Kennesaw State University
DigitalCommons@Kennesaw State University

Master of Science in Civil Engineering Theses

Department of Civil and Construction Engineering

Fall 10-8-2019

Spatio-Temporal Analysis of Highway Congestion and Accidents: A Case Study of Interstate 285 in Georgia

Melodi Zhan-Moodie

SPATIO-TEMPORAL ANALYSIS OF HIGHWAY CONGESTION AND ACCIDENTS: A CASE STUDY OF INTERSTATE 285 IN GEORGIA

A Thesis presented to the Faculty of

Civil and Construction Engineering of Kennesaw State University

by

Melodi Zhan-Moodie

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Civil Engineering (MSCE)

at the

Department of Civil and Construction Engineering

Southern Polytechnic College of Engineering and Engineering

Technology

Kennesaw State University, Marietta Campus

August 2019

Abstract

Road traffic crashes threaten thousands of drivers every day and significant efforts have been put forth to reduce the number of traffic crashes and their impact. Traffic congestion could be both a result of and a contributing factor to traffic crashes. The aim of this study is to investigate spatiotemporal traffic congestion and crash patterns to gain a better understanding of the causation of congestion and accidents, and their interaction. The Interstate 285 (I-285) in Georgia was used as a case study. With the aid of Geographic Information Systems (GIS), spatial clustering and densities of accidents were performed by following Anselin Local Moran's I method of spatial autocorrelation. The results indicated that the location of high-high accident clusters was in the northern half of the I-285 for all crash types. Additionally, geometric and traffic-related variables were correlated with accidents using logistic regression. The results showed that road segments involving merging, diverging, or weaving lanes had a positive correlation with the number of accidents. Specifically, the merging segments exhibited the highest crash frequency, followed by weaving and diverging segments. On the other hand, the road curvature did not play a significant role in crash occurrence, which is likely due to the gentleness of the road curvatures along the I-285 loop. However, the impact of acceleration on crash frequency remained inconclusive. It appeared that a lower average traffic speed correlated with a higher crash frequency, which may be due to a slow-down condition prior to crash occurrence.

Table of Contents

List of Figures

List of Tables

Introduction

Traveling by personal automobile has become the dominant mode of travel in the United States. As the population and the demand for mobility continue to grow, large metropolitan areas in the U.S. share two common issues: congestion and motor vehicle crashes. These two issues are closely related to each other because accidents cause congestion, and congestion could also cause accidents. The exact nature of this interaction is still somewhat unclear.

Motor vehicle accidents are the fourth most likely cause of injuries and deaths in the U.S. as of 2017 (National Safety Council, 2017). One of the first steps towards decreasing accident frequency is to examine the traffic conditions surrounding accidents. This thesis aims to better understand the relationship between road traffic conditions and accidents by focusing on the traffic congestion and rate of accidents occurring on Interstate 285 in the state of Georgia.

Interstate 285, when opened in 1969, was a four-lane beltway surrounding the city of Atlanta, Georgia. Today, I-285 has grown to be a large 63.98-mile highway consisting of eight to 12 lanes and having 62 exits ("Interstate 285," 2004; FHWA, 2019). Though I-285 has provided a greater degree of mobility for the residents in Atlanta and the surrounding area, the number of privately-owned vehicles registered in Georgia has also increased from around 2.5 million in 1969 to 9.6 million as of 2019 (OHPI, 2018; Department of Revenue, 2019). Unfortunately, the growth of Atlanta as a major metropolitan area in the last few decades has inevitably resulted in I-285 becoming well known for its traffic problems.

Atlanta ranked the 11th most congested city in the United States according to INRIX's 2018 Global Traffic Scorecard (INRIX, 2018). The overcrowded roads can be attributed to many factors such as a large commuting population, several bottleneck locations, and lack of carpooling and connectivity of public transportation. The congestion is generally associated with the entrance and exit lanes at interchanges. The interchange of I-285 and I-85 (North) has been ranked the number one worst truck bottleneck location in the U.S.A. It has the lowest daily speed of only 15 mph (ATRI, 2018). Several other Atlanta interchanges, such as I-75 and I-285 (North) interchange also made the 2018 top 100 for worst congestion rankings for truck-borne freight (ATRI, 2018).

Traffic accidents are related to congestion as both a potential cause and result of road congestion. Traffic crashes are a major cause of injury and death and are thus a major threat to public health. According to the World Health Organization, injuries from road traffic accidents are the leading cause of death for people aged 5 to 29 and around 1,350,000 people die yearly worldwide. In addition, an average of 20 to 50 million people suffer yearly non-fatal injuries from crashes (WHO, 2018). During the year of 2018, on Interstate 285 in Georgia there were 8,456 accidents, 2,230 injuries and 17 fatalities (GDOT, n.d.; National Center, 2019). The human pain and suffering associated with loss of life and injuries is well documented.

In addition to the cost of human life and injuries, traffic accidents also contribute to economic loss. During 2018 in the U.S.A. there were around 40,000 fatalities and an estimated cost of $412.8 billion for motor-vehicle deaths, injuries and property damage (NSC, n.d.). Atlanta and surrounding areas lost an average of 108 hours to congestion in 2018 at an average cost of $1,505 per driver (INRIX, 2018). By improving road conditions financial loss can also be reduced.

Government agencies have engaged in numerous efforts to improve road conditions and reduce traffic accidents. These efforts would benefit from a deeper and better understanding of the various potential factors that are associated with traffic accidents. Identifying these factors may help transportation agencies in their continuing work to improve the interstate highways to make them safer for motorists.

Objective

As discussed previously, both congestion and traffic crashes are issues of significance because they reduce the efficiency of transportation systems and threaten the safety of motorists. Traffic congestion is also linked to an increase in accidents. For example, rear-end collisions, which are the most common crash type, are more likely to occur during heavy traffic conditions due to traffic instability in congested road condition (Tanaka et al., 2008). This in turn could lead to more congestion from the lane being blocked by a crash, creating a cycle of accidents and congestion on the road.

The aim of the current study is to examine measurable variables that are related to traffic conditions, including congestions and driving speed, on interstate I-285.

The specific objectives of this research are:

1) Determine the spatial-temporal distribution of traffic speed on I-285.
2) Visually identify and examine its relationship with traffic accidents to determine potential patterns.
3) Analyze traffic speed data and historical crash data on I-285.
4) Examine correlations among variables such as acceleration of traffic, time, location, lighting condition, road surface condition, injuries, and the manner of collision.

Literature Review

Previous studies on the prediction of accidents examined contributing factors of accidents. Research focused on traffic flow dynamics, such as speed and acceleration. One main reason is because traffic flow data are not always collected with crash data, making it difficult to examine these variables. Understanding the relationship between traffic dynamics and traffic accidents is imperative to understand the relationship between traffic crashes and congestion.

The literature was analyzed by comparing different statistical models for prediction of crashes. The consensus was that determining the effect of both spatial and temporal variables in traffic crashes can be important in accounting for correlation due to unobserved heterogeneity. In other words, the data may not show why so many crashes are occurring, but if there is a significant correlation between crashes and other variables at a specific place and time, then that is helpful in finding out where the problem may lie. This can help narrow down where the road safety resources should be directed to be most effective (Liu & Sharma, 2018). A review of the literature indicated that most studies conducted data analyses of traffic crashes using both spatial and temporal effects, spatial or temporal, or neither. The benefits of each method are discussed.

Overview of factors

Many studies have been conducted to examine factors that contribute to road crashes. Identifying factors affecting road accidents help create informed decisions for making safer roads. Some factors may be difficult to quantify or collect large scale data on such as driver attentiveness or emotional states. This can lead to unobserved heterogeneity that requires a more sociological approach to study. Other factors such as traffic speed, time of day, or weather conditions are easier to identify and analyze. Many factors have been found by researchers to correlate with crash frequency; for example, curvature of road segment, freeway exit and entrance ramps, traffic volume, lighting, weather, work zones, speed, and deceleration rate (Kim et al., 2016; Zhang et al., 2016; Li et al., 2015; Glennon et al., 1983). Examining a wide range of potential factors surrounding accidents may lead to findings on which specific variables contribute to traffic crashes and need more focus in future studies.

The link between congestion and accidents is well documented. A study of the effect of the congestion charge on London traffic by Green et al. (2016) compared the accident rates and congestion of Central London before and after the implementation of a driving fee. The results show that after the city of London added a charge the congestion decreased, which in turn resulted in reduced accidents (Green et al., 2016).

A study by Pasidis (2019) on the two-way effect of accidents and congestion found there was a significant effect on each other. Pasidis examined the Highways Agency data in England, which included the average speed and journey times of ‘A’ roads on the Strategic Road Network from 2009 to 2015. There was evidence of a significant effect accidents had on congestion. There was a negative effect that accident occurrence had on traffic flow with a decrease in average speed on of 7.8 km/h (4.85 mph). A suggested reason is that drivers slow down to view an accident ('rubbernecking') (Pasidis, 2019).

Traffic congestion can be inferred from one of the three traffic stream variables: speed, traffic density, and traffic flow, whose relationship is often described by the classic traffic flow models.

Speed has an interesting association with accidents. Higher speeds are associated with more fatal crashes, but the studies have found both a positive and negative association for average speed with accident frequency (Wang, 2010). A possible explanation for this is that since most accidents involve another car, more cars lead to more accidents. But if there are more cars it results in the average traffic speed decreasing according to the flow-density relationship.

Another explanation is the variance of speed increases safety more than the speed itself (Lave, 1985). Part of the danger of congestion could be due to the deceleration of vehicles. Vehicles encountering congestion must sometimes make sudden decelerations to match the new speed. Sudden deceleration is a big correlating factor with rear-end type crashes (Kim et al., 2016). The National Highway Traffic Safety Administration advises that cars should be at least 100 feet away from each other at 30 mph and 400 feet away at 70 mph, but this advice is not followed enough.

