Missouri State Highway Patrol
Research and Development Division

Location Matters: A spatial and demographic analysis of Missouri State Highway Patrol driving while intoxicated stops.

SPECIAL REPORT

Missouri State Highway Patrol Statistical Analysis Center

December 2017

Introduction

Alcohol use contributed to more than 90,000 crashes in Missouri between 2005 and 2014. Fatal crashes constitute approximately 3,000 of those crashes with an average of 239 people killed each year. Interest groups, policy makers, and law enforcement agencies continually attempt to reduce drunk driving to save lives and prevent injuries. Informing the public on the dangers of drinking and driving and promoting stricter penalties for offenders may decrease alcohol related crashes. From a law enforcement perspective, it is imperative to determine where driving while intoxicated (DWI) stops frequently occur. Determining the location of the stops will enhance law enforcement's ability to properly allocate resources. This research specifically examines the location of Missouri State Highway Patrol (MSHP) DWI stops.

Defining the location of an incident can lend itself to different types of spatial and clustering analysis (Wieczorek & Hanson, 1997). For instance, Wieczorek and Naumov (2002) study the location of DWI offenders based on the offender's home address. They join the offender location to census blocks data and find significant spatial clustering of DWI offenders. They conclude that DWI offenders are not randomly displaced throughout the study area. Building off their study, Levine and Canter (2011) link DWI offender's home address with the location of their resulting crash. By connecting residence with crashes, the authors determine what streets are routinely used by potential DWI offenders. This information is essential for law enforcement personnel in terms of allocating resources to deter alcohol related crashes.

In addition, the location of an incident can be used in conjunction with other information in an area. Wieczorek and Hanson (1997) take DWI offenders' residences and join them to census tracts to create a rate, and find DWI rates are spatially dependent. Even though Brown, Sarasua, and Ogle (2016) did not specifically study DWI occurrences, they analyze fatal and injury crash locations in census blocks. They discover "high risk block groups had lower median incomes, higher levels of poverty, lower levels of educational attainment, and higher proportions of black population..."(p. 16). Furthermore, Levine and Canter (2011) aggregated census tracts to form traffic analysis zones and find that population, percentage of the population who are non-Hispanic White, and rural area lead to zones with more DWI stops.

Similarly, studies are using the location of other elements in association with DWI incidents. In connection with DWI arrests, research is analyzing the locations where people drink alcohol. For instance, Colon and Cutter (1983) find that the location of a drinking establishment is more important than determining the number of places to drink in an area. They contend if people must drive further to consume alcohol, there is a higher probability of getting into a crash as compared to walking to numerous bars located in a centralized district. However, Scribner, Mackinnon, and Dwyer (1992) find that alcohol outlet density is positively related to drinking and driving crashes. Gruenewald, Johnson, and Treno (2011) find similar results in a survey they conducted with people who consume alcohol and discover that greater restaurant densities correlate with more frequent drinking and driving behavior. Surprisingly, they find that DWI is unrelated to alcohol outlet density, but acknowledge this finding appears incomprehensible since alcohol outlet density is positively correlated to alcohol involved crashes.

The above studies suggest analyzing the relationship between drinking and driving in given geographic areas are situation specific. The specificity of drinking and driving in geographic areas should not be surprising since drinking and driving is an individual behavioral choice. For example, Carpenter (2004) finds that males are more likely to drink and drive than females, and younger people of both genders display different drinking and driving behaviors than their older counterparts. Additionally, "disproportionate numbers of alcohol impaired drivers have been found to be unmarried in their 20s and 30s," (Berger & Snortum, 1986, 139). Another possible factor can be employment. Karlsson, Hvitfeld, and Romelsjo (2000) do not find any significant difference between people employed in manual or non-manual labor in self-reported drinking and driving behavior. However, Green and Plant (2007) note that people employed in manual labor jobs consume more alcohol than those employed in non-manual labor, but do not indicate if this leads to more drinking and driving. Similarly, Romano, Peck, and Voas (2012) find that unemployed people are more likely to drive while impaired compared to employed people.

The literature indicates that location and demographic variables are important when studying driving and drinking. However, location studies imply that drinking and driving behavior can be area specific. Previous studies use states, cities, counties, census tracts, and census blocks with differing results. Additionally, the literature measures drinking and driving in different ways; driving while intoxicated arrest rates, alcohol related crashes, and surveys for self-reporting drinking and driving behavior. This study focuses on drinking and driving behavior in Missouri by analyzing DWI stops and how they correlate with the characteristics of a geographic area. The results may help law enforcement agencies allocate resources to areas that can have high volumes of impaired drivers.

Methods

This study uses a two-pronged approach to analyze Missouri State Highway Patrol's (MSHP) DWI stops. Both methods use the location of DWI stops and spatially join them to the American Community Survey (ACS) census tracts. After the spatial join, the methods for each approach differs. The first method utilizes population data to create a DWI stop rate for each census tract and then groups census tracts together to find similarities. Also, spatial autocorrelation is tested to visualize the impact of location of DWI stop rates. The second approach analyzes the actual count of DWI stops in each census tract to find correlation between the number of DWI stops and demographic data. Due to the distribution of the data, a zero-inflated negative binomial regression is performed to find statistical significance.

Unit of Analysis

Census Tracts: The United States Census Bureau distributes the ACS to collect population data. Per U.S. Census Bureau, the ACS "uses a series of monthly samples to produce annually updated estimates for the same small areas (census tracts and block groups) formerly surveyed via the decennial census long-form sample." The ACS provides a variety of data such as sex, occupation, income, etc... The census tracts are available geographically through TIGER/Line Shapefiles. This format allows the census data to be tied geographically with census tracts. The current study uses 2014 ACS five-year summary data, but there are some census tracts not utilized. First, 106 census tracts classified as St. Louis City are not used because MSHP did not allocate any manpower to St. Louis City, in the timeframe of this study. Second, an additional five census tracts in Jackson County, one in Platte County, and one in St. Charles County are excluded from the study because of missing population data. This selection process leaves 1,280 census tracts for observation.

Dependent Variable

Driving While Intoxicated (DWI) Stops: Anytime MSHP personnel make a traffic stop for driving while intoxicated, it is recorded in their daily log with a DWI tag. DWI stops performed by MSHP members during 2014 are collected for this study. All stops are given latitude and longitude coordinates enabling DWI stops to be geocoded in ArcGIS and then spatially joined to the 2014 ACS census tracts. This process gives each census tract a count of how many DWI stops occurred in it and allows rates to be calculated with the population data given by the ACS. A total of 7,341 DWI stops are geocoded and spatially joined to ACS census tracts.

Independent Variables

The following variables are provided by the 2014 ACS five-year summary, except for Missouri alcohol distributors, crashes, and the rural code. A description of the data follows.

Population: Total population of census tract. The count model uses the actual count of DWI stops in each census tract for analysis and; therefore, the actual population data instead of a rate. Levine and Canter (2011) find areas with larger populations have more DWI trips. Similarly, it is predicted that census tracts with high populations will equate to higher DWI stop counts in this study.

Male: Percentage of the population in each census tract who are male. Carpenter (2004) finds that males participate more in heavy drinking episodes than females, which could lead to more males being stopped for DWI. This study expects census tracts with higher percentage of males to have more DWI stops.

Age: Percentage of the population who are between the ages of 20 and 29 in each census tract. It is expected that a higher percentage of young people in the census tract will result in more DWI stops.

White: Percentage of the population who are white in each census tract. Brown, Sarasua, and Ogle (2016) find that high risk census blocks consisted of lower proportions of whites. Romano, Peck, and Voas (2012) find no significance in their final model between ethnicities and drinking and driving. However, the latter study is based on individual whereas the former study uses census blocks for their unit of analysis. Because the current study uses census blocks as the unit of analysis, it is anticipated that a higher percentage of whites will result in lower number of DWI stops.

Single: Percentage of population ages 15 years or older who are classified as either never married or divorced. Berger and Snortum (1986) did not include a marriage variable because it did not meet "the criterion for inclusion," (p 143). Valdez et al. (2007) finds that marriage can mediate substance abuse such as alcohol. Additionally, Gruenewald, Johnson, and Treno (2002) argue that single, divorced, or widowed people participate in drinking and driving more than their married counterparts. However, Wyse, Harding, and Morenoff (2014) contend marriage is a stressor that can lead to relapse and re-offending. However, this study expects the higher percentage of single individuals in the census tract, the more DWI stops are expected.

Education: Percentage of population age 18 years and older that have obtained only high school diploma or equivalent. Gruenewald, Johnson, and Treno (2002) contend that higher educated people drink more often, but this does not mean they drive while intoxicated. For instance, they could be drinking at home. Furthermore, Romano, Peck, and Voas (2012) argue that people with less educational attainment tend to drink and drive more than their counterparts. Because Romano, Peck, and Voas (2012) specifically, linked educational attainment with drinking and driving, it is expected that census tracts containing less educated people will have more DWI stops.

Occupation: Romano, Peck, and Voas (2012) analyzed the difference between employed and unemployed people, not specific type of jobs. Alternatively, Green and Plant (2007) examine specific types of jobs by comparing manual verses non-manual labor. They find manual labor employees tend to drink more than non-manual labor employees. This study uses a slightly different definition for occupation rather than the manual labor used in the previously mentioned studies. The current study uses occupations associated with blue-collar work. According to Merriam-Webster dictionary, blue-collar is "of, relating to, or constituting the class of wage earners whose duties call for the wearing of work clothes or protective clothing." The occupation categories available from the census data that fit this definition are natural resource, construction, maintenance, production, transportation, and material moving. The census only counts those 16 years of age and older for this variable. Following the same lines as Green and Plant (2007), it is expected that a higher percentage of blue-collar workers will result in more DWI stops in a census tract.

Income: The median income in the past 12 months in 2014 inflation adjusted dollars for the population 15 and over. Romano, Peck, and Voas (2012) argue that lower socioeconomic status results in more drinking and driving. This study expects a similar trend with lower median incomes resulting in more DWI stops.

Outlet: Numerous studies are now focusing on the location of places to drink alcohol (see Padilla & Morrissey, 1993; Gruenewald et al. 1996; Gruenewald, Johnson, & Treno 2002; Levine & Cantor, 2011). Missouri provides a list of all businesses and their geolocation that have a liquor license. Business with licenses that allow alcohol to be sold and consumed on their premises are included. These businesses were geocoded and aggregated up to the census tract level. A total of 12,192 distributors are mapped and joined to Missouri census tracts. It is expected that census tracts with higher amounts of alcohol distributors will have more DWI stops.

