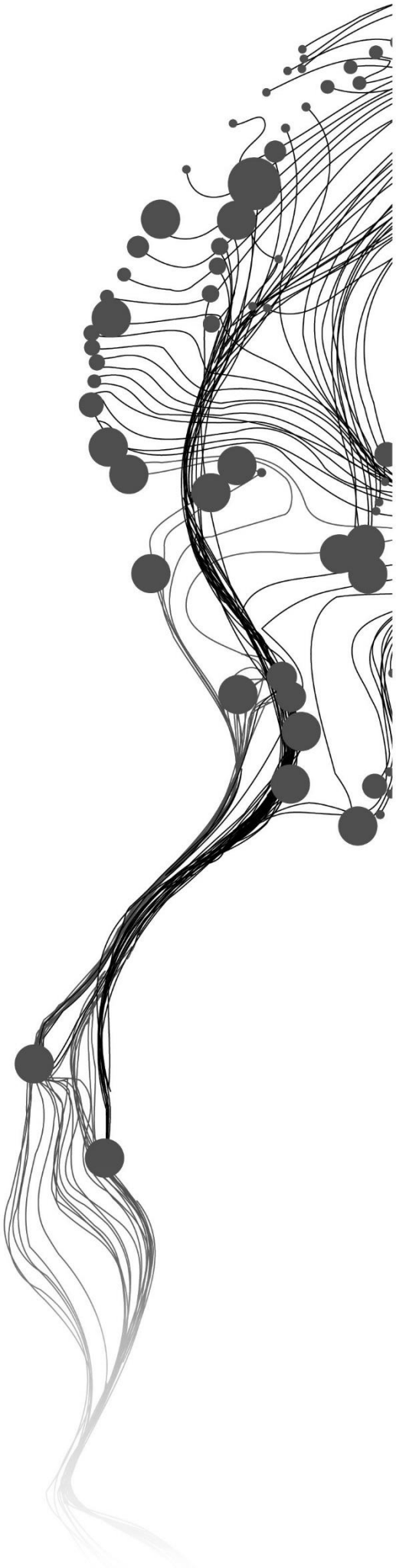
VISUALIZATION OF TRENDS IN OD MATRIX DATA

ILLUSTRATED BY UNITED NATIONS REFUGEE DATA

MARAL MALEKI

February, 2016

SUPERVISORS:
Prof. Dr. M. J. Kraak
Drs. B. J. Kobben



VISUALIZATION OF TRENDS IN OD MATRIX DATA

ILLUSTRATED BY UNITED NATIONS REFUGEE DATA

MARAL MALEKI
Enschede, The Netherlands, February, 2016

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

SUPERVISORS:
Prof. Dr. M. J. Kraak
Drs. B. J. Kobben

THESIS ASSESSMENT BOARD:
Prof. Dr. A.A. Voinov (Chair)
Dr. J. Von Engelhardt (External Examiner, University of Twente)

ABSTRACT

Every day in media, there is a news about refugee crisis. Controlling the movement of refugees and detecting their destination is a challenge for governments and organizations such as United Nations especially when the volume of data is large. The data representing these movements can be organized as an Origin and Destination. The OD-data which usually stores as an OD-matrix inherently has high complexity. Although OD-matrix itself is not suitable to produce an overview of the data, in this research we present an alternative way to illustrate OD-matrix data that facilitates the interpretation. Data visualization is one way of revealing patterns in the OD-matrix data. However, without a suitable simplification method the result will often be a clutter visualization which does not assist the user to obtain an insight of the data. Therefore, there is a need of reducing the big OD-matrix data while keeping the typical information, which, in this case, are the countries with high population of refugee. The method used in this research was a statistically defined threshold to find the outliers from the big OD-matrix. We developed a conceptual design based on user requirements and further adapted based on a focus group discussion. This resulted in a visualization environment with linked stacked bar graph, OD-Matrix, map with flow lines, and world matrix. Then, the application further evaluated using a heuristic evaluation by experts to be sure the application works well based on the user requirements.

Keywords: *OD-matrix, OD data, Refugees, Visualization, Threshold, Heuristic Evaluation, Spatiotemporal data, Focus Group Discussion, United Nations*

ACKNOWLEDGEMENTS

First and foremost, my deep gratitude to my first supervisor *Prof. Dr. M. J. Kraak*, thank you for your support, guidance throughout my thesis. It is my honour to be your student during this journey.

My special thanks to my second supervisor *Drs. B. J. Kobben*, thank you for all your encouragement and support. By your help I realized that I have family here also who always support you when you are sick.

My success dedicated to my parents who I never be here as I am without their love, motivation, and support.

I would like to thank my brother who never leaves me alone and reminds me I have assurance in my life.

My deep thanks and appreciation to my friend *Felipe De Carvalho Diniz* for his unconditionally help and precious guidance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Motivation and problem statement

According to United Nations Department of Economic and Social Affairs - UNDESA (2013), 232 million people, 3.2 percent of world's population lived outside of their country of origin. Most of them migrated to find the better economic opportunities and to have a better lifestyle. Others, like refugees, are forced to flee due to a natural disaster, violence, politics, war, and other reasons. Because of the increasing number of refugees and population displacements over time, controlling the movement of refugees and detecting their final destination is a challenge for organizations such as the United Nations.

Specifying where the refugees are from, where they are going, what kind of refugees they are, and the reason for their migration is essential for the United Nations in order to devise a careful plan for the refugee's comfort and protection (Migration Policy Institute, 2002). This information about refugee's data helps United Nations to get an insight of data especially when the volume of data is large.

Data that represent the movement, by referencing the origin and destination of, for example, money, energy, people, flight, and others, in a geographic space is called Origin-Destination data (OD-data). Organizing and exploring OD-data leads to a better understanding of the flow dynamics, with both changes over time and at varying volume (Boyandin, 2013).

OD-data are often stored in an OD-matrix in which the rows show the origins and columns the destinations. Understanding and analysing the OD-matrix data is a basis for decision makers to implement suitable policies, in particular for migration of refugees. However, analysing a large quantity of data in an OD-matrix in order to generate understandable information is difficult and presents challenges (Boyandin, 2013). The OD-matrix itself is not suitable to produce an overview of the data. Therefore, in this study we attempt to find an alternative way to present OD-matrix data that facilitates the interpretation. Data visualization is a possible way to accomplish this.

Data visualization is a form of data representation that helps users to see and understand the concepts of data which are not obvious and evident for them before (Kraak & Ormeling, 2013). One option of visually representing OD-matrix data to generate is to map the data, but this will quickly result in cluttered visualizations if the matrix is large (Kraak, 2011). Thus, finding efficient ways to represent and visualize a large amount of data is the main challenge for decision-makers and organizations.

With this number of difficulties, the motivation of this work lies in the need to design a visualization that provides an overview of OD-matrix flow over time. The visualization assists the user in understanding the trend in the data, such as refugees' migration, movement of refugees from the origin to destination, and its changes over time. The vision is that a visualization should provide interactivity of flow dynamics over time, be readable and represent the summary of OD-matrix data. The application is implemented based on open source technologies, which can increase the outreach of the application.

We evaluated the visualization based on user-centred design, considering the effectiveness, efficiency and satisfaction of the design. This approach for evaluation has the advantage of extracting the user's needs and ensure that those needs are being effectively represented in the visualization (Nielsen, 1993). The evaluation method was based qualitative measure, using techniques such as focus group and heuristic evaluation (Kraak & Ormeling, 2013).

1.2. Research identification

1.2.1. Research objectives

Main Objective:
The main objective of this research is to design, implement and evaluate an interactive map based representation, to visualize spatiotemporal OD-matrix data. Achieving this main objective requires the sub-objectives:

1. To determine and retrieve spatiotemporal trends from OD-matrix data.
2. To design and implement a visualization method for multi-temporal OD-matrix trends.
3. To evaluate the usability of the design (prototype).

1.2.2. Research questions

Questions related to Objective 1:

1. How to decide what are the trends in OD-matrix data?
2. How to extract the trends from the OD-matrix data?

Questions related to Objective 2:

1. What visualization representation(s) are suitable to visualize OD-matrix data?
2. Which representation(s) can show the trends or summary of these data?
3. Which questions should the visualization answer based on the user requirements?

Questions related to Objective 3:

1. How to set up and conduct usability study based on user tasks to evaluate the new representation(s)?

1.3. Innovation

The novelty of this research is to design, implement and evaluate a visualization capable of getting an insight into large spatiotemporal OD-matrix data. This was achieved by using a map-based representation capable of showing the summary of the data. As a case study data from refugees' migration dataset from United Nations is used.

1.4. Project setup

1.4.1. Method adopted

At the first stage of this research, we conducted literature based on the previous works and concentrated on the topics, OD-data and Visualization. The selected use cases were analysed and explained. The next stage was the conceptual design of a map based on OD-data visualization. In this step, the needs for conceptual design was taken from the literature review, OD-data and case study. Then, the representation that can show the summary or trends of OD-data was designed.

The next step was studying Data Driven Documents Library for JavaScript (d3.js). D3.js is an open library and tool for producing dynamic, interactive data visualization (Michael Bostock, Ogievetsky, & Heer, 2011).

Parallel to learn D3 a requirement analysis was executed resulting in questions that the user might need answers for. For instance, if user wants to know the number of refugees and their origin in a particular year, the question that would be suitable for that is "from which countries the most refugees come from in that particular year?".

We downloaded the data from United Nations Refugee Agency, framed to OD-Matrix structure and then organized in a database. Consideralbe data correction was required. A conceptual design was elaborated and implemented.

The next step was evaluating the usability of the proposed design. The evaluation of usability in general consists of assessing three main aspects: effectiveness, efficiency, and satisfaction. Effectiveness shows the quality and correctness; efficiency indicates the learning time and accuracy, and satisfaction illustrates the user preference and satisfaction of the work (Frøkjær, Hertzum, & Hornbæk, 2000). These three aspects are the goal of each visualization. To be able to reach this goal evaluating the visualization was done based on qualitative measurements. A pool of voluntary users was selected to evaluate the visualization. For the first qualitative evaluation a focus group was used. In this technique, a group of users was invited for a group discussion to find suitable and appropriate result for designing the visualization. In this discussion participants gave their idea and preferences about visualization and what interactivities among the visualizations are understandable for the users. The result of the discussion was used for creating the final conceptual design. This technique is cheap, easy and quick to see the reality from the user perspective. The heuristic evaluation as a final measure was used for evaluating the implemented design. In this method a small set of experts were asked to evaluate the application and find its strengths and weaknesses.

In the last stage, the conclusion was made. In this step, a summary of results and findings was shown and clarified. The visualization is applicable for organizations like United Nations, governments, and general people who would like to know about refugees' population and their changes over time which is based on the outcome of the evaluation. In addition, recommendation for future work is added. Figure 1-1 presents the flowchart of a project plan for the thesis.

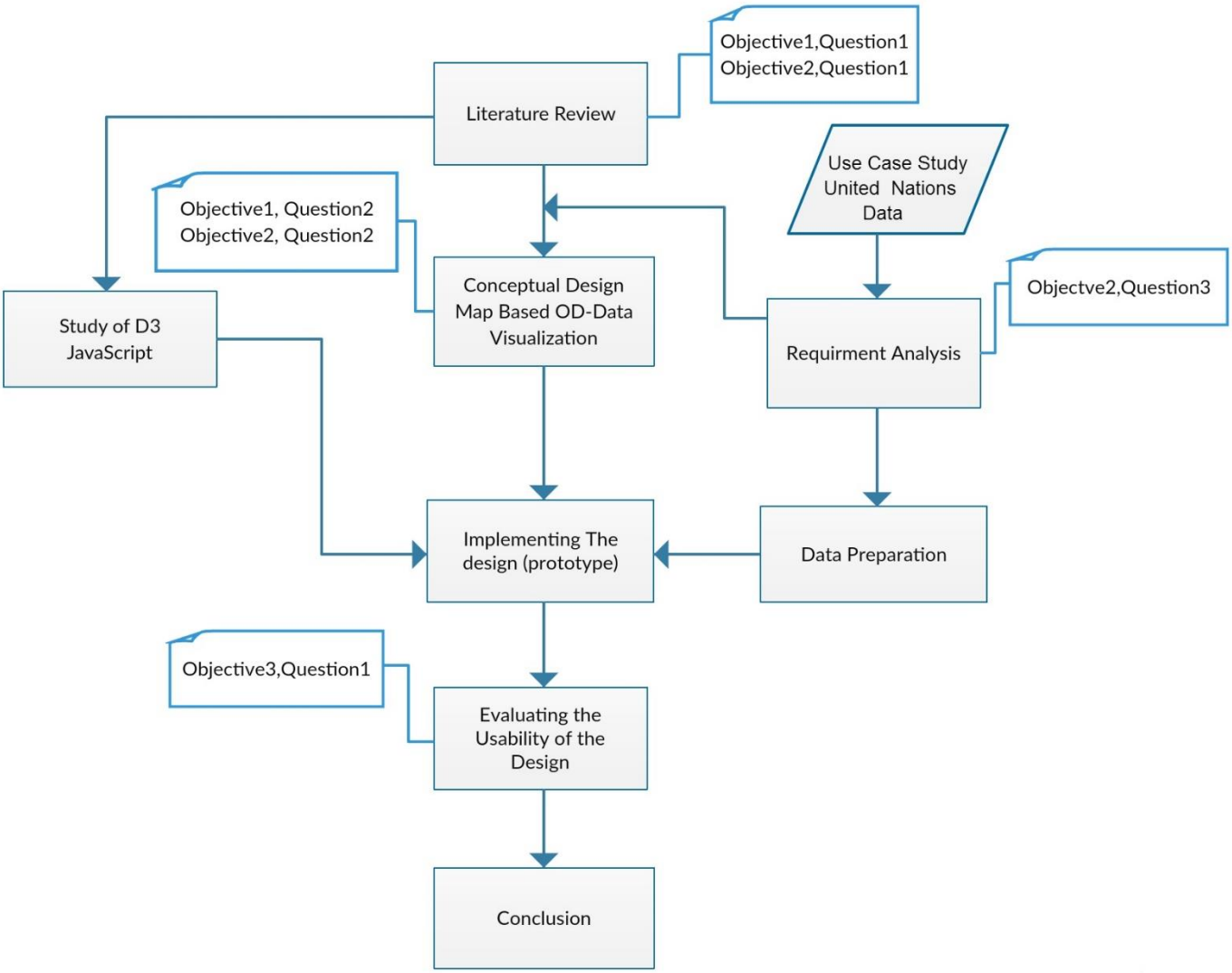


Figure 1-1 Project plan and methodology

1.5. Thesis structure

This thesis consists of eight chapters:

Chapter 1 – explains about motivation and problem statement, research objectives and questions, innovation, and project setup.

Chapter 2 – illustrates the refugees' problems and challenges, international concerns about the refugees. This includes general information about refugees and importance of consideration of refugees.

Chapter 3 – is about theoretical framework, it gives information about some of the basic theories of data visualization.

Chapter 4 – provides literature review.

Chapter 5 – describes the conceptual and implementation of the design. This chapter provides a sketch of the design and implementing the prototype. Moreover, it provides the focus group discussion.

Chapter 6 – evaluates the usability of the design based on usability testing methods, discusses the selected evaluation method, the result of evaluation, and consequence of the visualization.

Chapter 7 – provides conclusions and answers to the research questions in addition to recommendation of future work.

2. PROBLEMS OF REFUGEES

2.1. Introduction

This chapter first, explains general concepts about refugees, followed by section discussing refugees' problems, and the last section describes the international concerns about refugees. This chapter gives essential information about current problems that refugees are facing with which make them leave their country.

2.2. General information about refugees

Nowadays, there are many reasons that makes people leave their country in this hope that they can find better life and future. The core definition of refugees is based on 1951 United Nation Convention Relating to the Status of Refugees and its 1967 Protocol convention. They define refugees as a person who: "*owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality, and is unable and unwilling to avail himself of the protection of that country.*" (Wilkinson et al., 2001). Based on United Nations high commissioner Antonio Guterres: "*Today challenges are interconnected and complex. Population growth, urbanization, climate change, water scarcity and food and energy insecurity are exacerbating conflict and combining in other ways that oblige people to flee their countries*" ("Global Issues at the United Nations.," n.d.). One cannot easily understand the difficulties in refugee's life, where they can have a constant fear of being threatened, uncertain of their future, and having to leave the life in their country behind. Therefore, understanding who are refugees and their problems, which is explained in more details in this section, help governments, organizations and countries for the protections, both legally and physically.

2.3. Refugees and their problems

There are many brutalities against refugees over the years which shows the increasing trends. They are subjected to economic exploitation and social discrimination. Strong syndicates exploited the refugees for their political and strategic advantages. Although United Nation High Commissioner for Refugees (UNHCR) has done many works to assist refugees so far, still there are not a considerable improvement in the refugees' problem. The long term program with structural changes is needed to be able to improve refugees' situations. Based on the latest study, from 26 million refugees almost 70 percent in the world are women and children. There are many cases of women that have been raped, abused, and offended by people who had to support and protect them (Hindwan, 2010). Another problem that refugees are facing with is about the difficulty of speaking the language of the destination country especially when they have to fill the forms. Refugees indicated that is hard for them to fill the forms that they do not know the content of the forms. Raising children with new and unfamiliar culture is another problem for refugees. Parents often find their children quite different and strange with their culture. Children tend to learn a new language faster than their parents and use this to their advantage that this case is more popular among teenagers. Having a stable job also is another challenge for refugees. Although refugees are happy to take any job that be offered to them, finding a job and promoting in that is tough for refugees. For refugees who are educated and had a job is frustrating for them that they cannot find job in their area of study. Security housing is a problem of refugees as well. Finding a safe and affordable house is difficult with low-paying jobs; this makes refugees live together in a noisy and stressful environment that hardly allow them to rest and study (Nunez, 2014). There are also more problems and challenges that refugees have to deal

with like cultural barriers, transportation and accessing the services that are out of the scope of this research.

2.4. International concern about refugees

The international concern about refugees based on previous sections for governments and global bodies is "Legal protection". This is a major concern and enduring of international policy issues. The responsibility to protect the citizens is obligatory for governments to protect them from harms and troubles; but, when they are not able to protect their citizens, this responsibility goes to the international community. For organizations like UNHCR, the main command is the protection of refugees and has done many works for protecting the refugees. Other organizations are international humanitarian law and national law concerning refugees' legal protection.

3. THEORETICAL FRAMEWORK

3.1. Introduction

There are many different data type taxonomies for visualization; this chapter presents a taxonomy of task that support visualizing the Origin-Destination data. This taxonomy helps to formulate the questions in an overall level that can be asked by users. Following sections describes how integrating these taxonomies lead to creating spatiotemporal questions.

3.2. Peuquet triad scheme

Peuquet (1994) proposed a triad framework for spatiotemporal data which has three basic component: "Attribute", "Location" and "Time" and each of these components refers and links to specific questions; Attribute "What", Spatial "Where" and Time "When". According to Peuquet there are types of questions in this framework that are the following:

Table 3-1: Types of questions according to Peuquet (1994)

Type of Question	Description
When + where ⟶ what	Describe the objects or set of objects "what" that are present at a given locations or set of locations "where" at a particular time or set of times "when".
When + what ⟶ where	Describe location or set of locations "where" occupied by a given object or set of objects "what" at a given time or set of times "when".
Where + what ⟶ when	Describe the times or set of times "when" that a given object or set of objects "what" occupied a given location or set of locations "where".

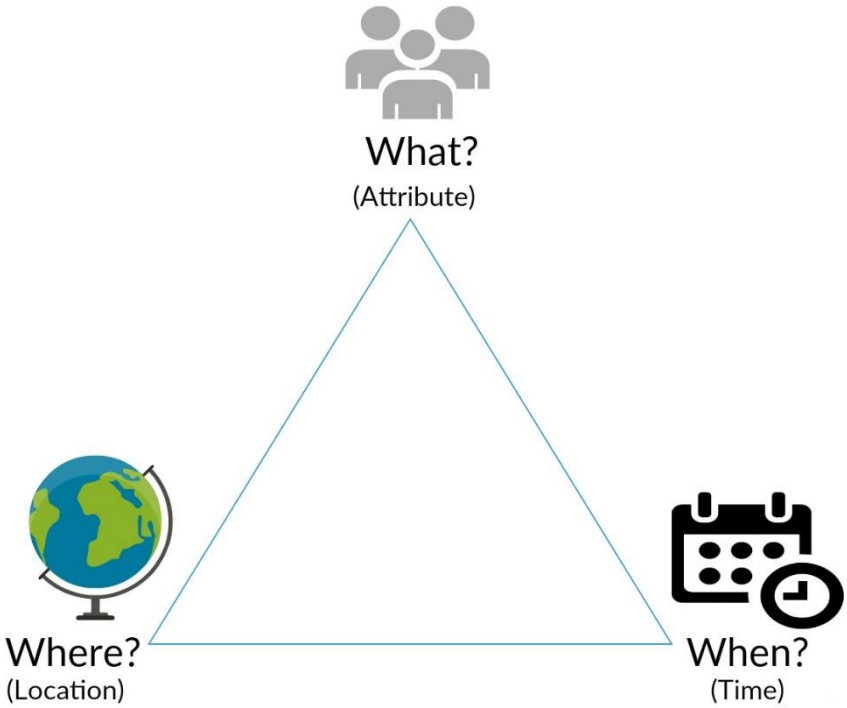


Figure 3-1: Basic concepts of triad framework

Figure 3-1 represents the fundamental components of triad framework that helps analysts to be able to formulate the basic questions that can apply to understand the nature of the data. Table 3-1 gives information regarding how the queries can be created. For instance, we want to describe the time or set of times when the object "What" or "Where" are given; then, the question of this type can be "When Afghanistan has the highest number of refugees?" Hence, Peuquet' question type leads to build the questions that can be asked by the user.

