UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

JOSÉ FLORENCIO DE QUEIROZ NETO

A VISUAL ANALYTICS APPROACH FOR GEOCODED CRIME DATA

FORTALEZA
2020

JOSÉ FLORENCIO DE QUEIROZ NETO

A VISUAL ANALYTICS APPROACH FOR GEOCODED CRIME DATA

Thesis submitted to the Programa de Pós-Graduação em Ciência da Computação of the Centro de Ciências at the Universidade Federal do Ceará, in partial fulfillment of the requirements for the degree of Doctor in Computer Science. Concentration Area: Computer Graphics

Supervisor: Profa. Dra. Emanuele Marques dos Santos

Co-Supervisor: Prof. Dr. Creto Augusto Vidal

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

Q44v Queiroz Neto, José Florencio de.
A Visual Analytics Approach For Geocoded Crime Data / José Florencio de Queiroz Neto. – 2020.
100 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Ciência da Computação , Fortaleza, 2020.
Orientação: Profa. Dra. Emanuele Marques dos Santos.
Coorientação: Prof. Dr. Creto Augusto Vidal.

1. Visualização analítica. 2. Visualização de dados. 3. Computação gráfica. I. Título.
CDD 005

JOSÉ FLORENCIO DE QUEIROZ NETO

A VISUAL ANALYTICS APPROACH FOR GEOCODED CRIME DATA

Thesis submitted to the Programa de Pós-Graduação em Ciência da Computação of the Centro de Ciências at the Universidade Federal do Ceará, in partial fulfillment of the requirements for the degree of Doctor in Computer Science. Concentration Area: Computer Graphics

Aprovada em:

EXAMINATION COMMITTEE

Profa. Dra. Emanuele Marques dos Santos (Supervisor)
Universidade Federal do Ceará (UFC)

Prof. Dr. Creto Augusto Vidal (Co-Supervisor)
Universidade Federal do Ceará (UFC)

Prof. Dr. José Antônio Fernandes de Macêdo
Universidade Federal do Ceará (UFC)

Profa. Dra. Carla Maria Dal Sasso Freitas
Universidade Federal do Rio Grande do Sul (UFRGS)

Dr. Wellington Clay Porcino Silva
Diretoria de Gestão e Integração de Informações do
Ministério da Justiça e Segurança Pública

I dedicate this thesis to the people who dreamed of being a scientist.

ACKNOWLEDGEMENTS

To my wife, Iana, and my daughter Sofia. For their love, which gives me the energy I need to keep my journey.

To Prof. Dr. Emanuele Marques dos Santos. For the firm guidance and constant encouragement.

To Prof. Dr. Creto Augusto Vidal. For the precise analyses and feedbacks.

To Professor Dr. David S. Ebert. For the lots of ideas.

To all my friends of the CRAB Lab, the VACCINE Lab, the Insight Lab, the Public Safety and Social Security Secretariat of Ceará, the Lafayette Police Department, the Purdue Police Department and the Brazilian Ministry of Justice and Public Safety. For their support.

To the funding institutions CAPES, CNPQ and FUNCAP. For their support.

"We use our reason for improving the sciences. Whereas we ought to use the sciences for improving our reason."

(Antoine Arnauld [1662])

RESUMO

Nos últimos anos, a violência aumentou consideravelmente no mundo. No Ceará, a taxa de homicídios passou de 16 por 100.000 habitantes em 2000 para 37 por 100.000 habitantes em 2016. Com a popularização de bancos de dados espaciais e Sistemas de Informação Geográfica (SIG), os departamentos de polícia começaram a criar vários tipos de mapas de crimes, gerados com diferentes técnicas, para analisar e combater o crime. Um dos tipos de mapas de crimes é o mapa de hotspots, que ajuda na identificação de áreas de alto risco e na alocação de recursos com mais eficiência. A análise de dados de crimes geralmente é uma operação complexa e indicada para os sistemas Visual Analytics (VA), que são sistemas que visam aumentar o raciocínio analítico humano por meio de interfaces visuais e interativas. A necessidade de interatividade exige que os sistemas VA incluam alto desempenho como uma das principais condições. Nesta tese, propomos o MSKDE - Marching Squares Kernel Density Estimation, uma técnica para gerar mapas de hotspot rápidos e precisos. Descrevemos o método e demonstramos suas qualidades superiores por meio de uma comparação cuidadosa com a Estimativa de Densidade de Kernel (KDE), amplamente usada para gerar mapas de densidade. Outra contribuição desta tese visa ajudar os departamentos de polícia em suas atividades de planejamento. Profissionais e pesquisadores concordam que rastrear o crime ao longo do tempo e identificar seus padrões geográficos são vitais para o planejamento eficiente dos recursos. Para ajudar a realizar essas atividades, freqüentemente, os departamentos de polícia têm acesso a sistemas muito complicados e excessivamente técnicos. Colaboramos com especialistas em departamentos de polícia do Brasil e dos Estados Unidos para identificar e caracterizar cinco tarefas inerentes à atividade de rastreamento de crimes e planejamento de alocação de recursos. Todas as tarefas estão relacionadas à análise de hotspots, um dos métodos mais importantes para combater o crime. Para facilitar a execução das tarefas, propusemos o SHOC, The One-Shot Comparison Tool, uma técnica que permite a comparação espacial imediata das superfícies de densidade de crimes. Incluímos o SHOC em um sistema VA, o CrimeWatcher, que permite aos usuários realizar operações de filtragem e visualizar mapas e dados. O CrimeWatcher se destaca pela simplicidade e permite que usuários, mesmo sem conhecimento técnico, executem as tarefas, anotem, salvem e compartilhem análises. Também demonstramos que o CrimeWatcher e o SHOC suportam efetivamente a conclusão de tarefas em dois estudos de caso reais.

Palavras-chave: Visualização analítica. Visualização de dados. Computação gráfica.

ABSTRACT

In recent years, violence has increased considerably in the world. In Ceará, a Brazilian state member, the homicide rate went from 16 per 100,000 inhabitants in 2000 to 37 per 100,000 inhabitants in 2016. With the popularization of spatial databases and Geographic Information Systems (GIS), police departments worldwide started to create various types of crime maps, generated with different techniques, to analyze and fight crime. One of the types of crime maps is the hotspot map, which helps decision-makers to identify high-risk areas and allocate resources more efficiently. The analysis of crime data is usually a complex operation, and a target of Visual Analytics (VA) systems, which are systems that aim to increase human analytical reasoning through visual and interactive interfaces. The need for interactivity requires that VA systems should include high-performance as one of the main conditions. In this thesis, we propose MSKDE - Marching Squares Kernel Density Estimation, a technique to generate fast and accurate hotspot maps. We describe the method and demonstrate its superior qualities through careful comparison with the Kernel Density Estimation (KDE), widely used to generate density maps. Another contribution of this thesis aims to help police departments in their planning activities. Professionals and researchers agree that tracking crime over time and identifying its geographic patterns is vital information for efficient resource planning. To help to perform these activities, frequently, police departments have access to systems that are too complicated and overly technical, leading to modest use at last. We collaborated with domain experts from police departments in Brazil and the United States to recognize and characterize five domain tasks inherent to the activity of tracking crime and resource allocation planning. All domain tasks are related to hotspot analysis and policing, one of the prominent approaches to fight crime. To facilitate the performing of the domain tasks, we proposed SHOC, The One-Shot Comparison Tool, a technique that allows immediate spatial comparison of crime density surfaces. We included SHOC into a VA system, CrimeWatcher, which allows users to perform filtering operations and visualize maps and data smoothly. CrimeWatcher strives for simplicity and will enable users, even without technical knowledge, to perform all tasks, annotate, save, and share analyzes. We also demonstrated that CrimeWatcher and SHOC effectively support the completion of domain tasks in two different real-world case studies.

Keywords: Visual analytics. Data visualization. Computer graphics.

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

KDE	Kernel Density Estimation
GIS	Geographic Information System
VA	Visual Analytics
PDF	Probability Density Function
SER	Search Efficiency Rate
PAI	Predictive Accuracy Index
RRI	Recapture Rate Index
VALCRI	Visual Analytics for Sense-Making in Criminal Intelligence Analysis
VALET	Visual Analytics Law Enforcement Toolkit
MSKDE	Marching Squares Kernel Density Estimation

CONTENTS

1 INTRODUCTION

This thesis presents a visual analytics solution that facilitates tracking crime over time and identifying its geographic patterns. In this chapter, I start discussing the motivation for this work in Section 1.1. In Section 1.2, I define the scope of the problem addressed by this thesis. In Section 1.3 I list and describe its objectives. I describe the methodology I used to achieve the objectives in Section 1.4 and the main contributions of this thesis in Section 1.5. In Section 1.6, I present my publications related to this thesis. Finally, I conclude this chapter with the outline of the rest of this manuscript in Section 1.7.

1.1 Motivation

Crime has become a central issue in many countries in the world. In particular, violent crimes have dramatically increased in some parts of Brazil in recent decades (LIMA *et al.*, 2014). This has challenged both governments and the entire society (IZIQUE, 2013).

Some countries invest heavily to combat crime. For instance, the USA, France, Germany, and Brazil spend more than 1% of their GDP annually on public safety (LIMA *et al.*, 2014). However, the total cost of crime is much higher than just the amount spent on public safety. In Brazil, the total cost associated with violent crimes in 2013, including public safety, state prison maintenance, and social costs, amounted to one-hundred billion dollars, corresponding to 5.4% of Brazil's GDP (LIMA *et al.*, 2014).

It is well known that crimes do not occur randomly, and do not spread uniformly through space. The offenders usually search the place and time for their actions carefully, trying to balance reward and their safety. Then, certain conditions make crime more frequent in some places than in others (SHERMAN *et al.*, 1989; ECK; WEISBURD, 1995; SHERMAN, 1995; CHAINEY; RATCLIFFE, 2005; RATCLIFFE, 2010). This has become known as "routine activity theory" and is studied by the branch of criminology called "environmental criminology" (WORTLEY; TOWNSLEY, 2016). The routine activity theory makes the spatial distribution of crimes one of the main features in crime analysis and prediction.

Police usually make their plans by exploiting the routine activity theory. One of the consequences of this theory is that a successful crime event increases the likelihood of similar crimes occurring in the same place and nearby regions (JOHNSON; BOWERS, 2004; WEISBURD *et al.*, 2012; JOHNSON *et al.*, 2007). Among the techniques for generating hotspots, continuous surface smoothing, based on Kernel Density Estimation (KDE) (ROSENBLATT, 1956) is the most prominent. Practitioners and researchers agree that the maps generated with these methods are more suitable for crime analysis (SANTOS, 2012) and provide good predictive performance (CHAINEY *et al.*, 2008).

In recent history, the world has faced increasing problems linked to criminal actions, whether by international terrorism or the rise of the power of criminal organizations, especially in the developing world. Following the September 11 attack in 2001, a class of systems, called Visual Analytics (VA) Systems[1], gained prominence and were increasingly adopted to address complex and dynamic issues, such as criminal analysis. Many researchers and practitioners chose this kind of system to deal with this unique scenario because they have evidence that the composition between computation and the human reasoning process would bring benefits in the efforts to work this intricate problem (THOMAS; COOK, 2005).

1.2 Problem Statement

Police departments throughout the world are trying to move from a reactive to a more proactive posture (SANTOS, 2012). The model based on a large number of officers, quick reaction to incidents and random patrols has been attacked by scholars and practitioners as expensive and ineffective (WEISBURD *et al.*, 2016).

One of the more successful policing techniques towards proactivity is hotspot policing (SHERMAN; WEISBURD, 1995; SHERMAN *et al.*, 1998; BRAGA *et al.*, 2014), which studies regions and allocates preventive resources based on hotspot maps.

In my collaboration with police departments, I identified that the sensemaking process on hotspot policing could be complicated. I identified that, to facilitate this process, the police needed not only an environment to identify and explore hotspots, but they also needed a tool to track and compare them, identifying their movements, expansions, and contractions. For example, they needed a tool to facilitate comparisons between different periods and distinct times of the day, and even compare mixed types of crimes. Successive meetings with the collaborators helped us to identify a set of domain tasks, composed of identifications and comparisons, that would have a significant impact on the police operations, and that could also benefit from an interactive visualization solution.

[1] "Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets" (KEIM *et al.*, 2010).

The literature already recognizes the importance of comparisons between crime density surfaces. Ratcliffe (2010), for example, states that “understanding the spatial dimensions of crime flux over time is a key component of cost-effective crime reduction in many situations.” Chainey e Ratcliffe (2005) point out that users need to know where crime is increasing or reducing, and where crime is fragmenting, expanding, or concentrating. Users also need to evaluate the effectiveness of specific crime prevention initiatives over time. All these examples indicate the benefits of a crime surface comparison tool.

A central task in police departments is crime analysis (SANTOS, 2012), and according to the International Association of Crime Analysts, the crime analyst gives support to most of the police activities. The activities include "the analysis of crime and criminals, crime victims, disorder, quality of life issues, traffic issues, and internal police operations, and its results support criminal investigation and prosecution, patrol activities, crime prevention and reduction strategies, problem solving, and the evaluation of police efforts" (International Association of Crime Analysts, 2014).

Another problem, according to our collaborators, is the mean to acquire information about the hotspots and making it consistently reach officers’ hands. In Brazil, for example, the police are divided into semi-autonomous regiments, which are frequently overloaded, with no dedicated team for elaborating their plannings. Although they have a general-purpose Geographic Information System (GIS) at their disposal, as it is a very technical piece of software, most of the officers do not use it. An officer said that “planning a routine task, like allocation of preventive resources, is a continuous process, and it should not be too laborious. In case of being a time-taking process, people will prefer to develop plans based on a priori knowledge rather than spending their little time trying to operate intricate systems.”

In the Visual Analytics (VA) systems field, we have not seen many efforts from researchers on providing systems that give evidence to, or assess the performance of, the interactive process. Another aspect that has been, in our opinion, neglected, is the set of features designed to improve team working (THOMAS; COOK, 2005).

1.3 Objectives

The overall goal of this thesis is to investigate how to facilitate reasoning and resource allocation planning in police departments by identification, exploration, and comparisons of crime densities using a visual analytics approach. To address the problems mentioned before, this thesis proposes CrimeWatcher, a visual analytics system that enables police departments to perform a set of domain tasks, identified as having a significant impact on their reasoning and planning activities.

To fulfill the main goal, the following specific objectives needed to be achieved:

1. Investigate environmental criminology to understand the relationship between place, time and crime;
2. Investigate the creation of crime maps, identifying the best techniques for crime analysis and prediction;
3. Investigate visual analytics for crime data to identify best practices and requirements;
4. Investigate police departments' reasoning and planning processes to understand their modus operandi and requirements;
5. Identify police departments' domain tasks that would benefit from a visual analytics system for crime data;
6. For each identified domain task, identify and design a visual analytics technique to address it;
7. Develop a prototype of a visual analytics system to address the domain tasks;
8. Evaluate the prototype, modeling use cases together with officers.

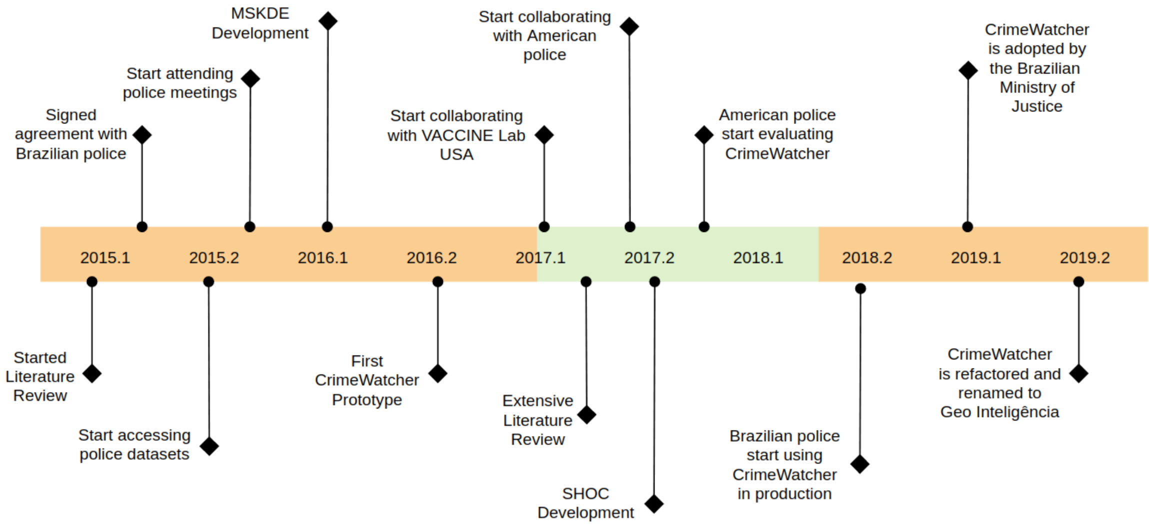
1.4 Methodology

The methodology of elaboration of this work had significant influence from the chronological sequence of my cooperation with police departments. Initially, I researched in Brazil for roughly two years. Then I spent fifteen months researching in the USA, and finally, I returned to Brazil to resume my cooperation with the Brazilian police department and finish my research. In addition, I have been continuously evaluating and adjusting the solution. These adjustments were based on literature review and feedback from officers and chiefs. Figure 1 shows the time line of the research, but I would like to emphasize the cyclical nature of this research, where I was always reviewing the literature, receiving feedback, and adjusting the solution.

The methodology could be summarized in the following steps:

- Conduction of a literature review: I focused on visualization, visual analytics, criminology, environmental criminology, crime mapping, and crime prediction.

Figure 1 – Timeline of the doctoral research



Source: the author.
The orange parts of the bar correspond to the time spent in Brazil, whereas the green part of the bar corresponds to the time spent in the USA.

- Establishment of cooperation with a Brazilian Police Department (BPD), which was composed of two parts:
  - Collaboration with the BPD Statistics Department: The BPD Statistics department was my first source of information inside the BPD as all the chiefs operational plannings were originally based on reports made by this department. Thus, they knew a lot about the Chiefs' needs.
  - Participation in the BPD meetings: I was allowed to attend dozens of their general planning meetings as an observer. I watched as chiefs presented statistics and discussed their jurisdictions. This was an excellent opportunity to understand how police departments plan their work, what kind of information they need, and the type of problems they face. As an ultimate intention, I tried to learn their reasoning process, for both Chiefs and officers.
- Identification of the domain tasks and requirements that the solution would be capable of accomplishing.
- Development of the conceptual approach, specifying the architecture components and the methodology steps.
- Construction of the first version of the prototype of my solution: a system capable of generating classic crime maps, as well as the MSKDE map. I have also started the annotation system and an analysis sharing feature. The statistics department

started using and evaluating it.

- Establishment of cooperation with two American police departments, which was composed of two parts:
  - Introduction of the approach: as I arrived in the USA with the first version of the prototype already working, I focused on presenting the solution and developing use cases together with American officers. I intended to investigate how the system matched their workflow to advance on the development of the tool, considering their domain tasks and requirements.
  - Participation in their strategy meetings: I was allowed to attend a few of their COMPSTAT (SCHOOL", 2019) meetings, and even present the solution in one of the meetings. In the COMPSTAT meetings, I could examine their needs and the suitability of the system.
- Adjustments in the domain tasks and requirements.
- Adjustments in the prototype to fit the changes in the domain tasks and the requirements.
- Deploy the system in the BPD and keep evaluating and adjusting it, according to feedbacks.

1.5 Contributions

This thesis is centered on providing a visual interactive environment for crime data that would facilitate performing a set of domain tasks identified as having a significant impact on police reasoning and plannings. It has three major contributions: a novel algorithm to calculate crime hotspots (Marching Squares Kernel Density Estimation-MKSDE), CrimeWatcher, a visual analytics solution for geocoded crime data that has been built in collaboration with police departments from the United States of America and Brazil, and a novel technique to facilitate spatial comparisons between density surfaces (One-Shot Comparison Tool-SHOC), also included in CrimeWatcher.

1.5.1 MSKDE – An improvement on the algorithm of generating crime hotspot maps based on KDE

The first contribution of this thesis is the MSKDE (Marching Squares Kernel Density Estimation), a fast technique to transform a low-resolution KDE hotspot map into a new hotspot map based on contour lines. The outcome of MSKDE is a hotspot map with rounded and smooth boundaries, which can be generated quickly, with an accuracy similar to KDE hotspot maps made with smaller cell sizes. Moreover, MSKDE maps, with their smoother boundaries, are a more natural representation of the real world's hotspots than the tiled looking original KDE maps. In an experiment, I will demonstrate that MSKDE maps preserve accuracy at lower computational cost when compared with traditional KDE maps with similar levels of accuracy.

To ensure a fair comparison between two hotspot maps, regarding their anomaly compared to a "perfect" hotspot map, we developed the Hotspot Anomaly Index (HAI). The HAI indicates the spatial percentage of misclassification of a hotspot map compared with a reference hotspot.

1.5.2 CrimeWatcher – A Visual Analytics System for crime data

To assist users in performing the hotspot analysis, we developed a web-based interactive visualization, and analysis environment for crime data. CrimeWatcher integrates traditional crime maps (point map, choropleth map, and KDE) with the novel MSKDE. This integration uses a layer-based approach, which allows sophisticated visual and interactive analyses.

To meet our collaborators' requirements and the visual analytic research agenda (THOMAS; COOK, 2005), we included a graphical and textual annotation system, as well as the feature of saving and sharing analyses. Thus, multiple users can communicate their ideas and interpretations to one another to encourage discussion. Those functionalities are novel among visual analytic systems for crime data.

With CrimeWatcher, users can compare different maps more efficiently and productively. They can rely on a powerful tool for tracking crime evolution that can improve crime prediction and help achieve crime reduction.

1.5.3 SHOC – One-Shot Comparison Tool

As I said before, I identified a set of domain tasks, evolving hotpot maps, that would have a significant impact on police operations. I further identified that a density surface comparison tool would promote the addressing of all domain tasks. For facilitating the spatial comparison between the surfaces, we developed SHOC, the One-Shot Comparison Tool. SHOC makes comparisons by superimposing two MSKDEs, each one representing one hotspot map. It has properties that make the comparison quick and effective, like freeing the user to make spatial comparisons visually and identification of spatial variation.

1.6 Publications

de Queiroz Neto, J. F., dos Santos, E. M., Vidal, C. A. Mskde-using marching squares to quickly make high quality crime hotspot maps. In: **2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. IEEE, 2016. p. 305-312.

Surakitbanharn, C., de Queiroz Neto, J. F., Wang, G., Ebert, D. S. Community Outreach Using Incident Records and Visual Analytics. In: **Community-Oriented Policing and Technological Innovations.** Springer, Cham, 2018. p. 19-27.

Nunes Junior, F. C., da Silva, T. L. C., de Queiroz Neto, J.F., de Macêdo, J. A. F., Porcino, W. C. A Novel Approach to Approximate Crime Hotspots to the Road Network. In: **Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility.** 2019. p. 53-61.

de Queiroz Neto, J. F., dos Santos, E. M., Vidal, C. A. Ebert, David S. A Visual Analytics Approach to Facilitate Crime Hotspot Analysis. In: **Computer Graphics Forum.** June 2020, vol. 39, iss. 3, p. 139-151.

1.7 Outline

While the first chapter motivates and gives some background information, the next chapters of this thesis are organized as follows.

Chapter 2: Fundamental Concepts. This chapter presents some information necessary to understand the thesis. First, it discusses Environmental Criminology and gives some evidence about the close relationship between crime and place, which is the reasoning behind hotspot maps. Second, it explains Kernel Density Estimation, the most used technique to make analytical and predictive crime maps. Third, it presents Marching Squares, a fast and precise method to create contour lines.

Chapter 3: Related Work. This chapter provides a summary and a critique of the most related state-of-the-art literature regarding crime maps, Geographic Information Systems, and Visual Analytics for crime data.

Chapter 4: Marching Squared Kernel Density Estimation (MSKDE). This chapter presents the first contribution of this thesis, which is a technique to calculate a crime hotspot map based on contour lines.

Chapter 5: CrimeWatcher - A WEB System for Visual Analysis of Crime Data. This chapter presents the visual analytics system that has been developed during this doctorate. It contains explanations of the domain tasks, the system interface, and how CrimeWatcher helps to address the domain tasks.

Chapter 6: Case Studies. This chapter describes two real case studies to demonstrate that SHOC helps to perform the domain tasks.

Chapter 7: Conclusions and Future Works. This chapter includes a summary of this thesis and an outlook of future work.

2 FUNDAMENTAL CONCEPTS

This chapter contains the basic concepts necessary to understand the main contributions of this thesis. The chapter is divided into four Sections. In Section 2.1, I discuss Environmental Criminology, which lead to the study of the relationship between crime and place. In Section 2.2, I discuss the relationship between crime and place, which is the foundation of the Hotspot Policing. In Section 2.3, I present the Kernel Density Estimation (KDE), the most used technique for making crime maps. In Section 2.4, I present Marching Squares, a technique used in CrimeWatcher to find isopleths. Finally, in Section 2.5, I conclude and summarize this chapter.

2.1 Environmental Criminology

According to Wortley e Townsley (2016), environmental criminology is a family of theories that share a common interest in criminal events and the immediate circumstances which they occur. The criminologist who follows this theory believes that environmental reasons could explain crime patterns, with the crime being the object of study, not the criminal. For them, the criminal is only a part of the crime equation, together with the place and the target.

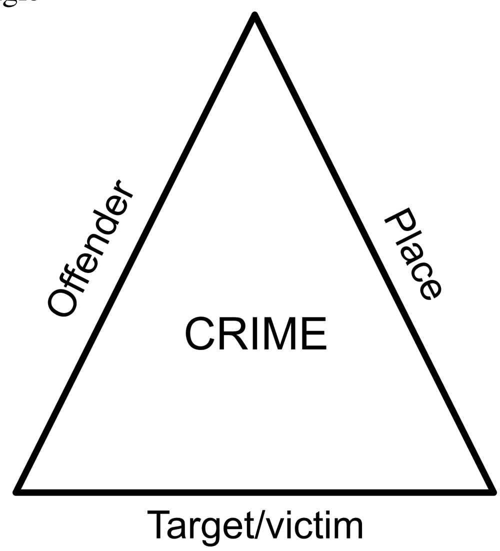
Environmental criminology is based on three aspects:

- The environment has a significant influence on criminal decisions. The motivation of the offender is not the only factor to trigger a crime. Features of the place and situation could have notable responsibility for the occurrence.
- Crime does not spread randomly in time and space. Criminals will look for areas with better opportunities, rather than well-guarded regions.
- Crime investigation, control, and prevention could benefit from the knowledge and manipulation of the environment. Changes in places can have a significant influence on crime incidence.

One of the main theories of the Environmental Criminology is the Routine Activity Theory (COHEN; FELSON, 1979), which states that a predatory crime occurs when a likely offender and a suitable target come together in time and space, without a capable guardian present (CLARKE; ECK, 2014).

