Spring 2015

Using Spatial Analysis to Identify High-Risk Driver Residential Areas in South Carolina

Kweku Brown
Clemson University

Using Spatial Analysis to Identify High-Risk Driver Residential Areas in South Carolina

Glenn Department of Civil Engineering

Kweku T. Brown

Objective

According to the Fatality Analysis Reporting System (FARS) of the National Highway Traffic Safety Administration (NHSTA), fatal crashes reduced from 39,252 to 29,757 between 2005 and 2011. Over the same time period vehicle crash fatality rates per 100 million vehicle miles travelled (VMT) and vehicle crash fatality rates per 100,000 population also reduced nationally. Although South Carolina has seen similar trends in all three crash statistics (according to FARS and South Carolina Department of Public Safety –SCDPS) within the same time period, fatality rates within the state are significantly higher than national rates. Addressing safety issues at high crash incidence locations through crash countermeasures or better geometric design helps to make roadways safer, however the most influential and ever-present factor in most crashes, the driver, is still not addressed. Studies have shown that the vast majority of all crashes involve human error, while a significantly lower percentage of crashes involve factors related to the roadway and/or the vehicle. This paper investigated the socio-economic and demographic characteristics of residential locations (found using 9-digit zip code data) of drivers involved in crashes in South Carolina aggregated to census block groups.

Traffic Safety Policy In South Carolina

Strategic Highway Safety Plan: The Road Map to Safety (SHSP): Recent efforts by SCDOT and SCDPS to reduce vehicle crashes, especially injury and fatal crashes within the state led to the development of the Strategic Highway Safety Plan: The Road Map to Safety (SHSP) in 2003 and published in 2007. Using 2004 as the baseline year two goals were adopted. The two goals were to reduce traffic fatalities from 1046 in 2004 to 784 or fewer in 2010 and to reduce the number of traffic crash injuries by 3% annually *(7)*.

SHSP Evaluation: In 2010, there were 810 traffic fatalities in South Carolina. This number although significantly (23%) lower than the number of fatalities in 2004 narrowly failed to meet the set goal of 784 or less. There was a 6% overall reduction from 51226 injuries in 2004 to 48303 injuries in 2009. However, year by year analysis shows that the annual 3% goal was not met within the five year span. Further, a 3% annual reduction in traffic crash injuries over the five year span would result in 43990 injuries in 2009 which was also not met. Although most of the proposed goals were not achieved, the evaluation of the effect of the SHSP showed that there has been a substantial reduction in total crashes, especially fatal and injury crashes leading to significant gains in transportation safety in the state

Factors Affecting Traffic Crash Frequency and Severity

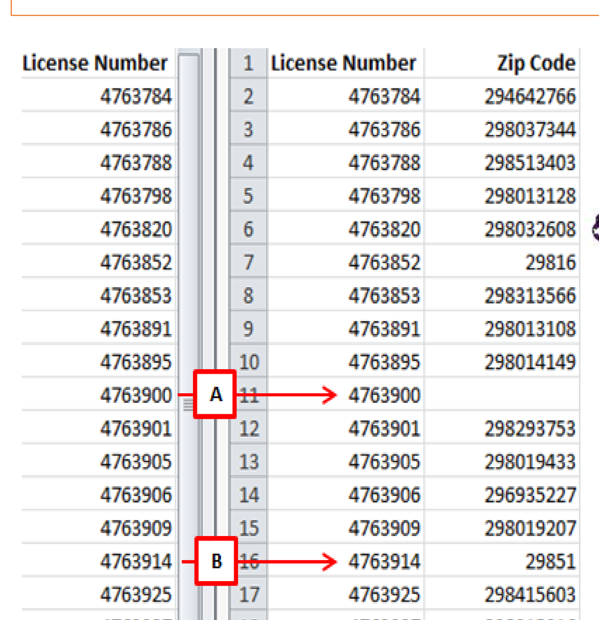
Demographic Factors: Research has shown that driver population characteristics like age and gender are significant determinants of traffic crashes. Studies have concluded that younger drivers, typically under the age of 30 are the most likely group to be involved in a crash. Within the young driver grouping, teenage drivers have the highest odds of being involved in a crash, specifically a fatal crash. Drivers over the age of 65 were the second most likely age group to be involved in a crash. Therefore middle aged drivers are the least likely to be involved in a crash. Several studies have also showed that more male drivers are involved in fatal crashes than female drivers. Statistics compiled by the National Highway Transportation Safety Administration (NHTSA) over the years show that male drivers have a higher fatal crash rate than female drivers across all driver age groups. This is especially prominent in younger drivers under 25 years of age.

