Applying Log-linear Models And GIS To Study The Safety Of Pedestrians And Bicyclists : A Case Study Of Orange County School Children

Sai Srinivas Chundi
University of Central Florida

APPLYING LOG-LINEAR MODELS AND GIS TO STUDY THE SAFETY OF PEDESTRIANS AND BICYCLISTS : A CASE STUDY OF ORANGE COUNTY SCHOOL CHILDREN

By

Sai Srinivas Chundi

B.E (Hons.), Birla Institute of Technology and Science, Pilani, India, 2002

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Civil and Environmental Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2005

Abstract

Pedestrian /bicycle safety of school children has been a growing menace that has been attracting attention from transportation professionals, school boards, media and the community all over the country. As such there has been a necessity to identify critical variables and assess their importance in pedestrian/bicycle crashes occurring in and around school zones. The current study is an endeavor in this direction.

The literature review identified some studies that were conducted on school zone safety related to pedestrian/bicyclist crashes. Most of the studies pertained to crashes with all age groups. There have been few studies with emphasis only on school aged children. In this study we focus on pedestrian age group (4 to 18 years), the time of the day when the school children are expected to be commuting (6:30 AM to 10:00 AM and 1:00 PM to 5:00.PM), the day of week (Monday through Friday) and the days when the school is opened (January 6th to May 31st and August 6th to December 21st). Geographical Information Systems was used to locate buffer zones around schools with higher crash incidence rates. The use of log-linear analysis has culminated in explaining the relationship between various variables and crash incidence or crash frequency

Crash data for this study was obtained in the form of crash database and GIS maps from the Department of Highway Safety and Motor Vehicles and the Orange County School Board respectively. Crash reports were downloaded from the CAR database of the FDOT mainframe website. The crash data was related to the GIS maps to visually depict the proximity of crashes to the school zones and thus identified risky schools and school districts. It was concluded from the spatial analysis that the incidence of crashes was higher at middle schools. In the log-linear analysis different models were

tested to explain the effects of driver characteristics, geometric characteristics and pedestrian characteristics on the crash frequency. It was found that driver age, number of lanes, median type, pedestrian age and speed limit are the critical variables in explaining crash frequency. By examining the levels of the variables that were significantly involved in the crashes we would get an insight on ways to explain and control pedestrian/bicyclists crashes at school zones. It is hoped that this thesis would make an active contribution in improving the safety of bicyclists and pedestrians in and around school zones and make the schools much safer for the children.

ACKNOWLEDGEMENTS

The quest for perfection is tough albeit an enjoyable one. I sincerely would like to thank my guide and advisor Dr. Mohamed Abdel-Aty, for paving the way for such a journey for perfection. This thesis stems up from the inspiration and advice that I have imbibed from him. I would also like to express my heartfelt gratitude to my thesis committee members, Dr. Essam Radwan, and Dr. Chris Lee, who have fine tuned my efforts that are reflected in this thesis. It would be an exaggeration to pass this effort as mine alone, as there are influencing efforts from my colleagues and friends, Ravi Chandra and Hari. I wish to thank them all. I would also like to convey special thanks to Ravi Chandra whose ideas and suggestions are the pillars to this thesis.

As a student, I was lucky to have found friends whose support encouraged me to get better as a student and a human being. I would like to thank Aparna, Rajashekar, Vidya, Sailaja, Sandeep, Piyush, Nishanth, Vamsi and Hima for their support and encouragement. Also, I would like to thank Bobby, Yan, and everyone else I have worked with at UCF. Last but not the least I would like to acknowledge the blessings of the Almighty, my parents and my brother back in India that helped me sail through my toughest times.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Background

The importance of pedestrian safety is particularly acute to transportation engineers and planners because of the unprotected nature of the individuals who embrace walking as their utilitarian or non-utilitarian mode of transportation. Pedestrian accidents are among the most common causes of death and serious injury to children in the developed world.

Road safety has increasingly been a vital topic for discussion, as traffic crashes have been identified as one of the top 10 causes of death in the United States of America (the World Almanac and Book of Facts, 1996). They are climbing up in the list of death causes, from no.9 in 1999 to an estimated no.3 in 2020. According to U.S. department of Transportation's National Highway Traffic Safety Administration (NHTSA) release on Highway fatalities in USA, for 2003, an estimated highest number of people were killed in traffic crashes since 1990 (NHTSA, 2004).

Florida has been consistently ranked as one of the worst states in terms of pedestrian and bicycle fatality rates. According to the Florida Department of Highway Safety and Motor Vehicles' (DHSMV) Traffic Crash Statistics Report, 2003 nearly 8000 pedestrian crashes have been reported only in the state of Florida. That is, an alarming 18% of all traffic fatalities in the state of Florida for the year 2003, were pedestrians. In addition to this, an estimated 5400 crashes, i.e., about 3% of the total fatalities were bicyclists. Also, about 11% of the total drivers involved in the crashes were under the age of 18. About 10% of the pedestrians and bicyclists killed in the crashes were under the age of 18, in other words school going age. Nearly 25% of the total pedestrians injured were under the age of 18.

This certainly creates a panic in the society as to how safe the streets are for the children to traverse. These figures reveal the seriousness of the problem and the need for immediate road safety measures. Also, the involvement of young school going children in such causalities needs more concern and attention to the current issue. There has been a substantial growth in the field of pedestrian and bicyclist's safety research in the recent past. Many road safety devices and studies haven been implemented across the nation, but in comparison to the gravity of the problem many more efforts need to be put in. There have been few studies with emphasis only on school going aged children. Very few studies mention about the effect of the time of day and age group on the frequency of crashes. No spatial analysis of the crashes involving pedestrians and bicyclists under the age of 18 using GIS has been done. Trying to understand the occurrence of the pedestrian and bicycle crashes in the younger age group would be very critical and challenging. Nevertheless, driver behavior is always a key factor in crash occurrence. These factors are supported and surrounded by many other factors like the roadway geometrics, traffic characteristics, environmental factors and even vehicle characteristics.

1.2 Problem Description and Objectives of Study

With increased media attention of crashes involving school children in Florida, related authorities are being more interested in trying to come up with various solutions to avoid further such crashes. The focus of this thesis is to identify the various factors that are critical to the crashes occurring in school zones. The outcome of this analysis would hopefully be used in making schools and pedestrians in school zones safer.

The problem that we are dealing with in this thesis is to *investigate and analyze the effect of various driver, pedestrian, traffic, geometric and vehicle characteristics on the risk of crash and crash frequency involving children (pedestrians and bicyclists) in Orange County, Fl.*

The research objectives of this thesis can be summarized as follows:

- Identifying pedestrian and bicycle crashes related to school children in the age group 4 to 18 years in Orange County, Florida.
- Analyzing the effect of various road characteristics on pedestrian and bicycle crashes.
- Identifying the significant factors that affect the crash frequencies involving pedestrians and bicyclists in the age group 4 to 18 years.
- Model crash frequencies for different types of crash categories using log linear analysis
- Creating a GIS based database that analyzes the crashes based on the school type and location of crashes

1.3 Organization of Thesis

This thesis report can be divided into three main analyses dealing with pedestrian and bicycle crashes in Orange County, involving children between the ages 4 and 18. The preliminary study (Chapter 4) describes the various driver, pedestrian, vehicle and geometric characteristics involved in the crashes. It states the facts and figures of the distribution of various factors on the crash frequency that happened in Orange County for the years 1999 to 2003. The chapter deals with the basic statistical analysis by developing various two- way contingency tables and carrying out some basic T-tests to check the effect of various factors on the crash frequency. It elaborates on the extraction of the data and selection of the important variables for further analysis.

Chapter 5 makes an effort to evaluate certain crash frequency modeling using log linear analysis which becomes the next logical extension of the study explained in chapter 5. The same crash data was used for building this model. The models developed in this chapter deal with splitting the crashes into two categories. They can be broadly classified as Driver related and Pedestrian related. That is, what kind of driver characteristics would be involved in a pedestrian crash and what kind of pedestrian characteristics would be involved in a pedestrian crash. Both the categories would involve geometric, traffic and vehicle characteristics. The interaction between driver characteristics and other characteristics excluding the pedestrian characteristics would be dealt with in the first model and the interaction between pedestrian characteristics and other characteristics excluding the driver characteristics would be dealt with in the second model. The models cannot be used to predict the chances of crash as we are not dealing with exposure. They can be used to analyze the relation of various factors in relation with drivers and pedestrians. The major advantages to be gained from the model-fitting techniques to be described are firstly that they provide a systematic approach to the analysis of complex multidimensional tables and secondly that they provide estimates of the magnitude of effects of interest; consequently they allow the relative importance of different effects to be judged. The study used crash data obtained from the Department of Highway Safety and Motor Vehicles, for 5 years, i.e. from 1999 to 2003.

The final analysis (Chapter 4) used the concepts of Geographical Information Systems. The Orange County base map and the school locations in Orange County were obtained from the Orange County Public schools website. The orange county base map was divided into the school districts based on the information provided in the school board website. Crash reports were obtained using the Florida Department of Transportation CAR database. The crash locations were marked on GIS. This process was done to look at the density of crashes in various school districts. Also, the effect type of school and the crashes was also seen. This concept of using GIS to understand better the crashes in and around school zones is very innovative and has not been used in any of the earlier studies.

2. LITERATURE REVIEW

Pedestrian and Bicycle Crashes have been a matter of discussion since a very long time. With the increase in the number of school children getting involved in crashes, this topic has gained even more momentum. The following literature review section mentions in brief about the various studies done till now related to pedestrian and bicyclist crashes particularly in the younger age group and also the statistical methodologies used to analyze them. The literature review section has been divided into two main categories. Studies related to Pedestrian and Bicycle Safety and Log-linear Analysis and GIS.

2.1 Pedestrian and Bicycle Safety

Gruenewald et al., 2004, proposed that multiple determinants of child pedestrian injury collisions such as rates of unemployment and locations of schools must be considered in evaluating the unique contributions of any one community feature to injury rates. They suggested that Schools are one stable geographic feature associated with regular, often concentrated periods of complex and congested traffic patterns. The objective of their study was to examine annual rates of child pedestrian injury in four California communities with a focus on the unique contribution of schools to injury risk. They predicted that annual numbers of child pedestrian injury collisions (both in-school and summer combined) would be greater in communities with higher youth population densities, more unemployment, fewer high-income households, and higher traffic flow. It was hypothesized that youth population density and its interaction with the number of schools in a given area would be related to greater rates of child pedestrian collisions during in-school months. An ecological approach was taken that divided the four communities into 102 geographic units with an average of 6321 people residing in each unit. Archival data on traffic flow, number of child pedestrian injury collisions and locations of schools were obtained from state agencies. Individual-level data were obtained from a general population survey conducted in the communities. The results showed that annual numbers of injuries were greater in areas with higher youth population densities, more unemployment, fewer high-income households, and greater traffic flow. Annual numbers of injuries during in-school months were greater in areas containing middle schools and greater population densities of youth.

Johnston et al., 2003 reported that pedestrian accidents are among the most common causes of death and serious injury to young children in the developed world. Recent pedestrian injuries figures from New Zealand, broken down into 5 years age groups, indicate that children aged 5–9 years accounted for the highest percentage of injuries (16%), followed by those aged 10–14 (13%) and 15–19 years (12%), with no other age group accounting for more than 7% (20–24 years). Pedestrian deaths, however, were highest for 15–19 years old (17%) and for 75–79 years old (14%) with the 5–9 and 10–14 years old age groups each accounting for only 3% of pedestrian fatalities. They employed a virtual reality (VR) system, using a head mounted display (HMD), to investigate road crossing behavior in children and young adults. There were 24 participants in total, 6 in each of the following age groups: 5–9, 10–14, 15–19 and >19 years. The experiment was composed of a total of 14 trials. In each trial there was a flow of traffic from the right-hand side. Younger children (aged 5–9 years) made the greatest number of unsafe road crossings and the oldest participants (aged >19 years) the fewest. Overall performance was better (fewer unsafe road crossings) in uniform speed than uniform distance trials, consistent with previous research suggesting that pedestrians base road crossing decisions on inter-vehicle distance rather than vehicle speed. ANOVA analysis was carried out to predict the results.

Harruff et al., 1998 performed a retrospective analysis of 217 pedestrian traffic fatalities in Seattle, WA, U.S.A. that occurred over a six-year period using medical examiner records with essentially all of the deaths examined by autopsy. The annual pedestrian fatality rate for the county averaged 2.0/100,000 for all ages and both sexes, and the age-specific rate varied from 1.0/100,000 for the 22-34 year age group to 1.5/100,000 for children under seven years and 7.0/100,000 for ages 70 years and older. Males had a 50% higher rate than females. Fatal accidents were most common during December and January and during the evening hours. Wednesday had the greatest number of accidents leading to death, 79% higher than the Saturday weekend rate. Of those tested, 24% had ethanol in their blood. 66% of the fatal injuries occurred on city or residential streets, and 29% occurred on major thoroughfares. A single urban highway accounted for 12% of pedestrian fatalities and represented a particularly hazardous traffic environment. Fatal head injuries and severe chest injuries were present in 73% of cases; injuries involving multiple sites were present in 60%. There were few significant differences in the extent of injuries with respect to vehicle speed or type of vehicle. Head injuries were much less common in the oldest age group, probably because elderly pedestrians were more vulnerable to death from less severe trunk and extremity injuries. Severe chest injury was the most important predictor of death occurring at the scene.

Isler et al., 1998 studied the child pedestrians' crossing gap thresholds. Three gender-balanced groups of 16 school children (5–6 years, 8–9 years, 11–12 years) participated in individual pretests of vision, hearing, and time to walk across a 12-m wide urban street and back. Each child then completed 10 roadside trials requiring judgement of the threshold point at which they would no longer cross in front of traffic approaching from their right. The judgments were made from a site immediately in front of a parked car at a point 2 m from the kerb and 4 m from the centre of the road. Traffic speeds and distances were measured using a laser speed and distance detector. The results indicated that, overall, distance gap thresholds remained constant regardless of vehicle approach speeds. Analysis of the thresholds for distance gap judgments for the 4-m half-street crossing showed that some of the older children could be expected to make safe decisions, but this was not so for the 5–6- and 8–9-year-olds at vehicle approach speeds above 60 kph. Almost two-thirds of the children reported using distance to judge gaps, which proved the least adequate strategy in terms of proportion of resultant safe decisions.

Fontaine et al., 1997 analyzed the main pedestrian characteristics- age and sex, movements, change of transport mode and alcohol impairment. A multiple correspondence analysis was used and followed by a hierarchical classification technique. The analysis was performed with 17 active variables that serve to define the axes of the factorial analysis. Nine describe the pedestrian involved in the accident: age, sex, reason for outing, type of trip (alone, in a group), pedestrian movement, position, obstacles to progress, change of mode, and impairment by alcohol. Eight are related to the environment or the type of accident: situation in or outside built-up area, whether the road was straight or curved, day of week, month, light, weather conditions, type of vehicle, and secondary accident. In the under-15 age group, the accident occurs in the child's town of residence in 72% of cases whereas for pedestrians as a whole the rate is 54%.

Preusser et al., 2002 reported the pedestrian crashes in Washington D.C. and Baltimore using the police crash reports for the years 1970 and 1998. They observed that pedestrians were more at fault than the drivers in 1998. Also, the driver's failing to yield right of way increased drastically over the years.

Stutts and Hunter, 1998 reported that the official crash records significantly under estimated the numbers of pedestrian and bicyclists involved in crashes with vehicles. Baltes, 1998 described the application of Florida specific crash data used to categorize pedestrian crashes according to a variety of factors including gender and age, time of day, pedestrian's contributing cause, injury severity, weather condition, road system identifier, etc., to the specific sequence of events perceived to influence the crashes involving pedestrians. He found out that a higher percentage of children between ages 0 and 18 were involved in crashes when "crossing not at Intersection". Also, about 13% of the pedestrians failed to yield right of way to the driver. He even observed that most of the incidents took place between 4:00PM and 7:59PM. Baltes, 1995 reported that between 1990 and 1994, in Florida, nearly 20% of the all traffic crash fatalities were pedestrian related. Of the 2484 pedestrian crashes, nearly 16% of the crashes were in the age group 0 and 19.

Noland et al., 2004 analyzed pedestrian and bicycle casualties using cross-sectional time series data for regions of Great Britain. A fixed effect negative binomial model was developed. Various factors associated with those killed and seriously injured as well as slight injuries are examined. These include the average age of vehicles in the region, the road length of various road classes, the amount of vehicle ownership in the region, per capita income, per capita expenditure on alcohol, age cohorts, and various proxies for medical technology improvements. Various specifications of the models are estimated. Generally, we find that more serious pedestrian injuries are associated with lower income areas, increases in percent of local roads, increased per capita expenditure on alcohol, and total population.

Hunter el al., 1996 found that roadway and environmental factors are associated with pedestrian crashes in the U.S. They analyze 5000 police-reported pedestrian crashes in six states. The results showed that 75% of pedestrian crashes occur where speed limits are less than 35 mph; 21% of crashes occur where there is a sidewalk on at least one side of the roadway. This suggests some effect from various road infrastructure elements.

Harruff et al., 1998 performed a retrospective analysis of 217 pedestrian traffic fatalities in Seattle, Washington. They concluded that elderly pedestrians were most vulnerable because they are more likely to be injured as a pedestrian and more likely to die of injuries that a younger person might survive. They found that nearly one quarter of the fatalities tested were positive for alcohol. There was little correlation of extent of injuries with vehicle type, speed zone where injury occurred, type of roadway and pedestrian activity. Minority populations were found to have a higher incidence of fatalities which may be due to environmental factors associated with lower income areas, such as high speed roads. Graham & Glaister , 2002 also find evidence to support this amongst patterns of childhood pedestrian fatalities which are strongly associated with more deprived areas.

Graham & Glaister, 2002 *analyzes* ward level data for England providing a high level of spatial resolution of area wide characteristics. They focus on child pedestrian casualties but also analyze all pedestrian casualties. Their results suggest that more deprived wards will have higher casualties as will those that generate more traffic (as measured via a proxy variable), more densely populated areas (with the most dense actually seeing a reduction in casualties), and the length of the road network.

LaScala et al., 2000 conducted a spatial regression analysis of pedestrian injuries associated with motor-vehicles in San Francisco, California. The results showed that pedestrian injuries were associated with increased traffic flow and population density (as measured per kilometer of road length). Areas with higher unemployment were associated with higher injury rates while areas with more high school graduates had lower injury rates. This is similar to the results of Graham & Glaister *(6)* who used an area-wide deprivation score in their analyses. Surprisingly, they found that more children (aged 0-15) in an area were associated with fewer pedestrian injuries.

Assailly, 1997 found that the two groups most 'at risk' in European countries are 5-9 years-olds and the elderly. The children are at risk in terms of high accident involvement, whereas the elderly are at risk in terms of mortality from injuries sustained, due to increased frailty. Pedestrian injury rates peak for those aged 7-9 years in almost all European countries and a secondary peak has been observed at ages 10-14 years in the United Kingdom. They suggested that environmental factors, including overcrowded housing, traffic density, absence of play areas, and parental monitoring practices are likely to contribute to children's vulnerability.

Derlet et al., 1989presented an epidemiological review of 217 pedestrian injuries treated at a level one-trauma center during a one-year period. Injuries that occurred in pediatric age group patients were reviewed separately from adults. Hospital length of stay and severity of injuries was found to be much worse in adults. Seven percent of adults and three percent of children died after arrival at the hospital. This study shows that the incidence of critical injuries to pedestrians is high, and adults sustain more severe injuries than children. This would perhaps make it difficult for improvements in medical technology to have much impact on actual outcomes.

Kingma, 1994 investigated the causes of pedestrian accidents (N=534) for patients treated for injuries at the emergency unit of a hospital. Accidents in collisions with motor vehicles were the main cause (87.8%). Young children (0-9 years old) and elderly (above 60 years of age) are the most vulnerable in terms of mortality rates observed in these age groups. Preponderance of males in pedestrian accidents was observed in the accident categories of collisions with motor vehicles and bicycles, whereas a slight preponderance of females was found in collisions with other traffic.

Johnson et al., 2005 proposed a model for understanding the relationship between socioeconomic status (SES) and the risk for injury among child pedestrians. The model was based on the general model that injury or disease that considers the rate of injury to be a function of both exposure and risk per unit of exposure. This general model has been adapted to the pedestrian context and specified for children. It presents SES as the primary variable. In the first relationship, SES influences a variety of environmental and social factors. These modifying factors, in turn, influence the behavior of pedestrians, drivers, and others. Resulting from behavior, injury risk can increase or decrease. The combination of exposure and risk per unit of exposure determines the rate of injury.

Tyrrell et al., 2004 made an extensive literature review on the effect of nighttime conspicuity from the pedestrians' perspective in pedestrian crashes. They found out that 64% of all the pedestrian crashes in the year 2001 took place in the nighttime. They concluded that pedestrians often lack sufficient conspicuity to approaching drivers at night.

In an effort to increase student safety, Ford, 2005 from the Mendocino County Department of Transportation (MDOT) undertook a trial school zone speed reduction project in 2003. They proposed to reduce speeds by using speed display signs to alert drivers to their actual speed as they entered the zones. These signs have built-in radar units that operate digital numeric speed displays and are permanent installations mounted on breakaway poles. With grant funding assistance from the California Office of Traffic Safety (OTS), they were able to install the signs on both approaches to three separate school zones. Four statistics were calculated for each of the eighteen radar speed surveys performed: mean speed, 85th percentile (prevailing) speed, highest observed speed and percent of vehicles exceeding 25 MPH. a collision fatality rate equation was developed as part of this study. This simple weighted average yields a single number that reflects the relative fatality risk to pedestrians from the speed distribution of the vehicles within each survey sample. It was found that by all measures, midday speeds were affected more than morning and afternoon speeds in all three school zones. This suggests that the greatest benefit may be derived from using digital speed display signs for school zones on multi-lane roads or in other situations where free flow traffic is present while students are being dropped-off and picked-up.

2.2 Log-Linear Analysis Modeling and GIS

For the analysis of contingency tables, hypothesis testing is usually considered. Alternatively, another approach can be used, namely that of fitting models and estimating the parameters in the models. The term model refers to some 'theory' or conceptual framework about the observations and the parameters in the model represent the 'effects' that particular variables or combinations of variables have in determining values taken by the observations. Such an approach is common in many branches of statistics such as regression analysis and the analysis of variance. Most common are linear models which postulate that the expected values of the observations are given by a linear combination of a number of parameters. Techniques such as maximum likelihood and least squares may be used to estimate parameters, and estimated parameter values may then be used in identifying which variables are of greatest importance in 'predicting' the observed values. The major advantages to be gained from the model-fitting techniques to be described are firstly that they provide a systematic approach to the analysis of complex multidimensional tables and secondly that they provide estimates of the magnitude of effects of interest; consequently they allow the relative importance of different effects to be judged.

Abdel-Aty et al., 2005 used log-linear models to analyze the vehicle-pedestrian crashes at intersections in Florida over the years, 1999-2002. The study made use of the driver, pedestrian and geometric road characteristics. It was found that pedestrian and driver characteristics are closely related to frequency of injury crashes. Also, the analysis proved that road geometric, traffic and environmental conditions are closely related to the injury severity and frequency. Two separate models were built and estimated- pedestrian crashes at pedestrians' fault and pedestrian crashes at drivers' fault. In case of crashes at driver's fault, the results showed that middle-age (25–64) and male drivers are more involved in crashes as causers than other driver groups.

