GEOGRAPHIC AND DEMOGRAPHIC PATTERNS OF ALCOHOL-RELATED FATAL TRAFFIC CRASHES: A SPATIAL-TEMPORAL ANALYSIS IN TEXAS, 1996-2005

A Thesis

by

GABRIEL A. ROLLAND

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

MASTER OF SCIENCE

December 2008

Major Subject: Geography

GEOGRAPHIC AND DEMOGRAPHIC PATTERNS OF ALCOHOL-RELATED

FATAL TRAFFIC CRASHES: A SPATIAL-TEMPORAL ANALYSIS IN TEXAS,

1996-2005

A Thesis

by

GABRIEL A. ROLLAND

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

MASTER OF SCIENCE

Approved by:

Chair of Committee,	D. Sui
Committee Members,	H. Liu
	T. Lomax
Head of Department,	D. Sherman

December 2008

Major Subject: Geography

ABSTRACT

Geographic and Demographic Patterns of Alcohol-related Fatal Traffic Crashes: A Spatial-Temporal Analysis in Texas, 1996-2005. (December 2008)

Gabriel A. Rolland, B.S., Texas A&M University

Chair of Advisory Committee: Dr. Daniel Sui

This thesis analyzes aggregated county-level data of fatal alcohol related traffic crashes where a driver was killed in the state of Texas during 1996 to 2005. Alcohol has constantly threatened drivers and passengers alike and continues to be a major cause of fatal crashes in Texas. Specifically, this paper targets those drivers that were killed while driving under the influence (0.01+ BAC). With an increase in manageable data and the ease of availability of aggregated crash records, accident analysis can provide a closer look into trends such as spatial-temporal patterns, clustering and correlations to various factors. Furthermore, Geographic Information Systems (GIS) have enabled researchers to more efficiently interpret and study a large amount of datasets using techniques that were previously difficult or inaccessible in applications related to traffic safety and transportation. Loose-coupling of GIS with other spatial analysis programs and/or statistical software packages can now provide important results that in turn relate vital information which can be used towards understanding and potentially alleviating problems in the transportation domain. The following sections concluded that aggregated datasets at the county level are currently incomplete and do not provide the

level of detail necessary to formulate a solid conclusion regarding relationships between the chosen factors and the crash dataset. Though this research was successful in mapping spatial variations and clusters, linking variables such as age, gender, location and population to the aggregated crash dataset requires more detailed information about the crash than was available. However, the objectives were successful in representing spatial-temporal patterns across the study period for all designated variables. This was an important step and solid contribution towards the representation of large datasets and their impact on policy, traffic safety, and transportation geography.

ACKNOWLEDGEMENTS

I would like to thank my committee chair Dr. D. Sui, as well as Becky Davies with the Texas Transportation Institute, Dr. H. Liu, Dr. T. Lomax and all of the other professors who dedicated much guidance and support throughout the course of this research. Over these past seven years I have had the opportunity of being able to build and enrich my knowledge through countless hours of teaching and lecturing by the professors of the outstanding department of geography at Texas A&M University. It was a pleasure and a privilege listening and learning from you all and I praise your knowledge and dedication towards teaching and research.

Thanks also go to my friends and colleagues, and to the department faculty and staff for making my time at Texas A&M University a memorable experience. Without their never ending support and guidance, this goal would have been out of reach. I also want to extend my gratitude to the National Highway Traffic Safety Administration, which provided the bulk of the data supporting this project and which continues to strive to improve safety on the roads.

Finally, I cannot thank my parents, brothers and family enough for their encouragement. Their patience, wisdom and love have always made it possible for me to pursue my goals and aspirations and I look forward to making them proud in the actions I have and will undertake in life.

NOMENCLATURE

GIS	Geographic Information System
BAC	Blood Alcohol Content (as a percentage)
NHTSA	National Highway Traffic Safety Administration
FARS	Fatality Analysis Reporting System
LISA	Local Indicators of Spatial Autocorrelation
IDW	Inverse Distance Weighting
TTI	Texas Transportation Institute
DBMS	Database Management Systems
GUI	Graphical User Interface
HDE	Heavy Drinking Episodes
MADD	Mothers Against Drunk Driving
NSCA	National Center for Statistics
SDC	(Texas) State Data Center

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION: TRAFFIC SAFETY

1.1 TRAFFIC SAFETY OVERVIEW

We have all experienced the horrors of alcohol-related traffic accidents. Some of us have been unfortunate witnesses to mingled steal and broken glass, while others have had the inopportune mishap of becoming a part of the physical, mental, and sometimes fatal event. We live in an era where employing a mode of transportation is seldom a choice, but often an obligation to fulfill the important needs of an increasingly mobile society. Unfortunately, though the detrimental effects alcohol imposes on drivers and passengers are proven, fatalities and crashes persist to occur at alarming rates. Transportation systems, researchers, and organizations are continually searching for ways to improve transportation while reducing accidents by employing an extensive range of approaches and available resources. Situated in the literature and context of traffic safety, this paper provides a study concerned with the extent particular demographics affect fatal alcohol related traffic crashes at the county-level for the state of Texas. Through the use of GIS, spatial autocorrelation techniques, and aggregated datasets at the state and county level, the following sections will study interpolation results as well as spatial clustering, graphs, tables and choropleth maps in an attempt to identify problematic areas and possible correlations.

This thesis follows the style of *Journal of Transport Geography*.

The U.S. Department of Transportation has reported that alcohol-related traffic crashes for the state of Texas have been declining, showing a significant decrease from 52.34% of total crashes resulting in fatalities for 1996 to 45.07% in 2005 (see Table 1). In order to try and reduce crash rates and occurrences, accident analysts and researchers from a variety of fields and backgrounds have attempted to study and relate many variables connected to transportation datasets. To productively understand these events, analysts typically look at five major factors that tend to influence traffic crashes: Driver behavior, vehicle types, roadway conditions, traffic characteristics, and environmental factors (Li et al., 2007). From these five major factors, this paper will focus on driver behavior and performance; specifically gender, age, population, and location.

Table 1. Total Fatal Crashes in Texas. Note the decrease in alcohol related crashes and high, average, low values respectively of 96', 00', and 05'.

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	Avg.	Total
Total Crashes	**3249**	**3084**	**3167**	**3108**	**3254**	**3316**	**3389**	**3368**	**3266**	**3095**	**3229.6**	**32296**
Crashes +BAC	**1701**	**1503**	**1546**	**1511**	**1585**	**1596**	**1602**	**1554**	**1519**	**1395**	**1551.2**	**15512**
% +BAC	**52.34**	**48.72**	**48.82**	**48.62**	**48.72**	**48.13**	**47.27**	**46.15**	**46.52**	**45.07**	**48.03**	

Researchers involved in accident analysis have recently begun to successfully integrate and apply Geographic Information Systems as a means of evaluating, processing, and displaying crash datasets (Pawlovich et al., 1998; Thill, 2000; Li et al., 2007; Loo, 2006). Today, several procedures and algorithms can be implemented within a GIS to produce several different types of results including interpolation of location-based point data, clustering through univariate and bivariate local indicators of spatial autocorrelation (LISA & BiLISA), and hot-spot analysis. The loose-coupling between a GIS with well-established statistical procedures and standalone spatial statistics software packages is a promising advancement for accident analysis and the hopeful prevention of future traffic crashes.

Geography plays a major role in contributing its relevance and contributions to traffic safety. Nearly every aspect of transportation is linked through space and time. This is not only important as a basic element to traffic safety, but also as a crucial study area in geography whereby complex spatial systems require an understanding about the movement of goods, people and goods across space and time. The geography of transportation is especially important to public transportation where spatial constraints and attributes with the origin, the destination, the extent, and the nature and purpose of the transport are directly linked. This research falls into the sub discipline of transport geography, which basically strives to understand the spatial relationships of particular events as they relate to transport systems. Essentially, geographical knowledge is imperative to traffic safety.

1.2 OBJECTIVES AND AIMS

This paper explores several methods related to crash count and rate analysis in an attempt to eventually fill-in or explore several gaps found in the related literature. The knowledge gaps are empirical since this particular dataset coupled with GIS and spatial statistics research has not been conducted at the county level for the state of Texas. Moreover, there is no apparent attempt at correlating the specified independent variables and their characteristics to alcohol-related driver fatalities at this designated scale (county level) and location. This research will provide information on whether or not these independent variables can be connected to alcohol-related traffic crashes temporally, statistically, and spatially by specifically focusing on Texas county level data.

The purpose of this research is to relate geographic information about the extent to which different demographic variables can relate to fatal alcohol related traffic crashes where a driver was killed with either a BAC of ≥ 0.01% or 0.00%. Table 2 provides a basic layout as to how the format of these objectives will be carried out. Specifically, these objectives will be accomplished by evaluating an extensive collection of aggregated data compiled from several sources. Through graphs, tables, and statistical relationships, specifically spatial autocorrelations using Moran's I & LISA, factors including age, gender, location and population will be evaluated for any patterns and correlations. A GIS will then be employed as a means towards the identification and production of the following: trend overtime and spatial patterns from 1996-2005, choropleth maps, and interpolated county centroids of dependent/independent variables. Finally, the results will be discussed in both the context of this paper as well as how they relate to previous studies discussed in the literature review chapter.

The results are predicted to demonstrate that age, gender, location, and population have an impact on fatal alcohol related crashes which can be identified and visually interpreted using a GIS and spatial autocorrelation at that level of aggregation. However, the extent to which these variables can be used to explain fatal crashes without knowing who was involved in the dataset may prove to be challenging and difficult to relate. Also, trends may provide some insight on declining crash rates in Texas. If these variables are found to be informative or of good use at highly aggregated levels, the methodology could be beneficial as a background or general study of county trends in Texas. Ultimately, this research attempts to provide another exploratory analysis to be used towards taking more refined precautionary measures, or at the very least, provide information regarding crash rates and trends in Texas.

Table 2. Summary of Research Objectives.

	Objectives:
1.	Evaluate crash, demographic, and population data using tables and graphs
2.	Explore Local Indicators of Spatial Autocorrelation using GeoDa on the same dataset
3.	Investigate spatial and temporal patterns using GIS through Interpolation and Choropleth mapping

CHAPTER II

LITERATURE REVIEW

2.1 TRAFFIC SAFETY AND ALCOHOL RESEARCH

Traffic analysts have studied a vast range of topics related to transportation systems in general. The heterogeneous nature of traffic crashes have lead scientists towards various matrices of research concerned with temporal, environmental, locational and maneuvering circumstances (McGwin and Brown, 1999). To name a few, these have included studies relating traffic safety to drug use, Socio-economic or demographic characteristics, alcohol or drunk-driving, and spatial or temporal patterns. This research not only attempts to improve the safety of American drivers and others around the world, but also strives to assist or persuade transportation engineers, policy makers, and government organizations dealing with transportation issues to make efficient and positive changes. Furthermore, the conclusions revolve around the common understanding and general acceptance that reducing any number of these degrading variables or implementing and enforcing appropriate laws may alleviate traffic crashes and improve safety. The broad range of topics studied under the common title of crash, accident, or traffic analysis gives significant evidence that crashes are a complex and multifaceted topic. Typically, traffic analysts employ methodologies and basic principles from physics, engineering, medicine, psychology, behavioral science, law, mathematics, logic, and philosophy (Evans, 2004). Though crashes are one subject of concern in the transportation domain, they are one of the more costly in terms of capital and personal losses, which makes analyzing their patterns, factors, and complexity an integral part of transportation studies. As population and road networks continue to grow to meet strenuous demands, so too will the likelihood of crash occurrences unless researchers and engineers maintain and improve their push for more efficient systems and preventative measures.

Correct terminology in traffic analysis and traffic safety is imperative as terms are sometime misused or ambiguous by nature. The following section will define important key terms, from Leonard Evan's book titled "Traffic Safety" (2004), which have been or will be discussed throughout this research. First and foremost, crashes are often defined as any vehicle with an engine striking anything along public roads. Amongst total crashes (fatal & non-fatal), fatal crashes are few in numbers and vary accordingly when dealing with multiple deaths within single vehicle, multi-vehicle crashes, and total distance traveled. Crashes are often further broken down by single, multi-vehicle, driver, passenger, gender, age, and roll-over incidences, to name just a few. In turn, records kept, such as those available from the FARS database, are explicit by default so that proper factors may be acquired for any given study. The term accident, though popular and more commonly employed amongst the general public when referring to a definition similar to that of a crash, is harder to define and frequently misused since it conveys a sense of fate and unpredictability. Fatal crashes or fatalities, frequently defined as crashes directly resulting in death within 30 days of the event, are more often employed in research than injuries or damages as they are straightforward (loss of life) and commonly well-documented. Non-fatal injuries, which comprise a wide range of definitions ranging from non to near death, are numerically categorized and harder to define compared to the ‘yes or no’ definite determination of fatal crashes. Employing the word ‘factor’ instead of ‘cause’ is also important in traffic analysis as ‘cause’ often appeals to a singular reason as to why an event may have happened. In reality, crashes commonly come as a result of several factors which typically work with or against one another, resulting in a specific event. As Table 3 portrays, even just a few factors grouped under one table can often be confusing, misleading, or even of similar definition to unfamiliar eyes.

Table 3. US Fatal Crashes in 2002. Displaying Differences Between Crashes, Fatalities, Survivors, and Involvement (Evans, 2004).

US fatal crashes in 2002:		
Quantity	Number	Ratio
Fatal crashes	38,309	
Fatalities	42,815	
Fatalities per fatal crash		1.12
Involved vehicles	58,113	
Vehicles per fatal crash		1.52
Total involved people	101,195	
Survivors of fatal crashes	58,380	
Survivors per fatal crash		1.52

The following chapter will review the literature concerned with the wide range of topics relevant to traffic analysis and pertaining to this particular research. The review is situated across various fields and topics concerning health, alcohol, safety, transportation, statistics, and behavioral studies. Commonly, these topics are found in journals including Accident Analysis and Prevention, Journal of Transport Geography, Journal of Safety Research, Journal of Studies on Alcohol, Computers, Environment and

Urban Systems; these topics are also the subject of several proceedings, conferences, and reports.

The use and abuse of alcohol is well documented across time as being detrimental when consumed with little regard to moderation. Though debatable, it is possible that one of the very first fatal crashes recorded might have been related to some alcohol consumption. The crash occurred in Harrow near London, on the 25th of February 1899, and it was believed that the driver who died on impact had consumed alcohol prior to the crash (Evans, 2004). Although alcohol was believed to have been a factor, it was concluded that the breaks had ultimately failed to stop the vehicle. Witnesses reported that the breaks had caused severe damage to the tires, which in turn led to a violent deceleration and eventual ejection of the driver.

A major problem with alcohol lies in the intertwined use of intoxicating beverages and the unreasonable decision to operate a motor vehicle before attaining a more appropriate BAC or finding a safer mode of transportation. Though the dangers of drinking and driving are understood by most, the decision can vary greatly across certain populations based on several factors. Certainly, the effects of alcohol on drivers are well established in the literature as being highly influential in promoting increased crash rates as opposed to sober drivers, with estimates that a driver with a BAC in excess of 0.08 % could be 50 times more costly on the road. (Miller et al., 1999; Smink et al., 2005). These statements are backed by the simple fact that, in dealing with traffic safety and alcohol, no single factor has led to more research, controversy, passion, and papers being produced (Evans, 2004). Nonetheless, even though it is also commonly known by the general public that alcohol consumption and abuse by drivers has been found to dramatically increase accident likelihood and severity, some drivers continuously make the decision to drink and drive (Holubowycz et al., 1994).

Alcohol-related studies in the transportation domain have attempted to link various independent variables to crash data in attempts to better understand factors for hopeful prevention. Some of these studies have sought to relate alcohol outlet locations to accident sites (Stevenson et al., 1998), BAC levels and their effect on drivers, passengers, and pedestrians (Holubowycz et al., 1994), and demographic/socio-economic characteristics ranging from gender and ethnicity, to poverty and unemployment, and the increase or decrease in alcohol consumption (Mercer, 1987; Holubowycz et al., 1994). All of these studies have reported a variety of results commonly concluding and accepting that there is a need to reduce alcohol-related traffic crashes by further studying, understanding, and validating the aforementioned variables and methodologies. The independent variables of concern at various geographic scales, which are discussed with more detail in forthcoming paragraphs, more often than not can be related to alcohol crashes and fatalities in some way.

Alcohol consumption specifically affects both a driver's behavior and performance in ways which can drastically increase an individual's risk on the road, while decreasing the likelihood of survival in a crash. Specifically, it has been estimated and accepted that alcohol-impaired drivers are 3.85 times more likely than sober drivers to die in crashes of comparable severity (Evans, 2004). Of course, the type of behavior and the decrease in performance varies according to an individual's weight, experience, gender, gastrointestinal content, comportment and several other factors. Commonly, drivers will express decreases in performance through reduced reaction times, vision, and other cognitive ailments. These decreases in performance are well-documented through various tests and studies where quantitative data has been produced as scientific evidence. Alongside performance, and equally as important although harder to quantify, behavior becomes problematic when normal decisions are inhibited (decision to drive after drinking), judgment is altered, and aggressiveness is increased, to name just a few (Evans, 2004). For further details entailing alcohol consumption and the resulting effects on performance and behavior, see Table 4.

Table 4. BAC and Effects on Behavior and Performance (Evans, 2004)

BAC %	Performance and Behavior Changes
0.01%	Normal actions, hardly influenced
0.02%	Changes in social behavior, mild euphoria, relaxation, increased gregariousness
0.05%	Feeling good, **less inhibited**, **altered judgment**, lowered alertness
0.07%	Judgment impaired, **likely to take risks and actions not taken when sober**, **release of inhibitions, impulsive behavior,** slight decrease in fine motor skills, **more bravado**, and less restraint for other behaviors such as eating, smoking, gambling, and drugs. Mood tends to shift from positive to negative
0.10%	Slower reaction times and impaired motor functions, **less caution**, slightly slurred speech, **increased aggressiveness**
0.20%	Major memory impairment -- "blackout" a possibility
0.27%	Confusion, staggering, slurred speech
0.30%	Double vision may occur, most drinkers become unconscious or fall asleep at this level and are difficult to awaken
0.40%	Barely conscious
0.45%	Death very likely

According to the Fatality Analysis Reporting System (FARS), alcohol related traffic accidents have generally been declining over the past several years, starting from 1982 (see Table 5). Though these numbers are reassuring and perhaps demonstrate that safety measures, vehicle safety, education, and road infrastructures are becoming more efficient, Texas consistently remains a top contender for alcohol-related traffic incidences. In March of 2008, the National Highway Traffic Safety Administration's National Center for Statistics and Analysis (NHTSA) released its 2006 Traffic Safety Facts focusing on Alcohol-Impaired Driving. Nationally, it was reported that 13,470 people were killed in alcohol-impaired-driving crashes (BAC $\geq$ 0.08%), or 32% of total motor vehicle traffic fatalities. According to the NHTSA review, Texas accounted for nearly 10% of all traffic fatalities which involved any alcohol (1544/15827=9.76%), and reported higher percentages of alcohol-related traffic fatalities than the national average for BAC categories of $\geq$ .01% (45% versus 37%), $\geq$ .08% (39% versus 32%), and $\geq$ .15% (26% versus 21%).

Table 5. Traffic Fatalities in Texas which were Alcohol-Related (FARS).

Fatalities in Texas

Year	Tot	Alc-Rel	%	0.08+	%
1982	4,213	2,801	66	2,570	61
1983	3,823	2,503	65	2,311	60
1984	3,912	2,457	63	2,232	57
1985	3,678	2,271	62	2,021	55
1986	3,567	2,206	62	1,932	54
1987	3,260	1,951	60	1,688	52
1988	3,392	2,011	59	1,775	52
1989	3,370	1,927	57	1,729	51
1990	3,250	1,989	61	1,769	54
1991	3,078	1,814	59	1,604	52
1992	3,059	1,818	59	1,624	53
1993	3,043	1,748	57	1,567	52
1994	3,187	1,725	54	1,546	49
1995	3,183	1,739	55	1,534	48
1996	3,742	1,967	53	1,745	47
1997	3,513	1,710	49	1,521	43
1998	3,586	1,745	49	1,550	43
1999	3,522	1,700	48	1,479	42
2000	3,779	1,841	49	1,642	43
2001	3,736	1,807	48	1,587	42
2002	3,823	1,810	47	1,610	42
2003	3,675	1,709	47	1,500	41
2004	3,583	1,642	46	1,417	40
2005	3,504	1,569	45	1,371	39
2006	3,466	1544	45	1,354	39

Successful reductions in alcohol related fatalities have been attributed to the enforcement of various laws, organizations, and campaigning. As was discussed in previous paragraphs, drunk drivers have plagued the roads since the introduction and popularization of the automobile. As a result, some of the earliest known traffic laws and punishments regarding drinking and driving simply, and often unsuccessfully, involved enacting various criminal laws (Evans, 2004). Until the 1930's, traffic violations involving alcohol were ambiguous by nature as they often relied on an officers judgment, which was frequently insufficient and in most cases hard to uphold in courts.

