GIMA

Geographical Information Management and Applications

Making sense of spatial perceptions

Analysing PPGIS data for planning purposes

Author: Michiel Prak

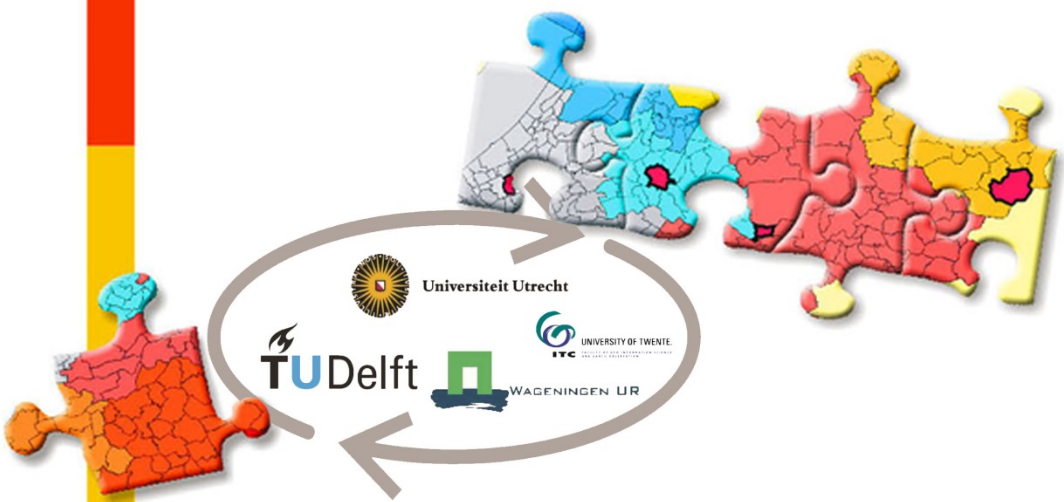
Supervisors: Fred Toppen

Eduardo Dias

Simeon Nedkov

Professor: Stan Geertman

March, 2014



Making sense of spatial perceptions

Analysing PPGIS data for planning purposes

MSc GIMA Thesis

Contact information

Student:	M.E. Prak
UU#:	3253333
ITC#:	s6003850
Email:	michiel.prak@gmail.com
Phone:	+31 6 15357398
Address:	M.P. Lindostraat 41-bis
	3532 XG Utrecht
	The Netherlands

Supervisors

Professor:	dr. S.C.M. Geertman (Stan)
UU:	drs. F.J. Toppen (Fred)
VU:	dr. E.S. Dias (Eduardo)
	ir. S.B. Nedkov (Simeon)

Summary

This project aims to address a hiatus in the academic literature on the analysis of data collected with the use of Public Participation Geographical Information Systems. The research question was: *How can geographical data gathered among citizens through the use of PPGIS be turned into information that is valuable in the spatial planning process?* PPGIS is defined as the intention of providing all stakeholders in spatial decision-making processes with equal access to the data and analytical capabilities provided through geographical information technology. As such, the role of PPGIS in spatial planning processes can be seen as a form of communication technology. One of the main obstacles in this exchange of information is the variation in knowledge of the stakeholders on the spatial issue at hand. The focus in this research is on the analysis of the experiences of citizens for use in participatory spatial planning processes.

First, a structured approach for the analysis of PPGIS data was developed. Appropriate analytical tools were assigned to each section of the approach. For the empirical part of the project data was collected on the experiences of citizens in the area of Hoensbroek in Heerlen. A PPGIS application was developed for data collection through use of a spatial survey. The issues addressed in the survey were decided upon in cooperation with the spatial planning department of the Heerlen municipality. The method for the analysis of PPGIS data was then applied to the collected data. As the data in this research was collected in the form of polygons instead of points, a procedure was also developed to prepare the data in such a way that the analyses could yield meaningful results. In the end the PPGIS cycle was completed almost in its entirety, with the development of a PPGIS application and associating questionnaire, the collection of data among citizens and the translation of that data into information through spatial analysis and visualization.

Representatives of the spatial planning department of the Heerlen municipality reported that they found the analyses that yielded information that was readily interpretable and provided concrete directions for spatial policies especially useful, i.e. the heat maps, the cluster analysis and the conflict analysis.

Contents

List of Figures and Tables

Figures

Tables

1. Introduction

With the rising trend of the amount of planning processes with a participatory character, the role of citizens in planning practice has changed dramatically. It is becoming clear that the traditional public gatherings, which are still the main technique used to construct the participatory process, are not effective in communicating the citizens' opinions and wishes. The emergence of public participatory geographical information systems (PPGIS) has provided the general public with more possibilities to actively engage in the spatial planning of their living environment. In the development of PPGIS the focus has mainly been on methods to involve citizens in the planning process and to induce them to voice their opinions by providing data on their preferences. Little attention, however, has been given to the step of translating the data collected to information useable for policymakers. This thesis discusses various ways in which valuable information for drawing up spatial plans can be derived from data gathered among citizens through the use of PPGIS.

In this chapter the theme of the thesis is introduced as well as some key concepts that form part of the theoretical framework. After the concise overview of the context in which the research takes place, the goals of the thesis are outlined. The goals have subsequently been transposed into concrete research questions. Next, the case-study for the research is described. Finally the societal and scientific relevance of this research are addressed.

1.1 Thesis outline

1.1.1 Background

Until not too long ago, spatial policy making was a task performed almost exclusively by urban and regional planners. These planners were responsible for drawing up a comprehensive spatial design for the region that was to undergo a certain kind of intervention (Hall, 2003). The information they deemed relevant to the project the planners acquired from experts in various fields such as environmental sciences, sociology and economy so as to be able to use that information as input when drawing up plans. Those experts most often did not actively participate in drafting the spatial strategies otherwise.

In the last few decades, however, the nature of spatial planning has changed. A major difference is that diverse stakeholders have become more involved in the process of spatial policy making. This is at least partly caused by the desire of Western governments that spatial developments contribute to a sustainable society in environmental as well as economic and social terms (Geertman & Stillwell, 2003). Consequently, both citizens living in and around the project area and experts from different policy fields are increasingly included in the planning process to ensure sustainability for these three aspects. The stakeholders all have their own interests to uphold, thereby increasing the complexity of planning tasks. The complexity is exacerbated by the fact that the representatives of the various disciplines typically use a specific jargon and have a different understanding and knowledge of the questions at hand. The diverging interests and dissimilarities in the language used by the parties involved make it difficult to effectively discuss spatial issues.

Spatial planners have been using geographical information systems to enhance their analytical and decision-making capabilities since their development. The shift towards a more inclusive approach of spatial planning has necessitated that computer-based tools support not only the analyses of data from the various disciplines, but also the communication of the relevant information between all the stakeholders (Geertman, 2002). Furthermore - as the different parties involved in the design of the spatial plans have different levels of experience with computer technology in general and GIS in particular – the need for easy to use and accessible geographical information systems arose (Haklay & Tobón, 2003).

1.1.2 Public Participation GIS

As a response, since the middle of the nineteen nineties a type of GIS has emerged that aims to make the analytical and communicational capabilities of geographical information systems accessible to non-expert users. Although numerous tools and computer programs have been developed for this purpose in a variety of different approaches, they are commonly termed Public Participation GIS or PPGIS (Sieber, 2006). PPGIS creates the possibility for stakeholders to share information on spatial issues and communicate their arguments for or against the adoption of certain spatial intervention strategies. This is usually done by using the project area as a common starting point and providing the participants with the possibility to visualize relevant data in the form of maps. Chapter 2 describes the emergence of PPGIS applications, their main features and their use in spatial planning in more detail.

Effectively communicating information on a spatial issue can be complicated enough when it concerns objective data on a phenomenon within a certain discipline. But when attempting to include the opinions of the citizens living in and around the project area this process becomes even more challenging. The reason is that the one of the main obstacles that is often encountered is the difficulty of convincing citizens to participate in the development of spatial policy in the first place. As a result most research on PPGIS and public participation in spatial policy making has focused on ways to engage citizens and incite them to think about and discuss spatial issues in their living environment. In other words, a lot of attention is given to the methods to gather data on the opinions of citizens (Krek, 2005).

In order to be able to utilize those opinions, the raw data has to be organized, analysed and interpreted, so that that spatial planners can actually use it when drawing up their designs (Thwaites & Simkins, 2007). That analytical step, however, has not yet garnered much attention from the scientific community. Some studies include a description of the analysis they performed on the data, but a method for the analytical approach of PPGIS data has not been developed until now.

A few researchers have introduced mathematical tools to analyse PPGIS data. An example is the set of social landscape metrics, developed by Brown & Reed (2012) and the GIS-Multicriteria Decision Analysis by Boroushaki & Malczewski (2010). Another method that has recently been developed to address the issue of the communication of citizens' knowledge and experience of their living environment is softGIS (Kahila & Kyttä, 2009). These studies do not offer a general view of how to deal with PPGIS data, but they include elements that are useful for choosing the actual analysis to be used. Chapter 3 will take a closer look at these methods and the elements that have been incorporated in this research.

1.1.3 Spatial analysis

Spatial analysis is an umbrella term that contains all the methods available to turn geographical data into useable information. As Longley *et al.* (2010) put it: "Spatial analysis […] includes all the transformations, manipulations and methods that can be applied to geographic data to add value to them, to support decisions and to reveal patterns and anomalies that are not immediately obvious."

Spatial analysis ranges from very simple and intuitive techniques - such as the identification of patterns in a map simply by observing it – to sophisticated mathematical models. The level of complexity of a technique, however, is not an indication of its importance or usefulness (Longley *et al.,* 2010).

Which method of analysis is most suitable to address a certain spatial issue depends on a number of factors. Firstly, the analysis methods that can be used on a dataset vary according to the type of data. The types of analysis that can be performed on vector data are different than those available for raster datasets. Additionally, points, lines and polygons provide different possibilities for the methods that can be used. Secondly, from within the range of analysis methods that a dataset allows, the analytical method that provides the most useful outcomes is dependent on the spatial question that is addressed and the type of answer that is sought.

In this thesis, the aim is to create valuable information for use in spatial planning.In order to achieve this, the questionnaire used for the data collection in the study-area was developed in cooperation with the spatial planning department of the Heerlen Municipality. Also, their assessment of the information generated from the results of the thesis is used to evaluate the value of the applied methods.

Chapter 4 presents the method that was developed for the structured approach to the analysis and interpretation of PPGIS data, based on the findings from scientific literature. The application that was created for the collection of the data among citizens and the design and contents of the survey are described in Chapter 5. Subsequently, a description of the empirical part of this research is given in Chapter 6, which consists of the collection of the data and the characteristics of the respondents. Chapter 7 explains the steps that were taken to prepare to make the data usable in spatial analyses. Finally, in Chapter 8 the results are presented of the analyses used to apply the structure presented in Chapter 4 on the collected data.

1.2 Goals

As is stated in the introduction, participatory processes are becoming more frequent in spatial planning. In the last few decades instruments and applications have been developed in the form of Public Participation Geographical Information Systems to collect data on the preferences and experiences of citizens regarding their surroundings. This thesis argues that, in order for this data to be of any value in spatial planning, methods have to be developed for the translation of this type of data into information usable by policymakers in the spatial planning process. So far, little research has been done on this topic.

Therefore, the goals of this thesis are twofold: First, it aims to develop a structured method for the exploration and analysis of PPGIS data. The second goal of the thesis is to apply this method by translating the general recommendations into practical analytical steps in order to turn data that was collected during this research into information that is valuable in the daily activities of spatial planners.

1.3 Research questions

Due to the issues covered in this thesis, the research has a distinct exploratory character. As described in the previous paragraphs, the goals of the thesis are aimed at addressing both a scientific and pragmatic hiatus that has been identified. In order to propose a practically feasible approach to the translation of PPGIS data into usable information, it is important to understand the context within which PPGIS has been developed and what its role is in the spatial planning process. Consequently, the following research question is addressed in this thesis, which comprises its associated subquestions:

How can geographical data gathered among citizens through the use of PPGIS be turned into information that is valuable in the spatial planning process?

- *What is Public Participation GIS?*
- *What is the role of PPGIS in spatial planning processes?*
- *What information is valuable in the spatial planning process?*
- *What spatial analysis are available for geographical data gathered among citizens?*
- *What analysis can be used to translate PPGIS data into valuable information?*

1.4 Societal and scientific relevance

As the number of stakeholders in the spatial planning process grows, the need for effective communication between the various parties involved becomes more important. This thesis focuses on the analysis of data on citizens' experiences of their living environments. By developing a structured approach for the translation of data collected with PPGIS tools into information that can be used in the planning process, the effectiveness with which citizens' opinions and preferences are communicated can be increased. Such an improvement would benefit the participatory process as a whole, and the position of citizens within it in particular.

Furthermore, the analysis of PPGIS data has not been the topic of many scientific studies. More focus has been put on the possible methods of enticing citizens to participate actively in the planning process by using PPGIS applications to capture their opinions and preferences on spatial issues. However, for the use of PPGIS to be effective, the analytical steps for the translation of the collected data into useful information is crucial. In this thesis a structured approach for the analysis of data collected among citizens with PPGIS is proposed. It builds on elements from other PPGIS studies that contained sections on their analysis of the data. The proposed structure can serve as a starting point for further discussion of the possible methods for the interpretation of PPGIS data and its translation into usable information.

1.5 Case study

In order to answer the research questions of this thesis, PPGIS data was needed to which the proposed structure of analysis could be applied. In the empirical part of this research, consequently, data was collected among citizens of the city-region of Hoensbroek in the municipality of Heerlen. A PPGIS application was developed for the collection of spatial data through the use of a spatial survey. The survey itself was drawn up in cooperation with the spatial planning department for the Heerlen municipality and was subsequently conducted among Hoensbroek inhabitants. In total 85 citizens filled out the questionnaire. A selection of the collected data was analysed in order to translate it into valuable information for use in the planning process.

The developed application and the survey are discussed in more detail in Chapter 5. Chapter 6 describes the data collection method and the choice for Hoensbroek as the case-study area for this research.

2. Public Participation GIS

Geographical information systems have been used by planners to help them in the decision making process since their emergence about half a century ago. Their capabilities for the analysis and visualization of geographical data have allowed planners to gain insight into the spatial constraints that are present at planning locations (such as restrictions due to environmental legislation), predict the effects of spatial interventions in the built environment and effectively communicate their findings and recommendations (Geertman, 2002). With the inclusion of a growing number of stakeholders, however, it was felt that the traditional GIS applications were no longer suited to address the needs of all parties involved in the planning process. The calls for the development of applications that could provide access to similar capabilities for stakeholders without a background in GI technologies became increasingly stronger (Haklay & Tobón, 2003). As a response public participation GIS applications were developed.

In this chapter the concept of PPGIS is introduced. The emergence of public participation Geographical Information Technology is discussed in relation to the criticism of the use of GIS in spatial planning. Issues associated with the use of PPGIS are presented as well as its role in participatory processes. Finally, the issues that might hinder the broader adoption of PPGIS technology in spatial planning are discussed.

2.1 Origins of Public Participation GIS

It is difficult to determine what exactly public participation GIS entails. There is no one concise definition of PPGIS that covers the great diversity of applications, uses and methodologies that fall under this umbrella term (Brown, 2012; Dunn, 2007). In their effort to delineate the 'public' and 'participation' in public participation GIS, Schlossberg and Shuford (2005, p. 15) state that "the more one looks to find a common thread or meaning about what PPGIS exactly means, one quickly realizes that guiding definitions are not to be found and that utilizing the term 'PPGIS' is inconsistent across applications and uses." The lack of a coherent description for PPGIS can be attributed to a number of causes.

First of all, although a large number of definitions of GIS have been proposed over the last decades, there is still no real consensus in the academic world on what constitutes a GIS (Chrisman, 1999). As the range and variety of spatial applications and technologies keeps expanding, the disagreement is not about to be resolved anytime soon. Since GIS forms three-fifths of its abbreviation, the effort to establish a systematic characterization for PPGIS is not helped by fact that there are multiple interpretations of geographical information systems (Dunn, 2007). The second reason is that many initiatives to make geographical information systems more accessible to the various stakeholders in spatial decision making processes originated around the same time in diverse disciplines. As a result many tools were developed almost simultaneously but completely separate from equivalent initiatives (Sieber, 2006). A lot of these tools thus deal with very different issues, have different capabilities and a different approach but can nonetheless all be considered PPGIS initiatives.

2.1.1 Definitions of GIS

Although there is no definition without its critics, there is a common set of characteristics that is applicable to (almost) all systems that could be termed a GIS. As Couclelis (2003) notes, most of the interpretations of GIS describe the capabilities of the software to store, manage and manipulate data and specifically stress the abilities of visualization and analysis. In a lot of cases these capabilities are interpreted as a "series of procedures that lead from input to output; from data sources through processing to displays" (Chrisman, 1999, p. 178). This is called the **process approach**. An often-used definition that utilizes this approach is:

"GIS is a system for capturing, storing checking , manipulating, analysing and displaying data which are spatially referenced to the earth" (Department of the Environment, 1987, p. 132).

Clearly, the processes describe a rather linear structure with an emphasis on their technological aspects. Another approach used to define GIS is the **toolkit approach**, which attempts to make a clear distinction between GIS software packages and other types of software systems by addressing capabilities specifically attributed to GIS (Chrisman, 1999). A clear example of this type of definition is provided by international GIS software supplier ESRI:

"A geographic information system (GIS) is a computer-based tool for mapping and analysing things that exist and events that happen on Earth. GIS technology integrates common database operations such as query and statistical analysis with the unique visualization and geographic analysis benefits offered by maps. These abilities distinguish GIS from other information systems and make it valuable for a wide range of public and private enterprises for explaining events predicting outcomes and planning strategies" (ESRI, 1997).

The fixation on the technological aspects of geographical information systems is not surprising as the adoption of GIS has generally been driven by the development and marketing of software by large companies such as ESRI (Sieber, 2004).

2.1.2 Criticism of GIS

From the moment the use of GIS became more widespread in the 1980's, criticism of its overly technical characterization began to grow (see for instance Taylor, 1990; Taylor & Overton, 1991; Openshaw, 1991; 1992). Although those voices were initially limited, a number of alternative definitions of GIS were proposed, in which its social and institutional context were included. A commonly cited example that was written up by a Delphi panel of GIS specialists reads:

"[GIS is] a system of hardware, software, data, people, organizations and institutional arrangements for collecting, storing, analysing and disseminating information about areas of the earth" (Dueker & Kjerne, 1989, p. 7).

Throughout the nineties the criticism of conventional geographical information systems increased. The main concern was the uneven distribution of access to spatial data, GIS technology and spatial analysis tools for different stakeholder groups in the decision making process. As the enthusiasm for the use of spatial data and analysis in guiding societal issues increased, the implications of GIS technology for the representation of people and place became more and more important (Elwood, 2006).

Conventional GIS were mostly used by “white males employed in academic and governmental institutions in North America and Europe” (Obermeyer, 1998, p. 65). Other groups with an interest in the issues that were managed through the use of GIS had only limited access to the relevant spatial data and software and, in the cases where they were available, lacked the expertise to utilise the technology. As a result many started to view GIS as a tool used by governments to shut out grass-root movements and other interest groups from the decision making process and thereby retaining a dominant position in policy making (Sieber, 2006).

Furthermore, critics cautioned that a reliance on GIS in spatial and environmental planning and other decision making processes would give way to a positivist approach in which intricate spatial issues were reduced to points, lines and areas. They claimed that GIS created the illusion of control and understanding of the issues where the real-world situation could be infinitely more complex (Sieber, 2006). Critics also pointed out that certain types of knowledge - mostly different types of ‘soft’ knowledge which is based on experience rather than measurements and research - were difficult to capture in a GIS. Consequently, when relying solely on the use of GIS technology, these types of knowledge would not be included in decision making processes, thereby also excluding the people and places it represents (Elwood, 2006). Elwood (2006, p. 695) sums up these objections by stating that “the epistemological and representational concerns about GIS focused on the limitations of the technology, and the potential social and political implications of these limitations, especially given the seemingly unmitigated enthusiasm for GIS at the time.”

The opinions of the drivers of the critique of GIS - which became known as GIS and Society (GISoc) – were voiced in a number of significant publications, the most well-known of which is Ground Truth (Pickles, 1995). In this book, many weaknesses of the use of GIS were scrutinized. A main point of criticism was that GIS only benefited those who had access to it (i.e. the decision makers) and marginalized those who did not. A 1995 special issue of the Cartography and Geographic Information Systems journal and a number of papers in Environment and Planning A were also important in the dissemination of their concerns.[1] Mostly as a reaction to the critique of GISoc several attempts were made to produce geographical information systems that were more socially aware (although some initiatives had started before the publication of the papers mentioned above) (Dunn, 2007). A growing number of researchers and GIS experts acknowledged the importance of the social and political implications that are inherent to the use of GIS in decision making processes and tried to gain insight into the way GIS technologies produce space, knowledge and power (Elwood, 2006). These efforts were the first steps in the development of a wide range of applications that looked to make GIS more accessible to all stakeholders in decision making and the drafting of policies and ultimately to “empower the less privileged groups in society” (NCGIA, 1996).

2.1.3 The emergence of Public Participation GIS

A wide range of initiatives to make GIS more democratic originated around the middle of the nineteen nineties. Some early efforts even saw collaborations between GIS researchers and prominent critics. A nice example is the work by Harris *et al.* in the aforementioned Ground Truth (Harris et al., 1995). With the development of more inclusive GIS tools the field of public

[1] For a detailed description of the debates and criticism surrounding the theory and practice of GIS see Pickles (1999) and Schuurman (2000).

participation geographical information systems emerged. But, although numerous conferences and special journal issues have been dedicated to PPGIS since that time, it remains a rather fragmented field (Sieber, 2006).

The term PPGIS was only first used in 1996 at two gatherings of the National Center for Geographic Information and Analysis (NCGIA). The conference was on the advancement of GIS technology and the development of the next generation of geographical information systems dubbed GIS/2. One of the main premises was that GIS/2 should position GIS tools more firmly in their social and political context (Sieber,2006). During the meetings it was revealed that a growing number of the participating GIS specialists had the desire to develop (or were already developing) GIS tools that were better able to represent the unofficial voices in decision making processes (NCGIA, 1996). Consequently the term PPGIS was used to describe the pragmatic approach of developing GIS technology to incite citizens to actively participate in the debates surrounding societal and other spatial issues and so providing them with tools to influence government policy. As a result, the definition of PPGIS that was drafted at the 1996 NCGIA conference reads: “[PPGIS is] a variety of approaches to make GIS and other spatial decision-making tools available and accessible to all those with a stake in official decisions” (Schroeder, 1996).

In an attempt to specify this vague definition, Aberley & Sieber (2002) compiled a list of fourteen guiding principles for PPGIS. This list includes themes such as the promotion of sustainability and social justice, the inclusion of marginalized groups in society, public access to official geographical data and the linking to theories and research methods from different social science disciplines. Although more specific than the definition from the 1996 NCGIA conference, the diverse set of characteristics described in the guiding principles is illustrative of the difficulty of containing the practically innumerable number of analogous applications in one concise definition. Consequently, many different labels have been given to describe different subsets of alternative GIS tools: Public Participation GIS (PPGIS; the mode widely used), Participatory GIS (PGIS), Community-integrated GIS, GIS/2, GIS for Participation (GIS-P), Participatory 3-Dimensional Modelling (P3DM), Bottom-UP GIS (BUGIS), Collaborative GIS and SoftGIS (Dunn, 2007). All these terms were thought up in different contexts by different developers and researchers and therefore represent distinct ideas on approaches and methods appropriate for the democratization of GIS (Sieber, 2006).

Not all of these varieties will be discussed in detail in this paper. [2] In the next section, however, the difference between PPGIS and PGIS will be elucidated. Furthermore, in the next chapter attention is given to the SoftGIS concept, as it has provided a lot of inspiration for the tool that was developed for this research.

[2] Relevant literature on the alternative GIS tools from Dunn (2007): Community-integrated GIS (Harris and Weiner, 1998), GIS-2 (Harris and Weiner, 1996; Schroeder, 1996; Pickles, 1999), GIS for participation (GIS-P; Cinderby, 1999), Participatory 3-Dimensional Modelling (P3DM; Rambaldi & Callosa, 2000; 2002), Bottom-Up GIS (BUGIS; Talen, 2000), and collaborative GIS/geocollaboration (Mac-Eachren and Brewer, 2004; Schaffers et al., 2005; Balram and Dragicevic, 2006; Jankowski et al., 2006).

2.2 The various aspects of PPGIS

There is huge variety of applications that can be considered Public Participation GIS. In this section an attempt is made to determine the characteristics that can help to distinguish between various alternative GIS initiatives.

2.2.1 Top-down and bottom-up

One of the main differences between various approaches to the development and use of alternative GIS is where the application originates and who is responsible for its development. As stated above, the term PPGIS was coined in 1996 at a meeting of the NCGIA in the United States. Applications that fall under the PPGIS label are commonly devised to address the desire of government agencies of developed countries to better engage the general public in spatial policy making (Brown, 2012). This means that the development and implementation of PPGIS technologies can be considered mostly a top-down process: they are the result of the use of GIS technology to support governments´ growing ambitions of participatory planning. PPGIS applications usually focus on the communication and usage of spatial information that is relevant to the spatial issues that need addressing among local stakeholders. It often concerns urbanized areas and its related spatial problems such as the discriminatory zoning of minorities in society (Rambaldi *et al.*, 2006).

PGIS, on the other hand, emerged in developing countries where GIS was used in Participatory Learning and Action initiatives. PGIS projects often involve the support of indigenous groups by non-governmental organizations and development scientists through the supply of GIS knowledge and technology to grassroots groups and community-based institutions (Brown, 2012). In those cases the aim is to empower the native inhabitants of a region by providing user-friendly and culturally sensitive applications to generate, manage and visualize spatial information in order to communicate their wishes and arguments in a spatial decision making process. These efforts to oppose undesirable policies by national or regional governments usually concern rural areas. As opposed to PPGIS initiatives, PGIS commonly represents bottom-up processes (Rambaldi *et al.*, 2006).

2.2.2 Who is the public in PPGIS?

Related to the question of whether an application is developed for and used in a top-down or bottom-up process is the issue of what specific parties should participate in a PPGIS project. As the criticism of the specialist character of conventional GIS technology - and consequently the number of more inclusive applications - grew, the discussion arose who should be involved. How to identify the public is of course relevant in any decision making process applying public participation elements and has been studied rather extensively (see for instance Day, 1997; Thomas, 1995). Schlossberg & Shuford (2005), however, have found that there is no clear definition of who is considered a stakeholder. Nevertheless, they argue that the choices made in delineating the public are of great importance in a public participation process as they affect and are affected by the objectives that are set for a project.

Thus, numerous methodologies have been described to determine who the most important stakeholders are in a project. One such approach is described by Rietbergen-McCracken and Narayan-Parker (1998) in their outlining of how to perform a stakeholder analysis. The first step in the analysis is to identify the key stakeholders. In order to do this, they pose the following questions:

1. Who are potential beneficiaries?
2. Who might be adversely affected?
3. Have vulnerable groups been identifi ed?
4. Have supporters and opponents been identifi ed?
5. What is the relationship among stakeholders?

(Rietbergen-McCracken & Narayan-Parker, 1998, p. 67)

Other methods are described by Willeke (1974), Thomas (1995), Aggens (1983), Mitchell *et al.* (1997) and Creighton (1973) among others.

In an attempt to frame the characterizations of the public, Schlossberg & Shuford (2005, p.18) state that there are three general classes into which stakeholders, as identified by the aforementioned researchers, can be allocated:

- "Those affected by a decision or program […],
- those who can bring important knowledge or information to a decision or program […] and
- those who have power to influence and/or affect implementation of a decision or program."

These classes are not mutually exclusive, as a group of stakeholders can fall into two or even all three of the categories. The categorization, however, substantiates the notion that the public is made up of a combination of actors with differing levels of skill and differing stakes (Sieber, 2006). Furthermore, categorizing the identified stakeholders in this way can provide some direction as it is usually beneficial to the process to involve stakeholders from all three classes.

In other words, when it is determined that a decision making process with public participation will be initiated, it is vital to define who the public is and what role the various parties involved will play. As there is no predefined group of people who constitute the public, this process will differ for each individual project. There are several methods to structure the delineation of the stakeholders. Who is eventually involved in the public participation process is mostly dependent on the intended outcomes of the project.

2.2.3 Indigenous spatial knowledge

In many instances the use of a PPGIS in the decision making process includes gathering data among the selected stakeholders. The collected data is used as input in the process of formulating one or more possible solutions to the issues at hand. The overarching goal of PPGIS is to facilitate the inclusion of all relevant knowledge and opinions in decision making processes. Specialists from various disciplinary backgrounds usually provide expert knowledge on one or more themes relevant to the project. Citizens living in the geographical vicinity of where the project is taking place or that are otherwise possibly affected by the outcomes of the project often don't possess such expert knowledge. What they do possess is knowledge and opinions that they have acquired and formed through the experience of living in the area in question. This type of knowledge is called experiential knowledge (Kyttä *et al.*, 2013), or indigenous spatial knowledge (McCall, 2003).

