# Geographic Information Systems for Spatial Analysis of Traffic Collision Locations in La Crosse, Wisconsin

Cory A. Brose
*Saint Mary's University of Minnesota, Graduate Studies in Resource Analysis, 700 Terrace Heights #10; Winona, Minnesota, 55987, USA*

*Key words:* Geographic Information System, traffic collisions, Structured Query Language, spatial analysis, Topologically Integrated Geographic Encoding Referencing files, geocode

## Abstract

Spatial distributions and densities of traffic collisions were defined through utilization of a Geographic Information System. Traffic collision data for La Crosse, Wisconsin were acquired from the Wisconsin Department of Transportation. Database and spreadsheet programs were used to edit and standardize the traffic collision data to index with 1995 Topologically Integrated Geographic Encoding and Referencing (TIGER) system files. Structured Query Language, an integrated functionality of the Geographic Information System, was used as the primary tool to initiate both spatial and statistical analyses. Traffic collision densities and trends, with respect to various road conditions, intersection control, driver circumstances, etc. were displayed as visual computer images. Statistical analyses, charts, and graphs were used to supplement the study.

## Introduction

An understanding of traffic collision patterns or trends is necessary when implementing efforts to improve traffic safety. Locational data, contributing circumstances, and collision rates are only a few of the essential elements required for planning or integrating preventative methods for decreasing traffic collisions. Other helpful features to consider may be locations of businesses, weather conditions or road conditions, and their correlation to collisions. An excellent system for saving and managing these types of data is a Geographic Information System (GIS). There is almost no limit as to how much data can be incorporated into a GIS. Banks, shops, restaurants and grocery stores are only a few of the features that can be built into the GIS. Outlines of residential, commercial and industrial areas could be added to the GIS as well. Demographic data such as age, education and income could also be integrated. The inclusion of these variables, and more, would add to the robustness of the GIS. Consequently, GIS not only allows for input, maintenance, and output of robust data sets, but has the capability for modeling spatial distributions and trends of these data sets. When factoring in this powerful capability, GIS appears to be a solution for numerous problems. One recent study noted that GIS makes possible things that were impossible or impractical before (Trend 1997). Considering the extent of spatial data associated with traffic collision locations, GIS seems to be a logical choice in defining spatial trends or patterns. The capabilities of GIS in the transportation field will permit assimilation, integration, and presentation of data collected and stored (Vonderohe *et al.* 1992).

Subsequently, it was apparent the traffic collision study would require a system that would promote geocoding of spatial data. Geocoding is a function that allows locations such as addresses or intersections to be indexed to a digital street map. As a result, ArcView GIS was chosen due to its straight forward Graphical User Interface (GUI), powerful spatial analysis capabilities and more specifically for its inherent capacity to geocode locational data.

Foundation of the study was supported by three phases including data acquisition, geocoding of data, and data analysis. Data acquisition encompassed two variables. The first, attaining an accurate basemap of the study area, and second, acquiring the actual collision data. Geocoding of the collision data was expected to be the most labor intensive of the phases and also the most unique as no definite procedures or methodologies for geocoding of street intersections existed. The data analysis would follow the geocoding process. Geocoded themes would be analyzed using the spatial and statistical tools built into ArcView GIS.

## Methods

### *Data Acquisition*

The initial step in conducting the study was formation of a basemap on which graphical elements and features would be displayed. Topologically Integrated Geographic Encoding Referencing (TIGER) system files from 1995 were chosen as the basemap. The 1995 TIGER files, which are the product of a joint effort by the Federal Bureau of Census and the United States Geological Survey, were available for download on the Environmental Systems Research Institute's (ESRI) web site. The TIGER files included several data layers for the area (Figure 1) necessary to deploy the study including La Crosse county streets, and minor civil division boundaries. Topology or intelligence is present in TIGER street files. In other words each line or arc in a TIGER street file is coded with a left and right side numbering system which allows for indexing of street addresses or street intersection data.