Crashes: Rookey (2012) uses alcohol involved crash fatality rate for a proxy of driving under the influence behavior and finds that counties with higher rates of alcohol involved crashes results in more DWI stops. This study will also use alcohol related crashes as a proxy for drinking and driving behavior. The state of Missouri collects detailed crash records. Included in these records is the geolocation of each alcohol involved crash. These crashes are geocoded, plotted, and spatially joined to the ACS census tracts. In 2013, there were 4,912 alcohol involved crashes that could be plotted. After spatially joining the crashes to the census tract, both the actual count of alcohol involved crashes and a rate of alcohol crashes per 1,000 persons was configured to each census tract. It is expected that a higher volume of alcohol related crashes will produce a higher volume of DWI stops.

Rural: Levine and Canter (2011) find that rural areas tend to have higher DWI crash occurrences than metropolitan areas. This finding may seem counterintuitive, but most motorists in rural areas must drive further than their metropolitan counterparts to consume alcohol at local establishments. In addition, people living in metropolitan areas have access to public transportation. The need to distinguish between rural and metropolitan areas is important when analyzing DWI stops. The United States Department of Agriculture maintains the Rural-Urban Continuum Codes (RUCC) for counties in the United States. Per the U.S. Department of Agriculture, the RUCC is "a classification scheme that distinguishes metropolitan counties by the population size of their metro area, and nonmetropolitan counties by degree of urbanization and adjacency to a metro area." Each county is given a value of one through nine, with one being completely urban and nine being completely rural. Table 1 displays the classification for each value and the number of Missouri counties in each category. Rookey (2012) uses the same RUCC in his DWI study to control for rural and metropolitan areas. Each census tract has a county code and is matched to the county's RUCC value. Unfortunately, this means that all census tracts in one county have the same RUCC value. Furthermore, the population values for the RUCC are based on the county level and therefore do not apply to the census tracts. However, the RUCC provides an efficient way to distinguish rural and urban areas than just having dichotomous variables. The previous literature suggests that rural areas will have more DWI stops and the same is expected in this study.

Table 1: Rural-Urban Continuum Codes and Descriptions

RUCC Value	***Description***
1	Metro - Counties in metro areas of 1 million population or more
2	Metro - Counties in metro areas of 250,000 to 1 million population
3	Metro - Counties in metro areas of fewer than 250,000 population
4	Nonmetro - Urban population of 20,000 or more, adjacent to a metro area
5	Nonmetro - Urban population of 20,000 or more, not adjacent to a metro area
6	Nonmetro - Urban population of 2,500 to 19,999, adjacent to a metro area
7	Nonmetro - Urban population of 2,500 to 19,999, not adjacent to a metro area
8	Nonmetro - Completely rural or less than 2,500 urban population, adjacent to a metro area
9	Nonmetro - Completely rural or less than 2,500 urban population, not adjacent to a metro area

Analysis

Part 1 - Location of DWI Stops

Figure 1 below illustrates the distribution of the DWI stop rate is extremely skewed to the left, which indicates it is a relatively rare event.

Figure 1: 2014 MSHP DWI Stop Rate Distribution among 2014 ACS Census Tracts (N=1,280)

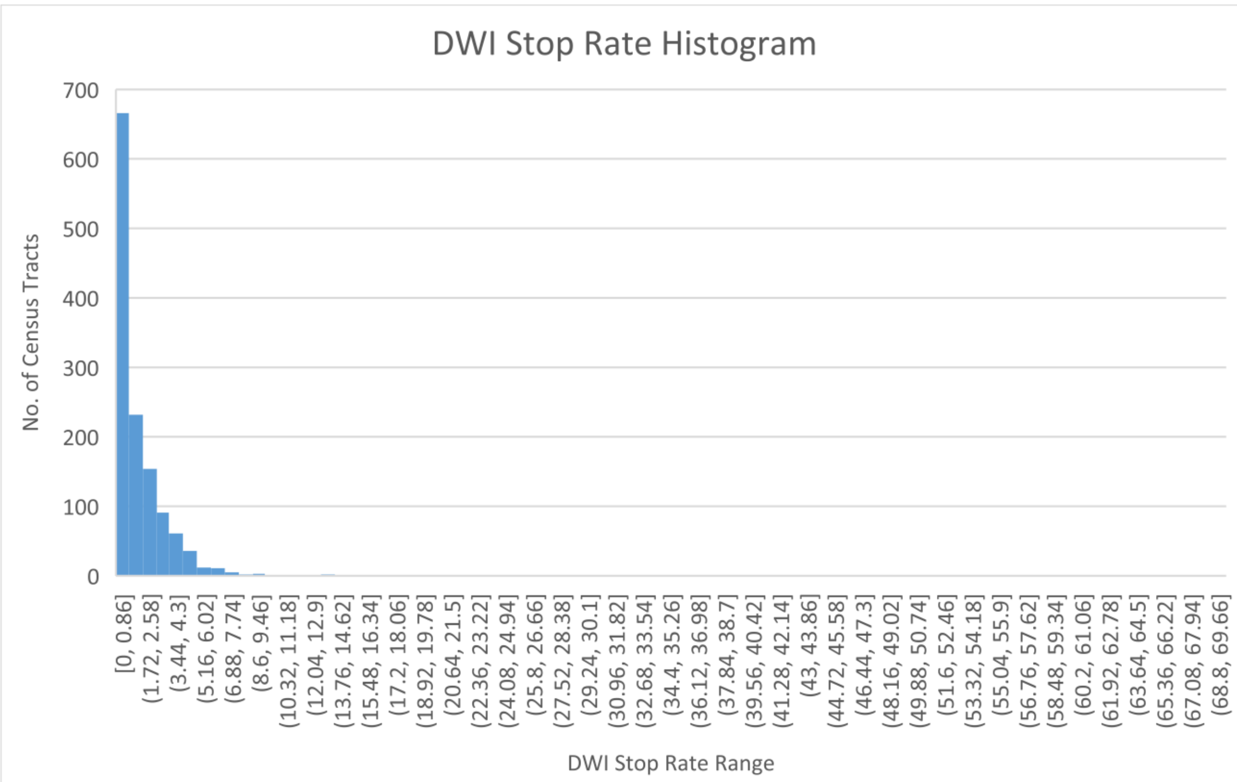
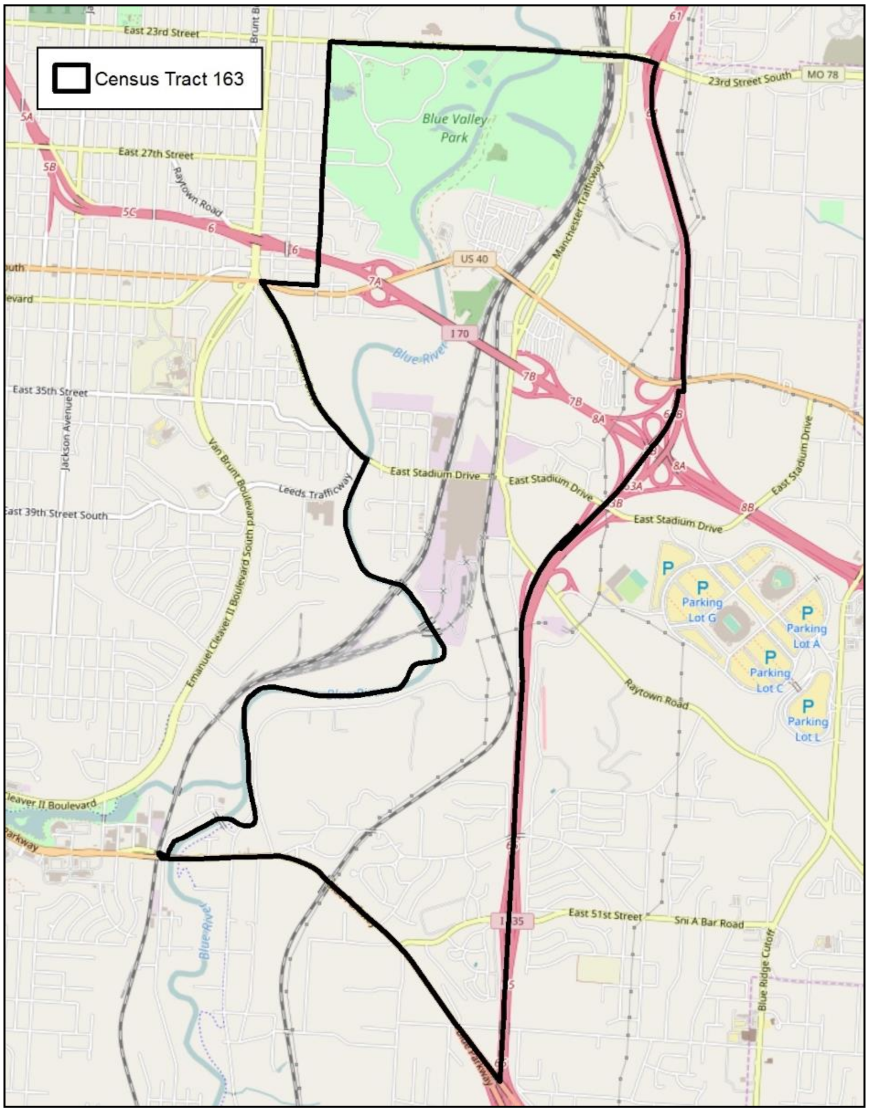


Figure 1 also indicates many census tracts with a DWI stop rate between zero and 0.86 per 1,000 persons. As the DWI stop rate increases, the number of census tracts decreases. This distribution makes it difficult to compare the data associated with other census tracts. Figure 1 does not include three outliers due to the extreme range on the y-axis. The three outliers are examined in further detail below:

Census Tract 163

The first census tract outlier is shown in Figure 2. This census tract has a DWI stop rate of 18.98 per 1,000 persons, an alcohol outlet rate of 3.16 per 1,000 persons, and an alcohol involved crash rate of 2.37 per 1,000 persons in an area consisting of 4.25 square miles. Interestingly, the population is only a few hundred more compared to the mean of all the census tracts, but has a high DWI stop rate. Inspecting the context and surrounding area of this census tract helps determine reasons for a higher DWI stop rate. Figure 2 illustrates major interstates such as I-70 and I-435 running through the census tract. These high-volume interstates provide more opportunities for DWI stops. Additionally, this census tract is located just west of Arrowhead and Kauffman Stadiums. Nearly three million people visit these stadiums each year. In recent years, stadiums have strongly promoted responsible drinking and the usage of designated drivers, but it is assumed people still get in their vehicles intoxicated after the game.