3.3. Shneiderman's Visual Information-Seeking Mantra

Shneiderman (1996) proposed a useful starting point for advanced graphic for user interface which called Visual Information-Seeking Mantra, these graphical user interface has seven tasks that are as follow:

Table 3-2: Graphical user interface tasks according to Shneiderman (1996)

Task	Description
Overview	Gain an overview of the entire collection
Zoom	Zoom in on items of interest
Filter	Filter out uninteresting items
Details-on-demand	Select an item or group and get details when needed
Relate	View relationships among items
History	Keep a history of actions to support undo, replay, and progressive refinement
Extract	Allow extraction of sub-collections and the query parameters

According to the Table 3-2 the seven data types of Shneiderman's data type taxonomy are: one, two and three-dimensional, temporal and multi-dimensional data, tree and network data, which each question can be sorted in one data type taxonomy. In another word, Shneiderman's Visual Information-Seeking Mantra assists how the created questions in each level can be shown in the interface. This helps the user to see an overview of the entire collection, if the user is intrested in an overview level of questions.

3.4. Bertin's reading levels

Bertin (1983) proposed three reading levels each with their question type. These three reading levels (Elementary, Intermediate, and Overall) for task systematization are distinguished as follow:

- **Elementary:** Questions concerning one single element
- **Intermediate:** Questions concerning a group of elements
- **Overall:** Questions concerning all the elements

Bertin's reading levels provide different reading levels elementary, intermediate, and overall. In overall level the question should concentrate on all the elements. For instance, "What is a trend of Afghanistan during 12 years?" this question is in overall level that considers the period of 12 years.

Based on the Bertin's reading levels of questions and Shneiderman's Visual Information-Seeking Mantra, analysts can create questions on an overall level to have the overview of the entire data collection. Bertin's reading levels and Shneiderman's Visual Information-Seeking Mantra together can lead analysts to create questions on different levels. For instance, if the Elementary level of questions wants to be created based on Bertin's reading levels, among the seven tasks of Schneiderman's Visual Information-Seeking Mantra the Details-on-demand has to be chosen. For the Intermediate level of questions, Zoom/Filter has

to be used, and for having an Overall level of questions, Overview task has to be applied. Figure 3-2 indicates this relation at a glance. Moreover, Peuquet triad scheme leads to having queries for spatiotemporal data with his three basic components; in this research "where" indicates the location of refugees, "what" refugees' attributes and "when" multiple years.

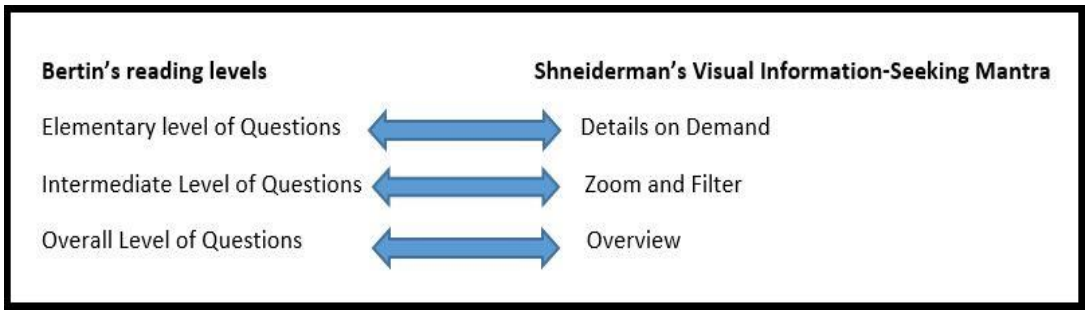


Figure 3-2: Relation of Bertin's reading levels and Shneiderman's Visual Information-Seeking Mantra

In conclusion, the theories integrate together to be able to create questions that maybe asked by the user. Peuquet, Shneiderman, and Bertin are theoretical framework to create spatiotemporal questions, how the questions should be represented in the interface, and how spatiotemporal questions should be created in each level respectively.

The next section is about creating the questions in different levels that user might be looking for. Although the concern of this research is formulating the questions in an overall view level, there has been provided different level of questions in an elementary and intermediate levels.

3.5. Data visualization and questions

Visualization tools help analysts to find the answers to these questions. These questions can be structured based on the need and concerns that analysts want to take from the data. Below we exemplify questions at different levels based on Shneiderman, Bertin, and Peuquet theories the questions at different levels have been provided below:

Elementary level:

1- How many people from country X registered as refugees in year Y?
2- What was the population of refugees in country X in year Y?
3- How many refugees from country X lived in country Z in year Y?

Intermediate Level:

1- Which country had the most number of refugees in year Y?
2- In which year country X had the highest refugee population?
3- How many refugees resided in continent X in year Y?
4- How many refugees from continent X resided in continent Z in year Y?

Overall Level (overview):

1- Did the refugees population increase or decrease during the 12 years?
2- In which Continent(s) did the refugee population increase?
3- Where are the major changes of the refugees into a continent X during the 12 years?

3.6. Threshold

Big OD matrices, especially when it is a spatiotemporal matrix, contains a large amount of data that cannot be visualized in an effective way. A method is required to reduce the amount of data in the OD matrix while keeping the most important aspects which in our case are the countries which had a high population of refugees. This can be realized by the application of a threshold. The method used to establish a threshold for this research explained beneath, and is applied on the attribute data.

In general data can be distributed into three categories to be able to describe the shape of the data: Symmetric, Left Skewed, and Right Skewed. According to the histogram of the data it can be determined if data is skewed or not. Figure 3-3 shows that when data is balanced and symmetric, mean is equal to the median which is indicated in the histogram. Moreover, the box plot and whisker of the symmetric data is shown in Figure 3-3. In boxplot can be seen that median which is second quartile is at the middle of first and third quartile which is a characteristic of the symmetric data (Siyavula, 2012). The refugee data is very skewed as can be seen in Figure 3.5.

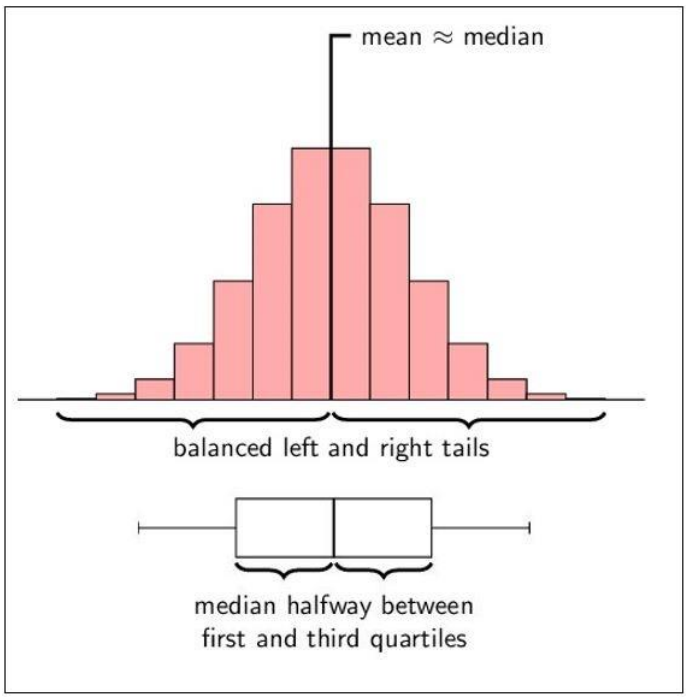


Figure 3-3: Indicates the Symmetric distribution histogram with Boxplot according to Siyavula (2012).

When the data is not symmetric, it is categorized in skewed data which can be left or right skewed.

Figure 3-4 shows the histogram of the right skewed data. It can be understood that mean is greater than median and also the tail of the histogram is longer to the right side in comparison with left side. For the boxplot for right skewed data median is closer to the first quartile than the third quartile. This process is the opposite for the histogram and box plot of the left skewed dataset.

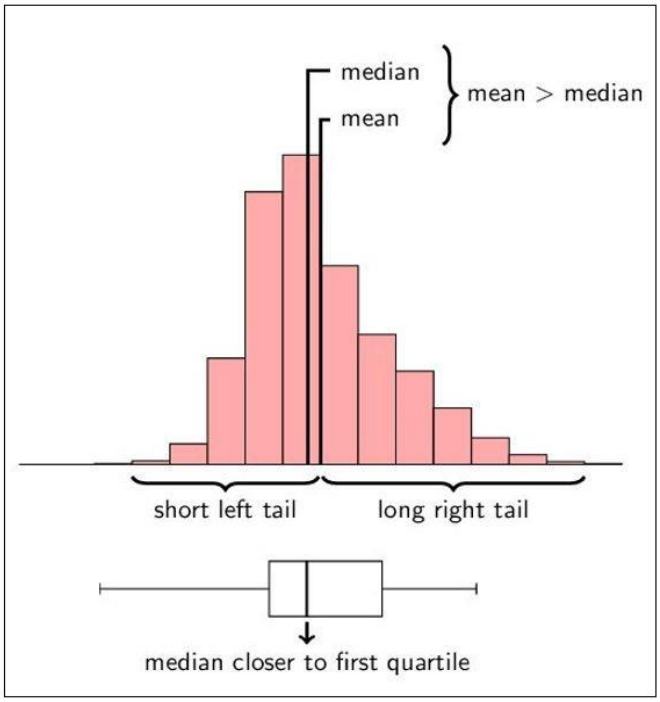


Figure 3-4: Illustrates the histogram and the Boxplot for Right Skewed data according to Siyavula (2012)

The idea of selection the countries from the big matrix is based on finding the countries that are outliers in the dataset. The outliers, in this case, are those countries that have a large number of refugees in a particular year, matching the trends which we wish to investigate. A common method of calculating outliers is based on the top whisker of boxplots, by the following formula:

$$Q3+ 1.5*IQR \quad (3.1)$$

$$IQR= Q3-Q1 \quad (3.2)$$

In this case, Q3+1.5*IQR is a top whisker of the boxplot, IQR is an interquartile range, and Q3 is the third quartile. However, the dataset used in this research is highly skewed as we shown in Figure 3-5 One can see from the histogram that almost 140 countries have a portion of data less than 0.001 (0.1%) of the data. This means that almost 140 countries do not have significant number of refugees and a few countries allocated the high portion data. The X axis indicates the portion of the refugees data based on the number of countries which is in Y axis. The histogram of the data that used in this research shows that data is skewed which can be seen from Figure 3-5. Therefore, formulas number 3.1 and 3.2 that are mentioned above which are for calculating the outliers for symmetric data, is not applicable for Skewed data. Therefore, not only the common boxplot method does not accurately describe outliers but also not being recommended to use in skewed distributions.

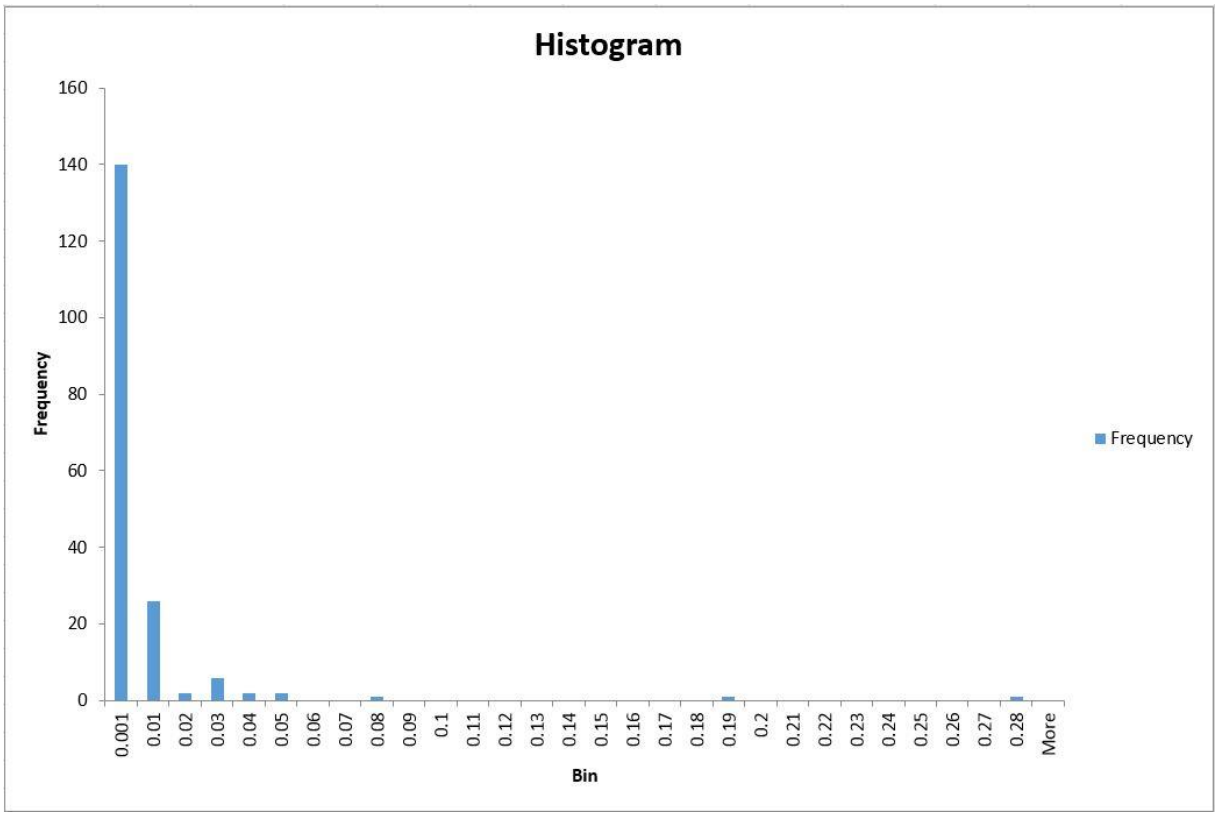


Figure 3-5: Shows the histogram of skewed data in 2014 for origin countries

Hence, the method that is intended for skewed distributions is to use an adjusted boxplot. The top whiskers are modified to take into account the measure of skewness of the distribution. In Hubert & Van der Veeken, (2008) are defined adjusted formulas for the whiskers of a boxplot:

$$[Q1 - 1.5\, e^{-4MC}\text{IQR},\ Q3 + 1.5\, e^{3MC}\ \text{IQR}] \tag{3.3}$$

If MC > 0 and

$$[Q1 - 1.5\, e^{-3MC}\ \text{IQR},\ Q3 + 1.5\, e^{4MC}\ \text{IQR}] \tag{3.4}$$

For MC< 0.

Where MC is the Medcouple Statistic for Skewness, a robust statistic to measure skewness of a univariate distribution. This measure can be calculated with by the use of the following formula according to the Brys, Hubert, & Struyf (2004):

Suppose a dataset with *n* observations from a continuous univariate distribution $X_n = \{x_1, x_2, x_3, \ldots, x_n\}$. Also, it has been assumed that the X_n is sorted in increasing way $x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n$. Moreover, let m_n be the median of X_n, defined as follow:

$$m_n = \begin{cases} (x_{n/2} + x_{n/2+1})/2, & if\ n\ is\ even \\ x_{(n+1)/2}, & if\ n\ is\ odd \end{cases} \tag{3.5}$$

The formula for calculating the medcouple (MC_n) is below:

$$MC_n = med_{x_i \le m_n \le x_j}\ h(x_i, x_j) \tag{3.6}$$

The kernel function h for $x_i \neq x_j$ is defined as:

$$h(x_i, x_j) = \frac{(x_j - m_n) - (m_n - x_i)}{x_j - x_i} \tag{3.7}$$

When $x_i = x_j = m_n$ is needed a special formula:

$$h(x_{m_i}, x_{m_j}) = \begin{cases} -1 & if\ i + j - 1 < k \\ 0 & if\ i + j - 1 = k \\ +1 & if\ i + j - 1 > k \end{cases} \tag{3.8}$$

Where k is the number of observations that are equal the median.

Calculation of outliers in the total dataset has been done for origin and destination countries. First, the data was aggregated by country by year (by summing all the possible destinations for the origin). This means that at the end there is a large dataset for each origin countries in 12 years (2003 to 2014). For instance, each country has a total number of refugees' population per year and 12 numbers for period of years. Then, the top whisker of the adjusted boxplot was calculated. This leads to seven countries as outliers for origin: Afghanistan, Iraq, Somalia, Syria, Sudan, Dem.Rep.Congo, and Vietnam. Then, the total dataset was reduced to only have data in which the origin is one of these seven countries, then the dataset was aggregated by destination country by year, and again the outliers calculated based on adjusted boxplot. This means, the outlier destination countries for the 7 origin countries that mentioned above are 11 countries: Pakistan, Iran, Syria, Jordan, Kenya, China, Chad, Turkey, Lebanon, Uganda, and Yemen. The same procedure was applied for destination also, the result of outliers are two countries: Pakistan and Iran and their Origin as an outliers are: Afghanistan, Iraq, and Somalia.

The application has an ability to show countries more than 2*3 matrix as the result of the destination. Therefore, in this way for the destination, the outliers were calculated with higher granularity, to verify smaller variation of data, and the outliers should be calculated by year. This means, instead of calculating the adjusted boxplot for total years, it has been done for each year separately and the result for destination countries are: Pakistan, Iran, Germany, United states of America, Syria, and Jordan and their outliers origin of these 6 countries are: Afghanistan, Iraq, Somalia, Sri Lanka, China, Colombia, Croatia, Eritrea, Haiti, Iran, Pakistan, Russian Federation, Serbia and Kosovo, Vietnam, Bosnia and Herzegovina, Sudan, Syria, Turkey, and Ukraine. Thus, in the application the same theory and method has been done for both of origin and destination with respect to that for origin it has been considered the total year at the same time and for destination per year to have not only the same destination that was on total year calculation but also more destination countries to be able to see in the application. In Figure 3-6 the left side of the figure presents the origin countries and their destinations based on adjusted boxplot. Moreover, on the right side of the figure indicates the destination countries with their origins.

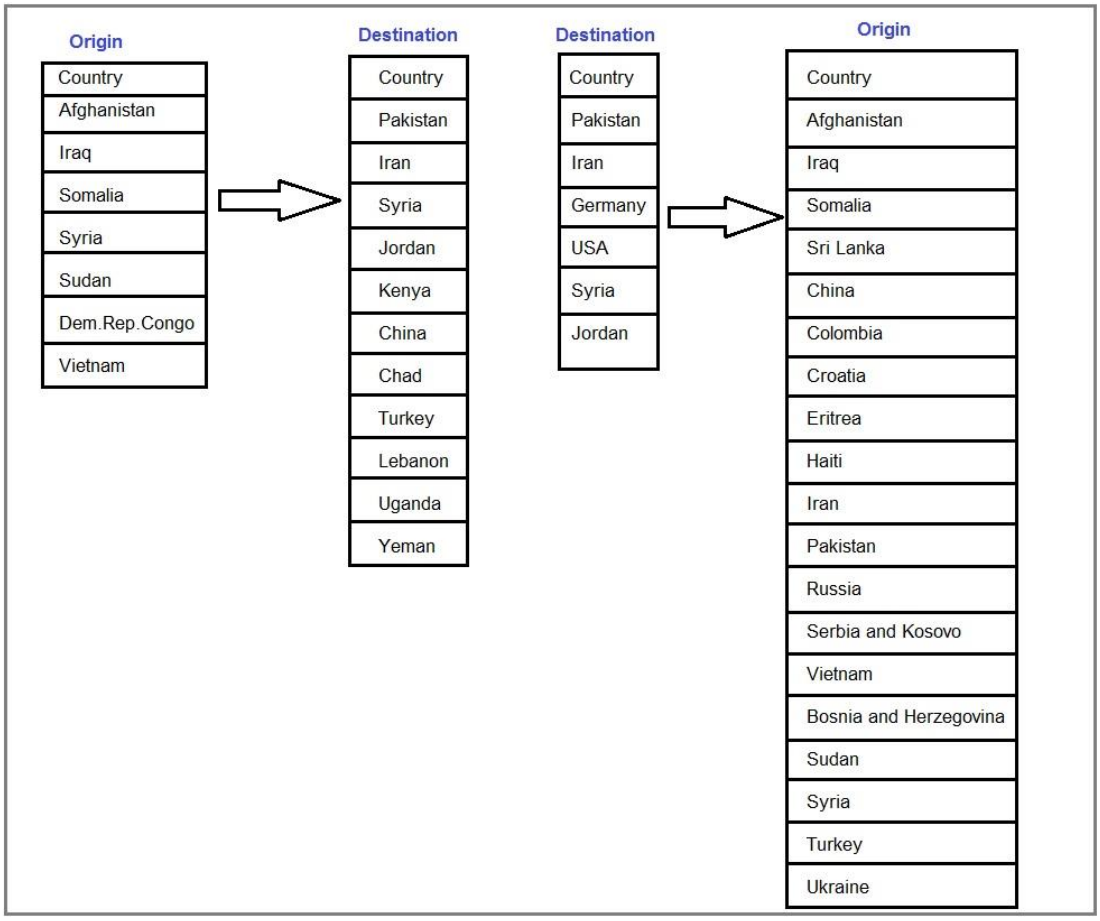


Figure 3-6: representing the Origin and Destination countries based on the threshold

4. RELATED WORK FOR VISUALIZATION

4.1. Introduction

The purpose of this chapter is to provide information about the previous works and researches regarding to this thesis topic. The first section starts with brief explanation about data visualization and provides some examples and theories of previous works that people have done for visualization of data. The second section discusses about geovisualization and how previously people work on this. The third section provides information about spatiotemporal visualization. This section considers about how previous works visualized and dealt with spatial and temporal data and visualized them. Visualization of Origin-Destination data is described in the section four. The last section is about usability which provides information about usability techniques and methods.