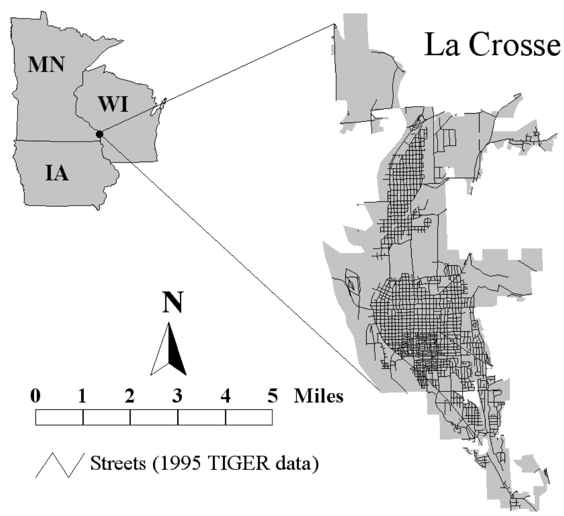


Figure 1. Location of study area.

After searching several resources for traffic collision data, the La Crosse regional office for the Wisconsin Department of Transportation (WiDOT) provided the most valuable information. They could attain traffic collision records from the Statistical Analysis System (SAS) in Madison, Wisconsin. Multitudes of traffic collision data exist on the SAS mainframe system. Data pertaining to weather, road and light conditions, driver circumstances, pedestrian involvement, manner of collision, accident location etc. were only a few that resided on the SAS.

Origination of data occurs in the event of a traffic collision at which a La Crosse police officer fills out a four page Wisconsin Motor Vehicle Accident Report. Completed accident reports are delivered to the WiDOT headquarters in Madison where they are keyed or scanned into the SAS. Unfortunately, SAS files are not in database format but rather the records are flat files.

"Typically, a DOT's highway inventory consists of a diverse collection of mainframe files that are difficult (or nearly impossible) to integrate. In many cases the data exist in flat files and are not incorporated under a database management system. These circumstances not only prohibit comprehensive GIS - Transportation application development, but they also limit the efficiency and effectiveness of an agency's overall information systems efforts" (Vonderohe *et al*. 1992).

Nevertheless, with no alternative data resources, the SAS files were obtained for 1994 through June of 1998. Accompanying the electronic files were several hardcopies from the SAS containing descriptive information about the flat files, i.e. character spacing and numbering with respect to organization of the different series of tabular information.

### *Data Format and Standardization*

Various problems were apparent upon examination of the traffic collision files extracted from the SAS. The first problem was format of the files. Having no fields, records, columns or rows, they were simply large series of numerical and textual data with little structure or uniformity. Relational technology inherent in ArcView GIS relies heavily on the premise that the tabular information it relates to is in a *relational* database format with rigid structure and organization of fields and records. Initial impetus of the study existed in transforming the flat files into rows and columns of traffic collision data.

An import utility in the Microsoft Excel97 spreadsheet program was utilized to begin the transformation process. First, the utility recognized the flat files as being in a fixed width file format. Recognition of fixed width was a crucial moment in the study as it yielded the opportunity to insert column breaks into the flat files. The hardcopy information provided by the WiDOT explained how many characters SAS used for the respective fields. This information was referenced to discern exact locations for inserting column breaks. With columns generated, the file was saved in the Excel spreadsheet (xls) format.

In addition to format changes, this study also necessitated that the traffic collision locations be recorded with respect to their intersections. Geocoding preferences in ArcView GIS require that locations be denoted in a very particular scheme. For example, a collision which occurred at the intersection of Main Street and State Street must exist in a single database cell as "Main St & State St" to index with the ArcView GIS geocoding system. However, street intersections for collision locations in the spreadsheet files are described using two fields, "on street" and "at street" and therefore need to be merged into one coherent cell.

