DESIGN AND IMPLEMENTATION OF A GEOSPATIAL DASHBOARD FOR CRIME ANALYSIS AND PREDICTION

By

Siyuan Liu

Bachelor of Geographic Information Science, Shandong University of Science and

Technology, Shandong, China, 2017

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Applied Science

in the program of

Civil Engineering

Toronto, Ontario, Canada, 2020

© Siyuan Liu, 2020

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, as the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Abstract

DESIGN AND IMPLEMENTATION OF A GEOSPATIAL DASHBOARD FOR CRIME ANALSIS AND PREDICTION

Master of Applied Science 2020

Siyuan Liu Civil Engineering Ryerson University

Dashboard has been around for a long time, and many have been developed as a governing and monitoring tool in city management, such as crime monitoring. However, the majority of crime dashboards function as a visualization tool and few of them has been specifically developed for crime analysis and prediction.

This thesis focuses on the development of geospatially-enabled crime dashboards with spatial analysis capabilities for supporting crime analysis and prediction. A prototype has been designed and implemented to support the understanding of crime events for crime reduction efforts. This dashboard will assist policy makers and leaders in crime fighting by visualizing basic statistical information of crimes, revealing their spatial and temporal patterns, identifying crime clusters, and analyzing relationships between crimes and other factors. Based on the criteria developed in this thesis, the prototype confirmed its ability of enhancing the understanding of crime events.

Acknowledgements

This thesis could not have succeeded without the love and support of my dearest family and friends. I would like to sincerely thank my supervisor Professor Songnian Li for his constant support, encouragement and guidance through the years. His knowledge and guidance were essential to the success of this research. I would also like to thank Dr. Wei Huang, Dr. WaiYeung Yan, and PhD student Richard Wen for their helpful advice and recommendations. I would also like to thank others from the Department of Civil Engineering at Ryerson University for their support to this thesis research and acknowledge the support of the City of Toronto Open Data for the crime data.

Author's Declaration	i
Abstract	iii
Acknowledgements	iv
List of Figures	. vii
List of Tables	viii
CHAPTER 1. INTRODUCTION	1
1.1 Background	1
1.2 Research Objectives	2
1.3 Approaches to the Research	3
1.4 Scope of Research	4
1.5 Significance and Contributions of the Research	4
1.6 Organization of the Thesis	5
CHAPTER 2. PREDICTIVE CRIME ANALYTICS AND GEOSPATIAL DASHBOARDS	6
2.1 Predictive Crime Analytics	6
2.1.1 Descriptive Statistics Approaches	6
2.1.2 Data Mining Approaches	7
2.1.3 GIS Approaches	8
2.2 Geospatial Dashboards for Crime Analysis and Prediction	. 10
2.2.1 Concept of Geospatial Dashboards	. 10
2.2.2 Applications of Geospatial Dashboard	. 10
2.2.3 Applications of Geospatial Dashboards for Crime Reduction and Prevention	. 16
2.3 Summary	. 18
CHAPTER 3 EVALUATION OF SELECTED CRIME DASHBOARDS	. 19
3.1 Evaluation Criteria	. 19
3.2 Selection of Suitable Candidates for Evaluations	. 20
3.3 Review and Evaluation of Selected Crime Dashboards	. 22
3.4 Evaluation Results and Discussion	. 23
CHAPTER 4 DESIGN AND IMPLEMENTATION OF A GEOSPATIAL CRIME DASHBOARD	. 28
4.1 Design of the Prototype	. 28
4.1.1 Key Design Considerations	. 28

Table of Contents

4.1.1.1 Intended Use 28
4.1.1.2 Users and User Requirements
4.1.2 Enabling Technologies
4.1.2.1 Tools and Libraries used for Building Dashboards
4.1.2.2 Third Party Libraries for Dashboard Front-end Development
4.1.3 System Architecture
4.2 Implementation of the Prototype
4.2.1 User Interface Development
4.2.2 Backend Development and Data Preprocess
4.2.3 Function Development
4.3 Summary
CHAPTER 5 DEMONSTRATION OF THE PROTOTYPE
5.1 Toronto Crime Data
5.2 Data Analysis Using the Prototype
5.2.1 Query and Visualization of Data
5.2.2 Basic Statistical and Spatial Analysis of Data
5.2.3 Predictive Crime Analysis
5.3 Summary
CHAPTER 6 CONCLUSION 69
REFERENCES

List of Figures

Figure 1 Overall process of the study
Figure 2 Functional requirements of geospatial dashboards for crimes
Figure 3 System architecture
Figure 4 Implementation architecture
Figure 5 P value and Z score table
Figure 6 K-distance graph
Figure 7 Max-min distance method
Figure 8 Max-min number method
Figure 9 Assault crime overview
Figure 10 Toronto police command division boundaries
Figure 11 Total crime distribution
Figure 12 Alert box to show analysis results
Figure 13 Assault crime mean centers
Figure 14 Assault crime standard division ellipse
Figure 15 Hotspot results
Figure 16 Assault hotspot and homeless shelters
Figure 17 Assault hotspot and police facilities

List of Tables

Table 1 Evaluation criteria	. 19
Table 2 Selected crime dashboard systems	. 21
Table 3 Evaluation results	. 23
Table 4 File system and database comparison	. 44
Table 5 Toronto crime data basic statistics	. 56
Table 6 Correlations between different crimes	. 62

CHAPTER 1. INTRODUCTION

1.1 Background

Toronto is one of the safest cities in Canada and North America. However, as shown by the statistics of crime events from Toronto Police, crime events have seen an upward trend in 2018 and 2019. The crime rate of Toronto has been increasing. Compared to last year in 2018, the assault crime rate has increased by 4%, the break and enter rate has increased by 14%, and the theft crime rate has increased by 12% (Toronto Police, 2019). To better understand crimes, reduce and prevent crimes more efficiently, various crime monitoring, and analytics tools have been developed and crime dashboard is one of them.

Dashboards have a wide usage in the government and management due to its intuitive data presentation and the assistance in decision-making process it provides. It is utilized frequently in understanding various crime scenarios. Nevertheless, most of the existing crime dashboard systems mainly focus on the visualization of crime data, which is insufficient for understanding crime events and obtaining valuable information, because the valuable information contained within crime data often needs to be analyzed and processed instead of simple visual inspection. Geospatial dashboard for crimes is a solution that combines dashboard visualization functions and crime analysis and prediction components with special considerations on spatial and temporal dimensions. It is necessary to include geospatial characteristics since crime events have inherent geographical nature. When a crime occurs, it happens at a place with a geographical location. Crime events could be explained in more depth by exploring their geographical location and distributions (Chainey & Ratcliffe, 2013).

The main limitation of existing crime dashboards is the lack of analysis components, specially the spatial analysis. Most of the current crime dashboard systems function as a data visualization and querying tool. The information is presented in the form of tables, graphs and/or indictors such as crime rate, amount of crime, and crime types. Compared to the dashboards that only provide text-based and graphical statistics information, the dashboard system that introduces the map element and the geospatial conception is a great improvement. However, it is still insufficient, the lack of analysis and interpretation of crime data from different perspectives and dimensions imped the deep understanding of crime events.

The main content of this research is to study and evaluate the usability and design of geospatial dashboards for crimes, develop a crime geospatial dashboard, and demonstrate the utilities of the prototype implementation with a case study.

1.2 Research Objectives

The goal of this study is to develop a geospatially-enabled dashboard for supporting crime visualization, analysis and prediction, and to study related design and implementation issues. This has been achieved by the following objectives:

- 1) design an architecture for crime analysis dashboards,
- 2) develop the prototype based on the architecture, and
- 3) study the usability of this prototype using a case study in Toronto.

To fulfill the objectives of developing a dashboard for crimes, several factors need to be considered. System-wise, the dashboard should include a basic visualization component and the crime data should be presented statistically from both spatial and temporal perspectives. It should support the visualization of crime statistics based on various time scales by hourly, daily, and monthly increments. From the spatial perspective, the data should be categorized by administrative units, and filtered by user's area of interest. There is also a comparison component between different datasets to help understand crimes in a better way. In addition to the basic visualization, the analysis component is another essential goal. To explore the underlying information and value, the objectives are further fulfilled by including both statistical and geo-statistical techniques such as mean center, standard deviation, hotspot analysis and causal relationship analysis.

1.3 Approaches to the Research

This study was unfolded in following approaches. First, the literature review of dashboard usage and development was done. It reviewed the composition of some representative geospatial dashboards and how they have been used in various scenarios to help with the decision-making process. Second, these existing crime dashboards were evaluated. Through the evaluations, their design patterns, common features, and current shortcomings were obtained as a reference of the development work in this study. Third, a review of dashboard development techniques was completed. This gives an overview of approaches to implement a geospatial dashboard, and these approaches were discussed and compared to deliver better options for dashboard development. A crime geospatial dashboard prototype was designed and developed based on the knowledge and experience obtained from the reviews and evaluation. Finally, the use of the developed prototype was demonstrated with sample data of a case study. Figure 1 shows the process of the study.



Figure 1 Overall process of the study

1.4 Scope of Research

This thesis research focused on design and development of geospatial dashboards in the context of managing crime data. The dashboard design was based on the results of requirements analysis through the literature review and the evaluation of some selected crime dashboards. Crime dashboard evaluation was subject to the available information of the systems selected. A research prototype was developed based on the design and the crime data from City of Toronto, available from Toronto Police Service Public Safety Data Portal-Open Data, was used to demonstrate the use of the prototype.

1.5 Significance and Contributions of the Research

In this study, selected crime-related geospatial dashboards in the context of general dashboards were reviewed. The review explored the trends and patterns of geospatial dashboard design and development and provides an overview and introduction of the geospatial dashboard development status. The review of geospatial dashboard development tools and techniques explored the capabilities and potentials of opensource resource on dashboard development.

A crime geospatial dashboard architecture was proposed in this study. This architecture provided guidance of geospatial dashboard development for crime scenarios, which made it easier to understand the whole system and the communication between all stakeholders. In addition, it enabled the reuse of requirements and experience. It defined the problems and considerations when it comes to implementation.

The crime geospatial dashboard prototype was implemented in this study. This prototype verified and testified the geospatial dashboard design and development patterns and principles. It was an implementation of the architecture. This prototype function as a visualization, analysis, and prediction tool to help with preventing and reducing crime events.

1.6 Organization of the Thesis

This thesis is organized into six chapters, starting with this introduction chapter, which presents the background of this research, research objectives, research approach, scope of research, and contributions of the research. Chapter 2 discusses the predictive crime analytics methods and introduces the geospatial dashboards. It provides the review of general geospatial dashboards as well as the usage of crime dashboards. Chapter 3 develops evaluation criteria for selecting crime dashboard and evaluates existing crime dashboards. Crime dashboard design patterns and principles are obtained. Chapter 4 presents the design and implementation of a crime geospatial dashboard, including its key considerations, user requirements, enabling technology, system architecture, and function development. Chapter 5 uses the City of Toronto as a study area and its crime data to walk through the usage of the prototype functions. Chapter 6 concludes this thesis by summarizing the work completed in this research and outlining opportunities for future research.

CHAPTER 2. PREDICTIVE CRIME ANALYTICS AND GEOSPATIAL DASHBOARDS 2.1 Predictive Crime Analytics

Predictive analytics is a set of business intelligence (BI) technologies that uncovers relationships and patterns within large volumes of data that can be used to predict behavior and events, it is forward-looking, using past events to anticipate the future (Eckerson, 2007). The predictive analytics is an essential component in predictive policing for predicting crimes (Rienks, 2015). With the advances of computer technology, the data is no longer only a record of crimes but contains useful information to reveal the patterns and underlying trends of crimes (Chen, et al., 2003). The predictive analytical methods for crime prediction fall into mainly three categories: descriptive statistics approaches, data mining approaches, and GIS approaches (Grover, Adderley & Brammer, 2006).

2.1.1 Descriptive Statistics Approaches

The descriptive statistics approach is a straightforward way to identify crime patterns. Statistics methods can be implemented from various perspectives, such as temporal, spatial, and correlational. For example, a study was done based on the relation between distance and crime events (Farrington & Lambet, 2000). The results showed that 69% of burglary and 55% of violence offenders lived within one mile of the scene of the crime. Only 8% of burglary and 15% of violence offenders lived more than five miles away from crime scenes. However, more advanced and complicated methods could be implemented to better understanding crimes, in addition to basic statistic measures such as mean centers for locating distribution centers and standard division ellipse for determine distribution directions. For example, statistic methods such as the Moran's I

analysis were used to detect the spatial autocorrelation of crime events to identify if it is distributed randomly and some other relation analyses can help determine if elements influence each other.

2.1.2 Data Mining Approaches

Data mining techniques empowers capacity of the crime analysis and prediction. They integrate statistics, artificial intelligence, criminology, and database technology. There are a couple of popular crime prediction methods based on data mining, such as Support Vector Machine (SVM), Fuzzy theory, and Artificial Neural Network (ANN). SVM is a nonlinear model and it performs well in time series prediction, classification tasks and regression (Wang, Matheiu, Ke & Cai, 2010). It predicts time series more stable and efficient with the ability of modeling nonlinear relations (Zhu & Zhang, 2012). Furthermore, SVM was also applied as a crime hotspot prediction method (Kianmehr & Alhajj, 2006). The methods based on fuzzy theory were also applied for forecasting crime (Shrivastav & Ekata, 2012) due to their high prediction efficiency and less computational efforts (Chang & Chen, 2009). ANN solve problems with a large number of components and mimic the operation of brain in decision making (Kumar, 2015). ANN can help analyze the historical data and wins over the traditional statistical methods by estimating the relationship between past values and future values without the limitation of the complexity of real systems (Xia & Rao, 2009). For example, it was used to predict the geospatial and temporal variations of crimes and disorders (Chitsazan, Rahmani & Neyamadpour, 2013).