The routine activity theory is the foundation of the crime triangle, shown in Figure 2. The crime triangle is a representation that indicates that for committing a crime, a criminal needs

Figure 2 – The Crime Triangle



Source: the author, based on Clarke e Eck (2014).

a target and a good place. For an officer or crime analyst studying a crime, the crime triangle symbolizes that all three dimensions must be present and should be considered.

Another critical theory of the Environmental Criminology is the Crime Pattern. It states that crimes do not occur randomly or uniformly in time or space or society (BRANTINGHAM *et al.*, 2017), but according to circumstances related to the environment. The theory proposes that the offender moves within familiar spaces, looking for potential targets, which also have their movement pattern. The intersection of the offender and victim spaces is identified as the criminal activity space of the offender. Moreover, the Crime Pattern Theory gives the theoretical foundation on criminology for the existence of crime maps.

The third basic theory of environmental criminology is the Rational Choice Theory, which states that criminals do not act because they would have a crime-oriented personality or predisposition. The theory proposes that criminals exploit opportunities and measure risks and rewards before committing a crime (CORNISH; CLARKE, 2017).

Finally, as a summary of the crime equation, according to the theories of environmental criminology, one can say that crime usually happens in a suitable and offender's familiar place, in an attractive situation, with a reward that repays the risks.

2.2 The relationship between crime and place

Over the past three decades, a new perspective has been unfolded in criminology, a perspective with the potential to greatly increase our understanding of crime. Researchers are finding strong evidence that very specific locations, such as blocks and addresses are more important for understanding crime than larger space units such as neighborhoods or districts (WEISBURD *et al.*, 2016).

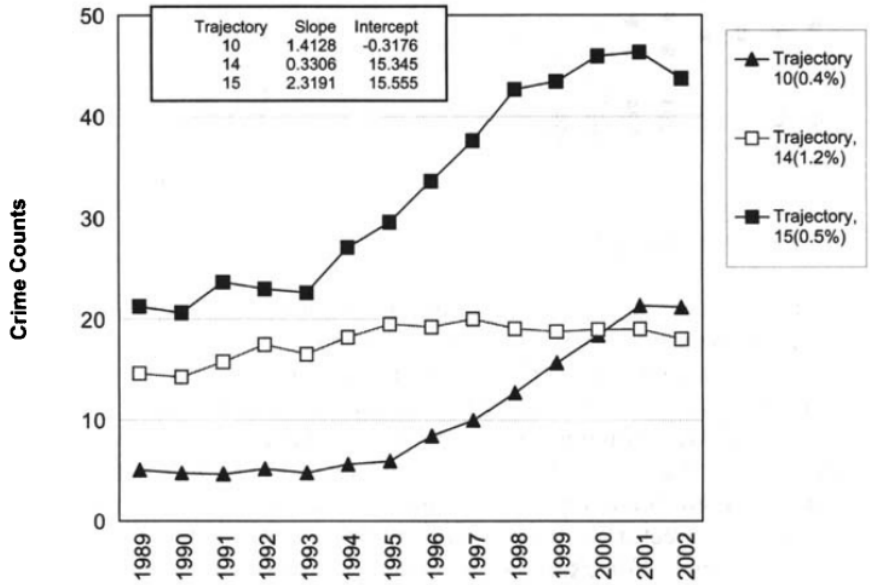
Empirical research consistently points out that crimes are concentrated in some regions of cities, regardless of whether cities are small or sizeable. Weisburd (2015) studied eight North American cities of different sizes, with populations ranging from 70,000 to 8,000,000 people, and analysis period ranging from 1 to 20 years. Weisburd found that despite the significant variability of cities, the concentration of crime shows little variation. He found that 50% of crimes were in the range 2.1% to 7.0% of the area, and 25% of crimes were in the range 0.4% and 1.6% of the area. Weisburd also investigated the amount of crime and the concentration of crimes over time and found that even though there was a wide variation in the number of crimes, the concentration varied very little over time. Weisburd identified that from the 1970s to the 1990s crime in the United States doubled, falling back to the same levels as the 1970s in 2010. In these 40 years, the concentration of crime has not changed significantly.

In a study of predatory crimes in Minneapolis, Sherman *et al.* (1989) found that only 3.3% of the places were responsible for more than 50% of the calls for police service, between December 15, 1985, and December 15, 1986. Evidencing the importance of place in crime, Sherman (1995) highlights that, in Washington, D.C, in Feb 14, 1989, four of the fourteen shot incidents at Capital were at the same block of a street.

Eck e Weisburd (1995) discussed the value of places in crime by considering five research areas: crime concentration on particular facilities; the high concentration of crime at some addresses and the absence of crime at others; the preventive effects of various place features; the mobility of offenders; and studies of how offenders select targets.

Weisburd *et al.* (2004) analyzed, using a statistical technique called group-based trajectory model (NAGIN, 2010), evidence about the concentration of crimes at some places. They had a database of fourteen years of criminal history in Seattle, the USA, and analyzed crime trend and the quantitative crime history (called “trajectory”) for each city street block (29,849 in total). After having all trajectories computed, they divided them into eighteen groups by trajectory similarity. After that, they analyzed whether the crime quantity at these group blocks was raising, decreasing, or stable. They found that, for most group blocks, the trajectory indicates stability over time, and the variability found in crime rate came almost all from a specific group of blocks. Figure 3 displays the count of crimes at three selected groups of blocks,

Figure 3 – Using group-based trajectory model to evidence the link between crime and place (WEISBURD *et al.*, 2004)



Source: Weisburd *et al.* (2004).
Weisburd *et al.* (2004) analyzed the historic quantitative crime trajectory of groups of street blocks in Seattle, USA. Each trajectory refers to a group of blocks and displays how many crimes were recorded in each year, producing the evolution of that group of street blocks.

where the crime count is increasing over time.

One can compare location-based crime statistics with people-based crime statistics, to see that by working with local criminology, we are dealing with fewer variables, and therefore would be more straightforward. For example, Wolfgang *et al.* (1987) found that, in the set of cities of their study, 6% of criminals are responsible for 50% of crimes. In the same cities of his study, they found that 50% of crimes occur in 3% of the city extension. Weisburd *et al.* (2016) consider that it is simpler to work with a percentage of an area of a city than a percentage of people; among other reasons, people move while areas do not.

Other works contributed to improve the theoretical framework of the criminology of place (KINNEY *et al.*, 2008; BRANTINGHAM; BRANTINGHAM, 1995; BRANTINGHAM; BRANTINGHAM, 1984) and to consolidate it as a relevant field of study (SHERMAN *et al.*, 1998; SHERMAN *et al.*, 1989; WEISBURD; ECK, 2004).

Given that crime is concentrated in a small part of the city, it would be quite reasonable to imagine that preventive actions taken in these areas of concentration would be more effective than those taken in other areas, or the city as a whole. This is the first principle of the criminology of place, and is widely explored by police departments through the Hotspot

Policing.

Hotspot policing is the strategy of deploying more crime prevention resources to areas identified as having higher crime likelihood (WEISBURD *et al.*, 2016; SANTOS, 2012). Considering the Routine Activity Theory and the Crime Pattern Theory, to identify the areas of higher crime likelihood, one must instead identify the areas of higher crime concentration. For this, crime mapping is one of the most prominent techniques in use (WORTLEY; TOWNSLEY, 2016). Hotspot policing has been assessed many times (BRAGA *et al.*, 2014; BRAGA, 2001; SHERMAN; WEISBURD, 1995) and researchers are convinced that place-based initiatives help to reduce crime and disorder (WEISBURD *et al.*, 2016).

The distribution of crimes in space is a Probability Density Function (PDF) that can be estimated. In the next section, I explain Kernel Density Estimation (KDE), the technique used to estimate the crime PDF, and that is one of the foundations of the proposed solution.

2.3 Kernel Density Estimation

Kernel Density Estimation (KDE) is a technique for nonparametric estimation of a PDF. It was initially developed by Rosenblatt (1956) for unidimensional random variables. However, it can be extended to more dimensions – in our case, we deal with two dimensions.

The idea behind KDE is that the random variable being studied has an inherent PDF that can be statistically estimated, based on sample data that is supposed to be generated by the real density function itself (BISHOP, 2006; SILVERMAN, 1986).

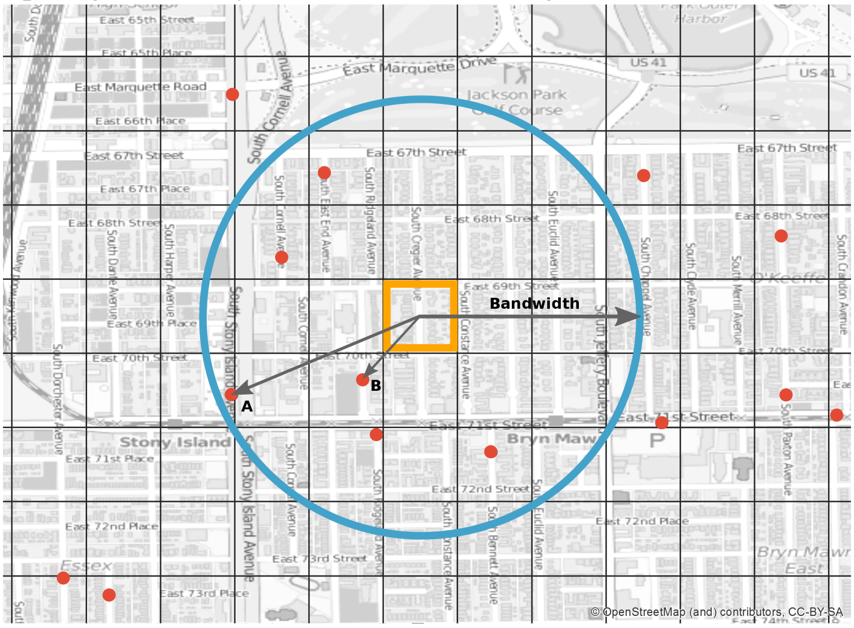
Algorithm 1: KDE($C, k, band, s$)

Input: C - dataset of events; k - kernel function; $band$ - bandwidth; s - cell size.
Output: A two-dimensional field

```
rect ← ComputeRectangleOfInterest(C, band)
field ← MakeEmptyGrid(rect, s)
for all j ∈ C do
  I ← CellsInsideCircle(j, band)
  for all i ∈ I do
    contrib ← ComputeDensityValue(i, j, k)
    field[i] ← field[i] + contrib
  end for
end for
return field
```

In the case of two dimensions, in which we have events scattered on a region, KDE

Figure 4 – Computing the density value for one cell using KDE



Source: the author.
How the density in one cell (in orange) is calculated using KDE: for each event, represented by a point, inside the bandwidth circle, the algorithm computes its distance to the cell center divided by the bandwidth and sends it to the kernel function. Common kernel functions consider that nearer events (e.g. B) contribute more to the cell density than distant events (e.g. A). Finally, the density is computed as the sum of contributions of all events that are inside the bandwidth.

estimates the PDF by dividing the area into a regular grid of cells, and calculating for each cell, using a kernel function, a density value. The kernel function calculates the density, summing contributions of all events that are located within a certain distance (bandwidth) from the cell center. The result is a two-dimensional scalar field that estimates the PDF, and it is the basis for a hotspot map. Figure 4 illustrates this process showing an example for a single cell in a two-dimensional field, whereas Algorithm 1 shows the pseudo-code for a standard two-dimensional KDE.

From the explanation above, it follows that KDE has three parameters that directly affect the resulting map: the cell size, the bandwidth and the kernel function.

Cell size: it impacts the map resolution, which affects appearance, accuracy and execution time. Small cells lead to more accurate, smooth and good-looking maps, which are slow to compute. Large cells, in turn, lead to less accurate, pixelated maps that are fast to compute;

Bandwidth: it impacts the degree of events aggregation and cluster formation. Small bandwidths lead to spotty maps, because the events will only contribute to nearby cells. On the other hand, large bandwidths aggregate more points, forming big clusters;

Kernel function: it defines how the events will contribute to each cell density. Common functions return a value inversely proportional to the distance between the event and the center of the cell, giving more weight to events that are nearer the cell. The most used and recommended kernel function is the quartic[1](HART; ZANDBERGEN, 2014; LEVINE, 2013).

The KDE algorithm can be implemented in two ways, which result in different computational complexities (n is the number of cells, e is the total number of events, s is the cell size, b is the bandwidth):

1. Based on cells: each cell is visited and its density is calculated using all events that are within the bandwidth. In this case, the complexity is $O(n \cdot e)$.
2. Based on events: each event is visited and all cells within its bandwidth are updated with its calculated contribution. In this case, the complexity is $O(e \cdot \pi \cdot ceil^2(\frac{b}{s}))$.

To optimize our implementation, we compare n with $\pi \cdot ceil^2(\frac{b}{cs})$ and choose the way that produces the fastest computation.

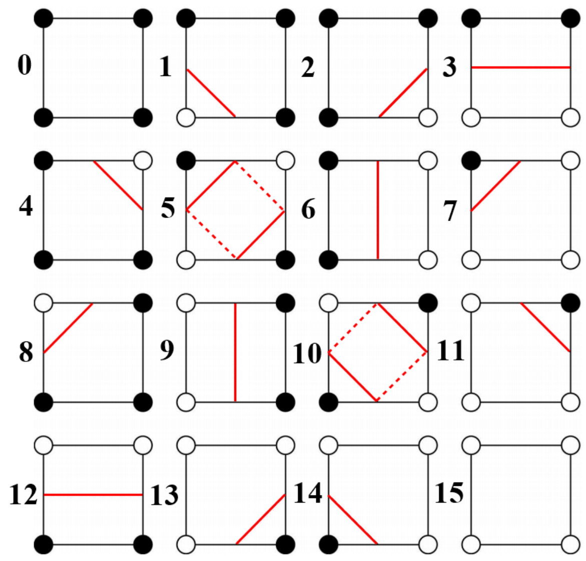
After being calculated, the crime PDF needs to be displayed on a map. In the next section, I explain Marching Squares, an algorithm that quickly generates high-quality isocontours, suitable for creating hotspot maps with some properties needed to build the proposed solution.

2.4 Marching Squares

Marching Squares is a very efficient algorithm for generating contour lines in a regular grid representing a two-dimensional scalar field. It is the 2D version of the 3D marching cubes algorithm (LORENSEN; CLINE, 1987). As its name implies, it works by *marching* the cells in the field, processing each one of them independently. The values of the scalar field on the cell's vertices (corners) are analyzed and the geometry of the cell is computed based on a pre-built lookup table (see Figure 5). The final position of the contour is obtained by linear interpolation along the cell's vertices. Notice in Figure 5 that cases 5 and 10 are ambiguous cases. For simplicity, the disambiguation in our implementation was obtained by averaging the values on the corners: if the average is greater than or equal to the contour value, the contour will pass between the two highest vertices (dotted lines), otherwise the contour will pass along the lowest vertices (solid lines). Other possible solutions for disambiguation are described in (NIELSON; HAMANN, 1991; LOPES; BRODLIE, 1998). The Marching Squares algorithm is

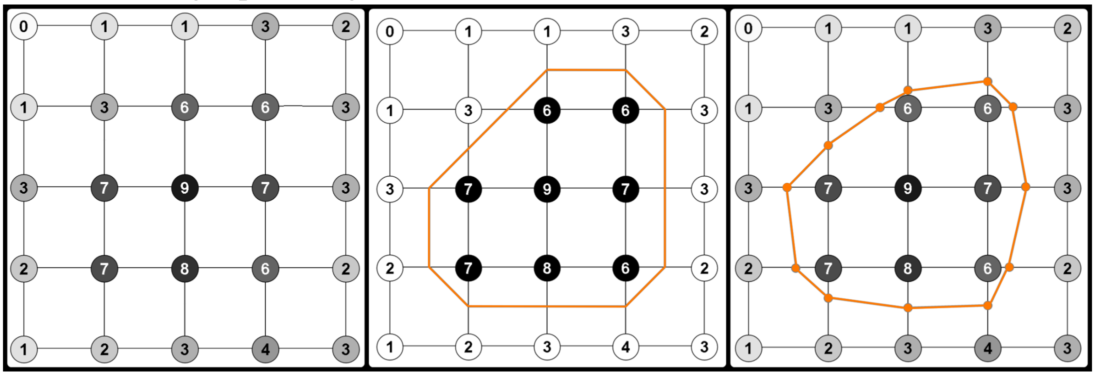
[1] $f(u) = \frac{15}{16}(1-u^2)^2$ for $0 \leq u \leq 1$, and u is a normalized bandwidth.

Figure 5 – Marching Squares' 16 cases



Source: the author.
Based on the cell's four vertex values, the contour can assume one of the 16 shown possibilities (0 to 15). Filled vertex dots mean that their values are greater than or equal to the contour value. Non-filled dots mean the opposite. In two cases (0 and 15), the contour line does not pass through the cell because the contour value is above or below all four vertex values. Special cases 5 and 10 are saddle points and hence, ambiguous.

Figure 6 – Marching Squares algorithm in action



Source: the author.
On the left, a field where marching squares will be used to generate an isocontour representing value $v = 5$. In the middle, the result after "marching" all cells, each cell assumes one of the 16 cases in Figure 5 without any interpolation. On the right, the result of the interpolation phase, where the contour vertices are adjusted to cell vertex values.

detailed in Algorithm 2.

Figure 6 displays a small example of Marching Squares being calculated. Note that the interpolation phase (on the right) is separated from the cases' classification (in the middle) only for illustration purposes; in fact, they occur at the same time to avoid another loop over the field.

Algorithm 2: Marching Squares

Input: A 2D scalar field with **h** rows and **w** columns, a contour value **v**.
Output: A set of edges **E**.

Load the lookup table L with the 16 possible cases (as in Figure 5)
for $i = 1$ **to** $h - 1$ **do**
  for $j = 1$ **to** $w - 1$ **do**
    Define a logical cell c and define $c.values$ with the field values of vertices (i, j),$(i+1, j)$,$(i, j+1)$ and $(i+1, j+1)$
    index ← build_index($c.values$, v)
    case ← lookup(L, index)
    if case == 5 **or** case == 10 **then** {Ambiguous cases}
      if average($c.values$) < v **then**
        Contour follows along lowest vertices (solid lines in cases 5 and 10 in Figure 5)
      else
        Contour follows along highest vertices (dotted lines in cases 5 and 10 in Figure 5)
      end if
    end if
    Build cell geometry for the case (edges and vertices)
    Linearly interpolate the edge vertices along $c.values$
    Store the interpolated edges in the set **E**
  end for
end for
return **E**

2.5 Summary

In this chapter, I presented some fundamental concepts used to build the proposed solution. First, I introduced Environmental Criminology, which aims to analyze crimes giving elevated importance to the circumstances of the incident, like place and time. Second, I discussed the relationship between crime and place, which are coupled together in most of the crimes. We saw that crime does not distribute randomly and is concentrated, with notable stability, in some parts of the city. Practitioners and researchers use this evidence to predict the place of future crimes and make interventions. Third, I explained Kernel Density Estimation (KDE), a technique to estimate a PDF. Due to its stability over time, researchers consider crime a PDF that can be conveniently predicted by a historical dataset using KDE. KDE will be used to calculate the density surface that will be the base of the hotspot map. Finally, I presented Marching Squares, a fast algorithm to create high-quality isocontours. We use Marching Squares for creating hotspot maps by applying it to a density surface calculated by KDE.

3 RELATED WORKS

We divided the related works into two sections. The first one is dedicated to crime mapping techniques, where we explore their historical evolution until reaching the current state. The second section is focused on GIS and VA systems for dealing with crime data.

3.1 Crime Mapping Techniques

In a visual analytics system for spatiotemporal data, a key component is a map, where spatiotemporal occurrences can be marked, relating them to administrative or natural divisions, or represented as clusters, like hotspot maps.

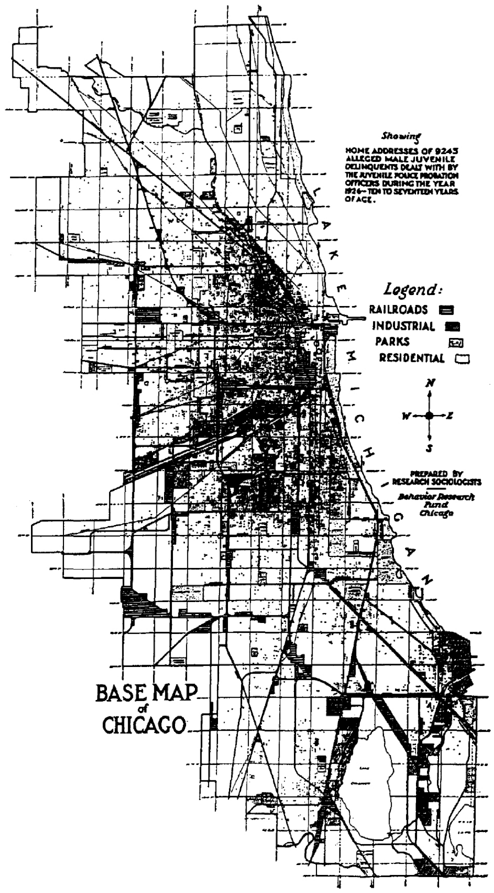
Consequently, the crime map being the starting point, the predominant view of the system, and the hotspot map being a primary visual instrument of analysis, its quality and construction technique would be fundamental. We have divided this Section into two parts, one discussing about the history of crime maps, and the other dedicated to the work and research in finding the best map for crimes, which could lead to better *insights*.

3.1.1 The history of crime maps

Weisburd e McEwen (1997) present an overall historical context about mapping crimes, from which we would like to highlight two initiatives that, due to their innovative features, have made history. First, introducing what is called a crime map, Balbi e Guerry (1829) built maps crossing crimes with the educational level of the population of France. They found that, in the best-educated regions, the most common crimes were against property, rather than violent crimes. In the less educated regions, the opposite effect has occurred. The second initiative was a spatial study of the juvenile delinquency in Chicago, in the 1920s and 1930s. A group of scholars, led by Clifford Shaw and Henry McKay, studied juvenile delinquency in Chicago, identifying a spatial correlation between crimes and areas with "economic disadvantage and social disorganization" (SHAW; MCKAY, 1942). They created crime maps similar to current computer-generated GIS maps (see Figure 7) and discovered patterns that suggest that delinquency rates vary according to community characteristics. Such characteristics are the "result of city expansion and migration patterns," in a very consistent way (RATCLIFFE, 2010).

It's important to note, that between the two presented initiatives, the research on crime mapping was reduced because the science of criminology, at that time, turned to emphasize

Figure 7 – Home addresses of allegedly delinquent youth (SHAW; MCKAY, 1942)



Source: Shaw e McKay (1942).

positivist theories, where the physiological characteristics of the offenders (such as head shape (HAYNES, 1935)) would explain their criminal behavior (WEISBURD; MCEWEN, 1997).

3.1.2 The search for the best crime map

The search for the best crime map went through two stages. First, the decision of what type, among several kinds of crime maps that have been historically used, would be the best for representing crimes. Second, researchers start searching for the best map for predictive tasks, including the search for an index that would measure the technique's predictive capability.

3.1.2.1 Comparison of several mapping techniques for use with crime applications

There are, basically, four types of crime maps: point maps, choropleth maps, ellipse maps, and KDE maps. Some works (ECK *et al.*, 2005; CHAINEY; RATCLIFFE, 2005; CHAINEY *et al.*, 2002) discuss the problems associated with point maps, choropleth maps,

Table 1 – Crime mapping applications and problems associated

Mapping Method	**Mapping Application**	**Problems of Method**
Point Maps	Small number of incidents, check incidents individually.	Inappropriate as method to interpret crime clusters
Choropleth Maps	Report of crime events by administrative area	Modifiable area unit problem
Ellipse Maps	Grouping of crime clusters	Crimes usually don't form spatial ellipses
KDE Maps	Crime analysis and prediction	All incidents must be geocoded

Source: the author, adapted from Chainey *et al.* (2002).

and ellipse map, on representing crimes. Authors indicate that, because of those issues, the community has been increasingly moving to KDE maps.

Table 1 shows a summary of the crime mapping types, applications and, problems for each technique.

3.1.2.2 Principal researches about predictive crime mapping

Regarding the search for the map with the best predictive skils, we selected the most groundbreaking works, which investigated the best map construction techniques and tried to define evaluation indexes to access the methods.

3.1.2.2.1 Prospective hotspot map and the Search Efficiency Rate (SER) index

In a work focused on burglary crimes, Bowers *et al.* (2004) presented a heat map, built using KDE, which contains special elements aimed at increasing its predictive ability. The first element was a change in the *kernel* function. The authors developed a rational kernel function, which involves both the spatial distance of the event to the center of the cell and the temporal distance between the day the event occurred and the heat map reference day. As a measure of spatial distance, the authors used the integer amount of cell side halves of the *grid*. For example, if the *grid* used to generate the KDE is 50 meters, and the event occurred 280 meters from the center of the cell being measured, the distance would be $integer(280/25) = 11$. As for the temporal distance, this is measured in whole weeks, discarding any remaining ones.

In this way, the equation to compute the density of a cell with index i is:

$$f(i) = \sum_{k=1}^{n} \frac{1}{(1+e_k)} \frac{1}{(1+t_k)} \tag{3.1}$$

Where n is the number of events within the *bandwidth* of cell i, e_k represents the spatial distance, and t_k represents the temporal distance (BOWERS *et al.*, 2004).

Note that in Equation 3.1, the number 1 was added to both the spatial distance and the temporal distance because, if the event had any zero distance, it would make the rational equation irregular. It was then established that the minimum distance of any measurement, whether spatial or temporal, is 1.

The idea of weighing time in reverse is that older crimes would have a minor influence on the occurrence of new crimes, thus reducing their contribution when calculating densities.

The second element that would improve the predictive capability of the model was choosing a more appropriate bandwidth. The model has two bandwidths, one spatial and one temporal; the spatial one was fixed at 400 meters, and the temporal one at 2 months. The choice of these two values comes from a previous work by Johnson e Bowers (2004), in which it was identified that burglary crimes positively influence the occurrence of new crimes of the same type up to a distance of 400 meters, and by up to 2 months.

Because of these two prediction-oriented elements, the authors named their technique as *prospective*, in contrast to earlier techniques, which in their opinion had a *retrospective* nature. From this work, the name *prospective* has also been used by other authors when referring to some technique that uses time gradient as a component of the *kernel* function.

