Master Thesis Business Informatics

The Effective Visualization of Mobility Information

Author:
Zaki Abdurrahman Koesoemahardja
z.a.koesoemahardja@students.uu.nl
5584205

Supervisors:
Dr. Joske M. Houtkamp
Dr. Floris J. Bex
Wim Steenbakkers

Master Business Informatics
Dept. of Information and Computing Science
Utrecht University
2017

Abstract

The aim of the present research is to investigate the effective ways to visualize information, especially in mobility information, for specific use and user groups. It focuses on the effectiveness of information visualization in two principles: accuracy and efficiency. Currently, our world is flooded with huge numbers of data which we believe contain important information that can be lead to knowledge. A successful method to understand and present the stories behind the data is by using visualization (Few, 2013). Ensuring the effectiveness of a visualization is divided into two principles: accuracy and efficiency becomes increasingly important as the use of visualizations continues to rise (Zhu, 2007). Nowadays, one of the challenges for business users like Mezuro is communicating information, especially in mobility information, as effective as possible through visualization. Mobility information can be used to answer questions about human location and movement (e.g. Where are people going? Where so people spend their time?). While other information only show general patterns of human settlement and migration (Palmer et al., 2013), mobility information also can be used to show people's mobility patterns. In this research we combine qualitative and quantitative studies.

This study has identified that the most effective visualization methods to present mobility information depend on the type of tasks and questions (comparison tasks, relation seeking tasks or lookup tasks) that will be answered. A combination of a bar chart and a table could be considered to be used in comparison tasks. In relation seeking tasks, a map can be used to enhance the efficiency and the accuracy when the information contains geographical areas or locations. In lookup tasks, it could be considered to use either a combination of a bar chart and a table or a map and a table. Subsequently, there are also visualization design elements that need to be considered like color, highlight, label, legend, layout, and explanation of information. Furthermore, this research shows that to achieve an optimal design, the design process should be structured as follows: determine user requirements, look at the type of question or task from the customers and select a visualization method that can be used in a certain task, test the effectiveness of a visualization during the development phase, and evaluation is very useful for understanding how the product was used in the context.

Acknowledgements

This thesis on The Effective Visualization of Mobility Information is the final result of my study in Business Informatics at Utrecht University. This thesis would not be completed without the help of many people and organization throughout the process. I want to take this opportunity to thank them for their assistance and support.

First of all, I would like to thank God, who has gave me health to accomplish this master thesis and finish my education. Secondly, I would like to thank my supervisors from Utrecht University, Joske Houtkamp and Floris Bex, for giving me support, feedback, and guidance throughout the project. Then, I would like to thank Mezuro for allowing me to perform this research in their company. My supervisor, Wim Steenbakkers, for providing me with discussion time, contacts, and supporting me to complete this master thesis. Subsequently, I would like to thank my colleagues at Mezuro for their support and friendliness.

I would also like to thank the partners of Mezuro that have provided me their time and constructive ideas. I would also like to thank the participants in my qualitative and quantitative measurements who have spared their time to participate in the experiment and the survey, which support my research. Furthermore, I would like to thank LPDP Scholarship of Indonesian Government for providing financial support to study at Utrecht University. And finally, I would like to thank my loved ones for their support throughout this process.

Table of Contents

List of Figures

List of Tables

1. Introduction

Currently, our world is flooded with huge amounts of data which we believe contain important information that can lead to knowledge. Nevertheless, retrieving this information is difficult when the data is contained in texts or traditional tabular forms (Chan, 2006). The information that is produced from many sources has become an important part of human life (North, 2005). To absorb a huge amount of information, humans have used visualizations as a successful method to discover, understand, and present the stories behind the data (Few, 2013).

Visualizations have been used as a tool for communication before the creation of written language. They can contain a large amount of information and are generally faster to understand than text (Ward, Grinstein & Keim, 2010). Visualizations use graphical representations of data to provide insight which is used to answer specific questions (Telea, 2014). Some research studies show the advantage of visualizations as seen in research by Burkhard (2004), which showed that visual images help people to interpret information easier than texts. Other research by Burkhard et al. (2005) shows that humans can absorb more information when they use their visual abilities.

Human perception affects the understanding of information and is therefore important in the design and development of information visualization (Alexandre & Tavares, 2010). According to Khan and Khan (2011), information visualization is "*the depiction of information using spatial or graphical representations to facilitate comparison, pattern recognition, change detection, and other cognitive skills by making use of the visual system*". The goal of information visualization is to give insight in complex information so it can easily be interpreted by the users (Spence, 2001).

According to Rhyne (2003), there are two main categories of visualization: scientific visualization and information visualization. Scientific visualization shows graphical models that are built from scientific data that represent objects—which means that it focuses on physical data, such as the earth, the human body, molecules, etc. Information visualization shows graphical models that express abstract concepts and relationships which means that it focuses on nonphysical data such as text, statistical data, hierarchies,

etc. Schematic or diagrammatic forms (e.g. line chart, bar chart, map, etc.) can be used to visualize information.

Therein lies an issue, and research problem, as there is no single streamlined, theoretically based method of determining visualization effectiveness. Further research is needed in this area. This research project was designed around a case study provided by Mezuro. Mezuro is a company that specializes in processing and analyzing data from mobile networks and then turning the data into useful information about the mobility of (groups of) people. They deliver the information in the form of reports that use visualizations. The company requires understanding how they should visualize their mobility information for specific questions and specific user groups. The following are examples of cases in which Mezuro uses information visualization:

- Traffic and transportation – By bicycle to Groningen: Every day about 70.000 people, living within 15 kilometers of the city of Groningen, visit the city on a frequent (commuters and students) or regular (shoppers) basis. In order to increase the number of cyclists among these visitors, the Groningen municipality required information on the number of visitors and their travelling distances. The municipality then asked Mezuro to investigate connections that have the best potential for a fast cycle path. The villages Bedum and Haren have the best potential: the number of daily frequent traveler is high and the distance is relatively short. From the visualization, we can see not only the numbers and the distances but also make a comparison. So it is easier to determine the locations and make a decision.

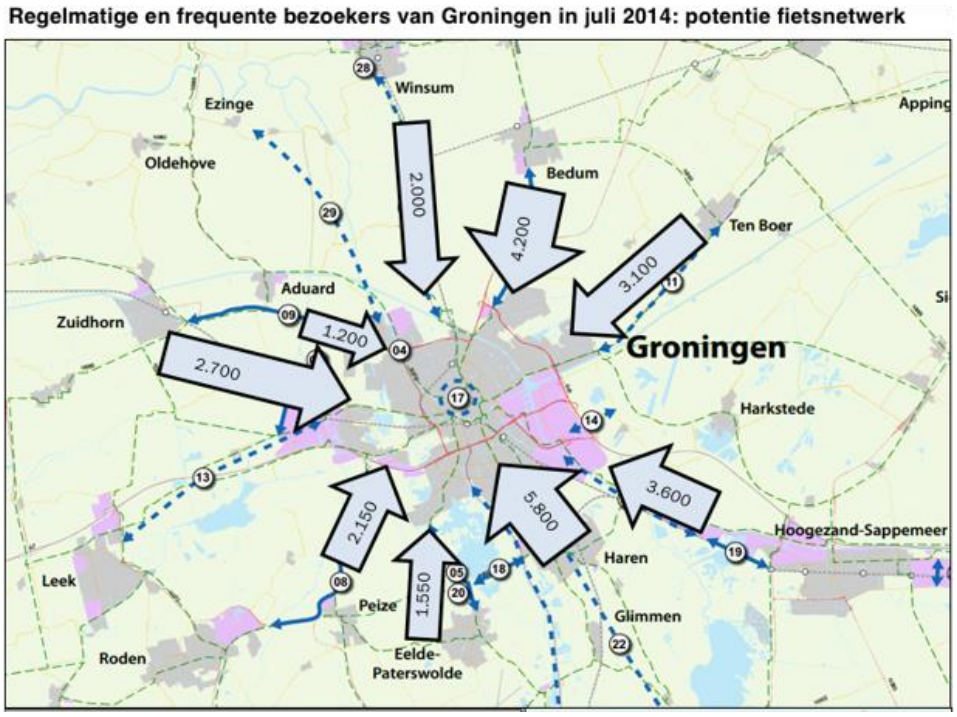


Figure 1.1 Traffic and transportation information visualization
(By bicycle to Groningen)

- Event – Queen's day 2013: A municipality wanted to know when is the peak hour or the busiest time Queen's day and they asked Mezuro to investigate it. From the visualization, we can quickly see that the peak hour during queen's night was between 24.00 and 01.00 and that during Queen's day the peak hour was in the afternoon.

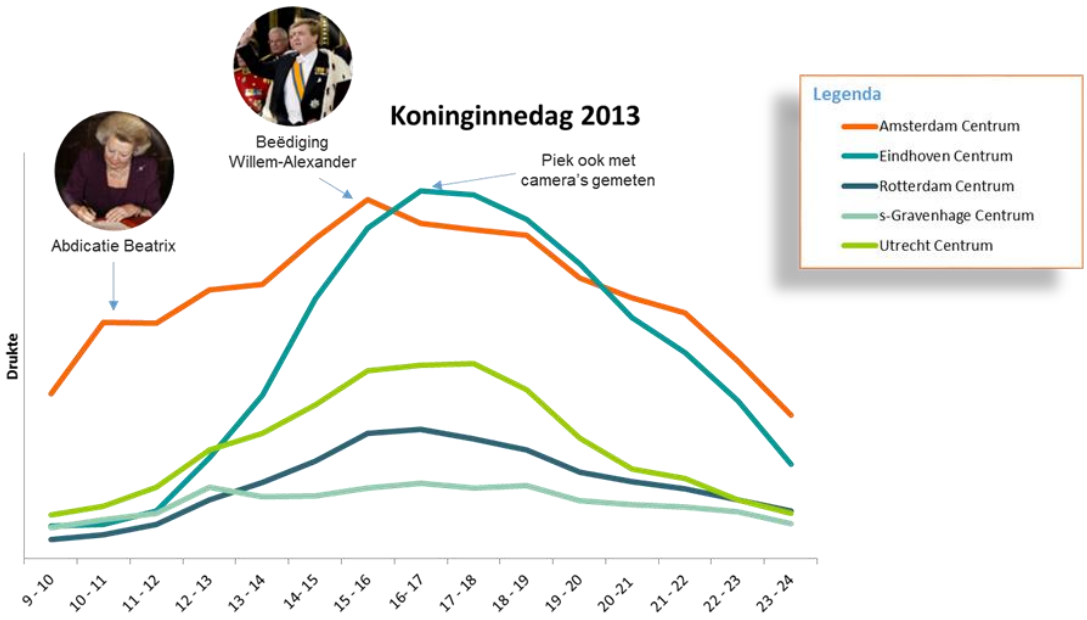


Figure 1.2 Event information visualization (Queen's day)

Currently, Mezuro produces several reports (e.g. events report, tourism report, and retail report) from the mobility information for their clients using different kind of visualizations. One of the issues they face is communicating the information as effective as possible to their customers. Therefore this research is needed to make guidelines and procedures in visualizing the mobility information not only for Mezuro, but also for other organizations that use mobility information.

1.1 Problem Statement

How an audience interprets information depends on the quality of its presentation. Maps, graphs, charts, etc.; have been used for several decades to communicate small and large amounts of information. Ensuring the effectiveness of a visualization is divided into two principles: accuracy and efficiency becomes increasingly important as the use of visualizations continues to rise. Few (2006) states that most visualizations fail to communicate effectively or efficiently, because of poor design of the visualization. Khan and Khan (2011) believe that it is still a challenge to make a perfect visualization which meets all the requirements of users.

A visualization is said to be effective if the desired information is delivered to the targeted audience (Ward, Grinstein & Keim, 2010). In making an effective visualization, visualization skills, along with a deep understanding of the data/information needs of the users are required. Many visualization research projects show interesting techniques and methods to visualize information but remain unaware of its level of effectiveness (Stone, 2009).

Nowadays, many organizations experience a phenomenon which is known as information overload. At the same time, their ability to analyze the information is not as fast as their ability to collect the data and information (Keim et al., 2006). In the competitive world, organizations can be successful if they can act faster than the competition. This may be achieved by transforming information into insights using visualization (Marr, 2016). Based on a survey in the United States and Europe among business executives, IT professionals, and consultants in 2013; 57% of respondents were implementing information visualization in their organizations (Stodder, 2013). Until now, one of the challenges for business users like Mezuro is communicating information as effective as possible through visualization. The type of information visualization will have an effect on how the information is interpreted (Data visualization-what it is and why it matters, 2016).

Mobility is a complex system because the trajectories human beings followed while awake needs to be examined both collectively, and on an individual basis. Understanding collective and individual movement behaviors are no easy feats because there are limited information sources (see section 3.4 for the details). Mobility information can be used to answer questions about human location and movement (e.g. Where are people going? Where do people spend their time? How do they travel?). Therefore, the objective of this research is to determine the most effective ways to visualize mobility information so that the data can be easily communicated and correctly interpreted.

1.2 Research Questions

This research focuses on the effectiveness of information visualizations in two principles: accuracy and efficiency (Zhu, 2007). It is conducted with the aim to investigate which

information visualization methods, especially in mobility information, can be designed to enhance their effectiveness for a specific use and user groups. Therefore, the main research question of this research is formulated as follows:

RQ: "Which visualization methods are most effective to present mobility information, in relation to specific user groups and their information needs? And how can the design process be structured to achieve an optimal design?"

To answer the main research question, the following sub-questions are derived:
SQ1: What is an effective information visualization?
SQ2: What are the existing visualization methods?
SQ3: What is mobility information?
SQ4: Which design patterns can be identified for specifying the most effective ways to present mobility information?
SQ5: Which techniques are required to determine user requirements, select the visualization methods, and evaluate the effectiveness of the visualization?

Information visualization has been, and still is, a popular topic in scientific research. However, research on this topic can have different levels of scope. It is therefore of great importance to define a clear scope of this thesis project. The first limitation regarding the scope of this research is the type of information. In this research we use mobility information that is produced by Mezuro from their data source which is mobile phone data. The second limitation is the user group. Mobility information can be used by many different groups (e.g. profit organization, non-profit organization, government, etc.). However, in this research we focus on the customers of Mezuro who work at municipalities in the Netherlands and use the mobility information.

1.3 Scientific and Practical Contributions

Scientific Contributions. Scientifically, this research is relevant as it focuses on information visualization, especially of mobility information. In recent years, information visualization has become an established area in academia (Few, 2007). Many research studies in information visualization not only come from computer science but also from

other disciplines, such as psychology. Although there has been much research in information visualization, the effective visualization of mobility information is not yet discussed extensively, in relation to specific user groups and their information needs. This research aims to determine the most effective ways to visualize mobility information. The result would broaden the knowledge domain of information visualization.

Practical Contributions. A Case study (Serious Request-Glazen Huis event) was used as a mechanism to better understand effective visualization of mobility information on a practical level. A challenge faced by Mezuro is their lack of knowledge on effective visualization techniques, which may lead to misinterpretation of information by the customer. This study contributes to an increase of understanding of useful techniques and methods in the company's domain.

1.4 Thesis Structure

In this section, the problem statement has been introduced along with the research questions, scientific and practical contributions. The remaining part of this thesis will be structured as follows. The following section describes the research approach we used to answer the research questions. In section 3, we provide literature regarding visualization, information visualization and mobility information. Subsequently, the data collection will be discussed. Section 5 presents the results and findings of this research. Finally, section 6 concludes this research by giving an overview of the things discussed, the limitations of the research, and provides improvement for future research.

2. Research Approach

In order to answer the research question, our approach in this research is a combination of qualitative and quantitative approaches. A qualitative approach is done by conducting a literature study, semi-structured interviews, and a case study. For the quantitative approach, we conduct an experiment to measure the accuracy and efficiency of a visualization using a survey. Hancock et al. (1998) state that qualitative research has the aim to gain an in-depth understanding of the social world and why things are the way they are. Qualitative research is in line with our aim because qualitative research is used to address "how" or "why" questions (Ritchie et al., 2013), and the aim of this research is to investigate which visualizations of information can be designed to enhance its effectiveness for a specific use and user group. For the quantitative research, Babbie (2010) states that it is used not only to gather numerical data but also to explain a particular phenomenon. Thus, the quantitative part of this research can provide strong evidence on the effectiveness of the information visualization designs.

In the second place, the research will lead to a proposal on the information visualization design and development process, that can be used in future projects. The process will be developed and designed in the case study. Figure 2.1 illustrates the process deliverable diagram (PDD) of this research. The first activity of this research is to define the problem. Subsequently, we conduct a literature review and select relevant literatures for this research. After that, we perform a case study by doing interviews, testing, and evaluation of visualizations. We do a case study to collect data about visualization requirements. After we have defined the case study conclusions, we finalize the thesis and present it in the final presentation.

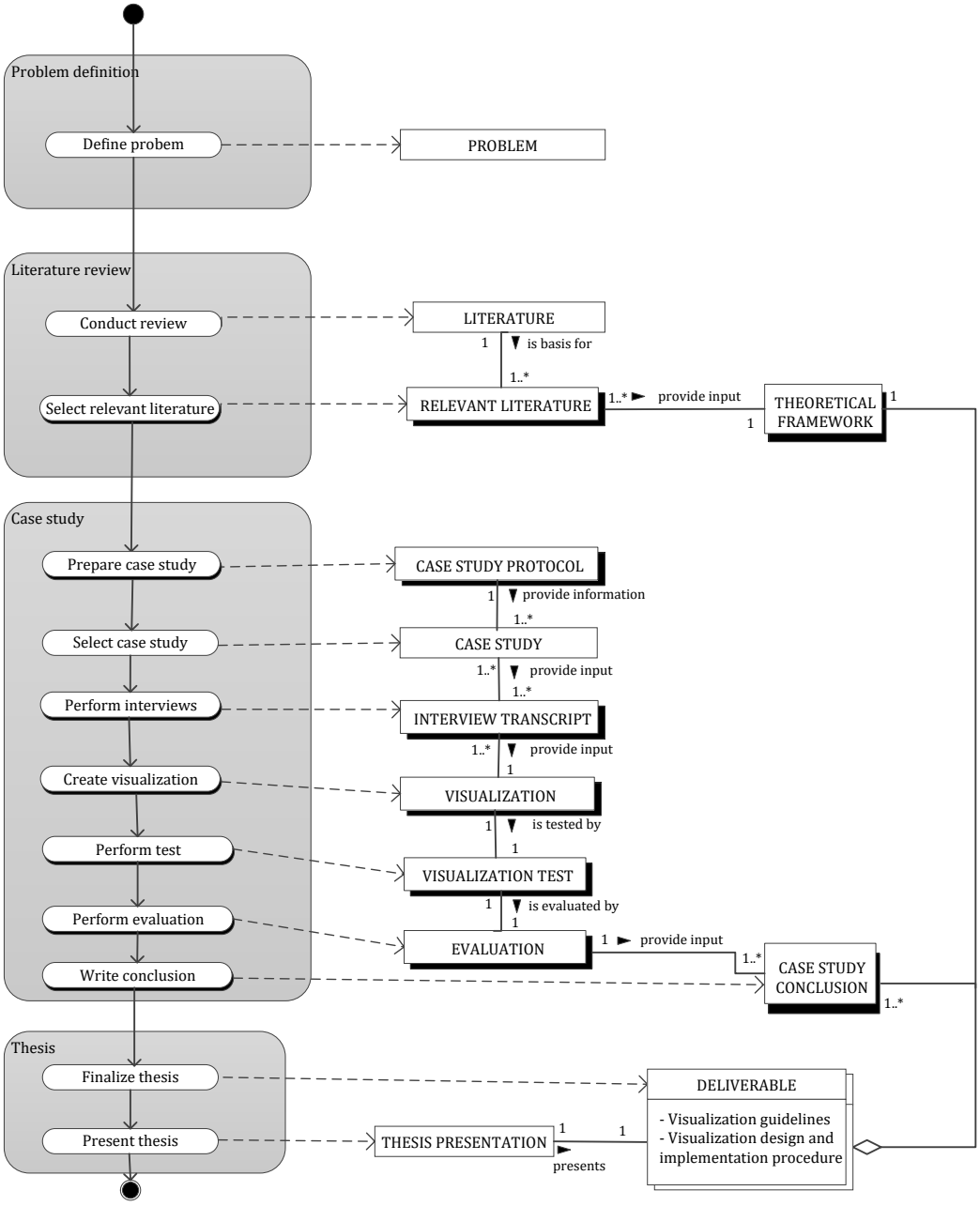


Figure 2.1 Process deliverable diagram of the research

2.1 Literature Study

The foundation for the thesis research is formed by the literature. Literature study is important because it helps the researcher to understand the topic and discover what has and has not been studied in that area (Denney & Tewksbury, 2012). The literature study

is conducted in the area of information visualization and mobility information. According to Webster and Watson (2002), the process of literature study is started by searching related knowledge within leading journals. In this research, we use the most common scientific journal databases and search engines (i.e. Google Scholar).

For each search result, we analyze the title and abstract. If the abstract indicates that the literature is appropriate for this research, we will continue to analyze the paper. Then, for each relevant paper, the next process is searching more literature using the "Snowballing" technique (Skoglund & Runeson, 2009). By using the snowballing technique, we can track references from one paper to another.

2.2 Semi-Structured Interviews

In this research, we perform a semi-structured interview method to gain data and information. Interviews are used to collect data on individual experiences and perspectives, especially when investigating a particular issue (Mack et al., 2005). Longhurst (2010) states that a semi- structured interview is "*verbal interchange where the interviewer attempts to elicit information from another person (interviewee) by asking questions*". It is organized around a set of major questions which is then followed by a series of sub-questions. Even though interviews take time, it can produce high quality results. We choose a semi-structured interview rather than a structured interview because a semi-structured interview gives the flexibility to approach different respondents and it offers a chance to explore a particular issue by discussing other questions emerging in the interview (Noor, 2008).

2.3 Case Study

Darke et al. (1998) state that in order to understand the interactions between information technology and organization contexts, case studies are the most used qualitative research method. A case study is helpful when the researcher wants to investigate a phenomenon in its real life context, applying different kinds of data collection methods to gather information (Yin, 2013). Therefore, a case study approach is chosen in this research to solve the research question. Figure 2.2 shows the whole process of case studies research by Yin (2013).

According to Yin (2013), there are three types of case study research: descriptive, exploratory and explanatory. The purpose of the case studies is to get insights on how mobility information can be designed to enhance its effectiveness for a specific use and user groups. Hence, in this research we choose the exploratory case study because it is used to explore the object of the study and is usually the beginning of new research. In this research we conduct a single case study which we focus on an event in the Netherlands (Serious Request-Glazen Huis).

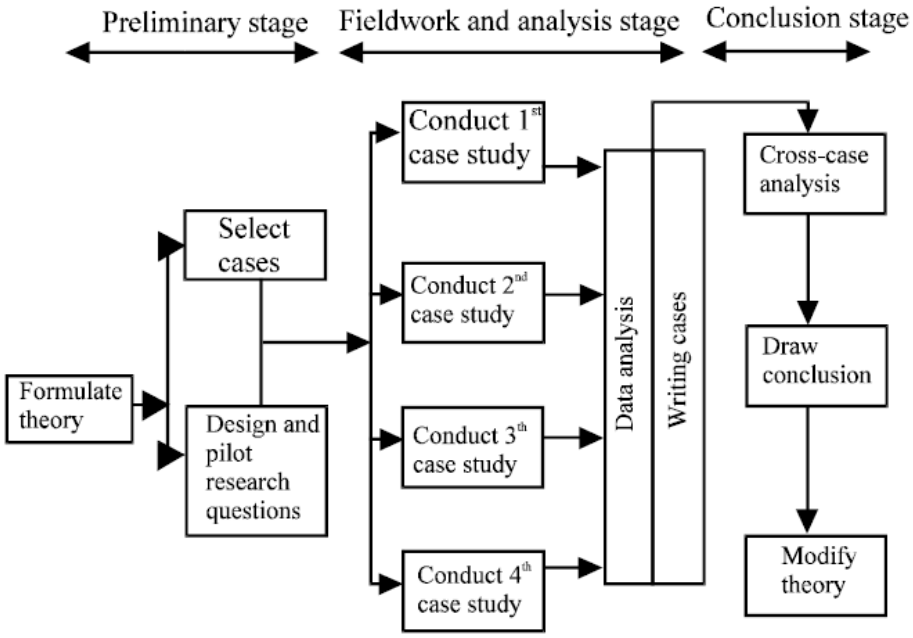


Figure 2.2 Case studies research process (Yin, 2013)

2.4 Experiment

In this research, we conduct an experiment to test the information visualization designs using qualitative and quantitative measurements. A qualitative measurement enables us to gain a deeper understanding of the cognitive process of information visualization viewers. In the experiment, we let the participants to think aloud while answering several questions using the visualizations (Krug, 2011; Someren et al., 1994). Next, in the quantitative measurement, we measure the accuracy (number of interpretation errors)

and the efficiency (time) of each participant in answering a single question (Zhu, 2007). More details on the experiment of this research are discussed in section 4.5. With a combination of qualitative and quantitative measurements, we empirically test the effectiveness of the visualizations.

3. Theoretical Background

In this chapter, we will discuss what information visualization is, the process of information visualization, what is considered effective visualization, mobility information, and visualization techniques that may be used for mobility information.

3.1 Information Visualization

Information visualization is an area of study which continues to garner more attention in the academic sector as well as the industrial environment. It is a region of research that spotlights the utilization of visualization procedures to help individuals comprehend and dissect data. Card et al. (1999) define information visualization as "*…the use of computer supported, interactive, visual representations of abstract data to amplify cognition.*" It does not involve data with a physical or geometric correspondence, rather it focuses on visual representations of abstract data in both numerical and non-numerical form. Put simply, it is designed to help us make sense out of facts. It can likewise help us to confirm our comprehension of previously analyzed data. Most commonly, information visualization is constructed graphically to support visual understanding.

The first step in understanding information visualization is to step back and focus on the fundamentals. According to The Dictionary of Computer Graphics and Virtual Reality, visualization is the "*process of representing data as a visual image*" (Latham 1995, p. 148). Visualization is used to answer a question, support a decision, communicate information, and increase efficiency. Visualizations are tools for seeing—for finding insights and then communicating those insights concisely and accurately. Often depicted graphically, they contain a huge amount of information which can be clearly understood and interpreted easily (Tufte, 1997). Information visualization is one of three categories of visualization and it depicts information using spatial or graphical representations to compare patterns, amongst other things. The purpose of visualizations is to discover, understand, and present the stories behind the data (Few, 2013). Tegarden (1999) asserts that visualization allows:

- Manipulation of the visual system to extract information from data.
- Achieve an overview of complex data sets.
- Identify patterns, structure, and relationships in data.

To further illustrate the possibilities of visualizations, Tufte (1997) shows a famous example that visualization has been used to make important decisions in London during the cholera epidemic in 1854. Figure 3.1 shows deaths from cholera and the location of water pumps. This explains that visual representations were used to solve problems and had a positive impact at that time. However, Tufte shows not only the positive impact of visual representations, but also the negative impact of a poor-quality visualization. He uses the launch of the space shuttle as an example to show the negative impact of a poor-quality visualization. When a poor-quality visualization is used in a technical decision making context, the interpretation of the visualization may be wrong and may result in the loss of lives.

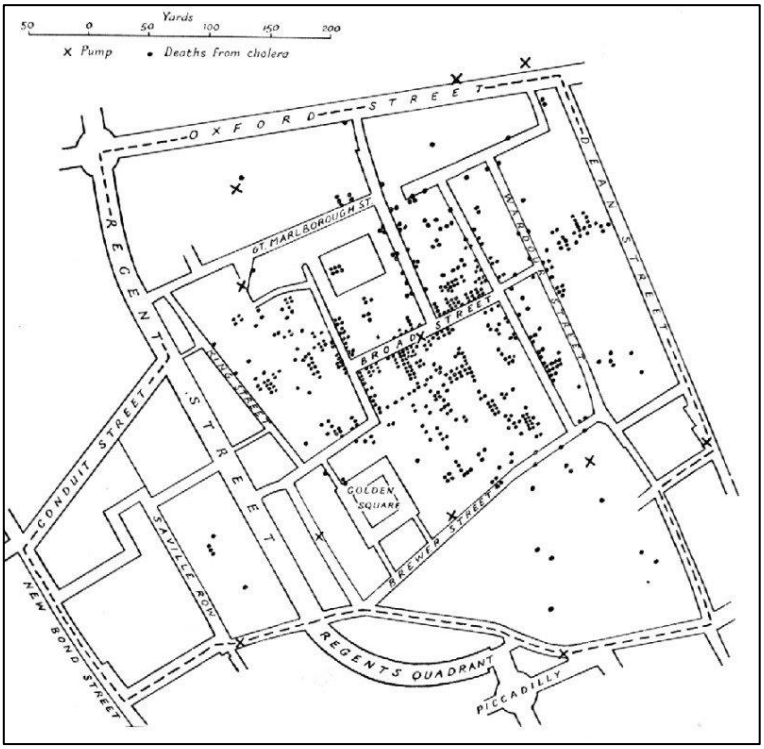


Figure 3.1 Map of London's Soho district in 1854

The field of information visualization has theoretical roots in multiple sub disciplines, such as psychology, sociolinguistics, graphic design, business processes, and computer design. It is applied in data mining, crime mapping, market research, financial studies, bench research, and more. For the purposes of this research, we have focused on determining the most effective ways to visualize mobility information so that the information can be easily communicated and accurately interpreted. A good visualization can transform multiple hours of hard work into an illuminating insight or can pass on an intricate story in a solitary minute.

Over the years, the understanding of information visualization changed from a construct of visual images in the people minds (internal) to external artifacts, such as a visualized graphs used for decision making. According to Khan and Khan (2011) information visualization is "*the depiction of information using spatial or graphical representations to facilitate comparison, pattern recognition, change detection, and other cognitive skills by making use of the visual system*". The aim of information visualization is to help the user understand and gain insight into complex information with little effort. Information visualization does not only involve collecting and processing information to be shown, but also considering the type of visualizations to present the set of information. Bar charts, network graphs, and scatter plots are information visualizations because they lay out in space data which do not have inherent spatiality.

3.2 Information Visualization Process

There has been a recent spike in interest regarding information visualization and its capability to communicate. Some of the recent interest stems from improved tools for the creation of the visualization. While historically, information visualization was seen as a vital device for analysts, the perception has shifted and broadened, and visualization is now seen as a fundamental part of research communication.

Card et al. (1999) show a high-level view on the information visualization process. This visualization process consists of three main activities: data transformation, visual transformation, and user interaction. Figure 3.2 shows the steps of this process. The first step is to process and transform the raw information into a data set. The next step is

mapping the data set into a visual structure. Then, the visual structure is transformed into views which show the visual representation of the information. The final step is the visual perception by the user.

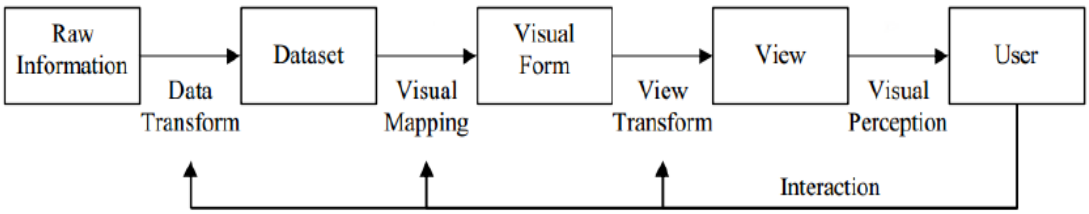


Figure 3.2 Information visualization process (Card et al., 1999)

Referring to Figure 3.2, Wunsche (2004) splits the process into two main stages in information visualization: encoding (convert information into visualization) and decoding (interpretation of the visualization). Figure 3.3 demonstrates the information visualization process. The encoding stage consists of data transformation, visualization mapping and displaying the visualization. The decoding stage, which involves perception and cognition, begins right after the encoding stage is complete.

Figure 3.3 Information visualization process (Wunsche, 2004)

3.3 Effective Visualization

A principle question in visualization is what constitutes an"effective" or "good" visualization and how the effectiveness can be measured? One of the most important issues with visualizations is making sure the visualization techniques are used in the most effective manner. A review of the literature however shows many different and inconsistent definitions of effectiveness. Also there is a need for improved standards in

measuring the effectiveness. Additionally, there is no clearly defined underlying theory to create a framework for excellent evaluation of effectiveness. The lack of standards for measurement and the need for a solid theoretical framework are noted as research issues.

So, how does one determine the "right" visualization that will be most effective in displaying information? Choosing a visualization depends upon what variables are used, whether they are qualitative or quantitative, and the way in which the designer would like to present the information. Then decide what colors to use, symbols, if there is a legend, etc.

Three analytic methodologies are used to evaluate the information visualization in this study: Heuristic evaluation, cognitive walkthroughs, and user testing. A heuristic evaluation involves an individual or group of individuals to walk through and evaluate the design of an interface and determine its usability (Zuk et al., 2006). There are no set heuristics for information visualization assessment and researchers must define their own. This method is cheap and easy, albeit slightly challenging, without a clearly defined set of heuristics to follow. A cognitive walkthrough is where a researcher acts on behalf of a user as an evaluator by working through preplanned scenarios. This is all in an attempt to mimic the behavior of the user to identify problems they may face. The last methodology is user testing where it is probably known as the most commonly used method to find usability problems and to evaluate user interfaces (Nielsen, 1994).