Further, artificial intelligence was used for identifying burglary crimes, for example (Lucas, 1986). Moreover, understanding of crime behaviors was obtained with string comparator, social network analysis and deviation detection (Chen, et al., 2004) and detection of abnormal

activities was realized by concept space algorithm (Hauck, Atabakhsb, Ongvasith, Gupta & Chen, 2002).

2.1.3 GIS Approaches

GIS has been widely applied in crime analysis and prediction. Various predictive methods have been developed to meet different purposes and situations. These methods include hotspot analysis, generalized linear model and additive model, repeat victimization, univariate methods, leading indicators and point process model.

A commonly used method is hotspots analysis. The crime hotspots could be labeled as crime generators, such as entertainment areas and shopping malls (Brantingham & Brantingham, 1995). Clustering methods have been used widely to obtain hotspots (Grubesic, 2006). These methods include density-based method, hierarchies that place objects in groups based on the characteristics they possess, and partitioning methods that split observations into a pre-specified number of groups, where the specified criterion was optimized globally over all possible splits (Grubesic & Murray, 2001). However, hotspot analysis is not the only one of crime predictive models, and has its limit; therefore, there are other models discussed below.

The hotspot method is an accurate predictor, but it is less effective due to its reliance on shorter time periods. As such, the conception of repeat victimization was brought up. The repeat victimization are those individuals or places that have been victimized once and are likely to be victimized again, and the time course to subsequent victimization is a few short months (Anderson, Chenery & Pease, 1995). Geo–referencing and searching crime locations using GIS allows for identifying repeats in a more efficient and accurate way (Ratcliffe & McCullagh, 1998).

The hotspot analysis for crime prediction has its limitations, of which the main one is that it only utilizes criminal incident data, such as the type of crime, locations, and time of crime events. It only shows the current patterns of crimes without the insight into the relationship between crimes and environment over time (Wang & Brown, 2012). Therefore, more advanced methods are applied to determine the relationships between crimes and environmental factors. It contains generalized linear models (GLM) that have better prediction performance than density models (Brown, Dalton & Hoyle, 2004), and nonlinear models that combined point process models with the Generalized Additive Model (GAM) to study the effects of installed security cameras on crimes (Rodrigues, Diggle & Assuncao, 2010). However, the spatial-temporal generalized additive models (ST-GAMs) brought by Wang and Brown (2012) discovered the underlying factors related to crimes to predict future crimes. There was a local spatial-temporal generalized additive model as well to better predict crimes in specific regions. It is verified by experiment that those two models outperformed the spatial GLM and hotspot models.

Univariate methods use previous values of one variable to predict its future value. There are several univariate methods available to predict crime range from simple random walk to sophisticated one that incorporates both seasonally and time trends. Sophisticated methods predict more accurately by including seasonality in the model, accounting for time trends using exponential smoothing and pooling data (Groff & La Vigne, 2002).

Leading indicators is a multivariate method that predicts the future value of dependent crime variables by current and past values of independent variables. The leading indicators refers to the specific characteristics of areas or neighboring areas for which their rise or fall in current and previous months can be used to predict future values of dependent crime variables. The point process model is based on the theory of point patterns and multivariate density estimation. Training is involved in this model, and past data are used to predict future events. It combines multivariate models together and uses notions from kriging and density estimation.

2.2 Geospatial Dashboards for Crime Analysis and Prediction

2.2.1 Concept of Geospatial Dashboards

A digital dashboard is a tool that manages information, and tracks, analyzes and displays the key performance indicators related to a subject matter that is monitored and assessed (Few, 2006). The function and design of dashboard systems depend on the needs. Existing dashboard systems mainly display and visualize data in the form of tables, charts, and gauges. Behind the scene, it connects files, attachments, and application programming interfaces (API). It centralizes on analysis and monitoring of business performance.

A geospatial dashboard can be considered as an extension of a digital dashboard, with some distinctive features. The key feature is the integration of geospatial data and spatial analysis with business intelligence (BI). Compared with a non-spatial dashboard system, a geospatial dashboard provides map interfaces, deals with geospatial data, and runs geospatial analysis.

2.2.2 Applications of Geospatial Dashboard

Geospatial dashboards have a wide usage in many areas such as city governing, disaster management, and resource allocation (Kazak, et al., 2017). In the following sections, the usage and development of dashboard in disaster management and city governing were discussed as examples to provide an overview of geospatial dashboard development.

A geospatial dashboard was created to function as a prediction in flood decision support system (Saha, Shekhar, Sadhukhan & Das, 2017). The map-based geographical visualization of this dashboard provided an integrated approach to extract, integrate, and view diverse information. It identified and visualized risk zones before disaster happens. This system enabled the efficient interpretation of bulk data from diverse sources, which allows a faster decision-making process. Geospatial visualization also allowed for a more interactive map interface. However, this dashboard could be improved by having real-time elements, and the visualization could be enhanced with the help of D3.js since JavaScript forms the basis of any visualization techniques.

Except for prediction, a geospatial dashboard could also perform as real-time monitoring system. This was represented by web GIS-based dashboards that are used to monitor marine water quality (Amiruddin, 2016). The dashboard functioned as a warning system that monitors the long-term status, trends of water quality, and determined the influence factors toward the water with a sensor called Wave Gilder. The real-time monitoring function was implemented by ArcGIS Geo Event Server, which enabled real-time event-based data streams to be integrated as a data source. The data could be filtered, processed, and sent to multiple destinations. Geo Event Server is capable of consuming event data from multiple real-time streams. The filters and processors of Geo Event Server allow analysts to discover and focus on the most interesting events, location, and thresholds for their operation. The event data was archived in the way of feature services, tables, and spatiotemporal big data store. Viewing and interacting with maps were fulfilled by ArcGIS viewer for Flex, which allowed users to customize their own product without any programming knowledge. The cooperation of Wave Gilder and Geo Event server enabled a dynamic real-time monitoring dashboard of marine water.

While dashboards like Wave Glider mainly used sensor data to explore the marine water conditions, the social media data was also a great data source in geospatial dashboards for disaster management. The dashboards that visualized social media sentiments in disaster scenarios was an example (Lu et al., 2015). Ebola twitter data was used in this study. The goal was to explore the underlying trends in positive and negative sentiments of Twitter users with respect to disasters and geographically related sentiments using geospatial-located Twitter data. This system was composed of two parts: sentiment modeling and geographic visualization. It provided an entropy-based metric to model sentiment contained in social media data with the help of sentiment analysis classifiers: CoreNLP, SentiStrength and SentiWordNet. The dashboard visualization component contained three parts: the geospatial comparison maps that observe the distribution of Tweets and area selection to show the most popular Tweets, and the entropy sentiment river designed to reveal the uncertainty of sentiment classification over time.

The usage of social media data analysis with geospatial dashboards was also represented in the dashboard system that identified disaster-affected areas using exploratory visual analysis of georeferenced tweets (Cerutti & Fuchs, 2016). This dashboard identified affected areas using geospatial footprints. It defined the geospatial footprint of a flood event that occurred in Italy in 2013. Data mining and exploratory visual analysis were used under time pressure to assist the rapid assessment of situation.

Geospatial dashboard not only played an important role in disaster scenarios, but also influenced the city governing process. The Smart Dashboard system was a geo-targeted search and analytic tool. It allowed researchers to search Twitter data based in different cities, filtered the data noise, analyzed social media from a spatiotemporal perspective, and visualized it in various way such as trends, top retweets, top URLs, and hashtags. It strongly emphasized on the geospatial characteristics of social media data. It filled the gap of social media analytics systems by enabling the target topics in specific geographical regions and analyzing the spatiotemporal patterns of social media message. The front-end of this dashboard displayed dynamic temporal (daily, weekly, and monthly) trends of social media data with interactive selection tools. The backend dealt with data and removed noise with data filtering and machine learning components. Multiple data mining programs, GIS methods, and advanced geo-targeted social media APIs were implemented to build up this system. Spatial analysis such as hotspot analysis and overlay with layers were also used. All these features and patterns were used for assisting researchers in better understanding and investigating various topics of cities (Tsou et al., 2015).

Another example was the real-time analytical Dublin urban dashboard (McArdle & Kitchin, 2016). This dashboard was an interactive web dashboard that collects, analyses, and visualize data from a variety of sources about Dublin, Ireland through a series of interactive maps, graphs and applications. It consisted of 12 modules. For examples, Dublin Mapped showed the results of two most recent Irish Censuses, Dublin Real-time showed the locations of main road segments, multi-stories car parks, and traffic cameras, and Dublin Bay dashboard provided visualizations about Dublin Bay. This dashboard was built up using a Model-View-Controller (MVC) architectural pattern. It used framework CakePHP and was supported by MySQL database. JavaScript and HighCharts were used for user interface design, and the base map was supported by Leaflet.

The geospatial dashboard could be in real-time in the city governing such as the City Eye dashboard system, which was a platform that brings together operations, sensors, and citizen

feedback data together through a web-based dashboard and mobile application (Lee et al, 2015). It mainly had two parts, the dashboard and lens. This dashboard enabled general users to overview some operations of the whole city in real-time. The lens was a smart phone app that allows citizens to view data and communicate their needs or satisfaction with services back to the city. Moreover, it allowed city managers to explore some underlying crosscutting problems by visualizing data from different sources. The data that were used in dashboard usually comes from three parts: environmental sensors, the provision of services, and citizens. The real-time view of the state of a city, such as weather conditions, air pollution, parking usage, and traffic condition, were provided by environment sensors. The information from the service providers helped predict the cleanliness of a certain part of the city and the timing of congestion and noise. Then the citizens' opinions told the manager what kind of short-term and long-term service insufficiencies they were facing.

The usefulness of geospatial dashboards is significantly improved when it combines with GIS analysis. The Smart Infrastructure Dashboard (SID) was a dashboard system that combines traditional BI (Business Intelligence) and GIS. It showed the present and past status of infrastructure condition as well as the future prediction with the 'what if scenario' (Wickramasuriya et al., 2014). SID was composed of an open-source ETL (Extract, Transform and Load) that handled various datasets, and a spatially-enabled data warehouse hosted in PostgreSQL/PostGIS. The data source came from private operators in SID project, including geometric datasets of utility networks, service usage or consumption at various geographic levels over time, water discharge at reservoirs and pumps, water quality at various points in the network, and waste collection. This dashboard allowed for spatial-temporal analysis, visualization of multiple datasets, uncovering of hiding patterns or problems, and identification of relationships between the use of infrastructure services, demographics, and weather parameters on a spatial and

temporal scale. The 'what-if scenario' was one of the most desirable components in this system. It enabled users to leverage on historical data and estimated the potential impact of changes of other variables to make better planning for the future decisions. The visualization of this dashboard was composed of three types of interactive reports: the map reports, radar chart report, and bubble chart report.

From these geospatial dashboards, some common features and patterns of geospatial dashboard design could be learned. First, a geospatial dashboard should have the capability of accommodating multiple data sources. Nowadays, the data has been generated from diverse sources in various formats. It is therefore desirable for a dashboard system to be compatible with multiple data sources to have a better interpretation of data. Second, a geospatial dashboard should work in a real-time manner. Real-time is a trend, and it provides a new perspective of viewing data. It is also necessary for emergency scenarios. Third, social media data has an increasing trend of usage in dashboard systems, including geospatial dashboards. Such data help interpret events from a new perspective. Finally, GIS analysis has become an essential component. Most of the data and information need GIS analysis to discover the value within it. With the support of GIS, geospatial dashboard systems can interpret data from both spatial and temporal perspectives. The historical data can be used most often to obtain patterns and rules, the real-time data represents the ongoing status, and the prediction allows for estimating the future and planning. The data itself could be interpreted with various temporal scales such as daily, weekly, and monthly. The data used in these dashboards are usually preprocessed by ETL (Extract, Transform and Load) and then delivered to a data warehouse.

2.2.3 Applications of Geospatial Dashboards for Crime Reduction and Prevention

Geospatial dashboards play an important role in various crime scenarios (Edwards & Thomas, 2005). To better design a crime dashboard system, it is important to understand how geospatial dashboards are used in crime scenarios for different purposes.

Geospatial dashboards help integrate crime data with other types of data from different sources. A crime dashboard can be useful in integrating crime data and crime-related data. Over the last decade, the growth in the amount and variety of digital data being created, documented and shared, has equally resulted in a growth of the use of dashboards in organizations. The big data revolution has created new possibilities to understand trends, spot patterns, and collect intelligence. The data source that a dashboard handles is also become more diverse. Typically, dashboards integrate data from different sources and exhibit them through informative graphics with explanatory captions (Ganapati, 2011). Social media data could be an essential data source in crime pattern analysis (Williams, Burnap & Solan, 2017). Useful information could be obtained in a dashboard could function as more than just data visualization and analysis, it could also work as a monitoring tool. The capability of processing real time data is important. The real time monitoring function provides updated information to user and let user have control.

