Masters Program in **GeoSpatial Technologies**

Investigating Crime Patterns in Egypt using Crowdsourced Data between 2011-2013

Thesis submitted in the partial fulfilment for the requirement for
The Degree of Master of Science in GeoSpatial Technologies

By:
Abbas Adel Ibrahim

Supervised by:
Prof. Dr. Edzer Pebesma
Prof. Dr. Jorge Mateu Mahiques
Prof. Dr. Marco Painho

DECLARATION

I hereby, certify that I have written this thesis independently with the guidance of my supervisors and with no other tools than the specified. The data was extracted primarily from Zabatak.com which founded which I have the write to use and publish the data. Sources used are referenced in the bibliography.

February 26, 2016
Muenster, Germany

ACKNOWLEDGMENTS

I begin with thanking God, The most Merciful, The Most Generous and The Most Gracious, for guiding me through this work.

Then, I would like to express my sincere gratitude to all my supervisors for their cooperation and aspiring guidance that allowed me to finalize this thesis and to make it as perfect as we can. I am grateful for the efforts you all made, your support, immense knowledge and your insightful comments that helped me to improve the work and finish the thesis in the best way possible. I am honored to have you as my supervisors.

Special thanks to Ph.D. Edith Gabriel for helping me applying her work on my data, Dr. Carl Peter Leslie for his valuable comment suggestions on my work during the thesis follow-up sessions, Dr. Ahmed Arafat for helping me understand essential concepts in this study and his continuous support, and Nourhan Khalifa and Haytham Atef for kind assistance and helpful advice.

Last but not the least, I would like to thank my wife Naglaa, my parents and my daughter for their patience and support.

Investigating Crime Patterns in Egypt using Crowdsourced Data Between 2011-2013

by Abbas Adel Ibrahim

Abstract

Crime is a social phenomenon that negatively impinges upon the society on various levels. Such phenomena are ought to be measured and analyzed to achieve control over its presence and consequences. One of the ways for measurement and analysis involves the use of crime maps as vital tools for visualising crime related data. Getting access to crime data is undoubtedly a challenged endeavour faced by hurdles of data collection, storage and making it available for public access. In addition, coming up with useful relationships for extracting information and patterns for crime data analysis is a significant challenge as well. This research investigates the link between the spatial and temporal variables in crime related data collected from crowdsourcing. The research will capitalize on crime data gathered throughout the operation of an online project called Zabatak founded by the author since January 2011 in Egypt. The dataset consists of more than 2000 crime incidents from various geographical areas across Egypt. The research considers an exploratory analysis in trying to interpret crime patterns and trends. The results of this study have identified various interesting trends and patterns in the dataset. One of the major findings of this research points out a strong relationship between the spatial and temporal variables in Car-Theft incidents. In addition, It was possible in the study to relate crime types to the type of the geographical area. The research considers Spatio-Temporal analysis using Inhomogeneous Spatio-Temporal K-function and pair-correlation functions which have identified a Spatio-Temporal cluster and interaction in crime data which can open new ways for crime maps data analysis.

Abbreviations

CAPMAS	Central Agency for Public Mobilization and Statistics
EDA	Exploratory Data Analysis
GIS	Geographical Information System
UNODC	United Nations Office on Drugs and Crime
UTM	Universal Transverse Mercator
VGI	Volunteered Geographic Information

Table of Contents

1. Introduction

Crime is a social phenomenon that negatively impinges upon a community on various levels. Such phenomenon is ought to be measured and analyzed in order to achieve control over its presence and consequences. According to Ratcliffe (2010), crime incidents are “neither uniformly nor randomly organized in space and time”. Ratcliffe’s analysis of crime mapping had opened the possibilities for in-depth crime data analysis, modeling and prediction of spatial and temporal variables. Such analysis could eventually lead to reduction and prevention in crime rates which is essential for urban security in modern societies.

Geographical Information System (GIS) has received significant development in the recent years. With the availability of GIS tools, mappers could discover patterns and trends to improve crime control policies and techniques. Such improvements could be boosted by the systematic prediction of geographical locations where crime is more likely to take place.

In this research, the relationship between the spatial and temporal variables highlighted by Ratcliffe (2010) is explored further in relevance to the Egyptian crime data collected through Zabatak.com. Zabatak is an online crowdsourcing project founded by the author after the 25th of January revolution to collect crime information after the failure of the law enforcements to help citizens to protect their assets (Ismail 2012).

The study is divided into six chapters. In the introduction chapter; a background to the study, related work and literature review, and the quality and limitation of the source of data used in the study. Chapter 2 presents and

describes the structure and classification of the dataset, the quality control procedures followed during the collection of the dataset and the study region. Chapter 3 describes two statistical analysis strategies: The first is exploratory analysis to understand the underlying architecture and patterns, and the second is a purely empirical approach that aims only to capture the Spatio-Temporal correlation structure. Chapters 4 and 5 perform and present the results of the analysis. The thesis ends with a discussion and conclusion in Chapters 6.

1.1. Crime Analysis

Many studies in the field of criminology explore and highlight the spatial relationship between crime incidents (Chainey and Ratcliffe 2013). Early attempts to employ crime maps in controlling the phenomena were ineffective due to the limitations of mapping and data storage technologies (Maltz, Gordon, and Warren 1991). However, in recent years, along the advances in technology and the focus on crime analysis in some police departments such as in the U.S., crime mapping techniques have been advancing notably (Taylor, Kowalyk, and Boba 2007). Ron Clarke refers to the future of crime mapping arguing that: “Quite soon, crime mapping will become as much an essential tool for criminological research as statistical analysis is at present” (Clarke 2004).

For centuries, criminal agencies have been reactive to the crime as proactivity requires collecting and analysing data to discover and predict possible crime hotspots as Ratcliffe (2010) stated: ”prevention requires proactivity requires predictability requires patterns”. Studies have proven that crime opportunities are not randomly distributed in space and time, Seasonal, monthly and even hourly trends have been documents, For instance, domestic violence tends and violent crime tend to increase during summer and commercial burglary in the winter (Clarke 2004). Vehicle theft mostly takes places at night in residential areas while nonresidential areas during the day (Ratcliffe 2000). With

these findings, better proactive measurements and policies could be held to reduce crime rate; for example, improve street lighting, better police patrol timing or use surveillance cameras.

1.2. Volunteered Geographic Information

The term Volunteered Geographic Information or for short VGI was introduced by Goodchild (2007) and it is widely disseminated afterwards. It refers to the voluntary efforts to contribute spatial data. Many other terms have emerged which have the same meaning such as web mapping, crowd-mapping, geo web, Public Participatory GIS and wiki-mapping (Elwood 2008).

The concept of involving the community to provide local data emerged from the limitation and the high cost of obtaining geographical information for a particular area. This involvement has enabled marginalized communities to influence the decision-making process (Deparday 2010).

Afterwards, the widespread of the online maps and mapping platforms such as Google Maps and Google Earth, and similar projects along with the accessibility and the availability of low cost of GPS-enabled devices, VGI becomes more available than before. It became easier for the public to create, organize and share spatial information and almost with no cost (Goodchild 2007).

Since 2014, Web 2.0 was developed to empower the internet user to contribute back to the web (O'Reilly 2009). Wikipedia[1] and openStreetMaps[2] are some of the famous examples that employed the power of crowdsourcing on the

[1] http://wikipedia.org
[2] http://www.openstreetmap.org

web. VGI found its way to capitalize on this new trend, and many terms have emerged which emphasis on the new abilities of the user to generate content.

"Neo-geography" was noted by Turner (2006) to describe the new characteristics of management and usage of the geographical information. Google Maps' term "mash-up map" became popular which describes the process of collecting and overlaying geographical information from different sources via the web, even the once created by non-professionals (Goodchild 2007). The 'crowd-mapping' term which was derived from 'crowdsourcing' by collecting information provided by the 'crowd'. Crowdsourcing and crowd-mapping are usually employed where large user groups perform tasks that are difficult or expensive to implement or automate (Ismail 2012).

1.3. User Generated Content

User generated content in the field of Geographic Information had evolved by many factors. The availability and the low cost of GPS enabled devices in cameras, mobiles, and portable computers, in addition to the Web 2.0 mapping platforms such as Google Maps and OpenStreetMaps enabled the user to make and distribute location aware content cheaply and effortlessly.

Consequently, the GPS enabled devices popularized 'Geo-tagging' which attach the geographical location to the generated content of pictures, videos or event text. The geo-tagging can provide structured information about the geographical location such as geodetic coordinates of Latitude and Longitude or in textual form as place names (Elwood 2008).

Furthermore, some other factors which are related to personal motivation such as self-promotion and personal satisfaction (Goodchild 2007) or dissatisfaction. For example, sharing geotagged photos in Flickr or others photo

sharing services could lead to self-promotion among friends or the Flickr community. Self-satisfaction could be achieved when the user sees the value of the contribution is solving a certain problem or helping others.

According to Coleman (2012), public participation in GIS is classified into three broad categories:

- The first is associated with economical factors, which means finding a cheap alternative to commercial spatial information providers by finding and contributing to another free one such as openStreetMaps.
- The second category is coupled with the common interest in groups to produce and exchange certain geographical information of their interest.
- The third category is related to a larger context where governmental or similar initiatives take initiatives to address specific issues, such as crises or crime issues. However, these categories are general and can have overlapped contents.

1.4. VGI Limitations and Challenges

While VGI sounds promising to provide a new type of information which were not available before with the multimedia content and accurate location, some challenges should be taken into consideration when implementing VGI project.

1.4.1. Digital Gap

While VGI was developed to overcome the lack of data in marginalized area and communities, it doesn't overcome the marginalization caused by the technology adoption where some groups of citizens do not have access to the internet. This issue could be tackled by using another available channel for information such as Short Messaging Service SMS or call-centers.

Seeger (2008) proposed another approach which is called facilitated VGI (f-VGI). This method transfers the knowledge from the individuals who don't have access to technology through the ones who are familiar with it.

1.4.2. Data Quality

Data quality is a major issue for VGI. There is some way to improve the quality, four of them are discussed in the following section: using group reviewers, user evaluation, automated data filtration, and correction by the crowd.

The first approach is to assign a group of trusted user to review the data. They should have more knowledge about the domain of the data to be able to review them. They could be from the local community who are in charge of the quality of the data in their area (Deparday 2010).

The second approach is data evaluation by allowing the users to evaluate the quality of the shared data. These assessments could be used to evaluate the data contributor who could affect the credibility of the contributor. This approach is also called peer-to-peer or crowdsourced credibility assessment (Deparday 2010).

The third approach is to combine the first two approaches and use sort of natural language processing where machines process and tries to understand the user provided content to automatic filtering data and flag records that need manual inspection. The distance between the user location and the reported data location or between other reported data location by the same user can also be used in the automatic evaluation (Deparday 2010).

The last approach is to leave the evaluation to the crowd by allowing them to flag and report missing or false information and enabling them to edit and confirm quality data. This technique has been used in software development and known by Linus law which is mainly used in open source development which allows other developers to flag and correct errors (Goodchild and Linna 2012) and it was introduced in Ushahidi, the leading crowdsourcing platform, to evaluate the user content (Goolsby and Rebecca 2010).

1.4.3. Data Diversity and Heterogeneous

The diversity of the data provided by VGI could lead to data heterogeneous due to the variety of data sources which happens when having different representations for the same entity. One method to tackle this problem is using standardization to provide automatic integration of the data which may contradict with the flexibility and richness of the contributed data (Elwood 2008). For example, providing a list of categories for the user to classify his or her contribution (Deparday 2010) or process and interpret the user's content and identify relevant keywords.

1.4.4. Security and Privacy Concerns

Privacy is another major concern of VGI data especially for the governmental project for the sensitivity and the mistrust of the generated information, mostly because of lack of awareness of the added-value of the citizen contribution (Cowan 2013). On the other hand, VGI implementer should ensure the protection of intellectual ownership of contributed data and the privacy of the contributor while allowing determination of trustworthiness of information (Metaxas, Panagiotis, and Eni 2013).

1.5. Crowdsourcing Crime Data

Crowdsourcing crime information is not new, and it is more interesting to study in developing countries where law enforcement fails to perform, and people resort to other ways to fill the gap. This type of data is actionable which means people can instantly use such information and react accordingly.

For Example, in Mexico, there is a rise of violence because of the drug industry and corruption. Despite the effort by the government and the armed forces to bring down the violence rate, criminals have “increased their terrorizing activities against random bystanders, tourists and even whole villages” (Metaxas, Panagiotis, and Eni 2013).

Citizens resorted to social media to share information about the danger they encounter every day. People used Twitter, in particular, to report, comment and confirm information about the violence and many Mexicans consulted Twitter every day on the hashtag #MTYfollow before planning their day (Metaxas, Panagiotis, and Eni 2013).

CROWDSAFE which is another novel approach using real-time crime data from the crowd and portable smart devices to provide the public with “Safety Router” in the metropolitan Washington DC area. Such real-time applications show the effectiveness and usefulness that crowdsourcing crime data can add to the citizen daily live (Shah et al. 2011).

In the Egyptian context, following the Egyptian revolution events and the fall of the ruling regime in Egypt on February 11th, 2011, many places in the country turned into police-less state. “People were urged to invent methods to regulate their lives” as, unfortunately, others saw the vulnerability of the state as

an opportunity to violate the law which became the norm for few months (Ismail 2012).

As a reaction to the law enforcement failure, we, a group of young activists who volunteered their time and expertise to solve the probelm stepped in using geo-enabled Web 2.0 tools to fill the enforcement vacuum (Ismail 2012). We created a web based model for citizen participation through monitoring, reporting and collaboration to track crime and outlaw activities and making the public aware of it.

The project relied on Ushahidi, an online platform provide crowdsourcing capabilities, to facilitate crowdsourced reporting from various channels including SMS, Twitter, Emails and Web Forms as well as visualization and interactive maps. The data collected through the project will be analysed in this study to identify crime patterns in Egypt.

For a successful interactive community mapping project, Shkabatur (2014) lists six factors to influence the success which some were considered during the project. These six factors are needed to “create a valuable participatory process and produce tangible outcomes”.

The first factor is information infrastructure, which was provided by Ushahidi platform, to collect the data from the crowd. To make reporting more efficient, we enabled crime reporting from various channels such as SMS, Twitter, Email and regular web forms. In addition to apps for popular smartphones brands at that time including Nokia, Blackberry, iPhone, Android and Windows devices.

The second factor is “Identified need for information” which resulted from the absence of law enforcements and the wide spread of crimes and rumors. As a reaction, people resorted to local committees for policing and gatekeeping and they used to exchange critical information with each other using personal connections. Providing interactive maps was more efficient to exchange and visualize information for the local communities and later for the law enforcement.

Civil society has to have the capacity to accept and use such tools which were available in the Egyptian context. People wanted to expose violations and needed to invent methods to regulate their lives (Ismail 2012). However, whoever have shared their knowledge about crime would have expectations, which makes it essential to seek for governmental cooperation, who are in turn eligible to meet those expectations, as in “If You See Something, Say Something” campaign created by the US Department of Homeland Security in the US in 2010 (Reeves 2013).

Therefore, “incentives” are needed to ensure continuous contribution but as Óscar Salazar (2009), the Mexican politician, said that “Mapping data is important for accountability. The fact that people see their report is the biggest motivation to engage them” (Ismail 2012).

The final factor in interactive community mapping project according to (Shkabatur and Jennifer 2014) is data quality which has been discussed earlier in depth.

2. Data Description

2.1. Introduction

Getting access to crime data is undoubtedly a challenged endeavour faced by hurdles of data collection, storage and making it available for public access, especially in countries where open-data laws are not available. Hence, crowdsourcing crime data could be an alternative to gain access to such data taking into consideration the limitation of that approach.

Information about crime is not publicly accessible in Egypt, and the need to such data became more critical in the absence of police forces after the Egyptian revolution, which lead to a wide spread of crime all over the country. This urged the people to find ways to share information about crime and take proactive measures against it. Web 2.0 projects such as Zabatak.com and social media, filled that need in order to:

- Share information instead of going to police stations,
- express frustration about crimes and the system,
- warn friends and others,
- or help each other.

Zabatak project was created to fill that gap, and provide the average citizen with a tool to report crime and corruption incidents anonymously using the internet, social media, and mobile technologies. The aim was to collect scattered information about criminal activities into structured centralized location for further analysis.

