Analysis of Texas Fatal Traffic Crashes: Progress Report

Project Manager: Mark Poitras
Assistant Manager: Paul Head
GIS Analysts:
Frank Martin
Catherine Perille
Larissa Matin

Prepared By: T.A.G.S
October 26, 2009

Date: October 26, 2009
To: Marcia Becker, MPH, CHES, PMP
From: T.A.G.S. – Traffic Accident Geospatial Solutions
Subject: Progress of Analysis of Texas Traffic Crash Data

Analysis of Texas Fatal Traffic Crashes

The following progress report provides an overview of the status of the requested analysis of Texas fatal traffic crashes, a project focusing specifically on examining particular driver age groups and the contributing factors leading to the crashes. This document will provide a status update on the work completed, the work currently in progress, and the work yet to be completed. In addition, this document provides for an overall understanding of the challenges T.A.G.S has encountered since the project's inception and for a clear layout as to how T.A.G.S plans to complete the project.

Project Description
T.A.G.S has been performing an analysis on the number of Texas fatal traffic crashes using the National Highway Traffic Safety Administration (NHTSA) Fatality Analysis Reporting System (FARS) database. This analysis is being done to identify any significant correlation between the Texas counties which have the highest number of traffic fatalities and specific contributing factors leading to the crashes such as:

- Age groups
- Road conditions
- Driver factor scenarios
- Roadway types
- Roadway status (urban versus rural)

This issue is of high importance because Texas has exceeded the national average for traffic fatalities over a four year period between 2004 and 2007. A better understanding through spatial analysis of what factors are involved in fatal crashes around the state will hopefully lead to the reduction of annual traffic fatalities, the identification of effective countermeasures for roadway deficiencies, and a reduction in the dollar amount spent annually on traffic fatality related expenses.

Project Scope
We have decided to concentrate our time and efforts to viewing the following problems that were discussed in your request for proposal at the county level:

70+ age group/ 70- age group/ 15-24 age group:

- Counties with the highest number of fatal crashes
- Urban vs. Rural breakdown of fatal crashes
- Roadway type breakdown
- Roadway condition breakdown
- Top 6 driver factors related to the fatal crash occurrences for each age group

Work Completed

Summary

The following describes what T.A.G.S has completed since the start of the project. After the proposal presentation, T.A.G.S immediately began exploring the online FARS data and query system hosted by the NHTSA. After learning how to use the system, we began to query and collect the data relevant to the project. T.A.G.S did run into some challenges during the data collection process. Examples of these challenges involving how to query and interpret the FARS data are discussed in the challenges section below. An important part of the study is the requested rate calculation needed to normalize the frequency data for the various factors related to fatal traffic crashes. This report shows the method by which the rates were calculated along with some examples. In addition, we have provided a set of sample maps that represent a similar format and layout that will be used for the final report. In addition to the data collection process, T.A.G.S has performed quality assurance on the data collected to ensure the data is accurate. A literature review of related material to the project has also been done in order to provide some answers to questions about the data in the FARS and to provide direction in the project. Below is a more detailed description of the work completed:

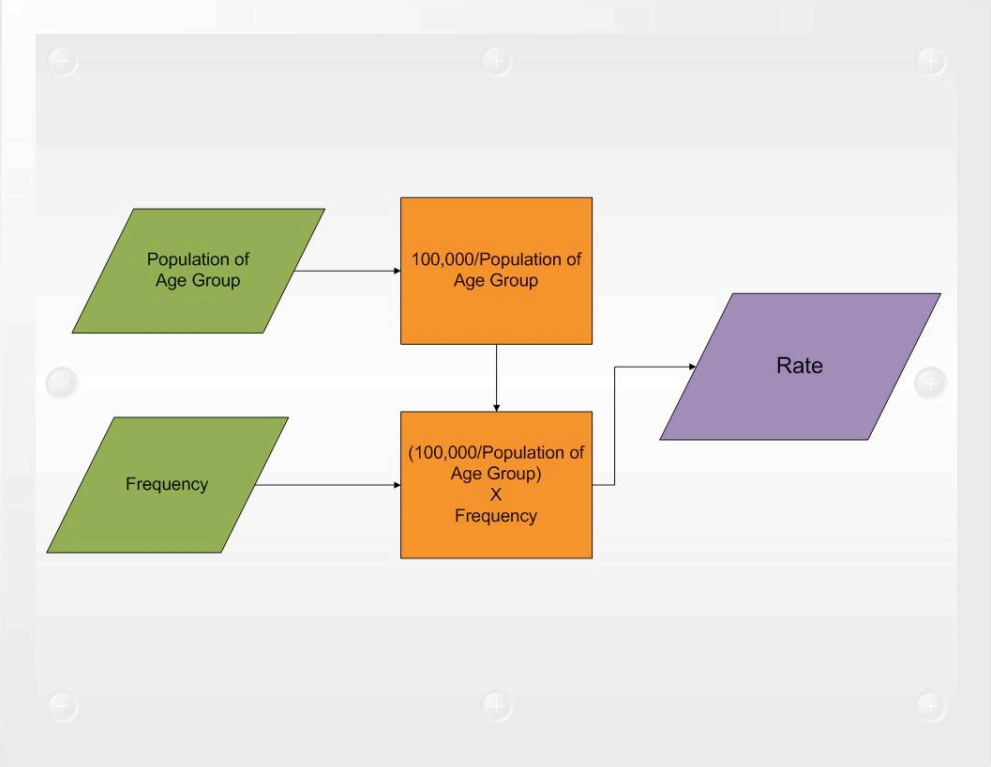
Query Process and Data Collection

- Access the FARS Encyclopedia website and select the query tab
- Select the year and "Submit"
- Select the data field(s) to query and "Submit"
- Select condition criteria for each data field and select the type of table you want to create
- Format the report by choosing the data fields to be displayed in the columns and rows and "Submit"

The resulting data table can then be exported into Excel where it can be manipulated and compiled with other data tables to create complete spreadsheet of related data to export into ArcGIS for map creation. The FARS queries needed to obtain the data for this project have all been executed and verified. The resulting data has been organized in databases created by T.A.G.S and will be the foundation for the project. The supplementary data such as population data and spatial data to be used in conjunction with the FARS data for analysis has also been collected and organized.

Rate Calculations

The rates requested in the Request for Proposal document have been calculated for all frequencies and factors of the FARS data. Below shows a flow chart on how the rates were calculated:



Below shows an example of a very small part of the data T.A.G.S has collected, organized and performed calculations:

				Driver Factors Frequencies						Driver Factor Rates per 100,000					
County	Population	Fatal Crashes	Rate (Fatal Crashes per 100,000)	DF (5)	DF (6)	DF (28)	DF (44)	DF (58)	DF (43)	DF (5)	DF (6)	DF (28)	DF (44)	DF (58)	DF (43)
Anderson	7,072	8	113.12	1	0	5	2	0	0	14.14	0	70.70	28.28	0	0
Andrews	2,113	2	94.65	0	1	0	1	2	0	0.00	47.33	0.00	47.33	94.65	0.00
Angelina	11,131	2	17.97	1	0	2	1	0	0	8.98	0.00	17.97	8.98	0.00	0.00
Aransas	2,809	1	35.60	1	0	0	1	0	0	35.60	0.00	0.00	35.60	0.00	0.00
Archer	1,303	0	0.00	0	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00
Armstrong	273	0	0.00	0	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00
Atascosa	6,427	1	15.56	0	0	1	0	1	0	0.00	0.00	15.56	0.00	15.56	0.00
Austin	3,574	2	55.96	0	0	1	0	0	0	0.00	0.00	27.98	0.00	0.00	0.00
Bailey	845	0	0.00	0	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00
Bandera	2,446	2	81.77	1	1	1	0	1	0	40.88	40.88	40.88	0.00	40.88	0.00
Bastrop	9,327	7	75.05	3	4	2	2	3	0	32.16	42.89	21.44	21.44	32.16	0.00
Baylor	434	0	0.00	0	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00
Bee	5,820	1	17.18	0	0	0	1	0	0	0.00	0.00	0.00	17.18	0.00	0.00
Bell	39,511	25	63.27	12	2	10	7	1	1	30.37	5.06	25.31	17.72	2.53	2.53
Bexar	242,371	53	21.87	20	16	16	22	5	0	8.25	6.60	6.60	9.08	2.06	0.00