4.2. Visualization of data

Data visualization is a mean of telling a story and send the message to the user by expressing data graphically. The purpose of visualizations is not just show the data in table or graph; it has to communicate the data in an simple way, which means it has to be clear, well-designed, accurate and powerful to make the complicated data simple and understandable. Data visualization can be simple like a bar graph to show the rising of the number of population, or an interactive visualization on the webpage (Laumans, 2009).

Shneiderman (1996) developed a data type taxonomy for data visualization with seven tasks (overview, zoom, filter, details-on-demand, relate, history, and extract) and seven data types (one, two and three-dimensional, temporal and multi-dimensional data, tree and network data). Shneiderman's data types taxonomy explains how the created questions which can be asked by user should be represented in the interface. In this research Shneiderman's idea has been used to create the application in an overview level.

Elzen & Wijk (2014) explored a method for visualization of multivariate network data, which is defined by nodes and links, such as telecommunication and transportation for both expert user and non-expert user. This method consists of high-level infographic overview and the low-level network elements. However, there is a limitation in scalability in selecting the number of attributes and number of selections in this method. For the future work, authors recommended the interactivity between the visualizations that can help to enhance the detail view and supporting different visualizations such as visualization of matrix that in this research has been attempted to provide an interactivity among the visualizations.

Elzen et al. (2012) presented a system for exploring and analysing an extensive amount of mobile phone data that let the user get insight of data. They identified user tasks and based on that they developed a system that supported visualization, and interaction; then, they evaluated it based on massive mobile phone dataset. However, they concentrated more on the information in details level which has less communication based on the time limitation that they had. They suggested that additional visualizations and automated techniques to explore and analyse the remaining datasets would be valuable. Moreover, they believe that for future work exploring and visualizing the non-aggregated data when the pattern is complex can lead to getting more insight of the data. They used different kind of visualizations which has interactivity with each other on their application such as graph, heat-map and lines. In this research we applied different kind of visualizations and having interactivity.

4.3. Geovisualization

Kraak (2006) declared that Geo-visualization should provide information that be useful for the user. He expressed that Geo-visualization tools can be used to support the main geo-problems, especially in problems that deals with a large amount of data. Moreover, it is mentioned that Geo-visualization tools

and applications when they have interactivity it assists the user to understand the application and reveal the information. In this research was applied this idea of interactivity in the visualization of the application to assist users to access the information from the interface quicker in a form that can easily use.

Korpi & Ahonen-Rainio (2013) presented different methods for reducing the clutter such as refinement, displacement, aggregation, selection, animation, and others. However, the result of their evaluation shows that not all the methods for clutter reduction are perfect. In this research we used animation as one of the methods for reducing the clutter.

Guo & Andris (2012) concentrated on spatial data like origin-destination mobility data. They presented a new approach in analysis of spatiotemporal pattern, by providing means of understanding and discovering the origin destination movement. Their approach involved two parts: first, spatial clustering of huge GPS points; second, mapping the cluster-based flow. Nevertheless, they suggested that the low level of clustering should be used. For instance, creating small spatial cluster taking care of spatial distribution. Moreover, the clusters can have changes over time and regions have overlap when the hub connected to both of them. The method that they used for extracting the information from large Origin-Destination data was clustering. However, in this research adjusted boxplot for skewed data has been used for establishing the threshold.

4.4. Spatio-temporal visualization

Andrienko et al. (2007) believe that making a decision is based on the need. For instance, it is necessary to identify the typical needs for visualization of data based on the user requirements. Although they have mentioned that a way for dealing space and time problems is a decline in size and complexity of information, these problems still are not solved by existing technologies like GIS and still we are far from attaining comprehensive solutions. In this research we defined the user requirements, their preferences for the visualization, and interactivity of the design.

Boyandin (2013) provided a visualization of temporal OD-data and evaluated the visualization. He concentrated more on visualizing a different kind of flow maps and comparing them. Despite multi-maps and visualization that have been provided, the nature of the evaluation study is qualitative, and the results are not statistically validated to get a generalizable result. Moreover, the author believes that not all the flow maps are completely understood. Thus, understanding the limitation of flow maps, preventing of having clutter flows are challenges that have been tried to consider in this research.

Ilya Boyandin, Bertini, Bak, & Lalanne (2011) presented an interactive visualization in two maps that consist of origin and destination maps. Between the two maps, they represented the flows with heatmap. However, the number of lines between origin and destination maps are high, which makes the map unreadable. In this research a threshold was used to remove the excess of lines on the map, without removing flows which represent a high population of refugees.

Ho, Nguyen, Åström, & Jern (2011) implemented a web enabled demonstrator for visualizing and disseminating of flow time series data. Based on the different approaches to dealing with clutter maps that authors mentioned in their paper, such as changing the curves that representing the flows, they preferred to use interactive queries to illustrate and visualize a small subset of data to be able to prevent the clutter map.

This visualization has interactivity with other components of the application. They concentrated more on animation and user interaction. However, when the number of arrow lines increases, they overlap at the destination point. In this research different kind of flow lines are offered to the user which not only showed the direction of the lines by range of colour and animation but also gives the opportunity to the user to be able to see a direction of the arrow when one mouse over.

4.5. Visualization of Origin-Destination (OD) data

Guo & Zhu (2014) presented a new approach to visualize a large amount of mobility data, such as taxi trip with GPS information, which allows the analysis of the big data to be able to understand the complexity of flow trends. The authors had three main objectives for the research: the first one was able to remove the differences of the size of spatial units; the second was to extract the main information from the data by filtering the duplicated information, and the third was to give an effective flow map by allowing and enabling the flow patterns among origins and destinations. Additionally, they recommended suggestions for future work, such as improvement of the intelligent flow layout, adding the different visualizations, and also having user interaction for designing and supporting the application would be useful. Thus, this research tried to achieve legible flow lines among origins and destinations using filtering to be able to show the main parts of information which provide reasonable information for this research.

Rae (2009) proposed an approach for mapping a large OD matrix that can help to understand geographical patterns. He presented different visualizations such as symbols, flow lines, and heatmap. Nevertheless, not only the visualization of each map does not have interactivity but also a lot of symbols and lines make the map unreadable and clutter. In this research we provided interactivity among the visualization components and avoiding the clutter by showing a subset of data.

Lautenschütz & Eds (2012) worked with different tools like Treemap (Human-Computer Interaction Lab, 2003) and Hierarchical Data Explorer (HiDE) (Slingsby, Dykes, & Wood, 2011) to visualize OD-data. They believe that treemap is effective at showing and analysing a large amount of data. Nevertheless, getting to know that how software manipulates, the data is vital because any minor changes could alter the landscape of a map. HiDE simplifies the data by presenting them graphically which helps to identify the patterns of geographical data. In this research we applied different kind of visualizations not concentration only on one type of visualization.

Wood, Dykes, & Slingsby (2010) presented a new technique for visualization of origin-destination maps. Instead of lines for OD-data, they proposed them as cells. Moreover, they showed the layout of OD-matrix as a gridded-two-level spatial treemap. This aggregation of OD vectors into a regular grid is costly, and this is a disadvantage of this aggregation. In this research OD-matrix data was visualized with different kind of visualizations such as Origin Destination matrix, flow lines, and stacked bar graph instead of aggregating the OD data into the grid.

Guo (2009) addressed the changes of spatial interactions such as flows of population migration and spreading the disease. He developed an interactive visualization for a large amount of data of migration flows among the counties of USA. This visualization was based on three components: the first one was to bound partitioned graph method, allowing more flows within regions than across the regions; the second component was to use multivariate visualization method to be able to find different patterns in the flows; the third component was applying the interactivity for the visualization. The author believes that these objectives lead to visualize a large amount of dataset. Nevertheless, user interaction strategies and visual interfaces have not been done for this research. Moreover, he recommended the further research to make a better understanding of having the fine-tuning result. This visualization does not have an evaluation before and after the implementation of the design that in this research was applied focus group and heuristic evaluation before and after the implementation of the design.

Nowadays, there are numerous visualizations regarding migration data, but still there is a need for appropriate tools and techniques for visualization of migration. Boyandin, Bertini, & Lalanne (2008) argued that the flow map, showing the movement of the object from one location to another, is the most extensively used way for representing the migration flows and OD-data. They presented a visualization of migration flows to evaluate and illustrate the changes of flows over time. The authors believe that the reasons for using the flow map were that it is readable by its nature and is familiar to most of the people. However, flow maps may have a lack of precise interpretation, especially for showing statistical data when the number of represented flows increases. The authors suggested the evaluation of the design of the

future work of their research which based on that they can get the feedback from the users. This suggestion has been applied in this research.

Zambotti, Guan, & Gest (2015) built an online platform which is designed based on user requirements to visualize human migration over space and time. They connected the centroids of the origin and the destination of the countries by circular arcs of differing color. Nevertheless, users wished to select different base maps, changing the symbology for the flow streams and shading the polygons by their population that due to the lack of time, they could not apply these properties. They recommended the future work more on user preferences. For instance, providing the options that user can select and choose by himself, be able to change the symbology, and create interactivity among the flows. In this research was attempted to provide an interactivity among the flows and with other visualizations as well.

Xiao & Chun (2009) presented a new approach for visualization called Kriskograms to visualize migration flows. For creating Kriskogram, geographical units considered as a set of points on the straight line that called a location line. In this approach migration flows showed in half-circle in a clockwise from the origin to the destination. However, this approach is not a mapping method from a strict cartographical outlook. It maybe provides convenient patterns for migration flow; but, it could be difficult to divulge the traditional map. The authors suggested that this approach can be done in more dynamic visualization environment, and temporal changes of migration flows can be shown with animation to show the changes over time. In this research we decided to use animation and map for the visualization that the authors recommended in their research.

4.6. Usability

Nielsen (1993) defines usability as a multidimensional concept with five attributes: learnability, the user should be able to learn the system easily; efficiency, once the user learned how to use it should be able to achieve high productivity; memorability, once the user learned how to use the system it should be easy to remember, even when a casual user returns to it after a period without using; errors, the system, should have a low amount of errors, and when the errors occur it should be easy to recover from it; and satisfaction, the user should be satisfied using it. In another word, application should be effective (be able to answer the user questions correctly), efficient (be able to answer fast), satisfaction (pleasant too get answer).

Dumas & Redish (1999) defines usability as a dual of functionality. In the one hand, functionality is associated with what the system can do. On the other hand, usability is if the user of the system can make use of the functionality in a quick and easy way to accomplish their tasks. The key concept in the usability is the intended user of the system. The user can vary in computer experience, in knowledge about the domain, and knowledge about the system (Nielsen, 1993). All those factors should be taken into account when building the system.

Abras, Maloney-Krichmar, & Preece (2004) the advantages and disadvantages of usability techniques such as focus group, heuristic evaluation, interviews, questionnaires, and others are summarized. As advantages of using usability techniques can be cited that the products are more likely to be efficient, effective and error-free; that the product will need less redesign; and providing means of managing user expectations and satisfactions with the system. As disadvantages can be cited that the system becomes more costly, and it takes more time to implement.

To verify the usability of a system is necessary to execute usability tests. According to Dumas & Redish (1999) usability testing has five steps: improve the system usability; use real tasks; use real users; observe and record the actions of the users in the testing; and analyse the data obtained in the test, making the adaptations in the system accordingly. Some methods for evaluating the product that can be cited are focus groups, interviews, thinking aloud, eye tracking, questionnaires, performance measures, user feedback, and others. Typically the usability testing methods evaluate the system by measuring time to users learn a specific function, time to execute one task, frequency and type of errors by users, and

subjective user satisfaction (Abras et al., 2004). In this thesis, the usability testing will be focused on a focus group, as an early qualitative evaluation in the design cycle, and heuristic evaluation, as a qualitative measure of the usability of the system.

In Nivala, Tiina Sarjakoski, Jakobsson, & Kaasinen (2003) a usability evaluation was conducted on the use of topographic maps in mobile devices for the Geospatial Info-mobility Service by REAL-TIME Data-integration and generalization project (GiMoDig project). They believe that evaluation of the prototype has to be done in an early stage to make the final application with less errors and be more based on user requirements. They introduced different methods for evaluations such as focus group which was applied for this research before implementation of the design based on user requirements.

5. CONCEPTUAL DESIGN AND IMPLEMENTATION OF THE DESIGN

5.1. Introduction

This chapter explains the conceptual design and implementation of the prototype. First the system requirements will be discussed, followed by the conceptual design. To evaluate the implemented prototype the principle of what focus group is explained. And lastly, was discussed how the conceptual design can be implemented.

5.2. System Requirements

The goal of the system is allow the user to extract trends in migration of refugees over different geographic and temporal scales. The system should focus on the origin and destination of refugee, allowing the user to identify trends over time of refugees entering into a particular country or leaving a country. Also the user should be able to perceive geographical patterns in the flows of refugees. The geographic scales considered are country level, continental level, and world level. At country level the system should assist the user by pre-selecting the countries with the most number of refugees (origin or destination) and show the number variation over the years. Also, it should be able to show the trend of a specific country that the user is interested in. At continental level the user should be able to see the migration intra and extra-continent; also it should be able to compare in a yearly basis the number of refugees in each continent. At world level the user should be able to see the total number of refugees (origin or destination) on a yearly basis, so they are able to evaluate an increase or decrease over a period.
The user should be able to perceive the difference of values for origin and destination for individual years or a range of years. Data for between the years of 2003 and 2014 is available.

5.3. Conceptual design

Conceptual design is a scheme for designing the idea. It shows the scenario and structure of the design. The main thought behind conceptual design is to increase the chance and probability of having practical final product. Therefore, conceptual design should work well in details, show the outline of the design, and at the end be applicable for the user (Middle East Technical University, 2015). Based on Choi (2008) the conceptual design should be written in understandable language, which means user be able to understand each and every task of the design. And as such it should describe the function of the system. Thus, the main goal of the conceptual design is to tell users what exactly the final design will do and how it will be.

Based on the system requirements, a user at a first stage has to select in which geographical scales he wants to analyse the data, having two options to choose: continents level and countries level. Moreover, as mentioned above, there is a world level also which is shown in time slider. It is the 1x12 matrix that presents the total number of refugees per year based on Origin or Destination. When the user selects one of these options, it shows the design for that choice. Then the user can select if he wants to focus on origin or destination of refugees. When a user clicks on origin, it shows the countries/continents refugees with lines connecting the origin and destination on a map. This map assists the user to visualize spatial patterns in the data. If the user chooses the origin and selects a country on the map, it shows the country as an origin country and it shows the destination countries with lines for that selected country. Also, the application presents a time slider, where the user can select the specific year (or range of years) that he is

interested into. By selecting a year the application updates to show the data accordingly. Moreover, there is an animation mode where the years are passed sequentially, assisting the user the perceive differences and patterns among years.

In order to provide additional information to the user, two other elements are added, an OD-Matrix and a Bar Graph. The OD-matrix allows the user to verify the origin/destination values for a pair of countries in a specific year. To reduce the complexity of the matrix the application filter out the irrelevant countries based on the threshold method explained in Chapter 3. The matrix uses a choropleth representation to indicate which country/continent and which year has most number of refugees in percentage which allows seeing the trend of the Origin-Destination countries/continents from 2003 to 2014. Also, the user can mouseover the cell to verify the actual value. The bar graph aggregates the data in the OD-matrix, allowing the user to verify the total number of refugees (origin or destination) for that specific country. In order to reduce the number of elements in the graph the threshold method is also applied displaying only the most important countries according to the criteria.

The last element is a 1x12 matrix presenting the aggregated data at world level. This matrix allows the users to verify trends in the total number of refugees in the world. All graphical elements are interactive linked to each other assisting the user to understand the data in different geographical and temporal scales. Figure 5-1 shows the overview of the design, Figure 5-2 Indicates the continental level when the user clicks on origin/destination displays the flow lines which by mouseover on the line, it will be highlighted and shows the number of refugees for that origin/destination. Moreover, the OD-matrix with the name of the countries which in this sketch is mentioned by alphabet, is simplified just to provide an overview of the design. Different colours in the OD-Matrix indicates the classified data with value as visual variable in four classes which illustrate the refugees data in percentage. The world matrix which is on time slider shows the total number of origin/destination refugees data in percentage. Figure 5-3 is for country level when user clicks on origin/destination. OD-matrix, bar graph, and flow lines show the information of refugees data based on the outliers countries.

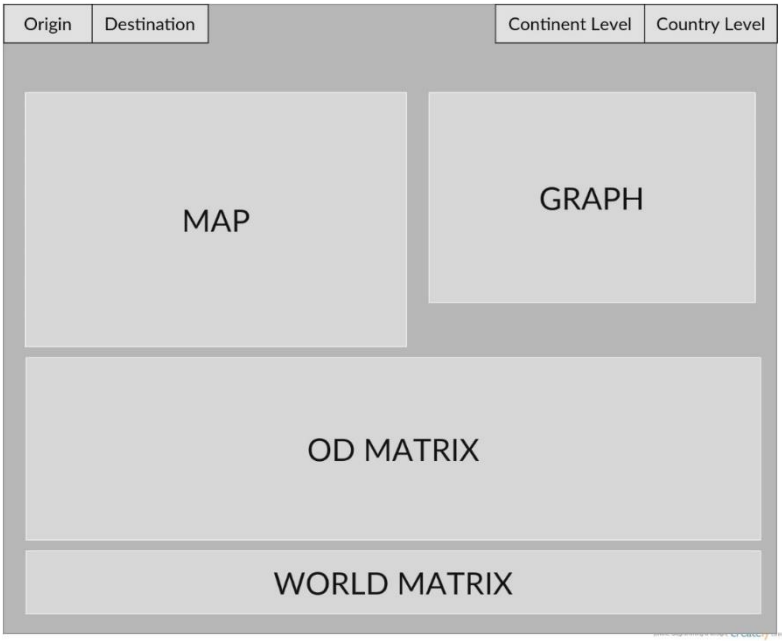


Figure 5-1: Design of the interface

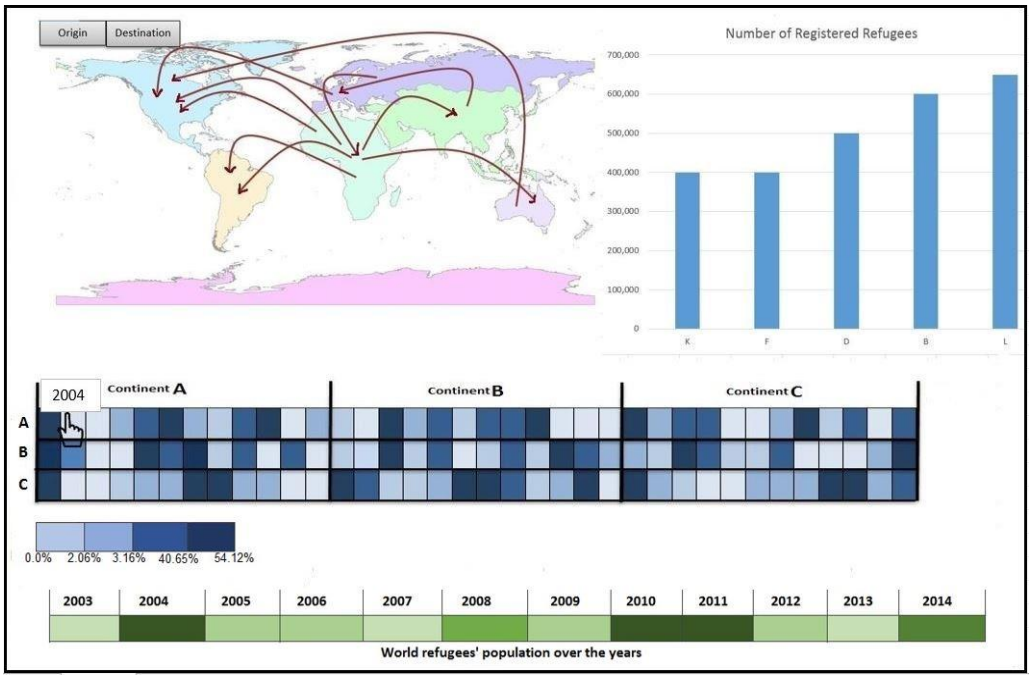


Figure 5-2: Conceptual design in continent level based on Origin and Destination data

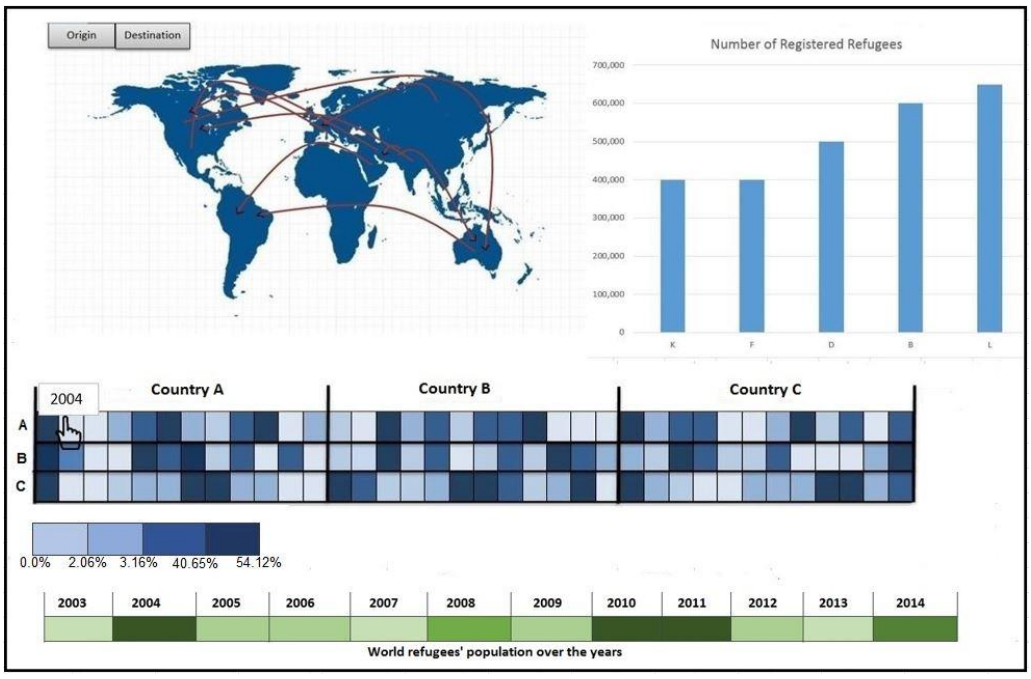


Figure 5-3: Conceptual design in country level based on Origin Destination data

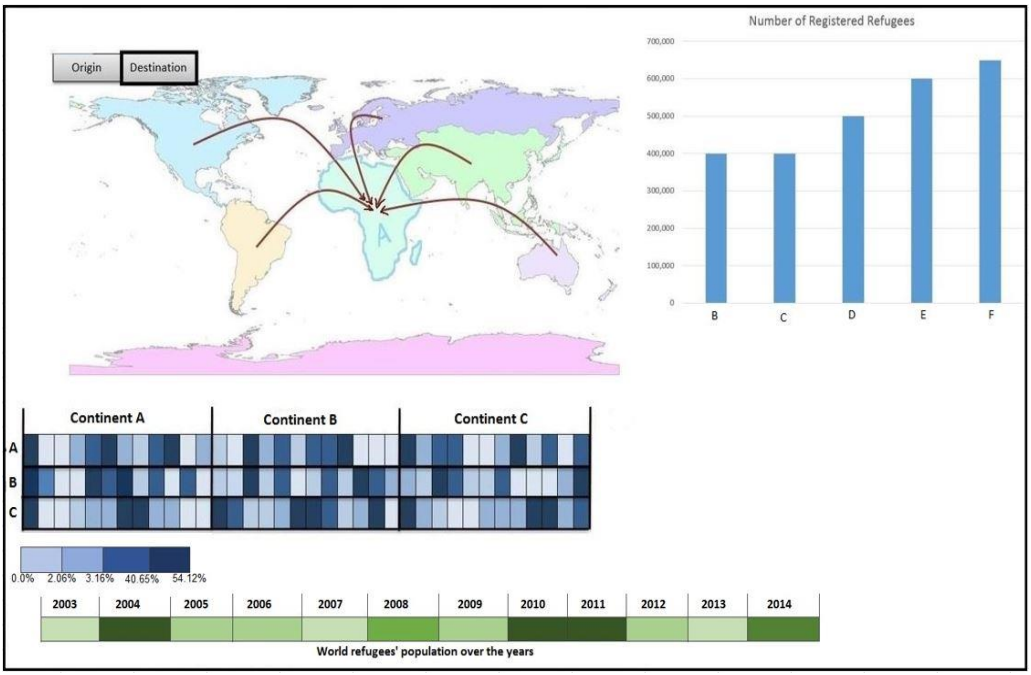


Figure 5-4: presents conceptual design at continent level based on selected continent

The figure 5.4 shows the continent A as a destination for the rest of the continents with the interactivity among the visualizations. This visualization is also applicable in countries level which means the user can select a country to see the flow lines in the selected country.

