EXAMINING CRASH LOCATION CHARACTERISTICS IN TEXAS BETWEEN 2003 AND 2017 TO ASSESS THE EFFECTS OF THE GREAT RECESSION ON FATALITIES

A Thesis

by

ANKIT JHAMB

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Dominique Lord
Committee Members,	Satish Bukkapatnam
	Francisco Olivera
Head of Department,	Robin Autenrieth

August 2020

Major Subject: Civil Engineering

ABSTRACT

Various studies have developed models to predict the number of crash fatalities based on the statistically most significant parameters in police-reported crash data during the Great Recession period (December 2007- June 2009 as per the National Bureau of Economic Research). However, no proper research has been conducted to study the spatial patterns for the fatal crash data during the same period. This study serves as an extension of the study conducted by Project 17-67 funded by NCHRP and aims to understand how the economic downturn affected the spatial patterns of crash fatalities between 2003 and 2017 in the state of Texas. The study divides the fatal crashes dataset into three time periods 2003-2007 (pre-recession era), 2008-2012 (recession era), and 2013-2017 (post-recession era). The study uses the optimized hotspot analysis tool, which first finds the critical distance (fixed distance band) at which spatial correlation between the crash locations is most significant for each dataset. It then finds the hotspots for all the datasets based on the critical distance. The study conducted a hotspot analysis based on two approaches: death counts per crash and death counts/AADT per crash. The use of AADT was considered necessary to remove the factor of traffic flow, which is one of the reasons for the higher number and severity of fatal crashes. The results obtained from the second approach are much more consistent in terms of the number of statistically significant points and their locations across the three time periods than the first approach and is, therefore, used for further analysis. The analysis showed that the number of hotspots and coldspots roughly doubled during the recession period compared to the pre-recession period, despite a reduction in the number of fatalities. Similarly, although the number of fatalities in the post-recession era increased, the number of hotspots and coldspots remained similar in numbers during the same period.

DEDICATION

I would like to dedicate my thesis to my family, my master, and numerous friends who have supported me through thick and thin. My parents, Mr. Anil Kumar and Mrs. Anju Jhamb blessed and encouraged me to continue to push harder whenever it seems challenging to complete the work. My brother, Jayant Jhamb, has encouraged me to think beyond my capabilities in order to come up with some good ideas.

I would also like to dedicate this to my master, Mr. Prabuddha Chaitanya Deepak, who has equipped me with the necessary strength to complete my work. Finally, I would like to dedicate this work to my innumerable friends, Vivek Gupta, Sanish Kumar Singh, Yatharth Vaishnani, Sachin Kesarwani, Sruthi Ashraf and so many others who have stood tirelessly behind me and helped me to complete this vast work.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Dominique Lord, who came up with this research idea, provided data and guided me a lot throughout the research. I would also like to thank Dr. Satish Bukkapatnam and Dr. Fransisco Olivera for serving as the members of my thesis committee and providing me inputs to improve my research work.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

I would like to thank my professor, Dr. Dominique Lord of the Zachry Department of Civil & Environmental Engineering, who has initiated the idea of this thesis and provided me with the data. He guided me throughout and helped me analyze the data and complete this work. I would also like to thank Dr. Fransisco Olivera of the Department of Civil & Environmental Engineering and Dr. Satish Bukkapatnam of the Department of Industrial & Systems Engineering for assessing my work and bringing new ideas to make this work more exciting and relevant.

Data were partially provided by Dr. Dominique Lord for conducting data and hotspot analyses. All the work conducted for the thesis was completed by the student independently.

Funding Sources

I would like to thank the Zachry Department of Civil & Environmental Engineering for granting me a fellowship for pursuing my Masters of Science – thesis program for the first year of my study at the university. I would also like to thank Dr. Dominique Lord for providing me with additional funding for a semester via the A.P. and Florence Wiley Faculty Fellow.

NOMENCLATURE

AADT	Annual Average Daily Traffic
AASHTO	American Association of State Highway Transportation Officials
ANN	Average Nearest Neighbor
BAC	Blood Alcohol Concentration
CRIS	Crash Records Information System
CRSS	Crash Report Sampling System
CSR	Complete Spatial Randomness
D&B	Design and Build
DUI	Driving under the influence
DWI	Driving while intoxicated
EDA	Exploratory Data Analysis
Esri	Environmental Systems Research Institute
FARS	Fatality Analysis Reporting System
FDR	False Discovery Rate
FHWA	Federal Highway Administration
GDP	Gross Domestic Product
GIS	Geographic Information System
Gi*	Getis-Ord Gi*
GRID	Geospatial Roadway Inventory Database
HSM	Highway Safety Manual
KDE	Kernel Density Estimation

MCS	Model considering states
MMUCC	Model Minimum Unifrom Crash Criteria
MNCS	Model not considering states
MNN	Median Nearest Neighbor
NASS GES	National Automotive Sampling System General Estimates System
NBER	National Bureau of Economic Research
NCHRP	National Cooperative Highway Research Program
NHTSA	National Highway Traffic Safety Administration
O&M	Operates and Maintains
OECD	Organisation for Economic Co-operation and Development
PARs	Police accidents reports
PDO	Property Damage Only
RI	Risk Index
SD	Standard Distance
TPP	Transportation Planning and Programming
TxDOT	Texas Department of Transportation
US	United States
VMT	Vehicle Miles Traveled
WGS	World Geodetic System
WTO	World Trade Organization

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

The number of crash fatalities on United States highways varied within 3.5% from 1994 to 2007. However, a sudden dip of over 9% each year over the next two years surprised many researchers. A similar trend was observed for the state of Texas, which roughly accounts for one-tenth of the total fatalities in the US, as shown in Figure 1.1 below (Texas Department of Transportation, 2018). The dip in the trend during the period coincides with the period of the economic downturn in the US. This trend led various researchers to suspect if the economic conditions of a country could have an impact on road safety.

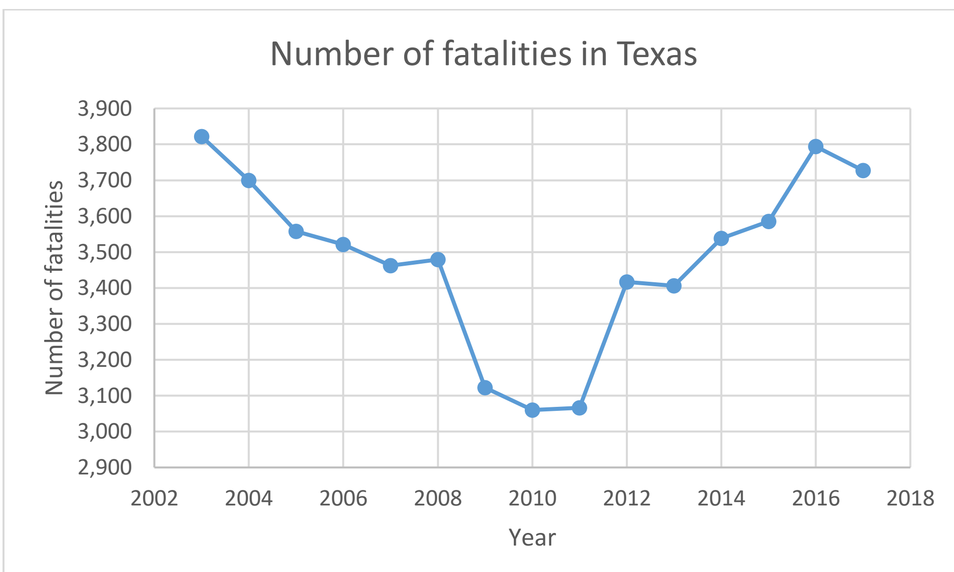


Figure 1.1: Number of fatalities in Texas from 2003-2017

As can be observed from Figure 1.2, the overall trend in percentage change in the number of traffic fatalities compared to their previous years remained similar for both the US and Texas. The fatalities data for United States was obtained from the report Traffic Safety Facts 2017 published by National Highway Traffic Safety Administration (NHTSA) in 2019. The percentage change first reduced during the years of the Great Recession, followed by a subsequent rise in the number of traffic fatalities over the years. However, there were small differences in the trends between both the US and Texas. The percentage change remained more or less similar for both Texas and the US till 2007. However, for the year 2008, the percentage changes increased a little for the state of Texas, while it reduced drastically for the US and continued till 2009, indicating other states got hit by the Great Recession in 2008. Texas got the impact in 2009, as can be observed from the sudden reduction in its percentage as compared to 2008. There was a sudden spike in both the trends in 2012, higher for the state of Texas as compared to the US, indicating the cessation of the impact of the Great recession after 2011. The trend remained positive overall for both the state of Texas and the US after 2012.

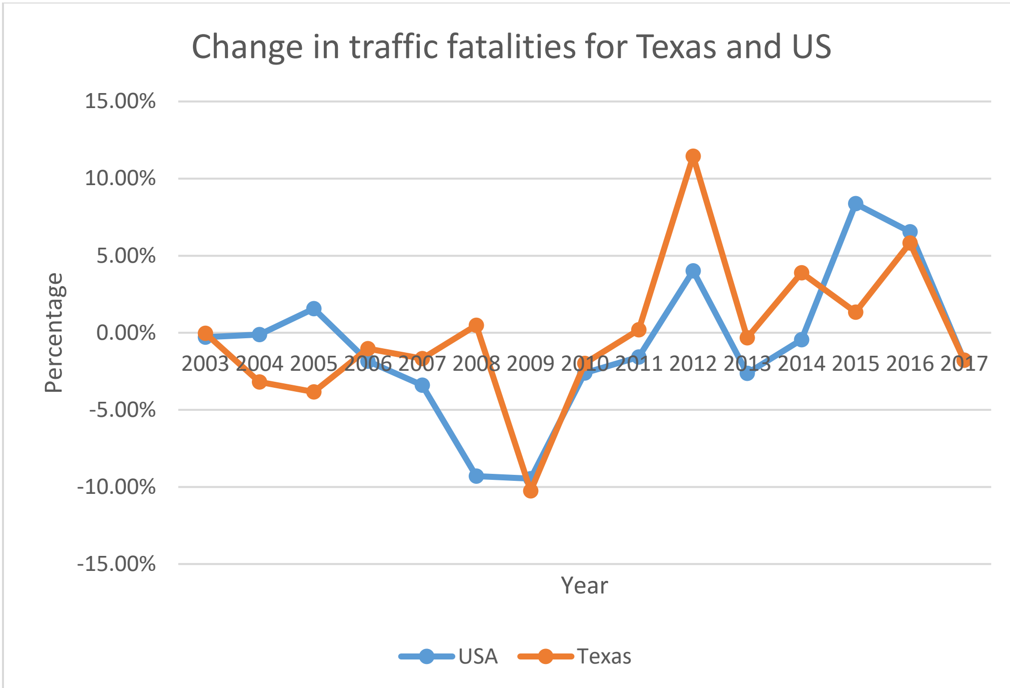


Figure 1.2: Change in the percentage of traffic fatalities compared to their previous years for the state of Texas and the US

Various studies have investigated the primary reasons behind the distribution of crashes during an economic recession. Some argued that the behavior was due to the fewer number of miles covered by the vehicles during lower-income periods (Wijnen and Rietveld, 2015), while others investigated other parameters such as the number of high-risk drivers on the road (Maheshri and Winston, 2016). There was also extensive research conducted in Project 17-67 funded by National Cooperative Highway Research Program (NCHRP), where various models were generated to predict the fatalities based on the dataset from 2001-2012 and Shimu (2019) in her study tested and recalibrated the models based on the additional dataset from 2013-2016. She further modified the model to predict the number of urban and rural fatalities based on the Vehicle Miles Traveled (VMT) in both the urban and rural areas. However, none tried to examine the space-time patterns to gain insights and further deepen our understanding of the impacts of the economic downturn on the crashes. This study aims to carry out the spatio-temporal analysis of the fatal crashes data between 2003 and 2017 and draw inferences.

The study used police-reported crash data between 2003 and 2017 obtained from the Texas Department of Transportation (TxDOT) Crash Query tool. The large dataset is trimmed down to include only the relevant columns required for analysis. The ArcGIS software provides various tools to analyze the data spatially and detect patterns and, therefore, is used in this study. The primary objective is to understand the impact of the economic changes of a region on spatial patterns of fatal crashes. The study further analyzes the effects of changes in spatial patterns of rural and urban fatalities separately during the same period.

I.1 Problem Statement

Based on the literature review, no research was found to visually and statistically analyze the change in patterns of fatal crashes spatially during 2003-2017. This study focus on performing spatial autocorrelation and hotspot analysis to examine the potential changes in the spatial distribution of fatal crashes data occurred before, during, and after the Great Recession. Also, it covers the identification of hotspots based on the death counts per crash locations and death counts divided by the traffic measured by Annual Average Daily Traffic (AADT) and then compares the results between both the approaches. Finally, the study generates and compares the hotspots for total fatalities, rural fatalities, and urban fatalities for the three time periods described.

I.2 Research Objectives

Given that no research was conducted to spatially analyze data and detect the patterns in the hotspots for the period that includes the economic recession to obtain the change in hotspot patterns, the study has the following objectives:

1. Examine how the distribution of fatalities changed before, during, and after the Great Recession period.
2. Determine how the effects changed between urban and rural areas
3. Determine if the spatial correlation changed over time between 2003 and 2017.
4. Determine if crash hotspots changed over time during pre and post the Great Recession period based on the number of fatalities per crash and the number of casualties per crash divided by AADT.

I.3 Study Outline

The outline of the study is as follows:

- Chapter II gives the details of the previous studies conducted by various researchers. It also provides a brief detailing of spatial analysis tools used for analysis.
- Chapter III provides the details of the source of different datasets used for the analyses.
- Chapter IV explores the data to find various trends for the US and the state of Texas in the exploratory data analysis chapter.
- Chapter V explains the methodology adopted for conducting the spatial data analysis.
- Chapter VI documents the results of the analyses and contains further discussions.
- Chapter VII provides a summary and conclusions of the study.

The next chapter presents a detailed literature review on the previous studies conducted to establish the relationship between the number of traffic fatalities and economic activities.

CHAPTER II

LITERATURE REVIEW

This chapter provides details about the previous studies to understand the impact of economic recessions on the number of traffic fatalities across countries, including the US. It also describes the tools that are used in this study to conduct the spatial analyses of traffic fatalities data. Section II.1 describes the effect of the Great Recession across the world, especially on the Organisation for Economic Co-operation and Development (OECD) countries. Section II.2 covers the factors associated with the reduction of the number of traffic fatalities during the period of the Great Recession in the US. Section II.3 gives a brief description of the study conducted by Shimu (2019). Section II.4 explains the methodology and reasons behind the use of spatial analysis tools to conduct the analysis. Section II.5 summarizes the chapter at the end.

II.1 Impact of the Great Recession on the World

The Great Recession of 2008 had dire consequences on the economies of various nations. It was initiated by the collapse of the housing market in the United States and led to the global economic slowdown in 2009 (Chappelow, 2020). As per the World Bank 2009 report, the world economy shrinked by 2.1% in the year 2009, which was a very significant fall in years after the war was over. The World Trade Organization (WTO) estimated the reduction of trade in terms of goods and services to be 12.2% in the year 2009.

The OECD is a group of 36 countries, including the United States, where the member countries review and discuss the policies adopted by their respective countries and look for scope for improvement. The group published a report "From Crisis to Recovery" authored by Keeley and Love (2010), discussing the causes and the consequences of the Great Recession across the OECD area. The report estimated a reduction of a total of 4.7% in the economies of OECD member countries between the 2008 first quarter and 2009 second quarter, as seen in Figure 2.1. The report also provided an estimation of roughly 17 million people who lost their jobs during the Great recession period. It brought down the employment rate for the young population (15-24 years-olds) by 8%.

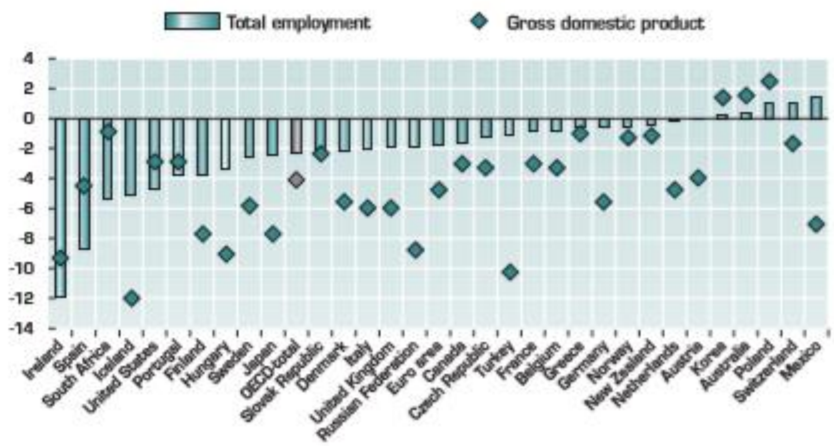


Figure 2.1: The chart showing Gross Domestic Product (GDP) decline and increase in unemployment between 2008 first quarter and 2009 third quarter (Reprinted from Keeley B., Love P., 2010)

Various studies were conducted to study the relationship between changes in the number of traffic fatalities and changes in the economic activity of a region. Wegman et al. (2018) reviewed six independent studies that were completed in the year 2014 on the relationship and drew various conclusions. The first two papers authored by Wijnen and Reitveld (2015) and Elvik (2015) were both review papers. However, the latter additionally studied and analyzed the data from 14 OECD member countries. Antoniou et al. (2015) and Bergel-Hayat et al. (2015) both worked on the data obtained from the European countries to establish the relationship between economic downturn and the number of traffic fatalities and constituted the third and fourth paper of the study. The last two papers examined the economic and crash data for two countries, namely Great Britain and Sweden, and derived the primary reasons accountable for the trends. (Forsman et al., 2015; Noble et al., 2015). All the studies at the end were compared, and the conclusions were drawn.

Subsection II.1.1 looks into the studies conducted across the world to understand the relation between the number of casualties and economic activities of a country. Subsection II.1.2 then covers the studies undertaken for the European countries over the years and discusses the results. Subsection II.1.3 covers other studies specific to countries such as Sweden and Great Britain, and conclusions are drawn.

II.1.1 Review Papers on Economic Recessions and Highway Safety

The study by Wijnen and Reitveld (2015) examined 49 estimates from 41 previous studies about the relationship between the number of traffic fatalities and economic recessions. Out of 41, 26 studies were from the United States, four from Victoria in Australia, two each from Sweden and Belgium, one each from Canada, Switzerland, Norway, New Zealand, Germany, Spain, and China. The time periods analyzed varied substantially between 1930 and 2008. The study used changes in GDP and unemployment rates as parameters to assess the economy of a region. The results showed 44 out of 49 estimates to provide a statistically significant relationship between the economic variables and the number of traffic fatalities. Out of 44, 34 estimates exhibited a positive relationship, and 10 estimates produced a negative relationship. There were 19 estimates identified to find the relationship between economic variables and the number of crashes per kilometer traveled (crash rate). 11 of them produced a positive relationship, seven of them exhibited a negative relationship, and there was no relation found in one estimate.

As per Figure 2.2, economic disturbances can have an impact on the number of casualties in four different ways. Firstly, the economic recession results in lesser growth or even decline in the vehicle kilometers traveled for the given period compared to previous years. Since traffic volume directly affects the crash rate (Hauer, 1995), the vehicle kilometers traveled has a direct influence on both the crash rate and the number of fatalities. Secondly, as the young drivers are high-risk drivers and are most affected during the period of economic recession, the reduction in the proportion of young drivers on the road lowers down the crash rate. Road user behavior also changes, for example, drinking and driving cases reduce during economic downturn owing to less money left in the hands of employees. This phenomenon, in turn, improves road safety. Safety investments also have an impact on the crash rate. Older cars are expected to run on highways more frequently due to the reduction in the sales of new cars and can affect the crash rate (Wijnen and Rietweld, 2015).

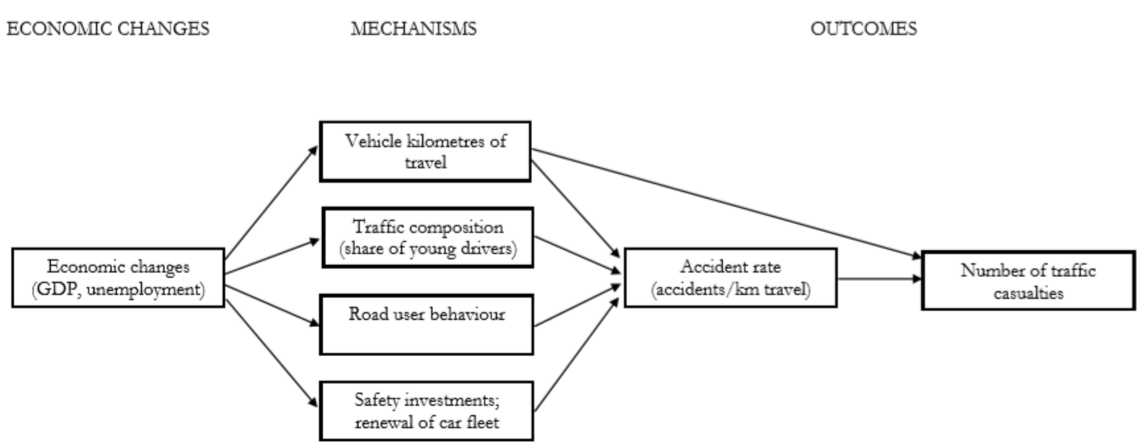


Figure 2.2: Causal diagram of how economic recessions may influence road safety (Reprinted from Wijnen and Reitveld, 2015)

Elvik (2015) was not able to use all the 41 studies for performing a meta-analysis because of their heterogeneity and, therefore, selected a subset of studies for analysis. He used 19 studies: seven from the United States, three from Australia, two from Belgium, one each from Canada, Switzerland, Norway, New Zealand, Germany, Spain, and Sweden. He estimated that a one percent increase in unemployment resulted in between 0.024 and 0.060 percent decrease in traffic fatalities. He then further examined the statistical data from 14 OECD member countries and evaluated the mean unemployment rate to increase by 40 percent between 2008 and 2010, which led to a reduction of between 1 and 2.4 percent in traffic fatalities.