Carlin et al., 1995, assessed the relationship of the risk of injury requiring hospital attendance in children riding bicycles to socio-demographic factors and to measures of exposure, in a large area of suburban Melbourne, Australia. Particular attention was given to the measurement of individual exposure in several dimensions. Log-linear Analysis was carried out for the study. Analysis of interim data from 109 cases and 118 controls shows that 51% of injuries occurred while the child was playing rather than making a trip on the bicycle and only 22% involved another vehicle. Exposure measures showed complex patterns of association with injury risk. Estimated time spent riding was more closely associated with risk than distance traveled, with an odds ratio of 2.2 (95% confidence interval 1.1-4.2) for children riding for more than 3 hours per week compared to children riding less than 1 hour. Riding more than 5 km on the sidewalk was also associated with increased risk (odds ratio 3.1, 95% CI 1.1-8.5).

Aultman-Hall et al., 1999 relating the route information of the 1196 respondents at Toronto to facility attributes in a Geographic Information System (GIS), defensible estimates of travel exposure on roads, off-road paths and sidewalks were developed. The rate of collision on off-road paths and sidewalks was lower than for roads. This result may confirm urban form, traffic levels and attitude do affect bicycle safety.

Abdel-Aty et al., 1998 assessed the effect of driver age on traffic accident involvement using log-linear models. Four log-linear models with three variables in each model were fitted and odd multipliers were computed to predict the effect of age on other variables. The results indicated significant relationships between the driver age and ADT, injury severity, manner of collision, speed, alcohol involvement, and roadway character.

Abdel-Aty et al., 2000 explored the relationship between alcohol and the driver characteristics in motor vehicle accidents in the state of Florida. Conditional probability and log linear models were developed to analyze the effect. The results showed that the 25–34 age group experiences the highest rate of alcohol, drug involvement in accidents. The rates decline with the increase in the age of the drivers.

From the literature review it can be observed that

1) Most of the studies pertained to crashes with all age groups. There have been few studies with emphasis only on school going aged children.
2) Very few studies mention about the effect of the time of day and age group on the frequency of crashes.
3) Developing log-linear models appears to be a very effective tool for analysis of various variables over each other.
4) Geographical Information Systems has been sparsely used in previous studies related to pedestrian crashes. No spatial analysis of the crashes involving pedestrians and bicyclists under the age of 18 using GIS has been done.

Hence, this study is unique as we are using a combination of spatial analysis using GIS and log-linear analysis to characterize bicyclists and pedestrian crashes under the age of 18.

3. DATA COLLECTION

Over the past few years, the issue of safety of school children, during the school hours, has gained a lot of attention in Orange County. There have been cases of crashes being reported, that involve school children.

3.1 Crash Data

The study used crash data obtained from the Department of Highway Safety and Motor Vehicles (DHSMV), for 5 years, i.e. from 1999 to 2003. The DHSMV traffic crash data base is a relational database consisting of seven files. Each file deals with a specific aspect of a traffic crash. The files are as follows:

1. Events file- this file contains general information about the crash characteristics and circumstances.
2. Vehicle file- this file contains information about the vehicles and vehicle actions in the traffic crash.
3. Driver file- this file contains information about the drivers and condition or action of the driver that contributed to the crash.
4. Pedestrian file- this file deals with information on any pedestrians involved in the traffic crash.
5. Violations file – this file lists the citations (if any) issued in connection with the traffic crash, by statute number.

6. Passenger file- this file provides information about any passengers involved in the traffic crash.
7. D.O.T. Site Location file- this file contains additional information about crash locations occurring on state roads only. This data is supplied by the department of Transportation.

Each crash is associated with a unique crash number. The Driver's file is the biggest amongst all the files and it has more than 400,000 entries each year. Handling such a large database would be very difficult. In order to be able to analyze our results better, it was decided that the following steps would be used to identify the crashes related to school.

- Limiting pedestrian and bicycle crashes to Orange County.
- The age of the persons involved in the crashes was limited between 4 and 18, that is, the school going age.
- Also, not all crashes involving children would be related to school. Hence it was decided that the crashes taking place in the time periods 6:30 AM - 10:00 AM and 1:00 PM – 5:00 PM (one hour earlier on Wednesday) only would be considered.
- Moreover, the crashes during weekends and school vacation were ignored. That is the crashes taking place on weekends, between 1st June to 7th August and 22nd December to 5th January, were not considered, mainly because most of the schools remain closed during these dates.

Microsoft Access, Microsoft Excel and SAS software were used in creating the final database consisting of only the pedestrian and bicycle crashes taking place in Orange County in the specified timings and dates. A total of 423 crashes involving school children were identified in which 451 pedestrians and bicyclists were involved. Of this 204 were bicyclists and 247 pedestrians. This dataset of 451 crashes with their related driver, pedestrian and vehicle characteristics was stored as a SAS dataset and used for further analysis (log-linear analysis). A single crash number could be associated with multiple pedestrians/bicyclist/drivers, and hence there are only 423 crashes for 451 pedestrians/bicyclists.

3.2 GIS Data

Of the total 423 pedestrian and bicycle crashes that took place in Orange County in the specified time periods and age groups, 262 (nearly 62%) of the actual crash reports were found and obtained. These crash reports were downloaded from the Florida Department of Transportation Mainframe website. The spatial analysis was based on these 262 crash reports (we can assume that they are a random sample representing the total 423 crashes). Most of the crash reports obtained were for crashes that took place near state roads as the FDOT Mainframe website contains only those reports that took place on state roads. And hence the analysis done in this study could be said to be based on crashes that took place on state roads mainly.

The first step in the process was locating the crashes on the Orange County streets map. The schools and locations, and the Orange County streets map were obtained from the Orange County School board. After a detailed inspection and review of the 262 crash reports, each crash was geo-coded onto the Orange County streets map. Figure 3-1 shows the GIS map of the Orange County streets and the crash locations.

The schools in Orange County are divided into 7 districts. This information was obtained from the Orange County Public Schools' website. This kind of division enables us to analyze the crashes at the school and district level. Figure 3-2 shows the 7 school districts in Orange County. There are a total of 157 elementary, middle and high schools identified in Orange County, according to the school locations file obtained from the School board of Orange County.

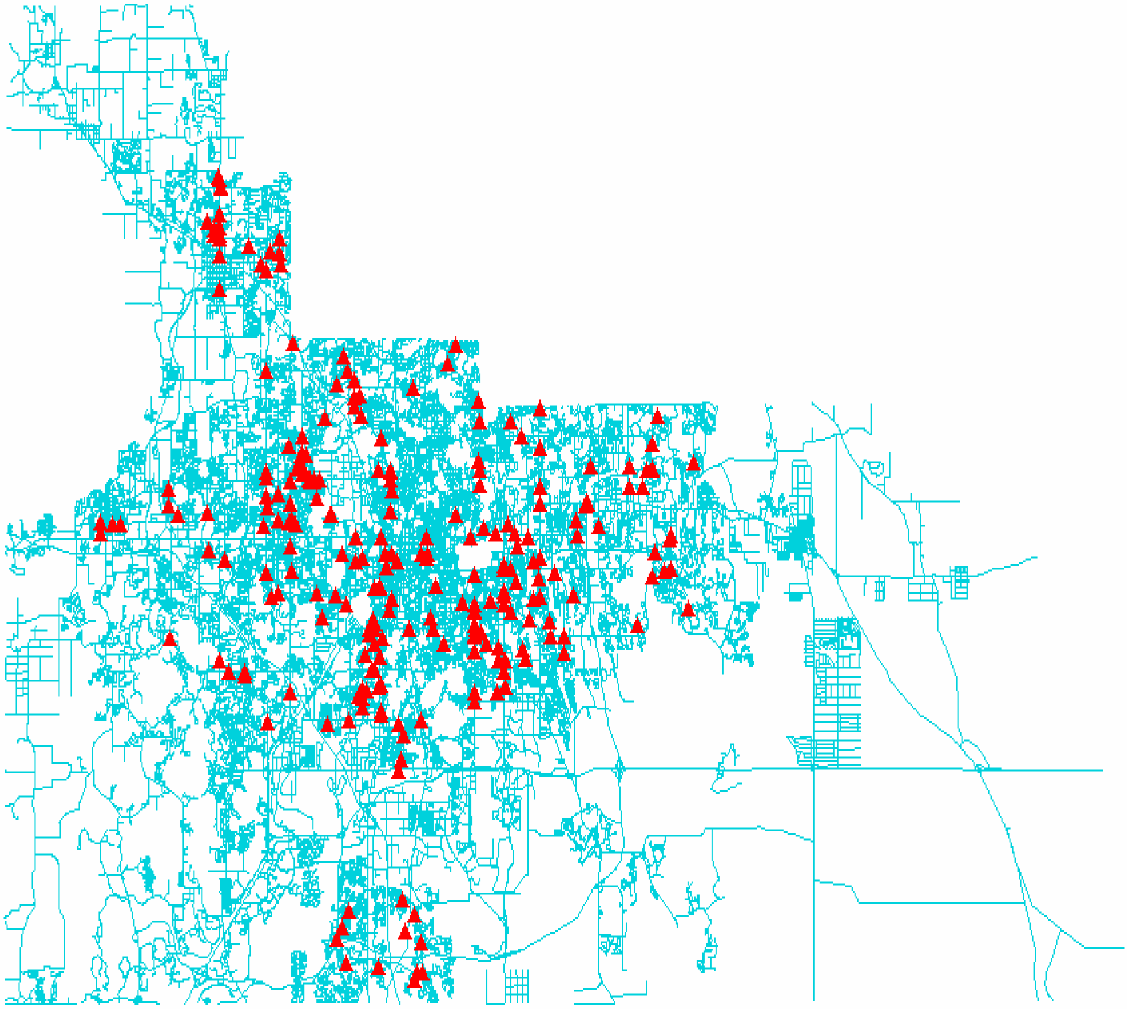

Figure 3-1: Orange County Streets and Crash Locations

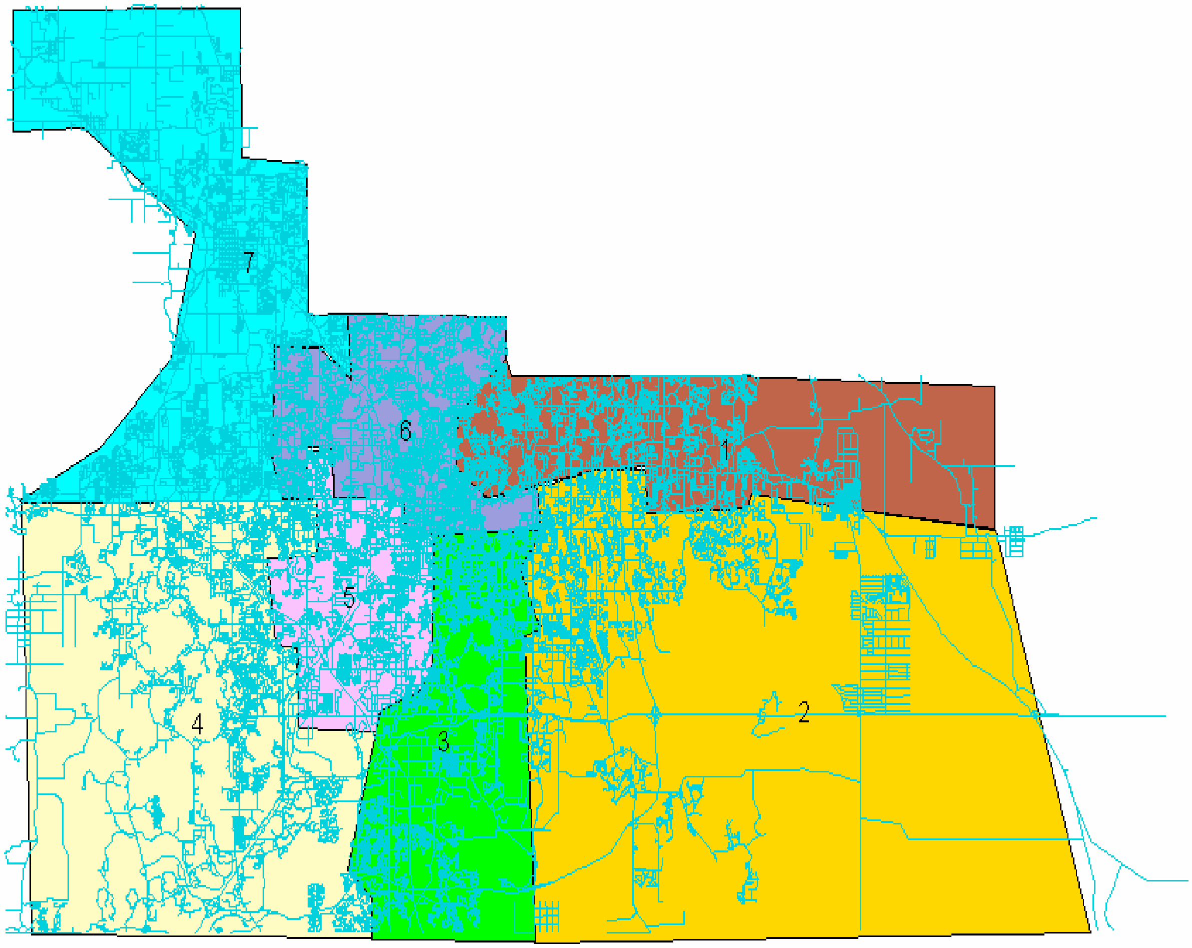

Figure 3-2: Orange County School Districts

4. PRELIMINARY DATA ANALYSIS & GIS

This thesis focuses on three kinds of analyses. The preliminary study describes the various driver, pedestrian, vehicle and geometric characteristics involved in the crashes. The variables were compared to each other using two- way contingency tables and bar graphs. The ages of the pedestrians and bicyclists involved in the crashes were divided into three categories based on the school going age. They were divided as follows: elementary school age (4 to 11), middle school age (12 to 14) and high school age (15 to 18). Basic t-tests were performed across various contingency tables to look at the inter-dependency among the variables. The crash data set containing 451 crashes was used for this analysis. Based on these preliminary analyses, certain significant variables were chosen that would be used for further analysis, i.e. Log –Linear Analysis. The crash data set of 451 crashes was further reduced to 444 crashes. The difference in the crashes was due to the removal of 7 crashes that were found dubious.

The study used crash data obtained from the Department of Highway Safety and Motor Vehicles, for 5 years, i.e. from 1999 to 2003. The data was extracted as per the following conditions to identify the crashes related to school.

- Limiting pedestrian and bicycle crashes to Orange County.
- The age of the persons involved in the crashes was limited between 4 and 18, that is, the school going age.

- Also, not all crashes involving children would be related to school. Hence it was decided that the crashes taking place in the time periods 6:30 AM - 10:00 AM and 1:00 PM – 5:00 PM (one hour earlier on Wednesday) only would be considered.
- Moreover, the crashes during weekends and school vacation were ignored. That is the crashes taking place on weekends, between 1st June to 7th August and 22nd December to 5th January, were not considered, mainly because most of the schools remain closed during these dates.

Analysis of Crash Data

Table 4-1 illustrates the total number of crashes that took place in Orange County, on weekdays and schooldays, between the time 6:30AM to 10:00AM and 1:00PM to 5:00PM. The choice of these times was mainly to identify the crashes related to school children, as they are normally the school opening and closing time periods. Also, the total number of drivers involved in the crashes as per the above constraints between the ages 4 and 18 has been determined. It shows that out of 21086 crashes that took place in all the 5 years according to the time and day in Orange County, 2660 of the drivers were between the ages 4 and 18. That is, 12.62% of the crashes had drivers under the age of 18.

Table 4-1: Total Crashes and Drivers involved in Orange County.

	1999		**2000**		**2001**		**2002**		**2003**		**Total**	
Total Crashes *(Time and Day)*	4541		4518		4421		3902		3704		21086	
Drivers *(Time, Day and Age)*	629	13.85%	645	14.28%	500	11.31%	438	11.23%	448	12.10%	2660	12.62%

Time indicates 6:30Am to 10:00Am and 1:00PM to 5:00PM. Day indicates weekdays and school working days.

Figure 4-1 illustrates the percentage of young drivers under the age of 18, each year. For example, 12.1% of the driver population in the year 2003, was under 18 on weekdays and schooldays, between the time 6:30AM to 10:00AM and 1:00PM to 5:00PM, in Orange County and this contributes towards 19% of the over all drivers during the 5 years.

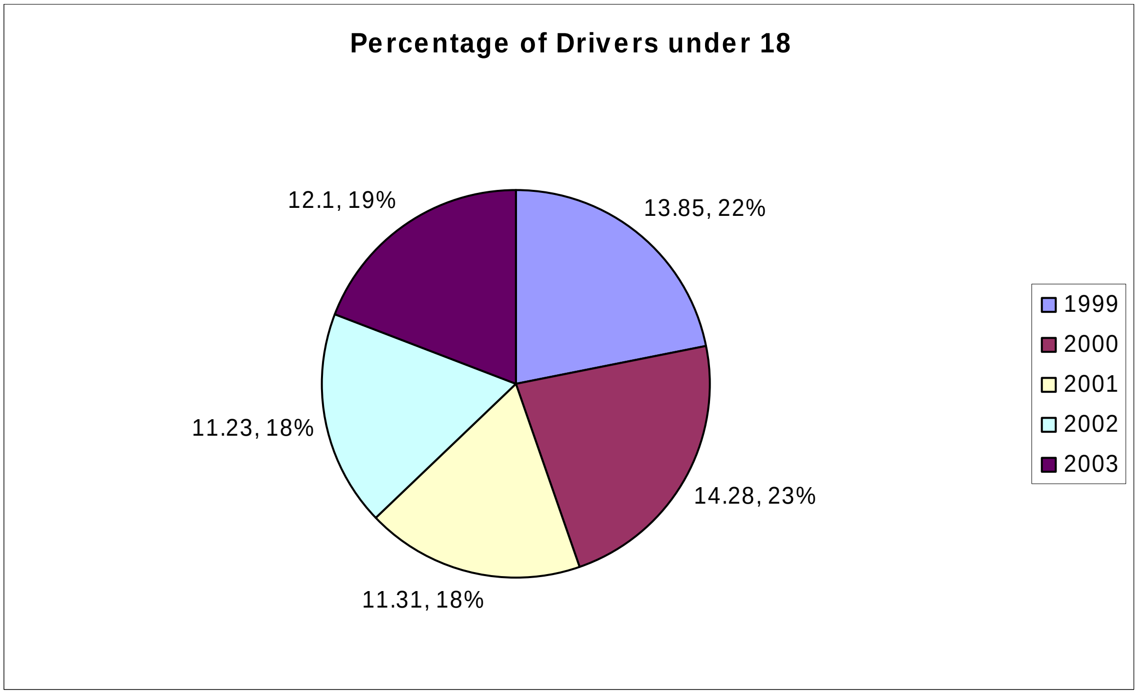

Figure 4-1: Percentage of Drivers under 18 in the 5 years

Nearly 33% of the pedestrian crashes in the age 4 to 18, occurred in the time periods of 6:30Am to 10:00PM and 1:00PM to 5:00PM, on weekdays and school working days, which is a significant percentage. Table 4-2 illustrates the total Pedestrian crashes in Orange County and there by classifies the crashes as per age first and then by age, time and day. Figure 4-2 illustrates the yearly distribution of the 247 pedestrian crashes mentioned in Table 4-2.

Table 4-2: Total Pedestrian Crashes

	1999		**2000**		**2001**		**2002**		**2003**		**Total**	
Total Pedestrian Crashes	689		620		625		564		538		3036	
Pedestrian Crashes *(Age)*	164	23.80%	163	26.29%	161	25.76%	133	23.58%	132	24.54%	753	24.80%
Pedestrians *(Age, Time and Day)*	58	35.37%	57	34.97%	41	25.47%	41	30.83%	50	37.88%	247	32.80%

Age indicates 4 to 18. Time indicates 6:30Am to 10:00Am and 1:00PM to 5:00PM. Day indicates weekdays and school working days.

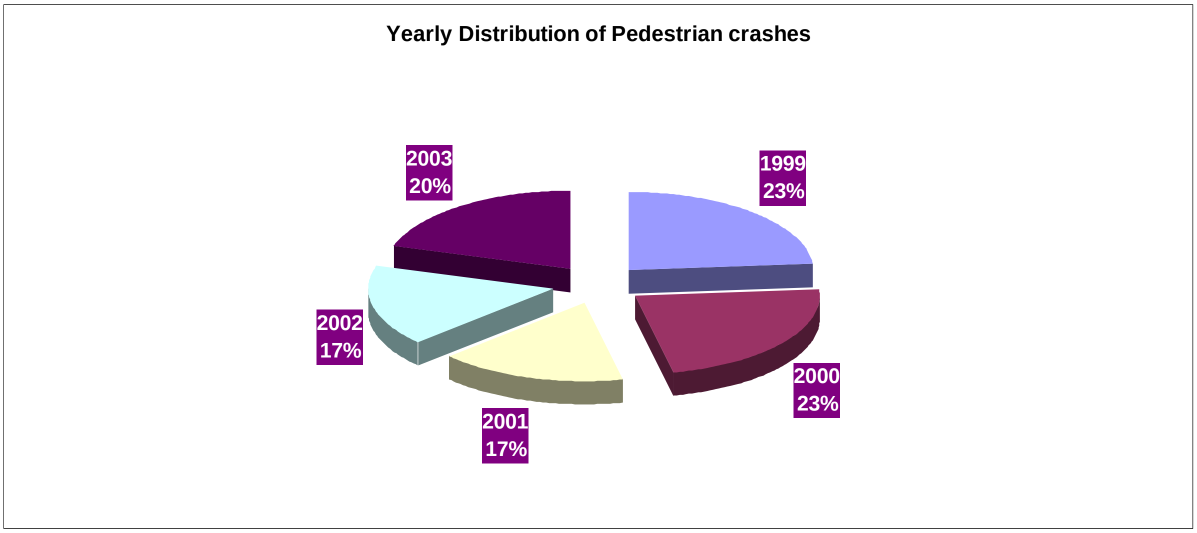

Figure 4-2: Yearly distribution of Pedestrian Crashes

Similar to the above table, Table 4-3 shows the percentage of bicyclist crashes in Orange County, in the Time, Age and Day constraints. On an average nearly 7.7% of the driver crashes were bicycle related.

Table 4-3: Total Bicyclist Crashes

	1999		**2000**		**2001**		**2002**		**2003**		**Total**	
Total Drivers	629		645		500		438		448		2660	
Bicyclists *(Age, Time and Day)*	58	9.22%	50	7.75%	38	7.60%	24	5.48%	34	7.59%	204	7.67%

Age indicates 4 to 18. Time indicates 6:30Am to 10:00Am and 1:00PM to 5:00PM. Day indicates weekdays and school working days.

Since the bicyclist and pedestrian crashes contribute to a significant amount of crashes, the analysis was focused on only for these two types of crashes. Table 4-4 elucidates the age group distribution of the pedestrian and bicyclist crashes. Totally, there are 423 crashes that involved 451 pedestrians and bicyclists, as per age, time and day constraints in Orange County. It can be noticed that about 23% of the total crashes are pedestrian crashes in the age group of 4 to 11, representing Elementary school children. Most of the bicycle crashes are in the middle and high school age group. Hence, the variation in the percentage of crashes among the three age groups- Elementary, Middle and High. It can also be observed that the Elementary school children have a higher percentage of crash rates among the three age groups. But at the same time, the Elementary school children share a higher pedestrian crash rate and a lower bicycle crash rate when compared to the middle and high school children.