Fortunately, the invention of the breathalyzer enabled officers to better estimate and document BACs which in turn produced new data and legislations. In 1936, Norway became the first country to uphold *per se* laws stating that no individual with a BAC greater than 0.05% could operate a motor vehicle (Evans, 2004). Other countries followed quickly and evidence of this move is alive and well in the US where it is considered illegal to operate a vehicle with a BAC greater than or equal to 0.08%.

The *per se* laws, which established thresholds where someone could be held accountable were they recorded at or above a certain BAC, were arguably successful at reducing fatalities and infractions. An evaluation of the 0.08% law in Texas by Gorman et al. found no statistical evidence that the law reduced alcohol-involved traffic crashes or fatalities. Gorman et al. studied several groups including race, gender, age and location (urban versus rural). They concluded that since the law was introduced in Texas on September 1st, 1996, a clear reduction in alcohol involvement was inconclusive. Unfortunately drivers overtime become desensitized by the notion that police officers are often few and far on the roads. Furthermore, word-of-mouth about ways to act properly or drive correctly if sighted or pulled over leads to questions about whether this law is enough to deter drinking and driving. In turn, it has been found that the key to maintaining low numbers of infractions would be to increase the belief that the probability of detection was high. Today, drunk drivers have approximately a 1/2000 chances of getting pulled over (Evans, 2004).

Random testing, Alcohol Ignition Device installation, zero-tolerance, Wet versus Dry counties, and several other laws were introduced following the aforementioned introduction of breathalyzers. Perhaps one of more important laws was the introduction of the minimum drinking age law, setting the minimum age at which a person could purchase alcohol to 21. The minimum age law was found by many to have significantly reduced the various negative effects associated with consuming alcohol and operating a vehicle at the relatively immature, and more importantly inexperienced, ages below 21 (Winn and Ciacopassi, 1993; Wagenaar et al., 2002; Evans, 2004). Another way that public awareness has increased has been attributed to the establishment of citizen organizations in the 1980's, such as Mothers Against Drunk Driving (MADD), which have consistently and tirelessly helped to raise public awareness through, for example, campaigning or even helping to persuade convictions of drunk drivers in court cases. Citizen organizations, alongside laws and strict enforcement, have positively helped to increase public awareness and belief that drinking and driving is not worth taking the risk.

Ultimately, reducing the various negative impacts alcohol imposes on traffic safety comes from a synthesis of efforts from all levels of government, organizations, businesses and corporations, and the individual driver. Though many of these efforts have been in place and continue to be enforced, there is room for improvement in regards to alcohol laws. According to Leonard Evans (2004), a national decrease in alcohol consumption would inarguably lead to fewer traffic deaths and related problems. More specifically, a national decrease in consumption would be accomplished by decreased advertising, increased prices and taxes through the Federal Excise Tax, decreased availability, and increasing sobriety lane checks and enforcement. Enforcing all of these, Evan argues, is not only possible and reasonable, but also well in reach and quite straightforward to attain and enforce.

2.2 GIS AND SPATIAL STATISTICS IN TRAFFIC SAFETY

Over the years, accident analysis has consistently transformed itself into a more technical and cutting-edge discipline in terms of statistical analysis and application, and through the adoption of various Geographic Information Systems that have become widely available and accepted (Kim et al., 1996; Thill, 2000; Li et al., 2007; Loo, 2006). Though GIS has been around since the early 1960's, it was not until the late 1980's that it became a popular tool for transportation research and management (Thill, 2000). The early 1990's were critical in that the enactment of a wide variety of federal legislations were introduced in hopes of changing the outlook of a previously uniform and homogeneous transportation domain, into a multi-disciplinary field involving neighboring sciences and associated domains. As is explained by Thill in a paper written for Transportation Research regarding GIS and transportation (2000), it can be argued that the present adoption of GIS in transportation brings the field to a full circle as it is rediscovering the primacy of space and place, two concepts that launched the systematic study of transportation in Geography and Regional Science in the 1950's.

In essence, the framework behind a GIS is not new to the transportation domain. Database Management Systems (DBMS) had been employed some time before GIS had been introduced and accepted as an integral component towards advancing research in transportation. For the most part, these DBMS had been implemented to store, reference, and carry out relatively simple studies on collected transportation datasets.

By contrast, GIS is all about the geographic description of the earth's surface, bringing forth the important element of time and space with thorough analytical and modeling capabilities (Thill, 2000). GIS in the transportation domain (GIS-T), as defined by Michael Goodchild who is commonly accepted as a pioneer in GIS literature, implements three classes of models including field, discrete, and network models. Though these are often used in other fields and applications, they are most important to transportation for continuous variation over space and time, accurate locations of static places (tolls, highway rest areas, traffic signals), and topologically connected and fixed linear entities such as roads (Goodchild, 1992; Thill, 2000).

As a result, interest in GIS from transportation engineers and researchers has paved the way towards the adoption and creation of several models, concepts, and important methodologies. For example, analysts have successfully produced various maps and projects depicting "hot-spots" or producing "black zone" analysis (Levine et al., 1995; Flahaut et al., 2003). This combination of maps, statistics, and tables identify clusters which are defined as a spatial pattern that differs in important respects from the geographic variation expected in the absence of the spatial processes that are being investigated. These clusters are commonly linked to spatial autocorrelation, namely when an above-average value in a location is surrounded by neighbors whose values are also above-average (high-high), or when a below average value is surrounded by neighbors with bellow average values (low-low). The opposite combinations are defined as negative spatial association, where a high (above-average) value is surrounded by low neighbors and vice versa (Messner & Anselin, 2004). As will be discussed in more detail in the methodology chapter of this paper, spatial autocorrelation and the identification of clusters is important in attempting to relate indirect independent variables to crash datasets where correlations across space and time are unknown.

Another reason the adoption of GIS and creation of GIS-T has flourished amongst transportation researchers has stemmed from the flood of spatial data available today. More importantly, this spatial data has been implemented, for example, in validating crash locations, locating motor vehicle crashes, modeling spatial-temporal patterns of crashes, and modeling highway geometry for traffic safety (Kim et al., 1996; Levine et al., 1998; Loo, 2006; Li et al., 2007). As is discussed by Levine et al., 1998, one of the more successful and primary uses of a GIS is the assigning of geographic coordinates to features that are not directly referenced by a geographic base map, commonly referred to as Geocoding. Though this feature is popular in domains other than transportation, it has nonetheless been appealing to this field for several reasons including representation of data, ease of calculating distances, buffers and the ability to store and retrieve this same data efficiently.

Though the consequences brought on by the adoption of GIS to the transportation domain have mostly been positive in nature, GIS has for the most part been used as a tool rather than treated as a science. In turn, several papers and discussions have been raised regarding the push for more comprehensive uses of GIScience in transportation (Goodchild, 2000). Furthermore, since GIS projects and results have been heavily dependent on data availability, several issues and challenges have been raised in discussing scale and accuracy. Concerning scale, the issue of the modifiable aerial unit problem (MAUP) has plagued studies dealing with aggregated data and the use of interchangeable boundaries. The problem lies in the lack of understanding that, although every level of aggregation is important, issues closely related to the ecological fallacy can arise if a background on scale and its effect on spatial data are not well understood (Fridstrom et al, 1991; Thomas, 1996). Simply put, an ecological fallacy is committed when a relatively smaller aggregated dataset of samples is assumed to apply to a larger population. As an example related to this study, stating that a county 'X' has more drunks because of higher crash incidences would be an ecological fallacy. A relatively high occurrence of fatal alcohol crashes does not necessarily extrapolate or reflect an entire counties drinking behavior. Furthermore, if a survey was conducted on a few individuals regarding their drinking and driving behavior, their results could not account or be applied to the rest of the population residing in the same county.

Accuracy has also been an issue in transportation studies that have relied on a variety of spatial datasets (Kim et al., 1996; Flahaut et al., 2003; Loo, 2006). Despite the fact that fatality datasets constitute some of the most reliable safety data, there is a need to be vigilant about any dataset which is brought into a GIS (Evans 2004). Even as data standards have been in place for some time, accident analysts often rely on data collected by several agencies which in turn may not reproduce or keep up data that would be up to a researcher's standard of quality. Since crash data, even data kept and maintained by FARS, is often first recorded on forms by police officers who are not necessarily trained or required to reproduce the necessary level of detail needed to most scientific studies, analysts are often obliged to rely on aggregated data with limited variables. Until all vehicles are equipped with, for example, a system of collision detection that records the time and location of first impact, crash data will have to mostly rely on incomplete and often erroneous police records (Kim et al., 1996).

Overall, maps and associated results stemming from the adoption of GIS have given a clearer insight to various statistics that would otherwise be difficult to comprehend due to the nature of the copious datasets available in accident analysis. In summation, GIS enables researchers to link crash data with travel information, land use, and social-economic information to better capture the relationship between crash occurrence and contributing factors (Li et al., 2007).

Although traffic fatality counts have, for the most part, increased since the early 1900's and had peaked by 1972 (54,589), Figure 1 portrays that their display in a table with data appearing in its original format (counts) has typically been difficult to interpret and draw conclusions from. For this reason, traffic analysts have employed ratios and normalized fatality counts to try and make sense of copious amounts of recorded data (rates). As an example, Figure 2 represents those same traffic fatality counts per billion km since 1921. In comparison to Figure 1, this table clearly indicates a downward trend (3.5% a year) in overall fatalities for the US. Though crash analysis is ultimately dependent on a count of some sort, analysts and researchers commonly apply a wide range of statistical and mathematical measures to those same raw variables in order to better understand the meaning of these events.

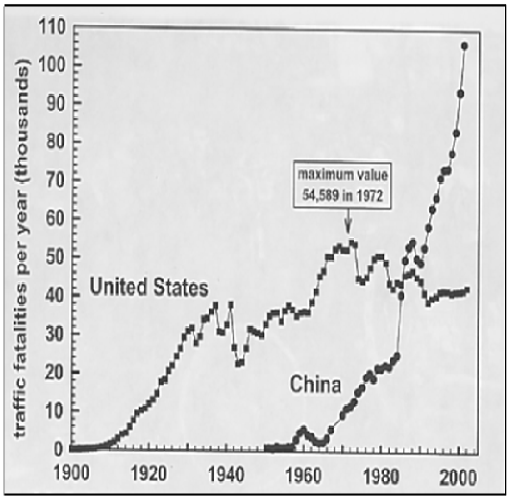

Figure 1. Traffic Fatalities since 1921.
(Evans, 2004)

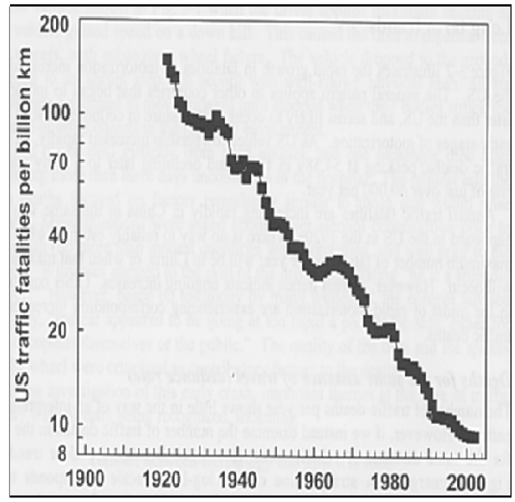

Figure 2. Traffic Fatalities/Billion km since 1921
(Evans, 2004)

Aside from tables and graphs, traffic analysts have used various methods to analyze a large variety of datasets available at different spatial scales. These statistical methods range from the Bayesian analysis (Hauer, 1992; Li et al., 2007) to Poisson regression, Analysis of Variance (ANOVA), linear regression, LISA, Moran's I, etc. Commonly, authors working with multivariate statistics, such as demographics and socio-economics, employ a regression or correlation method to try and understand an association between two or more different datasets (Mercer, 1987; Hauer, 1991; Holubowycz et al., 1994; Pawlovich et al., 1998).

Spatial statistics have also played an important role in attempting to further comprehend crash rates and imposing factors. Spatial statistics differ from the traditional practices and calculations in the adoption and intertwined use of formal techniques with topological, geometric, or geographic properties. Common methods include spatial autocorrelation, which was discussed in the GIS section of this literature review, spatial interpolation, and spatial regression. Although these techniques are complicated and require extensive computation, they are commonly executed in GIS software packages that can compute, manage, and display results efficiently. Furthermore, several standalone software developers have developed freeware packages which can carry out several important spatial statistical computations. Specifically, GeoDa (developed at the Spatial Analysis Lab, University of Illinois) has become a popular tool for visualization, exploration, and explanation of patterns in geographic data (Messner & Anselin, 2004). GeoDa's user-friendly graphical user interface (GUI) assumes that the user is unfamiliar with GIS and at an introductory level in regards to more complex statistical procedures, making it appealing to a wide variety of applicants wishing to compute rigorous spatial statistics. Together, though still considered to be loosely-coupled, the combination of these aforesaid softwares, techniques, and methodologies help to validate, assess, and even predict various traffic issues.

2.3 LOCATION, GENDER, AND AGE RESEARCH IN TRAFFIC SAFETY

With a current population of just over 30 million people and an estimated 1.9 vehicle per person, the US provides more traffic data than any single nation. However, researchers located in the accident analysis literature and closely linked fields have considered a wide variety of study sites in conducting their research. Some of these studies have produced several projects in Australia (Holubowycz, 1994), Honolulu (Levine et al, 1995), Hawaii (Kim et al., 1996), southern Georgia (Stevenson et al., 1998), French counties (Amoros et al, 2003), and Harris County or the city of Houston (Li et al, 2007). It was found that, for the most part, these authors chose the abovementioned locations for various reasons, though most were originally from those locations or had access to familiar data. Other reasons for choosing specific study sites included interests in Wet versus Dry County effects on alcohol-related vehicle crashes (Stevenson et al., 1998), exclusion of loosely defined spatial boundaries resulting from a study site being enclosed on an island (Levine et al., 1995), and accurate spatial distribution of crashes resulting from accurate reporting of crash occurrences at 100m interval markers on highways (Thomas, 1993).

Although study sites are chosen for various motives, specific results from these studies and a variety of others have established links between the locations of fatal crash occurrences and other factors. For example, several studies have concluded that the severity of a crash can often be attributed to its location in an urban versus rural environment, and to population density (Winn and Ciacopassi, 1993; Clark, 2003; Evans, 2004). More specifically, a study on the effects of population density on mortality after vehicle collisions found that, even after correcting for age, speed, and seatbelt-use, accidents outside of urbanized locations more often resulted in deaths (Clark, 2003). Reasons for attributing lower fatal crashes in urban areas have been correlated to lower speed limits, congestion, better lighting, better public transportation, shorter distances being traveled, and more efficient medical care or proximity to medical services (Winn and Ciacopassi, 1993; Evans, 2004). By comparison, rural areas have been shown to produce a greater number of fatal crashes due to opposing parameters relatively to those previously mentioned factors.

When comparing males versus females, studies tend to focus primarily on specific factors associated to behavior or performance. Instead of simplistically labeling one group as being worst at driving than the other, analysts examine complex relationships between gender characteristics and their potential effects on traffic safety. Studies have researched gender issues in traffic safety concerning influences on accident frequency and severity (Mercer, 1987), gender differences amongst new adolescent drivers (Farrow and Brissing, 1990), self-control and alcohol (Keane et al., 1993), BAC of male and female drivers, riders, and passengers (Holubowycz, 1994), risk and involvement in alcohol-related driver fatalities (Zador et al., 2000), and influence of driver sex on road crashes (Al-Balbissi, 2003). The following section will review the literature concerned with gender and traffic safety by interpreting separately, for males and females, distinct results which are discussed in more detail for both groups.

Regarding males and traffic safety, the literature is filled with various conclusions concerning alarming rates tied to several factors. A study by Mercer (1987) involving a correlation examination of 84 months and unemployment rates, alcohol involvement, restraint use, and driver demographics concluded that increases in unemployment levels resulted in the decrease of a significant percentage of male drivers on the roads. In turn, Mercer found that this reduced the frequency of drunk drivers and increased the proportion of seatbelt users. In turn, the reduction of male drivers reduced the frequency and severity of accidents. Research conducted by Farrow and Brissing (1990) on adolescent drivers and gender characteristics concluded that young males involved in legal difficulties, lower academic achievements, and high amounts of work hours per week were more likely to engage in driving while intoxicated. In addition, the author found that males were more likely to be involved in racing and thrill seeking. They concluded that the higher rates of adolescent male drivers in automobile crashes was partially attributed to culture; commonly brought on by a ‘rite of passage’ to ‘manhood’ partially manifested by unnecessary risk taking whilst driving. Holubowycz et al. (1994) established that, based on hospital admission after a fatal or injury crash, males were more likely to have a BAC greater than 0.08% as both drivers and passengers of single car crash vehicles (See Table 6). Furthermore, among the fatalities, the mean positive BAC of male drivers was significantly higher. Studies by Zador et al. (2000) also concluded that, on average, males have continually been involved in more fatal crashes and have displayed greater relative risks of crashing than females. Lastly, a recent study by Al Balbissi (2003) examining annual travel distance, economic participation, and different income region accident records concluded that male crashes were often more harmful, frequent, and dangerous than female crashes. The author concluded that perhaps males were more impatient, higher risk takers, and paid less attention than their female counterparts. Interestingly, the previous observation is mentioned in several conclusions where it is mentioned that in general male deaths generally exceed those of females, and males speed, exhibit risky behavior, indulge in violence, and anti-social behavior more than women.

Table 6. BAC of Admitted Crash Victims (Holubowycz, 1994).

Distribution of blood alcohol concentration by type of road user and sex: fatalities and hospital admissions aged 16 years or older

	Fatalities						Admissions					
	Drivers		Passengers		Riders		Drivers		Passengers		Riders	
BAC	Males	Females	Males	Females	Males	Total*	Males	Females	Males	Females	Males	Total*
Zero	49.8%	72.3	42.1	69.4	56.8	54.3	57.8%	71.9	51.0	73.8	67.3	63.2
.001–.079	7.0	8.8	8.6	10.0	8.5	8.2	8.0	5.7	11.7	11.3	10.5	9.9
.080–.149	9.3	5.4	19.1	11.9	11.3	11.5	15.4	10.9	21.9	8.3	10.3	13.0
.150+	33.9	13.5	30.1	8.8	23.5	26.0	18.8	11.5	15.3	6.5	11.9	13.9
Total known: %	100	100	100	100	100	100	100	100	100	100	100	100
Number	540	148	209	160	213	1292	415	192	196	168	370	1405
BAC unknown	37	9	19	17	14	97	34	21	41	33	31	168

*Includes all drivers, passengers, M/C riders and pillion passengers ≥16 years old.

Regarding females and traffic safety, the literature appears to indicate a lesser degree of negative conclusions in many of the aforesaid male related factors. Interestingly, however, issues were noted in several articles which pointed out unique harmful relationships between the females and traffic safety. A chapter dedicated to gender risks in traffic safety by Evans (2004) concluded that females were intrinsically more likely to die from physical crash impacts in general. The author concluded that this was due to a woman's physical traits such as height, for example, being a factor in airbag deployment accidents. A study by Massie et al. (1997) concluded that, although males had higher average annual mileage and higher risks of crash involvement, females had greater involvement rates when examining non-fatal crashes. The authors concluded that a higher non-fatal crash rate was attributed to more women driving around urban areas where crashes are often of a lesser degree in terms of injuries and fatalities. A recent study by Nyberg and Gregersen (2007), conducted in Sweden on 18 to 24 year old novice drivers, concluded that although females scored higher on written driving tests, males often drove better and performed the same maneuvers more correctly during actual driving lessons. Nonetheless, the authors concluded that males were more often involved in crashes, due to similar abovementioned reasons, and that driving schools should restructure their lessons accordingly so that male crashes could be reduced and female driving skills increased. Another study by Gmel et al. (2006) explored associations between heavy drinking episodes (HDE), consumption before injury, and volume drank separately for each gender accepted into a Swiss emergency rooms. Like many other studies, the researchers concluded that higher HDEs and volumes of alcohol increased the overall risk of injury in any activity. More importantly, it was found that low-volume drinking women without HDE experienced a high proportion of injuries. For this reason, the authors advised that women who fall in this category take special precautions in risky circumstances such as driving. Lastly, a study conducted in Finland by Laapotti and Keskinen (2004), examining differences between male and female driving patterns in 1984 versus 2000, found that females were more often involved in crashes connected to maneuvering, loss of control, and reversing (backing-up) accidents while sober and not speeding.

