Vehicular Crash Analysis in Washington County, Oregon

Nathan McNeil and Aaron Edelman

USP 531 Final Project

March 16, 2009

INTRODUCTION

A fatal vehicular crash occurs approximately every fifteen minutes somewhere in the United States. Vehicular crashes, also known as accidents, can typically be defined as one or multiple vehicles colliding with an object resulting in bodily injury or property damage. The purpose of this report was to create a useful method utilizing a geographic information system (GIS) to identify problem areas and potential causes within Washington County, Oregon. Crash data has been widely studied in the past, but this project focuses on how spatial crash data can present easily interpreted visual products.

In the United States from 2000 through 2002, a reported 127,146 auto fatalities occurred according to the United States Department of Transportation (DOT) National Highway Traffic Safety Administration (NHTSA) through the Fatality Analysis Reporting System (FARS). FARS also reported 1375 fatalities in the state of Oregon and 103 fatalities in Washington County during that same time period. A total of 58 of those 103 fatalities occurred on state highways. Several government agencies measure the economic value of statistical life (VSL) in order to quantify a usually unquantifiable loss so that cost benefit analyses can be run on policy options. At the time the project data was collected, different agencies estimated VSL fell into the range of $3 million to $8.04 million per life.[1] If a median value of $5.8 million was assumed, that would equate to a loss of $597.4 million for Washington County in a three year period due

---

[1] The Washington Post noted that in 2008, the Environmental Protection Agency decreased its figure for VSL from $8.04 million to $7.22 million, while the U.S. Department of Transportation increased its figure from $3 million to $5.8 million. The Consumer Product Safety Commission has remained steady for the past decade at $5 million. See "Value of Life," *Washington Post*, July 19, 2008, accessed 3/17/09 at http://www.washingtonpost.com/wp-dyn/content/graphic/2008/07/19/GR2008071900086.html.

to auto fatalities.[2] Although putting a value on a human life can be troublesome to many, these staggering numbers coupled with even a modest value for property damage associated with nearly every crash helps put the need for crash prevention in perspective. Even though safety in the auto industry and on the transportation system has progressed remarkably since the beginning of motorized transportation, there are still significant societal and personal costs to address.

PROBLEM STATEMENT

This research proposes to conduct an exploratory review of crash data in Washington County, Oregon, to identify the location of problem areas on state highway segments. Studies have shown that the analysis of crash information to identify problem areas using GIS can be a useful method in determining how a municipality can allocate funds aimed at ameliorating dangerous conditions.[3] Once problem areas, better known as "hotspots" or "blackspots" by researchers, are identified, further analysis can be conducted to identify if there are unique features in that area that increase the likelihood or seriousness of crashes. The goal was to apply previously considered applications of using GIS for identifying crash hotspots and apply them to one local area to assess potential areas for crash alleviation.

The role GIS played in the research evolved into the three major themes of data compilation, descriptive outputs and predictive solutions. As a tool for data compilation, GIS combined the data from the various sources into one central location. This allowed for easy

---

[2] A cost benefit analysis using VSL would assume, all other things being equal, safety measures that prevented each of the 103 fatalities in Washington County would be worthwhile if the price tag was $597.4 million or less. Further, a program that prevented one fatality per $5.8 million spent would be worthwhile. A higher VSL would make more expensive programs worthwhile.

[3] Mitra, Sudeshna. "Spatial Autocorrelation and Bayesian Spatial Statistical Method for Analyzing Fatal and Injury Crash Prone Intersections." Submitted for the Transportation Research Board annual meeting, 2009. Washington DC.

manipulation of the data and attributes to facilitate analysis and allow spatial comparison. Overall, the descriptive outputs provided spatial mapping and visualization to make the crash data analysis more effective and efficient. The need to produce maps and for spreadsheet software necessary to reach statistical conclusions was reduced into one operation. Analysis of attributes and a study of the spatial maps provided predictive solutions to reducing crashes in a hotspot.

DATA SUMMARY

Crash Data

The data used in the study was provided from several main sources. The central dataset was a collection of databases representing all crashes recorded on state highways in the state of Oregon. The data was acquired through accident reports and collected into a database by the Oregon Department of Transportation (ODOT). For this study, crash data was accessed through the Portland State University Intelligent Transportation Systems laboratory, which stores the information on a network drive.

The database includes columns for 54 variables, including crash date, highway number, milepost, crash type, number of fatalities, number of serious injuries, the influence of alcohol use or if speed was involved, and other factors. Each crash had already been geocoded to a specific location along a state route in Oregon. Databases from the years 2000, 2001 and 2002 were used in the study. Crashes on local and county roads (roadways less significant than major arterials) were not recorded in the ODOT data.

The quality of the crash data appeared to be high; however like most GIS data, the quality is dependent, in the end, on the input. In the case of crash records, reporting was an important factor because it required the police or the individual involved in the crash to submit the crash report. During the period under study, legally reportable crashes in Oregon were those involving death, bodily injury or damage to any one person's property in excess $1,000.[4] Two separate meta-analyses of accident studies each found a high-degree of accident under-reporting, finding that statistics reflect only 70%-80% of serious injuries from accidents and 95% of fatalities.[5] For the information that was reported and included in the database, only a few inaccuracies were identified (as will be discussed later, a few records were incorrectly geocoded, with one being labeled in the wrong county and others appearing slightly off the route of the state highway). However, there were omissions in the data in several cases. For example, over half the records did not have a time of day attributed to the crash. This omission would present substantial problems for analyses based on the time that crashes occurred; however, time was not used as a factor in this analysis.