Indigenous knowledge is a pool of information shared by the inhabitants of a certain area (McCall, 2003; 2004). Some indigenous knowledge is practical and aimed at identifying local issues and their possible solutions. As such, it forms the basis for the communication and decision-making of a community within a geographical region. This type of local knowledge is mostly of interest in developing countries, where native inhabitants pass on information on for instance local resources, resource management and ecosystem relationships from generation to generation (McCall, 2003). The indigenous knowledge can often be considered similar to or even more reliable than scientific (technical) knowledge, as it is acquired through practical experiences with the issues (McCall and Minang, 2005).

Another type of indigenous knowledge is more subjective in nature and focuses on the emotional relationship between people and their environment. On the one hand, this can refer to the attribution of spiritual and cultural values to certain lands by the native inhabitants (McCall, 2004). On the other hand, however, the term is used to indicate the associations of citizens with specific geographical locations. This interpretation is more applicable to urbanized regions in western society. It involves the values given by people to specific locations, the needs they experience in certain areas and the way they prioritize these associations (McCall & Dunn, 2012).

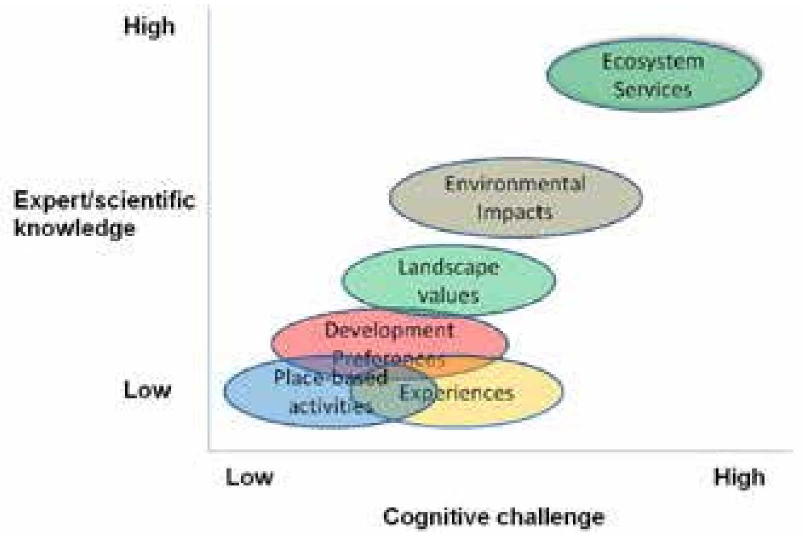


Figure 2.1: Categories of spatial attributes plotted against the cognitive and knowledge requirements for their identification (Brown, 2012)

Brown (2012) has performed many PPGIS research projects for a variety of regional and environmental planning programs in which different types of local knowledge were acquired by asking participants to identify various spatial attributes in their local urban or natural environment. In his experience with PPGIS initiatives in the United States, New Zealand and Australia he has categorized six different types of spatial attributes that participants can be asked to identify. The categories of spatial attributes require differing levels of expert knowledge on the relevant themes and also are a bigger or smaller cognitive challenge to locate spatially (Figure 2.1).

This research focuses on data collected among citizens on their experiences and perceptions of their environment , which puts them in the lower left corner of Figure 1: place based activities, experiences and landscape values. It is important to realize in what category the questions you ask of respondents fit and what implications that has for the expert knowledge required to answer those questions as it can affect the willingness of citizens to participate in the PPGIS project. This issue will be addressed in section 2.3.4 on the issues hampering the widespread use of PPGIS.

2.3 Participatory processes in spatial planning and PPGIS

2.3.1 A short history of public participation in spatial planning

In the first half of the twentieth century, participation in spatial planning was a privilege enjoyed only by the elite of society if any. Spatial decision making was a distinctive top-down process in which plans were drawn up by professional spatial planners and policymakers. Those plans were consequently executed by the civil servants and were deemed to be accepted by the citizens (Hall, 2003). As a number of comprehensive planning projects around the middle of the twentieth century turned out to have an less-than-positive outcome, however, the mistrust of professional planners quickly grew. Fuelled by the struggle for civil rights, citizens in the nineteen sixties throughout the developed world organized themselves and were able to appropriate an official position in the spatial decision making process (Ramasubramanian, 2010).

In fact, spatial planning was among the first policy fields in which the concept of public participation in decision making took root (Concilio & Molinari, 2011). In 1969 Arnstein presented a typology of the varying levels of citizen participation in policymaking in her seminal work on community participation titled ‘A Ladder of Citizen Participation’. She visualized the different degrees of participation as the rungs of a ladder, which became known as Arnstein’s Ladder (Figure 2.2). The eight rungs are divided into three categories, namely nonparticipation, tokenism and citizen power. The higher a project or process can be placed on the ladder, the more power citizens have in the decision making process through their participation (Arnstein, 1969).

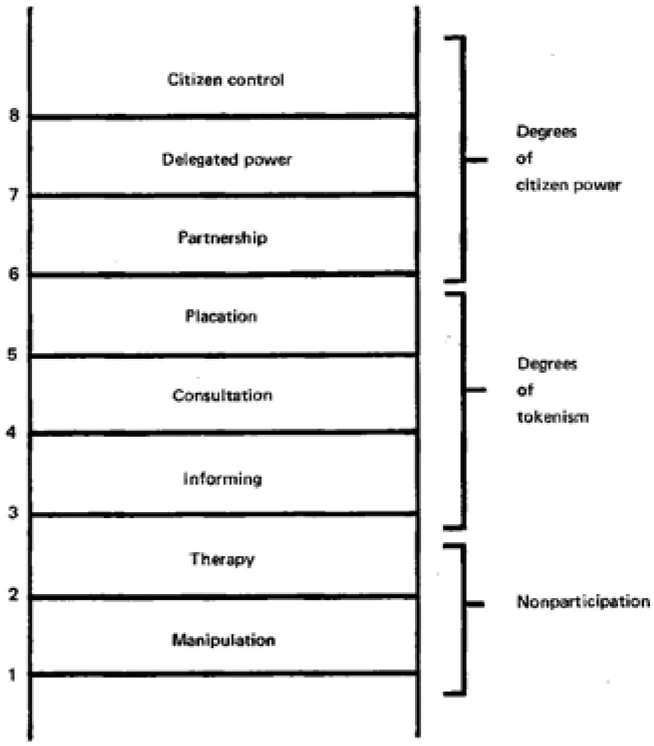


Figure 2.2: Arnstein’s Ladder of Participation (Arnstein, 1969).

Since the sixties, citizens have usually had a some role in spatial planning that can be represented by a rung from the tokenism category on Arnstein’s ladder. As of the middle of the nineteen nineties, however, interest has grown in a more collaborative approach to spatial planning in which the role of the citizen is more pronounced (Coenen et al., 2001)[3].

The Dutch context

In the Dutch planning context, one of the main reasons to initiate projects of interactive policy-making and participatory spatial planning was the rising criticism on the lack of efficiency and efficacy of the consultation methods adopted by local governments. Public hearings were (and still remain to be) the legally required and thus most commonly used tool to provide citizens with a chance to voice their opinion. These hearings, however, were perceived to come too late in the planning process, thereby denying the public the possibility to have any kind of input in the discussion other than voting on one of several designs or rejecting the proposals altogether. Consultation of the public was viewed more and more as a slowing of the decision-making process without adding much value. Participatory planning was thought to provide a solution to these issues by involving citizens and other stakeholders in an earlier stage in the spatial planning process (Coenen et al., 2001).

The United Kingdom

In the United Kingdom, the Royal Town Planning Institute (RTPI) also acknowledges the deficiencies in the organization of citizen participation in spatial planning in recent years (2007). The Institute states that the pace of changes in society has increased and that current planning practices are not capable of managing those changes if the goals of supporting environmental, economic and societal sustainability through planning are to be achieved. One of their recommendations is therefore that all interests that planning seeks to endorse, curtail or otherwise influence be integrated in the planning procedure by including all stakeholders at an early stage in the process. Just like in the Dutch context, the RTPI believes special attention should be given to the inclusion of citizens to counter the trend of growing imbalance between the decision makers and those most affected by the decisions: the citizens (Royal Town Planning Institute, 2007).

2.3.2 Arguments for (and against) participatory spatial planning

Arguments for

Advocates of participatory planning often claim that citizen involvement in decision making possesses a certain intrinsic value (Day, 1997). They posit that in a perfect society citizens are actively and directly involved in its governing, thereby recalling Artistotle's theory on the role of citizens in politics. Governmental legislation and regulations should provide the public with opportunities to become involved in decision-making more deeply, but the citizens themselves also have the responsibility to take the opportunities provided to them. Stivers (1990) argues that citizens must move away from the role of consumers or clients of the government and adopt the role of full participants in the decision-making process. It is his belief that a body of capable and engaged citizens will not only lead to a better representation of the interests of the general public by the inclusion of a larger number of people, but that active participation has beneficial consequences for the individual citizens by enlarging their recognition and appreciation that the community's

[3]For further reading see for instance Cooke & Kothari, 2001; Lovan *et al.*, 2004.

interests are as a matter of fact their own (Stivers, 1990). In *The Social Contract* Jean-Jacques Rousseau even went as far as to reason that it must be obligatory for all citizens to take part in the participatory processes for the benefit of both the individuals and the representativeness and quality of the decisions made (Pateman, 1970).

Apart from this rather philosophical plea, a number of more practical arguments can be made for the adoption of the participatory approach in spatial planning. Innes & Booher (2004) concisely present the reasons mostly used to substantiate the call for greater public involvement in policy-making. One of the purposes they have identified is that an increased use of local spatial knowledge when drafting policies potentially leads to improved solutions, as it adds valuable information not possessed by external planning experts and can lead to innovative approaches . Also, taking the preferences and priorities expressed by citizens into account from the early stages of the planning procedure arguably leads to decisions that better reflect the wishes of the citizens (Innes & Booher, 2004; Day, 1997).

Although actively involving citizens throughout a project requires an investment of time and effort, having insight into the needs of the public can lead to a more efficient progression in the later stages of the planning process as it (at least partially) eliminates the risk of unexpected opposition from the community and other stakeholders when nearing the implementation phase (Day, 1997).

Public participation also increases equality and fairness in the decision-making process (Innes & Booher, 2004; Royal Town Planning Institute, 2007). It gives many groups in society a platform to communicate their needs that would otherwise not be available to them. In the current situation the most disadvantaged people are often underrepresented if they are represented at all, even though they are usually the ones who could benefit most from the decisions being made. If participatory projects indeed succeed in involving citizens from all tiers of society, the outcomes of such projects are likely to gain support among a larger number of citizens and can serve as a substantiation of the legitimacy of the consequent plans or policies for the responsible governmental bodies (Innes & Booher, 2004).

Arguments against

Of course, citizen participation also has its potential negative effects. Stivers (1990), for instance, recognizes that it is long agreed in Western politics that it is impossible to involve citizens in many of the tasks and decisions for which governmental agencies are responsible. The workings of modern nation-states are simply too complex and too big to allow for the intimate engagement between governments and citizens that is sought in a truly participatory democracy (Stiver, 1990). Taking this notion into account, the question might be raised whether participatory planning is a desirable ambition in all situations, regardless of the context or the issue at hand. For instance, in issues where time is limited and swift action is required the involvement of citizens and other stakeholders may delay the execution of intervention strategies to such an extent that their effectiveness may be reduced.

Furthermore, doubts can be raised whether the representativeness of the process will actually grow with the participatory approach. The number of people willing to actively contribute to spatial planning discussions or other decision making processes appears not to have increased as much as anticipated and hoped with the introduction of participatory projects. The question then remains whether a wider variety of groups in society is represented by adopting participatory strategies. This issue will be discussed further in section 2.3.4.

2.3.3 The role of PPGIS in participatory planning

In order to gain insight into the possible ways of shaping the participatory process, exploratory research projects and case-studies have been performed. PPGIS is often used as a tool in these types of processes. But what role does PPGIS exactly play? In other words, what elements of PPGIS support parts of the participatory processes in a spatial planning context?

The ultimate goal of PPGIS is to provide citizens and disadvantaged groups in society with a means to claim a more substantial role in the decision-making process (Sieber, 2000; Haklay & Tobon, 2003; Elwood, 2002). This means that essentially a large part of the value of PPGIS has to do with is communicational capabilities. In section 3.2.3, for instance, the concept of indigenous spatial knowledge was explained. This knowledge is gained by people through the experience of living in a certain region for an extended period of time. When citizens are asked to participate in the development of a spatial plan, they usually rely on this experiential knowledge for their input. After all, most citizens do not possess any expert knowledge on subjects relevant to the planning issues. Moreover, for the most disadvantaged groups in society this knowledge might very well be the only means they have that they can bring to the table in order to defend their interests (McCall, 2003). It is, therefore, of importance that the indigenous spatial knowledge is gathered, analysed and represented in such a way that it carries the same validity as thematic knowledge provided by the other stakeholders. PPGISystems can be seen as a technology with which local and indigenous knowledge and expert data can be combined (Dunn, 2007).

One-way communication

The integration of these two types of data practically revolves around the communication of information between the government and its citizens and other relevant stakeholders. In the first applications that were developed PPGIS mostly consisted of the use of GIS tools to convey the specific requirements of different (usually urban) areas with regards to services, safety, transportation etcetera. This was done by visualizing the connection between certain phenomena or occurrences (e.g. number of crimes committed) in a region and the socio-economic status of inhabitants on the basis of census data (Ramasubramanian, 2010). As the representatives of community-based organizations tried to make a case for the needs of their citizens by applying PPGIS technology, the communication can be seen as one-way: from the citizens to the relevant governmental agency.

Communication from the government to the public also got a boost at the end of the nineties as the variety of applications supported on the web quickly grew. More and more people started to use the web for gathering information which spurred the possibilities of using it to disseminate data and information. In many developed countries governmental agencies from all tiers of government began the provision of data for citizens on for instance census information, environmental information and information on housing and urban development plans (Ramasubramanian, 2010). Through web-services citizens are able to view information relevant to their living environment. The data can be of use in situations where people want information on building permits and other legal constraints for for instance housing development in the area. Municipalities can, in turn, keep citizens up-to-date on building progress and its consequences for transport, noise disturbance etcetera.

Two-way communication

In participatory planning, however, the goal is ultimately to achieve some form of two-way communication, a substantive discussion, between stakeholders and the decision makers. This can be achieved in several ways with the use of Public Participation GI Tools. The two main categories that can be identified are the applications that can be accessed on the web and those that run locally on a PC, a tablet or even on a smartphone. The platform that is chosen (web-based or local applications) determines what methods can be used to achieve the communication between stakeholders and the government (Kingston *et al.*, 2000). If the choice is made to make use of a web-based PPGIS tool, citizens can be invited to use the application to voice their opinions on certain themes. In contrast to the example provided above, government representatives can respond to the input provided by citizens.

A nice example is the Dutch website 'www.verbeterdebuurt.nl' ('improve the neighbourhood.nl')[4]. On this website citizens can make a complaint about different forms of nuisances or disturbance by placing a pin on a digital map and adding a description of the problem. People can also add ideas for improving the area. The complaints and ideas are emailed to the municipality where an employee can respond to the input. The status of activities (to be) performed by the municipality is tracked and made visible to the public. Furthermore, other citizens can respond to and endorse the ideas and complaints. Essentially the website functions as a forum based on location on which public and government can engage into a debate.

In other cases a website may not be deemed a suitable platform to facilitate participatory processes. In those cases a community meeting or workshop type setting may be adopted in which the communication between citizens and decision-makers takes place. In practice such gatherings often lead to some form of community mapping. It has a lot of similarities with the web-applications in that it provides the attendees with the possibility of giving input on the issues at hand either by adding points of interest on a map or by presenting their preferred design for an area. In the context of a meeting, however, the suggestions can be discussed directly so as to cater to a genuine exchange of ideas.

Web-based applications versus community meetings

Both web-based applications and organized meetings have their advantages and their draw-backs. These mostly boil down to the notion of representativeness. Web-applications, on the one hand, give citizens the opportunity to participate from the comfort of their own home and at a time that suits them. The relative anonymity of the internet can entice people to provide their opinions that they may otherwise not have given in a situation where they are surrounded by others. The question remains, however, who can be reach with these types of applications. Not everybody has access to internet or is computer-literate enough to work with such applications. Also, the internet as a communicational platform can be experienced as rather impersonal and may not deliver a discussion with the depth that face-to-face meetings can organize.

Community meetings and organized workshops, on the other hand, are probably better suited to directly host an exchange of ideas and information between citizens and governments. But here the question is how many people will make the effort to actually show up. Experience has learned that community meetings are usually not very well attended which is of course detrimental to the representativeness of the voiced opinions (Bamberg, 2010). Furthermore, those that do

[4] http://www.verbeterdebuurt.nl/

attend are often people with a very strong (and usually negative) opinion on proposed plans and suggested solutions to the issues at hand. Their presence alone might keep others from attending these types of meetings and their opinions often drown out the views of other attendees. These types of citizens are referred to as 'the usual suspects' (Kingston *et al.*, 2000).

Acknowledging the pros and cons of both methods for the point of communication, the right choice wholly depends on the situation. It might even be desirable to adopt a mixed approach to offset some disadvantages of both tactics. Then again, this might be considered an unjustifiable amount of effort by the governing bodies. It is clear though, that the issue of representativeness is crucial in the debate on the added value of PPGIS in the spatial planning context.

2.3.4 Participatory planning has not taken flight

Although the potential usability and value of PPGI tools has grown tremendously thanks to the technological advances of the last twenty years, its effect on planning practice has not been as significant as hoped and maybe even expected (Brown & Kyttä, 2014). As a matter of fact, participatory planning processes are still very much the exception rather than the rule, despite that fact that it can potentially resolve a lot of issues that have been raised on the planning paradigm in the last few decades. So what are the reasons that participatory planning and consequently public participation GI tools have thus far failed to make a significant impact on planning practice?

The people

The issue of representativeness is of utmost importance when discussing participatory planning methods. It has become clear, however, that it is really difficult to get people from all groups in society involved in participatory planning projects. There are two main reasons why citizens are not participating: they are either unable or unwilling. These reasons are often linked but for clarity they will be described separately.

There are numerous ways to approach citizens in order to induce them to share their opinions. But not everybody you invite will be present or will fill out a questionnaire. In fact, depending on the method chosen to capture the opinions of the citizens, chances are that a group of people is unwittingly excluded from the process. Although the reasons for their exclusion may appear obvious, they form a serious obstacle in the justification of participatory planning as the preferred approach (Cinderby, 2010). Citizens might, for example, be unable to participate because they don't speak the language and do not understand the invitation to a community meeting or the questions in a questionnaire. Some people might be physically unable to attend due to an injury or a handicap. Others may not have access to the internet so that web based invitations, questionnaires or PPGI applications will not reach them. Some people are computer illiterate to such an extent that any tool on a PC or tablet is beyond their comprehension. More examples could be provided but it is clear that there are many reasons why certain people might be

The most important aspect of the unintentional exclusion of groups of people from participating is that it is often those exact people who should benefit the most from the participatory approach. They are usually the minorities who are having a hard time having their interests defended and their needs met through current planning procedures (Cinderby, 2010). Since participatory planning supposedly represents everyone's interests equally, this exclusion of disadvantaged groups poses a big problem.

But it is not just the disadvantaged groups who are hard to reach. Response rates for participatory planning initiatives have not been very promising in general. The conclusion can be drawn that people are not easily motivated to spend their time on participating in a decision-making process. Krek (2005) explains this phenomenon through the use of the concept of rational ignorance. His explanation is that people knowingly decide against partaking on these types of projects because the possible benefits that they will gain by participating do not outweigh the investment that they have to make in terms of time and effort in order to be able to participate at all. On the one hand the effort to participate grows as more knowledge is required in order to be able to join in the planning discussion.[5] On the other hand, the potential benefits of participation decrease if the impression exists among citizens that their input in the discussion will not influence the outcome of the discussion. In order to induce more people to participate in the planning process, it is important that the balance between the required invested effort and the experienced benefits for citizens is tipped in the other direction.

The procedures

But is not just a problem of convincing citizens to participate more. What is also missing is a change in the legislature and regulations surrounding spatial planning. Although participatory projects have been introduced more frequently, these initiatives are still to a large extent experimental in nature. In other words, the participatory approach is not yet recognized as an alternative to the current course of business. As a result, when there is a shortage in time and money – and there is always a shortage in time and money in spatial planning – the participation of citizens in the planning process is kept to the legally required minimum: the traditional public hearing.

Another result is that, as is the case with spatial planning support systems (Brown & Kyttä, 2014) PPGI tools are mostly used in research projects and are only rarely adopted in the field of practice. In order for PPGI tools to make an impact it is essential that the participatory planning approach is grounded more deeply into the planning doctrine of policymakers and eventually in legislature.

2.4 Conclusions

Spatial planning problems have grown increasingly complex over the last decades at least in part due to the involvement of a growing number of stakeholders. As a result, the communication of different types of knowledge between the various parties involved has become more important. The knowledge of citizens that is relevant to the planning process is usually based on the experiences they have had at certain locations in their daily activities. In order for participatory processes to be successful, it is therefore of importance that this type of knowledge is translated into information that can be understood and used by policy-makers.

PPGIS applications have the aim to provide citizens and other stakeholders with the means to better put forward their arguments in the spatial decision-making process. Depending on the specific application, this is done by providing access to geographical data and analytical tools, by collecting the opinions of different stakeholders or by functioning as a platform for the exchange of ideas.

[5] See section 2.2.3 for the different levels of knowledge required for answering certain questions that can be addressed using PPGIS technology.

Many studies have been performed on the collection of data among citizens and the various methods for enticing citizens to actively participate in the spatial planning process. Very little has been written, however, on the analysis of the data collected among citizens for use in the spatial planning process. This thesis argues that more research is needed on this translation of data collected with PPGIS applications into usable information for spatial planning.

3. Analyses in PPGIS studies

The analysis of PPGIS data and its translation to information for policymakers has not received a lot of attention from researchers. In order to be able to utilize data collected among citizens, however, it is essential that this data is communicated in such a way that it can be interpreted correctly and effectively by the other stakeholders involved in the project (Dunn, 2007; Kahila & Kyttä, 2009; Kyttä *et al.*, 2013). Despite the scarcity of scientific research devoted specifically to the analysis of PPGIS data, a large number of studies have been performed on projects involving the use of PPGIS to support (part of) spatial planning processes. In these studies the analysis of the data is not often described in great detail.

The next section describes a selection of PPGIS methods that have been developed within which the analysis of the data plays a significant role. The overview is not comprehensive. Many PPGIS studies have without doubt included analyses of the collected data but the amount of attention that this part of the research has gotten in scientific literature is meagre. It is therefore difficult to determine what methods have been used in those studies. This inventory thus focuses on literature in which the analysis amounts to an important part of the study and which addresses research that has some similarities to the research performed for this thesis: the use of PPGIS in spatial planning processes. These three studies are described as they have provided valuable input for the development of the structured method for the analysis of PPGIS data described in the next chapter. Furthermore, a number of analysis described in the results chapter are adaptations of approaches described in these articles.

3.1 SoftGIS

SoftGIS is developed as a response to Helsinki's urban densification planning approach (Kyttä *et al.*, 2013). Earlier research by the authors had shown that building density was inversely related to the perceived quality of the environment by citizens. Other studies, however, suggest that the relationship between urban density and environmental quality is a little more complex and highly dependent on the local context (McCrea & Walters, 2012). As a response to these findings, Kyttä *et al.* propose an approach to urban densification projects that is sensitive to the experiences and requirements of the local citizens. They suggest adding a social, ´soft´ layer to the planners' arsenal with which they could "perform contextually sensitive planning that could fundamentally increase understanding of the social acceptability of urban densification projects" (Kyttä *et al.*, 2013, p. 31). The SoftGIS method is a PPGIS tool used to provide insight into and communicate the residents' experiences.[6]

In the research citizens in Helsinki were asked through a web-survey to place points on a web-map indicating an association they had with that location. There were four categories to choose from when choosing an association: function, social life, appearance and atmosphere of the environment. These categories were further divided into eight sets of positive and negative attributes of the location of interest. So in total, respondents had a list of 64 associations to choose from. Respondents were free to place any number of experiences.

[6] Or ´location-based experiential knowledge´, as they have termed it.

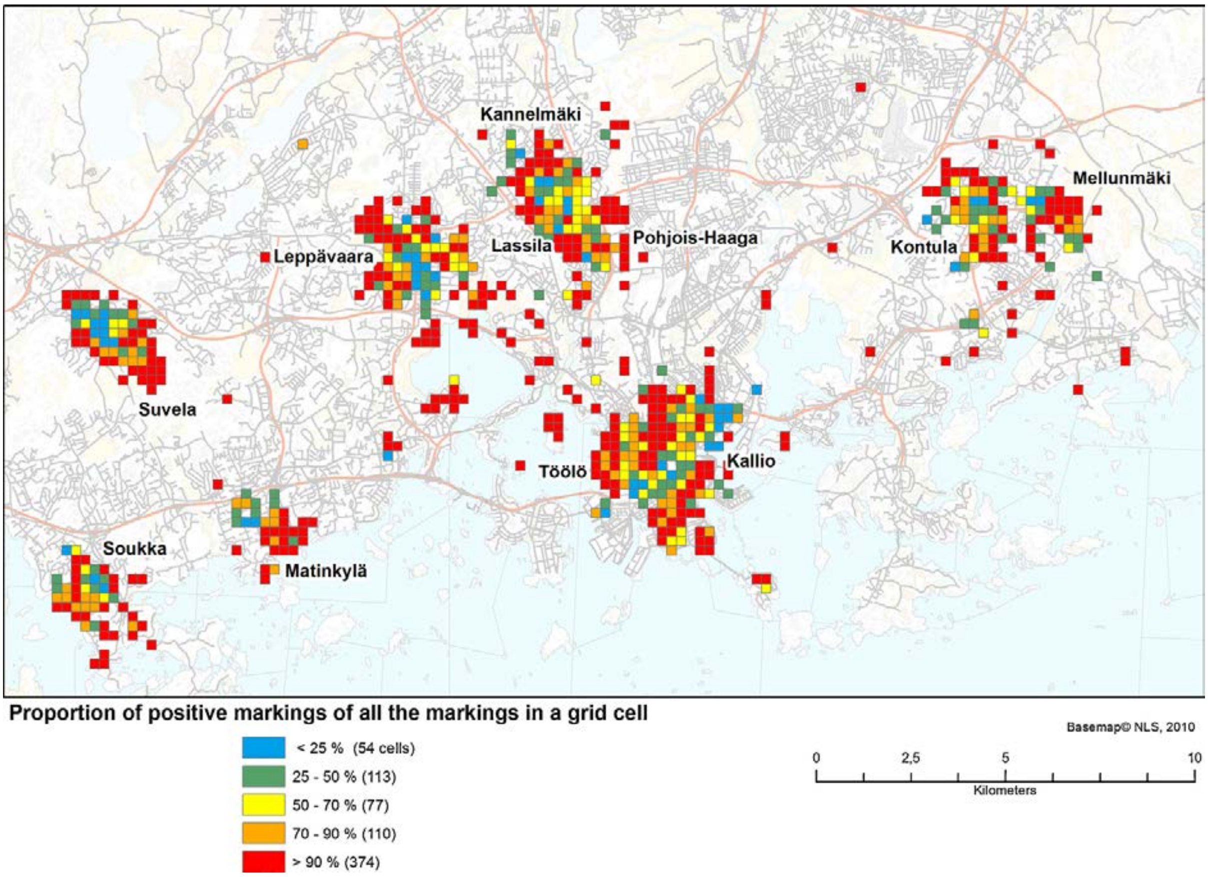


Figure 3.1: The percentage of positive associations of the respondents for each location (Kyttä *et al*., 2013)

In order to study the associations collected among the citizens, the researchers first looked at the most commonly placed positive and negative attributes. The results of this analysis are shown in Figure 3.1. Consequently, interpolation of the point-data was performed in order to provide an overview of the positively or negatively perceived areas within the Helsinki neighbourhoods. The authors state that the interpolated data can already provide planners with a global idea of what areas are in need of attention in the experience of the residents (Kyttä *et al.*, 2013).

They recognize, however, that further analysis of the data can increase its usefulness for planning purposes. Therefore, they examine whether the characteristics of the built environment have an influence on the possibility of a place being rated either positively or negatively. This is done by performing a regression analysis using urban structural measures within a 50 meter buffer surrounding the points placed by the respondents. The characteristics of the built environment that were used were the building density and the percentage of green. Furthermore, the researchers looked at the land-use types in which the environmental quality factors were placed.