5.4. Interactivity of the design

According to the Kristof & Satran (1995) everything that people do on computer has been interactive. For instance, helping people to interact with words, numbers and pictures. To have interactive design as a process, two things have to take into account. The first is when all the pictures, ideas and other materials are used together, they should be a usable interface, which the user can communicate. The second point is that a complex system should break into a simple and tangible components. Moreover, in Kraak (2003) is mentioned that when there is an interactive and dynamic environment it helps the user to access the data based on graphical representation. Hence, in this research has been tried to accomplish interactivity among the visualizations.

5.5. Choosing colours for the application

Suitable colours, appropriate texts with good design play a key role in visualization of application. Choosing how and where to use information and texts on the application is essential for building design. Based on Nielsen (1993) normally information in many cases where it is in top left which is in many cultures gets more attraction to the user. Also, highlighting the colours, making the bolder or larger test are the ways that provide attraction. The colours that use for the design should not have saturated and contrast, it has to have harmony. Moreover, color-coding needs to be between 5-7 different colours not more than that since it is difficult for human to remember lots of colours. For the background colour also

greys and muted colours are more suitable than high saturated colours. Also, it is recommended not to use colours for quantitative information, use the colours for categorizing, differentiating, and highlighting. As defined above these points have to be taking into account which in this research takes the consideration. Hence, a focus group discussion is used to review the proposed Conceptual Design and to ensure that the proposed system will have a user friendly visualization. In the next section, Focus Group has been discussed.

5.6. Focus group

5.6.1. Introduction

Focus group is an interview that is organized in a way that is quick and inexpensive. It exposes an attendance desire, target, and idea. Focus groups called "focused interview", was developed as a method for social research in the 1930s, and then it used to improve soldier's life during World War II. Based on this, probably this method is one of the oldest method for user experience (Kuniavsky, 2003).

5.6.2. When focus group is appropriate

Focus groups are a suitable way to find the desire, motivation, and target. In other words, it is a way to get at people's attitudes and perceptions. It helps people to feel comfortable in an environment that they can share their ideas and thoughts. This lets people share their assumption and their point of view that are based on their experience and real world situations. It causes that reality can be seen from the user perspective easily, cheaply and quickly in a short time (Kuniavsky, 2003).

5.6.3. How to conduct a focus group

Based on Kuniavsky (2003) before starting the focus group, there are some important things have to be determined.

- *A schedule:* To be able to have a better result is important to have a schedule with enough time to examine the probability. To have enough time to write and prepare for focus group discussion.
- *The target audience:* people that are going to be invited and likely give the best feedback are the target audience. Therefore, it is important to know who are going to be invited.
- *Topics:* it is important to talk about the topic that wants to be researched. It should be considered that not all groups feel comfortable to talk about all the subjects and also not all subjects participative in a discussion. The more carefully and thoughtfully one chooses a topic, the more it helps to get the most information out of the group discussion without talking about unnecessary subjects.

5.6.4. How to setup the focus group

To be able to set up the focus group some requirements have to be prepared. Nielsen (1994) provided some preparations for focus group discussion that are:

- Number of users should be about six to nine
- The discussion should take about two hours
- User should feel free to talk and give their idea and thought
- The room should be reserved at least for three hours
- The meeting should be recorded

The steering of the focus group is the responsibility of the moderator. According to Kuniavsky (2003) Moderator is a person who has responsibility for holding the focus group. Some of the moderator's responsibilities are mentioned below:

- Spending time like 5 minutes before discussion begins to give good feeling for the personalities of the participants.
- The moderator has to respect the participants, listen to them carefully, and be able to think fast.
- The moderator should stick to the guide that is better someone take care of that and reminds the moderator all things that have to be done. However, if the discussion turned to the questions that have to be asked late and participants be comfortable with it, let them ask the question; unless, the order of the questions is critical. Then, the discussion should be back to the track.
- Clarifying the comments. When the user says something which is not clear, the moderator should ask for clarification and restarting the idea.
- The moderator has to provide time to think specially for the participants who cannot express their idea.
- The moderator has to create atmosphere so that people feel free to tell their idea, but not to make fun of the topic.

5.6.5. Users

We selected the user based on the definition of the users that are people who use the final product to fulfil their goal (Abras et al., 2004) the users for this focus group discussion are selected. Users in this research regardless of United Nations and organizations that care about refugees are people who would like to know about refugee's population and want to work with the application. In this focus group, we invited eight people who are MSc students from ITC- Faculty of Geo-information Science and Earth Observation of the University of Twente. Moreover, there has been tried several times to contact with UN to find a user from United Nations; but, unfortunately, no responses has been taken from them. The students from ITC are from different domains with different nationalities and gender. Four students are from Geoinformatics domain, one from Water Resources and Management, one from Urban Planning and Management, one from Natural Resource Management, and one from Environmental Modeling and Management Domain.

5.6.6. Focus group discussion

After making a schedule and inviting the users the focus group discussion held. Someone was responsible for recording the discussion from the beginning to the end and also for checking all the steps that have to be done. At first, the users filled form regarding the background of the participant that is available in Appendix A. Then, information has been given to the users with a slide presentation, which the slides of that are in Appendix C. Based on the slides, the topic of the research and objectives were explained to the users. Then, to make the users understand the concept of OD and OD-matrix data, the definition of them by examples described and showed. After making a clarification to the users, the implementation of the design was shown to the users. For each visualization, users gave their idea and comments. They discuss how to have interactivity with the all the visualizations. For a sketch of each visualization, there were provided an example to the user that they can understand the interactivity and the visualization more tangible. After that, the questions regarding the visualization have been given to them, and the participants filled the form available in Appendix B. The results of the questions for Appendix B are provided in Figure 5-5.

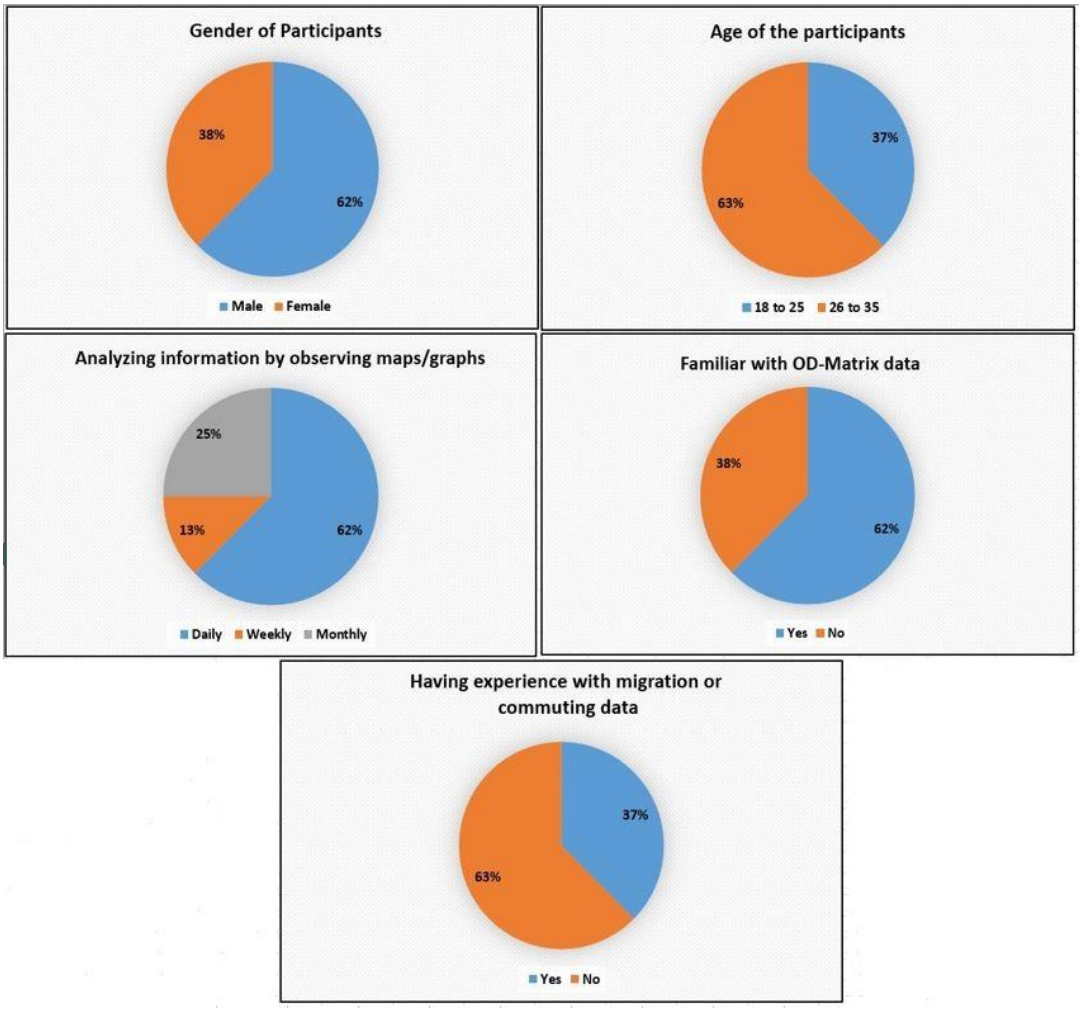


Figure 5-5: Pie chart representing the background of the participants in the group discussion

Figure 5-5 shows the background information from the participants in the group discussion. The information that has been taken from this evaluation is that 62% of the participants were familiar with OD-matrix data. Also, on the explanation that has been asked of them to describe the OD-matrix data, five out of eight people were familiar with OD-matrix data; but, not all of them could provide description for OD-matrix data. Three out of five participants explained and provided an example for OD-matrix data. For instance, one description of OD-matrix data from GFM student was "*it is a 2D matrix with one axis containing the place of origin and the other axis the destination, values contain a certain measure of volume and distance*". Another description that was given by GFM student is "*OD matrix gives transition of agents/commodities between two or more spatial locations at a given temporal distance*". Rest of the participants they were not familiar with OD-matrix data as mentioned above were three out of eight participants. The last question was about having an experience with migration or commuting data, based on that question, five out of eight participants did not have any experience with migration or commuting data. The participants who had experience with migration or commuting data were asked to give some examples. The examples were analysing the trends of commuters in Stockholm from their home to workplaces, analysing food productions import/export among countries, and doing exercise in Urban Planning classes for transport analysis.

The results of the questions regarding focus group discussion that is in Appendix B have been provided in the next steps. The first formulated question was how they prefer to see the visualization of Origin/Destination of their data. The output of the discussion for this question is shown in Figure 5-6. Five of the participants preferred to see the origin and destination by line. The users suggested that the lines with thickness would be more suitable than showing the lines with the same size. This idea helps the user to have a comparison among the lines.

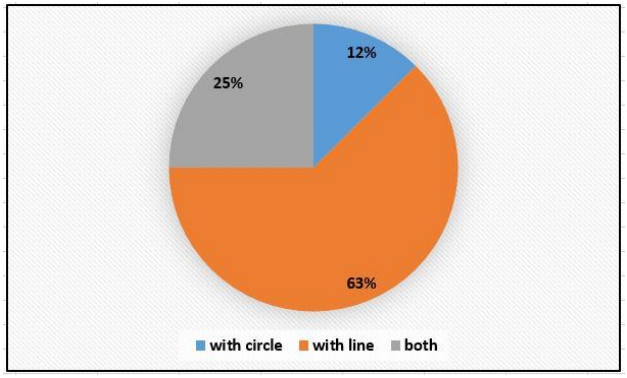


Figure 5-6: Participant preference of seeing the origin/destination of refugees' population in percentage.

The second question was asked from the users to discuss which kind of visualization they prefer when they want to see the path between origin and destination. It was asked the user the preference between line and slashed line. Moreover, the examples were provided to them to be able to see these types of visualization. The result below in Figure 5-7 shows their preference for this question. There was a suggestion that participants are seeking to see the lines with a thickness to compare which country or continent has more population of refugees. Moreover, animated lines which show the direction of origin and destination and thickness was another idea among the participants. We takes this idea into account in this research.

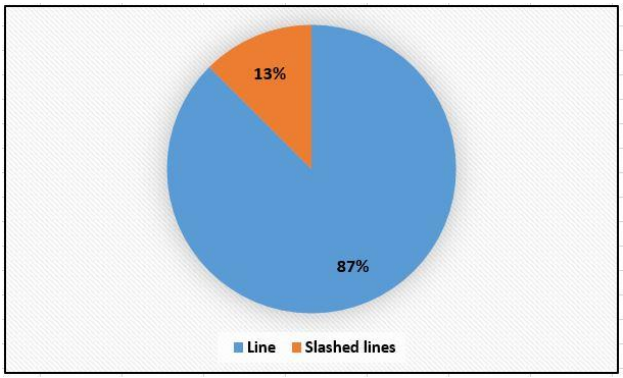


Figure 5-7: Participants preference of selecting line or slashed lines.

The next discussion about the third question was the preference of the user to see the changes over time in different ways. For instance, with animation or with a separate map for each year. Figure 5-8 indicates the outcome of the discussion of this question.

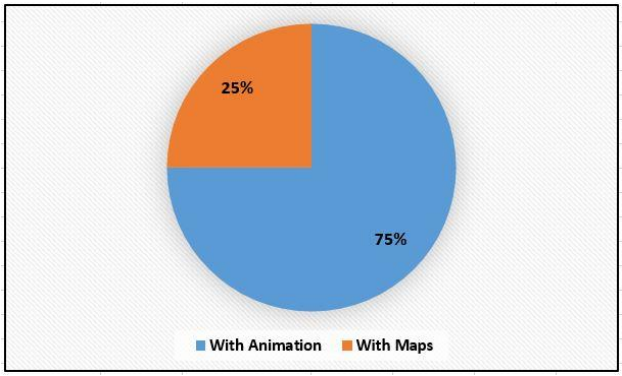


Figure 5-8: Preference of the users for seeing the changes in time

Most of the participants prefer to see time changes in time with animation that focuses on the origin and destination and has an interactivity with the rest of the visualizations. The next question was provided to discusse with the participant and to get their idea about bar graph and how they prefer to see the bar graph and that would be understandable for them to get the information easily. After the discussion, applicants provided an idea that consists of mixing of both types of bar graphs that had been explained to the participants. They preferred to have stacked bar graph for each year to be able to have both year and name of the country or continent also. Therefore, the outcome of this question is using stacked bar graph not the only suggested options in the question.

Finding which way of selection on the application is easier for the user, is another question that was explained and discussed in the group. Based on the result majority of the users preferred to have a button for selection. Figure 5-9 indicates the percentages of the users prefer the selection option. According to the discussion, most of the users prefer to have a configuration button to be able to select and choose the options.

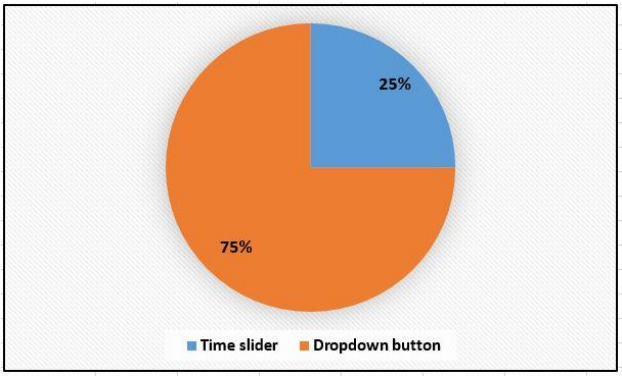


Figure 5-9: Percentage of users based on their preference of selecting options

The different colours for origins and destinations were another idea that has been discussed in the group that helps human vision to distinguish between origin and destination. This would be based on the colour of the countries or continents. Figure 5-10 shows the result of the discussion for these question.

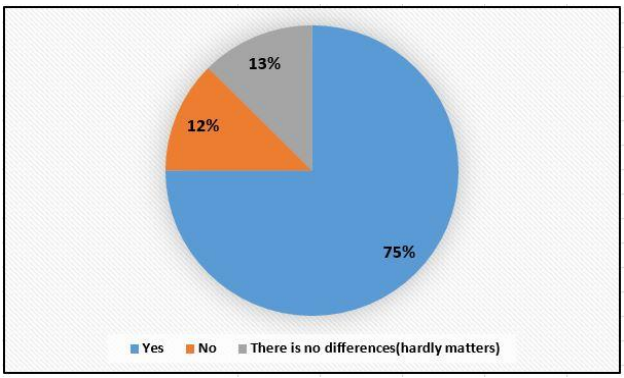


Figure 5-10: Preference to have different colours when they click on origin and destination.

Interactivity among different visualizations was discussed in the group. There were some idea that the users would like to see in the final visualization. These ideas are as follow which some of them have been decided in the conceptual design.

- When the user selects a country or continent on a map, then that continent or country has interactivity with its bar graph and with OD-matrix. This interactivity shows the country more clear for the user in comparison with rest of the countries or continents.
- When the user plays the animation while the year goes, each year simultaneously is identified in OD-matrix visualization. This is also the same for the destination.
- When the user selects origin, then define a country that country as origin becomes clear in OD-matrix.

5.6.7. Conclusion

Focus group discussion provided information for the user to know about their requirements, what interactivities they would like to see, and general understanding of OD-matrix data. These information and focus group discussion lead the conceptual design to be implemented in a way that final visualization is more user-friendly and has interactivity based on user preferences. According to the output of the focus group discussion, we decide to adapt the conceptual design to have interactivity with all the visualizations, fulfil the user requirements, have flowlines with thickness, show origin/destination with different colours, have animation on the application, have configuration button, and have stacked bar graph.

5.7. Implementation

5.7.1. Introduction

In this chapter mentioned earlier about conceptual design, and focus group discussion with the conclusion of the discussion. In this section implementation of the design has been explained. The prototype has visualizations that tries to fulfil the requirements and the outputs that has been achieved by the focus group discussion.