Geographic Data

Another important type of dataset was the geographic data against which crash data and other data would be mapped. This data was received from two main sources. First, shapefiles that ODOT produced for all state owned or maintained highways and mileposts were acquired from the Oregon Geospatial Enterprise Office online Spatial Data Library.[6] The Spatial Data Library also provided us with shapefiles for borders for the state of Oregon and county boundaries

---

[4] Oregon Department of Transportation website. "Trans Data - Crash Data". Accessed March 7, 2009. http://www.oregon.gov/ODOT/TD/TDATA/car/CAR_Main.shtml

[5] Elvik, R. and Mysen, A.B. (1999). Incomplete accident reporting: Meta-analysis of studies made in 13 countries. Transportation Research Record Journal of the Transportation Research Board, 1665: 133-140; and Hauer, E. and Hakkert, A.S. (1988). Extent and some implications of incomplete accident reporting. Transportation Research Record Journal of the Transportation Research Board, 1185: 1-10.

[6] Oregon Geospatial Enterprise Office online Spatial Data Library website accessed online at http://www.oregon.gov/DAS/EISPD/GEO/alphalist.shtml.

(produced by the Bureau of Land Management). A second source for geographic data was the Metro Data Resource Center maintained Regional Land Information System (RLIS). Shapefiles for cities and streets in the Metro area and an elevation raster for northwest Oregon were acquired from RLIS. Population and attribute data by census block produced by the U.S. Census Bureau was also acquired from RLIS. All RLIS data was acquired through Portland State University, which maintains a subscription to RLIS. There are no known problems with this data and the quality of the data appeared to be high.

Liquor License Data

Liquor license data was acquired from the Oregon Liquor License Commission (OLCC) for all on-premise drinking establishments in Washington County. The data was obtained in comma separated value (.csv) format and contained fields for license type, place name, address, zip code, as well as an identification number and date of issuance for the license. Unfortunately the date of issuance renews (perhaps when annual licenses are renewed) every April first, such that only licenses issued after that date can have their opening date identified. Since the OLCC data is current data from 2009 and the crash data was from 2000-2002, there was a built-in difficulty when comparing the data.

METHODOLOGY

Analysis Approach

As mentioned in the Problem Statement, one objective of the project was to utilize the data to identify crash hotspots so that those spots might be further studied. Several accepted methods of identifying problem areas were employed. Kim and Levine suggest three main approaches to conducting a spatial analysis of crashes: point analysis (which looks at crashes as events that happen at specific times and places), segment analysis (which examines crashes in relation the roadway segment on which they occur), and zonal analysis (which typically looks at crashes in relation to a census or other designated geographic zone).[7] The first two methods were used in this analysis; however, because the crash data consisted solely of accidents along state highways, the zonal analysis, with its added value in examining distributional aspects of data dispersed across zones, was not appropriate in this case. Other researchers have turned to more complex methods to identify hotspots. Mitra found the kernel density tool helpful in analyzing clustering of crashes, but found the method lacking in examining the intricacies of crashes that located in the attribute data, and suggests using spatial auto-correlation to examine specifically the attributes for fatalities and serious injuries.[8] While some of the more statistically complicated data analysis methods presented in the literature are beyond the scope of this project, the literature was used as a guide to get beyond simple point location analysis and to examine some of the important attributes, including fatalities, serious injuries, and alcohol use.

Fixing Data

The data acquired from various sources required a varying degree of editing, manipulating or geo-coding in order to be usable. The ODOT crash data tables for the years 2000, 2001, and 2002 were combined into one shapefile containing all crashes on state roads for those years. All crashes in Washington County were selected and exported to a separate shapefile. This new shapefile yielded one crash plotted well outside the county boundaries, indicating that either the geocoding was incorrect or the county attribute was incorrectly labeled as Washington County -

---

[7] Kim, K., and N. Levine. Using GIS to improve highway safety, *Computers, Environment and Urban Systems*, Vol. 20, No. 4, 1996, pp. 289-302.

[8] Mitra, 2009.

this crash was removed from the sample. As mentioned in the data summary, some records did not have data in all fields. The Wilson River Highway segment had 42 crashes that needed corresponding highway numbers manually inserted to properly analyze the data.

The OLCC list of establishments licensed to sell alcohol on-site was geocoded using the addresses provided by OLCC against an address locator populated by street and zip code information from the RLIS streets shapefile. The geocoding matched about 82% of the 647 records, leaving 118 unmatched. Of those 118, an additional 76 were matched through manually editing the address records to fix unrecognizable addresses.[9] Establishments licensed to serve alcoholic beverages, including those fully licensed to serve liquor and those with partial licenses to sell beer and wine only, were identified. Other licenses, such as those that only allow the manufacturing, importing, or shipping of alcohol, were removed from the dataset.

Several routes in the ODOT state routes shapefile were missing segments on which crashes were located. One such example was the Tualatin Valley Highway, which splits into a couplet on three occasions, once in Hillsboro, once in Cornelius, and again in Forest Grove. In each case, the westbound segment of roadway runs one block north of the eastbound segment. The ODOT shapefile in each case only followed the northern (westbound) section of the route. The editor tool was used to add line segments and capture these missing portions of roadway so that crashes could be mapped to them and analyzed.