A study in North Carolina on rear-end crashes used micro-scale driving behavioral data gathered through an in-vehicle sensing system. They were able to get speed data at a resolution of one-minute from cars with the INRIX GPS-enabled probes. The speed data of the location of identified crash hotspots was analyzed for travel time, acceleration, and trip or driver-based coefficient of variation. The study found high rates of deceleration correlated with road segments with high crash rates (Kim et al., 2016).

A higher number of vehicles present is associated with a greater risk for accidents. Previous research on the relationship between the density of cars and accidents have mixed conclusions. Lord et al. (2005) conducted an analysis on the freeways in Montreal, Quebec, on the relationship between accidents and density, where density is measured as number of vehicles per lane per kilometer. The analysis indicates a negative relationship with density and accidents that did not involve collision with another vehicle or resulted in fatalities. The accident rate was found to follow a U-shaped curve where as the density increases, accidents decrease, but then increase again at higher densities (Lord et al., 2005). Then once the traffic has reached absolute congestion where the density of vehicles has reached its critical point, there is no movement and no accidents occur.

The conclusions about the relationship between traffic flow and accidents are varied as there is a wide range of traffic flow and accident types. For example, hourly traffic flow was found to be positively correlated with accidents involving multiple vehicles, yet hourly traffic flow was found to be negatively correlated for single vehicle accidents (Wang, 2010). In a study on road bridges, traffic volume was the most important factor influencing road accidents (Elvik, 2019). The southeast region in China has a high traffic volume due to the major developed cities located there and is responsible for 25.6% of the serious casualty crashes from 2009-2013, providing support for a positive relationship between traffic volume and road accidents (Xu et al., 2018).

Certain types of road segments are more accident prone than others, especially the road segments on interchanges. The study by Kim et al. (2016) on the relationship between driver behavior and accidents on I-40 in North Carolina found 85% of all rear-end crashes examined occurred within 2000 feet of an entrance ramp to the freeway (Kim et al., 2016).

There is evidence that roads with more curvature are likely to have more accidents as well. Road alignment such as curves were a contributing factor in bridge accidents (Elvik, 2019). Vogt and Bared (1998) evaluated highway segments for Washington and found that increasing lane widths and less horizontal curvature reduces total crashes. Another study found 55.6% of serious casualty crashes occurred on roads with curves or slopes (Xu et al., 2018). Xu et al. postulated this could be due to drivers not being able to see ahead in curved roads as far or clearly as they could otherwise on straight stretches.

Weather condition and its impact on the road has also been studied by researchers. A study investigated the frequency of accidents in Finland from 2014 to 2016 by different weather and road conditions by road type and accident type (Malin et al., 2019). The analysis identifies five categories of dangerous road conditions based on road surface condition, visibility, and icy rain. Results suggested that adverse road conditions such as wet or icy surfaces, were correlated with increased accidents. The highest risk was for two-lane and multiple-lane roads and slippery roads. Multiple-vehicle accidents had the highest relative accident risk on multiple-lane roads (Malin et al., 2019).

Statistical Methods

There have been several methods developed that can be used to study traffic issues. Researchers use these measures in their studies to analyze the probability and severity of road accidents within a city or province using GIS and find hotspots using GIS spatial statistic tools, Moran's I, Getis-OrdGi* , and/or point kernel density function (Aghajani, 2016; Shafabakhsh, et al., 2017; Liu and Sharma, 2017; Prasannakumar et al., 2011). The Global Moran's I statistic determines if a pattern is clustered, dispersed, or random, compared to nearby features (Esri). Getis-OrdGi* is a spatial auto-correlation method finds high or low value clusters (Getis & Ord, 1992). Kernel density estimation (KDE) is the most common way to determine the density of each data point in a defined area (Plug et al., 2011). The results from these studies provide useful information. The drawback from the above methods is that the data is examined on spatial correlation or on temporal and spatial factors separately (Prasannakumar et al., 2011).

Research shows that the Local Moran's I tool is especially useful as it doesn't use the total aggregate of the data. By examining crashes as outliers compared to nearby road segments it can better illustrate potential correlation from unobserved heterogeneity, assuming nearby locations generally share similar qualities. It doesn't limit hotspots to road segments which have high traffic volume as compared to the mean volume of every segment (Abdulhafedh, 2017).

A method for analyzing crashes against other variables is ordered probit (OP) model, where the marginal effects of each explanatory variable such as curvature or weather conditions are computed based on if they are correlated with an increase in severity of crashes (Zhang et al., 2011). The researchers report that this is an effective model when examining more than one outcome with a natural ordering.

Other researchers find that for variables that have more than two outcomes, the logit model, also known as the logistic model, can be used to model the probability of an occurrence (Greene 2012; Xu et al., 2018). A multinomial logistic regression can assign more than two choices to a prediction model, making this a useful tool (Greene, 2012).

Studies comparing various traffic crash prediction models found that when comparing the multivariate Poisson-lognormal model, multivariate spatial Poisson-lognormal model, and multivariate spatiotemporal Poisson-lognormal model using WinBUGs software, the multivariate spatial-temporal model had the best fit for crash frequency prediction when the crashes were ranked by severity (Zhan et al. 2015; Ma et al. 2017). But univariate spatial-temporal modeling has been found to be the best fit when only one type of crash is examined (Liu & Sharma, 2017).

Most previous studies have examined the accident data at a yearly or daily level. It's been suggested that accident data may be more useful at the micro level (Plug et al., 2011). Further, using a higher resolution time scale can help better visualize the patterns found in the data. Therefore, the current research, focusing on data at 5-minute time interval, would provide a level of resolution that could result in a better understanding of traffic dynamics and the associated safety implications, which could be beneficial for future safety initiatives.

Data Acquisition and Processing

For this research, historical accident data from January 1, 2016 to December 31, 2018 was obtained from the Georgia Department of Transportation (GDOT). GDOT uses the Navigator's Video Detection System (VDS) as the primary source of real-time transportation information. Approximately 1,645 VDS stations were installed every 1/3 mile along most interstates around Atlanta. These VDS cameras provide continuous speed and volume data to the GDOT Transportation Management Center. The xml file containing traffic information from those VDS cameras was downloaded from the GDOT Navigator system. An excerpt of the XML file is shown in Figure 1.


```
▼<detectorData xmlns="tmdd">
   <organization-id xmlns="">GDOT</organization-id>
  ▼<collection-period xmlns="">
    ▼<collection-period-item xmlns="">
      ▼<detection-time-stamp xmlns="">
         <date xmlns="">20171210</date>
         <time xmlns="">08090000</time>
         <offset xmlns="">-0500</offset>
       </detection-time-stamp>
      ▼<detector-reports xmlns="">
        ▼<detector-report xmlns="">
           <detector-id xmlns="">2760</detector-id>
           <detector-name xmlns="">GDOT-STN-2856016</detector-name>
           <detector-status xmlns="">1</detector-status>
          ▼<lane-data xmlns="">
            ▼<lane-data-item xmlns="">
               <detector-lane-number xmlns="">10</detector-lane-number>
               <lane-vehicle-count xmlns="">1</lane-vehicle-count>
               <occupancy xmlns="">0</occupancy>
               <lane-vehicle-speed xmlns="">85</lane-vehicle-speed>
             </lane-data-item>
            ▼<lane-data-item xmlns="">
               <detector-lane-number xmlns="">01</detector-lane-number>
               <lane-vehicle-count xmlns="">0</lane-vehicle-count>
               <occupancy xmlns="">0</occupancy>
               <lane-vehicle-speed xmlns="">0</lane-vehicle-speed>
             </lane-data-item>
           </lane-data>
```


Figure 1. An example of the XML file of traffic data

The traffic data were collected from October 1, 2017 to June 25, 2018 at five-minute intervals. Then the specific values from the extracted xml files were written to an SQLite database using a Python program. An inventory file, containing information on camera name, the latitude and longitude of each camera, was used to map traffic information to a GIS roadway layer.

From the SQLite database, the speed data was visually illustrated by coding the speed as a pixel value using Python as seen in Figure 2. The y axis represents data organized by physical placement of cameras. The x axis represents the time of day by five-minute increments. Each pixel color represents the speed of traffic with white pixels meaning no movement and black pixels showing high speed with a range from 1 to 189 km/h. The cameras would not record speed data if there were no movement, so null speed values were set to -2 for ease of processing. Camera status was classified as 1, 2, or 3 to signify if the camera was operational at the time it was accessed. There was a total of 384 cameras but only 365 were consistently functioning for the duration of the eight months the data was collected. There are 180 cameras on counterclockwise traffic and 185 on clockwise traffic around I-285. In Figure 2, the top half of the image shows the clockwise traffic and the bottom half counterclockwise traffic. Due to the numbering system 11 clockwise cameras were placed in the bottom half, though only 5 were consistently functional. Additionally, there are 288 five-minute intervals in 24 hours but only 278 pixels are represented due to the cameras tending to provide no speed data from 23:35 to 00:25 for unknown reasons. The resulting image can be seen in Figure 2, where the red tinted area shows clockwise direction of I-285 and the purple tinted area shows counterclockwise direction of I-285.

Figure 2. Initial image created from collected speed data from I-285 cameras on March 28, 2018.

Though there would be deviation from day to day, a clear pattern regarding the ‘clouds’ of traffic congestion revealed itself based on day of week. It appears that Saturday and Sunday show far less congestion than weekdays on average. The averaged images (Figure 3, Figure 4, and Figure 5) based on each day of the week were created using the GNU Image Manipulation Program (GIMP) to show a clearer visual representation of congestion by weekdays and weekends. The color scale goes from white which represents the lowest speed and to black which represents the highest speed.