To solve this problem a new field was created between the "on street" and "at street" fields. Subsequently, the "&" symbol was entered into all cells in this new field using Excel's paste function. During the editing process it was discovered that many of the streets were misspelled and their suffixes were incorrect with reference to the TIGER street file. So, a new "suffix" field was created to the right of the "on street" and "at street" fields. The new fields were added to contain the suffixes for their respective "on" or "at" street, i.e. Ave., Blvd., Pl., St., Ter., etc. The spelling and suffix issue was further resolved through opening the TIGER street file and tiling it in the spreadsheet view with the traffic collision file. The "street name" field in the TIGER file was sorted alphabetically using the data sort function in Excel. The same procedure was applied to the "on street" field in the traffic collision file. At that point, street names and suffixes were copied simultaneously from the TIGER file and pasted to the coordinating "on street" and "suffix" fields in the traffic collision file. The same procedure was conducted to solve spelling and suffix issues for the "at street" field.

At this point the "on street", "at street", "&", and "suffix" cells were ready for merging. Merging was accomplished by first saving the spreadsheet file as a comma separated values (csv) file. Next, the csv file was opened in Microsoft WordPad. Successfully removing the desired commas without distorting the traffic collision data was accomplished by using the edit and replace function of WordPad. After commas were removed and replaced with spaces the file was saved and reopened in the Excel spreadsheet program. With spelling errors corrected, suffixes added and fields merged, the traffic collision data were saved in database format (dbf) and therefore ready for geocoding in ArcView GIS.

### *Geocoding*

Geocoding in ArcView GIS is a process that creates a layer of visual information based on locational data in tabular form and a reference feature theme. The traffic collision .dbf files were added to the GIS for each year 1994 through 1998 and geocoded to the TIGER street theme. Although traffic collision data had been modified and standardized in the aforementioned methods there were still several problems that interfered with the geocoding process.

The first problem encountered was a large number of intersections simply not geocoding to the TIGER street theme. Analysis of the situation proved that in numerous cases discrepancies still existed in the nomenclature of the traffic collision street names. For example, US Hwy 14, South Avenue and Mormon Coulee Road are all names that may possibly represent the same arc or line in the GIS. In such a case the geocoding failures were reviewed and the names were modified or changed to accommodate for the discrepancy.

In addition to discrepancies in nomenclature, there was another problem that hindered the geocoding process. Several arcs in the TIGER file had an empty street name field. The La Crosse Street map in the November 1998 CenturyTel phonebook proved to be a sufficient reference for updating the street name field of the TIGER file.

Another problem encountered in the geocoding process was the tendency for the GIS to create feature points for traffic collision locations in Onalaska, a municipality North of La Crosse. The explanation was that several intersections in Onalaska had similar names to those in La Crosse, for example, 2nd St and Pearl St. This problem was alleviated with a "theme on theme" selection.

***Spatial and Statistical Methods***

Spatial analysis capability is the hallmark characteristic of GIS. Being able to overlay multiple themes or layers of data and discern various spatial relationships among them is a technique widely applicable in government, engineering, natural resources, and utilities. This spatial functionality also proved to work well for analyses in this traffic collision study.

Previously mentioned was the sharing of the same name for several intersections in both La Crosse and neighboring Onalaska. Using a "theme on theme selection" traversed this obstacle. By selecting the arcs from the La Crosse County street theme that intersected the polygon representing the City of La Crosse, a new theme was produced containing streets for only the City of La Crosse. Refining of the La Crosse street file ensured a more accurate representation of the geocoded collision data.

Several questions were asked of the GIS in an attempt to estimate collision trends such as high collision areas, differences among intersection controls and weather conditions. High collision areas could be depicted by calculating point densities from the geocoded themes through use of the Spatial Analyst extension for ArcView GIS. Point density calculations were created as ArcView recognized the number of point features within a predefined radius. This density was displayed as a new theme. Density themes yielded a smooth output giving the user an estimation of intense, high or moderate collision areas (Figure 2).