Compiling Data

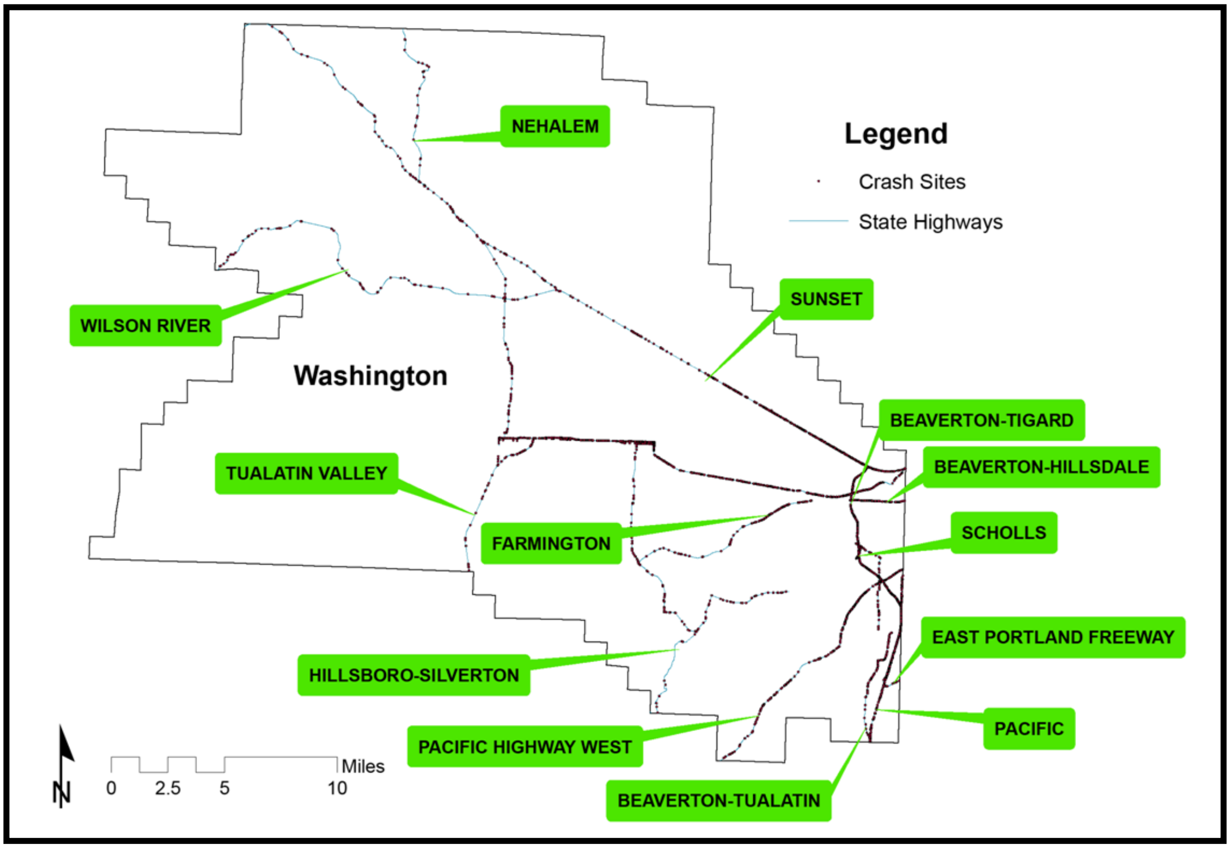
To begin the project research, the crash data was overlaid on a map of Oregon with all the state roadways displayed. The beginning study area extended to the Oregon borders and was reduced

---

[9] For example, "TV Hwy" was coded "Tualatin Valley Highway," yielding 32 additional matches, "BVRTN HLSDL HWY" was coded "Beaverton Hillsdale Highway," yielding 31 additional matches. Further, Google Maps was utilized to correct several incorrect zip codes.

to only include the Washington County area. The study area was refined to help define the project scope to a manageable size. This was accomplished by merging the three years of crash data, clipping it to Washington County border, as seen in Map 1.

Map 1: Washington County Study Area with 2000-2002 Highway Crash Locations



Once the base map was created, additional data was added to help determine if any spatial relationships existed. Census data was primarily used along with zoning to see if population or land use had an effect on the number of crashes in a particular area. This data was also clipped to the county level to keep files at a manageable size.

<u>Assessing Crashes by Highway</u>

A preliminary goal was to assess if particular highways were more dangerous than others. One way that assessed the relative safety was to examine how many crashes occurred per mile for each state highway in Washington County. "Catchment areas" were created by buffered all state routes in the county by two hundred feet and dissolving by highway number, yielding polygons to which crashes could be spatially joined. Next, crashes were joined to the buffered highways to give a total number of crashes over the three years. The length was calculated for each highway using the first and last milepost markings (contained in crash records), and verified using the measuring tool in ArcGIS (which allowed us to correct one error on Interstate 5/Pacific Highway). A simple analysis dividing the total number of crashes by the length in miles yielded an average number of crashes per mile, demonstrated in Table 1.