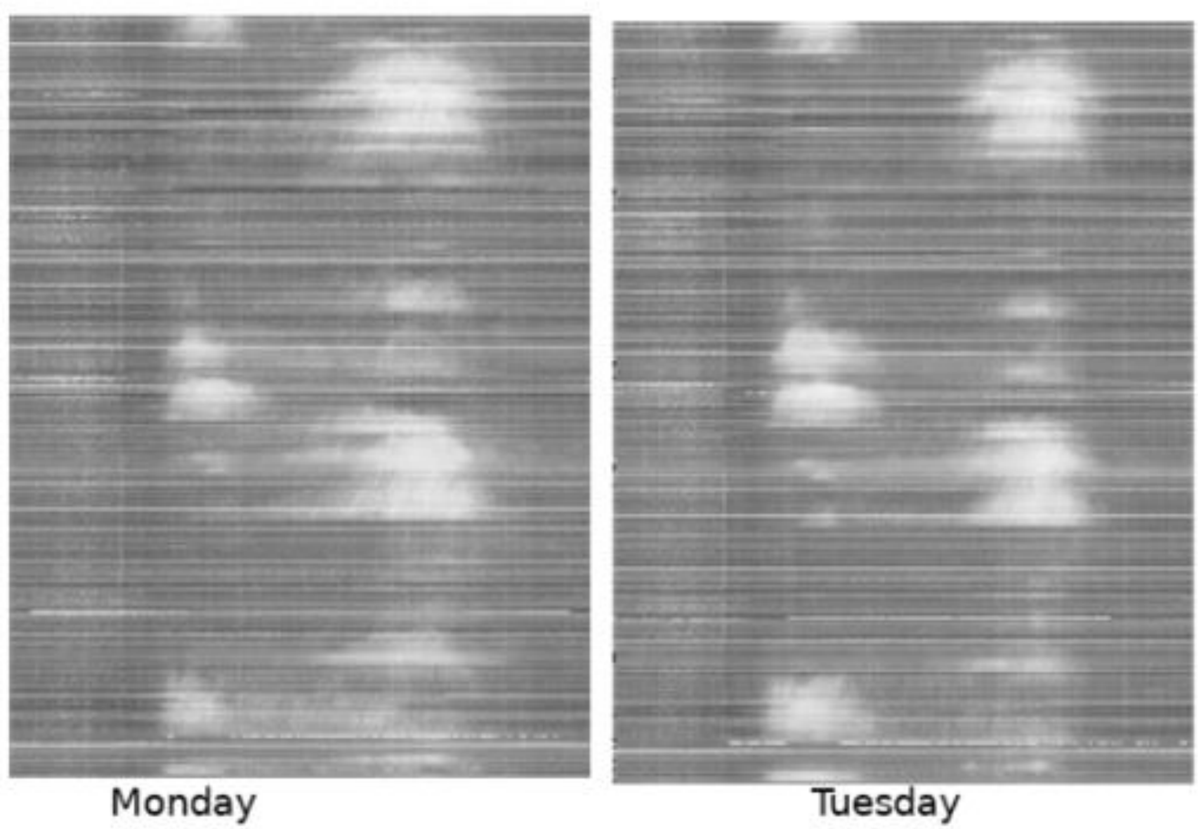


Figure 3. Average speed from October 1, 2017 to June 25, 2018 on Monday and Tuesday.

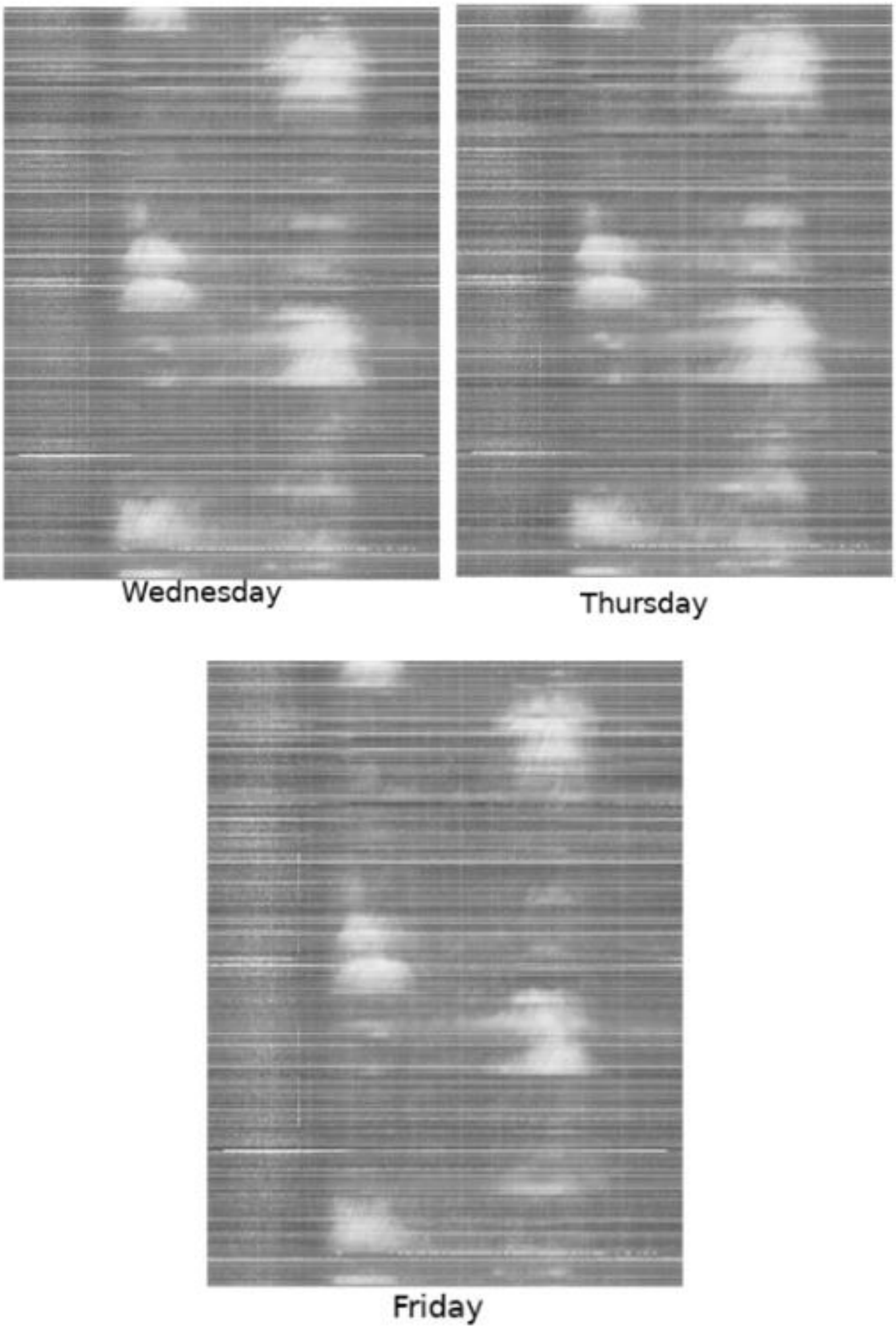


Figure 4. Average speed from October 1, 2017 to June 25, 2018 for Wednesday through Friday.

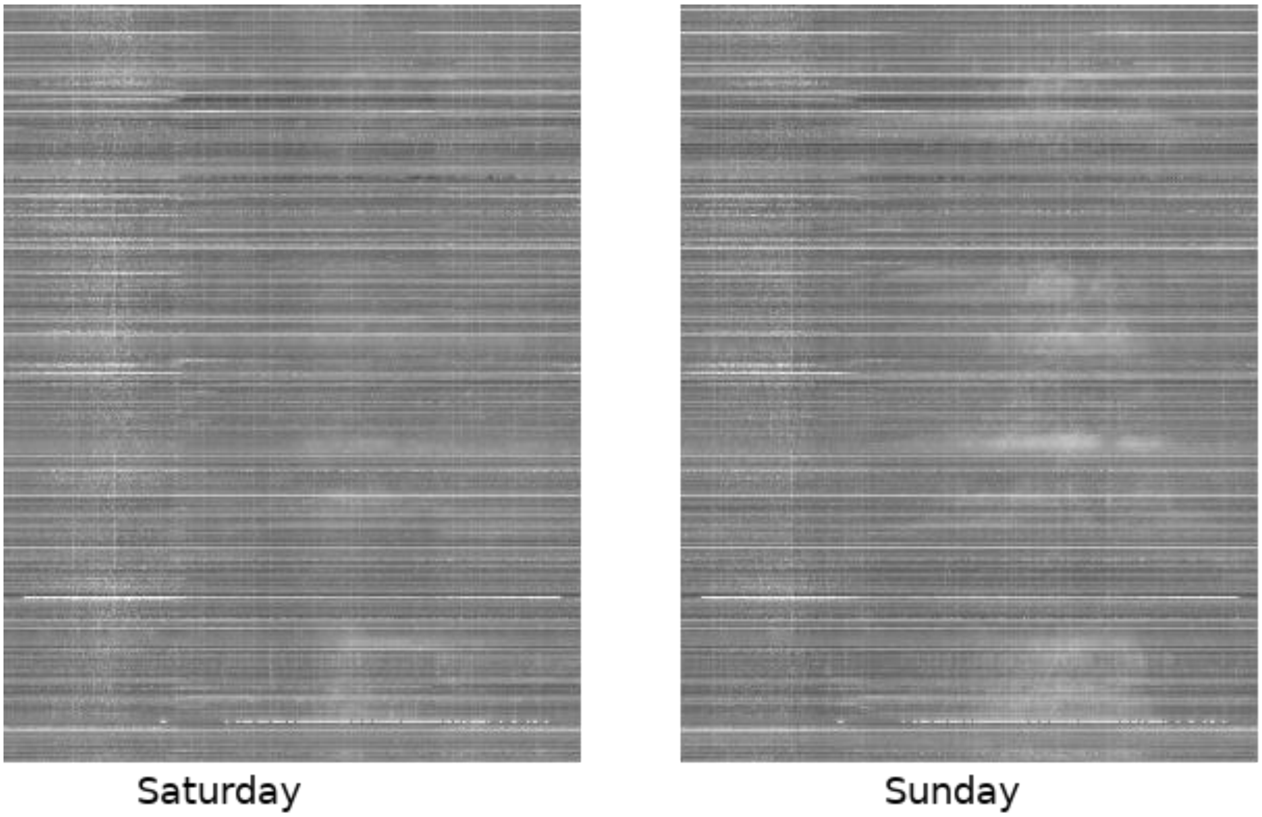


Figure 5. Average speed from October 1, 2017 to June 25, 2018 separated by Saturday and Sunday.