Table 4-4: Total Pedestrian and Bicyclist crashes

Age	**Pedestrians**		**Bicyclists**		**Total**	
Elementary (4 to 11)	102	*22.62%*	59	13.08%	161	35.70%
Middle (12 to 14)	67	14.86%	79	17.52%	146	32.37%
High (15 to 18)	78	17.29%	66	14.63%	144	31.93%
Total	247	54.77%	204	45.23%	451	100.00%

Figure 4-3, Figure 4-4 and Figure 4-5 pictorially represent the distribution of the pedestrian, bicycle and over all crashes over the three age groups. It can be noticed that

41% of the pedestrian crashes alone are related to the Elementary School Children, while 39% of the bicycle crashes are related to the Middle School Children. Also, the over all distribution shows that the Elementary school children have higher crash rates, roughly 10% higher than the other two groups.

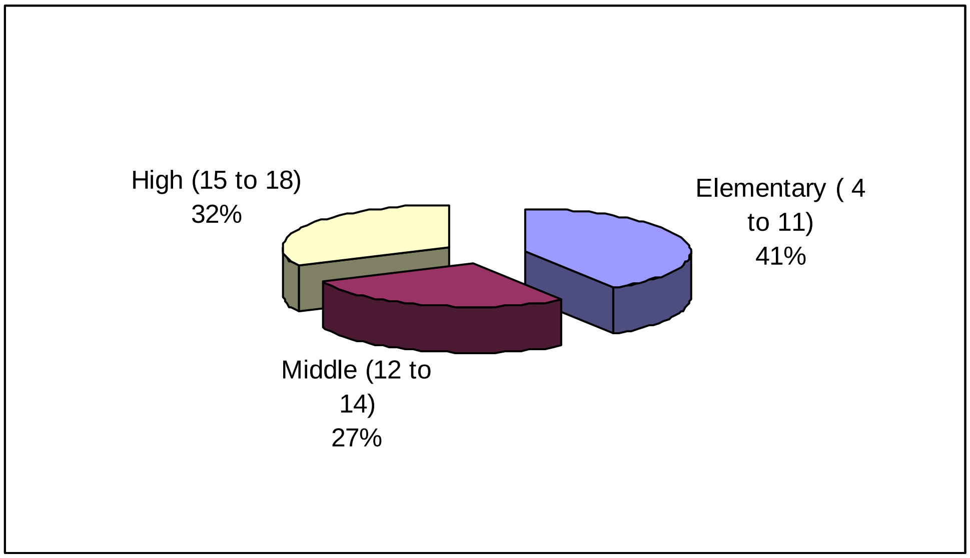

Figure 4-3: Distribution of the Pedestrian Crashes over the 3 age groups

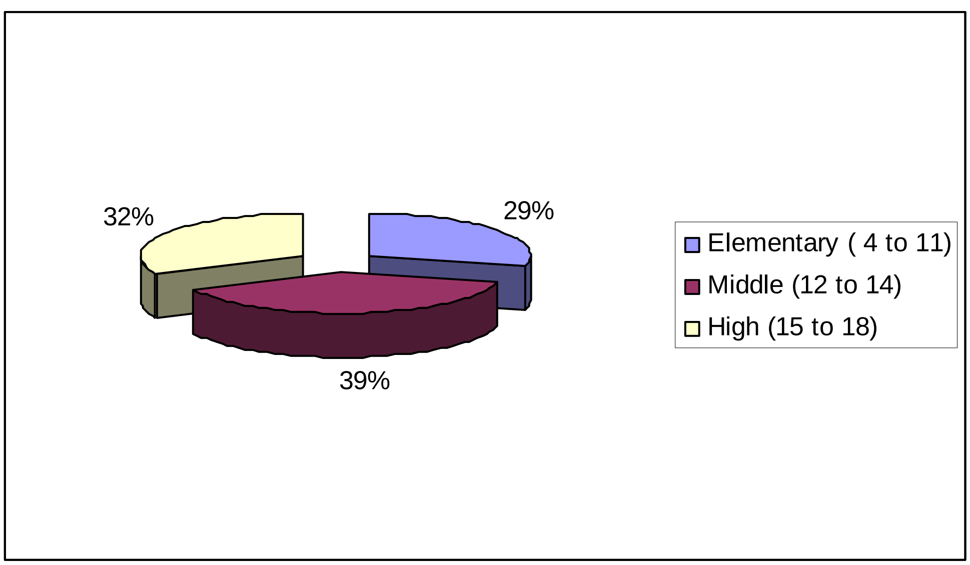

Figure 4-4: Distribution of the Bicycle crashes over the three age groups

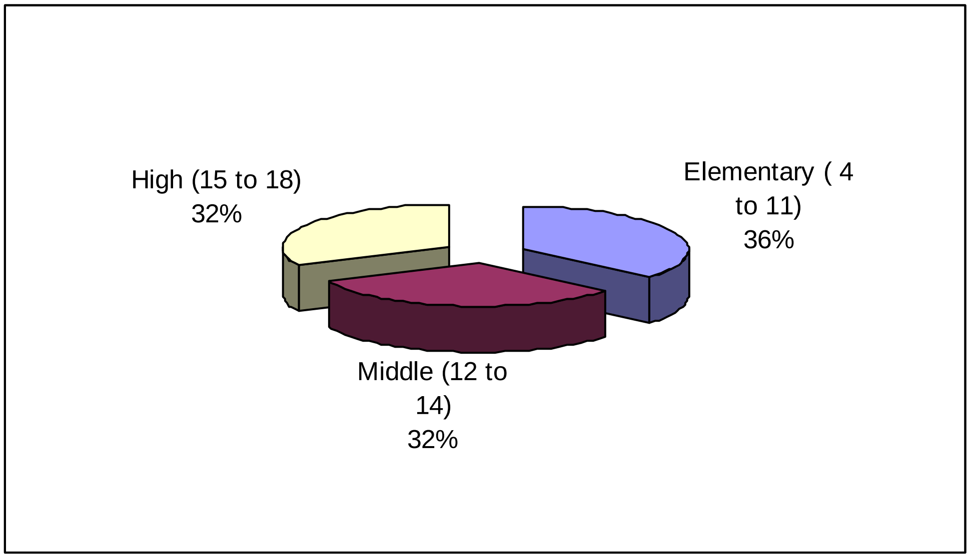

Figure 4-5: Distribution of the Total crashes (Pedestrian and bicycle) over three age groups

Table 4-5 shows the month distribution of the crashes. It can be noticed that the months of June and July, have almost 50% less crashes than the other months. We can attribute this to the fact that the schools are closed in those months. Hence, it is very apt to drop the crashes in these two months and also the crashes starting from 22nd December till the 5th of January, as the schools remain closed at this time too. Similarly table 6 depicts that Thursdays and Fridays have a slightly higher crash rate than the other weekdays.

Table 4-5: School related pedestrian and bicycle crashes

Month	Total Crashes	Crashes after excluding vacation
Jan	48	40
Feb	45	45
Mar	50	50
Apr	38	38
May	48	48
Jun	24	0
Jul	21	0
Aug	39	34
Sep	42	42
Oct	39	39
Nov	54	54
Dec	42	33
Total	490	423

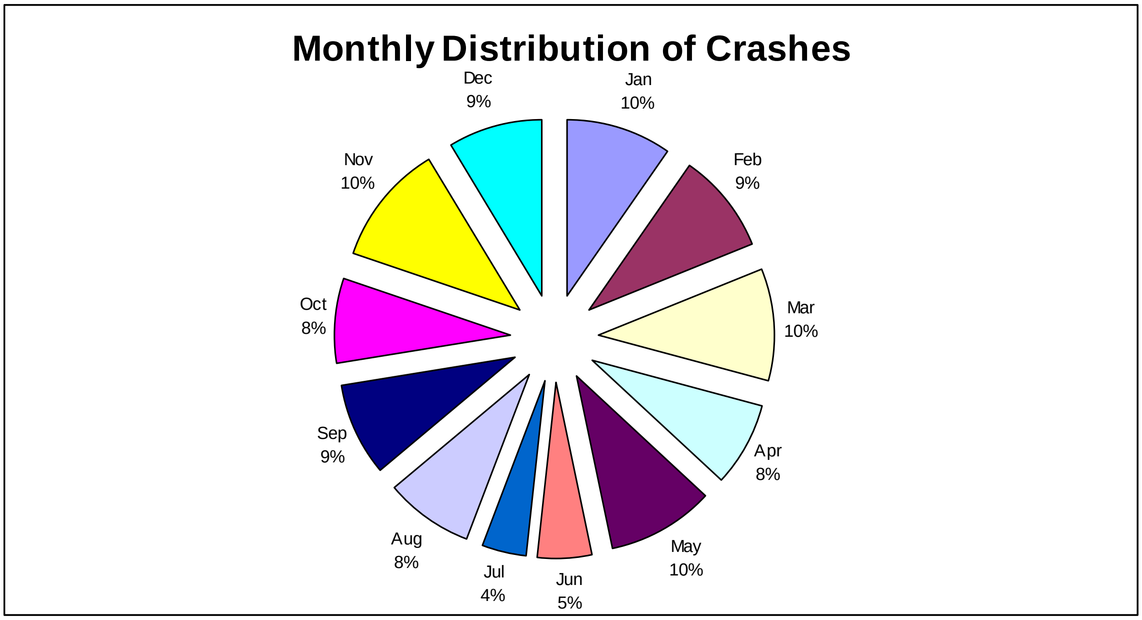

Figure 4-6: **Distribution of crashes monthly**

Figure 4-6 represents the monthly distribution of crashes. Clearly, the months of June and July recorded comparatively fewer percentages of crashes.

Table 4-6: Daily Crashes

Day	**Crashes**	**%**
Monday	87	19.29
Tuesday	89	19.73
Wednesday	86	19.07
Thursday	95	21.06
Friday	94	20.84
Total	451	100

One of the main factors that explain the crashes is the type of Traffic Control when the Crash occurred. Developing a two-way table with the three age groups and the various Traffic controls, it was found that 18% of the total crashes and 49% of the crashes in the age group 4 to 11 occurred at “No Traffic Control”. Also a significant about 22% of the total crashes occurred at “Stop Sign”. Table 4-7 illustrates the differences among the various age groups with respect to the traffic controls. Each record in the table has four entities, the crash numbers, the overall percentage, the row percentage and column percentage respectively. Pearson chi-square test was conducted to test the hypothesis of independence. The test statistic was found to be 0.0006, which is less than 0.005. This indicates that the rows and columns of this contingency table are dependent. So, it can be concluded that higher percentage of crashes occurred at “no traffic control” in the elementary age group.

Table 4-7: Traffic Controls as per age group

Age Group Vs Traffic Control												
	Traffic Control											
Age Group	**No traffic Control**	**School Zone**	**Traffic signal**	**Stop Sign**	**Yield Sign**	**Flashing Light**	**Guard**	**Special Speed Zone**	**Speed Control Sign**	**No Passing Zone**	**All Other**	**Total**
Elementary (4 to 11)	79	5	14	43	1	2	1	5	2	2	7	**161**
Total %	*17.52%*	1.11%	3.10%	9.53%	0.22%	0.44%	0.22%	1.11%	0.44%	0.44%	1.55%	*35.70%*
Row %	*49.07%*	3.11%	8.70%	26.71%	0.62%	1.24%	0.62%	3.11%	1.24%	1.24%	4.35%	100.00%
Col %	42.25%	15.15%	16.87%	42.57%	50.00%	50.00%	100.00%	45.45%	22.22%	100.00%	38.89%	35.70%

Middle (12 to 14)	47	16	31	35	0	1	0	3	3	0	10	**146**
Total %	10.42%	3.55%	6.87%	7.76%	0.00%	0.22%	0.00%	0.67%	0.67%	0.00%	2.22%	32.37%
Row %	32.19%	10.96%	21.23%	23.97%	0.00%	0.68%	0.00%	2.05%	2.05%	0.00%	6.85%	10000.00%
Col %	25.13%	48.48%	37.35%	34.65%	0.00%	25.00%	0.00%	27.27%	33.33%	0.00%	55.56%	32.37%
High (15 to 18)	61	12	38	23	1	1	0	3	4	0	1	**144**
Total %	13.53%	2.66%	8.43%	5.10%	0.22%	0.22%	0.00%	0.67%	0.89%	0.00%	0.22%	31.93%
Row %	42.36%	8.33%	26.39%	15.97%	0.69%	0.69%	0.00%	2.08%	2.78%	0.00%	0.69%	100.00%
Col %	32.62%	36.36%	45.78%	22.77%	50.00%	25.00%	0.00%	27.27%	44.44%	0.00%	5.56%	31.93%
Total	187	*33*	83	101	2	4	1	11	9	2	18	**451**
Total %	41.46%	7.32%	18.40%	22.39%	0.44%	0.89%	0.22%	2.44%	2.00%	0.44%	3.99%	100.00%

** Chi-square = 46.7, DF= 20, P-value = 0.0006*

Table 4-7 also shows that a meager 7.32% of the total pedestrian and bicycle crashes took place at the designated “School Zones”. This clearly states that the problem with the crashes is not entirely located near the schools.

Table 4-8 shows the distribution of crashes with respect to age group at various site locations. Nearly 37% of the crashes occurred “Not at intersection”. Almost an equal percentage, 35% of crashes took place “At Intersection”. A point to be noted here is that the crashes in the Elementary School age group i.e. ages between 4 and 11 are higher than the other two age groups.

Table 4-8: Crashes with respect to age group at various site locations

Age Group vs Site location											
	Site location										
Age Group	**Not at Intersection**	**At Intersection**	**Influenced by Intersection**	**Driveway Access**	**Entrance Ramp**	**Public Parking Lot**	**Pvt Parking Lot**	**Pvt Property**	**Pub Bus Stop Zone**	**All Other**	**Total**
Elementary (4 to 11)	55	55	11	12	1	8	9	2	0	8	**161**
Total %	12.2	12.2	2.44	2.66	0.22	1.77	2	0.44	0	1.77	*35.7*
Row %	34.16	34.16	6.83	7.45	0.62	4.97	5.59	1.24	0	4.97	100.00
Col %	33.33	34.59	35.48	34.29	100	44.44	81.82	50	0	32	35.7
Middle (12 to 14)	54	60	10	11	0	5	0	0	2	4	**146**
Total %	11.97	13.3	2.22	2.44	0	1.11	0	0	0.44	0.89	32.37
Row %	36.99	41.1	6.85	7.53	0	3.42	0	0	1.37	2.74	100.00
Col %	32.73	37.74	32.26	31.43	0	27.78	0	0	100	16	32.37
High (15 to 18)	56	44	10	12	0	5	2	2	0	13	**144**
Total %	12.42	9.76	2.22	2.66	0	1.11	0.44	0.44	0	2.88	31.93
Row %	38.89	30.56	6.94	8.33	0	3.47	1.39	1.39	0	9.03	100.00
Col %	33.94	27.67	32.26	34.29	0	27.78	18.18	50	0	52	31.93
Total	165	159	31	35	1	18	11	4	2	25	**451**
Total %	*36.59*	*35.25*	6.87	7.76	0.22	3.99	2.44	0.89	0.44	5.54	100

Continuing further on those lines, looking at the severity of the crashes for the pedestrians and bicyclists involved, more than 50% of them survived with "Non-Incapacitating Evident Injury"-i.e., any visible injuries such as bruises, abrasions, limping, etc. Table 4-9classifies the injuries according to age group.

Table 4-9: Severity of Pedestrians and Bicyclists as per age group

Age Group vs Severity						
	Severity					
Age Group	**No Injury**	**Possible Injury**	**Non-Incapacitating Evident Injury**	**Incapacitating Injury**	**Fatal Injury**	**Total**
Elementary (4 to 11)	7	52	76	23	2	**160**
Total %	1.56%	11.58%	16.93%	5.12%	0.45%	35.63%
Row %	4.38%	32.50%	47.50%	14.38%	1.25%	100.00%
Col %	20.59%	42.98%	33.48%	36.51%	50.00%	35.63%
Middle (12 to 14)	5	36	81	23	0	**145**
Total %	1.11%	8.02%	18.04%	5.12%	0.00%	32.29%
Row %	3.45%	24.83%	55.86%	15.86%	0.00%	100.00%
Col %	14.71%	29.75%	35.68%	36.51%	0.00%	32.29%
High (15 to 18)	22	33	70	17	2	**144**
Total %	4.90%	7.35%	15.59%	3.79%	0.45%	32.07%
Row %	15.28%	22.92%	48.61%	11.81%	1.39%	100.00%
Col %	64.71%	27.27%	30.84%	26.98%	50.00%	32.07%
Total	34	121	227	63	4	**449**
Total %	7.57%	26.95%	*50.56%*	14.03%	0.89%	100.00%

In Table 4-4, it was evident that there are 247 pedestrians involved in the crashes. A closer look at the pedestrian action, when the crash took place reveals that nearly 40% of the pedestrian crashes happened while “Crossing Not at Intersection”. Table 4-10 gives an idea about the Pedestrian action. Though the “Others” category seems higher, when looked at individually in the various categories under “Others”, each is insignificant.

Table 4-10: Pedestrian Action during the Crashes

Age Group vs. Pedestrian Action				
	Pedestrian Action			
Age Group	**Crossing Not at Int**	**Crossing at Int**	**All Others**	**Total**
Elementary (4 to 11)	38	24	40	**102**
Total %	15.38	9.72	16.19	41.3
Row %	37.25	23.53	39.22	100
Col %	39.18	41.38	43.48	41.3
Middle (12 to 14)	31	13	23	**67**
Total %	12.55	5.26	9.31	27.13
Row %	46.27	19.4	34.33	100
Col %	31.96	22.41	25	27.13
High (15 to 18)	28	21	29	**78**
Total %	11.34	8.5	11.74	31.58
Row %	35.9	26.92	37.18	100
Col %	28.87	36.21	31.52	31.58
Total	97	58	92	**247**
Total %	39.27	23.48	37.25	100

Table 4-11 indicates the Contributing cause of the Pedestrians during the crash. It can be seen that almost 34% of the pedestrians had “No Improper Action”. About 23% of the pedestrians “Failed to yield right-of-way”.

Table 4-11: Contributing Cause of the Pedestrians

Age Group vs. Contributing Cause				
	CCP			
Age Group	**No Improper Action**	**Failed to Yield Right-of-Way**	**All Other**	**Total**
Elementary (4 to 11)	29	28	44	**101**
Total %	11.89%	11.48%	18.03%	41.39%
Row %	28.71%	27.72%	43.56%	100.00%
Col %	34.94%	50.91%	41.51%	41.39%
Middle (12 to 14)	24	15	27	**66**
Total %	9.84%	6.15%	11.07%	27.05%
Row %	23.76%	14.85%	26.73%	65.35%
Col %	28.92%	27.27%	25.47%	27.05%
High (15 to 18)	30	12	35	**77**
Total %	12.30%	4.92%	14.34%	31.56%
Row %	29.70%	11.88%	34.65%	76.24%
Col %	36.14%	21.82%	33.02%	31.56%
Total	83	55	106	**244**
Total %	*34.02%*	*22.54%*	43.44%	100.00%

The above tables were more related to pedestrian and bicyclist characteristics. Now, looking at the geometric aspects of the road, where the crashes took place, the number of lanes and the type of median (divided or undivided) existing is definitely a major point to be considered. Clearly from Table 4-12 a large majority, 50%, of the crashes took place on Undivided two lane highways. Pearson's chi-square test was conducted to test the hypothesis of independence. The test statistic was found to be 0.001, which is less that 0.05. This indicates that the rows and columns of this contingency table are dependent, which asserts the finding that majority of the crashes took place on two lane undivided roads.

Table 4-12: Total Crashes with respect to No. of lanes and type of highway (Divided/undivided)

Highway Division Vs. No. of Lanes											
Highway	**No. of Lanes**										**Total**
	Parking Lot/Pvt Property	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**11**	
Divided	8	2	25	2	58	3	16	1	1	1	**117**
Total %	1.90%	0.47%	5.92%	0.47%	13.74%	0.71%	3.79%	0.24%	0.24%	0.24%	27.73%
Row%	6.84%	1.71%	21.37%	1.71%	49.57%	2.56%	13.68%	0.85%	0.85%	0.85%	100.00%
Col%	20.51%	20.00%	10.55%	28.57%	59.18%	75.00%	66.67%	100.00%	100.00%	100.00%	27.73%
Undivided	31	8	212	5	40	1	8	0	0	0	**305**
Total %	7.35%	1.90%	*50.24%*	1.18%	9.48%	0.24%	1.90%	0.00%	0.00%	0.00%	*72.27%*
Row%	10.16%	2.62%	69.51%	1.64%	13.11%	0.33%	2.62%	0.00%	0.00%	0.00%	100.00%
Col%	79.49%	80.00%	89.45%	71.43%	40.82%	25.00%	33.33%	0.00%	0.00%	0.00%	72.27%
Total	39	10	237	7	98	4	24	1	1	1	**422**
Total Col%	9.24%	2.37%	*56.16%*	1.66%	23.22%	0.95%	5.69%	0.24%	0.24%	0.24%	100.00%

** Chi-square = 115.05, DF =9, p-value= 0.001*

Figure 4-7 below clearly shows a demarcation between the divided and undivided highways. Also, it can be seen that most of the crashes, 72% occur on undivided highways than on divided highways.

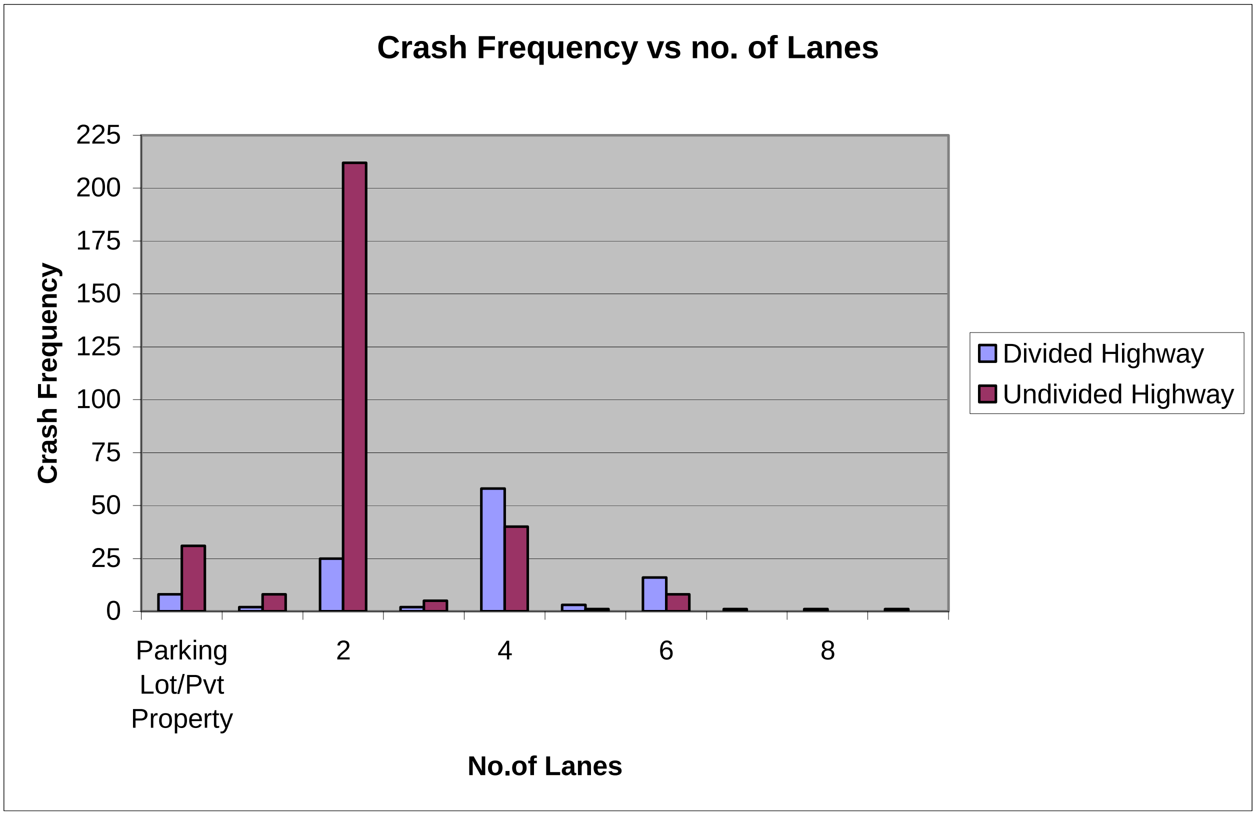

Figure 4-7 Crashes according to the no. of lanes and highway division

For crash analysis, considering the contributing cause for the crash from the driver's point is always essential. Table 4-13 shows that 52% of the drivers had "No improper driving action" but a significant 16% of the drivers were "careless".

