Map Design in a Neogeography Environment

Iriael Jackson Mlay
March, 2010

Map Design in a Neogeography Environment

by

Iriael Jackson Mlay

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialisation: Geoinformatics.

Thesis Assessment Board:

Chair: Dr. O. Huisman
External Examiner: Prof. Dr.Sc. P. De Maeyer
Supervisor: Prof.Dr. M.J. Kraak
Second supervisor: Dr. R. Zurita-Milla

INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
ENSCHEDE, THE NETHERLANDS

Disclaimer

Abstract

The availability and affordability of digital devices such as cell phones, camera, GPS and access to internet has brought a remarkable trend in mapping systems. Individuals use these digital devices to collect spatial data at very large scale and disseminate it on internet platforms such as Google Earth, Google Maps, Flickr etc. at small scale. This process of collecting and displaying spatial data on the internet platforms is called neogeography and the associated product is neogeography map. Neogeography data are increasing rapidly as observed in these platforms due to contribution by individuals. These data can be used for various purposes or applications. It is important to view the maps at small scale so as to get the spatial relationship between objects, making it suitable for the intended use. However, these map objects seem to be too cluttered when the maps are displayed at a small scale.

This research concentrates on finding alternatives for presenting the cluttered neogeography maps in an understandable manner. One of the alternatives is the cartographic approach. Producing cartographic maps consumes both time and money. Another alternative is through summarization of neogeography topics. This was done by partitioning the neogeography data points into clusters using the clustering algorithms namely: k – means and DBSCAN. The clustered data points are converted into polygons using the bounding containers. The bounding containers used are extent tool, rectangle tool, circle tool and the convex hull tool. The summarized neogeography topics are visualized using three visualization techniques namely; application of colour to polygons, transparency and use of outline colours (hollow polygons). Conceptual interface as an option of presenting the summarized neogeography topics is also designed. Focus group was used to test if the idea of summarizing the neogeography topics to polygon or having an interface works. The result from the focus group shows that the cluttering problem of neogeography multiple topics at small scale can be minimized by summarizing the neogeography topics.

Key words: Neogeography map, clustering algorithms, bounding containers, summarized neogeography topics, cartographic map and neogeography interface.

Acknowledgements

Let me take this opportunity to express my thanks to all people who made it possible to accomplish this research.

First and foremost, I offer my sincerest gratitude to Prof. Dr. M. J. Kraak who as my first supervisor encouraged, guided and supported me throughout this study. It is true that without him, this thesis would not have been completed or written. Thank you very much.

Secondly, I am forever grateful to my second supervisor Dr. R. Zurita-Milla, who has supported me throughout my research with his patience and knowledge whilst giving me the room to work in my own way. He used much of his time to read and comment on my report that helped me to have a readable thesis in a consistent manner. To him, I say thank you.

Also, I would like to offer my gratitude to Erasmus Mundus for the financial support that enabled the successful pursuit of my MSc studies.

Special thanks go to ITC for giving me the chance to study Geoinformatics. The knowledge I acquired will be beneficial to me, my university and my country in general. I appreciate the knowledge and I say thank you very much.

My deepest gratitude goes to my family for their untiring love and support throughout my life. I am indebted to my wife Christiner Eliamani for her love and care throughout this period though we were apart.

Last but not least, thanks be to God for my life and seeing me through all challenges in the past 18 months. You have made my life more fulfilling. May your name be praised.

Table of contents

List of figures

List of tables

1. Introduction

1.1. Motivation and problem statement

‘A map is defined as a graphic representation of our environment’ [1]. The major function of a map is to represent the geographic reality. Well designed symbols are used for selection and abstraction of features from the reality. In order to ensure that the geographic information is effectively communicated in a map, some guidelines exist to help one to use symbols efficiently although these guidelines are not the absolute laws such as those used in other fields of science [2]. A cartographer is a person who studies and applies those guidelines to make a map. With a proper application of these guidelines, maps that meet user requirements can be produced.

For a number of centuries, only Cartographers or official agencies were able of making proper maps, but since the early 2000s any individual can make maps [3]. This is due to the developments in technology such as Web 2.0 where maps can be created by a large number of private citizens who are not trained in making maps [4]. Web 2.0 is an environment that enables individuals to interact with and integrate data from various sources [5-6]. Because of these developments new terms have emerged. These terms are ‘Neogeography’ [7-8], ‘user generated contents’ [9] and ‘volunteered geographic information, VGI’ [3]. Goodchild (2007) defines VGI as a special case of the more general web phenomenon of user generated content [3].

Turner (2006) defines neogeography as a “new geography” and consist of a set of techniques and tools that fall outside the realm of traditional GIS, Geographic Information Systems” [7]. Turner says neogeography combines the complex techniques of cartography and GIS and place them within the reach of users and developers. The users who are the general public participate voluntarily in recording and exchanging of spatial data. The recorded or exchanged spatial data has also been used for making maps using the available neogeography tools and internet as a platform. Neogeographer is a term that is used to represent the user who makes map in neogeography environment.

An individual citizen can collect spatial data by using digital devices like GPS, GPS enabled cell phones and digital camera and disseminate it to the Web also known as Web 2.0 [10]. An example of spatial data voluntarily collected by citizens can be found in many neogeography platforms such as Flickr, Google Earth, OpenStreetMap etc. The spatial data displayed into these platforms includes photographs and video together with their location components. Another example of spatial data collected by citizens can be found in Waarneming.nl. This website shows the biodiversity of the Netherlands and covers all species groups. Birds, plants, mammals, fish, insects etc. are among the species found on that website. It is the result of real time observation where the citizens use instruments such as pocket personal computer and smart phones during collecting and dissemination of spatial data. Both technological development and innovative mapping trend have impacts in the field of cartography and the associated output. However, it is initially assumed that the outputs in neogeography are of poor quality according to cartographic guidelines. The poor output is the result of failure to follow cartographic approaches during designing. This brings a gap between those maps designed by either a cartographer or by neogeographer who uses the available web tools. A difference in the output provided by the two is an aggregation of many factors among them is poor design, use of visual variables and symbols in neogeography. “Cartography is not an expressive art form in which the graphic elements may be rearranged at will” as is currently the case in neogeography [11]. Cartwright (2009) gave an example of how the use of symbol can change the aesthetics of a map as shown in Figure 1.1 below [12]. In figure 1.1, the geographic phenomenon shown on the left has been represented using the default symbol while on the right the same geographic phenomenon has been presented using the unique symbols. When considering only symbols to judge the two geographic phenomena, one can realize that the unique symbols used to represent the geographic phenomena on the right are poorly placed such that the map becomes difficult to interpret.

Figure 1.1: Symbology from the New Times travel Web page

Again, looking on the figure above, one can think of how to represent the quantitative or qualitative information in this mapping environment at a very small scale. Furthermore, one can imagine how this geographic representation will look like in small size devices such as Location Based Services. Impact of the symbols to devices that are characterized by small screen and low resolution, and the way the information derived from such devices can influence the individual decision. From the above examples it clear that there is need to improve the way information is presented in neogeography maps.

The maps in figure 1.1 above looks cluttered in cartographic perspective [13]. This is based mainly on the scale at which the maps have been displayed. When the map is at small scale and cluttered as shown in figure 1.1, it may be a good idea to summarize or combine the point features to form a polygon. This can make the coverage or extent of the geographic feature to be realized when visualized at a small scale.

1.2. Research identification

There exist a gap between cartographers and the neogeographers in terms of designing quality product (map). This could be due to failure to consider cartographic approaches during map design.

1.3. Main research objective

To assist neogeographers in making maps by combining the available neogeography tools and cartographic approaches.

1.3.1. Specific research objectives

- To investigate current mapping procedures used in neogeography environment and look at both their positive and negative aspects.

- To compare neogeography with current cartography and to explore the possibility of transferring cartographic design principles into neogeography.

- To design an interactive neogeography environment for a 'proper' map representation.

- To test and implement the interactive neogeography environment.

1.3.2. Research questions

Table 1.1 shows the research questions and the associated specific research objective.

Research Questions	Specific research objective
1. What is neogeography and what are the characteristics of neogeographers? 2. What characterizes a typical neogeography map? 3. What are the advantages and disadvantages of mapping in neogeography environment? 4. Which tools do neogeographers use to make their maps and what are their cartographic functionalities?	• To investigate current mapping procedures used in neogeography environment and look at both their positive and negative aspects.
5. Are neogeographers aware of the fact that their maps are 'bad' and that they could be 'improved' if cartographic guidelines were considered? 6. What is 'wrong' from cartographic design perspectives with neogeography maps? 7. Are cartographic design principles applicable to neogeography maps?	• To compare neogeography with current cartography and to explore the possibility of transferring cartographic design principles into neogeography.
8. What are the requirements of an interactive neogeography environment? 9. What is a 'proper' map representation in neogeography environment? 10. How can neogeography and cartographic approaches be combined in the design of neogeography maps?	• To design an interactive neogeography environment for a 'proper' map representation.
11. How to test and validate the design?	• To test and implement the interactive neogeography environment.

Table 1.1: Research objectives and the corresponding research questions

1.4. Innovation aimed at

To design an interactive environment for summarizing and representing the neogeography topics by considering the cartographic approaches.

This research will explore the existing cartographic approaches and available neogeography tools in summarizing the neogeography multiple topics. Designing of neogeography interface for representing the summarized maps is also a part of this research. This will be among the steps to reduce the gap between the people building maps (neogeographers) with the new web tools and the cartographic community.

1.5. Related work

Neogeography means "new geography [7]. Designing a map in neogeography environment mostly is still one sided task in the sense that some of the cartographic outputs designed is inbuilt within the software [5]. This limits the individuals to apply and modify them during mapping. Individuals are the main source of data and information in neogeography environment. By using simple application programming interfaces, they can create online maps. Thus relying on tools available in a given neogeography environment which are already designed for them.

Designing good looking maps in a neogeography environment have attracted individuals even if they are not trained cartographers [14]. They are interested in designing readable and understandable maps rather than designing a good coloured maps or symbols. Map lettering and detail presentations are key issues considered by the individuals in designing a good map. Figure 1.2 is the map of United Kingdom for national cycling organization. This map is an example of good looking neogeography map. This map was designed nicely and able to bring attention to fine details in the eyes of neogeographers, but what about cartographer?

Figure 1.2: Example of good looking neogeography map Source[14]

1.6. Research methodology

In order to make sure that the research questions are answered the procedure will be as shown in table 1.2.

Research question to be answered	Method
1. What is neogeography and what are the characteristics of neogeographers?	• Literature review + looking the characteristics of their maps.
2. What characterizes a typical neogeography map?	• Literature review.
3. What are the advantages and disadvantages of mapping in neogeography environment?	• Literature review.
4. Which tools do neogeographers use to make their maps and what are their cartographic functionalities?	• Literature review (manuals + help function) + looking the characteristics of their maps.
5. Are neogeographers aware of the fact that their maps are 'bad' and that they could be 'improved' if cartographic guidelines were considered?	• Literature review + usability research (focus group).
6. What is 'wrong' from cartographic design perspectives with neogeography maps?	• Literature review + usability research (focus group).
7. Are cartographic design principles applicable to neogeography maps?	• Literature review + usability research (focus group).
8. What are the requirements of an interactive neogeography environment?	• Map server + open layers + scripting language.
9. What is a 'proper' map representation in neogeography environment?	• Usability research (focus group).
10. How can neogeography and cartographic approaches be combined in the design of neogeography maps?	• Designing of an interactive environment. (Considering visual design guidelines i.e. Overview first, zoom and filter, then details-on-demand).
11. How to test and validate the design?	• Use of forests, schools, hotels, and restaurants data for the Netherlands, Belgium and Germany.

Table 1.2: Research methodology

1.7. Organization of the thesis

The thesis comprised of 7 chapters. The structure of the next 6 chapters is organized as follows.
Chapter 2 contains the definition of neogeography and the characteristic of the neogeographers. Tools used in collection and creation of neogeography data are also discussed. This assists understanding the importance and application of neogeography data and map. Chapter 3 provides the cartographical procedures for making maps which have been used for number of years. Cartographic procedures can be used as a means of investigating the communicability of neogeography maps to its audience through new design approach. Chapter 4 explain the new design approach for representing neogeography maps through summarizing the neogeography multiple topics. Different methods for summarizing the neogeography multiple topics exist. Two methods are chosen and implemented on four neogeography topics as explained in chapter 5. The result of summarized neogeography topics are discussed in chapter 6 through the usability research. Chapter 7 provides the conclusion and recommendation for future work.

2. Neogeography environment

2.1. Introduction

Neogeography environment can be considered as environment for collection and dissemination of geographic and non geographic data which has been facilitated by technological advancement. The advancement in technology is observed in digital devices such as GPS, mobile cell phones etc, computer hardware and software, and high band width internet connections. The technologies are simple to use and inexpensive. Individuals are using those technologies to make their own maps or share their data on the internet.

This chapter describes the tern neogeography as currently used in mapping environment, the characteristics of neogeography datasets and importance of neogeography maps. The tools for mapping in this environment and the characteristics of map makers will be discussed. Furthermore, the advantages and disadvantages of mapping in neogeography environment are listed.

2.2. Neogeography

The mapping process passed various stages from paper maps to digital maps. Looking at recent evolution in mapping technology, it has been realized that 1980s and 1990s periods were respectively the eras of Desktop GIS software and internet mapping. Recently, the widespread engagement of large number of private citizens in the creation of geographic information is observed [3]. These private citizens are untrained in the field of mapping and their actions are voluntary. The term neogeography has been used to mark this trend. Currently, the term neogeography is not familiar to many people, including people in the field of mapping and other professional disciplines [13][URL 1]. In mapping environment, for the first time, the term neogeography was mentioned by Eisnor (2006), a co-founder of Platial.com [15][URL 2].

Eisnor and Wilson got an idea of creating site for sharing geographic information of the map on the web. The idea was to share hard copy maps on the web in a pined fashion way as shown in Figure 2.1. They thought that this would develop into something with millions of interests contributed and plotted by thousands of users all around the world [15]. In doing this they were not aware that it would develop into the so called mashups or pushpins, though they called it ‘neogeography’. According to Eisnor, neogeography meant the merging of user data and experiences with online mapping technologies. Online mapping using user data has been possible due to development of technology specifically web 2.0. People who do not have a formal qualification in mapping are able to create their own maps. This can be evidenced in the available platforms such as Google Earth, Flickr, OpenStreetMaps etc. A product manager for Google, Taylor (2006) confirms the ability of people to create maps on the available platforms. Taylor said “We provide the map, and other people put in the pushpins” [16]. In addition, the codes for making maps are simple to an extent that even the beginner can make a map. Neogeography has been defined by McFedries (2006) as the practice of combining online maps with data such as blog posts, websites and annotations related to specific places on those maps [17]. According to Turner (2006), "Neogeography is about people using and creating their own maps, on their own terms and by combining elements of an existing toolset" and added that neogeography is fun [7]. In its wider sense neogeography and GIS are distinct things though not mutually exclusive. From this concept, Turner modify the definition of neogeography to mean geographical techniques and tools used for personal activities or for utilization by a non expert group of users; not formal or analytical [18]. This can mean that there are no systematic procedures for data collection and creation of neogeography maps. Neogeography data and map can contain useful information which can be used by any individual not necessarily non expert group of users. This project opt the definition of neogeography given by Turner (2006), as a voluntary process of collecting, displaying, disseminating, editing, sharing and using of geographic data on the platform internet by individual(s) as an influence of technological development.

The volunteered effort of individuals in collecting and disseminating of geographic data has resulted into a massive collection of geographic data and maps on the web. The data collected can suit for a specific application. This makes neogeography to be important in the field of mapping as discussed in the next section. These individuals also describe the earth in terms of text and some of those descriptions may be available in Wikipedia website. In this website an individual can also provide or contribute ideas on a given subject.

Figure 2.1: Web shared maps created by Di-Ann Eisnor and Jason Wilson

2.3. Characteristics of Neogeographers

Neogeographers are those people who combine online maps with data such as blog posts, websites, and annotations related to specific places on those maps [URL 3]. One can think of neogeographers as people who lay information on top of existing maps. Chilton, the chairman of the UK Society of Cartographers noted two characteristics of neogeographers. First, neogeographers have no knowledge of cartographic principles, but more important they don't care about them. Second, neogeographers don't seem to have need for traditional support network. Example of traditional network support is like getting support from the Society of Cartographic [URL 4].

The receiving of information through blogs, forums and social bookmarking services , together with the use of geospatial technology and tools for personal activities without a wider scientific aim can also be considered as a typical characteristics of neogeographers [19].

2.4. Why neogeography mapping

Maps have been the source of geographic information for centuries. A good map is one that communicates information effectively to its audience. The purpose of the map, data and user requirements are the major things to be considered in creating maps that communicate information effectively. Due to advancement in technology specifically in computer hardware, software, and the internet, it is possible for the maps periodically to be current or up to date, real time, dynamic and even interactive. Use of satellite imagery and aerial photographs, can be considered as quickest way of obtaining large coverage data for mapping. However, not all information can be sensed or interpreted on a satellite images or aerial photographs. For example, consider earthquakes, floods, and forest fire disasters. It may happen that, no earth observing satellites pass over the affected area for a number of hours or days. Images from satellites and air craft may be obscured by clouds and smoke, and the condition on the ground may prevent downloading of digital imagery due to lack of power, internet connection or computer hardware and software [3]. Individuals who are neogeographers are intelligent and can report the condition through explaining, voice, mobile phones, pictures, text and audio or video recordings. Availability of this kind of data also called neogeography data may be among the means of reducing the problems stated above. The question is whether professionals and cartographers will accept the neogeography data and maps.

At the individual level, places affected by diseases, war, earthquake and floods are identified and presented on a map. This kind of mapping governed by individuals suit much in early warning and monitoring. Figure 2.2 is Typhoon Ondoy Map of Metro Manila. The map on the left show places of missing persons and the map on the right shows the situation in Metro Manila after the flooding [URL 5]. Information has been provided by an individual as read on pop up display 'Urgent Help Needed'. This kind of map is an example of a typical neogeography map that attempts to fulfil user needs at individual level.

In neogeography environment, the volunteered effort of an individual in collecting geographic data is also considered as the way of obtaining data to the places that do not have data. Example of this is OpenStreetMap coverage. OpenStreetMap is an international effort to create a free source map data through volunteer effort [3]. The following section describes the creation or collection of map data together with its characteristics.

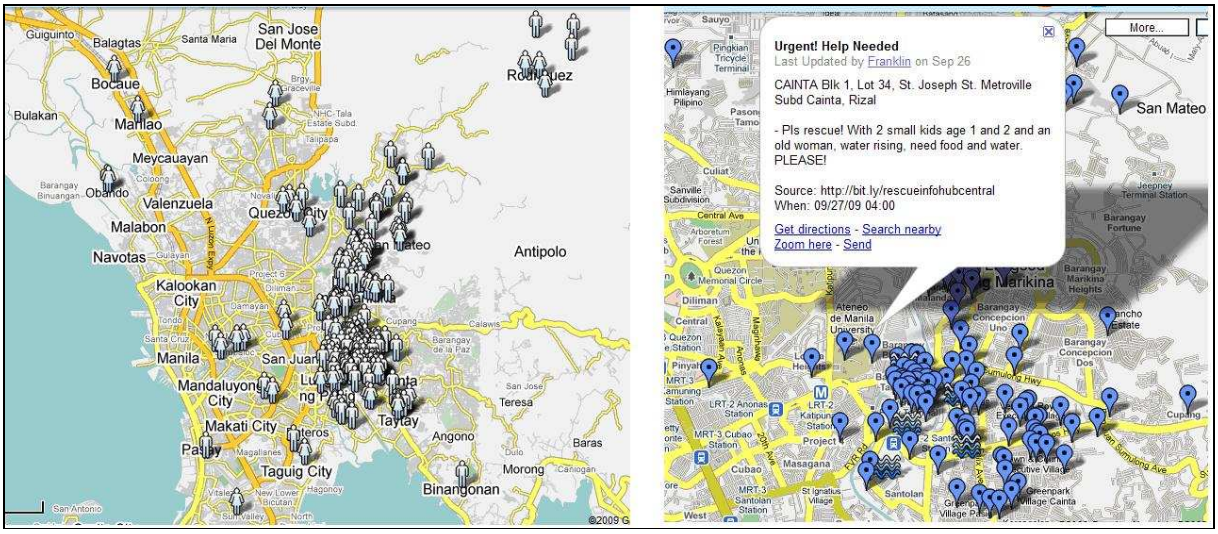


Figure 2.2: Typhoon Ondoy Map – Metro Manila

2.5. Neogeography dataset and its characteristics

2.5.1. Instruments used for data collection

Technological development has made an individual able to collect geographic data using the simple digital devices such as GPS, cell phones and cameras, etc. GPS is the only instrument that can enable direct measurement of geographic location. The instrument is fit for different purposes, the application for which the GPS is to be used dictates the type of GPS, precision, accuracy and hence the cost of the instrument. GPS data can be in GPX format (GPS eXchange format). This data format enables the GPS data to be used between programs or to be shared among users who are not using the same program.

GPS enabled digital cameras and cells phones are considered as instruments for collecting geographic data. This facilitate the individual in geotagging their photographs for sharing. Geotagging is the process of providing, either documents, photographs or other type of data with location information components like latitude and longitude [7]. A photograph acquired by a cell phone without an onboard GPS will need to be geolocated. In geolocating, the position of an object is obtained based on various other data such as cell phone tower. Also an individual can use the coordinates of a respective platform to spatially locate for the photographs acquired by camera or cell phone without a GPS. All of these data acquired by users voluntarily have distinguishable characteristics which are listed in the section below. These data collected by using the instrument mentioned above are then disseminated to the web where they are represented as map mashups or geotags.

2.5.2. Characteristics of neogeography data set

Coote and Rackham (2008), outline four characteristics of Neogeography datasets [20] as follows:

1. Collaborative: Individuals voluntarily capture, process and disseminate the geographic information.

2. Capturing, processing, creation and dissemination of geographic information is not tied to the existing and accepted standards and methodologies.
3. Creation has been stimulated by lack of available data, cost restrictions and limitation of existing conventional data sources.
4. Data is licensed using some open source approach, which allows for users to consume the data without charge provided the original creator is acknowledged and any user can do the same with anything you produce.

2.6. A typical neogeography map

Typical neogeography map refers to the representation of geographic information by new generation where websites are used as platforms. Web 2.0 is an environment that supports neogeography. With web 2.0 an individual can extend or edit contents disseminated by others. A typical neogeography map is characterized by web standards, customizable and interactive mapping (animated, real-time, distributed and dynamic contents).

UCrime is a real-time neogeography map (figure 2.3) that provides easy to read crime maps to the public. In order to get the alert of crime information around ones area through email, an individual is supposed to sign up. It incorporates web 2.0 and allows users to comment on a particular crime.

Figure 2.3: UCrime University of Wisconsin – Madison

Figure 2.4: Map of Namadgi National Park – Australia

Another example of a typical neogeography map is as shown in figure 2.4 above. It is a map of Namadgi National Park – Australia. This Park is estimated to be almost half of Australian Capital Territory [URL 6]. The data file is in KML format which was issued by a Google Earth community member. As it is free to add, edit and share contents in neogeography environment any individual is allowed to add information to it. The symbols and texts are used simultaneously to represent the geographic feature. This enables individuals to easily understand the nature of information represented on a map compared to maps where one can refer to a legend.

In figure 2.4, the Google Earth image has been used as a background where the data has been displayed. This user generated data can also be displayed on a platform or virtual environment where a background can be a map or an image. This process of displaying or combining user generated contents to form a map on already designed map or image is what characterizes the typical neogeography map and has been termed as map mashups.
The map mashups was facilitated by Google through the advent of Google Maps API. This opened the floodgate, allowing the individuals to mash all sort of data onto maps.

2.7. Neogeography tools

Formally cartographers were talking about different kinds of map projections when creating maps. For example, they could talk about Mercator versus Mollweide projections when creating maps for solving problems relating to land, but the map makers in a neogeography environment, who use mapping application programming interface (API), can talk about GML versus KML when geotagging their photos to create maps for summer vacation.[7]. Geotagging photography is also a web 2.0 application that has a spatial frame of reference. Figure 2.5 below show geotags and mashups as components in neogeography environment. API can be used as a means or a tool that enables user/neogeographer to display the generated contents which can also be a geographic data in a neogeography environment.

In order for the neogeographers to be able to create geographic data, there is a range of readily available tools for identifying the coordinates of location on the reality. The tools needed include the instruments for data collection and dissemination such as GPS, cell phones and cameras. These instruments have an inbuilt well defined interoperable system, and this make them not to talk about projections when collecting data that can be used for mapping.

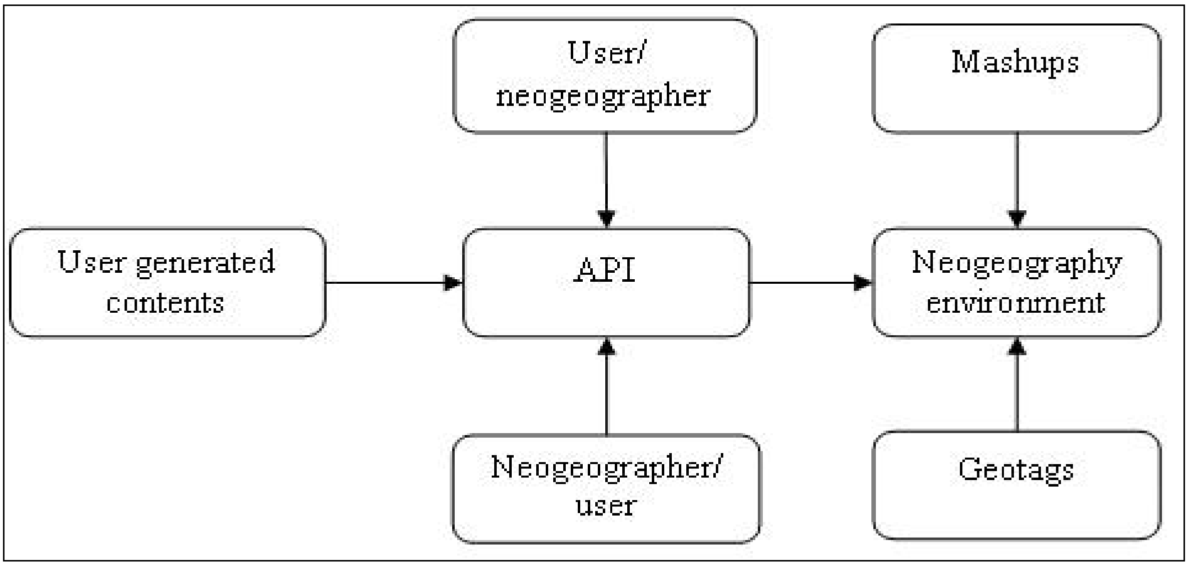


Figure 2.5: Neogeography tools and components

There are no specific tools (at this time of writing) designed for collecting data for neogeography mapping. The available tools have been made for specific purpose, but are also being used in neogeography. This shows that neogeographers experiment with varying tools during mapping. The varying tools also have different data formats that need distinct software or software that can handle all data formats. Due to this there exists a lower level tool that people can use to program their maps on their websites. These include mapmaking and map programming websites. Example of mapmaking websites are www.ning.com and www.platial.com while mapstraction API is an example of map programming website. During mapping, coordinates are very important for locating an object on the reality. The coordinates can be obtained using the instruments equipped with GPS, geolocation or through geocoding of the feature.

There are many tools at software level used in neogeography environment which are supported by web 2.0. Some of these tools include Google Map creator, MapTube, PhotoOverlay Creator, Google Map Image Cutter, Google Earth Creator. Geobase, Map point, Yahoo! Maps, Flickr, YouTube, Wikimapia, Wikipedia, etc. An individual can edit, contribute or share geographic information in text to Wikipedia website. Geotagging a photograph is simple in Flickr though it is possible in some other software. Video and audio data are mapped in YouTube which can also be displayed or mapped in Google Earth.
As observed from above, most of neogeography tools can be obtained from Google products because of its development in technology. This attracts many people in making their maps in Google. Google

community is also providing data which are well stored in Google Earth database. This makes it rich of user generated data. For that reason most of the examples of neogeography maps for this project will be referred to Google Earth or Google Maps. Data to be used for this project will be extracted from Flickr website as is the platform that allows automatic and manual data extraction.

2.8. Advantages of mapping in neogeography environment

1. Most geographic information systems unlike neogeography remain expensive programs and restricted to the use of highly trained specialists.
2. Neogeography is user centric and there exists sufficient up to date data for describing objects [URL 15][URL 7].
3. Sharing of information among members is common in neogeography[7].
4. Neogeography maps lower the cost of information production and dissemination and increases the number of information sources [URL16].
5. Interoperability problem is minimum as some software are underpinned by Open Geospatial Consortium (OGC) standards, i.e., easy to interlink with other information and media and accessibility across platforms [URL 7].
6. Information that requires individual experience can be provided, especially when one wants to understand particular physical environment [4].

2.9. Disadvantages of mapping in neogeography environment

1. The availability of massive data complicates assessing of the credibility of information. This means that it is difficult to check the quality of the data analytically and hence the trust and expertise are used as a measure of credibility [4].
2. Accuracy, meaning and relevance of information are left to the discretion of an individual [4].
3. Most of the information disseminated is made public (No privacy).
4. Sophisticated mapping themes such as soil class, land use, or land cover requires certain level of training and expertise limit participation of neogeographers to this task [3].
5. Neogeography mapping still lacks the ability to do geographic analysis; hence making basic decision difficult.
6. Bandwidth limitation and digital divide. It requires internet with high bandwidth and there is no much information to the areas where access to the internet is limited or internet services cost high like in Africa. [URL 7]
7. Regardless of scale, in many cases the maps display spatial point data. These maps are misleading as an areal feature distribution is summarized or represented as a point at all scales. Few neogeography software (example GMapCreator software) are available that can enable creation of thematic maps.
8. Copyright issues: who initiated the system, who owns the data and the maps? [URL 7].

3. Cartography

3.1. Basic cartographic design principles

Maps should be well designed in order for the user to be able answer question like where, what, how and when. Before designing a map or presenting the data into map format, the cartographer inspects thoroughly the nature of the data. These data are generally collected by third parties such as geodesists, photogrammetrists, geographers and statisticians [21]. The data can be described as qualitative or quantitative. 'Qualitative data is also called nominal data' [22]. When the data is nominal, there is no hierarchy of importance and no arithmetic manipulation can be done on the data. Different soil types such as clay, sand, silt etc. are examples of nominal data.

Quantitative data can be measured and expressed numerically along an interval scale or ratio scale [22]. When the data is measured on an interval scale the distance between values is known but there is no absolute zero. Simple arithmetic manipulation such as + and – can be performed on the data. An example of this is temperature which can be expressed in Fahrenheit or in degrees Celsius. When the data is measured on ratio scale, there is absolute zero point. All arithmetic manipulations such as +,-, /, and x can be performed on the data. Example of this is the number of inhabitants and tons of freight.

Quantitative data can be classified. Data classification methods include natural breaks, quantiles, equal interval, standard deviation method etc. Depending on the nature of the data, one method can be better than another. For example, Natural breaks method among others, can be selected when considering distribution of data along a number line.

Ordinal data is another type of data that can be measured along a relative scale. The data can be ranked, but the differences between values are not important. Example of this is the expression of urban density in terms of high, medium and low density.

After the cartographer inspects the data by considering scales of measurement (nominal, ordinal, interval and ratio), determination of the graphic options to use follows. These are described in the following section.

3.1.1. Graphic variables

Graphic variables are used to represent on maps, the locations and attributes of features existing on the Earth's surface using symbols. These symbols are: point symbols, line symbols and area symbols. Variation in symbol size, shape and colour can be used to represent quantitative and qualitative variation of attribute data. Bertin's six visual variables shown in figure 3.1 can be used in understanding the variation. These visual variables, namely size, value, texture, colour, orientation and shape can be applied to point, line and area symbols [22].