II.1.2 Relationship between the Number of Traffic Fatalities and Economic Activities

Antoniou et al. (2015) used the 37 years (1975-2011) data of 30 European countries to assess the relationship between GDP per capita (used as the indicator of economic activities) and the number of traffic fatalities. They developed the following model for analyzing the trend of traffic casualties for short term changes in GDP per capita (yearly) in countries i. The model stated that the annual change in the number of traffic fatalities was directly proportional to the change in its GDP for the same year.

$$\ln(Fat_{it}) - \ln\left(Fat_{i(t-1)}\right) = \propto + \beta_1\left[\ln(GDP_{it}) - \ln\left(GDP_{i(t-1)}\right)\right] + \beta_2 Country\ Group_{it} + \varepsilon_{it} \quad (1)$$

There were three country groups defined for this analysis: Northern Europe, Eastern Europe, and Southern Europe. The model estimated that the rise of one percent in GDP per capita led to an increment of between 0.42 and 0.66 percent in the number of traffic casualties. Similarly, the decline of one percent of GDP per capita resulted in a reduction of between 0.15 and 0.75 percent in the number of traffic casualties. A negative relationship was found between the annual change in traffic casualties and the elasticity of traffic casualties against GDP per capita. Figure

2.3 showed that countries where the number of traffic fatalities declined at a higher rate, tend to be more vulnerable to the fluctuations in the GDP per capita (Antoniou et al., 2015).

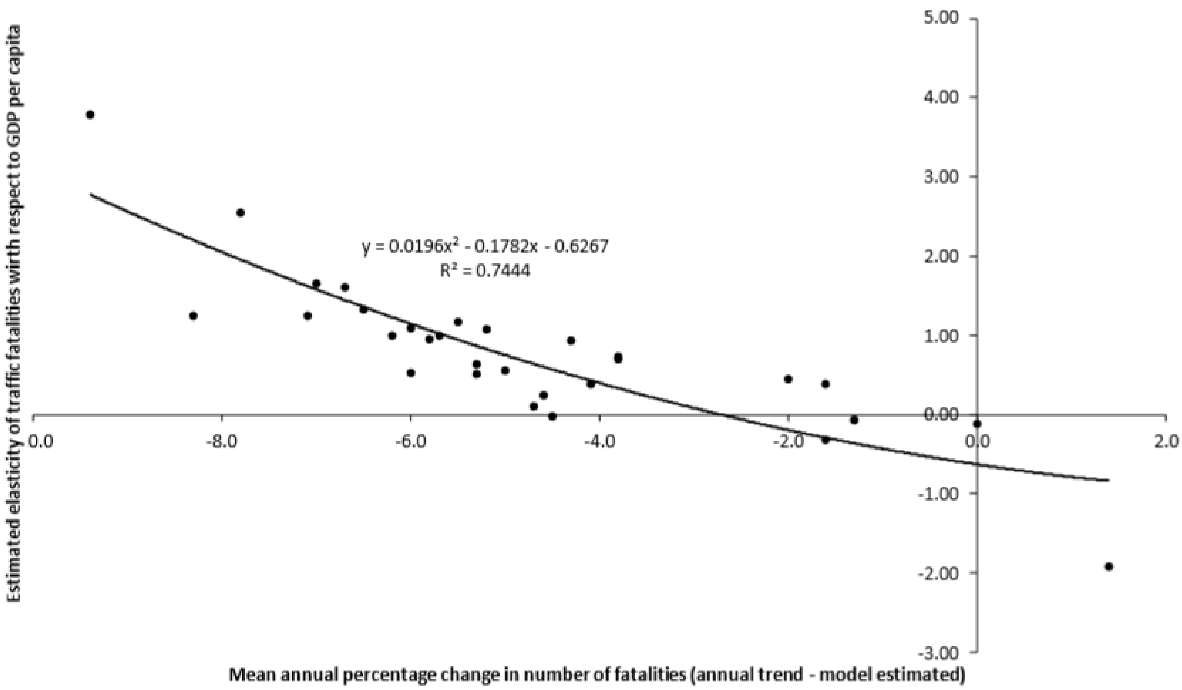


Figure 2II.3: The chart shows the negative relationship between the elasticity of traffic casualties against GDP per capita and the annual change in traffic casualties (Reprinted from Antoniou et al., 2015)

Bergel-Hayat et al. (2015) conducted a state-space time series analysis for three countries, France, Spain, and Greece, over the data for 30 years (1983-2012). They analyzed the monthly unemployment rate and estimated that a one percent rise in unemployment could reduce the number of traffic casualties by three percent, and vice versa. The results are consistent with all the previous studies conducted by different researchers in different countries with distinct methodologies for varying time periods. All the research studies showed that the number of traffic fatalities would decrease during the period of economic recession.

II.1.3 Other Studies

The study headed by Forsman et al. (2015) collected the data of traffic casualties for Sweden between December and March for the years 2006-2009 and compared them. They compared the data of 2006-2008 with 2009 (Great Recession period) and noticed the following observations:

1. The number of traffic casualties dropped more in the time period of 18-23 hours (evening) as compared to other remaining time periods in the day. This could be related to the decline in leisure activities in the evening.
2. The number of crashes involving drinking and driving (or driving under the influence (DUI) or Driving while intoxicated (DWI)) cases reduced more (47 percent) as compared to the crashes not involving such cases (32 percent).

The above observations could explain the reasons for the fewer number of fatal crashes during the Great Recession period.

Noble et al. (2015) estimated the decrease in traffic casualties of 37.2 percent between 2007 and 2010, which is much higher as compared to the decline of 3.5 percent for vehicle kilometers of travel in Great Britain. Similarly, they recorded a more significant drop in fatal crashes involving young drivers (17-24 years old) as compared to others. Also, fatal crashes involving drunk driving dropped more than crashes not involving drunk driving.

After analyzing all the results, Wegman et al. (2017) concluded that all the studies conducted before and during the Great Recession period consistently indicated a direct relationship between economic recession and the reduction in the number of traffic fatalities for all the countries. The main reasons for this were: less driving by young drivers, less speeding, and fewer drinking and driving cases.

II.2 Risk Factors Associated with the Great Recession

As defined by the National Bureau of Economic Research (NBER), the Great Recession is the period that started in December 2007 and ended in June 2009. It resulted in a reduction in the percentages of the VMT and the number of fatalities during the period. Table 2.1 shows that the decline was not proportional (NHTSA, 2019). There was a slight reduction in the VMT as compared to over nine percent reduction in fatalities in both years. This trend inspired many researchers to come up with reasonable explanations behind this behavior.

Table 2.1: VMT and fatalities data (Adapted from NHTSA, 2019)

Year	VMT (millions)	VMT change%	Fatalities	Fatalities change %
2007	3,031,124		41,259	
2008	2,976,528	-1.78%	37,423	-9.3%
2009	2,956,764	-0.67%	33,883	-9.46%

He (2016) chose the unemployment rate to be an indicator of the economic recession and estimated that a percent rise in the unemployment rate could result in a roughly three percent decline in the fatalities caused by motor vehicles. Another research conducted by Cotti and Tefft (2011) concluded that the reduction in fatal crashes occurred due to a decrease in drunk driving during the Great Recession period. Vikram and Clifford (2016) observed that the VMT of high-risk drivers, such as older drivers (age above 60) decreased, while the VMT of safer drivers increased during the Great recession, which could convincingly explain the sizeable reduction in the number of fatalities irrespective of little change in the VMT.

II.3 Modeling Crash Risk

This study is an extension of the study conducted by Shimu (2019), which developed four models to validate the models documented in Project 17-67 funded by the National Cooperative Highway Research Program (NCHRP) for the years 2001 to 2012. The models were recalibrated based on the additional dataset considered during 2013-2016 to predict the fatalities more accurately in the post-recession era. Three of the models followed Poisson-gamma regression models, one of them, MNCS (Model not considering states), considered only the time trend and not the spatial variations. In contrast, the other two models, MCS (Model considering states), incorporated a state-specific parameter to capture any state-specific effects on variables in the model as well as the time trend over the years. The two MCS models took two different parameters as exposure: VMT and population. Finally, the fourth model was a log-change regression model that considered the yearly change in the number of fatalities and did not take into account any state-specific parameter and base year. The models are as follows:

MNCS model $$\mu = VMT \times e^{\beta_0 + \sum_i (\beta_i x_i)} \tag{2}$$

Considering γ_s as the state-specific parameter and taking VMT and population as exposure,

MCS model with VMT $$\mu = VMT \times e^{(\beta_0 + \gamma_s) + \sum_i (\beta_i x_i)} \tag{3}$$

MCS model with population $$\mu = Population \times e^{(\beta_0 + \gamma_s) + \sum_i (\beta_i x_i)} \tag{4}$$

Log-change model considering yearly change

$$\ln(Fat_t) - \ln(Fat_{t-1}) = \beta_0 + \sum_{j=1}^{k} \beta_j z_t = \beta_0 + \sum_{j=1}^{k} \beta_j (\ln(x_t) - \ln(x_{t-1})) \tag{5}$$

where, $z_t = \frac{x_t}{x_{t-1}}$

μ = the estimated mean of the dependent variable

x_t = independent variables for year t

z_t = transformed change variables year to year

β = parameter coefficients

The authors found a total of seven parameters to be statistically significant in both the MCS with VMT as exposure and MNCS models with the unemployment rate among the youngsters aged 16 to 24 years, median household income, and gasoline price as the most statistically significant (at the 10% level). The MCS model having population as exposure and log-change models did not show good results and, therefore, were discarded. Further, the selected two models were used to predict fatalities based on land use type, that is, urban and rural fatalities. Keeping other variables the same as in total fatalities model, VMT was divided into urban and rural to predict the number of rural and urban fatalities, respectively. In this case, the MCS model with VMT as exposure provided better predictions as compared to the MNCS model.

II.4 Spatial Analysis Tools

Various studies have used different tools and approaches to conduct spatial analysis of the crash data. As per Lord and Mannering (2010), if one takes observations for a particular stretch of road at specified intervals, one can find the temporal correlations between the observations because many parameters governing these observations will remain the same for the period. A good example can be the restrictions in local sight-distance that may result in several crashes at a particular or surrounding area for a longer time (Mannering and Bhat, 2014). Also, Narayanamoorthy et al. (2013) and Bhat et al. (2014) attempted to combine spatial and temporal dependencies in the crash data to obtain multivariate models.

Li and Liang (2018) used the hotspot analysis on aviation crashes in the state of Florida to obtain airports with high-risk index (RI) in Miami and Tampa Bay. Shafabakhsh et al. (2017) employed nearest neighbor distance and K-function to identify clustering patterns and then used Kernel Density Estimation (KDE) tool to obtain crash hotspots, which then helped the traffic department to implement its resources better. Similarly, the hotspot study conducted on the teenage crashes in the city of Houston in Texas between 2006 and 2009 spotted the hotspots very close to the popular locations for teenagers and suggested strict enforcement and improved maintenance measures could result in the decline of the crashes involving teenagers (Goodwin et al., 2014).

Some researchers also worked on the spatiotemporal characteristics of traffic crashes. Kang et al. (2018) worked on traffic crashes involving older people as both the drivers and victims in Seoul and performed hourly as well as monthly analysis to detect the patterns in hotspots. Bil et al. (2019) investigated the crash hotspots on the Czech road network for the period 2010-2018 and divided them into three types: the emergence of hotspot, stability, and disappearance. The authors identified one hundred hotspots which remained stable over the study period. A study by Zhang and Shi (2019) used the wavelet decomposition method on traffic crash data and further used k-means clustering to understand spatial differentiation patterns.

The above literature review helped to identify tools to be used to generate spatial patterns and hotspots over crash data. The tutorials provided by the ArcGIS website also helped to understand and operate tools and how to interpret results. This study will use the two most essential tools for analysis: Spatial Autocorrelation and Hotspot Analysis tools. Subsection II.4.1 cover in detail about the Spatial Autocorrelation tool and its output. Subsection II.4.2 provides a detailed description of the Hotspot Analysis tool and the interpretation of the results.

II.4.1 Spatial Autocorrelation (Global Moran's I)

"How Spatial Autocorrelation (Global Moran's I) works," (n.d.) describes in detail about the spatial autocorrelation tool. The tool helps to find whether or not the location of points shows any clustered, dispersed, or random patterns. The Moran's I index, p-value, and z-value indicate the type of patterns formed based on the input field (death counts per crash used here) and the distance between the locations of features. Keeping the input field constant, the distance with the most significant p-value forms the most significant clusters. The Moran's I index is given by:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2} \tag{6}$$

Here, $z_i = x_i - \hat{X}$, $w_{i,j}$ denotes the spatial weight between i and j, n denotes the total features in the data, and $S_0 = \sum_{i-1}^{n} \sum_{j=1}^{n} w_{i,j}$

The tools calculate Moran's I index and the Expected Index value and then compares their values. The z-score and p-values tell whether the difference is statistically significant or not. The z_I–score is given by:

$$z_I = \frac{I - E[I]}{\sqrt{V[I]}} \tag{7}$$

where

$$E[I] = -\frac{1}{n-1} \tag{8}$$

$$V[I] = E[I^2] - E[I]^2 \tag{9}$$

The null hypothesis for Moran's I index is that all the points are randomly distributed. It is positive when the high (or low) values come spatially closer to other points of high (or low) values.

It is negative when the high values come spatially closer to points having low values. Table 2.2 shows the different outcomes of results and their interpretations:

Table 2.2: Result interpretation of spatial autocorrelation tool (Adapted from "How Spatial Autocorrelation (Global Moran's I) works," n.d.)

Sr. No.	p-value	z-value	Interpretation
1.	Not significant		Cannot reject the null hypothesis. The points are in complete spatial randomness (CSR)
2.	Significant	Positive	Reject the null hypothesis. High (or low) values are more spatially clustered.
3.	Significant	Negative	Reject the null hypothesis. High (or low) values are more spatially dispersed (that is, high (or low) values repel another high (or low) values).

II.4.2 Hotspot Analysis (Getis-Ord-Gi)*

Similarly, "How Hot Spot Analysis (Getis-Ord Gi*) works," (n.d.) describes in detail about the hotspot analysis tool. The hotspot analysis tool identifies the statistically significant hotspots, that is, the points having high (or low) values surrounded by high (or low) values. The tool calculates the local sum at a point and its neighboring points and compares it proportionally to the total sum having all the points. If the difference is significant, it results in higher z-score values, and those points are having z-scores higher than σ become hotspots. The tool computes Getis-Ord Gi* (G_i*) statistic for every point using the formula stated below:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{[n \sum_{j=1}^{n} w_{i,j}^2 - (\sum_{j=1}^{n} w_{i,j})^2}{n-1}}} \tag{10}$$

Here, x_j is the attribute value (death counts per crash) for point j; $w_{i,j}$ is the spatial weight (Euclidean distance) between points i and j; and, n is the number of points. The equation is defined as follows:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{11}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \tag{12}$$

Note that G_i* produces z-scores. z-scores can be both negative or positive. The higher the positive z-score is, the more statistically significant hotspot. The lower the negative z-score, the more statistically significant cold spot. Hotspots are the high-value points surrounded by high-value points, and the cold spots are the low-value points surrounded by low-value points. Gi_Bin shows the points present in the confidence interval based on the z-score of the total sum.

II.5 Chapter Summary

The purpose of the literature review was to explore all the relevant existing studies to understand their methodology adopted and their results obtained. This chapter aided in developing the idea and methodology for spatial statistical analysis used in the subsequent sections. The following conclusions were drawn:

- The results showed a positive relationship between the number of fatal crashes that occurred and the economics of a region. The study across the countries invariably showed that the number of traffic casualties decreased during the time of economic recession.

- It also revealed the leading factors across the regions that were responsible for the decline in the number of traffic fatalities during the same period.
- A brief description of the previous study conducted by NCHRP and Shimu (2019) found the MCS model with VMT as exposure to predict the urban and rural fatalities more accurately as compared to the other three models.
- The last section provided a brief description of various past researches conducting the spatial analysis and the tools to be used in this study.

The next chapter provides the sources of the different types of data used to perform exploratory data analysis and further spatial analysis.

CHAPTER III

DATA COLLECTION

Various sources were used to collect the data for the exploratory data analyses (EDA) for the fatalities and hotspot analysis of fatal crashes in the State of Texas. Section III.1 gives a brief description of the police-reported crash data collected from Crash Query tool provided by Crash Records Information System (CRIS). Section III.2 provides the source of data for the whole of the United States to conduct exploratory data analysis. Section III.3 gives the data source for the state of Texas to conduct spatial analysis. Section III.4 summarizes the vital components of the chapter.

III.1 Police-Reported Crash Data

Crash reports obtained from police forces are often used to develop models to predict the number of fatalities based on the most significant parameters. However, the dataset suffers from certain limitations. The research shows that the probability of reporting the fatal crashes is higher compared to other injury crashes and the least if the crash causes Property Damage Only (PDO) (Yamamoto et al., 2008; Ye and Lord, 2011). The reporting of crashes changes with the reporting agency as well (Hauer and Hakkart, 1988; James, 1991). Despite the limitations, the crash reports are still a popular source of data because of the vast amount of data it reveals, giving details of various parameters at the time of crashes.

Police-reported crash data uses the KABCO scale to define the severity of a crash, where K refers to a fatal crash, A denotes an incapacitating injury, B refers to a non-incapacity injury, C indicates a possible injury and O or Property Damage Only (PDO) means that no injury occurred during the crash (see Table 3.1 below). The Model Minimum Uniform Crash Criteria (MMUCC) fifth edition published by NHTSA in 2017 provides guidelines about the different data fields that the police force needs to report and record for every crash, which includes the KABCO scale as well. Every state describes the KABCO scale in a slightly different manner. Table 3.1 provides how Texas defines the KABCO scale (Federal Highway Administration (FHWA), 2017).

Table 3.1: KABCO scale of Texas (Reprinted from FHWA, 2017)

State	Injury Codes	Conversion	Definitions/Instructions/Notes
Texas	K-Killed	K	Died due to injuries sustained from the crash, within 30 days of the crash.
	A-Incapacitating Injury	A	Severe injury which prevents continuation of normal activities; includes broken or distorted limbs, internal injuries, crushed chest, etc.
	B-Non-Incapacitating Injury	B	Evident injury such as bruises, abrasions, or minor lacerations which do not incapacitate.
	C-Possible Injury	C	Injury, which is claimed, reported or indicated by behavior, but without visible wounds; includes limping or complaint of pain.
	N-Not Injured	O	The person involved in the crash did not sustain an A, B, or C injury.
	99-Unknown (2006 to present)	U	Unable to determine whether injuries exist. Some examples may include: Hit and Run, Fled Scene, FSRA, etc.

The Highway Safety Manual (HSM) (American Association of State Highway Transportation Officials (AASHTO), 2010) defines crash severity as the most severe injury caused to person(s) involved in the crash. Since the crash severity for all the fatal crashes will be killed (K), there is some parameter required to differentiate between the fatal crashes based on the damage the crash incur on society. So, the number of fatalities that occurred in a crash is used to serve this purpose.

In Texas, TxDOT records and maintains the crash data obtained from law enforcement police officers and publishes it from time to time. It also provides access to data via the TxDOT Crash Records Information System (CRIS). The data contain detailed information about various factors, some of them are environmental factors present at the time of the crash, type of vehicles involved, road conditions, crash date and time, the fatal crash locations (latitude and longitude) projected in the World Geodetic System (WGS) 1984 coordinate system. The crash data used in this study consists of crashes with severity level as fatal only.

The dataset obtained for the period 2003 to 2017 has a colossal number of columns and contains several parameters. The number of columns in police-reported data has changed over the years, and various new columns have been introduced from time to time. However, the columns required for the analysis have remained in the data. Table 3.2 shows the meaning of each column required for analysis:

Table 3.2: The dataset fields used for the analysis

Sr. No.	Column Name	Description
1	Crash_ID	A unique crash ID for each crash
2	Crash_Fatal_FI	Y if a person died or N if not
3	Rural_FI	Y if the crash happened in rural areas or N if not
4	City_ID	A unique identification number to denote cities
5	Cnty_ID	A unique identification number to denote counties
6	County Name	The name of the county of the crash location
7	Latitude	The latitude coordinates of the crash location
8	Longitude	The longitude coordinates of the crash location
9	Incap_Injry_Cnt	Number of persons having an incapacitating injury
10	Nonincap_Injry_Cnt	Number of persons having a non-incapacitating injury
11	Poss_Injry_Cnt	Number of persons having a possible injury
12	Unkn_Injry_Cnt	Number of persons having an unknown injury
13	Non_Injry_Cnt	Number of persons who are not injured
14	Tot_Injry_Cnt	Total number of persons getting injured
15	Death_Cnt	Number of persons killed

III.2 Data Source for the US

Various studies brought out many factors that could have affected the decline in the number of traffic fatalities during the time of the Great Recession. In order to confirm if the factors were right, it was decided to analyze each of the factors using the data compiled and published by the

NHTSA in an annual report, "Traffic Safety Facts". This annual report contains the data from three of its primary data sources: Fatality Analysis Reporting System (FARS), National Automotive Sampling System General Estimates System (NASS GES), and Crash Report Sampling System (CRSS). A brief description of all three sources are as follows:

1. FARS: It contains the fatal traffic crashes data from the 50 States, Puerto Rico, and the District of Columbia. The dataset contains greater than 100 data elements. Each year's annual report updates the numbers for the previous year.
2. NASS GES: The data are collected from a police-reported crash that has the most significant concern towards the safety of highways. A random sample of 55,000 police crash reports (PARs) is obtained per year from across the United States to extract roughly 90 data elements.
3. CRSS: CRSS replaced the NASS GES system in 2016. Similarly, the crash data taken from 53 sites across the US selected based on the population, geography, crashes, and miles driven are coded in roughly 120 data elements.

Chapter IV uses the data to identify the factors responsible for the lowering of fatal crashes during the period of the Great Recession.

III.3 Data Source for the State of Texas

This section provides the details of the two data sources which act as the source for all kinds of data required to conduct spatial analysis. Subsection III.3.1 describes the TxDOT Open Data portal, which is the most important source for obtaining Geographic Information System (GIS) files, such as the shapefiles and the database files. Subsection III.3.2 talks about Texas Demographic Center, the place to find the demographic data for the state of Texas.

