# Analysis and mapping of crime perception:

# A quantitative approach of sketch maps.

MARIANA VALLEJO VELÁZQUEZ

September, 2019

SUPERVISORS:

Dr. Rania Kounadi
Dr. Corné van Elzakker

Cartography M.Sc.

# Analysis and mapping of crime perception: A quantitative approach of sketch maps.

MARIANA VALLEJO VELÀZQUEZ

Enschede, The Netherlands, September 2019

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the joint Master of Science in Cartography

SUPERVISORS:
Dr. Rania Kounadi
Dr. Corné van Elzakker

THESIS ASSESSMENT BOARD:
Prof. Dr. Menno-Jan Kraak (Chair)
Dr. Ekaterina Chuprikova (External reviewer, Technical University of Munich)

ITC
UNIVERSITY OF TWENTE.

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

# Abstract

Crime perception is defined as the insight amount of criminal activity in a location or the risk of victimisation. Evidence exists that people's perception about the crime is not often consistent with the actual incidents statistics, and thus there is a tendency of underestimating or overestimating the safety. The misperception of crime can have repercussion on people's lifestyle, affect social behaviour and spatial and economic dynamics. Therefore, it is relevant for police agencies to develop strategies directed to reduce this perception gap.

To come up with efficient action plans is relevant to explore the different social, demographic and environmental factors that sway perception. Likewise, analyse them as a whole within a framework and not as individual and independent elements to have an overall understanding of the context.

Structured sketch maps are often used as a method to capture people's crime perception by collecting data about the places that are perceived as safe or unsafe. This type of sketch maps enables to keep the consistency of the reference context and thus extract spatial attributes out of the sketched features. The exploration of these features using GIS, spatial analysis and statistics methods could enable the understanding of the factors that influence perception. Consequently, allow the comprehension of the spatial arrangement of perceived safe and unsafe places.

This research aims to apply this approach in a case study. Some variables will be extracted from sketch maps and analysed to determine which of those variables are related to the perception of safe and unsafe areas. Moreover, a prototype of a GeoVisual Analytics environment is proposed. This type of interfaces enables the understanding of the complex relationships between multivariate and spatiotemporal datasets. In this case, a set of tools for the visualisation and analysis of perception data are integrated to support police agencies in the development of strategies to reduce misperception of crime.

# Acknowledgements

Throughout the MSc program, I was lucky to meet great Professors, colleagues and friends who turn into family. It has a great journal full of learning, new experiences and adventures.

I am grateful to my supervisor Dr. Rania Kounadi for supporting me throughout my research process. She guided me to explore an area that was unknown by me and encourage me always to go further. I thank my supervisor Dr. Corné van Elzakker, for always keeping the door open for any question I had and for his constant feedback. Thank you both for your time and help!

I also want to express my gratitude to Julianne for all her support she gave me, for her fast responses and her willingness to help us.

Moreover, I want to thank Dr. Pődör Andrea from Óbuda University for providing the datasets used in this research.

This goes to the friends that I am taking with me, everything is easier when you feel comfortable where you are. Thank for making this part of my life one of the best chapters!

Dr. Abraham Navarro thanks for infecting me with your passion for Cartography, thanks for planting a seed on me.

Jaime, I will always be thankful to you for encouraging me to follow this way.

And of course to my parents who constructed the beginning of this road and taught me how to keep building it by myself. To my sister who will always walk with me.

Thank you to all the people who make possible the Cartography M.Sc. program!

# Table of contents

# Table of contents

# List of figures

# List of figures

# List of tables

# Chapter 1

## Introduction

In this first chapter, the **research context** of the thesis is exposed. Basic concepts and target statements are defined in order to give a general background to the **problem** to be addressed (section 1.1).

Moreover, the general and specific **objectives** are listed, as well as the identified **research questions** (section 1.2) to be tackled along the research process. Each specific objective is addressed in a different chapter in which the methodology, results and preliminary conclusions are presented. Therefore, a general outline of the whole **methodology** (section 1.3) is described here in order to give an overall view of the workflow that was followed.

In the last section, the **structure** (section 1.4) of the thesis is presented with a brief description of the contents of each chapter.

## 1.1 Research context and problem statement

Human perception has been studied mostly by Psychology, but since the second half of the twentieth century, it is also of the interest of spatial sciences, such as Geography. The interest lies in the understanding of space based on people's insight. Cognitive mapping is the process of developing a mental map, based on the collection of information by sensorial perception. The geometry and attributes of each individual cognitive map are shaped by internal and external factors. The main tangible representation of a cognitive map is the sketch map.

A sketch map is the main mapping method to graphically depict the spatial knowledge of individuals. Relative location, geometric attributes, as well as impressions and beliefs about places can be portrayed on these representations. Sketch mapping is a recurrent method for the collection of data about people's perception. Researches have made use of this method in the study of diverse social geographical matters such as emergency management, hazard planning, land use planning and community safety (Sloan, Doran, Markham & Pammer, 2016).

Crime studies is one of the research areas in which structured sketch maps are utilized with the aim of collecting data about fear of crime (Curtis, 2012; Kohm, 2009) and perception of crime (Spicer, Song & Brantingham, 2014; Fuhrmann, Huynh & Scholz, 2013; Lopez & Lukinbeal, 2010) and they are usually explored with the use of GIS. The consistent spatial reference of structured sketch maps enables their analysis with spatial tools, due to the fact that they are drawn over a printed or digital base map. However, the analysis is often limited to overlay, aggregation and illustration purposes, mainly resulting in a visual and descriptive analysis (Curtis, 2012). The exploration, data extraction and analysis of sketch maps for crime perception studies may be improved with the integration of GIS, statistical methods and spatial analysis.

Although there are several theories that explain the factors that sway the perception of crime, the spatial component of it has not been explored in depth. Spatial characterization of those factors could result in a better understanding of the location of identified unsafe areas. The relevance of this lies in the fact that perception is not always similar to reality: perceptions of crime often mismatch the actual crime statistics. This disparity is known as the crime perception gap and it arises when a person has an inaccurate insight of safety. There are two types of crime perception inaccuracy: a person can perceive an area as unsafe, whereas it is safe or a person can conceive a place as safe but it is actually unsafe.

In this research, the term "accuracy/inaccuracy of crime perception" is utilized as the state of consistency between what is perceived and the reality defined by objective measurements. As perception cannot be described as "right" or "wrong", the concept of accuracy is employed to establish whether or not the perceived attribute matches with the actual value. Another term that must be defined is "level of crime perception accuracy/inaccuracy". This is not used as a measure of how close a people's perception is to reality, but it is defined as a means of comparison between the percentage of people who have an accurate or inaccurate perception among the total number of people who participate in a survey.

This inaccuracy can have an impact on people's daily life, social behaviour and spatial dynamics. The importance of narrowing this gap is the need for improving people's quality of life. Some studies have evidenced that fear can alter mental health due to anxiety (Foster, Giles-Corti, & Knuiman, 2010). It can restrict the people's daily activity area as a result of the desire to avoid unsafe areas by changing daily routines. As a consequence, the reduction of people transiting would force the relocation of services offered due to a lack of clients. All this would result in a reorganization of the spatial structure at micro-scale.

Thus, there is a need to increase perception accuracy with localized strategies that can narrow the gap. Although police agencies have developed actions to address this issue, they have mainly focused on reducing the fear of crime (Cordner, 2010; Grabosky, 1995; Bennett, 1991). They are particularly focussing on the inaccurate perception of high crime, in which people believe that the level of crime incidents is high, whereas in reality it is low. But then there is still the need to narrow the gap of an inaccurate perception of low crime in existing crime hotspots wherein the people are not aware of the risk of victimization.

Therefore, the strategies to narrow the crime perception gap must consider social, environmental and spatial factors that sway people's perception. These strategies must be implemented in the first place in priority areas characterized by the level of the people's perception accuracy. Police agencies are the bureaus responsible for developing plans of action (Cordner, 2010). Hence, they must be provided with tools that allow them to explore and relate multivariate datasets in order to ease the decision-making in the design of targeted strategies.

## 1.2 Objectives and research questions

The aim of the research is to present an integrated analysis of structured sketch maps in the study of crime perception, by performing a numerical and spatial analysis of the data extracted from the maps and by designing a geovisualization environment that supports their visual and analytical reasoning.

Thus, the general objective of this research is:

> "To quantitatively examine structured sketch maps to analyse and map crime perception. Moreover, to design a geovisual analytics environment that eases the decision-making in the development of strategies to amend the perception of crime"

In order to fulfil this, three specific objectives were defined:

1. To analyse the location of perceived unsafe areas in relation to a) the distribution of crime incidents and b) people's activity spaces.
2. To determine and explore the accuracy of people's crime perception and to map its spatial distribution.
3. To conceptually design a GeoVisual Analytic environment for the exploration and reasoning of perceptions of crime.

To tackle each specific objective five particular questions have to be answered along the research process:

1.1 What is the relationship between the people's daily activity spaces (neighbourhood and daily routes) and the location of the areas they perceive as unsafe?

1.2 What is the relationship between the location of the crime incidents and the perceived unsafe areas?

2.1 How to measure the accuracy of people's crime perception?

2.2 How can the location of inaccurately perceived unsafe areas be explained by the spatial distribution of another explanatory variable?

3.1 Which tools and representations could be integrated in a GeoVisual Analytic interface to explore and analyse crime perception?

The specific objectives are addressed in separate chapters in which the used methodology, the results and conclusions of each one are presented. An overview of the entire framework is described in the next section.

## 1.3 Methodology overview

The applied methodology is divided into three phases that correspond with each specific objective. The first phase comprises performing an exploratory modelling of the data extracted from structured sketch maps; the second phase is a spatial arrangement outline which includes determining the spatial distribution of crime perception and its analysis; the last stage is the geovisualization development in which an interactive environment will be designed (but not constructed). A summary of each phase is given below:

1) **Exploratory modelling:** consists of exploring the spatial relations between the location of the areas that people perceived as unsafe or safe and a) people's neighbourhoods and daily routes and b) the location of crime incidents. For this, five related spatial variables will be extracted from the sketch maps by performing spatial queries. These variables will be treated as covariates for a bivariate logistic regression analysis, with the purpose of defining which variables explain higher percentages of the variability of the likelihood of perceiving an area as unsafe. The resultant significant factors can afterwards be explored in a spatial context to uncover the spatial relations between them for an integral understanding of the crime perception accuracy spatial arrangement.

2) **Spatial delineation of the perception accuracy:** comprises the comparison between the perceived and a reference safety classifications. The 'perceived classification' will be derived from the structured sketch maps, through which participants basically classified the city into unsafe and safe areas by drawing polygons over a base-map. Meanwhile, a 'reference classification' will be based on the crime hotspots.

   Both classifications will be compared to define areas that are accurately or inaccurately classified. Then, the level of accuracy will be estimated with the percentage of 'correctly' perceived classifications.

   Last, a bivariate spatial correlation analysis will be performed to find out the possible relations between two variables that could explain the accuracy of people's crime perception.

3) **Development of a GeoVisual Analytics environment:** consists of designing the GeoVisual Analytics environment that encompasses: a) the description of the potential users, b) a statement of the problems that the tool is designed to solve, c) the questions that can be answered with it, d) and the list of functions and the design of the interface.

Figure 1.1 shows a diagram in which the connection between the three objectives is illustrated. The specific objectives follow three of the Principles of Geography that are the fundamental concepts for any geographical or spatial study (Hagget, 1979). The explanatory modelling will set the **causality or origin** of the problem, understood as the factors that sway crime perception; the spatial arrangement outline is focused on the **location or distribution** of the perception of crime and the geovisualization development is directed to exploring the **relations** between the factors and location.

Therefore, the results of objective one define the variables that are relevant for understanding crime perception. The main output of objective two is the location of the perceived safe and unsafe places and whether this perception is accurate or not. In order to understand this location, we go back to the results of the first objective. The analytical tool developed in objective three is meant to enable the visual analysis of the spatial distribution of the significant explanatory variables and the perceived unsafe places.

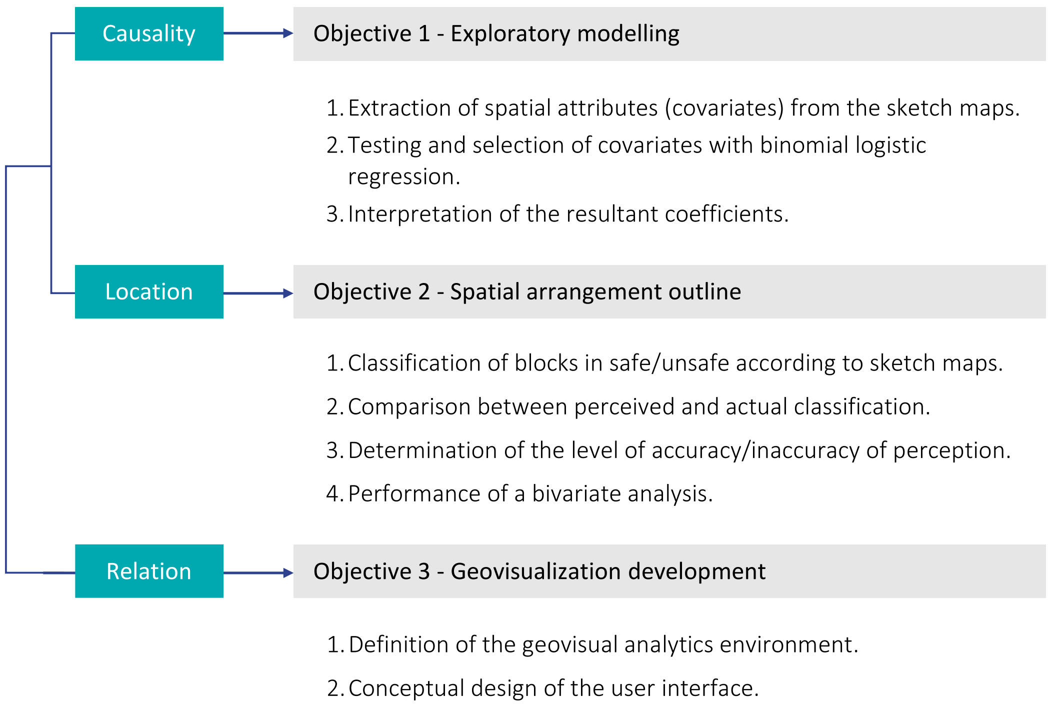



**Figure 1.1** Stages of the research process.

The three phases are intended to show the relevance of the extraction of data from sketch maps and its analysis in perception studies. Moreover, they aim to demonstrate how an integral analysis can contribute to the understanding of the spatial expression of people's perception. In the case of crime perception, the aim is to bridge the gap with located actions based on the comprehension of the related social and spatial factors that influence the perception of crime. The inquiry of the spatial attributes that are associated with those factors, such as location, distances, neighbouring elements and topological relations might contribute to the design of targeted strategies to narrow the crime perception gap.

This research makes use of a case study in which the presented methodology will be implemented. The data available correspond to an online survey conducted in Budapest, Hungary. It consisted of drawn structured sketch maps to depict the safe and unsafe areas in the city. The dataset was gathered in 2017. These sketch maps will be analysed in the exploratory modelling stage. For the second stage these sketch maps will be compared with the locations of the hotspots. The hotspots will be defined by the crime incidents that were reported in Budapest in 2017. In the last stage, a prototype of a geovisual analytics environment will be presented employing these data to exemplify the functionalities.

## 1.4 Thesis structure

This thesis is organized in seven chapters, including the presented introduction, in which a brief research context, the main and specific objectives, the identified research questions and a general outline of the implemented methodology were described. In the second chapter, a literature review is presented as the theoretical and conceptual framework that endorses this research project. The third chapter contains the description of the datasets, the preprocessing and geoprocessing procedures performed, as well as the list of software used. In the next three chapters, the methodology, results and conclusions of the exploratory modelling, spatial arrangement outline and the geovisualization environment development are presented. In the seventh chapter, the answers to the research questions will be summarized and general conclusions and recommendations for further research will be stated.

# Chapter 2

## Maps and perception of crime

This chapter is divided into four sections. In the first one (2.1) the distinction between the two main concepts **perception** and **cognition** will be addressed, as well as their connotation in spatial and cartographic terms. The difference between **cognitive mapping** as a process and a **cognitive map** as an internal cartographic product will be defined, to consequently continue with the introduction of **sketch maps** as the external representation of cognitive maps. The relevance of sketch maps as a data collection method for perception data will be discussed, especially in the perception of safeness, as well as the use of GIS for their analysis.

The second section (2.2) will be centered on the concepts of risk, fear and, mainly, the **perception of crime**. The last concept is the most relevant in this research. Several theories that explain the factors that sway the perception of crime will be expounded on briefly, including the influence of heuristics as one of the main emotional factors that mold perception.

Section 2.3 defines **perception accuracy** and a classification of it will be presented, differentiating between two types of accurate perception and two types of inaccurate perception. In crime studies, these latter are known as the **crime perception gap**. Some of the negative effects of the crime perception gap will be mentioned and, therefore, the need of narrowing the gap, as well as the importance of executing integral data analysis for the development of strategies that bridge the gap. The last section (2.4) is the conclusion of this chapter.

## 2.1 Perception, cognition and maps

Mapping is part of human nature. Maps as simplified spatial images have been in force even before the writing was developed (Raisz, 1985). Mental maps can be considered as the very first maps. The conception of this mental images of our surrounding environment based on individual perceptions and spatial knowledge have been created due to the need of being aware of where we are standing and the need of knowing our nearby space. This mental process implies being conscious of the attributes and relative location of objects and places.

Human perception refers to the process of acquiring information through the senses. The study of perception was initiated by psychologists, whose focus was the inquiry of the mental process of bringing together sensorial information, usually detached from the physical context. During the second part of the last century, this paradigm changed. Geographers pointed out the strong relationship between perception and geographic studies, because geographic space was considered as a mental conceptualization. Wood (1970) mentions that people's perception of space has a noticeable effect on their behaviour. Following Wood's theory, William Kirk proposed that Geography should be divided into two major areas: phenomenal environment and behavioural environment. For the latter, now known as Behavioural Geography, perception is defined as a selective gathering of images and ideas coming from the interaction with the environment and linked together with previous knowledge, memories and values (Wood, 1970).

It is not a trivial task to define the difference between perception and cognition because both of them are conceptually strongly linked and is difficult to mutually exclude them. There have been some attempts to outline them in a clearer way. Downs and Stea (1973) expound on that both are related to the organization and interpretation of the information, however the difference is that perception is "*the process that occurs because of the presence of an object, and that results in the immediate apprehension of that object by one or more of the senses*", while cognition is the process that happens in a second frame because is not linked with the immediate context. For Stea (as cited in Downs & Stea, 1973) the difference lies in the scale, as "*cognition occurs in a spatial context when the spaces of interest are so extensive that they cannot be perceived of apprehended at once*", and he suggests that cognition is a more complex process than perception, as the last is only concerned about "*briefer spatial perceptions*".

In Cartography, cognition is explained under the concept of cognitive mapping that is a mental process that consists on "*create and collect, organize, store, recall and manipulate information about the spatial environment*" (Downs & Stea, 1977). Space perception is then understood as an encompassed subprocess within the cognitive mapping major process, which embraces the creation of mental images of a given space. The output of the cognitive mapping process is a cognitive map that includes, not only information about relative locations, relative distances, geometries and directions, but also about non-visible characteristics of features and places (Matei, Ball-Rokeach & Qiu, 2001; Golledge, 1997; Downs & Stea, 1973).

Kevin Lynch in his book 'The Image of the City' (1960) describes five basic elements of urban structure that constitute the base of an urban cognitive map: paths, boundaries, districts, nodes and landmarks. Therefore, a cognitive map of an urban environment is an inner image that comprehends the urban base, defined by the five basic structures, with a non-metrical arrangement but a relational one, plus the physical and non-physical attributes, all defined by the individual cognitive process. Downs and

Stea (1973) identify two types of attributes: descriptive which are "*affectively neutral*" and evaluative that are "*affectively charged*". The selection and allocation of these attributes is conditioned by internal factors such as "beliefs, values, and attitudes" (Golledge, 1997) and external factors such as social responses, temporal, cultural and physical context.

Cognitive maps are individual, not tangible spatial models; to refer to their physical depiction, the term cognitive representation or cognitive configuration is used. Sketch maps appear to be the most use of the cognitive representations, especially to collect information from individual perceptual knowledge and individual reliable spatial information (Blades, 1990). They have been used as an implement to capture people's perception for different purposes, for instance, decision making, wayfinding, planning, risk management and marketing (Golledge, 1997).

There are two different types of sketch mapping: "*free recall*" (Figure 2.1) and "*structured sketch mapping*" (Figure 2.2). In the first method, the map is drawn on a blank paper, in the second, the features are sketched over a consistent printed or digital base map (Sloan et al., 2016).

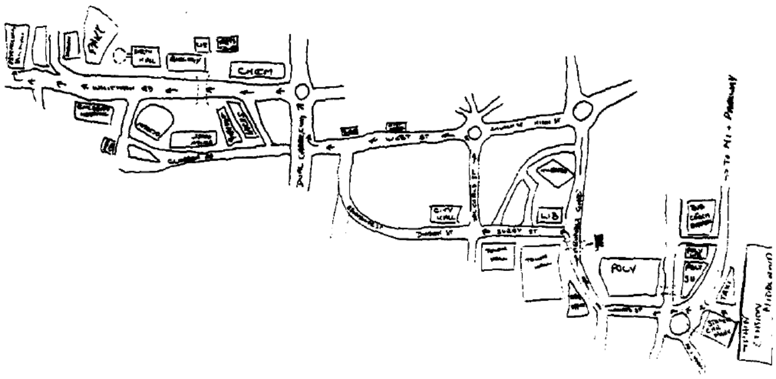



Figure 2.1 Free recall sketch map that shows the way to a railway station (Blades, 1990).

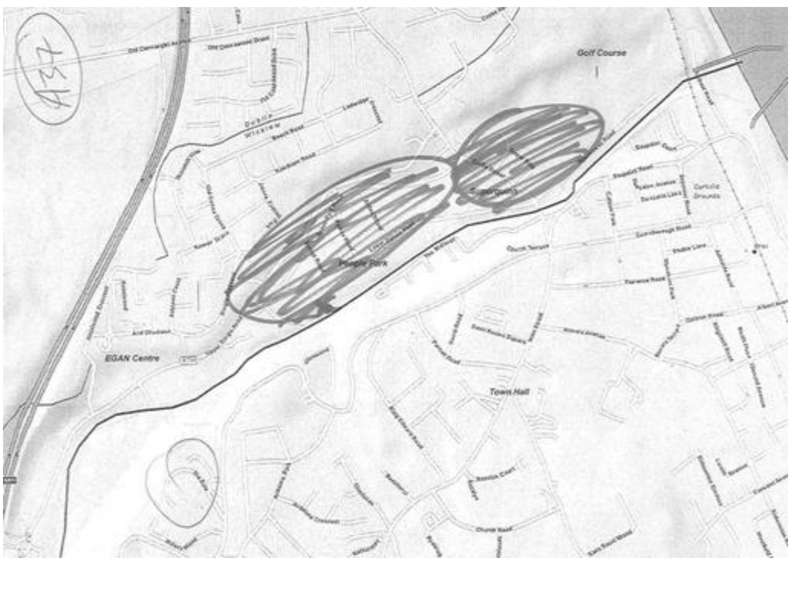



Figure 2.2 Structured sketch map that depict perceived risk of flooding (O'Neill et al., 2015).

The analysis of structured sketch map is eased by the use of Geographic Information Systems (GIS) as they have a consistent special reference. Curtis, Shiau, Lowery, Sloane, Hennigan and Curtis (2014) present a review of twelve resent studies that integrate GIS and sketch maps, where spatial processes like data aggregation, patterns analysis, overlapping operations and raster analysis are performed. The usage of GIS for the analysis of sketch maps allows an integral study of perception (Curtis, 2012) as more spatial data can be incorporated in the analytical process. On the other side, the studies of the spatial features and phenomena which incorporates perception data can lead to a more complete characterization of the space and bring rounded conclusions.

The use of sketch maps in spatial perception researches has been a common practice especially in crime perception studies (Curtis, 2012). In this case, sketch maps are used to capture the external form of people's safeness cognitive map, which usually depicts the safe or unsafe areas based on the person's insight. The most common method of data collection is to ask the survey respondents to identify, usually, in a printed base map (Curtis et al., 2014; Spicer, & Brantingham, 2014; Kohm, 2009; Matei et al., 2001), the places or areas where they think there is a higher risk of victimization. Often the sketch map goes with a questionnaire to characterize the drawn map or a think-aloud process is performed to add extra information (Lopez & Lukinbeal, 2010). Other method incorporates the use of digital media that allows an integral data collection as the data is recorded *in situ*, the volunteers are asked to carry a mobile device to record information while walking along an area where crime fear is triggered (Solymosi, Bowers and Fujiyama, 2015). In a similar way, Chataway, Hart, Coomber and Bond (2017) present an Ecological Momentary Assessments to collect context-dependent perception.

The aim of gathering and analysing safeness cognitive maps is to create a spatial model of the "*imaginable*" safety qualities of the environment (Pocock, 1979). It is relevant to know the collective perception that the inhabitants have assented of public spaces, as having an inaccurate perception of safety can have an impact on people's daily life, social behaviour and spatial dynamics. This is why crime perception has become an attention-grabbing area of study not only in Criminology but also in Geography and other Spatial Sciences.

In the next section three major concepts in crime perception studies are explained: perception of crime risk, fear of crime and crime perception. The limits between them are hard to set as they are closely linked together. Nevertheless, there are key ideas that can help distinguish them in a clearer way.

## 2.2 Crime: risk, fear and perception

Crime is an aspect of social life that has been studied from different perspectives as it involves an assortment of social, psychological and geographical aspects. Diverse sciences have taken part in the identification and understanding of the factors that are involved in a criminal event. Beyond from the events itself, different studies have tackled their social impact. One of the main social concerns is the risk of becoming a victim. The feeling of insecurity is mostly triggered by the fear of crime as an emotional response and the crime perception as a cognitive assessment (Foster, Knuiman, Wood & Giles-Corti, 2013).

Perception of crime risk and fear of crime are both related to worry and uncertainty, the difference between them lies on the temporality of the response. Fear of crime is an emotion shown as an immediate reaction in the face of a proximate threat, while the perception of crime risk is generated by "distant potential harm" (Jackson & Gouseti, 2014). The perception of risk is the result of a cognitive process that can combine two directions of thoughts, one based on objective information and the other one influenced by emotions. Loewenstein, Weber, Hsee and Welch (2001) defined these two directions, risk as a feeling, which is a perception bias by an intuitive and emotional way completely disregarding the real probability of victimization, and risk as analysis, which is defined by logical and impartial information.

Crime perception researches study crime risk as a feeling, the perception of risk is based then in the subjective probability of becoming a victim (Jackson & Gouseti, 2014) and commonly the risk has a negative connotation related to dangerous situations (Kemshall, 1997). Perception of crime risk is modelled by the envisioned vulnerability of becoming a victim of a criminal offense; it is an enduring, but not permanent, conception in time.