Table 4-13: Contributing causes of the Driver

Contributing Cause Driver		
CCD	**Frequency**	**Percent**
No Improper Driving	210	51.72
Careless Driving	63	15.52
Failed to yield right-of-way	49	12.07
Improper backing	10	2.46
Improper Turn	2	0.49
Disregarded traffic Signal	2	0.49
Disregarded Stop Sign	5	1.23
Improper Passing	3	0.74
Obstructing Traffic	1	0.25
Driving Wrong side	1	0.25
All Other	60	14.78
Total	406	100

Another significant factor is Speed. The crashes in Figure 4-8, show that the majority of the crashes nearly 27%, occurred at speeds of about 25MPH. This could be attributed to the fact that most of the residential areas have a speed limit of 30MPH.

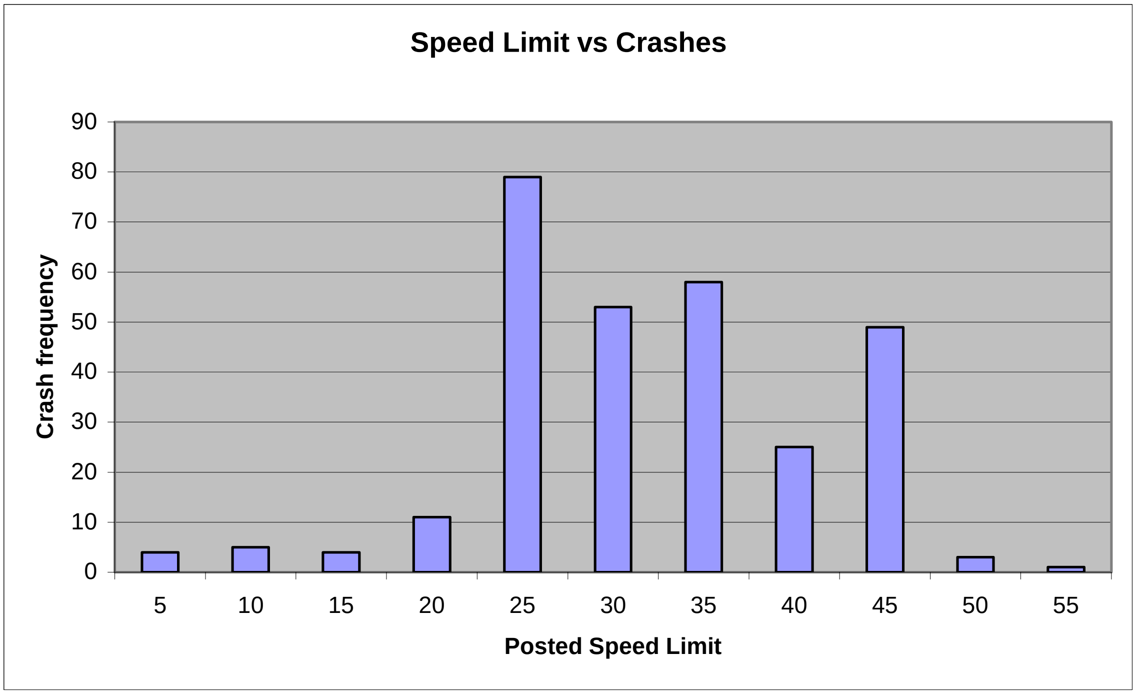

Figure 4-8: Crash Frequency according to the Posted Speed Limit**.**

Table 4-14 depicts the exact crash frequency values at posted speed limits.

Table 4-14: Crash Frequency and Posted Speed Limit

Posted Speed	**Frequency**	**Percent**
5	4	1.37%
10	5	1.71%
15	4	1.37%
20	11	3.77%
25	79	27.05%
30	53	18.15%
35	58	19.86%
40	25	8.56%
45	49	16.78%
50	3	1.03%
55	1	0.34%
Total	292	100.00%

Table 4-15, indicates that 64% of the pedestrian and bicycle crashes took place in "primarily residential areas". Since most of the schools are located in residential areas, it is not surprising to see such statistics.

Table 4-15: Crash Frequency and Location type

Location Type	**Frequency**	**Percent**
Primarily Business	158	35.03%
Primarily Residential	287	63.64%
Open Country	6	1.33%
Total	451	100.00%

Figure 4-9 is a pictorial representation of the crash frequency and location type values.

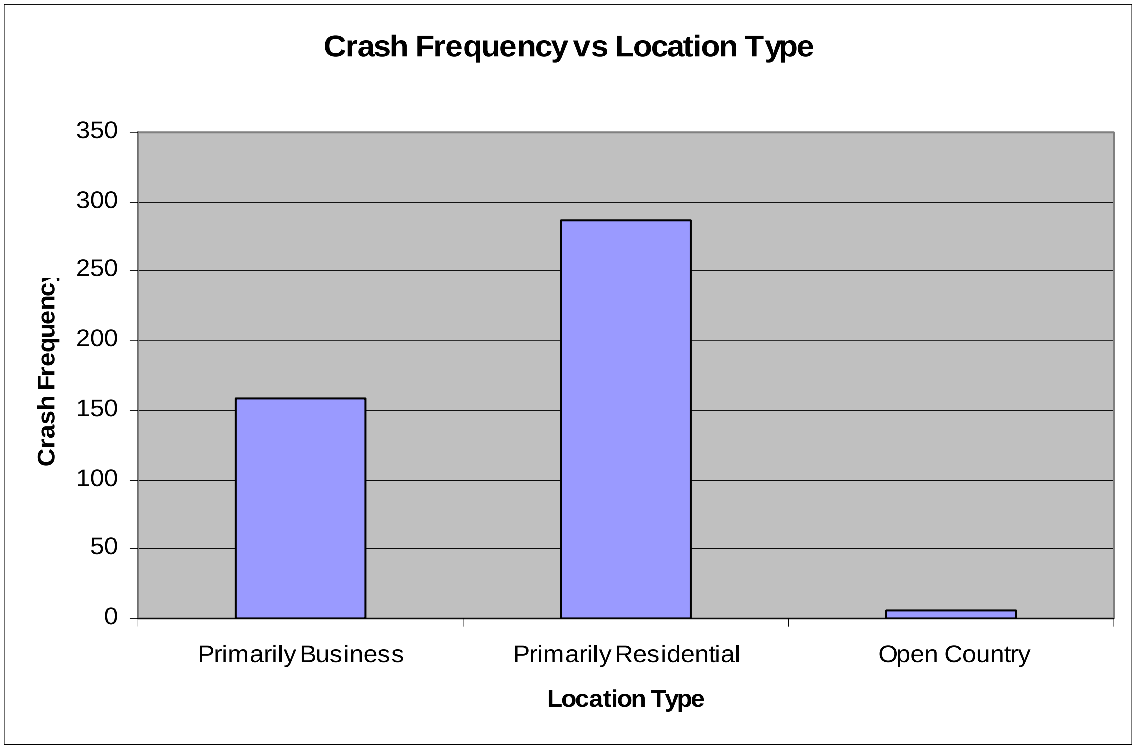

Figure 4-9: Location type and Crashes

More than the significance of crashes at the posted limit, its important for us to know, if the driver involved in the crash was speeding or not. Hence, by estimating a Speed ratio factor, i.e by dividing estimated speed by the speed limit, we would be able to say whether or not the driver was speeding. Table 4-16 illustrates the speeding factor. If the speed ratio <=1, that means the driver is with in or at speed limit. If it is more than 1, it implies that the driver is speeding.

Table 4-16: Crash frequency and Speed Ratio Factor

Speed Ratio Factor	**Frequency**	**Percent**
0	89	30.48%
0.15	51	17.47%
0.3	15	5.14%
0.45	22	7.53%
0.6	27	9.25%
0.75	31	10.62%
0.9	16	5.48%
1.05	36	12.33%
1.2	4	1.37%
1.35	0	0.00%
1.5	1	0.34%
Total	292	100.00%

Following the above rule, it was found that nearly 14% of the drivers were above the speed limit, i.e. the speed ratio is greater than 1. Figure 4-10 is a graphical representation of Table 4-16.

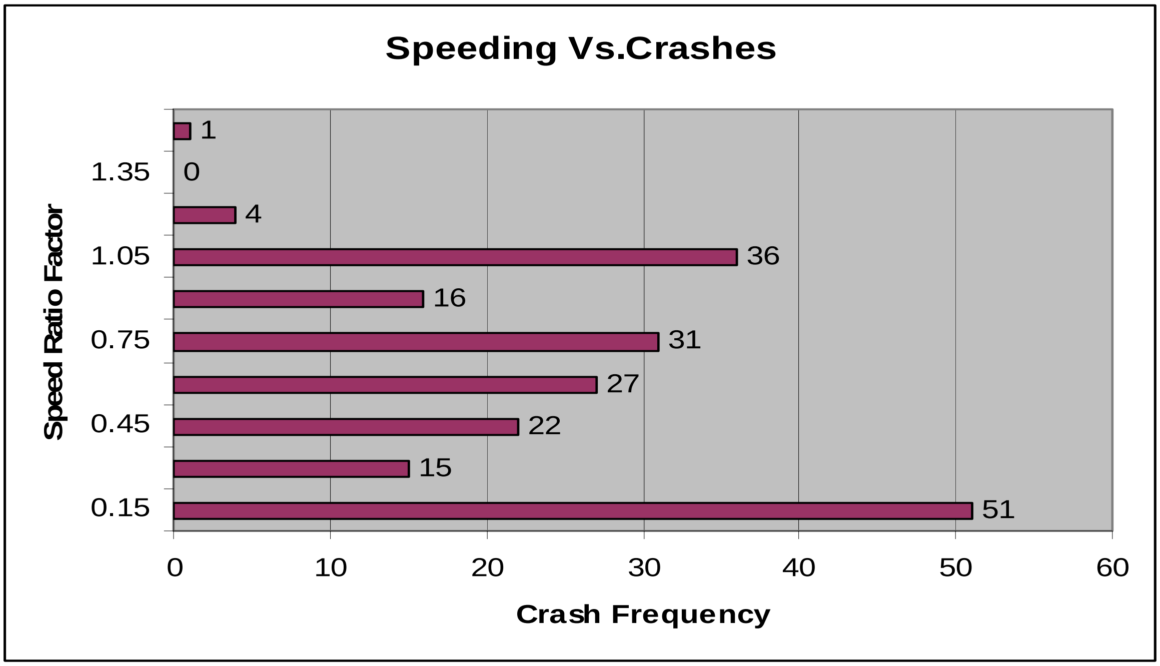

Figure 4-10: Speeding vs Crashes

Geo-Spatial Analysis

Geographical Information Systems (GIS) was used for the spatial analysis of the data. With the information about the total number of crashes and their locations, the schools and their locations, and also the Orange County base streets map, the spatial analysis was possible. Spatial analysis helps us understand a complex problem in a simpler way.

Of the total 423 pedestrian and bicycle crashes that took place in Orange County in the specified time periods and age groups, 262 (nearly 62%) of the actual crash reports were found and obtained. The spatial analysis was based on these 262 crash reports (we can assume that they are a random sample representing the total 423 crashes).

Methodology and Analysis

The first step in the process was locating the crashes on the Orange County streets map. After a detailed inspection and reading of the 262 crash reports, each crash was geo-coded onto the Orange County streets map. The schools and locations, and the Orange County streets map were obtained from the Orange County School board. Figure 4-11 shows the GIS map of the Orange County streets and the crash locations.

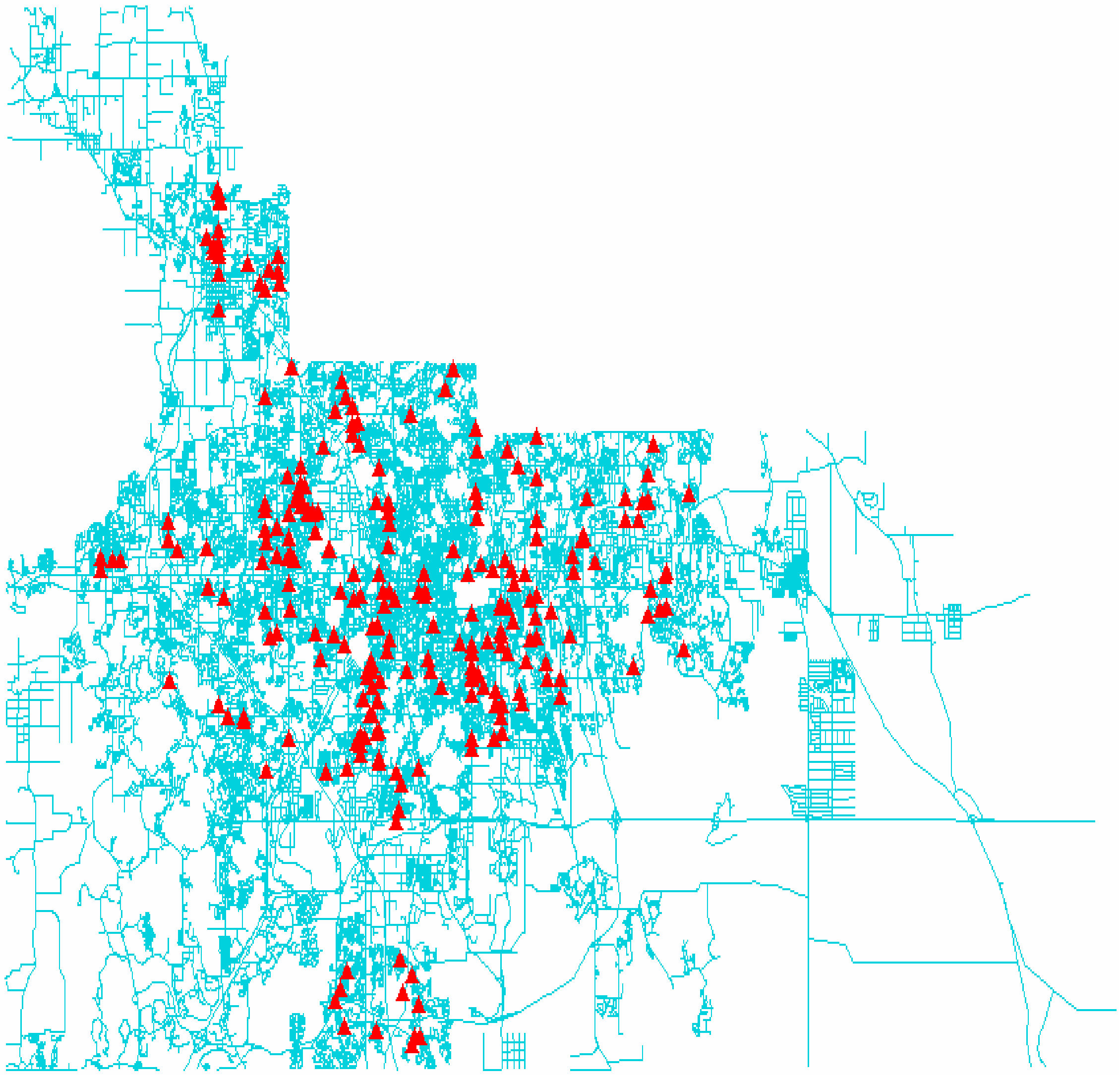

Figure 4-11: Orange County Streets and Crash Locations

The schools in Orange County are divided into 7 districts. To look at the microscopic picture of the schools and the crashes related to schools, the Orange County map was divided in to 7 districts with the schools in each district as the basis. This information was obtained from the Orange County Public Schools' website. Figure 4-12 shows the 7 school districts in Orange County.

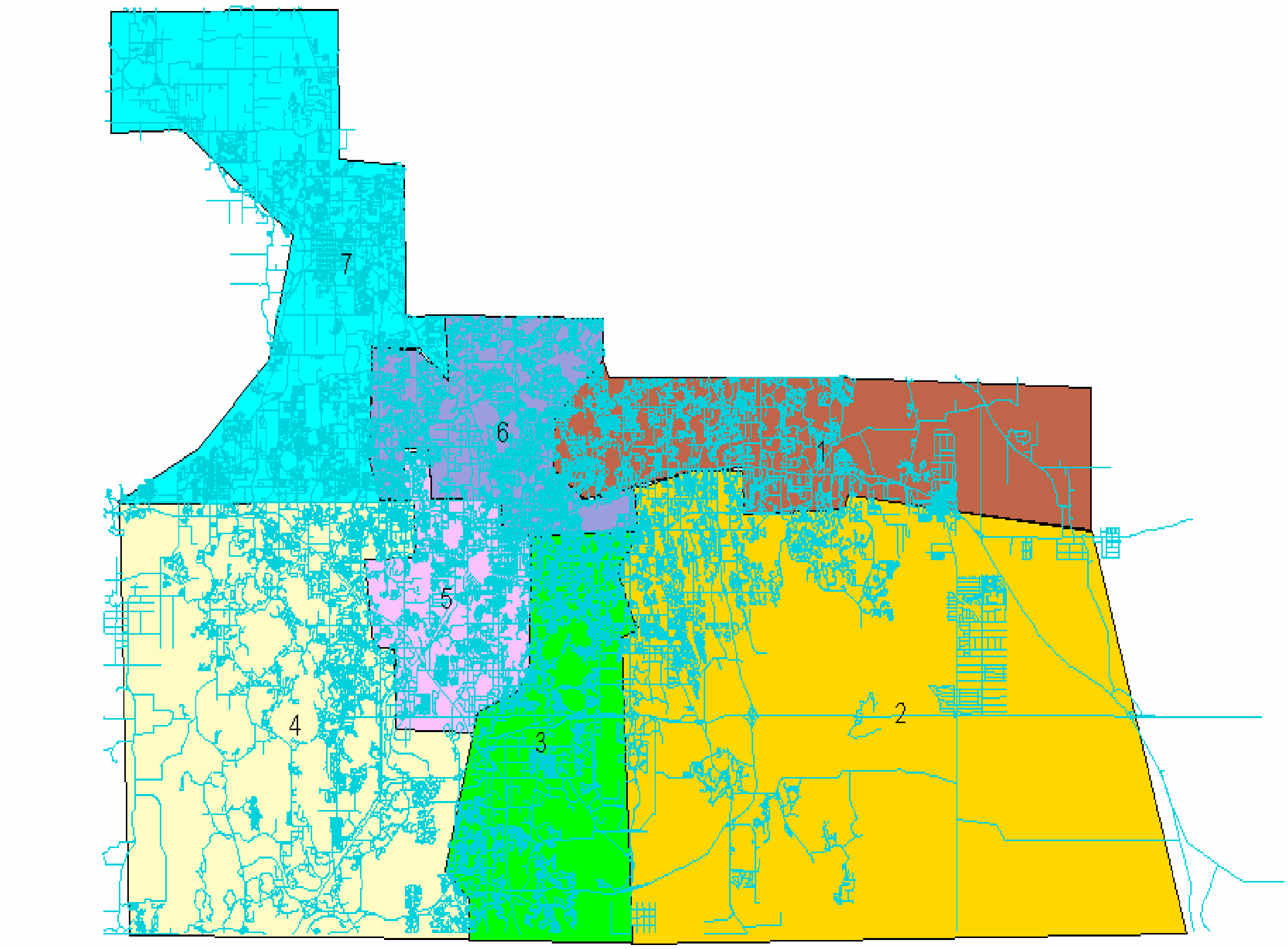

Figure 4-12: Orange County School Districts

There are a total of 157 elementary, middle and high schools identified in Orange County, according to the school locations file obtained from the School board of Orange County. Over laying the schools and the Crashes on the school district map above gives us an understanding of the district wise distribution of the crashes. Figure 4-13 shows the Orange County Districts map with the crash locations and the school zones.

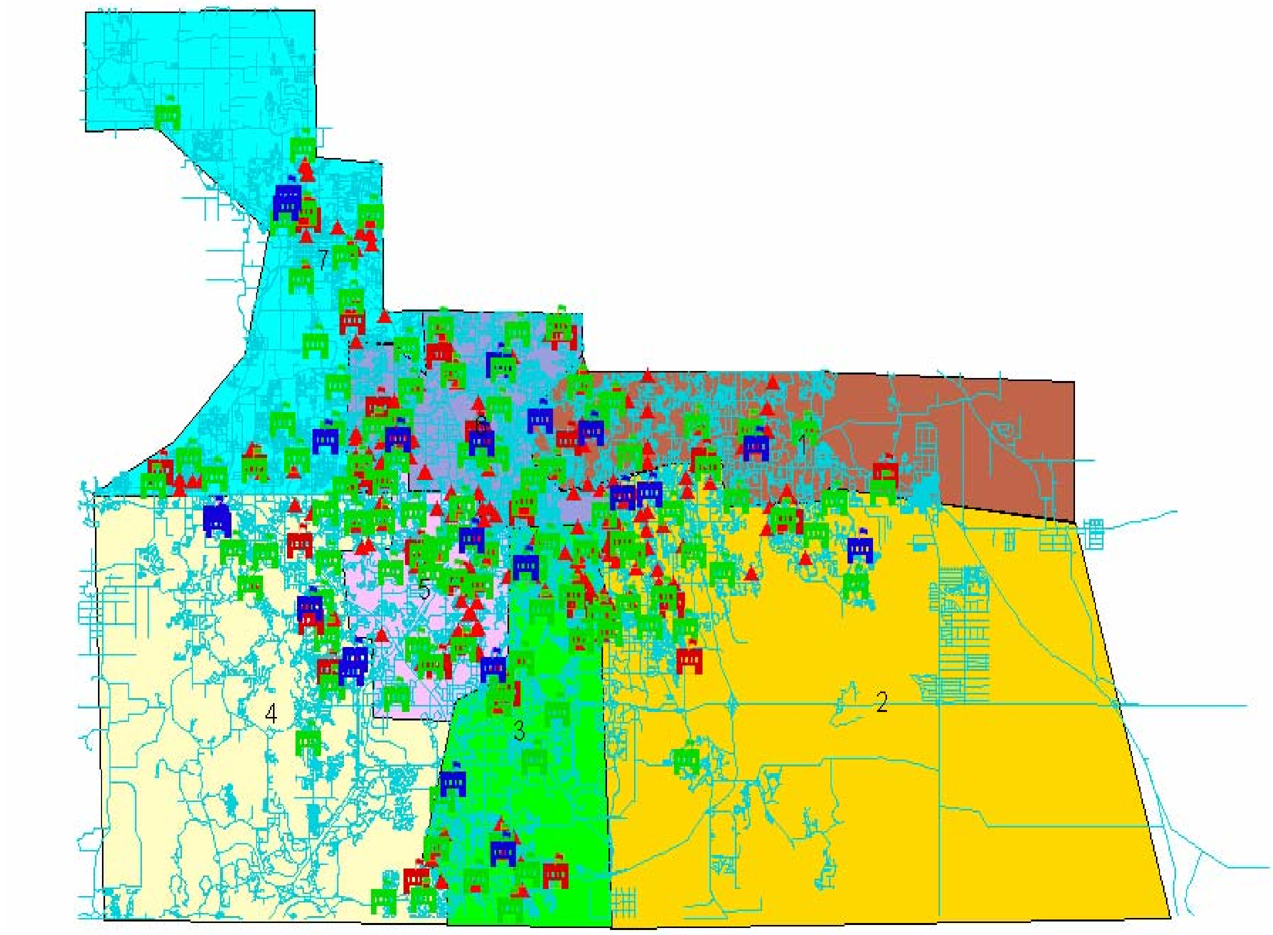

Figure 4-13: Orange County School districts with the school and crash locations

The green, red and blue buildings in Figure 4-13 depicts the elementary, middle and high school locations, respectively, while the red triangles represent the 262 crash locations.