Over 3.000 respondents participated in the research, who together provided over 10.000 experiences. Although this is an impressive amount, it means that the average respondent provided only three or four associations. As a result, little can be said on trends in the experiences provided by the individual respondents.[7] Furthermore, because the list of associations that can be chosen is so large, an enormous number of respondents is required in order to gain insight into the distribution

[7] A hypothetical trend that could be found is: the answers of respondents under 25 years old cover a much larger area than the answers of people over 25.

of the individual associations. The fact that the associations represent very local experiences and that the area within which respondents were invited to participate was quite large (the city of Helsinki) increased the risk of having very dispersed input for the various associations. The risk might have been averted by the number of respondents that participated in the research. However, the design of the research does not seem suitable for a study focused on a limited number of issues or experiences within a smaller study-area.

3.2 Social landscape metrics

Social landscape metrics is a term coined by Brown & Reed (2012). It is a variation on traditional landscape metrics, which is used in the discipline of landscape ecology to analyse the distribution and relative positions of features in the landscape. Rather than determining the spatial structure of such landscape elements as animals and plants, social landscape metrics are used to analyse the distribution of people's preferences and associations related to their physical surroundings. The authors have termed these associations 'landscape values' which indicate locations that are of cultural, social or personal importance or otherwise related to people's perceptions and preferences (Brown & Reed, 2012). Like in the SoftGIS method, the respondents' preferences are collected in the form of points. The analyses is split in two categories: inductive and boundary social landscape metrics.

For the calculation of the inductive social landscape metrics, the density of the collected associations is calculated by interpolating the point data. When the density of points representing the same landscape value is higher than a certain threshold value, the points are considered part of a cluster (See Figure 3.2).[8] These clusters are subsequently used for the calculation of the inductive social landscape metrics. Inductive metrics provide insight into the number and average size and shape of the clusters of different landscape values. Furthermore, metrics can be calculated that indicate whether one type of landscape value is dominant in an area and what proportion of the area is covered by clusters of the various landscape values (Brown & Reed, 2012).

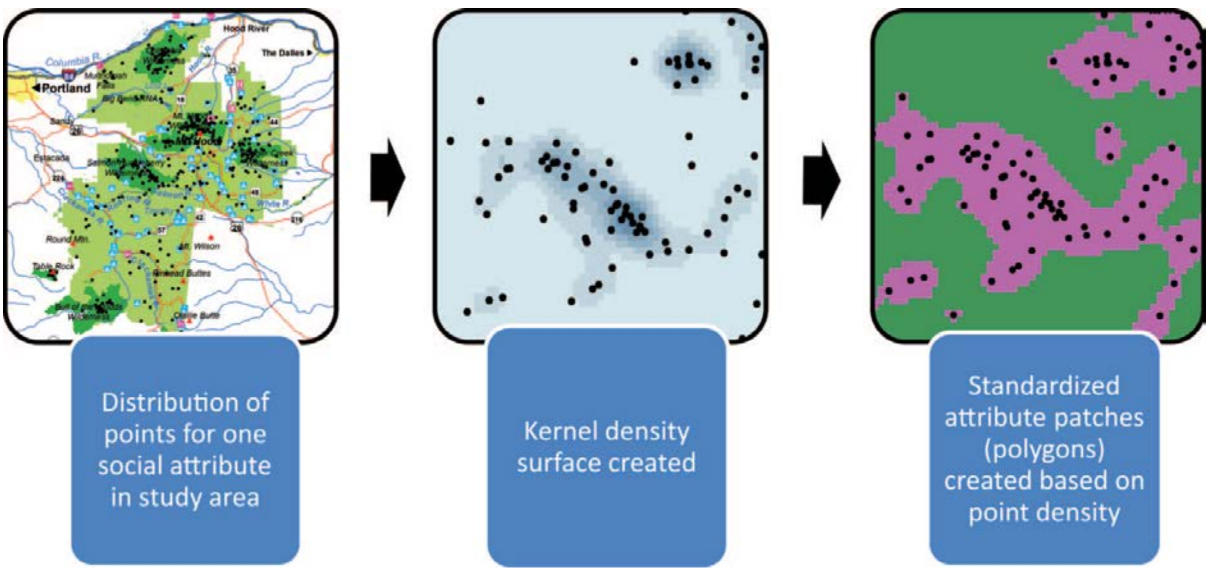


Figure 3.2: The creation of clusters (or 'patches') from point data representing landscape values (Brown & Reed, 2012)

[8] In both social and traditional landscape metrics these clusters are termed 'patches'.

The boundary category of social landscape metrics is, unlike the inductive metrics, not an adaption of traditional landscape metrics for use with PPGIS derived data. Rather, it is an attempt to provide insight into the distribution of landscape values within certain delineated areas of interest. For the calculation of the boundary metrics, the collected point data is intersected with administrative units. Subsequently, for each unit measures are calculated on the total number of collected landscape values and the ratio of the different landscape values within the unit. Also, on the basis of the estimated inherent conflict between different landscape values, a measure is calculated for the potential conflict within each unit. Furthermore, these measures are then compared between the administrative units in order to answer research questions such as “Where is the greatest concentration [of a certain landscape value]?”

The social landscape metrics can be used to gain insight into the locations and quantities of various landscape values. This information can be used in spatial decision making processes such as the allocation of resources. The authors see most potential in the use of social landscape metrics in the management of public lands, such as national parks and forests (Brown & Reed, 2012). With the collection of landscape values that better reflect urban environments, these metrics could also be of use in the context of urban planning.

Although the adaptation of traditional landscape metrics for use with peoples’ experiences and preferences of the landscape appears to be a useful approach to the use of PPGIS data in planning, some obstacles can be identified. Firstly, it remains to be seen whether the calculation of clusters on the basis of landscape values collected in the form of points results in meaningful outcomes. When the landscape values represent very specific experiences, the interpolation of the point data might obscure the variation in the data on a local scale. Secondly and in a similar vein, the aggregation of the collected data on the basis of administrative units for the calculation of the boundary social landscape metrics, might not provide information on the right scale for use in spatial planning. Administrative units usually don’t correspond well with the scale at which the processes underlying the respondents’ answers play.

3.3 GIS-MCDA

The research by Boroushaki & Malczewski (2010) focuses on another part of the participatory process than that of Brown & Reed, although some overlap exists. Where Brown & Reed are trying to gain insight into the spatial distribution of peoples’ values and preferences through a set of metrics and indices, Boroushaki & Malczewski suggest a procedure with which to measure the level of consensus between the various stakeholders on different planning alternatives. This procedure consists of a combination of GIS capabilities with multicriteria decision analysis, appropriately termed GIS-MCDA.

The implementation of GIS-MCDA in participatory planning can better the process in two ways: firstly, it aims to achieve consensus through the support of communication and the exchange of information between the various stakeholders. Secondly it provides analytical methods to come to a design that best reflects the requirements and interests of all the parties involved. The authors have termed these parts of the process the deliberative and the analytical dimension of collaborative spatial decision-making respectively (Boroushaki & Malczewski, 2010).

The GIS-MCDA procedure is developed on the basis of the notion that the capabilities of GIS and MCDA are complementary: the GIS components are utilized for the storage, management, analysis and visualization of geographical data. The multicriteria decision analysis methodology, on the other hand, provides “procedures, techniques and algorithms for structuring decision problems, and designing, evaluating and prioritizing decision alternatives” (Boroushaki & Malczewski, 2010, p. 323). Hence, the GIS elements of the procedure are mostly aimed at facilitating the deliberative aspect of the participatory planning process by providing citizens with access to relevant spatial information in the form of dynamic web-maps. In addition, WebGIS applications provide a platform for citizens and other stakeholders to exchange information, ideas, opinions and preferences regarding the spatial issues. The MCDA aspect supports the analytical dimension. It offers a structured environment within which the preferences stated by citizens can be analysed so as to gain insight into the levels of consensus and conflict in different areas. The level of consensus is indicated with a value between 0 and 1, where 0 means no agreement and 1 indicates full agreement between the stakeholders. The goal is to identify the source of the conflicts and draw op a solution that is acceptable to everybody involved (Boroushaki & Malczewski, 2010).

Methods for measuring consensus have been studied intensively in the field of multicriteria decision analysis. For their research the authors build on a model proposed by Herrera-Viedma *et al.* (2002) and have adjusted it to work with spatial data. They assert that an MCDA problem usually involves three main elements: a set of alternatives, a number of evaluation criteria and a group of stakeholders. In the spatial context, the various alternatives are interpreted as locations where a certain activity can be performed. The location can be represented by either a grid cell or a point, line or polygon. The preferences of the stakeholders are collected by allowing them to assign weights to the different criteria. Using multicriteria decision analysis the most appropriate solution is then sought by evaluating the alternatives on the basis of the weighted criteria. This way a group solution is generated on the basis of the preferences stated by the individual stakeholders. The group solution is usually presented in the form of a solution map.

Calculating the solution most acceptable to all stakeholders based on their individual preferences is a good way to look for common ground in the discussion between stakeholders. It should not aim to replace this discussion, however, as it does not take into account the fact that people can be convinced or inspired by arguments of others. Further research has to point out how well the calculated group solution corresponds with the final design.

3.4 Conclusions

This chapter describes three studies in which the analysis of PPGIS data plays a significant part. The context within which the approaches are developed vary drastically. Nonetheless a number of similarities between the analyses performed in the three studies can be identified. In the both the SoftGIS research and the social landscape metrics an important part of the analysis is aimed at providing insight into the spatial distribution of the provided answers. In both studies the collected point data is interpolated to visualize the areas where certain associations cluster spatially. This aspect is not addressed in the GIS-MCDA approach, where the interaction between the people’s preferences plays a central role. The relationship between the collected data however, can also be identified in the SoftGIS method and the social landscape metrics, in the analysis of the calculation of the percentage of positive values and the estimation of the conflict respectively.

These similarities have served as the inspiration for the development of the structured approach for the analysis of PPGIS data that is presented in the next chapter. Furthermore, a number of the analyses that are described in this chapter have been adapted and applied on the data that was collected during the empirical part of this research. These analyses are described in detail in chapter 8.

4. A structured approach to PPGIS analysis

As described in chapter 2, the increasing number of stakeholders in spatial planning has necessitated the development of methods to improve the communication between the parties involved. Depending on the specific application, PPGIS offers various capabilities that help stakeholders better communicate their preferences and opinions and as such improve their position in the decision-making process (Haklay & Tobon, 2003). This is especially important for citizens involved in participatory spatial planning processes, as the knowledge they possess - in the form of their experiences and preferences regarding their living environment – is very different form the expert knowledge most other stakeholders bring to the table. A lot of research has been done on the different methods for enticing citizens to actively participate in the spatial planning process. Only very few studies, however, focus on the translation of data collected among citizens into information that is usable in the spatial planning process.

This thesis argues that the analytical step in the interpretation of data collected among citizens is crucial to the effective application of PPGIS methods. A structured approach to the analysis of PPGIS data improves the efficiency with which the preferences and requirements of citizens are communicated to governmental agencies and other stakeholders in the planning process. However, such an approach does not appear to exist as of yet. Therefore, on the basis of the studies described in the previous chapter, a structured method for the analysis of PPGIS data was developed for the research reported in this thesis.

4.1 Three levels of analysis

The structure consists of three categories that have been termed the levels of analysis. These levels form a classification of the various analyses that can be applied to PPGIS data. The differentiation between the levels is based on the identification of three different types of relationships that have been identified within PPGIS datasets. In the following paragraphs the three levels will be explained in more detail. In the results chapter, analyses from each of the three identified levels are performed on the data that was collected during the empirical phase of this research.

4.1.1 Level 1

The analyses contained in level 1 are exploratory in nature. They are used to examine the spatial patterns of the data provided by respondents on one specific topic. The analyses in level 1 provide information on whether a certain phenomenon – such as a particular landscape value or a preference provided by citizens – is spatially clustered or dispersed. The boundary social landscape metrics indicating the density and frequency of a landscape value falls into this level of analysis. Also, the interpolation of the collected point data in the SoftGIS methodology and the subsequent visualization of clusters of positive or negative values can be considered a level 1 can be considered a level 1 analysis.

Insight into the spatial distribution of a phenomenon can be attained in several ways. The first approach is to use certain visualization methods when mapping the data, that provide an instant overview of where the density of the provided responses is highest. That way the data is prepared and depicted in such a way that it lends itself to quick visual examination by the user.

Decisions made in spatial planning processes, however, have to be substantiated. Consequently it is crucial that the distribution of the phenomenon is not only visualized, but is also quantified in some way. There are different analyses available to measure the level of clustering of geographical phenomena. These analyses determine whether the observed distribution of values encountered in the dataset deviates significantly from a distribution of the same values assigned by a random process. What analyses are suitable for the calculation of the amount of clustering depends on the type of data that is available (points versus polygons) and the type of information that is required. Global cluster analyses provide a single value to indicate the level of clustering for a complete dataset while local cluster analyses calculate a value for each input feature. The results of such analyses can substantiate claims about the level of clustering of experiences or preferences.

If background information is collected on the respondents during the data collection of the research, this allows for the examination of the answers provided by the various groups of respondents separately. In some instances, for instance, it is helpful to be able to distinguish between answers provided by men and women or people over or under 65 years of age. By applying a selection on the respondents on the basis of physical or socio-economic characteristics, the answers of groups of respondents can be compared and subsequently analysed.

4.1.2 Level 2

The analyses in level 2 can also be considered exploratory. However, instead of focusing on the distribution of a single phenomenon, these analyses aim to investigate the relationships between the collected data on different topics. By comparing the data collected on two or more phenomena, insight is gained on the level of consensus or the potential for conflict at various locations. Furthermore, the extent to which the answers to certain questions correlate provides an indication how experiences and their underlying processes are linked.

GIS-MCDA can be seen as a method that fits in the analyses of level 2. It compares the preferences of respondents based on their location and subsequently calculates a solution map that best fits the collected preferences and their assigned weights. The calculation of the percentage of positive values in the SoftGIS approach and the boundary social landscape metric measuring the level of diversity and potential conflict can be considered level 2 analyses.

There are several ways to examine the relationship between data on two or more phenomena. The first way is to make a comparison on the basis of two or more topics that are of particular interest. Such a comparison can focus on either the level of conflict or the degree of accordance. An example is the data on two contradictory experiences or preferences. By comparing the locations of the preferences expressed by the respondents, the locations can be identified where conflict is highest.

Another way to gain insight into the relationship between data on several topics is to focus on a specific area of interest and explore the ratios with which the various phenomena are represented. This approach is better suited for a general indication of what areas are viewed in similar fashion by most people and where people’s opinion and experiences of an area are more diverse.

Finally, the statistical relationships between data collected among citizens can be calculated using correlation analysis. The output of a correlation analysis describes the degree to which the relationship between to phenomena can be approximated by a mathematical equation. However, when significant correlation exists between two variables, this does not indicate a causal relationship between them. The results can be used to further investigate why certain phenomena show a clear statistical relation.

4.1.3 Level 3

The analyses of level 3 can be considered explanatory as they seek to answer questions on why the spatial distribution of phenomena and the relationships between them examined in level 1 and level 2 are observed. To answer those types of questions, the relationships are explored between the data collected among the citizens and data describing real-world phenomena. This can, for instance, be data on the characteristics of the physical environment (such as building density), distance to certain amenities or data on the locations of reported crimes.

The analysis in the SoftGIS approach calculating the effect of characteristics of the built environment on the possibility of a location being viewed positively or negatively is an example of a level 3 analysis. Also, the attempt to link the collected impressions of citizens to the land-use type at that location can be considered an analysis that fits within this level.

A firsts step in analysing the relationship between the collected data and characteristics of the environment is performing a correlation analysis. The results provide insight into the possible relationships that are present between the two datasets. As mentioned in the previous section, however, a significant correlation does not indicate causality.

In order to determine whether the preferences of citizens are actually caused by the characteristics of the environment, a regression analysis has to be performed. As opposed to a correlation analysis – where only the statistical coherence between two phenomena is calculated – regression analysis can give an indication of the influence of the real-world phenomena on the collected perceptions of the citizens. By describing these relationships, an attempt can be made to better understand the effects that certain spatial interventions have on the opinions and perceptions of the citizens.

4.2 Conclusions

The structured method described above and the accompanying analyses that fit within the three levels it comprises can be compared to a toolbox and the tools that it holds. The toolbox consists of a number of compartments that are designed to hold various tools used for similar purposes: one compartment holds the wood planers, chisels and files, while another holds flat head and Phillips head screwdrivers. As such, it provides a structure which helps the user identify what tools could be useful for the performance of a certain task. The method developed in this research for turning PPGIS data into valuable information for planning serves the same purpose: it structures the available analyses in a logical order. Based on the type of information that the users are interested in attaining from their data, they look into one of the three proposed levels for analyses that fit their purpose.

The analyses performed on the data can be seen as the tools. They are used to practically implement the measures that are desired by the user. Based on the characteristics of the collected data, some analyses are more suitable to attain certain information than others. Just as different types of screws require different types of screwdrivers. In the results chapter, analyses from each of the three identified levels are performed that are suitable for use with the data that was collected during research.

5. Methodology

In the previous chapter a structured approach for the analysis of PPGIS data was proposed. In order to assess the usability of the structure and its associated analyses it was necessary to apply it to a PPGIS dataset. Therefore, data was collected on the experiences of citizens of their living environment. This chapter describes the methods that were applied for the collection of this PPGIS data.

First, for the collection of the data, a PPGIS application was developed. This application allowed the conducting of a questionnaire, in which the respondents answered the questions by drawing areas into a virtual map. Its workings are described in detail in the next section.

Subsequently a survey was drafted that could be conducted through the use of the developed application. To increase the practical use of this research, the spatial planning department of the Heerlen municipality was contacted. A meeting was held to identify the topics that were of interest to them. On the basis of their interests, the issues to be addressed in the survey were identified. The contents of the survey are discussed in the second section of this chapter.

5.1 The PPGIS application

5.1.1 Phoenix software as the basis

For the collection of data for this research a Public Participation GIS application was developed.[9] The tool was based on the Phoenix software package which has been created at Geodan, a Geo-IT consultancy firm based in Amsterdam. The Phoenix software is itself a PPGIS application, aimed at providing a platform for discussions between the various stakeholders in spatial planning issues. This is achieved by providing its users access to geographical data in a wide array of digital formats through an easy-to-use interface. In addition, users can draw points, lines and polygons in the software, add labels to locations as a way to communicate arguments, and document the discussion for later use. These additions to the digital maps can be saved in formats that are supported by professional GIS software packages so that they can be used in analyses. The software can be used on PCs or laptops, as well as on devices controlled by touch, such as tablets and map-tables.

Since this survey focuses on the collection of spatial perceptions of citizens, the ease of use of the application was very important. Phoenix' interface is purposefully kept very simple for that same reason, making it ideally suited to serve as the basis on which to build the application for the data collection. The development of the application was necessary because, although Phoenix is in itself a PPGIS software package, it does not support the collection of data in the form of a survey. Therefore, the basic Phoenix functionality had to be extended with an application for conducting a questionnaire on spatial topics.

5.1.2 The 'SpatialQuestionnaire' application

For the purpose of this research an application was required that could function as a questionnaire on spatial issues: the user is presented with a set of questions, just like a regular web-based or paper

[9] The SpatialQuestionnaire application was developed in collaboration with Simeon Nedkov, a junior researcher at the Vrije Universiteit Amsterdam. He wrote the code for the software, while we jointly developed its functional design.

questionnaire. However, instead of providing answers in written text or by choosing an answer from a set of possible choices, users have to draw regions on the map that correspond with the question that is posed. One of the questions is, for instance, “Where in Hoensbroek do you find the greenery attractive?”. As an answer the respondents draw the area(s) on the map where the find the green characteristics of the environment attractive. The developed application was termed the SpatialQuestionnaire application.

In order for the application to function effectively and be comprehensible for its users, the possibilities and functionality had to be limited severely so that there could be no confusion as to what was required of the user in order to progress through the questionnaire. Thus, the application built for this research as an extension on the Phoenix software limited rather than expanded its basic functionalities. As such, the extension provided a framework within which users are guided through the questionnaire.

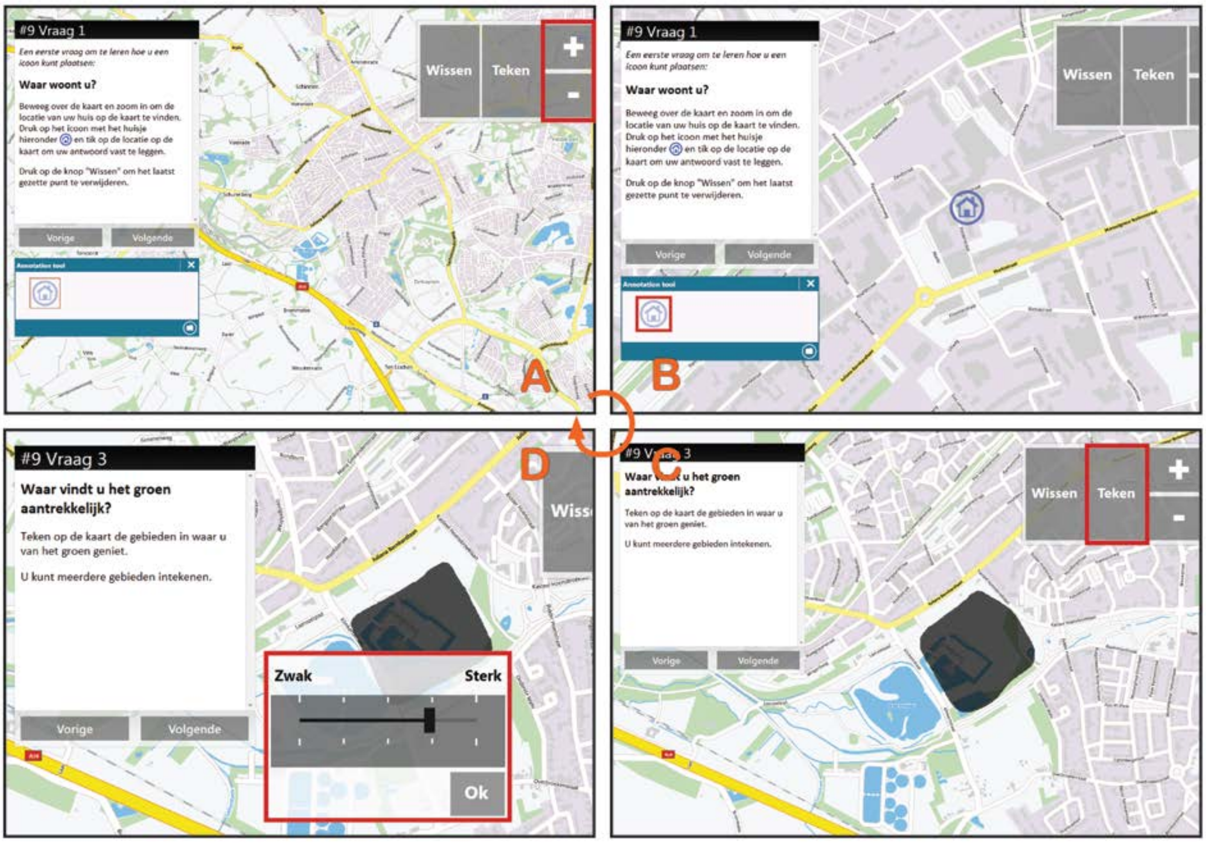


Figure 5.1: Functionality of the SpatialQuestionnaire application.
Important areas are highlighted in red
A: To answer the question, the respondent pans and zooms to the required location
B: To place a point, the respondent clicks the icon below the question and clicks at the required location
C: To draw an area, the respondent clicks the “Teken” (“Draw”) button and draws the contours of the area
D: An scale pops up, on which respondents can indicate how much they associate the chosen location with the experience addressed in the question, ranging from “Zwak” (“Weakly”) to “Sterk” (“Strongly”).

5.1.3 The functionality of the application

As can be seen in Figure 5.1: A, the SpatialQuestionnaire application starts with a very brief introductory text and a first question. It contains a very short explanation of the functionality of the tool and subsequently asks the respondent to identify their home location on the map. The text explains that a point feature, such as the home location, can be put on the map by clicking an icon representing the theme of the question which then lights up (Figure 5.1: B). By clicking the spot on the map where the user wishes to place the pin, it can be positioned on the map. Only a limited number of questions required the input of point locations.

The user presses the "Volgende" ("Next") button to continue to the next question. In the ensuing questions users are encouraged to draw as many areas as they can think of as answers to the posed questions. There is no limit to the number of areas that can be drawn or pins that can be positioned for a single question. Also – although the respondents were encouraged to provide answers within the Hoensbroek region - there is no defined area outside of which the input is considered invalid. However, it is not a requirement that an area is drawn in order to progress to the next question. As the aim is to gain insight into citizens' perception of their surroundings it would not make sense to force people to form opinions on certain themes. Therefore, if a respondent feels that a certain experience or association does not relate to any area in the Hoensbroek region, they can skip the question.

When answering a question, users can pan and zoom around the map in an intuitive manner, using the same controls as in popular applications such as Google Maps. When the user is ready to draw in an area on the map they press the "Teken" ("Draw") button (Figure 5.1: C). As soon as the drawing mode is activated the user can no longer pan or zoom. Any touch-input detected by the device is then registered as the drawing of the contours of an area. Consequently, users have to make sure that the full extent of the area that they want to draw is visible before they activate the drawing mode. When an area is drawn by the user, the polygon is automatically closed upon releasing the screen. It is therefore of importance that the user draws the contour of the area in one fluent motion, so as to prevent the polygon from being closed prematurely.

When an area is drawn or a pin is placed on the map, the users are asked to indicate to what degree they feel the area that they have chosen represents a certain experience or feeling. This feature is activated by letting a scale pop up when an area is closed or a pin placed (Figure 5.1: D). The scale is a semantic differential scale with five scale points reading from "Zwak" ("Weak") to "Sterk" ("Strong"). Areas that represent the experience, as indicated in the question, strongly receive a value of 5, while the areas that users only weakly associate with an area are given an value of 1. Somewhat strong and somewhat weak associations are given an value of 4 and 2 respectively whilst the middle scale point is scored as 3. The starting position of the scale is in the middle, so that the answers of people who have no intention to differentiate between them all get a value of 3. The scale can be closed by pressing the "OK" button, which saves the indicated scale point to that particular area or pin. Thus, users can adjust the level of intensity with which they associate the experiences for each area or pin individually.

Areas that are drawn incorrectly or misplaced pins can be removed by pressing the "Delete" ("Wissen" in Dutch) button. Users are also able to move back and forth between questions in order to add answers to certain questions at a later point in time. Once users reach the end of the survey, they are asked if they are indeed ready. If they confirm, their input is saved to an .phx file. This is a text file that saves the geometries of the provided answers and attribute data associated with each drawn shape or placed pin, such as the scale value provided by the user.

The number of questions can be adjusted through an .xml file. In this file it is also possible to indicate whether the required input from the user is in the form of areas or points. The content of the questions themselves is stored in separate .html files which are referenced in the .xml file containing the questionnaire's structure. Through these files the questionnaire can be modified to contain any number of questions concerning any number of themes. The content of the questionnaire is further discussed in the next section.

5.2 The survey

The survey comprised two separate parts. One part consisted of a paper questionnaire containing questions on a number of characteristics of the respondent, such as gender, age, education, income and how long they had lived in Hoensbroek (See Appendix I). This information was collected in order to be able to compare the answers provided by different groups of respondents. No names or addresses were asked so as to ensure the anonymity of the respondents. The paper also included a box into which the number of the electronic part of the questionnaire had to be written. It was essential that this number was copied correctly form the laptop or tablet used for the spatial part of the survey, as this was the way with which the respondents' characteristics could be linked to the preferences they expressed in the second part of the survey.

The second section of the survey required respondents to use the SpatialQuestionnaire application described in the previous paragraphs. Through the SpatialQuestionnaire the respondents were asked about a number of themes including services such as public transport, shopping and sporting facilities. Other questions were related to the respondents' experiences in the Hoensbroek area with regard to perceived beauty of the greenery or built environment and experienced nuisance, for instance as a result of inadequate maintenance of the surroundings or feelings of un-safety (see Appendix II for a full overview of the questions). The topics addressed in the questions asked through the SpatialQuestionnaire application were chosen after consultation of a representative from the spatial planning department of the Heerlen municipality.

The application and the questionnaire were tested among colleagues throughout the development process. Also a trial session with citizens was held in the Utrecht neighbourhood of Lunetten. The comments that were gathered during that session on the functioning of the application and the content of the survey were used to improve the application and the questionnaire.

5.3 The SpatialQuestionnaire vs SoftGIS

The data collection methods show similarities with the SoftGIS approach, although there are also a number of significant differences. The main resemblance between the approaches is the type of data that is collected among citizens. In both methods respondents are asked to indicate what places they associate with certain experiences. The questions that were asked using the SpatialQuestionnaire corresponded closely to the associations that could be chosen in the SoftGIS study. This means that similar information can be attained from the data collected with both methods.

Picture 5.1: Testing the application and questionnaire in Lunetten

However, with the SpatialQuestionnaire application respondents are asked explicitly on various experiences. They are asked to indicate what area they associated with an experience, instead of choosing an experience from a list of 64 possibilities. This means that the risk of getting very few locations for a certain experience is larger with the SoftGIS approach, as it does not actively encourage people to think about all the experiences addressed in the research.

