



Cartography M.Sc.

Master thesis

Enhancement of Density Visualization using Dot Density Maps

Mostafa Kheyrollahi



2020

Enhancement of Density Visualization using Dot Density Maps

submitted for the academic degree of Master of Science (M.Sc.)
conducted at the Department of Aerospace and Geodesy
Technical University of Munich

Author:	Mostafa, Kheyrollahi
Study course:	Cartography M.Sc.
Supervisor:	Murphy, Christian; Dr.-Ing.
Reviewer:	Prof. Dipl.-Phys. Dr.-Ing. habil. Dirk Burghardt

Chair of the Thesis
Assessment Board: Prof. Dr. Liqiu Meng

Date of submission: 10.10.2020

Statement of Authorship

Herewith I declare that I am the sole author of the submitted Master's thesis entitled:

"Enhancement of Density Visualization using Dot Density Maps"

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Munich, 10.10.2020

Mostafa Kheyrollahi

A handwritten signature in black ink, featuring a large, stylized loop at the beginning, followed by the name 'Mostafa Kheyrollahi' written in a cursive script.

Abstract

Density visualization of a phenomenon throughout a space is one of the usages of maps. Color coded heatmaps are a popular approach to visualize density. However, heatmaps do not give a quantitative view to the user and it is not possible to estimate quantitative values from their colors. Dot density maps are considered and adapted as an alternative in order to enhance the density visualization to overcome these issues. Two algorithms are developed to make a conventional dot map and a graduated dot map. A conventional dot map contains dots with equal size, while a graduated dot map has various classes of dot sizes on a regular grid. The resulting raster from running the KDE function contains a large number of cells which was used as a base to create dot maps. This base is a grid input in which dots are placed. The dot placement is pseudo random in the conventional dot map whereas it is non-random in the graduated dot map. Both map types avoid any dot overlap and coalescence, as this is a prerequisite to get a quantitative estimation from a dot density map. Two datasets were used to run the algorithms. A user test was carried out to evaluate the degree of enhancement. The test results show that both types of dot maps are significantly superior to the heatmap not only in presenting the overall density, but also in estimating quantitative values in any favorable part of the map. There is no definite superiority between the conventional and graduated dot maps. The test result stated that both of the maps are helpful based on the dataset, homogeneity of a phenomenon, the aim of the map reader, level of details and calculation time.

Key words: Density, Heatmap, Conventional dot map, Graduated dot map, Quantitative view, Enhancement

Table of Contents

1	Introduction	10
1.1	Motivation and problem statement	10
1.2	Research objectives and questions	11
1.2.1	Research objectives	11
1.2.2	Research questions.....	11
1.3	Outline of the thesis	12
2	Previous research and related works about Dot Mapping and KDE.....	13
2.1	General review of dot mapping methods and applications	13
2.2	A review on graduated dot maps	19
2.3	Related works about KDE	21
3	Research Methodology	27
3.1	Conventional dot mapping	28
3.1.1	The first step.....	29
3.1.2	The second step.....	30
3.1.3	The third step.....	30
3.1.4	The fourth step	32
3.2	Graduated dot mapping	33
3.2.1	The third step.....	33
3.2.2	The fourth step	34
3.2.3	The fifth step.....	35
4	User Test and Results	38
4.1	Results of running the algorithms.....	38
4.1.1	Maps of the POI dataset	38
4.1.2	Maps of the Twitter dataset	46
4.2	User test	50
5	Discussion	59
6	Conclusion.....	62
6.1	Thesis conclusion	62
6.2	Outlook	63
	References	64

Appendix	67
----------------	----

List of Figures

Figure 1.1. Dot density map of industrial workers in Baden-Württemberg, Germany	11
Figure 1.2. Research workflow	12
Figure 2.1 Dot density map of sheep in New Zealand.....	14
Figure 2.2 Mackay Nomograph	15
Figure 2.3 Dot density shading to monitor Topography in all over the world.....	16
Figure 2.4 Geographic distribution of outbreak related acute cases of toxoplasmosis in the Capital	17
Figure 2.5 Plots of the predicted number of randomly placed dots needed to obtain 10, 20, 30 and 40 percent overlap	18
Figure 2.6 Step-by-step population of spiral arms with different stages	19
Figure 2.7 US Oil Consumption per year by state(leistiar, 2012).....	20
Figure 2.8 Graduated dot map with three dot sizes for 200, 1000 and 5000 swine (©Atlas of Switzerland, Sheet 51, 2020).....	21
Figure 2.9 Kernel estimate showing individual kernels	22
Figure 2.10 Result of running KDE on a sample point dataset	23
Figure 2.11 Density curve from running KDE on POI to determine CBD	24
Figure 2.12 Result of running KDE on POI to determine CBD	24
Figure 2.13 An example of cluster heatmap matrix, showing social statistics of some countries	25
Figure 2.14 (a) Original points (b) Heatmap with a disordered rainbow (c) Heatmap with a dark-light color theme (c) Heatmap with a perceptual color range (warm-cold).....	26
Figure 3.1 Flowchart of the proposed approach.....	28
Figure 3.2 A sample heatmap made by the author	30
Figure 3.3 The defined grid around a sample data might cover some no data area	31
Figure 3.4 Pseudo code of conventional dot mapping algorithm	32
Figure 3.5 Workflow of the conventional dot mapping algorithm.....	33
Figure 3.6 The largest dot of the graduated dot map in a cell.....	35
Figure 3.7 Pseudo code of graduated dot mapping algorithm.....	36
Figure 3.8 Workflow of the graduated dot mapping algorithm	37
Figure 4.1 KDE heatmap for Density of POI	40
Figure 4.2 A comparison of dot maps for a more detailed or easier estimation.....	40
Figure 4.3 Conventional dot density map of POIs with dot value of 10.....	41
Figure 4.4 Conventional dot density map of POIs with dot value of 15.....	41
Figure 4.5 Conventional dot density map of POIs with dot value of 20.....	42
Figure 4.6 Conventional dot density map of POIs with dot value of 25.....	42
Figure 4.7 Conventional dot density map of POIs with dot value of 30.....	43
Figure 4.8 Graduated dot density map of POIs.....	44
Figure 4.9 Graduated dot density map of POIs.....	44
Figure 4.10 Graduated dot density map of POIs.....	45

Figure 4.11 Graduated dot density map of POIs.....	45
Figure 4.12 KDE heatmap for density of tweets	46
Figure 4.13 Conventional dot density map of Tweets with dot value of 5	47
Figure 4.14 Conventional dot density map of Tweets with dot value of 10	48
Figure 4.15 Conventional dot density map of Tweets with dot value of 15	48
Figure 4.16 Graduated dot density map of tweets	49
Figure 4.17 Graduated dot density map of tweets	49
Figure 4.18 Graduated dot density map of tweets	50
Figure 4.19 The specific region to quantify in KDE heatmap	51
Figure 4.20 The specific region to quantify in conventional dot map.....	52
Figure 4.21 The specific region to quantify in graduated dot map	52
Figure 4.22 The comparison of the preference of users in the first and second part of the survey	54
Figure 4.23 Individual results of the third part of the survey	57
Figure 4.24 The chart of the results of the user test for the third part of the survey	58
Figure 5.1 A graduated dot map for a detailed estimation	60
Figure 5.2 A conventional dot map for a detailed estimation	61

List of Tables

Table 4.1 Detailed parameters of made dot maps	39
Table 4.2 The results of the user test for the first part of the survey	53
Table 4.3 The results of the user test for the second part of the survey	53
Table 4.4 The comparison results in the third part of the survey	55
Table 4.5 The estimation results in the third part of the survey	55
Table 4.6 The results of the user test for the fourth part of the survey	58
Table 5.1 Advantages and disadvantages of both types of developed dot maps	61

1 Introduction

1.1 Motivation and problem statement

Cartographers have always tried to find a way to visualize a phenomenon in the best possible way. Their goal is to find the best method of mapmaking to monitor their data. There are several design styles of thematic mapping techniques such as choropleth map, proportional symbol map, dot map, heatmap, etc. Density estimation is one of the topics in which cartographers are interested. One of the applications of maps is to monitor the density of a phenomenon throughout an area.

Points of a phenomenon can be presented as a continuous function to create an effective and accurate impression of the incident distribution (Krisp & Spatenkova, 2010). For this purpose, interpolation methods such as Kernel Density Estimation (KDE) could be used to make another type of map which might enhance the visualization. KDE-based heatmap is one of the most commonly used tools for spatial point pattern visualization and analysis (Yuan et al., 2019). KDE tries to make the distribution of phenomena more continuous so that the user perceives them more accurately (Krisp & Spatenkova, 2010). The result of KDE is a raster which is known as a heatmap. Heatmaps use a relatively warm or cold color range within isolines to present the density distribution of point data (Duchowski et al., 2012), (Wilkinson & Friendly, 2009).

A heatmap made by KDE is considered as a usual density visualization method. However, there are still some issues with this method. Although a heatmap is a helpful approach to smoothing the input point data, there are still some visualization issues with this output. Firstly, the color scheme does not give a quantitative estimation to the user. In other words, there is no estimation of density differences between color coded classes. Secondly, there is no perception for the actual value of a phenomenon since final result of KDE is a cryptic density value.

Dot maps are useful for displaying data when an underlying phenomenon is not uniform throughout an area (Arnold et al., 2017). A dot map is a useful approach to show density. Figure 1.1 is an example of a dot map which shows density of industrial workers in the state of Baden-Württemberg in Germany. Regions with more dots show a higher density of workers and emptier regions show less density of industrial labor forces.

This thesis focuses mainly on improving visualization of density of a phenomenon by a dot map. The main purpose of the research is to remove drawbacks of this density visualization approach.

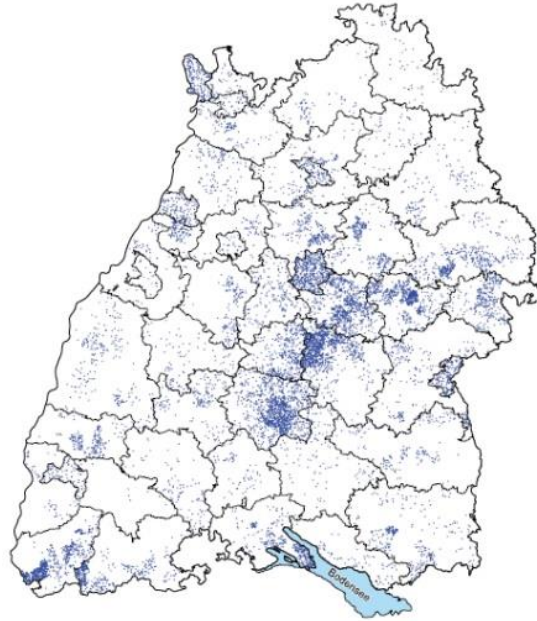


Figure 1.1. Dot density map of industrial workers in Baden-Württemberg, Germany
(GfK GeoMarketing GmbH, n.d.)

1.2 Research objectives and questions

1.2.1 Research objectives

The main objective of this thesis is to overcome the two main issues mentioned in section 1.1. Using KDE heatmap to make a dot map is suggested as a way to achieve this goal. The main task is to use the values of KDE output to place the required dots on a dot map. The results of the KDE function will be an input to the algorithms. The algorithm's output would be a conventional dot map to show density of a phenomenon. With some changes in the proposed algorithm, it is possible to develop a second algorithm to make a graduated dot map from the results of KDE. Then, it will be possible to compare the results of the two algorithms. The developed algorithms are used to show the density of points of interest in Munich, Germany. Also, the algorithms have been run on a twitter dataset related to a German traditional festival, Oktoberfest, in Munich.

1.2.2 Research questions

To achieve the research objectives, the following research questions need to be answered:

- Which criteria should be considered to make a dot map which reflects a density in the best possible way?
- What is the difference between the KDE heatmaps and dot maps in quantization and density presentation?
- Do final dot maps provide quantitative information to the user?
- What are the advantages and disadvantages of graduated and conventional dot maps?

1.3 Outline of the thesis

The thesis is structured with five chapters. The second chapter focuses mainly on the previously related works to different aspects of the research. The third chapter is the methodology of the research. In this chapter, all the steps of the project are presented in detail. The developed algorithms will be described completely in this chapter. The fourth chapter introduces the used data, the results of running algorithms on the data and the evaluation of the results which is conducted by a couple of users. The fifth chapter will be a detailed discussion about different parts of the research and the sixth and the last chapter is a conclusion. The research outline is described in a workflow below (Figure 1.2)

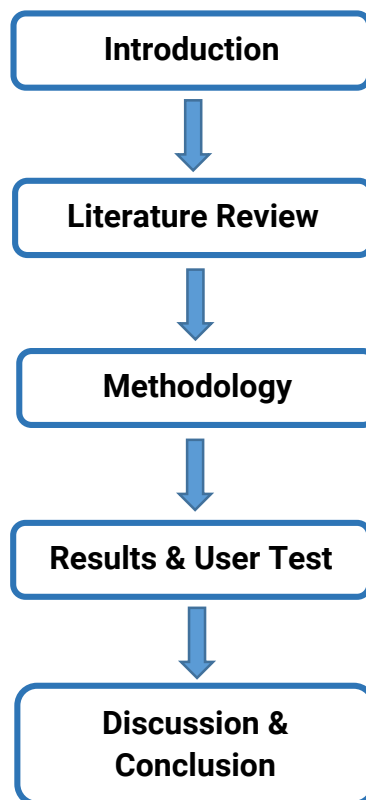


Figure 1.2. Research workflow

2 Previous research and related works about Dot Mapping and KDE

In this chapter, studies related to different aspects of the thesis are reviewed. The first part focuses on dot maps and their general application. The second section evaluates works about graduated dot maps. The last part of this chapter discusses papers about KDE.

2.1 General review of dot mapping methods and applications

Axismaps (2020) states that dot density maps are an old, simple and effective way of presenting density differences in geographic distributions across a landscape in which one dot stands for one or a number of things or a value. Three main advantages of dot maps over choropleth and color coded maps are: 1) it is possible to map counts or raw data (e.g., number of points of interest) and rates and ratios (e.g., number of points of interest per square kilometers); 2) no need to be tied to enumeration units; 3) no need to use color, as black and white is sufficient. The author provides a good example of a dot map (Figure 2.1), showing the density of sheep in New Zealand. The density of sheep is shown to be higher on the eastern side of the country.

Three categories of data can be used for dot mapping:

- Distribution of point data, e.g., distribution of wells in an agricultural area
- Distribution of a population within an area, e.g., distribution of homeless people in different districts of a city
- Distribution of areal data, e.g., distribution of corn harvested lands in hectare units in a farming region

Robinson et al. (1995) state that **dot size** and **unit value** are the two important parameters of a dot map. Too small dots lead to a sparse and insignificant distribution and too large dots show an excessive density. Very small unit values lead to dot overlapping and dark shaded regions, while a map with too large unit values cannot place dots in areas with low quantity (Mackay, 1949).

Slocum et al. (2009) conducted research about all aspects of thematic cartography and geovisualization. The seventeenth chapter of the book is dedicated to dot maps. The authors explained the principles of dot maps, including graduated and conventional dot maps in detail. According to this book (Slocum et al., 2009), dot maps are a good way to show a phenomenon which is not uniform throughout the enumeration units (unit values). One dot equals a certain amount of some phenomenon and is placed where that phenomenon is more likely to happen. In addition to dot size and unit values, they bring to attention two other parameters that are significant in dot mapping. The first parameter is the determination of regions in which dots should be placed. For example, a sea cannot

contain any dot when monitoring wheat lands. The second is dot placement. The book introduces some methods and formulas to calculate the placement of dots in dot maps. A number of these approaches are used in this thesis.



Figure 2.1 Dot density map of sheep in New Zealand

Mackey (1949) proposed one of the first methods of dot mapping. He proposed a graphical nomograph called *Mackey nomograph* to select the dot size and dot value. The primary step of using this graph is to calculate the number of dots per square centimeter on a map. The main components of the graph are a vertical and an oblique line. The former line represents a relatively dense area on the map and the second line is drawn from the origin to a desired dot size. Figure 2.2 shows the Mackey nomograph. However, a number of studies, such as Kimerling (2009), have highlighted disadvantages of this method. A detailed example is provided in the next paragraphs.

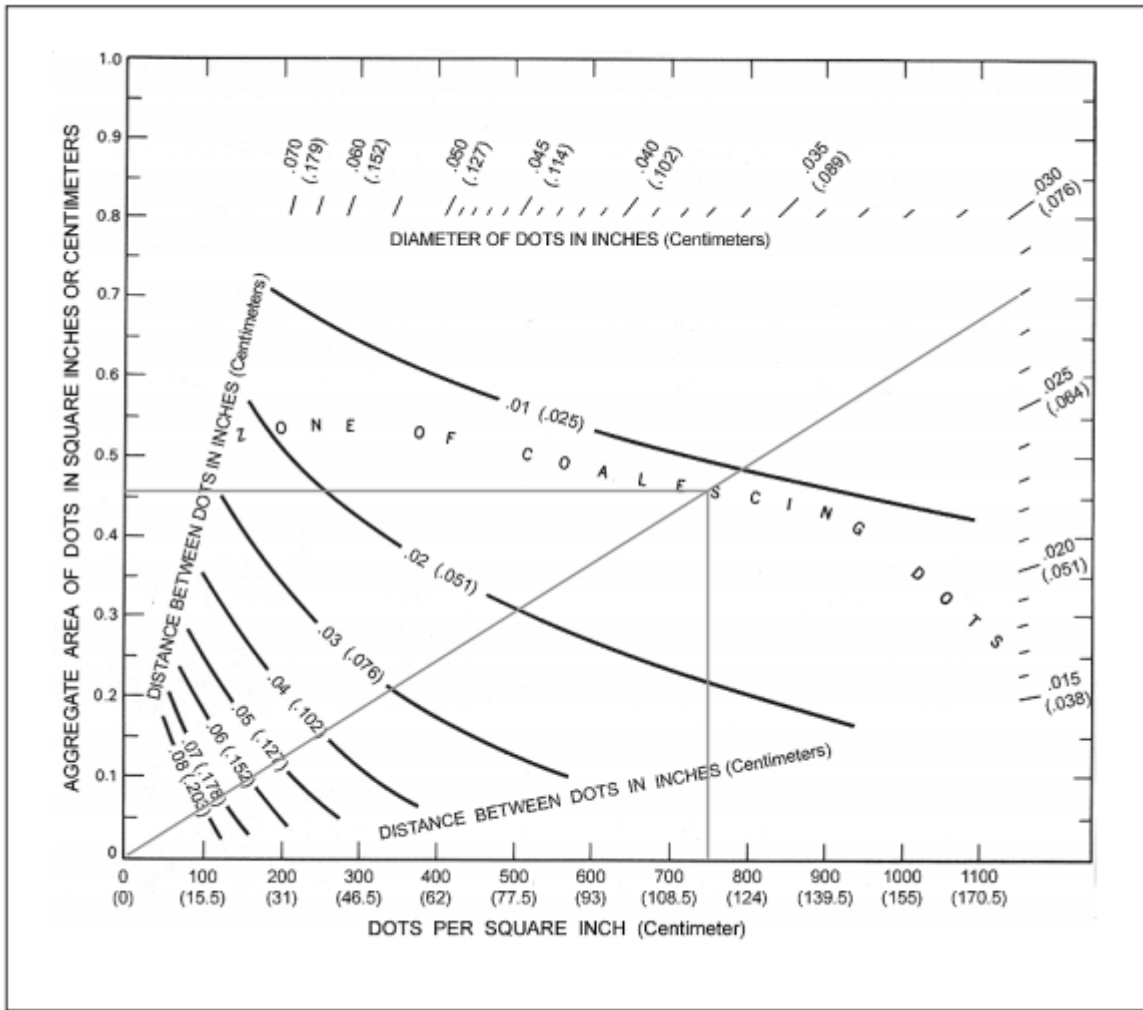


Figure 2.2 Mackay Nomograph

Lavin et al. (1986) are among the pioneer researchers who worked on the automatic placement of dots. The authors looked for an alternative approach to isolines to display continuous variables. They asserted that isolines present precise values, which are not useful for displaying spatial distribution. They proposed dot density shading as a technique which is much more practical than isolines. In this case, they used dot density shading to display the spatial distribution of continuous climatological variables for Colorado. The authors proposed an approach or a formula for placing dots, known as Lavin's dot density shading method. Lavin's method depends mainly on dividing the region into a grid of cells. Figure 2.3 shows the result of this approach, in which dot density shading is used to monitor topography all over the world. Denser areas pertain to more mountainous regions. As is obvious from the result, no attention is paid to dot coalescence and dot value. Density shading is the only purpose of this method.

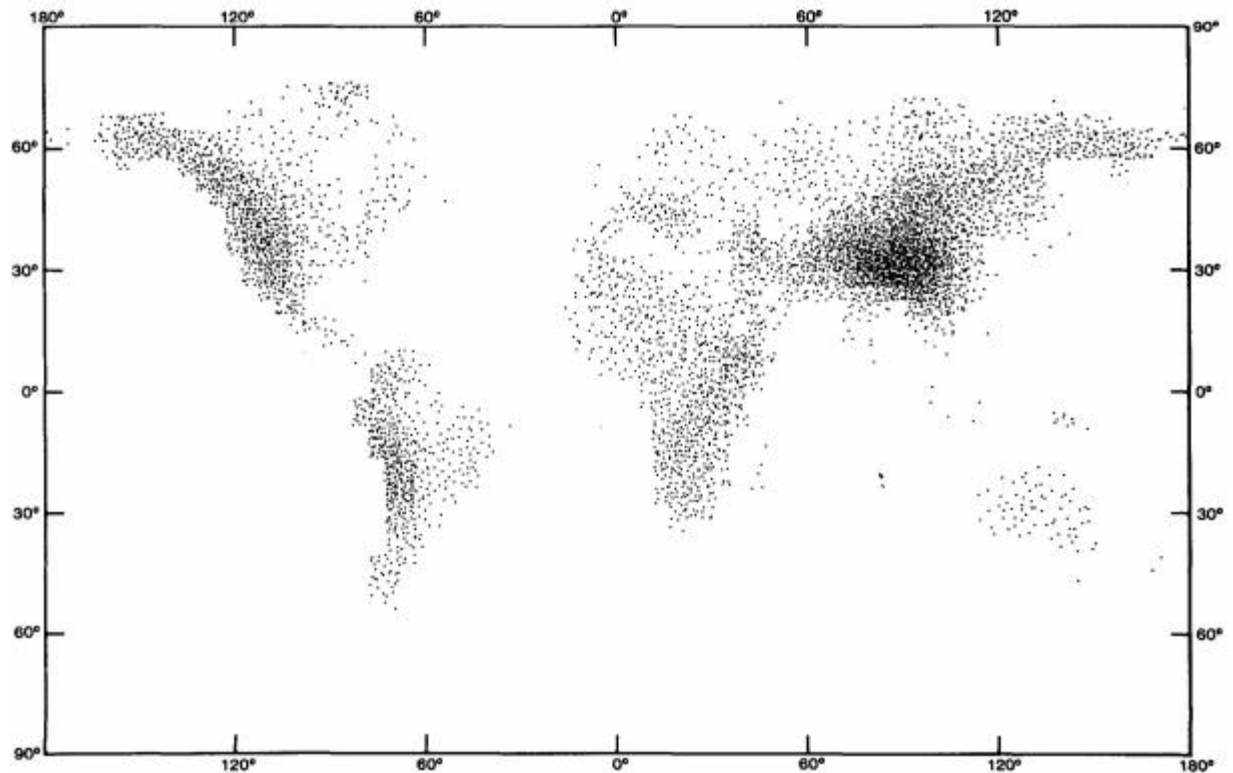


Figure 2.3 Dot density shading to monitor Topography in all over the world

Slocum et al. (2009) proposed a more complete approach based on Lavin's dot density shading method. This technique is used in one of the algorithms of this thesis. There are four major steps of this approach. Firstly, an equally spaced gridded network of data values should be defined over the studied area. Secondly, the values of each cell of the grid should be normalized and the percentage of occupied space by dots for each cell should be determined. Thirdly, the size and number of dots in each cell should be calculated. In the last step, dots are placed in each cell separately and randomly. If the cartographer sets the size of the dots correctly, it is possible to avoid overlap and coalescence.