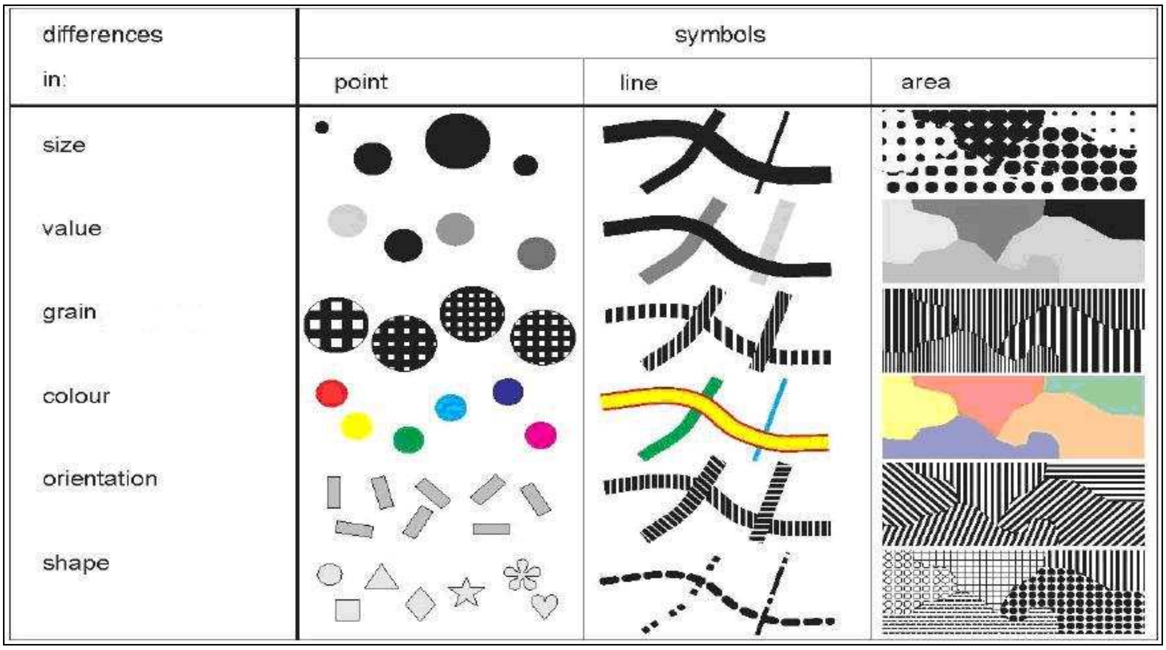


Figure 3.1: Bertin's six visual variables. Source:[22]

The variables shown on figure 3.1 can be used for different application based on scales of measurements of the data. For example the variable size is used for ratio data; value and grain are used for ordinal and interval data, while colour, orientation and shape are used for nominal data [22]. From figure 3.1, we can give the meaning of each of the visual variables as follows:

Size refers to the dimensions of the symbols.

Value refers the gray scale values ranging from white to black.

Grain refers to the variation in density of the graphic element.

Colour refers to hue.

Orientation refers to direction in which symbols are placed.

Shape refers to the shape of the symbol.

Each of the visual variables may have one or more perception properties namely: differentiation (and association), order, distance and proportional perception. Creating maps based on those properties aids the communicability of the maps. These perception properties are explained below.

Differentiation properties

Visual variables have differentiation perception properties if the symbols to which it has been applied are spontaneously perceived as different [23].

Association properties

Visual variables have association perception properties if the symbols to which it has been applied are spontaneously perceived as of equal importance. There is impression of uniformity where no symbol seem to be more important than another [23].

Order properties

Visual variables have order perception properties if spontaneously all symbols differentiated by a given variable can be placed in an unambiguous order [23].

Distance properties

Visual variables have distance perception properties if spontaneously all symbols to which it has been applied can be placed in unambiguous order and an estimate of the distance between the symbols within the range can be made [23].

Proportional properties

Visual variables have proportional perception properties if the differences between the symbols to which it has been applied can be expressed in distinct amounts [23].

3.1.2. Scale

Map scale is the ratio of a unit distance on the map to the corresponding distance on the ground. Maps are referred as being large scale or small scale. Large scale maps show more details than small scale maps. For example, a 1:10,000 paper map can show more details than a 1:50,000 paper map. A 1:10,000 paper map is considered as large scale map while a 1: 50,000 is small scale map. Depending on the scale of the map, data can be collected, conceptualized and treated differently [21]. At a certain scale a building can be treated as an area object while at some other scales as a point. Scale of the map enables determination of distances between two features, area or size of the mapped surfaces or objects. Scale can also dictate the accuracy levels associated with the data [URL 8].

In digital spatial data, the term resolution is used. Resolution is the mean dimension a cell covers on the object. Digital spatial data is stored without scale in a database. “Scale is a ratio notion associated with visual output, like a map or on screen display, not with the data that was used to produce the map” [22].

3.1.3. Symbolization

Symbolization is the process of creating graphic symbols to represent feature attribute values [24]. Geographic features such as point, line and area can be symbolized differently. Qualitative and quantitative point data can also be represented by point symbols. These point symbols can be either purely geometric or pictorial. Since both kinds of point representation have advantages and disadvantages, the cartographer is always careful when applying those symbols to a map. Symbols representing a quantitative data on the map can give a visual measure of its size, importance or number. This is different from qualitative where the symbols are of equal importance.

3.1.4. Typography and lettering the map

Text can express qualitative, quantitative, ordered and selective perceptual properties. This contributes to the identification of map features, legibility and understanding of maps. Text can be considered as the easiest way to show the geographic feature. Applying text to a map can be a challenge as it involves the choice of elements to be labelled, label design and placement. However, labelling can be automated, but still expertise and human intervention is needed. This is because text can contain other information like dictating the direction say of a river, distinction between a Ward or Postcode and Municipality etc. Based on their relevance, text on maps can contain primary information when applied within the mapped area, secondary and tertiary information when applied beyond the mapped area [25].

3.2. Web cartography

The evolution in cartography from paper to digital cartography is due to advancement in technology especially hardware and software. Furthermore, the networking ability between the digital devices through a World Wide Web (W3 also termed web) leads to another evolution in cartography called the web cartography. In web cartography the cartographic functions such as design, production and use of maps remain the same, but the difference is the medium used. Web cartography uses W3 as medium for the dissemination of geographic information. The advantages of disseminating the geographic information on W3 as outlined by Kraak (2001), Köben and Kraak (1999) are that the information on the web is virtually platform independent, unrivalled in its capacity to reach many users at minimum costs and easy to update frequently. Furthermore it allows for a dynamic and interactive dissemination of spatial data offering new techniques and use possibilities not seen before with the traditional printed maps [26-27]. Interactivity is very important as it gives the users the ability to pan, zoom, and scale the map. Furthermore it allows the user to access the spatial database for data querying.

3.3. The 'gap' between neogeography and cartography

When looking at neogeography and cartography we can realize that in neogeography, individuals are not strictly bound by rules when creating maps compared to cartography. For example, when selecting symbols to map the geographic information, the following three steps should be performed: determination of the nature of the data, identification of perception property and identification of suitable visual variables. Following the steps above will ensure the correspondence of the user's and cartographer's perception. Failure to follow the three steps could be among the factors contributing to the 'gap' between the neogeography and cartography. For example when one uses symbols of different sizes on a map, the map user can get the feeling of the feature represented. One can be able to identify that one object is bigger than another, more or less. The use of varying symbol sizes can stand in place of colour. Applying symbols properly on a map also called symbology can take into account the design qualities of a map such as legibility, figure-ground relationship and hierarchical organization.

The nature of neogeography data is its existence in different location being the result of collection by different individuals. It is difficult to consider the three steps mentioned above when using Neogeography data. This reveals that in neogeography, data analysis is not performed. Individuals are displaying their collected data to the respective web platforms without knowing whether the data are qualitative or quantitative. They are not aware to which measurement levels the data belong, whether nominal, ordinal, interval or ratio. An understanding of this will help them to represent their data on the map in a meaningful way. In order neogeography data to be used for different applications at small scale, they require to be summarized. It is hard for the neogeography software to perform that task.

The valuable data and/or information collected by the neogeographers, together with technology supporting the mapping trend in neogeography environment constitute the positive aspects of neogeography. On the other hand, some processes like the cartographic data analysis, symbol selection and cartographic text can be applied to neogeography to aid the communicability of the map. This also constitutes to the positive aspects of cartography.

Positive aspects in neogeography and cartography can be combined to minimize the gap between the neogeography and cartography. As cartography is based on rules, while neogeography is not, it is difficult to link directly cartography and neogeography. An intermediate step between neogeography and cartography, called 'summary map' is introduced as shown in figure 3.2 below. Summary map links the positive aspects from either neogeography and/or cartography, in design perspective. This is by using the spatial data obtained from neogeography environment and summarization and visualization techniques from cartography.

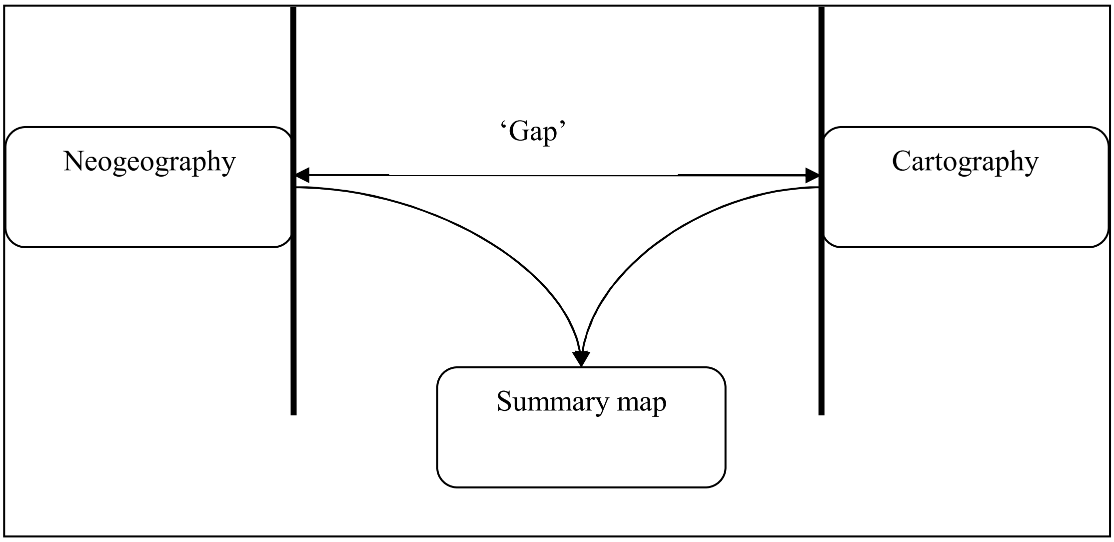


Figure 3.2: Summary map as a means of reducing a gap between Cartography and Neogeography

The 'gap' between the neogeography and cartography can be identified by observing the product (map) of neogeography and cartography. Thorough observing the two products one can find out what is wrong, happening or changing. In organizational or social fields, indicators are used for performance or quality outcome measuring [URL 14]. In this project the indicators will be used to compare the neogeography maps and cartography maps. This enables us to know the 'good' map and 'bad' map, and hence clear identification of the 'gap' between the neogeography and cartography maps. When reading the map, scale and symbols used have impact in understanding the map. For that reason, scale and symbology can be selected to compare neogeography, cartography and the summarized maps. However there exist many other indicators including text. The effect of scale and symbology as an indicator of a good map are explained in detail in chapter 6 based on the usability test results. In this project, the indicators will be referred to as the 'indicators of a good map'.

3.3.1. Scale as an indicator of a good map.

In neogeography environment the data are collected by individual using devices like GPS and digital camera. At this stage the collected data are considered to be at large scale. These data are displayed on neogeography platforms such as Google earth, Google maps, etc. as points where a scale is a notion associated with visual output of the screen display. Geographic feature can be represented as a point at certain scale while as an area at some other scales. For example, the built up area represented as point can be changed to an area if the scale of representation would increase. Similarly the built up area can be changed to a point if the scale of representation would decrease [21]. Scale determines the amount of details to be shown on the map. The larger the scale the more feature details to be shown on the map. The smaller the scale the less features details to be on the map [24]. Poor selection of scale to represent geographic features may cause a map to appear overcrowded.

In neogeography, any geographic feature details can be shown. The question is at what zoom level can an individual seem to represent geographic feature at lager scale? At what zoom level can an individual have the overview of the map? During detail viewing, can an individual observe the relationship between the geographic features or loses the spatial relationship? Answering the question above, it is good idea to design neogeography platform for visualization according to Shneiderman's visual information seeking mantra [28]. For the neogeography map, an overview can be at small scale. Zoom in for detail viewing to answer the question what, where and when may depend on users' need. Most of the neogeography maps used in this project are displayed at a scale of about 1:6000000 in Google Earth and also applies to maps created using ArcGIS. At this scale the maps are considered to be at small scale.

3.3.2. Symbology as an indicator of a good map

In order to select symbols for representing the geographic feature, among the factors to consider is the nature of the data. The data can be represented as qualitative or quantitative. Table 3.1 below shows the visual variables used for qualitative and quantitative data representation. This helps to realize the way data can be represented on a given map. The shaded cells represent the visual variables to use during data representation. For example, given the quantitative data, the visual variable to use can be the size, value or grain/texture.

Visual variables	**Data representation**	
	Qualitative	**Quantitative**
Size		
Value		
Grain/texture		
Colour		
Orientation		
Shape/form		

Table 3.1: Visual variables to consider during data representation

When the visual variables, perception properties and measurement level are considered during symbology, the correspondence of user's and cartographer's perception can be achieved. Table 3.2 shows the link between visual variable, perception properties, and measurement level.

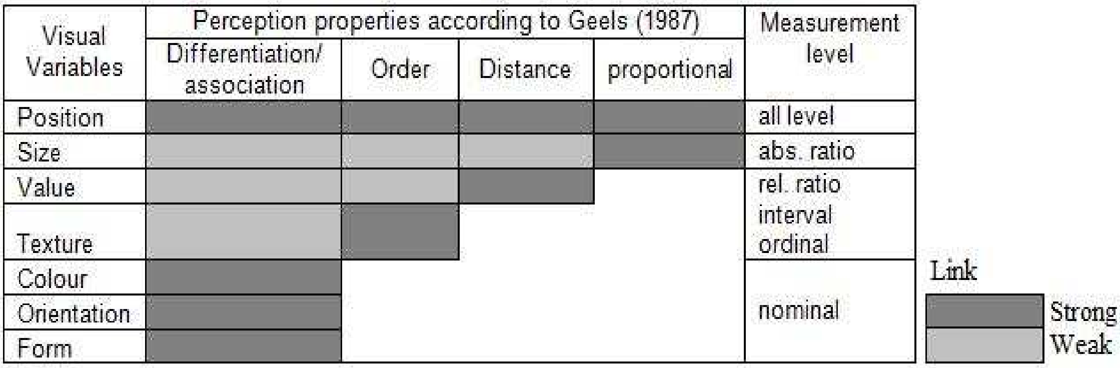


Table 3.2: Link between perception properties, visual variables and measurement level **Source:[23]**

3.3.3. Cartographic text as an indicator of a good map

The cartographic text is used for naming features on the map. It can also be used to reflect the character of geographic feature and confirm the location of places [29]. Size of the text need to be considered and is not supposed to be either too small or too large to be seen. Though there exists automatic labeling algorithm, human intervention is also needed as one can consider the scale, size of geographic object and the available space.

4. Summary Map

4.1. Introduction

Most neogeography maps only use point features. The neogeography map can represent single topic or multiple topics. At a given scale, those maps seem to be cluttered while at some other scale they are not cluttered as a result of zooming in. However, when applying zoom to neogeography maps only the level of detail of the background changes while the size of the point features remain the same. When one clicks on the neogeography symbol, an information window pops up describing the geographic feature. The above are some of criteria that describe the neogeography maps. In order to solve the problem of cluttering, the maps need to be summarized. Figure 4.1 shows how to create a summary map. Figure 4.1(a) is an original neogeography topic at small scale. It is a map of Metro Manila Philippines showing donation centres for the people who are affected by flooding [URL 5]. At this scale, the donation centres are represented by points. One can draw an ellipse as shown in Figure 4.1 (b) and 4.1 (c) to group or summarize those centres as an estimated area having the donation centres. All the donation centres enclosed by the ellipse in figure 4.1 (b) and 4.1 (c) can be converted to polygon as shown in figure 4.1 (d). Figure 4.1 (d) is the summary map obtained after this procedure. At small scale, the summary map can be considered as a simplification of the cluttered neogeography topic from points to polygon.

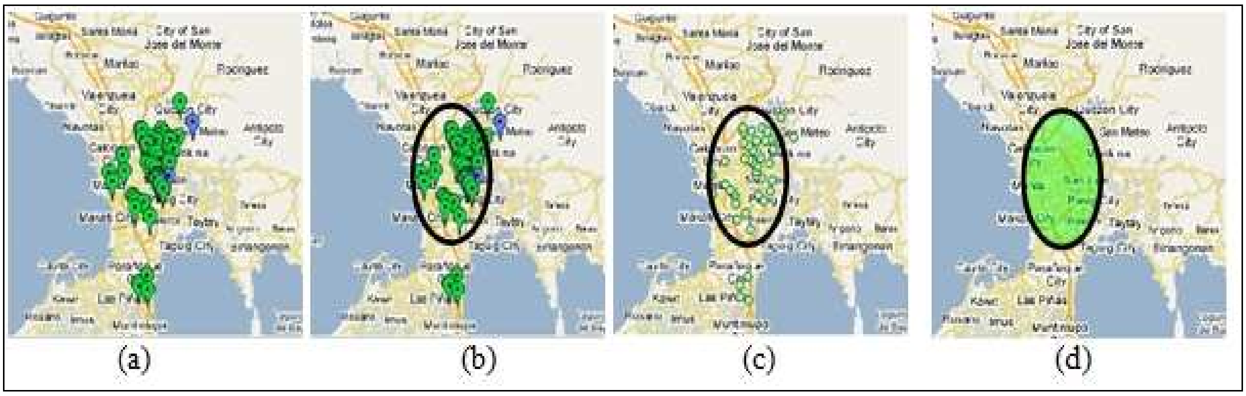


Figure 4.1: Creation of summary map from neogeography data

Therefore, at small scale summary map represents coverage while neogeography map represents points. Neogeography data are increasing rapidly as can be seen on the internet platforms such as Flickr, Google Earth etc. Applying cartographic approaches to the collected massive neogeography data to design a good neogeography map cost time and money. The idea of summarizing the neogeography topics to reduce cluttering can be simple, quick and less expensive. To implement the idea of summarizing the neogeography topics in this research, the following four steps were followed to summarize neogeography maps.

The steps are:

- o Extraction of neogeography data
- o Clustering of neogeography data
- o Symbolization
- o Visualization

4.2. Extraction of neogeography data

Extracting neogeography data from some of the neogeography platforms can be made automatically or manually. Extracting neogeography data automatically is preferred because the amount of data is very large, otherwise it will cost a lot of time to extract the data. Flickr website is among the neogeography platforms that allows both automatic and manual data extraction. This is makes Flickr an important data source for research [URL 8]. For automatic data extraction there is a python code for downloading the Flickr data. In order to run it, it only requires a Flickr key and a Flickr secret number. Both the key and secrete number can be obtained from Flickr. Running the code, it generate text file containing all the information requested. To be able to extract the information from the text file or to arrange them in a format that is familiar to many people such as in excel format, a python code is created. This code extract the data into excel format and is shown in Appendix D. Extracting the data using manual means, some steps need to be fulfilled. The way to fulfil the steps is explained in http://www.flickr.com/services/api/flickr.photos.search.html and a template for calling a method is obtained in API explorer [URL 9]. Defining the bounding box is required in order to get the data extraction coverage, and in order to get the latitude and longitude, the extras field in API explorer should be field with the word ‘geo’. It is important to understand the whole procedure for manual data extraction from Flickr as some parameters explained in it are required to be filled in the code for automatic data extraction. During automatic data extraction the bounding box values used are shown in table 4.1 below. These values were typed in the python code as whole numbers instead of decimal point numbers.

Position of bounding box	Longitude	Latitude
Lower left	3^{o} 07’ 48.60”E (3)	47^{o} 56’ 34.77”N (48)
Upper right	14^{o} 07’ 17.65” E (14)	54^{o} 54’ 58.13” N (55)

Table 4.1: Approximate coordinate used to define the bounding box

Other parameters such as minimum upload date, maximum upload date, text, page etc. are also defined within the code. This automatic data extraction made a massive data to be downloaded from Flickr website. This data could be used for various applications.

Figure 4.2 shows the neogeography map as a result of combining four neogeography topics: forests, hotels, restaurants and schools. These topics have been chosen by thinking that many people may have interest on them. For example, a tourist who looks for the forest can also look for a place to get restaurant and hotel. Also a student can look for a school, restaurant and hotel. The data for the four neogeography topics mentioned above were extracted from Flickr website (http://www.flickr.com/services/api/) and the bounding box for the search was as shown in table 4.1.

Using this bounding box values the coverage is mainly The Netherlands, Belgium and Germany. This data is displayed in Google Earth at a scale of 1:6000000.

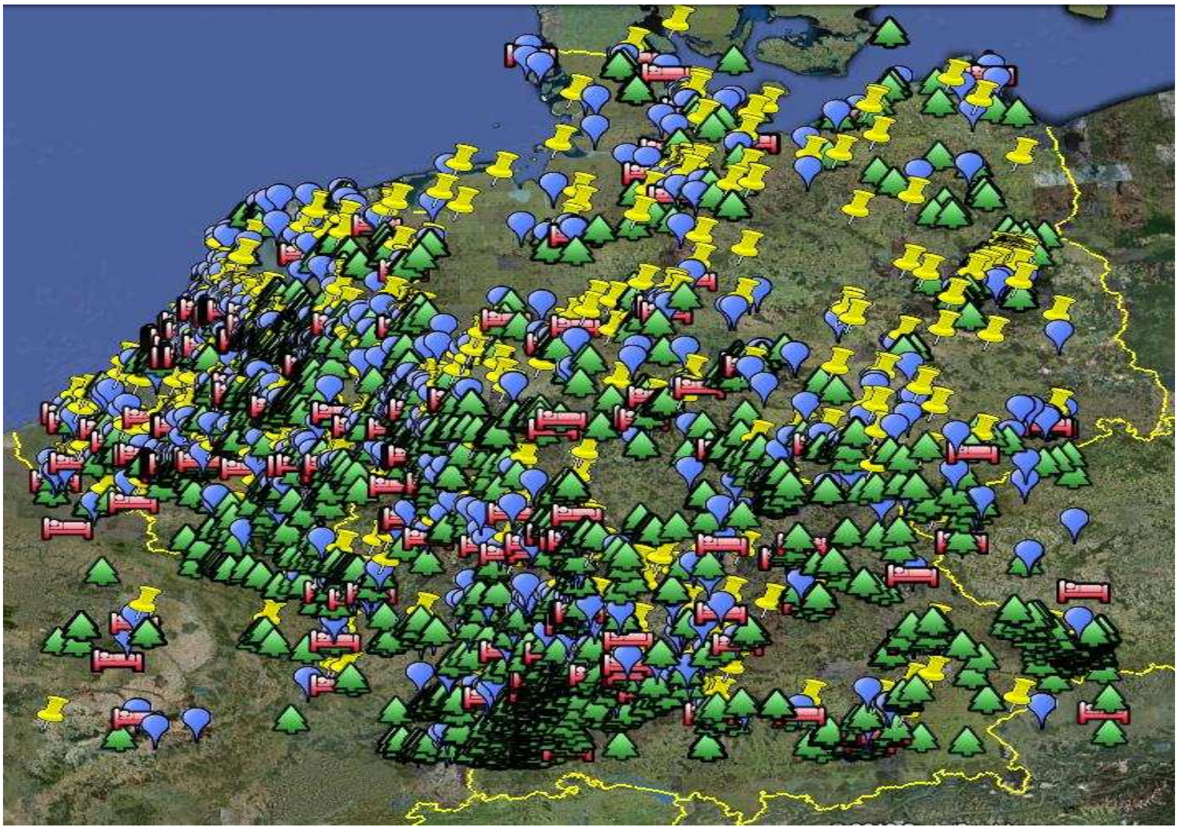


Figure 4.2: Neogeography map with four topics extracted from Flickr (scale 1:6000000)

At this scale the map is cluttered and is difficult to read. Looking on the figure above a pattern can be observed in North West, West and South West of the Netherlands and Germany. As the neogeography data are collected at very large scale, at this scale the maps are too cluttered to such an extent that they can be regarded as area features rather than point features. A cartographer who is expert in generalizing, may summarize the cluttered objects as one object as shown in figure 4.1. Summarizing the neogeography topics above through investigating the patterns observed can be considered as a means of reducing the cluttering. The next section describes some of the methods that can be used to summarize neogeography topics.

4.3. Clustering of neogeography data

Neogeography maps are made from data collected by any individual as described in section 2.5 and 2.6. This made data preparation stage difficult to be performed, yet be omitted. Data preparation stage is important as may require data set to be thinned and retaining the required features only. “The reason may be not all features are relevant for subsequent analysis or subsequent map production” [22]. Other neogeography data are from multiple sources and those too need be combined. These data may be of the same area but of different accuracy. Also data may be of the same area but represented differently. Data of adjacent areas may also need be combined to a single data set [22]. This project will not concentrate on combining data as a result of varying accuracy, but will look on how to combine or summarize neogeography data into a single data set. It tries to convert a cluster of points (neogeography data points) to a polygon by applying clustering techniques and other geometry processing functions that support spatial analysis. The processing functions to be used for this project may include buffer, length or distance, union and bounding containers. Examples of bounding containers are the extent, rectangle, circle, convex hull, ellipse and 2D alpha shape.

Clustering has been defined by Zalik (2008) as a "process of grouping data objects into disjointed clusters so that the data in each cluster are similar, yet different to the others " [30]. Clustering techniques are important as they can be applied to data analyses, pattern recognition and image processing. In grouping data objects, different types of clustering exist. These types of clustering are classified as Exclusive clustering, Overlapping clustering, Hierarchical Clustering and Probabilistic Clustering [URL 8]. The examples of clustering algorithms which are associated to each type of clustering mentioned above include: K – means clustering algorithm which is an example of exclusive clustering. Fuzzy C – means clustering algorithm which is an example of Overlapping clustering. Hierarchical clustering algorithm is an example of hierarchical clustering. Mixture of Gaussians clustering algorithm is also an example of Probabilistic clustering [URL 10].

4.3.1. K – means clustering

This algorithm characteristically determines all the clusters at once. This is through partitioning the input dataset into k clusters. The clusters are represented by adaptively changing cluster center [30]. By relying on some initial values, using random sampling, the k – means assigns each point to the cluster whose center is nearest. The center of the cluster is obtained by averaging all the points in the cluster. The main advantages of this clustering algorithm are its simplicity and speed which allows to run on large datasets [URL 11]. The main limitation of k – means algorithm is that the number of cluster must be predetermined and fixed. Selecting suitable number of clusters is still a challenge as it requires prior knowledge about the data. It may also require guessing the number of clusters [30]. This method is not suitable to discover clusters with non convex shapes [URL 12]. Pycluster is an open source clustering software that can be used to perform both hierarchical and k – means clustering [URL 13]. Pycluster is an extension module to Python and hence requires Python program in order to execute it. Using this software to perform k – means clustering, it requires the number of clusters as an input from the user. It requires also the data to be in a list form. Performing clustering using pycluster software for a large data set, cost time. The list of clusters generated from the pycluster can be visualized using GIS software such as ArcView. These clusters can be converted to polygons using geometry processing tools that are available in GIS software.

4.3.2. Fuzzy C – means clustering

Fuzzy c – means is a method of clustering which allows one piece of data to belong to two or more clusters. It is frequently used for pattern recognition. In this type of clustering, each point has a degree of belonging to clusters as in fuzzy logic rather than belonging completely to just one cluster [URL 10]. This method has become the most well known and powerful in cluster analysis. The limitation of this method is that it gives inaccurate results with a large number of different sample sized clusters especially when signal- to - noise ratio is low [31]. In clustering neogeography data it also requires one data to belong to one cluster only. This method is not suitable for clustering neogeography data when the crisp boundary is required.

4.3.3. Hierarchical clustering

Hierarchical clustering creates a hierarchy of clusters which may be represented in a tree structure called a dendrogram. The algorithms for hierarchical clustering are either agglomerative or divisive [URL 10]. Agglomerative clustering fuse n objects into groups while divisive clustering split n objects successively into finer groups [URL 11]. Hierarchical procedures that use subdivision are not practical unless the numbers of possible splittings are somehow be restricted. In agglomerative hierarchical clustering the number of stages is bounded by the number of groups in the initial partition [32]. Agglomerative clustering techniques partition the data by defining distances or similarities between clusters. This can be done using single linkage clustering, complete linkage clustering or average linkage clustering [URL 11]. 'The drawback of agglomerative method is that those that are practical in terms of time efficiency require memory usage proportional to the square of number of groups in the initial partition'[32]. Pycluster can also be used to perform hierarchical clustering. The nature of neogeography data as a points requires the spatial components Latitude and Longitude or Y and X to be described. It is difficult to take this in hierarchy. Performing hierarchical clustering using pycluster requires a single number and not a set or pair of coordinates. This makes hierarchical clustering not suitable for clustering the neogeography data.

4.3.4. Mixture of Gaussians clustering

This is a model based approach way of dealing with clustering problem. It considers clusters as Gaussian distribution centered on their barycentres [URL 8]. The advantages of this method are its component distributions have high peaks and the mixture model cover the data well. Other advantages are: a soft classification can be obtained, density estimation for each cluster, flexibility in choosing the component distributions and the statistical inference techniques used are well studied and are available [URL 13]. This algorithm can be used for clustering neogeography data when one has the notion of obtaining the density of each cluster in order to create isolines for representing the data at various levels.

4.3.5. Density based clustering

This clustering technique is based on density (local cluster criterion) such as density connected points. The main advantages of this are its capacity to discover cluster of an arbitrary shape and ability to handle noise. Density Based Spatial Clustering of Application with Noise (DBSCAN) and Ordering Points To Identify the Cluster Structure (OPTICS) are examples of density based clustering algorithm. DBSCAN is efficient for large spatial databases [33-34]. The summarized neogeography topics should maintain the original shape of the data at a given scale. DBSCAN can fit well in clustering the neogeography data due to its capability of handling large spatial databases and discovering the arbitrary shape of the cluster. OPTICS is similar to DBSCAN, but for a data of varying density, OPTICS can perform better than DBSCAN in detecting a meaningful cluster. Both methods can be used to cluster neogeography data. DBSCAN has been selected as it is fast in clustering the data compared to OPTICS.

4.4. Symbolization and Visualization

Bounding containers as one of geometry processing function do not provide an arbitrary polygon shape depending on the cluster of points. It outputs the polygons either in the form of extent, rectangle, circle or convex hull [6]. This will limit our symbolization processes based on shape. By exploring the limited shape of polygons and proper application of colours we can represents the multiple neogeography topics as a summarized map.

5. Implementation: conceptual framework and prototype

5.1. Introduction

From chapter four different clustering methods and bounding containers for clustered points are described. This chapter select and implement on a case study data, two methods namely k – means and DBSCAN together with four bounding containers in summarizing the neogeography map. The criteria used for selecting the methods based on its capability to handle large neogeography data set, speed in clustering, and ability to discover arbitrary shapes of the clusters. Bounding containers used are extent tool, rectangle tool, circle tool and convex hull tool.

5.2. Case study

Internet accessibility, cost and bandwidth capacity determine neogeography data availability. Places that lack access to internet or where internet services are expensive and the influence of bandwidth limitations make some places to have few neogeography data as is the case in Africa while areas with access to cheap and readily available internet services have massive neogeography data, for instance in Europe. This is why a study area that includes England, The Netherlands, Belgium and Germany was chosen. The four countries were chosen because they are among countries with massive geotagged data both for the Flickr platform and Google Earth. This is evidenced when the massive data for approximately 30 months were downloaded automatically from the Flickr. The downloaded data are for the supermarkets, playgrounds, airports, traffic lights, forests, hotels, restaurants and schools. The case study area was then divided into two cases. Case one deals only with England map with neogeography topics supermarkets, playgrounds, airports and traffic lights. The map made from these neogeography topics was mainly to test the way people understand the neogeography maps. Case two comprise of Netherlands, Belgium and German with the neogeography topics forests, hotels, restaurants and schools. The map made from these topics was mainly to test the new approach of summarizing the neogeography maps as the way of reducing the cluttering problem. Figure 5.1 is a work flow for prototype implementation showing all the steps from data collection to visualization for the two cases mentioned above.