Furthermore, with SoftGIS the experiences are collected in the form of points while the SpatialQuestionnaire is used to collect polygons. In order to create a reliable indication of the density of experiences collected in the form of points, many associations are required. By using polygons, the collected data can be transformed into a continuous surface for the assessment of the spatial distribution with much fewer respondents. How this is applied to the data collected using the SpatialQuestionnaire application is described in Chapter 7.

In both these cases the SoftGIS approach requires more respondents to attain useful information than the data collection method using the SpatialQuestionnaire. However, since the SoftGIS study is conducted over the internet the possibility of reaching many people in a short amount of time is much higher than with the SpatialQuestionnaire, which is conducted on site. Which method is more appropriate is therefore dependent on what the scale of the research is and whether specific issues are examined or a more general overview the people's experiences is desired.

5.4 Conclusions

For the empirical part of this research the SpatialQuestionnaire application was developed. In the application people are asked in the form of a survey to draw areas into a virtual map which they associate with various experiences. The issues addressed in the survey were determined in cooperation with the spatial planning department of the Heerlen municipality. The next chapter describes where and how the data was collected and provides information on the respondents that participated in the research. The results of the analyses performed on the data collected with the SpatialQuestionnaire application are presented in chapter 8.

6. Data collection and description

6.1 Case-study area: Hoensbroek, Heerlen

The data was collected in the Hoensbroek city-district in the Heerlen municipality. It is located in the Southern part of the province of Limburg in The Netherlands (Figure 6.1). This section introduces the area of Hoensbroek by briefly describing its history and geography. Subsequently the reasons for choosing Hoensbroek as the case-study area are explained. This is done by placing this research within the larger context of the AESUS research project and by introducing the concept of urban shrinkage and its link to the use of PPGIS data in spatial planning.

6.1.1 A short history of Hoensbroek

The village of Hoensbroek has its origins in the 13th century A.D. The castle is one of the biggest in The Netherlands and remains one of the most notable sights in the surrounding area. Hoensbroek remained rather unassuming until the beginning of the 20th century. It grew tremendously in size as a result of the two large coalmines: the Staatsmijn Emma in 1911 and the Oranje-Nassaumijn III in 1914. The Staatsmijn Emma was the second largest coalmine in The Netherlands. The arrival of the mines led to a great increase in jobs in the region and thus to an influx of miners and their families. The new neighbourhoods that were built to accommodate the miners' families were located between the village of Hoensbroek and the city of Heerlen, closing the distance between the two built-up areas. Nowadays Hoensbroek and Heerlen form one continuous urbanized zone. In 1981 Hoensbroek was integrated in the municipality of Heerlen, thereby reestablishing the political connection between them after almost six centuries (Rijckheyt, 2008).

The mines were closed in 1973, which meant that the primary economic force in the region disappeared. The impact of this event can still be felt today, as the region has struggled to find an alternative approach to stimulate the regional economy in the last decades (Intergemeentelijke Structuurvisie Parkstad Limburg, 2009). Currently the city-district of Hoensbroek has just under 20,000 inhabitants, a number expected to drop in the coming years (Dienst Landelijk Gebied, 2012).

6.1.2 Why Hoensbroek?

This research is part of the AESUS (Analyzing and Exploring Sustainable Urban Strategies) research project, which is in turn one component of the research program URD (Urban Regions in the Delta). In the AESUS project, the relationship between urban development and sustainability is examined. In this context, tools are developed to support the decision making process in a changing political and spatial context. Two regions were identified as case-study areas for the AESUS project: the city of Eindhoven and the city of Heerlen. These case-study areas were chosen for their contrasting characteristics. Eindhoven is a region with a growing economy, attracting highly skilled workers for its high-tech industry from around the globe. Heerlen, on the other hand, has seen a decline in population in the last couple of years. Its economy, once blossoming through the presence of the large coalmines, has been struggling ever since the mines closed in the beginning of the seventies. As a result, the citizens with higher education move away from the area in search of better employment opportunities in other parts of the country.

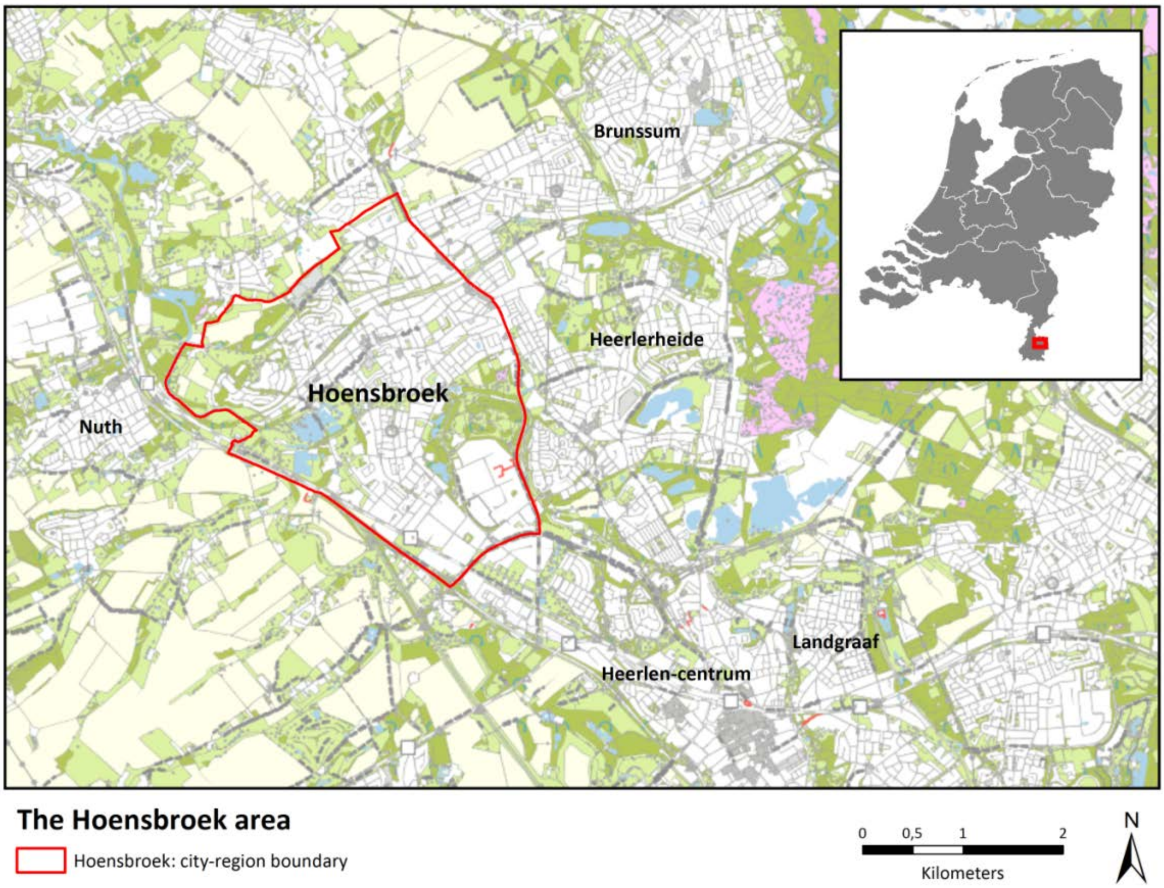


Figure 6.1: The area of Hoensbroek and its location in The Netherlands

It is clear that the economic and demographic contexts of Eindhoven and Heerlen result in very different spatial issues. In the one region choices have to be made regarding the location and type of housing and facilities for a growing population. In the other region the challenge is how to maintain the current level of public services with a declining number of inhabitants and what strategic locations can be identified for the demolition of houses to cope with the surplus in residences in the area. Theory suggests that public participatory approaches for spatial planning issues are particularly suitable in a context of population shrinkage. One of the main reasons for this is the financial burden that accompanies urban shrinkage: in regions where population decline occurs, the incomes in the form of taxes for local governments decrease while the cost of providing care and services for those who stay behind often increases. As a result, municipalities do not possess the funds required for the execution of elaborate redevelopment projects.

The inclusion of citizens from an early stage in the planning process offers a number of possible benefits in situations of population shrinkage. Firstly, partnerships can be formed between citizens, governments and other stakeholders to divide the costs of the development, execution and maintenance of projects over multiple parties. Citizens can not only provide monetary resources, but also provide labour or equipment to reduce the government's expenses. Secondly, the knowledge of citizens on the deficiencies of their living environment can be valuable in prioritizing for the allocation of the already scarce resources of the local government. And finally, actively including citizens in the spatial decision making process from an early stage can build acceptance of usually unpopular decisions that are the result of population decline such as the demolishing of houses

(Hospers, 2012). The occurrence of population shrinkage in the region of Heerlen, therefore made it an appropriate candidate to serve as study-area.

When the plan for the data collection was written, it became clear that the whole of Heerlen would be too large an area to cover in the limited time available. One of the anticipated consequences was that the answers provided by the respondents would be dispersed over such a large area that it would be difficult to draw any conclusions on the level of clustering of the perceptions of certain themes. This effect could possibly be increased seeing as this research is focused on very local phenomena such as the perception of people's surroundings. Consequently, the decision was made, together with representatives from the Heerlen municipality, to concentrate the efforts on the Hoensbroek area.

6.2 Data collection

The collection of the data in Hoensbroek was done over the course of four days which were spread out over three weeks. Since the setup used for the collection of data consisted of multiple laptops that required frequent charging and access to a wireless internet connection, the mobility was severely limited. As a result, the data was collected from a single location and the possible respondents consisted of the passersby. This method for collecting data much resembles the data collection techniques used in rapid appraisal participatory-GIS (RAP-GIS). In this method, participatory mapping is performed with groups or individuals in on-street events (Cinderby, 2010; Cinderby *et al.*, 2012). The difference between the approach used in this research and RAP-GIS, however, is that RAP-GIS commonly uses paper maps or physical models of the environment for the data collection instead of a virtual map or model.[10]

The data was collected from the Hoensbroek library. It provided enough sockets to charge multiple laptops and free internet access. Furthermore, the library is located in a central location on one of the neighbourhood's main roads near the market square. This assured a steady supply of Hoensbroek citizens to participate in our research.

6.3 Observations during data collection

During the collection of the data in Hoensbroek some observations were made. The main issue that was identified was that the time it took to fill out both parts of the survey varied drastically between the respondents. The paper part of the survey did not pose a problem for any of the respondents, most likely because it closely resembles the structure used by many other surveys in which some background information is asked. The section for which the SpatialQuestionnaire application was used, however, proved to be a bit more of a challenge for some. Because of the large differences in the time it took people to complete the survey, it was difficult to estimate the time it would take to collect the data from a certain number of respondents in Hoensbroek. The conclusion could be drawn fairly quickly that the first estimations of the required time per respondent were overly optimistic. As a result an additional visit to Hoensbroek had to be planned in order to collect data from a sufficient amount of citizens.

There are a number of reasons why some respondents had difficulty using the SpatialQuestionnaire application. First of all, the concept of answering questions by drawing areas

[10] The collected data is subsequently converted to a digital format for analysis using a GIS.

on a map is fairly new in general, so the chance that anyone had experience with similar methods was quite small. In other words, people could not build on previous experiences with regards to the concept of the questionnaire.

Secondly, the method requires people to orient themselves on a 2D map, which can be problematic in itself. Furthermore, respondents have to identify spaces they link to the experiences addressed in the questions on the map. Although this is of course inherent to the use of most PPGIS applications, it was not an easy task for many people. Adding familiar points of interest on the map could help improve this issue. Also the ability to search a location by its address would have been a useful addition to the SpatialQuestionnaire's functionality.

Thirdly, the difference in time required to complete the survey seemed to depend to an extent on the computer literacy of the respondent. Some respondents finished both the paper and the computer part of the questionnaire within eight minutes. Others, however, took up to 40 minutes to complete both parts of the survey. Although no records were kept on the time taken for the completion of the questionnaire, it was clear that younger people tended to take less time.

Finally, the issues addressed in the survey encouraged people to tell all kinds of stories that took place in Hoensbroek and the surrounding areas. So, aside from the degree of computer literacy, an explanation for the difference in time it took to fill out the questionnaire could be found in the amount of time spent on telling related stories. In the next section some more insight is provided into the people that participated in the research.

Picture 6.1: Data collection in the Hoensbroek library.

6.4 Data description

In total 85 respondents participated in the research. Together they provided 1391 answers in the form of drawn polygons. Due to the method used for the collection of the data, the group of respondents was not the result of random sampling of the Hoensbroek population. However, for the intent of this research random sampling of the population was not a necessity. What is of importance is whether the respondents are a good representation of the population so that the collected data is not biased due to the characteristics of the respondents. Therefore the group of respondents in this research should be both demographically and geographical representative for the population of Hoensbroek. The age and gender are considered when examining the demographic representativeness of the respondents, as these characteristics might influence people's perception of their surroundings more than for instance income or level of education.

Table 6.1 shows the gender of the group of respondents and the citizens of Hoensbroek in percentages. The table shows that the division of male and female inhabitants of Hoensbroek is almost perfectly balanced. In this research, 4 out of 85 respondents did not fill out their sex in the questionnaire. As a result the number of males accounts for just over 42 percent, while the females make up a little of half of the group. When only the people who provided an answer to the question on their sex are taken into account, men add up to 44,4 and women to 55,6 percent of the group of respondents. Although the division of gender in the respondents does not mirror the population exactly, it is not so skewed that the respondents are deemed unrepresentative of the population.

	Gender		
	Male	**Female**	**Not answered**
Respondents	42,4 %	52,9 %	4,7 %
Hoensbroek	49,7 %	50,3 %	-

Table 6.1: Gender of the respondents and the population of Hoensbroek in percentages.

In Table 6.2 the percentages of the different age groups of the respondents and the population of Hoensbroek are shown. As can be seen, the citizens under 25 are underrepresented in the research, while people over 65 years of age are overrepresented. The difference in percentages between the group of respondents and the population can be attributed to two things. Firstly, in the official statistics of Hoensbroek children are included in the calculation of the age categories. In this research, however, children did not participate (the age of the youngest respondent is 16 years old). This explains why the citizens under 25 appear underrepresented. Secondly, the research was conducted during weekdays mostly during working hours. This could be the reason why relatively many people over 65 years old are in the group of respondents, as they do not have to work during the day. It is conceivable that the overrepresentation of the age group of 65+ has impacted the collected data. Further research should be performed to assess whether this is the case. Since the rest of the age groups of the respondents correspond rather well with the division of the population, the assumption is made that the representativeness of the sample is not affected too much.

	Age category			
	Under 25	**25 - 45**	**45 - 65**	**65+**
Respondents	15,3 %	20,0 %	31,8 %	32,9 %
Hoensbroek	23,5 %	24,2 %	32,0 %	19,8 %

Table 6.2: Age groups of the respondents and the population of Hoensbroek in percentages.

Where people live in Hoensbroek might also impact their experiences and perception of their living environment for different topics. In order to assess the geographical representativeness of the respondents, therefore, the location of their house was used. Figure 6.2 shows the homes of the respondents throughout Hoensbroek. As can be seen, the respondents' homes are rather evenly distributed across the possible living areas of Hoensbroek. On the basis of this distribution, the conclusion was drawn that the respondents are also geographically representative of the population.

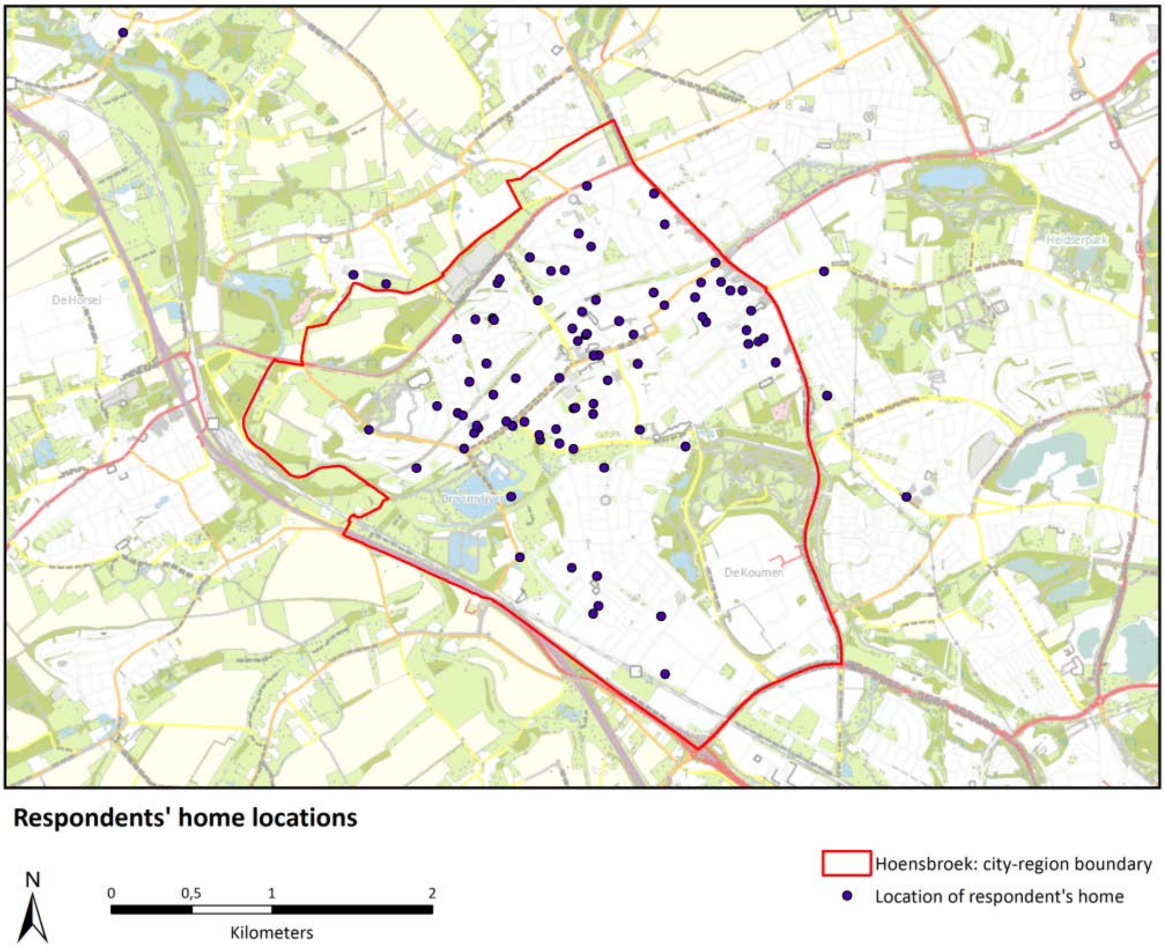


Figure 6.2: The locations of respondents' homes in (and around) Hoensbroek.

6.5 Conclusions

The empirical phase of this research was performed in the city-region of Hoensbroek in Heerlen. The demographic and economic processes that are at play there made it a suitable study-area for this research. 85 respondents provided over 1300 answers. The group of respondents were representative for the population of Hoensbroek on the basis of their gender and home locations. However, citizens over 65 years old appeared overrepresented among the respondents.

The data collection phase yielded a lot of data. It was necessary to prepare this data in such a way that analyses could be performed on it. How this was done is described in the next chapter.

7. Data preparation

The data collected during the empirical phase of the research consisted of polygons representing areas that people associated with certain experiences. When examining the spatial distribution of the data on a specific experience, the answers provided by the respondents thus consisted of overlapping polygons. These polygons had to be prepared in some way, in order to be able to use the data in spatial analyses. Making sense of these overlapping polygons turned out to be a rather complex issue. In this chapter the procedure of the preparation of the data is described.

7.1 Overlapping polygons

When it was decided what kind of analyses would be performed, the collected data had to be prepared in order to be usable in those analyses. The first step in this process was to find a way to gather the data from the .phx files containing the answers from each individual respondent.
A tool was developed to 'scrape' the information from the text-based .phx files and use the data to create a separate .csv file for points and polygons.[11] .csv files are a type of matrix with a row for each drawn polygon or point and columns for the different attributes for those features.[12] The .csv files can be imported into a GIS software package, where, consequently, it can be saved as a shapefile with all the respondents' answers as features and their associated attributes described above. Having the data visualized and projected in a GIS environment provided the opportunity to get familiar with the data, it structure and its possibilities for analyses.

Most of the required analyses revolve around the amount of people who associated a certain perception with a specific location. Therefore, the next step in the process was to manipulate the data in such a way that it was possible to see where the provided answers overlapped and subsequently how many respondents had that association for each given location. In other words, a way was sought to stack the overlapping polygons and count the number of polygons that overlapped each area. This could be achieved by adding an attribute to all features in the dataset with a value of 1, representing the area as an association with a certain question for a single respondent. [13]

Then a Union operation was performed, essentially slicing the polygons of a selection of the dataset into parts on the basis of the boundaries of the other polygons of the set. After performing the union operation, all the areas where multiple polygons overlap are identified in separate features. Each feature has all the attributes of the polygon it originally belonged to. However, in order to count how many polygons are in each location, the Dissolve operation has to be used. It aggregates features on the basis of equal values for chosen features and summarizes selected attributes of those features. In order to aggregate all features at a given location, the size and shape of the features have to be calculated. These attributes are consequently used as input on the basis

[11] The 'scrape' tool was written by Simeon Nedkov, junior researcher at the Vrije Universiteit Amsterdam.
[12] In this case the attributes consisted of the number of the questionnaire (to identify the respondent, as explained in section 5.2 on the survey), the number of the related question, a number to distinguish between multiple answers provided to one question by a respondent, the scale value the respondent attributed to that answer, and the coordinates of the polygons' vertices or the points.
[13] Although the respondents could indicate a degree to which they felt an area corresponded to a certain experience, it was decided to count each provided answer as a value of 1. In other words, the value at a location for a certain experience represented the number of people that indicated that they associated that location with the specific experience. This made the interpretation of the answers and the outcomes of analysis much more understandable.

of which the Dissolve operation is performed. When deciding which attributes to summarize the attribute with a value of 1 (added to all features as mentioned above) is chosen to be summed. The result is a polygon dataset with a value for the amount of respondents that have associated each location with a specific theme or experience. Figure 7.1 shows this result for question 3 of the survey on where the greenery in the area was considered attractive.

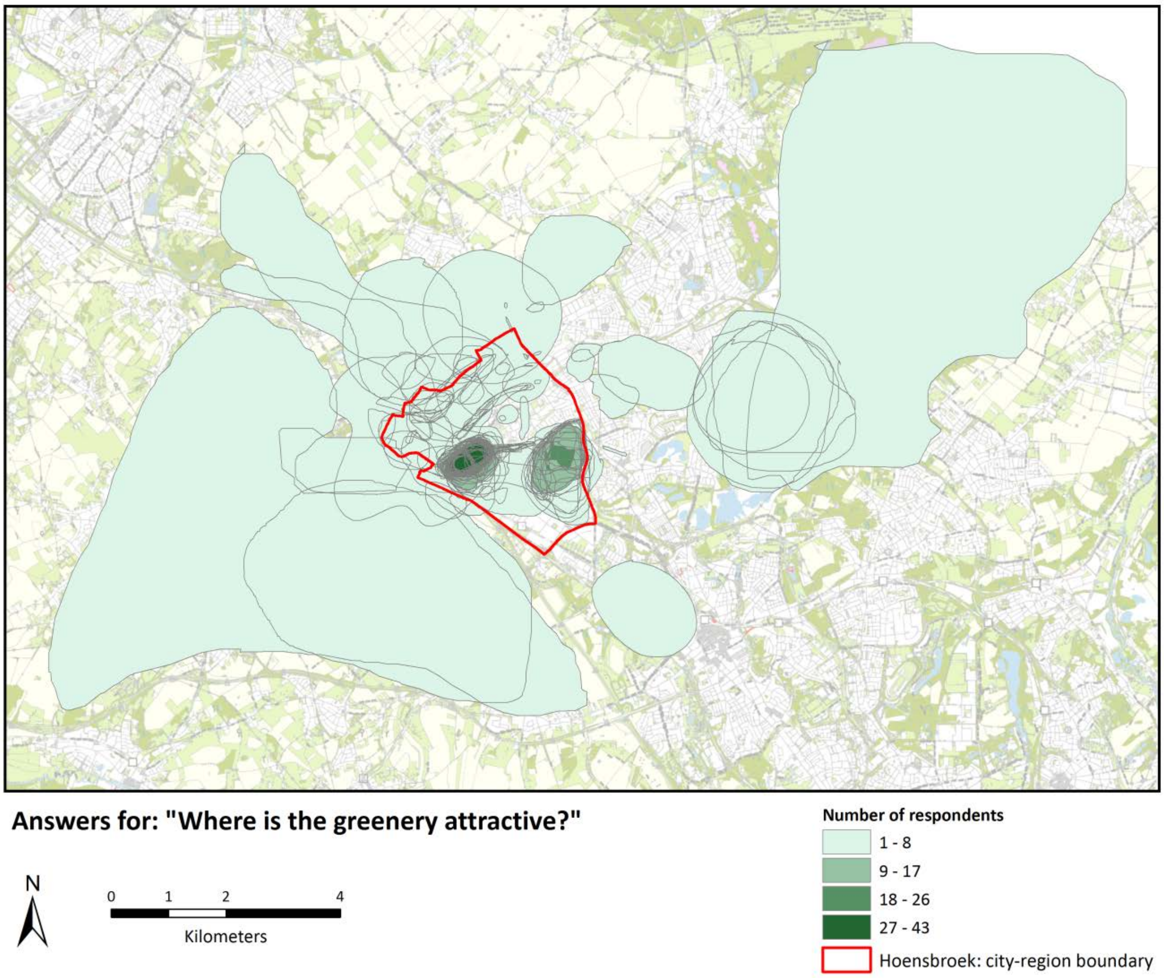


Figure 7.1: Answers provided to the question "Where do you find the greenery attractive?": full extent of the answers.

7.2 Slivers

Figure 7.2 depicts a close-up of the same dataset, focusing on the area around the Hoensbroek castle where many of the respondents' answers overlap. The map shows how the boundaries of the different polygons interact. The method was originally attempted because using the polygons drawn by the respondents was considered a good way to stay as close to the original collected data as possible. Also, since many spatial analyses in the ArcGIS software package require features as input, this appeared to be an efficient way to move from the raw data to datasets usable in various analyses.

However, as Figure 7.2 illustrates, the high number of overlapping polygons around a single area causes many very small polygons to be created. The fact that the polygons drawn by the respondents are divided into smaller features in places where they overlap is obviously inherent to this approach. A problematic consequence is that the number of small features and the size of those features is dependent on the number of overlapping polygons in an area. For instance, in the dataset depicted in Figure 7.1 the original number of polygons drawn by the respondents for question 3 is 130. After the Union and Dissolve operations were performed, this number rose to 4905 features.

Of those features, around 10% were smaller than 1 m^2 and the area of around 20% was smaller than 5 m^2. These small polygons are called ´slivers´. Seeing as the smallest area drawn as an answer to that question was 3275 m^2, the discrepancy between the sizes of the original polygons and the areas resulting from the operations could be considered too large. Furthermore, since these slivers are formed in areas with a high number of overlapping polygons, their average value is relatively high. When performing spatial analysis on the created dataset, the conclusion was drawn that the average value was influenced disproportionally by the high number of slivers and their associated high values. Finally, the realization that this problem is intrinsic to this particular approach of dealing with overlapping polygons and that the problem is only exacerbated when dealing with datasets containing more polygons, led to the decision to prepare the data for analysis in another way.

