e-05 | 0.04065 | 0.01698 |
| W-statistic | 0.753958 | 0.754891 | 0.754286 | 0.7548 | 0.750893 | 0.76172 | 0.753542 | 0.75188 |

Shapiro-Wilk Test for Normal Distribution per Subtask

| | Task1.1 | Task1.2 | Task1.3 | Task1.4 | Task1.5 | Task1.6 | Task1.7 | Task1.8 |
|---|---|---|---|---|---|---|---|---|
| p-value | 0.4783 | 0.3008 | 0.02295 | 0.03373 | 0.766 | 0.1914 | 0.8489 | 0.5971 |
| W-statistic | 0.75706 | 0.756402 | 0.752158 | 0.03373 | 0.757837 | 0.755322 | 0.758291 | 0.757397 |

Shapiro-Wilk Test for Log-Normal Distribution per Subtask

| | BM1CRF1 | BM1CRF2 | BM2CRF1 | BM2CRF2 |
|---|---|---|---|---|
| p-value | 7.28e-05 | 2.182e-07 | 2.789e-08 | 0.001093 |
| W-statistic | 0.750559 | 0.82714 | 0.77834 | 0.75318 |

Shapiro-Wilk Test for Normal Distribution per Stimuli

**Task 2**

| | BM1CRF1 | BM1CRF2 | BM2CRF1 | BM2CRF2 |
|---|---|---|---|---|
| p-value | 1.082e-05 | 0.1201 | 0.01632 | 0.3333 |
| W-statistic | 0.78567 | 0.755083 | 0.752203 | 0.756556 |

Shapiro-Wilk Test for Normal Distribution of Average Time to New Fixation

| | BM1CRF1 | BM1CRF2 | BM2CRF1 | BM2CRF2 |
|---|---|---|---|---|
| p-value | 0.0648 | 0.8675 | 0.5041 | 0.1265 |
| W-statistic | 0.754209 | 0.758358 | 0.757212 | 0.755155 |

Shapiro-Wilk Test for Log-Normal Distribution of Average Time to New Fixation

| | Df | F-value | Pr(>F) |
|---|---|---|---|
| group | 3 | 0.38 | 0.77 |
| | 136 | | |

Levene's Test for Average Time to New Fixation (log)

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| as.factor(Stimuli) | 3 | 13.41 | 4.469 | 39.16 | >2e-16 |
| Residuals | 136 | 15.52 | 0.114 | | |

ANOVA for Average Time to New Fixation (log)

| Stimuli Combination | diff | lwr | upr | p adj |
|---|---|---|---|---|
| BM1CRF2-BM1CRF1 | 0.70551135 | 0.4954638 | 0.751555891 | 0.0000000 |
| BM2CRF1-BM1CRF1 | 0.06339211 | -0.1466554 | 0.27343967 | 0.8611565 |
| BM2CRF2-BM1CRF1 | 0.57957767 | 0.3695301 | 0.78962523 | 0.0000000 |
| BM2CRF1-BM1CRF2 | -0.64211924 | -0.8521668 | -0.43207168 | 0.0000000 |
| BM2CRF2-BM1CRF2 | -0.12593368 | -0.3359812 | 0.08411388 | 0.4052588 |
| BM2CRF2-BM2CRF1 | 0.51618556 | 0.3061380 | 0.72623312 | 0.0000000 |

Tukey HSD for Average Time to New Fixation (log)

| | BM1CRF1 | BM1CRF2 | BM2CRF1 | BM2CRF2 |
|---|---|---|---|---|
| p-value | 0.0002349 | 0.005911 | 0.0005169 | 4.095e-05 |
| W-statistic | 0.85035 | 0.750642 | 0.86508 | 0.81512 |

Shapiro-Wilk Test for Normal Distribution of Task Duration

| | BM1CRF1 | BM1CRF2 | BM2CRF1 | BM2CRF2 |
|---|---|---|---|---|
| p-value | 0.03852 | 0.75302 | 0.01049 | 0.2005 |
| W-statistic | 0.753464 | 0.75862 | 0.751534 | 0.755811 |