Eng et al. (1999) used computer generated dot maps to examine the spatial distribution of an infection called *Toxoplasma gondii* in British Columbia, Canada. They used computer-generated dot maps to monitor the potential sources of infection and to analyze the data. Municipal water sources were critical to the evaluation because they were suspected as the source of the outbreak. The addresses of all pregnant women were placed on the map of the region. The registered women were classified to three types: *acute* cases, *never infected* and *non-acute* cases. The probability of the effect of the suspicious water reservoir was calculated by the density of cases in the area served by that reservoir in comparison to other areas. To eliminate other potential causes, age of the women was taken into account. The age distribution of the women in these analyses did not differ when grouped by infection status. According to this research,

using computer generated dot maps had five benefits: easier verifications and placement of addresses, saving time, analyzing several sets of data, enabling queries and printouts or overheads for presentation. Figure 2.4 shows the density of acute cases in the whole region and the region served by the suspicious water reservoir.

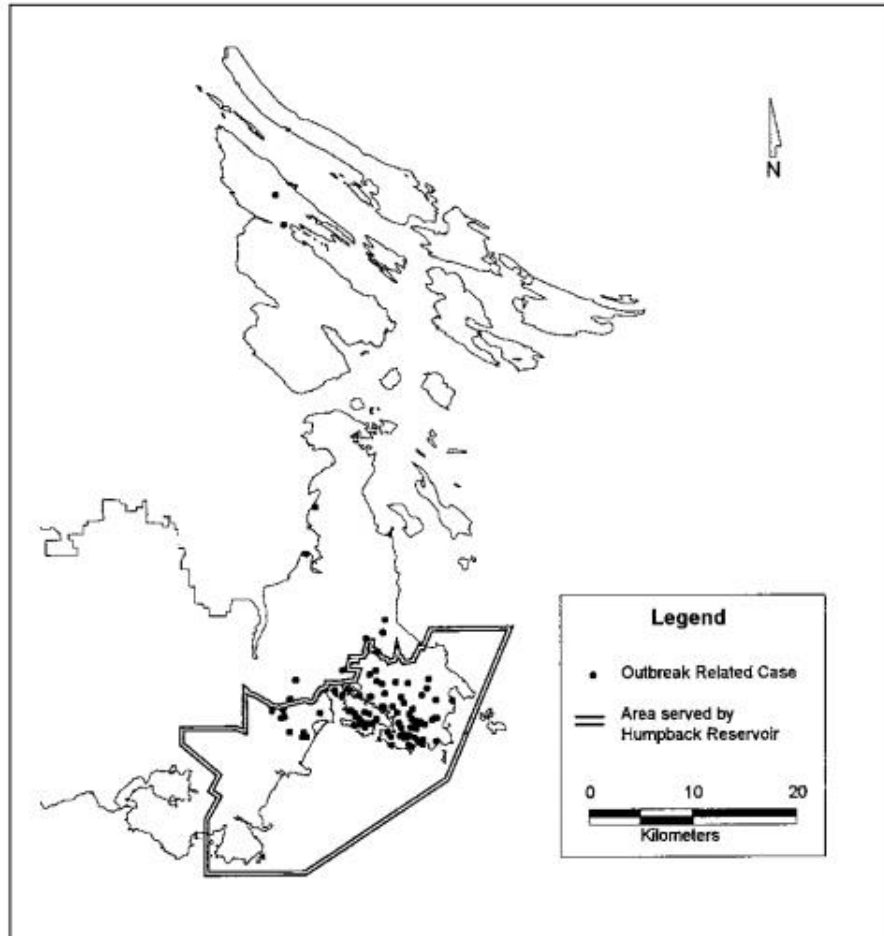


Figure 2.4 Geographic distribution of outbreak related acute cases of toxoplasmosis in the Capital

Kimerling (2009) mainly worked on reducing overlap of dots in dot maps. He believed that Mackey's nomograph is not efficient to avoid overlap. A new graphical approach has been proposed to calculate the dot unit value, which is created by modeling aggregate area of dots. The main focus of this research was to decrease dot overlaps. The studied region is divided into cells of unit areas and the covered area within each cell is calculated. There are two useful hints in this research. Firstly, there is an inverse relation between the size and the number of dots within a cell. Secondly, overlap between dots has a direct correlation with the percent of the aggregated covered area within a cell. Figure 2.5 presents these two hints very well. The researcher reached the conclusion that pseudo ransom dot placement has a big superiority over totally random dot placement. Their graphical comparison between these two shows that the more the dot placement is pseudo random, the less is the overlap among dots.

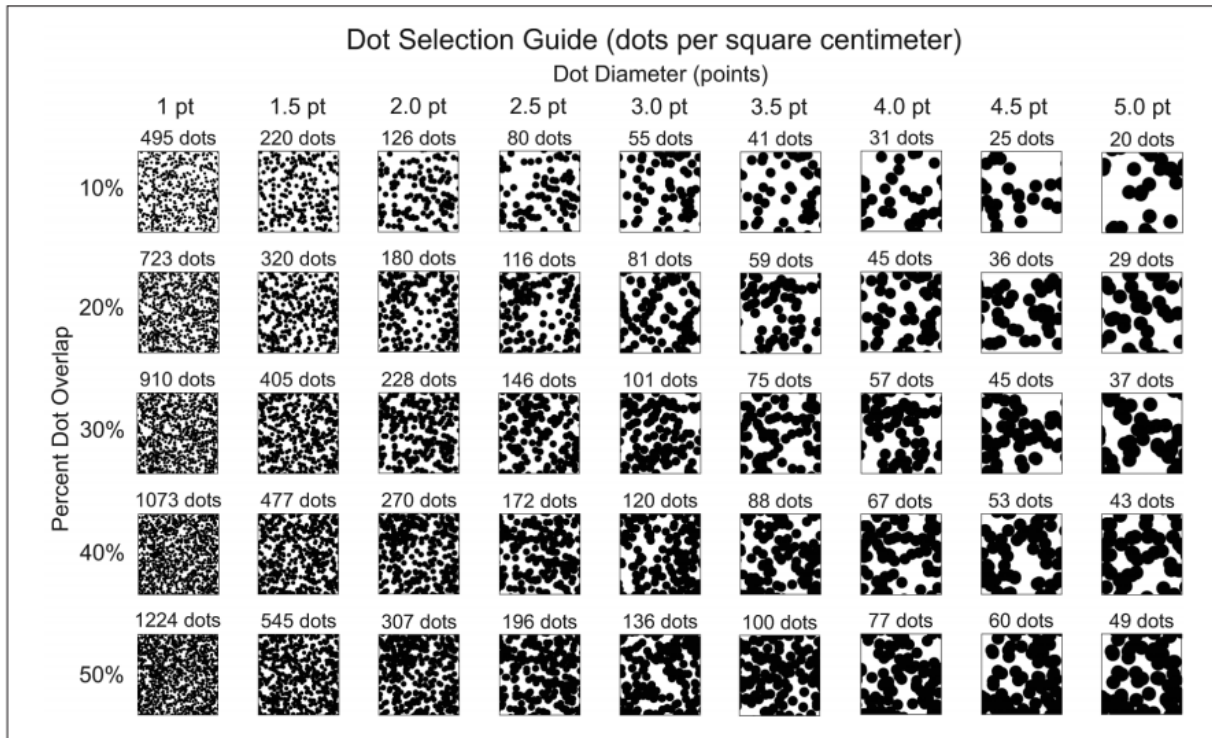


Figure 2.5 Plots of the predicted number of randomly placed dots needed to obtain 10, 20, 30 and 40 percent overlap

Hey (2012) questioned the previous approaches for placing dots for two reasons. Firstly, she claims that a random placement of dots avoids repetitive results in different times of making a dot map from the same dataset. Secondly, overlaps among dots is another problem for counting dots and getting a quantitative overview of spatial distribution. She proposes a new and complex method named the *Spiral Whirl* technique. The Spiral Whirl is a center based approach in which dot clusters group around one central point, e.g., the centroid of an enumeration area. According to the author, the main advantage of this approach is that the maximum number of dots with a specific dot size can be placed in an area. Therefore, placing dots would be in the most efficient way possible. Figure 2.6 presents a short summary of dot placement by the Spiral Whirl.

Hey and Bill (2014) built upon the previously mentioned research conducted by Hey in 2012. The improved Spiral Whirl technique is easy to use, does not need any new software, supports every map symbol, distribution and exclusion areas are included and the output shape-file can be combined with other map elements using a GIS to form a complete dot map.

After a review of some previous works and studies on principle, methods and applications of dot mapping, we need to consider useful hints from the previous works in the thesis. The main tips which should be considered for making a dot map are as follows:

- Four main parameters of dot maps are:

- Dot value
- Dot size
- Dot placement
- Determination of regions for placement
- The size of dots has an inverse relation with the number of dots in a cell.
- Increasing covered area by dots in a cell leads to more overlap of dots.
- Pseudo random dot placement has a big superiority over totally random dot placement.

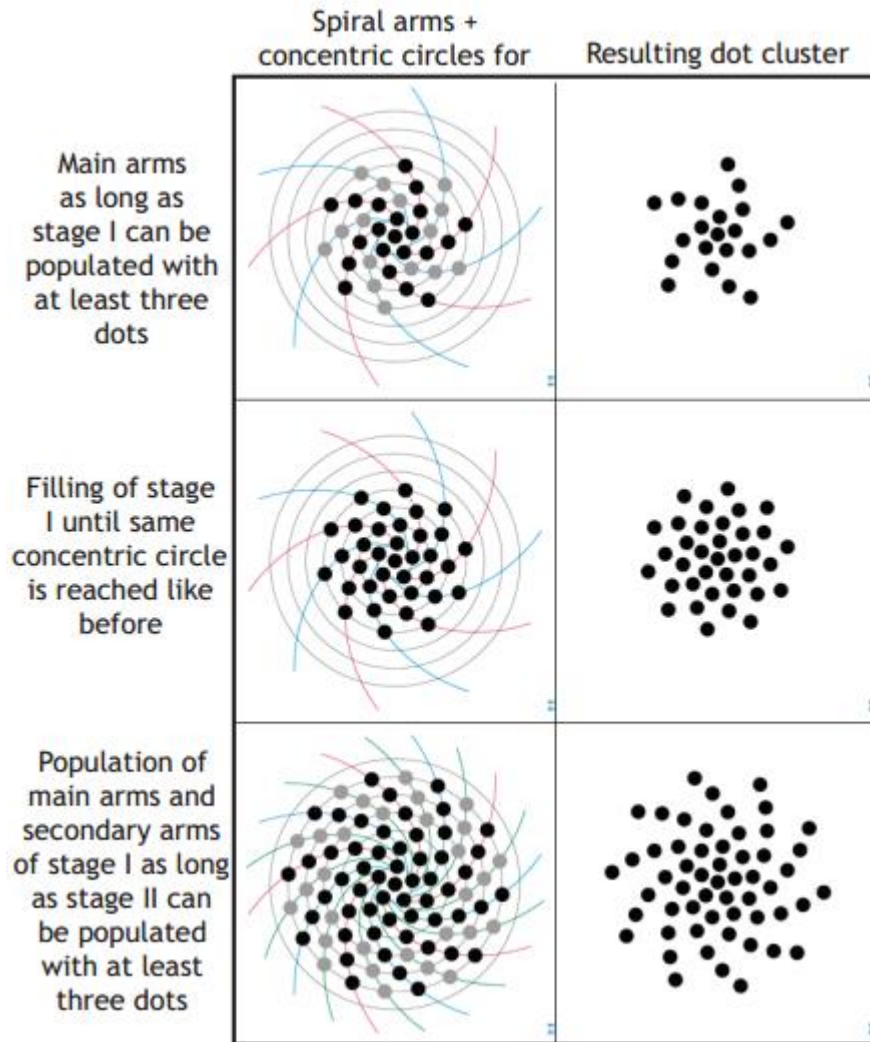


Figure 2.6 Step-by-step population of spiral arms with different stages

2.2 A review on graduated dot maps

Slocum et al. (2009) worked on the concept and application of proportional symbol maps. This type of map focuses on raw data as they are collected at point locations or data over areas which can be located at a point. Although a proportional symbol map is different from a graduated dot map, a part of their formula to set the size of symbols can be used

for a graduated dot map. In the formula, a size would be assigned for the largest circle (symbol) and the size of other circles is defined based on the value which the circle is supposed to present. The relation of the size of a circle to the largest circle is equal to the relation of its value to the value of the largest circle. The technique of using this formula for graduated dot map is explained in the third chapter. Figure 2.7 is an example of proportional symbol map which shows US oil consumption per year by state.

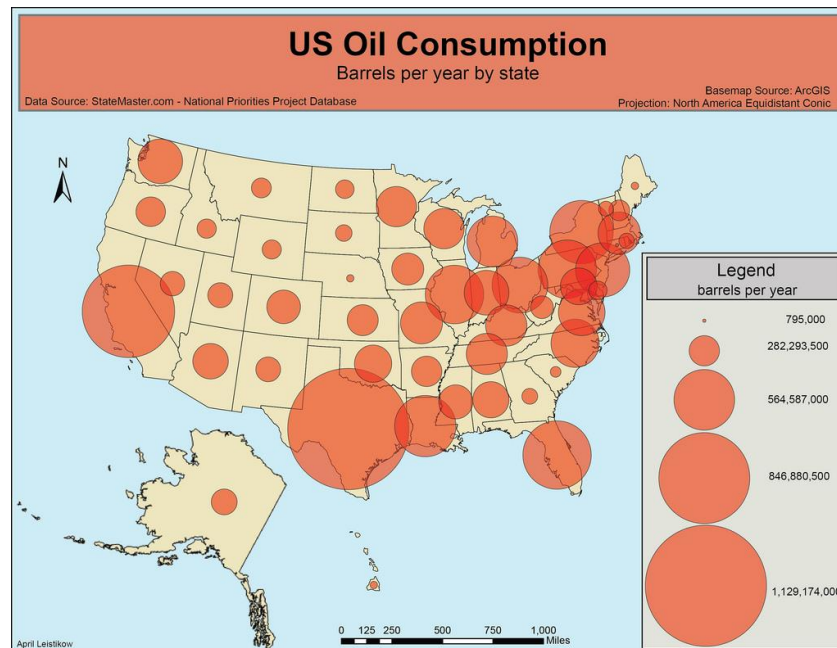


Figure 2.7 US Oil Consumption per year by state(leistiar, 2012)

Arnold et al. (2017) stated that graduated dot maps use dots with different sizes that represent different values. With graduated dot maps, the number of dots on a map is fewer, reducing the likelihood of overlapping dots. Figure 2.8 shows a graduated dot map with three dot sizes for 200, 1000 and 5000 swine. Arnold et al. (2017) questioned some popular dot map techniques such as pseudo random dot maps and indicated that these methods have some weaknesses such as much overlap and difficulties for users to estimate values. Since the number of dots is lower in graduated dot maps, the likelihood of overlapping is less. Based on a user test with almost 300 participants, Arnold et al. (2017) found that graduated dot maps give more accurate quantitative overview to a user than conventional dot maps, and area-proportional circle maps. However, Arnold et al. (2017) agree that if the only goal is to get the spatial density patterns, a conventional dot map is appropriate.

Gomez (2017) made population dot density maps of some areas, including mainland Portugal, Oporto and Lisbon from statistical data of the 2011 Census. He used dots with constant values and sizes for less than a threshold of population and also proportional circles for more than that threshold. The output of the investigation displays the density of population throughout the mentioned studied area very well. However, it seems that

not enough attention was paid to overlaps among dots, which show the population less than the threshold.



Figure 2.8 Graduated dot map with three dot sizes for 200, 1000 and 5000 swine (©Atlas of Switzerland, Sheet 51, 2020)

Following this brief review, we may reach several conclusions about graduated dot maps:

- There is a difference between proportional symbol maps and graduated dot maps.
- Graduated dot maps can be used instead of conventional dot maps to reduce the number of dots, reduce overlapping and give a more quantitative view to the user.
- Graduated dot maps give more accurate quantitative overview to a user than conventional dot maps, and proportional circle (symbol) maps.
- If the only goal is to get the spatial density patterns, conventional dot map is appropriate.

2.3 Related works about KDE

Silverman (1986) states that *probability density function* is a fundamental concept in statistics. Any random quantity X that has probability density function f which gives a natural description of X and allows probabilities associated with X to be found from the relation:

$$P(a < X < b) = \int_a^b f(x)dx \quad \text{for all } a < b \quad \text{Eq. (2.1)}$$

According to Silverman (1986), density estimation is the construction of an estimate of the density function from the observed data. Tang et al. (2016) defined KDE as a method to analyze distribution of point and linear features over 2-D planar space. Terrell and Scott (1992) mentioned equation 2.2 as the base equation for KDE as follows:

$$f(y) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - y}{h}\right) \quad \text{Eq. (2.2)}$$

where h is the window width, also called smoothing parameter or bandwidth. n is the random sample size, $(x_1, x_2, x_3, \dots, x_n)$ is a univariate independent and identically distributed sample drawn from some distribution and K is the kernel.

Figure 2.9 from Silverman (1986) shows an example of running KDE function on a sample size of seven and a band width of 0.4. There are seven curves with the same size which show all of the seven observations. h in equation 2.2 determines the width of curves and K determines the shape of each curve. Function f constructs the density curve (the large one) by adding all of the curves up. The density curve is higher in the area in which the density of sample curves is higher (in the middle) and it is lower when the density of sample curves is less (in the right and left). To clarify the concept, Figure 2.9 shows the result of running KDE function for seven samples. When running KDE function on a dataset, each element of a dataset gets a curve with the same size (size of curves depend on K and h) and a density curve is constructed to show the density in the whole dataset.

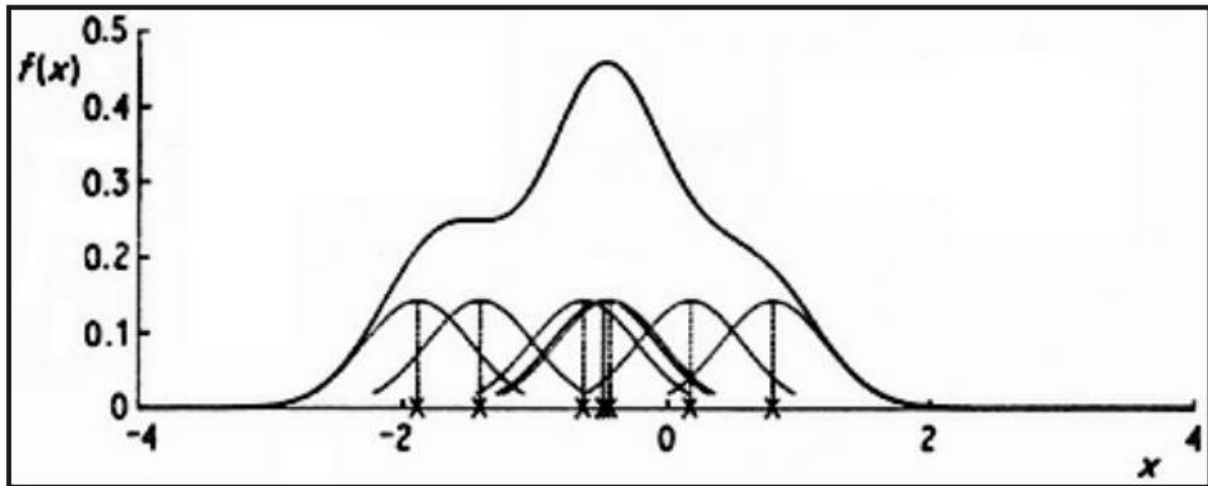


Figure 2.9 Kernel estimate showing individual kernels

Yu et al. (2015) conducted research to show the result of running KDE function on a dataset. The authors are interested in finding the central business district (CBD) in Shenzhen and Guangzhou cities in China. They used KDE to determine density of points of interest in the whole city. Figure 2.10 is the result of running KDE on an assumptive point data set. Based on the explanations in the previous paragraph and Figure 2.9, 15 equal-sized curves would be constructed for this sample dataset because there are 15 points in Figure 2.10. Based on the constructed density curve after running the function,

the higher the density curve is, the darker the heatmap would be in this visualization example. For example, the density curve is much higher around the three close points in the lower left quadrant of the picture than the point in the upper right corner.

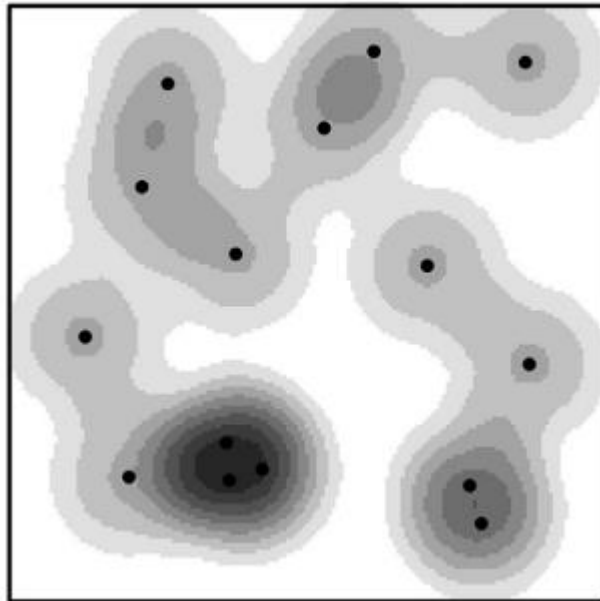


Figure 2.10 Result of running KDE on a sample point dataset

Figure 2.11 shows the density curve for the studied area. This density curve is the same as the explained case in Figure 2.9 but for a real dataset. The density curve is higher in the regions in which the density of business points is higher. Figure 2.12 is the result of their research on the mentioned project. In this paper, the final result of running KDE function is visualized with a heatmap with some isolines which separates different colors and values of the heatmap. Higher parts of the density curve in Figure 2.11 lead to darker parts of the heatmap in Figure 2.12.

Wilkinson and Friendly (2009) worked on cluster heatmaps. According to their research, heatmap uses a relative warm and cold/light and dark colored continuous surface to reflect the density distribution of point data. The heart of the heatmap is a color-shaded matrix. A data matrix with some rows and columns represents a heatmap. Each element of the matrix contains a value. Figure 2.13 is an example of cluster heatmap matrix, showing social statistics of some countries (i.e., urbanization, literacy, GDP, etc.) from a United Nations survey of world countries. The color of each cell depends on its value. Warmer cells show more values and colder ones show fewer values.