To compare their work with earlier models, the authors generated heat maps for a region in the south of the city of Liverpool, England, using both retrospective (not weighting the time of the event) and prospective techniques. The retrospective map was made with a 50 meter grid, 200 meter bandwidth, and a quartic function. The prospective map was made with a grid of 50 meters, a bandwidth of 400 meters, and the rational function already presented. A new form of heat map evaluation has been created, called the Search Efficiency Rate (SER), which measures the number of correctly predicted crimes in each square kilometer of the map hotspot. The *hotspot* region was defined in both maps as the area with the highest 20% of density. The hit rates were compared for two groups: crimes up to two days and up to seven days from the date of map generation. The results indicated a better performance of the prospective map for both the 2-day and 7-day groups. In the 2-day group the prospective map hit 62% of the crimes while the retrospective map hit 46% of the crimes. In the 7-day group, the prospective map hit 64% of the crimes while the retrospective map hit 56% of the crimes. The SER index values were presented for the 2-day group only, yielding 2.96 crimes per km^2 for the prospective map and 2.22 crimes per km^2 for the retrospective map.

The work of Bowers *et al.* (2004) was the first predictive crime map in criminology that uses temporal distance as a variable of the *kernel* function; previous works used time only as a *bandwidth*, since they limited predictive events to a range, but gave the same weight treatment for both older and newer events. The authors also presented a metric for heat map predictive quality (the SER), which although feasible for comparisons within the same region, may yield incorrect results if the comparison is made between regions of different sizes. This problem occurs because the calculation of the SER does not include the ratio of the hotspot to the whole region, and situations with totally different criminal densities may have the same value for the index, which would be a nonsense. Chainey *et al.* (2008) provides a textual counter-example, demonstrating that hotspot generation techniques that lead to totally different realities can have the same SER. Chainey's textual example is not simple to follow so we present an adaptation of his example, which can be easily understood through Figure 8.

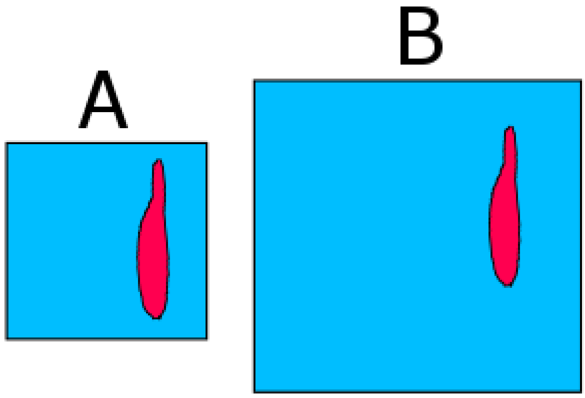
The authors have not provided a rationale for using the rational function in their kernel from the various options available (quartic, Gaussian, triangular, etc.). It is noteworthy that the rational function used has a rapid decay when it begins to move away from the center of the cell. For example, concerning spatial distance, just 2 cells away from the center (4 half cells), the event contribution value already drops to $1/(1+4) = 1/5$. I believe that a comparison with another function that does not have such a rapid decay (triangular for example) could be more informative.

3.1.2.2.2 Assessment of the Kernel Density Estimation (KDE) on prediction and the Predictive Accuracy Index(PAI)

Chainey *et al.* (2008) presented a study assessing the predictive ability of some of the most traditional crime mapping techniques, focusing on 2 main objectives:

- Identify if the predictive accuracy of crime maps varies, considering the different types of crimes;

Figure 8 – Counterexample for the SER



Source: the author, based on the work of Chainey *et al.* (2008).
Consider regions A and B, with 100 km^2 and 200 km^2 of area, respectively. We created heat maps for both regions using different techniques (techniques A and B), which happened to generate *hotspots* of the same size: $10km^2$ for both regions (highlighted in red). In both regions there were 400 crimes, with 200 crimes being correctly predicted in both hotspots. In this context, the calculation of the SER for both regions has the same value ($200/10 = 20$ crimes per km^2), which would indicate the same quality between the two techniques. However, it can be seen that technique B marked a much smaller hotspot area ($10km^2/200km^2 = 5\%$ of the area) proportionally compared to the hotspot generated by technique A ($10km^2/100km^2 = 10\%$ of the area). This fact indicates that technique B generated a hotspot that is more efficient than that generated by technique A (a proportionally smaller hotspot map would facilitate, for example, resource allocation). Finally, maps with the same SER may have different quality, indicating that this index is not suitable for every situation.

– Compare the predictive accuracy of various types of crime maps.

The authors had 2 years (01/01/2002 to 12/31/2003) of crime data in central north London, totaling 52,844 events. Crimes were classified into 4 types: residential burglary, street crime (robbery or theft), theft from a vehicle, theft of vehicles. The prediction was made by separating crimes into two groups by temporal criteria, using the oldest group to generate the predictor model, testing its predictive accuracy against the second group of crimes (more recent crimes).

The data were initially divided with the predictor group assuming the period from 01/01/2002 to 12/31/2002 and the predicted group assuming the period from 01/01/2003 to 12/31/2003. In an attempt to isolate some seasonal factors, typical of that region of London, and also as another way of checking consistency, another date division was created, with the predictor group between 03/13/2002 and 03/12/2003, and group to be evaluated between 03/13/2003 and 03/31/2003. The analyzes of both groups were performed in parallel, independently. In explaining the methodology, we will use the dates of the first division because the activities done with the second division were the same.

From the predictor data, ten non-mutually exclusive subsets of data were produced, comprising different periods from one day to one year, all periods ending 12/31/2002, with each subset creating a predictor map. Similarly, from the data to be validated, ten non-mutually exclusive subsets were elaborated, comprising different periods from one day to one year, with all periods starting on 01/01/2003. The one-day map was compared to the one-day future data, the one-week map was compared to the one-week future data, and so on, yielding ten comparisons. The final result of the evaluation was the average of the prediction accuracy of the ten maps with the ten evaluation subgroups.

In the prediction evaluation step, the authors presented a new index called Predictive Accuracy Index (PAI). The index was created to consider both the hit rate and the percentage of hotspot in the evaluation of the predictive map. It is essential to consider the hotspot area in the calculation because if the evaluation was done only through the Hit Rate, it would be enough that the area defined as hotspot would be the entire map to obtain a 100% hit. The authors then defined the PAI index as the ratio between the percentage of Hit Rate and the percentage of the hotspot within the total area under study, with equation:

$$PAI = \frac{(\frac{n}{N})100}{(\frac{a}{A})100} = \frac{\frac{n}{N}}{\frac{a}{A}} \tag{3.2}$$

Where n means the number of “upcoming” crimes that occurred within the *hotspot* region, N the total number of crimes, a the area considered as *hotspot* and A the total area under analysis (CHAINEY *et al.*, 2008).

Finally, the *PAI* is a number that can be interpreted as the number of times that the *hotspot* region of the predictive map turned out to be criminally denser at the time of prediction than the total area. The *PAI* index is a number that the larger the better, revealing which maps are more accurate for predictive activity.

The types of crime maps that were evaluated in the article were: spatial ellipses, neighborhood thematic map, quadratic thematic map and Kernel Density Estimation. For each of these 4 map types, 10 predictor maps were generated for each type of crime, out of a total of 160 maps (4 map types x 4 crime types x 10 time intervals). All maps were compared with their corresponding “upcoming” data and calculated the *PAI* indexes.

After aggregating and averaging the *PAI* indexes, the authors reported that, in terms of type of crime, street crime (theft and robbery) is significantly more predictable than other types of crime. Regarding the type of map, the authors reported that the maps generated by Kernel Density Estimation are significantly better for prediction than the other. Table 2 provides a summary of the results.

Table 2 – Results of *PAI* index calculations

Map Type	**RB**	**SC**	**TFV**	**TOV**
Spatial Ellipsis	1.32	2.59	2.15	2.93
Neighborhood Thematic	1.25	3.32	2.93	2.01
Quadratic Thematic	1.95	4.14	2.55	1.89
KDE	2.33	4.59	3.66	3.05

Source: Chainey *et al.* (2008). Legend: RB => Residential Burglary. SC => Street Crime. TFV => Theft from Vehicle. TOF => Theft of Vehicle.

The work of Chainey *et al.* (2008) brought mathematical evidence for the consolidation of KDE as the best for crime prediction. In a work on criminal analysis, Santos (2012) supplements that the KDE-based map is the choice for analytical activities, the others being more suited to administrative tasks. As for the types of crimes, the evidence that street crime (robbery and theft) would be more predictable is exciting and indicates that this type of crime is more linked to the spatial component than the other types. The lack of homicide crime in this analysis leaves this gap for other researchers to invest in this line of research.

The Prediction Accuracy Index is simple and easy to calculate, and it is an advance on the SER by Bowers *et al.* (2004). There is a criticism about *PAI* that lies in the fact that it ignores the way the predictor heat map was generated, and may be very inaccurate. A work by Levine (2008) analyzes *PAI* and seeks to address its possible deficiency by introducing another index, called by the author as Recapture Rate Index (RRI), which will be discussed later.

An important observation is that *PAI* alone does not indicate whether the model is suitable or not for practical use, because the same index can be obtained in countless ways, with the same data, varying the amplitude of the *hotspot*. Ideally, we should build a predictive map with the customer in mind and how much hotspot area they want. For example, making a heat map with a *hotspot* of twenty percent of the area, considered good for having a high *PAI* index, may be unfeasible for the police of a city that would only have the resources to act on ostensibly in ten percent of the area.

3.1.2.2.3 The evaluation of the map precision and the Recapture Rate Index (RRI)

The RRI arose in a response from Levine (2008) to Chainey *et al.* (2008), in which Levine, while recognizing the value of the *PAI* index as a means of comparing predictive capability between crime mapping techniques, warns that *PAI* ignores the way the heat map has been generated and this may refer to a situation of great inaccuracy. The best way to understand Levine's argument is by an example. Take a hypothetical region where there were, in a predictive period, 1500 crimes and, having made a heat map and segregated the 10% denser surface in crimes, this surface comprised 50% of all crimes (750 crimes). Applying this heatmap to a more recent period, to identify its predictive capability, it was found that of the 1600 crimes of the new period, a total of 600 crimes were located in the *hotspot* region of the map. Calculating its *PAI* reveals a value of 3.75[1] but ignores that there was a loss of accuracy in the model between heat map generation and use, since the hotspot region comprises 50% of predictor data and only 37.5% of future data under analysis.

To solve this problem, Levine (2008) proposed the RRI, whose formula is defined as:

$$RRI = \frac{\frac{n_b}{n_a}}{\frac{N_b}{N_a}} \tag{3.3}$$

Where n_b is the number of "future" crimes that occurred in the *hotspot* region, n_a the number of predictive crimes present in the *hotspot* region, N_b the total number of "future" crimes and N_a the total number of predicted set crimes.

The RRI can also be understood as the ratio between the hit rate of the prospective period and the hit rate of the retrospective period. An RRI index of 1 reveals that future crimes occurred in the *hotspot* region with the same precision as the predictive data. An RRI index of less than 1 (somewhat expected) indicates that there has been a loss of accuracy while an index of greater than 1 indicates that the prediction has reached a higher value than in the creation of the *hotspot*, which in Levine's opinion is unreal and not to be expected. Levine indicates that what is expected for a model with good accuracy is an RRI smaller than 1 but close to it; An RRI index much smaller than 1 indicates a model with poor accuracy.

We understand that RRI adds interesting information about the prediction. However,

---

[1] $PAI = \frac{\frac{600}{1600}}{\frac{10}{100}} = 3.75$

it cannot be used in isolation because the ratio of the hotspot region to the total area is not contemplated in its calculation, and may lead to misinterpretations, depending on of how the model was generated.

As attested by Chainey *et al.* (2008), KDE is considered the best choice in predictive crime mapping. However, the predictive power of models generated with KDE may vary, depending on how its parameters are adjusted. Hart e Zandbergen (2014) assessed how much variation in KDE parameters influences their predictive ability. The authors worked on data from the US city of Arlington, Texas, and studied four types of violent crime: theft, aggravated theft, commercial burglary, and vehicle theft.

KDE has 3 parameters to vary: grid cell size, bandwidth size, and interpolation function. For cell size, the authors selected three values: the smallest side of the map divided by 250, the smallest side of the map divided by 150 and one-third of the medial side of the city blocks (equivalent to the smallest side of the map divided by 320, to the city under study). The authors varied bandwidth by 1/4 mile, 1/2 mile, and 1 mile. The interpolation function was varied in 4 ways: Gaussian, triangular, quartic, and uniform. Using the 2007 data, they produced one KDE map for each type of crime, for the 36 parametric variations, totaling 144 maps. In each map, the authors selected the 2.5% of the highest density cells, forming the hotspot of each map. The hotspots were compared with the 2008 crimes, where Hit Rate, PAI (CHAINEY *et al.*, 2008) and RRI (LEVINE, 2008) indices were calculated.

The authors presented index results for parameter variations and also listed a set of recommendations. The authors found no significant differences between the various cell sizes, found that the smallest bandwidth showed the best performance and that the quartic and triangular interpolation functions created the best predictive maps. The aggravated assault was the most predictable crime.

This work was the first to investigate KDE parameters specifically, and yielded impressive results, and we believe they could be even better if the authors had extended the work a little further as it will be discussed below. The authors reported no significant difference between cell sizes. In fact, there was no difference because the size of the cell interferes only if it is very large. In this case, the model loses accuracy because of the lack of resolution caused by the large cells; it could lead to misclassification of the crimes on the edges. As the cell sizes tested were all small, there was no inaccuracy. The authors evaluated only three *bandwidths* and pointed out that the smallest one presented the best results. The fact that the best is the smallest raises a question: if the bandwidths were even smaller, would it get even better results? Interesting would be to test with a wider range of bandwidths, starting with very small numbers (possibly with bad results) to large numbers (also possibly not the best results), and thus finding the maximum point. The authors generated the KDE data from 2007 and compared it to the full data from 2008. Since KDE maps are generated to make predictive police in a much smaller time span than one year, predictive maps could be compared against data from a shorter period, such as one month or 1 fortnight for example. In this scenario, the results would be more realistic and make it possible to generate more models with the same data.

3.1.2.2.4 Self-exciting point process modeling of crime

Mohler (2014) presented a prediction model for homicides for the city of Chicago that combines the concepts of retrospective and predictive hotspots. The author drew on a work by Gorr e Lee (2012) to split the hotspot into two components, one chronic, which would be stable, persistent over time and commonly known to the police, and a temporary one, linked to the phenomenon of repeat crime, which occurs when success in one crime triggers, for a while, other similar crimes in the same region. Mohler argues that abundant data are usually available for estimating the chronic hotspot since police usually have years or more of data, but for the temporary hotspot, calculated through the prospective estimation, the concept of bandwidth in the temporal dimension (usually months) limits the amount of data, leading to low accuracy models. To work around this problem, Mohler proposes to use other minor crimes as predictors of more serious crimes. Although arguing for the use of less serious crimes for the temporary hotspot, the author also uses this data for the calculation of the chronic hotspot (MOHLER, 2014).

The author has defined a model for homicide prediction composed of a mixture of two other models, the first for the chronic part and the second for the temporary part. The chronic part of the hotspot was estimated by a Gaussian kernel where the contribution of each crime was weighted by its type. The temporary part was estimated through a kernel with Gaussian concepts for space and exponential for time decay. In the temporary part, another parameter is the number of repeat crimes triggered by each type of crime. To estimate this large number of parameters (temporal bandwidth, chronic part spatial bandwidth, temporary part spatial bandwidth, weights for each type of crime and number of repeat-triggered crimes for each type of crime), the author applied a variant of the algorithm Expectation-Maximization.

Mohler tested his model with Chicago data from 2007 to 2009 to predict the crimes of 2010, data from 2008 to 2010 to predict the crimes of 2011, and data from 2009 to 2011 to predict the crimes of 2012. At the prospective stage, to calculate the hotspot for the day d all data up to day $d-1$ were used. Aggregated results show that the retrospective-prospective hybrid model offers better predictive performance than chronic, homicide-only, and pure prospective hotspots for both 75-meter and 150-meter grids.

Mohler's work uses the chronic and temporary hotspot division, which seems to be coherent, and develops a model for each division. The main contribution seems to be the weighted input of various types of crimes into the calculation of the two parts of the hotspot, and the way that the parameters were estimated: a sophisticated variation of the Expectation-Maximization algorithm. The author offers a comparison with the traditional and prospective (BOWERS *et al.*, 2004) techniques, but does not offer the PAI (CHAINEY *et al.*, 2008) index of his models, which would make it easier to compare with other models.

3.2 Geographic Information Systems and Visual Analytics Systems

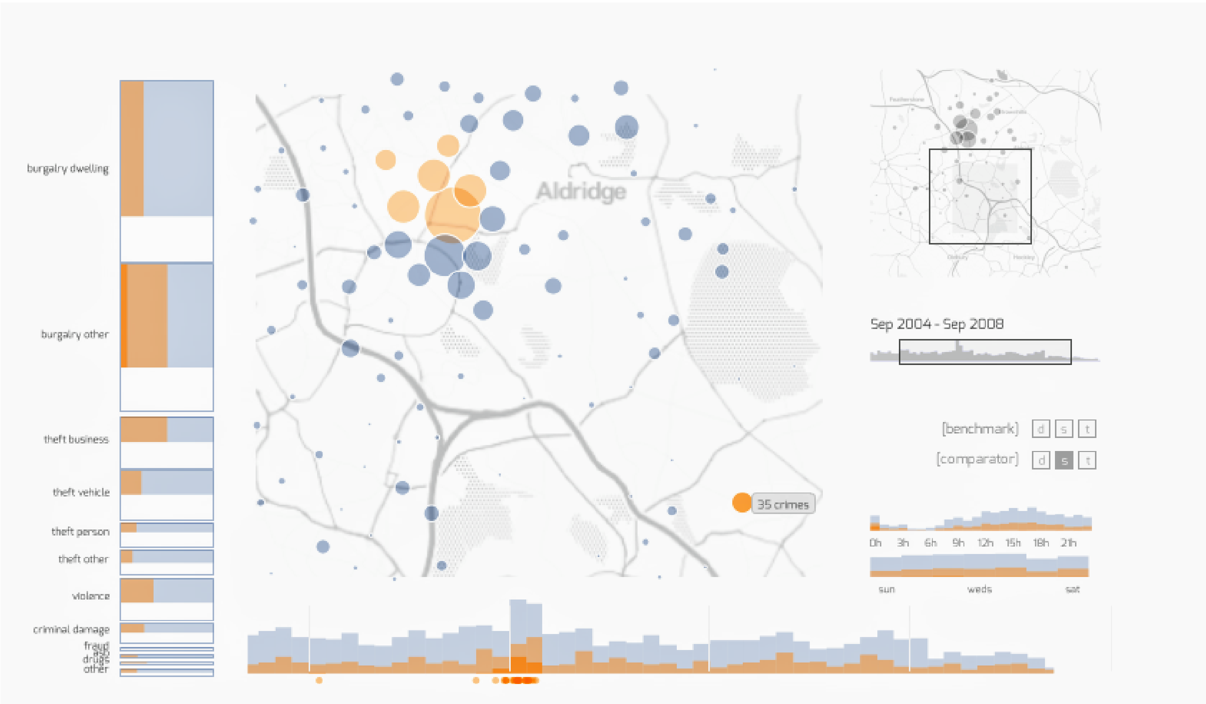
Most police departments rely on a general-purpose Geographic Information Systems (GIS) (ESRI.COM, 2015; Quantum GIS Development Team, 2017) to conduct spatial analysis of crime data. A general-purpose GIS usually have a very wide-ranging functionality, offering an integrated environment to perform all tasks of a production cycle, like data acquisition, management, analysis, and reporting.

General-purpose GIS usually can generate static KDE hotspot maps using historical crime databases. However, since such maps are suitable for visual inspection of a single situation, to compare several KDE maps at once is difficult (SANTOS, 2012) because users cannot overlap them efficiently, due to occlusion. An alternative solution would be to examine maps side by side, but, for that to be somewhat acceptable, systems have to provide a way of synchronizing points of view, colors and densities among maps (CHAINEY; RATCLIFFE, 2005). Unfortunately, those functionalities are not standard in GIS systems.

A few visual analytics systems were developed specifically to deal with crime data. CityProtect (Motorola Solutions, 2020) is a system that allows one to visualize crime incidents that are stored in a database, which is fed and maintained by law enforcement agencies from the USA and Canada. As it is focused on the general public, it has a limited number of visual components: base map, icons representing the incidents, circles depicting the clusters and line and bar charts showing the trends.

Visual Analytics for Sense-Making in Criminal Intelligence Analysis (VALCRI) is a semi-automated visual analytics system for crime intelligence analysis (WONG, 2018). VALCRI's user interface is based on the concept of thinking landscape (WONG; KODAGODA, 2017), where near regions give the details, and further away regions give the context. Figure 9 shows the VALCRI interface for crime analysis, which shows concentrations of events as a set of circular clusters, accompanied by a bar chart that shows the quantitative variations over time (BEECHAM *et al.*, 2015). As the VALCRI circular-shaped clusters are similar to spatial ellipsis, they "do not represent the actual spatial crime distribution" (CHAINEY *et al.*, 2002). Therefore, it is not the usual representation of hotspots used by police departments.

Figure 9 – An interface for Crime Analysis present in VALCRI project



Source: Beecham *et al.* (2015).

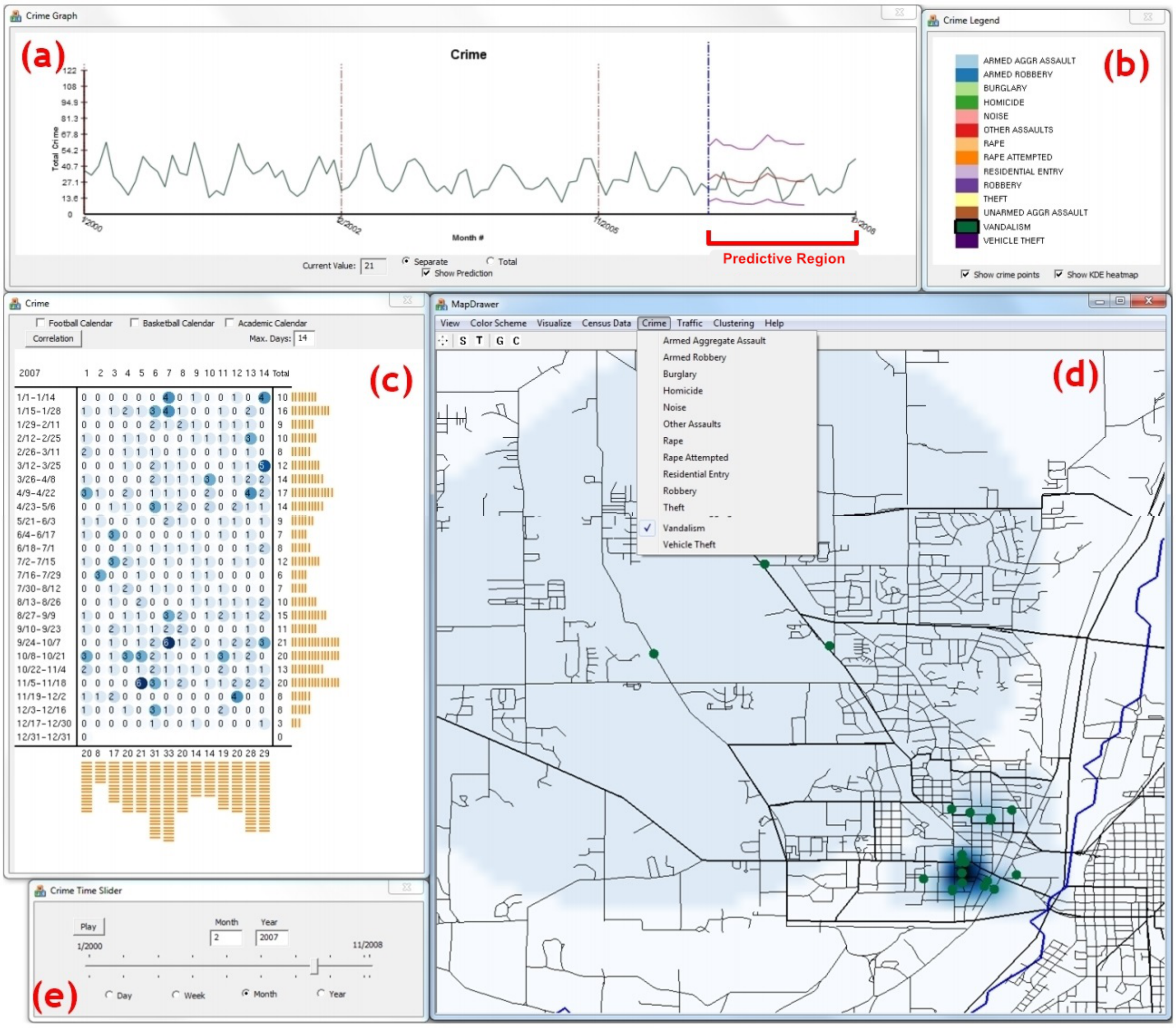
CrimeVis is an interactive visualization system for analyzing criminal data (SILVA *et al.*, 2017). The system is based on choropleth maps for verification of concentrations, clustering algorithms for classification of the zones according to their level of criminality, and parallel coordinates charts for multi-dimensional analysis comparing jurisdictions. CrimeVis is an ongoing project that still lacks tools such as KDE, spatial filters, and annotation system.

Malik *et al.* (2010) introduce Visual Analytics Law Enforcement Toolkit (VALET), a system that offers users, through a visual and interactive interface, the ability to perform explorations and analyses on spatiotemporal data related to police activity. Moreover, the system is capable of absorbing data sets from various areas, such as population senses, urban zoning, weather reports, and event calendars, to assist in analytical activities. Visually, the system presents itself as a set of interactive and mutually interconnected views (see Figure 10), all of them responding synchronously to user interaction. In the spatial dimension, the system includes mapping of police occurrences, with point maps, thematic maps using the city administrative divisions and heat maps, which are generated through KDE with the Epanechnikov kernel[2] and a variable bandwidth, found by the distance to the k-th nearest neighbor (Figure 10d). Criminal occurrences are displayed on the map as colored dots with their colors following a legend coded by type of crime (Figure 10b). In the temporal dimension, one view is a flexible calendar developed based on a work by Van Wijk e Van Selow (1999), which indicates shadow-shaped trends in days and histograms for each cycle (week, fortnight, etc.) (Figure 10c). Continuing in the temporal dimension, the system presents a line chart with the number of crimes relative to both retrospective and predictive period (Figure 10a). Finally, the system has a sliding bar component where the user can quickly change the data under analysis based on a time aggregator (day, week, month or year) (Figure 10e). In the prediction area, the system uses the Seasonal-Trend decomposition based on Loess (CLEVELAND *et al.*, 1990) method to decompose the time series of crime into its components linked to the day of the week, the time of year and the inter-year variation and the remainder as a means of identifying predictable variations. Finally, VALET has a clustering module, where it uses the nonparametric spatial proximity algorithm AMOEBA (ESTIVILL-CASTRO; LEE, 2000) to segregate crime into groups.