A more commonly used approach is through user testing, which evaluates how effective and efficient a visualization was user during a controlled experiment with participants who are representative of the users. Perceptual speed and visual working memory can be evaluated through the completion time of tasks compared to a baseline and refer to effectiveness and efficiency of the visualizations (Conati & Giuseppe, 2011). This type of approach can be done using quantitative or qualitative approach. Wunsche (2004), believes a visualization is considered effective when the decoding stage is performed correctly and the perceived data quantities and the relationship between data reflect the actual data. Similarly, Wattenberg and Fischer (2004) outline that a visualization is

effective when it has a structure that fits with the structure of the data. Other researchers believe that a visualization becomes effective because of a specific task, like Casner (1991), who explains that to have a more effective visualization, the visualization should be made for a specific task. Moore and Purchase (2011) propose utility and attractiveness as principles of an effective visualization. In their paper, they propose to add attractiveness, which may encourage the user to interact with the data and allow more effective communication of the information.

From a psychological standpoint, researchers explain that the effectiveness of visualization depends on user's ability, experience, and knowledge in reading a visualization (Zhu, 2007). The article goes on to define the effectiveness of visualization in two principles: accuracy and efficiency. Accuracy means that the structure of the visualization should match the structure of the data. It explains the relationship between the data and the visualization. Following the efficiency principle, a visualization should be easy to understand and improve task efficiency. It explains the relationship between the user and the visualization. Zhu (2007) also mentions methods to measure the effectiveness of a visualization. Table 3.1 shows the quantitative and qualitative measures for accuracy and efficiency.

Table 3.1 Quantitative and qualitative measures for visualization effectiveness (Zhu,2007)

Principles	**Quantitative measurements**	**Qualitative measurements**
Accuracy	Measures the number of interpretation errors	- Interview - Observation
Efficiency	Record task completion time	- Interview - Observation

3.4 Mobility information

The universal utilization of mobile devices brings opportunities to collect vast amounts of persistent data about human conduct. Moreover, while ultimately researchers intend to solve the issue of how, when, where, and why living things move, finding an answer requires a complex, multistep process. During the last several decades, technology advances such as GIS, global positioning, and remote sensing have transformed the way in which human beings are studied in time and space. Technology advances are helping research people's movements to answer the following questions about human location and movement:

a. Where are people going?
b. Where do people spend their time?
c. Where have they come from?
d. How do they travel?

Mobility is defined as "the movement of people in a population, as from place to place" (Mobility, n.d.). In this research, we use mobile phone data as our source to visualize mobility information. Almost everyone owns a personal mobile phone and they are used in all facets of our lives (e.g. at home, at work, and while traveling). Therefore, it is now possible to gather information about human mobility patterns and behaviors by mining phone data (Pu et al., 2011). People's movements can be traced by analyzing phones transition from one cell tower to another (Palmer et al., 2013). Wireless providers continuously collect mobile phone data for billing and for improving the functionality of the network (Becker et al., 2013). This type of data is excellent for studying human behavior with precision and on a vast scale not possible using other data collection techniques. "This data helps to overcome the limitations of censuses and surveys and is appropriate for capturing information about people's mobility in real time." (Palmer et al., 2013; Wu et al., 2016).

Mezuro processes mobile phone data into mobility information. The company uses the mobility information to determine human trajectories and provide continuous information that can be of great value for a variety of applications such as traffic and transportation, tourism, events, retail and urban planning. The information is usually

used by municipalities, retailers, tourism (amusement parks, museums, etc.), and event organizers. From the following explanations we can see why mobility information is useful:

- Traffic and transportation: mobility information can be used to answer these questions:
    - Where do the inhabitants of a city go and what means of transport do they use?
    - What are the effects of a new road?
- Retail: mobility information provides the catchment area of a store or shopping center, the frequency of visits, the number of visitors, and trend analyses of visitor numbers.
- Tourism: mobility information provides a clear analysis of tourism in a city or region. Insights in where tourists come from, how they move around through the Netherlands, and also the effect measurement of media campaigns (advertisements).

3.5 Visualization Techniques

In recent years, technology has advanced tremendously and opened a new world of opportunity for Big Data—and anyone else seeking visualization for exorbitant amounts of information. Lengler and Eppler (2007) define a visualization techniques as "a systematic, rule-based, external, permanent, and graphic representation that depicts information in a way that is conducive to acquiring insights, developing an elaborate understanding, or communicating experiences". In their paper, they categorize six types of visualization techniques, of which two were pertinent for this research:

- *Data Visualization*, this shows quantitative data in schematic or diagrammatic forms such as bar charts, line charts, scatter plots, etc. It is used for getting an overview of data.
- *Information Visualization*, this is a visual representation of data to increase cognition. It transforms data into an image, e.g. tree map, semantic network, timeline, etc.

The Visual Information-Seeking Mantra summarizes multiple visual design rules and gives structure to planning data perception applications (information visualization). It is characterized by: overview first, zoom and filter, then details-on-demand (Shneiderman, 1996). An information visualization starts with an overview, it is the first thing a viewer sees and it is used to help him or her get the whole picture of the entire visualization. The overview should not overload the user with too much information. Then, the viewer will focus on particular areas of interest which will be done by zooming in on the information. After that, the viewer will filter out uninteresting information and finally make a selection and get more details as needed.

From section 3.4, we see that mobility information consist of spatiotemporal data, and is best understood through visualization (Pu et al., 2011). Andrienko et al. (2003) distinguish three components of spatial and time data: objects (what), time (when), and space (where). Therefore, three basic questions can be made from those components:

- Where + when -> what: Define an object or group of object that is present at one location or set of locations at a given times or set of times.
- When + what -> where: Define a location or set of locations filled by an object or group of objects at a given times or set of times.
- Where + what -> when: Define a time or set of times that an object or set of objects present in one location or set of locations.

Furthermore, Andrienko et al. (2003) created a task model that is defined by a set of questions that viewers might look for to answer with the help of visualizations. Figure 3.4 shows the taxonomy of visualization tasks. In the first level there are two classes of tasks: elementary and synoptic tasks. Elementary tasks contain individual data elements or individual values and synoptic tasks involve a general view (sets of values or groups of data). Both tasks further divided into lookup, relation seeking, and comparison.

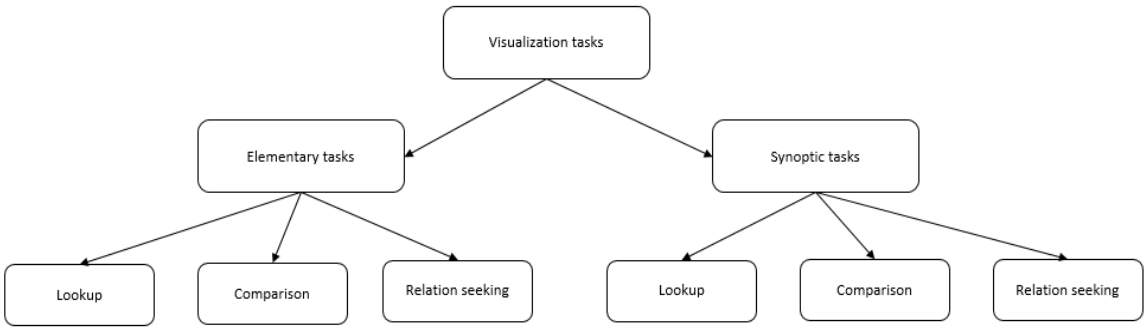


Figure 3.4 Taxonomy of visualization tasks

Lookup tasks address the search for data values and search for points in space and time. Relation seeking tasks look for the relations specified between data characteristics. Comparison tasks are almost the same as relation seeking tasks, but the difference is that relations to be decided are not specified beforehand. Below are lists of illustrative examples of the different type of tasks:

Elementary tasks:

- Lookup
  What was the price of Microsoft stocks on December 25?
  On which day was the highest stock price for Google in December 2015?
- Comparison
  Compare the stock prices of Samsung and Apple on December 25.
  Did the price of Samsung stocks reach $500 before or after December 25?
- Relation seeking:
  On which days was the price of Samsung stocks higher than the price of Apple stocks?

Synoptic tasks:

- Lookup
  What was the trend of Microsoft stocks during April?
  Find months in which the price of Apple stocks decreased.
- Comparison
  Compare the behavior of the stock price of Microsoft in January and December.

How is an increasing trend of Apple stocks related to the Christmas?

- Relation seeking

  Find two contiguous months with opposite trends in the stock price of Samsung.

The visualization techniques chosen for this research project support the above tasks, display geospatial data, and can be shown in static mapping (2D), which means that the visualizations are on one page so viewers can find the answer without having to browse through other pages. Because of positioning technologies, we can use the visualization techniques that are categorized as listed below (Andrienko et al., 2003):

a. "Universal" techniques, these techniques are applicable to all data types.
   - Focusing, linking and arranging views

     Focusing techniques contain the selection of variables of information in the visualization. This results in showing only partial information, which is compensated by showing different aspects of data in multiple views. Then, linking techniques are used to connect the multiple views so that the information can be integrated as a whole. The most popular method for linking views is identical marking, e.g. using the same color or size. Finally, arranging views is used to simplify comparisons. Cartographic maps and space-time cubes are universal ways to view spatiotemporal objects (including moving objects, spatial events, and time series of attribute values).
   - Map iteration

     A technique used for the representation of spatial-temporal data is using maps and map iteration or "small multiples". An example is the juxtaposition of several maps where each map shows the state of an event at a different time.

b. Techniques suitable for data about existential changes.
   - Time labels

     Time labels show when particular locations are visited. The moment when some events take place can also be identified by time labels. Moments may be specified imprecisely, such as in intervals rather than points in time.

c. Techniques for studying numeric changes.

- Change map

  This technique is used to show for each location or area the amount of change between two time moments. Different colors represent the changes. The degree of darkness shows the size of the change.

Table 3.2 presents the overview of general visualization techniques or visualization design guidelines that applicable for mobility information and the examples of its application in our visualizations.

Table 3.2 List of visualization design guidelines

Visualization design guidelines	**Explanations**	**Examples (see section 4.3 for the details on the visualizations)**
1. Type of chart		
a. Bar chart	Using a bar chart, in general, is a straightforward way to visually communicate information, specifically discrete data. For this project, bar charts were used to support a quick comprehension of values and show comparisons of values between different categories (Few, 2004; Data visualization 101: How to design graphs and charts, 2014).	In our visualizations, the bar chart is used to show comparisons between the number of visitors, the preference of transportation type that is used by the visitors, and the number of visitors in the city center for every hour.
b. Maps	Maps are important for this research project because mobility information contains spatial data. A map can be used to show the geographical location or area of the information (Mazza, 2009).	Maps are used to show the percentage of inhabitants that visit the area of the event per city, the visitors preference of transportation per city,

		and the visitors average staying time in the city center per province.
2. Color	Because colors play a focal role in data visualization, being able to choose the colors that will provide adequate differentiation during a task is essential. Distinguishing one element from another is one of the most important uses for color and a function, Tufte (1997) referred to as "to label". Few (2008) suggest several rules for using color in visualizations: • Because some people have color vision deficiency, try to avoid using a combination of red and green in the same visualization. • Only use different colors when they correspond to differences of meaning in the data. • Use soft, natural colors to display most information and bright or dark colors to highlight information that requires greater attention. • When using color to encode a sequential range of quantitative values, use a single hue (or a small set of closely related hues) and differentiate the intensity	• We choose the colors from the "color blind colors" group in Tableau software. • We use dark gray or dark blue to highlight the information that is required greater attention. • We use a single hue (blue or green) on the maps to encode a sequential range of quantitative values.

	from pale colors for low values to increasingly darker and brighter colors for high values.	
3. Highlight	Carenni et al. (2014) present a study to evaluate a variety of visual prompts (bolding, de-emphasizing, reference lines, and connected arrows) which are called 'interventions' that can be used in a visualization to help viewers process it. Their results show that all the highlighting interventions, except reference lines, can improve visualization processing compared to no interventions. Although no single highlighting intervention is most effective in general, de-emphasis (fades all non-relevant bars or areas on the map) always has the best results regarding performance and rated usefulness.	We highlight the information on the total number of extra visitors that visit the city center.
4. Label each bar or area on the map	In this way, we can see exactly which value each bar or area is representing while also getting a general visual comparison of it with all other bars or areas. Labels should be used sparingly to identify important aspects of the matter, otherwise it will make the visualization hard to read (Your Graphs Look Like Crap: 9 Ways to Simplify and Sexify Data, 2013; Ten Things to Consider When Making a Map, 2011).	We label the bar with the values of the data and we show the name of the cities on the maps.

5. Legend	Maps often use symbols or colors to represent information and legends are used to explain what they mean. For the legend of the map, we use 3-5 numerical ranges that enable even distribution of data between them (Data visualization 101: How to design graphs and charts, 2014; 7 Basic Rules for Making Charts and Graphs, 2010).	An example of the legend of the maps in our visualizations is the percentage of inhabitants: >10.1 % 5.8-10% 1.7-5.7% 0.6-1.6% 0.1-0.5%
6. Structure/ Layout	Users look first for information on the top and the left sides of a visualization and this is the natural path of the eye. The left side of the visualization is the section that gains the maximum attention from the viewers and where the most important information should be placed. The other information should be the distributed based on their order of importance (A Guide to Creating Dashboards People Love to Use, 2009; Four Cognitive Design Guidelines for Effective Information Dashboards, 2014).	We order the visualizations from left to right. Then, we start to put the most important information on the left part of the visualizations. For example, we start by showing the total number of visitors to the city center.
7. Explanation of information	Context and explanation are needed to understand new and unfamiliar events. Letting the data speak for itself can generate misinterpretation and confusion (A Guide to Creating Dashboards People Love to Use, 2009).	We add information about the date of the event, the date of the regular day, the area of the event, the privacy disclaimer and the definition of visitors frequency.

3.6 The processes of visualization design

In designing a visualization, Sharp et al. (2007) indicate the need for a user-centered approach to development which means that the users and their goals are the driving force behind visualization development. They mention the process of visualization design into four activities:

1. Establishing requirements

   To have a well design visualization that is useful for people, the first thing that we must do is knowing and understanding who our target users and what is their goal. It forms the basis of the product's requirements and support design and development phases.
2. Designing alternatives

   There are two sub activities in this activity: conceptual design and concrete design. Conceptual design describes what people can do with a visualization and what visualization techniques are needed to visualize an information. Concrete design considers the elements of the visualization like colors, structures/layouts, etc.
3. Prototyping

   During the development, the performances and the reactions of intended user to understand visualizations are observed and measured. It can be done in many ways, for example through observing users, interviewing them or having a discussion with them.
4. Evaluating

   In this activity, the usability and acceptability of a visualization design is measured in terms of usability and user experience criteria. Evaluation activity complements and enhances the activities related to quality assurance and testing to ensure that the visualization is fit for purpose.

4. Data Collection

In this section, the data collection strategy is explained. Data collection in this research consisted of five procedures: semi-structured interviews, document studies, information visualization design, information visualization pre-testing, and information visualization effectiveness testing.

4. 1 Interviews

As a first step for this research, semi-structured interviews were conducted with three interviewees that have experience in information visualization especially using mobility information from Mezuro. Yin (2013) states that interviews are usually found in case study research and that they are one of the important sources of evidence. The semi-structured interview is chosen because the interviews are intended to be exploratory and this technique gives more flexibility in directing the interview. Each interview took time around 30 to 40 minutes. An interview protocol (see appendix A) was made in order to guide the interviews.

The semi-structured interviews were used to obtain insights and information as a comprehensive preparation for the visualization effectiveness testing. The goals of the interview were to discover criteria or components that are required for /in an effective information visualization, to identify the type of questions from customers and the information they need on mobility information, and to identify the important procedures in designing and implementing effective mobility information visualization. The first part of the interview contained questions focused on the effective information visualizations and mobility information. In the second part, the questions focused on the type of questions that customers ask and information they need especially on mobility information, based on the interviewee experiences. The last part of the interview contained questions from the information visualization examples that the interviewee had made. The intention here was to identify the reasons behind the information visualizations that they have created. For instance, they could explain why they choose a certain type of visualization to visualize the information. The interview was digitally recorded to avoid the loss of any essential information.

In table 4.1, the overview of interviewees involved are listed. We interviewed three participants in total who all were partners of Mezuro. They had experience with information visualization, mobility information, and event project cases. The interviewees were approached either by phone or email. The interviewees were asked for verbal consent which reconfirmed their rights and were asked for permission to record the interview for research purposes. In this thesis, we do not disclose the names of the interviewees. We only show their company's names and their positions.

Table 4.1 Overview of selected interviewees

Interviewee	Position	Company
1	Senior consultant	Impact information management
2	Senior researcher	Marble research
3	Senior consultant	Decisio

4.2 Document studies

Besides conducting semi-structured interviews, this research also uses visualization reports and presentations as data sources. The visualization reports and presentations were gathered from Mezuro and their partners (the interviewees). Before the interview, we asked the interviewees to send their visualization reports or presentations that were used in project cases with Mezuro, especially in event cases. Fortunately, all of the interviewees had documents that were related to this research. This was helpful because these documents gave the researcher knowledge about the type of visualizations that are mostly used in visualizing mobility information. Moreover, their visualizations were also used as the basis for making the visualizations of mobility information in this research.

4.3 Information visualization design

The visualizations that were used in this research were designed based on the theoretical background in section 3.5, the results from the interviews, and the document studies. Before we made the visualizations, we defined the information that should be shown and was usually asked by customers of Mezuro (see table 4.2). From the list below, we knew the information that customers might want to know from the mobility information of Mezuro especially in an event case.

Table 4.2 Mobility information that customers want to know from an event case

No	Information
1	Number of visitors and their origin
2	Type of transportation that is used by the visitors
3	Visitors average staying time

After we had defined the questions and information that were needed by the customers, we made three visualizations using the Tableau 10 software. Each visualization shows different information relevant to the questions from the customer. We showed the number of visitors, their origin, their average travelled distances, the type of transportation that they used, and their average staying time. We used a single case study which we focus on an event in the Netherlands (Serious Request-Glazen Huis) that was held on December 2015. We considered the following elements when creating the visualizations: visualization technique, type of chart, color, highlight, label, legend, structure/layout, and explanation of information.

The first thing we did was choosing the type of chart that we wanted to use in visualizing the information. We used bar charts because we wanted to show comparisons between the information that we have. Few (2004) states that bar charts can be used to support a quick comprehension of values and show comparisons of values between different categories. Besides bar charts, we used maps for representation of spatial-temporal data. In maps we used two techniques from Andrienko et al. (2003), the first technique was "map iteration" which we used to show differences between two moments in time. The second technique was the "changed map" which we used to show for each location or area the amount of change between two time moments and where different colors represent the changes.

After we had decided the type of chart for the visualizations, then we went to the colors selection for the visualizations. In selecting the colors, we chose colors from the “color blind colors” group in Tableau software to make sure that people who are colorblind can

recognize the data that were color coded. Next, we used soft or light colors to show most information and bright or dark colors to highlight information that is required greater attention (see table 3.2 for the details). De-emphasis highlighting intervention was also used on the visualizations to highlight the total number of extra visitors that visit the city center. Then, when we used colors to show a sequential range of quantitative values, in particular on a map, we used a single hue (or a small set of closely related hues) and varied intensity from light colors for low values to increasingly darker colors for high values.

To help viewers to see which value each bar or area represented we labeled the bar with the values of the data and we showed the name of the cities on the map. For the map, we made a legend to represent the colors which contained 3-5 numerical ranges that enabled fairly even distribution of data. We did not only use labels to help viewers to recognize the information clearly, but we also paid attention to the structure or layout of the visualizations. We ordered the visualizations from left to right where the left part of the visualization is expected to receive the highest attention from the viewers so the most important information should be put in this part. The other information was distributed based on their order of importance. To finalize the visualizations, we added information about the date of the event, the date of the regular day, the area of the event, the privacy disclaimer and the definition of visitors frequency.

The first visualization (Fig. 4.1) was created to show the effect of an event in city center on the number of visitors; and their origin. In the visualization, we showed the number of visitors per frequency type using bar charts and the percentage of inhabitants that visit the area of the event per city using maps. The next visualization (Fig. 4.2) was created to show the effect of an event in a city center on the preference of transportation type that was used by the visitors. In the visualization, we showed the visitors preference of transportation using bar charts and maps (per city). The last visualization (Fig. 4.3) was made to show visitors average staying time in the city center. In the visualization, we show the number of visitors in the city center for every hour using bar charts, and visitors' average staying time in the city center per province of origin using maps. For the larger images see appendix D.

The visualizations were custom made for the single case study (Serious Request-Glazen Huis). It took quite some time for us to make it not only because Tableau software was a new tool for us, but there were also some elements that have not been supported by Tableau. For example, to make the Netherlands map we made the custom geocoding (country, zip code, latitude and longitude).

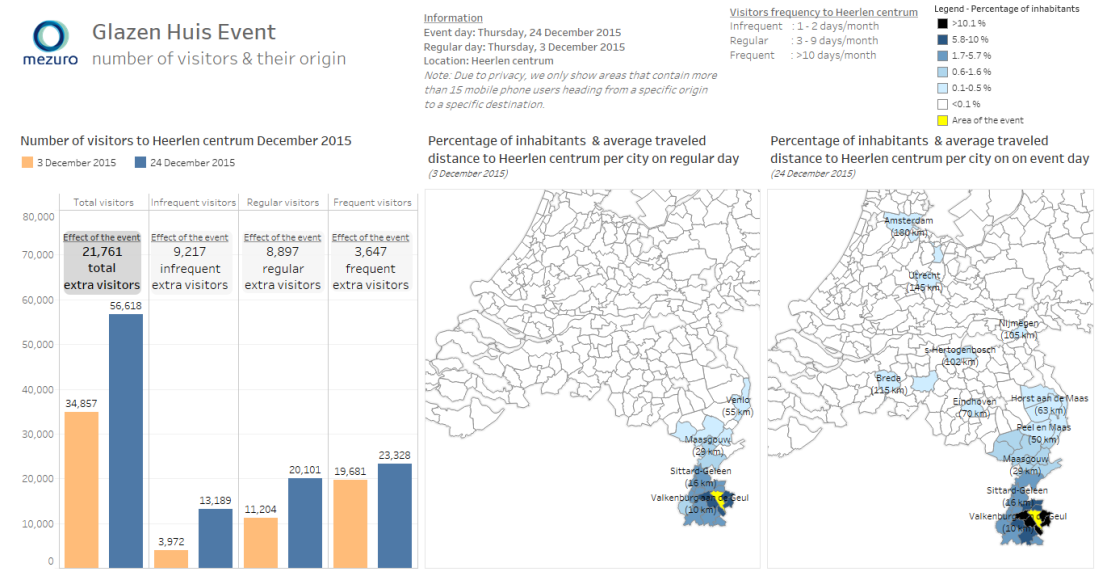


Figure 4.1 Visualization page 1

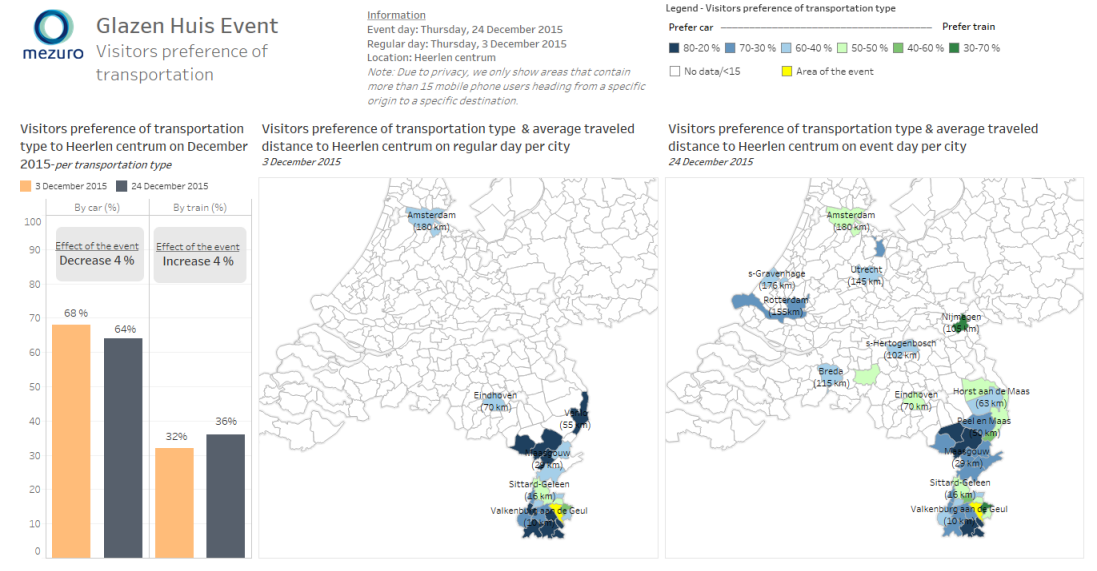


Figure 4.2 Visualization page 2

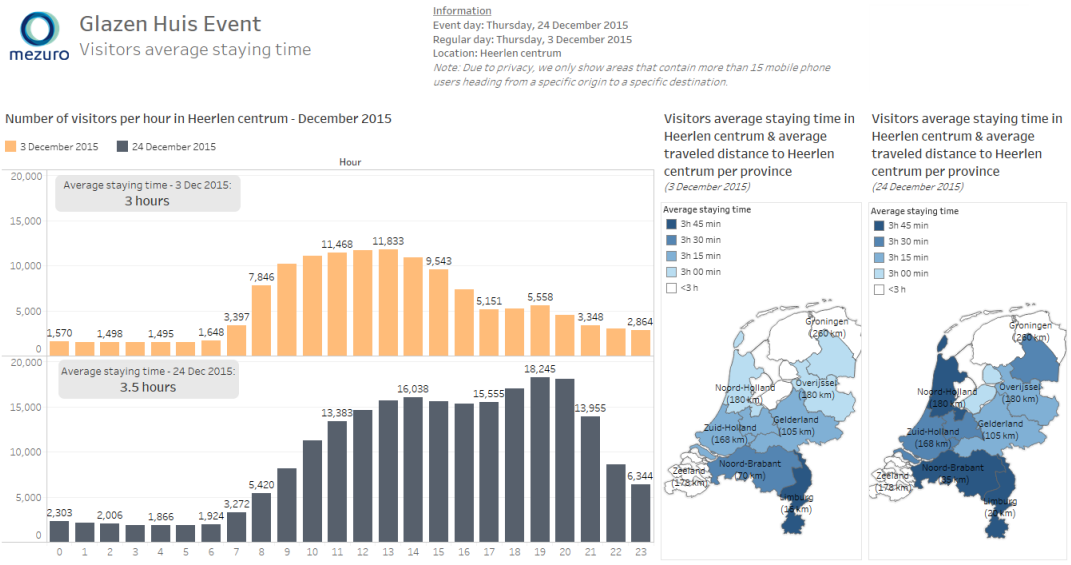


Figure 4.3 Visualization page 3

4.4 Information visualization pre-test

Before conducting the actual experiment for information visualization effectiveness testing, the visualizations, and the experiment protocol were evaluated using a pre-test. We did the pre-test with four participants from Mezuro who were excluded from the actual experiment. In the pre-test, each participant was given a set of 3 different visualizations and a list of questions. Using the information that was shown in each visualization, the participants had to answer questions that would also be used in the information visualization effectiveness testing. Each visualization displayed different types of information relevant to the question presented to them.

In the visualization evaluation, we asked the participants whether they had difficulties in understanding the visualizations or could easily extract the relevant information. We also asked them to give us suggestions how the visualizations should be improved. For the experiment protocol, we asked the participants to assess the protocol and the questions to determine whether the protocol and the questions were clear enough. Finally, the pre-test, we continued to conduct the actual information visualization effectiveness testing described in section 4.5.

4.5 Information visualization effectiveness testing

In this section, we discuss the qualitative and quantitative measures for visualization effectiveness that were used in the experiments. We tested the effectiveness of the visualizations using a scenario from the case study "Glazen Huis", described above. In the qualitative part, we asked the participants to think aloud during the process of answering the questions. Then in the quantitative part, we measured the accuracy (number of interpretation errors) and the efficiency (time) of each participant in an online survey.

4.5.1 Qualitative study

4.5.1.1 Participants (qualitative study)

In the experiments, 7 participants took part in the qualitative study. They were customers of Mezuro working at municipalities in the Netherlands. Table 4.3 shows the overview of the seven participants.

Table 4.3 Qualitative measurement participants

Participant (P)	Position	Municipality
1	Program officer	Sittard
2	Advisor / smart city innovation officer	Zeist
3	Economist in the spatial and economic developments department	Heerlen
4	Senior advisor research and statistic	Heerlen
5	Geoinformation advisor	Zeist
6	Geographic information specialist	Zeist
7	Marketer	Utrecht

4.5.1.2 Materials (qualitative study)

The participants were presented three information visualization designs discussed in section 4.3. They were also given a scenario from a case study in Mezuro and a list of 16 questions that they to answer using the visualizations. The scenario and questions are shown in appendix B.

4.5.1.3 Measurements (qualitative study)

In order to understand how the visualizations were analyzed and interpreted, the interpretation process of the participants was observed using the think aloud method. The think-aloud method is a method where the participants verbalize their thoughts during the process of answering the questions (Krug, 2011; Someren et al., 1994).

4.5.1.4 Procedures (qualitative study)

The participants were given a situation where they did a role play as an event planner in a municipality in the Netherland and there was an event "Serious Request (Glazen Huis)" that was held in their city on December 2015. From the information visualizations of the event made for the experiment the participants were asked to answer several questions. Each visualization showed different types of information that were relevant to the question presented to them. The participants were asked to voice their thoughts, feelings, and opinions while answering the questions. After observing the interpretation process of the participants, we asked some personal opinion questions regarding the visualizations. This qualitative measurement was digitally recorded.

4.5.1.5 Results (qualitative study)

Below, we present a summary of the interpretation processes of the participants when the three pages of the visualizations were shown to them. Overall, all of the participants felt that it was important for them to have or to use an information visualization in their current job. They thought that an information visualization was important because of it easier for them and for other people to understand the information. "*In my opinion, an excel sheet with a lot of data is difficult to read and understand, so a visualization is helpful*" (Participant 2). "*I notice that a visualization of information makes it easier to interpret. In my experience, it was very important for me to use a visualization when I worked with a huge amount of data like facts, figures, and numbers*" (Participant 1). They also said that an information visualization helped them to get the information quickly. "*The people here have very busy agendas and they need to get the information very quick, so it is very supporting to use visualizations*" (Participant 4). "*It is much easier to understand if you have something in a graphic than a text. So for me, it is easier and faster when I used the visualization*" (Participant 3).

From the interviews, we found that all of the participants were still using a static (on paper) information visualization and mostly they used it in reports or presentations. However, they planned to have the dynamic information visualization like a dashboard in the future. "*So currently we are still using a static visualization, but hopefully in the future we will have a dynamic visualization*" (Participant 2). *"For now it still the static one, but we like to develop a dashboard so we can use it interactively"* (Participant 7).

Of the tasks that the participants usually conduct with an information visualization (table 4.4), three were most frequently mentioned by the participants: 'comparison', 'evaluation', and 'exploration'. Comparison tasks are tasks in which the participants use an information visualization to compare different data, for example the number of people in the city center on a normal day with the day of a special event. The second task was evaluation which means that the information visualization was used to evaluate the goals or targets that they have made, for example they evaluate the effect of a certain project, policy or special event in their city. The last task was exploration which means that they used an information visualization to see and explore trends in the information.

Table 4.4 Type of tasks from an information visualization

Participant	**Tasks that the participants usually conduct with an information visualization**
P1	See trends and comparison
P2	See trends, comparison, and evaluation
P3	Comparison
P4	Evaluation and monitoring
P5	Exploration and evaluation
P6	Exploration
P7	Give information and comparison

Visualization page 1

P: Participant
Q: Question
Process: The way that is used by each participant in answering each question.

The process is expected & the answer is correct
The process is expected, but the answer is wrong
The process is unexpected, but the answer is correct
The process is unexpected & the answer is wrong

Correct answers							
	P1	P2	P3	P4	P5	P6	P7
Visualization 1							
Q1							
Q2							
Q3							
Q4							
Q5							
Q6							

Figure 4.4 The processes and correct answers (visualization page 1)

In figure 4.4, overall we see that the process used by the participants to answer the questions has been as expected. However, there were some wrong answers which meant that the visualization was not interpreted correctly by the participants. In order to improve the visualization, we looked at how the participants used and understood the elements in the visualization. We also looked at their comments on the visualization (see appendix C for the full analysis table of visualization 1).

