International Cartographic Association
Association Cartographique Internationale

The visual complexity of bike maps

Donat Büchel [a], Tumasch Reichenbacher [a,*]

[a] *Department of Geography, University of Zurich, Switzerland, donat.buechel@uzh.ch, tumasch.reichenbacher@uzh.ch*

* Corresponding author

Abstract: More and more cities try to encourage residents to cycle more. Therefore, governments are developing comprehensive bike maps to facilitate trip planning and increase the popularity of cycling. 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. To understand how base maps and the display of cycling related features affect the visual complexity of bike maps and thus their effectiveness, we used different metrics (GMLMT, Subband Entropy, Edge Density, Feature Congestion, and Distinct Object-Type Counts) on four bike maps with four different visual complexity levels. We ran an eye-tracking experiment with 35 participants solving four different everyday tasks with these four bike maps. 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 colour were found to have the biggest influence on the applied metrics. The analysis of eye-tracking data revealed that the display of cycling related features can affect the time needed for successfully completing a task. To deepen the gained understanding, further research should in more detail investigate how base maps influence bike maps efficiency.

Keywords: map complexity, visual complexity, bike maps, eye tracking

1. Introduction

In the course of mitigating traffic problems, reducing CO2 emissions, climate change impacts, noise immissions, etc. cities worldwide increasingly promote cycling as means of transport and invest in respective infrastructure. Compared to other means of private transportation cycling has many advantages. It causes neither noise nor pollution and requires significantly fewer public resources and space. The energy required for cycling is provided by the traveller, resulting in additional health benefits for the traveller. Cycling is also much cheaper than a private car or public transportation. In sum, it is the most environmentally, socially, and economically sustainable means of transportation (Pucher and Buehler, 2008). Different aspects can help promote bicycling, such as direct routes, the presence of bicycle facilities, road safety, and others (Rybarczyk, 2014). Additionally, cities publish dedicated bike maps that emphasize features of particular importance to cyclists. The design of bike maps shows great variation in the number of shown map elements and symbology. While some maps display all streets, others depict only major routes. Furthermore, urban cyclists are not a homogeneous group and show large differences in terms of abilities, destinations, purposes, and needs. While some use their bicycles to commute to work, others prefer to ride in their leisure time (Rybarczyk, 2014). Different types of cyclists may need specific information and hence are interested in different aspects of a bike map, such as the availability of bicycle lanes, paved or unpaved streets, one-way streets, dangerous crossings, terrain, bicycle service stations, or pumping stations. This versatility of possible features makes bike maps more visually complex than other maps. With this paper, we aim at studying the visual complexity of bike maps, specifically from a perceptual angle. Our objective is to contribute to an improvement of the overall design and efficiency (i.e., the speed with which a task is completed successfully) of bike maps. Thus, we show the results of an eye-tracking experiment to answer the following research questions, where 1a) and 2a), as well as 1b) and 2b), are related:

1a) How visually complex are different bicycle base maps?
1b) How does visual complexity of base maps affect the 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 the efficiency of bike maps?

From these research questions, we formulate following hypotheses:
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

Bike maps differ from topographic maps or general city maps, by deliberaltly depicting features that or important to cyclists, such as bikeabilioty, dedicated bike lanes, terrain steepness, bike parking opprtunisties, etc. While

Proceedings of the International Cartographic Association, 5, 3, 2023.
31st International Cartographic Conference (ICC 2023), 13–18 August 2023, Cape Town, South Africa. This contribution underwent single-blind peer review based on submitted abstracts. https://doi.org/10.5194/ica-proc-5-3-2023

such bike maps exist, Wessel and Widener (2015) criticise that most bike maps issued by city governments assume a typical cyclist. However, there is no such thing as a typical cyclist. The purpose of cycling, skills, and needs differ substantially. Since cycling experience is a major factor for the suitability of routes for different cyclists, Dill and McNeil (2016) propose four categories of cyclists. *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 steep terrain, inability, or a lack of interest. For the city of Portland, Oregon the inhabitants were categorised 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).

