IN-DEPTH ANALYSIS OF TEXAS ACCIDENTS USING DATA-MINING TECHNIQUES AND GEO-STATISTICAL ANALYST TOOLS

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Faiz ul Islam

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Construction Management and Engineering

March 2018

Fargo, North Dakota

North Dakota State University
Graduate School

Title

In-Depth Analysis of Texas State Accidents' Dataset Using Data-Mining Techniques and Geostatistical Analyst Tools

By

Faiz ul Islam

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Eric Asa
Chair

Dr. Matthew Stone

Dr. Stephanie Day

Approved:

April 6th, 2018	Dr. Jerry Gao
Date	Department Chair

ABSTRACT

Traffic accidents have been a consistently growing problem in the United States. The road-safety issues have not been completely resolved and pose danger to people driving on the roadways. This research used various approaches and techniques to evaluate and analyze the Texas State traffic-accident dataset profoundly and meticulously. Data-mining techniques were used to analyze the accident dataset for Texas statistically, and information were collected. The resulting information from the analysis suggested that the city of Houston, Texas, was the point of persistent accidents and accounted for most accidents in all Texas cities. Therefore, Houston was analyzed further by using the geostatistical and geo-analyst tools in ArcGIS. The Geostatistical Analysis tools including Space-Time identified the key hotspot locations within the city to study the overall behavior, and developed prediction maps from the kriging tool. A similar approach can apply to other parts of Texas and any location in the United States.

ACKNOWLEDGEMENTS

I would like to give my thanks to Dr. Eric Asa, my adviser, for his genuine support and effort in reshaping my thoughts and for his help with working towards my thesis. My special thanks and love go to my parents and siblings for all their prayers, motivations, and continuous mental and financial support to achieve my goals and dreams. My thanks go to my committee members, Dr. Matthew Stone and Dr. Stephanie Day, for helping and providing timely guidance whenever I needed it. I thank Ingrid Scarski and Ann Denny for listening and resolving some of the major issues in the department and always standing up for all the international students and me. I am grateful to the faculty of the Department of Construction Engineering and Management and NDSU for giving me the honor and opportunity to complete my graduate degree at a prestigious institution. I am very thankful to all my friends who always stood behind me and cheered me up during any hardship. Finally, I am grateful to anyone who helped me to complete the degree with ease.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

1. INTRODUCTION

1.1. Background

The United States has one of the largest roadway infrastructures in the world; this system is constructed with modern technological tools and techniques. The roadway infrastructure accommodates billions of vehicles per year, and numbers are on the upsurge. According to online statistics, about 261.8 million vehicles, including cars and light trucks, were registered in 2016. With the advancement of the roadway system and vehicles, it is predictable that the risks for roadway traffic safety have been amplified severely over time, which is reflected with the statistical facts issued by the Fatality Analysis Reporting System (FARS). According to FARS, about 6.2 million crashes were reported to the police in 2015. In assessing the previous year, road crashes increased by 3.8 percent from the last year. Of the total crashes in 2015, 35,092 people died, and 2.44 million people were injured. Similar to the crashes in 2015, fatalities and injuries also increased. In 2015, 96 people died daily from U.S. traffic accidents, up from 90 people per day during 2014.

The accident statistics from the National Highway Traffic Safety Administration (NHTSA, 2016) website stated that almost 32,675 fatalities were reported in the United States during 2014. Of those 32,675 deaths, around 10.8% of the fatal accidents were in Texas, the highest percentage for any U.S. state. Comparing with the previous year's fatalities, the positive percentage change of 4% was relatively significant. However, most states indicated a significant percentage decline for fatal accidents in 2014 than in 2013. From all the fatal crashes in Texas in 2014, there were 3,538 people killed. Knowing that Texas is the second-largest U.S. state by population, the statistical figures are still significant when equated with all other U.S. states.

The U.S. Department of Transportation (USDOT) is estimated to have a budget of $98.1 billion for fiscal year 2017. According to the USDOT website, the transportation budget is targeted to support and to complete infrastructure projects; to rehabilitate roads, bridges, transit systems, railways, and the aviation system; to make safety improvements; and to perfect the overall budget's spending practices. By observing the statistical facts and figures, the goal to improve and to enhance roadway safety still needs to be accomplished. There is also a need to invest the budget in the core, identified issues for transportation safety which can be verified with the FARS and NHTSA's annual reports.

1.2. Problem Statement

Traffic accidents pose a great concern and threat to road safety. This anxiety is heightened in the United States where roughly 17,000 traffic accidents took place daily in 2015. With such a high number of accidents, 96 people died every day. Knowing the traffic accidents' severity and importance, the U.S. Department of Transportation is projected to spend about $98.1 billion in fiscal year 2017 in order to improve the overall transportation system; this figure also includes money to enhance and ameliorate the safety. Despite spending nearly $94.7 billion on the transportation budget in 2016, safety issues and problems still exist. Traffic accidents have displayed a positive trend for a decade, and it is expected that the trend would continue. Moreover, the traffic accidents' consequences are long-lasting and could easily take lives. Many people die from the accidents yearly, and if the accident does not result in death, injuries may cause a person to live a handicapped life. Researchers have developed various techniques and tools for road safety with in-depth study about the foremost reasons for the accidents. The responsible authorities have succeeded in adopting and implementing these methods and practices for their transportation systems, but these officials are not able to control the U.S.

traffic-safety hazards. Therefore, there is a necessity to present transportation authorities with novel techniques so that understanding the main problems and realistic models can serve as a source to counter the traffic-safety threats.

1.3. Research Questions

The research is intended to address the following questions:

i. What is the trend for traffic accidents in the state of Texas?

ii. How has the trend changed statistically through time at the Texas state?

iii. In comparison to total number of accidents, what is the rate of injuries and deaths for each year?

iv. What regions or locations account for the record number of accidents?

v. What percentage of accidents have occurred at intersections? Also, identify the factors that contributed to these accidents.

vi. Does the urban and rural population have any effect on the accidents?

vii. What factors have caused the accidents to take place by population group?

viii. Does the average daily traffic have any influence on the accidents? If yes, what are the factors that have caused numerous accidents to happen?

ix. Do the accident data provide a more holistic picture when they are analyzed visually by geographical representation?

x. What geostatistical analysis tools are appropriate to identify the hotspot locations for accident-count data for a defined region?

xi. What kriging methods can be implemented on the accident-count data for a significant outcome from the hit-and-trial method?

1.4. Aims and Objectives

The study's goals are as follows:

i. Review the literature to identify what has been done and what needs to be studied on road accidents, specifically in Texas.
ii. Display the trend, statistically, for road accidents, injuries, and deaths in Texas.
iii. Identify the accident factors that have a severe effect on road accidents.
iv. Investigate the intersections to detect the major causes of road accidents.
v. Create maps to represent the accidents for each year.
vi. Develop geostatistical hotspot maps for each year's accident data to graphically represent the accident growth.
vii. Establish a prediction map for a specific region that would be applicable to the overall state.

1.5. Research Methodology

Figure 1.1 provides the steps performed for this research. The research activities were conducted in a sequential manner. The processes used for the data analysis, such as data mining and geostatistical analysis, are also labeled in detail and in systematic order in Figure 1.1.

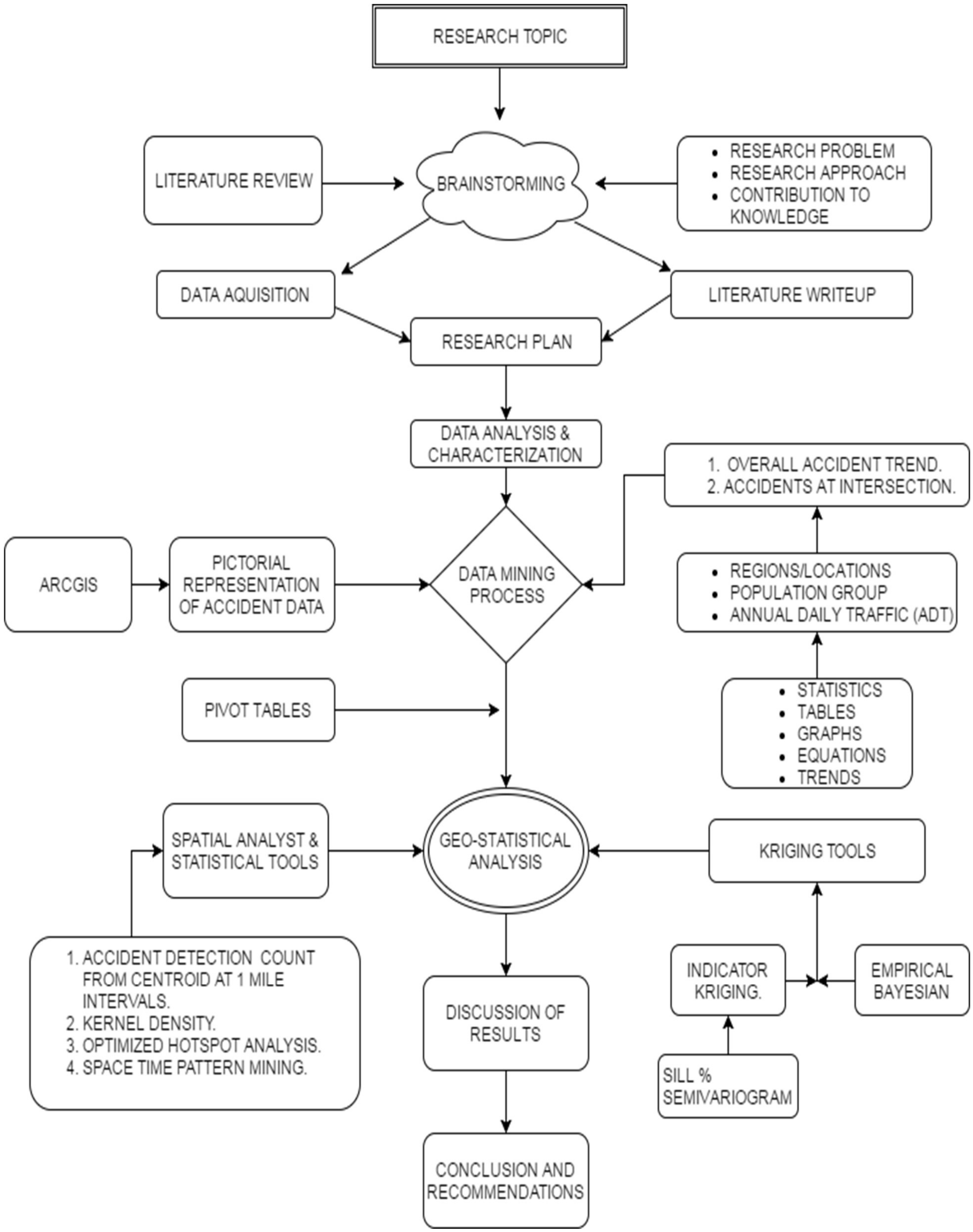


Figure 1.1. Research Methodology Flowchart.

2. LITERATURE REVIEW

2.1. Introduction

Traffic accidents are a global phenomenon and cause severe concerns about the safety of individuals and government property. In Japan, there has been a report of 800,000 road accidents annually (Matsuzaki, Nitta, & Kato, 2008). With the expanding socio-economic growth around the world and an emphasis on building infrastructures, especially in developing countries and with population growth, there has been a notifiable increase in road-traffic coincidences with the yearly upsurge in traffic volume. According to the World Health Organization (2004), road accidents are the source of 1.2 million worldwide deaths annually (Peden et al., 2004). Millions have suffered from death-defying injuries with some permanent disabilities. In Europe, about 40,000 people die from road fatalities on a yearly basis (Shen et al., 2008). Moreover, the accidents are increasing drastically every year with significant numbers in Europe. With this much chaos on the roads and the uncertainty about traffic protocols and procedures, high responsibilities have been put on the law-enforcement agencies as well as the transportation department to lessen the incidents that are growing at a rapid pace. The facts and numbers show the obvious need to improve the system for the sake of humanity and the socio-economic development for any country. Although tremendous work has been done to enhance road safety and traffic operation, a gap remains, and there has not been notable improvement in the sense of diminishing the road fatalities, posing a significant concern for people and society's well-being (H. Wu, Gao, & Zhang, 2013).

Over the years, the traffic has created the public's immense vexation on deaths and injuries resulting from global road crashes. Apart from fatalities, the traffic issue has "by hook or by crook" aggregated the stream of road traffic and inducted road congestion in the traffic-flow maneuver. Explained another way, traffic accidents can be defined as “the nonrecurring events that cause reduction of roadway capacity” (H. Zhang & Khattak, 2010).

The decreased road capacity may not have big effects on the roads in rural zones, but the traffic in urban areas is influenced by the catastrophe of traffic accidents. The traffic flow in urbanized areas is immense, and a minute disturbance on the roads could affect the overall chaos of the road traffic. Imagine yourself in an ambulance on the way to a rescue or a hospital when people are stuck in traffic congestion due to a road crash. The loss of lives, in that case, could be multiplied, i.e., the loss of life from the ambulance’s inability to rescue the person and the loss of life from the accident itself. Typically, traffic accidents account for 30-50% of the traffic congestion within a municipal jurisdiction (Kwon, Mauch, & Varaiya, 2006; Ozbay & Kachroo, 1999; Skabardonis et al., 1995). Accidents are not limited to crashes with vehicles, but may include abandoned or broken vehicles as well as road debris (H. Zhang & Khattak, 2010). Road debris is a form of road hazard that is on or off the road. The debris includes materials, objects, or substances that exist on the traffic path and cause a flow disruption and is against the typical street atmosphere. The first accidents not only cause traffic-flow delays, but are also responsible for secondary incidents due to the distraction and traffic backups from the original accidents.

The condition of traffic accidents in the United States is not any different than other countries. With the growing U.S. economy as well as closing the gaps between people and cities, there has been a significant and noticeable rise in road crashes. According to the statistics, 35,092 people lost their lives from accidents in the United States during 2015, compared to the 2014 figure which was 32,744 people (National Highway Traffic Safety Administration, 2016). Additionally, the increase for traffic crashes between 2014 and 2015 was 7.22%, the largest in 50 years of history for U.S. traffic accidents. According to the National Highway Traffic Safety

Administration (2016), the most substantial increase for traffic crashes was observed from 1965 to 1966 with an estimated percentage upsurge of 8.1%. The increased percentage of accidents is alarming and devastating for any country's transportation authority, specifically for a country like the United States which spends billions of dollars on road safety and improving the transportation system. According to the USDOT, the requested budget for fiscal year 2017 is $98.1 billion. In contrast, the amount spent in 2016 was $78 billion. The amount that focused on the federal highway administration alone was $42.7 billion. The difference in the amount requested for 2017 and the money spent in 2016 is enormous, illustrating the need to improve the nation's transportation system. The question about the credibility on the traffic safety system is still undefined and has all sorts of concern.

Besides, the deaths from road accidents are not the only worrisome thing for the responsible authorities. The crashes' injury consequence is the primary and foremost issue for the transportation authorities to address. The National Highway Traffic Safety Administration (2016) stated that the injuries during 2015 were felt by 2.44 million people, compared with about 2.34 million people in 2014. Some people were disabled from the severe injuries. The cost for rehabilitating the people who require medical treatment and special care exceeds billions. Hospitalization and rehabilitation are not the only expenses that authorities need to consider. The property, wage, and productivity losses contribute immensely towards the authorities' expenditures. The Association for Safe International Road Travel states that road crashes cost the United States $230.6 billion per year, an average of $820 per person. The crashes are not the loss of an individual, but the overall loss of the people affected by the accidents and the respective authorities.

Knowing the number of accidents, it is imperative to say that the United States is one of the deadliest places to drive when compared to the world's traffic reports. Alarmingly, the numbers as well the transportation department's budget are increasing yearly. The traffic concerns are still in place to provide a better environment for drivers and families to travel in a safe manner and without distress. There are many things that federal and state officials and transport agencies have done to improve the overall safety and stricken the traffic laws. However, the results have not been achieved, and better plans are required as the growth in road crashes is significant. Additionally, researchers have done various studies, analyzing the factors involved with fatalities in order to come up with a system to improve the public's road safety for the future.

2.2. Analysis of Traffic Accidents and Safety

As stated earlier, the academic participation to develop an improved road-safety structure to reduce crashes has been enormous. Universities have conducted a plethora of research on various subjects that are linked to the road's traffic accidents and safety. Every state's department of transportation has provided universities with research funds to conduct studies to help improve the overall structure of the traffic system. In an accident, there are multiple contributing factors which account for a crash's risk and unforeseen mishaps. These factors include, but are not limited to, vehicle design, road design, road environment, the driver's speed, weather conditions, light conditions, and the driver's demeanor. Road crashes are the consequence of multiple factors and scenarios that cannot be avoided at the time of performing analysis and cannot be incorporated to yield an improved product to reduce accidents.

Accident analysis is often done to examine crashes and to prevent similar incidents in the future. The study aids the understanding of factors which are related to a traffic accident (Kumar

& Toshniwal, 2016). The best explanation is provided by Bhalla, Tripathi, and Palria (2014): "Accident analysis is carried out to determine the cause or causes of an accident or series of accidents so as to prevent further incidents of a similar kind. It is also known as accident investigation" (Bhalla et al., 2014).

With plenty of statistical techniques available, all sorts and sizes of crash-data analysis can be performed. One outcome could be different and might be not as accurate as another, but each method provides its benefit. With the growing academic and spreading technology, the smooth race of betterment in a traffic system is on the role. One of the most common ways to analyze data is by performing linear models (LM). This method is popular because of the benefits it provides. This technique transforms the data into a linear form, hence the means and variances could easily be derived from the linearly modified data (Oppe, 1992). However, this technique cannot be applied if the crash data do not follow the linear-model trend and if the utilized parameters have no direct relationship. Various researchers have used this approach to build models for analyzing the traffic accidents' factors (Akoz & Karsligil, 2010; Greibe, 2003; Haque, Chin, & Debnath, 2012; P. Y. Park, Miranda-Moreno, & Saccomanno, 2010). P. Y. Park et al. (2010) used multiple linear models to evaluate the rural highways' speed factor that is responsible for more accidents and stated that the traditional speed-differential measure is not useful with the early design phase for road construction. Greibe (2003) examined the urban-junction and public-road links to produce accident-prediction models. These models were intended to precisely forecast traffic accidents at the road junctions and links. Haque et al. (2012) utilized the log-linear model to study motorcycle crashes in Singapore. The research used various factors, such as environment and roadway characteristics, to determine what caused vehicle accidents at different locations.

Sequential binary logit models are another method for analyzing accident severity. Various factors were identified in the research and were used to determine the accidents' severity. Nassar, Saccomanno, and Shortreed (1994) identified the factors as accident dynamics, seating position, vehicle condition, vehicle size, etc. To predict road fatalities in China, Qing and Zhongyin (2015) analyzed crash data by performing a co-integration analysis. The statistical property used to create a model helps to provide a more reliable and precise method in determining the probability of road accidents. The other statistical technique, quantile regression, was examined to develop a methodology to estimate the crash rate of recurrences (H. Wu et al., 2013). Also, this statistical tool provided a more comprehensive way to study the accident data.

Data analysis is a hectic, long process and requires a longer period, specifically for crash data, because the road-accident volume rises yearly. Techniques such as data mining and cluster analysis have been used to study the factors and conclusions (Kumeta, Miyake, & Ogawa, 2006; Rui, Zhaosheng, & Maolei, 2010; Shanthi & Ramani, 2012). These studies examined aspects such as accident frequency and the crash's attributes to produce a novel methodology for predicting road fatalities. The studies showed that sufficient data-analysis research has been done for road crashes. The need is to utilize these techniques in the most effective manner.

2.3. Tools to Improve Road Safety and to Evade Accidents

As with earlier discussion, the Literature Review highlighted the data-analysis practices for road accidents that were applied in the academic research. This section discusses the tools and models which were developed to improve the overall road safety by reducing accidents and mishaps on the highways. The problem to minimize traffic accidents has existed for decades, and several models and tools have been developed to counter that problem. The issue is gaining importance with the passage of time; hence, there is a need to develop more techniques and tools by the aid of latest technology. For that reason, researchers have utilized the latest tools and technology to create a better model for an improved road environment.

Recent applications, such as Adaptive Neuro-Fuzzy Inference System (ANFIS), have been applied in the field of accident data to anticipate the uncertainty and unpredictability of accident data (Hosseinpour, Yahaya, Ghadiri, & Prasetijo, 2013). In the research, ANFIS was utilized to create a model by using the identified indicators; later, ANFIS was compared with the Poisson, negative binomial, and non-linear exponential regression models. The results demonstrated that the ANFIS model provided more accuracy and exactness than the other models. Hence, the ANFIS model can be used by transportation authorities because of its ability to provide better prediction for enhanced road safety, negating indecision with traffic data (Hosseinpour et al., 2013).

Technology, such as video cameras and photographs, has been used to record crash data in recent years because accurate accident data are a necessity for building realistic prediction models. However, the current practice of recording the accident scene is old fashioned and does not satisfy today's demand (Z. Guo, Shang, Wang, & Sun, 2000). In order to resolve the problem, (Z. Guo et al., 2000) used the photogrammetry and computer-vision techniques to capture data using video-camera tools that are integrated with the system. With this method, the data-recording technology was improved and provided accuracy which was matched with the practical measurement requirement. Moreover, technology tools were applied to regulate the vehicles' maneuvering and speed. For research conducted in Japan, (Matsuzaki et al., 2008) developed an intelligent traffic-light system. Because the rate of accidents for pedestrians and vehicles at a blind intersection is comparatively high, the researchers limited their study to that particular case. To verify the results and effectiveness, the researchers experimented with the intelligent traffic-light system at a blind intersection and installing sensors and receivers on a pedestrian and used a mobile robot car. The result displayed the positive output, and the system was entirely applicable to install for a real-time situation (Matsuzaki et al., 2008).

Apart from using technology to predict road accidents or models to reduce the crashes, several other tools and systems were developed with the passage of time and by using alternative methods. A traffic-accident prediction system based on fuzzy logic was proposed by Driss, Saint-Gerand, Bensaid, Benabdeli, and Hamadouche (2013). From the study, the researchers investigated to measure influences on accidents from the local road network. This system offered prediction of risk exposure for road crashes and an analysis of the complex factors involved (Driss et al., 2013). This system helps to identify risk factors for the highways and is fully applicable as a road-safety tool. Furthermore, the Transport Research Laboratory (TRL) developed the SafeNet tool that models the risk of a traffic accident and estimates the injuries per year for a particular road network (Basbas, 2005). This tool was tested at the researcher's local premises and suggested that the tool is valuable for the traffic engineers' work. Another model to predict traffic accidents, developed by Q. Wang and Liu (2009), is known as the GNN forecasting model. This model is a combination of two models built on the Grey prediction model and ANN. With this tool, the model provides more precision for predicting traffic accidents and a simple tool that is practically applicable (Q. Wang & Liu, 2009).

2.4. The Empirical Bayesian (EB) vs. Full Bayesian (FB) Method and Their Application on Traffic-Accidents' Data

The Bayesian method is a statistical inference tool of combining prior and current information in the form of data to describe an event's probability. With the Bayesian method, there are various approaches to combine and analyze information from the given data. Two approaches are Full Bayesian and Empirical Bayesian, different methods of combining the prior and current information. With Empirical Bayesian, the prior distribution is calculated from the given data, whereas the Full Bayesian (FB) uses the approach of fixed prior distribution without observing the current data.

From a research perspective, the Bayesian approach for analyzing the data is not novel. This method has been used to analyze data in all application categories, including traffic (Carriquiry & Pawlovich, 2004). The earliest work of applying the Bayesian method to analyze traffic safety started with Hauer and others (Hauer, 1986, 1996a, 1996b; Higle & Witkowski, 1988; B. N. Persaud, 1988). These researchers utilized the Empirical Bayesian (EB) method to analyze the crash data and consider Empirical Bayesian an advantage over other traditional statistical-analysis approaches. The EB approach is well accepted for analyzing traffic data (Carriquiry & Pawlovich, 2004).

In contrast with the advantages for the EB approach, various authors have argued for the FB method to analyze data. Studies that analyzed the traffic-crash data and safety evaluation by comparing the FB and EB methods favored FB over the EB approach (Carriquiry & Pawlovich, 2004; B. Persaud, Lan, Lyon, & Bhim, 2010). They argued for the FB method and concluded that the FB approach has more advantages than EB method. Also, the authors stated that the FB method is more practical and is a more realistic approach for the data process.

As mentioned earlier, the EB method is an acceptable approach to analyze the traffic-crash data. Recent research work has been done to study the traffic-crash data and to model them with the EB method (Azizi & Sheikholeslami, 2013; Huang, Chin, & Haque, 2009; Schubert & Wanielik, 2011; Srinivasan, Ullman, Finley, & Council, 2011; Zhou, Zhao, Hsu, & Huang, 2013). These studies were conducted to observe the safety effects of some security applications and the crash factors with the EB method. The results stated that the EB method is useful for identifying crash factors and helping to evaluate the safety techniques. The procedure is successfully applicable and acceptable for traffic-data evaluation with some limitations.

2.5. Geographic Information System (GIS): An Effective Tool for Traffic Data

With more development in the field of technology, the geographic information system (GIS) has become a powerful tool to visualize and analyze the data collected from the world graphically. GIS has gained enormous acceptance in all sort of fields and has been part of almost every type of educational study and research. GIS usage is increasing with the passage of time and has been in demand. Likewise, there have been numerous recent studies that utilized GIS to analyze and to manage the accident data as well as to provide geographical information (Y. Chen, Liu, Wu, & Sun, 2011). The advantage of using GIS is that it has the capability for analyzing and processing a huge amount of data with ease (Durduran, 2010). The data are managed and analyzed with various software, such as ArcGIS which is a software and GIS tool that performs numerous functions with the information collected from the GIS. ArcGIS has a framework that works with the world's map and geographical information.

The GIS' uniqueness is that it has the capability to provide geographic information along with all the other information to perform a task. For this reason, a GIS is the most efficient tool for managing, organizing, and analyzing the traffic data. Bhalla et al. (2014) used a similar approach to perform a traffic-accident analysis of Ajmer City (India) and suggested that applying a GIS is a useful tool to produce a road-accident database system. Moreover, the factors that account for the traffic accidents are also identified using the GIS, hence helping to improve the traffic conditions (Y. Chen et al., 2011). The traffic-crash data are complex with all sorts of uncertainties and variables that cause accidents. With the increasing number of accidents per year, the data are getting larger, causing difficulty with managing the database, and performing the analysis is very challenging. The GIS with the ArcGIS tool makes it easier to visualize the data and to form a better understanding of the database, hence improving the data's overall organization, and the information is utilized in the most effective manner (Jinlin Wang, Chen, Zhou, Wang, & Zhang, 2008). Another advantage GIS provides, is forming prediction models and decision-making system from the accident database evaluation, which could help in taking precautionary measures in dangerous conditions (Durduran, 2010). Overall, the benefits of using the GIS are evident in the traffic data from previous research, and it is one of the best techniques to analyze the data.

2.6. An Overview of Traffic-Accident Factors

The trend for road accidents in Texas has grown in recent years. However, the accidents can restrain or confine the proper analysis of the factors that contribute to the traffic fatalities, and on that basis, the prediction models developed (B. Chen, He, & Wang, 2011). The evaluation of traffic-accident factors is an important subject and must be taken seriously to overcome the problem with traffic fatalities. Various studies were conducted to find the factors that contribute to road accidents. X. Li, Lord, and Zhang (2010) studied the application of generalized additive models to conduct research on the crash-modification factors. Likewise, the implication of the binomial regression models was used to construct an accident-modification factor and crash-prediction models for the horizontal curves in the United States (Fitzpatrick, Lord, & Park, 2010; Knecht, Saito, & Schultz, 2016). An algorithm, such as the support vector machine (SVM), is employed to study the classification of the traffic condition involve striking before the accident occurrence (Qu, Wang, & Wang, 2011). Furthermore, there are possible factors for sideswipe accidents in the off-peak hours when vehicles are traveling in a straight line on a multilane highway (Jiangfeng Wang, Zhang, Wang, Weng, & Yan, 2016). The traffic accidents have massive socio-economic consequences, harming road productivity and increasing Medicare expenditures (Naumann, Dellinger, Zaloshnja, Lawrence, & Miller, 2010). Therefore, the value linked to the crashes demands a deep understanding and a model for cost estimation (Hancock, Zhang, Sardar, & Wang, 2016).

In recent years, there have been various studies to analyze the database for traffic-accident severity. The goal for all the studies is to help control road accidents by providing possible solutions. Statistical tools and methods are used and considered to assess and to analyze road accidents and fatalities (Savolainen, Mannering, Lord, & Quddus, 2011). Examples of the multiple statistical methodologies and models which are employed for the analysis are Poisson regression (Guohui Zhang, Zheng, & Wang, 2012), the simple micro approach (Medina, Shen, & Benekohal, 2014), multiple logistic regression (Guopeng Zhang, Sun, Lou, Xu, & Jiang, 2013), the logistic-regression model (R. Chen, Zhang, Li, & Wang, 2012; MacLeod, Griswold, Arnold, & Ragland, 2012; Schultz, Farnsworth, & Saito; Tefft, 2013), the Bayesian model (Huang & Abdel-Aty, 2010; J. Ma & Li, 2010; Schultz, Black, & Saito, 2014; Schultz et al.), the multinomial logit model (F. Chen & Chen, 2011; H. Chen, 2014), the bivariate Poisson-lognormal model (X. Ma, Chen, & Chen, 2016), the multivariate Poisson-lognormal model (Bai, Liu, Li, & Xu, 2011), and quantile regression (H. Wu et al., 2013).

2.7. Road-Accident Factors

2.7.1. Age and Gender

Age and sex are factors which are connected with fatal accidents and are related to each other. Female drivers follow the same accidental behavior in driving skills parallel to males. However, a study suggests that females are more likely to be in an accident than males due to maneuvering skills when speeding (Kelley-Baker & Romano, 2010). The behaviors and factors that cause crashes vary with age groups and sex (Hao, Kamga, & Daniel, 2015). It has been observed and investigated that older people and teens are prone to more fatalities than middle-aged individuals (Alam, 2011; Masten, Foss, & Marshall, 2011). Teens with learner permits are more likely to have accidents than adults with a similar vehicle. Also, teens are the primary reason for severe crashes (Curry, Hafetz, Kallan, Winston, & Durbin, 2011; Lee, Simons-Morton, Klauer, Ouimet, & Dingus, 2011; Simons-Morton et al., 2011).

2.7.2. Construction Zones

Highway construction zones are hotspots for predictable road accidents. The recurrent incidents at intersections demand thorough consideration to improve safety measures (Elghamrawy, El-Rayes, & Liu, 2010; Higa & Kim, 2013; Pulugurtha & Nujjetty, 2011; Y. Zhang, Zhu, Wang, Hu, & Liu, 2011). The characteristics of work-zone areas are a vital source of information for transportation authorities to measure construction-zone safety and traffic management (Akepati & Dissanayake, 2011). The findings suggest that nighttime work in construction zones provides less visibility for the workers and creates hazardous conditions (Valentin, Mannering, Abraham, & Dunston, 2010). Z. Wang, Lu, Wang, Lu, and Zhang (2010) model severe crashes in work zones using the ordered Probit regression. The practical implication made by applying the work-zone barrier system in Oregon City to enhance the safety of workers at labor area (Tymvios & Gambatese, 2014).

2.7.3. Freeways and Vehicle Type

To avoid traffic congestion and maintain the traffic flow, freeways were constructed in the United States. The freeways' purpose is to help drivers reach their destination quickly and on time. However, freeways create great concern for traffic authorities because of the recurrent crashes and fatalities (Y. Guo & Sun, 2013). Accidents can be categorised by vehicle type, such as buses (Kaplan & Prato, 2012), trucks, etc., because the risks and factors involved with each vehicle type vary. The presence of large vehicles on the highways causes fear and is responsible for a lot of fatalities. The potential risk for truck accidents on the roads is a primary concern, specifically in construction zones. Therefore, possible risk factors are analyzed for a smooth traffic flow with no impediment on the motorways (Elvik, 2016; Y. Li, Cheng, & Bai, 2012; Vadlamani, Chen, Ahn, & Washington, 2010).

2.7.4. People's Carelessness

Often, people's negligence leads to major accidents. Abandoned or disabled vehicles on the highways may not pose a danger. However, abandoned and disabled vehicles create adverse safety conditions on the road and account for 78% of the traffic accidents in Tennessee (Chimba, Kutela, Ogletree, Horne, & Tugwell, 2013). Likewise, long driving routes may cause drowsy feelings and fatigue, particularly when you have not had enough sleep, if you have sleep apnea (Philip et al., 2010; Tregear, Reston, Schoelles, & Phillips, 2010), or if you have taken drugs (Pressman, 2011). Drowsy driving accounts for numerous accidents and fatalities. Hence, the drowsy-driving advisory system was developed to overcome the fatalities origin from the drowsiness (Kang, Momtaz, & Barnett, 2015).

2.7.5. Weather Conditions

Similarly, climatic factors cause some accidents. Weather conditions, such as rain and snow, create hazardous conditions for drivers. The United States has plenty of places which receive lots of rain and snow, depending on the season. Snow creates an unsatisfactory road surface, and visibility is affected (Seeherman & Liu, 2015). Rain causes a slippery road and produces unfavorable conditions for drivers. Thus, adverse conditions from rain cause severe crashes and demand an analysis to identify the accident-prone wet locations (Ye, Shi, Huang, & Wang, 2015). A study suggests that drivers respond differently, depending on their age and gender, to changing road-surface conditions (Morgan & Mannering, 2011). Jung, Qin, and Noyce (2011) used the sequential logistic-regression approach to study crashes that resulted from rain on the high-speed highways. The low visibility and less physically active of a driver may lead to drive in wrong direction. The wrong-way driving may not cause severe damage as other accidents, but it still accounts for crash reports, and majority of these accidents go unreported (Rogers, Al-Deek, & Sandt, 2014). Moreover, Nourzad, Salvucci, and Pradhan (2014) proposed a computational model to examine how driver distraction accounts for accidents. Texting and mobile-phone usage in driving are new issues for the authorities to mitigate and are causes of distraction for drivers (Ige, Banstola, & Pilkington, 2016; Wilson & Stimpson, 2010), especially teens and new drivers (Klauer et al., 2014). Therefore, studies, such as ones that use the computational model to determine the effect of distracted drivers on the extensive road networks (Nourzad et al., 2014), provide wide-ranging safety information to implement.

2.7.6. Alcohol

Consuming alcohol before or while driving is considered a major issue in the United States, and it poses a real concern for traffic safety. Alcohol-related accidents are responsible for a record number of U.S. crashes, and study suggests that people who are drugged while driving create hazardous road conditions (G. Li, Brady, & Chen, 2013). Therefore, researchers focused their studies on this particular subject. Hajizamani, Shrubsall, and Viegas (2011) designed a device which is installed inside the vehicle to keep an alcohol-impaired person from driving the automobile, and the procedure was evaluated using agent-based modeling (ABM). However, an incentive, such as tax increments and passing laws which have zero tolerance, should be implemented by the government to control the use of alcohol before and while driving (Chang, Wu, & Ying, 2012). Previous studies suggested that increased taxes for alcohol would help to significantly reduce alcohol-related fatalities (Elder et al., 2010; Wagenaar, Tobler, & Komro, 2010).

2.7.7. Technology

Technology has its advantages and has flourished recently. Using technology, such as sensors (Hallowell, Teizer, & Blaney, 2010) and the Advanced Transportation Management Information System (ATMIS; (Choe, Gordon, & Martinez, 2013), has proven benefits for the operation of traffic flow and the overall safety. The adaptive traffic signal control (ATSC) system has been adopted by various places in the United States. The ATSC generates the crash-mediation factors that were used in the research to study intersection accidents (J. Ma et al., 2016). A study found that using technology such as side-view assist, forward-collision warning/mitigation, lane-departure warning/prevention, and adaptive headlights prevents and mitigates road accidents (Jermakian, 2011). Over the years, increased use of the geographic information system (GIS) has had a significant effect on the analysis of crash data. Geospatial analysis using the GIS provides an easy understanding of the data and predicts areas which are inclined to have traffic accidents (Mehta, Li, Fields, Lou, & Jones, 2015; Pulugurtha & Pasupuleti, 2013). Data about the locations which are prone to accidents provide helpful information that is related economic loss, hence the loss can be estimated (Yang, Lu, & Wu, 2013).

2.7.8. Intersections

Crashes at intersections are of bigger concern for the U.S. transportation authorities. Traffic data at intersections have been studied (X. Wang, Chen, & Sun, 2010), and models have been developed to estimate the crashes at intersections (F. Guo, Wang, & Abdel-Aty, 2010; Pulugurtha & Nujjetty, 2011). One standard practice to mitigate accidents at intersections is to install signals (Shams & Dissanayake, 2014). However, accidents at signalized intersections comprise a large percentage of the crashes (Haleem, Gan, & Alluri, 2014; Xu, Teng, Kwigizile, & Mulokozi, 2014; Guopeng Zhang et al., 2013). Therefore, systematic studies have been done in this particular area to develop numerous systems, such as the integrated dilemma zone protection system (IDZPS; (S. Y. Park, Lan, Chang, Tolani, & Huang, 2016), the access management technique (Xu, Teng, & Kwigizile, 2011), and the Geographically-Weighted Regression (GWR) technique (Z. Li, Lee, Lee, & Valiou, 2011). Various factors result an incident, thus the factors can be classified for the accidents at intersections as well. Lighting is one factor studied to investigate the accidents at intersections (Zhao, Jiang, & Li, 2016). Moreover, the finding implicates that roadway lighting provides better vision for drivers and helps to reduce accidents (Isebrands et al., 2010). Controlling accidents at intersections will lessen the rate of accidents in the United States immensely. Thus, there is a need to provide enough resources at the intersection, and multiple options are required to improve safety (Mishra & Khasnabis, 2011). Appiah, Rilett, Naik, and Wojtal (2012) used the actuated warning system to investigate the effectiveness on the intersectional accidents.

2.7.9. Planning and Design

The best possible solution to reduce accidents is by incorporating the accident factors while planning and designing highway networks. The roadway's geometric design controls the operational speed; therefore, the relationship between the planning and design requires consistency for traffic safety (K.-F. Wu, Donnell, Himes, & Sasidharan, 2013). Similarly, the design for roundabouts has various factors to control and prevent accidents (Zirkel, Park, McFadden, Angelastro, & McCarthy, 2012). Recent studies have found that shoulder paving had a positive influence on the highways' traffic safety (Z. Li et al., 2013; Z. Li, Lee, Lee, Zhou, & Bamzai, 2011). Research by Anderson and DeMarco (2013) provides a valid tool and accurate information for designing new slopes. During the planning stage, the median is an important aspect of the road's infrastructure and is a dividing line between the two roads with traffic flowing in both directions. The dividing line needs to be prominently visible for drivers to stay on the given pathway and not intersect with oncoming traffic, especially at night. High-tension cable barriers are a way to avoid vehicle crashes on the median, hence the costs and benefits of installing cable barriers are studied in comparison with median-related crashes (Villwock, Blond, & Tarko, 2010).

2.7.10. Pedestrians

The consequence of road accidents is high for pedestrians who are crossing the streets or waiting to cross to the other side of the road. Drivers' inattentiveness as well as broken or poorly implemented laws are the leading causes for pedestrians' crashes (Zegeer & Bushell, 2012). The vast number of these accidents involve children. Previous studies suggested that regular practice and training to cross streets can lead to safer environments (Schwebel, Combs, Rodriguez, Severson, & Sisiopiku, 2016). Schwebel et al. (2016) experimented with cognitive behavior by providing a virtual semi-mobile and semi-immersive environment at schools and community centers. However, pedestrians' sloppiness and negligence, such as texting, talking on cell phones, listening to music, etc., lead to distraction for drivers and accounts for road accidents (Neider, McCarley, Crowell, Kaczmarski, & Kramer, 2010; Schwebel et al., 2012).

2.8. Conclusion

From the Literature Review, it is clear that traffic accidents are a huge concern for transportation organizations with the increment each year at rapid growth. The demand to take action is getting stronger. The authorities have worked on various projects with researchers to study and to analyze the accident database, creating better systems and models to improve traffic safety. These measures are helpful, but there is still a lot to be done. The models and studies must be updated yearly in the database, and more accident factors are required to identify. Also, using technology requires merging the latest studies for a beneficial outcome and better models. Hence, the overall target should be to improve the safety and well-being for a better environment.

3. EXPLORATORY DATA ANALYSIS

3.1. Introduction

This chapter provides a holistic picture of the information that was discovered in the database; this information is discussed in detail. Various methods and techniques were used to analyze the data. The analysis method and technique were based on the hit-and-trial method by first trying different elements and parameters on the data in order to obtain the relationships among parameters. The method provided essential information from the database which was extracted to answer crucial research questions. The results argued and discussed in detail that lead to multiple conclusions. The results are presented in the form of tables, statistical models, and graphical methods. ArcGIS was used to observe trends in the database and to identify the key locations and attributes for the research study. The ArcGIS trend is shown in a separate section, and the significant trends which were mined from the ArcGIS models are elaborated expansively.

3.2. Data Acquisition

The database was obtained from the Texas Department of Transportation (TX DOT) by completing an online request form to provide the database for Texas accident data from 2010 to 2016. The database was huge and consisted of multiple Excel CSV (comma separated value) files for each year, providing geographical information in the form of XYZ coordinates. For analysis purposes, each year’s data were amalgamated into a single Excel file. The crash database provided by the TX DOT was expansive in detail, with different elements and parameters that included a database dictionary to describe the codes and standards in the accident database. Examples of some data parameters and characteristics for the database acquired from the TX DOT are listed in Table 3.1.

Table 3.1. Example Dataset of Accidents from the Texas DOT.

Crash_ID	Crash_Date	Crash_ Time	Case_ID	Rpt_CRIS _Cnty_ID	Rpt_City_ ID	Rpt_Street_ Name
11152318	1/19/2010	10:02 AM	1001140016	165	291	WILLIAMS
11152319	1/19/2010	11:30 AM	1001190017	165	291	ANDREWS
11152320	1/19/2010	4:36 PM	1001190037	165	291	WADLEY
11152327	1/23/2010	1:53 PM	2010-003226	220	31	SHIRLEY
11152328	1/24/2010	10:55 PM	10001529	43	468	COUNTY LINE ROAD
11152331	1/20/2010	3:55 PM	10000768	61	1626	I35E
11152333	1/26/2010	10:44 PM	2010-3912	101	29	ALEXANDER
11152338	1/8/2010	1:44 AM	19072	15	379	IH35N
11152344	1/8/2010	7:41 AM	19408	15	379	FLORES
11152346	1/10/2010	2:04 AM	10-00026	161	442	PARK LAKE

3.3. Accident Trends, 2010-2016

The overview section elaborates the first perspective after looking at the database by compiling the separate files into one Excel file. The overview offers conclusive information that was evident at the beginning and that lead to further research for more obvious conclusions. Table 3.2 provides a summary of the yearly crash data, 2010 to 2016, and the percentage increase for accidents in Texas.

Table 3.2. Percentage Increase for Traffic Accidents, 2010-2016.

Years	Total Crashes	Percentage Increase in Crashes from Prior Year
2010	472298	0
2011	456034	-3.44
2012	495778	8.72
2013	521101	5.11
2014	554934	6.49
2015	598791	7.90
2016	624734	4.33

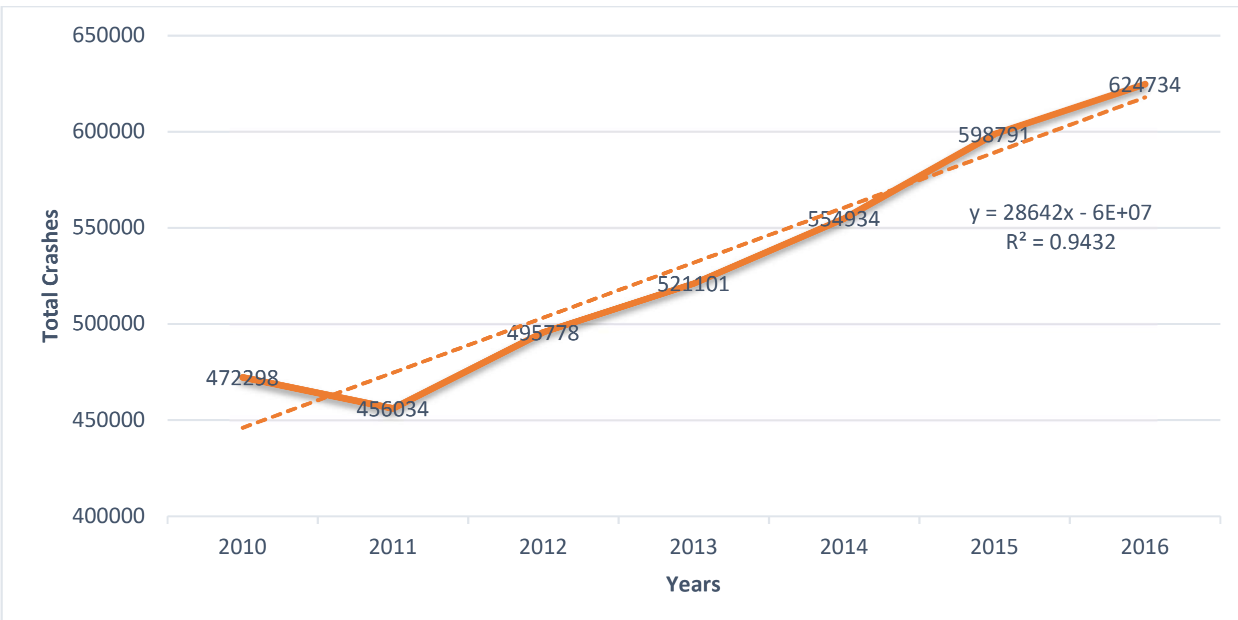


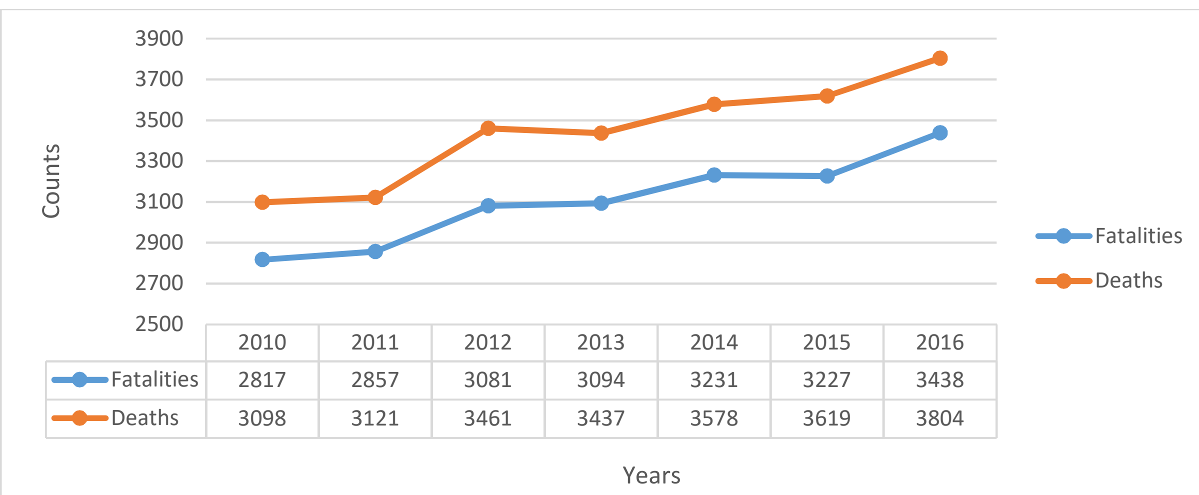
Figure 3.1. Trend for Annual Traffic Accidents in Texas, 2010-2016.