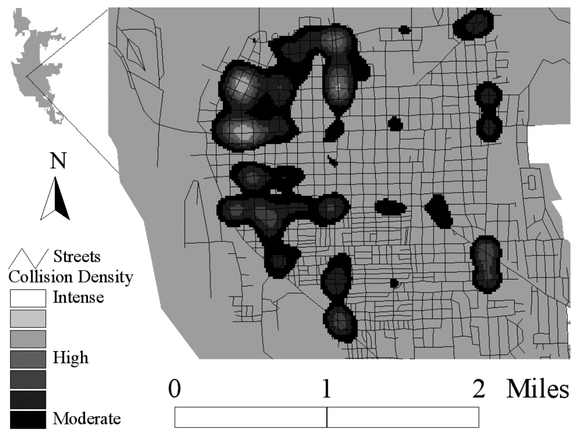


Figure 2. Collision densities in a portion of south La Crosse.

Furthermore, spatial relationships of traffic collisions with respect to various businesses, weather conditions, intersection controls or driver contributing circumstances could be displayed. For example, several traffic collisions were recorded as being caused due to driver condition as the contributing circumstance. No implication existed as to whether or not alcohol or drugs were involved. Speculating that alcohol may have been involved, a new .dbf file was constructed that contained all La Crosse taverns listed in the November 1998 CenturyTel phonebook. The .dbf file was geocoded to the La Crosse street theme and displayed as a new theme in the view. Using Structured Query Language (SQL), the expression ( [At_st_driv] = "Driver condition" ) was used for selection of all accidents caused by driver condition from the traffic collision database file. The selected records appeared in the view highlighted, as yellow points. These points were used to create a new data layer consisting of accidents caused by driver condition.

Intersection control was another aspect of the study. The question was asked, "what intersections produce higher collision averages? No control or traffic signal control?" To answer this question, similar SQL techniques were used to query the data for locations of intersections with different types of control, for example traffic signals, no control etc. Unfortunately, many intersections were recorded as having both traffic signals and no control. Captain Tom Jacobs, (La Crosse police department - Administrative Services), stated intersections should never be recorded as having more than one type of control i.e. traffic signals and no control. He further added that records indicating this type of situation were incorrect. For this reason no spatial analyses were conducted with respect to traffic signal verses no control, though the study did include descriptive and comparative statistics between intersections with four-way stop control and intersections with no control.

A situation was hypothesized that the state of Wisconsin was allocating funds to the City of La Crosse for installation of some new street lighting equipment. The new equipment would be installed at intersections that needed it most. So, ArcView GIS was used to locate intersections that consistently produced nighttime collisions. Evidence is mounting that municipalities building a geographic information system (GIS) to organize their public works data more efficiently have advantages over other communities trying to win funding approval for infrastructure programs (Cowden 1990). The goal was to find areas at which traffic collisions have occurred every year for the past five years. Light condition data are recorded in the event of a traffic collision and consequently were included in the SAS data set. All traffic collisions that occurred during night conditions without street lighting were selected from the accident files. From these selected point locations, density themes were created for each of the five years. Next, a map query was used to separate out the areas of commonality, among the five density themes. The final result was a map (Figure 3) showing where traffic collisions have occurred every year for the past five years at night without streetlights.