Wessel and Widener (2015) propose a design for a bike map for the city of Cincinnati from a cyclist's perspective, the Cinncinati Bike Map. The roadways should be presented in such a way that cyclists can see the possibility of friction with cars. Included is the speed indicated by colours, the width of streets, bike lanes, and elevation. Also included is additional useful information such as track signals, water, and bicycle shops. This approach on a bike map is novel. However, many map readers did not examine the legend and just guessed the meaning of roads based on the colours. Since the map symbology is not intuitive, the interpretation was frequently inaccurate. For example, some readers thought that colours mean good or bad roads (Wessel and Widener, 2015).

2.2 Map Complexity

Map complexity has been studied since the 1970s to grasp the process of map-reading, find out what makes maps difficult to read, and ultimately improve their design and effectiveness (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, Bektaş and Çöltekin, 2018). Despite the different perspectives, a consensus on two major categories has emerged: visual (or graphic) and intellectual complexity (MacEachren, 1982). According to Ciołkosz-Styk and Styk (2011), the two complexity aspects correspond to two fundamental aspects of a map: syntactic and semantic. Visual complexity is concerned with the complexity of the map symbology, while intellectual complexity refers to the intrinsic complexity of the features or phenomena that are represented by the map. Visual complexity is determined by the degree of extensiveness, generalisation, and visual variable order (Ciołkosz-Styk and Styk, 2011).

(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. Alongside the development of metrics, user experiments using eye-tracking became an experimental approach to estimating map complexity (Barvir and Vit, 2021).

2.3 Approaches to Quantify Map Complexity

Methods to assess the complexity of maps with quantitative measures can be divided into two broad categories: object counting based and image processing based on raster pixels.

Counting the total number of objects in a map for quantifying map complexity is a basic approach. However, it depends on the definition of a single object, which is not so easy. However, determining what an object is, is not straightforward. I.e., it is unclear if a road as a whole or a certain segment of it count as an object. Harrie and Stigmar (2007) describe different measurements for evaluating map complexities, such as the number of objects, number of points in the objects, object line length, object length, the spatial distribution of objects, and spatial distribution of points. As an extension of the **Object Counting** approach distinct object-type count was introduced by Schnur, Bektaş and Çöltekin (2018). Complexity might be affected not only by the total number of individual items on a map but also by the number of distinct types or categories. 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, Bektaş and Çöltekin, 2018).

The first image processing method applied to quantifying map complexity was edge detection. Oliva et al. (2004) used this technique to calculate the **Edge Density**, i.e. the percentage of pixels in the map that are edge pixels. The assumption is that the higher the number of edge pixels, the higher the map complexity.

Rosenholtz, Li and Nakano (2007) propose **Subband Entropy** to measure spatial uniformity in an image as a proxy for complexity. This is based on the notion that when a picture becomes more crowded, the number of bits required for subband image coding will increase. It measures how hard it is to encode the information that is present in the image. As subbands of the image features like brightness, chrominance, colour, and edge orientation are taken into account (Speed *et al.*, 2017). As a measure for visual clutter Rosenholtz, Li and Nakano (2007) suggested **Feature Congestion**. Clutter is defined as ”...the state in which excess items, or their representation or organisation, lead to a degradation of performance at some task” (Rosenholtz *et al.*, 2005; Rosenholtz, Li and Nakano, 2007). When more and more items are added to a map, there is less space to add new items, leading to feature congestion. There are already too many colours, sizes, and shapes that make up a crowded area. Based on feature (co)variance computation and integration over scales an overall feature congestion value is output for an image.

Fairbairn (2006) suggested other measures, such as contrast weighted edge density, double log fractal dimension, landscape shape index, Simpson’s and Shannon’s Index, and others.

Following up on image-based complexity measures Barvir and Vit (2021) developed a **Graphic Map Load Measurement Tool (GMLMT)** to calculate a map-load value as a proxy for map complexity. As a first step, an edge detection filter (Sobel) is applied to the map extract and then the image is transformed to monochrome mode. Next, the image histogram is used for calculating the average pixel value of the monochromatic image. The value range is normalised into percentages (0 – 100%). Finally, the map-load level is computed as the average of the map’s current structures. For better illustration, also a grid is created that indicates which parts of the image have a higher load (bright tones) and which have a lower load (dark tones).