Figure 3.1 represents the trend for traffic accidents in Texas. The data suggest that the number of traffic accidents rose every year except 2011. In 2011, road accidents decreased by 3.44% from the preceding year. Figure 3.1 shows the positive trend for the relationship between the number of total crashes and the year, with an average of 531,953 accidents per year between 2010 and 2016. From Figure 3.1 and Table 3.2, we see the alarming situation in Texas: the increased percentage of traffic accidents is significant. The highest percentage increases for traffic crashes are in 2012 and 2015 with higher increments of 8.72% and 7.9%, respectively. The positive trend in Figure 3.1 provides a significant R^2 value of 0.9432, and a fitted linear equation is provided in Eq. 3.1:

$$Y = 28642x - 6E{+}07 \qquad \text{(Eq. 3.1)}$$

3.3.1. Annual Fatalities, Injuries, and Deaths, 2010-2016

The important aspect for analyzing any crash data is by counting the fatalities, deaths, and injuries for each year. This information illustrates the importance of traffic-safety issues for a state compared to other states and measures the performance of the respective state authorities. However, Texas portrays a traumatic picture of a dangerous U.S. state with the most fatal accidents. Figure 3.2 elaborates the number of traffic fatalities, deaths, and injuries in the Texas from 2010 to 2016.



	2010	2011	2012	2013	2014	2015	2016
Fatalities	2817	2857	3081	3094	3231	3227	3438
Deaths	3098	3121	3461	3437	3578	3619	3804

Figure 3.2. Annual Traffic Accident Fatalities and Deaths in Texas from 2010-2016.

In Figure 3.2, "fatalities" refers to accidents that have resulted in one or more deaths, whereas "deaths" denote the total number of deaths for an individual year. Figure 3.2 has a positive trend with ups and down, and there is a significant increase in deaths and fatalities by

2016. The fatalities dropped in 2015, yet the decrease is insignificant with just four fewer fatalities than the previous year. The numbers also show a decline in deaths for 2013 with a decrement of less than 1%. The highest number of fatalities and deaths is in 2016, with an increase of 22% and 22.8%, respectively.

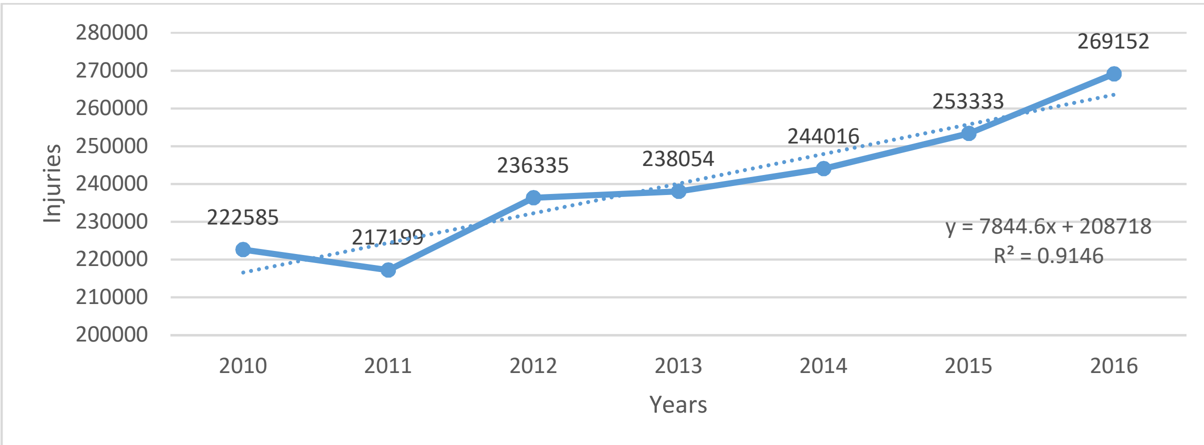


Figure 3.3. Texas' Annual Count for Traffic Injuries, 2010-2016.

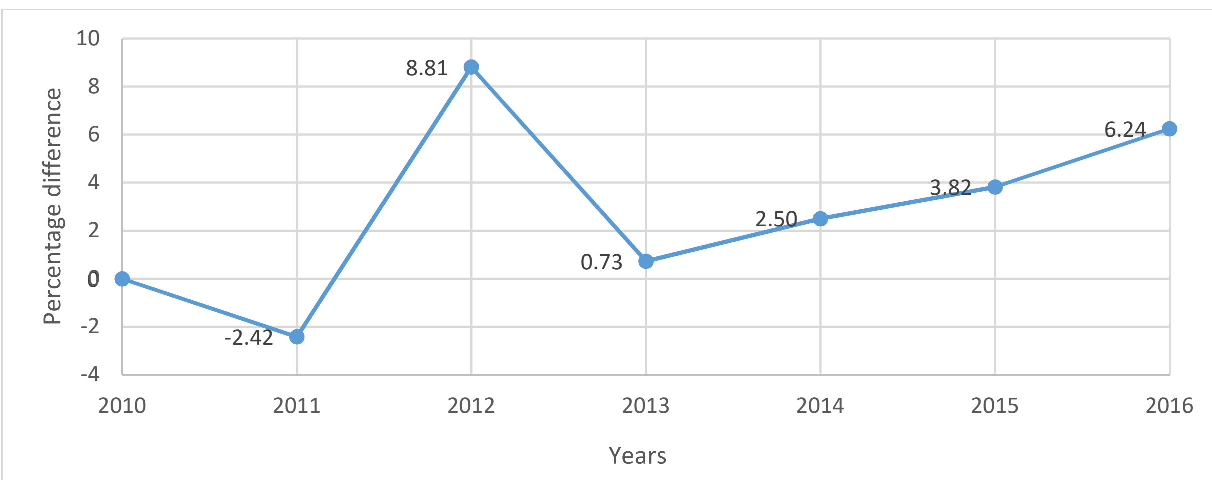
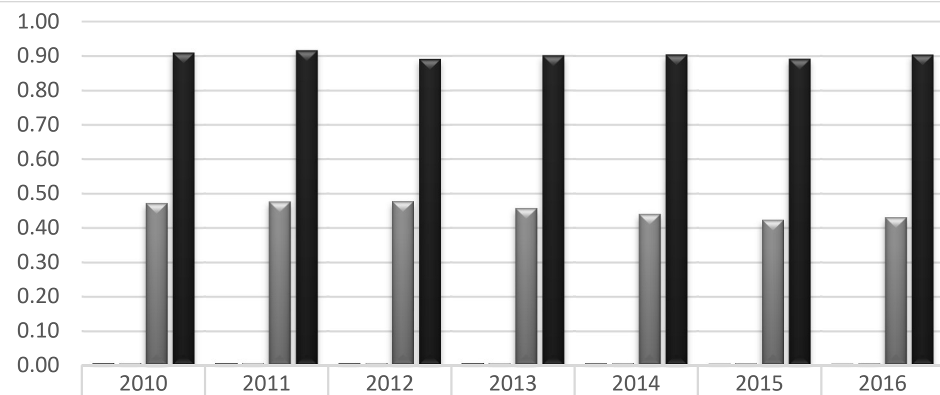


Figure 3.4. Percentage Difference for Injuries by Year, 2010-2016.

Figures 3.3 and 3.4 represent the number of injuries and the percentage of difference for 2010 to 2016. Figure 3.3 is skewed towards the left because it shows the positive tendency for the number of injuries which increased yearly except in 2011. The highest number of injuries is noted in 2016, with a value of 269,152, and the lowest value is observed in 2010 (222,585 injuries). The mean for injuries for the 7-year period is 240,097. A fitted linear equation (Eq. 3.2) is plotted for each year's injuries, and an R^2 value of 0.9146 is attained, a relatively significant finding.

$$Y = 7844.6x + 208718 \tag{Eq. 3.2}$$

Figure 3.4 demonstrates that the greatest percentage increase (8.81%) for injuries was in 2012. In 2011, the injuries from traffic accidents went down 2.42%. The average for the 7-year accident database is a 3.28% increase for an individual year. Despite the decreased number of injuries in 2011, there are more deaths and fatalities than the previous year. A minimal injury increase, with a value of 0.73%, is noted in 2013.



	2010	2011	2012	2013	2014	2015	2016
Fatality rate per crashes	0.0060	0.0063	0.0062	0.0059	0.0058	0.0054	0.0055
Death rate per crashes	0.0066	0.0068	0.0070	0.0066	0.0064	0.0060	0.0061
Injury Rate per crashes	0.471	0.476	0.477	0.457	0.440	0.423	0.431
Fatality to Death Ratio	0.909	0.915	0.890	0.900	0.903	0.892	0.904

Figure 3.5. Various Ratios for Traffic Accidents, Injuries, Deaths, and Fatalities, 2010-2016.

Figure 3.5 provides another perspective to examine the extracted data about fatalities, deaths, and injuries for the 7 years. Figure 3.5, along with statistical numbers, offers the fatality, death, and injury rate (per accident) and the fatality-to-death ratio for an individual year. Irrespective of recordable accidents for 2015 and 2016, the injury rate per crash has values of 0.423 and 0.431, respectively, which is lower than the other years. On the other hand, the injury rate per crash is highest in 2012 with a 0.477 value; the next closest value is in 2011 when it is 0.476. Interestingly, 2011 has a higher fatality rate per crash than all subsequent years and close to the highest number for the death rate per crash. The highest mortality rate per crash is in 2012. To find a conclusive result to determine which year was the deadliest regarding fatalities and accidents, we have calculated the ratio of fatalities and deaths. Irrespective of the smaller number of accidents and injuries, the fatality-to-death ratio suggests that 2011 was the most lethal year in the 7-year period with a prominent value of 0.915; the next-closest ratio is 0.909 in 2010. The results conclusively illustrate that a year with a lower number of accidents has a higher fatality-to-death ratio compared to years with a higher number of accidents.

3.3.2. Analysis by Region and Location Type

It is vital to determine the accident-prone counties and cities in order to implement active safety measures and to protect people’s lives. The purpose is to promote a healthy transportation environment and to safely maneuver on the road. The approach will aid in identifying the key areas and locations where problems and issues exist. Moreover, this in-depth research will lead to conclusive evidence for recognizing the existing problem and the accident parameters for the following places.

Table 3.3. Traffic Accidents, Deaths, and Injuries, Counted by Counties with the Highest Number.

Years	Parameters	Counties						Ratios
		Harris	Bexar	Dallas	Tarrant	El Paso	Travis	
2010	Accidents	84378	45337	42170	29977	17414	15532	0.497
	Deaths	373	153	190	138	62	82	0.322
	Injuries	35770	18547	25174	16348	6418	10647	0.507
2011	Accidents	78025	44229	41378	28979	17456	15712	0.495
	Deaths	387	154	185	155	98	88	0.342
	Injuries	34191	18661	24412	15653	6494	10904	0.508
2012	Accidents	91411	48087	42550	30855	18564	17947	0.503
	Deaths	374	174	212	124	76	109	0.309
	Injuries	40752	20820	24778	15644	7147	12794	0.512
2013	Accidents	101794	48908	46456	33609	18164	17291	0.511
	Deaths	378	191	227	146	114	61	0.325
	Injuries	43126	20250	25472	16391	12153	6827	0.522
2014	Accidents	119044	52252	48626	34034	18026	18395	0.523
	Deaths	425	190	240	150	70	96	0.327
	Injuries	47028	21003	25567	16508	6843	11744	0.527
2015	Accidents	129472	58422	55124	36906	21993	19466	0.537
	Deaths	400	193	265	160	62	147	0.339
	Injuries	47790	23672	28013	17340	7302	11847	0.537
2016	Accidents	127330	63171	62318	41332	22688	21419	0.541
	Deaths	454	224	320	161	83	114	0.356
	Injuries	47612	25512	32400	19783	7659	12810	0.542

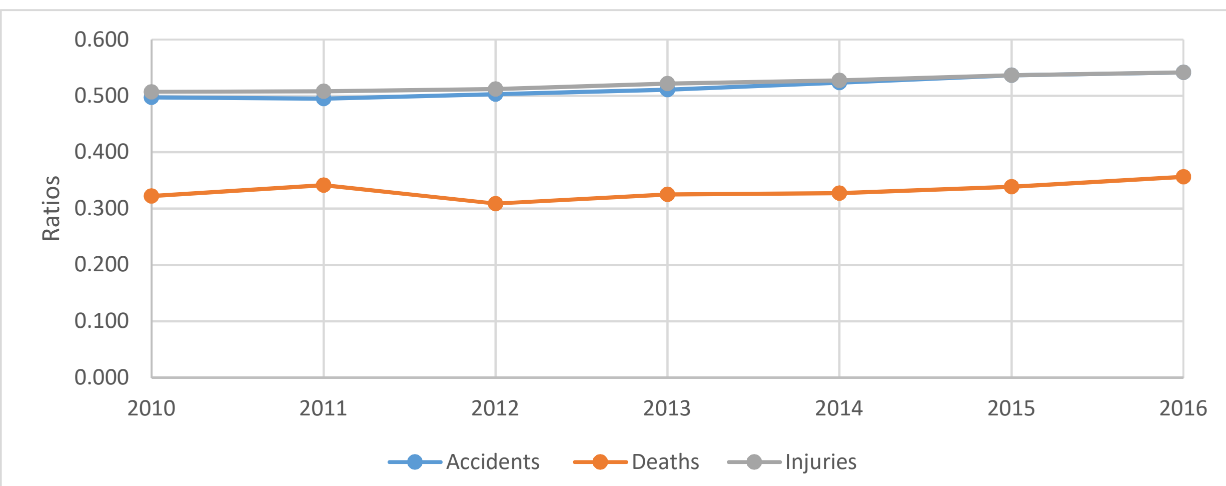


Figure 3.6. Ratios for the Counties' Accidents, Deaths, and Injuries, 2010-2016.

Table 3.3 shows the counties that had the most accidents in the 7-year period, in descending order from left to right, and statistics for the resulting deaths and injuries. Texas has 254 counties; the 6 counties with the most accidents from 2010 to 2016 are given. The designated counties, in the order of the most to least accidents, are Harris, Bexar, Dallas, Tarrant, El Paso, and Travis, respectively. The statistics for accidents, injuries, and deaths reflect the fact that accidents have mostly increased in each county by year, and as the numbers suggest, transportation authorities' proactive measures have failed to cope with the issues and problems. Moreover, deaths and injuries have an increasing trend in each county. The table's important aspect is to identify the portion that these counties contribute towards the overall accidents, deaths, and injuries each year; for this purpose, the ratios are recognized and displayed Table 3.3. Figure 3.6 shows the trend for all years. The rates are calculated by dividing the accidents, death, and injuries of each year with the value of total accidents, deaths, and injuries in the respective year. The values in Table 3.3 and Figure 3.6 show that the 6 counties are responsible for most accidents in Texas, with an average value of 0.52 and a median of 0.51. The ratio provides the value for the standard deviation (0.0174), which is relatively low. The numbers suggest that half of the accidents in the Texas occur in 6 of the 254 counties. The mean and median for the injury rate are 0.52 and 0.522, respectively, endorsing and restating the same facts discussed earlier. The standard deviation for the injury rate is 0.0128. However, the death rate is lower than the accidents and injuries, and displays the mean and median as 0.33 and 0.327, respectively, with a standard deviation of 0.0144. The previous statement expresses the dissimilarity with the accident and injury rate for the 7-year period, illustrating that the death rates in these counties are relatively low for the number of accidents and injuries. However, the mortality of only 0.33 in 6 counties is a significant value and cannot be neglected. Figure 3.6 shows the slight positive increase for all three parameters, although there was a substantial upsurge for the death rate in 2011.

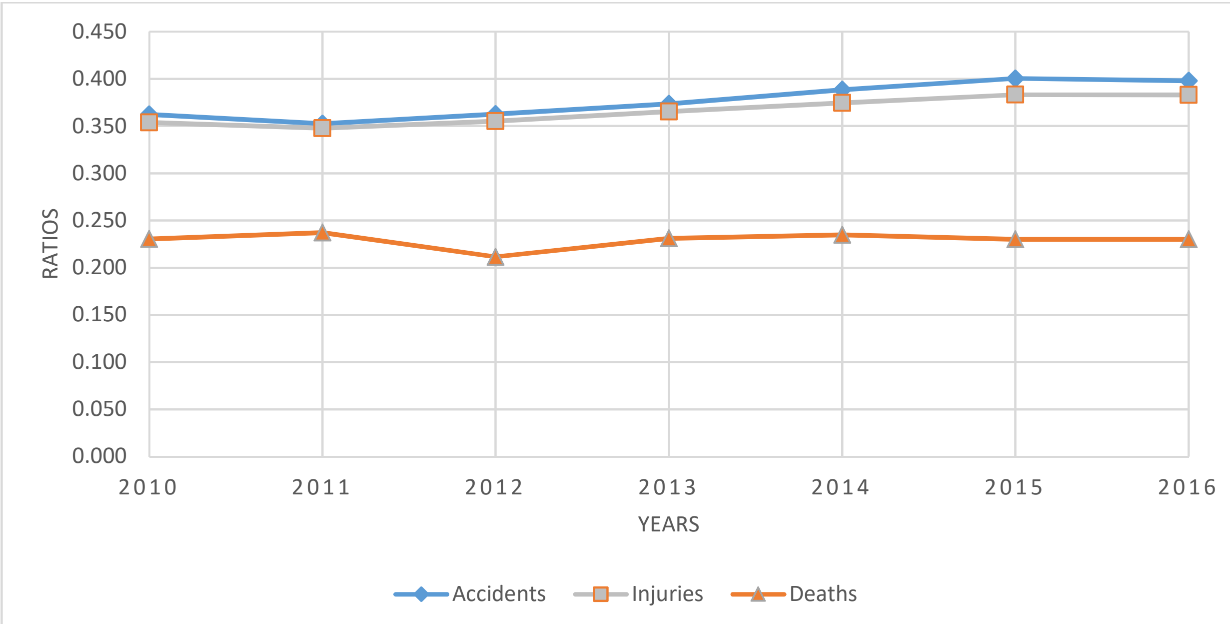


Figure 3.7. Ratios for the Cities Accidents, Deaths, and Injuries, 2010-2016.

Table 3.4. Traffic Accidents, Deaths, and Injuries by City (Descending Order).

Years	Parameters	Cities						Ratios
		Houston	San Antonio	Dallas	Rural Harris County	El Paso	Fort Worth	
2010	Accidents	51412	40597	27675	23122	15355	12993	0.362
	Injuries	24660	16522	16837	7549	5699	7473	0.354
	Deaths	221	120	129	118	52	74	0.230
2011	Accidents	45223	38780	27247	22131	15394	11983	0.353
	Injuries	22753	16242	16359	7503	5747	6922	0.348
	Deaths	212	120	119	143	76	70	0.237
2012	Accidents	54790	42958	27611	25566	16469	12416	0.363
	Injuries	28754	18387	16031	8194	6306	6317	0.355
	Deaths	203	141	141	132	55	60	0.211
2013	Accidents	59855	43342	30112	30715	15308	15344	0.374
	Injuries	29627	17892	16624	9688	6066	7052	0.365
	Deaths	197	170	145	159	50	73	0.231
2014	Accidents	68054	46254	30891	38096	15921	16435	0.389
	Injuries	32563	18702	16349	10230	6058	7492	0.375
	Deaths	238	152	156	158	56	80	0.235
2015	Accidents	74984	52083	34385	40151	19786	18397	0.400
	Injuries	33717	21447	17709	9846	6466	7906	0.383
	Deaths	215	157	176	141	52	91	0.230
2016	Accidents	75319	56244	38840	37169	20329	20752	0.398
	Injuries	33382	23251	20683	9747	6803	9257	0.383
	Deaths	258	196	195	162	69	82	0.253

The cities are an alternative approach to view the data from a different perspective. There are more than a thousand cities and rural areas in Texas. Table 3.4 displays the statistics about accidents, deaths, and injuries for the top 6 accident-prone cities in Texas. Houston has the highest number of accidents for all cities in Texas. It is important to know that these cities are in the 6 counties with a record number of accidents. The statistics reflect how accidents and injuries have increased with a significant increase for every city, yet the death figures depict ups and downs for the trend. In general, Table 3.4 suggests that there is a rise for all 3 parameters in the cities' jurisdiction. The accidents, injuries, and deaths in the six cities are added to determine the rate of accidents, deaths, and injuries for each year. The mean and median for the accident and injury rate are 0.37 and 0.36, with a standard deviation of 0.0175 and 0.0134, respectively. The statistics explain that these cities account for about 37% of the accidents and injuries in Texas. As expected, the death rate for these cities is relatively low compared to the number of accidents and injuries. The mean and median for these cities are calculated as 0.23 with a standard deviation of 0.011. Therefore, 23% of all the traffic deaths in Texas occurred in these major cities. Figure 3.7 illustrates the significant positive trend for accidents and injuries as each year progressed; the mortality figure reflects a slight increase for each year with a significant drop in the death rate in 2012. Texas is the second-largest U.S. state in terms of size, yet the six cities and counties are the epicenter of the accidents. This information can help the authorities to focus their efforts on improving safety for these places.

3.3.3. Analysis by Population Group

This section provides more thorough information about the deaths, fatalities, and injuries that result from traffic accidents as they correspond to the population. Studying and identifying accidents with the population lead to substantial, conclusive evidence because the road structure varies and depends on the population group of the corresponding area. From the statistical analysis, we conclude that the cities and counties have higher injury and accident rates with relatively lower death rates. This evidence recommends further study about changing trends in the less-densely populated zones in order to evaluate the finding that the less-populated zones have higher death ratios.

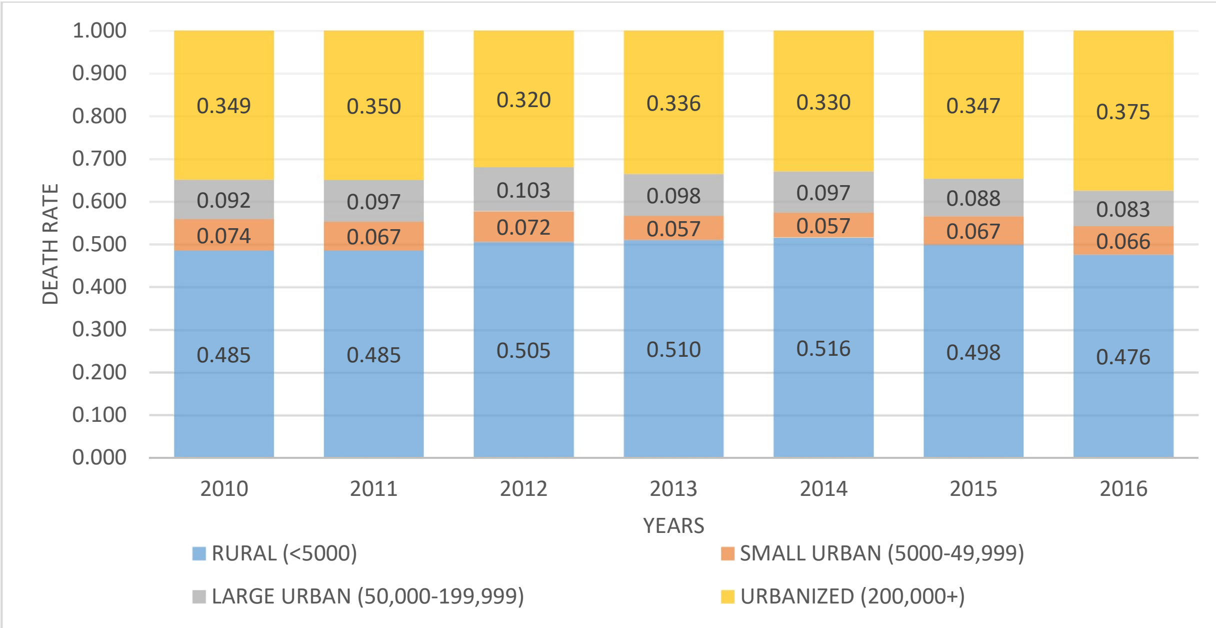


Figure 3.8. Traffic Death Rate in Texas by Population Group.

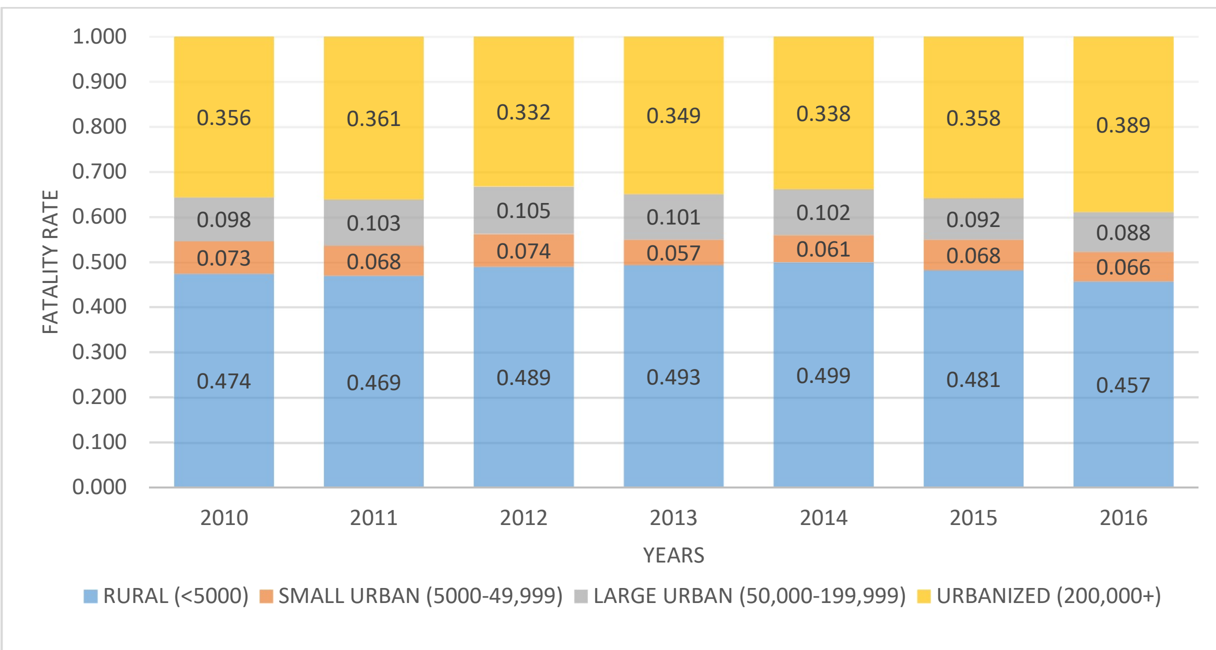


Figure 3.9. Traffic Fatality Rate in Texas by Population Group.

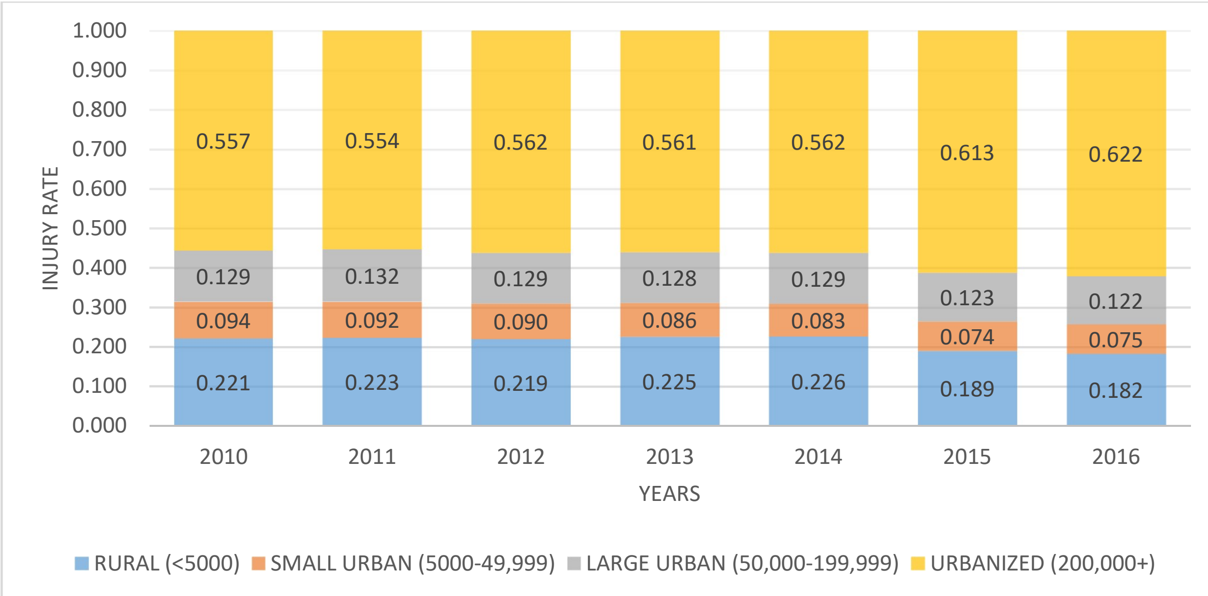


Figure 3.10. Traffic Injury Rate in Texas by Population Group.

In response to the question, Figures 3.8, 3.9, and 3.10 are plotted with the yearly death, fatality, and injury rates for urban and rural areas. The urban zones are categorized in three groups according to their population ranges. However, the database has urban as a population higher than 5,000 people, and everything below that number is described as a rural area. Figures 3.8 and 3.9 provide a holistic picture and convincing evidence that the rural area has the greatest percentage of deaths and fatalities, with a mean of 49.7% and 48%, respectively, for the 7-year duration. The death and fatality rate is moderately constant throughout this period. The injury rate suggests that an urban population that is greater than 200,000 people accounts for most injuries, with a mean of 57.6%. It is important to note that the rural population has a smaller injury rate and a higher death rates. The statistics provide an absolute fact that accidents in the country result in more deaths than accidents in urban places and provides a smaller margin of survivability in accidents. Hence, these values offer an alternative fact that driving in rural zones is deadlier than driving in the more-populated region.

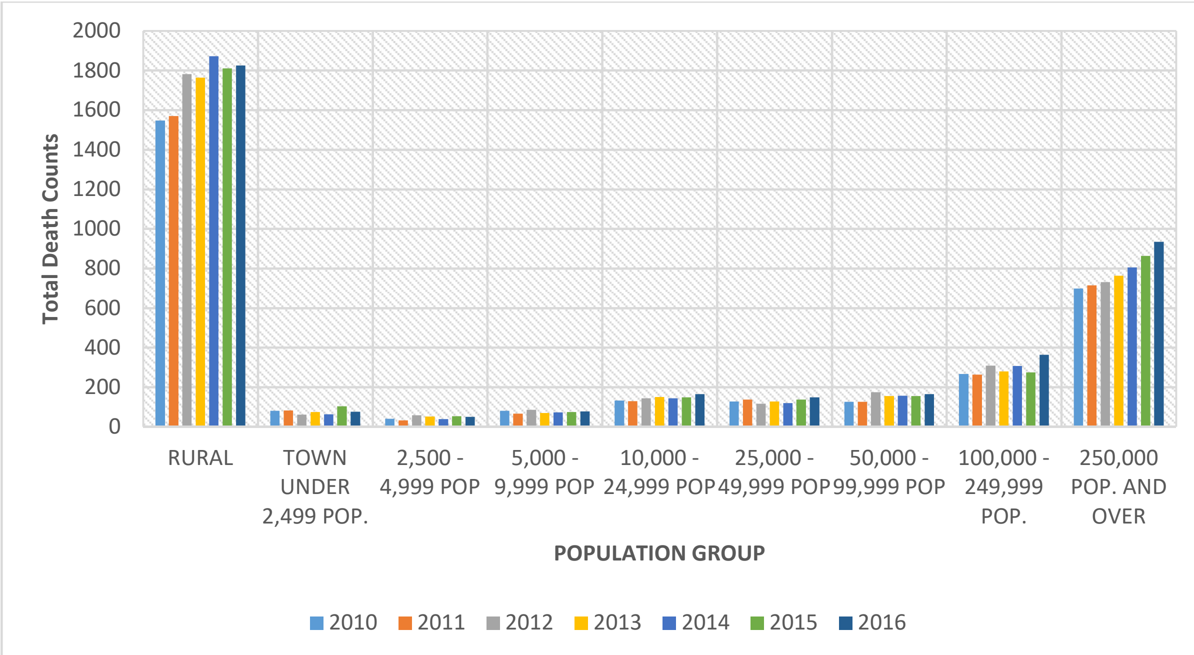


Figure 3.11. Road Accidents' Death Count by Population Group (2010-2016).

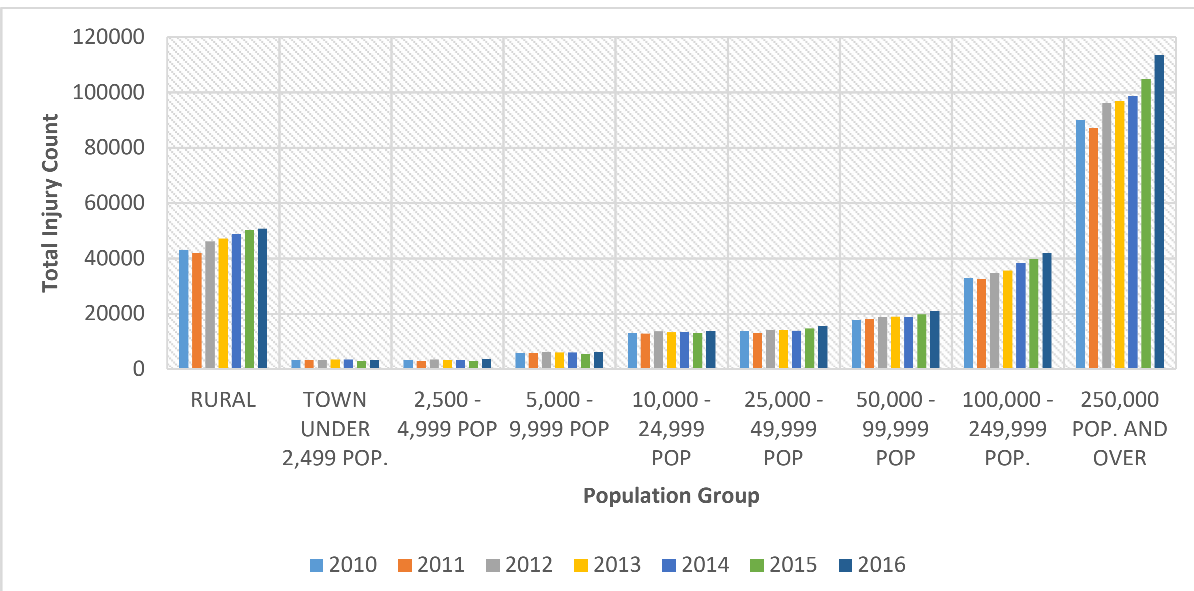


Figure 3.12. Road Accidents' Total Injury Count by Population Group (2010-2016).

Figures 3.11 and 3.12 provide more comprehensive information by splitting the rural and urban zones into broader population groups. The rural and urban areas are divided into three and six categories, respectively, depending on the population ranges. Figure 3.11 illustrates that the death count is the highest in the less-populated region that has fewer than a thousand residents, and the trend is constant for the entire study period. The death count is second highest for the highly populated locations. However, the trend fluctuates for injuries and reveals a different image. The injuries are highest in most-populated areas, and the second-highest injuries are found in the less-populated location. The important conclusion is that the highest and lowest populated places account for most deaths and injuries.

Table 3.5. Accident Parameters for Injuries in the Rural Population Group.

Rural injury parameters	Years						
	2010	2011	2012	2013	2014	2015	2016
Weather condition	Clear	Clear	Clear	Clear	Clear	Clear	Clear
Light condition	Daylight	Daylight	Daylight	Daylight	Daylight	Daylight	Daylight
Road type	2 lane, 2 way	2 lane, 2 way	2 lane, 2 way	2 lane, 2 way	2 lane, 2 way	2 lane, 2 way	2 lane, 2 way
Road alignment	Straight, level	Straight, level	Straight, level	Straight, level	Straight, level	Straight, level	Straight, level
Surface condition	Dry	Dry	Dry	Dry	Dry	Dry	Dry
Traffic control	Center stripe/divider	Center stripe/divider	Center stripe/divider	Center stripe/divider & marked lanes	Marked lanes & center stripe/divider	Marked lanes & center stripe/divider	Marked lanes & center stripe/divider
Road part	Main/proper lane	Main/proper lane	Main/proper lane	Main/proper lane	Main/proper lane	Main/proper lane	Main/proper lane
Roadway system	Farm to market	Farm to market	Farm to market	Farm to market, local roads/streets & state highway	Farm to market, county road & state highway	Farm to market, county road & state highway	Farm to market, county road & state highway
Surface type	High type flexible	High type flexible	High type flexible	High type flexible	High type flexible	High type flexible	High type flexible
Speed-Limit Ranges	55-70	70-45	70	55-75	55-75	45-75	45-75

Table 3.5 displays information about the various accident factors for injuries in rural areas during the 7-year period. Each element's properties were selected based on the maximum number of crashes that resulted in injuries. Table 3.5 highlights the roadway's condition, road structure, environment, and speed limit at the time of an accident which involved an injury or

injuries. The most injuries were encountered in rural areas when vehicles were making their way on a farm to market road. When an accident that caused an injury or injuries occurred, the weather condition was reported as clear, and the light condition was rated as excellent. Looking at Table 3.5 suggests that the conditions at the time of accidents which caused injuries were typically standard and portrays a similar style for all years. At the time of an injury, the roads were in perfect shape, but the relevant argument to notice is that the traffic-control category at the scene of a crash which caused injuries was center stripe/divider. Combining this fact with the speed limit and the road type creates an argument and questions the roadway's possible structure. By viewing the factors in Table 3.5, a case can be made that injury crashes could happen due to drivers trying to pass the vehicle in front of them by speeding, thus leading to an injury or injuries. If the casual driving is the case, a possible solution could be to change the road structure by expanding the roads and providing a median between the two-way traffic.

Table 3.6. Accident Parameters for Fatalities in the Rural Population Group.

Rural Fatality Parameters	Years						
	2010	2011	2012	2013	2014	2015	2016
Weather Condition	Clear	Clear	Clear	Clear	Clear	Clear	Clear
Light Condition	Daylight & Dark, Not Lighted	Daylight & Dark, Not Lighted	Daylight & Dark, Not Lighted	Daylight & Dark, Not Lighted	Daylight & Dark, Not Lighted	Daylight & Dark, Not Lighted	Daylight & Dark, Not Lighted
Road Type	2 Lane, 2 Way	2 Lane, 2 Way	2 Lane, 2 Way	2 Lane, 2 Way	2 Lane, 2 Way	2 Lane, 2 Way	2 Lane, 2 Way
Road Alignment	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level
Surface Condition	Dry	Dry	Dry	Dry	Dry	Dry	Dry
Traffic Control	Center Stripe/Divider	Center Stripe/Divider	Center Stripe/Divider	Center Stripe/Divider	Center Stripe/Divider & Marked Lanes	Marked Lanes & Center Stripe/Divider	Marked Lanes & Center Stripe/Divider
Road Part	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane
Roadway System	Farm To Market	Farm To Market, State Highway & Us Highway	Farm To Market, Us Highway & State Highway	Farm To Market, State Highway & Us Highway	Farm To Market, Us Highway & State Highway	State Highway, Farm To Market & Us Highway	Farm To Market, Us Highway & State Highway
Surface Type	High Type Flexible	High Type Flexible	High Type Flexible	High Type Flexible	High Type Flexible	High Type Flexible	High Type Flexible
Speed Limit	70 & 65	70 & 65	70	55 - 75	75	75	75

In comparison with the injuries, Table 3.6 depicts a slight difference in terms of the values for the factors which cause fatal accidents. These accidents happened in the daylight and in dark conditions. Hence, vision was an issue with the fatal accidents. As anticipated, the speed-limit range was significantly higher for the fatal accidents than in the injuries table. However, the other factors in the category remained the same, such as clear weather, a dry road surface, and flexible road type. A slight change in the roadway system supports that the majority of the fatal accidents took place on State and U.S. highways, apart from vehicles which were maneuvering from the farm to market.

Table 3.7. Accident Parameters for Injuries in the Urban (250,000+) Population Group.

Urban (250,000+ Pop.) Injury Parameters	Years						
	2010	2011	2012	2013	2014	2015	2016
Weather Condition	Clear	Clear	Clear	Clear	Clear	Clear	Clear
Light Condition	Daylight	Daylight	Daylight	Daylight	Daylight	Daylight	Daylight
Road Type	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided
Road Alignment	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level
Surface Condition	Dry	Dry	Dry	Dry	Dry	Dry	Dry
Traffic Control	None & Signal Light	Signal Light, Marked Lanes & None	Signal Light, Marked Lanes & None	Marked Lanes, Signal Light & None	Marked Lanes, Signal Light & None	Marked Lanes, Signal Light & None	Marked Lanes, Signal Light & None
Road Part	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane
Roadway System	Local Road/Street	Local Road/Street	Local Road/Street	Local Road/Street	Local Road/Street	Local Road/Street	Local Road/Street
Surface Type	High Type Flexible & Rigid	High Type Flexible & Rigid	High Type Flexible & Rigid	High Type Flexible & Rigid	High Type Flexible & Rigid	High Type Flexible	High Type Flexible & Rigid
Speed Limit	35	35	35	35	35	35	35

Table 3.7 discusses the accident factors that resulted in an injury or injuries for urban populations with 250,000 or more people. For the urban population, the injuries happened most

on the local roads/streets. The weather, lighting, and surface conditions are similar to the situations in rural areas. However, the contrast showed in the factors for speed limit, road type, and traffic control. Interestingly, injuries occurred on 4-or-more-lane roads with a speed limit of 35 mph. Therefore, these circumstances make traffic control equipment an important factor to study. This factor proposes that signal lights and marked lanes are the key areas for the injury accidents. Combining all elements together, we can conclude that the accidents occurred due to traffic congestion and that drivers were unable to switch lanes safely. Moreover, the signal light suggests that drivers were either in a rush and crossed the traffic signal or were unable to stop at the signal. The other possibility could be that some drivers applied the break suddenly to stop at traffic lights while other drivers, who were coming from behind, could not react hastily; therefore, crashes with injuries happened.

Table 3.8. Accident Parameters for Fatalities in the Urban (250,000+) Population Group.

Urban (250,000+ Pop.) Fatality Parameters	Years						
	2010	2011	2012	2013	2014	2015	2016
Weather Condition	Clear	Clear	Clear	Clear	Clear	Clear	Clear
Light Condition	Dark, Lighted	Dark, Lighted & Daylight	Dark, Lighted & Daylight	Dark, Lighted & Daylight	Dark, Lighted & Daylight	Dark, Lighted & Daylight	Dark, Lighted
Road Type	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided
Road Alignment	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level
Surface Condition	Dry	Dry	Dry	Dry	Dry	Dry	Dry
Traffic Control	Marked Lanes	Marked Lanes	Marked Lanes	Marked Lanes	Marked Lanes	Marked Lanes	Marked Lanes
Road Part	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane
Roadway System	Local Road/Street	Local Road/Street	Local Road/Street	Local Road/Street	Local Road/Street	Local Road/Street	Local Road/Street
Surface Type	High Type Rigid	High Type Rigid	High Type Rigid	High Type Rigid & Flexible	High Type Rigid & Flexible	High Type Rigid & Flexible	High Type Rigid
Speed Limit	60 & 35	60 & 35	35 - 60	35 - 60	35 - 60	30 - 60	30 - 60

In contrarst, Table 3.8 provides a depiction of factors for a fatal accident that are significantly different than an accidental injury. As shown in Table 3.8, it is imperative that vision did not play a major factor in the fatal accidents due to light condition. The roadway was dark, but it was lit. Second, most fatalities occurred at a time with daylight. The weather condition was clear and thus endorse the vision was relatively good when a fatal accident ensued. Moreover, the roads had four lanes, were divided, were straight, and were dry. However, the parameters for traffic endorses that most fatal accidents happened due to a higher volume of traffic. Because urban areas are highly populated, we can anticipate the higher volume for road traffic. When viewing the table, we can conclude that the speed-limit ranges have increased drastically, from 35 to 60 mph, on the local street/roads. This statement proposes that the greater vehicle speed in congested traffic condition triggered fatal accidents. The type of road surface is equally rigid and flexible at the fatal accident's location.

3.3.4. Using the Average Daily Traffic to Analyze the Accident Parameters

Average daily traffic (ADT) measures the mean number of vehicles which pass a particular point in 24 hours and is typically calculated throughout the year. ADT helps identify and analyze the relationship between traffic volume and accidents. Studying the average daily traffic (ADT) determines if the increased ADT contributed to more accidents, fatalities, or injuries. In other words, does the increased ADT cause more safety-related issues with the traffic's movement? The other important question that we need to answer is the overall trend and if there is an observable ADT shift with the accidents for each year. These questions are addressed in the remaining discussion.

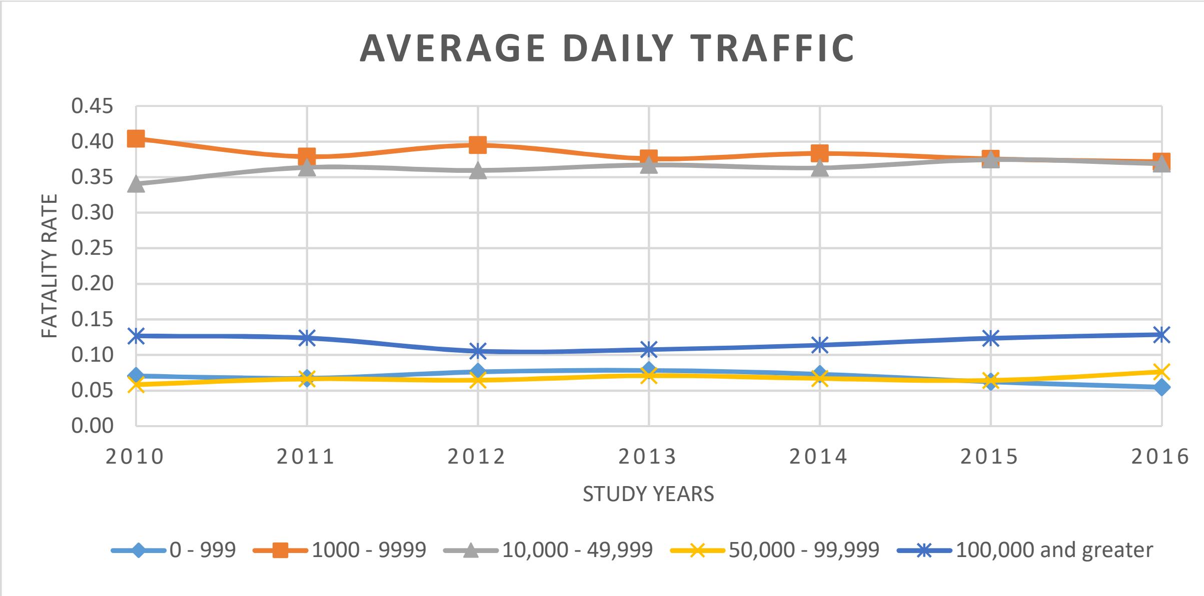


Figure 3.13. Traffic-Accident Fatality Rate by Average Daily Traffic (ADT) Category.

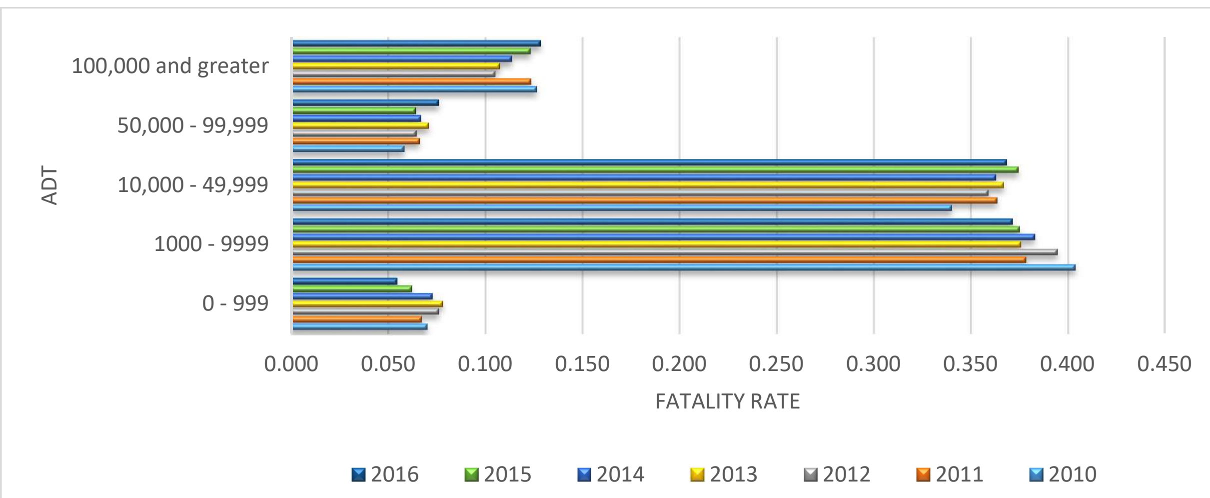


Figure 3.14. Comparison of the Yearly Traffic-Accident Fatality Rate by the Average Daily Traffic (ADT) Ranges.