Finally, it is important to state that not all factors mentioned in the literature point towards one gender being any more significant than the other, in regards to traffic safety and alcohol consumption. Keane's (1993) article, addressing primarily with self-control and DUI, concluded that convicted male and female drivers shared common traits regarding self-control. Specifically, it was found that based on several tests, sampling, and datasets, low self-control could commonly indicate a relationship between drinking and driving for both genders. Thus, Keane dismissed any common issue raised by other analysts regarding separate laws for genders, reiterating that such legislation would be unfitting and unnecessary. For this reason, it has been difficult to try reducing drinking and driving by enforcing laws or programs on one gender and not the other, except for increased insurance premiums for young males compared to females. However, the majority of the literature points towards gender as being a significant determinant of several factors involving traffic safety. Moreover, males, especially young males, are commonly seen as being more detrimental to traffic safety when it comes to drinking and driving.

Much like gender, age is often a strong determinant and factor in understanding and assessing traffic safety. Conversely however, the age of a driver has been subject to several laws, such as the zero tolerance law, for drivers under the age of 21 in the U.S. According to the NHTSA, the 21 year old purchase age has been responsible for a significant decrease in fatal crashes, injuries, homicides, suicides, and other alcohol-related problems under that age limit. Nonetheless, the drinking and driving problem persists in different age brackets and fatal crashes continue to take the lives of mostly young and healthy individuals, some of which do not necessarily exhibit problematic drinking behaviors. Although fatal alcohol-related crashes are declining in general, problems persist where the age of a driver as a factor continues to produce discouraging rates. Reinforcing the younger driver problem, Evans (2004) reports that young male drivers have the highest fatality and crash rates, and pose the greatest threats in traffic safety.

For these aforementioned reasons, the following paragraphs will explore age specific factors related to traffic safety and alcohol. Firstly, young drivers will be discussed (approximately < 30), followed older drivers (approximately > 55). Although specific age brackets vary between different studies, crash analysts commonly designate ten year intervals in which studies can focus conclusions on individual groups with similar characteristics. Older and younger drivers are the point of focus as a result of a dominant amount literature relating their impacts on traffic safety. Commonly, middle-aged drivers (approximately 30 to 55) are considered to contain lower rates when considering overall crashes (see Figure 3). The reason why middle-aged drivers are at the bottom of the curve on figure 3 lies in their aversion to risk, strong judgments, understanding of consequences, driving experience, maturity, and overall awareness of traffic (McGwin and Brown, 1999). Additionally, various articles and research attempts have concluded that, all other traffic factors being equal, younger and older drivers are comparably hazardous per mile driven compared to middle-aged drivers (McGwin and Brown, 1999).

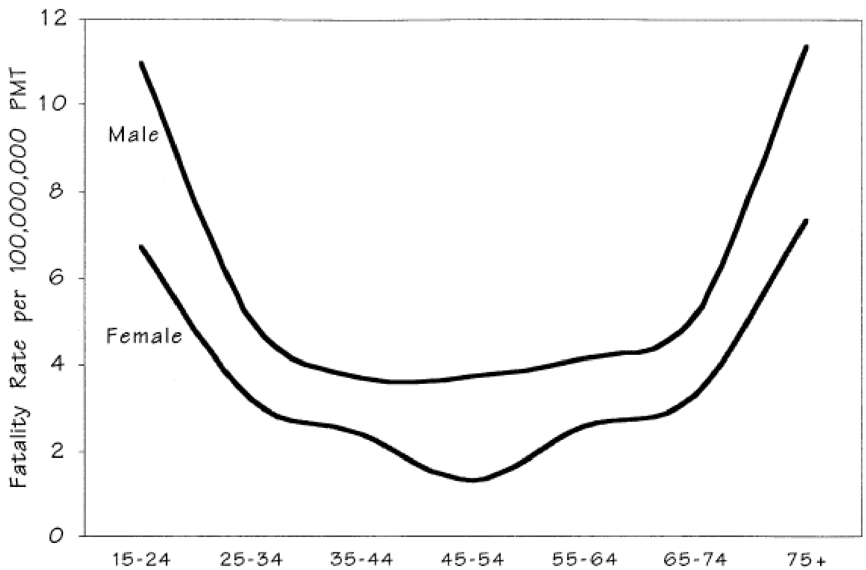

Figure 3. Fatality Rates per 100 Million per-Person Miles of Travel (McGwin and Brown, 1999).

According to Figure 4 (NHTSA's 2005 Traffic Safety Facts report) drivers in their early to late 20's are problematic in terms of alcohol consumption and their decision to drink and drive. This is unfortunate considering the fact that most drivers killed in alcohol-related crashes are often healthy and have much to give, considering their relatively young ages. Perhaps more alarming are statements made by McGwin and Brown (1999) that younger drivers incur the majority of fatal crashes and this is likely to remain this way in the future. According to a study by Dissanayake and Lu (2002) performed on the severity of young driver crashes, high crash rates and poor safety performance can be attributed to three major factors: inexperience, risk taking behavior and immaturity, and greater risk exposure. Regarding behavior, Farrow and Brissing (1990) attempted to explore various factors leading up to an adolescent's decision to drink and drive. By assessing several scores acquired through written confidential questionnaires, the authors were able to gather useful resources which helped to formulate several conclusions. The authors found that certain factors present prior to attaining a drivers license could give insight towards predicting, or at the very least understanding, why young drivers exhibited greater rates in alcohol related crashes. Some of these factors included the presence of legal difficulties, lower academic achievement, engaging in dangerous activities, drug use, family dysfunction, and even earlier dating. Research by Lam (2002) concerned with distractions inside and outside of a vehicle found evidence of links between certain age groups and the ability to overcome distractions. The author assimilated that since the recent efforts to assess cell phone distractions and their effects on crash rates, more attempts have been made towards understanding distractions and their consequences on drivers of different ages. Lam's results related that young drivers, due to immaturity, developing cognitive capacity, and greater attention and concentration spent on driving the vehicle, were at a greater risk of crashing during distracting events. Coincidentally, the state of California recently (July 1st, 2008) successfully implemented a cellular phone law which prohibits drivers from using a cell phone while driving. The law states that drivers over the age of 18 may employ a hands-free device only. Any use of a cellular device for a reason other than an emergency service call now results in a specified fine ranging from $ 20 to 50. Another study by Yagil (1998) on a Northern Israeli university using questionnaires concluded that, regarding gender and age-related differences in attitudes towards traffic laws and violations, younger drivers expressed a lower level of normative motivation to comply and respect traffic laws. A paper by Miller et al. (1998) estimating 1993 highway crash costs using several data sources such as medical bills, police enforcement, and per mile driven found several factors related to young drivers. The authors concluded that novice drivers, designated as drivers under the age of 18, were overrepresented in crashes and had higher estimated crash costs per mile driven compared to other drivers in different age brackets. Even more alarming, the authors found that an estimated 53% of all crash costs, but just 5.5% of all vehicle miles, could be related back to young drivers and drunk driving.

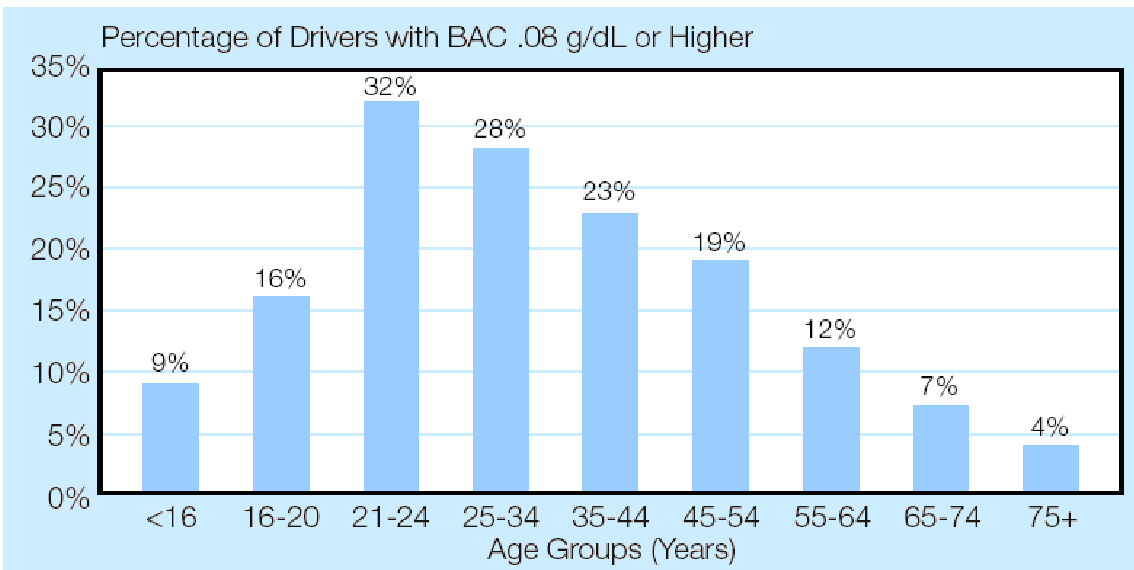

Figure 4. Age Groups and Drivers Killed in 2005 at BAC .08 g/dL or More.

Recently, there has been an increase in the amount of attention paid to older drivers. According to McGwin and Brown (1999), this attention shift has been partially attributed to the numerical increase of aging drivers on the road and a greater availability of rates and studies regarding their impacts on traffic safety. These authors enforce those statements by pointing out that, by 2030, the proportion of US population aged 65 and above will be nearly 22%, compared to about 13% back in 1996. In general older drivers tend to refrain from consuming alcohol, and if they do drink are more responsible and deterred from driving. However, an aging population generates several unique factors, compared to those of younger and middle-aged drivers that decrease their overall traffic safety. According to Evans (2004), the literature has produced several strong arguments revolving around the understanding and acceptance that specific driving skills decrease with age. Although older drivers are less of a safety concern than the younger ones, Evans argues with an example that an 80 year-old woman driver is 7 times more likely to be killed than a 45 year-old women in trips of equal distances. Moreover, aside from being more fragile and thus more likely to die from severe crashes, the risk of crashing increases steeply with age at the oldest ages. As discussed above in the introductory paragraph of the 'age' section, aging correlates to risk aversion, though to a certain extent. According to research by McGwin and Brown (1999) on characteristics of traffic crashes based on age, perceptual problems and difficulty judging and responding to traffic flow unfortunately can counterbalance risk aversion. Using police-record crash data from 1996 in the state of Alabama, the same authors found that older drivers (55+) are more likely than middle-aged drivers to have crashes at intersections, fail to yield or heed stop signs, strike unseen objects, and be at fault.

2.4 PLACE WITHIN THE LITERATURE

According to what has been recorded in the literature, county-level research has not been conducted in regards to the effects of age, gender, location and population on alcohol-related crashes for the state of Texas. Furthermore, a spatial-temporal approach attempting to study driver fatalities where alcohol was recorded at the county level in Texas has not yet been conducted. The spatial autocorrelation will be established at the county level since data availability at finer scales is limited. It is important to state that, regarding aggregated data, crash datasets available at the county level have several advantages over other types of scales; one of the most important ones is the availability of transportation and socioeconomic data (Valverde et al., 2006). Furthermore Thomas (1996) discusses that it is commonly accepted by scholars that all levels of aggregation are meaningful in road safety analysis: The choice simply depends on the aim of the study. Finally, the use of a GIS and GeoDa for choropleth, isoplethic mapping, and statistical analysis varies within the literature; especially in terms of map outputs and analysis regarding various results. Thus, an emphasis in this research will be placed on producing choropleth, cluster (spatial autocorrelation), and isoplethic maps as exploratory and explanatory mediums. In turn, the results will attempt to provide insights on possible issues related to demographic and socio-economic characteristics as factors as they may be associated to fatal alcohol-related crashes where a driver had a BAC greater than or equal to 0.01%. Aside from attempting to fill these knowledge gaps, this research attempts to provide another approach and methodology in conducting this type of research that can be used as a foundation for further research.

CHAPTER III

DATA AND METHODS

3.1 DATA AND STUDY AREA

This research places a greater part of its emphasis on analyzing and drawing spatial information from ongoing crash records compiled across the 254 counties located in the state of Texas. For this reason, most of the data discussed in the following sections was typically acquired at the county level through several websites and organizations. Firstly, Texas was chosen as the study site for a few reasons including practical and personal preferences related to the context of this paper. Personal reasons included familiarity, availability of data through various local contacts, and experience concerning several counties, cities, and locations across the state.

Practical reasons revolved around the fact that all of the research was conducted on site at Texas A&M University where county level crash datasets were readily available for further processing. Other decisions for choosing Texas as the study region included more specific or unique reasons including its distinctive spatial variability in the context of population distribution, the Texas Urban Triangle, demographic and socio-economic clusters, and the fact that Texas is the second largest and populous state in the U.S. Finally and perhaps most importantly considering the idea that traffic safety aims to reduce and understand crash rates, Texas consistently reports higher traffic fatality and alcohol related crash rates in contrast to nationwide averages.

Data was gathered from several sources including the Fatality Analysis Reporting System (FARS), Texas Commission on Alcohol and Drug Abuse, National Center for Statistics and Analysis (NSCA), Texas State Data Center (SDC), Texas Department of State Health Services, and the U.S. Census Bureau. FARS is a popular and often-employed compilation of various traffic crash datasets covering a wide variety of factors which mainly originate from police records and on-site documentations. Thus far, through the collaboration of NHTSA and US Department of Transportation, the online encyclopedia has kept a census of all US fatal crashes since the 1st of January 1975. For this study, FARS provided several datasets including fatalities per age groups, fatal alcohol crashes, and estimated licensed drivers. Regarding fatal alcohol crashes, the FARS database began to apply a Multiple Imputation method in 2001. The Multiple Imputation, which estimates missing BAC values using ten imputed variables, employs Rubin's method of scalar estimands (Subramanian, 2002). In comparison to the Discriminant Analysis which was previously applied to BAC datasets to provide estimates of missing variables, the Multiple Imputation followed a similar trend temporally, although some rates were up to 2% higher in some years (See Figure 5). Other data sources mentioned above either relied on census data where counts or averages were compiled for age, gender, population etc.

Figure 5. Comparison of Imputation Methodology. The Bold Line Represents the Multiple Imputation Method (Subramanian , 2002).

Once all of the necessary data was accounted for, counts per county were compiled into several Excel files. Concerning alcohol, drivers who were killed in a fatal crash with a recorded or estimated BAC $\geq$ 0.01% (any alcohol) were compiled within Excel. Similarly, total drivers killed in fatal crashes, regardless of alcohol content, were also compiled so that rates could be compared between both definitions. The independent datasets, or demographic factors, were also gathered and compiled from various sources. For this particular study, the chosen independent variables or factors were gender (male versus female), age, and population. Similarly to the crash dataset, these variables were compiled into several Excel sheets and arranged so that they were matched to their appropriate crash rates at the county level. The population data was found on the U.S. Census website where estimates of population by county were available from 1996 to 2005. Gender and age data was gathered from the Texas Department of Health Services website where it was possible to select gender counts and age counts for individual counties from 1996 to 2005. For the age data, 10 year intervals starting at 15 years of age and ending at 65+ were chosen to correlate to age distributions studied in previous articles.

Finally, though a 10 year period of data is available, 3 years were chosen to reduce the amount of data being processed and account for a greater range of change over several years rather than comparing 10 separate years. Of the ten years, 1996, 2000, and 2005 were chosen on the basis that they provided five year intervals and appropriate highs/lows respectively where: 1996 proved to be the most severe year in terms of alcohol-related fatal crashes and percent of total that were alcohol-related, 2000 delivered average or near-average values and was a census count year, and 2005 signified a steady and overall drop in percent crashes and alcohol crashes.

3.2 DESCRIPTIVE STATISTICS

In its original format, the data was presented as thousands of cells including counts for factors which were displayed and separated within excel for each individual county. In order to make more sense of that data, various calculations and charts or tables were carried out and created in Excel to assess the overall trend and gather basic information. These included percentages based on population, normalization, and summations or averages. As an example, some tables created were used to represent drivers killed per 100,000 licensed drivers where any alcohol was either present or absent. Rates were often chosen over raw counts in order to more appropriately account for any underlying trends such as increases in population, for example. Other tables were used to represent rates where estimated licensed drivers by age groups were compared to their corresponding fatality rates (total killed in traffic accidents per county totals) for 1996.

Since individual counties' reports of total licensed drivers was not available for this study, licensed driver data from the Federal Highway Administration was extracted for individual age groups and gender at the state level for the according years. Once the age groups were combined to match those of the Excel crash dataset, the state level rates were calculated with the county level population (ages 16 +) by multiplying the percentage of the total licensed driving population of an individual age group or gender by the total population of a county for the same gender or age group. For example, in 1996 there were 31,162 males (of all ages) in Anderson County. In order to estimate licensed drivers or drivers 16 years of age and above, the population count was multiplied by 67.59% (percent of total population in Texas that is a male and licensed to drive) to get an estimated 21065.51 male drivers in Anderson County. This same methodology was applied to both gender totals, total population per county, age groups, and genders within age groups to account for estimated drivers rather than total populations.

Fatality counts were calculated in the same fashion since normalized rates for individual counties were not available from FARS. For example, in 1996 Anderson County reported an estimated 15 drivers killed, including 7 of which were found to have had a positive BAC. In order to compare those counts to estimated drivers in Anderson County, each crash count was divided by the estimated total estimated licensed population and multiplied by 10,000. In other words, Anderson County reported 4.39 ((15/34189.57)*10,000)) drivers killed in traffic related crashes, and 2.05 ((7/34189.57)*10,000)) drivers killed and having positive BAC per 10,000 licensed drivers in that county.

3.3 GEODA DATA ANALYSIS

In order to assess spatial autocorrelation, shapefiles were created within ESRI's ArcMap 9.2 GIS software so that they could be imported into GeoDa. The shapefiles were created by importing the 1996, 2000, and 2005 Excel tables as *.csv files (comma separated values) into ArcMap. The selected years were then linked to a premade shapefile of Texas containing borders, boundaries, and county IDs by selecting two similar attributes from both tables and linking them as a single shapefile. Finally, the newly created shapefiles were exported as individual datasets where their projection and coordinate systems were defined and projected in ArcCatalog as a North American Datum 1983 Texas Centric Mapping System Albers and GCS North American 1983, respectively.

The abovementioned shapefiles were then imported into GeoDa so that local indicators of spatial autocorrelation could be studied across the three selected years and their corresponding datasets. Firstly, GeoDa requires its users to create weights dependent on the nature and spatial characteristics of the data and study area. Since Texas counties are contiguous, two options were given that could compute spatial autocorrelation based on a Rook or Queen Contiguity matrix and a desired order of contiguity. As Figure 6 shows, Rook and Queen Contiguities can be thought of as matrices where two different orders of direction can be implemented to define neighboring cells, or in this case a county. For a Rook Contiguity, adjacent neighbors located at cardinal intervals, in other words above, below, left, and right of the focus cell, are considered. A Queen Contiguity considers those same neighbors as a Rook Contiguity, and also includes cells located at each ordinal or intercardinal direction (NW, SW, etc). The contiguity order determines the extent of influence a spatial value has at various different scales. Essentially, a 1^{st} order considers those cells which are adjacent to the focus or center cell. In other words if considering a 3x3 matrix with center cell A1, a 1^{st} order will include adjacent cells or immediate neighbors A2-A9 when implementing spatial autocorrelation. A 2^{nd} order contiguity considers the spatial values which neighbor the previous neighbors or those cells adjacent to A2-A9 in the previous example. Since the level or order chosen for the Rook or Queen Contiguity is dependent on the spatial characteristics of the study and the nature of the data, a Queen contiguity was implemented for this study with a 2^{nd} order including lower orders.

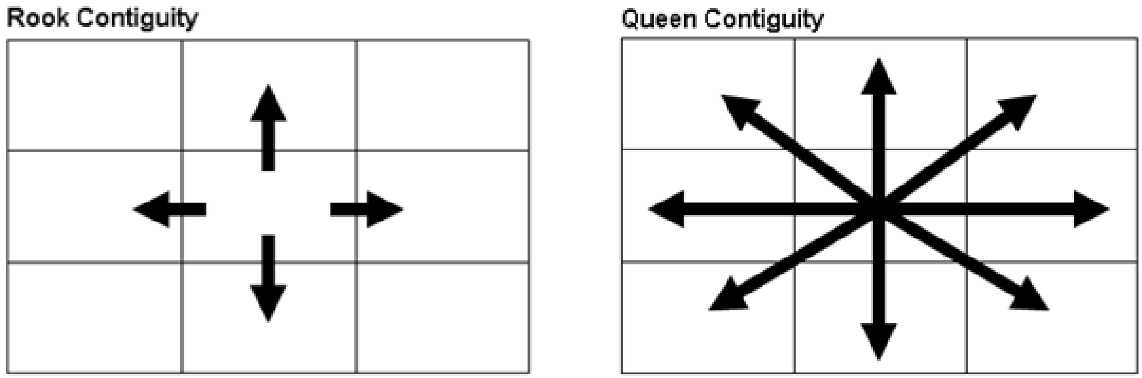

Figure 6. Rook versus Queen Contiguity.