3. Methods

3.1 Participants

As bike maps should be useful for everyone, we had no exclusion criteria for our experiment. Participants were recruited at the University and from the personal network of the first author. Initially, 44 people signed up to take part in the experiment. Unfortunately, for 9 participants the eye-tracker could not record data reliably. As a result, this study includes data from 35 people participants between 21 and 34 years old (M=25; SD 2.3; F:11; M: 24). None of these participants indicated colour blindness. 67% of the participants are geography students, 8% are teachers, 25% have another academic background. All participants gave written consent to the experiment after being informed.

3.2 Experimental Design

Research questions 1a and 1b relate to base maps, whereas 2a and 2b focus on cycling related features. For this study, a two-by-two factorial design was adopted (Figure 1). The four cells will be abbreviated with BM1CRF1, BM1CRF2, BM2CRF1, and BM2CRF2. BM relates to the base map, while CRF relates to cycling related features. The numerals 1 and 2 represent the degree of complexity. 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).

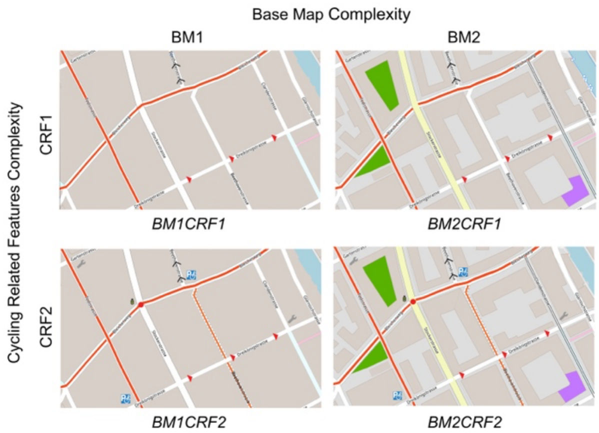

Figure 1. 2x2 factorial design of the experiment with two complexity levels for base map and bike related features.

The independent variables ‘visual map complexity of the base maps’ and the ‘complexity of cycling related features’ have two levels each (see Figure 1 for an example). The dependent variables are eye-tracking metrics (time to first fixation, fixation count, and task solving time). Age, gender, prior familiarity with bike maps, and maps in general are controlled in the experiment. To obtain this data a questionnaire was created using LimeSurvey[1].

3.3 Materials

To reach a higher ecological validity with our experiment, we decided to use a real city with an existing bike map rather than use simple design mock-ups. However, to address the possibility of familiarity, we chose the city of Nashville, Tennessee that is likely unknown to our study group. The city has sufficient bicycle infrastructure to allow the creation of more complex maps and is also very likely to be unfamiliar to the participants of the experiment.

To inform the stimuli design, we looked at bike maps from twelve of the top twenty cities listed by the Copenhagenize Index, which assesses cities worldwide in terms of their bicycle friendliness (Zayed, 2016). For stimuli creation, the Nashville bike map was imported to Affinity Designer and modified to produce our testing map set.

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): streets, highways, parks, water bodies, street labels, and park labels (Figure 2).

The base map of higher complexity (BM2) displays the same features as the low complexity base map, but additionally includes hospitals and buildings from OpenStreetMap, and a higher number of labels (Figure 3).

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.

[1] https://www.limesurvey.org

 https://doi.org/10.5194/ica-proc-5-3-2023 |

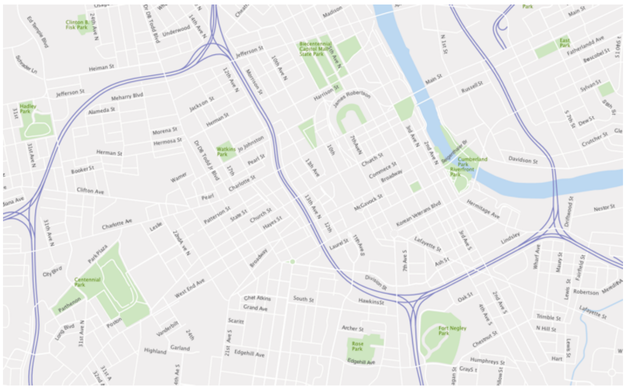

Figure 2. Low complexity base map.

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 coloured. 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.

Figure 3. High complexity base map.

A plethora of cycling related features could be depicted on a bike map. The most common were undoubtedly cycle tracks, bicycle lanes, shared lanes, bicycle stations, pedestrian areas, and repair shops. The decision on which features to display and what term to use is based on the Nashville bike map. For the lower complexity (CRF1) following cycling related features were considered: bicycle racks, physically protected bicycle lanes, bicycle lanes, bicycle routes, and non-cycling roads (Figure 4).