Quality Assurance

To perform quality assurance on the data sets that were collected, a spreadsheet was created listing all the FARS data set queries, the individual assigned to complete each query, and a validation column for each query. As each query was completed by the assigned individual, a validation of the query and the data set created was done by a different member of the team. Quality assurance for all the FARS data collected has been completed and is shown below.

70+ Age Driver Group	Number of Maps	Scope	Person	Started	Complete	Checked
Frequency of Fatal Crashes	1 Map	All Counties	Mark	✓	✓	FM
Rate of Fatal Crashes	1 Map		Frank	✓	✓	MP
Frequency of Fatal Crashes for each top 6 Major Driver Factors	6 Maps - 1 for each driver factor		Frank	✓	✓	MP
Rate Fatal Crashes for each top 6 Major Driver Factors	6 Maps - 1 for each driver factor		Frank	✓	✓	MP
Frequency of Fatal Crashes for Roadway Types (Roadway Function Class)***	5-6 Maps - 1 for each roadway type		Cat	✓	✓	FM
Rate of Fatal Crashes for Roadway Types (Roadway Function Class)	5-6 Maps - 1 for each roadway type		Frank	✓	✓	MP
Frequency of Fatal Crashes for Urban/Rural (Roadway Function Class)*	1 map for Urban / 1 map for rural		Larissa	✓	✓	PH
Rate of Fatal Crashes for Urban/Rural (Roadway Function Class)	1 map for Urban / 1 map for rural		Frank	✓	✓	MP
Frequency Fatal Crashes and Roadway Conditions	** # of Maps will vary		Paul	✓	✓	MP
Rate of Fatal Crashes and Roadway Conditions	** # of Maps will vary		Frank	✓	✓	MP
<70 Age Driver Group						
Frequency of Fatal Crashes	1 Map	All Counties	Mark	✓	✓	FM
Rate of Fatal Crashes	1 Map		Paul	✓	✓	FM
Frequency of Fatal Crashes for each top 6 Major Driver Factors	6 Maps - 1 for each driver factor		Mark	✓	✓	PH
Rate Fatal Crashes for each top 6 Major Driver Factors	6 Maps - 1 for each driver factor		Paul	✓	✓	FM
Frequency of Fatal Crashes for Roadway Types (Roadway Function Class)***	5-6 Maps - 1 for each roadway type		Cat	✓	✓	FM
Rate of Fatal Crashes for Roadway Types (Roadway Function Class)	5-6 Maps - 1 for each roadway type		Frank	✓	✓	MP
Frequency of Fatal Crashes for Urban/Rural (Roadway Function Class)*	1 map for Urban / 1 map for rural		Paul	✓	✓	MP
Rate of Fatal Crashes for Urban/Rural (Roadway Function Class)	1 map for Urban / 1 map for rural		Paul	✓	✓	FM
Frequency Fatal Crashes and Roadway Conditions	** # of Maps will vary		Paul	✓	✓	MP
Rate of Fatal Crashes and Roadway Conditions	** # of Maps will vary		Paul	✓	✓	FM
15-24 Age Driver Group						
Frequency of Fatal Crashes	1 Map	All Counties	Mark	✓	✓	PH
Rate of Fatal Crashes	1 Map		Mark	✓	✓	PH
Frequency of Fatal Crashes for each top 6 Major Driver Factors	6 Maps - 1 for each driver factor		Larissa	✓	✓	MP
Rate Fatal Crashes for each top 6 Major Driver Factors	6 Maps - 1 for each driver factor		Mark	✓	✓	PH
Frequency of Fatal Crashes for Roadway Types (Roadway Function Class)***	5-6 Maps - 1 for each roadway type		Cat	✓	✓	LM & MP
Rate of Fatal Crashes for Roadway Types (Roadway Function Class)	5-6 Maps - 1 for each roadway type		Mark	✓	✓	LM
Frequency of Fatal Crashes for Urban/Rural (Roadway Function Class)*	1 map for Urban / 1 map for rural		Larissa	✓	✓	PH
Rate of Fatal Crashes for Urban/Rural (Roadway Function Class)	1 map for Urban / 1 map for rural		Mark	✓	✓	LM
Frequency Fatal Crashes and Roadway Conditions	** # of Maps will vary		Larissa	✓	✓	MP
Rate of Fatal Crashes and Roadway Conditions	** # of Maps will vary		Mark	✓	✓	LM