Table 1: Crash Rates by Washington County Highway, 2000-2002

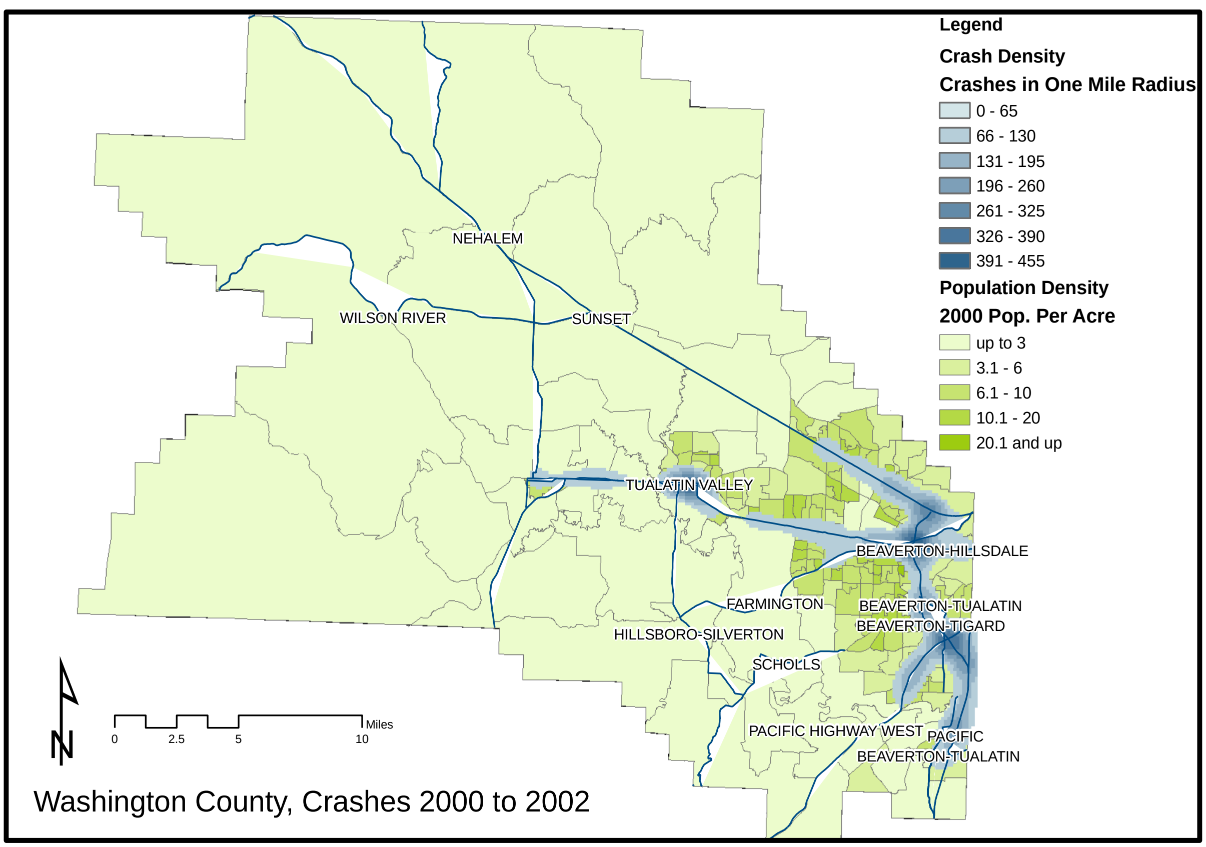
Highway Name	Crashes	Miles	Crashes/Mile
Beaverton-Hillsdale Highway	397	2.43	163
Beaverton-Tigard Freeway	932	7.46	125
Beaverton-Tualatin Highway	348	12.53	28
East Portland Freeway	29	1.28	23
Farmington Road	238	8.80	27
Hillsboro-Silverton Highway	192	15.00	13
Nehalem Highway	171	21.52	8
Pacific Highway	757	7.58	100
Pacific Highway West	925	12.27	75
Scholls Ferry Road	163	6.01	27
Sunset Highway	1148	35.95	32
Tualatin Valley Highway	2134	26.43	81
Wilson River Highway	81	18.82	4
Totals	7515	176.08	43

However, because some highways were long and traversed both dense urban areas and sparse rural areas, the results smoothed out localized differences in crash numbers. Many highways which had relatively few crashes per mile appeared to have localized segments of high crash activity. For example, this early analysis suggested the Wilson River Highway was the safest state highway in the county at only 4 crashes per mile; however, certain segments actually experienced up to 10 crashes per mile with extremely high fatality rates.

Identifying Hotspots

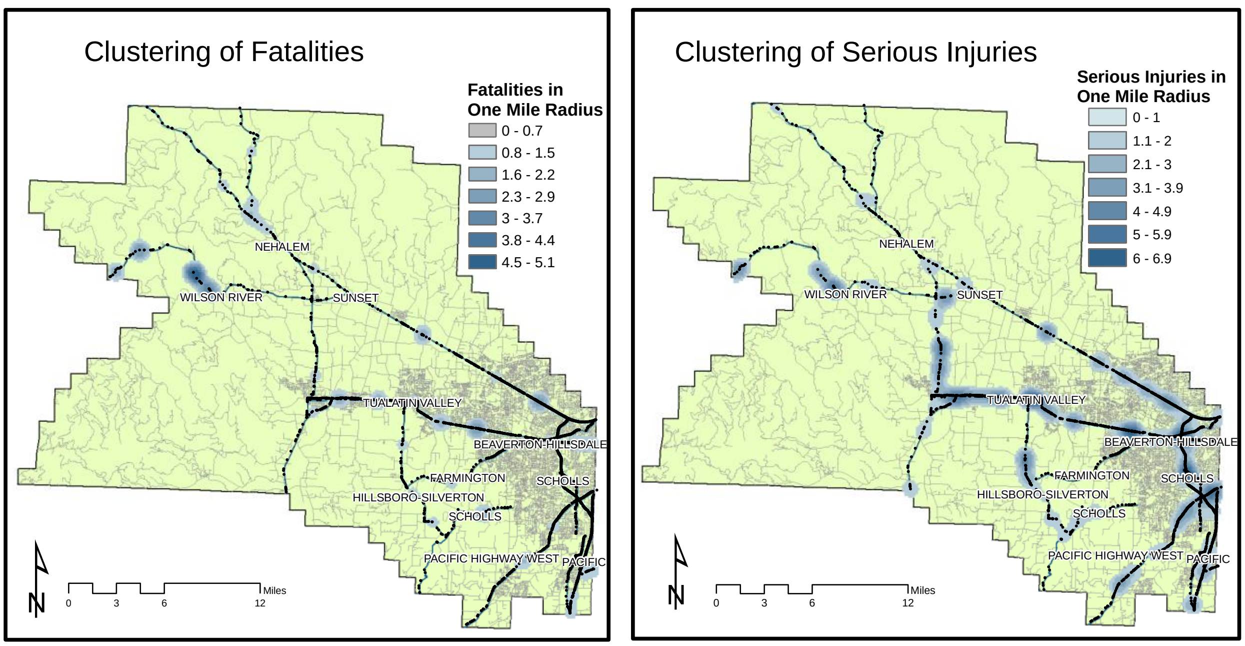
To get beyond the potentially misleading base number of crashes per mile along Washington County highways, point analysis, including a kernel density test, and segment analysis were used to identify hotspots of increased crash activity or elevated occurrences of dangerous crashes involving fatalities or serious injuries. A spatial analysis was done utilizing kernel density with a 650 foot kernel and one mile search radius. This test was performed on all crashes and mapped against block group population density. As seen in Map 2, this revealed clustering of crashes in areas of denser population and more intersections.