The perception of crime risk can be altered by an internal and external stimulus in a given situation, triggering an immediate emotional response of fear of crime. The fear of crime is defined by Brantingham and Brantingham (1995) as a "*condition created by a certain spatial and temporal context in which a person feels vulnerable to become a victim of criminal attack*". This context is not necessarily defined by a high crime environment, and various researches have concluded that the perceived crime risk and people's fear are not related with an actual high probability of victimization (Lewis & Maxfield, 1980), which means there are other factors that sway this type of people's responses.

Dorand and Burgess (2011) present a review of different theories that explain the causes that trigger fear of crime and having a high crime risk. The theories are grouped according to the factors that may explain the trigger of these emotional responses (Figure 2.3).

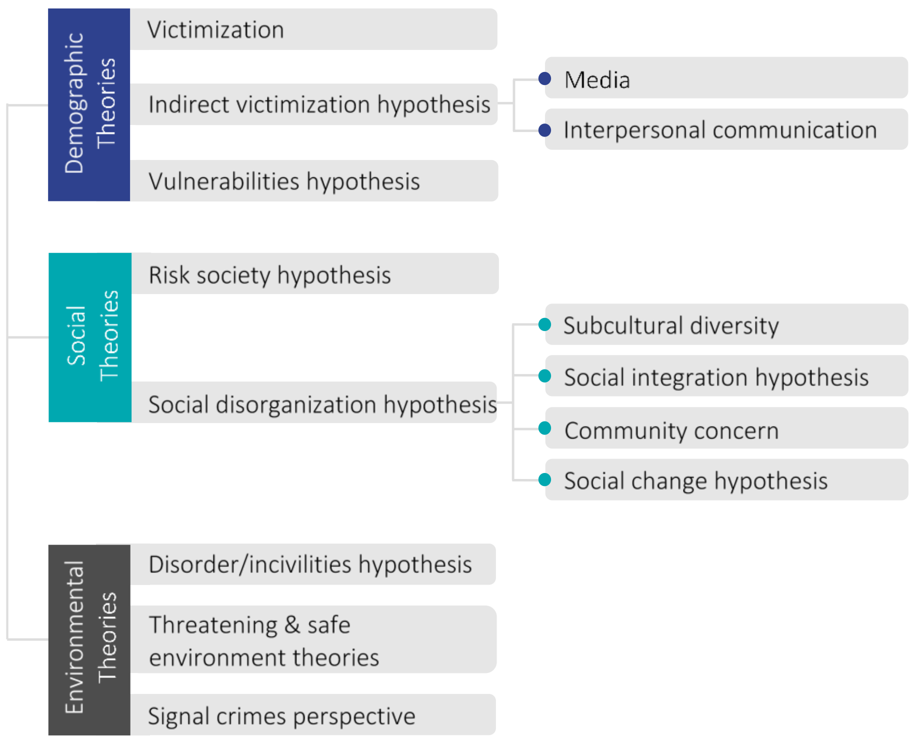



**Figure 2.3** Theories that explain fear of crime and perception of crime risk (based on Dorand and Burgess, 2011).

The demographic theories associate previous experiences of victimization and some demographic characteristics with a higher crime fear, for instance, women and elder people tend to feel more vulnerable. The social theories relate fear and risk with the social disintegration and lack of organization in a community. Meanwhile, the environmental theories expound on the social and physical characteristics of the landscape as a factor of crime fear and perception of a higher risk of victimization. This categorization does not mean that the theories are mutually exclusive and although there are researches that support these theories, there are others that contrast them. A brief description of the hypothesis is presented below (see Dorand and Burgess, 2011):

### Demographic theories

- Victimization hypothesis: people with previous experience of direct victimization tend to feel more vulnerable and perceive a higher level of risk (Crank et al., 2003; Mesch, 2000; Skogan & Maxfield, 1981).
- Indirect victimization hypothesis: non-victims sense the same fear as a direct victims when they know about someone's crime encounters usually through the media and interpersonal communication (Clark, 2003; Hanson, Smith, Kilpatrick & Freedy, 2000)

  The Media: media aggravates perceptions of risk of victimization, through different approaches: cultivation, substitution, resonance, social comparison and interpersonal-diffusion (Lane and Meeker, 2003).

  Interpersonal communication: victims' experience of victimization spreads through communication networks, non-victims increases their fear of crime and the perceived risk of victimization (Mawby, Brut & Hambly, 2000; Taylor and Hale, 1986).
- Vulnerabilities hypothesis: the level of fear of crime varies for every sociodemographic group; each one believes is more vulnerable to criminal victimization, for example, women and elderly (Warr, 2000; Liska, Sanchirico & Reed, 1988).

### Social theories

- Risk society hypothesis: people tend to feel in danger and threatened from unknown situations as result of anxiety condition; their fear is extended to other (Lianos & Douglas, 2000; Beck, 1992).
- Social disorganization hypothesis: segregation of social organization breaks communication channels preventing the maintenance of public order, which derives into crime and delinquency (Sun, Triplett & Gainey, 2004; Cochran, Bromley & Branch, 2000; Taylor & Covington, 1993).

  Subcultural diversity hypothesis: fear of crime is developed when people live close to someone from different culture (racial diversity), due to, their "unknown" behavior (Lane & Meeker, 2003).

  Social integration hypotheses: the lack of social integration, communication and support within a community increases the fear of crime (Crank et al., 2003; Markowitz, Bellair, Liska & Liu, 2001).

Community concern hypothesis: when a community declines socially and physically, inhabitants and external people develop a state of caution and fear of crime (Lane & Meeker, 2003; Covington & Taylor, 1991).

Social change hypothesis: fear of crime results from people resents the process of social changes, for instance, diversification of races, declination of economy and increase of unemployment. Fear develops because of the changes in space (Clark, 2003; Furstenberg, 1971).

Environmental theories

- The disorder/incivilities hypothesis: the social -drug users, gangs, beggars- and physical -abandoned cars, damaged buildings, graffiti- characteristics of an environment have an influence over people's fear of crime. Disorder and incivilities generate an image related to criminal activity and vandalism (Millie & Herrington, 2005; Crank et al., 2003; Tulloch, 2000; Nasar, Fisher & Grannis, 1993).
- Threatening and safe environments theories: areas with certain characteristics, not necessarily disorder or incivilities, are considered apt for criminal victimization; for instance, a street with poor lighting, overgrown vegetation and alleyways (Cozens, 2002; Kuo & Sullivan, 2001).
- Signal crimes perspectives: crime and disorder affect the people in a different way and with a dissimilar intensity; also each person interprets them with different connotations (Innes, Fielding & Langan, 2002).

Perception of crime risk and fear of crime play an important role in the cognitive mapping process. As mentioned before, this process is about designing a mental image of the geometry and characteristics of the space. The perception of risk and the fear of crime define the descriptive and evaluative attributes (Golledge, 1997) related to safeness in our cognitive map.

This attribution is based on the subjective probability of victimization that is associated with the perceived amount of criminal activity, this is known as crime perception. While, spatial crime perception defines the characteristics of a location in term of safeness, commonly categorized as safe or unsafe. Hence, perception and fear set attributes to objects and places shaping the space in a cognitive map as an arrangement of perceived and real spatial characteristics.

Brantingham and Brantingham (1981) identify four theories that explain the criminal activity based on environmental criminology, which theorizes about the influence of the environment on victimization and criminality. Spicer, Song and Brantingham (2014) took these theories and diverted them to explain the crime perception from a spatial point of view:

- Routine activity theory: it considers three elements: the victim, the offender and the location. During the daily activities, there are scenarios with "non-capable guardians" where the perceived offender can find an opportunity to victimize an individual. These scenarios are produced in certain routes and time of the day, which can trigger a fear feeling as a response to the perceived situation.

- Rational choice theory: it lies over the supposition that people's actions are based on a previous decision-making process. Under this theory, the perceived risk of victimization influence people's actions; decisions are taken considering potential risk and possible consequences, but not always risky situations or spaces can be avoided.
- Geometry of crime: people build-up an activity space in which their daily routines and routes happen, as far as possible, it will correspond with an awareness space where situations and places of perceived high risk of victimization can be avoided.
- Crime Pattern theory: people define 'safety templates' to avoid victimization creating a cognitive map.

These schemes place the perception of crime in a spatial context, they expound on how perception is related to decision-making, awareness space and activity spaces. This means there is an intrinsic relationship between perception and space. The foundations of the environmental theories explain how the physical context of public spaces is relevant for shaping perception of safety. Researches that support environmental theories have identified the level of incivility as one of the principal features that underpin the relation perception-space (Kohm, 2009; Lewis & Maxfield, 1980).

Millie and Herrington (2005) recognize two aspects of incivility: disorderly physical surroundings -such as graffiti, abandoned buildings, litter- and disruptive social behavior –beggars, gangs, street drinkers and drug addicts-. This kind of physical conditions and social conducts prompts concern and fear as they reflect apt scenarios for criminal offenses to occur, and that the image is that people perceived, creating a sense of danger and thus, setting an unsafety attribute.

Although physical attributes plays an important role in crime perception, there are also social factors that have an impact on it. Lora (2016) points out that safety perception is "*strongly influenced by the affect and availability heuristics*". Heuristics are mental shortcuts for decision-making based on the promptly available information.

The affect heuristics are "*due to proximal cues and due to feelings of trust*" (Lora, 2016), they describe how the assignation based on emotions or attachment feelings can affect the judgment of risk. In this direction, Carvalho and Lewis (2003) explain that the crime perception is also shaped by how distant or linked people are related to security issues, although it is an aspect of social and daily life, some people consider them as a more salient problem than others. These problems then overshadow people's daily life, generating a higher unsafe feeling. Contrary to people who feel more distant or detached to these problems, their reactions tend to be neutral or more objective.

Meanwhile, the availability heuristics (Tversky and Kahneman, 1974, as cited in Jackson & Gouseti, 2014) "*predicts that the probability of an event tends to be judged by the ease with which instances of it can be retrieved from memory*" (Jackson & Gouseti, 2014). These memories mainly referred to direct and indirect victimization (Figure 2.3 and the information available in the media (Lora, 2016) which produce a constant image of risk. Consequently, the probability of victimization tends to be perceived as high in such a way that increase the identification of fear spots, which are the places where people feel more vulnerable to criminal attacks but there is a low crime rate (Fisher & Nasar, 1995).

It is more common then, that people tend to overestimate the crime rate or to misidentify the unsafe areas. However, there is also the case when people underestimate this rate and are not aware of the high risk of victimization. Both misconceptions can have negative outcomes as people can develop cognitive maps that are far away from reality, indicating a gap between the perceived attributes and the actual ones. This is known as perception gap. When these attributes are related to safeness and crime incidents it is called the crime perception gap.

## 2.3 The accuracy of crime perception

Both social and physical environmental factors have an impact on people's crime perception. The safety cognitive map is an individual spatial model in which each person attributes an area subjective characteristics in term of safeness. Usually, this attribution does not coincide with the actual one. It is frequently the case where there is a misconception of the current crime rate. This is known as crime perception gap and is defined as the difference between the level of insecurity that is conceived by a person and the actual level of insecurity based on actual crime incidents (Mohan, Twigg & Taylor, 2011).

The gap can be present in two ways: it is believed that the crime rate is higher than the actual rate or, inversely, it is thought that the crime rate is lower than the actual rate. Instead of in terms of an ordinal scale, the crime perception gap can also be referred to on a nominal one, by misclassifying the safe areas as unsafe and the unsafe areas as safe. Usually, these terms are more current in use as in this case, perception is a qualitative assessment rather than a quantitative one. Figure 2.4 shows this binary classification.

| PERCEPTION \ REFERENCE | SAFE | UNSAFE |
|---|---|---|
| SAFE | Accurate perception of safe area (AS) | Inaccurate perception of safe area (IS) |
| UNSAFE | Inaccurate perception of unsafe area (IU) | Accurate perception of unsafe area (AU) |

**Figure 2.4** Types of crime perception accuracy based on safety attributions.

Research of crime perception performed in different countries like the United Kingdom, Australia, South Africa, Colombia and the United States have identified the existence of gaps in people's perceptions (Mohan et al., 2011*).* It is more common to find people who believe that crime rate, either stays more or less constant or keeps rising, even though the statistics can prove these ideas wrong (Millie & Herrington, 2005), than people thinking that there has been a reduction of criminal events. This overestimation is not only related with the number of incidents but also depend on the type of crime; Pfeiffer, Windzio and Kleimann (2005) found that people think that the type of crime that they are more vulnerable at, is the one which has increased the most.

In spatial crime perception studies, this gap is usually distinguished by comparing safety sketch maps where people identify unsafe areas and maps that depict the location of crime hotspots. When overlapping both, the identified areas frequently do not correspond with the hotspots. Which means, people tag low crime rate areas as unsafe ones, or also may happen that participants categorize unsafe areas as safe.

As explained before, IU is usually related to heuristics or incivilities, some people make this type of misclassification in areas where they are less familiar with or where they are not related. While the IS happens usually in people's own neighbourhood, that can be explained by the “endowment effect” that consist in assigning a higher or better value to the objects we possess, than to the same objects that we do not own (Kahneman et al. 1990 as cited in Lora, 2016). In crime perception, this can be applied when people tend to characterize their neighbourhood as safe, under the assumption that where people belong to, exist better conditions than the surroundings. People tend to have a perceptual bias due to a feeling of attachment toward the own community or neighbourhood (Duffy, Wake, Burrows & Bremer, 2008).

These social misperceptions can have an impact on different aspects. For the case of an IS, people are not aware of the high risk of victimization, thus appropriate precautions are not being taken and the probability of an attack increases. For the IU the impact can be on a bigger scale and its effects can last for a longer time, as it has repercussions on people’s lifestyle (Ardanaz, Corbacho, Ibarraran & Ruiz-Vega, 2013), health due to anxiety (Foster, Giles-Corti, & Knuiman, 2010), social behaviour and the spatial and economic dynamics (Doran & Burgess, 2012). A high crime perception can restrict the individual daily activity area of a person due to the avoidance of unsafe areas or streets at certain hours, thus people might have to change their daily routes. On those identified areas, fewer people would transit there, eventually forcing the relocation of shops, restaurants, or any business that could be affected, which will lead to a reorganization of the spatial activities.

Narrowing the perception gap also relevant as reducing the crime rate. Some researchers have determined that publishing data about crime statistics can have significant effects on people’s perceptions (Lore, 2016). Ardanaz et al. (2013) conclude that people improve the perception of safety by being more positive about police effectiveness and by reducing the perceived risk of victimization, due to people who feel well informed tend to be more confident about their safeness (Ardanaz et al., 2013), improving, in this sense, their quality of life.

The aim of mapping the crime is to communicate spatial crime data in the most objective way without increasing, the fear of crime to the readers. The first attempts to map the crime events were back in the nineteen century in France by Adriano Balbi and Andre-Michel Guerry: they mapped the incidents recorded from 1825 to 1827 (Weisburd & McEwen, 1998) (Figure 2.5) . While this kind of data, for many decades, was on the governmental domain, nowadays the philosophy of open-data have released it to the public.

The communication strategies must present meaningful information with a “*greater narrative-based*” (Duffy, Wake, Burrows & Bremer, 2008) other than plain statistical data. The way it is presented can have a different impact on the reader; it can be used as a fear-reduction strategy, but also can have an inverse effect, as people can become anxious about the high crime areas shown in the map.

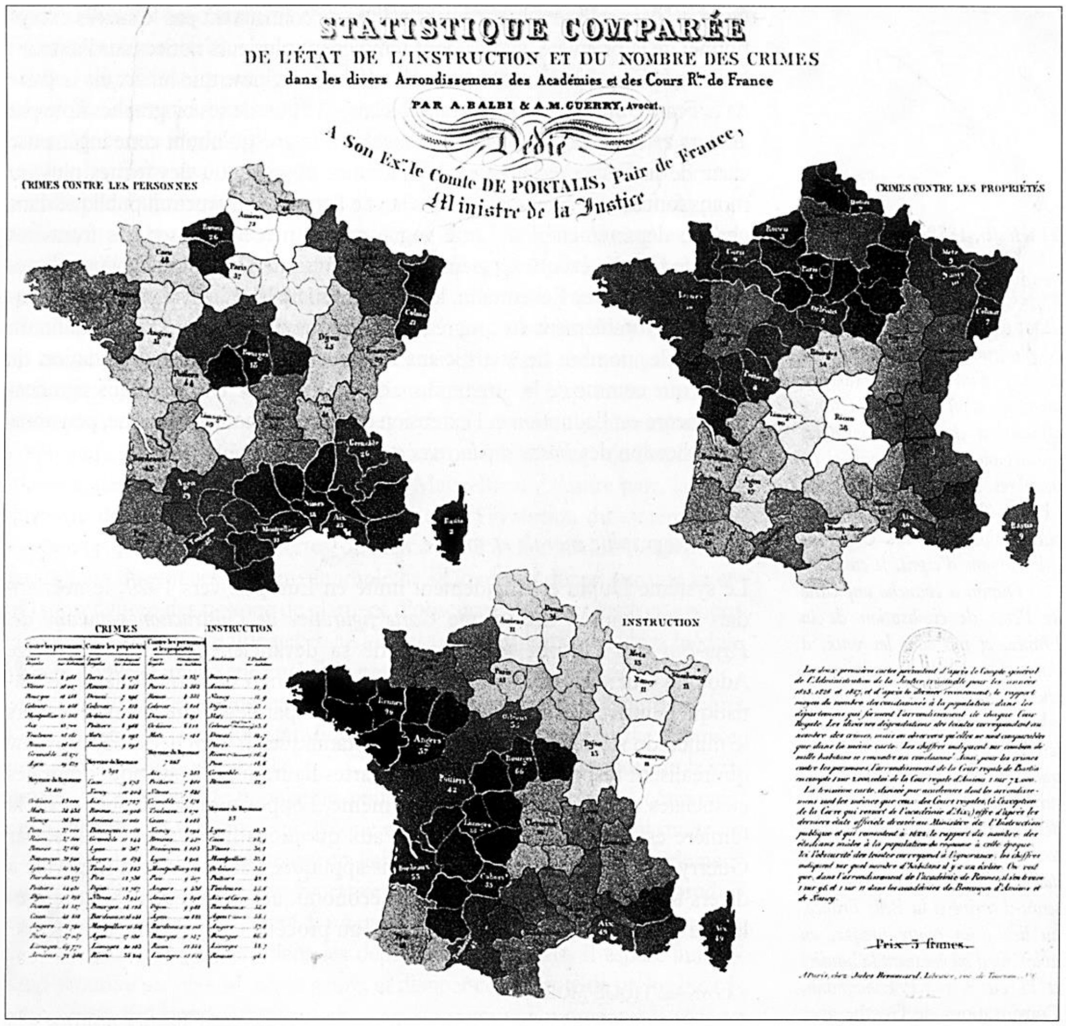



Figure 2.5. One of the first crime maps, by Adriano Balbi and Andre-Michel Guerry in 1829: "Statistique comparée de l'état de l'instruction et du nombre des crimes" (Comparative statistics of the state of education and the number of crimes) (Friendly, 2007).

Maps are one of the most common representations to visualize crime data, but still, there is still a lack of studies about the impact of maps on citizen perceptions. The most representative research in this domain was conducted by Groff et al. (2005). Based on the experiments, where they compare the impact of different representation methods –tables, graduated symbol maps and density maps-, they conclude that "graduated symbol maps as the overall preferred method of crime information transmission to citizens without significantly increasing fear of crime".

Mapping the perception gap can also have an impact on people's perception, but it could be more apt for decision making as the competent authorities would be able to identify the places where the gap perception is, and then develop an action plan to narrow it by informing the citizens about the current situation. The actions should be towards recovering the confidence of the inhabitants, not only by reducing the crime rate but also by increasing the feeling of safeness.

Police agencies have always been responsible for designing strategies for the reduction of fear of crime (Cordner, 2010; Grabosky, 1995; Bennett, 1991). Cordner (2010) identified 12 "*fear-reduction hypotheses*" which are strategies that could work as possible solutions. These include reduction and prevention of crime, policing located actions, police-citizen contact and reducing disorder. In order to turn these strategies into action plans, it is necessary to have a general perspective of the location of the crime perception gap and the factors that sway perception. The integration of spatial information could ease the design of strategies with defined target actions and extension.

Therefore, understanding and mapping crime perception and its origins is the primarily needed for developing target strategies to narrow down the crime perception gap.

## 2.4 Conclusion

A cognitive map is an inner image that combines subjective information based on the way a person perceives and mentally structures the surrounding environment. The physical representation of a cognitive map is the sketch map. In spatial crime studies structured sketch maps are the main source of data collection to capture people's insights of safety.

People's perception of crime can be swayed by numerous factors that have been studied and gathered in demographic, social and environmental theories. Some of them consider the spatial attributes as explanatory variables. The importance of analysing these factors is to better understand how crime perception works as it is not always consistent with reality. Having an inaccurate perception of crime, that is a disparity between perception and reality, could retract from the quality of life, change social behaviour and spatial dynamics.

Therefore, it is needed to implement actions focused on increasing the accuracy of perception and consequently narrow the gap. Police agencies are the competent authority to develop strategies directed to reduce and prevent crime, but also to implement actions to reassure or make people aware of the current state of safety, depending of the case.

To develop effective strategies first is needed to recognize the overall spatial context where the problem is. This approach is followed to expound on the case-study presented in this research. In this case, two components will characterize the spatial context: the factors that sway people perception of crime in Budapest and the spatial arrangement of the crime perception accuracy. In Chapter 4, the first component is addressed by exploring structured sketch maps. The second component is discussed in Chapter 5 by comparing the location where people identified unsafe/safe places and the actual unsafe/safe places.

The methodology used in Chapters 4 and 5 is different and will be treated separately, but as the data used in the case-study is the same, the dataset, as well as the software used in the analysis will be discussed first in the next chapter.

# Chapter 3

## Data description and software

In this chapter, the **datasets** that are used in the case-study will be described, as well as the **preprocessing and geoprocessing** (section 3.1) transformations which were required to perform the analysis later on. These include selection, classification, aggregation, data correction and geocoding. Some processes are described in detail as they are relevant for the selected methodology.

Additionally, the **software and web application** (section 3.2) that were used in this thesis are briefly described, including the stages in which they were executed.

## 3.1 Data: description, pre-processing and geo-processing

The data used in this study is derived from an ongoing participatory online survey (http://bunmegelozes.amk.uni-obuda.hu/) conduct at a national level in Hungary by the Institute of Geoinformatics from the Óbuda University. The initial participants were students from the University, afterwards, the survey spread out by a snowball effect. The data used in this research was collected in 2017 and is constrained to the city of Budapest.

The survey consisted to draw a digital structured sketch map over a web-based map, in which the participant indicated the areas that he or she perceives as unsafe or safe, similarly, they marked with lines their daily routes. Furthermore, they were asked to give some identifying information such as their age, sex, postal code where they live (Figure 3.1) and the main mean of transportation they use.

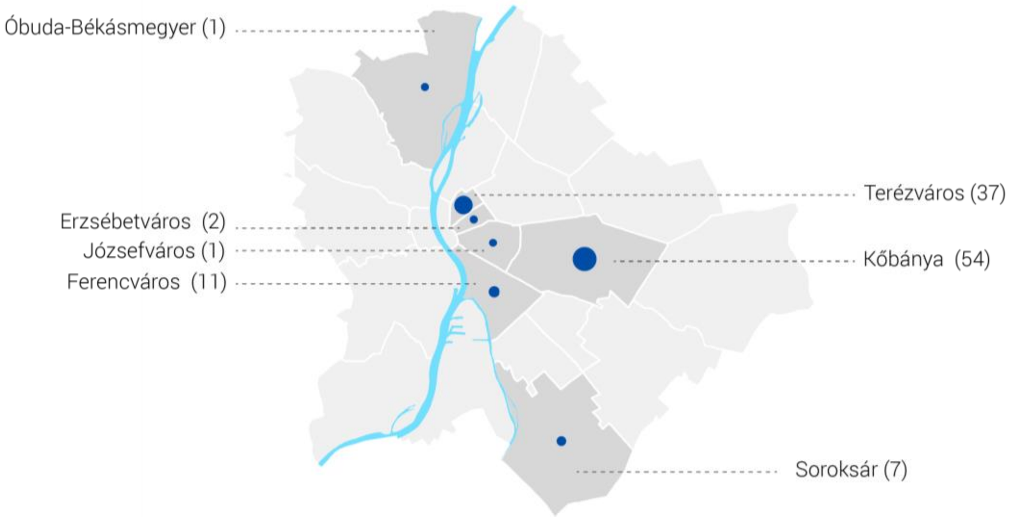



**Figure 3.1** Number of participants per district in Budapest (total number of participants = 113).

Table 3.1 is a summary of the structured sketch maps of the 113 participants. From the resultant digital sketch maps, three vector files were extracted: perceived safe areas (97 polygons) (Figure 3.2 A), daily routes (214 lines) (Figure 3.2 B), and perceived unsafe areas (231 polygons). The data sets were filtered by attributes and location. For the first type of filter, if the participants did not provide personal information (age and sex), their sketch maps were not considered as a quality control measure. For the second filter, if the polygons or lines exceeded the boundaries of Budapest, they were clipped to limit the analysis within the city. In total, there were 113 participants, 39 women and 74 men, between 18 and 76 years old who drew their daily route(s) and at least one polygon.

| Gender | Total participants | Daily routes | Polygons | |
|---|---|---|---|---|
| | | | Safe | Unsafe |
| Women | 39 | 69 | 31 | 84 |
| Men | 74 | 128 | 66 | 147 |

**Table 3.1** Summary of the structured sketch maps by gender and sketched element.

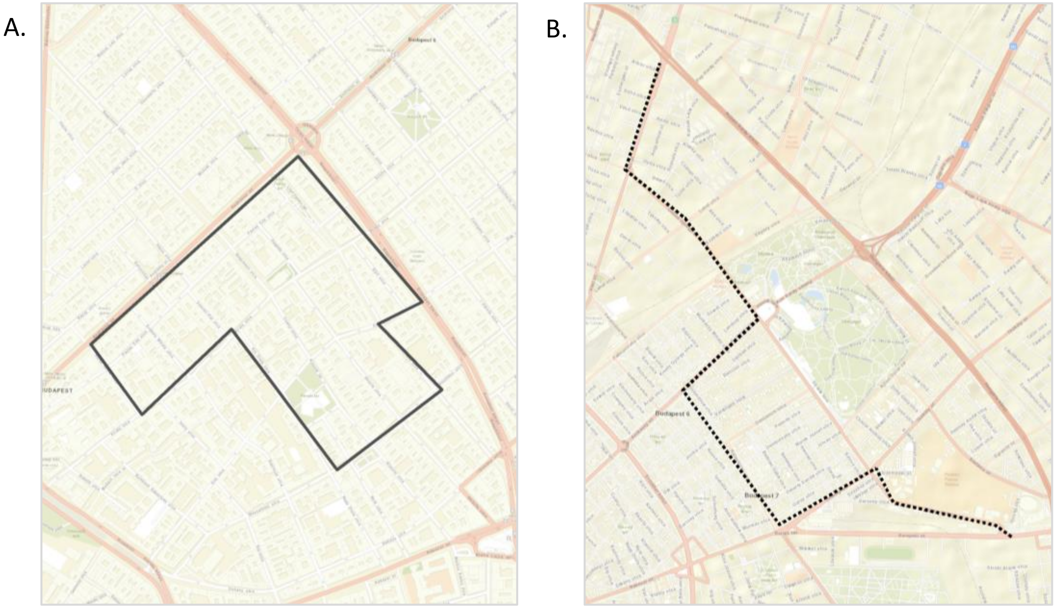



**Figure 3.2** Examples of the structured sketch maps from the online survey. A) Perceived safe area sketched with a polygon. B) Daily route sketched with a line (black dotted line).