The images were first denoised by averaging the images by weekday. Then, the horizontal white lines of missing data were masked and inpainted using the surrounding pixels. Finally, the selective gaussian tool from the GNU Image Manipulation Program (GIMP) was applied. Those successive steps result in a 278x365 image, as shown in Figure 6.

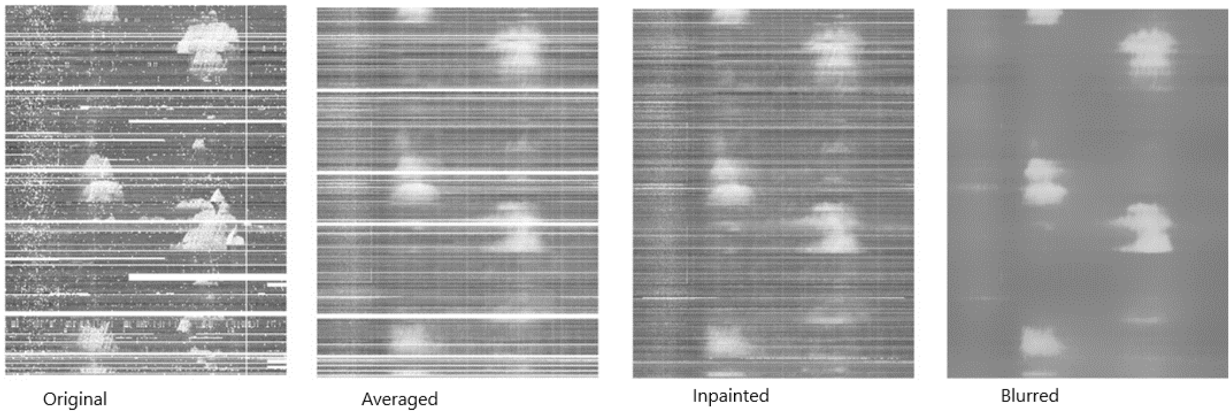


Figure 6. The averaging and denoising of Wednesday traffic.

The eight ‘clouds’ of congestion found were marked and numbered in Figure 7, occurring mostly in the northern half of the I-285 loop as seen in Figure 8 and Figure 9. Clouds 1, 4, 5, and 8 appeared from around 6:00 to 10:00 and clouds 2, 3, 6, and 7 appeared from roughly 13:00 to 19:00. Based on this pattern, it would be most beneficial to direct congestion relieving policies to these specific locations and times.

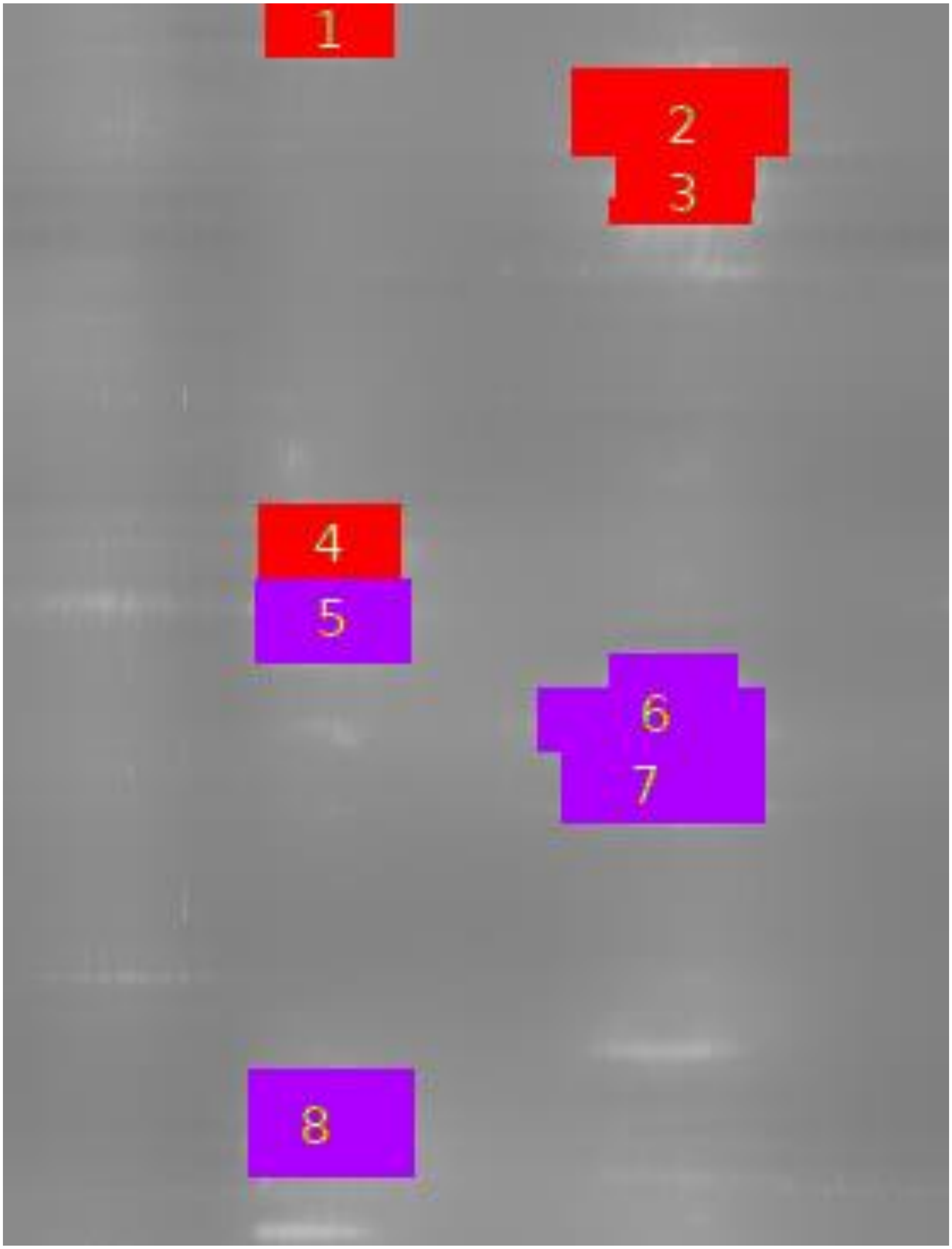


Figure 7. Averaged weekday speed map with identified congestion.

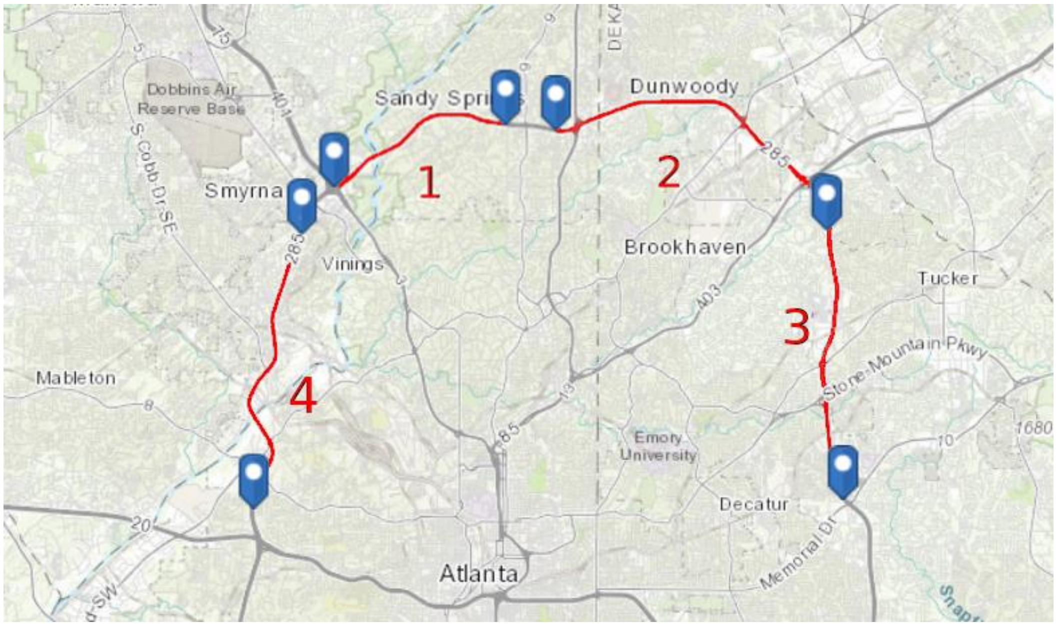


Figure 8. Location of congestion clouds for Clockwise traffic.