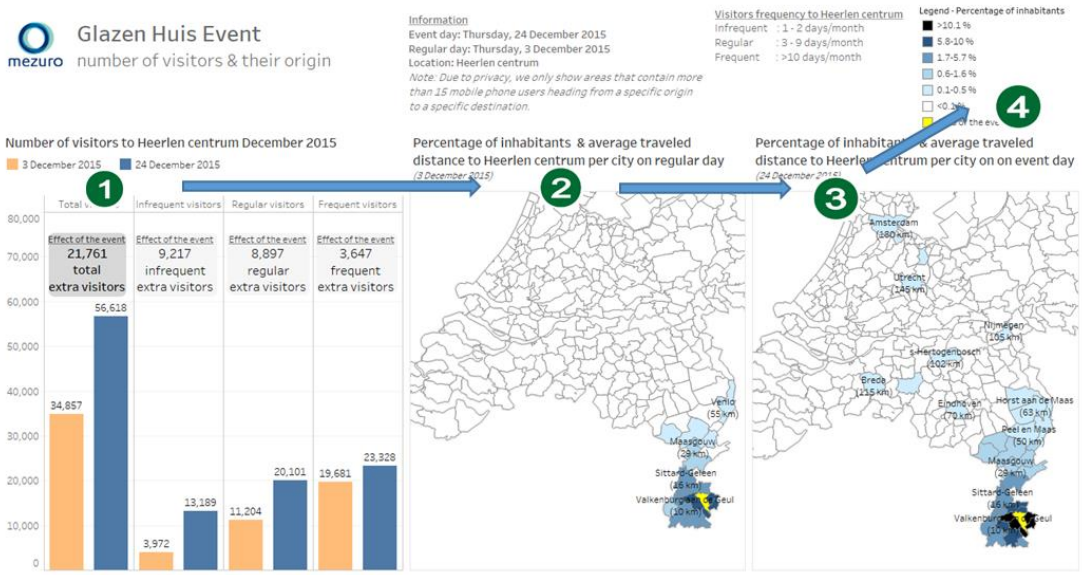


Figure 4.5 Parts that gained attention by the participants (visualization page 1)

First, we found that all participants started to look at the left side of the visualization, where the bar chart was located (fig.4.5). Then they continued to read the whole visualization from left to right (as expected). Some of the participants indicated that they had difficulties to understand the visualization because the visualization had too many texts or redundancy in texts. *"The visualization had too many texts*" (Participant 3). Also, we found that they did not pay much attention to the information titled "Visitors frequency to Heerlen centrum". *"The position of the information about the visitor's frequency is a little bit too far"* (Participant 4). The participants gave a solution for this issue by moving the information closer to the bar chart. So, on the improved visualization page 1 we simplified the title on the maps and moved the "Visitors frequency to Heerlen centrum" information closer to the bar chart. Lastly, we asked the participants to grade the visualization on the ease of use and the aesthetic quality from 1-10. For the ease of use, on average was 7.5/10 where the lowest grade was 6 and the highest grade was 9. For the aesthetic quality, on average the participants assigned 7.4/10 where the lowest and the highest grade were also 6 and 9.

Visualization page 2

P: Participant
Q: Question
Process: The way that is used by each participant in answering each question.

The process is expected & the answer is correct
The process is expected, but the answer is wrong
The process is unexpected, but the answer is correct
The process is unexpected & the answer is wrong

Correct answers							
	P1	P2	P3	P4	P5	P6	P7
Visualization 2							
Q1							
Q2							
Q3							
Q4							

Figure 4.6 The processes and correct answers (visualization page 2)

Although we do not see any red colors in figure 4.6, we still see that the visualization was not interpreted correctly and needed to be improved. Nevertheless, the process used by all of the participants to answer the questions was as expected. Therefore, we looked at how the participants used the elements in the visualization and at their comments (see appendix C for the full analysis table of visualization 2).

First, we found that 5 out of 7 participants started to look at the left side of the visualization, where the bar chart (fig.4.7). Then they continued to read the whole visualization from left to right (as expected). The other 2 participants started to look at the right side of the visualization which showed the legend of the maps (fig.4.8). Then they continued to read the maps and did not pay much attention to the bar chart on the left side of the visualization. "*I just went straight to the map and I think it is because the amount of space it takes on the screen"* (Participant 1). Next, we saw that almost all participants had difficulties to understand the legend of the maps. "*I think it is a little bit difficult for me to understand the legend, maybe because of the percentage group*" (Participant 2). They gave a solution for this issue by changing the percentage into an absolute number, for example only 60% instead of a range, 40-60%. Subsequently, we again found that some of the participants felt that the visualization contained too many

texts. So, on the improved visualization 2 we simplified the title on the maps and we changed the legend of the maps into an absolute number. For the ease of use, on average was 6.8/10 where the lowest grade was 5 and the highest grade was 9. For the aesthetic quality, on average the participants assigned 7.6/10 where the lowest and the highest grade were also 6 and 9.

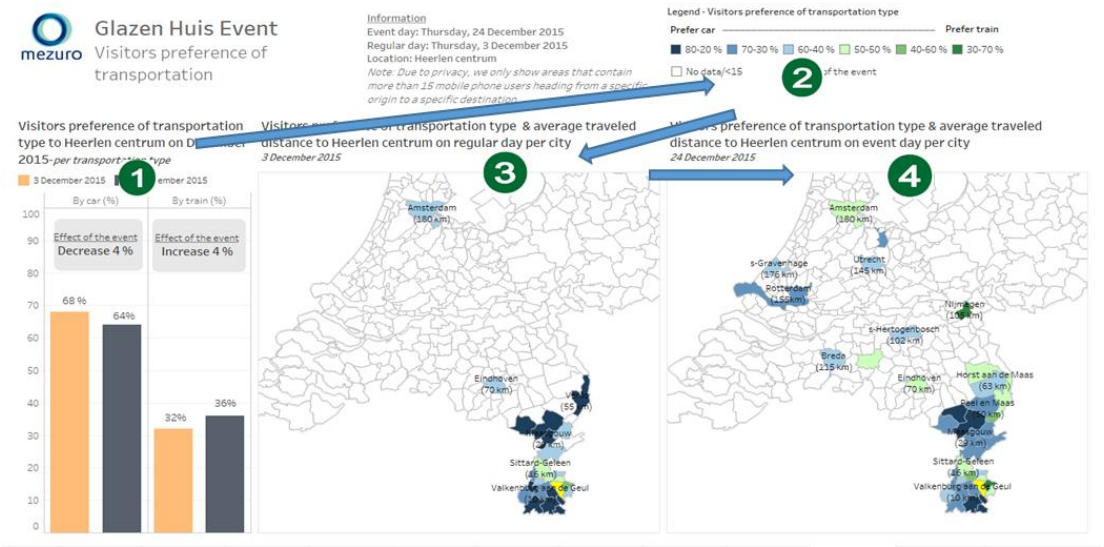


Figure 4.7 Parts that gained attention by the 5 participants (visualization page 2)

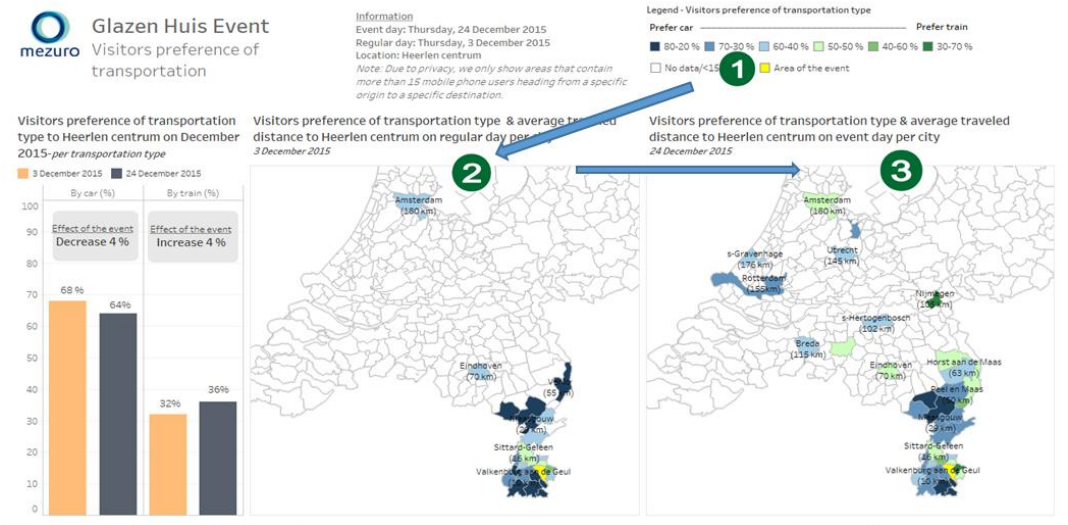


Figure 4.8 Parts that gained attention by the 2 participants (visualization page 3)

Visualization page 3

P: Participant
Q: Question
X: Cannot be compared because the question has been revised for P3-P6.
Process: The way that is used by each participant in answering each question.

The process is expected & the answer is correct
The process is expected, but the answer is wrong
The process is unexpected, but the answer is correct
The process is unexpected & the answer is wrong

Correct answers							
	P1	P2	P3	P4	P5	P6	P7
Visualization 3							
Q1							
Q2							
Q3	X	X					
Q4							
Q5							
Q6							

Figure 4.9 The processes and correct answers (visualization 3)

From figure 4.9, overall we see that almost all of the questions can be answered correctly by the participants and also the process used by them has been as expected. However, we still improved the visualization applying the same changes as suggested by the participants for the first two pages (see appendix C for the full analysis table of visualization 3).

First, we found that there was no difference in the way that the participants started to look at the visualization. They all started to look at the left side of the visualization (fig.4.10). Although the participants felt that this visualization is clearer than the other two visualizations, they still gave us comments to improve the visualization. Three participants said that it would be better if the hours (y-axis) were in both bar charts. "*For me this visualization is much clearer than the other two and I think you can put the hours (y-axis) in both bar charts*" (Participant 3). So, on the improved visualization 3 we put the hours (y-axis) in both bar charts. We also simplified the title on the maps. Lastly, the

evaluation of the ease of use, on average 7.8/10, where the lowest grade was 6 and the highest grade was 9. For the aesthetic quality, the results were an average of 7.8/10 where the lowest grade was 6.5 and the highest grade was 9.

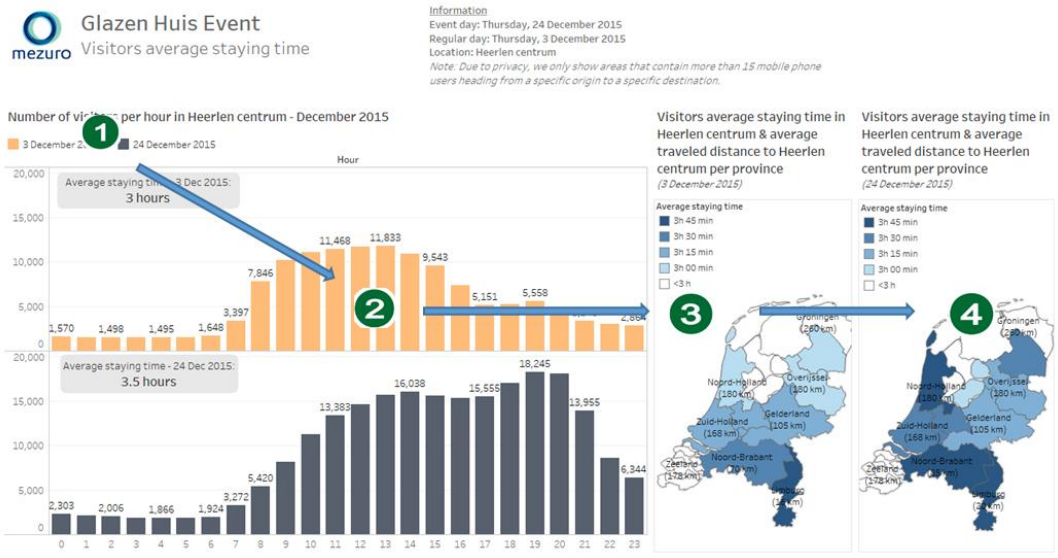


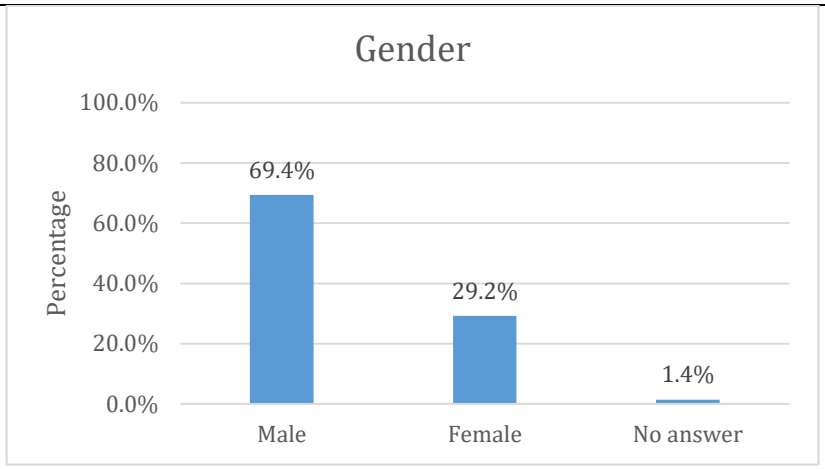
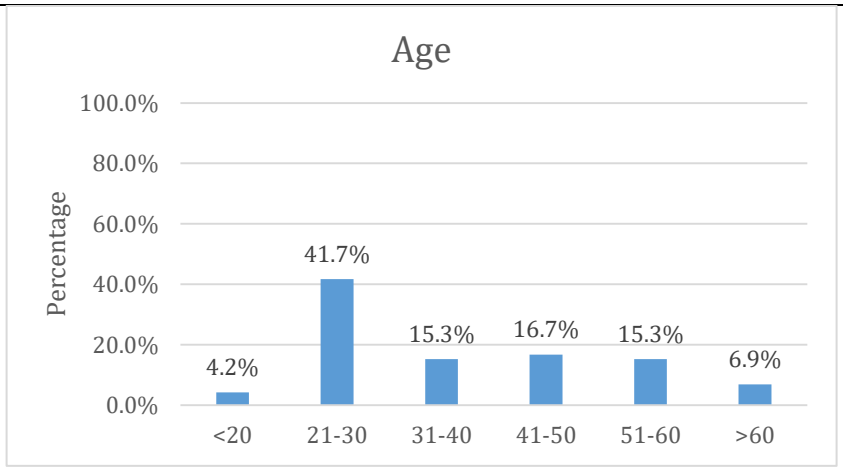
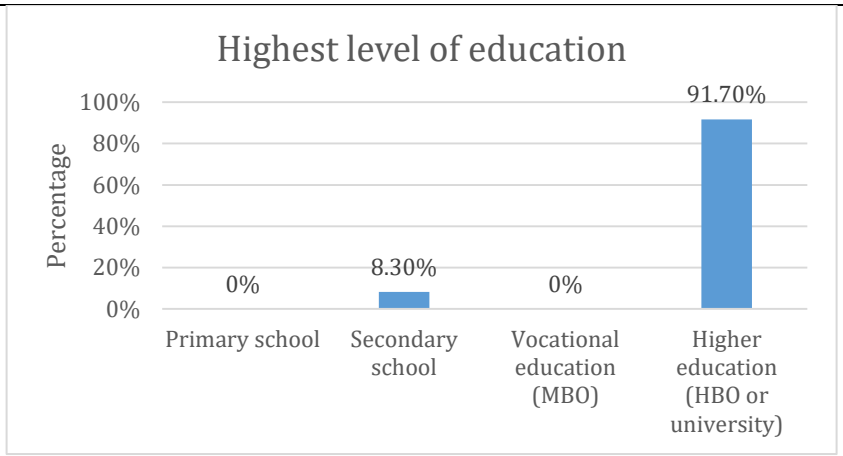
Figure 4.10 Parts that gained attention by the participants (visualization page 3)

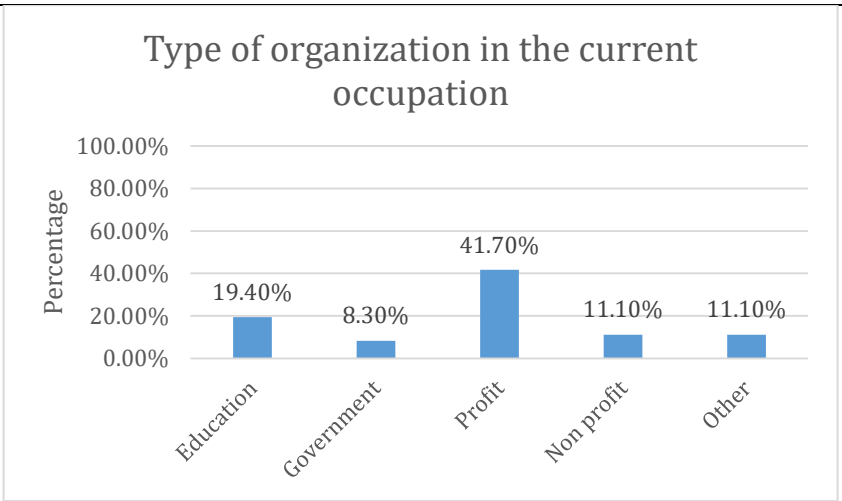
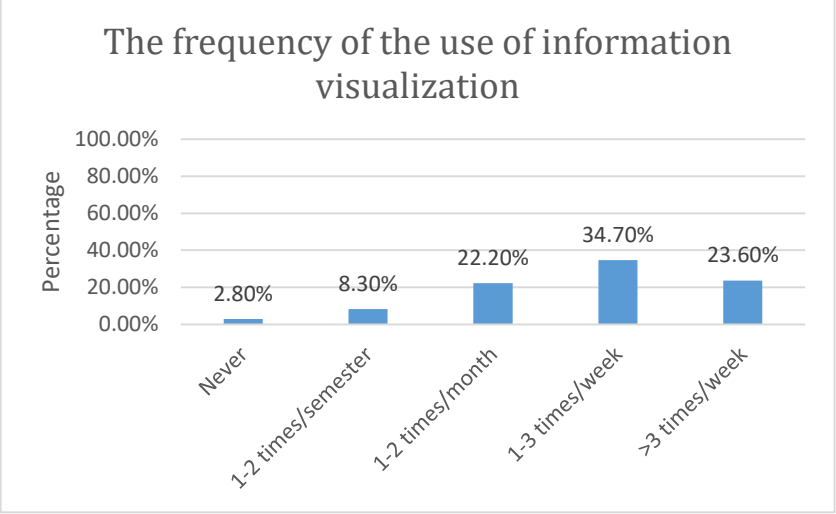
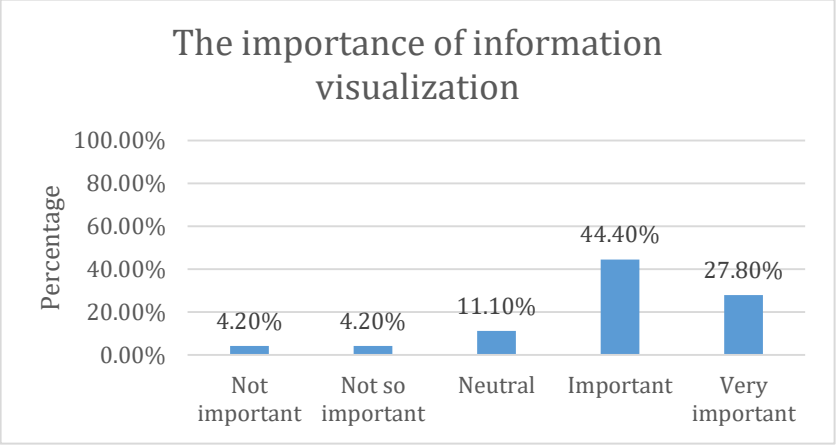
4.5.2 Quantitative study

4.5.2.1 Participants (quantitative study)

The online survey generated 72 responses. The participants of the survey reacted on the invitation that was published on LinkedIn and in Mezuro's newsletter. We also sent the online survey to students at Utrecht University. We divided them into 3 groups, one group used the visualization designs described in section 4.3, one group used the visualization designs that had been improved based on the result of the qualitative measurement in section 4.5.1.5, and another group used reports that only had tables and numbers on which the visualization in section 4.3 were based. Table 4.5 summarizes the demographic profile of the participants.

Table 4.5 Demographic survey participants

Item	Percentage
Gender	
Age	
Highest level of education	

Type of organization in the current occupation	
The frequency of the use of information visualization	
The importance of information visualization in their daily jobs	

4.5.2.2 Materials (quantitative study)

In the quantitative part, three different sets of materials were used:

1. The three visualization designs from section 4.3.
2. The three visualization designs that had been improved (fig. 4.11-4.13) based on the result of the qualitative measurement in section 4.5.1.5. The differences between the two visualization designs are:
    - Improved visualization page 1: we simplified the title on the maps and moved the "Visitors frequency to Heerlen centrum" information closer to the bar chart.
    - Improved visualization page 2: we simplified the title on the maps and we changed the legend of the maps into an absolute number.
    - Improved visualization page 3: we simplified the title on the maps and we put the hours (y-axis) in both bar charts.
3. Another three information reports (fig. 4.14-4.16) that only had tables and numbers on which the visualization in section 4.3 were based.

For the larger images see appendix D. To collect the quantitative data, we made an online survey using Qualtrics. The 16 questions are shown in appendix E.

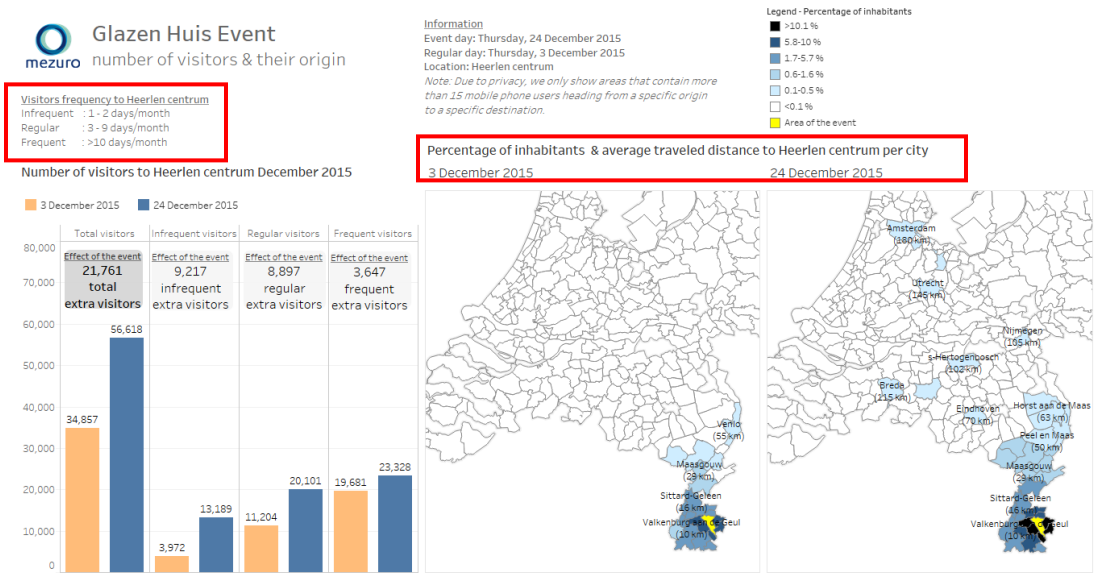


Figure 4.11 Improved visualization page 1

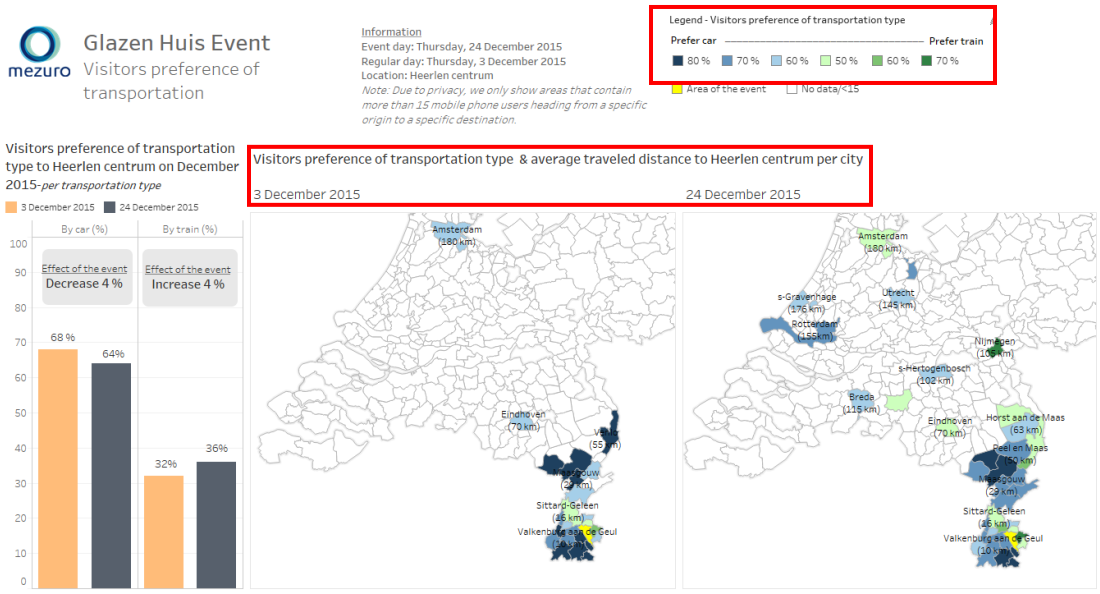


Figure 4.12 Improved visualization page 2

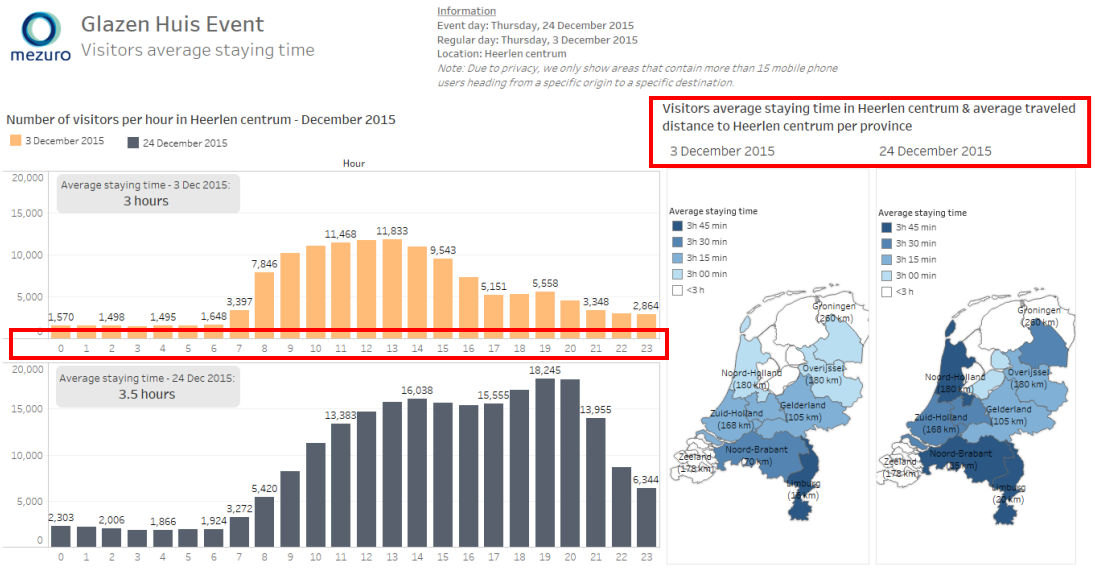


Figure 4.13 Improved visualization page 3

Glazen Huis Event

mezuro Number of visitors & their origin

Information

Event day: Thursday, 24 December 2015
Regular day: Thursday, 3 December 2015
Location: Heerlen centrum
Note: Due to privacy, we only show areas that contain more than 15 mobile phone users heading from a specific origin to a specific destination.

Visitors frequency to Heerlen centrum

Infrequent	: 1 - 2 days/month
Regular	: 3 - 9 days/month
Frequent	: >10 days/month

Number of visitors to Heerlen centrum on December 2015

Total visitors		Infrequent visitors		Regular visitors		Frequent visitors	
21,761 extra visitors		9,217 extra visitors		8,897 extra visitors		3,647 extra visitors	
3 December	24 December	3 December	24 December	3 December	24 December	3 December	24 December
34,857	56,618	3,972	13,189	11,204	20,101	19,681	23,328

Percentage of inhabitants that visit Heerlen centrum per city on December 2015

Origin (traveled distance)	Percentage of inhabitants		Origin (traveled distance)	Percentage of inhabitants	
	3 December 2015	24 December 2015		3 December 2015	24 December 2015
Amsterdam (180 km)	No data/<15 phone users	0.1-0.5 %	Nijmegen (105 km)	No data/<15 phone users	0.1-0.5 %
Beek (10 km)	1.7-5.7 %	1.7-5.7 %	Nuth (7 km)	5.8-10 %	>10.1 %
Beesel (42 km)	No data/<15 phone users	0.6-1.6 %	Onderbanken (9.5 km)	1.7-5.7 %	5.8-10 %
Breda (115 km)	No data/<15 phone users	0.1-0.5 %	Peel en Maas (50 km)	No data/<15 phone users	0.1-0.5 %
Brunssum (6 km)	1.7-5.7 %	5.8-10 %	Roerdalen (27 km)	No data/<15 phone users	0.6-1.6 %
Echt-Susteren (23 km)	0.6-1.6 %	1.7-5.7 %	Roermond (34 km)	0.1-0.5 %	0.6-1.6 %
Eijsden-Margraten (14 km)	1.7-5.7 %	1.7-5.7 %	Schinnen (8.5 km)	1.7-5.7 %	5.8-10 %
Eindhoven (70 km)	No data/<15 phone users	0.1-0.5 %	s-Hertogenbosch (102 km)	No data/<15 phone users	0.1-0.5 %
Gulpen-Wittem (10 km)	1.7-5.7 %	5.8-10 %	Simpelveld (7 km)	5.8-10 %	>10.1 %
Heerlen	Area of the event		Sittard-Geleen (16 km)	1.7-5.7 %	1.7-5.7 %
Hilversum (158 km)	No data/<15 phone users	0.1-0.5 %	Stein (16.5 km)	1.7-5.7 %	1.7-5.7 %
Horst aan de Maas (63 km)	No data/<15 phone users	0.1-0.5 %	Tilburg (97 km)	No data/<15 phone users	0.1-0.5 %
Kerkrade (7 km)	5.8-10 %	>10.1 %	Utrecht (145 km)	<15 mobile phone users	0.1-0.5 %
Landgraaf (5 km)	5.8-10 %	>10.1 %	Vaals (14 km)	No data/<15 phone users	1.7-5.7 %
Leudal (39 km)	0.1-0.5 %	0.6-1.6 %	Valkenburg aan de Geul (10 km)	1.7-5.7 %	5.8-10 %
Maasgouw (29 km)	0.6-1.6 %	0.6-1.6 %	Venlo (55 km)	0.1-0.5 %	0.1-0.5 %
Maastricht (19 km)	0.6-1.6 %	1.7-5.7 %	Venray (72 km)	No data/<15 phone users	0.1-0.5 %
Meerssen (15 km)	1.7-5.7 %	1.7-5.7 %	Voerendaal (4 km)	5.8-10 %	>10.1 %
Nederweert (48 km)	No data/<15 phone users	0.6-1.6 %	Weert (45 km)	0.1-0.5 %	0.6-1.6 %

Figure 4.14 Information report page 1

mezuro

Glazen Huis Event

Visitors preference of transportation

Information

Event day: Thursday, 24 December 2015
Regular day: Thursday, 3 December 2015
Location: Heerlen centrum
Note: Due to privacy, we only show areas that contain more than 15 mobile phone users heading from a specific origin to a specific destination.

Visitors preference of transportation type to Heerlen centrum on December 2015

By car		By train	
Decrease 4%		Increase 4%	
3 December	24 December	3 December	24 December
68%	64%	32%	36%

Visitors preference of transportation type to Heerlen centrum per city on December 2015

Origin (traveled distance)	Transportation preferences (% car - % train)		Origin (traveled distance)	Transportation preferences (% car -% train)	
	3 December 2015	24 December 2015		3 December 2015	24 December 2015
Amsterdam (180 km)	60%-40%	50%-50%	Nuth (7 km)	70%-30%	70%-30%
Beek (10 km)	60%-40%	60%-40%	Onderbanken (9.5 km)	60%-40%	60%-40%
Beesel (42 km)	No data/<15 phone users	40%-60%	Peel en Maas (50 km)	No data/<15 phone users	70%-30%
Breda (115 km)	No data/<15 phone users	60%-40%	Roerdalen (27 km)	No data/<15 phone users	70%-30%
Brunssum (6 km)	50%-50%	50%-50%	Roermond (34 km)	60%-40%	70%-30%
Echt-Susteren (23 km)	60%-40%	70%-30%	Rotterdam (155 km)	No data/<15 phone users	70%-30%
Eijsden-Margraten (14 km)	80%-20%	70%-30%	Schinnen (8.5 km)	50%-50%	40%-60%
Eindhoven (70 km)	60%-40%	50%-50%	s-Gravenhage (176 km)	No data/<15 phone users	60%-40%
Gulpen-Wittem (10 km)	80%-20%	80%-20%	s-Hertogenbosch (102 km)	No data/<15 phone users	60%-40%
Heerlen	Area of the event		Simpelveld (7 km)	80%-20%	80%-20%
Hilversum (158 km)	No data/<15 phone users	70%-30%	Sittard-Geleen (16 km)	50%-50%	50%-50%
Horst aan de Maas (63 km)	No data/<15 phone users	60%-40%	Stein (16.5 km)	70%-30%	70%-30%
Kerkrade (7 km)	60%-40%	50%-50%	Tilburg (97 km)	No data/<15 phone users	50%-50%
Landgraaf (5 km)	40%-60%	30%-70%	Utrecht (145 km)	No data/<15 phone users	60%-40%
Leudal (39 km)	80%-20%	80%-20%	Vaals (14 km)	80%-20%	80%-20%
Maasgouw (29 km)	80%-20%	70%-30%	Valkenburg aan de Geul (10 km)	70%-30%	70%-30%
Maastricht (19 km)	70%-30%	60%-40%	Venlo (55 km)	80%-20%	50%-50%
Meerssen (15 km)	80%-20%	70%-30%	Venray (72 km)	No data/<15 phone users	50%-50%
Nederweert (48 km)	No data/<15 phone users	80%-20%	Voerendaal (4 km)	80%-20%	70%-30%
Nijmegen (105 km)	No data/<15 phone users	30%-70%	Weert (45 km)	80%-20%	70%-30%

Figure 4.15 Information report page 2

mezuro **Glazen Huis Event**
Visitors average staying time

Information
Event day: Thursday, 24 December 2015
Regular day: Thursday, 3 December 2015
Location: Heerlen centrum
Note: Due to privacy, we only show areas that contain more than 15 mobile phone users heading from a specific origin to a specific destination.