Geospatial dashboards allow for filtering crime data for simple view, visualization and analysis. Data filtering is also an essential component in crime dashboards. Filters allows users to specify one or more search conditions to limit the features from sources and can be used to power the data visualization of the dashboard. The filters that are often used in a crime dashboard system is spatial filter, temporal filter, and type filter. The temporal filter that usually implemented in crime dashboard systems allows user to filter crime data by time, such as hourly, daily or weekly, or by setting a time period through dragging time axis. The purpose of temporal filtering is to provide users with a better understanding of crime events from temporal perspective. Filtering crime data from different time chunk or scales depicts the ongoing information of crime events as time goes by, and reveals crime information like trending and comparison, which is helpful for crime prediction and prevention. The spatial filtering that are often used in crime dashboards are in the form of selecting specific areas or drawing polygons on the map interface. This filtering allows users to check the data that they are interested in from a spatial perspective. It enables users to inspect crime data from both micro and macro perspectives. Further, the filters could be additive, which means multiple filters could be used together to apply on the data. The data filtering enables users to inspect data based on their interests.

Geospatial dashboards are used to visualize crime data and insights obtained from analyzing and mining crime data. It provides simple and powerful analyses of crime data. The crime events that are used in dashboard systems is mainly categorized into three types, the crime rate, real time crime monitoring and specific crime type. The crime rate gives an overview of crime amount data, such as the crime rate distribution in different states, the transformation of crime rates over years, and the ranking of crime rates among various areas. The 'Crime in United States 2012 edition' by Wee Wee Sim (Sim, 2012) was a good example of crime rate dashboard. It provided statistical compilation of violent crime like murder, non-negligent manslaughter, forcible rape, robbery, and aggressive assault and property crime such as burglary, larceny-theft, vehicle theft, and arson. Thorough crime rate data analysis, useful information like crime trends, crime statistics could be obtained. The real time crime data are mainly used in dashboard systems for monitoring. The real time data source could be urban sensors, cameras, and police reports. The VPD GeoDASH of Vancouver and Philadelphia real time dashboard (Vancouver Police Department, 2019) both grabbed real time crime data to monitor the ongoing of city crimes. In these dashboards, analysis was unfolded from both temporal and spatial perspectives and usually implemented by clustering methods. Crime dashboards, as an analytical dashboard, should be able to mine insights from crime data through statistical, spatial and temporal analyses. The descriptive statistics is the most basic approach to reveal crime patterns. The descriptive statistics are usually implemented by charts, graphs, diagrams, and maps. Through the visualization of the data, basic statistical information could be obtained such as the maximum, minimum, mean and so on. Besides that, crime dashboards can have the support from GIS, which enhances their analysis function. The most common GIS method that are used in crime dashboard systems is heat map. The heat map shows locations of higher density of geographic entities. However, the hotspot analysis is considered as a more reliable approach since it has statistical significance. The spatial temporal hotspot analysis was used in HotSketch to help with police rapid route planning (Godwin & Stasko, 2017).

2.3 Summary

This chapter began with a discussion and review of predictive crime analytics, which includes approaches like descriptive statistics, data mining, and GIS. Hotspot analysis is the most common and popular method of GIS methods. However, hotspot analysis has its own limits and there are other methods which overcome the limits of this analysis. This chapter went on to discuss the normal geospatial dashboard usage and development. The discussion summarized geospatial dashboard design patterns and trends. Subsequently, this chapter discussed the geospatial dashboard applications for crime reduction and prevention. The crime dashboard has wide usage in integrating crime data, filtering crime data, and mining insights from crime data.

CHAPTER 3 EVALUATION OF SELECTED CRIME DASHBOARDS

This chapter developed the evaluation criteria for assessing geospatial dashboards for crimes and evaluated some selected crime dashboards against the defined criteria. The evaluation results were presented and discussed for the purpose of later prototype development.

3.1 Evaluation Criteria

A dashboard is a visualization tool which provides awareness, trending, and both planning and actual comparisons, frequently visualized in a slick simplified user interface. The dashboard should have goals set by users and consistently meet their expectations and end user experience is one of the major features of dashboard software. To better understand the design of crime geospatial dashboards and their development as well as identify the limits of existing applications, it is necessary to evaluate existing crime geospatial dashboards. A set of key criteria for evaluating dashboards were proposed for all type of dashboards (Karami, Langarizaden & Fatehi, 2017). These criteria were classified into 7 main categories include user customization, knowledge discovery, security, information delivery, alerting, visual design, and integration and system connectivity. The evaluation criteria for crime dashboards were developed based on these criteria, considering some special requirements for being a crime dashboard. These criteria were listed and briefly described in Table 1.

Table	1	Eval	luation	crit	eria

Criteria	Brief descriptions	Criteria
category		
User	Dashboards empower the users to	(C1). Restricted access to specific
customization	response rapid changes in	metrics to a group of users
	environment and turn their	
	departments to agile department.	

System	Dashboard systems must be	(C2). Connectivity to variety of data
connectivity	coordinated with the systems	sources like OLAP cubes, Lists and
and	infrastructure in organization.	Spreadsheets and support different
integration		operating systems.
_		(C3). The data must have geospatial
		attributes.
		(C4). Capable of dealing with real
		time data.
Knowledge	Dashboards should allow users to	(C5). Drill down features, it enables
discovery	conduct analysis and discover	the user to go from general view to
	cause.	detailed view by clicking of mouse.
		(C6). Data visualization from
		multiple perspectives and scales.
		(C7). Predictive analysis.
		(C8). Advanced analysis from spatial
		and temporal perspectives.
Information	Dashboards are not a reporting tool,	(C9). Data filtering
delivery	but the main use is reporting.	
Visual design	Dashboards should be visually	(C10). Having metal data and help. It
	appealing and engaging without	informs users the background
	overwhelming the user but make	information about this dashboard and
	them feel comfortable about design.	the instruction for its use.
		(C11). It must have mapping
		interface and charting elements.
		(C12). Interactive visualization.
		(C13). Auto refresh/re-query of
		dashboard objects.
Alerting	Dashboards should have a	(C14). Visual intelligence to
	mechanism to turn the focus to the	highlight areas and values.
	expectations, outliers and data	
	highlights	
Security	It is important to protect dashboards	(C15). Protecting data from change.
	by tracking and monitoring any	
	changes on the source of dashboard	
	system.	

3.2 Selection of Suitable Candidates for Evaluations

Table 1 above listed 9 crime dashboards selected for the evaluation. An application must at last meet three selection criteria to be qualified as a suitable candidate. First, it must be an online

application. Second, it must be operational and available for testing when the evaluation was conducted. Third, the dashboard system should focus on crimes, ideally with mapping interface and analysis components. The detailed information of the selected crime dashboards such as name, brief descriptions, developer, and access URL were provided in Table 2.

Dashboard name	Brief descriptions	Developer	Access URL
2012 US crime analysis dashboard (D1)	This dashboard system presents the crime rate information of US in 2012. It functions as a crime rate querying and filtering tool by selecting different crime type and location. It also provides trending and ranking components.	Wee Wee Sim	https://public.tabl eau.com/profile/w ee3190#!/vizhom e/USCrimeAnaly sisDashboard120 814- 1/MainDashboard
Philophobia police Dashboard (D2)	This dashboard is a real time crime querying system. It presents real time from last hour, 24 hours, 3 days, 7 days. The crime event is reflected on the map interface by real time. Trends and comparison view are provided as well.	Operations Dashboard Team (Esri)	https://www.arcgi s.com/apps/opsda shboard/index.ht ml#/9ef296f66f72 4c36bcaf01fc697 68ecd
Metropolitan police dashboard (D3)	This dashboard focuses on the statistical visualization of crime data in London area. It visualizes crime data by type and time. Data filtering is implemented by selecting location and setting time period. Attribute table is available in this dashboard.	METROP OLITAN POLICE	https://www.met. police.uk/sd/stats- and- data/met/crime- data-dashboard/
Toronto Police Service- Crime App (D4)	This crime app is composed of a crime map and analytical charts. The crime map visualizes crime data by crime type. The crime data can be quired by police division, date range and spatial filters.	Toronto police service	http://data.toronto police.on.ca/page s/major-crime- indicators
VPD GeoDash Vancouver (D5)	This dashboard system works on the crime data visualization and searching towards Vancouver area. It displays crime on the map and the traffic camera data on the map is available.	Vancouver Police department	https://geodash.vp d.ca/Html5Viewe r/?disclaimer=on &viewer=VPDPu blicRefresh_gvh &x=33&y=38
Toronto crime analysis Cartio Vista	This dashboard visualizes and filters crime data in Toronto. Data table of chart is	CartoVista	https://cartovista. com/examples/cri

(D6)	available. Heat map is implemented to describe crime density.		<u>me-analysis-</u> dashboard
Hot Sketch	Hot sketch is designed to allow police	Alex	NaN
(D7)	officers to more rapidly utilize predictive	Godwin,	
	models for crime in their neighborhood. It	John	
	is composed of interactive spatiotemporal	Stasko	
	host spot analysis and rapid routing		
	planning analysis.		
Security force	This dashboard system provides	Miguel de	NaN
data	meaningful insights regarding security and	Castro	
visualization	in the location-allocation of security forces	Neto, et.al.	
(D8)	in the communication with the		
	SIM4SECURITY results.		
Los Angeles	This crime dashboard visualizes crime data	Microsoft	https://communit
Crime	in Los Angeles from 2010 to 2015 by age	Power BI	y.powerbi.com/t5
Analysis	group, most dangerous weapon, and Top		/Data-Stories-
(D9)	crimes. It has detailed and comprehensive		Gallery/Los-
	visualization of crime data from various		Angeles-Crime-
	perspectives.		Analysis/td-
			<u>p/403118</u>

3.3 Review and Evaluation of Selected Crime Dashboards

This section reviewed and evaluated these selected crime dashboards based on the evaluation criteria developed in the previous section. The results were summarized in Table 3. A cross (\times) in the table denoted that such feature or functions are offered whereas an empty space represented the opposite. The name of the system was represented by index as noted in the previous section.

Systems/	D1	D2	D3	D4	D5	D6	D7	D8	D9
Criteria									
C1			×				×		×
C2		×		×	×			×	
C3	×	×	×	×	×	×	×	×	×
C4		×					×		
C5	×	×	×	×	×	×	×	×	×
C6	×	×	×	×	×	×	×	×	×
C7							×	×	
C8						×	×	×	
C9	×	×	×	×	×	×	×	×	×
C10	×			×			×		×
C11	×	×	×	×	×	×	×	×	×
C12	×	×	×			×	×		×
C13	×	×	×	×	×	×	×		×
C14	×	×	×	×	×	×	×		×
C15	×	×	×	×	×	×	×	×	×

Table 3 Evaluation results

3.4 Evaluation Results and Discussion

The evaluation result was straightforward, and all the dashboard systems were evaluated based on the criteria. None of them could met all the criteria and fulfilled the objective that having a better understanding of crime events. The discussion of the evaluation results was unfolded with the discussion of most common features that appeared in dashboard systems, least common features that appeared in dashboard systems, and what need to be further improved from existing dashboard systems.

One of the most critical and fundamental elements of being a geospatial dashboard is the base map, which is also a unique feature that distinguishes geospatial dashboards from other business dashboards. The base map plays a central role in many dashboards (ESRI, 2019), and it is the most effective and intuitive way to display geographic and geographically referenced information. All the evaluated crime dashboard systems contained a base map. The base map is

mainly used to display crime events and present a spatial overview of crime events. However, the base map might function slightly differently depends on the content and purpose. Some of these dashboards visualizes specific crime events and present the overview of crime events on the base map. For example, the Toronto crime analysis Carto Vista (D6) displayed all the crime data of Toronto include Theft, assault, homicide, drug, and robbery on the map. Users are allowed to check each event on the map by clicking on the map. Except displaying specific crime data, the base map was also used to visualize crime rate, such as the 2012 US Crime Analysis Dashboard (D1) used different color scale and symbol to indicate the crime rate and crime number in different states of US. The base map also contained other information like street camera that allows user to real -time monitor crime situation in VPD GeoDASH system of Vancouver. A qualified map interface should include these elements: the pop-ups to display the attributes of selected crime events such as crime type, occurrence timing, and related information. It should also include a map legend to identify which of the elements are being shown on the map; the base map switcher that allows users to change the base map during run time, and zoom in and out function that allows users to view the map and elements in various scale. In addition, other useful elements would be the default extent function that returns the map to the initial extent while browsing, and a search function that allows the user to find a particular feature on the map. The map should not only display data, but also take the responsibility of interacting with other elements of the dashboard. For example, dynamically presenting the analyzed data and triggering the other visualization elements by map actions.