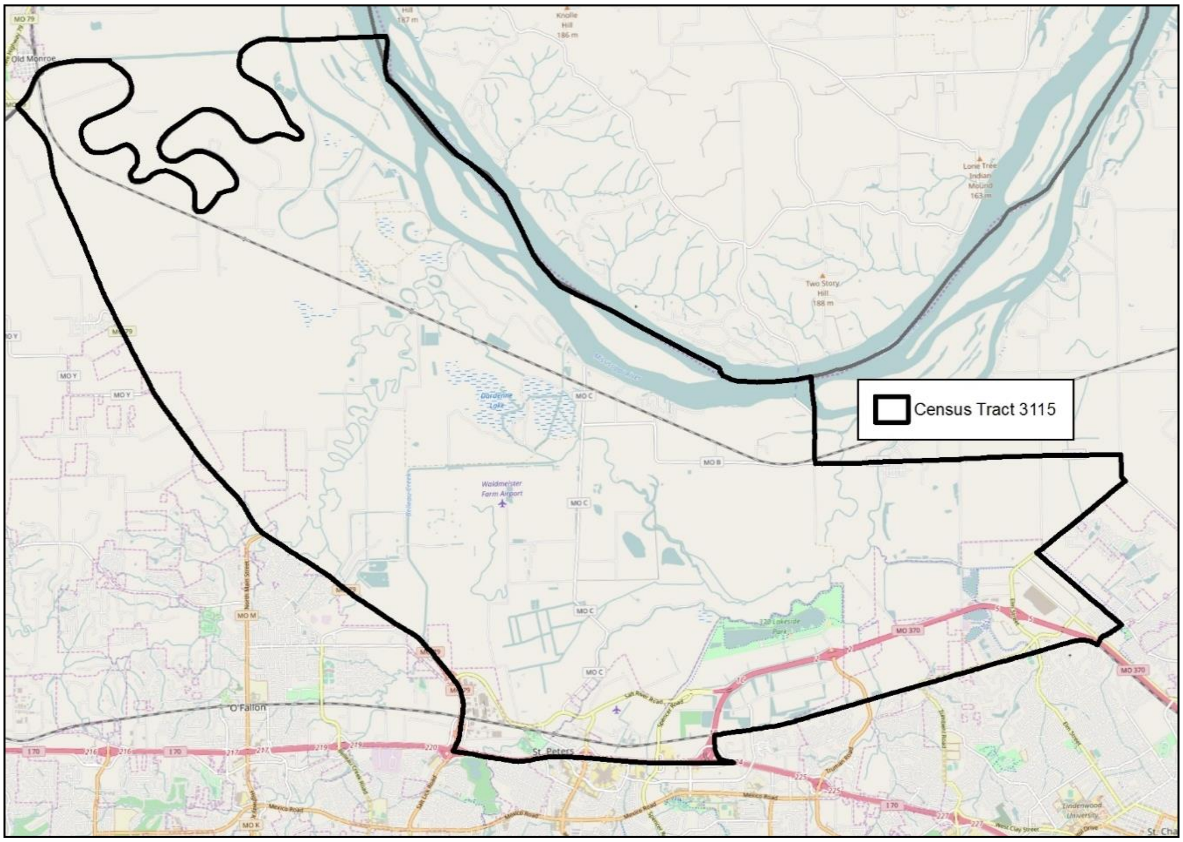
Figure 2: A picture of Census Tract 163 generated from ESRI'S ArcMap 10.4.1



Census Tract 3115

The second outlier is located just outside the St. Louis metropolitan area and is shown in Figure 3. This census tract has a much larger area compared to census tract 163 with 61.23 square miles. The population is just over 1,900 with an alcohol outlet rate of 14.58, an alcohol involved crash rate of 8.33, and a DWI stop rate of 27.07 per 1,000 persons. This census tract contains I-70 and MO-370 ensuring a high-volume traffic area. Additionally, about half of the city of St. Peters is located within this tract, which may contribute to the higher rate of alcohol outlets compared to other census tracts. Finally Figure 3 displays a wide swath of land proving ample opportunity for DWI stops even though the population may not be as large as other census tracts.

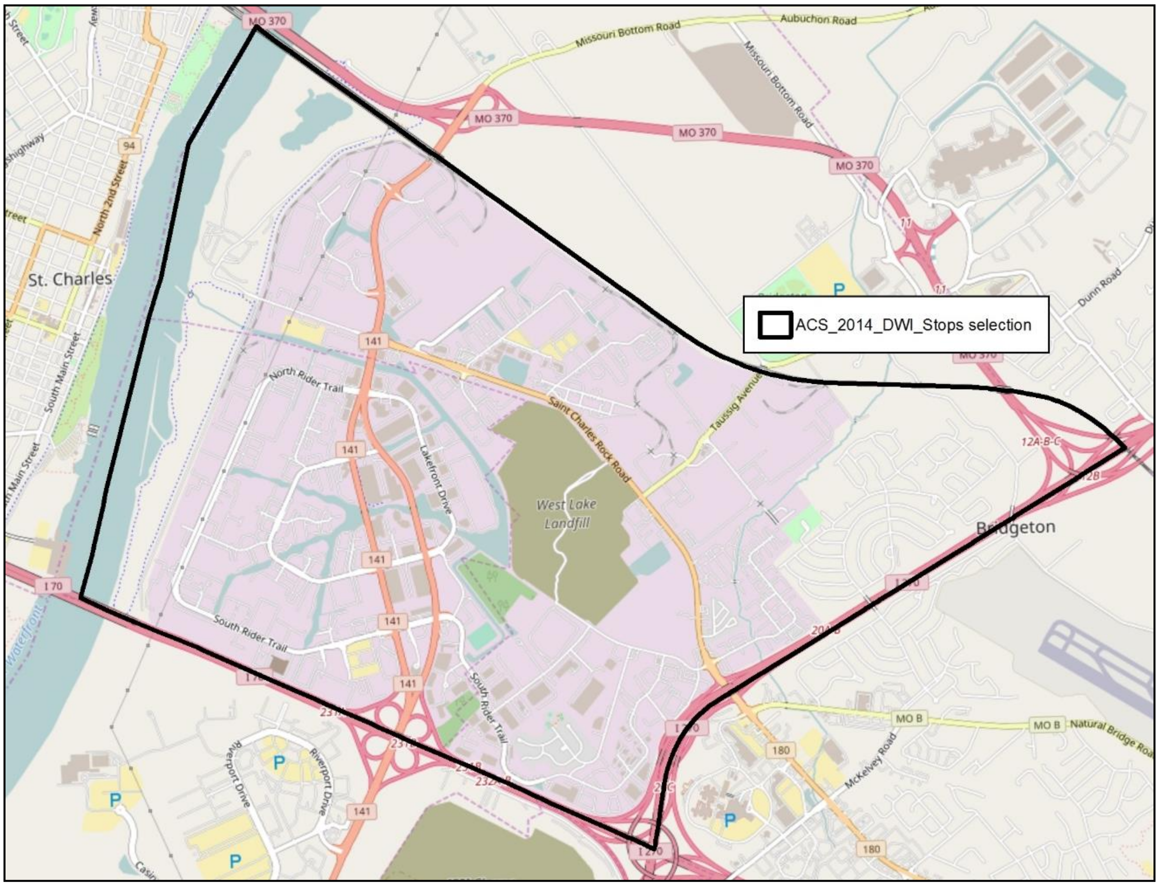
Figure 3: Census Tract 3115 generated from ESRI's ArcMap 10.4.1



Census Tract 2131.02

Figure 4 displays Census Tract 2131.02, the smallest of any tract included in this study with an area of only 6.68 square miles. Yet, the DWI rate is 69.54. There are 18 alcohol distributors but only 2 alcohol involved crashes in this census tract. With the small population, the rates per 1,000 persons are overly inflated. Investigating the tract further reveals an area that is mostly industrial with a landfill in the middle of it. Furthermore, this census tract is located in between St. Louis and St. Charles with many major roadways running through it such as I-70 and I-270. Moreover, the Ameristar Casino is directly west of this census tract across the river. This proximity to the casino presents ample opportunity for DWI stops.

Figure 4: Census Tract 2131.02 generated from ESRI's ArcMap 10.4.1



To make meaningful comparisons, census tracks are grouped based on similar DWI stop rates. First, the outliers and DWI rate of zero census tracts are separated from the analysis. Even after separating these census tracts the DWI stop rate still ranges from 0.10 to 13.12 per 1,000 persons with a standard deviation of 1.66 indicating a degree of variance within the data.

ArcGIS provides multiple ways to group data together such as Natural Breaks (Jenks). Per ESRI, "Natural breaks classes are based on natural groupings inherent in the data. Class breaks are identified that best group similar values and that maximize the differences between classes." Furthermore, natural breaks "minimize[s] within-class variation and maximize[s] between-class variation in an iterative series of calculations" (Brewer & Pickle, 670). Using the natural breaks method helps group similar census tracts based on the DWI stop rate. Figure 5 shows a choropleth of Missouri census tracts based on DWI stop rate using natural breaks.