Average daily traffic (ADT) is grouped into five categories. Figures 3.13 and 3.14 show the yearly fatality rates for the various ADT ranges. In Figure 3.13, it is important to notice that ADT ranges from 1,000 to 49,999 have a higher fatality rate than areas with more ADT; the fatality rate drops when the ADT increases. A similar drop in the fatality rate is observed when the ADT decreases. However, there is an increased fatality rate when the ADT value increases from 100,000. The fatality rates' mean values for ADT ranges from 1,000-9,999 and 10,000-49,999 are 0.383 and 0.362, respectively. The trend line in Figure 3.13 suggests that the fatality rate remained constant, with slight variations, throughout the study period. The numbers give more indication on the responsibility for authorities to put more effort to manage Texas' traffic problems. A slight decrease in the fatality rate observed in Figure 3.14 for the ADT ranges from 1,000 to 9,999, and fatalities has probably shifted to the ADT ranges from 10,000 to 49,999 due to the fatality rate increased over the time in this range. There is a drop in the fatality rate observed during the study's middle years, but the trend shifts to growth in fatality rate in 2015 and follows the same pattern into 2016.

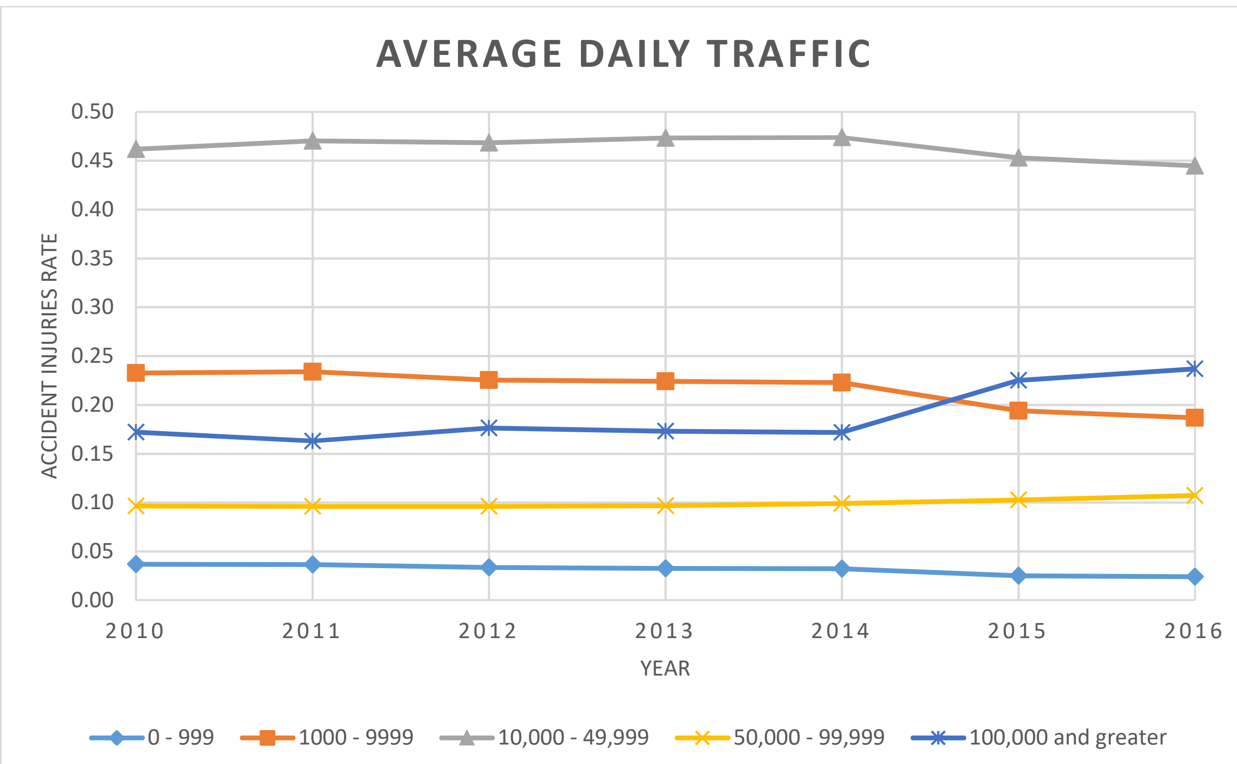


Figure 3.15. Traffic-Accident Injury Rate by Average Daily Traffic (ADT) Category.

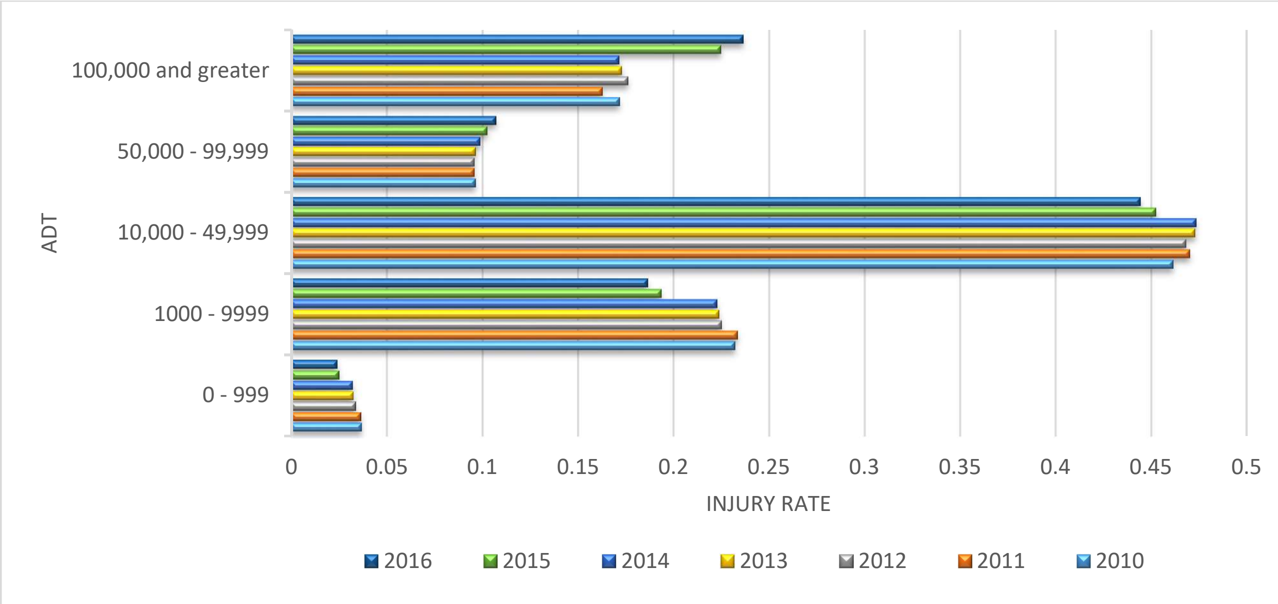


Figure 3.16. Comparison of the Yearly Traffic-Accident Injury Rate for Various Average Daily Traffic (ADT) Ranges.

Figures 3.15 and 3.16 show that the injury rate for different ADT ranges reflects a different trend than the fatality rate. By viewing Figures 3.15 and 3.16, we conclude that the ADT range of 10,000-49,999 accounted for the highest number of injuries and had the maximum injury rate. The injury rates mean value is 0.464. This value suggests that almost 50% of the injuries occur in the ADT ranges from 10,000-49,999. Figure 3.15 indicates that the ADT range of 100,000 and higher has a drastic increase of 38% for the injury rate by the end of the study period. However, the opposite trend, a decreased injury rate, is observed for the ADT ranges of 1,000-9,999 and 0-999; there is a negative value of 20% and 35%, respectively. The remaining ADT ranges show a constant value for the overall study period.

3.4. Intersections

Intersections are considered hotspots for traffic accidents in the United States and have contributed enormously towards the crashes. This part of the analysis attempts to understand the condition at intersections by the help of statistical figures and numbers for various accidental parameters. The discussion will try to dive in deep to precisely locate the issues that cause trouble at the intersections. Moreover, the statistical models are constructed to determine the relationships and their significance among multiple accident factors.

Table 3.9. Ratios for Accidents, Injuries, Fatalities, and Deaths, 2010-2016.

Year	Accident Ratio	Injury Ratio	Fatality Ratio	Death Ratio
2010	0.257	0.342	0.166	0.167
2011	0.264	0.346	0.169	0.168
2012	0.271	0.354	0.167	0.167
2013	0.278	0.361	0.181	0.184
2014	0.282	0.362	0.171	0.172
2015	0.284	0.371	0.183	0.183
2016	0.291	0.377	0.176	0.175

Table 3.9 provides the statistical figures for intersection accidents on the roadways. The values are listed in rate form which is calculated by dividing the accident, death, injury, and fatality value at the intersection with a corresponding overall value for the individual year. The crash ratio's mean for each year is calculated as 0.275 with a standard deviation of 0.012 and a median of 0.278. Therefore, intersections accounted for 27.5%, on average, of the accidents each year. The highest accident ratio is in 2016 with a value of 0.291, and the lowest one is in 2010 with a value of 0.257, suggesting that accidents at intersections significantly increased during the study period. The injury rate is significantly large with a mean of 0.359 and a standard deviation of 0.013. The injury ratios are amplified every year. Intersections are responsible for about 36% of the total injuries in Texas, and the injuries account for 62.5% of the total accidents at the intersections. The statistical values in Table 3.9 are more than predictable and pose a scary situation for the authorities to take initial steps in order to improve the Texas intersections. The

fatality and death ratios reflect the fluctuating trend and have declined for some previous years. The mean values for the fatality and death ratios are 0.173 and 0.174, respectively, with a standard deviation of 0.007. The average rates for both casualties and deaths are same. Still, 0.3% of the deaths resulted from accidents at intersections.

Table 3.10. Casualty Count and Year-to-Year Growth at Intersections by Month and Year (2010-2016).

Years	Casualty Count	Casualty Count YoY Growth	Years	Casualty Count	Casualty Count YoY Growth	Years	Casualty Count	Casualty Count YoY Growth
2010	121543		2013	144741	7.80	2016	181543	6.86
Jan	8960		Jan	10708	5.60	Jan	14311	10.36
Feb	9574		Feb	10926	7.57	Feb	14877	18.66
Mar	10881		Mar	12707	10.86	Mar	15669	10.87
Apr	10700		Apr	12155	11.60	Apr	15482	10.07
May	10576		May	12599	10.38	May	15544	8.42
Jun	10239	-	Jun	11844	7.41	Jun	14782	5.51
Jul	9835		Jul	11528	7.45	Jul	14441	4.64
Aug	9937		Aug	12283	7.26	Aug	15378	7.13
Sep	9930		Sep	12268	7.51	Sep	15683	8.25
Oct	10758		Oct	13407	10.56	Oct	16397	3.81
Nov	9921		Nov	12394	6.69	Nov	14985	4.72
Dec	10232		Dec	11922	0.68	Dec	13994	-7.29
2011	120284	-1.04	2014	156572	8.17			
Jan	8979	0.21	Jan	11761	9.83			
Feb	8776	-8.34	Feb	11430	4.61			
Mar	10322	-5.14	Mar	12784	0.61			
Apr	10753	0.50	Apr	13322	9.60			
May	10557	-0.18	May	13625	8.14			
Jun	9979	-2.54	Jun	12531	5.80			
Jul	9134	-7.13	Jul	12283	6.55			
Aug	9777	-1.61	Aug	13335	8.56			
Sep	10067	1.38	Sep	13244	7.96			
Oct	11101	3.19	Oct	15122	12.79			
Nov	10076	1.56	Nov	13601	9.74			
Dec	10763	5.19	Dec	13534	13.52			
2012	134269	11.63	2015	169894	8.51			
Jan	10140	12.93	Jan	12967	10.25			
Feb	10157	15.74	Feb	12538	9.69			
Mar	11462	11.04	Mar	14133	10.55			
Apr	10892	1.29	Apr	14065	5.58			
May	11414	8.12	May	14337	5.23			
Jun	11027	10.50	Jun	14010	11.80			
Jul	10729	17.46	Jul	13801	12.36			
Aug	11452	17.13	Aug	14355	7.65			
Sep	11411	13.35	Sep	14488	9.39			
Oct	12126	9.23	Oct	15795	4.45			
Nov	11617	15.29	Nov	14310	5.21			
Dec	11842	10.03	Dec	15095	11.53			

Table 3.10 offers more detailed information about the accidents at intersections for each year; the information is further categorized by month. Table 3.10 gives a holistic picture to identify the deadliest months. Every year, October is recognized as when most accidents occurred at intersections. Table 3.10 also gives the statistical figures for the casualty count of year-to-year growth. The statistics illustrate that the number of accidents has risen each year, in most cases within a range from 0-18% from previous years' month to month values. However, 2011 displays a decrease in accidents in few months compared to the prior year's months. Overall, it is imperative from the table 3.10 values that the accident trend has presented a significant positive increment in the study period compared from month to month for each respective year. Table 3.10 also shows the increases and decreases for year-to-year growth displayed with direction signs, where a red sign represents a decrease, the yellow sign shows a slight increase (up to 10%), and the green sign represents more than 10% growth.

3.4.1. Intersection Analysis by Population Group

As discussed earlier and overviewed on the accidents sensitivity at intersections, this part of the analysis has an in-depth review of the factors which cause crashes at an intersection in relation to the population groups. The population is divided into nine categories in the Texas DOT database. The population groups are used to determine the injury and fatality counts for each year and are linked with various factors to investigate the conditions and circumstances that caused an injury or fatality during an accident.

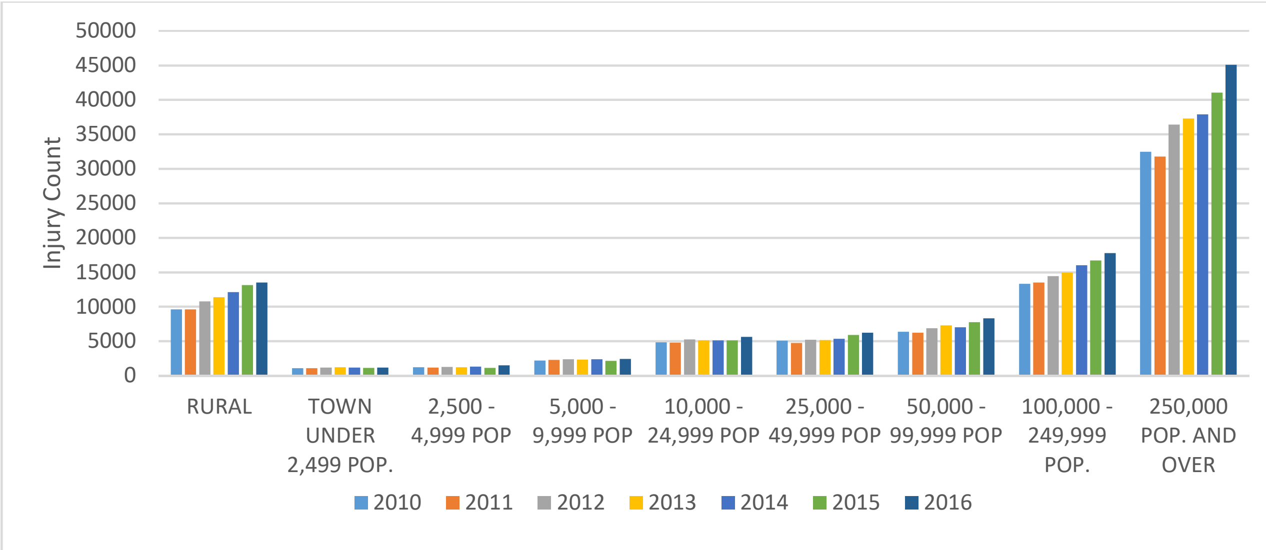


Figure 3.17. Each Year's Traffic-Accident Injury Count, by Population Groups, at Intersections.

Figure 3.17 shows the statistical figures for accident injuries that correspond to the population groups. Figure 3.17 shows no additional differences than the previous analysis for the overall accidents in Texas during the individual years. The values in Figure 3.17 illustrate that the highest population group has more injuries than any other group at intersections. The largest population group represents Texas' urban cities. However, there is a slight change in Figure 3.17 based on what was discussed in the earlier section. Figure 3.17 displays that the second-largest population group has a second-highest count for injuries while the rural group was in third place. Therefore, the trend has shifted from the countryside towards the highly populated region for intersections. The injury difference at intersections, from first highest to second highest, is significantly high, although the second and third highest have a slight variation compared to injuries.

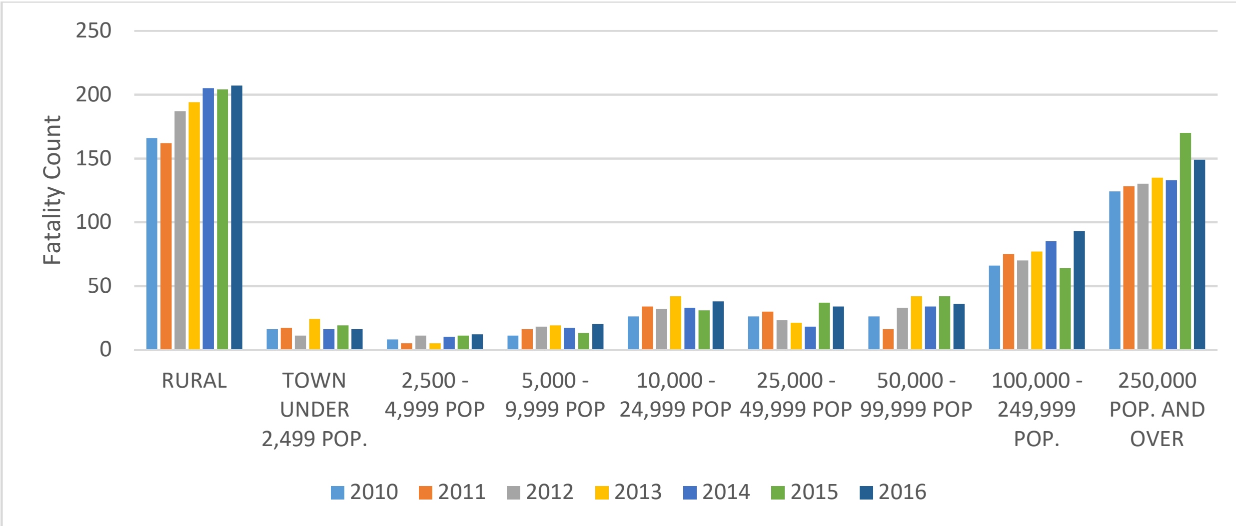


Figure 3.18. Each Year's Traffic-Accident Fatality Count, by Population Groups, at Intersections.

The fatality count in Figure 3.18 shows a different outlook than the injury counts in Figure 3.17, however, the fatality overview in Figure 3.18 is similar when compared with the total accidents in Texas, by population group, for a 7-year period. As expected, the rural area has the most fatalities, followed by the highest-population group. The trend in Figure 3.18 illustrates that the intersection fatalities have grown significantly for the rural population group since 2010 and have remained constant in the most recent 2 to 3-year period.

Table 3.11. Intersection Accidents' Fatality Parameters for the Rural Population Group.

RURAL FATALITY PARAMETERS	YEARS						
	2010	2011	2012	2013	2014	2015	2016
Weather Condition	Clear	Clear	Clear	Clear	Clear	Clear	Clear
Light Condition	Daylight	Daylight	Daylight	Daylight	Daylight	Daylight	Daylight
Road Type	2 Lane. 2 Way	2 Lane. 2 Way	2 Lane. 2 Way	2 Lane. 2 Way	2 Lane. 2 Way	2 Lane. 2 Way	2 Lane. 2 Way
Road Alignment	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level
Surface Condition	Dry	Dry	Dry	Dry	Dry	Dry	Dry
Traffic Control	Stop Sign	Stop Sign	Stop Sign	Stop Sign	Stop Sign	Stop Sign	Stop Sign
Road Part	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane
Roadway System	US Highway, Farm to Market & State Highway	US Highway, Farm to Market & State Highway	Farm to Market, State Highway & US Highway	US Highway, State Highway, Farm to Market & County Road	State Highway, Farm to Market & US Highway	State Highway, US Highway & Farm to Market	US Highway, Farm to Market & State Highway
Surface Type	High-Type Flexible	High-Type Flexible	High-Type Flexible	High-Type Flexible	High-Type Flexible	High-Type Flexible	High-Type Flexible
Speed Limit	55 - 70	55 - 70	70	55 - 75	55 - 75	55 - 75	55 - 75

Table 3.11 shows various conditions and parameters contribute the most towards road fatalities and examined in co-relation to the rural places which were identified formerly for highest fatality count among all population groups. Table 3.11 displays the identified factors for each parameter which are significantly constant throughout the study; the table does not show

much change for the pattern. When a fatality occurred, the weather condition was identified as clear, happened in a daylight condition, and the road-surface condition was dry. Therefore, the weather did not contribute to the fatality because the overall status was good and the vision was satisfactory when an accident occurred. The road structure pronounces that the roads were two lanes, two-way, and straight and that they were constructed with high-type flexible material. However, two important factors are traffic control and speed; most fatalities occurred at stop signs and due to speeding because the speed-limit range was higher during a fatality. This fact gives information about how stop signs are lethal for traffic and demand a revolutionary effort to reduce fatalities.

Table 3.12. Intersection Accidents' Injury Parameters for the Urban (250,000+) Population Group.

URBAN (250,000+ POP.) INJURY PARAMETERS	YEARS						
	2010	2011	2012	2013	2014	2015	2016
Weather Condition	Clear	Clear	Clear	Clear	Clear	Clear	Clear
Light Condition	Daylight	Daylight	Daylight	Daylight	Daylight	Daylight	Daylight
Road Type	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided	4 Or More Lanes, Divided
Road Alignment	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level	Straight, Level
Surface Condition	Dry	Dry	Dry	Dry	Dry	Dry	Dry
Traffic Control	Signal Light	Signal Light	Signal Light	Signal Light	Signal Light	Signal Light	Signal Light
Road Part	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane	Main/Proper Lane
Roadway System	Local Road/Street	Local Road/Street	Local Road/Street	Local Road/Street	Local Road/Street	Local Road/Street	Local Road/Street
Surface Type	High Type Flexible & Rigid	High Type Flexible, Rigid & Composite	High Type Flexible & Rigid	High Type Flexible & Rigid	High Type Flexible & Rigid	High Type Flexible & Rigid	High Type Flexible & Rigid
Speed Limit	30 - 35	30 - 35	35	35	35	35	35

As explained previously, the urban population of greater than 250,000 was identified as the group responsible for most injuries from accidents each year, and Table 3.12 presents information about the factor elements for the identified population group from the crash database. The weather, surface, and lighting conditions are identical to the factors discussed for fatalities. However, the road has increased from 2 to 4 lanes and is divided as expected in the major cities due to the amount of traffic on roadways. The relevant information in Table 3.12 identifies that the location for most injuries was a signal light at an intersection and that the speed limit was low as 30 mph when injuries were logged. The speed projects on how people were not paying attention in driving to their destination and troubled at the signal light for one of the following reasons: not following the signal lights, speeding at the signal, or hitting the brakes to make a sudden stop that triggered vehicle collisions.

3.4.2. Intersection Analysis by Average Daily Traffic (ADT)

Average daily traffic (ADT) is used in this research to study the amount of daily traffic on highway intersections for total crashes, accidents, and fatalities at various range. The results give an additional perspective about how various ranges of Average daily traffic destabilize the roadway's safety conditions and illustrate the variation for crashes or accidents at different ADT ranges. The ADT ranges are typically drawn from low-to-semi-medium and from medium-to-high ADT. We classify the ADT in 5 different ranges which rely on the minimum and maximum ADT value. The added ADT ranges provide a more wide-ranging position for accidents at intersections.

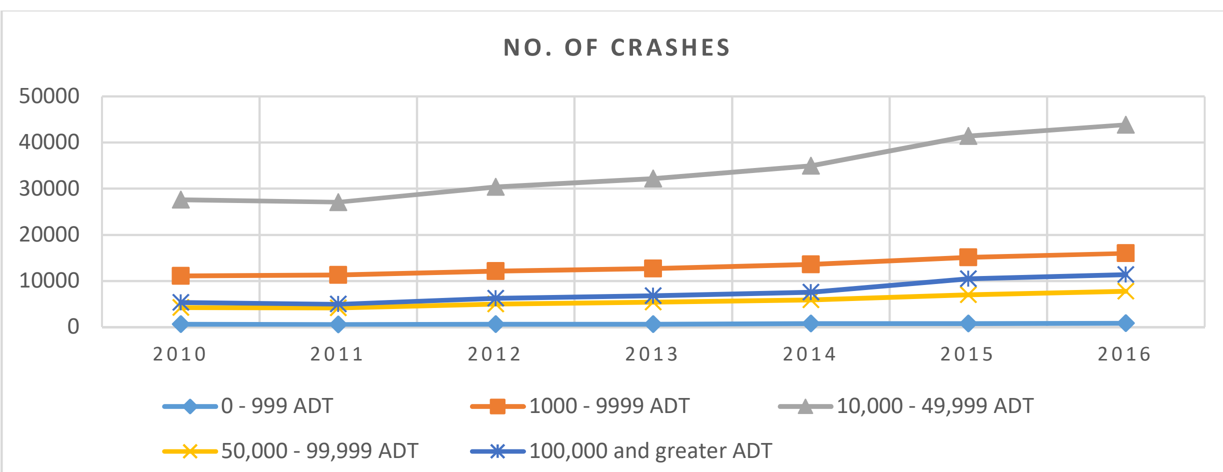


Figure 3.19. Traffic-Intersection Accident Counts by ADT Category Ranges.

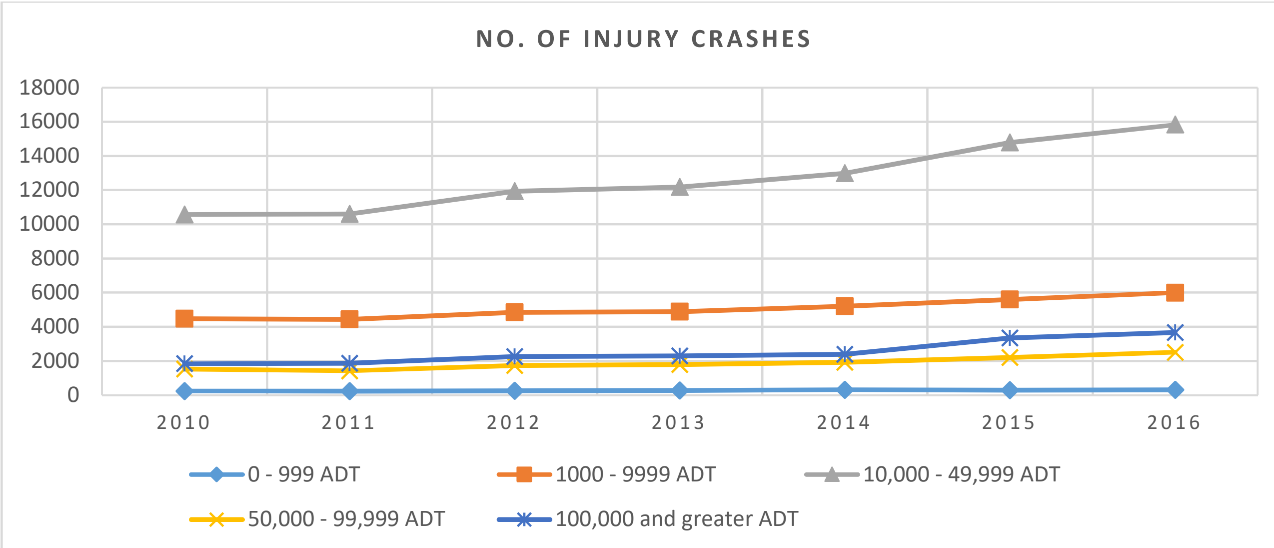


Figure 3.20. Traffic-Intersection Injury Counts by ADT Category Ranges.

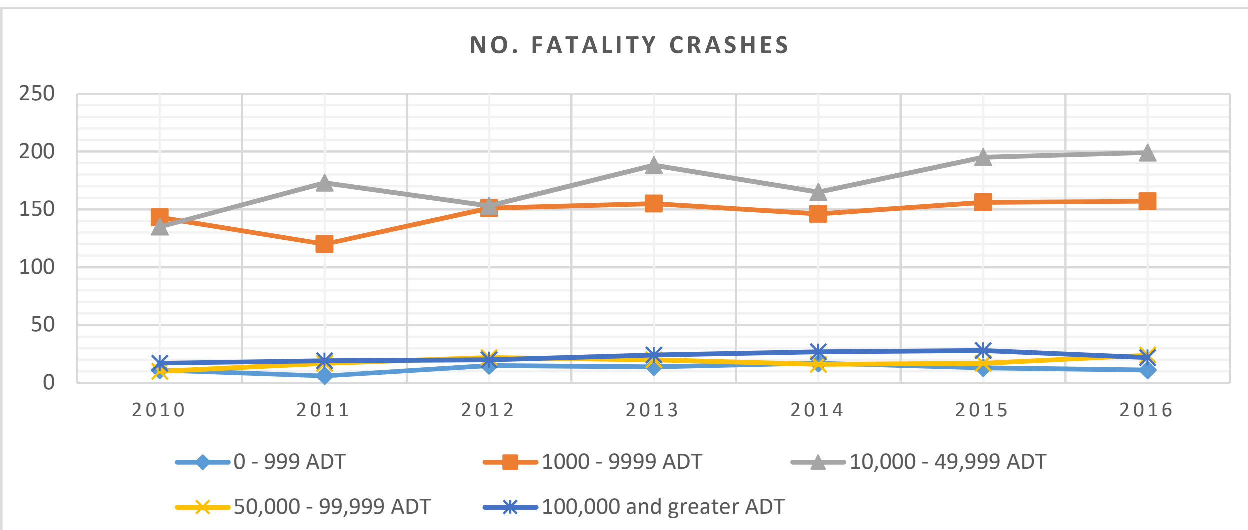


Figure 3.21. Traffic-Intersection Fatality Counts by ADT Category Ranges.

Figures 3.19, 3.20, and 3.21 portray the values for accidents, fatalities, and injuries for various ADT ranges at intersection points. The graph for accidents and injuries displays a symmetrical trend for the trend line and illustrates that the middle-range ADT, i.e., 10,000-49,999, has the highest number of crashes and injuries. Figure 3.21 provides a different look at fatal crashes; the ADT ranges of 1,000-9,999 and 10,000-49,999 have the highest number of fatal crashes, with slight variations. Nevertheless, the ADT range of 10,000-49,999 has slight superiority over the others and accounts for the most fatal accidents at intersections. On average, there have been 33,915 crashes and 12,697 injuries with an ADT range of 10,000-49,999 during the 7-year period. Figures 3.19, 3.20, and 3.21 also indicate that, on average, the lowest ADT range (0-999) has the lowest number of crashes and injuries as well as the smallest number of fatal accidents. The smallest ADT does not contribute as significantly as the medium-to-high average daily traffic. The mean value for the ADT ranges that caused most fatal accidents are 173 and 147 for the ranges of 10,000-49,999 and 1,000-9,999, respectively.

Table 3.13. Crash Injury Rate by Average Daily Traffic Category Ranges.

Average Daily Traffic (ADT)	2010	2011	2012	2013	2014	2015	2016	Mean	Median	Standard Deviation
0 - 999	0.401	0.420	0.405	0.405	0.402	0.379	0.379	0.399	0.402	0.015
1000 - 9999	0.402	0.393	0.399	0.384	0.381	0.370	0.375	0.386	0.384	0.012
10,000 - 49,999	0.383	0.391	0.394	0.379	0.372	0.357	0.361	0.377	0.379	0.014
50,000 - 99,999	0.363	0.343	0.348	0.331	0.325	0.316	0.323	0.336	0.331	0.017
100,000 and greater	0.345	0.375	0.362	0.340	0.316	0.321	0.322	0.340	0.340	0.022
Trend Line										

Table 3.14. Injury Fatality Rate by Average Daily Traffic Category Ranges.

Average Daily Traffic (ADT)	2010	2011	2012	2013	2014	2015	2016	Mean	Median	Standard Deviation
0 - 999	0.046	0.025	0.058	0.053	0.054	0.046	0.036	0.045	0.046	0.012
1000 - 9999	0.032	0.027	0.031	0.032	0.028	0.028	0.026	0.029	0.028	0.002
10,000 - 49,999	0.013	0.016	0.013	0.015	0.013	0.013	0.013	0.014	0.013	0.002
50,000 - 99,999	0.007	0.012	0.013	0.011	0.008	0.008	0.010	0.010	0.010	0.002
100,000 and greater	0.009	0.010	0.009	0.010	0.011	0.008	0.006	0.009	0.009	0.002
Trend Line										

From a different viewpoint, the crashes' injury rate and the injuries' fatality rate are calculated for each year based on the ADT ranges, and the statistical figures are provided in Tables 3.13 and 3.14. The tables also provide the data bars and the trend lines for each year's values. The statistical values in Tables 3.13 and 3.14 illustrate that the lowest ADT (0-999) has the highest injury fatality rate and a slight higher from values in the crash injury rate. The mean value for the accidents' injury rate and the injuries' fatality rate for the 0-999 ADT range is 0.39 and 0.045, with a standard deviation of 0.015 and 0.012, respectively. The numbers identify that, on average, 39% of the crashes result in injuries and that 4.5% of the injuries cause fatalities. However, the crashes' injury-rate values for each ADT range have a slight variation, and the rate difference, on average, is 6%. Therefore, we can conclude that the crashes' injury rate trended equally among all ADT ranges. The difference for the injuries' fatality-rate values is significantly large, and the rate suggests that the injuries' fatality rate decreased as the ADT values increased. The drop-in values are seen in the trend-line column for both rates.

3.5. Pictorial Representation of the Accident Database, 2010–2016

Initially, ArcGIS is utilized to analyze and to study the pictorial and graphic trends for accidents in Texas. The ArcGIS map provides a convenient tool for the audience and viewers to gain insight and awareness about the crash database. Without going into complex statistics, the readers can have a better understanding and in-depth overview on the argument the writer is making about Texas' crash data. The previous statistical analysis illustrated that most accidents occur in the densely populated regions. The maps offer the locations for places with the most accidents, enhancing the results identified by the statistical analysis with the help of geographic positional mapping. Moreover, the accidents' locations are important aspects when analyzing the data, specifically when studying situations such as accidents at urban intersections, to detect the causes and issues. The general conclusion for the accidents' root causes can be studied from the toughest accidental scenario-based studies. The results can be applied to improve the transportation system in Texas.

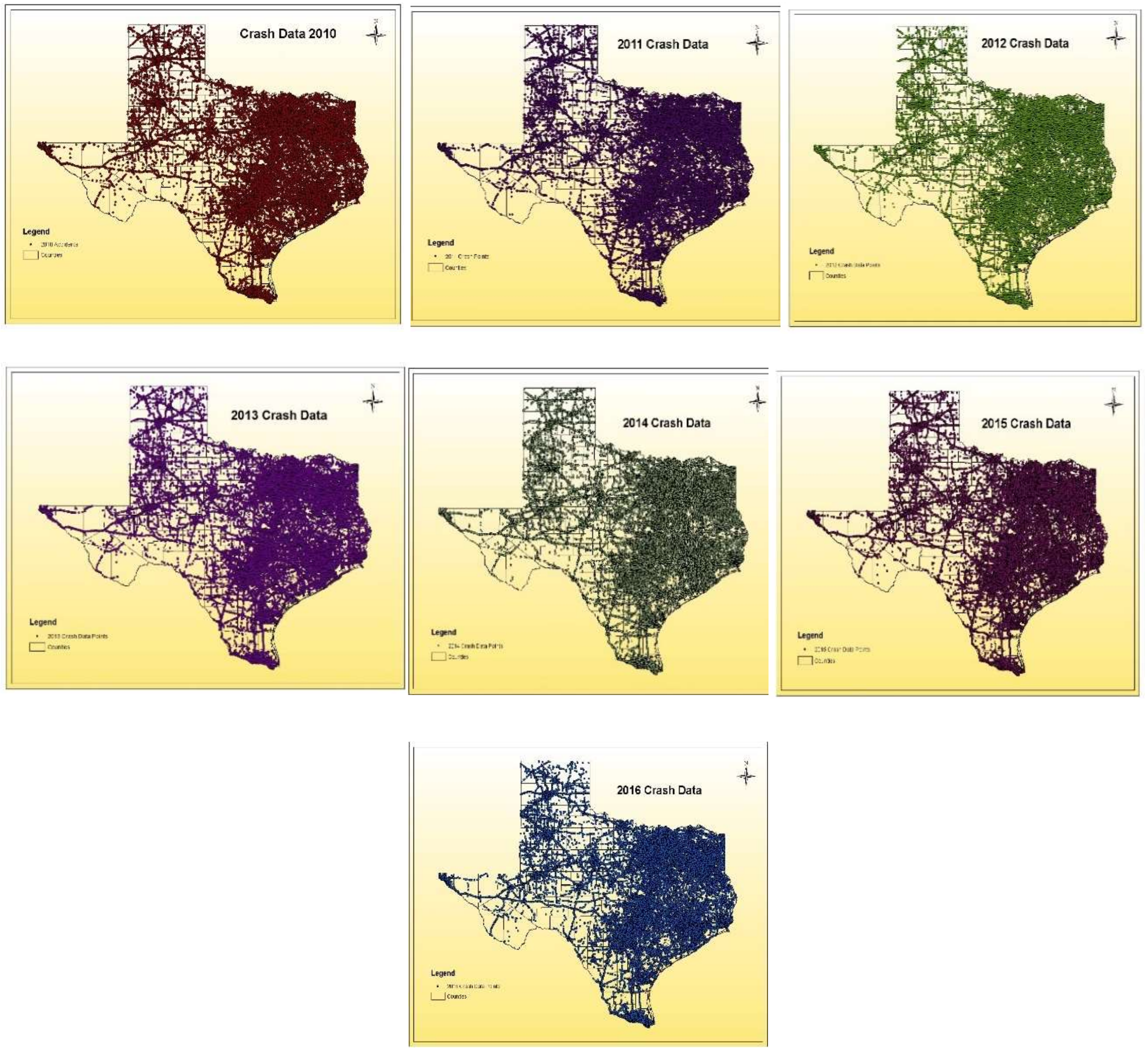


Figure 3.22. Pictorial Representation of Texas' Accident Counts Using ArcGIS, 2010-2016.

The graphical representation of the accident data from 2010-2016 is shown in Figure 3.22. The accident locations are shown in varying colors for each year. Figure 3.22 illustrates the trend that many accidents occur in the east and southeast parts of Texas. The other side of the state is blank with no accidents, and the points shadow the polyline's trend, signifying the network of highways. With the passage of time, the images show more clusters for accident locations. The graph shows the growth of accidents in Texas with the passage of time during the study period. Also, the cluster of numerous accidents in Texas is seen more in the densely populated region as anticipated and analyzed statistically. Overall, the trend is constant for the 7-year study period, and there is not a swing in the accident data.

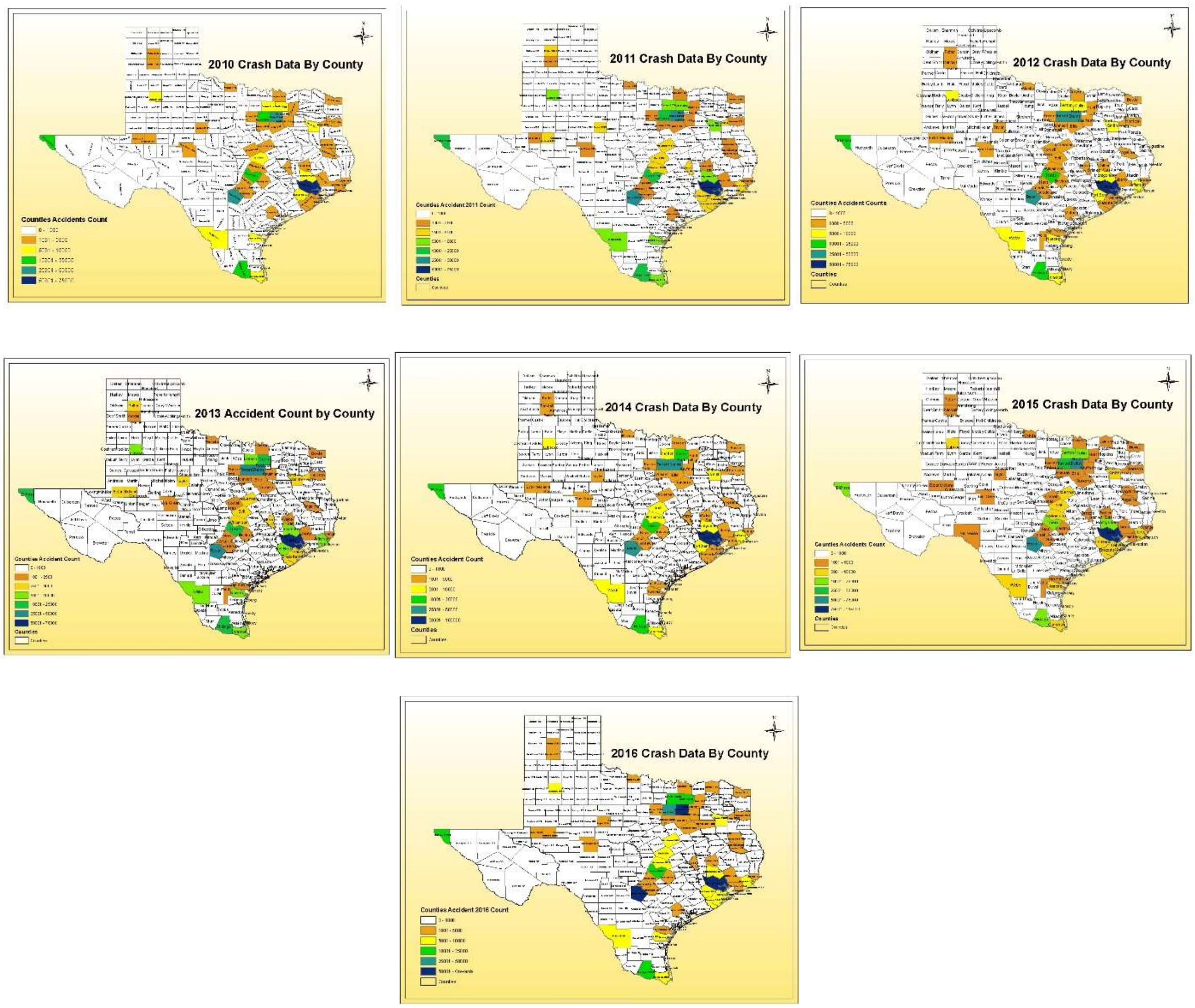


Figure 3.23. Pictorial Representation of Texas' Accident Counts, by County, Using ArcGIS, 2010-2016.

The accident database for the 7 years from 2010-2016 is spatially joined based on the location within the counties' jurisdiction, and the accident counts for each county are displayed in Figure 3.23. The symbolic representation in graduated colors shown in the maps to display ranges for various accident counts. The pictorial maps help classify the counties with some accidents in different population ranges. These maps give a holistic picture of the 254 counties that cannot be presented statistically. The statistical work identified 6 counties with recorded accidents, whereas these maps provide the accident count for the entire state. Hence, using the maps efficiently enable the viewer to detect the counties' critical spots per accident counts. Moreover, the maps identify counties adverse traffic condition and point towards the authorities to take effective actions to enhance the overall situation for the counties.

3.6. Conclusion

The chapter talked, in detail, about Texas' accident database and tried to create a perspective in the readers' mind about the critical roadway conditions. Furthermore, the chapter attempted to give multiple perspectives by analyzing the database from various directions, using parameters and factors, for the reader's better understanding. The purpose of the exploratory data analysis was to present the accident database in the simplest form and to exaggerate key points in order to highlight the issues for the responsible authorities. There are other ways that the database can be manipulated and utilized to deliver more information and viewpoints; however, more details were not included due to the constraints of time and research scope.

4. GEO-STATISTICAL ANALYSIS

4.1. Introduction

The chapter discusses information about the spatial analysis, along with its statistical tools, that was performed using ArcGIS to find decisive outcomes and eventful information about accidents in Houston, Texas. Geostatistics is a technique that incorporates two components: spatial location and statistical analysis. The geostatistical analysis creates a statistical model from the given point and interpolates the accident co-ordinates to develop a continuous surface through a spatial estimation and simulation technique. The importance of the tools and methods is identifying the locations which are more prone to accidents and need attention. The geostatistical-analysis tools used to evaluate the accident data for this thesis are kernel density, optimized hotspot analysis, space-time pattern mining, and kriging. The chapter is divided into the following sections:

i. Accident-Detection Count from the City Centroid at 1-Mile Intervals
ii. Kernel Density
iii. Optimized Hotspot Analysis
iv. Space-Time Pattern Mining
    a. Create Space-Time Cube
    b. Hotspot Analysis
    c. Local-Outlier Analysis
v. Kriging Tools
    a. Indicator Kriging
    b. Bayesian Kriging

The chapter evaluates each tool, in detail, based on how the tools performed. There is also a discussion about the result output.

4.2. Accident-Detection Count from the City Centroid at 1-Mile Intervals

4.2.1. Introduction

The first approach used to evaluate Houston's accident data is by creating a map that provides accident counts at 1-mile intervals from the city's centroid. The centroid is a geometric center of a feature or center of a mass for polygon. The purpose of developing these maps is to see trends for the number of accidents, starting from Houston's centroid. Moreover, the maps help to identify areas within the city where there are problems, providing an estimate for the location of most accidents. The maps give an overview of the location changes for Houston's traffic accidents from 2010-2016.

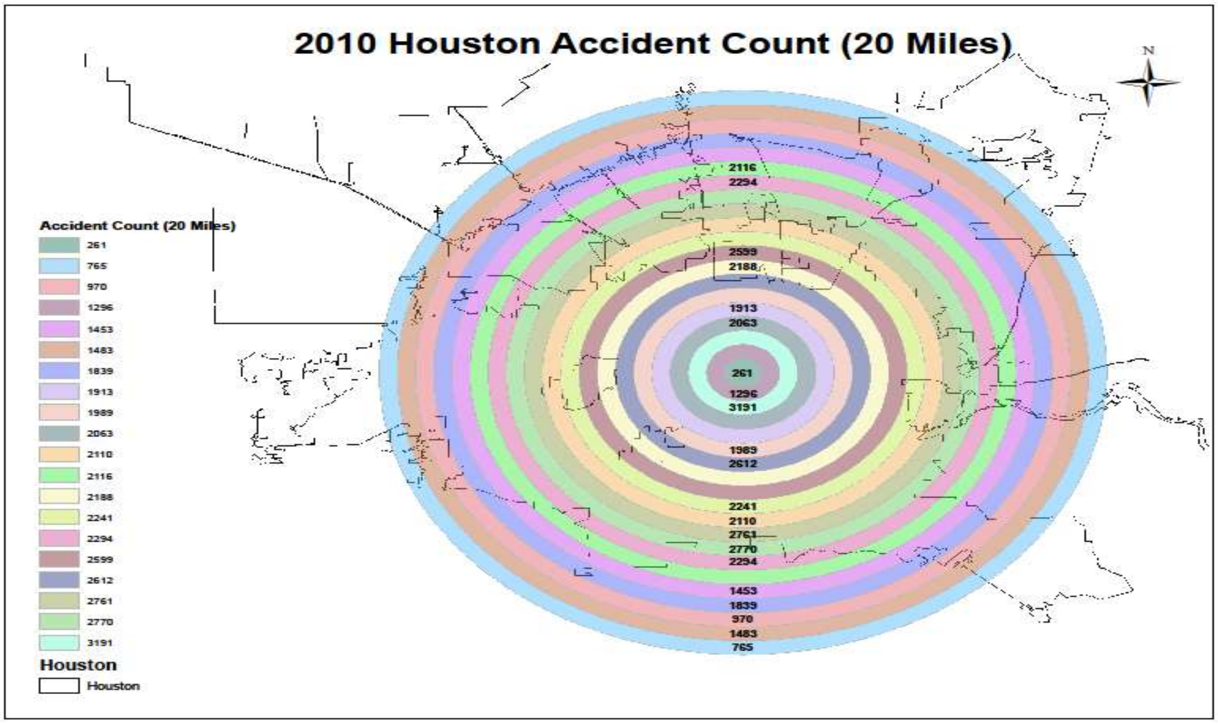
4.2.2. Procedure

The following steps are used to develop the accident-detection maps:

i. Determine the centroid for Houston. The centroid is the center of mass for a geometric object with a uniform density. ArcGIS calculates a polygon's centroid from the attribute table of the feature class. The ESRI (Environmental Systems Research Institute) website defines the procedure to find the centroid for a given polygon.
ii. Create rings from centroid in 1-mile intervals, going up to 20 miles.
iii. Use the clip tool to calculate the accident count for each ring.
iv. Use the Symbology Tab to showcase the accident count for each ring.