The authors do not present an explanation for the use of the Epanechnikov kernel, since the market, through the manufacturer ArcGis (ESRI.COM, 2015), and other publications about crime mapping (WILLIAMSON *et al.*, 1999; CHAINEY *et al.*, 2008; ECK *et al.*, 2005; HART; ZANDBERGEN, 2014) make no mention of the *kernel* Epanechnikov and mainly indicate the *kernel* quartic for this purpose. The authors do not specify whether VALET already brings the predefined k (for the k-th nearest neighbor) or whether it as an input parameter in the process of finding the bandwidth for the density calculation. VALET also makes no mention about the time of the crime, which could be used on the assumption that the criminal would select the most appropriate time for his actions. An important caveat about the kernel with variable bandwidth by nearest k-neighbors is the possibility of generating 0 (zero) density "valleys"

[2] Epanechnikov's function is defined as $k(u) = \frac{3}{4}(1-u)^2$.

Figure 10 – Visual Analytics Law Enforcement Toolkit (VALET)

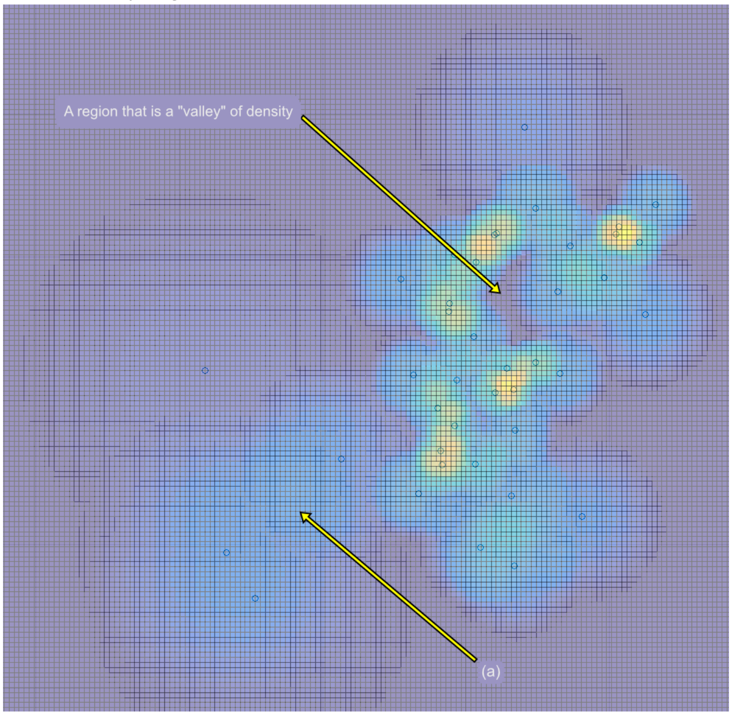


Source: Malik *et al.* (2010).
VALET is a criminal data analytic visualization system composed of a set of interrelated views. **a**: Online graph of crime count per time; presents both the retrospective and the predictive periods. **b**: Color caption for points on the map that represent crimes. **c**: Calendar used to select both the period under review and look for insights. **d**: map of the region under analysis, where VALET plot crimes, heat and thematic maps. **e**: Time frame sliding selection tool, where the analyst selects the time frame unit through a sliding component.

between event-filled regions (see Figure 11).

Malik *et al.* (2012) presented a new increment on their analytic visualization environment (MALIK *et al.*, 2010; MALIK *et al.*, 2011), now aimed at improving the search and exploration of spatiotemporal correlations (see Figure 12). As an evaluation criterion of correlation, the authors maintained Pearson's product-moment correlation coefficient, validating its significance through the two-tailed *t-student* test at 95%. Exploration of temporal correlations occurs initially by selecting two time series to be confronted and selecting the temporal window of calculations, which is the period between dates that will be considered. The tool allows users

Figure 11 – Kernels with varying bandwidth



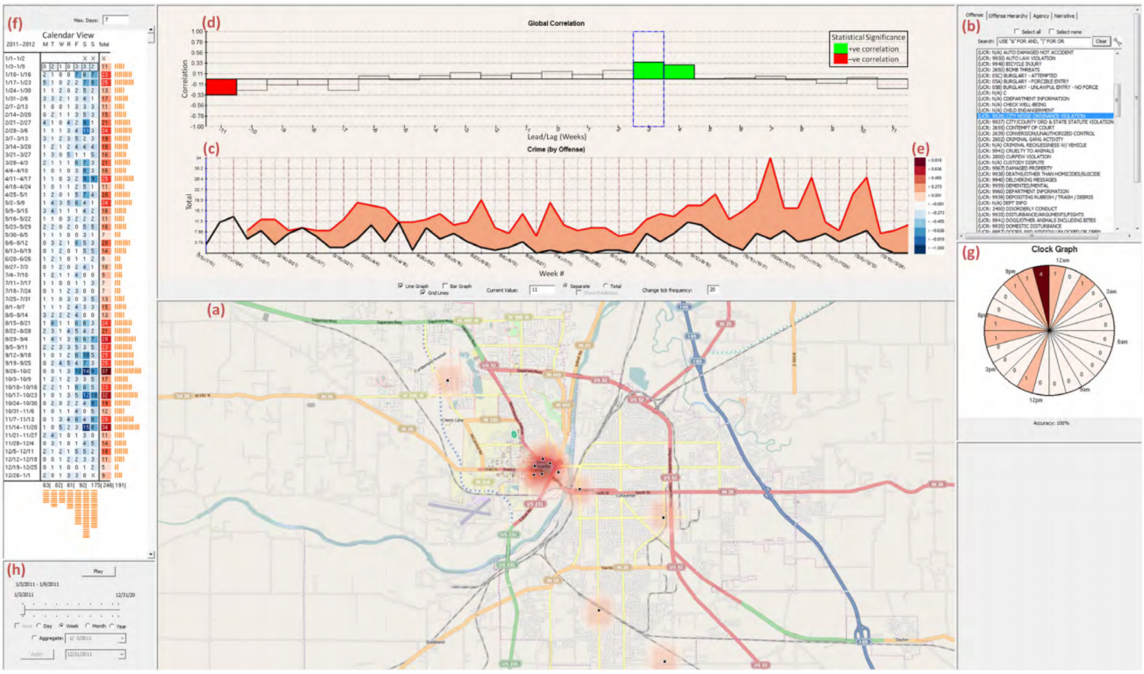
Source: the author.
Kernels with bandwidth varying with the k-th nearest neighbor may generate “density valleys” (regions without density). Note that region (a) is perceived by the kernel to have a higher density than the “valley” region, even though it is close to crime-ridden regions. The problem can be worked out with a better selection of the *k* factor, causing the *bandwidth* to increase, ending with the “valley.” The analyst should be careful about excessive large *k* factors, which can lead to very flat maps, with low contrast.

to adjust the time series, taking into account the *lead / lag*[3], which occurs when a time series has consequences on another, however with time delay.

The *lead / lag* adjustment can be done manually through a drag-and-drop interaction with the series, or it can be done automatically by the system, searching for the adjustment that offers the maximum correlation. Finally, the correlation is calculated, and a significance test result is presented. The system also has an automated correlation finder, searching for correlations through brute force, from a set of time series, also automatically adjusting *lead / lag*. Regarding the search for spatial correlations, the system allows arbitrary selection of regions, either through user-drawn polygons or even selected through pre-established administrative divisions. Events can be presented within a time frame offset by a user-selected lead / lag or

[3] A hypothetical example of the *lead / lag* phenomenon would be the correlation between explosive theft and use of explosives in break-ins. If the sequence of explosive theft is compared with the sequence of burglary using explosives without adjustment, the correlation would probably be low because the use of the explosive requires some planning by the criminals as there may be delays between obtaining the explosive and its use. If we consider an average delay of 1 week between the actions of obtaining the explosive and its use, higher correlation rates can be obtained by temporally advancing the explosive use sequence by 1 week.

Figure 12 – Improvements in VALET by Malik *et al.* (2012)



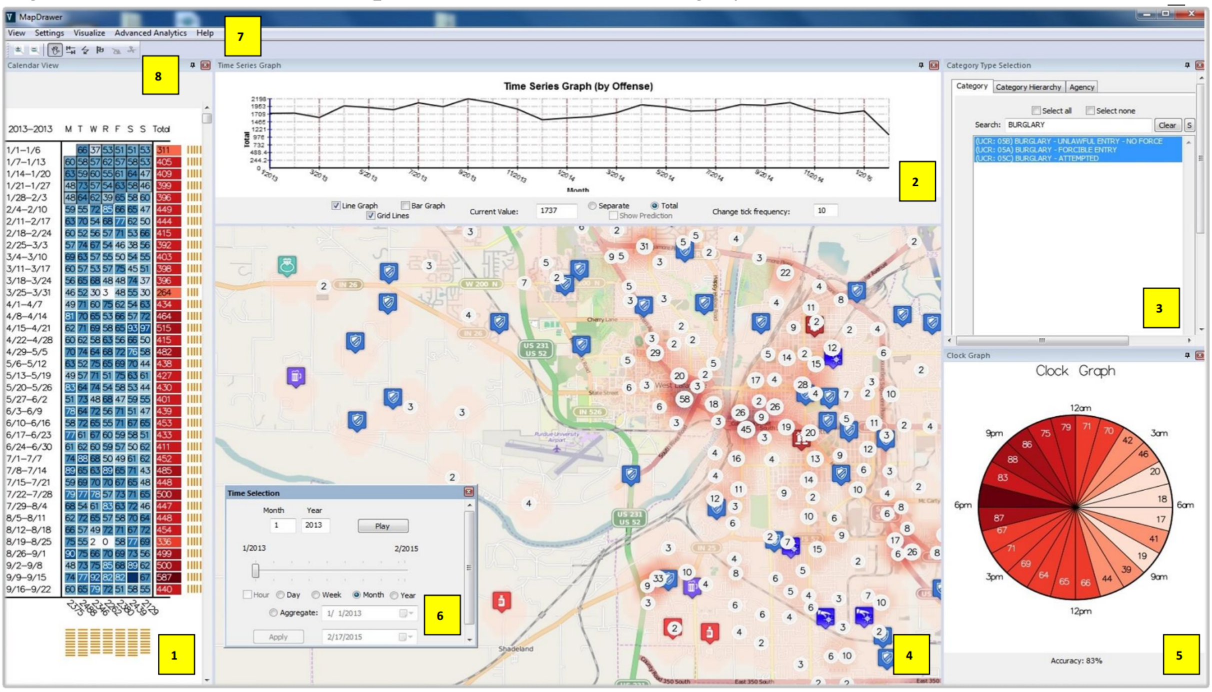
Source: Malik *et al.* (2012).
The analytical visualization system of Malik *et al.* (2012) now brings improvements in the search for spatiotemporal correlations. The system maintains the occurrence map (a), a new dataset picker (b), the linear occurrence count viewer (c) which now displays in the space between the two time series the color that indicates the correlation coefficient (e), the *lead/lag* phenomenon analysis tool (d), time selection calendar (f), a clock tool (g), and the temporal aggregation selection tool (h).

from the time correlation search tool. The search for spatial correlations occurs through visual inspection by the analyst. Finally, covering a lack of previous work, the authors included the concept of the time of the crime, and analysts can also segregate events by this variable, enriching the possibility of creating hypotheses.

The authors' work incorporates into an existing product a simple but quite complete framework for searching for correlations, especially temporal ones. The system also collaborates in the search for spatial correlations but does not offer any automated tools for identifying such correlations.

Malik et al. (MALIK *et al.*, 2014) presented an analytical visualization system for predictive activities and resource allocation for use in law enforcement agencies. The framework is placed as an evolution of previous works (MALIK *et al.*, 2010) (MALIK *et al.*, 2012), adding more features and positioning the product as an environment for proactive decision making (Figure 13). The system maintains its visual interactive characteristics for spatiotemporal data through a set of interconnected views, maintains prediction through the Seasonal-Trend decomposition based on Loess (STL) (CLEVELAND *et al.*, 1990) but uses it in a new concept, called Natural Scale Templates, which allows users to analyze their data at the spatiotemporal granularity level they find most appropriate (daily, weekly, monthly, etc.). The authors kept their Kernel Density Estimation with a variable bandwidth based on the distance from the nearest k-th neighbor, but added another kernel that calculates the density of a point based on its position with respect to a two-dimensional Gaussian kernel, with covariance matrix determined by the covariance between the latitudes and longitudes of the nearest k-th occurrences and the mean vector formed by the latitudes and longitudes averages of the same occurrences - the user decides which of the 2 *kernels* to use. The authors also added a neighborhood similarity detection feature based on data from demographic senses. Similarity detection is useful if analysts find, in some situations, for lack of police occurrences, regions with low statistical significance. In this case, the authors argue that demographically similar regions usually have the same trends for some types of crime, and analysis of one region may be used in anothers.

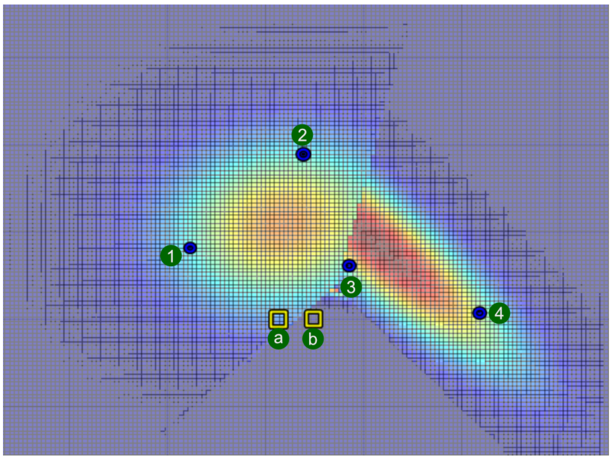
Figure 13 – Environment for proactive decision making by Malik *et al.* (2014)



Source: Malik *et al.* (2014).
In (4) we can see the map, which shows the places of the occurrences, hotspot analyzes, and thematic maps. In (1) we can see the calendar, where one selects and can analyze the temporal variables (periods under analysis, for example). In (3) we can see the dataset selection tool. In (2) we can see the line chart with counts per time. In (6) we can see the time aggregation unit selection tool. In (5) we can see the time selection tool. Image adapted from the manual of VALET (VACCINE, 2015).

The system added the feature of using a Gaussian kernel to calculate kernel density estimation, where the authors argued that this new kernel would reduce stochastic variation. To the best of our knowledge, there is no kernel density estimation study to support the Gaussian kernel in the presented configuration. Anyway, calculating cell density based on a Gaussian created by the nearest k-th neighbors can lead to abruptly broken surfaces, caused by changing which k occurrences were used top compose the Gaussian, as the space window advances. Figure 14 illustrates a hypothetical example where there is a sharp break in the scratch surface, which is unnatural.

Figure 14 – Example of distortion in the crime density when using the Gaussian kernel defined by the k-th nearest incidents.



Source: the author

Example of distortion in the crime density when using the Gaussian kernel defined by the k-th nearest incidents. The example contains four incidents, and the Gaussian uses the three closest incidents for its calculation. Note that when the kernel evaluates region (a), the three closest incidents are number 1, 2, and 3. When the kernel evaluates window (b), which is very close to (a), the three closest incidents become numbers 1, 3 and 4, totally changing the density value, even though the two regions are very close. Note that where elements of the composition change, a border is created with abrupt variations, which are unnatural deformities, and occur almost to their full extent.

Other approaches (BRUNSDON *et al.*, 2007; MACIEJEWSKI *et al.*, 2010; HU *et al.*, 2018; NAKAYA; YANO, 2010) proposed predictive crime hotspot maps based on KDE variations, while Godwin and Stasko (GODWIN; STASKO, 2017) proposed a method for augmenting crime data analysis in urban spaces. They explored the concepts of paths, nodes and edges (LYNCH,) in creating mental maps, which are built in cooperation with the community, and use raw data from police departments. None of those approaches facilitate visual comparisons and do not meet the whole set of domain tasks and requirements that arouse from our collaboration with domain experts (see Subsection 5.2).

4 MARCHING SQUARES KERNEL DENSITY ESTIMATION (MSKDE)

This chapter describes Marching Squares Kernel Density Estimation (MSKDE). Section 4.1 explains the technique. Section 4.2 contains an experiment comparing MSKDE to KDE, in which it is introduced a new index, the Hotspot Anomaly Index (HAI). The results of the experiment are discussed in Section 4.3. Finally, this chapter concludes in Section 4.4, with a summary.

4.1 The MSKDE Technique

Marching Squares Kernel Density Estimation (MSKDE) is a novel technique, to quickly generate high quality contour crime hotspot maps, based on a pre-built KDE hotspot map.

The majority of the works in the literature about crime hotspot maps based on KDE (HART; ZANDBERGEN, 2014; CHAINEY, 2013; CHAINEY; RATCLIFFE, 2005; CHAINEY *et al.*, 2008) calculate the KDE surface and separate the "hot" area from the "non-hot," using a threshold, ignoring all other levels. The researchers usually were interested only in the border generated by the threshold, which classifies the regions of the map. We identified this aspect as an opportunity to improve the algorithm of calculating this particular kind of hotspot map. We already knew the algorithm Marching Squares (see Section 2.4), which can quickly generate high quality smooth isocontours for two-dimensional fields, and we also knew that the primary computational cost of KDE is proportional to the number of cells. Then our idea was to speed up the KDE algorithm by reducing the number of cells, using a bigger cell size, and applying Marching Squares to calculate the border. This way, we could previously create a low-resolution KDE to be the base of the map, replacing the pixelated KDE border with a smooth one, calculated by Marching Squares.

An MSKDE hotspot map is generated according to the Algorithm 3. Notice that as MSKDE is based on KDE, it inherits all KDE parameters. Consequently, for creating an MSKDE, the user must provide the cell size, the bandwidth, and the kernel function, in the same way as is required by KDE.

If the analysts do not want to select the contour values manually, they can use existing techniques for automatic selection. A usual choice is to equally divide the range of values of the field into a few segments and select the border values as contour values. Another option

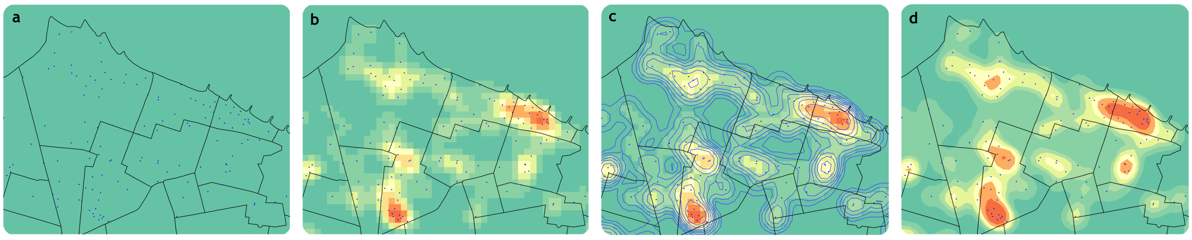
---

Algorithm 3: Generation of an MSKDE hotspot map

Input: A dataset of events **E**, the cell size **c**, the bandwidth **b**, the kernel function **k**, the threshold value **t**.
Output: A set of polygons **P**.

Load the dataset of Events
KDE = kde(E, c, b, k)
MSKDE = mskde(KDE, t)
return MSKDE

Figure 15 – Overview of Marching Square Kernel Density Estimation



Source: the author.
In (a) a set of spatial events is selected for the hotspot map. In (b) the KDE hotspot map is generated from the event set. In (c) we see a set of contour lines (in blue) generated after running Marching Squares. In (d) we see the final MSKDE map, after applying an appropriate colormap to regions defined by the contours.

is to apply a one-dimensional clustering algorithm to the field values, forgetting the spatial components, partitioning them into groups, and again taking the border values as contours. In the example shown in Figure 15, the contour values were determined by one-dimensional k-means algorithm. Notice that a clustering algorithm is a good alternative to select contour values, because the border values will work as natural breaks. However, depending on the purpose of the map, percentiles could also be a good choice.

So far, when comparing the final result of MSKDE in Figure 15d with KDE in Figure 15b, it may seem that the only advantage of MSKDE is to draw smooth maps, but there are other advantages. In fact, the use of MSKDE in crime hotspot map generation is encouraged by three main aspects:

1. Research in crime hotspot map indicates interest in analysis of definite ranges of crime density (like the highest 5% or 10% of the field). That makes sense because crime hotpot maps are mainly used in predictive policing and resource allocation, where definite boundaries are easier to work with. MSKDE fulfills this requirement because the contour boundaries are definite delimiters of density levels.
2. Sometimes the pixelated appearance of KDE maps can lead to incorrect classifi-

cations of regions near boundaries. That could be solved by using a smaller cell size, but at the expense of increasing the execution time. MSKDE is better in that regard because the interpolation used to build the contours generates more realistic regions without a significant increase in execution time.

3. The smooth MSKDE delimiters bring a significant gain in accuracy. This accuracy can be converted to a gain in time, by applying MSKDE to low resolution KDE maps, speeding up the map generation and making possible to use better hotspot maps in interactive systems.

4.2 Experiment and Evaluation

This section consists of a comparison between MSKDE and KDE in a typical task of the police department of a large city: generating a hotspot map to guide resource allocation and to define police patrol routes. The considered hotspot region is the highest 5% of the field and the maps are compared with respect to two aspects: generation time and accuracy.

The experiment consists in creating a set of MSKDE and KDE maps of the same region and events, in which the values of the bandwidth and of the kernel function were fixed and the cell size was varied within a certain range to measure its effect on the execution time and on the appearance of the map.

For accuracy evaluation purposes, the maps generated by both techniques were compared with a reference hotspot map. This reference map represents the real crime probability density of the region. The ideal reference map would be a KDE map using a cell with infinitesimal size. However, since that is not practical, we use a KDE hotspot map with a very small cell size as reference map; in our case, 10 meters. All the data, parameter configuration and computer environment used are described below.

4.2.1 Data and parameters for the experiment

Our experiment was conducted with the following data and parameters:

- Spatial events: 2,916 homicide crimes occurred in a two-year period (2014 and 2015) in a large city in Brazil.
- KDE parameters:
  - Cell sizes: a range of 50 to 200 meters, with increments of 10 meters. This

wide range enables more detailed comparisons between the techniques.

    ◦ Bandwidth: 1000 meters. This value was chosen for two reasons. First, for applications such as resource allocation, a spotty small bandwidth map renders the decision-making process more difficult due to the excess of possibilities. Larger bandwidth values, such as 1000 meters, generate more continuous clusters, which are more suitable for strategic tasks (ECK *et al.*, 2005). Second, a bigger bandwidth would be more appropriate to the city size used in the experiment, which has more than 300 square kilometers of area.
    ◦ Kernel function: a quartic kernel function, following the trend of most studies in the field.

- Considered hotspot: the highest 5% of the field.
- Reference map for accuracy: a KDE hotspot map, generated from the same data, with the same bandwidth and kernel function, but with a cell size of 10 meters.[1]

4.2.2 Computational Resources:

The computer and software used to run the experiments were:

- Computer: X64 compatible PC;
- Processor: Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz, 4 Cores;
- Memory: 16 GB of RAM;
- Operational System: Microsoft Windows 10 Pro X64;
- Software: Mathworks' Matlab v. 2015b.

4.2.3 Experiment

Our experiment consists of the following steps:

1. Generate the KDE hotspot map with cell size of 10 meters, that is used as a reference in accuracy analysis.
2. Classify the highest 5% cells of the KDE as hot and the remaining 95% as non-hot.
3. Generate a MSKDE map and a KDE map for each cell size, in the range of 50 to 200 meters, with a 10-meter increment step, making a total of 32 maps.
4. For each cell of the standard KDE reference map:

[1] The generation of the KDE map for 2916 events and a 10-meter cell size took 28 minutes.

- ○ get the coordinates of point located at the cell's center;
- ○ find out whether the cell is classified as hot or non-hot;
- ○ for each of the 32 generated maps:
  - – increment a counter whenever the classification of the cell in which that point falls is different from the classification of its cell in the standard KDE reference map.

After that experiment, we have, for each map, its generation time and the number of cells of the reference map that did not receive the same classification in it. Those misclassified cells indicate the degree of malformation of the hotspot in a generated map. The hot cells that are misclassified as non-hot represent regions that will not be covered in the resource allocation, with, possibly, serious consequences. On the other hand, the non-hot cells that are misclassified as hot represent regions to which resources will be allocated unduly, indicating a waste of resources.

The map accuracy will be evaluated, indirectly, by evaluating the degree of anomaly of the map with respect to the reference map. High anomaly indicates low accuracy, and vice versa.

To evaluate the anomaly between hotspot maps, we present the Hotspot Anomaly Index (HAI), an index that indicates how a hotspot map is anomalous with respect to a reference hotspot map. HAI is defined as 100 times the ratio of the number of cells in the reference map that are misclassified in a given map to the total number of cells classified as hot in the reference map. Note that we cannot evaluate a hotspot map only by its coverage of the hot cells of the reference map, because it would be easy to obtain a "100% accurate" map by just enlarging the hot area of the map under evaluation to cover its whole area. However, a hotspot map whose hot area covers the whole map is not really a hotspot map. For this reason, a hotspot map evaluation index must be with respect to the size of the hot area in the reference map.

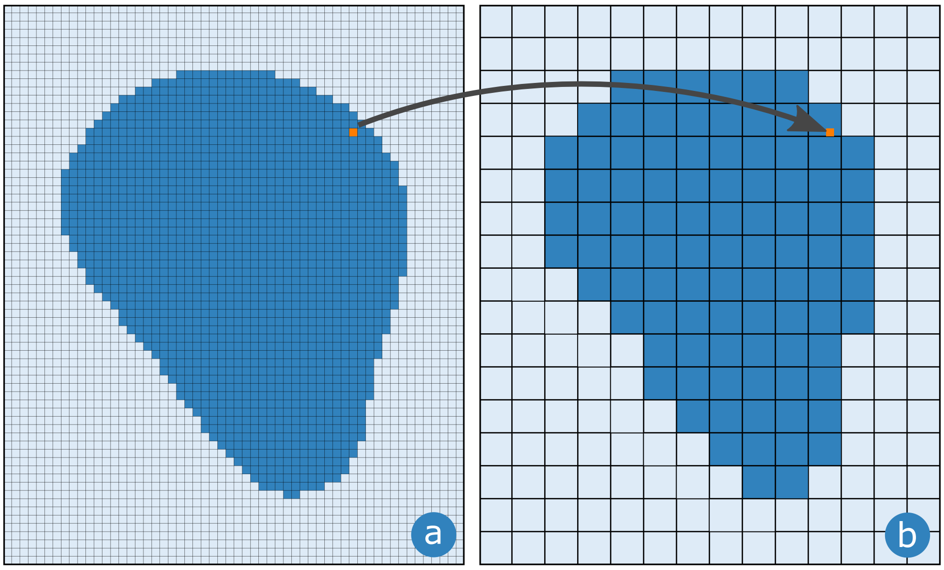
The HAI formula is defined as

$$HAI = \frac{m}{n} \cdot 100 \tag{4.1}$$

Where m is the number of cells in the reference map that are misclassified in the map being evaluated (considering the cell center), and n is the number of cells classified as hot in the reference map.