Data filtering function is a common feature appears in crime dashboard systems. The data filtering could be implemented from either spatial perspective, temporal perspective, or attribute perspective. The data filtering allows user to inspect crime data based on their interest. The spatial filtering in crime dashboard is usually implemented by user-given polygon, such as the VPD

GeoDASH (D5) and Toronto Police Service crime dashboard (D4) that allowed users to draw rectangle and circle to filter the crime data. The polygon gives a boundary, which filters data that located inside the boundary. The temporal filtering is usually given by setting time period or selecting time point like the time axis in Toronto crime Cartio Vista (D6), and the time box in Los Angeles Crime analysis (D9). The attribute filtering is usually fulfilled by selecting crime type or related information. The filters of crime dashboard are usually aggregated. The combination of filters from different perspective present more precise and accurate delivery of information. For example, the Metropolitan Police Dashboard (D3) allowed users to filter data by setting time period, choosing specific area, and selecting crime type at the same time. However, the data filtering feature function mutually with the interactive visual design, which is another common feature that appears in crime dashboard. In most crime dashboard systems, the map is the center of interaction. The map element and charting element correspond with each other, map and charting elements themselves function as filtering tool as well. They are the receptor and the effector of dashboard visualization simultaneously.

Drilldown features is one of the extremely powerful features in business intelligence. It is a capability that takes user from a more general view of the data to a more specific one at the click of a mouse. The drill down feature allows users gain instant knowledge of different depths of data, it gives users a deeper insight of data by letting them see what makes up the figure he is analyzing. This is a widely used feature in dashboard design. The clicking object could be various like points in VPD GeoDASH (D5) that delivered detailed information about this event, or polygon clicking in 2012 Crime US Crime analysis (D1) and Metropolitan Police Dashboard (D3) that triggered the overview of selected area and visualization of crime events in whole selected area. In addition, most of the crime dashboard systems integrate crime data from multiple perspective and scales. Crime event commit in specific place at specific timing. So, it is important to look into crime events from both spatial perspectives. The crime data is also sensitive to scales, for some pattern detection and identification, different scales might cause different results. For example, the Philophobia Police Crime dashboard (D2) allowed users to inspect crime data in Philophobia by day of week, hour of day, police division, and police district.

All features listed above were the common features appeared in crime dashboard. However, there were still some great features to have in crime dashboard that appeared less frequent in existing crime dashboard systems. Having these features would enhance the functionality of a crime dashboard system and enables better and deeper understanding of crime events. Those features were discussed in following sections. These features were also the ones that need to be further improved in existing crime dashboard systems.

Spatial analysis is an essential component, but not widely available in the evaluated and existing crime dashboards. Only the Hot Sketch (D7) system implemented the hot spot analysis method and the Toronto Crime Cartio Vista (D6) implemented the heatmap. The basic visualization gives user intuitive presentation of data, where information could be obtained through visual inspection. However, the patterns and rules inside data is hard to observe when data is getting big and complex. Spatial analysis methods and algorithms need to be implemented to better discover the data. There are many algorithms and methods in geographic information science from different perspectives to reveal the patterns or rules in data, such as spatial autocorrelation, spatial interpolation, spatial interaction, simulation and modeling, and density mapping. These methods interpret the data and produces valuable information (Ding & Fotheringham, 1992).

Being real-time is another ideal feature of a geospatial dashboard, as well as crime dashboard for monitoring crime occurrences. Previously, the data was employed primarily to understand what had already happened and to predict what would happen in the future. Data management for decision support has moved through three generations, with the latest being real-time data warehousing. This latest generation is significant because of its potential for affecting tactical decision making and business processes (Waston, Wixom, Hoffer, Anderson-Lehman & Reynolds, 2006).

Another desirable component is the correlation between crime and environmental factors. The occurrence of crime is influenced by environment, such as city layout, transportation, and police resources. The places that have more crime events are usually denser residential areas, within older neighborhoods, or on streets carrying more traffic and commercial land use mixed with residences (Tylor, 2003). Those correlations between crime and environments helps to understand and predict crimes to combat crime more efficiently.

The reviewed dashboards in general contained a base map, which functioned as the container for crime data and the center of interactive visualization. Those crime dashboards generally supported data filtering and interpretation from multiple perspectives and scales. The data filtering was usually done from spatial, temporal, and attribute perspectives. Different filter requirements usually could function together with each other. However, the main lack of these crime dashboards was the analysis and predictive components. Those elements were essential to a crime dashboard if the objective is to thoroughly understand crime. Advanced methods and analysis need to be implemented to have a better discovering of knowledge and information.

CHAPTER 4 DESIGN AND IMPLEMENTATION OF A GEOSPATIAL CRIME DASHBOARD

This chapter started with the design and development methodology that were used in prototype development. It discussed the design of the prototype, which included the key design considerations, intended use, user and user requirements, and enabling technologies. Further, the chapter presented the implementation of this prototype, including interface design, backend design, and function development of the prototype.

4.1 Design of the Prototype

4.1.1 Key Design Considerations

4.1.1.1 Intended Use

The prototype is intended to facilitate users like law makers, policy makers to have a better understanding of crime events from spatial and temporal perspectives and at different scales. It is used to visualize crime data, reveal crime patterns, and predict crime events.

4.1.1.2 Users and User Requirements

The main target users of the prototype are the policy makers and lawmakers. The general public may use this dashboard if necessary statistical knowledge is equipped. As the number of expected users would be in the range of hundreds to thousands, Internet is considered an appropriate implementation platform. Ideally, all users can have their own choice of time and location to use this dashboard. In reality, a certain portion of public users may not own a computer or access to Internet, so they will be using public computers. The prototype must be operational without installation of additional plug-ins beyond a web browser.
In the evaluation of existing crime dashboards, crime dashboards design and development knowledge were discussed such as the common features and less uncommon features that appeared in crime dashboard systems. It provided a reference and guidance for the development of this crime dashboard system. The user requirements of this study were defined based on the results and experience obtained from the evaluation as well as literature review. The most common features and the features that are ideal for a crime dashboard but appeared less frequent in existing crime dashboard are referenced in the list of user requirements. For example, the crime dashboard should contain a map interface; the dashboard should integrate data from multiple perspectives and scales; the visualization of crime dashboard should be interactive; the analysis component should be included in crime dashboard to reveal patterns, correlations and so on. All the user requirements were detailed and extended from those features.

The user requirements of this crime dashboard system are listed below:

- The dashboard should visualize crime data hourly, daily, monthly by charts.
- The dashboard shall give comparison view of different crime by stack chart
- The map interface should display crime data with location and detailed information by mouse clicking.
- The map interface shall filter data by user-given rectangle and reflect on charts elements.
- The dashboard shall able to calculate crime autocorrelation and provide information about it if it is auto-correlated.
- The dashboard shall able to calculate crime correlation between various crimes and reflect if there are correlated.

- The dashboard shall able to calculate the crime mean center movement of each division and the whole area and visualize it on the map interface with line elements.
- The dashboard shall able to calculate the crime standard division ellipse of each division and visualize it on the map interface with ellipse elements.
- The dashboard shall able to analyze crime hotspot and reflect crime hotspot on the map with point elements.
- The dashboard shall able to calculate correlation between crime hotspot and environmental factors with distance measurement and reflect it.

A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved. A use case diagram can identify the different types of users of a system, and the different use cases. It is valuable for visualizing the functional requirements of a system that will translate into design choices and development priorities. It also helps identify any internal and external factors that may influence the system. It specifies how the system interacts with actors without worrying about details of how that functionality is implemented. The use case diagram of this system is shown in Figure 2.



Figure 2 Functional requirements of geospatial dashboards for crimes

4.1.2 Enabling Technologies

Various technologies can be used to build a dashboard system. The commercial products such as Tableau and Power BI provide good development platforms, taking care of each step of dashboard building. However, they are not free. Dashboard could be implemented by using other technical resources such as open source software tools and libraries. They provide tons of templates and models, and allow users to build dashboards with simply dragging, adding and necessary coding. Further, there are technologies that allows for building dashboards from scratch. In the following sections, the open source frameworks, dashboard front-end development techniques, and back end techniques were discussed.

4.1.2.1 Tools and Libraries used for Building Dashboards

Both commercial and open source software tools and libraries have been used for developing dashboard systems. ArcGIS is a good example of commercial platforms. Examples of open source plugins, solutions, framework, and platforms include Kibana, Grafana, Mozaik, and Dashbuilder. This section reviewed some of these tools and libraries for the purpose of the prototype development.

Grafana¹, open observability platform, provides open source analytics and monitoring solutions for databases. It has various visualization methods including heatmaps, histograms, graphs and maps. The main panel in Grafana is called Graph, which provides a very rich set of graphing options. It has multiples options to allow users to customize the design. The Axes tab controls the display of axes, the Legend tab controls the visibility of legend and its position, and the Display tab controls the visual properties of the graph. The Time Range tab allows the users to override the dashboard time range and specify a panel specific time. It also has a single panel, which allows users to show one main summary statistic of single series and reduce the series into a single number such as maximum, minimum, average, or sum. The Table panel is flexible and supports both multiple modes for time series as well as table, annotation, and raw JSON data. It provides data formatting, value formatting and coloring options as well. The Alert list displays

¹ https://grafana.com/

dashboard alerts and dashboard lists that provide links to other dashboards. The data source components can work with multiple time series data stores.

Grafana includes built-in integration with Graphite, Prometheus, InfluxDB, MySQL, PostgreSQL, Elasticsearch and additional data sources using plugins. Referring to the access control and authentication, Grafana allows users to restrict and control access to the dashboards, such as using the external SQL or LDAP server. In addition, Grafana's API can be used for tasks such as saving a specific dashboard, creating users, and updating data sources. Users can also create specific API keys and assign them to specific roles (Yigal, 2018).

Kibana² is an open-source data visualization and analysis tool that is beneficial for log, time series analysis, and application monitoring. It works together with a powerful search engine, ElaticSearch. Kibana becomes a default choice for visualizing data stored in ElaticSearch. Kibana is popular for its powerful and easy to use histogram features, line graphs, pie charts, heat maps, and built-in geospatial support. Kibana has multiple features that make it special. The first one is the interactive charts and visualizations. They offer intuitive charts and reports that users can use to navigate vast data interactively, such as dragging time windows, zooming into specific data sets, zooming out to see big pictures, and drilling down on reports to extract insight. The visualization components include the basic graphs, such as line, area and bar graphs, heat maps that shade cells within a matrix, and pie charts that display the contribution to a total of each source. It also includes methods such as Data Type visualization, such as Data table, Metric and Goal, and Gauge. The Coordinate map associates the results of an aggregation with geographic locations and the Region map is a thematic map where the metric value is presented by shape's color, which is the third components of the visualization methods. Time series Visual Builder visualizes time series data

² https://www.elastic.co/products/kibana

using pipeline aggregations. Other visualization components include Markdown widgets to display free -form information or instructions, Tag cloud to display words as a cloud and the size of words present importance, and Vega graph that allows customized graph, external data source, images, and user-defined interactively.

Except for these open source frameworks, dashboards also can be developed by professional GIS software such as ESRI series products. The geospatial dashboard element is an embedded conception in ESRI products. The geospatial dashboard used to be both native based products and web-based products but are now all transformed to mainly web-based (ESRI, 2019). Geospatial dashboard could be implemented by either web app builders or operation dashboards without coding. The limitation of these two products is that the data source must be a feature service or map service published by ArcGIS server in order to be displayed by the visualization charts (ESRI, 2019). The data sources such as Tile layers, Shapefiles, Feature collections, GeoRSS, CSV, txt, and gpx files, KML, OGC, Streaming feature layers, and Cached map services can only just be displayed on the dashboard base map. None of the charts and visualization elements can interpret the data directly. The dashboard is just a visualization in ESRI products, and it must cooperate with other components to generate useful information. If the operation dashboard is still considered too rigid, there are APIs that users can use to develop their own dashboard. As for being real-time, ESRI also provides real-time server, which is called Geo Event Server. The Geo Event Server itself contains the dashboard element. The key capabilities of Geo Event Server are ingesting data from almost any type of data. It has real-time data analysis such as filtering and geospatial fence, store data as feature records in the geodatabase and visualize feature records (GeoEventTeam, 2014).

The above software tools and libraries are powerful, and they are more user friendly and easy to use. However, there are limitations. Those dashboard frameworks do not have analysis component and it is hard to embed outside components. For ArcGIS products, it allows user to build dashboards with ArcGIS resources, nevertheless, more advanced analysis component and functions are only available through ArcGIS Enterprise, which is not free and pricy. Furthermore, this study also aims to study the utilities of open source resources on dashboard development, to discuss and compare various open source resource in dashboard development. The sections below discuss and compare various techniques and resources for dashboard development.

4.1.2.2 Third Party Libraries for Dashboard Front-end Development

Unlike frameworks, a library provides a set of helper functions/objects/modules called by application codes for specific functionality. The main component of the front components of a geospatial dashboard are the mapping elements and charts elements. A review of web mapping and charting libraries is discussed below.