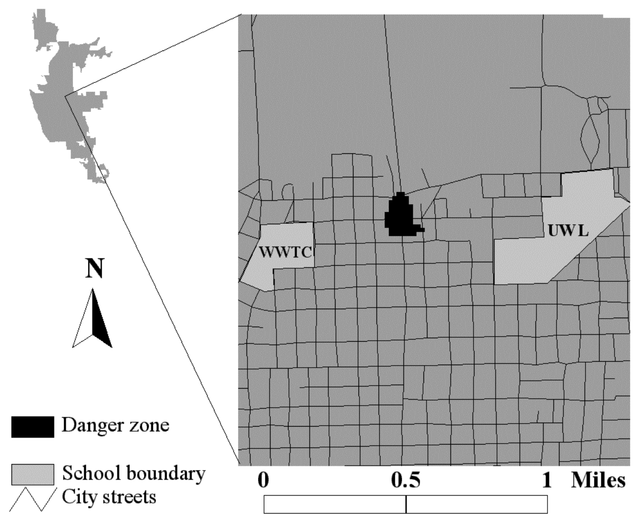


Figure 3. Night collisions, with no street lighting, occur most often in college residential areas.

As pedestrian safety is of tremendous importance and synchronizes with traffic safety, a similar type of commonality query was conducted for traffic collisions that involved pedestrians over the past five years. Databases were created and geocoded to produce data layers containing different businesses and organizations. To investigate if spatial trends or relationships existed among the data, records were queried, point densities calculated and data themes were overlaid for spatial analysis.

Visualization was an excellent learning technique but numerical data were also necessary to support findings of the study. The SQL functionality of ArcView GIS made easy the selection and summation of various data fields. For example, the study appeared to show that more accidents occurred during dry road conditions than any other road condition. ArcView GIS allowed for easy summation of the "road condition" field by counting the number of traffic collisions for each road condition, i.e. dry, wet, snowy, etc. ArcView GIS then summarized numbers of traffic collisions to their respective road conditions into a new database. Several of these databases were created for different variables including light conditions and manner of collisions. From these databases, various bar and pie graphs were made.

In addition to various graphs, statistical comparisons were performed on two different portions of the study. The first comparison was between collisions that occurred during dry and snowy road conditions. Structured Query Language was used to select out the desired records. Then the road condition field was summed with respect to the traffic collision date. These summations were converted into new databases that included all collisions that occurred on their respective day. Using the field statistics option in ArcView GIS, a daily mean for each road condition, i.e. snowy or dry was computed. Ultimately, the means for the years 1994 through 1998 were statistically compared to show which conditions produced higher accident rates. The second comparison was between accident rates for different intersection controls, i.e. no control vs. four-way stop control. Statistical analyses were computed with reference to Biostatistical Analysis (Zar 1996), and then checked using the data analysis option in Microsoft Excel97.

## Results

Further investigation of traffic collisions caused by driver condition revealed that most were sited within a quarter mile of taverns. Theme on theme selection was deployed to ask which accidents were within a quarter mile of any tavern. The points, which appeared in the view, seemed to be aggregated in a tight pattern around the tavern point features. The density calculation of accidents caused by driver condition also seemed to support that the collisions may have been alcohol related. Another theme on theme selection revealed that over seventy percent of the collisions caused by driver condition were within one-quarter mile of a tavern. A quarter mile buffer region around the taverns (Figure 4) shows the spatial relationship among the collisions and taverns.

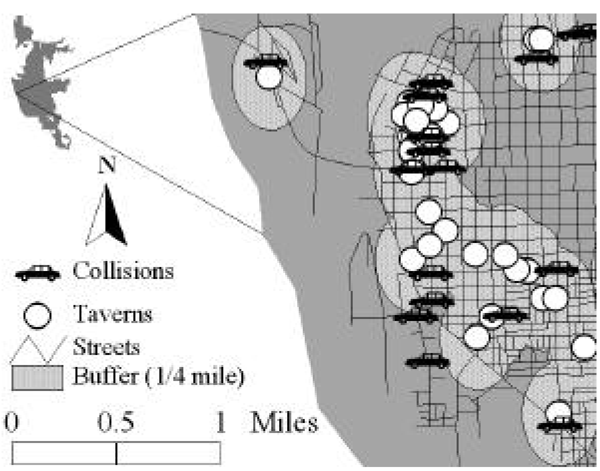


Figure 4. Most “driver condition” collisions are less than a few city blocks from a tavern.

