Aalto University
School of Science
Master's Programme in Computer, Communication and Information Sciences

Ninh Truong

Parking fines, and where to find them

An interactive visualization of parking violations in Helsinki

Master's Thesis
Helsinki, May 23, 2020

Supervisor: Professor Petri Vuorimaa, Aalto University

Aalto University
School of Science
Master's Programme in Computer, Communication and Information Sciences

ABSTRACT OF MASTER'S THESIS

Author:	Ninh Truong
Title:	Parking fines, and where to find them - An interactive visualization of parking violations in Helsinki
Date:	May 23, 2020
Pages:	88
Major:	Computer Science
Code:	SCI3042
Supervisor:	Professor Petri Vuorimaa

Visualization is proven to be an effective way to communicate data to the readers. In Helsinki, open data on parking violation is available and updated frequently. However, the prior attempts on the data set, which are open to the public, do not meet the expectation with the up-to-date information.

This thesis first reviews and summarizes the knowledge on the field of information visualization. Inspired by prior works on the process of creating a computer generated visualization, a model is adapted which include both technical aspects and human participation.

Following the process, the parking violation data in Helsinki from 2018 to 2019 is handled through *pandas* library, then represented as an interactive visualization on the web using *D3JS* and *Mapbox GL JS*. With the intention of making public data meaningful to each individual reader, a feature named "explorer" is built, allowing the analysis of the data in a radius around a specific location.

Through a series of optimization and evaluation with the focus on rendering performance and insight creation, the work demonstrates the capability to combine with the parking experiences of its users to bring excitement and create new knowledge.

Keywords:	parking violation, information visualization, geovisualization, personalized visualization, web visualization
Language:	English

Acknowledgements

A typical workday, for me, can fall into one of three types: have nothing done, finish a random portion of work in a random time, or complete the workload of the whole week in a couple of hours. Lacking the ability to focus, I cannot imagine that I can have the thesis done without the support of the following people:

Professor Petri Vuorimaa for his lightning-fast replies and his guidance during the work, especially when I run into a mental problem, which I did not think that I can proceed. HCID programme lead Mika Nieminen for his advice and patience to a four-year Master student like me. My colleagues at work Jarkko, Juuso, Katriina, Matias, Per and Teemu, as well as my friends Ha, Thanh, Trang, Son and Nguyen for helping me in various way, from encouraging, translating, throwing ideas, to spending time for the usability test. My partner Thuy Ngo for being the mood booster, sharing the ups and downs of the long and hard thesis process. My mom and dad for their constant pressure and the belief that their son will certainly make it.

And finally, for my grandma.

Helsinki, May 23, 2020

Ninh Truong

Abbreviations and Acronyms

HCI	Human-Computer Interaction
GIS	Geographic information System
EDA	Exploratory data analysis
CRS	Coordinates reference system
SRS	Spatial reference system
GCS	Geographic coordinate system
PCS	Projected coordinate system
EPSG	European Petroleum Survey Group
Pedshed	Pedestrian shed
HTML	HyperText Markup Language
CSS	Cascading Style Sheet
JS	JavaScript
SVG	Scalable Vector Graphics
CSV	Comma-separated values

Contents

Chapter 1

Introduction

Nobody enjoys parking fines. For the drivers, money is taken out of their pockets, or their cars might be temporarily banned from traffic. The effect is even worse on the "victims". Illegal parkings on the spots for the disabled make their lives much harder. Leaving the car blocking the bus stop delays the regular routines of the bus, along with all the people travelling with it. Parking exceeding the time allowed prevents other drivers to find a space, indirectly results in traffic congestion.

The city of Helsinki, in the effort of raising the knowledge of its citizens and encouraging the utilization of technology, opens the regional data for public usage at no cost. Parking violations data are among the available resources. Unfortunately, reading a text file containing millions of records is not an attractive activity. This matter raises the demand for applications, which communicate the data better.

One effective method is turning the data into an interactive visualization - the graphical form, which people can interact with.

Around the world, the idea is popular in urban planning in general. In New York, visualizations are used to explore the living quality of the, such as air pollution[1], or tree coverage[2]. Efforts from Dominic Royé in *Visualize urban growth*[3] and the Morphocode team in *Urban layers*[4] are examples of transforming the history of Valencia and Manhattan into colorful and insightful images. The two projects map the built age of each building with a corresponding color, and the users can choose different periods to examine

[1] `http://senseable.mit.edu/urban-exposures/`
[2] `https://www.coord.com/blog/tree-coverage-on-streets-from-open-data-in-nyc`
[3] `https://dominicroye.github.io/en/2019/visualize-urban-growth/`
[4] `http://io.morphocode.com/urban-layers/`

how the city develops through the flow of time. Representing the home and work presence of the population throughout 24 hours of the day as an animated visualization, the *Breathing city*[5] project creates an impression that a city is also a living being.

In the field of parking in particular, the work from Jimmy Steinmetz and Maureen O'Donnell[6] allows the user to analyze the parking violation in Manhattan by selecting the neighborhood, vehicle type, violation reasons, and time of day. In 2014, Matt Chapman experiences the frustration of receiving a parking ticket in Chicago, which motivates him to make his own visualization[7], from where he manages to draw a number of notable and interesting patterns around the city.

Presented and distributed on the web, the mentioned projects have reached a wide audience and proven their usefulness.

The Helsinki parking violation data set has a page to showcase its applications[8]. However, the works there are either inaccessible, or no longer maintained. All available are screenshots or demonstration videos. The knowledge about parking violations is not learnt once and valid forever. Therefore, the tool representing it needs to be updated with the latest data, as the policies change from time to time. Besides, these projects are available in Finnish only, which is a barrier for foreign people living in the city, who are responsible for a significant amount of parking violations.

This raises a challenge in visualizing public data. The large number of records and the need of frequent updates requires a comprehensive procedure to process the raw data. The goal is to have the hard work done once, and easing the maintenance effort in the future.

Illegal parking is the result of poor communication. This might come from an ill-placed parking sign from the city, the rocket-science instructions of the parking lot owner, or from drivers who do not know they are parking against the rules. If people have more awareness about the parking violations around them, will the number of fines reduce? The work of this thesis is motivated by having the mentioned doubt in mind.

---

[5]https://www.darkhorseanalytics.com/blog/breathing-city

[6]https://interworks.com/blog/modonnell/2016/03/18/dont-get-ticketed-nyc-parking-violations-data-viz/

[7]https://mchap.io/parking-ticket-visualization-in-chicago.html

[8]https://hri.fi/data/en_GB/dataset/showcases/pysakointivirheet-helsingissa

1.1 Research questions

1.1.1 Visualization challenges

How to visualize massive amounts of data, but at the same time, preserve visual quality and acceptable system performance? The city's parking data set contains geolocation information for each violation. Thus, maps visualization is considered a natural way of representation. However, with about 150 000 records per year, visual cluttering is inevitable. The huge number of data mentioned also cause a detrimental impact on the rendering speed. In web, slow is dangerous, since the average human attention span is just 8 seconds [16]. On desktop, with only 3 second delay, a website could lose 21% of its visitors leaving in frustration [60].

To answer this question, this project utilizes different technologies and methods. Pandas - a Python library is used for combining, aggregating and filtering text as well as numerical data. The violation and the supporting charts are generated by D3JS, into a combination of SVG and HTML Canvas to keep the rendering process catching up the metrics.

1.1.2 Personalize visualization

How can visualization help users gain insights, which are meaningful to their personal interests? This question is raised with the assumption that people are not aware about parking violations around them well.

To answer this question, the visualization prototype includes an element named "Explore mode", which allow an individual to examine the parking violation data within 350m, which is equivalent to 5-minute-walk radius around a location of choice. The study will approach a number of people who drive in the city, then make a comparison between their estimation and the actual data.

1.2 Thesis structure

"Theoretical background" chapter introduces theories about parking violation, and information visualization. One of its branches - geographic visualization - will be discussed in depth. This chapter also clarifies the idea of personalizing a visualization - the action of making data meaningful to every single user. Chapter 2 finishes with a brief explanation of how visualization works on the Web.

Next chapter - "Framework" - discusses two models of visualization workflow from Fry (2004) and Munzner (2009), then adapts a model which is compatible with the context of this thesis.

Based on the knowledge above, "Implementation" chapter describes the working process of transforming raw city parking violation data into an interactive map visualization tool.

Later, the usability of the tool are evaluated in "Evaluation" chapter, focusing on technical performance and insight creation.

The last chapter "Conclusions" summarizes the findings of this thesis, analyzes the limitations, and discusses the opportunities for further research.

Chapter 2

Theoretical background

This chapter starts with the introduction of the backgrounds behind the domain - parking violation and the technique - information visualization. Then, several related theories which motivate and inspire the work of this thesis including geographic visualization, personalizing visualization and visualization on the Web, will be discussed.

2.1 Parking, enforcement and violation

"Parking", as defined in The Cambridge dictionary is the act of "leaving a vehicle in a particular place for a period of time" [11]. In the context of this thesis, the terms "parking" is restricted to cars, rather than other kinds of vehicles. Young (1990) points out that a car is not always in use, so parking is an inevitable need [82]. Shoup (1997) adds a specific number: cars spend on average 95 percent of their lifetime not moving [65]. With the mentioned high demand, parking policy and enforcement play an important role in urban management [17].

2.1.1 The need of parking enforcement

When the first automobile run by gasoline was sold in America in 1896 [28], cars were parked at the tethering spots of horses and carriages for free [65]. Nearly 20 years later, in 1913, the number of cars in the world grew rapidly to more than 600,000 units [28]. Parking at that time was still free, and its demand became higher than the land could supply.

Shoup (1997) gives two arguments on how free parking is costly in urban planning [65]. First, minimum requirement makes drivers take parking for granted, so they leave their cars longer than needed. As a result, other people driving around at lower speed than normal, struggled to search for an available parking spot, and at the same time polluted the environment and delayed other traffic. Second, like other properties, parking spaces need maintenance too. How to cover this cost if drivers park for free? The answer lies in the hidden raised cost of the products involved, from higher apartment rent to more expensive movie tickets. Consequently, everyone, regardless of driving or not, has to pay for free parking indirectly.

To tackle the problem, urban planners introduced parking control, aiming to optimize the number of available spots in a parking location, thus to reduce traffic congestion and accidents [46]. Different methods of parking control include the limitation of space, time, and convenience. No-parking signs and off-street parking areas like parking lots or garages restrict drivers to park wherever they want (figure 2.1(a)). Maximum parking time and allowed parking hours during a day are specified in many areas, especially in large cities so drivers cannot park whenever and as long as they want (figure 2.1(b)). Instead of free, parking areas are encouraged to charge drivers (figure 2.1(c)). Optimal pricing strategy is already addressed in various researches [66] [1] [67].

(a) No-parking sign in New York

(b) Maximum parking time restriction in Helsinki

(c) Parking rates of a garage in London

Figure 2.1: Different types of parking restriction

In 1935, the Park-O-Meter No. 1 was installed in Oklahoma City as the first physical form of parking control [29]. The five-cent-per-hour at that time marked the beginning of the multi-billion parking industry today [68].

Cullinane and Polak (1992) summarized parking enforcement as the policies for the "allocation of roadspace between moving and parked vehicles" [17]. Having parking policies introduces a new job - parking warden - who make sure the rules are followed and the offenders are punished.

2.1.2 Parking violation

Parking violation is the act of parking against the rules. The rules, can be specified by the signs before the driver enters the parking spot (e.g., paid parking from 08:00 to 20:00, reserved parking spot for the disabled, etc.), or defined in the traffic law (e.g., cars must park towards the traffic direction on the street, no parking on the zebra crossing, etc.). When the parking control officer - normally parking warden or the police - notices a parking violation, commonly, a parking ticket will be issued and attached on the windshield of the violated car as Figure 2.2. The parking ticket includes several information like the time of inspection, the plate number of the car, the reason of violation, the fine amount and the payment instruction. The fine amount is different in each region, but it is much more expensive compared to the original parking price.

Figure 2.2: A parking ticket in Helsinki, Finland

The study from Thanh and Friedrich (2017) reveals the top motivations for people who violate the rules are: short walking distance from the parking space to their destination, they have no time to search for a spot, no parking charge on the chosen spot, and no patrolling officers to issue parking tickets. Having more knowledge about the area can trigger the attempt of illegal parking more easily [73]. O'Malley (2009) argues that the flat parking fine affects the rich and the poor differently [57]. Hence, for a number of wealthy people, the penalty cost is too small so they can just ignore it. When the violations happened accidentally, poor communication from the enforcers to the drivers, for example, missing or confusing parking signs, is the subject to blame [8] [44].

Money from parking fines goes to the city fund or the parking operator, which can be considered a source of revenue [13]. However, the effort of collecting this money is not always smooth. The rate of unpaid fine is high, especially in the case when the drivers cannot be identified [81]. Besides, illegal parking disrupts the normal traffic flow, hampers other vehicles (e.g., buses, fire trucks, etc.) from fulfilling their tasks [20], and involves directly in traffic accidents [26]. These costs are unmeasurable, and cause frustration on the drivers who violate the rules accidentally or intentionally, as well as the cities who have to deal with the consequences.

"Illegal parking is a pattern" [17]. Therefore, finding this pattern can help cut down the number of parking violations.

2.2 Information visualization

2.2.1 What visualization?

There are different terms related to visualization, which are extensively used and might cause confusion. They are "data visualization", "information visualization", and "knowledge visualization".

According to Chen et. al. (2008), "Data" is the representation of entities which is almost incomprehensible by humans, like the bits and bytes on the computer, or a log consists of codes and ids. "Information" is the result of interpreting and assigning meaning to the data. Now readable, humans can learn facts and draw conclusions from the information. That is knowledge. [14]

Masud et. al. (2010) presents six disciplines of visualization. Among them, "data visualization" and "information visualization" both help finding insight of the data set. The former is a broader term for graphical representation of data in every form, while the later focus on interactive, computer-supported visualizations. "Knowledge visualization" serves a different purpose: communicating and transferring knowledge [45].

The work related to this thesis applies web technology to generate and present visualizations from readable open data. Therefore, "information visualization" is the most suitable term, and will be used throughout this document for the sake of consistency.

2.2.2 Definition

The evolution of information visualization can be summarized from the work of Friendly (2010), from the one-dimension distance map of Langren in 1644, to the invention of basic graphs such as pie chart or bar chart by Playfair in the beginning of the 19th century, until the modern time where visualizations are interactive and generated by computer [24]. In its rich history, there are many different viewpoints on defining the term. The following section discusses several notable definitions.

> Information visualization: The use of computer-supported, interactive, visual representations of abstract data to amplify cognition

Above is a popular interpretation of the term, proposed by Card, Mackinlay and Shneiderman in 1999 [12]. The weakness in this definition is the limitation on the means of creation (computer-supported) and the requirement of the work (interactive). Classic hand-drawn, static visualizations are excluded. However, this definition shines in the current era, where a data set might contain millions of records, which is impossible to process manually.

Kosara (2007) proposes visualization definition in a process-oriented way. He specifies a visualization is the work "based on (non-visual) data, which "produces an image", with "the result is readable and recognizable". [39].

Other researchers approach the term in a goal-oriented way. Few (2013) considers visualization "the graphical display of abstract information for two purposes: sense-making (also called data analysis) and communication" [22]. Card et. al. (1999) also agrees that "The purpose of visualization is insight, not pictures" [12].

Even though there are various viewpoints on the definition of information visualization, they all revolve around one core idea: representing abstract data using graphics to support human understanding.

2.2.3 Why visualize?

The human brain can store and process a limited amount of information at the same time [48]. To release the burden of memorizing, people write down their to-do list, publish books to share knowledge, or log activities in a computer file in the form of text. For ages, text has been used as an efficient and precise way to store information. However, the brain capacity problem starts coming back when the amount of content to read and understand become larger. Some strategies to tackle this issue are: summary - use a short paragraph to explain the meaning of a long one; or statistical analysis - use numbers and models to represent the characteristics of a data set containing large amount of records.

If these strategies are effective enough, why visualize still needed? A classic example from Anscombe (1973) introduces four data sets sharing identical statistical values (Figure 2.3) [2]. However, when graphed, the differences between them can be quickly spotted (Figure 2.4).