Li et al. (2014) visualized a large point dataset interactively. They used KDE to estimate point distribution throughout the area and presented a density image. The final result is a heatmap, displayed interactively in three different color schemes, including rainbow, dark-light and warm-cold. The purpose of this variety is to compare efficiency of each theme category. Borland and Tailor (2007) found that rainbow (Figure 2.14.b) was

harmful for visualization because of its lack of clear distribution. Given that rainbow is disordered, it can easily confuse users. The overall process of work of Li et al. (2014) was

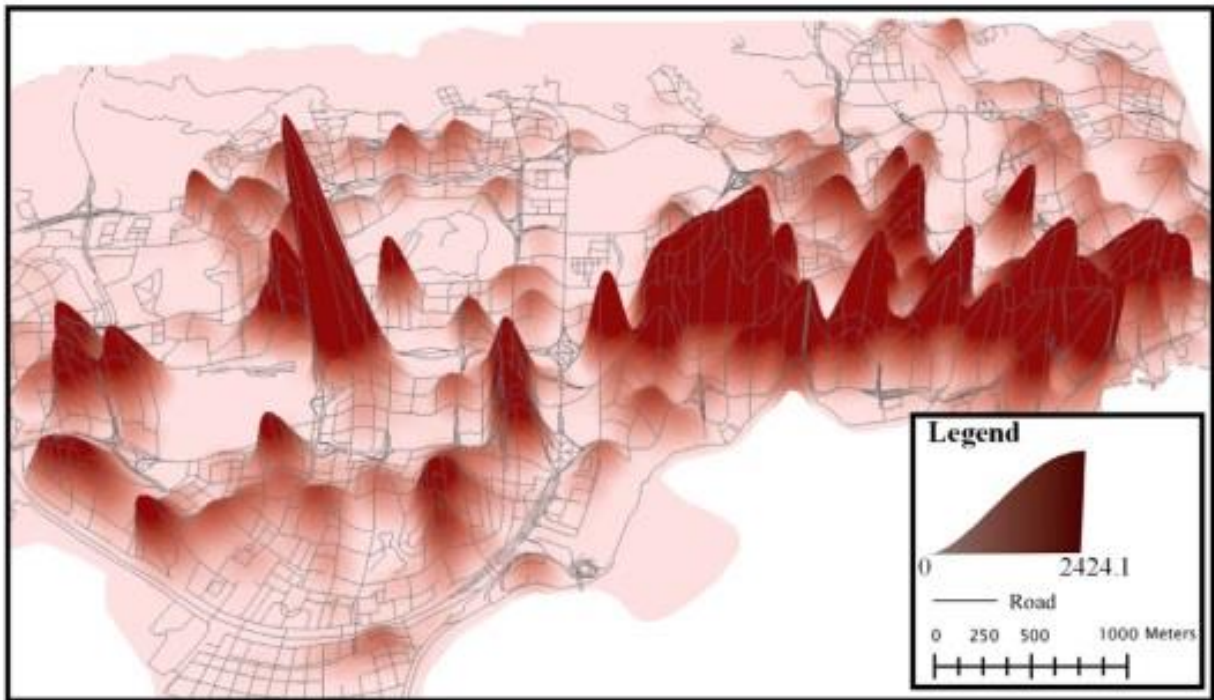


Figure 2.11 Density curve from running KDE on POI to determine CBD

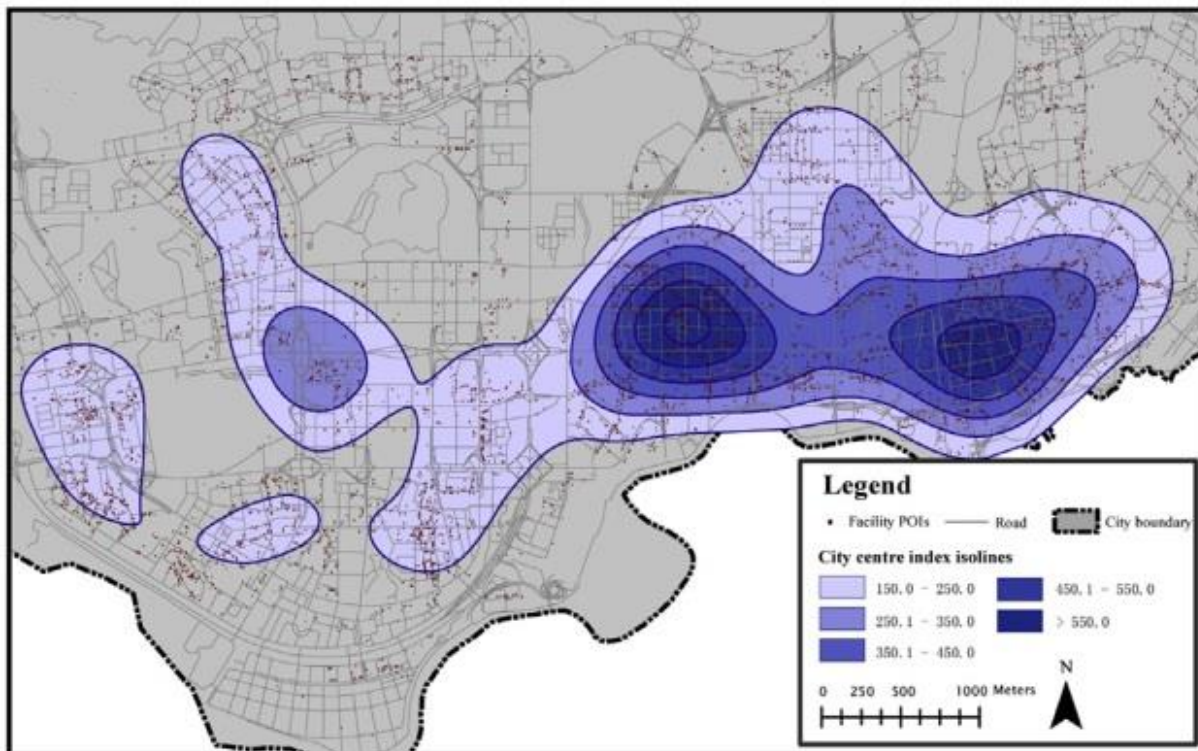


Figure 2.12 Result of running KDE on POI to determine CBD

almost the same as the introduced works, but instead the visualization tool is interactive web maps. Figure 2.14 shows three sample heatmaps (b, c, d) with different color schemes, which shows the density of the original sample points data (a).

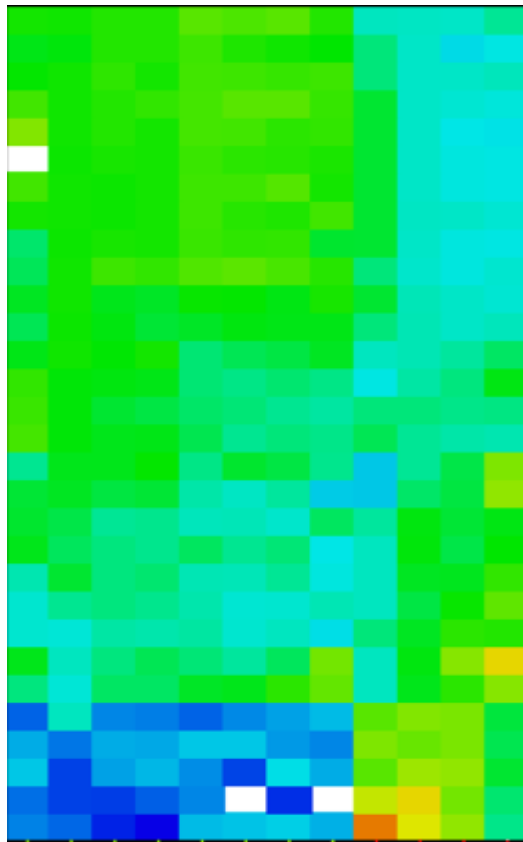
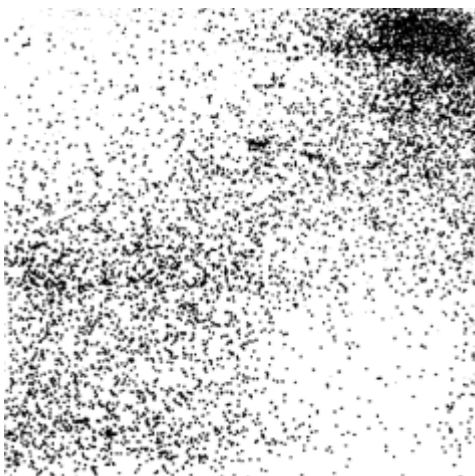
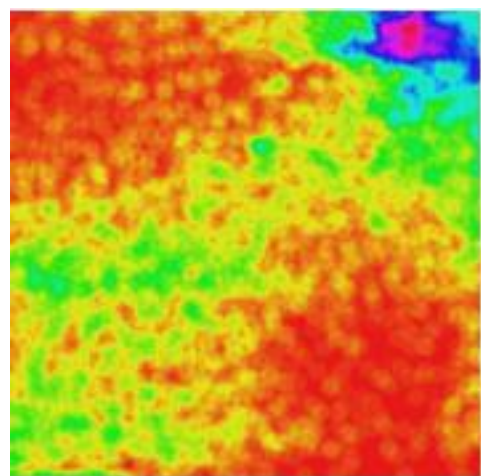


Figure 2.13 An example of cluster heatmap matrix, showing social statistics of some countries

a)



b)



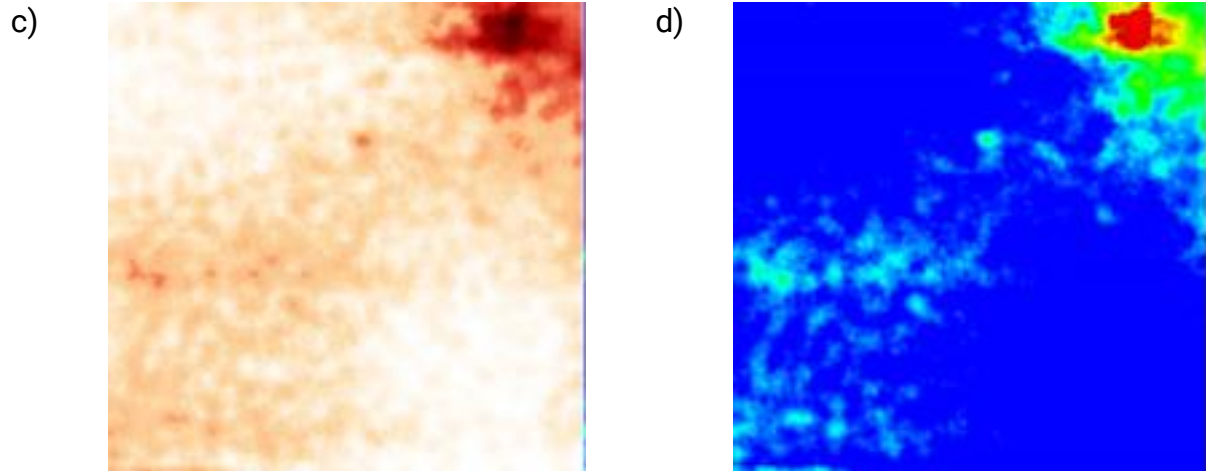


Figure 2.14 (a) Original points (b) Heatmap with a disordered rainbow (c) Heatmap with a dark-light color theme (d) Heatmap with a perceptual color range (warm-cold)

Spatenkova and Krisp (2010) investigated to find a solution to display fire incident distribution in Helsinki to support rescue preparedness planning and rescue allocation. All of the incident points are presented by point data with an exact location. By applying KDE function on the mentioned phenomenon, it is possible to present the density of incidents all over the city smoothly and continuously.

In this section, a review was conducted to perceive the principle, result and application of KDE. The final conclusions about it are as follows:

- KDE is a non-parametric way to estimate the probability density function of a random variable.
- It is used for point or linear datasets.
- The result can be a heatmap with a range of colors and some isolines.
- The product of KDE is a matrix (raster) with some pixels with a specific probability value.

3 Research Methodology

In this chapter, the workflow of the thesis is explained in detail. The scientific workflow, the developed algorithms and used software and platform are described. The chapter has two main parts. The first part reports the process of making a conventional dot map and the second section focuses on the graduated dot map.

As was discussed in the first chapter, the main purpose of this thesis was to overcome the quantitative issues with the heatmap resulting from running KDE on a dataset to visualize the density. Dot map was chosen to solve this issue. Based on the discussions in the previous chapter, the final dot map should have three standard characteristics as follows:

- Dot size and dot value should be determined.
- Dot coalesce and overlap should be avoided.
- Random dot placement should be avoided as much as possible and regular or at least pseudo random approaches should be used.

Splitting the studied area into a regular grid seemed to be a good way to start making the dot mapping more regular. There were some benefits with dot mapping in a grid framework as follows:

- Dot placement would be done separately in each cell of the grid, which is helpful to get to a less random result.
- The size of each cell is a limitation which prevents placing too large dots.
- It is possible to set the proportion of dot coverage in each cell and preventing too crowded cells, which was helpful to avoid dot coalesce.

The dot number and size should have been gained from a source to visualize the density of a phenomenon correctly. Although the heatmap resulting from running KDE function on a dataset had some shortages to visualize the density, it contained the mentioned prerequisites to make a regular or pseudo random dot map. The output of KDE was a raster which contained a huge number of equal size pixels and each pixel had a value. Thus, the KDE function provided a helpful input to make a dot map based on the proposed approach in this thesis. The raster would have been the required grid with many pixels based on its resolution. This raster provided the needed cells and the value of each cell could determine the required dot number and size. A grid was useful to make both categories of dot maps because both dot size and dot value were fixable by this approach.

Both conventional and graduated dot maps have advantages and disadvantages to visualize density. A user test was necessary to evaluate whether the dot maps could solve the mentioned issues. It assessed whether the dot maps give a better quantitative

estimation than the heatmap. Apart from that, users were required to compare graduated and conventional dot maps. Additionally, they were asked to compare dot maps with different dot sizes in each category separately. Figure 3.1 presents the flowchart of the proposed approach.

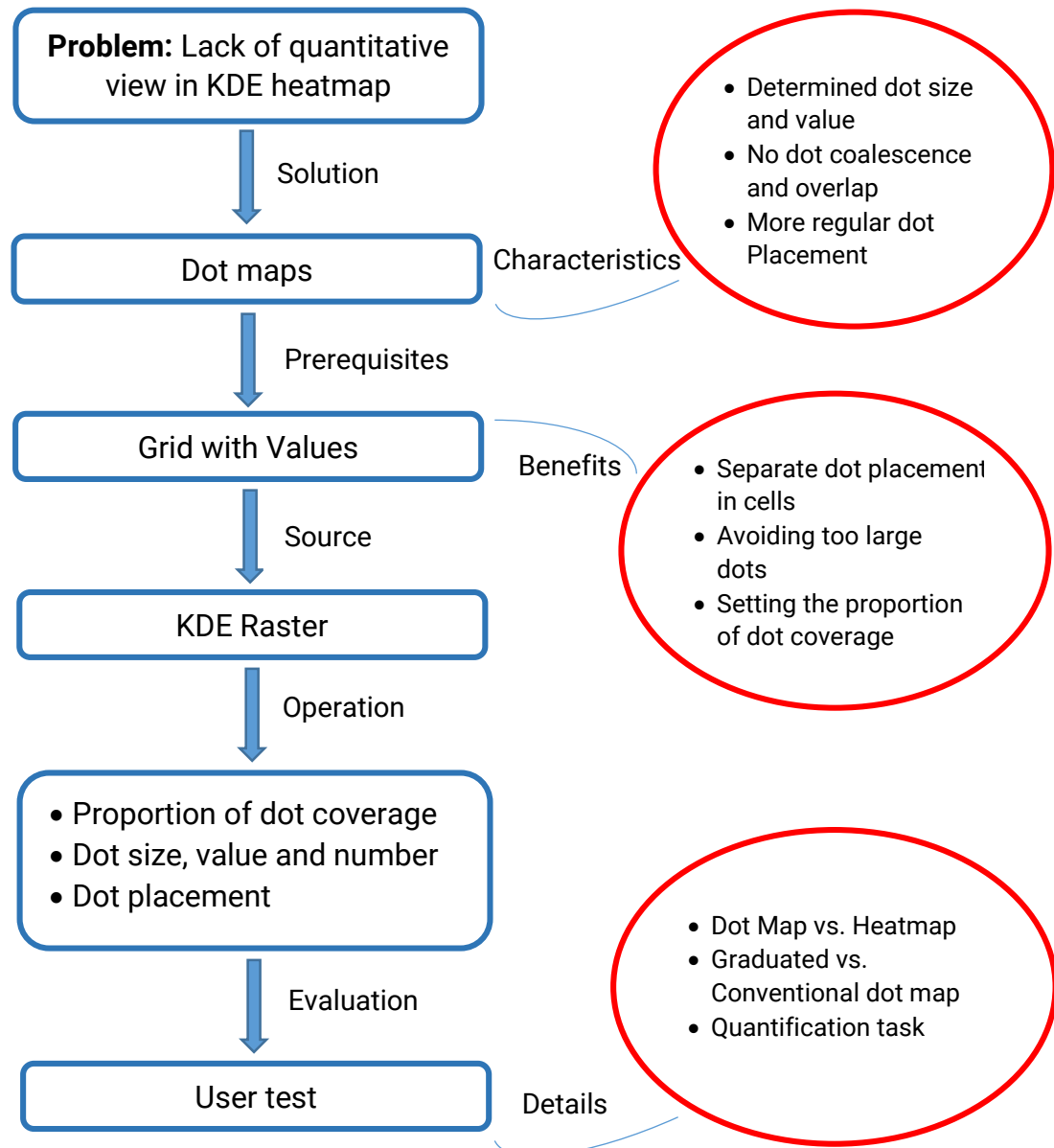


Figure 3.1 Flowchart of the proposed approach

3.1 Conventional dot mapping

I developed a conventional dot mapping algorithm using Python 3.7. The process of making a conventional dot map are explained in the following four main steps.

3.1.1 The first step

This part started with importing a dataset to the programming environment as a geodataframe. In this thesis, only point datasets were used as an input to the algorithm. The coordinate system of the input data had to be changed to a metric coordinate system, because the size and unit of dots were set based on meters. The KDE function could be run on the input data. Working on the categories of KDE was not the focus of this thesis and varied types of KDE did not affect considerably the quality of the outputs. Gaussian KDE was chosen in this research. There were two prerequisites which should have been set before running the function:

Firstly, the bounding box of data should have been recognized. The minimum and maximum of latitudes and longitudes defined the bounding box of the data. The reason for this definition was to set a rectangular frame for the KDE function, so that it visualizes the results in this frame and covers the whole dataset. Also, adding some values to four sides of the bounding box improved the quality of visualization, but did not change the quantity of results. The reason was that isolines were not constructed completely in the sides of studied region and adding some values helped to make efficient isolines.

Secondly, the resolution of the output raster should have been determined. Resolution has a secondary role in a conventional dot map. A number for both width and length of the output raster was required. The resolution of the raster should have been determined by the cartographer based on the data and desired dot sizes. It was an iterative process to determine the resolution and recommending a precise solution in advance proved difficult. The larger the studied area is, the higher the resolution should be. Moreover, the resolution should have been adapted based on the dot size. When the dot size was increased, the resolution should have been decreased and vice versa. This adaption resulted in harmony between the dot size and the resolution and avoided any wrong estimation. For example, a resolution of between $20 * 20$ to $50 * 50$ is suitable for an area with the same size as Munich city in Germany. It is better to give equal values to the length and width of the raster to obtain a square to be able to manage the data well. A square raster makes the process easier for the cartographer. As was explained in part 2.3, the result of KDE could be a heatmap which was a rectangular raster with cells and isolines. Each region has a specific color within the isolines. The number of cells depends on the resolution, which was defined before running the algorithm. Each pixel of a raster had a value and those values had a key role in the algorithm. This value had a direct relation to the density of the studied phenomenon. The denser a region is, the higher a pixel value in that region would be. Figure 3.2 is a sample heatmap made by the user to show the density of a sample phenomenon. The pixels of the black area have the highest pixel value because that region is the most dense area. The pixels of the light red regions have the lowest value as those are the least dense regions. The value of each pixel is called the z-value of a cell of the grid in this thesis.

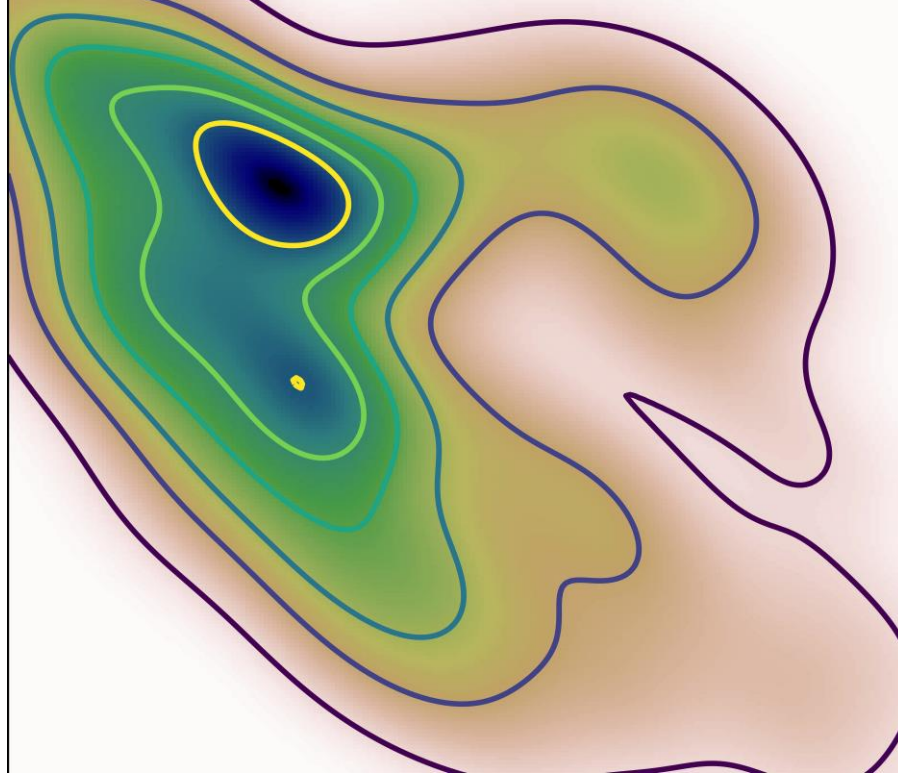


Figure 3.2 A sample heatmap made by the author

3.1.2 The second step

The second step of the algorithm was to convert the raster to a polygon vector. The vector was analyzed as a geodataframe with the same assigned coordinate system for the input data. The number of rows in the geodataframe equals the number of cells and depends on the resolution of the raster. Z-values of cells were obtained from corresponding pixels and were listed as a new column in the geodataframe. The rows with a z-value of zero had to be removed for two reasons: 1) cells with a value of zero showed a zero density and there was no need to place any dots within them; 2) a number of the zero value cells were out of the study area because the raster was a circumscribed rectangle around the dataset and might have covered some no data regions, especially in the corners. Figure 3.3 shows the defined grid around a sample data, containing some cells out of the studied area.