Once the weights were created and saved, univariate LISA was calculated for the following variables: Total drivers killed, drivers killed with positive BAC, normalized total drivers killed (crash per 10,000 licensed drivers), normalized drivers killed with positive BAC, estimated licensed drivers, total population, population density (people per square miles), % males, % females, and % of both genders, and % of ages 16 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 and above for individual counties. Univariate LISA is related to Tobler's first law of geography which essentially states that "everything is related to everything else, but closer things are more related to each other". This phenomenon, which explores relationships in two-dimensional geographic space, can be derived from Moran's I which attempts to find departures from randomness across space. In turn, these departures from randomness are assessed numerically to determine whether or not clusters can be determined in the context of a designated dataset. Moran's I is defined as:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} \left(X_i - \overline{X}\right)\left(X_j - \overline{X}\right)}{\sum_i \left(X_i - \overline{X}\right)^2}$$

Where N is the number of spatial units (in this case 254 counties) indexed by i and j; $\bar{x}$ is the variable of interest (population, crashes, licensed drivers etc…), $\overline{X}$ is the mean of $\bar{x}$; and Wij is a matrix of spatial weights (as created in the aforementioned paragraph).

Univariate LISA results are yielded in 3 separate outputs including a cluster map, significance map, and Anselin's Moran's I scatter plot. The cluster map displays any significant findings within 4 unique categories based on the weighted characteristics and variables chosen per county. The key elements displayed in the cluster map revolve around clustering where X equals a variable in the county centroid, and Y equals a variable surrounding a designated county centroid. The 4 unique results are color coded so that positive spatial autocorrelation is represented either by a "high/high" or "low/low" event where: "high/high" (red) equals a significantly high value in X (center of cluster or county) surrounded by a significantly high values in Y (surrounding clusters or counties). A "low/low" (blue) event equals a significantly low value in X surrounded by significantly low values in Y. Negative spatial autocorrelation occurs both when "high/low" (pink) represents a significantly high value in X surrounded by significantly low values in Y and "low/high" (light blue) equals a significantly low value in X surrounded by significantly high values in Y. Figure 7 displays a resulting cluster map of total crash counts for Texas counties in 1996.

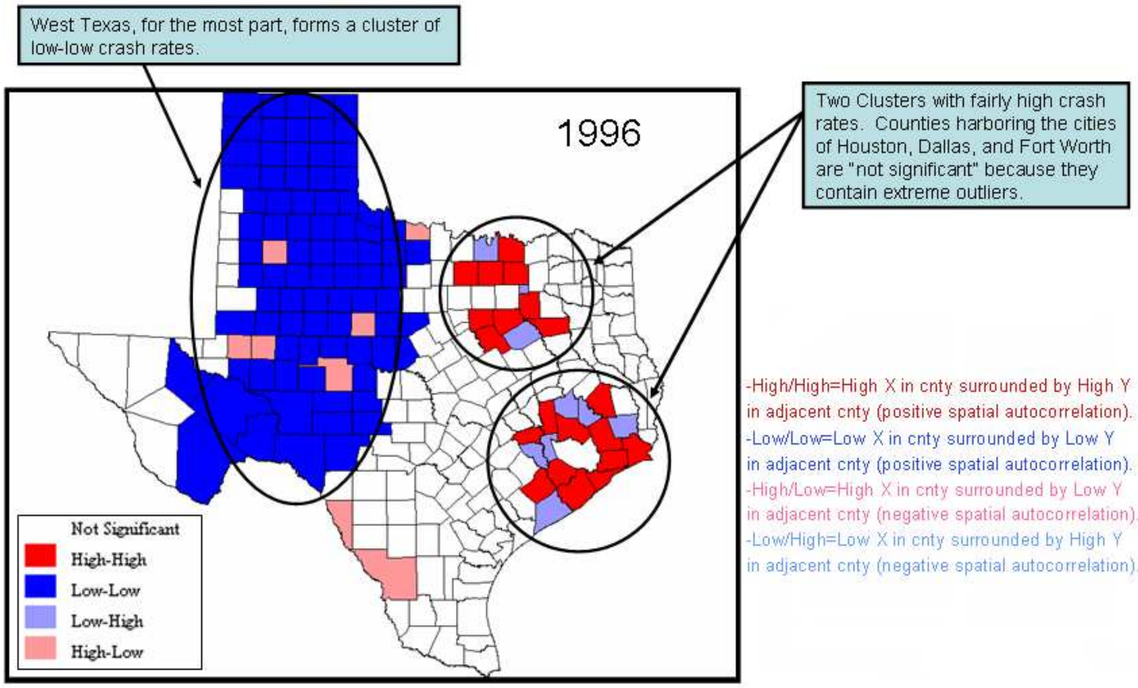

Figure 7. Fatal Crashes in Texas for the Year 1996; Cluster Map from GeoDa.

The significance map and Anselin's Moran's I scatter plot display significant values and negative or positive spatial autocorrelation, respectively (see Figure 8). For the significance map, the default p-value is 0.05 so that any value above that threshold is designated as being insignificant (see figure 8 as an example). Anselin's Moran's I Scatter plot displays all of the county standardized variables alongside the x-axis. The y-axis represents the spatial lag of those same county standardized variables in order to reflect the impacts neighboring counties have on each individual county centroids. If the slope reported at the top of Anselin's Moran's I scatter plot is positive there is evidence that, based on the assigned weights and algorithm output, a positive spatial autocorrelation is present (high/high, low/low). If the slope is negative, than the opposite is true where a negative spatial autocorrelation exists within the clustered results (high/low, low/high).

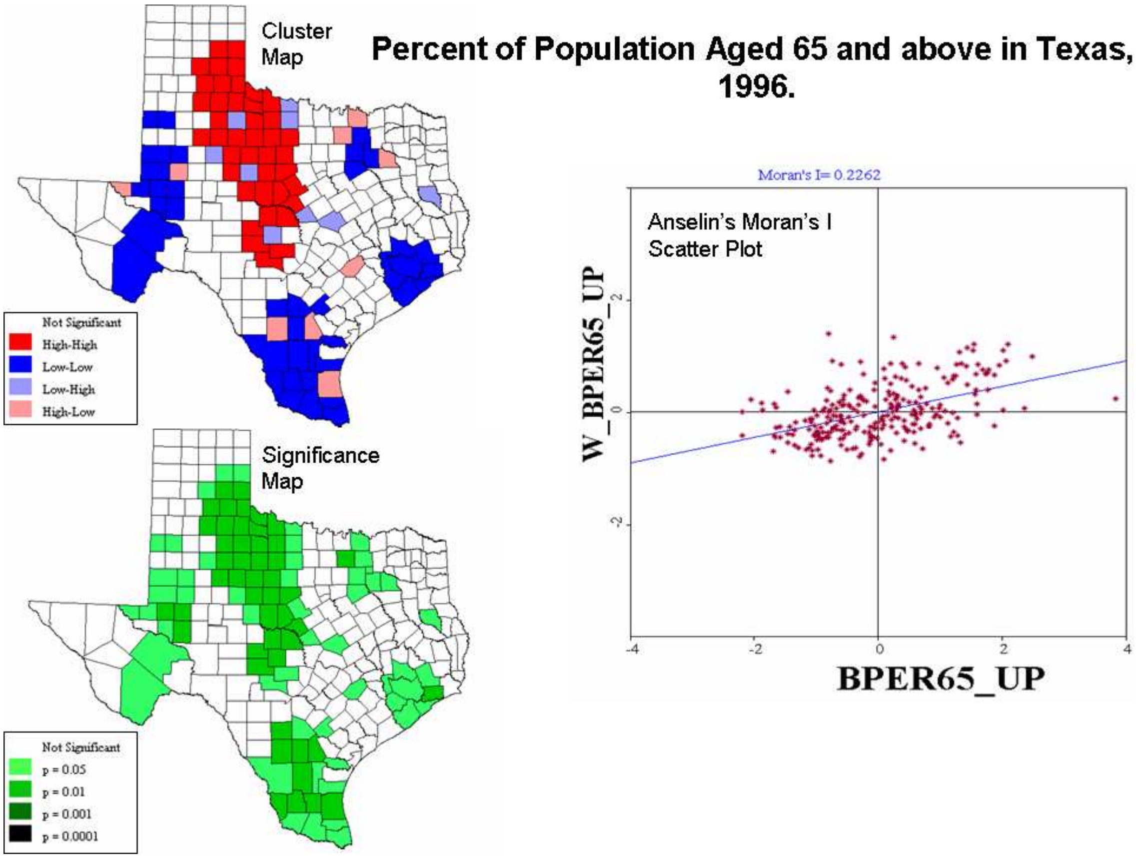

Figure 8. Example of Results from GeoDa.

3.4 GIS DATA ANALYSIS

The same shapefiles created from the Excel data sheets, which were previously used in GeoDa, were once more opened in ArcMap 9.2 for the creation of choropleth and interpolated maps. Although GeoDa had already been used to produce cluster maps in the form of choropleth results, the GIS software was used to depict the same datasets from a different point of view. Choropleth maps were produced for nearly all of the variables present in the Excel data sheets by following the necessary steps required in the graphical user interface within ArcMap 9.2. However, manual numeric classification ramps were created for certain variables in order to account for subtle changes and to assure a constant scale throughout the three separate time periods. When a manual classification was not used, a Natural Breaks (Jenks), which is available as an option in ArcMap and is typically the standard or default choice, was applied to represent the data. The Natural Breaks classification scheme simply examines all variables and determines natural breaks by statistically investigating adjacent pairs of values and assessing whether or not there is a large difference between the two. Figure 9 shows an example of all of the 2005 Males per county percentages classified under 6 unique categories in a Natural Breaks (Jenks) classification ramp with corresponding statistics. As can be seen from the figure, when a large difference in values between a pair occurs, a classification boundary is applied and a group is created.

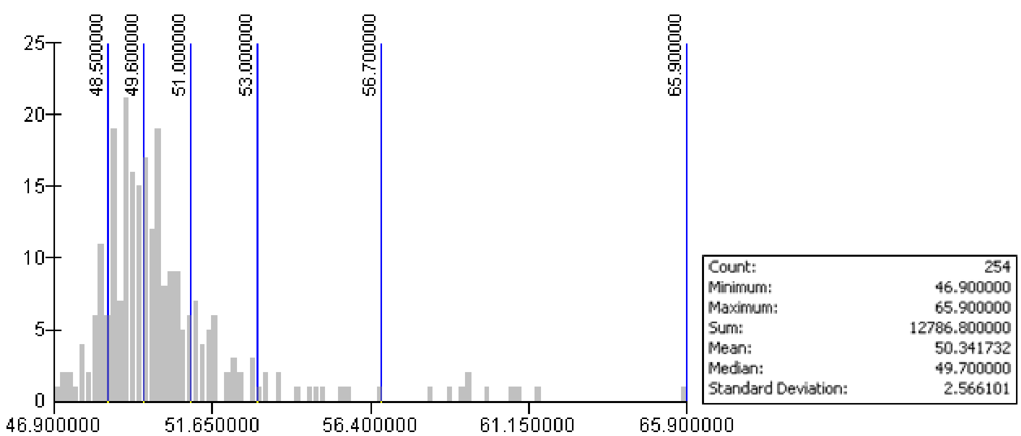

Figure 9. Example of Natural Breaks (Jenks) Classification.

Figure 10 represents a final output of estimated licensed drivers in Texas for the year 2000. As can be seen, the main focus of the choropleth map was to create an easy-to-read and comprehensible visual analysis of Texas counties and their unique values in contrast to studying numbers and tables compiled in Excel. From a choropleth map, spatial distribution and spatial patterns of chosen variables can be visually assessed rapidly, efficiently, and compared to other results found in GeoDa, tables, and interpolated maps. As can be inferred from the legend on the right hand side of the graphic, the outputs are color coded so that deviations from green towards the color red indicate an increase in value within 6 distinct colors (in this case licensed drivers). 6 categories were used for all of the choropleth maps in order to minimize confusion between color variances and to ensure that the 254 counties were distinguishable from one another without too much blending appearing. Further information regarding distances or spatial scale, map information, and date of creation can be gathered from other areas of the figure to improve and standardize the related information.

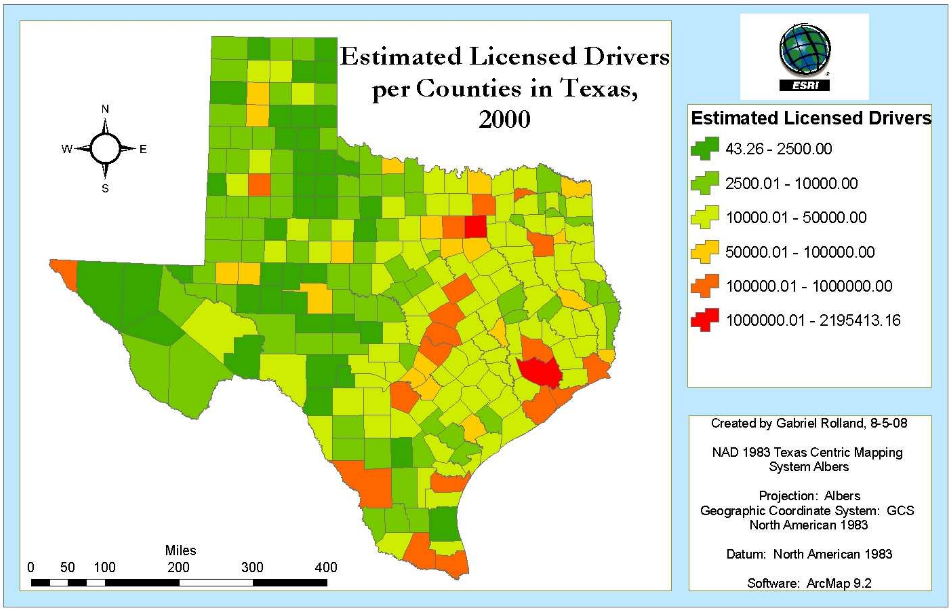

Figure 10. Example of Choropleth Map Output and Its Various Map Elements.

Interpolation of variables across space is often used in spatial sciences and GIS in order to better understand and represent data distribution, or account for missing values through various algorithms. Simply put, interpolation techniques take the set of known variables in a raster dataset and compute missing points based on neighboring distances, weights, and other spatial characteristics. Figure 11 gives an example of collected data points (left) and their interpolated values (right) after running a specified interpolation procedure. Interpolation in the context of this research is useful as another approach towards identifying areas and spatial distribution of high crash rates and related variables. For example, an interpolated map of crash rates will color-code those areas closer together in space under one defined color. This is due to the fact that, in a raster map generated from crash data, the creation and identification of points that are closer together and weighted accordingly are grouped under one defined interval. The resulting map gives the reader insight on the predicted spatial distribution of crashes over a continuous surface. Since the crash data is restricted to county boundaries in a choropleth map, due to the level of data chosen for this research, an interpolated map allows for the creation of a surface that lifts restrictions which are otherwise apparent in choropleth maps.

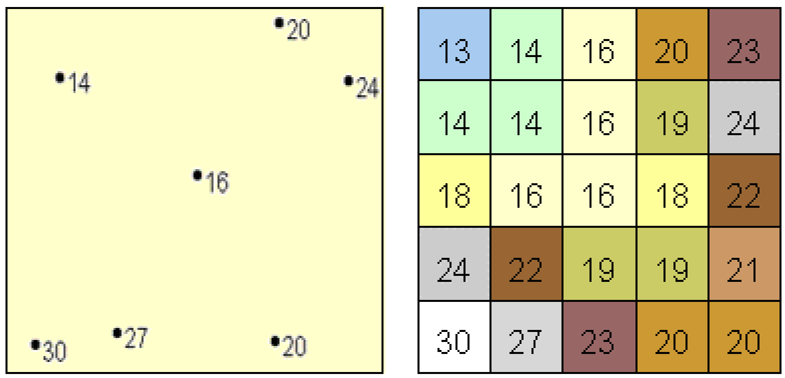

Figure 11. Example of Interpolated Values (Graphic Source: ArcGIS 9.2 Desktop Help).

A commonly used interpolation method for isoplethic mapping is Inverse Distance Weighting or IDW. Isoplethic representation of data using IDW interpolation determines cell values using a linearly weighted combination of a set of sample points. The weight assigned to the set of points is a function of inverse distance. The surface being interpolated should be that of a locationally dependent variable where points are regularly and densely distributed in the context of the local variations. In essence, as the distance between a scatterpoint and an interpolated point increases across space, the weight of the scatterpoint decreases accordingly thus reducing the influence those two points have on one another. Although there are several variations of IDW, it is commonly and more simply derived through:

$$F(x,y) = \sum_{i-l}^{n} w_i f_i$$

Where n is the number of sample points in the dataset, f_i are the values assigned to those individual sample points, and w_i are the assigned weight functions assigned to the sample points in the dataset.

In order to run IDW on the shapefiles previously used in choropleth mapping and GeoDa, several necessary steps were employed. Firstly, the county polygons which had originally contained the individual variables were converted into points. Rather than employing ArcGIS 9.2's toolbox which contains various related options, the shapefiles were converted in GeoDa. This option was comparably easier and less demanding than those presented in ArcGIS 9.2 as it involved simply running the 'polygon to point' option under the 'tools' drop-down menu in the GeoDa GUI. The resulting maps (1996, 2000, and 2005) displayed 254 points located in approximated county centroids containing those same variables that were previously studied. The newly created maps were then imported into ArcGIS 9.2 where the 'Interpolation' submenu under the 'Spatial Analyst Tools' option in ArcToobox was chosen. Once IDW was chosen as the interpolation technique, several parameters had to be defined including: input point features, Z value field, output raster, output cell size, power, search radius, and input barrier polyline feature. For point features and Z field, the newly created point files and individual variables were selected, respectively. The cell size was determined as the shorter of the width or the height of the extent of the input point features divided by 250. This value was automatically determined to be 4552.43 meters (top-bottom: (1569425.524038 m– 431318.574639 m)/ 250) and was not manually altered for any of the 3 years. The power option, which determines the significance of surrounding variables or points on the interpolated values, was left as 2. ArcGIS Desktop Help states that a higher power results in less influence. If further explains that most reasonable results are gathered when values ranging from 0.5 to 3.0 are selected as exponents of distance. A variable search radius was chosen under the search radius option. This was selected on the basis that it was the default option and was appropriate for this study. The number of nearest input sample points was set to 12, also the default value, considering the fact that for many counties this included all contiguous neighbors. Figure 12 represents the general the 3 different map outputs.

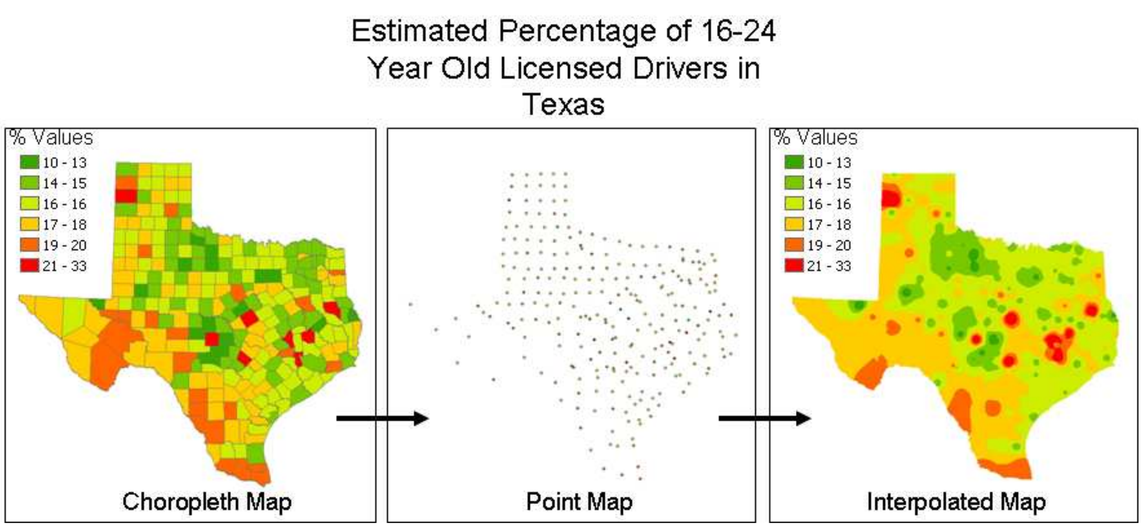

Figure 12. Example of Choropleth to Point to Interpolated Map outputs.