In order to identify the crashes around a school better, district wise distribution of the crashes and the schools was considered. Using geo processing wizard and the spatial join tool, this distribution was achieved. A factor called "Crashes per Schools" was deduced in each district. It was calculated by dividing the total number of crashes taking place in each district with the total number of schools in that district. Table 4-17 summarizes the number of schools, crashes and crashes per school in each district. If the factor crashes per school is greater than the mean, for a particular county, it can be said that the particular district has a safety problem.

Table 4-17: District wise distribution of total no. of schools and crashes

District	No.Of Schools	No.Of Crashes	Crashes per School
1	16	30	1.88
2	23	38	1.65
3	25	38	1.52
4	22	15	0.68
5	21	43	2.05
6	24	65	2.71
7	26	33	1.27

Here, the mean crashes per school are 1.68. Clearly, districts 5 and 6 show a higher crash rate. At the same time, it can be noted that the no. of schools in districts 5 and 6 are higher when compared to the other districts. This distinction in the crashes per school can be seen in Figure 4-14, where darker the color implies more crashes per school rate.

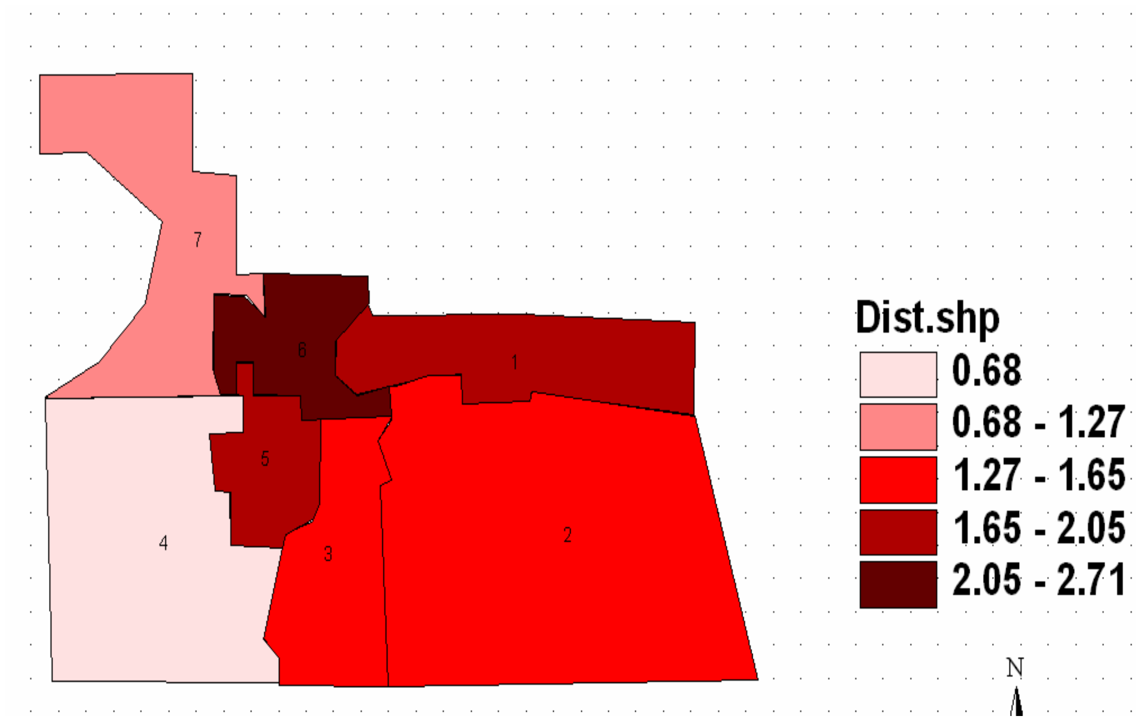

Figure 4-14: Crashes per School representation

The legend for districts shows the crashes per school attribute. Figure 4-14 shows that, districts 1, 5 and 6 are at a higher risk, in that order.

Further analysis, was done by looking at the schools in particular. Buffers were created around each school with a half mile radii distance. It was assumed that the crashes taking place with in half a mile distance from the school, would be influenced by the school location. Hence, all the crashes in the buffer zone around each school were counted. Using the create buffer zones, geo-processing wizard tool and the spatial join analysis tool, this analysis was carried out. Only those buffer zones with crash locations were identified. There were 66 such buffer zones with a total of 147 crashes and 87 schools.

That is there are some buffer zones which covered more than more one school. A factor called Crashes per school was deduced similar to the one mentioned in the earlier section. Table 4-18 describes the total no. of crashes and schools located in each buffer zone.

Table 4-18: Buffer zone with the no. of crashes and the schools

Buffer Identity	No. Of Crashes	No. Of Schools	Crashes per school
12	1	1	1.00
14	1	1	1.00
16	2	2	1.00
24	2	1	2.00
26	1	1	1.00
30	1	2	0.50
31	1	1	1.00
32	3	1	3.00
33	8	1	8.00
35	3	1	3.00
36	2	1	2.00
39	2	1	2.00
44	1	1	1.00
45	1	1	1.00
47	1	1	1.00
49	2	2	1.00
51	1	1	1.00
57	1	1	1.00
58	1	1	1.00

Buffer Identity	No. Of Crashes	No. Of Schools	Crashes per school
62	1	1	1.00
64	1	1	1.00
66	3	1	3.00
70	1	1	1.00
72	2	1	2.00
75	2	1	2.00
76	2	1	2.00
77	2	1	2.00
78	1	1	1.00
82	1	1	1.00
83	2	1	2.00
85	2	1	2.00
90	5	1	5.00
98	2	1	2.00
102	1	1	1.00
104	1	1	1.00
107	3	1	3.00
108	1	1	1.00
110	1	1	1.00
112	1	1	1.00
113	6	1	6.00
114	1	1	1.00
115	2	1	2.00
118	4	1	4.00
123	2	3	0.67

Buffer Identity	No. Of Crashes	No. Of Schools	Crashes per school
125	3	1	3.00
126	3	1	3.00
129	2	1	2.00
134	2	2	1.00
136	3	2	1.50
137	1	2	0.50
141	2	2	1.00
142	7	2	3.50
143	1	2	0.50
144	4	2	2.00
147	3	2	1.50
148	3	3	1.00
149	3	2	1.50
150	1	1	1.00
151	3	1	3.00
152	5	2	2.50
153	4	2	2.00
154	2	2	1.00
155	2	2	1.00
156	3	1	3.00
157	4	2	2.00

The mean crashes per school rate is 0.94. That is, the buffer zones with crashes per school rate higher than 0.94 could be said to be risk prone. Figure 4-15 demonstrates the crashes per school rate with respect to the buffer zones. The darker the buffer, the higher the crash rate it has.

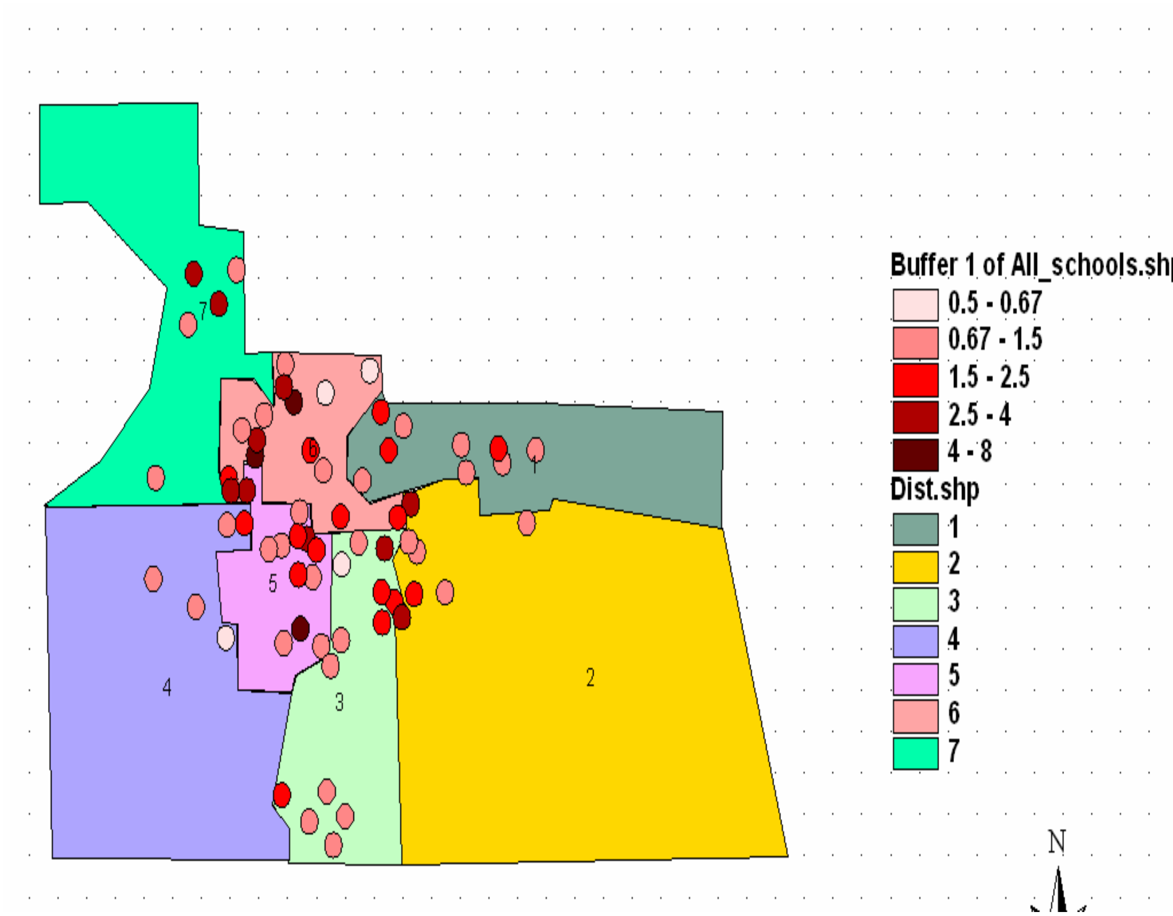

Figure 4-15: Buffer zones around schools representing the crashes per school rate

The legend for the buffer zone represents the crashes per school attribute. Observing Figure 4-15, we can see that the concentration of the "darker" buffer zones is in and around districts 5 and 6. In the earlier section, looking at the district wise crash rate, it was found that districts 5 and 6 have higher crashes per school rate. This is a clear indication that districts 5 and 6 have high potential of crash occurrence than the others.

The spatial analysis was carried out on each type of school separately. Out of the 157 schools located in Orange County, 108 are elementary schools, 27 are middle schools and 22 are high schools. Separate buffer zones were created around each type of school. It was done in order to check the effect of the presence of each school type on the crashes.

In the Elementary schools 109 crashes were inside buffer zones of half a mile radius. Figure 4-16 elaborates the elementary schools buffer zones. The legend for buffer zone represents the crashes per school attribute.

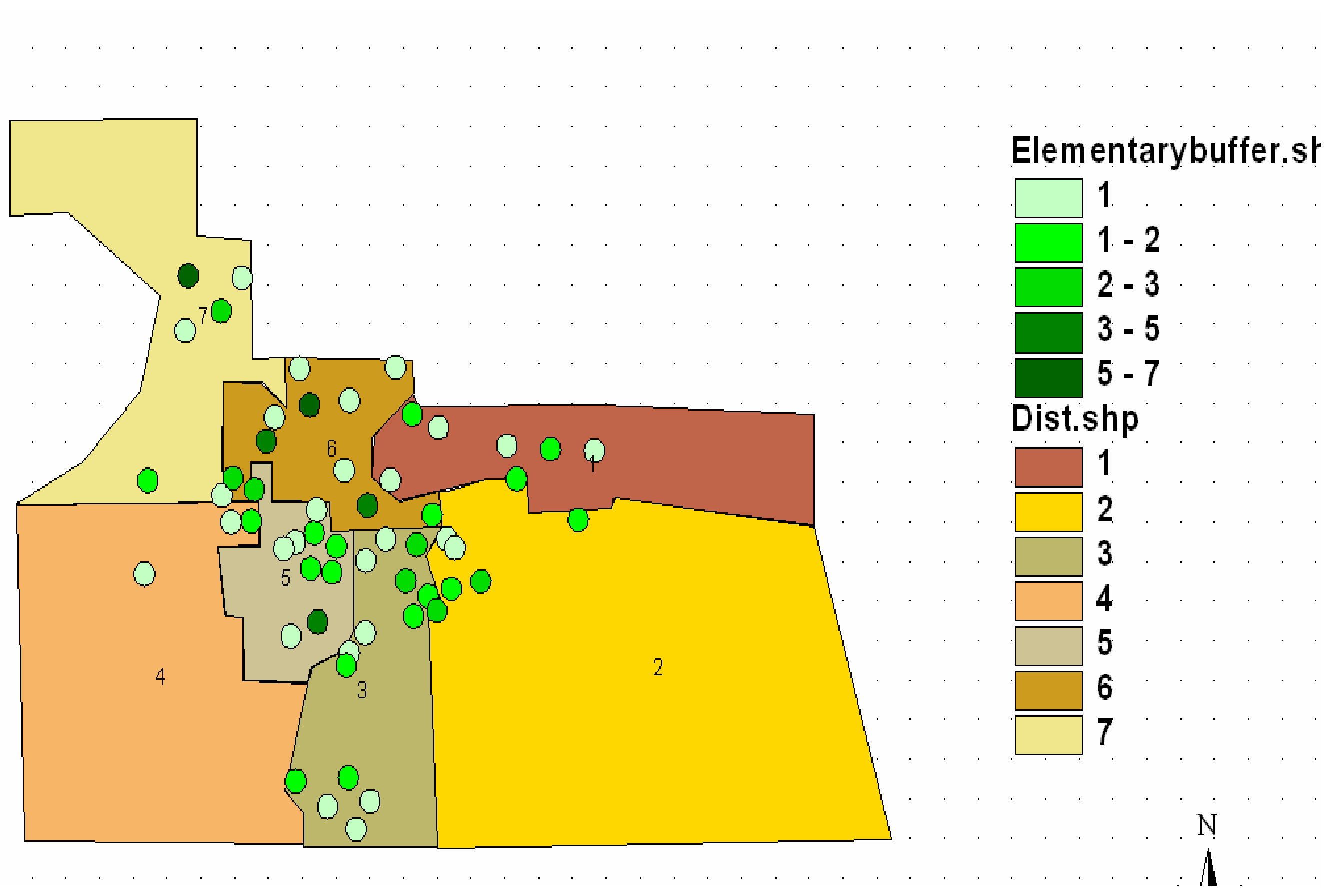

Figure 4-16: Elementary School Buffer Zones

Figure 4-16 indicates that a majority of the crashes influenced by the elementary schools fall in districts 3, 5 and 6. The mean crashes per school rate for elementary schools alone is calculated to be 1.00, which is marginally higher than the overall mean crashes per school rate of 0.94.

For the Middle Schools 41 crashes were located inside buffer zones of half a mile radii. Figure 4-17 indicates the middle schools buffer zones. The legend for buffer zone represents the crashes per school attribute.

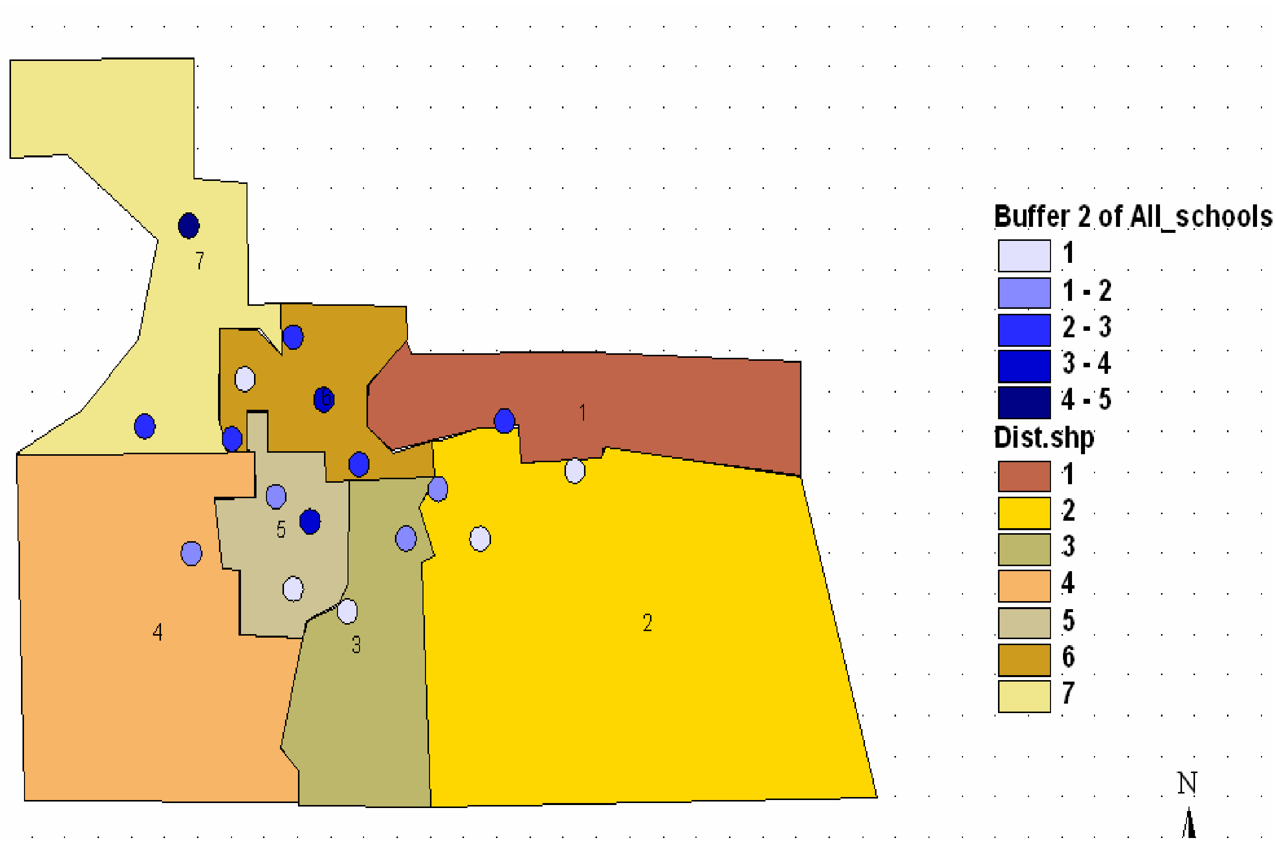

Figure 4-17: Middle School Buffer Zones

By close observation, it can be seen that even for middle schools, the crashes per school rate is higher in district 6. The mean crashes per school rate for middle schools alone is calculated to be 1.52, which is higher than the over all mean crashes per school rate of 0.94.

Finally, for high schools 35 crashes were located inside the buffer zones of half a mile radii. Figure 4-18 shows the high school buffer zones. The legend for buffer zone represents the crashes per school attribute.

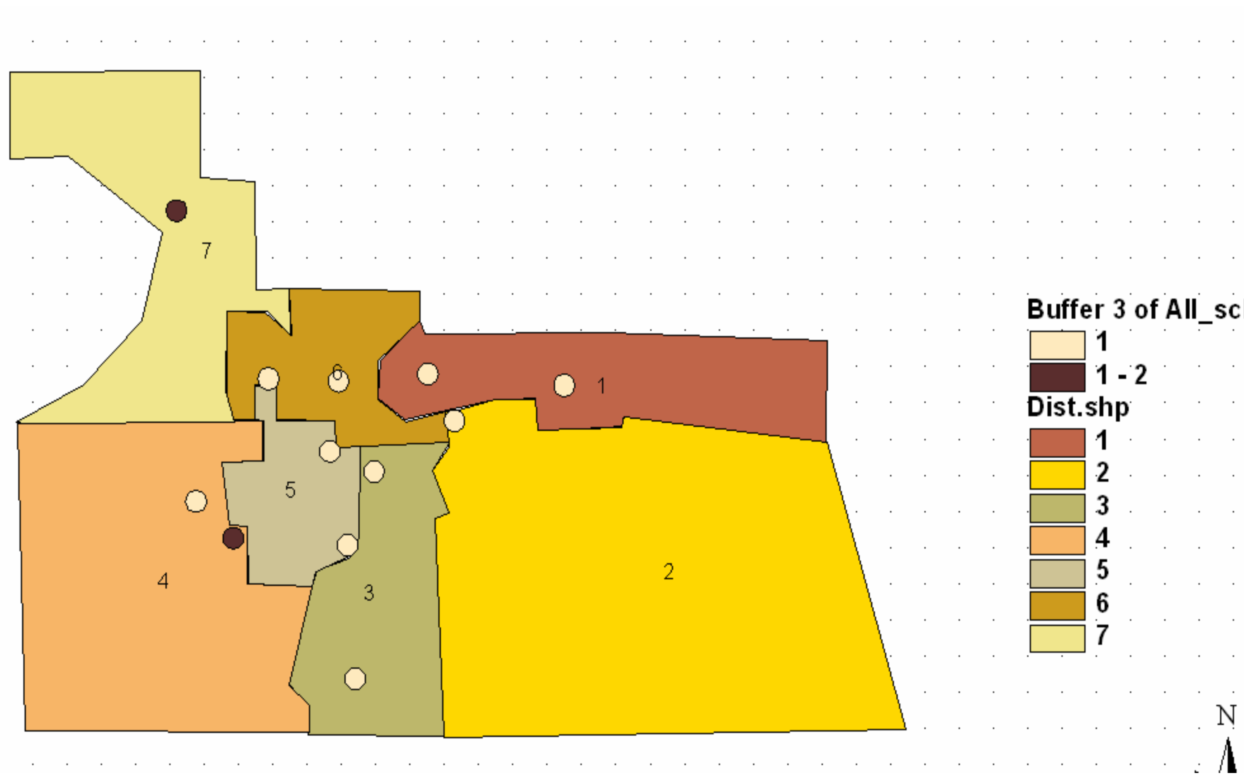

Figure 4-18: High School Buffer Zones

Unlike the previous two cases, the crashes influenced by the high school buffer zones are located in districts 4 and 7. The mean crashes per school rate for high schools alone are calculated to be 1.59, which is also higher than the over all mean crashes per school rate of 0.94.