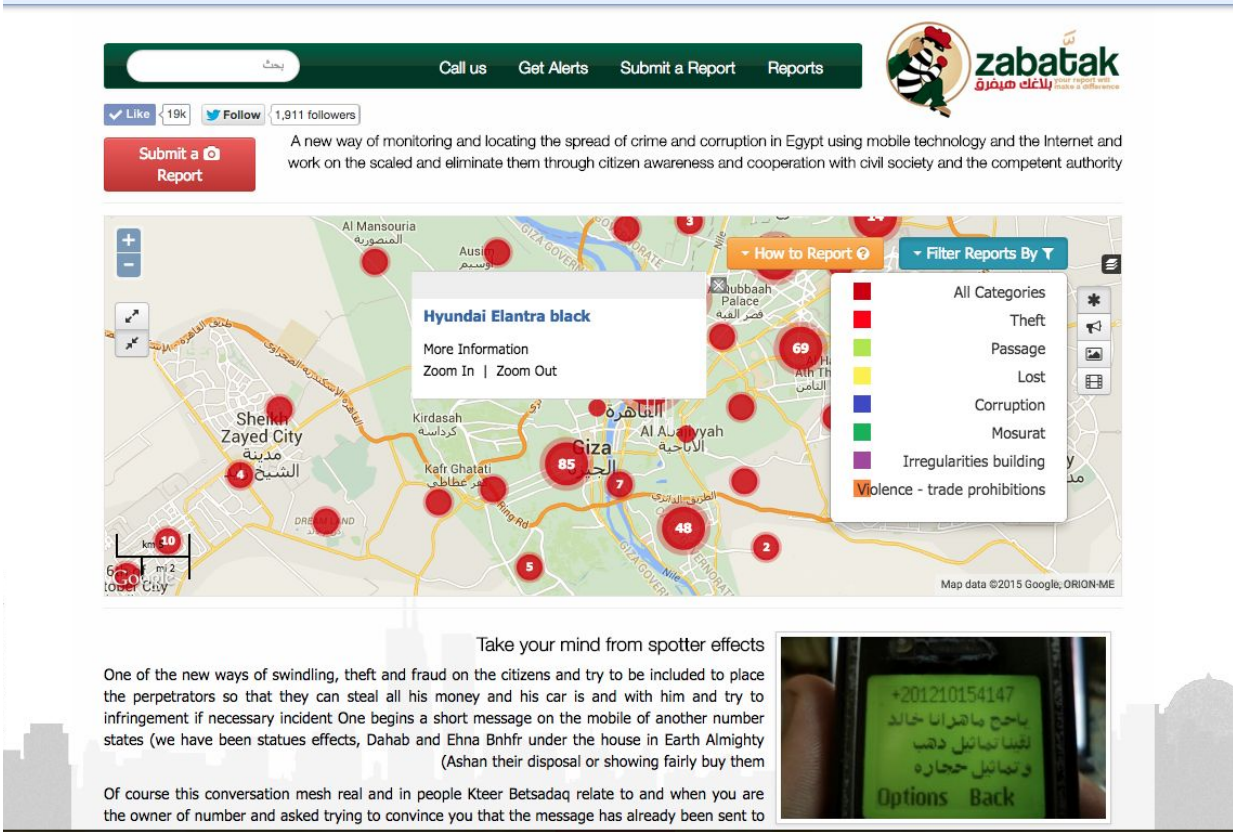


Figure 2.1: Screenshot of Zabatak.com taken at 10/01/2016

Zabatak uses the leading crowdsourcing platform Ushahidi[3] which is an online platform designed to collect information from citizens for crisis response, providing a simple and efficient way. Ushahidi was introduced for the first time during Kenyan elections in 2007 to collection information about post-election violence. Since then, it has been used to cover many events and provided insights into crises that are inaccessible or ignored by mainstream media.

Deparday (2010) noted that "even though the data are supposed to be objective, in an emotional context like a crisis event, some data may be biased or exaggerated". Users can also interact with reported data by commenting, voting up or down to assess the credibility of a report. The collected data is aggregated

[3] http://ushahidi.org

and visualized on an interactive map. The users can filter the data by category, dates and the attached media types (Deparday 2010).

Ushahidi platform utilizes Web 2.0 technologies to collect and visualize real-time public reports on a map via SMS, email, and the web. In addition, it supports Twitter and smartphone applications to broaden the possibilities of inputs. It can incorporate photos and video as well to provide richer information (Goolsby and Rebecca 2010).

2.2. Data Structure

For an incident to be reported successfully, a set of required inputs needs to be filled accordingly as in Figure 2.2. The “Title” fields are used for a summary about the reported incident while the description is used to provide more detailed information. The “Location name” is used to provide the address information for the incident. However, the user is required to provide the geographical location using an interactive map with the possibility to search by an address which is then geocoded using Google Maps APIs. The user is required to provide the date and the exact time of the incident and to select one category from a list of predefined categories.

To increase the credibility of the report, the contributor can provide additional information such as photos, videos or links. Even though users can report anonymously, the user is encouraged to leave his or her name and contact information which is not displayed to the public and stored in a secure location inside the system.

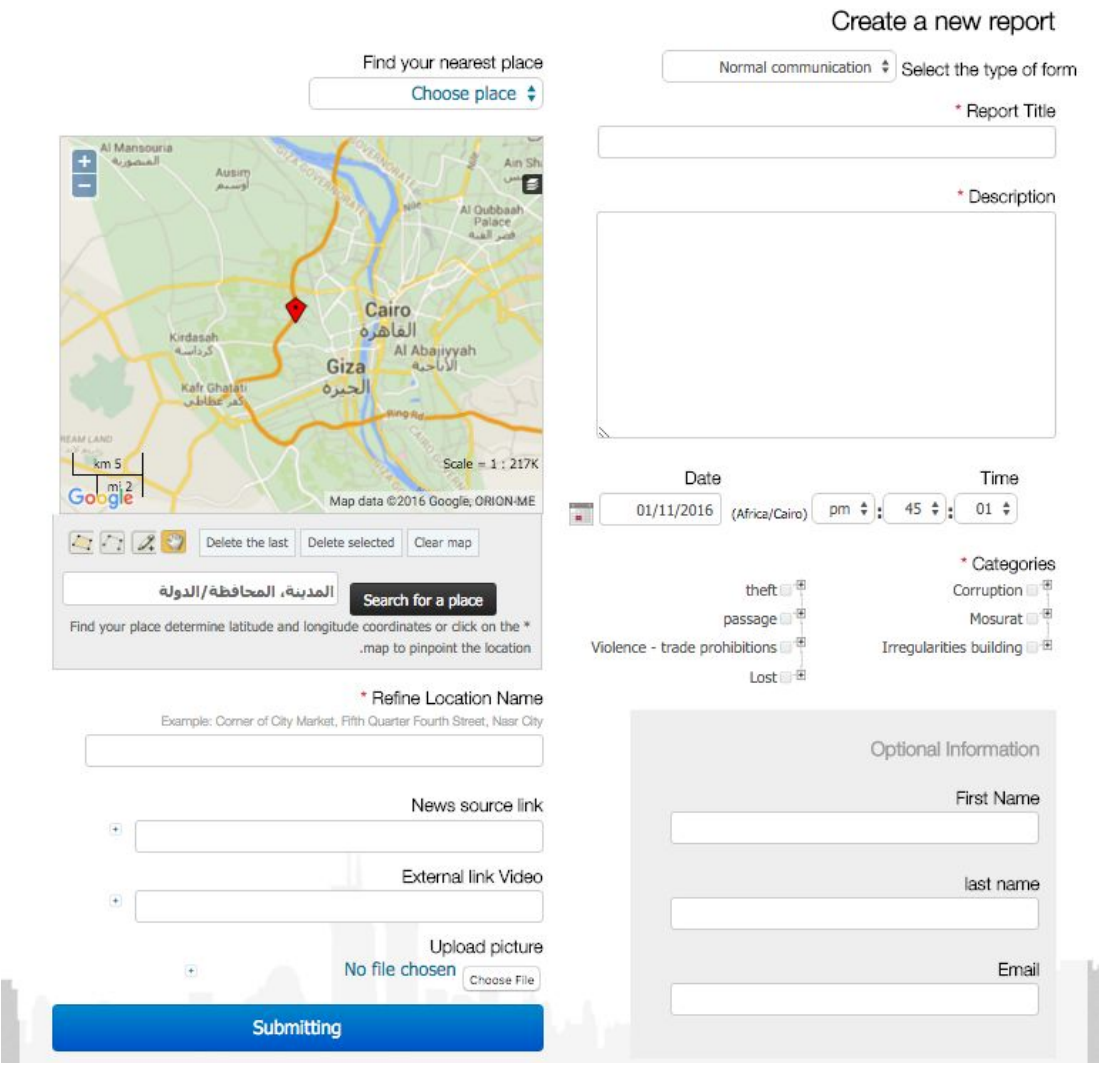


Figure 2.2: Recent screenshot for the online reporting form.

The project also provided two specialized forms for reporting Car-Theft and Land Violations incidents. For the former case, the users are required to provide the “Car Model” and the “Plate Number”. For the latter case, the user is asked to give more information about the “Type” and violated “Area” of the land along with the number of floors if available. The second form was created for a joint project with Fayoum University in Egypt lead by Ismail (2012) to be used during field work conducted by graduate research students of the mentioned university.

Only the mandatory fields will be used during this study, and two datasets will be prepared. The first one includes the following fields and to be used during the spatial, temporal and spatial-temporal analysis.

Location	GPS coordinates of latitude and longitude
Date of incident	The date and time of the incident
Category ID	The record ID of the category
Subcategory ID	The record ID of the subcategory

The second dataset will be used during the exploratory analysis and will use the same fields as the first dataset plus textual information.

Title	A brief description of the incident
Description	Detailed information about the incident
Address	Description of the location

2.3. Data Categorization

The data are categorized into eight main categories and 26 subcategories. However, only the following four categories will be used during the study as the others don’t contain enough information.

Category	**Subcategory**
Theft	Car theft
	Fraud and swindling
	Theft of private and public property

Violence - Illegal Trade	Arms or drug trade
	Terrorizing citizens, torture, murder or firing shots
Building Violations	Building on agricultural land
	New building without permit
	Building modification without permit
	Using the public land in any private use
Corruption	Bribery, job profiteering, and favoritism
	Careless Behavior
	Corruption in Localities (Throwing trash - Electricity - Water - Paving roads)
	Commercial Fraud (Price Cheating - Corrupted Products)

2.4. Input channels

Since the launch of the project in 2011, we tried to facilitate the reporting to the public by making it less time-consuming and more convenient. The website is the main channel for reporting. Smartphone apps were developed for major mobile apps platforms at that time including iOS, Android, Nokia, BlackBerry and Windows phones. An SMS short-number was also provided late 2011 to enable reporting by SMS. The project utilizes the functionality provided by Ushahidi platform to accept reports through emails and from Twitter using the hashtag #zabatak.

Despite the variety or reporting channels, almost all reports came through the online portal. Although the mobile apps were downloaded over 500 times but hardly used as smartphone penetrations were very low of 8.4% from the total handsets according to (Egypt Smartphone Survey 2012).

2.5. Data Quality Control

Crowdsourcing is known for providing low quality and biased data. During the course of the project, the following criteria were used to classify the quality of the data into three categories; trusted, not-trusted and not-accepted. However, there is no guarantee for the absolute validness of the collected data.

The reported incident is considered “Trusted” when:

- The contributor is known,
- the contributor or the incident can be verified by a volunteer,
- the incident has enough information for verification,
- or receiving similar information from multiple contributors.

The reported incident is considered Not-Trusted when:

- No enough information about the incident,
- no media or other documents to support the contributor claim,
- or no similar incidents available.

The reported incident is considered “Not-Accepted” when:

- False to too general information,
- or mentioning names of specific people, brands or entities.

In this study, we will focus only on the first two categories

2.6. Area of Study

Egypt is located in the northeastern corner of the African continent. It is surrounded from west to north by Libya, Sudan, the Red Sea, and the Mediterranean Sea. It has approximated area of one million square kilometers.

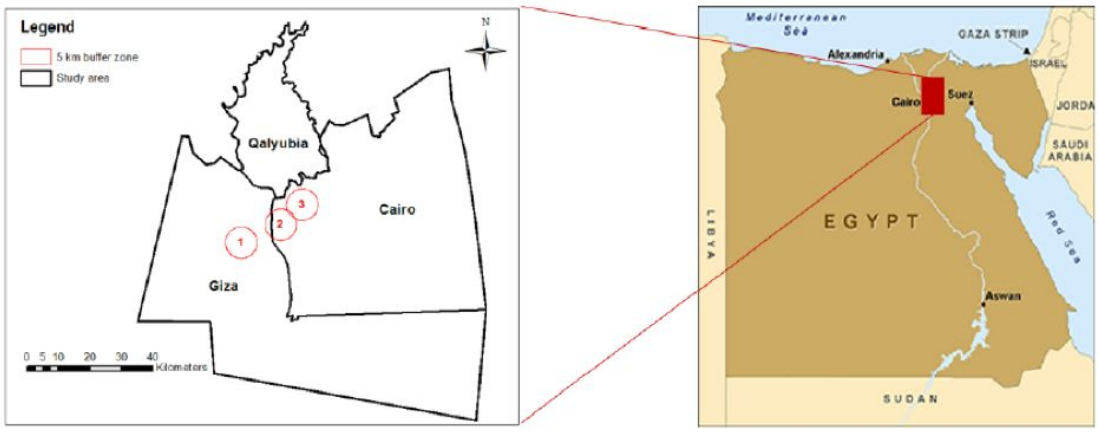


Figure 2.2: Greater Cairo area (Left). Egypt (right) (Megahed et al. 2015)

Egypt is the most populous country in North Africa and the Arab World with a total population of 86 million with a density of 1100 capita/km^2 in populated areas, and 8 million of outside population (CAPMAS 2014). The focus area in the study is the metropolitan area of Cairo which consists of the political capital and Qalyubiyah City, and parts of Giza and Sharqiyah cities that belong to Greater Cairo. The study area is located at 30° 02' N and 31° 21' E with an area of 8,942 km^2.

Greater Cairo is considered one of the fast growing megacities worldwide, with the highest population and population density among other Egyptian governorates (SIS, 2015). Cairo City was the most populous among other Egyptian cities (SIS, 2015), with almost 9 million inhabitants, representing 10.7% of total

population recorded in the same year (Megahed et al. 2015), followed by Giza, Sharkia, and Dakahlia with 7.6, 6.5, 5.9 million respectively as in Figure 2.3.

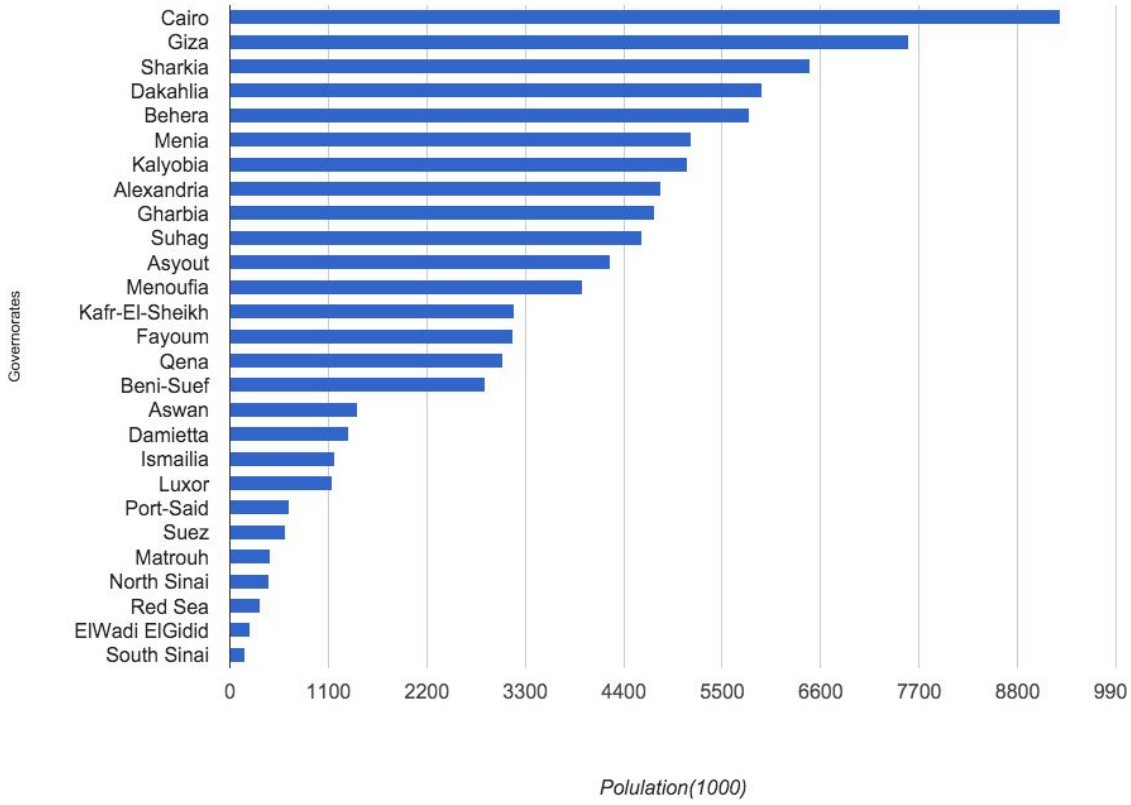


Figure 2.3: Population per 1000 in Egypt per Governorate (SIS 2015)

3. Methodology

The study is composed of generally three main phases. We will start by data preparation and validation followed by Exploratory Data Analysis and finally Spatio-Temporal Analysis. The three phases are described in details in the following sections.

3.1. Data Preparation and Validation

The collected data is stored in an online relational database. During this phase, data will be extracted from the project database and each field will be examined to determine its usefulness in the analysis. Two datasets will be prepared for the next two analysis phases.

Due to the low quality of crowdsourced or VGI data as noted earlier, the provided dataset will need to be carefully inspected and cleaned. The data will be check for duplicate, incomplete and invalid information.

3.2. Exploratory Data Analysis

Once the data is prepared, the analysis will be performed using statistical method Exploratory Data Analysis (EDA) which employs a variety of techniques that aims to understand the underlying structure, patterns and trend in the data.

EDA is a tradition way in data analysis that stems from John Tukey's work in the 1960s (Tukey 1969). EDA can be characterized as an understanding of the data that answer back this main question “What is going on here?"; and an emphasis on graphic representations of the data (Behrens 1997). Consequently, a pattern discovery is also revealed by this type of analysis which is often linked to detective works as described by Behrens (1997).

The researcher should listen to the data in as many ways as possible until a "plausible" story of the data is apparent. EDA has many approaches for graphical representation, such as scatter plot, frequency distribution, histogram plot, etc. This analysis aims to maximize insight into a data set, understand the underlying structure, extract important variables, detect outliers and anomalies and verify underlying assumptions.

During this study, various graphs and visualizations will be used to illustrates the distribution of data among the categories and subcategories. We will also look into the data distribution in years, months, the day of week and the time of day. Geographical distribution of the data over the Egypt's governorate will be explored as well.

3.3. Spatial-Temporal Analysis

As the geographical location is attached to each reported incident, the dataset is considered to be irregularly distributed spatial point pattern. The dataset has temporal component as well that will be considered during the study of the underlying phenomenon. Spatio-Temporal Point Process, rather than only Spatial Point Process, will be considered as potential model.

Spatio-Temporal Analysis holds the major part of the study in which we will try to identify clustering in time and space. We first consider an empirical analysis of the Spatio-Temporal Point Process of all crime incidents, in which we model the first-order intensity function as a product the spatial and the temporal intensities, and use the inhomogeneous Spatio-Temporal K-function to estimate second order intensity.