CHAPTER IV

RESULTS AND DISCUSSIONS

4.1 CRASHES

Table 7. Driver Statistics in Texas.

1996	12568000	2134	964	45.17	16.98	7.67
1997	12834000	2013	834	41.43	15.68	6.5
1998	13323000	2083	887	42.58	15.63	6.66
1999	13359000	2090	894	42.78	15.64	6.69
2000	13462000	2229	956	42.89	16.56	7.1
2001	13046000	2250	946	42.04	17.25	7.25
2002	13185000	2331	975	41.83	17.68	7.39
2003	13498000	2369	965	40.73	17.55	7.15
2004	14544000	2193	869	39.63	15.08	5.97
2005	14659000	2107	892	42.34	14.37	6.08
Years	**Licensed Drivers**	**Drivers Killed**	**Drivers Alcohol**	**% Alcohol**	**Drivers_Killed (per 100,000 licensed drivers)**	**Drivers_Alcohol (Per 100,000 licensed drivers)**

Table 7 is a summary which was compiled from the collected datasets concerned with crashes starting in 1996 and ending in 2005. The table represents all drivers killed in fatal crashes having either no alcohol or any alcohol (BAC $\geq$ 0.01%) in their bloodstream at the fatal event. The table also lists the total licensed drivers per year (Licensed_Drivers), total drivers killed in fatal crashes (Drivers_Killed), drivers killed in fatal crashes with any alcohol (Drivers_Alcohol), drivers killed in fatal crashes per 100,000 licensed drivers, and drivers killed with positive BACs per 100,000 licensed drivers. Licensed drivers increased by 14.26% or 2,091,000 drivers. Though this overall increase of just over 2 million drivers was no surprise, it was unexpected that in 2001 licensed drivers decreased by nearly 426,000 and then slowly increased beyond the year 2000 count by the year 2003. Similarly, drivers killed per 100,000 licensed drivers which recorded either no alcohol or any alcohol, exhibited a particular above-average trend for those 3 years starting in 2001 and ending in 2003. Paradoxically, the total population in Texas consistently increased by a total of 17.11% during the study period, thus not correlating with the unexpected decrease from 2001 to 2003. A closer look was given at the different age groups in trying to analyze the decrease by nearly half a million of the licensed drivers in 2001. The 20 to 24 and 55+ year old groups increased as expected while surprisingly the 16 to 54 year old groups decreased by a total of 416,296 licensed drivers (see Table 8).

Table 8. Difference in Licensed Drivers by Age from 2000 to 2001 (Federal Highway Administration).

Age Group	2000	2001	Change
16-19	771,147	721,553	-49,594
20-24	1,257,322	1,290,591	33,269
25-29	1,438,847	1,324,647	-114,200
30-34	1,447,224	1,404,419	-42,805
35-39	1,587,237	1,403,207	-184,030
40-44	1,537,402	1,427,055	-110,347
45-49	1,331,012	1,272,014	-58,998
50-54	1,114,379	1,097,354	-17,025
55-59	826,880	865,350	38,470
60-64	623,641	644,019	20,378
65-69	520,208	520,210	2
70-74	428,422	436,295	7,873
75-79	317,749	332,358	14,609
80-84	170,342	200,334	29,992
85+	90,211	106,321	16,110
Total	13,462,023	13,045,727	-416,296

Unfortunately, the number and rate of drivers killed with the presence of either none or any alcohol also increased even though the number of licensed drivers decreased sharply in 2001 and slowly increased up until 2003. However, that increase in fatalities was first apparent in the year 2000 where the number of driver fatalities increased by 139 and 62, respectively, for sober drivers or drivers with any alcohol present in their bloodstream. Fortunately for the State of Texas, the percent of drivers killed with BACs

≥ 0.01% decreased overall from 1996 to 2005. Furthermore, by 2005 the rates and numbers by both sober drivers and drivers with any alcohol present had somewhat decreased from previous years to lower values.

Drivers killed in fatal crashes (no alcohol present) results from GeoDa for 1996, 2000, and 2005 are shown in Figure 13. Generally, as the Moran's I value approaches + 1.0 or – 1.0, clustering or dispersion becomes apparent and more significant. As can be seen from Figure 13, the Moran's I value was nearly 0.12 for all significant values (p-value ≤ 0.05). For counts considering drivers killed with either none or any alcohol in the bloodstream at the time of the crash, cluster maps related similar information. Likewise for the map categories, an east versus west extent of clustering was apparent. Counties in western and northwestern Texas reported low crash counts surrounded by counties with similar values. On the other hand, parts of eastern Texas reported high crash counts surrounded by counties with similar values. Not surprisingly, these clusters were nearly identical to cluster maps produced from selecting population or estimated licensed drivers as the value of concern.

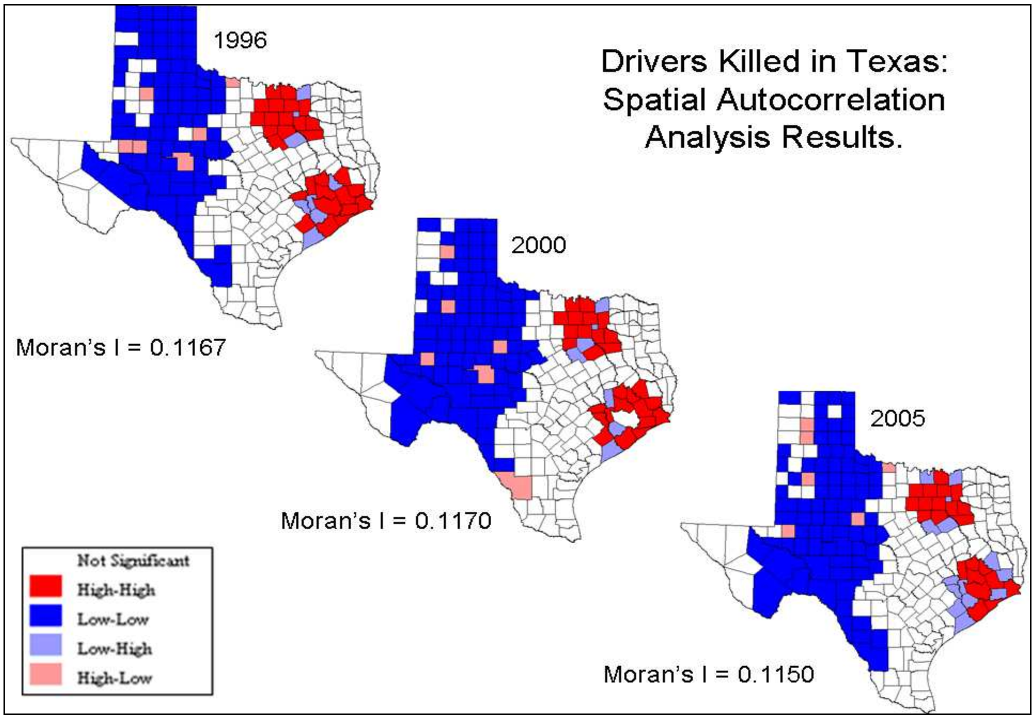

Figure 13. Spatial Autocorrelation Results for Drivers Killed in Texas.

The spatial autocorrelation crash count results brought particular attention to counties including and surrounding the cities of Houston, Dallas/Fort Worth, and San Antonio. Although most of these high-population counties exhibited below average crash rates considering their large populations, the top 20 counties in terms of fatal crash counts concerned with drivers killed with any alcohol present in their system, accounted for approximately 50 percent of all crashes in Texas for the three designated years (50.62 % in 1996, 49.69% in 2000, and 52.58% in 2005). Furthermore, within those 20 counties, the counties harboring the cities of Houston, Dallas/Fort Worth, and San Antonio (Harris, Dallas, and Bexar County) accounted for just over 20 percent of all drivers killed in fatal crashes with a positive BAC. To put that percentage in a broader picture, 4 out of the 254 Counties in Texas accounted for more than 20 percent of all fatal crashes where a driver was killed with either none or any alcohol in their bloodstream. Those same 20 counties also reported high driver fatality counts (accounting for 45.60%, 46.88%, and 46.65% of total crashes in the state for 1996, 2000, and 2005 respectively). Because those 20 counties accounted for roughly 65% of the population, any measure which could be applied to reduce fatality counts in particular counties such as Harris, Dallas, and Bexar could dramatically reduce total counts and overall state rates. Clustering was evident in the surrounding counties of those 3 major cities which happen to construct the outer edges of the Texas Urban Triangle. GeoDa reported high counts of fatal crashes within and surrounding those cities for both sober drivers and drivers having any alcohol present in their bloodstream. Since these counties contain a large proportion of the population in Texas, special attention should be given to those urban areas in order to reduce state fatality counts.

Figures 14, 15, and 16 compare the locations of the top 20 counties that reported the highest numbers in terms of drivers killed in traffic crashes with any alcohol present in their bloodstream at the crash. The tables to the left of the maps represent the same dataset which was introduced and discussed in the aforementioned paragraphs. The headings represent drivers killed in fatal crashes (Killed), drivers killed in fatal crashes where a positive BAC was reported (Killed BAC), drivers killed in fatal crashes per 10,000 estimated licensed drivers (killed Norm), drivers killed in fatal crashes where a positive BAC was reported per 10,000 estimated licensed drivers (BAC Norm), and estimated licensed population (pop). The choropleth maps add an element of spatial relationships and relate extra information when coupled with their original dataset. This visualization approach to understanding where the most crashes occur relates various information including spatial patterns which were apparent in the similar outputs created in GeoDa. For the most part, the maps show that the same counties are producing high fatal counts for all three years. Also, the maps show a clear distinction between east and west Texas. The distinction is obvious enough that one could almost draw a vertical line through central Texas separating the two halves in terms of crashes involving any alcohol. Furthermore, the Texas Urban Triangle is clearly distinguishable in all three choropleth maps. As for spatial variation between the three years it appeared that, with the exception of 1996 showing more counties with less traffic crashes than 2000 and 2005, the three years were similar regarding fatal alcohol crash distribution.

CNTY	Killed	Killed BAC	Killed (Norm)	BAC (Norm)	Pop.
HARRIS	189	104	0.92	0.51	3117376
DALLAS	129	59	0.98	0.45	1999926
BEXAR	91	52	1.05	0.6	1318431
TARRANT	60	30	0.70	0.35	1306287
TRAVIS	48	27	1.07	0.6	680541
EL PASO	40	22	0.90	0.5	673893
SMITH	41	19	3.79	1.76	164547
HIDALGO	43	18	1.32	0.55	496485
MONTGOMERY	41	17	2.58	1.07	241855
CAMERON	28	16	1.37	0.78	312064
DENTON	33	16	1.43	0.69	350905
COLLIN	34	15	1.39	0.61	373095
JEFFERSON	29	14	1.80	0.87	245849
MC LENNAN	27	14	2.03	1.05	202679
HILL	20	12	10.30	6.18	29538
FORT BEND	24	11	1.21	0.55	302017
GRAYSON	24	11	3.63	1.66	100611
NUECES	25	11	1.23	0.54	310561
GALVESTON	25	10	1.57	0.63	241981
KAUFMAN	22	10.00	5.43	2	61646
Total	973	488	2.24	1.12	12530287
% total crashes	45.60	50.62	__	__	65.51

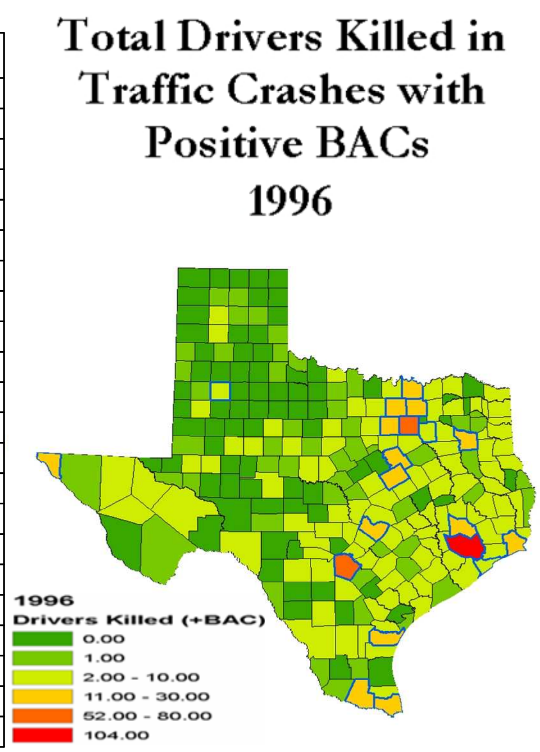

Figure 14. 1996 Top 20 Counties with the Most Drivers Killed with Any Alcohol.

CNTY	Killed	Killed BAC	Killed (Norm)	BAC (Norm)	Pop.
HARRIS	219	108	1	0.49	3400578
DALLAS	152	66	1.06	0.46	2218899
BEXAR	81	40	0.9	0.44	1392931
TARRANT	87	38	0.93	0.41	1446219
TRAVIS	73	28	1.39	0.53	812280
HIDALGO	46	22	1.25	0.6	569463
CAMERON	26	18	1.2	0.83	335227
BELL	33	16	2.15	1.04	237974
MONTGOMERY	37	15	1.95	0.79	293768
JEFFERSON	32	14	1.97	0.86	252051
NUECES	28	13	1.38	0.64	313645
BASTROP	23	12	6.17	3.22	57733
BRAZORIA	31	12	1.99	0.77	241767
SMITH	28	12	2.48	1.06	174706
KAUFMAN	25	11	5.43	2.39	71313
COLLIN	40	10	1.26	0.32	491675
DENTON	30	10	1.07	0.36	432976
GALVESTON	18	10	1.11	0.62	250158
HUNT	19	10	3.84	2.02	76596
PARKER	17	10	2.98	1.75	88495
Total	1045	475	2.08	0.98	13158454
% total crashes	46.88	49.69	__	__	63

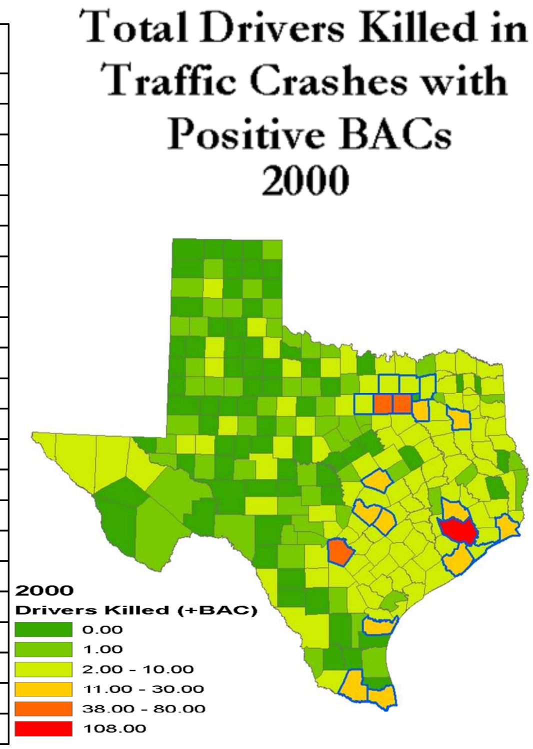

Figure 15. 2000 Top 20 Counties with the Most Drivers Killed with Any Alcohol.

CNTY	Killed	Killed BAC	Killed (Norm)	BAC (Norm)	Pop.
HARRIS	201	111	0.85	0.47	3693816
DALLAS	103	51	0.7	0.35	2300359
TARRANT	88	38	0.85	0.37	1621055
BEXAR	88	28	0.91	0.29	1510556
TRAVIS	49	27	0.85	0.47	896753
HIDALGO	49	23	1.13	0.53	677902
MONTGOMERY	48	23	1.99	0.95	375689
COLLIN	36	18	0.86	0.43	655687
EL PASO	30	18	0.64	0.39	726006
SMITH	35	16	2.87	1.31	190019
DENTON	37	15	1.03	0.42	558450
GALVESTON	23	14	1.3	0.79	275338
LUBBOCK	21	12	1.31	0.75	250276
BRAZORIA	30	11	1.69	0.62	276956
CAMERON	26	11	1.07	0.45	378074
FORT BEND	22	11	0.75	0.38	455991
JEFFERSON	35	11	2.21	0.69	247322
MC LENNAN	24	11	1.68	0.77	222313
HARRISON	19	10	4.68	2.46	63315
POTTER	19	10	2.48	1.31	119377
Total	983	469	1.49	0.71	15495254
% total crashes	46.65	52.58	__	__	67.78

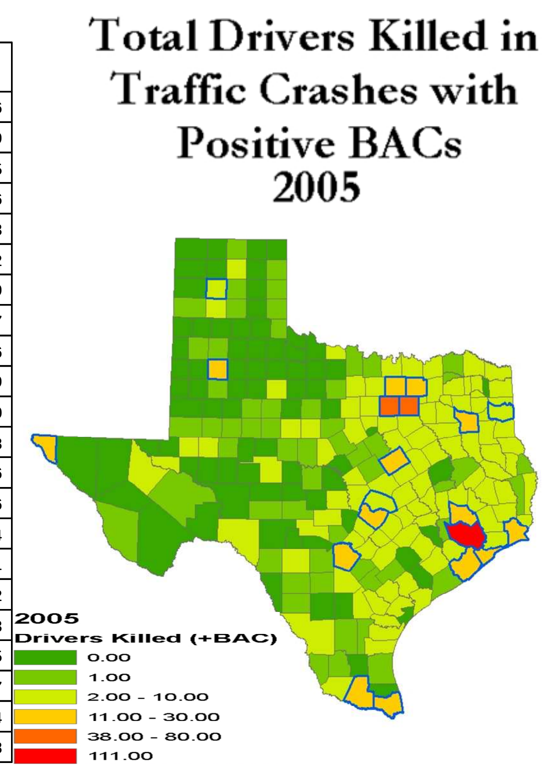

Figure 16. 2005 Top 20 Counties with the Most Drivers Killed with Any Alcohol.

Although unique spatial patterns were clearly distinguishable when looking at crash counts, rates which were computed per estimated 10,000 licensed drivers yielded very different results. Rates computed from GeoDa results of drivers killed per 10,000 estimated licensed drivers provided very little evidence in the context of significant clustering. The Moran's I results decreased from 0.0455 in 1996, to -.02560 and -0.0207 for 2000 and 2005. This signified that once normalization by estimated licensed drivers was applied to the crash counts, crashes became random events with no apparent influences from their surrounding county rates. In other words, the clusters returned that were significant were highly dispersed (Figure 17). Nonetheless, there were a few clusters which appeared to persist over the years. Clusters of low-low and high-low relationships were apparent in the western panhandle of Texas for all three years. Specifically, counties in northwest Texas where low population counts and corresponding low crash fatality rates were apparent returned significant low-low clusters. In the southern tip of Texas, the years 2000 and 2005 returned a cluster comprised of several low-high clusters due to high fatal crash rates with positive BACs for those years in Kenedy and Jim Hogg counties. Unfortunately, these clusters were difficult to interpret considering the fact that many of these counties which reflected high crash rates had very low populations. For example, Jeff Davis County located in western Texas returned a rate of 14.77 traffic fatalities per 10,000 licensed drivers. Though this was the highest rate for Texas in 1996, Jeff Davis County was found to have had 2061 inhabitants with a total fatal crash count of 4 and 2 respectively for crashes recording none or any alcohol present for the driver killed. Thus, the 1996 cluster in west Texas of high-high fatal crash counts where a positive BAC was established was solely due to Jeff Davis County having a much higher rate than surrounding counties.

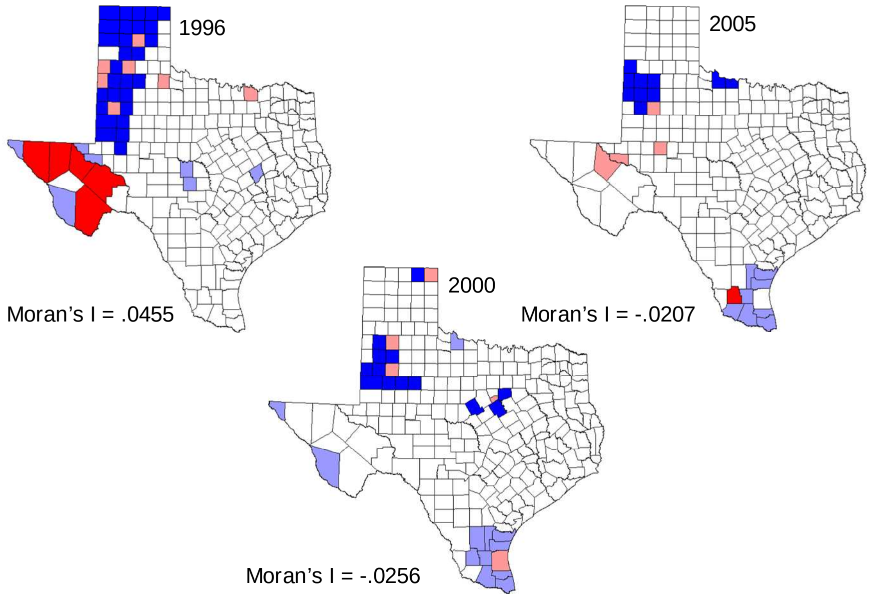

Figure 17. LISA Results for Drivers Killed in Traffic Crashes with Positive BACs (per 10,000 Estimated Licensed Drivers).