Figure 5: Choropleth of Missouri Census Tracts based on DWI Stop rate using Natural Breaks (Jenks) Classification

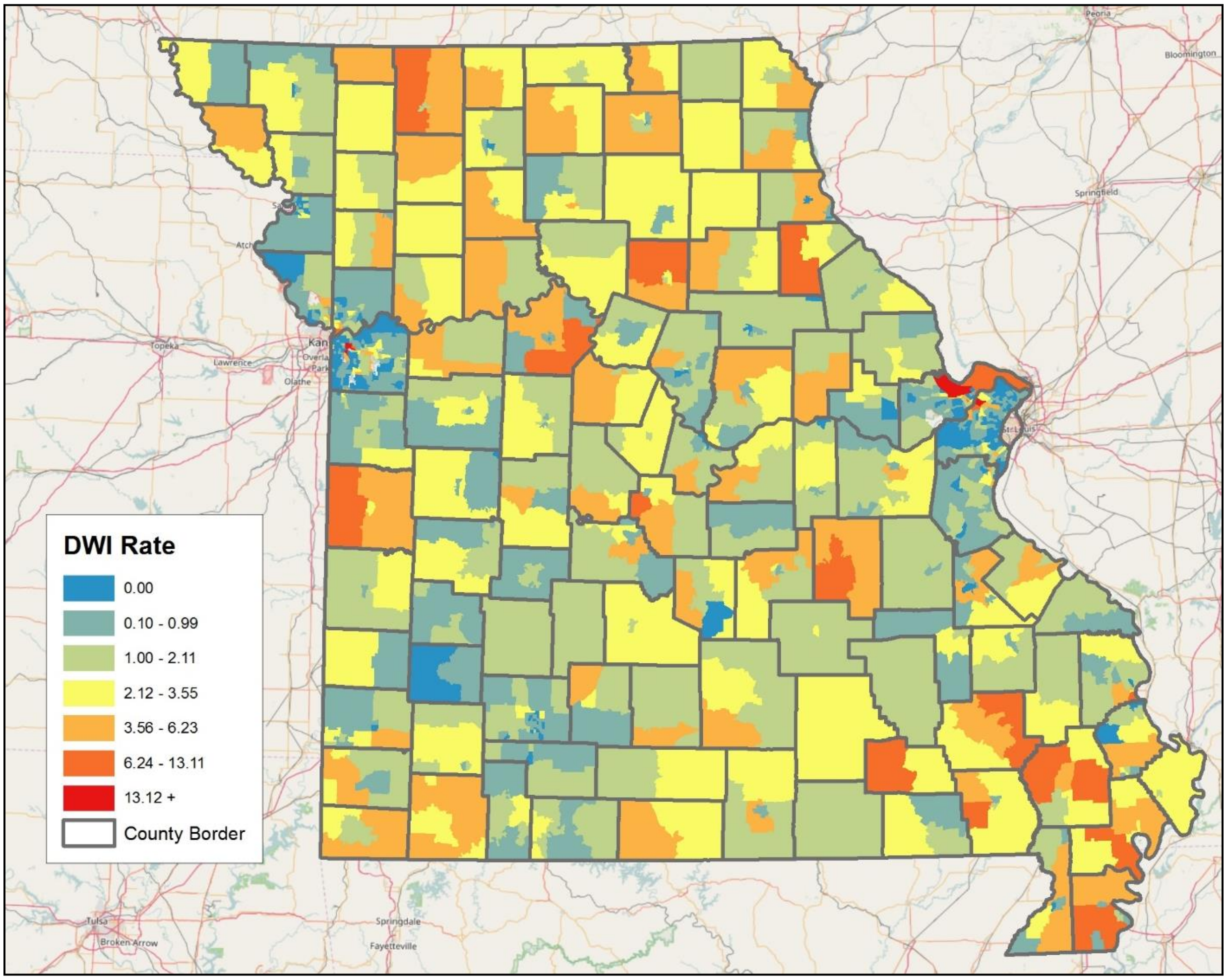


Table 2: Descriptive Statistics of Census Tracts grouped by DWI Stop Rate

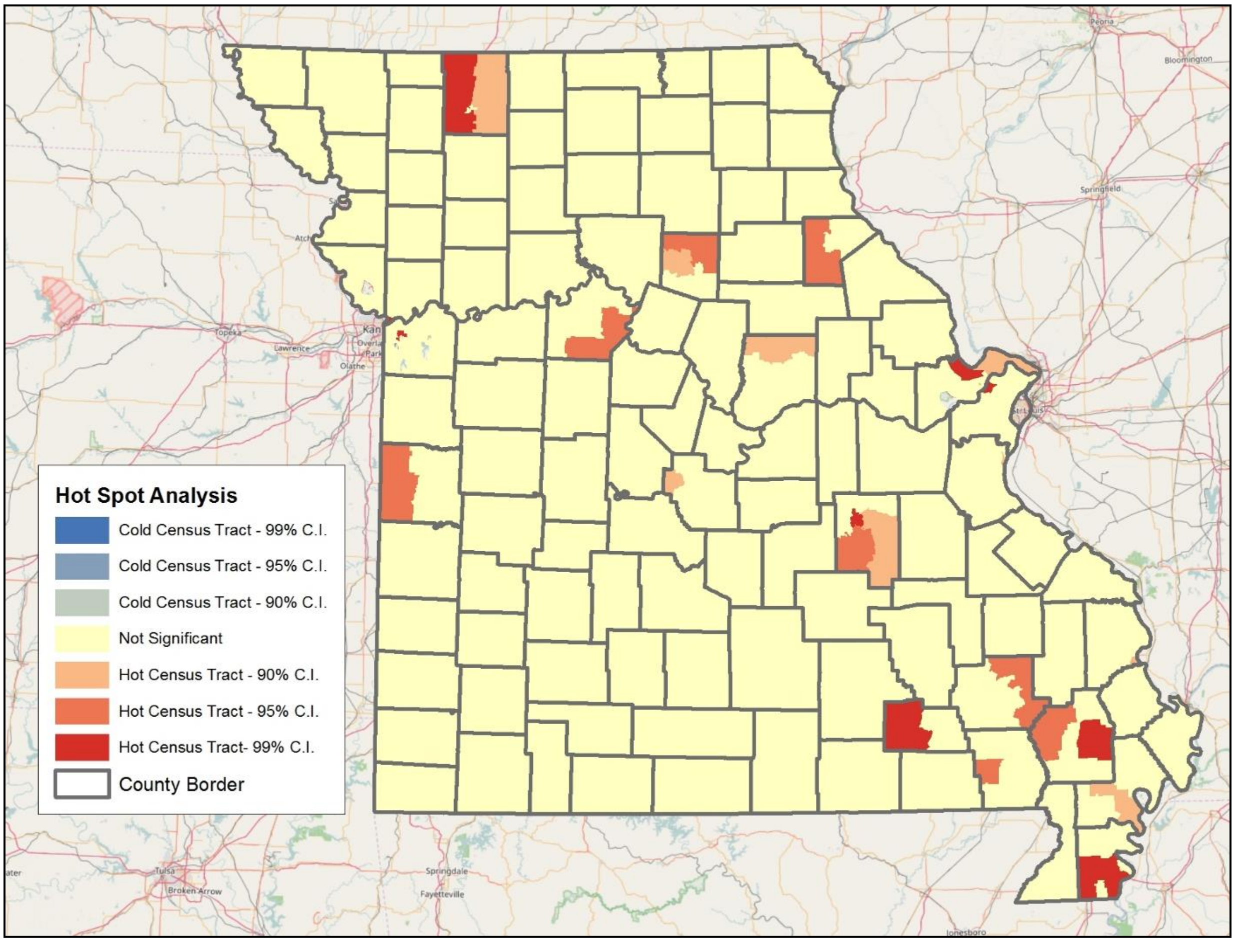
DWI RATE RANGE (PER 1,000 PERSONS)	0	0.10 - 0.99	1.00 - 2.11	2.12 - 3.55	3.59 - 6.23	6.35 - 13.11
NUMBER OF CENSUS TRACTS	327	386	264	177	101	22
TOTAL AREA (SQ. MI)	1,342.3	9,979.8	22,685.3	20,187.1	12,039.9	3,237.9
TOTAL POPULATION	1,318,685	1,922,310	1,240,306	733,363	414,368	75,516
POPULATION DENSITY (POP/SQ.MI)	982.4	192.6	54.7	36.3	34.4	23.3
AVERAGE MEDIAN INCOME ($)	29,136.22	26,902.52	24,591.30	23,028.72	24,564.82	21,700.77
EDU ATTAINMENT - HIGH SCHOOL %	0.35	0.41	0.49	0.52	0.52	0.57
OCCUPATION: BLUE-COLLAR WORKERS %	0.16	0.21	0.26	0.28	0.29	0.29
MARITAL STATUS (NEVER MARRIED/DIVORCED) %	0.45	0.41	0.38	0.38	0.36	0.39
MALE %	0.48	0.49	0.50	0.49	0.50	0.50
AGE 20 - 29 %	0.14	0.14	0.13	0.13	0.12	0.10
WHITE %	0.74	0.84	0.91	0.92	0.92	0.91
ALCOHOL DISTRIBUTORS RATE (PER 1,000 PERSONS)	1.8	1.9	2.1	2.7	2.8	3.3
2013 ALCOHOL INVOLVED CRASH RATE (PER 1,000 PERSONS)	0.6	0.7	0.9	1.2	1.3	1.9
DRIVING WHILE INTOXICATED RATE (PER 1,0000 PERSONS)	0.0	0.5	1.5	2.7	4.4	8.1

Several data trends emerge when viewing Table 2. First, the DWI stop rate increases as the total population declines, which is probably attributed to the number of census tracts decreasing in each category. Second, census tracts with lower median incomes have a higher DWI stop rate, except for the tracts with a DWI stop rate range of 3.11 through 6.23 per 1,000 persons. Third, census tracks that contain more people with an education of less than a high school diploma and those census tracts with more people employed in blue-collar jobs have a higher DWI stop rates. Fourth, tracks with a higher percentage males or whites also have higher DWI stop rates. Fifth, census tracts with increased rates of both alcohol distributors and alcohol involved crashes results in tracts having a higher the DWI stop rate. Finally, census tracks with individuals classified as single and people ages 20 through 29 have higher DWI rates.

While these trends are interesting, causation and even correlation cannot be determined. However, it is still possible to determine if the census tracts are spatially correlated. "Spatial autocorrelation may be loosely defined as the property of random variables taking values at pairs of locations a certain distance apart, that are more similar (positive autocorrelation) or less similar (negative correlation) than expected for randomly associated pairs of observations," (Legendre, 1993, 1659). ESRI suggests using Moran's *I* to test for spatial autocorrelation. Moran's I is "a single test statistic that indicates two types of spatial autocorrelation - positive autocorrelation and negative autocorrelation," (Zhang & Lin, 2007, 6123). Performing Moran's I shows a positive autocorrelation indicating the spatial distribution of the census tract DWI stop rates is not random. (See Appendix A for Moran's I results.)

Since positive autocorrelation is detected, hot spot and clustering analysis is performed. The hot spot analysis tool provided by ESRI is the Getis-Ord Gi* and will assist in explaining why census tracts have an unexpectedly high or low DWI rate. Per ESRI, the hot spot analysis tool "works by looking at each feature within the context of neighboring features...To be a statistically significant hot spot, a feature will have a high value and be surrounded by other features with high values as well." Figure 6 displays the results of the Getis-Ord Gi* for the census tracts based on DWI stop rate.