The HAI index quantifies the anomaly of a hotspot as a percentage of the hot area in

Figure 16 – Hotspot Anomaly Index (HAI) calculation example



Source: the author.
In **a**, we have a 68x56 reference hotspot map (3,808 cells) whose hot area, colored in blue, consists of 1,484 cells. In **b**, we have a 17x14 KDE hotspot map whose hot area consists of 94 cells. The HAI index of **b** with respect to **a** is calculated by counting how many of the 3,808 cells in **a** (e.g. the orange cell in **a** whose center is also marked as an orange spot in **b** pointed by the arrow) are classified differently in the map shown in **b**, dividing that count by the number of hot cells in the reference map in **a**, and multiplying that ratio by 100. In this example, 121 cells of **a** were misclassified in **b**, therefore, $HAI = \frac{121}{1484} \cdot 100 \approx 8.15\%$, which indicates that the hotspot map in **b** has an anomaly, corresponding, in size, to approximately 8.15% of the hot area of the reference map. Notice that for HAI calculation, the only information needed from map **b** is how this map classifies every cell of map **a**; the remaining information is taken from map **a**.

the reference map. Figure 16 displays an example of HAI calculation.

4.3 Results and Discussion

The hotspot map used as reference was generated and it has 2802 rows and 2422 columns; with a total of 6,786,444 cells. The highest 5% values (338,423 cells), were classified as the hot part of the map.

All 16 KDE maps and 16 MSKDE maps were generated, with execution times presented in Table 3. The same information is displayed graphically in Figure 17.

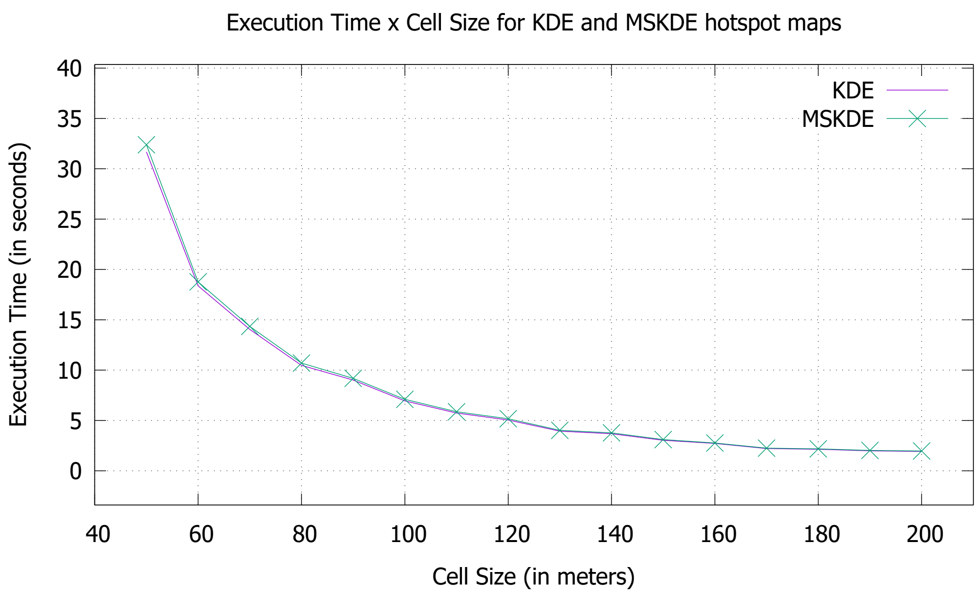
Table 4 contains HAI computed for all 32 maps and Figure 18 displays the same information graphically.

Table 3 – Execution time of each type of hotspot map according to cell size. Timings are expressed in seconds

Cell Size	KDE	MSKDE
50	31.68	32.39
60	18.37	18.76
70	14.05	14.34
80	10.47	10.70
90	9.01	9.18
100	6.94	7.09
110	5.74	5.86
120	5.04	5.17
130	3.93	4.02
140	3.68	3.76
150	3.03	3.10
160	2.71	2.76
170	2.20	2.25
180	2.13	2.17
190	1.98	2.02
200	1.92	1.97

Source: the author.

Figure 17 – Execution time of each type of hotspot map according to cell size



Source: the author.

4.3.1 Discussion

4.3.1.1 Performance

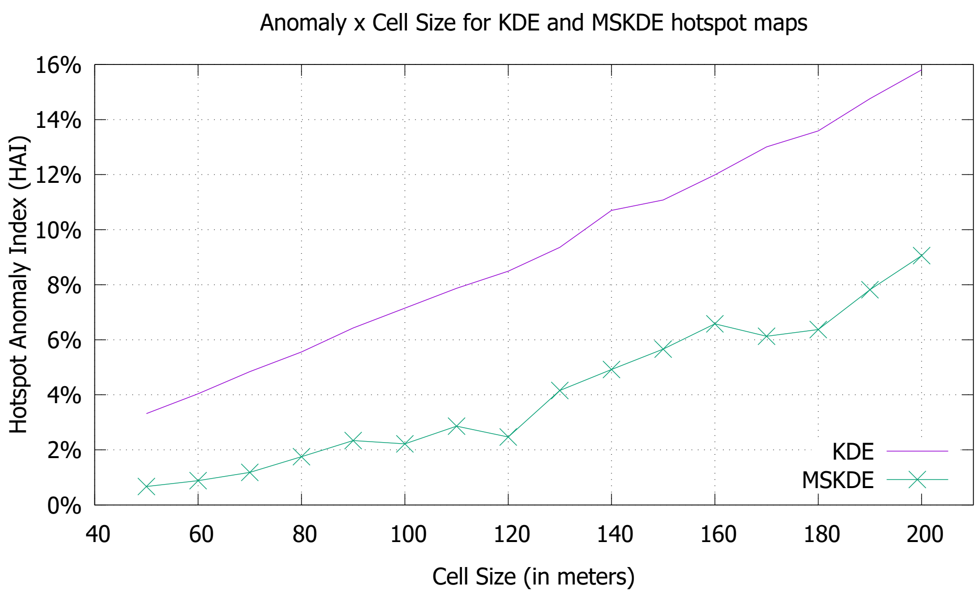
The execution times of MSKDE and KDE, for a range of cell sizes, are reported in Table 3. We observe that, on average, the execution time of MSKDE is only 2.18% higher than

Table 4 – Hotspot Anomaly Index: expressed in percentage of the hot region in the reference hotspot map

Cell Size	KDE	MSKDE
50	3.32%	0.67%
60	4.04%	0.88%
70	4.84%	1.18%
80	5.56%	1.75%
90	6.43%	2.34%
100	7.15%	2.22%
110	7.87%	2.86%
120	8.49%	2.47%
130	9.36%	4.16%
140	10.70%	4.92%
150	11.08%	5.66%
160	11.99%	6.58%
170	13.01%	6.13%
180	13.59%	6.37%
190	14.76%	7.82%
200	15.81%	9.06%

Source: the author.

Figure 18 – Anomaly (HAI) of each type of hotspot map according to cell size



Source: the author.

that of KDE. This increase is not significant for most applications and is hardly noticeable in Figure 17.

The execution time, in both kinds of maps, has exponential decay as the cell size increases, and the time cost, when the cell size is small, is so high that its use in interactive systems is not recommended. For use in interactive systems, the cell size should have a minimum

Table 5 – Execution time (in seconds) and anomaly (HAI) computed for KDE and MSKDE hotspot maps

KDE			MSKDE		
Cell Size	**Time**	**HAI**	**Cell Size**	**Time**	**HAI**
50	31.68	3.32%	120	5.17	2.47%
60	18.37	4.04%	130	4.02	4.16%
70	14.05	4.84%	140	3.76	4.92%
80	10.47	5.56%	150	3.10	5.66%
90	9.01	6.43%	160	2.76	6.58%
100	6.94	7.15%	170	2.25	6.13%
110	5.74	7.87%	180	2.17	6.37%
120	5.04	8.49%	190	2.02	7.82%
130	3.93	9.36%	200	1.97	9.06%

Source: the author.

size that does not hinder the user's experience.

4.3.1.2 Accuracy

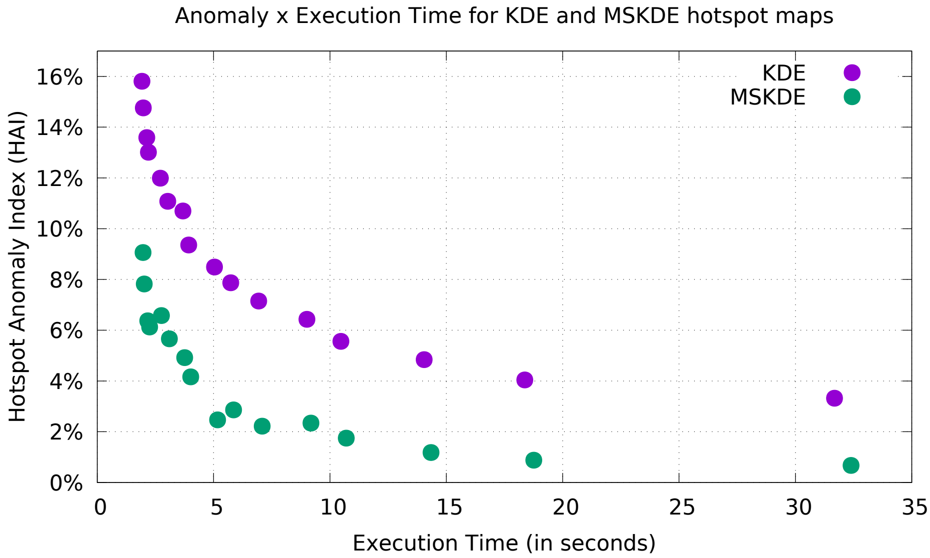
As observed in Table 4 and in Figure 18, for every cell size, MSKDE generates less anomalous (more accurate) maps than KDE.

Analyzing the chart in Figure 18, we can see that, for KDE and MSKDE maps with similar HAI value, the cell size on MSKDE maps is significantly bigger than the cell size on KDE maps. This is very advantageous for MSKDE because maps with bigger cell sizes can be generated in less time (Table 3 and Figure 17).

For example, comparing a KDE map with cell size of 70 meters and a MSKDE map with cell size of 140 meters, both maps have similar HAI values (4.84% and 4.92% respectively). However, while the KDE map is generated in 14.05 seconds, the MSKDE map is generated in only 3.76 seconds (including the time to compute its KDE). Other cell sizes can be compared in Table 5.

The relationship between HAI and execution time for KDE and MSKDE is displayed in Figure 19. Notice that MSKDE outperforms KDE in every combination, whether providing a better execution time for similar anomaly or giving a more accurate map at the same execution time.

Figure 19 – The relationship of Hotspot Anomaly Index (HAI) and execution time for KDE and MSKDE maps



Source: the author.
Notice that MSKDE outperforms KDE in every scenario, offering less anomaly at a fixed time and being faster for a fixed anomaly.

4.3.1.3 Limitations

In this experiment, we have compared MSKDE and KDE in generating hotspots map with 5% of hot area and a bandwidth of 1000 meters. The MSKDE map with cell size of 50 meters is composed by 16 polygons, with a total of 2,362 edges.

Bandwidth is a parameter that has a major influence on the number of polygons, because as we saw in Section 2.3, small bandwidths generate spotty maps, formed with more polygons. For example, if our experiment were done with a bandwidth of 400 meters, the MSKDE map would have been composed by 76 polygons, with a total of 4,502 edges. This larger number of polygons and edges can increase the execution time and delay operations such as coloring and rendering.

Other point that deserves attention is the additional cost incurred when it is necessary to plot an MSKDE map with more than one contour level. For example, in the case of Figure 15 on the right, which has 6 contour levels, the part of the MSKDE process responsible for calculating polygons, must run 6 times and the final map will have many more polygons and edges. This extra steps will add some cost to the final execution time.

4.4 Summary

In this chapter, we presented Marching Squares Kernel Density Estimation (MSKDE), a technique to make crime hotspot maps quickly. It is faster than KDE without losing precision. We also presented the Hotspot Accuracy Index (HAI). An index that measures the anomaly of a hotspot map, compared with a hotspot map considered "perfect."

5 CRIMEWATCHER - A WEB SYSTEM FOR VISUAL ANALYSIS OF CRIME DATA

This chapter describes CrimeWatcher, a web-based system that facilitates visual analyses of crime data and tracking the crime phenomena. Section 5.1 introduces the collaboration with both Brazilian and American police. Section 5.2 defines the domain tasks related to hotspots analyses. Section 5.3 presents CrimeWatcher and Section 5.4 describes SHOC, a visualization tool that strives for simplicity and ease of use in helping users to perform all the domain tasks defined in Section 5.2. Finally, this chapter concludes in Section 5.5.

5.1 The collaboration with the police

We worked with two police departments in the United States of America (USA) and one police department in Brazil to identify a set of domain tasks that would have a significant impact on police operations and to develop a visual analytics solution to address them. Our framework's design is a result of the collaborative work among my colleagues and I, hereafter called researchers, and analysts and police officers hereafter called domain experts.

We started working with the Social Security and Public Safety Department of Ceará (SSPDS-CE) in Brazil to try to improve public safety in our state. We agreed on scheduling a set of meetings where we could go over their analytic methods to have a better understanding of their work, learn their difficulties, and identify opportunities to collaborate.

After a few meetings, we identified their primary requirement: a decentralized tool for quick and easy generation of hotspot maps. They routinely used standard Geographic Information Systems (GIS) to create static hotspot maps by applying Kernel Density Estimation (KDE) (ROSENBLATT, 1956) to a historical data set of Events. Nevertheless, the process of generating and distributing the maps was admittedly inefficient. Usually, a small team of statisticians helps 22 regiments of approximately 500 police officers each to patrol 20 designated areas in the state. Even though some of the regiments have skilled personnel that is capable of using GIS tools, the time they take to produce hotspot analyses is prohibitive.

So, we proposed to develop a collaborative environment that would allow officers not only to create hotspot maps but also to share them and discuss their analyses and insights in a standardized tool. While we developed the system, we would go more in-depth in local police methods and learn other requirements.

Later on, while I was at Purdue University in the USA as a visiting Ph.D. student, two local police departments joined the project: the Lafayette Police Department (LPD) and the Purdue Police Department (PUPD). Our collaborators from the USA already made use of visual analytics tools. Those systems combine map generation tools with other spatiotemporal analysis and interactive visualization techniques (MALIK *et al.*, 2010; LUKASCZYK *et al.*, 2015; MALIK *et al.*, 2014) to enable users to explore different scenarios to obtain rich analyses (THOMAS; COOK, 2005). They adhered to the project because the solutions they had at their disposal did not fully meet their requirements.

During the first two years of collaboration, we realized that crime hotspot analysis would benefit from a tool for enabling spatial comparisons. We also recognized that such a tool should be as simple as possible in order for non-technical users to perform hotspot analysis tasks easily.

To help the police to do those tasks, we developed a visual analytics system – CrimeWatcher that also includes SHOC (one-SHOt Comparison Tool). This tool allows users to quickly make clear spatial comparisons of hotspots characterized in five domain tasks. SHOC applies set-like operations on superimposed geometries to provide interpretable analyses that are ready to be explored, annotated, and shared with others (see Figure 20).

5.2 Domain tasks and requirements

We created a system to perform the analytic tasks related to hotspots that domain experts conduct regularly. In this section, we present those tasks, and the requirements for the system.

We defined the tasks based on multiple semi-structured interviews and meetings with different people working with police agencies in both countries. Initially, we worked with the team of statisticians of the Brazilian police department, exploring their data and understanding their methods and current tools. Next, we attended the officers' periodic meetings, which are similar to a Compstat's meeting – "Crime Control Strategy Meeting." (WEISBURD *et al.*, 2003), and learned the different officers' perspectives on the data and how they planned their work. After about ten sessions and several rounds of visualization prototypes that took about one year, we determined the officers' needs and proposed a VA solution. Then, I spent 15 months in the USA to formalize a set of domain tasks, and to refine the proposed VA solution. Last, we perfected the domain tasks and the VA prototype with the Brazilian officers. The resulting domain tasks are:

DT1 **Hotspot Identification and Comparison** consists in identifying priority areas of relative importance, the hotspots, and comparing pairs that represent different periods. They need to know which areas are changing from priority to non-priority and vice-versa, and which areas continue to be a priority. This task helps the police to reevaluate the allocation of their forces continually.

DT2 **Hotspot Evolution** consists in tracking the hotspot over time, using an absolute level of likelihood as the criterion. This comparison allows officers to identify expansions, contractions, and general movements of the crime, by the level of probability. This task helps the police to track more effectively the crime in both spatial and likelihood dimensions.

DT3 **Multi-Time Analysis** consists in identifying how crime is distributed in different periods of the day. Police usually consider this information when planning their movements during the day or when designing their shifts.

DT4 **Multi-Type Analysis** consists in identifying how different types of events correlate with one another spatially. This kind of identification can be useful in different ways. First, to deal with different types of crimes may require different approaches or specialized teams. Second, it is helpful to test hypotheses, such as the correlation between drug abuse and thefts in a given region. Third, it is useful in verifying whether police initiatives are synchronized with incidents.

DT5 **Multi-Level Analysis** consists in identifying regions based on increasing levels of likelihood. This task helps the police to reevaluate the number of resources deployed in each region.

In addition to the above tasks, during the process of designing our visual analysis tool, we identified the following requirements based on the domain experts' input:

R1 The tool should allow the types of events to be analyzed individually or in groups.

R2 The tool should allow analyses to be performed over selected regions of the city (the whole city or a part of it, e.g., a neighborhood or a district).

R3 The tool should be intuitive and straightforward, so that any officer can operate it easily, with minimum training and basic computer proficiency.

R4 The tool should allow annotations and sharing, so the analyses should be persistent and shareable.

We observed that officers in both countries had difficulties in performing the Domain Tasks using their available tools. In Brazil, for example, chiefs had to present and comment on crime evolution and resource allocation every other week in a meeting. After attending more than ten of these meetings, we recognized how difficult it was to discuss and answer questions using only charts, tables, and a base map. We never saw a KDE or even a point map in their presentations. They claimed technical difficulties for not developing more elaborated analyses. In the USA, on the other hand, the police worked closely with the community, and besides tracking crime in general, they were particularly concerned about the effectiveness of interventions. For example, they would like to track the impact of a new surveillance system in an apartment complex, with recurrent episodes of housebreaking, assault, and even shot. Their VA system did not provide an easy way of comparing time frames and of tracking whether the crime episodes were diminishing or moving. That kind of task – evaluating and adjusting interventions – can take months of work. So, annotating, saving, and sharing analyses would be useful features for their system to have.

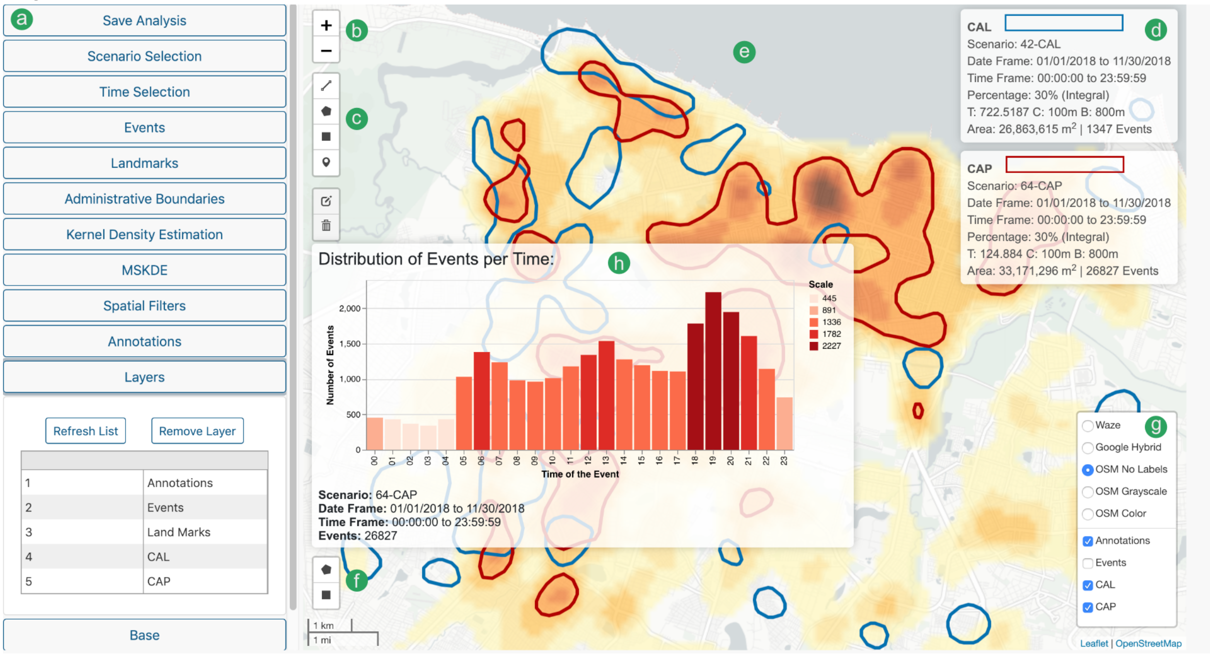
Next, we describe our approach and the design choices involved in its development.

5.3 CrimeWatcher

We developed a visual analytics system (shown in Figure 20) that provides an interactive visual environment for encouraging data exploration. Users can visualize crime events in different ways, such as: point maps, choropleth maps, and density maps (KDE and MSKDE). The maps are organized in layers and can be made visible or invisible using the layer control shown in Figure 20(g). Those maps are fully configurable (control panels are omitted in Figure 20), i.e., the users can control colors and transparency, point size and shape, line borders etc. Also, our tool includes spatial and temporal filtering components for selecting the dataset to be analyzed, and an annotation component, which enables users to annotate their findings and plans directly on the map.

CrimeWatcher has a front-end GUI and a back-end server. The front-end (Figure 20), which can be loaded on any modern Web browser, communicates with CrimeWatcher's server and accesses external graphical services and libraries as well. The server, on the other hand, is responsible for accessing the database and performing all the processing tasks, such as calculating KDE and MSKDE. Both, front-end and back-end, are described in more detail below.

Figure 20 – CrimeWatcher's main interface.



Source: the author.
(a) Control panel for filtering and visual components. (b) Map zooming tool. (c) Annotation toolbar. (d) Legends. (e) Map. (f) Spatial filter tool. (g) Layer control. (h) Histogram of events by time of the day.

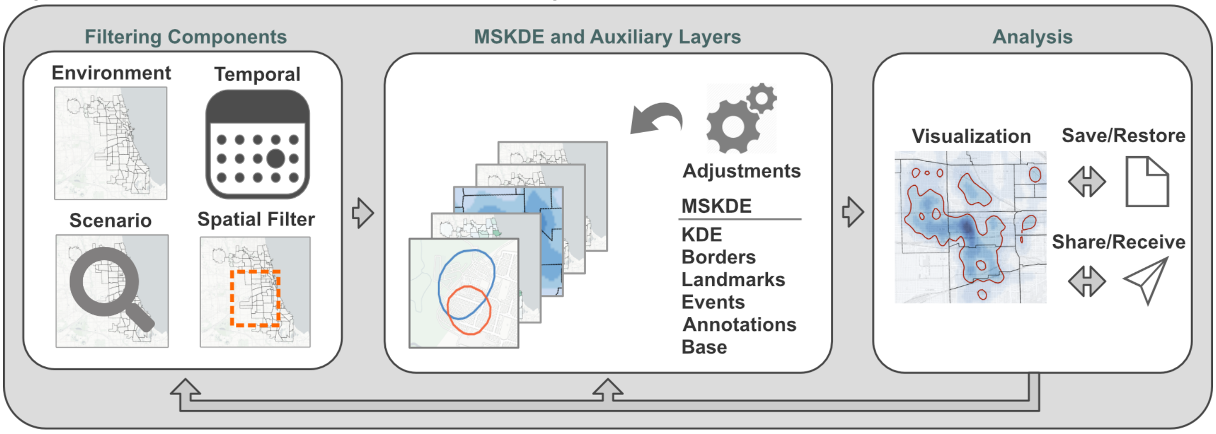
5.3.1 *Front-end*

To improve interactive exploration and analysis of crime maps, CrimeWatcher provides a layer-based approach for creating visualizations. For example, in Figure 20, the front-end shows a visualization that results from the combination of five different layers: an image-based geographic context map, a KDE layer, two MSKDE layers, and a histogram. A user can manipulate the layers at will, e.g., by reordering them, by making them visible or invisible, etc.

When using CrimeWatcher, the user follows an interactive visual-process (see the flow diagram in Figure 21) to create analyses and visualizations. First, the user selects four elements of the dataset before composing the visualization: an *Environment*, a *Scenario*, a temporal selection (date and time range) and an optional spatial filter. The *Environment* defines a spatial region (e.g., a city) whereas the *Scenario* defines the selection of a crime type (e.g., crimes against property or homicide).

Next, the user creates visualizations interactively, with no rigid sequence of steps. The user is free to include and remove layers, to change the selected *Scenario*, to modify the time range, to choose a different spatial filter and to adjust other parameters. Including more

Figure 21 – CrimeWatcher's workflow diagram.



Source: the author
The creation of visualizations is an iterative and interactive process of selections and adjustments, in a multi-layer visual approach.

than one scenario in the same analysis enables making spatiotemporal comparisons between different types of Events or different periods.