Similarly to results gathered from GeoDa, maps from ArcMap 9.2 returned comparable spatial-temporal outcomes when examining rates over crashes. As Figure 18 shows, interpolated results showed similar spatial patterns for all three years. Evidence of those same clusters depicted in the GeoDa results is also apparent in Figure 18. For example, the same county in 1996 (Jeff Davis) which returned high rates for fatal crashes where a driver was killed and found to have a positive BAC is clearly distinguishable by its orange color in western Texas. Unfortunately, the same problem was apparent in the interpolated maps where counties with excessively high rates often were attributed to low populations and relatively low crash counts. However, the interpolated maps proved that for the most part, most counties across Texas returned average crash rates. This was especially true in east Texas and along the Texas Urban Triangle where crash counts had been excessively high in those areas where population concentrations were much higher than average.

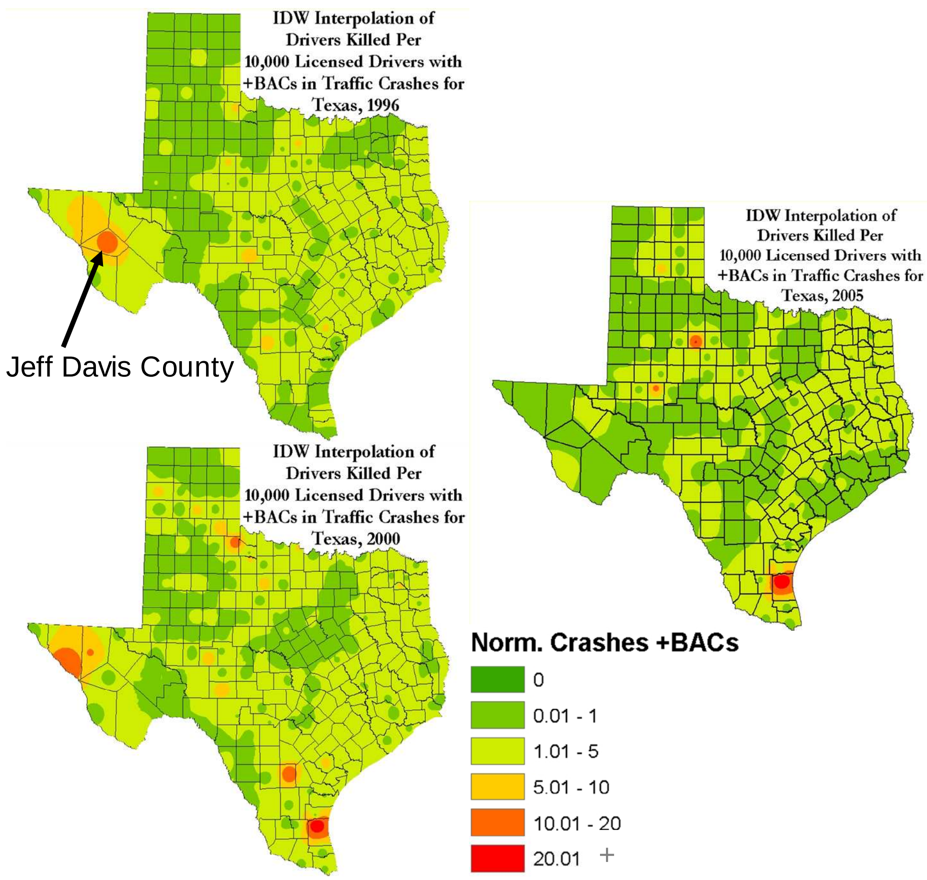

Figure 18. IDW results for Drivers Killed in Fatal Crashes. Displaying per 10,000 Licensed Drivers where a Positive BAC was established for the Driver.

4.2 AGE

As mentioned in the previous chapter, the age distribution of estimated drivers ranged across 6 groups (G1-G6). Each group was separated by 10 year intervals, except for 16 to 24 (G1) and 65+ year olds (G6), starting from 16 years of age and ending with G6 for ages 65 and above. Moreover, each group was broken down into three categories explaining whether the results accounted for male, female, or both sexes.

Figures 19, 20 and 21 represent total fatalities in Texas for 1996, 2000, and 2005. Clearly, a disparity exists between the population counts and the estimated fatality rates recorded during all three years for those young drivers aged 16 to 24. As is apparent by the fatality trend line spanning across the 6 age groups for all 3 years, younger and less experienced drivers clearly pose a safety problem similar to what was referenced and discussed in the literature. Although young drivers represent a relatively small percentage of all drivers, they generate a major problem in the sense that they consistently produce above average crash rates. As the drivers mature, their numbers increase while their fatality rates generally decrease (G2, G3, and G4 for 2000 and 2005). Particularly, drivers aged 45 to 54, the baby boom generation, showed an interesting trend across the three years whereby their total populations increased dramatically while their fatality rates decreased for all 3 years into 2005. Finally, G5 and G6 displayed an upwards movement along the fatality rate line, also agreeing with the literature about the fact older drivers often showed an increase in crash rates in proportion to their percentage of the total driving population.

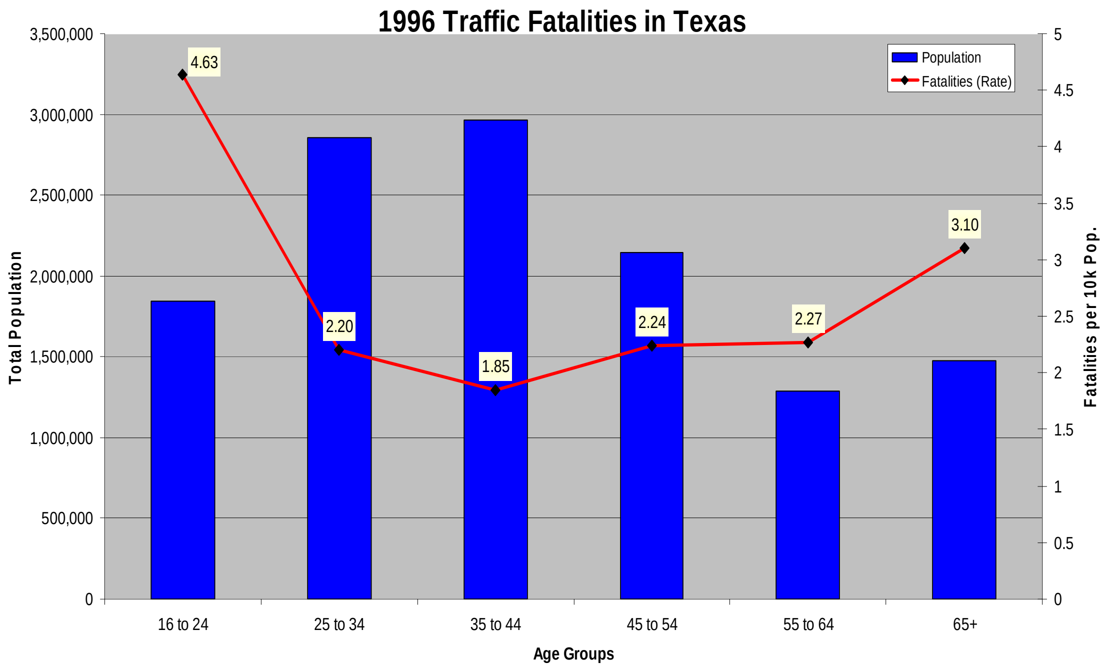

Figure 19. 1996 Traffic Fatality Rates versus Population Counts for Age Groups.

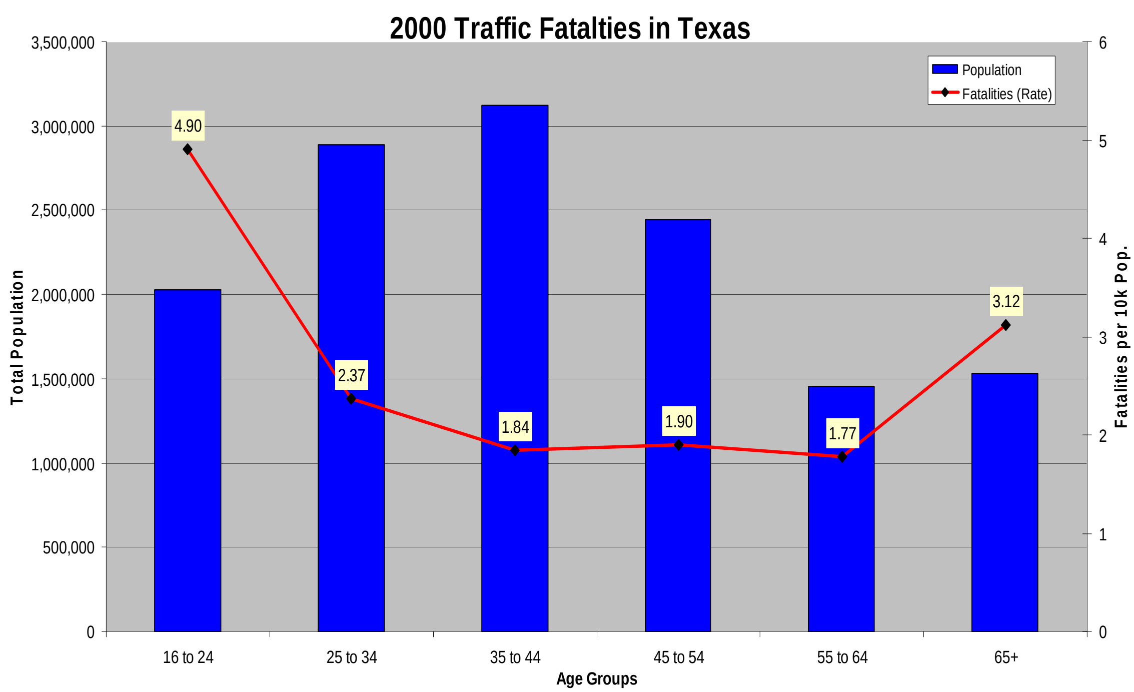

Figure 20. 2000 Traffic Fatality Rates versus Population Counts for Age Groups.

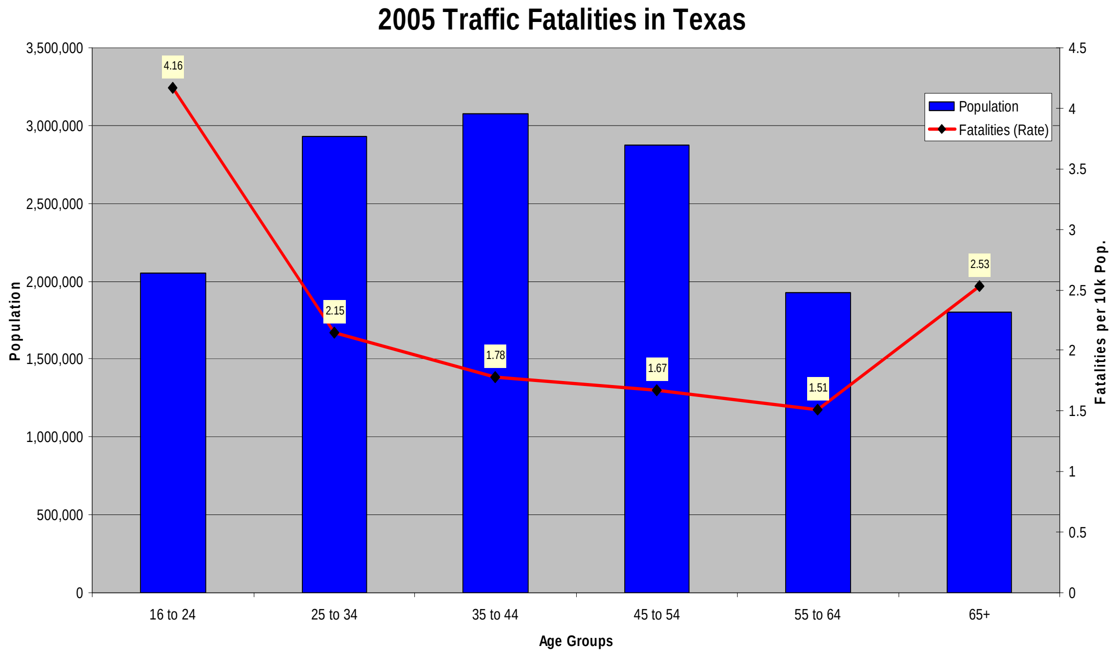

Figure 21. 2005 Traffic Fatality Rates versus Population Counts for Age Group.

Similarly to the tables depicting the top 20 counties in term of crash counts, tables listing the top counties with the 20 highest percentages of estimated licensed drivers aged 16 to 24 were created for all three years. Among those counties, Brazos County, home of Texas A&M University which is one of the most populous universities in the nation, consistently reported the highest percentage of estimated licensed drivers for that age group. However, the crash rates reporting either none or any alcohol present in the driver killed were well below average compared to the state averages (0.88, 0.71, and 0.39 drivers killed with a +BAC per 10,000 licensed drivers). Walker County, which is the home of Sam Houston State University, was similarly found within the top 3 counties for highest proportions of young drivers. In 2000 Walker County returned an estimated 2.01 drivers killed per 10,000 licensed drivers who had a positive BAC and had similar fatal crash counts compared to Brazos County even though its total population was just over 1/3rd of Brazos Counties'. Nacogdoches County also reported rates, where a positive BAC was established for the killed driver, which were worth of attention (2.05 and 2.35 driver killed per 10,000 licensed drivers). Interestingly, Nacogdoches County, which is just shy of Walker counties population, is the home of Stephen F. Austin State University, coinciding with Brazos and Walker county in terms of having a higher percentage of younger drivers likely due to the large number of students in state universities.

Figure 22 shows the spatial distribution of the 20 counties with the highest percentage of estimated drivers aged 16 to 24. Also listed in that figure are the counties which coincidentally have state universities. Spatial patterns concerning population distribution indicated higher concentrations of 16 to 24 year old licensed drivers along the Texas-Mexico border. Although not explored in this research, this was understood to correlate to the higher percentages of Hispanic populations which tend to be younger, on average, in comparison to counties in Central and north/northeastern Texas where the percentages in that age group were much lower. Similar results were reported in the interpolated maps which showed high concentrations of 16 to 24 year olds along the border and into west Texas, as well as those counties near large universities (i.e., Brazos and Walker County). According to the choropleth and interpolated maps, the general trend was a decrease in 16 to 24 year olds towards the upper central portions of west Texas and along the northern border shared by Oklahoma and Texas.

The Moran's I results concerning 16 to 24 year olds from GeoDa ranged from 0.0718 in 1996, to 0.0692 in 2000 and, interestingly, 0.1315 in 2005. In agreement with the results displayed by ArcMap, the significant clusters for 1996, 2000, and 2005 were as follows: Low-low in northern Texas, high-high around Brazos and Walker County as well as along the Texas-Mexico border, and high-low or low-high around Brazos and Walker Counties. The GeoDa results for 2005, which reported a relatively strong amount of clustering, showed 2 major areas of high-high clustering and 2 areas of low-low clustering located in west Texas and south Texas, and northeastern Texas, respectively.

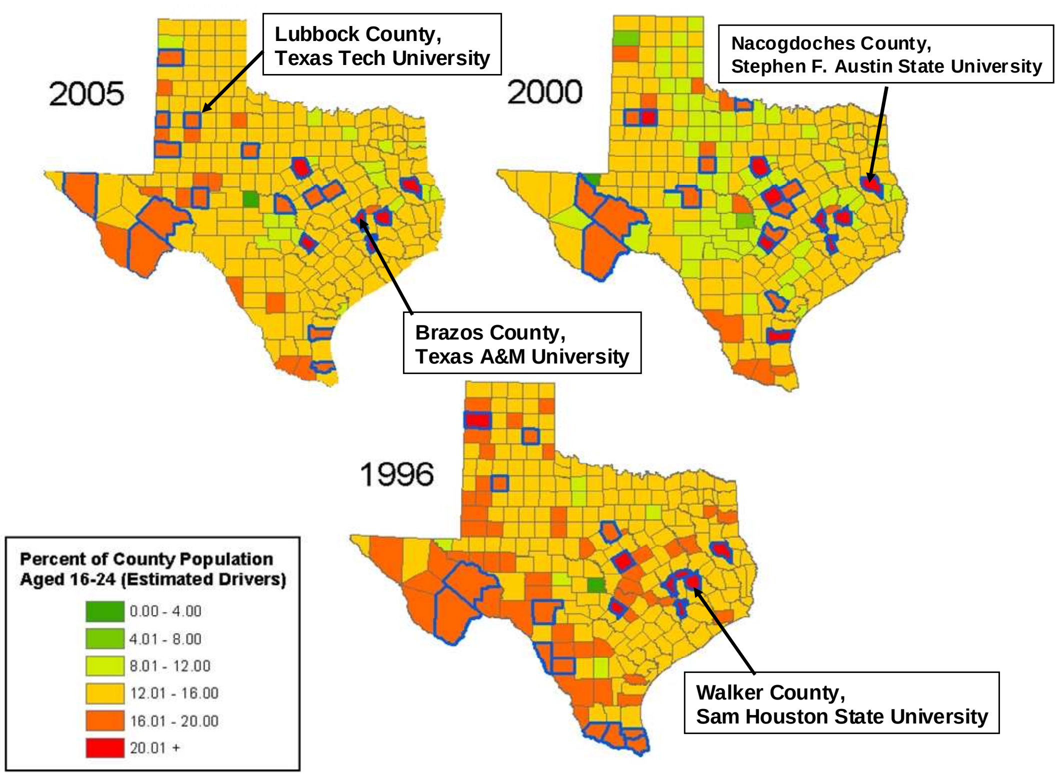

Figure 22. 1996 Estimated Driver Percentages Aged 16 to 24..

Lastly, in order to try to better understand whether a possible relationship existed between the percentage of young drivers and the corresponding crash rates where a positive BAC was established, an attribute query was carried out. The query was similar for all three years: it searched for results whereby above average crash rates AND high percentages of young drivers aged 16 to 24 were located across the choropleth maps. The query was as follows for all three years: "Nrm_DrkDrv" >= (average crash rate for 'X' year) AND "Bper16_24" >= 16.01%. 16.01% was used as the above average value since it signified a high concentration of younger drivers in the choropleth maps. Interestingly, 1996 returned 23 counties which had above average crash rates as well as high percentages of younger drivers. Of the 23 counties, 12 were located along the Texas-Mexico border, an area which was found to consistently represent a higher percentage of younger drivers. Nacogdoches County reported the highest crash count where any alcohol was found and represented the highest percentage of estimated drivers aged 16 to 24. As for the counties with the highest crash rates, they contained very low populations to begin with and thus were overrepresented considering the fact that they had very low crash counts. For example, Jeff Davis County had an astounding crash rate of 14.77 drivers killed per 10,000 licensed drivers where a positive BAC was established. However, although that same county had 16.48% of the estimated driving population aged 16 to 24, the estimated total licensed population count was a mere 1355 people: this shows the limitation of the method and the need to accommodate population in the analysis.

For the year 2000, 8 counties were returned under the abovementioned query. Nacogdoches once again was accountable for having the highest crash count where a positive BAC was recorded for the deceased driver. With an estimated 23.82% of the driving population being between the ages of 16 to 24, Nacogdoches also reported a total crash count of 21 drivers killed and rates of 5.49 and 2.35 for drivers killed with no alcohol or any alcohol reported, respectively. Not so far behind was Walker County with a similar population count, crash rates and counts, and the similar situation whereby a State University was located within the county. Although it is not possible to conclude from the current data whether or not the location of a university is solely responsible for the high rates and counts, the higher percentages of younger drivers and the implication that some college students engage in destructive drinking behaviors should not be ruled out.

For 2005, 5 counties were returned as containing relatively high crash rates and a high percentage of drivers aged 16 to 24. However, Erath County was the only county with a population above 10,000 estimated licensed drivers or even total population. The other 4 counties had a combined crash count of 5 drivers killed with a positive BAC. Fortunately for Texas though, the rates returned were all high only because those counties have relatively low populations.