Data set	1		2		3		4	
Variable	x	y	x	y	x	y	x	y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Sum	99.0	82.51	99.0	82.51	99.0	82.50	99.0	82.51

Figure 2.3: The data set in Anscombe's quartet

- Mean of the x's: 9.0
- Mean of the y's: 7.5
- Variance of x: 9.0
- Variance of y: 7.5

- Correlation coefficient: 0.816
- Equation of regression line: y = 3 + 0.5x

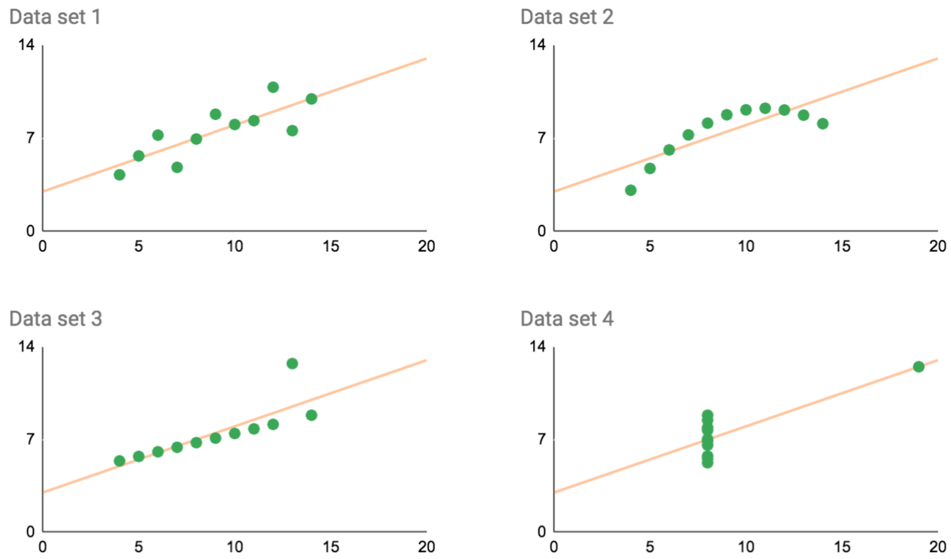


Figure 2.4: Anscombe's quartet visualized as scatter plot

Why can the insight be gained quickly and precisely through images? Research by Zimmerman (1989) estimates that every second, our vision system sends to our brain the amount of data approximately 8 times of other senses (hearing, touch, smell, taste) combined [83]. As a result, an overall impression is immediately formed, and the stand-out features are effortlessly detected. Few (2013) explains that the human brain is working based on two systems: seeing - fast, efficient but unconscious, and thinking - slow, inefficient, but takes part in most of the activities. Visualizations help balancing the brain usage towards the former one in order to utilize our powerful vision [22].

Despite having its advantages, visualization is only strong at showing certain aspects of the whole data. Therefore, it does not completely replace other means of data communication. Usually a visualization work is the mixture of graphics and text, where the graphic part supports quick understanding of the overall picture, while text or number provides the precise detail of the information.

The ultimate goal of visualization is still communication. Koponen and Hildén (2019) point out when it is possible to clearly explain the information by words (such as Figure 2.5), visualization is not necessary, or can even be misleading [38].

Figure 2.5: Just a simple sentence like "Common legal drinking ages in the Nordics is 18, except Denmark (16) and Iceland (20)" is enough to cover the content of this graph.

2.2.4 Visual encoding and variables

Records in the data set have characteristics, which are called variables [38]. For example, each parking has its start time or duration. Similarly, the attributes of a data point representing a record in the visualization are called visual variables. In a 2D graph consisting of circles, each circle can be at a different position and has different size. If a parking is shown as a circle, its distance to the y axis might reflect the start time, or the area of the circle could be used to describe parking duration. The act of mapping data structures into graphical structures as described is an example of visual encoding.

The effectiveness, or accuracy, of each visual encoding method is different, and depends on not only the variable it represent, but also the data scale. The data scale, or scale of measurement, is formed by possible values of a variable. Stevens (1946) proposed four types of scale: nominal, ordinal, interval and ratio [70]. Bertin (1983) shares similar ideas, but groups the scales into three different categories: qualitative, ordered, and quantitative (or numeric) [6]. Börner (2019) encapsulates the definition and example of each scale in figure 2.6 [9].

Data Scales	Logical Math Operations				Measure of Central Tendency	Examples
	= ≠	< >	+ -	x ÷		
Nominal	✓				mode	
Ordinal	✓	✓			median	
Interval	✓	✓	✓		arithmetic mean	0–6 | 7–12 | 13–18
Ratio	✓	✓	✓	✓	geometric mean	0 1 2 3

Figure 2.6: The summary of data scales

Bertin first introduces seven visual variables in 1983, including position, size, shape, value, color, orientation, and texture [6]. Inspired by his work, Cleveland and McGill (1984), Mackinlay (1986), Roth (2016), and Koponen and Hildén (2019) further develop the theory by adding more variables and verifying their effectiveness against different data scales [15] [43] [62] [38]. Figure 2.7 summarizes common variables and their order of accuracy based on the mentioned prior works.

A significant factor to visualization success is to choose the right variable to encode the data. Ideally, the most efficient method should be used to represent the most important characteristic. It is notable that occasionally the ideal choices are not available. For example, position (or location), in figure 2.7, proves its superior across all data scales. However in geographic visualization, position is inevitably taken to encode the physical coordinates of each data record. This situation also applies in the work of this thesis.

2.2.5 Types of visualization

In terms of communication purpose, a visualization based on data falls into one of these two categories: explanatory or exploratory [32] [38].

Explanatory visualization is commonly used to declare, explain and communicate the information. In this case, the creator already knows the facts behind the data. A decent amount of editorial effort is needed to select the focused data, and filter the irrelevant ones out. Explanatory visualizations are

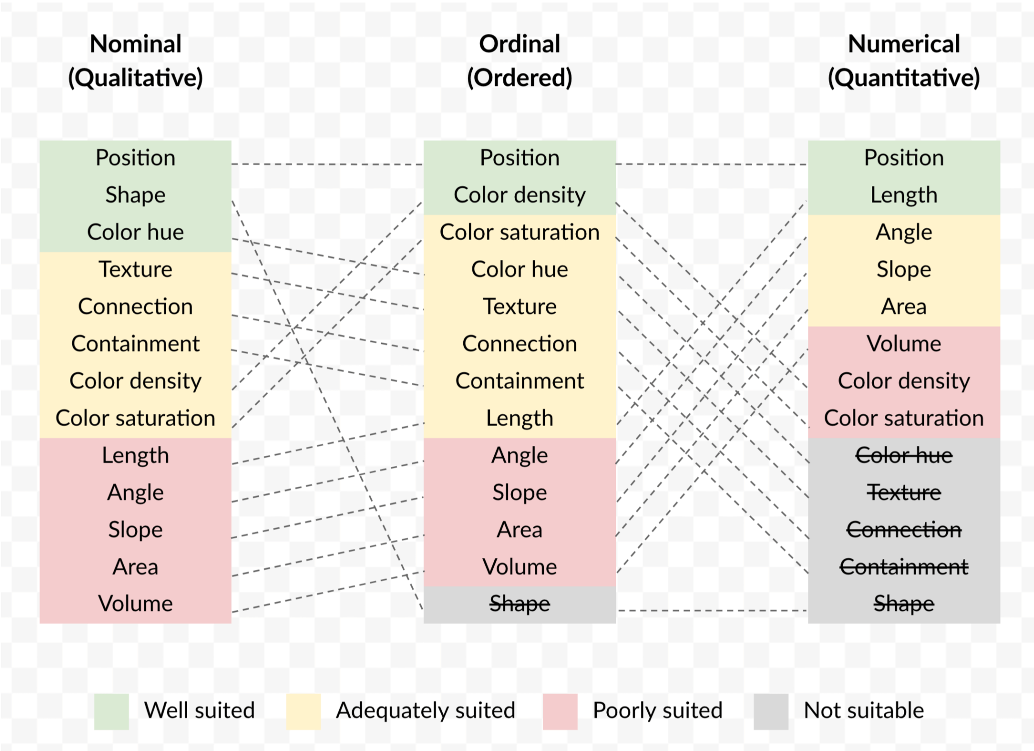


Figure 2.7: Visual variables and their order of effectiveness

commonly handcrafted, and usually accompany and support the information presentation phase.

On the other hand, **exploratory** visualization is suitable for information discovery. The creator in this case does not know what insight the final work might reveal. As filtering too much might lead to the missing of important information, the effort here is spent on making the exploration tool, instead of data selection. Visualizations fall into this category usually created with the aids of computers by using a set of rules to convert data into graphics. Exploratory visualization is widely used as a tool in the data analysis phase.

Besides, there are hybrid visualizations, where the designer curates a certain section of the data for communication, but also leaves chances for the reader to discover more insight from the other parts. These works are commonly presented in the form of interactive visualization, which will be discussed in the next section.

2.2.6 Interactive visualization

As its name implies, interactive visualization is dynamically presented based on the actions of readers. The graphics can be manipulated directly (by interacting with the graph itself using hand or computer mouse), or indirectly (via a user interface consisting of controls like buttons, sliders, etc.) [21] [71]. An action where humans perform against a system via a physical input/output device is called an interaction technique [23].

There are various researches working on collecting and classifying interaction techniques. According to Shneiderman (1996), the starting point of visualization interaction technique design is the visual information seeking Mantra: "overview first, zoom and filter, then details on demand" [64]. Yi et. al. (2007), through their comprehensive review of previous works, propose seven main groups to categorize interaction technique based on user intent [80]:

- Select (mark something as interesting) allows users to keep track of specific items by making them visually different from the others.
- Explore (show me something else) allows users to examine different portions of the whole data set.
- Reconfigure (show me a different arrangement) provides users the ability to change the order of items.
- Encode (show me a different representation) enables users to change the visual appearance of data items.
- Abstract/Elaborate (show me more or less detail) helps users adjust the amount of information displayed in the visualization. Techniques in this category are usually used in cooperation with Select techniques.
- Filter (show me something conditionally) allows users to show/hide items which match their defined conditions.
- Connect (show me related items) highlights the relationship between specified items, or presents hidden items which are relevant to the selected one.

In the computer-generated data era, the content being visualized becomes more complex and larger in scale. Therefore, with the limitations in the size of the displays (like paper or screen) and the information processing capability of humans, not all information is presented at the same time.

Consequently, readers "interact" with the actual static visualization by using a pen to highlight the data point, moving closer / further or rotating the paper itself [80]. Adding interactions to the image gives readers the tool to manipulate the graphics by their personal interest. This enables different viewpoints, thus generates more insights from the underlying data.

2.3 Geovisualization

Geographic visualization refers to a branch of information visualization, which supports the representation, exploration and analysis of data with location attached. Geovisualization works usually include a map interface, where readers can interact with the geospatial data behind [41]. This field enhances traditional cartography with modern computing technology. According to MacEachren and Kraak (2001), geovisualization is the combination of "scientific visualization, (exploratory) cartography, image analysis, information visualization, exploratory data analysis (EDA) and GIS to provide theory, methods and tools for the visual exploration, analysis, synthesis and presentation of geospatial data" [42].

Geovisualization has become a popular field in recent years. Within the same study, MacEachren and Kraak (2001) estimate 80% of digital data involve geospatial references such as position or address [42]. Nöllenburg (2007) summarizes three main reasons leading to the increasing interest in geovisualization [55]. First, the rapid development of technology makes graphics / display hardware cheaper but better. Second, the amount of data skyrocketed due to how effortless and cost-effective the collection process became. And third, the popularity of the Internet enables geovisualization to reach to the wider audience.

Besides sharing common knowledge with information visualization in general, geovisualization has its own special characteristics. It is possible for a geovisualization to produce wrong results, or to be unable to proceed if these points are left unnoticed. In the context of this thesis, two among them, namely "map projection" and "coordinates reference system" will be discussed.

2.3.1 Map projection

Map projection is the technique of "flattening" the Earth [69] in order to represent it on a two-dimensional surfaces like papers or computer screens.

Name	Azimuthal equidistant	Mercator
Method	Azimuthal	Cylindrical
Preserves	Distance from the central point	Shape & direction
Distorts	Shape. The further from the central point, the stronger the distortion is	Area & distance
Known usage	Polar area or small scale continent mapping	Nautical navigation, web map
Example	 United Nation includes the world map with azimuthal equidistant projection in the organization's flag	 The world map from OpenStreetMap, using Mercator projection

Figure 2.8: Two examples of map projection

A map projection is formed based on complex mathematical calculations [79] to define the relationship between a point from the Earth's surface with its equivalent location on the map [5].

To visually demonstrate how map projection works, Battersby (2017) used the term "developable surface" - "a surface that can be flattened to a plane without introducing distortion from compression or stretching" [5]. Similarly, Koponen and Hildén (2019) describe the process as using a "projection surface" - an imaginary paper and wrap around the globe [38]. Therefore, three different methods of constructing a map projection are proposed: azimuthal - projecting the Earth on a plane, cylindrical - on a cylinder, and conic - on a cone.

The two-dimensional surface cannot deliver all the details from the three-dimensional spaces. The act of preserving certain features of the Earth such as distance, area or shape, etc. differently results in infinite ways of map projection [69]. Each projection serves a particular purpose, therefore none is considered the best. Besides, they all result in distortion, which is one of the causes to map miscalculations and misinterpretations like Figure 2.9.

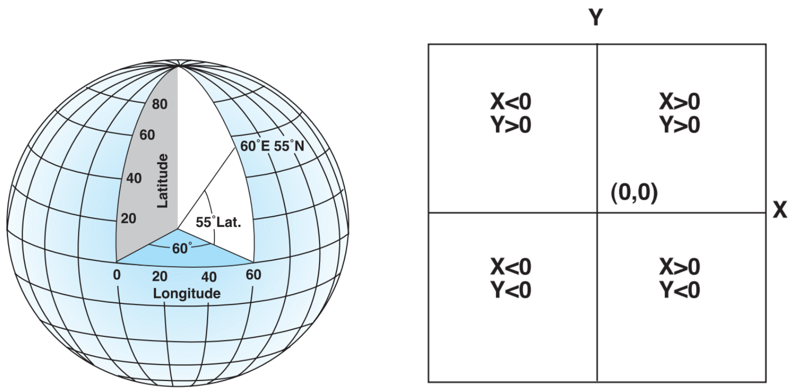
Figure 2.9: Map distortion example - Greenland looks as large as Africa, while it is about 14 times smaller in real life

2.3.2 Coordinates reference system

Coordinates Reference System (CRS) or Spatial Reference System (SRS) defines the relationship between the location on the map and the location in the real world. This relationship, namely coordinates, is represented by a set of numbers (e.g: the location of Kamppi Center, Helsinki is 60.1702358, 24.9352622 on Google Maps). Map projection is one of the factors forming a CRS [56]. Therefore, as there is no limitation on the number of possible map projections, the number of CRS is infinite also. Kennedy (2000) classifies CRS into two categories [36]:

- The Geographic coordinate system (GCS) defines a location based on the three-dimensional space. CRS of this type locates a point on the Earth's surface based on its two angles to the Earth's center. The angles are called latitude and longitude, measured by degree. A popular example of GCS is WGS84, which is widely used in navigation and mapping services, like the Google Maps in the example above.

- The Projected coordinate system (PCS) identifies a location using a flat, gridded surface with a center point. In PCS, the coordinates are in the format of x, y values, referring to the distance of the specific location to the center point, and are measured using linear units of measurement such as metre or feet. Two among the factors forming a PCS are a GCS as the base, and a map projection. The current PCS of Helsinki Metropolitan Area is ETRS-GK25, based on the ETRS89 system.



(a) GCS (b) PCS

Figure 2.10: How a location is defined in different CRS [36]

Geovisualizations commonly involve multiple data sources (as the base map has its own data source too), therefore it is important to investigate the CRS each of them are using. Obe and Hsu suggest that any two sources of data sharing the same EPSG number result in a ”perfect overlay” [56]. The EPSG number is the identifier assigned to each available CRS by the European Petroleum Survey Group. In the EPSG system, WGS84 is identified as EPSG:4326[1] and ETRS-GK25 is EPSG:3879[2].