Socioeconomic Factors: The socioeconomic characteristics of driver households for example income, educational attainment, and poverty level have been supported by research to have a high correlation with fatal crash rates. Specifically, socioeconomic analysis on south eastern US states find that drivers living in areas with low household income are more likely to be involved in a single vehicle crash and vehicle crash fatality rates are much higher in lower income areas than in more affluent areas. A spatial analysis of fatal and injury crashes in Pennsylvania also showed that most socioeconomic variables, including poverty levels, were significant in analyzing crashes. Counties with a higher percentage of the population living under the poverty level were found to have significantly increased crash risk .

Other Factors: Other safety related studies aside from those investigating crash frequency, severity and type have also shown the significance of demographic and socioeconomic characteristics on outcomes. For example a study on impaired (especially drunk) driving by Edwardo concludes that age is a significant determinant on whether a driver involved in a crash will have been impaired or not. Drivers under the age of 25 were more likely to be impaired in a crash than any other age group. Results from seatbelt usage studies show that older drivers, women, Caucasians and individuals with higher incomes are more likely to use a seatbelt when driving . In addition, young drivers and male drivers are less likely to use a seatbelt.

Residential Nine-Digit Zip Code Data

A list of driver license numbers of drivers involved in crashes from 2007 to 2012 were extracted from South Carolina crash data and was provided to the SCDMV to procure locations where drivers live. To minimize privacy issues, a request for 9-digit zip codes was made rather than actual home addresses. The resolution of 9-digit zip is at the neighborhood level which was deemed sufficient for block group analysis. The resultant encrypted list of 9-digit zip codes provided by SCDMV was decoded and preprocessed. Figure 1 shows a sample of the drivers' license list and zip code data received from SCDMV. Arrows 'A' and 'B' show issues with the received SCDMV data



License Number		License Number	Zip Code
4763784	2	4763784	294642766
4763786	3	4763786	298037344
4763788	4	4763788	298513403
4763798	5	4763798	298013128
4763820	6	4763820	298032608
4763852	7	4763852	29816
4763853	8	4763853	298313566
4763891	9	4763891	298013108
4763895	10	4763895	298014149
4763900 (A)	11	4763900	
4763901	12	4763901	298293753
4763905	13	4763905	298019433
4763906	14	4763906	296935227
4763909	15	4763909	298019207
4763914 (B)	16	4763914	29851
4763925	17	4763925	298415603
4763927	18	4763927	298015016

Figure 1: SCDMV 9-Digit Zip Code Data

Figure 2 shows the geocoded 9-digit zip code locations for the whole state and a blow up at the street network level.