III.3.1 TxDOT

TxDOT divides Texas into four regions, East, West, North, and South regions and further into 25 districts. Each district plans, D&B (design and build), and O&M (operates and maintains) the state roadways within its boundary. Each region is a conglomeration of some districts, which are, in turn, a conglomeration of the different counties. The details of the regions and the districts corresponding to the regions is shown in Figure 3.1 and Table A.1 of Appendix. As most of the data sources provide the data on the county level, it is crucial to identify the counties that were a part of the 25 districts which make the four districts of the Texas region (Texas Department of Transportation, n.d.). The shapefiles used for spatial analysis in ArcGIS for counties and districts were obtained from the TxDOT Open data portal.

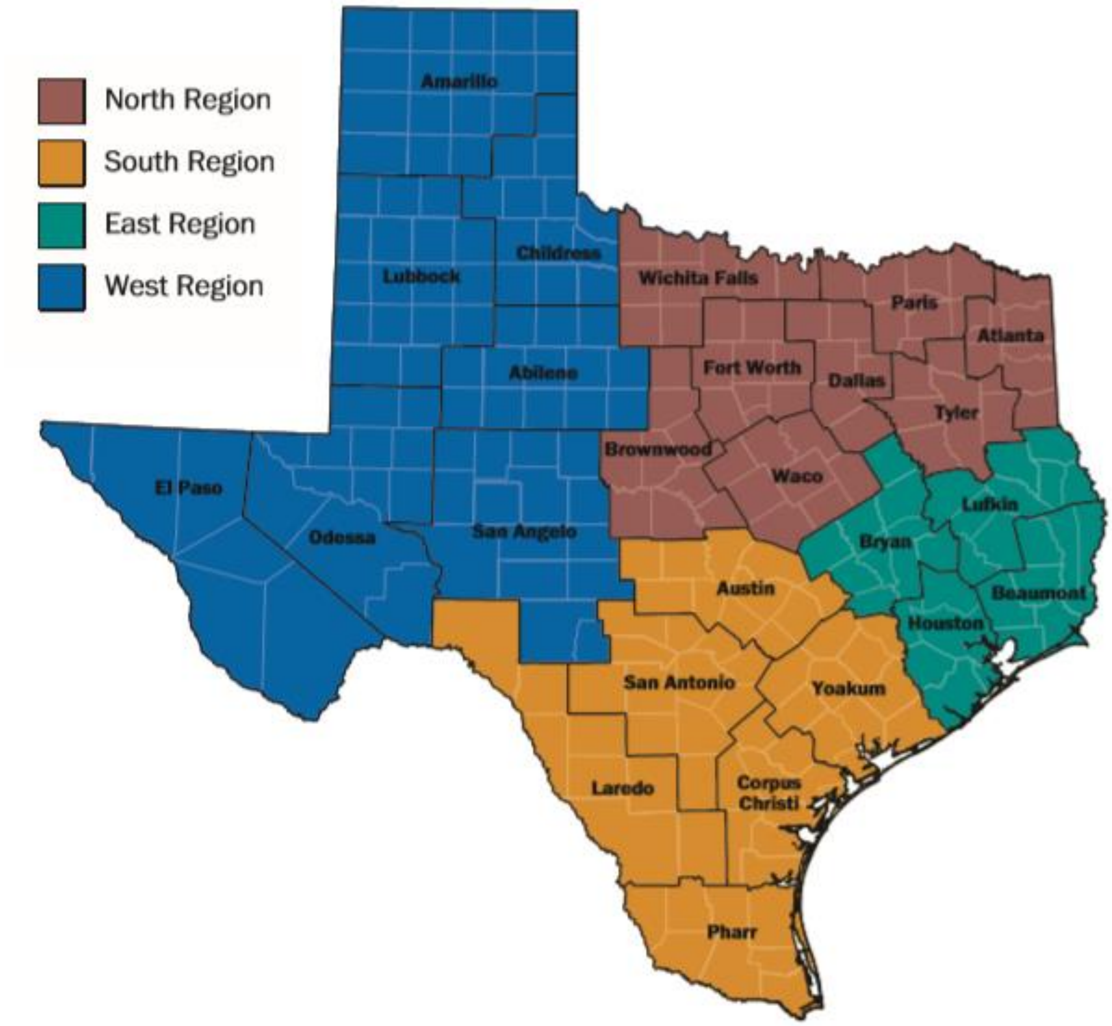


Figure 3.1: Texas divided into four regions and 25 districts (Reprinted from TxDOT, n.d.)

Annual Average Daily Traffic (AADT) data present near each fatal crash location is used to normalize the death counts at each location so that the hotspots take both the death counts and AADT of the crash location into account. The Transportation Planning and Programming (TPP) Division at TxDOT is responsible for collecting the AADT data at various traffic stations. The axle counts are collected at these locations by short-term pneumatic tubes to provide AADT. The data are displayed at the centerline of the roadway and is equal to the summation of traffic counts on both mainlanes and frontage roads. AADT is calculated as follows (TxDOT, 2019):

$$AADT = axles * axle\ factor * seasonal\ factor \tag{13}$$

The roadway data are required to understand the type of roads where the fatal crash locations and their hotspots are present. TPP division of TxDOT again maintains the dataset of the roadways in the form of polylines for planning and visualization purposes. It includes both on-system routes (the routes which TxDOT maintains), such as, US, interstate and state highways, and ranch and farm roads, and off-system routes, that is, county and local roads. Geospatial Roadway Inventory Database (GRID) is the source of the data (TxDOT, 2020).

III.3.2 Texas Demographic Center

The population data for each year between 2003 and 2017 are required to understand how changes in the number of traffic fatalities fared against an increase in population during the three time periods. Texas Demographic Center has Texas Population Estimates Program as well as Texas Population Projections Program to produce estimates and projections respectively of the population for all the counties and the state of Texas as a whole based on sex, age, and race/ethnicity. Various characteristics related to change in population on the county level and change in the age distribution and race/ethnicity over the years can be identified (Lloyd, n.d.).

III.4 Chapter Summary

This chapter has described in detail the sources of the various datasets required for performing exploratory data analysis and further spatial statistical analysis in the subsequent sections. The key points to note were as follows:

- TxDOT provided the dataset of fatal traffic crashes for the state of Texas using CRIS.

- Data to obtain various national trends for the US were collected from the Traffic Safety Facts report of 2017, published by NHTSA.
- Various GIS files required to analyze the traffic fatalities data geographically were obtained from the TxDOT Open Data portal. Counties' and districts' data, roadways data, and AADT on various points on the different types of roads were all obtained from the TxDOT website.
- Texas Demographic center provided the population data used for further examination of the fatal crash dataset.

The next chapter presents the results of the exploratory analyses of the data that were collected in this study.

CHAPTER IV

EXPLORATORY DATA ANALYSES

This chapter describes the exploratory analyses of the data between 2003 and 2017. These exploratory analyses will provide insights for the GIS analyses described in subsequent chapters. For the sake of understanding things better, the dataset is divided into three time periods: pre-recession (2003-2007), during the Great Recession (2008-2012), and post-recession (2013-2017). Also, in order to understand the spatial data better, the state of Texas is further divided into four regions: East, West, North, and South regions, as shown in Figure 3.1. This chapter first explores the national trends of the United States for 15 years, followed by the trends of fatalities for the state of Texas.

Section IV.1 presents the trends followed by significant factors that affected the sudden decrease in the number of traffic fatalities during the period of the Great Recession in the United States. Section IV.2 shows the results of the exploratory data analyses on the fatal crash data for the state of Texas. Section IV.3 provides the summary and conclusions of the chapter.

IV.1 National Trend

This section summarizes the national trend based on already published statistics by NHTSA in Traffic Safety Facts 2017. The section is divided into three subsections. Subsection IV.1.1 covers the age distribution of the people that died in a fatal crash over the years. Subsection IV.1.2 provides details of the type of vehicles that were more involved in fatal crashes between the years 2003 and 2017. Subsection IV.1.3 gives details of the number of traffic casualties where alcohol impairment can be one of the primary causes of the fatal crash.

IV.1.1 Age Distribution

Figure 4.1 shows the fatality rate per 100,000 population for the eight age groups: 16-20 years, 21-24 years, 25-34 years, 35-44 years, 45-54 years, 55-64 years, 65-74 years, and greater than 74 years for the period between 2003 and 2017. As can be seen in Figure 4.1, the fatality rate per 100,000 population decreased at a faster rate for the younger generation between 2007 and 2010 as compared to other age groups. The highest drops were recorded for the age groups 16-20 and 21-24 years old with 28.2% and 29.6% respectively in the fatality rates per 100,000 population between the years 2007 and 2009 as compared to other age groups. The age group of 25-34 years old was a distant third with a 19.2% reduction for the same years. The dropping percentages were lowest for the >74 years old. These observations indicate that the young people of 16-24 years old were driving and traveling less as compared to other age groups. This trend is in accordance with the results obtained by Noble et al. (2015), which showed a higher reduction in fatal crashes involving young people (17-24 years old) in Great Britain. The results could be explained by the fact that the unemployment rate was higher among the young population as compared to any other age group (Keeley and Love, 2010).

After 2010, the curve had flattened for all the age groups. The fatality rates per 100,000 population fluctuated between 12% and 14% for 16-20 years old and 15.5% and 18% for 21-24 years old, which was significantly higher than all other age groups.

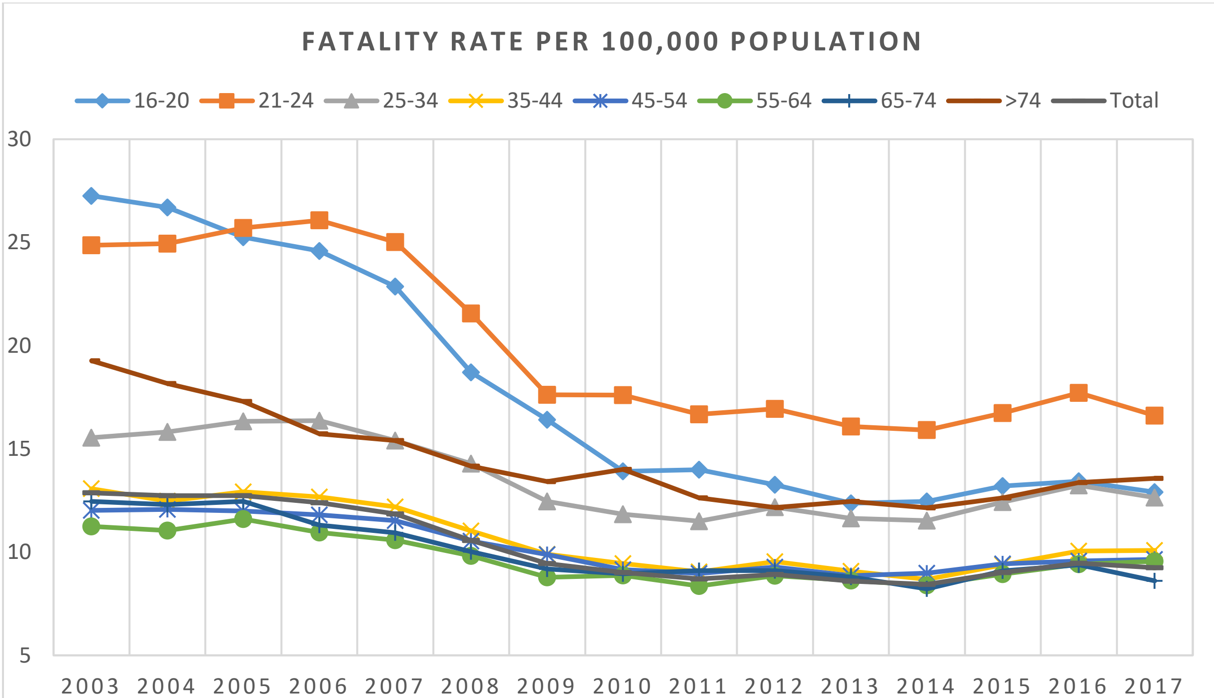


Figure 4.1: Fatality rate per 100,000 population by age group

IV.1.2 Vehicle Type

Figure 4.2 shows the percentage change in the number of vehicles involved in fatal crashes per 100 million VMT that occurred every year between 2003 and 2017 to the number of vehicles involved in fatal crashes per 100 million VMT that occurred in the year 2002 for four different types of vehicles, namely, passenger cars, light trucks, large trucks, and motorcycles. Figure 4.2 depicts that the involvement rate per 100 million VMT increased for both the large trucks and motorcycles till 2006, and then it drastically dropped down to their lowest figures by the year 2009. After 2009, the numbers remained more or less constant for motorcycles involved in fatal crashes, while the number of large trucks per 100 million VMT increased steadily till 2017. The decline in the numbers was most significant in the case of large trucks, and the results were consistent with the findings of Elvik (2015).

The number of passenger cars present in fatal crashes per 100 million VMT declined till 2010, and the most significant decline coincides with the economic recession period. However, the trend reversed after the year 2010, and the numbers rose steadily until 2017. The involvement of light trucks in fatal crashes per 100 million VMT reduced continuously till 2011, the decline being higher in 2008 and 2009 before the curve flattened.

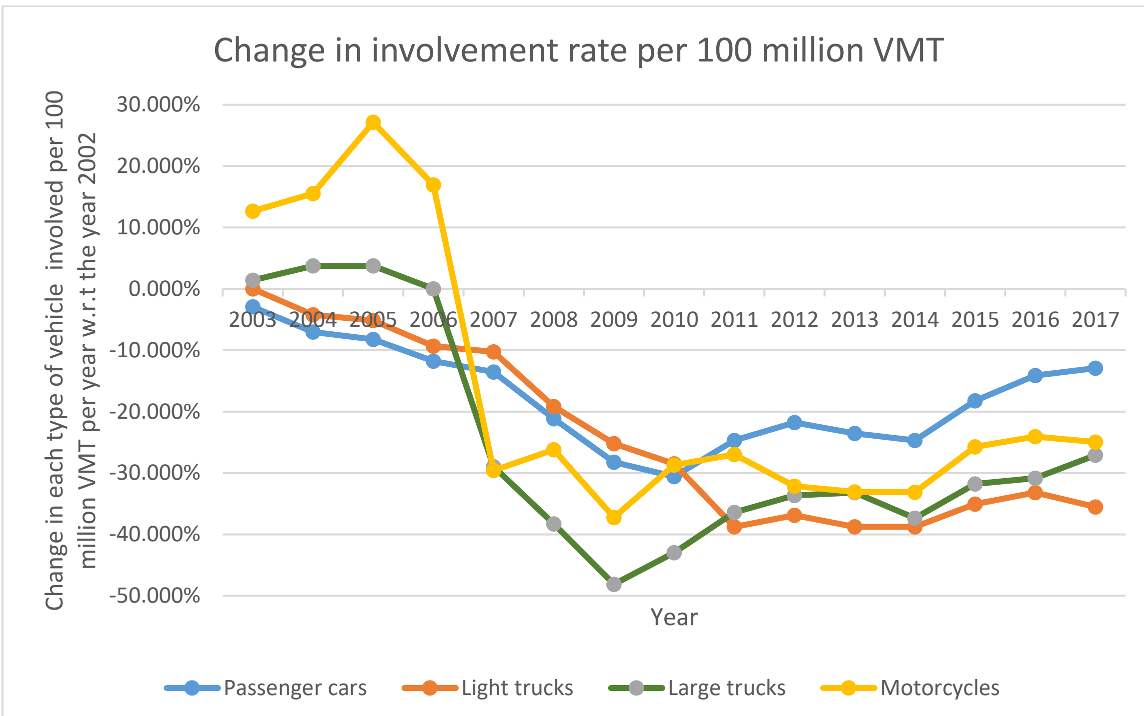


Figure 4.2: Percentage change in the number of fatalities per vehicle type against the number of fatalities in the year 2002

IV.1.3 Alcohol-Impaired Fatalities

Consumption of alcohol before or during driving has been one of the leading causes of crashes resulting in casualties over the years. Many pieces of research pointed to a much lesser number of drinking and driving cases during the period of the Great Recession. For example, Wijnen and Rietveld (2015) concluded that lower incomes during the economic recession impacted the road user behavior, and led to a fewer number of drinking and driving cases. The numbers obtained from Traffic Safety Facts 2017 published by NHTSA were a testimony to this (Figure 4.3).

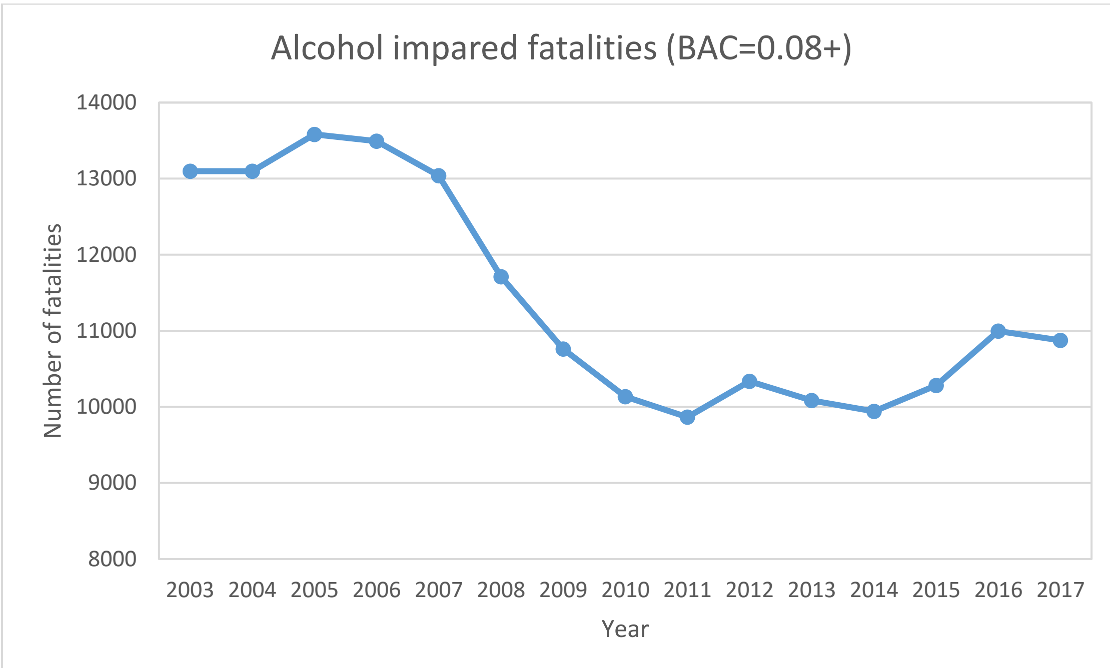


Figure 4.3: Fatalities caused by drinking and driving

Here, if the blood alcohol concentration (BAC) of the driver involved in a fatal crash is found higher than 0.08, the consumption of alcohol is considered to be a factor in the cause of the crash. Figure 4.3 demonstrates a sudden reduction in the number of fatalities between 2007 and 2010, concurring to the period of the Great Recession. The percentage reduction in the number of fatalities was reported as 11% in 2008 and 8% in 2009 compared to their previous years.

IV.2 Trends for the State of Texas

In this section, various trends and patterns were extracted from the fatal crash dataset for the state of Texas obtained from CRIS before analyzing the data spatially and finding out the hotspots. As mentioned in the previous chapter, in order to analyze the spatial trends for the fatalities in the state, Texas is divided into four regions, namely, East, West, North, and South (Figure 3.1). This section analyzes the fatalities in the four regions based on the three five-year periods (pre-recession period, recession period, and the post-recession period) and then compare the data based on the region and time period to conclude.

Subsection IV.2.1 first looks into the death counts for each of the four regions and how they vary over the three time periods. Subsection IV.2.2 then takes into account the fatalities that occurred in the urban and rural areas and provide more profound insights into the data. Subsection IV.2.3 finally provides the number of casualties that occurred in each district and their variation across the three time periods.

IV.2.1 Death Counts in Each Region

Figure 4.4 shows a comparison of the number of deaths that occurred due to fatal crashes in the four regions for the three time periods. As can be seen, the North region registered the highest number of fatalities and the West region the least number of fatalities. Figure 4.4 also indicates that the number of casualties reduced during the recession period in all regions. The most substantial reduction in the number of casualties was recorded in the North region at 11.5% while the lowest in the West region at 4.9% during the recession period. However, the jump of 31.8% in the numbers in the West region was very significant in the post-recession period. The lowest rise was recorded in the East region at 4.6%.

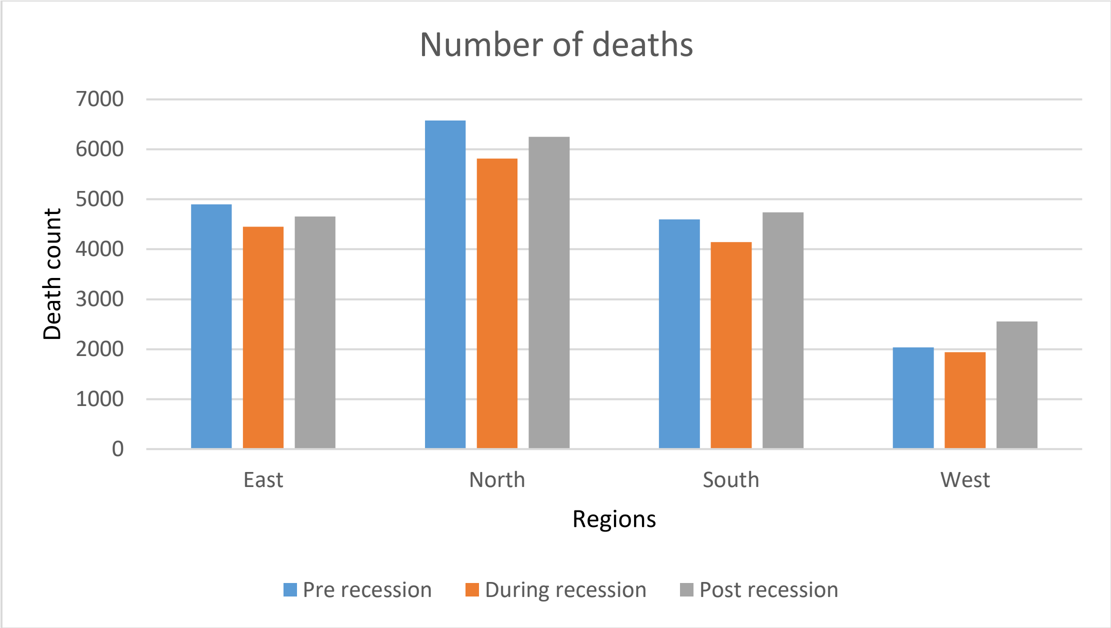


Figure 4.4: Number of deaths in four regions in the state of Texas for three time periods

IV.2.2 Fatalities in the Urban and Rural Areas

The fatal crash dataset has a column Rural_Flag which provides the location of the crash, whether it has occurred in the rural or urban areas. Figure 4.5 illustrates a fewer number of urban fatalities compared to the number of rural fatalities during most of the period between 2003 and 2017. The number of urban and rural fatalities was similar between the years 2003 and 2005. After that, the urban and rural fatalities followed the opposite directions. The number of traffic casualties increased drastically in rural areas, whereas it dropped down suddenly in urban areas in 2006. The year 2009, the period of the Great Recession saw a considerable reduction of 12.2% and 6.9% in the number of urban fatalities and rural fatalities, respectively, compared to the previous year. The number of casualties continued to increase after the Great Recession period until it achieved similar numbers by the year 2017 as in 2003.