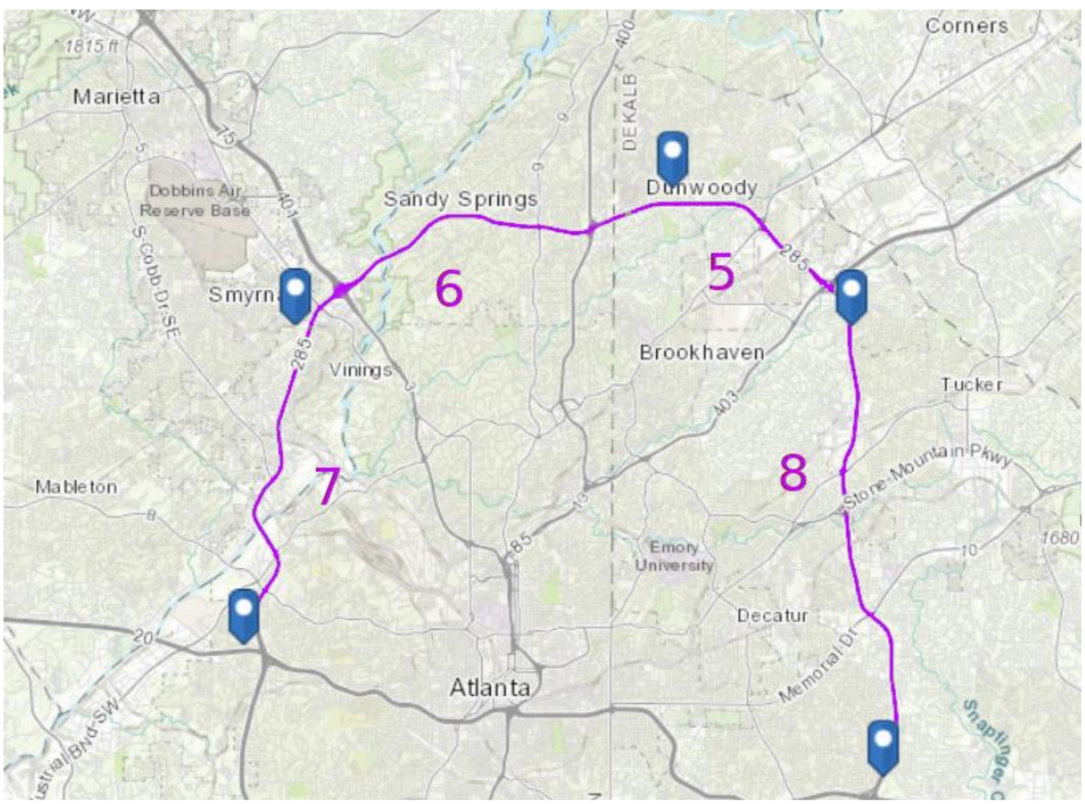


Figure 9 Location for congestion clouds for Counter-clockwise traffic.

In order to visualize the spatial-temporal relationship between congestion and accident occurrence, red pixels denoting crashes were placed in the “cloud” map based on the date, time and nearest camera of the crash. An example can be seen in Figure 10. As shown, the majority of crashes occurred within or at boundaries of the clouds.

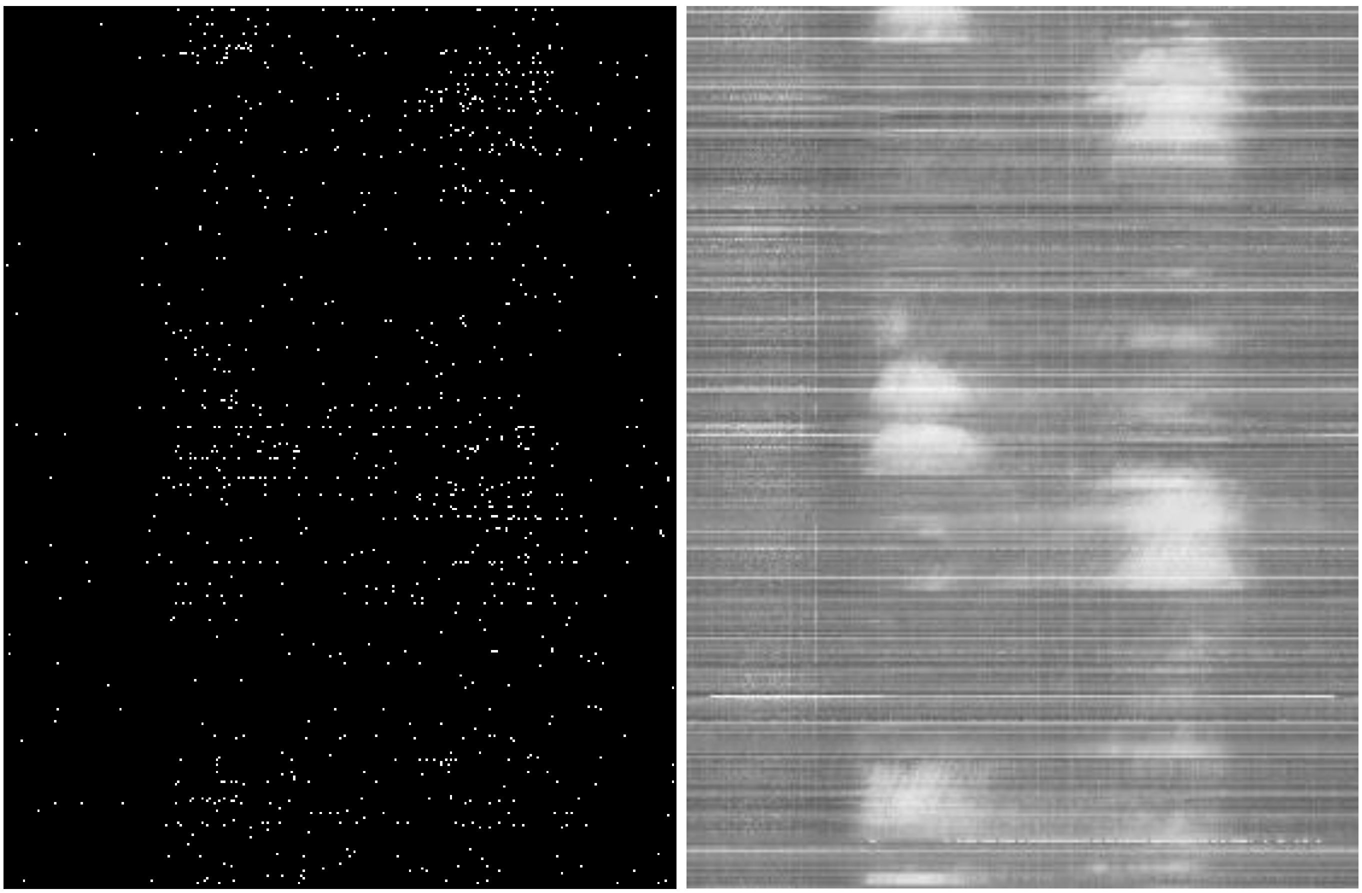


Figure 10. A visualization of the 42 traffic accidents (in red, size increased for clarity) on a denoised cloud map representing speed of traffic on March 28, 2018.

To further examine the relationship between traffic speed and accidents, a random weekday, Monday, was picked as the point of analysis. The total number of crashes for Monday from 2016 to 2018 was compared to the averaged traffic speed map for Monday in Figure 11.

There appeared to be a correlation between traffic speed and accidents as evidenced by the accidents appearing to be positioned near or within the clouds, as shown in Figure 11.