2.4 Personalize a visualization

This section is inspired by two concepts. Firstly, ”news application”, defined by ProPublica and mentioned by Alberto Cairo (2016) in his book ”The truthful art” [10]:

[1]`https://epsg.io/4326`
[2]`https://epsg.io/3879`

> A news application is a special kind of visualization that lets people relate the data being presented to their own lives.

Each reader has an unique background and interest, therefore the insight perceived by everyone is not the same. As the visual information seeking mantra (overview first, zoom and filter, then detail on demand) explains [64], people only pay attention to the overall information and a particular part of the detail, not the whole data set. These interesting parts can be what the reader can relate to (e.g. their birthplace) or the group that he or she belongs to (e.g. age, occupation, or income status, etc.).

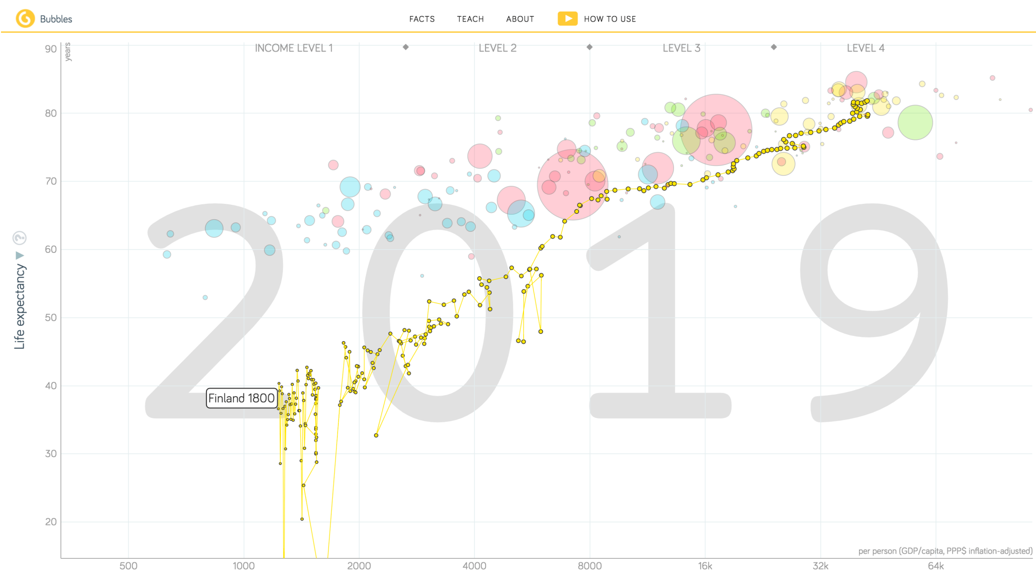


Figure 2.11: The Gapminder tool, comparing the income and life expectancy of Finland from 1800 with the rest of the world in 2019

"News application" is usually seen as the form of exploration tool for public data set. For example, the Gapminder tool[3] first provides the chart of income and life expectancy of countries in the world. The user can choose a country of interest, and observe the position of that country changing every year from 1800, in comparison with the whole world as Figure 2.11. Another field known for utilizing news applications is data journalism - where journalists attract the readers and enhance report quality by accompanying their stories with visualization [30]. "Where your job is most popular"[4] (figure 2.12) does

[3] https://www.gapminder.org/tools/
[4] https://flowingdata.com/2019/02/27/where-your-job-is-most-popular/

not provide the overview, and only becomes meaningful when users input their own information. This visualization has done well in placing the hint to its feature by randomly fills the text box when it starts. Another good example of relating personal life with the data is "What Bloomberg's half-billion dollars in campaign spending would cost you on your budget"[5], even though it is not an information visualization.

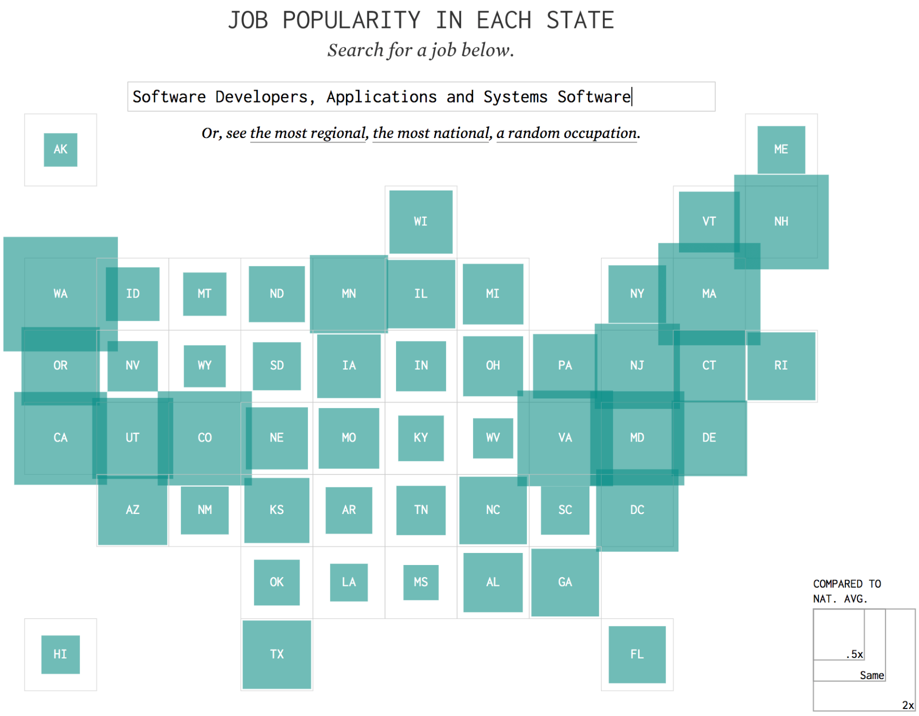


Figure 2.12: "Where your job is most popular" from Flowing Data

The second inspiration comes from a popular term in urban planning: pedestrian shed, or pedshed in short. The pedshed around a certain place (like a metro station or a supermarket) stands for the distance that people are willing to walk from that place to other destinations before considering other means of transportation (e.g. taking the bus, or driving the car, etc.).

The range of a pedshed is measured by walking time or the distance in the form of an as-the-crow-flies circle. The common value is 5 minutes, which is equivalent to a circle of 400m (or a quarter of a mile) radius [47] [72]. Pedshed is a useful factor to determine the life quality of a neighborhood, such as the

[5] https://www.washingtonpost.com/graphics/2020/politics/wealth-comparison/

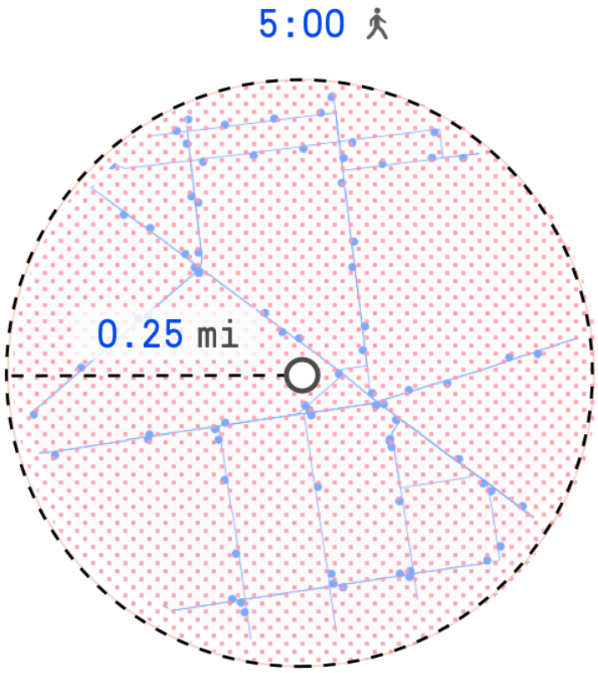


Figure 2.13: The pedestrian shed, illustrated by Morphocode

availability of public transport or healthcare facilities [58]. Explained in a more data-oriented way, "the five minute walk sets a scope for collecting both quantitative and qualitative data at a human scale" [50].

In the context of parking, Ji et. al. (2007) recognize that walking distance is the most important criteria when choosing a parking spot [34]. Based on Ji's work, the study from van der Waerden et. al. (2017) points out the acceptable walking distance based on different purposes ranging from 50 to 500m [75]. These results are related, and perhaps can shed some light on the visualization of parking violations.

2.5 Information visualization on the Web

Visualization, like all other content on the web, reaches its audience through a program called browser. On the very first days, browsers can display just text and simple images. Then, graphics, interactive content and animations are available through the support of third party plugins such as Java or Flash. This dependence ends in recent years with the emergence of new technologies like HTML5, CSS3 or WebGL [7]. The development of computing hardware and internet technology further empower web browsers to deliver complex information faster and wider, including visualizations.

This section introduces the basic components used to display information on the modern web as well as popular toolkits for web visualization.

2.5.1 Building blocks of the Web

Looking at the source code of a website, it is just text written in a certain structure. The browser then takes these instructions and translates them into what humans see on the web. Nowadays most web pages involve components made by the following languages:

- Hypertext Markup Language (HTML) provides means to define the content on a web page. One HTML element, or HTML tag, consists of the tag name written between a pair of angle brackets, attributes of the tag, and its content. For example, the tag in Figure 2.14 will be displayed as a clickable link with the text Aalto University. The content can be plain text, or another nested HTML tag. Multiple elements with specific structure and order form a HTML document, which the browser will base on to render the content instead of displaying the tags [78].

Figure 2.14: An HTML tag

- Cascading Style Sheet (CSS) describes how the web page should be presented, such as size, color, layout, etc. CSS can be included as an HTML attribute (inline), or between the <*style*> and </*style*> tag. The latter is recommended as it allows the rules to be reusable between different elements and pages. A CSS block written in this way contains a set of rules which apply to all tags that match its selector. The example in Figure 2.15 change the looks of the link in Figure 2.14 into the red, slanted *Aalto University*

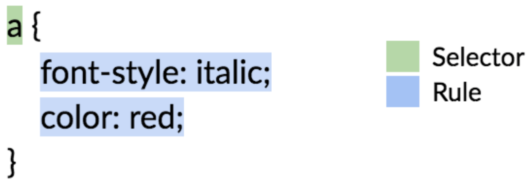


Figure 2.15: A CSS block

- JavaScript (JS) controls the behavior of the elements on the web page. Unlike HTML or CSS which are markup languages, JS is a scripting language which allows logic and algorithms to be implemented. JS code is written within the <script></script>tag in a HTML document. JS can be used to create dynamic web content, by its ability to mass generate and manipulate other elements like HTML, CSS or SVG. JS also handles the interaction between users and the web, as well as between the web elements themselves. A huge advantage of JavaScript in comparison to other scripting languages is it can be interpreted by all popular browsers nowadays. The browser can open and run a JS file directly, but not a Python file.

```
<script>
  document.getElementById('aalto').style.color = 'green'
</script>
```

Figure 2.16: This script will change the color of the link in Figure 2.14 into green

- Scalable Vector Graphics (SVG) uses the same structure as HTML to define graphics on the Web. Unlike traditional raster images (like JPG, PNG, etc.) which become blurred when enlarged [76], SVG are vector images, which can preserve the same quality no matter how they are resized. Storing as the form of computer-generated code, a SVG file is generally lighter than a raster counterpart. On the web, SVG drawings can be styled by CSS, and manipulated by JS for animation and interaction.

```
<svg width="200" height="250" version="1.1" xmlns="http://www.w3.org/2000/svg">
    <line x1="10" x2="70" y1="100" y2="100" stroke="orange" stroke-width="5"/>
    <polygon points="70,90 70,110 80,100" stroke="orange" fill="orange" />
</svg>
```

(a)

(b)

Figure 2.17: The SVG code above generates the orange arrow below

2.5.2 Web visualization libraries

In programming, a task is accomplished through multiple lines of code, grouped in a function. The function is made to avoid code repetition: instead of writing the same sub tasks again, the developer just needs to call the function name. A library is the collection of well-written functions, solving a set of tasks, and shareable from developers to developers.

JavaScript libraries for information visualization on the web mainly provide the shortcuts to output web graphical content based on the input data. It is irrelevant for the scope of this thesis to have a comprehensive list of visualization libraries, therefore some notable names are introduced below:

- D3[6] is considered one of the most popular libraries for visualization. Its strength is based on the ability to structure and bind the data to HTML or SVG. Users then can ”draw” the visualization graphics by creating, selecting and manipulating these elements. Modules for building animation, interaction and handling geographic data are also included. The programming effort is decent and the learning curve is quite steep, but there is almost no limitation of what can be created with this library.
- Charting-specializing libraries are collections of ready made basic graphs like line charts, doughnut or scatter plot, etc. With these libraries, user just input the data through a function and the visualization is generated. As the components are built-in, the customization of their

[6] https://d3js.org/

animation, interaction or styles is limited to some extent. Besides, visualization components in charting libraries have fixed data structure, therefore more data processing effort is required beforehand. Despite mentioned disadvantages compared to D3, charting libraries are easier to learn and require less coding, which is suitable for beginner or quick development works. Popular names based on the number of GitHub stars in this categories are ChartJS (47 800), eCharts (40 100), Chartist (12 000), etc.[7]

- Mapping libraries focus on the creation of interactive maps. The main advantage compared to the geographic module of D3 is, these libraries provide developers their own base maps to work on. A base map contains real world context such as boundaries, street, or even satellite images, and acts as the background layer for the main data. Base maps improve the ability to comprehend and validate the data, as humans normally relate to "Helsinki central railway station" better than (60.1718756, 24.939233). Mapping libraries also provides basic interactions with their base maps such as zoom, pan, location search, or navigation, etc. Widely used mapping libraries on the web include Google Maps, Leaflet, MapboxGL JS, ArcGIS, etc.

As each library has its own strengths and weaknesses, it is common to combine multiple libraries together in complex visualization projects, for example, MapboxGL JS can be used for the base map, with overlay data created from D3, and supporting charts made with ChartJS.

[7] Rounded numbers to the nearest 100, recorded at the time of visit (24.03.2020)

Chapter 3

Framework

Based on the knowledge mentioned from the previous chapter, this chapter reviews two frameworks for developing computer generated visualization:

- The seven stage process of Computational Information Design, proposed by Ben Fry (2004) in his doctoral thesis
- The nested process model for Visualization Design and Validation, introduced by Tamara Munzner (2009)

Throughout this document, the names of the above frameworks will be omitted as "the 7-stage process" and "the nested process" respectively, to make the reference more convenient. The terms "level", "phase", "step", "stage", and are used interchangeably to refer to a stage in Fry's process or a level in Munzner's model.

Taking the strengths and weaknesses of each framework as foundation, later this chapter presents an adapted framework suitable for the scope of this thesis - visualizing parking violation data.

3.1 The 7-stage process by Ben Fry

Fry (2004) argues that data visualization is a process, which requires the knowledge from many different fields: Computer science, mathematics, statistics, data mining, graphics design, information visualization, and human-computer interaction. His paper points out that the solutions provided by these fields are disconnected. Design and visualization alone provide visual aids, yet lack the ability of dealing with immense amounts of data. Statistics and data mining tackle this problem, but do not support any interaction with the data. Programming focuses on equipping visualization creators with means of mass representation and interaction, while takes aesthetic aspects lightly.

To approach the problem, which occurs especially in complex data visualization today, Fry combines the necessary skills and knowledge from the mentioned areas into a single process. In the context of his dissertation, he collaborates with Casey Reas in developing a tool named "Processing" to demonstrate the capability of the process, and further, to aid other developers and designers in their work [61].

According to Fry, the process of creating a visualization with the aid of computer includes the following stages: *Acquire, Parse, Filter, Mine, Represent, Refine, and Interact*. He explains the order of the stages with relevant skills and background knowledge in Figure 3.1.

Figure 3.1: The 7-stage process (Fry, 2004, p. 13) [25]

Acquire is the step of obtaining raw data, as the visualization creator is not always the data owner. Fry identifies the significant knowledge in this stage is understanding how each data source works and their limitations, in order to decide the most suitable way of storing the data. For example, files are easier to read and write, but must loaded as a whole so large data could become a bottleneck. Databases are specialized in handling large data and allow user to "query" a part within, but too many connections could decrease the performance.