The higher complexity level of cycling related features (CRF2) comprises additional features, i.e., pumping stations, railway crossings, bicycle signs, bicycle rental shops, main bicycle routes, off-street bicycle routes, and easy-riding zones (Figure 5). Moreover, the number of symbols overall is higher than in CRF1. Second, the bicycle paths have been further separated into two categories.

As the four stimuli should be a well-balanced set of complexities, we calculated the image-based complexity measures and performed a distinct object count. This allowed us to modify the stimuli in such a way that the order of complexity is preserved and known.

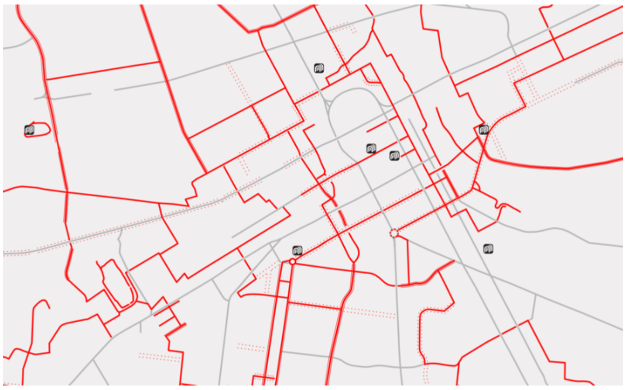

Figure 4. Low complexity cycling related features layer.

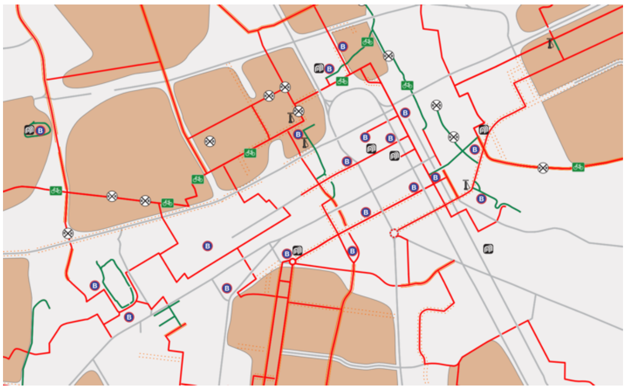

Figure 5. High complexity cycling related features layer.

We first calculated GMLMT, Features Congestion (FC), Subband Entropy (SE), and Edge Density (ED) for all four layers. These are metrics proposed to measure complexity. However, GMLMT, FC, and ED are based on the same principle of measuring how many edges are within a picture. Thus, having a high congruence when applied. SE measures the bits required to encode an image, taking different subbands (brightness, chrominance, colour, and edge orientation) into account.

For every measure, BM1 gets lower values than BM2, and CRF1 gets lower values than CRF2. Furthermore, base map layers are more complex than cycling related feature layers (BM1 > CRF1, and BM2 > CRF2).

The quantification for all four stimuli is shown in Table 1. For every measure, BM1CRF1 > BM1CRF2 & BM2CRF1 > BM2CRF2. As a result, the overall goal of balancing was achieved. The goal was for BM1CRF2 and BM2CRF1 to be as similar as possible. This was not possible because base maps have a higher complexity than cycling related features. Another essential aspect was to make sure that the spacing between BM1CRF1 and BM1CRF2|BM2CRF1 is close to BM1CRF2|BM2CRF1 and BM2CRF2.

Stimuli	GMLMT	FC	SE	ED
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 1: Quantification of stimuli

31st International Cartographic Conference (ICC 2023), 13–18 August 2023, Cape Town, South Africa. This contribution underwent single-blind peer review based on submitted abstracts. https://doi.org/10.5194/ica-proc-5-3-2023

We also performed a distinct object-type count on the produced stimuli. 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 show 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.

Some of the stimuli had to be altered for the tasks (see section 3.4). The task-specific stimuli were also quantified to ensure that the different metrics did not change too significantly. Although for some tasks the values of the measures for the individual stimuli vary, they are still valid and follow the main order of complexity.

3.4 Experiment Tasks

All participants had to complete four tasks in the experiment. The experiment follows a within-subject design, meaning that all the participants see the same stimuli.