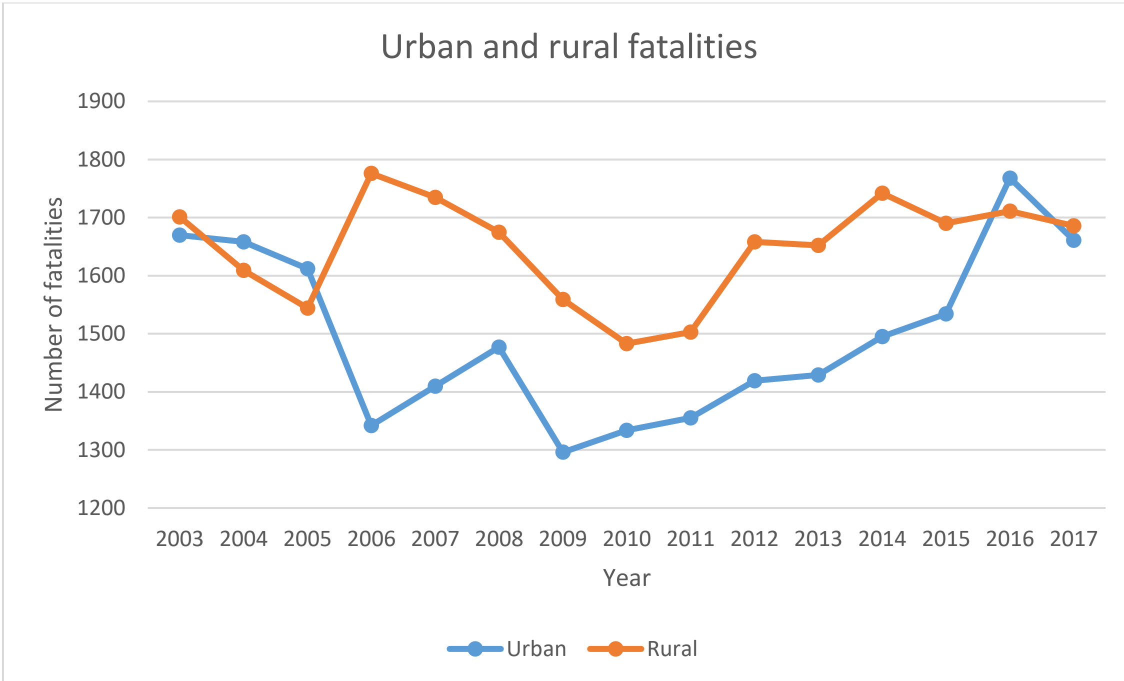


Figure 4.5: Urban and rural fatalities for the state of Texas

IV.2.3 Fatalities in Each District

Figures 4.6 and 4.7 show the number of urban and rural casualties that occurred before, during, and after the period of the Great Recession. Houston district in the East region, Dallas and Fort Worth districts in the North region and San Antonio district in the South region have higher death rates in urban areas as compared to rural areas. All the other districts have a higher number of rural fatalities. Also, an interesting observation to note was that the total number of fatalities was less for the East and North regions and more for the South and West regions in the post-recession period as compared to the pre-recession period.

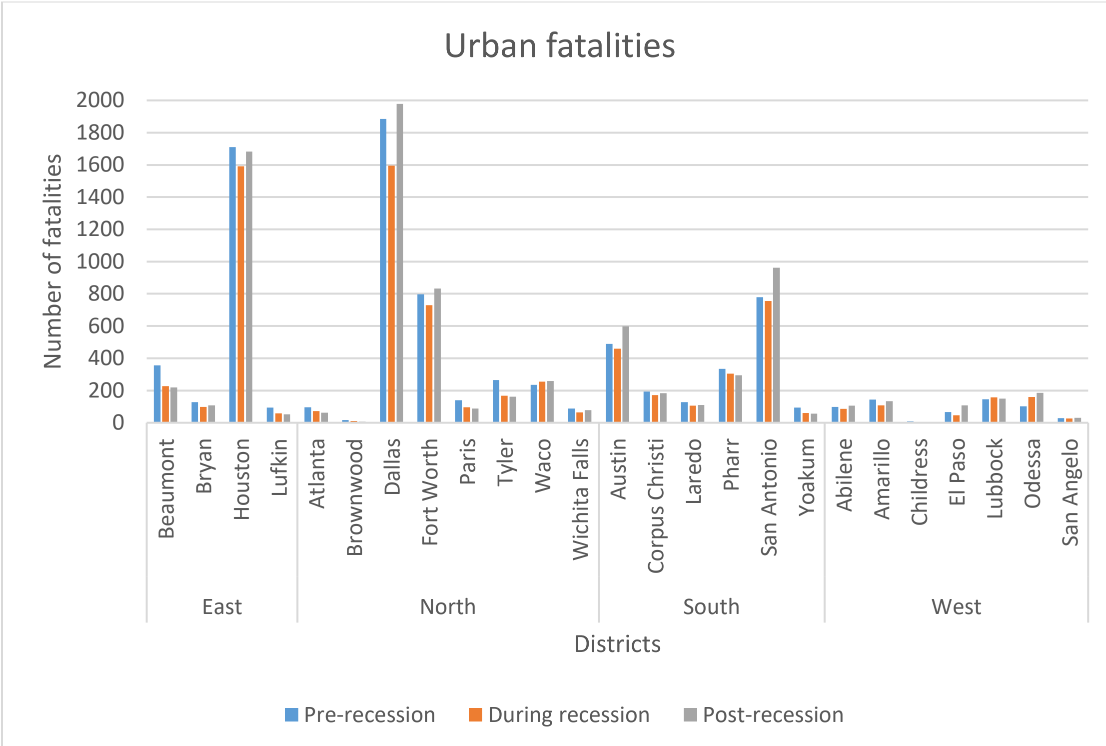


Figure 4.6: Total number of urban fatalities for each time period

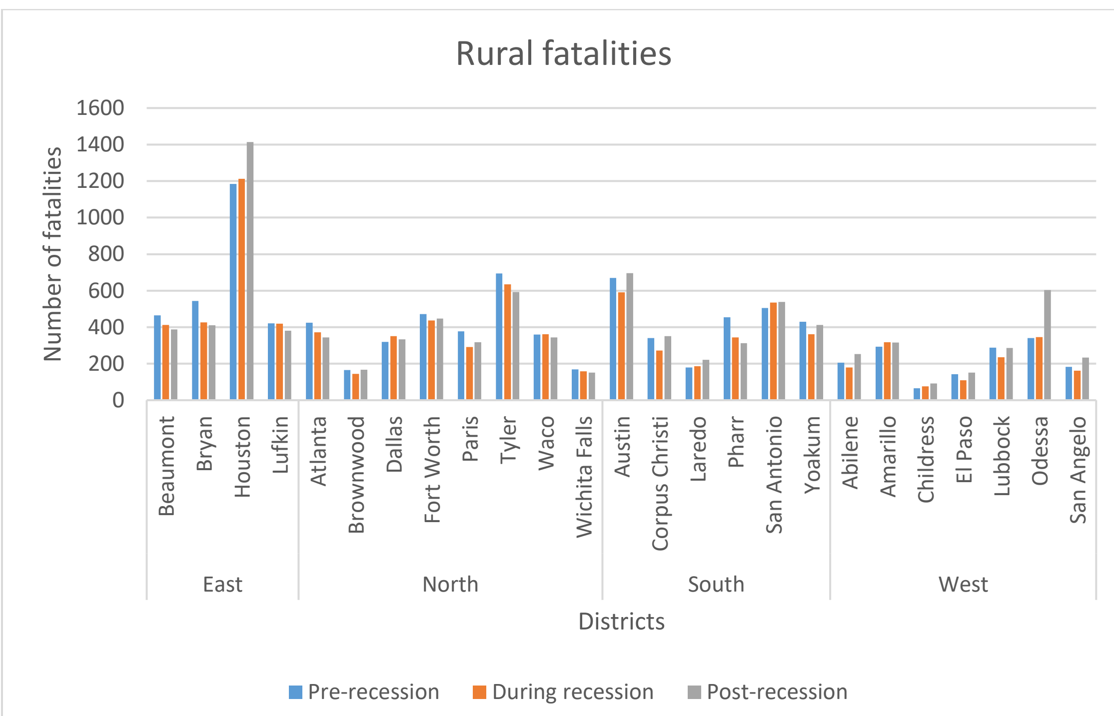


Figure 4.7: Total number of rural fatalities for each time period

The East region registered a significant reduction in the total number of casualties at 9.2% during the Great Recession period and an increment in casualties at 4.6% in the post-recession period. The increase is attributable to the Houston district, accountable for roughly 60% of the total casualties of the region, showed a significant rise in the number of casualties in the post-recession period while other districts registered a negative trend. The region recorded a considerable decline in the number of urban casualties by 13.6% during the Great Recession and a standard increment of 4.3% in the post-recession period. It was worth noting that Beaumont and Lufkin districts showed a continuous drop in the number of urban fatalities in both during and post-recession periods, though the change did not affect much in the overall numbers of the region. The number of rural casualties for the region decreased by 142 during the Great Recession period and increased by 120 after the Great Recession.

The districts having a very high urban population usually covers more than half of the crashes in the region. Here, Dallas and Fort Worth districts comprised more than half of total deaths and over three-fourths of the number of urban fatalities in the North region. The North region showed the highest decline of 11.5% among all the four regions during the Great recession and a modest increase of 7.5% after the Great Recession. The number of urban casualties decreased from 3,532 in the pre-recession period to 2,992 during the Great Recession period and then rebounded back to 3,467 in the post-recession period. However, the number of rural casualties declined by 8.4% between the pre-recession and the post-recession period.

The South region recorded an increment in the total number of fatalities from the pre-recession period to the post-recession period. It went down by 9.8% during the Great Recession before increasing by 14.3% in the post-recession period. As the numbers were similar for the rural regions in the pre and post-recession era, the main reason for the rise in the number of fatalities was attributed to the urban regions. San Antonio and Austin contributed to more than half of the fatalities in the region, and both of them recorded more than 27.4% and 30% increment respectively in the number of urban casualties in the post-recession period as compared to 0.7% and 17.8% respectively in the number of rural casualties.

The West region was unique in the sense that no district was big enough to account for the bulk of the region's fatalities. This region showed the least decrement of 4.6% from the pre-recession to the Great Recession and the most dramatic increase in the post-recession period by 31.4% compared to the post-recession period among all the four regions. The increment in the number of rural fatalities was much higher than the number of urban fatalities, that is, by 35.4% (505 in numbers) in rural areas as compared to 21.6% (127 in numbers) in urban areas. There was only one district, Childress, in the whole state of Texas that had not reported any urban death in ten years (2008-2017).

Figure 4.8 presents the chart capturing the change in the total number of fatalities per 100,000 population of the respective districts and regions between the pre-recession era and the post-recession era. As can be seen, the three districts, namely, East, North, and South, showed a negative change in the number of fatalities even post the Great Recession era, while the situation is opposite for the West region. A brief examination of the results for each region was provided in the following paragraphs.

The East region showed the negative trend for change from both the pre-recession to the Great Recession and the Great Recession to post-recession times. Bryan recorded close to 23.2% decline, the most significant decline in the East region in the post-recession era. For the North region, all the regions recorded a negative growth during the Great Recession period, Dallas being the most affected region. For the post-recession period, Dallas is the only region that showed a definite increase in the number of fatalities per 100,000 population, close to 10%. Overall, the North region registered a decline in the number of 15.9% during the Great Recession and a further 4.3% post the Great Recession.

The South region registered an overall negative trend for both the time series. All the regions showed a negative trend in the first time series, while half of them showed a positive increase in the number of fatalities when moved to the post-recession period. The West region showed an overall different trend. The number of deaths per 100,000 population decreased by 11.2% in the Great Recession period and subsequently increased by 23.6% in the period after the Great Recession. El Paso showed 56.7% increase and Odessa displayed a close to 40% rise in the fatalities in the post-recession era. In fact, all the districts in the West region showed positive growth in the number of deaths per 100,000 population in the post-recession era.

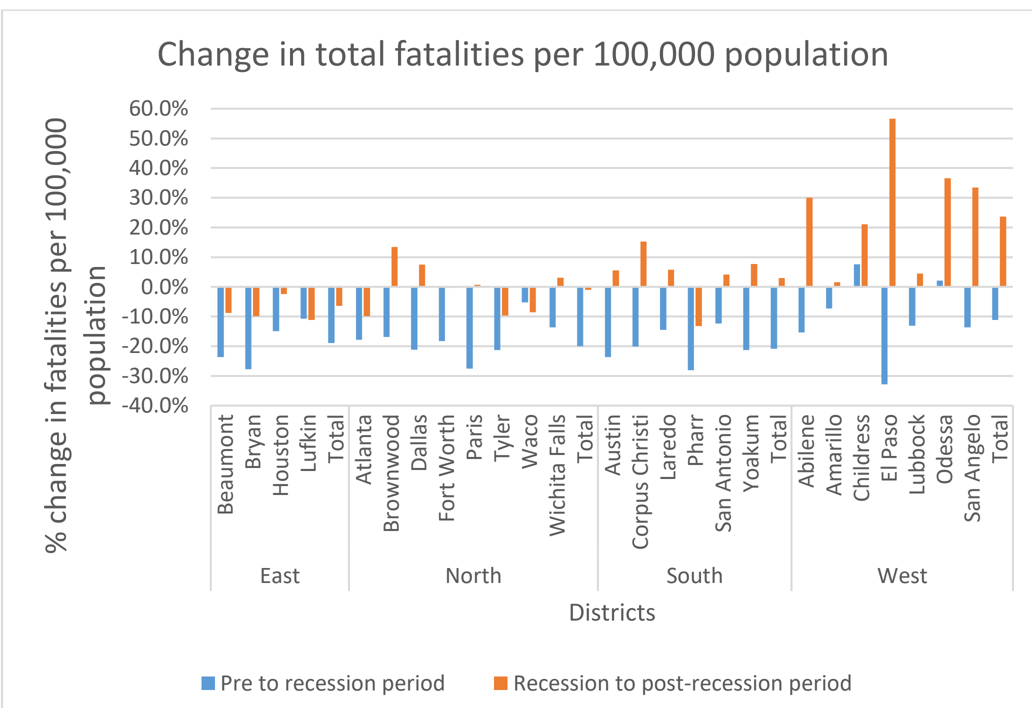


Figure 4.8: Change in fatalities per 100,000 population of each district for each time period

IV.3 Chapter Summary

This chapter has presented a preliminary data analysis and interpreted the results before conducting the spatial analysis. The results provided more profound insights into the data which were summarized as below:

- The national trends section explored various factors pointed out by many researchers in their studies for the real-world data provided for the United States by NHTSA. It showed the decrease in the younger driving population, fewer heavy trucks, and less alcohol-related crashes as the primary reasons for registering a decrement in the number of traffic casualties during the period of the Great Recession.
- The exploratory data analysis on the fatal crash data for the state of Texas showed exciting trends. The North region registered the highest number of fatal crashes while the West region recorded the lowest number of fatal crashes during the three time periods. However, the increment in the number of casualties between the pre-recession and the post-recession periods was highest in the West region.
- The number of urban fatalities fell more significantly compared to the number of rural traffic fatalities, and the trend continued for a long time before again reaching the same numbers by 2017 as in 2003.
- The total number of fatalities reduced for the North and East regions in the post-recession period as compared to the pre-recession period, while the case is opposite for the West and South regions.

- Change in the number of fatalities per 100,000 population was negative for all the regions during the Great Recession. However, during the post-recession period, the fatalities per 100,000 population decreased for the North, East, and South regions, and increased for the West region.

The next chapter describes the spatial statistical tools that were used for the spatial analysis of the traffic casualties' data.

CHAPTER V

SPATIAL STATISTICAL TOOLS FOR ANALYSES

This chapter describes the GIS tools that would be used for analyses in the subsequent chapters. A software called ArcGIS, developed by Environmental Systems Research Institute (Esri) was used to perform the spatial analyses of the three five-year periods of fatal crash data between 2003 and 2017, which were: the pre-recession era, the recession era, and the post-recession era. The spatial autocorrelation between the locations of traffic crashes and the hotspots generated during the three time periods would provide deeper insights into understanding the role of the positional parameters in improving traffic safety at a given location.

Section V.1 gives information on the concept of Geographic Information System (GIS) and the spatial analyses. Section V.2 describes the methodology followed by the tools provided in the spatial statistics toolbox and analysis toolbox used for analyzing the geographic data spatially. Section V.3 concludes the chapter and summarizes the use of the different tools at the end.

V.1 Geographic Information System (GIS)

GIS is a method to gather, manage, and analyze data. It integrates different types of data and analyzes the geographic locations provided in the data. It aids in visualizing different layers of information using maps and, thus, helps to find and examine spatial patterns and correlation between various variables (“What is GIS?,” n.d.)

There are two ways to represent the real-world data: the feature (or vector) data, which stores the exact geographic location, and the raster data, which stores the continuous data in the form of regular grids. Raster grids are usually used to represent the natural environment, such as the elevation, temperature, and precipitation data of an area. In contrast, the vector data are usually used to represent the built environment, such as the roadways data, as well as the administrative data, such as census and countries. The detailed information of each location can be further stored as “attributes” (Esri Press Team, 2018).

In GIS, every dataset is represented in the form of a layer, and these can be combined using various analytical operators to find patterns and correlations. This process is known as overlay analysis. Figure 5.1 explains the basic concept of spatial analysis. Different features provided by their respective datasets represented in the form of layers when stacked together at a geographical location can provide answers to many critical questions. Also, the geometric information, such as direction, length, area, and volume and topological information, such as connectivity, adjacency, and inclusion, can be extracted from the spatial data (Esri Press Team, 2018).

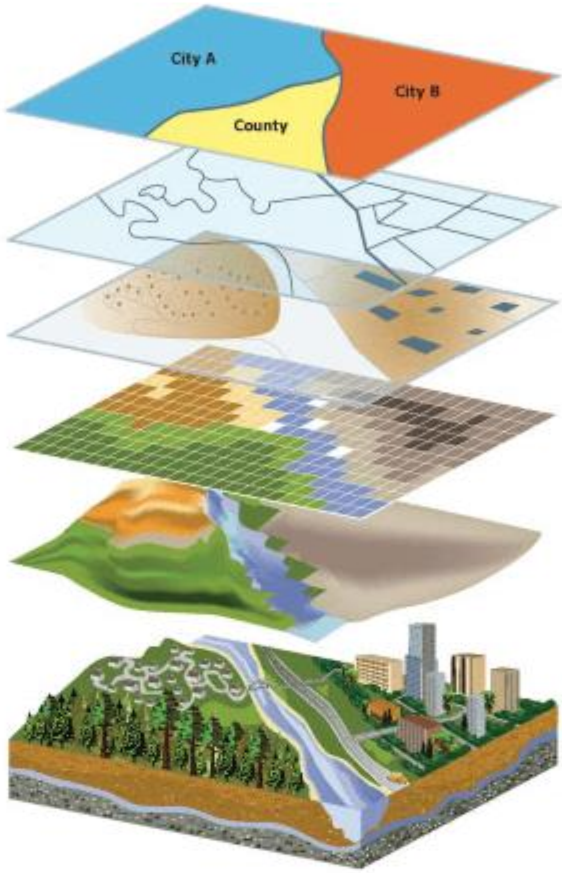


Figure 5.1: Stacking of different layers in GIS (Reprinted from Esri Press Team, 2018).

Spatial analysis is a method to analyze the data geographically by using various tools and techniques and then analyze the results. This analysis has found applications in examining the suitability of a particular location for specific purposes, estimating and forecasting spatial outcomes, understanding the reasons behind gradual or sudden changes in the data, and detecting specific patterns. In this study, the tools present in the Spatial Statistics toolbox in ArcGIS were used for analysis (Esri Press Team, 2018).

V.2 Spatial Statistics and Analysis Toolbox

The fatal crash data were obtained for each year between 2003 and 2017. The dataset was divided into three time periods, pre-recession (2003-2007), during the recession (2008-2012), and post-recession (2013-2017). The county, as well as the rural and urban area details, were available for each fatal crash location. The counties' data availed to identify the districts and regions for each fatal crash geographic location.

To prepare the fatal crashes dataset that could be used for analyses, other shapefiles were required. These were AADT, roadways, counties, and districts data, all of them available on the TxDOT website. Once every dataset was accessible, the next step consisted of identifying the essential tools required for the spatial analyses. Two of them were Optimized Hotspot Analysis tool and Near tool. Subsection V.2.1 discusses the Optimized Hotspot Analysis tools and its components. Subsection V.2.2 discusses both the Near tool and the Generate Near Table tool used in the subsequent chapters.

V.2.1 Optimized Hotspot Analysis Tool

"How Optimized Hot Spot Analysis Works," (n.d.) describes in detail about the spatial autocorrelation tool. This tool generates the hotspots similar to hotspot analysis (Getis-Ord Gi*) tool. However, the difference lies in the initial work done on the input data based on its characteristics to obtain settings that yield the optimal hotspot outcome. The Optimized Hotspot Analysis tool runs three components to identify the correct settings before running the Getis-Ord Gi* statistic. They are *Initial data assessment*, *Incident aggregation*, and *Scale of analysis*. All three components are described in detail below.