Map 2: Crash Density and Population Density, 2000-2002



To get beyond this intuitive finding, the kernel density was run again using the same parameters but isolating the attribute of number of fatalities. Illustrated in Map 3, this test indicated that, although crashes occurred most frequently in the densest parts of the county, crashes with fatalities were clustering in other areas. The clearest cluster of fatalities was on the Wilson River Highway in the western portion of the county, with smaller clusters occurring along the Tualatin Valley Highway, among other places. The kernel density test was also run for number of serious injuries to identify if there were particular areas where dangerous, but non-fatal crashes were occurring. The Tualatin Valley Highway again was identified as a hotspot, as seen in Map 4.

Maps 3 and 4: Clusters of Crash Fatalities and Serious Injuries, 2000-2002



Spatial statistical tests were also performed to locate patterns in the crashes along the roadways, but did not yield conclusive results.[10]

<u>Methods of Assessing Hotspots</u>

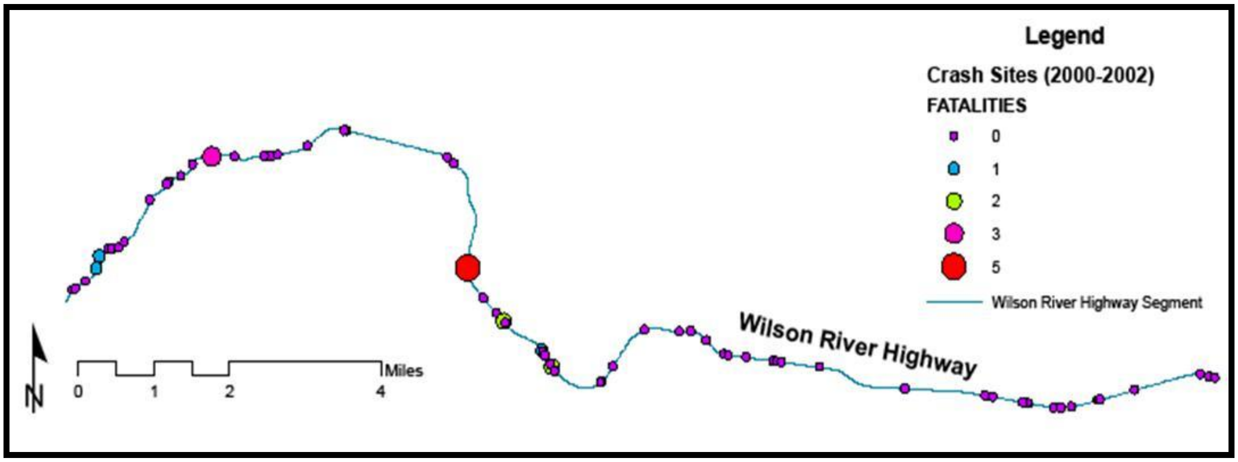
---

[10] The first test was a spatial auto-correlation tool (Morans I) which found that all fatalities in Washington County were random. This test was then used at the Wilson River Highway level which yielded the same results with a Moran's Index of 0.07 and a Z score of 0.2 standard deviations. The second test was a cluster and outlier analysis, which found no statistical patterns.

One way to incorporate some of the detailed attribute data in the analysis was to run through the attributes, identifying attributes that might provide insight, and selecting all crashes by that attribute. The resulting working map documents reflected the frequency and location of the selected attributes. In this way crashes were selected that involved alcohol, that involved fatalities (and those that involved multiple fatalities), head-on collisions, weather-related crashes, crashes by driver age, and more. While this was an interim step, it provided an important method to start exploring the data, and is a method that can provide some quick, "back-of-the-envelope"-type, assessment for local governments looking at crash data.