Shapiro-Wilk Test for Log-Normal Distribution of Task Duration

| | BM1CRF1 | BM1CRF2 | BM2CRF1 | BM2CRF2 |
|---|---|---|---|---|
| p-value | 0.005804 | 0.0001078 | 0.0004125 | 4.545e-05 |
| W-statistic | 0.750613 | 0.8351 | 0.86094 | 0.81733 |

Shapiro-Wilk Test for Normal Distribution of Mean Fixation Count

| | BM1CRF1 | BM1CRF2 | BM2CRF1 | BM2CRF2 |
|---|---|---|---|---|
| p-value | 0.02462 | 0.1469 | 0.01672 | 0.1037 |
| W-statistic | 0.752813 | 0.755367 | 0.752239 | 0.754874 |

Shapiro-Wilk Test for Log-Normal Distribution of Mean Fixation Count

**Task 3**

| | BM1CRF1 | BM1CRF2 | BM2CRF | BM2CRF2 |
|---|---|---|---|---|
| p-value | 0.5466 | 0.1861 | 0.1503 | 0.2622 |
| W-statistic | 0.757351 | 0.755704 | 0.754029 | 0.756199 |

Shapiro-Wilk Test for Normal Distribution of Time to New Fixation

| | Df | F-value | Pr(>F) |
|---|---|---|---|
| group | 15 | 1.15 | 0.32 |
| | 123 | | |

Levene's Test for Average Time to New Fixation

| | Df | F-value | Mean Sq | F-value | Pr(>F) |
|---|---|---|---|---|---|
| Num | 1 | 0.02 | 0.02 | 0.03 | 0.85 |
| Residuals | 136 | 26.82492.15 | 0.67 | | |

ANOVA for Average Time to New Fixation

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| BM1CRF2-BM1CRF1 | 0.31710837 | -0.1858518 | 0.82006850 | 0.3597842 |
| BM2CRF1-BM1CRF1 | -0.14243254 | -0.6490774 | 0.36421233 | 0.8843678 |
| BM2CRF2-BM1CRF1 | 0.08669983 | -0.4162603 | 0.58965995 | 0.75698307 |
| BM2CRF1-BM1CRF2 | -0.45954091 | -0.75661858 | 0.04710395 | 0.0900422 |
| BM2CRF2-BM1CRF2 | -0.23040854 | -0.7333687 | 0.27255158 | 0.6331539 |
| BM2CRF2-BM2CRF1 | 0.22913237 | -0.2775125 | 0.73577723 | 0.6427220 |

Tukey HSD for Average Time to New Fixation

| | BM1CRF1 | BM1CRF2 | BM2CRF | BM2CRF2 |
|---|---|---|---|---|
| p-value | 3.644e-10 | 0.2164 | 0.1381 | 5.856e-05 |
| W-statistic | 0.46426 | 0.75592 | 0.755279 | 0.82264 |

Shaprio-Wilk Test for Normal Distribution of Duration of Task Completion

| | BM1CRF1 | BM1CRF2 | BM2CRF | BM2CRF2 |
|---|---|---|---|---|
| p-value | 0.0003051 | 0.8833 | 2.826e-09 | 0.2129 |
| W-statistic | 0.85532 | 0.758417 | 0.54324 | 0.755897 |

Shapiro-Wilk Test for Log-Normal Distribution of Duration of Task Completion

| | BM1CRF1 | BM1CRF2 | BM2CRF1 |
|---|---|---|---|
| BM1CRF2 | 0.11 | - | - |
| BM2CRF1 | 0.57 | 0.11 | - |
| BM2CRF2 | 0.41 | 0.41 | 0.11 |

Mann-Whitney U Test for Duration of Task Completion

**Task 4**

| | BM1CRF1 | BM1CRF2 | BM2CRF | BM2CRF2 |
|---|---|---|---|---|
| p-value | 1.668e-06 | 6.609e-05 | 5.839e-07 | 1.944e-07 |
| W-statistic | 0.72617 | 0.82515 | 0.71241 | 0.68148 |