V.2.1.1 Initial data assessment

The purpose of this component is to scrutinize whether the dataset provided satisfies the minimum criteria required for a proper analysis by the tool. It checks the dataset thoroughly to find if there are enough number of data points, and sufficient variation is present in the values of the field specified for analysis ("How Optimized Hot Spot Analysis Works," n.d.). In this study, the input features were fatal crashes data for the three time periods, one at a time. The study would be conducted for two parameters: a) high density of casualties per crash location when the death counts would be specified in the Analysis field and, b) high density of casualties per crash location per 100,000 vehicles/day when the death counts/AADT would be specified in the Analysis field.

This component also looks for locational outliers and removes them from the analysis. Locational outliers are those geographic locations whose distance from the nearest neighboring points is much larger than most of the distances between the pair of points. To identify them, this component calculates the average nearest neighbor distance for each data point and then accumulates them to draw the distribution curve of all the distances obtained. The points which are at a distance greater than three standard deviations away from the nearest non-coincident neighboring point are locational outliers.

V.2.1.2 Incident Aggregation

This component aggregates the data. There are two approaches which can be used in the Incident Data Aggregation Methods field described as follows

1. Count_Incidents_within_Fishnet_Polygons: All the coinciding points collapses to create one point at every unique location. Then, the tool computes the average nearest neighbor (ANN), and the median nearest neighbor (MNN) distance at all unique points, barring the locational outliers which were removed in the previous step. The values of

ANN or MNN provides the cell size of the fishnet polygon mesh to be used in the analysis. Finally, it locates the data points in each polygon cell and finds out the hotspots accordingly.

2. Snap_nearby_incidents_to_create_weighted_points: Similar to the first method, this method also calculates the values of ANN or MNN. However, the difference lies in the calculation of the snapping distance from data points based on ANN or MNN values. All the data points within the adjusted snap distance collapse into one point and the number of data points snapped together decides the weights.

V.2.1.3 Scale of Analysis

This component computes the distance between the data points at which the clustering is most significant. It uses the incremental spatial autocorrelation method to find the distance. This method employs Global Moran's I statistic that increases the distance gradually and returns the z-score, which acts as a measure of the intensity of clustering between the points at that distance. Usually, z-scores increase with the distance and peaks at a certain distance before going down. The peak distance becomes the scale of analysis. If multiple peaks are present, then the first peak distance is used.

If the above method does not show any no peak distance, then the tool uses other approaches. It scrutinizes the spatial distribution of input features and decides the K number of neighbors for every input feature. K is calculated as 0.05 * Number of input features. K is adjusted to have a value between three and 30. In this study, 0.05 * Number of input features is always greater than 30. So, K is always equal to 30. Now, the tool calculates the average distance of the K neighbors and check if it is within one standard distance. If it is within one standard distance, then the value of the average distance calculated is taken; otherwise, one standard distance is taken as the average distance between the neighboring points.

The standard distance (SD) is the radius of a circle around a feature, and it measures the compactness of a bunch of features. The formula is as follows:

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n} + \frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{n} + \frac{\sum_{i=1}^{n}(z_i - \bar{Z})^2}{n}} \quad (14)$$

Here, x_i, y_i, and z_i are geographic coordinates of feature i, $(\bar{X}, \bar{Y}, \bar{Z})$ provides the Mean center calculated for the set of features, and n represents the number of input features.

V.2.1.4 Hotspot analysis

After determining all the settings for the data, the tool is ready to apply the Getis-Ord Gi* statistic to find the optimal number of hotspots. The resultant z-scores will help to identify the hotspots (the cluster of points having high values) and the coldspots (the cluster of points having low values). It also applies the False Discovery rate (FDR) correction method to adjust for two problems: multiple testing and spatial dependency. Both the problems are described in detail below ("What is a z-score? What is a p-value?," n.d.):

1. Multiple Testing: This is an issue when a large number of tests are conducted based on the number of input features in the dataset. As for the fatal crashes dataset having more than 8,500 data points in each time period, the statistical tool will perform tests on each point. If the confidence level is 95 percent, there are more than 425 chances that a spatial pattern may appear structured (dispersed or clustered), although the pattern could be random. Due to statistically significant p-values, the null hypothesis could be falsely rejected ("What is a z-score? What is a p-value?," n.d.)

2. Spatial Dependency: The hotspot analysis tool evaluates each location based on the values associated with the neighboring data points. It means that the points near to each other share the same neighboring points, and therefore, the tests on the data points are not independent as required by the statistical tools to provide correct outcomes. This problem is known as spatial dependency ("What is a z-score? What is a p-value?," n.d.).

The solution is to apply FDR correction. This technique first finds out the number of false positives obtained from the tests for a specific confidence interval and adjusts the critical p – values accordingly. Once generating the statistically significant points, the tool then arranges their p-values in increasing order and removes the p-values in the lower order. All the statistically significant p-values are arranged in increasing order, and as per the false positives numbers, the p-values in the lower order are removed ("What is a z-score? What is a p-value?," n.d.).

V.2.2 Near and Generate Near Table Tools

Initially, the hotspot analysis was conducted using the death counts of the crash locations to generate the hotspots. However, as more traffic on a particular location could result in a higher number of deaths, it was not fair to declare them as hotspots. A new factor of AADT for normalizing the results was introduced. The data was accessed from the TxDOT website, which provided the AADT at various locations throughout the state of Texas.

Now the problem lied in the fact that the data point locations providing the AADT and fatal crashes locations were adjacent to each other but did not coincide. The Near tool helped to match the fatal crash locations and nearest AADT points. The Near tool assigns and calculates the distance between the input feature in one dataset and near feature in another dataset. It adds two columns in the input feature dataset: Near_FID, the object ID of neighboring near feature, and Near_Dist, the distance between near and input feature ("Near (Analysis)," n.d.).

Generate Near Table tool works similarly as the Near tool. However, the difference is that this tool assigns more than one near features to the input data. Secondly, it generates a separate table instead of modifying the input attribute table. Similar to the Near tool, in this study, the input features are the geographic locations of the fatal crash, and the near features are AADT locations. The maximum number of closest matches option can be used to provide the maximum number of nearest features required for a particular point in the output table. In this study, the maximum number of closest matches was three. The output table has four columns; their description is provided in Table 5.1 ("Generate Near Table (Analysis)," n.d.).

Table 5.1: The output columns of the Generate Near Table tool (Adapted from "Generate Near Table (Analysis)," n.d.)

Column Name	Description
In_FID	The object ID of input point feature
Near_FID	The object ID of nearest point feature
Near_Dist	The distance between the near and the input feature. The distance is provided in the units of the coordinate system of input features, or meters if GEODESIC is used for method parameter with the geographic coordinate system.
Near_Rank	The ranking is based on the distance between the near and input feature. The closest near feature is ranked 1, and the second closest near feature is ranked 2, and so on.

As the AADT values were assigned to each of the fatal crash locations, a new column dividing the death counts by nearest correct AADT values was added to normalize the death counts at a location. The Optimized Hotspot Analysis tool was again used to obtain hotspots. The results were compared with the previous approach without normalizing the results. Also, the hotspots and coldspots generated were divided into rural and urban areas to understand how these numbers varied for the three periods.

V.3 Chapter Summary

This chapter has covered a detailed explanation of the methodology adopted by the different tools to conduct the analysis spatially. The main points are as follows:

- GIS performs analysis on the geographic data spatially
- The Optimized Hotspot Analysis tool first analyzes the data to decide the parameters to obtain the optimal number of hotspots.
- The Near tool assigns the geographic locations of one dataset with the nearest neighboring geographic data points of another dataset.

The next chapter describes the characteristics of the datasets generated to perform analysis and the results of the hotspot analysis.

CHAPTER VI

RESULTS AND DISCUSSIONS

This chapter provides deeper insights into the results obtained from the hotspot analysis based on two parameters: death counts and death counts/AADT. Before examining the results, the data needed to be cleaned up by removing the data points, which either did not have geographic locations or nearest AADT values. Section VI.1 covers the details of the screening of data. Section VI.2 highlights the results for all four regions and 25 districts obtained for each of the three time periods and provides comparisons between them. Section VI.3 discusses the significance of the results and the inferences drawn. Section VI.4 summarizes the main points of this chapter at the end.

VI.1 Data Cleaning

As discussed in the previous chapters, the hotspot analysis was performed on the three five-year time periods. These time periods were decided based on the fatal crash data for the state of Texas and were defined as follows:

1. *Pre-recession period*: The period between 2003 and 2007 is defined as the time period before the period of the Great Recession for this study.
2. *Recession period*: The period between 2008 and 2012 is defined as the time period when the number of traffic fatalities experienced a dip compared to the previous time period, a result of the impact of the Great Recession.

3. *Post-recession period*: The period between 2013 and 2017 is defined as the time period when the number of traffic fatalities bounced back to the pre-recession period, an indication that the influence of the Great Recession is over.

The traffic fatalities data of each year in 15 year study period were merged into three datasets of five years, as discussed above. The hotspots were generated based on the two variables, death counts in each fatal crash, and the death counts divided by AADT for each fatal crash. As more traffic over the roads could lead to more fatal injuries, it was considered necessary to generate the second variable normalizing the death counts with AADT. To do that, AADT data and the roadways data needed to be first downloaded from the TxDOT's Open data portal. The AADT dataset was obtained in a similar format as the traffic fatalities data, that is, point features representing AADT value recorded at each location, while the roadways dataset provided the road locations as line features. AADT value of a point for each time period was calculated as the average of the AADT values for five years in that time period. The Near and Generate Near Table tools were then used to assign the nearest AADT points with each fatal crash point. The steps were given below:

1. The nearest AADT points to the fatal crash locations were assigned using the Near tool. However, at some places, the AADT points on the other roads were assigned to the fatal crash. For example, in Figure 6.1, the fatal crash with crash ID 12945078, shown with the pushpin sign, has three nearest AADT points having object IDs 17767, 17768, and 17769 shown with the asterisk sign. As the AADT point, object ID 17769, is closest to the fatal crash location, the Near tool assigns this point to the fatal crash point. However, this is present on the other intersecting road and is not the AADT of the road of the fatal

crash location. There could be numerous similar cases like that, and therefore, it was essential to locate and assign the correct nearest AADT points to the fatal crash locations.

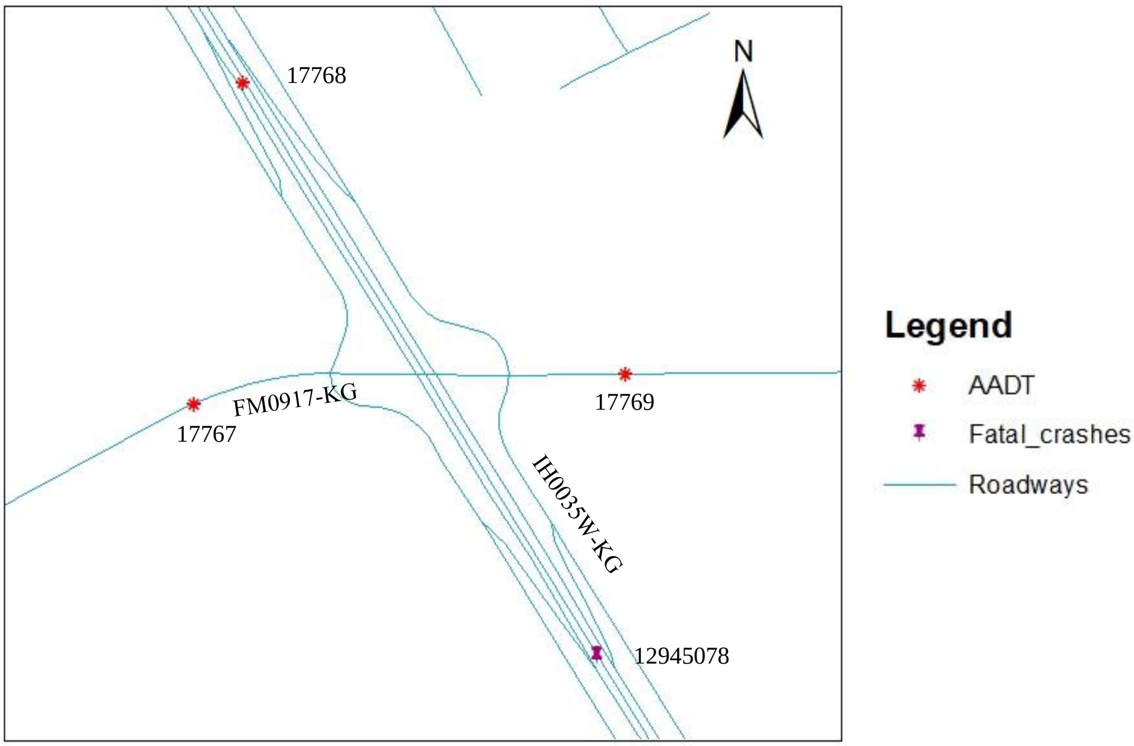


Figure 6.1: Example showing fatal crash location and their nearest three AADT locations

2. The Generate Near Table tool was used to produce the nearest three AADT points for each fatal crash location to find out the correct AADT point. The roadway dataset was roped in to identify the correct AADT points for the crash locations.
3. The Near tool was then utilized to identify the roadway segments over which both the AADT and the fatal crash locations were present. In Figure 6.1, the fatal crash is located on roadway segment IH0035W, the first two nearest AADT points are located on FM0917, and the third closest AADT location is on the same road as the fatal crash location.

4. Two new columns providing roadways for both the nearest AADT points and the fatal crash location were added. For each fatal crash location, if the closest AADT point had the same roadway segment as the fatal crash location, that AADT location would be assigned to the fatal crash location, otherwise the next closest location would be checked. All the three closest AADT locations were checked sequentially by their rank, closest ones having the lowest ranks. Table 6.1 shows details for the fatal crash location in Figure 6.1. AADT location having object ID 17768 and rank three is located on the same road as fatal crash location and is, therefore, matched with the fatal crash location.

Table 6.1: Example to show how correct AADT is chosen

Sr. No.	Fatal crash			AADT				Match
	FID	Crash ID	Roadway	AADT FID	Object ID	Rank	Roadway	
1	13155	12945078	7674	16768	17769	1	506752	N
2				16766	17767	2	506752	N
3				16767	17768	3	7674	Y

5. If none of the three AADT locations had the same road segment as the fatal crash location, then it was highly likely that the AADT value was not recorded on that whole road segment and should be removed. In Figure 6.2, the fatal crash location is present on the road segment CS259508. However, all the three nearest AADT locations are present on different road segments, two on FM2252 and one on SL1604. So, this fatal crash point could not be used for the analysis and, therefore, was ignored.

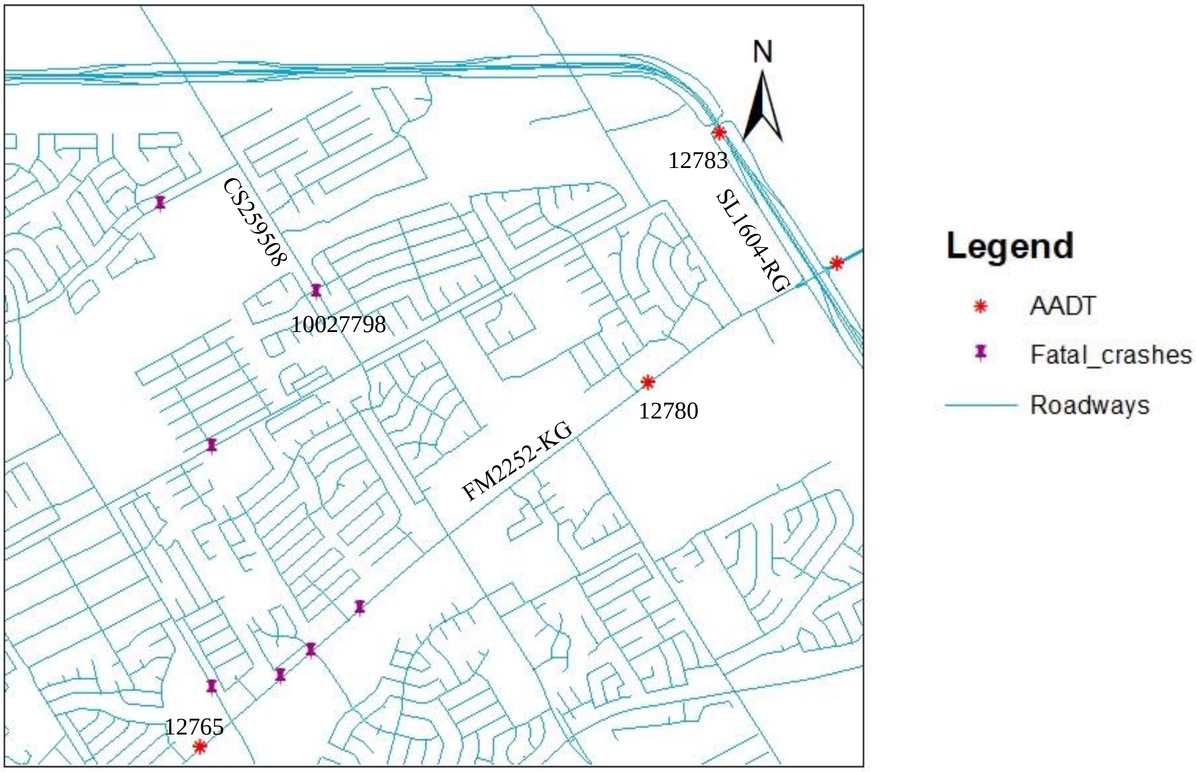


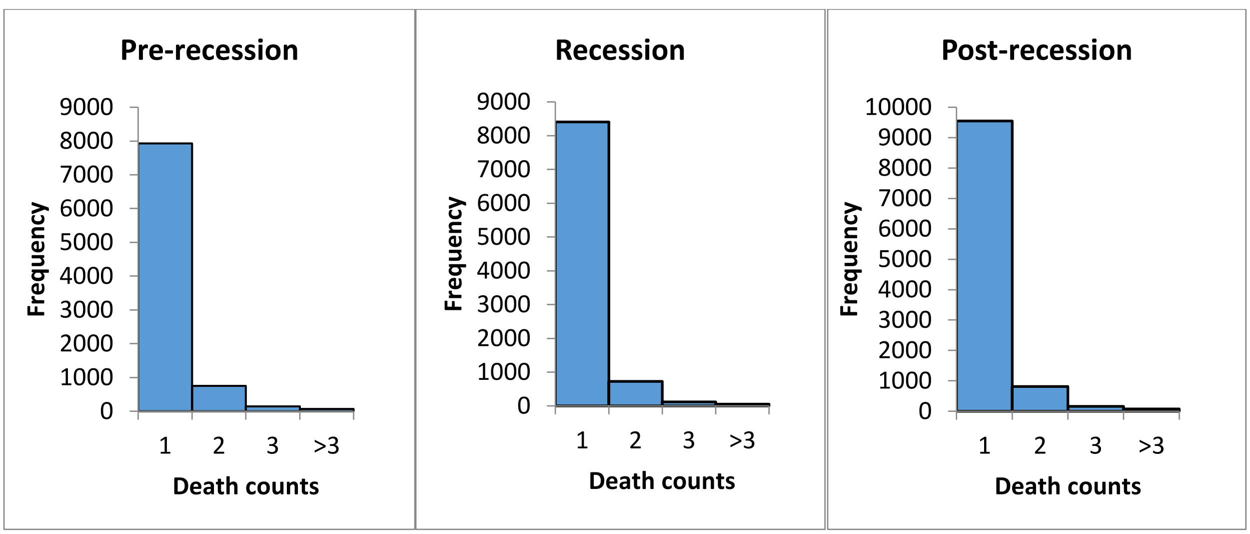
Figure 6.2: Example showing the fatal crash location with no nearby correct AADT location

6. After eliminating all the fatal crashes which did not have any AADT values matched, when the AADT values matched were checked for the remaining fatal crash data points, it was observed that many locations had AADT values as zero. Upon further examination, it was noticed that various AADT points were not present before and during the Great Recession periods but were recording the AADT values in the post-recession periods. So, for them, AADT values were zero for the first two time periods sequentially. It was decided to check for the next ranked AADT points for them and repeat step 4 and step 5.

After assigning the closest correct AADT value to each fatal crash location and removing all the fatal crash values which did not have any AADT value matched or AADT value as zero, the final datasets for the pre-recession, during the recession and the post-recession periods had 8,894, 9,298, and 10,588 data points respectively. The next section provides the frequency distribution of the values of the two parameters: death counts and death counts/AADT before delving into hotspot analysis.

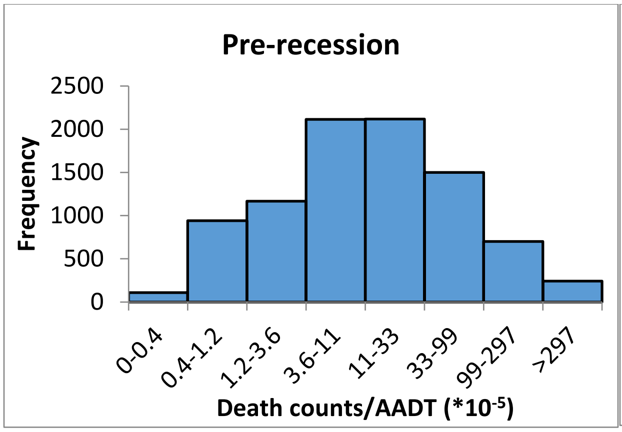
VI.2 Results for the Three Time Periods

The hotspot analysis was performed on two parameters: a) the number of fatalities occurred in each crash, and b) the number of fatalities divided by their AADT values. Before discussing the hotspot analysis, it is crucial to look into the frequency distribution of the values of both the parameters for the three time periods. The results are shown in Figures 6.3 and 6.4.