The package STPP in R will be used during the Spatio-Temporal Analysis which provides many of the models in applications of point process methods to

the study of Spatio-Temporal phenomena following the analysis conducted by Gabriel et al. (2013).

3.3.1. Defining Spatio-Temporal Point Processes

In any Spatio-Temporal Point Process dataset, the incidents of a countable set of points, $P = \{(s_i, t_i) : \ i = 1, 2, ...\}$ in which $s_i \in \mathbb{R}^2$ is the location and $t_i \in T \subset \mathbb{R}^+$ is the time of occurrence of the i-th incident within a bounded Spatio-Temporal observation window $S \times T$, where S is the space polygon and T is the time in a single closed interval and $Y(A)$ is the number of incidents in an a region A.

3.3.1.1. First and second-order intensities

First-order intensity are described by the intensity of the Spatio-Temporal Point Process.

$$\lambda(s,t) = \lim_{|ds|\to 0, |dt|\to 0} \frac{E[Y(ds,dt)]}{|ds||dt|},$$

where ds a small spatial region around the location s, $|ds|$ is its area, dt is a small interval containing the time t, $|dt|$ is the length of this interval and $Y(ds, dt)$ denotes number of incidents in $ds \times dt$. The intensity $\lambda(s, t)$ is the mean number of incidents per unit volume at the location (s, t) (Gabriel et al. 2013).

Second-order intensity describe the relationship between numbers of incidents in pairs of subregions within $S \times T$ as outlined below:

$$\lambda_2((s_i, t_i), (s_j, t_j)) = \lim_{|D_i|, |D_j|\to 0} \frac{E[Y(D_i)Y(D_j)]}{|D_i||D_j|},$$

where $D_i = ds_i \times dt_i$ and $D_j = ds_j \times dt_j$ are small cylinders containing the points (s_i, t_i) and (s_j, t_j), respectively (Gabriel et al. 2013).

3.3.2. Analyzing space-time point process data

Second-order intensity are used to analyze the Spatio-Temporal structure of a point process. The space-time inhomogeneous pair-correlation function and K-function can be used to measure of Spatio-Temporal clustering/regularity and as a measure of Spatio-Temporal interaction (Gabriel, Edith, and Diggle 2009).

3.3.2.1. Space-time inhomogeneous K-function

For a second-order intensity Spatio-Temporal Point Process, the Space-Time Inhomogeneous K-function (STIK-function):

$$K_{ST}(u,v) = 2\pi \int_0^v \int_0^u g(u',v')u'du'dv',$$

where $g(u,v) = \frac{\lambda_2(u,v)}{\lambda(s,t)\lambda(s',t')}$, $u = ||s - s'||$ and $v = |t - t'|$.

For the intensity of Inhomogeneous Spatio-Temporal Poisson Process less than zero, $K_{ST}(u,v) = \pi u^2 v$. Values of $K_{ST}(u,v)$ greater than $\pi u^2 v$ indicate aggregation at cumulative spatial and temporal separation less than u and v, otherwise it indicates regularity. The STIK function can also be used to verify for space-time clustering and space-time interaction (Gabriel et al. 2013).

3.3.2.2. Space-time inhomogeneous pair-correlation function

An estimator of the inhomogeneous space-time pair-correlation function is

$$\widehat{g}(u,v) = \frac{1}{|S \times T|} \frac{1}{4\pi u} \sum_{i=1}^{n} \sum_{j \neq j} \frac{1}{w_{ij} v_{ij}} \frac{k_s(u - ||s_i - s_j||) k_t(v - |t_i - t_j|)}{\lambda(s_i, s_i)\lambda(s_j, t_j)},$$

where w_{ij} and v_{ij} are the spatial and temporal edge correction factors and $k_s(.)$, $k_t(.)$ are kernel functions for space and time (Gabriel et al. 2013).

3.3.2.3. Implementation in R

The space-time inhomogeneous K-function and pair-correlation function calculation in this study will use the implementation in R provided by (Gabriel et al. 2013) in STPP package using the following functions:

- `STIKhat(xyt, space, time, dist, times, lambda)`
- `PCFhat(xyt, space, time, dist, times, lambda)`

Where xyt is the list of Spatio-Temporal point pattern. The parameters $space$ and $time$ are for the interval of space and time respectively. The parameters dist and times are vectors of distances u and times v at which $K(u,v)$ is computed. The parameter $lambda$ is a vector of values of the space-time intensity function evaluated at each points of x,y and time t.

4. Exploratory Data Analysis

The aim of EDA is to understand the data and try to discover the patterns and trends in the data. We will analyze the distribution of the data in categories, subcategories and in years, months, the day of week and the time of day and visualize the distribution of data on the geographical locations.

4.1. Data Preparation

This phase of analysis started by obtaining and cleaning up the data from the project database. The original dataset contains 2168 valid incidents. The dataset was filtered to include only the 4 categories with the most number of incidents. The study will only focus on reported incidents between 2011 and 2013. All incidents reported after 2013 will be ignored.

During the data cleanup process, we discover that 25% of incidents where mislocated. The online project used Google Geocoder to convert written addresses to coordinates. Some people were not able to locate the incident location, so they pointed to the nearest known location which resulted in some aggregation in few locations. We went through all the incidents and re-calculated the actual location using the address provided with the report. Example of the results are shown in Figure 4.1.

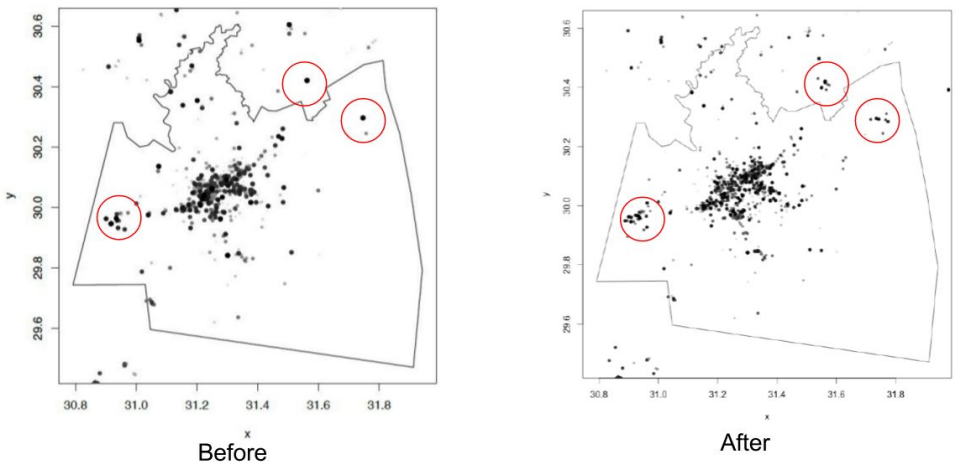


Figure 4.1: On the left, some incidents used the same location. On the right, the updated accurate location

The filtration process resulted in 1995 incidents classified into four categories and 13 sub-categories. The “Theft” Category contains the largest proportion of data of 40% of incidents slightly followed by “Building Violations” with 35% incidents. The other two categories “Violence - Illegal Trade” and “Corruption” account for 25% of roughly the same number of incidents as in Figure 4.2.

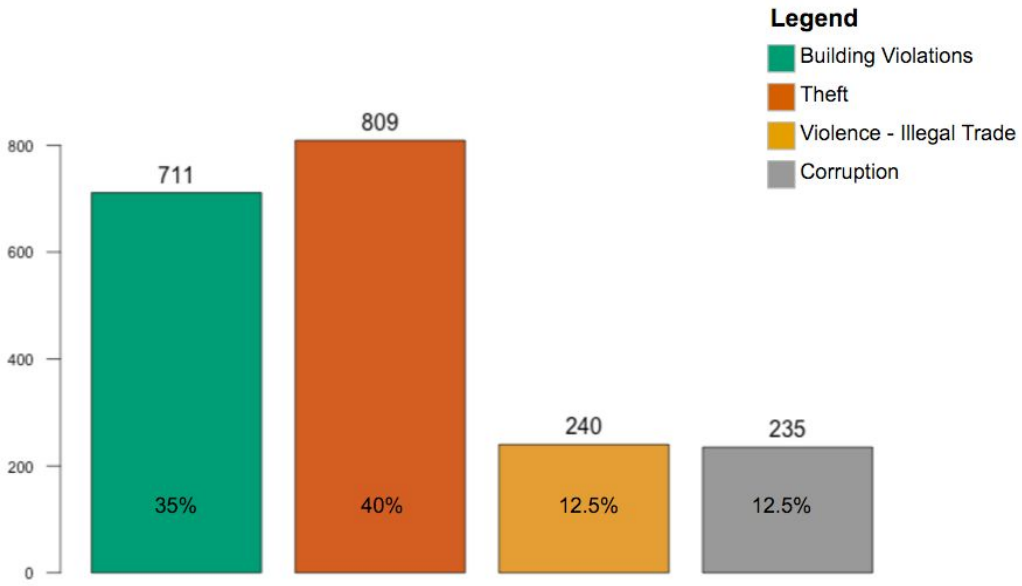


Figure 4.2: Incidents per categories.

The following color convention will be using during the analysis:

Category	Color
Building Violations	Green
Theft	Orange
Violence - Illegal Trade	Yellow
Corruption	Grey

Figure 4.3 illustrates the distribution of data in Subcategories. The highest number of the reported incidents in the sub-categories is in the “Car Theft” subcategory in “Theft” category which accounts for 24% followed by “Building on Agricultural Land” from “Building Violations” category with 21%. “Theft of property” ranked the 3rd by 15% of total incidents. The other sub-categories have low percentages of incidents.

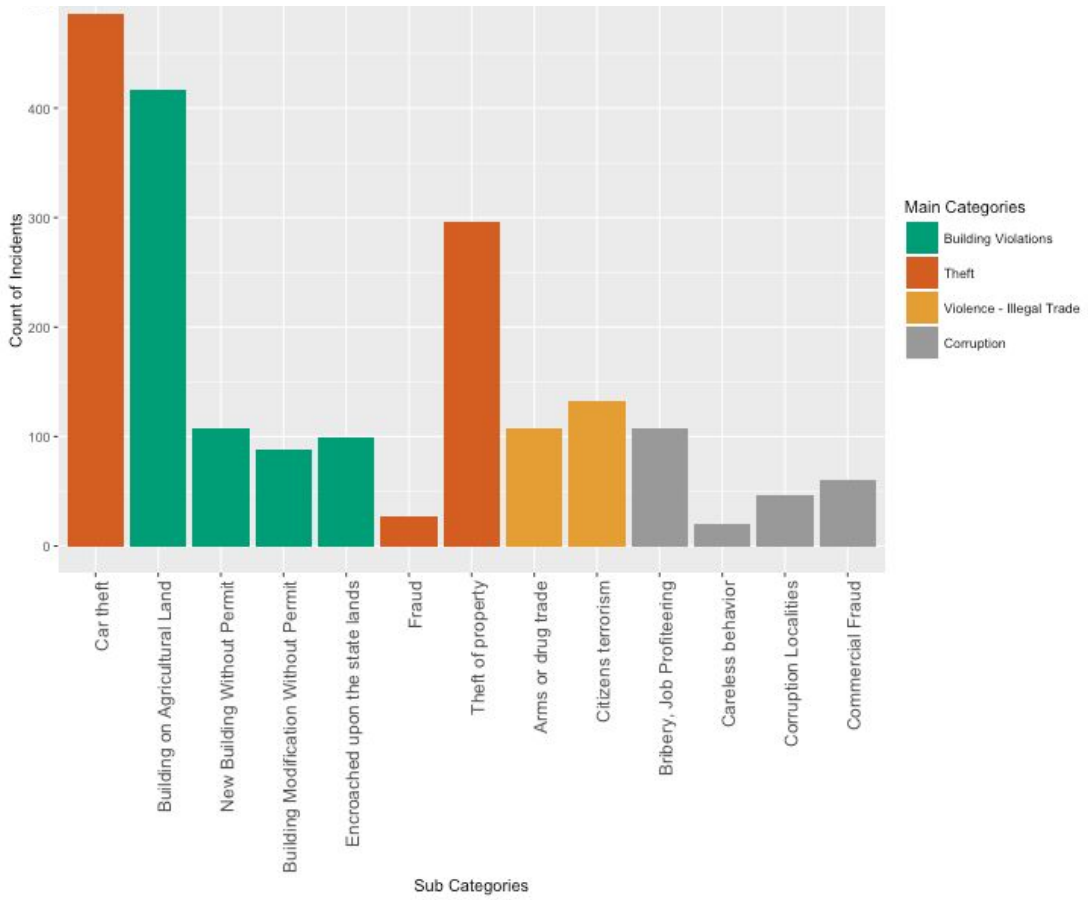


Figure 4.3: Incidents per sub-categories.

4.2. Crime Trend in Egypt

As stated earlier, there is no official data from the government for comparison. However, United Nations Office on Drugs and Crime published some statistics about crime trend in Egypt from 2006 to 2011 which shows in Figure 4.4 slight increase in Burglary and Car-Theft before 2011 with a massive increase in 2011 because of the events of the revolution and the absence of the law enforcement.

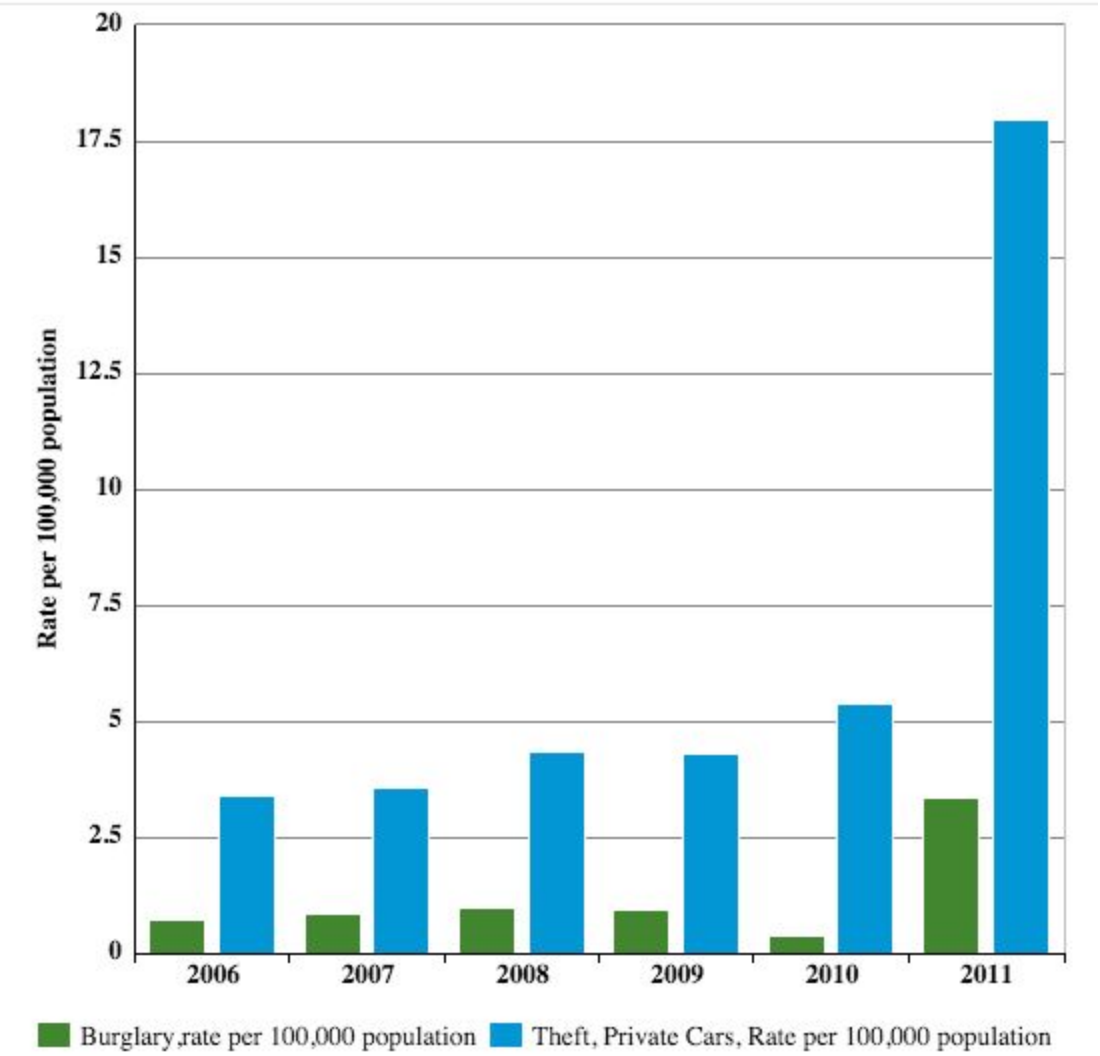


Figure 4.4: Burglary, Car-Theft rates in Egypt between 2006 - 2011 from (UNODC 25 June, 2015)

Figure 4.4 shows Burglary, Theft of private cars rates in Egypt between 2006 and 2011 according to United Nations Office on Drugs and Crime (UNODC 25 June, 2015). Statistics reported to the United Nations by to the authorities of countries which could also be subject to accuracy issues. Anyhow, the crowdsourced data confirms and complements the increasing trend in 2011 and 2012 as in Figure 4.5.