Figure 6: Results of ArcGIS's Getis-Ord Gi tool based on Census Tracts DWI Stop Rate*



The Getis-Ord Gi* analysis does not indicate much clustering of "hot" census tracts, only 29 to be exact. Some of the "hot" tracts are in areas that are expected, Kansas City, St. Louis County, and St. Charles County. Still, there are some surprising hot areas such as Harrison County, Crawford County, along with other census tracts peppered throughout southeast Missouri. Looking specifically at the three "hottest" census tracts in southeast Missouri, renders little similarities other than population and DWI stop rates. The alcohol outlet rate ranges from less than one to almost five per 1,000 persons, and the alcohol involved crashes from 2013 range from 1.8 to 4.3 per 1,000 persons. Table 3 gives the descriptive statistics for all "hot" census tracts observing the standard deviation for these variables indicates wide variation in alcohol distributors rate, total population, and the area of the census tracts.

Table 3: Descriptive Statistics for all "hot" census tracts

	TOTAL	MEAN	ST DEV	MIN	MAX
NUMBER OF CENSUS TRACTS	29	-	-	-	-
TOTAL AREA (SQ. MI)	4,007.28	138.18	120.71	0.64	378.86
TOTAL POPULATION	97,872	3,374.90	1,124.31	302	5,784
POPULATION DENSITY (POP/SQ. MI)	24.42	367.92	781.20	9.74	3423.69
MEDIAN INCOME	-	$21,523.31	$3,724.65	$13,850.00	$31,259.00
EDU ATTAINMENT - HIGH SCHOOL %	0.58	0.58	0.10	0.30	0.76
BLUE COLLAR WORKERS %	0.29	0.29	0.09	0.11	0.45
MARITAL STATUS (NEVER MARRIED/DIVORCED) %	0.40	0.41	0.11	0.28	0.64
MALE %	0.49	0.49	0.04	0.39	0.57
AGE 20 - 29 %	0.11	0.12	0.05	0.05	0.28
WHITE %	0.91	0.89	0.14	0.41	0.99
ALCOHOL DISTRIBUTORS RATE (PER 1,000 PERSONS)	3.70	5.76	10.63	0.69	59.60
2013 ALCOHOL INVOLVED CRASH RATE (PER 1,000 PERSONS)	1.99	2.36	1.83	0.26	8.33
DRIVING WHILE INTOXICATED RATE (PER 1,0000 PERSONS)	8.51	10.91	11.94	5.96	69.54

Part 2 - Statistical Testing

The previous section explores the spatial correlation for the census tracts' DWI stop rate. Spatial concepts and techniques are becoming increasingly popular to use in social sciences (Logan et. al., 2010) Despite the increase in usage there is growing debate on how to include spatial analysis in the social sciences. Including spatial autocorrelation can help reduce model misspecification (Getis, 2008). Furthermore, "whereas correlation statistics were designed to show relationships between variables, autocorrelation statistics are designed to show correlations within variables, and spatial autocorrelation shows the correlation within variables across space," (Getis, 2007, 493). The below section analyzes the DWI stops using traditional statistical methods.

The distribution of the data determines what statistical methods should be used to find correlation and significance between variables. The DWI stop rate distribution shown in Figure 1 suggests a count model approach. Count models count the occurrences of events. Therefore, instead of using rates, this section uses the actual number of DWI stops in each census tract as the dependent variable. Figure 7 shows the distribution of MSHP DWI stops, which is similar to the distribution of the DWI stop rates. Additionally, the actual number of alcohol outlets, and alcohol involved crashes for each census tract are used as independent variables, rather than the rates used in part one. The total population for each census tract is included as a control variable

Figure 7: 2014 MSHP DWI Stops Distribution among 2014 ACS Census Tracts (N=1,280)

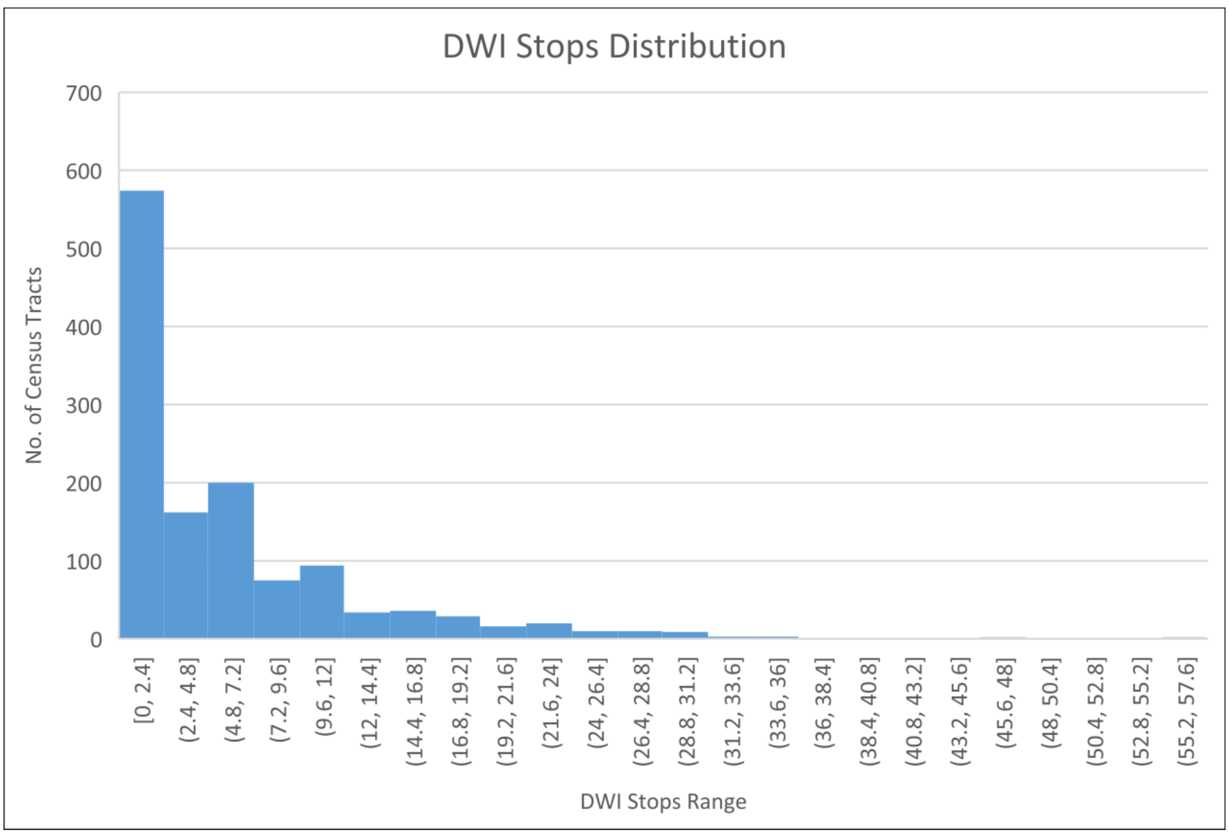


Table 4: Descriptive Statistics for Part 2

Variable	Mean	St. Dev	Min	Max
DWI Stops	5.75	7.26	0.00	56.00
Total Population	4,460.39	1,823.25	302.00	12,743.00
Male (%)	0.49	0.40	0.29	0.75
Age (%)	0.14	0.07	0.03	0.72
White (%)	0.83	0.22	0.01	1.00
Single (%)	0.41	0.13	0.17	0.99
Education (%)	0.45	0.16	0.05	0.83
Occupation (%)	0.24	0.10	0.01	0.53
Median Income ($)	26,167.79	9,276.30	2,724.00	78,625.00
Alcohol Distributors	9.47	8.84	0.00	74.00
Crashes	3.76	3.22	0.00	23.00
Rural Code	2.97	2.53	1.00	9.00

The proper count model must be determined and there are several to choose from. The Poisson Regression Model (PRM) is the first choice, but the PRM "rarely fits due to over dispersion," (Long & Freese, 2006, 372). After the PRM, the Negative Binomial Regression Model (NBRM) is considered. Both the PRM and NBRM are "the foundation for other count models," (Long & Freese, 349). Other count models help when the data consists of many zero counts, such as the current data. These models are zero-inflated count models and assume two reasons for zero counts, no instances occurred or there can be no occurrences due to certain restrictions in the data. Fortunately, STATA helps identify which count model fits the data best using the countfit command. The countfit command runs through myriad tests for count models, one of which is comparing mean probabilities. Figure 8 displays the under and over prediction of the count models when comparing mean probabilities. The best model will be at zero, no over or under predictions, for each count value.