3.1.3 The third step

Now we have a grid with a number of cells which have the same size. Based on the z-values, a number of dots should have been inserted in each cell. In this step, Lavin's approach (Slocum et al., 2009) was used to determine the number of dots in each cell. The approach had three formulas from equation 3.1 to equation 3.3. The first task was normalization of z-values of each cell (Eq. 3.1):

$$z_i = \frac{v_i - v_s}{v_L - v_s} \quad \text{Eq. 3.1}$$

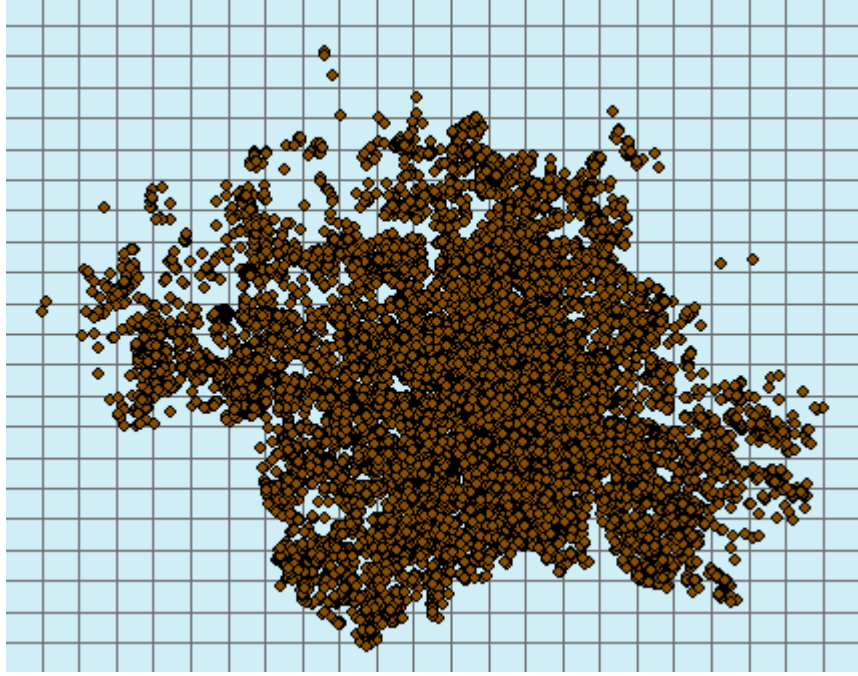


Figure 3.3 The defined grid around a sample data might cover some no data area

where v_i is the z-value of a cell, v_s is the smallest and v_L is the largest z-value among cells and z_i is the normalized z-value of a cell which is a number between 0 and 1. Then, definition of dot coverage in each cell was required. Dot coverage is the area in a cell covered by dots. Equation 3.2 calculates data coverage as follows:

$$P_i = z_i (P_{max} - P_{min}) + P_{min} \quad \text{Eq. 3.2}$$

where P_i is the proportion of data coverage for each cell. P_{max} and P_{min} are the minimum and maximum desired proportions of dot coverage, respectively. The mentioned source of Lavin's approach recommended 0.50 and 0.08 for these elements, respectively. The algorithm was run with different P_{max} . But because the dot placement was a random process within each cell, the algorithm could not use a P_{max} more than approximately 0.50 because the running time will be infinite in this case. It means that in this dot placement method, at least 50 percent of the area of a cell remains empty and with no dots. Therefore, (Slocum et al., 2009) recommended 0.50 for P_{max} since a higher number is impossible and a lower number makes a cell too empty. However, P_{min} could change with respect to the data. It is up to the cartographer to decrease or increase it. A higher P_{min} decreases the difference of number of dots in high z-value cells and lower ones. The number of dots in each cell was calculated by equation 3.3:

$$N_d = \frac{P_i * A_c}{A_d} \quad \text{Eq. 3.3}$$

where N_d is the number of dots in each cell, A_c is the area of one cell and A_d is the area of one dot. There was an inverse relationship between dot size and the number of dots in a cell as was concluded in part 2.1. Dot size was determined based on an iterative process because it is highly related to the data. It was not possible to set a dot size suitable for every data. It was possible to decrease the dot size to increase the number of dots which leads to a more crowded dot map; one dot shows fewer points of interest from the original data in this case. Also, it was possible to increase the dot size to decrease the number of dots which led to a less crowded dot map; one dot showed more number of points of interest from the original data in this case.

3.1.4 The fourth step

The last step of the algorithm was the placement of dots in each cell. The placement of dots was a pseudo-random placement which means not completely random and not completely regular. It places the determined number of dots in each cell randomly but avoids any dot overlap and coalescence within and between cells. Based on the discussion in the second chapter, dots should have a value in a dot map. The dot value was calculated simply by dividing the number of points in the input dataset by the whole number of dots. Figure 3.4 shows the pseudo code of the proposed algorithm. Also, Figure 3.5 presents the workflow of the algorithm.

```

1: input = point data
2: kde = running Gaussian KDE function on input
3: grid = vectorize (kde)
4: if z-value of a cell equals zero:
5:     remove the cell
6: for all of cells:
7:      $z_i = (z - z_{min}) / (z_{min} - z_{max})$  //  $z_i$  refers to normalized z-value
8: for all of cells:
9:      $p_i = z_i * (p_{max} - p_{min})$  //  $p_i$  refers to proportion of data coverage
10: for all of cells:
11:      $dn = (p_i * ca) / da$  //  $dn$  refers to dot number,  $ga$  refers to cell area,  $da$  refers to dot size
12: for all of cells:
13:     while the number of dots in each cell is less than its determined number:
14:         find a random position within the cell with a distance of radius of dots to the border
15:         if the dot is the first dot in a cell:
16:             insert it in the found position
17:         otherwise if the dot doesn't overlap any of existing dots:
18:             insert it in the found position

```

Figure 3.4 Pseudo code of conventional dot mapping algorithm

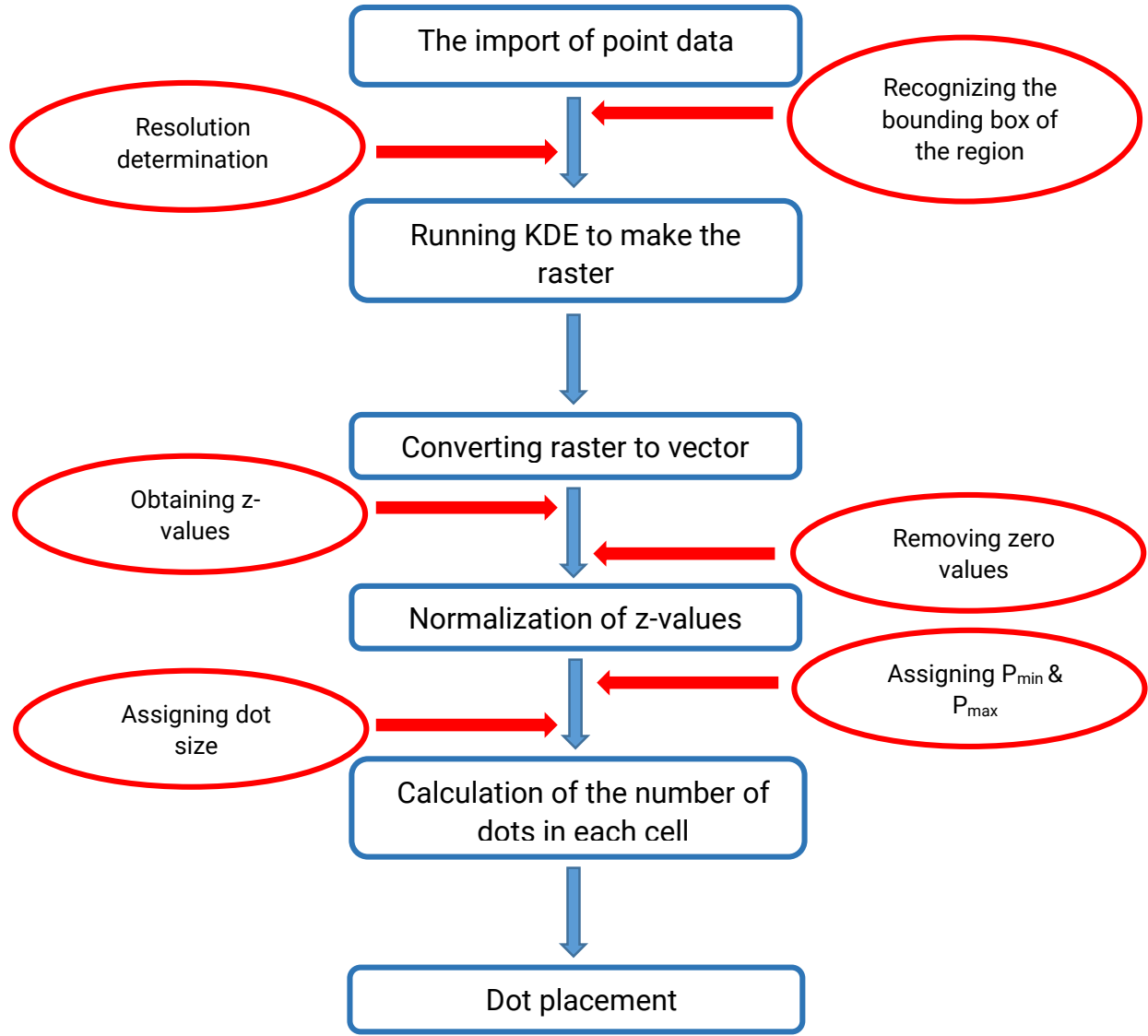


Figure 3.5 Workflow of the conventional dot mapping algorithm

3.2 Graduated dot mapping

We also developed a graduated dot mapping algorithm using Python 3.7. The process of making a graduated dot map from the beginning until the end of the second step was exactly the same as the conventional dot mapping algorithm. Steps 3 to 5 will be explained below.

3.2.1 The third step

The third step was normalization of each z-value by dividing each one by the maximum z-value. Equation 3.4 shows this formula:

$$z_i = \frac{v_i}{v_{max}} \quad \text{Eq. (3.4)}$$

where v_i is the z-value of a cell, v_L is the largest z-value among cells and z_i is the normalized z-value of a cell which is a number between 0 and 1. There was a huge number of unique z-values among cells and it was not possible to use this huge number of dot sizes in a map. Minimal values were required for a graduated dot map because a limited number of dot sizes were needed for such maps. Therefore, there was a need to classify the normalized z-values to the desired number of sizes of dot that we needed. Since the z-values were very heterogeneous, a classification based on some round values (e.g., 0.2, 0.4, 0.6, ...) and equal intervals was not a guarantee for a good classification. *Jenks natural breaks optimization* is a method to classify one dimensional data based on their density. It is a data clustering method designed to determine the best arrangement of values into different classes. This method seeks to reduce the variance within classes and maximize the variance between classes (McMaster, 1997). After classification of the normalized z-values, each cell got a new z-value which was the maximum number of each cluster of data. Hence, there were a limited number of unique z-values among cells.

3.2.2 The fourth step

Unlike the conventional dot map, only one dot was inserted in each cell, while the size of dots was varied in graduated dot map. The size of the cells was based on the resolution of the heatmap determined by the cartographer in an iterative process. Setting the resolution was highly related to the data and should have been adapted based on that. Larger cell sizes (lower resolution) should have been used when the studied area was heterogeneous. The reason was that we could show the big density difference by very large and very small dots. On the other hand, if the region was more homogeneous, we could use smaller cell sizes (higher resolution) because we did not need a big difference between the largest and smallest dot. Resolution has a primary role in a graduated dot map because it is the main criterion to set the size of dots. There is no dot radius determination in advance in graduated dot mapping. The size of the largest dot was the base of the calculation of other dot sizes. The largest dot(s) was the inscribed circle in a cell which was placed in the cell(s) with the highest normalized z-value (equal to 1). The radius of the largest dot equaled half of the side of a cell. Figure 3.6 shows the largest dot in a cell in a sample picture. The size of the rest of dots was calculated by equation 3.5 (Slocum et al., 2009):

$$r_i = \sqrt{z_i} * r_L \quad \text{Eq. (3.5)}$$

where r_i is the radius of dot in cell i and r_L is the radius of the largest dot in cell(s) with normalized z-values equal to 1. As was stated in the second chapter, equation 3.5 was used for proportional symbol maps in the mentioned source but was used for graduated dot maps in this research. Thus, the number of dot sizes equaled the number of classes in classification of normalized z-values by Jenks natural breaks optimization method which is determined by the operator. Finally, based on the determined dot size, only one

dot was placed in the center of each cell. This graduated dot mapping technique was a completely regular and non-random method, in which there was neither overlap nor coalescence between dots. Since the largest dot(s) was an inscribed circle in their cells and every dot was placed exactly in the center of a cell, overlap and coalescence between dots was impossible.

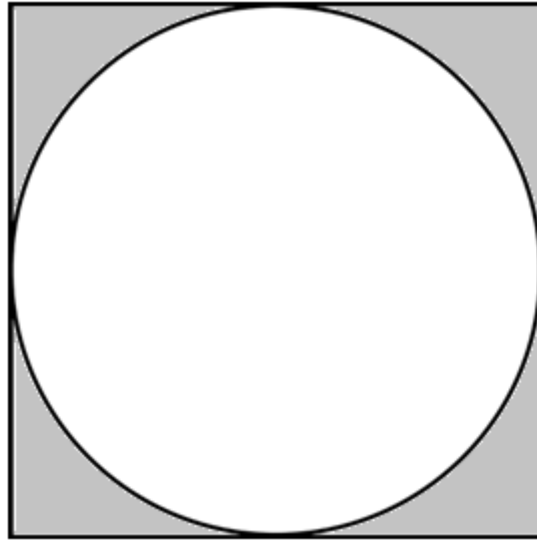


Figure 3.6 The largest dot of the graduated dot map in a cell

3.2.3 The fifth step

The last step of this part was to find dot value for each class of dots. The key to finding values was isolines of the heatmap product of the KDE function. After running the KDE function, KDE defines some levels and inserts isolines based on those levels on the heatmap automatically. However, it is possible and necessary to define new levels for isolines because a harmony is required between the position of isolines and the transitions of varied dot sizes. Since the levels were based on z-values of cells, Jenks natural breaks optimization was again a solution for definition of levels for isolines. In this part of the project, the number of defined levels for isolines should have been equal to the number of clusters of z-values in the previous section. After obtaining the raster with defined z-values, it should have been converted to a vector and a geodataframe. The number of polygons in the vector equaled to the number of classes/levels. Then, there was a need to count the number of points from the input data which lay within each polygon or isoline. The division of this number by the number of already inserted dots with a specific and corresponding size gave the dot value for each size of dots in the produced dot map. Figure 3.7 shows the pseudo code of the proposed algorithm. Figure 3.8 presents the workflow of the algorithm.

```

1: input = point data
2: kde = running Gaussian KDE function on input
3: grid = vectorize (kde)
4: if z-value of a cell equals zero:
5:     remove the cell
6: for all of cells:
7:      $z_i = (z)/(z_{\max})$  //  $z_i$  refers to normalized z-value
8: for all of cells:
9:     the whole  $z_i$  are classified into some classes by Jenks natural break optimization
10: for all of cells:
11:     Every  $z_i$  is changed to the head of the cluster in which lies
12: for all of cells:
13:     radius equals to square root of  $z_i$  multiplied by radius of the largest dot
14: for all of cells:
15:     insert just one dot with the assigned radius to the center of the cell
16: assign the levels of isolines by Jenks natural break optimization
17: polygonise the heatmap based on the isolines
18: count the number of points of original data within each isolines
19: division of result of line (18) by the number of dots in corresponding cluster gives the value of dots

```

Figure 3.7 Pseudo code of graduated dot mapping algorithm

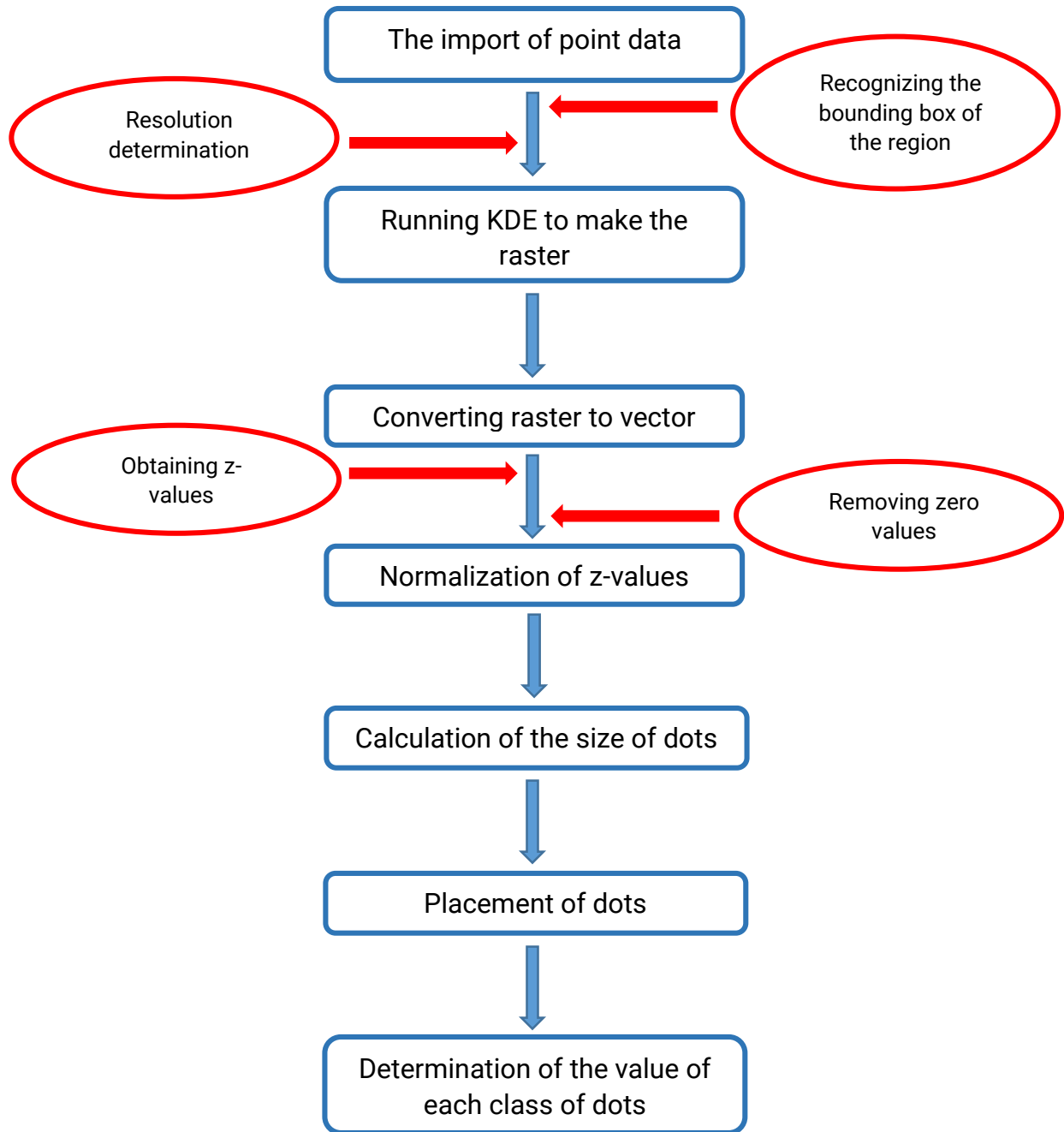


Figure 3.8 Workflow of the graduated dot mapping algorithm

4 User Test and Results

This chapter describes the results of execution of the two algorithms on two datasets. Moreover, a user test was conducted to test, evaluate and compare the results of the execution and respond to the research questions. The first dataset was the Points of Interest (POIs) data of Munich city in Germany, which contains 29,232 POIs. This dataset contains the location of any type of point of interest, e.g., restaurants, super markets, bicycle shops, book stores, etc. in the city. This data was obtained from OpenStreetMap (OSM) open source project. The second dataset was tweets about the traditional Bavarian festival, called Oktoberfest, obtained through Twitter API. This dataset contains 2,745 tweets with Oktoberfest relevant keywords. These tweets were extracted based on hashtags that were assumed to be related to Oktoberfest. Only tweets with a location in Munich were downloaded.

4.1 Results of running the algorithms

Three categories of maps were provided for each data. Firstly, a heatmap was created to show the density of the phenomena, because it is the usual method of showing the density. Secondly, some conventional dot maps with varied dot sizes were developed for each data. Thirdly, some graduated dot maps with varied dot size classes were built for the data. The results of quantization of each conventional and graduated dot map were almost the same. However, there was a difference between the dot sizes and the number of dots. Since the coordinate system of the input data should have been metric, "*Gauss-Kruger zone 4*" with *EPSG: 31468* was selected for both of the used datasets. It is a transverse Mercator projection, which preserves angles. This coordinate system is the former West Germany onshore between 10°30'E and 13°30'E - states of Bayern, Berlin, Niedersachsen, Schleswig-Holstein. The projected coordinate system is "*DHDN_3_Degree_Gauss_Zone_4*" and the geographic coordinate system is "*GCS_Deutsches_Hauptdreiecksnetz*". It is a suitable metric coordinate system for Germany. A different conformal coordinate system should be chosen based on the location of data in the world. Some other parameters, including *resolution*, *dot size* and P_{min} , should have been fixed based on the data (Table 4.1).

4.1.1 Maps of the POI dataset

One heatmap (Figure 4.1), five conventional dot maps (Figure 4.3 to Figure 4.7) and four graduated dot maps (Figure 4.8 to Figure 4.11) were made for the POI dataset of Munich. The goal was to make maps with the defined standards in the beginning of the third chapter. Maps with fewer dot sizes could lead to complicating the estimation process and maps with greater dot sizes could lead to impossible estimation for small regions on the map. These sizes were distinguished as the best size of dots.

Data	Map type	Figure	Resolution	Dot size (m)	P _{min}
POI	Conventional	Figure 4.3	42*42	45	0.01
POI	Conventional	Figure 4.4	35*35	56	0.01
POI	Conventional	Figure 4.5	29*29	65	0.01
POI	Conventional	Figure 4.6	25*25	79	0.01
POI	Conventional	Figure 4.7	23*23	85	0.01
POI	Graduated	Figure 4.8	30*30	---	---
POI	Graduated	Figure 4.9	40*40	---	---
POI	Graduated	Figure 4.10	50*50	---	---
POI	Graduated	Figure 4.11	60*60	---	---
Tweets	Conventional	Figure 4.13	52*52	60	0.02
Tweets	Conventional	Figure 4.14	39*39	100	0.02
Tweets	Conventional	Figure 4.15	30*30	120	0.02
Tweets	Graduated	Figure 4.16	40*40	---	---
Tweets	Graduated	Figure 4.17	50*50	---	---
Tweets	Graduated	Figure 4.18	57*57	---	---

Table 4.1 Detailed parameters of created dot maps

The heatmap of POI (Figure 4.1) was presented only to make it possible to compare the created dot maps with the heatmap and judge the potential superiority of each map. In this heatmap, the density of POIs was presented by some isolines and a color scheme. Each color within an isoline visualizes the density of POIs per square kilometer. For example, the dark blue part within the yellow isoline contains 1,106 points per square kilometer.

Figure 4.3 to Figure 4.7 display five conventional dot maps with different dot sizes. I tried to set the parameters of the algorithm so that we could set round numbers (10, 15, 20, 25, 30) for the value of dots. Round numbers help the user to quantify the number of POIs more easily. In each map, one dot represents a number of points based on its value. For instance, each dot represents 20 points of interests in Figure 4.5. As was explained in section 3.1.3, the number and the value of dots have an inverse relationship. The number of dots decrease and their value increase in these five maps, respectively. However, the result of estimation should be the same in all the five maps. The estimation process seems harder in the map with smaller dots (Figure 4.3) because the number of dots is greater and the estimation process in a specific area is more time consuming. However, if the user is interested in estimating the quantity in more detail for a certain area, the more crowded maps with the greater number of dots and lower dot value is more useful. When the map reader is interested in estimating the quantity in a very small, specific region on the map, she needs a dot map with smaller values. On the other hand, estimation is easier in the map with larger dots (Figure 4.7) because the number of dots is fewer. However, a less detailed estimation is possible. For example, Figure 4.2 shows a sample area with small (a) and large (b) dot sizes. The user is able to estimate the quantity faster in (b) because she has to count two dots. However, if the user is interested

in knowing the quantity in the upper/lower side of the area, she has to use the first map (a).