Moreover, the Óbuda University provided a CSV file with 60,784 addresses of the recorded crime incidents in Budapest during 2017. The addresses were geocoded in QGIS using the plugin MMQGIS with the web service Nominatim, a search engine for OpenStreetMap data. Part of the pre-processing of the data was the replacement of two Hungarian characters, ő and ű, because they could not be recognized by the UTF-8 encoding required by Nominatim. The characters were replaced with "o" and "u" respectively, subsequently, five addresses that contained these characters were searched in the website of OpenStreetMap in order to compare the output location of the addresses that included the original characters and the same addresses but with the mentioned change. The result was that both addresses were geocoded in the same location. So the changed of characters did not affect the geocoding.

From the data cleaning process, 1,218 records (2%) were deleted due to the lack of an address. The process was run with a set of 59,566 records, from which 58,379 addresses were geocoded, that equals to a hit rate of 98%. According to Ratcliffe (2004), a minimum geocoding hit rate (percentage of record successfully recorded) of 85% is needed to produce an accurate map which reflects the actual distribution of the criminal events. In total, there were 1,187 (2%) addresses that could not be geocoded, some of the reason were due to misspellings mistakes, the use of non-recognized abbreviations, mistaken street types or because the record was not an address but a location or the name of a place instead.

The original dataset contained crime incidents of spatial or non-spatially-explicit nature such as fraud, crimes against computer system and data, health related, misuse of documents and blackmail. Thus, the crime data passed through another filtered process as the research is directed to the analysis of street crimes, which are the criminal offenses that happen in public places. The data was reduced to 42,805 reports of 41 types of crime, which were grouped in 9 street crimes classes listed in Table 3.2.

| Type of crime | Count |
|---|---|
| Theft | 19,352 |
| Disturbance and vandalism | 11,747 |
| Distribution and drug consumption | 4,356 |
| Larceny | 2,670 |
| Assault | 2,454 |
| Harassment | 2,074 |
| Rape | 102 |
| Homicide | 47 |
| Kidnapping | 3 |

**Table 3.2** Number of incidents per street crime type.

The geocoded points were spatially joined to the city blocks. The blocks vector layer was extracted from OpenStreetMap. The streets were manually digitalized in order to assure that all the streets were connected, so that afterwards the line features could be transformed into polygons. The process was done manually due to the fact that the line vector file from OpenStreetMap contains a large variety of types of lines which makes the selection process difficult as one street can combine different types of lines. If one type is excluded, a segment of the street would be missing and consequently the shape of the polygon would change.

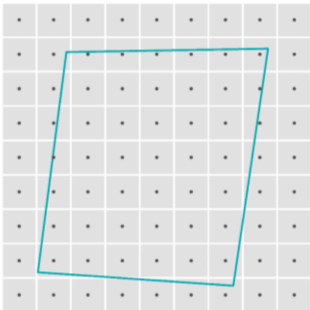
The point aggregation in blocks was done due to the quality of the geocoding results. For some addresses the points were located in the centroid of a block, mainly when the address corresponded to a specific public place such as a mall, park, airport or train station. The difficulty with these points is that this type of places tends to be the scenario of multiple incidents. So, in the same pair of coordinates there could be more than one hundred points. Thus, grouping the points by block allows a characterization of the block in which the place is contained and not of a single point location.

Addressing research objectives one and two implied some analysis processes and the datasets of the sketch polygons were transformed as the type of analysis required. The exploratory modelling stage (Objective 1) consists of the extraction of data from the structured sketch maps and their analysis. Meanwhile, the spatial delineation of the perception accuracy: (Objective 2) comprises the identification of the spatial distribution of the accuracy of perception of crime.

The first objective, the exploratory modelling, involved the extraction of attributes from the sketch polygons in order to understand the factors that could be involved in the selection of the areas which the participants identified as a safe or an unsafe area. As the sizes and shapes of the polygons were diverse, the aim was to characterize the polygon not as a whole entity with generalized attributes but to capture the different attributes within the area that the polygon covered.

Therefore, the drawn polygons were segmented into small analysis units. Working with the polygons as single samples would result in a non-precise analysis because a large area could be influenced by the attributes of that sample. Thus, polygons were split into cells with a rectangular grid in which the centroid of each cell was obtained (Figure 3.3. A), so that each centroid could represent one data sample. The cells length is 45x45m, the size was selected based on the smallest drawn polygon. So instead of analysing 328 polygons, 68,032 cells' centroids that were within the polygons (Figure 3.3. B) were explored. The centroids' data set includes an identification number, the participant's and polygon's ID, and the type of polygon to which that centroid belongs to, either a safe or unsafe identified area.

A.



B.

**Figure 3.3** A) Segmentation of the sketch polygons with a rectangular grid and its centroid. B) Selection of the centroids within the polygons.

By dividing the polygons into smaller units, it was possible to capture the heterogeneity of the spatial attributes within the area limited by the sketch polygon. To assure that this method was more appropriate for the aim of this research, the analysis was performed considering the whole polygons and the segmentation of them. The results were not satisfactory for working with the whole polygon.

The approach of the polygon segmentation is suitable for the analysis of sketch maps in the context of perception. As a sketch map is the external representation of an individual cognitive map, it has to be considered that each mental map has a different scale. From the sketched polygons, it can be assumed that the participants were working at different scales, due to, some of them visualized the problem in a big scale as they traced their polygons following the city blocks of the base maps, meanwhile, other participants saw the problem in a smaller scale, as their polygons do not have a structured shape and they did not followed the geometry of the city having a comparably bigger size. To dispel this differentiation and elude generalizations it was convenient to work with the smallest possible analysis unit. The purpose of segmenting the polygons was to characterize as precise as possible the sketch maps, due to, the polygons drawn are mainly irregular figures that cannot simply be generalized.

For the second objective, the spatial delineation of the perception accuracy, the aim is to identify the spatial distribution of the crime perception accuracy. This will be done by comparing two datasets the "reference classification" and the "perceived classification" of safe and unsafe areas in Budapest. To perform the comparison both datasets have to be in the same spatial unit. As the reference classification is defined by the actual number of crime events and these were aggregated by blocks, the perceived classification, defined by the sketch polygons, have to be transformed also in blocks.

The transformation of safe and unsafe sketch polygons into blocks was done with an intersection operation. The first step was to count the number of safe and unsafe polygons that intersect each block.

The second step was to label each block as safe or unsafe according to the highest percentage of intercepted polygons by type. Thus, the unit of both datasets, the perceived classification and reference classification, was set in blocks and this allow the comparison between them.

## 3.2 Software and web application

The software and web application used for the respective analyses performed in the exploratory modelling and the spatial delineation of the perception accuracy stage is presented below:

- PostgreSQL (with PostGIS extension in QGIS)

  The vector files were treated as spatial data tables with the aim of performing spatial queries for the data extraction of the vector files. Working with SQL eases the data analysis as it enables relating more than two vector datasets and allows the integration of different geometries in the same table, which facilitates the manipulation of the data. The queries were performed with the open-source relational database management system, PostgreSQL with the PostGIS extension. To visualize the results, this database management system was attached to QGIS.

- ArcMap

  Although working with PostGIS has many advantages, there are spatial operations that have less computational cost if they are performed in a GIS, due to, in a GIS a vector file contains the topology information which allows to perform the spatial queries in a more efficient way. ArcMap is the GIS used in this research. Also the maps were designed in this software, whereby, in some cases, some elements were modified or added using Adobe Illustrator.

- GeoDA

  This free and open-source software was developed by the Spatial Analysis Laboratory, University of Illinois, and lately its development continued in the University of Chicago (https://geodacenter.github.io/). It was used to perform a spatial autocorrelation analysis to define the crime hotspots that determined the unsafe areas. This software was also used to meet the second objective while performing a bivariate local Moran’s I analysis.

- Jupyter notebook with Python

  The binary logistic regression was executed in Python with the Scikit-learn machine learning library (https://scikit-learn.org) and the module for statistical models, StatsModels. The script was coded in the open-source web application Jupyter Notebook (https://jupyter.org).

## 3.2 Conclusions

In this chapter, the datasets were described as well as the pre-processing and geo-processing performed. Each stage of the analysis requires a different data transformation process according to the set objectives.

For the exploratory modelling, it is required to work with a small analysis unit that enables to capture the spatial heterogeneity of the perceived unsafe and safe areas. As each sketched polygons vary in shape and size, a single data sample for each polygon would not allow performing a detailed extraction of attributes.

In the case of the spatial delineation of the perception accuracy, the used datasets must be in the same spatial analysis unit as is a comparative analysis. Thus there has to be a data consistency between the layers of information. Working with city blocks is a common practice in crime studies. For the aim of objective two, the data aggregation of the crime events in blocks is more significant than working with single points. Besides, the nature of the chosen type of analysis is based on the adjacency between spatial units. Therefore it is not possible to consider the crime events as independent observations.

I am aware of the pitfalls of data aggregation and data segmentation in spatial analysis. That is why some data testing was done to assure the quality of the transformations performed. The selection of the data transformation methods were picked after comparing the results of the analysis performed with different datasets. The chosen methods were more efficient or accurate for the aim of each objective.

The described data will be the input for the analysis performed in this research. In the three forthcoming Chapters, the three specific objectives will be addressed. The following Chapter explains the exploratory modelling in which different variables are extracted from the structured sketch maps and explore by a regression method.

# Chapter 4

## Crime perception: Exploratory modelling

In this chapter the first specific objective:

*To analyse the location of perceived unsafe areas in relation to a) the distribution of crime incidents and b) people's activity spaces.*

This chapter will address the first specific research objective. It is organized in three main sections. The first one is the methodology followed (4.1), in which the **binary logistic regression** is explained (subsection 4.1.1) as the selected method to explore the factors (covariates) that have an impact on the perception of crime. As in any regression method the **covariates** or independent variables have to be defined. Four covariates were chosen and their values were extracted from the sketch maps. The methods used for the data extraction are described afterwards (subsection 4.1.2).

The second section (4.2) covers the **Exploratory Data Analysis** (EDA) (subsection 4.2.1) performed with the values of the covariates. Based on the EDA, four hypotheses linked to each covariate were defined and tested with the results of the regression. The analysis is supported with tables, graphs and maps which ease the explanation of the variables. The exploration of the **output of the model** and its interpretation are presented in the second subsection (4.2.2) of this section.

The **conclusions** with respect to research objective one are addressed in the third section (4.3).

## 4.1 Methodology

The first research objective covers the explanatory modelling stage. Its aim is to examine the impact that people's activities areas (daily routes and neighbourhood) and the crime hotspots have on crime perception. Even though some studies have concluded that perception of safeness is not related to the criminal incidents, one of the specific objectives of this research is to explore these variables from a spatial focus. Therefore, four related variables (the distance from the sketched polygon to a) people's neighbourhood, b) daily activities routes, c) high crime intensity areas and d) crime hot spots) were examined to find out their possible relation with the location of the areas that people perceived as unsafe or safe.

The goal is to explore the impact of these four variables on the classification of the space into perceived safe/unsafe areas. One of the possible approaches to address it is by using a supervised classification machine learning method.

Machine learning is a discipline related to Computer Science, which focuses on developing systems that can learn from the data and subsequently use this knowledge for future related tasks. The machine learning methods are divided into two major types of learning: supervised and unsupervised. In supervised learning, the input ($X$) and output ($Y$) is known, the aim is to define a function $f(X) = Y$ that relates both. The process of defining this function is done with a set of training labelled data, which means, that the output ($Y$) is known. The resultant model is used to calculate $Y$ for forthcoming unlabelled data where only the input is known. Supervised learning methods are divided into two groups: classification and regression. The classification methods group objects or features into categorical classes, based on their characteristics (independent variables); meanwhile, the regression methods predict a numerical continuous variable.

The unsupervised learning is used when the classes are not defined, the output ($Y$) is unknown but the input ($X$) is known, so the aim is to group the sample data according to its characteristics, thus there is no training data needed. Which means, the learning process is done with the given data. Unsupervised learning methods are classified into two types: clustering and association. The clustering methods group objects by similarity in attributes or location and the association are rule-based methods.

The aim of this research is to find the relation between the location of the a) people's activities and b) the crime hotspots, and the location of the perceived unsafe and safe areas. Because the input and output are known, a supervised method was selected. The target ($Y$) variable is the centroid of the cell with the binary class label of "safe" or "unsafe"; thus, it is a classification problem. The input data ($X$) are calculations derived from spatial analysis between the target and the four variables mentioned earlier.

There are different classification algorithms in machine learning, such as logistic regression, nearest neighbour, support vector machines, decision trees, random forest, and neural networks. The logistic regression method differentiates from the rest, as the output is not only the resultant class but also an expression of the relationship between the independent variable(s) and the output class. This method performs a classification based on a regression. It defines a classification function $f$ that sorts an object into one of the two given classes $Y$, $f(X) = Y$ (Mello & Ponti, 2018), as $Y$ has to be a dichotomous class. The process consists of evaluating the impact of a set of characteristics of the objects or events $X_n$ on the probability of classifying them into one or other defined class.

Hence this method is suitable to tackle the first objective of the research, as the problem deals with binary classification (safe and unsafe) and the coefficients of the regression indicate the relationship of the explanatory variables and the dependent variable.

Like in any regression method, there must be a set of independent variables (characteristics) to be evaluated and a dependent variable, in this case, a binary one. In the next sections, I explain the chosen method (see 4.1.1) and the selection and extraction of the independent variables and their values (see 4.1.2).

### 4.1.1 Binary Logistic Regression

The binary logistic regression is used when the number of output classes is reduced to two. This method defines the relationship between a dichotomous nominal variable and one or more independent variables, which can be nominal, ordinal or interval. It has become one of the most frequently used inference methods in crime research (Weisburd & Britt, 2007), especially to identify and compare the effects of the extensive number of factors that influence criminal activity (Weisburd & Piquero, 2008). This method was chosen to define the impact that some spatial variables have in perceiving an area as unsafe or safe.

Contrary to the linear regression in which $Y$ must be a continuous value, in logistic regression instead of predicting $Y$, the predicted value is "*the natural logarithm (ln) of the odds of getting 1 on the dependent variable*" (Weisburd & Britt, 2007). Getting 1 on the dependent variable means classifying a feature or object in one of the two defined classes.

An odd is the relative rate between the probability of an event to occur (success, $Y = 1$), related to the probability of not occurring (not success, $Y = 0$). The range of the odd value goes from 0 to +∞ as probability $P$ varies in values closer to 0 and 1. As shown in Figure 4.1, working with odds would mean finding a relation between variables using a non-linear function, which makes it more complex. Thus, the odds must be transformed into values between -∞ to +∞, similarly to the linear regression where the $Y$ axis can have any number. In order to get this result, it is necessary to get the natural logarithm of the odd; this is called the Logit of Y or the Logit function [1].

$$\ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = b_0 + b_1 X_1 \qquad [1]$$

Now the values of the dependent variable $Y$ go from -∞ to +∞, and it establishes an equality between the Logit function and the linear equation.


**Figure 4.1** Graphic representation of the regression functions.

In order to transform these continuous values into a probability between 0 and 1, it is needed to calculate the cumulative logistic probability function [2]:

$$P\ (Y = 1) = \frac{1}{1+e^{-b_0+b_1X_1}} \qquad [2]$$

The dependent variable is defined as the probability of classifying an object into a target class ($Y = 1$), thus, this method constraints the dependent variable to be in the range from 0 to 1, as it is a probabilistic value.

While for the linear methods the adjustment of the regression line is done by least squares, for the logistic regression the Maximum Likelihood Estimation (MLE) is performed to obtain the coefficients $b_n$. This method evaluates a group of coefficients and selects the parameters that have the highest probability of being the ones that could have generated the observed data. The probability of the coefficients or the model to have generated an observed data is known as the likelihood $(L)$. The probabilities goes from 0 to 1, but the likelihood values are so small that they are calculated by a natural logarithms. However, because the logarithm of a number smaller than 1 is negative, the likelihood is then calculated by $-2Ln(L)$. The likelihood of the model is obtained by the ratio between the likelihood of the saturated model (the model with all the variables) and the base model, which only considers the constant $b_0$.

The coefficients are interpreted by their exponent $e^b$. Contrary to linear regression the effect of $b_n$ is not a constant.

### 4.1.2. Covariates

As in any other regression, binary logistic regression requires to be defined the dependent and the explanatory variables aka covariates. For the aim of this research, the dependent variable of analysis is whether the areas are perceived as unsafe or safe; meanwhile, the explanatory variables are different measurements extracted from the participatory data by spatial queries. The selected covariates were chosen based on the available datasets and the theories that try to explain fear of crime and crime perception.

The data sets consist of four vector files: a) the sketched polygons of the perceived safe areas, b) the sketched polygons of the perceived unsafe areas, c) the participants' daily routes and d) the crime incidents. The sketch polygons were divided into small cells to work with the minimal spatial unit. Each cell centroid represents a data sample, from which four distance-based measurements were calculated from the centroid of each cell to: a) the participant's neighbourhood (postal code area), b) his daily route, c) a crime hotspot and d) high crime intensity areas. These four measurements are the chosen independent variables, which are described hereunder:

A. <u>Neighbourhood</u>

The purpose of this variable is to explore whether the people tend to perceive their own neighbourhood and the surrounding area as safe or unsafe. The participants' neighbourhood was defined by the area of the postal code that each of them reported in the online survey. The postal code areas were download from OpenStreetMap.

To explore this relation, it was measured the distance between the centroid of the cells within the sketched polygon and the nearest point of the respective participant's postal code area. Thus, if the distance is zero, this would mean that the target centroid is located inside the participant's neighbourhood.

B. <u>Daily route</u>

This variable describes if people follow "safe routes" traced in their cognitive maps to avoid high crime perceived areas. Based on the "Geometry of crime" and "Crime Pattern theory", diverted by Spicer, Song and Brantingham (2014), people would design daily routes through which they can stay off situations and places where they perceive as unsafe.

Each cell's centroid within a sketch polygon was linked with the route or routes drawn by the same participant who draw that polygon. The minimum distance between the cell and the lines(s) was measured. From all the relations, the shortest distance was selected.

Table 4.1 shows an example of the query result: the cell with the id 71,410 belongs to the polygon 4,258, sketched by the participant 230. This participant draw four daily routes (87, 88, 89 and 90). The minimum distance between the centroid of the cell and the four lines was calculated and then the closest line was selected. In this example, the closest line was number 90 with a distance of 317.83 meters.

Figure 4.2 shows the graphic representation of this example; the grey points are all the cells' centroids within the polygon 4,258, the big black point is the centroid of cell 72,410. The coloured lines are the routes participant 230 traced while the dotted black lines are the distances measured to the closest point of each line, which are shown with the black crosses.

The corresponding shortest distance between the centroid and the daily route line was the second independent variable.

| Cell ID | User ID | Type | Polygon ID | Line ID | Distance |
|---|---|---|---|---|---|
| 72 410 | 230 | unsafe | 4 258 | 87 | 422.81 |
| 72 410 | 230 | unsafe | 4 258 | 88 | 439.88 |
| 72 410 | 230 | unsafe | 4 258 | 89 | 654.12 |
| 72 410 | 230 | unsafe | 4 258 | 90 | 317.83 |
| | | | | | |
| 72 410 | 230 | unsafe | 4 258 | 90 | 317.83 |

**Table 4.1** Extract of a spatial query result to select the closest daily route from the centroid of each grid cell.

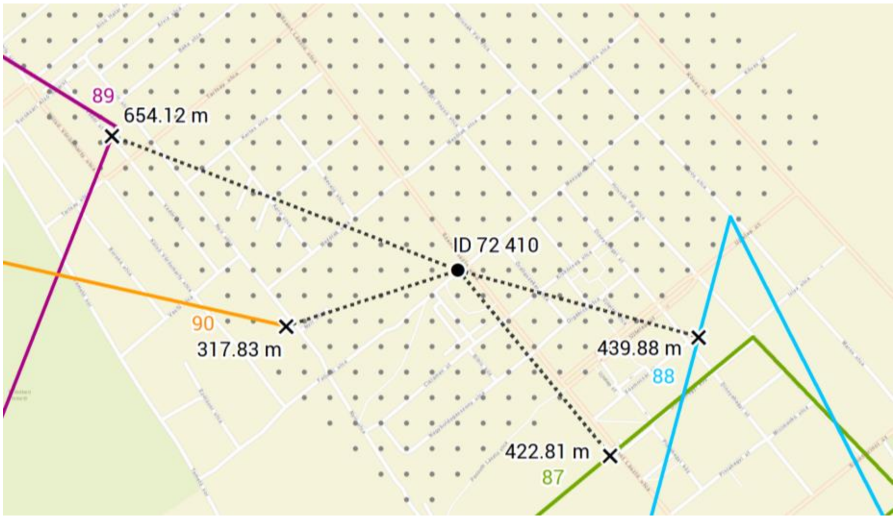



**Figure 4.2** Distances measured (dotted lines) from the centroid of the cell 72,410 (black point) to the nearest point of the correspondent sketched daily routes (solid lines).

C. Hotspots,

The aim of this covariate is to explore how the location of crime hotspots is related to the location of the areas people perceive as unsafe. Some researchers have concluded that the actual level of crime is slightly connected to safeness perception, as there are other factors that could have more impact on this kind of perception. One of the objectives of this research is to explore the spatial relations of the location of actual hotspots and perceived unsafe/safe areas by measuring the distance between these two places.

A hotspot should not only be defined by a high amount of events happening in a location but also when the “local structure is sufficiently unusual” (Ord & Getis, 1995). The local spatial autocorrelation statistics indicate where unexpected values are located in comparison with a random distribution. There are three main local measures of spatial autocorrelation: local Moran’s I, local Geary, Getis-Ord statistics (Table 4.2).

| Spatial autocorrelation methods | Logic | Comparison | Clusters | Outliers | Type of association |
|---|---|---|---|---|---|
| Local Moran's I | cross product | slope | ✓ | ✓ | ✓ |
| Local Geary | square distance | distance | ✓ | | |
| Getis-Ord statistics | point pattern | ratio | ✓ | | ✓ |

**Table 4.2** Comparison of spatial autocorrelation methods.

The local Moran's Index is a correlation measurement estimated for each data observation. The advantage of using Moran's I is that it estimates the four types of local association: two spatial clusters and two spatial outliers. The clusters are identified when the attribute value of the observed feature is significantly similar (positive autocorrelation), high or low from the mean (high-high, low-low), as its neighbours. The spatial outliers are those features which attribute values is significantly different (negative autocorrelation) from the mean than its neighbours (high-low, low-high). The high-high, low-low, high-low, low-high indicate the kind of cluster of outlier, the first attribute refers to the target feature, and the second attribute to the value of the neighbouring features. For instance, the high-low class means that the target feature have a higher value than the expected mean, while its neighbours have a lover value than the expected mean.

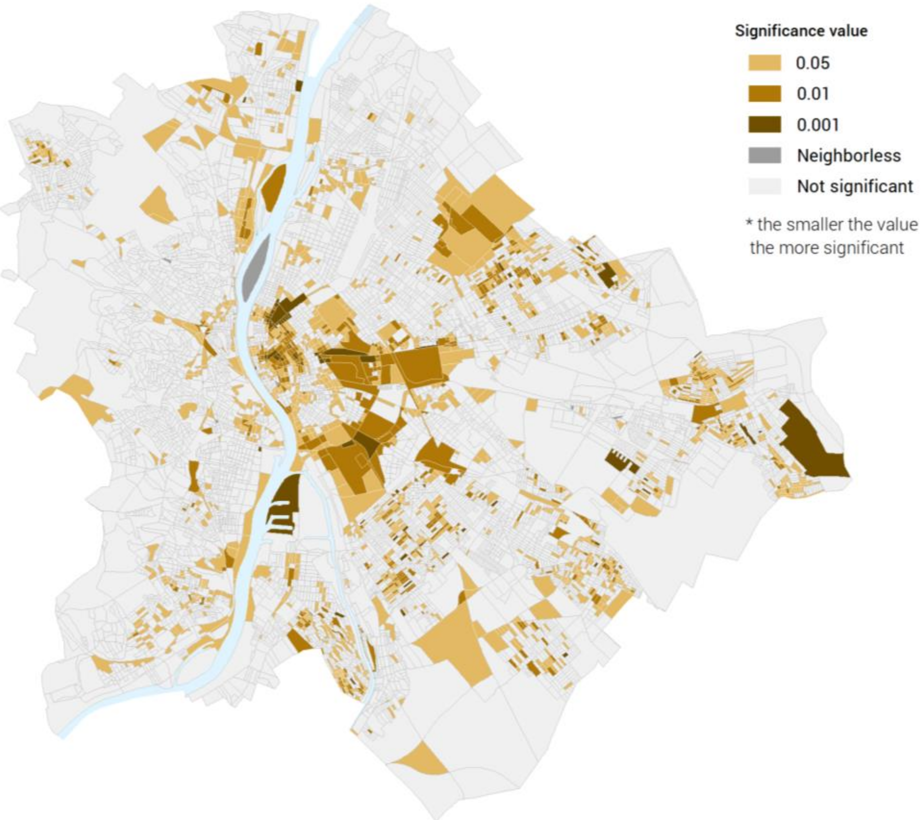
The Geary's statistic measures the square difference in relation to the mean between the value of the target feature and the one of its neighbours, it is a distance in attribute space. If the result is a large square difference indicates negative spatial autocorrelation or dissimilarity, meanwhile small square differences means positive spatial autocorrelation or similarity. The Geary's statistic only shows if the association is positive or negative but it does not indicate whether the relation is high-high or low-low in the case of similarity, or high-low or low-high for dissimilarity as is a square difference and the sign is lost.

The Getis-Ord statistic is based on a point pattern logic. This statistic counts the features' value within an area and compares it, as in a ratio, with the addition of all the features in the dataset. If this ratio is higher than it would be on the randomness, it is identified as a cluster of high values, if it is lower is a cluster of low values. The advantage is that the interpretation is really simple as the hotspots and coldspots are given by positive and negative G-statistic. The disadvantage is that this method does not detect spatial outliers.

In crime studies, the detection of hotspots must consider both clusters and spatial outliers, as both show the location of unusual values in comparison with the ones that a random distribution would have. Thus, the local Moran's I was selected to perform the hotspot analysis. This was executed in the Geoda software as the outputs of the analysis are a Moran's I scatter plot, a significance map and a cluster map that is the combination of the Moran's I scatter plot and the significance map. The cluster map show the four types of local association.