Google Maps API, Leaflet, and OpenLayer are three web-mapping libraries that were compared below. Google maps is the most popular interactive mapping platform, while Leaflet is an open source and a free mapping plugin for JavaScript and has many add-ons and plugins at user's disposal. There are some factors to consider upon deciding which library to use. Google Maps API offers many useful services such as geolocation, autocomplete boxes, traffic and transit (Temprano, 2016). The traffic and transit services are especially useful as it extends to China and the Middle East, which is not the case of other service providers. However, Leaflet does not have their own services. It relies on third-party services. Sometimes the third-party sources' quality cannot be guaranteed. As for functionality, both Google Map and Leaflet are able to create custom markers, add geoJSON, and have layer interactions, popups, etc. Google Maps API is more efficient in handling large numbers of markers with the help of two must-have plugins: MarkCluster and Spiderfier (Google Maps Platform, 2019). Referring to the map layer, Google maps API forces the user to use the default base layer. In order to customize maps, such as by adding and removing names of city, changing the color of features, users can edit map source codes. However, Leaflet is more flexible since it is built to use multiple types of base layers and relies on the third party to supply users with a variety of map layers. In the end, Leaflet outperforms Google maps on customization (Netmag, 2014).

OpenLayer is another open source library that is used to load, display, and render maps from multiple sources on web pages. First, users are not tied to any particular map provider or technology with OpenLayers. Users are allowed to change map at any time and not have to rewrite the codes. Second, Google, Yahoo, Microsoft and ArcGIS servers are all supported by OpenLayers, which allows for the combination of data from different sources. Finally, OpenLayers provides better support for vector data, such as points, polylines, and polygons. For the precision in the maps, OpenLayers would be a more suitable map server backend than Google maps API to get a better map projection since Google maps assume that the earth is a perfect sphere (Liptak, 2018).

For charting libraries, there are some popular free libraries such as Chartjs³, Google charts⁴ and D3.js⁵, as well as commercial libraries such as Highchart⁶ and Fusionchart⁷. A review of charting libraries is discussed below.

³ https://www.chartjs.org/

⁴ https://developers.google.com/chart

⁵ https://d3js.org/

⁶ https://www.highcharts.com/

⁷ https://www.fusioncharts.com/

To begin, Chartis is one of the most popular libraries that is aesthetically pleasing, minimalistic, and right out of the box. In comparison, D3.js has a variety of visualization methods and can easily generate non-standard visualizations for large datasets. It is more than a chart library as it usually acts as the foundation of other libraries. Furthermore, D3.js is well known for its high levels of interactivity. All the charts are rendered in Scalable Vector Graphics (SVG), which enables the user to interact with each element of the chart. This also helps to display tooltips and legends for charts. Chartis only carries non-interactive canvas-based charts that are like a picture. It is just a single HTML element and cannot be modified. The charts provided by Chartjs are limited to only six types. However, D3.js has a single disadvantage, which is the steep learning curve. The code is so complicated because it involves creating each element of the chart, while Chartis is comparably easy to implement (Mark, 2014). Another popular and powerful free library is Google charts API. In comparison with D3.js, it contains various charts ranging from the simplest line graph to a complex hierarchical tree-map. It has ready-to-use examples for all the charts which can be easily implemented by embedding and exporting Portable Network Graphics (PNG).

In addition to these popular libraries, there are also visualization libraries under specific frameworks. For examples, Flot⁸ is very specific to jQuery, Ember charts⁹ are built on top of D3.js under embers framework, n3 charts¹⁰ are built on top of D3.js under Angular framework, and Victory¹¹ is designed for React. Some libraries focus on one chart, such as Sigmajs, which is good for graph drawings like the representation of node and networks. Springy.js focuses on force

⁸ https://www.flotcharts.org/

⁹ https://opensource.addepar.com/ember-charts/

¹⁰ https://n3-charts.github.io/line-chart/

¹¹ https://formidable.com/open-source/victory/docs/victory-chart/

directional graphs that offer a very "springy" animation and SmoothieCharts specializes in displaying streaming data (Krishna, 2014).

4.1.3 System Architecture

This prototype followed the MVC (Model View Controller) design pattern in a browserserver 3-tier architecture. It is a web-based application. The services run on server and accessed by the browser. MVC divides a software application into three interconnected parts. The Model specifies the logical structure of data and the high-level classes associated with it. View component is used to present the UI logic of the application. It renders the Model into a form suitable for interaction. Controller functions as an interface between Model and View components, which handles all incoming request and process all business logic. It receives input and initiates a response by calling model objects. This architecture performs well for small and monolithic software (Cazton, 2019). Figure 3 illustrates the system architecture.



Figure 3 System architecture

The browser-server architecture is an improvement of the client-server structure. The client-server architecture has its advantages in application development. The server helps in administering the whole set-up, accessing rights, and allocating resources. All the files are stored in same place, which simplify the management and searching of files. Further, new resources and changes can be edited easily by upgrading the server. In addition, browser-server architecture is platform-independent because it only requires a browser. The user interaction in web application is through a web browser that hosted in a browser-controlled environment, which makes it cost-effective for development. In addition, the user interface of web-based applications is easier to

customize and incorporate with modern design schemes. Furthermore, it is easier to share data between different computer systems. For web-based applications, installation and maintenance becomes less complicated. The three tiers structure of this dashboard system, which includes dashboard UI, data storage, and function services are introduced below.

As mentioned above, the data storage was implemented at the local file system. The data was required to contain attributes like geospatial location information, occurrence timing, and crime type. Data format conversion was done automatically using the system. All data were converted to GeoJSON format in the convenience of processing. The cloud storage service like Amazon S3 was implemented to extend the scalability of the local file system.

Once the data discussed above were appropriately integrated, functions were required to better utilize them. The services of this system could be summarized into three categories: spatial analysis service for depicting crime spatial pattern, correlation analysis service for presenting the correlation between and within crimes, and prediction analysis service for forecasting future crime occurrence. The spatial analysis service included the detection of crime direction and crime movement. The correlation analysis contained crime autocorrelation, the correlation within different crimes, and the correlation between crime and environmental factors. The prediction analysis service used clustering method to detect crime hot spots, which can be considered as the future crime occurrence.

The dashboard UI presented the interface tier, where the end users can directly interact with the system. Interactive mapping allowed the user to interact with the dashboard system and charting elements by map interface. It is an essential component of interactive visualization. The charting elements presented macro view of crime data, while the drill-down feature enabled the micro explanation. Since the data is sensitive to scales, visualization integration was also implemented from various perspective and scales, which delivered an overview of crime data as well as a detailed explanation.

4.2 Implementation of the Prototype

The following sections described the development of a geospatial dashboard prototype for crimes. The techniques used in this development were explained in detail. For each development technique, how it contributed to this system and the consideration for choosing it were discussed.

Figure 4 illustrated a schematic of the general system implementation architecture. The dashboard UI layout was implemented by Bootstrap, which is the most popular front-end development CSS framework. The map interface was fulfilled by Leaflet.js, and the charting visualization elements were done by D3.js. Node.js Express framework was used in this prototype as the web server. Since Python is essential for data analysis, and it has well-supported analytical resources, this prototype incorporated with Python by using child process. The data was stored at the local file system and cloud storage service was implemented to extend its scalability.



Figure 4 Implementation architecture

4.2.1 User Interface Development

The user interface is composed of a base map, charting elements, and a side menu bar. The base map is supported by Leaflet.js, and the charting elements are implemented with D3.js. The interface structure is fulfilled by Bootstrap, which provides a gird layout to form the interface. The side menu bar locates on the left side of the dashboard. It contains all the buttons and controls to running analysis. To the right of the menu bar, the map interface displays all crime data and analysis results. Below the map and right next to the map are the charting elements. There is a bar chart to visualize crime day information, a line chart to visualize hour information, a bubble chart to visualize division information, a pie chart to visualize month information, and a table to present statistics information.

The base map is the most fundamental element in a geospatial dashboard. It displays crime events, visualizes analysis results, and interacts with charts. All layers were controlled and managed by a Leaflet LayerGroup in this system. The LayerGroup ensured that different layers do not influence each other. The data filter function was applied on map interface and it was activated by a user-drawn rectangle. Every time the rectangle is edited, the browser sends a request to the server with the parameters of the bounding box, and the server-side filters the data by coordinates and then send the qualified data back to the browser. Another concern was the visualization of huge amounts of points on the map. In this system, Leaflet MarkerCluster was implemented to dynamically cluster all the points displayed on the map as the map is zoomed in and out to give a smooth experience. Without clustering markers, the web page can be easily frozen due to the loading of too many points.

The visualization of data was implemented by charts. The visualization could be implemented either from the client-side or server-side. Each method has its pros and cons. The advantages of server-side rendering are that it has the search engine that can crawl the site for better SEO (search engine optimization); the initial page is loaded faster and beneficial for statistics sites. The disadvantages are also apparent. It has frequent server request, overall slow page rendering, full-page reloads, and non-rich site interactions. However, the client-side advantages include rich site interactions, fast website rendering after initial load, great for web applications, and robust selection of JavaScript libraries. Its disadvantages are the low SEO if it is not implemented correctly, and that the initial load might require more time and external library (Vega, 2017). In this system, the visualization and rendering were done on the browser side due to the rich site interactions and fast website rendering. The HTTP requests only bring back either the data or processed result to the browser, and the browser implements the visualization by D3.js.

4.2.2 Backend Development and Data Preprocess

The data used in this system was stored in the file system, not a database. All data need to be converted to GeoJSON format to be stored. This crime dashboard prototype is also able to run the operations on the external data, which means the user can upload their own data to the dashboard and perform analyses. This function is supported by the Amazon S3 service. Instead of storing the user data directly to the server, which will take up too much server space as the data grows, the data from a user will be uploaded to the AWS S3 storage server. Then the server-side will read the data from S3 whenever it is needed.

In most cases, the database can be a good option for storing and managing data. However, it still depends on the situation. The pros and cons of choosing file system or database are listed in Table 4 (Thahir, 2017).

	File system		Database	
Pros	1. 2. 3. 4.	Performance can be better when store large files. A simple query will load the whole data if you select * in your query in database, while file system access files simpler and more lightweight. Saving and downloading is much simpler than database It is easier to migrate data by filesystem. It is easier to be compatible with cloud storage.	1. 2. 3. 4. 5.	ACID consistency which includes a rollback of an update that is complicated when the files are stored outside the database Files will be in sync with the database so cannot be orphaned from it which gives you an upper hand in tracking transactions. Backup automatically include file binaries More secure than Saving in file system
Cons	1.	Loosely packed, which means no ACID (Automatically, Consistency, Isolation, Durability) operations.	1.	Data may need to be converted to be stored in database

1 able 4 File system and database compariso

2. Low security	2. Memory ineffective

Besides the data storage and process, the system modularization and complex analysis calculation were another two concerns in the process of development. The implementation and concerns of these two factors were discussed in following sections.

As a weak typing language, there is no conception of class like Java and C in JavaScript. However, JavaScript modularization allows small units of independent, reusable code to be added and removed without disrupting the system. It mimics how class functions in Java or Python. There are different specifications to implement JavaScript modularization, such as AMD, CMD, CommonJS, and ES6. Node.js module system on the server-side is developed based on the CommonJS specification. The challenge is that browsers nowadays support ES6 specification instead of CommonJS. Therefore, it either makes the browser compatible with backend or makes backend compatible with front-end. In the backend, the npm that is the package manager provides lots of resources, and it only can be modularized with CommonJS. This process is implemented by JavaScript bundle tools Browserify. Browserify allows developers to write Node.js style modules that compile for use in the browser, so the developer can 'require' modules on the browser the same way as in the server. With the help of Browserify, the module system management becomes consistent from front-end to back end.

The analysis function is supported by the backend code. The ST-DBSCAN was implemented by python code and then integrated into this system due to the better performance in heavy computation tasks. The processing time is a consideration for the dashboard system design when it comes to big data. Long-time processing and waiting results in bad user experience. The default time for an HTTP request is 120 seconds. If there are massive calculation task runs in the backend and the request is not responded within default time, the browser will disconnect the request. This could be solved by extending the time-out for an HTTP request, but this method goes against the async principle and it will block the thread, which causes the browser to freeze. The solution is to use long polling. Long polling is to put the ajax request in a loop, where the browser keeps sending the request and receiving the response to check the processing status, then to end the loop when the operation is finished. Since Node is single threaded, the user needs to open an extra process for this backend tasks, which is fulfilled by Node.js ChildProcess. One should be cautious when looping the request, since only the first-time request should trigger the operation of ChildProcess, and all the other requests are only for checking the status. It is necessary to check if it is the first-time request or not. If the request is sent, which in turn triggers the operation every time, it will end up opening as many ChildProcesses as the number of requests, which makes the operation much slower. For any heavy calculation tasks like that, this would be the proper way to solve the long-time processing problem.

However, there is another way to solve this problem with more efficiency and memory effectiveness, the WebSocket. The REST style HTTP request has its limitations, such as the creation of overhead due to opening and closing, the inability to handle rapid response or real-time interactions, and the 'poll' paradigm that means the server cannot push data to the browser when it wants to (PubNub, 2015). WebSocket is a communication over a TCP connection that can overcome some limitations of an HTTP request, such as it can provide full-duplex communication, and it does not compromise with the security of the web. The choice of REST or WebSocket is depended on the situation. The WebSocket is usually used in a real-time application. WebSocket

is better with high loads, while REST is excellent for occasional communication. In this dashboard system, WebSocket could be an alternative due to the server-push function, so there is no need to send a request to check the status, and the server will send back the result whenever it is ready.