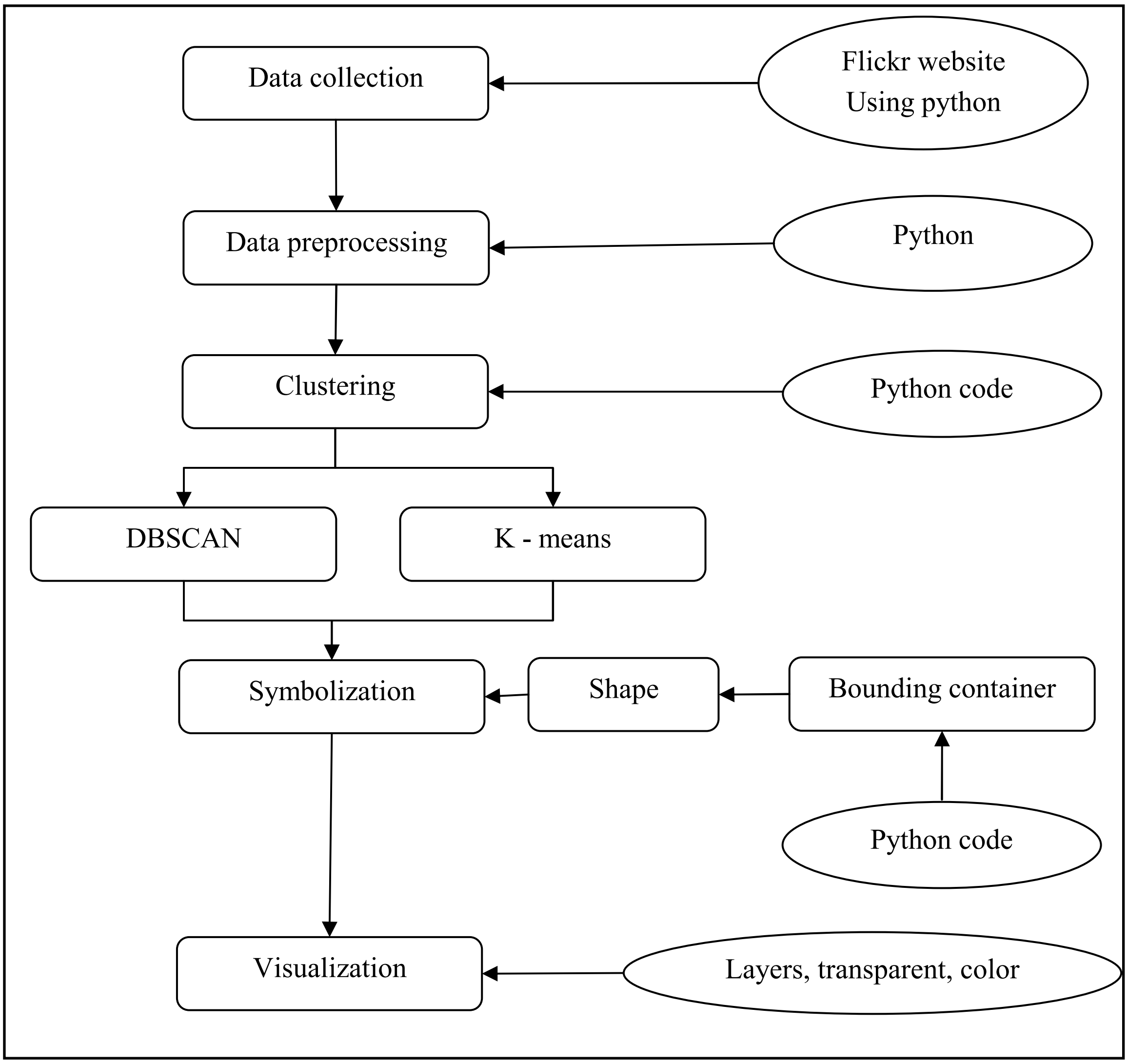
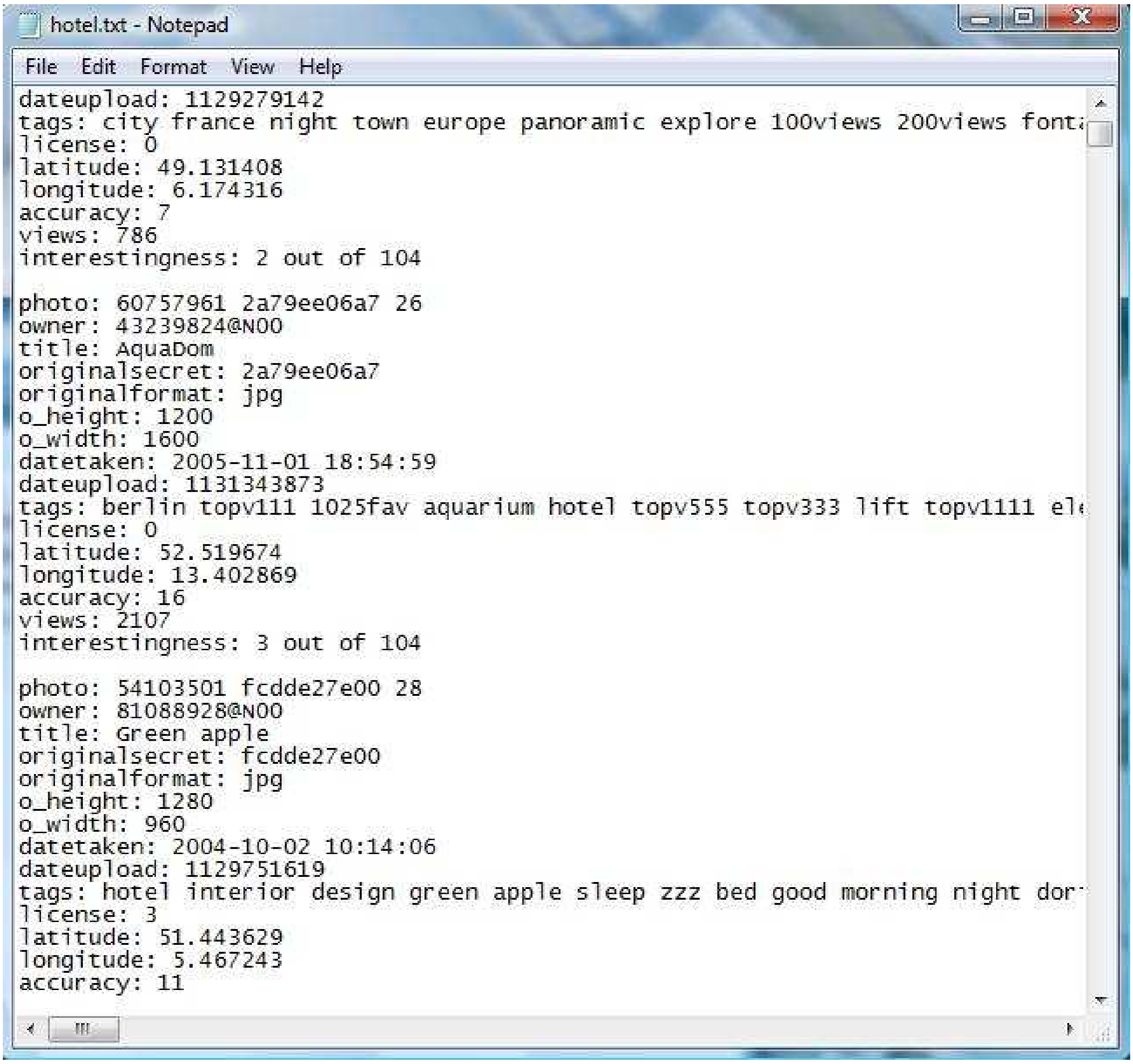


Figure 5.1: Work flow for prototype implementation

5.3. Data collection and pre-processing

Data used in this project were downloaded from Flickr website using python code as explained in section 4.2. The extracted data is in text format with the following information and arrangements as shown in figure 5.2 below. Since neogeography data is huge, the extracted data was only from 12/10/2005 to 12/06/2008. The processing of it includes writing a python code that reads the text file and out put the excel file which can be converted to dbase format or other format. Converting data to dbase format, enables observing the attribute information in the table when displaying the data in GIS software. The python code is modified to extracts the latitude and longitude only from the text file and convert it to a list which can then be used as an input data during clustering using either the k – means or DBSCAN clustering methods.



hotel.txt - Notepad

File Edit Format View Help

```
dateupload: 1129279142
tags: city france night town europe panoramic explore 100views 200views font
license: 0
latitude: 49.131408
longitude: 6.174316
accuracy: 7
views: 786
interestingness: 2 out of 104

photo: 60757961 2a79ee06a7 26
owner: 43239824@N00
title: AquaDom
originalsecret: 2a79ee06a7
originalformat: jpg
o_height: 1200
o_width: 1600
datetaken: 2005-11-01 18:54:59
dateupload: 1131343873
tags: berlin topv111 1025fav aquarium hotel topv555 topv333 lift topv1111 el
license: 0
latitude: 52.519674
longitude: 13.402869
accuracy: 16
views: 2107
interestingness: 3 out of 104

photo: 54103501 fcdde27e00 28
owner: 81088928@N00
title: Green apple
originalsecret: fcdde27e00
originalformat: jpg
o_height: 1280
o_width: 960
datetaken: 2004-10-02 10:14:06
dateupload: 1129751619
tags: hotel interior design green apple sleep zzz bed good morning night dor
license: 3
latitude: 51.443629
longitude: 5.467243
accuracy: 11
```

Figure 5.2: Data format and arrangement as extracted from Flickr website

Further processing of the data is also required as there are data containing 0 values for latitude and longitude. Data with 0 values for latitude and longitude are deleted as the study area for case two mentioned in section 5.2 do not contain 0 values for latitude and longitude. For the case one study area data may have the 0 values for the longitude, this makes data with 0 values to be considered carefully.

5.4. Clustering

After pre processing of the collected data, the data can be clustered using the k – means or DBSCAN methods. Figure 5.3 shows the conceptual framework for clustering, summarizing and visualization of neogeography maps.

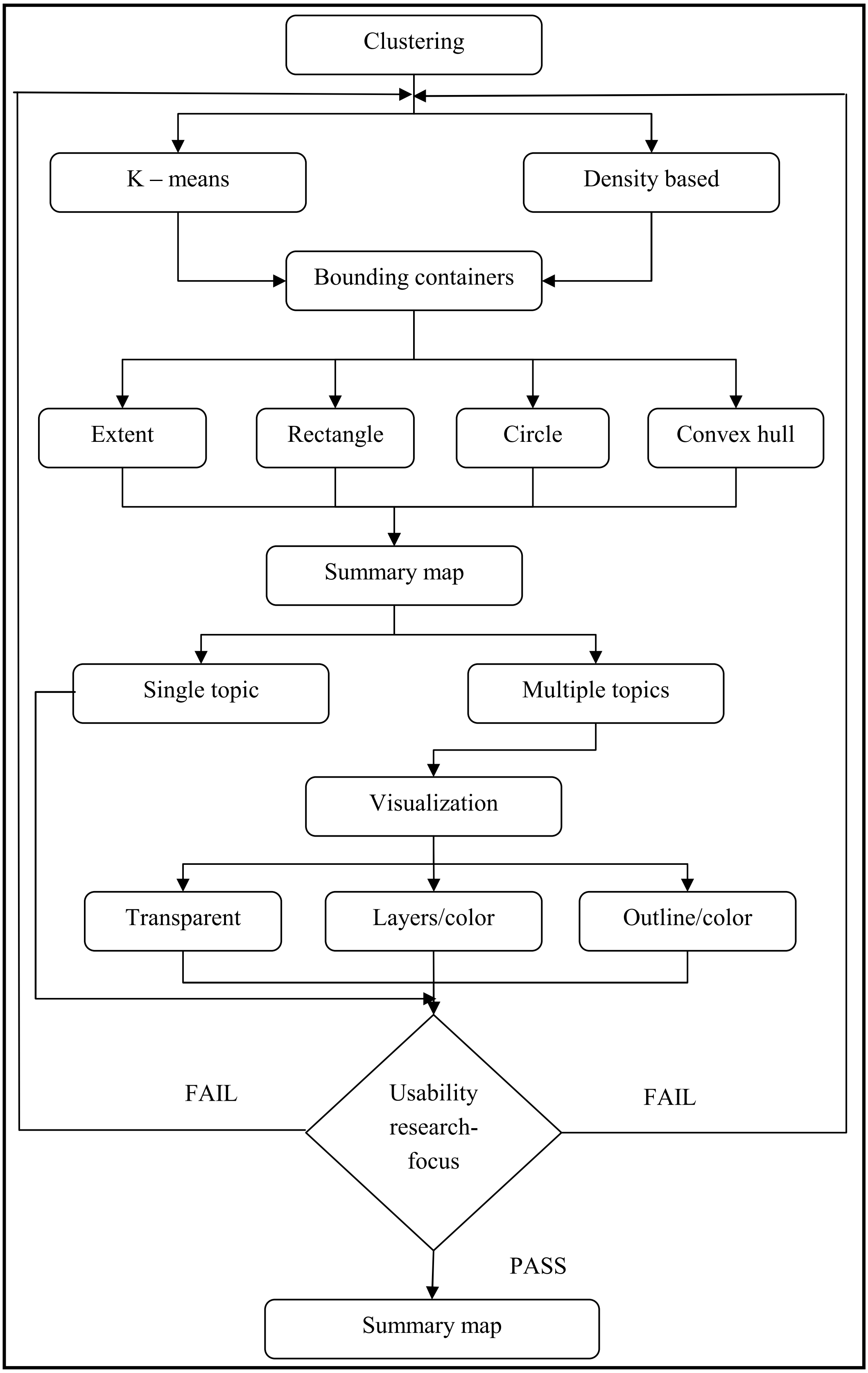


Figure 5.3: Conceptual framework for prototype implementation

Using four neogeography topics namely hotel, forest, school and restaurant, the first step in summarizing the map was to use the bounding containers for each neogeography topic. The use of bounding containers for bounding the clusters of each neogeography topic helps to get the visual impression of how the original data points relate with the clustered points. This also enables us to understand how well the algorithms cluster the data points into clusters. This is due to the fact that the performance of the clustering algorithms depends also on the distribution of the data. Depending on user needs and purpose of the map and observing the clustering algorithm that gives a good result, we can combine together all the topics to get the summarized neogeography map. Applying the cartographic cosmetics (aesthetics) to the summarized neogeography map will result to the so called the "Summary map". Figure 5.4 shows the processing steps of all four neogeography topics.

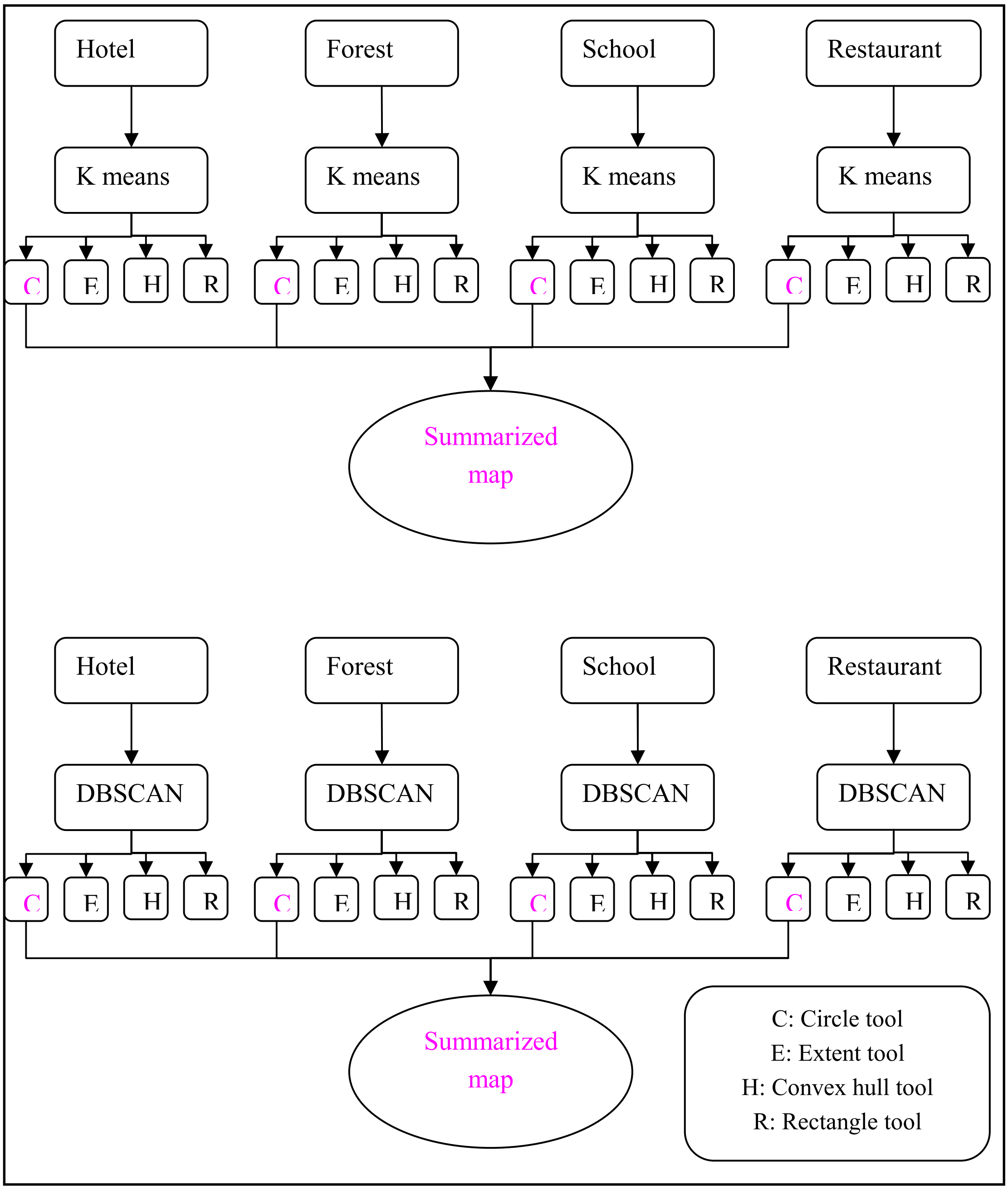


Figure 5.4: Steps for creating summary map from various bounding containers and 2 clustering methods

There exist various methods of clustering neogeography maps as explained in chapter four. This project uses two methods mentioned in section 4.3.1 and in section 4.3.5, i.e. k– means and DBSCAN.

After clustering the data, the points belonging to each cluster are converted to polygon using the bounding containers described as C, E, H, and R shown in figure 5.4. The combining of the corresponding bounding containers represented using letters will result into several maps that summarize the neogeography topics. With 4 neogeography topics and using four bounding containers, this results in 16 summarized neogeography topics which when combined, 4 different summarized maps are observed.

To show that the neogeography multiple topics can be summarized, this project took forest as an example in figures 5.5, 5.6, 5.7, 5.8 and summarizes it using k – means algorithm where the number of cluster were chosen as eight. Figure 5.9 is also a summarized neogeography multiple topics using k – means algorithm. The original neogeography maps which have been summarized using k – means and DBSCAN are shown in Appendix A, figure A_a, A_b, A_c and A_d. In observing the capability of this clustering algorithm in clustering the neogeography topics, the white background has been used for all the figures in this section. This may be contrary to neogeography maps where the satellite images are used as the background. The DBSCAN clustering algorithm also applied to the neogeography topic (forest) as shown in appendix A by figure A_S1_a, figure A_S2_a, figure A_S3_a and figure A_S4_a. For the DBSCAN algorithm to perform better two parameters are required namely, the number of objects in a neighborhood of an object (k) and the neighborhood radius (Eps). The number of objects in a neighborhood of an object, k=7 and the neighborhood radius, Eps = 0.2 were used in this project as they seem to give good results for this case study data. Different values for the parameters k and Eps can be changed directly on the python code. With the use of interface, these parameters can be changed by using a slider bar as shown in figure 5.10. By considering figure 5.4, all the maps in appendix A were created using the DBSCAN algorithm.

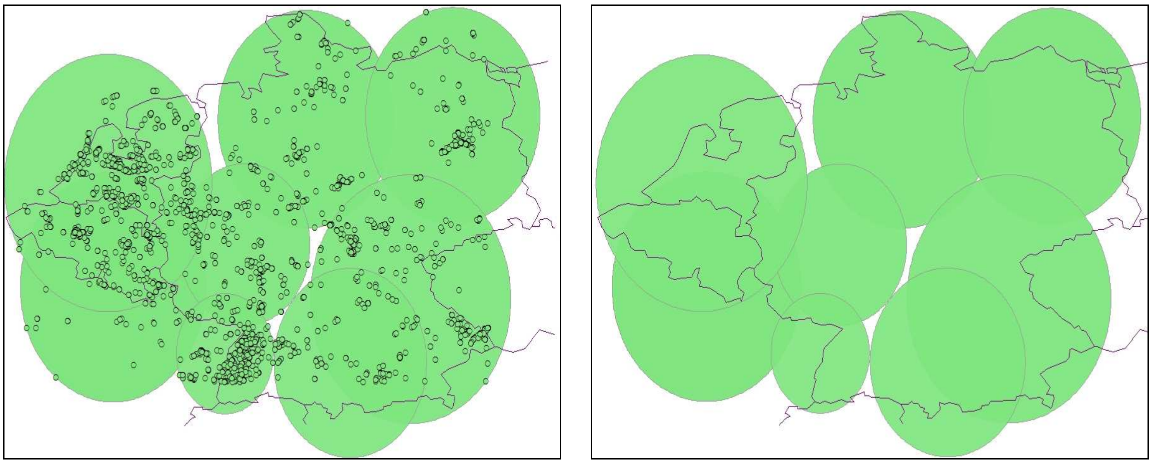


Figure 5.5: Forest data points converted to polygon using circle tool

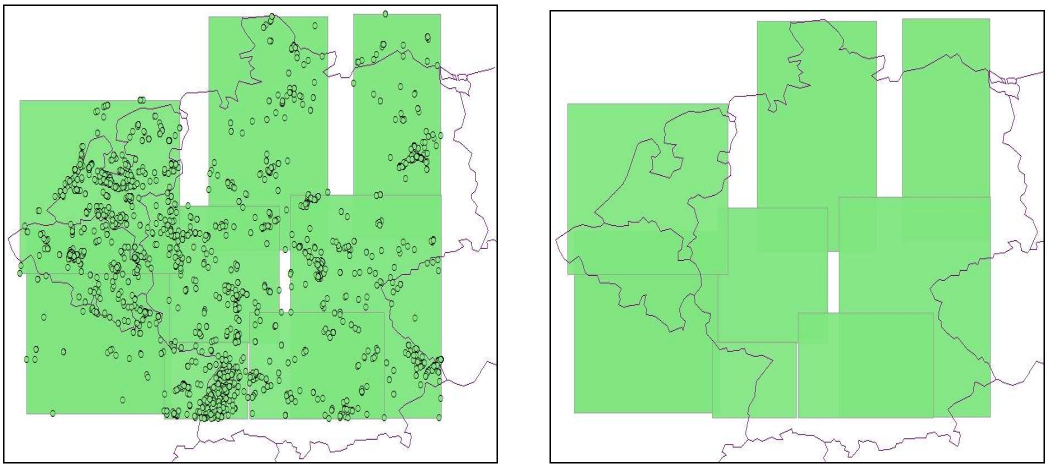


Figure 5.6: Forest data points converted to polygon using extent tool

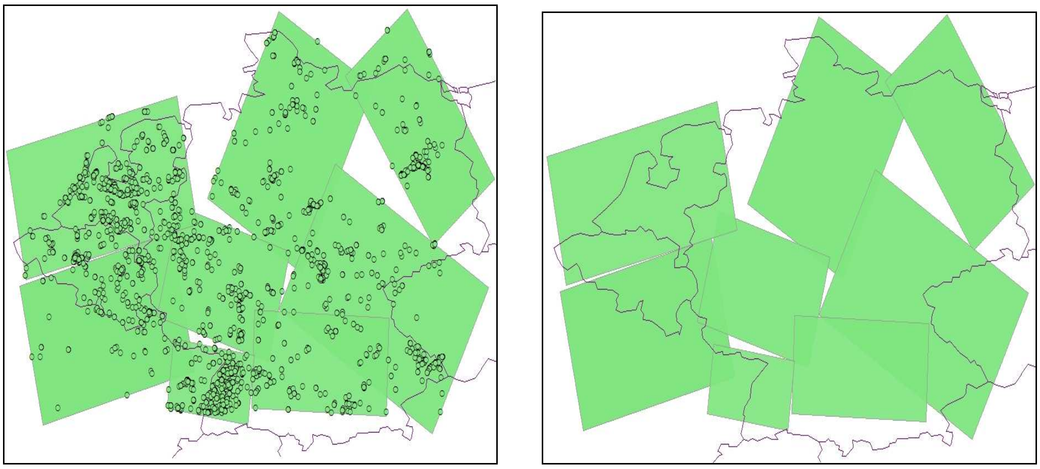


Figure 5.7: Forest data points converted to polygon using rectangle tool

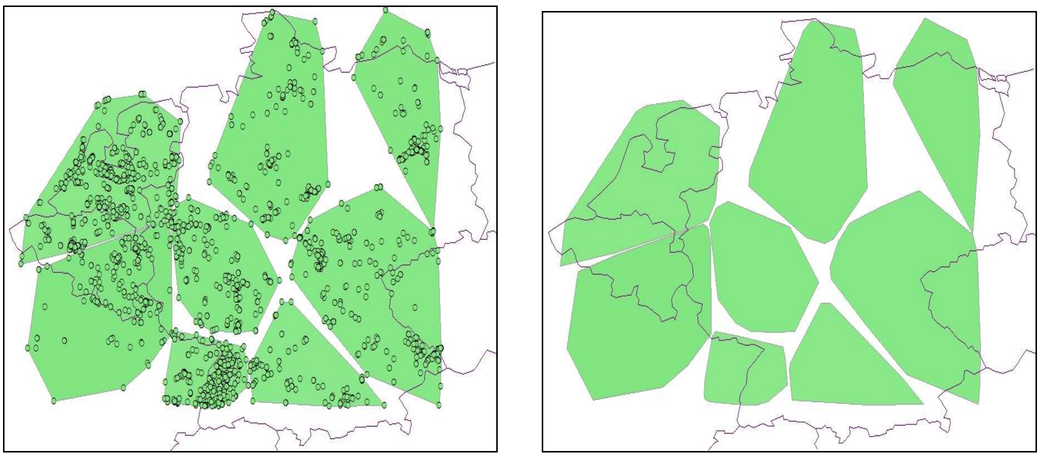


Figure 5.8: Forest data points converted to polygon using convex hull tool

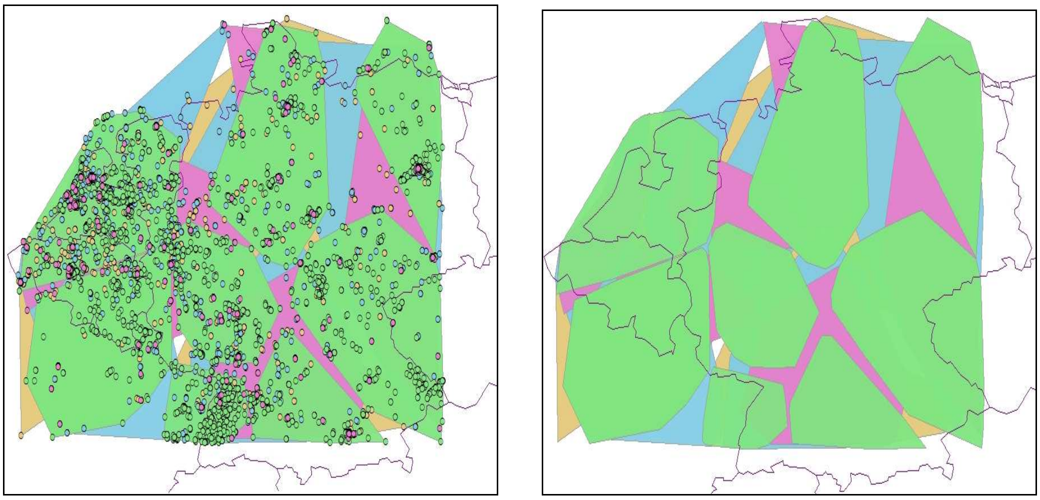


Figure 5.9: Multiple neogeography data points converted to polygon using convex hull tool

By looking on figure 5.5, 5.6, 5.7, 5.8 and 5.9 one can realize that the convex hull tool best fit the clusters compared to other tools such as extent, rectangle, and circle. These tools, extent, rectangle, and circle can fit well only when the cluster's shape look like the tool itself. This implies that the clustered points can be converted to polygons of varying shapes. In order not to distort the original shape of a cluster at a given scale the polygon generated should fit the shape of the cluster. However, user can decide which polygon type to use. When the multiple topics are displayed together as shown in figure 5.9, some features are obstructed. This problem can be minimized by applying visualization techniques explained in section 5.5. To facilitate user to choose type of a polygon tool to represent the clustered points, an interface shown below in figure 5.10 can be used. This interface gives the user an option of selecting the clustering method, classification by using the slider bar, layers to display and a polygon type tool for representing the clustered data at a given scale.

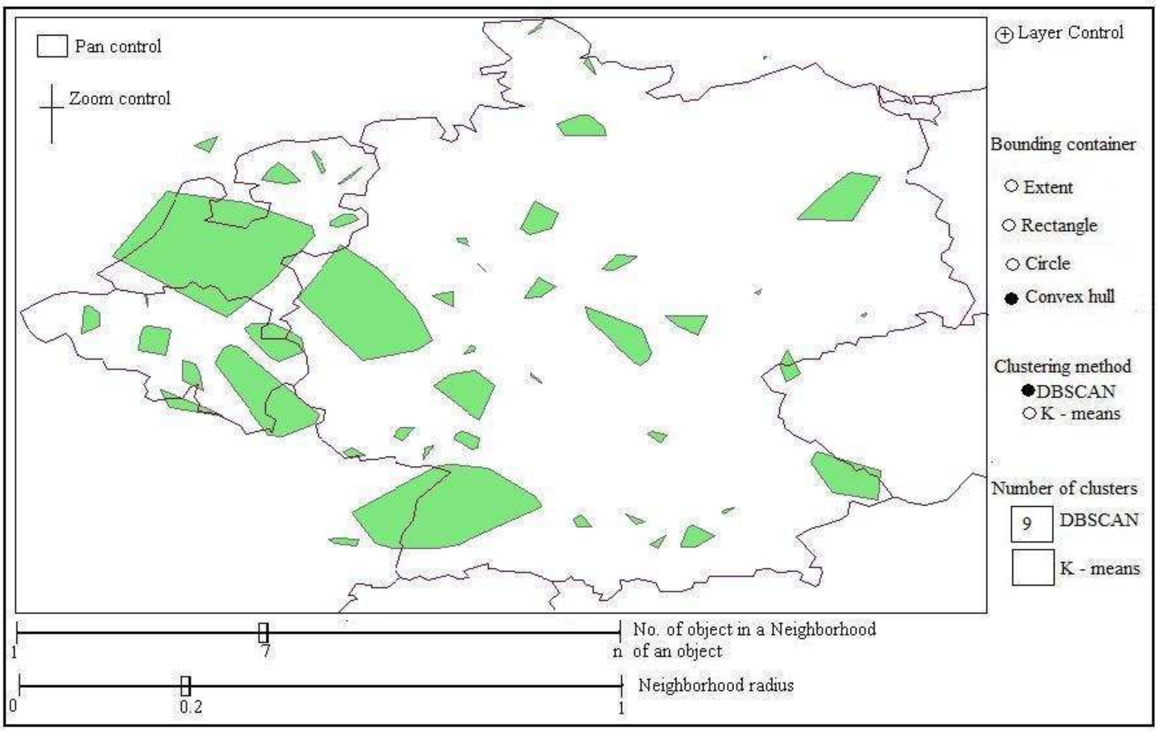


Figure 5.10: Cartographic interface control with white background

From figure 5.10, a user can display one or all the neogeography topics by switching on/off the respective layers. These layers contain the clustered neogeography data points. Using the DBSCAN algorithm, the number of clusters and the shape of clusters depends on the number of objects in a neighborhood of an object and the neighborhood radius. Figure 5.10 is a clustered neogeography map, representing a neogeography topic 'forest' as a layer. Clustering the data using DBCAN, number of objects in a neighborhood of an object and the neighborhood radius used are 7 and 0.2 respectively. Using these parameters the resulting number of clusters are 9 and the best fitting polygon tool used to convert the resulting clusters to polygon appears to be the convex hull. By varying either neighborhood radius or the number of objects in a neighborhood of an object different number of clusters will be obtained. Depending on the nature and distribution of the data, at a given scale, varying those parameters the resulting clusters will remain points or one cluster coverage of all data range. When the clusters remain as individual point, converting it to a polygon will not be possible. When the whole data range is taken as one cluster, one polygon will be generated.