Parse converts the data into a useful format. Usually the raw data is unstructured, or structured to some extent and does not meet the visualization needs. Occasionally, the data is pre-filtered (by techniques such as decrypting or decompressing), before categorized and restructured to be suitable and easier to consume for the tool used in the later stages.

Filter analyzes and prioritizes the fields within a data set, then omits unnecessary parts, only keeping the data of interest. Filtering acts as the preparation, because a large data set might serve several purposes, so the whole can be redundant for a certain visualization. Besides, Fry noted that this stage has a strong connection with "Interact" stage, when the data might be filtered further based on the demand of the reader.

Mine can be interpreted as to "summary" the data content. This step applies mathematics and statistics to calculate basic figures such as minimum/maximum value, mean, median, normalization, or standard deviation. Data mining methods are also used in the need of reducing the number of dimensions in a data set, or classifying the data into groups.

Represent aims to select a basic graphics form to visualize the data based on how the data set was filtered and mined in the previous stages. Beside the catalog of visualization types provided by Fry in his thesis, the comprehensive guide from Heer, Bostock, and Ogievetsky (2010), or Koponen and Hildén (2019) could be a good starting point for this step [27] [38].

Refine applies visual design techniques to communicate the data content better. This stage includes making the graphics elements become clean and clear so the difference between important features and less relevant ones can be easily detected by the readers. Fry takes example of various strategies for differentiation: contrast, size, weight, color, and placement. The usage, pros and cons of visual variables are discussed and summarized in section 2.x.x in this thesis.

Interact serves two purposes. First, this stage defines the rules for the data to interact with itself (like the size of elements when zooming, or the position of elements having connection). Second, it involves the design of controls for readers to directly manipulate the visualization on demand. The importance of limiting the range of user actions is emphasized in Fry's text, for the goal of better guidance and error prevention.

Acquire	Obtain and store the data	The data set and data storage method
Parse	Organize the data	The easier-to-consume data format
Filter	Remove unnecessary data	The relevant part of the data
Mine	Analyze the data using mathematics, statistics, and data mining methods	The overview of information within the data
Represent	Select the appropriate visualization type	The most basic version of the visualization

Refine	Improve the visual elements	The better version of the visualization in terms of aesthetics and effectiveness
Interact	Define rules for data self-interaction, and adding controls to the visualization	The components which allow readers to interact with the visualization

Table 3.1: The summary of Ben Fry's process

Even with the declared order, the stages are iterative. The output of each stage affects the others in different ways. Thus, it is very rare to take each stage just once in the whole process. Instead, the earlier stages are revisited and optimized based on the needs of the later ones.

3.2 The nested process by Tamara Munzner

Munzner (2009) recognizes a problem among the visualization frameworks at that time: each either focuses on how to create a visualization, or how to evaluate it. There are a lot of "how", but a guidance about when to choose what was missing. These problems motivate the author to introduce her own framework which splits visualization process into four levels as Figure 3.2

Munzner explains the metaphor of "nesting" as the output of outer levels (upstream) is the input of the inner ones (downstream). This model challenges visualizers from the early stages of the process, as "an upstream error inevitably cascades to all downstream level". The levels are described based on "threats" - factors, which cause the errors and "validity" - methods to verify if errors happen. There are two approaches to validation: immediate - taken right at the current level, or downstream - at the inner levels.

Domain level aims to understand the problem of the field, where the data context is in. The designers and users might not speak the same language, since each domain has its own terminology. This is where the errors start to emerge. According to the author, the primary threat in this level is proceeding to solve a problem, which users do not have. This threat can be verified immediately by gathering qualitative responses through interviewing observing target audience. For downstream validation, Munzner suggests analyzing the adoption rate when the product is ready, even she admits this method is not always reliable.

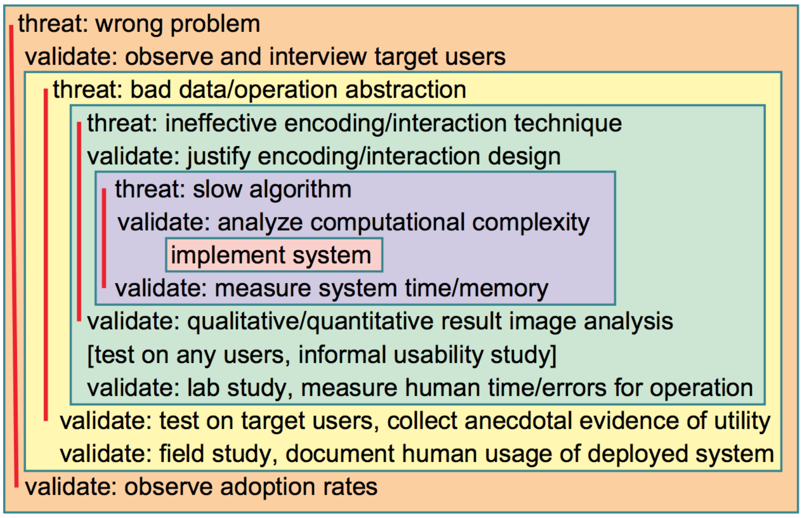


Figure 3.2: The nested model (Munzner, 2009, p. 923)

Abstraction level consists of two aspects. First is to find out, which operations users wish to take with the data. Second is to convert raw data into the types which can be processed by visualization techniques. According to Munzner, it is common that designers take the former step of domain researching lightly and skip this step with the assumption that they know the right problems to solve. As a result, wrong data type and operations are chosen. Observing target audiences using the product prototype in their own environment is the only validity method proposed. Without any immediate check applicable, this step is the most difficult among the four.

Encoding and Interaction level visually represents the data and how users can interact with the data. Ineffective communication between the design and the readers is the main threat in this step, and can be validated by applying cognitive principles or analyzing the work based on design guidelines such as heuristic evaluation [54], expert review [74]. More intensive downstream validation suggested are laboratory experiment and result discussion with images and video.

Algorithm level executes the work of generating the visual elements and interaction methods. Two threats might present in this stage: dissatisfying performance of the system in terms of speed and memory, or the system renders unexpected features compared to the former step. Algorithm complexity analysis is proposed as the immediate validation. Downstream validation utilizes system benchmarking techniques. Both rely heavily on computer science knowledge.

The examples in Munzner's paper point out that not all levels are present in a visualization project. Due to the input-output feedback between the levels, they are not executed in a strict order and can be returned after the validation of other steps. Low fidelity prototypes [77] without algorithm implemented such as paper testing, sketches, or still images can be developed to speed up the validation process.

3.3 Discussion on the two frameworks

Through their work, Fry and Munzner introduced two practical frameworks as guidelines to computer-generated visualization process.

In general, the content of both frameworks agree with each other in many aspects. "Abstraction" level in Munzners process shares the same ideas with Fry's "Parse", "Filter" and "Mine" stage in stating the importance of pre-processing data before visualization. Meanwhile, "Encoding and Interaction" level from the nested process covers the action taken in "Represent", "Refine" and "Interaction" of the 7-stage process. Besides the similarities, there are notable differences which result in the advantages and limitations of each framework.

3.3.1 The focus

Scope of the nested process is wider compared to the 7-stage process as described in table 3.3. With the focus on evaluation - the "why", Munzner sees the project as a series of problems and solutions, threats and validations. Through the nesting metaphor, she pointed out the importance of solving early-emerged threats since they can affect all the steps after.

However, when explaining the implementation aspects, Munzner's terminology is too board. With the argument of interconnection between tasks, many components are included into one level. A noticeable example is the third layer - Encoding & Interaction - combines visual representation with interaction method as its name implies. Similarly, the Abstraction level covers the workload equivalent to three stages Parse, Filter and Mine in Fry's process.

There are two drawbacks on this approach. First, by broadening the scope, the primary threats for each level become too general, as Munzner recognized that she "oversimplifying the problem". Second, as each task in the visualization process requires different backgrounds to approach and has own set of problems to solve, merging these tasks makes the explanation of how to carry out the work unclear. Thus, it is hard to use the nested process as a sole guideline for data visualization.

Focusing on providing a tool for visualization - the "how", the 7-stage process fills up the mentioned hole. [Note: write a sentence to explain this]

3.3.2 The human factor

According to Kirk (2013), "human is the most impactful factor in visualization success" [37]. In a visualization project, the domain expert [51], the visualization implementer, and the end user are not always the same person. The human factor implies the difference in viewpoints, and the importance of communication between them.

The 7-stage process focuses on the workflow for "single practitioner" as the author stated. As a result, the human factor is not mentioned there. On the other hand, most of the content in the nested process is functioning around the relationship between the main implementer and other participants. In three outer levels among the four, Munzner explains the role of domain expert and end users in validating if a threat presents, when to involve them, and in which method.

3.3.3 Project evaluation

Evaluation, in the 7-stage process, is just slightly mentioned in the "Refine" stage, and has two major drawbacks.

First, while the refinement of visual elements is discussed, system performance did not receive the same attention. In the scope of this thesis, which visualize large amount of parking violation data on the web, this point needs to be addressed because a slow system hurts user satisfaction severely. People are impatient, and long waiting time might kill their interest [49].

Second, it is common in exploratory visualization that the visual designer is not the user, but helping the users solve their problems via data [38]. The "Refine" stage only evaluates and improves the visualization based on rules and guidelines from graphics design knowledge. As Nielsen (1993) stated "Your best guess is not good enough" [52], Fry's model also needs to verify the effectiveness of the visualization with its actual users.

The nested process covers these absent aspects by analyzing system efficiency in the innermost level, and guiding the process of user evaluation in every stage of the project.

To summarize, the nested process focus on the "why", while the 7-stage process focuses on the "how" in a visualization project. They are not competing, but complementary to each other. Table 3.3 describes the differences, similarities and relationships between the two processes. The idea for this table is inspired by the Master Thesis of Presnyakova (2017) [59], where green boxes mean the appropriate component is introduced in the process, and red boxes mean the component is not mentioned.

Component	The nested process	The 7-stage process
Domain analysis	Domain	
Data acquisition & storage		Acquire
Data transformation	Abstraction	Parse Filter
Data analysis		Mine
Visual encoding	Encoding & Interaction	Represent
Visual refinement		Refine
Interaction		Interact
System optimization	Algorithm	
User evaluation	Domain Encoding & Interaction	

Figure 3.3: Comparison of the 7-stage process and the nested process

3.4 The adapted process

The mentioned reflections raise the need of formulating a framework based on the two introduced processes. The adapted process takes the 7-stage process as the starting point, and extends it by including the human factor from the nested process.

The context of this thesis fits the two-role project model introduced by Koponen and Hildén (2019) [38], where the Parking Control and Services of Helsinki (Helsinki Pysäköinninvalvonta ja pysäköintipalvelut) acts as the content owner, and Ninh Truong as the implementer.

The adapted process groups the stages in four categories: Requirement, Preparation, Execution and Improvement, with the following addition:

- "Understand" is added as the second stage. This thesis takes open data to make a tool for exploring it. The data context belongs to a special domain (parking), and is in a language which is not familiar to the implementer (Finnish). Therefore, after acquiring the data, it's important to communicate with the owners to understand every term in the data content. Gathering their opinions about which features should be focused, or which interaction they prefer can help the execution of later stages to some extent.

- "Draft evaluation" is conducted after finishing the most basic visual form. This stage involves the content owners and a one or two users who understand the domain, with the ultimate goal is to test if the intended features or patterns are easy to recognize. The worst case, when "Representation" stage need to be completely reworked, happens when the target users cannot identify the differences, or misunderstand the data. Otherwise, opinions and comments at this stage can be taken as input for the next stage - "Refine".

- "Optimize" stage is included to improve data rendering algorithm and methods if needed, after visual elements refinement. The output of this stage might result in the revisit to "Parse", "Filter" and "Mine", where data is reformatted and reanalyzed for better performance.

- The workflow is finalized with "Usability test" stage - which checks the effectiveness of the visualization against end users when it is ready technically and visually.

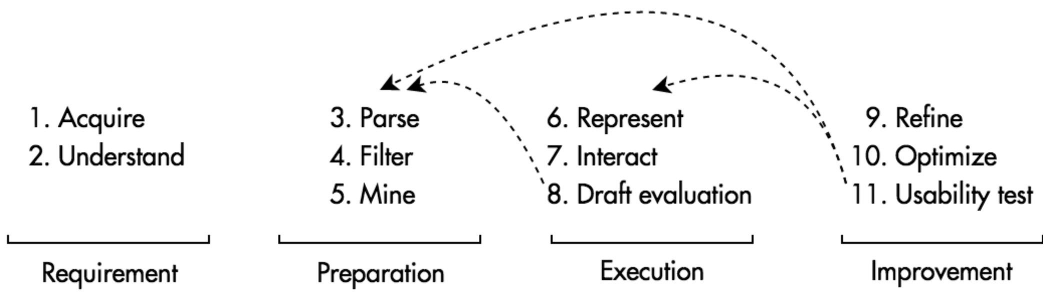


Figure 3.4: The adapted process

Chapter 4

Implementation

The data set from Helsinki Region Infoshare (HRI) named "Parking violations in Helsinki" records the violations noticed by the officers, which result in a warning or a parking fine. Undetected violations are not in the context of this data set. The visualization works with the full year of 2018 and 2019.

Appendix **??** contains the detailed list of violation reasons and their explanation.

4.1 The tool

The Helsinki parking violation exploration visualization, which will be referred to as "the tool", is accessible at `https://tmhn89.github.io/master_thesis` . The tool has five main components (Figure 4.1)

- The map (1), in which each location having parking violation is shown as a bubble. Two factors determine the bubble size: number of violations occurred in the selected time period, and the current zoom level of the map. Each reason of violation belongs to a category (see Appendix **??**), and the bubble color is calculated based on the reasons involved. Under the violation layer is the base map with land boundaries. Street names and building blocks are available when zooming in, which helps users aware of the space better. Users can perform basic map actions such as changing the zoom level or moving the map around. Zooming closer (with zoom level more than 14) enables clicking on each bubble for its detail.
- The period selector (2), appears as a bar chart, representing the months

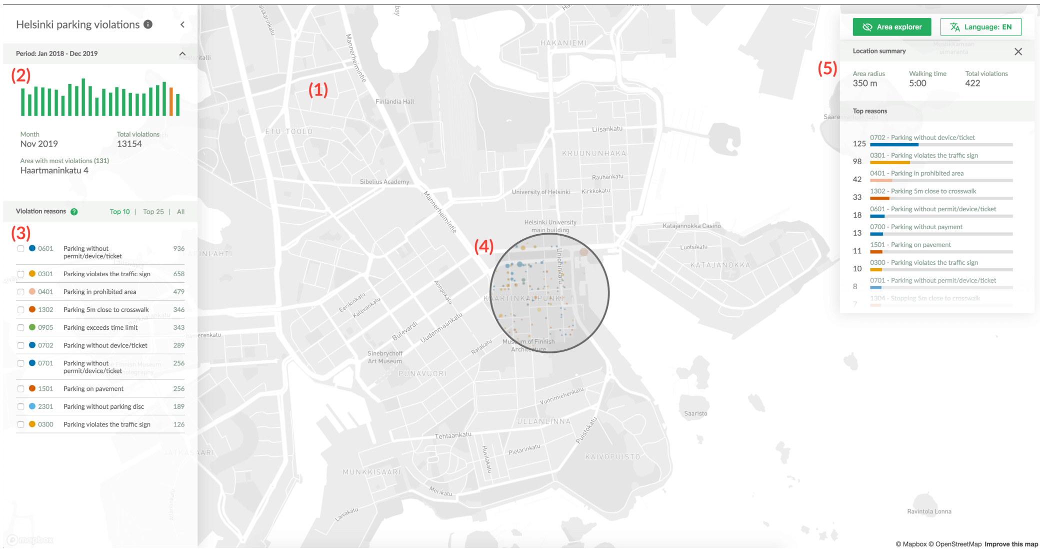


Figure 4.1: The Helsinki parking violation exploration visualization tool

and total violations happened in each. Clicking on a bar selects a month, while brushing[1] through the chart selects a period of consecutive months. The data of the map and reason list will be repopulated according to the selected period.

- The reason list (3), displays all of the reasons leading to a violation in the current selecting time. The list is sorted by number of occurrences for each reason. The user can choose to show only interested reasons on the map and the summary by clicking the checkbox on its row.