4.2.3. Results

This section presents the maps which result after performing the steps given in Section 4.2.2. The following maps are the results:



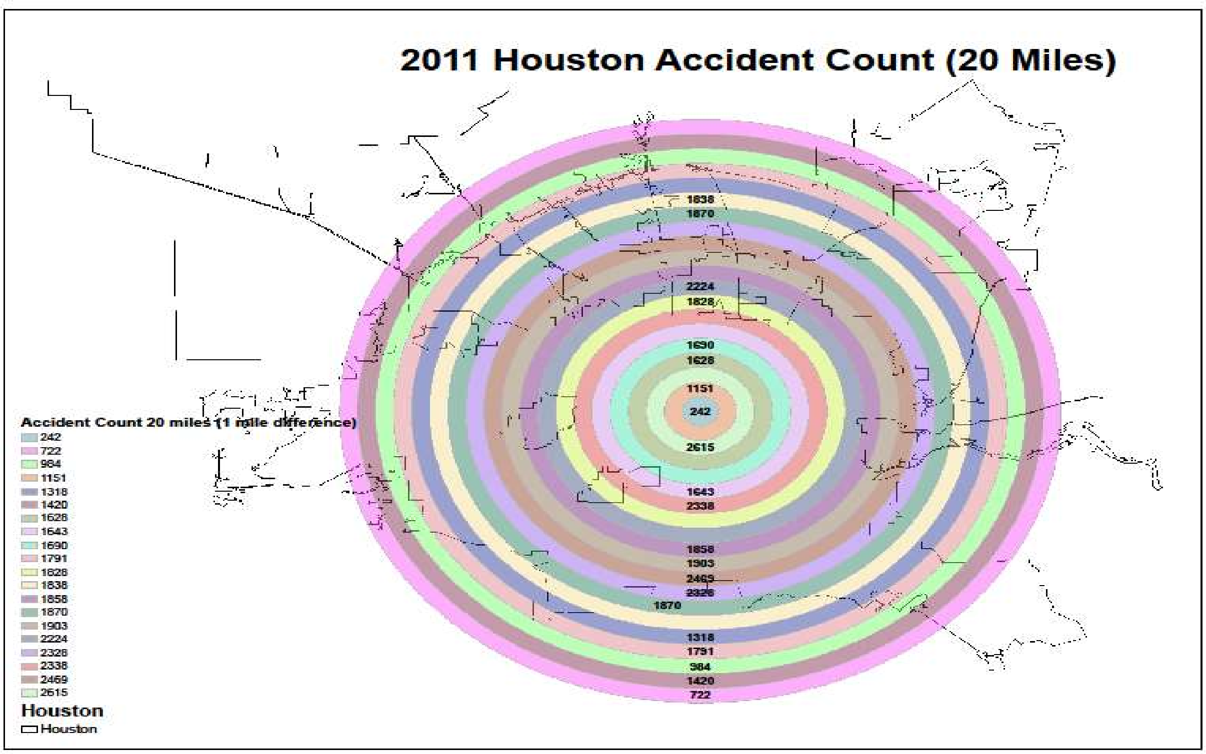
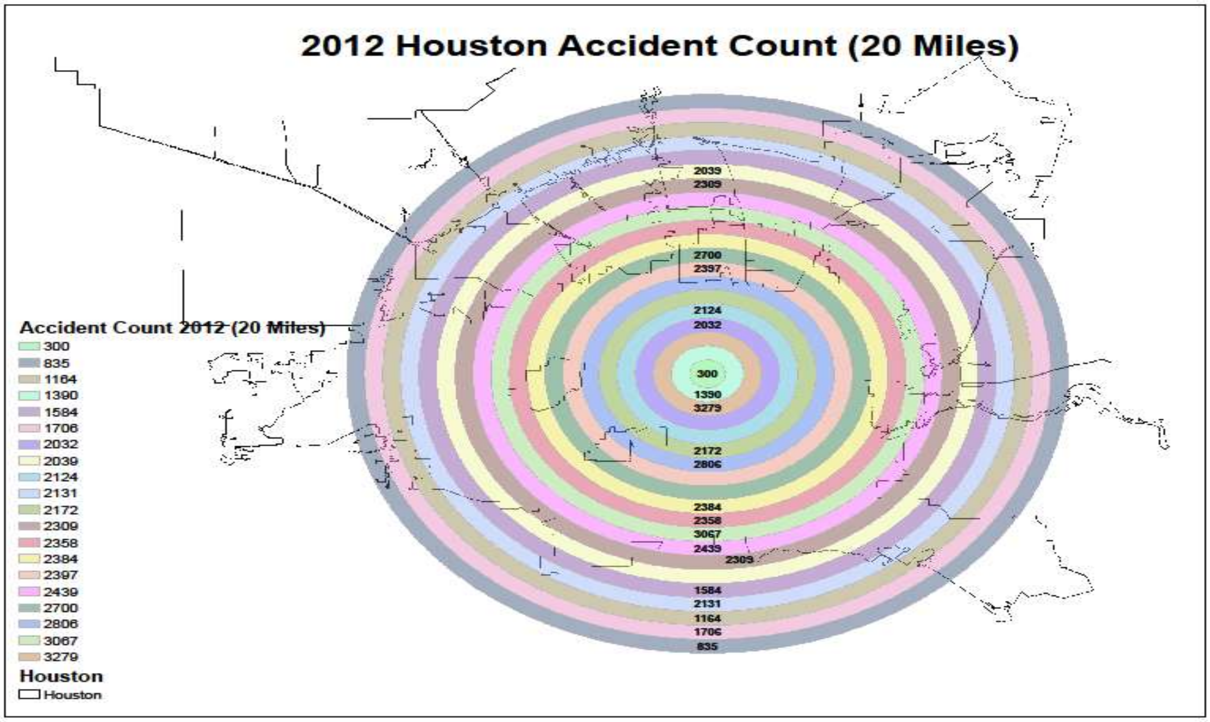


Figure 4.1. Maps for Houston's Accident Counts (20-Mile Radius) at 1-Mile Intervals, 2010-2016.



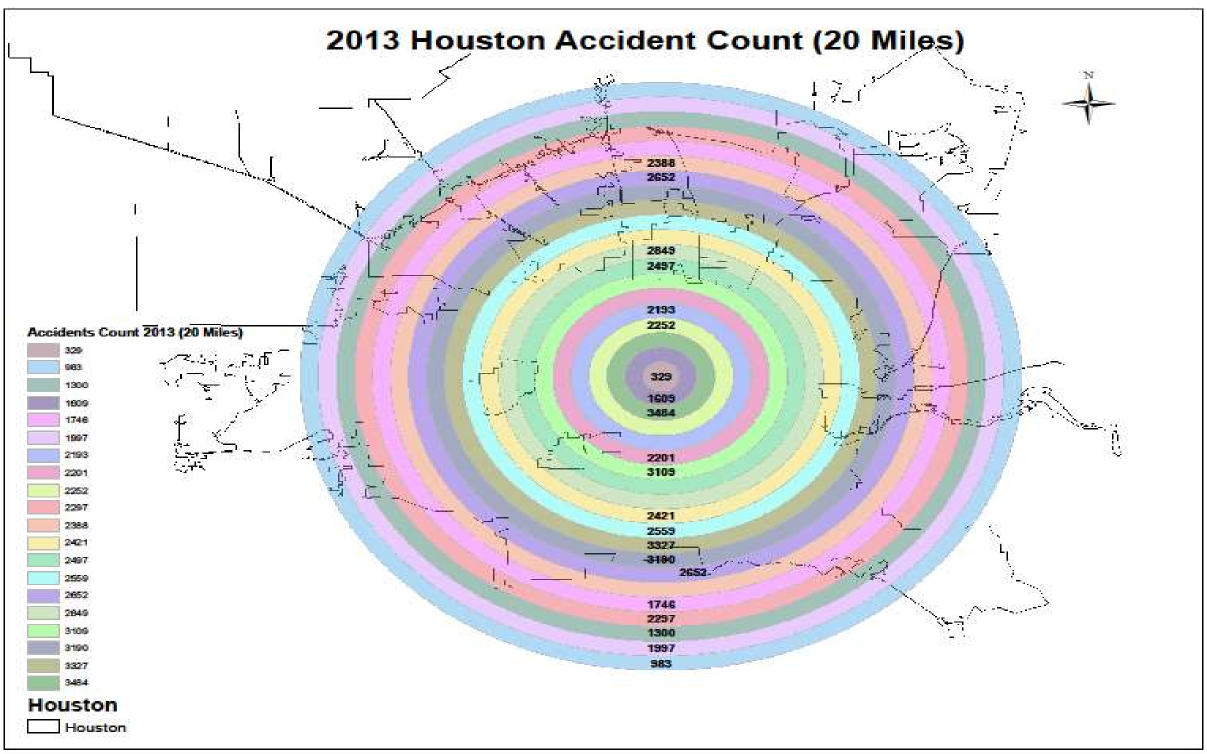
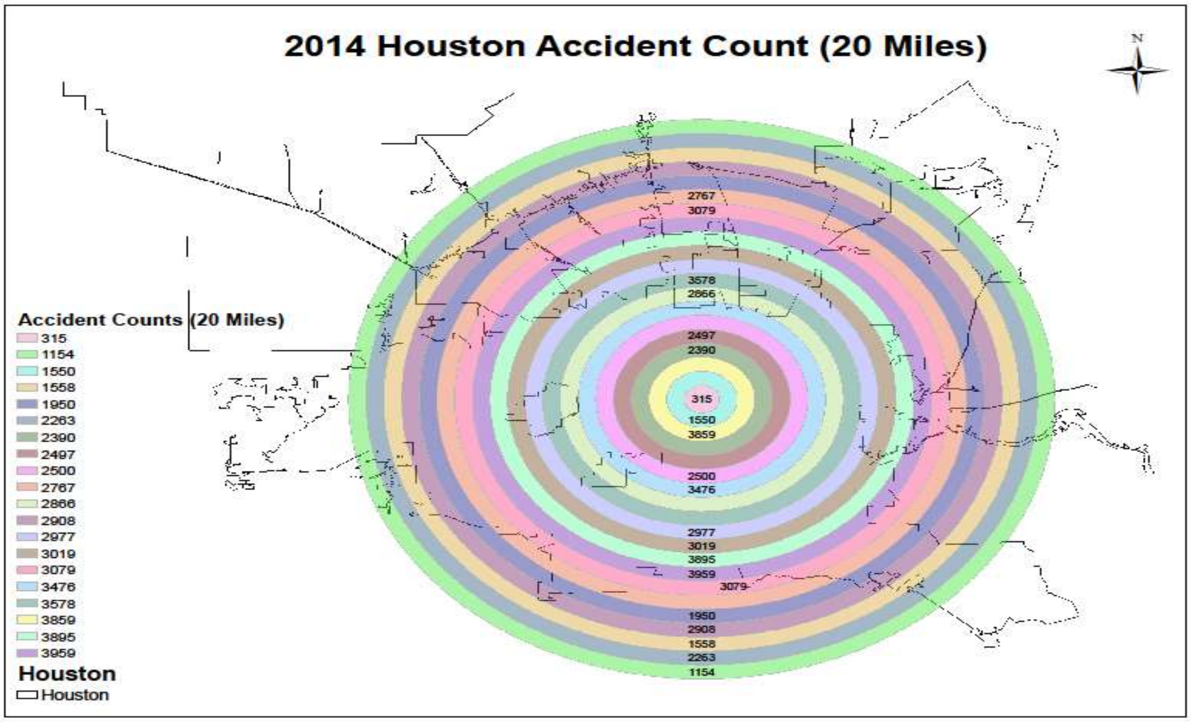


Figure 4.1. Maps for Houston's Accident Counts (20-Mile Radius) at 1-Mile Intervals, 2010-2016. (Continued)



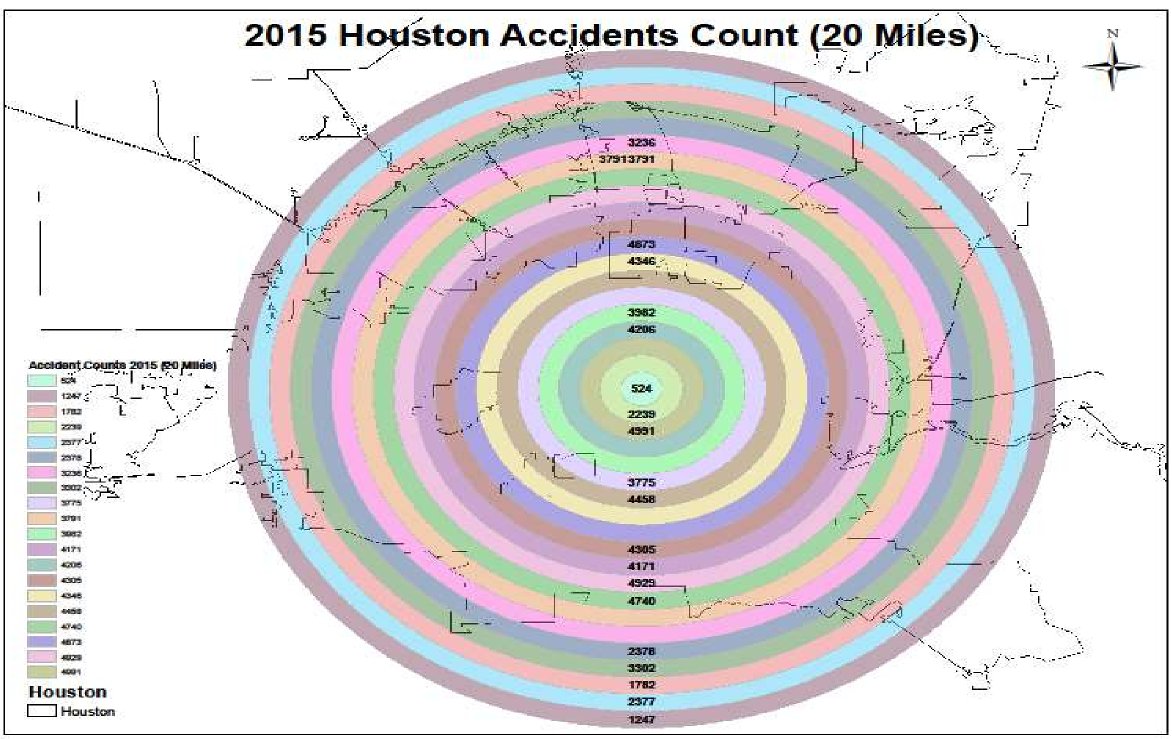


Figure 4.1. Maps for Houston's Accident Counts (20-Mile Radius) at 1-Mile Intervals, 2010-2016. (Continued)

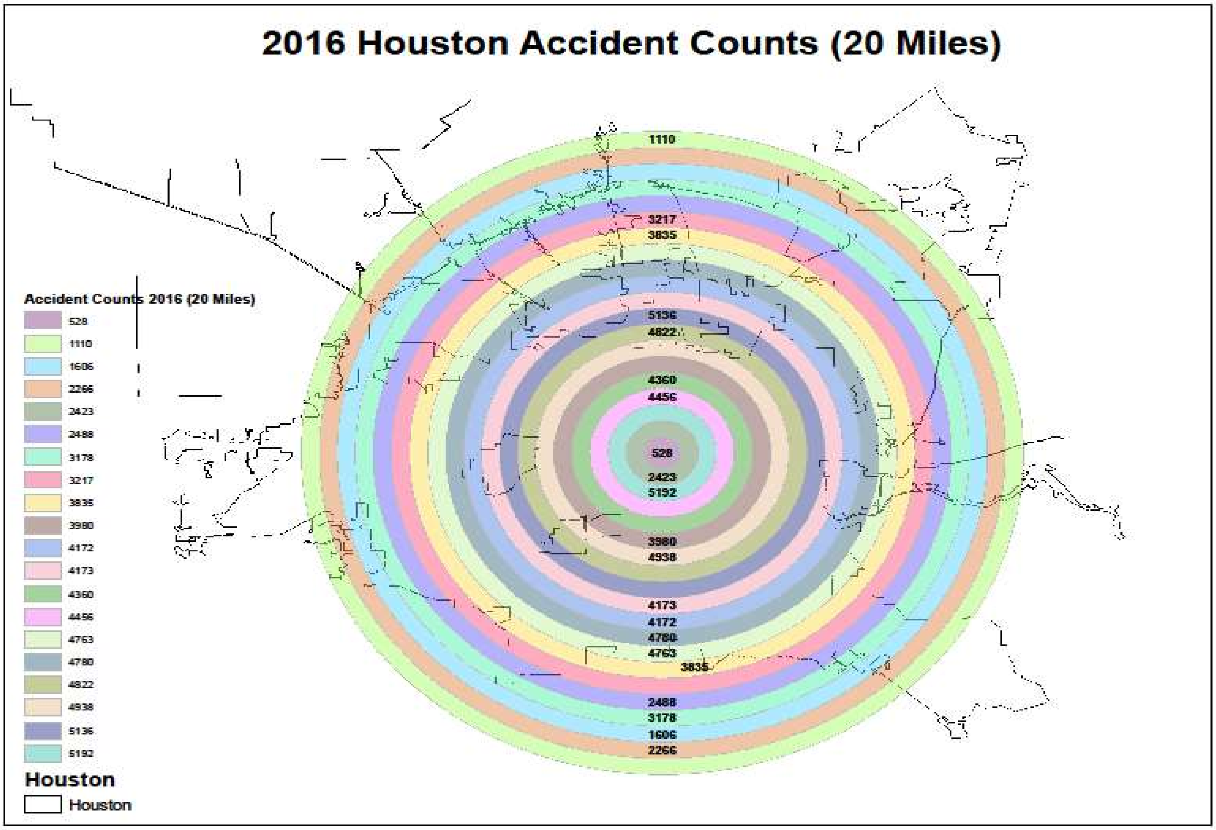


Figure 4.1. Maps for Houston's Accident Counts (20-Mile Radius) at 1-Mile Intervals, 2010-2016. (Continued)

Figure 4.1 provides the accident counts that are needed to create the graphs that evaluate traffic-accident trends in Houston. Figure 4.1 gives an elaborate picture for each year's accident trend and illustrates how the patterns have changed over time. The graphs are plotted between the accident counts for each year and radial distance interval given in miles. Figure 4.2 shows the graphs obtained by plotting the values.

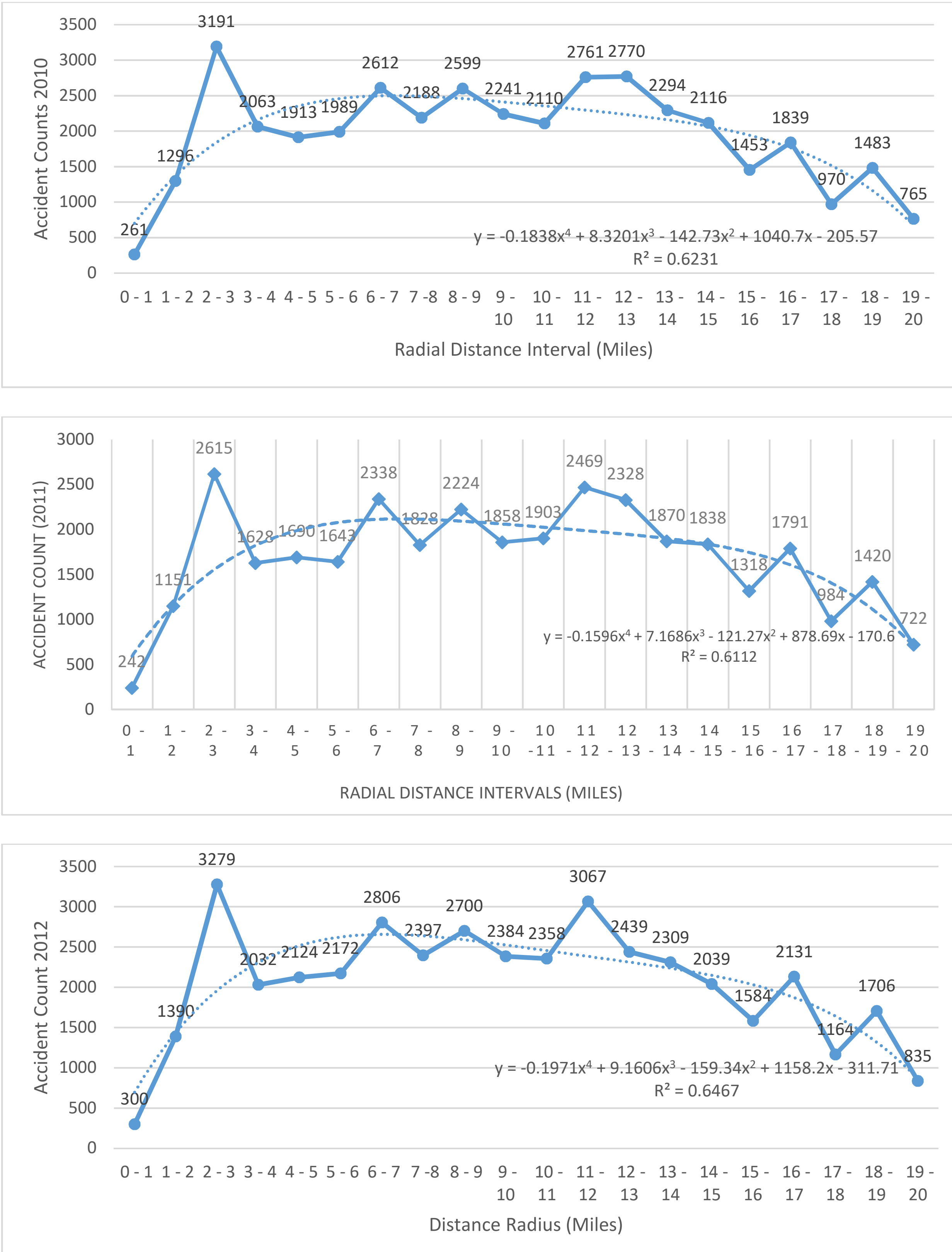
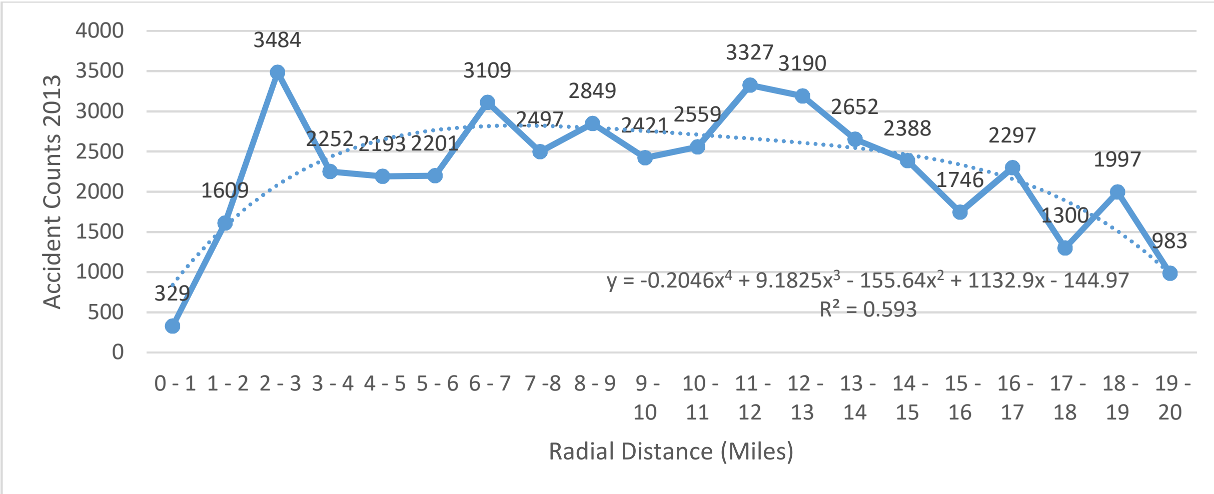
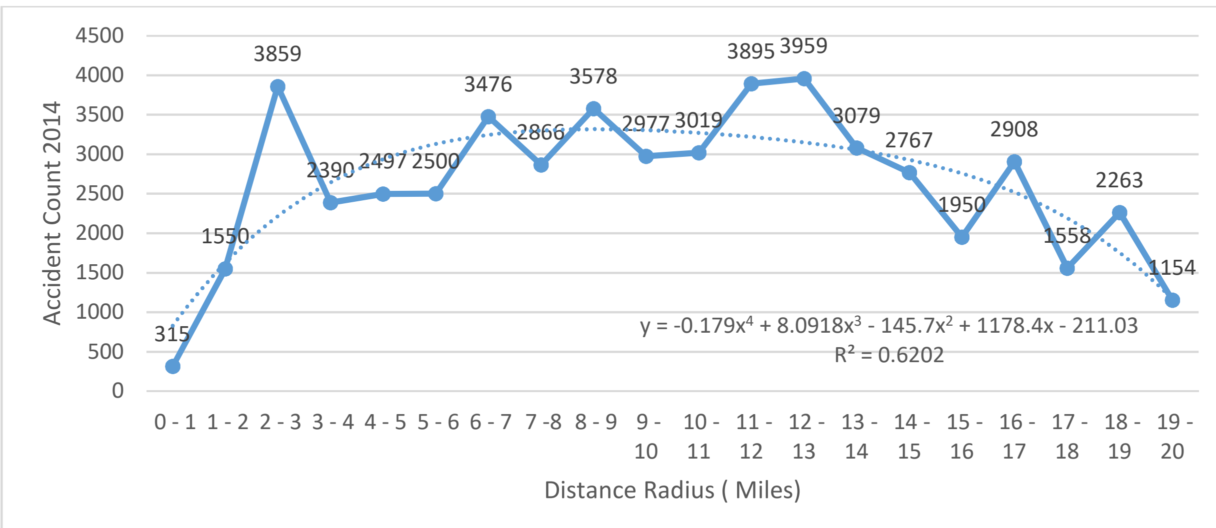


Figure 4.2. Houston's Accident Counts (20-Mile Radius at 1-Mile Intervals), 2010-2016.





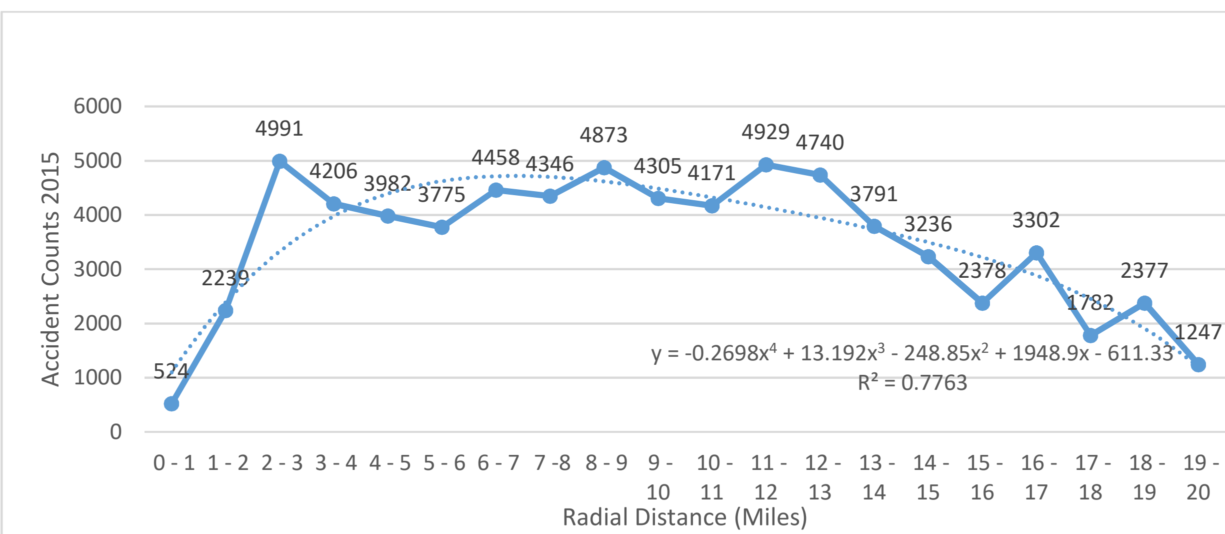


Figure 4.2. Houston's Accident Counts (20-Mile Radius at 1-Mile Intervals), 2010-2016. (Cont.)

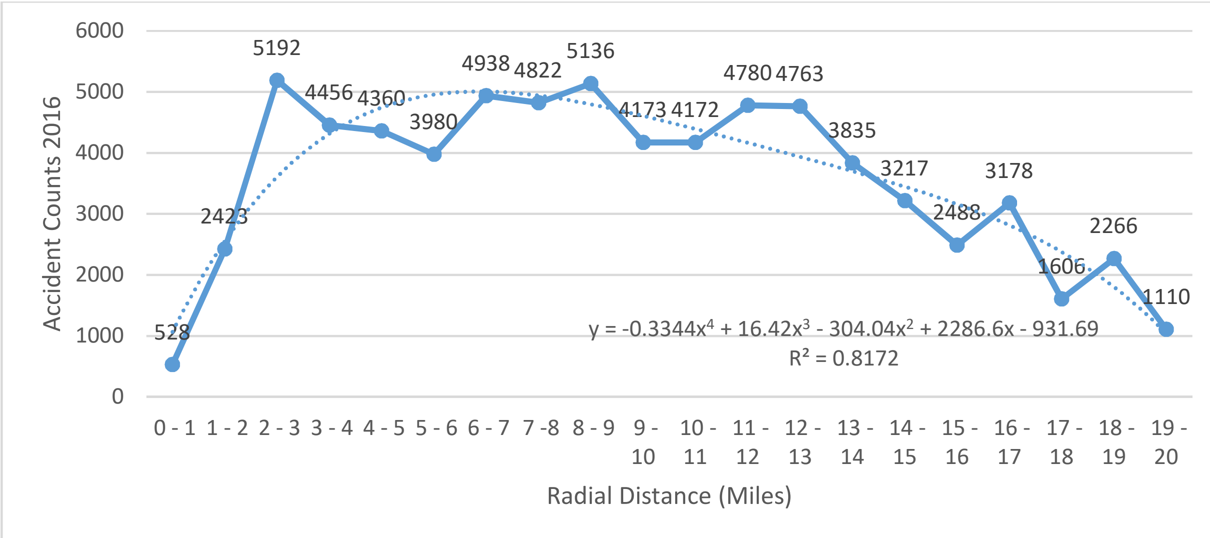


Figure 4.2. Houston's Accident Counts (20-Mile Radius at 1-Mile Intervals), 2010-2016. (Cont.)

Figure 4.2 displays the auto-fit smooth curves for each year's accidents. From the graphs, a rising trend of accidents is seen at the beginning and is, later, skewed towards the right. The skewness in Figure 4.2 for each year's accidents is slight or significant for some of the plotted values. The curve-fit equations and the R^2 values are also plotted and displayed in Figure 4.2. The R^2 values range from 0.6 to 0.8, which is significantly useful. The trend in Figure 4.2 demonstrates that the hotspot for most accidents is within 2-3 miles of the city's centroid. However, this trend has varied over time and presents a significant change with increased values for accidents 8-9 and 11-13 miles from the city's centroid. The highest value for traffic crashes is found at the 2-3-mile interval for each year's accident graph, except 2014 where the highest number of accidents is 12-13 miles from the centroid. Figure 4.1 suggests that these intervals have the most accidents due to the presence of major intersections as well as vehicles entering or exiting the city's center.

4.3. Kernel Density

4.3.1. Introduction

Kernel density is a spatial-analysis tool in ArcGIS that uses the kernel function to calculate the neighborhood density from the given point features or polyline. The kernel-density function uses a magnitude-per-unit-area calculation from each data point and results in a smooth, continuous surface that categorizes the density values. Kernel density is a handy statistical tool to convert the data into a continuous surface in order to offer locations that have a high-density value. The high-density aids with classifying areas which are significantly different from the other. Hence, the critical places that are identified and targeted can be studied and investigated for any purpose. In this study, the density values depict the magnitude, or intensity, for a sum of accidents in Houston. The high-density values indicate the locations which are hotspots for accidents and are critical for authorities to study and investigate. More reinforcement and precautionary measures from the authorities can facilitate the reduction of accidents in the identified hotspots. Moreover, the trend for each year can be examined to identify the growth of accidents in a specific direction.

4.3.2. Procedure

The following inputs are required to utilize the kernel-density tool:

i. Locate the point or polyline feature.

ii. Provide the population field; if population field has not provided, then select NONE. The population field defines the volume under the surface. If NONE is selected, the field assumes the value of 1 and for point data as 1.

iii. Locate the output for the raster data.

iv. The tool automatically detects the area unit, depending on the map's unit and the coordinate system.

v. The rest of the functions are optional and can be used, depending on the given dataset.

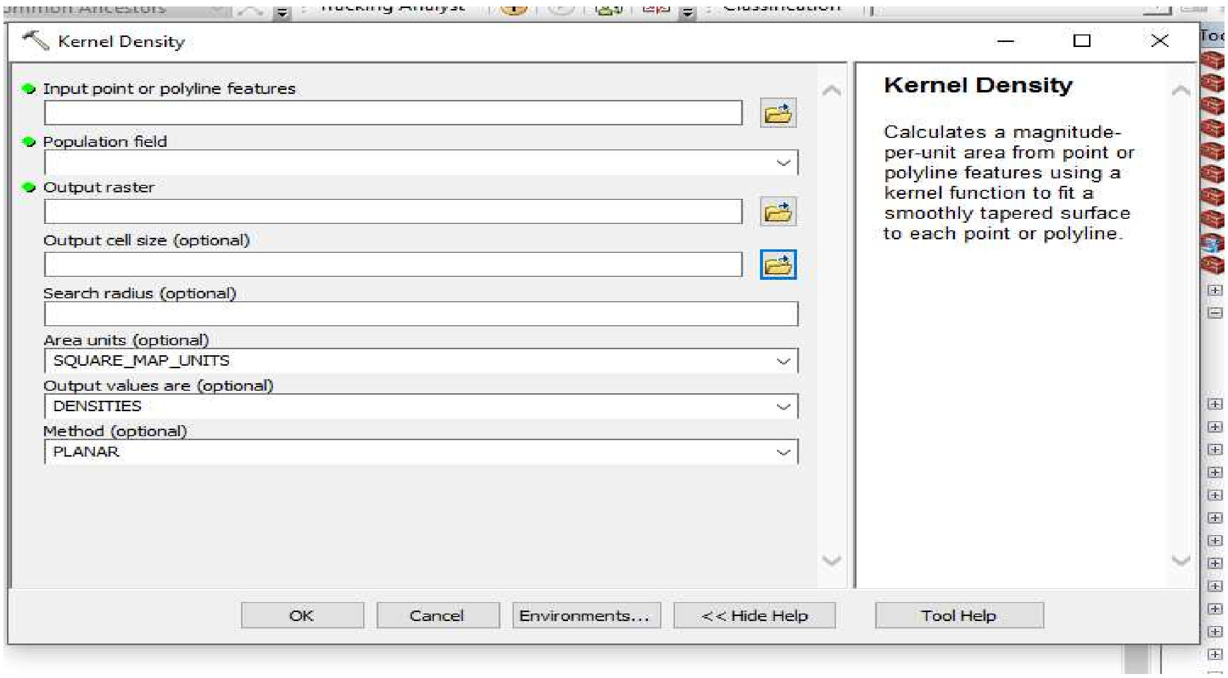


Figure 4.3. Framework for the Kernel-Density Tool in ArcGIS.

Figure 4.3 gives a better picture of the kernel-density tool in ArcGIS. The picture displays the framework to perform the kernel density. The tool is situated in the Spatial Analyst tool under the Density category of ArcGIS.

The tool requires the calculation of the default search radius, which is also known as bandwidth. The search radius is calculated with the following formula:

$$search\ radius = 0.9 * min\left(SD, \frac{\sqrt{1}}{\sqrt{\ln(2)}} * DM\right) * n^{\wedge} - 0.2 \qquad \text{(Eq.4.1)}$$

Where,

SD = Standard Distance

DM = Median Distance

N = Number of points if no population field used. If the population field is used, N is the population-field values.

In formula, the minimum (min) tells that the equation uses the minimum value result from two options inside the bracket. By using equation, the ArcGIS tool calculates the points within search radius from the accident locations. The surface value is highest at the accident points, decreasing if moving away from the points and reaching zero at the search-radius distant from the point.

4.3.3. Results

After performing the kernel density, the results are presented in the form of the maps given in Figure 4.4. The following are the Kernel Density Analysis results for the Houston city:

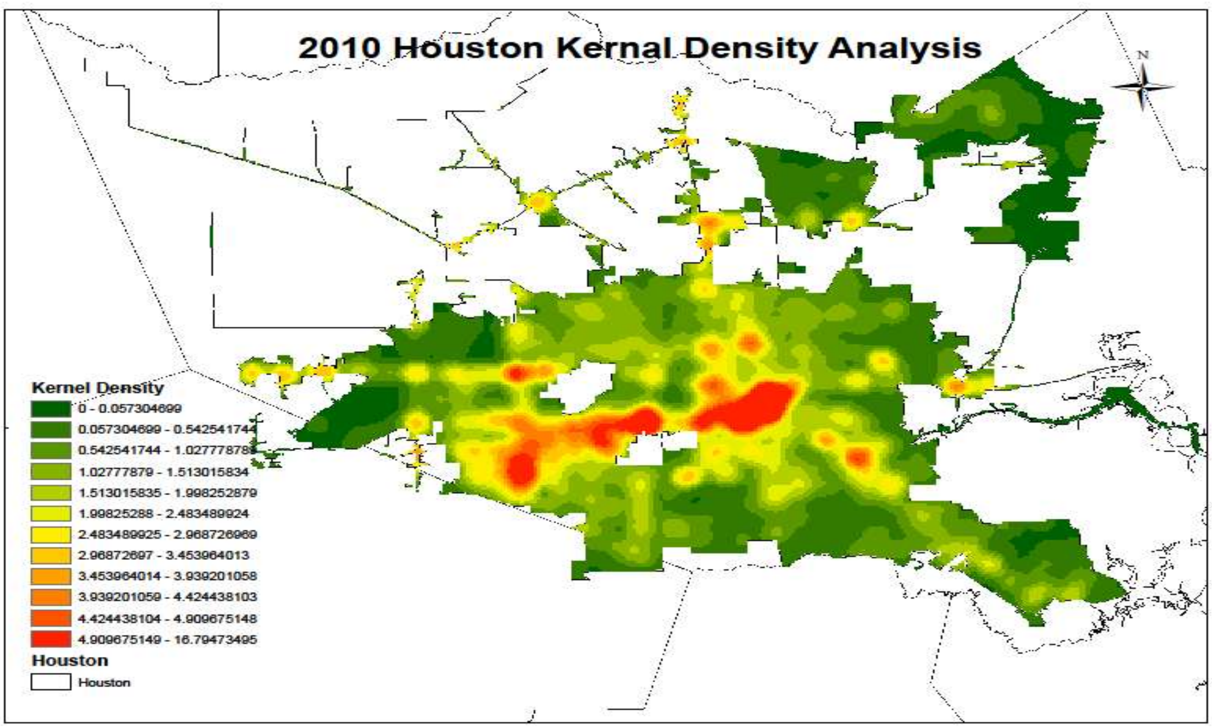
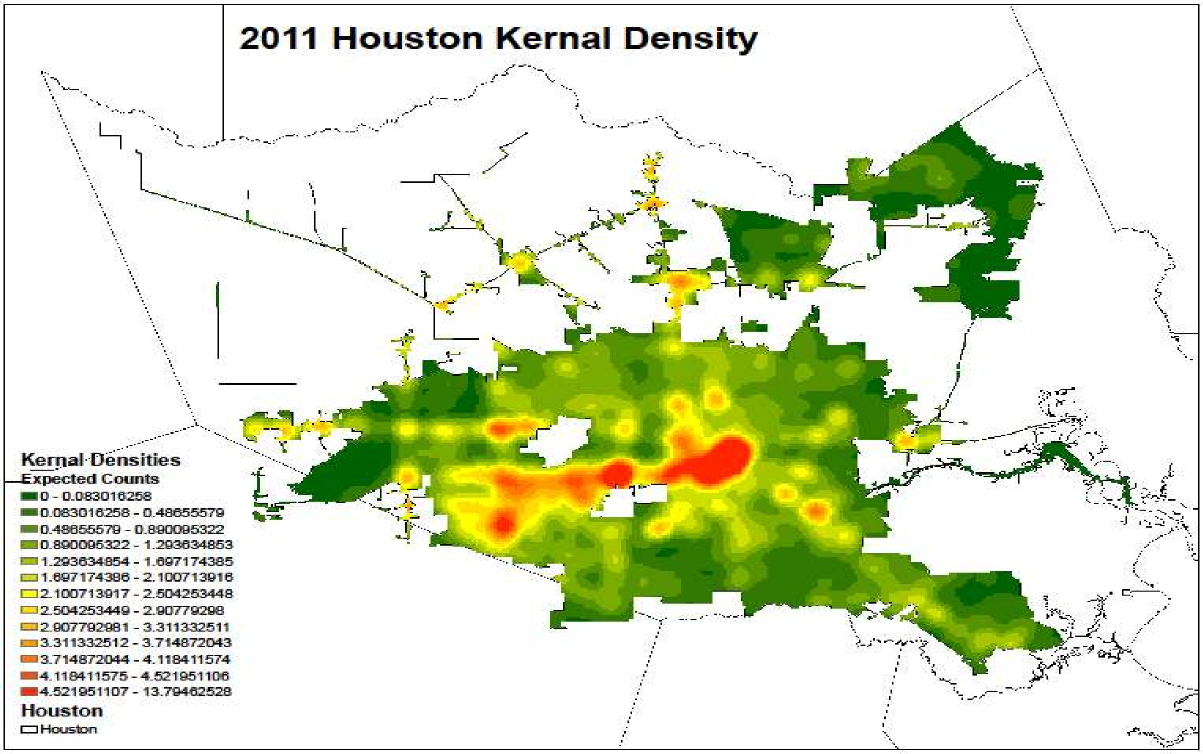


Figure 4.4. Houston’s Kernel-Density Analysis Maps, 2010-2016.