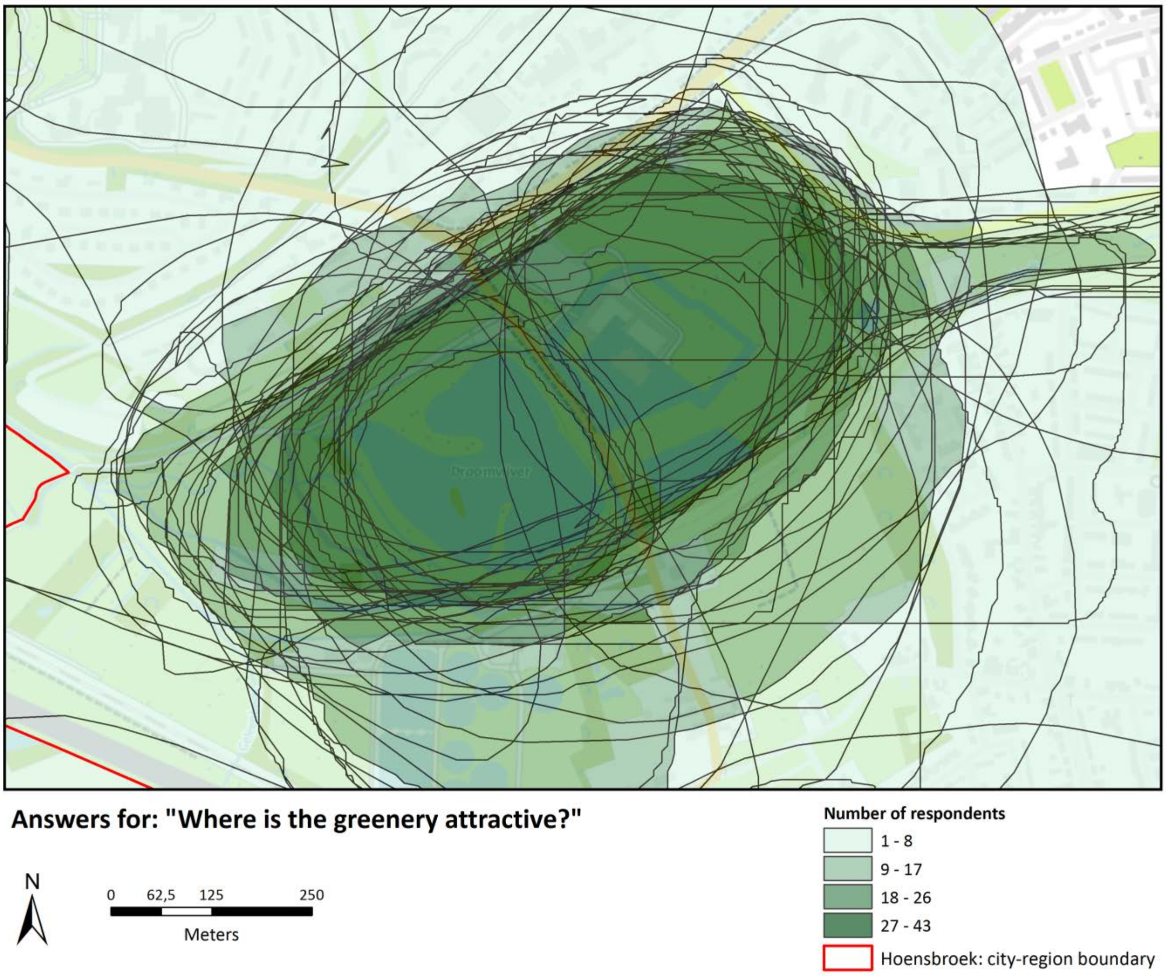


Figure 7.2: Answers provided to the question "Where do you find the greenery attractive?": zoomed in.

7.3 Rasterizing the data

When it was decided that using the overlapping polygons in the manner described above was not a viable option, a new way of spatially dividing the area was sought. One possibility was to use administrative boundaries in the area, such as the neighbourhood boundaries in Hoensbroek or the postcode areas to aggregate the data. However, the conclusion was drawn that these geographical divisions were too crude and would therefore not yield usable results for spatial planners. Consequently the decision was made to make a spatial division of the study-area by applying a grid.

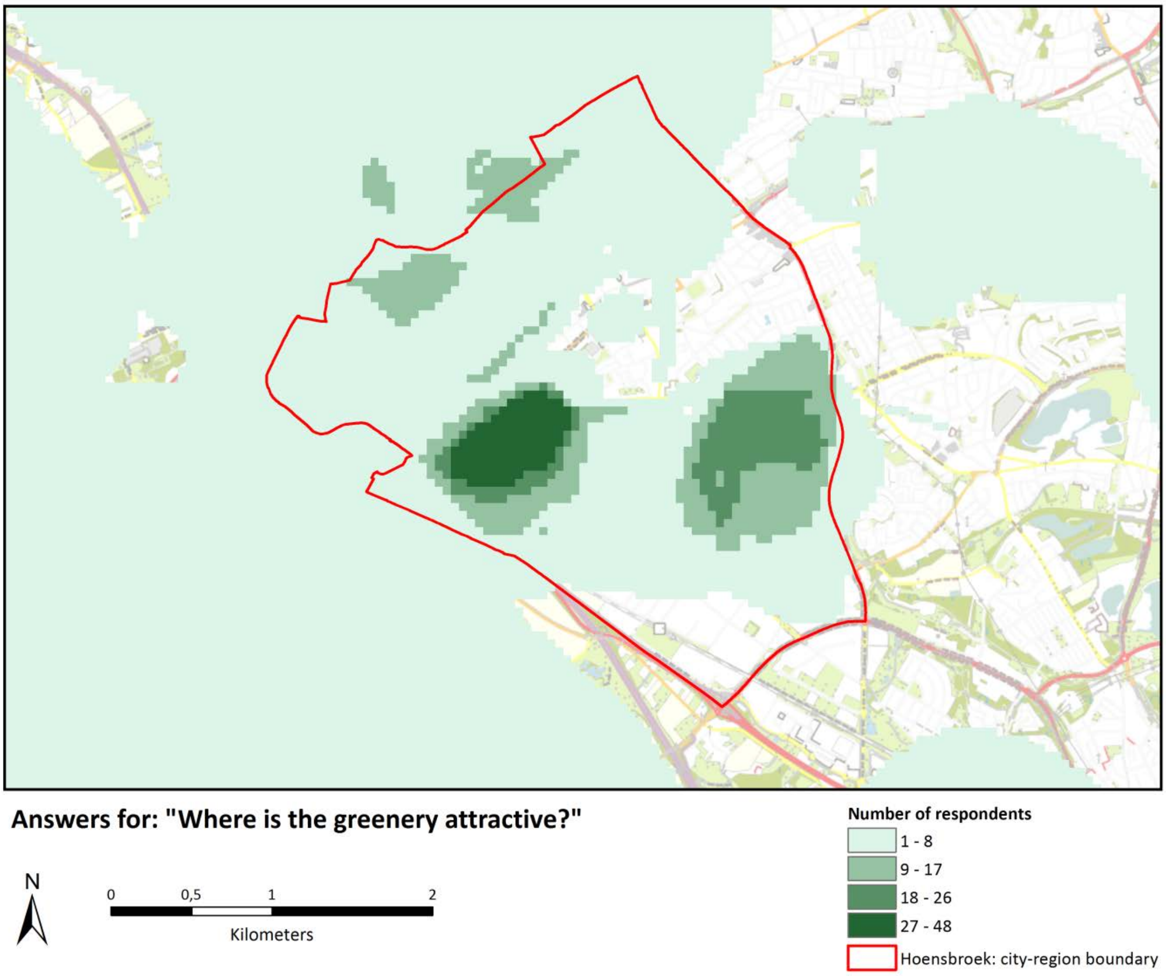


Figure 7.3: Answers provided to the question "Where do you find the greenery attractive?": rasterized.

This has the advantage of providing the possibility to examine the data at various spatial levels. Figure 7.3 shows the same data as shown in Figure 7.1 after transforming it to a 50 by 50 meter grid.

Rasterizing the collected data has another advantage. Point data is usually interpolated in order to make claims on the density of incident data. This mean that a continuous surface is created which indicates the amount expected incidents at each location on the basis of collected incidents in a specified area surrounding that location. When the data consists of polygons instead of points, however, it is not necessary to use interpolation techniques in order to create a continuous surface. When the original answers are rasterized and the rasters for a certain question are summed (similar to the approach described with the overlapping polygons) a continuous surface is created where each cell has a value corresponding to the number of rasters that are stacked on top of each other.

Furthermore, raster data allows calculations that are not possible to perform on polygons and these calculations are much less computationally intensive than many analyses on shapefiles. On the other hand, certain analyses require polygons as input. Therefore it was important to produce both a raster dataset and a gridded polygon dataset based on the rasters. That way the advantages of both data types could be utilized and all analyses could be used on the data.

To achieve this, a tool was developed in python to rasterize all individual features on the basis of the coordinates of their vertices in the .csv files.[14] This allows the setting of the required resolution of the raster and the selection of the attribute used as the cell value. It generates a set of raster files for all selected answers. The tool was then extended to immediately sum the individual rasters to create a raster file representing the number of rasters present at each location. A selection can be made to indicate which features are used in the calculation of the summed raster.

As a polygon can store many more values per feature than a grid cell by ascribing it a shear unlimited number of attributes, it is possible to create a polygon grid that contains all the data available in the separate rasters. In order to create such a polygon grid, first the grid structure has to be generated on the basis of the original rasters. In order to extract the values of the rasters, a point dataset needs to be generated based on the centroids of the polygon grid. These points then also form the centroids of the rasters (assuming that all the rasters used are perfectly aligned). It is then possible to extract the values of all the desired rasters to the point dataset simultaneously. Finally, the attributes of the point dataset (now containing all the data from the chosen rasters) can be joined to the polygon grid cells on the basis of their geographical location.

When both of these steps have been carried out, both a set of rasters is available with the their values representing the sum of the number of respondents that have identified that area with a certain experience as well as a polygon grid containing all of the rasters data as values for its attributes.

7.4 Conclusions

When it was concluded that using the original polygons of the respondents' answers was not a viable way of preparing the data for analysis, the decision was made to rasterize the collected data. For this purpose a tool was developed to extract the polygons to a raster on the basis of their vertices described in the .csv files. Subsequently these rasters were summed to create a raster image with a value for each cell representing the number of people that associated that location with the experience addressed in the survey. A number of analyses,, however, require input in the form of features instead of raster files. Therefore the values of the constructed rasters were extracted and added to a matching polygon grid. In the end the prepared dataset thus consisted of a polygon grid containing all the rasterized data for the separate questions of the survey and raster files for the individual questions. The next chapter describes what analyses were performed on this data and presents their outcomes.

[14] This rasterization tool was again written by Simeon Nedkov, junior researcher at the Vrije Universiteit Amsterdam.

8. Results

In this chapter the structure proposed in the Methodology chapter is implemented by translating it into concrete analyses. The analyses are described and applied to the data collected in Hoensbroek. The outcomes of the analyses are then interpreted and evaluated on their value for use in spatial planning. Due to the amount of questions in the survey and the resulting amount of data, it is impossible to discuss in this thesis the results of the analyses for all the data collected. These results will be presented in a separate report to the policymakers in Hoensbroek.

Instead, in order to provide coherence in the performed analyses and their associated outcomes, one theme was selected: the areas where respondents found the greenery attractive. Since the analyses follow the structure from the Methodology, the results tell a story in which the steps follow each other in a logical order. This is beneficial to understanding and interpreting the found results.

8.1 Level 1

As described, the analyses of Level 1 of the structure revolve around interpreting the answers to the individual questions provided by respondents. Exploring the spatial distribution of the answers can provide insight into the areas that are of interest for the specific themes addressed in the various questions. Three methods were selected to examine the spatial distribution of the data: heat maps, cluster analysis and typology. In the following sections these methods are described and the results from their application to the collected data are interpreted.

8.1.1 Heat maps

Introduction

The first step in the data exploration is to create heat maps from the respondents' input. Heat maps are similar to choropleth maps in that they represent the value of an variable with graded differences in shading or colour. Choropleth maps, however, represent data by enumeration units (such as municipalities, neighbourhoods or postal code areas) while heat maps are used for the presentation of continuous data.[15] Another difference is that the colour scheme for choropleth maps are usually limited to a range of lighter to darker shades of a single colour, whereas heat maps use colour schemes where both the hue and the lightness of the colours vary. As the term heat maps suggests, their colour schemes typically mimic a temperature-like visualization.

Heat maps are an effective method to visually represent the distribution of values for a certain variable because it shows where areas with high and low values are located at a glance. A reason for this is that people can distinguish between different colours better than they can between different shades of lightness. This means they can identify more details in a map using a colour scheme with multiple colours (Kraak & Ormeling, 2010).

However, a few problems arise with the use of heat maps. The first is that, while people associate levels of lightness of a colour with a natural ordering of the values, no such ordering is perceived between different colours. As a matter of fact, some colours appear more prominent than others in people's eyes (such as yellow and cyan), creating the risk that attention is drawn to areas that are of no particular importance. It is therefore wise not to use too many colours in a heat map and to select colours with a similar level of perceived intensity. Another problem that is encountered

[15]Choropleth comes from the Greek χώρος and πλήθος meaning 'area' and 'multitude'.

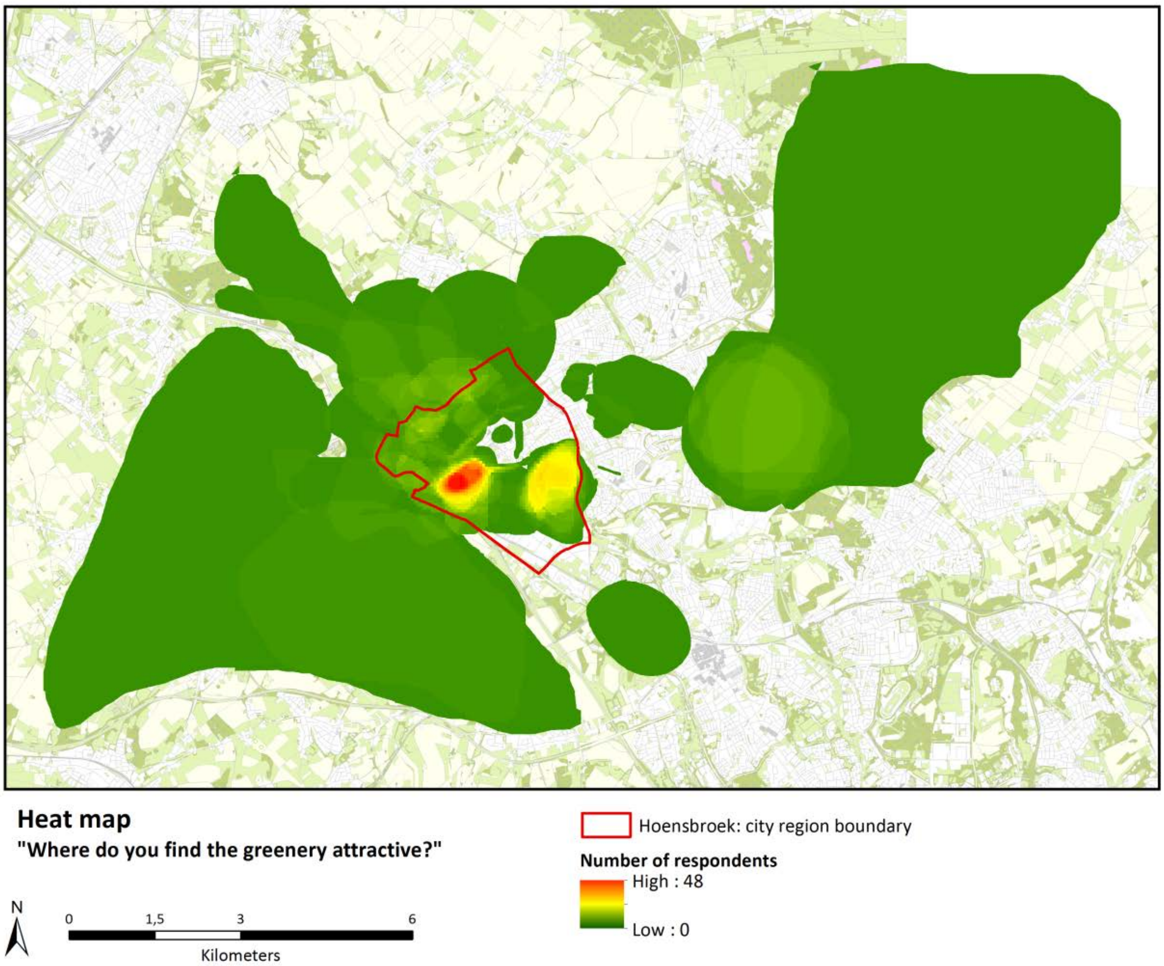


Figure 8.1: Heat map of the answers to the question "Where do you find the greenery attractive?": full extent.

with both heat maps and choropleth maps is that the representation of the data can be altered dramatically with changes in the colour scheme and the chosen classification (Monmonier, 1996). This must be kept in mind when interpreting data on the basis of heat maps or choropleth maps.

Hoensbroek results

Figures 8.1 and 8.2 show the data collected in Hoensbroek represented in the form of a heat map. As stated in the Data preparation chapter, the value of the grid cells indicates how many respondents have drawn a polygon as an answer to the question "Where do you find the greenery attractive?" at that location. It is, as such, a count of the answers provided by people for each location.

The green areas indicate low values and cells with high values are coloured red. Figure 8.1 shows the complete extent of the answers provided by the respondents to the question. A relatively large number of the answers lay outside the boundary of the city-region of Hoensbroek, with a single respondent even including a chunk of German territory in the answers. This was possible as there was not limit set on the level to which respondents could pan or zoom when answering the questions.

Although the focus of the research is on Hoensbroek itself, it is worth looking into the answers provided outside of the study-area. The areas with the highest values are within the Hoensbroek boundary, but important information might be overlooked if the answers are automatically clipped to the boundaries of Hoensbroek. If, for instance, people consistently travel

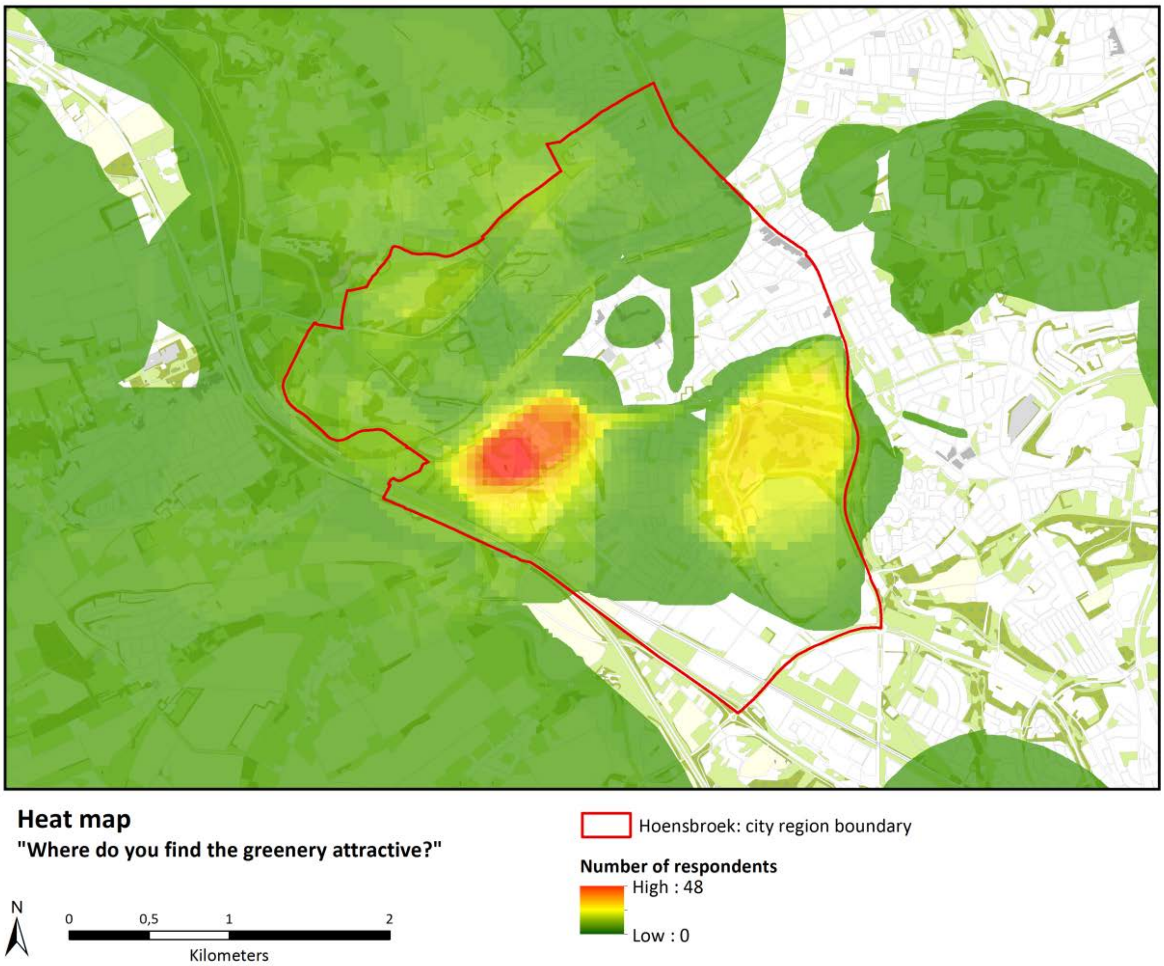


Figure 8.2: Heat map of the answers to the question "Where do you find the greenery attractive?": Hoensbroek.

outside of Hoensbroek for the use of certain amenities, this would not show up on a map representing solely the data within the Hoensbroek boundaries. In this case the yellow area to the East of Hoensbroek (which corresponds with the Brunssummerheide) indicates that quite a few Hoensbroek citizens enjoy the nature in that area. Also, the area neighboring Hoensbroek in the North-West in the direction of Vaesrade and Amstenrade shows some rather high values.

Figure 8.2 shows us the same dataset, but zoomed in to the study-area. Also, the transparency of the collected data is increased to allow for easier identification of the geographical places that correspond to the answers. It is apparent that the highest values are concentrated around the Hoensbroek castle and the Koumenberg, incidentally the two largest green areas within the Hoensbroek boundaries. The small protrusion in the North-East corner of the red area around the castle lines up with a deer enclosure and a neighbouring narrow park, located along the south of the KasteelHoensbroeklaan. This narrow strip of greenery connects the park around the castle with the Koumenberg. Other areas that show relatively high values for the appreciation of their greenery are the green belt along the old mine railway tot the North of the castle and the patches of greenery just North of the ring road near the Northern boundary of Hoensbroek.

Heat maps offer a quick overview of the collected data. The heat maps can be used to visually examine the spatial distribution of the data and identify the areas with clusters of high and low values. These areas can, in turn, be used as starting points for further investigation. Figures 8.1 and 8.2 can provide the answer to the question “Where do the respondents find the greenery attractive?”. A follow-up question might for example be “Why do people find the green at some locations more attractive than at others?”. After examination of the heat maps, it could be decided that an investigation is made of the features that distinguish the area around the Hoensbroek castle from the green spaces perceived as less attractive.

Although the heat maps provide a certain amount of insight into the spatial distribution of a phenomenon, they do not indicate where the areas with high or low values exactly begin or end. In order to answer those questions, a statistical analysis of the data is required.

8.1.2 Cluster analysis

Introduction

In order to get a more objective measure of what areas exhibit spatial clustering of values, several spatial analyses can be performed. A number of decisions have to be made when deciding which analysis to use. Some analyses examine the clustering of incident data, while others are created for use on the values of features that make up a geographical division of the study-area. Since the respondents’ answers were turned into a continuous raster surface during the data preparation, the second type of analysis is appropriate for the Hoensbroek data .

Two much used analyses to determine whether the values of features are spatially clustered, dispersed or depict a random pattern, are the Moran’s I and Getis-Ord G indices. Both analyses are inferential, which means that the null-hypothesis of the analyses is that the values of the features are distributed randomly across space. The outcomes of the analyses show measures which indicate whether or not the null-hypothesis can be rejected (the z-score)[16], and the associated level of confidence with which this can be done (p-value)[17]. Though the Moran’s I and Getis-Ord G analyses are capable of answering comparable questions with regard to the spatial distribution of values, there are some differences in their approach and consequently in the results that they produce. In this research the Moran’s I analysis was used. This decision is explained in the following paragraphs.

The Moran’s I analysis has both a global and a local variant. The global measures calculate a single value to indicate the level of clustering within the dataset between -1 and 1. When the Moran’s I analysis yields a statistically significant result in the form of a sufficiently low p-value, the null-hypothesis of a random distribution of the values is rejected. When the associated Index value is negative, it indicates a dispersion of the values. A positive index value signifies a clustering of values. In other words, a Moran’s Index of -1 indicates perfectly dispersed values and an index of 1 indicates

[16]The z-value can be a positive or a negative value. It represents the number of standard deviations the observed distribution of values deviates from the expected distribution on the basis of complete randomness.

[17]The level of confidence is indicated with the p-value, or probability. It indicates the chance (with a value between 0 and 1) that the observed spatial distribution of values is the result of some random process. A small p-value, as a result, indicates a small chance that the distribution of values resulted from a random process. When the p-value is lower than a selected threshold value (usually 0,05 or 0,01) the null-hypothesis of the analyses can be rejected. For example: when the p-value is 0,05, the null-hypothesis of random spatial distribution of values can be rejected with a confidence of 95%.

perfectly clustered values (see Figure 8.3).[18] Global cluster analyses can be used to answer questions such as "are the values of the features spatially clustered?".

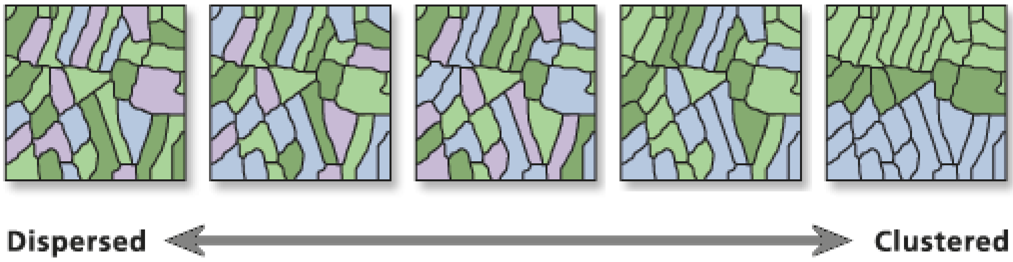


Figure 8.3: Moran's I analysis: a negative value indicates a dispersed pattern of values, a positive value clustering.

A local cluster analysis produces a value for each individual feature, indicating whether it is considered part of a statistically significant cluster or considered an outlier. In this calculation the value of the feature under scrutiny is compared to the values of features in its surrounding area.[19] The local Moran's I Index depicts for each feature whether its value is similar or dissimilar to the values of neighbouring features, and whether the values of the neighbourhood deviate significantly from the average value for that variable in the whole dataset. The null-hypothesis of the analysis is that the value of the feature is the result of a random process. Rejecting the null-hypothesis thus means that the value of the feature is either part of a statistically significant cluster of similar values or an outlier compared to the values of neighbouring features.

If the local Moran's I analysis produces a low p-value in combination with a positive z-value, it indicates that the neighbouring features have similar values to the analysed feature, meaning the feature is part of a cluster. This cluster can comprise of features with either high or low values. A low p-value with a negative z-value, on the other hand, shows that the feature is an outlier compared to its neighbouring features: either a high value surrounded by low values or a low value surrounded by high values. In other words, a statistically significant p-value for a local Moran's I analysis indicates that a feature forms either an outlier or a cluster (shown in Figure 8.4 and 8.6). What type of outlier or cluster the feature represents, is calculated on the basis of the observed values of the feature itself and its neighbouring features. The possible cluster and outlier types are: HH (a cluster of high values), LL (a cluster of low values), HL (a high value surrounded by low values) and LH (a low value surrounded by high values) (as shown in Figure 8.5 and 8.7). As this research focuses on very local processes such as people's perception of an area, the ability to identify spatial outliers was a reason to select the Moran's I analysis over the Getis-Ord G.

---

[18]The results from the Getis-OrdG analysis are interpreted a little differently: a significant p-value with a positive z-value indicates the spatial clustering of high values, while a negative z-value demonstrates the clustering of low-values.

[19]For the calculation of both the global and local Moran's I, it is necessary to determine what features are considered part of the neighbourhood of another feature. The best estimation of the conceptualization of the spatial relationships of the data is based on the underlying processes that determine the distribution of the data. As the data is based on people's perception of a location, it was determined that the neighbourhood should not be too large (around 100 meters). As a result, it was decided that the 24 cells closest to the feature are considered part of its neighbourhood.

Before the Moran's I analyses could be performed, the area to be included in the analyses had to be determined. The area is of importance as choosing to include or exclude certain features impacts the average value of the dataset in the calculations. Changes in the average value affect what features are identified as statistically significant clusters and outliers. The respondents were encouraged to focus their attention on the Hoensbroek area when providing answers to the questions, but areas drawn outside of this region were not considered invalid. This made the choice as to what area to select as delineation of the area to consider in the analysis a little intricate.

After consultation with an expert on spatial statistics[20] the decision was made to address the issue of determining the area used in the cluster analyses in two separate ways. The first approach was to look at all the input provided by the respondents as answers to a question. This meant that the study area was formed by all the polygon grid cells that had a value of at least 1 for a certain topic (meaning that at least one person associated that location with a certain experience). Practically, this meant that the polygon grid – which contained the data on all the questions – was clipped based on the answers provided to a specific question as shown in Figures 8.4 and 8.5.

As a result, no grid cells with a value of zero were considered when looking into the clustering of the data of that theme. This decision was made because it proved to be impossible to define a study area containing all the areas relating to a specific question in a structured manner without creating an enormous number of cells with a value of zero.[21] By clipping the polygon grid on the basis of the polygons of the original input provided by the respondents, clustering within the answers can be calculated.