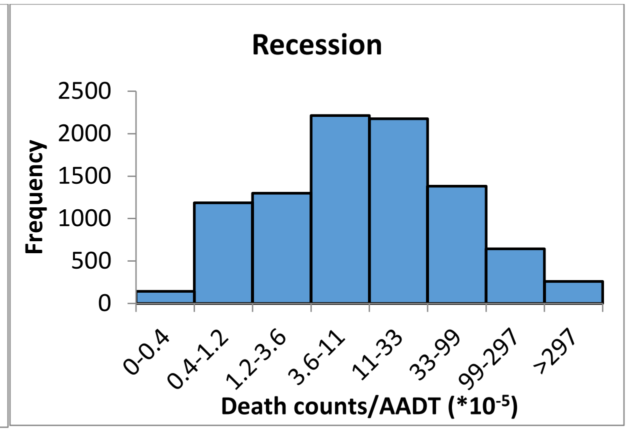


A. Pre-recession period B. Recession period C. Post-recession period

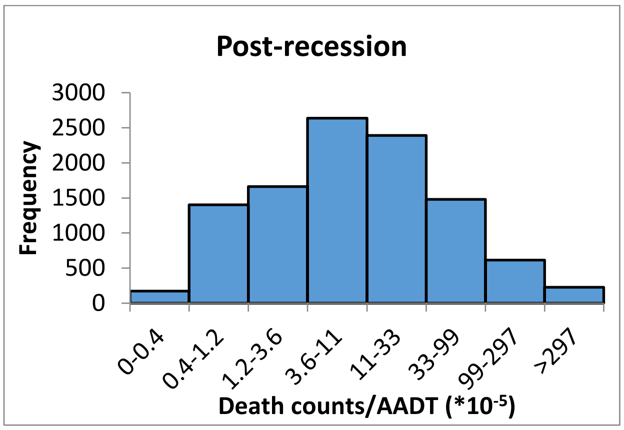
Figure 6.3: The frequency distribution of the death counts for all the three time periods



A. Pre-recession period



B. Recession period



C. Post-recession period

Figure 6.4: The frequency distribution of the parameter death counts/AADT for all the three time periods

As can be seen in Figure 6.3, roughly 90% of the fatal crashes recorded have only one fatality in all three time periods, while rest 10% of the fatal crashes have more than one fatality. Also, a sudden spike of roughly 2.5% in the numbers in the lower bands between zero and 1.2 can be noted in the frequency distribution of the parameter death counts/AADT during the recession period as compared to the pre-recession period. For Figure 6.4, the number of fatalities divided by AADT at the crash location is shown as a value multiplied by 10^5. For example, if the value of the death counts/AADT at a given location is 11, it means 11 deaths occurred at the location on the road having 100,000 vehicles traveling per day. As can be seen from Figure 6.4, the values vary widely across the spectrum. Half of the values are present below 11, and the rest half of them range from 12 to 6000.

The Optimized Hotspot Analysis was performed for the three time periods for both the parameters. Tables 6.2 and 6.3 provide results of hotspot analysis on both the death counts and the death counts/AADT. For the hotspot analysis based on the death counts, the scale of analysis achieved peak clustering in the post-recession period, however, there was no peak clustering obtained for the pre-recession and the recession periods and the optimal fixed distance band was calculated based on the average distance calculated for 30 nearest neighbors of input features. For the other parameter, the Optimized Hotspot Analysis tool obtained the peak clustering distance in all the three time-periods.

Table 6.2: Results for hotspot analysis based on the death counts

Description	Pre-recession	Recession	Post-recession
Number of input features	8,894	9,298	10,588
Locational outliers	181	190	204
Optimal fixed distance band	20,658 meters	19,953 meters	17,145.68 meters
Statistically significant clusters	46	183	148

Table 6.3 Results for hotspot analysis based on the death counts/AADT

Description	Pre-recession	Recession	Post-recession
Number of input features	8,894	9,298	10,588
Locational outliers	181	190	204
Optimal fixed distance band	16,682.46 meters	20,257.27 meters	15,271.45 meters
Statistically significant clusters	2,029	3,847	3,788

Figures 6.5 and 6.6 show the number of statistically significant points that form either hotspots or coldspots with Gi_Bin ranging from -3 to 3, not including 0, where the points having negative values of Gi_Bin are called coldspots and the points having positive values of Gi_Bin are called hotspots. Gi_Bin having value as zero represent points that are not statistically significant, values as -1 and 1 represent points at 67% confidence level, values as -2 and 2 represent points at 95% confidence level, and values as -3 and 3 represent points at 99% confidence level. Tables 6.2 and 6.3 give the number of statistically significant points obtained for both the parameters and Figures 6.5 and 6.6 show their distribution across the regions for all the three time periods, pre-recession, recession and post-recession periods.

As can be seen in Figure 6.5, the numbers do not show any significant patterns for the three time periods. The two regions, East and North, showed a drastic increment in the number of statistically significant points in the recession period before coming down again to the previous levels in the post-recession period. The numbers increased drastically in the post-recession period in the South region and increased consistently in the West region throughout the three time periods. Overall, it was challenging to find a consistent pattern in four regions taking the death counts as the parameter of hotspot analysis. However, if hotspots were taken based on the death counts/AADT value, as in Figure 6.6, some consistent patterns can be found. The number of statistically significant points suddenly increased by around 90% during the recession period as compared to the pre-recession period suggesting the concentration of the high fatal crash points during the recession period. The numbers remained more or less similar in the post-recession periods as in the recession period for all the four regions.

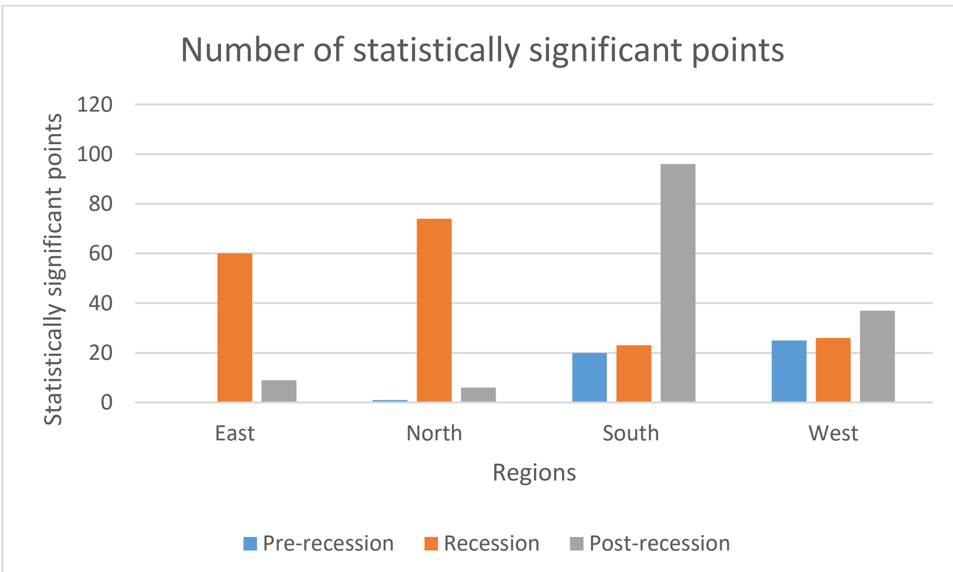


Figure 6.5: Number of statistically significant points for hotspots based on the death counts

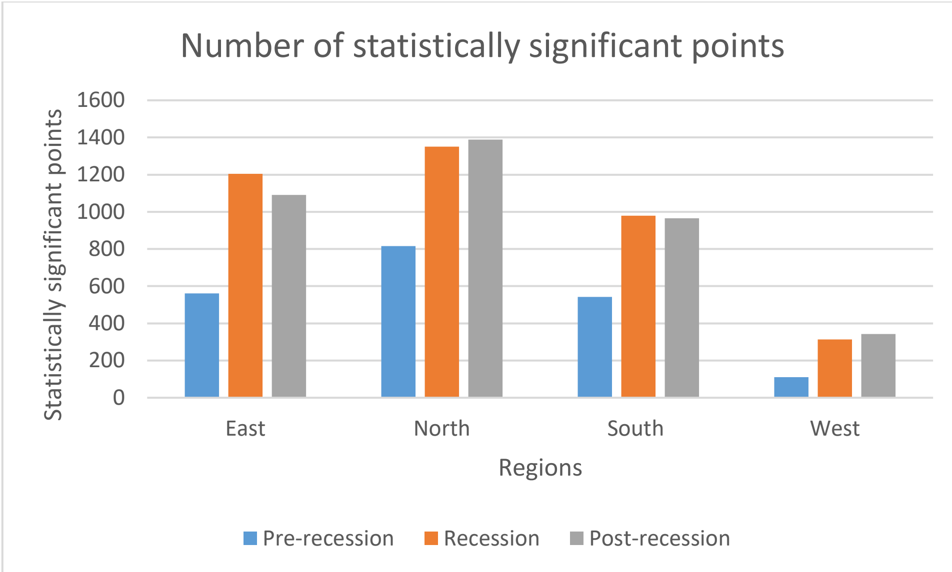
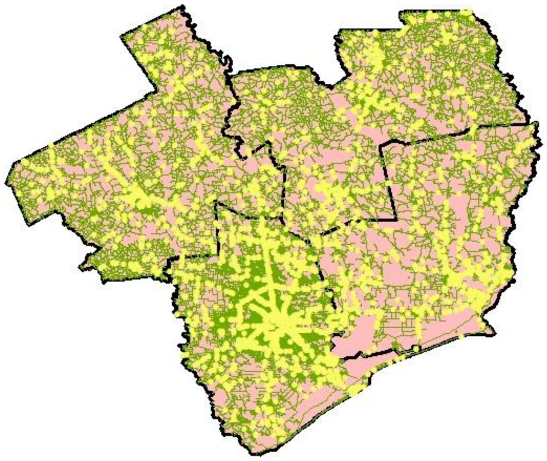
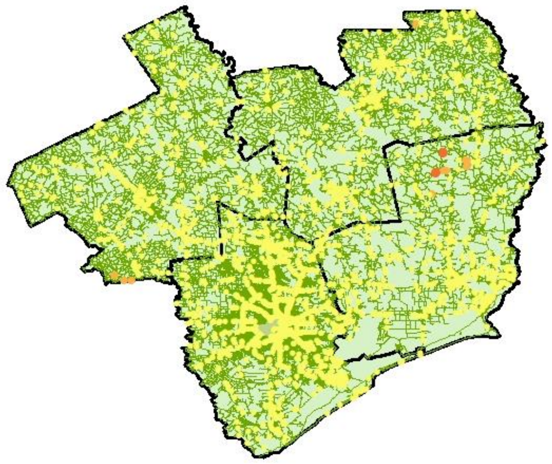


Figure 6.6: Number of statistically significant points for hotspots based on the death counts/AADT

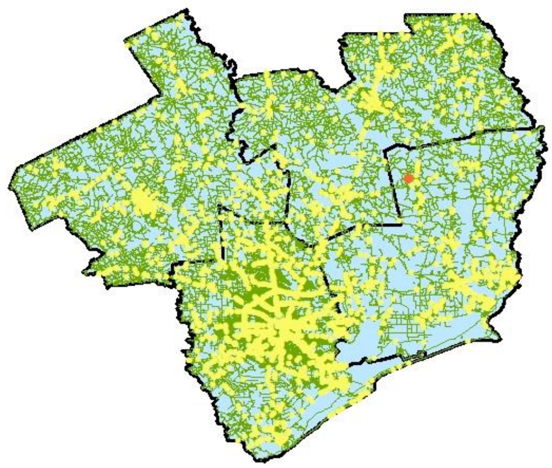
Figures 6.7-6.14 show the hotspots for all the four regions and 25 districts for both the parameters. Red circles and blue circles represent hotspots and coldspots, respectively. Darker the color of hotspots and coldspots, higher the absolute value of Gi_Bin. Figures 6.7-6.10 represent the results of the hotspot analysis based on the number of fatalities in the four regions, while Figures 6.11-6.14 represent the results of the hotspot analysis based on the number of fatalities divided by AADT value at that fatal crash location. As can be seen in Figures 6.7-6.10, no coldspots are present in the first four figures. Also, the hotspots have changed randomly over the years in terms of both numbers and geographic locations. For Figures 6.11-6.14, there are both hotspots and coldspots present. Most of the hotspots are located around the same area over the years, and new hotspot locations are added, especially during the recession period. The coldspots are concentrated in the cities and urban areas in each region.



A. Pre-recession period

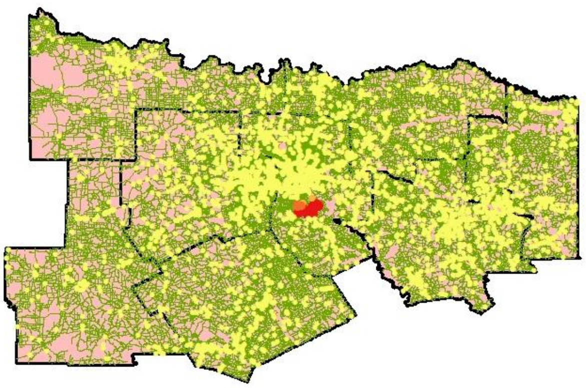


B. Recession period

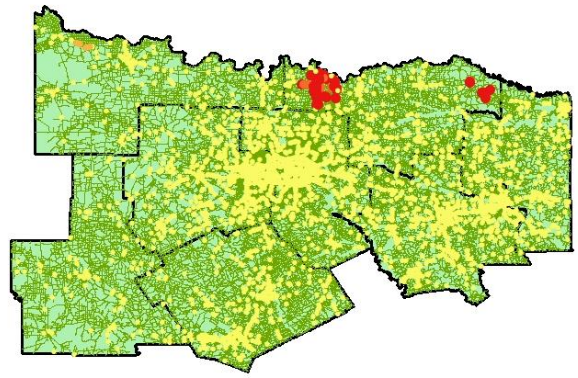


C. Post-recession period

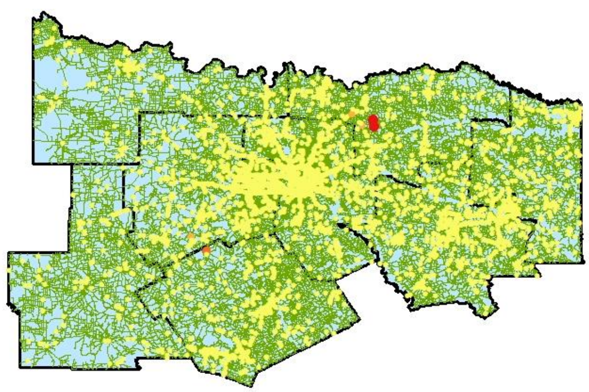
Figure 6.7: Hotspots in the East region based on the death counts



A. Pre-recession period

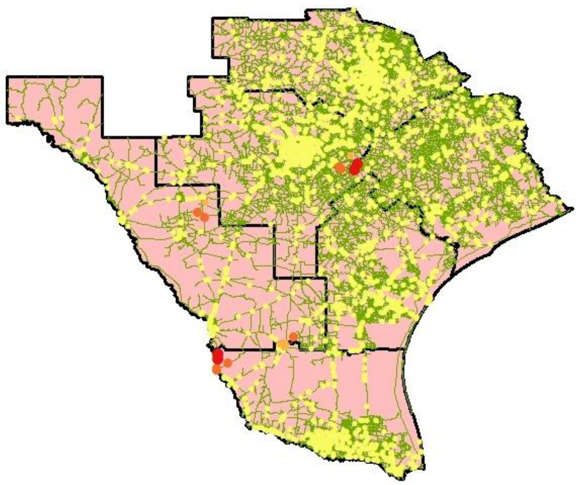


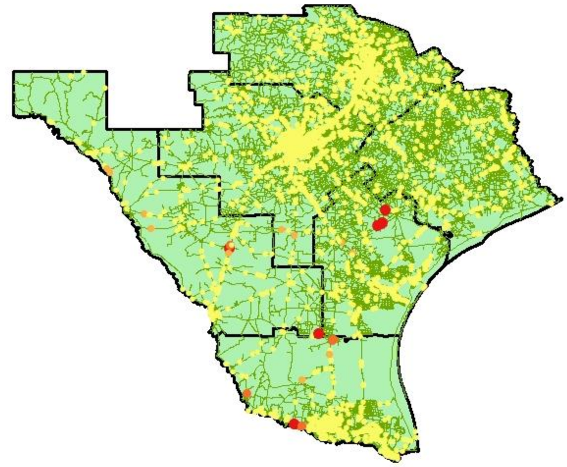
B. Recession period



C. Post-recession period

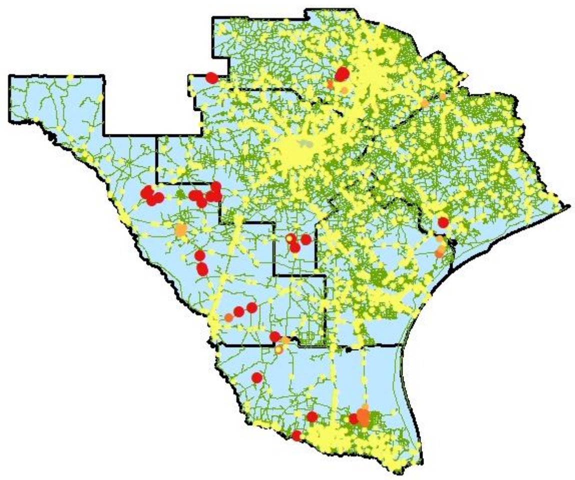
Figure 6.8: Hotspots in the North region based on the death counts





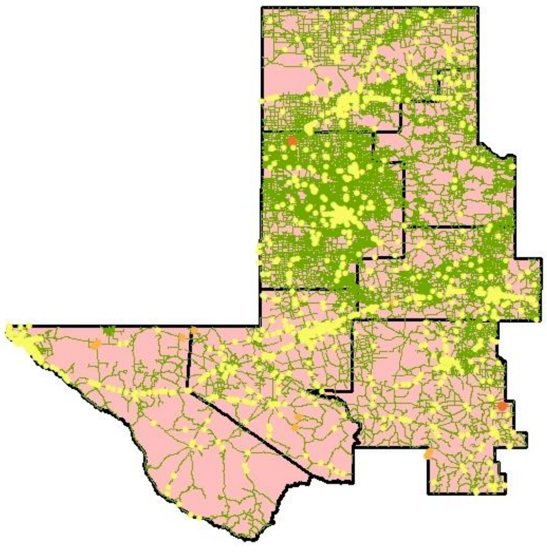
A. Pre-recession period

B. Recession period

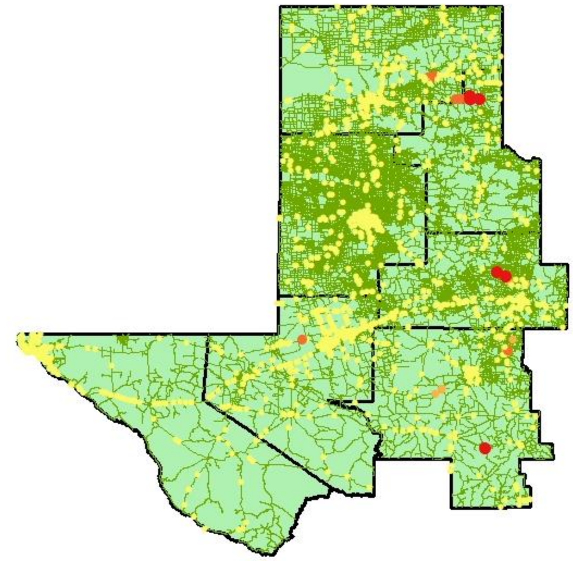


C. Post-recession period

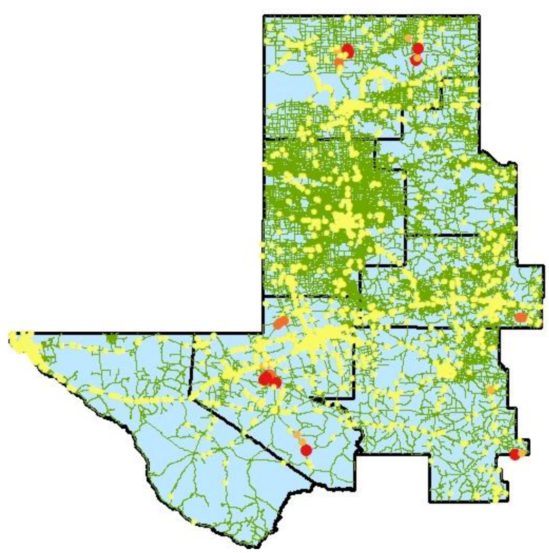
Figure 6.9: Hotspots in the South region based on the death counts



A. Pre-recession period

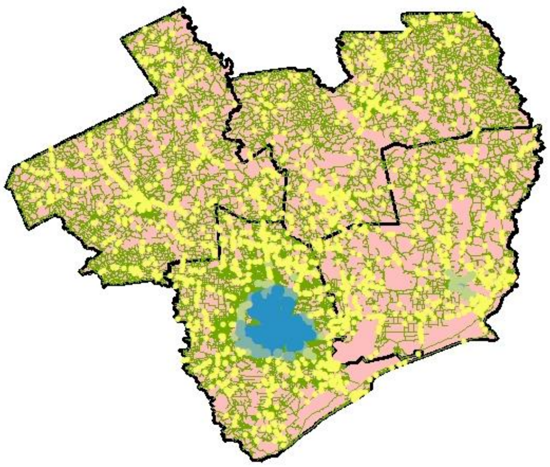


B. Recession period

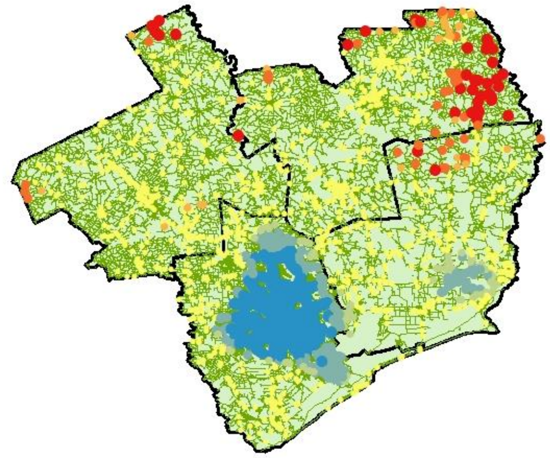


C. Post-recession period

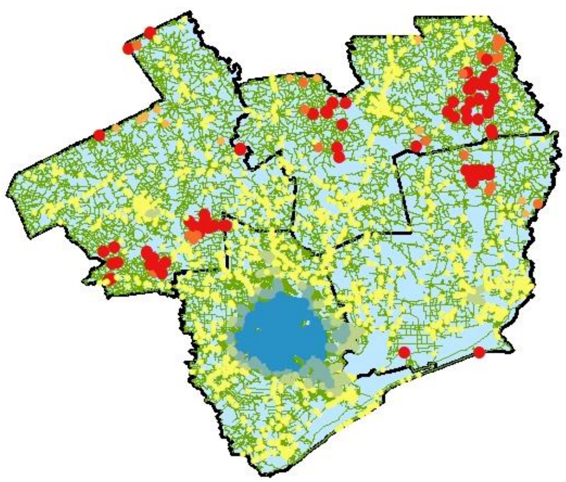
Figure 6VI.10: Hotspots in the West region based on the death counts