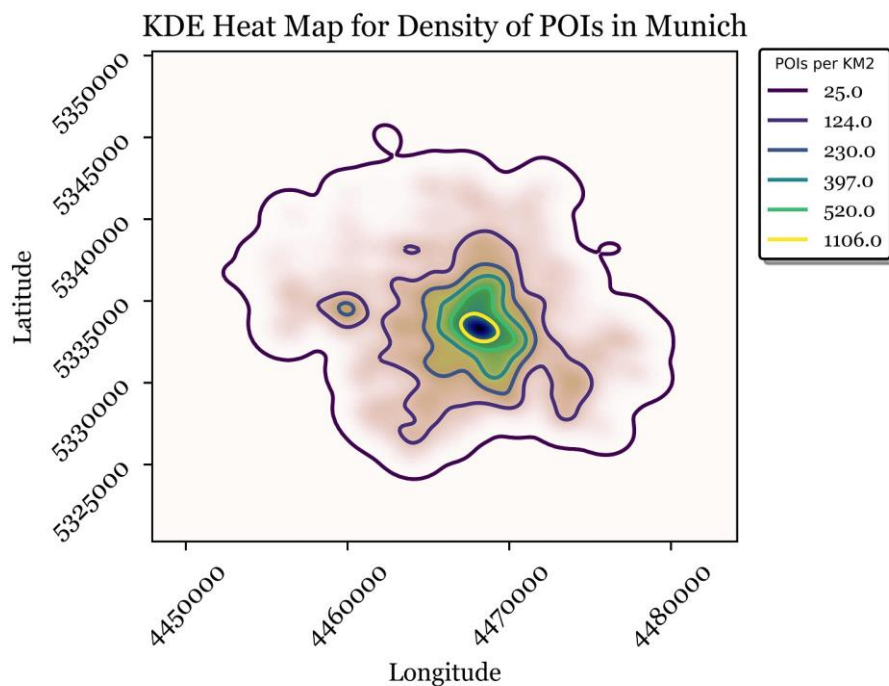


Figure 4.1 KDE heatmap for Density of POI

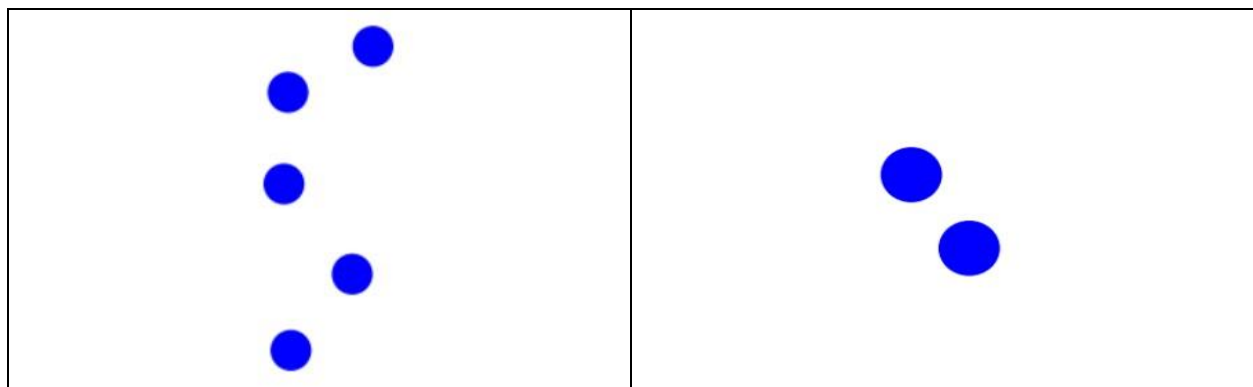


Figure 4.2 A comparison of dot maps for a more detailed or easier estimation

As was explained in section 3.1.1, the resolution is an effective factor in a conventional dot map. Although it does not change directly the dot size and value, it should be adapted based on the determined dot size. Based on the Table 4.1, the resolution and dot size should have an inverse relationship to make a harmony between these two factors. The cartographer needs to adapt these two parameters to avoid any bias in the quantification. If the resolution is not adapted based on the dot size, the density estimation doesn't give any true results. In these maps, the density of the dots is more in the center of the area. The center of the area is the city center of Munich, in which there are more shops, events, historical places, etc.

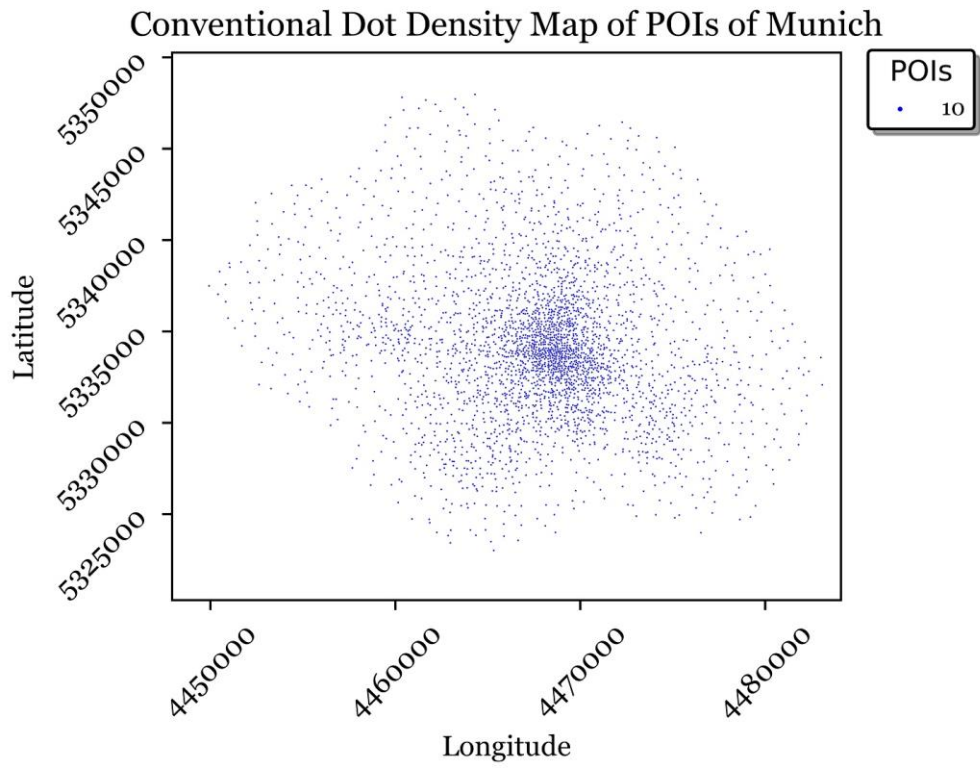


Figure 4.3 Conventional dot density map of POIs with dot value of 10

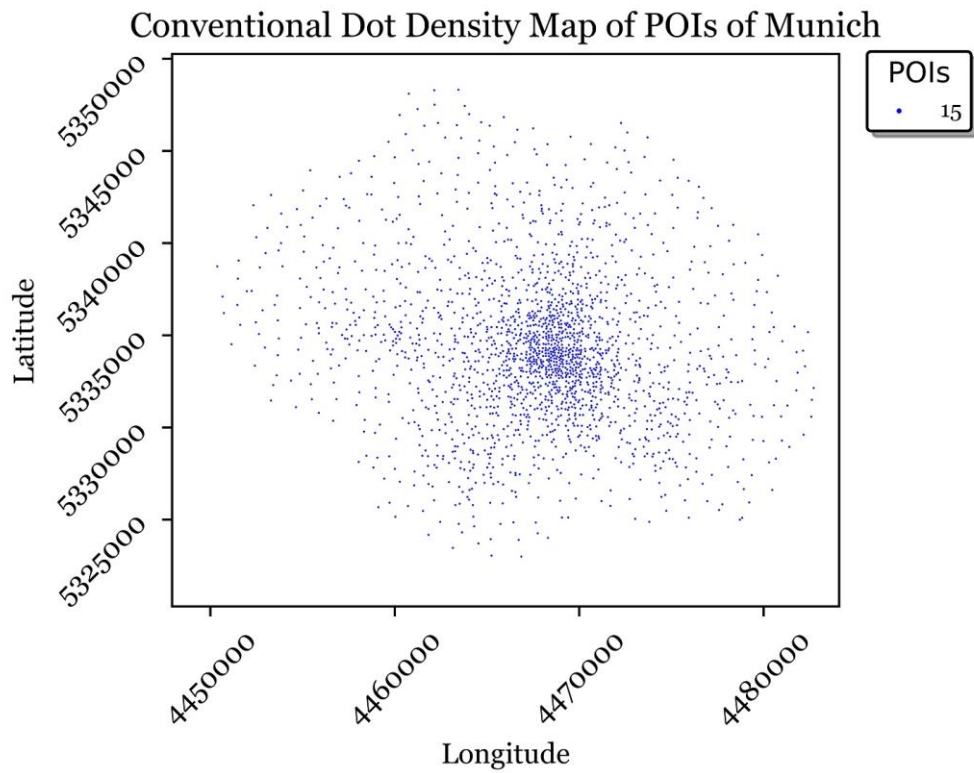


Figure 4.4 Conventional dot density map of POIs with dot value of 15

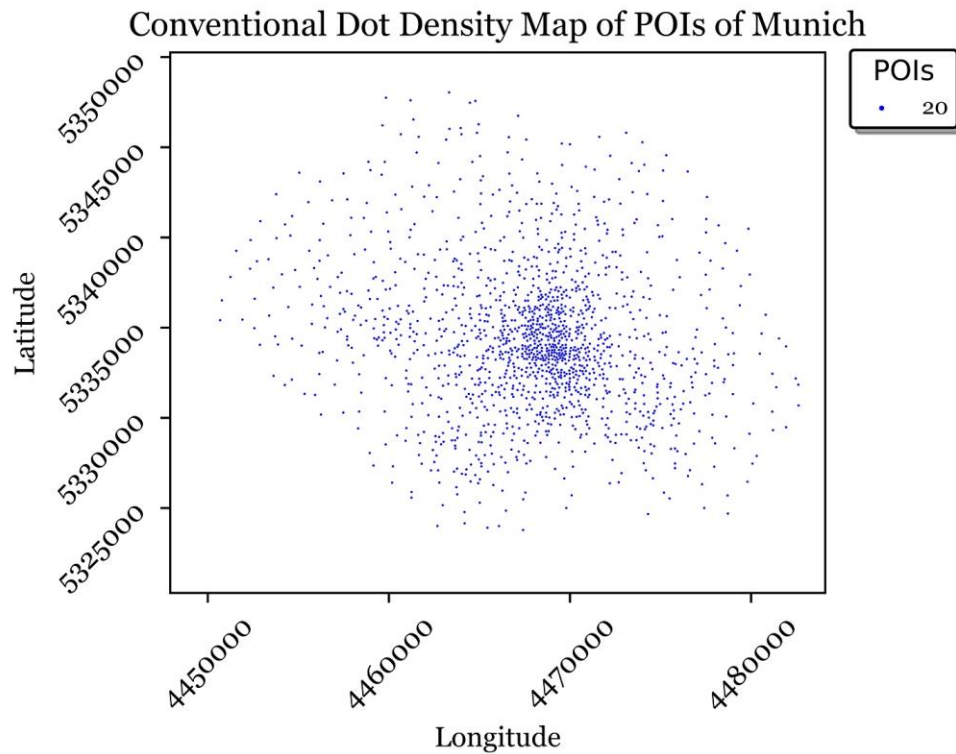


Figure 4.5 Conventional dot density map of POIs with dot value of 20

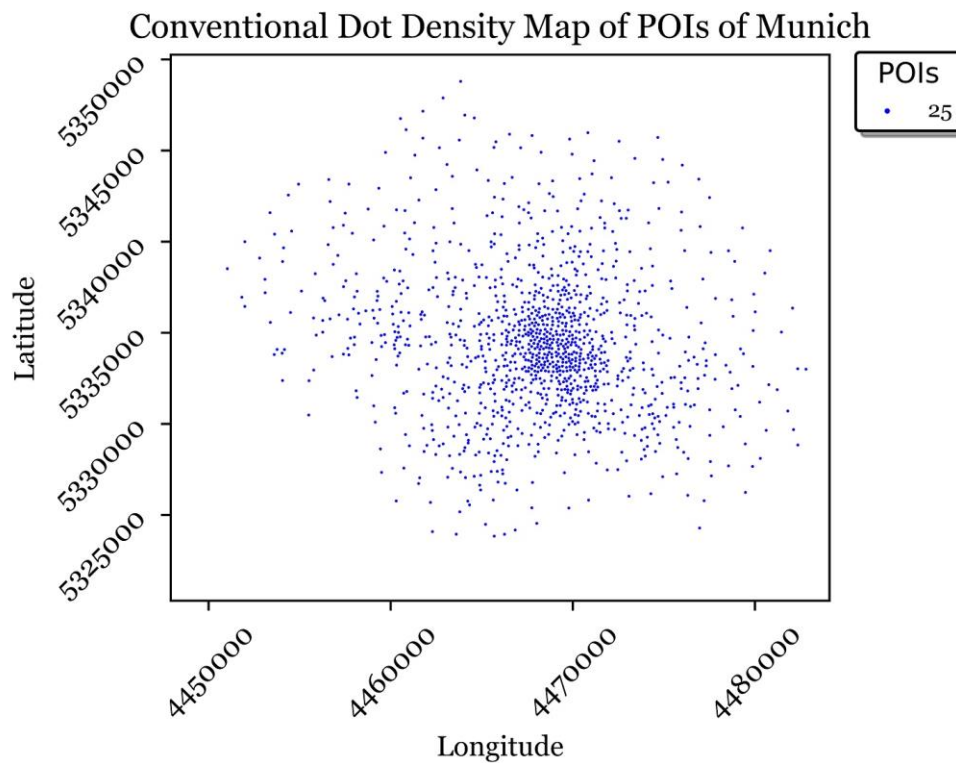


Figure 4.6 Conventional dot density map of POIs with dot value of 25

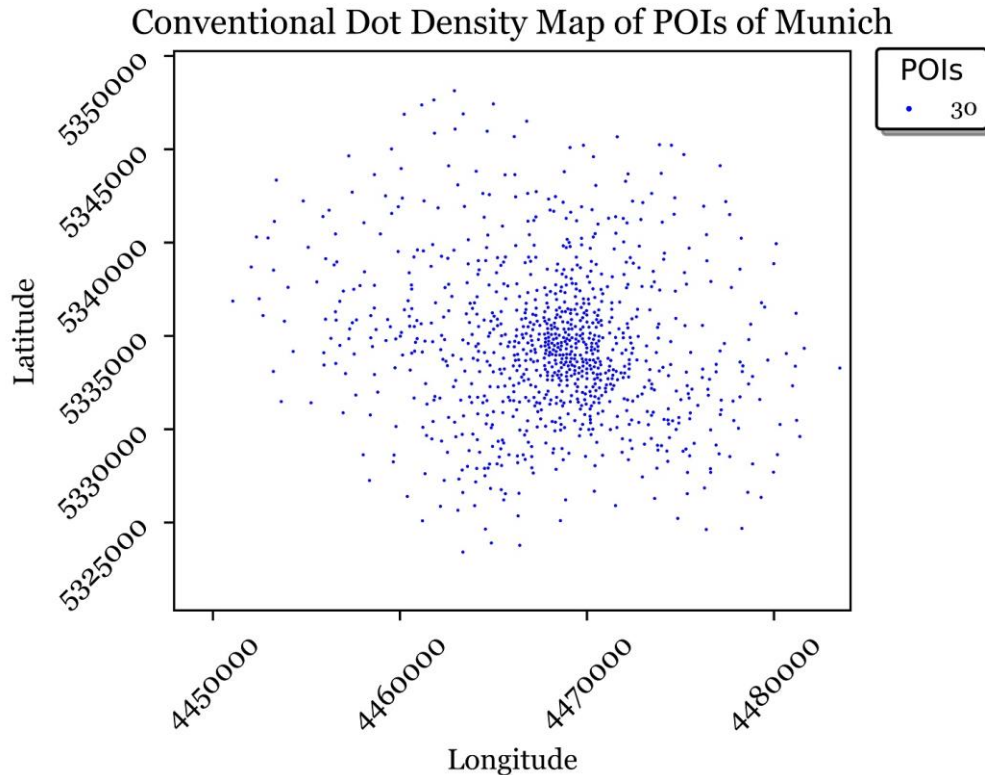


Figure 4.7 Conventional dot density map of POIs with dot value of 30

Figure 4.8 to Figure 4.11 present four graduated dot maps for the density of POIs. As was indicated in the section 3.2.2, the placement of dots was completely non-random and regular. As was proved in the related studies in the second chapter, e.g. (Hey, 2012), (Kimerling, 2009) and (Hey & Bill, 2014), regularity of dot placement helps users to easily estimate the quantity. The more the dot placement is regular and non-random, the easier is the estimation. These four maps are presented in four proportional scales. The same as conventional dot maps, the number of dots and dot values have an inverse relationship in graduated dot maps. The number of dots increase and the value and size of the classes of dots decrease from Figure 4.8 to Figure 4.11, respectively. For example, the largest dot shows 745 points and smallest dot presents 3 points in Figure 4.8, while the largest dot shows 227 and the smallest one presents 1 point in Figure 4.11. It is possible to use the same analysis as conventional dot maps for the graduated dot maps. The estimation process is easier in a map with larger dots (lower resolution) and a more detailed estimation is better acquired in a map with smaller dots (higher resolution). The determined resolution in the algorithm is the main factor to change the dot size and value in a graduated dot map. The resolution has an inverse relationship with the dot size and dot value. A lower resolution leads to larger dots with higher values. However, the lower the resolution is, the harder it is to get a detailed estimation of a specific region. As was explained in section 3.2.2, resolution is the main criterion to change the dot sizes in a graduated dot map. Changing the resolution leads to a change in the dot sizes.

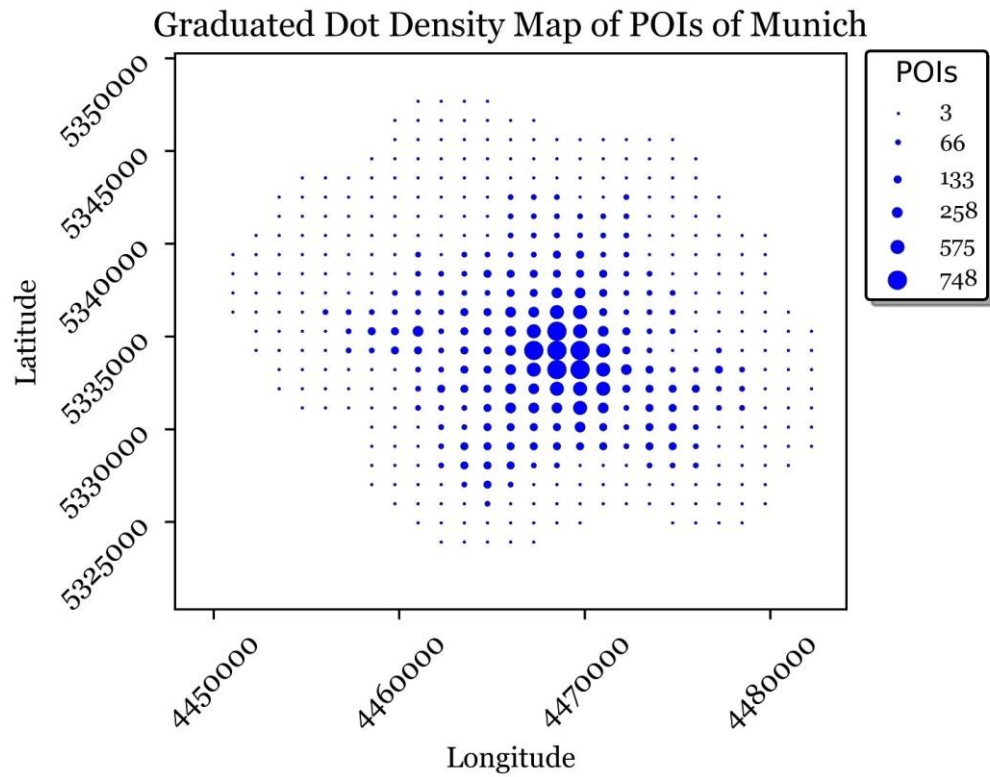


Figure 4.8 Graduated dot density map of POIs

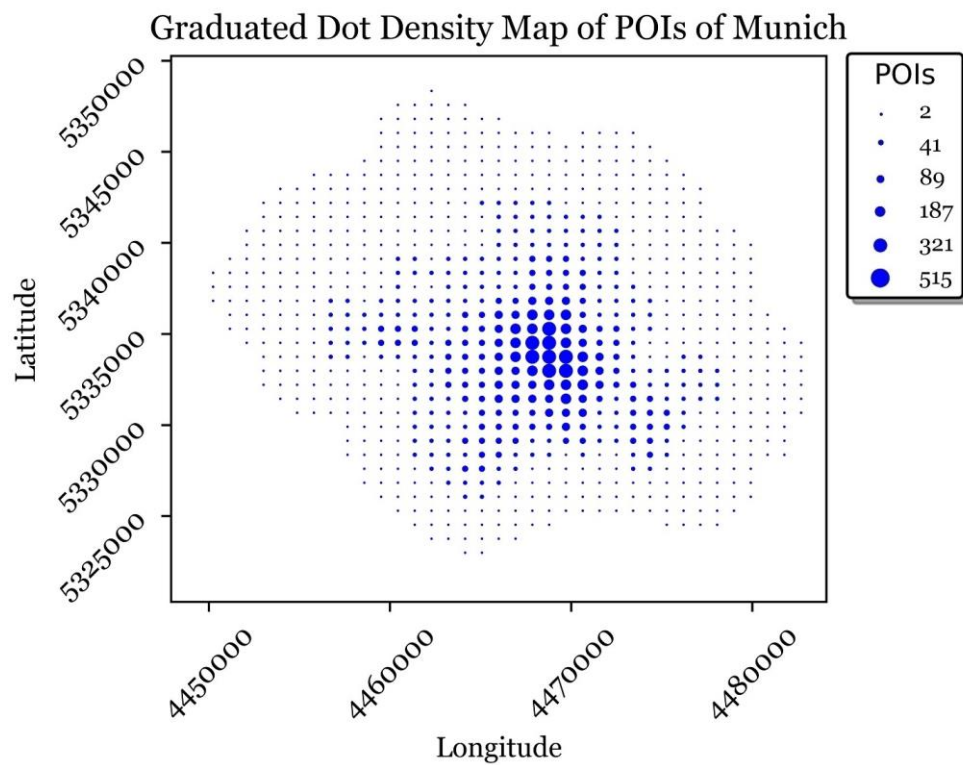


Figure 4.9 Graduated dot density map of POIs

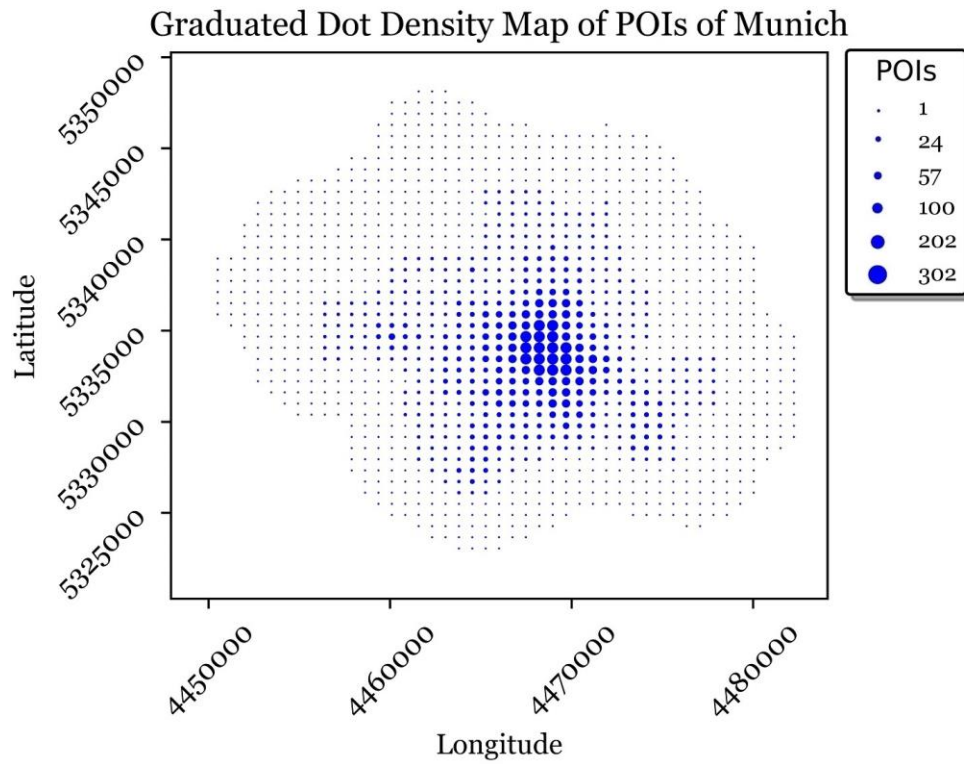


Figure 4.10 Graduated dot density map of POIs

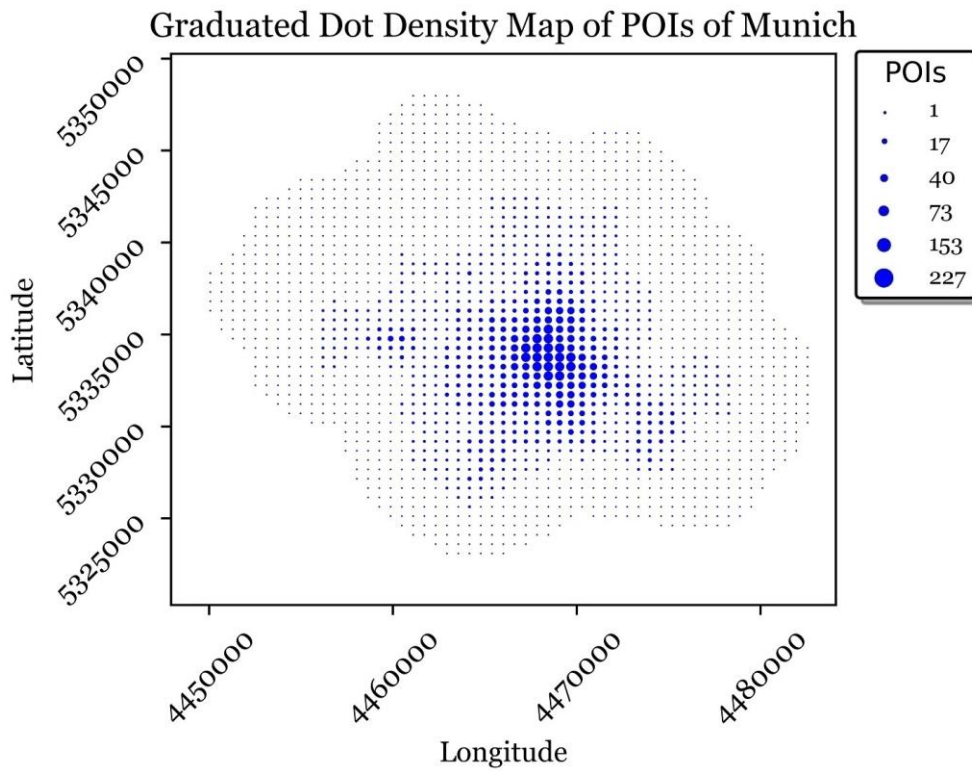


Figure 4.11 Graduated dot density map of POIs

4.1.2 Maps of the Twitter dataset

Similarly, one heatmap (Figure 4.12), three conventional dot maps (Figure 4.13 to Figure 4.15) and three graduated dot maps (Figure 4.16 to Figure 4.18) were made for the twitter dataset of Oktoberfest. The reasons for making maps with different dot values and sizes was the same as the section 4.1.1 for the POI dataset.