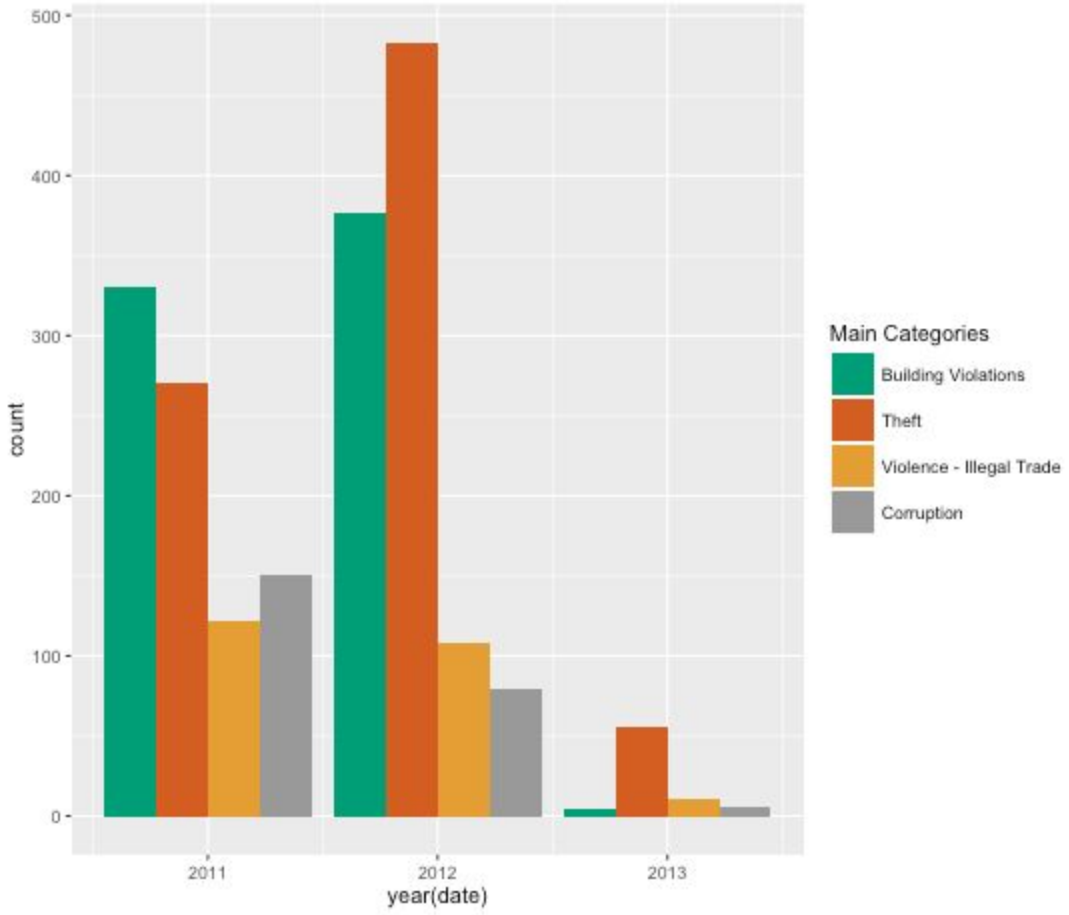


Figure 4.5: Incidents per year.

The distribution of data in years shows an interesting trend. While the data crowdsourcing activity started early 2011, more than 50% of incidents were reported in first half of 2012. Still, 43% were reported in 2011 and a slight percentage in 2013. 38% of “Building Violations” incidents were reported in 2011 with a slight increase in 2012. The second half shows a decline in the number of reports in Figure 4.6 as Egypt had the presidential election and the form of the new government where law enforcement activities were gradually restored and fewer reports were received online.

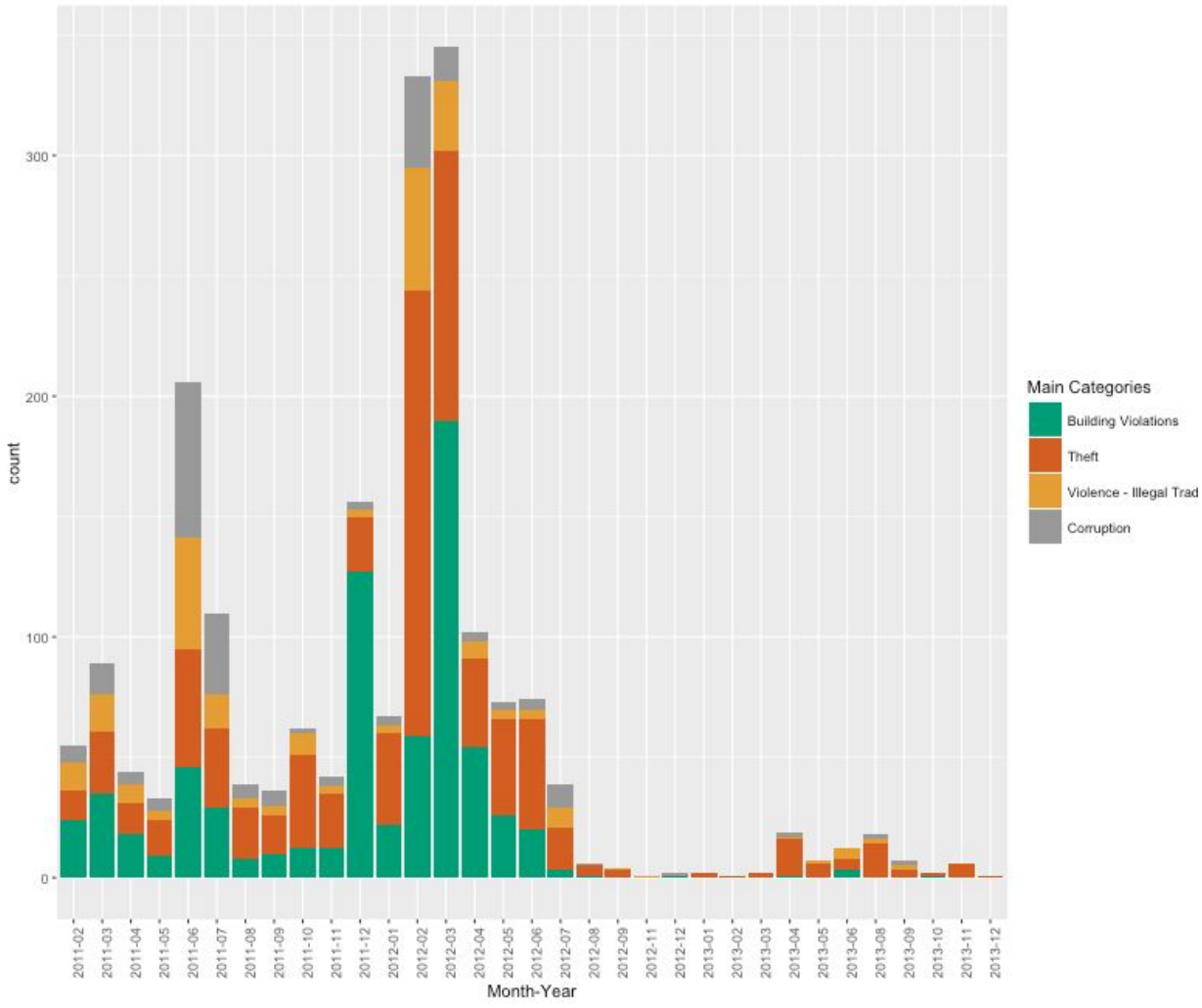


Figure 4.6: Incidents per Month-Year.

4.3. Geographical Distribution of Crime Incidents

On a geographical level, comparing the percentage of reports on each governorate in Egypt with population percentage shows interesting relation. There is a positive relation between the number of reported incidents and the population density which is logical as the more people live in a geographical location; the more reports are received from that location, many other factors affect this relation such as education level, the internet and mobile penetration.

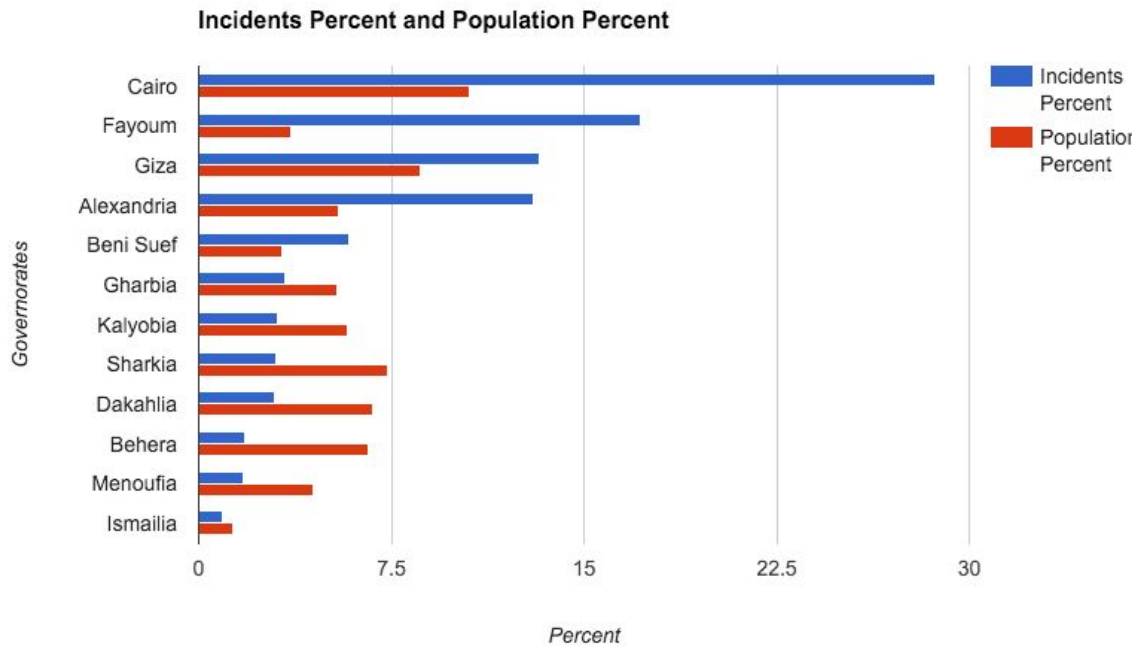


Figure 4.9: Incidents percentage and population percentage in Governorates

Figure 4.9 compares the population percentage with reported incidents rate. The city of Cairo as expected has the majority of population percentage of 10.7% (estimated 9 million inhabitants) which also has the majority of reported incidents with 28%. Fayoum, which is a rural governorate, has the second most reported incidents of 13% while it has 3.6% of total population. Giza and Alexandria, which are urban cities, have almost the same percentage of reported incidents of 13% while they have different population percentage of 8.6% and 5.4% respectively. On the other hand, Beni Suef, which is a rural area, has incidents rate of 5.8% while it has small population percentage of 3.2%.

While population percentage has an enormous influence on the number of reports as illustrated in Figure 4.10, other factors should be considered. Education level is relatively significant as only highly educated citizens would use the internet and social media to report outlaws activities. The Internet and

mobile penetration are important as well as the crowdsourcing project emphasis of the use of the internet, mobile app and social media to report a crime.

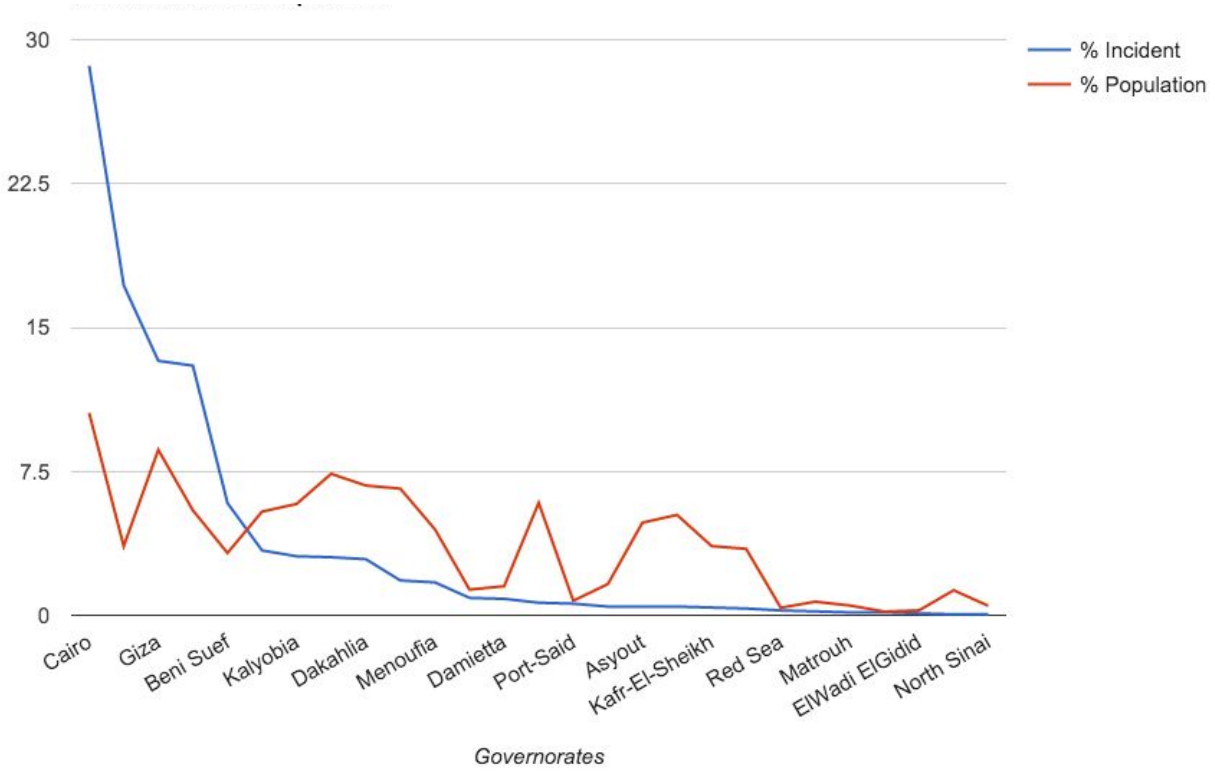


Figure 4.10: The relation between population and incidents percentages.

Distribution the incidents using the selected color code over the map of Egypt shows a very interesting pattern as in Figure 4.11. The Building Violations incidents, which are represented by green dots, are highly concentrated in the rural areas in Fayoum and Beni Suef, on the north coast in Alexandria and in the northern parts of Cairo and Giza. Theft incidents are highly concentrated in the Greater Cairo area with less concentration towards the Mediterranean Sea in the north.

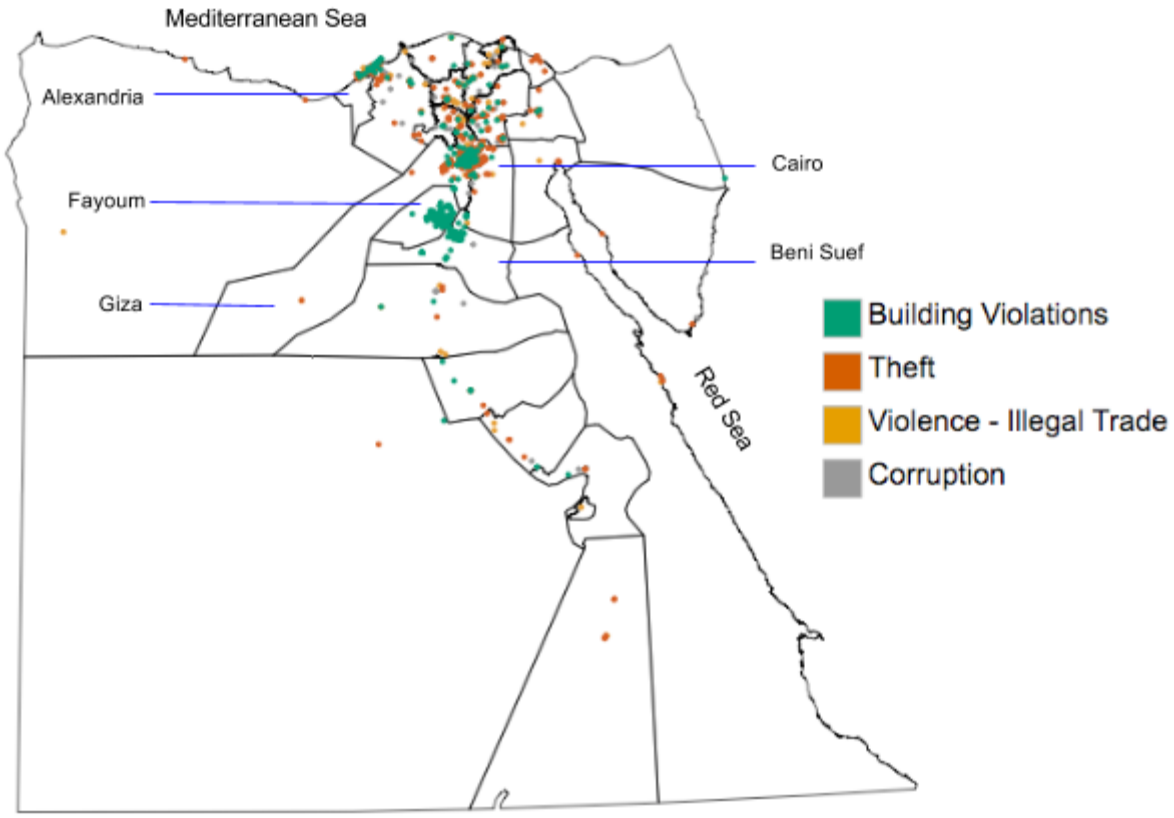


Figure 4.11: Incidents distribution over Egypt.

From Figure 4.12, It is clear that Theft tends to take place in urban governorates such as Cairo while Giza while Building Violations happens in rural areas in Fayoum and Beni Suef which is logical and confirmed by the data. There is one exception to this pattern with Alexandria the coastal city has high numbers of both Theft and Building Violation as well.

Alexandria, unlike other coastal cities, has tight regulations on new constructions and the number of floors in the buildings at seaside which explains why some people might have taken the opportunity of the absence of the law enforcements and broke the law by increasing the number of floors illegally or removing old themed buildings and constructing new ones which are also prohibited. The majority of Building Violations reports in Alexandria had

multimedia content such as pictures and videos to document and to increase the credibility of the information which is mostly reported by neighbors who are directly affected by such violation.