By performing T-tests on the three means, it was observed that there was a significant difference between the means of middle schools and high schools with a P-value of 0.0008. Because of the difference in the sample size between the elementary schools and middle or high schools, the statistical significance couldn't be achieved. There are 108 elementary schools, 27 middle and 22 high schools. But considering a comparable, around 20, random sample size from the elementary schools, the P-value was found to be 0.021. This implies that the middle schools are at a higher risk than the elementary or high schools (please note that the differences in student population among the different school levels were not considered).

Figure 4-19 shows the overlay of major streets on the school buffer zones and the crash locations, on the orange county district map. From the above analysis, it was found that the majority of the crashes, which occurred within half a mile of the school locations, occurred in districts 1, 5 and 6. Some of the major streets that could have contributed to these crashes are sections of Hiawassee Rd, Powers Dr, Silver star Rd, Orange Blossom Trail, Hastings Rd, Balboa Dr, Forest City Rd, Edgewater Dr, Semoran Blvd and Colonial Dr, in districts 1, 5 and 6.

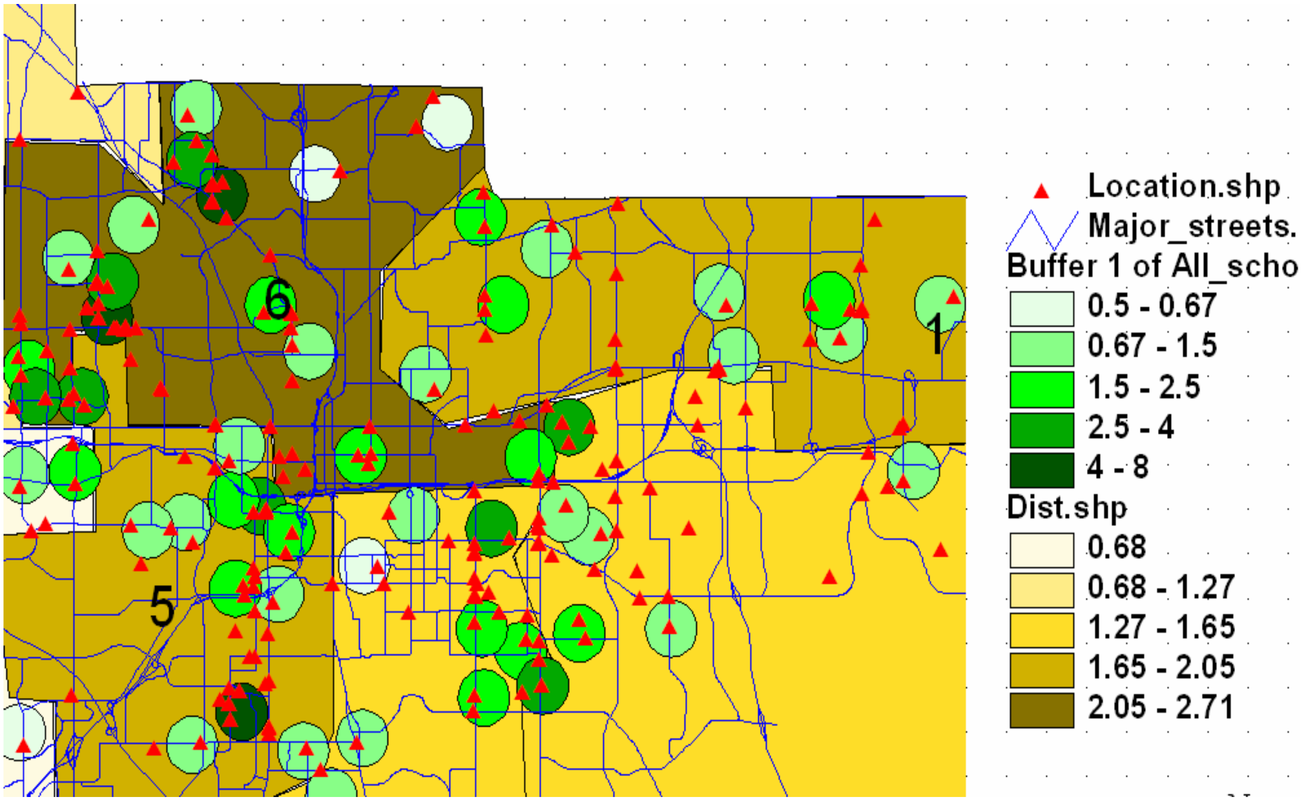

Figure 4-19: Overlay of major streets and crash locations on district map

Based on the buffers created around the elementary, middle and high school and by carrying out the above analysis, the following schools in table were found to be critical with a crash rate per school = 2.

Table 4-19: Critical Schools (Crashes per school =2)

Winter Park High School
Riverdale Elementary School
Lake George Elementary School
Lakemont Elementary School
Ventura Elementary School
Grand Av Elementary School
Orange Center Elementary School
Azalea Park Elementary School
John Young Elementary School
Hiawassee Elementary School
Shenandoah Elementary School
Orlo Vista Elementary School
Howard Middle School
Hillcrest Elementary School
Memorial Middle School
Catalina Elementary School
Conway Middle School
Conway Elementary School
Lee Middle School
Edgewater High School

Similar to the above list of schools, the schools with crashes per school equal to 3 is mentioned below in Table 4-20.

Table 4-20: Critical Schools (Crashes per school = 3)

Colonial High School
Robinswood Middle School
Pine Hills Elementary School
Dover Shores Elementary School
Lockhart Middle School
Mccoy Elementary School
Jones High School
Lovell Elementary School

Table 4-21 below gives the list of schools which could be said to be the most critical as their crashes per school rate is greater than 3.

Table 4-21: Most critical schools (crashes per school > 3)

Apopka Memorial Middle School
Dream Lake Elementary School
Rolling Hills Elementary School
Palmetto Elementary School
Lake Weston Elementary School
Evans High School

Findings and Conclusions

From the preliminary analysis, using simple two-way tables and graphs, we conclude the following.

- There are a significant number of pedestrian and bicyclist crashes in Orange County, related to school children.
- Of the three age groups, the children between the ages 4 to 11 seem to be at higher risk, as they share a larger crash population than the remaining two groups. That is, the number of crashes involving Elementary School Children is higher when compared to the Middle and High school children.
- Majority of the crashes took place on undivided two lane highways. Also most of the crashes took place in Primarily Residential areas where the posted speed limits are around 30 MPH.
- Most of the Crashes happened while the pedestrians were “crossing not at Intersection” and because of “Carelessness” of the Driver.

From the Spatial analysis, it can be concluded that

- Middle and High schools are at a higher risk when compared to elementary schools.
- Districts 1, 5 and 6 seem to have higher crashes per school ratio for crashes occurring within half a mile of the schools.

5. LOG-LINEAR ANALYSIS

5.1 Introduction

Log-linear models are used to analyze categorical data. They model the means of cell counts in contingency tables by describing the association patterns among a set of categorical variables without specifying any variable as a response (dependent) variable. The term log-linear comes from the form of the model; the natural logarithm of cell counts is modeled as a linear function of the effects of categorical variables and their interactions. For example, suppose that we want to investigate relationships between three categorical variables, X, Y and Z, where X has I categories, Y has J categories and Z, K categories. Then the full (saturated) log-linear model is

$$\log (m_{ijk}) = \lambda + \lambda_i^{X} + \lambda_j^{Y} + \lambda_k^{Z} + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ},$$

for each combination of the levels $i=1,2,\ldots, I$, $j=1, 2, \ldots, J$ and $k=1, 2, \ldots, K$, of the categorical variables X, Y and Z. In many situations, simpler models, containing a subset of the parameters from a saturated model may be adequate. For example,

(i) If all three categorical variables are mutually independent, then the following, much simpler model will describe the relationships between X, Y and Z:

$$\log(m_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z,$$

(ii) If X and Z are associated but both are independent of Y, then the following model will describe the relationships between X, Y and Z:

$$\log(m_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ},$$

(iii) If X and Y are conditionally independent, that is, X and Y are independent if we control for Z, then the following model will describe the relationships between X, Y and Z:

$$\log(m_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ},$$

(iv) If there is no three-factor interaction (a no three-factor interaction or homogeneous association model) then the model is

$$\log(m_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ},$$

and it implies that the conditional odds ratios between any two variables are the same at each level of the third variable.

The examples of log-linear models given above, (i) – (iv), are called hierarchical models. Hierarchical models include all lower order terms composed from variables in the highest terms in the model.

The parameters of log-linear models are usually estimated by the maximum likelihood (ML) method. Poisson or multinomial distributions of the cell counts are most commonly used in the log linear model analysis. The ML estimates are identical for both of these distributions.

The parameters of a log-linear model can be estimated by ML using PROC CATMOD in SAS. Proper interpretation of results depends on understanding the identifiability constraints on the parameters imposed by PROC CATMOD.

In order for log-linear models to be identifiable, that is, to have unique parameter estimates, certain constraints have to be imposed on the parameters. In PROC CATMOD, the constraints require that the sum of parameters over all categories of each variable be zero. For example, for a 2 X 2 table, that is, two variables, X and Y, both having two categories:

$\lambda_1^X + \lambda_2^X = 0,$

$\lambda_1^Y + \lambda_2^Y = 0,$

$\lambda_{11}^{XY} + \lambda_{12}^{XY} = 0,$

$\lambda_{21}^{XY} + \lambda_{22}^{XY} = 0,$

$\lambda_{11}^{XY} + \lambda_{21}^{XY} = 0,$

$\lambda_{12}^{XY} + \lambda_{22}^{XY} = 0,$

which implies that $\lambda_1^X = -\lambda_2^X$, $\lambda_1^Y = -\lambda_2^Y$ and $\lambda_{11}^{XY} = -\lambda_{21}^{XY} = -\lambda_{12}^{XY} = \lambda_{22}^{XY}$. It means that one parameter will be estimated for X (corresponding to the first category), one parameter for Y (also corresponding to the first category) and one for the interaction of X and Y (corresponding to the first categories of X and Y). The values of the remaining parameters can be determined from the equations above.

Model Building

The output of the first PROC CATMOD with the loglin statement `loglin x| y| z continues` the saturated model, main effects and all interactions. The "Maximum Likelihood Analysis of Variance" table displays significance tests for each effect in the specified model. The Chi-Square test for each effect is a Wald Test based on the information matrix from the likelihood calculations. The Likelihood Ratio statistic at the bottom is a goodness-of-fit test for the model. It compares the specified model with the saturated model and is equal to -2 times the difference of the log likelihoods for the specified and the saturated models. For the saturated models the difference is 0.

A saturated model is not very useful because of its complexity. For example, a four variable saturated model would have four main effects, six two-way interactions, four three-way interactions and one four-way interaction terms (15 terms in all). Analyzing such a model would be cumbersome when simpler models could provide almost similar amount of information with lesser number of variables (parameters). Therefore a saturated model needs to be reduced such that the model on the whole remains significant (overall p value > 0.05), the model effects (main and interaction) are significant (p<0.05) and the likelihood ratio statistic is as small as possible (least loss of information due to removal of effects from the saturated model). While removing effects from a saturated model, we begin by eliminating the highest order interaction effects with the highest p-value and continue like wise till an acceptable model fit is obtained.

Odds Ratio is a way of comparing whether the probability of certain event in the same for two groups. An odds ratio of 1 implies that the event is equally likely in both groups. An odds ratio greater than 1 implies that the event is more likely in the first group. An odds ratio less than 1 implies that the event is less likely in the first group. For example for a given 2 X 2 table,

Disease

Factor		+	-	
	+	a	b	a+b
	-	c	d	c+d
		a+c	b+d	N

The Analysis of Contingency Tables, 1991, pp.30

Figure 5-1: Example of Odds Ratio
Based on B.S.Everitt (1991)

the odds ratio for row + are a/b. The odds for row – are c/d. The odds ratio is simply the ratio of two odds and is equal to ad/bc.

The parameter estimates using PROC CATMOD can be used to compute odds ratios in log-linear models.

For example, if we are considering a log-linear model with three variables X, Y and Z each at levels I, J and K respectively and has the following form

$$\log m_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}$$

We want to compute the odds of Z=1 vs. Z=4 for X=1 against the odds of Z=1 vs. Z=4 for X=2.

The log of the desired odds ratio is

$$\begin{aligned}
\log((m_{1j1}\, m_{2j4})/(m_{2j1}\, m_{1j4})) &= \log m_{1j1} + \log m_{2j4} - \log m_{2j1} - \log m_{1j4} \\
&= \lambda_1^X + \lambda_1^Z + \lambda_{1j}^{XY} + \lambda_{11}^{XZ} + \lambda_2^X + \lambda_4^Z + \lambda_{2j}^{XY} + \lambda_{24}^{XZ} \\
&\quad - \lambda_2^X - \lambda_1^Z - \lambda_{2j}^{XY} - \lambda_{21}^{XZ} - \lambda_1^X - \lambda_4^Z - \lambda_{1j}^{XY} - \lambda_{14}^{XZ} \\
&= \lambda_1^X + \lambda_1^Z + \lambda_{11}^{XZ} + \lambda_2^X + \lambda_4^Z + \lambda_{24}^{XZ} \\
&\quad - \lambda_2^X - \lambda_1^Z - \lambda_{21}^{XZ} - \lambda_1^X - \lambda_4^Z - \lambda_{14}^{XZ}.
\end{aligned}$$

Now, taking into account the identifiability constraints of PROC CATMOD, which in this example are:

$$\lambda_1^X + \lambda_2^X = 0$$

$$\lambda_{11}^{XZ} + \lambda_{12}^{XZ} + \lambda_{13}^{XZ} + \lambda_{14}^{XZ} = 0$$

$$\lambda_{21}^{XZ} + \lambda_{22}^{XZ} + \lambda_{23}^{XZ} + \lambda_{24}^{XZ} = 0$$

$$\lambda_{11}^{XZ} + \lambda_{21}^{XZ} = 0$$

$$\lambda_{12}^{XZ} + \lambda_{22}^{XZ} = 0$$

$$\lambda_{13}^{XZ} + \lambda_{23}^{XZ} = 0$$

$$\lambda_{14}^{XZ} + \lambda_{24}^{XZ} = 0,$$

we get,

$$\begin{aligned}\log ((m_{1j1}\, m_{2j4})/(m_{2j1}\, m_{1j4})) &= \lambda_{11}^{XZ} + \lambda_{24}^{XZ} - \lambda_{21}^{XZ} - \lambda_{14}^{XZ} \\ &= 2\lambda_{11}^{XZ} - 2\lambda_{14}^{XZ} \\ &= 4\lambda_{11}^{XZ} + 2\lambda_{12XZ} + 2\lambda_{13}^{XZ}\end{aligned}$$

Hence, the odds ratio is exp $(4\lambda_{11}^{XZ} + 2\lambda_{12XZ} + 2\lambda_{13}^{XZ})$.

5.2 Methodology, Model Building and Estimation

In the preliminary analysis section, the variables considered were driver age, pedestrian/bicyclist age, driver sex, pedestrian/bicyclist sex, month of the crash, day of crash, Traffic Control, Site locations, injury severity, pedestrian action, alcohol use of driver, alcohol use of pedestrian/bicyclist, vehicle type, vehicle movement, contributing cause of pedestrians, no. of lanes, type of median, contributing causes of driver (CCD), posted speed limit, speed ratio and location type. Building a model with all these variables would have been cumbersome and even unnecessary when the same information could be obtained by using lesser number of variables. Hence, it was decided to build log-linear models only with those variables which would have a significant impact on the model fit.

Choosing the required variables from among the above listed variables was done by looking at the joint occurrences of various variables. Individual frequency tables were also considered. Those variables with no significant occurrences in the frequency tables were ignored from further analysis. For instance, the effect of alcohol on pedestrians/bicyclists was divided into two categories – Under Influence and Not under Influence. By looking at the frequency distribution it was observed that out of the 444 crash cases, 432 (97.3%) were Not Under Influence and the remaining 12 cases (2.7%) were under influence. This clearly is an indication that Pedestrians under influence will not be affecting our model in any way as they are of negligible size. And hence, this variable was removed from the log-linear analysis. Similar logic and reasoning was followed in the decision making process of eliminating or keeping a variable and finally, the following variables were considered to be having some significant importance. The variables are Driver age, Pedestrian/bicyclist age, Driver sex, Pedestrian/Bicyclist sex, Alcohol usage by driver, CCD, posted speed limit, Speed ratio, vehicle type, vehicle movement, Traffic Control, Site Location, no. of lanes and median type.

The next step in model building is the categorization of the variables. It is essential for a log-linear model that all the variables involved in the model are categorical. They were classified in such a way that there exist a comparable number of crashes among all categories of a variable. This was achievable by looking at their percentile values. The above variables were categorized as shown in Table 5-1 based on their frequency distribution and percentile values.

Table 5-1: Categorization of Variables

Variable Name	Categories
Driver Age	1. <= 25 (Young) 2. 26 to 50 (Middle) 3. >50 (Old)
Driver Sex	1. Male 2. Female
Pedestrian/Bicyclist Age	1. 4 to 11(Elementary) 2. 12 to 14 (Middle) 3. 15 to 18 (High)
Pedestrian/Bicyclist Sex	1. Male 2. Female
CCD	1. No Improper Driving 2. Careless, failed-to-yield, improper behavior 3. Others
Traffic control (TC)	1. No Traffic control 2. Traffic Signal, Stop sign, Yield sign 3. Others 4. Special Speed zone, school zone, flashing light
Vehicle Type	1. Passenger cars 2. SUV's , pickup trucks etc
Site Location (SL)	1. Not at intersection 2. At intersection, Influenced by intersection 3. Driveway access, Parking lots, Private Property 4. Others
Alcohol Usage by driver	1. Under Influence 2. Not Under Influence
Lanes	1. <=2 2. >2 3. Parking lots etc.
Median Type	1. Divided 2. Undivided
Vehicle Movement	1. Going Straight Ahead 2. Slowed/stopped/Making right/left turn 3. Others
Speed Limit	1. <=25 MPH 2. >=26 and <=40MPH 3. >40MPH
Speed ratio	1. <=0.9 2. >0.9 and <=1.1 3. >1.1

A crash occurring at a place would normally involve a driver(s), vehicle, and pedestrian/bicyclist(s). Other characteristics like road characteristics, traffic characteristics, environmental effects and even pedestrian/bicyclist characteristics (if a pedestrian/bicyclist is involved) are also a part of the crash and crash conditions. As we are dealing with pedestrian/bicyclist crashes in particular, the pedestrians/bicyclists would be involved in all crashes. Since the overall objective is to identify what geometric, traffic, driver and pedestrian/bicyclist characteristics are critical in pedestrian/bicyclist crashes, it is essential to divide the available variables as belonging to either driver, pedestrian/bicyclist or other (including geometric, vehicle and traffic) characteristics. By dividing the variables into these classes, it would be interesting to interpret the various combinations of characteristics that are more likely to be involved in pedestrian/bicyclist related crashes. Also, building a model would be easier and better understood when lesser variables would be involved. The variables are classified in the three groups as follows.

Table 5-2: Variable Groups

Driver Characteristics	**Pedestrian/bicyclist Characteristics**	**Other Characteristics**
Driver Age Driver sex CCD Alcohol use by driver Speed Ratio Vehicle movement	Pedestrian/bicyclist Age Pedestrian/bicyclist sex	Lanes Median type Traffic Control Site Location Speed Limit Vehicle type

Two kinds of log-linear models would now be built using the three above mentioned groups of variables – driver related (Driver Models) and Pedestrian/bicyclist related (Pedestrian/bicyclist models). The driver related models would involve a combination of variables from the Driver group and other characteristics, while the Pedestrian/bicyclist models would involve a combination of variables from the Pedestrian/bicyclist group and other characteristics. The purpose of doing this is to understand microscopically the effect of driver characteristic variables on other characteristics in pedestrian/bicyclist related crashes. Similarly, the effect of pedestrian/bicyclist characteristics on the other characteristic variables in pedestrian/bicyclist related crashes would be known.

5.3 Driver models

As mentioned in the discussion above, the driver models consist of variables from the driver characteristics and the other characteristics of Table 5-2. After trying out various combinations of model fits using the different combinations of variables from both the groups, the following models had significant p-values (over all p-value>0.05) and were used to estimate the odds multipliers.

Model 1- Driver age, Vehicle type, Median

Model 2- Driver age, Lanes, Median, Speed limit

Model 3- Driver age, Lanes, Median, Speed ratio

Model 4- Driver age, Lanes, Median, Driver sex

Model 5- Driver age, Alcohol, CCD

Model 6- Driver age, Site Location, Driver sex

Model 7- Driver age, Vehicle type, Vehicle Movement, Alcohol

Model 8- Driver age, Speed limit, Traffic Control, Driver Sex

The methodology for deciding on these models has been given below. Also, the calculation of odds multipliers for the variables in the model has been elucidated.

5.3.1 Model 1

Model 1 includes the variables driver age (drivage), vehicle type (vehtype) and median (median). The output of the first PROC CATMOD with the loglin statement `loglin Drivage| Vehtype| median;` (the saturated model, main effects and all interactions are included) contains the following table.

Table 5-3: Likelihood Analysis of variance for saturated Model

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
Drivage	2	31.08	<.0001
Vehtype	1	22.97	<.0001
drivage*Vehtype	2	31.43	<.0001
median	1	27.45	<.0001
drivage*median	2	0.90	0.6365
Vehtype*median	1	0.05	0.8275

drivage*Vehtype*median	2	2.27	0.3212
Likelihood Ratio	0	.	

The Likelihood Ratio statistic at the bottom is a goodness-of-fit test for the model. It compares the specified model with the saturated model and is equal to -2 times the difference of the log likelihoods for the specified and the saturated models. Since in this example the specified and the saturated models are the same, the difference is 0.

The Wald test for drivage*vehtype*median does not indicate significance of the third order interaction. We can remove it from the model and rerun PROC CATMOD with the loglin statement loglin `drivage| vehtype| median @2;` which specifies a model with the main effects and all possible interactions of order 2. Here is the "Maximum Likelihood Analysis of Variance" table for this model.

Table 5-4: Likelihood Analysis of Variance -second order model

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
drivage	2	44.68	<.0001
Vehtype	1	31.92	<.0001
drivage*Vehtype	2	39.08	<.0001
median	1	61.81	<.0001
drivage*median	2	1.65	0.4392

Vehtype*median	1	0.04	0.8483
Likelihood Ratio	2	2.31	0.3144

Since the Likelihood Ratio statistic compares the model without the third order interaction with the saturated model, it is the likelihood ratio test for the significance of drivage*vehtype*median, that is, it tests the same hypothesis as the Wald test for drivage*vehtype*median in the previous “Maximum Likelihood Analysis of Variance" table (for the saturated model). The Wald test indicates that the vehtype*median interaction is not significant. We will remove it from the model and test if the model of conditional independence of vehtype and median fits the data.

Table 5-5: Likelihood Analysis of variance without Vehtype*median interaction

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
drivage	2	45.32	<.0001
Vehtype	1	36.00	<.0001
median	1	69.26	<.0001
drivage*Vehtype	2	39.13	<.0001
drivage*median	2	1.81	0.4045
Likelihood Ratio	3	2.35	0.5029

The Wald test indicates that the drivage*median interaction is not significant. We will remove it from the model and test if the model of conditional independence of drivage and median fits the data.

Table 5-6: Likelihood analysis of variance of best fit

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
Drivage	2	45.55	<.0001
Vehtype	1	36.00	<.0001
Median	1	77.34	<.0001
drivage*Vehtype	2	39.13	<.0001
Likelihood Ratio	5	4.23	0.5162

The model fits the data reasonably well, as indicated by the Likelihood Ratio test (p-value = 0.5162). Also the Chi-Square value is below 11.071. The output includes also a table containing parameter estimates as shown below.