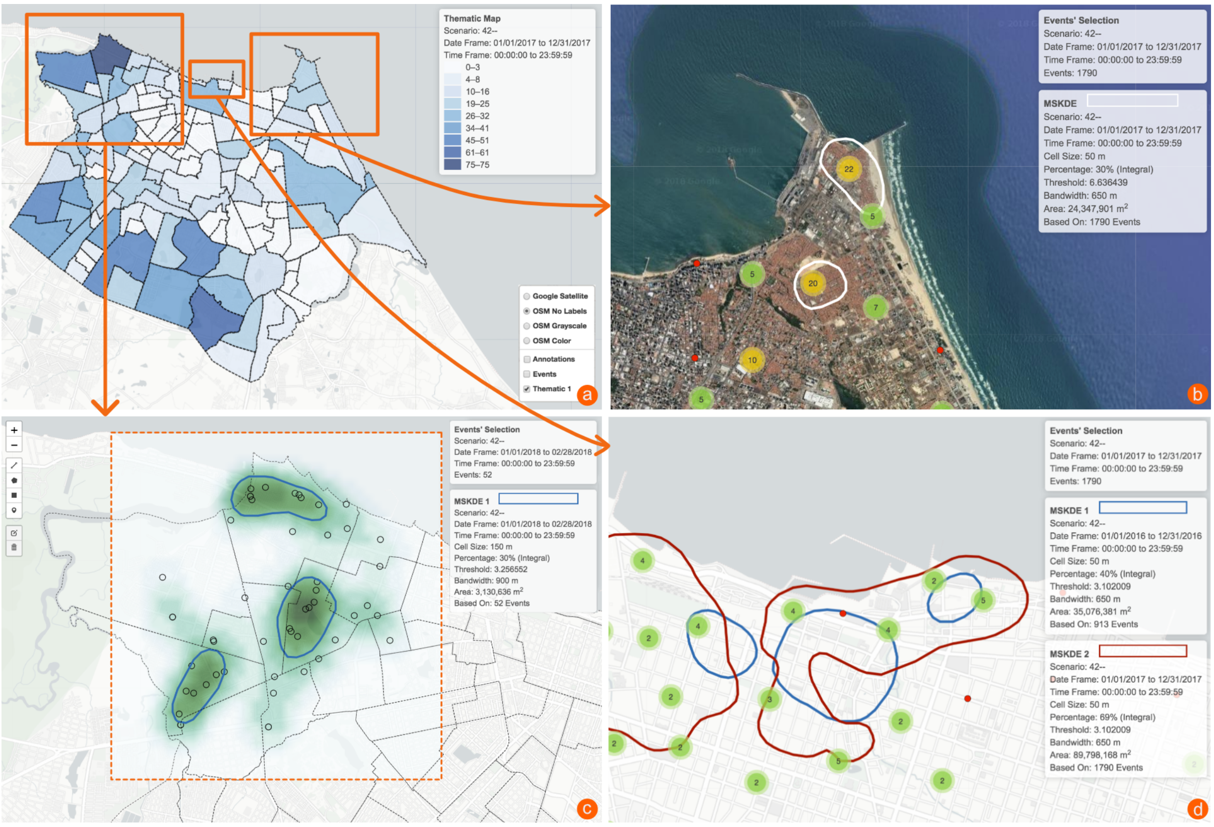
Finally, users can make graphical and textual annotations of their findings, which they can save, at any time, for future use (continuing the investigation) or for sharing with other analysts. Currently, the front-end supports five types of layers:

Base layer: CrimeWatcher places a map of the investigation region under all other layers to provide geographic context. That base map can be a satellite image (Figure 22b) or any representative image of the region (Figure 22a). Police investigators frequently use the satellite view to inspect the region where the crimes occur.

Dot map layer: Dot maps are used for displaying *Crime Events* and other geographically located elements (landmarks) that help contextualize the crimes, such as schools and ATMs. The location of each crime event receives a point marker, whose color distinguishes the type of crime. A click on a marker brings more information about the associated crime, such as date and time of occurrence, the weapon used in the crime, and any complementary description. Standard icons are used to represent the other geographically located elements. To minimize overlapping of events, markers can be clustered in larger circles, which are labeled with the number of events they represent (see figures 22b and 22d). Zooming-in on the map causes the clusters to expand. If two or more events are located in the same place, clicking on the circle will show all cases as a radial graph in which the events are connected with a line to their location, and the user can interact with them separately.

Choropleth map layer: Choropleth crime maps are commonly used when investi-

Figure 22 – CrimeWatcher's flexibility in creating and combining layers.



Source: the author

In (a), a combination of a geographic map with a choropleth map. In (b), a satellite geographic map is used for better understanding the spatial context of the region. MSKDE contours are displayed in white, improving the contrast with the background. Crimes are represented as red circles or circular clusters. In (c), a combination of a KDE map below an MSKDE map with a blue border and no fill. In (d), a more detailed image with a gray-scaled background. In this case, we show two MSKDE maps regarding two different periods of time (2016 and 2017).

gators want to communicate with a nontechnical audience (SANTOS, 2012). They are easy to understand because the colors of well-defined spatial regions represent ranges of crime occurrences. In CrimeWatcher, the user chooses any of the predefined regions (grids or administrative boundaries), or loads new spatial divisions as GEOJSON files. In Figure 22a, the choropleth map was built using the regions defined by the district divisions map.

Hotspot map layer: This layer provides two types of hotspot maps: KDE and MSKDE. KDE generates a smooth visualization of the hotspot regions; and MSKDE draws contour lines delimiting the hotspot regions, and fills them with a given color if requested. Both types of maps can be used in conjunction to build more detailed maps such as the one shown in Figure 22c. Users can choose from a variety of color schemes to build their hotspot maps.

Annotation layer: This layer offers a useful feature in CrimeWatcher, which enables the user to add annotations to the visualizations. The interface provides a powerful tool (see

Figure 20(c)) for graphical and textual annotations, with which, users can draw different shapes and lines, and place markers directly on a layer in the visualization. All graphical annotations can be enriched with text provided by the users.

5.3.2 Back-end

The back-end deals with all the analysis algorithms and functions requiring geometry processing, e.g., computing values inside regions for choropleth and hotspot maps, such as KDE and MSKDE.

A distinctive feature of the CrimeWatcher back-end is the cache of matrices. As MSKDE is based on KDE, it might suffer from performance issues, depending on the dataset and parameters. The main computational cost of MSKDE is the calculation of the two-dimensional matrix, at the KDE phase, which the system uses to calculate all subsequent MSKDEs that are based in the same dataset, cell size, and bandwidth. Moreover, every time that the system opens a previously saved analysis, it used to calculate the corresponding matrices again.

To avoid the repetitive calculations, we implemented a cache of arrays, which stores all matrices calculated by the system, with an index liked to a hash of the composition of the dataset, the bandwidth, and the cell size. This way, before calculating the matrix, the system checks if the matrix was already calculated, and uses it, in this case. This strategy revealed adequate because after the first calculation, all subsequent times that the officer uses the analysis, it opens almost instantaneously. Another example is when the officer changes the density percentage of an MSKDE; as the matrix will be the same, it is not necessary to calculate it again. The matrix cache works behind the curtains, with the database of matrices shared among all users, accelerating their work.

5.3.3 Limitations

When officers started using our system to perform the domain tasks, they spent a lot of time for completing the comparative analyses. For example, consider a simplified DT3 scenario faced by every police department: to deploy resources for preventing robbery, considering daytime and nighttime shifts.

To help solving the root problem above using hotspot analysis, officers usually raise questions such as:

1. Where are the robbery hotspots during daytime and nighttime?

2. How many robbery incidents happened during daytime and nighttime last year?
3. What regions are common to both hotspots?

The answer to Question 1 defines where to deploy resources in each particular shift. The answer to Question 2 provides quantitative information, which is related to which proportion of resources they should consider allocating in each shift. The answer to Question 3 identifies regions where robbery is stable over time – no doubt; knowing where robbery is transient over time is useful, as well.

This simple analysis of a hypothetical DT3 case reveals that it could benefit from a spatial comparison tool, where users could examine the density level and spatial distribution of the hotspots of the shifts, separately or together, to identify correlations and coincidences.

Careful analysis of all other domain tasks shows that, similarly, they benefit from a surface comparison tool. In CrimeWatcher, these types of comparisons would sometimes take a lot of cognitive processing from the users. They would build separate maps and make computations inside their heads. To avoid this cognitive overload, we developed a novel technique, called SHOC (the One-SHOt Comparison Tool), to make hotspot comparisons in a more direct way.

Next we describe how SHOC works, including the parameters and operations involved, and also describe how to perform a SHOC analysis and how SHOC helps to achieve the domain tasks.

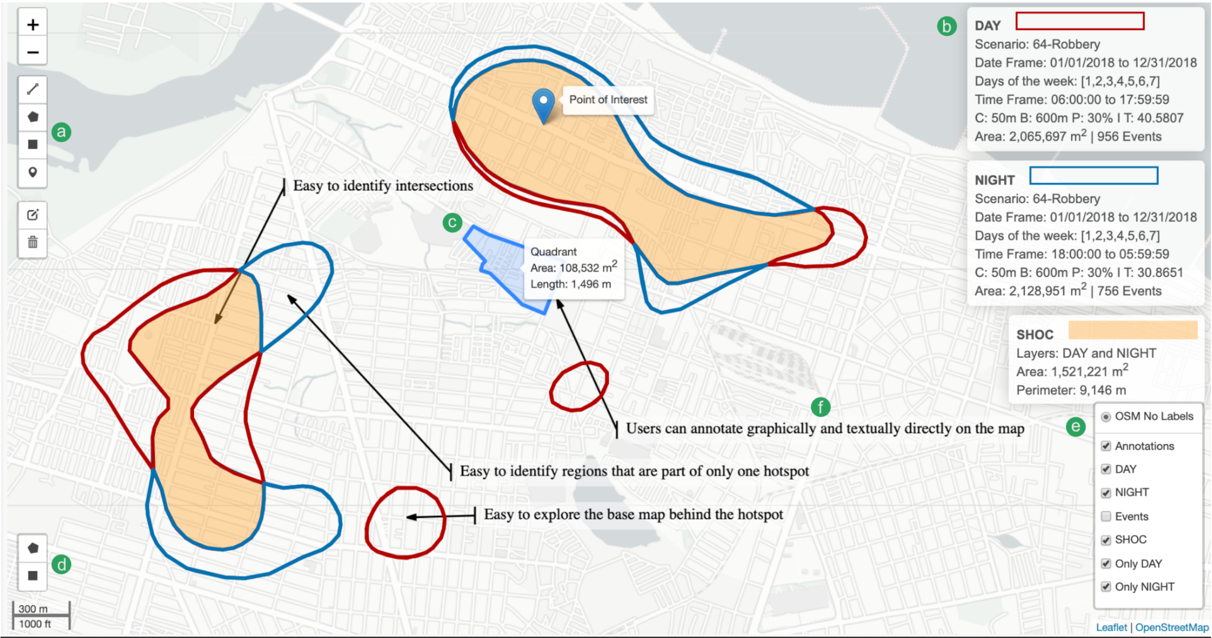
5.4 SHOC: the One-SHOt Comparison tool

Figure 23 shows a typical SHOC analysis for comparing robbery in daytime and nighttime. SHOC performs the spatial comparison directly by superimposing two MSKDEs (see Chapter 4), each one representing a hotspot map. The following properties of SHOC's design make it simple and effective:

- Computation of precise crime hotspot maps with MSKDE (QUEIROZ NETO *et al.*, 2016).
- Using of contour lines to show, clearly, where the hotspot begins and finishes.
- Superimposition of MSKDEs with minimal occlusion to enable simultaneous analysis of the hotspots, identification of correlations, and exploration of the base map that is behind the polygons.
- Displaying of immediate spatial variations between the hotspots to release the user from making any visual comparison.

- Integrating of the analysis' results in one frame (shot), including hotspots, base map, and annotations, in order to facilitate printing and taking of pictures.
- Displaying of the difference and intersection of hotspots in a familiar Euler diagram form, in order to facilitate handling and interpretation. Stable regions (intersections) are shown as solid light yellow polygons, and regions that experience changes are shown fully transparent with their original borders (Figure 23).

Figure 23 – A typical SHOC analysis.



Source: the author.
Red polygons represent hotspots for robbery in the daytime, whereas blue polygons represent hotspots for robbery in the nighttime. Other visual components: (a) the annotation toolbar, (b) legends, (c) a graphical annotation example, (d) the spatial filter toolbar, (e) the layers control, and (f) a textual annotation example.

5.4.1 *SHOC's Parameters*

SHOC has three parameters: cell size, bandwidth and contour threshold. In this section, we explain those parameters and their default values.

5.4.1.1 Cell Size

The cell size determines the resolution of the geographical space, i.e., its level of refinement, and affects only the map's visual appearance and the system's performance (see subsection 4.3.1.1 for a detailed comparison). Because cell size is independent of the bandwidth, it does not influence the levels of crime density in the cells.

When defining the cell size, the user has to take into account the trade-off between visual appearance and performance. Thus, the smaller the cell size, the higher the resolution of the map (smoother visual appearance and more refined polygons), and the worse the performance of the system will be. Our experience shows that 50 m or 100 m are values that provide a good balance between performance and visual quality. The cell size's default is 100 m.

5.4.1.2 Bandwidth

The bandwidth controls how far an event spreads its influence to nearby cells, i.e., the bandwidth defines the support of the KDE kernel function. Cells receive more density in maps with larger bandwidths, and those maps tend to form fewer and bigger density clusters, more suitable for strategic planning. On the other hand, a smaller bandwidth creates spotty maps, more appropriate for tactics. In SHOC, all hotspot maps in the same comparative analysis should use the same bandwidth so that the spatial comparison will be fair.

There is no consensus in the literature on which bandwidth value to use in general situations. Many works show values ranging from 150 m to 500 m (CHAINEY *et al.*, 2008; CHAINEY, 2013; ECK *et al.*, 2005; HART; ZANDBERGEN, 2014; CHAINEY; RATCLIFFE, 2005; BOWERS *et al.*, 2004). SHOC adopts a default bandwidth value of 400 meters, which can be adjusted depending on the analysis' purpose: if more strategic, bigger values should be used; and if more tactic, smaller values should be used.

5.4.1.3 Contour Threshold

Users can set the contour threshold in two indirect ways:

- The user specifies the integral percentage of the density that the shapes should surround, and SHOC calculates the associated contour threshold; or
- The user specifies an MSKDE already created, and SHOC extracts the contour threshold level from it.

The choice of which option to use depends on the analysis the user wants to perform. Therefore, the user must understand how SHOC calculates the contour threshold to have full control over the study. That will be explained in Section 5.4.2.

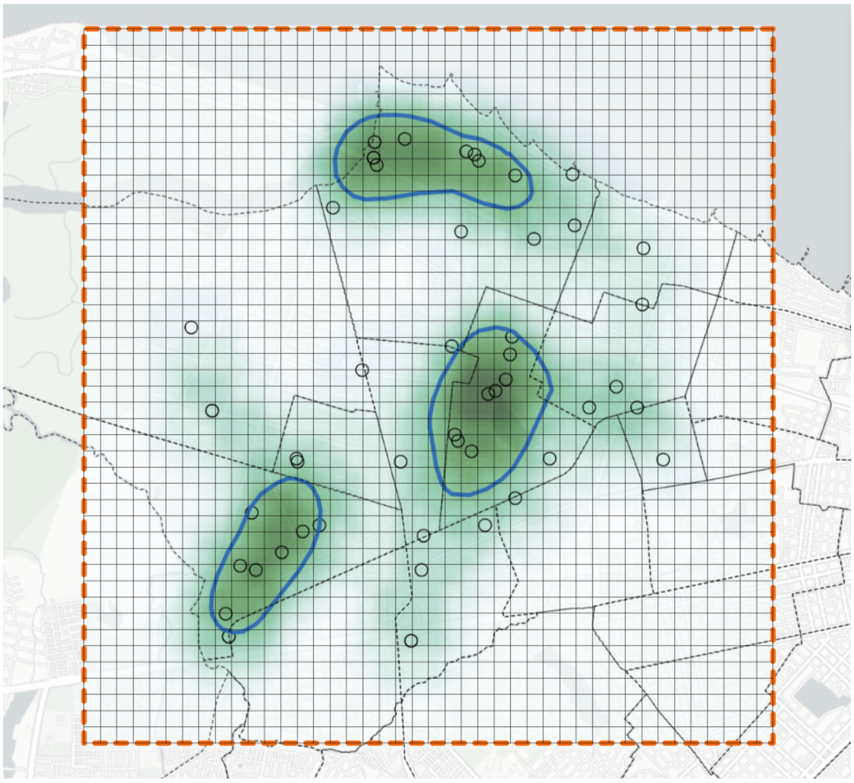
The default value of the contour threshold is associated with an integral percentage of the density equals to 30%. That value leads to a hotspot map coverage between 1% and 5% of the total area of the map, which is similar to the coverage of most hotspot maps found in the literature.

5.4.2 Computation of the Contour Threshold

In the original formulation of MSKDE (described in Chapter 4), the threshold was calculated so that polygons surrounded a certain percentage of the map area. This thesis extends the original MSKDE technique by providing two new ways of determining its threshold: first, the value is determined so that the MSKDE area includes a pre-specified percentage of the total incidents (density); and in the second one, it is a specific value, defined by another MSKDE.

Figure 24 shows a hypothetical situation to illustrate how SHOC computes the threshold. The figure shows a region with a set of fifty-two events (small black circles); a colored KDE surface computed using the set of events; and an MSKDE map (three polygons) that was generated based on the KDE surface. The region of the KDE field is delineated by a dotted orange box and contains 42×44 grid cells, 139 of which are inside the MSKDE polygons.

Figure 24 – A hypothetical situation to illustrate how SHOC computes thethreshold.



Source: the author.
An MSKDE map is composed of a set of polygons (shown in blue) enclosing hotspot regions. A KDE map is a 2D field of densities, colored according to a colormap (shown in green).

The total contribution of the incidents of the field is approximately 183.8. The *threshold* delimits three disjoint regions inside three closed level curves. The *threshold* value could have been determined by one of the two different ways previously mentioned. If the *threshold* were to be determined using the percentage of the total contribution of the incidents, the user would have specified 30%, which means that the three level curves surround a total area containing approximately 55.1 of the contribution of the incidents. If the *threshold* were to be specified directly, the user would have specified a value of 3.25, which is the value of the *threshold* corresponding to the three level curves shown in Figure 24. Since this is just an illustration, all the two alternatives would give the same *threshold* of 3.25, and, inside the level curves, there would be 139 cells that account, together, for 55.1 as the contribution of the incidents (30% of the total in the dotted region).

5.4.3 Superimposing MSKDEs

SHOC gives information that helps to perform the domain tasks by identifying variations on superimposed MSKDEs, which represent different situations. For each domain task, SHOC needs a particular change between the MSKDEs, which is obtained by varying their input parameters.

Table 6 shows all the MSKDE input variations that SHOC needs to help to perform the domain tasks, one row for each domain task. Each row of the table informs, when superimposing two MSKDEs, which inputs should vary and which ones should be kept the same to produce the right contrast between them. For example, the first row shows that, if the user superimposes two MSKDEs that share the same Type of Event, Time Frame, and Percentage of Density (indicated by the value “same” in the corresponding columns in the table), varying only the Date Frame (indicated by the value “different” in the column “Date Frame”), the resulting visualization will expose differences between the hotspots of two different periods of time (column “Result”), which will help to perform DT1 (column “Task”).

Users should be aware that, when comparing MSKDE maps regarding different date frames or time frames, periods should have the same range for comparisons to be fair. When comparing date frames, if possible, they should refer to the same period of the year, avoiding seasonality variations (MCDOWALL *et al.*, 2012).

At the end of the superimposition, SHOC also applies polygon intersection operations to compute the 3 sets of polygons that are of interest to analysts. Those operations are a subset

Table 6 – Variations on the Input Parameters to Perform SHOC

Task	data set			MSKDE	Result
	Type of Event	Date Frame	Time Frame	Threshold	
DT1	Same	Different	Same	Same Density Percentage	MSKDEs showing hotspots about different periods of time. Suitable for identifying regions with high concentration of density.
DT2	Same	Different	Same	Same Value	MSKDEs showing hotspots about different periods of time, with MSKDEs delineating at the same threshold. Suitable for tracking expansions, contractions and general movement of the density.
DT3	Same	Same	Different	Same Density Percentage	MSKDEs showing densities regarding different time frames throughout the day. Suitable for comparing shifts.
DT4	Different	Same	Same	Same Density Percentage	MSKDEs showing densities regarding different types of events. Suitable for identifying the spatial correlation between different types of incidents.
DT5	Same	Same	Same	Different Density Percentages	Similar to a topographic map, with MSKDEs showing, progressively, levels of priority.

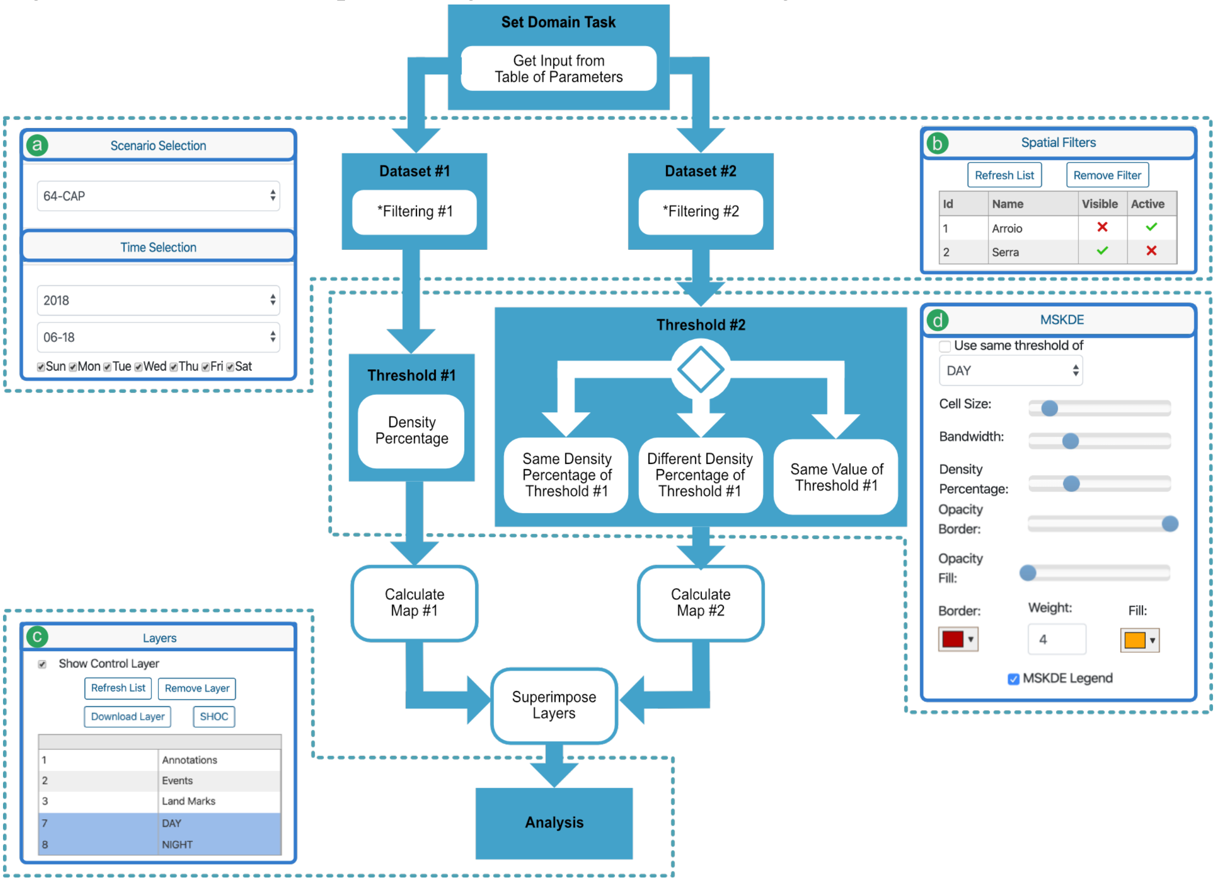
Source: the author.

of the topological relations framework proposed by Egenhofer e Franzosa (1991). Given two MSKDEs A and B, the following sets are computed:

1. Set 1: A - B, the polygons corresponding to the regions exclusive to hotspot A;
2. Set 2: B - A, the polygons corresponding to the regions exclusive to hotspot B;
3. Set 3: A ∩ B, the polygons corresponding to the regions common to both A and B.

Each set receives an appropriate visual encoding and is assigned to its own layer. Set 1 receives the same visual encoding as MSKDE A, Set 2 receives the same visual encoding as MSKDE B, and Set 3 receives a solid light yellow color. This design choice works best when only two MSKDEs are superimposed (a discussion is provided in Section 6.3.2) and favors the easy identification of the common regions without losing the context of the hotspots (ALSALLAKH *et al.*, 2014). Users can always emphasize the other regions by manipulating the visible layers, as they can hide or show layers on demand.

Figure 25 – Workflow for performing the Domain Tasks using SHOC



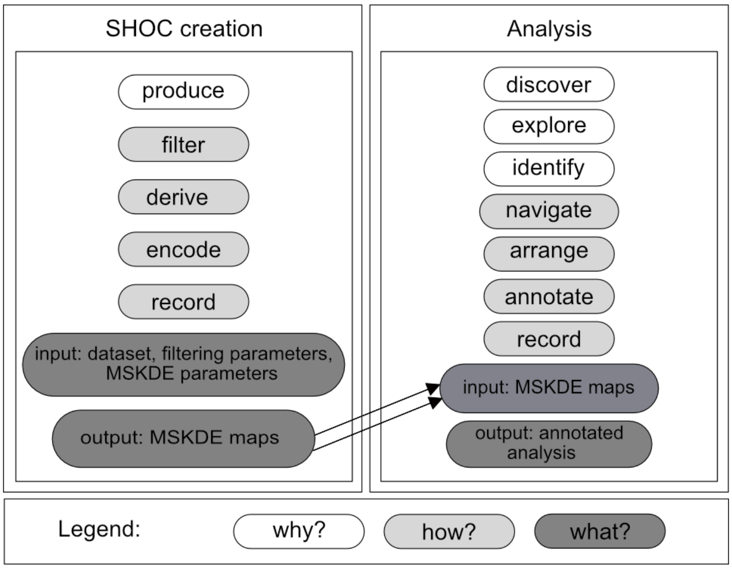
Source: the author.
The *Filtering process is the application of filters regarding Type of Incidents, Date Frame, Time Frame, Days of the Week (a), and Spatial filters (b). Users create MSKDE layers using the MSKDE panel (d), all variations in only one page. In the Layers panel (c), users select which ones they want to apply set-like polygon operations (subtractions and intersections), facilitating hotspot analysis.

5.4.4 Performing SHOC's analysis tool

In this section, we explain, in detail, how to perform any of the five Domain Tasks. Although this paper focuses on specific domain tasks, which were identified as the most relevant, we created a system that is flexible and easy of use. Therefore, we designed the interface and its components, so they are not tied to the five domain tasks, allowing users to create variations of those tasks or completely new ones.

Figure 25 shows SHOC's basic workflow and main interface components to perform any of the domain tasks as follows. First, the user chooses the domain task, learns the setup for it from Table 6, and decides what cell size and bandwidth to use in the analysis. Second, the user selects the first dataset by applying the following filters: a Scenario Selection filter, indicating the types of incidents; Time Selection filters, indicating date frame, time frame and days of the

Figure 26 – Description of the tasks using Brehmer and Munzner's typology (BREHMER; MUNZNER, 2013).



Source: the author.

week (Figure 25a); and an optional Spatial Filter (Figure 25b). Third, the user creates the first hotspot by filling out the MSKDE panel (Figure 25d) with: cell size, bandwidth value, percentage of density for computing contour threshold, and some graphical choices (colors, opacity etc.). Fourth, after the first MSKDE is created, the user selects the dataset for the second hotspot, using the same Scenario Selection, Time Selection, and Spatial Filter tools. Fifth, with the second dataset on hand, the user goes to the MSKDE panel to create the second hotspot map by using one of the three strategies of defining the threshold. Sixth, after creating the hotspot maps, the user selects the two MSKDE layers in the Layers Panel and presses the "SHOC" button (Figure 25c).

After those six steps, SHOC computes the polygon operations and visually encodes them, generating three new layers. At this point, the user can begin exploring and annotating the analysis.