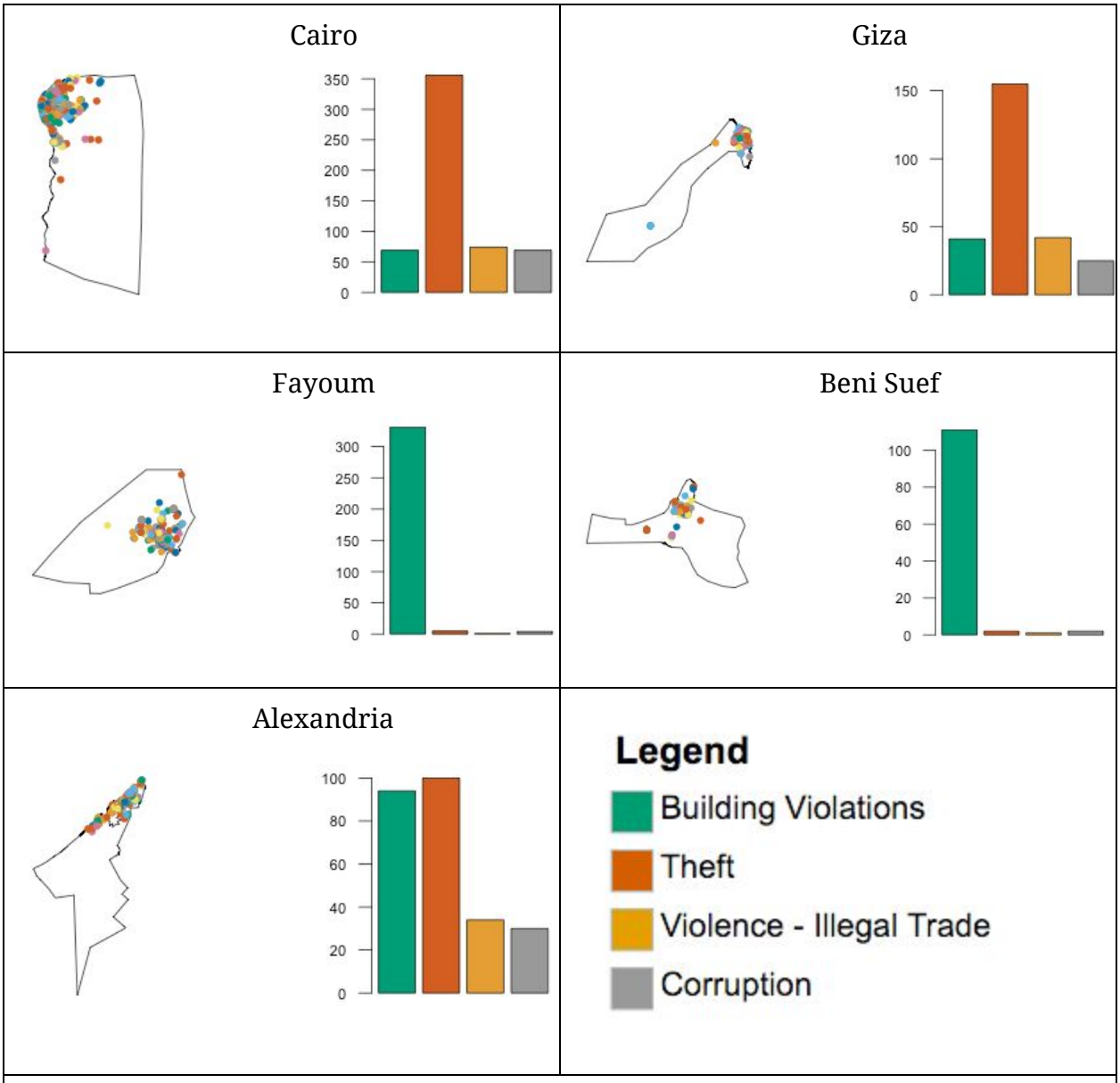


Figure 4.12: Distribution of incidents in Categories in top 5 governorates.

4.4. Building Violations

Distributing the data over the week showed an interesting trend regarding Building Violations during the weekends in Egypt where the weekend days are Fridays and Saturdays. As shown in Figure 4.12, there is an increasing trend in Building Violations reported incidents start the day before the weekend, which is Thursday, and continue increasing on Friday with a slight decrease in Saturday then goes low again during the weekdays.

This behavior could be explained as the law attention of the law enforcement during weekends and holidays. Hence, the offender tries to impose the adherence to the status quo which could lead later to a paying a fine. However, paying a penalty is much less than the profit of violating the building laws. Such statistics could be a good alarming for the policymakers to amend the current legislation to penalize such act.

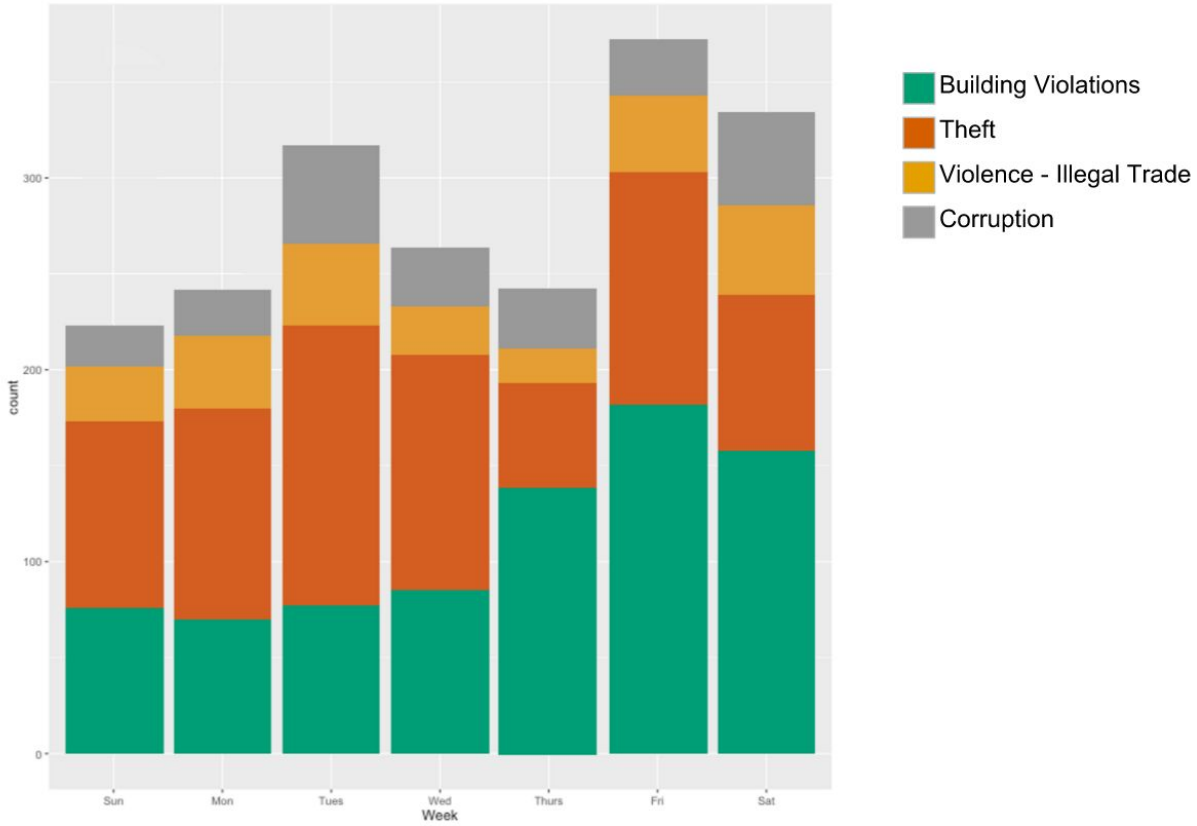


Figure 4.12: Distribution of the data over weekdays.

4.5. Time of Car-Theft

All reported incidents have time component which is obligatory for the reporter to indicate the date and time of the incidents. However, time is not relevant in every categories. Therefore, we will only consider the time information with Theft and more specifically Car-Theft.

Analyzing the relation between Car-Theft incidents and the time and location of the incidents shows interesting findings in Figure 4.12. Car-Theft is the most reported crime by 24% of all reports. All the reported incidents have the time component which the reporter believes the car was stolen at which happens between the last sight of the car and the discovery of the theft. Only the hour of the day is considered in this analysis.

The function of the location was obtained by examining the geographical location using a Google Maps and the familiarity of the author of the provided location. The address information and textual description of incidents were also examined which sometimes indicates the function of the location such as “next to my home”, “nearby the club or a court” and “In front of work pace”.

The location of the incident is classified into five classes; Residential, Business, Entertainment, Highways, and Unknown. The incident is considered to be in Residential class if the incident took place in famous residential areas, near hotels or the reporter mentions explicitly that the car was stolen near where he or she lives. The incident is classified as “Business” if it took place near famous business areas, city centers, near hospitals, mosques and supermarkets or explicitly mentioned by the reporter. For clubs, malls, coffee shops and restaurants, the incident is considered in “Entertainment” class. “Highway” category is only considered if the incident took place outside the city or on highway roads. Otherwise, it is classified as “Unknown”.

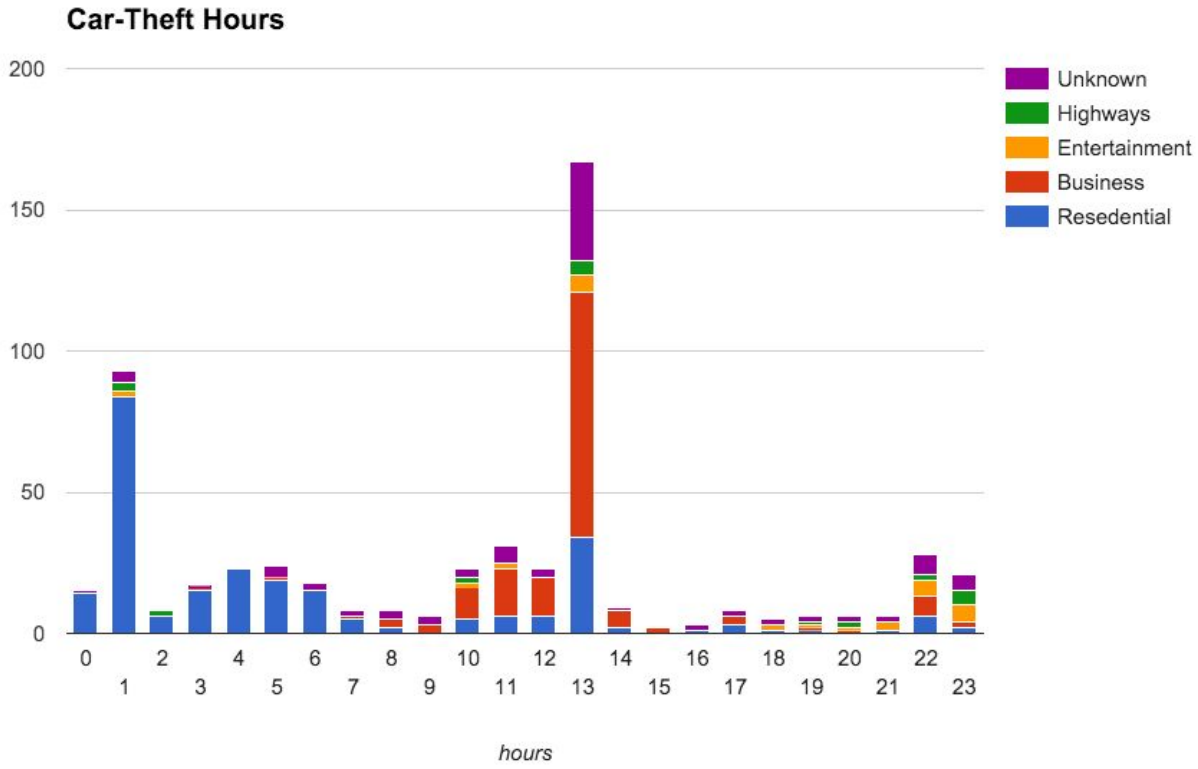


Figure 4.13: The relation between the time of car-theft incident and the function of the location.

By visual interpretation to Figure 4.13, it is clear that there are two peak hours at 1:00am and 1:00pm indicating the time of most Car-Theft reported crimes happened and fewer cases reported during other times. The majority of incidents in Residential areas occur after midnight which is logical and confirmed by the data.

People tend to think that their cars were stolen at 1:00am which could be influenced by the daily street activity patterns. Usually in Egypt, street activities tend to end between 11:00 pm and midnight and the streets slightly get crowded again between 3am-4am for Fajr prayer, then the normal activities starts after 5:00 am. This clearly gives an opportunity to Car-Theft incidents to take place

from midnight to just before the Fajr prayer. This explains why people tend to think why their car could be stolen at this hour of the night.

The other peak at 1:00 pm is more interesting and needs careful analysis. The majority of incidents at that time is classified as Business which is logical as it is in the middle of the day. The typical private sector work day in Egypt starts at 8:00-9:00am and ends at 5:00-6:00pm with one hour break in between which usually take place at 2:00-3:00pm. During the lunch break, the person may interact again with his or her car which leads to the discovery of the theft. The car could potentially be stolen at any time from the last interaction to the time of the discovery. However, people tend to think that the theft took place in the middle of the business day where less attention and interaction with the car. The findings here is also confirmed by Ratcliffe(Ratcliffe 2000) in his study where he confirmed vehicle theft mostly takes places at night in residential areas while nonresidential areas during the day.

5. Spatio-Temporal Analysis

The dataset contains the location and time components which make it suitable for Spatio-Temporal Analysis. While the provided dataset contains data from all over Egypt, the Spatio-Temporal Analysis will begin by analyzing the all reported incidents over the area of Greater Cairo, which has roughly half of all reported incidents with defined geographical boundaries as illustrated in Figure 5.1 and discussed in details in Section 2.6. Later on, we will perform the Spatio-Temporal Analysis of a newly defined window called “Central Cairo” to exclude the areas where no reported incidents.

All reported incidents were georeferenced and mapped using GPS coordinates of latitude and longitude. Euclidean distance will be used for distance measurement so that all coordinates were converted to two-dimensional Cartesian coordinate system (UTM projection, WGS84 Datum, Zone 35-N).

5.1. Spatio-Temporal for the “Greater Cairo” area

The Spatio-Temporal Analysis started by examining the spatial and temporal intensity in the selected dataset inside Greater Cairo boundaries in Figure 5.1.

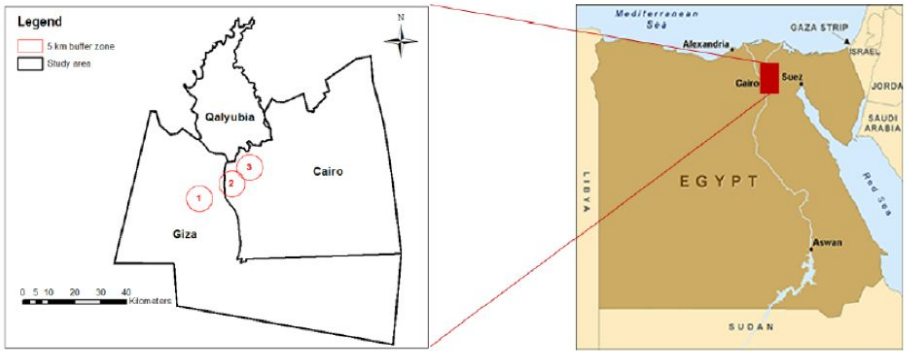


Figure 5.1: The location and boundaries of Greater Cairo. (Megahed et al. 2015)

From Figure 5.2, we can see a clear spatial cluster in the center of Greater Cairo, which is the joining point between the three governorates that form Greater Cairo as illustrated in Figure 5.1. From the map, we can see fewer incidents on the west and north and almost no incidents at the Southeastern part of the map.

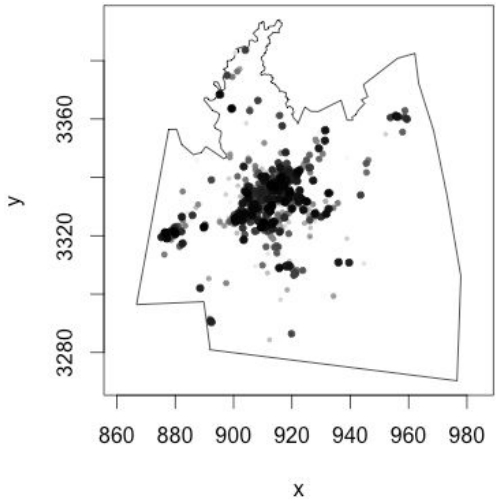
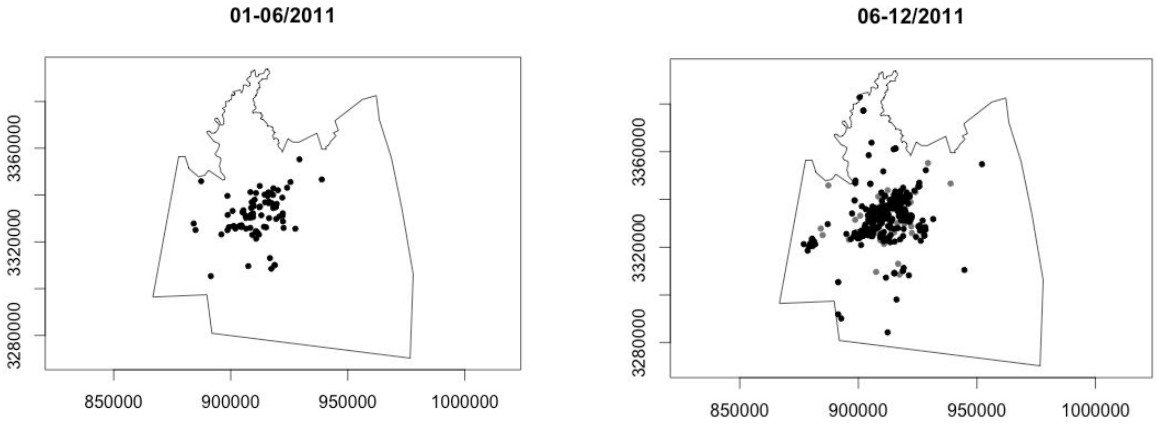


Figure 5.2: Spatial distribution of reported incidents over Greater Cairo area.