Task 1 was to spot the you-are-here (YAH) symbol and then click on it when they have found it. Eight bike maps (two for each factorial design cell) with randomly placed YAH symbols were shown to the participants.

In **Task 2**, participants had to look for and count bicycle racks randomly distributed on a bike map. In total, four maps are presented, two displaying six and two depicting seven bicycle racks. Depending on the distribution, the visual map complexity changes, which was tested before utilising the stimulus. On finishing the task, participants had to click on the screen and select the counted number from a scale.

In **Task 3**, participants had to search for a specific park on the map. 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 each of the four cells of the factorial grid, four varied maps were created. To avoid a learning effect, the names of the parks to detect were randomised and the names of ten labelled parks always changed. For each cell, only one of the four crated maps was randomly picked and presented to the participant.

For **Task 4,** four different maps with varying degrees of complexity were shown to participants, including a start and end point and a legend for the various bicycle infrastructure displayed on the map. 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. Two routes with almost equal distance and complexity were chosen and start and end points were swapped and assigned to two of the four maps.

3.5 Procedure

The experiment took place in an eye-tracking lab at the University of Zurich. Stimuli were presented to participants on a 23-inch screen with a resolution of 1920 x 1080 pixels. Participants' gaze was recorded with the binocular Tobii TX300 eye tracker at a sampling rate of 300Hz. We decided to run the experiment in the lab, for safety reasons and to control for varying environmental conditions. While a field study would be closer to real world scenarios and yield higher ecological validity, changing conditions between participants would also lead to a bias.

A pilot test was run before the main experiment to ensure that tasks are clear and understandable and stimuli work. This allowed us to make minor changes to the experiment design and the stimuli.

The main experiment started with an introduction to the experiment objective and the signing of the consent form. Next, the participant's gaze was calibrated with the eye tracker. Then, a short training run for all four tasks was conducted to ensure participants understand the tasks. There was only one task per map to complete and the displayed maps were small extracts of a fictional map. The main data collection started, and participants looked at a total of 20 maps in 4 tasks.

At the end of the experiment, participants remained seating and filled in the questionnaire in a browser tab.

4. Results

4.1 Eye-Tracking Data

To analyse the collected eye-tracking data statistically, we defined areas of interest (AOI) around the symbols, legends, etc. in the different stimuli. Analysis of variance is utilized to test for variance in the gathered data. Before conducting ANOVA, tests for normal distribution (Shapiro-Wilk test) and homoscedasticity (Levene's test) must be conducted. If ANOVA reveals significant differences Tukey HSD shows where differences within the groups exist. When the data is not (log-)normally distributed, a non-parametric test (Mann-Whitney U) is used. All tests are conducted with a significance level of 0.05.

4.1.1 *Task 1: Locate the YAH Symbol*

The time to first fixation (TTFF) data aggregated for the sub-tasks shows that BM2CRF1 has the lowest average value, followed by BM1CRF2 and BM1CRF1, which have very similar values. BM2CRF2 has the highest values.

A Mann-Whitney U test shows significant differences in TTFF between the stimuli, except for BM1CRF1 and BM1CRF2 which show very similar values. It is worth noting, however, that BM2CRF1 is significantly smaller than BM1CRF1, which was unexpected.

4.1.2 *Task 2: Count Bicycle Racks*

The average time to first fixation on the bicycle racks is the shortest for BM1CRF1, 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. This order is also manifested for the average time needed to fixate the next bicycle rack.

An ANOVA reveals that there are significant differences between some stimuli ($p < 2e-16$). The Tukey HSD shows that BM1CRF1 and BM2CRF1 are not significantly different, just like BM1CRF2 and BM2CRF2. All the other combinations are significantly different from each other.

The time needed for task completion shows a similar order for the stimuli as for TTFF. A Mann-Whitney U test confirms that BM1CRF1 and BM2CRF1 are significantly different from BM1CRF2 and BM2CRF2. Differences between BM1CRF1 and BM2CRF1, like BM1CRF2 and BM2CRF2, are not significant.

A Mann-Whitney U test for the number of fixations, i.e. how often do the participants look on average on a bicycle rack, before moving on, shows the same results as for the task completion duration.