Figure 26 shows a description of the domain tasks using the typology of tasks from Brehmer e Munzner (2013). According to their typology, SHOC is a production task - it produces a set of MSKDE geometries that are used as input to the analytical part of the domain tasks. The analytical parts of the domain task are dedicated to exploration, identification, and discoveries.

For every Domain Task, SHOC immediately gives the spatial differences between the datasets. Officers do not need to make any extra visual comparisons. They can proceed to the next steps of the analysis, exploring the base map, together with the hotspot maps, their coincidences, and differences. SHOC, to facilitate the analysis process, also provides a legend for each hotspot map with the parameters that the user had set, the calculated threshold, the method for calculating the threshold, the area of the hotspot, and the number of incidents used to calculate the hotspot.

5.5 Summary

In this chapter, we provided a characterization and abstraction of the hotspot analysis problem into five domain tasks and presented CrimeWatcher, a visual analytics system for geocoded crime data and SHOC, a tool that facilitates performing the domain tasks by giving a straightforward way of making spatial comparisons, using set-like operations on superimposed geometries.

6 CASE STUDIES

This chapter presents two real-world case studies to demonstrate that SHOC helps to perform the domain tasks. Section 6.1 describes the first case study concerning the problem of drug abuse in Tippecanoe County, Indiana, USA, during the years of 2016 and 2017. Section 6.2 describes the second case consisting of two analyses used by a police department in Brazil to plan the patrol of a beat (a geographic area designated by the police department) to fight Crimes Against Property (CAP) and Crimes Against Life (CAL). All cases deal with crime problems that are relevant in many cities around the world nowadays. Section 6.3 presents feedback received from domain experts and discusses our approach's limitations. Section 6.4 contains a few remarks about the differences and similarities between the Brazilian and the American police environments. In Section 6.5 we discusses the current status of CrimeWatcher. Finally, this chapter concludes in Section 6.6.

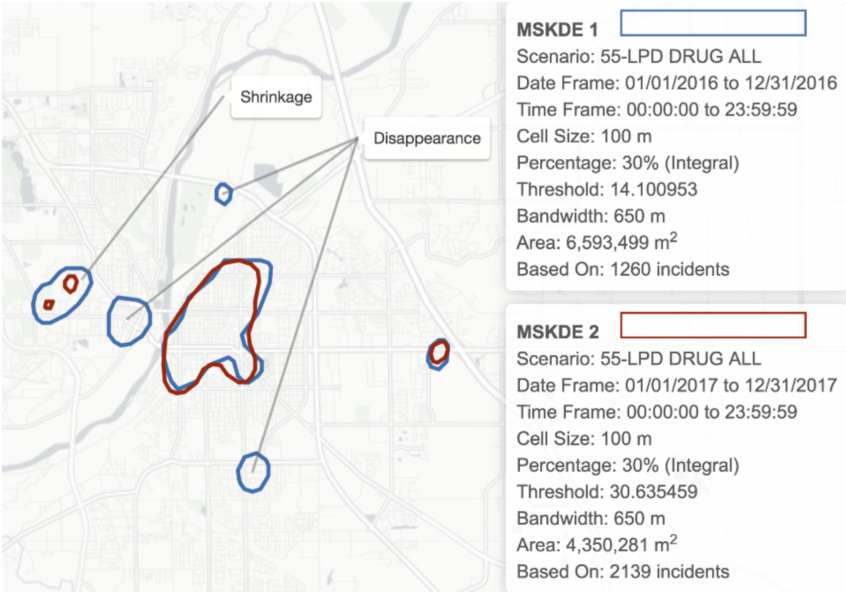
6.1 Drug Abuse in Tippecanoe County, USA (2016-2017)

This example is one of the several study cases that we developed together with the Lafayette Police Department (LPD) and the Purdue Police Department (PUPD), in a series of 8 meetings between November 16, 2017, and May 23, 2018. It used a prototype version of SHOC that did not display the differences and intersections using different visual encoding.

The police knew that the recorded number of drug abuse incidents was increasing, and they wanted to know more about how and where that increase was taking place. For making the comparisons, we partitioned the data set, which included all the incidents recorded from January 1, 2016, to December 31, 2017, into two partitions. The first partition contained the 1,288 incidents of 2016, and the second partition contained the 2,129 incidents of 2017.

We performed domain tasks DT1 and DT2, comparing the two partitions. For that, we used the following parameters: percentage of the total weighted incidents equal to 30% (that value was used to compute the threshold for both MSKDEs of DT1 and the first MSKDE of DT2); and bandwidth of 650m (that bandwidth was appropriate because the study region comprised the whole county).

Figure 27 – Hotspor Identification and Comparison in Tippecanoe County, USA.



Source: the author.
The blue polygons represent the hotspot of 2016, while the red polygons represent the hotspot for 2017. The larger number of incidents in 2017 and the smaller area of the red polygons indicate that there was a concentration of the crime density from 2016 to 2017.

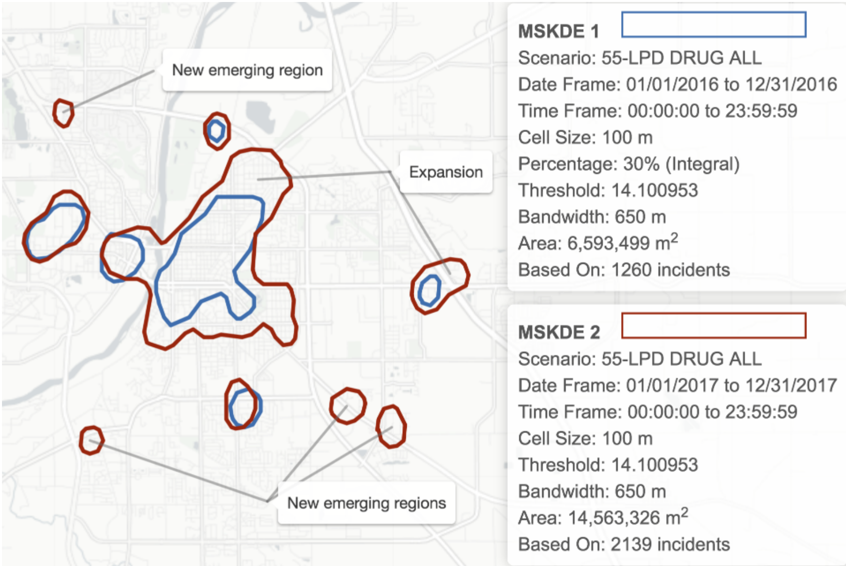
6.1.1 Hotspot Identification and Comparison (DT1)

Figure 27 shows MSKDE maps for partitions 1 (2016, in blue) and 2 (2017, in red). Despite the increase of nearly 70% in the number of incidents from 2016 to 2017, the smaller total area in 2017 indicates that drug abuse incidents are more concentrated in 2017 than in 2016. The increase in the *threshold* from 14.10 to 30.64 confirms that concentration. Notice that, in Figure 27, there are two larger polygons (one red and another blue), whose shapes are almost coincident. However, the crime density level on the red border is more than twice that on the blue border. The disappearance and shrinkage of polygons from 2016 to 2017 do not necessarily indicate a reduction of crime density in those areas, but, rather, a decrease in their relative importance concerning the entire community. Finally, the analysis suggests that the central region of the map is where the priority with regard to actions against drug abuse should be focused in the county.

6.1.2 Hotspot Evolution (DT2)

Figure 28 depicts partitions 1 (2016, in blue) and 2 (2017, in red). The comparison shows that the MSKDE of 2017 has more than twice the area of the MSKDE of 2016. That indicates the spreading, in 2017, of the regions with crime density above the level of 14.10.

Figure 28 – Hotspot Evolution in Tippecanoe County, USA.



Source: the author.
The blue polygons represent the hotspot for 2016, while the red polygons represent the hotspot for 2017, using the same threshold level of the blue polygons. The larger area of red polygons indicates the expansion of the crime density from 2016 to 2017.

The MSKDE maps also indicate changes in shape and number of polygons. The larger central polygon is expanding and enclosing the other polygons. Additionally, in 2017, there are four new emerging regions of high crime density, which did not exist in 2016. Finally, the comparison shows that, using 2016 as a reference, drug abuse in 2017 has expanded geographically, with indications of where the expansion took place.

6.1.3 Conclusions

Based on both domain tasks, we can conclude that drug abuse crime was prevalent in the central area of the city and expanded there from 2016 to 2017 with even higher concentration. Some new high concentration areas emerged in other points of the city. The substantial increase in the number of incidents indicates the need of a strategy change.

6.2 Crimes Against Life and Property in Brazil

SSPDS-CE divided one of its cities into ten beats, each with its police resources. In this use case, officers of one of the beats (Beat F) used SHOC to improve the understanding of the criminal context and define priority areas and time frames to patrol two kinds of incidents: Crimes Against Life (CAL) and Crimes Against Property (CAP).

Concerning social and economic development, Beat F can be roughly divided into three zones (see Figure 31): Zone A, in the northwest, which is the most developed area; Zone B, alongside the right border, with intermediate development; and Zone C, in the middle and southwest, which is the less developed sector. In the officers' opinion, the different social and economic levels have a significant impact on the distribution of events in the Beat.

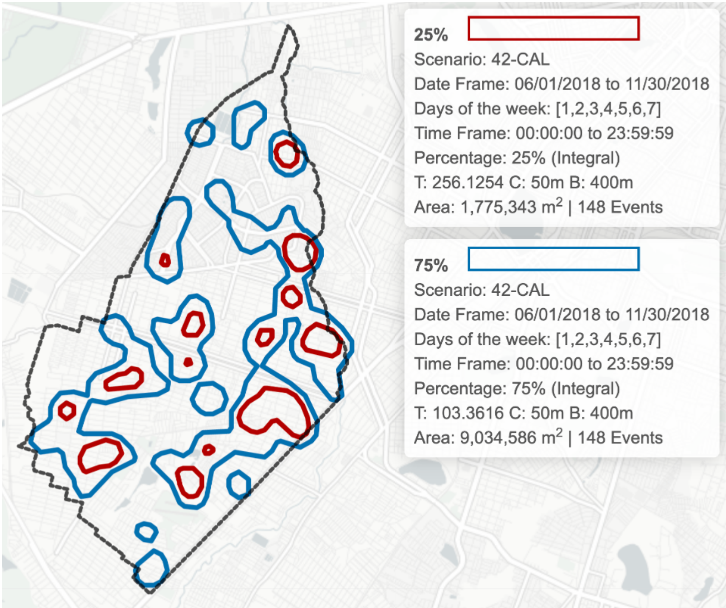
In all the analyses, officers applied a spatial filter to extract only the events inside Beat F. They also used a temporal filter: for CAP analyses they restricted events to the period from January to November of 2018; and for CAL analyses, from June to November of 2018. Due to the size of the Beat (21.01 km^2), they choose a bandwidth of 400 m.

6.2.1 Crimes Against Life (CAL)

In our initial engagement in the activities of Beat F for analyzing CAL, due to a critical situation with a consistently high number of homicides per month, the Police Department had already started a project to reduce the homicide rate, named the "Stanch Project." That project's strategy was to select parts of the Beat (called "quadrants") and a single time frame (limited to a continuous range of 6 hours) and deploy much more resources than usual there, borrowed from other Beats. The duration of the Stanch Project was only 15 days, but, due to good initial results, the Police Department was committed to launching a new phase of equal duration immediately after the first phase was finished. When we started our activities, the first phase was already in progress. Therefore, we only helped the police to define the quadrants and the timeframe for the second phase. We decided not to perform the analyses ourselves, but, instead, to offer training sessions to the officers, and let them perform the analyses and define quadrants and timeframes, with our support. Following instructions from the Chief, despite being CAL the focus of the project, whenever possible, CAP was also taken into consideration. The analytical process happened as follows:

Exploratory analysis: Following our recommendation, officers conducted an activity in which they were free to create any analyses for the CAL and CAP scenarios. The purpose was to make the officers more aware of the spatiotemporal distribution of incidents, and to promote discussions among themselves and to let them draw comparisons with their a priori knowledge. The result was surprising: they created dozens of analyses, even exploring some possibilities that we had not included in the training sessions, such as an expanded spatial filter to examine the vicinity of the Beat. As an example of their exploration, Figure 29 presents

Figure 29 – DT5 analysis for the CAL scenario in Beat F.



Source: the author.
Red polygons surround areas that comprise 25% of the crime density, while blue polygons surround regions that include 75% of the crime density. The spaces between the red and blue borders account for 50% of the crime density and the rest of the Beat, 25%. Officers can identify regions at three different levels of crime density, facilitating resource allocation.

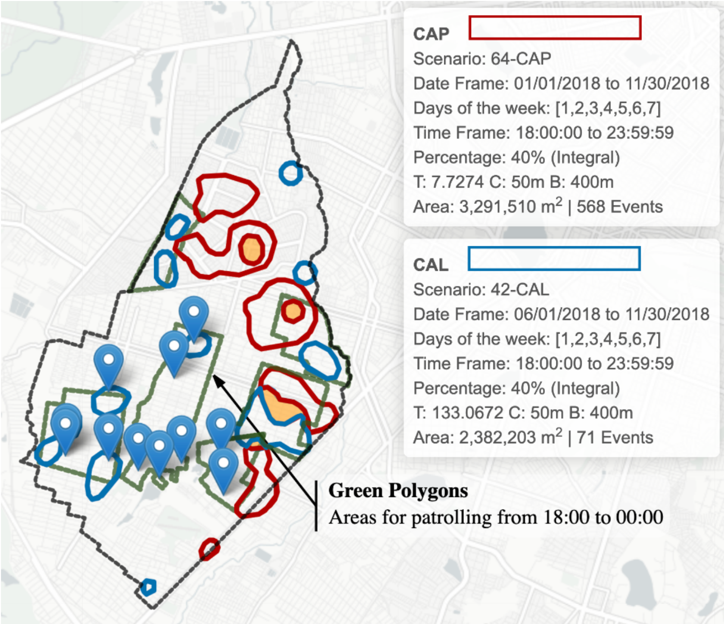
a multi-level analysis (DT5) for the CAL scenario with two percentages of the total weighted incidents: 25% and 75%, where they could compare SHOC outlines on resource allocation to their current policy.

Date frame and time frame definitions: In order not to lose the perception of the most recent homicides' dynamics in Beat F, the officers decided to use only the last six months of available data: from June to November of 2018. They defined the time frame as the hottest six-hour continuous interval of the CAL histogram, which was the 18h-24h time frame.

Quadrants' definition: The officers created a multi-type analysis (DT4) that included both CAL and CAP and used 40% of the total weighted incidents. Next, they compared the DT4 study with the quadrants of the first phase of the project. Finally, they determined the new quadrants based mainly on the CAL layer of the DT4 analysis. However, they adjusted those quadrants to contemplate some parts of the CAP layer as well as a few regions outside the DT4 polygons that would benefit from the police's presence because of the high tension between rival gangs.

Figure 30 shows the officers' analysis for the definition of quadrants. The officers used the annotation system to draw the new quadrants (green borders) and to mark the regions of high tension (blue icons). Due to the small size of the image, we omitted the textual annotations

Figure 30 – Multi-type analysis (DT4) to help define quadrants.



Source: the author.
Quadrants have green borders. Blue icons indicate high-tension places that were included based on a priori knowledge.

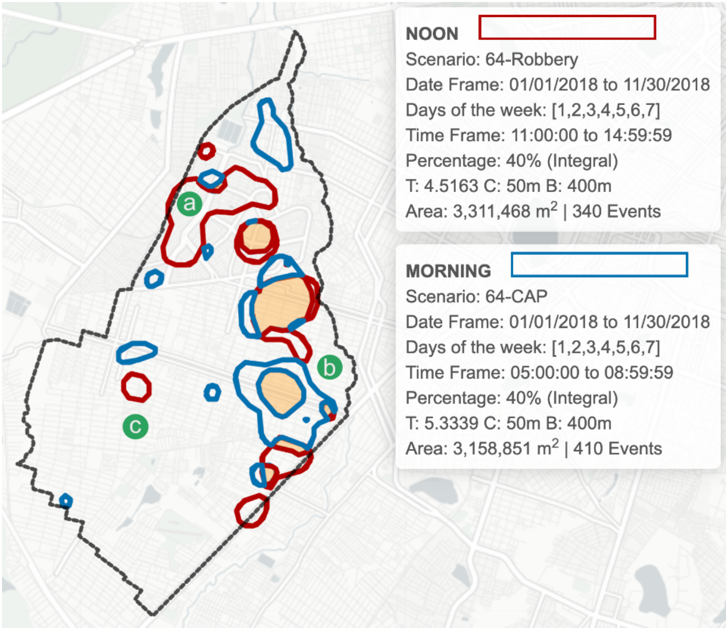
to make it easier to interpret. Notice the spatial difference between the hotspots for CAL and CAP. In this analysis, crimes against properties are more concentrated in affluent regions while crimes against life are more frequent in impoverished regions (see the zones in Figure 31), a situation similar to the one observed by Balbi and Guerry in their pioneer study (BALBI; GUERRY, 1829).

6.2.2 Crimes Against Property (CAP)

For analyzing CAP, officers created a histogram that shows the number of crimes at each hour and selected the two hottest four-hour periods to concentrate their efforts: the first period, in the morning, from 5:00 to 9:00 (people going to work and school) and the second, around noon, from 11:00 to 15:00 (people going out for lunch or returning from school). For the exploration of the differences between the two time frames, they created a multi-time analysis (DT3) with a percentage of 40% of the total contribution of weighted incidents. Figure 31 shows the DT3 analysis.

From the DT3 analysis, officers acquired some new information. For example, they observed CAP concentration in Zone A in the morning but not around noon. They thought that happened because people living in Zone A usually do not need to move too far away from their

Figure 31 – Multi-time analysis (DT3) for crimes against property in Beat F.



Source: the author.
The different zones, based on their social and economic development, are indicated with labels: Zone A, in the northwest, is the most developed area; Zone B, alongside the right border, is intermediate; and Zone C, in the middle and southwest, is the less developed area.

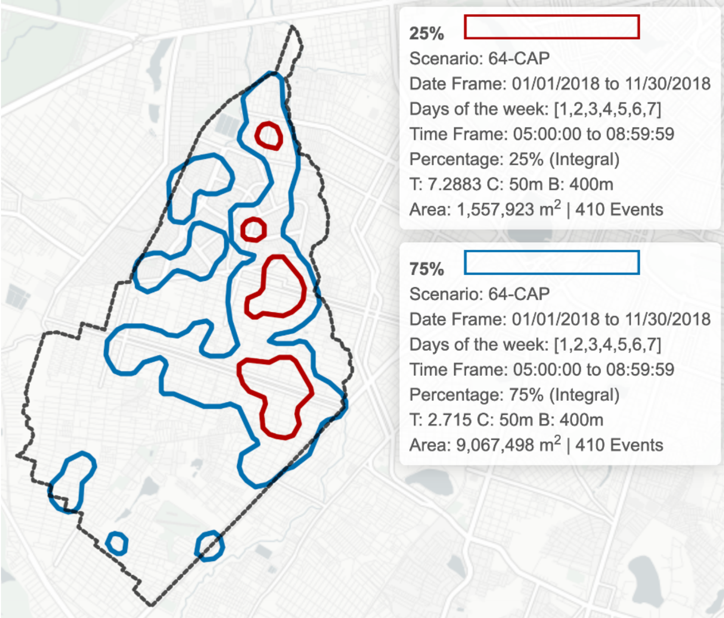
homes to go to school or to go to work early in the morning, so they are less vulnerable in the first period. On the other hand, people of zone B do need to move in the morning and become more vulnerable. Officers thought that CAP rises around noon in Zone A because of the numerous businesses and schools in the area, which attract criminals in these periods. They believed that Zone B also presented a high CAP rate around noon because of its peculiar geographical location, which connects Beat F to the rest of the city (right border). So, there is always a significant flow of people and vehicles through Zone B. Zone C is almost free of CAP hotspots because of the weak economy, a place less suitable for crimes against property.

To help planning the resource allocation, they created a multi-level analysis (DT5) for each period, using 25% and 75% of weighted incidents. Figure 32 shows the multi-level analysis for the period from 5:00 to 9:00, which gives indications on where are, proportionally, the CAP density.

6.3 Discussion

In this section, we present the feedback received from domain experts during the development process and execution of the use cases, followed by a discussion on the differences between the police of the two countries and the impact of using the system. We also discuss the

Figure 32 – Multi-level analysis (DT5) for crimes against property in Beat F, in the morning.



Source: the author.
The four red polygons include 25% of the density whereas the big blue polygon encloses 75%.

limitations of the tool.

6.3.1 *Domain Expert's Feedback*

Officers expressed excitement with the possibility of comparing, on the same screen, hotspots for different time frames or different kinds of incidents. In their opinion, it is easier and more accurate than comparing two hotspot maps side by side.

The LPD's Chief and the Crime Analyst told us that SHOC would be useful in the frequent task of evaluating the impact of initiatives (DT2 is particularly useful). They thought that quick and precise visual comparisons facilitated the assessment of actions and strategy changes.

Officers considered that SHOC would have a positive impact on routine activities. They believed that tracking the crime phenomena with SHOC would lead to improvements in patrol planning and, therefore, to better resource allocation.

The LPD's Chief and both Captains in charge of patrolling pointed out that easily making comparative analyses in SHOC is useful for decision-making regarding shifts. To keep the shifts as efficient as possible, they need, for example, to compare crime density between daytime and night-time and between working days and weekends. They agreed that SHOC facilitates those comparisons.

The LPD's Specialist in Crime Prevention also remarked that the annotation layer would improve communication and teamwork. For example, an officer A would annotate a given analysis and, later, an officer B could have insight into the analytical process used by A, even without any help from officer A.

The LPD's crime analyst requested us to simplify the interface regarding the number of input parameters, allowing, for example, the use of the city's or office's profiles to set some parameters automatically. He believed that officers would feel more confident in using the system when they understood everything in the interface.

6.3.2 Limitations

Although there are specific rules about parameters, SHOC is still not pure enough concerning this point. Officers have to set cell size, bandwidth, and the density percentage, which is not natural for some of them.

As we pointed out, SHOC computes polygon operations, such as intersections and subtractions, for two MSKDE layers, and the resulting pairwise comparisons satisfy the requirements of the officers. Nevertheless, this is more a matter of taste than of a system limitation, since SHOC is capable of creating and visualizing as many layers as desired. Moreover, the user can easily hide and show layers using the layer control interface. In fact, users often create many layers per analysis, showing some of them on demand. However, in our experience with the officers, visualizing more than three MSKDEs at the same time is not recommended because of clutter.

SHOC is based on MSKDE, which is a contour and, therefore, it does not include the spatial distribution of the weighted incidents contribution inside the polygons. To minimize that limitation, we included in the visual analytics system, a KDE generator. With that, users can, on-demand, show, and hide a KDE layer and look for noticeable concentrations inside the polygons that could have an impact on the analysis.

6.4 Impressions about differences and similarities between the Brazilian and the American police environments.

In general, we observed more similarities than differences between the two environments. Regarding the distribution of incidents, the spatial and temporal concentrations are similar, varying the kind of events. We noted that, in the USA, the police are closer to the community, sharing more data, and making some discussions together (SURAKITBANHARN *et al.*, 2018). In both countries, they are conscious of the benefits of hotspot policing. Regarding officers' profiles, we found officers with a high level of interest and willingness to use the system in every police department. However, they were always pressed for time, and prioritized being on the streets.

We believe that SHOC would be useful in many police departments around the world since hotspot policing is suitable for fighting most types of crimes. Our approach is straightforward, and with only a few clicks, an officer can create an analysis that is easy to be interpreted. They can also fully annotate and share it with colleagues, improving the knowledge of the police department.

6.5 CrimeWatcher Current Status

The Brazilian Ministry of Justice and Public Safety has been developing a big data platform and a set of applications related to public safety, to be deployed all over the country. The project is called "Big Data and Artificial Intelligence." The platform and systems would be available to all public safety institutions of Brazil, whether federal, state, or municipal (Ministério da Justiça e Segurança Pública, 2019). The Ministry has evaluated the academic version of CrimeWatcher, and the evaluators have found that, due to its simplicity and effectiveness, CrimeWatcher could be adequate to be part of this project.

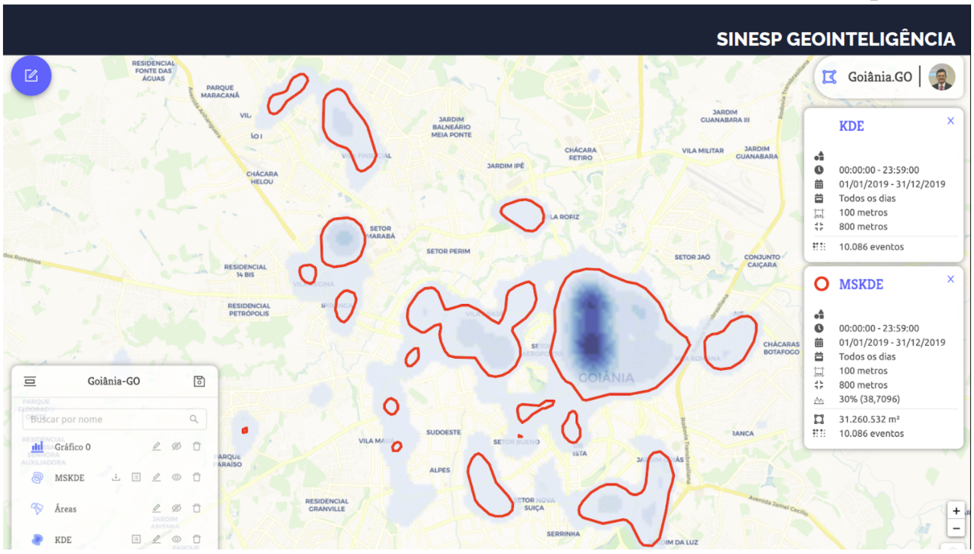
The Ministry refactored CrimeWatcher and gave it a new name: Geo Inteligência. Geo Iteligência is capable of dealing with millions of incidents, it is multitenant, and it has improved security and architecture. Since that, the Ministry started deploying it all over the country, and twelve States have received training sessions and technical assistance so far. The system is now in a permanent cycle of improvements, and part of the Ministry's portfolio of applications.

Figure 33 shows an example of the current interface of the system.

6.6 Summary

In this chapter, we demonstrated SHOC effectiveness in two real case studies that helped police departments on planning their preventive initiatives for high relevance public safety

Figure 33 – Current interface of CrimeWatcher, with its new brand name, Sinesp GeoInteligência



Source: the author.

issues. Moreover, we presented some feedback from domain experts taken when conducting the case studies. Finally, we discussed the limitations of the technique, and, in the intention of giving a better context, we pointed out some observations on differences and similarities between the Brazilian and the American police.

7 CONCLUSIONS AND FUTURE WORKS

This chapter summarizes the contributions of this thesis, their limitations, and provides an outlook for future work.