- The explorer component (4), triggered by the "Area explorer" button on the top right of the view. When exploring mode is active, clicking on the map will place a circle at the same location, summarizing all the violations happening within its boundary. The default radius of the exploring circle is 350m, which can be modified up to 1km by dragging the circle's border.

- The summary popup (5) lists all of the violations involved in the selected area, and only visible in explorer mode, or when a bubble is clicked.

[1]Brushing: highlight to select a subset of the data [19]

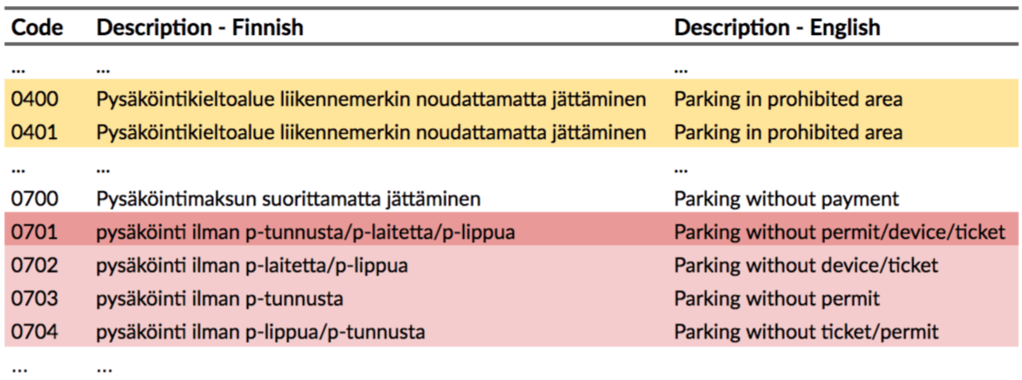
4.2 The process

4.2.1 Acquire

"Helsinki parking violations" dataset is open for public from the website of Helsinki Region Infoshare[2]. The files are in CSV format, and updated three times per year.

4.2.2 Understand

An important field to categorize the records in the data set is "Virheen pääluokka / pääsyy" or "Main reason". A reason consists of a four-number code and the description text. Investigating the reason list encounters two issues. First, there are multiple reasons sharing the same description, but having different code. Second, a number of reasons cover the content of others. An example is provided in Figure 4.2.



Code	Description - Finnish	Description - English
...	...	...
0400	Pysäköintikieltoalue liikennemerkin noudattamatta jättäminen	Parking in prohibited area
0401	Pysäköintikieltoalue liikennemerkin noudattamatta jättäminen	Parking in prohibited area
...	...	...
0700	Pysäköintimaksun suorittamatta jättäminen	Parking without payment
0701	pysäköinti ilman p-tunnusta/p-laitetta/p-lippua	Parking without permit/device/ticket
0702	pysäköinti ilman p-laitetta/p-lippua	Parking without device/ticket
0703	pysäköinti ilman p-tunnusta	Parking without permit
0704	pysäköinti ilman p-lippua/p-tunnusta	Parking without ticket/permit
...	...	

Figure 4.2: Example of confusing violation reasons

Project manager Oskari Rantanen from the Urban Environment Division of Helsinki helps clearing the confusion by providing the form currently used by the parking warden in the city. The form specifies that all reasons for violation are organized in 30 groups. All the reasons with the same first two digits in their code can be grouped together. The group names match the reasons which end in "00".

[2]https://hri.fi/data/en_GB/dataset/pysakointivirheet-helsingissa

Expressing his opinions on a tool based on the data set, he expects a fast and efficient way "to see where and why parking fines come from".

4.2.3 Parse and Filter

This stage use *d3-dsv*[3] for converting the text encoding, pandas[4] for restructuring the data, and Google Geocoding service[5] to translate addresses into coordinates.

First, special characters from the Finnish language (ö, ä, and å) are unreadable with the default encoding of the original CSV. The file encoding is converted into UTF-8 which is more compatible with common data processing tools.

Next, one year's data has approximately 150 000 records, with the size about 23MB, which is too large for a file loading on a web page. As in Figure 4.3(a), there is a huge amount of repeating data in full text. After removing columns without data or containing unnecessary data, the following fields are kept and reformatted:

- Virheen tekokuukausi (month) is stored as number (1, 2, etc.) instead of (Tammikuu, Helmikuu, etc.)
- Virheen tekovuosi (year) is meaningful, but is omitted from the dataset as the data file name itself contains the year.
- Osoite (address) is reverse geocoded into coordinates in the format of "latitude, longitude" following the WGS84 system. As the address text can grow the file size a lot, but less accessed in the visualization (only shows in the summary or detail panel), all unique addresses in text form are collected into a separated CSV file for referencing afterwards.
- Virhemaksun vaihe (type) contains two values: Pysäköintivirhe (parking fine) and Huomautus (warning). To simplify, the field is renamed into "is_warning", where the former value is represented as 1, and the latter is 0.
- Virheen pääluokka / pääsyy (main reason) originally includes both reason code and text. Only the codes are kept in the main dataset, and the text is stored in a different file.

[3]https://github.com/d3/d3-dsv/
[4]https://pandas.pydata.org/
[5]https://developers.google.com/maps/documentation/geocoding/start

The above actions reduce the file size for each year from 23MB to about 4MB. Analyzing the newly created file, there are rows with the same content, which means violations with a particular reason can happen multiple times at a location in the same period. Grouping these rows together and counting the number of repeating records into the "occurrence" field further shrink yearly data files to roughly 2.5MB each.

Virheen tekokuukausi	...	Toukokuu	Toukokuu	...
Virheen tekovuosi	...	2019	2019	...
Osoite	...	Kumpulantie 1	Teollisuuskatu 18	...
Virhemaksun vaihe	...	Pysäköintivirhe	Pysäköintivirhe	...
Virheen pääluokka / pääsyy	...	1504 Pysäyttäminen jalkakäytävälle	1501 Pysäköinti jalkakäytävälle	...
Virheen kirjaaja	...	Pysäköinnintarkastaja	Pysäköinnintarkastaja	...
y*	...			...
x*	...			...
Postinumero*	...			...
Postitoimipaikka*	...			...
Alue*	...			...
Kunta*	...			...
Kunta_nro*	...			...
Kaupunginosa	...	VALLILA	VALLILA	...
Kaupunginosa_sv	...	Vallila	Vallila	...

↓

(a)

month	...	5	5	...
is_warning	...	0	0	...
reason	...	1504	1501	...
coords	...	60.1965727 24.9411528	60.1947724 24.9454242	...
occurrence	...	1	1	...

(b)

Figure 4.3: The data structure, before and after processed. Fields with * mark are left empty in the original data set

In the parsed data file, there are rows without coordinates. Unable to be placed on the map, these records need to be filtered out. These rows in the original CSV either have no address, or the address is filled as "-". Excluded features account for around 2% of the total yearly data, as summarized in Figure 4.4.

Year	2018	2019
Total records	154907	150097
No address	2942	3091
% fitered out	1.90%	2.06%

Figure 4.4: Number of violations which cannot be visualized

4.2.4 Mine

The original data set has a total of 140 unique reason codes leading to a parking fine or warning, arranged into 30 groups. Analyzing the meaning of the group names allows further organizing them into 9 categories, to be more suitable for the human short term memory capability [48].

To prepare for representing bubble size, a necessary step is to investigate how many violations might occur at the same location in one month. The maximum number is recorded at Mäntymäenkenttä in February 2019, where drivers violated the rules 311 times.

4.2.5 Represent and Interact

The most basic form of the tool projects the coordinates of each record to the center point of a SVG circle on the screen. The category which a violation reason belongs to determines the color of the circle, while its radius is proportional to the number of occurrences. This step uses D3 to generate the SVG elements and calculate the projection.

A layer of base map by Mapbox is added under the violation layer in order to confirm the points are correctly placed. The base map layer also provides basic map navigation, including zoom - make the map wider or closer to reveal more or less detail, and pan - move the map position to examine different parts of the whole data. Every time the user interacts with the map, the projection of the violation layer is recalculated based on the center point, zoom level and scale of the base map. This action also redraws all the points in the violation layer to maintain their location preciseness.

Clicking a location on the map displays the detailed data of the bubble: what are the reasons for violation happening and how many times each reason occurs in the form of horizontal bar charts. Horizontal bar charts is chosen over the traditional pie chart as comparing lengths is more accurate than comparing area (Figure 2.7).

The explore feature summarizes the same information, but for a specified area. Activate explore mode hide all violations but the ones inside the explore area. The explore area is also a SVG circle, which has a thick border to distinguish from violation data points. Dragging this border resizes the explore area, while dragging inside the area moves it around.

For the sake of performance, the tool does not display the whole data set, but a month's data by default. The overview of monthly total violations is represented as a vertical bar chart, where the user can hover the mouse over a column to examine the number, and click a column to load and display the data for the corresponding month. The user can further filter the data through the list of all reasons which happen in the selected month.

The changes in any filter, including month, reason, selected location, or exploring area are handled by *d3.event*, which triggers the redraw and update of all other sections.

4.2.6 Draft evaluation

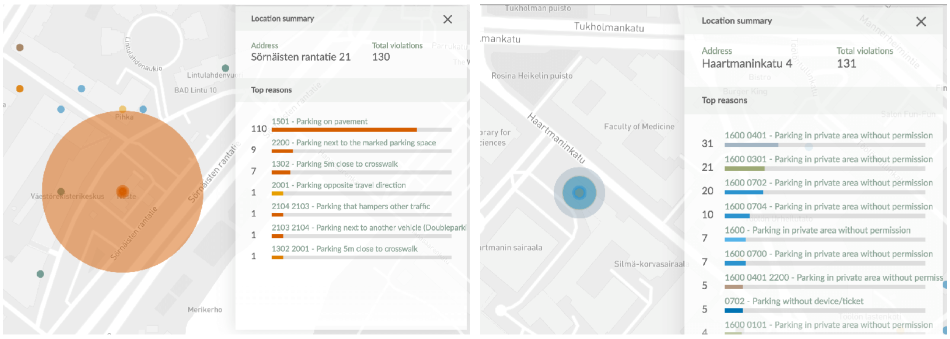
The draft evaluation includes two participants. One has been working 3 years in the parking industry as Customer Success Manager, the other drives frequently, so they already have experience with the domain. The participants are asked to perform a set of predefined tasks as well as freely interact with the tool. The setup is similar to usability testing stage, which is explained in depth on Chapter 5.

Visual encoding: Both participants can immediately make comparisons based on the size of the bubble, and express their excitement when interacting with the explorer tool. A participant notices that there are two locations with similar large numbers of violations but their sizes look obviously different. When zooming out to a certain level, the bubbles become extremely large, blocking the whole view.

Usability: While the parking expert completes the tasks with almost no assistance, the driver has trouble of recognizing the available features at first. He does not know the columns in the timeline are clickable, and struggles on pointing the mouse exactly to the explorer's border to resize it. Both participants experience delays when zooming and panning the map, especially at wide zoom levels when large numbers of bubbles are visible. The parking expert, when moving the map outside Helsinki, notices there are violation bubbles in other cities as well.

4.2.7 Refine

Originally, each reason is represented as a bubble at the coordinates where the violations took place. When a location has multiple violation reasons in the same period, multiple bubbles will be drawn at the same place. This encoding results in one of the draft testers interpreting the number of violations incorrectly. In the case demonstrated in Figure 4.5, there are approximately the same amount of violations happening in both locations (130 and 131 respectively). However location (b) has several reasons stacking together, the user thinks that it has much less violation than location (a), which has one dominant reason. He also needs to click exactly on the center of the circles to get the precise summary of the location.



(a) (b)

Figure 4.5: Stacking reason bubbles makes the user interprets the number of violation wrong

To solve this problem, multiple violations at the same location are grouped into one bubble. The color of the bubble is defined by taking the average color of the reasons and their number of occurrences. This calculation is supported by the color-specializing library *chroma.js*[6].

Map zoom level and number of months in the selected period are involved when calculating the radius in order to prevent cluttering when the map is wide-zoomed. Minimum zoom level is set to 11, which is enough to cover the whole Helsinki area.

[6]`https://gka.github.io/chroma.js/`

To tackle the hard-to-resize explorer issue, making the border thicker can draw the attention from the violation bubbles. An alternative solution is adding a transparent layer along with the explore circle to receive the drag event, where its interaction area is 16px outside the visible border. Moving the mouse over the interaction area changes the pointer icon to let the user know that actions can be taken.

The violation bubbles are the center of the visualization therefore they can draw attraction from the controls. When the tool is loaded for the first time, a tutorial is shown to give the user prior knowledge and expectation of what they can do and where to perform each action.

4.2.8 Optimize

Reverse geocoding: Violation bubbles appear outside Helsinki means a number of addresses are converted to coordinates incorrectly. The reverse geocoding in the Parse stage is performed by passing the address field in the original data set to Google Geocoding API, then assigning the coordinates of the first result to the address. It is common that different regions can share the same street name. That fact explains why searching for "Sibeliuksenkatu 1" returns a place in Helsinki, but "Sibeliuksenkatu 6" points to Lahti - a city 100 km away.

The search result can be controlled by adding the city name after the address when searching. In the mentioned example, "Sibeliuksenkatu 6 Helsinki" returns the expected location. The problem is solved after applying the same solution for the rest of the addresses.

Rendering performance: Every time the user interacts with the map, a code block is called to remove and render again all the SVG elements representing violation bubbles. The number of elements in a month's data is approximately 5000, and the map event can be triggered multiple times in just one second, resulting in the delay experienced by the draft evaluation participants.

Compared to SVG when working with heavy graphics, HTML Canvas is proven to be faster and less resource consuming [35] [18], as it renders all features in one raster element. The downside of Canvas is its content becomes pixelated when scaling, and not capable to neither manipulate nor respond to user interaction.

The solution for improving performance is the mix usage of both techniques. A layer of HTML Canvas is rendered when the map is at wide zoom levels or is moving, or explained differently, when the user does not interact with the violation bubbles. When the map is not moving and zoom level is greater than 14, the SVG layer is available, invisible and fits the features of the canvas layer under, just to receive mouse events from the user. The delay is remarkably reduced, as the number of SVG elements to redraw is relatively less.

4.2.9 Usability test

The setup, conduction and results of usability testing will be explained in detail in the next chapter.

Chapter 5

Evaluation

This chapter aims to measure how well the prototype answers to the two research questions: *"How to visualize massive amounts of data, but at the same time, preserve visual quality and acceptable system performance?"* and *"How can visualization help users gain insights, which are meaningful to their personal interests?"*. Therefore, the evaluation is conducted with the focus on the tool performance, and the effectiveness of the tool on information discovery and insight creation.

5.1 Performance benchmark

5.1.1 Criteria and measures

After the "Optimize" stage, the performance benchmark is conducted to analyze how effective the new rendering mechanism is against the old one. Take the draft evaluation result as the base where the testers experience a noticeable delay between each action, this benchmark examines the improvement of two metrics:

- **Execution time** measures how fast the tool responds to the user interaction, in second. The lower this number, the better. Hoxmeier and DiCesare (2000) suggest 12 seconds is the threshold before user satisfaction deteriorates [31].
- **Frame rate** indicates how smooth the interaction is, measured by frame per second (FPS). 60 FPS is the ideal number, and 24 FPS is an acceptable goal [3].

5.1.2 Setup and conduction

The test is taken using a Macbook Pro retina Mid 2015, in Chrome browser version 81.0.4044.138.

One set of performance trials includes: *data selection* with different period length, *map navigation* (zooming and panning at different zoom level), and *explorer interaction* (moving and resizing the exploring area). Instead of manually done, each action is performed programmatically via the built in function of Mapbox such as *panTo* or *zoomTo*, to ensure the consistency between all tests. The numbers in Figure 5.3 are the average of five trials taken.

The execution time is measured by taking the difference of two timestamps, generated when the action starts and finishes by creating a *Date* object in Javascript. Obtaining the FPS is relatively trickier, as the actions need to be recorded using the Chrome developer tools first, then the numbers are manually read from there [4].