Number of visitors per hour in Heerlen centrum – December 2015

Average staying time 3 December 2015	Average staying time 24 December 2015
3 hours	3.5 hours

Hour	Day type		Hour	Day type	
	3 December 2015	24 December 2015		3 December 2015	24 December 2015
0:00	1,570	2,303	12:00	11,730	14,668
1:00	1,537	2,157	13:00	11,833	15,723
2:00	1,498	2,006	14:00	10,938	16,038
3:00	1,473	1,840	15:00	9,543	15,596
4:00	1,495	1,866	16:00	7,339	15,359
5:00	1,519	1,889	17:00	5,151	15,555
6:00	1,648	1,924	18:00	5,259	17,062
7:00	3,397	3,272	19:00	5,558	18,245
8:00	7,846	5,420	20:00	4,521	18,128
9:00	10,174	8,190	21:00	3,348	13,955
10:00	11,104	11,234	22:00	3,015	8,647
11:00	11,468	13,383	23:00	2,864	6,344

Visitors average staying time in Heerlen centrum per province

Province (avg. traveled distance)	Average staying time	
	3 December 2015	24 December 2015
Drenthe (237 km)	3h 00 min	3h 30 min
Flevoland (183 km)	<3 h	3h 00 min
Friesland (255 km)	<3 h	<3 h
Gelderland (105 km)	3h 15 min	3h 15 min
Groningen (260 km)	<3 h	<3 h
Limburg (18 km)	3h 45 min	3h 45 min
Noord-Brabant (80 km)	3h 30 min	3h 45 min
Noord-Holland (180 km)	3h 00 min	3h 45 min
Overijssel (180 km)	3h 00 min	3h 15 min
Utrecht (150 km)	3h 15 min	3h 30 min
Zeeland (178 km)	<3 h	<3 h
Zuid-Holland (168 km)	3h 15 min	3h 30 min

Figure 4.16 Information report page 3

4.5.2.3 Measurements (quantitative study)

In this part, we measured the accuracy based on the number of interpretation errors; the lower the number of interpretation errors, the higher the accuracy of the visualization is. We also determined the efficiency of the visualizations measuring the time each

participant took to solve a single question in seconds. The lower the time to answer the questions, the higher the efficiency of the visualizations is.

4.5.2.4 Procedures (quantitative study)

The participants were presented with the same situation as in the qualitative study (section 4.5.1.4). They were assigned to three groups. The first group used the original visualization designs (section 4.3), the second group used the improved visualization designs (section 4.5.2.2), and the last group used the information report that only had tables and numbers. We informed the participants that they should complete the questions as accurately as possible in one session.

4.5.2.5 Results (quantitative study)

Below we show the quantitative data results from the online survey. Before we conducted the analysis, we checked the missing values, the outliers and the distribution of the results. In total there were two missing values (one in version 2 and one in version 3). For the outliers in the task completion time, we used the methods suggested by Field (2009). First we removed the outliers that had extreme values. We define extreme values as the task completion time which we believe is not a reasonable result for the population that we intended to sample. For example, we found one question that was answered in 3564 seconds which indicates that a factor not related to the survey affected the result. Then, we did a z-score test and we used the *"mean plus two standard deviations"* method (see appendix F for the details). Table 4.6 shows whether the results were normally distributed or not, which we checked using Kolmogorov- Smirnov test (see appendix F for the details).

Table 4.6 Normal distribution results (Kolmogorov- Smirnov test)

	Version 1	**Version 2**	**Version 3**
Page 1			
Q1	Not normal	Normal	Normal
Q2	Not normal	Not normal	Normal
Q3	Not normal	Not normal	Not normal
Q4	Not normal	Normal	Normal
Q5	Not normal	Normal	Normal
Q6	Not normal	Normal	Normal

Page 2			
Q1	Normal	Not normal	Not normal
Q2	Normal	Normal	Normal
Q3	Not normal	Normal	Normal
Q4	Not normal	Not normal	Normal
Page 3			
Q1	Not normal	Not normal	Not normal
Q2	Not normal	Normal	Normal
Q3	Not normal	Not normal	Normal
Q4	Not normal	Normal	Normal
Q5	Not normal	Normal	Normal
Q6	Not normal	Normal	Not normal

The first analysis performed was to determine the time needed by the participants to answer the questions (the task completion time) and the number of incorrect answers. For each version (group), the average task completion time was calculated from the last click time in seconds. The number of incorrect answers was the total number of incorrect answers received. Table 4.7 shows the results.

Table 4.7 The average task completion and the number of incorrect answers

Version 1: Group of participants that used the original visualization.

Version 2: Group of participants that used the improved visualization.

Version 3: Group of participants that used the report (only tables and numbers).

Page 1: Visualization shows information on the number of visitors and their origin.

Page 2: Visualization shows information on the type of transportation.

Page 3: Visualization shows information on the visitors average staying time.

Green: The fastest average task completion time between the 3 versions, per question.

Yellow: The lowest number of incorrect answers between the 3 versions, per question.

	Version 1 (N=28)		Version 2 (N=23)		Version 3 (N=21)	
Page 1	Avg. time in second (last click)	Total incorrect answers	Avg. time in second (last click)	Total incorrect answers	Avg. time in second (last click)	Total incorrect answers
Q1	57.24	1	70.24	3	94.20	0
Q2	32.44	5	55.40	3	108.89	6
Q3	28.57	3	30.69	6	38.60	2
Q4	43.94	12	47.25	8	116.63	12
Q5	50.18	2	53.32	7	64.55	1
Q6	31.23	22	29.31	18	37.49	12
Sub total	243.60	45	286.21	45	460.36	33
Page 2						
Q1	53.48	2	68.61	5	57.79	3
Q2	35.47	9	46.18	7	56.27	6
Q3	49.47	9	37.29	10	47.48	5
Q4	31.85	21	29.70	6	25.91	4
Sub total	170.26	41	181.77	28	187.45	18
Page 3						
Q1	50.66	0	38.43	2	36.64	2
Q2	39.25	0	33.26	2	55.31	1
Q3	26.44	1	45.59	2	25.14	2
Q4	30.50	2	30.52	6	30.19	2
Q5	57.00	10	65.42	9	82.01	16
Q6	27.99	1	27.37	3	28.18	0
Sub total	231.84	14	240.58	24	257.46	23
TOTAL	645.70	100	708.57	97	905.27	74

4.5.2.5.1 Analysis of the efficiency

To determine if the differences in efficiency, or average task completion time, between the versions is statistically significant, we conducted a statistical analysis. First, we did a one way ANOVA. The results revealed statistically significant differences between versions, F (2, 69) = 5.180, p < .05. Furthermore, we conducted a statistical test for each question and between versions. We did a parametric and non-parametric test (see

appendix F for the details). Table 4.8 shows the results of the tests. In the parametric test, we used the one-way ANOVA for page 2 – Q2 because:

- It has a normal distribution.
- Our independent variables are nominal and more than 2 (version 1, 2, 3).
- Our dependent variables are ratio (time in seconds).

In the non-parametric test I used the Kruskal-Wallis test for the other questions because:

- They have a non-normal distribution.
- Our independent variables are nominal and more than 2 (version 1, 2, 3).
- Our dependent variables are ratio (time in seconds).

Table 4.8 The results of parametric and non-parametric test (efficiency)

Page 1	
Q1	Significant
Q2	Significant
Q4	Significant
Page 2	
Q2	Significant
Page 3	
Q2	Significant

After we had found the significant results, we continued to do the post hoc test using the Tukey test and the Mann-Whitney test (see appendix F for the details). Field (2009) states that in order to ensure that the Type I errors do not build up to more than .05, some adjustment needs to be made. The method that can be used is a Bonferroni correction for which we use a critical value of .05 divided by the number of tests we have conducted. So then we made a set of comparisons together with the hypotheses:

- Test 1 : The current visualization (version 1) compared to the improved visualization (version 2).
  H0 : There is no difference between version 1 & version 2
  H1 : There is a difference between version 1 & version 2
- Test 2 : The current visualization (version 1) compared to the report (version 3)

H0 : There is no difference between version 1 & version 3

H1 : There is a difference between version 1 & version 3 (version 1 is more efficient than version 3)

- Test 3 : The improved visualization (version 2) compared to the report (version 3)

  H0 : There is no difference between version 2 & version 3

  H1 : There is a difference between version 2 & version 3 (version 2 is more efficient than version 3)

This results in three tests, so rather than use .05 as our critical level of significance, we use .05/3 = .0167. Table 4.9 presents the results of the analysis per question that were significant.

Table 4.9 The results of the analysis in the efficiency

Page 1	**Results**
Q1 How did the event affect the total number of visitors to Heerlen city center? The increase of visitors is:	**Time needed to answer the question/task completion time (efficiency) was significantly affected by the version of visualizations**, $H(2) = 7.50$, $p < .05$. Mann–Whitney tests were used to follow up this finding. A Bonferroni correction was applied and so all effects are reported at a .0167 level of significance. It appeared that task completion time (efficiency) were no different when version 1 and 2 ($U = 276$, $r = -.12$) or version 2 and 3 ($U = 166$, $r = -.27$). However, **there was a significant difference between version 1 and 3 ($U = 160$, $r = -.39$). It means that a visualization is more efficient than a table.**
Q2 What differences do you notice between the event day and the regular day with regard to the visitor's origin?	**Time needed to answer the question/task completion time (efficiency) was significantly affected by the version of visualizations**, $H(2) = 22.93$, $p < .05$. Mann–Whitney tests were used to follow up this finding. A Bonferroni correction was applied and so all effects are reported at a .0167 level of significance. It appeared that task completion time (efficiency) were different between version 1 and 2 ($U = 181$, $r = -.37$) or version 1 and 3 ($U = 73$, $r = -.64$) or version 2 and 3 ($U = 128$, $r = -.40$). **The effect size in version 1 and 3 is the biggest. It means that a visualization has more effect in the efficiency than a table (a map has more effect in the efficiency than a table.)**

Q4 What is the difference in the longest traveled distance on the event day compared to the regular day?	**Time needed to answer the question/task completion time (efficiency) was significantly affected by the version of visualizations**, $H(2) = 22.25$, $p < .05$. Mann–Whitney tests were used to follow up this finding. A Bonferroni correction was applied and so all effects are reported at a .0167 level of significance. It appeared that task completion time (efficiency) were no different between version 1 and 2 ($U = 304$, $r = -.05$). However, **there was a significant difference between version 1 and 3 ($U = 71$, $r = -.62$) or version 2 and 3 ($U = 71$, $r = -.58$). It means that a visualization is more efficient than a table (map is more efficient than table).**
Page 2	
Q2 Which cities showed an increase in the visitors preference of public transportation (train) to Heerlen city center, on the event day, compared to the regular day?	**The results revealed statistically significant differences between versions**, $F(2, 68) = 4.154$, $p < .05$. A post hoc Tukey test revealed statistically significant differences between version 1 and version 3 (M= 20.8, SD=7.25). **It means that a visualization is more efficient than a table (map is more efficient than table).**
Page 3	
Q2 When was the busiest time in Heerlen city center on the regular day and on the event day?	**Time needed to answer the question/task completion time (efficiency) was significantly affected by the version of visualizations**, $H(2) = 11.38$, $p < .05$. Mann–Whitney tests were used to follow up this finding. A Bonferroni correction was applied and so all effects are reported at a .0167 level of significance. It appeared that task completion time (efficiency) were no different when version 1 and 2 ($U = 275$, $r = -.13$). However, **there was a significant difference between version 1 and 3 ($U = 154$, $r = -.40$) or version 2 and 3 ($U = 115$, $r = -.45$). It means that a visualization is more efficient than a table (bar chart is more efficient than table).**

4.5.2.5.2 Analysis of the accuracy

After we analyzed the efficiency, we continued with the analysis of the accuracy. The first analysis performed was to determine the score of the incorrect answers for each version as shown in table 4.7. However, statistical tests are required to determine if differences between versions are significant. Thus, we did statistical tests in each question and between versions to determine a statistically significant difference in the accuracy.

To determine if the differences in accuracy, or number of incorrect answers, between the versions is statistically significant, we conducted a test in each question and between versions. We did a Kruskal-Wallis test (table 4.10) because:

- They have a non-normal distribution.
- Our independent variables are nominal and more than 2 (version 1, 2, 3).
- Our dependent variables are ratio (number of incorrect answers).

Table 4.10 The results of Kruskal-Wallis test (accuracy)

Page 1	**Significant/ not significant**
Q5	Significant
Page 2	
Q4	Significant
Page 3	
Q5	Significant

After we had found the significant results, we continued to do the post hoc test using the the Mann-Whitney test like we did for efficiency (see appendix F for the details). Then we made a set of comparisons together with the hypotheses:

- Test 1 : The current visualization (version 1) compared to the improved visualization (version 2).
  H0 : There is no difference between version 1 & version 2
  H1 : There is a difference between version 1 & version 2
- Test 2 : The current visualization (version 1) compared to the report (version 3)
  H0 : There is no difference between version 1 & version 3
  H1 : There is a difference between version 1 & version 3
- Test 3 : The improved visualization (version 2) compared to the report (version 3)
  H0 : There is no difference between version 2 & version 3
  H1 : There is a difference between version 2 & version 3

This results in three tests, so rather than use .05 as our critical level of significance, we use .05/3 = .0167. Table 4.11 presents the results of the analysis per question that were significant.

Table 4.11 The results of the analysis in the accuracy

Page 1	**Results**
Q5 What is your estimation of the average traveled distance from cities that had 5.8-10 % of inhabitants in Heerlen city center on the event day?	**Number of incorrect answer (accuracy) was significantly affected by the version of visualizations,** $H(2) = 7.69$, $p < .05$. Mann–Whitney tests were used to follow up this finding. A Bonferroni correction was applied and so all effects are reported at a .0167 level of significance. From the one tailed test, it appeared that accuracy were no different between version 1 and 3 ($U = 287$, $r = -.04$). However, **there was a significant difference between version 1 and 2 ($U = 247$, $r = -.30$) or version 2 and 3 ($U = 180$, $r = -.32$). The biggest effect size is between version 2 and 3. It means that the table is more accurate than the visualization (the table is more accurate than the map).**
Page 2	
Q4 What percentage of the people who came from Nijmegen used public transportation (train) to Heerlen city center on the event day?	**Number of incorrect answer (accuracy) was significantly affected by the version of visualizations,** $H(2) = 19.02$, $p < .05$. Mann–Whitney tests were used to follow up this finding. A Bonferroni correction was applied and so all effects are reported at a .0167 level of significance. It appeared that accuracy were no different between version 2 and 3 ($U = 225$, $r = -.08$). However, **there was a significant difference between version 1 and 2 ($U = 165$, $r = -.48$) or version 1 and 3 ($U = 130$, $r = -.54$). The biggest effect size is between version 1 and 3. It means that the table is more accurate than the visualization (the table is more accurate than the map).**
Page 3	
Q5 Which answer describes the relation between the average traveled distances and average staying time on the regular day and on the event day, the best?	**Number of incorrect answer (accuracy) was significantly affected by the version of visualizations,** $H(2) = 8.96$, $p < .05$. Mann–Whitney tests were used to follow up this finding. A Bonferroni correction was applied and so all effects are reported at a .0167 level of significance. It appeared that accuracy were no different between version 1 and 2 ($U = 311$, $r = -$

	.03). However, **there was a significant difference between version 1 and 3 (*U* = 175, *r* = -.40) or version 2 and 3 (*U* = 152, *r* = -.37). It means that the visualization is more accurate than the table (the map is more accurate than the table).**

4.5.2.5.3 Analysis based on the type of chart

Now, we present the analysis of the efficiency and the accuracy based on the type of chart Table 4.12 shows the questions per type of chart. In some questions both bar charts and maps were used to determine an answer. These questions were not analyzed.

Table 4.12 The questions based on the type of chart

Bar chart	Map
Page 1 – Q3	Page 1 – Q2
Page 3 – Q2	Page 1 – Q4
Page 3 – Q3	Page 1 – Q5
	Page 2 – Q2
	Page 2 – Q3
	Page 2 – Q4
	Page 3 – Q4
	Page 3 – Q5

Efficiency of bar charts and maps compared with data tables

In the efficiency between bar charts and data tables, we found a significant difference between bar chart and table in page 3 – Q3 (*When was the busiest time in Heerlen city center on the regular day and on the event day?*). Table 4.9 presents the significant result of the analysis in the efficiency between bar charts and data tables.

In the efficiency between maps and data tables, we found three results showed a significant difference between map and table (page 1 – Q2, page 1 – Q4, and page 2 – Q2). Table 4.9 presents the significant results of the analysis in the efficiency between maps and data tables.

Accuracy of bar charts and maps compared with data tables

There was no significant result of the analysis in the accuracy of a bar chart compared to data tables. In the accuracy between maps and data tables, we found three results showed a significant difference between map and table (page 1 – Q5, page 2 – Q4, and page 3 – Q5). Table 4.11 presents the significant results of the analysis in the accuracy between maps and data tables.

After answering the questions using the visualizations, the personal preference of the participants between the visualizations and reports was evaluated (fig. 4.17). 68% of the participants said they preferred the visualizations. Their comments indicate, it is easier and faster for them to get the information by using a visualization.

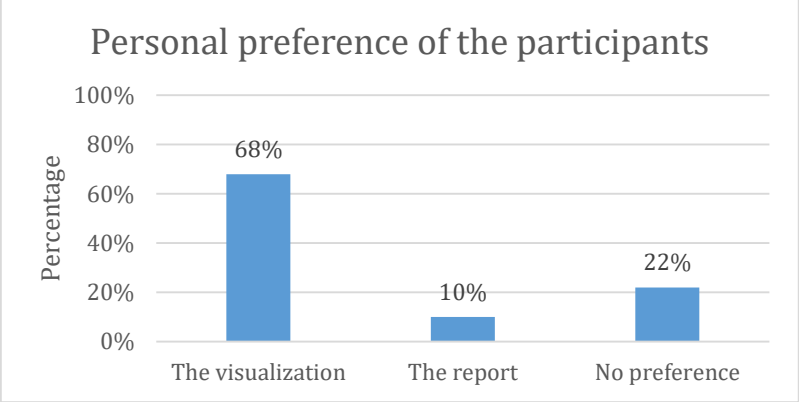


Figure 4.17 Personal preference of the participants (quantitative)

5. Discussions

In this chapter, all findings from the qualitative and the quantitative measurements in the experiments are discussed. First, the interpretation process of each visualization is presented. This is done using the results from the think aloud method. Second, the accuracy (number of interpretation errors) and the efficiency (time) of each visualization is discussed. This is done using the results from the survey in each participant group.

5.1 Discussions qualitative study

When we looked at the results of the qualitative study, it shows that information visualization is important to people who work at municipalities in the Netherlands. It helps them to understand the information and to get the information quickly. Currently, they are still using a static (on paper) information visualizations and mostly used them in reports or presentations. However, they do not rule out the possibilities to have dynamic information visualization in the future.

Of the tasks that the participants usually conduct with an information visualization, three were most frequently mentioned by the participants: 'comparison', 'evaluation', and 'exploration'. Similar findings came from a study by Andrienko et al. (2003). They created a task model that is defined by a set of questions that viewers might look for to answer with the help of visualizations. They divided the visualization tasks into 'comparison', 'relation seeking', and 'lookup' (details refer to section 3.5). Comparison tasks are tasks in which individuals use an information visualization to compare different data. The exploration tasks mentioned by the participants are similar to the relation seeking tasks, because in both tasks the users look for patterns and explore trends in the information. However, the evaluation tasks are not the same as the lookup tasks. Evaluation tasks are used to evaluate the goals or targets that users have, while in lookup tasks they search for data values.

From the think aloud evaluation, it became clear that the left side of an information visualization is the first part that is seen by the participants. This indicates that the most important information should be placed on the left side of an information visualization. This is in line with a case study on web pages by Nielson (2006) in which an F-shaped

pattern of reading in identified. The study shows that people's reading order of web pages look like an F, where the bottom right area of the pages get the least attention (fig. 5.1). From the experiment, we found 2 participants did not start from the left side of the visualization. This indicates that some people have different strategies in reading a visualization.

Figure 5.1 Jakob Nielsen F-shaped pattern

Next, comments that stand out are about a number of texts. The participants preferred to have fewer texts on the visualization and the texts should be located close to the visualization. For example, the participants mentioned that the title on the maps could be shorter. The information that is shown in a legend on a map, should be clear whether it is an absolute or a categorical value. Lastly, the evaluation of the aesthetic quality of the visualizations varied between 7.4 and 7.8. It means that the visualizations are considered quite attractive. Moore and Purchase (2011) state that attractiveness allow more effective communication of information.

5.2 Discussions quantitative study

Efficiency

When we summarizes the results, 5 out 16 questions show a significant difference regarding efficiency between a visualization and a report (only tables and numbers). This

indicates that a visualization (bar charts, maps or when both are used) is more efficient or faster to interpret than a report.

We looked at the visualization design guidelines applied in the visualization, the type of charts, and the type of tasks. First, in visualization page 1 – Q1 we find that 'highlighting intervention' helps the participants to get the information quickly. So, on the visualization page 1 we used the de-emphasis highlighting intervention on the total number of extra visitors that visit the city center. This is in line with a study from Carenni et al. (2014) which shows that the highlighting intervention can improve visualization processing.

In one instance, we found that a bar chart is more efficient than a table (*When was the busiest time in Heerlen city center on the regular day and on the event day?*). In the question, the participants are asked to compare between the regular day and the event day (comparison task) and to determine the busiest time (lookup task). From the bar chart, the participants can easily see the busiest time from the size of the bar chart. The highest bar chart is the busiest time in the city center.

Lastly, 3 questions showed that a map is more efficient than a table (see table 4.14 for the details). In the questions, the participants are asked to find patterns (relation-seeking task) and to find certain values (lookup task) from certain geographical locations. From the map, the participants not only can see the patterns but they can also see the geographical locations or areas of the information. Although this conclusion is based on only 3 questions, it indicates that when users want to see and find a pattern, a map can be used to find the information faster than a table.

Accuracy

When we summarize the results, 3 out of 16 questions show a significant difference regarding accuracy between a visualization and a report (only tables and numbers). First, we find that when the information on a legend especially on a map is not clear, then the visualization is interpreted incorrectly. This is in line with the results of the qualitative study where the participants had a difficulty with the legend.

Next, we find 2 questions show a table is more accurate than a map. In these questions (see table 4.15 for the details), the participants are asked to find certain values (lookup task) in the information. It shows that when the participants want to look for a specific value, they can get it more accurately using a table. However, we only conclude it from two questions.

The results of 1 question shows a map is more accurate than a table (*Which answer describes the relation between the average traveled distances and average staying time on the regular day and on the event day, the best?).* In the question, the participants are asked to find patterns (relation-seeking task). Although it is only 1 question, we could say a map can be used to find the patterns more accurately using a map than a table. We did not find any other significant differences to support the advantage of visualizations or tables for specific tasks.

6. Conclusions and limitations

Due to the rise of the use of information visualizations, ensuring the effectiveness of a visualization in two principles, accuracy and efficiency, becomes increasingly important (Few, 2013; Zhu, 2007) This study focused on the most effective visualization methods to create visualizations using mobility information that can easily communicate the information and can be interpreted correctly. The main research question of this research is "*Which visualization methods are most effective to present mobility information, in relation to specific user groups and their information needs? And how can the design process be structured to achieve an optimal design?* ". Qualitative and quantitative approaches were used in order to answer the main research question. In section 6.1, the conclusion of this research will be presented. Then, the limitations and further research are discussed in section 6.2.

6.1 Conclusions

This research tries to answer the research question and sub-questions described in section 1.2 through literature studies, interviews, document studies, qualitative studies and quantitative studies. To answer the main research question, five sub-questions contributed as a building block of this research. Each sub-question will be answered in order to answer the main research question.

SQ1: What is an effective information visualization?

From the literature studies, it was found that there is no one strict definition of the effective information visualization. Different research has different views on how to define an effective information visualization. However, there are the same principles that have been summarized by Zhu (2007) which are accuracy and efficiency. Accuracy is about how accurate an information visualization can be interpreted and efficiency is how easy and fast an information visualization can be understood. So, an effective information visualization is a visualization of information where people can easily read and correctly understand it.

SQ2: What are the existing visualization methods?

Firstly, literature studies were conducted to answer this question followed by document studies and interviews with the partners of Mezuro. Based on the literature reviews, there are many different visualization techniques depending on the type of data and the type of task. However, in this research we looked at visualization techniques that can be applied using mobility data which consist of spatial and time data and can be used to support comparison tasks, relation-seeking tasks, and lookup tasks that were mentioned by Andrienko et al. (2003). Document studies and interviews followed this step to identify the type of visualizations that they mostly used. From the literature studies, it was found that a bar chart can be used for comparison tasks and lookup tasks (Few, 2004; Data visualization 101: How to design graphs and charts, 2014). Then, a map can be used for relation seeking tasks and also can be used to show the geographical location or area of the information (Mazza, 2009). From the document studies and the interviews conducted, it is discovered that mostly they use a bar chart, a map and a table to present mobility information to their customers.

SQ3: What is mobility information?

In order to answer this sub-question, literature reviews and interviews with Mezuro's partners were performed. From the literature reviews, in general mobility information is the information on the movement of people in a population, as from place to place. Furthermore, the interviews gave us a more clear explanation of mobility information that Mezuro has (which is collected from large numbers of mobile phones data and is taken within a specific period of time). So, we can define the mobility information in Mezuro as information on the movement of people from one place to another in a certain time together with their frequency of visits and motives.

SQ4: Which design patterns can be identified for specifying the most effective ways to present mobility information?

To answer this sub-question, a literature study, a qualitative study and a quantitative study were performed. From the studies, we made guidelines that can be used to provide the most effective ways to present mobility information. First, we show the recommendation for showing mobility information based on the type of tasks and the

type of charts (table 6.1). From the recommendation column, we can see that it may be useful to present the information both as visualizations and tables.

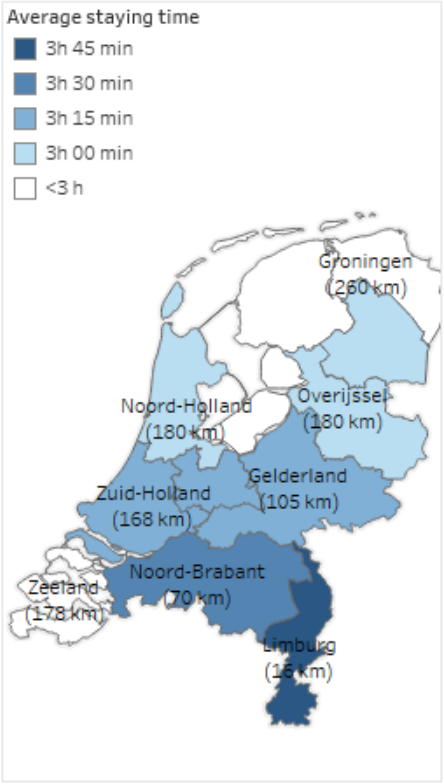
For a bar chart, our findings in the quantitative study support our findings in the literature study where a bar chart can be used to support a quick comprehension of values and show comparisons of values. For a map, our findings in the quantitative study also support our findings in the literature study where a map can be used to show patterns and to help people find certain values from certain geographical locations. For a table, our findings in the quantitative study add detail in our research where a table can be used to help people look for a specific value.

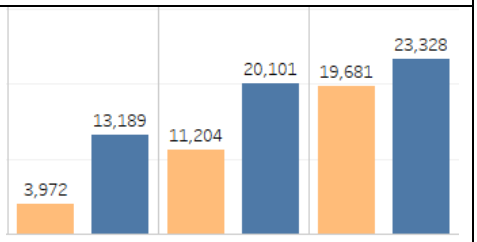
Table 6.1 The recommendation for showing a mobility information

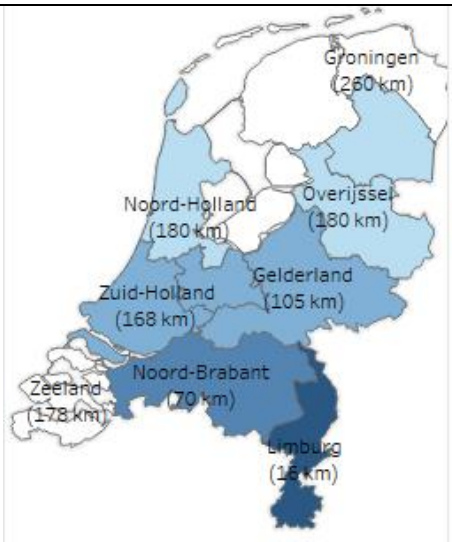
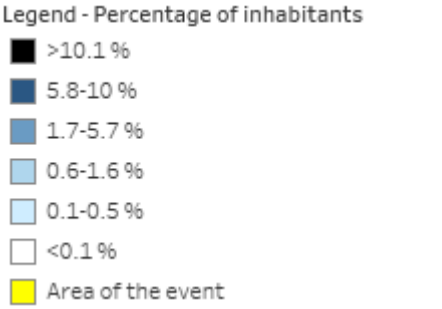
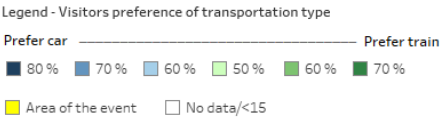
Type of tasks	**Efficiency**	**Accuracy**	**Recommendations**	**Reasons**
Comparison	Bar chart	-	Bar chart and table	Information can be interpreted quickly by a bar chart and interpreted accurately by a table.
Relation seeking	Map	Map	Map	Information can be interpreted quickly and accurately by a map.
Lookup	Bar chart and map	Table	Bar chart and table Or Map and table	Information can be interpreted quickly by a bar chart or a map and interpreted accurately by a table.

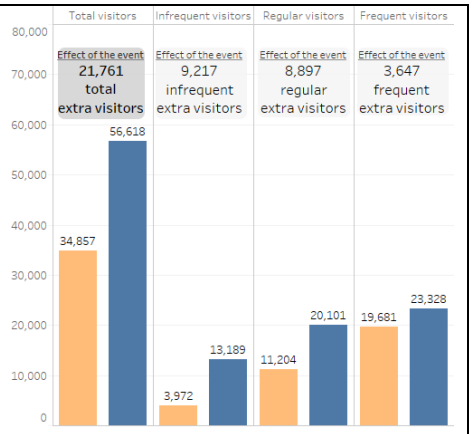
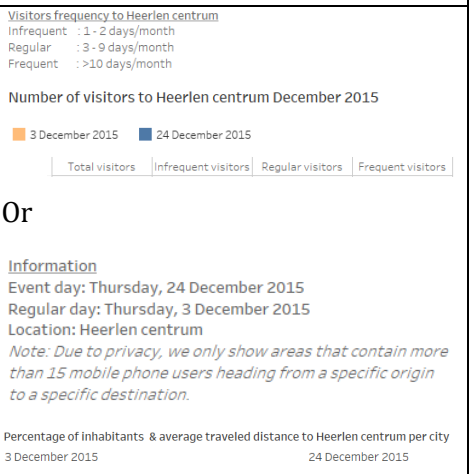
Second, in table 6.2 we present visualization design elements that need to be considered when making a mobility information visualization.

Table 6.2 Visualization design elements

Design elements	Explanations	Examples (screenshots)
1. Color	This was not explicitly tested in the experiment. However, it needs to be considered and from the literature studies we recommend: • Because some people have color vision deficiency, try to avoid using a combination of red and green in the same visualization. Online tools are available to check color schemes for this problem. • Only use different colors when they correspond to differences of meaning in the data. • Use soft, natural colors to display most information and bright or dark colors to highlight information that requires greater attention. • When using color to encode a sequential range	

	of quantitative values, use a single hue (or a small set of closely related hues) and differentiate the intensity from pale colors for low values to increasingly darker and brighter colors for high values.	
2. Highlight	Our findings in the quantitative study support our findings in the literature study where the highlighting intervention can be used to help people to get an information quickly. So, our recommendation is the important information should be highlighted using de-emphasis highlighting intervention.	
3. Label each bar or area on the map	Although this was not explicitly tested in the experiment, from the literature studies we recommend that labels should be used to help people identify certain values, otherwise it will make the visualization hard to read.	 Or

		
4. Legend	Our findings in the qualitative and quantitative study add detail to our findings in the literature study where the information that is shown on a legend should be presented as clearly as possible. Our recommendation is the information that is shown on a legend should be clear whether it is an absolute or a categorical value.	 or 

5. Structure/ layout	Our findings in the qualitative study support our findings in the literature study where the left side of an information visualization is the first part that is seen by the people. So, our recommendation is the most important information should be placed on the left side of an information visualization and the other information should be distributed based on their order of importance.	
6. Explanation of information	This was not explicitly tested in the quantitative study. However, our findings in the qualitative study add detail to our findings in the literature study where the people need the explanation of information but they preferred to have fewer texts on an information visualization. So, our recommendation is to use a simple and precise explanation of information.	

SQ5: Which techniques are required to determine user requirements, select the visualization methods, and evaluate the effectiveness of the visualization?

From the literature study and our experience in the qualitative and the quantitative studies that we have done, we can answer this sub-question. First, in order to determine user requirements, an interview or a discussion with the customers should be conducted. This is essentials to understand what the customers really want from the mobility information and should be undertaken in the first stage of the project.

Next, before selecting a visualization method, we look at the type of question or task from the customers. For example, it is a comparison task, a relation seeking task or a lookup task. Then, we look at the visualizations that are commonly used in a certain task. For example, a bar chart is commonly used in a comparison task.

During the development, customers should be requested to give feedback on prototype visualizations to adjust the visualizations where required, within the project constraints (budget and planning). To test the effectiveness of a visualization during development, the think aloud method can be used. The advantages of this method are:

- We can observe the interpretation processes and whether the visualization interpreted correctly or not.
- We can determine which elements or parts of the visualization should be kept or improved.