Figure 4.12 is the heatmap to present the density of Oktoberfest tweets. Like the section 4.1.1, the purpose of creating the heatmap as the most common visualization technique was to make a benchmark for the user test. Isolines and color scheme were utilized to show the density of tweets per kilometer square.

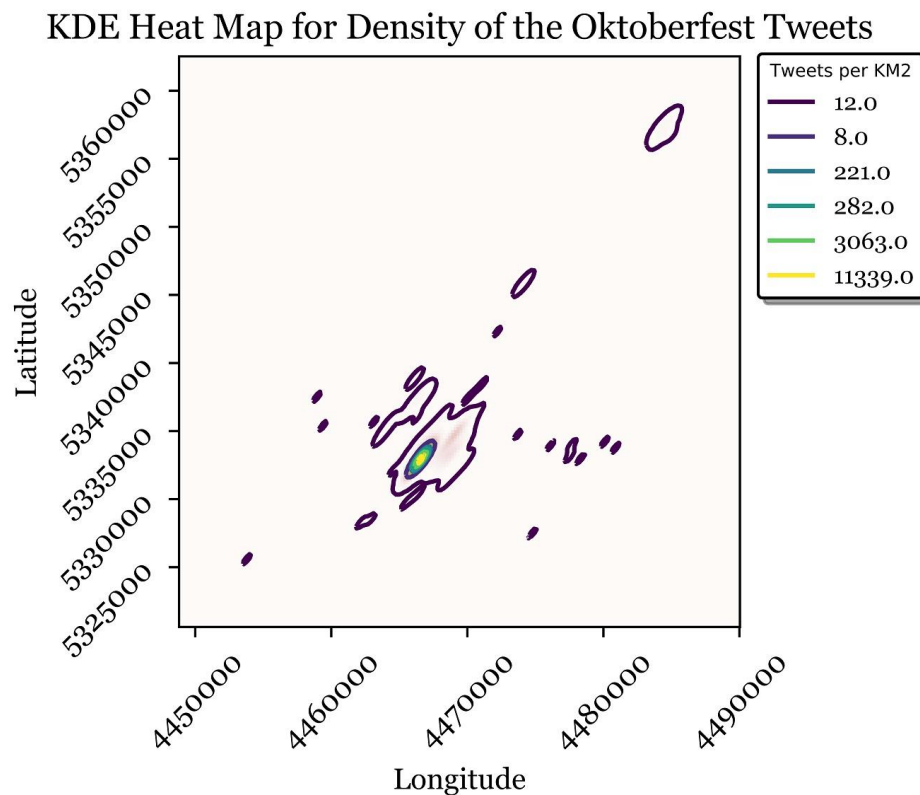


Figure 4.12 KDE heatmap for density of tweets

Figure 4.13 to Figure 4.15 are three conventional dot maps to show the density of tweets. The same process and rules of conventional dot maps are applicable for this dataset and any other dataset, as well. The number and value of dots have an inverse relation. The rounded numbers (5, 10 and 15) were used for dot values. In this dataset, there is a big contrast in density among different parts of the area. The area is very heterogeneous because the difference of the density between the most dense region and the least dense region is very high. The regions with a medium density are very rare. The highly dense areas have a coincidence with the area in which Oktoberfest is held every year. With growing distance from the Oktoberfest spot, the density of tweets considerably

decreases. It indicates that most of the tweets about the Oktoberfest have been posted in the festival spot. The big contrast of density in this dataset made another issue. The pixelated effect in the most dense areas are observable in figures 4.13 and 4.14. The high difference in the density z-values of central pixels and their surrounding pixels made this problem. Figure 4.15 is a better result for this dataset and compensate this issue. Like the POI dataset, the resolution and dot size have the inverse relation. When the dot sizes changes, the resolution should be inversely adapted.

Figure 4.16 to Figure 4.18 display three graduated dot maps to show the density of Oktoberfest tweets. The same process and rules of graduated dot maps of the POI dataset are applicable for this dataset and any other dataset, as well. Six classes of dot values were used to show the density of tweets. For example, the largest dot(s) presents 2,011 and the smallest dot(s) show 1 tweet in Figure 4.16. Like the POI dataset, the resolution is the main factor to change the dot sizes. A change in resolution directly changes the size of dots.

Conventional Dot Density Map of the Oktoberfest Tweets

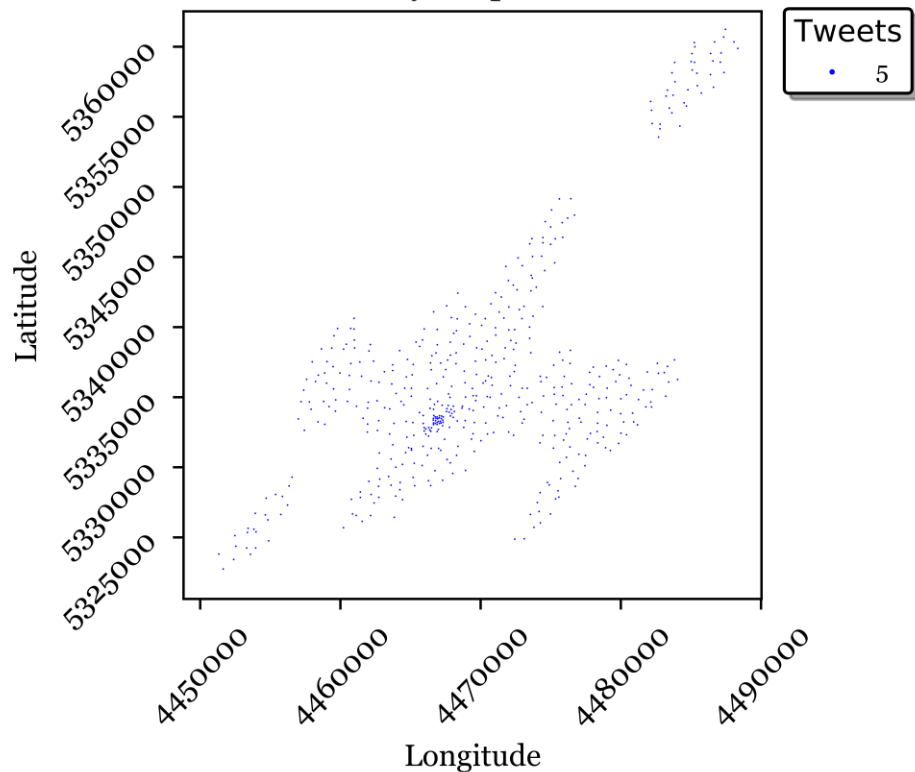


Figure 4.13 Conventional dot density map of Tweets with dot value of 5

Conventional Dot Density Map of the Oktoberfest Tweets

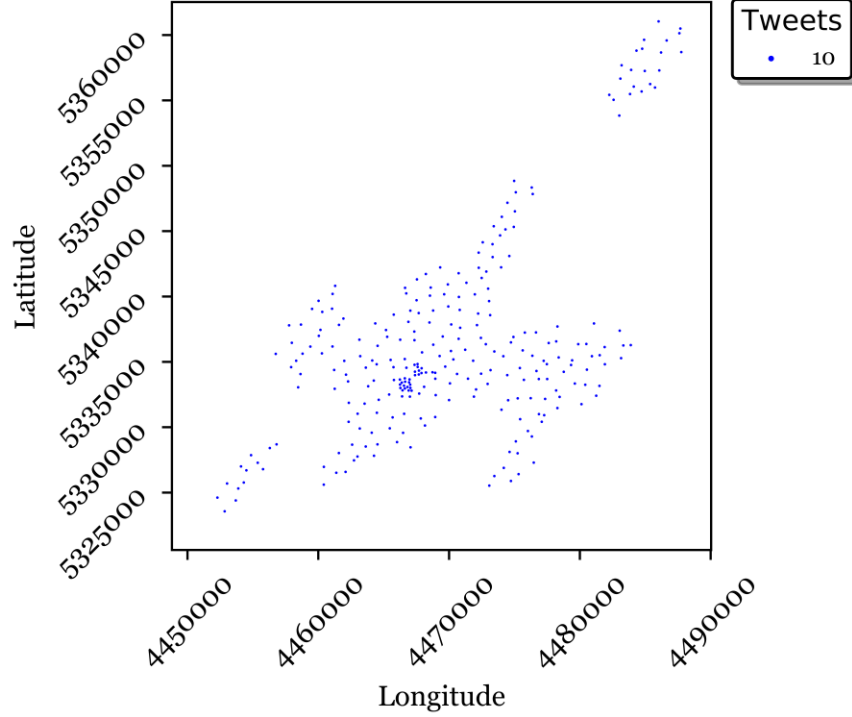


Figure 4.14 Conventional dot density map of Tweets with dot value of 10

Conventional Dot Density Map of the Oktoberfest Tweets

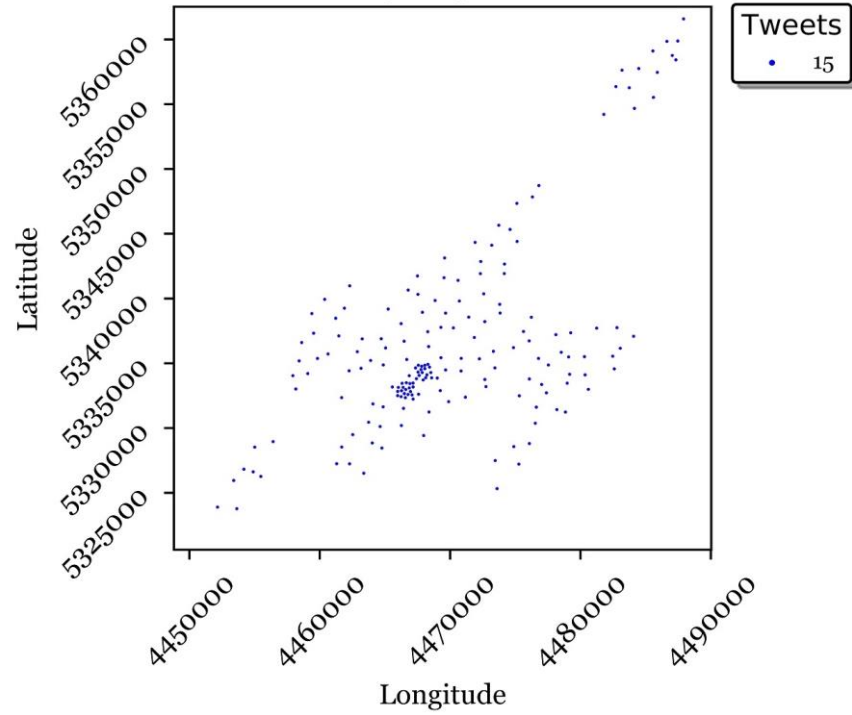


Figure 4.15 Conventional dot density map of Tweets with dot value of 15

Graduated Dot Density Map of the Oktoberfest Tweets

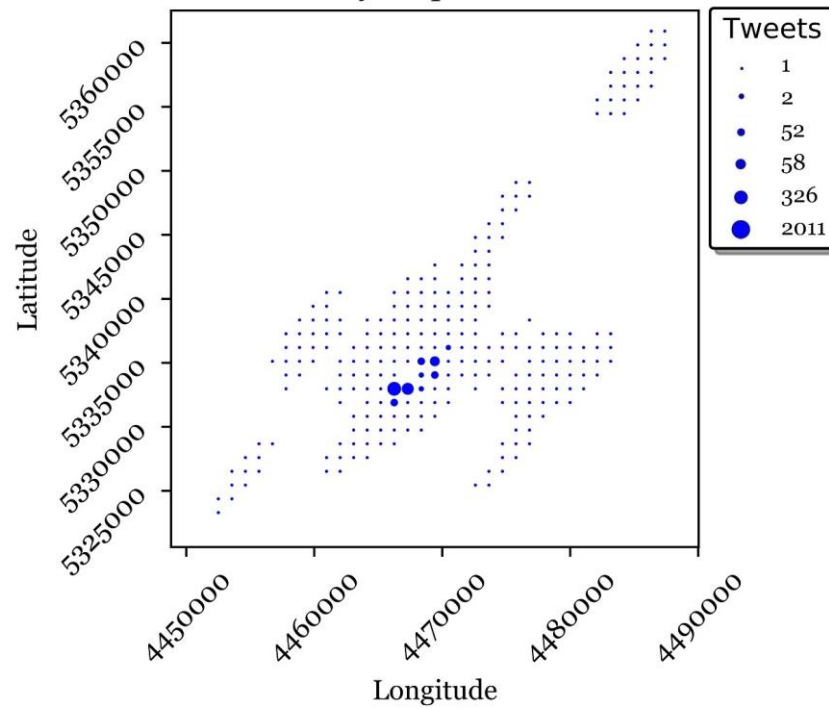


Figure 4.16 Graduated dot density map of tweets

Graduated Dot Density Map of the Oktoberfest Tweets

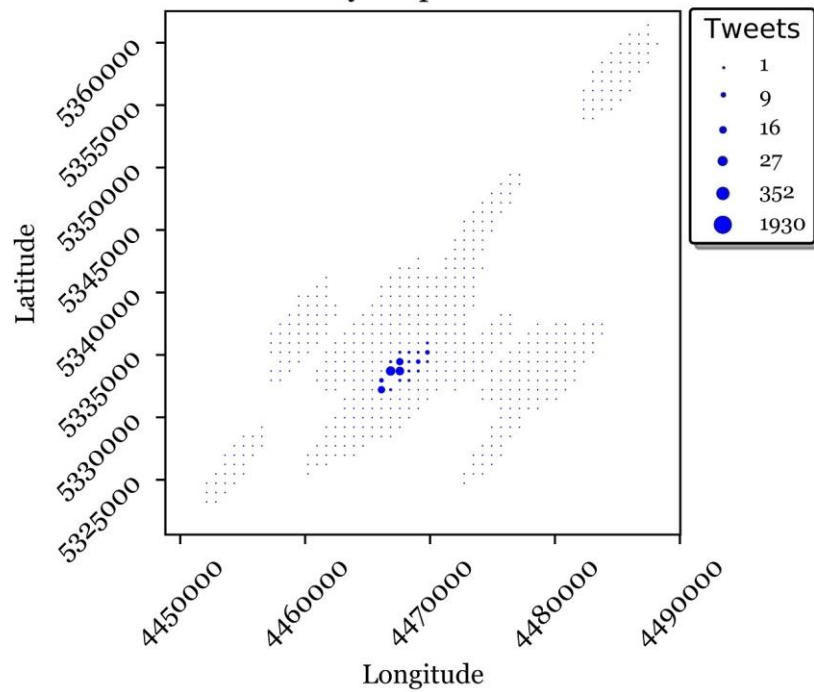


Figure 4.17 Graduated dot density map of tweets

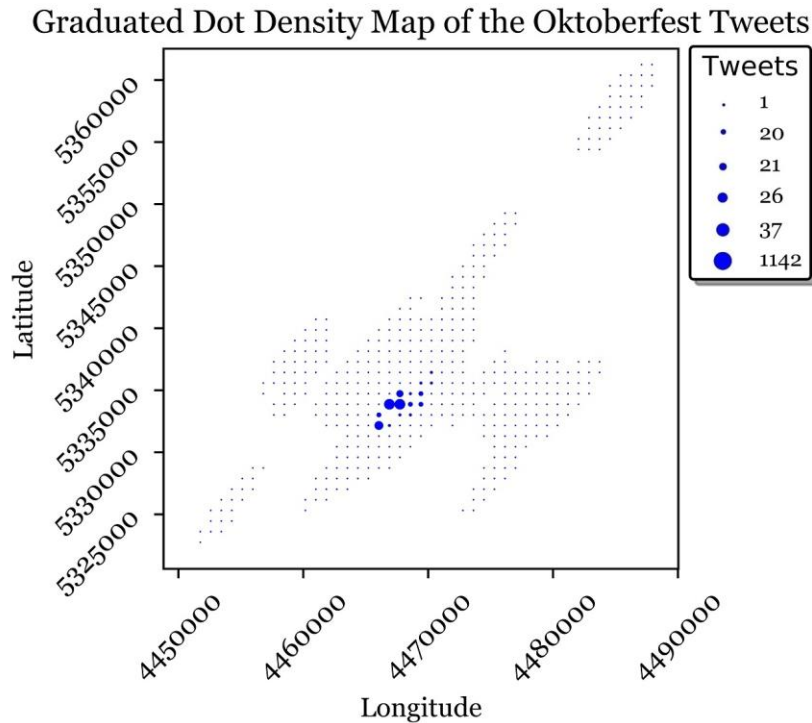


Figure 4.18 Graduated dot density map of tweets

4.2 User test

In order to evaluate the resulting maps and answer the research questions, an online survey was designed to ask some questions from the users. The users were 14 professional cartographers. The survey began with some comparison questions and finished with some tasks to evaluate the success of the dot maps and quantify the density. The questions were asked based on the heatmap, conventional and graduated dot maps of both datasets. The main purpose of the questions of the survey were as follows:

1. Comparison of the heatmap and dot maps:

The users were supposed to compare the heatmap with both types of dot maps separately and state; 1) which one better presented the overall density of the phenomenon; 2) which one provided a better quantitative estimation of the density in a specific part of the region (See questions 1 to 4 in the appendix).

2. Comparison of the conventional and graduated dot map:

The users were asked to compare the conventional and graduated dot maps separately and answer the same questions as in the previous part (See questions 5 and 6 in the appendix).

3. Estimation of the quantity in a specific region:

A specific region on the three map types was given to the users to indicate: 1) how easy was the estimation of the quantity of the phenomenon in the specified

region; and 2) an estimation of the quantity. (See questions 7 to 18 in the appendix).

4. Choosing the best type of map:

At the end of the survey, users were asked to choose the best type of map among heatmap, conventional and graduated dot maps to give a quantitative view of a phenomenon. Users were free to select none or more than one of the choices (See question 19 in the appendix).

For the first and second part of the survey, figures 4.1, 4.7 and 4.8 were used as heatmap, conventional and graduated dot maps of POI dataset, respectively. Likewise, figures 4.12, 4.15 and 4.16 were used as heatmap, conventional and graduated dot maps of the twitter dataset, respectively. I used maps with larger dots to make the survey easier and less time consuming for the cartographers taking part in the survey.

Figure 4.19 to Figure 4.21 show the specified regions for the maps of POI data for the third part of the survey. The red box was the chosen region in each map. The same process was followed for the twitter data. The size and location of the red box was chosen randomly to avoid any bias in the estimation task.