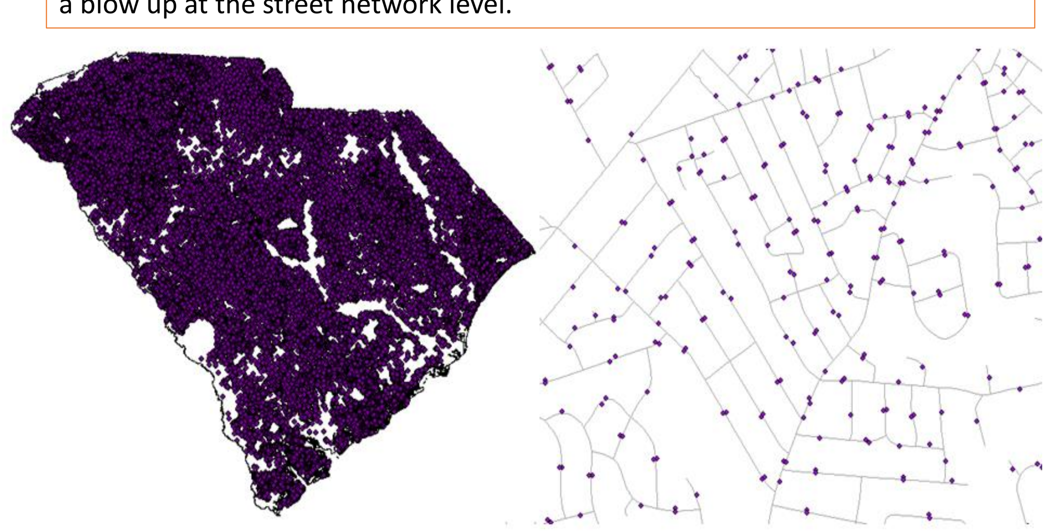


Figure 2: Geocoded 9-Digit Residential Zip Code Data

Spatial Analysis - *Block Group Analysis*

Drivers involved in crashes over the two analysis years were aggregated to South Carolina census block group boundaries using a spatial overlay procedure in the GIS. Initial grouping of census block groups was done using the number of drivers involved in fatal and injury crashes per 1000 population of driving age. Block groups were classified as Low-risk (Below 5 Risky-Drivers Per 1000 Driving Population), medium-risk (5 – 12) and High-risk (Above 12) using Jenks natural breaks optimization generated by ArcGIS. Figure 3 shows a thematic map of the 3 classes.

Category	Low Risk	Medium Risk	High Risk
Number of Block Groups	1138	1358	558
Total Area (Sq.Mi)	13149.1	12347.1	5341.2
Total Population	1775209	2131250	718905
Population Density (Pop/Sq.Mi)	135.0	172.6	134.6
Average Median HH Income ($)	47862.0	44421.7	36944.0
% of Individuals In Poverty	14.5	16.2	21.5
Enrollment % - HS and Above	12.1	10.3	10.6
Edu Attainment % - At least HS Diploma	54.0	54.0	51.0
Age 15-35 %	34.7	32.8	33.3
Age 35-65 %	48.3	50.1	50.3
Age Above 65 %	17.0	17.0	16.4
White %	71.2	66.4	53.2
Black %	23.5	27.1	41.2
Hispanic %	4.5	5.6	4.9
Asian %	0.1	1.5	0.8
Fatal/Injury Crashes Per 1000 Drivers	5.4	7.1	8.9

Table 1: Characteristics of 3 Risky-Driver Groupings

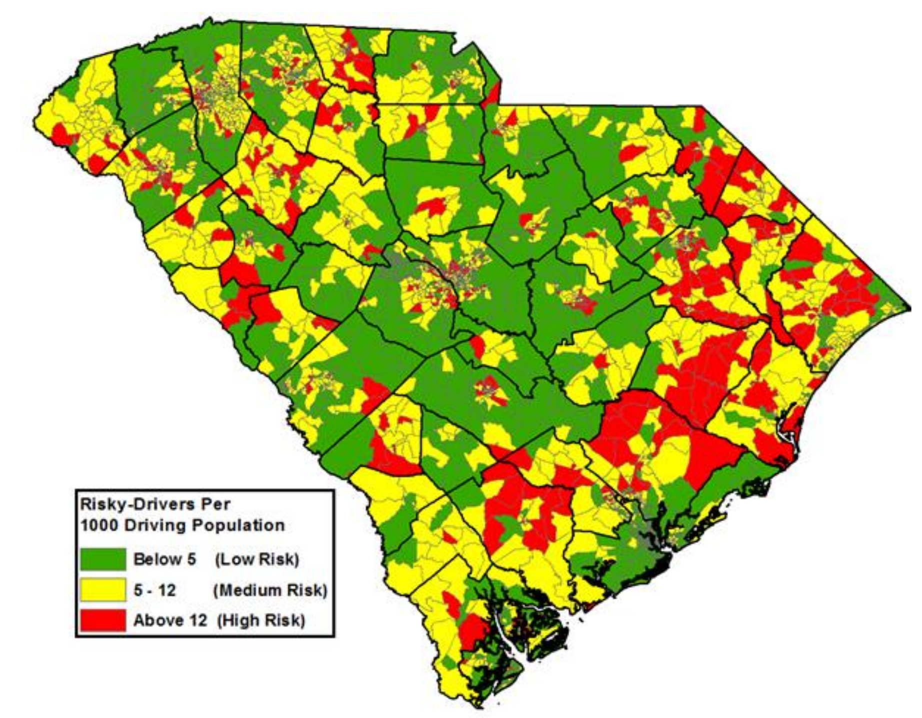


Figure 3: Risky-Drivers Per 1000 Driving Population

The initial grouping of block groups by risky-drivers per 1000 driving population began to show concentrations of high numbers of risky-drivers across the state (Figure 3). Concentrations of medium and high risk block groups were evident in the eastern and north western parts of the state. The summary demographic and socio-economic statistics for these risk groupings (Table 1) show a reduction in average median household income moving from low risk to the high risk groups. A similar but opposite trend is seen in the percent of people living in poverty, where more people live in poverty in the high risk grouping. Also, there was evidence of a positive correlation between risky-drivers and fatal and injury crashes. This trend suggests that risky-drivers are involved in more crashes closer to home.