A. Pre-recession period

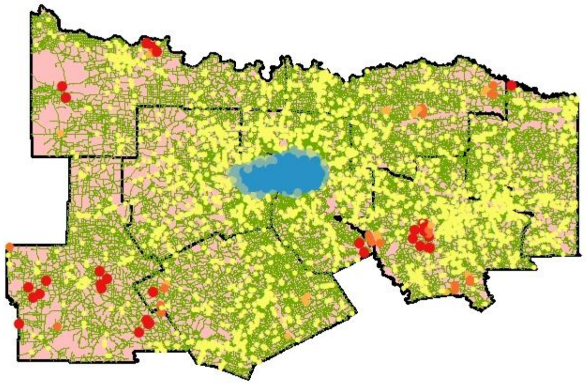


B. Recession period

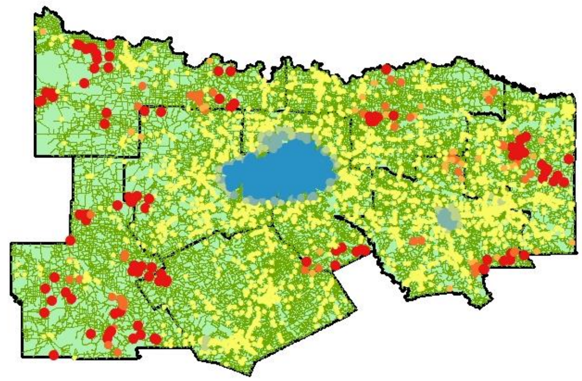


C. Post-recession period

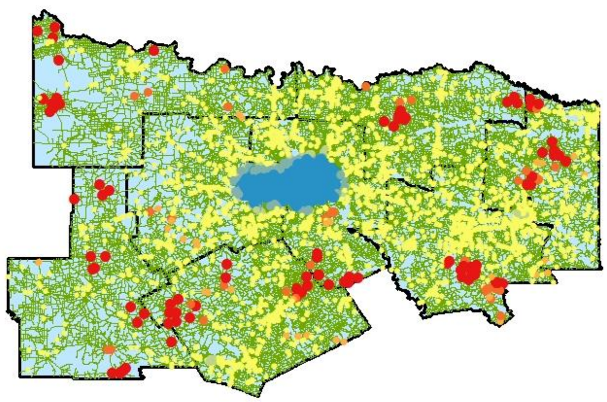
Figure 6.11: Hotspots in the East region based on the death counts/AADT



A. Pre-recession period

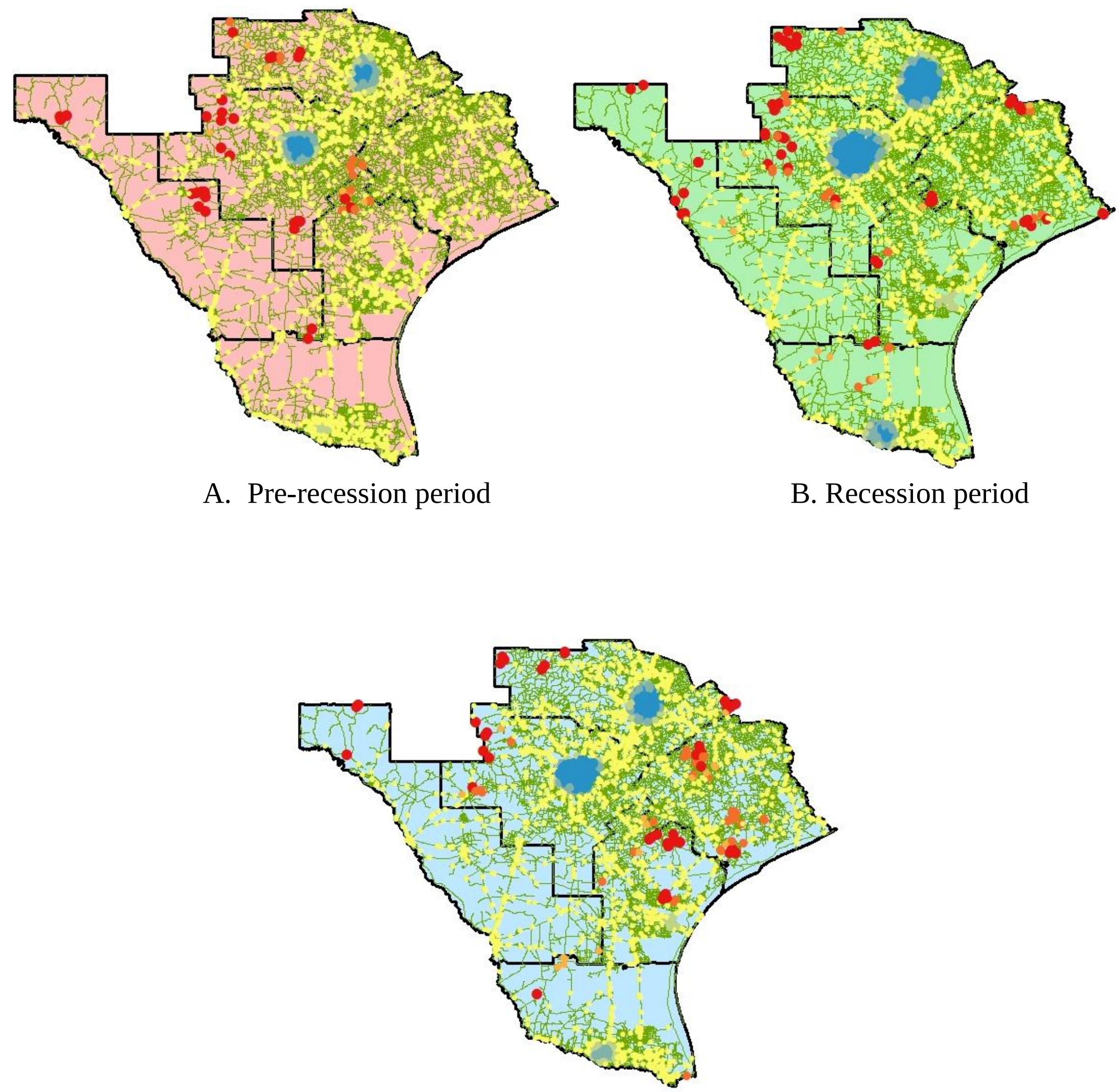


B. Recession period



C. Post-recession period

Figure 6VI.12: Hotspots in the North region based on the death counts/AADT

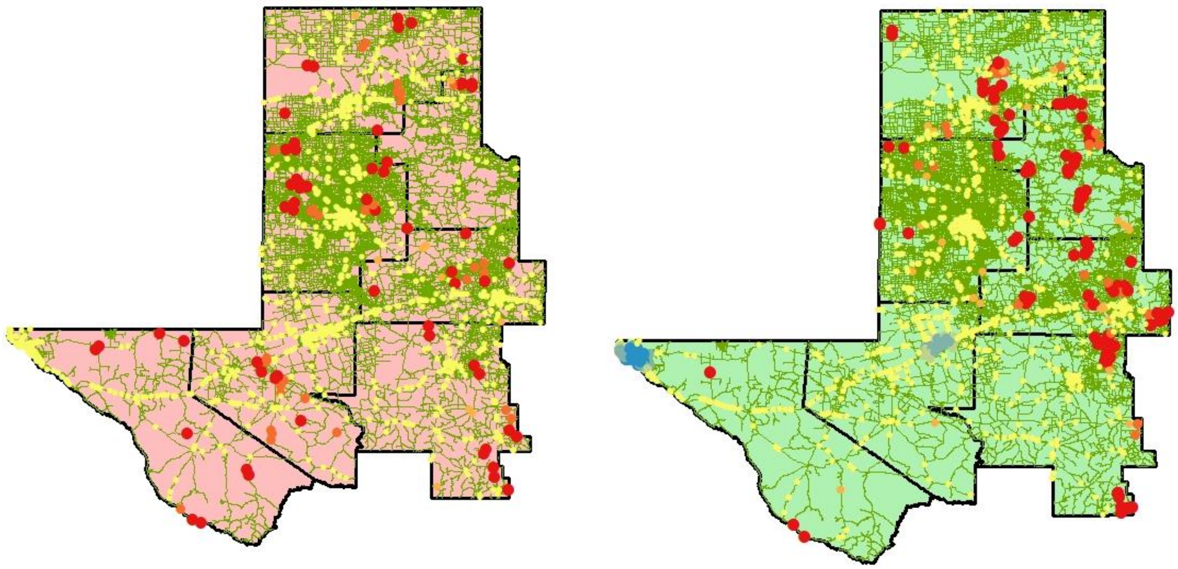


A. Pre-recession period

B. Recession period

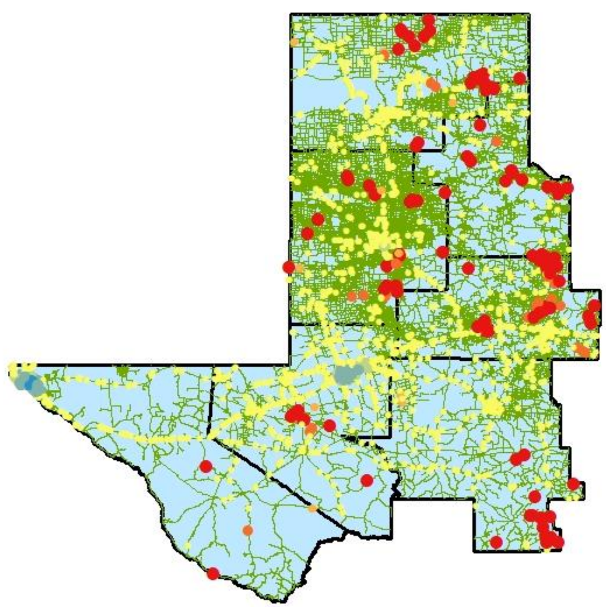
C. Post-recession period

Figure 6VI.13: Hotspots in the South region based on the death counts/AADT



A. Pre-recession period

B. Recession period



C. Post-recession

Figure 6.14: Hotspots in the West region based on the death counts/AADT

Tables 6.4-6.6 show the hotspots and coldspots that are present at 95% and 99% confidence levels for both the parameters: death counts and death counts/AADT.

Table 6.4: Hotspots for 95% and 99% confidence level based on the death counts

Region	District	Pre-recession			Recession			Post-recession		
		Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total
East	Beaumont					2	2		1	1
	Bryan									
	Houston									
	Lufkin									
East Total						2	2		1	1
North	Atlanta									
	Brownwood					2	2			
	Dallas									
	Fort Worth									
	Paris				38	29	67		2	2
	Tyler									
	Waco								1	1
	Wichita Falls									
North Total					38	31	69		3	3
South	Austin								7	7
	Corpus Christi					4	4		1	1
	Laredo		3	3		1	1		16	16
	Pharr	1	5	6		7	7	1	10	11
	San Antonio		4	4					5	5
	Yoakum								1	1
South Total		1	12	13		12	12	1	40	41
West	Abilene					2	2		3	3
	Amarillo					7	7	1	5	6
	Childress					4	4		1	1
	El Paso	3	10	13						
	Lubbock		1	1						
	Odessa					1	1		13	13
	San Angelo		1	1		2	2		1	1
West Total		3	12	15		16	16	1	23	24
Grand Total		4	24	28	38	61	99	2	67	69

Table 6VI.5: Hotspots for 95% and 99% confidence level based on the death counts/AADT

		Pre-recession			Recession			Post-recession		
Region	District	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total
East	Beaumont					14	14	2	14	16
	Bryan					10	10		35	35
	Houston								2	2
	Lufkin					64	64		50	50
East Total						88	88	2	101	103
North	Atlanta		1	1		25	25		22	22
	Brownwood		11	11		42	42		17	17
	Dallas		2	2		8	8		11	11
	Fort Worth					8	8		2	2
	Paris		11	11		14	14		12	12
	Tyler	6	15	21		22	22	5	22	27
	Waco		3	3		6	6		20	20
	Wichita Falls		6	6	2	34	36		15	15
North Total		6	49	55	2	159	161	5	121	126
South	Austin		9	9		9	9		12	12
	Corpus Christi		8	8		4	4	1	19	20
	Laredo		12	12		10	10		4	4
	Pharr					6	6		2	2
	San Antonio	1	15	16		23	23		11	11
	Yoakum	2		2		21	21		38	38
South Total		3	44	47		73	73	1	86	87
West	Abilene		10	10		34	34		40	40
	Amarillo		14	14	3	28	31		20	20
	Childress	1	6	7		28	28		18	18
	El Paso		10	10		3	3		6	6
	Lubbock	2	23	25		15	15	2	23	25
	Odessa		18	18					12	12
	San Angelo		15	15		31	31		15	15
West Total		3	96	99	3	139	142	2	134	136
Grand Total		12	189	201	5	459	464	10	442	452

Table 6.6: Coldspots for 95% and 99% confidence level based on the death counts/AADT

		Pre-recession			Recession			Post-recession		
Region	District	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total
East	Beaumont	3		3	40	14	54			
	Houston	343	98	441	593	312	905	539	262	801
East Total		346	98	444	633	326	959	539	262	801
North	Dallas	443	4	447	613	40	653	738	7	745
	Fort Worth	241	4	245	312	22	334	380	17	397
	Tyler				31	11	42			
North Total		684	8	692	956	73	1029	1118	24	1142
South	Austin	143	28	171	186	85	271	204	62	266
	Pharr				77	30	107	54	3	57
	San Antonio	230	5	235	333	61	394	401	15	416
South Total		373	33	406	596	176	772	659	80	739
West	El Paso				8	3	11	20	1	21
	Odessa				52	29	81	52	41	93
West Total					60	32	92	72	42	114
Grand Total		1403	139	1542	2245	607	2852	2388	408	2796

VI.3 Discussion on the Results

In this study, the hotspot analysis was conducted for two variables: the number of fatalities that occurred in the crash and the number of fatalities at a location per 100,000 vehicles/day. First, the dataset was prepared and then compared for both approaches. As can be seen, the number of input features was 8,894, 9,298, and 10,588 for the pre-recession, the recession, and the post-recession periods respectively. To find the optimal fixed distance band, the tool has to identify the beginning distance and increase it gradually using an incremental autocorrelation tool to find the distance at which the peak clustering is obtained. Beginning distance is by default the minimum distance for which each feature in the dataset has at least one neighbor. However, some points have much higher neighboring distances than most of the other points. Starting with this distance would assign thousands of neighbors to other points that are relatively much closer to each other and unnecessarily increase the computational time. To avoid this, the tool identifies the points having a neighboring distance greater than three standard deviations as locational outliers and removes them temporarily for computing the optimal fixed distance band.

The optimal fixed distance band was computed and hotspots were generated for all three time periods. While the number of input features was highest for the post-recession periods, the number of hotspots generated was highest during the recession period. This observation held for both approaches. A necessary inference drawn is that the number of hotspots generated is independent of the number of input features.

For the first approach, 90% of the fatal crashes had only one person who died, and the rest of the crashes had two or more people who died. So, when the hotspot analysis was conducted, it is not surprising to observe that there were no coldspots as 90% of the data points had the same lowest value. Wherever there are high-value points, the software awarded all the neighboring points surrounding the high-value point as hotspots. That is why the results have been erratic and are not consistent over the three time periods, as can be noticed in Figure 6.5.

However, the case is different with the hotspot analysis conducted on the death counts/AADT variable. In this case, the values of the data points vary from 0.27 to 6000, as shown in Figure 6.4. As the high-value points and the low-value points have a range of values, it is possible to generate both the coldspots and hotspots for the fatal crash locations. As can be observed from Figures 6.11-6.14 and Table 6.5, the coldspots are more concentrated in the densely populated regions, that is, cities and urban areas, which indicate a majority of the crashes occurring in the urban areas have lower fatality rates. Most of the hotspots are present in the rural areas and very few of them (<6%) in the urban areas. Also, the results are more consistent for all the regions over the three time periods displaying fewer hotspots in the pre-recession era, the number of hotspots almost doubled during the recession period, and the numbers remained in the similar range in the post-recession period. As the results show consistent trends for all the regions for the three time periods, therefore, from here on, the analysis will be more focused on the results obtained on this second variable.

The hotspot analysis helps to identify the spatial distribution patterns of the fatal crashes. If a comparison is drawn between Figures 4.4 and 6.6, the first point to realize is that although the number of deaths was higher for the time period before the Great Recession, the number of hotspots was almost half of the number of hotspots during the recession period. Therefore, the results suggest that the geographic locations of higher severity of fatal crashes were randomly distributed over the state of Texas before the Great Recession. During the recession period, while the number of fatal crashes decreased, the higher and lower severity of fatal crashes got more concentrated in some particular areas.

Further examination of the results in Tables 6.5 and 6.6 highlights that both the number of coldspots and hotspots have risen during the recession period. Figure 4.5 shows that the number of urban fatalities reduces drastically during the period of recession, and, therefore, it results in a sudden spike of around 2.5% in the lower band between zero and 1.2 as can be seen in Figure 6.4. These factors combined results in a larger spread of the coldspots region to include surrounding areas of the urban cities during the recession period as can be seen in Figures 6.10 to 6.14. On the other hand, Figure 6.4 shows no remarkable difference in the frequency distribution of death counts/AADT higher band greater than 99 between pre-recession and recession periods. A possible explanation for the doubling of the hotspots in the recession period as compared to the pre-recession period can be the concentration of high-value points in rural areas during the recession period resulting in the emergence of the new hotspot locations. Most of the hotspot locations remained consistent in the post-recession period.

Figures 6.15 and 6.16 show the distribution of the statistically significant points obtained for all the districts for rural and urban areas, respectively. When compared with Figures 4.6 and 4.7, the results seem to follow similar trends. For urban areas, the number of hotspots was highest in large cities, they were: Houston and Beaumont in the East, Dallas and Fort Worth in the North, Austin, San Antonio, Pharr in the South, and Odessa in the West region, following similar trends as the number of fatalities. There has been a gradual increase in the number of spatially statistically significant points over the years in most of the urban areas.

For rural areas also, Houston and Lufkin in the East, Tyler in the North, Austin, Pharr and San Antonio in the South, and Odessa in the West region are the main contributors to the hotspots and coldspots similar to the number of fatalities shown in Figure 4.7. For rural regions, it is interesting to note that 14 districts showed more than 100% growth in the number of statistically significant points, and three districts got new statistically significant points that were not present in the pre-recession era. Again, 17 districts recorded a decrement in the number of statistically significant points after the recession period. It can be concluded that a sudden spike in the number of rural hotspots and coldspots was recorded during the recession period.

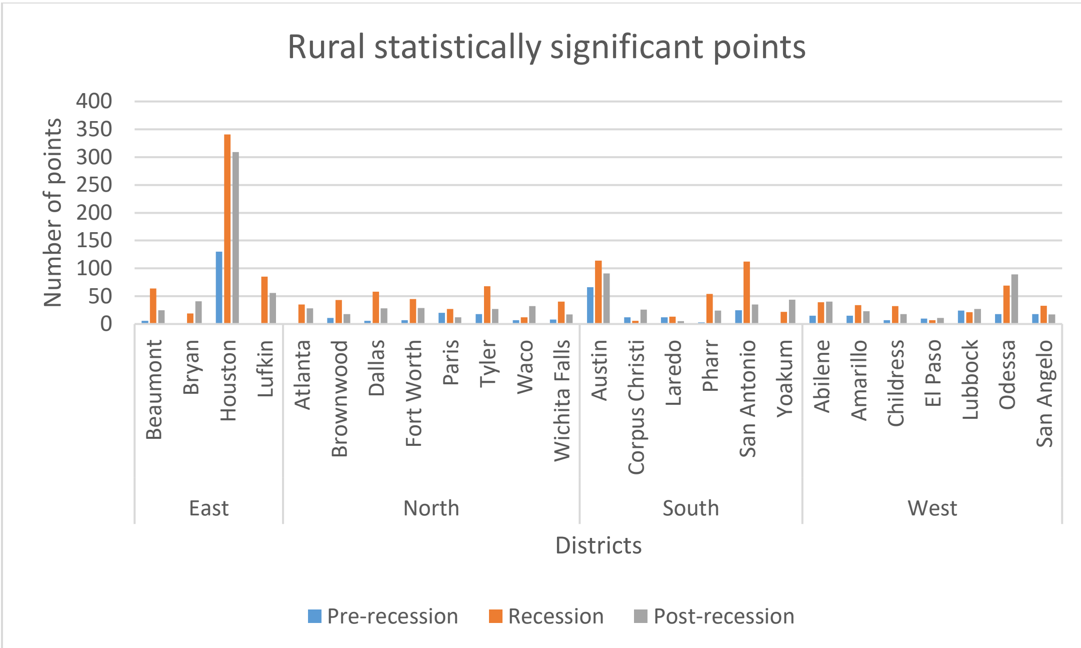


Figure 6.15: Number of statistically significant points present in the rural areas of each district

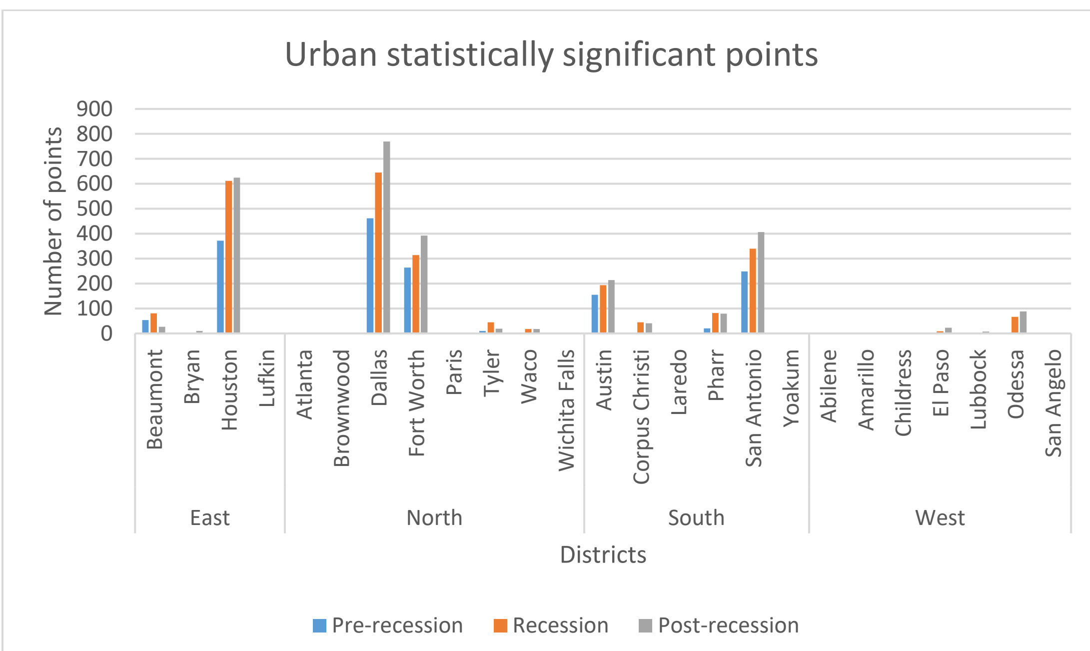


Figure 6.16: Number of statistically significant points present in the urban areas of each district

VI.4 Chapter Summary

This chapter has covered in detail the results obtained and drawn conclusions. The major points are as follows:

- Many fatal crash locations that did not have AADT value defined on the road segment ended up being removed from the dataset used for the analysis.
- The number of statistically significant points was independent of the number of input features used for hotspot analysis.
- The hotspots obtained from the death counts were inconsistent across the three time-periods and regions and, therefore, not reliable.
- The hotspots and coldspots obtained from the death counts/AADT were consistent and showed that the numbers had almost doubled during the recession period as compared to the pre-recession period. This meant that the higher and lower severity of the crashes were concentrated in certain areas during the recession period rather than randomly distributed across the areas as in the pre-recession period.
- The hotspots and coldspots remained similar in numbers in the post-recession periods as they were in -recession periods.
- The coldspots were present in the densely populated areas, and the hotspots were found predominantly in rural areas. The coldspots regions expanded and the new hotspots emerged during the recession period.
- The sudden spike in the number of statistically significant points was observed in the 17 districts out of 25 districts of rural areas while the number of statistically significant points gradually increased over the years in most of the urban districts during the recession period.