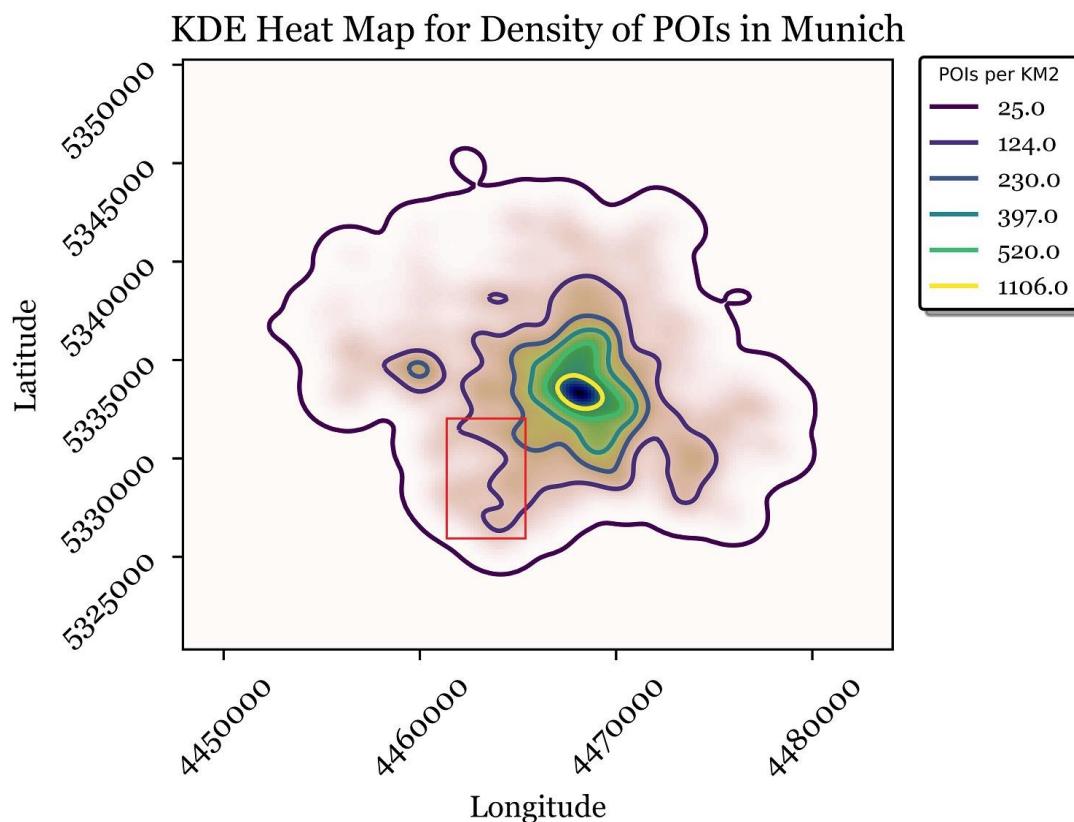


Figure 4.19 The specific region to quantify in KDE heatmap

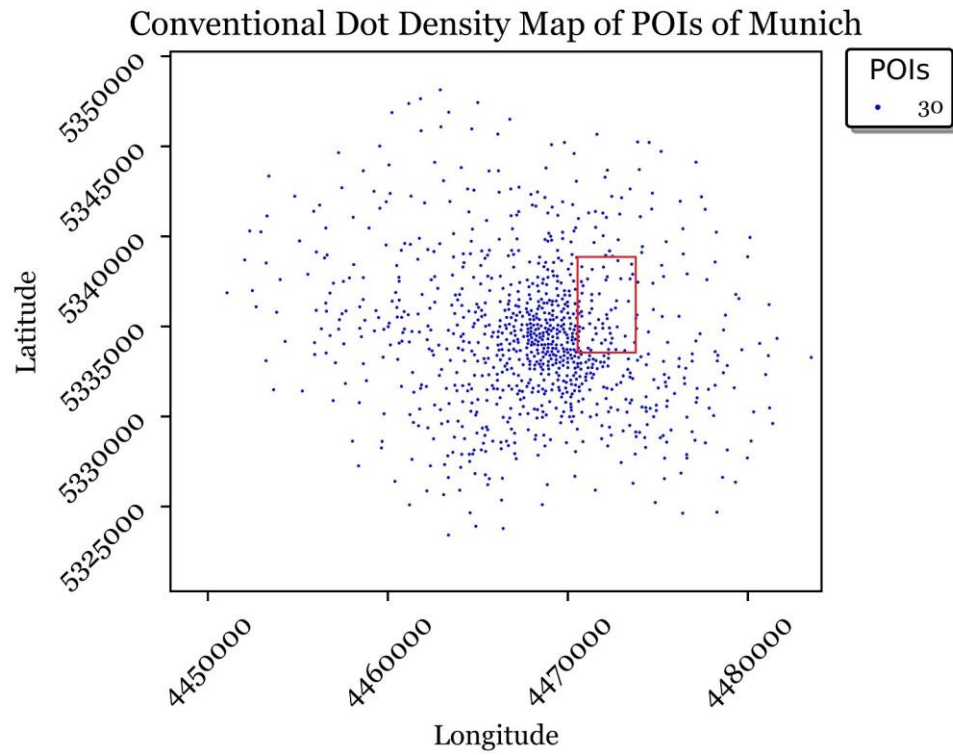


Figure 4.20 The specific region to quantify in conventional dot map

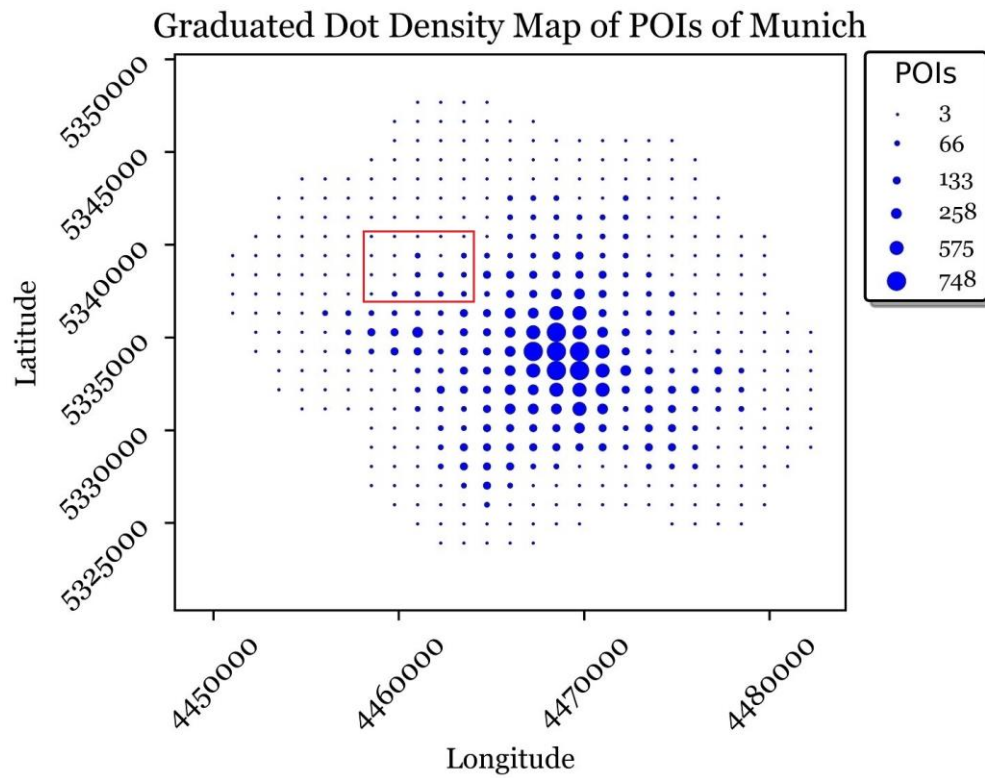


Figure 4.21 The specific region to quantify in graduated dot map

The results of the user test for the first and second part of the survey can be seen in Table 4.2 and Table 4.3. The numbers indicate the number of the participants who selected the corresponding choice. The numbers in parentheses are the percentage of the participants. Users had five choices based on the comparison of the first and second maps. For example, with respect to the first row, four users highly believed that the conventional dot map was better than the heatmap to get the overall density of POIs; six people thought that the conventional dot map was relatively better; there was no difference between them for one of the users; three of them stated that the heatmap was relatively better, and nobody thought that the heatmap was highly better. All of the percent numbers in parentheses were rounded.

			Comparison results (%)				
Maps	Comparison type	Data	C or G (highly)	C or G (relatively)	Equal	H (relatively)	H (highly)
C vs. H	Overall density	POI	4 (28)	6 (43)	1 (7)	3 (22)	0
		TW	2 (14)	8 (58)	2 (14)	1 (7)	1 (7)
	Quantification	POI	4 (28)	3 (22)	4 (28)	2 (14)	1 (7)
		TW	1 (7)	9 (65)	2 (14)	1 (7)	1 (7)
G vs. H	Overall density	POI	3 (22)	8 (57)	1 (7)	2 (14)	0
		TW	2 (14)	8 (58)	3 (22)	1 (7)	0
	Quantification	POI	2 (14)	8 (58)	1 (7)	2 (14)	1 (7)
		TW	2 (14)	7 (50)	2 (14)	2 (14)	1 (7)
1Conventional dot map = C Graduated dot map = G Heatmap = H Tweets = TW							

Table 4.2 The results of the user test for the first part of the survey

Maps	Comparison type	Data	Comparison results (%)				
			C (highly)	C (relatively)	Equal	G (relatively)	G (highly)
C vs. G	Overall density	POI	2 (14)	6 (43)	1 (7)	5 (35)	0
		TW	1 (7)	6 (43)	3 (22)	4 (28)	0
	Quantification	POI	1 (7)	5 (35)	3 (22)	3 (22)	2 (14)
		TW	2 (14)	5 (35)	5 (35)	2 (14)	0

Table 4.3 The results of the user test for the second part of the survey

The user test was focused on overall density and quantification in the dot maps. As can be seen from the results of the first step of the user test, at least 70% of the users believed that the dot maps were more appropriate to get an overall density of the phenomenon regardless of the data. A maximum of 22% of the users preferred a heatmap to the dot maps to get an overview of the overall density of the phenomena. Depending on the dataset and dot map type, at least 50% percent of the users thought that the dot maps are a better visualization technique to get a quantitative estimation from both of the

¹ This legend is applicable for all of the tables and figures in this chapter.

phenomena. This number reached 72% for different datasets and dot maps. A maximum of 21% of the users chose a heatmap to get a quantitative estimation of a specific area.

Conventional and graduated dot maps were also compared in the second step of the survey. Up to 57% of the users preferred conventional dot maps to graduated dot maps for overall density, while up to 35% chose graduated dot maps in the two datasets. A maximum of 49% thought that conventional dot map is better for estimation, while this number was not more than 36% for people who favored the graduated dot map.

Furthermore, the overall results of the first and second part of the survey can be observed in a radar chart, as in Figure 4.22. The tendency of the users based on the map type and also the usage of the map for overall density and quantification of a specific area is presented in the chart.

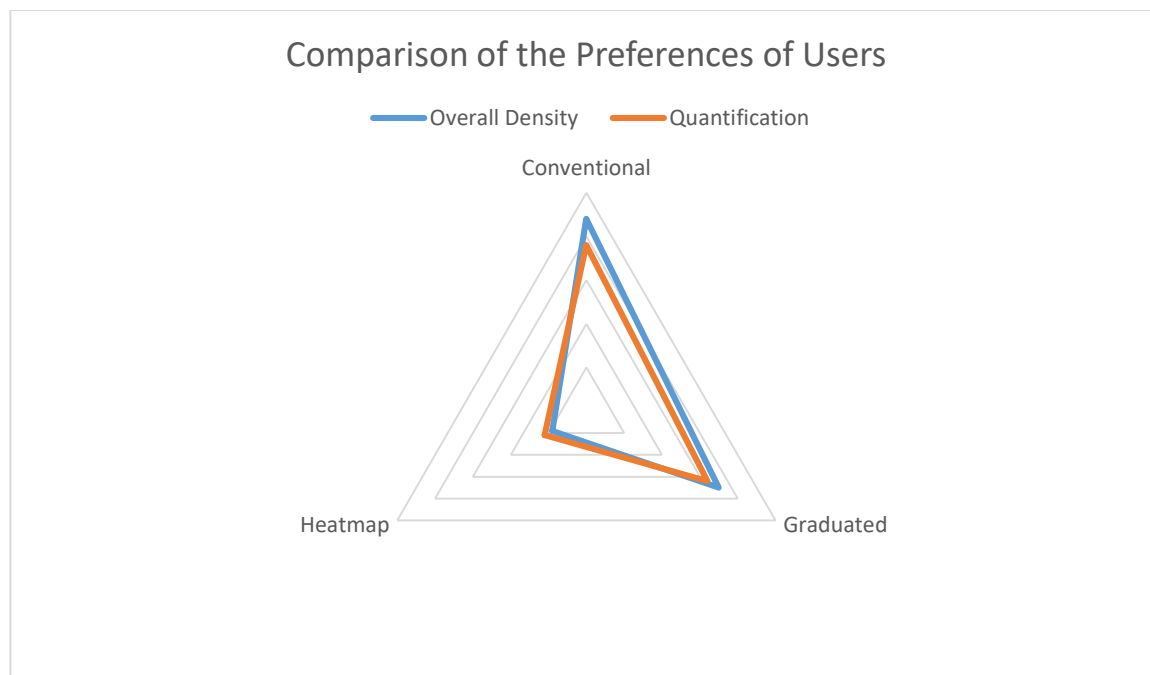


Figure 4.22 The comparison of the preference of users in the first and second part of the survey

The results of the third part of the test are depicted in Table 4.4 and Table 4.5. The numbers indicate the number of the participants who selected the corresponding choice. The numbers in parentheses are the percentage of the participants. Users stated how easy the estimation of the quantity of the phenomenon in the red box was and also gave their estimated number if applicable. For example in Table 4.4, with respect to the first row, one of the users thought that it was very easy to estimate the quantity of POIs in the red box in the conventional dot map; four people found it easy; two cartographers believed that it was neither easy and nor difficult; it was hard for six people to do the task and rather difficult for one of them. In the first row of Table 4.5, 13 out of 14 users completed the estimation task. The exact number of POIs in the red box was 1,740. The average counted number was 1,546, and the standard deviation of estimations was 765

based on the true number. The minimum and maximum of estimations were 1,000 and 4,000, respectively. Error percentage was 11% which is calculated by equation 5.1. All of the numbers were rounded. The inapplicable fields were filled by dashes (---).

$$\text{Error percentage} = \left| \frac{\text{True number} - \text{Mean}}{\text{True number}} \right| \quad \text{Eq. (5.1)}$$

		Comparison results (%)				
Maps	Data	Very easy	Easy	Medium	Hard	Very hard
C	POI	1 (7)	4 (28)	2 (15)	6 (43)	1 (7)
	TW	0	4 (28)	4 (28)	5 (37)	1 (7)
G	POI	1 (7)	4 (28)	3 (22)	6 (43)	0
	TW	2 (14)	6 (43)	0	6 (43)	0
H	POI	0	0	0	1 (7)	13 (93)
	TW	0	0	0	1 (7)	13 (93)

Table 4.4 The comparison results in the third part of the survey

Maps	Data	Successful counters	True Number	Mean	Std. Dev.	Error Percentage	Min.	Max.
C	POI	13	1,740	1,546	765	11%	1,000	4,000
	TW	13	594	499	319	16%	400	630
G	POI	14	525	730	108	39%	539	1,400
	TW	14	2,402	1,544	1,429	36%	350	4,050
H	POI	---		---	---		---	---
	TW	---		---	---		---	---

Table 4.5 The estimation results in the third part of the survey

In the third step, they were asked to estimate the number of points of a phenomenon in a specific region. From 28% to 35% of the users stated that estimation is an easy task in conventional dot maps, while from 44% to 50% believed that it is a hard task. Depending on the dataset, from 35% to 57% found it easy to estimate using the graduated dot map and 43% selected the hard choice for this type of map. Everyone believed that the task is hard in heatmaps.

Only one user could not do the estimation task on different dot maps of the two datasets. However, the level of satisfaction was different between graduated and conventional dot maps based on the datasets. Additionally, the quality of estimation was different based on the map and data. The average of estimation for the POI dataset in the conventional dot map was 1,546, which is close to the true number. However, the standard deviation was 765, which proves that there are estimations far from the average and the true number. The error percentage was 11%. The average of estimation for the twitter dataset in the conventional dot map was 499, which is very close to the true number. The standard deviation of the estimations was 319, which is not very low and not very high. The error percentage was 16%. The average of estimation for the POI dataset in the graduated dot map was 730, which was a good estimation in comparison to the true number. The

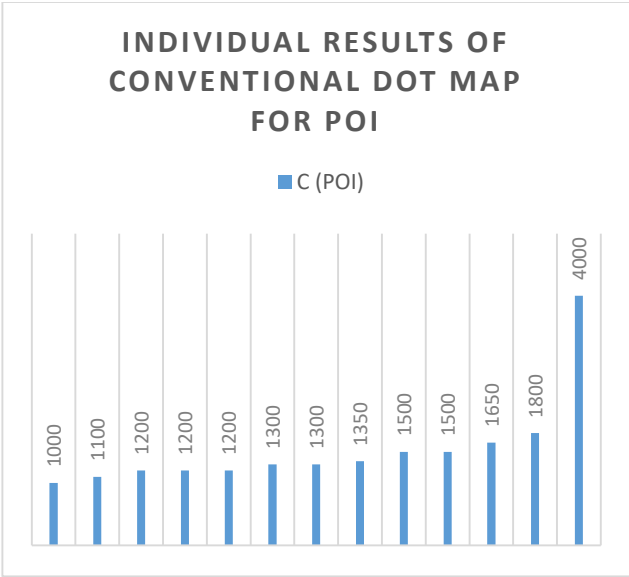
standard deviation of 108 proved that the estimations were not too scattered. The error percentage was 39% for this dataset. The average of the estimation for the twitter dataset in the graduated dot map was 1,544, which was not close to the real number and was not considered as a good estimation, although some of the users were able to estimate correctly. Similarly, the high standard deviation of 1,429 proved that the estimation was scattered and not very precise. The error percentage was 36%. It can be inferred from the survey that both the conventional and graduated dot maps were preferable to the heatmap for both the overview of the density and the quantification of the points. However, each one might have a superiority over the other one based on the dataset. Estimation task shows more error percentage in graduated dot maps. The homogeneity, the contrast between the most dense areas and the least dense areas of the map are the effective parameters for this dominance. In addition, the number of points is another factor to choose one of these maps. Too crowded maps might persuade a user to choose the graduated dot map due to less number of dots.

Moreover, Figure 4.23 presents all the individual results of the estimation for both of the map types and datasets. Figure (a) proves that only one of the users estimated the number of POIs far from the other users', the average and the true number using the conventional dot map. Figure (b) shows that almost all of the users could quantify the number of tweets close to the real number using the conventional dot map. Figure (c) states that 12 out of 14 people were able to estimate the number of POIs close to the real number using a graduated number. Figure (d) indicates that only 6 people out of 14 could guess a number around the true number. Figure 4.23 verifies that most of the participants were successful in estimating the number of points in the first three tasks and about half of the users were successful in the last task.

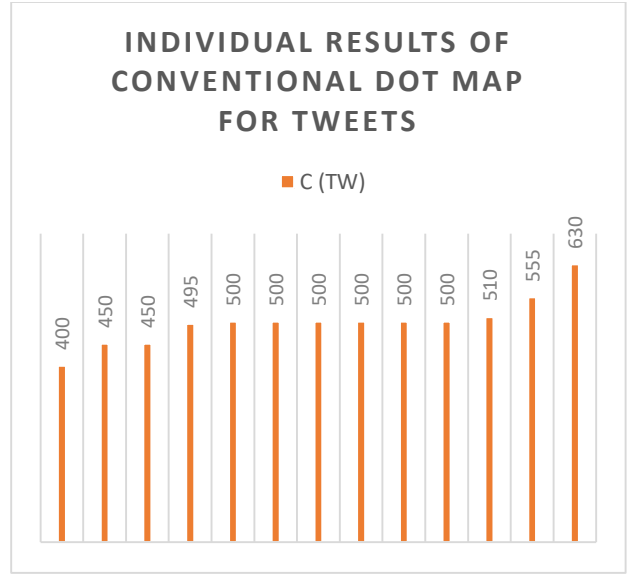
It is possible to compare the results of the third part of the survey in Figure 4.24. The average of estimations, minimum and maximum of the estimations, the standard deviation based on the average and the true numbers can be observed and compared for each map and dataset.

The results of the fourth part of the survey are represented in Table 4.6. Users chose which density map was the best to give a quantitative view to the user.

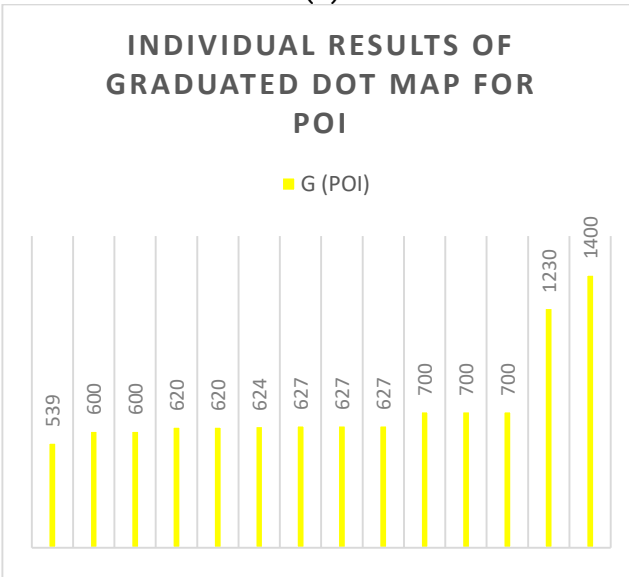
No one believed that the heatmap was a better approach to estimate the quantity of a phenomenon than the dot maps. Although the ideas were different based on the type of dot map and dataset, a large majority of the users preferred dot maps to heatmaps to visualize the overall density and get a quantitative estimation from the map. About 64% of the users favored graduated dot maps, 14% favored the conventional dot maps and 22% believed that both of them are good density maps to get a quantitative view from the region.



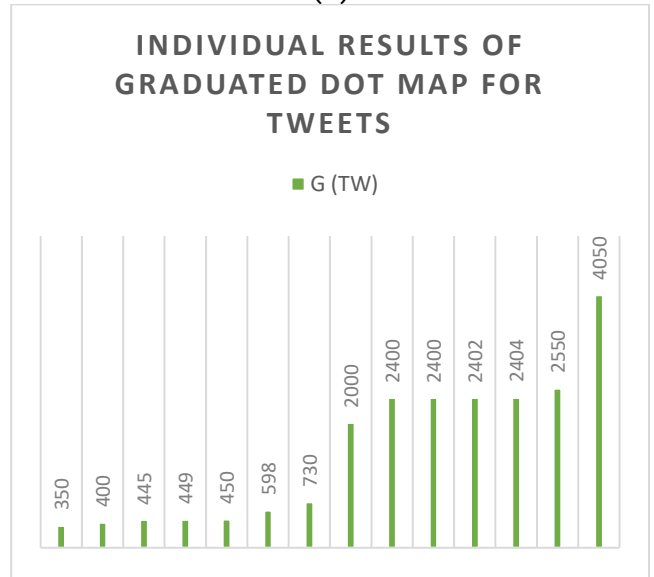
(a)



(b)



(c)



(d)

Figure 4.23 Individual results of the third part of the survey

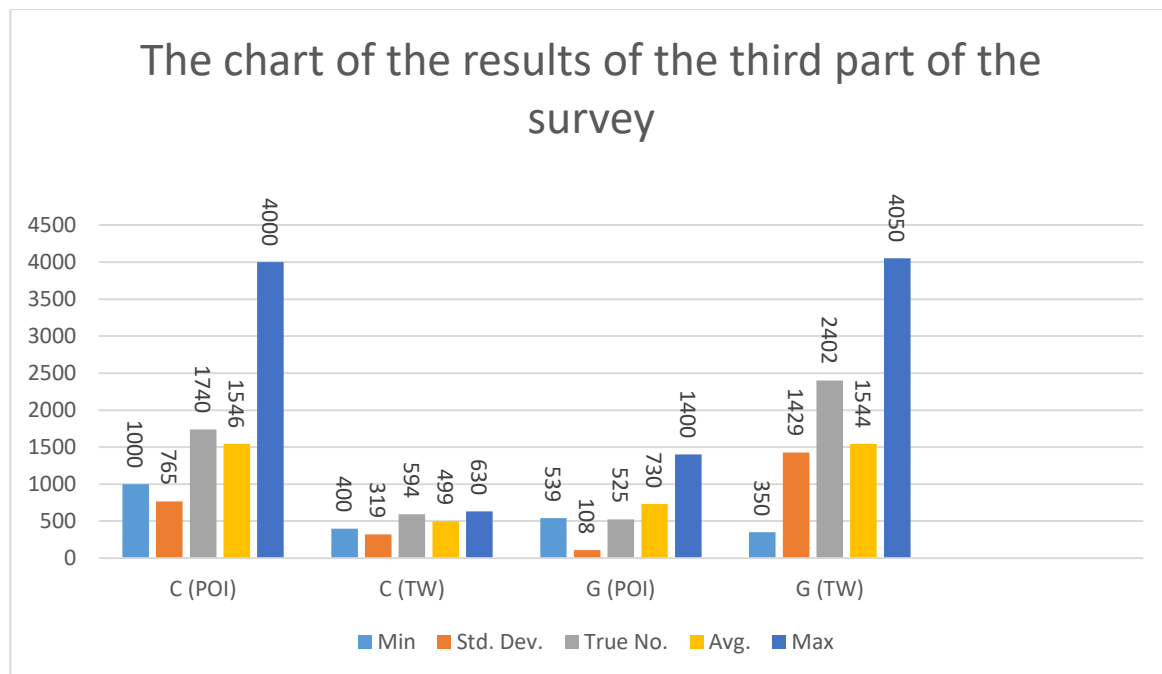


Figure 4.24 The chart of the results of the user test for the third part of the survey

Map type	C	G	H	None of them
Single selections (%)	2 (14)	9 (64)	0	0
Multiple selections (%)	3 (22)		0	

Table 4.6 The results of the user test for the fourth part of the survey

5 Discussion

In this chapter, it is possible to answer the research questions.