Figure. 5.3 shows the evolving spatial pattern of the crime incidents in Greater Cairo area where roughly half of all incidents were reported. The Figure illustrates several temporal snapshots; the original data record both the location and date of each of 795 incidents over the period 1 February 2011–31 December 2013.



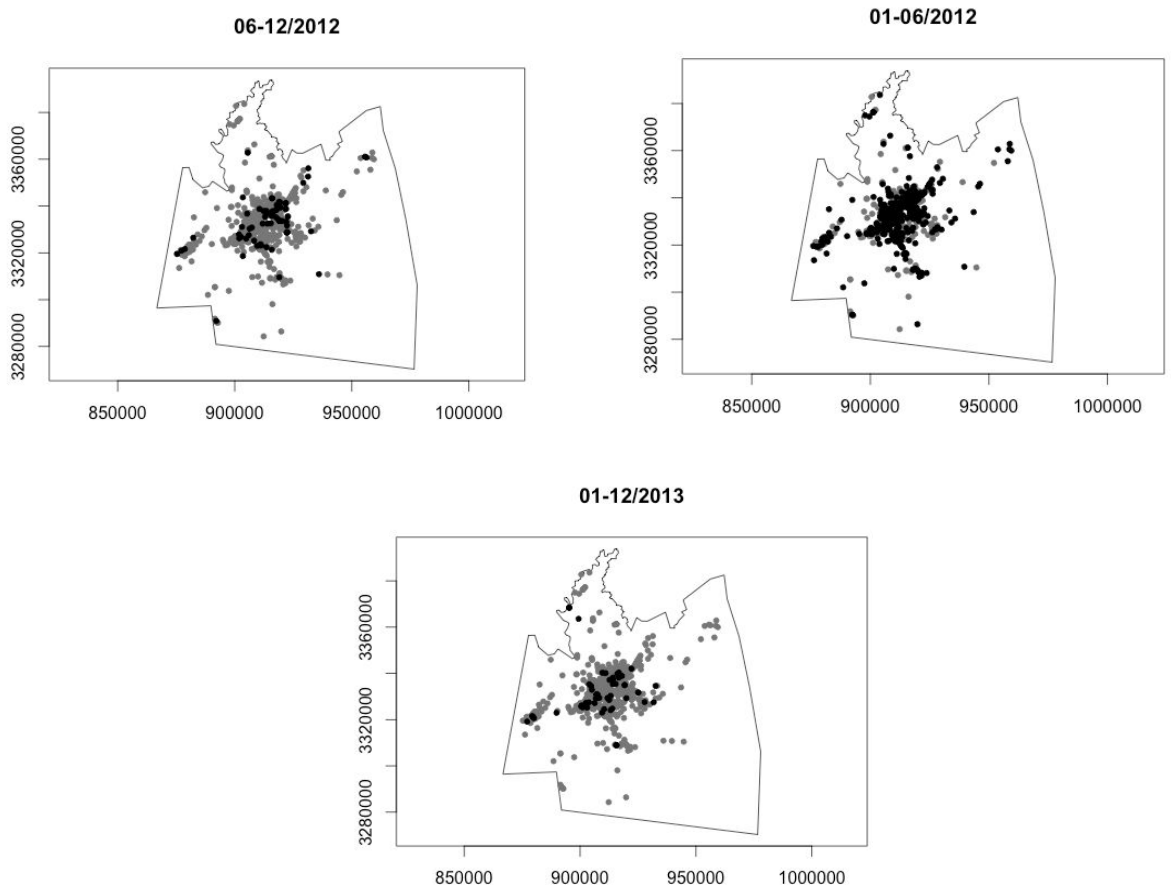


Figure 5.3: The evolving pattern of current incidents in black and past incidents in grey in a discrete-time sequence from 2011 to 2013.

The basic format of the data is $\{(x_i, t_i) : \ i = 1, \ldots, n\}$, where $x_i \in W \subset R^2$ denotes the location and $t_i \in (0, T) \subset R^+$ the corresponding time in days since 1th of February, 2011.

A Spatio-Temporal Point Process can be analyzed as a spatially or temporally marked. Therefore, it can be considered as temporally marked to locate a certain pattern in space and time (Gabriel, Edith, and Diggle 2009). The resulting plot is shown in Figure 5.4.

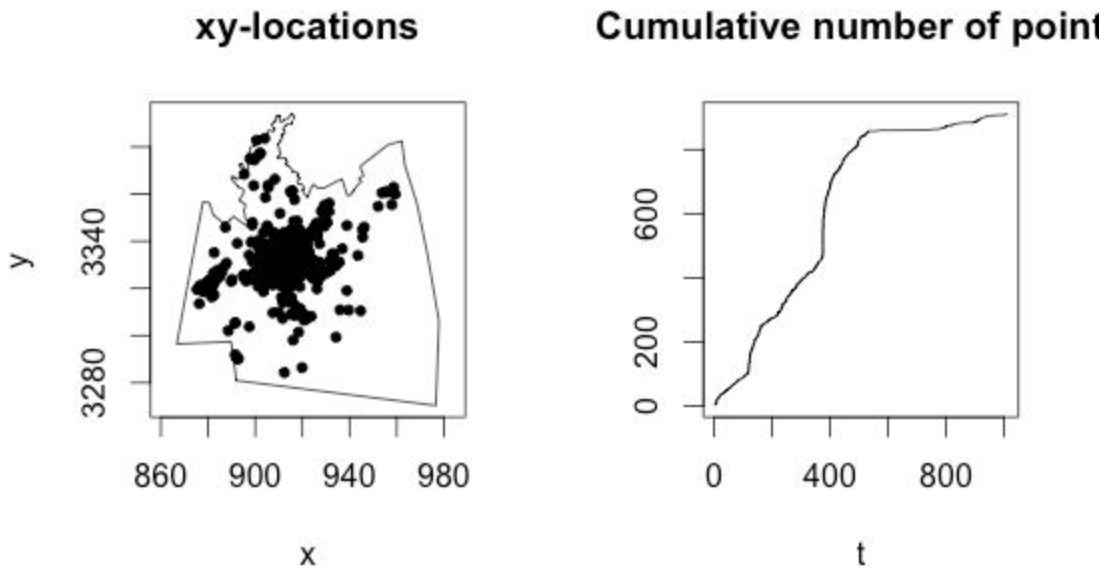


Figure 5.4: Left-hand panel shows a static display of the data consisting of locations. Right-hand panel shows cumulative distribution of the times.

The spatial pattern displayed in Figure 5.4 showing a clear “star” structure in the very upper right corner to the center. In order to investigate the Spatio-Temporal cluster, we calculated the corresponding K-function and pair-correlation function using the available implementation in STPP package in R provided by (Gabriel et al., 2013).

The result of the Space-Time Inhomogeneous K-function `STIKhat` is presented in Figure 5.5. To test if Spatio-Temporal clustering or regularity of the point pattern exists we computed $\hat{k}(u,v) - 2\pi u^2 v$. Thus, negative values indicate $\hat{k}(u,v) < 2\pi u^2 v$ as in Figure 5.5.

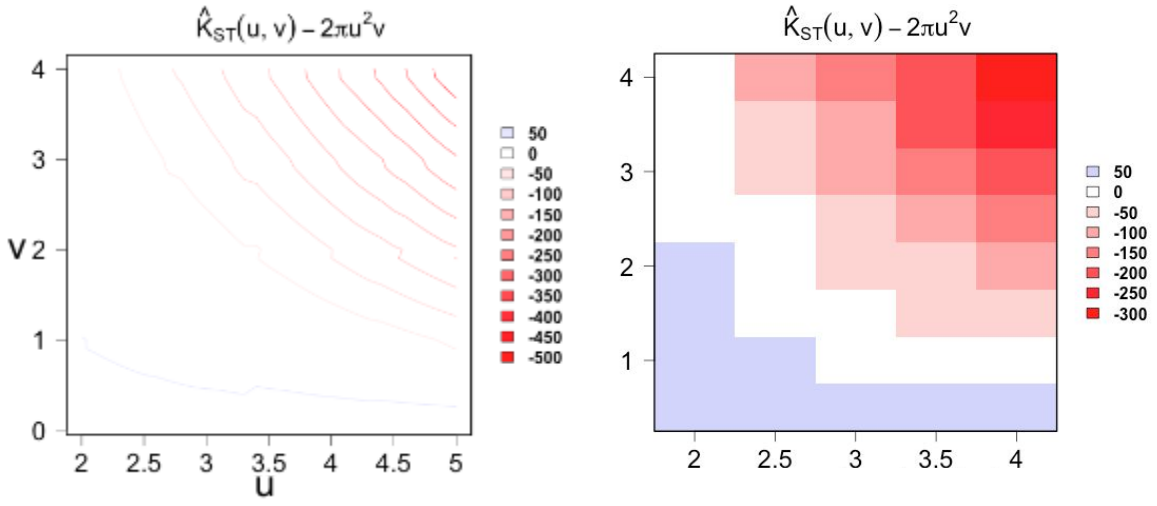


Figure 5.5: The spatio-Temporal inhomogeneous K-function contour plot applied on "Greater Cairo" area (left). Comparison between $\hat{k}(u,v) - 2\pi u^2 v$ and tolerance envelopes indicating Spatio-Temporal clustering in blue shading (right).

The resulting plot is shown here, where the v distances in kilometers are displayed on the y axis while the u distances are times in days shown on the x axis. Clearly, inspecting the contour lines visualizes some variation of $\hat{k}(u,v) - 2\pi u^2 v$. The blue colours indicate aggregation, red colours indicates regularity. Thus, this plot emphasized a highly aggregated point pattern within $u = [2, 4.5]$ and $v = [0, 2]$. In contrast, we observed regularity only in the very upper right corner of Figure 5.5. This means that a crime incident is more likely to happen once more within a period of 2 days at a distance of 4.5km from an incident in Greater Cairo.

Investigating the interaction between incidents in space and time is crucial to the analysis to see if the crime is navigating in space and time from one place to another. We will apply the Space-Time Inhomogeneous pair-correlation function `plotPCF` using the available implementation in STPP package in R provided by (Gabriel et al. 2013) as well which resulted in the plot in Figure 5.6.

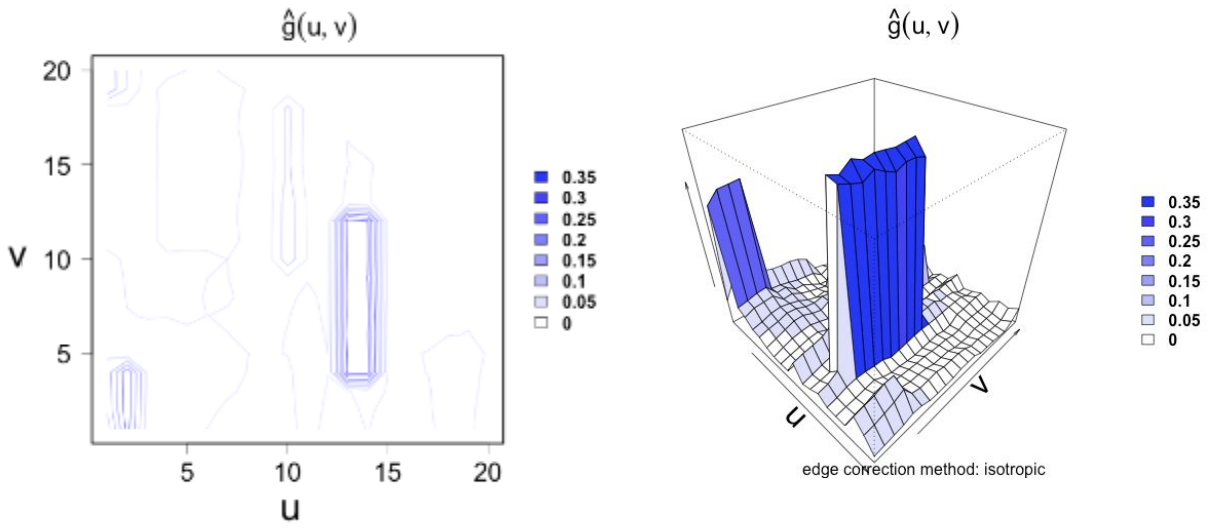


Figure 5.6: The space-time inhomogeneous pair-correlation function for Greater Cairo area. Contour plot (left), perspective plot (right)

Here, corresponds to the contour plot and perspective plot of the pair-correlation function applied on Greater Cairo area. Values greater than one indicate clustering and values less than one indicates regularly. In our case, all values are below 1 which means there is no interaction between incidents in space and time over Greater Cairo area.

As shown earlier in Figure 5.2, there is a vast area where no data is available. Both Space-Time inhomogeneous K-function and pair-correlation function depend on the spatial and temporal intensity which make them sensitive to areas with no data which will be investigated in the next section.

5.2. Spatio-Temporal for “Central Cairo” area

The problem with investigating Spatio-Temporal clustering in Greater Cairo is the vast area where no reported incidents. Figure 5.7 shows the distribution of reported incidents among the three governorates which define the Greater Cairo area. It is clear that the majority of incidents are aggregated on the shared borders of Cairo, Giza, and Kalyoubia governorates.

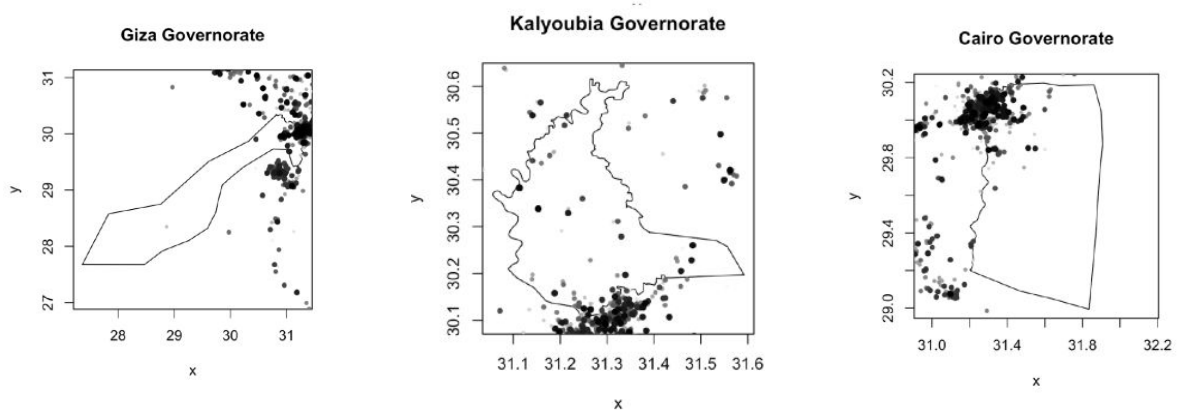


Figure 5.7: Shows the distribution of incidents among Giza, Cairo and Kalyoubia governorates.

To further investigate the Spatio-Temporal clustering, we will define a new window which will be a subset of Greater Cairo area excluding the South Eastern parts which have less reported incidents. We will call the new window “Central Cairo” area as in Figure 5.8.

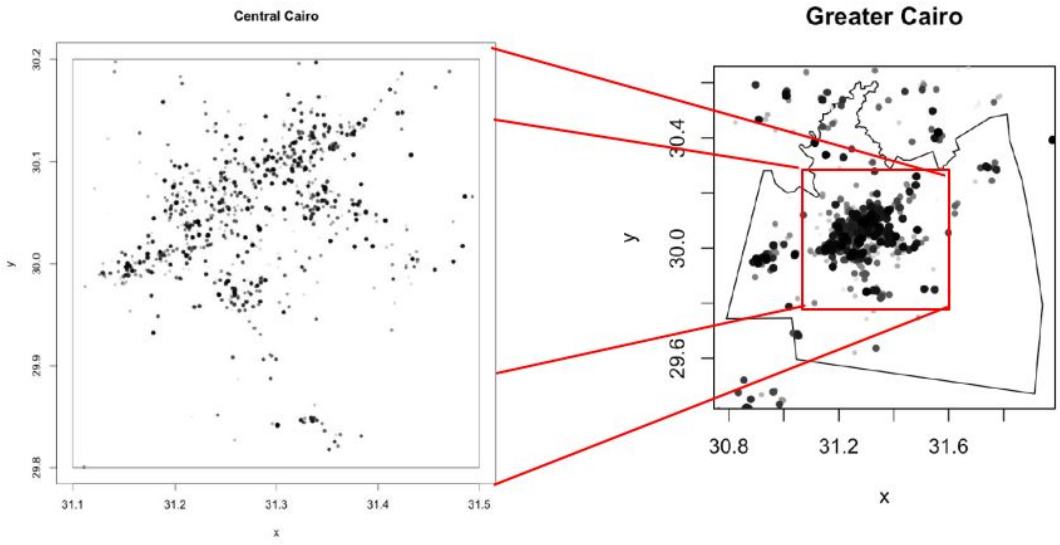


Figure 5.8: Defining “Central Greater Cairo” area.

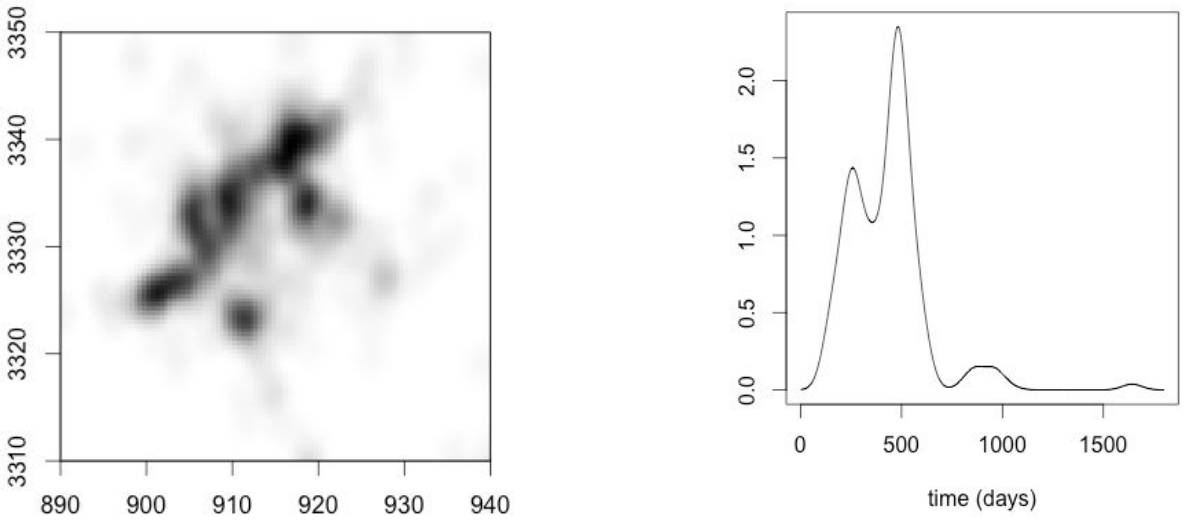


Figure 5.9: Spatial (left) and temporal (right) intensity functions estimated. Time is treated as a quantitative mark; light grey small dots correspond to the oldest incidents and dark grey/large dots correspond to the most recent incidents.