However, this method also has some disadvantages:

- Not everyone is talkative or easy to voice their thoughts, feelings or opinions. So, we need extra efforts when we want to do this method with this kind of people.
- This method requires time.

After development, evaluation is very useful for understanding how the product was used in the context. For instance, asking user feedback after a period of time or asking user satisfaction. It can be done by having a discussion with the users or having feedbacks from a newsletter or LinkedIn. Mezuro will have to consider how these activities fit in their own project management principles.

MRQ: Which visualization methods are most effective to present mobility information, in relation to specific user groups and their information needs? And how can the design process be structured to achieve an optimal design?

Based on the findings from previous sections and the answer of each sub-question in this section, we can conclude that the most effective visualization methods to present mobility information depend on the type of tasks or type of questions (comparison tasks, relation seeking tasks or lookup tasks) that will be answered.

In comparison tasks, it could be considered to use a combination of a bar chart and a table. In relation seeking tasks, a map can be used to enhance the efficiency and the accuracy when the information contain geographical areas or locations. In lookup tasks, it could be considered to use either a combination of a bar chart and a table or a map and a table. So, it is useful to present both visualizations and tables. Furthermore, there are also visualization design elements that need to be considered like color, highlight, label, legend, structure/layout, and explanation of information.

Then, in order to achieve an optimal design, the design process should be structured as follows:

- First, an interview or a discussion with the customers should be conducted to determine user requirements.
- Second, look at the type of question or task from the customers and then select a visualization method that can be used in a certain task.
- Third, during the development, customers may be requested to give feedback on prototype visualizations, to adjust the information needs or expectations of the visualization. The think aloud method can be used to test the effectiveness of a visualization during this phase.
- Last, evaluation is very useful for understanding how the product was used in the context. For instance, asking user feedback after a period of time or asking user satisfaction.

6.2 Limitations and further research

There were some limitations to the conducted research where a different approach or further research could provide more insights. The first limitation of this research is the type of information. In this research we focus on the effective visualization of mobility information where we used mobile phone data. For a different type of information, different types of data and different questions could be used.

The second limitation is the user group. Mobility information can be used by many different groups (e.g. profit organization, non-profit organization, government, etc.). However, in this research we focus on the customers of Mezuro who work at municipalities in the Netherlands and use the mobility information. Different results will occur in different groups, for instance researchers.

The third limitation is the participants of the survey. In this research we put the survey not only on LinkedIn and Mezuro's newsletter, but we also sent the survey to the students at Utrecht University. 41.7% of the participants were the aged between 21-30 years old, and are probably students, who are not primary users of the visualizations. Last but not least, we focus on a static (on paper) mobility information visualization. Further research could be conducted on a dynamic mobility information visualization.

References

7 Basic Rules for Making Charts and Graphs. (2010). Retrieved from http://flowingdata.com/2010/07/22/7-basic-rules-for-making-charts-and-graphs/

A Guide to Creating Dashboards People Love to Use. (2009, November). Retrieved from http://www.cpoc.org/assets/Data/guide_to_dashboard_design1.pdf

Al-Kassab, J., Ouertani, Z. M., Schiuma, G., & Neely, A. (2014). Information visualization to support management decisions*. International Journal of Information Technology & Decision Making*, 13, 407-428.

Alexandre D. S., & Tavares, J. M. R. S. (2010). Introduction of human perception in visualization. *International Journal of Imaging and Robotics*, 4, 60–70.

Andrienko, N., Andrienko, G., and Gatalsky, P. (2003). Exploratory spatio-temporal visualization: an analytical review. Journal of Visual Languages & Computing, 14(6), Elsevier: Pages 503-541

Becker R, Cáceres R, Hanson K, Isaacman S, Loh JM, Martonosi M. (2013). Human mobility characterization from cellular network data. Communications of the ACM. 2013;56:74–82. doi: 10.1145/2398356.2398375.

Burkhard, R. (2004). Learning from Architects: The Difference between Knowledge Visualization and Information Visualization. In Proceedings of *Eighth International Conference on Information Visualization* (IV04), pp. 519-524. London: IEEE.

Burkhard, R., Spescha, G. & Meier, M. (2005). "A-ha!": How to Visualize Strategies with Complementary Visualizations". *Conference on Visualising and Presenting Indicator Systems*, pp.1–10.

Card, S. (2003). Information visualization. In A. Sears, & J. A. Jacko (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications*. New York, NY: Lawrence Erlbaum Assoc Inc.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: using vision to think*: Morgan Kaufmann.

Carenini G., Conati C., Hoque E., Steichen B., Toker D., Enns J.(2014).Highlighting interventions and user differenc-es: Informing adaptive information visualization support. *In Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, ACM.

Casner, S. M. (1991). A task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics*, 10, 111–151.

Chan W. W.-Y. (2006). A survey on multivariate data visualization. *Department of Computer Science and Engineering. Hong Kong University of Science and Technology 8*, 6, 1-29.

Darke, P., Shanks, G., & Broadbent, M. (1998). Successfully completing case study research: combining rigour, relevance and pragmatism. Information Systems Journal, 8 (4), 273-289.

Data visualization 101: How to design graphs and charts. (2014, May 29). Retrieved from https://visage.co/data-visualization-101-design-charts-graphs/

Data visualization – what it is and why it matters. (2016, April 29). Retrived from http://www.sas.com/en_us/insights/big-data/data-visualization.html

Denney, A. S., & Tewksbury, R. (2012). How to write a literature review. *Journal of Criminal Justice Education*. Advance online publication. doi:10.1080/10511253.2012.730617

Few,S. (2004). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Oakland, CA: Analytics Press.

Few, S. (2006). *Information dashboard design: The effective visual communication of data*. Sebastopol, CA: O'Reilly Media, Inc.

Few, S. (2007). Data Visualization: Past, Present, and Future. IBM Cognos Innovation Center.

Few, S. (2008). Practical rules for using color in charts. Visual Business Intelligence Newsletter. Retrieved from http://www.perceptualedge.com/articles/visual_business_intelligence/rules_for_using_color.pdf

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

Field, A. (2009). *Discovering statistics using SPSS*. Thousand Oaks, CA: Sage publications.

Four Cognitive Design Guidelines for Effective Information Dashboards. (2014, June 18). Retrieved from https://uxmag.com/articles/four-cognitive-design-guidelines-for-effective-information-dashboards

Graves, A., and Hendler, J. (2013), "Visualization tools for open government data". In *Proceedings of the 14th Annual International Conference on Digital Government Research* (136-145).

Hancock, B., Ockleford, E., & Windridge, K. (1998). *An introduction to qualitative research*. Nottingham: Trent focus group.

Jansen, S., & Brinkkemper, S. (2008). Applied multi-case research in a mixed-method research project: Customer configuration updating improvement. *Information Systems Research Methods, Epistemology and Applications*, 1-29.

Keim, D., Mansmann, F., Schneidewind, J., Ziegler, H. (2006). Challenges in visual data analysis. In *Proceedings of Information Visualization. IEEE Computer Society*, London, UK, pp. 9-16

Khan, M., & Khan, S. S. (2011). Data and information visualization methods, and interactive mechanisms: A survey. *International Journal of Computer Applications*, 34(1), 1-14.

Krug, S. (2011), *Rocket Surgery Made Easy: The Do-it-yourself Guide to Finding and Fixing Usability Problems*, New Riders, Berkeley, CA.

Latham, R. (1995). *The dictionary of computer graphics and virtual reality*. New York: Springer-Verlag.

Longhurst, R. (2010). Semi-structured interviews and focus groups. In N. Clifford, S. French, & G. Valentine (Eds.), *Key methods in geography* (2nd ed.) (pp. 103-115). London, EN: SAGE Publications Limited.

Lengler, R. & Eppler, M. J. (2007). Towards a Periodic Table of Visualization Methods of Management. In IASTED *Proceedings of the Conference on Graphics and Visualization in Engineering*, Florida, USA, pp. 1-6

Mack, N., Woodsong, C. M., MacQueen, K.M., Guest, G., & Namey, E. (2005). *Qualitative research methods: A data collector's field guide*. Research Triangle Park, NC: Family Health International.

Marr, B. (2016, May 18). How to Turn Big Data into Insights and Action. Retrieved from http://data-informed.com/7-key-ingredients-for-knock-out-data-visualizations/

Mazza, Riccardo. (2009). *Introduction to Information Visualization*. Springer: London.

Mobility. (n.d.). *Dictionary.com Unabridged*. Retrieved June 28, 2016 from Dictionary.com website http://www.dictionary.com/browse/mobility

Nielson, J. (1994). Usability Inspection Methods. New York: Wiley. ISBN: 0-471-01877-5

Nielson, J. (2006, April 17). F-Shaped Pattern for Reading Web Content. Retrieved from https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/

Noor, K.B.M., (2008). Case study: A strategic research methodology. American Journal of Applied Sciences, 5 (11), 1602–1604.

North, C. (2005). Information visualization. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 1222-1246). New York: John Wiley.

Odenhoven, S., & Worring, M. (2014). Modelling the alignment gap between academic and practical information visualization design approaches.

Palmer, J. R. B., Espenshade, T. J., Bartumeus, F., Chung, C. Y., Ozgencil, N. E., & Li, K. (2013). New approaches to human mobility: Using mobile phones for demographic research. *Demography, 50,* 1105–1128.

Pu, J., Xu, P., Qu, H., Cui, W., Liu, S., & Ni, L. (2011). Visual analysis of people's mobility pattern from mobile phone data. *In Proceedings of the 2011 Visual Information Communication – International Symposium*, VINCI

Rhyne, T.M. (2003). Visualization Viewpoints: Does the difference between information and scientific visualization really matter? *IEEE Comput Graphics and Applications,* 23(3),6–8.

Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (Eds.). (2013). *Qualitative research practice: A guide for social science students and researchers*. Sage.

Sharp, H., Rogers, Y., & Preece, J. (2007). *Interaction design: Beyond human-computer interaction*. Hoboken, N.J.: Wiley.

Shneiderman, Ben. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *IEEE Xplore. IEEE Computer Society Washington*.

Skoglund, M., & Runeson, P. (2009). Reference-based search strategies in systematic reviews. *Proceedings of the 13th international conference on Evaluation and Assessment in Software Engineering,* Swinton, UK, 31-40.

Spence, R. (2001). "Information Visualization". New York: Addison Wesley.

Stodder, D. (2013). Data visualization and discovery for better business decisions. TDWI Best Practices Report. Renton, WA: The Data Warehouse Institute.

Stone, M. (2009). Information visualization: Challenge for the humanities. In *Working together or apart: Promoting the next generation of digital scholarship* (43-56). Washington, DC: Council on Library and Information Resources.

Tegarden, D.P. (1999) "Business Information Visualization" Communications of the Association for Information Systems, (1)4.

Telea, A. (2014). *Data visualization: principles and practice.* Boca Raton: CRC Press.

Ten Things to Consider When Making a Map. (2011, September 10). Retrieved from https://www.gislounge.com/ten-things-to-consider-when-making-a-map/

Tufte, Edward R. (1997). Visual Explanations: Images and Quantities, Evidence and Narrative. CT: Graphics, 1, Cheshire.

van Someren, M. W., Barnard, Y. F. and Sandberth, J. A. C. (1994). The think aloud method: a practical guide to modelling cognitive processes. Academic Press, London.

van de Weerd, I., & Brinkkemper, S. (2009). Meta-modeling for situational analysis and design methods. In M. Syed & S. Syed (Eds.), *Handbook of research on modern systems analysis and design technologies and applications* (pp. 35-38). Hershey: Idea Group Publishing.

Ward, M. O., Grinstein, G., and Keim, D. A. (2010). *Interactive Data Visualization: Foundations, Techniques, and Application*. A. K. Peters, Ltd., Natick, 4, 129-170.

Wattenberg, M., and D. Fisher. 2004. "Analyzing Perceptual Organization in Information Graphics." *Information Visualization*, Vol. 3, p. 123–133.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), xiii-xxiii.

Wu, W., Xu, J., Zeng, H., Zheng, Y., Qu, H., Ni, B., Yuan, M., and Ni, L. M. (2016). Telcovis: Visual exploration of co-occurrence in urban human mobility based on telco data. *Visualization and Computer Graphics, IEEE Transactions on*, 22(1):935.944.

Wunsche, B. (2004). A survey, classification and analysis of perceptual concepts and their
application for the effective visualization of complex information, *Proceedings of the 2004 Australasian symposium on Information Visualisation, Darlinghurst, Australia, Australian Computer Society*, Inc 2004:17-24.

Your Graphs Look Like Crap: 9 Ways to Simplify and Sexify Data. (2013, October 7). Retrieved from https://blog.hubspot.com/marketing/data-graph-design-powerpoint-tips-ht#sm.0000z66n4c7u0f3dw5g1gelwxkg30

Yin, R. K. (2013). *Case study research: Design and methods*. Sage publications.

Zhu, Y. (2007). Measuring effective data visualization. *Proceedings of the 3rd International Symposium on Advances in Visual Computing*. 652–661. New York: Springer.

Zuk, T., Schlesier, L., Neumann, P., Hancock, M. S., & Carpendale, S. (2006). Heuristics for information visualization evaluation. *Bibliometrics*. doi:10.1145/1168149.1168162

Appendix A – Mezuro partner interviews

I. Impact

Master thesis – information visualization

Introduction

Dear interviewee, my name is Zaki, a master student from Utrecht University. At this moment I am doing my thesis and internship in Mezuro. The aim to investigate which type of visualization of information, especially in mobility information, is most effective in a specific use and for specific user groups.

All information gathered through this interview will be valuable to my research. Moreover, gathered information will be recorded and treated with respect and will only be used for scientific purposes. If you do not want the name of your company in the thesis document, we can replace the name with a short description. In addition, you are allowed to stop the interview at any time, if you want, since participating in the interview is completely voluntary.

The interview takes approx. 40 minutes and the questions focus on information visualization. If we notice that there is a chance of running out of time, we might interrupt you and move on to the next question.

Topics for interview:

- Discovering guidelines for effective mobility information visualization.
- The important procedures in designing and implementing effective mobility information visualization.

General information

Date	: 4 November 2016
Time	: 10.00
Position/job title	: Consultant at impact management
Since	: 2005

Questions

1. Can you please tell me about your position? (Main tasks, responsibilities, etc.)

   In my position I build data mods (Metadata Object Description Schema), data warehouse solutions and visualizations of KPI dashboard for government.

2. In your current job or position, do you use (as a user) information visualizations? How often do you use it? (how often: once a month, a week, a day, more than a day)

   In a week, 70% of my time I'm building data mods and for 30% I'm building visualizations (e.g. Information visualizations, reports, etc.).

   a. What types; what kind of data, for what kind of questions.

   Usually I visualize financial data that is used by the government. For instance, I visualize information about comparison of budget between estimation and real numbers that are used by the government in the past 12 months.

   b. And why? How does you use it (in making decisions, search information?)

   The information visualization is used for making decisions. So in the government there are 3 layers:

   - *Top: They need information to make decisions and directions. For example, they can see the comparison of budget between the estimation numbers and real numbers (e.g. what did I do this year?). So they know their position whether is it good or bad.*
   - *Middle: They want to see details (e.g. numbers or figures). The timeframes may be wider than the top layer. For instance, they can see the comparison of budget in their department between the estimation numbers and real numbers in each month or week.*
   - *Low: They only want to see lists (e.g. what should I do today, what is my workload, etc.).*

   Different layers will have different visualizations.

3. Do you create information visualizations? Which software do you use (standard software, Excel etc)?

   I'm using a tool in creating the visualizations. I only use Cognos IBM software.

<u>Information visualization</u>

4. In your opinion, what are the advantage of information visualization (in comparison to text, numbers, etc.)?

   In comparison to text, it is much easier to see in one view where my problems are. So where do I have to react. That is the main idea of using visualizations. For instance you can see the sickness report where is and in which department that the

sickness is too high compare to the estimation and by seeing the details we can see in which department the problem lies.

So the top layer does not interest in seeing a bunch of lists, but they are more interest in seeing the overview of the problems and solutions. For the middle and low layer, they want to see more on the details of the data or numbers.

5. Based on your experience, what is an effective information visualization?

   It is a simple visualization of information where you easily read and understand it.

6. What criteria or components are required for /in an effective information visualization?

   - *The first place is facts, there can be only one truth (figure/number). So you need a tool that everybody in the company uses it, which means that everybody in all layer can see the facts.*
   - *Keep the visualization simple, do not use too many visualizations in one screen. Because then you cannot read it, it is too much information.*
   - *Put some instructions in the visualizations so that the readers or audiences know what to do with the visualizations.*

7. In your opinion, what is mobility information?

   It is a future and new way to see where people are and what are they doing. In this case we are not talking about legal and privacy.

<u>Customers (Event)</u>

8. What information, in this case using event data, do customers want to know?

   They want to see the effectiveness of the event. So for example:

   - *Do I get more people?*
   - *From where/ from which region do the people come from?*

   Usually by using information from Mezuro, you can use or see not only 1 city but more than 1 cities. What will be very interesting in the future is how do I get people from those cities to my city? Is that possible?

9. What kind of questions do customers ask from event data, like mezuro has? Can you please give examples?

   - *Do I get more people? (event day compare to regular day)*
   - *From where/ from which region do the people come from?*
   - *Do the people come from the same area on a normal day?*

10. From your experience, what are good techniques or guidelines on how to present or visualize the information to them?

    So if you want to compare the event day and regular day, you should compare it with the same day. For instance, if the event is on Friday then you must compare it with Friday on regular days. Do not compare it with other days.

11. Do your clients always know exactly which information they need, or do you have to assist them in finding out the most relevant information for their problem? (E.g. availability of data, data processing, integration, filtering, validation).

    In my area, I need to assist them. Because most of the time, the customer (government) always ask what kind of visualization do I need for their problems.

12. How would you determine the best type of information visualization for the customer?

    So, usually I:

    - *Use our best practices in order to visualize the information.*
    - *Use the suggestions of the information visualization type from the visualization tool (Cognos IBM).*

<u>Visualization examples</u>

13. From the visualization report on event (Glazen huis) that I got from my supervisor in Mezuro, I would like to know:

    a. What are the reasons behind the visualizations? (E.g. the type, the color, etc.)

    For the visualization type, actually it depends on what information are you going to show. For instance, do I want to present the difference between estimation and reality or do I want to present the difference number in the last 12 months.

    The reason why we put a lot of texts is because the theory about Mezuro data is totally unknown by the viewers, so we have to explain. And the important thing about the colors are they should be constant in your whole report.

    b. I see in some of the visualizations, you mention the numbers. Can you please explain the reasons behind it?

Visualizations only present the high level like whether the estimation is already good or bad. By using numbers or figures, we can see more details on the problems. We can determine where the problem lies are.

c. How useful is a map in visualizing the information?

Map can be used to point something and determine where the people are come from. You need other visualization that can show the numbers.

II. Marble Research

Master thesis – information visualization

Introduction

Same as in the first Mezuro's partner interview.

General information

Date	: 17 November 2016
Time	: 10.00
Position/job title	: Senior researcher in Marble Research
Since	: 2009

Questions

1. Can you please tell me about your position? (Main tasks, responsibilities, etc.)

   Right now I'm a researcher and my main business is market research. I do a lot of researches in public transport and especially I'm interested in opinions and attitudes from travelers. For instance I do a lot of research project for Dutch Railways and Prorail. I also do data analysis, apply statistical method and visualize the data.

   Previously I made a long report of our research to the customers, but nowadays they want short report. So I change my way of reporting to my customers, I start to use visualizations because it is easy for them to interpret the data.

2. In your current job or position, do you use (as a user) information visualizations? And why?

   I use visualizations when I analyze data, because I notice when I visualize the data then sometimes it is easier and faster to understand what is actually happening. The visualizations is used for making decisions and searching information.

3. Do you create information visualizations? How often? (how often: once a month, a week, a day, more than a day)

   I create information visualizations every time I finished analyzing the data.

   a. For whom? Which software (standard software, Excel etc)?

   I made the visualization for people who work in municipalities and event organizers. I use Tableau software as my tool.

Information visualization

4. In your opinion, what are the advantage of information visualization (in comparison to text, numbers, etc.)?

   Well, it is easier to interpret the results and it takes less time because customers do not have to read the whole report. Most of the customers do not really interested in details but maybe just a few things or indicators that are important for them.

5. Based on your experience, what is an effective information visualization?

   It depends on the target group you have, because some they only want see the conclusion on one sheet. For instance, they want to see the results on one sheet and they want to know what to do. And for the other, they also want that but they also want to be able to have a look at the data on a deeper level.

6. What criteria or components are required for /in an effective information visualization?

   You have to think about the type of data/type of variables you have, for instance whether you have nominal, ordinal or categorical data. Then you need to combine it with the type of graphics which are effective in bringing all the messages. So for instance like pie chart, when you have a pie chart and ten categories with 3-4 very small categories you cannot visually distinguished them in a pie chart. So it is better for you to use a bar chart.

   So, what you have to think about is:
   - *What kind of variables do we have?*
   - *Which type of graphs that fit with the data?*

   The second thing is that you should use colors to support the message you want to tell. For instance, when you have a good result you should use a green color compare to a bad result which you could present it with a red color. You should take care that the colors really support the message you want to tell.

The last thing is keep the visualizations simple so when they see it, they can go directly to the conclusion.

7. What should be avoided in information visualization designs (for instance: many colours, too much data, too much text, etc.)

   The principal is "less is more", so try not to show too much. Because when you show too many figures in one page, then it becomes less easy to interpret it.

8. In your opinion, what is mobility information?

   In principal it is a trip with destination and origin. That is just the starting point and from there you can do a lot of kind of analysis. You have travel time, you can compare how important certain destination or certain origin are, you can have a look at what kind of activity people do at a destination, etc.

   There is an approach called "the activity based approach of mobility" which is very good especially from sociological point of view because people make a trip because they want to do some activities at a certain place and the kind of activity also influences the amount of time that people spend and the type of transportation that they use. So, it is all about the behavior of people. How do they behave depend on the activity they do at the destination

Customers (Event)

9. What information, in this case using event data, do customers want to know?

   Mostly they want to know:
   - *How many people go to the event?*
   - *Where do the people come from?*
   - *How long do they stay?*
   - *When is the peak hour of the event?*
   - *What type of transportation do they use?*

10. What kind of questions do customers ask from event data, like mezuro has? Can you please give examples?

    The same as above

11. From your experience, what are good techniques or guidelines on how to present or visualize the information to them?

    People orient themselves in time and space, and when you show a data on a map people will understand it. For instance, when you talk about the origin where do most people come from you can show the cities on a map using the dark and light colors.

12. What should be avoided?

Do not show too many information and too many visualizations on one page.

13. Do your clients always know exactly which information they need, or do you have to assist them in finding out the most relevant information for their problem? (E.g. availability of data, data processing, integration, filtering, validation).

    Sometimes they do not know the possibilities of what can they do with the data or what they need. So my task is not only to tell them what is in the data but also what they can do with it.

14. How would you determine the best type of information visualization for the customer?

    Usually I use the suggestions from Tableau software, but I also use common visualization types like bar chart, line chart, etc. It is because people already know about it, so it is easy for them to interpret.

15. What kind of complaints did you usually get from the customers regarding the information visualization that you have created or showed to them? (e.g. they do not like the type of the visualization, they do not like the colors, etc.)

    In my experience, sometimes they need a little bit of information about the visualization.

III. Decisio

Master thesis – information visualization

Introduction

Same as in the first Mezuro's partner interview.

General information

Date	: 9 Dec 2016
Time	: 09.00
Position/job title	: Senior consultant at Decisio
Since	: 2008

Questions

1. Can you please tell me about your position? (Main tasks, responsibilities, etc.)

I am a senior consultant at Decisio. I responsible for many projects especially in economic analysis like social economic analysis. I also do analysis of government investment:

- *I try to calculate what brings to society as a whole and which projects or other kind of policies that the government should put their money in and which not.*
- *Monitoring and evaluation of events using data from Mezuro so I can determine the effect of an event for a city like the number of visitors, etc.*

2. In your current job or position, do you use (as a user) information visualizations? How often do you use it? (how often: once a month, a week, a day, more than a day)

   a. What types; what kind of data, for what kind of questions.

   In my current position, I use information visualization almost every day. I use mostly graphs like bar charts, line charts, and map. For the data, most of the time I use economic data.

   b. And why? How does you use it (in making decisions, search information?)

   For me, I rarely use the information visualization for myself. Most of the time I work with plain data like numbers and I analyze the numbers. But when I want to show the results of my analysis, I make information visualization. I make information visualizations to explain the data that I have in a better way.

3. Do you create information visualizations? How often? (how often: once a month, a week, a day, more than a day)

   a. What types; what kind of data, for what kind of questions.

   I create information visualization almost every day and I create it mostly using graphs like bar charts, line charts, and maps, but sometimes my colleagues also use infographics. I use information visualization to make report for people who work in municipalities. So I do economic analysis and to explain the results of my analysis I create information visualizations.

   My customers are not directly use the information visualization to make a decision but they use it as insights before they can make a decision in the future. One of the decision that they usually make is investment decision for a certain project or policy.

   b. For whom? Which software (standard software, Excel etc)?

My customers are people who work in municipalities and I use Ms. Excel and Qgis for my softwares.

Information visualization

4. In your opinion, what are the advantage of information visualization (in comparison to text, numbers, etc.)?

   Some people are really bad at reading numbers and they just don't see it or they just can't remember it and restore it in their brain and use it in their decision making. So if you visualize it, it is easier for them to understand. They can directly see the differences between the numbers because the visualizations can give them insights in scales. When you have a visualization, it is nice to have a quick view on trends and development of several figures or numbers at first glance.

   I also make visualizations in geographical way using maps to give insights about where are people coming from, where they are going to, etc. In my opinion, map is the best way to show geographical information. Because then you can directly the position and you can also put numbers with colors, shapes, etc.

   If you have geographical information and especially when it is from several areas that are close to each other then it is really good to have that visualize. For instance, I want to show a map of Amsterdam area and hundreds of areas where people are coming from, then I think I have to show it on a map. Because I can show the connection between number of people and geographical areas. I can show whether the people are coming from nearby areas or not.

5. Based on your experience, what is an effective information visualization?

   It is hard for me to define an effective visualization because I think it depends on the story you need to tell and the data you have which you want to share.

6. What criteria or components are required for /in an effective information visualization?

   I think the amount of data is really important. In most stories sometimes you need to have a very complex figure but try to make it simple which means that you should try to minimize the amount of variables that you're going to show and highlight the important figure that is needed by the customer.

   And I don't know what colors are important in a way that you use to visualize information, but maybe don't use wrong color in explaining the results of your analysis like red color for positive numbers and green color for negative number.

7. In your opinion, what is mobility information?

   I think every available information about the movement of people or it might also be the number of people that move from one place to another in a certain time together with their frequency of visits and motives.

Customers (Event)

8. What information, in this case using event data, do customers want to know?

   Most of the time they want to know:
   - *How many people has visited their event in total?*
   - *What is the effect of the event to the city?*
   - *Did the event attract people only from nearby areas or also further areas?*
   - *How long did they stay?*

9. What kind of questions do customers ask from event data like mezuro has? Can you please give some examples?

   Same as above

10. From your experience, what are good techniques or guidelines on how to present or visualize the information to them? Can you please explain it step by step?

    Actually now we don't have any guidelines on how to visualize the information. But we are in the process of making it. Now, we use our experiences and suggestions from the software to make an information visualization for our customers.

11. Do your clients always know exactly which information they need, or do you have to assist them in finding out the most relevant information for their problem? (E.g. availability of data, data processing, integration, filtering, validation).

    No, they don't know. Most of the time the process is like this, so they have questions and we are going to answer them and we are going to show them the information using the visualization and some text about the explanation of the visualization. So we take the initiative and then afterwards we are in the discussion with our clients to ask them whether the information is clear or not.

12. How would you determine the best type of information visualization for the customer?

I think it has to deal with experience. I want to keep the visualization as simple as possible, so I don't want to have too many variables. For instance, if I have a chart number of visitors during a month then I'm going to use a bar chart because I know that it is the one that customers can read easily because they can see the difference between several days.

For geographical information, I always visualize the origin of the visitors on a map and sometimes I combine it with a table that consists of the top 10 visitor's origin so they can see the actual numbers and really know the most important origin and how much they differ. I also use heat map because it can show where the hotspots are or where the most important areas are.

If I have data with percentages, I use a pie chart or stacked column chart. I also use stacked column chart when I have several moments in time and I want to show differences between them because you can see them exactly next to each other and it is easier to compare.

It depends on the number of dimensions. So if you have number of visitors and time frame then I use bar charts and if I also have that for several areas than I have to use line charts. If I have the geographical components I will make maps. So I think the number of dimensions you have those are very important in which type of visualization is best.

13. Can you please explain the process on how you create a visualization? maybe you can explain it step by step. (e.g. when you want to answer the questions from a customer using a visualization, what will you do? Then how you determine the color, the structure, etc.)

    So I have a question from a customer for instance our customer ask how many visitors come to the event and where do they come from, etc. Then we see the data and analyze it. Then we are going to think about what is the best way to present the information and answer the customer's question.

    I always check the visualization with my colleagues, so if I make something I always let my colleagues to see it and ask them whether they understand it or not. I also ask some suggestions from them if there is a part that I should improve.

14. What kind of complaints did you usually get from the customers regarding the information visualization that you have created or showed to them? (e.g. they do not like the type of the visualization, they do not like the colors, etc.)

    Most of the time, the most complex thing is that they need to know what the data behind the visualization or what are the numbers actually saying. So first they

have to read some texts and some backgrounds to be able to interpret the visualization.

Appendix B – Qualitative studies (experiment protocol & results)

I. Participant 1

Master thesis – information visualization

Usability test script *reprinted from Rocket Surgery Made Easy © 2010 Steve Krug*

Introduction

Dear participant, my name is Zaki, and I'm going to lead you through this session today.

You probably already have an idea of why we asked you here, but let me go over it again briefly: We are asking people to try using visualizations that we are working on so we can see whether they work effectively or not.

The session should take approximately 30 minutes. I want to make it clear right away that we are testing the visualizations, not you. You can't do anything wrong here. In fact, this is probably the one place today where you don't have to worry about making mistakes. Also, please don't worry that you are going to disappoint us. We are doing this to improve the visualizations, so we need to hear your honest reactions. As we go along, I am going to ask you to answer questions about the visualizations. All of the numbers that we show in this session are not real numbers. The numbers are made only for the research.

You may have noticed the sound recorder. With your permission, we are going to record our conversation. The recording will only be used to help us to interpret the results of this session and will only be used for scientific purposes. If you would, I am going to ask you to sign for your participation and your permission to record you.

Do you have any questions before we begin?

Background information

Before we look at the visualizations, I'd like to ask you a few quick questions on your background information:

1. What is your name?

2. How old are you? <20 21-30 31-40 41-50 51-60 >61

3. What is the highest level of education you have completed? *University (Master of Science)*

4. What is your occupation? *Program officer*

5. How long have you been working in your current position? *1.5 years*

6. Is it important to have or to use a visualization in your current job? If yes:

- How important is the visualization?

Yes, I would say it is important to use a visualization in my current job. I use it mostly as a tool to do part of my job which is currently I have to give advice about political environment to eldermen and the council. I notice that a visualization of information makes it easier to interpret. In my experience, it was very important for me to use a visualization when I worked with a huge amount of data like facts, figures, and numbers.

- Why is it important?

  Have already been answered in above question.

- What type of visualization do you usually use?

  I use graphs quite often. I can say I use pie charts and line charts quite often. And what I really like is maps that are used to visualize information. The reason why that works for municipality is that eldermen and council members can relate the amount of data (e.g. number of people) with specific locations, so that is the way to engage them more.

- Are the visualizations static (on a paper) or dynamic (e.g. a dashboard)?

  Currently, the visualization is usually static and I would like it to be more dynamic. I'm striving towards it being more dynamic.

- What type of data that are shown on the visualization that you usually use? (E.g. numerical data, categorical data, geographical data, etc.)

  It is mostly numerical data. I use financial data (e.g. number of subsidies, amount of money that the municipality spent, etc.) and also data from Mezuro like number of visitors to the city center.

- What type of tasks do you usually do using the visualization? (e.g. explore, compare, find information, etc.)

  I mostly use the visualization to see trends and to measure the goals or targets that we have made. For example, it would be comparing the results of the targets between current year and last year. So I can say mostly what I did was seeing comparisons and trends.

  Additional info:
  Based on your experience, do people who work in the municipality like to use a visualization or they just want to see the numbers, tables, etc.?

I think in the municipality because there is so much information that passes and so many decisions and such a broad theme we working in, so the key is that it is easy to understand information fast. So you can directly see what is happening and that depends from person to person what it is of course, but I do think certain graphs or visualizations can be very helpful to make people understand the information.

Great, we are done with the background information questions and we can go to the next part.

Scenario and questions

In this part, I'm going to ask you to answer some questions using visualizations. We need you to **"think aloud"** during the process of answering the questions. So you need to speak out what you see, what you think, etc. Before we continue to the real tests, I would like to start with a small exercise so you get the idea of what you need to do in this session.

You can start by looking at the page (www.uu.nl) and try to answer below question without using the search box. Don't forget to think aloud during the process of answering the question and please write down the answer.

"When does the European Law master study program in Utrecht University start?"

After you finish the small test, we can start the real questions using a role play and below situation.

Situation:

We are going to do a role play. Imagine that you work as an event planner in a municipality in the Netherlands, and there was an event (Glazen Huis) that was held in your city in December 2015. A company in the Netherlands analyzed the visitors event and made a report of the event. From the report of the event, you want to know the effect of the event on your city like the number of visitors, the type of transportation they used to your city, etc. Using the visualizations, please answer the questions below

Visualization 1

1. How does the event affect the total number of visitors to Heerlen city center?

   So we are talking about total amount of visitors and then this a regular day and this one an event day. So you can see from this bar that it is about 22,000 more visitors on the event day.