The second approach focuses on the area of Hoensbroek specifically. It considers the city-region to be the study-area and so the answers are clipped on the basis of its boundaries. The results for this approach are shown in Figures 8.6 and 8.7. This means that all the provided data outside the Hoensbroek boundaries are excluded from the cluster analysis and that the grid cells within the study area with a value of zero are taken into account (as opposed to the previously described approach). In this case the fact that no respondents associated certain areas within Hoensbroek with a particular experience was considered a choice, justifying the decision to include cells with a value of zero. It also resolves the issue caused by large polygons drawn by one or two respondents in the calculations of the clusters. These large areas with very low values influence the average value of the complete dataset and might thereby impact the analyses performed (similar to the impact of large regions with a value of 0 as described before). This method was adopted in order to achieve greater insight into the distribution of the experiences in Hoensbroek itself.

[20] Consultation took place with Daniel Arribas-Bel. He is currently a postdoctoral researcher at the Vrije Universiteit Amsterdam and has previously worked as a Postdoctoral Research Associate at the GeoDa Center for Geospatial Analysis and Computation.

[21] An attempt was made to define the study-area by calculating the minimum bounding geometry (envelope, rectangle or convex hull) but the geographical distribution of the answers was such that this did not yield usable results as the amount of resulting grid cells with a value of zero influenced the average value of the dataset too much. The grid cells of zero form a problem in this situation, because it cannot be determined whether all respondents considered the same area when providing answers to the questions. By including the cells with a value of zero, those areas might very well be overrepresented and not reflect the respondents opinions correctly.

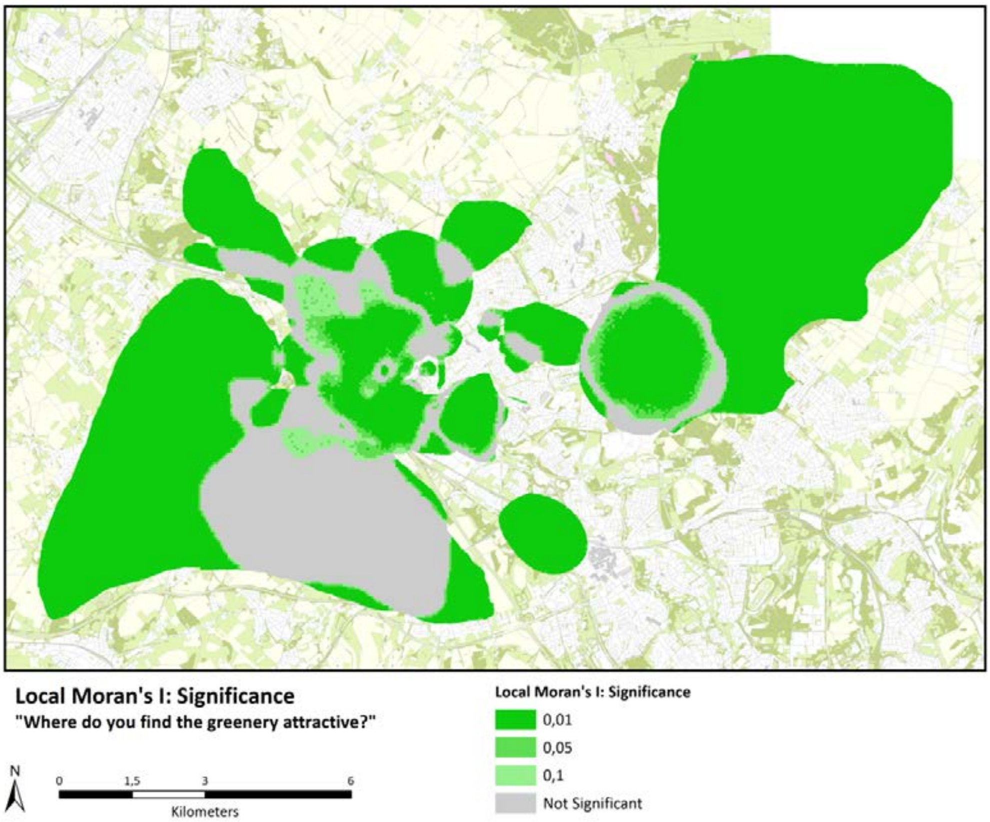


Figure 8.4: Local Moran’s I significance: full extent

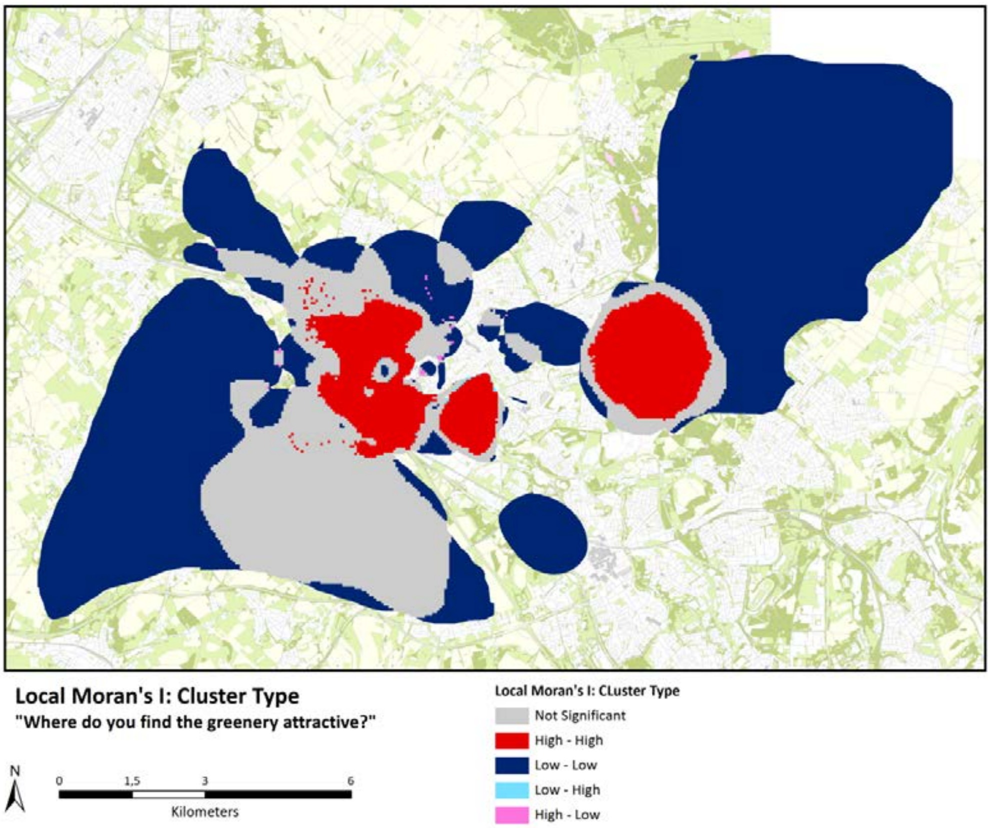


Figure 8.5: Local Moran’s I cluster types: full extent

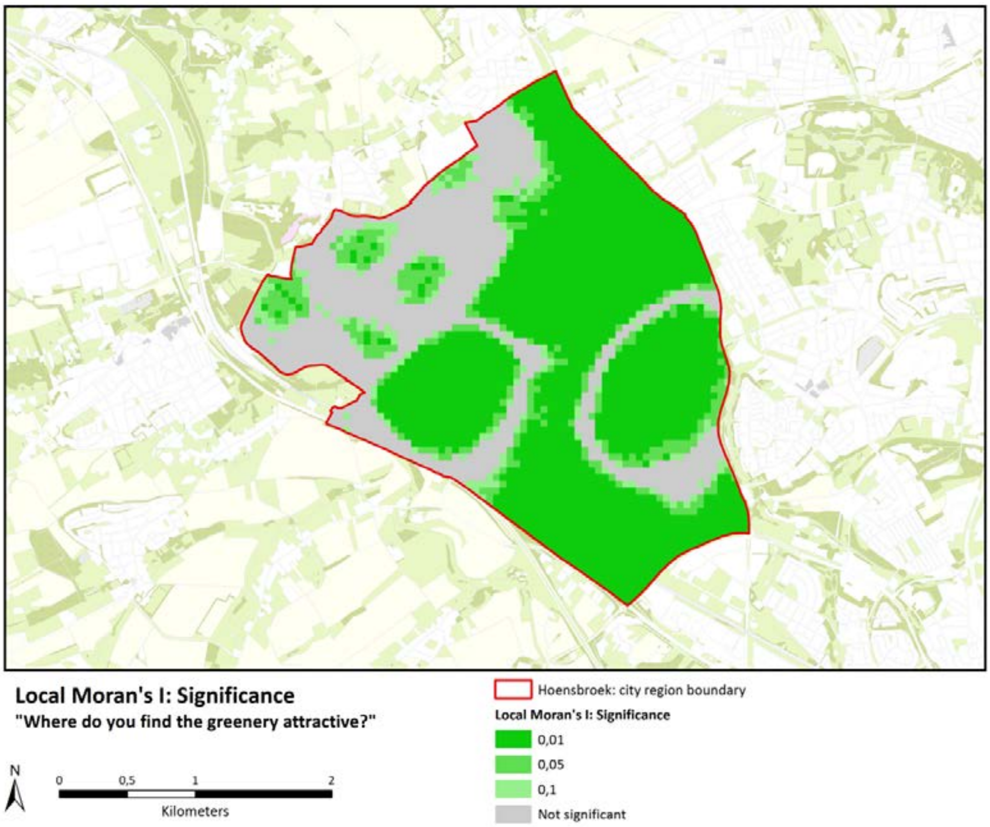


Figure 8.6: Local Moran's I significance: Hoensbroek

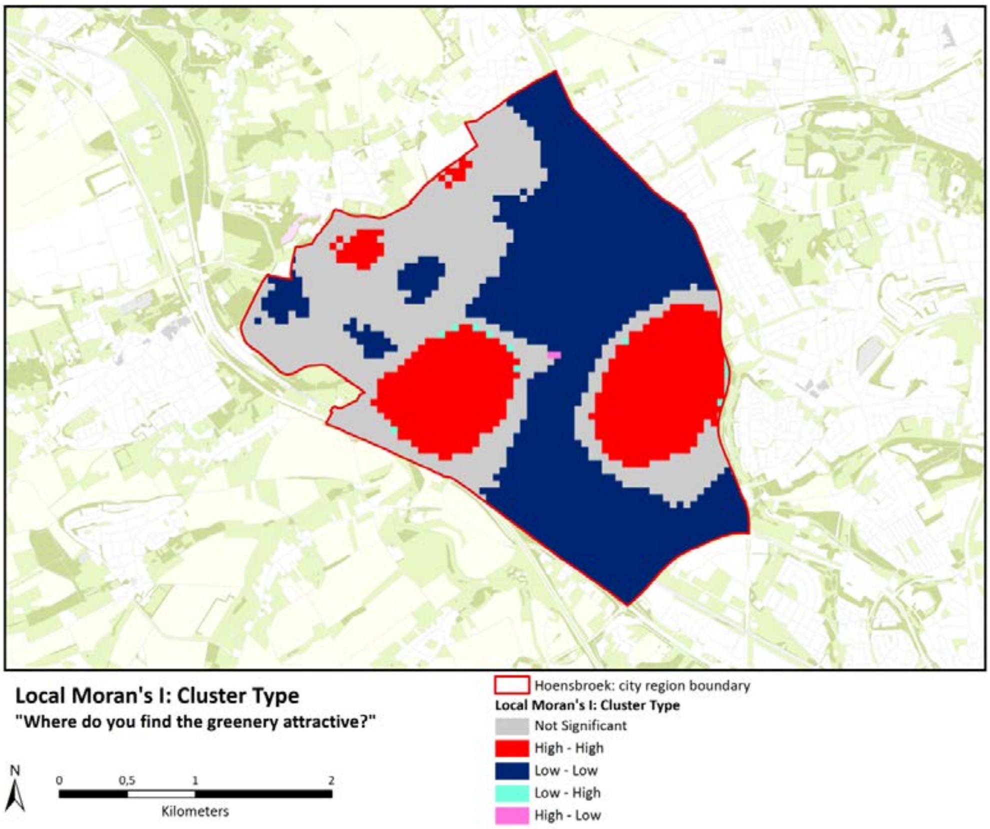


Figure 8.7: Local Moran's I cluster types: Hoensbroek

Hoensbroek results

For the performance of the cluster analyses on the collected data the GeoDa application was used. It is a user-friendly software package for exploratory spatial data analysis developed at the GeoDa Center for Geospatial Analysis and Computation (for a detialed description see Anselin *et al.*, 2005).

For the dataset using the entire extent of the given answers the index of the global Moran's I is 0,94 and its associated p-value is 0,00. The Moran's I index value indicates that the values are indeed spatially clustered. The p-value specifies that the null hypothesis of Moran's I (a spatially random distribution) can be rejected with over 99% certainty. These values signify a high degree of spatial clustering within the answers provided by the respondents. However, for the purpose of this research the location of the clusters is highly important.

The results from the local Moran's I analysis are shown in Figures 8.4 and 8.5. Figure 8.5 only shows the cluster types of clusters that have a significance of 0.05 and lower. It Is clear from Figure 8.4 that the large areas outside of Hoensbroek with a value of 1 are significant at the 0.01 level, indicating clustering, or the presence of outliers. The map depicting the cluster types indicates that these areas are considered clusters of low values. Three large clusters of high values are observable, focusing around the Hoensbroek castle, the Koumenberg and the Brunssummerheide. Part of the cluster surrounding the castle is the area to its North, where areas of relatively high values were also identified on the basis of the heat maps. A number of high outliers surrounded by low values can be identified in the cluster type map in the form of pink specks. Although these were not identifiable in the heat maps, upon further inspection they correspond to very specific locations indicated by one or two respondents.

The index of the global Moran's I for the dataset clipped on the basis of the Hoensbroek boundary is 0,92 and its associated p-value is again 0,00. Unsurprisingly, these values also indicate a statistically significant level of clustering in the data. This finding is underpinned by the outcomes of the local Moran's I shown in Figures 8.6 and 8.7. Figure 8.6 shows the significance level of the local Moran's I analysis on the dataset, while the cluster types are visible in Figure 8.7 When looking at the dataset clipped on the Hoensbroek boundary, some similarities are visible compared to the dataset discussed above. The main parallels are the clusters of high values around Hoensbroek castle and the Koumenberg. Upon closer inspection, however, a few differences emerge.

First of all, the clusters of high values are smaller compared to the other dataset. The prominent high-value clusters in the North-Western part of Hoensbroek, as seen in Figure 8.5, are now reduced to a few small areas in Figure 8.7. In fact, some areas that were part of the cluster of high values are now considered low-value clusters. In addition, the statistically significant areas of low values in the Hoensbroek area are much larger.

The differences in the outcomes can be explained by the two changes in the dataset. Because of the inclusion in the analysis of all the cells within the Hoensbroek area with a value of 0, these areas are likely to be considered a cluster of low values. This explains the larger area of clusters of low value cells in Figure 8.7. On the other hand, by clipping the dataset on the Hoensbroek boundaries and thereby excluding a much larger number of cells with a value of only 1 or 2, the average value in the dataset has actually increased as compared to the data clipped on the extent of the answers. As a result, higher values are necessary in order to stand out as a statistically significant cluster of high values, thus resulting in the reduction in size of the high value clusters as identified by the local Moran's I.

By applying a global cluster analysis to a dataset, the distribution of the values can be quantified. If the null-hypothesis of spatial randomness of the cluster analysis is rejected, this means spatial clustering of values occurs within the data. Using a local cluster analysis, the location of the various cluster and outlier types can subsequently be identified as well as their statistical significance. The location of the clusters can be used to substantiate arguments for the prioritization of some areas when addressing spatial issues. The level of confidence serves as a statistical underpinning of those arguments. Furthermore, the statistically significant clusters and outliers can be used to compare the spatial distribution of values between datasets, which is impossible to do based on just the information provided by the heat maps. Finally, the location of the clusters can provide insight into the underlying processes that are at the cause of the clustering of the values.

8.1.3 Typology

Introduction

A disadvantage of aggregating the data by rasterization - or by performing a union and dissolve operations on the input polygons successively for that matter - is that the information linking the individual answers to the respondents is lost. This means that a selection has to be made on the basis of respondents' characteristics and the topics of interest before the aggregation of the data is performed. Visualizing the answers of a selection of the respondents provides information on the preferences or experiences of a group that is of particular interest. Also, the selected data can be used as input in other analysis such as the conflict analysis described further down. Being able to select the answers on the basis of the characteristics of the respondents provides the opportunity to explore how the experiences of certain groups of people relate to one another.

Hoensbroek results

In Figure 8.8 the answers are shown for the question "Where do you sometimes feel unsafe?" divided by the gender of the respondents. What is immediately apparent is that there are many more answers drawn by women than by men, even when the ratio of male and female respondents is taken into account. What is also interesting is that many women have indicated sometimes feeling unsafe in the centre of Hoensbroek, in the area near the market square and the main shopping street. Most answers from men, on the other hand, are located further away from the centre, instead corresponding to more remote and less populated areas in Hoensbroek.

The results from the typology on the basis of gender of the answers to the question "Where do you sometimes feel unsafe?" can be used when deciding what areas to focus on when addressing the issues of feelings of un-safety of women. Also, a comparison can be made on the basis of this information between the experiences of male and female respondents and the amount and type of crimes that are committed in the Hoensbroek area. This can provide insight into the relationship between actual occurrences of crime and the feelings of unsafety at those locations. Furthermore, the information gained from the division of the provided answers on the basis of personal characteristics of the respondents can be used as input for further analysis, for instance by examining the spatial distribution of the answers of the separate identified groups using a cluster analysis.

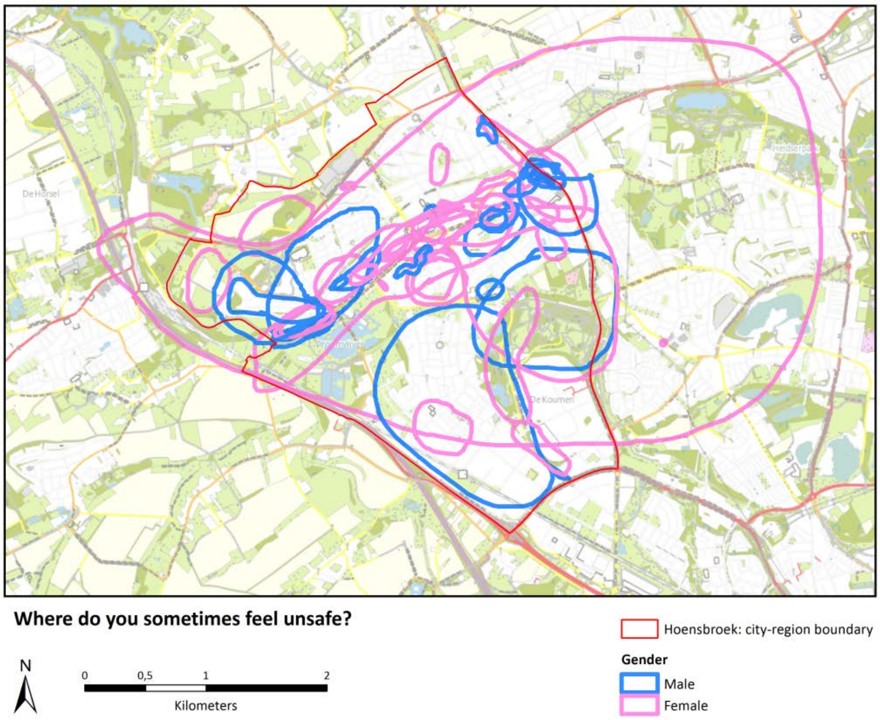


Figure 8.8: Typology of the answers to the question “Where do you sometimes feel unsafe?” based on gender.

8.2 Level 2

After having looked into the answers to the various questions, we want to know how the answers relate to one another. The association that one respondent has with a certain area could be completely different from another respondent’s opinion or experience with that same location. Also, the data might show a certain connection between the answers provided to different questions. The relations between these experiences can be quantified, to gain insight into the degree of conflict or accordance between respondents’ associations with an area.

Several methods are available to investigate how and to what degree the experiences and perception of the respondents are related. The first way is to make a selection of a set of topics in which the data is either contradictory or corresponding and identify where the answers on these topics overlap. Another possibility is to select a location within which all the provided answers are divided into positive or negative associations, which uncover what types of experiences it evokes. The third method is to calculate the correlation between two answers.

8.2.1 Conflict analysis

Introduction

With the use of the conflict analysis the level of dissonance between two variables can be examined. It is based on the assumption that certain combinations of experiences or preferences contain a level of disagreement. By calculating and visualizing this disagreement, insight can be gained into the degree of conflict that is apparent in the respondents' answers and where this conflict occurs.[22] The conflict analysis is performed by applying the following calculation for each location:

$$Degree\ of\ conflict\ = \frac{lower\ value}{higher\ value} * c$$

The c represents the conflict factor inherent in the chosen set of variables. It is a value between 0 and 1, with 0 indicating complete agreement and 1 indicating complete conflict. The conflict factor is implemented since the level of conflict inherent to sets of variables is not constant. For instance, answers to an opposing set of questions such as "Where do you find the buildings attractive?" and "Where do you find the buildings unattractive?" are interpreted as being completely opposing and would therefore receive a conflict factor of 1. Many sets of experiences, on the other hand, or not opposites of each other and are thus assigned conflict factor with a lower value, thereby limiting the maximum attainable value of the degree of conflict.

Subsequently, the ratio between the two conflicting datasets is calculated on the basis of the values at each location. If the number of people associating an area with a certain experience is far greater than the number of people associating it with an opposing experience, it makes sense that the level of conflict is considered lower than when the number of people associating a location with both opposing experiences is almost equal. In order to keep the proportions consistent, despite the possibility of one experience of the opposing set having a higher value in some locations and the other experience in others, the ratio is calculated by dividing the lowest of the two values by the highest.

Hoensbroek results

Figure 8.9 shows the conflict map of the answers provided to the questions "Where do you find the greenery attractive?" and "Where do you sometimes feel unsafe?". The maximum value for the level of conflict that can be identified in the map is 0,5. This is also the value of the conflict factor that has been assigned to this set of variables. The assigned conflict factor is based on the assumption that experiences of the appreciation of greenery and the feeling of unsafety are not completely conflicting. It is, for instance, conceivable that green areas such as parks are appreciated by people during the daytime but are avoided by the same people during the night. As such, the experiences do not rule each other out, but rather have an impact in each other because of a certain level of dissonance. As a result , a conflict factor of 0.5 was selected for the description of the relationship between these experiences.

The map shown in Figure 8.9 is therefore calculated with the equation:

$$Degree\ of\ conflict = \frac{lower\ value}{higher\ value} * 0.5$$

All the areas for which the degree of conflict has been calculated are areas where the answers for the questions "Where do you find the greenery attractive?" and "Where do you sometimes feel unsafe?" overlap. The large areas with very high levels of conflict around the edges of the mapped

[22] This conflict analysis is an adaption of the analysis of potential conflict mentioned in the section on the boundary social landscape metrics in section 3.2.

area are due to the low values for both the variables at those locations (usually a value of 1 for both variables).[23]

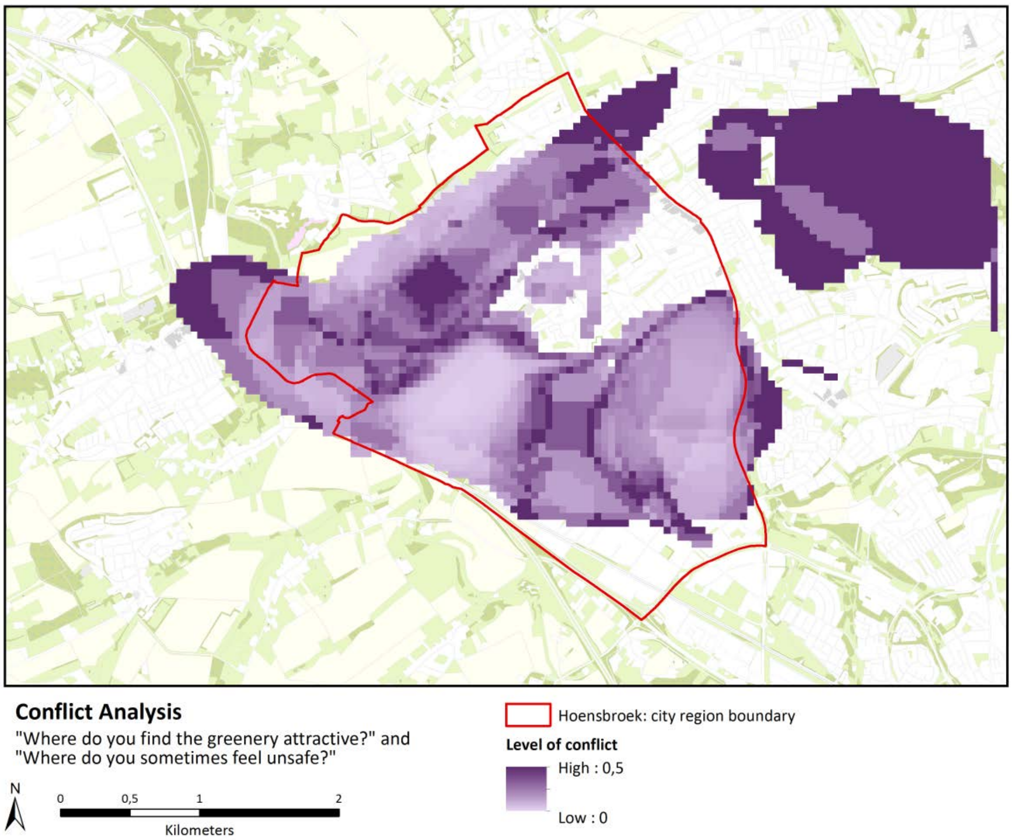


Figure 8.9: Results from the conflict analysis between the answers to the questions "Where do you find the greenery attractive?" and "Where do you sometimes feel unsafe?".

Due to the structure of the equation the areas with very high values for the appreciation of the greenery show low levels of conflict as the ratio between the variables decreases. This becomes apparent as the level of conflict increase towards the edges of the clusters of high values for the greenery appreciation. Within these areas, however, a number of locations can be identified where the level of conflict is higher, apparently due to a higher value for the experienced feeling of un-safety. A good example is the banana shaped area on the west side of the eastern cluster for the appreciated greenery. The level of conflict increases towards the Southern edge of the data.

Other areas showing higher levels of conflict are located in more built-up areas towards the West of the Hoensbroek city-region. A recognizable square depicts a cluster of high conflict values. This area corresponds with part of a neighbourhood called Maria-Gewanden. A number of people appreciate this area for its greenery, probably due to its spacious design and abundance of trees. Apparently, however, this area also evokes feelings of un-safety which could possibly be attributed to the same characteristics. The Hoofdstraat and the trailer park at the Western end of the

[23] A threshold value could be determined to exclude the locations where the value for both or either variable is considered to be not high enough.

Hoofdstraat also show high levels of conflict with regards to the appreciated greenery and feelings of un-safety.

The conflict analyses shows us the extent to which conflicting experiences overlap and in what ratio the two variables are present at each location. This provides insight into the areas where disagreement between the experiences or preferences is largest. These areas can be used as a starting point to further examine the relationship between the two variables. Furthermore, having information on the level of disagreement between sets of variables for different locations can be used to predict what topics are likely to cause disputes between stakeholders in the spatial planning process.

8.2.2 Percentage of positive values

Introduction

When an area is considered of particular interest, we want to know more about the satisfaction of the respondents with certain features of that location. A method to gain some insight into this phenomenon is to calculate the summed value of the positive associations as a percentage of the total summed value of the answers provided for that area, as shown below[24]:

$$General\ satisfaction = \frac{sum\ of\ positive\ values}{sum\ of absolute\ positive\ and\ negative\ values}$$

This approach works best when a several questions deal with a central theme.[25] In this research a several questions were asked about people's satisfaction with the quality of certain services in the Hoensbroek area. Using data on the answers to a set of such questions concerning a particular theme, information can be acquired as to how positive the location is generally regarded for that subject.

Hoensbroek results

In the survey eight questions were included concerning the perceived quality of services in the Hoensbroek area. The respondents were asked to identify the areas where they found the quality of four types of services to be either good or lacking. The services were chosen together with the planning department of the Heerlen municipality on the basis that they could be expected to be available in an urbanized area such as Hoensbroek. The four services were sports facilities, shopping possibilities, public transport and bars and restaurants.

Figure 8.10 shows the results of the calculation of the percentage of positive values for the Hoensbroek region. Two main areas can be identified where the respondents have indicated to have mostly negative associations with the services. The area in the South of Hoensbroek corresponds to the neighbourhood of NieuwLotbroek-Zuid. It is flanked to the South-East by an industrial park called De Koumen. This area is quite far from the shopping facilities in the center of Hoensbroek and has not been reachable by public transport since the bus connection was discontinued a number of

[24] This analysis is comparable to the percentage of positive values calculated in the SoftGIS method, which is presented in section 3.1.

[25] The calculation requires that input can be given on a positive phenomenon and its negative counterpart. When respondents are not asked specific questions but rather can freely place associations of opposing sets (such as in the SoftGIS research), it is not possible but not necessary to select a theme to perform this analysis.

years ago. A comparable situation applies for the region in the North-West of Hoensbroek with a low percentage of positive values. This has probably contributed to the negative associations regarding the level of services in those neighbourhoods.