Literature Review

Throughout the data collection process T.A.G.S has referred to many different types of literature to learn, study, and verify information for the project. There are many ways that the FARS data can be queried and analyzed. T.A.G.S has reviewed past reports, NHTSA previous studies, FARS term definitions, manuals, forms, exercise examples, and guides in order to properly query and analyze the data to obtain the information desired for the project. This also has helped verify the quality of the data obtained. Additional literature will, without a doubt, be reviewed as the project progresses.

Challenges

- **Driver Factor Usage** – A question arose concerning the usage and prioritization of the four Driver Related Factor (DF1, DF2, DF3, DF4) data fields associated with each "vehicle-driver" record. Our conclusions were that each driver is assigned four Driver Related Factor data fields, that there were no differences in priority among them, and that their values could be one of one hundred and two condition criteria values. In addition, it was concluded that a condition criteria value could not be repeated within any single "vehicle-driver" record, however, the condition criteria value could be assigned to any one of the four Driver Related Factor data fields within that record. We confirmed the above conclusions with the Information Technology Manager of the National Highway Traffic Safety Administration States Data Reporting Systems Division. With this understanding, a separate query using each Driver Related Factor was created to obtain the total number of any particular condition criteria value which could possibly be associated with any of the four Driver Related Factors. Below is a small example:

  vehicle-driver record 1 - DF1 = (5), DF2 = (44)
  vehicle-driver record 2 - DF1 = (6). DF2 = (28)
  vehicle-driver record 3 - DF1 = (44)
  vehicle-driver record 4 - DF1 = (6), DF2 = (44), DF3 = (28)

  By querying each of the four Driver Related Factors separately and summing the counts for each condition criteria value, the totals below are obtained.

  condition criteria value (5) = 1
  condition criteria value (6) = 2
  condition criteria value (28) = 2
  condition criteria value (44) = 3

  These are the frequencies of occurrence for each condition criteria value associated with any fatal crash.

- **Unequal total for the number of crashes using specific data fields versus the actual number of total crashes.** - The resulting query produced a total of 1896 crashes with the following statement "Because you are counting Number of Crashes with the variables County and Driver Related Factors (1), the sum of the totals (1896) may not equal the Number of Crashes (1831)." Our conclusion is that the resulting total of 1896 is due to the fact that, in some cases, multiple drivers (vehicles) can be associated with each crash, and therefore, resulting in an increased number of Driver Related Factors associated with each crash producing a higher total.

Current Work

Data Preparation
Since the FARS data collection is complete and quality of the data has been checked by T.A.G.S, we have been working on formatting and preparing the data to properly join with other spatial data involved in the project. This preparation is essential in order to start mapping the data and performing the spatial analysis for the project. Provided at the end of this report are some sample maps that show some of the current FARS data mapped and analyzed. These maps also represent a similar format of how the data will be presented in the final report.