5.5. Visualization

The summarized neogeography topics from point to polygon always take the shape of the tool used to convert it. For example when the circle tool (figure 5.5) used as bounding box, the resulting polygons will also be a circle. It is also true for the case of convex hull tool, extent tool and rectangular tool. This limits the range of symbols to represent the summarized neogeography topics.

Neogeography topics namely forests, hotels, restaurants and schools represent quantitative information. Summarizing these neogeography topics to polygon changes them to qualitative. From chapter three, it is clear that qualitative information can better be represented using Bertin's visual variable namely colour, orientation, shape/form as shown in table 3.1. Colour has both the differentiation properties and association properties. Applying colour to the summarized neogeography topics, the topics need to be perceived as different but of equal importance. Therefore, use of colour, transparency and outline colour for polygons seems to be the alternatives for increasing the symbology and assist in visualization.

By using the DBSCAN method, the neogeography map shown in figure 4.2 can be summarized using the convex hull tool and visualized as shown in Figure 5.11 (a), 5.11 (b) and 5.11 (c). The neogeography topics are represented as coloured polygons in figure 5.11 (a), transparency in figure 5.11 (b) and as hollow polygons with outline colours in figure 5.11 (c). Depending on the purpose of the map user can visualize the map using colour, transparency or outline colour. Looking on these figures, one can observe that the satellite image from Google Earth has been used as a background showing the administrative boundaries within the country. When using the white background the map appears as shown in appendix A Figure A_M4. The question whether to use a white background or the satellite image in visualizing the summarized neogeography maps is left in the hands of the user.

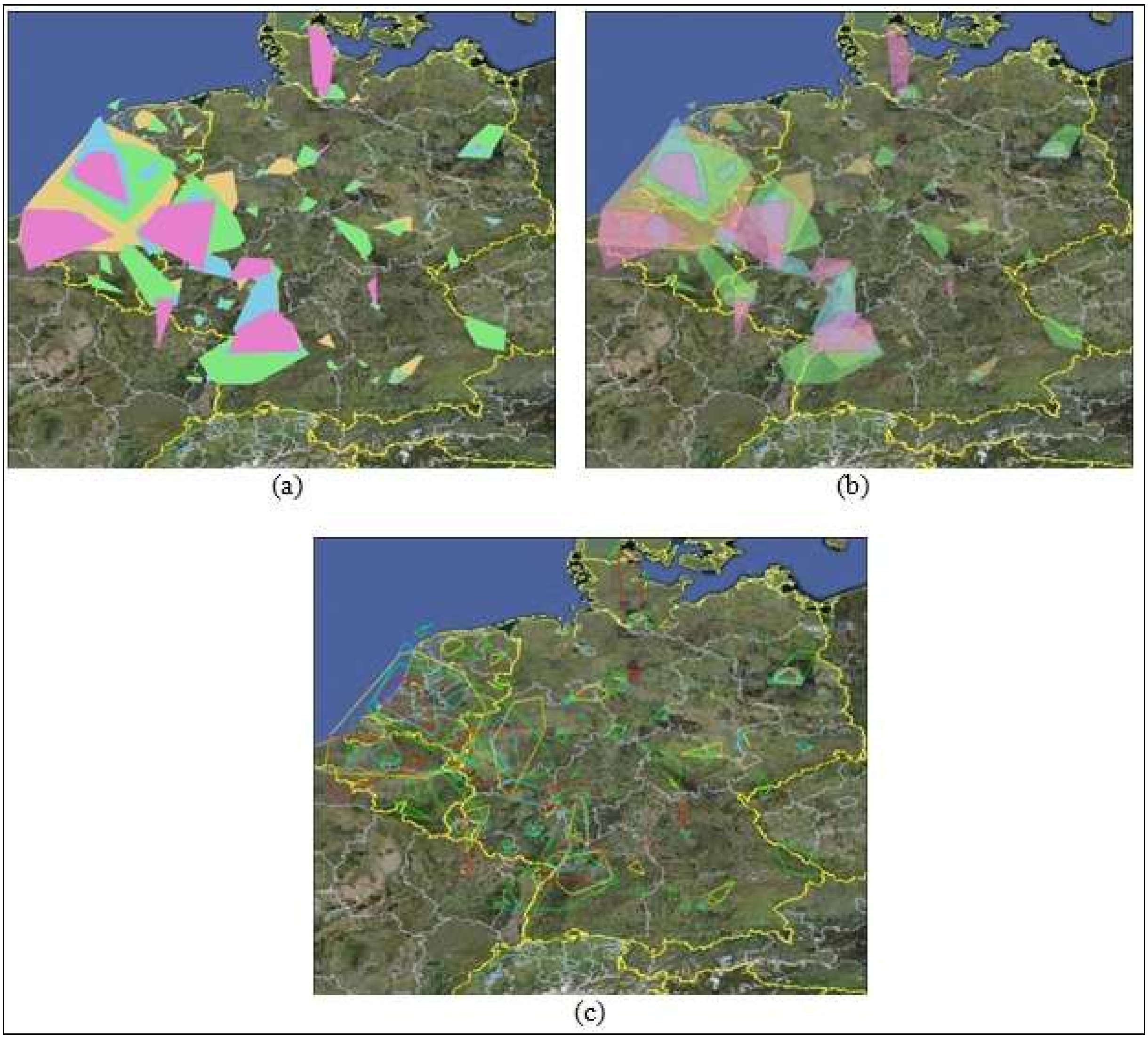


Figure 5.11: Visualization of neogeography topics: (a) Coloured polygons; (b) Transparency; (c) Hollow polygons with outline colours.

5.6. Cartographic map

The cartographic maps shown in Appendix B were made from the same neogeography data used to create the neogeography map shown in Appendix A which includes figures A_a, A_b, A_c and A_d. Figures B1_a, B2_a, B3_a and B4_a in Appendix B are the neogeography topics: forest, hotel, school and restaurant displayed by changing only the symbol type. The grids of size 30′ x 30′ were created and draped on the neogeography topics in place of administrative units because of the difficulty in acquiring administrative units covering the entire study area. Counts of the neogeography data points per square grid were made to create the cartographic maps shown in Appendix B: B1_b, B2_b, B3_b and B4_ b. Figure B5_a represents all the four neogeography topics displayed by changing the symbol type. Figure B5_b is a cartographic map created by displaying together maps in Appendix B: figure B1_b, B2_b, B3_b and B4_ b.

6. Evaluation and discussion of the output

6.1. Introduction

This chapter compares the original neogeography maps with the summarized maps and with the cartographic maps. The summarized cartographic maps were created using the two clustering methods namely k – means as shown in figure 5.5 – figure 5.9 and DBSCAN as shown in Appendix A, Figure A_S1_a – Figure A_S4_d. By considering figure 5.4, we can observe that several maps can be created as a result of applying only two clustering methods. These maps seem to be different in terms of coverage because of the property of bounding containers, and even the message they deliver to the users. To select an optimal map from the many maps, a test has been conducted by using a small group of users called focus group. The group, as a representation of users was given some tasks and their response noted.

6.2. Evaluation

The evaluation of the maps produced was based on comparing three maps. The evaluation was done through discussion with the focus group. The maps shown in appendix C were discussed by the focus group where their thoughts and feelings have been considered in ascertaining whether the idea of summarizing the neogeography maps from points to polygon works. Another task given to the focus group was the idea of designed neogeography interface. In making the discussion active during interviewing, and to enable the focus group to perform the visual analysis on the produced maps, they were asked some questions as listed in appendix C.

6.2.1. Focus group

The focus group was interviewed based on the two cases described in chapter 5. Case one was an individual task. They were asked to rank the different maps of England. Case two consists of three tasks and covers The Netherlands, Belgium and Germany. The three tasks were: Selection of bounding containers that best represent the neogeography topics, symbology for representing the data and discussing the functionality of the conceptual neogeography interface. The purpose of evaluating the maps is to test if people understand neogeography maps, together how they perceived the proposed summarized maps being the way of reducing the cluttering problem. The whole interview session was recorded on video and audio. Eight participants from different countries and specialization were invited. Their names, country of origin and specialization are as shown in the table below. Seven participants were interviewed in case one and eight participants were interviewed for case two.

	Name	Country	Specialization
1	Habtom Tsega	Ethiopia	Geography
2	Juliet Gwenzi	Zimbabwe	Georology
3	Duc Ha Nguyen	Viet Nam	Informatic
4	Cut Susant	Indonesia	Geodetic Engineering
5	Ahmed Elmadhoun	Palestinian Territory, Occupied	Civil Engineering
6	Rodrigo Sperb	Brazil	Environmental Engineering
7	Maia Zumbulidze	Georgia	Geography
8	Lizda Iswari	Indonesia	Systems developer

Table 6.1: Names, country and specialization of the focus group members

Case One

Arrangement of England maps

The maps shown in figure 6.1 were used during the interview with the users to rank the neogeography maps as 'best', 'good', 'bad' or 'worse'. These maps are also shown in Appendix C in large size. The neogeography topics data used for designing the map were extracted from Flickr website. The neogeography topics were the supermarkets, playgrounds, airports and traffic lights.

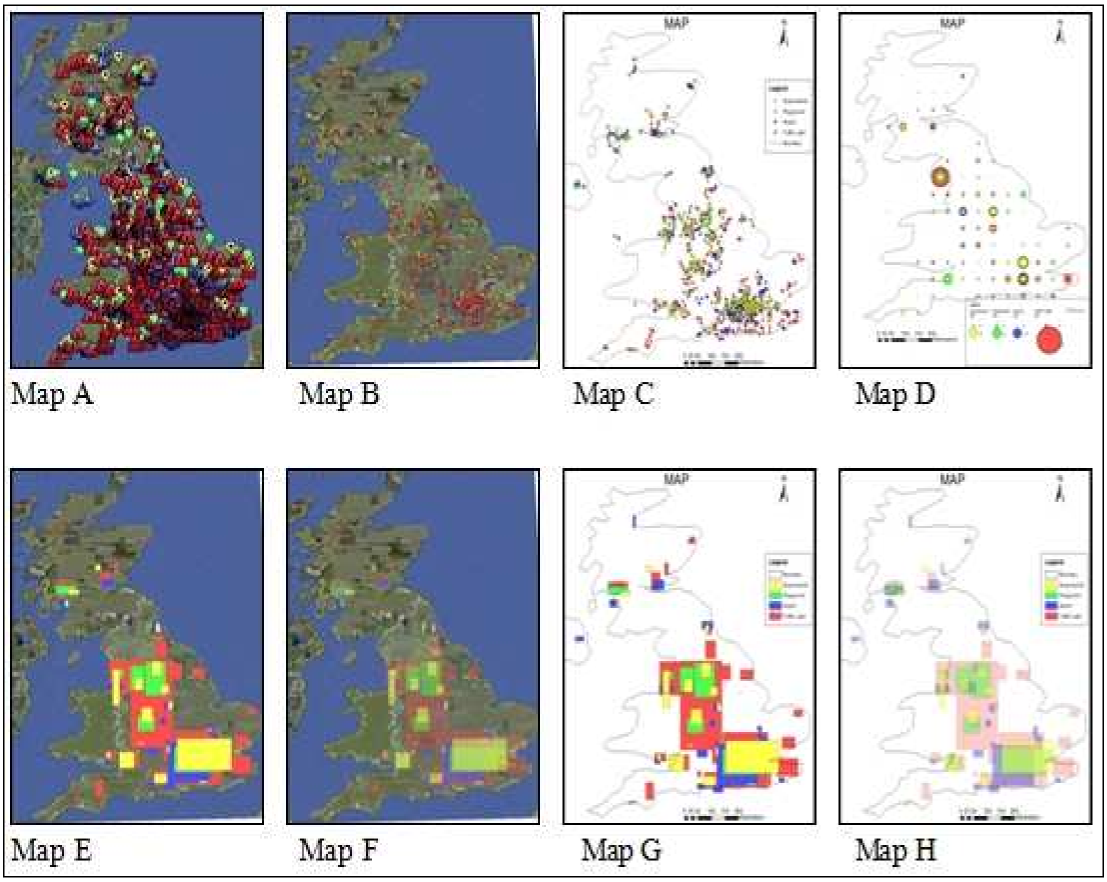


Figure 6.1: Designed maps for England using the same neogeography data set

The ranking of eight maps shown in figure 6.1 by the users gives the following results, shown in table 6.2 below.

Map/Rank	Best	Good	Bad	Worse
Map A		1		6
Map B		5	1	1
Map C	1	4	1	1
Map D	1	3	2	
Map E	1	1	5	
Map F	1	3	3	
Map G	3		4	
Map H		3	4	

Table 6.2: Ranking results for the England maps

Observing the table above, we can see that different answers were obtained from the users. For instance, Map B was ranked by 5 users as a 'Good' map while 1 user ranked it as 'bad' and 1 user ranked it as a 'worse' map. Most maps that were selected as 'best' seem to be simple to interpret when referring to the legend. Cartographic maps seem to represent the message best but the symbol size was considered by the users as very small. Some users would prefer the map to be digital and interactive rather than paper map to give the zoom in and zoom out functionality during the test. In reality when one provides zoom functionality, then neogeography needs be compared with the web cartography. At a small scale, the users also suggest the neogeography topics be changed from point representation to polygon representations. Map A, which seems to be a typical neogeography map was mainly described as 'worse' map because the map objects were too cluttered when displayed at a very small scale.

Case two

Comparison of bounding boxes with the neogeography map:

The users were provided with the neogeography map of Western Europe covering the Netherlands, Belgium and Germany. The neogeography map shown in Appendix C as figure 1, was summarized by using bounding containers namely rectangle tool, circle tool, extent tool and convex hull tool as shown in figure 2(a), 2(b), 2(c) and 2(d) respectively. Most of the users selected the convex hull tool as a bounding box which represents the given data set well. They did select the map summarized using the convex hull tool because the tool appears to cover the neogeography map well and the shape of the clustered data is maintained.

The use of circle tool was challenging to the users, as some realized it as a bounding container while others realized it as a symbol for presenting the quantitative information (proportional symbol). The discussion among the users raised the notion that the use of any tool will depend on the purpose of the map.

Comparison of bounding boxes with the cartographic map:

The users were also given a task to compare the cartographic map with the bounding containers. Most of the users propose that the cartographic map should have the satellite images as a background to have a better representation of information in the map. When the users were required to compare the neogeography maps with the cartographic maps, they preferred the cartographic maps and even proposed the change of symbol size and inclusion of a background for the cartographic maps to be more meaningful.

The users were also provided with the interface shown in Appendix C in figure 7 and figure 8 as a means of creating and presenting all the maps that were given to them indicated as figure 2(a), 2(b), 2(c) and 2(d). Figure 7 was created with a white background while figure 8 used a satellite image as a background. Discussions revealed that the interface will be the only option to solve the problem that arises between the user's different opinions. The users preferred the idea of having an interface to represent the summarized neogeography map as it would give them the freedom for use and ability to visualize what they want at any point in time. Users would also be able to classify the data and also realize into how many classes the data have been partitioned. Users preferred the interface using a satellite image as a background to the interface with a white background as it gives more information about the layers and administrative boundaries.

The users were asked if it is a good idea to summarize the neogeography data and use the interface to represent it. Again, most of the users preferred the idea of summarizing the neogeography data and the use of an interface. Summarizing the neogeography data points to polygon was preferred by the users only when the data are visualized at a very small scale. However, at a large scale they prefer the neogeography map so as to get much detail.

Users were also eager to understand the difference between summarizing the neogeography maps and the map generalization. The main difference explained was summarization is not subjective compared to generalization. In generalization, one geographic feature can be accentuated while another can be deemphasized depending on the purpose of the map.

Symbology to represent the data

The summarized map was visualized by applying either full colour to polygons, transparency, or hollow polygons with outline colours as shown in Appendix C and by figure 3(a), 3(b) and 3(c) respectively. Users identified that full coloured polygons obstruct others when displayed as layers on top of each other. Hollow polygons with outline colours seem to be more confusing when there are many lines. Applying hollow polygons with outline colours made their boundaries visible but not the information. The transparency was preferred as the users were able to see all the layers and determine the coverage. When the users were asked to choose a map for visiting forest sites in Europe, they also selected the transparent one as the coverage was clearly indicated.

In the discussion with the users, they proposed that the summarized maps should show the amount or quantity of the features summarized. This is observed as the size of the bounding boxes depends much on the distribution of the features and not the density of the features. They also proposed that the cartographic map to have a satellite image background as there is loss of information when using the plain or white background. The users agree that any bounding container to summarize the original data is a good idea but the shape to use will depend on the purpose of the map. Summarizing the neogeography maps using any bounding container can be accepted if one has the notion of coverage. A notion of quantity need to be considered through interactivity.

6.3. Limitations

Although the users found favour with the ideas of summarizing the neogeography maps together with the neogeography interface still there are some draw backs outlined. These are:

1. Summarizing the neogeography map not suitable for large scale.
2. The scale range at which the neogeography maps need to be summarized is not clear.
3. As the whole process is based on algorithms, some important features may be omitted.
4. Grouping of features is mainly due to distance and not density.
5. Selection of the clustering method and the bounding containers to use are left in discretion by users.
6. Use of transparency is limited when the number summarized of neogeography topics is large and there is a maximum overlay of the topics.
7. There is also a loss of information when overlaying two or more summarised neogeography features of equal sizes.
8. Bounding containers chosen are not suitable for bounding linear neogeography data.

6.4. Conclusion

The arrangement of the map of England by the users in terms of 'best', 'good', 'bad' and 'worse' reveal that the neogeography map A shown in Appendix C is 'worse' map due to cluttering. The cluttering of the map is due to symbol size and the scale at which the map is viewed. The neogeography data used in this research, which covers the Netherlands, Belgium and Germany appeared to be better described using the DBSCAN algorithm when the parameters are well set. Using this algorithm and the four bounding containers namely extent tool, rectangle tool, circle tool and convex hull tool; the convex hull tool seems to represent the study case data well. The realization of good representation of the data by using the convex hull tool is through visual inspection. This is observed in figures A_S4_a, A_S4_b, A_S4_c and figure A_S4_d in Appendix A. Inspecting the figures thoroughly, we can realize that when the points are converted (bounded) to form a polygon, the shape of the resulting polygon depends on the way the features or points are distributed outside the cluster. As these polygons are considered to be the elements of the summary map, they can have an impact on the appearance of the summarized map. When the clustered points are poorly defined by any bounding containers, the summary map can deviate much in approximating the original neogeography data. This has been observed when the tools like extent, rectangular and circle poorly bounded the case study data set. This does not mean that the tools are not good for describing the neogeography data but they were not suitable with reference to nature of distribution of the data, the number of clusters selected and the parameter settings used.

The three techniques used to visualize the summarized maps were perceived differently by the users. It was difficult to understand the summarized multiple topics displayed as a map using the full coloured polygons. However for the single topic as shown in appendix C figure 7 and figure 8, the full coloured polygons can be used. Displaying multiple topics using coloured polygons made it possible to analyze only the polygon that is on top of the others. The layer ordering is important in presenting the summarized neogeography multiple topics. The polygons displayed at the top dominate the whole map and are observed as being more important than other polygons. The use of outline colour defines well the coverage of the summarized neogeography maps. The complexity of understanding these coloured lines increases as the number of neogeography topics under consideration increase. The use of transparency appears as the optimal solution as the coverage of the features can be identified. The limitation of this is imposed by the visual isolation capacity of the user when the number of neogeography topics increases.

The idea of summarizing neogeography data was preferred by the users when the maps are to be viewed at small scale. The shape of the bounding container to use will depend on the purpose of the map. Users are able to perform classification and select the bounding container that describes well the data, using the proposed interface. The use of neogeography maps were also preferred by the users when the maps are to be displayed at large scale for detail viewing. Choosing either the summarized neogeography map or the neogeography map will depend on the purpose of the map.

7. Conclusion and recommendations

7.1. Conclusion

In this study, neogeography multiple topics maps have been summarized. The original point representation has been changed into a polygon representation as a means of reducing the cluttering problem. This also changes the data representation from quantitative to qualitative. In summarizing the neogeography maps, two clustering methods namely; k – means and DBSCAN, together with various bounding tools namely extent, rectangle, circle and convex hull have been tested. Both clustering methods give good results when the parameters are well set. The bounding tool to use depends on nature of data distribution.

The maps of England have been presented to the users in trying to find out a ‘good’ and ‘bad’ map. Users chose the map displayed as points on a satellite image background as ‘good’ map because satellite image adds more information to a map. Users identified Map E in figure 6.1 as a ‘bad’ map because most of the information were obstructed by the polygons, including background information contained in a satellite image. In figure 6.1, Map A was identified by the users as a ‘worse’ map because of cluttering.

Most neogeography data used in this research were clustered using DBSCAN algorithm. DBSCAN can discover the shape of cluster quickly by varying the number of objects in a neighborhood of an object (k) and the neighborhood radius (Eps). The summarized neogeography maps based on DBSCAN algorithm were then presented to the users for discussion.

The map products, namely; the neogeography maps, summarized neogeography maps and cartographic maps were compared at a small scale. The summarised neogeography map using a convex hull tool as a bounding container with a satellite image background was the ‘best’ using a transparency visualization technique. The summarized neogeography maps using the circle tool was challenging to interpret. The users perceived the varying circle sizes as representation of quantitative information where proportional symbols have been applied, which is not the case since the extent of the circle represents distribution of data. The cartographic maps were considered ‘good’ by the users who also suggested the symbol sizes need to be increased. Using satellite image as a background in cartographic maps is preferred to increase map communicability.

Using the DBSCAN method to summarize neogeography maps is ideal for perception of coverage (distribution) when the maps are viewed at small scale. For the perception of density of information, OPTICS is preferred as it can detect clusters of varying densities. At large scale, original neogeography maps are preferred in order to view details.

Cartographic maps seem to represent well the quantitative information by use of proportional symbols. Users encountered difficulty in realizing the quantity when the symbol sizes were too small.

Users also faced difficulties in differentiating colours when the symbol sizes were too small. This made them to propose a change in symbol sizes for the cartographic maps to be more user friendly.

Time limitation made the implementation of interactive neogeography interface incomplete. This caused the users to be interviewed and conduct discussions on a conceptual interactive neogeography interface based on paper maps. The conceptual interactive neogeography interface was tested by the users and found to be able to present the summarized neogeography maps. The summarization of neogeography maps and their visualization is the way one can combine neogeography with cartographic approaches.

The use of an interface offers the users a means to select the clustering methods and the bounding container to use. Furthermore, it gives the user an option of switching on and off, one or more summarized neogeography topics with the associated quantities.

7.2. Recommendations for further research

This research recommends the following for further research:

- Hierarchical clustering may be good in presenting the neogeography data at varying scale. Therefore, this research proposes further study and implementation of hierarchical clustering.
- Users requested the summarized maps to show the amount of details (density of information). OPTICS appears to be good for identifying the clusters of varying density and hence need to be implemented for clustering the neogeography topics.
- Due to limitation of time, the implementation of the designed conceptual interactive neogeography interface was incomplete and therefore there is a need to implement it.
- Use of ellipse and 2D alpha-shapes as bounding containers seem to be a promising idea for preserving the shape of the cluster and so need to be implemented and tested
- The use of clustering algorithms and bounding containers for polygons explore important patterns from the neogeography data set, this needs to be implemented on linear features such as roads, rivers, railways etc and a tool for converting clustered points to a line needs to be used so as to explore the linear neogeography data set for pattern recognition.

URL:

1. http://geothought.blogspot.com/2009/07/are-we-now-in-post-neogeography-era.html
2. http://www.mcwetboy.net/maproom/2006/04/neogeography.php
3. http://googleearthdesign.blogspot.com/2007/08/steve-chilton-interview.html
4. http://www.signtific.org/en/signals/traditional-cartography-vs-neo-geography-and-gis
5. http://www.google.com/landing/typhoon-ondoy.html
6. http://bbs.keyhole.com/ubb/ubbthreads.php?ubb=showflat&Number=1259914#Post1259914
7. Web Mapping for Everyone: http://www.nado.org/conference_files/cockburngis.pdf
8. http://code.flickr.com/blog/2009/05/21/flickr-shapefiles-public-dataset-10/
9. http://www.flickr.com/services/api/explore/?method=flickr.photos.search
10. http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/index.html
11. http://en.wikipedia.org/wiki/Cluster_analysis
12. http://webdocs.cs.ualberta.ca/~zaiane/courses/cmput690/slides/Chapter8/sld001.htm
13. http://www.stat.psu.edu/~jiali/course/stat597e/notes2/mix.pdf
14. http://www.proveandimprove.org/new/meaim/developgoodindicators.php
15. http://www.slideshare.net/sderle/gis-and-the-neogeographer

REFERENCES:

1. Kraak, M. and A. Brown, *Web cartography: developments and prospects*. 2001: CRC.

2. Kramers, R.E., *Interaction with maps on the internet - A user centred design approach for the atlas of Canada.* Cartographic Journal, 2008. **45**(2): p. 98-107.

3. Goodchild, M., *Citizens as sensors: the world of volunteered geography.* GeoJournal, 2007. **69**(4): p. 211-221.

4. Flanagin, A. and M. Metzger, *The credibility of volunteered geographic information.* GeoJournal, 2008. **72**(3): p. 137-148.

5. Haklay, M., A. Singleton, and C. Parker, *Web mapping 2.0 : the neogeography of GeoWeb.* In: Geography compass, pp. 2011-2039, 2008.

6. Sangawe, G., *Using geographic names and geotagging for determining vague boundaries of geographic regions*. 2009, ITC: Enschede. p. 97.

7. Turner, A., *Introduction to neogeography*. 2006, CA: O'Reilly Media.

8. Sui, D.Z., *The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS.* Computers, Environment and Urban Systems, 2008. **32**(1): p. 1-5.

9. Sieber, R. *Geoweb for social change.* 2007 [cited 2009 August 12]; Available from: http://www.ncgia.ucsb.edu/projects/vgi/docs/supp_docs/Sieber_paper.pdf.

10. Goodchild, M., *Commentary: whither VGI?* GeoJournal, 2008. **72**(3): p. 239-244.

11. Dent, B.D., *Cartography : thematic map design*. 5th ed. 1999, Dubuque, IA: WCB/McGraw-Hill. xiv, 417.

12. Cartwright, W., *Considerations for using Google maps and imagery with mobile devices.* 2009.

13. Kraak, M. and N. He, *Organising the neo-geography collections with annotated space-time paths.*

14. Allan, A. *Advanced OSM cartography* 2009 [cited 2009 12 August]; Available from: http://www.vimeo.com/5609679.

15. Dr. Satyaprakash. *NEOGEOGRAPHY: Goodbye to GIS?* 2008 [cited 2009 October 06]; Available from: http://www.gisdevelopment.net/magazine/global/2008/february/70.htm.

16. National Geographic News. *"Neogeography" Blends Blogs With Online Maps*. [cited 2009 October 06]; Available from: http://news.nationalgeographic.com/news/2006/04/0425_060425_map_blogs.html.

17. McFedries, P. *neogeography*. 2006 [cited 2009 October 13]; Available from: http://www.wordspy.com/words/neogeography.asp.
18. Turner, A. *Neogeography - towards a definition*. 2007 [cited 2009 October 4]; Available from: http://highearthorbit.com/neogeography-towards-a-definition/.

19. FISCHER, F. *Learning in Geocommunities an explorative view on geo-social network communities*. 2009 [cited 2009 October 07]; Available from: http://www.schule.at/dl/Learning_in_Geocommunities_-_Florian_Fischer_-_GI_Forum_2009.pdf.

20. Coote, A. and L. Rackham. *Neogeographic data quality – is it an issue?* 2008 [cited 2009 October 30]; Available from: http://www.consultingwhere.com/resources/Neogeography+Data+Quality+-+is+it+an+issue+-+V1_1.pdf.

21. Kraak, M.J. and F.J. Ormeling, *Cartography : visualization of geospatial data*. Second edition ed. Also available in Indonesian : Kartografie visualisasi data geospasial. 2003, Harlow: Addison Wesley Longman. 205.

22. de By, R.A., P.Y. Georgiadou, R.A. Knippers, M.J. Kraak, Y. Sun, M.J.C. Weir and C.J. van Westen, *Principles of geographic information systems : an introductory textbook*. Third edition ed. ITC Educational Textbook Series;1. 2004, Enschede: ITC. 226.

23. Block, C., *Design:graphic variables*. 2009, International Institute for Geo-Information Science and Earth Observation (ITC). p. 1-42.

24. Robinson, A.H. R.D Sale and J.L Morrison, *Elements of cartography*. Sixth edition ed. 1995, New York etc.: Wiley & Sons. 674.

25. Mank, T., *Typography*. 2008, International Institute for Geo-Information Science and Earth Observation (ITC). p. 25.

26. Kraak, M., *Settings and needs for web cartography*. Web cartography: Developments and prospects. 2001: CRC. 1-7.

27. Köbben, B.J. and M.J. Kraak, *Web cartography : dissemination of spatial data on the web.* In: Proceedings of the 2nd AGILE conference on Geographic Information Science, Association of Geographic Information Laboratories in Europe, Roma (1999), pp. 14-18, 1999.

28. Ben, S. *The eyes have it: A task by data type taxonomy for information visualizations*. 1996. Proceedings IEEE Symposium on Visual Languages, Boulder, Colorado, IEEE Computer society.

29. Sheesley, B. *Elements of Design*. 2009 [cited 2009 November 15]; Available from: http://cartography2.org/Chapters/page5/CartographicText.html.

30. Zalik, K.R., *An efficient k'-means clustering algorithm.* Pattern Recognition Letters, 2008. **29**(9): p. 1385-1391.

31. Wu, K.-L. and M.-S. Yang, *Alternative c-means clustering algorithms.* Pattern Recognition, 2002. **35**(10): p. 2267-2278.

32. Fraley, C. and A.E. Raftery, *How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis.* The Computer Journal, 1998. **41**(8): p. 578-588.

33. Viswanath, P. and V. Suresh Babu, *Rough-DBSCAN: A fast hybrid density based clustering method for large data sets.* Pattern Recognition Letters, 2009. **30**(16): p. 1477-1488.

34. Ester, M., H.P. Kriegel, J. Sander and X. Xu, *A density-based algorithm for discovering clusters in large spatial databases with noise*. 1996.