Figure 4.3 and 4.4 depict the results of the local spatial autocorrelation analysis performed with the Univariate Local Moran's I method with the total count of crimes per block (Figure 4.3) and with the crime density value per block (Figure 4.4).

A)




B)

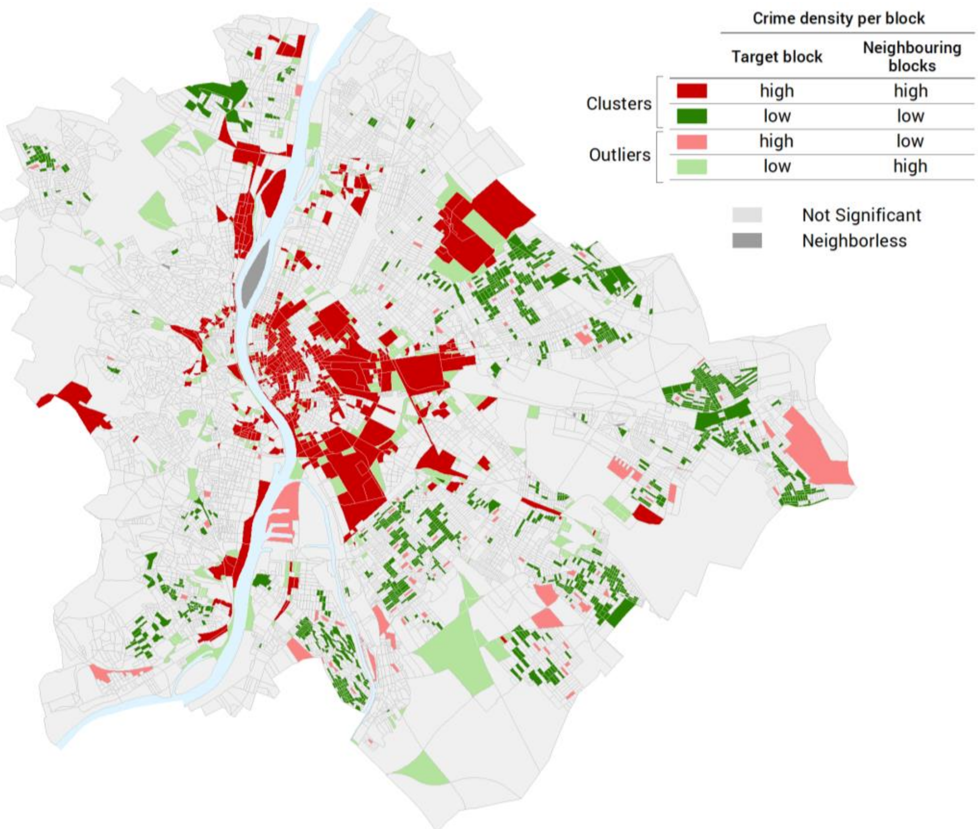
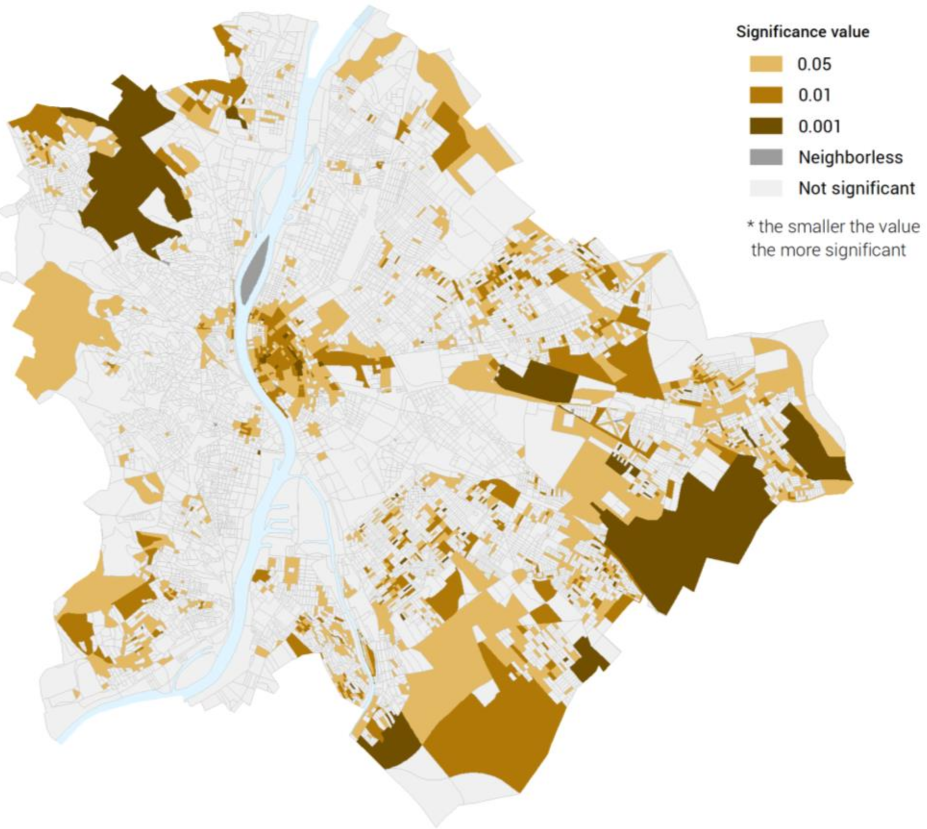



Figure 4.3 Local spatial autocorrelation analysis (Moran's I) performed in GeoDA with the total count of crime incidents per block. A) Significant local statistics per block. B) Spatial association per block (clusters and outliers).

A)




B)

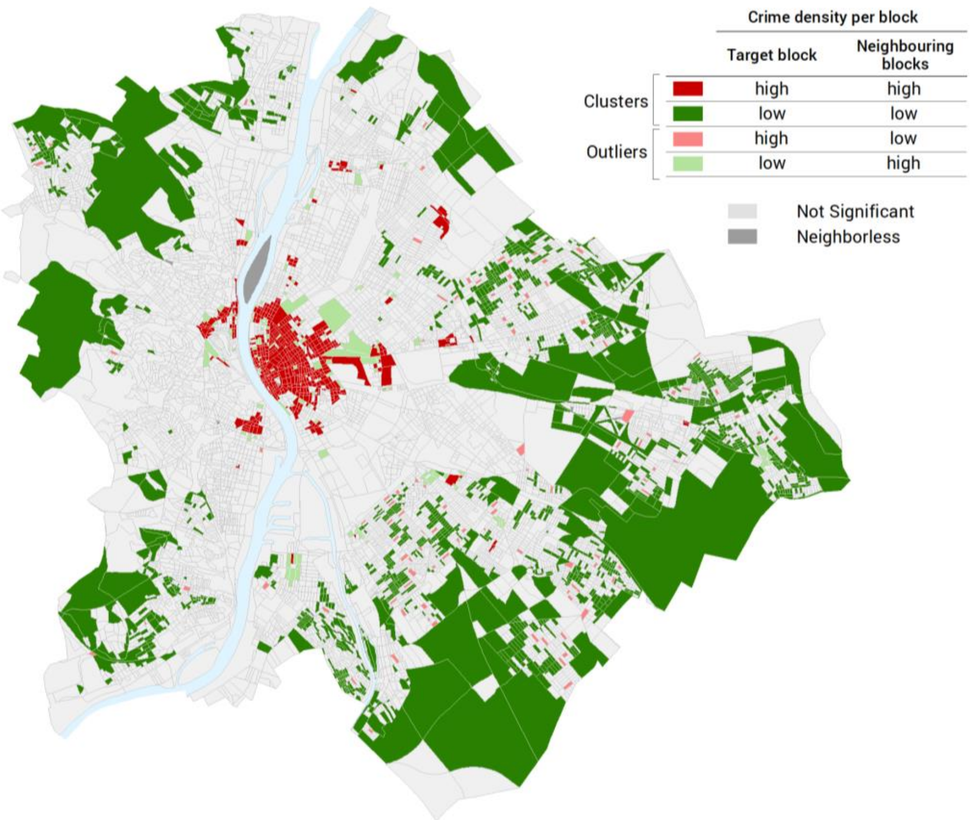



Figure 4.4 Local spatial autocorrelation analysis (Moran's I) performed in GeoDA with the density value of crime incidents per 100 m2 per block. A) Significant local statistics per block. B) Spatial association per block (clusters and outliers).

The Moran's I statistic was obtained with 999 permutations and a queen contiguity neighborhood relation, which means that the blocks evaluated were the target block in relation with the surrounding blocks with common sides and vertices. The blocks without surrounding blocks are labeled as "neighborless".

The output significance map (Figures 4.3. A and 4.4 A) shows the significance values per block given by the pseudo p-values. The maps show the blocks wich values that are not significant and those which are. The latter are divided into classes according to the the significance pseudo p-values which are related to how extreme the observed value was in comparison with the referenced values (conditional permutation). The lower the p-value, the more extreme the observation is and thus the most significant. That is why in the maps (Figures 4.3. A and 4.4 A) the smaller p-values are shown in a darker tint as they are more significant.

The cluster map (Figures 4.3. B and 4.4 B) shows the location of the clusters (hotspots and coldspots) and the outliers.

In order to define the third covariate for the binary logistic regression that is the distance between the cells' centroid within each sketch polygon and the closest hotspot, there were considered as hotspots the "high-high" clusters and the "high-low" outliers. The selected features with a pseudo p-value equal or smaller than 0.05 were selected, as the features with p-value=0.001 mostly correspond with the core block of each cluster, meanwhile, those with 0.05 and 0.01 p-value approximately correspond with the neighbours of the cores.

D. <u>High crime intensity area</u>

For the fourth covariate the distance between each cells' centroids and the closest high crime intensity area was measured. The high crime intensity areas were identified by the total count of crime per block. In order to select the range of values that would define the most intense areas. The four blocks with the highest total count of incidents that are within the range of 304 to 614. This covariate will allowed us to explore whether the people is aware of this most significant hotspots or their crime perception is more focused on the other hotspots, perhaps those which are located closer to their neighbourhood or daily routes.

## 4.2 Results

The analysis was performed with the data obtained from a participatory online survey where the participants identified the areas that they perceive as unsafe and safe in Budapest. The resultant polygons were segmented into cells of 45x45 meters, each cell's centroid was considered a data sample. Four Euclidean distances where measured from each centroid: a) to the participant's neighbourhood, b) to the participant's daily route, c) to the closest crime hotspot and d) to the closest high crime intensity area. These values were analysed as covariates with a binary logistic regression to get the impact of each distance on the location of the areas that the participants perceive as safe and unsafe.

In the next two sections the results of the performed analysis are described. The first subsection (4.2.1) comprises the result of a brief Exploratory Data Analysis (EDA) performed with the results of the four explored covariates in order to better understand the resultant regression model. The interpretation of the output coefficients will be presented in subsection 4.2.2.

### 4.2.1. Exploratory data analysis (EDA)

The first tested variable was the distance between the perceived unsafe/safe area and the participants' neighbourhood, in order to explore how the people tend to perceive their own neighbourhood and its surrounding area. Figure 4.5 shows two cumulative relative frequency plots with the distance between each sketched polygon (centroids) and the respective participant's neighborhood (nearest point). The light blue line depicts the distances to safe areas, and the dark blue line to unsafe areas. The points were grouped in classes of every 200 meters.

From the 231 unsafe sketched polygons, 41 (17.7%) have at least 50% of their area within the respective participant's neighborhood, meanwhile, from the 97 safe areas, 13 (13.4%) presented this characteristic. Thus, some participants defined some areas of their own neighborhood as not safe and others as safe.

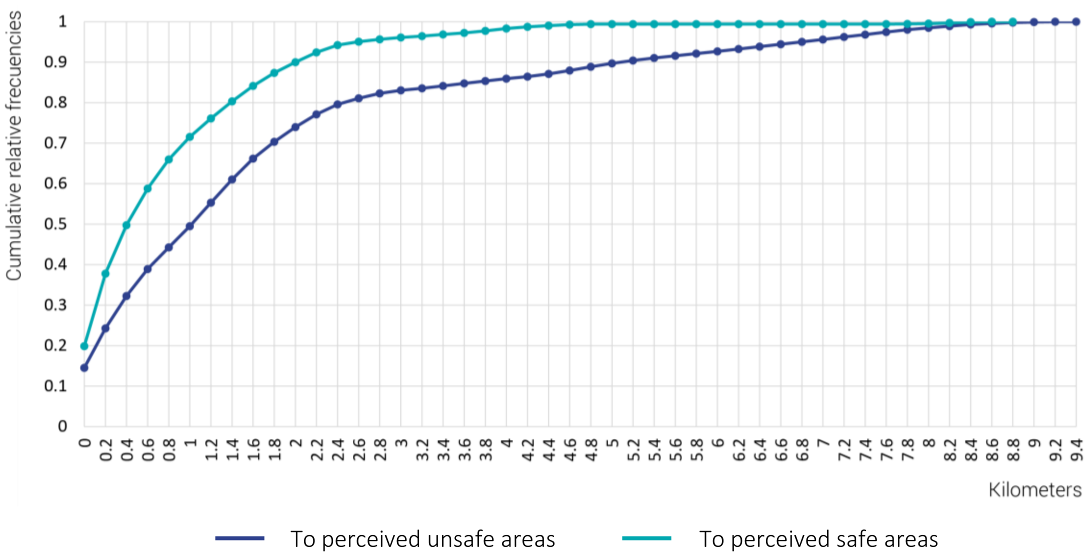



**Figure 4.5** Cumulative relative frequencies of the measured distances (km) from each cell's centroid within a participant's sketch polygon and the nearest point to the participant's neighbourhood.

Without considering the points that were within the neighborhood area (zero kilometers), the minimum distance between a participant's neighborhood and a perceived unsafe area, was less than one meter and the maximum 9.3 km. Meanwhile, the shortest distance to a perceived safe area was 5 m and the longest 8.7 km.

50% of the centroids within an identified unsafe areas are less than 1 km away from the participants' neighborhood, meanwhile, half of the centroids of the safe areas are less than 400 m away. In general, the participants identified safe areas closer to their neighborhood.

The second chosen variable was the minimum distance between the perceived unsafe/safe area and the participants' daily route, the aim of it is to identify the impact that the crime perception has on the daily navigation of people. The *Geometry of Crime* and *Crime Pattern* theories explain how people create a "safety spatial templates" in their cognitive maps to locate awareness space where the crime rate is perceived as high, therefore people would try to avoid it (Spicer et al., 2014).

Figure 4.6 shows a cumulative relative frequency graph with the distances from participants' daily routes to the location of their sketched polygons. It can be observed that around 38% of centroids within sketched safe area and around 13% centroids within unsafe areas were located less than 200 m away of the corresponding daily route. The graph shows that more participants identified unsafe areas further away from their daily routes.

The minimum distances to safe areas goes from less than one meter to 4.4 km, while to unsafe areas the longest minimum distance was 7.4 km. Participants identifies unsafe areas further away from their daily routes, which could be related to the 'safety spatial templates' theory. 50% of the participants sketched safe areas in a distance no longer than 200 m, meanwhile, the unsafe areas were identified in a distance up to 1.2 km by 50% of the participants.

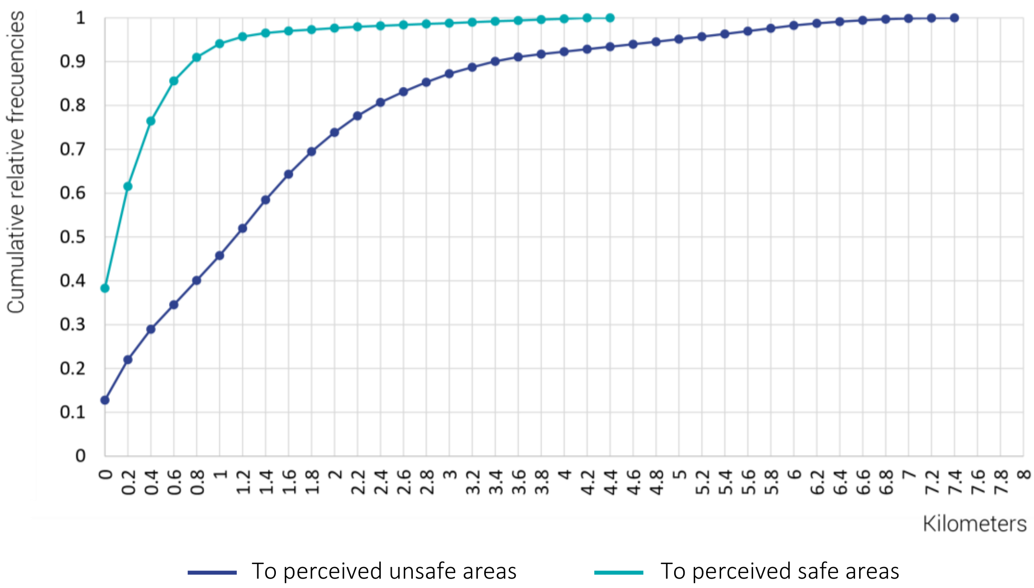



**Figure 4.6** Cumulative relative frequencies of the measured distances (km) from each cell's centroid within a participant's sketch polygon and the nearest point to the participant's daily route(s).

The third variable that was explored is the distance from a crime hotspot to a perceived safe/unsafe area. The hotspots were identified based on the Local Moran's Index of each block, the blocks classified as cluster or outliers with a pseudo p-value equal or smaller than 0.5 were considered. The distance was measured between each cell's centroid and the nearest point of the closest block labeled as a hotspot.

The types of hotspots were analyzed, the first one based on the density values (number of crimes per 100m$^2$) and the second considering the total count of incidents per block.

Figure 4.7 shows the cumulative relative frequencies of the distances measured in kilometers from each cell's centroid within a safe and unsafe sketch polygon and the closest block identified as hotspots defined by the total count of incidents. The minimum distance from the identified safe areas ranges from less than 0.02 m up to 998 m, meanwhile, from the unsafe areas, they range from less than 0.02 m to 2.73 km.

For 122 (52.8%) of the sketched unsafe areas at least 50% of their area is within a crime hotspot, and for 40 (17.3%) out of them, the entire polygon is contained by a hotspot block. For 62 (63.9%) of the sketched safe areas, at least 50% of their area overlaps with crime hotspots and 10 (10.3%) are completely within a hotspot. This finding is a first indication of the crime perception gap that is further modelled and quantified in the next sections.

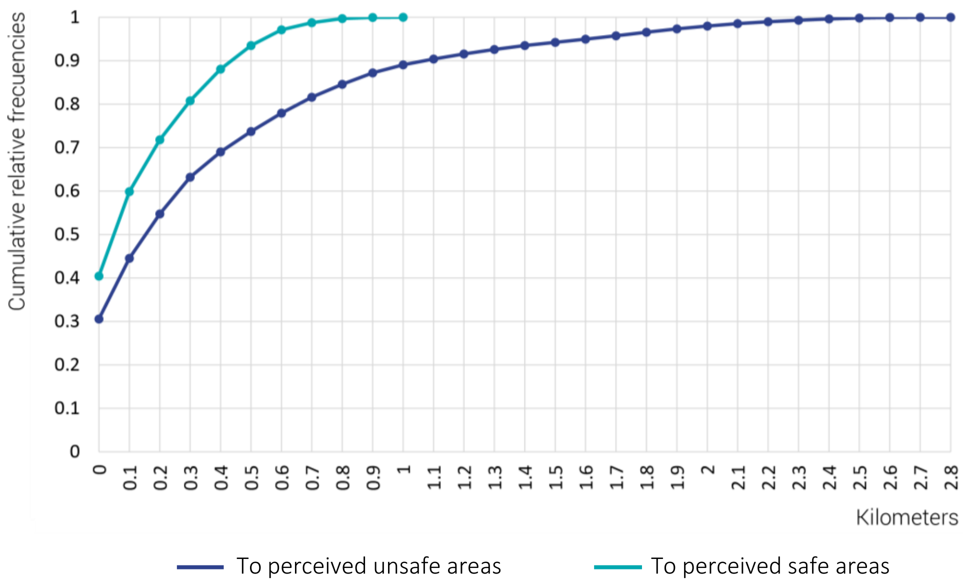



**Figure 4.7** Cumulative relative frequencies of the distances measured from each cell's centroid within a sketch polygon and the closest block identified as a hotspots defined by the total count of incidents.

Figure 4.7 is comparable to Figure 4.8, but Figure 4.8 shows the distance to closes hotspot defined by the crime density per block, it shows that the values of minimum distances measured to a perceived safe area range from 0.05 m to 2.13 km, and for unsafe areas from 0.14 m to almost 2.82 km.

For 77 (33.3%) of the total perceived unsafe areas more than 50% of the area is located inside a hotspot, and for 44 (19%) the area is completely contained by a hotspot. From the sketched safe areas, 68 (70%) of them have at least 50% of their surface laying inside a hotspot and 18 (18.5%) are completely within one.

The variation of distance could be explained by the difference of the spatial distribution of the two types of hotspots identified. This can be seen by comparing the maps shown in Figure 4.3 B and 4.4 B: the hotspots by total count are more scattered than the hotspots by crime density, which are gathered in the center of the city. Thus, the distances are shorter to the nearest hotspots by total count as they are located in distributed parts of the city.

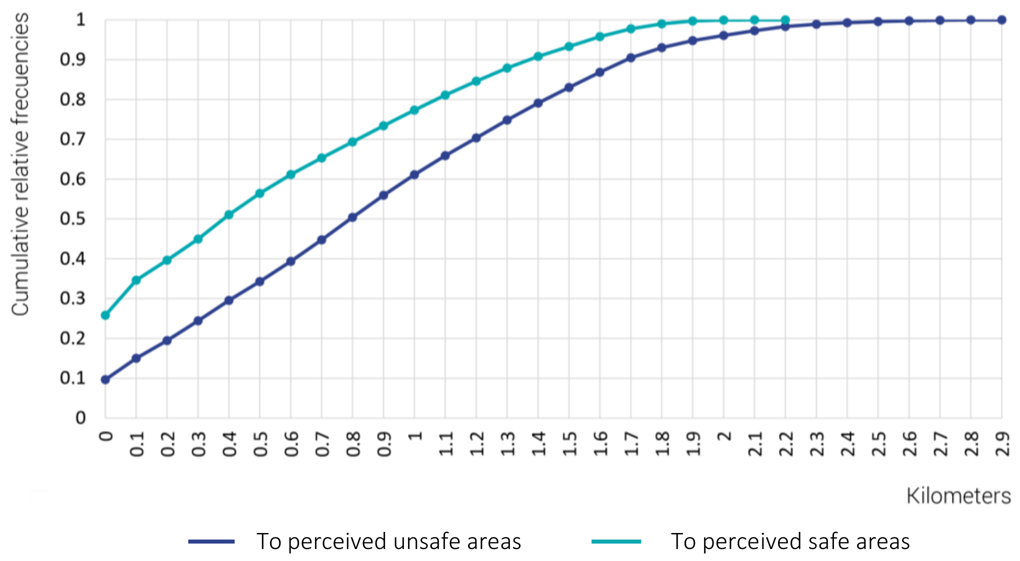



**Figure 4.8** Cumulative relative frequencies of the distances measured from each cell's centroid within a sketch polygon and the closest block identified as a hotspots defined by the density values per block.

The fourth tested variable was the distance from the cell's centroid within the sketched polygons and the nearest high crime intensity areas (HCIAs), defined by the four blocks with the highest number of incidents, that are located either in the city centre or in an adjacent urban residential area. Figure 4.9 shows the distance from the centroids within an unsafe and safe perceived area to the closest HCIA.

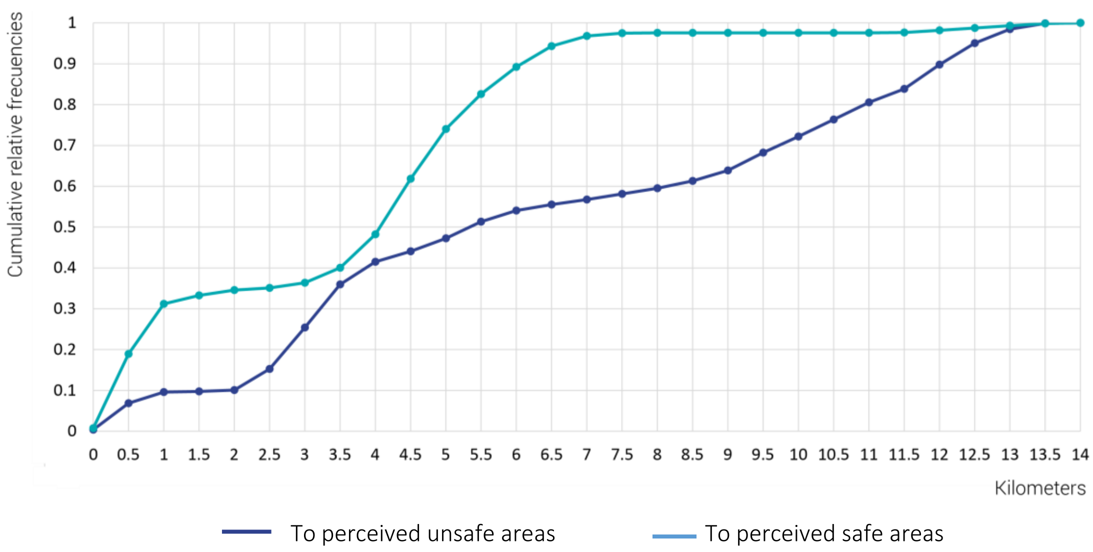



**Figure 4.9** Cumulative relative frequencies of the distances measured from each cell's centroid within a sketch polygon and the closest block identified as a high crime intensity area (HCIA).

The distances from perceived safe areas were in a range of 0.28 m and the maximum of 13.7 km. Meanwhile from perceived unsafe areas the distances went from 3.2 m to 13.6 km. Figure 4.9 show that a higher percentage of the centroids within safe polygons were located closer to a HCIA than centroids within unsafe perceived areas. For instance, 30% of the centroids from safe polygons were located less than one kilometre away from the closest HCIA, and only 9% of the centroids from unsafe polygons.

This means that people perceived a higher percentage of safe areas around the HCIAs. The could be explain by the high percentage of participants that live in the surrounding areas where the HCIAs are located, and according to Figure 4.5 participants In general identified safe areas closer to their neighbourhood.

The preliminary conclusions, based on the EDA of each covariate, are:

- The participants identified safe areas near to their neighbourhood.
- The participants identified unsafe areas further away from their daily routes.
- Participants perceived safe areas near to the hotspots.
- More participants identified safe areas closer to the HCIAs than unsafe areas.

Based on these results, there is evidence of an inaccurate perception of crime among the participants that led in the formulation of two hypotheses.

> Hypothesis 1: *the likelihood of people perceiving an area as unsafe increases when the target area is far away from their neighbourhood and their daily routes*.
> This hypothesis is tested in subsection 4.2.2
>
> Hypothesis 2: *People's misconception of crime reality involves both the overestimation of safe areas (inaccurate perception of safe areas- IU) and the underestimation of unsafe areas (inaccurate perception of safe areas -IS).*
> This hypothesis is tested in Chapter 5.