Shapiro-Wilk Test for Normal Distribution of Time to First Fixation Symbol A

| | BM1CRF1 | BM1CRF2 | BM2CRF | BM2CRF2 |
|---|---|---|---|---|
| p-value | 0.01617 | 0.3668 | 0.0481 | 0.01985 |
| W-statistic | 0.751804 | 0.756702 | 0.753784 | 0.752495 |

Shapiro-Wilk Test for Log-Normal Distribution of Time to First Fixation Symbol A

| | BM1CRF1 | BM1CRF2 | BM2CRF | BM2CRF2 |
|---|---|---|---|---|
| p-value | 3.892e-06 | 1.296e-09 | 5.093e-09 | 3.426e-08 |
| W-statistic | 0.74863 | 0.51417 | 0.56443 | 0.62858 |

Shapiro-Wilk Test for Normal Distribution of Time to First Fixation Symbol B

| | BM1CRF1 | BM1CRF2 | BM2CRF | BM2CRF2 |
|---|---|---|---|---|
| p-value | 0.1323 | 0.008392 | 0.4808 | 0.00377 |
| W-statistic | 0.754992 | 0.75119 | 0.757133 | 0.89923 |

Shapiro-Wilk Test for Log-Normal Distribution of Time to First Fixation Symbol B

| | BM1CRF1 | BM1CRF2 | BM2CRF1 |
|---|---|---|---|
| BM1CRF2 | 0.75347876 | - | - |
| BM2CRF1 | 0.1730721 | 0.1895660 | - |
| BM2CRF2 | 0.4485234 | 0.5930706 | 0.07382893 |

Mann-Whitney U Test for Time to First Fixation Symbol A

| | BM1CRF1 | BM1CRF2 | BM2CRF1 |
|---|---|---|---|
| BM1CRF2 | 0.00873 | - | - |
| BM2CRF1 | 0.00189 | 2e-07 | - |
| BM2CRF2 | 0.85492 | 0.02507 | 0.00062 |

Mann-Whitney-U Test for Time to First Fixation Symbol B

| | BM1CRF1 | BM1CRF2 | BM2CRF | BM2CRF2 |
|---|---|---|---|---|
| p-value | 0.4716 | 0.4524 | 0.3677 | 0.0003642 |
| W-statistic | 0.757101 | 0.757033 | 0.756706 | 0.85863 |

Shaprio-Wilk Test for Normal Distribution of Duration of Task Completion

| | BM1CRF1 | BM1CRF2 | BM2CRF | BM2CRF2 |
|---|---|---|---|---|
| p-value | 0.2305 | 0.714 | 0.009843 | 0.5114 |
| W-statistic | 0.756011 | 0.757863 | 0.751436 | 0.757236 |

Shapiro-Wilk Test for Log-Normal Distribution of Duration of Task Completion

| | BM1CRF1 | BM1CRF2 | BM2CRF1 |
|---|---|---|---|
| BM1CRF2 | 0.5230 | - | - |
| BM2CRF1 | 0.0258 | 0.0622 | - |
| BM2CRF2 | 0.0258 | 0.0056 | 8.8e-06 |

Mann-Whitney U Test for Duration of Task Completion

| | BM1CRF1 | BM1CRF2 | BM2CRF | BM2CRF2 |
|---|---|---|---|---|
| p-value | 2.46e-08 | 6.272e-07 | 2.476e-08 | 4.735e-07 |
| W-statistic | 0.6179 | 0.71436 | 0.61812 | 0.70666 |

Shapiro-Wilk Test for Fixation Count on Legend

| | BM1CRF1 | BM1CRF2 | BM2CRF1 |
|---|---|---|---|
| BM1CRF2 | 9.1e-06 | - | - |
| BM2CRF1 | 6.5e-08 | 0.016 | - |
| BM2CRF2 | 2.7e-10 | 6.0e-06 | 0.002 |