4.2.3 Function Development

The crime analysis component contains a couple of functions to reveal the spatial and temporal patterns of crime events. The spatial analysis methods used in this system were Moran's I analysis that calculate crime events autocorrelation, mean center analysis that tracks crime center movements, standard division ellipse analysis that depicts crime direction, crime correlation analysis that describes influences between each other, ST-DBSCAN analysis that predicts future crime occurrence, and correlation analysis between crime and environmental factors. The reason to choose these analysis models and how they function were discussed one by one in the following paragraphs.

The first model is the Moran's I calculation. The reason to use Moran's I is that it calculates the spatial autocorrelation of crime events. The spatial autocorrelation describes the degree of two crime events at spatial locations that are similar to each other. This analysis could be used to reveal information like if the crime events in this area are organized or if they are gang crime activities. The spatial autocorrelation helps to define how important spatial characteristics in affecting given objects in space and if there is a clear relationship (Altaweel, 2017). In this study, the spatial weights of the Moran's I was decided based on if the elements are adjacent to other elements. The whole study area was divided into multiple parts based on the division, and if division A and B are adjacent, then the spatial weight between them is 1, otherwise is 0. The spatial weights could also be determined by the distance between two elements. In this case, the distance is considered as the distance between the mean center of those two elements. This operation presents an initial overview of the spatial characteristics like if it is clustered or dispersed, it can be used as the first step to detect if data has some spatial patterns. However, more advanced models need to be applied to detect patterns.

The results obtained from this function are three values, Moran's I indicator, the P-value, and Z score. The Moran's I indicator ranges from -1 to 1. It is a perfect dispersion when the value is -1, and it is perfect randomness when the value is 0, it indicates perfect clustering of similar values when the value is 1 (Tiefelsdorf, 2006). The null hypothesis is the complete spatial randomness (Mitchell, Andy, 2005). The p-value is the possibility that you have falsely rejected the null hypothesis, Z-score is a test of statistical significance that helps to decide whether or not to reject the null hypothesis. The P-value and Z-score usually appear together and form the trend of the distribution. The null hypothesis is only accepted when p is larger than 0.05 or less than - 0.05. The P-value and Z-score distribution are shown in Figure 5.



Figure 5 P value and Z score table

The calculation of Moran's I is listed below. S_0 is the sum of all the spatial weights, z_i is the difference between the element i's attribute and the mean which is $(x_i - X)$, $w_{i,j}$ is the spatial weights between i and j, n is the sum of all elements.

$$V[I] = E[I^{2}] - E[I]^{2}$$
$$E[I] = -1/(N - 1)$$
$$z_{I} = \frac{I - E[I]}{\sqrt{V[I]}}$$
$$I = \frac{n}{S_{0}} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} Z_{i} Z_{j}}{\sum_{i=1}^{n} z_{i}^{2}}$$

The second analysis component used in this dashboard is the mean center and standard division ellipse. The mean center method is to track the crime movement path within a single day. The mean center of crime events is calculated every 6 hours, from 0:00~ 6:00, 6:00 ~12:00, 12:00 ~ 18:00, 18:00 ~ 24:00. There would be four mean centers in total, and a line could be generated by connecting all mean centers. This line could be considered as the path of the crime movement within a day. The method could be used in allocating the police resources based on the switching of the crime center. The formula of the mean center is listed below, x_i is the value and w are the weights, x is the mean center value:

$$\mathbf{x} = \frac{\sum_{1}^{n} (x_i * w_i)}{\sum_{1}^{n} w_i}$$

The standard division ellipse is to reveal the distribution direction of crime events in the belonging division. It was implemented in this system in the unit of each division and the whole area. The direction information can be used to analyzed together with the information such as roads, buildings, and infrastructures, to determine the relationship between crime direction and these factors, and how the factors influence the direction. The standard deviation ellipse is given as:

$$SDE_x = \sqrt{\frac{\sum_{i=1}^n (x_i - X)^2}{n}}$$
$$SDE_y = \sqrt{\frac{\sum_{i=1}^n (y_i - Y)^2}{n}}$$

where x_i and y_i are the coordinates of feature I, {X, Y} represent the mean center for the features, and n is equal to the total number of features.

The third analysis component is Spatial-Temporal DBSCAN. This clustering method detects the hotspots of crime events from both spatial and temporal perspective. It presents the places that crime occurs most dense, and these places can be considered as the places that future crime will most possibly locate. The results of hotspot analysis are also used as a reference for analyzing correlations between crimes and environmental factors. There are clustering methods such as partitioning methods, hierarchical methods, density-based methods, and partitioning methods. The representation of partitioning methods is K-means, it divided the n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster (Hartigan & Wong, 1979). The drawback of partitioning methods is that it cluster objects based on the distance between them, which results in finding only spherical-shaped cluster and having the difficulty of discovering clusters in arbitrary shapes. The density

method solves the limitation of partitioning methods. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering algorithm is one of the most commonly used density methods for clustering in data mining. DBSCAN can find the clusters of variable sizes and shapes, and it will detect the noise (Dharni & Bnasal, 2013). The principle of this method is that the cluster is kept growing until they fail to fulfill the requirement of being a core point. In a scenario of crime analysis, it is necessary to detect hotspots from both spatial and temporal perspectives.

The principle of DBSCAN is discussed a lot in all kind of papers and online materials as it is the most popular density-based clustering method. There are a couple of things to notice in the DBSCAN method. Two parameters that influence the algorithm's performance need to be carefully taken care of. The first one is the EPS, which is also the radius ε , and the second one is the minimum-points. The minimum number of points is usually considered as 4. When the minimum of points is less than 2, the result will be as same as the hierarchical clustering with a single link metric, so the minimum number of points must be greater than 3. The bigger minimumpoints value usually applied to datasets that have more noise or has a large size or contains duplicates (Schubert, Sander, Ester, Kriegel, & Xu, 2017). The EPS is defined by the minimumpoints through the k-distance graph, the diagram is shown in Figure 6.



Figure 6 K-distance graph

For each point in the dataset, the distance from the point in question to all other points in the dataset is obtained. These are then sorted by distance to obtain the Kth distance. For dataset that contains n points, there are n Kth distance, sorted all the Kth distance and then the Kth distance of these Kth distances is the EPS distance.

In this study, ST-DBSCAN (Birant & Kut, 2007) method was implemented to identify crime hotspots. The method functions as the prediction method since the hotspot results could be considered as the future crime spots (Block & C, 1995). The temporal EPS is 6 hours, which means if two crime events occur within 6 hours, these two are temporally neighbored. The minimum number of points is still 4 here as this is usually the most suitable value that is verified by large amounts of experiments. This method was not applied to the full-size data in this study. This is due to the limitation of the method. The complexity of the original DBSCAN algorithm is O(n)~O(n2). The DBSCAN is known to work well as small datasets; it is time-consuming to run a large dataset. The fast region query method, like k-d trees or ball trees, could be implemented as future study to decrease execution times and lower the time complexity to $O(n \log n)$.

The fourth analysis part is the correlation analysis between different kinds of crimes. This part of study is to verify if there is correlation between different crime types. For example, if the occurrence of a crime would somehow cause another crime to happen. It calculates the possibility that crime type A will occur under the occurrence of crime type B. These two approaches are original methods that extend the Tobler's First Law of Geography (Miller, 2004), which is "everything is related to everything else, but near things are more related than distant things."

The first method is called max-min distance. For example, for crime instance 'a' in crime type A, there is one crime instance that nearest to 'a' from each of the others crime type, and all

the nearest instances form a dataset C. The crime instance that is nearest to 'a' in dataset C is the crime type that most related to crime instance 'a'. For the whole crime type A, there will be n instances 'a', therefore there are n dataset C, and n instances that are nearest to each 'a' in A. Find the crime type that takes the largest percentage of n, then this is the most related crime type towards A. The result has to be significant enough to be considered that these two crimes are related. For example, 50 percent means half-and-half, which means the possibility of this crime happen under the other crime is random. Figure 7 is listed below to illustrate the method, b, c, d, and e are the nearest crime events from B, C, D, E to a.



Figure 7 Max-min distance method

The second method is called the max-min score. Assume there are N types of crimes in total, for crime instance "a" in A, there are N-1 points that are located most close to a. If the dataset is distributed pure randomly, and each crime have equal influence on each other, then the N-1 points should be composed of one instance from each other crime type. There are two requirements need to be fulfilled to decide which crime type in this N-1 points is most related to "a", the distance and amount. The close the distance, the higher the relation; the larger amount of the instances, the higher the relation. The result can be presented as a score S, which calculates the influence of each crime type in the N-1 points towards instance 'a'. S is the score, and n is the number of instances of the crime that are located within the N-1 points. d(i,a) is the distance between instance "i" and "a", the equation is given as:

$$S = \sum_{i=1}^{n} \frac{1}{d_{i,a}}$$

The crime type with the highest S score can be considered as the most related crime type to "a". Sum up the S value for all the "a" in A, and get a final result to decide which crime type is most related to A. The figure 8 is listed below to help understand this method. For example, it looks 'c' is closer and more related to 'a', but 'b' might win over c since b has more instances included.



Figure 8 Max-min number method

The next component is the analysis of influence of environmental factors on crimes. The environmental factors include but not limited to transportation, infrastructure, weather, city layout and so on. In this study, homeless shelter and police facilities are taken as two examples to test the environmental factors influence toward crimes. Some people think that homeless shelter might be a potential security risk, and it is the hotbed of crime; also, the position of police facilities influence the occurrence of a crime event. This hypothesis will be tested in this study. In this study, the correlation analysis between crime and police resources, the correlation analysis between crimes and homeless shelters are implemented by distance measurement. The correlation between crimes events and environmental factors are calculated based on distance. The conception 'walking distance' is brought up here. It is the distance people are willing to walk before opting to drive (Morphocode, 2019).

4.3 Summary

The first half of this chapter presented the design of this prototype. The prototype was intended to help decision maker have a better understanding of crime events, and it was developed as an online application which only requires a web browser. The main objectives of this prototype are to visualize and interpret crime data from various perspectives and scales, such as spatially, temporally, correlation and predictive aspects. Based on these considerations, the design consideration, user and user requirements, and enabling technology were derived.

The second half of this chapter covered the implementation of the prototype. Since this is a dashboard system, it is essential to visualize and present data and results. The implementation began with the user interface and visualization elements, followed by the implementation of backend analysis component based on the corresponding interface elements.

CHAPTER 5 DEMONSTRATION OF THE PROTOTYPE

In this chapter, Toronto city crime data is used to illustrate the functionality of the prototype that facilitates effective analysis of crime events. This chapter first introduces the Toronto crime data and then walks readers through the prototype by showing how data analyses are performed using the prototype, including data visualization, spatial analysis, and predictive crime analysis.

5.1 Toronto Crime Data

In this study, City of Toronto is selected as the study area. The crime data is obtained from Toronto Police Service Public Safety Data Portal Open Data¹². The data includes 71441 events of assault crimes from 2014 to 2017 containing break and enter from 2014 to 2017 with 27354 events, robbery from 2014 to 2017 with 14592 events, theft from 2014 to 2017 with 4210 events, and homicide from 2004 to 2017 with 920 events. Table 5 shows the basic statistics of the data.

Crime dataset name	Number of crime events	Time period
Assault 2014 to 2017	71441	2014 ~ 2017
Break and Enter 2014 to 2017	27354	2014 ~ 2017
Robbery 2014 to 2017	14592	2014 ~ 2017
Theft Over 2014 to 2017	4210	2014 ~ 2017
Homicide	920	2004 ~ 2017

Table 5 Toronto crime data basic statistics

¹² http://data.torontopolice.on.ca/search?q=crime

5.2 Data Analysis Using the Prototype

5.2.1 Query and Visualization of Data

Assault crimes are used as an example to show data query and visualization. In the user interface shown in Figure 9, the top section displays a base map, monthly pie chart, and data statistics, while the bottom section shows a daytime line chart, weekday bar chart, and division bubble chart.



Figure 9 Assault crime overview

From the monthly pie chart, the crime events occurred in similar numbers each month. The greatest number of crimes occurred in July with 6438 events, and the minimum occurrence happened in February with 5099 events. The bar chart of weekday distribution shows that weekends have more crime events, with an upward trend beginning from Friday. From the daytime distribution by the line chart, there is a distinct drop around 5 and 6 o'clock in the early morning, and the peak is at midnight. The division distribution is implemented by the bubble chart. The

divisions are divided by police resource distribution, the maximum division is D43, and the minimum division is D13. The Toronto divisions map is shown in Figure 10. Besides the data visualization, the division data is also used in Moran's I analysis as spatial weight units to calculate autocorrelation within all divisions. The total crime distribution delivers an overview and comparison of the crime data in Toronto as displayed in Figure 11.