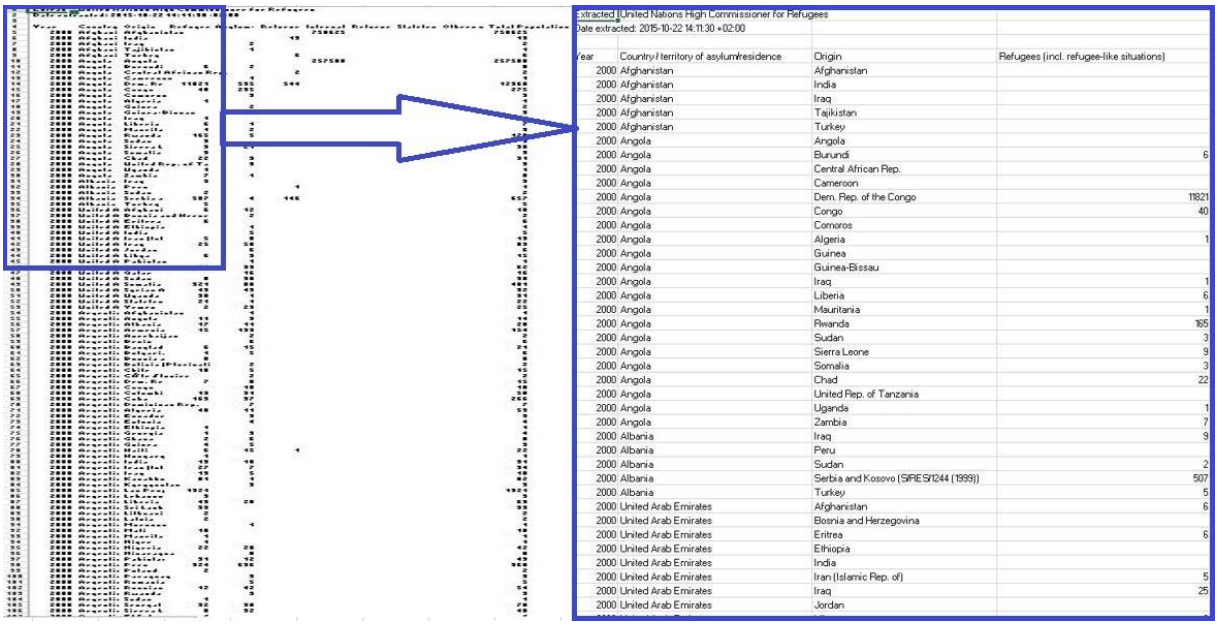
Implementation is executing the plan, idea, method, and design a model. It follows a preliminary idea or concept in order to make it happen. It is a step that idea and conceptual design implements. To be able to have a successful implementation many tasks have to be done to accomplish the successful implementation. For instance, it should be based on user requirements (Rouse, 2015).

5.7.2. Data preparation and processing

The type of the data that has been used in this research is United Nations Refugees data. This data was taken from United Nations Refugee Agency for 12 years (United Nations Refugee Agency, 2015). Data has been saved in CSV file and then prepared, corrected and imported to the database. Data at the beginning had some problem based on their Origin and Destination countries. This error has been reported to the UN refugee agency, and they corrected the data. Then, after taking the corrected version of data the preparation, correction and importing the data have been done again. Based on the information from the United Nation refugees Agency and Literature review of previous works with UN data refugee, the number of refugees from the data indicates the population of registered refugees. The data shows the people who registered as a refugee in that specific country in that particular year. Therefore, from the differences between the data during the different years, this may indicate that some of the refugees got accepted, some of them went out, some of them sent back and more reasons that we cannot be sure to say how many went or how many came.

5.7.3. Large amount of data

Human vision is not to be able to get all information from big matrix data at the same time. Moreover, presenting a large amount of data on a page is not readable, and the reader cannot get any information. Besides that, showing the OD-matrix data in values also does not demonstrate overall view and general information to the reader. Figure 5-11 indicates this problem.



Extracted | United Nations High Commissioner for Refugees
Date extracted: 2015-10-22 14:11:30 +02:00

Year	Country/Territory of asylum/residence	Origin	Refugees (incl. refugee-like situations)
2000	Afghanistan	Afghanistan	
2000	Afghanistan	India	
2000	Afghanistan	Iraq	
2000	Afghanistan	Tajikistan	
2000	Afghanistan	Turkey	
2000	Angola	Angola	
2000	Angola	Burundi	6
2000	Angola	Central African Rep.	
2000	Angola	Cameroon	
2000	Angola	Dem. Rep. of the Congo	11821
2000	Angola	Congo	40
2000	Angola	Comoros	
2000	Angola	Algeria	1
2000	Angola	Guinea	
2000	Angola	Guinea-Bissau	
2000	Angola	Iraq	1
2000	Angola	Liberia	6
2000	Angola	Mauritania	1
2000	Angola	Rwanda	165
2000	Angola	Sudan	3
2000	Angola	Sierra Leone	9
2000	Angola	Somalia	3
2000	Angola	Chad	22
2000	Angola	United Rep. of Tanzania	
2000	Angola	Uganda	1
2000	Angola	Zambia	7
2000	Albania	Iraq	9
2000	Albania	Peru	
2000	Albania	Sudan	2
2000	Albania	Serbia and Kosovo (S/RES/1244 (1999))	507
2000	Albania	Turkey	5
2000	United Arab Emirates	Afghanistan	6
2000	United Arab Emirates	Bosnia and Herzegovina	
2000	United Arab Emirates	Eritrea	6
2000	United Arab Emirates	Ethiopia	
2000	United Arab Emirates	India	
2000	United Arab Emirates	Iran (Islamic Rep. of)	5
2000	United Arab Emirates	Iraq	25
2000	United Arab Emirates	Jordan	

Figure 5-11: Shows the UN refugees data from 2000 to 2014

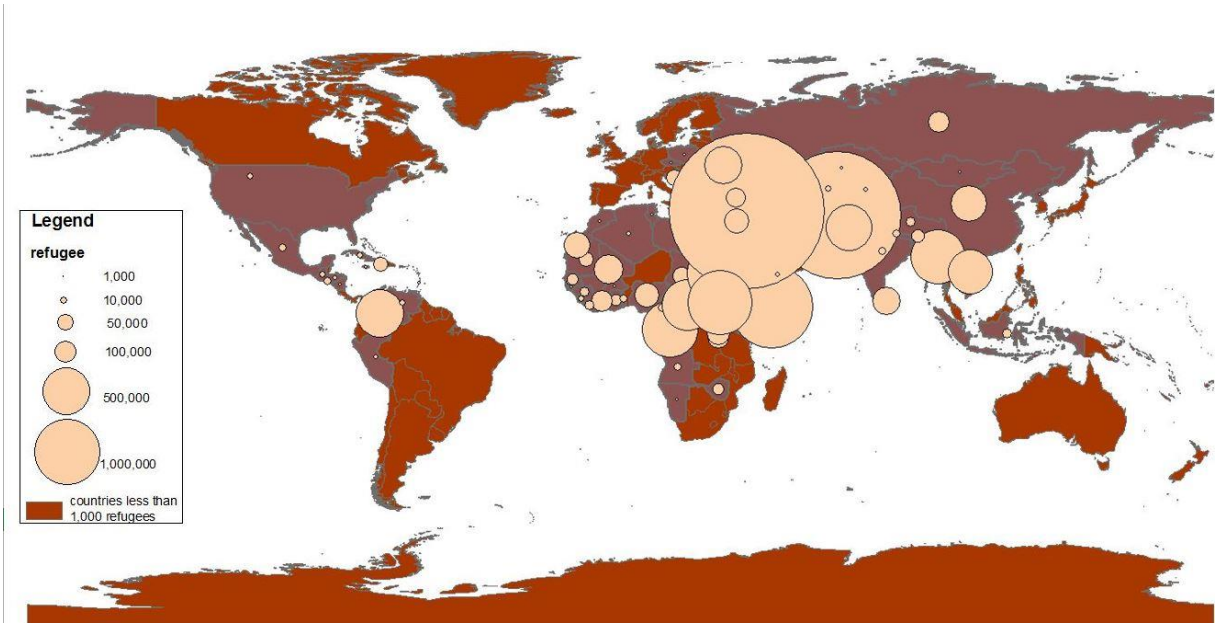


Figure 5-12: UN data in 2014 with refugees more than 1,000.

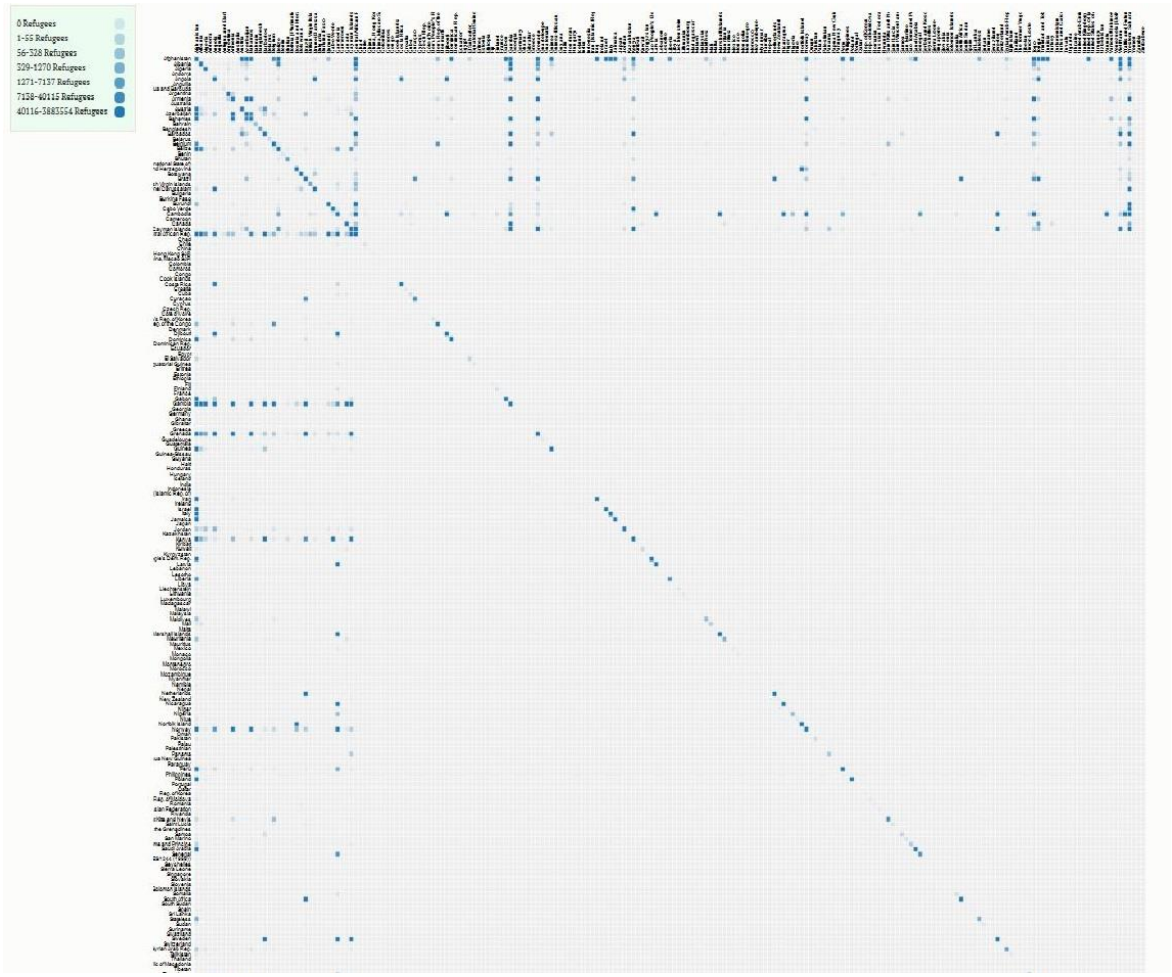


Figure 5-13: UN refugees OD-matrix data in value 2014 for all countries.

Figure 5-11, Figure 5-12, Figure 5.13 indicates a different kind of ways to show a large amount of data. One with data itself, one with visualizing on the map. One can notice the diagonal pattern in Figure

5.13 that is due to the data also consider internal displacement, which will not be the focus of this thesis. However, still getting general information a big OD-matrix getting information and comparison among the countries is difficult.

5.7.4. Programming language and using the library

To be able to have an interactive map, be able to sort and select there is a need for programming. In this research, JavaScript was used. The libraries that were used for this research are D3.js and C3.js. As advantages of D3.js can be cited that it is a JavaScript library that manipulates scalable vector graphics (SVG) elements, which are supported by every modern browser, it is a flexible library, not being constrained to a set of available visualizations. C3.js is a wrapper around the D3 library that provides pre-constructed charts.

5.7.5. Implementing the prototype

According to Houde & Hill (1997) prototype of an application is an early implementation that allows to represent and explore a conceptual design. Based on the conceptual design and conclusion of the group discussion, the prototype was implemented. The implementation of the design fulfils the output of the group discussion and implement the design with respect to the user requirements. In previous sections mention that application has different visualizations that has interactivity with each other. Figure 5-14 illustrates the visualization of refugee data in a continent level based on the origin. It can be seen from the Figure that different visualizations have interactivity with each other. For instance, the time slider with matrix and the stacked bar graph are shown by highlighting the selected year with red colour. In the following part explains about each of this visualizations.

The continent map is a GeoJSON file with the geometries and names of the continent. This was created from the countries polygons by merging all the countries that belong to the same continent. Later it was simplified and removed small islands that are not relevant for this visualization. This procedure was performed in Feature Manipulation Engine (FME). When the user clicks on the continent, the selected continent with the total population of refugees for each year will be shown in the stacked bar graph. Moreover, simultaneously the related flow lines and its interactivity with matrix will be illustrated. AppendixE shows more visualization of continent level with the interactivity among the visualizations.

The stacked bar graph is from C3.js with the original colours (Tanaka, 2014). The stacked bar graph for continent level consist of all the continents with the total population of refugees for each one separately by clicking on the continent and by indicating the portion of each continent on the stacked bar when all the continents are stacked. Although this procedure is the same for countries level, the selected countries for origin/destination on the stacked bar graph and matrix are extracted from a big OD-matrix by applying adjusted box plot.

All the data that have been used, were prepared for each of the components in the user interface. Then, data were extracted from database (MySQL) into JSON. Hence, there is specific file for each visualizations. One for stacked bar graph (stackBarPrepared.json), one for OD-matrix (odMatrixPrepared.json), one for the flowlines (mapPrepared.json), and one for cartogram/choropleth (data/aggregatedData.json). The time slider data comes from the Stacked bar graph.

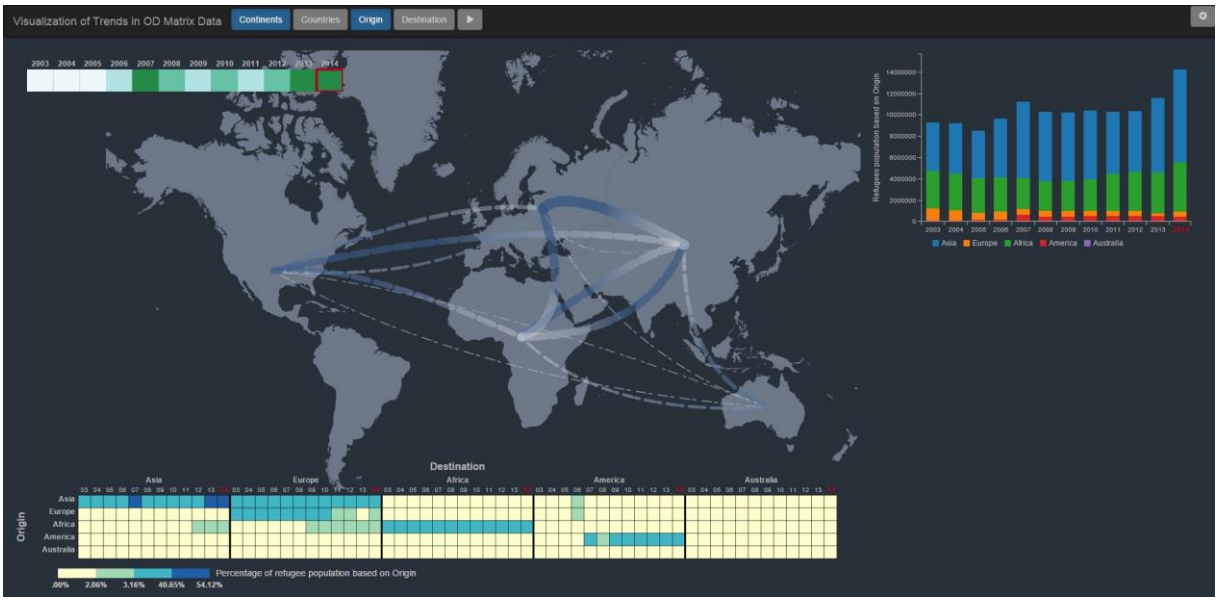


Figure 5-14: Shows the refugees' population in continent level

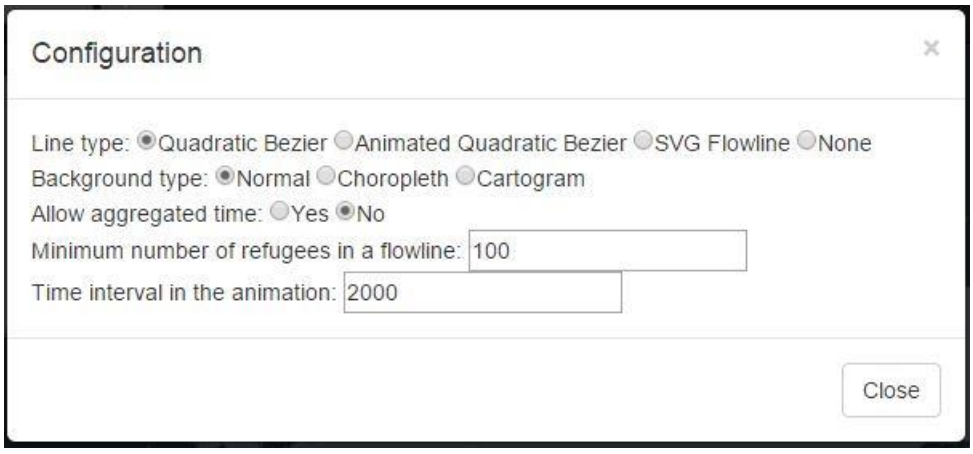


Figure 5-15: Indicates the configuration button

The application has configuration button that user can select different options which was shown in 5.15. It is observed that there are three types of lines "Quadratic Bezier", "Animated Bezier", and "SVG Flow lines". Regarding the Quadratic Bezier curved lines based on W3C (2011) quadratic Bezier takes 3 points, the initial, the final, and one auxiliary, the auxiliary point has been calculated between the initial and final points in a way that manipulates the orientation of the line. Moreover, if the two lines (the initial and the final) are close to each other, they calculated as a straight line. The animated dashed lines was done by only Cascading Style Sheets (CSS) from this reference (Abrate, 2013). It can be seen that lines also have gradient from light blue colour to dark blue colour which this help users to be able to see the direction of the flow along with the arrow and animation. As an comparison with quadratic Bezier flow line, SVG flow line from Boyandin was implemented which can be found at (Ilya Boyandin, 2015). We found cartogram and choropleth map two interesting options that we implement them also.

Another option from configuration button is Cartogram which can be seen in Figure 5.16. For cartogram there is the need of geometry which should be in TopoJSON. Therefore, TopoJSON library

should read the geometry which this TopoJSON library has been used by the following website (Mike Bostock, 2016). Moreover, the original library for cartogram uses the old versions of D3 and TopoJSON, and it does not work with new versions. To be able to solve this problem Allen (2015) adapted the cartogram library to work with the new version of TopoJSON and D3. Moreover, there is a need of higher simplification of the world map countries for the cartogram to have a reasonably fast calculation.

Cartogram also has interactivity with the stacked barograph and OD-matrix. By selecting or mouse overing the countries, the information about that country will be shown on stacked bar graph and on the cartogram. Moreover, by mouse overing on the OD-matrix the information about the refugees, year, and origin/destination for that country will be shown. The more visualization about the cartogram and the interactivities based on the origin/destination has been provided in AppendixF.

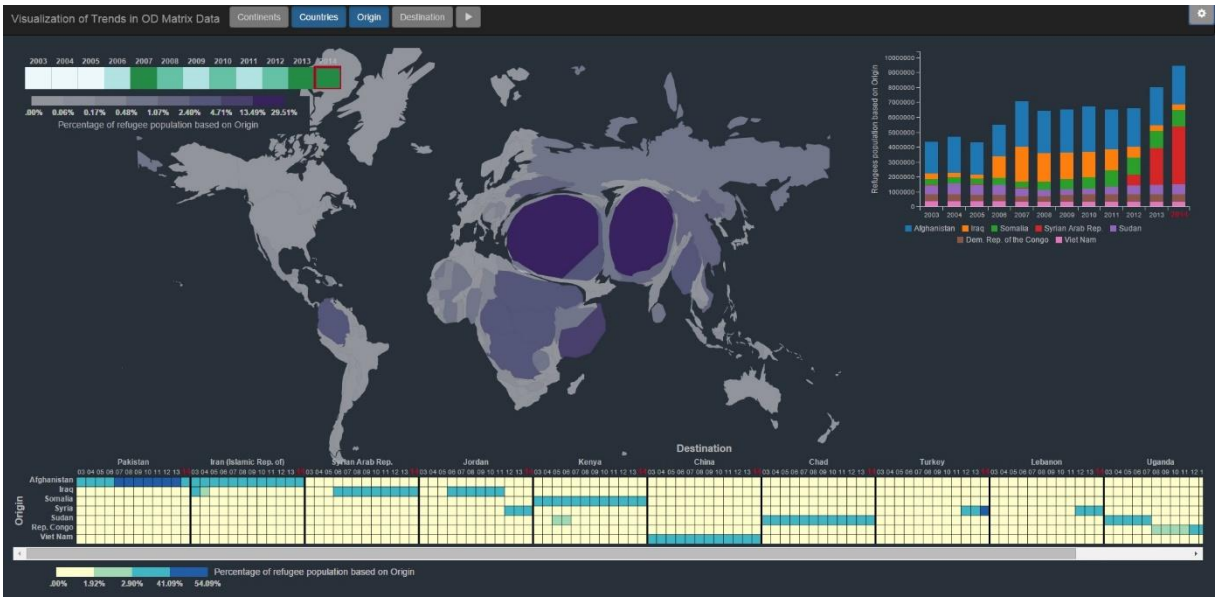


Figure 5-16: Shows the cartogram map for origin countries in 2014

Choropleth map is another type of visualization that was provided to show the trend of refugees by countries or continents over time. Value as visual variable in eight classes was applied in Choropleth map which represents in percentage. Figure 5-17 indicates the Choropleth map of countries which can be seen simultaneously with flow lines. Also, in Figure 5-18 Choropleth map has interactivity with all the visualizations and has different colour classification like cartogram and OD matrix for origin and destination. The colour selection was used from the Colour Brewer which can be found at (Brewer, Harrower, Sheesley, Woodruff, & Heyman, 2013).