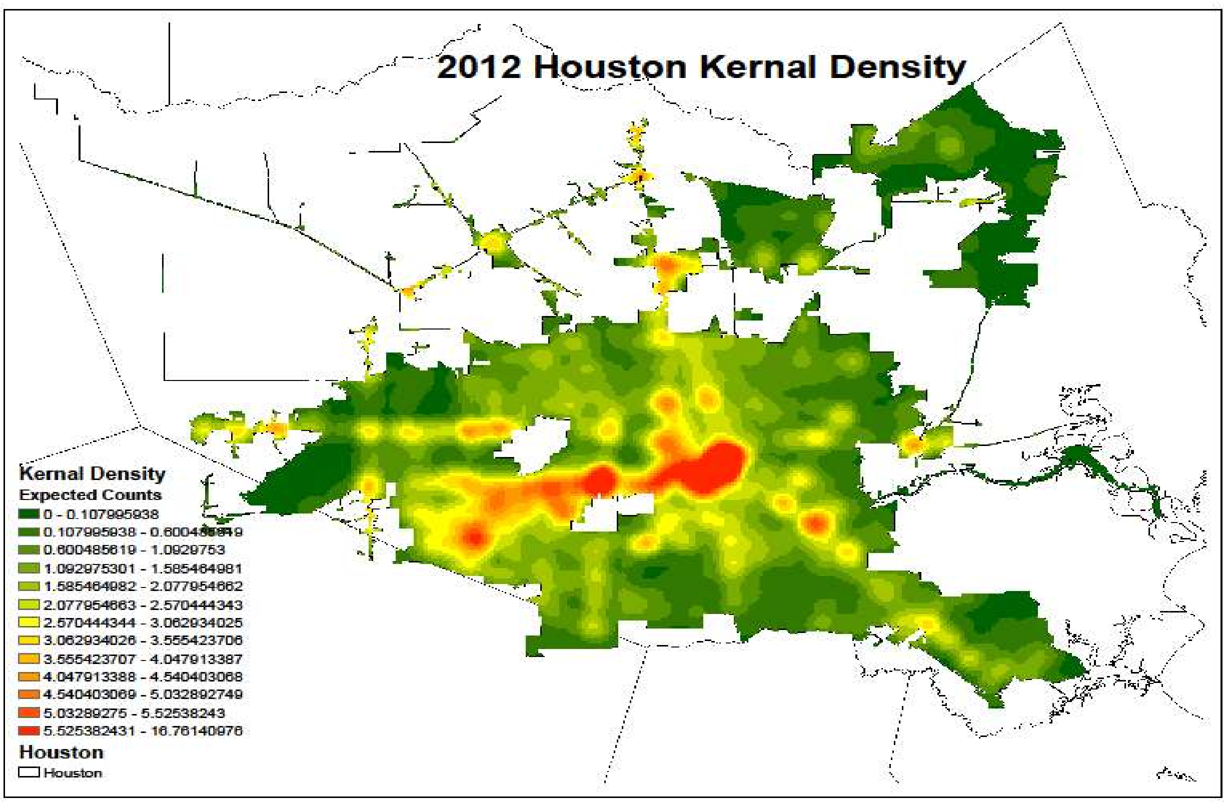


Figure 4.4. Houston's Kernel-Density Analysis Maps, 2010-2016. (Continued)

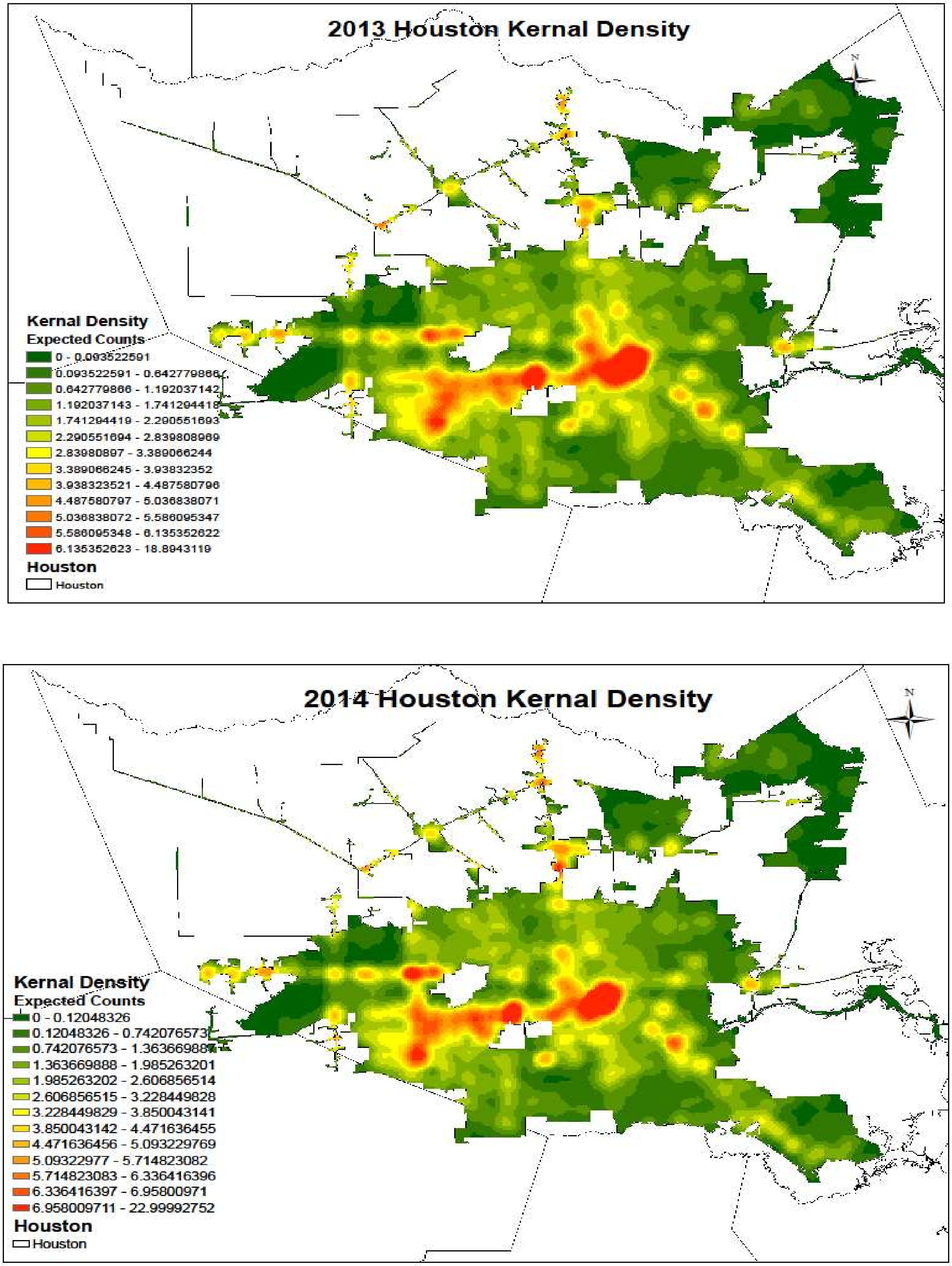
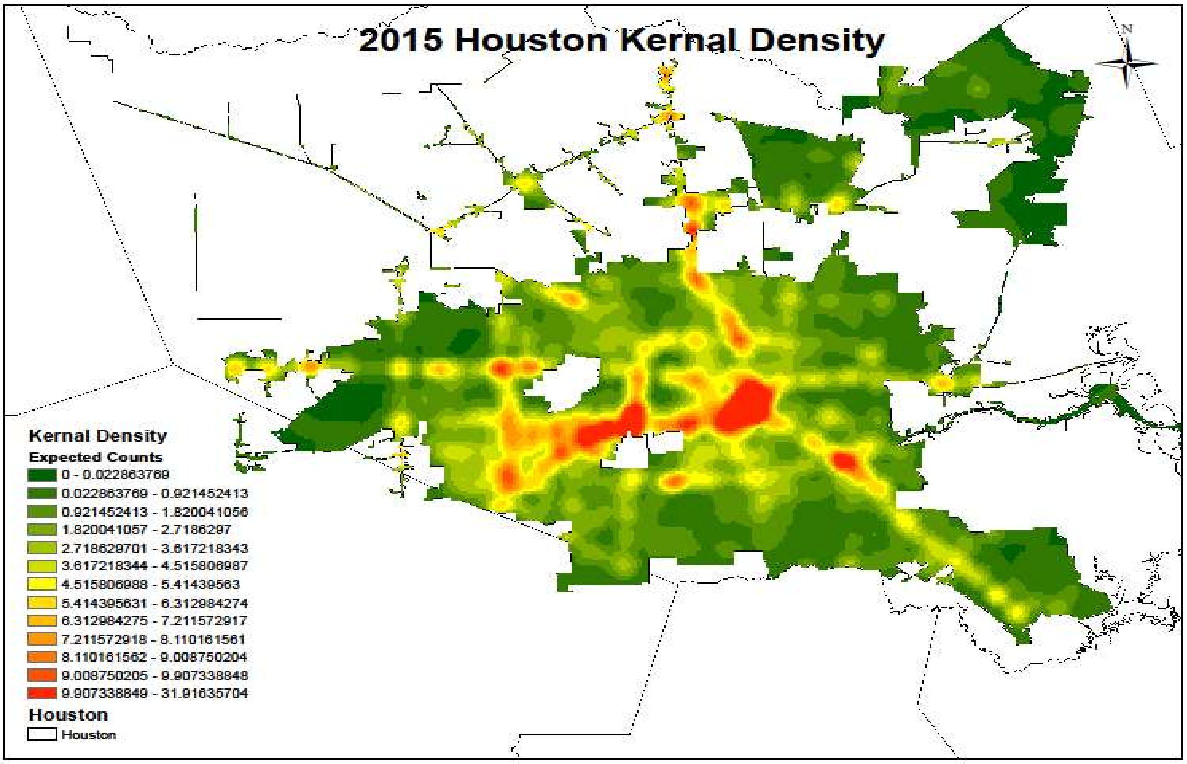


Figure 4.4. Houston's Kernel-Density Analysis Maps, 2010-2016. (Continued)



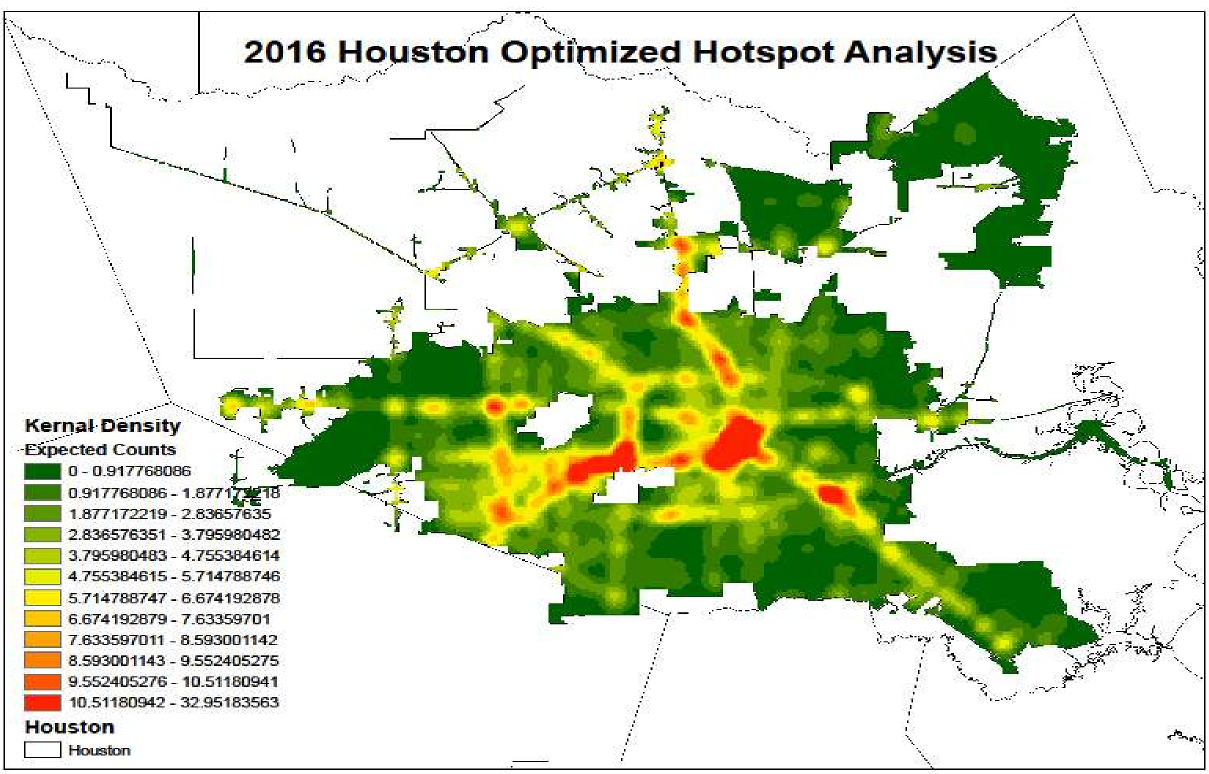


Figure 4.4. Houston's Kernel-Density Analysis Maps, 2010-2016. (Continued)

The produced kernel-density maps, shown in Figure 4.4, display information about the expected counts for accidents in Houston. By looking at Figure 4.4, the output suggests that the highest-expected traffic-accident count areas are located at the city's center and spread towards the west side. The map for each year illustrates that dense accident locations have grown in the city's north and the southeast directions with the passage of time. The high-accident trend draws a polyline structure network from the city's center, suggesting that the densest locations are at the intersections and main highways. The highly dense areas are marked with red. The densest regions ranged from 16-32 expected accidents each year from 2010-2016. The urban areas have low dense-accident sums, and those locations are marked with green on the maps. Figure 4.4 identifies the spatial locations where the accidents have occurred the most during the period. Moreover, necessary measures can be taken to alleviate accidents by using the given information. Table 4.1 details the statistical information about the maps which resulted from the kernel-density tool shown in Figure 4.4.

Table 4.1. Statistical Values for the Kernel-Density Maps.

Year	Min. Value	Max. Value	Mean	Stand. Deviation
2010	0	16.794	1.2703	1.4557
2011	0	13.794	1.0918	1.2106
2012	0	16.761	1.339	1.4774
2013	0	18.894	1.4667	1.6477
2014	0	22.999	1.6744	1.8647
2015	0	31.9163	2.2693	2.6957
2016	0	32.9518	2.3568	2.8782
Average	0	22.0157	1.6383	1.8900

4.4. Optimized Hotspot Analysis

4.4.1. Introduction

The Optimized hotspot analysis is a spatial-statistics tool that creates a map of hot- and cold-spots using the Getis-Ord Gi* statistics from the given incident point or weighted data. The tool uses the optimal settings which are derived from the characteristics of provided accident data and are automatically adjusted for spatial dependency and multiple testing when using the False Discovery Rate (FDR) correction method. In the density analysis, the density provides the cluster locations for the dataset, whereas the hotspot analysis calculates the statistical significance of the clusters. The tool uses the aggregation method to count the incidents within the aggregate grids. There are three possible methods of data aggregation for this tool. At least 30 data points are needed to perform this statistical tool efficiently. The resulting map provides the z and p values for each incident dataset aggregated in a grid. A feature's high Z-score and small P-value indicate a significant hotspot location. A low negative Z-score and a small P-value indicate a significant cold spot. The higher (or lower) the Z-score, the more intense the clustering. A Z-score near zero means that there is no spatial clustering. The Optimized hotspot analysis is used for each year's traffic-accident data to provide the statistically important hotspot locations for Houston. The constructed maps also help to identify the trend and to evaluate the progress for each year's accidents in Houston.

4.4.2. Procedure

The tool uses the following input components to perform the analysis:

i. Provide the input point or polygon feature.

ii. Locate the output feature that provides the z and p values as well as the Gi_Bin results.

iii. Provide the analysis field from the data's characteristics. This function is optional.

iv. Select the incident-data aggregation method to create the weighted feature for analysis. This field is optional; however, the aggregation methods may cause the outcome to be different.

v. The rest of the functions are optional.

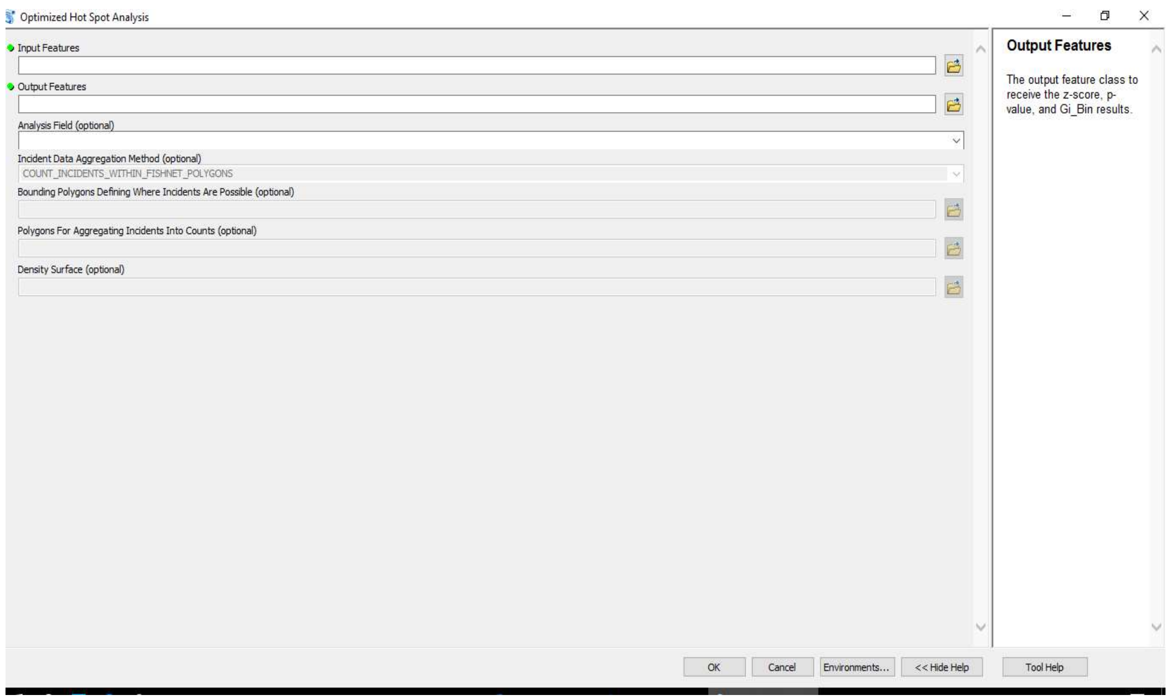
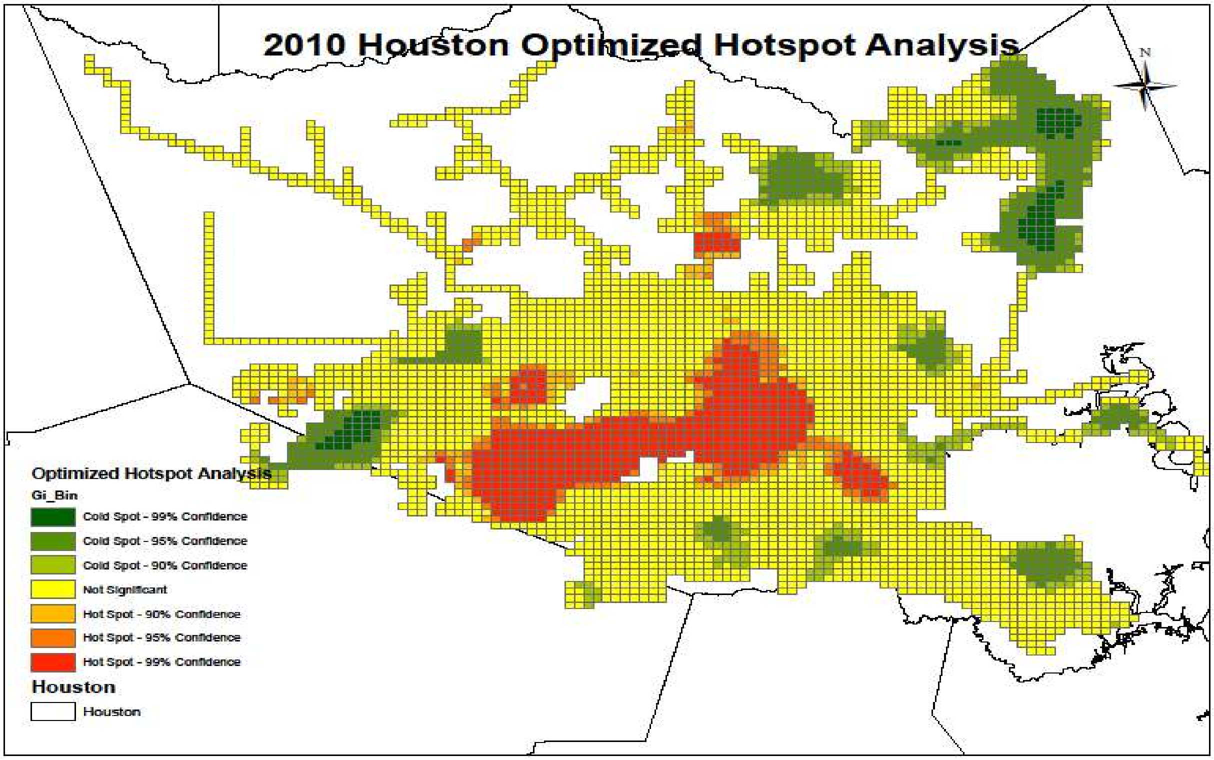


Figure 4.5. Framework for the Optimized Hotspot Analysis Tool in ArcGIS.

Figure 4.5 shows the framework for the optimized hotspot analysis tool and displays each field options require to perform the spatial statistic tool. This ArcGIS tool is situated in the spatial statistics tool under the Mapping Cluster category in the Arc Toolbox of ArcGIS.

4.4.3. Results

Figure 4.6 shows the outcome maps for Houston after the Optimized hotspot analysis tool in ArcGIS is used. The resulted maps from Optimized hotspot analysis tool are given:



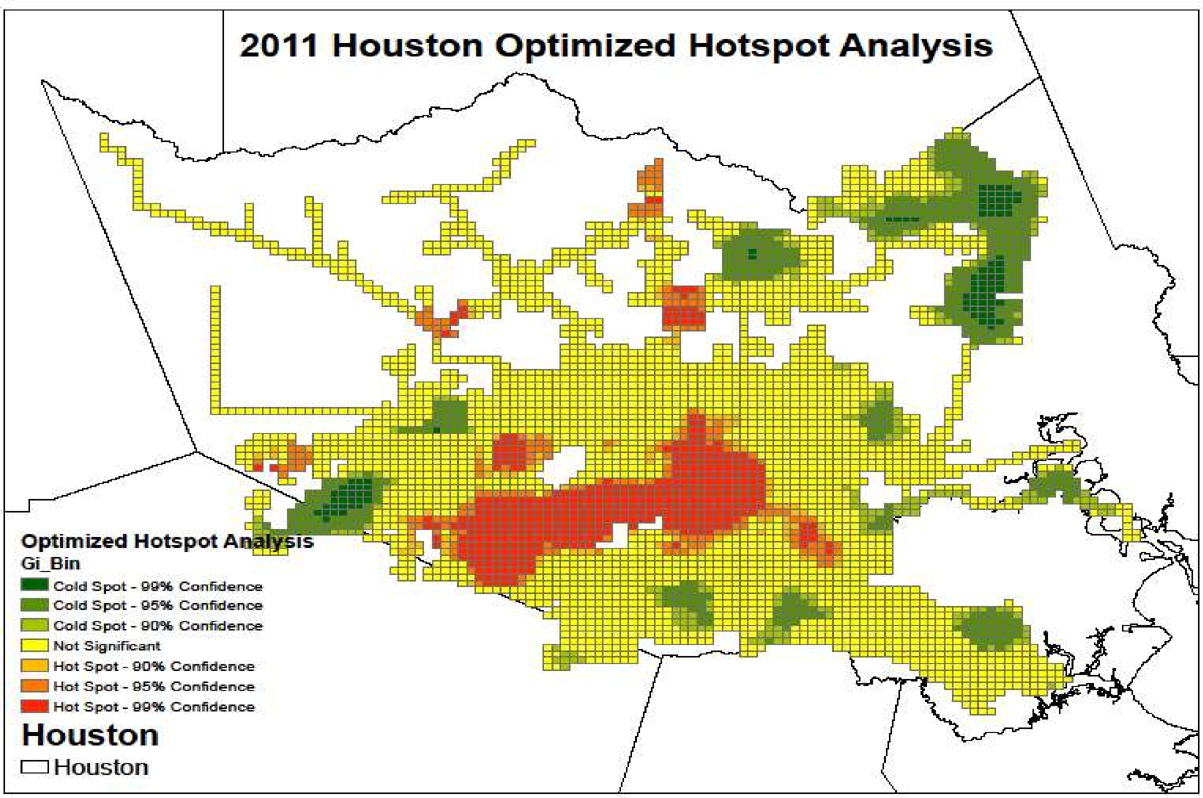
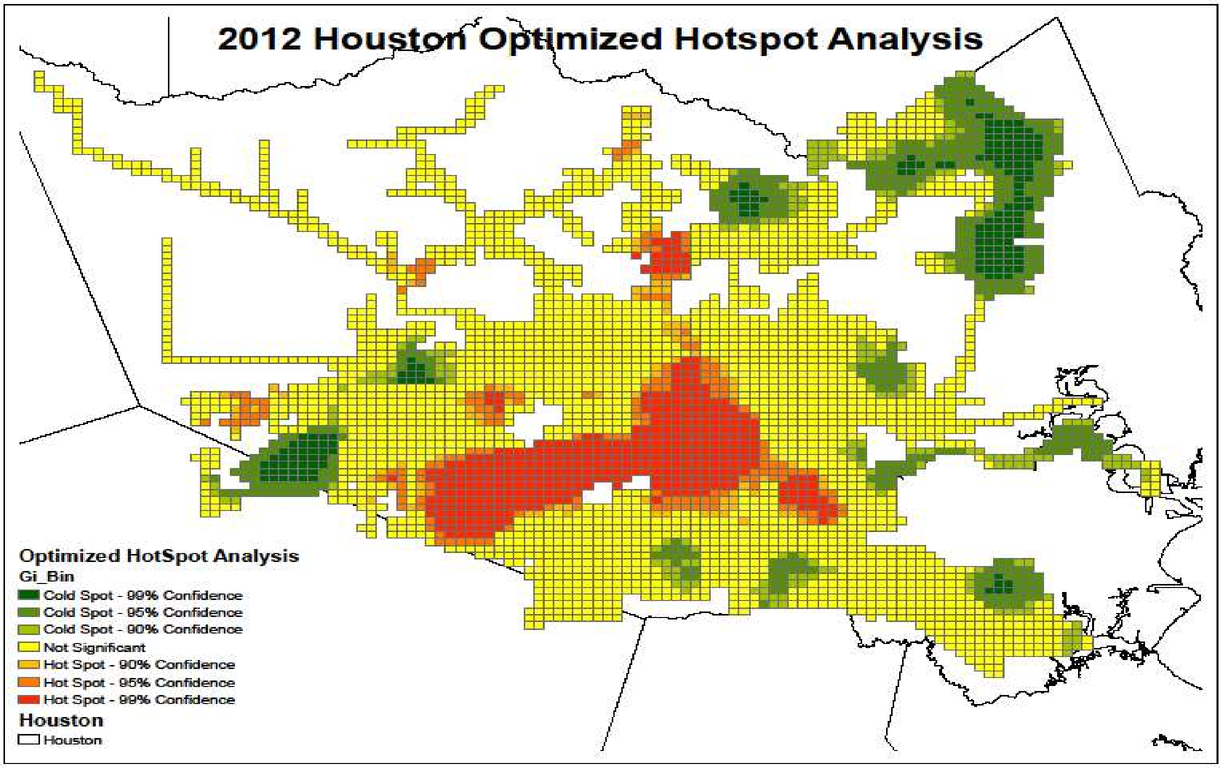


Figure 4.6. Houston's Optimized Hotspot-Analysis Maps, 2010-2016.



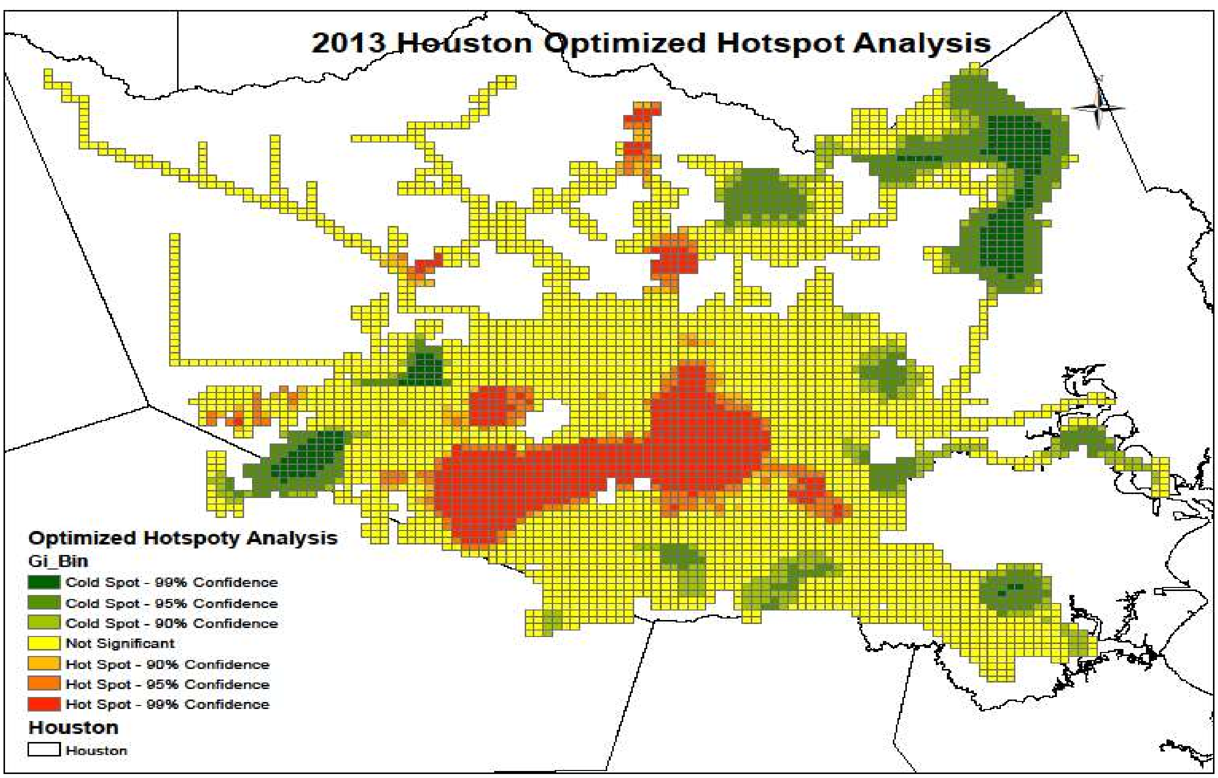
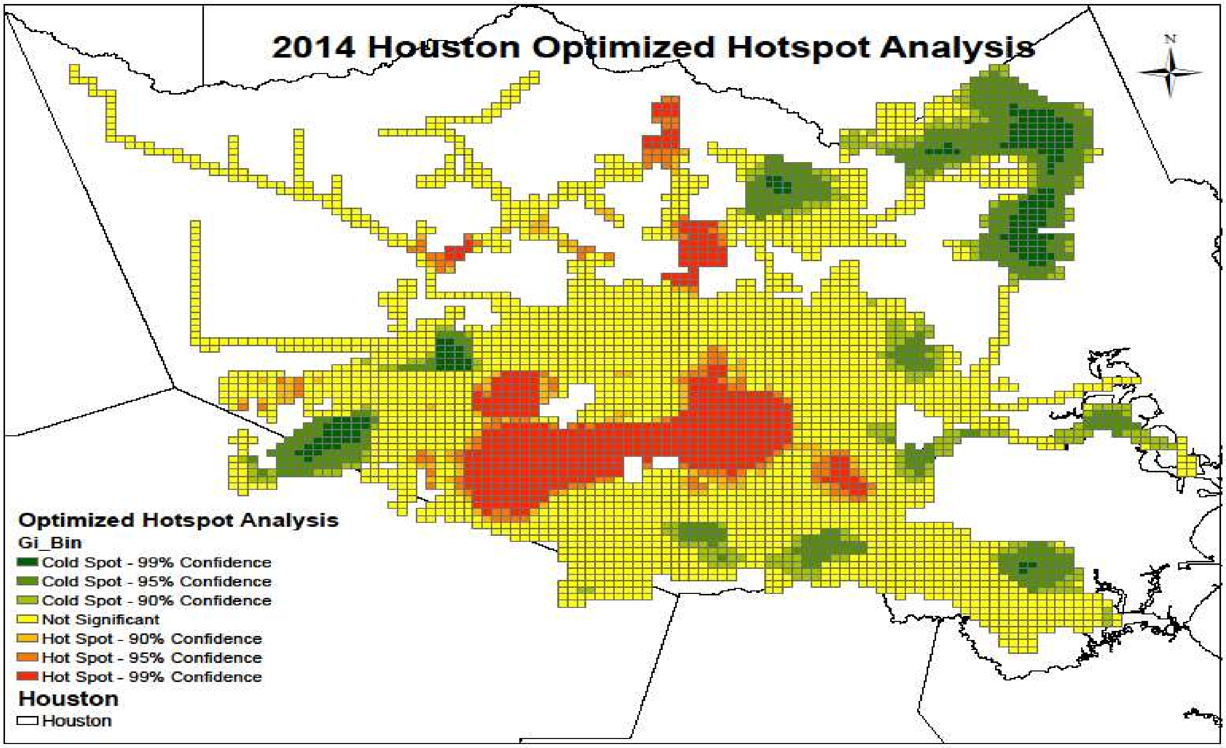


Figure 4.6. Houston's Optimized Hotspot-Analysis Maps, 2010-2016. (Continued)



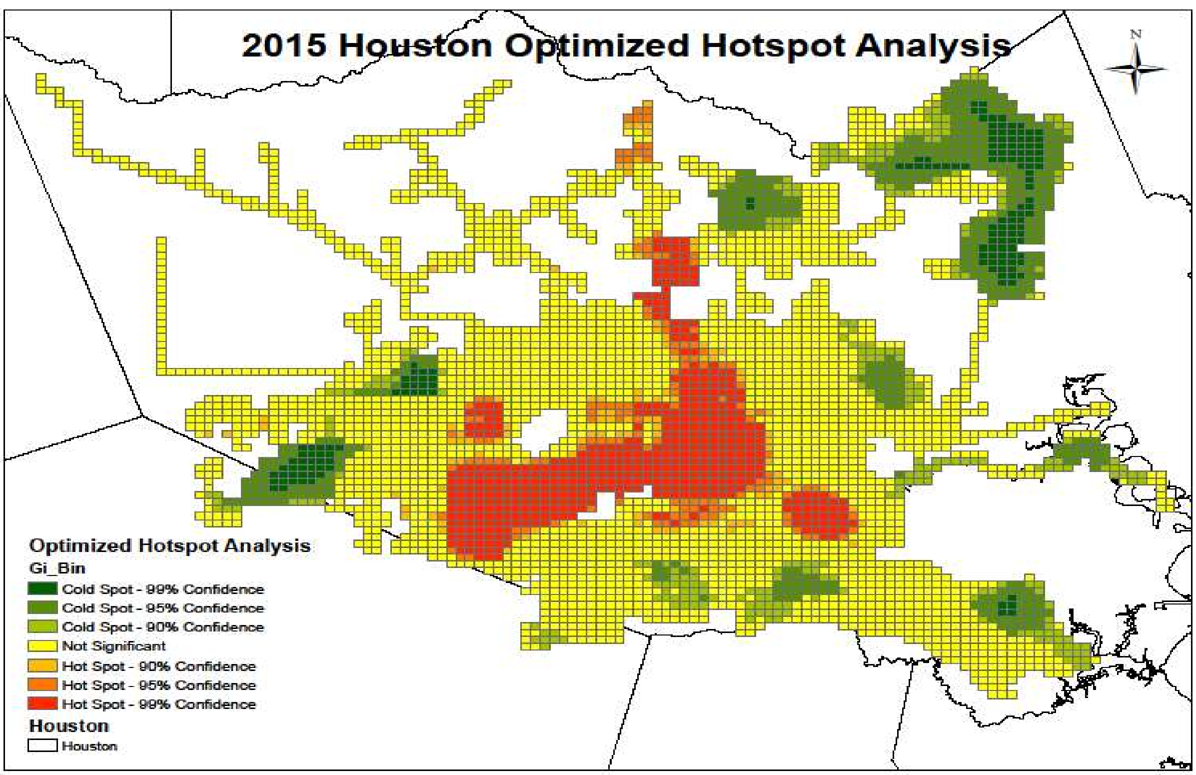


Figure 4.6. Houston's Optimized Hotspot-Analysis Maps, 2010-2016. (Continued)

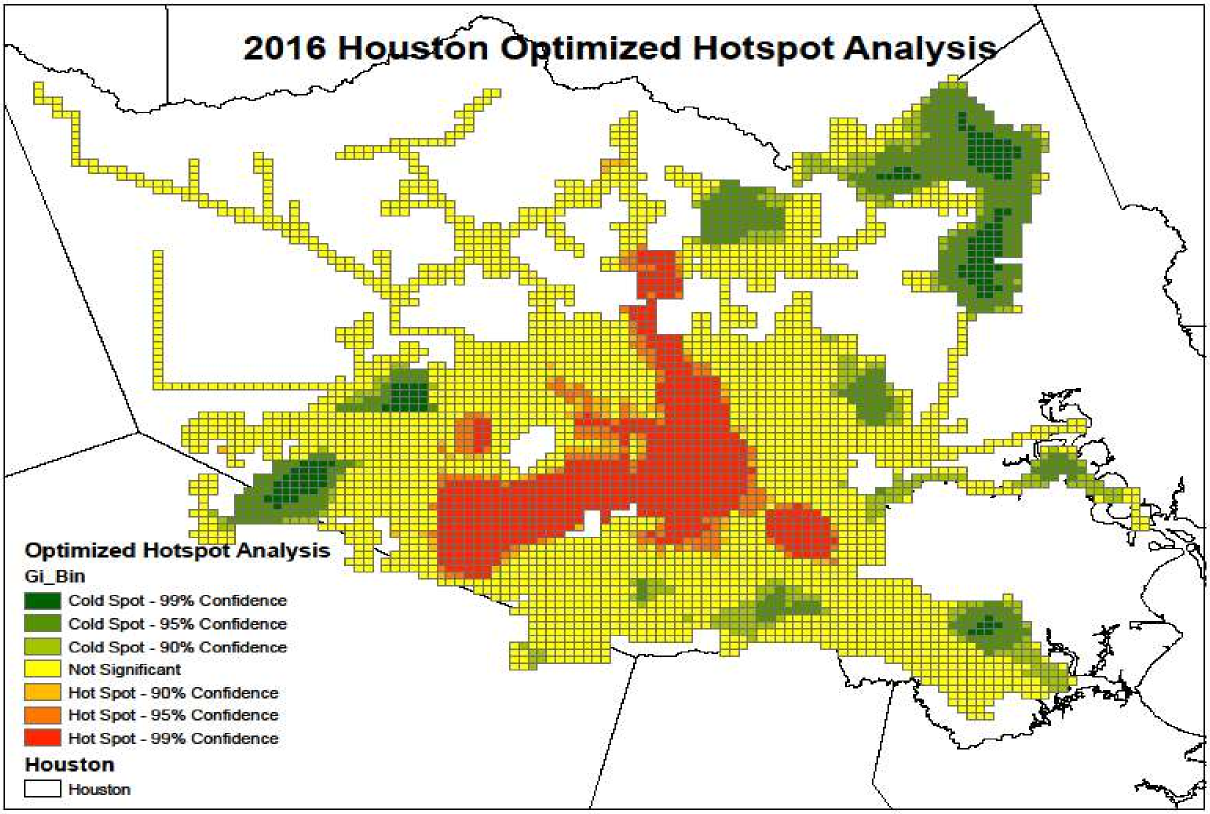


Figure 4.6. Houston's Optimized Hotspot-Analysis Maps, 2010-2016. (Continued)

Figure 4.6 displays the hotspots, cold spots, and non-significant values for Houston's accident counts in each fishnet polygon. The hotspot values are presented in red color; the cold spots are green; and the non-significant locations are yellow. The hotspot polygons provide that the value of z is high and that the p-value (90-99% confidence interval) is small, making those polygons significant. From the information, we can conclude that these polygons are significantly dissimilar from the overall neighborhood values in Houston. Hence, the accident rate accounted for in these polygons is high. Similarly, the cold-spot polygons mean that the z value is lower and negative and that the p values (90-99% confidence interval) are small. The cold-spot polygons are significantly different from other polygons, and the cold spots have a lower accident rate than other places. Moreover, non-significant legend in Hotspot maps describes the polygons where trends typically spread with no major change in accident rates. The resulting maps provide the z and p values for each fishnet polygon.

While observing the maps for each year's accident dataset in Figure 4.6, it is imperative that the hotspot locations are identified in the center of Houston's city, which was earlier discovered in maps for kernel density. The important hotspot locations at the city's center and west side remain constant during the study period. However, the trend is significant growth from Houston's center towards the city's north and southwest. Likely, the cold-spot polygons located at the edge of the city limit are recognized as suburban areas. The high number of polygons on the maps is not significant. However, these polygons have diminished with time and have been taken over by the hot- and cold-spot polygons. Tables 4.2 and 4.3 give the ranges and statistical values for the Z-score and P-score for each year's accident data points that were used to identify the hot- and cold-spot locations on resulting optimized hotspot analysis maps.

Table 4.2. Z-Score Statistical Values for Each Year's Accident Dataset from the Optimized Hotspot Analysis Maps.

Year	Min. Value	Max. Value	Mean	Stand. Deviation
2010	-3.498419	21.455481	0.183559	3.076772
2011	-3.550085	20.911537	0.163045	3.00717
2012	-3.753729	21.378704	0.19618	3.138704
2013	-3.596119	21.52591	0.171079	3.070986
2014	-3.61312	19.283406	0.139802	2.930598
2015	-3.5564	21.010703	0.17057	3.079145
2016	-3.494846	21.285249	0.187022	3.128668
Average	-3.580388286	20.9787	0.1730	3.0617

Table 4.3. P-Score Statistical Values for Each Year's Accident Dataset from the Optimized Hotspot Analysis Maps.

Year	Min. Value	Max. Value	Mean	Stand. Deviation
2010	0	0.999815	0.254333	0.295708
2011	0	0.99925	0.25078	0.29255
2012	0	0.99893	0.242072	0.289072
2013	0	0.999983	0.255302	0.300685
2014	0	0.999796	0.250909	0.294293
2015	0	0.999312	0.247116	0.29864
2016	0	0.999195	0.247796	0.2964
Average	0	0.9995	0.2498	0.2953

The Optimized hotspot analysis outcomes displayed a similar trend as seen in the kernel-density maps. The tools endorsed each other to find results and provided grounds to verify the results from the different tools.

4.5. Space-Time Pattern Mining

4.5.1. Introduction

The space-time pattern mining is a relatively new toolbox that was introduced in ArcGIS. This toolbox incorporates time as the model's third dimension with the Cartesian coordinate system. By integrating space and time, the tools analyze the data distribution statistically. The toolbox also allows for visualizing the data in both two and three dimensions. There are three tools in this toolbox:

i. Create Space-time Cube
ii. Emerging Hotspot Analysis
iii. Local Outlier Analysis

4.5.1.1 Create Space-Time Cube

The Create Space-time cube tool aggregates the dataset given in space-time to a cube-like structure. The multi-dimensional structure has bins that are defined as several small cubes in a large cube, both horizontally and vertically, which are created with the user's provided information. The information requires to create bins boxes based on the time-interval input and the spatial-grid size. The spatial-grid size depends on the dataset's denseness and closeness. The time-interval input relies on the user input how the dataset is to envision and analyzed. Each bin contains aggregated points which occurred in that space-time interval. Each location has multiple bins that share the spatial extent but are comprised of many temporal extents. The bins at each location provide information about what's been happening at that location over time as well as the time-series information. Later, the space-time cube is used to perform the time-series processes, such as hotspot and local-outlier analysis.

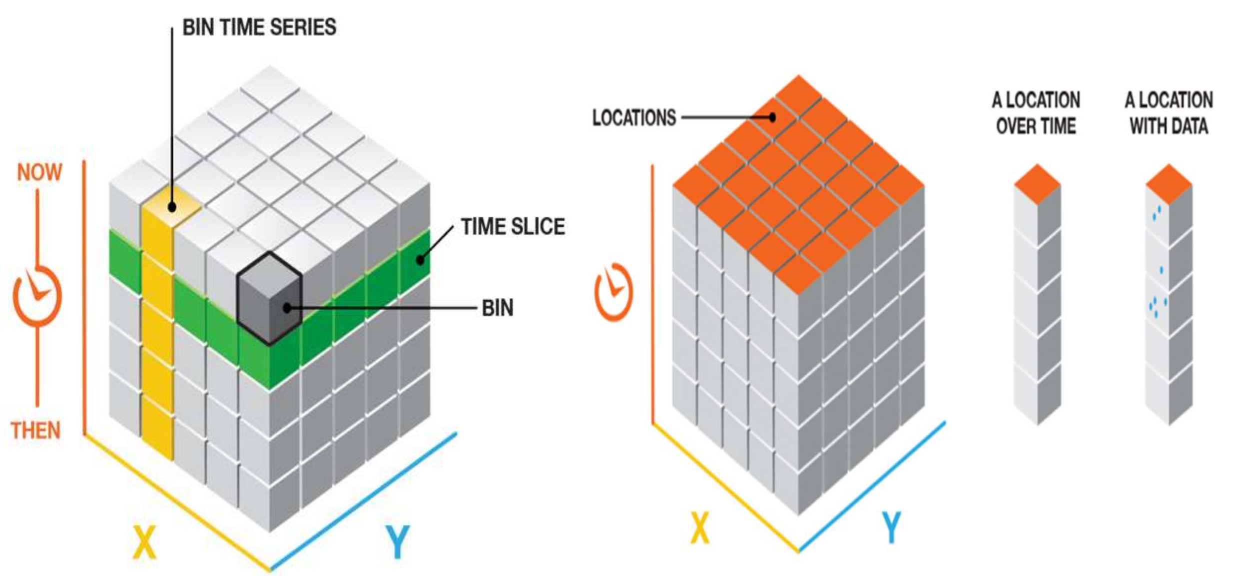


Figure 4.7. Symbolic Representation of the Space-Time Cube Model in ArcGIS. (ArcGIS, 2017)

4.5.1.2. Emerging Hotspot Analysis

The Emerging Hotspot Analysis is a tool in space-time pattern mining that indicates the statistically significant hot- and cold-spot trends over time. As the name suggests, the patterns are analyzed as they emerged over the time intervals. The Emerging Hotspot Analysis provides a surface that indicates the locations with new, intensifying, persistent, or sporadic hotspot patterns at diverse time-step intervals. The space-time cube serves as an input for the Emerging Hotspot-Analysis tool. The trends analyzed for each bin have aggregated points or summary fields in space and time. The tool also provides necessary information about the bins and locations as well as the p and z values in the dataset result output.

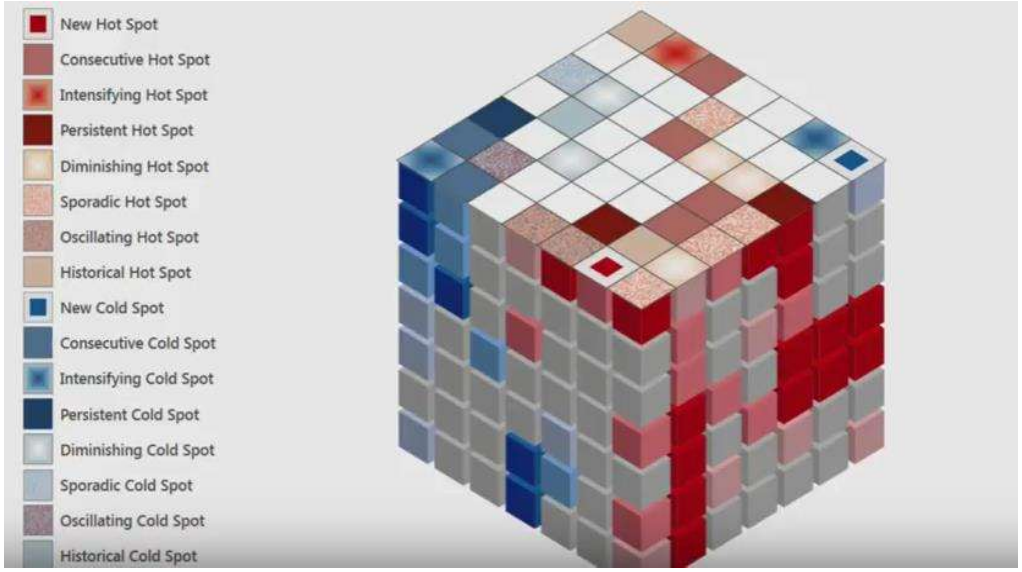


Figure 4.8. Symbolic Representation of the Emerging Hotspot-Analysis Model in ArcGIS. (Prescott, 2016)

Each category shown in Figure 4.8 has the following definitions as described by ESRI:

i. Last time step is hot:

- New: the most recent time-step interval is hot for the first time.
- Consecutive: a single, uninterrupted run of hot time-step intervals, comprised of less than 90% of all intervals.

- Intensifying: at least 90% of the time-step intervals are hot and becoming hotter over time.
- Persistent: at least 90% of the time-step intervals are hot, with no trend up or down.
- Diminishing: at least 90% of the time-step intervals are hot and becoming less hot over time.
- Sporadic: some time-step intervals are hot.
- Oscillating: some time-step intervals are hot, and some are cold.

ii. Last time step is not hot:

- Historical: at least 90% of the time-step intervals are hot, but the most recent time-step interval is not.

iii. Last time step is cold:

- New: the most recent time-step interval is cold for the first time.
- Consecutive: a single, uninterrupted run of cold time-step intervals, comprised of less than 90% of all,
- Intensifying: at least 90% of the time-step intervals are cold and becoming colder over time.
- Persistent: at least 90% of the time-step intervals are cold, with no trend up or down.
- Diminishing: at least 90% of the time-step intervals are cold and becoming less cold over time.
- Sporadic: some time-step intervals are cold.
- Oscillating: some time-step intervals are cold, and some are hot.

iv. Last time step is not cold:

- Historical: at least 90% of the time-step intervals are cold, but the most recent time-step interval is not.

4.5.1.3. Local-Outlier Analysis Tool

The local-outlier analysis tool also uses the space-time cube to determine the clusters that are statistically significant for low and high values as well as the outliers that are significantly different from their neighbors. The tool uses both space and time to identify the locations that are significantly different than their neighbors in the study area. This space-time tool implements the concept of Anselin Local Moran's I statistic. The tool primarily answers the following question: Is this bin significantly different from all other bins, or is the neighborhood significantly distinct from all other neighborhoods? The p and z values for each bin are calculated, and on that basis, the values determine whether the bin is high-high, low-low, high-low, or low-high. These categories represent that the bin is statistically significant, and categories such as high-high signify that the bin is high and is surrounded by high neighborhood values. This information is applicable to all other classes such as low-low and depicts similar meaning to the provided information. The results can be visualized in 2 or 3 dimensions by using the visualizer tool within the space-time toolbox. The 3D visualize tool provides a more elaborate understanding of model by slicing more deeply into the resulting model to analyze the interval trends.