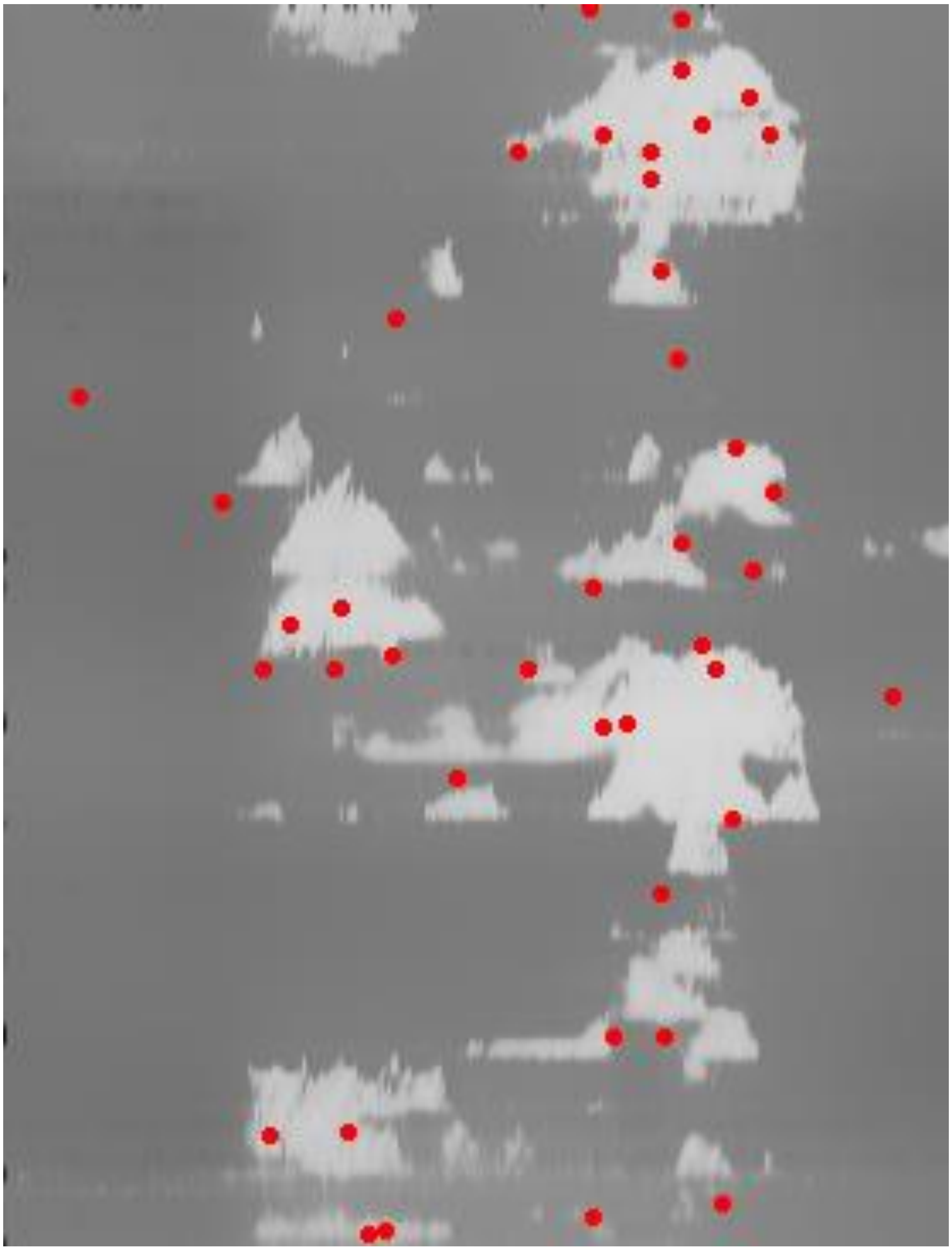


Figure 11. All Monday accidents from 2016 to 2018 on I-285 (left) and averaged Monday I-285 traffic speed map (right).

To investigate possible correlation between acceleration and accidents, an acceleration plot of the speed data was created using Python. It was hypothesized that areas of high deceleration would result in more crashes on freeways.

Then the crashes were graphed to the image as seen in Figure 12. Dark blue represents high deceleration and yellow represents high acceleration.

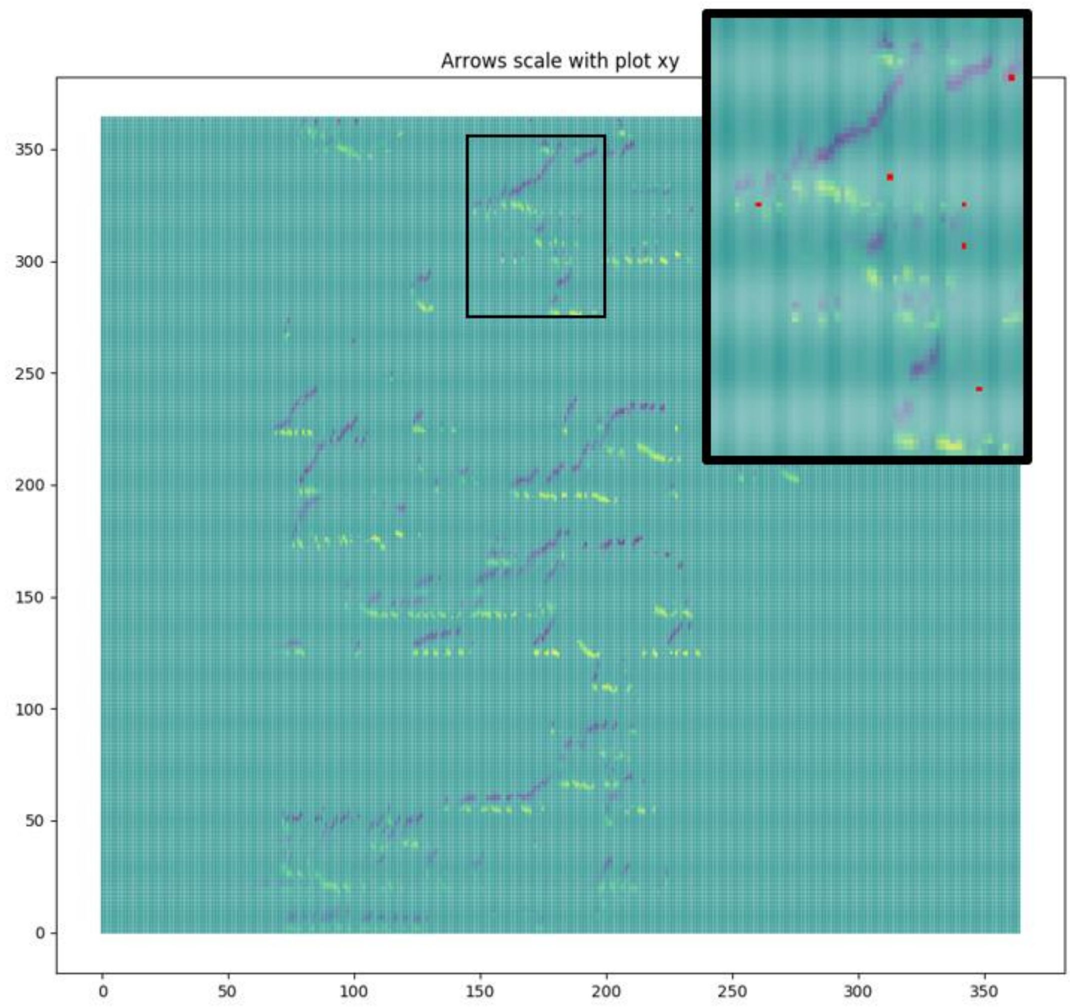


Figure 12. Acceleration plot of March 28, 2018 with image in scale to show relation of crashes (red pixels)

Applying crash data to averaged data did not illustrate an accurate image of the possible relationship between traffic acceleration and accidents. A simple visual observation did not reveal an obvious connection between acceleration and crashes. This might be because the spatial resolution is not high enough due to each camera being roughly three miles away from each other. Additional analysis would provide insight on the possible significance of a correlation between accidents and traffic acceleration.

The camera and crash data were transferred to a MySQL database for additional clean up and processing. The acceleration values for each segment of time at a given location were calculated and accident data was merged to the table based on time and location.

The crash data from 2016-2018 contained 25,875 accidents that happened on the I-285. Most of the accidents, with 14,390 instances, are categorized as 'Rear End', meaning a vehicle impacted the rear bumper of the vehicle in front of it. The second most numerous crash type, with 6,728 instances, was 'Sideswipe-Same Direction', where the sides of two vehicles traveling in the same direction make impact. The category of 'Angle' collisions is third most common with 2,387 occurrences and is when a vehicle impacts another vehicle at an angle. The 'Not A Collision with Motor Vehicle' category is when a vehicle hits something like the road railing or a tree and happened 2,135 times. 'Head On' collisions are when two cars in opposite directions collide. ‘Head On’ collisions are relatively uncommon with only 138 reported cases but are the deadliest type of collision (Georgia Governor’s, n.d.). 'Sideswipe-Opposite Direction' crashes are the same as the 'Sideswipe-Same Direction' crashes, but the vehicles are traveling in opposite directions and only happened 48 times. For 49 accidents there was no manner of collision recorded and was labeled as 'Unknown'. Figure 13 presents these data.

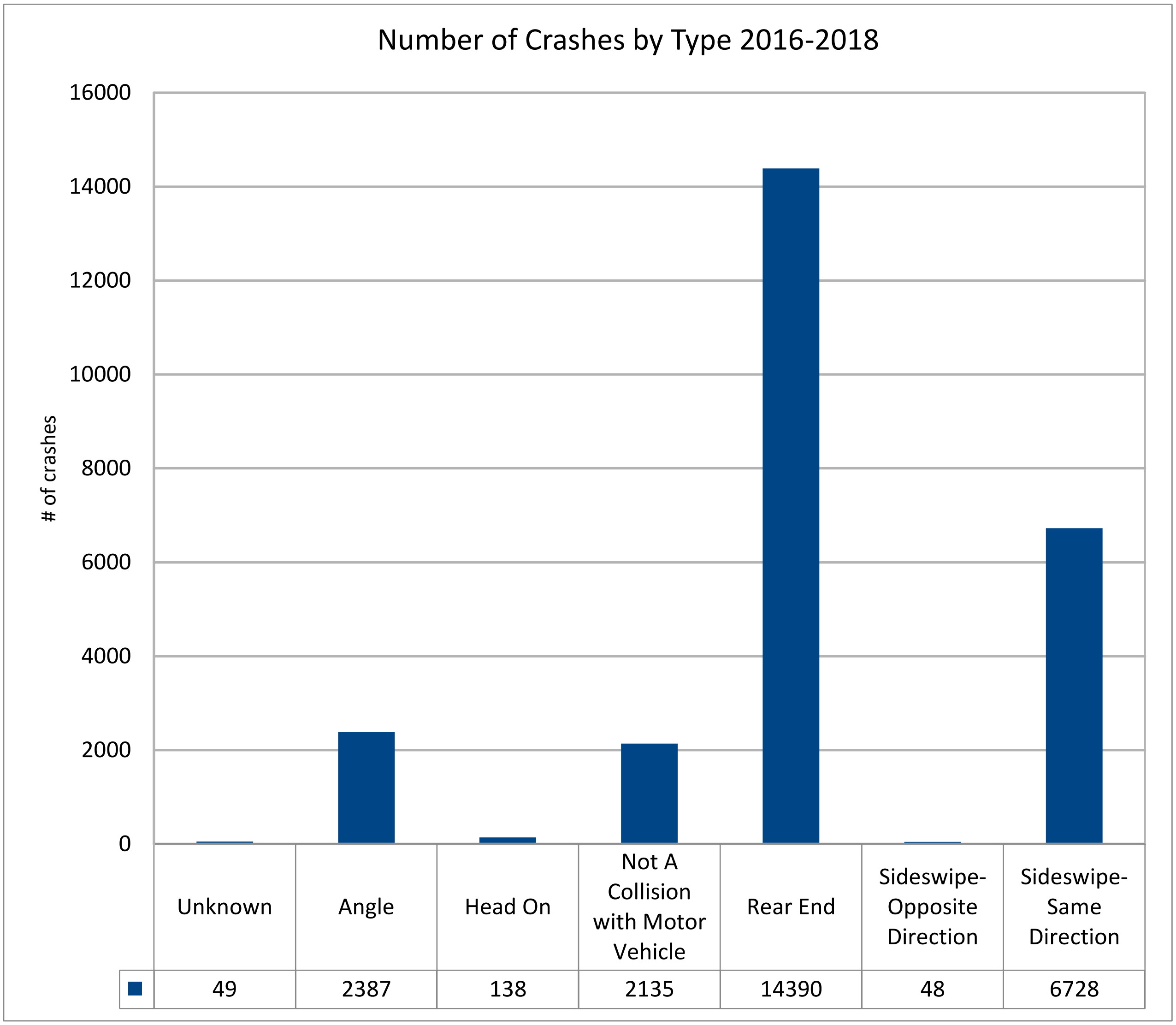


Figure 13. Distribution of I-285 Crashes from 2016 to 2018 by collision category.