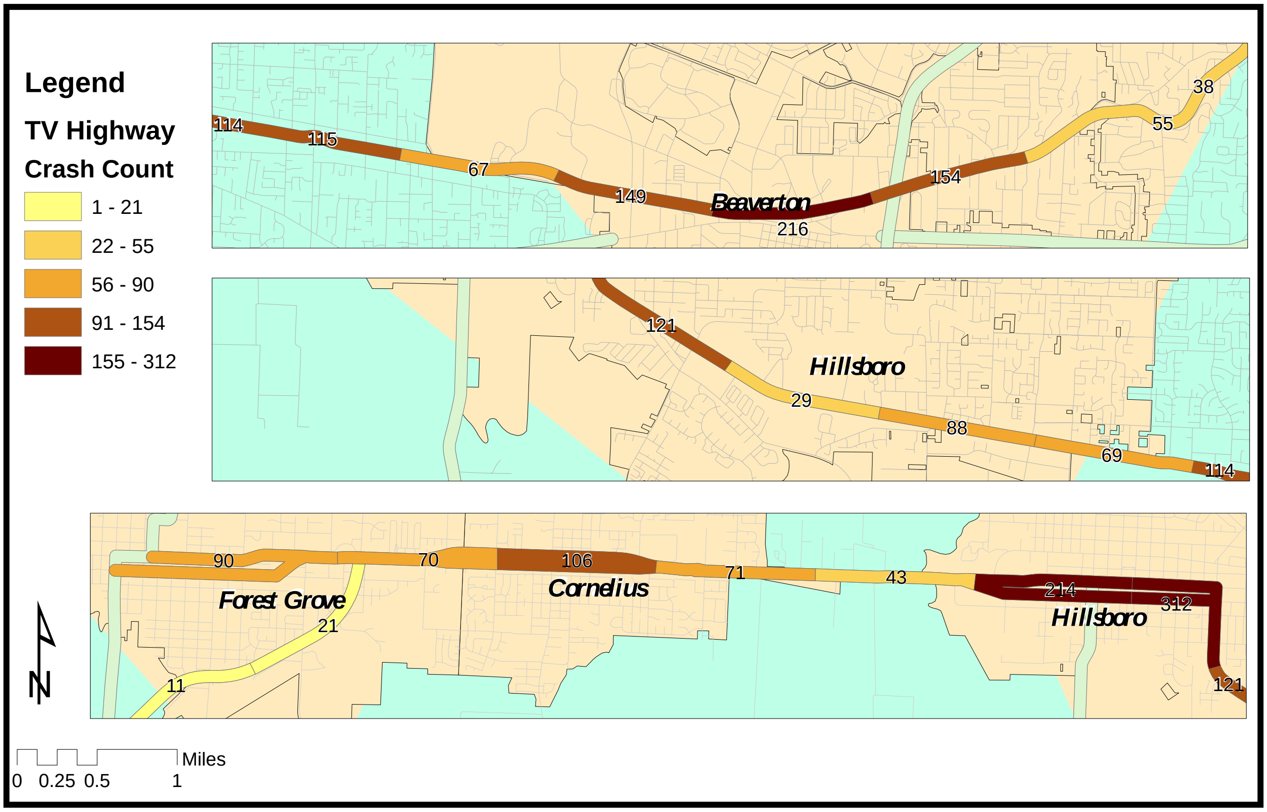
Point/Cluster Analysis: A section of the Wilson River Highway near the town of Glenwood was identified in the kernel density test as being a hotspot for crash fatalities. As shown in Map 5, this stretch experienced several crashes that resulted in multiple fatalities. After this hotspot was located, the focus was to determine if the cluster had a recurring theme or if each individual crash had a specific cause. The recurring theme discovered by looking at each individual crash's attributes pointed to weather and road conditions. These two attributes obviously play a role together because it can be deduced that if the weather conditions are unfavorable, there will be a high likely-hood that the road conditions will also be unfavorable. Once this was discovered, the next step was to determine why this stretch of roadway experienced a majority of the crashes during inclement weather and why the crashes were not evenly dispersed over the entire roadway. A close look pointed in a number of directions such as elevation change, poor alignment, poor shoulder conditions and a tree canopy over the roadway.

Map 5: Wilson River Highway Fatalities, 2000-2002



Segment Analysis Incorporating Liquor License data: The Tualatin Valley Highway was also identified as being a hotspot for both fatalities and serious injuries. In order to assess the length of the highway for specific crash hotspots, the highway buffer was broken down into one mile segments and crashes were then joined to these segments to reveal areas with higher concentrations of crashes. As with the county-wide kernel density test for all crashes, these segments, shown in Map 6, revealed that a greater number of crashes were occurring within cities and towns, generally in areas with more people and streets.