The other age group of concern, G6 or estimated licensed drivers aged 65 +, was studied in the same fashion as 16 to 24 year olds. Moran’s I results showed fairly strong clustering indicated by high values corresponding to 0.2383 for 1996, 0.2558 for 2000, and 0.2334 for 2005. Across those three years, GeoDa cluster maps indicated a large area of counties with high-high clustering located in the central and northern Texas Regions. Other areas of interest were those low-low clusters centered around Dallas/Tarrant and Harris Counties, and counties located along the Texas-Mexico border in the southern portion of Texas. From the GeoDa results it was apparent that an east versus west division of ages was occurring whereby a younger population was present along the Texas-Mexico border and eastern regions, while the west and northwest portions were represented by a higher percentage of older drivers. This information was summarized on Figure 23 where cluster map results are displayed for 1996 to 2005.

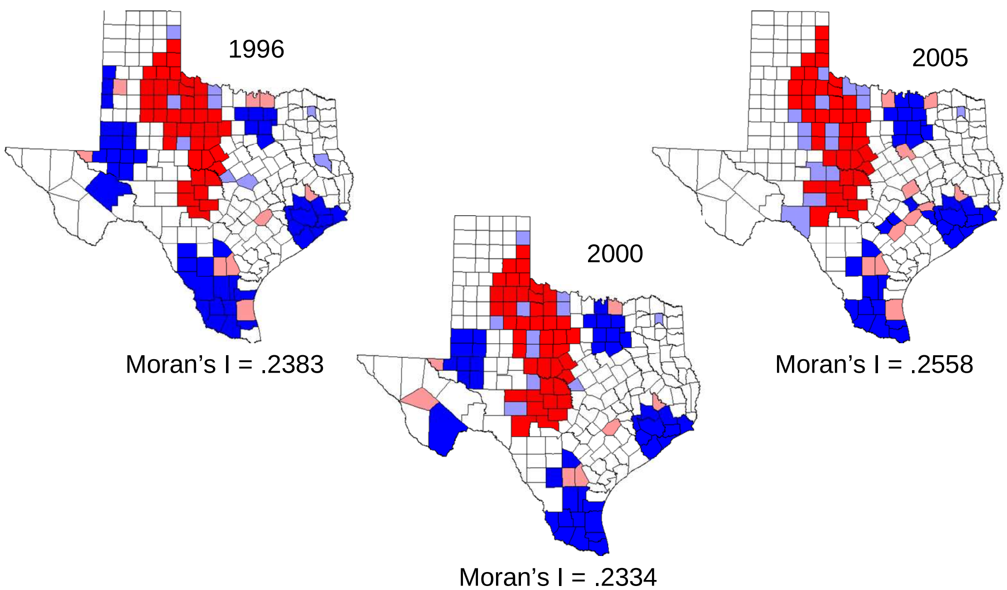

Figure 23. GeoDa Clusters and Moran's I Values. Estimated Percent of Licensed Drivers Aged 65+ Per Counties.

The interpolation and choropleth maps displayed various interesting trends regarding the population distribution of estimated licensed drivers aged 65+. Firstly, the number of estimated licensed drivers in this age group increased by 339,434 people over the 10 year period studied. This increase in older drivers was accompanied by a decrease in crash rates for that age group from 1996 to 2005 (Baby Boomers with driving experience joining this age group over time). However, the rates remained proportionally high considering the fact that as a group, the estimated number of drivers aged 65+ represented themselves as a low percentage of the total estimated licensed drivers in Texas. This increase in the number of estimated licensed older drivers was apparent in the maps created in ArcMap for both the choropleth and interpolation maps. In general, Counties including and surrounding Harris/Tarrant, Dallas, Travis, Bexar (Texas Urban Triangle), and those counties located along the Texas-Mexico border had lower percentages of estimated licensed drivers aged 65+. Counties centered within the Texas Urban Triangle or located in central, western, and northern Texas generally represented higher percentages. In 1996, ArcMap labeled 82 counties as having 20.01% or more estimated licensed drivers in that same age group. However, that number had fallen to 66 by 2000 and had increased again in 2005 where 105 counties were returned.

In order to try to represent this situation in the context of crashes where a driver was killed and reported to have had a positive BAC, a query similar to the one employed for estimated licensed drivers aged 16 to 24 was typed in ArcMap. Furthermore, the 20 counties with the highest percentages of drivers in G6 were arranged in order to identify any trends. Contrary to the results discussed in the 16 to 24 age groups where drinking and driving was reported as a serious problem in the literature, the older drivers were expected to show just the opposite assuming they would refrain from being involved in that type of destructive behavior. However, even though drivers aged 65+ represented a smaller percentage of both the total driving population and directly compared to drivers aged 16 to 24, there were many more counties returned under the query due to the fact that most counties with high percentages of older drivers had fairly low population counts. Also, the 16 to 24 group included 9 possible years of age whilst the 65+ group represented many more possible years. In other words, even if the query employed 20.01 % rather than 16.01 % under the percentage of drivers required by the query, a larger number of counties were returned in contrast to the query applied for younger drivers. For example, 23 counties were returned as having greater than 16% of their population aged 16 to 24 and having an above average crash rate for 1996. In contrast, 62 counties were returned for the exact same query applied to drivers aged 65 and above. However, of those 62 counties, more than half had total estimated licensed driver counts below 10,000.

For 1996, Hill County, with 22.32% of its estimated licensed drivers aged 65+, topped the list with 12 fatalities in terms of total counts for drivers killed where a positive BAC was established. That same number represented a rate of 6.18 drivers killed per 10,000 licensed drivers, although the estimated licensed population was below 20,000. However Llano County, with an estimated licensed driver population aged 65+ at 39%, had the highest percentage of that age group. With an estimated total licensed population of 8445, Llano County reported 4 fatalities where alcohol was reported for the driver killed, and a rate of 4.74 drivers killed per 10,000 licensed drivers. Besides those two Counties, Grayson, Polk, Rusk, Wharton, and Van Zandt Counties were significant in terms of their high crash counts, older driver percentages, and population being greater than 25,000 estimated licensed drivers. The rest of the counties had both significantly lower population counts and crash counts, although their crash rates were misleadingly high due to their lower population counts.

For 2000, Van Zandt County reported the highest number in terms of crash counts where a driver was killed and reported to have had a positive BAC. With an estimated 19.48% of the Counties population being aged 65+ and having a driver's license, 7 fatalities representing a rate of 2.25 drivers killed per 10,000 licensed drivers was recorded. Interestingly, the top 5 counties in terms of total crash counts where a +BAC was recorded more than doubled when alcohol was not found in the deceased drivers system. For example, Van Zandt County reported 16 drivers killed where no alcohol was present versus 7 where a +BAC was recorded. This trend was also apparent in 1996 for several counties, as well as for several more counties which were reported by the query for 2000 and 2005.

For 2005, Llano County reported just below 40% of its population in the 65+ age group estimated to have a driver's license. With an estimated total licensed population of 11,908, Llano County reported a rate of 2.52 drivers killed per 10,000 licensed drivers where a +BAC was established. Gillespie County also placed itself in the query with an estimated 31.92% of its licensed population aged 65+. Interestingly, Gillespie County was another occurrence where more than twice the crashes recorded did not involve any alcohol. In fact, crashes where no alcohol was recorded amounted to 11 versus 2 for crashes where a positive BAC was established. This trend was again apparent in Van Zandt and Henderson Counties. Interestingly, this seemed to correlate to the literature where it was explained that older drivers often refrained from making poor decisions before driving. Rather, older drivers were more often involved in crashes due to causes unrelated to having consumed alcohol prior to operating a motor vehicle. Unfortunately, this conclusion cannot be set in stone considering the fact that the data did not specify any information concerning the drivers exact age, gender, etc.

4.3 GENDER

Since gender was discussed as a subject under which crashes could be related to or better understood, the same procedures employed above for the age and crash data were applied to the estimated licensed drivers who were male or female. However, interpreting gender as a possible connection to crash rates or crash counts was rather difficult considering the fact that, once more, the gender of the driver killed was not known. Furthermore, the percentages which represented gender at the county level were often so close to 50% that a possible connection was made difficult to establish.

Firstly, the gender difference in Texas, as a percent representing the total population, displayed a 1% difference between males and females for 1996. In other words, the percentage of each gender for 1996 was 49.5% for males and 50.5% for females. By 2005, that percentage was nearly 50%, similar to 1996. Thus, in that regard Texas was comparable to the national average and considering any kind of conclusion regarding a potential correlation to the count or rate of traffic crashes at this level was not possible. Nonetheless, some differences were established at the county level.

Regarding gender, GeoDa returned 6 possible different results where males and females were analyzed for possible clusters and their significance. However, since the reciprocal results were true for females, only males are discussed in regard to the GeoDa results. Males returned Moran's I values of 0.0853, 0.0357, and 0.0358 respectively for 1996, 2000, and 2005. Although those values indicated a fairly weak association of clusters, a few specific clusters appeared over the 3 separate years. In 1996, a large cluster of counties located along the Texas-Louisiana border indicated a low-low percentage of males. In 2000, that same cluster had diminished in terms of county counts although it was still discernable from the cluster map. By 2005, that same cluster of low-low relationships had once more diminished, although it was still comprised of 10 counties, down from 20 in 1996 and 12 in 2000. Another notable cluster included a high-high and low-high cluster of counties along the Gulf Coast. Specifically, those counties surrounding Brazos and Anderson County were discerned as having high-high percentages of males whilst counties included in the 2nd order of the Queens Contiguity were comprised of low-high relationships with their neighboring counties. By 2005, the cluster of high-high and low-high relationships had diminished in county counts. Finally, a few scattered counties in west Texas were represented as high-high, low-low, high-low, and low-high. However, those last clusters were likely insignificant considering the fact that they appeared as single counties, thus representing themselves accordingly in contrast to the low Moran's I values.

In 1996, 3 counties had greater than 60% of their estimated licensed driver population as males. Hartley County, with an estimated 62.3% male drivers, had approximately 2,060 males out of a total population of 4,895. Located in north Texas near the town of Amarillo, Hartley County reported no fatal crashes for the 3 years where a driver was killed and a positive BAC was established for the victim. Except for Anderson County, the other counties either had very low population counts or had percentages so close to 50% that it did not make sense to analyze them further. However, Anderson County was worth noting due to the fact that it had 60% of its population listed as being male and licensed to drive. Anderson County reported a population of 52,301 people and 34,189 estimated licensed drivers. Anderson County reported an above average crash rate of 4.39 and 2.05 for drivers per 10,000 licensed drivers killed where either none or any alcohol was reported, respectively. Those numbers translated to 15 and 7 fatalities where drivers were killed in traffic crashes with either none or any alcohol reported.

In 2000, regarding high percentages of estimated male drivers per county, 6 counties had an estimated 60% or more percentage of their licensed population as being male. Of those 6 counties, 3 were rural and 3 were suburban. Of the rural counties, 2 had no fatalities where alcohol was recorded (Concho and Hartley counties). The third one, Mitchell County, had 1 fatal crash, but its population was only 3,966 people so it was not meaningful. The suburban counties with more than 60% males included Anderson, Walker, and Jones Counties. Walker County, with 60.2% males estimated to be licensed drivers, reported above average crash rates of 4.26 and 2.01, for fatal crashes with either none or any alcohol. Interestingly, this county was also discussed as having a high percentage of young drivers in previous sections of this research. As for Anderson and Jones County, their crash rates were comparable to the state average for the year 2000.

In 2005, 4 counties were reported as having more than 60% of their population as estimated males and licensed drivers. Anderson County, with 60.8% males, was returned as being the only suburban county whilst the other 3 were rural with low populations and a total of 1 fatal crash where alcohol was reported (for Mitchell County with 9,576 total population). Though only slightly above the state of Texas' average for that year, Anderson County reported 14 and 6 crashes, as well as 3.9 and 1.67 drivers killed per 10,000 licensed drivers where either none or any alcohol was reported, respectively for counts and rates.

In trying to find trends with the female population, counties with high percentages of estimated licensed females never exceeded 53.8% of their estimated licensed population across all 3 separate years. Thus, attempting to explore their crash rates, counts, and possible correlations to high percentages was not feasible. However, it is worth noting that counties which were reported as being in the top 20 for higher proportions of female drivers across all three years had noticeably lower crash counts and rates. Specifically, nearly 10 out of 20, for each of those separate years, had 0 crashes where a driver was killed and a positive BAC was reported. Nonetheless, the percent difference between males and females was too close to make any kind of conclusion or assumption.

Finally, since the state average was so close to 50/50 across all three years, the choropleth and interpolated maps showed little spatial and temporal variation. This was especially true for estimated female drivers whereby most of the counties were represented by a single color definition. The choropleth and interpolated maps for male drivers were also similar in the sense that little variation was apparent across all three years. These results can be seen in Figure 24 where it is evident that, aside from a few clusters discussed in the GeoDa results, the population distribution was quite simply evenly distributed across all three years.

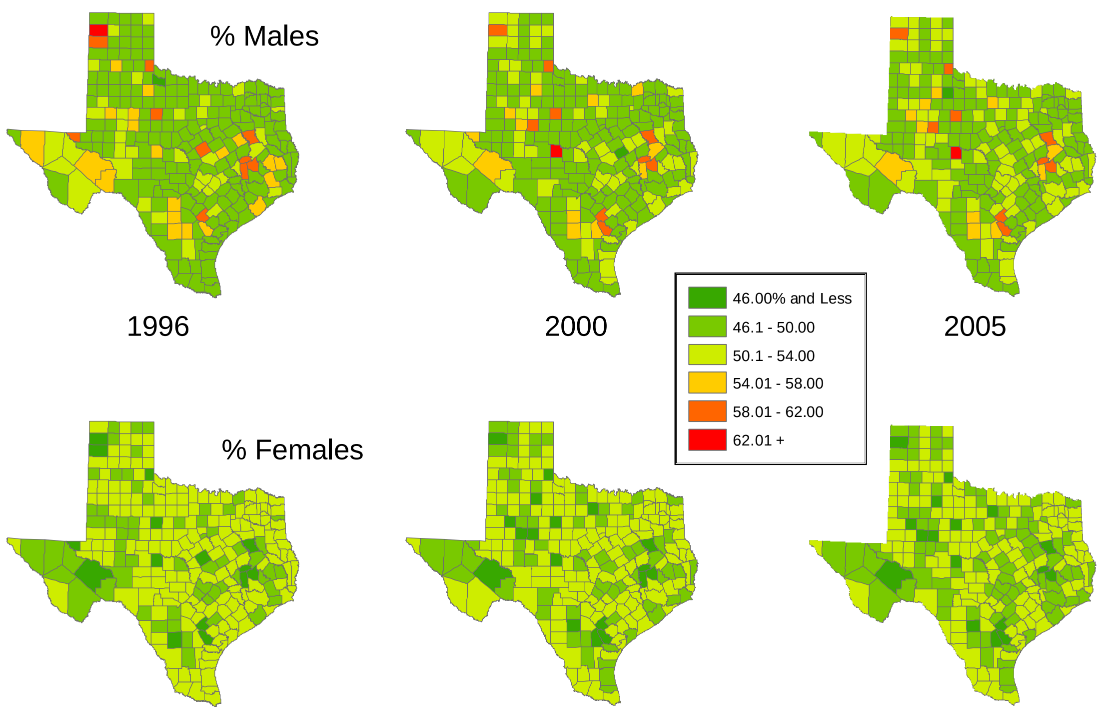

Figure 24. Percent of Estimated Drivers by Gender. Top = Male, Bottom = Female.

4.4 DISCUSSIONS

This study clearly defines spatial patterns of both crash counts and crash rates across the state of Texas. Through the incorporation of spatial statistics, choropleth maps, and interpolated data, a clear understanding of the spatial variation is discernable across both space and time. Concerning crash counts, there is clearly a difference between the western and eastern half of the state whereby the eastern half of the state accounted for over half of the driver fatalities. Also, counties along the Texas Urban Triangle consistently report high crash counts across the entire study period. Furthermore, the top 20 counties, in terms of crash occurrences where either none or any alcohol was recorded, account for approximately 50% of the state's total driver fatalities. That number may not seem surprising considering the fact that most of the population resides within those same counties. However, the perspective that if 20 of those 254 counties made a greater effort to reduce crash counts, it would likely have a huge impact on the state's total outcome. After all, Texas is consistently reported as being at or near the top in terms of total fatalities on the road. A push for more stringent laws in those 20 counties could greatly reduce those fatalities.

Although the spatial variation was similar for all three years in regards to crash counts where either none or any alcohol was reported, some counties with low population were represented as having particularly high crash rates in comparison to those counties which had high counts due to high populations. For example, in 1996 Hill County was reported as having had a crash rate, where a driver was killed and alcohol present, of 6.18. In most cases, a rate significantly higher than the state average of 1.35 drivers killed per 10,000 estimated licensed drivers in 1996 might have been due in fact to a low crash count being relative to a low population count. Thus, Hill County, which was the 15th highest ranked county in terms of drivers killed with a positive BAC reported, had a total population of 29,538 people, compared to McLennan County which was 14th with a much greater population of 202,679 people. Furthermore, Hill County had a crash rate where no alcohol was reported of 10.30, more than 3 times the average for Texas that year. This special situation whereby a county with a relatively low population produced high crash counts and rates was present across the three years studied for several different counties. A further study of individual counties having high rates across a greater number of individual years would be useful in the sense that possible spatial and temporal patterns may exist and be identifiable. However, for this study it was difficult to establish any patterns in regards to individual counties since 3 separate years were studied at separated intervals rather than every single year for a specified amount of time. Overall, crash rates appeared less clustered and less significant both statistically and visually through GeoDa and ArcMap, relating the information that although there were counties with excessive crash counts every year (Harris County), their rates were often at or below the state average.

Concerning age as a possible factor, the state-level data clearly correlated to some of the discussions found in the literature for that matter. Specifically, the tables representing traffic fatalities in comparison to total population counts for those same age groups clearly indicated a wide gap between the populations of 16 to 24 and 65+ year olds and their corresponding crash rates per 100,000 population counts. This was similar to results discussed by NHTSA reports, McGwin and Brown (1999), Evans (2004) where younger and older drivers were often singled out as having higher crash counts and rates in comparison to their actual licensed populations. However, the county data employed for this study was less apparent and harder to interpret for possible causations.

Specifically, counties which contained higher percentages of estimated drivers between the ages of 16 to 24 or 65+ either had low population and high rates or vice versa. In others words, since the age of the driver killed with either any or no alcohol was not reported, conclusions or even assumptions were difficult to make.

Regarding 16 to 24 year olds, several counties discussed in the results section were found to have large universities. However, although some of those counties had above average rates in some years, it is hard to assume that those crashes which involved either none or any alcohol were related to the fact that a higher percentage of 16 to 24 year olds were present in those counties. Thus, with the aggregated data employed for this study, it was not possible to correlate counties with higher percentages of younger drivers with higher crash rates where either none or any alcohol was used.

The same was true for 65+ year olds. However, a difference was apparent in an overall lower count per county of total populations where older drivers resided, in comparison to counties which contained higher percentages of younger drivers and had higher population estimates. Also, though possibly related to the fact that less people resided in those counties, rates were higher and counts were lower for both crash types for older drivers in comparison to younger drivers. Nonetheless, assumptions in this regard were dismissible. The exception existed for those few counties discussed in the results section that clearly had high rates in comparison to their estimated licensed drivers and total populations. Even then, this study did not take into account enough years or detail to really understand connections between age and traffic crashes based on the available aggregated data set.

With reference to GeoDa and AcrMap results, the age distributions were successfully mapped and assessed statistically in respect to spatial autocorrelation. GeoDa results varied from low to high Moran's I values for younger and older drivers, respectively. This indicated that younger drivers were more spread out and less clustered than older drivers in the context of a counties age distribution. Spatially, higher percentages of younger drivers were mostly concentrated along the Texas-Mexico border as well as several clusters scattered around Brazos and Lubbock Counties, and the I-35 side of the Texas Urban Triangle. A lower concentration of younger drivers was apparent in Central and far north Texas around the city of Amarillo. Temporally, 1996 and 2005 were similar, while 2000 appeared to have had a period where portions of central Texas had less young drivers than the other two years. For the 65+ year old estimated licensed drivers, a high concentration of older drivers was apparent in Central Texas and along some portions of the Texas-Louisiana border, as well as in some parts of southeastern Texas. Their percentages were lower for counties along the Texas-Mexico border and along the Texas Urban Triangle. Temporally, it appeared that the population represented itself spatially by showing an overall aging effect whereby many counties were bumped up into higher percentage classes in the choropleth and interpolated maps for older rather than younger age groups over the years. Figure 25 shows a clear pattern whereby 35 to 44 year olds decreased both in their numbers and the spatial variation under which they were categorized as comprising 20% or more of the counties total population. The opposite was true for 45 to 54 year old both in their numbers and the spatial distribution in higher percent compositions per counties.