5.1.3 Result

As the result in Figure 5.3 shown, rendering using HTML Canvas is faster compared to SVG, in both loading the data and performing interactions. The change is most noticeable when the action involves a large number of data points, such as zooming the map wider, or selecting a longer period. When more than three consecutive months of data is selected, the tool only renders after the map stops moving, therefore the execution time is even shorter than in the case of a single month. This change makes the actions like panning the map when selecting 6 month data possible, which was unusable before.

Unfortunately, the FPS after optimized is slightly raised, but still below the acceptable level in the majority of the basic actions. Only explorer related interactions for one month data satisfies the standard. "Jitter", or inconsistent frame rate is also a problem, which affects user satisfaction [3]. The large difference between lowest and highest fps is detected when navigating the map at 6 month data, while moving the map is smooth as no violation bubbles is rendered but the performance falls dramatically when the browser starts drawing.

Action	Before optimize			After optimize		
	Time*	FPS min	FPS max	Time*	FPS min	FPS max
First time load**	9.24	-	-	5.39	-	-
Select a period						
1 month	3.85	-	-	1.99	-	-
3 month	5.1	-	-	2.15	-	-
6 month	5.58	-	-	2.31	-	-
12 month	6.72	-	-	2.69	-	-
Interaction - 1 month data						
Zoom out from level 12 to 11	11.56	2	3	1.9	4	5
Move the map at zoom 11	12.73	1	2	1.84	4	6
Zoom in from level 11 to 14.5	6.45	2	3	2.4	4	8
Move the map at zoom 14.5	2.96	11	24	1.33	4	12
Move the explorer		14	27		27	50
Resize explorer to 700m		18	33		23	44
Interaction - 6 month data						
Zoom out from level 12 to 11	23.78	1	3	1.54	2	40
Move the map at zoom 11	24.86	1	2	1.04	2	29
Zoom in from level 11 to 14.5	12.68	1	6	2.98	2	46
Move the map at zoom 14.5	5.43	6	9	3.17	2	10
Move the explorer	-	14	21	-	16	30
Resize explorer to 700m	-	7	17	-	10	24

* Execution time measured in seconds
** From the page starts loading until the map idles, and the user can interact

Figure 5.1: Performance benchmark result

5.2 Usability test

5.2.1 Criteria and measures

The insight creation and the usability of the tool are analyzed based on the framework of Koua and Kraak (2004) on geovisualization evaluation, and the methodology on insight-based evaluation proposed by Saraiya et. al. (2005) [40] [63].

The insights from the interactions and verbal expression of the testers are encoded with the following characteristics:

- **Observation** - The fact which the participant discovers from the data.
- **Category and complexity** - An insight can fall into three categories: *classify* (the awareness of the data grouping and arrangement), *compare* (the recognition of the relationships between different records or parts of the data), and *reflect* (the construction of rules and patterns, or the connection to personal experience).
- **Time** - The time taken to reach the first insight *into information* after completing the tutorial.
- **Correctness** - The inaccurate observation caused by misinterpreting either the visual encoding or on-screen components

Three criteria for evaluating the user interface includes *effectiveness* (the functionalities of the tool), *usefulness* (how the tool satisfies user expectations), and *user reaction* (the personal attitude of the testers towards the tool). Figure 5.2 lists the criteria and their involved metrics.

Criteria	**Measurement**
Effectiveness	Total time spent, including introduction
	Number of features covered
	Assistance frequency
	Insight discovered - number and quality
	Tasks completion time and accuracy
	Usability issues encountered
Usefulness	Pre-test versus post-test result
	Number of features suggested
User reaction	User attitude and expression
	Subjective opinion and comments

Figure 5.2: Criteria and measurement for user interface evaluation

5.2.2 Participants

The profile of the participants is divided into three categories: parking experts, drivers and non-drivers. "Parking experts" are people working in the parking industry, with the assumption that they have good awareness about the parking fines in general and the situation around their frequently visited places in particular. "Drivers" are people who have driven and parked in Helsinki. As people, who are not driving, but often travel by personal car, can be potential users of the tool, they are also included in the "non-drivers" group.

Nielsen (2000) suggests testing with five users is enough to discover the majority of usability problems, which are likely to happen in a product [53]. Among five participants recruited, there are 1 parking expert, 3 drivers and 1 non-driver, with the average age of 31. Two of them are native Finnish speakers, the others know the language at a very basic level.

5.2.3 Setup and conduction

Due to the social distancing situation caused by COVID-19, all test sessions are taken online. To support the observation, the participants are asked to share their screen and verbally express their thoughts, expectations and intentions while interacting with the tool, following the thinking aloud protocol [33]. The sessions are recorded using a smartphone on a tripod, with the consent of all testers.

Before the test, the participants are given a set of questions to assess their experience with parking fines. The final pre-test question asks each tester to name a popular place of their choice, and estimate the likelihood of receiving a parking fine there. Then, the participant received the URL to access the tool, and asked to freely interact with it. Neither instructions nor guidance is provided, as the introduction section is shown when the tool loads for the first time and the testers can choose to read or skip it. When the participant finishes, two tasks are given:

- "Tell where the fines with reason *Sallitun pysäköintiajan ylitys kiekkopaikalla* (Parking exceeds time limit) happened commonly during *the three last months of 2019*". This task covers the period and reason filter, as well as checks if the participant interprets the visual encoding correctly.
- "Examine the location you mentioned in the last pre-test question again with the tool. Does your estimation about the possibility of receiving

parking tickets there changed?". This task partially verifies the insight created by the tool.

The test sessions conclude with the discussion on the experience of the participant with the tool, and suggestions for improvement.

5.2.4 Result - on insight

Measurement	Min	Average	Max
Time spent	14:47	17:20	22:33
Introduction	0:12	1:31	2:27
Free interaction	6:35	9:05	11:51
Task #1	0:37	1:08	1:36
Task #2	0:17	0:57	2:26
Discussion	1:25	2:36	4:52
Number of features covered	6	9	11
Number of questions asked	3	5	9
Average assistance duration	0:30	1:40	5:03
Time until first insight	0:08	1:31	4:45
Number of insights	6	11.4	16
Classify	2	2.8	5
Compare	2	3.8	8
Reflect	2	4.8	8
Number of wrong insights	1	1.4	4
Number of usability issues	2	5	7
Number of features suggested	4	6.8	13

Figure 5.3: Usability evaluation result, in number

Category and complexity

After freely interacting with the tool, three testers gained more than 10 insights. Two others with 6 and 8 insights, both spent a long time on figuring the feature out by themselves, as well as running into several usability issues.

Participant	1	2	3	4	5	All
Classify	5	2	2	2	3	**14**
Compare	8	4	2	2	3	**19**
Reflect	3	7	2	4	8	**24**
Total insights	**16**	**13**	**6**	**8**	**14**	**57**

Figure 5.4: Number of insights collected

> ”440 is how many people got fined, right?”

Classify: The expression above is an example of a Classify insight, emerging when the user identifies how the data is arranged on the screen, or say in another way, the logic of visual encoding. This is the simplest kind of insight, which arrives early. If most of the insights fall into this category, that means the participants need more effort to understand the data rather than to gain actual information from it. Fortunately, that did not happen in any of the five test sessions.

> ”It seems like there are more fines here in the winter than in the summer”

Compare: The number of insights in this category measures how effective the tool can be used to detect the similarities and differences between data points. Among the six relationship types suggested by Koponen and Hildén (2019) [38], only *connection* is not available. The participants made the most comparison on *location* and *number*, via the bubble size and bubble density in certain areas. Perceiving the *rank* of violations by popularity is trivial according to one participant, as the tool automatically sorts the reason with the most occurrence to top.

Time insights require multiple actions involving the timeline component. There were comments about its considerably smaller size compared to the map view, which caused several problem in action recognition and interaction. As a result, this comparison was relatively rare. One among two insights recorded brought excitement to a participant as she observed the bubbles around her office changed after selecting different time period.

Comparing *category* via color for violation is commonly taken, but only accurate when user applies the reason filter. The error started happening when at least two reasons with the same color appeared, or there are multiple reasons in a location, which result in the bubble having a color, which is totally different from the reasons in the info box.

> "Our CEO's parking ticket might be among one of these"

Reflect: These are the most complex and valuable insights in the context of this thesis, where the participants combine the understanding about parking violation data through the tool and their own parking experiences before. Among five participants, the parking expert and two drivers found the insights belong to this category more than any other types.

All of the testers tried to analyze the situation around their home at least once. One of the drivers knew exactly around an area there is no parking time limit, so he checked with the tool if there are any of the mentioned violations recorded. The only driver who received parking tickets before, tried to look whether there were many others having the same problems. One heard the story of a colleague getting parking fine near her office, and managed to point out when and where it was.

The parking expert, on the other hand, spent more time examining at the map at wide zoom level than his frequently visited places. He usually gives more detail or guessed why the problems happened before zooming in any specific area.

While testing, the participants discovered various use cases for the tool. "A lot of people think that the warden is not active in this area so they tried their luck. If they see this map, they might change their mind" - the parking expert said. One driver tried to check an area he plans to go, and found parking at the spots reserved for buses were common there. Another driver would like to analyze the fines around where he parks the car in the future.

Time

> "How about I just skip the tutorial and just play with it to see how it goes?"

One test session varied from 14:47 to 22:33, with the majority of duration being on free interaction.

In most test sessions, it took just less than a minute to reach the first insight. There is an exceptional case where the first insight came after nearly 5 minutes. That happened to the non-driver when she requested the explanation for the meaning of some violation reasons and played with the supporting features (collapsing the sidebar) before actually taking the first action.

The action, which the participants spent the most time on, was locating the desired place on the map. Plus the delay on moving and zooming the map, the longest search took 38 seconds until one tester recognized his place based on a popular pizzeria nearby.

Correctness

> "Why I can't see this red? there are only two colors in that box.
> May be some data is missing?"

A common issue was reported by 3 out of 5 participants: they could not find the reference for several colors that appeared on the map at either the info box or the reason list. When explained, they all understood about the idea of color mixing, but still expressed their confusion when using it. A participant thought that the data is incomplete as he failed to find the corresponding color. However, one positive effect is out-of-the-list colors triggered the curiosity of the testers more.

The title of the timeline makes two testers think that it indicates the selected time period instead of the whole time coverage of the data set. One participant tried to click around it to change the time, but ended up with collapsing the panel instead.

Task execution

> "What, there is not that much? So they mainly just get the ticket
> for dumping their cars here."

The participants manage to quickly locate the actions they need to complete the tasks. The first task - asking the testers to filter a specific reason and look for its commonly violated area - took roughly a minute to complete, and the tester easily zoomed in to the location where the density of the bubbles are high. The execution time for the second task varied, as one of the participants struggled to locate her frequent grocery store on the map. During this task, all testers managed to compare their estimation in the last pre-test question. Two participants were surprised to discover the amount of parking tickets around their places is much less than their expectation. One participant who was excited with the explorer feature before, used it to draw his conclusion in just 17 seconds.

5.2.5 Result - on user interface

The instruction

In the draft evaluation, it took a considerable amount of time to explain about the tool. Also, one of the testers struggled to find several features. Therefore the introduction is added with the intention that the test facilitator does not have to repeat himself, and the tester can always open it when they are stuck.

On average, the participants spent from 1 to 2 minutes on the on-screen instruction. One tester closed it almost right away, but she was the person who asked questions and required assistance the least. The participant complaint about the text length the most was the one spending most time reading. There were two users who actually opened the instruction again later.

Feature coverage

This metric examines if the participants managed to take all of the action supported by the tool during the free interaction time. The low number means certain features are unimportant to the users, or some are hard to recognize. There are 11 features in total, including:

1. Introduction
2. Select a time period
3. Filter the reason
4. Open the popup to read the reason category
5. Move (pan) the map
6. Zoom the map
7. Click on a bubble for detailed information

8. Activate explorer mode
9. Move the explorer circle
10. Resize the explorer circle
11. Change the language

Participant	1	2	3	4	5
Number of features covered	11	10	11	6	7
Unused features	-	#4	-	#2 #4 #9 #10 #11	#1 #3 #4 #11

Figure 5.5: Summary of feature coverage by each tester

Supporting features are not used by all users. As mentioned, a participant closed the introduction (1) almost right away, so it is considered unused once. One person did not notice that the timeline (2) is clickable, while several others had troubles using it. Four testers tried the reason filter (3) but their attitude varied. The small button for displaying the reason category (4) is ignored in three sessions, making it the feature left unnoticed the most. While the language switcher (11) is completely untouched by native Finnish users, the other three testers clicked it the earliest.

As having experience with map view, the map related interactions (5, 6 and 7) are obvious to all participants. They also managed to bring the explorer into the map (8) several times for each person. Four out of five testers figured out the explorer circle can be modified (9 and 10), and are excited when watching the bubbles and the info box updated accordingly. Only one participant did not recognize it. He changed the exploring area by clicking different points on the map instead.

Usability issues

Navigating the map manually is the most complaint feature, as several users have trouble with finding exactly the location of even their familiar places. In addition, the delay and lag mentioned in the Performance benchmark result also extends the time needed to complete map related actions. Unnecessary features such as map rotation, or 3D mode should be disabled, as it is unable to revert the map to the workable state and the tool needs to be restarted.

The timeline is reported as ”hard to notice”, or ”not recognized that it is clickable”. Also, the title which says ”Period: Jan 2018 - Dec 2019” is misleading. Two out of five testers thought it specifies the current period, not the whole timeline of the data set.

How the reason filtering works - leaving all check boxes empty to show all reasons - is a controversial feature, as three participants had no problem, while the other two strongly said it is counter intuitive. They suggested it should be better to have a ”Deselect all” button, and selecting all checkboxes will display every reason on the map.

With the explorer, the testers had trouble at resizing it to an exact size (e.g., 500m). One participant would like to have the center point visible to ”aim” the explorer precisely to the place he wanted.

User reactions and suggestions

Participant	1	2	3	4	5	All
UI	6	1	6	1	2	**16**
Visual encoding	1	1	-	-	1	**3**
New feature	6	3	1	3	2	**15**
Total suggestions	**13**	**5**	**7**	**4**	**5**	**34**

Figure 5.6: Number of suggestions collected

Before the discussion time, the participants often expressed their expectation right after they encountered any issues. As Figure 5.6, the suggestions are about the improvement of either user interface or visual encoding, or the proposal of new features for the tool.

Many of the ideas reflect from the experience of the user with similar tools. An example is the most requested feature - adding an address search box to locate a specific place easier, which based on the everyday usage of all participants with Google Maps. Two participants would like to have the zoom buttons instead of using the two-finger swiping with the touch pad or the scroll wheel of the mouse.

Another category of suggestions is adding alternative options for current controls. As observed in some test sessions, dragging actions such as brushing the timeline to select multiple months, or modifying the explore circle is only fast, but not accurate. There were several demands for a drop down to change the time period, or text box to write down precisely the desired explorer radius.

The least but most interesting are the suggestions of integrating with other products. One driver imagined he could see the fines happen along with the available parking spots in Parkopedia service[1]. The parking expert, who currently works for a mobile application, opened up the opportunity to adapt the tool to mobile platforms.

[1] https://www.parkopedia.com/

Chapter 6

Conclusions

This chapter discusses the contribution of this thesis, analyzes the success of the work based on the research questions, following with its limitations, and opportunities for further development.

6.1 Contribution

With the goal of answering two research questions raised in the first chapter, this thesis first collects and summarizes the background related to the domain of parking, along with information visualization theories and practices. To further support the implementation, the next step reviews the prior works from Fry (2004) and Munzner (2009), and adapts a process, which involves both technical aspects and human participation. An interactive visualization is created as a tool for exploring the parking violation in Helsinki using the open data set from the city. The performance and capability of creating insights of the tool are evaluated based on both benchmarking metrics and user feedback.

For public usage, the tool can be added to the showcase gallery of the data set, in the official Helsinki Region Infoshare website[1].

[1] `https://hri.fi/`

6.2 Responses to the research questions

The Helsinki parking fine exploration tool only partially answers the first research question: *"How to visualize massive amounts of data, but at the same time, preserve visual quality and acceptable system performance?"*. By using different strategies in pre-processing the raw data, the input of the visualization is minimized, which significantly lessens the loading effort. Rendering speed is improved to a certain level, but not enough to pass the standard metrics and the notice of several sharp-eyed users. Expected visual cluttering is reduced by grouping the violations by location and changing their size based on the map zoom level. During the usability test, the users can quickly make comparisons across the city map. However, the choice of using colors is not effective and leads to considerable confusion.