For every stimulus, the majority counted the objects correctly. All participants (100%) gave the correct answer for BM1CRF1 (with seven bicycle racks). The second greatest accuracy has BM2CRF1 (with six bicycle racks) (94%), followed by BM1CRF2 (with six bicycle racks) (91%) and the lowest accuracy (74%) of counted bicycle racks has BM2CRF2 (with seven bicycle racks).

4.1.3 *Task 3: Locate a Park*

Although the average time to fixate a next park is longest for BM1CRF2 an ANOVA revealed no significant differences between the four stimuli. For the time needed to complete the task, a Mann-Whitney U test also shows no significant differences between the four stimuli.

4.1.4 *Task 4: Search for a Route*

For stimuli BM1CRF1 and BM2CRF2 A and B had the same position but were inverted. BM1CRF2 and BM2CRF1 were treated in the same way. In all stimuli, the legend was fixated often.

TTFF on point A is highly similar within the stimuli. BM1CRF1 and BM2CRF1, the maps with a low number of symbols, have lower TTFF values than BM1CRF2 and BM2CRF2. The same appears to be true for symbol B.

A Mann-Whitney U test is applied to both datasets. For the start symbol, no significance was found. For the second symbol, significant differences could be found. Table 2 shows the calculated values for the Mann-Whitney U test.

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

Table 2: Mann-Whitney U Test for TTFF for the Second Symbol

It was expected that the participants would take the least amount of time to complete the task in BM1CRF1, but instead, BM2CRF1 took them the least amount of time, followed by BM1CRF2. For the task completion duration, a Mann-Whitney U test revealed that BM1CRF1 is significantly different from BM2CRF1 and BM2CRF2. BM2CRF2 significantly differs from the other three stimuli.

5. Discussion

We formulated two hypotheses with respect to the visual complexity of base maps and cycling related features:

H1a) More detailed base maps are visually more complex.

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

From the quantitative analysis of the experiment's base maps, it can be concluded that the detail of base maps and visual complexity are positively correlated. The same is true for showing more cycling related features. For all measures GMLMT, FC, and ED, the computed complexity values were higher for the more detailed base map and the layer displaying more cycling related features.

This finding is intriguing since only two items are added to the more detailed base map: hospitals and buildings. 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. Adding six additional objects to CRF2 compared CRF1, seems to have a smaller effect on the visual complexity than adding more buildings to the base map. Pedestrian areas and easy-riding zones are on the one hand widely distributed, resulting in a bigger impact on complexity than other features. On the other hand, they have a smaller impact than buildings, since they typically have fewer edges.

Another important factor to consider is contrast. High contrasts (e.g., between dark buildings and white background) create stronger edges that then influence the complexity measure. Reducing contrast may influence complexity. For our base maps we applied some transparency to the buildings to reduce contrast. We could observe in some cases, that the depiction of zones can minimise complexity, when it is represented with a colour leading to less contrast. However, we suspect that the lower contrast in our stimuli caused some participants to perceive this zone worse.

Overall, both hypotheses H1a) and H2a) can be accepted. A more detailed base map or adding more cycling related features leads to a visually more complex cycling map. When the cycling related characteristics are integrated with the base map, however, it's vital to remember that zones might minimise visual complexity.

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 hypotheses address the efficiency of bike maps:

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

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

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.

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 centre. In both BM1CRF1 and BM2CRF2, the symbols were about the same distance from the centre. The symbols in BM2CRF2 are located in areas where map information is rather low. As a result, the high values for stimuli BM2CRF2 can also be explained by the position of the symbols. 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 when creating maps. Selecting an extract with the YAH symbol in the centre of the map can help users avoid searching for a long time.

Task 2: Count Bicycle Racks

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. 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.

The response 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

This task was meant to address the various levels of base map complexities. However, despite the pilot testing and randomisation efforts, participants recognised where the parks are located 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 measured 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.

Task 4: Search for a Route

The assumption for this task is that symbols of less complex maps are detected faster and the time needed for task completion is shorter.

TTFF 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.

Both hypotheses H1b) and H2b) cannot be accepted. An influence of map complexity on efficiency could only be found partially, but without clear significance. Although the complexity of cycling related features had a greater influence on efficiency, we got mixed results. Reasons for these findings may be inappropriate tasks and possible learning effects (task 3).