In order to perform any statistical analysis on the relationship between traffic accidents and road variables, the ArcGIS software was selected due to familiarity and widespread use for spatial statistics. In ArcGIS Desktop the edited accident data was uploaded and further modified by adding a column with a binary choice, created for identifying direction of traffic, position regarding interchange location, severity of crash, and other variables. Then each accident was mapped according to crash type. Figure 14. Heat map of Crashes on I-285 in 2018 by Collision type 1-4. and Figure 15. Heat map of Crashes on I-285 in 2018 by Collision type 5 and 6 presents the results.

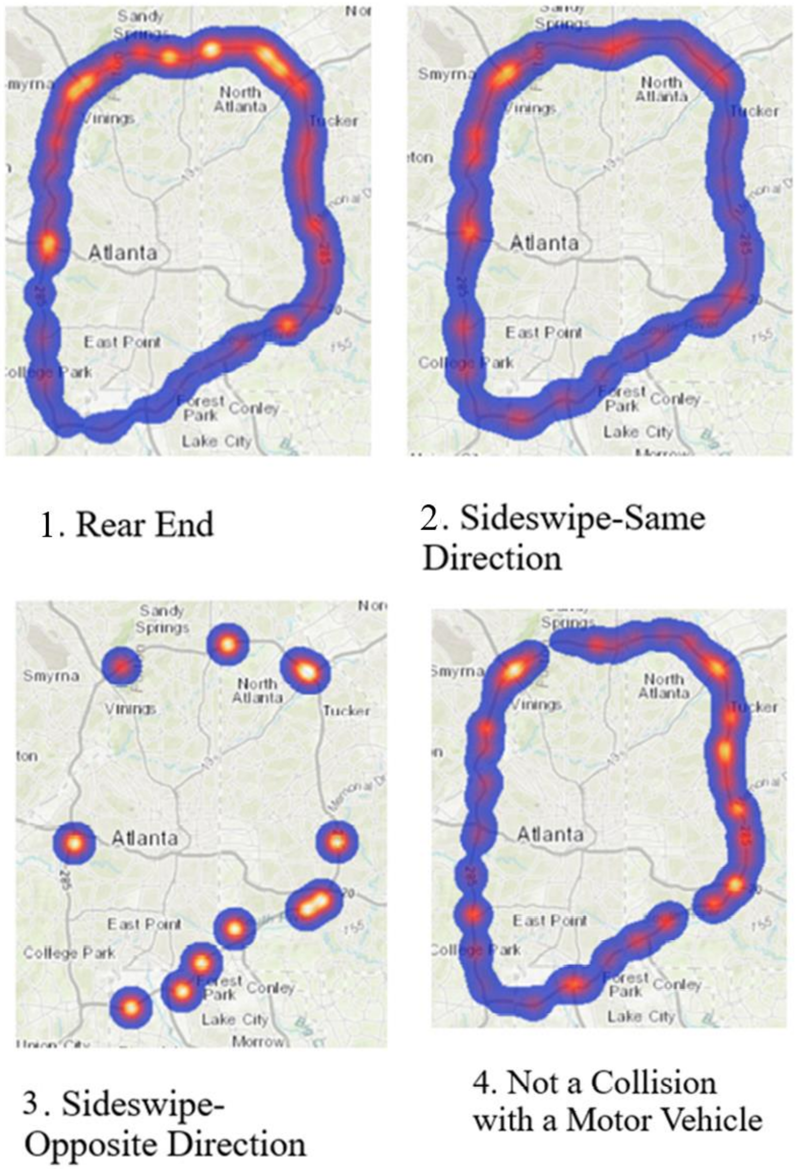


Figure 14. Heat map of Crashes on I-285 in 2018 by Collision type 1-4.

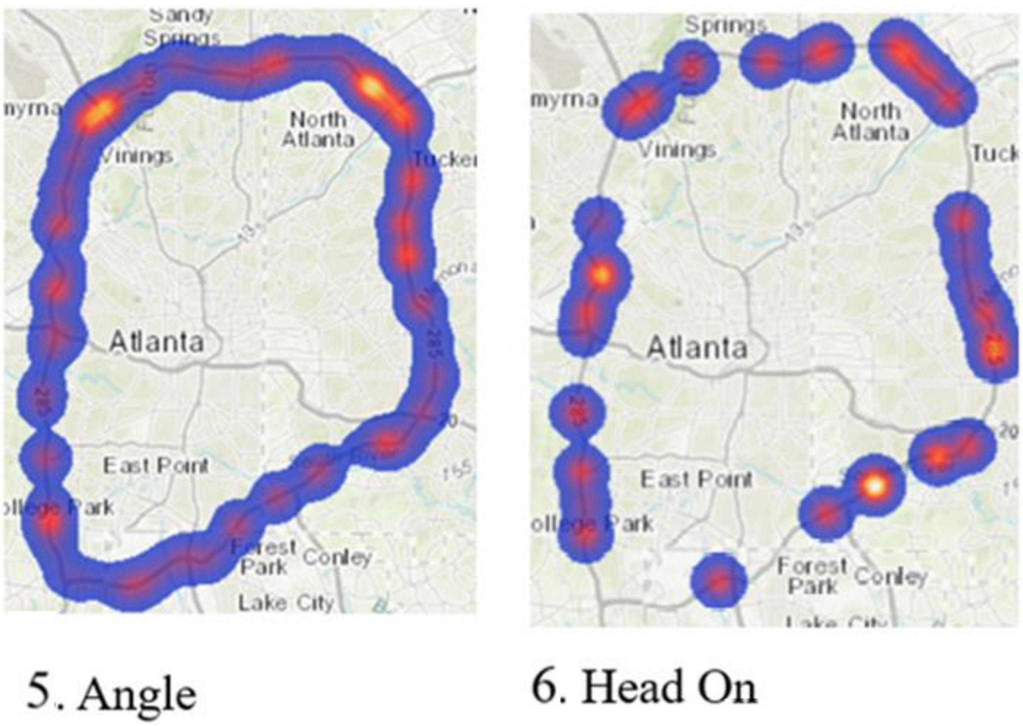


Figure 15. Heat map of Crashes on I-285 in 2018 by Collision type 5 and 6.

Due to limitations in file upload size, only the crashes on I-285 from 2018 were mapped to ArcGIS. There were not enough data points for Head On, Sideswipe-Opposite Direction and Unknown type crashes to provide a comprehensive idea of high concentration areas. However, Angle, Not a Collision, Rear End, and Sideswipe-Same type crashes showed some similarities in areas of highest crash concentration, most notably the northwestern corner of the I-285 loop.

The heatmap generated for each type of collision showed different areas of high concentrations with some patterns emerged. There appeared to be a tendency for accidents to happen around interchanges, especially rear end crashes. An example of how the crashes appeared to congregate near an interchange with empty areas before and after is shown in the close-up of an interchange below in Figure 16