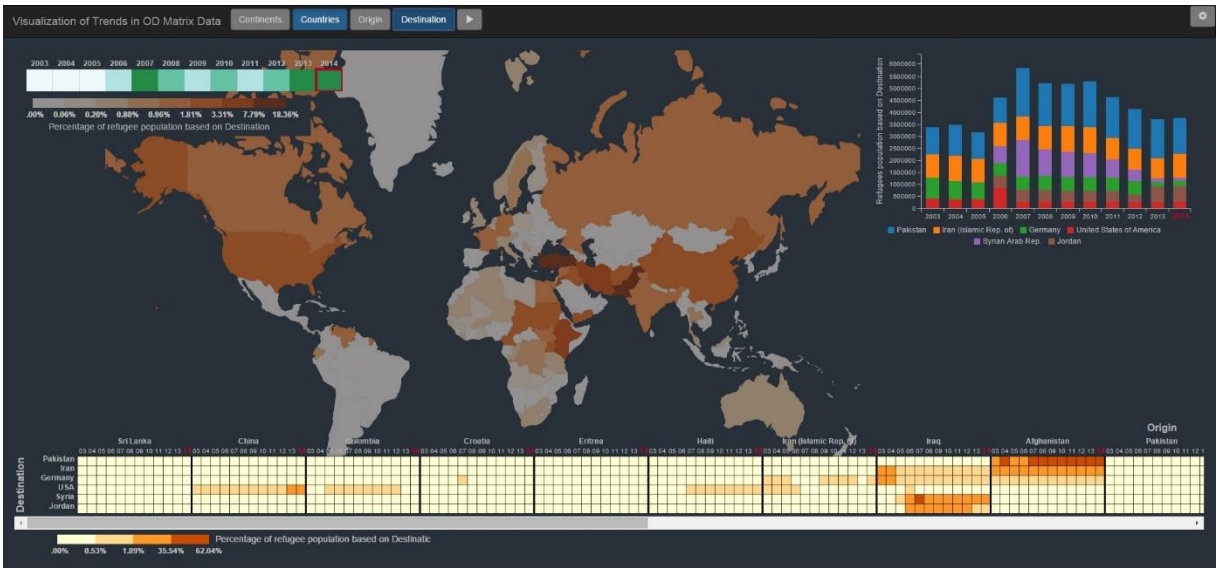


Figure 5-17: Illustrates the Choropleth map for countries in 2014

Figure 5-18: Shows the selected country as an Origin with its flow lines' destinations.

The classification of the data for OD matrix is with four classes, eight classes for Choropleth, cartogram, and Flow lines. These classification because of highly skewed data was done with Nested classification. Also, the total world matrix data in the time slider is not skewed, so for data classification equal frequency with four classes has been used. Moreover, as mentioned earlier to have a comparison between the lines, there are SVG flow lines which was implemented based on Ilya Boyandin (2013) Ph.D.

Thesis. Figure 5-19 indicates the SVG flow lines in continent level for a selected continent based on destination which has interactivity with stacked bar graph and OD matrix as well. It can be understood from the Figure and also configuration button that time slider is able to be aggregated. This permits user to select period in this case shows the refugees flow from 2011 to 2014 from Asia to Europe.

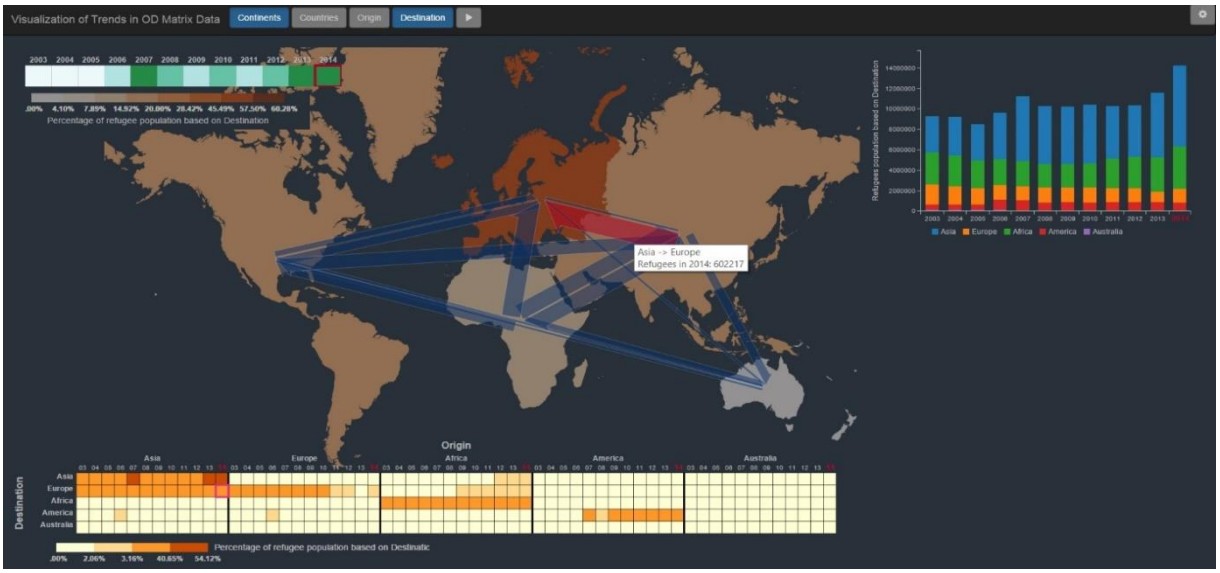


Figure 5-19: Indicates the SVG flow lines of the selected continent based on destination.

6. EVALUATING THE USABILITY OF THE DESIGN

6.1. Introduction

In the previous chapter, conceptual design and implementation of the design have been explained and discussed. To test if the design works well for the user, an evaluation has been applied. It should reveal weaknesses and strengths of the application. Heuristic evaluation is one of the evaluation methods which has been used in this research. The following section describes this approach in more details.

6.2. Heuristic evaluation

Usability is a question about how well system works for the user. It has different components such as being easy to use, easy to learn, being efficient for the user, easy to remember, having few errors, and be pleasant in the subject (Nielsen, 1993). Heuristic evaluation according to Nielsen (1994) is an engineering usability method to be able to find the problem, weaknesses and strength of usability from the interface. In this method a small set of evaluators which are experts, examines the interface based on recognizing the usability principles.

In Usability.gov (2013) heuristic evaluation has advantages and disadvantages which are defined as follow:

The advantages of Heuristic evaluation:

- The feedback that designers can get from this evaluation is quick and almost inexpensive.
- The feedback can be achieved in at an early process of the design.
- Allocating the correct evaluation can lead to the best measurement to the designers.
- It can be used with other usability testing methods.

The disadvantages of Heuristic evaluation:

- To have an effective evaluation it needs knowledge and experience.
- Finding the usability experts is difficult and sometimes is expensive.
- The result of multiple experts should be aggregated.

There are some steps that Danino (2001) specified as first steps of the heuristic evaluation that have to be taken into account.

- Develop tasks that are essential to the application's success. In another word, these tasks lead evaluators to know what should be expected to get from the application.
- Explain to the evaluators the goal of the application and allow them to provide their tasks and feedback from the application.
- Request evaluators to go through the application several times and give their comments about the application.

6.2.1. How to conduct the evaluation

Heuristic evaluation should be performed with each one of the evaluators inspecting the application alone. The findings of the evaluation should be documented as written reports for each evaluator, or by the conductor of the evaluation summarizing all the comments of the evaluators as they use the interface. The advantage of written reports is that there is a formal personal record of the evaluation. The advantage of the conductor of the evaluation summarize the findings is that it reduces the workload in each evaluator.

During the evaluation the evaluator should test and assess the interface elements several times, judging their compliance with recognized usability principles. It is recommended to the evaluators assess the interface at least twice, where the first assessment is to the evaluator understand the general scope and flow of interactions of the application while the second assessment allows the evaluator to focus on

specific elements. A section usually lasts between one and two hours for each evaluator. In Nielsen (1995a) it exposed ten usability heuristics that can be used for user interface design, they are summarized as follows:

- *Visibility of system status:*
  The system has to provide information about the application, what is it about in an appropriate time.
- *Match between system and the real world:*
  The system should be based on user language. It has to follow the logic and natural order. Moreover, the concept of the system should be familiar for the user.
- *User control and freedom:*
  The system should has function when the user makes mistake and needs to go back to the previous position or change the selection.
- *Consistency and standards:*
  There should not be the same meaning, words or action in the system to make users confuse.
- *Error prevention:*
  The system has to be carefully designed to prevent of having error. If it has error, it has to be explained to the user before user does an action or eliminate the error.
- *Recognition rather than recall:*
  The system does not have to have options, actions or objects that user needs to keep them in mind. Hence, the information should be visible for the user and limit the memory of the user.
- *Flexibility and efficiency of use:*
  The system has to have interactivity and be flexible to be usable for experienced and inexperienced users.
- *Aesthetic and minimalist design:*
  The system should contain relevant information and it does not have to contain information which is irrelevant and rarely needed.
- *Help users recognize, diagnose, and recover from errors:*
  The error of the system should be in a plain language that user be able to read and understand the meaning of the error and it provides the solution as well.
- *Help and documentation:*
  Although it is better for the system to not have documentation, sometimes is essential for the system to have help and documentation. However, this information should be easy to search and not be too large.

Since there is always major individual differences between the evaluations is recommended a group discussion in the end of the evaluation. In this group discussion the different opinions can be exposed and verified if any consensus about the problems found can be achieved. Also in this meeting can be discussed the good aspects of the application, not being restricted to the problems found.
The output of a heuristic evaluation is a list of usability problems found in the interface and their correspondent usability principles that was violated based on the opinion of each evaluator. With this list the application can be redesigned following the guidelines provided by the usability principle it violated.

6.2.2. Setup of the Heuristic evaluation

The first step in setting up the Heuristic evaluation is to decide on the number of evaluators. Figure 6-1 exposes the relation between the number of evaluators and the ratio of benefits to costs. In the case of the thesis also, the number of evaluators were constrained by the availability of usability experts.

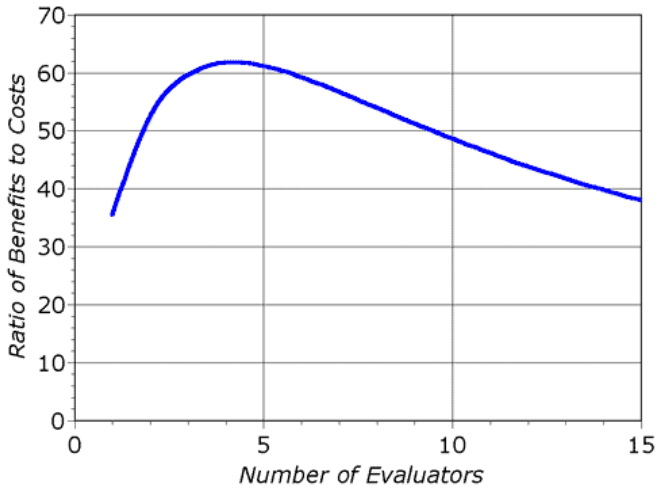


Figure 6-1: Number of evaluators versus ratio of benefits and costs (Nielsen, 1995b)

In Nielsen (1993) states that with three evaluators about 60% of the usability problems can be found. However, since the evaluators usually finds different problems in the application is possible to achieve a better performance. Also it recommends that the minimum number of evaluators should be three, and based on a cost-benefit analysis define the number used.

It is given an introduction about the application and the topic. Also a series of questions exposing the application requirements are given to the evaluators in order to help to understand the typical usage scenario. This questions are provided since the application is domain-dependent and the evaluators are not necessarily familiar with refugees' migration. The following questions were used to exemplify the typical usage scenario of the application:

Questions can be applied for Origin or Destination.

1. Did the refugees population increase or decrease during 12 years?
2. In which continent(s) did the refugee's population increase in 12 years?
3. Which region has the most number of refugees during 12 years?
4. What is the trend of continent X during 12 years?
5. What is the trend from continent X to continent Y during 12 years?
6. Which year Country X has a highest number of refugees?
7. What is the trend for country X during 12 years?
8. Which country has a highest number of refugees in year Y?
9. What is the trends of continent X with all its destination/origin countries during 12 years?
10. To which continent did the majority of refugee go from continent X in year Y?
11. Which Continent has the highest refugees' population during 12 years?

6.2.3. Heuristic evaluation of the application

This section explains the evaluation of the application. Experts for this evaluation were invited from the University of Twente, ITC-Faculty of Geoinformation Science and Earth Observation. After making an

appointment with each one individually, the evaluation was held. At first, the brief explanation of the thesis and objective of the evaluation have been given and explained to them which the brief explanation is available in Appendix 4. Then experts used the application and tried to answer the questions that have been provided for them. Experts examined the interface and found different weaknesses and strength points of the application based on the evaluation. The results of the evaluation with the first expert are listed as follow:

The weaknesses of the application:

- Adding general information on the application about the visualizations would help the user to be able to use the application in an easier way.
- The SVG flow lines hardly indicate the arrow especially when the magnitude of the line is small.
- The changes of flow lines over years are not as smooth as changing the choropleth map and cartogram.
- To mention the reasons of the trend changes, such as political situations and war helps the user to know more about the data and the reasons for these changes.
- The interactivity between the stacked bar graph and time slider to represent the year is not clear enough at the first moment. Increasing the font size of the year or making it bold make application more clear to the user.

The strength points of the application:

- It is be able to answer the questions based on user requirements.
- It has different visualizations that user is able to answer the questions in different ways.
- It has interactivity with each other which causes user understand the relation among the application.
- It is possible to have different flow lines that user can compare and select the most suitable one based on the user preference.

Discussion about the first evaluation:

This part explains about the weaknesses of the application based on the evaluation and provides the reasons that some of the feedbacks of the evaluation are not applicable to the application. The idea of adding more information on the application would be useful since the information does not make the application. Moreover, the SVG flow line, is based on the Ilya Boyandin (2013) Ph.D. thesis as a comparison with two different flow lines which in this application user preferred the Quadratic Bezier and animated Quadratic Bezier flow lines.

The outcome of the second expert provides below:

The shortcoming of the application:

- The boundary of the countries can be more reliable.
- The year of the default selected of the application in the time slider is not visible for the user to be able to define the years.
- Internal migration except with OD matrix in continental level, does not show in the application.
- The location of the configuration is not suitable for the user to be able to find it at the first moment.
- The total number of refugees' population for all 12 years has not been shown on the application.
- The scale of the vertical stacked bar graph should remain the same for all selected countries or continents.
- From the matrix most of the data belongs to one class. In this way the trend of refugees is not clear.

The advantages of the application:

- The origin and destination both have different colour which this make user have different visualization.

- Application is able to answer the questions that was provided for the expert.
- The different flow lines was a good idea that helps user to be able to have more options to select and compare.
- Cartogram is a suitable way to represent the refugee's population on the map especially when user wants to see the changes over time.

Discussion about the second evaluation:

The data that was used in this research is refugees' population who are registered and internal migration which is based on the United Nations considers as "*internally displaced persons*" is another attribute of refugees that is not within the scope of this research. Therefore, internal migration does not show in this application. Moreover, the refugees' population cannot be added because, the number of the refugees are registered refugees, not the exact value of refugees for each year separately. Hence, showing the summation of data during 12 years is not applicable. Regarding the Stacked bar graph, it combines rich information at the same time. If the scale remains the same, user is not able to see the countries which have less number of refugees. For instance, if the total number of refugees is 9,000,000 on the stacked bar graph and when user clicks on the country to be able to see the trend if the scale remains the same, the country with less than 100,000 refugees on the stacked bar graph will have small portion of the graph. Thus, to make graph with visible bars the scale has been tried to change based on the refugees' population for countries or continents. Also, the data is extremely skewed and the most suitable data classification was nested. This differences between the numbers of refuges causes the data does not show the tangible trend on the matrix.

The result of the evaluation of the third expert are as follow:

The weakness of the application:

- When the user selects the continent and then change to the origin or destination, the selected continent does not remain in the selected position
- Red and green colours are not suitable colours to be close together which is in time slider world matrix. Red and green are on the list of colour blind colours which a person who is colour blind may not be able to distinguish these colours.
- The SVG flow line does not show the direction of the arrow clearly specially when the number of lines increases.

The strength of the application:

- The matrix helps the user to be able to see the trend during 12 years at the same time based on their origin and destination.
- It has different visualizations that user can compare and examine.
- The application can answer the questions that is based on user requirements.

Discussion about the third evaluation:

The SVG flow lines as described earlier is based on the Ph.D. thesis, and the evaluator mentions that the arrow of the flow line which is not clear and two different flow lines are preferable.

6.2.3.1. Conclusion

To put everything in a nutshell, the interface can accomplish the user tasks. Based on the user tasks each of the visualizations of the application is suitable for specific type of questions. For instance, stacked bar graph in continent level is able to answer the questions like "did the refugees population increase or decrease during 12 years". Although world matrix on the time slider answers this question also, evaluators prefer to find the answer from the stacked bar graph. From the cartogram and choropleth map evaluators could answer the questions such as "In which countries did refugee population increase during 12 years"

and "which region has the most number of refugees during 12 years". The questions like "what is the trend of continent X during 12 years" can be answered by OD-matrix and stacked bar graph which in OD-matrix user can see the trend of origin/destination during 12 years at a glance. From the flow lines the questions like "To which continent the majority of the refugees go from continent X in year Y". Thus, application fulfils the user tasks and could answer the questions of the evaluation. According to the outcomes of the evaluation which explained with details in previous sections, in general, the application adapted the conceptual design and the weaknesses of the application are recommendation as future work.

7. CONCLUSION AND RECOMMENDATION

7.1. Conclusions

Origin-Destination matrix data inherently has high complexity. Extracting information from a large quantitative OD-matrix data is challenging. Although data visualization is one way of representing and interpreting the data, without a proper simplification method large matrices result into clutter visualization, not assisting the user to obtain an insight of the data. To be able to tackle this challenge, there is a need for alleviating the data while keeping the main and interesting information, which, in our case, are the countries with high population of refugees. The method that we used in this thesis was statistically defined a threshold. This simplifies a big OD-matrix to a small one while preserving the countries that have the maximum portion of data. Based on the threshold the conceptual design was built and discussed in the focus group. It is important to note that the outcome of focus group discussion such as how to have interactivity among the visualizations, using stacked bar graph instead of a simple bar graph, and others help to improve the conceptual design. Then, from the conceptual design we implemented the application with different types of visualizations which interactivity linked the components. This interactivity gives users the opportunity to see the relation among the visualizations. Then, the heuristic evaluation applied to assess the usability of the application. Based on the positive and negative feedback from the heuristic evaluation and overall outcome from the experts application seems to fulfil the user requirements.

According to the data that has been used in this research, visualization shows the refugees' population who are registered, not the exact number of refugees in each year. Although the visualization shows the trend of refugees' population, it cannot explain how many refugees are exactly in specific year in particular country. The flows of individuals are not taken into account, only considering the total number of refugees at the moment of data collection. In another word, if one year has more number of refugees in comparison with previous year, it shows more refugees; but, there may be some registered refugees that are from previous year, some were sent back, some did not register and more cases which do not show how many refugees are exactly in that year. Next, we answer the research questions.

Revisiting the research questions based on the objectives

Questions Related to Objective 1:

1- How to decide what are the trends in OD-matrix data?

We answer this research question at chapter 3. Visualizing a big OD-matrix data in an efficient way is challenging and cannot be visualized in effective way. The data that is used in this research is for 12 years from 2003 to 2014. We need a method to reduce the data from a big OD-matrix and make it in summary OD-matrix while keeping the main interesting information from the data that user would like to get from the data. In this case the countries which had a high number of refugees. Based on the literature there are different methods such as clustering, threshold, and others to have summary of data from the lots of data.

2- How to extract the trends from the OD-matrix data?

We answer this research question at chapter 3. The idea of choosing the countries that have high population of refugees from the big matrix is finding the outliers. In this case is the countries with having a large number of refugees in a particular year, as mentioned with more details in section 3.7. The data is highly skewed and the method for defining the outliers is applying a threshold for high skewed data.

Therefore, based on the adjusted boxplot, a method for calculating the highly skewed data, the outlier for origin/destination was defined.

Questions related to Objective 2:

1- What visualization representation(s) are suitable to visualize OD-matrix data?

We answer this question in Chapter 4. Based on the literature different visualizations were discussed and each one has its strengths and weaknesses. Flow line is one of the visualizations that in general is suitable for representing the OD-matrix data and it indicates the origin and destination. The interactive map is one of the visualizations that links the components to each other and makes the visualization understandable for the user. Stacked bar graph is rich in information and helps user to get different information at the same time.

2- Which representation(s) can show the trends or summary of these data?

We answer this question in chapter 5. Based on the conceptual design and outcome of focus group discussion, the conceptual design possess the user requirements and is able to show the trends of OD-matrix data. This visualization consists of different kind of flow lines "Quadratic Bezier", "Animated Quadratic Bezier", and "SVG flow lines" which represent on the map with three different choices "Normal map", "Choropleth map", and "Cartogram". Moreover, it has stacked bar graph, OD-matrix, which all of the visualizations have interactivity with each other.

3- Which questions should the visualization answer based on the user requirements?

We answer this question in chapter3 and formulated overall level questions for Origin or Destination, and used in the heuristic evaluation.