Table 5-7: Parameter Estimates

Analysis of Maximum Likelihood Estimates					
Parameter		**Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**
Drivage	**1**	-0.9862	0.1462	45.52	<.0001
	2	0.4856	0.0907	28.64	<.0001
Vehtype	**1**	0.4758	0.0793	36.00	<.0001
Median	**1**	-0.4629	0.0526	77.34	<.0001
drivage*Vehtype	**1 1**	0.7526	0.1462	26.51	<.0001
	2 1	-0.1886	0.0907	4.32	0.0377

These parameter estimates in the Table 5-7 can be used to compute odds ratios. The calculation of odds ratio has been discussed earlier in this chapter.

The final model where drivage has I levels, vehtype has J levels and median has k levels is given as

Log $m_{ijk} = \lambda +$ drivage $_i$ + vehtype $_j$ + median $_k$ + drivage*vehtype $_{(i,j)}$

The odds multipliers for the main effect are calculated to be as follows.

Table 5-8: Main effect Odds multipliers

Parameter	**Odds Multipliers**
Median	
1. Divided	1
2. Undivided	2.52

The odds multipliers for the interaction term Driver age * Vehicle type are calculated to be as follows.

Table 5-9: Interaction terms odds multipliers

Parameter	**Vehtype**	
Driver Age	**Cars**	**SUV/Trucks**
1. Young (<=25)	1	1
2. Middle (>25 and <=50)	1.69995	11.1674
3. Old (>50)	1.18554	16.5007

Interpretation of odds multipliers:

The odds multipliers in Table 5-8 and Table 5-9 can be interpreted as follows. In table 8, the odds multiplier for the undivided highway is 2.5 times higher than the divided highway. This means that the chances of a pedestrian/bicyclist crash occurring on undivided highway is at least 2.5 times higher than that of a divided highway. From Table 5-9, it can be observed that the chance of a crash happening is highest when the driver is over 50 years and is driving a vehicle other than a passenger car.

5.3.2 Model 2

Model 2 included the variables Driverage, lanes, median and speed limit. Similar methodology as followed in developing model 1 was carried out here. The final model is

Log $m_{ijk} = \lambda +$ drivage $_i$ + lanes $_j$ + median $_k$ + speed limit $_l$ + drivage*lanes $_{(i,j)}$ + lanes * median $_{(j,k)}$ +lanes * speed limit $_{(j,l)}$

The over-all p-value was significant at 0.6312. The chi-square value of 21.13 is below 36.415. The Odds multipliers for the interaction terms are given in Table 5-10 and Table 5-11.

Table 5-10: Odds multipliers of model 2

Parameter	Lanes	
Driver age	**<=2**	**>2**
Young	1	1
Middle	1.84	5.21
Old	1.81	2.92

Table 5-11: Odds multipliers of model 2 (contd.)

Parameter	**Median**		**Speed Limit**		
Lanes	**Divided**	**Undivided**	**<=25**	**>=26 - <=40**	**>40**
<=2	1	1	1	1	1
>2	2.65	0.25	0.35	0.93	0.36

Interpretation of Odds multipliers:

From Table 5-11, the odds multipliers for median type show that there is a higher chance for crash occurrence on two or less than two lane undivided roads, while there's atleast 2.7 times higher chance of crash occurrence on greater than 2 lanes divided highways. Considering the speed limits, the chance of occurrence of crash on a two lane road with speed limit < 25 MPH is higher than that of greater than 2 lanes road at the same speed limit. In general the crash occurrence when there are less than or equal to two lanes is higher than the other lane group categories across all the speed limits.

It can be observed from Table 5-10, the involvement of middle aged drivers across all lane groups is higher than the other two age groups.

5.3.3 Model 3

Model 3 is very much similar to model 2. The only exception being speed limit has been replaced by speed ratio. By doing such a thing, the model fit turned out to be better and more efficient. The final model by using speed ratio instead of speed limit is

Log $m_{ijk} = \lambda +$ drivage $_i$ + lanes $_j$ + median $_k$ + speed ratio $_l$ + lanes * median $_{(j,k)}$ +lanes * speed ratio$_{(j,l)}$

The over all p-value of the model turned out to be 0.1016. Chi-square value for the model is 28.34 which is less than 31.41. The odds multipliers for the model fit are given Table 5-12 and Table 5-13.

Table 5-12: Odds Multipliers (Model-2)

	Median		**Speed ratio**		
Lanes	**Divided**	**Undivided**	**<=0.9**	**>0.9-<=1.1**	**>1.1**
<=2	**1**	**1**	1	1	1
>2	1.07	0.54	1.06	0.55	2.11

Table 5-13: Odds Multipliers (Contd.)

Parameter	**Odds Multipliers**

Speed Ratio	
<=0.9	1
>0.9-<=1.1	0.344
>1.1	1.25

Interpretation of Odds multipliers:

The odds multipliers of speed ratio are higher when the driver is speeding (speed ratio >1.1). Also, the chances of crash occurrence are higher when speeding occurs on more than two lanes.

By using speed ratio instead of speed limit, we are able to reduce the interaction terms and also able to interpret the odds multipliers in a better way. But the disadvantage in using speed ratio in place of speed limit in our analysis is the lack of sample size. Nearly 43% of the sample size had unknown estimated speeds. And hence the sample size reduced from 444 to 251. Just because this particular model could be fitted with this sample size, doesn't necessarily make it convenient to use speed ratio to fit other models. The sample size criterion does not fit. And so, speed ratio could not be used a significant variable for the other models.

5.3.4 Model 4

Model 4 included the variables Driverage, lanes, median and driver sex. Similar methodology as followed in developing model 1 was carried out here. The final model is

Log m_{ijk} = λ+ drivage $_i$ + lanes $_j$ + median $_k$ + driver sex $_l$ + drivage*lanes $_{(i,j)}$ + lanes * median $_{(j,k)}$

The over-all p-value was significant at 0.6159. The chi-square value for the over all model is 20.43 which is less than 35.173 at 23 degrees of freedom. The Odds multipliers for the interaction terms are given in Table 5-14, Table 5-15 and Table 5-16.

Table 5-14: Odds multipliers (Model-4)

Parameter	Odds Multipliers
Driver Sex	
1. Male	1
2. Female	0.65

Table 5-15: Odds multipliers - interaction terms

Parameter	**Lanes**	
Driver age	**<=2**	**>2**
Young	1	1
Middle	1.79	4.81
Old	1.02	1.75

Table 5-16: Odds multipliers-interaction terms (contd.)

Parameter	**Median**	
Lanes	**Divided**	**undivided**
<=2	1	1
>2	2.55	0.26

Interpretation of Odds multipliers:

From Table 5-14, the odds of male drivers getting involved in crashes is 35% times higher than females.

5.3.5 Model 5

Model 5 included the variables Driverage, alcohol and CCD. Similar methodology as followed in developing model 1 was carried out here. No interaction terms were found to be significant. Hence the final model was left with just the main effects.

$$\text{Log } m_{ijk} = \lambda + \text{drivage}_i + \text{alcohol}_j + \text{CCD}_k$$

The over-all p-value was significant at 0.1662. The chi-square value of 9.13 is less than 12.592 at 6 degrees of freedom. The Odds multipliers for the interaction terms are given in Table 5-17.

Table 5-17: Odds multipliers

Parameter	**Odds Multipliers**
Driver Age	
1. Young	1
2. Middle	2.45
3. Old	1.13
CCD	
1. No improper Action	1
2. Careless, failed to yield right-of-way,	0.43
3. Others	0.18
Alcohol	
1. Under Influence	1
2. Not under influence	0.1

Interpretation of Odds multipliers

From Table 5-17, the odds of middle aged drivers getting involved in crashes is in general 2.5 times higher than younger aged drivers. Also, the risk of being a part of crash is 10 times higher when the driver is under influence.

5.3.6 Model 6

Model 6 included the variables Driverage, Site Location (SL) and driver sex. Similar methodology as followed in developing model 1 was carried out here. The final model is

$$\text{Log } m_{ijk} = \lambda + \text{drivage}_i + SL_j + \text{driver sex}_k$$

The over-all p-value was significant at 0.2937 for the chi-square of 19.63 and 19 degrees of freedom. The Odds multipliers for the interaction terms are given in Table 5-18.

Table 5-18: Odds Multipliers

Parameter	Odds Multipliers
Driver Age	
1. Young	1
2. Middle	2.45
3. Old	1.25
Site Location	
1. Not at Intersection	1
2. At/influenced by intersection	0.75
3. Driveway Access, Parking Lots	0.45
4. Other	0.24
Driver Sex	
1. Male	1
2. Female	0.65

Interpretation of Odds multipliers

From the above table, it can be seen that the odds for a crash to occur is higher "Not at Intersection". Also the odds of a male driver being involved in a crash are higher than that of a female driver. As seen earlier, the chances of middle aged driver being involved is higher when compared to the other two age groups.

5.3.7 Model 7

Model 7 included the variables Driverage, vehicle type (vehtype), vehicle movement (vehmov) and alcohol. Similar methodology as followed in developing model 1 was carried out here. The final model is

$$\text{Log } m_{ijk} = \lambda + \text{drivage}_{i} + \text{vehtype}_{j} + \text{vehmov}_{k} + \text{alcohol}_{l} + \text{drivage*vehtype}_{(i,j)} + \text{vehtype * alcohol}_{(j,l)}$$

The over-all p-value was significant at 0.3942. The chi-square value for 14 degrees of freedom was 14.77 which is significantly less that 23.685 for p=0.05. The Odds multipliers for the interaction terms are given in Table 5-19 and Table 5-20.

Table 5-19: Odds multipliers main effects

Parameter	**Odds Multipliers**
Vehicle movement	
1. Straight Ahead	1
2.Slowed/Stopped/Making Right/Left Turns	0.65
3. Others	0.8

Table 5-20: Odds multipliers-interaction terms

Parameter	Vehtype	
Driver Age	**Cars**	**SUV/Trucks**
1. <=25	1	1
2. >25 and <=50	1.7	11.2
3.>50	0.77	6.17

Interpretation of Odds multipliers

Table 20 indicates that the chance of a crash occurrence is the highest when a middle aged driver is using an SUV/Truck. Also, the probability of a crash occurrence is higher when the vehicle movement is "Straight Ahead".

5.3.8 Model 8

Model 8 included the variables Driverage, speed limit, Traffic control (TC) and driver sex. Similar methodology as followed in developing model 1 was carried out here. The final model is

$$\text{Log } m_{ijk} = \lambda + \text{drivage}_i + \text{speed limit}_j + \text{TC}_k + \text{driver sex}_l$$

The over-all p-value was significant at 0.8038. The chi-square value for the model was found to be 40.42 which was lesser than 67.5 for p=0.05. The Odds multipliers for the interaction terms are given in tables 11 and 12.

Table 5-21: Odds multipliers for model 8

Parameter	**Odds**
Driver age	
1.Young	1
2. Middle	2.50
3. Old	1.35
Speed Limit	
<=25MPH	1
>25 to <=40MPH	1.20
>40 MPH	0.47
Traffic Control	
1.No Traffic Control	1
2. Traffic Signal/Stop Sign/ Yield Sign	1.16
3. Others	0.14
4. Special Speed Zone/School Zone/Flashing Zone	0.41
Driver Sex	
1.Male	1
2.Female	0.63

Interpretation of Odds multipliers

Table 5-21 depicts that the chances for a crash occurrence are higher when the driver is traveling at around 30 MPH to 40MPH.

5.4 Pedestrian/Bicyclist Models

Just as the driver models were built, the pedestrian/bicyclist models were also developed. The variables in pedestrian/bicyclist characteristics group and other characteristics group were mixed and matched and the following models were developed.

Model 1- Pedestrian/bicyclist age, Lanes, Median, Speed ratio

Model 2- Pedestrian/bicyclist age, SL, Lanes, Pedestrian/bicyclist sex

Model 3- Pedestrian/bicyclist age, Speed Limit, Pedestrian/bicyclist sex

The model building procedure for these three models is very similar to the driver models procedure.

5.4.1 Model 1

Model 1 included the variables Pedestrian/bicyclist age (Pedage), Lanes, Median and speed ratio. Similar methodology as followed in developing driver model 1 was carried out here. The final model is

Log $m_{ijk} = \lambda +$ pedage $_i$ + lanes $_j$ + median $_k$ + speed ratio $_l$ + Pedage * Lanes $_{(i,j)}$ +Lanes * Median $_{(j,k)}$ + pedage*speed ratio$_{(i,l)}$

The chi-square value was 21.45 and was way below the required 24.996 for p=0.05. The over-all p-value was significant at 0.4696. The Odds multipliers for the interaction terms are given in Table 5-22.

Table 5-22: Odds multipliers for Pedestrian/bicyclist Model -1

Parameter	Lanes		Speed Ratio		
Pedage	**<=2**	**>2**	**<=0.9**	**>0.9 & <=1.1**	**>1.1**
Elementary	1	1	1	1	1
Middle	1.44	2	0.79	0.45	2.31
High	0.99	4.59	0.44	0.14	4.47

Table 5-23: Odds multipliers Lanes against Median

Parameter	Median	
Lanes	**Divided**	**Undivided**
<=2	1	1
>2	3.32	0.32

Interpretation of Odds multipliers

The odds multipliers in Table 5-22 show that the chances of crash occurrence for middle school children is higher than high school or elementary school, when there are either 2 or les than 2 lane roads. In the case of more than 2 lane roads, the probability of crashes involving high school children is much higher than the middle school children. Also, from Table 5-23, it can be seen that the similar to the earlier results related to lanes and median type, the chances of crashes occurring on undivided and less than or equal to two lane roadways is three times higher than that of more than 2 lane undivided roadways.

5.4.2 Model 2

Model 2 included the variables Pedestrian/bicyclist age (Pedage), Site Location, Lanes and pedestrian/bicyclist sex (pedsex). Similar methodology as followed in developing driver model 1 was carried out here. The final model is

$$\text{Log } m_{ijk} = \lambda + \text{pedage}_i + \text{sl}_j + \text{lanes}_k + \text{pedsex}_l + \text{Pedage} * \text{Lanes}_{(i,k)} + \text{SL} * \text{Pedsex}_{(j,l)}$$

The over-all p-value was significant at 0.6050. The over all chi-square value was found to be 36.05 at 36 degrees of freedom. It was significantly less than 50.964 at p=0.05. The Odds multipliers for the interaction terms are given in Table 5-24 and Table 5-25.

Table 5-24: Odds multipliers

Parameter	**Lanes**	
Pedestrian/bicyclist age	**<=2**	**>2**
Elementary	1	1
Middle	1.09	1.06
High	0.68	1.53

Table 5-25: Odds multipliers contd.

Parameter	**Pedestrian/bicyclist Gender**	
Site Location	**Male**	**Female**
1. Not at Intersection	1	1
2.At/influenced by intersection	1.06	1.36
3. Driveway Access, Parking Lots	0.25	0.26
4. Other	0.07	0.27

Interpretation of Odds multipliers

The odds multipliers in Table 5-24.show that the chances of crash occurrence for middle school children is higher than high or elementary schools, when there are 2 or les than 2 lanes or less. In the case of more than 2 lane roads, the odds of crashes involving high school children are much higher than the middle and elementary school children. Also, from Table 5-25, Gender seems to be interesting criteria. Both the male and female pedestrians/bicyclists are at higher risk to be involved in a crash "At or Influenced by an Intersection" than the rest of the site location categories.

5.4.3 Model 3

Model 3 included the variables Pedestrian/bicyclist age (Pedage), Speed limit and pedestrian/bicyclist sex (pedsex). Similar methodology as followed in developing driver model 1 was carried out here. The final model is

Log $m_{ijk} = \lambda +$ pedage $_i$ + speed limit $_j$ + pedsex $_k$ + Pedage * speed limit $_{(i,j)}$ + speed limit * pedsex $_{(j,k)}$

The over-all p-value was significant at 0.6758. The chi-square value was 4.01 at 6 Degrees of freedom and was less than 12.592 (p=0.05). The Odds multipliers for the interaction terms are given in Table 5-26 and Table 5-27.

Table 5-26: Odds multipliers-Pedestrian/bicyclist Model 3

Parameter	**Speed Limit**		
Pedestrian Age	**<=25**	**>=26-<=40**	**>40**
Elementary	1	1	1
Middle	0.76	1.32	1.25
High	0.57	1.07	2.58

Table 5-27: Odds Multipliers model 3 Contd.

Parameter	**Pedestrian Sex**	
Speed Limit	**Male**	**Female**
<25	1	1
26-40	0.97	1.55
>40	0.46	0.46

Interpretation of Odds multipliers

By careful observation of the two tables above, it can be said that the odds of a middle aged school child crash are higher when the child is female and when the speed limit is between 26 and 40 MPH.

Another Crash variable –Crash type (levels 1) pedestrian 2) bicyclists) was introduced to check the significance of type of crash (pedestrian or bicyclist) on crash frequency. The overall model significance deteriorated in almost all the models when this variable was introduced. Therefore this variable was discarded from the final models. This means that number of crashes did not depend greatly on the type of crash.

Finally, a model consisting of all the variables considered was built. Such a model will not be very useful because of the sample size restrictions and most of the effects will be insignificant. A reduction of effects was carried out so that there were atleast a few main effects left in the model. Ignoring the sample size constraints, the best model that could be arrived at is the one with the variables drivage, pedage, vehtype, speedlimit, lanegroup and median type. The significance of these effects has been displayed in the appendix. Since this model still doesn't comply with the sample size constraint its utility is severely restricted because the parameters have high standard errors. Further reduction of effects, in order to comply with the sample size constraints, results in model identical to model 2 of the driver models. The effects and results of this model have been discussed in the earlier paragraphs. In other words, the best comprehensive model that satisfies our sample size constraint is model 2.

$$\text{Log } m_{ijk} = \lambda + \text{drivage}_{i} + \text{lanes}_{j} + \text{median}_{k} + \text{speed limit}_{l} + \text{drivage*lanes}_{(i,j)} + \text{lanes * median}_{(j,k)} + \text{lanes * speed limit}_{(j,l)}$$

6 RESULTS AND CONCLUSIONS

From the results of log-linear models it has been seen that some levels of specific variables are more likely to be related to crashes. This was concluded based on the values of odds multipliers for each level of the variable compared to the base level of the variable. Since the objective was to analyze the relationship between the driver/pedestrian characteristics with geometric characteristics it makes more sense to interpret the odds multipliers for the variables under driver and pedestrian characteristics separately along with their interactions with the other characteristics.

6.1 Driver Characteristics

6.1.1 Driver Age

Driver age entered in all the models for the driver characteristics. It is the most important driver characteristic with which we relate all other characteristics. In the models considered driver age has significant interactions with Vehicle Type, Lane Group, Speed Limit and Speed ratio. From the odds multipliers for driver age in models with no interactions (models 5, 6 and 8) we find that middle ages drivers are 2.5 times more likely to be involved in pedestrian crashes than other age groups. This can be explained by the fact that middle aged drivers are mostly commuters and therefore more likely to be on the road in morning and afternoon school beginning and ending timings. Since our crashes are being modeled for the same time periods it is not surprising to see that the middle aged drivers are more involved than the other age groups.

When we consider interaction with vehicle type (model 1) we find that for cars middle aged drivers are more susceptible to crashes than other groups, while older aged drivers are more involved when driving SUV’s or trucks. This is attributed to the fact that the perception-reaction time of older aged drivers is more when compared to the other age groups. Also, maneuvering bigger vehicles like SUV’s or trucks is comparatively more difficult than cars. A combination of these two factors makes older drivers driving SUV’s and Trucks more involved in crashes.

Considering the interaction with speed ratio (model 3), speed limit (model 2) and lanes group (models2 and 4) the odds multipliers indicate that the middle aged drivers are the most involved in pedestrian/bicyclist crashes due to the reasons elucidated above. Older drivers relatively have lesser distractions (Distracted Driver report, December 2004) and their involvement level is lesser compared to other age groups.

6.1.2 Driver Sex

Driver sex appears just as a main effect in models 4, 6 and 8. It bears no significant interaction with any other variables. From the odds multipliers for driver sex from respective models it can be inferred that male drivers are more involved in pedestrian/bicyclist crashes than female drivers. In general, the male driver population is more prone to crashes (DHSMV, Traffic Crash Statistics Report, 2003). This trend is continued even when pedestrians are involved.

6.1.3 Alcohol Usage

Alcohol usage is significant in models 5 and 7. It has been established that a person under influence is more involved in crashes than a normal person. The odds multipliers from the respective models show that the odds of a person under influence are 10 times higher than a person not under influence. This is in accordance with the stated fact that persons in inebriated state are more likely to be involved in crashes.

6.1.4 CCD

Contributing cause of the driver appears in model 5. It does not bear any significant relation with any other variables. It is observed that the odds of “No improper Action” by the driver are at a higher level than the other levels. This means that the pedestrian crashes tend to occur at no fault of the driver’s. This might be expected as the crashes involving children took place mainly “not at intersection”. The children could be expected to be crossing the streets not at intersections while estimating the driver’s speed wrong.

6.1.5 Speed ratio

Speed ratio appears in model 3 and it has a significant interaction with lanes. The odds multipliers for this interaction indicate that higher number of crashes occurs when there is greater than two lanes and the driver is speeding.

6.2 Pedestrian Characteristics

6.2.1 Pedestrian Age

Pedestrian age enters in all the pedestrian models. Similar to the driver age, the pedestrian age is a very critical variable that needs to be tested for interaction with all other variables in order to understand the crash characteristics better. Pedestrian age has interactions with lanes (model 1), speed limit (model 3) and pedestrian sex (model 3). Interactions with the lanes shows that middle school aged children have higher odds of crash occurrence than the other age groups when the lane type is two or les than two lanes. Also, the high school aged children seem to be at higher odds when there are greater than 2 lanes. This is quite possible as the middle school aged children most often would be prepared to cross street with 2 or less than 2 lanes than streets with more lanes. High school children might easily cross two lane roads but are open to mis-judgement while crossing wider roads with more than 2 lanes.

Interaction with speed limit indicates that middle school aged children have higher odds of crash occurrence than high school children when the speed limit is less than 40 MPH. For speeds with greater than 40 MPH the high school children are more likely to be involved in crashes. This could also be explained in the same way as above. Middle aged children would be prepared to cross streets with lesser speed limits. High school children being much faster would be able to cross the streets and hence middle school children are more susceptible. Interaction with pedestrian sex shows that boys are more likely to be involved in crashes than girls in all the age groups.

6.2.2 Pedestrian Sex

Pedestrian sex appears in models 2 and 3 with interactions with site location and speed limit respectively. The odds multipliers with speed limits show that females are less likely to be involved in crashes in any of the speed limits. So also is the case with site Locations. Irrespective of the site locations the odds are higher for male pedestrians than females.

6.3 Other Characteristics

One of the other significant interactions noticed in the models was the lane group and median type interactions. The odds multipliers for lane group show that a divided highway with greater than two lanes is more likely to experience crashes than a divided highway with less than or equal to 2 lanes. More over, the odds for lane group for undivided highway with lanes less than or equal to 2 lanes is much higher than the undivided greater than two lanes category.

6.4 Conclusions

This thesis attempts to describe three different analysis levels- preliminary analysis, geo-spatial analysis and log-linear analysis of crashes involving pedestrians and bicyclists between the ages 4 and 18, in Orange County, Florida. The research investigated the effect of various driver, pedestrian, traffic, vehicle and geometric variables on crash occurrence.