2. What differences do you notice between the event day and the regular day with regard to the visitor's origin?

   So you see from the maps, the darker the color the more visitors. You can see that on the 24th of December more visitors come from further away.

3. Are the people in Heerlen city center during Glazen Huis, people who often go to Heerlen city center, or are they infrequent visitors?

   I'm looking to the bar again to see the difference in the infrequent, regular, and frequent visitors. From the bar chart you can see that the highest increase is explained by the infrequent and regular visitors.

4. What is the difference in the longest traveled distance (as reported) on the event day compared to the regular day?

   Then I look to the maps again because that where we saw where people came from, and on the event day the furthest away I'm assuming it is Amsterdam. Then the furthest away on the regular day is Venlo, so the difference is 125 km.

5. What is your estimation of the traveled distance from cities which have 6-10 % of inhabitants in Heerlen city center on the event day?

   Then I'm looking at the colors group & I'm surprised that I didn't read the legend correctly. I just notice the colors mean the percentage of inhabitants that were in Heerlen on that day. So from what I can see it is basically up until 10 km radius.

6. How many people come from Amsterdam to Heerlen city center on the event day?

   So then you can see here that is between 0.1-0.5 % of the inhabitants of Amsterdam. Then Amsterdam population is around 800,000 inhabitants, so the answer probably around 80,000 people come from Amsterdam or is there another way to answer it?

Personal opinion – Vis 1:

- Which parts of the visualization could be improved, why & how?

  Well, the percentage of inhabitants of the city require you to have background knowledge. I have some sort of idea of how big Utrecht is and how big Eindhoven is so I can read this map but only because I have that information that I shouldn't need to have to read the map. So if it was in real numbers instead of the percentage that would be good. For the bar chart is already clear, I wouldn't change anything.

- What grade would you give to the ease of use of the visualization? (1-10)
  Bar chart: 8, I think it is quite easy.
  Map: 6, I could figure it out but it took me a while and it is not intuitive.

- What grade would you give to the aesthetic quality of the visualization? (1-10)
  Bar chart: 7
  Map: 7

Visualization 2

1. How does the event affect the utilization of public transportation (train) to Heerlen city center?

   Now I'm going to the legend immediately because I didn't do it on the first visualization. So what I want to know is do more people use the train on the map. So from the maps I can see dark blue and dark green, then I would say that there were more people that used the train because you can see more green colors on the map.

2. Which cities show the increase in the utilization of public transportation (train) to Heerlen city center, on the event day, compared to the regular day?

   This is actually difficult to see because there are some cities that are added, but I would say like I can see Venlo clearly, Maastricht just a little bit and Eindhoven.

3. What is your estimation of the preference of transportation type which is used by the visitors that come >150 km from Heerlen city center on the event day?

   So now I see the cities above 150 km from the map and the colors would mean that like in Amsterdam it is 50-50, then in Den haag it is 60-40, and in Rotterdam it is 70-30. So I would say that on average 60 % of the people from those cities prefer car than train on the event day.

4. What percentage of the people which come from Nijmegen use public transportation (train) to Heerlen city center on the event day?

   I'm looking at the legend and the map, then I would say 70% of the people from Nijmegen prefer to use train.

Personal opinion – Vis 2:

- Which part of the visualization could be improved, why & how?

  The colors definitely, so more contrast between the colors. Another one is I think the legend with 60-40 or 70-30 got me confused, I think it would have been easier if it said here only the whole number (e.g. only 60 % train instead of 40-60 %).

  Additional info:
  Most of the time I didn't see you use the bar chart, why?
  Yes that's true, I went straight to the map. I think it is because the amount of space it takes in the screen. Also I could answer these questions using the map so I didn't need to look further.

- What grade would you give to the ease of use of the visualization? (1-10)

  6

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  7

Visualization 3

1. How does the event affect the visitor's average duration of stay in Heerlen city center?

   So we have number of visitors per hour in a bar chart, then the average time they stay in Heerlen compare to the province where they from. From the question you hope it increases, and it does increase because you can see in the average staying time that is higher.

2. When is the busiest time in Heerlen city center on the regular day and on the event day?

   I start by looking the bar chart again, and normally it is at 13.00 or between 13.00-14.00. On the event day it's between 19.00 & 20.00.

3. What is your estimation of the increase in visitor numbers during the evenings (after 18.00 hrs) on the event day compared to the regular day?

   Well, my estimation is about 10.000 – 12.000. So I just subtract the number of visitors for each hour after 18.00.

4. Which province(s) show(s) the biggest difference in average staying time, between the regular day and the event day?

   So I go to the map, then I need to find out the biggest difference in color from light blue to dark blue which I see Noord Holland and Drenthe.

5. What differences do you notice based on the relation between the distances to Heerlen city center on average staying time on the regular day? And on the event day?

   So you would assume on the regular day the people that come from further away didn't stay as long as the closer areas. For the event day the staying time was different for each province.

6. How many visitors come from Noord Holland and stay in Heerlen city center only for 1 hour?

   I can't get the answer using the visualization.

Personal opinion – Vis 3:

- Which parts of the visualization could be improved, why & how?

I think the bar chart is unclear. It is not clear whether a bar represent an hour or a duration (e.g. is it 11.30-12.30 or 11.00-12.00 or 12.00-13.00). I think that could be improved.

For the staying time, I think it is more interesting to show the number of people that stay for <1 hour, >3hours, >4 hours, etc. So what I mean is bigger gaps in the staying time. Because these 15 minutes are not very crucial information.

- What grade would you give to the ease of use of the visualization? (1-10)

  6

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  7

II. Participant 2

Master thesis – information visualization

Usability test script *reprinted from Rocket Surgery Made Easy © 2010 Steve Krug*

Introduction

Same as in the first participant.

Background information

Before we look at the visualizations, I'd like to ask you a few quick questions on your background information:

1. What is your name?

2. How old are you? <20 21-30 31-40 41-50 51-60 >61

3. What is the highest level of education you have completed? *WO (university)*

4. What is your occupation? *Advisor / smart city innovation officer*

5. How long have you been working in your current position? *1.5 years*

6. Is it important to have or to use a visualization in your current job? If yes:

   - How important is the visualization?

     In my current job I use a visualization to show other people like my colleagues the projects that what we are doing. I think it is important to have a visualization because it is easier for the people to understand the information. In my opinion an

excel sheet with a lot of data is difficult to read and understand, so a visualization is helpful.

- Why is it important?

  Have already been answered in above question.

- What type of visualization do you usually use?

  I usually use graphs like bar chart, line chart, and map. I sometimes also use infographics.

- Does the visualization static (on a paper) or dynamic (e.g. a dashboard)?

  So currently we are still using a static visualization but hopefully in the future we will have a dynamic visualization. So for the dynamic visualization, we still in the discussion phase with a partner.

- What type of data that are shown on the visualization that you usually use? (E.g. numerical data, categorical data, geographical data, etc.)

  Maybe most of the time I can say numerical data like number of people who lives in Zeist, number of people that come to Zeist, etc.

- What type of tasks do you usually do using the visualization? (e.g. explore, compare, find information, etc.)

  I use a visualization to see trends, comparison, and evaluation. So for example we compare the number of people in Zeist city center on the normal day with the day where there is a special event. For the evaluation, what I mean is like we want to see the effectiveness of an event or a project (e.g. infrastructural project) in our city.

Great, we are done with the background information questions and we can go to the next part.

<u>Scenario and questions</u>

Same as in the first participant.

Situation:

Same as in the first participant

Visualization 1

1. How does the event affect the total number of visitors to Heerlen city center?

   I'm looking at the effect of the event and total extra visitors, and I think that is the answer. So there are 21,761 extra visitors.

2. What differences do you notice between the event day and the regular day with regard to the visitor's origin?

   Now I'm looking at the maps because I think that where the people come from. Then I read and try to understand the title "percentage of inhabitants to Heerlen". I notice that the left map is the regular day and the right map is the event day. So I could say that on the event day there are people not only come from Limburg, but also from outside of Limburg like Amsterdam, Utrecht, etc.

3. Are the people in Heerlen city center during Glazen Huis, people who often go to Heerlen city center, or are they infrequent visitors?

   So I'm looking at the visitor's frequency information. Then I go back to the bar chart and I think the answer is there are more infrequent visitors than the regular and frequent visitors.

4. What is the difference in the longest traveled distance (as reported) on the event day compared to the regular day?

   I'm looking at the maps again and on the event day the longest traveled distance is from Amsterdam which is 180 km. Then on the regular day the longest traveled distance is from Venlo which is 55 km. So 180 – 55 = 125 km.

5. What is your estimation of the traveled distance from cities which have 6-10 % of inhabitants in Heerlen city center on the event day?

   I look at the color legend and 6-10% is this one "dark blue". Then I look the map on the event day and I think it is 8 or 10 km.

6. How many people come from Amsterdam to Heerlen city center on the event day?

   I think the percentage is about 0.1-0.5 % but I can't directly answer the number of people from Amsterdam because I don't have the information on the number of inhabitants in Amsterdam.

Personal opinion – Vis 1:

- Which parts of the visualization could be improved, why & how?

  - *So for me I didn't see the "Glazen Huis event" title, I just directly went to the visualization because each graph or map already has their own title. Then I*

understand what the visualizations are about. For me the title on the maps are too long.

- *For the bar chart is already clear, but maybe the position of the information about the visitor's frequency is a little bit too far. Maybe you can move it closer to the bar chart. For the "information" part I will just skip it or delete it.*

- *For me, instead of showing the percentage I prefer to show the actual number of people or visitors to Heerlen city center.*

- What grade would you give to the ease of use of the visualization? (1-10)

  I would say 8

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  7.5 and I prefer to have less blue on the color.

Visualization 2

1. How does the event affect the utilization of public transportation (train) to Heerlen city center?

   So first I'm looking at the bar chart and I see effect of the event on the transportation type. I think there more people that use the train by 4%.

2. Which cities show the increase in the utilization of public transportation (train) to Heerlen city center, on the event day, compared to the regular day?

   I start by looking at the maps. Then I think I should look at the difference in the color between the left and the right maps. So I need to look cities which have blue color on the left map and green color on the right map. I think the cities are Venlo, Eindhoven and Amsterdam.

3. What is your estimation of the preference of transportation type which is used by the visitors that come >150 km from Heerlen city center on the event day?

   So I look at the right map and I see these 3 cities (Amsterdam, Den Haag, and Rotterdam). Then I can say they prefer to use the car.

4. What percentage of the people which come from Nijmegen use public transportation (train) to Heerlen city center on the event day?

   Nijmegen has dark green color, so I think it is 30-70 %.

Personal opinion – Vis 2:

- Which part of the visualization could be improved, why & how?

- For me, it is easier to understand the 2nd visualization because I have seen the 1st visualization. Again I would say the title on each graph is too long.

- The legend part, I like it but I think it is a little bit difficult for me to understand when I first saw it maybe because of the percentage group like 40-60 %, 30-70 %, etc.

- Still in the legend part, for me the green color in 40-60 % and 30-70 % are almost the same.

- What grade would you give to the ease of use of the visualization? (1-10)

  I would give 9

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  I like this one and more colors so I would give 9

Visualization 3

1. How does the event affect the visitor's average duration of stay in Heerlen city center?

   So the bar chart shows the number of visitors per hour in Heerlen. Then the above one is the regular day and the below one is the event day. I see that from the box the average staying time increase 0.5 hour.

2. When is the busiest time in Heerlen city center on the regular day and on the event day?

   From the bar chart, on the regular day I would say at 13.00 and on the event day at 19.00.

3. What is your estimation of the increase in visitor numbers during the evenings (after 18.00 hrs) on the event day compared to the regular day?

   My estimation will be 10,000-12,000 visitors. I just subtract the number of visitors like 13,000 – 3,000.

4. Which province(s) show(s) the biggest difference in average staying time, between the regular day and the event day?

   So I'm looking at the maps and I think I need to see province or provinces which have darker blue color on the right map than on the left map. I think Noord Holland has the biggest difference in average staying time for at least 45 minutes.

5. What differences do you notice based on the relation between the distances to Heerlen city center on average staying time on the regular day? And on the event day?

   From the maps, I think when people are living closer to Heerlen city center they do not have much differences in the staying time between the regular day and the event day.

6. How many visitors come from Noord Holland and stay in Heerlen city center only for 1 hour?

   For this one, I can't see the answer from the visualization.

Personal opinion – Vis 3:

- Which parts of the visualization could be improved, why & how?

  - *For the bar chart, I think it will be better if you can put the hours (y axis) in both charts.*

  - *For the map, it is ok for me. I like it.*

- What grade would you give to the ease of use of the visualization? (1-10)

  I would give 9

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  Also 9

III. Participant 3

Master thesis – information visualization

Usability test script *reprinted from Rocket Surgery Made Easy © 2010 Steve Krug*

Introduction

Same as in the first participant.

Background information

Before we look at the visualizations, I'd like to ask you a few quick questions on your background information:

1. What is your name?

2. How old are you? <20 21-30 31-40 41-50 51-60 >61

3. What is the highest level of education you have completed? *University (master degree)*

4. What is your occupation? *Economist in the department spatial and economic developments.*

5. How long have you been working in your current position? *16 years*

6. Is it important to have or to use a visualization in your current job? If yes:

- How important is the visualization?

  It is important to explain or it is helpful to explain numbers or figures like number of jobs in the city, education degrees, income levels, etc. It is much easier to understand if you have something in a graphic than a text. So for me it is easier and faster when I used the visualization.

- Why is it important?

  Have already been answered in above question.

- What type of visualization do you usually use?

  So I usually use bar chart, pie chart, maps, and line chart. I think that are the most common graphs that I used.

- Are the visualizations static (on a paper) or dynamic (e.g. a dashboard)?

  For me now I still use the static visualizations. Because usually I use the visualizations in a presentation or a meeting.

- What type of data that are shown on the visualization that you usually use? (E.g. numerical data, categorical data, geographical data, etc.)

  I'm an economist so most of the time I can say financial data. I also use data about income levels, education degrees, and type of houses.

- What type of tasks do you usually do using the visualization? (e.g. explore, compare, find information, etc.)

  I think mostly I use for comparison. For example, I compare the income levels, education degrees, number of jobs between Heerlen and other cities.

Great, we are done with the background information questions and we can go to the next part.

Scenario and questions

Same as in the first participant.

Situation:

Same as in the first participant.

Visualization 1

1. How does the event affect the total number of visitors to Heerlen city center?

   I'm looking at the bar chart and I see that the number of infrequent visitors increase and the number of frequent visitors is about the same.

2. What differences do you notice between the event day and the regular day with regard to the visitor's origin?

   So I look at the information part and I know that the event day is on 24th of December and the regular day is on the 3rd of December, but for this question I can't find the answer from the visualization. It's not clear and too much text for me.

3. Are the people in Heerlen city center during Glazen Huis, people who often go to Heerlen city center, or are they infrequent visitors?

   I look at the bar chart and I think the answer is frequent visitors.

4. What is the difference in the longest traveled distance (as reported) on the event day compared to the regular day?

   So now I'm looking at the left map which is the regular day and the longest traveled distance is 55 km. Then I'm looking at the right map which is the event day and the longest traveled distance is 180 km. So the difference is 125 km.

5. What is your estimation of the traveled distance from cities which have 6-10 % of inhabitants in Heerlen city center on the event day?

   From the legend, 6-10 % is the second color. Then I go to the right map and I think it is about 10 km.

6. How many people come from Amsterdam to Heerlen city center on the event day?

   So Amsterdam has light blue color so it is 0.1-0.5 %, but I do not have the number of people that live in Amsterdam or do you have it?

Personal opinion – Vis 1:

- Which parts of the visualization could be improved, why & how?
  - *For me it is too much texts.*
  - *Too much difference information on one sheet.*
  - *Maybe the colors, it is not very distinctive.*

- What grade would you give to the ease of use of the visualization? (1-10)

  I think 6

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  Also 6

Visualization 2

1. How does the event affect the utilization of public transportation (train) to Heerlen city center?

   So this is about transportation, then I look at the bar chart. I see the transportation by car and by train, so the answer will be increase of 4 %.

2. Which cities show the increase in the utilization of public transportation (train) to Heerlen city center, on the event day, compared to the regular day?

   So now I'm going to the maps and I look at the cities that have green color. I think the cities are Amsterdam and Eindhoven.

3. What is your estimation of the preference of transportation type which is used by the visitors that come >150 km from Heerlen city center on the event day?

   I look at the right map and I think it is train for Amsterdam.

4. What percentage of the people which come from Nijmegen use public transportation (train) to Heerlen city center on the event day?

   From the legend and the color I think it is 30-70 %.

Personal opinion – Vis 2:

- Which part of the visualization could be improved, why & how?

  Too much information and texts on 1 sheet and for me I need to look quite long before I can see the information that I'm looking for.

- What grade would you give to the ease of use of the visualization? (1-10)

  5

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  6

Visualization 3

1. How does the event affect the visitor's average duration of stay in Heerlen city center?

   So the above bar chart is the regular day and the below bar chart is the event day. Then I can see on the regular day is 3 hours and on the event day is 3.5 hours. So it is increase by half an hour. It is not much.

2. When is the busiest time in Heerlen city center on the regular day and on the event day?

   Still in the bar chart I see that on the regular day the busiest time is at 12.00 and on the event day is at 19.00 or 20.00.

3. What is your estimation of the increase in total number of visitors during the evenings (after 18.00 hrs) on the event day compared to the regular day?

   The increase is large if you compare it. After 18.00 the increase is about 30,000.

4. Which province(s) show(s) the biggest difference in average staying time, between the regular day and the event day?

   I look at the maps now and from the color I think it is Noord Holland.

5. What differences do you notice based on the relation between the distances to Heerlen city center on average staying time on the regular day? And on the event day?

   So from the maps I can see that on the regular day, the shorter you travel the longer you stay. For the event day, the longer you travel the longer you stay.

6. How many visitors come from Noord Holland and stay in Heerlen city center only for 1 hour?

   I don't that because I don't have the number.

Personal opinion – Vis 3:

- Which parts of the visualization could be improved, why & how?

  - *I think you can make the hours (y axis) in both charts.*
  - *For me this visualization is much clearer than the other two because I can see the information in one look.*

- What grade would you give to the ease of use of the visualization? (1-10)

  8

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  8

IV. Participant 4

Master thesis – information visualization

Usability test script *reprinted from Rocket Surgery Made Easy © 2010 Steve Krug*

Introduction

Same as in the first participant.

Background information

Before we look at the visualizations, I'd like to ask you a few quick questions on your background information:

1. What is your name?

2. How old are you? <20 21-30 31-40 41-50 51-60 >61

3. What is the highest level of education you have completed? *University (master degree)*

4. What is your occupation? *Senior advisor research and statistic*

5. How long have you been working in your current position? *15 years.*

6. Is it important to have or to use a visualization in your current job? If yes:

   - How important is the visualization?

     It can helps you to explain things or to get things understood for either someone else or for yourself. So from a visualization you can give more information when you do a visualization. So I think the visualization is important because it can helps you, it can support you or it can make it easier for the one you are explaining things to.

     The people here have very busy agendas and they need to get the information very quick, so it is very supporting to use visualizations.

   - Why is it important?

Have already been answered in above question.

- What type of visualization do you usually use?

  So we usually use all kinds of charts like line chart, bar chart, and pie chart. Not only the charts, but we also use maps.

- Does the visualization static (on a paper) or dynamic (e.g. a dashboard)?

  The visualizations are mostly static in this time. So maybe there are some possibilities to have the dynamic visualization like using an application. That could be nice but it is not necessary now.

- What type of data that are shown on the visualization that you usually use? (E.g. numerical data, categorical data, geographical data, etc.)

  Maybe I can say like statistical data and numerical data. Also data that we get from Mezuro.
  So for example you can see for a year the number of visitors in Heerlen and you can compare it with the other cities.

- What type of tasks do you usually do using the visualization? (e.g. explore, compare, find information, etc.)

  We are doing evaluation, so for example we want to see the effect of a certain project or policy to our city. We are also do monitoring, so we know the progress of a certain project.

Great, we are done with the background information questions and we can go to the next part.

Scenario and questions

Same as in the first participant.

Situation:

Same as in the first participant.

Visualization 1

1. How does the event affect the total number of visitors to Heerlen city center?

   Ok, so this is about the number of visitors to Glazen Huis. I'm looking at the bar chart now, I guess what I see is the difference between the 3rd of December and 24th of December. So I think it is 21,761 extra visitors.

2. What differences do you notice between the event day and the regular day with regard to the visitor's origin?

   Origin is where they come from, so I look at the maps. From the maps what I see is normally the visitors are coming from the province of Limburg and on 24th of December there are more visitors that coming from outside Limburg.

3. Are the people in Heerlen city center during Glazen Huis, people who often go to Heerlen city center, or are they infrequent visitors?

   So I look at the infrequent, regular, and frequent visitors on the bar chart. I think the answer is more frequent visitors.

4. What is the difference in the longest traveled distance (as reported) on the event day compared to the regular day?

   I don't know where I should look for the distance, but I think it is the one with bracket on the maps. So when you need the longest traveled distance then on the 24th of December it is 180 km, and on the 3rd of December it is 55 km. So the difference is 125 km.

5. What is your estimation of the traveled distance from cities which have 6-10 % of inhabitants in Heerlen city center on the event day?

   So I look at the legend and 6-10 % is the dark blue. From the maps, I think the answer is 30 km.

6. How many people come from Amsterdam to Heerlen city center on the event day?

   It is between 0.1-0.5 %, then I think I need to calculate it with the number of visitors in the bar chart (56,618) but I'm not sure. So if it possible I need to see the data set to answer this question.

Personal opinion – Vis 1:

- Which parts of the visualization could be improved, why & how?

  - *I think the information is already good and clear.*
  - *I think the visitor's frequency information could be moved closer to the bar chart.*

- What grade would you give to the ease of use of the visualization? (1-10)

  I would say it is 7

- What grade would you give to the aesthetic quality of the visualization? (1-10)

8

Visualization 2

1. How does the event affect the utilization of public transportation (train) to Heerlen city center?

   So this is about the transportation preferences and from the legend I see the range for the one that prefer car and the one that prefer train. What I see from the map is on the event day, most of the people from this area (area close to Amsterdam) prefer to use car. Then people that come from Amsterdam and Eindhoven prefer to use train.

2. Which cities show the increase in the utilization of public transportation (train) to Heerlen city center, on the event day, compared to the regular day?

   From the maps, I guess it is Nijmegen and Amsterdam.

3. What is your estimation of the preference of transportation type which is used by the visitors that come >150 km from Heerlen city center on the event day?

   Still from the maps, I guess they use car but I don't know the number. It is only the percentage. So I would say they prefer car.

4. What percentage of the people which come from Nijmegen use public transportation (train) to Heerlen city center on the event day?

   For Nijmegen, I think it is 30-70 %.

Personal opinion – Vis 2:

- Which part of the visualization could be improved, why & how?

  - For me, the visualization is a little bit difficult to understand. I need to look at the visualization and understand it for quite long.
  - I guess the legend can be better. The percentage (%) range is quiet strange for me.

- What grade would you give to the ease of use of the visualization? (1-10)

  5

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  8

Visualization 3

1. How does the event affect the visitor's average duration of stay in Heerlen city center?

   The duration that is the very important one, when they are spending more time in the city they can spend more money. So from the information on the bar chart I see that the average staying time increases half an hour and also more people are staying longer on the event day.

2. When is the busiest time in Heerlen city center on the regular day and on the event day?

   From the bar chart, the busiest time on the regular day is at 13.00 and on the event day is at 19.00.

3. What is your estimation of the increase in total number of visitors during the evenings (after 18.00 hrs) on the event day compared to the regular day?

   So I see the number of visitors after 18.00 and I think the difference between the event day and the regular day is about 2 or 3 times more in each hour.

4. Which province(s) show(s) the biggest difference in average staying time, between the regular day and the event day?

   So I'm looking at the differences between the two maps and also the legend. So like the province of Noord Holland, on the regular day the average staying time is 3 hours and on the event day it is 3 hours and 45 minutes so that is a difference of 45 minutes. I think from the color, I would say Noord Holland.

5. What differences do you notice based on the relation between the distances to Heerlen city center on average staying time on the regular day? And on the event day?

   On the regular day, the further the visitors come from, the shorter they stay. And in this case on the event day it is very different. I can also say that on the event day, the people that come from further distance have chosen to stay longer. They really took a decision to stay in Heerlen for the event I guess. So the event is the reason why they come and stay longer in Heerlen.

6. How many visitors come from Noord Holland and stay in Heerlen city center only for 1 hour?

   I can't see that.

Personal opinion – Vis 3:

- Which parts of the visualization could be improved, why & how?

- *I think the legend for the white color should be revised. It is not only less than 3 hours but it can be also no data. So like "<3 or no data". Because now I have to look carefully and then I see that they are staying for less than 3 hours. But I can also interpret it like they are not coming.*
- *For the the maps, maybe it could be helpful if you also give a color for the area that have average staying time of "<3 or no data". Because from what I see, the surrounding areas (e.g. Belgium, Germany, and the sea) are also white, so you need to differentiate it.*

- What grade would you give to the ease of use of the visualization? (1-10)

  Bar chart: I think it is very helpful so I would say an 8
  Map: 6

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  7

V. Participant 5

Master thesis – information visualization

***Usability test script** reprinted from Rocket Surgery Made Easy © 2010 Steve Krug*

Introduction

Same as in the first participant.

Background information

Before we look at the visualizations, I'd like to ask you a few quick questions on your background information:

1. What is your name?

2. How old are you? <20 21-30 31-40 41-50 51-60 >61

3. What is the highest level of education you have completed? *Secondary vocational education (MBO)*

4. What is your occupation? *Geoformation advisor*

5. How long have you been working in your current position? *9 years*

6. Is it important to have or to use a visualization in your current job? If yes:

- How important is the visualization?

  We are from geo department so it is important for us to show the data on the map. It is important because it makes the data visible when you have a huge amount of data and it is also easier to understand.

- Why is it important?

  Have already been answered in above question.

- What type of visualization do you usually use?

  Most of the time I would say map, like heat map. Sometimes also the charts like line chart, bar chart, and pie chart.

- Are the visualizations static (on a paper) or dynamic (e.g. a dashboard)?

  Right now it still on a paper or the static one because we use the visualization for reports and we need to print it.

- What type of data that are shown on the visualization that you usually use? (E.g. numerical data, categorical data, geographical data, etc.)

  I would say demographic data (e.g. gender, age, etc.), geographical data, and also the data from Mezuro (e.g. number of visitors, etc.).

- What type of tasks do you usually do using the visualization? (e.g. explore, compare, find information, etc.)

  So maybe I can say that we use the visualization for exploration and evaluation. So for example we use the visualization to evaluate the effect of King's day in Zeist city center and also explore the origin of the visitors.

Great, we are done with the background information questions and we can go to the next part.

Scenario and questions

Same as in the first participant.

Situation:

Same as in the first participant.

Visualization 1

1. How does the event affect the total number of visitors to Heerlen city center?

   I start look at the legend to see what the colors mean then I see the maps. From the maps I see that on the regular day most of the people are coming from Limburg, but on the event day there also other people that come from outside Limburg. So I think the effect of the event is the increase of visitors from cities that close to Heerlen.

2. What differences do you notice between the event day and the regular day with regard to the visitor's origin?

   Still from the maps, I think the difference is on the event day there are more people that come from the surrounding areas of Heerlen.

3. Are the people in Heerlen city center during Glazen Huis, people who often go to Heerlen city center, or are they infrequent visitors?

   I can't see the answer in the maps, so I have to go to the bar chart. Then I see the total number of visitors, infrequent, regular, and frequent visitors. I think the answer is there are more infrequent visitors.

4. What is the difference in the longest traveled distance (as reported) on the event day compared to the regular day?

   So the difference is on the regular day there are no visitors coming from further distance or outside Limburg.

5. What is your estimation of the traveled distance from cities which have 6-10 % of inhabitants in Heerlen city center on the event day?

   So 6-10 % is the dark blue, then I have to look the dark blue color on the map. So it is around 10 – 16 km.

6. How many people come from Amsterdam to Heerlen city center on the event day?

   Amsterdam has a light blue, so it is 0.1-0.5 %. But I don't know the number of visitor.

Personal opinion – Vis 1:

- Which parts of the visualization could be improved, why & how?

  For the maps, if it is possible maybe you can zoom in the maps. So what I mean is maybe you can have 1 map that shows the whole Netherland and on the corner you can make a small map that shows only the area of the event with the surrounding cities.

  Overall, the visualization is already good. In one screen you give so much information.

- What grade would you give to the ease of use of the visualization? (1-10)

  8

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  8

Visualization 2

1. How does the event affect the utilization of public transportation (train) to Heerlen city center?

   I look at the legend and I see prefer car and prefer train. Then 50-50 means 50 % prefer car and 50 % prefer train. From the map on the event day, there are more green color. So I think there are more people use the train.

2. Which cities show the increase in the utilization of public transportation (train) to Heerlen city center, on the event day, compared to the regular day?

   It is difficult a one for me, I cannot see clearly the answer from the screen but maybe it is Amsterdam.

3. What is your estimation of the preference of transportation type which is used by the visitors that come >150 km from Heerlen city center on the event day?

   This one also difficult, because I see like Amsterdam is 50-50 and like Rotterdam prefers car. So I guess they prefer both transportation.

4. What percentage of the people which come from Nijmegen use public transportation (train) to Heerlen city center on the event day?

   I look at Nijmegen from the map and from the color I think it is 30-70 %.

Personal opinion – Vis 2:

- Which part of the visualization could be improved, why & how?

  When I see the maps at the first time, I interpreted the color differently. I thought it was continuous instead of categorical. But after I read it again, now I understand it.

- What grade would you give to the ease of use of the visualization? (1-10)

  7

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  8

Visualization 3

1. How does the event affect the visitor's average duration of stay in Heerlen city center?

   So it is about the average staying time. The above bar chart is the regular day and the below bar chart is the event day. I think the effect of the event is there are more visitors on the event day especially in the evening.

2. When is the busiest time in Heerlen city center on the regular day and on the event day?

   I look at the bar chart, on the regular day is at 13.00 o'clock and on the event day is at 19.00 o'clock.

3. What is your estimation of the increase in total number of visitors during the evenings (after 18.00 hrs) on the event day compared to the regular day?

   So I think I need to calculate the number after 18.00 and I think the increase is about 45,000 visitors.

4. Which province(s) show(s) the biggest difference in average staying time, between the regular day and the event day?

   The left map is 3rd of December and the right map is 24th of December. From the color I think the answer is Noord Holland.

5. What differences do you notice based on the relation between the distances to Heerlen city center on average staying time on the regular day? And on the event day?

   What I see from the maps, there are increase in the average staying time for people that come from Noord Brabant, Zuid Holland, and Noord Holland.

6. How many visitors come from Noord Holland and stay in Heerlen city center only for 1 hour?

   I can't find it so I guess it is not possible.

Personal opinion – Vis 3:

- Which parts of the visualization could be improved, why & how?

  For me this visualization is easier to understand than the others. I also like it and I don't have any comments for the improvement.

- What grade would you give to the ease of use of the visualization? (1-10)

  8

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  8

VI. Participant 6

Master thesis – information visualization

Usability test script *reprinted from Rocket Surgery Made Easy © 2010 Steve Krug*

Introduction

Same as in the first participant.

Background information

Before we look at the visualizations, I'd like to ask you a few quick questions on your background information:

1. What is your name?

2. How old are you? <20 21-30 31-40 41-50 51-60 >61

3. What is the highest level of education you have completed? *Bachelor degree*

4. What is your occupation? *Geographic Information Specialist*

5. How long have you been working in your current position? *6 months*

6. Is it important to have or to use a visualization in your current job? If yes:

   - How important is the visualization?

     I would say visualizing data is important because it is the basic or main task of my occupation especially in geographic data. There is a saying that "a picture is worth a thousand words", so by using a visualization it helps me to present the information to other people.

   - Why is it important?

     Have already been answered in above question.

   - What type of visualization do you usually use?

     In my current job, mainly I use maps. Sometimes I also use graphs like bar chart, pie chart, etc.

   - Are the visualizations static (on a paper) or dynamic (e.g. a dashboard)?

     It is hard to say, some people want to have the printed version of the visualization so it is static but sometimes it is also dynamic. So we make the dynamic one when

the data changes. So what I mean is like when you have data that change in time. For example when I have data about number of people in Zeist city center in one month, I can show it on the map and it will change day by day.

- What type of data that are shown on the visualization that you usually use? (E.g. numerical data, categorical data, geographical data, etc.)

  Most of the time I use geographical data.

- What type of tasks do you usually do using the visualization? (e.g. explore, compare, find information, etc.)

  I would say we usually use the visualization for exploration. So for example by using a map we can see and explore where are the visitors of Zeist come from.

Great, we are done with the background information questions and we can go to the next part.

Scenario and questions

Same as in the first participant.

Situation:

Same as in the first participant.

Visualization 1

1. How does the event affect the total number of visitors to Heerlen city center?

   So the event is Glazen Huis and from the information it was on 24th of December. Then I look at the bar chart and the total visitors were a lot more on the event day. So I think the event has a big effect.

2. What differences do you notice between the event day and the regular day with regard to the visitor's origin?

   So I'm looking at the infrequent and the regular visitors. I think there is quite a difference in the infrequent visitors and in the regular visitors between the event day and the regular day.

3. Are the people in Heerlen city center during Glazen Huis, people who often go to Heerlen city center, or are they infrequent visitors?

   I look at the bar chart and I think most of the people are frequent visitors because the number is the highest.