The next chapter provides a summary of methodology and results and discusses limitations.

CHAPTER VII

SUMMARY AND CONCLUSIONS

This chapter presents a summary and conclusions of the study related to the spatial analysis of traffic fatalities in Texas between 2003 and 2017. Before commencing this chapter, it is apt to have a brief look into all the chapters covered before. Chapter II delved into the literature present across the world related to the relationship between the number of traffic fatalities and the economic activities and covered the spatial analysis tools used besides. Chapter III cited various websites and the source of the data used in the study. Chapter IV provided preliminary analyses of the data to give a better understanding of the data. Chapter V looked into the spatial statistical analysis tools used to conduct spatial analysis. Chapter VI discussed the results and compared the results with the exploratory data analysis results to understand them better. Chapter VII provides the details of the limitations of the study and the future scope of the study.

Section VII.1 provides a summary of the work accomplished in this study. Section VII.2 lists the various limitations related to the data and methodology. Section VII.3 covers the scope of works that can be taken up in the future to investigate the results obtained in this study, or a similar methodology can be used on other data.

VII.1 Summary

This research study served as an extension of the study conducted by the NCHRP 17-67 and the study conducted by Shimu (2019). Various previous studies had indicated that the period of the economic downturn harmed the number of fatalities, which meant higher the unemployment rate, lower the number of crashes. However, no study was conducted in understanding how the regional factors or geographic locations of the crashes could affect the number of fatalities during the period of the Great Recession. This study aimed to understand spatial patterns associated with the change in the number of fatalities over the years. An extensive literature review first examined how the trends of traffic fatalities varied for different countries across the world during the period of economic recessions. Most of them concluded a positive relationship between the number of traffic crashes and economic activities. The main factors to cause this relationship were: fewer number of high-risk drivers (younger generation) on the road, less use of heavy trucks on the road, and fewer drink and drive cases.

Previous studies conducted to perform hotspot analysis, and Esri's ArcGIS website guided to develop the methodology for this study. Optimized Hotspot Analysis tool was identified to be used to primarily understand the spatial trends over the geography of the region and perform hotspot analysis. After the identification of the datasets required to perform spatial analysis, it was crucial to find out the sources of data. The police-reported crash data was collected from TxDOT's CRIS for the state of Texas to obtain details of fatal crashes for 15 years between 2003 and 2017. Then, the dataset was divided into three time periods of five years, pre-recession era, recession era, and the post-recession era. TxDOT's Open Data portal provided all the GIS files like AADT, roadways, counties, and districts data required for the study.

After collecting the data, the unnecessary columns present in the fatal crash data, which would not be used for any analyses, were dropped. Exploratory data analyses were conducted to understand the data better. The data were analyzed based on geographic locations, and Texas was divided into four regions, North, East, West, and South, and further, into 25 districts. The number of fatalities associated with each district and region and the patterns of urban and rural fatalities were obtained and analyzed. The data that did not have geographic locations and correct AADT values associated with each fatal crash location was removed to prepare it for hotspot analysis.

Now, the hotspot analysis was conducted on two variables: death counts and death counts/AADT values, and the results obtained were examined. The primary finding of the study was that despite the number of deaths reduced during the recession and increased after the recession period, the spatially statistically significant points (hotspots and coldspots) almost doubled during the period of the recession period and slightly reduced in the post-recession period. More detailed analysis is required to understand the proper reasons behind this phenomenon. Another vital point to note was that the results were independent of the number of input features.

The results can be used by the government agencies to understand and identify the change in spatial patterns during the period of economic recessions. The results will aid the government to allocate optimal funds to potential hotspots. As the crash fatalities tend to reduce during the recession periods, the government can divert a portion of the funds to other economic activities. The results can help the insurance agencies to identify the potential hotspot areas and charge the drivers accordingly.

VII.2 Limitations of the Study

There are various limitations associated with this study described as follows:

- The fatal crashes from the police-reported crash data that did not have the geographic locations had to be removed from the analysis.
- Although the number of locations where AADT was recorded increased over the years, still, the data did not cover all the roads and, therefore, could not be matched with all the

fatal crash locations. So, some fatal crashes which did not show the correct AADT values had to be removed from further analysis.

- Based on the optimal fixed distance band selected by the Optimized Hotspot Analysis tool, there should be at least eight neighbors for all the data points. However, due to the presence of locational outliers and the spatial distribution of the input features, it was not possible to have all input features having eight neighbors. In this study, roughly 13% of the input features had less than eight neighboring points.
- This study assigns correct AADT values to each fatal crash location based on the roadway dataset provided by TxDOT. So, the accuracy of this analysis is dependent on the quality of the GIS file of roadway data.

VII.3 Further Research

- The study focused on fatal crashes only and could be extended to include non-fatal injury crashes for the hotspot analysis.
- Various factors such as income levels, younger age group, and others, causing the change in the number and location of the hotspots during the three time periods need to be analyzed further. Also, the causes of the sudden change in the number of hotspots during the recession periods need to be examined more closely.
- The police-reported crash data have a large number of variables that were not used. They can be used to categorize the fatal crash, such as day and night time crashes, or the factors which caused the crash, such as the collision type (e.g., fixed object, overturning, etc.). A similar methodology can be developed to conduct the hotspot analysis.

- Many previous studies have investigated the factors associated with higher fatality rates in rural areas as compared to urban areas. Beck et al. (2017) found that more age-adjusted occupant death rates, lesser use of seat belts, and a higher percentage of unrestrained passenger deaths in the rural areas are the leading causes of higher fatality rates in the rural areas. However, further research is required to explain the variation in the number of urban and rural fatalities that occurred between the years 2003 and 2017.

REFERENCES

American Association of State Highway and Transportation Officials, Washington DC. (2010). Highway Safety Manual.

Antoniou, C., Yannis G., Papadimitriou, E., and Lassarre, S. (2015). Improving fatalities forecasting in times of recession in Europe. In *Why does road safety improve when economic times are hard?* (pp. 143-168). Paris: ITF/IRTAD.

Beck, L., Downs, J., Stevens, M., and Sauber-Schatz, E. (2017). Rural and Urban Differences in Passenger-Vehicle-Occupant Deaths and Seat Belt Use among adults- United States, 2014. *Morbidity and Mortality Weekly Report: Surveillance Summaries, Vol. 66, No. 17,* pp. 1-13.

Bergel-Hayat, R., Christoforou, Z., and Ferriere, S. (2015). The impact of the economic crisis on road mortality: an exploratory approach for some countries in Europe. In *Why does road safety improve when economic times are hard?* (pp. 169-182). Paris: ITF/IRTAD.

Bhat, C., Paleti, R., and Singh, P. (2014). A spatial multivariate count model for firm location decisions. *Journal of Regional Science, Vol. 54, No.3*, pp. 462-502.

Bil, M., Andrášik, R., and Sedonik J. (2019). A detailed spatiotemporal analysis of traffic crash hotspots. *Applied Geography*, pp. 82-90.

Blower, D., Flannagan, C., Geedipally, S., Lord, D., and Wunderlich, R. (2019). *Identification of factors contributing to the decline of traffic fatalities in the United States from 2008 to 2012.* Washington, DC: Final Report NCHRP Project 17-67, Transportation Research Board.

Business Cycle Dating Committee, National Bureau of Economic Research. (2010). Retrieved from https://www.nber.org/cycles/sept2010.html. Last accessed June 28, 2020

Chappelow, J. (2020). *The Great Recession*. Retrieved from Investopedia: https://www.investopedia.com/terms/g/great-recession.asp. Last accessed June 28, 2020

Cotti, C., and Tefft, N. (2011). Decomposing the relationship between macroeconomic conditions and fatal car crashes during the Great Recession: Alcohol and Non-Alcohol related accidents. *The B.E. Journal of Economic Analysis & Policy, De Gruyter, Vol. 11, No. 1*, pp. 1-24.

Elvik, R. (2015). An analysis of the relationship between economic performance and the development of road safety. In *Why does road safety improve when economic times are hard?* (pp. 43-142). Paris: ITF/IRTAD.

Esri Press Team. (2018). *How to Perform Spatial Analysis.* Retrieved from https://www.esri.com/arcgis-blog/products/product/analytics/how-to-perform-spatial-analysis. Last accessed June 28, 2020

Federal Highway Administration (FHWA). (2017). *KABCO Injury Classification Scale and Defintions.*

Forsman, Å., Simonsson, L., Wiklund, M., and Berg, Y. (2015). Analysis of fatal accidents during a period of severe economic recession in Sweden. In *Why does road safety improve when economic times are hard?* (pp. 183-193). Paris: ITF/IRTAD.

Generate Near Table (Analysis). (n.d.). Retrieved from https://pro.arcgis.com/en/pro-app/tool-reference/analysis/generate-near-table.htm. Last accessed June 28, 2020

Goodwin, G., Schoby, J., and Council, W. (2014). *A hot spot analysis of teenage crashes: an assessment of crashes in Houston, Texas.* College Station: Southwest Region University Transportation Center.

Hauer, E. (1995). On exposure and accident rate. *Traffic Engineering and Control, Vol. 36, No. 3*, pp. 134-138.

Hauer, E., and Hakkart, A.S. (1988). Extent and some implications of incomplete accident reporting. *Transportation Research Record 1185*, pp. 1-10.

He, Monica M. (2016). Driving through the Great Recession: Why does motor vehicle fatality decrease when the economy slows down? *Social Science & Medicine*, pp. 1-11.

How Hot Spot Analysis (Getis-Ord Gi) works*. (n.d.). Retrieved from https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm. Last accessed June 28, 2020

How Optimized Hot Spot Analysis Works. (n.d.). Retrieved from https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/how-optimized-hot-spot-analysis-works.htm#GUID-CDA6ACB4-C2CD-4B04-9B84-C408120D2313. Last accessed June 28, 2020

How Spatial Autocorrelation (Global Moran's I) works. (n.d.). Retrieved from https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm. Last accessed June 28, 2020

James, H. (1991). Under-reporting of road traffic accidents. *Traffic Engineering and Control, Vol. 32, No. 12*, pp. 574-583.

Kang, Y., Cho, N., and Son, S. (2018). Spatiotemporal characteristics of elderly population's traffic accidents in Seoul using space-time cube and space-time kernel density esitmation. *PLoS ONE, Vol. 13, No. 5*.

Keeley, B., and Love, P. (2010). *From Crisis to Recovery: The Causes, Course and Consequences of the Great Recession.*

Li, Y., and Liang, C. (2018). The analysis of spatial pattern and hotspots of aviation accident and ranking the potential risk airports based on GIS platform. *Journal of Advanced Transportation*.

Lloyd, Potter. (n.d.). Retrieved from Texas Demographic Center: https://demographics.texas.gov/Data/TPEPP. Last accessed June 28, 2020

Lord, D., and Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A, Vol. 44, No. 5*, pp. 291-305.

Maheshri, V., and Winston, C. (2016). Did the Great Recession keep bad drivers off the road? *Journal of Risk and Uncertainty, Vol. 52*, pp. 255-280.

Mannering, F., and Bhat, C. (2014). Analytic methods in accident research: Methodological frontier and future directions. *Analytic methods in accident research, Vol. 1*, pp. 1-22.

Narayanamoorthy, S., Paleti, R., and Bhat, C. (2013). On accomodating spatial dependence in bicycle and pedestrian injury counts by severity level. *Transportation Research Part B, Vol. 55*, pp. 245-264.

National Highway Traffic Safety Administration (NHTSA). (2017). *Model Minimum Uniform Crash Criteria Guideline Fifth Edition.*

National Highway Traffic Safety Administration (NHTSA). (2019). *Traffic Safety Facts 2017.* Washington, DC.

Near (Analysis). (n.d.). Retrieved from https://pro.arcgis.com/en/pro-app/tool-reference/analysis/near.htm. Last accessed June 28, 2020

Noble, B., Lloyd, D., Fletcher, J., Lloyd, L., Reeves, C., Broughton, J., and Scoons, J. (2015). Fatal road casualties in Great Britain: two studies relating patterns to wide-ranging exposure factors. In *Why does road safety improve when economic times are hard?* (pp. 194-227). Paris: ITF/IRTAD.

Shafabakhsh, G., Famili, A., and Bahadori, M. (2017). GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran. *Journal of Traffic and Transportation Engineering, Vol. 4, No. 3*, pp. 290-299.

Shimu, T. H. (2019). *Examining the factors causing drastic reduction and subsequent increase of roadway fatalities on United States Highways between 2005 and 2016.*

Texas Department of Transportation (TxDOT). (2018). *Comparison of motor vehicle traffic deaths, vehicle miles, death rates, and economic loss 2003-2018.*

Texas Department of Transportation (TxDOT). (2019). *TxDOT AADT Annuals*. Retrieved from http://gis-txdot.opendata.arcgis.com/datasets/txdot-aadt-annuals. Last accessed June 28, 2020

Texas Department of Transportation (TxDOT). (2020). *TxDOT Roadways*. Retrieved from https://gis-txdot.opendata.arcgis.com/datasets/txdot-roadways. Last accessed June 28, 2020

Texas Department of Transportation (TxDOT). (n.d.). *Districts*. Retrieved from https://www.txdot.gov/inside-txdot/district.html. Last accessed June 28, 2020

Texas Department of Transportation (TxDOT). (n.d.). *Open Data Portal: Your Gateway to TxDOT GIS Data*. Retrieved from http://gis-txdot.opendata.arcgis.com. Last accessed June 28, 2020

Texas Department of Transportation (TxDOT). (n.d.). *Welcome to the TxDOT Crash Query Tool*. Retrieved from https://cris.dot.state.tx.us/public/Query/app/welcome. Last accessed June 28, 2020

Wegman, F., Allsop, R., Antoniou, C., Bergel-Hayat, R., Elvik, R., Lassarre, S., Lloyd, D., and Wijnen, W. (2017). How did the economic recession (2008-2010) influence traffic fatalities in OECD-countries? *Accident Analysis and Prevention, Elsevier, Vol. 102*, pp. 51-59.

What is a z-score? What is a p-value? (n.d.). Retrieved from https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/what-is-a-z-score-what-is-a-p-value.htm#ESRI_SECTION1_2C5DFC8106F84F988982CABAEDBF1440. Last accessed June 28, 2020

What is GIS? (n.d.). Retrieved from https://www.esri.com/en-us/what-is-gis/overview#image1. Last accessed June 28, 2020

Wijnen, W., and Rietveld, P. (2015). The impact of economic development on road safety: a literature review. In *Why does road safety improve when economic times are hard?* (pp. 22-42). Paris: ITF/IRTAD.

World Bank. (2009). *Global Economic Prospects 2009: Commodities at the Crossroads.* Washington, D.C.

World Trade Organization. (2010). *World Trade Report 2010.*

Yamamoto, T., Hashiji, J., and Shankar, V. (2008). Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. *Accident analysis and prevention, Vol. 40, No. 4*, pp. 1320-1329.

Ye, F., and Lord D. (2011). Investigation of effects of underreporting crash data on three commonly used traffic crash severity models: multinomial logit, ordered probit, and mixed logit. *Transportation Research Record: Journal of the Transportation Research Board, No. 2241*, pp. 51-58.

Zhang, J., and Shi, T. (2019). Spatial analysis of traffic accidents based on WaveCluster and vehicle communication system data. *EURASIP Journal on Wireless Communications and Networking 2019, No. 1*, pp. 1-10.

APPENDIX A

DETAILS OF COUNTIES IN EACH DISTRICT

Table A-1 Details of all counties in each district of a region

Region	District	County Name and ID
East	Beaumont	Chambers (36), Hardin (100), Jasper (121), Jefferson (123), Liberty (146), Newton (176), Orange (181), Tyler (229)
	Bryan	Brazos (21), Burleson (26), Freestone (81), Grimes (93), Leon (145), Madison (154), Milam (166), Robertson (198), Walker (236), Washington (239)
	Houston	Brazoria (20), Fort Bend (79), Galveston (84), Harris (101), Montgomery (170), Waller (237)
	Lufkin	Angelina (3), Houston (113), Nacogdoches (174), Polk (187), Sabine (202), San Augustine (203), San Jacinto (204), Shelby (210), Trinity (228)
North	Atlanta	Bowie (19), Camp (32), Cass (34), Harrison (102), Marion (155), Morris (172), Panola (183), Titus (225), Upshur (230)
	Brownwood	Brown (25), Coleman (42), Comanche (47), Eastland (67), Lampasas (142), McCulloch (160), Mills (167), San Saba (206), Stephens (215)
	Dallas	Collin (43), Dallas (57), Denton (61), Ellis (71), Kaufman (129), Navarro (175), Rockwall (199)
	Fort Worth	Erath (72), Hood (111), Jack (119), Johnson (126), Palo Pinto (182), Parker (184), Somervell (213), Tarrant (220), Wise (249)
	Paris	Delta (60), Fannin (74), Franklin (80), Grayson (91), Hopkins (112), Hunt (116), Lamar (140), Rains (190), Red River (194)
	Tyler	Anderson (1), Cherokee (37), Gregg (92), Henderson (107), Rusk (201), Smith (212), Van Zandt (234), Wood (250)
	Waco	Bell (14), Bosque (18), Coryell (50), Falls (73), Hamilton (97), Hill (109), Limestone (147), McLennan (161)
	Wichita Falls	Archer (5), Baylor (12), Clay (39), Cooke (49), Monatague (169), Throckmorton (224), Wichita (243), Wilbarger (244), Young (252)

Region	District	County Name and ID
South	Austin	Bastrop (11), Blanco (16), Burnet (27), Caldwell (28), Gillespie (86), Hays (105), Lee (144), Llano (150), Mason (157), Travis (227), Williamson (246)
	Corpus Christi	Aransas (4), Bee (13), Goliad (88), Jim Wells (125), Karnes (128), Kleberg (137), Live Oak (149), Nueces (178), Refugio (196), San Patricio (205)
	Laredo	Dimmitt (64), Duvall (66), Kinney (136), La Salle (139), Maverick (159), Val Verde (233), Webb (240), Zavala (254)
	Pharr	Brooks (24), Cameron (31), Hidalgo (108), Jim Hogg (124), Kenedy (131), Starr (214), Willacy (245), Zapata (253)
	San Antonio	Atascosa (7), Bandera (10), Bexar (15), Comal (46), Frio (82), Guadalupe (94), Kendall (130), Kerr (133), McMullen (162), Medina (163), Uvalde (232), Wilson (247)
	Yoakum	Austin (8), Calhoun (29), Colorado (45), Dewitt (62), Fayette (75), Gonzales (89), Jackson (120), Lavaca (143), Matagorda (158), Victoria (235), Wharton (241)
West	Abilene	Borden (17), Callahan (30), Fisher (76), Haskell (104), Howard (114), Jones (127), Kent (132), Mitchell (168), Nolan (177), Scurry (208), Shackleford (209), Stonewall (217), Taylor (221)
	Amarillo	Armstrong (6), Carson (33), Dallam (56), Deaf Smith (59), Gray (90), Hansford (98), Hartley (103), Hemphill (106), Hutchinson (117), Lipscomb (148), Moore (171), Ochiltree (179), Oldham (180), Potter (188), Randall (191), Roberts (197), Sherman (211)
	Childress	Briscoe (23), Childress (38), Collingsworth (44), Cottle (51), Dickens (63), Donley (65), Foard (78), Hall (96), Hardeman (99), King (135), Knox (138), Motley (173), Wheeler (242)
	El Paso	Brewster (22), Culberson (55), El Paso (70), Hudspeth (115), Jeff Davis (122), Presidio (189)
	Lubbock	Bailey (9), Castro (35), Cochran (40), Crosby (54), Dawson (58), Floyd (77), Gaines (83), Garza (85), Hale (95), Hockley (110), Lamb (141), Lubbock (152), Lynn (153), Parmer (185), Swisher (219), Terry (223), Yoakum (251)
	Odessa	Andrews (2), Crane (52), Ector (68), Loving (151), Martin (156), Midland (165), Pecos (186), Reeves (195), Terrell (222), Upton (231), Ward (238), Winkler (248)
	San Angelo	Coke (41), Concho (48), Crockett (53), Edwards (69), Glasscock (87), Irion (118), Kimble (134), Menard (164), Reagan (192), Real (193), Runnels (200), Schleicher (207), Sterling (216), Sutton (218), Tom Green (226)