### 4.2.2. Modelling

The preliminary conclusions presented above are based on the visual data exploration of the chosen covariates. The next step was to perform a binary regression analysis to define the quantitative impact of the four covariates on people's safeness perception. The 68,032 samples (cells' centroids within the sketched polygons) of the dataset were divided 80% for training data and 20% for testing.

The personal data provided by the participants (age, sex and main mean of transportation) was explored in the binary regression analysis, but none of these three variables have a significant impact on crime perception. The participants' mean of transportation had a p-value higher than 0.05 which means they were not significant, meanwhile, the age and the sex only explained around 2% of the variability in likelihood of perceiving an area as unsafe. As these covariates did not explain much of the differences of the likelihood and as the aim is to explore the spatial variables, age and sex were not considered for the final model.

While performing the analysis and choosing the most suitable covariates for the model, the results attest that measuring the distance between the sketched polygons to the block classified as a cluster hotspot and to those classified as an outlier hotspot increased the $R^2$ of the regression than grouping both clusters and outliers as hotspots and measure the distance to the closest one. To consider two different covariates rather than one, explained a higher percentage of the variability in $Y$, which at the same time increased the accuracy values of the final model. Thus, the five covariates were evaluated, Table 4.3 shows the results of the logistic regression.

| Covariate | Coefficients | Standard error | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **Distance to:** | | | | | | |
| Neighbourhood | 0.0002 | 8.28 e-05 | 29.284 | 0.000 | 0.0002 | 0.0003 |
| Daily route | 0.0010 | 1.93 e-05 | 53.569 | 0.000 | 0.001 | 0.0011 |
| Cluster hotspot | 0.0007 | 1.93 e-05 | 35.643 | 0.000 | 0.0006 | 0.0007 |
| Outlier hotspot | 0.0009 | 1.79e-05 | 52.384 | 0.000 | 0.0009 | 0.0010 |
| High crime intensity areas | 0.0001 | 6.55 e-06 | 12.625 | 0.000 | 0.0001 | 0.0001 |

Constant -3.1573 Pseudo $R^2$ 0.312 Log-Likelihood -25,792 Log-Null -37,466

**Table 4.3** Result of the binary logistic regression.

The resultant regression model is represented in equation 3.

$$Xb = -3.1573 + 0.0002X_1 + 0.001X_2 + 0.0007X_3 + 0.0009X_4 + 0.0002X_5 \qquad [3]$$

The interpretation of the coefficients differs from a linear regression method, in multivariate binary logistic regression the coefficients represent the "*estimated change in the logarithm of the odds of* $Y = 1$ *occurring when all other independent variables are held as constant*" (Weisburd & Britt, 2007). For this model $Y = 1$ means classifying an area as unsafe, therefore the coefficients are interpreted over this variable.

The p-values indicate that the five variables are related to the classification of unsafe areas. The resulting coefficients are explained in terms of their odds ration which is usually expressed by the exponent of $b, e^b$ (Table 4.4), "*the odds ratio represents the impact of a one-unit change in* $X$ *on the ratio of the probability of an event occurring to the probability of the event not occurring*" (Weisburd & Britt, 2007). When the odds ratio is greater than 1 it means that the odds of getting $Y = 1$ increases when the $X$ increases, with values less than 1 the odds of getting $Y = 1$ decreases when $X$ decreases.

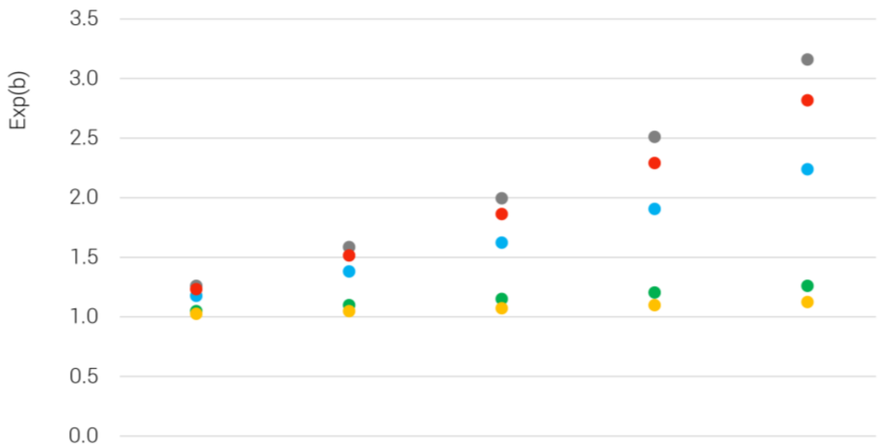
| Explanatory variable | Coefficient | $e^b$ |
|---|---|---|
| **Distance to:** | | |
| Neighbourhood | 0.0002 | 1.0002 |
| Daily route | 0.0010 | 1.0010 |
| Cluster hotspot | 0.0007 | 1.0007 |
| Outlier hotspot | 0.0009 | 1.0009 |
| High crime intensity areas (HCIA) | 0.0001 | 1.0001 |

**Table 4.4** Exponents $e^b$ of the resulting coefficients.

Due to the fact that the odds ratio is not a linear function of the coefficients, it is necessary to estimate the coefficients with the specific number of units $X$ and then get the exponential of the coefficient. In this case, the covariates $X$ were estimated in meters (distances measured) and, therefore, the likelihood is given in reference to one meter distance.

In order to make the interpretation of the resultant coefficients more significant, the independent variable (i.e. the distance in meters) would be transformed into the number of blocks. Thus, the likelihood of perceiving an unsafe area would be given according to the number of blocks that we move far away from the target locations (in this case, the participant's neighbourhood and daily route, a crime hotspot and the HCIA.)

Considering the average block size in Budapest (Atlas of Urban Expansion, 2016) of 5.3 ha or 0.053 $km^2$. it can be estimated that the average length of a block is 230 meters. Figure 4.10 shows how the likelihood of perceiving an area as unsafe changes while moving away from the people's neighbourhood, daily route, a crime hot spot and high crime intensity areas.




| | 1 block (230 m) | 2 blocks (460 m) | 3 blocks (690 m) | 4 blocks (920 m) | 5 blocks (1150 m) |
|---|---|---|---|---|---|
| ● Neighbourhood | 1.05 | 1.10 | 1.15 | 1.20 | 1.26 |
| ● Daily route | 1.26 | 1.58 | 1.99 | 2.51 | 3.16 |
| ● Cluster hotspot | 1.17 | 1.38 | 1.62 | 1.90 | 2.24 |
| ● Outlier hotspot | 1.23 | 1.51 | 1.86 | 2.29 | 2.82 |
| ● High crime intensity areas | 1.02 | 1.05 | 1.07 | 1.10 | 1.12 |

**Figure 4.10** Exponents of the coefficients $e^b$ obtained for different ranges of distances or number of blocks from the target locations.

The results of the interpretation of the coefficients show that the likelihood of perceiving an unsafe area increases when moving away from the five selected referenced locations. As shown in Figure 4.10, the increment of the likelihood value is not linear. For the covariates daily route, cluster hotspot and outlier hotspot the gradient changes faster than for neighbourhood and high crime intensity areas.

The likelihood of perceiving an unsafe area highly increases while moving away from the people's daily route. This result can be linked to Figure 4.6 in which it is visible how people defined unsafe areas at longer distances than safe areas. Also the likelihood increases with increasing distances to peoples' neighbourhoods. This result concurs with Figure 4.5 in which it is visible how participants defined safe areas closer to their neighbourhoods. This confirms hypothesis 1, *the likelihood of people perceiving an area as unsafe increases when the target area is far away from their neighbourhood and their daily routes*.

In the case of the high crime intensity areas the variation of the likelihood over distance presents a smooth increase. The increment of the likelihood is consistent with the fact that people identified unsafe areas 14 km away from this zone of crime rates.

For both types of hotspots the likelihood increases more or less in the same proportion considering them as two separate variables explained a greater percentage of the variation of the likelihood than defining the model with a single hostpot variable combining both types. This result is consistent with the EDA. Based on the minimum distance between perceived unsafe and safe areas and a hotspot it indicated that the people perceived unsafe areas further away from a hotspot than from safe areas, Figure 4.7 shows the wide differences in distances from a safe and unsafe perceived area: this means people perceived safe areas closest to a hotspot.

The results from the regression model are consistent with the preliminary conclusion presented from the EDA. The results of the covariates neighbourhood and daily routes can be explained by the "geometry of crime" and "crime pattern" theories (see subsection 2.2). Meanwhile, the results of the covariates hotspots and high crime intensity areas are conclusive for inaccurate crime perception.

## 4.3 Conclusions

In this chapter, the methodology and results of dealing with the first specific objective which consisted in analysing the location of perceived unsafe areas in relation to a) the distribution of crime incidents and b) people's activity spaces were presented. The methodology was divided into two parts. The first one consisted of extracting spatial variables from the sketch maps and the second one involved the exploration of those variables with a binary logistic regression. The aim was to explore whether those variables are related to the perception of crime.

Subsection 4.2.1 focused on the EDA of the extracted variables, the conclusions for the four variables are:

- <u>The participants identified safe areas closer to their neighbourhood.</u> This can be explained by the "endowment effect" as described by Kahneman et al. (as cited in Lora, 2016). In the bases of the study of crime perception it can be applied as people tend to characterize their neighbourhood as safe, under the assumption that where people belong to, better conditions exist than in the surroundings. People tend to have a perceptual bias due to a feeling of attachment towards their own community or neighbourhood.

- <u>A higher percentage of participants identified unsafe areas further away from their daily routes.</u> This result endorses the "geometry of crime" and "crime pattern" theories which refer to awareness space and safety templates that people tend to trace in their cognitive maps in order to avoid situations and places that are perceived as having a high risk of victimization.

- Participants identified more safe areas than unsafe ones closer to the blocks categorised as hotspots (by total count of crime incidents per block and crime density per block). This confirms that the participants, in general, have an inaccurate perception of crime.
- People perceived a higher percentage of safe areas around the HCIAs. An explanation can be that people identified also unsafe areas around the places they are familiar with.

Based on this conclusions, two hypothesis were formulated. The outputs from the developed model in section 4.2.2 confirmed the first of these two hypothesis: *the likelihood of people perceiving an area as unsafe increases when the target area is far away from their neighbourhood and their daily routes*.

The quantitative exploration of the structured sketch maps makes their analysis with statistical and special methods possible. The data extraction and exploration enabled to conclude that the location of the neighbourhoods and the people's daily routes play a role in the spatial perception of crime. And, therefore, in order to narrow the crime perception gap, it is needed to consider these two factors in the design of strategies.

Once the factors that sway crime perception are identified and so the presence of perception gaps, the next step is to locate the places where people have an inaccurate perception of safety. For the aim of this research, an area is classified as safe or unsafe only based on the number of crime incidents reported in that area, due to the limited information available.

In the next chapter the methodology, results and conclusions of objective two (determining the level of accuracy of people's crime perception and its spatial distribution) will be presented.

# Chapter 5

## Crime perception accuracy: Spatial delineation

In this chapter the second specific research objective will be addressed:

*To determine and explore the accuracy of people's crime perception and to map its spatial distribution.*

The methodology, results and conclusions of dealing with this second objective will be presented, each of them will be discussed in a separate sections. The methodology (section 5.1) is divided into two parts. In the first one, the **spatial distribution of the accurately and inaccurately perceived safe and unsafe areas** will be determined, as well as the **level of accuracy** in each area. In the second part, the **Bivariate Local Moran's I statistic** will be explained in order to perform a spatial bivariate analysis of which the output could explain the locations of perception gaps.

In the second section 5.2, the results of the analysis performed will be presented. The outputs of the spatial distribution of the perception's accuracy and the bivariate analysis are presented in a **series of maps** which are described and interpreted.

The third part of this chapter (section 5.3) contains the conclusions related to the second objective.

## 5.1 Methodology

The aim is to perform a quantitative analysis of the sketched maps in order to define the spatial distribution of the perception gap and determine the accuracy level. The data description presented in the explanatory model showed evidence of a misperception of the crime rate. Figure 4.7 shows that some participants identified safe areas within a crime hotspot and some others marked unsafe areas more than one kilometre away from a hotspot, which indicates that there is an inaccuracy in the perception of low crime or high crime.

The methodology to address the second research objective is divided into two parts. The first one consists of defining the spatial distribution of the level of crime perception accuracy by comparing the class (safe/unsafe) to which each block belongs according to the perceived classification and the reference classification based on actual crime data. The perceived classification is defined by counting the sketch maps by type (safe/unsafe) that overlap within one block. Meanwhile, the reference classification of the blocks was defined by crime hotspots based on total count of incidents. Thus, if a block was labelled as a hotspot then it belongs to the unsafe class, and if it was not, then it was categorized as safe.

In the second part a bivariate spatial autocorrelation analysis is proposed as a method to explore the spatial relationship between two variables. In this case the inaccurate perception and the number of crime incidents.

### 5.1.1 Types and level of crime perception accuracy

City blocks are the spatial analysis unit that has been employed in this stage of the research. In order to determine the spatial distribution of the crime perception accuracy two variables will have to be compared: the "perceived" and the "reference" safe/unsafe areas. So, the blocks were classified as safe or unsafe based on the perception of people and based on the actual number of crimes per block. These two classifications were compared in order to check whether the block was correctly classified or not.

The "reference classification" of the blocks was done according to the crime hotspots. In the previous chapter, one of the explored covariates was the distance between a sketch polygon and the closest hotspot. According to the regression analysis that was executed, the hotspots as defined by the count of incidents per block explained a higher percentage of the likelihood of perceiving an area as unsafe, than the hotspots that were defined by the crime density per block. Therefore, in this stage of the research the hotspots identified by count of incidents formed the basis of the "reference classification". If a block is a hotspot then it belongs to the "reference unsafe class", if the block is not a hotspot, then it belongs to the "reference safe class".

The "perceived classification" of the blocks was based on the structured sketch maps. The first step was to count per type (safe/unsafe) the number of participants who sketched a polygon that has at least one cell's centroid within a block. Then, the percentage of participants who classified the block as unsafe from the total number of participants who sketched on that block was calculated. Obviously, the result ranged from 0 to 100, where 100 indicates that all participants agreed on classifying the block as unsafe and zero indicates that everybody agreed on categorizing the block as safe. 50

Indicates that the same number of persons identified the block as safe or as unsafe. Thus, when the percentage was higher than 50 the block was labelled as “perceived unsafe”, when it was smaller than 50 it was labelled as “perceived safe”, and when the percentage was 50 the block was “undefined”.

Table 5.1 is an extraction of the blocks’ attribute table to exemplify the way the blocks were classified based on the participants’ perception:

| Block ID | Hotspot | Participants who classified the block by type | | Total | % Unsafe | Perceived Class |
|---|---|---|---|---|---|---|
| | | Safe | Unsafe | | | |
| 1 | no | 13 | 6 | 19 | 31.5 | safe |
| 2 | yes | 15 | 36 | 51 | 70.6 | unsafe |
| 3 | yes | 3 | 3 | 6 | 50 | undefined |

**Table 5.1** Example of the block dataset including the “perceived classification”.

Figure 5.1 shows a scatter plot of the percentage of participants per block who identified it as unsafe and its corresponding number of crime incidents; each point represents a block. This plot shows the presence of a crime perception gap in the study area; as three of the four blocks with the highest amount of crime incidents were identified as safe areas by the majority of participants who sketched over those blocks. Contrary, some blocks were classified as unsafe where there were no reported incidents. On the other hand, there are also blocks that people are aware of the high and low crime rate.

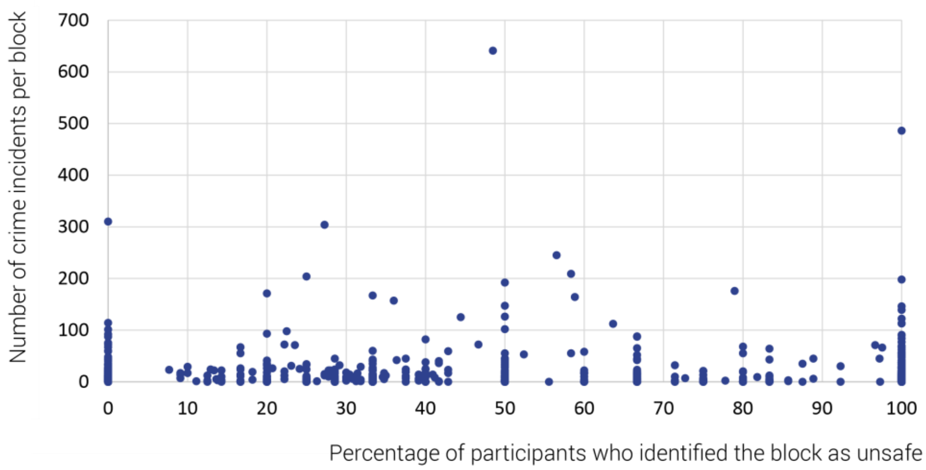



**Figure 5.1** Crime incidents per block and percentage of participants who identified the block as unsafe.

Thus, the blocks vector file contains, among other attributes, the values of the perceived and the reference classification. Both values were compared and the blocks were classified into one of the four types of crime perception accuracy: accurate perception of safe (AS) or unsafe areas (AU) and inaccurate perception of safe (IS) or unsafe areas (IU) (Figure 2.4).

The next step consisted of defining the level of accuracy or inaccuracy of people's perception. If the block was accurately classified (reference classification = perceived classification), the level of accuracy was defined by the percentage of participants who correctly classified the block by the total number of participants who classified the block. If the block was inaccurately classified (reference classification <> perceived classification), the accuracy level was defined by the percentage of participants who incorrectly classified the block by the total number of participants who classified the block.

Based on the percentage values, an ordinal classification was defined to determine three levels of accuracy: low (>50% - 65%), medium (>65% - 85%) and high (>85% - 100%). For the two accurate classes (AS and AU) this scale represents the proportion of participants who are aware of the safety situation, in this case the blocks that were labelled as low accuracy means that the proportion between the people who were accurate is slightly higher than those who were not.

For the case of the two inaccurate classes (IS and IU) the scale represents the proportion of people who are not aware of the crime situation. In this case the blocks that were classified as high require more attention than those that were labelled as low. There are two main possible answers that could explain the low accuracy: either people started to feel unsafe because of a recent event or start to feel safe as a response to some police activity. For both cases the question could be answered by a spatiotemporal analysis where it could be compared to the speed of the change and also the criminal activity.

Table 5.2 shows an example of the accuracy type classification and the level of accuracy.

| Block ID | Participants who classified the block by type | | Total | % of participants who classified the block by type | | Classification of the block | | Accuracy type | Level of accuracy |
|---|---|---|---|---|---|---|---|---|---|
| | Safe | Unsafe | | Safe | Unsafe | Reference | Perceived | | |
| 1 | 4 | 7 | 11 | 36.4 | 63.6 | unsafe | unsafe | AU | Low |
| 2 | 7 | 1 | 8 | 87.5 | 12.5 | safe | safe | AS | High |
| 3 | 13 | 4 | 17 | 76.5 | 23.5 | unsafe | safe | IS | Medium |
| 4 | 3 | 19 | 7 | 13.6 | 86.4 | safe | unsafe | IU | High |

**Table 5.2** Example of an accuracy type and level classification of 4 blocks.

A matter that must be considered as well is the number of people who classified each block, as the level of accuracy is based on the total amount of participants who classified a block but not on the overall total amount of participants. This could be solved by a threshold selection that indicates the minimum number of the total participants who classified the block, for instance, "show only the blocks that were classified by more than 10 participants".

According to the final classification of blocks, the crime perception gap is identified where the blocks were classified as "inaccurate perception of safe areas" and "inaccurate perception of unsafe areas", as in these blocks the perception does not correspond with reality. The relevance of distinguishing between these types of inaccuracy lays in the fact that the strategies needed to narrow the perception gap are different for each type of inaccuracy: whereas in the IS people need to be aware of the risk of victimization, in the IU the strategies must be focused on reassuring the people. In order to develop plans of action it is required to explore the possible causes that explain the inaccuracy in those specific locations.

Therefore, once the spatial distribution of the perception gap is identified, the next step is to explore the possible causes. As this research has the aim to explore the spatial attributes of the factors that sway perception, the proposed analysis has the purpose of defining the spatial relation between the location of the perception gap and the spatial distribution of the values of a second variable. For instance the block where an IU exists and the location of hotspots. So the aim is to explore how the perception of safeness in one block could be related to the number of crime events in the neighbouring blocks.
The Bivariate Local Moran's I method is a bivariate spatial autocorrelation statistic which analyses two variables in different locations, thus is apt for this type of analysis. In the next section this method will be explained.

### 5.1.2 Bivariate Local Moran's I

The bivariate local Moran's I is a spatial autocorrelation measurement which relates the value of one variable in a given location and the average value of the neighbouring features of a second variable, which means the two variables are not analysed in the same location. The value of the first variable in one location is compared with the average value of a conditional permutation performed with the neighbouring features.

The outputs of the bivariate Local Moran's I analysis in GeoDA is a cluster map which classified the significant spatial units into high-high, low-low, high-low and low-high, where the first attribute corresponds to the value of the first variable and the second the value of the second variable in the neighbouring areas.

As the input file must contain numerical values, the two tested variables were the percentage of participant that defined the area as unsafe from the total participants who classified that block (<50%=safe area and >50%=unsafe area) and the number of events occurred per block. The aim is to define if there is a relation between the locations of IS and the surrounding high crime rate areas or IU and the surrounding low crime rate. The analysis is not meant to explain the inaccuracy of perception but it will show the spatial relations between the two input variables.

## 5.2 Results

In total there are 9,655 blocks in Budapest; from which 1,706 lie within the sketched polygons, and thus they were classified by the participants as unsafe or safe. Only these blocks were examined in the crime perception gap analysis. From these classified blocks, 302 are actual crime hotspots and they were classified as "reference" unsafe areas. The rest (1,404), for the purpose of this research, were considered as "reference" safe areas as they are no hotspots.

Below the results of the identification of the perception gap and the bivariate spatial autocorrelation analysis are described.

### 5.2.1 Spatial arrangement of crime perception

The classification of the blocks was done based on the four types of accuracy of crime perception (Figure 2.4). The result of the analysis is two maps, one that shows the blocks that were accurately classified and another one that depicts the inaccurately classified blocks. The maps are centered in the part of the city where the participants sketched the polygons.

Figure 5.2 shows the blocks that were accurately perceived. 37.7% (114) of the blocks identified as hotspots were accurately perceived as unsafe, meanwhile, 37.3% (524) of the non-hotspots were accurately perceived as safe.

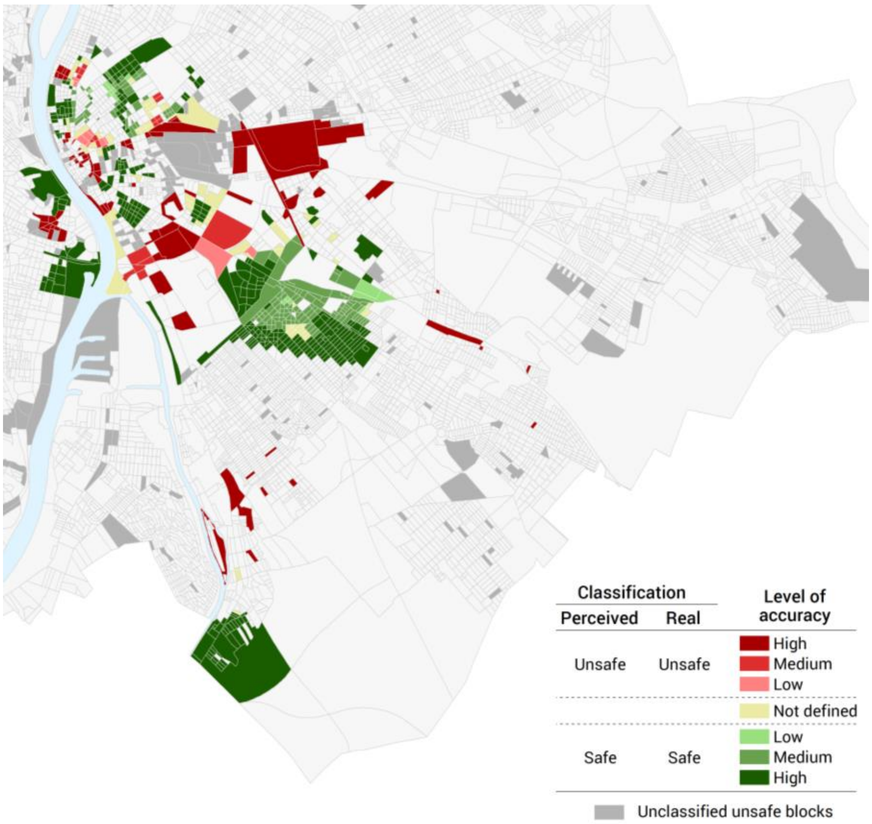



**Figure 5.2** Level of accurate perception of unsafe (AU) and safe (AS) areas per block.

The map shows some visible clusters of safe and unsafe areas where people are aware of the crime rate. The lightest green areas are those blocks where prevention actions must be taken, as in comparison with the total number of participants who sketched over those blocks, the percentage of those who are aware that the area is a crime hotspot is low, which means the ratio between the people who perceive it as safe is slightly higher than those who perceived it as unsafe.

Figure 5.3 depicts the blocks that were inaccurately classified, thus this map show the actual crime perception gap. 54% (163) of the hotspots blocks were inaccurately perceived as safe, and 58.5% (822) of the safe blocks were inaccurately perceived as unsafe block. In the centre of the city, people tend to have an IS, meanwhile the IU happens in the south and southeast part of the city.

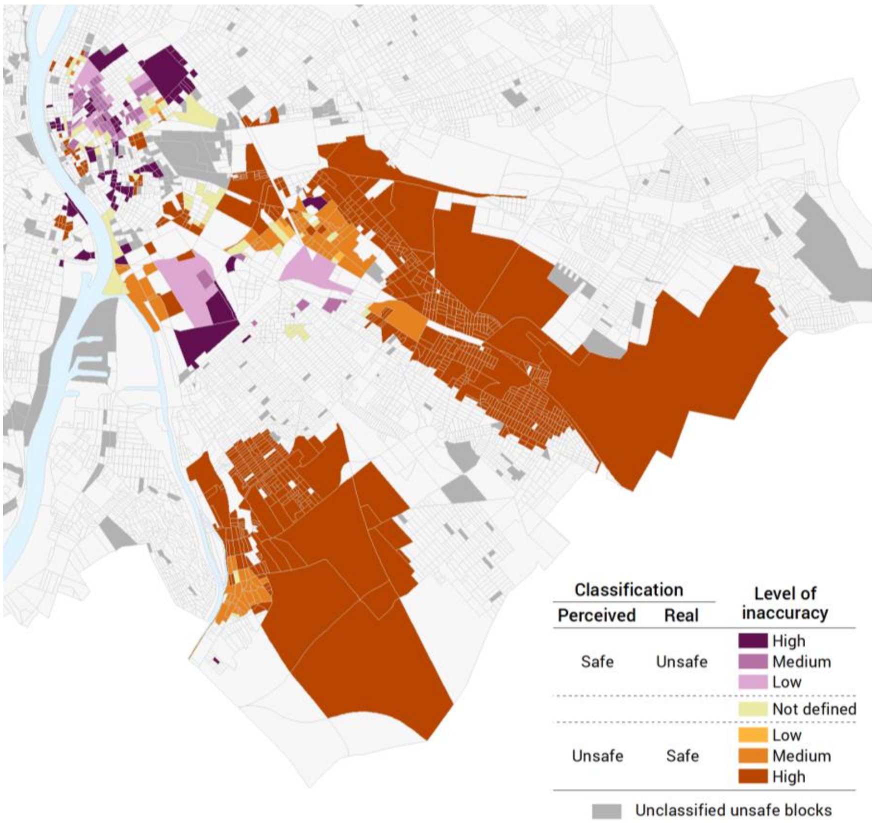



Figure 5.3 Level of inaccurate perception of safe (IS) and unsafe (IU) areas per block.