APPENDIX A: DBSCAN clustering

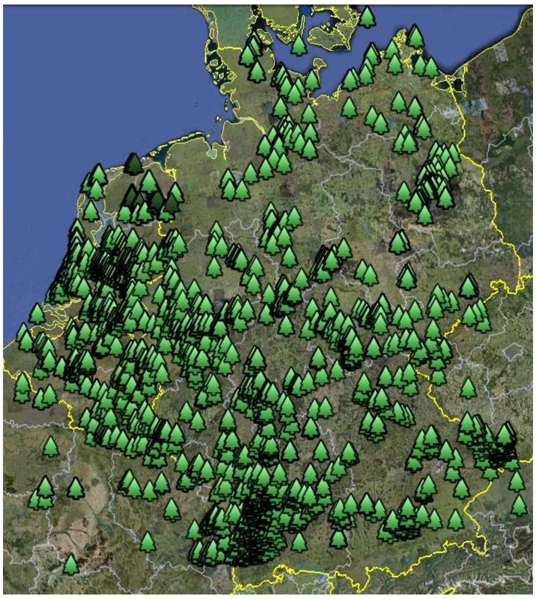


Figure A_a: Neogeography topic (Forest)

Figure A_b: Neogeography topic (Hotel)

Figure A_c: Neogeography topic (Restaurant)

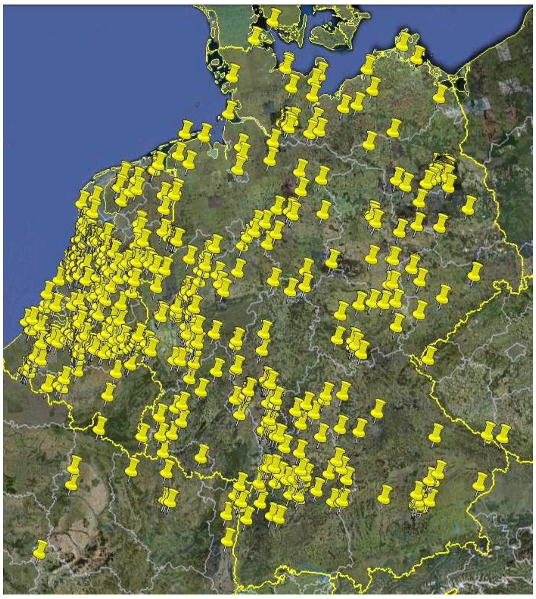


Figure A_d: Neogeography topic (School)

Figure A: Neogeography topics represented at small scale

Clustering Single (S) neogeography topic

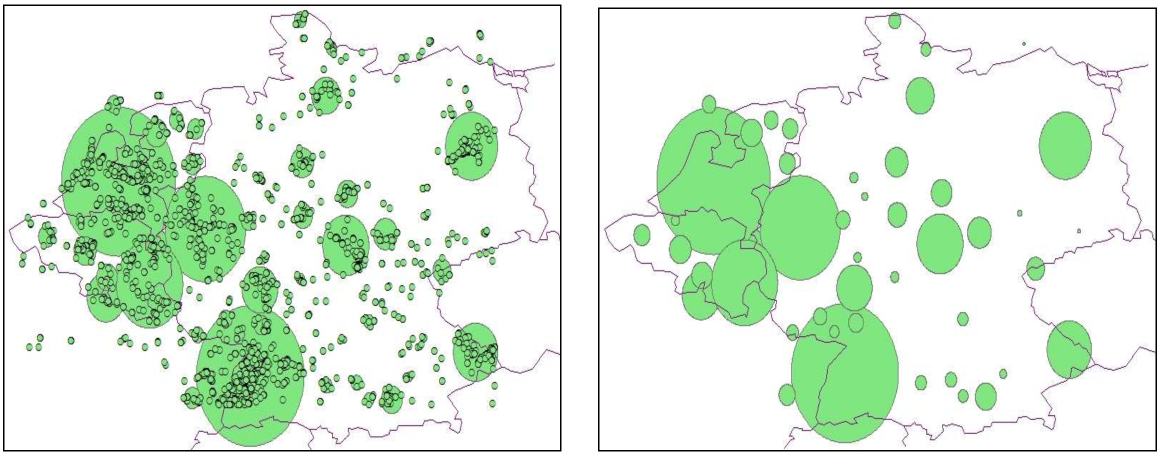


Figure A_S1_a: Forest data points converted to polygon using circle tool

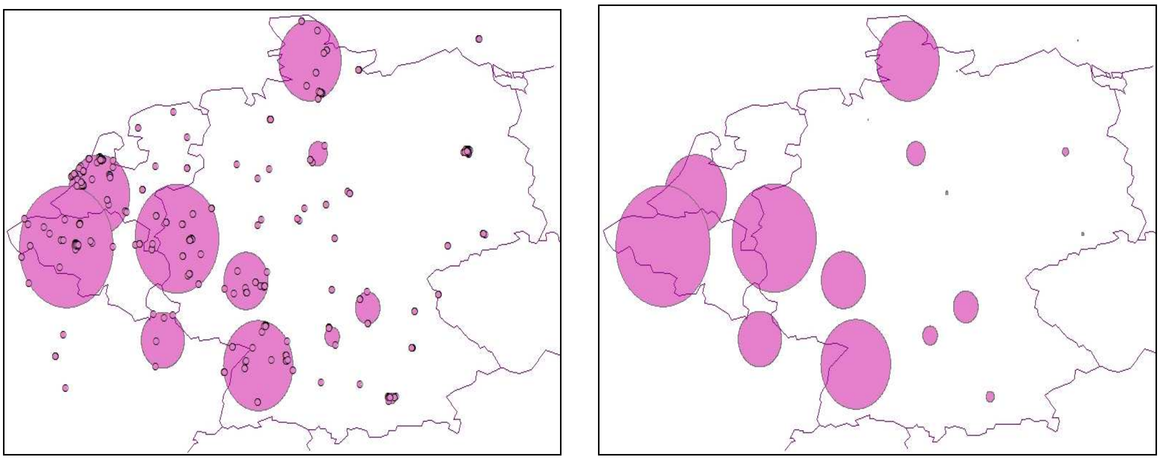


Figure A_S1_b: Hotel data points converted to polygon using circle tool

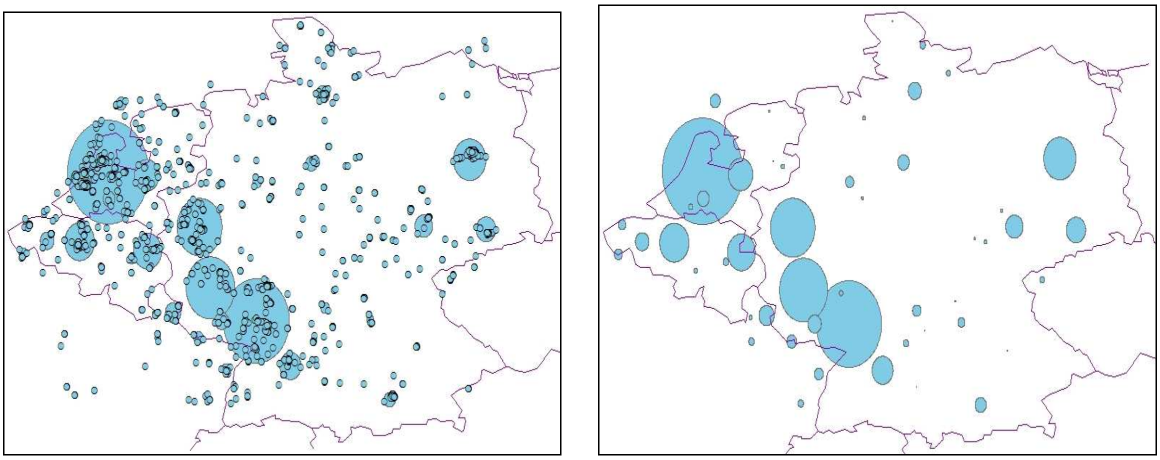


Figure A_S1_c: Restaurant data points converted to polygon using circle tool

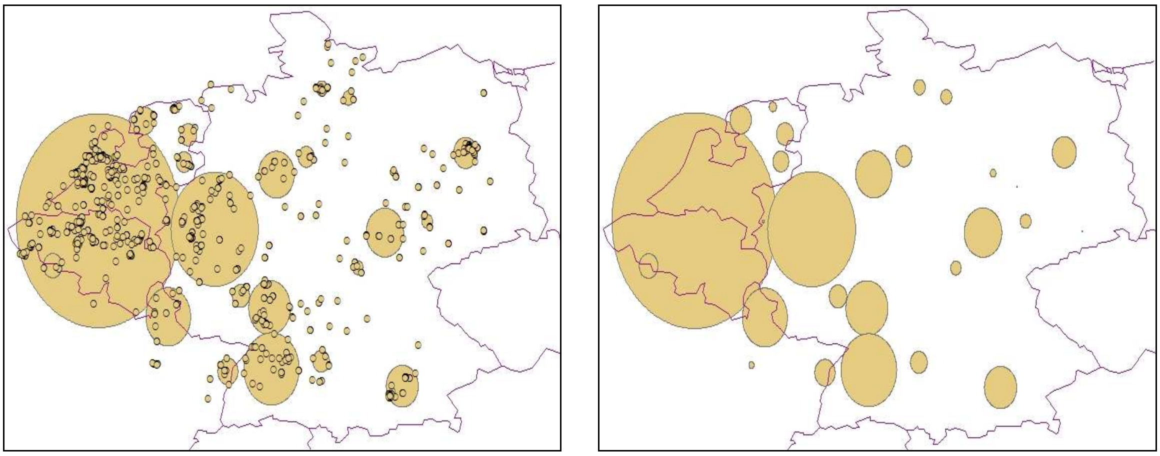


Figure A_S1_d: School data points converted to polygon using circle tool

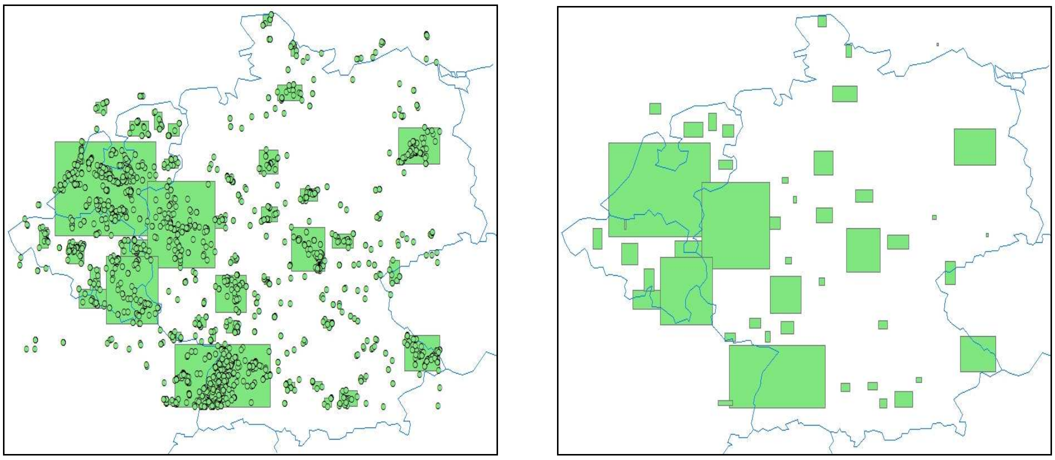


Figure A_S2_a: Forest data points converted to polygon using extent tool

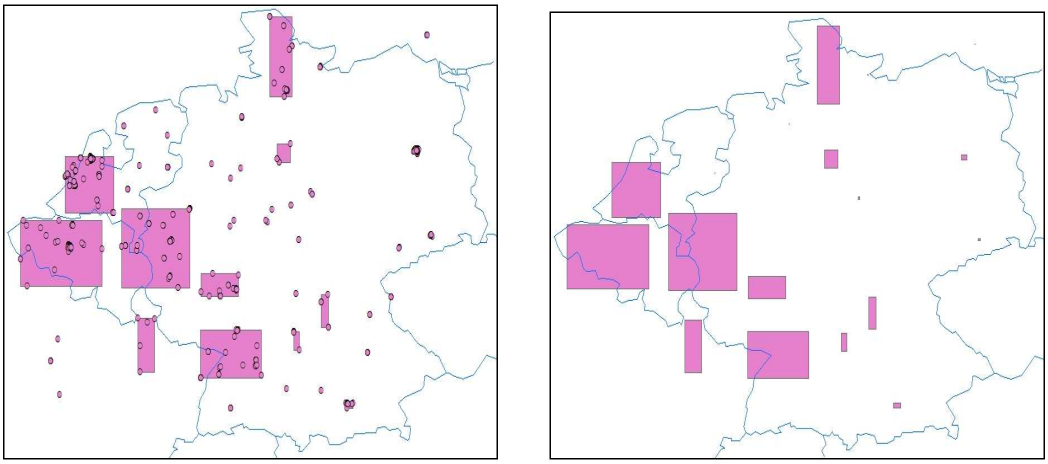


Figure A_S2_b: Hotel data points converted to polygon using extent tool

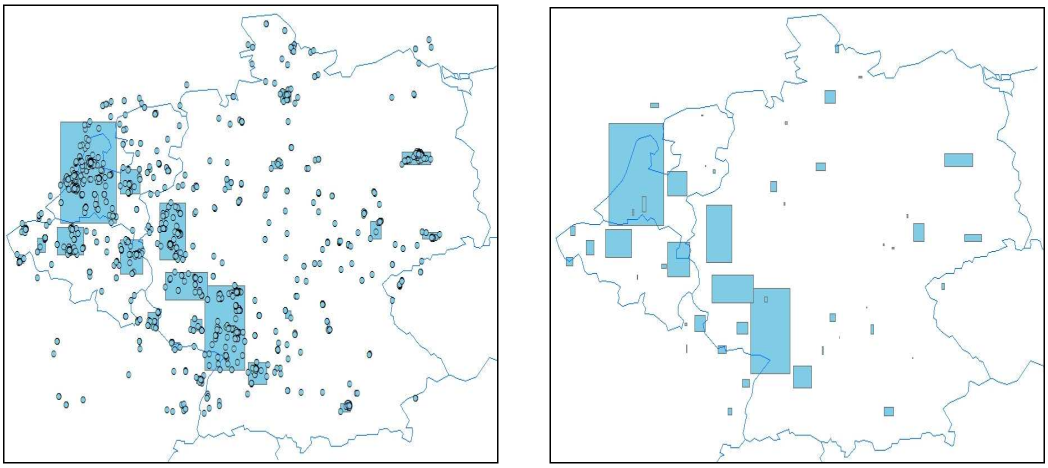


Figure A_S2_c: Restaurant data points converted to polygon using extent tool

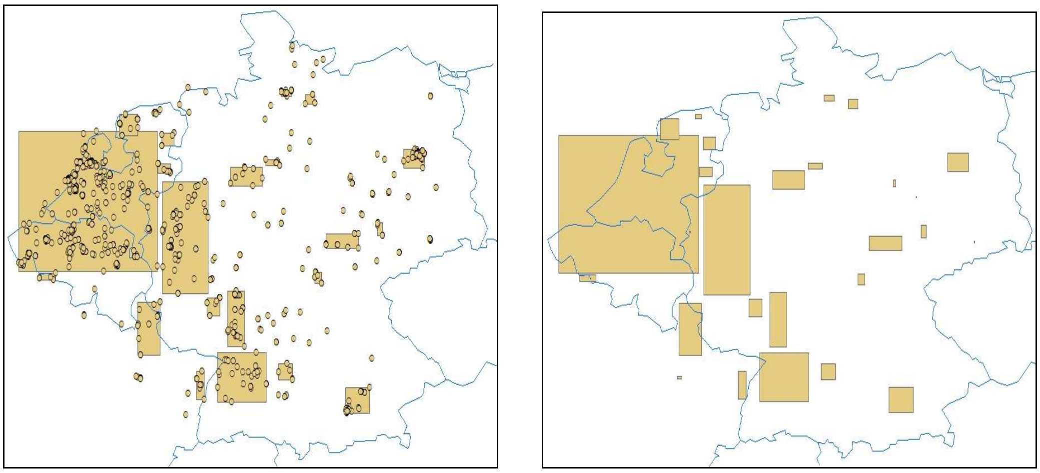


Figure A_S2_d: School data points converted to polygon using extent tool

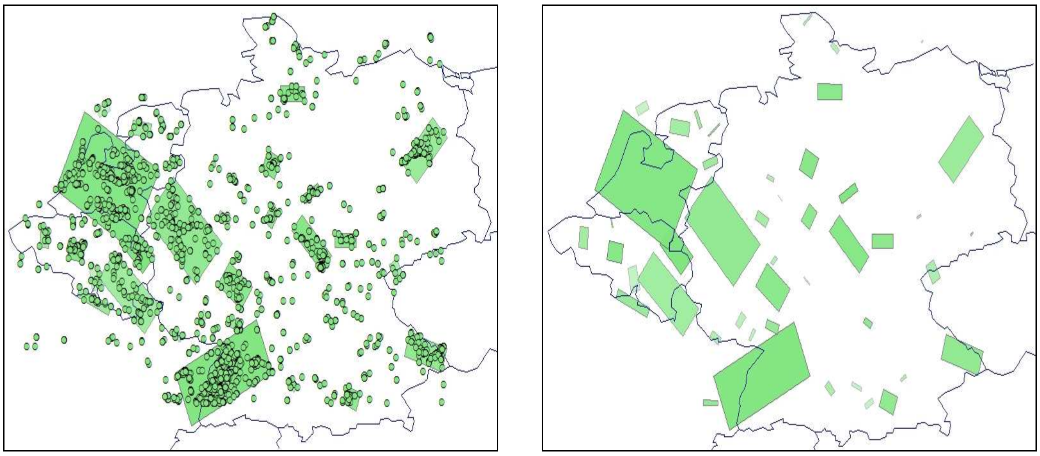


Figure A_S3_a: Forest data points converted to polygon using rectangle tool

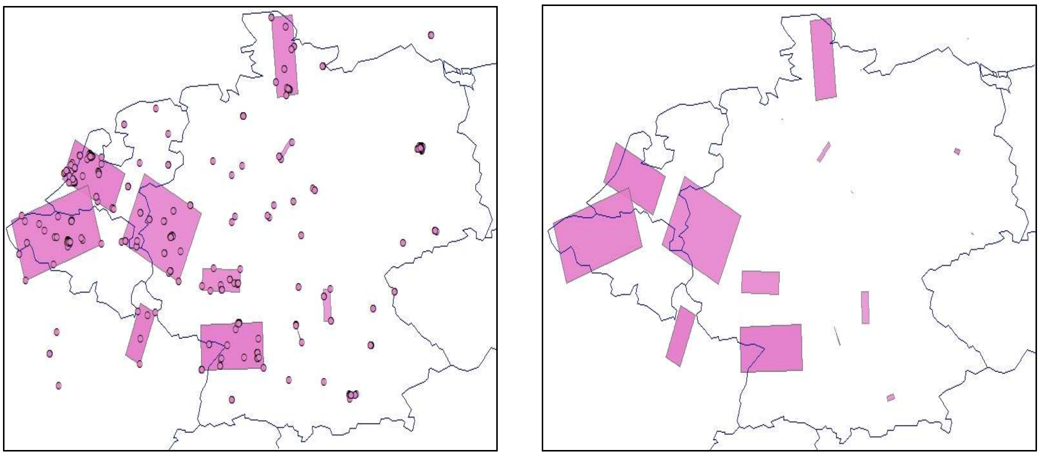


Figure A_S3_b: Hotel data points converted to polygon using rectangle tool

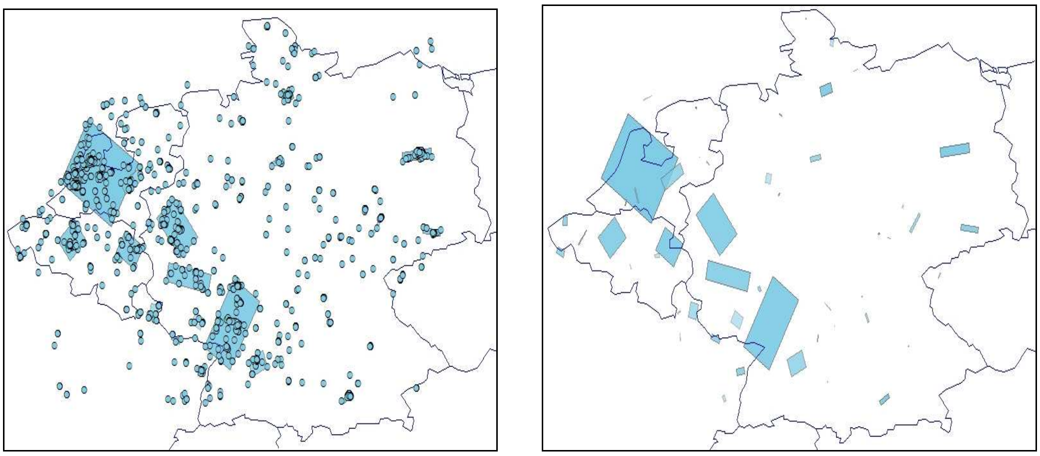


Figure A_S3_c: Restaurant data points converted to polygon using rectangle tool

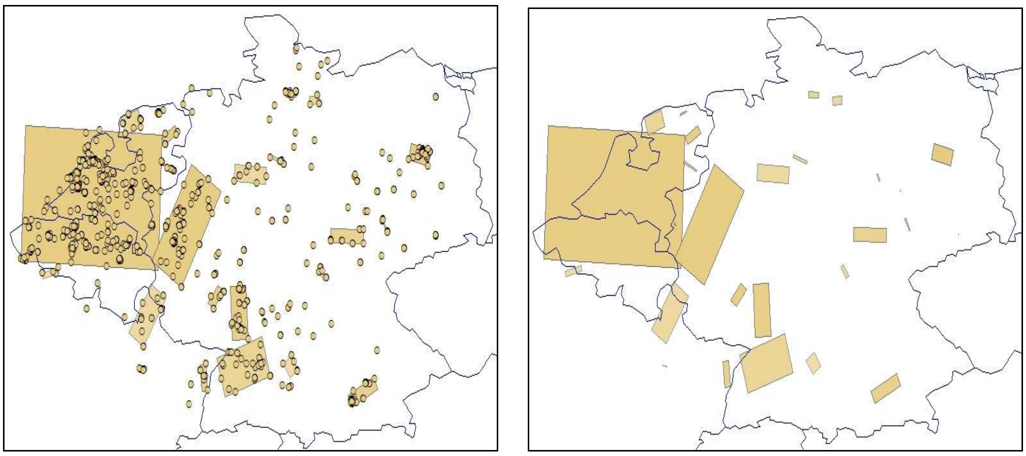


Figure A_S3_d: School data points converted to polygon using rectangle tool

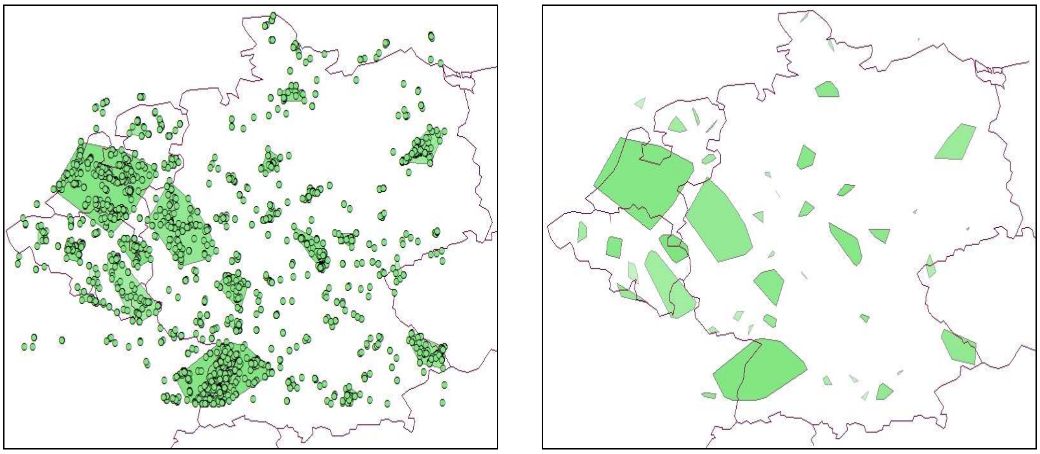


Figure A_S4_a: Forest data points converted to polygon using convex hull tool

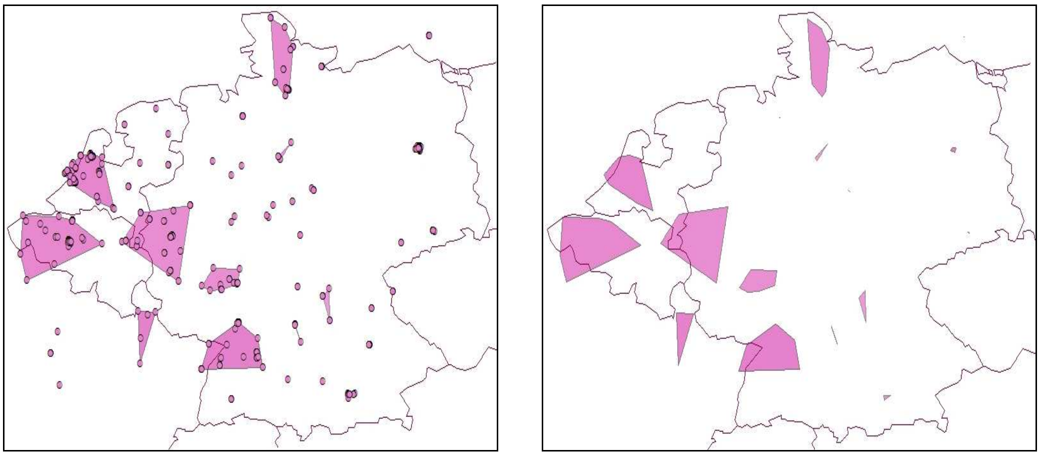


Figure A_S4_b: Hotel data points converted to polygon using convex hull tool

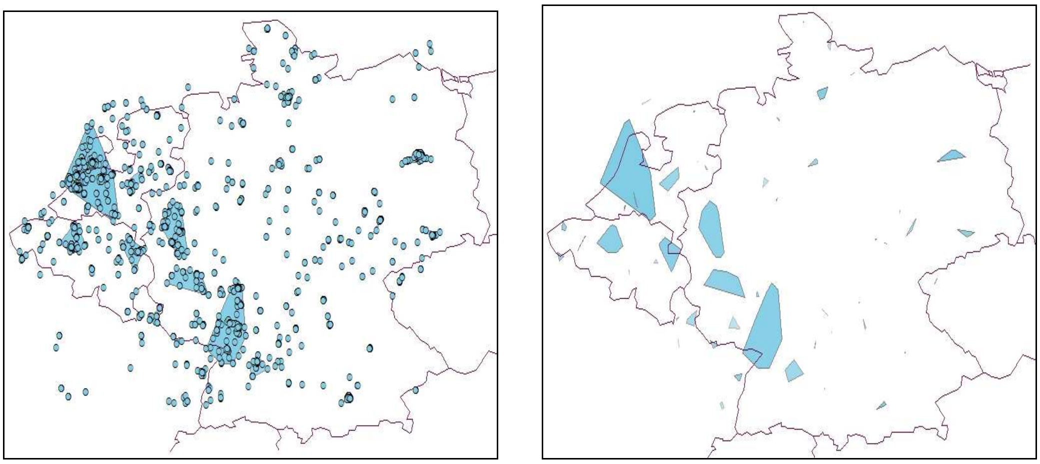


Figure A_S4_c: Restaurant data points converted to polygon using convex hull tool