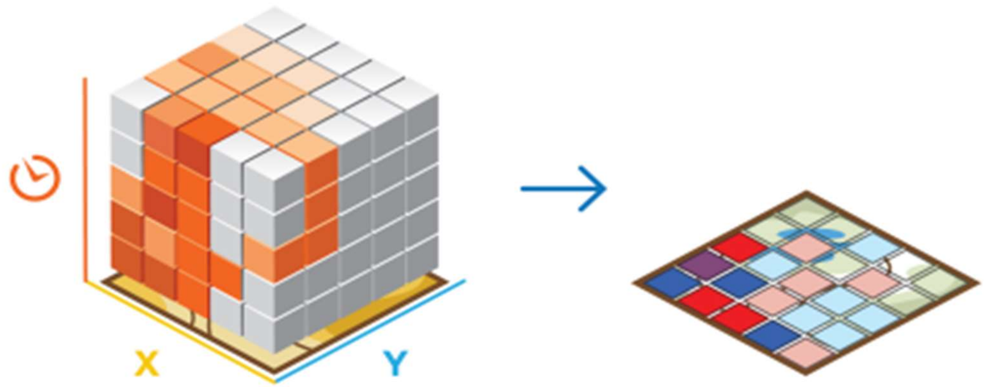


Figure 4.9. Symbolic Representation of the Local-Outlier Analysis Tool Cube in ArcGIS. (ArcGIS, 2017)

4.5.2. Procedure

The first step towards space-time pattern mining is to create a space-time cube for the dataset. The following information is required to utilize this tool:

i. Input the feature point data.
ii. Specify the output location for the space-time cube. The resulting file will be a NetCDF data cube.
iii. Select the time field. The dataset must have the time field in the properties. Otherwise, this tool will not perform.
iv. Identify the time-step interval. The minimum time interval required by the tool is 10. The time interval can be in weeks, days, months, or years, depending on the dataset's length.
v. Select the time-step alignment. This function helps the dataset to divide the time intervals equally. The best practice should be to provide a time alignment that covers the dataset equally for each time interval. Hence, each time interval has the same amount of days and has data points for the start and end dates.

vi. Provide the distance interval. It is the spatial extent of the bins to aggregate the input's feature points. The value is something the user must decide based on the dataset's area and denseness. This function is optional, and if it is not provided, the software will calculate the distance depending on the dataset's denseness. If the dataset is spread out, the distance interval will be larger and vice versa.

vii. Provide a summary field, if needed. If it is not provided, the tool will automatically count the points in each bin. This option depends on the type of dataset. For example, the road-accident dataset only needs to be counted, whereas the water-usage dataset point requires addition. If the bins are empty, there are fill options available to calculate the empty-bin dataset.

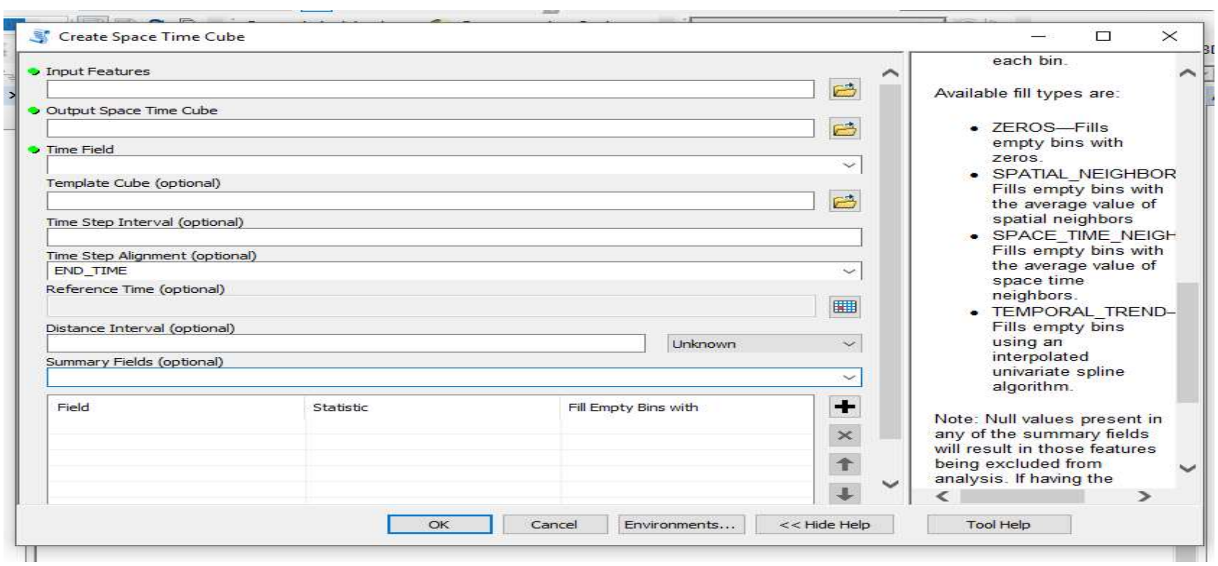


Figure 4.10. Framework of the Create Space-Time Cube Tool in ArcGIS.

Figure 4.10 provides the framework for the Create Space-Time Cube tool in ArcGIS. Once the space-time cube is built, the rest of the tools in the space-time mining pattern can be used. The Optimized hotspot analysis needs the following details to accomplish the task:

i. Locate the space-time cube created with the create space-time cube tool.

ii. Select the analysis variable. It is the numeric value in the NetCDF file that you want to analyze.

iii. Locate the directory for the output feature.

iv. Provide a value for the neighborhood distance. This function value will perform the space-time clustering to assess which feature to analyze for the spatial extent of the neighborhood analysis. This feature is optional.

v. Enter the value for the time-step interval to include in the neighborhood analysis. This function is optional.

Figure 4.11 displays the framework for the Optimized hotspot analysis in ArcGIS. This tool is placed in the Arc Toolbox under the category of space-time pattern mining.

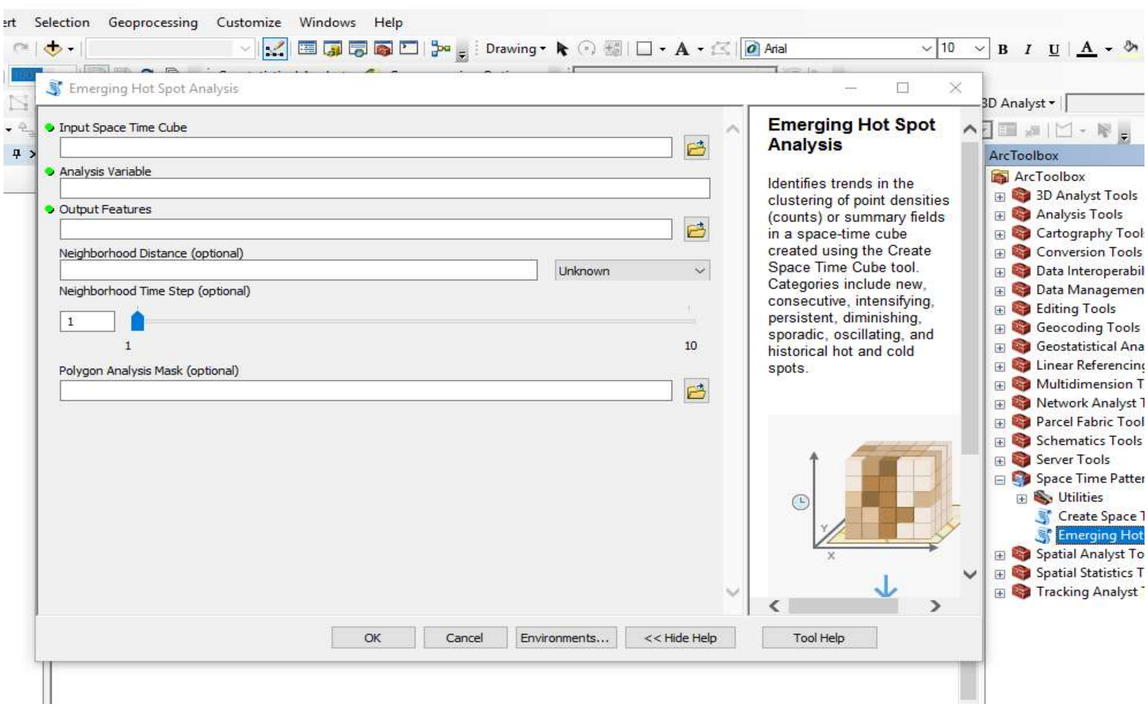


Figure 4.11. Framework for the Emerging Hotspot-Analysis Tool in ArcGIS.

Another tool has been used to analyze the data outliers in Space-Time Pattern toolbox is local-outlier analysis. The required inputs for this tool are similar to the emerging hotspot-analysis tool except that this tool requires the number of permutations. Although this input feature is optional, it may have a significant effect on the output result. Permutation is a process of analyzing the dataset values to find the actual spatial distribution. Each permutation process reshuffles the values for the neighborhood around the bins and calculates the local Morgan I's. The values are then compared with the original Morgan I values, and probability is determined based on an observed value which corresponds to a random distribution. The tool has a default permutation value of 499. However, the permutation value can be increased to improve the random-sample distribution.

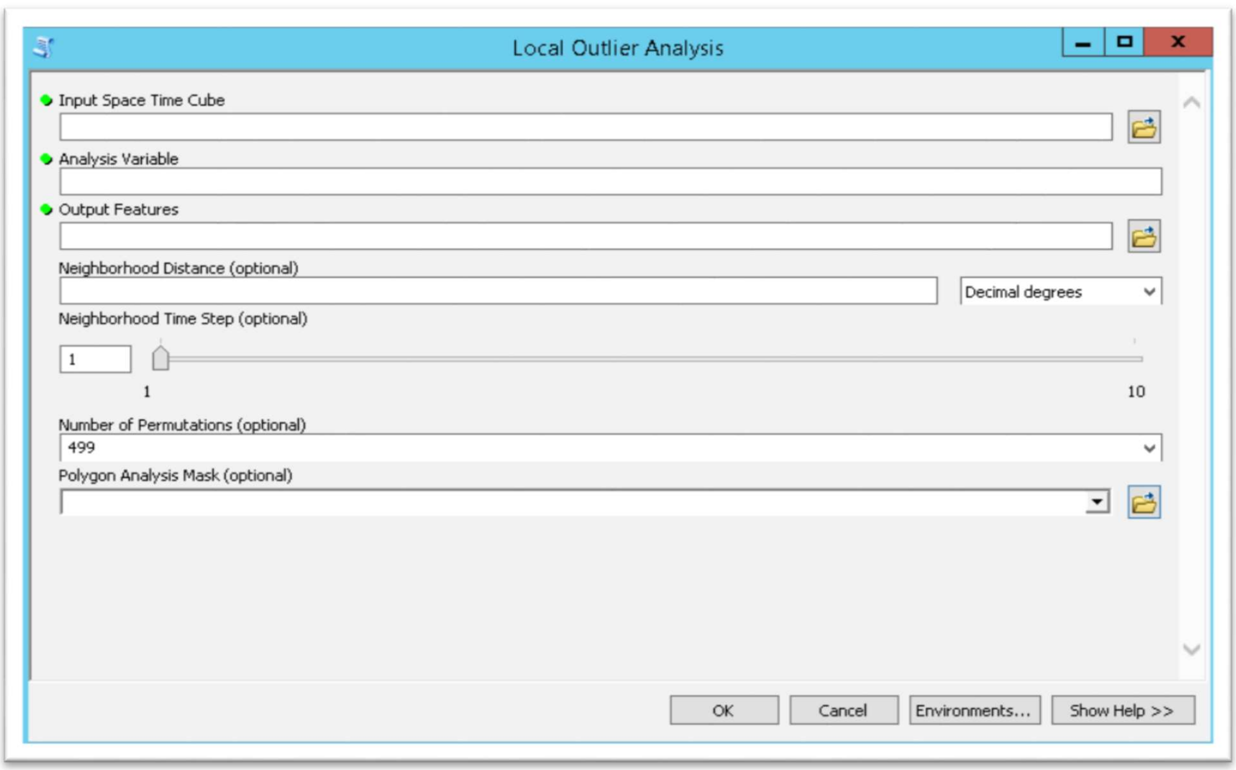


Figure 4.12. Framework for the Local-Outlier Analysis Tool in ArcGIS.

4.5.3. Results

4.5.3.1. Create Space-Time Cube

The space-time cube has aggregated 384,597 points into 7,100 fishnet grid locations over 14-time step intervals. Each location is 2,895 feet by 2,895 feet square. The entire space-time cube spans an area that is 289,500 feet from west to east and 205,545 feet from north to south. Each time-step interval is 6 months in duration, so the entire period covered by the space-time cube is 84 months. Of the 7,100 total locations, 2,371 (33.39%) contain at least one point for at least one-time step interval. These 2,371 locations comprise 33,194 space-time bins, out of which 27,444 (82.68%) have a point count greater than zero. There is a statistically significant increase for point counts over time. The trend direction is growing with a statistical value of 3.7227 and a trend p-value of 0.0002.

4.5.3.2. Emerging Hotspot Analysis

Table 4.4 and Figure 4.13 offer a comparison and evaluation between the hot and cold spots for Houston's accident dataset during the study period, 2010-2016. The graphical representation in Figure 4.13 is a more elaborative picture to observe the comparison for the hot and cold spots. Table 4.4 suggests that 17 new hotspots developed from 2010-2016. Moreover, there are 424 locations where the hotspots are consecutive, and 246 locational polygons are intensifying. However, there are no diminishing and sporadic hotspots identified in the accident database's 7-year period. The 90 hotspots locations spotted as the sporadic and 280 values for the oscillating hotspots. In historical, no hotspots are identified.

The cold-spot values that emerged in the 7-year study period are 0 for the new, consecutive, intensifying, and oscillating locations. However, there are persistent cold spots with a value of 86, and there is a value of 372 for diminishing cold-spot locations. There are 13 historical cold-spot places, and the sporadic cold spot has a value of 18.

From the total spots in the accident dataset, approximately 68% are hotspots, and 32% are cold spots. The values are intriguing and support the emerging trend of more traffic accidents in Houston as discussed and studied previously with the different statistical analyses. The results of the emerging hotspot analysis endorse the statistical-analysis results. The value for diminishing cold spot is 372 which poses a threat to concern authorities and these cold spots can turn into hot spots locations within a few years. However, there are 0 diminishing hotspots. All the hotspot categories have significantly higher values than the cold-spot categories, significantly showing that the trend for traffic accidents in Houston has risen and is expected to increase unless thoughtful corrective actions are implemented.

Table 4.4. Results for Hot and Cold Spots by Accident-Data Category, 2010-2016.

Category	Hot	Cold
New	17	0
Consecutive	424	0
Intensifying	246	0
Persistent	0	86
Diminishing	0	372
Sporadic	90	18
Oscillating	280	0
Historical	0	13

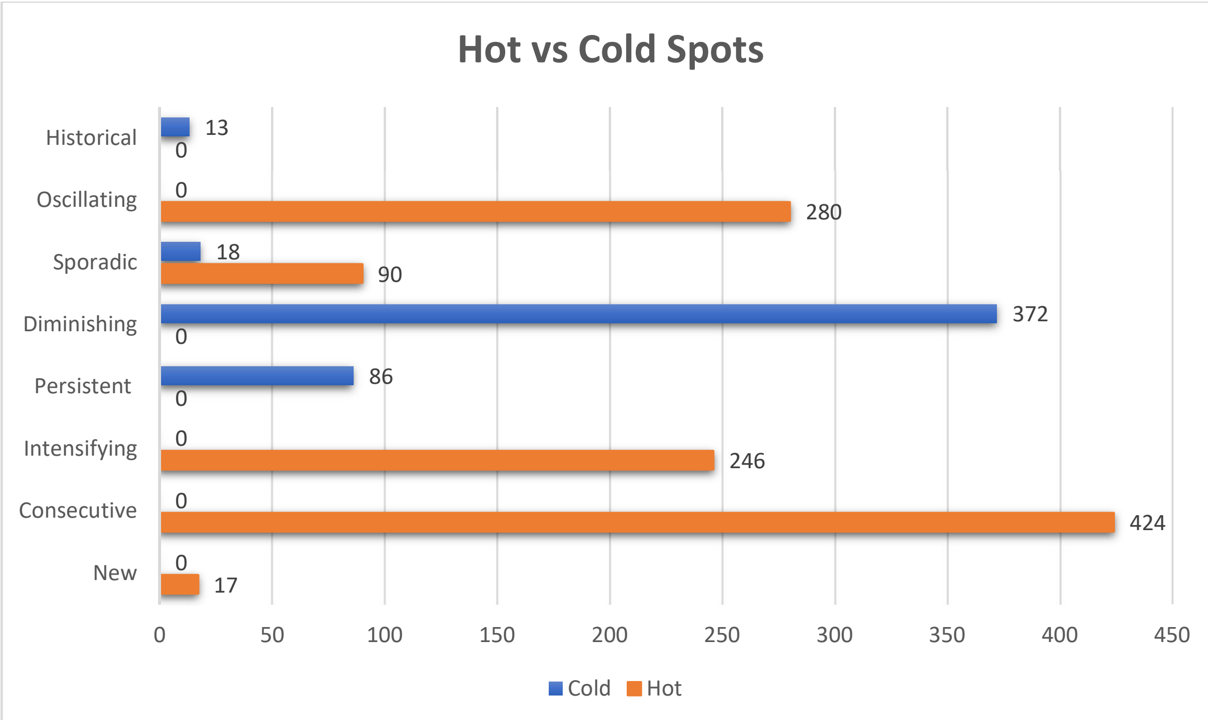


Figure 4.13. Graphical Representation for the Hot and Cold Spots by Accident-Data Category.

The outcome for the emerging hotspot analysis is presented in Figure 4.14. From the hotspot map, it is evident that new hotspot locations have emerged over time, from Houston central position towards the west and east directions. The new hotspots growth in the 7-year period are few outer to the oscillating hotspots and the Houston center are identified as intensifying and persistent hotspot locations. All coldspots are in the city's outskirts with cold-spot patterns. However, some successive hotspot places have been detected with the analysis of Houston's north side.

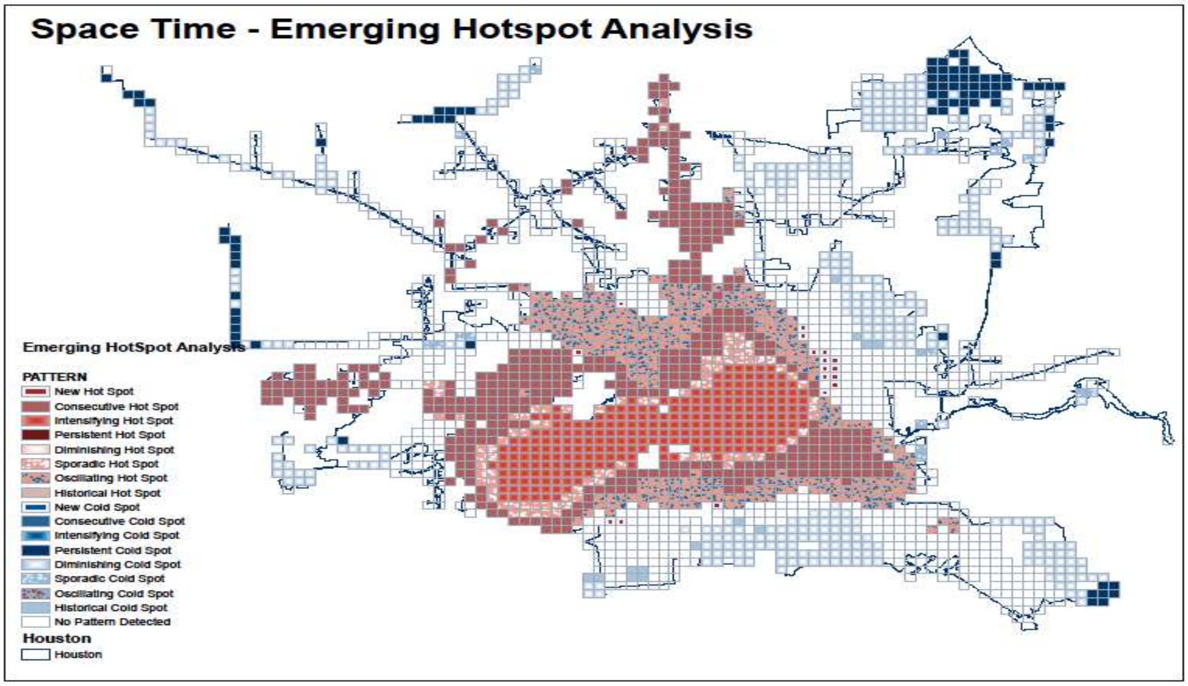


Figure 4.14. Output Map of the Emerging Hotspot-Analysis Tool.

4.5.3.3. Local-Outlier Analysis

Table 4.5 displays the key time steps that were attained after performing the analysis using the local outlier. The key analysis steps offer valuable information about the key findings regarding outliers and clusters in the various time intervals.

Table 4.5. Highest Values for the Local-Outlier Analysis Parameters at Various Time Intervals.

Parameters	Time Interval	Values
Highest number of outliers	2016-06-30 00:00:01 to 2016-12-31 00:00:00	517
Lowest number of outliers	2010-06-30 00:00:01 to 2010-12-31 00:00:00	247
Highest number of high-low outliers	2011-12-31 00:00:01 to 2012-06-30 00:00:00	191
Lowest number of high-low outliers	2016-06-30 00:00:01 to 2016-12-31 00:00:00	58
Highest number of low-high outliers	2016-06-30 00:00:01 to 2016-12-31 00:00:00	459
Lowest number of low-high outliers	2011-06-30 00:00:01 to 2011-12-31 00:00:00	86

Table 4.5 suggests that the highest number of outliers (517) was discovered in the most-recent time interval, and the lowest number of outliers (247) was in the second-time interval for 2010. Remember that each year is divided into 2 time intervals, which is of 6 months. The highest trend of decreasing for an outlier from high to low is noticed in the first time-interval of 2102 with a value of 191, and the lowest number is in the most recent time interval with a minimum value of 58. Interestingly, the highest number of low-to-high outliers (459) is identified in the most-current time interval, a finding which is verified by the previous statistical analysis. Moreover, the minimum trend of high-to-low outliers has a value of 86 which occurred in the second-time interval of 2011.

Table 4.6. Number and Percentage of Locations by Local-Outlier Analysis Parameters.

Category	# of locations	% of locations
Never Significant	89	3.75
Only high-high clusters	292	12.32
Only low-high outlier	241	10.16
Only Low-low cluster	907	38.25
Only high-low outlier	61	2.57
Multiple types	781	32.94

Table 4.7. Number and Percentage of Bins by Local-Outlier Analysis Parameters.

Category	# of Bins	% of Bins
high-low outlier	1501	4.87
low-high outlier	2929	9.50
High – High cluster	4310	13.98
Low - Low Cluster	11252	36.51
Multiple types	10831	35.14

Tables 4.6 and 4.7 detail the numeric information about the locations and bins for each local-outlier analysis category. The follow-up percentages in Tables 4.6 and 4.7 are obtained from each category’s numeric value divided by the total number of locations and bins, respectively. The relevant information that Tables 4.6 and 4.7 provide is the number of high and

low outliers and the number of cluster points. The high-low outlier's value is only 61, which is 10.16%. However, the high-high clusters and the low-high outliers have a total percentage value around 15 which is significantly high for any city. Thus, 15% of Houston is identified as critical and dangerous for drivers to operate vehicles. These points require further investigation. These locational-interval polygons could help study the factors about how the values have become low-high or high-low in the time intervals. Further, an investigative analysis on the high-high and low-low clusters has remained constant over time. These studies can help to mitigate the accident issues in Houston. The percentage for non-significant location is only 3.75%. However, this information does not imply that the other categories have been non-significant in that particular time interval.

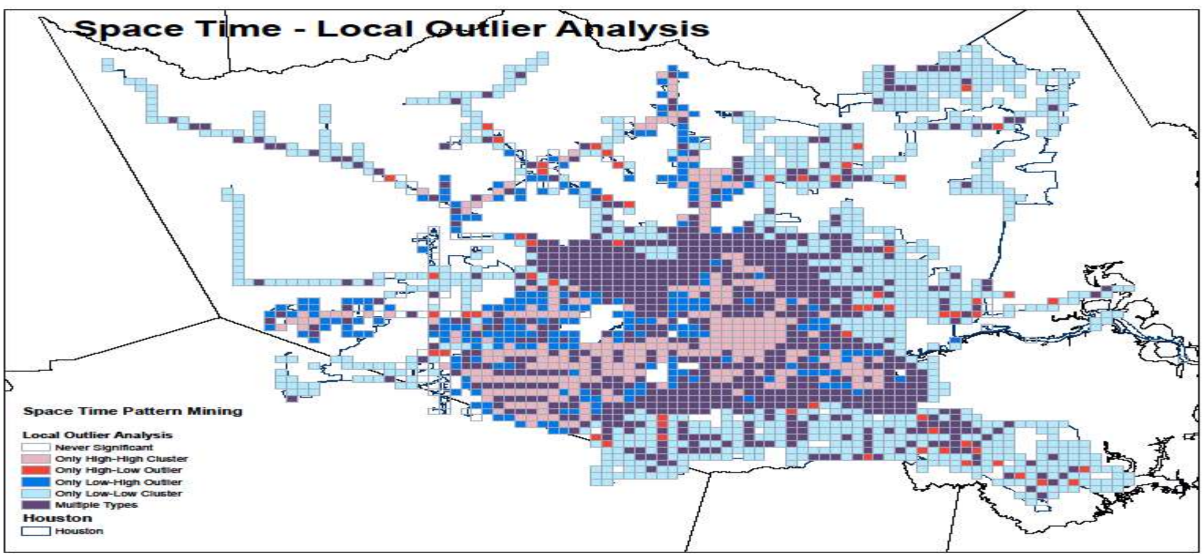


Figure 4.15. Output Map for the Local-Outlier Analysis Tool.

The outcome for the local-outlier analysis is presented in Figure 4.15 Again, the central part of Houston is identified as being a high-high outlier mostly, and this trend spread towards the southwest section of the city. The locations surrounding the city's center have multiple outlier and cluster patterns. Houston's outskirt areas mostly have a pattern with low-low clusters. However, some locational polygons only have low-high outliers.

4.6. Kriging Analysis

Kriging is an interpolation method or process to estimate the values from the surrounding data points' Z values based on a regression that is spatially correlated. Kriging is a geostatistical method that uses statistical methods to find relationships among the measured points. ArcGIS has various kriging tools that can be implemented with accident data to obtain the prediction and probability maps which are based on spatial correlation. The researcher varies the kriging tools in ArcGIS to produce a realistic model of the crash-point data. The following kriging tools provide the best and most acceptable outcome for Houston's accident database.

4.6.1. Indicator Kriging

4.6.1.1. Introduction

Indicator kriging uses the probability function to calculate the predicted values for the unknown points. The Indicator kriging redefines the categories to indicator function by computing semi-variogram, fits the models, and interprets the results as probabilities. The probability function predicts the unknown points by transforming values into a binary number of either 0 or 1, depending upon the spatial relationship. The binary coding of data relies upon their relationship with the designated-limit z values. Transforming values to the binary code is non-linear. The information states that the values greater than designed limit will receive the same indicator values as those with slightly higher values from designated limits. The result provides the prediction values in range of 0 and 1. Indicator kriging can be designed for more than one threshold value; therefore, the cumulative distribution function (CDF) technique is used to approximate each point on the grid. The drawbacks for indicator kriging are that the distribution may result in unexpected values, such as negative values, a value greater than one, etc. Also, it is hard to model a variogram or covariance system with indicator kriging. However, the indicator kriging in ArcGIS incorporates all these factors to provide smooth and realistic results for the predicted values. Indicator kriging assumes the following model:

$$I(s) = \mu + \varepsilon(s) \tag{Eq. 4.2}$$

Where,

μ = unknown constant

I(s) = binary variable

The indicator kriging tool in ArcGIS requires modeling a semi-variogram to examine the autocorrelation and to quantify the predicted values. The semi-variogram reflects the spatial autocorrelation among the measured values. A model of the average values for each location point is plotted and fit using various models. The model has certain characteristics that are commonly used to describe sill, nugget, partial sill, and range. Figure 4.16 illustrates the model and aids the reader's understanding of its features.

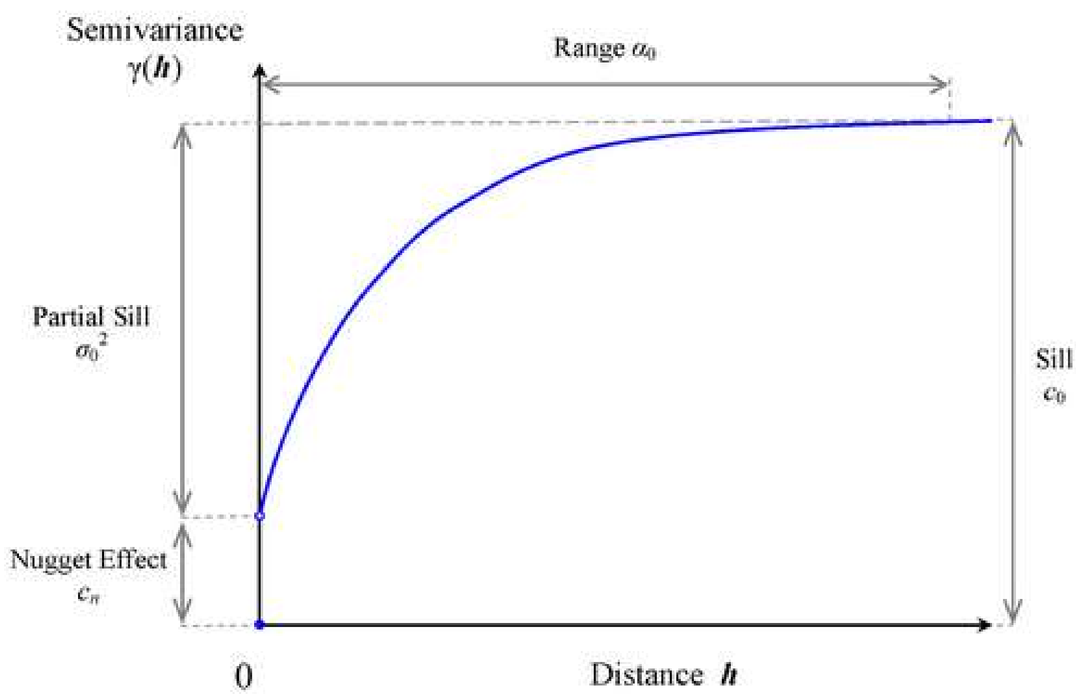


Figure 4.16. Graphical Explanation of the Semi-Variogram Characteristics. (SAS, 2013)

To gauge the best-fit model and spatial dependency, the nugget-to-sill ratio is commonly calculated and expressed as a percentage.

$$nugget\!:\!sill\ ratio = \left(\frac{nugget}{sill}\right) * 100 \qquad \text{(Eq. 4.3)}$$

Where,

Sill = nugget effect + partial sill

The sill:nugget ratio accounts for the total variance that is successfully incorporated into the model and indicates what percentage of the overall variance is found at the nearest lag interval. When the data are normally distributed, there is little difference in the variance for any distance comparison. However, the variance tends to increase when there is a pattern present in the distribution that is spatially autocorrelated. When the design is straight and level, a sill is formed, telling you that the sample points become independent at this point and beyond. The low percentage (< 25%) indicates that a large part of the variance has been introduced to the model, implying a strong spatial correlation among the variables. A higher percentage (> 75%) suggests a weak spatial independence among the variables.

4.6.1.2. Procedure

These steps are used to build the semi-variogram and to perform the kriging:

i. Select Kriging/Co-Kriging in the Geostatistical Methods.
ii. Provide the data input by selecting the dataset source and the data field that will be used for modeling and interpolation.
iii. After the data input, click Next to go to the next step.
iv. Select Indicator Kriging in the Kriging Type, and click Next. Before going to the next step, choose the output surface type that you want after performing the geostatistical wizard.

v. This step lets you perform the semi-variogram/covariance modeling. This frame provides the nugget and partial-sill values. There is an optimize model option to improve the model parameters using cross-validation. Also, there are multiple model types available that can be selected to improve the model and to observe which one provides the best fit. Modeling the semi-variogram is an experimental process and involves various parameters to produce the best result. Figure 4.17 gives an overview of the settings which are used to develop these models. The optimize model function is optional and depends on the semi-variogram type. The semi-variogram is modeled using the exponential model type for the results because this sort ensued the best fit semi-variogram for given accident dataset.

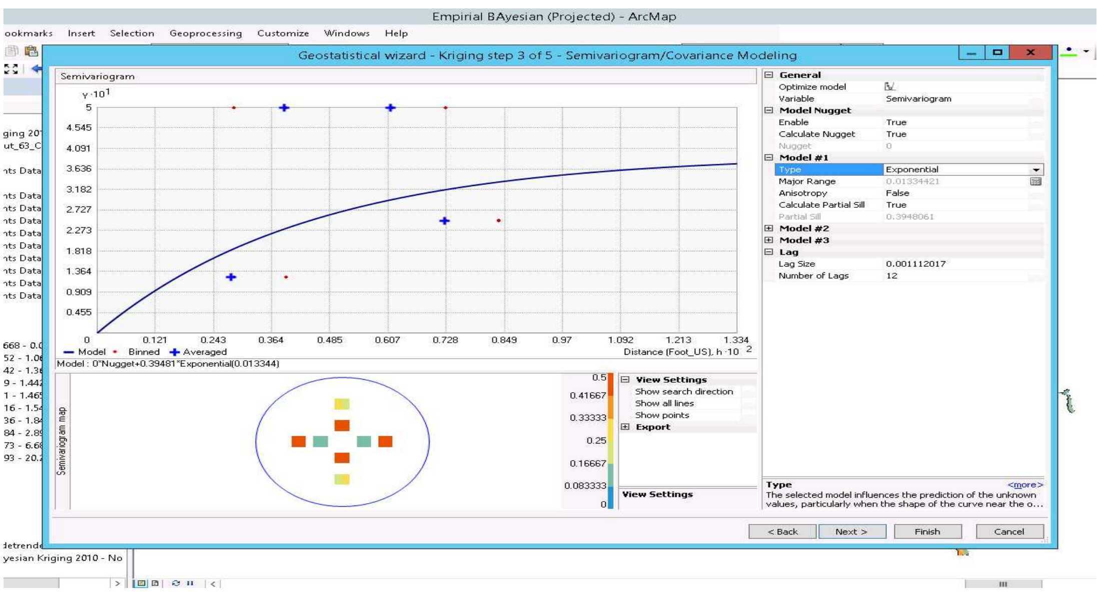


Figure 4.17. Framework for the Semi-Variogram/Covariance Modeling in the Geostatistical Analysis.

vi. In the next step, there are options about the search neighborhood. For this study's results, these options have used the default setting.

vii. The last step of the geostatistical wizard is cross-validation. Here, the tool details the prediction error's statistical values and the graph that could be helpful to display the model's validation results and to compare the best interpolation method in ArcGIS.

4.6.1.3. Results

By using Indicator kriging in the geostatistical tool, semi-variogram models for each year's accident data are obtained. The models are shown in Figures 4.18-4.24. The resulting semivariogram models are the finest that have been created after trying the model's various characteristics.

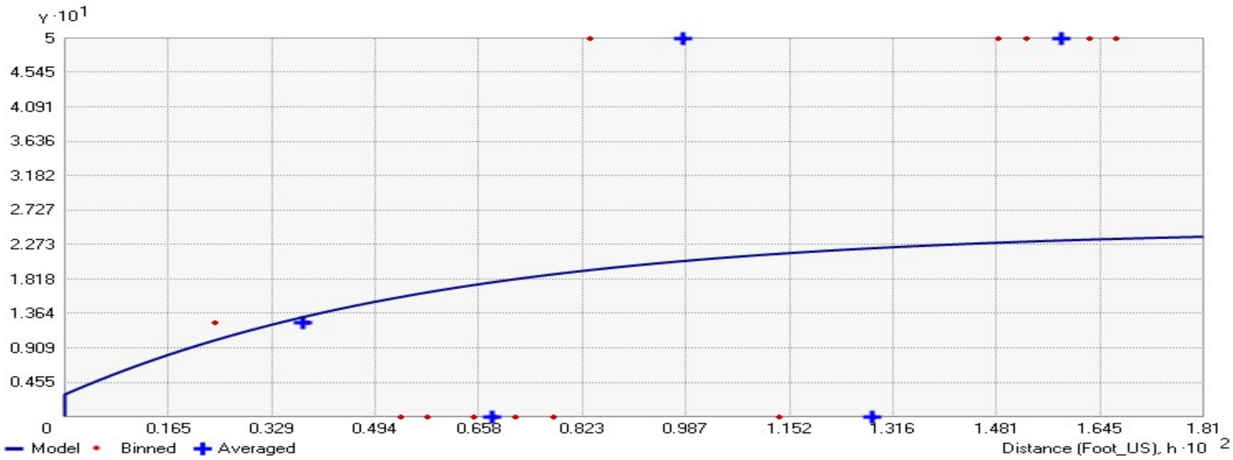


Figure 4.18. Semi-Variogram Model for Houston's 2010 Accident Dataset.

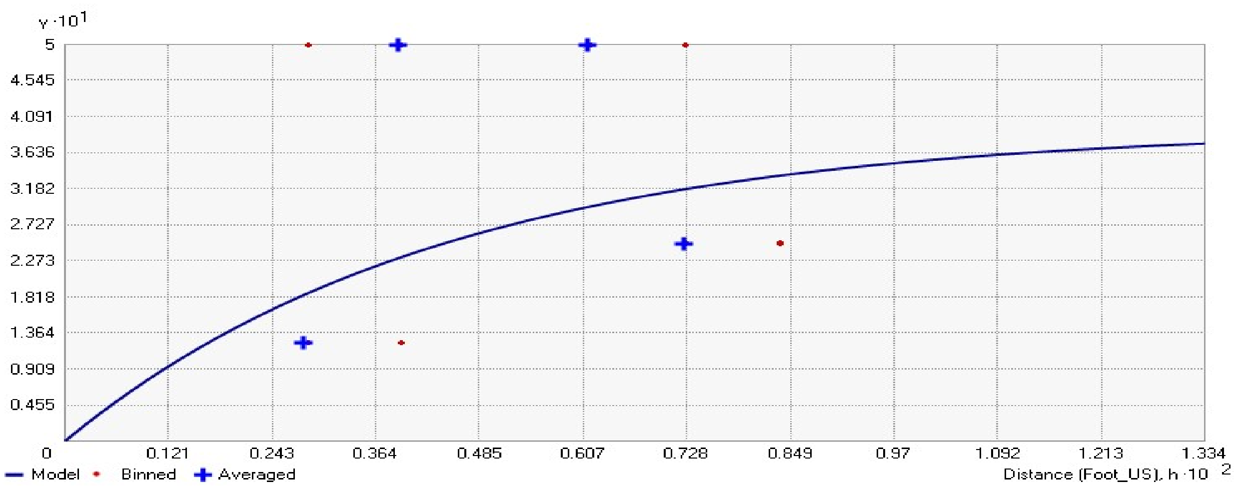


Figure 4.19. Semi-Variogram Model for Houston's 2011 Accident Dataset.

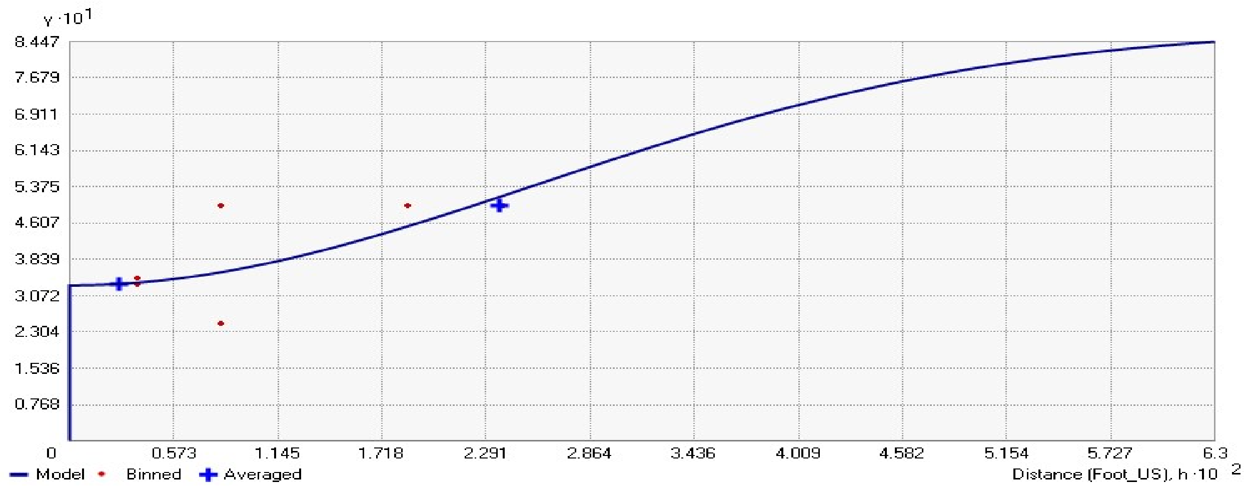


Figure 4.20. Semi-Variogram Model for Houston's 2012 Accident Dataset.

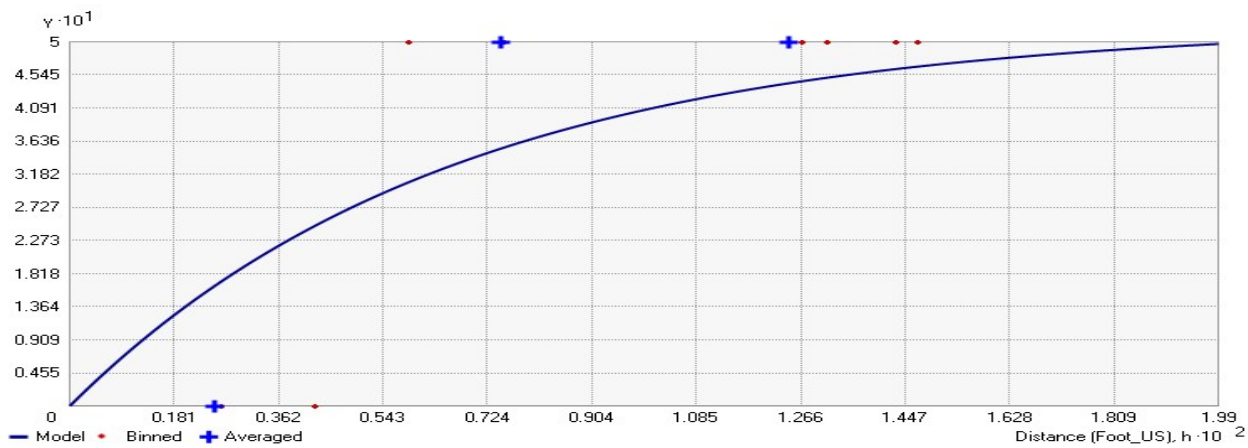


Figure 4.21. Semi-Variogram Model for Houston's 2013 Accident Dataset.

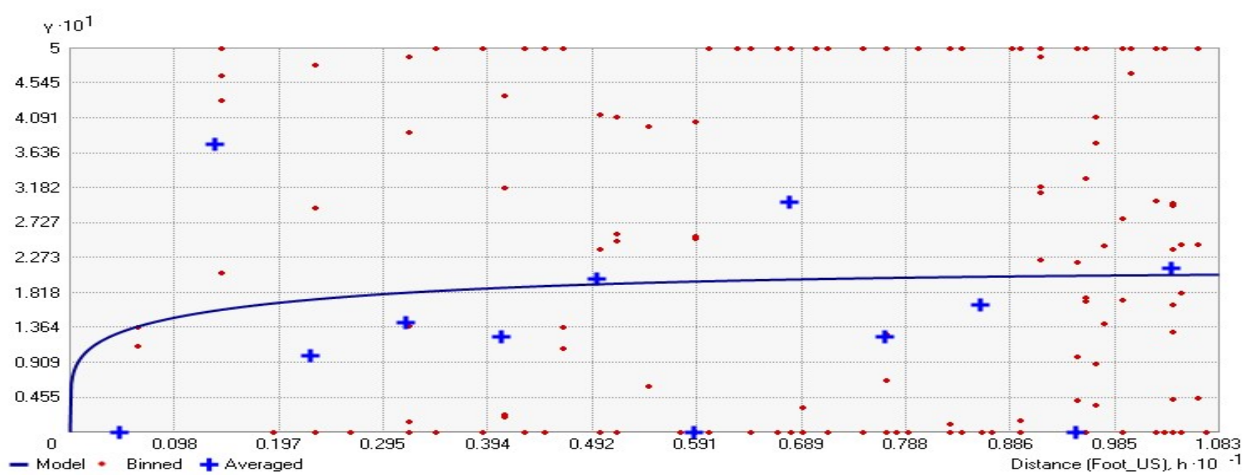


Figure 4.22. Semi-Variogram Model for Houston's 2014 Accident Dataset.

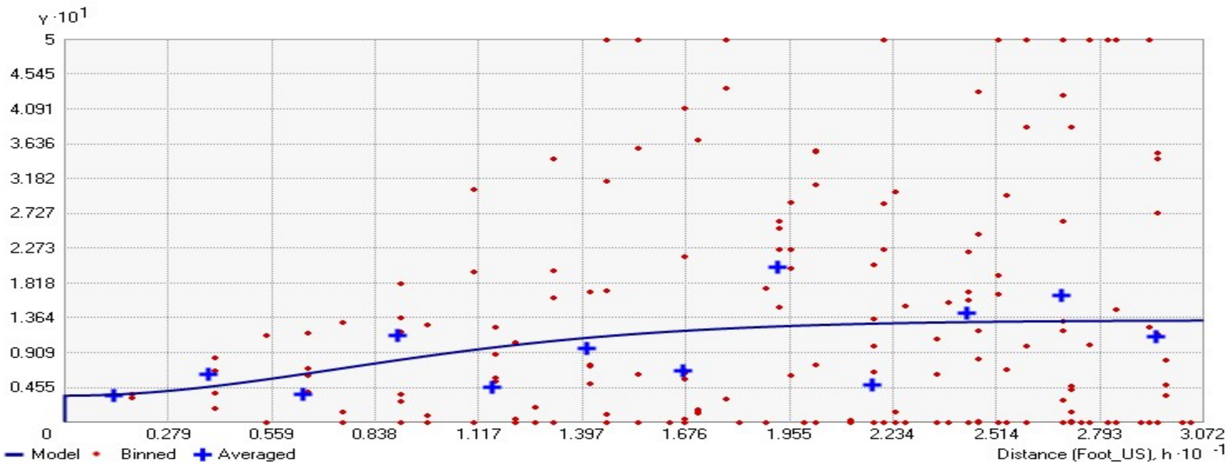


Figure 4.23. Semi-Variogram Model for Houston's 2015 Accident Dataset.

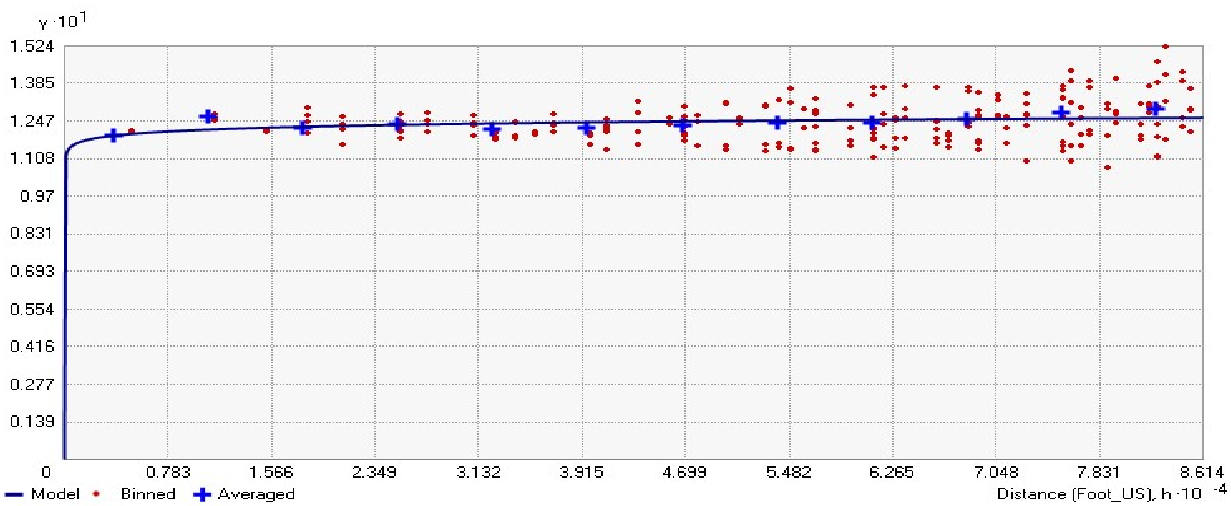


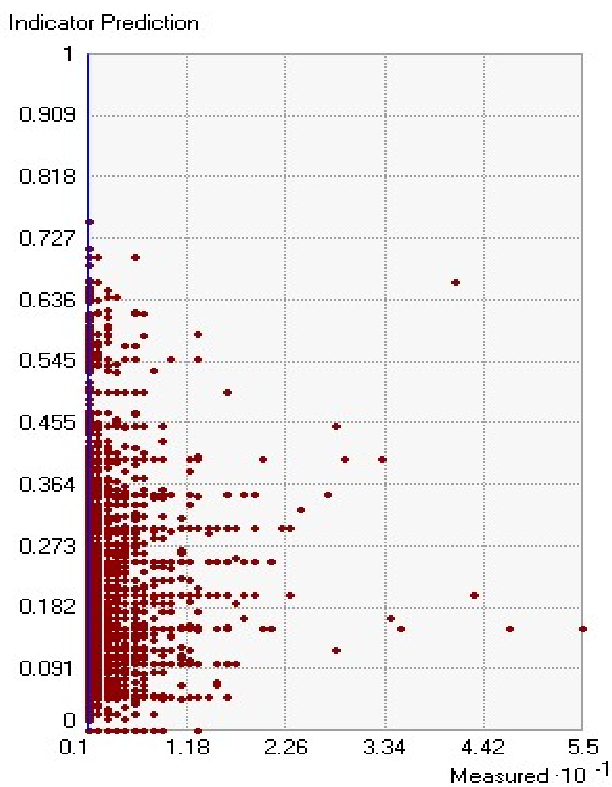
Figure 4.24. Semi-Variogram Model for Houston's 2016 Accident Dataset.