Map 6: Tualatin Valley Highway One Mile Segments, 2000-2002



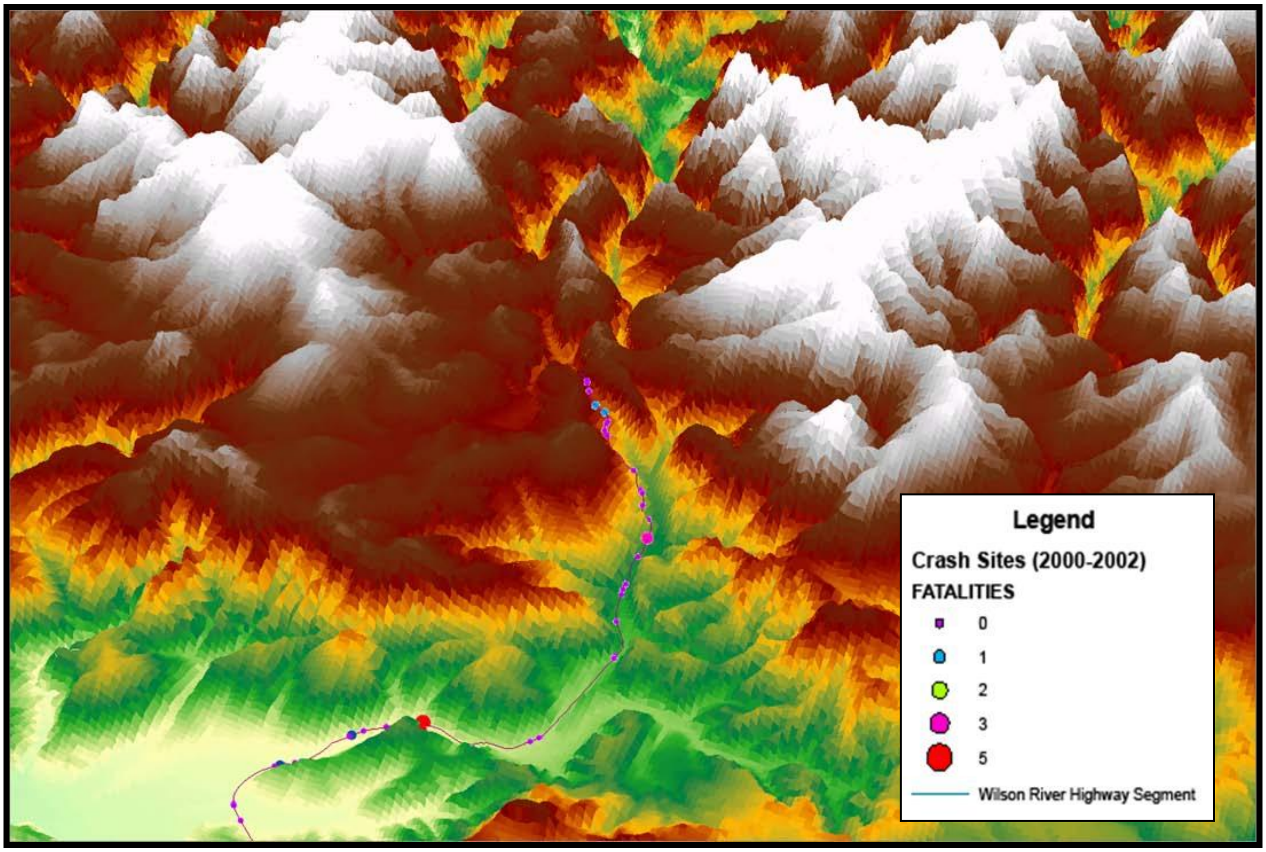
Using the "back-of-the-envelope" approach described above, it was determined that nearly twenty five percent of alcohol related crashes in the dataset occurred on the Tualatin Valley Highway. The next step was to determine if those crashes were more likely to occur near establishments with alcohol licenses. Establishment with full alcohol licenses were given 0.1 mile dissolved buffers, and crashes that were marked as having "alcohol involved" were joined to both the license establishment buffers and the highway segment buffers. Each of the buffers was mapped with graduated colors based on the number of alcohol related crashes. The graduations were coordinated to show equivalent crash rates (the establishment buffers were normalized by the length in miles of highway passing through the segment, while the highway segments were already in one mile increments). Finally, the actual locations of the nineteen alcohol related crashes along the highway were added to the map, which appears in the Analysis Results section.

ANALYSIS RESULTS

Wilson River Analysis

Using the point analysis of crashes in Washington County and utilizing graduated symbols to reflect the number of fatalities occurring, one major hotspot was identified along the Wilson River highway in the western section of Washington County. Along a two mile stretch of highway, five accidents had a fatality (out of 47 fatal crashes county-wide over the three year stretch). Of those four accidents, four had more than one fatality (out of seven total multiple fatality crashes countywide over the three year period). Further, this stretch of road involved a clustering of head-on collisions. A raster elevation rendering looking easterly over this stretch of road in ArcScene revealed that this clustering of crashes occurred in an area where the roadway climbs into the Coast Mountains through a narrow valley, as shown in Scene 1. A close look at this segment of roadway revealed a rural meandering two lane undivided highway shouldered by a guardrail protecting a steep drop-off on the north side. Numerous stretches of this tree line road have no openings in the canopy to allow sunlight to hit the pavement and many large pines fall within the clear zone along sides of the roadway. The tree canopy could be directly linked to the forty nine percent of crashes that occurred during times when the road surface had some form of moisture present. Only two crashes involving alcohol and two crashes where it was unknown if alcohol was involved were reported in the three-year period studied for this stretch. Speed did play a role in thirty nine percent of the crashes making this a significant contributing variable.