Statistical analyses for comparing traffic collision rates between snowy and dry road conditions yielded an interesting outcome (Table 1). The null hypothesis ($H_0$: snowy road condition accident rates <= dry road condition accident rates) was not rejected as analysis revealed that accident rates in the City of La Crosse did not statistically increase during snowy road conditions. La Crosse drivers apparently adapt to snowy road conditions and adjust their driving habits accordingly.

Table 1. Snowy roads verse dry roads.

| | Snowy | Dry |
|---|---|---|
| Collisions/Day | 3.640 | 3.400 |
| Variance | 0.780 | 0.060 |
| Observations (years) | 5 | 5 |
| Degrees freedom | 5 | |
| $t_{calculated}$ | 0.600 | |
| Probability of error | 29.0% | |
| $T_{0.05,(1),5}$ [table value] | 2.020 | |

Results of the comparison between different intersection controls revealed a surprising outcome. Previous speculation was that uncontrolled intersections were not as dangerous as other types of intersections due to their locations in residential areas having lower traffic flow or density. But, analysis showed with a very high level of confidence that traffic collision rates were significantly higher at uncontrolled intersections than intersections with four-way stop control ($P<0.01$).

Graphs were created to show collision distributions for each of the five years. A very strong consistency across the years existed. Collision percentages for each year were either the same or differed by negligible amounts. For example, seventy two percent of traffic collisions occur during dry road conditions - consistent throughout 1994 to 1998. Likewise, the various types of collisions i.e. angle, rear end etc. also occurred in the same proportions over all years analyzed (Figure 5).

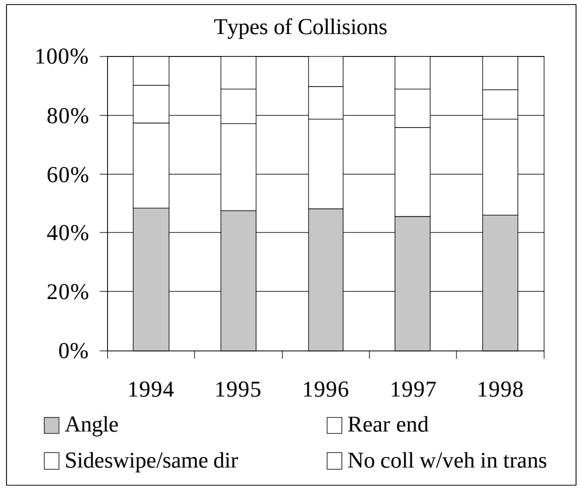


Figure 5. Collision types occur in consistent proportions.

Collision densities were also very consistent for 1994 through 1998. The visual images produced with GIS showed patterns that were nearly identical for each year. The intersections of 3rd St & Cass St, Gillette St & State Hwy 16 and US Hwy 53 & State Hwy 16 appeared to produce the highest density each year. This was believed to be due to the fact that each of these intersections is near a major corridor of travel. For example, the intersection of 3rd St and Cass St is at the base of the bridge that leads into Minnesota. It is the point of entry and exit for multitudes of commuters every day. Likewise, the other two intersections are passageways for drivers commuting to and from Onalaska or North La Crosse. The corridors appear to act as funnels in the traffic system. A strong relationship also appeared to exist between high collision areas and gas station locations (Figure 6).

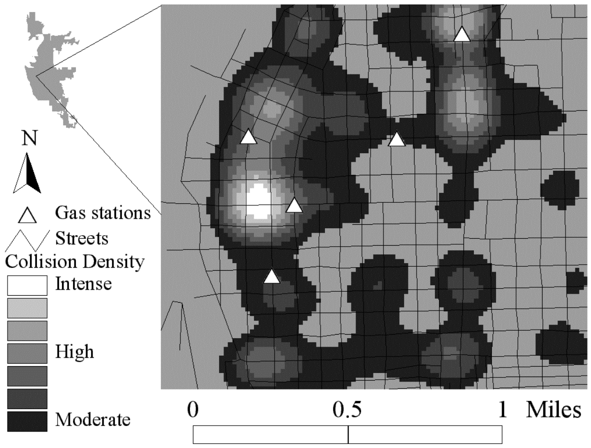


Figure 6. Traffic congestion near gas stations may contribute to high collision rates.