Mann-Whitney U Test for Fixation Count on Legend

**Task 5**

| | BM1CRF1 | BM2CRF2 |
|---|---|---|
| p-value | 0.003438 | 0.1032 |
| W-statistic | 0.83178 | 0.750677 |
| p-value (log) | 0.355 | 0.75152 |
| W-statistic (log) | 0.754729 | 0.757528 |

Shapiro-Wilk Test (Log-)Normal Distribution for Time of Task Completion

| | Df | F | Pr(>F) |
|---|---|---|---|
| group | 1 | 0.346 | 0.5604 |
| 33 | | | |

Levene's Test for Time of Task Completion

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Stimuli | 1 | 1.475 | 1.4751 | 7.716 | 0.00896 |
| Residuals33 | 6.309 | 0.1912 | | | |

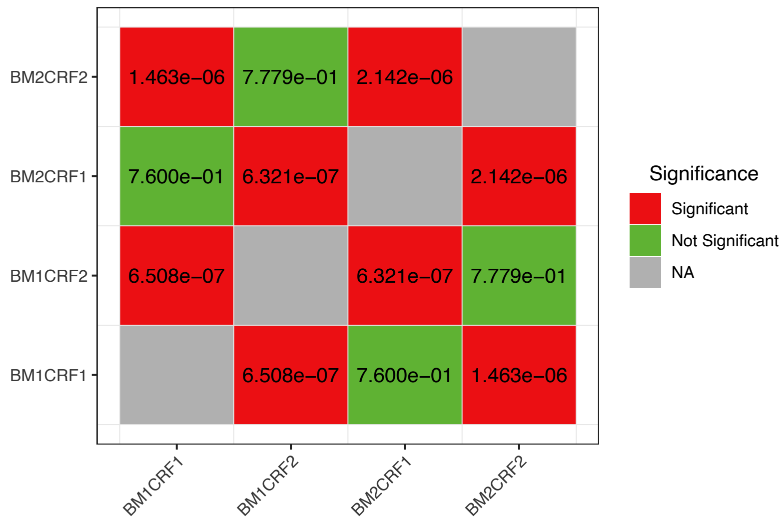
ANOVA for Time of Task Completion

| | BM1CRF1 | BM2CRF2 |
|---|---|---|
| p-value | 0.005749 | 0.3099 |
| W-statistic | 0.84586 | 0.753664 |
| p-value (log) | 0.197 | 0.009232 |
| W-statistic (log) | 0.753304 | 0.83838 |

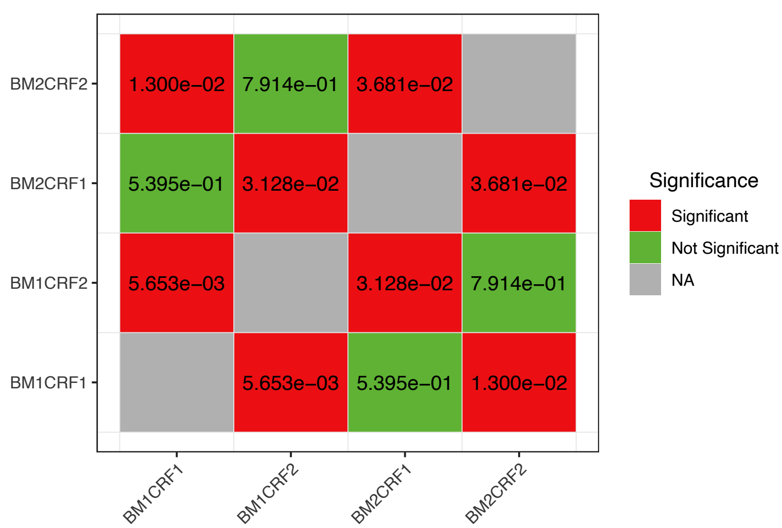
Shapiro-Wilk Test for (Log-)Normal Distribution of Estimations

# H Statistical Figures

## Task 2




Matrix for Mann-Whitney U Test for Duration of Task Completion




Matrix for Mann-Whitney U Test for Fixation Count

# I Personal Declaration

I hereby declare that the submitted Thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the Thesis.