4. What is the difference in the longest traveled distance (as reported) on the event day compared to the regular day?

   For the longest traveled distance on the event day it would be Amsterdam by 180 km and on the regular day it would be Venlo by 55 km. So the difference is 180-55: 125 km.

5. What is your estimation of the traveled distance from cities which have 6-10 % of inhabitants in Heerlen city center on the event day?

   So the traveled distance is the one that in the brackets. Then from the legend 6-10 % is the second color, so I guess it is about 10 km.

6. How many people come from Amsterdam to Heerlen city center on the event day?

   I think it is about 0.1-0.5 % but I can't say the absolute number.

Personal opinion – Vis 1:

- Which parts of the visualization could be improved, why & how?

  Maybe there is too much texts so you have to read a lot, but overall it is quite clear I think.

- What grade would you give to the ease of use of the visualization? (1-10)

  9

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  9

Visualization 2

1. How does the event affect the utilization of public transportation (train) to Heerlen city center?

   It is about the visitors preference of transportation type. From the bar chart I see that there is increase of public transportation by train.

2. Which cities show the increase in the utilization of public transportation (train) to Heerlen city center, on the event day, compared to the regular day?

   So I think I should compare the color between the two maps. Then I guess the cities that show the increase are Amsterdam and Eindhoven.

3. What is your estimation of the preference of transportation type which is used by the visitors that come >150 km from Heerlen city center on the event day?

   I look at the maps and I think they are coming by train.

4. What percentage of the people which come from Nijmegen use public transportation (train) to Heerlen city center on the event day?

   From the color, I think it is about 40-60 % or 30-70 % I'm not sure. I will just choose 40-60%.

Personal opinion – Vis 2:

- Which part of the visualization could be improved, why & how?

  For the legend I would say use only one number of percentage. Maybe something like this:

  Car *Train*
  80% 70% 60% 50% 60% 70% 80%

  Or

  Prefer car
  0% 30% 50% 80% 100%

  Prefer train
  0% 30% 50% 80% 100%

- What grade would you give to the ease of use of the visualization? (1-10)

  9

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  9

Visualization 3

1. How does the event affect the visitor's average duration of stay in Heerlen city center?

   I look at the bar chart and I think the above one is the regular day and the below one is during the event. So I think the people stay longer on the event day especially in the evening.

2. When is the busiest time in Heerlen city center on the regular day and on the event day?

   On the regular day, it is around 12.00. On the event day, it is around 19.00.

3. What is your estimation of the increase in total number of visitors during the evenings (after 18.00 hrs) on the event day compared to the regular day?

   So the total number after 18.00, then I need to calculate it first. I miss some figures on the bar chart but I will try it. So I guess the increase is about 60,000 visitors.

4. Which province(s) show(s) the biggest difference in average staying time, between the regular day and the event day?

   From the colors on the maps, the biggest difference would be Noord Holland and Noord Brabant.

5. What differences do you notice based on the relation between the distances to Heerlen city center on average staying time on the regular day? And on the event day?

   I would say that on the event day people from further away would stay longer but it is not entirely true, because Noord Brabant is quite close to Heerlen and they stay quite long. For the regular day, people from closer areas would stay longer.

6. How many visitors come from Noord Holland and stay in Heerlen city center only for 1 hour?

   For only 1 hour, I can't answer it.

Personal opinion – Vis 3:

- Which parts of the visualization could be improved, why & how?

  It is clear for me, so nothing to change.

- What grade would you give to the ease of use of the visualization? (1-10)

  9

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  9

VII. Participant 7

Master thesis – information visualization

***Usability test script** reprinted from Rocket Surgery Made Easy © 2010 Steve Krug*

Introduction

Same as in the first participant.

Background information

Before we look at the visualizations, I'd like to ask you a few quick questions on your background information:

1. What is your name?

2. How old are you? <20 21-30 31-40 41-50 51-60 >61

3. What is the highest level of education you have completed? *Bachelor degree*

4. What is your occupation? *Marketer at Utrecht municipality*

5. How long have you been working in your current position? *3 years*

6. Is it important to have or to use a visualization in your current job? If yes:

- How important is the visualization?

  First of all it is important for internal communication, I think it is very important to visualize the things we are busy with like the amount of visitors in our city or overnight stays or something like that because then we become more aware of the actual situation and how we can improve it. Secondly, I think it is very interesting to present the information to the outside world and show them like tourism.

- Why is it important?

  It is important for me because it is easier for me to present the information and also it is easier for the customer to understand it. I think when it is more attractive to see, people will have more attention to it.

- What type of visualization do you usually use?

  I usually use graphs like pie chart, bar chart, line chart, etc. But sometimes we also only use tables. It is not attractive but maybe useful.

- Are the visualizations static (on a paper) or dynamic (e.g. a dashboard)?

  For now it still the static one, but we like to develop a dashboard so we can use it interactively.

- What type of data that are shown on the visualization that you usually use? (E.g. numerical data, categorical data, geographical data, etc.)

Maybe like the amount of spending that people do in our city, number of visitors, where they come from, etc.

- What type of tasks do you usually do using the visualization? (e.g. explore, compare, find information, etc.)

  I could say we use it not only to give information to people, but we also use it to compare different information for example in the first quarter of this year how many people come to our city.

Great, we are done with the background information questions and we can go to the next part.

Scenario and questions

Same as in the first participant.

Situation:

Same as in the first participant.

Visualization 1

1. How does the event affect the total number of visitors to Heerlen city center?

   So I start to see effect of the event and total number of visitors. Then I think the answer is this one 21.761 extra visitors.

2. What differences do you notice between the event day and the regular day with regard to the visitor's origin?

   I'm looking at the maps which show the regular day and the event day. Then my answer will be there are more people from outside of Limburg province on the event day.

3. Are the people in Heerlen city center during Glazen Huis, people who often go to Heerlen city center, or are they infrequent visitors?

   The yellow one is the area of the event and the area that close to the area of the event on the regular day and event day have intense color. So I would say the answer is both of them.

4. What is the difference in the longest traveled distance (as reported) on the event day compared to the regular day?

   From the maps I would say bigger distance because of the event.

5. What is your estimation of the traveled distance from cities which have 6-10 % of inhabitants in Heerlen city center on the event day?

So 6-10 % I think it is this one, the dark blue, and I would say around 10-16 km.

6. How many people come from Amsterdam to Heerlen city center on the event day?

   I don't think it is possible to be answered from the visualization.

Personal opinion – Vis 1:

- Which parts of the visualization could be improved, why & how?

  - I think it is better to work with more different colors because it is a little bit hard to differentiate it.
  - Overall I think it is a clear sheet, but maybe it can be a little more attractive by using an image

- What grade would you give to the ease of use of the visualization? (1-10)

  7.5

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  6.5

Visualization 2

1. How does the event affect the utilization of public transportation (train) to Heerlen city center?

   So I start by looking at the left side of the sheet and I can see that the preference of transportation, by car and by train. So my answer is more people use the train.

2. Which cities show the increase in the utilization of public transportation (train) to Heerlen city center, on the event day, compared to the regular day?

   I'm looking at the legend and the dark blue means more car and dark green means more train. From the maps I would say Amsterdam and the east of Breda.

3. What is your estimation of the preference of transportation type which is used by the visitors that come >150 km from Heerlen city center on the event day?

   Still from the maps, I see on the event day they use more cars but in Amsterdam it 50-50. So I would say mostly they prefer car.

4. What percentage of the people which come from Nijmegen use public transportation (train) to Heerlen city center on the event day?

For Nijmegen I would say 30-70 %.

Personal opinion – Vis 2:

- Which part of the visualization could be improved, why & how?

  The percentages like 50-50 %, 30-70 % are not clear.

- What grade would you give to the ease of use of the visualization? (1-10)

  6.5

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  6.5

Visualization 3

1. How does the event affect the visitor's average duration of stay in Heerlen city center?

   So from the bar chart I see that on the event day the number of visitors increase and also the average staying time increase.

2. When is the busiest time in Heerlen city center on the regular day and on the event day?

   So, on regular day the busiest time is at 1 pm and on the event day is at 7 pm.

3. What is your estimation of the increase in total number of visitors during the evenings (after 18.00 hrs) on the event day compared to the regular day?

   The total increase on the event day after 18.00 I would say maybe it doubled the regular day.

4. Which province(s) show(s) the biggest difference in average staying time, between the regular day and the event day?

   From the maps, maybe it is Noord Holland.

5. What differences do you notice based on the relation between the distances to Heerlen city center on average staying time on the regular day? And on the event day?

   Now you can see that on the regular day, the closer they traveled the longer they stay. On the event day it is difference.

6. How many visitors come from Noord Holland and stay in Heerlen city center only for 1 hour?

   You can see that the white color means less than 3 hours, but only for 1 hour I can't answer it.

Personal opinion – Vis 3:

- Which parts of the visualization could be improved, why & how?

  - I think you can make the hour axis on both bar charts.
  - I think you can put background color instead of just white color.

- What grade would you give to the ease of use of the visualization? (1-10)

  7.5

- What grade would you give to the aesthetic quality of the visualization? (1-10)

  6.5

Appendix C – Qualitative study analyses
I. Parts that gain the attention from the participant (structure/layout)

Parts that gain the attention from the participant (structure/layout)							
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
Visualization 1	P1 started to look at the bar chart which was on left side of the visualization, then continue to read the whole visualization from the left to right (as expected). **P1 didn't put much attention on the legend which was located on the top right of the visualization.**	P2 started to look at the bar chart which was on left side of the visualization, then continue to read the whole visualization from the left to right (as expected). **P2 didn't put much attention on the "information" part.**	P3 started to look at the bar chart which was on left side of the visualization, then continue to read the whole visualization from the left to right (as expected).	P4 started to look at the title on the top left side to understand what the visualization is about (as expected). Then P4 went to the bar chart which was on left side of the visualization and continued to read hole visualization from the left to right (as expected).	**P5 started to look at the right side of the visualization which was the legend. Then P5 went straight to the maps. So P5 put much attention on the right side of the visualizations and the maps.**	P6 started to look at the title on the top left side to understand what the visualization is about (as expected). Then P6 went to the bar chart which was on left side of the visualization and continued to read hole visualization from the left to right (as expected).	P7 started to look at the bar chart which was on left side of the visualization, then continue to read the whole visualization from the left to right (as expected).
Visualization 2	**P1 started to look at the right side of the visualization which was the legend. Then P1 went straight to the maps. So P1 put much attention on the right side of the visualizations and the maps.**	P2 started to look at the bar chart which was on left side of the visualization, then continue to read the whole visualization from the left to right (as expected). **Still, P2 didn't put much attention on the "information" part.**	P3 started to look at the title on top left of the visualization, then P3 went to the bar chart which was on left side of the visualization (as expected).	P4 started to look at the title on the top left side to understand what the visualization is about (as expected). Then P4 went to the legend of the map which was on the right side of the visualization. Then P4 went to the maps.	**P5 started to look at the right side of the visualization which was the legend. Then P5 went straight to the maps. So P5 put much attention on the right side of the visualizations and the maps.**	P6 started to look at the title on the top left side to understand what the visualization is about (as expected). Then P6 went to the bar chart which was on left side of the visualization and continued to read hole visualization from the left to right (as expected).	P7 started to look at the bar chart which was on left side of the visualization, then continue to read the whole visualization from the left to right (as expected).
Visualization 3	P1 started to to look at the bar chart which was on left side of the visualization, then continue to read the whole visualization from the left to right (as expected).	P2 started to look at the bar chart which was on left side of the visualization, then continue to read the whole visualization from the left to right (as expected).	P3 started to look at the bar chart which was on left side of the visualization, then continue to read the whole visualization from the left to right (as expected).	P4 started to look at the title on the top left side to understand what the visualization is about (as expected). Then P4 went to the bar chart which was on left side of the visualization and continued to read hole visualization from the left to right (as expected).	P5 started to look at the title on the top left side to understand what the visualization is about (as expected). Then P5 went to the bar chart which was on left side of the visualization and continued to read hole visualization from the left to right (as expected).	P6 started to look at the bar chart which was on left side of the visualization, then continue to read the whole visualization from the left to right (as expected).	P7 started to look at the bar chart which was on left side of the visualization, then continue to read the whole visualization from the left to right (as expected).

II. Understandability of the visualizations (e.g the visualization is clear or confusing)

Understandability of the visualizations (e.g the visualization is clear or confusing)							
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
Visualization 1	Overall, the visualization was clear for P1. **P1 prefered the information on the maps was in the actual numbers of people or visitors to Heerlen than the percentage.**	Overall, the visualization was clear for P2. **P2 prefered the information on the maps was in the actual numbers of people or visitors to Heerlen than the percentage.**	**P3 had a diffulty to understand the visualization because the amount of information and texts in one screen. P3 felt the visualization had too much information and texts.**	Overall,the visualization was clear for P4.	Overall,the visualization was clear for P5.	Overall,the visualization was clear for P6. **But P6 thought maybe the visualization had too much texts.**	Overall,the visualization was clear for P7.
Visualization 2	Overall, the visualization was clear for P1 **but P1 got a little bit confused with the legend "60-40" or "70-30".**	Overall, the visualization was clear for P2 **but P2 thought that the legend for the maps was a little bit difficult.**	**P3 had a diffulty to understand the visualization because the amount of information and texts in one screen. P3 felt the visualization had too much information and texts.**	**P4 said that the visualization was a little bit difficult to understand. P4 needed to look at the visualization and understood it for quite long.**	Overall, the visualization was clear for P5 **but P5 got a little bit confused with the legend "60-40" or "70-30".**	Overall P6 understood the visualization.	Overall, the visualization was clear for P7 **but P7 got a little bit confused with the legend "60-40" or "70-30".**
Visualization 3	Overall P1 understood the visualization **but P1 thought the bar chart is a little bit unclear. It was not clear whether a bar represent an hour or a duration.**	Overall P2 understood the visualization. **P2 thought that it would be better if the hours (y axis) were in both charts.**	Overall P3 understood the visualization. **P3 thought that the visualization was much clearer than the other two because P3 can see the information in one look. P3 also thought that it would be better if the hours (y axis) were in both charts.**	Overall P4 understood the visualization.	Overall P5 understood the visualization.	Overall P6 understood the visualization.	Overall P7 understood the visualization.

III. Common strategy or personal preference in answering the questions

Common strategy or personal preference in answering the questions							
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
Visualization 1	P1 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions.	P2 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions.	P3 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions. **But P3 had a diffulty to understand the visualization because the amount of information and texts in one screen. P3 also didn't put much attention on the "visitor's frequency information"**	P4 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions. **But P4 didn't put much attention on the "visitor's frequency information"**	P5 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions, **but most of the time he used the information from the maps to answer the questions.**	P6 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions. **But P6 didn't put much attention on the "visitor's frequency information".**	P7 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions. **But P7 didn't put much attention on the "visitor's frequency information". P7 also put more attention to the maps.**
Visualization 2	**P1 didn't use the bar chart at all, P1 went straight to the maps.**	P2 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions.	P3 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions. **But P3 had a diffulty to understand the visualization because the amount of information and texts in one screen.**	**P4 didn't use the bar chart at all, P1 went straight to the maps.**	**P5 didn't use the bar chart at all, P5 went straight to the maps.**	P6 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions.	P7 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions.
Visualization 3	P1 used all the elements (e.g. bar charts, maps, legend, information,titles, etc.) on the visualization while answering the questions.	P2 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions.	P3 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions.	P4 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions.	P5 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions.	P6 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions.	P7 used all the elements (e.g. bar charts, maps, legend, information, titles, etc.) on the visualization while answering the questions.

IV. Understandability of legends (e.g. interpretation, colors, position, etc.)

Understandability of legends (e.g. interpretation, colors, position, etc.)							
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
Visualization 1	**At the beginning, P1 didn't notice the meaning of the colors on the legend for the maps. P1 thought that the darker the color, the high number of visitors. But then on the 5th question, P1 noticed the meaning of the colors which was the percentage of inhabitants.**	P2 interpreted and used the legend as expected. **P2 would like to have more different colors.**	P3 interpreted and used the legend as expected. **P3 thought the colors on the legend for the maps were not distinctive.**	P4 interpreted and used the legend as expected.	P5 interpreted and used the legend as expected.	P5 interpreted and used the legend as expected.	P7 interpreted and used the legend as expected. **P7 would like to have more different colors.**
Visualization 2	P1 interpreted the legend for the maps as expected, **but P1 would liked to have only one or whole number on the legend for the maps (e.g. only 60 % train instead of 40-60 %).**	P2 used the legend for the maps as expected **but P2 interpreted it incorrectly, because on the 4th question P2 didn't answer the question correctly. P2 thought that the legend for the maps was a little bit difficult and P2 would like to have only one or whole number on the legend (e.g. only 60 % train instead of 40-60 %). P2 also said that the green colors were almost the same.**	P3 used the legend for the maps as expected **but P3 interpreted it incorrectly , because on the 4th question P3 didn't answer the question correctly.**	P4 used the legend for the maps as expected **but P4 interpreted it incorrectly, because on the 2nd, 3rd, and 4th questions P4 didn't answer the question correctly. P4 thought that the legend could be better. The percentage range was quiet strange for P4.**	P5 used the legend for the maps as expected **but P5 interpreted it incorrectly, because on the 3rd and 4th questions P5 didn't answer the questions correctly. P5 also said that P5 saw the maps, at first P5 interpreted the color differently. P5 thought it was continuous instead of categorical.**	P6 used the legend for the maps as expected **but P6 interpreted it incorrectly, because on the 3rd and 4th questions P6 didn't answer the questions correctly. P6 also suggested to use one number of percentage for the legend.**	P7 used the legend for the maps as expected **but P7 interpreted it incorrectly , because on the 4th question P7 didn't answer the question correctly.**
Visualization 3	P1 interpreted and used the legend as expected.	P2 interpreted and used the legend as expected.	P3 interpreted and used the legend as expected.	P4 interpreted and used the legend as expected. **P4 added that the white color on the legend for the maps should be revised. It was not only less than 3 hours, but it could be also no data. So it could be changed into "<3 or no data".**	P5 interpreted and used the legend as expected.	P6 interpreted and used the legend as expected.	P7 interpreted and used the legend as expected.

V. Understandability/legibility of texts (e.g titles, information, explanation, labels, etc.)

Understandability/legibility of texts (e.g titles, information, explanation, labels, etc.)							
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
Visualization 1	P1 abled to understand the texts on the visualization.	P2 abled to understand the texts on the visualization, **but p2 thought that the title on the maps were too long. Also the information about the "visitor's frequency" was a little bit too far. Maybe it would be better if it closer to the bar chart.**	**P3 felt the visualization had too much information and texts.**	P4 thought that the information was already good and clear. **P4 suggested that the "visitor's frequency" was a little bit too far. Maybe it would be better if it closer to the bar chart.**	P5 abled to understand the texts on the visualization.	P6 abled to understand the texts on the visualization. **But P6 thought maybe the visualization had too much texts.**	P7 abled to understand the texts on the visualization.
Visualization 2	P1 abled to understand the texts on the visualization.	P2 abled to understand the texts on the visualization, **but p2 thought that the title on the maps were too long.**	**P3 felt the visualization had too much information and texts.**	P4 abled to understand the texts on the visualization.	P5 abled to understand the texts on the visualization.	P6 abled to understand the texts on the visualization.	P7 abled to understand the texts on the visualization.
Visualization 3	P1 abled to understand the texts on the visualization.	P2 abled to understand the texts on the visualization.	P3 abled to understand the texts on the visualization.	P4 abled to understand the texts on the visualization.	P5 abled to understand the texts on the visualization.	P6 abled to understand the texts on the visualization.	P7 abled to understand the texts on the visualization.

VI. Comments

	Comments						
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
Visualization 1	- The bar chart was already clear. **- The information on the maps, P1 prefered to have the real number of visitors instead of the percentage.**	**- The title on the maps were too long** **- The position of the information about the visitor's frequency is a little bit too far. P2 suggested to move it closer to the bar chart.** **- The information on the maps, P2 prefered to have the real number of visitors instead of the percentage.**	**- The visualization had too much texts.** - Too much difference information on one sheet. - The colors on the map were not distinctive.	- The information was already good and clear **- The position of the information about the visitor's frequency is a little bit too far. P4 suggested to move it closer to the bar chart.**	- For the maps, maybe it would be good if one map showed the whole Netherlands and on the corner another small map (zoom in map) that showed only the area of the event with the surrounding cities.	- **Maybe the visualization had too much texts**, but overall it was quite clear for P6.	- Overall P7 thought it was a clear sheet, but maybe it could be a little more attractive by using an image . - P7 thought it was better to work with more different colors because it was a little bit hard to differentiate it.
Visualization 2	- P1 wanted to have more contrast between colors. **- P1 prefered to have a whole number for the percentages on the legend.**	**- The title on the maps were too long** **- P2 prefered to have a whole number for the percentages on the legend.**	**- Too much information and texts on one sheet.**	- The visualization was a little bit difficult to understand. **- The legend could be better. The percentage range was quiet strange.**	**- The legend could be better because at first P5 interpreted the color differently.**	- **For the legend, P6 suggested to use one number of percentage like:** Car Train 80% 70% 60% 50% 60% 70% 80% or Prefer car 0% 30% 50% 80% 100% Prefer train 0% 30% 50% 80% 100%	**-The percentages on the legend like 50-50 %, 30-70 % were not clear.**
Visualization 3	- The bar chart was unclear whether a bar represent an hour or a duration.	**- P2 thought that it would be better if the hours (y axis) were in both charts.** - The map was already clear.	- The visualization was much clearer than the other two visualizations. **- P3 thought that it would be better if the hours (y axis) were in both charts.**	- The white color on the legend for the maps should be revised. It was not only less than 3 hours, but it could be also no data. So it could be changed into "<3 or no data". - For the maps, P4 thought that maybe it could helpful if the area that have average staying time of "<3 hours or no data" had also a color. Because P4 saw the surrounding areas (e.g. Belgium, Germany, and the sea) were also white. So P4 suggested to differentiate it.	- The visualization was easier to understand than the other two.	- The visualization was clear for P6, so nothing to change.	- **P7 thought that it would be better if the hours (y axis) were in both charts.** - P7 thought to put background color instead of only white color.

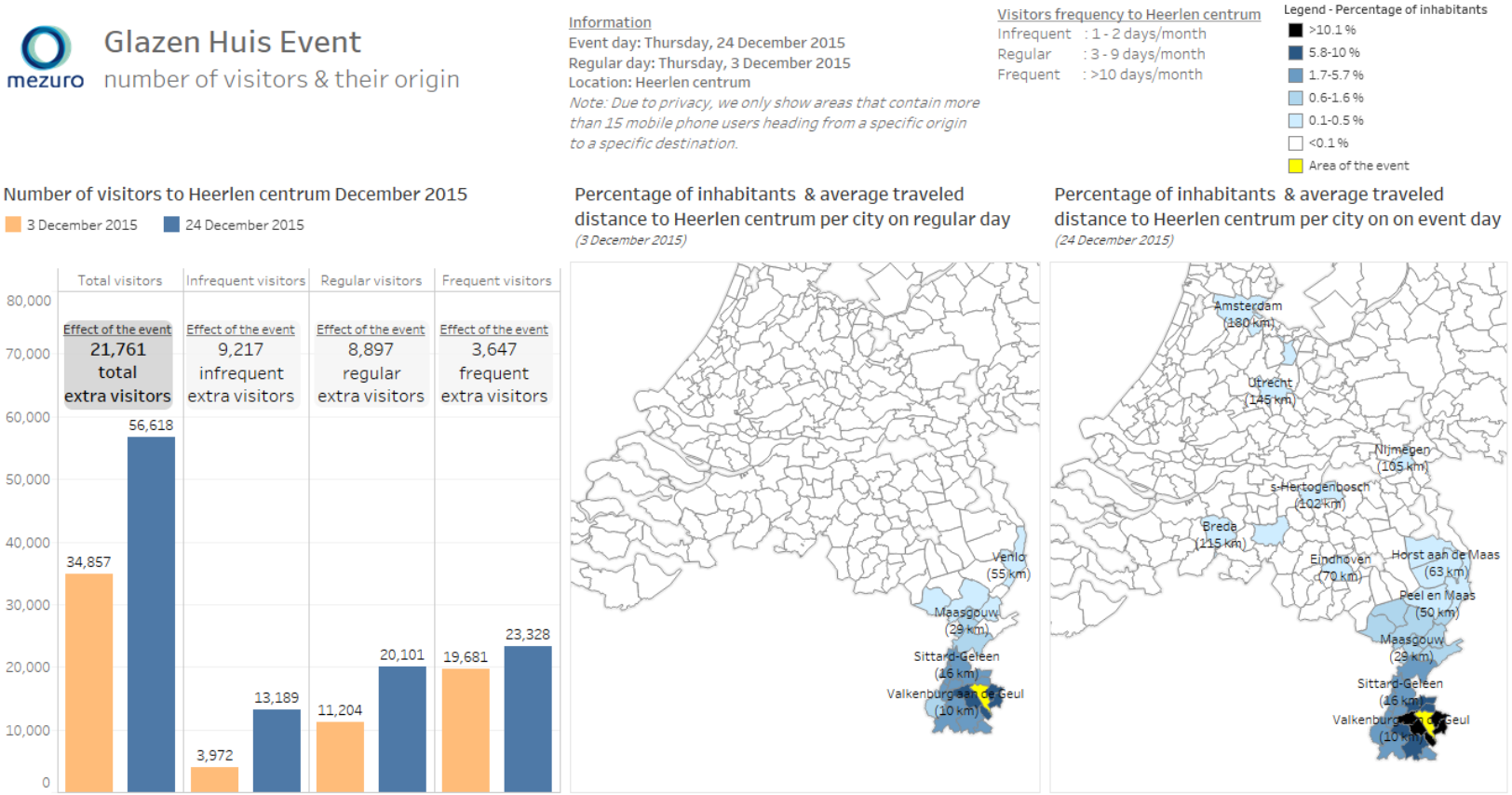
VI. Ease of use and aesthetic quality

Ease of use (grade 1-10)								
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**Average**
Visualization 1	7	**8**	6	7	**8**	**9**	**7.5**	7.5
Visualization 2	6	**9**	5	5	**7**	**9**	6.5	6.8
Visualization 3	6	**9**	**8**	7	**8**	**9**	7.5	7.8

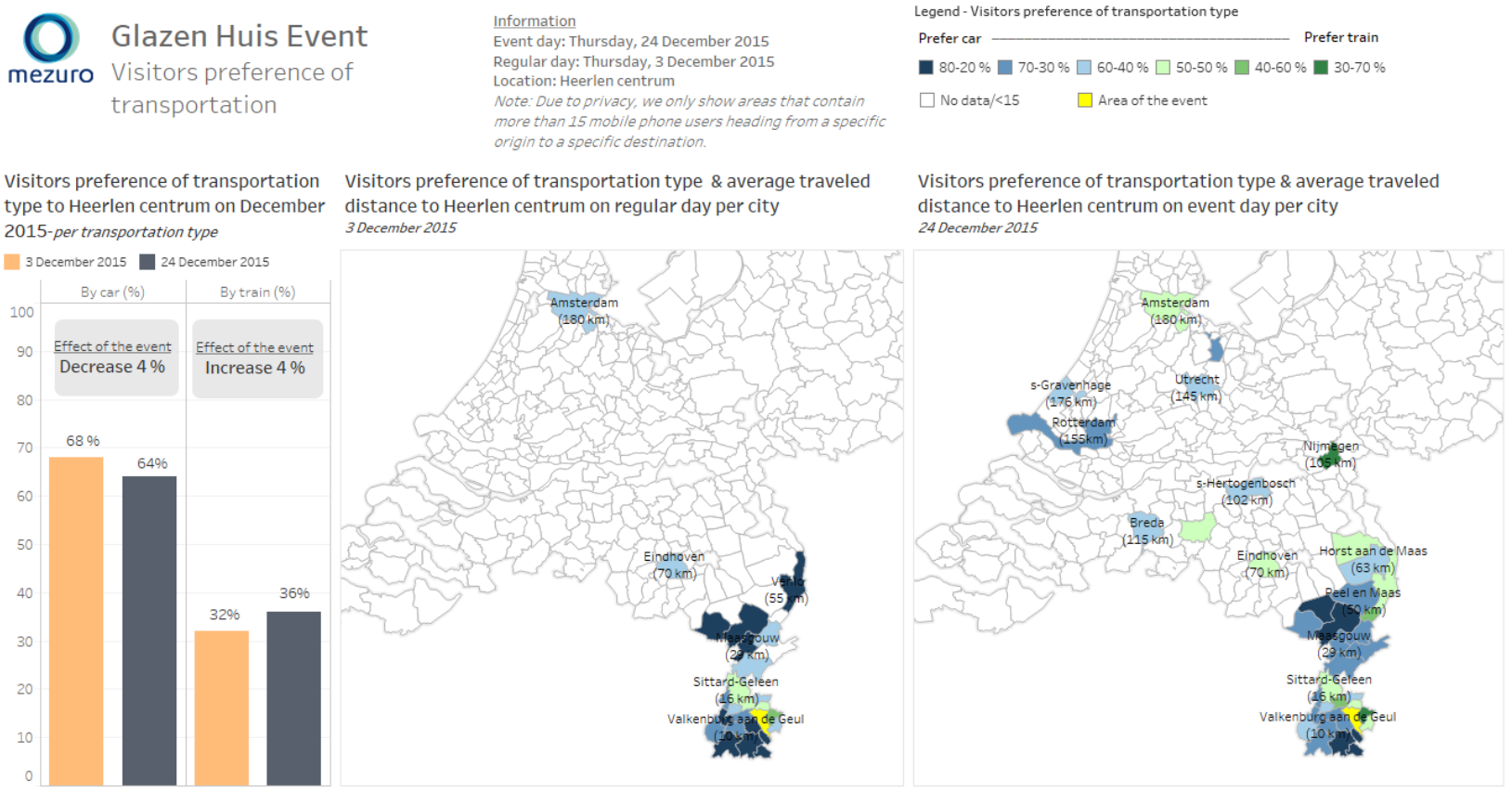
Aesthetic quality (grade 1-10)								
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**Average**
Visualization 1	7	**7.5**	6	**8**	**8**	**9**	6.5	7.4
Visualization 2	7	**9**	6	**8**	**8**	**9**	6.5	7.6
Visualization 3	7	**9**	**8**	7	**8**	**9**	6.5	7.8