Per the hypothetical situation in which a new street light was to be installed, the commonality queries showed one specific location between 12th St and West Ave along Badger St. representing areas that consistently produced traffic collisions at night with no street lighting. The University of Wisconsin - La Crosse, and Western Wisconsin Technical College exist to the east and west of this area, respectively. Close proximity of these two, large educational institutions drew speculation that the area in question may be a college housing area. Large numbers of students coupled with factors such as night courses and nightlife may necessitate a new streetlight for additional safety.

Pedestrian safety coincides with traffic safety. There were several areas yielding traffic collisions with pedestrian involvement over the past five years. One of these areas is located directly in La Crosse's downtown. With stores, businesses, banking, taverns etc, available in the downtown area it was no surprise that traffic collisions involving pedestrians had occurred there. Another area that produced pedestrian involvement was the intersection of Losey Blvd and State St. Several circumstances may have contributed to this situation, such as banks, fast food, restaurants, shopping and groceries are available near this intersection. Another area between Mississippi St and Jackson St along West Ave showed pedestrian involvement with traffic collisions. Spatial analysis did not suggest that gas stations, grocery stores or banks contributed to the accidents, instead they appeared to correlate with several taverns two or three blocks to the southwest and, St Francis Hospital and clinic to the north. These were the main factors believed to contribute to higher concentrations of pedestrians and pedestrian involvement with collisions in the area (Figure 7).

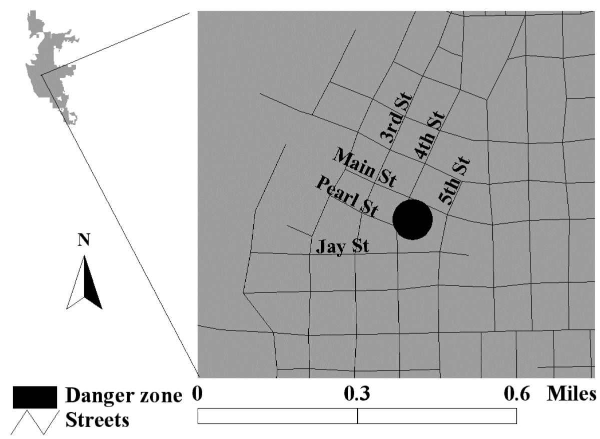


Figure 7. Collisions involving pedestrians during each of the past five years.

## Discussion

Purity of traffic collision data is extremely important when conducting a traffic analysis study, especially when the study involves GIS and procedures such as geocoding. The traffic collision records used in this study showed several instances where error could occur. As with most computer technology, input data should contain as few errors as possible in order to achieve the full potential of the programs analyzing the data. Initially, the geocoding process was rather unsuccessful as forty percent of the traffic collision incidents did not index with the TIGER street file. In efforts to improve accuracy of results for spatial and statistical analyses, editing and standardization processes were repeated until approximately eighty-percent of the traffic collision records geocoded with the TIGER street file. Approximately ten-percent of recorded traffic collisions could not be studied due to their occurrence in parking lots, private properties or alleys, which have no standard nomenclature or specified correlation with street intersections. The remaining ten percent of the traffic collision records had errors beyond repair.