1. Which criteria should be considered to make a dot map which reflects a density in the best possible way?

Based on the results in the fourth chapter, a number of criteria, including resolution, dot size, dot value, maximum and minimum of occupation of a cell by dots and the randomness of dot placement affect the quality of dot maps. Adaption of these criteria highly depends on the data and the desired result of the cartographer.

2. What is the difference between the KDE heatmaps and KDE dot maps in quantization and density presentation?

The survey proved that the developed dot maps have some advantages over commonly used heatmaps in showing density. Dot maps do not use colors and cryptic density values; they present a quantitative estimation for any desired specific region on the map and give a better and more understandable overall density of a phenomenon. In other words, dot maps are able to present the density better both in quantification objectives and getting a density overview.

3. Do the final dot maps provide quantitative information to the user?

Based on the results of the survey, a vast majority of the users found dot maps a better visualization tool to get quantitative information from a desired area. The participants were able to estimate numbers for specific areas using the dot maps. Although not all of their quantifications were exact, most of the estimations were close to the true numbers. These estimations were obtained, while none of the users could get a number from the heatmaps. It shows that a minority of users who preferred a heatmap in the first part of the survey, they could not give any estimation based on the heatmap in the third part of the survey. In the fourth part of the survey, none of them chose heatmaps anymore. This change in their idea shows the dominance of dot maps over heatmaps for quantification of the density.

4. What are the advantages and disadvantages of graduated and conventional dot maps?

Although the users enjoyed the graduated dot maps more than the conventional dot maps in the last part of the survey, the estimated numbers and answers to some questions in the second and third part of the survey proved that both might have superiority based on the data and objective of density presentation. The heterogeneity of the data is an important factor to choose one of these types of dot maps. In the estimation part of the survey for the twitter dataset, the number of wrong estimations were very high. It showed that a graduated dot map was not the best choice for density presentation for this dataset and a conventional dot map was a better option. The objective of the cartographer is another significant factor in this selection. It depends on the interests of the cartographer.

Easier estimation, detailed estimation, calculation process and the effective parameters are vital, effective criteria on the objective of a user. Graduated dot maps contain fewer dots. The different classes of dot sizes could be helpful to show the density to the map reader. The developed graduated dot maps are completely regular and without overlap and coalescence. On the other hand, since the conventional dot maps only contain one dot value, the user can easily estimate the quantity by a simple multiplication. The size of dots allows a detailed estimation more than graduated dot maps because the large size of dots in some parts does not allow a very detailed estimation. For example, if the user is interested in estimating the number of points in the yellow box in Figure 5.1. it is difficult to quantify the number of points in the right side due to the exclusion of parts of the large dots from the box. A conventional dot map (Figure 5.2) is a better solution for this problem.

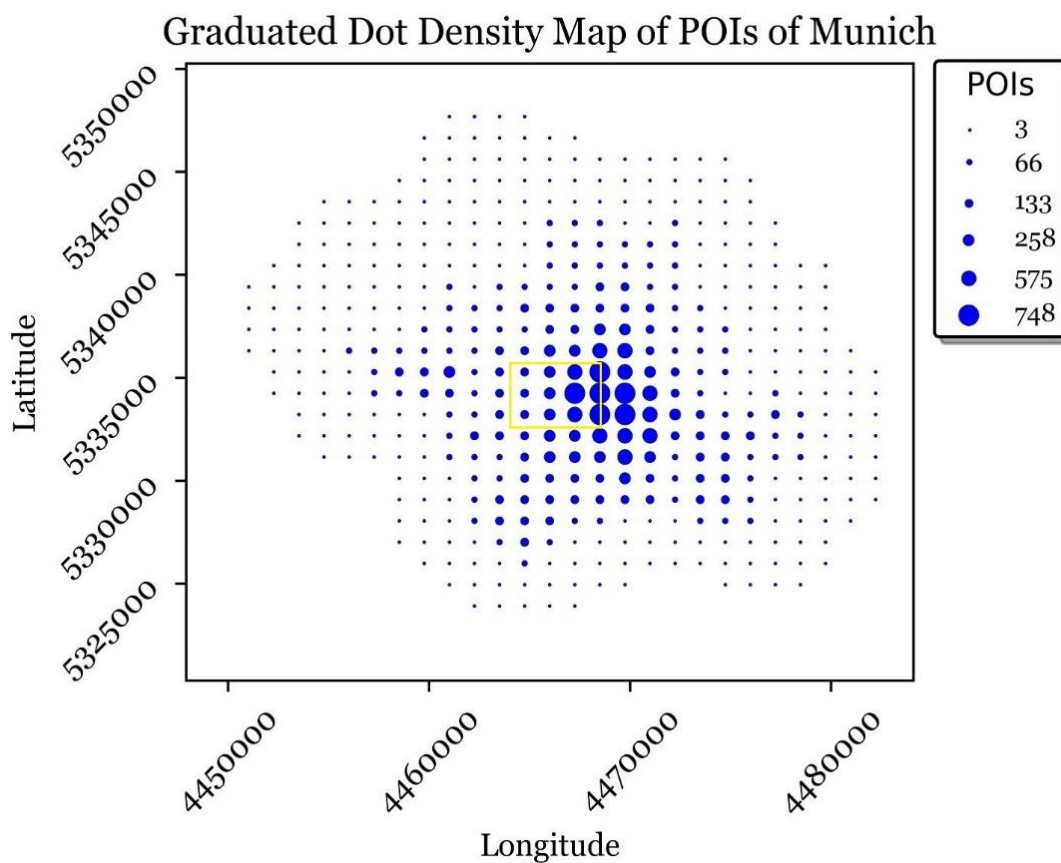


Figure 5.1 A graduated dot map for a detailed estimation

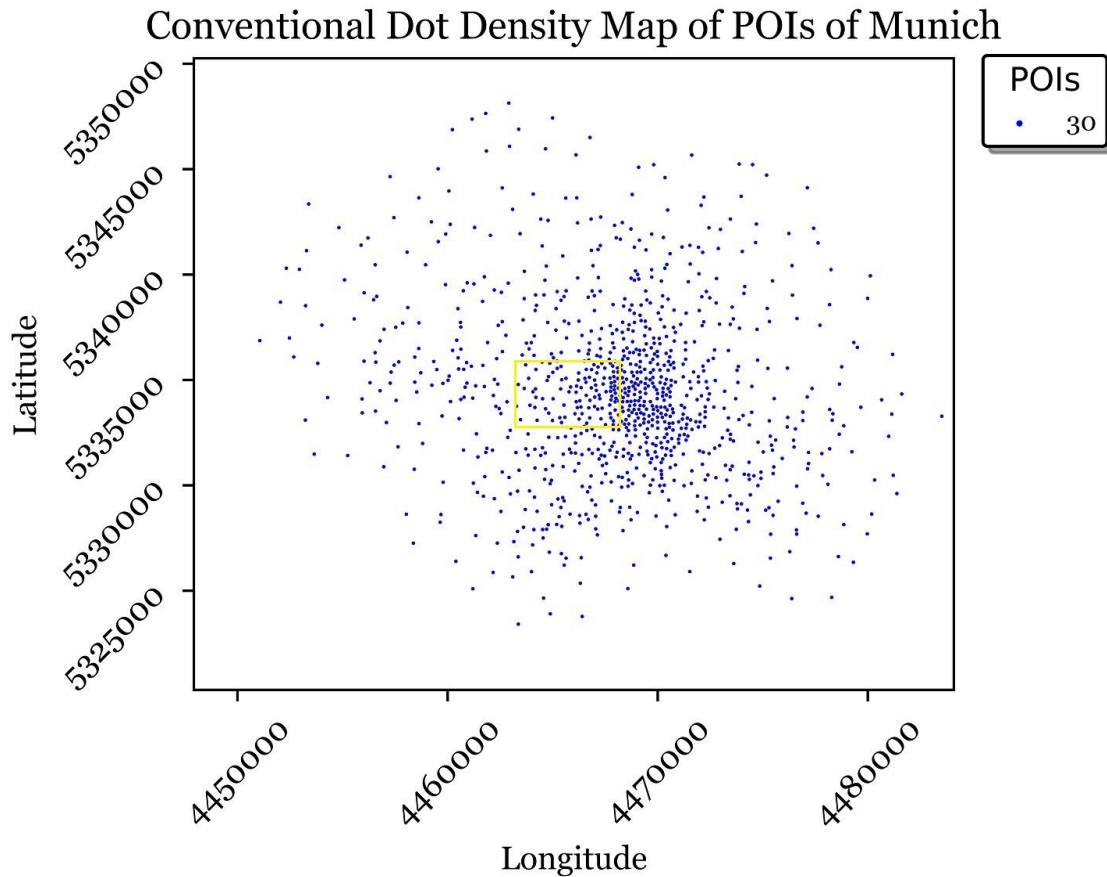


Figure 5.2 A conventional dot map for a detailed estimation

Table 5.1 shows a summary of advantages and disadvantages of both types of developed dot maps.

Conventional Dot Map		Graduated Dot Map	
Pros	Cons	Pros	Cons
Fast calculation	Greater number of dots	Fewer number of dots	More calculation
More detailed estimation	Randomness of dots	Regularity of dots	Less detailed estimation
Only one dot value	More parameters in the algorithm	Fewer parameters in the algorithm	Various dot values
No need to classification	More iterative processes for settings	Fewer iterative processes for settings	Requirement to a classification method

Table 5.1 Advantages and disadvantages of both types of developed dot maps

6 Conclusion

6.1 Thesis conclusion

This thesis aimed to propose an approach to enhance the visualization of the density of a phenomenon quantitatively to overcome the issues of commonly used KDE heatmaps. Although heatmaps are a common visualization tool to show the density of a phenomenon, they do not present a quantitative estimation for a specific region with their color schemes and the cryptic density values. Heatmaps are not considered as the best method of density visualization. However, running KDE function produces a raster. This raster contains a large number of cells with different values which was used as a base to create dot maps. The raster is a grid with many equal sized cells and different values, which are helpful to make a dot map. Two algorithms were developed to make two types of enhanced dot maps, including graduated and conventional dot maps. Both are made based on the grid provided by the heatmap. The number of cells of the grid were based on the chosen resolution. Based on the value of each cell, a number of equal sized dots were placed pseudo randomly on a conventional dot map. Only one dot was placed regularly on each cell of a graduated dot map. There is no overlap and coalescence among dots.

Based on previous studies e.g., (Slocum et al., 2009) and (Lavin et al., 1986), results of this thesis and the evaluation of the results in the survey, enhanced dot density maps are suggested as a better method than heatmaps both to get an overview of the density and quantification of the density. The statistics of the survey show a definitive dominance of the proposed dot maps over heatmaps, which is a common map type to show density. Giving a better quantitative view from different parts of a region is the main reason for the superiority of dot maps. Dot maps help users to quantify the number of points of a phenomenon in any favorable region on the dot map. The user test proved that although there were some differences in the estimated numbers, those numbers were close to the true number in most of the cases. Those results verified the dominance of dot maps over heatmaps which could not give any quantitative estimation to the participants of the survey in any part of the studied area. It is possible to change the values and size of dots, the resolution and even switch from one type to another type of dot map based on the data and the objective of the cartographer to create the most suitable density maps.

The dot maps have good characteristics in map reading. They are simpler. There is no need to distinguish the cryptic density values within isolines using colors. Dot maps use only one color and one geometrical shape. An easy arithmetic calculation gives the estimation to the user. Dot maps are more helpful when detailed quantifications are required. A user is able to quantify the number of points in very small part on the map. If the map does not help to estimate, adaptation of resolution, dot value, dot size and even map types are other alternatives to create the most favorable dot density map. Dot maps

do not use any cryptic approach to show the density. Their density presentation is much more obvious since they use precise values.

There is no definite advantage between graduated and conventional dot maps because the results of this research show varied opinions based on the questions and tasks of the survey and the data. However, most of the users preferred the graduated dot map to the conventional dot map at the end of the experiment. The estimation process might be easy in both types of the dot maps. It is easy in conventional dot maps because all of the dots have equal size and value, but the number of dots is greater and the estimation task might take more time. On the other hand, the advantage of a graduated dot map is that the number of dots is fewer and the dot positions are completely regular and separate. These advantages make the estimation easier. However, there are disadvantages in graduated dot maps. The calculation of the number of different sizes of dots is harder. Furthermore, it might be difficult to distinguish the exact class of a dot based on the legend of the map.

6.2 Outlook

The results of the project were evaluated by professional cartographers and benefits of the proposed approach were proved. However, the precise relationship between some of the criteria such as dot value, resolution, P_{\min} should be investigated further. Although we found the inverse relation between the dot size and the resolution in a conventional dot map, it might be possible to create a formula to calculate their exact numeric relation. Moreover, the effect of adapting P_{\min} in the conventional dot mapping can be more precisely investigated. In this thesis, it was constant for each dataset. Future research in this field is recommended to expand more logical correlations among varied criteria in this topic. Researchers might be able to develop the algorithm and make it more automatic so that cartographers do not have to set many parameters in an iterative process. The future research should make the process simpler, shorter and more automatic.

References

- ©Atlas of Switzerland, Sheet 51. (2020, July 30). *Graduated dot map*. <https://atlasderschweiz.ch/>
- Arnold, N. D., Jenny, B., & White, D. (2017). Automation and evaluation of graduated dot maps. *International Journal of Geographical Information Science*, 31(12), 2524–2542. <https://doi.org/10.1080/13658816.2017.1359747>
- Axismaps. (2020, July 29). Axis Maps. <https://www.axismaps.com/guide/univariate/dot-density/>
- D, B., & Mr, T. (2007, April). *Rainbow color map (still) considered harmful*. IEEE Computer Graphics and Applications; IEEE Comput Graph Appl. <https://doi.org/10.1109/mcg.2007.323435>
- Duchowski, A. T., Price, M. M., Meyer, M., & Orero, P. (2012). Aggregate gaze visualization with real-time heatmaps. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 13–20. <https://doi.org/10.1145/2168556.2168558>
- Eng, S. B., Werker, D. H., King, A. S., Marion, S. A., Bell, A., Issac-Renton, J. L., Irwin, G. S., & Bowie, W. R. (1999). Computer-generated dot maps as an epidemiologic tool: Investigating an outbreak of toxoplasmosis. *Emerging Infectious Diseases*, 5(6), 815–819. <https://doi.org/10.3201/eid0506.990613>
- GfK GeoMarketing GmbH. (n.d.). *Dot density map of industrial workers in Baden-Württemberg, Germany* [Map].
- Gomes, E. (2017). Creating a Dot Density Map: Resident Population in Mainland Portugal. *The Cartographic Journal*, 54(2), 157–162. <https://doi.org/10.1080/00087041.2016.1148106>
- Hey, A. (2012). Automated Dot Mapping: How to Dot the Dot Map. *Cartography and Geographic Information Science*, 39(1), 17–29. <https://doi.org/10.1559/1523040639117>
- Hey, A., & Bill, R. (2014). Placing dots in dot maps. *International Journal of Geographical Information Science*, 28(12), 2417–2434.
- Kimerling, A. J. (2009). Dotting the Dot Map, Revisited. *Cartography and Geographic Information Science*, 36(2), 165–182. <https://doi.org/10.1559/152304009788188754>
- Krisp, J. M., & Spatenkova, O. (2010). Kernel Density Estimations for Visual Analysis of Emergency Response Data. In M. Konecny, S. Zlatanova, & T. L. Bandrova (Eds.), *Geographic Information and Cartography for Risk and Crisis Management: Towards Better Solutions* (pp. 395–408). Springer. https://doi.org/10.1007/978-3-642-03442-8_27

Lavin, S., Hobgood, J., & Kramer, P. (1986). Dot-Density Shading: A Technique for Mapping Continuous Climatic Data. *Journal of Climate and Applied Meteorology*, 25(5), 679–690. JSTOR.

leistiar. (2012). *Proportional Symbols Map* [Photo]. <https://www.flickr.com/photos/89769525@N08/8242929652/>

Li, C., Baciú, G., & Han, Y. (2014). Interactive visualization of high density streaming points with heat-map. *2014 International Conference on Smart Computing*, 145–149. <https://doi.org/10.1109/SMARTCOMP.2014.7043852>

Mackay, J. R. (1949). Dotting the dot map: An analysis of dot size, number, and visual tone density. *Surveying and Mapping*, 9(1), 3–10.

McMaster, R. (1997). In Memoriam: George F. Jenks (1916-1996). *Cartography and Geographic Information Systems*, 24(1), 56–59. <https://doi.org/10.1559/152304097782438764>

Robinson, A. H., Morrison, J. L., Muehrcke, P. C., Kimerling, A. J., & Guptill, S. C. (1995). *Elements of Cartography* (6 edition). Wiley.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2009). *Thematic Cartography and Geovisualization, 3rd Edition* (3rd ed.). Pearson. <https://www.pearson.com/us/higher-education/program/Slocum-Thematic-Cartography-and-Geovisualization-3rd-Edition/PGM94810.html>

Tang, L., Kan, Z., Zhang, X., Sun, F., Yang, X., & Li, Q. (2016). A network Kernel Density Estimation for linear features in space–time analysis of big trace data. *International Journal of Geographical Information Science*, 30(9), 1717–1737. <https://doi.org/10.1080/13658816.2015.1119279>

Terrell, G. R., & Scott, D. W. (1992). Variable Kernel Density Estimation. *The Annals of Statistics*, 20(3), 1236–1265. JSTOR.

Wilkinson, L., & Friendly, M. (2009). The History of the Cluster Heat Map. *The American Statistician*, 63(2), 179–184. <https://doi.org/10.1198/tas.2009.0033>

Yu, W., Ai, T., & Shao, S. (2015). The analysis and delimitation of Central Business District using network kernel density estimation. *Journal of Transport Geography*, 45, 32–47. <https://doi.org/10.1016/j.jtrangeo.2015.04.008>

Yuan, K., Cheng, X., Gui, Z., Li, F., & Wu, H. (2019). A quad-tree-based fast and adaptive Kernel Density Estimation algorithm for heat-map generation. *International Journal of*

Geographical Information Science, 33(12), 2455–2476.
<https://doi.org/10.1080/13658816.2018.1555831>

Appendix

The User test:

The study focuses on visualization of the density of a phenomenon. Two phenomena are used in this survey: density of points of interest (POI) and density of tweets about Oktoberfest both in Munich, Germany.

1. You see a heatmap (Figure 4.1) and a conventional dot map (Figure 4.7) which show the density of POIs in Munich. The comparison of these two indicates that the dot map

...

... is better to perceive the overall density than the heat map.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
-------------------	----------	----------------------------	-------	----------------

... provides a better quantitative estimation of the density in any part of the map than the heat map.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
-------------------	----------	----------------------------	-------	----------------

(If you think that the inverse of the statements is correct, you can choose "disagree" or "strongly disagree" and if they are equal, you can choose "Neither agree nor disagree")²

2. You see a heatmap (Figure 4.1) and a graduated dot map (Figure 4.8) which show the density of POIs in Munich. The comparison of these two indicates that the dot map ...

... is better to perceive the overall density than the heat map.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
-------------------	----------	----------------------------	-------	----------------

... provides a better quantitative estimation of the density in any part of the map than the heat map.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
-------------------	----------	----------------------------	-------	----------------

² This explanation is applicable for questions 1 to 6.

3. You see a heatmap (Figure 4.12) and a conventional dot map (Figure 4.15) which show the density of Oktoberfest tweets in Munich. The comparison of these two indicates that the dot map ...

... is better to perceive the overall density than the heat map.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
-------------------	----------	----------------------------	-------	----------------

... provides a better quantitative estimation of the density in any part of the map than the heat map.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
-------------------	----------	----------------------------	-------	----------------

4. You see a heatmap (Figure 4.12) and a graduated dot map (Figure 4.16) which show the density of Oktoberfest tweets in Munich. The comparison of these two indicates that the dot map ...

... is better to perceive the overall density than the heat map.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
-------------------	----------	----------------------------	-------	----------------

... provides a better quantitative estimation of the density in any part of the map than the heat map.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
-------------------	----------	----------------------------	-------	----------------

5. You observe a conventional (Figure 4.7) and a graduated dot map (Figure 4.8) which show the density of POIs in Munich. The comparison of these two indicates that the conventional dot map ...

...is better than the graduated dot map to perceive the overall density.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
-------------------	----------	----------------------------	-------	----------------

...provides a better quantitative estimation of the density in any part of the map than the graduated dot map.

Strongly
disagree

Disagree

Neither agree
nor disagree

Agree

Strongly agree

6. You observe a conventional (Figure 4.15) and a graduated dot map (Figure 4.16) which show the density of Oktoberfest tweets in Munich. The comparison of these two indicates that the conventional dot map ...

...is better than the graduated dot map to perceive the overall density.

Strongly
disagree

Disagree

Neither agree
nor disagree

Agree

Strongly agree

...provides a better quantitative estimation of the density in any part of the map than the graduated dot map.

Strongly
disagree

Disagree

Neither agree
nor disagree

Agree

Strongly agree

7. How is it possible to estimate the number of POIs in the red rectangular sample in this conventional dot map (Figure 4.20)?

Very easy

Easy

Not very easy
and not hard

Hard

Too hard

8. Please estimate the quantity of POIs in the red rectangular sample in the conventional dot map (Figure 4.20).

If is not possible to obtain a number, please leave it empty.³

9. How is it possible to estimate the number of POIs in the red rectangular sample in this graduated dot map (Figure 4.21)?

Very easy

Easy

Not very easy
and not hard

Hard

Too hard

10. Please estimate the quantity of POIs in the red rectangular sample in the graduated dot map (Figure 4.21).

³ This explanation is applicable for question 8, 10, 12, 14, 16 and 18.

11. How is it possible to estimate the number of POIs in the red rectangular sample in this heatmap (Figure 4.19)?

Very easy	Easy	Not very easy and not hard	Hard	Too hard
-----------	------	-------------------------------	------	----------

12. Please estimate the quantity of POIs in the red rectangular sample in the heatmap (Figure 4.19).

13. How is it possible to estimate the number of tweets in the red rectangular sample in this conventional dot map?

Very easy	Easy	Not very easy and not hard	Hard	Too hard
-----------	------	-------------------------------	------	----------

14. Please estimate the quantity of tweets in the red rectangular sample in the conventional dot map.

15. How is it possible to estimate the number of tweets in the red rectangular sample in this graduated dot map?

Very easy	Easy	Not very easy and not hard	Hard	Too hard
-----------	------	-------------------------------	------	----------

16. Please estimate the quantity of tweets in the red rectangular sample in the graduated dot map.

17. How is it possible to estimate the number of tweets in the red rectangular sample in this heatmap?

Very easy	Easy	Not very easy and not hard	Hard	Too hard
-----------	------	-------------------------------	------	----------

18. Please estimate the quantity of tweets in the red rectangular sample in the heatmap.

19. After this experiment, which map was more helpful to estimate the quantity of a phenomenon?

You can choose more than one choice.

The conventional
dot map

The graduated dot
map

The heatmap

None of them

If you have any feedback to this questionnaire, please write in the box below.