Remarkably, the area around Hoensbroek center, with all its shops and restaurants, still has less than 50 percent positive values. Although the services are present at those locations, apparently they do not meet the expectations of many respondents regarding their quality. The areas with a high percentage of positive values, on the other hand, can be considered up to the standards of the respondents and possibly be used as examples on which to base the services necessary in the other regions.

Information on the percentage of positive values thus provides insight into the general satisfaction with a certain theme for the different areas. The information can serve as a starting point for further investigation into reasons for the perceived quality in different areas. Furthermore, it can be used in the allocation of resources and efforts for the improvement of the quality of the various services.

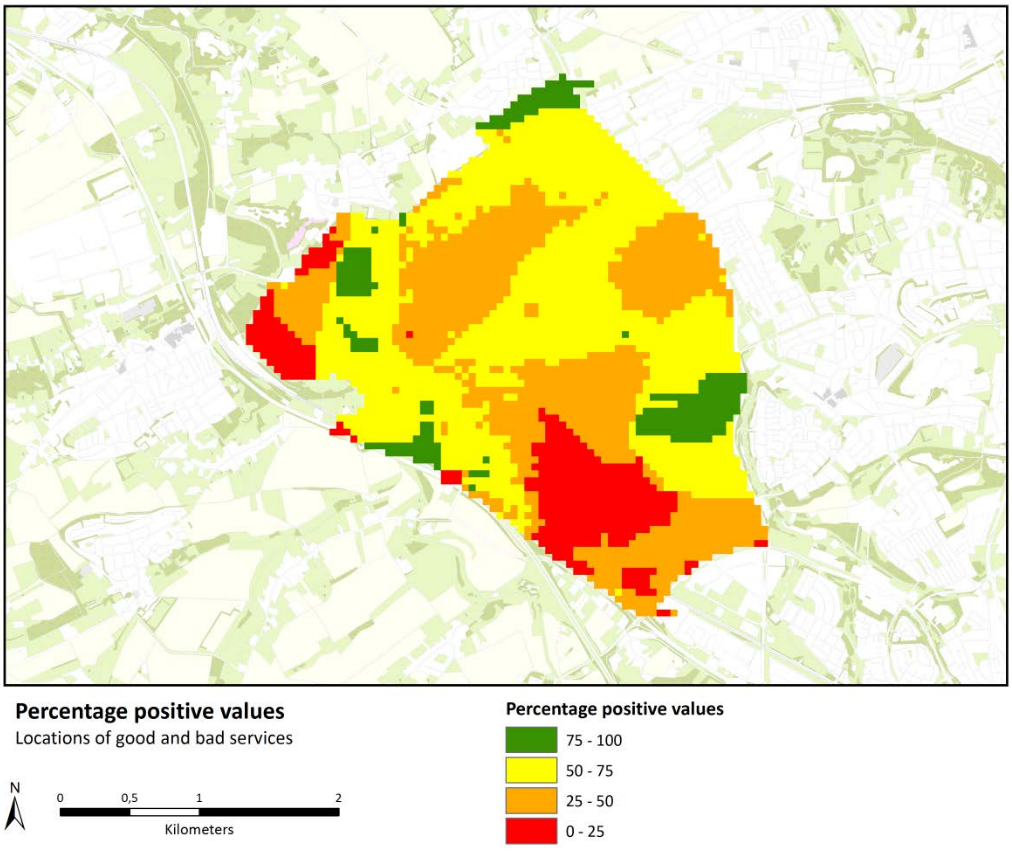


Figure 8.10: Results from the analysis of the percentage of positive values of questions regarding the quality of services.

8.2.3 Correlation analysis

Introduction

Using a correlation analysis, the degree of association between two variables can be calculated. If changes in the values of one variable correspond consistently to changes in the values of another variable it can be said that the variables are correlated. The outcome of a correlation analysis is a unitless number called the correlation coefficient. Its value gives a global indication of the coherence between two variables and ranges from +1 to -1. A value of +1 indicates a perfect positive correlation, which means that the value of one variable always increases with a fixed amount as the value of the other variable rises. A correlation coefficient of -1 demonstrates a perfect negative relation between the two variables, where the one variable always decreases as the other increases.

It is important to note that correlation between two variables does explicitly not prove a causal effect of one variable on the other. The fact that a statistical relationship exists between two variables does not imply that the value of one valuable is a result of the value of the other. Also, the correlation coefficient does not give an indication of the amount with which a value increases or decreases as a result of a change in the value of the related variable. It solely indicates the degree of coherence between the two variables. The most commonly used correlation is Pearson's correlation coefficient, which describes linear relationships between variables. In other words, the Pearson's correlation coefficient describes how well the plotted relationship between two variables is approximated by a straight line.

When exploring the correlation between two variables, a scatterplot-matrix is a good starting point. A scatterplot is a graph in which the values of two variables are plotted against each other for each location (in this case each grid cell). A scatterplot-matrix provides an overview of scatterplots of all the sets of selected variables (as seen in Figure 8.11). These scatterplots can give a first indication of a possible relation between sets of variables. On the basis of the scatterplot-matrix the relevant variables can be selected for further investigation.

To identify possible relations between the answers, their correlation can be calculated using a statistical package. ArcGIS provides the possibility of exporting the data of multiple raster files into a .dbf file, which can in turn be imported into many statistical software packages. Using such a package (SPSS 16 was used) the correlation coefficients of the sets of variables can be calculated. Besides the correlation coefficient, the statistical significance of the relation is provided in the form of a p-value. The correlation coefficient of the relationship between two variables and the p-value of that coefficient give us insight into how the values sets of variables are related statistically.

Hoensbroek results

For the correlation analysis in Level 2, the relationships between the various answers provided by the respondents were of interest. Therefore, a scatterplot matrix of the variables was created. A scatterplot with a selection of the variables is shown in Figure 8.11.

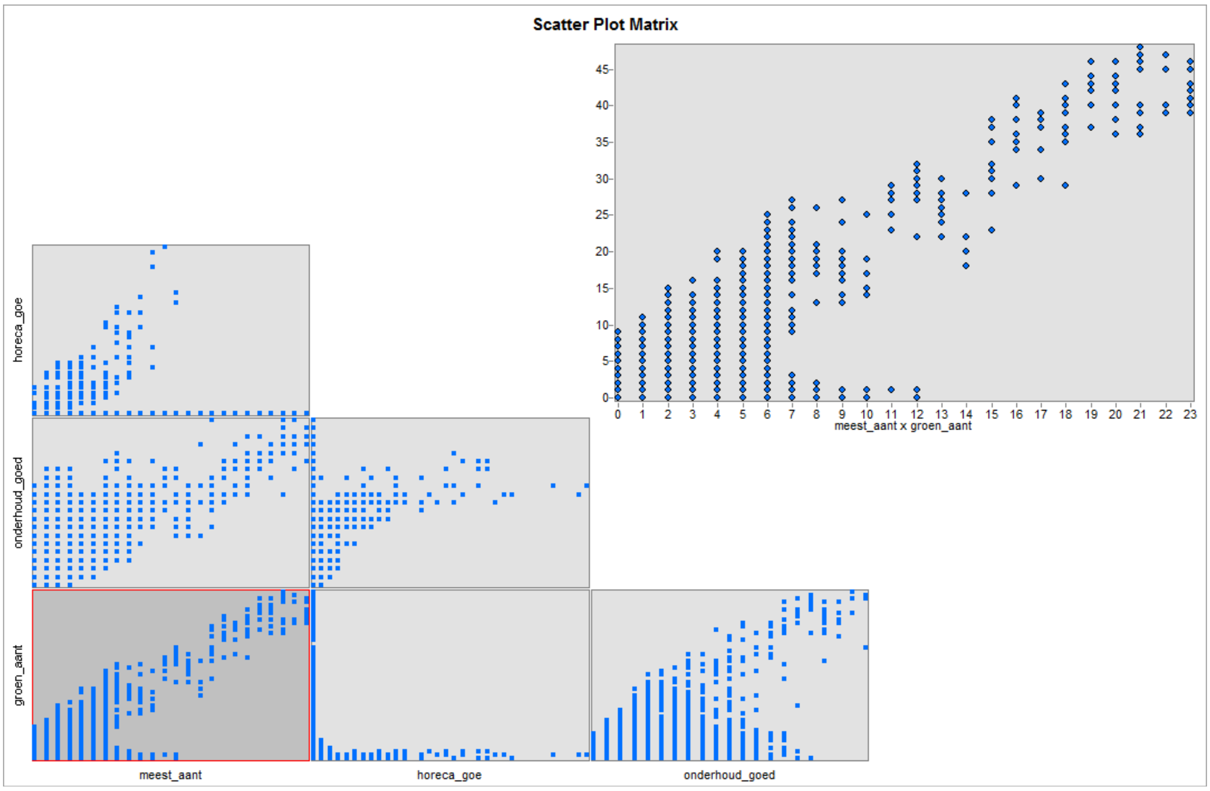


Figure 8.11: Scatterplot-matrix of the data on "What is the most attractive area in Hoensbroek?" and positive experiences of the greenery, the level of maintenance, and bars and restaurants.

For this analysis the focus was on the answers provided by respondents to the question "What is the most attractive area in Hoensbroek?". The relationship between the data provided as an answer to this question and to the other questions can provide insight into the themes that are of importance when people choose a most attractive location. The scatterplots of the data of the locations respondents deemed most attractive are shown on the left of Figure 8.11. The more the data points in the scatterplots are situated along a straight line, the better their relationship can be expressed in a linear relation. A number of scatterplots show a clear linear relation between the data on the most attractive locations and other variables. The strongest relationship appears to be with the data on the appreciation of the greenery. The scatterplot of these two variables is enlarged in the top right corner of Figure 8.11

Having assessed the scatterplots visually for apparent relationships between the variables, the SPSS statistical package was used to calculate the correlation coefficients and their statistical significance. The results for the analysis is shown in Table 8.1. The variables are ordered in descending values of the correlation coefficient. The results show that the relationship between the most attractive places and the areas the respondents appreciate for their greenery can indeed be described to a large extent by a positive linear function. This means that in general, when the value for the appreciation of the greenery increases at a location, so too does the value at that location for the amount of respondents that have identified it as the most attractive location and vice versa.

Apart from the appreciation of the greenery, the variables on the maintenance of public spaces, on the bars and restaurants, the shopping possibilities and the attractiveness of the buildings also have statistically significant positive relationships with the variable for most attractive areas.

Variable	Pearson Correlation Coefficient	Significance
Attractive greenery	0,882**	0,000
Well maintained	0,462**	0,000
Good bars and restaurants	0,116**	0,000
Good shopping possibilities	0,108**	0,000
Attractive buildings	0,058**	0,001
Good sporting facilities	0,032	0,067
Good public transport	0,014	0,424

**. Correlation is significant at the 0.01 level (2-tailed).

Table 8.1: Correlation coefficient and significance of variables with "Most attractive areas in Hoensbroek"

Performing a correlation analysis provides insight into the relations between sets of variables. It can help identify connections between certain phenomena and their underlying processes that might not be readily apparent from looking at the provided answers for each question individually. The scatterplot-matrix offers a method for a visual examination of the relationship between sets of variables. The correlation analysis, in turn, quantifies the found relationships and tests their statistical significance. The information can be used to examine the causal relations between the respondents' associations further. It can also help in decisions regarding the prioritization of spatial interventions that aim to serve a specific purpose.

8.3 Level 3

The analyses for Level 3 are aimed at finding the relationship between characteristics of the built environment and the experiences provided by the respondents. More specifically, the goal is to gain insight into the impact of spatial features on the experiences of citizens. Information on this relationship can be used for understanding the experiences of the respondents by calculating what spatial characteristics are of influence and to what degree they impact their associations. Also, predictions can be made on the basis of the models that are developed. These can be used to anticipate citizens responses to certain spatial interventions, a feat which can greatly benefit spatial planners during the planning process.

8.3.1 Correlation analysis

Introduction

The correlation analysis for Level 3 is performed like that for Level 2. The difference is in the variables chosen for use in the analysis. Since the analyses of Level 3 are aimed at understanding the connection between the physical environment and the experiences of the citizens, the correlation is calculated for variables depicting the answers provided by the respondents and the variables describing characteristics of the built environment. The correlation analysis, in this case, is used as a steppingstone towards the analysis which contain greater explanatory capabilities: the regression analysis.

Hoensbroek results

When a correlation analysis is performed on the data collected among the respondents in Hoensbroek and the data on spatial characteristics, the relationship is examined without the implication of a causal relationship between the two. The correlation coefficient gives an indication of the degree of coherence between the variables. The associated p-value depicts the statistical significance of the relationship. Since the aim of the Level 3 analyses is to model the relationship between the physical characteristics and citizens' perceptions, the correlation analysis is performed in order to make a selection of the variables that might be included in the subsequent regression analysis.

The choice was made to attempt to model the answers provided to the question "What are the most attractive areas of Hoensbroek?" based on the data available on the spatial characteristics of Hoensbroek. The results of the correlation analysis are shown in Table 8.2 The correlation between the collected PPGIS data on the most attractive areas in Hoensbroek and the spatial characteristics at those locations is found to be statistically significant in all three cases. The number of people who identify a location as the most attractive area in Hoensbroek grows as the percentage of the area covered with greenery increases. The value for most attractive areas also rises with a lessening of the percentage of built-up area and with an increase in the age of the buildings of which the built-up areas consist.

Variable	Pearson Correlation Coefficient	Significance
Percentage green	0,442**	0,000
Percentage built-up area	-0,294**	0,000
Year constructed	-0,354**	0,000

**. Correlation is significant at the 0.01 level (2-tailed).

Table 8.2: Correlation coefficient and significance of variables with "Most attractive areas in Hoensbroek"

As was mentioned, the correlation coefficient does not provide any insight into the amount of change in the value of one variable as a result of changes in the other variable. A regression analysis calculates a model for the relationship between the two or more variables, which does provide this type of information. Which characteristics of the physical environment can be included in the regression can be identified on the basis of the outcomes of the correlation analysis.

8.3.2 Regression analysis

Introduction

By applying a regression analysis, the research process can move beyond the exploratory stage and attempt to provide explanations for the observed phenomena. Regression analysis is aimed at describing the relationships between a dependent and one or more independent variables through the development of a regression function. Consequently, an important prerequisite for performing a regression analysis is that a causal relationship is expected between the dependent and independent variables. The assumption is that people, when asked to identify the most attractive areas in their living environment, make this decision at least in part on the basis of the physical characteristics of a place. This assumption justifies the performance of a regression analysis on the collected data.

In this research the ordinary least squares method is used for the estimation of the parameters describing the relationship between the chosen variables.[26] This means an attempt is made to model that relationship with a linear function that follows a structure such as this:

$$y = \alpha + \beta * x$$

The y represents the value of the dependent variable, while the x is the value of an independent variable. The α is the intercept: this is the value for y when the value of x is equal to zero (or: the value at which the line crosses the y-axis). The β is the coefficient of the independent variable. It determines the slope of the line as it indicates the change in the value of y that results from a change in the value of x.

To attain meaningful results from the linear regression, it is important that the correct independent variables are selected to model the dependent variable. This is where the outcomes of the correlation analysis are used, as these results provide insight into the relationship between the different variables (see Table 8.2).

When the potentially relevant variables are chosen for the development of the ordinary least squares regression, the best combination of independent variables has to be sought. ESRI's ArcInfo software offers a tool called Exploratory Regression which automatically examines all possible combinations of the chosen variables and provides a table with diagnostics relating to their capacity to model the dependent variable effectively and a set of statistical diagnostics (see Appendix III). When a combination of the independent variables meets all the criteria set by the user, it is flagged as a 'passing model'.

On the basis of the diagnostics produced by the exploratory regression tool, the ordinary least squares regression analysis can be performed. The analysis produces a formula such as the example depicted above, which describes the influence of the different independent variables on their linear relation with the dependent variable. Once the model is calculated it can be used to predict the value of the dependent variable at locations where no observed values are available but the values for the independent variables are known.

Here the aim is to model the experiences of the respondents on the basis of characteristics of the built environment. This means that the answers provided by the Hoensbroek citizens are considered the dependent variables and that the data on spatial features collected by governmental agencies is used as the independent variables.

Hoensbroek results

Based on the result s from the correlation analysis, the decision was made to include the three variables describing the environment in the regression analysis. The fact that the correlation coefficients are statistically significant does not, however, guarantee that all the variables are of added value to the regression model. Using the exploratory regression tool all the combinations of the independent variables were tested to examine how they influence the predictive value of the regression model. The output of the tool is visible in Appendix III.

The predictive quality of the model is expressed in the coefficient of determination, or R^2. This value indicates to what extent the model explains the variation in the values of the dependent dataset. In the case of a linear regression model such as the one calculated here, it indicates how

[26] This means the equation is based on the smallest squared sum of the distance between the observed data and the predicted data on the basis of the model. Hence the term ordinary least squares regression.

well the modeled function fits the plotted datapoints. The highest R^2 of the calculated models is 0,19, which means less than 20 percent of the variation in the values of the variable for most attractive areas in Hoensbroek can be explained on the basis of the independent variables used.

The exploratory regression tool also produces a list of other diagnostic values of the tested models. These values indicate whether the models meet the criteria to be considered properly specified ordinary least squares regression models. The diagnostics are concerned with whether the values of the residuals of the model follow a normal distribution (Jarque-Bera Statistic), whether their variances are constant (Koenker Statistic) and to what degree the residuals are spatially clustered (Moran's I). The tool's output indicates that none of the models meets all of the criteria required to be flagged as a 'passing' model. This means that an ordinary least squares regression model using these independent variables to model the values of the most attractive areas variable will not yield reliable results.

For the sake of the exercise, however, an ordinary least squares regression model was calculated with these variables nonetheless. Appendix IV shows the outcomes of the ordinary least squares regression analysis. The output of the analysis includes the same set of diagnostics as the exploratory regression tool showing whether or not the model is reliable based on the characteristics of the independent variables and the models residuals.[27] The diagnostics again indicate that the model does not yield reliable results. This is visualized in the map shown in Figure 8.12. The map depicts the spatial distribution of the residuals of the model on the basis of the number of standard deviations they depart from the mean. A clear cluster of residuals with high positive values can be identified in the area around the castle in Hoensbroek. This means the value for the variable of the most attractive area is consistently under-predicted with a substantial amount in that region. A Moran's I value of over 75 with a p-value of 0.00 confirms that the residuals are severely spatially clustered, underlining the notion that the results of the model are unsound. The spatial distribution of the residuals can provide an indication as to which explanatory independent variables are missing in the regression model.

The spatial distribution of the residuals can be explained by the coefficients calculated for the three independent variables. Performing the ordinary least squares regression analysis on the three independent variables yields the outcomes shown in Table 8.3.

Variable	Coefficient
Intercept	*1,192*
% Green	0,031
% Built-up area	0,006
Year of construction	0,000

Table 8.3: Parameters for the regression equation for the independent variables

[27] The residuals are calculated by subtracting the value predicted by the regression model from the observed value at location. This means that at locations with positive residuals, the observed values are higher than the predicted values indicating an underprediction of the model. At locations with negative residuals, on the other hand, the values are overpredicted by the model.

This means the value of the dependent variable is modeled with the equation:

$$Value\ most\ attractive\ area = 1{,}192 + (\ 0{,}031 * \%\ Green\ ) + (\ 0{,}006 * \%\ Builtup\ area) + 0$$

The coefficients tell us that the value of the variable most attractive areas at a given location is dependent almost entirely on the value of the percentage of green at that location. As these two datasets do not coincide very well spatially, it is unsurprising that the model does not predict the values of the dependent variable effectively and that the residuals are spatially clustered.

The fact that the regression analyses did not yield reliable results, could be due to a number of reasons. An explanation could be that the assumption of a causal relationship between characteristics of the physical environment and the value for the most attractive location in Hoensbroek was wrong. This would mean that trying to model the relationship with a regression model is wrong in itself. Another reason could be that the relationship between the dependent and independent variables is not modeled correctly by a linear equation, but rather follows another mathematical coherence. In that case another type of regression analyses could yield better results. Finally, it could be that the independent variables available for the description of the characteristics of the physical environment do not capture the features that are of importance to people when assessing the beauty or appeal of a location effectively. In other words, the essential independent variables were missing in the equation. Which of these explanations holds true can only be assessed through further research. The outcomes of the performed regression analysis can provide clues as to where the answers may lie.

Although no satisfactory model was found for the explanation and prediction of the value for the most attractive area, the application of regression analyses on this type of collected data holds a lot of potential. As has been emphasized, the aim is to be able to predict (to an extent) how citizens will react to certain spatial interventions. Just as models exist for the forecasting of noise pollution when building a road, so too would a model be valuable which predicts the effect of a spatial design on the citizens' perceptions and experiences. Using a regression analysis assumes that there is indeed a causal relationship between characteristics of the physical environment and the experiences citizens associate with those locations. Further research is necessary to determine whether this is the case and, if so, to what extent features of the environment influence people's perceptions.

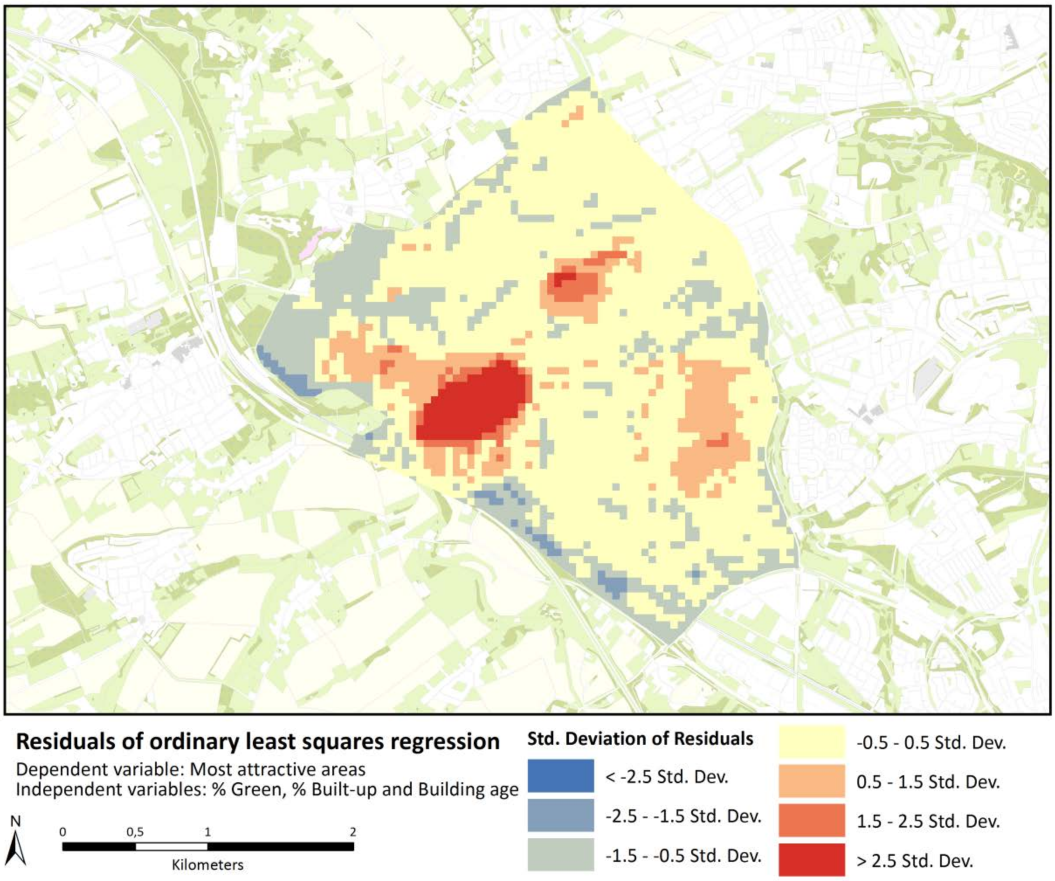


Figure 8.12: The spatial distribution of the residuals of the regression model on the basis of the number of standard deviations they depart from the mean.

8.4 Evaluation of the results

After the performance of the analyses, the results were evaluated by a small group of spatial planners from the Heerlen municipality. Overall, their impression of the usability of the collected data and the information gathered from that data was positive.

In general, the visualization of the information in the form of maps was regarded as a great improvement over the traditional representation of similar data as tables or charts. Many municipalities collect data on the citizens' opinions of their living environments through questionnaires distributed by mail or over the internet and Heerlen is no exception. Although the data collected with these questionnaires also provide insight into the experiences and perceptions of the inhabitants of the municipality, they lack the geographical aspect that was achieved by using the SpatialQuestionnaire application for the data collected in this research. The appreciation of the planners of this spatial aspect of the information can be regarded as a feat of the PPGIS application rather than the used analysis.

The spatial element in the collected data provided the opportunity to use the analyses that were applied in this chapter. The planners were most enthusiastic about the information gained from the analyses of Level 1. The quick insight that the heat maps and the cluster analysis provide into the spatial distribution of the collected answers were deemed especially valuable. The typology of the answers on the basis of certain characteristics of the respondents was assessed a little less favourable. This might be explained by the fact that the analysis itself does not yield results that can be interpreted at a glance. Rather it is stepping stone on the way to other analyses.

The conflict analysis in Level 2 was also seen as a valuable tool. The possibility of comparing the input to two questions spatially was regarded as useful. The results of the analysis calculating the percentage of positive results for a certain theme was deemed a little less beneficial for use in the planning process. A possible explanation could be that the output that the analysis yields is too general to base any conclusions on. The more specific results gained from the conflict analysis might therefore be easier to interpret and thus draw conclusions from.

The correlation analysis (in both Level 1 and Level 2) and the regression analysis were considered least usable by the planners, although the potential value of the results was recognized. It was thought that the output could be valuable for further research into specific issues in the area. However, the usability of the calculated correlation coefficients and regression equation in the spatial planning process directly was not considered very high. The logical explanation is that the relationships between the data that are found with these analyses cannot be translated into spatial policy directly, but rather must be examined further to produce results that can be interpreted easily.

One important side note was added during the evaluation of the results of this thesis. The spatial level at which the analysis were performed were considered useful for the development of the 'Structuurvisie', a plan for the entire municipality containing the spatial policy of the area for the following 10 years. In order for the gathered information to be useful for the daily planning activities it should be delivered at a larger scale. Planners often work on the scale of a single neighbourhood or housing block. Consequently, the results should be able to distinguish between two sides of the same street to make them instantly usable in the planning process.

Nevertheless, the notion of an automated approach of analyzing the collected data and visualizing it in the form of a map would be considered of great added value to the planning process.

9. Conclusions

The aim of the research was twofold: firstly we hoped to develop a structured general approach for the analysis of PPGIS data. Secondly, we wanted to translate this structure into sets of analyses in order to put the proposed method into practice. These research goals resulted from an apparent hiatus in the academic literature on the analysis of data collected with the use of Public Participation Geographical Information Systems. The research question to be answered was therefore:

How can geographical data gathered among citizens through the use of PPGIS be turned into information that is valuable in the spatial planning process?

In order to find an answer to this question, we had to get insight into the contexts within which the transformation from PPGIS data to information takes place. The following sub-questions were articulated to provide a framework for the study of the theoretical, societal and organizational context of the main research question:

- *What is Public Participation GIS?*
- *What is the role of PPGIS in spatial planning processes?*
- *What information is valuable in the spatial planning process?*
- *What spatial analyses are available for geographical data gathered among citizens?*
- *What analysis can be used to translate PPGIS data into valuable information?*

The study of the scholarly literature on PPGIS showed that it is difficult to capture the phenomenon in a single, comprehensive definition. The diversity of applications and software tools that fall under this umbrella term is just too large. PPGIS can therefore probably best be described by the intention of providing all stakeholders in spatial decision-making processes with equal access to the data and analytical capabilities provided through geographical information technology. Not only should the technology allow stakeholders to access relevant information and tools in order to be able to make informed decisions, it should also provide stakeholders with the possibility to communicate their opinions effectively.

As such, the role of PPGIS in spatial planning processes can be seen as a form of communication technology. One of the main obstacles in this exchange of information is the variation in knowledge of the stakeholders on the spatial issue at hand. In most planning processes experts of various disciplines are involved. Their arguments in favour of the adoption of a certain spatial solution are based on their knowledge of a specific subject relevant to the problem. Citizens, on the other hand, base their opinion and preferences on their experience of the area as such, sometimes termed ‘indigenous knowledge’. The challenge is to bridge the gap between these two worlds. As Dunn (2007) stated: “PPGISystems can be seen as a technology with which local and indigenous knowledge and expert data can be combined.”

What role PPGIS plays exactly is dependent on the phase of the spatial planning process in which it is used. The participation of citizens can occur at any stage. As a result, PPGIS might be used to capture the experiences of citizens for the identification and inventory of problems in the area. If PPGIS is instead used during the design stage of the process, it might be used as an analytical tool for stakeholders to present their personal preferred spatial solution.