On the bright side, the tool satisfies the second question *"How can visualization help users gain insights, which are meaningful to their personal interests?"* by representing the data in a familiar map view and introducing the explorer feature. During free interaction, the participants are able to intuitively examine their well-known places such as home or office. The users can also quickly make comparison between different areas, or check the correctness of their assumptions. All the testers report that the tool is fun and exciting to use, as they manage to discover several "A-ha!" moments while interacting, despite some usability issues encountered.

6.3 Limitations and further development

Reusability / Extendability

With the adapted process and implementation, it is possible to extend the scale of the data, like adding the records from other years. Similar user interfaces can be used for different data set having geolocation information and categorized.

However, for a larger amount of data compared to approximately 150 000 records in this thesis, more performance issues might emerge. The current implementation loads the whole at once. For every action the user takes, the browser solely executes three tasks: filtering the records, summarizing metrics, then painting the data points and charts to the screen. To optimize the process, a potential solution is to store the data in a separate database, and perform expensive calculations from there. The rendering step can be done through WebGL, as there are many libraries based on the technology, which can draw millions of data point smoothly such as *PixiJS*[2] or *deck.gl*[3].

Evaluation

The number of testers involved is enough to discover usability issues. However, only one non-driver or one parking expert is too small to represent their persona group. As a result, it is possible that running the usability test with more users of the same category might produce different results. The owners of the data set, representing the urban planner profile, is not available for the usability test. Therefore, how the tool matches their expectation is still an open question.

Usage

The work of the thesis only reveals where and why there are violations. Certain knowledge could be deduced from it, such as the wardens are more active, or the traffic signs are poorly placed in specific areas, etc. A false insight, which have not appeared during the test sessions, but noticeable, is the users might think the map shows where it is "safe" to park. Before releasing to the public, there should be clear communication on the tool itself about its capability, and what the users can expect from it. Further, the cooperation with other parking service providers, such as the mobile parking application suggested by one of the testers, can improve the comprehensiveness of the information, and open the opportunities for novel use cases.

[2] `https://www.pixijs.com/`
[3] `https://github.com/visgl/deck.gl`

Bibliography

[1] Anastasiadou, M., Dimitriou, D. J., Fredianakis, A., Lagoudakis, E., Traxanatzi, G., and Tsagarakis, K. P. Determining the parking fee using the contingent valuation methodology. *Journal of Urban Planning and Development 135*, 3 (2009), 116–124.

[2] Anscombe, F. J. Graphs in statistical analysis. *The american statistician 27*, 1 (1973), 17–21.

[3] Bakaus, P. The illusion of motion, 2014. `https://paulbakaus.com/tutorials/performance/the-illusion-of-motion/` Accessed 14.05.2020.

[4] Basques, K. Performance analysis reference, 2019. `https://developers.google.com/web/tools/chrome-devtools/evaluate-performance/reference#fps` Accessed 14.05.2020.

[5] Battersby, S. Map projections. *The Geographic Information Science & Technology Body of Knowledge* (2017).

[6] Bertin, J. Semiology of graphics; diagrams networks maps. Tech. rep., 1983.

[7] Boulos, M. N. K., Warren, J., Gong, J., and Yue, P. Web gis in practice viii: Html5 and the canvas element for interactive online mapping, 2010.

[8] Burgess, S. Confusing signs make parking tickets a scam, driver argues, 2019. `https://www.cbc.ca/news/canada/ottawa/confusing-parking-sign-unfair-ticket-ottawa-1.4684319` Accessed 20.11.2019.

[9] Börner, K., Bueckle, A., and Ginda, M. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments.

Proceedings of the National Academy of Sciences 116, 6 (2019), 1857–1864.

[10] Cairo, A. *The truthful art: data, charts, and maps for communication.* New Riders, 2016.

[11] Cambridge Dictionary. Parking. https://dictionary.cambridge.org/dictionary/english/parking/ Accessed 18.11.2019.

[12] Card, M. *Readings in information visualization: using vision to think.* Morgan Kaufmann, 1999.

[13] CarRentals.com Blog. The real cost of parking tickets: 16 u.s. cities make over $1.4 billion in annual parking ticket revenue. https://www.carrentals.com/blog/parking-tickets-cost-americans Accessed 20.11.2019.

[14] Chen, M., Ebert, D., Hagen, H., Laramee, R. S., Van Liere, R., Ma, K.-L., Ribarsky, W., Scheuermann, G., and Silver, D. Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications 29*, 1 (2008), 12–19.

[15] Cleveland, W. S., and McGill, R. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association 79*, 387 (1984), 531–554.

[16] Consumer Insights, M. C. Attention spans.

[17] Cullinane, K., and Polak, J. Illegal parking and the enforcement of parking regulations: causes, effects and interactions. *Transport Reviews 12*, 1 (1992), 49–75.

[18] Dengler, P. Thoughts on when to use canvas and svg, 2011. https://docs.microsoft.com/en-us/archive/blogs/ie/thoughts-on-when-to-use-canvas-and-svg/ Accessed 05.05.2020.

[19] Eick, S. G., and Wills, G. J. High interaction graphics. *European journal of operational research 81*, 3 (1995), 445–459.

[20] Fanning, G. Woman driver calls for action to tackle illegal parking a bus stop in wishaw's stewarton street, 2017. https://www.dailyrecord.co.uk/news/local-news/woman-driver-calls-action-tackle-9729234/ Accessed 20.11.2019.

[21] Ferster, B. *Interactive visualization: Insight through inquiry.* MIT Press, 2012.

[22] Few, S. Data visualization for human perception. *The Encyclopedia of Human-Computer Interaction, 2nd Ed.* (2013).

[23] Foley, J. D., Dam, A. v., Feiner, S. K., Hughes, J. F., and Carter, M. P. Computer graphics: Principles and practice, in c. *Color Research and Application 22*, 1 (1997), 65–65.

[24] Friendly, M. Milestones in the history of data visualization: A case study in statistical historiography. In *Classification: The Ubiquitous Challenge*, C. Weihs and W. Gaul, Eds. Springer, New York, 2005, pp. 34–52.

[25] Fry, B. J. *Computational information design.* PhD thesis, Massachusetts Institute of Technology, 2004.

[26] Galatioto, F., and Bell, M. Simulation of illegal double parking: quantifying the traffic and pollutant impacts. In *4th International SIIV Congress, Palermo, Italy* (2007).

[27] Heer, J., Bostock, M., Ogievetsky, V., et al. A tour through the visualization zoo. *Commun. Acm 53*, 6 (2010), 59–67.

[28] History.com Editors. Automobile history, 2018. `https://www.history.com/topics/inventions/automobiles` Accessed 18.11.2019.

[29] History.com Editors. World's first parking meter installed, 2019. `https://www.history.com/this-day-in-history/worlds-first-parking-meter-installed` Accessed 18.11.2019.

[30] Howard, A. B. The art and science of data-driven journalism.

[31] Hoxmeier, J. A., and DiCesare, C. System response time and user satisfaction: An experimental study of browser-based applications. *AMCIS 2000 Proceedings* (2000), 347.

[32] Iliinsky, N., and Steele, J. *Designing data visualizations: Representing informational Relationships.* " O'Reilly Media, Inc.", 2011.

[33] Jääskeläinen, R. Think-aloud protocol. *Handbook of translation studies 1* (2010), 371–374.

[34] Ji, Y., Deng, W., Wang, W., and Liu, G. Two-phased parking choice model for pre-trip parking guidance system. Tech. rep., 2007.

[35] Johnson, D. W., and Jankun-Kelly, T. A scalability study of web-native information visualization. In *Graphics Interface* (2008), pp. 163–168.

[36] Kennedy, M., Kopp, S., et al. *Understanding map projections.* Esri Redlands, CA, 2000.

[37] Kirk, A. Discussion: Storytelling and success stories, 2013. `https://www.visualisingdata.com/2013/04/discussion-storytelling-and-success-stories` Accessed 01.11.2019.

[38] Koponen, J., and Hildén, J. *Data visualization handbook.* Aalto korkeakoulusäätiö, 2019.

[39] Kosara, R. Visualization criticism-the missing link between information visualization and art. In *2007 11th International Conference Information Visualization (IV'07)* (2007), IEEE, pp. 631–636.

[40] Koua, E. L., and Kraak, M.-J. A usability framework for the design and evaluation of an exploratory geovisualization environment. In *Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004.* (2004), IEEE, pp. 153–158.

[41] Kraak, M.-J. Geovisualization illustrated. *ISPRS journal of photogrammetry and remote sensing 57*, 5-6 (2003), 390–399.

[42] MacEachren, A. M., and Kraak, M.-J. Research challenges in geovisualization. *Cartography and geographic information science 28*, 1 (2001), 3–12.

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

[44] Martino, M. Can you decipher these confusing parking signs?, 2019. `https://www.abc.net.au/news/2019-03-21/quiz-sydney-confusing-parking-signs/10683014` Accessed 20.11.2019.

[45] Masud, L., Valsecchi, F., Ciuccarelli, P., Ricci, D., and Caviglia, G. From data to knowledge-visualizations as transformation processes within the data-information-knowledge continuum. In *2010*

14th International Conference Information Visualisation (2010), IEEE, pp. 445–449.

[46] May, A. Parking enforcement[seminar report]. *Traffic Engineering and Control 23* (1982), 345.

[47] Mehaffy, M., Porta, S., Rofe, Y., and Salingaros, N. Urban nuclei and the geometry of streets: The 'emergent neighborhoods' model. *Urban Design International 15*, 1 (2010), 22–46.

[48] Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review 63*, 2 (1956), 81.

[49] Mishunov, D. Why perceived performance matters, part 1: The perception of time, 2015. `https://www.smashingmagazine.com/2015/09/why-performance-matters-the-perception-of-time/` Accessed 01.11.2019.

[50] Morphocode. The 5-minute walk. `https://morphocode.com/the-5-minute-walk/` Accessed 10.03.2020.

[51] Munzner, T. A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics 15*, 6 (2009), 921–928.

[52] Nielsen, J. *Usability engineering*. Elsevier, 1994.

[53] Nielsen, J. Why you only need to test with 5 users, 2000.

[54] Nielsen, J., and Molich, R. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (1990), ACM, pp. 249–256.

[55] Nöllenburg, M. Geographic visualization. In *Human-centered visualization environments*. Springer, 2007, pp. 257–294.

[56] Obe, R., and Hsu, L. Postgis in action. *GEOInformatics 14*, 8 (2011), 30.

[57] O'Malley, P. Theorizing fines. *Punishment & Society 11*, 1 (2009), 67–83.

[58] Oppio, A., Bottero, M., Giordano, G., and Arcidiacono, A. A multi-methodological evaluation approach for assessing the impact of neighbourhood quality on public health. *Epidemiol Prev 40*, 3-4 (2016), 249–256.

[59] Presnyakova, A. Investigating the visualization process. development of an integrated framework, 2017.

[60] Rapoza, J. The very real costs of bad website performance.

[61] Reas, C., and Fry, B. *Processing: a programming handbook for visual designers and artists*. Mit Press, 2007.

[62] Roth, R. E. Visual variables. *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology* (2016), 1–11.

[63] Saraiya, P., North, C., and Duca, K. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE transactions on visualization and computer graphics 11*, 4 (2005), 443–456.

[64] Shneiderman, B. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE symposium on visual languages* (1996), IEEE, pp. 336–343.

[65] Shoup, D. C. The high cost of free parking. *Journal of planning education and research 17*, 1 (1997), 3–20.

[66] Shoup, D. C. Cruising for parking. *Transport Policy 13*, 6 (2006), 479–486.

[67] Simićević, J., Milosavljević, N., Maletić, G., and Kaplanović, S. Defining parking price based on users' attitudes. *Transport Policy 23* (2012), 70–78.

[68] Smith, B. Escaping the'60s. *Parking Professional 30*, 4 (2014).

[69] Snyder, J. P. *Flattening the earth: two thousand years of map projections*. University of Chicago Press, 1997.

[70] Stevens, S. S., et al. On the theory of scales of measurement.

[71] Swayne, D. F., and Klinke, S. Introduction to the special issue on interactive graphical data analysis: What is interaction? *Computational Statistics 14*, 1 (1999), 1–6.

[72] Talen, E. The geospatial dimension in urban design. *Journal of Urban Design 16*, 1 (2011), 127–149.

[73] Thanh, T. T. M., and Friedrich, H. Legalizing the illegal parking, a solution for parking scarcity in developing countries. *Transportation Research Procedia 25* (2017), 4950–4965.

[74] Tory, M., and Moller, T. Evaluating visualizations: do expert reviews work? *IEEE computer graphics and applications 25*, 5 (2005), 8–11.

[75] van der Waerden, P., Timmermans, H., and de Bruin-Verhoeven, M. Car drivers' characteristics and the maximum walking distance between parking facility and final destination. *Journal of transport and land use 10*, 1 (2017), 1–11.

[76] Vaughan-Nichols, S. J. Will vector graphics finally make it on the web? *Computer 34*, 12 (2001), 22–24.

[77] Virzi, R. A., Sokolov, J. L., and Karis, D. Usability problem identification using both low-and high-fidelity prototypes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (1996), ACM, pp. 236–243.

[78] W3Schools. Html introduction. `https://www.w3schools.com/html/html_intro.asp` Accessed 20.03.2020.

[79] Yang, Q., Snyder, J., and Tobler, W. *Map projection transformation: principles and applications.* CRC Press, 1999.

[80] Yi, J. S., ah Kang, Y., and Stasko, J. Toward a deeper understanding of the role of interaction in information visualization. *IEEE transactions on visualization and computer graphics 13*, 6 (2007), 1224–1231.

[81] Yle. Helsinki losing millions in unpaid parking fines issued to foreign vehicles, 2018. `https://yle.fi/uutiset/osasto/news/helsinki_losing_millions_in_unpaid_parking_fines_issued_to_foreign_vehicles/10281783` Accessed 20.11.2019.

[82] Young, W. *A compilation of research into parking policy.* US Department of Commerce, National Techncal Information Service, 1990.

[83] Zimmermann, M. The nervous system in the context of information theory. In *Human physiology.* Springer, 1989, pp. 166–173.

Appendix A

Image source

Figure 2.1(a): Extracted from http://www.brooklynsigns.com/hpd-signs-nyc/no-parking-signs

Figure 2.1(b): Self created material

Figure 2.1(c): Extracted from https://www.parkme.com/lot/102350/brunswick-square-car-park-london-uk

Figure 2.2: Extract from https://yle.fi/uutiset/osasto/news/helsinki_losing_millions_in_unpaid_parking_fines_issued_to_foreign_vehicles/10281783. Credit: Antti Haavisto / Yle.