Other Relevant Data Analysis
Now that the FARS data has been collected and is ready for spatial analysis, we are exploring another related data source that could have some relevancy to the project in order to ascertain whether it feasible to add this data to the current FARS data analysis. The additional accident report data provided to us is currently being explored for this purpose. Using this data, other data sets, or even more of the FARS data, we are exploring the possibilities of some further analysis. This could involve focusing on a county that has some fatal crash trends and examining it in greater detail in order to add further analysis for the project. If discovered that such further analysis is sensible and related to the project, T.A.G.S will conclude accordingly whether we will have time to expand the project scope.

Future Work

Map Creation
Approximately fifty-two maps will be created based on the results of our data acquisition process. These maps will represent by county the fatal crash frequency and the normalized frequency rate (fatal crashes per 100,000 population) data sets of each age group and condition selected.

Metadata Creation
Metadata (data about the data) will be provided as a summary document of the content, quality, type, creation, and spatial information for each data set used in the map creation process. Below is a small subset example:

Identification Information:
Citation:
Citation Information:
Originator: Traffic Accident Geospatial Solutions (Texas State University Fall, 2009 GEO 4427)
Publication Date:
Title: Texas Fatal Traffic Crash Frequencies for Drivers over 70 Years Age
Geospatial Data Presentation Form: vector digital data
Publication Information:
Publication Place: San Marcos, TX

Publisher: Traffic Accident Geospatial Solutions (Texas State University Fall, 2009 GEO 4427)
Description:
Abstract: Fatal Analysis Reporting System (FARS) traffic fatality data for crashes involving drivers over 70

Data Analysis and Interpretation
Analysis and interpretation of the fatal crash frequency and the normalized frequency rate (fatal crashes per 100,000 population) maps will continue through mid-November. In addition to providing a visual representation, our analysis may reveal counties and conditions of interest that will warrant a further detailed investigation to pinpoint driver related issues and potential roadway condition improvements.

Website Development
The T.A.G.S. website will include a project description which includes the purpose and background information, all project documentation, fatal crash frequency and the normalized frequency rate (fatal crashes per 100,000 population) maps, data sets, contact information, and links to related websites.

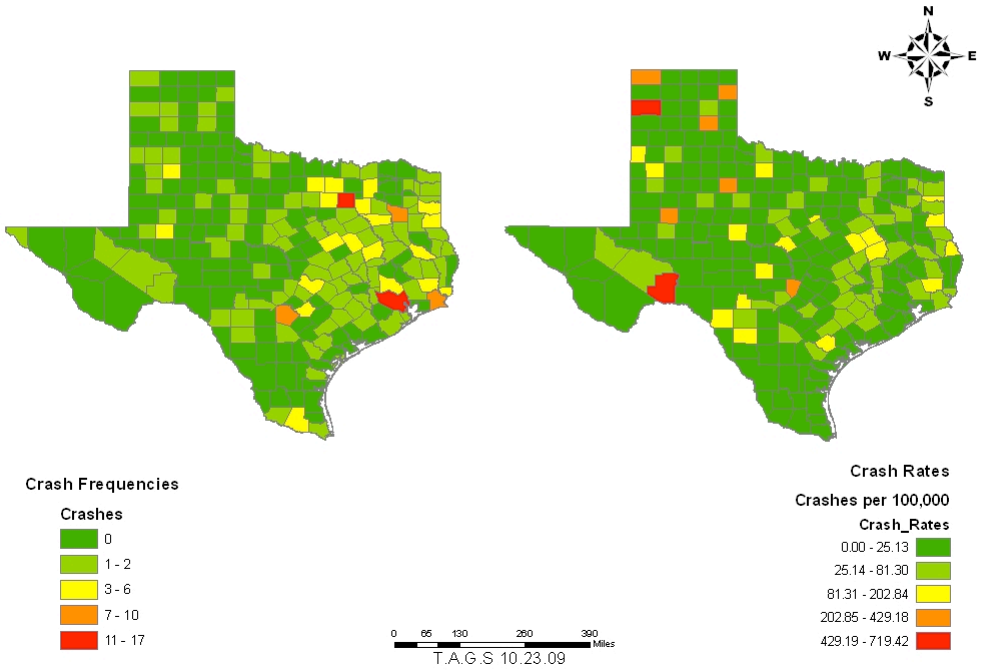
Final Deliverables
The final deliverables will include a detailed final report of our analysis, project specific website, a professional poster summarizing the results of our analysis, and a DVD containing all project data.