The spatial pattern displayed in Figure 5.8 is found in Figure 5.9 showing a clear "star" pattern in the very upper right corner to the center similar to Greater Cairo. In order to investigate the Spatio-Temporal cluster, we calculated the corresponding K-function and pair-correlation function.

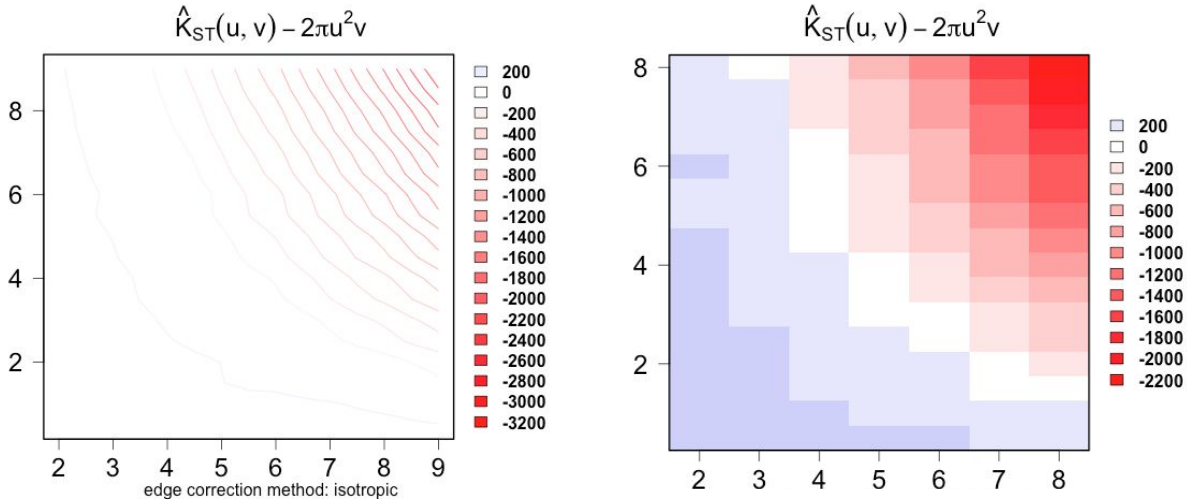


Figure 5.10. The Spatio-Temporal inhomogeneous K-function Contour plot (left), and tolerance envelopes indicating clustering in blue shading (right) over Central Cairo area.

To test for Spatio-Temporal clustering or regularity of the point pattern exists we will calculate $\hat{k}(u,v) - 2\pi u^2 v$. Thus, negative values indicate $\hat{k}(u,v) < 2\pi u^2 v$.

The resulting plot is shown here in Figure 5.10, where the v distances in kilometers are displayed on the y axis while the u distances are times in days shown on the x axis. Clearly, inspecting the contour lines visualizes some variation of $\hat{k}(u,v) - 2\pi u^2 v$. The blue colours indicate aggregation, red colours indicates regularity. Thus, this plot emphasized a highly aggregated point pattern within $u = [2,8]$ and $v = [0,8]$ which is much better than the results from Greater Cairo area.

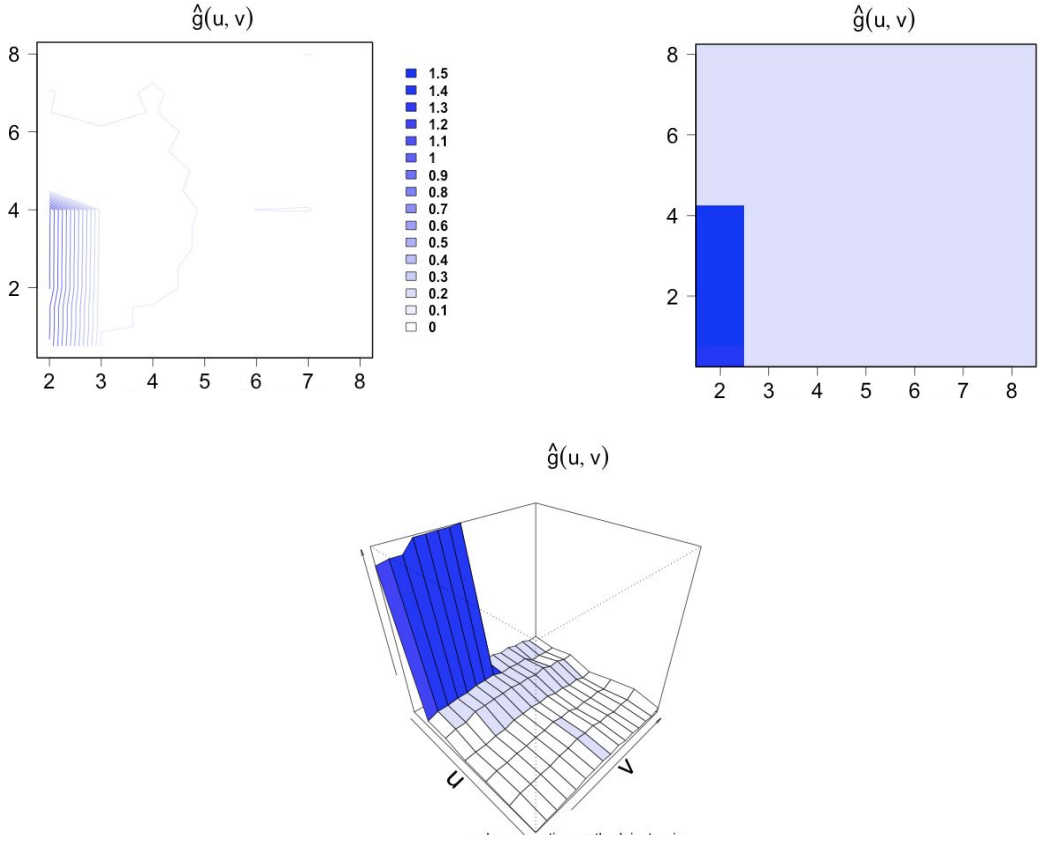


Figure 5.11: The space-time inhomogeneous pair-correlation function, contour plot (top-left) and tolerance envelopes (top-right), perspective plot (bottom)

In Figure 5.11, the pair-correlation function is plotted, where the v distances in kilometers are displayed on the y axis while the u distances are times in days shown on the x axis. We see a high level pair-correlation within small spatial and temporal distances of 4 days and radius of 2km. This relation slightly decreased from bottom to top and left to right and vanished completely when inspecting the upper boundaries of u and v.

The findings concerning the scales of Spatio-Temporal clustering and interaction do not admit a precise mathematical interpretation, but rather suggest the approximate ranges where we might seek to develop a more explanation for the Spatio-Temporal relation in the data (Tamayo-Uria et al. 2014).

We observed that incidents recur in New Cairo neighborhood which is a newly constructed residential area east of Cairo. Four days after the first incident, new incidents occur within a radius of 2km around the first incident and they take place a few days later.

6. Discussion and Conclusions

Crime is a negative social phenomena which can be tackled by many approaches. One approach was presented in this study by utilizing the power of the crowd and richness of crowdsourced data. This research has presented a case study on that manner.

This study investigated the crime patterns and trends in Egypt, and tried to find the relation between the spatial and temporal variables in crowdsourced data collected from Zabatak.com, an online project founded by the author after the 25th of January 2011 revolution collecting crime information after the downfall of the regime. The dataset consists of more than 2000 crime incidents from various geographical areas across Egypt.

This study argues the usefulness of the crowdsourced data despite the inherited limitations regarding the quality, digital gap, data diversity and privacy which could be tackled by several approaches as presented. Crowdsourced data can also be a mean to provide richer and diverse data, such as multimedia content and accurate GPS location, which was not available nor accessible with the traditional means. Such actionable information could help the citizens to take proactive measures protecting their assets, avoid certain places or plan their daily activities such as travel and work. In addition, it could help policymakers to understand better crime evolution which may lead to reduction eventually.

The research started with applying exploratory analysis methodologies to the data trying to interpret crime patterns, trends, and distribution in time and space. The results of this method have identified various interesting trends and patterns in the dataset. One of the major findings of this research points out a

strong relationship between the spatial and temporal variables in Car-Theft datasets where the possible time of Car-Theft has been identified. According to the data, Car-Theft incidents in Residential areas usually take place after midnight with a peak at 1 am. During the day, the owners of the stolen cars think that their cars were stolen at 1 pm in business areas.

We were able to confirm the increasing trend of crime in Egypt between 2011 to 2013 which complements the data provided by UNODC between 2006 and 2011. However, the trend went down after the presidential election in the second half of 2012 and continued to drop after the form of the new government and restoring policing activities.

The data also detected the day of the week where Building Violations incidents mostly happen which usually take place during weekends (Fridays and Saturdays) and the day before the weekend due to the low attention of the police. Hence, imposing the adherence to the status quo which results in the tendency of violators preferring to pay a fine which is much less than the profit they made off the violation.

Visualizing the data distribution over the map of Egypt showed an interesting pattern. It was possible in the study to relate crime types to the type of the geographical area. Theft tends to happen in urban areas while Building Violations tend to take place in rural areas, with exception to Alexandria, the coastal city, which had both types due to the relatively tighter building regulations which some people broke when they had the opportunity.

Furthermore, the research applied Spatio-Temporal analysis using the inhomogeneous Spatio-Temporal K-function and pair-correlation function. The results also confirmed a Spatio-Temporal clustering where crime is likely to

happen once more within two days and within a radius of four and half kilometers in Greater Cairo area. The analysis also identified a slight Spatio-Temporal interaction in four days within a radius of two kilometers where crimes may recur in Central Cairo area which was observed in New Cairo neighbourhood in East of Cairo.

These findings in crime data can open new ways for crime maps data analysis. The data confirmed many of logical relations between crime and space and time. These results can be used to document the current crime scene in Egypt and could be used later as a benchmark for further comparison and analysis when more data is available. There are some limitations to the techniques employed in this study, which are discussed in the next section.

6.1. Limitations

While the findings presented in this study confirm the usefulness and richness of crowdsourced data, there may be a few drawbacks using crowdsourced data as the only source of information, which was the case of this study, as no other sources of data was available for comparison.

As we have presented the patterns of crime extracted from the crowdsourced data, we have to be careful not to generalize this all over the country beyond the years of the study. Half of the reported incidents were coming from a subtle geographical area and from a fraction of the population compared to the country size and population density, whereas there is no obligation from the users to provide correct and complete information about incidents.

Visual interpretation and experience with the area of study and familiarity with the political and critical events were crucial during the exploratory analysis

which may be subjective and incomplete if examined by unfamiliar reader. However, we have been cautious to find suitable citation whenever possible.

The findings concerning the scales of Spatio-Temporal clustering and interaction do not admit a precise mathematical interpretation, but rather suggest the approximate ranges where we might seek to develop a more explanation for the Spatio-Temporal relation in the data (Tamayo-Uria et al. 2014).

Although the scope of this study is limited, this case study can serve as a basis for other studies using the same methodological framework and guide policymakers in governmental and non-governmental organizations in crime mapping using crowdsourced data to understand and combat crime in certain geographic locations.

7. References

Behrens, John T. 1997. “Principles and Procedures of Exploratory Data Analysis.” *Psychological Methods* 2 (2): 131–60.

CAPMAS. 2014. “Central Agency for Public Mobilization and Statistics.” http://www.capmas.gov.eg/.

Chainey, Spencer, and Jerry Ratcliffe. 2013. *GIS and Crime Mapping*. John Wiley & Sons.

Clarke, Ronald V. 2004. “New Challenges for Research: Technology, Criminology and Crime Science.” In *Crime and Technology*, 97–104.

Coleman, David J. 2012. “Potential Contributions and Challenges of VGI for Conventional Topographic Base-Mapping Programs.” In *Crowdsourcing Geographic Knowledge*, 245–63.

Cowan, Terri. 2013. “A Framework for Investigating Volunteered Geographic Information Relevance in Planning.” *University of Waterloo, Ontario, Canada*.

Deparday, Vivien. 2010. “Enhancing Volunteered Geographical Information (VGI) Visualization with Open Source Web-Based Software.” *University of Waterloo, Ontario, Canada*.

“Egypt Smartphone Survey 2012.” 2016. Accessed January 7. http://arabadvisors.com/node/13876.

Elwood, S. 2008. “Geographic Information Science: New Geovisualization Technologies -- Emerging Questions and Linkages with GIScience Research.” *Progress in Human Geography* 33 (2): 256–63.

Gabriel, Edith, Gabriel Edith, Rowlingson Barry, and Diggle Peter. 2013. “Stpp : An R Package for Plotting, Simulating and Analyzing Spatio-Temporal Point Patterns.” *Journal of Statistical Software* 53 (2). doi:10.18637/jss.v053.i02.

Gabriel, Edith, Gabriel Edith, and Peter J. Diggle. 2009. “Second-Order Analysis of Inhomogeneous Spatio-Temporal Point Process Data.” *Statistica Neerlandica* 63 (1): 43–51.

Goodchild, Michael F. 2007. “Citizens as Sensors: The World of Volunteered Geography.” *GeoJournal* 69 (4): 211–21.

Goodchild, Michael F., and Li Linna. 2012. “Assuring the Quality of Volunteered Geographic Information.” *Spatial Statistics* 1: 110–20.

Goolsby, Rebecca, and Goolsby Rebecca. 2010. “Social Media as Crisis Platform.” *ACM Transactions on Intelligent Systems and Technology* 1 (1): 1–11.

Ismail, Ayman M. 2012. “People GIS A Web2.0 Approach to Confronting Landuse Violations.”

Maltz, Michael D., Andrew C. Gordon, and Friedman Warren. 1991. “Thoughts on Communication in Police Departments.” In *Mapping Crime in Its Community Setting*, 144–47.

Megahed, Yasmine, Megahed Yasmine, Cabral Pedro, Silva Joel, and Caetano Mário. 2015. “Land Cover Mapping Analysis and Urban Growth Modelling Using Remote Sensing Techniques in Greater Cairo Region—Egypt.” *ISPRS International Journal of Geo-Information* 4 (3): 1750–69.

Metaxas, Panagiotis, Metaxas Panagiotis, and Mustafaraj Eni. 2013. “The Rise and the Fall of a Citizen Reporter.” In *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci ’13*. doi:10.1145/2464464.2464520.

O’Reilly, Tim. 2009. *What Is Web 2.0*. “O’Reilly Media, Inc.”

Ratcliffe, Jerry H. 2000. “Aoristic Analysis: The Spatial Interpretation of Unspecific Temporal Events.” *International Journal of Geographical Information Science: IJGIS* 14 (7): 669–79.

———. 2010. “The Spatial Dependency of Crime Increase Dispersion.” *Security Journal* 23 (1): 18–36.

Reeves, Joshua. 2013. *If You See Something, Say Something: Surveillance, Communication, and Citizenship in American Life*.

Seeger, Christopher J. 2008. “The Role of Facilitated Volunteered Geographic Information in the Landscape Planning and Site Design Process.” *GeoJournal* 72 (3-4): 199–213.

Shah, Sumit, Shah Sumit, Bao Fenye, Lu Chang-Tien, and Chen Ing-Ray. 2011. “CROWDSAFE.” In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS ’11*. doi:10.1145/2093973.2094064.

Shkabatur, Jennifer, and Shkabatur Jennifer. 2014. “Interactive Community Mapping: Between Empowerment and Effectiveness.” In *Closing the Feedback Loop: Can Technology Bridge the Accountability Gap?*, 71–106.

SIS. 2015. “State Information Service.” http://www.sis.gov.eg/Ar/Templates/Articles/tmpArticles.aspx?ArtID=64485#.VpH0f66rRE5.

Tamayo-Uria, Ibon, Tamayo-Uria Ibon, Mateu Jorge, and Peter J. Diggle. 2014. “Modelling of the Spatio-Temporal Distribution of Rat Sightings in an Urban Environment.” *Spatial Statistics* 9: 192–206.

Taylor, B., A. Kowalyk, and R. Boba. 2007. “The Integration of Crime Analysis Into Law Enforcement Agencies: An Exploratory Study Into the Perceptions of Crime Analysts.” *Police Quarterly* 10 (2): 154–69.

Tukey, John W. 1969. “Analyzing Data: Sanctification or Detective Work?” *The American Psychologist* 24 (2): 83–91.

Turner, Andrew. 2006. *Introduction to Neogeography*. “O’Reilly Media, Inc.”

UNODC. 25 June, 2015. “UNODC Statistics Online.” https://data.unodc.org/?lf=1&lng=en.