The same applies to the type of information gathered among citizens. This again depends on the phase in the planning process in which the input from citizens is required. In general, the ability to analyse the PPGIS data in such a way that it yields results that can be used to substantiate decisions based on that information, is the crucial thing. This means that the information should not only provide insight into the distributions and relations found in the collected data, but also allow quantification of these findings.

On the basis of the preliminary assessment of scientific and professional literature on the subject of the analysis and use of PPGIS data for the purpose of spatial planning, the absence of a structured approach to the issue became obvious. A number of scholars have produced studies on PPGIS methodology in a spatial planning context in recent years, although their focus was not specifically on the analysis aspect of PPGIS data. The examples provided in this thesis are studies on the use of Social Landscape Metrics by Brown, the development and application of the SoftGIS methodology by Kyttä, and the adoption of the GIS-Multicriteria Decision Analysis technique by Boroushaki & Malczewski. Elements from these studies provided valuable input for the development of the structured strategy for dealing with PPGIS data and at the same time underpinned the need for its creation.

The answer to the main research question was sought in a combination of the findings from the literature study with an empirical study. First, a structured approach for the analysis of PPGIS data was developed. Appropriate analytical tools were assigned to each compartment of the approach. For the empirical part of the research the decision was made, after consultation of the spatial planning department of the Heerlen municipality to collect data on the experiences of citizens for the exploration phase of the spatial planning process. A PPGIS application was developed for the data collection, which subsequently took place in the Hoensbroek area of Heerlen. The method for the analysis of PPGIS data was then applied to the data.

The two elements developed in this research can be compared to a toolbox and the tools that it holds. The toolbox consists of a number of compartments that are designed to hold various analytical tools used for similar purposes. As such, it provides a structure which helps the user identify what tools could be useful for the performance of a certain task. The method developed in this research for turning PPGIS data into information that can be used for spatial planning serves the same purpose: it structures the available analyses in a logical order. Based on the type of information that the users are interested in recovering from their data, they can look into one of the three proposed levels of analysis. Based on the characteristics of the collected data, some analyses are more suitable to attain certain types of information than others.

As the data in this research was collected in the form of polygons instead of points, a procedure was also developed to prepare the data in such a way that the analyses to be performed yielded meaningful results. The outcomes of the analyses are reported in the Results chapter.

Representatives of the spatial planning department of the Heerlen municipality provided feedback on the usefulness of the analyses and the underlying structure associated with it. Their assessment of the attained information from the collected data was generally positive. Especially the analyses that yielded information that was readily interpretable and provided concrete directions for spatial policies were considered useful. Therefore, the heat maps, the cluster analysis and the conflict analysis were most appreciated. Although the value of the correlation and regression analysis was acknowledged, they were considered more as stepping stones for further investigation into specific topics. Overall, a structured approach for the analysis and spatial visualization of PPGIS data was considered a valuable addition to the spatial planning process.

10. Discussion

I encountered public participation GIS applications for the first time during my internship. When the time came to choose a topic for the thesis I decided that I wanted to continue my work on the use of PPGIS in the context of spatial planning. After an extensive study of the scholarly literature the hiatus of the analysis of PPGIS data became apparent. It quickly became clear that a study of this topic would have to include an empirical element as there are no PPGIS datasets readily available for examination. In the end the PPGIS cycle was completed almost in its entirety, with the development of a PPGIS application and associating questionnaire, the collection of data among citizens and the translation of that data into information through spatial analysis and visualization.

Due to its exploratory nature, some decisions had to be made during the execution of this research based on limited theory. Many colleagues and experts on various fields were consulted during the development of the SpatialQuestionnaire application and the survey and the execution of a number of analyses. Also, to ensure the practical viability of the eventual output of the research – both in the form of the collected data and the results from the various analyses – cooperation with the planning department of the Heerlen municipality was sought.

In hindsight, a number of choices that were made in the course of this research turned out well, while sometimes another course could maybe have yielded better results. One of the issues that was debated extensively was whether or not to let respondents draw polygons instead of points as answers to the questions. Eventually the decision was made to give it a try, in part because it did not appear to be attempted in previous studies. Although working with the collected polygons provided its own problems (as described in the chapter 7), the continuous data that resulted from the rasterization process was nicely suited for use in many spatial analyses.

Also the decision to not limit the area within which citizens could draw their answers was a point of discussion. Again, it presented problems in the use and interpretation of a number of analyses (see for instance section 8.1.2). As the feedback from the Heerlen planning department included remarks about requiring data on a larger scale for their day-to-day activities, a solution in future research could be to select the area of interest and allow citizens to only provide answers within its boundaries. On the other hand, data on the activities of citizens outside of the study-area could prove very valuable when trying to understand their preferences. Also, larger trends between certain answers or themes, such as addressed with the correlation and regression analysis, could be harder to identify when focussing the survey on a very small area. Thus, a recommendation for future research using PPGIS applications for collecting data would be to carefully consider this issue and base the decision on the specific aims of the project.

Finally, one of the main pitfalls of this research was the amount of data that was collected among the citizens in Hoensbroek. Although it has resulted in a very rich dataset, the shear amount of variables makes it difficult to distinguish between the important and less important connections within the data. As was stated in the introduction of the Results chapter, it was necessary to select a specific theme for which to describe the outcomes of the analyses in order to prevent readers from getting lost. A recommendation for future research is therefore to choose a specific topic that is relevant and of interest and thereby limit the number of issues addressed in the data collection phase. A correlation and regression analysis on data from a round of preliminary data collection could help identify what variables are of actual value by providing insight into the relationships within the data.

A framework for the approach of PPGIS data can increase the efficiency of participatory processes in spatial planning as it improves the effectiveness with which the experiences and preferences of citizens can be communicated to other stakeholders. Although the approach of data collection applied in this research does not represent a two-way communication between the citizens and the local authorities, a structured inventory of the citizens' perceptions can serve as the starting point for a discussion on the issues that are identified. Although the use of PPGIS can conceivably reach many more people than the traditional community meetings, the issue of assuring representativeness would need further investigation. With the proposed structured method for the translation of PPGIS data into useable information, a first attempt was made to develop a framework on which to base the division of the various analyses that are available.

Future research could go into three distinct directions: critical, supplementary and deepening. I will provide an example of each. Critical research could focus on the advantages (or the lack of them) of input of preferences using polygons vis-à-vis points. Supplementary research could expand the range of data on visual characteristics of the built environment and their relation with preferences expressed by respondents. It could also look into the use of fuzzy logic on PPGIS data. Deepening research could look into the implementation of the proposed structure in a real-world participatory planning process, to assess its practical viability.

Literature

Aberley, D. & R. Sieber (2002). Public Participation GIS (PPGIS) Guiding Principles First International PPGIS Conference. Held by URISA at Rutgers University, New Brunswick, New Jersey, pp. 20–22.

Aggens, L. (1983). Identifying different levels of public interest in participation. Fort Belvoir, VA: The Institute for Water Resources, U.S. Army Corps of Engineers.

Anselin, L., Syabri, I. & Y. Kho (2005). GeoDa: An Introduction to Spatial Data Analysis. Geographical Analysis, 38(1), pp. 5-22.

Arnstein, S.R. (1969). A ladder of citizen participation. In: Journal of the American Institute of Planners, 35, pp. 214-229.

Balram, S. & S. Dragicevic (2006). Collaborative geographic information systems: origins, boundaries, and structures. In: Balram, S. & S. Dragicevic (Eds.). Collaborative Geographic Information Systems. Hershey, PA: Idea Group Publishers, pp. 1–22.

Bamberg, J. (2010). Ambiguities in knowledge production: multimodal analysis of discourse and dramaturgy in public participation GIS experiments. Environment and Planning B: Planning and Design, 37(5), pp. 895–910.

Boroushaki, S. & J. Malczewski (2010). Measuring consensus for collaborative decision-making: A GIS-based approach. Computers, Environment and Urban Systems, 34(4), pp. 322–332.

Brown, G. (2012). Public Participation GIS (PPGIS) for regional and environmental planning: reflections on a decade of empirical research. Journal of Urban and Regional Information Systems Association, 25(2), pp. 7-18.

— & M. Kyttä (2014). Key issues and research priorities for public participation GIS (PPGIS): A synthesis based on empirical research. Applied Geography, 46, pp. 122–136.

— & B.A. Reed (2012). Social Landscape Metrics: Measures for Understanding Place Values from Public Participation Geographic Information Systems (PPGIS), Landscape Research, 37, pp. 73-90.

Chrisman, N.R. (1999). What does 'GIS' mean? Transactions in GIS, 3, pp. 175–186.

Cinderby, S. (1999). Geographic information systems (GIS) for participation: the future of environmental GIS? International Journal of Environmental Pollution, 11, pp. 304–15.

— (2010). How to reach the "hard-to-reach": the development of Participatory Geographic Information Systems (P-GIS) for inclusive urban design in UK cities. Area, 42(2), pp. 239–251.

—, Bruin, A. de, White, P. & M. Huby (2012). Analyzing Perceptions of Inequalities in Rural Areas of England Using a Mixed-methods Approach. Journal of Urban and Regional Information Systems Association, 25(2), pp. 33 – 42.

Coenen, F., Peppel, R. van de & J. Woltjer (2001). De evolutie van inspraak in de Nederlandse planning. Beleidswetenschap, 15, pp.313-332.

Cooke, B. & U. Kothari (2001). Participation: The New Tyranny? Zed, London.

Couclelis, H. (2003). The certainty of uncertainty: GIS and the limits of geographic knowledge. Transactions in GIS, 7, pp. 165–75.

Concilio, G., & F. Molinari (2011). Citizen Participation in Urban Planning: Looking for the "E" Dimension in the EU National Systems and Policies. In: The Proceedings of the 11th European Conference on E-government: Faculty of Administration, University of Ljubljana, Ljubljana, Slovenia, 16-17 June 2011. Academic Conferences Limited.

Creighton, J. L. (1983). Identifying publics/staff identification techniques. Fort Belvoir, VA: The Institute for Water Resources, U.S. Army Corps of Engineers

Day, D. (1997). Citizen participation in the planning process: an essentially contested concept? Journal of Planning Literature, 11, pp. 421-34.

Department of the Environment (1987). Handling Geographic Information. London, Her Majesty's Stationary Office.

Dienst Landelijk Gebied (2012). Gebrookerbos: Plan van Aanpak voor Realiseren Standslandbouw.

Dueker K.J. & D. Kjerne (1989). Multipurpose cadastre: Terms and definitions. Falls Church, VA: American Society for Photogrammetry and Remote Sensing.

Dunn, C.E. (2007). Participatory GIS – a people's GIS? Progress in Human Geography, 31, pp. 616-637.

Elwood, S. (2002). GIS use in community planning: a multidimensional analysis of empowerment. Environment and Planning A, 34, pp. 905–922.

— (2006). Critical Issues in Participatory GIS: Deconstructions, Reconstructions, and New Research Directions. Transactions in GIS, 10, pp. 693–708.

ESRI (1997). About GIS. Accessed 17 Novermber 2013,
<http://healthcybermap.org/HGeo/res/esri_what_is_a_gis.pdf>

Geertman, S.C.M. (2002). Participatory Planning and GIS: a PPS to Bridge the Gap. Environment and Planning B, 29, pp. 21-35.

— & J. Stilwell, eds. (2003). Planning Support Systems in Practice. Berlin: Springer.

Haklay, M. & C. Tobon (2003). Usability evaluation and PPGIS: towards a user-centred design. International Journal of Geographical Information Science, 17, pp. 577–592.

Hall, P. (2003). Cities of tomorrow: An intellectual history of urban planning and design in the 20th century. London: Blackwell.

Harris, T.M., Weiner, D., Warner, T.A. & R. Levin (1995). Pursuing social goals through participatory geographic information systems. In Pickles, J., editor, Ground truth: the social implications of geographic information systems. New York: Guilford Press, pp. 196–222.

— & D. Weiner (1996). GIS and Society: The Social Implications of How People, Space and Environment are Represented in GIS. Scientific Report for NCGIA Initiative # 19 Specialist Meeting, University of California at Santa Barbara, November.

— & D. Weiner (1998). Empowerment, marginalization, and 'community-integrated' GIS. Cartography and Geographic Information Systems, 25, pp. 67–76.

Hospers, G.J. (2012). Urban Shrinkage and the Need for Civil Engagement. In: Haase, A., Hospers, G.J., Pekelsma, S. & D. Rink (Eds.). Shrinking Areas: Front-runners in Innovative Citizen Participation. The Hague: European Urban Knowledge Network, pp. 7-28.

Innes, J.E. & D.E. Booher (2004), Reframing Public Participation: Strategies for the 21st Century. Planning Theory and Practice, 5, pp. 419-436.

Intergemeentelijke Structuurvisie Parkstad Limburg (2009). Ruimte voor Park en Stad. Parkstad Limburg.

Jankowski, P., Robischon, S., Tuthil, D., Nyerges, T. & K. Ramsey (2006). Design Considerations and Evaluation of a Collaborative, Spatio-Temporal Decision Support System. Transactions in GIS, 10, pp. 335-354.

Kahila & Kyttä (2009). Soft-GIS as a Bridge-builder in Collaborative Urban Planning. In: Geertman S.C.M. & J. Stilwell (Eds.). Planning Support Systems: Best Practice and New Methods. Berlin: Springer, pp. 389-411.

Kingston, R., Carver, S., Evans, A. & I. Turton (2000). Web-based public participation geographical information systems: an aid to local environmental decision-making. Computers, Environment and Urban Systems, 24(2), pp. 109–125.

Kraak, M.J. & F. Ormeling, (2010). Cartography, Third Edition: Visualization of Spatial Data. New York: Guilford Press.

Krek, A. (2005). Rational Ignorance of the Citizens in Public Participatory Planning. Paper published in the conference proceedings of CORP 2005, Vienna.

Kyttä, M., Broberg, A., Tzoulas, T. & K. Snabb (2013). Towards contextually sensitive urban densification: Location-based softGIS knowledge revealing perceived residential environmental quality. Landscape and Urban Planning, 113, pp. 30-46.

Longley, P.A., Goodchild, M.F., Maguire, D.J. & D.W. Rhind (Eds.) (2010). Geographic Information Systems and Science. New York: Wiley.

Lovan, W.R., Murray, M. & R. Shaffer (Eds.) (2004). Participatory Governance: Planning, Conflict Mediation and Public Decision-Making in Civil Society. Aldershot: Ashgate.

Mac Eachren, A. & I. Brewer (2004), Developing a conceptual framework for visually-enabled geocollaboration. International Journal of Geographical Information Science, 18 , pp. 1-34.

McCall, M.K. (2003). Seeking good governance in participatory-GIS: a review of processes and governance dimensions in applying GIS to participatory spatial planning. Habitat International, 27, pp. 549–573.

— (2004). Can Participatory-GIS Strengthen Local-level Spatial Planning? Suggestions for Better Practice. Paper prepared for GISDECO 2004.

— & C.E. Dunn (2012). Geo-information tools for participatory spatial planning: Fulfilling the criteria for ‘good’ governance? Geoforum, 43 pp. 81–94.

— & P. Minang (2005). Assessing participatory GIS for community-based natural resource management: claiming community forests in Cameroon. Geographical Journal, 171, pp. 340–356.

McCrea, R., & P. Walters (2012). Impacts of urban consolidation on urban liveability: Comparing an inner and outer suburb in Brisbane, Australia. Housing, Theory and Society, 29(2), pp. 190–206.

Mitchell, R.K., Agle, B.R. & D.J. Wood (1997). Toward a theory of stakeholder identification and salience: defining the principle of who and what really counts. The Academy of Management Review, 22, pp. 853-886.

Monmonier, M.S. (1996). How to Lie with Maps. Chicago: University of Chicago Press.

National Center for Geographic Information and Analysis (NCGIA) (1996). Summary report: Public participation GIS workshop, Orono, ME, 10–13 July.

Obermeyer, N.J. (1998). The evolution of public participation GIS. Cartography and Geographic information Systems, 25, pp. 65–66.

Openshaw, S. (1991). A view on the GIS crisis in geography, or using GIS to put Humpty-Dumpty back together again. Environment and Planning A, 23, pp. 621–628.

— (1992). Further thoughts on geography and GIS: a reply. Environment and Planning A, 23, pp. 463–466.

Pateman, C. (1970). Participation and Democratic Theory. Cambridge: Cambridge University Press.

Pickles, J. (Ed.) (1995). Ground truth: the social implications of Geographical Information Systems. New York: Guilford Press.

Pickles, J. (1999). Arguments, debates, and dialogues: the GIS-social theory debate and the concern for alternatives. In: Longley, P.A., Goodchild, M.F., Maguire, D.J. & D.W. Rhind (Eds.). Geographical Information Systems: principles and technical issues, New York: Wiley, pp. 49–60.

Ramasubramanian, L. (2010). Geographic information science and public participation. Berlin: Springer.

Rambaldi, G. & J. Callosa. (2000). Manual on participatory 3- dimensional modelling for natural resources management. Vol. 7 of Essentials of Protected Areas Management in the Philippines. NIPAP, PAWB-DENR, Philippines.

— (2002). Participatory Three-dimensional Modelling: Guiding Principles and Applications, 2002 edition. Asean Regional Centre for Biodiversity Conservation (ARCBC), Los Banos, Phillipines.

Rambaldi G., Kyem Kwaku, A.P., Mbile, P., McCall, M. & D. Weiner (2006). Participatory Spatial Information Management and Communication in Developing Countries. Electronic Journal of Information Systems in Developing Countries, 25.

Rietbergen-McCracken, J. & D. Narayan-Parker (Eds.) (1998). Participation and Social Assessment: Tools and Techniques. Washington: World Bank.

Rijckheyt – Centrum voor regionale geschiedenis (2008). Accessed 20 December 2013, < http://www.rijckheyt.nl/sjablonen/rijckheyt/pagina.asp?pagina=465>

Royal Town Planning Institute (2007). A New Vision for Planning: Delivering Sustainable Communities: An Agenda for Action. Royal Town Planning Institute, London.

Schaffers, H., Komninos, N. & M. Pallot (Eds.) (2012). Smart Cities as Innovation Ecosystems Sustained by the Future Internet. Brussels: European Commission.

Schlossberg, M. & E. Shuford (2005). Delineating ‘Public’ and ‘Participation’ in PPGIS. Journal of the Urban and Regional Information Systems Association, 16, pp. 15–26.

Schroeder, P. (1996). Criteria for the design of a GIS/2. Specialists' meeting for NCGIA Initiative 19: GIS and society, Summer 1996.

Schuurman, N. (2000). Trouble in the heartland: GIS and its critics in the 1990s. Progress in Human Geography, 24, pp. 569–590.

Sieber, R.E. (2000). Conforming (to) the opposition: the social construction of geographical information systems in social movements, In: International Journal of Geographical Science, 14, pp. 775–793.

— (2004). A PPGIScience? Cartographica, 38(3-4), pp. 1-3.

— (2006). Public Participation Geographic Information Systems: A Literature Review and Framework. Annals Of The Association Of American Geographers, 96, pp. 491-507.

Stivers, C. (1990). The Public Agency as Polis: Active Citizenship in the Administrative State. Administration & Society, 22, pp. 86-105.

Talen, E. (2000). Bottom-up GIS – a new tool for individual and group expression in participatory planning. Journal of the American Planning Association, 66, pp. 279–294.

Taylor, P.J. (1990). GKS. Political Geography Quarterly, 9(3), pp. 211–212.

—, & M. Overton (1991). Further thoughts on geography and GIS. Environment and Planning A, 23, pp. 1087–1094.

Thomas, J.C. (1995). Public participation in public decisions: new skills and strategies for public managers. San Francisco: Jossey-Bass Publishers.

Thwaites, K. & I. Simkins (2007). Space and Place. London: Taylor.

Willeke, G.E. (1974). Identification of publics in water resources planning. Atlanta: Department of City Planning: Environmental Resources Center.

Appendix I: Survey Part 1

Beste respondent,

Wij stellen het zeer op prijs dat u wilt meewerken aan dit onderzoek van de Vrije Universiteit van Amsterdam. Met deze studie willen wij de meningen van bewoners over hun woonomgeving in kaart brengen en hun ervaring van de gevolgen van de bevolkingsontwikkeling op die omgeving bekijken.
Het onderzoek duurt ongeveer 15 minuten en is volledig anoniem.
In het eerste deel van dit onderzoek is u gevraagd om op een virtuele kaart verschillende positieve en negatieve associaties een plaats te geven. Hieronder wordt u gevraagd wat achtergrondinformatie te verstrekken.

Achtergrondinformatie:
Omcirkel aub wat op u van toepassing is.

Geslacht: Man / Vrouw

Leeftijd: ……… jaar

Hoeveel jaar woont u al in Hoensbroek?
- o Minder dan 5 jaar
- o 5 – 9 jaar
- o 10 – 14 jaar
- o 15 – 20 jaar
- o Meer dan 20 jaar
- o Ik ben een bezoeker

Hoe ziet uw huishouden eruit?
- o Eenpersoons huishouden
- o Eenouder huishouden
- o Paar zonder kinderen
- o Paar met kinderen

Wat is uw hoogst voltooide opleiding?
- o Basisonderwijs
- o VMBO
- o HAVO/VWO
- o Middelbaar beroepsonderwijs (MBO)
- o HBO/WO

1/2 **Z.O.Z.**

Wat is uw werksituatie?

- Voltijd baan
- Deeltijd baan: meer dan 50%
- Deeltijd baan: 50%
- Deeltijd baan: minder dan 50%
- Ik heb geen werk

Wat is uw netto jaarlijkse inkomen?

- Ik heb geen eigen inkomen
- Minder dan € 7.500
- € 7.500 - € 15.000
- € 15.000 - € 25.000
- € 25.000 - € 40.000
- Meer dan € 40.000
- Hier geef ik liever geen antwoord op

Heeft u thuis internet?

- Ja
- Nee

Hartelijk dank voor uw deelname! De resultaten van dit onderzoek zullen onder anderen worden gebruikt om een vervolgonderzoek op te zetten. Als u bereidt bent aan het vervolgonderzoek deel te nemen, geef dan hieronder uw emailadres op:

Email-adres: ………………………………

2/2

Appendix II: Survey Part 2

Questions asked through the use of the SpatialQuestionnaire application:

Number	Question
1	Waar woont u?
2	Welk gebied beschouwt u als uw woonomgeving?
3	Waar vindt u het groen aantrekkelijk?
4	Waar vindt u de bebouwing aantrekkelijk?
5	Waar vindt u de bebouwing onaantrekkelijk?
6	Welke plaatsen worden goed onderhouden?
7	Welke plaatsen worden slecht onderhouden?
8	Waar in Hoensbroek neemt u leegstand waar?
9	Op welke plaatsen ervaart u overlast door leegstand?
10	Waar zijn goede winkelvoorzieningen?
11	Waar zijn de winkelvoorzieningen gebrekkig?
12	Waar zijn goede openbaar vervoer voorzieningen?
13	Waar is de openbaar vervoer voorziening gebrekkig?
14	Waar zijn goede horeca voorzieningen?
15	Waar zijn de horeca voorzieningen gebrekkig?
16	Waar zijn goede sportvoorzieningen?
17	Waar zijn de sportvoorzieningen gebrekkig?
18	Waar voelt u zich soms onveilig?
19	Waar doen zich onveilige verkeerssituaties voor?
20	Waar vindt u de omgeving het meest aantrekkelijk?
21	Waar vindt u de omgeving het meest onaantrekkelijk?
22	Waar in Hoensbroek zou u het liefst willen wonen?

Appendix III: Output exploratory regression

```
****************************************************************************
Choose 1 of 3 Summary
     Highest Adjusted R-Squared Results
  R2      AICc    JB    BP   VIF    MI    Model
0,18 17903,96 0,00 0,00 1,00 0,00  +PERC_GROEN***
0,15 18036,76 0,00 0,00 1,00 0,00  -AGGREGA_BA***
0,07 18306,39 0,00 0,00 1,00 0,00  -PERC_BEBOU***
    Passing Models
R2 AICc JB BP VIF MI    Model
****************************************************************************
Choose 2 of 3 Summary
             Highest Adjusted R-Squared Results
  R2      AICc    JB    BP   VIF    MI    Model
0,19 17854,51 0,00 0,00 2,07 0,00  +PERC_GROEN***   -AGGREGA_BA***
0,18 17905,91 0,00 0,00 1,69 0,00  +PERC_GROEN***   -PERC_BEBOU
0,15 18022,13 0,00 0,00 1,49 0,00  -PERC_BEBOU***   -AGGREGA_BA***
    Passing Models
R2 AICc JB BP VIF MI    Model
****************************************************************************
Choose 3 of 3 Summary
                    Highest Adjusted R-Squared Results
  R2      AICc    JB    BP   VIF    MI    Model
0,19 17854,65 0,00 0,00 2,45 0,00  +PERC_GROEN***   +PERC_BEBOU*   -AGGREGA_BA***
    Passing Models
R2 AICc JB BP VIF MI    Model
****************************************************************************
************* Exploratory Regression Global Summary (MEEST_AANT) *************

        Percentage of Search Criteria Passed
       Search Criterion Cutoff Trials # Passed % Passed
 Min Adjusted R-Squared > 0,50      7        0      0,00
Max Coefficient p-value < 0,10      7        6     85,71
          Max VIF Value < 7,50      7        7    100,00
Min Jarque-Bera p-value > 0,10      7        0      0,00
  Min Moran's I p-value > 0,10      7        0      0,00

------------------------------------------------------------------------------

Summary of Variable Significance
Variable    % Significant
PERC_GROEN        100,00
```

PERC_BEBOU	75,00
AGGREGA_BA	100,00

Summary of Multicollinearity

Variable	VIF	Violations	Covariates
PERC_GROEN	2,45	0	--------
PERC_BEBOU	1,77	0	--------
AGGREGA_BA	2,17	0	--------

Summary of Residual Normality

JB	R2	AICc	BP	VIF	MI	Model
0,000000	0,145360	18036,760375	0,000000	1,000000	0,000000	-AGGREGA_BA***
0,000000	0,073840	18306,389341	0,000000	1,000000	0,000000	-PERC_BEBOU***
0,000000	0,178529	17903,956406	0,000000	1,000000	0,000000	+PERC_GROEN***

Summary of Residual Autocorrelation

MI	R2	AICc	JB	BP	VIF	Model
0,000000	0,145360	18036,760375	0,000000	0,000000	1,000000	-AGGREGA_BA***
0,000000	0,073840	18306,389341	0,000000	0,000000	1,000000	-PERC_BEBOU***
0,000000	0,178529	17903,956406	0,000000	0,000000	1,000000	+PERC_GROEN***

Table Abbreviations

R2 Adjusted R-Squared
AICc Akaike's Information Criterion
JB Jarque-Bera p-value
BP Koenker (BP) Statistic p-value
VIF Max Variance Inflation Factor
MI Moran's I p-value
Model Variable sign and significance
* = 0.10, ** = 0.05, *** = 0.01

Appendix IV: Output ordinary least squares regression

```
                                Summary of OLS Results
  Variable Coefficient StdError t-Statistic Probability Robust_SE  Robust_t
Robust_Pr  VIF [1]
 Intercept    1,192032 0,138827    8,586457   0,000000*  0,104539 11,402714
0,000000* --------
PERC_GROEN    0,030866 0,002342   13,176663   0,000000*  0,002506 12,314823
0,000000* 2,451462
PERC_BEBOU    0,006302 0,004614    1,365621   0,172161   0,003385  1,861377
0,062778  1,767541
AGGREGA_BA   -0,000000 0,000000   -7,323051   0,000000*  0,000000 -6,897375
0,000000* 2,169264

                                                     OLS Diagnostics
Input Features:             PolyGrid_50m_Clip_Hoensbroek_alles   Dependent
Variable:                                  MEEST_AANT
Number of Observations:                                        3355   Akaike's
Information Criterion (AICc) [2]:   17854,653059
Multiple R-Squared [2]:                                    0,191721   Adjusted
R-Squared [2]:                          0,190997
Joint F-Statistic [3]:                                  264,948592   Prob(>F),
(3,3351) degrees of freedom:           0,000000*
Joint  Wald  Statistic  [4]:                                        495,801615
Prob(>chi-squared), (3) degrees of freedom:      0,000000*
Koenker  (BP)  Statistic  [5]:                                      209,648497
Prob(>chi-squared), (3) degrees of freedom:      0,000000*
Jarque-Bera  Statistic  [6]:                                      16057,310982
Prob(>chi-squared), (2) degrees of freedom:      0,000000*


                           Notes on Interpretation
 *  Statistically significant at the 0.05 level.
[1]  Large  VIF  (>  7.5,  for  example)  indicates  explanatory  variable
redundancy.
[2] Measure of model fit/performance.
[3] Significant p-value indicates overall model significance.
[4] Significant p-value indicates robust overall model significance.
[5]  Significant  p-value  indicates  biased  standard  errors;  use  robust
estimates.
[6]  Significant  p-value  indicates  residuals  deviate  from  a  normal
distribution.
```