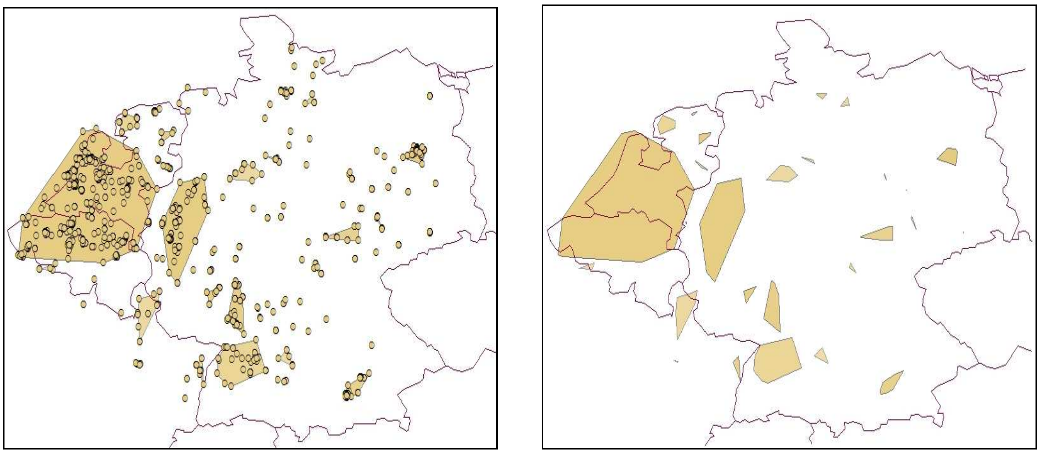


Figure A_S4_d: School data points converted to polygon using convex hull tool

Clustering multiple (M) neogeography topics

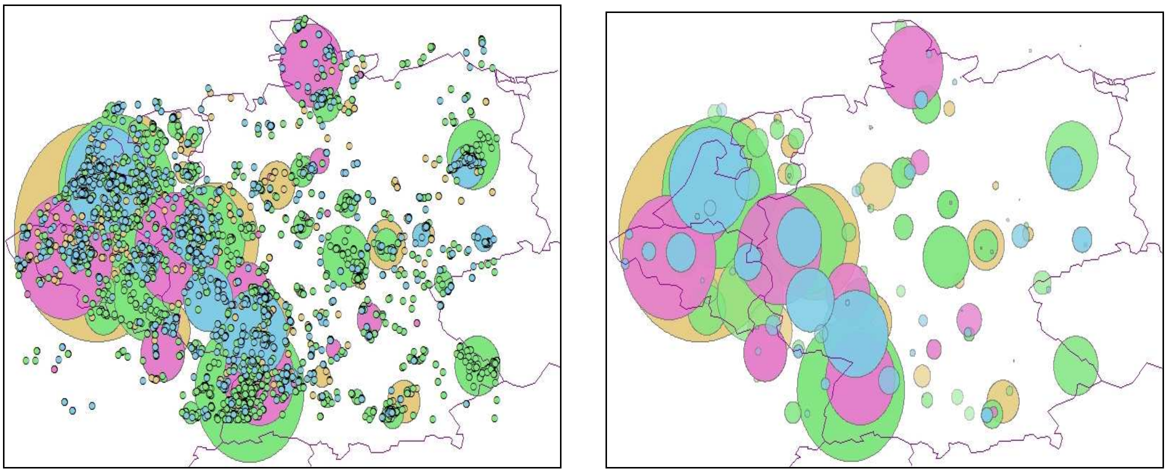


Figure A_M1: Multiple data points converted to polygon using circle tool

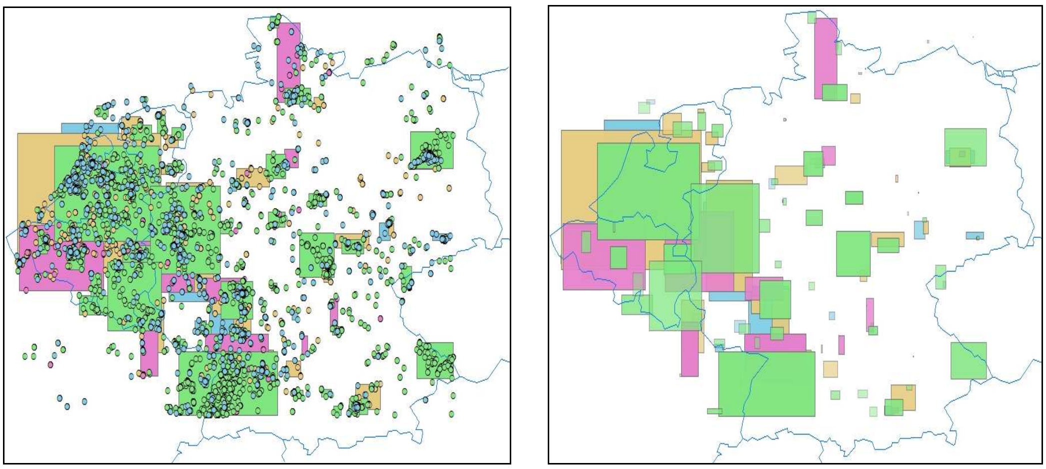


Figure A_M2: Multiple data points converted to polygon using extent tool

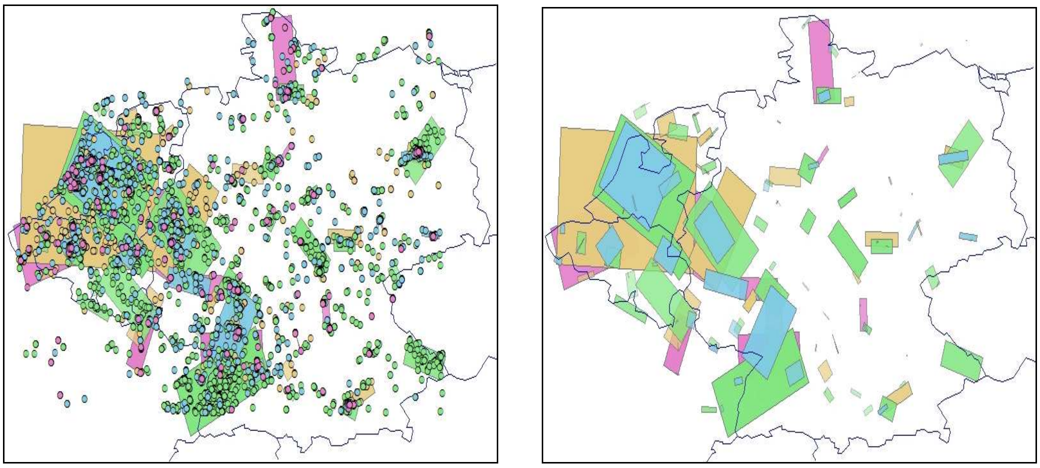


Figure A_M3: Multiple data points converted to polygon using rectangle tool

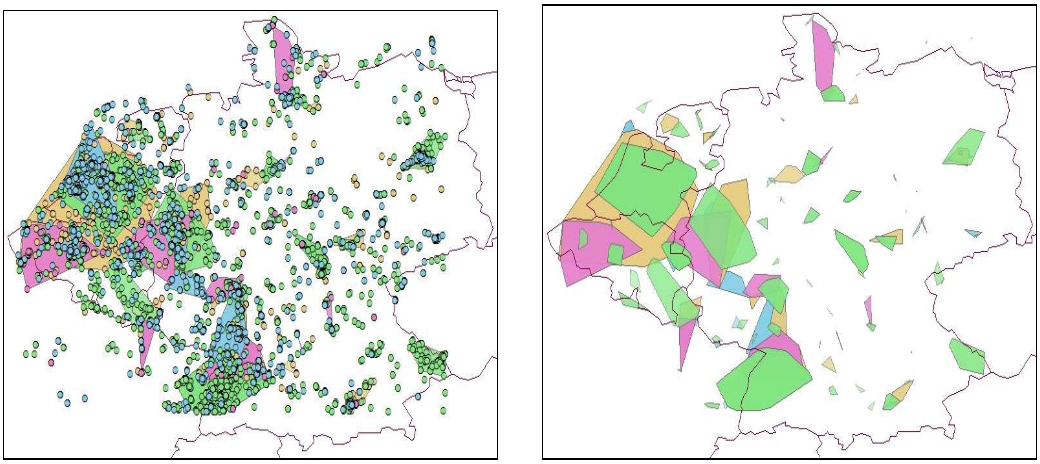


Figure A_M4: Multiple data points converted to polygon using convex hull tool

APPENDIX B: Cartographic maps

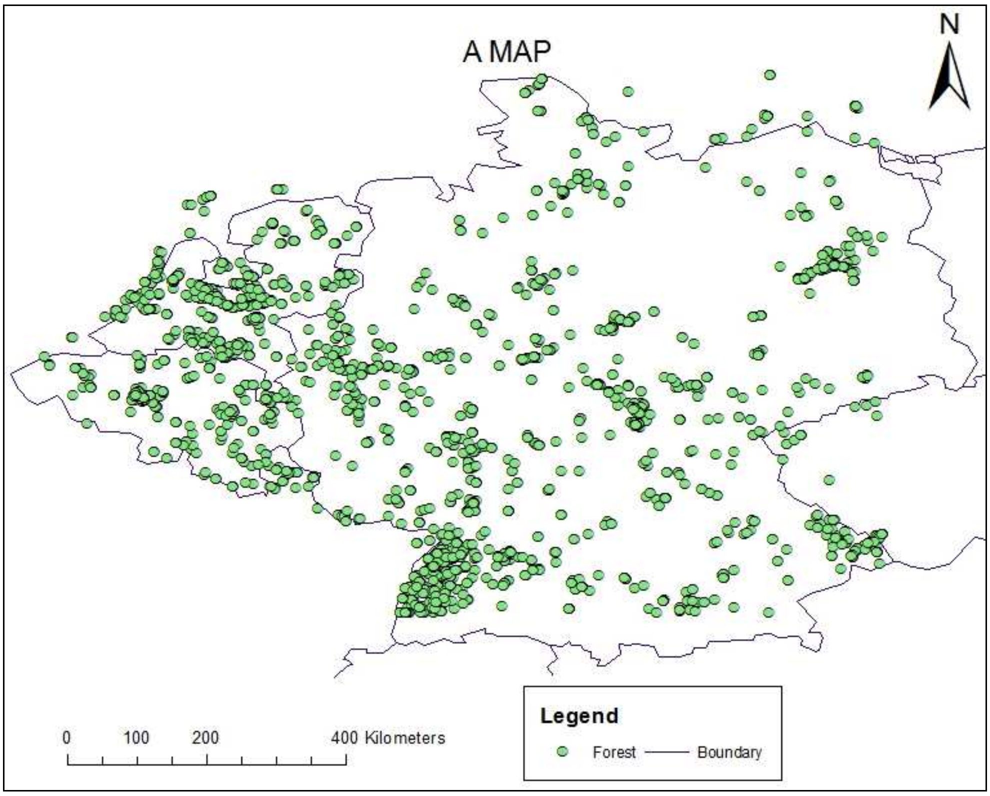


Figure B1_a: Neogeography forest topic displayed by changing the symbol type

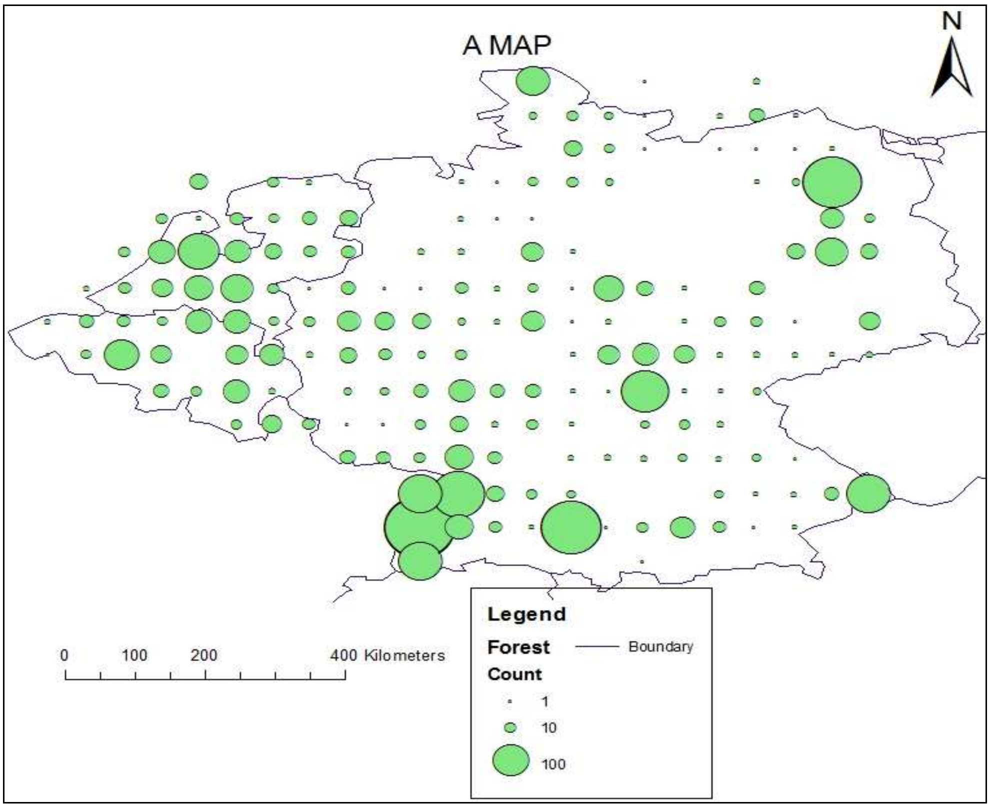


Figure B1_b: Neogeography forest topic displayed using proportional symbols

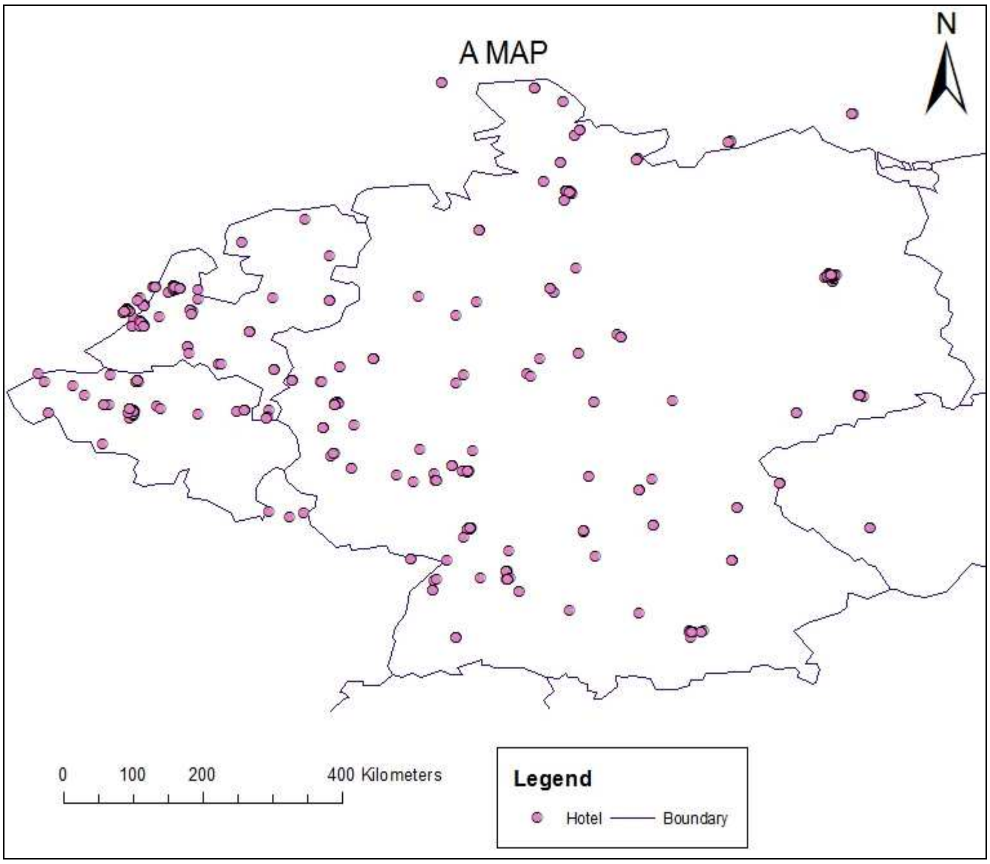


Figure B2_a: Neogeography hotel topic displayed by changing the symbol type

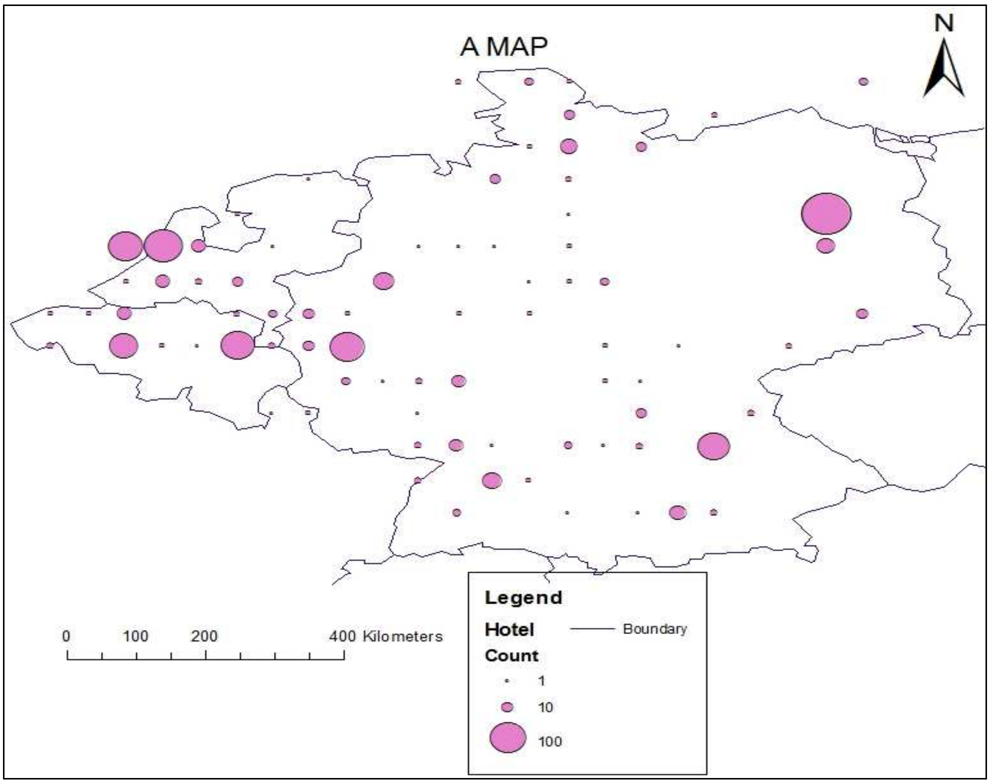


Figure B2_b: Neogeography hotel topic displayed using proportional symbol

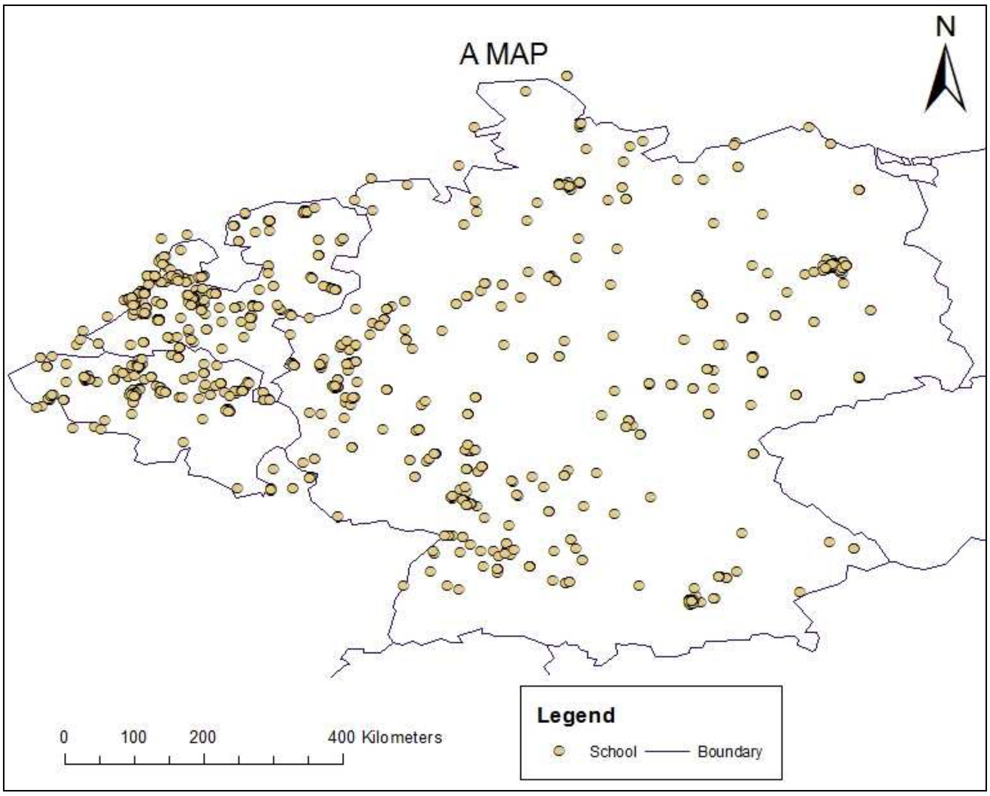


Figure B3_a: Neogeography school topic displayed by changing the symbol type

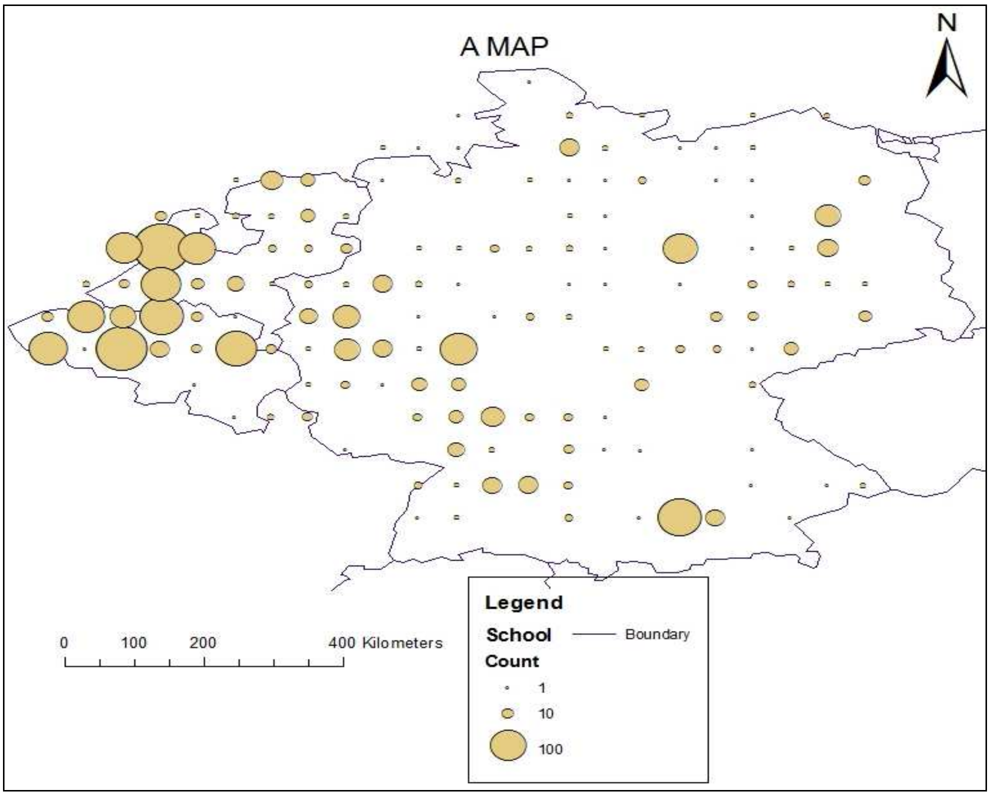


Figure B3_b: Neogeography school topic displayed using proportional symbols

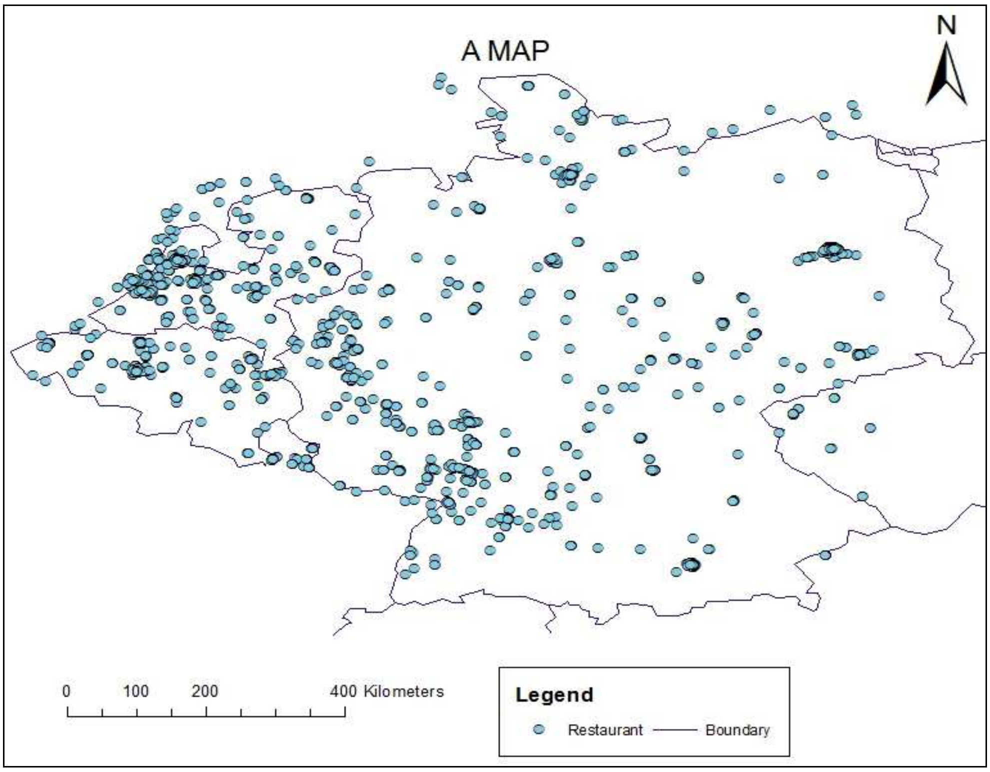


Figure B4_a: Neogeography restaurant topic displayed by changing the symbol type

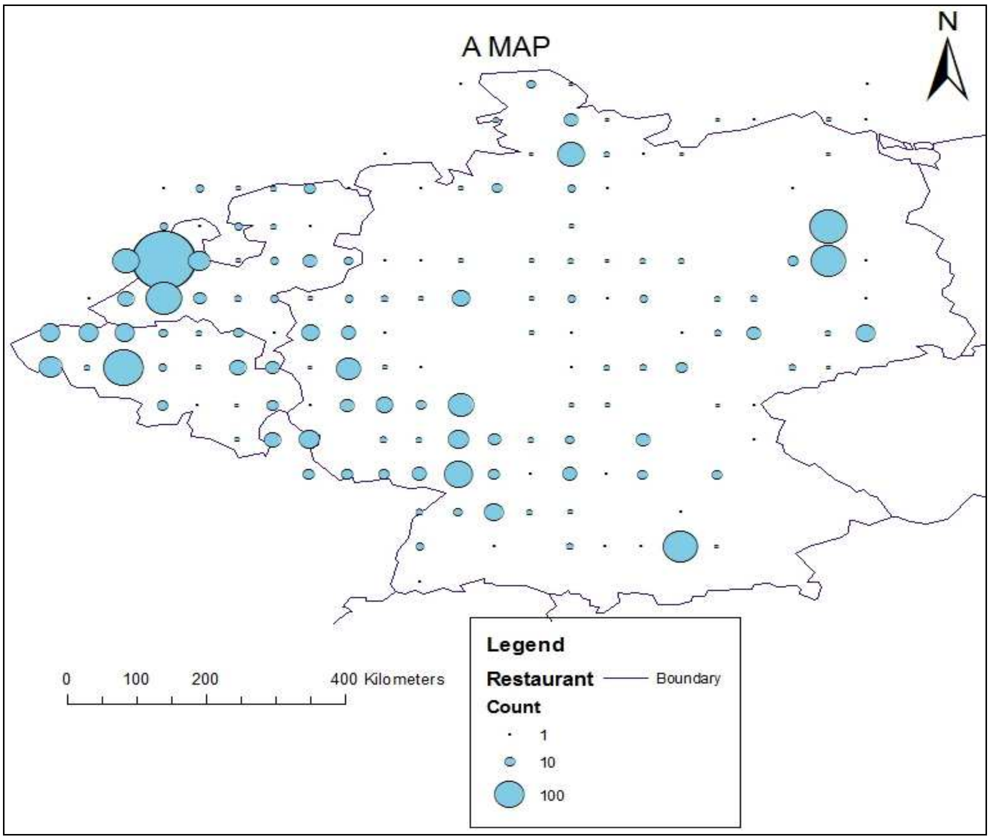


Figure B4_b: Neogeography restaurant topic displayed using proportional symbols

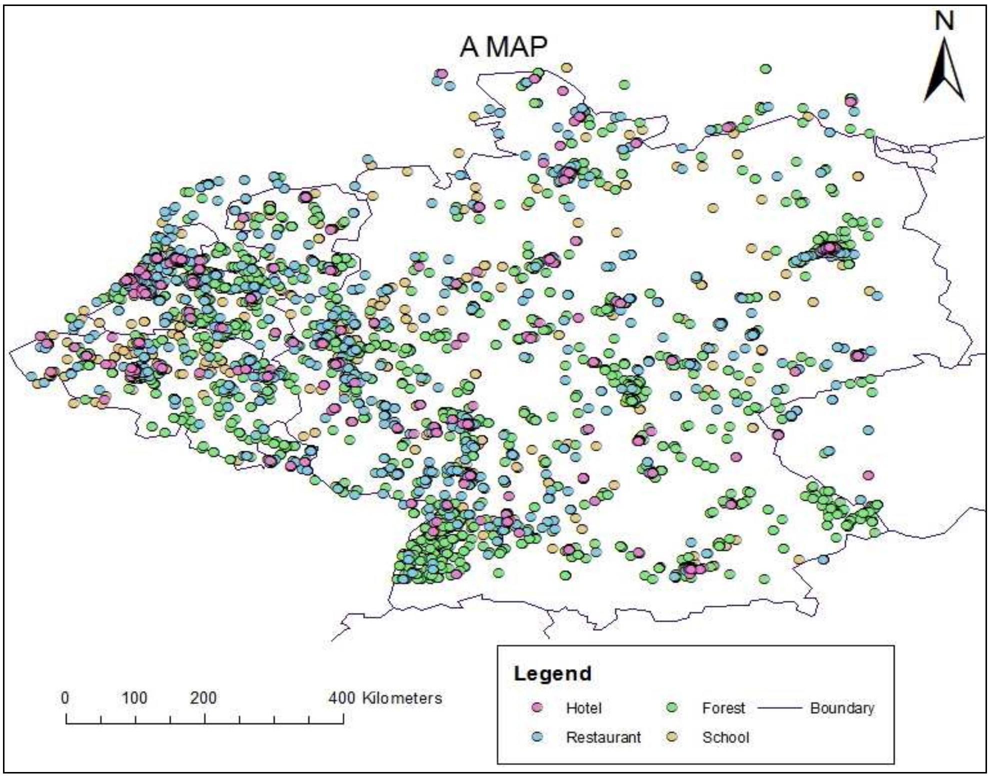


Figure B5_a: Neogeography multiple topics displayed by changing the symbol type

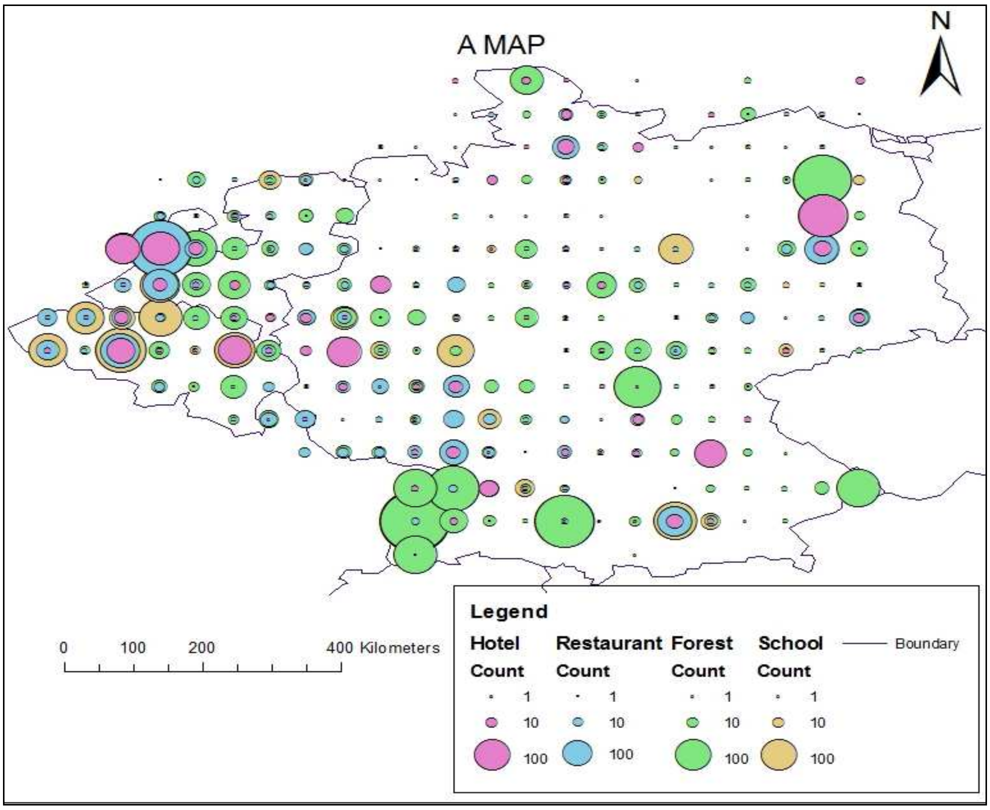
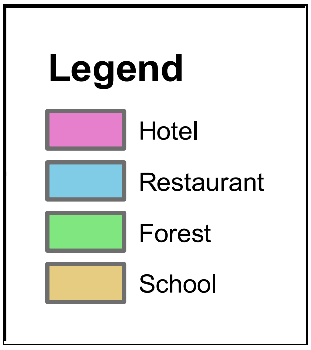


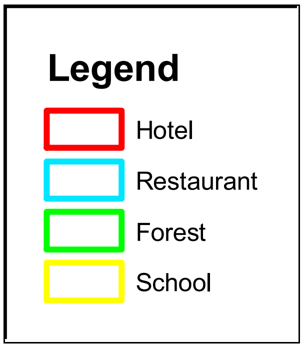
Figure B5_b: Neogeography multiple topics displayed using proportional symbols

APPENDIX C: Usability testing

Introduction

Advancement in technology made individuals to be able to collect, display, share and edit data on the available internet platforms such as Google Earth, Google maps, Flickr etc. This process is called neogeography and the associated product is called neogeography map. Generally, neogeography map is made up from multiple topics such as hotel, rivers, churches, restaurants, roads, etc. Both data and topics are increasing daily, and when displayed at small scale the output map appear to be cluttered. As these data contains useful information, and in order to use these data or maps a new approach has been suggested. The approach is to summarize the topics through clustering, and apply bounding boxes to each cluster obtained and visualize them. Since these maps are not interactive, this legend will be used as a reference when describing neogeography maps.





Major objective:

To investigate if people understand the neogeography maps and look for alternatives for representing the neogeography maps.

Tasks:

In order to accomplish the objective above, we have to perform three tasks

- To arrange the map of England (individual task)
- To select which bounding box best represents the neogeography map
- To find symbology to represent the data
- To discuss the functionality of neogeography interface

- **Arranging the map of England (individual task)**

Arrange the neogeography maps, summarized maps and the cartographic maps of England (Map A, B, C, D, E, F, G and H) in the order such as best, good, bad and worse, in terms of good design and communication. (Reasons for such an arrangement is required). Use number as shown below to rank the maps and number repetition is allowed. Map represented without a legend are the neogeography maps as they were supposed to be interactive. However the color and shape remains the same as it on the maps having the legend.

1. Best
2. Good
3. Bad
4. Worse

- **Selection of bounding boxes and discussing the interface**.

The neogeography multiple topics shown in figure 1 have been clustered using the DBSCAN method and four bounding boxes namely circle, extent, rectangle and convex hull has been applied as shown in figure 2 (a) – 2(d). Figure 4, 5 and 6 are cartographic maps drawn using the same data

Question 1: Comparison of bounding boxes with the neogeography map

(a) Which bounding box (geometry) represents best neogeography map?

(b) Are the bounding boxes dependent on data distribution?

Question 2: Comparison of bounding boxes with the cartographic map

(a) Which bounding box (geometry) represents best cartography map?

(b) Is there any difference between 1(a) and 2(a) and Why?