Do to the extensive amounts of time and labor used to correct WiDOT data impurity, a customized program was written to demonstrate the functionality of a standardized data entry system. Avenue, the object-oriented program language running in ArcView, was used to create the data entry system. Through Avenue, several dialogue boxes were created that contained predefined lists of variables, i.e. intersections, weather conditions, road conditions, etc. As a result, manual key entry of records would be replaced by selection of variables from the customized, error-free, pull down menus. For example, the different types of collisions were integrated into the code (Table 2). When a variable such as “head on collision” is chosen, Avenue will find the respective field name, which happens to be “Mnrcoll”, and insert “head on” into a new record in the Mnrcoll field. The code would not only eliminate error, but time as well. Choosing several variables from pulldown menus is more efficient than keying them in by hand.

Table 2. Sample of computer code for prevention of data entry errors into ArcView GIS

```
MnrcollField=theVTab.FindField(“Mnrcoll”)
NewMnrcoll=MsgBox.ChoiceAsString({
“No coll w/veh in trans”,
“Rear end”,
“Head on”,
“Rear to rear”,
“Angle”,
“Sideswipe/same dir”,
“Sideswip/opposite dir”,
“Unknown”,
},”...choose the type of collision”, “Manner
of collision:”)
theVTab.SetValue(MnrcollField,rec,newMn
rcoll)
theVTab.Refresh
```

A similar type of code was used for road condition, weather condition etc. In each case the particular variables were integrated into the code and visible in their respective "pulldown" menus. This type of data entry system would be strongly recommended in the implementation of GIS for traffic collision analysis, as error prevention is crucial in achieving accurate results.

Another recommendation for a traffic collision analysis system would be the creation of a more accurate basemap. Although the 1995 TIGER files proved sufficient for this study, it is not feasible that a city or county agency would successfully implement GIS for traffic analysis without involvement with other divisions of government. When considering budgets, funds would be better spent on one accurate basemap for all county or city agencies to utilize rather than several different basemaps. For example, city assessors and surveyors often use streets or their centerlines as benchmarks for conducting their work. Interchanges and curves in the TIGER files appeared sharp blocky or "kinked" and deviated from the actual real world geometry of their respective intersections by feet and even yards in some cases. Unfortunately, TIGER files are by no means accurate enough for land assessment purposes or for measurements for deed descriptions. While conducting a project much like this study, the Delaware Department of Transportation arrived at a similar conclusion.

"TIGER was nationally based, the files were amazingly complete, and represented a quantum leap forward for GIS, but for transportation and road alignment, TIGER files left a lot to the imagination, especially at the local level" (Tufo 1998).

Computer Assisted Drafting (CAD) systems in the La Crosse City Engineering department should have high accuracy detailed line work representing city streets. These CAD street files could be imported into the GIS as one alternative in creation of an accurate basemap. Another option may be to scan or digitize hard copy street blueprints into the GIS using a high resolution, large format, scanning device or automated digitizing table. The ultimate goal would be a basemap that includes all streets in La Crosse, open architecture for adding new streets in the future, and enough accuracy for all divisions of government involved.

Analysis in this study served in identifying and displaying various trends and relationships of traffic collisions in the City of La Crosse. The study also supported the theory that GIS may serve as an adequate decision support tool in the event of a specific need of the city or municipality. The trends and patterns defined in this study through the use of a Geographic Information System may be used to maximize efficiency of efforts to improve traffic and pedestrian safety.

**References**


Cowden, R. 1990. Using a GIS to Win Funding. *American City & County* 8: 56

Trend. 1997. County, cities collaborate on use of GIS technology. *American City & County* 6: 18

Tufo, M. 1998. Capital Improvement Program (CIP) Sample. *Geo Info Systems* 5: 27

Vonderohe, A., Travis, L., and Smith, R. 1992. Implementing GIS at State DOTs, Part One. *Public Works* 2: 52

Vonderohe, A., Travis, L., and Smith, R. 1992. Implementing GIS at State DOTs, Part Two. *Public Works* 3: 38

Zar, J.H. 1996. Biostatistical Analysis, pp. 127. In J.H. Zar, *Testing for Difference Between two Means*. Prentice Hall, New Jersey.