Example of overall level questions are:

- Did the refugees population increase or decrease during 12 years?
- In which continent(s) did the refugee's population increase in 12 years?
- Which region has the most number of refugees during 12 years?
- What is the trend of continent X during 12 years?
- What is the trend from continent X to continent Y during 12 years?
- Which year Country X has a highest number of refugees?
- What is the trend for country X during 12 years?
- Which country has a highest number of refugees in year Y?
- What is the trends of continent X with all its destination/origin countries during 12 years?
- To which continent did the majority of refugee go from continent X in year Y?
- Which Continent has the highest refugees' population during 12 years?

Question related to Objective 3:

1- How to set up and conduct usability study based on user tasks to evaluate the new representation(s)?

In chapter 5 and 6 we answer this question. At the first stage we adapt a conceptual design based on user requirements a focus group discussion. Number of users, students from University of Twente, invited for the discussion to give their idea about the design. Focus group discussion is a guideline to make sure the conceptual design is on a right track before be implemented. Then, we set up a heuristic evaluation to evaluate the application. In this evaluation, experts examine the application to find the weaknesses and strengths, regarding the usability of the system. This information can be later used to improve the application.

7.2. Recommendations for future work

Further analysis of the outcome of the heuristic evaluation could be done and implemented Additionally if the UN staff would have been able to participate we would have sharper user requirements. Therefore other recommendations are as follow:

- The interactivity can be shown with increasing the font size of the years or make them bold to be clear for the user.
- To have flexible formula that users can decide how many countries they want to see. In another words, users can choose the threshold.
- Currently the application only show stacked bar graph and OD-matrix for countries above the threshold. Ideally the application should be able to show data for any selected country.

LIST OF REFERENCES

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, *37*(4), 445–56. http://doi.org/10.1.1.94.381

Abrate, M. (2013). Flowlines. Retrieved February 3, 2016, from http://bl.ocks.org/nitaku/6354551

Allen, S. (2015). D3-Cartogram. Retrieved February 4, 2016, from https://github.com/shawnbot/d3-cartogram

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M. -J., MacEachren, a., & Wrobel, S. (2007). Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, *21*(8), 839–857. http://doi.org/10.1080/13658810701349011

Bertin, J. (1983). *Semiology of Graphics*. Madison: The University of Wisconsin Press. Retrieved from http://catalogue.nla.gov.au/Record/2340325

Bostock, M. (2016). TopoJSON. Retrieved January 28, 2016, from https://github.com/mbostock/topojson

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D 3 : Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2301–2309. Retrieved from http://vis.stanford.edu/files/2011-D3-InfoVis.pdf

Boyandin, I. (2013). *Visualization of Temporal Origin-Destination Data*. University of Friborg(Switzerland). Retrieved from http://ilya.boyandin.me/assets/thesis.pdf

Boyandin, I. (2015). SVG-Flowline. Retrieved February 3, 2016, from https://github.com/ilyabo/svg-flowline

Boyandin, I., Bertini, E., Bak, P., & Lalanne, D. (2011). Flowstrates: An Approach for Visual Exploration of Temporal Origin-Destination Data. *Computer Graphics Forum*, *30*(3), 971–980. http://doi.org/10.1111/j.1467-8659.2011.01946.x

Boyandin, I., Bertini, E., & Lalanne, D. (2008). *Visualizing Migration Flows and their Development in Time: Flow Maps and Beyond. Citeseer*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.187.2102&rep=rep1&type=pdf

Brewer, C., Harrower, M., Sheesley, B., Woodruff, A., & Heyman, D. (2013). ColorBrewer. Retrieved February 4, 2016, from http://colorbrewer2.org/#

Brys, G., Hubert, M., & Struyf, A. (2004). A Robust Measure of Skewness. *Journal of Computational and Graphical Statistics*, *13*(4), 996–1017. http://doi.org/10.1198/106186004X12632

Choi, B. (2008). CSC230 Designing the System. Retrieved from http://benchoi.info/bens/Teaching/Development/Software/PDF/CH05.pdf

Danino, N. (2001). Heuristic Evaluation - a Step By Step Guide Article. Retrieved January 19, 2016, from http://www.sitepoint.com/heuristic-evaluation-guide/

Dumas, J. S., & Redish, J. C. (1999). *A Practical Guide to Usability Testing* (Revised Ed). Portland, USA: Intellect Books. Retrieved from https://books.google.nl/books?id=4lge5k_F9EwC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Elzen, S. Van Den, Blaas, J., Holten, D., Buenen, J., Wijk, J. J. Van, Spousta, R., … Chan, S. (2012). Exploration and Analysis of Massive Mobile Phone Data : A Layered Visual Analytics approach. Retrieved from http://www.win.tue.nl/~selzen/

Elzen, S. Van Den, & Wijk, J. J. Van. (2014). Multivariate Network Exploration and Presentation : From Detail to Overview via Selections and Aggregations. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 2310–2319. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6875972

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring Usability : Are Effectiveness , Efficiency , and Satisfaction Really Correlated ? *ACM CHI 2000 Conference on Human Factors in Computing Systems*, *2*(1), 345–352. http://doi.org/10.1145/332040.332455

Global Issues at the United Nations. (n.d.). Retrieved August 12, 2015, from http://www.un.org/en/globalissues/briefingpapers/refugees/nextsteps.html

Guo, D. (2009). Flow mapping and multivariate visualization of large spatial interaction data. *IEEE Transactions on Visualization and Computer Graphics*, *15*(6), 1041–1048. http://doi.org/10.1109/TVCG.2009.143

Guo, D., & Andris, C. (2012). Discovering Spatial Patterns in Origin-Destination Mobility Data, *16*(3), 411–429. http://doi.org/10.1111/j.1467-9671.2012.01344.x

Guo, D., & Zhu, X. (2014). Origin-Destination Flow Data Smoothing and Mapping, *20*(12), 2043–2052. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6875983

Hindwan, S. (2010). The Dynamics of Refugee Problems. *Journal of Peace Studies*, *17*(2). Retrieved from http://www.icpsnet.org/adm/pdf/1291715201.pdf

Ho, Q., Nguyen, P. H., Åström, T., & Jern, M. (2011). Implementation of a Flow Map Demonstrator for Analyzing Commuting and Migration Flow Statistics Data. *Procedia - Social and Behavioral Sciences*, *21*, 157–166. http://doi.org/10.1016/j.sbspro.2011.07.029

Houde, S., & Hill, C. (1997). What do Prototypes Prototype? In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (Second). Amsterdam: Elsevier. Retrieved from https://books.google.com/books?id=WuQbERgXR10C&pgis=1

Hubert, M., & Van der Veeken, S. (2008). Outlier detection for skewed data. *Journal of Chemometrics*, *22*(3-4), 235–246. http://doi.org/10.1002/cem.1123

Human-Computer Interaction Lab, U. of M. (2003). Treemap: Home page. Retrieved August 14, 2015, from http://www.cs.umd.edu/hcil/treemap/

Korpi, J., & Ahonen-Rainio, P. (2013). Clutter Reduction Methods for Point Symbols in Map Mashups. *The Cartographic Journal*, *50*(3), 257–265. http://doi.org/10.1179/1743277413Y.0000000065

Kraak, M.-J. (2003). Geovisualization illustrated. *ISPRS Journal of Photogrammetry and Remote Sensing*, *57*(5-6), 390–399. http://doi.org/10.1016/S0924-2716(02)00167-3

Kraak, M.-J. (2006). Beyond Geovisualization: visualization viewpoints. *Ieee Computer Graphics And Applications*, (August), 6–9. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1652917

Kraak, M.-J. (2011). Is There a Need for Neo-Cartography? *Cartography and Geographic Information Science*, *38*(2), 73–78. http://doi.org/10.1559/1523040638273

Kraak, M.-J., & Ormeling, F. (2013). *Cartography: Visualization of Spatial Data* (3rd ed.). New York, USA: Taylor and Francis.

Kristof, R., & Satran, A. (1995). *Interactivity by design*. Mountain View, CA: Adobe Press. Retrieved from http://www.abebooks.com/9781568302218/Interactivity-Design-Kristof-Ray-Satran-1568302215/plp

Kuniavsky, M. (2003). *Observing the User Experience: A Practitioner's Guide to User Research*. Morgan Kaufmann Publishers Inc. Retrieved from http://dl.acm.org/citation.cfm?id=1197887

Laumans, J. (2009). An Introduction to Visualizing data. *Portal.Acm.Org*. Retrieved from http://piksels.com/wp-content/uploads/2009/01/visualizingdata.pdf

Lautenschütz, A., & Eds, S. T. (2012). *Migration and Mobility Data Visualization Workshop*. Retrieved from http://www.dsgz.geo.uzh.ch/pdfs/Visual_Migration_Workshop_May2012.pdf

Middle East Technical University. (2015). Conceptual Design [EE493-494 Design Course]. Retrieved December 28, 2015, from http://www.eee.metu.edu.tr/~design/doku.php?id=conceptual_design

Migration Policy Institute. (2002). *Refugees: Risks and Challenges Worldwide | migrationpolicy.org*. Retrieved from http://www.migrationpolicy.org/article/refugees-risks-and-challenges-worldwide

Nielsen, J. (1993). *Usability Engineering*. Burlington: Elsevier Science. Retrieved from http://ezproxy.utwente.nl:2343/patron/FullRecord.aspx?p=1190977

Nielsen, J. (1994). Heuristic Evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability Inspection Methods*. New York: John Wiley and Sons, Inc.

Nielsen, J. (1995a). 10 Heuristics for User Interface Design. Retrieved January 12, 2016, from https://www.nngroup.com/articles/ten-usability-heuristics/

Nielsen, J. (1995b). Heuristic Evaluation: How-To. Retrieved January 12, 2016, from https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/

Nivala, A.-M., Tiina Sarjakoski, L., Jakobsson, A., & Kaasinen, E. (2003). Usability Evaluation of Topographic Maps in. *2003 the 21St International Cartographic Conference*, 1903–1913. Retrieved from http://icaci.org/files/documents/ICC_proceedings/ICC2003/Papers/237.pdf

Nunez, C. (2014). The 7 biggest challenges facing refugees and immigrants in the US. Retrieved November 10, 2015, from https://www.globalcitizen.org/en/content/the-7-biggest-challenges-facing-refugees-and-immig/

Peuquet, D. . (1994). It's about Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems on JSTOR. *Annals of the Association of American Geographers*, *84*(3), 441–461. Retrieved from http://www.jstor.org/stable/2563777?seq=1#page_scan_tab_contents

Rae, A. (2009). From spatial interaction data to spatial interaction information? Geovisualisation and spatial structures of migration from the 2001 UK census. *Computers, Environment and Urban Systems*, *33*(3), 161–178. http://doi.org/10.1016/j.compenvurbsys.2009.01.007

Rouse, M. (2015). Implementation Definition. Retrieved January 5, 2016, from http://searchcrm.techtarget.com/definition/implementation

Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*. http://doi.org/10.1109/VL.1996.545307

Siyavula. (2012). Everything Maths and Science. Retrieved February 9, 2016, from http://www.everythingmaths.co.za/maths/grade-11/11-statistics/11-statistics-05.cnxmlplus

Slingsby, A., Dykes, J., & Wood, J. (2011). Hierarchical Data Explorer (HiDE) for Exploring Data and Tweeting Insights. *IEEE Symposium on Information Visualization*, 1–2. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.353.3351&rep=rep1&type=pdf

Tanaka, M. (2014). C3.js | D3-based reusable chart library. Retrieved February 3, 2016, from http://c3js.org/samples/chart_bar_stacked.html

United Nations Department of Economic and Social Affairs. (2013). *International Migration Report 2013*. Retrieved from http://www.un.org/en/development/desa/population/publications/pdf/migration/migrationreport2013/Full_Document_final.pdf#zoom=100

United Nations Refugee Agency. (2015). UNHCR Population Statistics - Data - Persons Of Concern. Retrieved December 27, 2015, from http://popstats.unhcr.org/en/persons_of_concern

Usability.gov. (2013, October 9). Heuristic Evaluations and Expert Reviews. Retrieved January 18, 2016, from http://www.usability.gov/how-to-and-tools/methods/heuristic-evaluation.html

W3C. (2011). Scalable Vector Graphics (SVG) 1.1 (Second Edition). Retrieved February 3, 2016, from https://www.w3.org/TR/SVG/paths.html

Wilkinson, R., Brill, W., Karsenty, N., Tigere, P., Zekrya, V., & Associés, V. W. (2001). the Wall Behind Which Refugees Can Shelter. *World War II*, *2*. Retrieved from http://www.unhcr.org/pages/49da0e466.html

Wood, J., Dykes, J., & Slingsby, A. (2010). Visualisation of Origins, Destinations and Flows with OD Maps. *Cartographic Journal*, *47*, 117–129. http://doi.org/10.1179/000870410X12658023467367

Xiao, N., & Chun, Y. (2009). Visualizing Migration Flows Using Kriskograms. *Cartography and Geographic Information Science*, *36*(2), 183–191. http://doi.org/10.1559/152304009788188763

Zambotti, G., Guan, W., & Gest, J. (2015). Visualizing Human Migration Through Space and Time. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *II-4/W2*(July), 155–161. http://doi.org/10.5194/isprsannals-II-4-W2-155-2015

APPENDIX A: BACKGROUND OF THE PARTICIPANTS IN THE GROUP DISCUSSION

Name:
Country:
Email Address:

1- **What is your gender?**
   a. Male
   b. Female

2- **What is your age?**
   a. 18 to 25
   b. 26 to 35
   c. 36 to 45
   d. 46 to 55
   e. 56 to 65

3- **How often do you analyse information by observing the graphs/maps?**
   a. Daily
   b. Weekly
   c. Monthly
   d. Rarely
   e. Almost never

4- **Are you familiar with Origin Destination-matrix data? If your answer is yes please describe OD-matrix data.**
   a. Yes
   b. No

5- **Do you have any experience with migration or commuting data or patterns? If yes, give some examples.**
   a. Yes
   b. no

APPENDIX B: QUESTIONS REGARDING FOCUS GROUP DISCUSSION

1- Do you prefer to see the population of refugees (based on Origin/Destination) with which type of visualization? It has to be mentioned that while user mouseover on the line it shows the number of refugees.

http://citydynamicbrowser.datacollider.io/

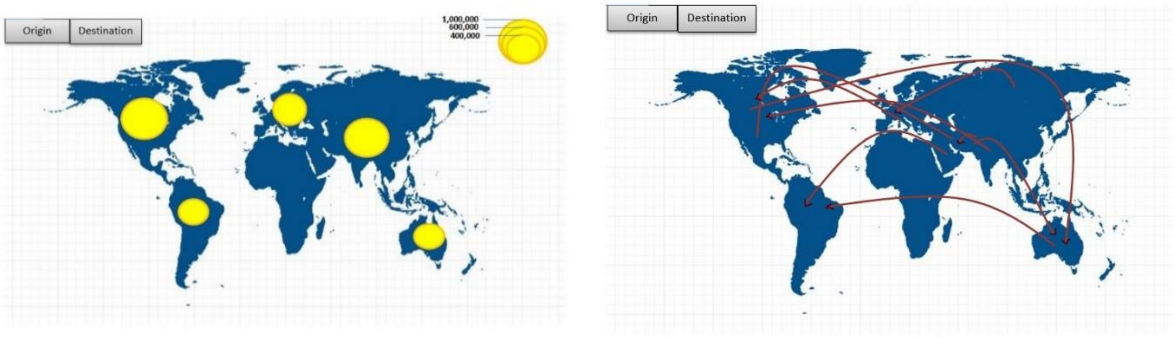


Figure B 1: two different visualization types for representing the refugee population

A) With circle
B) With line
C) Both

2- If your previous answer to question 1 is Line, do you prefer to see the direction (origin/destination) with a line or slashed lines?

http://www.lucify.com/the-flow-towards-europe/
http://air.nullschool.net/

A) Line
B) Slashed lines

3- Do you prefer to see the changes over time with animation or with maps in 12 years?

- With animation

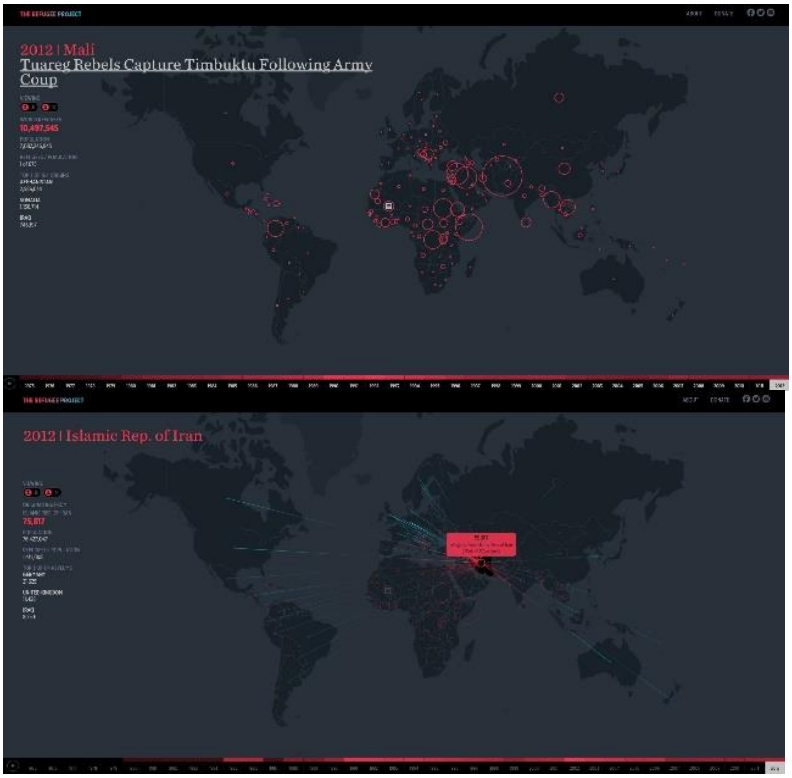


Figure B 2: an example of animation with time slider to indicate the time

Source: http://www.therefugeeproject.org/

- With different maps

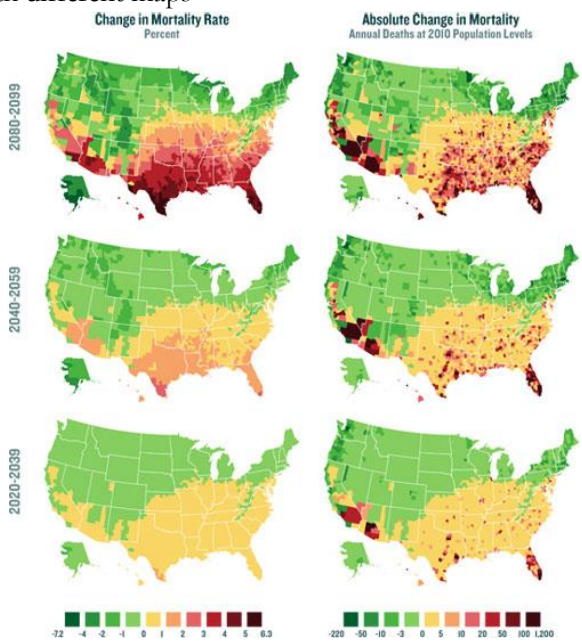


Figure B 3: maps representing the changes over time

Source: http://phys.org/news/2014-06-economic-impacts-climate-flashlight-night.html

A) Animation
B) Different maps

4- Do you prefer to have bar graph that shows the 12 years by having interactivity with map and matrix or you prefer to have a bar graph with the name of the countries/continents by having interactivity with map and matrix?

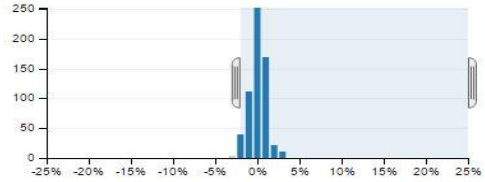
http://www.datavizcatalogue.com/

A) Bar graph that shows the years
B) Bar graph that shows name of the countries/continents

5- Do you prefer to use a time slider or to click on the button to choose a threshold?

https://dc-js.github.io/dc.js/
http://data.unhcr.org/dataviz/

- Time slider



- Dropdown Button

A) Time slider
B) Clicking on the button

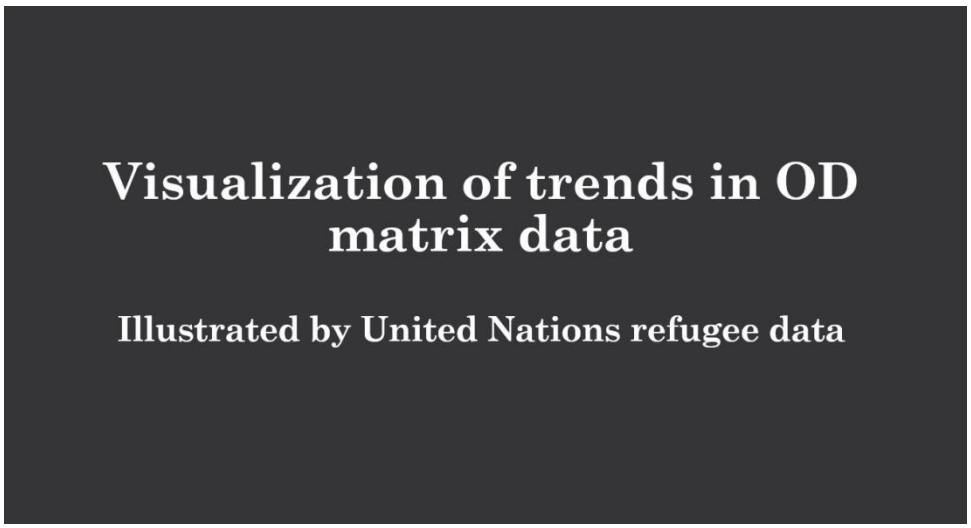
6- Do you prefer to have a different colour for origin and destination?