Appendix D – Visualizations

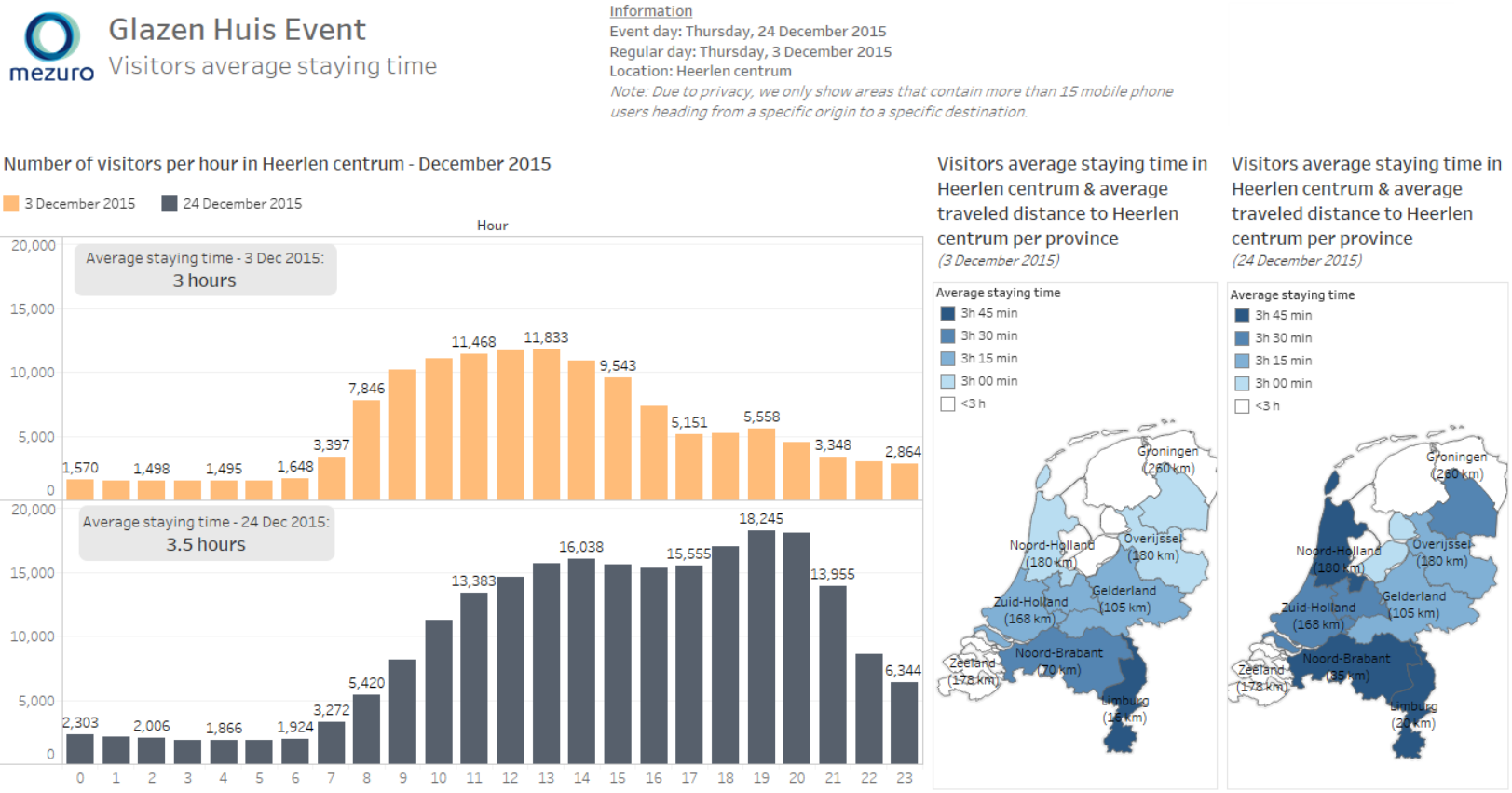
Visualization page 1



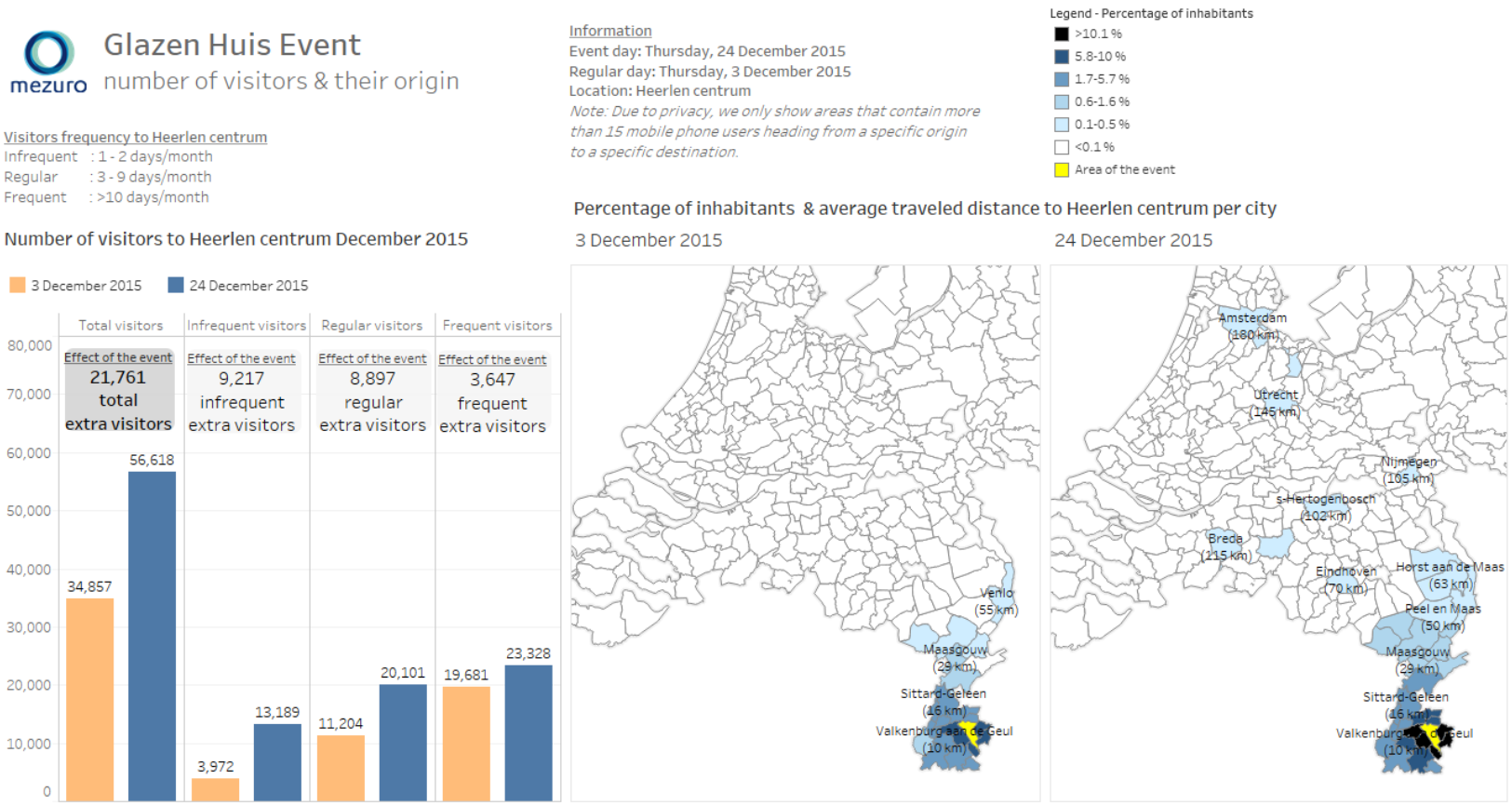
Visualization page 2



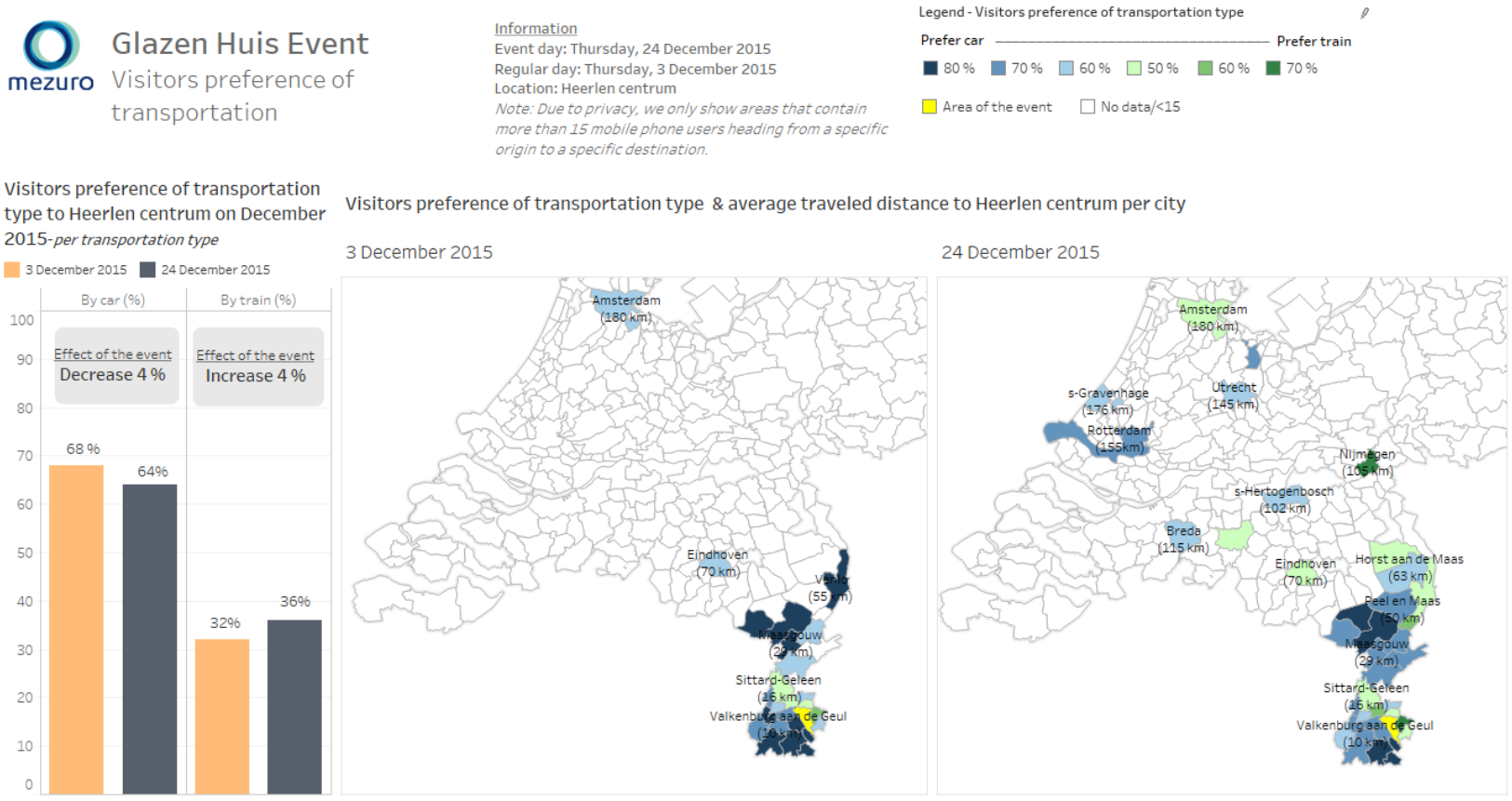
Visualization page 3



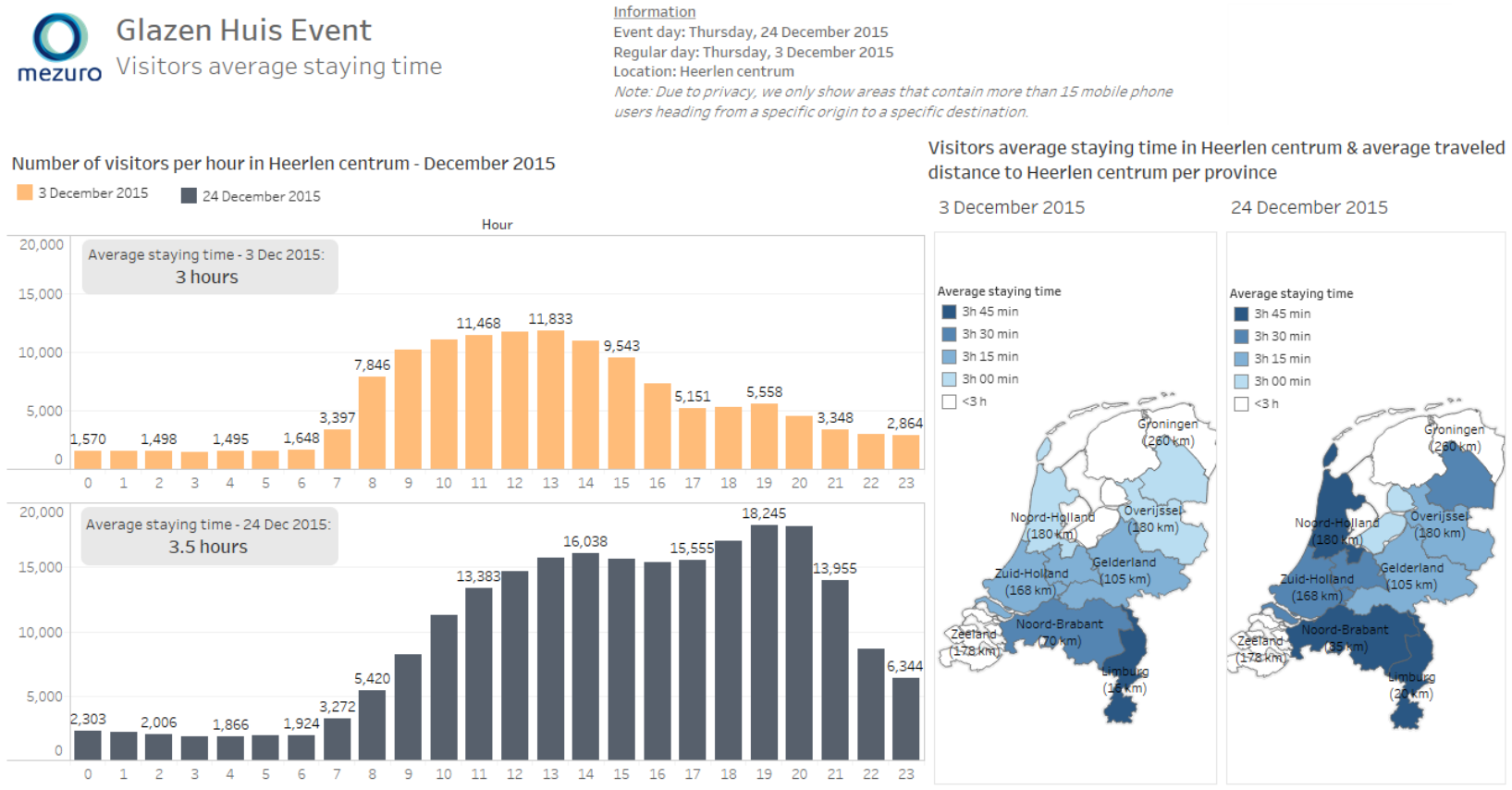
Improved visualization page 1



Improved visualization page 2



Improved visualization page 3



Appendix E - Quantitative studies (experiment protocol)

Welcome to the online survey for my master thesis.

Dear participant, thank you for agreeing to take part in this survey for my master thesis project. My name is Zaki, and I'm doing my thesis project on information visualization at Mezuro. The goal is to determine effective methods for visualizing mobility information. I am going to ask you questions using visualizations that we are working on. There are 3 visualizations and 16 questions in this survey.

The session will take approximately 15-20 minutes. Your responses are used only for this research project and will be anonymous and confidential. If you have any questions or concerns, please contact me through my email: z.a.koesoemahardja@students.uu.nl. Before we start the survey, I would like to ask you a few questions for background information:

1. Age: <20 21-30 31-40 41-50 51-60 >60

2. Gender: Male Female No answer

3. Country of residence: Netherlands Other

4. The highest level of education you have completed:

- Primary school
- Secondary school
- Vocational education (MBO)
- Higher education (HBO or university)

5. Current job or occupation?

- A student
  Field of education:
- An employee
  Work role:
- Currently not employed or studying

6. What best describes the type of organization you work for?

- Education
- Profit
- Nonprofit (e.g. religious, arts, social assistance, etc.)
- Government
- Other

7. How often do you use an information visualization? (5 likert scales)

- Never
- 1-2 times/semester
- 1-2 times/month
- 1-3 times/week
- >3 times/week

8. How important are information visualizations in your daily job? (5 likert scales)

- Not important
- Less important
- Neutral
- Important
- Hardly important

If you would like to receive the results of this survey, please fill in your email address:

We now start the questions about the visualizations using a scenario.

Scenario:
Imagine that you work as an event planner in Heerlen, and an event "Serious Request (Glazen Huis)" was held in your city in December 2015. A company in the Netherlands analyzed the event and made a report. From the report, you want to know the effect of the event on your city like the number of visitors, type of transportation they used to your city, etc. Using the visualizations, please answer the questions in this survey.

<u>Visualization 1</u>

1. How did the event affect the total number of visitors to Heerlen city center? The increase of visitors is:
   Answer: A. 56,618 B. 23,328 **C. 21,761**

2. What differences do you notice between the event day and the regular day with regard to the visitor's origin?

   Answer:
   A. On the event day, the origin of the visitors was further away than on the regular day.
   B. On the event day, the origin of the visitors was closer than on the regular day.
   C. On the event day, the origin of the visitors was the same as the regular day.

3. Which type of visitors has the largest relative increase between the event day and the regular day?

Answer:
A. Frequent visitors
B. Infrequent visitors
C. Regular visitors

4. What is the difference in the longest traveled distance on the event day compared to the regular day? *(You can ignore areas with a percentage of inhabitants <0.1)*
   Answer: A. 180 km **B. 125 km** C. 55 km

5. What is your estimation of the average traveled distance from cities that had 5.8-10 % of inhabitants in Heerlen city center on the event day?
   Answer: **A. 10 km** B. 30 km C. 50 km

6. How many people came from Amsterdam to Heerlen city center on the event day?

   Answer:
   A. 0.1–0.5%
   B. 0.6-1.6 %
   C. It is not possible to answer from the visualization

Personal opinion – Vis 1:

- What grade would you give to the ease of use of the visualization? (1-10)

<u>Visualization 2</u>

1. How did the event affect the visitors preference of public transportation (train) to Heerlen city center?

   Answer:
   A. The event increased the visitors preference of public transportation (train) to Heerlen city center
   B. The event decreased the visitors preference of public transportation (train) to Heerlen city center
   C. There was no effect on the visitors preference of public transportation (train) to Heerlen city center

2. Which cities showed an increase in the visitors preference of public transportation (train) to Heerlen city center, on the event day, compared to the regular day?

Answer:
A. Amsterdam and Eindhoven
B. Amsterdam and Nijmegen
C. Rotterdam and Utrecht

3. Which type of transportation did visitors prefer who travelled further than 150 km to Heerlen city center on the event day?
   Answer: A. Most preferred the train **B. Most preferred their car** C. Both

4. What percentage of the people who came from Nijmegen used public transportation (train) to Heerlen city center on the event day?
   Answer: A. 30 % **B. 70 %** C. 30-70 %

Personal opinion – Vis 2:

- What grade would you give to the ease of use of the visualization? (1-10)

<u>Visualization 3</u>

1. How did the event affect the visitors' average staying time in Heerlen city center?

   Answer:
   A. The event increased the average staying time
   B. The event decreased the average staying time
   C. The event had no effect on average staying time

2. When was the busiest time in Heerlen city center on the regular day and on the event day?

   Answer:
   A. Regular day: 08.00-10.00 and event day: 12.00–14.00
   B. Regular day: 11.00-13.00 and event day: 12.00-14.00
   C. Regular day: 11.00–13.00 and event day: 18.00-20.00

3. What is your estimation of the increase in the total number of visitors at 19.00 on the event day compared to the regular day?
   Answer: A. 8,000 **B. 13,000** C. 20,000

4. Which province shows the biggest difference in average staying time, between the regular day and the event day?
   Answer: A. Noord-Brabant **B. Noord-Holland** C. Zuid-Holland

5. Which answer describes the relation between the average traveled distances and average staying time on the regular day and on the event day, the best?

   Answer:
   A. Regular day: the further the traveled distance, the shorter the staying time. Event day: it was diverse
   B. Regular day: the further the traveled distance, the shorter the staying time. Event day: the shorter the traveled distance, the longer the staying time.
   C. No differences between the regular day and the event day

6. How many visitors came from Noord Holland and stayed in Heerlen city center only for 1 hour?

   Answer:
   A. 10,000
   B. 15,000
   C. It is not possible to answer from the visualization

Personal opinion – Vis 3:

What grade would you give to the ease of use of the visualization? (1-10)

Final part – personal preference

Improved visualization 1

Report 1

From the visualization and the report (numbers and tables) above, which one do you prefer and why?

Answer: A. The visualization B. The report C. No preference

Why:

Appendix F – Quantitative studies analyses and SPSS Output

This appendix presents the output of SPSS on which we based Section 5.2.

Outliers

These are the number of outliers that we removed or changed per version, per page, and per question:

	Version 1	**Version 2**	**Version 3**	**Changed/removed**
Page 1				
Q1	0	1	0	Changed
Q2	0	0	1	Changed
Q3	0	1	1	Changed
Q4	0	0	1	Removed
Q5	0	1	1	Removed
Q6	1	0	1	Changed
Subtotal	1	3	5	
Page 2				
Q1	1	2	1	Version 1: Removed Version 2 & 3: Changed
Q2	0	1	0	Removed
Q3	1	0	0	Removed
Q4	0	1	0	Changed
Subtotal	2	4	1	
Page 3				
Q1	2	0	0	Changed
Q2	1	0	1	Changed
Q3	1	0	0	Removed
Q4	0	1	0	Removed
Q5	1	1	1	Changed
Q6	0	1	1	Version 2: Removed Version 3: Changed
Subtotal	5	3	3	
TOTAL	8	10	9	

Normal distribution (Kolmogorov Smirnov test)

SPSS results – Version 1, 2 & 3

Page 1 (Q1-Q6)

Tests of Normality

	Survey version	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
LastClick_Q1_Vis1	Version 1	.184	28	.017	.877	28	.003
	Version 2	.175	22	.078	.884	22	.014
	Version 3	.112	19	.200*	.957	19	.513
LastClick_Q2_Vis1	Version 1	.317	28	.000	.514	28	.000
	Version 2	.255	22	.001	.777	22	.000
	Version 3	.173	19	.139	.923	19	.131
LastClick_Q3_Vis1	Version 1	.169	28	.040	.870	28	.002
	Version 2	.266	22	.000	.711	22	.000
	Version 3	.284	19	.000	.763	19	.000
LastClick_Q4_Vis1	Version 1	.238	28	.000	.649	28	.000
	Version 2	.154	22	.190	.864	22	.006
	Version 3	.092	19	.200*	.953	19	.442
LastClick_Q5_Vis1	Version 1	.177	28	.024	.868	28	.002
	Version 2	.217	22	.008	.813	22	.001
	Version 3	.179	19	.110	.859	19	.009
LastClick_Q6_Vis1	Version 1	.187	28	.013	.814	28	.000
	Version 2	.130	22	.200*	.924	22	.094
	Version 3	.180	19	.109	.878	19	.020

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Page 2 (Q1-Q4)

Tests of Normality

	Survey version	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
LastClick_Q1_Vis2	Version 1	.123	26	.200*	.961	26	.405
	Version 2	.219	22	.007	.806	22	.001
	Version 3	.295	21	.000	.701	21	.000
LastClick_Q2_Vis2	Version 1	.106	26	.200*	.977	26	.801
	Version 2	.103	22	.200*	.953	22	.359
	Version 3	.172	21	.104	.942	21	.244
LastClick_Q3_Vis2	Version 1	.245	26	.000	.566	26	.000
	Version 2	.101	22	.200*	.959	22	.470
	Version 3	.097	21	.200*	.972	21	.767
LastClick_Q4_Vis2	Version 1	.210	26	.005	.791	26	.000
	Version 2	.228	22	.004	.719	22	.000
	Version 3	.176	21	.089	.879	21	.014

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Page 3 (Q1-Q6)

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Survey version	Statistic	df	Sig.	Statistic	df	Sig.
LastClick_Q1_Vis3	Version 1	.172	27	.039	.814	27	.000
	Version 2	.214	22	.010	.741	22	.000
	Version 3	.258	21	.001	.670	21	.000
LastClick_Q2_Vis3	Version 1	.170	27	.043	.906	27	.018
	Version 2	.110	22	.200*	.965	22	.598
	Version 3	.121	21	.200*	.959	21	.487
LastClick_Q3_Vis3	Version 1	.250	27	.000	.738	27	.000
	Version 2	.277	22	.000	.602	22	.000
	Version 3	.149	21	.200*	.962	21	.552
LastClick_Q4_Vis3	Version 1	.203	27	.006	.890	27	.008
	Version 2	.128	22	.200*	.908	22	.043
	Version 3	.150	21	.200*	.955	21	.428
LastClick_Q5_Vis3	Version 1	.177	27	.030	.842	27	.001
	Version 2	.124	22	.200*	.925	22	.096
	Version 3	.106	21	.200*	.977	21	.875
LastClick_Q6_Vis3	Version 1	.193	27	.011	.838	27	.001
	Version 2	.173	22	.086	.797	22	.000
	Version 3	.300	21	.000	.551	21	.000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

<u>One way ANOVA - Between the versions (in general)</u>

ANOVA

Total_Avg_LastClick

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3221.497	2	1610.748	5.180	.008
Within Groups	21455.386	69	310.948		
Total	24676.883	71			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Total_Avg_LastClick

Tukey HSD

		Mean Difference (I-J)			95% Confidence Interval	
(I) Survey version	(J) Survey version		Std. Error	Sig.	Lower Bound	Upper Bound
Version 1	Version 2	-3.89554	4.96233	.713	-15.7819	7.9908
	Version 3	-16.02877*	5.09041	.007	-28.2219	-3.8356
Version 2	Version 1	3.89554	4.96233	.713	-7.9908	15.7819
	Version 3	-12.13323	5.32226	.065	-24.8817	.6153
Version 3	Version 1	16.02877*	5.09041	.007	3.8356	28.2219
	Version 2	12.13323	5.32226	.065	-.6153	24.8817

*. The mean difference is significant at the 0.05 level.

Parametric test (One way ANOVA)

Page 2 – Q2

ANOVA

LastClick_Q2_Vis2

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5242.300	2	2621.150	4.154	.020
Within Groups	42908.106	68	631.002		
Total	48150.406	70			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: LastClick_Q2_Vis2

Tukey HSD

(I) Survey version	(J) Survey version	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Version 1	Version 2	-10.70888	7.15665	.299	-27.8568	6.4391
	Version 3	-20.80362*	7.25145	.015	-38.1787	-3.4285
Version 2	Version 1	10.70888	7.15665	.299	-6.4391	27.8568
	Version 3	-10.09474	7.66352	.391	-28.4572	8.2677
Version 3	Version 1	20.80362*	7.25145	.015	3.4285	38.1787
	Version 2	10.09474	7.66352	.391	-8.2677	28.4572

*. The mean difference is significant at the 0.05 level.

Non parametric test (Kruskal Wallis test)

Page 1

Test Statistics[a,b]

	LastClick_Q1_Vis1	LastClick_Q2_Vis1	LastClick_Q3_Vis1	LastClick_Q4_Vis1	LastClick_Q5_Vis1	LastClick_Q6_Vis1
Chi-Square	7.500	22.925	2.387	22.244	1.465	.315
df	2	2	2	2	2	2
Asymp. Sig.	.024	.000	.303	.000	.481	.854

a. Kruskal Wallis Test

b. Grouping Variable: Survey version

Page 2 - Q1, Q3 & Q4

Test Statistics[a,b]

	LastClick_Q1_Vis2	LastClick_Q3_Vis2	LastClick_Q4_Vis2
Chi-Square	.099	2.806	.597
df	2	2	2
Asymp. Sig.	.952	.246	.742

a. Kruskal Wallis Test

b. Grouping Variable: Survey version

Page 3

Test Statistics[a,b]

	LastClick_Q1_Vis3	LastClick_Q2_Vis3	LastClick_Q3_Vis3	LastClick_Q4_Vis3	LastClick_Q5_Vis3	LastClick_Q6_Vis3
Chi-Square	2.863	11.379	1.907	.216	4.985	.763
df	2	2	2	2	2	2
Asymp. Sig.	.239	.003	.385	.898	.083	.683

a. Kruskal Wallis Test

b. Grouping Variable: Survey version

Post Hoc Kruskal Wallis test (Mann Whitney test)

Page 1 – Q1

Test 1

Test Statistics[a]

	LastClick_Q1_Vis1
Mann-Whitney U	276.000
Wilcoxon W	682.000
Z	-.871
Asymp. Sig. (2-tailed)	.384

a. Grouping Variable: Survey version

Test 2

Test Statistics[a]

	LastClick_Q1_Vis1
Mann-Whitney U	160.000
Wilcoxon W	566.000
Z	-2.707
Asymp. Sig. (2-tailed)	.007

a. Grouping Variable: Survey version

Test 3

Test Statistics[a]

	LastClick_Q1_Vis1
Mann-Whitney U	166.000
Wilcoxon W	442.000
Z	-1.774
Asymp. Sig. (2-tailed)	.076

a. Grouping Variable: Survey version

Page 1 – Q2

Test 1

Test Statistics[a]

	LastClick_Q2_Vis1
Mann-Whitney U	181.000
Wilcoxon W	587.000
Z	-2.669
Asymp. Sig. (2-tailed)	.008

a. Grouping Variable: Survey version

Test 2

Test Statistics[a]

	LastClick_Q2_Vis1
Mann-Whitney U	73.000
Wilcoxon W	479.000
Z	-4.465
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Survey version

Test 3

Test Statistics[a]

	LastClick_Q2_Vis1
Mann-Whitney U	128.000
Wilcoxon W	404.000
Z	-2.667
Asymp. Sig. (2-tailed)	.008

a. Grouping Variable: Survey version

Page 1 – Q4

Test 1

Test Statistics[a]

	LastClick_Q4_Vis1
Mann-Whitney U	304.000
Wilcoxon W	710.000
Z	-.341
Asymp. Sig. (2-tailed)	.733

a. Grouping Variable: Survey version

Test 2

Test Statistics[a]

	LastClick_Q4_Vis1
Mann-Whitney U	71.000
Wilcoxon W	477.000
Z	-4.371
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Survey version

Test 3

Test Statistics[a]

	LastClick_Q4_Vis1
Mann-Whitney U	71.000
Wilcoxon W	347.000
Z	-3.872
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Survey version

Page 3 – Q2

Test 1

Test Statistics[a]

	LastClick_Q2_Vis3
Mann-Whitney U	275.000
Wilcoxon W	551.000
Z	-.890
Asymp. Sig. (2-tailed)	.374

a. Grouping Variable: Survey version

Test 2

Test Statistics[a]

	LastClick_Q2_Vis3
Mann-Whitney U	154.500
Wilcoxon W	560.500
Z	-2.818
Asymp. Sig. (2-tailed)	.005

a. Grouping Variable: Survey version

Test 3

Test Statistics[a]

	LastClick_Q2_Vis3
Mann-Whitney U	115.000
Wilcoxon W	391.000
Z	-2.972
Asymp. Sig. (2-tailed)	.003

a. Grouping Variable: Survey version

Kruskal Wallis test (accuracy)

Page 1

Test Statistics[a,b]

	Answer_Q1_Vis1	Answer_Q2_Vis1	Answer_Q3_Vis1	Answer_Q4_Vis1	Answer_Q5_Vis1	Answer_Q6_Vis1
Chi-Square	3.849	1.739	3.021	2.238	7.685	3.314
df	2	2	2	2	2	2
Asymp. Sig.	.146	.419	.221	.327	.021	.191

a. Kruskal Wallis Test

b. Grouping Variable: Survey version

Page 2

Test Statistics[a,b]

	Answer_Q1_Vis2	Answer_Q2_Vis2	Answer_Q3_Vis2	Answer_Q4_Vis2
Chi-Square	2.222	.071	1.913	19.024
df	2	2	2	2
Asymp. Sig.	.329	.965	.384	.000

a. Kruskal Wallis Test

b. Grouping Variable: Survey version

Page 3

Test Statistics[a,b]

	Answer_Q1_Vis3	Answer_Q2_Vis3	Answer_Q3_Vis3	Answer_Q4_Vis3	Answer_Q5_Vis3	Answer_Q6_Vis3
Chi-Square	2.672	2.384	.807	4.202	8.960	3.849
df	2	2	2	2	2	2
Asymp. Sig.	.263	.304	.668	.122	.011	.146

a. Kruskal Wallis Test

b. Grouping Variable: Survey version

Mann Whitney test (accuracy)

Page 1 – Q5

Test 1

Mann-Whitney Test

Ranks

	Survey version	N	Mean Rank	Sum of Ranks
Answer_Q5_Vis1	Version 1	28	23.32	653.00
	Version 2	23	29.26	673.00
	Total	51		

Test Statistics[a]

	Answer_Q5_ Vis1
Mann-Whitney U	247.000
Wilcoxon W	653.000
Z	-2.150
Asymp. Sig. (2-tailed)	.032

a. Grouping Variable: Survey version

Test 2

Mann-Whitney Test

Ranks

	Survey version	N	Mean Rank	Sum of Ranks
Answer_Q5_Vis1	Version 1	28	25.25	707.00
	Version 3	21	24.67	518.00
	Total	49		

Test Statistics[a]

	Answer_Q5_ Vis1
Mann-Whitney U	287.000
Wilcoxon W	518.000
Z	-.341
Asymp. Sig. (2-tailed)	.733

a. Grouping Variable: Survey version

Test 3

Mann-Whitney Test

Ranks

	Survey version	N	Mean Rank	Sum of Ranks
Answer_Q5_Vis1	Version 2	23	25.20	579.50
	Version 3	21	19.55	410.50
	Total	44		

Test Statistics[a]

	Answer_Q5_Vis1
Mann-Whitney U	179.500
Wilcoxon W	410.500
Z	-2.180
Asymp. Sig. (2-tailed)	.029

a. Grouping Variable: Survey version

Page 2 – Q4

Test 1

Mann-Whitney Test

Ranks

	Survey version	N	Mean Rank	Sum of Ranks
Answer_Q4_Vis2	Version 1	28	31.63	885.50
	Version 2	23	19.15	440.50
	Total	51		

Test Statistics[a]

	Answer_Q4_Vis2
Mann-Whitney U	164.500
Wilcoxon W	440.500
Z	-3.448
Asymp. Sig. (2-tailed)	.001

a. Grouping Variable: Survey version

Test 2

Mann-Whitney Test

Ranks

	Survey version	N	Mean Rank	Sum of Ranks
Answer_Q4_Vis2	Version 1	28	30.88	864.50
	Version 3	21	17.17	360.50
	Total	49		

Test Statistics[a]

	Answer_Q4_ Vis2
Mann-Whitney U	129.500
Wilcoxon W	360.500
Z	-3.838
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Survey version

Test 3

Mann-Whitney Test

Ranks

	Survey version	N	Mean Rank	Sum of Ranks
Answer_Q4_Vis2	Version 2	23	23.24	534.50
	Version 3	21	21.69	455.50
	Total	44		

Test Statistics[a]

	Answer_Q4_ Vis2
Mann-Whitney U	224.500
Wilcoxon W	455.500
Z	-.550
Asymp. Sig. (2-tailed)	.582

a. Grouping Variable: Survey version

Page 3 – Q5

Test 1

Mann-Whitney Test

Ranks

	Survey version	N	Mean Rank	Sum of Ranks
Answer_Q5_Vis3	Version 1	28	25.61	717.00
	Version 2	23	26.48	609.00
	Total	51		

Test Statistics[a]

	Answer_Q5_Vis3
Mann-Whitney U	311.000
Wilcoxon W	717.000
Z	-.249
Asymp. Sig. (2-tailed)	.804

a. Grouping Variable: Survey version

Test 2

Mann-Whitney Test

Ranks

	Survey version	N	Mean Rank	Sum of Ranks
Answer_Q5_Vis3	Version 1	28	20.75	581.00
	Version 3	21	30.67	644.00
	Total	49		

Test Statistics[a]

	Answer_Q5_Vis3
Mann-Whitney U	175.000
Wilcoxon W	581.000
Z	-2.781
Asymp. Sig. (2-tailed)	.005

a. Grouping Variable: Survey version

Test 3

Mann-Whitney Test

Ranks

	Survey version	N	Mean Rank	Sum of Ranks
Answer_Q5_Vis3	Version 2	23	18.61	428.00
	Version 3	21	26.76	562.00
	Total	44		

Test Statistics[a]

	Answer_Q5_ Vis3
Mann-Whitney U	152.000
Wilcoxon W	428.000
Z	-2.451
Asymp. Sig. (2-tailed)	.014

a. Grouping Variable: Survey version