Although we got insightful results, our work has some limitations. The experiment is primarily set up as a within-subject design, which may have introduced a learning effect. Participants may be able to learn from presented maps, and hence possibly perform better after seeing the stimuli numerous times. Overall, we could not observe a major learning effect. Only in task 3, where the names of the parks change, but the park areas remain the same, we could observe a possible learning effect from the outcome. We randomised the order of stimuli within a task. It would have been even better to randomise the stimuli even across tasks. However, this would lead to major confusion for the participants, having to solve alternating tasks.

Another issue arises from the position of the symbols in tasks 1, 2, and 4. Symbols placed in the centre of the map are likely to be recognised 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 centre of the screen.

Overall, our sample size is adequate, but the sample is rather unbalanced with respect to gender, age, background, and bike use. A minority of the participants are female (11 female and 24 male) and the age range is skewed towards young people. With most participants having a geography background, their background is different from the broader public. 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).

We used static bike maps in our experiment for better control. Likewise, we conducted the experiment in a lab. Future studies in the field with interactive bike maps on smartphone screens could bring more ecological validity and would be beneficial.

6. Conclusions

To foster cycling in urban environments, an increasing number of bike maps are being created. Due to lacking design guidelines for bike map approaches are very diverse. 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 analysed. 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 colour 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 visual complexity. The colour contrast is crucial, 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 an eye-tracking experiment with 35 participants. Four tasks that were similar to those faced by cyclists on a daily basis had to be completed.

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, counting 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 cannot be given.

For the complexity of cycling related features, the 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 randomised, 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.

7. References

Barvir, R. and Vit, V. (2021). Graphic Map Load Measuring Tool – development and verification, *International Journal of Cartography*, 7(3), pp. 285–303.

Barvir, R. and Vozenilek, V. (2020). Developing Versatile Graphic Map Load Metrics, *ISPRS International Journal of Geo-Information*, 9(12), p. 705.

Castner, H. and Eastman, J. (1984). Eye-Movement Parameters and Perceived Map Compexity – I, *The American Cartographer*, 11(2), pp. 107–117.

Ciołkosz-Styk, A. and Styk, A. (2011). Measuring maps graphical density via digital image processing method on the example of city maps, *Geoionformation Issues*, 3(1), pp. 61–76.

Dill, J. and McNeil, N. (2016). Revisiting the Four Types of Cyclists: Findings from a National Survey, *Transportation Research Record*, 2587(1), pp. 90–99.

Fairbairn, D. (2006). Measuring Map Complexity, *The Cartographic Journal*, 43(3), pp. 224–238.

Harrie, L. and Stigmar, H. (2007). An evaluation of measures for quantifying complexity of a map, *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(3), pp. 266–274.

MacEachren, A. (1982). Map complexity: comparison and measurement, *American Cartographer*, 9, pp. 31–46.

Oliva, A. *et al.* (2004). Identifying the Perceptual Dimensions of Visual Complexity of Scenes, in *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, pp. 1041–1046.

Pucher, J. and Buehler, R. (2008). Making Cycling Irresistible: Lessons from The Netherlands, Denmark and Germany, *Transport Reviews*, 28(4), pp. 495–528.

Rosenholtz, R. *et al.* (2005). Feature Congestion: A Measure of Display Clutter, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (CHI 05), pp. 761–770.

Rosenholtz, R., Li, Y. and Nakano, L. (2007). Measuring visual clutter, *Journal of Vision*, 7(2), pp. 1–22.

Rybarczyk, G. (2014). Simulating bicycle wayfinding mechanisms in an urban environment, *Urban, Planning and Transport Research*, 2(1), pp. 89–104.

Schnur, S., Bektaş, K. and Çöltekin, A. (2018). Measured and perceived visual complexity: a comparative study among three online map providers, *Cartography and Geographic Information Science*, 45(3), pp. 238–254.

Speed, A. *et al.* (2017). *Applying Image Clutter Metrics to Domain-Specific Expert Visual Search*.

Wessel, N. and Widener, M. (2015). Rethinking the Urban Bike Map for the 21st Century, *Cartographic Perspectives*, 0(81), pp. 6–22.

Zayed, M.A. (2016). Towards an index of city readiness for cycling, *Unmanned Aerial Vehicles and Remote Sensing*, 5(3), pp. 210–225.