Figure 8: Comparisons of count models using STATA 13 countfit command

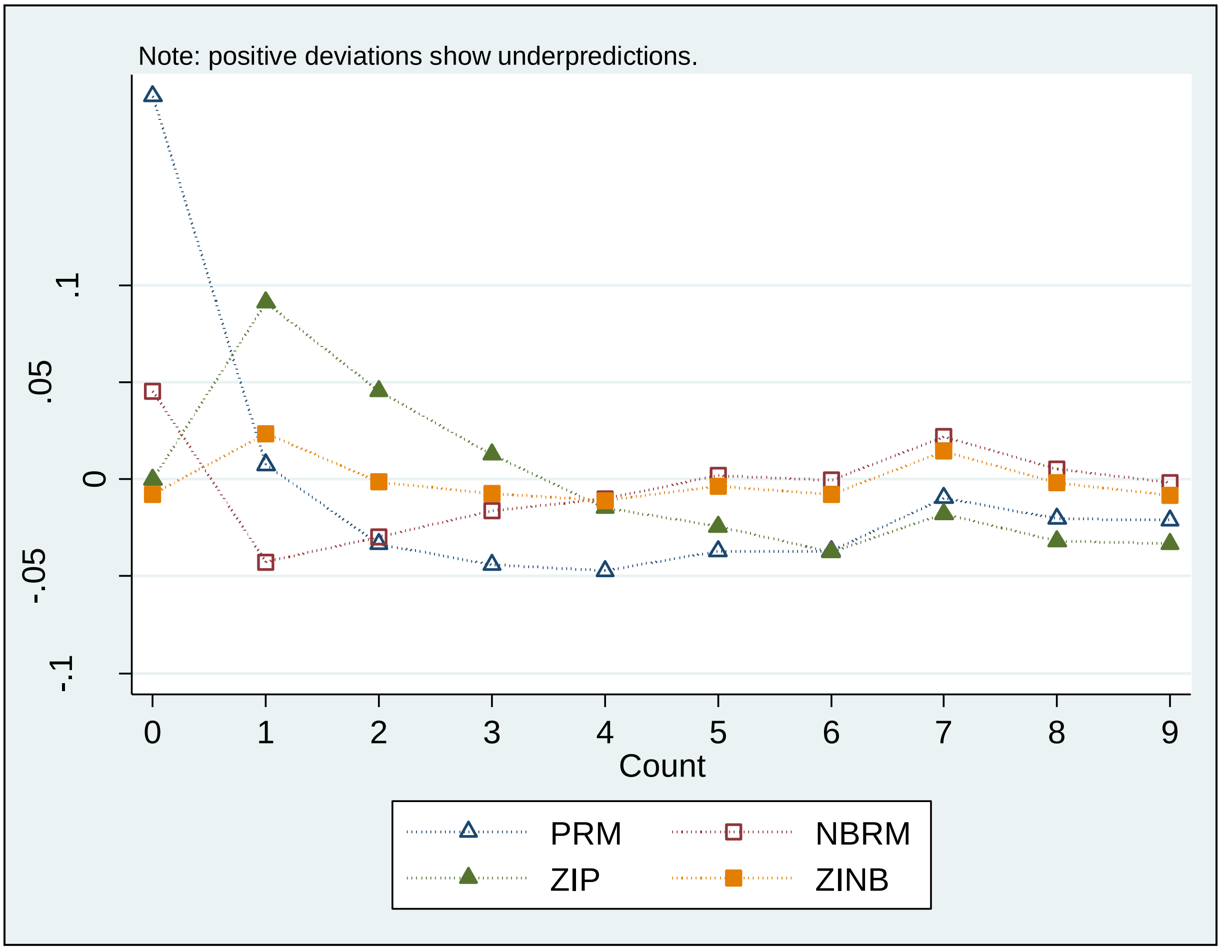


Figure 8 indicates the Zero-Inflated Negative Binomial Model (ZINB) producing the most accurate model compared to the others. The ZINB slightly over predicts and under predicts counts, but not to the severity of the other models such as the Zero-Inflated Poisson (ZIP). The ZINB handles zero values better than the other models and can be explained theoretically. The PRM and NBRM assume that every census tract has the potential for at least one DWI stop. The ZINB model relaxes this assumption. Remember the DWI stop is predicated on where MSHP members are patrolling. While the MSHP's jurisdiction includes the entire state, metro areas are handled primarily by local agencies. For example, census tracts in the Kansas City area that do not have a state road or interstate have a much higher probability of being worked by the Kansas City Police Department. Additionally, MSHP members patrol interstates, state roads, and county roads, but do not typically patrol city streets. Thus, it is possible that some census tracts do not have any MSHP DWI stops due to patrol patterns. The ZINB splits the data into two groups, an always zero group and a not always zero group. A census tract in the always zero group has an outcome of zero with a probability of one, whereas a census tract in the not always zero group might have a zero count but there is a probability greater than zero for the census tract to have a positive count of DWI stops (Long & Freese, 2006). The results of the ZINB regression are shown in Appendix B. Table 5 shows an interpretation the results.

Table 5: Results of Zero Inflated Negative Binomial Model with Percent Interpretations

Count Equation: Percentage Change in Expected Count for Those Not Always 0

DWI Count	b	z	P>|z|	%	%StdX	SDofX
Total Population	0.000	3.546	0.000**	0.000	11.900	1823.253
Male	-0.346	-0.428	0.669	-29.200	-1.400	0.040
Age	0.747	0.958	0.338	111.000	5.400	0.071
White	0.903	3.143	0.002**	146.700	22.200	0.222
Single	-0.267	-0.484	0.628	-23.400	-3.400	0.131
Education	1.794	3.767	0.000**	501.100	33.100	0.160
Occupation	0.192	0.342	0.732	21.100	1.900	0.100
Income	0.000	1.612	0.107	0.000	12.000	9276.302
Distributors	0.016	3.934	0.000**	1.600	15.400	8.845
Crashes	0.088	9.247	0.000**	9.200	32.900	3.215
Rural	0.030	1.950	0.051	3.000	7.800	2.527

Binary Equation: Factor Change in Odds of Always 0

Always 0	b	z	P>|z|	%	%StdX	SDofX
Total Population	0.000	-2.405	0.016*	0.000	-26.100	1823.253
Male	6.220	2.062	0.039*	50180.700	28.500	0.040
Age	-0.877	-0.331	0.740	-58.400	-6.000	0.071
White	-0.478	-0.631	0.528	-38.000	-10.100	0.222
Single	5.045	2.495	0.013*	15417.300	93.300	0.131
Education	-0.821	-0.457	0.648	-56.000	-12.300	0.160
Occupation	-5.607	-2.128	0.033*	-99.600	-43.000	0.100
Income	0.000	1.224	0.221	0.000	29.500	9276.302
Distributors	-0.036	-2.325	0.020*	-3.600	-27.500	8.845
Crashes	-3.672	0.000	-62.900	-17.300	-45.600	3.215
Rural	-0.599	-3.229	0.001**	-45.000	-78.000	2.527

*p < 0.05, **p< 0.01

The significant variables in both equations are total population, the number of alcohol outlets, and the number of 2013 alcohol involved crashes. Despite its significance, the population has little impact on the occurrences of DWI stops because the coefficient is extremely small. For example, the variable's interpretation is for a one unit increase in the population the number of DWI stops increases by 0%. A one standard deviation increase in population increases the amount of DWI stops by about 12%.

Contrasting with population, the number of alcohol outlets and the number of alcohol involved crashes from 2013 can have a larger impact on the number of DWI stops within a census tract. Among those census tracts with a higher opportunity of DWI stops, each additional alcohol outlet or alcohol involved crash increases the number of DWI stops by 1.6% and 9.2%, respectively. For those census tracts that have a low propensity of MSHP DWI stops, for each additional alcohol outlet or alcohol involved crash, the odds of not having any DWI stops decreases by 3.6% and 17.3%, respectively. Figure 9 displays the probability of a census tract with zero DWI stops given a certain amount of alcohol outlets. The line labeled "0s from Both Equations" gives the probability of a census tract having no DWI stops when combining both the binary equation and the count equation from the ZINB regression. As the figure shows, with zero alcohol outlets, the probability of a census tract having zero DWI stops is only about 20%. As the amount of alcohol outlets increases, the probability of a census tract having zero DWI stops decreases to well below 10%.

Figure 9: Probability of 0 DWI stops in a Census Tract based on the number of Alcohol Distributors

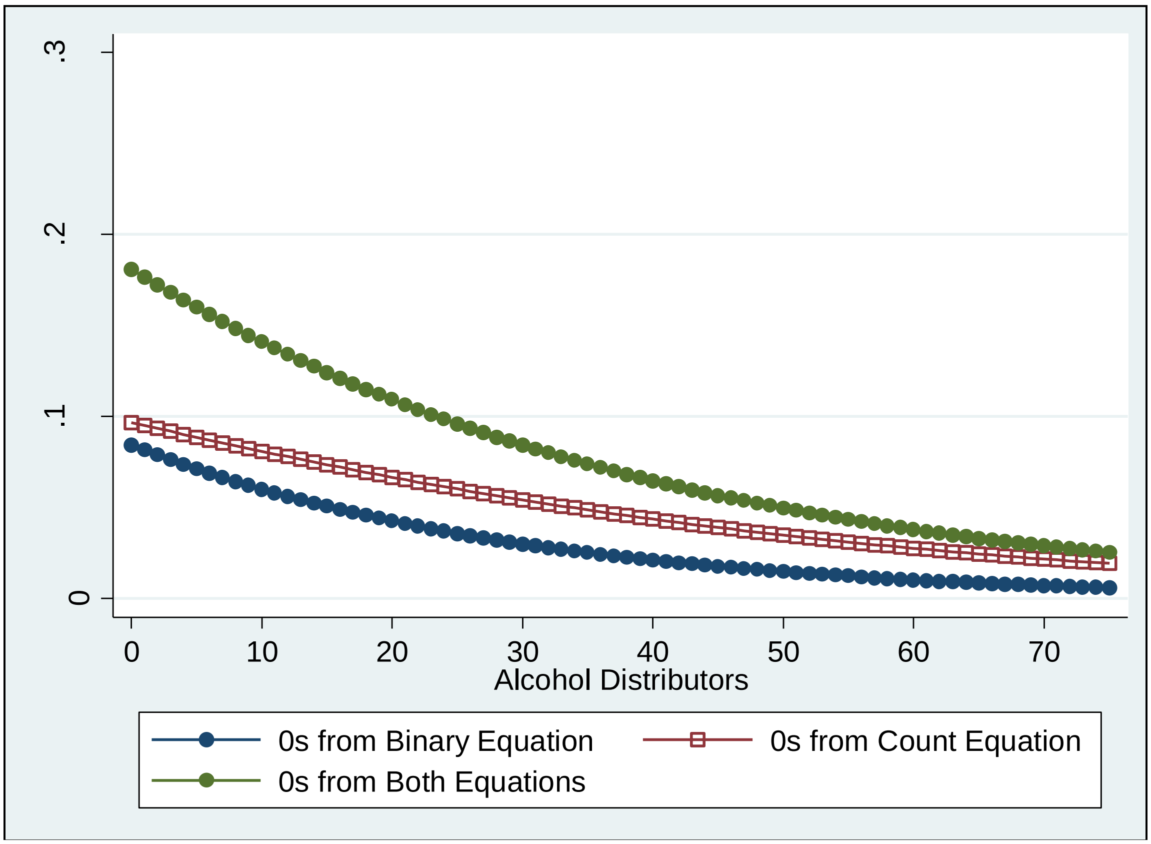
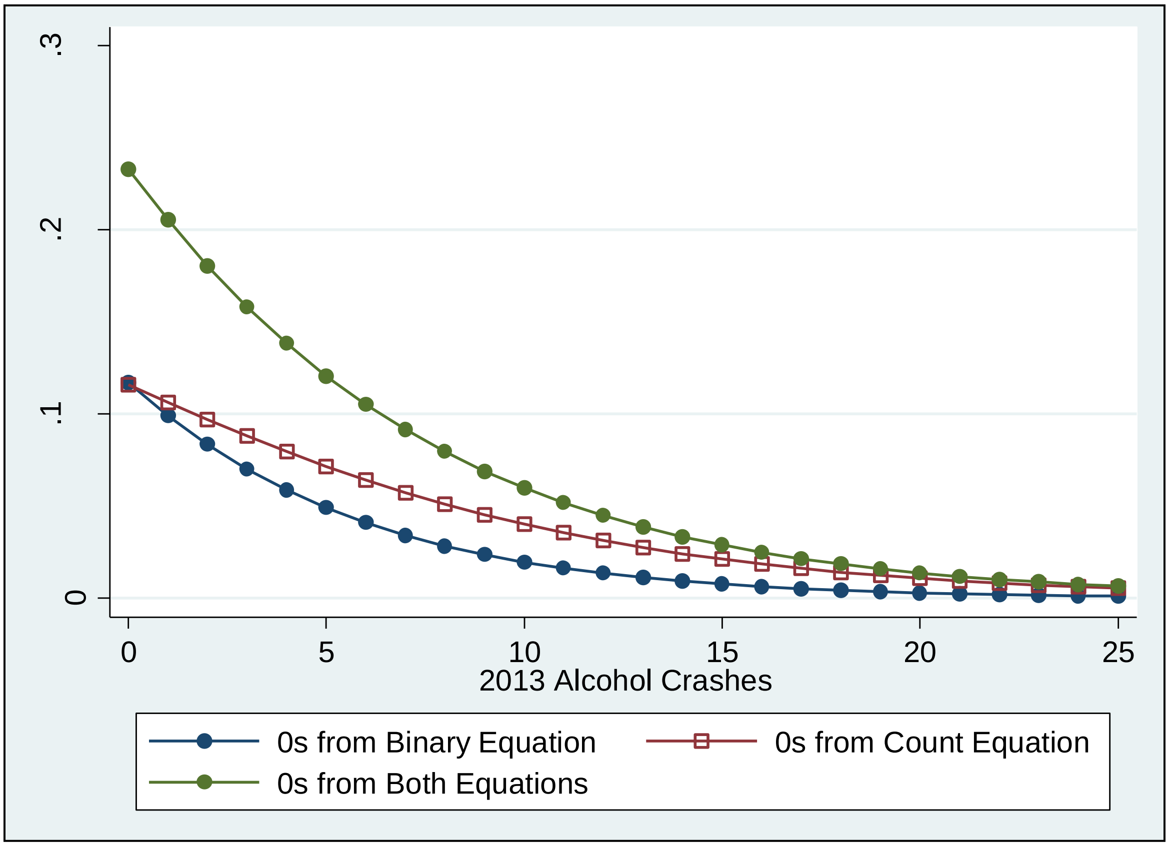