In the preliminary analysis, various driver, pedestrian, vehicle and geometric variables were considered. Two-way contingency tables were used to look for the effects of one variable over the other. The preliminary study was done to establish the crash frequency of the pedestrian and bicyclists crashes in Orange County pertaining to a fixed age group and time period. From the preliminary analysis, using simple two-way tables and graphs, we can conclude that there are a significant number of pedestrian and bicyclist crashes in Orange County, related to school children. Of the three age groups, the children between the ages 4 to 11 seem to be at higher risk, as they share a larger crash population than the remaining two groups. That is, the number of crashes involving Elementary School Children is higher when compared to the Middle and High school children. Majority of the crashes took place on undivided two lane highways. Also most of the crashes took place in Primarily Residential areas where the posted speed limits are around 30 MPH. Most of the Crashes happened while the pedestrians were "crossing not at Intersection" and because of "Carelessness" of the Driver.

This information was useful in creating log-linear models. The results of the log-linear analysis have been described above. By building such models it would easier to estimate the effect of various variables over each other. Classifying the variables into three categories- driver, pedestrian and other made the task a little easier for the results could be analyzed in a better with lesser interaction terms. The effects of driver characteristics with other characteristics and pedestrian characteristics with other characteristics were elucidated. It was observed that middle aged drivers in particular were more likely to be involved in crashes involving pedestrians and bicyclists in age group 4 and 18. Also the effects of lanes and median type on the crash frequency have been discussed.

Finally, GIS was used to look closer at the crash distribution in Orange County. With the information about the total number of crashes and their locations, the schools and their locations, and also the Orange County base streets map, the spatial analysis was possible. Spatial analysis helps us understand a complex problem in a simpler way. Of the total 423 pedestrian and bicycle crashes that took place in Orange County in the specified time periods and age groups, 262 (nearly 62%) of the actual crash reports were found and obtained from the FDOT Mainframe website. The spatial analysis was based on these 262 crash reports. After a detailed inspection and review of the 262 crash reports, each crash was geo-coded onto the Orange County streets map. By creating buffers around different school locations (elementary, middle and high schools), the density of crashes were found for all schools together, and each kind of school separately. From the Spatial analysis, it can be concluded that middle schools are at a higher risk when compared to the other schools. Districts 1, 5 and 6 seem to have higher crashes per school ratio for crashes occurring within half a mile of the schools.

6.4.1 Suggested Recommendations

Based on the study, the following recommendations/suggestions have been made. Promoting road crossing education among school children is an important factor. The schools need to put in more effort in educating the children regarding traffic sense. It must be made mandatory for the children to cross the street only at the specified school crossing zones. The children should be encouraged to walk only on the sidewalks and not on the road. Also, it must be ensured by the school that there is a provision for sidewalk and crossing zones at sufficient gaps around the school zones which would caution the driver to slow down. Another issue that needs attention is to investigate the crash prone school zone areas that were concluded from the study. Teams of surveyors and engineers need to be sent out to the field to observe the scenarios around the mentioned school zones and even the streets where these crashes are taking place.

6.5 Limitations and Future Scope

This study was limited mainly because of the small sample size and the involvement of large number of variables. The more the sample size, the better it would have been to build efficient models. Also, some of the variables like Estimated Speed had lots of missing values, nearly 43%. This clearly had an impact on the model building, as speed ratio could not be used in all the models because of this reduced sample size. Hence alternate data collection sources should be investigated.

A possible extension to this study is to expand the data set to the entire state. Since the approach in this thesis in trying to pin point pedestrian and bicycle crashes in school going children during the commute hours has been elucidated for Orange County, similar study could be done at the state level, which would mean, a bigger sample size with almost the same variables. This would surely result in an efficient and better model fit. Also the spatial analysis mentioned in the thesis could be extended to various other places and the crash density around each school could be found out. This would help in making the schools safer for the school children. The only concern would be the availability of the base map and the details of the school locations.

APPENDIX A

ESTIMATES OF LOG-LINEAR MODELS

DRIVER MODELS

MODEL 1

Analysis of Maximum Likelihood Estimates					
Parameter		Estimate	Standard Error	Chi-Square	Pr > ChiSq
drivage	1	-0.5513	0.1198	21.18	<.0001
	2	0.362	0.0965	14.08	0.0002
	3	0.1893			
lanegroup	1	0.7472	0.2979	6.29	0.0121
	2	0.5579	0.2828	3.89	0.0485
	3	-1.3051			
median	1	-0.476	0.0768	38.45	<.0001
	2	0.476			
speedlimit	1	0.6098	0.2889	4.46	0.0348
	2	0.044	0.2991	0.02	0.883
	3	0.6026	0.1401	18.5	<.0001
	4	-1.2564			
drivage*lanegroup	1 1	0.2098	0.1391	2.28	0.1313
	1 2	-0.4414	0.1666	7.02	0.0081
	2 1	-0.3016	0.1139	7.01	0.0081
	2 2	0.2755	0.1264	4.75	0.0293
	13	0.2316			
	2 3	0.0261			
	3 1	0.0918			
	32	0.1659			
	3 3	-0.2577			
drivage*speedlimit	1 1	-0.03	0.1647	0.03	0.8557
	1 2	-0.2557	0.1603	2.55	0.1106
	1 3	0.0371	0.1397	0.07	0.7903
	1 4	0.2486			
	2 1	-0.2958	0.1345	4.84	0.0278
	2 2	0.2902	0.1261	5.29	0.0214
	2 3	0.0963	0.112	0.74	0.3897
	2 4	-0.0907			
	3 1	0.3258			
	3 2	-0.0345			
	3 3	-0.1334			
	3 4	-0.1579			
lanegroup*median	1 1	-0.5394	0.0963	31.39	<.0001
	2 1	0.6455	0.0903	51.1	<.0001
	3 1	-0.1061			
	1 2	0.5394			
	2 2	-0.6455			
	3 2	0.1061			

lanegroup*speedlimit	1 1	-0.6682	0.2895	5.33	0.021
	1 2	0.5254	0.3047	2.97	0.0847
	1 3	-0.085	0.1735	0.24	0.6242
	14	0.2278			
	2 1	-0.5144	0.3012	2.92	0.0876
	2 2	-1.0054	0.3487	8.31	0.0039
	2 3	0	0	0	0
	2 4	1.5198			
	3 1	1.1826			
	3 2	0.48			
	3 3	0.085			
	3 4	-1.7476			

MODEL 2

Analysis of Maximum Likelihood Estimates					
			Standard	Chi-	
Parameter		**Estimate**	**Error**	**Square**	**Pr > ChiSq**
drivage	1	-0.6397	0.2557	6.26	0.0123
	2	0.4184	0.1786	5.49	0.0191
	3	0.2213			
lanegroup	1	0.8371	0.3614	5.37	0.0205
	2	0.6312	0.3439	3.37	0.0664
	3	-1.4683			
median	1	-0.372	0.131	8.06	0.0045
	2	0.372			
speedlimit	1	0.2517	0.3142	0.64	0.423
	2	0.6155	0.1342	21.02	<.0001
	3	-0.8672			
drivage*lanegroup	1 1	0.2385	0.2641	0.82	0.3666
	1 2	-0.2673	0.2816	0.9	0.3426
	1 3	0.0288			
	2 1	-0.2114	0.1873	1.27	0.2591
	2 2	0.3244	0.1955	2.75	0.097
	2 3	-0.113			
	3 1	-0.0271			
	3 2	-0.0571			
	3 3	0.0842			
lanegroup*median	1 1	-0.6483	0.1458	19.76	<.0001
	1 2	0.6483			
	2 1	0.5317	0.142	14.03	0.0002
	2 2	-0.5317			

	3 1	0.1166			
	3 2	-0.1166			
lanegroup*speedlimit	1 1	0.3429	0.3219	1.14	0.2866
	1 2	-0.1309	0.1747	0.56	0.4538
	1 3	-0.212			
	2 1	-1.0608	0.365	8.45	0.0037
	2 2	0	0	0	0
	2 3	1.0608			
	3 1	0.7179			
	3 2	0.1309			
	3 3	-0.8488			

MODEL 3

Analysis of Maximum Likelihood Estimates

Parameter		**Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**
drivage	1	-0.4877	0.1133	18.53	<.0001
	2	0.4445	0.0876	25.78	<.0001
	3	0.0432			
lanegroup	1	-0.0282	0.3363	0.01	0.933
	2	0.0415	0.3564	0.01	0.9074
	3	-0.0133			
median	1	-0.3986	0.1956	4.15	0.0416
	2	0.3986			
spdratio	1	0.2819	0.304	0.86	0.3538
	2	-0.7866	0.2475	10.1	0.0015
	3	0.5047			
lanegroup*median	1 1	-0.6249	0.2122	8.67	0.0032
	2 1	0.5574	0.2046	7.42	0.0064
	3 1	0.0675			
	1 2	0.6249			
	2 2	-0.5574			
	2 3	-0.0675			
lanegroup*spdratio	1 1	0.8352	0.3129	7.13	0.0076
	1 2	0.6693	0.2948	5.16	0.0232
	1 3	-1.5045			
	2 1	0.8266	0.3539	5.46	0.0195
	2 2	0	.	.	.
	2 3	-0.8266			
	3 1	-1.6618			
	3 2	-0.6693			
	3 3	2.3311			

MODEL 4

Analysis of Maximum Likelihood Estimates					
Parameter		Estimate	Standard Error	Chi-Square	Pr > ChiSq
drivage	1	-0.4427	0.1133	15.27	<.0001
	2	0.5059	0.0915	30.59	<.0001
	3	-0.0632			
lanegroup	1	0.5207	0.1075	23.48	<.0001
	2	0.3108	0.1064	8.54	0.0035
	3	-0.8315			
median	1	-0.5056	0.08	39.99	<.0001
	2	0.5056			
dsex	1	0.2133	0.0541	15.52	<.0001
	2	-0.2133			
drivage*lanegroup	1 1	0.2425	0.1295	3.51	0.0611
	1 2	-0.2676	0.1546	3	0.0834
	2 1	-0.1247	0.1066	1.37	0.2423
	2 2	0.355	0.118	9.05	0.0026
	1 3	0.0251			
	2 3	-0.2303			
	3 1	-0.1178			
	3 2	-0.0874			
	3 3	0.2052			
lanegroup*median	1 1	-0.5341	0.1032	26.77	<.0001
	2 1	0.6135	0.0958	41	<.0001
	3 1	-0.0794			
	1 2	0.5341			
	2 2	-0.6135			
	3 2	0.0794			

MODEL 5

Analysis of Maximum Likelihood Estimates					
Parameter		Estimate	Standard Error	Chi-Square	Pr > ChiSq
drivage	1	-0.3395	0.088	14.9	0.0001
	2	0.5555	0.0717	59.99	<.0001
	3	-0.216			
aldriv	1	1.1286	0.1752	41.51	<.0001
	2	-1.1286			
ccd	1	0.8477	0.0771	120.85	<.0001
	2	0.0139	0.0889	0.02	0.8759

	3	-0.8616			

MODEL 6

Analysis of Maximum Likelihood Estimates					
Parameter		Estimate	Standard Error	Chi-Square	Pr > ChiSq
drivage	1	-0.3727	0.0873	18.24	<.0001
	2	0.5223	0.0709	54.33	<.0001
	3	-0.1496			
SL	1	0.5187	0.0901	33.11	<.0001
	2	0.674	0.0869	60.16	<.0001
	3	-0.2843	0.1143	6.19	0.0129
	4	-0.9084			
dsex	1	0.2133	0.0541	15.52	<.0001
	2	-0.2133			

MODEL 7

Analysis of Maximum Likelihood Estimates					
Parameter		Estimate	Standard Error	Chi-Square	Pr > ChiSq
drivage	1	-0.7505	0.1484	25.58	<.0001
	2	0.7212	0.0943	58.56	<.0001
	3	0.0293			
Vehtype	1	0.2837	0.0957	8.79	0.003
	2	-0.2837			
vehmov	1	0.4096	0.0658	38.76	<.0001
	2	-0.596	0.0845	49.71	<.0001
	3	0.1864			
aldriv	1	0.0264	0.0775	0.12	0.7338
	2	-0.0264			
drivage*Vehtype	1 1	0.6602	0.1484	19.79	<.0001
	2 1	-0.281	0.0943	8.89	0.0029
	3 1	-0.3792			
	1 2	-0.6602			
	2 2	0.281			
	3 2	0.3792			
Vehtype*aldriv	1 1	0.2845	0.0775	13.47	0.0002
	1 2	-0.2845			
	2 1	-0.2845			
	2 2	0.2845			

MODEL 8

Analysis of Maximum Likelihood Estimates					
Parameter		**Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**
drivage	1	-0.4066	0.1009	16.25	<.0001

	2	0.5088	0.0808	39.64	<.0001
	3	-0.1022			
speedlimit	1	0.1863	0.0861	4.68	0.0305
	2	0.3689	0.0828	19.85	<.0001
	3	-0.5552			
TC	1	0.6688	0.1201	31.03	<.0001
	2	0.8163	0.1174	48.37	<.0001
	3	-1.2696	0.2569	24.42	<.0001
	4	-0.2155			
dsex	1	0.2342	0.0618	14.36	0.0002
	2	-0.2342			

PEDESTRIAN MODELS

MODEL 1

Analysis of Maximum Likelihood Estimates					
Parameter		**Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**
pedage	1	0.1616	0.4383	0.14	0.7124
	2	0.0935	0.232	0.16	0.687
	3	-0.2551			
lanegroup	1	1.0171	0.4077	6.22	0.0126
	2	1.0491	0.4047	6.72	0.0095
	3	-2.0662			
median	1	-0.5343	0.19	7.91	0.0049
	2	0.5343			
spdratio	1	1.2571	0.1471	73.02	<.0001
	2	-0.1366	0.1727	0.63	0.4288
	3	-1.1205			

pedage*lanegroup	1 1	-0.2804	0.3761	0.56	0.456
	1 2	-0.7619	0.4131	3.4	0.0652
	1 3	1.0423			
	2 1	0.1555	0.1957	0.63	0.4269
	2 2	0	.	.	.
	2 3	-0.1555			
	3 1	0.1249			
	3 2	0.7619			
	3 3	-0.8868			
pedage*spdratio	1 1	0.1876	0.2511	0.56	0.4551
	1 2	0.7513	0.2724	7.61	0.0058
	1 3	-0.9389			
	2 1	0.0235	0.188	0.02	0.9004
	2 2	0.012	0.2309	0	0.9586
	2 3	-0.0355			
	3 1	-0.2111			
	3 2	-0.7633			
	3 3	0.9744			
lanegroup*median	1 1	-0.4941	0.2061	5.75	0.0165
	2 1	0.6748	0.1993	11.46	0.0007
	3 1	-0.1807			
	1 2	0.4941			
	2 2	-			

		0.6748			
	3 2	0.1807			

MODEL 2

Analysis of Maximum Likelihood Estimates					
Parameter		**Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**
pedage	1	0.1454	0.0888	2.68	0.1016
	2	-0.1694	0.1112	2.32	0.1277
	3	0.024			
SL	1	0.7887	0.0968	66.4	<.0001
	2	0.9733	0.0931	109	<.0001
	3	-0.5746	0.1215	22.4	<.0001
	4	-1.1874			
lanegroup	1	0.1365	0.0875	2.44	0.1186
	2	-0.3188	0.096	11	0.0009
	3	0.1823			
pedsex	1	0.1515	0.0626	5.85	0.0156
	2	-0.1515			
pedage*lanegroup	1 1	-0.0137	0.1022	0.02	0.8935
	1 2	-0.3078	0.1141	7.28	0.007
	1 3	0.3215			
	2 1	0.2904	0.1222	5.65	0.0175
	2 2	0.0686	0.131	0.27	0.6005
	2 3	-0.359			

	3 1	-0.2767			
	3 2	0.2392			
	3 3	0.0375			
SL*pedsex	1 1	0.2044	0.0863	5.62	0.0178
	2 1	0.0827	0.0821	1.01	0.3143
	3 1	0.1696	0.1123	2.28	0.1309
	4 1	-0.4567			
	1 2	-0.2044			
	2 2	-0.0827			
	3 2	-0.1696			
	4 2	0.4567			

MODEL 3

Analysis of Maximum Likelihood Estimates

Parameter		**Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**
pedage	1	-0.0757	0.093	0.66	0.4153
	2	0.00131	0.0897	0	0.9884
	3	0.07439			
speedlimit	1	0.1485	0.09	2.72	0.0991
	2	0.4827	0.0828	33.95	<.0001
	3	-0.6312			
pedsex	1	0.3173	0.0647	24.05	<.0001
	2	-0.3173			
pedage*speedlimit	1 1	0.3533	0.1174	9.06	0.0026
	1 2	-0.0383	0.1159	0.11	0.7409
	1 3	-0.315			

	2 1	0.00803	0.1177	0	0.9456
	2 2	0.1609	0.1106	2.12	0.1458
	2 3	-0.16893			
	3 1	-0.36133			
	3 2	-0.1226			
	3 3	0.48393			
speedlimit*pedsex	1 1	0.1204	0.0869	1.92	0.1659
	2 1	-0.2433	0.0802	9.2	0.0024
	3 1	0.1229			
	1 2	-0.1204			
	2 2	0.2433			
	3 2	-0.1229			

PARAMETERS OF COMPREHENSIVE MODEL

```
              Maximum Likelihood Analysis of Variance

Source                        DF   Chi-Square    Pr > ChiSq
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
drivage                        2         6.29        0.0431
pedage                         2         1.34        0.5130
drivage*pedage                 4         2.24        0.6914
lanegroup                      2         1.96        0.3754
drivage*lanegroup              4         3.92        0.4167
pedage*lanegroup               4        10.29        0.0359
Vehtype                        1         7.13        0.0076
drivage*Vehtype                2         1.37        0.5045
pedage*Vehtype                 2         0.18        0.9154
lanegroup*Vehtype              2         1.96        0.3758
median                         1         5.07        0.0243
drivage*median                 2         0.71        0.6999
pedage*median                  2         0.40        0.8205
lanegroup*median               2        12.29        0.0021
Vehtype*median                 1         0.03        0.8559
speedlimit                     3         7.39        0.0603
drivage*speedlimit             6         3.67        0.7214
pedage*speedlimit              6         5.55        0.4749
lanegroup*speedlimit           5*       16.56        0.0054
Vehtype*speedlimit             3         1.66        0.6454
```

median*speedlimit	3	1.61	0.6565
Likelihood Ratio	104	65.81	0.9987

REFERENCES

Ford, S H., Using Collision fatality rates to estimate pedestrian safety in school zones, *Presented at the Transportation Research Board Conference*, Washington D.C., January 2005.

Noland, R.B., and Quddus, M.A., An Analyses of Pedestrian and Bicycle Casualties Using Regional Panel Data, *Presented at the Transportation Research Board Conference*, Washington D.C., January 2004.

Hunter, W., Stutts, J., Pein, W., and Cox, C. Pedestrian and bicycle crash types of the early 1990's. U.S. Department of Transportation, Federal Highway Administration, *FHWA-RD-95-163*, 1996.

Harruff, R.C., Avery, A., and Alter-Pandya, A.S. Analysis of circumstances and injuries in 217 pedestrian traffic fatalities. *Accident Analysis and Prevention*, Vol. 30, No. 1, 1998, pp. 11-20.

Graham, Daniel J., and Stephen Glaister. Spatial variation in road pedestrian casualties: the role of urban scale, density and land use mix, Imperial College Centre for Transport Studies, working paper, 2002.

LaScala, E.A., Garber, D., and Gruenewald, P.J. Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. *Accident Analysis and Prevention*, Vol. 32, 2000, pp. 651-658.

Assailly, J.P. Characterization and prevention of child pedestrian accidents: an overview. *Journal of Applied Developmental Psychology*, Vol. 18, 1997, pp. 257- 262.

Derlet, R W; Silva, J, Jr; and Holcroft, J. Pedestrian accidents: adult and pediatric injuries. *The Journal of Emergency Medicine*, Vol 7, No. 1, 1989, pp. 5-8.

Kingma, J. Age and gender distributions of pedestrian accidents across the life span. *Perceptual and Motor Skills,* Vol 79, No. 3, 1994, pp. 1680-1682.

Tyrrell, R. A., Brooks, J. O., Wood, J.M., and Carberry, T.P., Nighttime Conspicuity From the Pedestrian's Perspective, Presented at the Transportation Research Board Conference, Washington D.C., January 2004.

Johnson, E., Geyer, J.A., Rai, N., and Ragland, D.R., Low Income Childhood Pedestrian Injury: Understanding the Disparate Risk, Presented at the Transportation Research Board Conference, Washington D.C., January 2005.

Stutts, J.C. and Hunter, W.W., *Police-reporting of pedestrians and bicyclists treated in hospital emergency rooms,* Presented at the 77th Annual Transportation Research Board Meeting, Washington D.C., 1998.

Michael R. Baltes, *Descriptive Analysis of Crashes Involving Pedestrians in Florida, 1990-1994*, Presented at the 77th Annual Transportation Research board Meeting, Washington D.C, 1998.

Michael R. Baltes, *A Study of Fatal Pedestrian Crashes in Florida*, Presented at the 77th Annual Transportation Research Board Meeting, Washington D.C, 1998.

Carlin, J.B., Taylor, P. and Nolan, T., A Case-Control Study Of Child Bicycle Injuries: Relationship of Risk to Exposure, Accident Analysis and Prevention, Vol.27, No.6, pp. 839-844, 1995.

Gruenewald, P.J., LaScala, E.A. and Johnson, F.W., An Ecological Study of the Locations of schools and Child Pedestrian Injury Collisions, Accident Analysis and Prevention, Vol 36, pp.569-576, 2004.

Johnston, J., Simpson, G. and Richardson, M., An Investigation of Road Crossing in a Virtual Environment, Accident Analysis and Prevention, Vol 35, pp.787-796, 2003.

Isler, R.B., Connelly, M.L., Conaglen, H.M. and Parsonson, B.S., Child Pedestrians' Crossing Gap Thresholds, Accident Analysis and Prevention, Vol 30, pp.443-453, 1998.

Abdel-Aty, M. and Lee, C., Comprehensive Analysis of Vehicle-Pedestrian Crashes at Intersections in Florida, Accident Analysis and Prevention, Vol 37, pp.775-786, 2005.

Fontaine, H. and Gourlet, Y., Fatal Pedestrian Accidents in France: A Typological Analysis, Accident Analysis and Prevention, Vol 29, pp.303-312, 1997.

Aultman-Hall, L., and Kaltenecker, M.G., Toronto Bicycle Commuter Safety Rates, Accident Analysis and Prevention, Vol 31, pp.675-686, 1999.

Abdel-Aty, M.A., Chen, C.L. and Schott, J.R., An Assessment of the Effect of Driver Age on Traffic Accident Involvement Using Log-Linear Models, Accident Analysis and Prevention, Vol 30, No. 6, pp.851-861, 1998.

Abdel-Aty, M.A., and Abdelwahab, H.T., Exploring the relationship between alcohol and the driver characteristics in motor vehicle accidents, Accident Analysis and Prevention, Vol 32, pp.473-482, 2000.

Preusser, D.F., Wells, J.K., Williams, A.F. and Weinstein, H.B., Pedestrian Crashes in Washington, DC and Baltimore, Accident Analysis and Prevention, Vol 34, pp.703-710, 2002.

Everitt, B.S. The Analysis of Contingency Tables, Published by Chapman & Hall, London, 1991.

Distracted Driver Report, Department of Highway safety and Motor Vehicles, Office of Management Research and Development, Tallahassee, Fl, December 2004.

Traffic Crash Statistics Report, Department of Highway safety and Motor Vehicles, Office of Management Research and Development, Tallahassee, Fl, December 2004.