Though there is no definite answer which can be extracted within this project, it is possible that the aging population might play a part in the fact that the percent of fatal crashes reporting alcohol have been declining. The literature explained that older and more mature drivers were less likely to drink and drive then the younger ones, who tend to make poor decisions and have less driving experience.

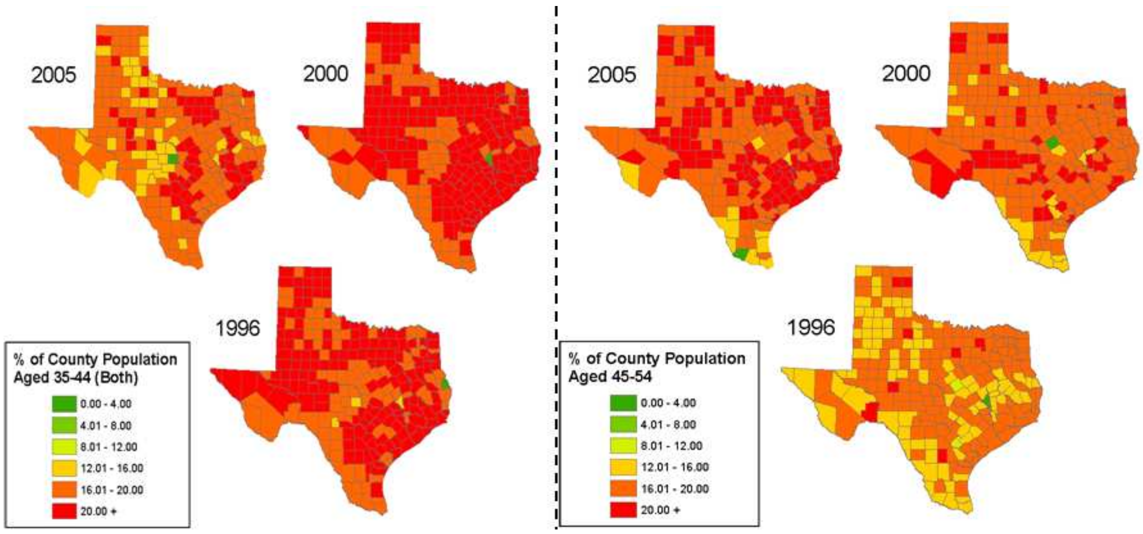

Figure 25. Percent per County Comparison of 35 to 44 and 45 to 54 Year Olds.

Compared to age as a possible factor, gender was even harder to interpret considering the fact that most counties were so close to having a 50/50 spread between male and female estimated licensed drivers. At the state level, it was not rational to link a possible explanation as to having a greater percentage of one gender possibly affecting the overall outcome of fatal crashes. Regarding spatial autocorrelation, the Moran's I values were low, thus confirming the idea that spatial clustering of one gender was not significant. However, the cluster maps did reveal a few spots where males and females were represented by higher percentages as compared to their surrounding neighboring counties. The largest cluster was located along the Texas-Louisiana border in north eastern Texas. That cluster was represented by low-low clustering for males and high-high clustering for females. However, the higher percentage of females present in that cluster was misleading considering the fact that the percentage of females living in any given county rarely exceeded 50%. Aside from that cluster, several others existed where males were more numerous in some counties compared to their surrounding neighboring counties. This high-low or high-high instance was often the case in west Texas as well as a few counties inside the Texas Urban Triangle.

When counties with the 20 highest percentages of either males or females were compared, a few observations were made. For males, the top 20 counties had a greater percentage than females. For example, the highest percentage of females living in a given county was 53.8% for Falls County in 2000. For males, the highest percentage was 65.9% for Concho County in 2005. Also, the rates and counts for both types of crashes were very similar when counties with higher percentages of males versus females were compared. However, the top 20 counties corresponding to females had much greater population counts for 2000 and 2005 than those containing greater percentages of males. For 2000, counties with a high percentage of males had more crashes where no alcohol was reported than counties with a higher percentage of females, but the same number of crashes where a positive BAC was established for the deceased driver (77 versus 69 and 39 versus 39). However, the rates were higher for counties with higher percentages of females than males with 3.57 versus 2.84, and 2.36

versus 1.49. Although rates were higher for females in 2000, they resided in more populous counties. Also, their crash rate averages for the top 20 counties were skewed by Cottle County, which reported rates of 16.27 with only 2 fatal crashes and a total county population of 18,576. For 2005 and 1996, the situations were similar except that 1996 had slightly fewer estimated female drivers than males but very similar counts and rates.

Either way, since it was not known whether the driver killed was male or female and the county percentages were so close to 50%, it was difficult to link those variables to crash counts and rates from the tables as well as the maps created. As was discussed in the results section, the interpolated centroids and choropleth maps were similar spatially and temporally. Spatially, those counties which had 54% or more estimated licensed male drivers were scattered around Texas, though most of those approaching 60% were in west Texas. For females, the spatial variation was spread evenly across all of Texas. Temporally, there was very little variation across the three years studied whereby clusters appeared or faded.

Overall, though the independent and dependent variables had distinct patterns across space and time, there was not enough evidence within the aggregated dataset to clearly identify a link between age or gender and fatal crashes where a driver was killed with none or any alcohol present in their system. Even though the literature clearly reports these variables as factors related to traffic crashes, the nature of this dataset did not contain enough detail to draw significant evidence. However, this research was successful in analyzing spatial, temporal, and spatial autocorrelation patterns across the selected years. Spatial clustering of crashes provided clear areas of concern whereby reoccurring high counts or high rates should be given further attention. Although it was apparent that the location and population of a county could be directly related to crash counts and rates, the results were expected except for several counties discussed above. However, in order to draw specific conclusions concerning those counties exhibiting higher rates and counts, a greater succession of years should be studied.

Based on these results, the overall decline in the percentage of crashes involving alcohol during the study period cannot be attributed to the chosen variables alone. Even though there is an overall aging of the population occurring across space and time, that factor alone is not enough to draw such a powerful conclusion. Thus, the decline is likely a combination of the studied variables with several others including, for example, race/ethnicity, traffic laws, increasing awareness and education, more efficient vehicles available, and possibly the decline of rural populations towards urban and suburban ones.

CHAPTER V

SUMMARY AND CONCLUSIONS

5.1 SUMMARY

The aim of this research was to analyze geographic and demographic patterns of alcohol-related fatal traffic crashes from 1996 to 2005. By linking age, gender, location, and population aggregated datasets to similarly aggregated crash counts where a driver was killed in a traffic crash where either none or any alcohol was present in the driver's bloodstream, an analysis of those independent factors spatially, temporally, and as they might correlate to one another was assessed. This study was conducted at the county level for the State of Texas which has continuously reported higher counts and rates than the national average. This was done by first providing a descriptive statistical analysis on the datasets. Secondly, an analysis of the global Moran's I and LISA was conducted on the datasets in order to explore spatial clustering of counties and their corresponding variables. Finally, choropleth and isoplethic maps were created to further understand spatial and temporal patterns in the datasets.

The results found a clearly defined and unique spatial pattern of alcohol-related fatal crashes in Texas. Much like what was anticipated through the course of the research, crash counts were positively correlated to population counts and county memberships within the Texas Urban Triangle. Rather than display a large number of rates, trends through time, graphs, and tables, this thesis displayed crash counts and crashes as they appeared across space and time. Though there was not a whole lot of variation through time in regards to drivers killed in fatal alcohol-related crashes, this research found several important points to be raised. For one, in the context of traffic safety, this is a great opportunity for raising awareness considering the fact that if high crash counts or rates are consistently occurring in the same counties over a particular amount of time, this can point out that something can and must be done to alleviate the situation. For example, the largest number of crashes, which were directly tied to the Texas Urban Triangle Counties, not only occurred in this region but also re-occurred every year within the study period for nearly the same number of counties. With that said if one was to take action and try and further reduce alcohol-related fatalities by campaigning a specific message or enforcing a particular law or set of laws, enacting them could have a positive effect on those counties which reported high crash counts through the study period. This is important considering the fact that Texas was repeatedly reported as being at the top or very close to the top during the study period in comparison to the national averages and other states.

This made this particular spatial and temporal analysis of aggregated counties a critical study in traffic safety. There were clusters of counties that had high crash counts which were surrounded by counties with similarly high counts. This helped to further point out that if action was to be taken to reduce alcohol-related fatalities, those counties located in those specific clusters could and should be primary targets for further research.

For the aforementioned reasons, this project fills in several gaps, one of which includes the fact that this is the first known project of its kind whereby aggregated datasets were studied in this fashion using those specific softwares and techniques in this particular combination. Furthermore, the techniques discussed in regards to the use of mapping and spatial autocorrelation techniques, broaden the amount of information about the selected crash dataset for the state of Texas. The information provided by this research is especially useful if considering the spatial and temporal variations at the county level in respect to crash counts and crash rates. Where most of the traffic safety literature refrains from incorporating digital maps as powerful exploratory and explanatory mediums, this research emphasized the importance of these significant innovations and their contributions. Finally, this research provided a framework for future studies.

5.2 FURTHER RESEARCH AND LIMITATIONS

This research contributes several techniques which can be attributed to traffic safety mapping in general. The concepts regarding spatial autocorrelation and interpolation, though known in the field, are coupled in a way that spatial and temporal information can be readily drawn from and studied for similar datasets. This research could be taken further by incorporating a more specific dataset containing more information about the crash locations. Also, knowing more specific information about the crash regarding, for example, age, gender, and time could greatly alleviate the steps taken towards drawing solid conclusions.

Until every vehicle is equipped with a crash monitoring and recording system, police officers are trained to record the level of data needed by GIS analysts, and organizations are organized to the point that data is up to date and readily available, traffic safety in regards to the research conducted here will have several limitations. This need of a more detailed dataset was perhaps the greatest limitation in this study. The fact that aggregated datasets were used at a relatively small scale, for example with more detail than state or national-level studies but less than census tracts, made it difficult to conclude in favor of the proposed hypothesis that crashes were declining as a result of changes spatially and temporally partially respect to those chosen variables. Another limitation was due in part to the complexity of crashes. For example, simply attributing age, gender, location, and population as independent variables cannot fully explain the spatial and temporal variations of traffic crashes. Another limitation stems from the fact that, although crashes were a positive BAC was established were considered alcohol-related, the possibility that the driver died from consequences completely unrelated to alcohol is a possibility. It is believed that a BAC as low as 0.01% can have adverse effects for some, although for most a BAC that low often does not impair driving. It is also important to point out that Texas has wet, partially wet, and dry county laws in regards to allowing or prohibiting the sale of alcoholic beverage. Though the definition of each, respectively, seems rather obvious whereby wet counties, the technicalities and legislations involved in formulating which county falls under which category is rather difficult. For example, various counties may be partially wet in that some of the cities within it are wet or dry. This makes it difficult to assign a single category to various counties. This research also did not put into account the various laws which were discussed in the literature review (21 and up, zero tolerance etc). Finally, aside from Moran's I values indicating the significance of clustering, conclusions were drawn without the use of rigorous statistical procedures commonly discussed within the traffic safety literature.

5.3 CONCLUSION

Although traffic safety is a broad and extensive field of research, geography plays an important role in understanding complex relationships between the many variables that play a role in traffic crashes. Understanding the spatial variation of the dependent and independent variables through time and space is a critical part of traffic safety research. GIS is becoming more than a simple tool adopted in traffic safety research and the advances that have been made in recent years are a sure sign that geography continues to contribute imperative information to the field. Digital maps continue to gain importance and respect as representations of space and designated factors which are critical to traffic safety. These methodologies built from advances in geographic knowledge and techniques will continue to contribute a great amount of information for a variety of sectors from civil engineering to traffic safety administration.

REFERENCES

Al-Balbissi, A. H., 2000. Role of gender in road accidents. Traffic Injury Prevention 4, 64-73.

Amoros, E., Martin, J. L., Laumon B., 2003. Comparison of road crashes incidence and severity between some French counties. Accident Analysis and Prevention 35, 537-547.

Clark, D. E., 2003. Effect of population density on mortality after motor vehicle collisions. Accident Analysis and Prevention 35, 965-971.

Dissanayake, S., Lu, J., 2002. Analysis of severity of young driver crashes: Sequential binary logistical regression modeling. Transportation Research Record: Journal of the Transportation Research Board 1784, 108-114.

Evans, L., 2004. Traffic Safety. Science Serving Society, Bloomfield Hills, MI.

Farrow, J. A., Brissing P., 1990. Risk for DIW: A new look at gender differences in drinking and driving influences, experiences, and attitudes among new adolescent drivers. Health Education Quarterly 17 (2), 213-221.

Flahaut, B., Mouchart, M., Martin, S. E., Thomas, I., 2003. The local spatial autocorrelation and the kernel method for identifying black zones a comparative approach. Accident Analysis and Prevention 35, 991-1004.

Fridstrom, L., Ingebrigtsen, S., 1991. An aggregate accident model based on pooled, regional time-series data. Accident Analysis and Prevention 23 (5), 363-368.

Gmel, G., Bissery, A., Gammeter, R., Givel, J-C., Calmes, J-M., Yersin, B., Daeppen, J-B., 2006. Alcohol-attributable injuries in admissions to a Swiss Emergency room – An analysis of the link between volume of drinking, drinking patterns, and preattendance drinking. Alcoholism: Clinical and Experimental Research. 30 (3), 501-509.

Goodchild, M. F., 1992. Geographical data modeling. Computers and Geosciences 18, 401-408.

Goodchild, M. F., 2000. GIS and transportation: Status and challenges. Geoinformatica 4 (2), 127-139.

Gorman, D. M., Huber JR J. C., Carozza, S. E., 2006. Evaluation of the Texas 0.08 BAC law. Alcohol & Alcoholism 41 (2), 193-199.

Hauer, E., 1992. Empirical Bayes approach to the estimation of "Unsafety": The multivariate regression method. Accident Analysis and Prevention 24 (5), 457-477.

Holubowycz, O. T., Kloeden, G. N., McLean A. J., 1994. Age, sex and blood alcohol concentration of killed and injured drivers, riders and passengers. Accident Analysis and Prevention 26 (4), 483-492.

Keane, C., Maxim, P. S., Teevan, J. J., 1993. Drinking and driving, self-control, and gender: Testing a general theory of crime. Journal of Research in Crime and Delinquency, 30 (1), 30-46.

Kim, K., Levine, N., 1996. Using GIS to improve highway safety. Computer, Environment, and Urban System 20 (4/5), 289-302.

Laapotti, S., Keskinen, E., 2004. Has the difference in accident patterns between male and female drivers changed between 1984 and 2000?. Accident Analysis and Prevention 36, 577-584.

Lam, L. T., 2002. Distractions and the risk of car crash injury: The effect of drivers' age. Journal of Safety Research 33, 411-419.

Levine, N., Kim, K., Nitz, H. L., 1995. Spatial analysis of Honolulu motor vehicle crashes: 1. spatial patterns. Accident Analysis and Prevention 27 (5), 663-674.

Levine, N., and K. E. Kim, 1998. The location of motor vehicle crashes in Honolulu: A methodology for geocoding intersections. Computers, Environment and Urban Systems, 22 (6), 557-576.

Li, Linhua., Zhu, Li., Sui, Daniel Z., 2007. A GIS-based Bayesian approach for analyzing spatial-temporal patterns of intra-city motor vehicle crashes. Journal of Transport Geography 15, 274-285.

Loo, B. P. Y., 2006. Validating crash locations for quantitative spatial analysis: A GIS-based approach. Accident Analysis and Prevention 38 (5), 879-886.

Massie, D. L., Green, P. E., Campbell, K. L., 1997. Crash involvement rates by driver gender and the role of average annual mileage. Accident Analysis and Prevention 29 (5), 675-685.

McGwin, G. Jr., Brown, D. B., 1999. Characteristics of traffic crashes among young, middle-aged, and older drivers. Accident Analysis and Prevention 31, 181-198.

Mercer, W. G., 1987. Influences on passenger vehicle casualty accident frequency and severity: Unemployment, driver gender, driver age, and restraint device use. Accident Analysis and Prevention 19 (3), 231-236.

Messner, F. S., and Anselin, L., 2004. Spatial analysis of homicide with areal data. Spatially Integrated Social Science Ed. Goodchild, M., and Janelle, D., 127-144.

Miller, T. T., Lestina D. C., Spicer R. S.. 1998. "Highway crash costs in the united states by driver age, blood alcohol level, victim age, and restraint use". Accident Analysis and Prevention. 30(2): 137-150.

Miller, T. T., Spicer, R. S., Levy, D. T.. 1999. "How intoxicated are divers in the United States? Estimating the extent, risks and costs per kilometer of driving by blood alcohol level". Accident Analysis and Prevention. 31: 515-523.

Nyberg, A.. Gregersen, N. P., 2007. Practicing for and performance on drivers license tests in relation to gender differences in crash involvement among novice drivers. Journal of Safety Research 38, 71-80.

Pawlovich, D. M., Souleyrette R. R., Strauss, T., 1998. A methodology for studying crash dependence on demographic and socioeconomic data. Transportation Conference Proceedings, 1-7.

Smink, B. E., Ruiter, B., Lusthof K. J., Gier, de J. J., Uges, D. R. A., Egberts, A. C. G., 2005. Drug use and the severity of a traffic accident. Accident Analysis and Prevention 37, 427-433.

Stevenson, M., R. D. Brewer, and Lee, V., 1998. The spatial relationship between lincesnsed alcohol outlets and alcohol-related motor vehicle crashes in Gwinnet County Georgia. Journal of Safety Research 29 (3), 197-203.

Subramanian, R., 2002. Transitioning to Multiple Imputation – A new method to estimate missing blood alcohol concentration (BAC) values in FARS. Mathematical Analysis Division, National Center for Statistics and Analysis http://www-nrd.nhtsa.dot.gov/Pubs/809-403.PDF (Accessed 7-10-08)

Thill, J. C., 2000. Geographic information ystems for transportation in perspective. Transportation Research Part *C* 8, 3-12.

Thomas, I., 1996. Spatial data aggregation: exploratory analysis of road accidents. Accident Analysis and Prevention 28 (2), 251-264.

US Department of Transportation, National Highway Traffic Safety Administration, 2005. Traffic Safety Facts: 2005 Data Alcohol. *Washington DC*. http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/ncsa/tsf2005/2005tsf/810_616/images/alcohol.pdf (accessed 4-29-07)

US Department of Transportation, National Highway Traffic Safety Administration, 2006. Traffic Safety Facts: 2006 Alcohol-Impaired Driving. Washington DC http://www-nrd.nhtsa.dot.gov/Pubs/810801.PDF (Accessed 5-30-08)

Valverde, J., and Jovanis, P., 2006. Analysis of fatal and injury crashes in Pennsylvania. Accident Analysis and Prevention 38 (3), 618-625.

Wageneaar, A. C., Toomey, T. L., 2002. Effects of minimum drinking age laws: Review and analyses of the literature from 1960 to 2000. Journal of Studies on Alcohol/Supplement 14, 1-21.

Winn, R. C., Giacopassi, D., 1993. Effects of county-level alcohol prohibition on motor vehicle accidents. Social Science Quarterly 74 (4), 1-11.

Yagil, D.. 1998. “Gender and age-related differences in attitudes toward traffic laws and traffic violations”. Transportation Research Part F. 1: 123-135.

Zador, P. L., Krawchuk, S. A., Voas, R. B., 2000. Alcohol-related risk of driver fatalities and driver involvement in fatal crashes in relation to driver age and gender: An update using 1996 data. Journal of Studies on Alcohol 61 (3), 387-395.

VITA

Gabriel A. Rolland received his Bachelor of Science degree in geography from Texas A&M University in December 2005. Prior to graduation, he studied geography while focusing on GIS and cartography. Furthermore, he earned several awards including an induction into the International Geographical Honor Society in 2004 and a Distinguished Student Award in 2003. He entered the geography program at Texas A&M University in August 2006 and received his Master of Science degree in geography on December 2008. His research interests include GIS analysis, traffic safety, remote sensing, and both physical and human geography. He plans to publish one or more research articles on these topics, focusing on alcohol crashes in Texas. Finally, he hopes to pursue an interest in meteorology concerning natural disasters and modeling, forecasting, and emergency preparedness.

Name:	Gabriel A. Rolland
Address:	10696 Edenoaks St., San Diego, California 92131
Email Address:	rllndg18@neo.tamu.edu
Education:	B.S., Geography, Texas A&M University, 2005 M.S., Geography, Texas A&M University, 2008