The map also shows those hotspot blocks that were not classified by the participants, these are considered as another block type due to the fact that they are conceptually part of the perception gap. But as they do not have a value in the "perceived classification" attribute and the level of inaccuracy cannot be measured, they are not considered to belong to the "inaccurate perception" class because they were not classified by the participants. They are shown in the map because they should be considered in the design of strategies to narrow the perception gap.

The accuracy of the participants' perception is presented in Table 5.3 which is a confusion matrix showing the blocks that were correctly and incorrectly classified, as well as the commission and omission errors.

Out of the 1,706 classified blocks, 83 were 'not defined' due to half of the participants classified those blocks as safe and the other half as unsafe; thus, only 1,623 blocks were labelled as safe or unsafe. From the labelled blocks, 61% of the safe blocks were identified as unsafe and 59% of the unsafe blocks were identified as safe. The overall accuracy of the classification is 39%, which is the percentage of accurately classified blocks.

| | | Reference | | | Error of commission |
|---|---|---|---|---|---|
| | | Safe | Unsafe | Total | |
| Perception | Safe | 524 | 163 | 687 | 0.24 |
| | Unsafe | 822 | 114 | 936 | 0.88 |
| | Total | 1,346 | 277 | 638 | |
| Error of omission | | 0.61 | 0.59 | | **0.39** |

**Table 5.3** Confusion matrix of the crime perception classification.

One thing that must be considered is the percentage of participants who classified each block (Figure 5.4). Although the level of accuracy was categorized as high in Figures 5.2 and 5.3 that does not necessarily mean that a meaningful number of participants classified the block concerned. Therefore, it is import to take into account the number of participants who sketched over each block to decide whether or not the level of accuracy is significant: the bigger the amount of participants the more meaningful the classification would be. For instance, whereas Figure 5.3 shows all blocks that were incorrectly classified, Figure 5.5 shows only the blocks that were classified by more than 10 participants.

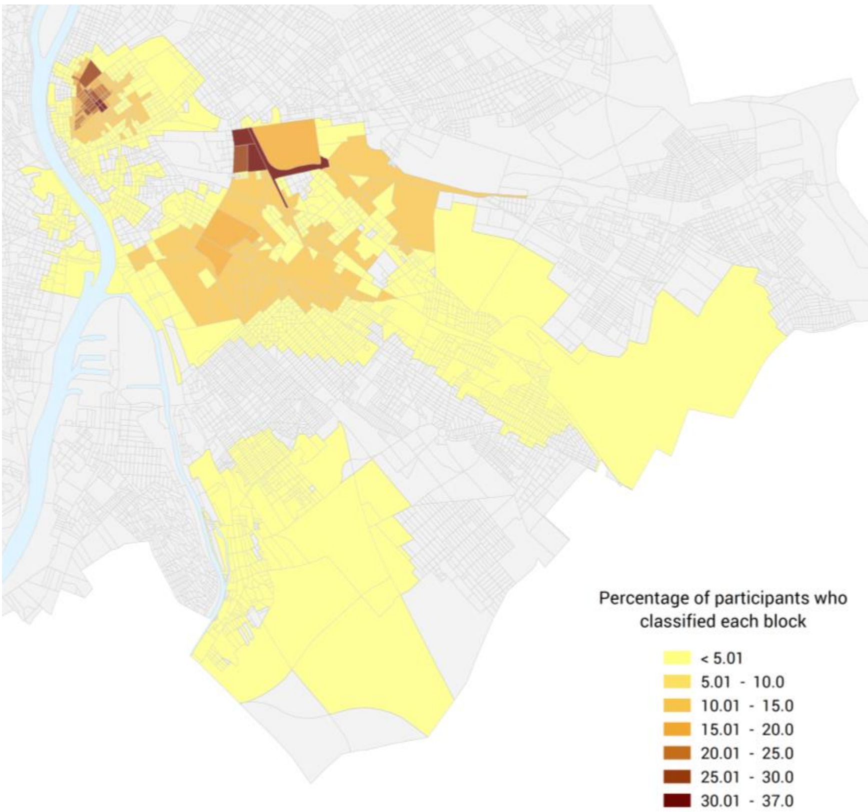



**Figure 5.4** Percentage of the total number of participants who classified a block.

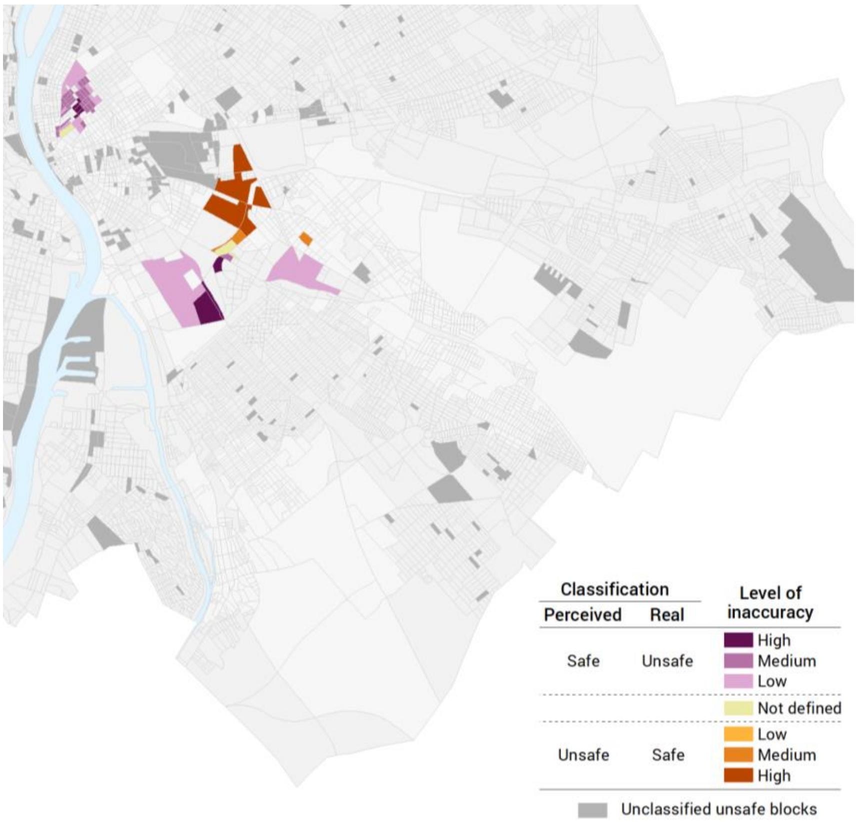



**Figure 5.5** Level of inaccurate perception of safe (IS) and unsafe (IU) areas per block classified by more than 10 participants.

In figure 5.4 can be observed that most of the participants focus on the centre of the city. A reason could be that most of them indicated to live in this area (Figure 3.1), in contrast with the south / south-east of Budapest where less than five participants sketched over these areas. If Figure 5.3 and 5.4 are overlaid, it can be observed that the large clusters of IU blocks were classified by less than five persons, whereas the blocks in the centre were classified by up to 41 participants. Although the level of accuracy does not evidence these differences, it is meaningful in terms of depicting the ratio of people perceiving an area as safe to those who perceived it as unsafe. A visual comparison of both maps (in Figures 5.3 and 5.4) shows that the priority area is the city centre, next to the river, as a high amount of people is not aware of the high crime rate there.

### 5.2.2 Local spatial autocorrelation analysis

Although crime perception depends on several intrinsic and extrinsic factors, location plays an important role in terms of the spatial association that people could make among contiguous places. The way attributes of a location can be easily transferred to nearby sites could explain the misperception of safeness.

The bivariate spatial autocorrelation analysis aims to explore the relationship between the perception of safeness in one place and the criminal activity in the surrounding areas. This relation could provide some explanation of the location of the perception gap.

The method used to perform the analysis was the bivariate local Moran's I, with queen contiguity of first order and 999 permutations. The two input variables were the perceived classification given by the percentage of participants who identified a block as unsafe (>50% = unsafe area and <50% = safe area) and the number of events in the surrounding blocks. The output map is shown in Figure 5.6.

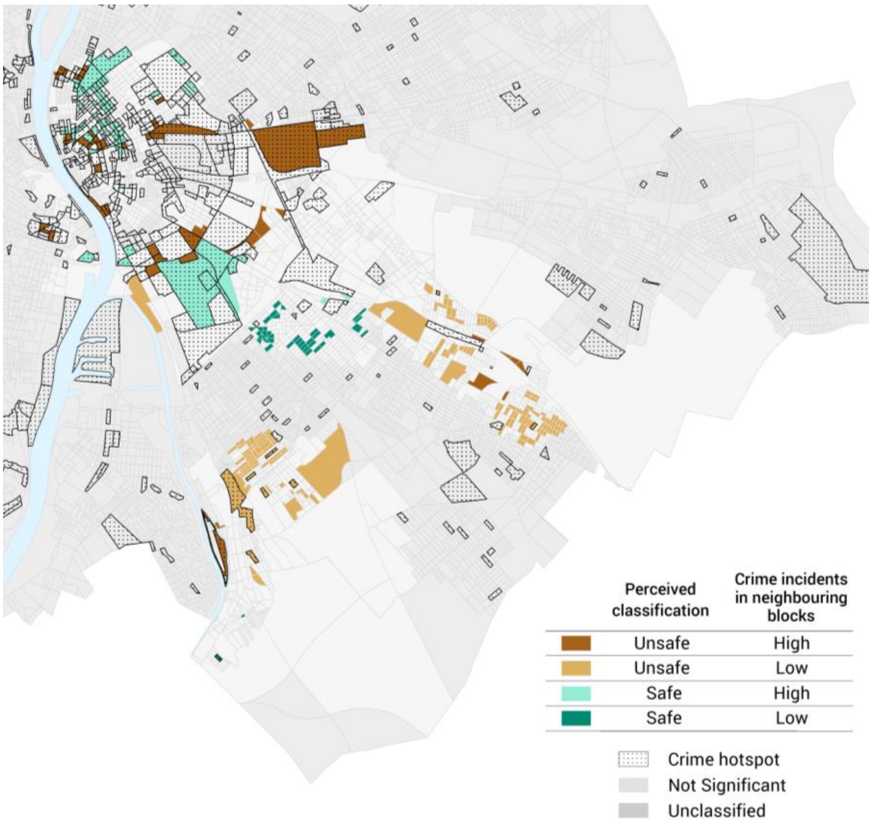



**Figure 5.6** Cluster map result from a Bivariate Local Moran's I analysis in GeoDa, performed with the perceived classification in the target block and the number of crime incidents in the neighbouring blocks.

The turquoise colour represents the perceived safe blocks, the dark ones are surrounded by blocks with low crime incidences, and the light ones are bounded by blocks with high crime incidences. Meanwhile, the brown blocks are perceived unsafe areas and contrary to the turquoise blocks, the neighbouring blocks of the dark browns have high crime incidences and the light browns low crime incidences. Additionally, the hotspot blocks are shown for a better reference of the relationship between both variables. The light grey blocks are areas that are not significant, that means, those are blocks which neighbours' values are not significantly different from the value resultant from a random permutation. The dark grey areas are those blocks that were not classified by the participants.

The following step was to select, from the significant identified blocks in the bivariate spatial autocorrelation analysis (turquoise and brown blocks in Figure 5.6), those which were previously labelled as "inaccurately perceived" (Figure 5.3). Figure 5.7 shows the result of the selection.

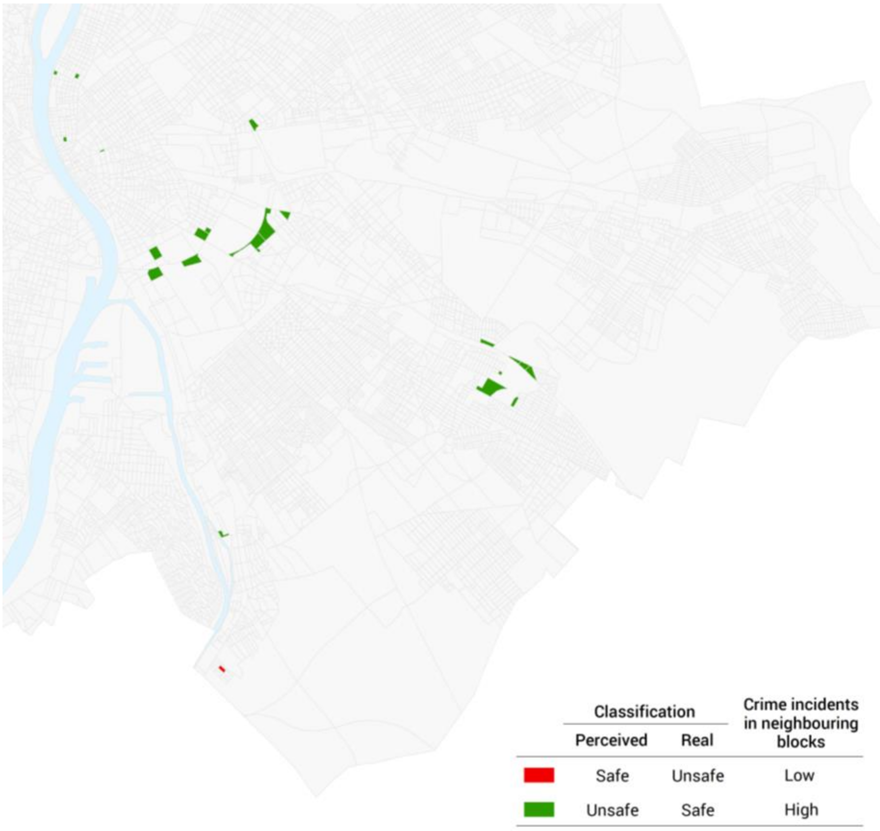



Figure 5.7 Inaccurately perceived blocks and actual crime rate of their neighboring blocks.

This map depicts in green those blocks that were inaccurately perceived as unsafe and of which neighboring blocks have high crime incidences. This relation could explain the inaccurate perception of safe areas, as the surroundings of the blocks perceived as unsafe could have had an impact on people's perception. They could believe, by spatial association that those selected blocks were actually unsafe areas due to the characteristics of the enclosing blocks. If this is this case, then the perception gap in those inaccurately classified block could be narrowed down by reducing the criminal activity in the surrounding areas.

The red block was inaccurately perceived as a safe area, whereas it is safe in reality. Similarly to the previous case, this could be explained by the fact that the surrounding areas have low crime incidents and that due to the closest distance to low crime areas, the block is perceived as safe.

This type of analysis are usually performed for spatiotemporal studies, in which the same variable is compared in two different time moments (Anselin, 2019). This type of analysis could also be perform for analyzing perception, to explore how perception change along time according to past events. Unfortunately, the perception dataset (sketch maps) that was available for the case-study did not contain temporal information.

## 5.3 Conclusions

This chapter presented the analysis to address the second objective. In the first part of the analysis the perceived and reference classification of safety were compared to define the type of crime perception accuracy in each block. Then, based on the number of participants who classified them "correct" or "incorrect", depending on the type of accuracy, the level of accuracy was specified. This classification allows determining the blocks that could be priority areas for strategies directed to narrow the perception gap. It also must be considered the total number of participants who classified the block, for it, a selection per attribute can be done over the classified blocks and only select, for instance, those that were classified by more than 10 participants; the higher the number of participants the more meaningful the classification is.

For the whole analysis only the classified blocks by the participants were considered. The location of these block coincides with the districts where they live, which can explain why they focused on those areas of the city.

The total accuracy of the participants' classification (perceived classification) was 0.39, which means 39% of the blocks were accurately classified. Besides from those unsafe areas (hotspots) that were inaccurately classified, there are those that were not considered by the participants and therefore also not in the analysis, nevertheless, they must be taken into account for the design of strategies as people are not aware of them.

Whit this result the second established hypothesis (*People's misconception of crime reality involves both the overestimation of safe areas (inaccurate perception of safe areas- IU) and the underestimation of unsafe areas (inaccurate perception of safe areas -IS))* is accepted.

These findings imply that the people are not well informed of the safety situation in their city and/or that there are other factors affecting of the inaccuracy of perception that should be detected in order to narrow the gap. From a criminology theory perspective we should be concerned (and further explore) if and how such a situation (i.e. a crime perception spatial gap) affects the current crime prevalence of an area.

The second part of the analysis consisted in calculating the bivariate local Moran's I statistic to identify the relation between the location of the perception gap (blocks inaccurately classified) and the number of crimes reported in the surrounding areas. This was done in order to explore the impact that the crime events have on the perception of safety in adjacent areas. The output map shows only a few blocks where this relation is statistically significant mainly located near to the city centre.

This type of analysis is suitable to explore the spatial association that people tend to do by transferring attributes from one location to adjacent areas. This could be performed eventually with a different pair of variables to explore their spatial relations.

The maps presented here could be used as analysis layers for further exploration, they can be filtered, overlap, compared or perform the bivariate analysis with a different selection of layer. The number of possible combinations of attributes and conditions is already large and if the variable "time" was to be included, then space turns into a complex scenario of multiple relations between features and spatial and non-spatial attributes. These relations can be tough to discern and therefore require suitable methods and analytical tools.

In the next chapter, a prototype proposal of a GeoVisual Analytics environment is presented that is intended to ease the analysis of crime perception data.

# Chapter 6
# Development of a GeoVisual Analytics environment

In this chapter, the third specific objective will be addressed:

*To conceptually design a geovisual analytic environment for the exploration and reasoning of perception of crime.*

In this chapter a brief theoretical background in Visual Analytics and **GeoVisual Analytics** (GVA) is presented as a frame for data integration, exploration and understanding of complex spatial data relations. These foundations are considered to present a proposal of a GVA interface prototype that supports the visualization and analysis of spatiotemporal data related to the perception of crime.

This chapter is divided into five sections. The first two comprise a brief description of GVA (section 6.1) and the **user-centred design** (section 6.2) approach. In the third section a **requirement analysis** (section 6.3) is presented that includes the potential users of the interface and the main objective of it.

The fourth section contains a low-fidelity **prototype** (section 6.4) of the proposed interface, including the description of the tools, as well as some examples of the outputs.

The **conclusions** of this chapters are summarized in the fifth section (6.5).

## 6.1 Geovisual Analytics Environment

In the two previous chapters, the aim of analysing the crime perception was presented through exploratory modelling (Chapter 4) and the spatial arrangement outline (Chapter 5) of the data extracted from the structured sketch maps. These two stages defined the causality and location of the perceived safe and unsafe places in Budapest, which additionally resulted in the estimation of people's perception accuracy.

Knowing the aspects that sway peoples' crime perception is not sufficient for its understanding, and conceptualizing them as isolated factors could result in a rather focused study. On the other hand, identifying the location of the accurately and inaccurately perceived safe and unsafe areas is not enough to develop strategies to reduce the fear of crime or to make people aware. Causality and location must be studied by their relations, in this case their spatial relations, to capture a holistic overview of the context of perception.

Therefore, in order to understand the spatial behaviour of a phenomenon, it must be addressed from a system approach in which its parts are interrelated. The three components of a system are elements, states and the relations between these two. The elements are the collection of physical objects and their states or properties are the attributes (Huggett, 1980) that they are given or they possess. In geographic and spatial studies the relationships are determined by the location of the elements and states, employing maps as the main tool, among other data representations, with the purpose of solving two central questions: where and why.

The "way" from element and status to relations is usually not straightforward and additional exploration data methods and tools are required to analyse the complexity of those associations thoroughly. An interactive interface that integrates multivariate datasets can enable the user to explore and understand the complex relationships within the data (Kveladze, Kraak & van Elzakker, 2017).

Visual Analytics (VA) has emerged as an integrating science that combines '*automated analysis techniques with interactive visualizations for an effective understanding and reasoning of multiple datasets*' (Keim, Andrienko, Fekete, Görg, Kohlhammer & Melançon, 2008). It integrates data mining, data fusion, graphic representations and statistics, among others, in which visualization is '*the medium of semi-automated analytical processes*' (Kohlhammer, May & Hoffmann, 2009), as '*the analyst observes and interacts with the current data representation, interprets and makes sense of what he or she sees*' (Cook & Thomas, 2005).

In terms of spatial data, the concept GeoVisual Analytics follows the same principals as Visual Analytics but specifically for the analysis of georeferenced data that may have temporal attributes too. Kveladze et al. (2017, p. 207) define it as:

> "A GeoVisual Analytics (GVA) environment is based on highly interactive and dynamic visualization techniques intending to reveal knowledge in complex and multivariate geodatasets. By depicting information, these techniques amplify human capabilities and facilitate the performance of cognitive tasks for pattern recognition, decision-making or analytical reasoning."

The main foundations of this science are visualization and analytical reasoning; the latter is defined by Cook and Thomas (2005) as the method in which '*users obtain deep insights that directly support situation assessment, planning and decision making*'. The user is involved as an active component within the environment, as the core of GVA enhances the synergy of computational approaches and human reasoning to answer questions directed to solve spatial problems.

The extracted answers from the visual and analytical reasoning would be in the direction of a better understanding of the target problem, which will consequently lead to solutions that are more efficient for the decision-makers. Therefore, the interface requires to be designed for target users with particular needs and characteristics. The user-centred design approach, which places the users as the reference point in the development process, must be the model to follow in the GVA design process.

## 6.2 User-centred design

The aim of a user-centred design approach is to achieve an *interface success* in terms of the user, utility and usability, following an iterative process. Utility refers to how useful the interface is for solving a specific task and usability describes how easy it is to complete a target task using the interface. And the users are the main focus of this approach as the design, evaluation and revision processes of the interface are based on the users' profile and needs assessments (Roth, Ross & MacEachren, 2015). Roth, Ross, Finch, Luo and MacEachren (2010) modified the user-centred design approach presented by Robinson et al. (2005 as cited in Roth et al., 2010) into a six iterative stages process in which the users' feedback plays the main role (Figure 6.1). They applied this approach in the development of *GeoVISTA CrimeViz* '*an interactive and web-based mapping application supporting visual analytics of criminal activity in space and time*' (Roth et al., 2015).

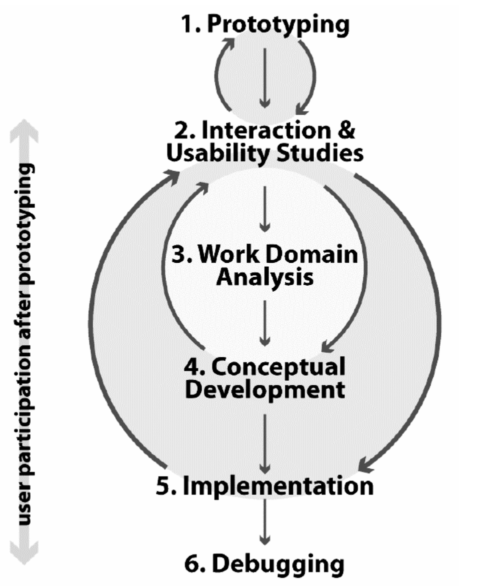



Figure 6.1 Modified user-centred design approach by Roth et al. (2010).

The modified process starts with a prototyping stage, followed by interaction and usability studies performed on the prototype. After this, a work domain analysis is done to capture impressions and ideas from the target users; the work domain comes after the prototype as it is common that in a project the initial prototyping is usually performed by designers and developers and later on it has to be tried with the actual user group. This feedback is integrated to the initial prototype to implement the interface. The process requires constant modifications which make this process an iterative one, as the users are involved by giving feedback and remarks after the prototyping stage in order to assure the functionality and design by effective means of interaction.

To address the third objective of this research, the design of the GVA interface will only cover the first stage of the process as a proposal for future studies. For the design of the prototype a requirement analysis to know the users' profiles and needs is necessary and is explained further in the next sections.

## 6.3 Requirement analysis

The design of a prototype is the first stage in the user-centred design approach. There are two classes of prototyping: low-fidelity and high-fidelity. The former refers to an exploration product that is simple and quick to produce, mainly paper-based, as it has to be flexible for alternative design and ideas that may modify it. The latter is a product that looks similar to the final one and it usually allows user interaction (Preece, Sharp & Rogers, 2002).

In the prototyping stage, the target users are not directly involved, instead, the developers make a user requirement analysis that is a four steps process (Figure 6.2) in order to have a context on which to base their interface wireframe.

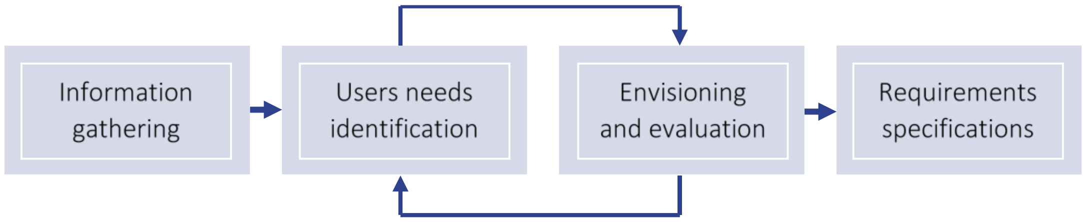



**Figure 6.2** General process for user requirements analysis (Maguire & Bevan, 2002).

In the information-gathering stage, the developer builds up a background of the target users and the problem to address. Then the needs of the users are listed and with these elements, a prototype can now be designed. The third stage consists of revising the prototype, which can result in the redesign of it. In the final stage, all the requirements are documented.

Due to the limits of this thesis research, only the two first stages of the process will be performed. The aim is to present a general context where the interface can be base on. Hereunder the problem to be addressed will be stated, the target user group and the questions that are intended to be answered with the use of the interface will be described.

- Problem

  Exploring crime perception data means dealing with multivariate spatiotemporal datasets. This requires an environment that integrates different tools, data representations and views that ease the understanding of the complex relationships between the spatial and non-spatial attributes of the analysed features.

  Therefore, the aim is to develop a GeoVisual Analytics interface that supports spatiotemporal crime perception and related data, in order to explore and analyse the spatial arrangement of the crime perception accuracy within a contextual scenario. The interface is envisioned to ease the identification of spatial patterns, structures, changes and relations in a multiscale environment that will assist the users in the formulation of strategies and action plans directed towards increasing the accuracy of perception and thus, narrowing the crime perception gap.