Challenges
Individual county analysis may be limited based on the difficulty in connecting outcome data sets to the FARS data sets.

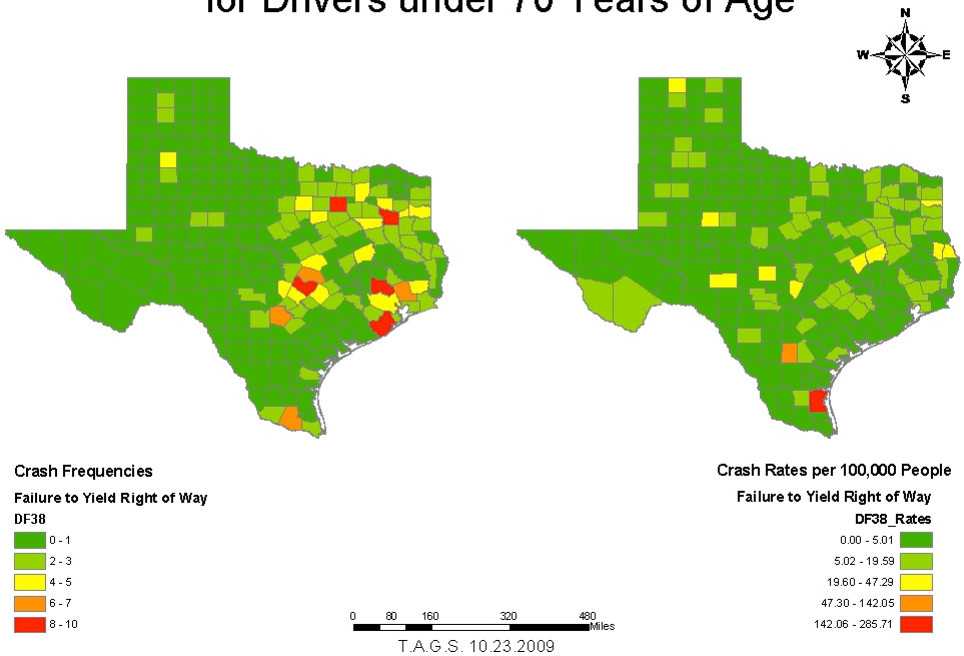
Overall Project Assessment

Overall, T.A.G.S. is confident about completing the project by the previously chosen deadline. We have been able to overcome all challenges that we have encountered throughout the course of our data collection process, and soon we will be entirely involved in the interpretation of our acquired data. We are very confident that the project will be concluded and ready for presentation on December 9, 2009, and look forward to presenting our completed work and findings to you on this date.

Fatal Traffic Crashes for Drivers Over 70 Years of Age



Fatal Crashes due to Failing to Yield Right of Way for Drivers under 70 Years of Age



Participation

Mark Poitras – Summary, Rate Calculations, Literature Review, Current Work Section
Paul Head – Map Creation and pilot study
Catherine Perille – Query Process and Quality Assurance
Frank Martin – Future Work Section and Challenges
Larissa Matin – Introduction, Project Description, Scope, Overall project assessment

All participated in the slideshow production