Figure 10 demonstrates a similar pattern with 2013 alcohol involved crashes. The probability of a census tract with zero DWI stops when no alcohol involved crashes occurred the previous year is above 20%. As the number of alcohol involved crashes increases toward its max value of 23 crashes in one census tract, the probability of that census tract having zero DWI stops is virtually 0%.

Figure 10: Probability of 0 DWI stops given a certain number of 2013 Alcohol Involved Crashes



Other variables present more difficult interpretations as shown in Table 5. The percentage of whites in the population of each census tract is significantly positive in the count equation, but not significant in the binary equation. Therefore, it could be said that an increase in the percentage of whites increases the expected number of DWI stops by about 150%. Similarly, an increase in the percentage of people who only hold a high school diploma or less education increases the expected count of DWI stops by 500%. However, neither white nor education attainment was significant in the binary equation.

Table 5 displays four variables that are significant in the binary equation, but not in the count equation; male, marital status, occupation, and the rural-urban continuum code for each census tract (see also Appendix B). One standard deviation increase in the percentage of males in the census tract increases the odds of always of having zero DWI stops by 28.5%. Similarly, one standard deviation increase in the percentage of people who are single increases the odds of the census tract having zero DWI stops by about 93.3%. Conversely, with each standard deviation increase in the percentage of the population with blue-collar jobs the odds of having zero DWI stops within that census tract decreases about 43%. The rural-urban continuum code variable is also negatively significant. This means that with each unit increase, or as the census tracts become more rural, the odds of having zero DWI stops within a census tract decreases by about 45%.

DISCUSSION

This study analyzes MSHP DWI stops in two different ways. Part one analyzes DWI stops by location within census tracts. After grouping common census tracts based on DWI stop rates, certain patterns emerge. Moreover, performing spatial analysis with Moran's I and Getis-Ord Gi* identifies "hot" census tracts with high DWI stop rates that are not randomly displaced throughout the state. This result is similar to the studies by Wieczorek and Hanson (1997) and Wieczorek and Naumov (2002) in that DWI stops and offenders are not randomly displaced within a geographic area. Identifying these "hot" census tracts may help with resource allocation in terms of the number of law enforcement personnel needed in certain areas of the state. Therefore, location and examining individual census tracts is crucial to the understanding of DWI stops.

Part two of this study analyzes the number of DWI stops within each census tract. It is surprising population did not have a large impact. This could be explained by adding or subtracting one additional person from the census tract, which would have zero impact on the dependent variable. The mean population for all the census tracks in this study is approximately 4,460 people with a standard deviation of approximately 1,823. The results indicate increasing by the standard deviation, 1,823 people, will increase the number of DWI stops. However, it is not known how quickly and by how much census tract populations change. It is more likely that type of change would occur over a long period of time.

Other results indicate that census tracts that are homogeneous and less educated increase the expected count of DWI stops, supporting the findings of Romano, Peck, and Voas (2012). Simultaneously, the results of the binary equation would indicate that census tracts with more people who are single would increase the odds of having zero DWI stops. This finding goes against the hypothesis that marriage has a tempered effect on those likely to drink and drive. It also partially supports the findings of Wyse, Harding, and Morenoff (2014) that marriage can become a stressor and result in more offending. However, their finding had the prerequisite that individuals already had a history of drinking and offending, thus marriage could cause them to relapse. In the current study, the prior drinking and offending behavior of individuals is not known so it cannot definitively conclude why census tracts with more single people have higher odds of having zero DWI stops. Another issue could be with the variable itself. The single variable includes anybody over the age of 15 and is capturing people who are not legally allowed to drink. Assuming these people follow the law, the variable inflates the number of people who do not drink and drive, possibly creating a false distinction between the number of people who can and cannot drink and drive and whether they are single. The variable is also capturing people who are generally never married before they can drink.

The racial makeup of a census tract is also significant but only in the count model. The higher percentage of whites a census tract contains, the occurrence of DWI stops will increase. This finding could indicate that a more homogenous population produces more DWI stops. However, this study did not distinguish between ethnicities, just between white and non-white. Looking back at the descriptive statistics in Table 4, the mean for the variable white is 83%. This finding could just be capturing the fact most census tracts have a larger portion of whites than non-whites. More research is needed to investigate this finding.

The occupation variable is significant in the binary equation but not the count equation. Therefore, the odds of always having zero DWI stops decreases with more people employed in the study's selected employment categories. This finding could support Green and Plant's (2007) finding that people employed in certain job categories drink more and thus have a higher potential to drink and drive. Yet, the occupation variable is not significant in the count model, which may lend support for Karlsson, Hvitfeldt, and Romeljso’s (2000) study. Thus, the results are mixed whether the type of occupation plays a role in where DWI stops occur.

Similarly, the rural variable is significant in the binary equation but not the count equation. This means that the more rural a census tract becomes, the odds of always having zero DWI stops in that census tract decreases. This supports Levine and Canter's (2011) finding of rural areas having more DWI crashes. People living in rural areas often drive further to drinking establishments, giving a higher opportunity for driving while intoxicated. This finding also supports Rookey's (2012) research that enforcement patterns of drinking and driving vary between rural and non-rural areas. However, the significance and non-significance between the two equations needs to be explained. As stated previously, the MSHP's jurisdiction technically encompasses all roads within the entire state, but most metro areas are patrolled by local law enforcement. Thus, it makes sense the MSHP would have more DWI stops in rural census tracts. Again, because the variable is not significant in both equations, it is more difficult to make any definitive conclusions. Furthermore, the rural-urban continuum codes are designed as county classifications. It could be that parsing out these codes to all the census tracts within a county, takes away some of the statistical power that would be present at the county level.

The male variable is also positively significant in the binary equation but not significant in the count equation. The results indicate a higher percentage of males in a census tract increases the odds of that census tract always having zero DWI stops. In other words, more males mean less DWI stops. This finding is not expected especially given the research, such as Carpenter's (2004) study finding that males drink more than females. This finding should be used with caution because the variable is not significant in the count equation. The binary equation notes how these census tracts do not have an opportunity for a DWI stop. This finding could just be a product of certain census tracts not having roads patrolled by the MSHP. It could be that other agencies make DWI stops in these census tracts, which are not captured in this study.

Differing from the previous variables, educational attainment is positively significant in the count equation but not significant in the binary equation. This suggests that the higher percentage of lower educated people in the census tract the more DWI stops will occur. This finding echoes Romano, Peck, and Voas's (2102) finding that less educated people tend to drink and drive more than their more educated counterparts.

The most conclusive findings are that the number of alcohol outlets and alcohol involved crashes within census tracts significantly attribute to the number of DWI stops. The more alcohol outlets and alcohol involved crashes in a census tract, the more DWI stops. This extends the research of Scribner, MacKinnon, and Dwyer (1992), outlet density is related to alcohol involved crashes and now DWI stops. This finding is noteworthy because Gruenewald, Johnson, and Treno (2011) state that bars attract local patrons rather than people from surrounding areas. While census tracts are not as big a geographical area as a county, it could be that people being stopped for a DWI within a census tract live in the same tract or relatively close by. Furthermore, the relationship between the previous year's alcohol involved crashes and DWI stops demonstrate both drinking and driving behavior and law enforcement behavior. The more alcohol involved crashes and the more DWI stops could indicate that more people tend to drink and drive in a certain census tract. It can also be an indication that MSHP personnel patrol places known for alcohol involved crashes. The fact MSHP personnel are detecting DWI offenders in the same geographic region as the crashes still indicates a pattern of drinking and driving behavior in the area.

The policy goals are to reduce the number of drunk drivers on the road to prevent crashes that injure and kill innocent people. The demographic variables, education attainment and occupation, suggest more outreach programs and education may decrease drinking and driving. A more attainable goal from a law enforcement perspective is the allocation of resources. Identifying census tracts with more alcohol outlets and more alcohol involved crashes indicates a higher level of drinking and driving behavior. To prevent or even curtail this behavior, law enforcement should be present in these areas. Future research can examine whether placing law enforcement personnel in these tracts deters or lowers the number of alcohol related crashes, or observe if drinking and driving behavior moves to other less patrolled areas.

Future research should also continue to examine the relationship between DWIs and census data. Recall that the DWI stops examined in this study are only those made by MSHP members. If DWI stops made by local agencies were included, the results may differ. Adding in other agency stops may impact the results of the rural and urban classification. Another avenue for research that can be explored in Missouri is looking at offender data. Identifying where DWI offenders come from in relation to where they are traveling could help with more precise positioning of law enforcement personnel. With the improvement of GPS data, research can look at exact locations of offenders and other variables that can be linked to drinking and driving behavior.