Figure 10 Toronto police command division boundaries

Figure 11 Total crime distribution

From all the statistics and results in this dataset, crime distributions vary when compared to each other, but correlations between distributions may also exist. The crime event patterns are influenced by environmental and societal factors rather than the actual crime. Each element in the criminal event has some historical relevance shaped by experience and future intention, by routine activities and rhythms of life, and by constraints of the environment (Brantingham & Brantingham, 1993). From the single crime chart and the overall charts, patterns could be obtained from the visualization. Those patterns appear to be reasonable when they are analyzed together within the context of actual situations. For example, the occurrence of assault crimes is most often at its peak around 12:00 and 24:00, and it stayed at a relatively high level after 16:00 when most people are getting off work and are starting to hang out on the street. The weekends also have more assault events since people go out more on the weekends more than weekdays. The temperature has an influence on crime events as well. July is the peak month of all crimes. The weather generally becomes nicer, people take vacations, there are lots of entertainment events in the summer. Good weather drives people to go out and get together, which is a factor in cultivating the generation of assault crime. Break and enter crimes happened mostly at 12:00 and 24:00, which is the time that people go to sleep, while the lower point is around 6:00, which is the time when people start to get up. Weekends had the least number of break-in events since people usually stay at home, making it a bad time for a successful break and enter. Robberies occur the most at 21:00 and the least at 7:00. From the line chart, the morning distribution is lower than the afternoon, which means robbery usually happens on people's way home. Theft crimes are also sensitive to timing. Generally, crime is more likely to happen in nice weather, and the possibility of crime events decreases as the temperature goes down. Everything discussed above is from the temporal view. From the police division distribution, the divisions located around the border and far away from

the city center have the most crime events, regardless of the type of crime; however, the divisions in town have fewer crimes. This can also be explained by the police resource distribution; the outer divisions are larger and have comparably fewer police resources.

5.2.2 Basic Statistical and Spatial Analysis of Data

This crime dashboard system has basic statistical and spatial analysis methods applied to reveal crime patterns from various perspectives and scales, such as autocorrelation analysis to detect if crime events are distributed randomly or clustered, crime correlations analysis to see if different crimes influence each other, and crime mean center as well as crime ellipse to reveal crime movements and directions. The demonstration of these analyses toward Toronto crime data is discussed below.

Autocorrelation is a measure of similarity between nearby observations. Global Moran's I is implemented in this study to calculate the autocorrelation. The result of the analysis is displayed with a pop-up window on the screen as shown in Figure 12.



Figure 12 Alert box to show analysis results

Spatial autocorrelation is implemented in the unit of division, and the attribute of the weight matrix is the number of crime events of each division. The spatial weight matrix is defined when those two divisions are adjoined. If they are adjoined, then the weight is 1 and otherwise is 0. The

Moran's I indicator, p-value and z scores of those crimes are listed below; the p-value is the result that the corresponding p-value of z score in the Normal table subtracted by 1:

- Assault crime: Moran's I indicator: -0.005625911822884292, z-score:
 0.539569127647453, p-value: 0.2981
- Break and Enter crime: Moran's I indicator: -0.13580617315673188, z-score:
 -0.6975197506740045, p-value: 0.2451
- Robbery crime: Moran's I indicator: 0.11477717891542025, z-score:
 1.9836065086161278, p-value: 0.0465
- Theft crime: Moran's I indicator: -0.02409868954429486, z-score:
 0.3658051491605641, p-value: 0.3594

The p-value represents the significance of the null hypothesis. The null hypothesis is not to be rejected when p is less than 0.05. Assault, break, and theft all have a negative autocorrelation. Assault, break, and theft all have P-value of around 0.2, which is too big to reject the null hypothesis. Only the robbery crime has a positive indicator, and the p-value is 0.0465, which is less than 0.05, allowing the rejection of the null hypothesis.

This component detects the correlation between crimes. The correlation analysis is applied between assault, break and enter, robbery, and theft crime. The analysis results are displayed similar to the autocorrelation in Figure 19. Method 1 is the max-min distance method and method 2 is the max-min score method. A sample that contains 4000 events of each crime type in addition to a sample that contains 1000 events of each crime type are used in this analysis. The number 4000 is chosen because the maximum amount of Theft Over is 4000. The number 1000 is chosen to form a comparison with 4000 if there is no correlation obtained from 4000. The comparison of different samples reveals if and how the correlation changes with the sample amount. The results are listed in Table 6.

	Method1	Method2	Method 1	Method2
	4000	4000	1000	1000
Assault-break	53.11%	32.61%	39.37%	28.74%
Assault-robbery	28.16%	37.25%	31.46%	37.73%
Assault-theft	18.72%	30.13%	29.15%	33.52%
Break-assault	44.24%	33.52%	36.17%	31.89%
Break-robbery	25.09%	32.29%	28.65%	34.92%
Break-theft	31.39%	34.18%	35.17%	33.17%
Robbery-assault	47.92%	35.14%	39.67%	37.97%
Robbery-break	33.61%	32.73%	33.66%	32.70%
Robbery-theft	19.17%	32.12%	26.55%	29.31%
Theft-assault	43.38%	32.94%	38.17%	36.19%
Theft-break	40.99%	35.30%	40.28%	31.46%
Theft-robbery	16.30%	31.74%	21.54%	32.33%

Table 6 Correlations between different crimes

From the result, it shows that there is no important correlation between the crimes analyzed. When the crime sample amount is 4000 and is analyzed by method 1, the most related crime to assault is break and enter, with a percentage of 53.11%. Assault crimes have the highest correlation with break and enter crimes, with a percentage of 44.24%. Similarly, assault crimes have the highest relation to robberies, with a percentage of 47.92%. Again, assault is also highly correlated with theft, with a percentage of 43.38%. All these correlations are around 50% and none of them is over 80%. The results might be different if the data is not limited to a sample size of 4000. For now, the conclusion is that there is no correlation between these four crimes.

Autocorrelation provides information about the crime on whether it is spatially clustered or scattered. The mean center is used to track the transformation of the crime center within one single day, and the standard deviation ellipse reveals the direction of crime event distribution in a specific area.

In this study, the crime events that happened within one single day are divided into four parts, 0:00 to 6:00, 6:00 to 12:00, 12:00 to 18:00, and 18:00 to 0:00. For each time window, there is one mean center calculated. Four mean centers in total track crime center movements in a day. This could be useful for the assignment of police resources during different periods. Each division of Toronto city is tracked, in addition to a tracked crime center for the entire Toronto area.

Figure 13 shows the crime mean center movement of assault crimes. The mean center transformation of the entire Toronto area is displayed in green; the mean center transformation of each division is displayed in blue.



Figure 13 Assault crime mean centers

From the mean center results, some patterns could be observed. The red arrow on the track path implies the direction. The first pattern is that the path from 0:00~6:00 to 6:00~12:00. All four crimes' mean centers moved toward either the east or the northeast direction during this period. This could be related to factors like transportation, infrastructure, and the city layout. The second pattern observed reveals that all the trajectories of the mean centers took a turn. None of them traveled in a straight line or toward the same direction, and they all returned to the direction where they started as time goes by. This also verifies the first pattern. Those results could be studied and researched together with other city data such as infrastructure data to analyze the correlation between these factors, the results from which are useful for future city management.

The mean center transformation delivers the dynamic movement of crime within one day, while the standard deviation depicts the direction of spatial distribution. The standard deviation distribution is applied to each division of Toronto city. The result of assault crime is shown in Figure 14.


Figure 14 Assault crime standard division ellipse

Interestingly, the standard deviations looks similar to each other. Comparing the result with what is shown in Figure 9, which is the division map of Toronto police resource, one can see that the shape of each ellipse or the direction of each ellipse follows the size and shape of each division area. These results reveal that there is no obvious direction of crime events in each division. However, this does not mean that there is no directional pattern at all since the data is sensitive to scale. The result might vary if the standard division is applied with a different scale.

5.2.3 Predictive Crime Analysis

As mentioned before in the predictive crime analytics, there are many methods to predict crime events, and hotspot analysis is one of the most popular and common approaches. In this study, the hotspot analysis is implemented by the spatial-temporal DBSCAN (ST-DBSCAN) method. It detects the hotspot of crime events in Toronto. Furthermore, the clustering result is analyzed together with police resources and shelter locations to obtain the correlation between crime and those factors. The results of ST-DBSCAN and the correlation are listed below.

Assault crime data is taken as an example to show the results of ST-DBSCAN and correlation as Figure 15, Figure 16, and Figure 17.



Figure 15 Hotspot results

The blue points are the crime hotspots of assault in Toronto, the pink points are homeless shelter locations, and the grey points are police facilities in Toronto. It can also be seen that most assault crime hotspots are far from the police facilities.



Figure 16 Assault hotspot and homeless shelters



Figure 17 Assault hotspot and police facilities

The correlation of crime with the locations of homeless shelters and police facilities are calculated based on distance. The walking distance is given as 400m (Morphocode, 2019). It is assumed that the homeless shelter residences do not have a car or other transportation methods. For the assault crime data, only 2% percent of crime hotspots have at least one homeless shelter within 400 meters. This value changes to 4% for break and enter, 5% for robbery, and 1.8% for theft. From these results, the conclusion is that homeless shelters are not a hotbed of crime as was speculated before. Recalled the correlation between crime hotspot and police facility, there are 21% of assault crime hotspots located within 1000 meters of at least one police facility. For break and enter crimes, this percentage is 2%, 5% for robberies, and 9% for theft. From these results, the conclusion is that police facilities indeed deter break and enter, robbery, and theft crimes.

5.3 Summary

This chapter presents the demonstration of this prototype with the crime data from Toronto city, which proves that this crime geospatial dashboard is able to facilitate the understanding of crime events. It enables users to have a better understanding of crime events from spatial patterns, temporal patterns, and correlations. It is also capable of predicting future crime events by the implementation of crime hotspots.

CHAPTER 6 CONCLUSION

The objective of this research was to design and implement a crime dashboard prototype to demonstrate that geospatial dashboards can be a useful tool to derive a better understanding of crime events and support crime reduction and prevention. The following tasks were implemented to accomplish this objective.

First, a literature review of predictive crime analytics and geospatial dashboard usage in crime analysis was conducted to acquire background knowledge about this research. It introduced the predictive crime analytics including descriptive statistics approaches, data mining approaches, and GIS approaches. The geospatial dashboard applications in general usage and in crime prevention were analyzed, and crime dashboard design patterns were summarized.

An assessment on selected crime dashboards was then conducted to find out the extent to which they can help enhance understanding of crime events. First, the purpose of the evaluation was defined. It helps to understand the development and limitation of existing crime dashboard systems. Then, a set of evaluation criteria from the principles of evaluation of effective dashboards with the consideration from geospatial and crime perspectives were developed. After the evaluation criteria were established, nine crime geospatial dashboard applications were selected for evaluation. A detailed review of each application in terms of the availability of crime analysis and prediction was conducted to serve as background information to facilitate the evaluation. The evaluation results served as the baseline for later prototype design and development. The evaluation revealed that existing applications have not yet taken full advantage of GIS analysis capabilities in aiding the understanding of crime events. The above evaluation provided insight into the direction a preferred crime geospatial dashboard system should proceed. Other key design considerations for the prototype were identified, including the intended use, target user group, and user requirements. Based on these considerations, a crime dashboard architecture was developed. For the purpose of proof-of-concept, a crime dashboard prototype system was implemented based on the dashboard architecture with available components. The interface design was completed by Leaflet.js and D3.js, followed by the implementation of the system backend with the help of Node.js Express framework. The analysis function design that used in this prototype was later discussed. This crime dashboard system delivered a deeper and thorough understanding of crime events from multiple perspectives.

Through this study, observations of crime dashboard design and development was obtained. The crime geospatial dashboard is an extension of dashboard system. It follows the most basic and fundamental dashboard design rules but also varies on some aspects with some special considerations, which exists in mainly following contents.

The first is the dashboard visualization. It is a trend that crime dashboards have interactive data visualization. The map element usually functions as the core of interactive visualization. It works not only as the container of geospatial data, but also interacts with users and other visualization components. Further, crime dashboards tend to visualize crime data from multiple perspectives and scales, which delivers a more complete and thorough interpretation.

Refers to the data, the data of crime dashboard could be stored and obtained either on a file system or database depend on the situation. The database is not always the better choice due to crime dashboard is not designed for heavy querying job. However, in the consideration of scalability, a cloud storage service could be implemented as an extension of file system to accommodate all size data. Besides, the data should be converted into GeoJSON format in the convenience of developing and processing.

Crime geospatial dashboards in general contains base map, charting elements, drill-down features, and data filtering. However, the main lack of existing crime dashboard is the crime analysis and predictive components. A desired crime dashboard should be cable of interpreting and analyzing crime data from multiple perspectives and scales, such as spatial perspective, temporal perspective, and correlation perspective. Predictive methods include descriptive statistics, data mining, and GIS approaches could be implements in crime dashboard to assist the analysis process. These analyses are usually heavy calculation tasks that are time consuming and memory consuming. Therefore, calculation optimization and performance optimization techniques are needed in the consideration of better user experience.

Certain things could be improved for future development. The real-time component could be implemented as an extra component to monitor ongoing crime events, which makes crime dashboard not only an analysis platform but also a monitoring tool. The output result of dashboard analysis could be formatted as a HTML report instead of just pop-up window to give more detailed and complete presentation. Refer to the analysis functions, the autocorrelation analysis could be implemented with different scales and different weight matrix to get a more precise result. The weight matrix could be obtained by the inverse-distance method, and the size of each unit could be smaller instead of just using districts. Besides, the ST-DBSCAN algorithm needs to be improved by accelerating the searching speed by ball tree or KD tree to be compatible with big data. In addition, the lack of consideration of environmental factors towards crime limits the generalization of the crime prediction model. In future studies, more advanced and generalized approaches such as GLM, ST-GAM and LST-GAM could be established to achieve crime prediction, which predicts the crime with the consideration of environmental factors.