Question 3: Using the interface

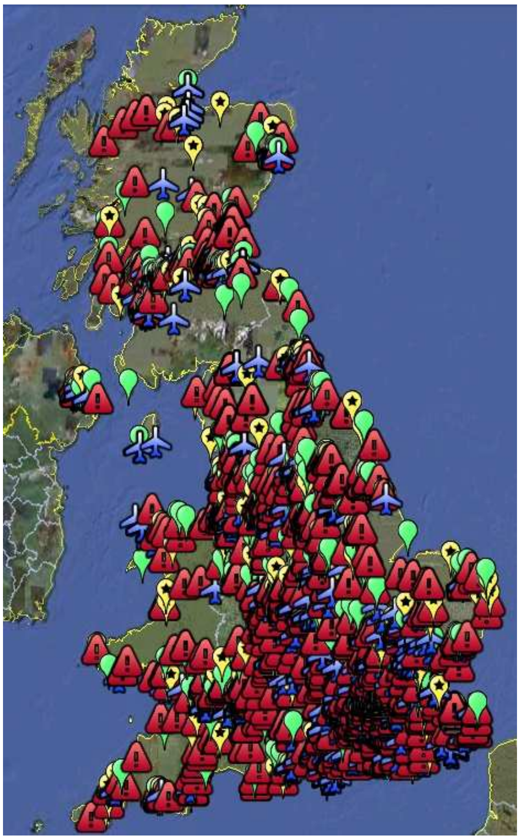
Would it be helpful to use the interface in producing this kind of maps as shown in figure 2(a) – 2(d).

- **Symbology to represent the data**

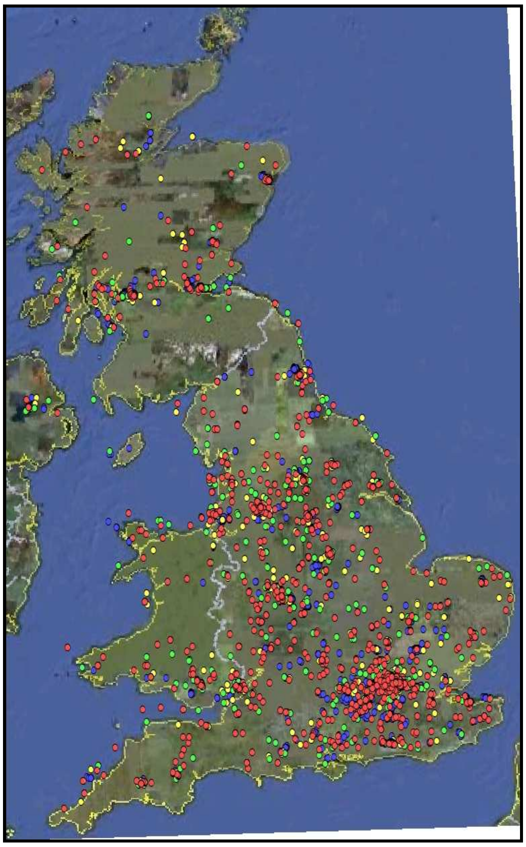
Figure 3(a) – 3(c) is one of the summarized maps using convex hull tool in bounding the cluster of points. Different visualization conditions such as colored polygons, transparency and full outline color has been applied.

Question 4: (a) Which visualization condition does relate the information best? (Colored polygons, transparency or full outline), which way(s) of presenting the map will you prefer?

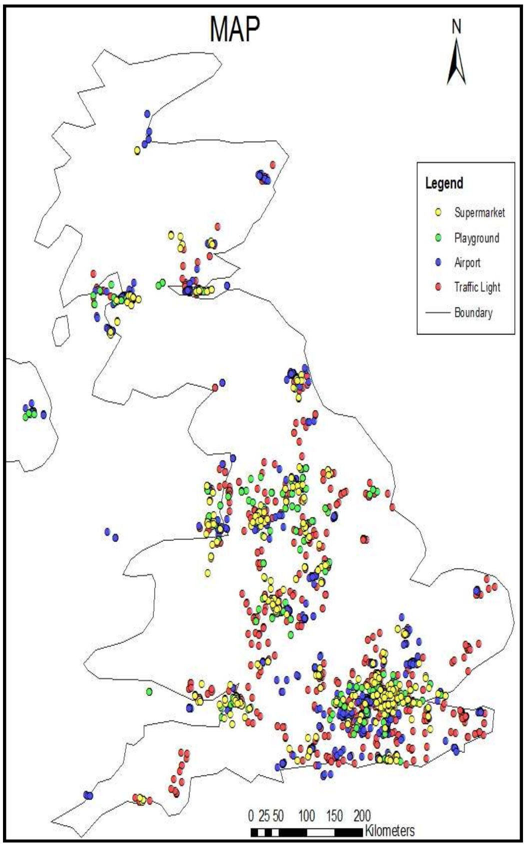
(b) If you are a tourist and you are planning to survey for the forested places, which map will you use? Give reason(s)



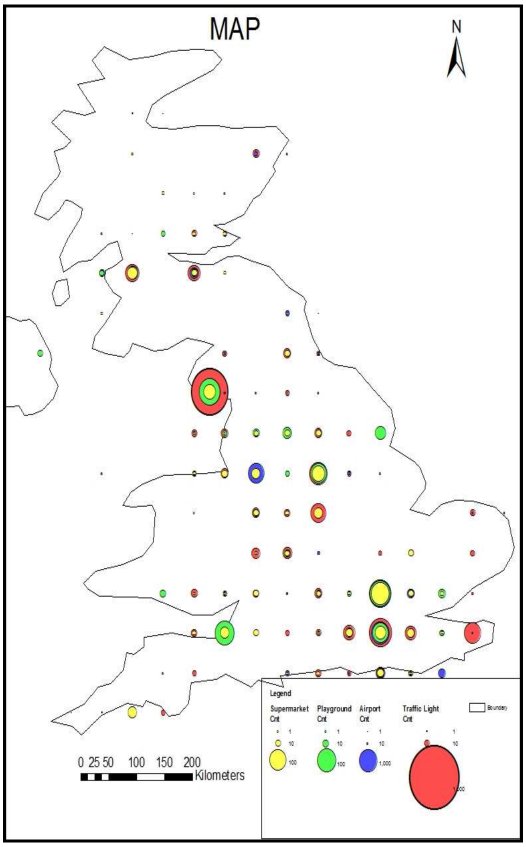
Map A



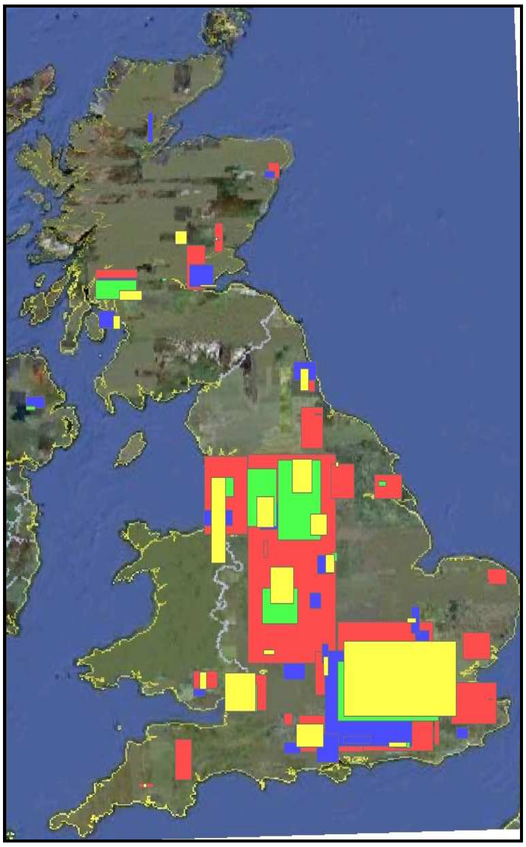
Map B



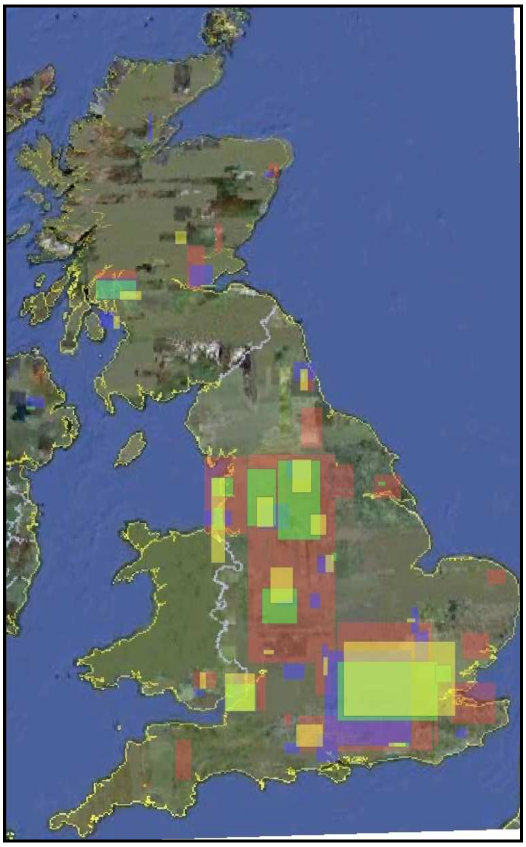
Map C



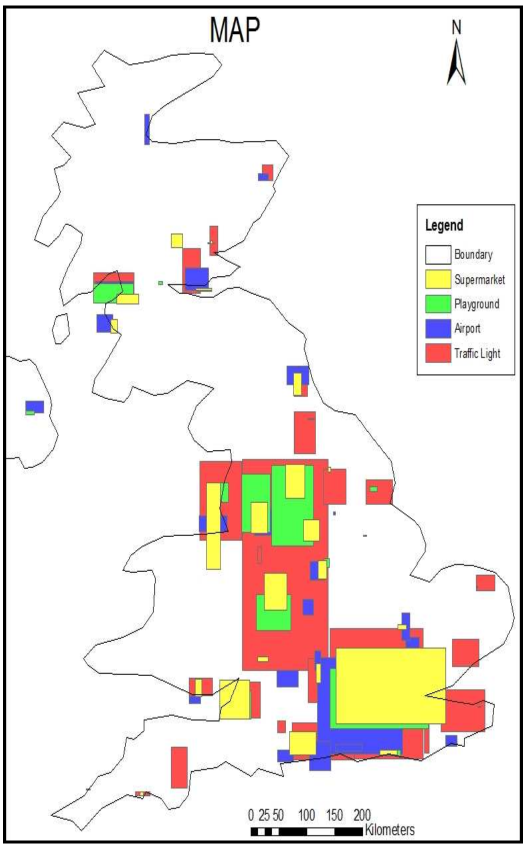
Map D



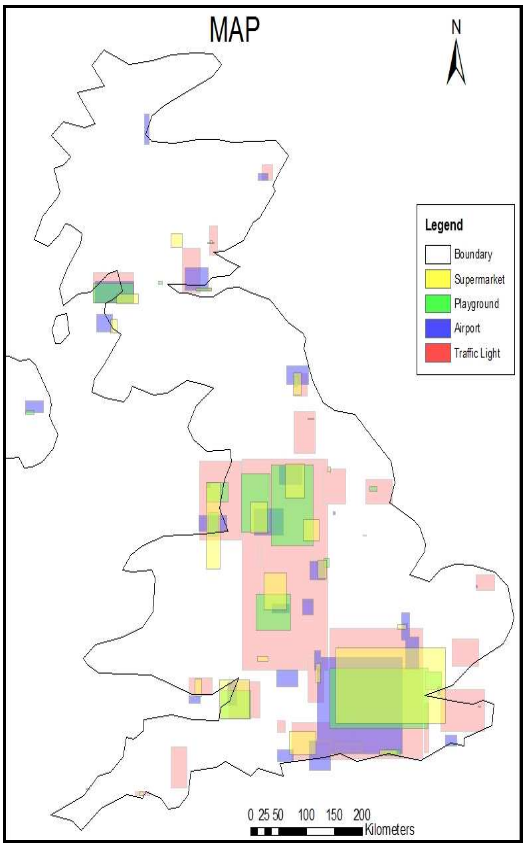
Map E



Map F



Map G



Map H

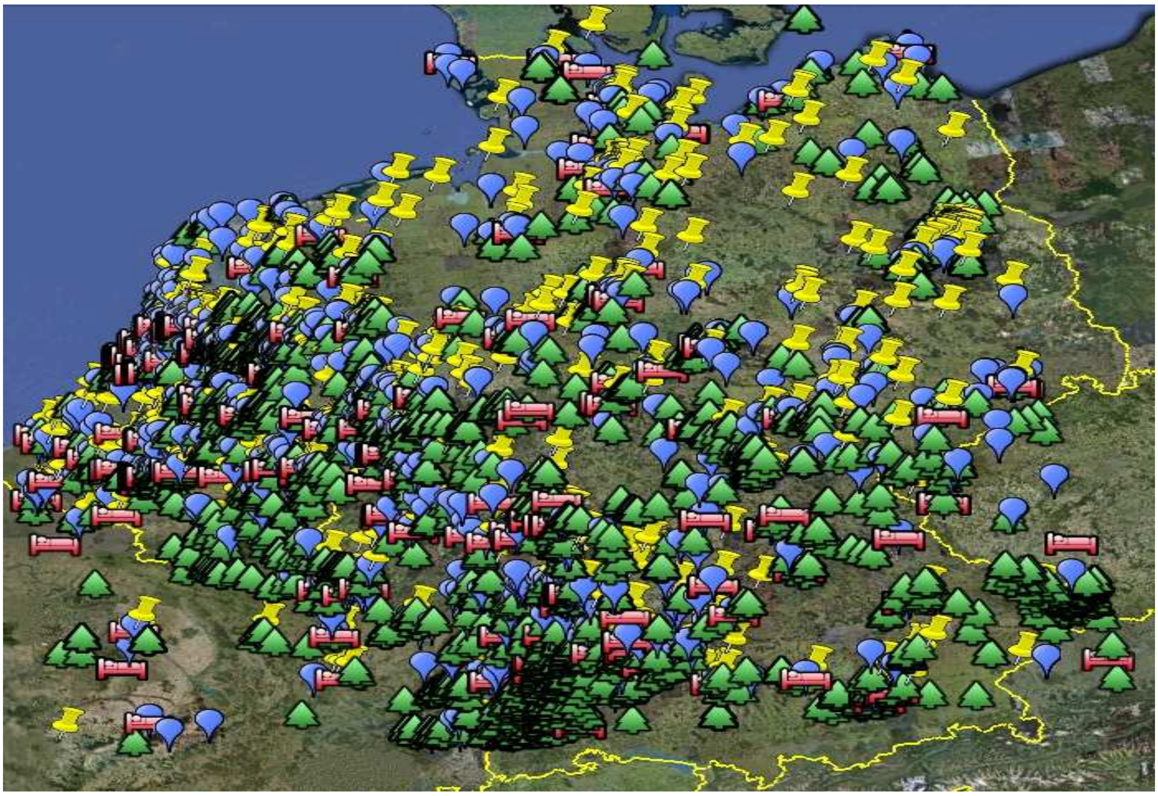


Figure 1: Neogeography map with 4 topics (hotel, restaurant, forest and school)