Spatial Analysis - *Cluster Analysis*

Cluster analysis identifies and groups statistically significant high or low values of a variable or attribute in a dataset. Cluster analysis was done for several variables across the state. The variables with the most significant clustering patterns were 'risky-drivers per 1000 driving population', 'average median household income' and 'fatal/injury crashes per 1000 driving population'

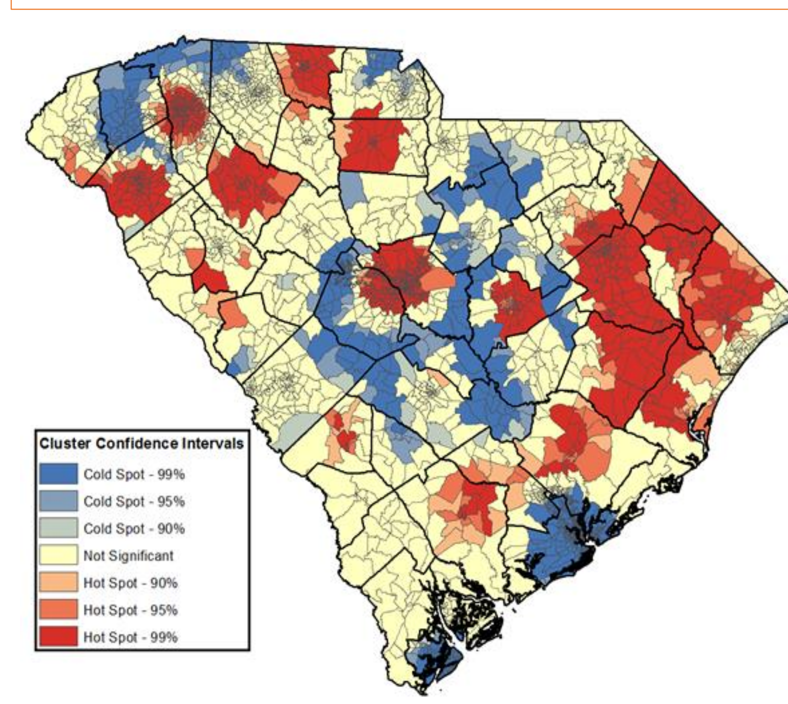


Figure 4: Risky-Driver Per 1000 Drivers

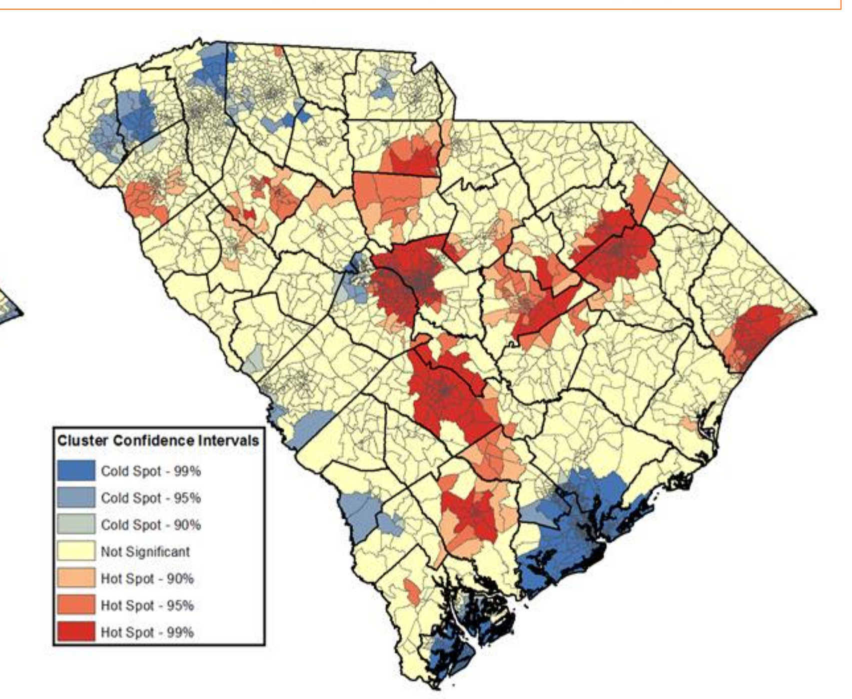


Figure 5: Average Median Household Income

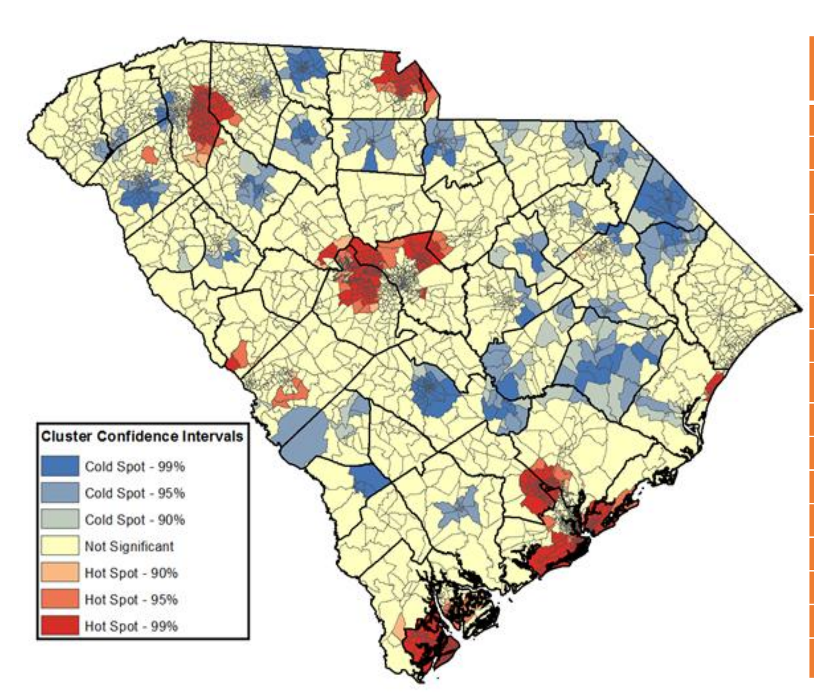


Figure 6: Fatal and Injury Crashes Per 1000 Drivers

Category	Hot Clusters	Cold Clusters	State
Number of Block Groups	1160	722	3054
Total Area (Sq. Mi)	8653.132	6350.46	30843.2
Total Population	1691860	1112150	4624463
Population Density (Pop/Sq.Mi)	195.5	175.1	149.9
Average Median HH Income ($)	42558	50535	44337
% of Individuals in Poverty	17.4	15.1	16.4
Enrollment % -HS and above	11.5	11.6	11.0
Edu Attainment % - At least HS Diploma	52.8	54.8	53.5
Age 15-35 %	35.1	34.6	33.6
Age 35-65 %	48.7	48.4	49.5
Age Above 65 %	16.2	17.0	16.9
White %	61.3	71.6	66.2
Black %	32.7	21.9	27.9
Hispanic %	4.8	5.9	5.1
Asian %	0.1	1.5	0.8
Crash Rate (Crash/1000 Drivers)	7.8	4.8	6.7

Table 2: Characteristics of Cluster Groupings

Spatial Analysis - *Cluster Analysis*