8. Appendix

8.1 Dataset

The dataset is available to download from the following link:
https://github.com/abbasadel/zabatak-data

8.2 Exploratory Analysis R Code

```
require("ggplot2")
require("lubridate")
require("rgdal")
library("stpp")
library("rgl")
library("rpanel")
library("maptools")

WORK_DIR = '~/data'
setwd(WORK_DIR)

######################### LOADING DATA ########################

colors <- c("#009E73", "#D55E00",  "#E69F00",   "#999999")
events = read.csv('zabatak-14-12-15.csv', sep = ",")

#format variables
events$date = strptime(events$date, '%Y-%m-%d %H:%M:%S')
events$cat = as.factor(events$cat)
events$subcat = as.factor(events$subcat)

#compute days starting from min(events$incident_date)
events$days = as.numeric(as.Date(events$date) - as.Date(min(events$date))) ;

#select hour of day
```

```
events$hour = as.numeric(format(events$date, "%H")) ;

#REMOVE UNSED CATEGORIES
events = events[!events$cat%in% c("111", "34", "118"),]
events = events[events$date < dmy("01-01-2014"),]
events <- droplevels(events)


levels(events$cat) = list(
  "Building Violations"= "32",
  "Theft" = "101",
  "Violence - Illegal Trade" = "108",
  "Corruption" = "113",
  "Missing" = "111",
  "Founds" = "34",
  "Traffic" = "118");

levels(events$subcat) = list(
  "Car theft" = "22",
  "Missing Persons" = "26",
  "Traffic Violation" = "31",
  "Found Cars" = "35",
  "Other Founds" = "36",
  "Building on Agricultural Land" = "37",
  "New Building Without Permit" = "38",
  "Building Modification Without Permit" = "39",
  "Encroached upon the state lands" = "43",
  "Fraud" = "102",
  "Theft of property" = "104",
  "Arms or drug trade" = "110",
  "Citizens terrorism" = "109",
  "Other Missing" = "112",
  "Bribery, Job Profiteering" = "114",
  "Careless behavior" = "115",
  "Corruption Localities" = "116",
  "Commercial Fraud" = "117",
```

```
  "Road Accidents" = "119");


################################################

events <- droplevels(events)



#draw barplot
#incidents per category

opar <- par(no.readonly=TRUE)
par(mai= c(3.1,1,0.5,1), cex.axis=0.7 )


## CATEGORIES
counts = table(events$cat);
bplt = barplot(counts, las = 2, main="Categories",  col=colors)
text(y=counts+35, x= bplt, labels=as.character(counts), xpd=TRUE)

## SUB - CATEGORIES

counts = table(events$subcat);
bplt = barplot(counts/1995, las = 2, main="Sub-Categories")
text(y=counts/1995, x= bplt, labels=as.character(counts), xpd=TRUE)

par(opar)

#incidents per category parent
opar <- par(no.readonly=TRUE)
par(mai= c(3.1,1,0.5,1) )
barplot(sort(table(events$cat)), las = 2)
par(opar)
```

```
attach(events)

qplot(x=subcat, data = events[order(cat),] , fill = cat, xlab = "Sub
Categories", ylab = "Count of Incidents" ) +
  scale_fill_manual(values=colors, name = "Main Categories") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 12));


#incident per month
qplot(
  month(date, TRUE), data = events, fill = cat, xlab = "Months"
) +
  scale_fill_manual(values=colors, name = "Main Categories")


#incident per year
qplot(year(date), data = events, fill = cat, xlab = "Years", position="dodge")
+
  scale_fill_manual(values=colors, name = "Main Categories") +
  geom_bar(position="dodge")

ggplot(events, aes(year(date), fill=cat)) +
  geom_bar(position="dodge") +
  scale_fill_manual(values=colors, name = "Main Categories")

### table to x,y
dfTab <- as.data.frame(table(year(events$date)))
colnames(dfTab)[1] <- "x"
dfTab$lab <- as.character(100 * dfTab$Freq / sum(dfTab$Freq))

#incident per week
qplot(
  wday(date, TRUE), data = events, fill = cat, xlab = "Week"
) +
  scale_fill_manual(values=colors, name = "Main Categories") ;
```

```
#incident per month
qplot(
  format(date, "%Y-%m"), data = events, fill = cat, xlab = "Month-Year"
) +
  scale_fill_manual(values=colors, name = "Main Categories") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1));

events_cars = events[events$subcat%in% c("Car theft"),]

qplot(
  format(date, "%H"), data = events_cars,  xlab = "Hour", color=colors[2]
) +
  scale_fill_manual(values=colors[2], name = "Main Categories") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1));

plot(table(format(events_cars$date, "%H")),  xlab = "Hour", color=colors[2])


#car theft time
qplot(format(date, "%H"), data = events_cars, xlab = "Car-theft time",
fill="red") +
  theme(legend.position="none")


#######---------------------------------------
WORK_DIR = '~/data/'

CSR_LATLNG = CRS("+proj=longlat +ellps=WGS84 +datum=WGS84")
CSR_UTM = CRS("+proj=utm +zone=35N +datum=WGS84 +units=m")

setwd(WORK_DIR)


library("rgdal")
library("stpp")
```

```
library("rgl")
library("maptools")
require("spatstat")
library("colorspace")

#Transform events to UTM
coordinates(events) <- c("x", "y")
proj4string(events) <- CSR_LATLNG
events = spTransform(events, CRS=CSR_UTM)

## LOAD POLYGONS gcairo
gcairo = readOGR("gcairo/greaterCairo.shp", layer = "greaterCairo")
gcairo = spTransform(gcairo, CRS=CSR_UTM) #Transform to UTM



egypt <- readOGR("egypt/EGY_Governorates.shp", layer = "EGY_Governorates")
egypt <- spTransform(egypt, CRS=CSR_UTM) #Transform to UTM

##PLOT INCIDENTS OVER EGYPT
plot(egypt)
  for (i in length(levels(cat)):1 )  {
    print(levels(cat)[i])
    print(colors[i])
    points(events[cat %in% c(levels(cat)[i]),], pch=20, col=colors[i], cex=0.5)
  }



#count incidents over all governorates
sapply(over(egypt, geometry(events), returnList = TRUE), length)

#Counts if Incident in each top goves
#21 cairo #22 fayoum #27 giza #25 alex 21 Beni Suef
geo = geometry(egypt)
opar <- par(no.readonly=TRUE)
par(mai= c(3.1,1,0.5,1), cex.axis=0.7, oma = c(0, 0, 2, 0) )
for(i in c(21,22,27,25,12)){
```

```
  par(mfrow=c(1,2))
    area = geo[i]
    ok <- !is.na(over(events, as(area,"SpatialPolygons")))
    only = events[ok,]
    plot(area)
    points(only, pch=20, col=colors)
    counts = table(only$cat);
    bplt = barplot(counts, las = 2,   col=colors)
    #text(y=counts+12, x= bplt, labels=as.character(counts), xpd=TRUE)
    title(paste(length(only), "Incidents in", egypt$Gov_En[i] ), outer=TRUE)
}
par(mfrow=c(1,1))
par(opar)


#The evolving pattern of incident (current rat sightings)
#and prevalent (past rat sightings) cases as a discrete-time
#sequence from 2006 to 2008. Incident cases within the time interval between
successive frames are shown in black. Prevalent
#cases on the dates indicated are shown in grey.
ok <- !is.na(over(events, as(gcairo,"SpatialPolygons")))
events = events[ok,]

plot(gcairo, main="01-06/2011",axes = TRUE)
points(events[events$date < dmy("01-06-2011"),], pch=20)

plot(gcairo, main="06-12/2011",axes = TRUE)
points(events[events$date < dmy("01-06-2011"),], pch=20, col="#777777")
points(events[events$date < dmy("01-01-2012") & events$date >
dmy("01-06-2011"),], pch=20)

plot(gcairo, main="01-06/2012",axes = TRUE)
points(events[events$date < dmy("01-01-2012"),], pch=20, col="#777777")
points(events[events$date < dmy("01-06-2012") & events$date >
dmy("01-01-2012"),], pch=20)
```

```
plot(gcairo, main="06-12/2012",axes = TRUE)
points(events[events$date < dmy("01-06-2012"),], pch=20, col="#777777")
points(events[events$date < dmy("01-01-2013") & events$date >
dmy("01-06-2012"),], pch=20)

plot(gcairo, main="01-12/2013",axes = TRUE)
points(events[events$date < dmy("01-01-2013"),], pch=20, col="#777777")
points(events[events$date > dmy("01-01-2013"),], pch=20)


 #### YEARS

plot(gcairo, main="2011",axes = TRUE)
points(events[events$date < dmy("01-01-2012"),], pch=20)

plot(gcairo, main="2012",axes = TRUE)
points(events[events$date < dmy("01-01-2012"),], pch=20, col="#777777")
points(events[events$date < dmy("01-01-2013") & events$date >
dmy("01-01-2012"),], pch=20)

plot(gcairo, main="2013",axes = TRUE)
points(events[events$date < dmy("01-01-2013"),], pch=20, col="#777777")
points(events[events$date < dmy("01-01-2014") & events$date >
dmy("01-01-2013"),], pch=20)

##### CONTINUE HERE



## DEFINE CAIRO CENTRAL AREA
central = cbind(x= c(890000, 890000, 940000, 940000), y=c(3310000, 3350000,
3350000, 3310000))
Sr1 = Polygon(central, hole=as.logical(NA))
central = SpatialPolygons(list(Polygons(list(Sr1), "s1")), 1:1,
proj4string=CSR_UTM)
```

```
## create greater cairo polygon
cairo <- readOGR("cairo/greaterCairo.shp", layer =
"greaterCairo")@polygons[[1]]@Polygons[[1]]@coords

## create Egypt governorate polygon
egypt <- readOGR("egypt/EGY_Governorates.shp", layer = "EGY_Governorates")
egypt_govs.x = egypt@polygons[[1]]@Polygons[[1]]@coords[,1]
egypt_govs.y = egypt@polygons[[1]]@Polygons[[1]]@coords[,2]
#append(egypt_govs[,1], egypt@polygons[[2]]@Polygons[[1]]@coords[,1])

for( i in 2:length(egypt@polygons)){
  egypt_govs.x <- c(egypt_govs.x, egypt@polygons[[i]]@Polygons[[1]]@coords[,1])
  egypt_govs.y <- c(egypt_govs.y, egypt@polygons[[i]]@Polygons[[1]]@coords[,2])
}

egypt = list(egypt_govs.x, egypt_govs.y)
names(egypt) <- c('x','y')


#prepare all incident locations

xyt = events[,c(8,7,1)];
names(xyt) <- c('x','y', 't')
#xyt[,3] = as.Date(xyt[,3])
##convert date to days starting fron minimum date
xyt[,3] = as.Date(events$incident_dateadd) - as.Date(min(events$incident_date))
;
points <- as.3dpoints(xyt);

##ploting events for Egypt
plot(points, s.region = egypt, pch = 20, mark = TRUE, mark.col = 1)
title('Spatial Distribution over Egypt')

egypt.shp = readShapeSpatial("egypt/EGY_Governorates.shp")
```

```
egypt.owin= as(egypt.shp, "owin")
par(new=TRUE)
plot(egypt.owin)
par(new=FALSE)


##ploting events for cairo
plot(points, s.region = cairo, pch = 20, mark = TRUE, mark.col = 1)
title('Spatial Distribution over Greater Cairo')

#get categories
categories = unique(events$cat_parent_id);
categories_titles = unique(events$cat);

for (i in 1:length(categories)) {
  selectedData = events[events$cat_parent_id == categories[i],];
  xyt = selectedData[,c(8,7,1)];
  names(xyt) <- c('x','y', 't')
  xyt[,3] = as.Date(selectedData$incident_dateadd) -
as.Date(min(selectedData$incident_date)) ;
  points <- as.3dpoints(xyt);
  plot(
    points, s.region = cairo, pch = 20, mark = TRUE, mark.col = 1
  )
  title(main = categories_titles[i])
}
str

for (i in 1:length(categories)) {
  selectedData = events[events$cat_parent_id == categories[i],];
  xyt = selectedData[,c(8,7,1)];
  names(xyt) <- c('x','y', 't')
  xyt[,3] = as.Date(selectedData$incident_dateadd) -
as.Date(min(selectedData$incident_date)) ;
  points <- as.3dpoints(xyt);
  plot.stpp2(
```

```
    points, s.region = cairo, pch = 20, mark = FALSE
  )
  title(main = categories_titles[i], outer = TRUE, line = -1 )
}

egypt.shp = readShapeSpatial("egypt/EGY_Governorates.shp")
egypt.owin= as(egypt.shp, "owin")


plot(points, pch = 20, mark = TRUE, mark.col = 1)
plot(egypt.owin)
```

8.3 Spatio-Temporal Analysis R-Code

```
library("rgdal")
library("stpp")
library("rgl")
library("maptools")
require("spatstat")


WORK_DIR = '/Users/abbasadel/Google Drive/Thesis/data/'
CSR_LATLNG = CRS("+proj=longlat +ellps=WGS84 +datum=WGS84")
CSR_UTM = CRS("+proj=utm +zone=35N +datum=WGS84 +units=m")
######################### LOADING DATA #########################

setwd(WORK_DIR)


## DEFINE CAIRO CENTRAL AREA
central = cbind(x= c(890000, 890000, 940000, 940000), y=c(3310000, 3350000,
3350000, 3310000))
Sr1 = Polygon(central, hole=as.logical(NA))
central = SpatialPolygons(list(Polygons(list(Sr1), "s1")), 1:1,
proj4string=CSR_UTM)


## LOAD POLYGONS gcairo
gcairo = readOGR("gcairo/greaterCairo.shp", layer = "greaterCairo")
gcairo = spTransform(gcairo, CRS=CSR_UTM) #Transform to UTM


egypt <- readOGR("egypt/EGY_Governorates.shp", layer = "EGY_Governorates")
egypt <- spTransform(egypt, CRS=CSR_UTM) #Transform to UTM
```

```
## SET STUDY AREA
area = gcairo

######################### PREPARE DATA ##########################

#### EVENTS ####

events = read.csv('zabatak-14-12-15.csv', sep = ",")

#Transform events to UTM
coordinates(events) <- c("x", "y")
proj4string(events) <- CSR_LATLNG
events = spTransform(events, CRS=CSR_UTM)

#format variables
events$date = strptime(events$date, '%Y-%m-%d %H:%M:%S')
events$cat = as.factor(events$cat)
events$subcat = as.factor(events$subcat)

#compute days starting from min(data$incident_date)
events$days = as.numeric(as.Date(events$date) - as.Date(min(events$date))) ;

#calculate hour of day
events$hour = as.numeric(format(events$date, "%H")) ;

######################### FILTER DATA ##########################

# select points inside area
ok <- !is.na(over(events, as(area,"SpatialPolygons")))
events.xyt = events[ok,]

#reverse coordinates anti-clockwise to be combatible with owin
area.poly = area@polygons[[1]]@Polygons[[1]]@coords
area.poly <- area.poly[dim(area.poly)[1]:1,]
```

```
#prepare xyt
#xyt =  cbind(x=events.xyt$x, y=events.xyt$y, t=events.xyt$days)
xyt = as.3dpoints(events.xyt$x/1000, events.xyt$y/1000, events.xyt$days)

#remove dublicated points
xyt = xyt[!duplicated(xyt[,1:2]),]

######################### CONVERT TO KILO-METERS

area.poly = area.poly/1000

#xyt[,1] = xyt[,1]/1000
#xyt[,2] = xyt[,2]/1000

######################### PROCESS DATA #########################

#plot polygon with points using SP package
plot(area)
points(events.xyt, pch=20)


### PLOT MAP
points3d <- as.3dpoints(cbind(xyt[,1], xyt[,2], xyt[,3]))
plot.stpp(points3d, s.region=area.poly,  pch = 20, mark = TRUE, mark.col = 1)
title(main="Crimes in Greater Cairo")


#Calculate kernel bandwidth using bw.diggle
points_ppp <- ppp(x=xyt[,1],y=xyt[,2], window = owin(poly=area.poly))
h = bw.diggle(points_ppp)

#Calculate temporal intensity
Mt <- density(xyt[, 3], n = 1000, bw="nrd")
mut <- Mt$y[findInterval(xyt[, 3], Mt$x)] * dim(xyt)[1]

plot(Mt, main="")
```

```
h <- mse2d(as.points(xyt[,1:2]), poly=area.poly, nsmse = 100, range = 300)
h <- h$h[which.min(h$mse)]


#Kernel smoothing function
Ms <- kernel2d(as.points(xyt[,1:2]), area.poly, h = h, nx = 5000, ny = 5000)
#image(Ms) #image kernel
atx <- findInterval(x = xyt[, 1], vec = Ms$x)
aty <- findInterval(x = xyt[, 2], vec = Ms$y)

mhat <- NULL
for(i in 1:length(atx)) mhat <- c(mhat, Ms$z[atx[i], aty[i]])

#choose U and V values
#gcairo
#u <- seq(2, 5, by = 0.1) #distance in KM
#v <- seq(0, 4, by = 0.1) #days

#centeal
u <- seq(2, 8, by = 1) #distance in KM
v <- seq(0, 8, by = 0.5) #days


# Calculate K function
stik <- STIKhat(xyt = xyt , s.region = area.poly, t.region = c(0, 1800), lambda
= mhat * mut/dim(xyt)[1], dist = u, times = v, infectious = FALSE)

plotK(stik,type="image", L=TRUE, xlab="u", ylab="v") #image k() - pi*r^2*t

plotK(stik, n=30) #plot k()
plotK(stik, n=30, L=TRUE) #plot k() - pi*r^2*t
plotK(stik,type="persp",theta=-45,phi=45) #plot k() in 3D
plotK(stik,type="persp",theta=-45,phi=45, L=TRUE) #plot k() - pi*r^2*t in 3D
plotK(stik,type="image", L=TRUE) #image k() - pi*r^2*t
```

```
# Calculate G function
#greater cairo
#u2 <- seq(0, 20, by = 1) #distance in KM
#v2 <- seq(0, 20, by = 1) #days


#central Cairo
u2 <- seq(0, 10, by = 0.5) #distance in KM
v2 <- seq(0, 10, by = 0.5) #days

g <- plotPCF(xyt = xyt, lambda = mhat * mut/dim(xyt)[1], dist = u2, times = v2,
s.region = area.poly, t.region = c(0, 1800))

plotPCF(g) #plot PCF
plotPCF(g, type = "persp", theta=25,phi=35) #plot PCF in 3D
plotPCF(g, n=30, L=TRUE) #plot k() - pi*r^2*t
plotPCF(g,type="image") #image k() - pi*r^2*t
```