- Users

Police agencies have been identified as the bureaus in charge of creating strategies to reduce fear of crime (Cordner, 2010; Grabosky, 1995; Bennett, 1991). Usually, their action plans are focused on increasing the confidence towards the police department and reducing fear and they are mainly directed towards reducing the inaccurate perception of unsafe areas. As mentioned before, attention is also required in the areas where the people have an inaccurate perception of safe areas and a low percentage of accurate perception.

The proposed interface is intended to support police agencies in the decision-making process of designing strategies to narrow the crime perception gap. The organization of each police agency is usually divided in units, departments and divisions that have different responsibilities in developing action plans and policies to enhance public safety. To identify specific target users is necessary to get in touch with the respective police agencies because the organizational chart of each police department differs. This was not possible to do in this research but it is strongly recommended for further iterative prototype development.

- Questions

The questions listed here are some of those that the interface is intended to assist in answering. Some of them are mainly exploratory but others arose while interpreting the outputs of the analysis performed in objective one and two. The answers may enrich the understanding of perception when the output maps are not sufficient.

- Where do people have an accurate and inaccurate perception of crime?
- Is crime perception limiting people's daily activities routes?
- Do people have an inaccurate crime perception of the own neighbourhood?
- How does perception change over time?
- Is perception change related to the increase or decrease of crime incidents?
- Is inaccurate perception of a place related to the level of crime incidents in surrounding areas?

Based on these general ideas a heuristic prototype will be presented. The proposed functionality and tools are based on visualization methods that could solve the questions that were identified and may lead to a better understanding of the spatial dynamics of the perception of crime. In the next section, the conceptual design of the interface will be presented with a low-fidelity prototype that depicts its structure and organization. Additionally, a more detailed prototype will be presented to clarify the navigation scheme.

## 6.4 GVA prototype

The aim of a conceptual design is to show the overall organization of an interface and the relations between the different functionalities; prototypes are the main products that illustrate the design. This proposal follows a heuristic approach. The organization of the interface is similar to a GIS in which the map panel is the main one. Some tools are based on GIS functions and others are inspired by the different steps performed in the process of data extraction from the sketch maps (spatial queries ran when dealing with research objective one, see section 4.2.1) and the data analysis (spatial operations performed when dealing with research objective two, see sections 5.1.1 and 5.1.2).

Figure 6.3 is the low-fidelity prototype that depicts the organization of the proposed GVA interface, including three toolbars and four panels. Figure 6.4 shows a more detailed representation.

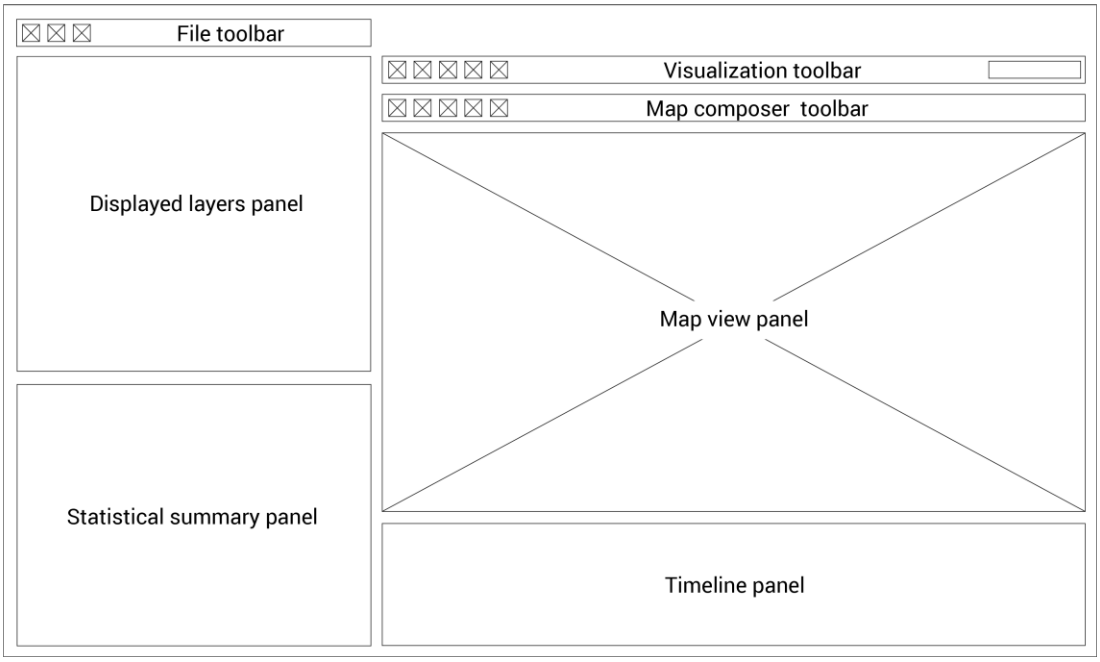



Figure 6.3 Low-fidelity prototype of the proposed GVA interface for crime perception data that shows the organization of the toolbars and panels.

Each component of the interface is described below:

- File toolbar

  This toolbar is placed in the top left corner. Like in most interfaces, it holds three buttons: one to create a new project, one to open an existing project and one to save the current project.

- Visualization toolbar

  This toolbar contains visualization options and “selection” tools. The aim of these tools is to focus attention on a specific attribute or location; the outputs of the selection can be turned into a map layer for further analysis.

  1. **Information** (button): retrieve attribute information when clicking on a feature within the map view panel.
  2. **Area of study** (button): centres the map panel in the area of study (in this case: Budapest).
  3. **Charts** (window): create a plot of the selected attributes. Unlike the statistical summary panel, this tool allows complex charts to be created. The created graph will be shown in a separate window.

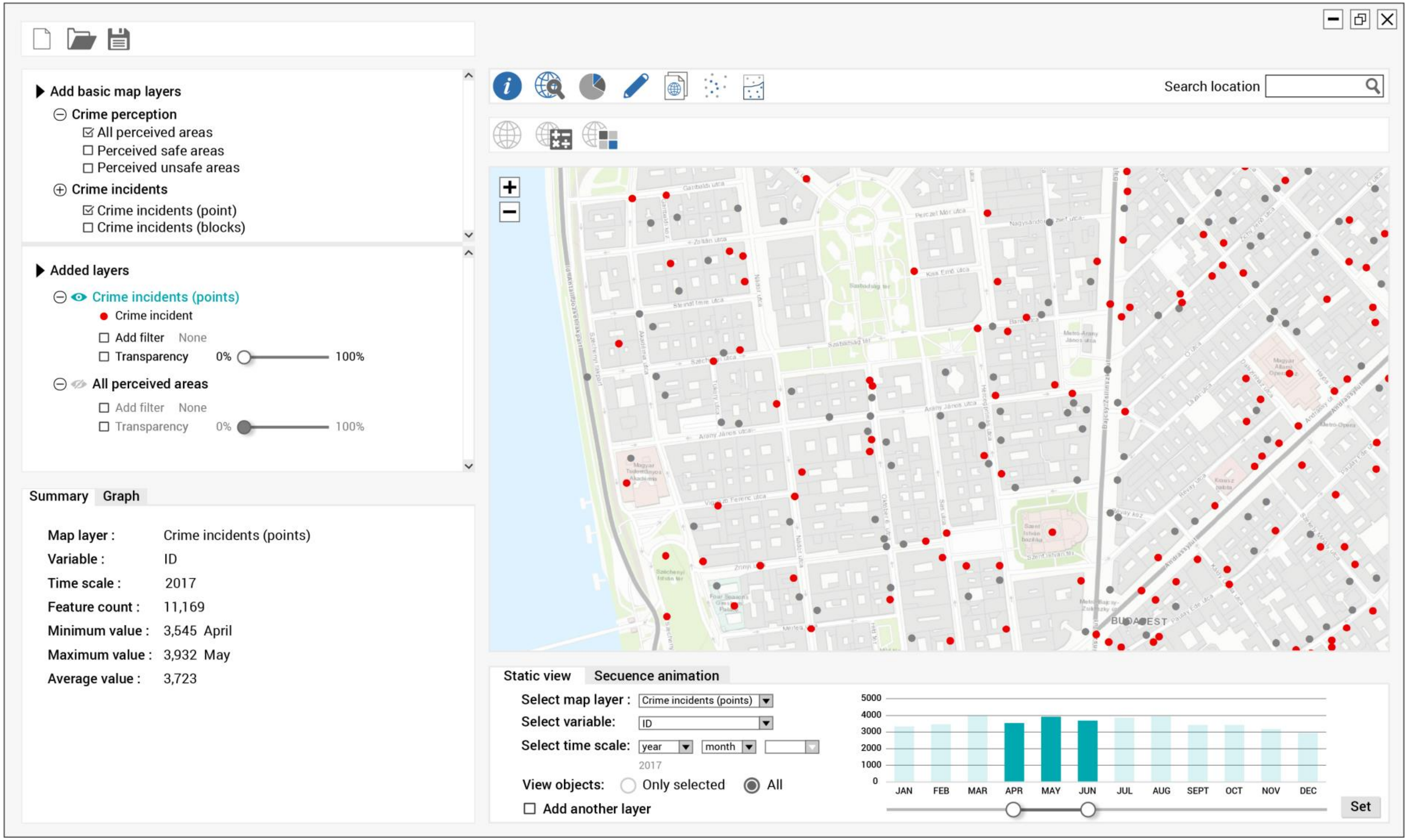



Figure 6.4 Detailed low-fidelity prototype of the proposed GVA interface for crime perception data.

4. **Draw** (dropdown button): allows sketching on the selected layer. The traced element is shown as an image over the layer, so no new features are added to the layer. The dropdown list contains four shapes of drawing: a square, a circle, a polygon or freehand style. This tool is useful to highlight elements of interest for faster identification. The sketches will be done in the selected layer (the selected layer is indicated in a blue colour in the displayed layer panel).

5. **Layer comparison** (window): compare two or three map layers. In the window options, two methods of comparison can be selected: combined view and swipe view. The combined view option divides the map view panel into two or three linked views in which the selected map layers are displayed. In the swipe view, two panels are overlaid: one stays static and the second one can be dragged to make visible the one that is underneath.

6. **Selection on the map** (button): enables the selection of more than one feature from the same layer and shows a descriptive statistics summary of the selected feature(s).

7. **Selection by intersection** (window): selection of features from two layers that spatially intersect. In the pop-up window, first the target area is chosen, either by selecting it by attribute or directly in the map view panel. Then the layer that contains the features of interest is selected. The output is the features from the second layer that intersect the target area, also a summary of the output features is shown in the statistical panel. Figure 6.5 shows an example of the tool where all the assault events are is shown that were reported in the Erzsebetvaros district.

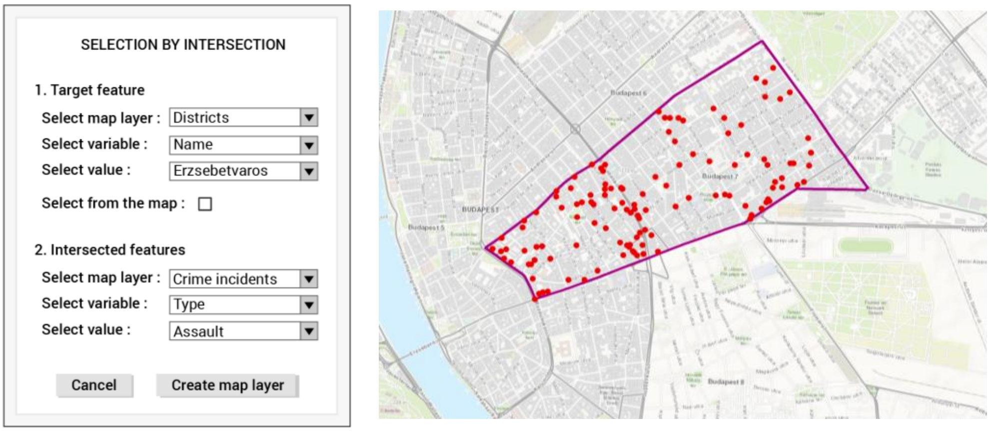



Figure 6.5 Modal window and map view example of the "selection by intersection" tool. The map shows the assaults reported in Erzsebetvaros district (Budapest).

8. **Selection by attribute** (window): the user can build standard SQL queries by selecting the layers of information and the operators to easily build the query statement.

9. **Linked mix selection** (window): selection of features from two layers that share at least one attribute column. First, a target area is selected by attribute or directly on the map (the layer must be added first, for instance the district layer), then a second layer is selected. Finally, one common attribute that appeared in both layers is selected. Figure 6.6 shows an example of the output of the tool, in it, the Erzsebetvaros district was selected as the target feature, the second layer contained the sketch maps/polygons (all perceived areas) and the common attribute

between both layers was the ID of the participant. Thus the map shows the polygons (purple areas) traced by the participants who live in Erzsebetvaros (purple line).

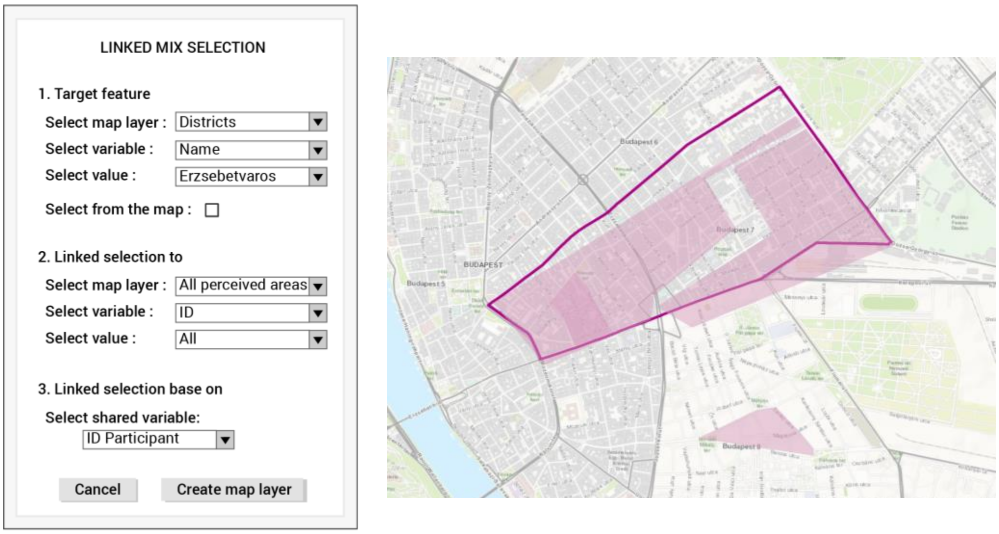



Figure 6.6 Modal window and map view example of the "linked mix selection" tool. The map shows the sketch maps (purple area) traced by the participants who live in Erzsebetvaros.

The tool can be useful to explore whether the people classified areas in other districts different from the one where they live. This can also be done with the neighbourhoods and the daily routes, for instance.

10. **Search by address or place** (text field): search an address or the name of a place and the map will be centred in that location.

- Map composer toolbar

With these tools, the user can select the unit of analysis: either block, postal code zone or districts. The maps layer created will be added to the displayed layers panel where the legend will also be shown.

1. **Univariate map** (window): creates a univariate choropleth layer map by selecting the unit of analysis and one attribute from a map layer. The values in the map are given in percentages.

   Figure 6.7 shows an example of the modal window and the resultant map. In this example, the percentage of incidents reported per district is shown. In the window, only one filter can be selected. In case another filter has to be applied, this can be done directly in the displayed layers panel (Figure 6.4).

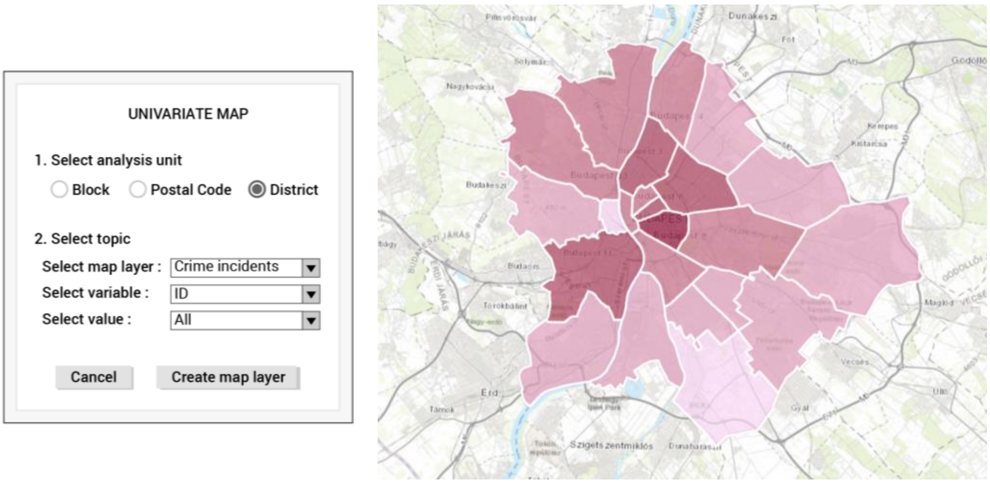



Figure 6.7 Modal window and map view example of the “univariate map” tool. The map shows the percentage of incidents reported per district.

2. **Index map** (window): Creates a choropleth layer map with the values of a constructed index. In the displayed modal window, the user selects the layers and their attributes that are needed to develop the index, as well as the math operators. For instance, a choropleth map of the crime rate (number of incidents per number of inhabitants) per district.

3. **Bivariate-adjacency map** (map): Creates a Bivariate Local Moran’s statistic map. The first step is to select the unit of analysis. Thereafter, the first variable is selected (for the target area unit) and then the second one (for the surrounding area units). The queen contiguity type is selected by default but the user must select the order of contiguity. The output map would be similar to the one of Figure 5.6, in which the unit of analysis is blocks. The first variable is the perceived classification, and the second the crime incidents.

   This type of maps is useful to explore the relation of perception and spatial association between two variables. A bivariate-adjacency map layer can also be displayed in the timeline panel by aggregating the information per time frame in order to explore how the changes in one variable can have a spatial impact on the values of a second variable.

- Map view panel

This the central panel of the interface. The rest of the panels are linked to what is displayed and selected on the map. The map layers are displayed here for their spatial exploration and visual analysis.

The map supports zooming and panning to change the current scene of the map. These actions can easily be done with the scroll wheel mouse, which facilitates the navigation. But, still, the zoom in and out icons are within the map view panel.

- Displayed layers panel

This panel is divided into two subpanels. The first one shows a list of available preselected layers to display in the map like the sketched unsafe/safe polygons, crime incidents, daily routes, districts and postal code areas. Once one layer is selected, it is added to the second panel which contains the list of displayed layers. The created map layers from the map composer toolbar are added directly in the second subpanel.

All the layers listed in the panel can be switched on/off and used as information layer for other analysis. They can also be filtered and the transparency percentage can be set (see Figure 6.4).

- Statistical summary panel

  Displayed here are either graphs or a written summary of the selected features. Basic descriptive statistical information, such as the minimum and maximum values, average values and total counts of the selected features are displayed in the panel. It is also possible to display a plot like a histogram or pie chart.

- Timeline panel

  In this panel, the user can filter the map layers by their temporal attributes and depict it in a static map or in a sequence animation. The static view pane showed in Figure 6.8 is divided into two parts. In the left part, there are three selected variables with dropdown-list menus. The first two selections are the map layer of interest and the variable to be displayed. The dropdown-list only contains the map layers that are added in the displayed layers panel. In this example, the “crime incidents (points)” were chosen, and the ID is the selected variable that means that all the events will be shown.

  The third element to be selected is the time frame. Three filters can be chosen by selecting the temporality from three dropdown-lists. The first step is to select the first filter; in the example shown in Figure 6.8 “year” was chosen. Once the filter is selected, the horizontal axis in the histogram will be changed to years. Then, by moving the slider the user can choose the range of years he/she is interested in, and then press the “set” button. Once set, the second dropdown-list is activated and the same steps have to be done again. In this example, “months” were chosen and the slider was set to the period from April to May.

  The next option is to view only the selected point or all the records. In this case “All” was selected. Therefore, in the map the events that happened in 2017 from April to May are shown in red and the grey points show the rest of the incidents that happened in 2017.

  Figure 6.9 shows the sequence animation option panel. The left options work the same as in the static view option. What is different is the right part in which a media control panel is shown instead of the histogram. In this case, a sequence of static maps depicting the disturbance and vandalism incidents that happened from April to August 2017 and 2018 are selected (this is an example of what it could be done, the original dataset did not contain this attributes). As the animation plays, the month that is currently shown is coloured in turquoise in the slider. In this case, a third filter, for instance, the hour, could be added as the month was “set” from April to August.

  Additionally, in both options, static view and sequence animation, it is possible to display one or two map layers in the same view. For this, the checkbox “Add another layer” has to be checked and the dropdown-lists will be “cleaned” so the user can make a new selection. This makes possible the comparison in time of the variability of two datasets

The conceptual design presented here shows some examples of the potential tools that could be included in a GVA directed to explore crime perception data. The next step would be to implement the prototype, make changes if needed, and pass to the second phase, which are the interaction and usability studies.


Figure 6.8 Example of the “static view” option in the timeline panel. In the map, the red points are the crime incidents that happened in 2017 between April and June. The grey points are the other events recorded in 2017.

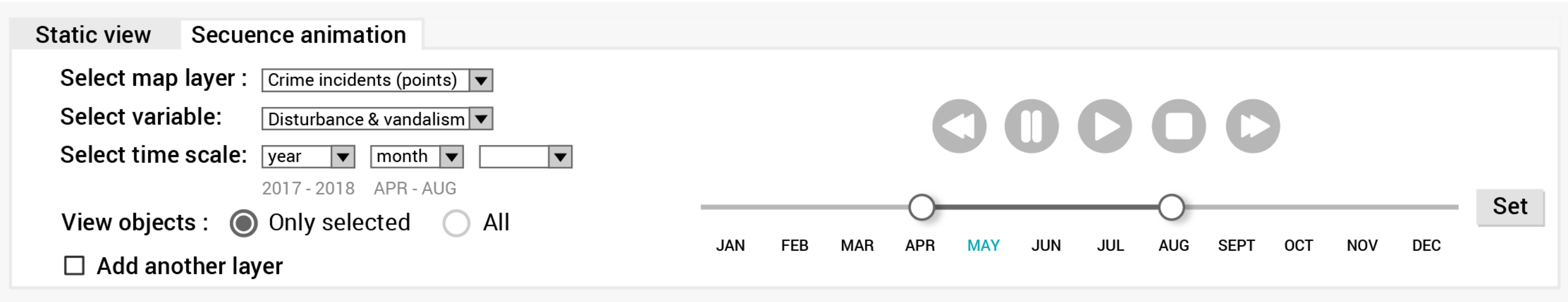



Figure 6.9 Sequence animation option from the timeline panel.

## 6.5 Conclusions

A Geovisual Analytics (GVA) environment is an interactive system that holds elements and states in the shape of spatial datasets and their attributes. Its functionalities are intended to support the user in the exploration, analysis and understanding of the complex spatial relationships between components. The design of a GVA interface using a user-centered approach starts with the presentation of a low-fidelity prototype that shows the structure and functionalities of the interface. The step preceding the actual prototype design is to perform a requirement analysis to define the profile of the users' group and their needs.

The presented prototype has the aim to support police agencies to develop efficient strategies to narrow the crime perception gap by getting to know the spatial context of the problem. Due to the limits of this research, the requirement analysis performed only describes the aim and users of the interface.

The main goal of this proposal was to show potential tools and functionalities that a GVA for crime perception data could include. There are no similar interfaces developed; the existing GVAs are focused on the visualization and analysis of crime incidents, such as GeoVISTA CrimeViz (Roth et al., 2015) and VIS-STAMP (Guo & Wu, 2013).

The tools are intended to answer the identified questions (see section 6.3) that during the analysis process and interpretation of the results arose. They were designed based on existing spatial analysis methods and tools. The visualization and map composer tools enable the user to create map layers by selecting the attributes he/she is interested in to explore. The design of the functions is simple, as the users are non-technical experts in the use of spatial data.

The next step of a subsequent study would be to define precisely the users of the interface and involve a small user group to evaluate the prototype, make the respective changes and develop a high–fidelity prototype to perform the interaction and usability studies.

# Chapter 7

## Conclusions and final remarks

This chapter is divided into five sections. In the first one (section 7.1), a summary of the three stages of the research process is presented. In section 7.2, the five research questions presented in Chapter 1 will be answered in summary. Section 7.3 comprises the general conclusion of this research. In the following section (7.4) additional observations are discussed, and to conclude with final recommendations (section 7.5) for future researches are described.

## 7.1 Summary

Sketch mapping is a method frequently used in crime studies to gather perceptual information. An analysis and an integral interpretation of sketch maps can be done by incorporating the use of GIS, spatial analysis, and statistics. In the case-study, a methodology that enabled a quantitative exploration of structured sketch maps to analyse the perception of crime was implemented. This methodology consisted of three stages, an exploratory modelling, a spatial arrangement outline of the perception accuracy, and the development of a GeoVisual Analytics (GVA) environment directed to support the exploration and understanding of crime perception. A summary of the three stages is presented below.

- Exploratory modelling

In crime perception studies, the use of structured sketch maps is frequently directed to define the location of the perceived safe and unsafe places. In this research, the analysis was not only focused on the distribution of the spatial geometry of the sketched polygons but also on data extraction and its analysis.

The data extraction process consisted of segmenting the sketched polygons into small cells to capture the spatial heterogeneity of the covered area. Each cell was treated as a data sample from which a set of variables were measured. These variables were analysed using Exploratory Data Analysis (EDA), and its results were confirmed by the outputs of a Machine Learning model using binary logistic regression.

The regression approach was chosen to explore the variables that influence the perception of unsafe areas. During the modelling process, different variants of data extraction methods and different covariates were tested to select those that better explain the perception of crime. The final model was built with the most significant covariates, according to the p-values and the $R^2$ value. The interpretation of the resultant coefficients (see subchapter 4.3) indicated that the likelihood of a person perceiving an area as unsafe increases as he/she moves away from his/her neighbourhood and daily route, or move away from a crime hotspot or a high crime intensity area.

Moreover, the results from the EDA and the regression pointed out the existence of a crime perception gap. However, as this type of analysis did not indicate the location of the gap, the following stage consisted of determining the spatial distribution of the perception accuracy.

- Spatial arrangement outline

This stage comprised the identification of the spatial arrangement per block of the crime perception accuracy. The first step was to label the blocks as safe or unsafe according to the overlapping sketched polygons (perceived classification) and according to the number of incidents per block that defined the crime hotspots (reference classification). Both values were compared to define whether the perception of the participants was accurate or inaccurate.

The analysis of the comparison showed that out of 1,706 blocks that overlapped with the sketched polygons, 39% were accurately classified (perceived classification=reference classification); 24% of the perceived safe areas are actually unsafe. Meanwhile, 88% of the perceived unsafe areas are actually safe (see Table 5.3).