Table 4.8 shows information about the model type used to develop the semi-variograms as well as their nugget and partial-sill values. By using the numerical values for sills and nuggets, the nugget-to-sill ratio is calculated. The percentages for each year's accident database suggest that the model has incorporated most of the variances and reflect a model with a high spatial relationship among the variables. Moreover, model types such as exponential, Gaussian, and K-Bessel have shown the best fit with the lowest percentages for the sill ratio.

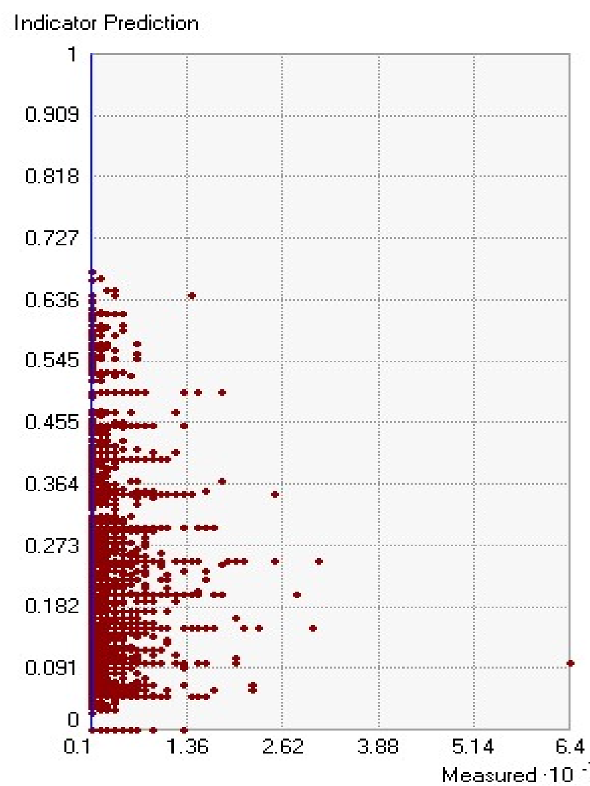
Table 4.8. Indicator Kriging Characteristic Values of the Semi-variogram Models for Each Year's Accident Dataset.

Year	Model Type	Model Optimized	Nugget	Partial Sill	Sill	Percentage
2010	Exponential	YES	0.029271135	0.219262932	0.248534066	11.778%
2011	Exponential	NO	0	0.394806144	0.394806144	0.000%
2012	Gaussian	YES	0.329318566	0.542365543	0.871684109	37.780%
2013	Exponential	NO	0	0.52321509	0.52321509	0.000%
2014	K-Bessel	NO	0	0.208472166	0.208472166	0.000%
2015	K-Bessel	NO	0.034603586	0.09858832	0.133191906	25.980%
2016	K-Bessel	YES	0.017713312	0.112959199	0.130672511	13.555%

After executing the semivariogram models, cross-validation allows checking the indicator probability for each feature point used for the kriging models. The cross-validation step provides a graph for the probability-indicator prediction which is displayed in Figure 4.25.

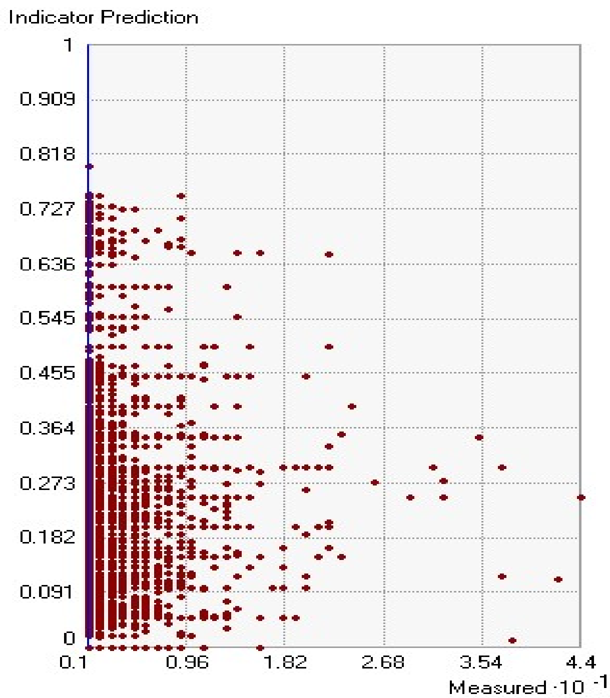


(2010)

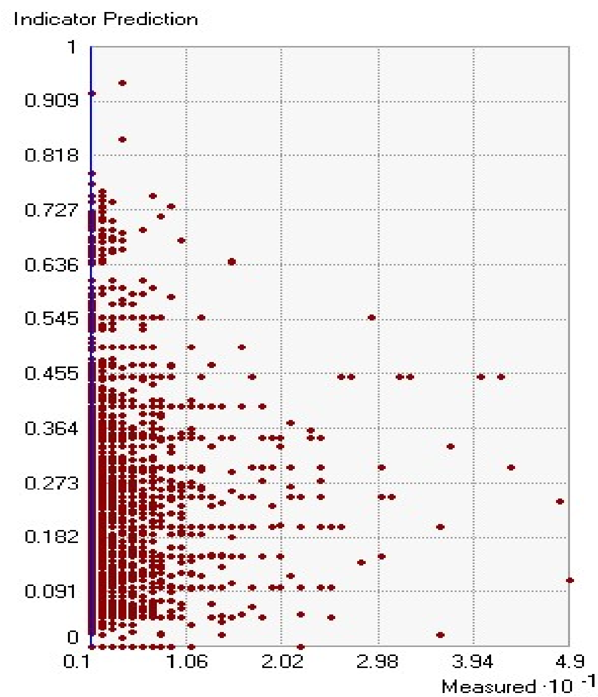


(2011)

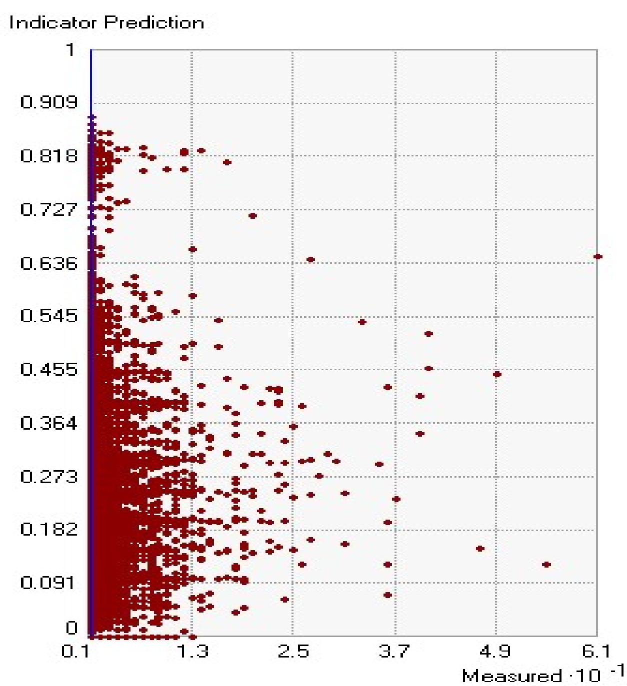
Figure 4.25. Indicator Prediction Graphs for Accident-Point Data (2010-2016).



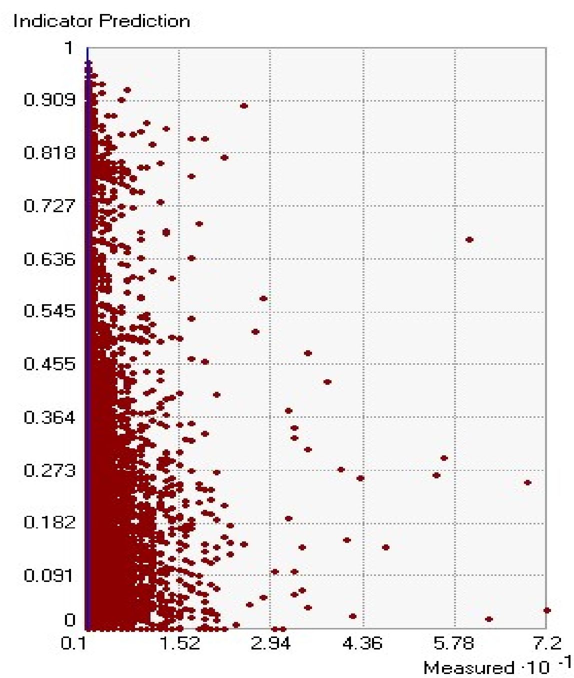
(2012)



(2013)

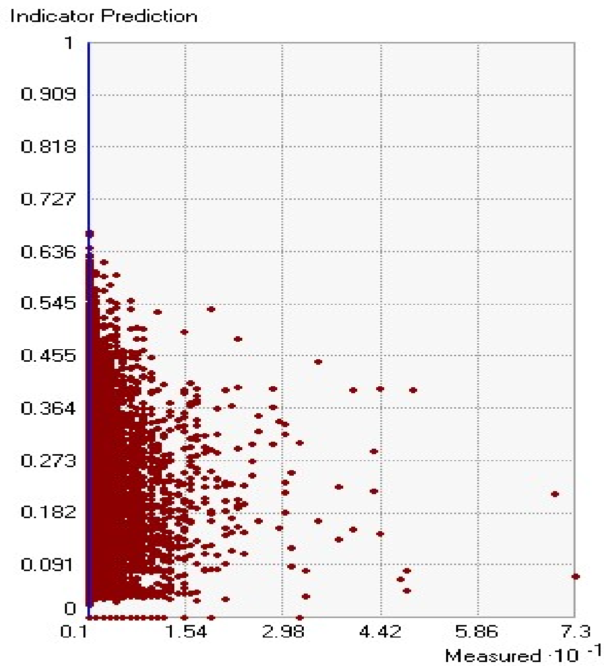


(2014)



(2015)

Figure 4.25. Indicator Prediction Graphs for Accident-Point Data (2010-2016). (Continued)



(2016)

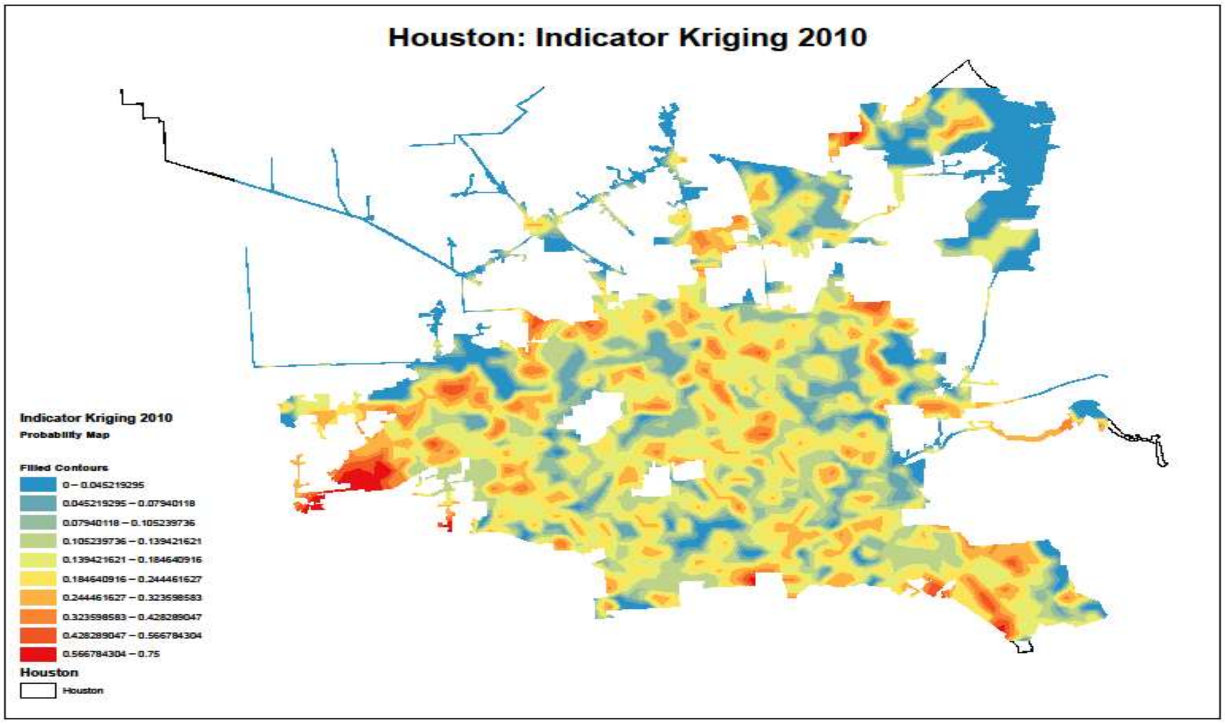
Figure 4.25. Indicator Prediction Graphs for Accident-Point Data (2010-2016). (Continued)

Table 4.9 details the statistical information about the model's prediction errors. These values provide the grounds for a comparison between the parameters and among the different models for each year and to understand how these models can be measured based on performance. The targeted values for each comparison parameter are also given in Table 4.9. If the parameter values are closer to the target values, the model is said to be a good fit and acceptable.

Table 4.9. Comparison Parameters for Each Year's Semi-Variogram Models.

Comparison Parameters	Prediction Errors	RMSS	Mean Standardized	RMS	Average Standard Error
	Target Values	1	0	as low as possible	as close to RMS error as possible
Years	2010	0.779555104	-0.009871036	0.397328235	0.511553085
	2011	0.624945744	-0.009706496	0.402697974	0.645849301
	2012	0.42238587	-0.0057711	0.401402569	0.957085543
	2013	0.557333131	-0.008676688	0.411333474	0.741607537
	2014	0.914041033	-0.014897626	0.417484269	0.466258464
	2015	1.143042969	0.015346926	0.378606561	0.345846568
	2016	1.03612686	0.03333929	0.358394241	0.34546983

The probability maps which are produced after performing the indicator kriging are shown in Figure 4.26. The maps have values in the range of 0 to 1. Figure 4.26 shows the accidents probability maps in Houston interpolated from the accidents database. The probability limits for each kriging map differ, but the appearance is typically similar for all the indicator-kriging maps. The outskirt areas have some places with a lower probability of accidents, whereas the city's central part and its surrounding area have the possibility of multiple accidents.



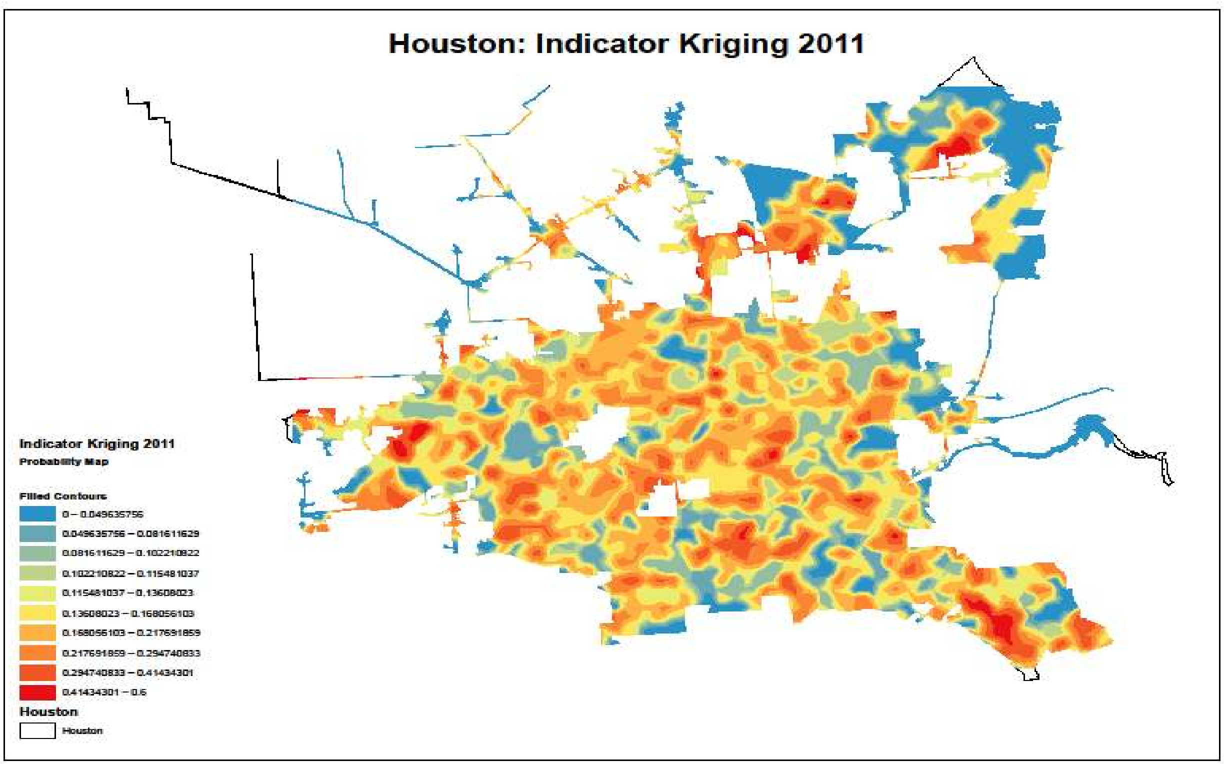
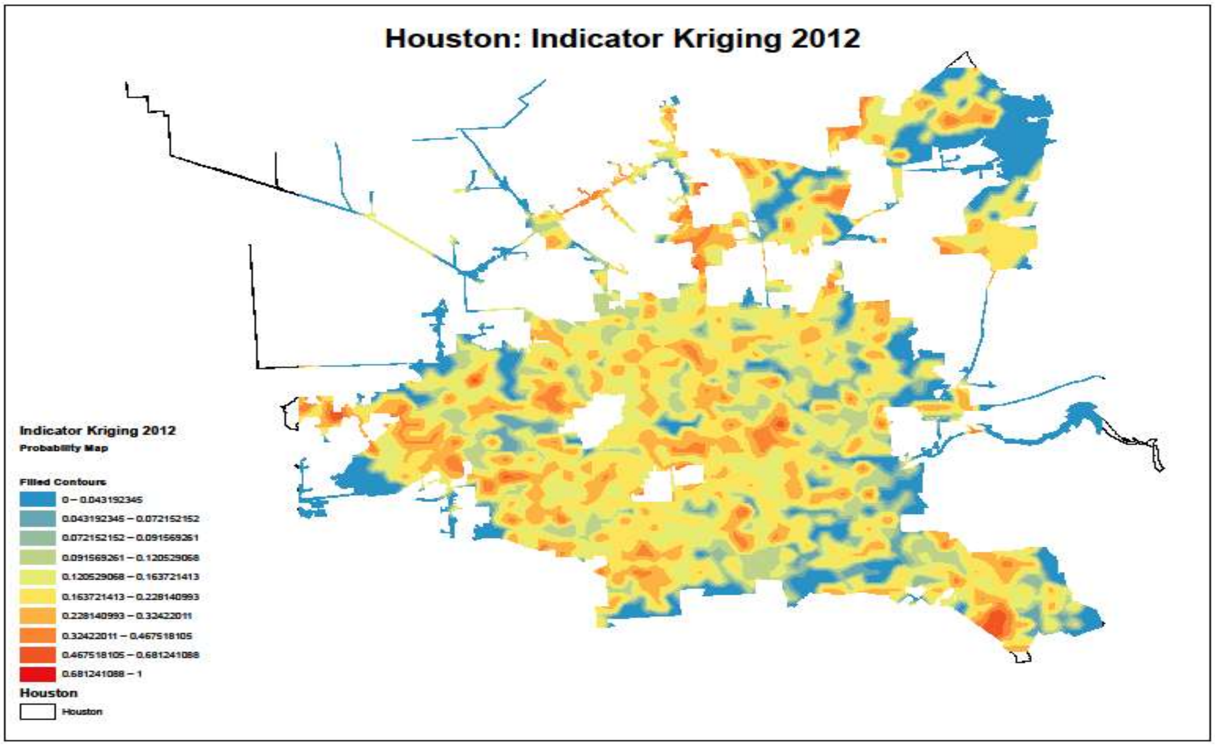


Figure 4.26. Result Maps from Indicator Kriging with Houston's Traffic-Accident Data, 2010-2016.



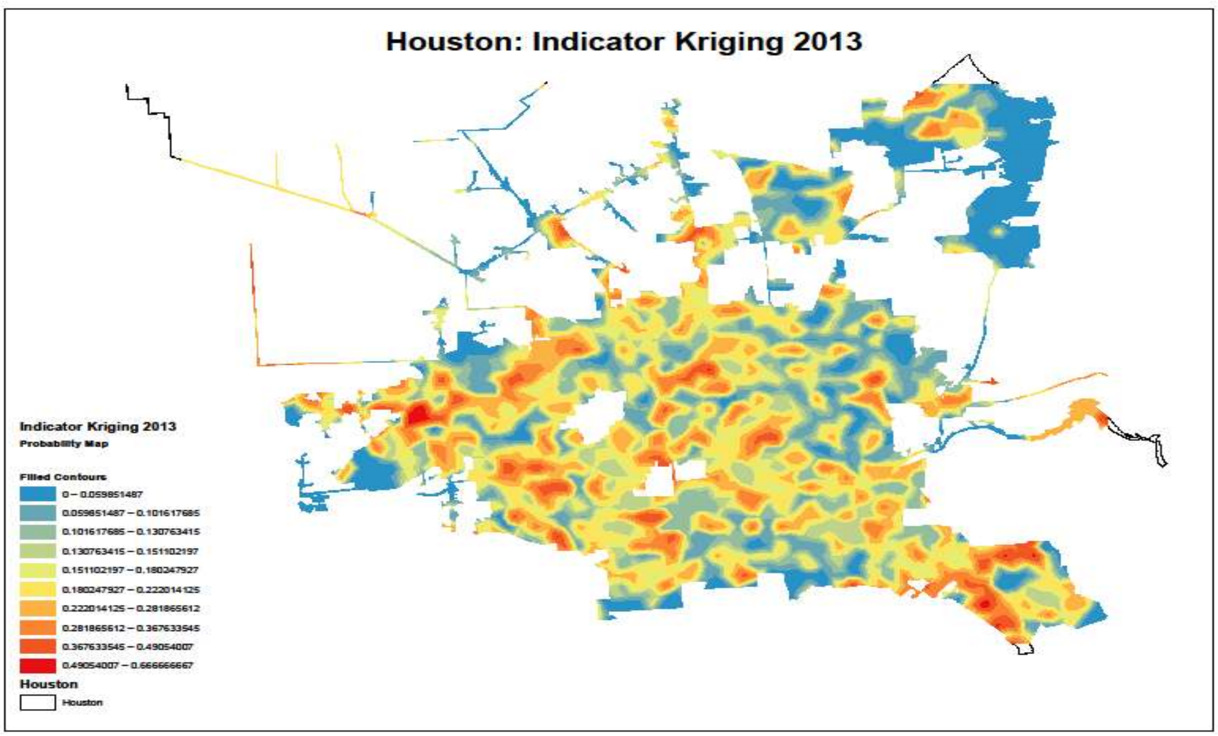
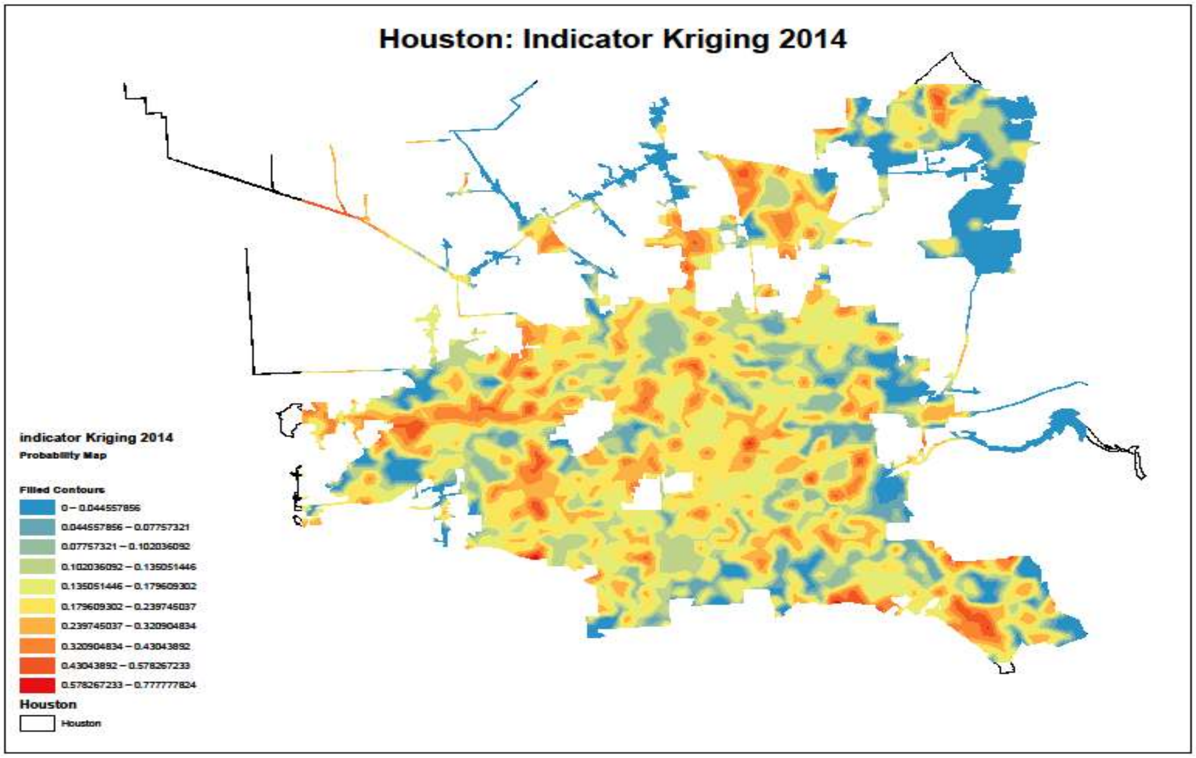


Figure 4.26. Result Maps from Indicator Kriging with Houston's Traffic-Accident Data, 2010-2016. (Continued)



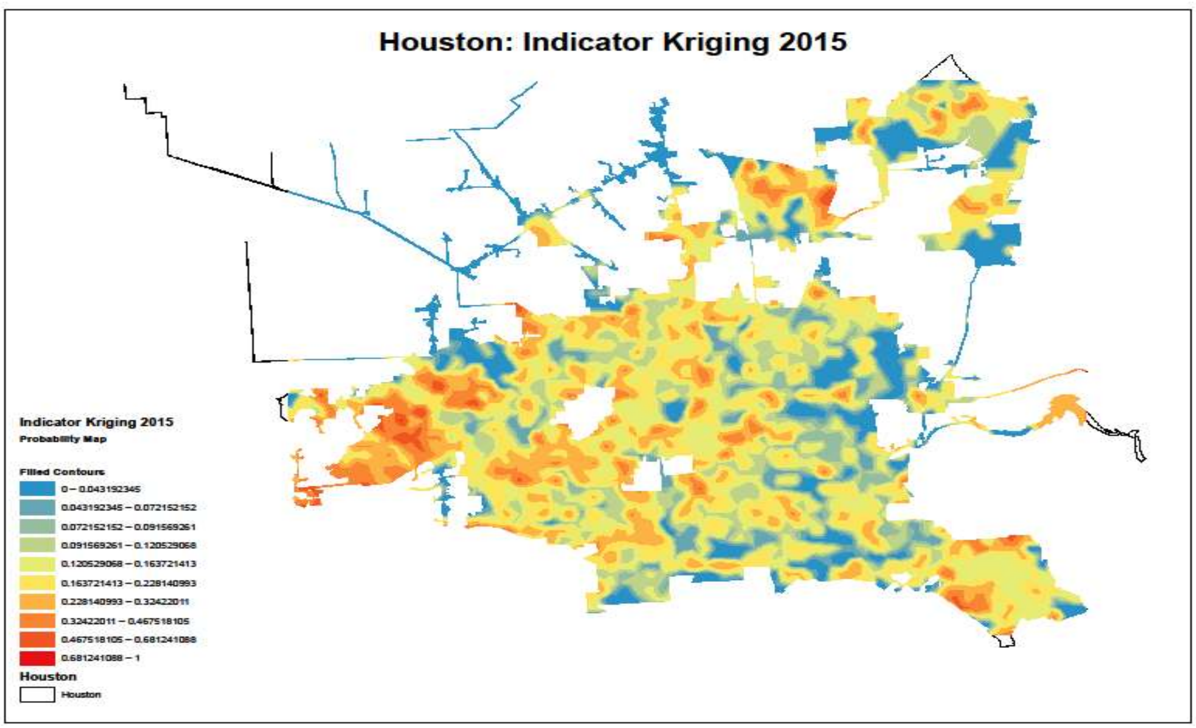


Figure 4.26. Result Maps from Indicator Kriging with Houston's Traffic-Accident Data, 2010-2016. (Continued)

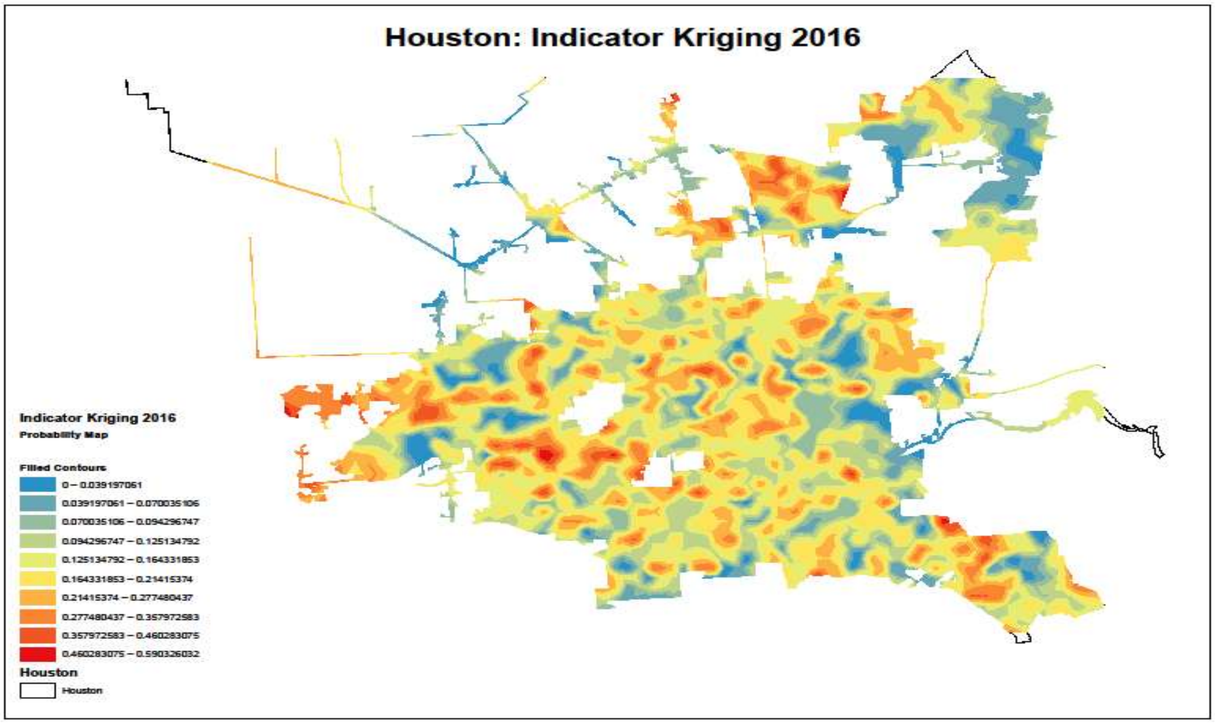


Figure 4.26. Result Maps from Indicator Kriging with Houston's Traffic-Accident Data, 2010-2016. (Continued)

4.6.2. Empirical Bayesian Kriging

4.6.2.1. Introduction

Empirical Bayesian Kriging (EBK) is a newly introduced geostatistical-interpolation analysis tool in ArcGIS. On the ArcGIS tutorial webpage, Empirical Bayesian is defined as an interpolation process or method that applies repeated simulation techniques to negate the errors introduced by the semi-variogram. EBK provides a useful kriging model by automating the entire process that other tools perform manually. This kriging differs from the other ArcGIS geostatistical operations for the following reasons:

i. Automates the complete process: the other geostatistical tool needs manual adjustment for a better result.
ii. Concludes the error rise from the semi-variogram.

iii. Interpolates the region with the estimated semi-variogram while other interpolation methods execute the semi-variogram from the given data, thus creating uncertainty and a standard error in prediction.

iv. This tool delivers more accurate standard errors for prediction than the other kriging tools.

The research used the Empirical Bayesian Kriging tool to determine the prediction model for the accident counts in Houston, Texas. The accidents were presented as data points from XY coordinates on the ArcGIS maps geographically.

4.6.2.2. Procedure

There are necessary parameters to implement the Empirical Bayesian Kriging tool in ArcGIS, whereas the other parameters are optional. The EBK tool is located in the Interpolation tab under the Geostatistical Analysis Tools. The following procedure is applied for this research to use the EBK tool in ArcGIS; the first two parameters are necessary while the rest are optional:

i. Provide input features that define the data points to interpolate.

ii. Allocate point feature that provides the height or magnitude information for each field.

iii. Select Data Transformation Type if you want to transform the data. There are two types of transformation available in this tab: Empirical and Log-Empirical. The log-empirical transformation is recommended when you have positive values of magnitude or height for each count, such as the rainfall. However, the log-empirical transformation would not perform correctly if your data contain outliers; therefore, the result will provide prediction values which are considerably larger or smaller than the input magnitude. If a transformation is applied to the point feature, the software uses a simple kriging model along with a parametric distribution for the nugget, partial sill, and range. This research used the log-empirical data-transformation type for the accident-count data.

iv. Change the transformation type which allows the user to have multiple semi-variogram models. They have disadvantages and advantages which are defined in the user tutorial on the ArcGIS website. The options for the semi-variogram model type count on the sort of transformation you select for the database. This research chose the log-empirical transformation that features a list of six semi-variogram model types: exponential, exponential detrended, whittle, whittle detrended, K-Bessel, and K-Bessel Detrended. K-Bessel detrended was selected for this research data count because this method delivers the most flexible and accurate model while eliminating the first-order trend.

v. For the output parameter, you have the option to select the output surface type. In this category, ArcGIS lists four parameters: prediction, quantile, probability, and prediction standard error. The selection solely depends on the outcome to perceive the best fit resulting model. Initially, the research database used the prediction output to observe the effect; the outcome model was not satisfactory, and the values were comparatively higher than the defined magnitude. The outcome model was tested using probability using output surface type, and the result presented more acceptable values than the previous model.

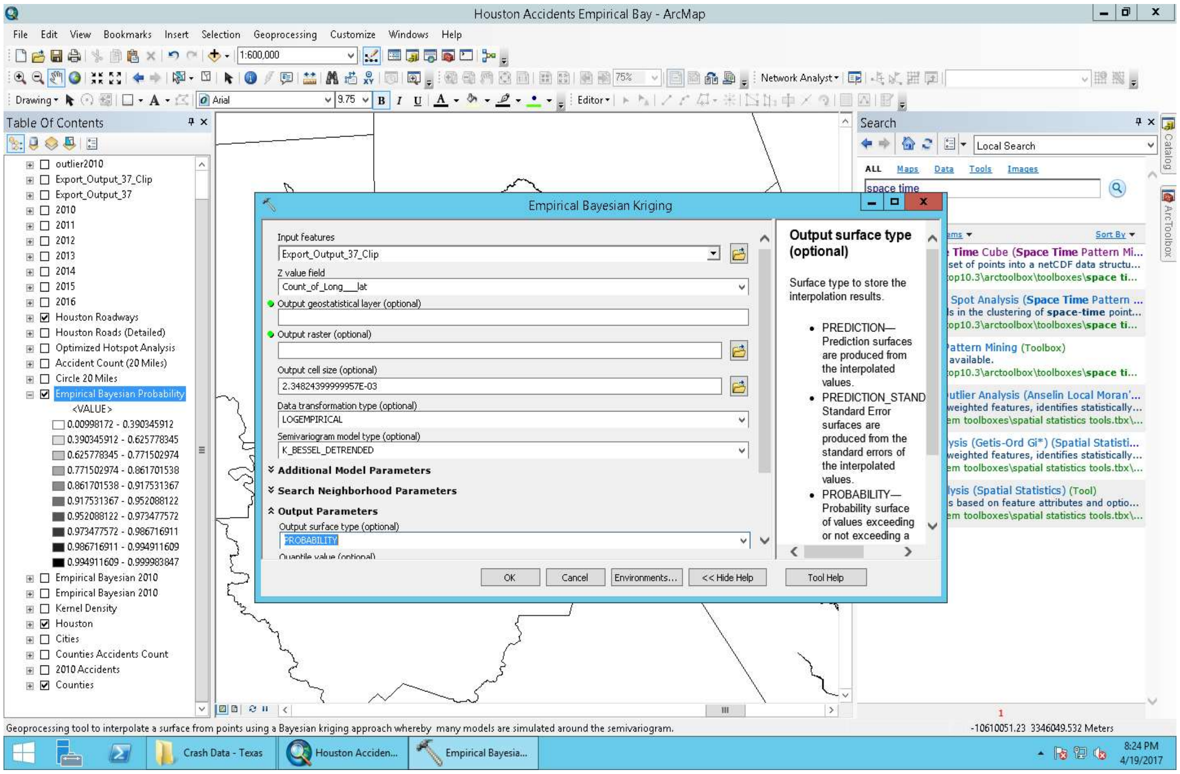


Figure 4.27. Input Parameters for the Empirical Bayesian Kriging Model.

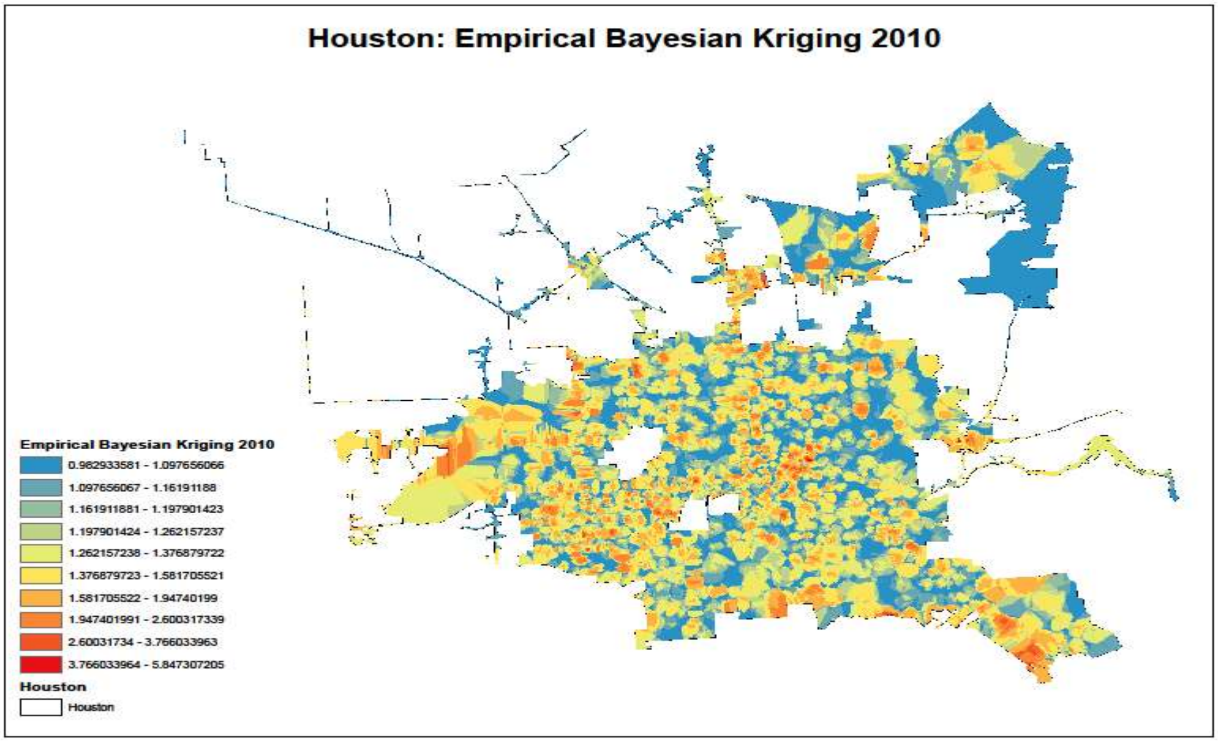
4.6.2.3. Results

The statistical values in Table 4.10 are obtained after executing the Empirical Bayesian Kriging for each year from 2010-2016. Table 4.10 displays values for the minimum, maximum, mean, and standard deviation of each year kriging analysis maps. In 2015 and 2016, the prediction map shows a negative value in the minimum-value column that can be ignored or rejected. The reason to disregard the negative values is merely understanding that the accident count cannot be a negative number. Other than 2015 and 2016, the values are positive and acceptable. The accident prediction map average is 1.3496 for all years, with a standard deviation of 0.3794.

Table 4.10. Statistical Values for the Resulting Empirical Bayesian Maps.

Year	Min. Value	Max. Value	Mean	Stand. Deviation
2010	0.983	5.847	1.309	0.3026
2011	1	7.238	1.3366	0.3268
2012	1	10.29635	1.343	0.3577
2013	1	7.057	1.383	0.411
2014	1	13.96	1.394	0.4257
2015	-0.9511	15.845	1.3393	0.4138
2016	-3.773	20.2958	1.342	0.4181
Average	0.0370	11.5056	1.3496	0.3794

The Empirical Bayesian Kriging model's outcome offers predicted accident values for Houston; detailed symbology maps are shown in Figure 4.28. Statistically, the predicted values for the overall maps range from 0 to 20 were classified in geometric interval with 10 classes each. The prediction map has a wide spread for the values at all Houston locations and displays a sporadic and scattered map. We can conclude from the map that the high value areas are spread at certain locations and scattered completely in maps. Similarly, the low-accident predicted sites spread, too, but their locality in the city's northeast outskirts represents the trend for few accidents in almost all maps. Figure 4.28 provides the outcome model for the Empirical Bayesian Kriging tool.



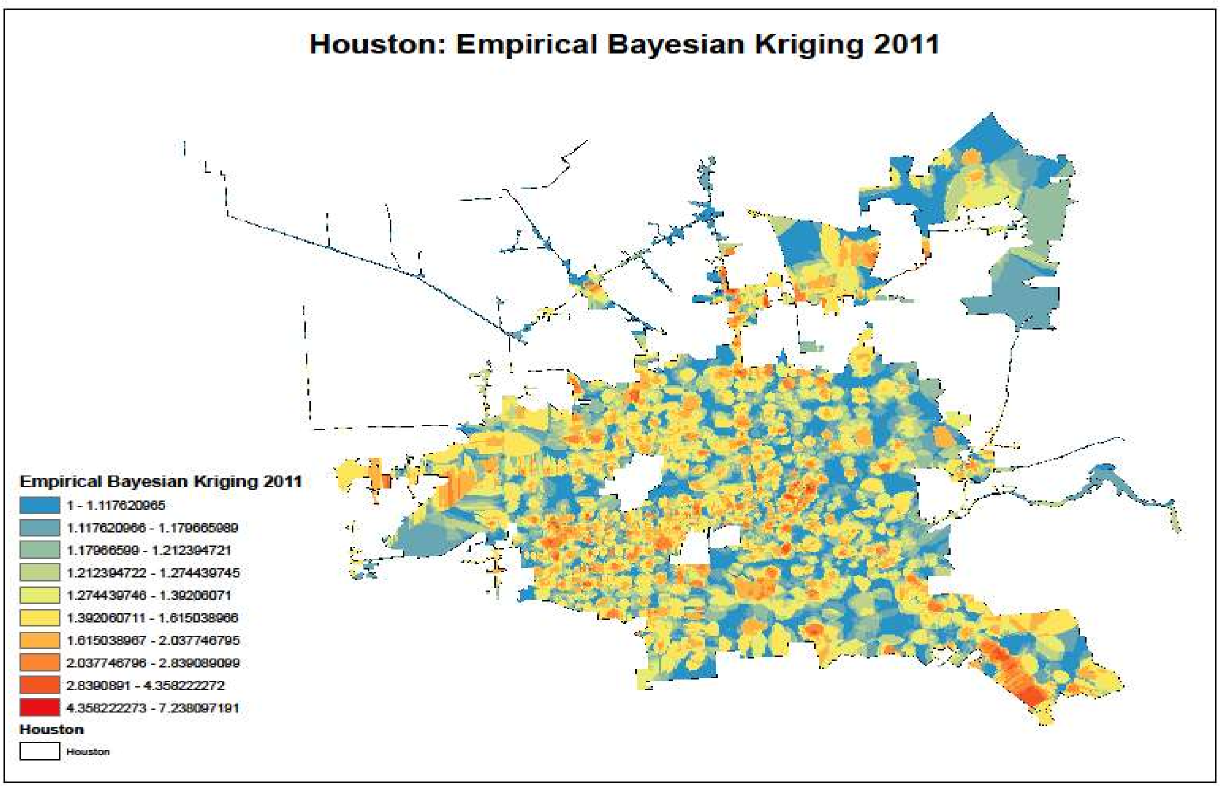
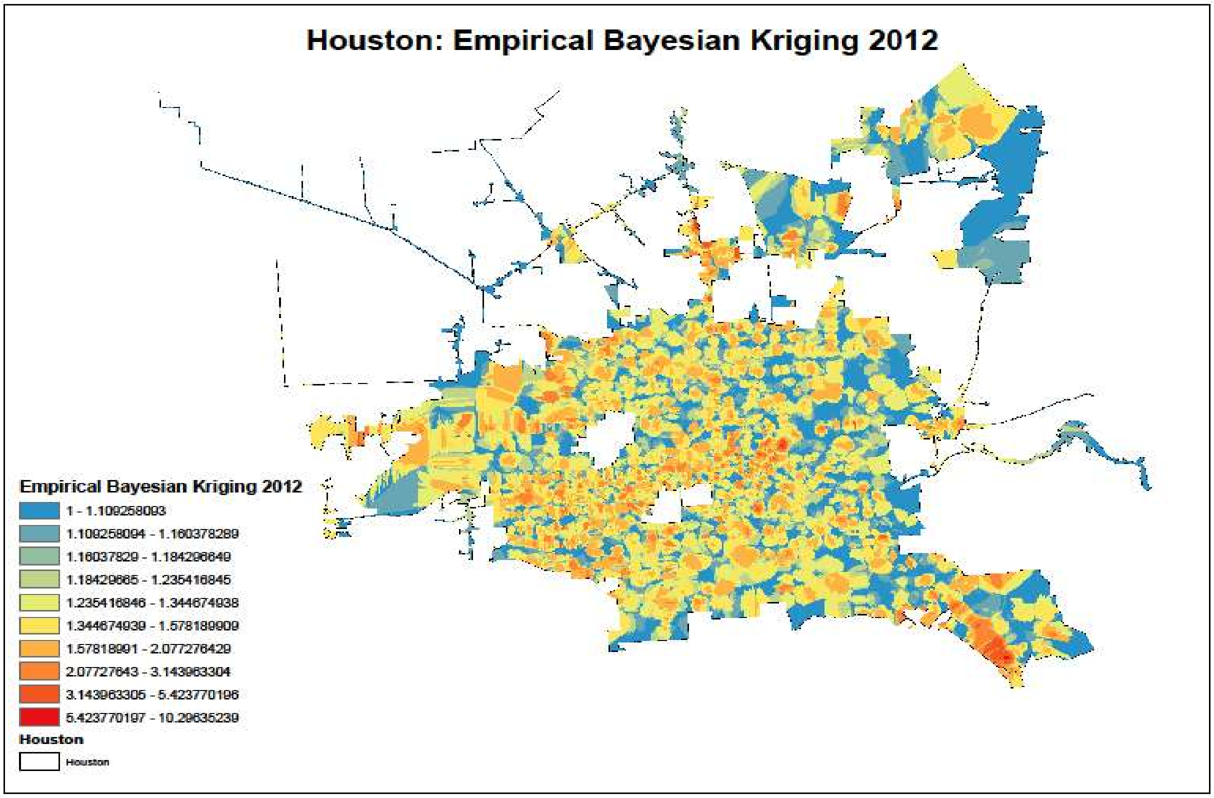


Figure 4.28. Maps that Result from Using the Empirical Bayesian Tool in ArcGIS.