A) Yes
B) No
C) There are no differences (hardly matters)

http://data.unhcr.org/dataviz/

APPENDIX C: PRESENTATION SLIDES FOR FOCUS GROUP DISCUSSION

Slide1:



Slide2:

- **Main Objective**

Is to design, implement and evaluate an interactive map based representation, to visualize spatiotemporal OD-matrix data.

Slide3:

Overall Level questions

- Questions can be applied for Origin or Destination.

1- Did the refugees population increase or decrease during 12 years?

2- In which continent(s) did the refugee's population increase in 12 years?

3- Which region has the most number of refugees during 12 years?

4-What is the trend of continent X during 12 years?

Slide4:

Origin/Destination(OD) data
Origin/Destination(OD) Matrix

- What is OD data?

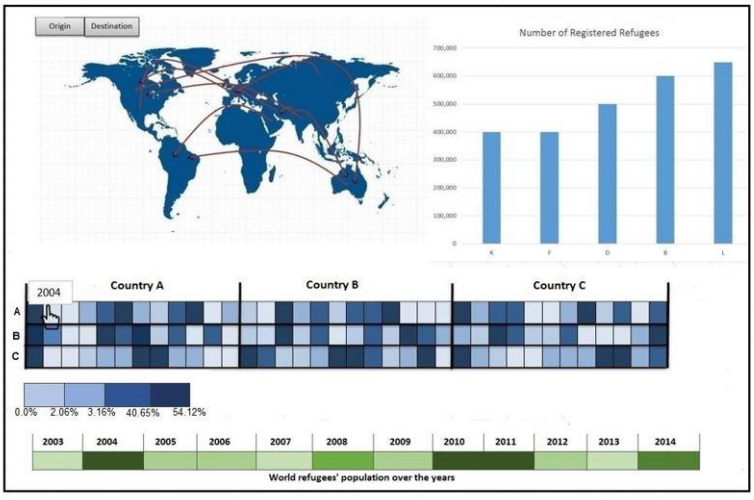
Is a way to show the transition between origin and destination of spatial data in geographic location.

- What is OD matrix data?

Is 2D matrix which rows show the origin and columns show the destination.

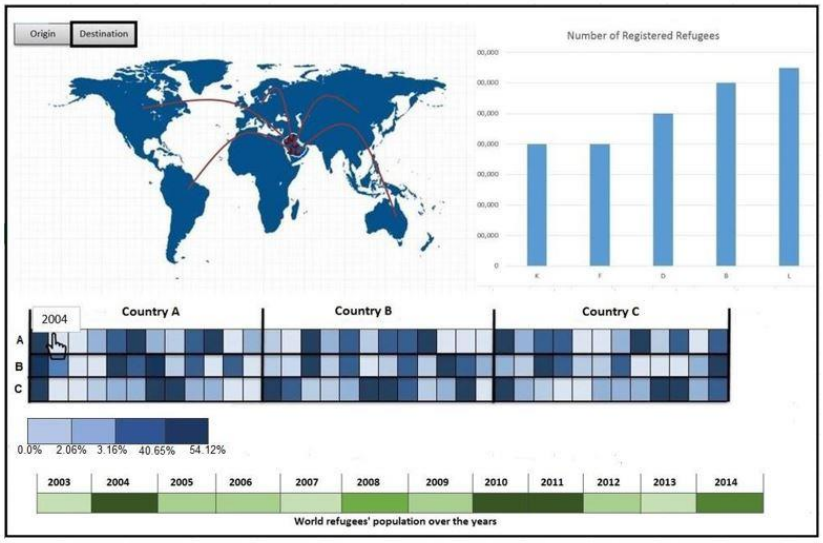
Slide5:

If user clicks on origin/destination and country level



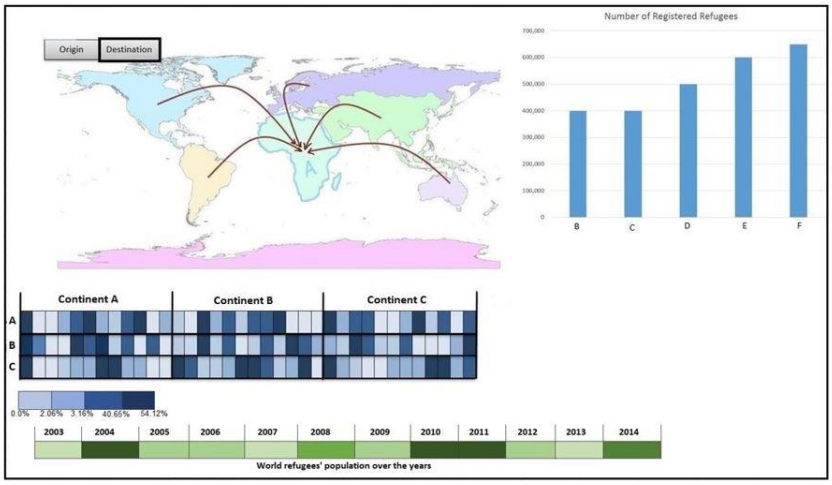
Slide6:

By selecting a country



Slide7:

By selecting the continent



APPENDIX D: BRIEF EXPLANATION OF THE OBJECTIVE OF THE EVALUATION

Brief explanation of the objective of the evaluation:
My thesis topic is the visualization of trends in OD Matrix data. Origin-Destination data is about spatial interaction between for instance people, or goods. The data is stored in matrix in which rows represent the origin and columns show the destination.

The main goal of this research is to design, implement and evaluate an interactive map based representation, to visualize the trends in spatiotemporal OD data. As an application United Nations refugee data from 2003 to 2014 is used. The prototype representation consist of three main views, the map, a stacked bar graph, and the OD matrix. Users are able to select origin and destination in the map.

The purpose of this evaluation is to get feedback about the usability of the prototype. For this you have to answer a set of questions while 'playing' with the prototype. It is also expected you list the major advantages and disadvantages. After the individual evaluation there is a group discussion to evaluate the common experience to find out how the prototype can be improved to reach its objective.

To be able to have an overview about the prototype, there are some explanations about the application. FigureD1. Shows the overview of the application which by default is 2014 with quadratic Bezier flow lines in continent level. This application has selection buttons which let user select the origin or destination at country or continent level. Moreover, it has a configuration button which FigureD2. Indicates the options that can be selected to change between different visualizations.

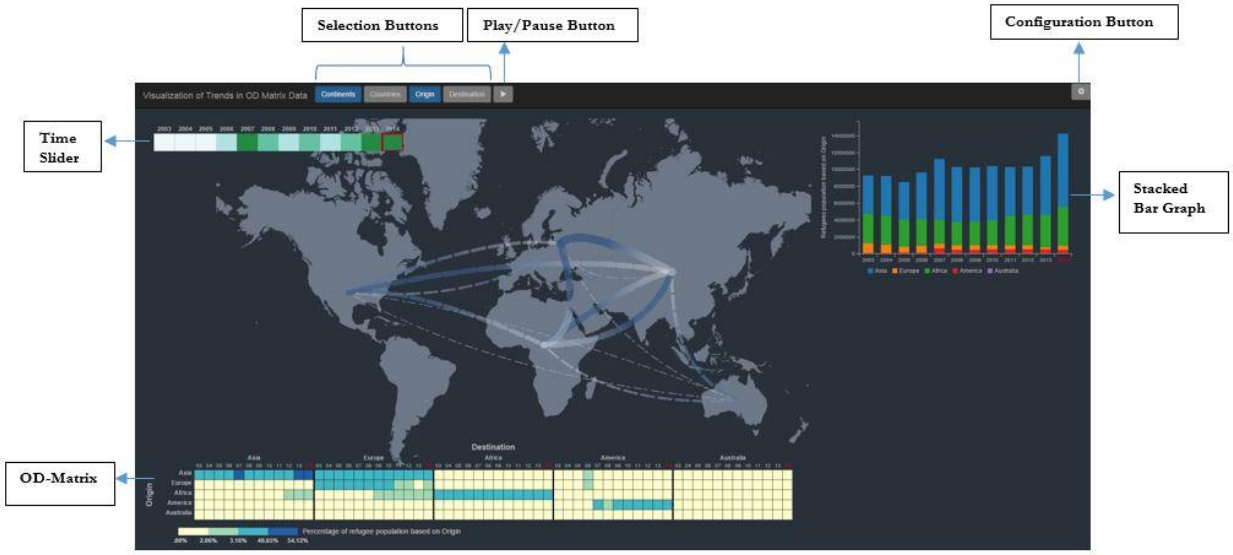


Figure D 1: the application with different visualizations

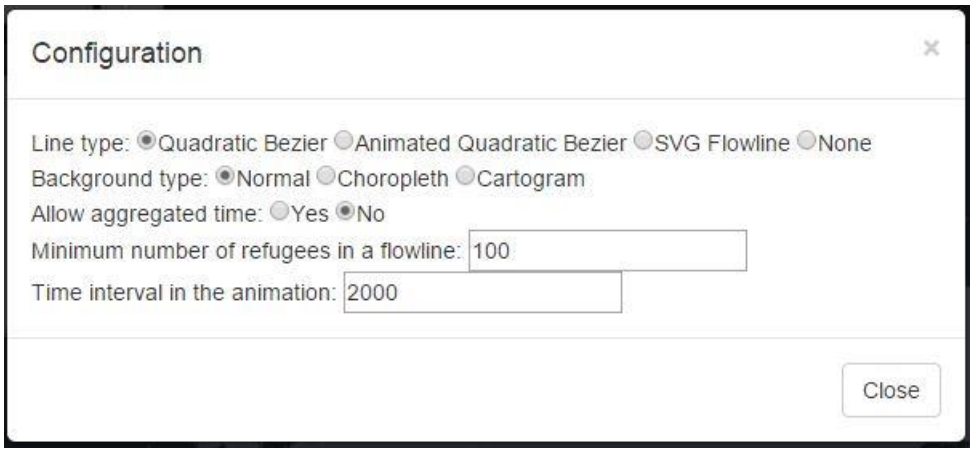


Figure D 2: configuration options

FigureD3 showed the interactive link between the stacked bar graph and the map. When user mouse over a country or continent it shows the portion of that country and when the country is selected, the refugee's population is shown for each year on the stacked bar graph. The stacked bar graph is also linked to the time slider and matrix. When user click on a continent, the year, the origin and destination of that country will be coloured on matrix. Moreover, when user mouse over on the flow line the corresponding cell of that flow line on the matrix will be coloured. This is also applicable in country level.

Figure D 3: Examples of the interactivity of the prototype

FigureD4. Illustrate the prototype with map with cartogram. The right one is based on origin and the left is based on the destination. One is also able to select aggregated time to be able to see the cartogram with aggregated values. This also is applicable for chorpleth map also. The cartogram is only in country level and choropleth map is for both country and continent level.

Figure D 4: the cartogram version for origin (right) and destination (left)

The application also has different flow line options. The left image in FigureD5. Shows the SVG flow line and the right image shows the Quadratic Bezier.

Figure D 5: SVG and Quadratic Bezier flow line

After the brief explanation and overview about the application. You should now use the application and try to answer the questions below.

1. Did the refugees population increase or decrease during 12 years?
2. In which continent(s) did the refugee's population increase in 12 years?
3. Which region has the most number of refugees during 12 years?
4. What is the trend of continent X during 12 years?
5. What is the trend from continent X to continent Y during 12 years?
6. Which year Country X has a highest number of refugees?
7. What is a trend for country X during 12 years?
8. Which country has a highest number of refugees in year Y?
9. What are the trends of continent X with all its destination/origin countries during 12 years?
10. To which continent did majority of refugee go from continent X in year Y?
11. Which Country has the highest refugees' population during 12 years?

APPENDIX E: SELECTED CONTINENT WITH ITS INTERACTIVITY BASED ON ORIGIN/DESTINATION

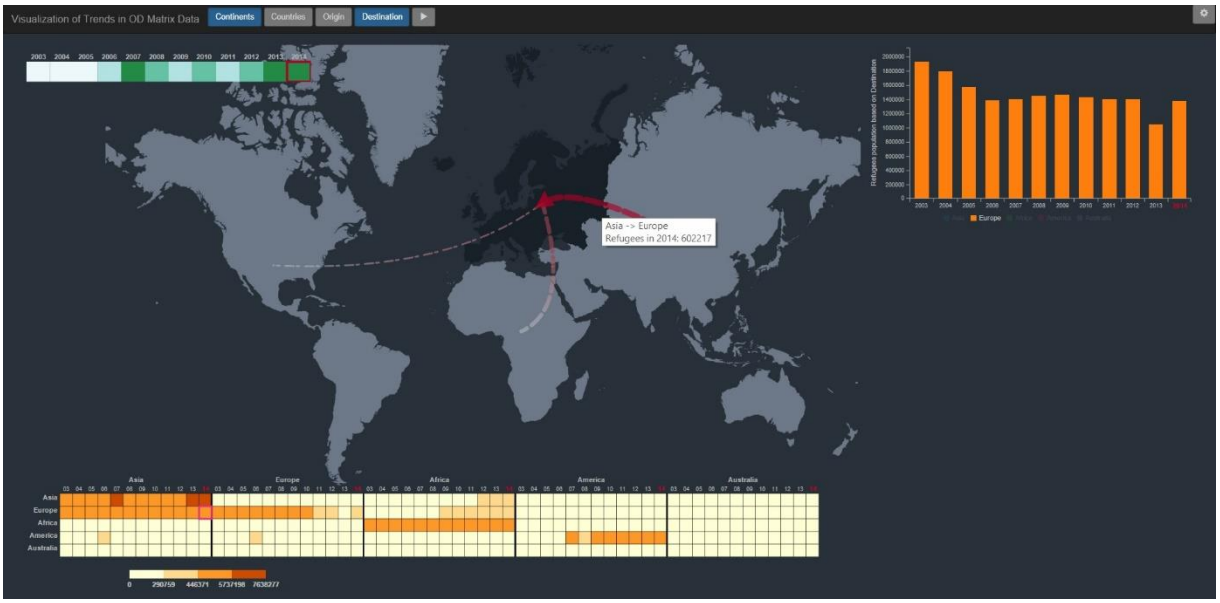


Figure E 1: the selected Destination continent with its interactivity with OD-matrix and stacked bar graph.

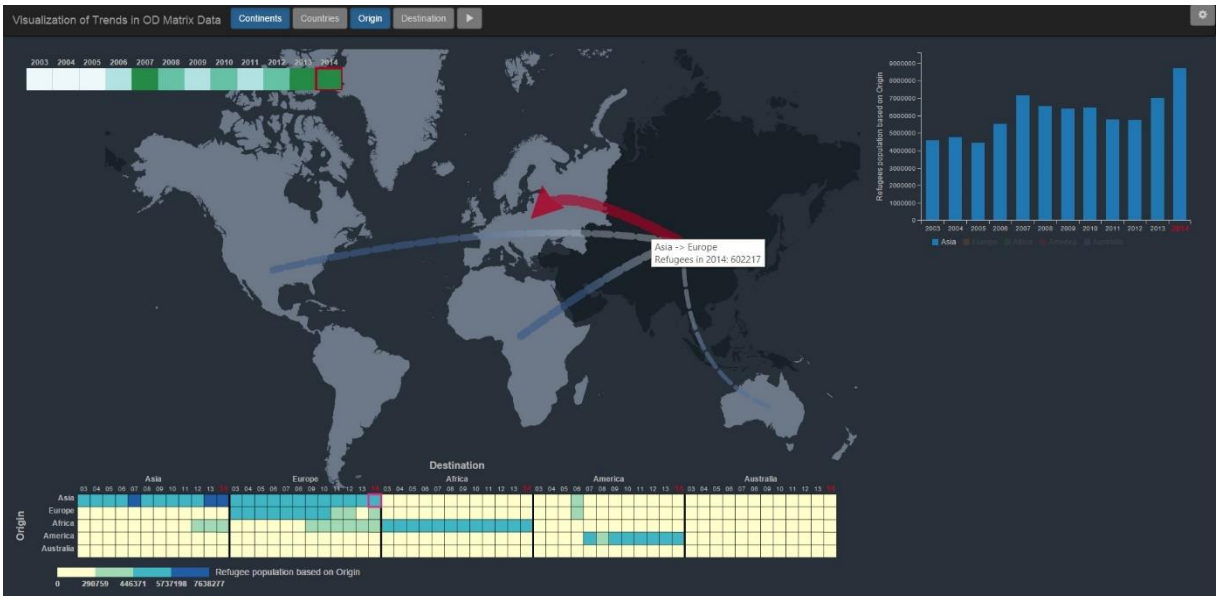


Figure E 2: indicates the selected Origin continent with its interactivity with OD-matrix and stacked bar graph.

APPENDIX F: CARTOGRAM FOR COUNTRIES BASED ON ORIGIN AND DESTINATION WITH THE INTERACTIVITY

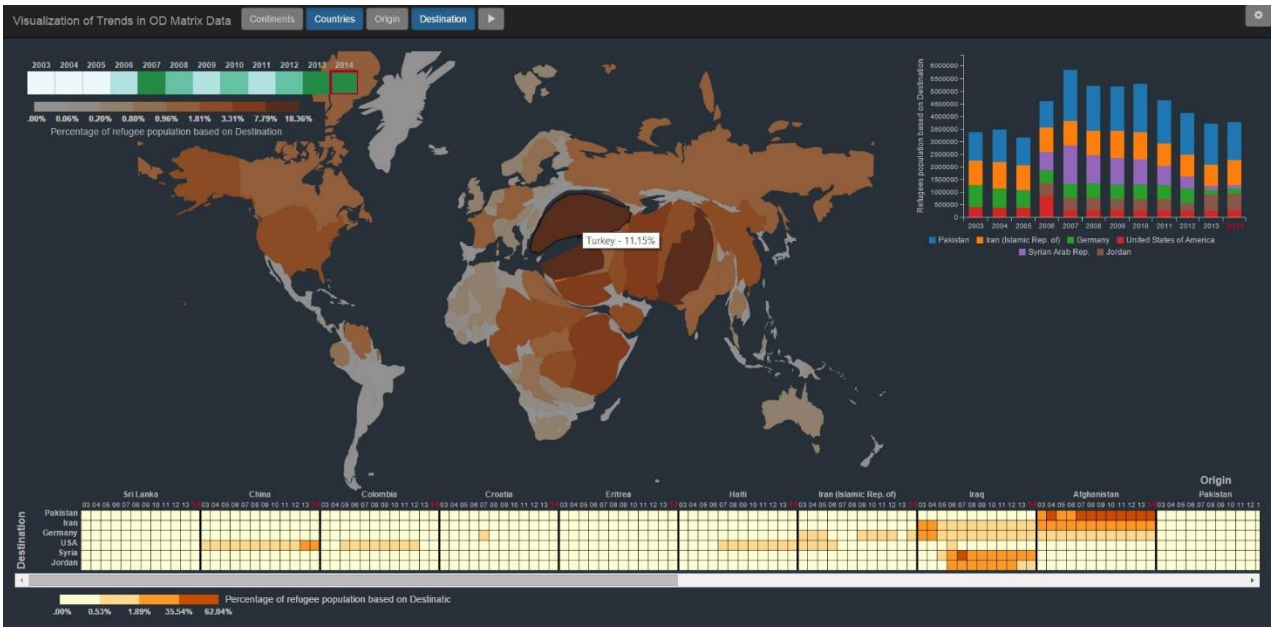


Figure F 1: the countries with cartogram when mouseover

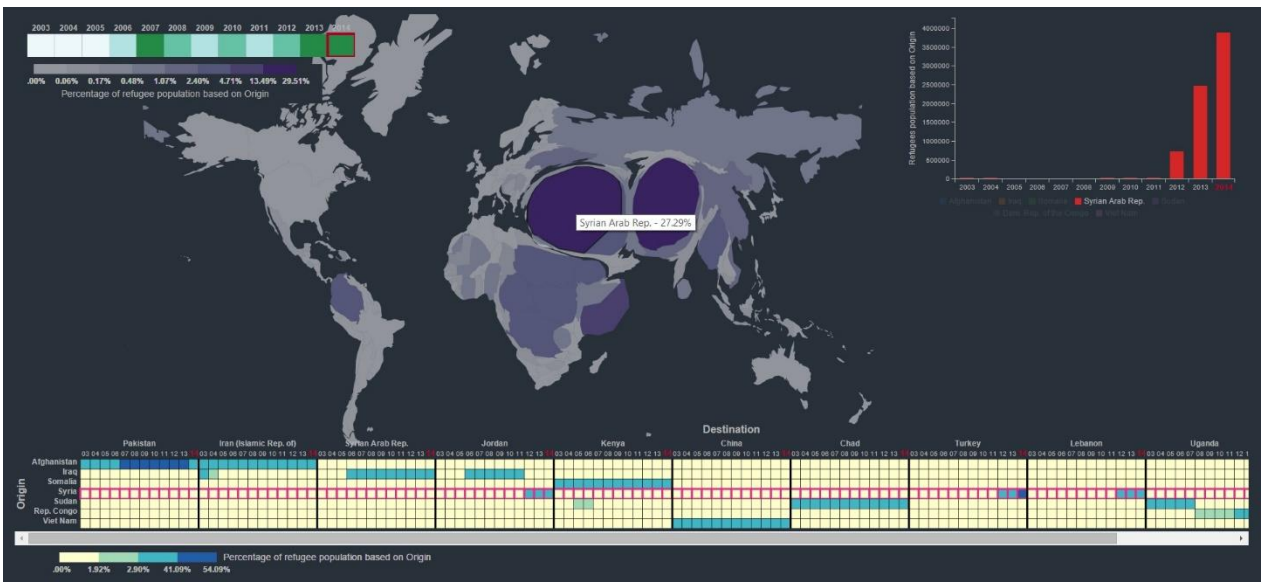


Figure F 2: the selected country with cartogram and its interactivities with stacked bar graph