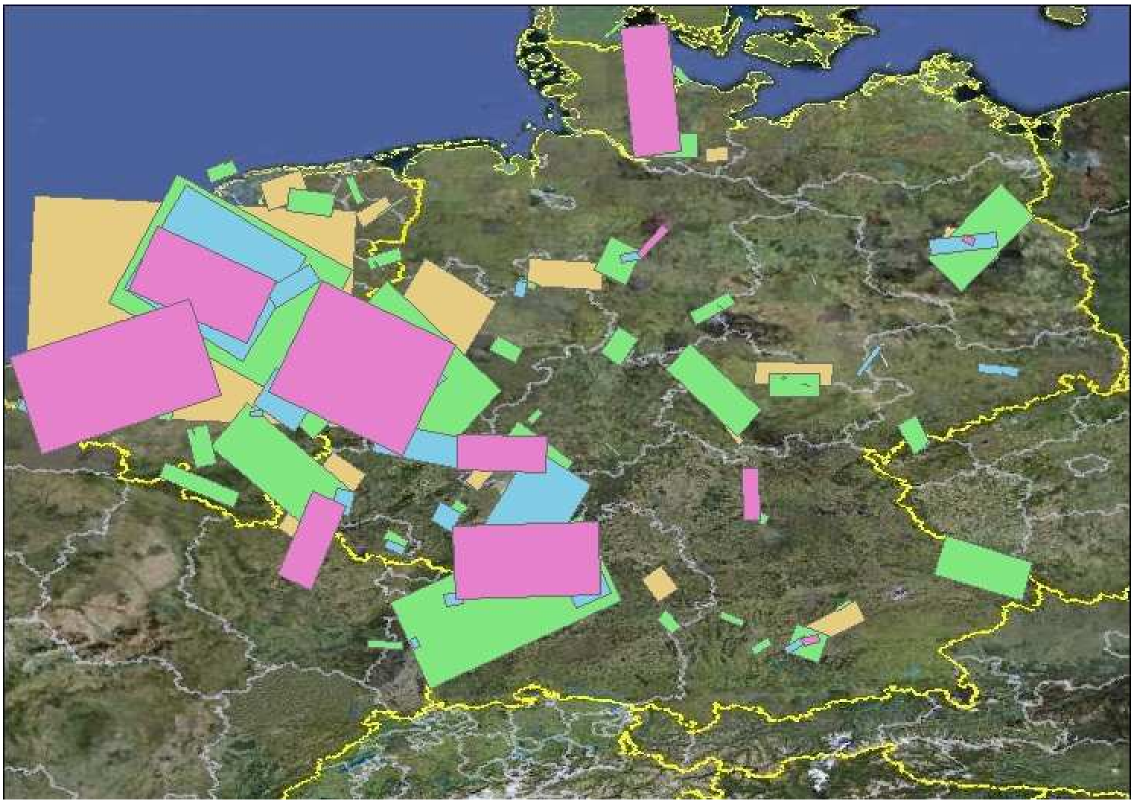


Figure 2 (a): Rectangle tool

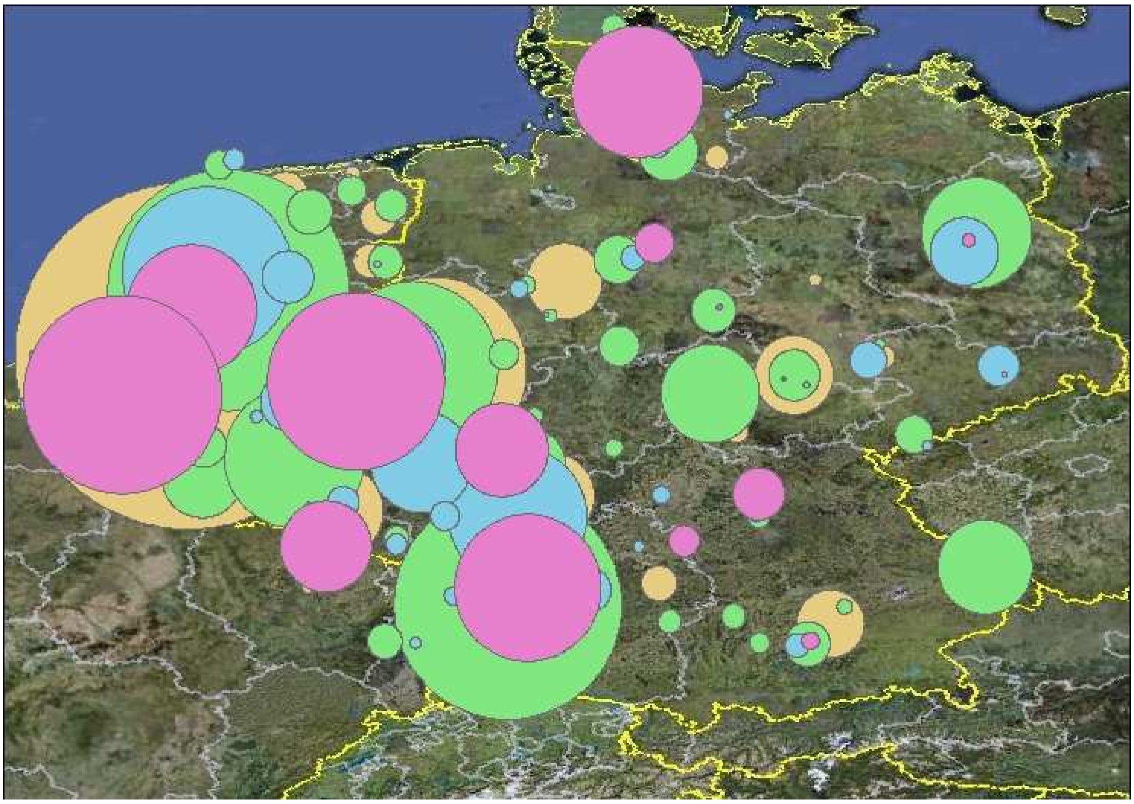


Figure 2(b): Circle tool

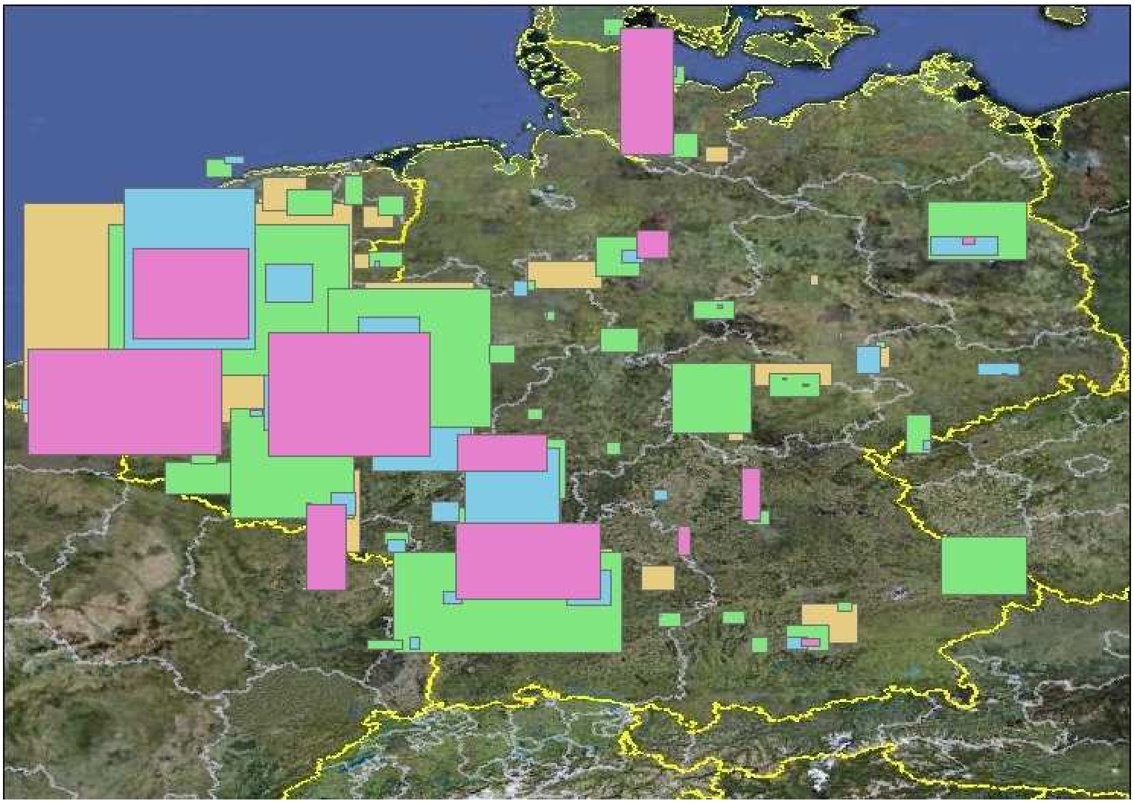


Figure 2(c): Extent tool

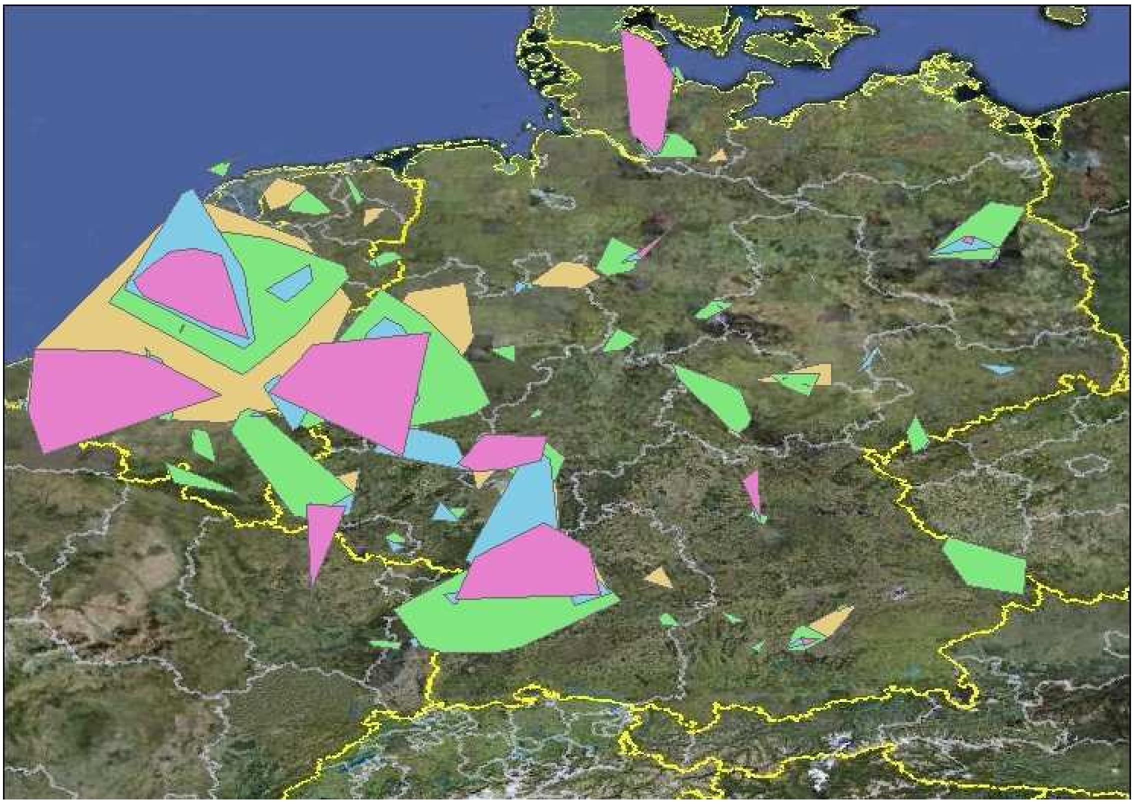


Figure 2(d): Convex hull tool

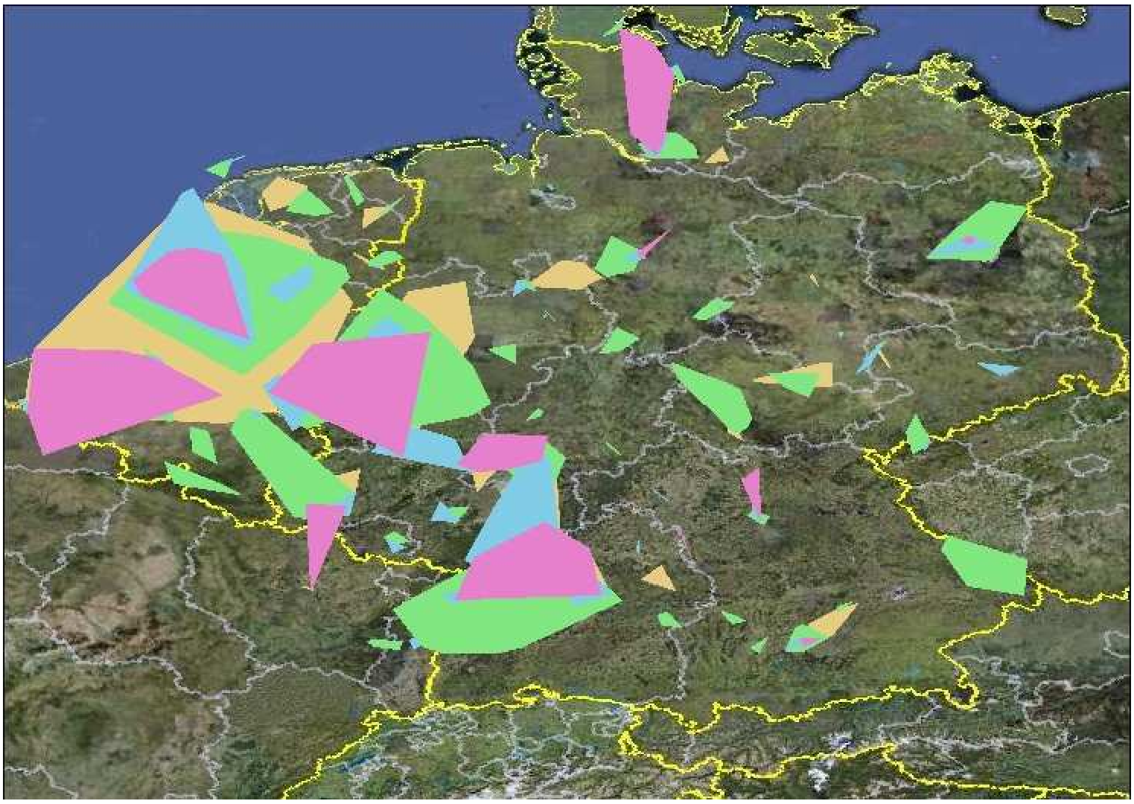


Figure 3 (a): Colored polygons

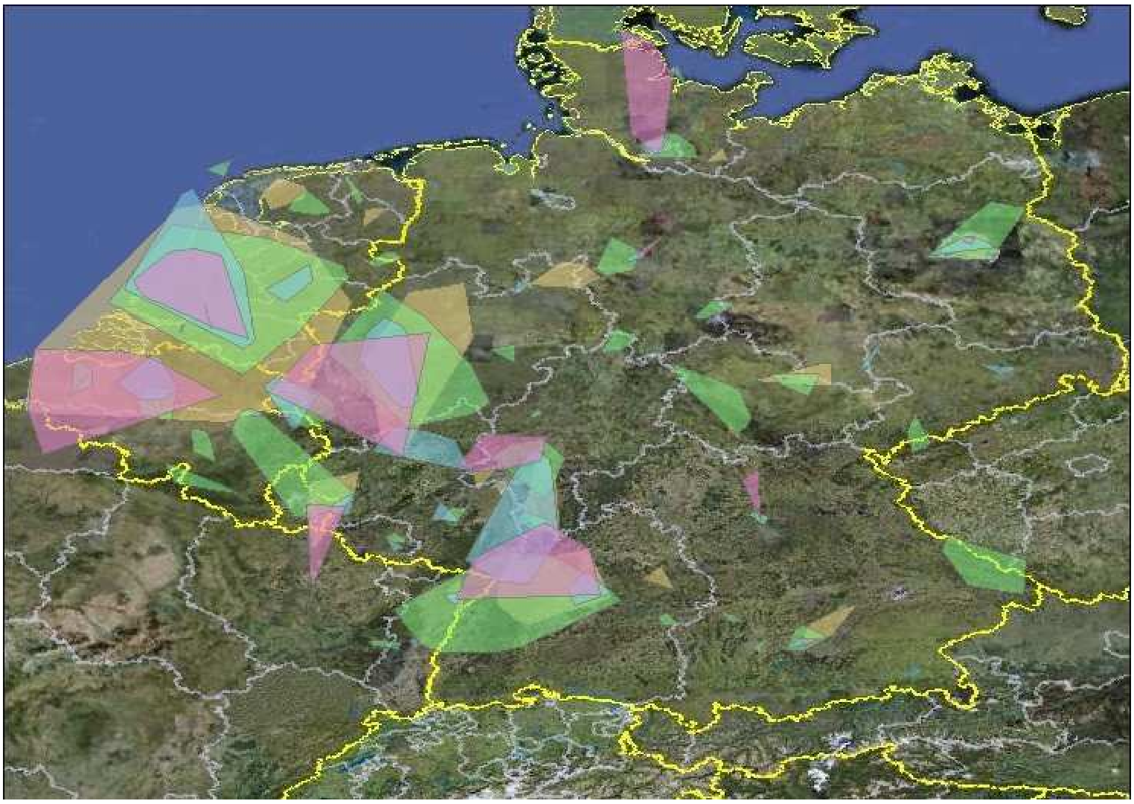


Figure 3 (b): Transparency

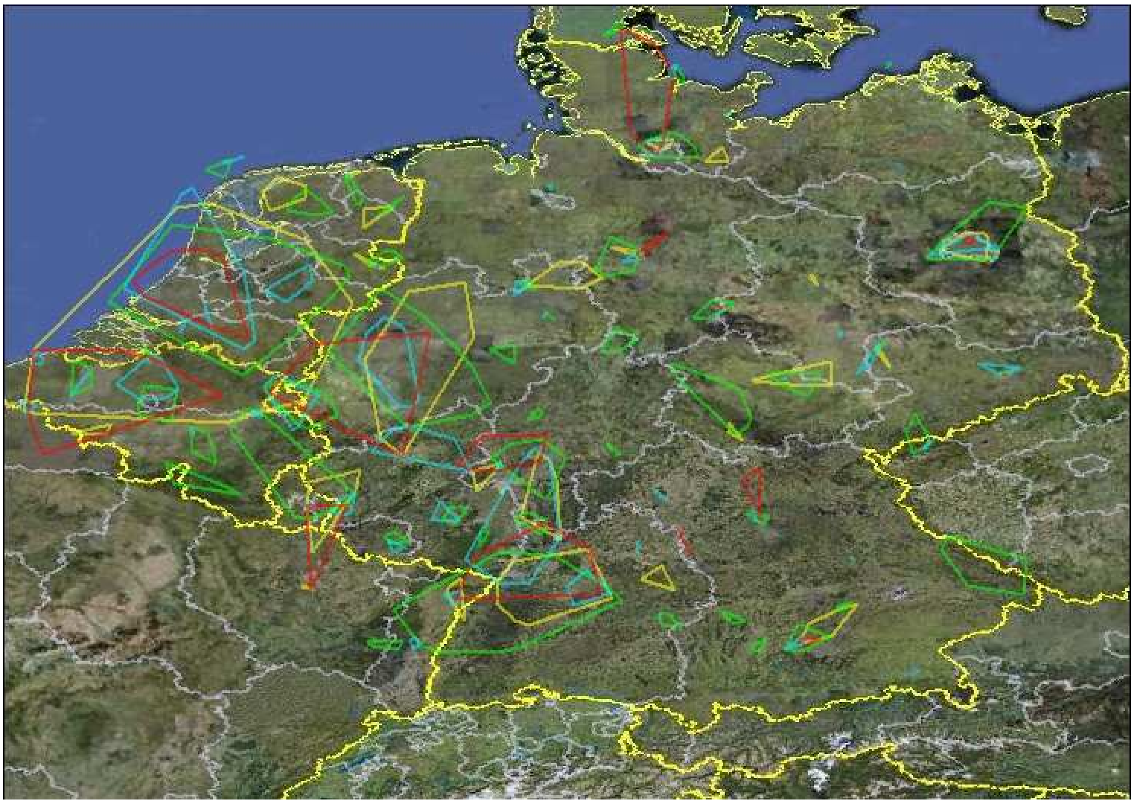


Figure 3 (c): Outline colors

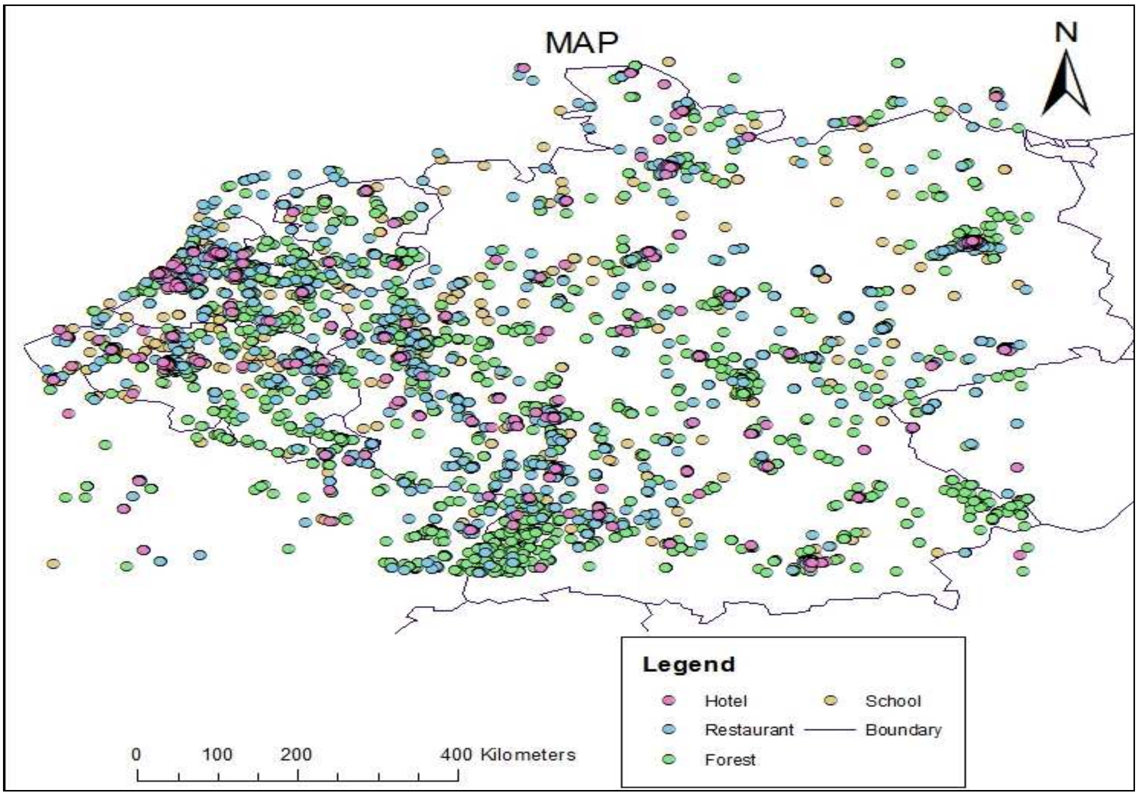


Figure 4

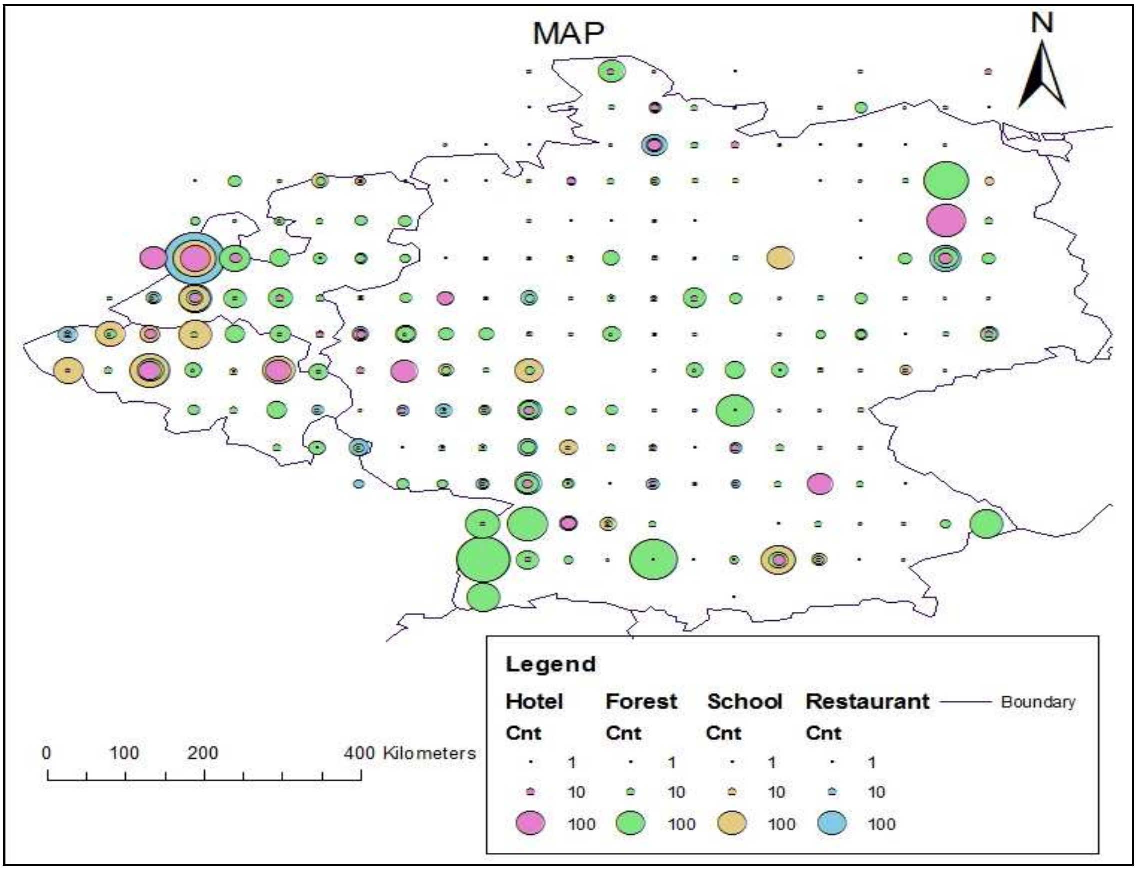


Figure 5

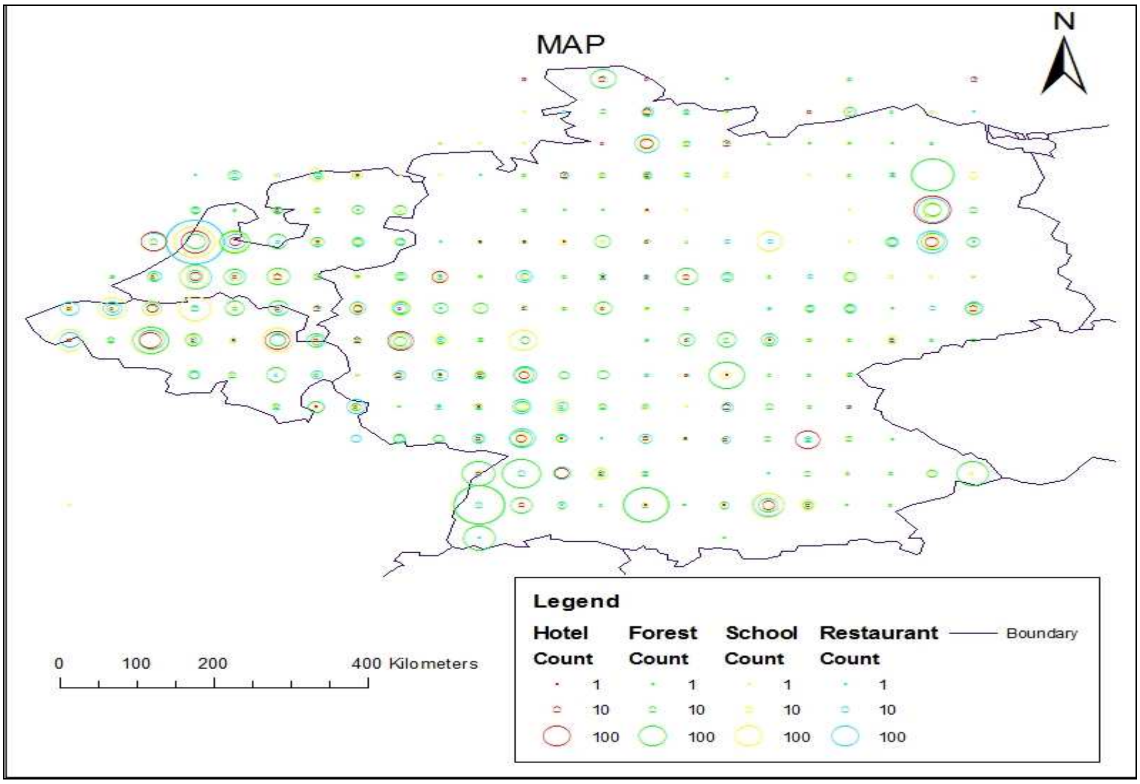


Figure 6

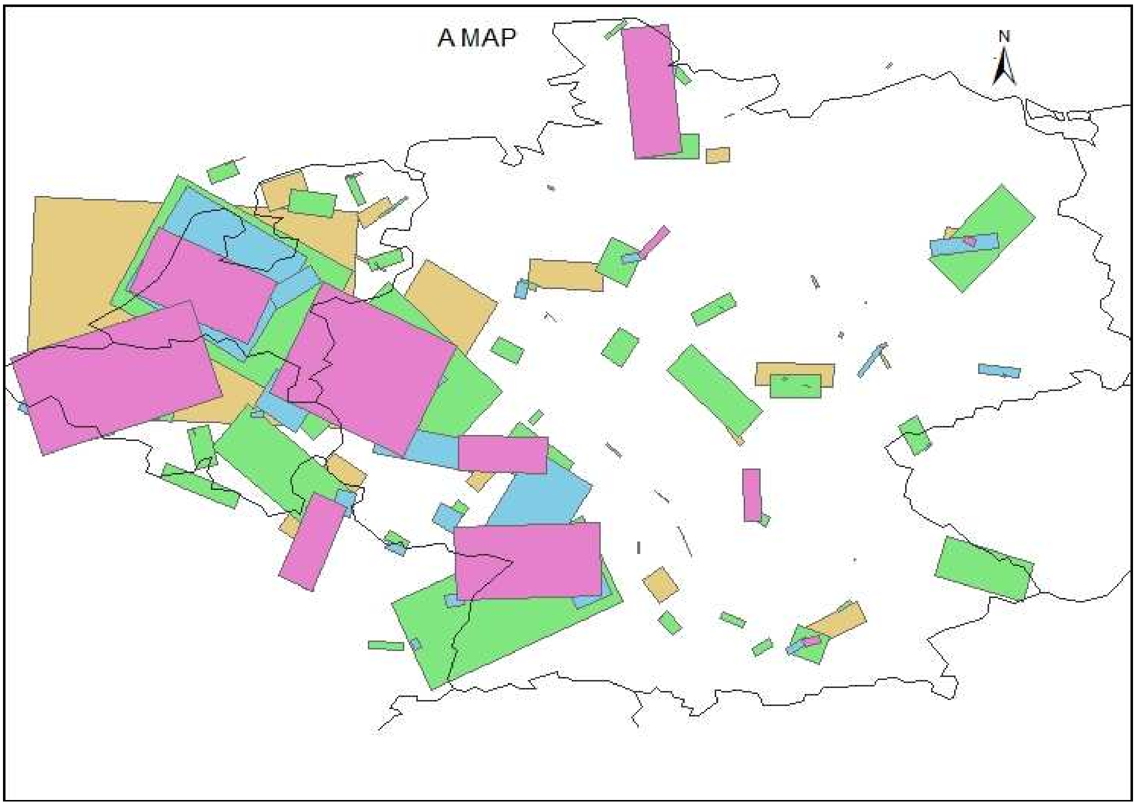


Figure 2(d): Convex hull tool

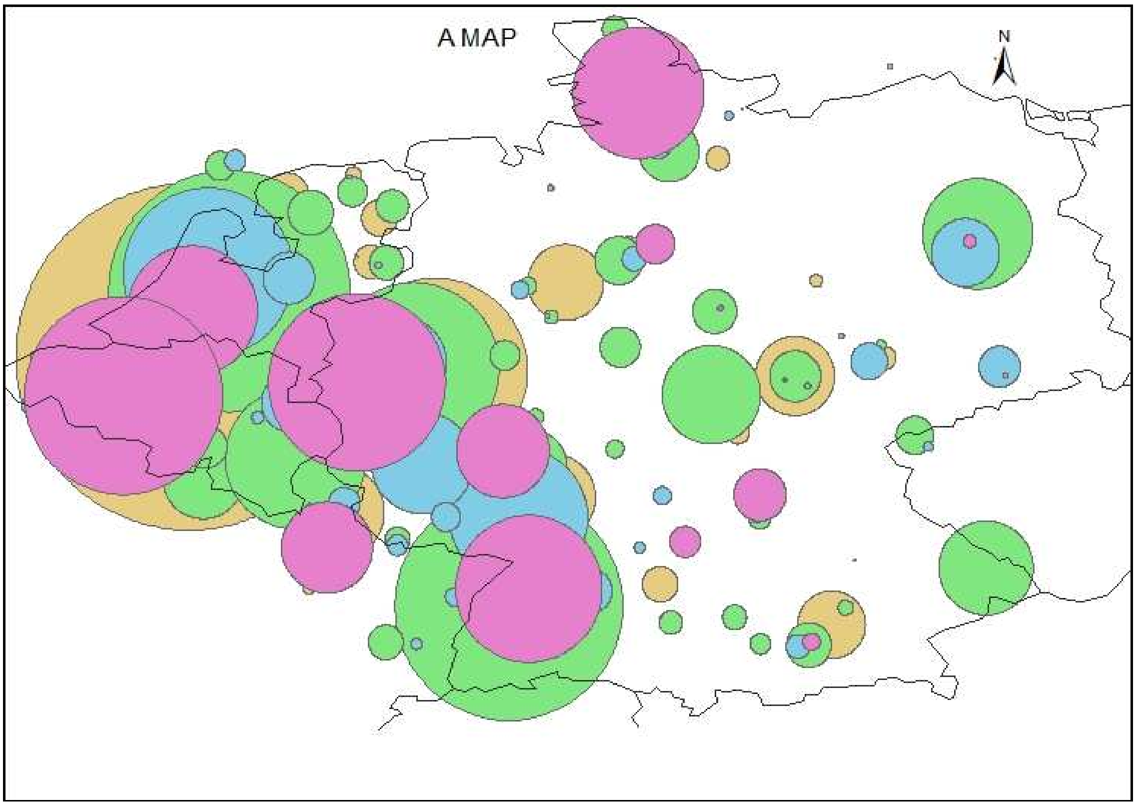


Figure 2(d): Circle tool

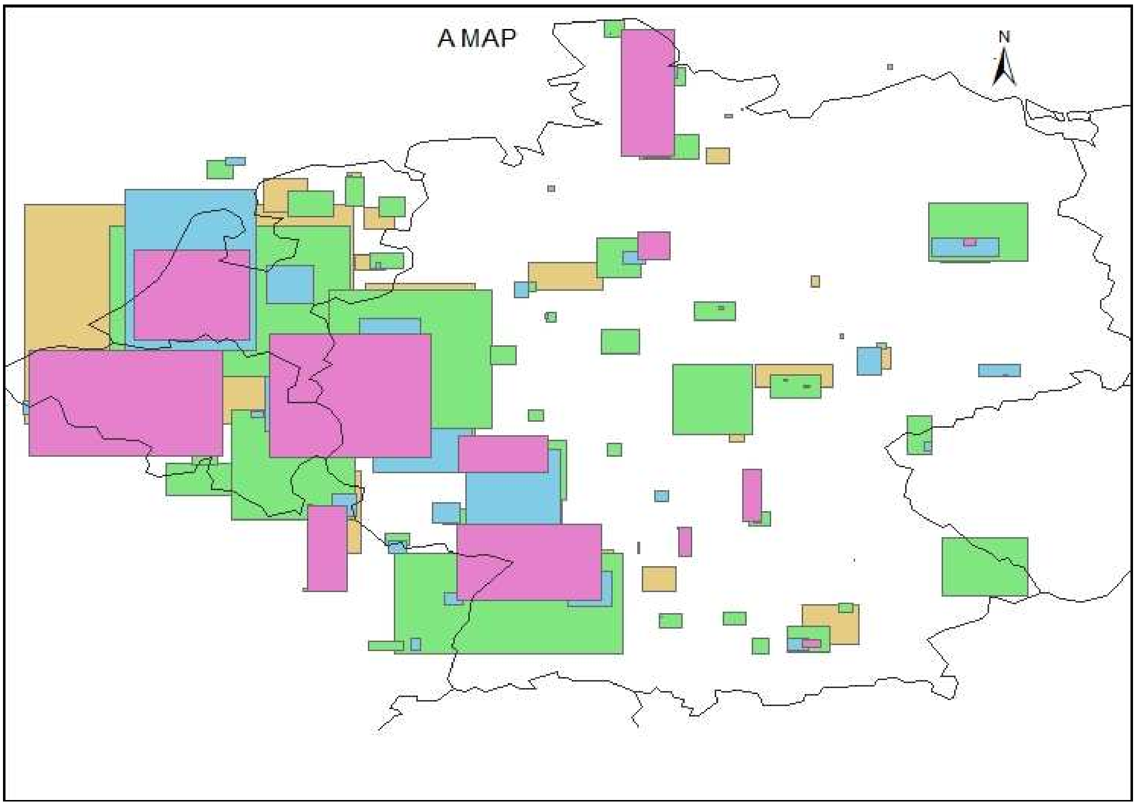


Figure 2(d): Extent tool

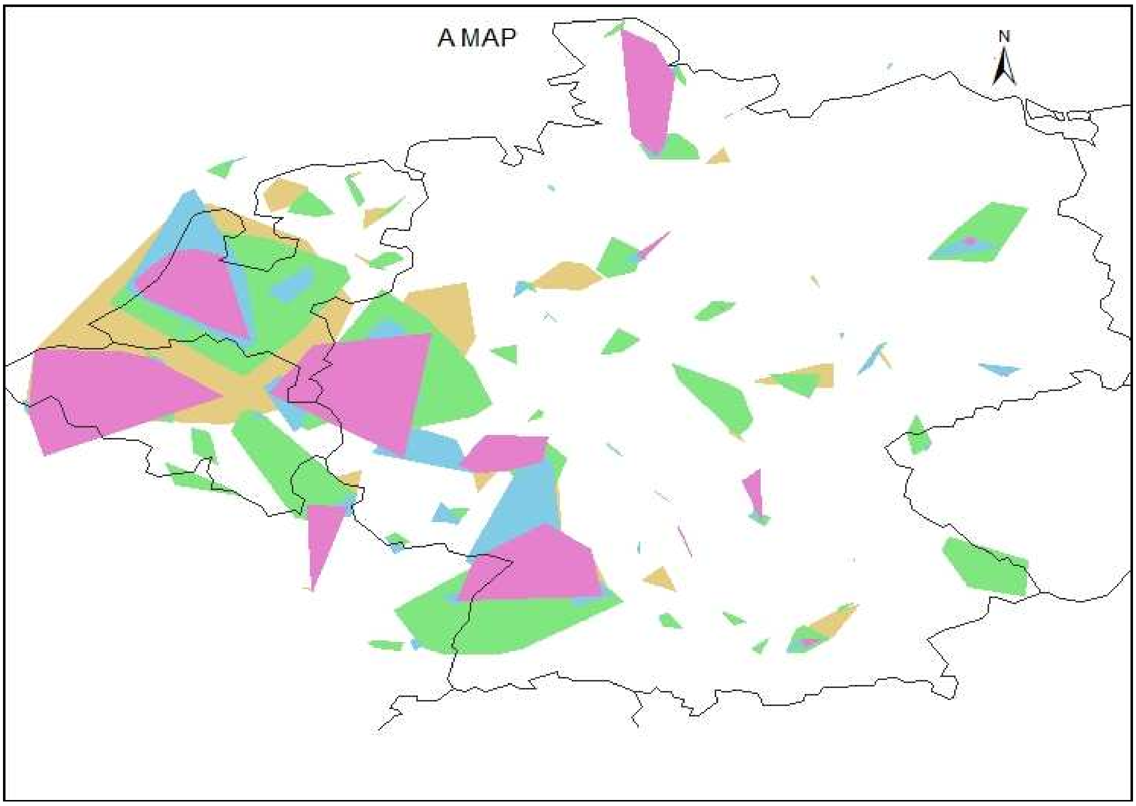


Figure 2(d): Convex hull tool

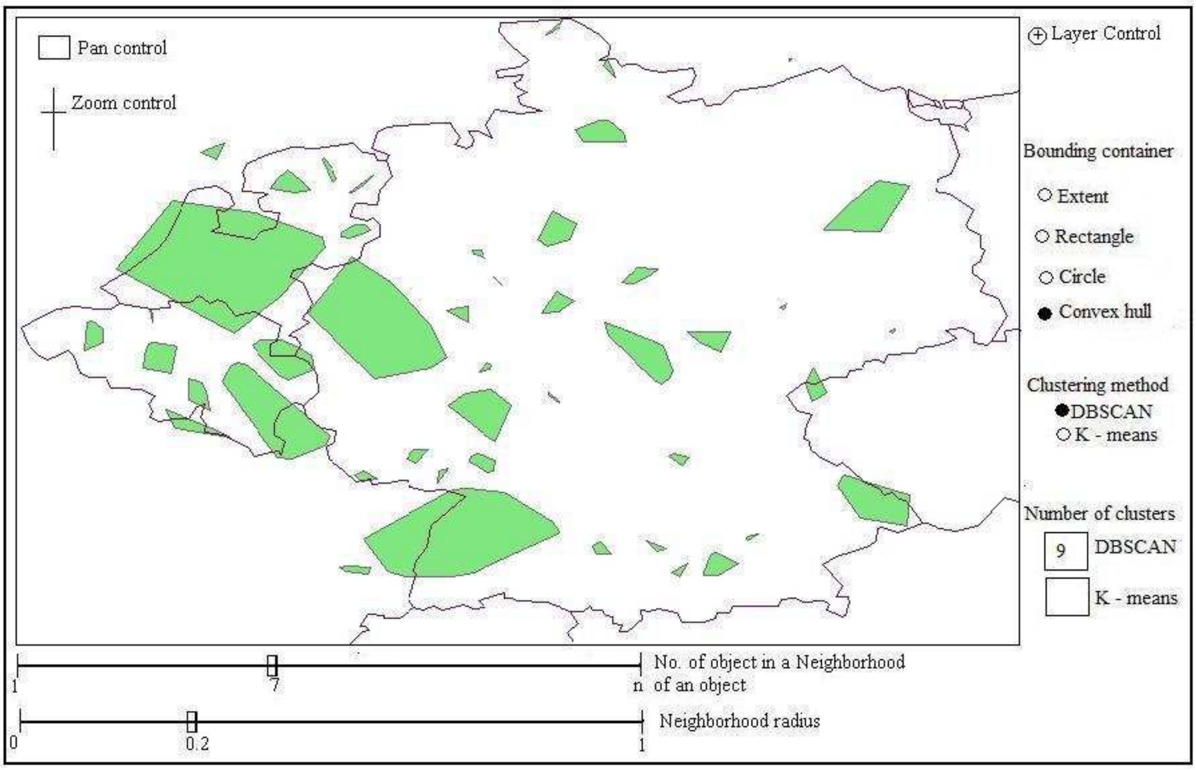


Figure 7: Cartographic interface with white background

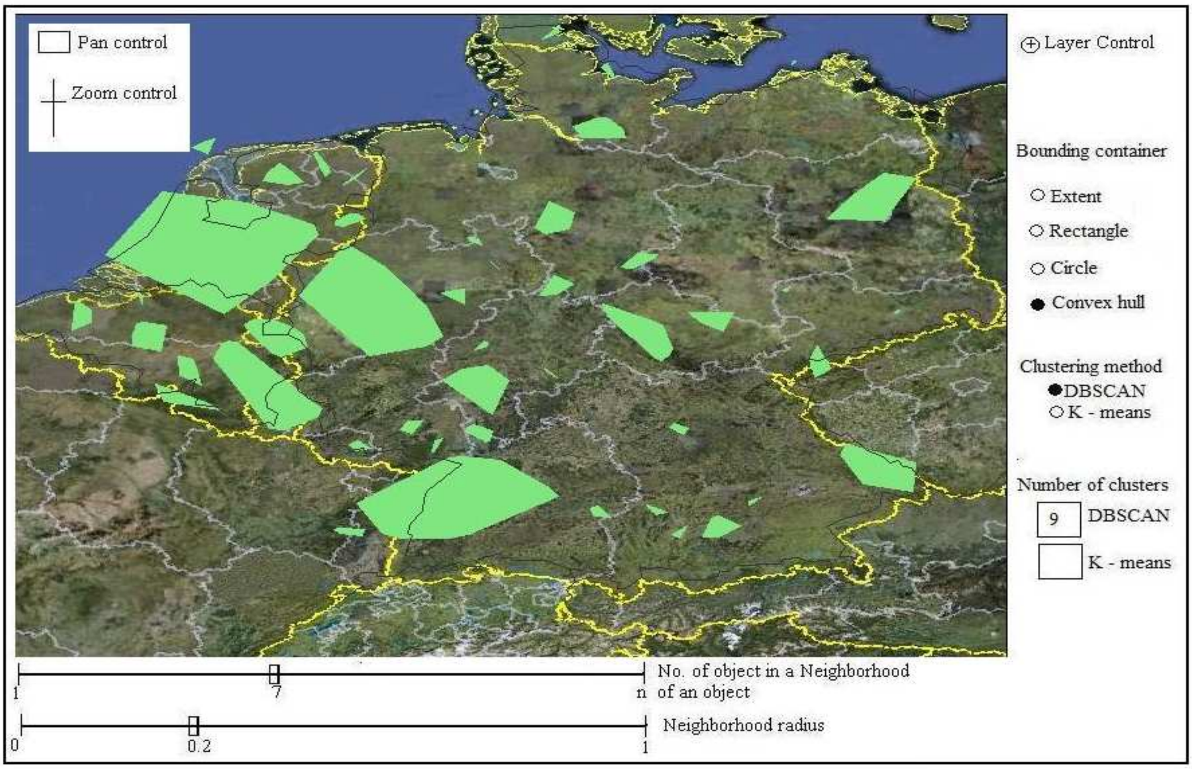


Figure 8: Neogeography interface

APPENDIX D: Python code for extracting data

Python code for extracting data from text format (txt) to excel format (csv)

```
#function to read file
def readfile(filename):
    f=open(filename)
    data=[]
    while 1==1:
        line=f.readline()
        if line=="":
            break
        else:
            data.append(line)
    f.close()
    return data

def writefile(outdatalines,ofile):
    # outdatalines is a list of lines of strings
    try:
        outfile=open(ofile,'w')
        for foo in range(len(outdatalines)):
            outdataline=str(outdatalines[foo])+'\n'
            outfile.write(outdataline)
        outfile.close()
        return
    except IOError:
        print "Sorry! ",ofile, "not well specified!"
        return

def extract_data(data_list):

result_list=["photo,owner,title,originalsecret,originalformat,o_height,o_width,datetaken,dateupload,ta
gs,license,latitude,longitude,accuracy, views, interestingness"]
    clean_line=""
    for i in range(len(data_list)):
        line=data_list[i].strip()
        #convert all commers to -
        for j in range(len(line)):
            if line[j]==",":
                line=line[:j]+"-"+line[j+1:]
        #put a commer after every item
        if line[:6]=="photo:":
            clean_line+=line[7:]+","
        elif line[:6]=="owner:":
```

```
            clean_line+=line[7:]+","
        elif line[:6]=="title:":
            clean_line+=line[7:]+","
        elif line[:15]=="originalsecret:":
            clean_line+=line[16:]+","
        elif line[:15]=="originalformat:":
            clean_line+=line[16:]+","
        elif line[:9]=="o_height:":
            clean_line+=line[10:]+","
        elif line[:8]=="o_width:":
            clean_line+=line[9:]+","
        elif line[:10]=="datetaken:":
            clean_line+=line[11:]+","
        elif line[:11]=="dateupload:":
            clean_line+=line[12:]+","
        elif line[:5]=="tags:":
            clean_line+=line[6:]+","
        elif line[:8]=="license:":
            clean_line+=line[9:]+","
        elif line[:9]=="latitude:":
            clean_line+=line[10:]+","
        elif line[:10]=="longitude:":
            clean_line+=line[11:]+","
        elif line[:9]=="accuracy:":
            clean_line+=line[10:]+","
        elif line[:6]=="views:":
            clean_line+=line[7:]+","
        elif line[:16]=="interestingness:":
            clean_line+=line[17:]+","
            result_list.append(clean_line)
            clean_line=""
    return result_list

input_file_name=raw_input("InputFile:")
data_from_file=readfile(input_file_name+".txt")
data_list=extract_data(data_from_file)
writefile(data_list,input_file_name+"_out.csv")
```