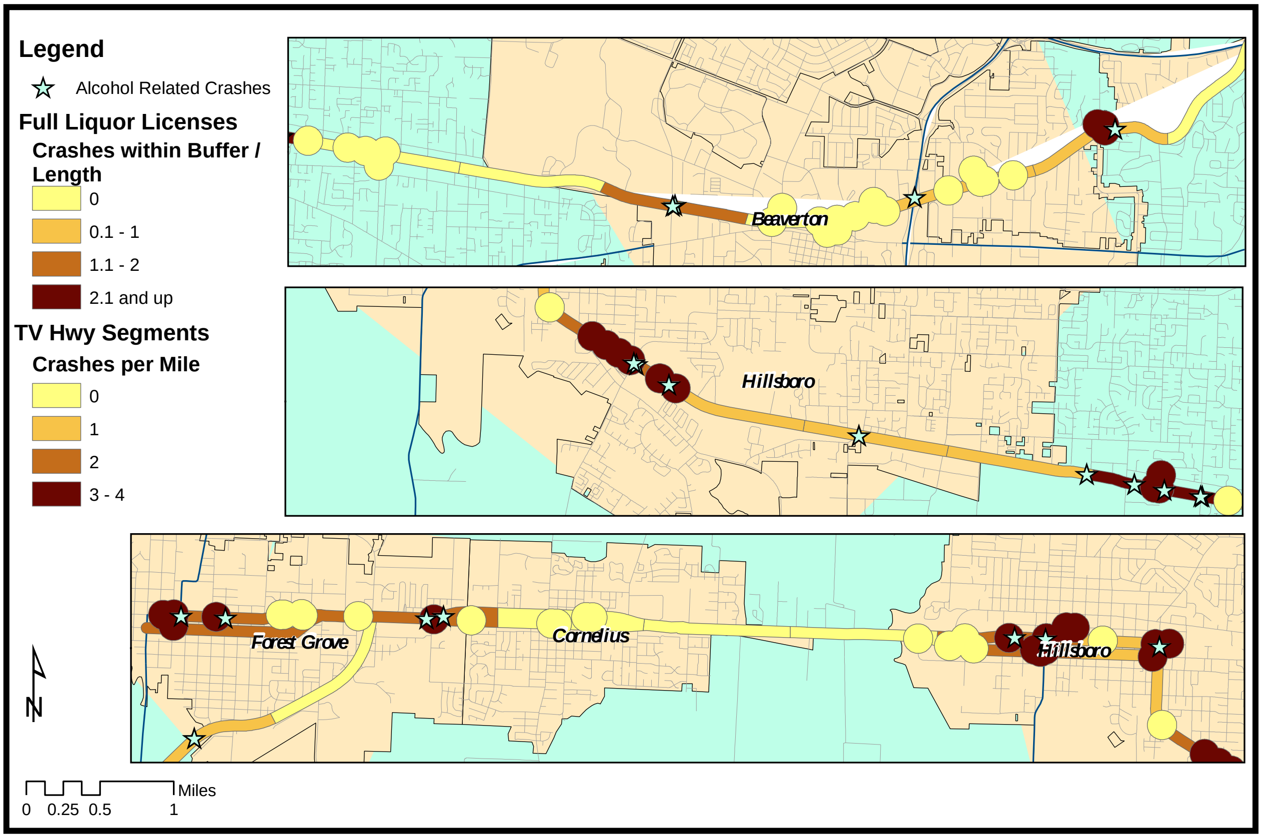
Scene 1: Raster Rendering of Wilson River Highway Segment Elevation



<u>Tualatin Valley Analysis</u>

Using a simple point analysis and segment analysis, and selecting for any fatal crashes and then crashes involving alcohol, the Tualatin Valley Highway was identified as being a hotspot. This highway stretch had nine fatal accidents, two of which had multiple fatalities. In fact, the short stretch of Wilson River Highway and the Tualatin Valley Highway combined account for six of the seven multiple fatality accidents in Washington County over the three-year period. There was also a concentration of head-on collisions here, which may be part of the reason for the clustering of serious injuries seen in the kernel density test. Nineteen crashes along the Tualatin Valley Highway were alcohol related, which accounted for nearly 25% of the 81 alcohol related crashes in Washington County over the three year period. Map 7 was created to demonstrate the occurrence of these crashes in relation to establishments with full liquor establishments, overlaid on the one mile segments to provide a baseline for the area.

Map 7: Tualatin Valley Highway Alcohol Related Crashes



This map revealed that on eastern half of the highway, there does not appear to be a connection between the license locations and alcohol related crashes. However, on the western portion of the highway there does appear to be a relationship: starting Midway through Hillsboro, every alcohol related crash occurred within the 0.1 mile buffers around the licensed establishment. These results merit further study, but one explanation may be that the more urbanized eastern half of the county closer to Portland may be seeing people coming from a wider collection of places and thus alcohol related crashes are more dispersed. In the western half of the county, as the population density decreases, there appears to be a higher correlation between these places of alcohol sales and alcohol related crashes. Funding aimed at targeting specific locations for battling drunken driving might be best suited for these places further east in the county.

CONCLUSIONS

As with all research, each project experiences limitations throughout the course of study. This project was no different and several drawbacks were noted as they surfaced. The first item was that the crash data was seven years old at the time of the analysis. Secondly, the crash data was missing data in several attributes (e.g. nearly half of records were missing time of incident). An additional record that would have provided useful for network analysis would have been to have origin and destination information for the commuters involved in each crash. Finally, if traffic volumes were available for each roadway, the relationship between density and crash frequency could have been validated. One area of difficulty experienced was isolating influence of factors such as licensees and census data.

However, there are important and potentially life-saving lessons to be learned here. Using kernel density tests to identify point and segment hotspots for analysis, specific locations or stretches of road with poor safety records were noted. GIS was used both in identifying those hotspots using methods suggested in prior studies, in the "back-of-the-envelope" analysis of crash attributes, and in the visualization methods adopted to display these results. Although crash causation is uncertain without further results, potential causes have been identified for further assessment. The results suggest funding priorities for safety improvement projects, such as along the section of Wilson River Highway, or additional law enforcement, such near fully licensed establishments on the western half of the Tualatin Valley Highway. A safety project could be as simple as adding a "Slippery When Wet" sign to a section of roadway that has a tree canopy and never receives sunlight to dry the pavement. Further, the methods outlined in this paper, of using accepted techniques of identifying hotspots, along with the simple visualization techniques, can provide an avenue for Washington County or other counties to assess and act on crash data. As many agency budgets grow even tighter, these methods may be used by the county or even concerned citizen groups to assess priorities or lobby for improvements to the agencies. Given the high cost of fatal crashes, the modest costs associated with these methods can surely be justified.