County	Rank	Hot Cluster BGs	County BGs	Area of Hot Clusters	County Area	Area %	Hot Cluster Pop	County Pop	Pop %
Dillon	1	28	28	406.5	406.5	100.0	32062	32062	100.0
Florence	2	105	107	751.7	803.4	93.6	133987	136885	97.9
Cherokee	3	39	41	350.4	397.2	88.2	50500	55342	91.3
Richland	4	224	245	377.0	771.4	48.9	343079	384504	89.2
Marion	5	28	31	323.7	494.0	65.5	29454	33062	89.1
Sumter	6	56	68	280.6	681.8	41.1	88058	107456	81.9
Laurens	7	47	58	534.9	723.7	73.9	51844	66537	77.9
Williamsburg	8	23	32	726.9	936.7	77.6	26301	34423	76.4
Colleton	9	20	30	456.4	1069.2	42.7	29193	38892	75.1
Georgetown	10	33	45	524.8	837.0	62.7	42812	60158	71.2

Table 3: County Ranking – Hot Risky-Drivers Clusters

The results of the cluster analysis of risky-drivers shown in Figure 3 identified spatial patterns and distributions of risky-drivers similar to the results of the initial block group analysis. However the cluster analysis produced more profound groupings of block groups that had relatively high numbers of risky-drivers and areas with a low number of risky-drivers. Summary descriptive statistics based on the cluster groups of risky-drivers per 1000 drivers are shown in Table 2

As part of the cluster analysis, an aggregation of hot and cold clusters of risky-drivers was done for counties within the state to better identify counties with areas that have a high concentration of risky-drivers. Tables 3 and 4 show the ranking of counties for hot and cold clusters respectively.