Figure 2.3: Self created material. Data from Anscombe (1973)

Figure 2.4: Self created material. Data from Anscombe (1973)

Figure 2.5: Koponen and Hildén (2019), p.34

Figure 2.6: Börner (2019)

Figure 2.7: Self created material. Based on Mackinlay (1986), p.125

Figure 2.8: Self created material

Figure 2.9: Self created material

Figure 2.10: Kennedy (2000), p.8 and p.16

Figure 2.11: Extracted from https://www.gapminder.org/tools/

Figure 2.12: Extracted from https://flowingdata.com/2019/02/27/where-your-job-is-most-popular/

Figure 2.13: Extracted from https://morphocode.com/the-5-minute-walk/

Figure 2.14: Self created material

Figure 2.15: Self created material

Figure 3.1: Fry (2004), p.13

Figure 3.2: Munzner (2009), p.923

Figure 3.3: Self created material. Adapted from Presnyakova (2017)

Figure 3.4: Self created material

Figure 4.1: Self created material

Figure 4.2: Self created material

Figure 4.3: Self created material

Figure 4.4: Self created material

Figure 4.5: Self created material

Figure 5.3: Self created material

Figure 5.2: Self created material

Figure 5.3: Self created material

Figure 5.4: Self created material

Figure 5.5: Self created material

Figure 5.6: Self created material

Appendix B

Parking violation reasons in Helsinki

Code	Finnish	English
0100	Pysäyttäminen kielletty liikennemerkin noudattamatta jättäminen	Stopping violates the traffic sign
0101	Pysäyttäminen kielletty liikennemerkin noudattamatta jättäminen	Stopping violates the traffic sign
0102	Ei koske cd-autoja	Does not apply to CD (Club Diplomat)-cars
0103	Ei koske kuormaavia eikä purkavia kuorma- ja pakettiautoja	Does not apply to loading and unloading lorries/trucks and vans
0104	Ei koske city car clubin autoja	Does not apply to city car club cars
0105	Ei koske linja-autoja	Does not apply to buses
0106	Ei koske sähköautoja	Does not apply to electric cars
0107	Ei koske ministeriön virka-autoa	Does not apply to ministy's (official) car
0108	Ei koske kaupungin virka-autoja	Does not apply to city's (official) cars
0109	Ei koske ambulansseja	Does not apply to ambulances
0110	Ei koske hälytysajoneuvoja	Does not apply to emergency vehicles

0111	Ei koske lähetysautoja ei koske valtion virka-autoja	Does not apply to delivery vehicles does not apply to goverment (official) cars
0112	Ei koske Z-tunnuksia	Does not apply to Z-permits
0113	Ei koske kuormaavia eikä purkavia ajoneuvoja	Does not apply to loading and unloading vehicles
0200	Taksin pysäyttämispaikka/Taksiasema-alue -liikennemerkin noudattamatta jättäminen	Taxi-stop / Not following (or violating) taxi area traffic sign
0201	Taksin pysäyttämispaikka	Taxi-stop
0202	Taksiasema-alue liikennemerkin noudattamatta jättäminen	Not following (or violating) taxi area traffic sign
0300	Pysäköinti kielletty liikennemerkin noudattamatta jättäminen	Parking violates the traffic sign
0301	Pysäköinti kielletty liikennemerkin noudattamatta jättäminen	Parking violates the traffic sign
0302	Ei koske linja-autoja	Does not apply buses
0303	Ei koske ministeriön virka-autoja	Does not apply ministy's official cars
0304	Ei koske poliisiautoja	Does not apply police cars
0306	Ei koske henkilöautoja	Does not apply passenger (normal) cars
0307	Ei koske kaupungin virka-autoja	Does not apply city's (official) cars
0308	koskee kuorma-ja linja-autoja	Lorries/trucks and buses
0309	koskee kuorma-autoja	Lorries/trucks
0310	koskee linja-autoja	Buses
0312	Ei koske kirjastoautoja	Does not apply to library cars/trucks (mobile library)
0313	Vuoropysäköinti, kielletty parittomina päivinä	Prohibited on odd days
0314	Vuoropysäköinti, kielletty parillisina päivinä	Prohibited on even days

0400	Pysäköintikieltoalue liikennemerkin noudattamatta jättäminen	Parking in prohibited area
0401	Pysäköintikieltoalue liikennemerkin noudattamatta jättäminen	Parking in prohibited area
0402	Koskee kuorma-autoja	Lorries/trucks
0404	Koskee kuorma- ja linja-autoja	Lorries/trucks and buses
0405	Pysäköintitunnus ei luettavissa/maksullisuus	Parking permit not readable
0500	Pysäköinti pelastustielle	Parking on emergency access / exit
0600	Pysäköintimaksun suorittamatta jättäminen	Parking without payment
0601	pysäköinti ilman p-tunnusta/p-laitetta/p-lippua	Parking without permit/device/ticket
0602	pysäköinti ilman p-laitetta/p-lippua	Parking without device/ticket
0603	pysäköinti ilman p-tunnusta	Parking without permit
0700	Pysäköintimaksun suorittamatta jättäminen	Parking without payment
0701	pysäköinti ilman p-tunnusta/p-laitetta/p-lippua	Parking without permit/device/ticket
0702	pysäköinti ilman p-laitetta/p-lippua	Parking without device/ticket
0703	pysäköinti ilman p-tunnusta	Parking without permit
0704	pysäköinti ilman p-lippua/p-tunnusta	Parking without ticket/permit
0800	Sallitun pysäköintiajan ylitys	Parking exceeds time limit
0801	Sallitun pysäköintiajan ylitys kiekkopaikalla	Parking exceeds time limit on parking disc area
0803	Sallitun pysäköintiajan ylitys	Parking exceeds time limit
0900	Sallitun pysäköintiajan ylitys	Parking exceeds time limit
0905	Sallitun pysäköintiajan ylitys kiekkopaikalla	Parking exceeds time limit
0907	Sallitun pysäköintiajan ylitys	Parking exceeds time limit on parking disc area

1000	Pysäyttäminen/Pysäköinti kävelykadulle	Stopping / parking on pedestrian zone
1001	Pysäyttäminen kävelykadulle	Stopping on pedestrian zone
1002	Pysäköinti kävelykadulle	Parking on pedestrian zone
1100	Pysäköinti pihakadulle	Parking on yard street
1200	Pysäköinti invalidin ajoneuvolle varatulle paikalle	Parking on disabled driver's vehicle reserved spot
1201	Pysäköinti invapaikalle pysäköintikiellossa	Parking on disabled driver reserved spot
1202	Pysäköinti invapaikalle P-paikalla	Parking only for disabled drivers
1300	Pysäköinti/Pysäyttäminen suojatielle/5 m matkalle ennen suojatietä	Stopping / Parking on / 5m close to crosswalk
1301	Pysäköinti suojatielle	Parking on crosswalk
1302	Pysäköinti 5 m matkalle ennen suojatietä	Parking 5m close to crosswalk
1303	Pysäyttäminen suojatielle	Stopping on crosswalk
1304	Pysäyttäminen 5 m matkalle ennen suojatietä	Stopping 5m close to crosswalk
1400	Pysäköinti/Pysäyttäminen risteykseen/5 m lähempänä risteävän ajorajan lähintä reunaa tai sen ajateltua jatkoa ajoradalle	Parking / Stopping on / 5m close to intersection
1401	Pysäköinti risteykseen	Parking on intersection
1402	Pysäköinti 5 m lähempänä risteävän ajoradan lähintä reunaa tai sen ajateltua jatkoa ajoradalla	Parking 5m close to intersection
1403	Pysäyttäminen risteykseen	Stopping on intersection
1404	Pysäyttäminen 5 m lähempänä risteävän ajoradan lähintä reunaa tai sen ajateltua jatkoa ajoradalla	Stopping 5m close to intersection
1500	Pysäköinti/Pysäyttäminen jalkakäytävälle/pyörätielle	Parking / Stopping on pavement / bikepath
1501	Pysäköinti jalkakäytävälle	Parking on pavement

1502	Pysäköinti pyörätielle	Parking on bikepath
1503	Pysäköinti yhdistetylle jalkakäytävä/pyörätielle	Parking on combined pavement/bikepath
1504	Pysäyttäminen jalkakäytävälle	Stopping on pavement
1505	Pysäyttäminen pyörätielle	Stopping on bikepath
1506	Pysäyttäminen yhdistetylle jalkakäytävä/pyörätielle	Stopping on combined pavement/bikepath
1600	Luvaton pysäköinti yksityisalueelle	Parking in private area without permission
1700	Luvaton maastopysäköinti	Parking in non-traffic area without permission
1800	Pysäköinti/Pysäyttäminen pysäkkialueelle	Parking / Stopping at stop-zone (meant for bus/public transportation)
1801	Pysäköinti pysäkkialueelle	Parking at stop-zone (meant for bus/public transportation)
1802	Pysäyttäminen pysäkkialueelle	Stopping at stop-zone (meant for bus/public transportation)
1803	Pysäköinti matkailuliikenteelle varatulle pysäkkialueelle	Parking at zone reserved for tourist transportation
1804	Pysäyttäminen matkailuliikenteelle varatulle pysäkkialueelle	Stopping at zone reserved for tourist transportation
1900	Pysäköinti kiinteistölle johtavan ajotien kohdalle / siten, että ajoneuvoliikenne kiinteistölle tai sieltä pois on oleellisesti vaikeutunut	Parking on the access road or at the gateway to a building that hamper vehicular access to the property
1901	Pysäköinti kiinteistölle johtavan ajotien kohdalle	Parking on the access road to a property
1902	Pysäköinti kiinteistölle/kiinteistöltä johtavan ajotien kohdalle	Parking on the access/exit road to a property
1903	Kiinteistölle/kiinteistöltä ajoa haittaava pysäköinti	Parking that hampers the access/exit road to/from a property
2000	Pysäköintiohjeen noudattamatta jättäminen	Parking not following the specified direction
2001	Pysäköinti ajosuunnan vastaisesti	Parking opposite travel direction

2002	Pysäyttäminen ajosuunnan vastaisesti	Stopping opposite travel direction
2003	Vinopysäköinti	Slant parking
2004	Poikittaispysäköinti	Diagonal parking
2007	Uusi	New
2100	Pysäköinti/Pysäyttäminen liikennettä haittaavasti/tarpeettoman kauas ajoradan reunasta/muun kuin kaksipyöräisen ajoneuvon rinnalla	Parking/Stopping that hampers traffic/unnecessarily far from the edge of the road/next to a non-two-wheel-vehicle
2101	Pysäköinti tarpeettoman kaukana ajoradan reunasta	Parking unnecessarily far from the edge of the road
2102	Pysäyttäminen tarpeettoman kaukana ajoradan reunasta	Stopping unnecessarily far from the edge of the road
2103	Rinnallepysäköinti	Parking next to another vehicle (Doubleparking)
2104	Muuta liikennettä haittaava pysäköinti	Parking that hampers other traffic
2200	Pysäköinti merkityn pysäköintipaikan viereen	Parking next to the marked parking space
2300	Pysäköinti ilman P-kiekkoa/P-tunnusta	Parking without P-disc / P-permit
2301	Pysäköinti ilman pysäköintikiekkoa	Parking without parking disc
2302	Pysäköinti ilman pysäköintikiekkoa/p-tunnusta	Parking without parking disc / P-permit
2303	Pysäköintikiekko ei luettavissa	Parking disc not readable
2400	Pysäköinti/Pysäyttäminen ryhmitysalueelle/siten, että ryhmitys vaikeutuu	Parking / stopping in groupping lane/so that it hampers grouping
2401	Pysäköinti ryhmitysalueelle	Parking in grouping lane
2402	Pysäköinti siten että ryhmitys vaikeutuu	Parking so that it hampers grouping
2403	Pysäyttäminen ryhmitysalueelle	Stopping at grouping lane
2404	Pysäyttäminen siten että ryhmitys vaikeutuu	Stopping so that it hampers grouping

2500	Ei tosiasiallisesti liikennekäytössä	(Vehicle) no longer registered for traffic
2501	Ei saa säilyttää katualueella	(Vehicle) not allowed to be stored/kept on street area
2600	Tarpeeton joutokäynti	Unnecessary idling
2700	Pysäköinti lisäkilvessä mainituille ajoneuvoille varatulle paikalle	Parking on spot reserved for vehicles noted in the additional sign
2701	Pysäköinti linja-autoille varatuille paikoille	Parking on spot reserved for buses
2702	pysäköinti kuorma- ja linja-autoille varatuille paikoille	Parking on spot reserved for lorries/trucks and buses
2703	pysäköinti kuorma- ja pakettiautoille varatuille paikoille	Parking on spot reserved for lorries/trucks and vans
2704	pysäköinti henkilö- ja pakettiautoille varatuille paikoille	Parking on spot reserved for passenger cars and vans
2705	pysäköinti henkilöautoille varatuille paikoille	Parking on spot reserved for passenger cars
2706	pysäköinti moottoripyörille varatuille paikoille	Parking on spot reserved for motorcycles
2707	pysäköinti torikauppiaille varatuille paikoille	Parking on spot reserved for market square vendors
2708	pysäköinti hallikauppiaille varatuille paikoille	Parking on spot reserved for market hall vendors
2709	pysäköinti polkupyörille varatuille paikoille	Parking on spot reserved for bicycles
2710	pysäköinti kaupungin virka-autoille varatuille paikoille	Parking on spot reserved for city's (official) cars
2712	pysäköinti metron huoltoautoille varatuille paikoille	Parking on spot reserved for metro's maintenance cars
2713	pysäköinti kuorma-autoille varatulle paikalle	Parking on spot reserved for lorries/trucks
2801	pysäköinti kaksisuuntaisen tien keskellä olevan puistikon päätyyn	Parking at the end of a park in the middle of a two-way road,
2802	pysäköinti kaksisuuntaisen tien keskikorokkeen päätyyn	Parking at the end of a platform in the middle of a two-way road,
2803	pysäköinti kaksisuuntaisen tien keskikorokkeelle	Parking on a platform in the middle of a two-way road,

2805	pysäköinti kaksisuuntaisen tien keskikorokkeen viereen	Parking next to a platform in the middle of a two-way road,
2900	Muuta liikennettä haittaava pysäköinti	Parking that hampers other traffic
2901	Raitiotieliikennettä estävä pysäköinti	Parking that obstructs tram traffic
2902	Raitiotieliikennettä haittaava pysäköinti	Parking that hampers tram traffic
2903	Pysäköinti ilman pysäköintilupaa	Parking without parking permit
2904	Toisen ajoneuvon pois paikalta siirtämisen estävä pysäköinti	Parking that hampers moving another vehicle
2905	Pysäköinti etuajo-oikeutetun tien ajoradalle taajaman ulkopuolella	Parking on the right of way (priority) lane outside the urban area
2906	Liikennemerkin tai muun liikenteenohjauslaitteen havaittavuuden estävä pysäköinti	Parking that obstructs visibilty of traffic sign or other traffic direction device
2907	Pysäköinti moottoritielle	Parking on highway
2908	Pysäköinti sulkuviivan kohdalle	Parking between traffic lanes
2909	Pysäyttäminen sulkuviivan kohdalle	Stopping between traffic lanes
2912	Pysäköinti alikäytävässä	Parking at an underpass
2914	Pysäköinti mäenharjalla tai näkyvyydeltään rajoitetussa kaarteessa tai niiden läheisyydessä	Parking at a hill crest or close to a road curve (so that it obstructs visibility)
3000	Vapaamuotoinen	Free form (violations that cannot be explained by predefined text)
9999	Muut	No reason specified

Appendix C

Evaluation result

Participant	1	2	3	4	5
Time spent (including idle)	**22:33**	**14:47**	**16:14**	**17:00**	16:08
Introduction	1:48	2:02	2:27	1:09	0:12
Free interaction	11:51	6:43	6:35	9:07	11:10
Task #1	1:36	0:54	1:07	0:37	1:25
Task #2	2:26	0:22	0:17	1:12	0:26
Discussion	4:52	1:38	1:25	3:35	1:30
Number of features covered	**11**	**10**	**11**	**6**	**7**
Unused features	-	#4	-	#4, $9, #10, #11	#1, #3, #4, #11
Number of questions asked	**9**	**5**	**5**	**3**	**3**
Data content	3	1	0	1	0
Violation text meanin	3	-	-	-	-
Data coverage	-	1		1	-
UI	4	3	3	1	3
Language	1	1	1	-	-
Maps	-	-	2	-	1
Timeline	-	-	-	-	1
Reason filter	1	1	-	-	-
Explorer	-	-	-	1	1
Info box	2	1	-	-	-
Visual encoding	2	1	2	1	0
Color	2	-	2	1	-
Info box	-	1	-	-	-
Assistance duration	**5:03**	**1:01**	**1:04**	**0:45**	**0:30**
Time until first insight	**4:45**	**0:24**	**0:24**	**0:08**	**1:55**
Number of insights - UI	**3**	**9**	**8**	**3**	**3**
Number of insights - data	**16**	**13**	**6**	**8**	**14**
Classify	5	2	2	2	3
Compare	8	4	2	2	3
Reflect	3	7	2	4	8
Number of misinterpretation	**1**	**2**	**4**	**0**	**0**
Visual encoding	1	2	4	-	-
Number of usability issues	**3**	**2**	**7**	**6**	**7**
Against expectation	2	1	6	4	6
Delay	-	1	-	-	1
Crash	1	-	1	2	-
Number of suggestions	**13**	**5**	**7**	**4**	**5**
UI	6	1	6	1	2
Visual encoding	1	1	-		1
New feature	6	3	1	3	2

Figure C.1: Number of qualitative information encoded from five usability test sessions