REFERENCES

- Altaweel, M. (2017, May). GIS and Spatial Autocorrelation. Retrieved from https://www.gislounge.com/gis-spatial-autocorrelation/
- Amiruddin, S. M. (2016, June). Real-time Web GIS to monitor marine water quality using wave glider. In *IOP Conference Series: Earth and Environmental Science* (Vol. 37, No. 1, p. 012074). IOP Publishing.
- Anderson, D., Chenery, S., & Pease, K. (1995). *Biting back: Tackling repeat burglary and car crime* (pp. 1-57). London: Home Office Police Research Group.
- Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering*, 60(1), 208-221.
- Block, C. R. (1995). STAC hot-spot areas: A statistical tool for law enforcement decisions. In Crime analysis through computer mapping. Washington, DC: Police Executive Research Forum (pp. 15-32). Citeseer.
- Brantingham, P. L., & Brantingham, P. J. (1993). Environment, routine and situation: Toward a pattern theory of crime. *Advances in criminological theory*, *5*(2), 259-94.
- Brantingham, P., & Brantingham, P. (1995). Criminality of place. *European journal on criminal policy and research*, *3*(3), 5-26.
- Brown, D., Dalton, J., & Hoyle, H. (2004, June). Spatial forecast methods for terrorist events in urban environments. In *International Conference on Intelligence and Security Informatics* (pp. 426-435). Springer, Berlin, Heidelberg.
- Cazton. (2019). Software Architecture Consulting. (n.d.). Retrieved from https://cazton.com/consulting/enterprise/software-architecture
- Cerutti, V., Fuchs, G., Andrienko, G., Andrienko, N., & Ostermann, F. (2016). Identification of disaster-affected areas using exploratory visual analysis of georeferenced Tweets: application to a flood event. Association of Geographic Information Laboratories in Europe: Helsinki, Finland, 5.
- Chainey, S., & Ratcliffe, J. (2013). GIS and crime mapping. John Wiley & Sons.
- Chang, Y. C., & Chen, S. M. (2009, October). Temperature prediction based on fuzzy clustering and fuzzy rules interpolation techniques. In 2009 IEEE International Conference on Systems, Man and Cybernetics (pp. 3444-3449). IEEE.
- Chen, H., Chung, W., Qin, Y., Chau, M., Xu, J. J., Wang, G., ... & Atabakhsh, H. (2003, May). Crime data mining: an overview and case studies. In *Proceedings of the 2003* annual national conference on Digital government research (pp. 1-5). Digital Government Society of North America.
- Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: a general framework and some examples. *computer*, (4), 50-56.

- Chitsazan, M., Rahmani, G., & Neyamadpour, A. (2013). Groundwater level simulation using artificial neural network: a case study from Aghili plain, urban area of Gotvand, south-west Iran. *Geopersia*, *3*(1), 35-46.
- Dharni, C., & Bnasal, M. (2013, December). An improvement of DBSCAN Algorithm to analyze cluster for large datasets. In 2013 IEEE International Conference in MOOC, Innovation and Technology in Education (MITE) (pp. 42-46). IEEE.
- Ding, Y., & Fotheringham, A. S. (1992). The integration of spatial analysis and GIS. *Computers, Environment and Urban Systems*, *16*(1), 3-19.
- Eckerson, W. W. (2007). Predictive analytics. *Extending the Value of Your Data Warehousing Investment. TDWI Best Practices Report*, 1, 1-36.
- Edwards, D., & Thomas, J. C. (2005). Developing a municipal performance-measurement system: Reflections on the Atlanta Dashboard. *Public Administration Review*, 65(3), 369-376.
- ESRI. (n.d.). Operation dashboard: Map element. Retrieved from https://doc.arcgis.com/en/operations-dashboard/help/map-element-and-tools.htm
- ESRI. (n.d.). Understand data sources. Retrieved from https://doc.arcgis.com/en/operationsdashboard/help/understand-data-sources.htm
- ESRI. (n.d.). What is a dashboard. Retrieved from https://doc.arcgis.com/en/operationsdashboard/help/what-is-a-dashboard.htm
- Farrington, D. P., & Lambert, S. (2018). Statistical approaches to offender profiling. In *Profiling property crimes* (pp. 243-284). Routledge.
- Few, S. (2006). Information dashboard design.
- Ganapati, S. (2011). Key features for designing a dashboard. *Government Finance Review*, 0883-7856.
- GeoEventTeam. (2014, Dec 30th). Tutorial-Introduction to GeoEvent Server. Retrieved from http://www.arcgis.com/home/item.html?id=b6a35042effd44ceab3976941d36efcf
- Godwin, A., & Stasko, J. (2017, January). Hotsketch: Drawing police patrol routes among spatiotemporal crime hotspots. In *Proceedings of the 50th Hawaii International Conference on System Sciences*.
- Google Maps Platform. (2019n.d.). Marker Clustering. Retrieved from https://developers.google.com/maps/documentation/javascript/marker-clustering
- Groff, E. R., & La Vigne, N. G. (2002). Forecasting the future of predictive crime mapping. *Crime Prevention Studies*, 13, 29-58.

- Grover, V., Adderley, R., & Bramer, M. (2006, December). Review of current crime prediction techniques. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence* (pp. 233-237). Springer, London.
- Grubesic, T. H. (2006). On the application of fuzzy clustering for crime hot spothotspot detection. *Journal of Quantitative Criminology*, 22(1), 77.
- Grubesic, T. H., & Murray, A. T. (2001, December). Detecting hot spothotspots using cluster analysis and GIS. In *Proceedings from the fifth annual international crime mapping research conference* (Vol. 26).
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100-108.
- Hauck, R. V., Atabakhsb, H., Ongvasith, P., Gupta, H., & Chen, H. (2002). Using Coplink to analyze criminal-justice data. *Computer*, 35(3), 30-37.
- Karami, M., Langarizadeh, M., & Fatehi, M. (2017). Evaluation of effective dashboards: key concepts and criteria. *The open medical informatics journal*, *11*, 52.
- Kazak, J., Chalfen, M., Kamińska, J., Szewrański, S., & Świąder, M. (2017, March). Geodynamic decision support system for urban traffic management. In *Proceedings of GIS Ostrava* (pp. 195-207). Springer, Cham.
- Kianmehr, K., & Alhajj, R. (2006, March). Crime Hot-spots prediction using support vector machine. In *IEEE International Conference on Computer Systems and Applications*, 2006. (pp. 952-959). IEEE.
- Krishna. (2014, Sep). *D3.js vs Google Charts: A data scientist's review*. Retrieved from https://humansofdata.atlan.com/2014/09/d3-js-versus-google-charts/
- Kumar, A. (2015). Designing of controller based on artificial neural network for liquid level system. International Conference on Advances in Electronics, Electrical and Computer Engineering.
- Lee, D., Felix, J. R. A., He, S., Offenhuber, D., & Ratti, C. (2015, July). CityEye: Real-time visual dashboard for managing urban services and citizen feedback loops. In *Proc.* 14th Int. Conf. on Comp. in Urban Plann. & Urban Manag. CUPUM (Vol. 2015).
- Liptak. (2018). Google Maps now depicts the Earth as a globe. Retrieved from https://www.theverge.com/2018/8/5/17653122/google-maps-update-mercatorprojection-earth-isnt-flat
- Lu, Y., Hu, X., Wang, F., Kumar, S., Liu, H., & Maciejewski, R. (2015, May). Visualizing social media sentiment in disaster scenarios. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 1211-1215). ACM.
- Lucas, R. (1986, September). An expert system to detect burglars using a logic language and a relational database. In *on Proceedings of the Fifth British National Conference on Databases (BNCOD 5)* (pp. 43-54). Cambridge University Press.

- Mark. (2014). Comparison between d3.js and chart.js (only for charts). (n.d.). Retrieved from https://stackoverflow.com/questions/27347798/comparison-between-d3-js-and-chartjs-only-for-charts
- McArdle, G., & Kitchin, R. (2016). The Dublin Dashboard: Design and development of a real-time analytical urban dashboard.
- Miller, H. J. (2004). Tobler's first law and spatial analysis. *Annals of the Association of American Geographers*, 94(2), 284-289.
- Mitchell, Andy. (2015). How Spatial Autocorrelation (Global Moran's I) works. Retrieved from https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm
- Morphocode. (2019). *the 5-minute walk*. Retrieved from https://morphocode.com/the-5-minute-walk/
- Netmag. (2014, Dec 31th). Is Leaflet a better tool than Google Maps? Retrieved from https://www.creativebloq.com/web-design/leaflet-google-maps-121413738
- OpenLayers vs Google Maps. (n.d.). Retrieved from https://stackoverflow.com/questions/1237187/openlayers-vs-google-mapsLiptak, A. Google Maps now depicts the Earth as a globe. (2018). Retrieved from https://www.theverge.com/2018/8/5/17653122/google-maps-update-mercatorprojection-earth-isnt-flat
- PubNub. (2015, Jan 25th). *WebSocket vs REST: Understanding the Difference*. Retrieved from https://www.pubnub.com/blog/2015-01-05-websockets-vs-rest-api-understanding-the-difference/
- Ratcliffe, J. H., & McCullagh, M. J. (1998). Identifying repeat victimization with GIS. *The British Journal of Criminology*, *38*(4), 651-662.
- Rienks, R. (2015). *Predictive Policing: Taking a chance for a safer future*. Politieacademie, Lectoraat Intelligence.
- Rodrigues, A., Diggle, P., & Assuncao, R. (2010). Semiparametric approach to point source modelling in epidemiology and criminology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(3), 533-542.
- Saha, S., Shekhar, S., Sadhukhan, S., & Das, P. (2018). An analytics dashboard visualization for flood decision support system. Journal of Visualization, 21(2), 295-307.
- Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. ACM Transactions on Database Systems (TODS), 42(3), 19.
- Shrivastav, A. K., D. (2012). Applicability of soft computing technique for crime forecasting: a preliminary investigation. *International Journal of Computer Science & Engineering Technology*, 9(9), 415-421.

- Sim, W. (2012). 2012 US Crime Analysis Dashboard. Retrieved from https://public.tableau.com/profile/wee3190#!/vizhome/USCrimeAnalysisDashboard1 20814-1/MainDashboard
- Taylor, R. B. (2003). Crime Prevention through Environmental. *Handbook of environmental psychology*, *413*.
- Temprano, V. G. (2016, Nov 2th). Google Maps API or Leaflet: What's the Best for your Project? Retrieved from: https://www.codementor.io/victorgerardtemprano/googlemaps-api-or-leaflet--what-s-best-for-your-project-faaev60vm
- Thahir, A. (2017, Feb 28th). *Which is better? Saving Files in Database or in File system.* Retrieved from https://habiletechnologies.com/blog/better-saving-files-database-file-system/
- Tiefelsdorf, M. (2006). *Modelling spatial processes: the identification and analysis of spatial relationships in regression residuals by means of Moran's I* (Vol. 87). Springer.
- Tsou, M. H., Jung, C. T., Allen, C., Yang, J. A., Gawron, J. M., Spitzberg, B. H., & Han, S. (2015, July). Social media analytics and research test-bed (SMART dashboard). In *Proceedings of the 2015 international conference on social media & society* (p. 2). ACM.
- Vancouver Police Department. (2019). VPD GeoDASH Crime Map. Retrieved from https://geodash.vpd.ca/Html5Viewer/?disclaimer=on&viewer=VPDPublicRefresh_gv h&x=33&y=38
- Vega, J. (2017, Feb 28th). Client-side vs server-side rendering: why it's not all black and white. Retrieved from https://medium.freecodecamp.org/what-exactly-is-client-siderendering-and-hows-it-different-from-server-side-rendering-bd5c786b340d
- Wang, P., Mathieu, R., Ke, J., & Cai, H. J. (2010, August). Predicting criminal recidivism with support vector machine. In 2010 International Conference on Management and Service Science (pp. 1-9). IEEE.
- Wang, X., & Brown, D. E. (2012). The spatio-temporal modeling for criminal incidents. Security Informatics, 1(1), 2.
- Watson, H. J., Wixom, B. H., Hoffer, J. A., Anderson-Lehman, R., & Reynolds, A. M. (2006). Real-time business intelligence: Best practices at Continental Airlines. *Information Systems Management*, 23(1), 7.
- Wickramasuriya, R., Ma, J., Somashekar, V., Perez, P., & Berryman, M. (2013). SMART Infrastructure Dashboard: A Fusion between Business Intelligence and Geographic Information Systems.
- Williams, M. L., Burnap, P., & Sloan, L. (2017). Crime sensing with big data: The affordances and limitations of using open-source communications to estimate crime patterns. *The British Journal of Criminology*, 57(2), 320-340.
- Xia, H., & Rao, Q. (2009, December). Flood Level Prediction on the Basis of the Artificial Neural Network. In 2009 First International Conference on Information Science and Engineering (pp. 4887-4890). IEEE.

- Yigal, A. (2018 June 6). *Grafana vs Kibana: The key difference to know*. Retrieved from https://logz.io/blog/grafana-vs-kibana/
- Zhu, K., & Zhang, J. (2012). Predicting the potential locations of the next crime based on data mining: A case study. *International Journal of Digital Content Technology and its Applications*, 6(20), 574.