The county ranking of the percentage of the population living in hot clusters (Table 3) could serve as an initial screening of counties for safety analysis

County	Rank	Cold Cluster BGs	County BGs	Area of Cold Clusters	County Area	Area %	Cold Cluster Pop	County Pop	Pop %
Charleston	1	224	234	464.3	945.0	49.1	335868	350209	95.9
Kershaw	2	29	43	516.8	740.2	69.8	39594	61697	64.2
Pickens	3	46	72	406.5	512.1	79.4	70517	119224	59.1
Calhoun	4	6	12	270.4	392.3	68.9	8833	15175	58.2
Bamberg	5	8	17	118.7	395.4	30.0	8996	15987	56.3
Beaufort	6	65	111	228.0	580.7	39.3	83411	162233	51.4
York	7	53	109	192.8	695.9	27.7	112033	226073	49.6
Dorchester	8	23	67	87.3	575.6	15.2	57433	136555	42.1
Berkeley	9	42	100	97.6	1218.3	8.0	70113	177843	39.4
Lee	10	7	17	209.3	411.1	50.9	7359	19220	38.3

Table 4: County Ranking – Cold Risky-Drivers Clusters

Negative Binomial Analysis

Statistical analysis of some socioeconomic and demographic variables for hot cluster block groups was done to further understand the relationships between risky-drivers and the characteristics of their residential locations.

Model 2 - Dependent Variable: Number of Risky-Drivers			
Independent Variables	Estimate	P-Value	Significance
Intercept	2.49E+00	< 2e-16	0.001
Population Density	-4.57E-06	0.70096	Not Sig
Fatal/Injury Crashes	9.03E-03	0.000118	0.001
Median Household Income	-6.15E-06	1.09E-10	0.001
Population:15- 24	-3.55E-04	3.03E-05	0.001
Population:25- 34	1.10E-03	9.88E-06	0.001
Population:35- 44	-7.98E-04	0.05096	0.1
Population:45- 54	1.78E-03	0.000898	0.001
Population:55- 64	1.99E-03	2.12E-06	0.001

Table 6: Negative binomial Analysis – Model 2

The negative binomial regression analysis results were mostly consistent with the identified trends and relationships between risky-drivers and the socio-economic and demographic variables in the block and cluster analysis

Model 1 - Dependent Variable: Number of Risky-Drivers			
Independent Variables	Estimate	P-Value	Significance
Intercept	2.43E+00	< 2e-16	0.001
Population Density	-4.38E-05	0.000107	0.001
Fatal/Injury Crashes	8.37E-03	0.000378	0.001
Median Household Income	-1.23E-06	0.211808	Not Sig
White Population	3.39E-04	< 2e-16	0.001
Black Population	6.02E-04	< 2e-16	0.001
Hispanic Population	5.34E-04	0.006303	0.01

Table 5: Negative binomial Analysis – Model 1

Model 3 - Dependent Variable: Number of Risky-Drivers			
Independent Variables	Estimate	P-Value	Significance
Intercept	2.27E+00	< 2e-16	0.001
Population Density	5.96E-06	0.59676	Not Sig
Fatal/Injury Crashes	9.56E-03	3.04E-05	0.001
Median Household Income	-1.34E-06	0.243817	Not Sig
No High School Diploma	6.14E-04	0.000231	0.001
High School Diploma	1.16E-03	< 2e-16	0.001
College Degree and more	2.02E-04	0.019271	0.05

Table 7: Negative binomial Analysis – Model 3

Conclusion

The results of the analyses done in this research have spatially and statistically shown the relationship between risky-driver clusters and some socio-economic and demographic characteristics of these drivers. Whereas not all statistical correlations are causations, combining the spatial analysis with the statistical analysis provides a stronger argument with regard to the validity of the findings in this research. Of particular interest is the identification of hot clusters that have significantly higher proportions of high risk drivers than other areas. These locations should get highest consideration for targeted programs to educate drivers about driver safety. The SCDOT has sponsored a number of projects in recent years that focus on making roads safer. However the greatest contributor to a crash is the driver. The results of this research provide justification for state decision makers and officials to support safety programs and research that target drivers while providing a method for prioritizing areas of the state with greatest need from a high risk driver standpoint.

Acknowledgement

The authors would like to thank the South Carolina Department of Transportation (SCDOT) for providing data for this research and funding this research project