Firstly we present the contributions, taking into account the objectives stated in the introduction, Section 1.3. Secondly, we summarize the limitations our techniques and finally, we discuss our future works. In terms of the specific objectives, Table 7 shows the relationship of each one with the chapters where they were covered in this thesis.

Table 7 – Specific Objectives and chapters of this thesis

#	Specific Objectives	Chapters
1	Investigate environmental criminology to understand the relationship between place, time and crime	2,3
2	Investigate the creation of crime maps, identifying the best techniques for crime analysis and prediction	2,3
3	Investigate visual analytics for crime data to identify best practices and requirements	2,3,4
4	Investigate police departments' reasoning and planning processes to understand their modus operandi and requirements	4,5
5	Identify police departments' domain tasks that would benefit from a visual analytics system for crime data	5
6	For each identified domain task, identify and/or design a visual analytics technique to address it	5
7	Develop a prototype of a visual analytics system to address the domain tasks	5
8	Evaluate the prototype, modeling use cases together with officers	6

Source: the author.

7.1 Contributions

This thesis explores the adoption of visual analytics to support the work of police departments regarding planning their activities. This adoption involves improving a classic algorithm to generate hotspot maps, the extracting of domain tasks related to hotspot analysis, and the addressing of them.

7.1.1 Marching Squares Kernel Density Estimation (MSKDE)

We introduced Marching Squares Kernel Density Estimation (MSKDE), a technique that quickly transforms low resolution KDE hotspot maps into a more attractive and accurate hotspot map. We demonstrated that MSKDE outperforms KDE not only by offering a less anomalous map in the same execution time, but also by generating a map with a predefined anomaly faster. Furthermore, we present the Hotspot Anomaly Index (HAI), an index to evaluate the degree of anomaly between a hotsport map and a reference hotspot map.

7.1.2 CrimeWatcher - A Visual Analytics System for Crime Data

We provided a characterization and abstraction of the hotspot analysis problem into five domain tasks, presented SHOC (one-SHOt Comparison Tool) and CrimeWatcher. SHOC is a technique that facilitates performing the domain tasks by giving a straightforward way of making spatial comparisons, using set-like operations on superimposed geometries. CrimeWatcher is a solid yet easy visual analytics system that ables users to perform SHOC and create other classic crime maps. We demonstrate SHOC effectiveness in two real case studies that helped police departments on planning their preventive initiatives for high relevance public safety issues.

7.2 Limitations

This thesis proposes new methods and technologies that have some limitations. The main constraints are:

- As MSKDE is based on KDE, it experiences from KDE issues, as computational cost.
- MSKDEs are visual components that can be overlapped for facilitating comparisons. However, analysis usually starts suffering from cluttering when they have four or more layers.
- As MSKDE is a contour, and it brings limited spacial information about the area inside the borders.
- The current implementation of SHOC can create auxiliary geometries, from the set operations (intersections and subtractions), based on only two layers.
- CrimeWatcher demands that users set some parameters, like bandwidth and cell size, which might not be particularly natural for them.

7.3 Future Works

Based on the feedback of the users during the initial deployments of CrimeWatcher, there are different opportunities for future work.

Regarding the interface, we would like to simplify parameter selection, creating parameter profiles for the environments that, working together with the user profiles, would automatically select cell size and bandwidth. Also, when sharing analyses with non-analysts and other stakeholders, it would be desirable to display a simplified version of the interface with fewer controls.

We identified that as users' analyses start to become more sophisticated, they will need a more sophisticated annotation system. We plan to enable provenance, encryption, and support semantic queries on the annotations. We also plan to support multiple annotation layers in the analyses.

We plan to incorporate animation techniques and controls to display the evolution of MSKDEs over time as an animation that users can pause and playback at selected time frames to help them perform further investigation. In this subject, there is derivated research that investigates animated MSKDEs (Ramos *et al.*, 2019). We also plan to examine, in research about perception, which visualization technique (static MSKDEs or animated MSKDEs) would be better for insights regarding hotspot tracking over time.

We plan to incorporate the calculation of hotspots projected to the road network. We would like to include layers with roads and street segments highlighted as hotspots. This would improve planning police patrolling, which, most of the time, is performed on roads. There is a derivated, ongoing research on this subject (JUNIOR *et al.*, 2019) with great possibilities of high-grade results.

We have plans to make CrimeWatcher a comprehensive environment for dealing with, besides crime, most of the Spatio-temporal events, like fire, car crashes and traffic stops. For this, we have to incorporate other algorithms, particular to other domains, and include more attributes, like starting date and ending date, starting position, and ending position. Making CrimeWatcher richer on dealing with more domains and different data would facilitate it to be a broader anomaly detector for Spatio-temporal events.

Finally, we plan to incorporate other custom analyses involving machine learning techniques for the prediction of events.

BIBLIOGRAPHY

ALSALLAKH, B.; MICALLEF, L.; AIGNER, W.; HAUSER, H.; MIKSCH, S.; RODGERS, P. Visualizing Sets and Set-typed Data: State-of-the-Art and Future Challenges. In: BORGO, R.; MACIEJEWSKI, R.; VIOLA, I. (Ed.). **EuroVis - STARs**. [*S. l.*]: The Eurographics Association, 2014.

BALBI, A.; GUERRY, A. M. **Statistique Comparée - De L'Etat De L'Instruction et du Nombre des Crimes**. [*S. l.*]: Ministre de la Justice, 1829.

BEECHAM, R.; DYKES, J.; SLINGSBY, A.; TURKAY, C. Supporting crime analysis through visual design. In: **VIS 2015**. Chicago, USA: [*S. n.*], 2015.

BISHOP, C. **Pattern Recognition and Machine Learning**. [*S. l.*]: Springer, 2006. (Information Science and Statistics). ISBN 9780387310732.

BOWERS, K. J.; JOHNSON, S. D.; PEASE, K. Prospective hot-spotting the future of crime mapping? **British Journal of Criminology**, Oxford, England, v. 44, n. 5, p. 641–658, 2004.

BRAGA, A. A. The effects of hot spots policing on crime. **The Annals of the American Academy of Political and Social Science**, Thousand Oaks, CA, USA, v. 578, p. 104–125, November 2001.

BRAGA, A. A.; PAPACHRISTOS, A. V.; HUREAU, D. M. The effects of hot spots policing on crime: An updated systematic review and meta-analysis. **Justice quarterly**, Taylor & Francis, Abingdon-on-Thames, England, UK, v. 31, n. 4, p. 633–663, 2014.

BRANTINGHAM, P.; BRANTINGHAM, P. Criminality of place. **European journal on criminal policy and research**, Springer, Berlin/Heidelberg, Germany, v. 3, n. 3, p. 5–26, 1995.

BRANTINGHAM, P. J.; BRANTINGHAM, P. L. **Patterns in crime**. [*S. l.*]: Macmillan New York, 1984.

BRANTINGHAM, P. J.; BRANTINGHAM, P. L.; ANDRESEN, M. A. The geometry of crime and crime pattern theory. **Environmental criminology and crime analysis**, Taylor & Francis, Abingdon-on-Thames, England, UK, v. 2, 2017.

BREHMER, M.; MUNZNER, T. A multi-level typology of abstract visualization tasks. **IEEE Transactions on Visualization and Computer Graphics**, IEEE Educational Activities Department, Washington, DC, USA, v. 19, n. 12, p. 2376–2385, dez. 2013. ISSN 1077-2626.

BRUNSDON, C.; CORCORAN, J.; HIGGS, G. Visualising space and time in crime patterns: A comparison of methods. **Computers, environment and urban systems**, Elsevier, Amsterdam, Netherlands, v. 31, n. 1, p. 52–75, 2007.

CHAINEY, S. Examining the influence of cell size and bandwidth size on kernel density estimation crime hotspot maps for predicting spatial patterns of crime. **BSGLg**, Bingley, West Yorkshire, England, v. 60, n. 1, p. 7–19, 2013. ISSN 07707576.

CHAINEY, S.; RATCLIFFE, J. **GIS and crime mapping**. [*S. l.*]: Wiley, 2005.

CHAINEY, S.; REID, S.; STUART, N. When is a Hotspot a Hotspot? A procedure for creating statistically robust hotspot maps of crime. In: KIDNER, D.; HIGGS, G.; WHITE, S. (Ed.). **Innovations in GIS 9: Socio-economic applications of geographic information science**. Abingdon-on-Thames, England, UK: Taylor & Francis, 2002.

CHAINEY, S.; TOMPSON, L.; UHLIG, S. The utility of hotspot mapping for predicting spatial patterns of crime. **Security Journal**, London, England, v. 21, p. 4–28, 2008.

CLARKE, R.; ECK, J. **Become a Problem-Solving Crime Analyst**. Abingdon-on-Thames, England, UK: Taylor & Francis, 2014. ISBN 9781135898946.

CLEVELAND, R. B.; CLEVELAND, W. S.; MCRAE, J. E.; TERPENNING, I. STL: A seasonal-trend decomposition procedure based on loess. **Journal of Official Statistics**, Stockholm, Sweden, v. 6, n. 1, p. 3–73, 1990.

COHEN, L. E.; FELSON, M. Social change and crime rate trends: A routine activity approach. **American Sociological Review**, American Sociological Association, Washington, D.C., USA, v. 44, n. 4, p. 588–608, 1979.

CORNISH, D. B.; CLARKE, R. V. The rational choice perspective. In: **Environmental criminology and crime analysis**. Abingdon-on-Thames, England, UK: Taylor & Francis, 2017. p. 29–61.

ECK, J. E.; CHAINEY, S.; CAMERON, J. G.; LEITNER, M.; WILSON, R. E. **Mapping crime: Understanding hotspots**. [*S. l.*]: National Institute of Justice, 2005.

ECK, J. E.; WEISBURD, D. Crime places in crime theory. **Crime and Place**, Lynne Rienner Publishers, Boulder, Colorado, USA, v. 4, p. 1–33, 1995.

EGENHOFER, M. J.; FRANZOSA, R. D. Point-set topological spatial relations. **International Journal of Geographical Information System**, Taylor & Francis, Abingdon-on-Thames, England, UK, v. 5, n. 2, p. 161–174, 1991.

ESRI.COM. **Esri - GIS Mapping Software, Solutions, Services, Map Apps, and Data**. Redlands, CA, USA: [*S. n.*], 2015. Available at http://www.esri.com. Accessed on 10 jan. 2020.

ESTIVILL-CASTRO, V.; LEE, I. Amoeba: Hierarchical clustering based on spatial proximity using delaunay diagram. In: **Proceedings of the 9th International Symposium on Spatial Data Handling. Beijing, China**. [*S. l.*: *s. n.*], 2000. p. 1–16.

GODWIN, A.; STASKO, J. T. Nodes, Paths, and Edges: Using Mental Maps to Augment Crime Data Analysis in Urban Spaces. In: KOZLIKOVA, B.; SCHRECK, T.; WISCHGOLL, T. (Ed.). **EuroVis 2017 - Short Papers**. [*S. l.*]: The Eurographics Association, 2017. ISBN 978-3-03868-043-7.

GORR, W. L.; LEE, Y. Longitudinal study of crime hot spots: dynamics and impact on part 1 violent crime. In: **Proceedings of the 32nd international symposium on forecasting**. [*S. l.*: *s. n.*], 2012. p. 24–27.

HART, T.; ZANDBERGEN, P. Kernel density estimation and hotspot mapping: Examining the influence of interpolation method, grid cell size, and bandwidth on crime forecasting. **Policing: An International Journal of Police Strategies & Management**, Emerald Group Publishing, Bingley, West Yorkshire, England, v. 37, n. 2, p. 305 – 323, 2014. ISSN 1363951X.

HAYNES, F. E. **Criminology**. [*S. l.*]: McGRAW-HLLL BOOK COMPANY, INC., 1935.

HU, Y.; WANG, F.; GUIN, C.; ZHU, H. A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation. **Applied geography**, Elsevier, Amsterdam, Netherlands, v. 99, p. 89–97, 2018.

International Association of Crime Analysts. **Definition and types of crime analysis [White Paper 2014-02]**. Overland Park, KS, USA: [*S. n.*], 2014. Available at https://iaca.net/white-papers. Accessed on 08 feb. 2020.

IZIQUE, C. **Crescimento da violência no país surpreende pesquisadores**. São Paulo, SP, Brazil: [*S. n.*], 2013. Available at: http://exame.abril.com.br/brasil/noticias/violencia-democracia-e-direitos-humanos. Accessed on 20 jan. 2020.

JOHNSON, S. D.; BERNASCO, W.; BOWERS, K. J.; ELFFERS, H.; RATCLIFFE, J.; RENGERT, G.; TOWNSLEY, M. Space–time patterns of risk: A cross national assessment of residential burglary victimization. **Journal of Quantitative Criminology**, Springer, Berlin/Heidelberg, Germany, v. 23, p. 201–219, 2007.

JOHNSON, S. D.; BOWERS, K. J. The burglary as clue to the future the beginnings of prospective hot-spotting. **European Journal of Criminology**, Newbury Park, California, USA, v. 1, p. 237–255, 2004.

JUNIOR, F. C. N.; SILVA, T. L. C. d.; NETO, J. F. d. Q.; MACÊDO, J. A. F. d.; PORCINO, W. C. A novel approach to approximate crime hotspots to the road network. In: **Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility**. [*S. l.*: *s. n.*], 2019. p. 53–61.

KEIM, D.; KOHLHAMMER, J.; ELLIS, G.; MANSMANN, F. **Mastering the Information Age - Solving Problems with Visual Analytics**. [*S. l.*]: Eurographics Association, 2010. ISBN 9783905673777.

KINNEY, J. B.; BRANTINGHAM, P. L.; WUSCHKE, K.; KIRK, M. G.; BRANTINGHAM, P. J. Crime attractors, generators and detractors: Land use and urban crime opportunities. **Built environment**, Alexandrine Press, Marcham, England, v. 34, n. 1, p. 62–74, 2008.

LEVINE, N. The hottest part of a hotspot: comments on the utility of hotspot mapping for predicting spatial patterns of crime. **Security journal**, London, England, 2008.

LEVINE, N. **CrimeStat IV - A Spatial Statistics Program for the Analysis of Crime Incident Locations**. [*S. l.*]: National Institute of Justice, 2013.

LIMA, R. S.; BUENO, S.; RODRIGUES, B.; PEKNY, A. C.; FIGUEIREDO, L.; PRÖGLHÖF, P. N.; SOBRAL, I.; MONEO, V.; APARÍCIO, C. A. P.; KAHN, T.; RICARDO, C.; CERQUEIRA, D.; OLIVEIRA, F. L.; SILVA, F. S.; PIRES, L.; RAMOS, L. O.; CUNHA, L. G.; BAIRD, M. F.; PERES, M. F. T.; POLLACHI, N.; ALCADIPANI, R.; CUSTÓDIO, R.; MIKI, R.; MUGGAH, R.; PERES, U. **Anuário Brasileiro de Segurança Pública 2014**. São Paulo, SP, Brazil, 2014. Available at: http://www.forumseguranca.org.br/storage/8_anuario_2014_20150309.pdf. Accessed on 20 jan. 2020.

LOPES, A.; BRODLIE, K. Accuracy in contour drawing. In: CITESEER. **Eurographics UK**. [*S. l.*], 1998. v. 98, p. 301–311.

LORENSEN, W. E.; CLINE, H. E. Marching cubes: A high resolution 3d surface construction algorithm. **SIGGRAPH Comput. Graph.**, ACM, New York, NY, USA, v. 21, n. 4, p. 163–169, ago. 1987. ISSN 0097-8930.

LUKASCZYK, J.; MACIEJEWSKI, R.; GARTH, C.; HAGEN, H. Understanding hotspots: A topological visual analytics approach. In: **Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems**. New York, NY, USA: ACM, 2015. (SIGSPATIAL '15), p. 36:1–36:10. ISBN 978-1-4503-3967-4.

LYNCH, K. **The Image of the City**. [*S. l.*]: Harvard U.P.; Oxford U.P. (Harvard-MIT Joint Center for Urban Studies Series). ISBN 9780262620017.

MACIEJEWSKI, R.; RUDOLPH, S.; HAFEN, R.; ABUSALAH, A.; YAKOUT, M.; OUZZANI, M.; CLEVELAND, W. S.; GRANNIS, S. J.; EBERT, D. S. A visual analytics approach to understanding spatiotemporal hotspots. **IEEE Transactions on Visualization and Computer Graphics**, Washington, DC, USA, v. 16, n. 2, p. 205–220, March 2010. ISSN 1077-2626.

MALIK, A.; MACIEJEWSKI, R.; COLLINS, T. F.; EBERT, D. S. Visual analytics law enforcement toolkit. In: **2010 IEEE International Conference on Technologies for Homeland Security, HST 2010**. [*S. l.*: *s. n.*], 2010. p. 222–228. ISBN 9781424460472.

MALIK, A.; MACIEJEWSKI, R.; ELMQVIST, N.; JANG, Y.; EBERT, D. S.; HUANG, W. A correlative analysis process in a visual analytics environment. In: **IEEE Symposium on Visual Analytics Science and Technology 2012, October 14-19, Seatle, WA, USA**. [*S. l.*: *s. n.*], 2012. p. 33–42.

MALIK, A.; MACIEJEWSKI, R.; HODGESS, E.; EBERT, D. S. Describing Temporal Correlation Spatially in a Visual Analytics Environment. In: **2011 44th Hawaii International Conference on System Sciences**. [*S. l.*: *s. n.*], 2011. p. 1–8. ISSN 1530-1605.

MALIK, A.; MACIEJEWSKI, R.; MCCULLOUGH, S.; EBERT, D. S.; TOWERS, S. Proactive Spatiotemporal Resource Allocation and Predictive Visual Analytics for Community Policing and Law Enforcement. **IEEE Transactions on Visualization and Computer Graphics**, Washington, DC, USA, v. 20, n. 12, p. 1863–1872, December 2014.

MCDOWALL, D.; LOFTIN, C.; PATE, M. Seasonal cycles in crime, and their variability. **Journal of Quantitative Criminology**, Springer, Berlin/Heidelberg, Germany, v. 28, n. 3, p. 389–410, 2012.

Ministério da Justiça e Segurança Pública. **Ministério entrega aos estados primeiras ferramentas de Big Data e Inteligência Artificial para combater a criminalidade**. 2019. Available at: https://www.justica.gov.br/news/collective-nitf-content-1566331890.72. Accessed on 10 jan. 2020.

MOHLER, G. Marked point process hotspot maps for homicide and gun crime prediction in Chicago. **International Journal of Forecasting**, Elsevier, Amsterdam, Netherlands, v. 30, n. 3, p. 491–497, 2014.

Motorola Solutions. **CityProtect**. Chicago, IL, USA: [*S. n.*], 2020. Available at: https://www.cityprotect.com. Accessed on 10 jan. 2020.

NAGIN, D. S. Group-Based Trajectory Modeling: An Overview. **Handbook of quantitative criminology**, Springer, Berlin/Heidelberg, Germany, p. 53–67, 2010.

NAKAYA, T.; YANO, K. Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. **Transactions in GIS**, Wiley, Hoboken, New Jersey, USA, v. 14, n. 3, p. 223–239, 2010.

NIELSON, G. M.; HAMANN, B. The asymptotic decider: Resolving the ambiguity in marching cubes. In: **Proceedings of the 2Nd Conference on Visualization '91**. Los Alamitos, CA, USA: IEEE Computer Society Press, 1991. (VIS '91), p. 83–91. ISBN 0-8186-2245-8.

Quantum GIS Development Team. **QGIS - A Free and Open Source Geographic Information System**. 2017. Available at: http://www.gqis.org. Accessed on 20 jan. 2020.

QUEIROZ NETO, J. F.; SANTOS, E.; VIDAL, C. A. Mskde - using marching squares to quickly make high quality crime hotspot maps. In: CAPPABIANCO, F. A. M.; FARIA, F. A.; ALMEIDA, J.; KöRTING, T. S. (Ed.). **Electronic Proceedings of the 29th Conference on Graphics, Patterns and Images (SIBGRAPI'16)**. São José dos Campos, SP, Brazil: [*S. n.*], 2016. Available at: http://gibis.unifesp.br/sibgrapi16. Accessed on 20 jan. 2020.

Ramos, A. R. C.; Santos, E.; Cavalcante Neto, J. B. A partition approach to interpolate polygon sets for animation. In: **2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. [*S. l.*: *s. n.*], 2019. p. 139–146.

RATCLIFFE, J. Crime mapping: spatial and temporal challenges. In: PIQUERO, A. R.; WEISBURD, D. (Ed.). **Handbook of quantitative criminology**. Berlin/Heidelberg, Germany: Springer, 2010. cap. 2, p. 5–24.

ROSENBLATT, M. Remarks on Some Nonparametric Estimates of a Density Function. **The Annals of Mathematical Statistics**, Institute of Mathematical Statistics, Ann Arbor, Michigan, USA, v. 27, n. 3, p. 832–837, 1956.

SANTOS, R. B. **Crime Analysis With Crime Mapping**. [*S. l.*]: SAGE Publications, 2012. ISBN 9781452202716.

SCHOOL", H. K. **Compstat: A Crime Reduction Management Tool**. Cambridge, MA, USA: [*S. n.*], 2019. Available at: https://www.innovations.harvard.edu/compstat-crime-reduction-management-tool. Accessed on 20 jan. 2020.

SHAW, C.; MCKAY, H. **Juvenile Delinquency and Urban Areas**. [*S. l.*]: University of Chicago Press, 1942.

SHERMAN, L. W. Hot spots of crime and criminal careers of places. **Crime and place**, Willow Tree Press, Monsey, NY, v. 4, p. 35–52, 1995.

SHERMAN, L. W.; GARTIN, P. R.; BUERGER, M. E. Hot spots of predatory crime: Routine activities and the criminology of place. **Criminology**, Wiley, Hoboken, New Jersey, USA, v. 27, n. 1, p. 27–55, 1989.

SHERMAN, L. W.; GOTTFREDSON, D. C.; MACKENZIE, D. L.; ECK, J.; REUTER, P.; BUSHWAY, S. D. **Preventing crime: What works, what doesn't, what's promising**. [*S. l.*], 1998.

SHERMAN, L. W.; WEISBURD, D. General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial. **Justice quarterly**, Taylor & Francis, Abingdon-on-Thames, England, UK, v. 12, n. 4, p. 625–648, 1995.

SILVA, L. J. S.; GONZÁLEZ, S. F.; ALMEIDA, C. F. P.; BARBOSA, S. D. J.; LOPES, H. Crimevis: An interactive visualization system for analyzing crime data in the state of rio de janeiro. In: **19th International Conference on Enterprise Information Systems**. [*S. l.*]: SCITEPRESS - Science and Technology Publications, 2017. p. 193–200.

SILVERMAN, B. W. **Density estimation for statistics and data analysis**. [*S. l.*]: Chapman and Hall, 1986.

SURAKITBANHARN, C.; NETO, J. F. de Q.; WANG, G.; EBERT, D. S. Community outreach using incident records and visual analytics. In: **Community-Oriented Policing and Technological Innovations**. Berlin/Heidelberg, Germany: Springer, 2018. p. 19–27.

THOMAS, J.; COOK, K. A. **Illuminating the Path: The Research and Development Agenda for Visual Analytics**. [*S. l.*]: IEEE Computer Society Press, 2005. ISBN 9780769523231.

VACCINE. **Visual Analytics Law Enforcement Toolkit (VALET) User Manual**. [*S. l.*]: Visual Analytics for Command, Control, and Interoperability Environments, a Department of Homeland Security Center of Excellence, 2015.

Van Wijk, J. J.; Van Selow, E. R. Cluster and calendar based visualization of time series data. In: **Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis'99)**. [*S. l.*: *s. n.*], 1999. p. 4–9.

WEISBURD, D. The law of crime concentration and the criminology of place. **Criminology**, Wiley, Hoboken, New Jersey, USA, v. 53, n. 2, p. 133–157, 2015.

WEISBURD, D.; BUSHWAY, S.; LUM, C.; YANG, S.-M. Trajectories of crime at places: A longitudinal study of street segments in the city of seattle. **Criminology**, Wiley, Hoboken, New Jersey, USA, v. 42, n. 2, p. 283–322, 2004.

WEISBURD, D.; ECK, J.; BRAGA, A.; TELEP, C.; CAVE, B.; BOWERS, K.; BRUINSMA, G.; GILL, C.; GROFF, E.; HIBDON, J. *et al.* **Place Matters: Criminology for the Twenty-First Century**. [*S. l.*]: Cambridge University Press, 2016. ISBN 9781316483152.

WEISBURD, D.; ECK, J. E. What Can Police Do to Reduce Crime, Disorder, and Fear? **The ANNALS of the American Academy of Political and Social Science**, Sage, Thousand Oaks, California, USA, v. 593, n. 1, p. 42–65, set. 2004.

WEISBURD, D.; GROFF, E.; YANG, S. **The Criminology of Place: Street Segments and Our Understanding of the Crime Problem**. [*S. l.*]: OUP USA, 2012. ISBN 9780199928637.

WEISBURD, D.; MASTROSFSKI, S. D.; MCNALLY, A. M.; GREENSPAN, R.; WILLIS, J. J. Reforming to preserve: Compstat and strategic problem solving in american policing*. **Criminology & Public Policy**, Wiley, Hoboken, New Jersey, USA, v. 2, n. 3, p. 421–456, 2003.

WEISBURD, D.; MCEWEN, T. **Crime Mapping and Crime Prevention**. [*S. l.*]: Criminal Justice Press, 1997.

WILLIAMSON, D.; MCLAFFERTY, S.; GOLDSMITH, V.; MOLLENKOP, J.; MCGUIRE, P. A better method to smooth crime incident data. **ESRI ArcUser Magazine**, ESRI, Redlands, CA, USA, 1999. Available at: http://www.esri.com/news/arcuser/0199/janmar99.html. Accessed on: 20 jan. 2020.

WOLFGANG, M. E.; FIGLIO, R. M.; SELLIN, T. **Delinquency in a Birth Cohort**. [*S. l.*]: University of Chicago Press, 1987.

WONG, C. R. W.; KODAGODA, N. **Analyst User Interface: Thinking Landscape as Design Concept**. London, UK, 2017. (VALCRI White Paper Series, VALCRI-WP-2017-001). Available at: http://valcri.org/our-content/uploads/2017/01/VALCRI-WP-2017-002-AUI-Thinking-Landscape.pdf. Accessed on 20 jan. 2020.

WONG, W. **VALCRI: Visual Analytics for Sense-Making in Criminal Intelligence Analysis**. London, UK: [*S. n.*], 2018. Available at: http://valcri.org. Accessed on 20 jan. 2020.

WORTLEY, R.; TOWNSLEY, M. **Environmental Criminology and Crime Analysis**. Abingdon-on-Thames, England, UK: Taylor & Francis, 2016. (Crime Science Series). ISBN 9781317487098.