A typology and a scale of accuracy were proposed to characterize people's perception. Four classes of perception accuracy were defined by the comparison of both classifications (perceived vs. reference) (see Figure 2.4). The scale of accuracy was defined by three levels of "agreement" (low, medium and high) based on the number of participants who accurately or inaccurately classified the block (see Table 5.2).

This stage also consisted of estimating the Bivariate Local Moran's I statistic. This is a local spatial autocorrelation method that compares the values from two variables in adjacent locations. In the case-study, the blocks that were inaccurately perceived due to the high or low criminal activity in neighbouring areas were identified (see Figure 5.7).

The resulting maps as presented in Chapter 5 are some of the possible layers of information that can be further explored to uncover the complex relations between elements to ease the analytical reasoning about the perception of crime. Based on this assumption, a GVA interface that integrates these types of information layers, additional context data and analytical tools has been proposed. It is intended to be directed to police agencies which are the ones in charge of developing strategies to reduce the fear of crime and reassure people about the safety state.

- Geovisualization development

The purpose of the proposed GVA environment is to gather the previous information and make it available to police agencies in an easily manageable interface that enables its visualization and analysis.

The tools included in the GVA interface prototype as proposed in section 6.3 are intended to support the police agencies in the detection and understanding of spatial relations, patterns, or changes of perception of crime. A holistic knowledge with a system approach would result in the development of efficient and effective strategies to narrow down the crime perception gap.

The overall functionality of the proposed prototype is founded on GIS interfaces. The tools, such as map comparison, selection by attribute and location, or more complex types of selection tools such as based on intersection or linked selection, were designed based on the spatial analysis method and data extraction performed in the exploratory modelling and the spatial arrangement outline stage.

The principle of the presented prototype is that the user can create map layers for further, analysis. The creation of the map layers is straightforward and it only requires the selection of the variables(s) to be shown. The design of the tools avoids technical vocabulary for easy use.

Performing an exploratory modelling and a quantitative analysis set the contextual background that allowed the design of a GVA prototype that potentially could support the analytical reasoning process or crime perception.

By knowing beforehand which variables are relevant in a given problem, the analysis process will be speeded up by reducing the complexity as only the correlated variables will be explored.

The methodology presented addressed the three specific objectives set to reach the general objective of this thesis. Along the research process five questions were answered related to each objective; the answers to them are given in the next section.

## 7.2 Answers to the research questions

The answers to the five research questions are presented below:

*1.1 What is the relationship between the people's daily activity spaces (neighbourhood and daily routes) and the location of the areas they perceive as unsafe?*

Based on the EDA, it was concluded that, in general, the participants identified safe areas closer to their neighbourhoods. Meanwhile, the daily routes' analysis generated that the participants identified unsafe areas further away from their daily trajectory. The interpretation of the resulting covariates from the regression model confirmed the EDA findings. The regression model's output interpretation indicated that the likelihood of a person perceiving an area as unsafe increases while moving away from his or her neighbourhood or daily routes.

*1.2 What is the relationship between the location of the crime incidents and the perceived unsafe areas?*

The crime incidents were analysed with two variables, the hotspots and the high crime intensity areas (HCIAs).

Two variants of hotpots were analysed, one defined by the number of incidents per block (Figure 4.3) and the other by the density of crimes per block (Figure 4.4). The output maps show a different spatial arrangement between both types. The hotspot locations by total number of incidents are scattered in comparison with the hotspots by density, which are mainly clustered in the city centre. This distinction in the spatial distribution can explain the results of the EDA. According to the EDA the distances between perceived unsafe areas and a hotspot by total count were shorter than by density.

Moreover, the EDA showed that, for both types of hotspots, for more than 50% of the sketched safe polygons, half of their area was within a hotspot. This means that people demonstrated an inaccurate perception of safe areas.

For the HCIAs, based on the EDA, it was determined that people perceived a higher percentage of safe areas around the HCIAs.

In the regression modelling process, both types of hotspots were explored; both of them were significant to explain the perception of unsafe areas. Nevertheless, the distance measured to the hotspot by the total count of incidents per block explained a higher percentage of the likelihood variability of perceiving an area as unsafe. Therefore, this type of hotspot was selected over the other one. Additionally, it was tested whether exploring the clusters and outliers hotspots as two separate variables was more significant that grouping both and exploring them as a single variable. Based on the $R^2$ value, two covariates were more significant than one.

The interpretation of the output coefficients shows that the likelihood of perceiving an area as unsafe increases while moving away from the crime hotspots and the HCIAs, which suggests that participants had an inaccurate perception of unsafe areas. Thus there is a crime perception gap.

*2.1 How to measure the accuracy of people's crime perception?*

A perception typology has been proposed to determine the type of accuracy: it consists of two classes of accurate perception and two classes of inaccurate perception of safe and unsafe areas each. The classification method used in this research comprises three steps. First, select the analysis unit and aggregate or transform the perception data and the crime data into the chosen spatial unit. Then the conditions to label a unit as perceived safe/unsafe and reference safe/unsafe must be set. Each unit will then have two additional attributes: the perceived and the reference status of safety. Both values are compared to define whether the perception was accurate or inaccurate. If both values concord as "safe" then the unit is labelled as AS (i.e accurate perception of safe areas), but if they are "unsafe" then they are tagged as AU (i.e. accurate perception of unsafe areas). The units for which the values differ will be labelled as IS (i.e. inaccurate perception of safe areas, if the perceived class is "safe" whereas the reference one is "unsafe"), or if the perceived class is "unsafe" but the reference class is "safe" the unit will be tagged as IU (i.e. inaccurate perception of unsafe areas) (see Figure 2.4).

For the case-study, the city blocks were the unit of analysis. The "perceived classification" was defined by the number of participants who sketched a polygon (safe/unsafe) over each block; the type with the highest percentage was set as the perceived class. In the case of the "reference classification," the condition was that if the block was distinguished as a crime hotspot (by the count of crime incidents) it was labelled as unsafe, otherwise as safe.

The level of accuracy was defined by the number of participants who classified a block, i.e., who sketched a polygon over a block. If the block was accurately classified (reference classification = perceived classification), the level of accuracy was defined by the percentage of participants who classified the block correctly by the total number of participants who classified the block. If the block was inaccurately classified (reference classification <> perceived classification), the accuracy level was defined by the percentage of participants who classified the block incorrectly by the total number of participants who classified the block.

Based on the percentage values, an ordinal classification was proposed to determine three levels of accuracy: low (>50% - 65%), medium (>65% - 85%) and high (>85% - 100%) (Table 5.2). As this scale is based on the number of people who sketched over a block, although the level is 100% it does not necessarily mean a high value because it can be the case that over that block only one person sketched, then whether accurate or not the level will be 100. Therefore, the blocks could be filtered by setting a minimum value of a number of participants who sketched on a block, so that the level of accuracy can be more meaningful.

*2.2 How can the location of inaccurately perceived unsafe areas be explained by the spatial distribution of another explanatory variable?*

People tend to characterize a feature or place by association with the attributes of nearby elements. The relation between the locations of two related variables can explain the location of the first one.

The bivariate local Moran's I is a local spatial autocorrelation measurement which relates the value of a first variable in a target place and the average value of a second variable in the neighbouring areas compared with the average value of a conditional permutation. The aim is to define which places have a higher or lower average value in the surrounding location compared with the expected value result from the conditional permutation.

In the case-study, the bivariate local Moran's index was measured to detect where people have an inaccurate perception of safeness as response to spatial association. For instance, the places that are labelled as IU and which are surrounded by actual unsafe place, could be explained by the adjacency with unsafe places. The spatial arrangement of crime perception can be explained by individual social, demographic and environmental factors, but also by the spatial relationship between them. This type of analysis shows how the distribution of one feature can explain the distribution of a second one, which confirms the importance of spatial studies with a system approach.

*3.1 Which tools and representations could be integrated in a GeoVisual Analytics interface to explore and analyse crime perception?*

The heuristic selection of the presented tools was based on the methods used in the data extraction and analysis performed in the exploratory modelling and spatial arrangement outline (see Chapter 4 & 5).

Exploring and analysing spatiotemporal perception data requires an interface that enables the visualization of multivariate datasets. To work with data that include different attributes and spatiotemporal information increases the complexity of understanding the relationships among the datasets. Therefore, it is required that the tool enables the selection and filtering of the data of interest.

The idea of the tools presented is that the user can create map layers by quickly selecting the variables and applying special and temporal filtering. After the user has created the map layer, the analysis tools can be executed.

Three maps layers can be created: univariate maps, index maps (in which a variable is constructed by the combination of different attributes) and a bivariate-adjacency map (bivariate local Moran's statistic map). Temporal attributes and locations can filter these map layers. Some visual analysis can be performed by maps comparison to detect changes and selecting features within a polygon. Also by linked selection that enables the user to select features from two different map layers that have an attribute in common to explore the spatial relation between features. Additionally, a timeline panel was included for visualizing static or sequential animation views for change and pattern detection.

## 7.3 General conclusion

The main conclusion of this research is the relevance of an integral analysis of sketch maps in the study of perception. The methodology performed showed some approaches of data extracting, exploration and quantitative analysis that enabled an understanding of the perception of crime. The segmentation into cells of the sketched polygons gave good results for the analysis of the maps. Working with multiple data samples instead of one allowed capturing the diversity of the features and spatial characteristics covered by the polygons. By this segmentation, the polygons were analysed in the same way regardless of their size or geometry.

A statistical analysis of the data extracted from the cells enabled an objective exploration of the structured sketch maps. The binary logistic regression is a useful method for the analysis of perception where there are only two possible answers; "good or bad", "true or false", "safe or unsafe". Narrowing the answers to binary options can reduce the level of ambiguity that perception has *per se*.

This research gave a general idea of how incorporating quantitative and spatial analysis methods for the study of spatial perception from structured sketch maps can result in a more complete and objective interpretation.

## 7.4 Discussion

The analysis performed in the first two stages of the research required different data aggregation and transformation processes, which can bring a certain level of uncertainty into the results.

In this research, the crime events were not treated as independent points because it was not meaningful for the type of analysis performed; therefore, they were aggregated within blocks. Two main concerns come with this type of generalization. The first one is how to deal with a point located on the boundary shared by two polygons. When performing a spatial join operation in a GIS between a point and polygon vector layer this problem would be solved by counting the point twice, one for each polygon.

The second issue is the effect of the Modifiable Areal Unit Problem (MAUP) that deals with the shape and size of the units of aggregation. This pitfall affects the analysis results due to the size differences of the blocks, especially for the hotspots analysis as the number of events per block defined them. Additionally, the aggregation of points depends on the block digitalization process, on which the geometry of the blocks depends.

Another point of discussion is the quality of the geocoding process. One issue found was the accuracy of the geocoding results, as for some records the points were located in the same pair of coordinates for similar addresses. Table 7.1 shows six addresses, 2 of them without an address number and 4 with it. In spite of this difference, the six points were placed in the same location (Figure 7.1). The difficulty about this problem is that it is a problem related to the geocoding tool, because although four addresses are complete they were not located accurately.

| Address | Number of records |
|---|---|
| Szent Imre tér | 314 |
| Szent imre tér | 17 |
| Szent Imre tér 3 | 6 |
| Szent Imre tér 6 | 4 |
| Szent Imre tér 10 | 3 |
| Szent Imre tér 2 | 1 |

**Table 7.1** Data example of addresses that were geocoded at the same location.

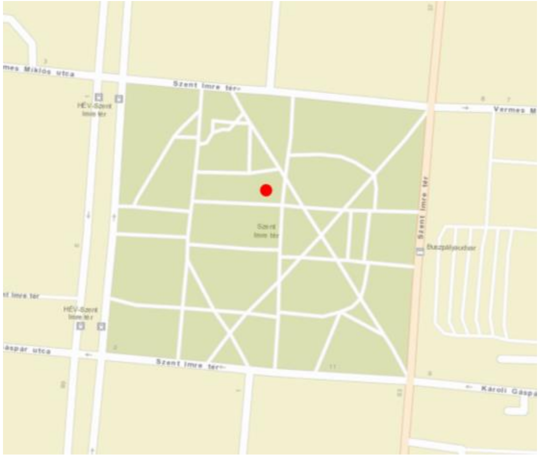



Figure 7.1 Geocoded location (red point) of the data example shown in Table 7.1.

## 7.5 Recommendations

This thesis intended to contribute to the spatial studies of perception by the data extraction and analysis of sketch maps. During the research process some difficulties arose that lead to the following recommendations for future research:

- The perception data collection by sketch maps must include a questionnaire or a think-aloud process that can provide more information to the interpretation of the map. Although, the analysis of the data extracted can reveal valuable information, having an additional context of the participants' cognitive map can add more variables to explore that would lead to a better characterization of the people's perception.
- The logistic regression model can be improved by exploring additional contextual variables, for instance the land use or the average income. Although there are theories that explain the factors that sway the perception of crime, each city has different social dynamics where those factors may not have the same impact. Therefore, it is necessary to explore them to get a more precise overview of the context.
- Due to the time limits of this thesis research, the requirement analysis presented for developing the GVA prototype was based on a literature review only. The requirement analysis should be improved to enhance the GVA prototype presented, which could be used as a reference to develop a high-fidelity prototype that can go through interaction and usability studies. It should then be tested whether the proposed interface could be implemented as a tool in a decision-support system for developing strategies directed towards narrowing down the crime perception gap.

# References


Anselin, L. (2019). *Local Spatial Autocorrelation*. https://geodacenter.github.io/workbook/6b_local_adv/lab6b.html [Retrieved July 2019]

Ardanaz, M., Corbacho, A., Ibarraran, P., & Ruiz-Vega, M. (2013). Mind the gap: Bridging perception and reality with crime information. *Unpublished Research Report. Washington, DC, United States: Inter-American Development Bank*.

Atlas of Urban Expansion (2016). *Budapest*. Retrieved from [Retrieved July 2019] http://www.atlasofurbanexpansion.org/cities/view/Budapest

Beck, U. (1992). *Risk society: towards a new modernity*. London, Sage.

Bennett, T. (1991). The effectiveness of a police-initiated fear-reducing strategy. *The British Journal of Criminology*, 31(1), 1-14.

Blades, M. (1990). The reliability of data collected from sketch maps. *Journal of Environmental Psychology*, *10*(4), 327-339.

Brantingham, P. J. & Brantingham, P. L. (1981). Notes on the geometry of crime. In P. J. Brantingham & P. L. Brantingham (Eds.), *Environmental Criminology* (pp. 27-54). Beverley Hill, CA: Sage.

Brantingham, P. L., & Brantingham, P. J. (1995). Criminality of place: Crime generators and crime attractors. *European Journal on Criminal Policy and Research*, 3, 5-26.

Carvalho, I., & Lewis, D. A. (2003). Beyond community: Reactions to crime and disorder among inner-city residents. *Criminology*, *41*(3), 779-812.

Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security journal*, *21*(1-2), 4-28.

Clark, J. (2003). Fear in fear-of-crime. *Psychiatry, Psychology and Law*, *10*(2), 267-282.

Cochran, J. K., Bromley, M. L., & Branch, K. A. (2000). Victimization and fear of crime in an entertainment district crime “hot spot:” A test of structural-choice theory. *American Journal of Criminal Justice*, *24*(2), 189-201.

Cook, K. A. & Thomas, J. J. (2005) *Illuminating the path: the research and development agenda for visual analytics*. IEEE Computer Society.

Crank, J. P., Giacomazzi, A., & Heck, C. (2003). Fear of crime in a nonurban setting. *Journal of criminal justice*, *31*(3), 249-263.

Cordner, G. (2010). *Reducing Fear of Crime Strategies for Police.* U.S. Department of Justice Office of Community Oriented Policing Services.

Covington, J., & Taylor, R. B. (1991). Fear of crime in urban residential neighborhoods: Implications of between-and within-neighborhood sources for current models. *The Sociological Quarterly*, *32*(2), 231-249.

Cozens, P. M. (2002). Sustainable urban development and crime prevention through environmental design for the British city. Towards an effective urban environmentalism for the 21st century. *Cities*, *19*(2), 129-137.

Curtis, J. W. (2012). Integrating sketch maps with GIS to explore fear of crime in the urban environment: A review of the past and prospects for the future. *Cartography and Geographic Information Science*, *39*(4), 175-186.

Curtis, J. W., Shiau, E., Lowery, B., Sloane, D., Hennigan, K., & Curtis, A. (2014). The prospects and problems of integrating sketch maps with geographic information systems to understand environmental perception: A case study of mapping youth fear in Los Angeles gang neighborhoods. *Environment and Planning B: Planning and Design*, *41*(2), 251-271.

Doran, B. J., & Burgess, M. B. (2011). *Putting fear of crime on the map: Investigating perceptions of crime using geographic information systems*. Springer Science & Business Media.

Downs, R. M., & Stea, D. (Eds.). (1973). *Image and environment: Cognitive mapping and spatial behavior*. Transaction Publishers.

Downs, R. M. & Stea, D. (1977). *Maps in minds - Reflections on cognitive mapping*. New York: Harper & Row.

Duffy, B., Wake, R., Burrows, T. and & Bremer, P. (2008). Closing the gaps – crime and public perceptions. *International Review of Law Computers & Technology*, *22* (1-2), pp. 17-44.

ESRI (n.d.) *How Hot Spot Analysis: Getis-Ord Gi* (Spatial Statistics) works.* Retrieved from http://resources.esri.com/help/9.3/arcgisengine/java/gp_toolref/spatial_statistics_tools/how_hot_spot_analysis_colon_getis_ord_gi_star_spatial_statistics_works.htm [Retrieved July 2019]

Fisher, B., & Nasar, J. L. (1995). Fear spots in relation to microlevel physical cues: Exploring the overlooked. *Journal of Research in Crime and Delinquency*, *32*(2), 214-239.

Foster, S., Giles-Corti, B., & Knuiman, M. (2010). Neighbourhood design and fear of crime: A social-ecological examination of the correlates of residents' fear in new suburban housing developments. *Health & place*, *16*(6), 1156-1165.

Foster, S., Knuiman, M., Wood, L., & Giles-Corti, B. (2013). Suburban neighbourhood design: Associations with fear of crime versus perceived crime risk. *Journal of Environmental Psychology*, *36*, 112-117.

Friendly, M. (2007). A.-M. Guerry's" Moral Statistics of France": Challenges for Multivariable Spatial Analysis. *Statistical Science*, 368-399.

Fuhrmann, S., Huynh, N. T., & Scholz, R. (2013). Comparing fear of crime and crime statistics on a university campus. In *Crime Modeling and Mapping Using Geospatial Technologies*(pp. 319-337). Springer, Dordrecht.

Furstenberg, F. F. (1971). Public reaction to crime in the streets. *The American Scholar*, 601-610.

Golledge, R. G. (1997). *Spatial behavior: A geographic perspective*. New York: Guilford Press.

Grabosky, P. N. (1995). Fear of crime, and fear reduction strategies. *Current issues in criminal justice*, 7(1), 7-19.

Groff, E. R., Kearley, B., Fogg, H., Beatty, P., Couture, H., & Wartell, J. (2005). A randomized experimental study of sharing crime data with citizens: Do maps produce more fear?. *Journal of Experimental Criminology*, *1*(1), 87-115.

Guo, D., & Wu, J. (2013). Understanding spatiotemporal patterns of multiple crime types with a geovisual analytics approach. In *Crime modeling and mapping using geospatial technologies* (pp. 367-385). Springer, Dordrecht.

Hanson, R. F., Smith, D. W., Kilpatrick, D. G., & Freedy, J. R. (2000). Crime-Related fears and demographic diversity in Los Angeles county after the 1992 civil disturbances. *Journal of Community Psychology*, *28*(6), 607-623.

Haggett, P. (1979) *Geography: a modern synthesis.* Harper & Row, New York.

Huggett, R. J. (1980). *Systems analysis in Geography*. Oxford University Press.

Innes, M., Fielding, N., & Langan, S. (2002). Signal crimes and control signals: towards an evidence-based conceptual framework for reassurance policing. *A Report for Surrey Police*.

Jackson, J., & Gouseti, I. (2014). Fear of crime. *The encyclopedia of theoretical criminology*, 1594-1602.

Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In *Information visualization* (pp. 154-175). Springer, Berlin, Heidelberg.

Kemshall, H. (1997). Sleep safely: crime risks may be smaller than you think. *Social Policy & Administration*, *31*(3), 247-259.

Kohlhammer, J., May, T., & Hoffmann, M. (2009). Visual analytics for the strategic decision making process. In *GeoSpatial Visual Analytics*, pp. 299-310. Springer, Dordrecht.

Kohm, S. A. (2009). Spatial dimensions of fear in a high-crime community: Fear of crime or fear of disorder? *Canadian Journal of Criminology and Criminal Justice*, 51(1), 1-30.

Kuo, F. E., & Sullivan, W. C. (2001). Environment and crime in the inner city: Does vegetation reduce crime? *Environment and behavior*, *33*(3), 343-367.

Kveladze, I., Kraak, M.J., van Elzakker, C. (2017) The Usability of a GeoVisual Analytics Environment for the Exploration and Analysis of Different Datasets. In: Peterson M. (eds) *Advances in Cartography and GIScience. ICACI 2017. Lecture Notes in Geoinformation and Cartography.* Springer, Cham.

Lane, J., & Meeker, J. W. (2003). Ethnicity, information sources, and fear of crime. *Deviant Behavior*, *24*(1), 1-26.

Lewis, D. A. & Maxfield, M. G. (1980). Fear in the neighborhoods: An investigation of the impact of crime. *Journal of Research in Crime and Delinquency*, 17(2), pp. 160–189.

Lianos, M., & Douglas, M. (2000). Dangerization and the end of deviance: the institutional environment. *British Journal of Criminology*, *40*(2), 261-278.

Liska, A. E., Sanchirico, A., & Reed, M. D. (1988). Fear of crime and constrained behavior specifying and estimating a reciprocal effects model. *Social Forces*, *66*(3), 827-837.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological bulletin*, *127*(2), 267-286.

Lopez, N., & Lukinbeal, C. (2010). Comparing Police and Residents' Perceptions of Crime in a Phoenix Neighborhood using Mental Maps in GIS. *Yearbook of the Association of Pacific Coast Geographers*, 33-55.

Lora, E. (2016). The Distance between Perception and Reality in the Social Domains of Life. In *Handbook of Happiness Research in Latin America*, pp. 531-555. Springer, Dordrecht.

Lynch, K. (1960) *The Image of the City*. Cambridge: The MIT Press.

Maguire, M., & Bevan, N. (2002). User requirements analysis. In *IFIP World Computer Congress, TC 13,* pp. 133-148. Springer, Boston, MA.

Markowitz, F. E., Bellair, P. E., Liska, A. E., & Liu, J. (2001). Extending social disorganization theory: Modeling the relationships between cohesion, disorder, and fear. *Criminology*, *39*(2), 293-319.

Matei, S., Ball-Rokeach, S. J., & Qiu, J. L. (2001). Fear and misperception of Los Angeles urban space: A spatial-statistical study of communication-shaped mental maps. *Communication Research*, *28*(4), 429-463.

Mawby, R. I., Brunt, P., & Hambly, Z. (2000). Fear of crime among British holidaymakers. *British Journal of Criminology*, *40*(3), 468-479.

Mello, R. F., & Ponti, M. A. (2018). *Machine Learning: A Practical Approach on the Statistical Learning Theory*. Springer.

Millie, A., & Herrington, V. (2005). Bridging the gap: understanding reassurance policing. *The Howard Journal of Criminal Justice*, *44*(1), 41-56.

Mohan, J., Twigg, L., & Taylor, J. (2011). Mind the double gap: Using multivariate multilevel modelling to investigate public perceptions of crime trends. *The British Journal of Criminology*, *51*(6), 1035-1053.

Nasar, J. L., Fisher, B., & Grannis, M. (1993). Proximate physical cues to fear of crime. *Landscape and urban planning*, *26*(1-4), 161-178.

O'Neill, E., Brennan, M., Brereton, F., & Shahumyan, H. (2015). Exploring a spatial statistical approach to quantify flood risk perception using cognitive maps. *Natural Hazards*, *76*(3), 1573-1601.

Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. *Geographical analysis*, *27*(4), 286-306.

Pfeiffer, C., Windzio, M., & Kleimann, M. (2005). Media use and its impacts on crime perception, sentencing attitudes and crime policy. *European journal of criminology*, *2*(3), 259-285.

Pocock, D. C. (1979). The contribution of mental maps in perception studies. *Geography*, 64(4), 279-287.

Preece, J., Sharp, H., & Rogers, Y. (2002) *Interactive design: Beyond human-computer interaction.* John Wiley & Sons.

Raisz, E. (1938). *General Cartography*. McGraw-Hill Book Company.

Ratcliffe, J. H. (2004). Geocoding crime and a first estimate of a minimum acceptable hit rate. *International Journal of Geographical Information Science*, *18*(1), 61-72.

Roth, R. E., Ross, K. S., Finch, B. G., Luo, W. and MacEachren, A. M., 2010. A user-centered approach for designing and developing spatiotemporal crime analysis tools. Proceedings of GIScience 2010. Zurich, Switzerland.

Roth, R., Ross, K., & MacEachren, A. (2015). User-centered design for interactive maps: A case study in crime analysis. *ISPRS International Journal of Geo-Information*, *4*(1), 262-301.

Sloan, N., Doran, B., Markham, F., & Pammer, K. (2016). Does base map size and imagery matter in sketch mapping?. *Applied geography*, *71*, 24-31.

Solymosi, R., Bowers, K., & Fujiyama, T. (2015). Mapping fear of crime as a context-dependent everyday experience that varies in space and time. *Legal and Criminological Psychology*, *20*(2), 193-211.

Spicer, V., Song, J., & Brantingham, P. (2014). Bridging the perceptual gap: variations in crime perception of businesses at the neighborhood level. *Security Informatics*, 3(1), 14.

Sun, I. Y., Triplett, R., & Gainey, R. R. (2004). Neighborhood characteristics and crime: A test of Sampson and Groves' model of social disorganization. *W. Criminology Rev.*, *5*, 1.

Taylor, R. B., & Covington, J. (1993). Community structural change and fear of crime. *Social problems*, *40*(3), 374-397.

Taylor, R. B., & Hale, M. (1986). Testing alternative models of fear of crime. *J. Crim. L. & Criminology*, *77*, 151.

Tulloch, M. (2000). The meaning of age differences in the fear of crime. *British journal of criminology*, *40*(3), 451-467.

Warr, M. (2000). Fear of crime in the United States: Avenues for research and policy. *Criminal justice*, *4*(4), 451-489.

Weisburd, D., & Britt, C. (2007). *Statistics in criminal justice*. Springer Science & Business Media.

Weisburd, D. & McEwen, T. (1998). *Crime mapping and crime prevention*. New York: Criminal Justice Press.

Weisburd, D., & Piquero, A. R. (2008). How well do criminologists explain crime? Statistical modeling in published studies. *Crime and Justice*, *37*(1), 453-502.

Wood, L. J. (1970). Perception studies in geography. *Transactions of the Institute of British Geographers*, 50, 129-142.