REFERENCES

Berger, D. E. & Snortum, J. R. (1986). A structural model of drinking and driving: Alcohol consumption, social norms, and moral commitments. *Criminology*, 24(1), 139-153.

blue-collar. 2017. In *Merriam-Webster.com*. Retrieved (Insert Date), from https://www.merriam-webster.com/dictionary/blue%20collar

Brewer, C.A., & Pickle, L. (2002) Evaluation of Methods for Classifying Epidemiological Data on Choropleth Maps in Series, *Annals of the Association of American Geographers*, 92(4), p. 662-681.

Brown, K., Sarasua, W. A., & Ogle, J. H. (2016). Safety Planning Analysis of the Socio-Economic and Demographic Characteristics of At-Risk Driver Residential Areas in South Carolina. In *Transportation Research Board 95th Annual Meeting* (No. 16-0534).

Carpenter, C. (2004). How do zero tolerance drunk driving laws work? *Journal of Health Economics*, 23, 61-83.

Colon, I. & Cutter, H. S. (1983), The relationship of beer consumption and state alcohol and motor vehicle policies to fatal accidents, *Journal of Safety Research*, 14, 83-89.

ESRI, "How Hot Spot Analysis (Getis-Ord Gi*) works" Retrieved from http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm

ESRI, "Natural Breaks (Jenks)" Retrieved from http://pro.arcgis.com/en/pro-app/help/mapping/symbols-and-styles/data-classification-methods.htm#ESRI_SECTION1_B47C458CFF6A4EEC933A8C7612DA558B

Getis, A. (2007) Reflections on spatial autocorrelation, *Regional Science and Urban Economics*, 37, 491-496.

Getis, A. (2008) A History of the concept of spatial autocorrelation: A geographer's perspective, *Geographical Analysis*, 40, 297-309.

Green, J. & Plant, M. A., (2007). Bad bars: A review of risk factors, *Journal of Substance Use*, 12(3), 157-189.

Gruenewald, P. J. (1996). The geography of availability and driving after drinking, *Addiction*, 91(7), 967-983.

Gruenewald, P. J., Johnson, F. W., & Treno, A. J. (2002). Outlets, drinking and driving: a multilevel analysis of availability, *Journal of Studies on Alcohol*, 63(4) 460-468.

Karlsson, G., Hvitfeldt, T., & Romelsjo, A. (2000). The association between drinking habits, socioeconomic status and drinking and driving in Stockholm County. In *International Conference on Alcohol, Drugs and Traffic Safety, Stockholm*.

Legendre, P. (1993), Spatial Autocorrelation: Trouble or new paradigm? *Ecology*, 74(6), pp 1659-1673.

Levine, N. & Canter, P. (2011) Linking origins with destinations for DWI motor vehicle crashes: an application of crime travel demand modeling, *Crime Mapp*, 3, 7-41.

Logan, J. R., Zhang, W., & Xu, H. (2010) Applying spatial thinking in social science research, *GeoJournal*, 75(10), 15-27.

Long, J. S. & Freese, J. (2006), *Regression Models for Categorical Dependent Variables Using Stata*, 2nd ed, Stata Press Publication, College Station, TX.

Padilla, A. M. & Morrissey, L. (1993). Place of last drink by repeat DUI offender: a retrospective study of gender and ethnic group differences, *Hispanic Journal of Behavioral Sciences*, 15(3), 357-372.

Romano, E. O., Peck, R. C., & Voas, R. B. (2012), Traffic environment and demographic factors affecting impaired driving and crashes, *Journal Safety Research*, 43(1), 75-82

Rookey, B. D. (2002). Drunk Driving in the United States: An Examination of Informal and Formal Factors to Explain Variation in DUI Enforcement across U.S. Counties, *Western Criminology Review*, 13(1), 37-52.

Scribner, R. A., MacKinnon, D. P., & Dwyer, J. H. (1994), Alcohol outlet density and motor vehicle crashes in Los Angeles county cities, *Journal of Studies on Alcohol*, 55(4), 447-453.

State of Missouri Data Portal, ""Missouri Active Alcohol License Data" Retrieved from https://data.mo.gov/Regulatory/Missouri-Active-Alcohol-License-Data/yyhn-562y/

United States Census Bureau (2016), "American Community Survey (ACS)" Retrieved from https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html

United States Department of Agriculture Economic Research Service, (2013), "Rural-Urban Continuum Codes Documentation," Retrieved from https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation/

Valdez et al. (2007). Aggressive crime, alcohol and drug use, and concentrated poverty in 24 U.S. urban areas, *The American Journal of Drug and Alcohol Abuse*, 33, 595-603.

Wieczorek, W. F. & Hanson. C. E. (1997) New Modeling Methods: Geographic Information Systems and Spatial Analysis, *Alcohol Health & Research World*, 21(4), 331-339.

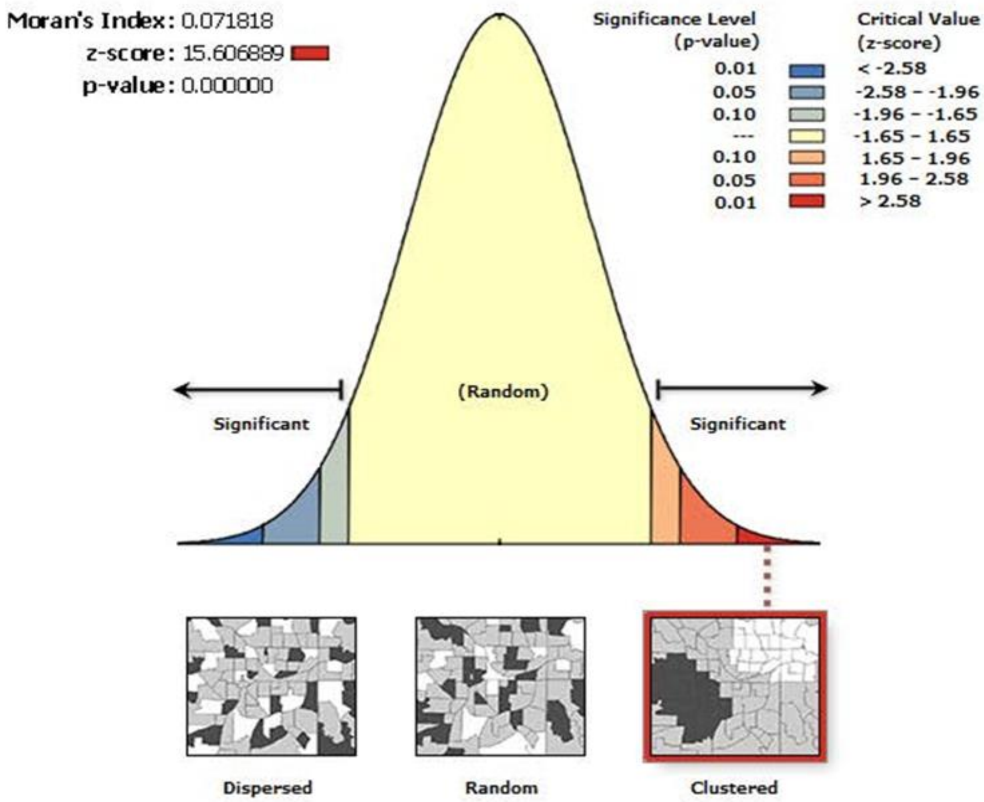
Wieczorek, W. F. & Naumov, A. Y. (2002) Geographic Analysis of DWI Offenders, In *Proceedings of the 16th International Conference on Alcohol, Drugs and Traffic Safety*.

Wyse, J. J. B., Harding. D. J., & Morenoff, D. (2014). Romantic Relationships and Criminal Desistance: Pathways and Processes, *Sociological Forum,* 29(2), 365-385.

Zhang, T. & Lin, G (2007), A decomposition of Moran's *I* for clustering detection, *Computational Statistics & Data Analysis*, 51, pp 6123-6137.

Appendix A

Spatial Autocorrelation Report



Given the z-score of 15.6068886745, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Global Moran's I Summary

Moran's Index:	0.071818
Expected Index:	-0.000782
Variance:	0.000022
z-score:	15.606889
p-value:	0.000000

Dataset Information

Input Feature Class:	ACS_2014_DWI_Stops

Input Field:	VARIABLES_FINAL.CSV.DWI_RATE
Conceptualization:	INVERSE_DISTANCE
Distance Method:	EUCLIDEAN
Row Standardization:	False
Distance Threshold:	29460.2958 Meters
Weights Matrix File:	None
Selection Set:	False

Appendix B

DWI Count	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**95% Conf. Interval**	
Total Population	0.000	0.000	3.55	0.000**	0.000	0.000
Male	-0.346	0.808	-0.43	0.669	-1.929	1.238
Age	0.747	0.780	0.96	0.338	-0.782	2.275
White	0.903	0.287	3.14	0.002*	0.340	1.466
Single	-0.267	0.552	-0.48	0.628	-1.348	0.814
Education	1.794	0.476	3.77	0.000**	0.860	2.727
Occupation	0.192	0.561	0.34	0.732	-0.908	1.292
Income	0.000	0.000	1.61	0.107	0.000	0.000
Distributors	0.016	0.004	3.93	0.000**	0.008	0.024
Crashes	0.088	0.010	9.25	0.000**	0.070	0.107
Rural	0.030	0.015	1.95	0.051	0.000	0.060
Constant	-0.896	0.641	-1.40	0.162	-2.152	0.360
Inflate	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**95% Conf. Interval**	
Total Population	0.000	0.000	-2.40	0.016*	0.000	0.000
Male	6.220	3.017	2.06	0.039*	0.308	12.133
Age	-0.877	2.647	-0.33	0.740	-6.066	4.311
White	-0.478	0.757	-0.63	0.528	-1.962	1.006
Single	5.045	2.022	2.49	0.013*	1.081	9.008
Education	-0.821	1.797	-0.46	0.648	-4.344	2.702
Occupation	-5.607	2.635	-2.13	0.033*	-10.772	-0.443
Income	0.000	0.000	1.22	0.221	0.000	0.000
Distributors	-0.036	0.016	-2.33	0.020*	-0.067	-0.006
Crashes	-0.189	0.052	-3.67	0.000**	-0.291	-0.088
Rural	-0.599	0.185	-3.23	0.001**	-0.962	-0.235
Constant	-2.816	2.194	-1.28	0.199	-7.117	1.485
/lnalpha	-0.520	0.069	-7.55	0.000	-0.654	-0.385
alpha	0.595	0.041			0.520	0.681

*p < 0.05, **p< 0.01