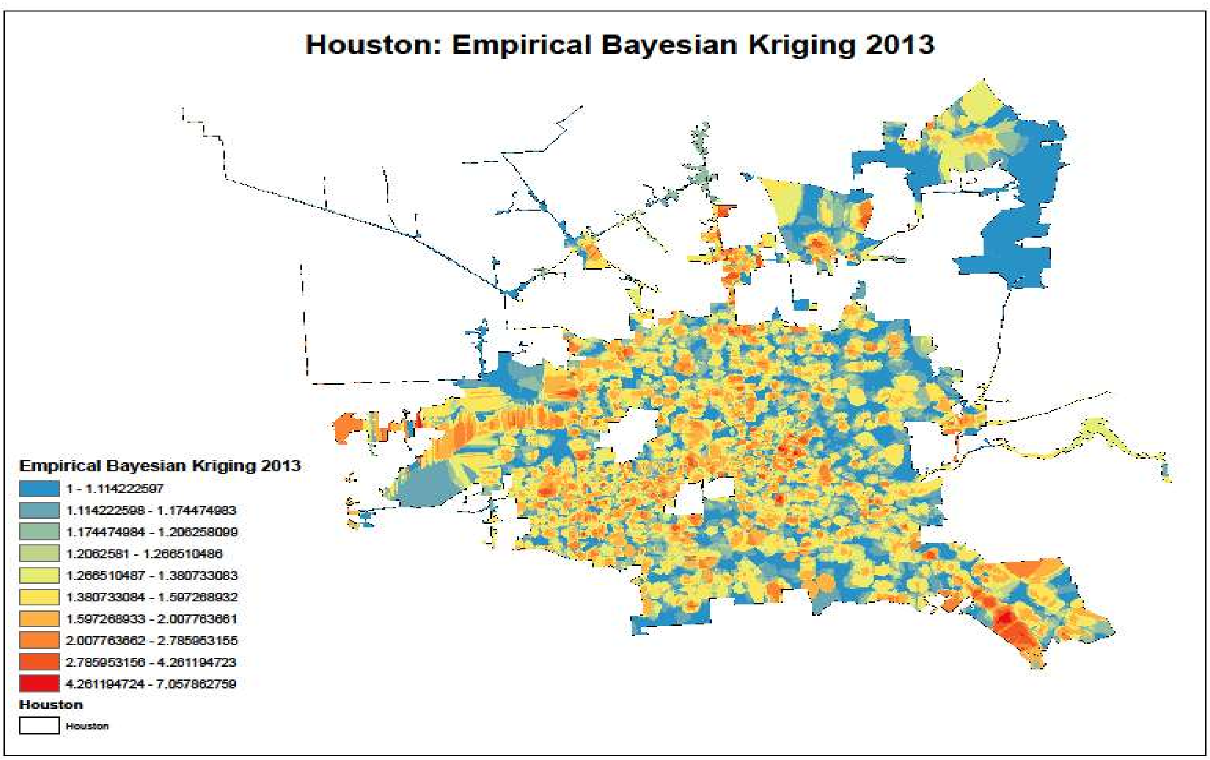
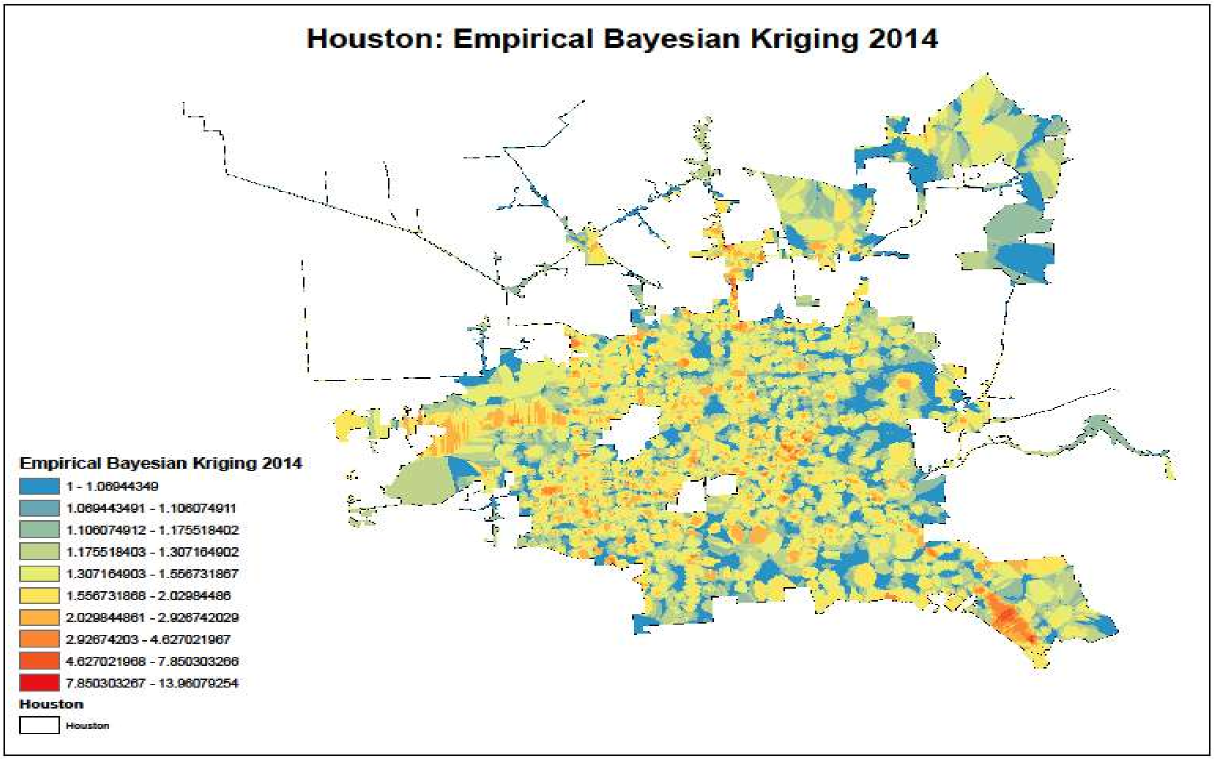


Figure 4.28. Maps that Result from Using the Empirical Bayesian Tool in ArcGIS. (Continued)



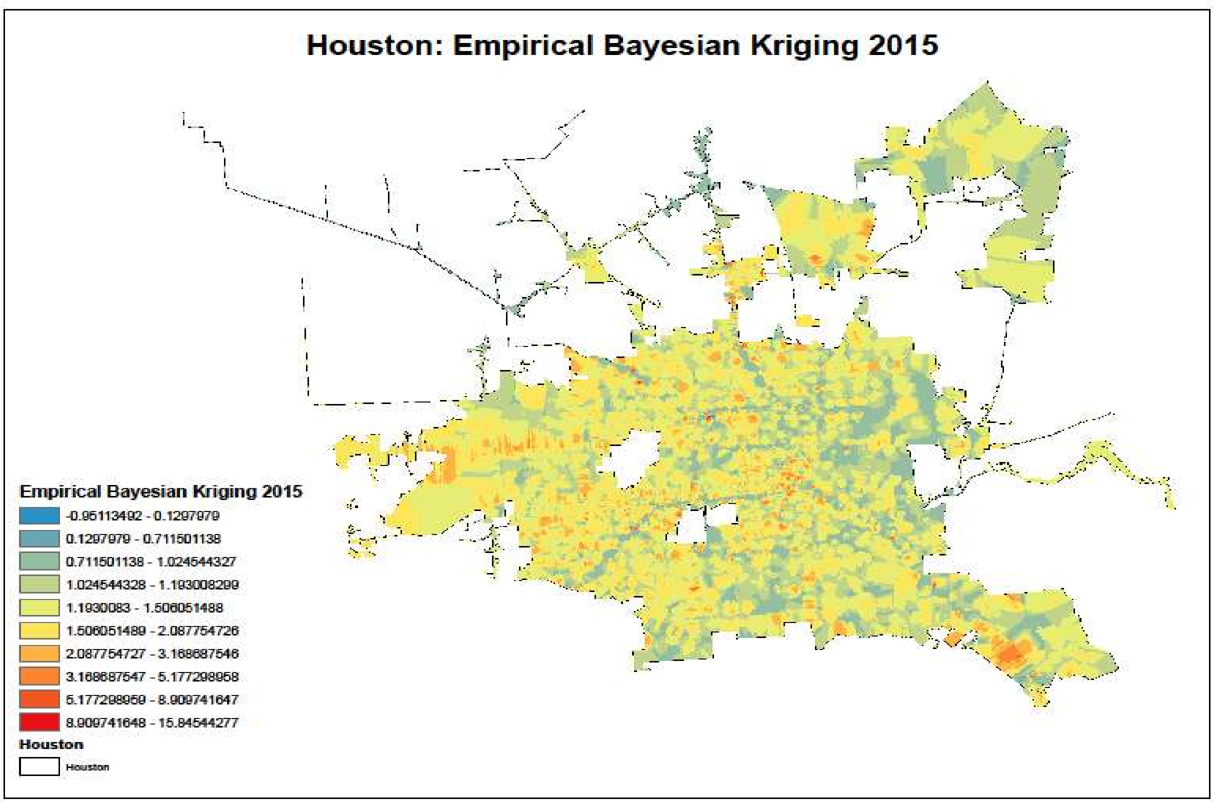


Figure 4.28. Maps that Result from Using the Empirical Bayesian Tool in ArcGIS. (Continued)

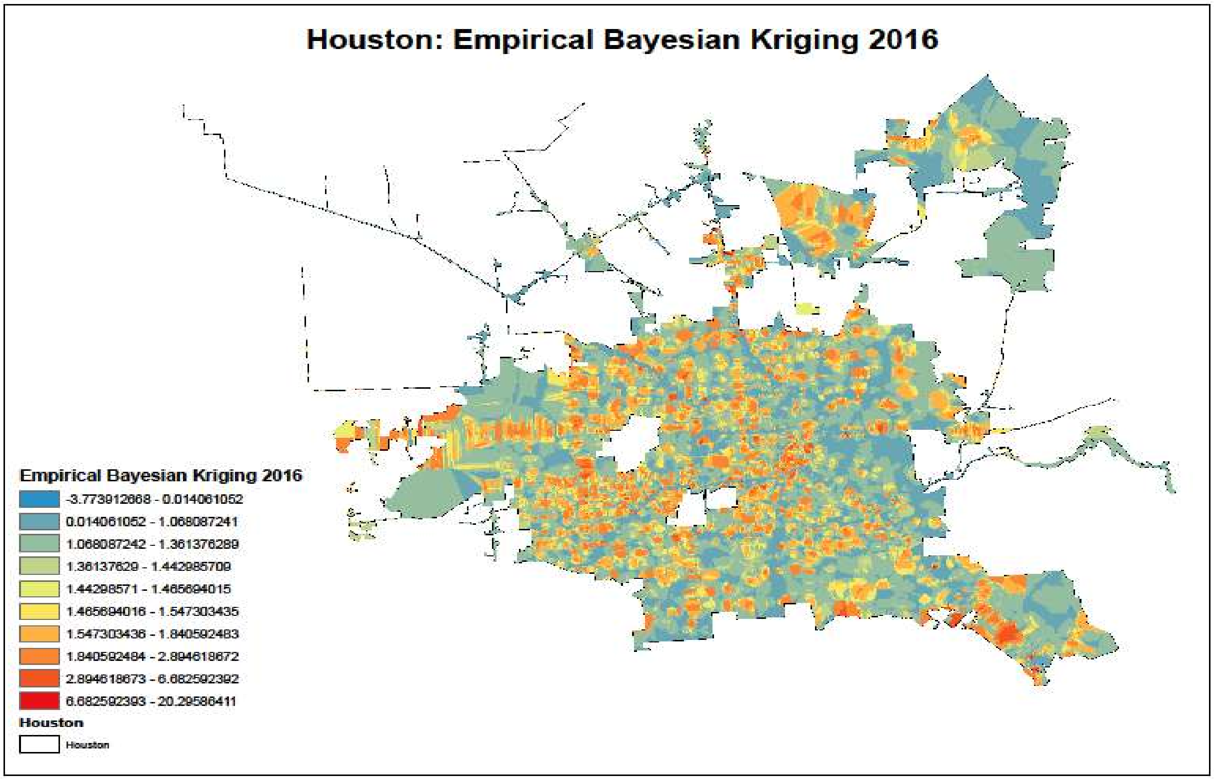


Figure 4.28. Maps that Result from Using the Empirical Bayesian Tool in ArcGIS. (Continued)

5. CONCLUSION AND RECOMMENDATION

5.1. Conclusion

The research was intended to provide a comprehensive viewpoint about the road conditions in Texas. A little information about Texas' fatalities and traffic accidents are availabe and the research was targeted to delve into issues about traffic endangerment in Texas by analyzing the accident database. By studying the Literature Review, it is evident that Texas has been the epicenter of chaotic and disordered road traffic that accounts for the most U.S. accidents and fatalities. Moreover, the less efforts have been exercised to counter the traffic issues as the NHTSA facts and figures advocate. Despite having a large expenditure for the yearly budget to improve road safety, the problem remains baffling, and more importantly, it has increased immensely in a couple of years. Having said that, the authorities' expenditure for traffic safety has not addressed the real problems and concerns about the state's accidents and fatalities. The reason could be because there is not enough information or statistics about the accidents' locations or characteristics. This research tried to communicate the more incite information and evidence on this issue.

It is imperative that Texas give serious attention to the traffic-accident problem and probably create an emergency safety plan. The research could help fashion a more precise safety and precautionary plan for road accidents. The authorities can incorporate significant findings from this study's analysis. This research focuses on a particular location in Texas and serves as a model that authorities can employ for similar approaches at different localities.

This study focuses on providing a realistic and holistic picture of the traffic conditions in Texas. The facts and figures are investigated for the 7-year period and are presented in a way that makes the results understandable to the public. The statistical values portray the importance of the given information, depending on comparable criteria and standards. From the findings, the injury-to-accident ratio was 0.453, and the death-to-fatality ratio for every year of the accident dataset was greater than 1. Moreover, novel findings also evolved from the statistical values, such as rural areas accounting for more fatalities than the populated places, whereas the accident and injury rates are higher in the major cities. After finding the above facts and information, the research determined the factors and parameters that triggered the record amount of accidents, injuries, and fatalities in particular conditions in Texas from 2010-2016. Of the total accident deaths in Texas, rural areas accounted for almost 50%, whereas the injury rate was highest in the urban locations that had nearly 60% of the overall injuries from 2010-2016. Similarly, the fatality and injury rate was highest for the roadway systems with low to mid-range average daily traffic (ADT).

Furthermore, the thorough Literature Review identified that intersections create major issues for traffic safety and contribute to a significant number of accidents. Therefore, the study also explored the accidents at Texas intersections and performed a statistical analysis to discover how much intersections contribute to traffic accidents, fatalities, and injuries. The results from the study anticipated and validates the contribution of traffic accidents on large scale at intersections. According to the values, the accident and injury rate at intersections was 0.275 and 0.359, respectively, for the study period. A similar approach was used to examine the accidents at intersections, and significant findings were discovered. The parameters initially identified after a systematic study at intersections were stop signs, speed, and signal lights.

Although the statistical analysis provided valuable information, the researcher was also interested in determining the spatial relationships among the accident counts and offering prediction models and hotspot locations by using ArcGIS. The dataset was too large to analyze the entire state of Texas; hence Houston was selected for the geostatistical analysis because the city is in the category for the highest number of accidents for Texas cities from 2010-2016. Therefore, the significance and relevancy to investigate the accidents in Houston is verified from the stated information. The research utilized various geostatistical analyses that include hotspot analysis, density analysis, space-time analysis, and kriging tools. These studies helped identify the problematic location within city that requires to be considered as high-accident areas and prediction maps for future consideration and reference in road safety planning.

Overall, this study's results verified the previous factual information and reached to new findings confirming that Texas is one of the deadliest U.S. states in road accidents; including fact that Texas is the 2nd biggest U.S. state by population. Therefore, the accidental problem exists on a larger scale. Texas accounts for most accidents in the United States and hence requires alarming consideration from the government authorities to initiate significant steps for corrective actions for friendly and safer environment in Texas highways.

5.2. Recommendation

The following recommendations can be considered for future research:

i. This study evaluated two factors: population and AD for key accidental findings from accident dataset. Therefore, a more thorough study can be conducted with a similar dataset for more perspectives about Texas' road accidents. Additional factors that can be studied include, but are not limited to, age, driving under the influence (DUI), and vehicle type.

ii. There are other geostatistical and geo-analysis tools available in ArcGIS; these tools can be used to analyze the traffic-accident dataset from a different perspective. A potential tool to examine the crash dataset is outlier analysis. However, there is no single tool that can suggest finest for the accident dataset. The researcher should compare the tool's performance by using the resulting maps and statistical information.

iii. The research focused on indicator and empirical kriging. Other kriging tools in ArcGIS were tested, but the outcomes were not conclusive and convincing. The accident dataset is discrete, rather than continuous data. Therefore, the dataset can be transformed in a way that the other kriging tools will produce acceptable results. Data-mining techniques can be utilized to transform the accident dataset.

iv. The research used a specific data-transformation and semi-variogram model to execute the Empirical Bayesian Kriging tool. In the future, other transformation types and semi-variogram models can syndicate to perform the Empirical Bayesian Kriging and to examine the prediction map.

v. It is important to mention that Texas is also one of the fastest growing states in the US and Houston specifically attracts people from all over the country and the world each with their own driving culture. This might be a challenging problem for that reason alone.

REFERENCES

Akepati, S. R., & Dissanayake, S. (2011). *Characteristics of the Work Zone Crashes.* Paper presented at the Transportation and Development Institute Congress 2011: Integrated Transportation and Development for a Better Tomorrow.

Akoz, O., & Karsligil, M. E. (2010). *Severity detection of traffic accidents at intersections based on vehicle motion analysis and multiphase linear regression.* Paper presented at the 13th International IEEE Conference on Intelligent Transportation Systems, ITSC 2010, September 19, 2010 - September 22, 2010, Funchal, Portugal.

Alam, B. M. (2011). Case-Based Analysis of Age and Sex Distribution of Drivers Causing Fatal Crashes: Evidence from Florida, USA *ICTIS 2011: Multimodal Approach to Sustained Transportation System Development: Information, Technology, Implementation* (pp. 1113-1121).

Anderson, S. A., & DeMarco, M. J. (2013). Use of Rockfall Rating Systems in the Design of New Slopes *GeoChallenges: Rising to the Geotechnical Challenges of Colorado* (pp. 37-53).

Appiah, J., Rilett, L. R., Naik, B., & Wojtal, R. (2012). Driver response to an actuated advance warning system. *Journal of transportation engineering, 139*(5), 433-440.

ArcGIS. (2017). Create Space Time Cube By Aggregating Points. Retrieved from *http://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/create-space-time-cube.htm*

ArcGIS. (2017). Local Outlier Analysis. Retrieved from *http://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/localoutlieranalysis.htm*

Azizi, L., & Sheikholeslami, A. (2013). Safety Effect of U-Turn Conversions in Tehran: Empirical Bayes Observational Before-and-After Study and Crash Prediction Models. *Journal of transportation engineering, 139*(1), 101-108. doi:10.1061/(asce)te.1943-5436.0000469

Bai, L., Liu, P., Li, Z.-b., & Xu, C.-c. (2011). Using Multivariate Poisson-Lognormal Regression Method for Modeling Crash Frequency by Severity on Freeway Diverge Areas *ICCTP 2011: Towards Sustainable Transportation Systems* (pp. 2385-2394).

Basbas, S. (2005). *Road accident risk models: The use of SafeNET in Thessaloniki.* Paper presented at the 11th International Conference on Urban Transport and the Environment in the 21st Century, Urban Transport 2005, April 12, 2005 - April 14, 2005, Algarve, Portugal.

Bhalla, P., Tripathi, S., & Palria, S. (2014). *Road traffic accident analysis of Ajmer City using remote sensing and GIS technology.* Paper presented at the ISPRS Technical Commission VIII Mid-Term Symposium 2014, December 9, 2014 - December 12, 2014, Hyderabad, India.

Carriquiry, A., & Pawlovich, M. (2004). From empirical Bayes to full Bayes: methods for analyzing traffic safety data. *White Paper, Iowa State University*.

Chang, K., Wu, C.-C., & Ying, Y.-H. (2012). The effectiveness of alcohol control policies on alcohol-related traffic fatalities in the United States. *Accident Analysis & Prevention, 45*, 406-415.

Chen, B., He, C., & Wang, J. (2011). Freeway accident detection model based on support vector machine *ICTE 2011* (pp. 3104-3109).

Chen, F., & Chen, S. (2011). Injury severities of truck drivers in single-and multi-vehicle accidents on rural highways. *Accident Analysis & Prevention, 43*(5), 1677-1688.

Chen, H. (2014). Exploring Contributing Factors to Driver Injury Levels during Nighttime at Local Collectors *CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation Systems* (pp. 2455-2462).

Chen, R., Zhang, M., Li, Z., & Wang, W. (2012). Development of Crash Prediction Model for Rear-End Collisions at Recurrent Bottlenecks on Freeways *CICTP 2012: Multimodal Transportation Systems—Convenient, Safe, Cost-Effective, Efficient* (pp. 2591-2602).

Chen, Y., Liu, C., Wu, H., & Sun, W. (2011). *Identification of black spot on traffic accidents and its spatial association analysis based on geographic information system.* Paper presented at the Natural Computation (ICNC), 2011 Seventh International Conference on.

Chimba, D., Kutela, B., Ogletree, G., Horne, F., & Tugwell, M. (2013). Impact of abandoned and disabled vehicles on freeway incident duration. *Journal of transportation engineering, 140*(3), 04013013.

Choe, T., Gordon, I., George E, & Martinez, J., Guillermo. (2013). Port of Long Beach and Port of Los Angeles Advanced Transportation Management and Information System (ATMIS) *Ports 2013: Success through Diversification* (pp. 1434-1443).

Curry, A. E., Hafetz, J., Kallan, M. J., Winston, F. K., & Durbin, D. R. (2011). Prevalence of teen driver errors leading to serious motor vehicle crashes. *Accident Analysis & Prevention, 43*(4), 1285-1290.

Driss, M., Saint-Gerand, T., Bensaid, A., Benabdeli, K., & Hamadouche, M. A. (2013). *A fuzzy logic model for identifying spatial degrees of exposure to the risk of road accidents (Case study of the Wilaya of Mascara, Northwest of Algeria).* Paper presented at the Advanced Logistics and Transport (ICALT), 2013 International Conference on.

Durduran, S. S. (2010). A decision making system to automatic recognize of traffic accidents on the basis of a GIS platform. *Expert Systems with Applications, 37*(12), 7729-7736.

Elder, R. W., Lawrence, B., Ferguson, A., Naimi, T. S., Brewer, R. D., Chattopadhyay, S. K., . . . Services, T. F. o. C. P. (2010). The effectiveness of tax policy interventions for reducing excessive alcohol consumption and related harms. *American journal of preventive medicine, 38*(2), 217-229.

Elghamrawy, T., El-Rayes, K., & Liu, L. (2010). *Analysis of Injury and Fatal Crashes in Highway Construction Zones.* Paper presented at the Construction Research Congress 2010: Innovation for Reshaping Construction Practice.

Elvik, R. (2016). Does the influence of risk factors on accident occurrence change over time? *Accident Analysis & Prevention, 91*, 91-102.

Fitzpatrick, K., Lord, D., & Park, B.-J. (2010). Horizontal curve accident modification factor with consideration of driveway density on rural four-lane highways in texas. *Journal of transportation engineering, 136*(9), 827-835.

Greibe, P. (2003). Accident prediction models for urban roads. *Accid Anal Prev, 35*(2), 273-285. doi:10.1016/S0001-4575(02)00005-2

Guo, F., Wang, X., & Abdel-Aty, M. A. (2010). Modeling signalized intersection safety with corridor-level spatial correlations. *Accident Analysis & Prevention, 42*(1), 84-92.

Guo, Y., & Sun, Q. (2013). Modeling Crash Frequency of A Typical Mountainous Freeway *ICTIS 2013: Improving Multimodal Transportation Systems-Information, Safety, and Integration* (pp. 1417-1425).

Guo, Z., Shang, Z., Wang, F., & Sun, Z. (2000). Photogrammetric Model and Technology Used in Road Traffic Accident Scene Measurement *Traffic and Transportation Studies (2000)* (pp. 88-92).

Hajizamani, M., Shrubsall, S. C., & Viegas, J. M. (2011). Evaluation of In-Vehicle Alcohol Intake Control Devices Using Agent-Based Modeling (ABM) *ICTIS 2011: Multimodal Approach to Sustained Transportation System Development: Information, Technology, Implementation* (pp. 1555-1561).

Haleem, K., Gan, A., & Alluri, P. (2014). *Exploration of Fractal Characteristics in Crash Data.* Paper presented at the T&DI Congress 2014: Planes, Trains, and Automobiles.

Hallowell, M. R., Teizer, J., & Blaney, W. (2010). *Application of sensing technology to safety management.* Paper presented at the Construction Research Congress 2010: Innovation for Reshaping Construction Practice.

Hancock, K., Zhang, W., Sardar, H., & Wang, Y. (2016). *Underlying Relationships between Fatal Crashes and All Other Non-Fatal Crashes.* Paper presented at the International Conference on Transportation and Development 2016.

Hao, W., Kamga, C., & Daniel, J. (2015). The effect of age and gender on motor vehicle driver injury severity at highway-rail grade crossings in the United States. *Journal of safety research, 55*, 105-113.

Haque, M. M., Chin, H. C., & Debnath, A. K. (2012). An investigation on multi-vehicle motorcycle crashes using log-linear models. *Safety Science, 50*(2), 352-362. doi:10.1016/j.ssci.2011.09.015

Hauer, E. (1986). On the estimation of the expected number of accidents. *Accident Analysis & Prevention, 18*(1), 1-12.

Hauer, E. (1996a). Detection of safety deterioration in a series of accident counts. *Transportation Research Record: Journal of the Transportation Research Board*(1542), 38-43.

Hauer, E. (1996b). Identification of sites with promise. *Transportation Research Record: Journal of the Transportation Research Board*(1542), 54-60.

Higa, L., & Kim, J.-L. (2013). Evaluating Accident Data for the Safety of Nighttime Construction in Southern California *ICSDEC 2012: Developing the Frontier of Sustainable Design, Engineering, and Construction* (pp. 711-718).

Higle, J. L., & Witkowski, J. M. (1988). *Bayesian Identification of Hazardous Locations (with Discussion and Closure)*.

Hosseinpour, M., Yahaya, A. S., Ghadiri, S. M., & Prasetijo, J. (2013). Application of Adaptive Neuro-fuzzy Inference System for road accident prediction. *KSCE Journal of Civil Engineering, 17*(7), 1761-1772. doi:10.1007/s12205-013-0036-3

Huang, H., & Abdel-Aty, M. (2010). Multilevel data and Bayesian analysis in traffic safety. *Accident Analysis & Prevention, 42*(6), 1556-1565.

Huang, H., Chin, H., & Haque, M. (2009). Empirical Evaluation of Alternative Approaches in Identifying Crash Hot Spots. *Transportation Research Record: Journal of the Transportation Research Board, 2103*(2103), 32-41. doi:10.3141/2103-05

Ige, J., Banstola, A., & Pilkington, P. (2016). Mobile phone use while driving: Underestimation of a global threat. *Journal of Transport & Health, 3*(1), 4-8.

Isebrands, H. N., Hallmark, S. L., Li, W., McDonald, T., Storm, R., & Preston, H. (2010). Roadway lighting shows safety benefits at rural intersections. *Journal of transportation engineering, 136*(11), 949-955.

Jermakian, J. S. (2011). Crash avoidance potential of four passenger vehicle technologies. *Accident Analysis & Prevention, 43*(3), 732-740.

Jung, S., Qin, X., & Noyce, D. A. (2011). Injury severity of multivehicle crash in rainy weather. *Journal of transportation engineering, 138*(1), 50-59.

Kang, M.-W., Momtaz, S. U., & Barnett, T. E. (2015). Crash Analysis and Public Survey for Drowsy-Driving Advisory Systems. *Journal of transportation engineering, 141*(9), 04015016.

Kaplan, S., & Prato, C. G. (2012). Risk factors associated with bus accident severity in the United States: A generalized ordered logit model. *Journal of safety research, 43*(3), 171-180.

Kelley-Baker, T., & Romano, E. (2010). Female involvement in US nonfatal crashes under a three-level hierarchical crash model. *Accident Analysis & Prevention, 42*(6), 2007-2012.

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England journal of medicine, 370*(1), 54-59.

Knecht, C., Saito, M., & Schultz, G. G. (2016). *Development of Crash Prediction Models for Curved Segments of Rural Two-Lane Highways.* Paper presented at the International Conference on Transportation and Development 2016.

Kumar, S., & Toshniwal, D. (2016). A novel framework to analyze road accident time series data. *Journal of Big Data, 3*(1), 8.

Kumeta, K., Miyake, A., & Ogawa, T. (2006). Prediction of accident frequency for road transport of dangerous goods using cluster analysis. *Science and Technology of Energetic Materials, 67*(1), 17-22.

Kwon, J. Y., Mauch, M., & Varaiya, P. (2006). Components of congestion - Delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand. *Freeway Operations and High Occupancy Vehicle Systems 2006*(1959), 84-91.

Lee, S. E., Simons-Morton, B. G., Klauer, S. E., Ouimet, M. C., & Dingus, T. A. (2011). Naturalistic assessment of novice teenage crash experience. *Accident Analysis & Prevention, 43*(4), 1472-1479.

Li, G., Brady, J. E., & Chen, Q. (2013). Drug use and fatal motor vehicle crashes: a case-control study. *Accident Analysis & Prevention, 60*, 205-210.

Li, X., Lord, D., & Zhang, Y. (2010). Development of accident modification factors for rural frontage road segments in Texas using generalized additive models. *Journal of transportation engineering, 137*(1), 74-83.

Li, Y., Cheng, F., & Bai, Y. (2012). Characteristics of Truck-related Crashes in Highway Work Zones *Sustainable Transportation Systems: Plan, Design, Build, Manage, and Maintain* (pp. 364-371).

Li, Z., Kepaptsoglou, K., Lee, Y., Patel, H., Liu, Y., & Kim, H. G. (2013). Safety effects of shoulder paving for rural and urban interstate, multilane, and two-lane highways. *Journal of transportation engineering, 139*(10), 1010-1019.

Li, Z., Lee, S. H., Lee, Y., Zhou, B., & Bamzai, R. (2011). *A Methodology for Assessing Safety Impacts of Highway Shoulder Paving.* Paper presented at the Transportation and Development Institute Congress 2011: Integrated Transportation and Development for a Better Tomorrow.

Li, Z., Lee, Y., Lee, S. H., & Valiou, E. (2011). *Geographically-weighted regression models for improved predictability of urban intersection vehicle crashes.* Paper presented at the

Transportation and Development Institute Congress 2011: Integrated Transportation and Development for a Better Tomorrow.

Ma, J., Fontaine, M. D., Zhou, F., Hu, J., Hale, D. K., & Clements, M. O. (2016). Estimation of Crash Modification Factors for an Adaptive Traffic-Signal Control System. *Journal of transportation engineering, 142*(12), 04016061.

Ma, J., & Li, Z. (2010). Bayesian modeling of frequency-severity indeterminacy with an application to traffic crashes on two-lane highways *ICCTP 2010: Integrated Transportation Systems: Green, Intelligent, Reliable* (pp. 1022-1033).

Ma, X., Chen, S., & Chen, F. (2016). Correlated random-effects bivariate poisson lognormal model to study single-vehicle and multivehicle crashes. *Journal of transportation engineering, 142*(11), 04016049.

MacLeod, K. E., Griswold, J. B., Arnold, L. S., & Ragland, D. R. (2012). Factors associated with hit-and-run pedestrian fatalities and driver identification. *Accident Analysis & Prevention, 45*, 366-372.

Masten, S. V., Foss, R. D., & Marshall, S. W. (2011). Graduated driver licensing and fatal crashes involving 16-to 19-year-old drivers. *Jama, 306*(10), 1098-1103.

Matsuzaki, K., Nitta, M., & Kato, K. (2008). *Development of an intelligent traffic light for reducing traffic accidents.* Paper presented at the Control, Automation and Systems, 2008. ICCAS 2008. International Conference on.

Medina, J., Shen, S., & Benekohal, R. (2014). *Microscopic Analysis for Accident Data at Railroad Grade Crossings.* Paper presented at the T&DI Congress 2014: Planes, Trains, and Automobiles.

Mehta, G., Li, J., Fields, R. T., Lou, Y., & Jones, S. (2015). Safety Performance Function Development for Analysis of Bridges. *Journal of transportation engineering, 141*(8), 04015010.

Mishra, S., & Khasnabis, S. (2011). Optimization model for allocating resources for highway safety improvement at urban intersections. *Journal of transportation engineering, 138*(5), 535-547.

Morgan, A., & Mannering, F. L. (2011). The effects of road-surface conditions, age, and gender on driver-injury severities. *Accident Analysis & Prevention, 43*(5), 1852-1863.

Nassar, S. A., Saccomanno, F. F., & Shortreed, J. H. (1994). Road accident severity analysis: a micro level approach. *Canadian Journal of Civil Engineering, 21*(5), 847-855.

National Highway Traffic Safety Administration, U. (2016). 2015 motor vehicle crashes: overview. *Traffic safety facts research note, 2016*, 1-9.

Naumann, R. B., Dellinger, A. M., Zaloshnja, E., Lawrence, B. A., & Miller, T. R. (2010). Incidence and total lifetime costs of motor vehicle–related fatal and nonfatal injury by road user type, United States, 2005. *Traffic injury prevention, 11*(4), 353-360.

Neider, M. B., McCarley, J. S., Crowell, J. A., Kaczmarski, H., & Kramer, A. F. (2010). Pedestrians, vehicles, and cell phones. *Accident Analysis & Prevention, 42*(2), 589-594.

Nourzad, S. H. H., Salvucci, D. D., & Pradhan, A. (2014). Computational Modeling of Driver Distraction by Integrating Cognitive and Agent-based Traffic Simulation Models *Computing in Civil and Building Engineering (2014)* (pp. 1885-1892).

Oppe, S. (1992). A Comparison of Some Statistical Techniques for Road Accident Analysis. *Accident Analysis and Prevention, 24*(4), 397-423. doi:Doi 10.1016/0001-4575(92)90052-K

Ozbay, K., & Kachroo, P. (1999). Incident management in intelligent transportation systems.

Park, P. Y., Miranda-Moreno, L. F., & Saccomanno, F. F. (2010). Estimation of speed differentials on rural highways using hierarchical linear regression models. *Canadian Journal of Civil Engineering, 37*(4), 624-637. doi:10.1139/L10-002

Park, S. Y., Lan, C.-L., Chang, G.-L., Tolani, D., & Huang, P. (2016). Design and Predeployment Assessment of an Integrated Intersection Dilemma Zone Protection System. *Journal of transportation engineering, 142*(12), 04016063.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., & Mathers, C. D. (2004). World report on road traffic injury prevention: World Health Organization Geneva.

Persaud, B., Lan, B., Lyon, C., & Bhim, R. (2010). Comparison of empirical Bayes and full Bayes approaches for before–after road safety evaluations. *Accident Analysis & Prevention, 42*(1), 38-43.

Persaud, B. N. (1988). *Do traffic signals affect safety? Some methodological issues*.

Philip, P., Sagaspe, P., Lagarde, E., Leger, D., Ohayon, M. M., Bioulac, B., . . . Taillard, J. (2010). Sleep disorders and accidental risk in a large group of regular registered highway drivers. *Sleep medicine, 11*(10), 973-979.

Prescott, M. (2016, October 18). Empowering HGIS Research at ESRI Canada User Conference October 2016. Retrieved from *https://empiretimber.wordpress.com/project-blog/page/3/*

Pressman, M. R. (2011). Sleep driving: sleepwalking variant or misuse of z-drugs? *Sleep medicine reviews, 15*(5), 285-292.

Pulugurtha, S. S., & Nujjetty, A. P. (2011). *Crash Estimation Models for Intersections.* Paper presented at the Transportation and Development Institute Congress 2011: Integrated Transportation and Development for a Better Tomorrow.

Pulugurtha, S. S., & Pasupuleti, N. (2013). Geo-Spatial and Statistical Methods to Model Intracity Truck Crashes *Green Streets, Highways, and Development 2013: Advancing the Practice* (pp. 251-261).

Qing, Y., & Zhongyin, G. (2015). *Road Traffic Accident Forecasts Based on Cointegration Analysis.* Paper presented at the ICTE 2015. Fifth International Conference on Transportation Engineering, 26-27 Sept. 2015, Reston, VA, USA.

Qu, X., Wang, W., & Wang, W. (2011). Identification of traffic conditions leading to sideswipe crashes on freeways *ICCTP 2011: Towards Sustainable Transportation Systems* (pp. 2092-2101).

Rogers, J. H., Al-Deek, H., & Sandt, A. (2014). *Wrong-Way Driving Incidents on Central Florida Toll Road Network, Phase-1 Study: An Investigation into the Extent of this Problem?* Paper presented at the T&DI Congress 2014: Planes, Trains, and Automobiles.

Rui, T., Zhaosheng, Y., & Maolei, Z. (2010). *Method of Road Traffic Accidents Causes Analysis Based on Data Mining.* Paper presented at the 2010 International Conference on Computational Intelligence and Software Engineering (CiSE 2010), 10-12 Dec. 2010, Piscataway, NJ, USA.

SAS. (2013, December 17). Characteristics of Semivariogram Models. Retrieved from *http://support.sas.com/documentation/cdl/en/statug/66859/HTML/default/viewer.htm#statug_variogram_details02.htm*

Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. *Accident Analysis & Prevention, 43*(5), 1666-1676.

Schubert, R., & Wanielik, G. (2011). *Empirical evaluation of a unified Bayesian object and situation assessment approach for lane change assistance.* Paper presented at the 14th IEEE International Intelligent Transportation Systems Conference, ITSC 2011, October 5, 2011 - October 7, 2011, Washington, DC, United states.

Schultz, G. G., Black, C. W., & Saito, M. (2014). *GIS Framework for Hierarchical Bayesian based Crash Data Analysis.* Paper presented at the T&DI Congress 2014: Planes, Trains, and Automobiles.

Schultz, G. G., Farnsworth, J. S., & Saito, M. *Mitigating Safety in Utah Using the Hot Spot Identification and Analysis Methodology.* Paper presented at the International Conference on Transportation and Development 2016.

Schwebel, D. C., Combs, T., Rodriguez, D., Severson, J., & Sisiopiku, V. (2016). Community-based pedestrian safety training in virtual reality: A pragmatic trial. *Accident Analysis & Prevention, 86*, 9-15.

Schwebel, D. C., Stavrinos, D., Byington, K. W., Davis, T., O'Neal, E. E., & De Jong, D. (2012). Distraction and pedestrian safety: how talking on the phone, texting, and listening to music impact crossing the street. *Accident Analysis & Prevention, 45*, 266-271.

Seeherman, J., & Liu, Y. (2015). Effects of extraordinary snowfall on traffic safety. *Accident Analysis & Prevention, 81*, 194-203.

Shams, A., & Dissanayake, S. (2014). *Improving Safety at Unsignalized Rural Intersections in Kansas.* Paper presented at the T&DI Congress 2014: Planes, Trains, and Automobiles.

Shanthi, S., & Ramani, R. G. (2012). *Gender specific classification of road accident patterns through data mining techniques.* Paper presented at the Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on.

Shen, Y., Hermans, E., Ruan, D., Wets, G., Vanhoof, K., & Brijs, T. (2008). *Development of a composite road safety performance indicator based on neural networks.* Paper presented at the Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on.

Simons-Morton, B. G., Ouimet, M. C., Zhang, Z., Klauer, S. E., Lee, S. E., Wang, J., . . . Dingus, T. A. (2011). The effect of passengers and risk-taking friends on risky driving and crashes/near crashes among novice teenagers. *Journal of Adolescent Health, 49*(6), 587-593.

Skabardonis, A., Noeimi, H., Petty, K., Rydzewski, D., Varaiya, P., & Al-Deek, H. (1995). Freeway service patrol evaluation. *California Partners for Advanced Transit and Highways (PATH)*.

Srinivasan, R., Ullman, G., Finley, M., & Council, F. (2011). Use of Empirical Bayesian Methods to Estimate Crash Modification Factors for Daytime Versus Nighttime Work Zones. *Transportation Research Record: Journal of the Transportation Research Board*(2241), 29-38.

Tefft, B. C. (2013). Impact speed and a pedestrian's risk of severe injury or death. *Accident Analysis & Prevention, 50*, 871-878.

Tregear, S., Reston, J., Schoelles, K., & Phillips, B. (2010). Continuous positive airway pressure reduces risk of motor vehicle crash among drivers with obstructive sleep apnea. *Sleep, 33*(10), 1373-1380.

Tymvios, N., & Gambatese, J. (2014). *Evaluation of a Mobile Work Zone Barrier System.* Paper presented at the Construction Research Congress 2014: Construction in a Global Network.

Vadlamani, S., Chen, E., Ahn, S., & Washington, S. (2010). Identifying large truck hot spots using crash counts and PDOEs. *Journal of transportation engineering, 137*(1), 11-21.

Valentin, V., Mannering, F. L., Abraham, D. M., & Dunston, P. S. (2010). Evaluation of the visibility of workers' safety garments during nighttime highway-maintenance operations. *Journal of transportation engineering, 136*(6), 584-591.

Villwock, N. M., Blond, N., & Tarko, A. P. (2010). Cable barriers and traffic safety on rural interstates. *Journal of transportation engineering, 137*(4), 248-259.

Wagenaar, A. C., Tobler, A. L., & Komro, K. A. (2010). Effects of alcohol tax and price policies on morbidity and mortality: a systematic review. *American Journal of Public Health, 100*(11), 2270-2278.

Wang, J., Chen, X., Zhou, K., Wang, W., & Zhang, D. (2008). *Application of spatial data mining in accident analysis system.* Paper presented at the Education Technology and Training, 2008. and 2008 International Workshop on Geoscience and Remote Sensing. ETT and GRS 2008. International Workshop on.

Wang, J., Zhang, Q., Wang, Y., Weng, J., & Yan, X. (2016). Analysis of Sideswipe Collision Precursors Considering the Spatial-Temporal Characteristics of Freeway Traffic. *Journal of transportation engineering, 142*(12), 04016064.

Wang, Q., & Liu, S. (2009). *An information renewal GNN model for road traffic accident forecasting.* Paper presented at the 2nd International Conference on Transportation Engineering, ICTE 2009, July 25, 2009 - July 27, 2009, Chengdu, China.

Wang, X., Chen, X., & Sun, H. (2010). Analysis of the Safety Influence Area for 4-Legged Signalized Intersections *ICCTP 2010: Integrated Transportation Systems: Green, Intelligent, Reliable* (pp. 903-913).

Wang, Z., Lu, J. J., Wang, Q., Lu, L., & Zhang, Z. (2010). Modeling Injury Severity in Work Zones using Ordered PROBIT Regression *ICCTP 2010: Integrated Transportation Systems: Green, Intelligent, Reliable* (pp. 1058-1067).

Wilson, F. A., & Stimpson, J. P. (2010). Trends in fatalities from distracted driving in the United States, 1999 to 2008. *American Journal of Public Health, 100*(11), 2213-2219.

Wu, H., Gao, L., & Zhang, Z. (2013). Analysis of crash data using quantile regression for counts. *Journal of transportation engineering, 140*(4), 04013025.

Wu, K.-F., Donnell, E. T., Himes, S. C., & Sasidharan, L. (2013). Exploring the association between traffic safety and geometric design consistency based on vehicle speed metrics. *Journal of transportation engineering, 139*(7), 738-748.

Xu, X., Teng, H., & Kwigizile, V. (2011). Safety impact of access management techniques at signalized intersections *ICTIS 2011: Multimodal Approach to Sustained Transportation System Development: Information, Technology, Implementation* (pp. 403-416).

Xu, X., Teng, H., Kwigizile, V., & Mulokozi, E. (2014). Modeling signalized-intersection safety with corner clearance. *Journal of transportation engineering, 140*(6), 04014016.

Yang, S., Lu, S., & Wu, Y.-J. (2013). GIS-based Economic Cost Estimation of Traffic Accidents in St. Louis, Missouri. *Procedia-Social and Behavioral Sciences, 96*, 2907-2915.

Ye, Z., Shi, X., Huang, J., & Wang, S. (2015). Quality Control of Precipitation Data for Wet Pavement Accident Analysis *Environmental Sustainability in Transportation Infrastructure* (pp. 25-40).

Zegeer, C. V., & Bushell, M. (2012). Pedestrian crash trends and potential countermeasures from around the world. *Accident Analysis & Prevention, 44*(1), 3-11.

Zhang, G., Sun, L., Lou, J., Xu, S., & Jiang, X. (2013). Driver Factors Analysis of Rear-End Accidents at Signalized Intersections *ICTE 2013: Safety, Speediness, Intelligence, Low-Carbon, Innovation* (pp. 196-201).

Zhang, G., Zheng, J., & Wang, Y. (2012). Numerical Examinations of Traffic Accident Characteristics Using Analytical Statistical Methods *CICTP 2012: Multimodal Transportation Systems—Convenient, Safe, Cost-Effective, Efficient* (pp. 3546-3557).

Zhang, H., & Khattak, A. (2010). What is the role of multiple secondary incidents in traffic operations? *Journal of transportation engineering, 136*(11), 986-997.

Zhang, Y., Zhu, S., Wang, H., Hu, Y., & Liu, B. (2011). Variance Analysis of Factors That Affected Traffic Safety in Highway Work Zones *ICTIS 2011: Multimodal Approach to Sustained Transportation System Development: Information, Technology, Implementation* (pp. 567-574).

Zhao, G., Jiang, Y., & Li, S. (2016). Safety Effects of Intersection Lightings. *Bridging the East and West*, 115.

Zhou, H., Zhao, J., Hsu, P., & Huang, J. (2013). Safety Effects of Median Treatments Using Longitudinal Channelizers: Empirical Bayesian Before-and-After Study. *Journal of transportation engineering, 139*(12), 1149-1155. doi:10.1061/(asce)te.1943-5436.0000585

Zirkel, B., Park, S., McFadden, J., Angelastro, M., & McCarthy, L. (2012). Analysis of sight distance, crash rate, and operating speed relationships for low-volume single-lane roundabouts in the United States. *Journal of transportation engineering, 139*(6), 565-573.