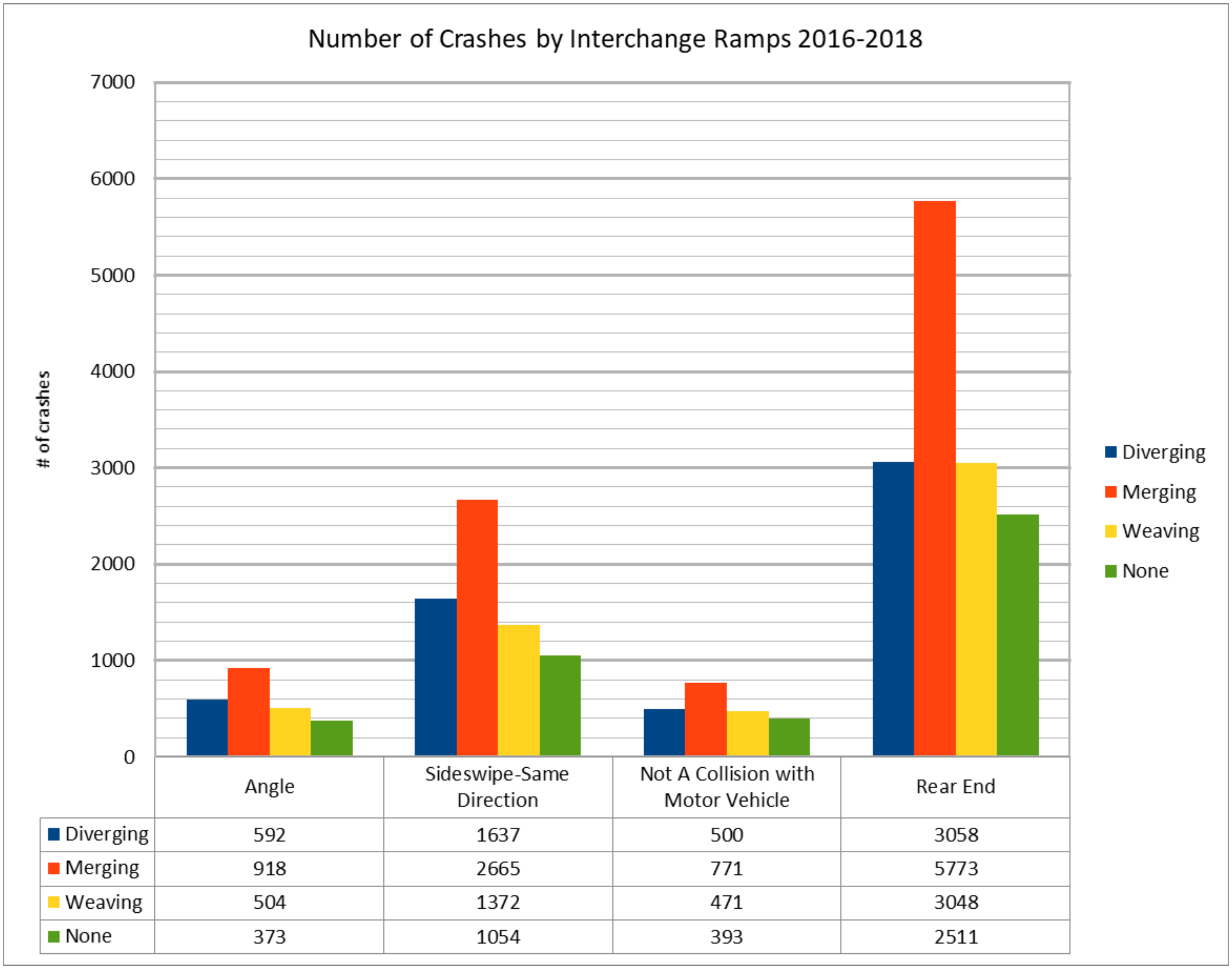
Figure 16. Map of I-285 accident data with closeup of interchange.

This observation was confirmed by grouping the crashes based on road segment type. There were 377 separate road segments with 158 segments not having any interchange association and 219 that did. In ArcGIS each segment that was observed to contain either an exit and/or an entrance ramp(s) was marked. In the attribute table a new column was named intersect and if a road segment had an exit ramp it was given a value of -1 to signify a diverging lane, a 1 if there was an entrance ramp resulting in a merging lane, a 2 if the segment contained both an exit and entrance ramp or the segment was within an interchange between exit and entrance ramps, and left 0 for all other road segments.

An illustration of the classification of different road segments can be found in Figure 17. Merging road segments with an entrance lane are marked in dark blue, diverging road segments with an exit lane are marked in red, weaving road segments that have both exit and entrance lanes or are within an interchange between diverging and merging road segments are marked in light blue. All other road segments are in beige.

Figure 17. Classification of road segment based on position in interchange.

It is clear from Figure 18, most crashes occurred in a road segment with an entrance ramp creating a merging lane. Rear end crashes especially showed a dramatic amount of crashes in a merging road segment compared to the other crash types.



	Angle	Sideswipe-Same Direction	Not A Collision with Motor Vehicle	Rear End
Diverging	592	1637	500	3058
Merging	918	2665	771	5773
Weaving	504	1372	471	3048
None	373	1054	393	2511

Figure 18. Bar graphs of number of crashes by crash type and road segment type.

Analysis

Two statistical techniques, the Local Moran's I measure and logistic regression, are used to examine the relationship between congestion and road accident frequency. The Local Moran's I measure determines if there are significant clusters or outliers from the accident frequency and the location of the accident hotspots found. The logistic regression tested the strength of the association between chosen variables and their significance level.

Local Moran's I

The ArcGIS spatial statistics Local Moran's I tool analyzes data for hotspots based on the number of crashes. Local Moran's I is computed by Eq. 1. The results show low-low clusters, where low value features are surrounded by low value neighbors, and high-high clusters, where high value features are surrounded by high value neighbors. Low-high clusters signify significant low value when compared to high value neighboring features and high-low clusters signify significant high value when compared to low value neighboring features ("How Cluster," n.d.).

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^{n} w_{i,j}(x_j - \bar{X}) \quad (1)$$

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^{n} (x_j - \bar{X})^2}{n-1} \quad (2)$$

where,

$\bar{x}$: mean of x;

x_i: value of x at location i;

x_j: value of x at location j;

w_{ij}: the elements of the weight matrix;

n: number of observations;

Index values range from -1.0 to 1.0, with a negative value indicating a negative spatial correlation and a positive value indicating a positive spatial correlation. The expected Moran's I_i is computed by Eq. 3.

$$E[I_i] = -\frac{\sum_{j=1,j\neq i}^{n} w_{ij}}{n-1} \tag{3}$$

When Ii, the Moran's I value, is greater than E[Ii], it indicates a positive spatial correlation. If Ii is less than E[Ii] then it indicates a negative spatial correlation. If location j is next to location i, then Wij has a weight of 1.0, if else 0.0. The z-score of I_i is calcuated by Eq. 4,

$$z_{I_i} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}} \tag{4}$$

The variance of I_i is calculated by Eq. 5.

$$V[I_i] = E[I_i^2] - E[I_i]^2 \tag{5}$$

Logit Regression

Logit regression, also known as logistic regression, is a popular algorithm for solving a classification problem. A binomial logistic regression is when the dependent variable is dichotomous. The logistic regression estimates the probability of one of two outcomes. This method was chosen since the crash data has a discrete outcome, either a crash happens, or it doesn't. If a crash occurred, it was coded as 1, if not, then it was coded as 0.

A logit regression model was fitted using the statsmodel package from Python. The resulted coefficients were considered significant when $p<0.05$.

Explanatory variables would include curvature of the road, if the road segment has an exit or entrance ramp, current lighting, and road condition. For this experiment, data from March 2018 was used. March 2018 had the most accidents of the eight-month period of camera data recording. This resulted in 649 crashes from the police data and 2 million rows of traffic speed from the cameras.

The vehCount per lane was calculated by taking the number of vehicles recorded by the camera every five minutes divided by the number of lanes at the road segment the camera was located.

Lag_speed was the speed of traffic during the previous five-minute interval before the crash was recorded to have occurred. Lag_acc was the acceleration of traffic during the previous five-minute interval before the crash.

Table 1. Descriptive statistics of variables

VARIABLES	Min	Mean	Max
LAG_SPEED (m/s)	0	28.64	52.5
LAG_ACC (m/s^2)	-0.1574	0.0014365	0.1481
CURVA	0	0.0009032	0.01042
VEHCOUNT/LANENUM	0	4.667	24.5

If the road segment was positioned along an interchange, it was sorted into three categories. Merging was denoted with a 1 and meant the road segment was receiving incoming traffic from an entrance ramp. Diverging was coded as a -1 and meant traffic was decreased from an exit ramp leading off the interstate. Weaving was coded as a 2 and meant the road segment was either between a Merging and Diverging segment within the interchange or contained both entrance and exit ramps. Non-interchange aligned road segments were coded as 0.

Results

The hot-spot analysis of the spatial pattern of the traffic crashes on the I-285 from 2016 to 2018 focused on identifying high-high clusters, road segments with high accident numbers and neighboring segments with high accident numbers, or high-low outliers, road segments with high accidents numbers and neighbors with low accident numbers. Both would identify areas with a concentration of high accidents with can then be compared to the areas of congestion mapped above.

The logistic regression for crash occurrence on I-285 used crash data from October 1, 2017 to June 25, 2018 to match the time limits of the collected speed data. The speed and acceleration of traffic, traffic density, and three road segment types based on their position regarding interchanges were the variables modeled for correlation with accident occurrence per road segment. The information was segregated spatially.

Using the Local Anselin Moran's I method, the results showed the high-high clusters in Table 2 and Figure 19 below. Interestingly there were no high-low clusters, meaning there were no road segments with a high number of accidents compared with neighboring low accident number road segments.

The positive Local Moran's I index (LMiIndex) value indicates the value is part of a cluster, the higher the number the more similar the neighboring values. All the clusters had positive LMiIndex values which meant no outliers. The Count_1 column shows number of accidents that occurred in that road segment. The NNeighbors value is the number of neighbors, all neighbors were limited to road segments within 1km to maintain local adjustment weight. Only values with a Local Moran's I p-value (LMiPValue) of less than 0.05 were considered significant.

Table 2. High-high clusters of accidents.

SOURCE_ID	Count_1	LMiIndex	LMiZScore	LMiPValue	COType	NNeighbors
75	113	1.078949	3.44736	0.01	HH	5
123	409	5.404191	1.880554	0.048	HH	5
134	243	5.654452	5.19767	0.006	HH	5
137	426	6.72903	2.633219	0.028	HH	5
141	150	1.709725	2.42811	0.032	HH	5
195	115	1.021248	2.728536	0.03	HH	5
237	116	1.763449	4.067834	0.02	HH	5
286	157	1.976411	3.159171	0.02	HH	5
308	255	3.505724	2.461291	0.042	HH	5
345	128	1.769842	3.595203	0.016	HH	5
362	261	5.28387	3.876627	0.008	HH	5

The location of the clusters are highlighted in Figure 19. Road segment location of high-high clusters of crashes highlighted in blue. below. None of the identified clusters were in the southern half of the I-285 loop. There are 11 high-high clusters total but only nine appear show on the map. This is due to the proximity of two pairs of segments, where one is clockwise and one counter-clockwise that leads to overlapping when the map scale adjusts. The two overlapping segments are located to the east on the far right of Figure 19 near Doraville. Every segment was located on a road segment with an entrance or exit ramp.

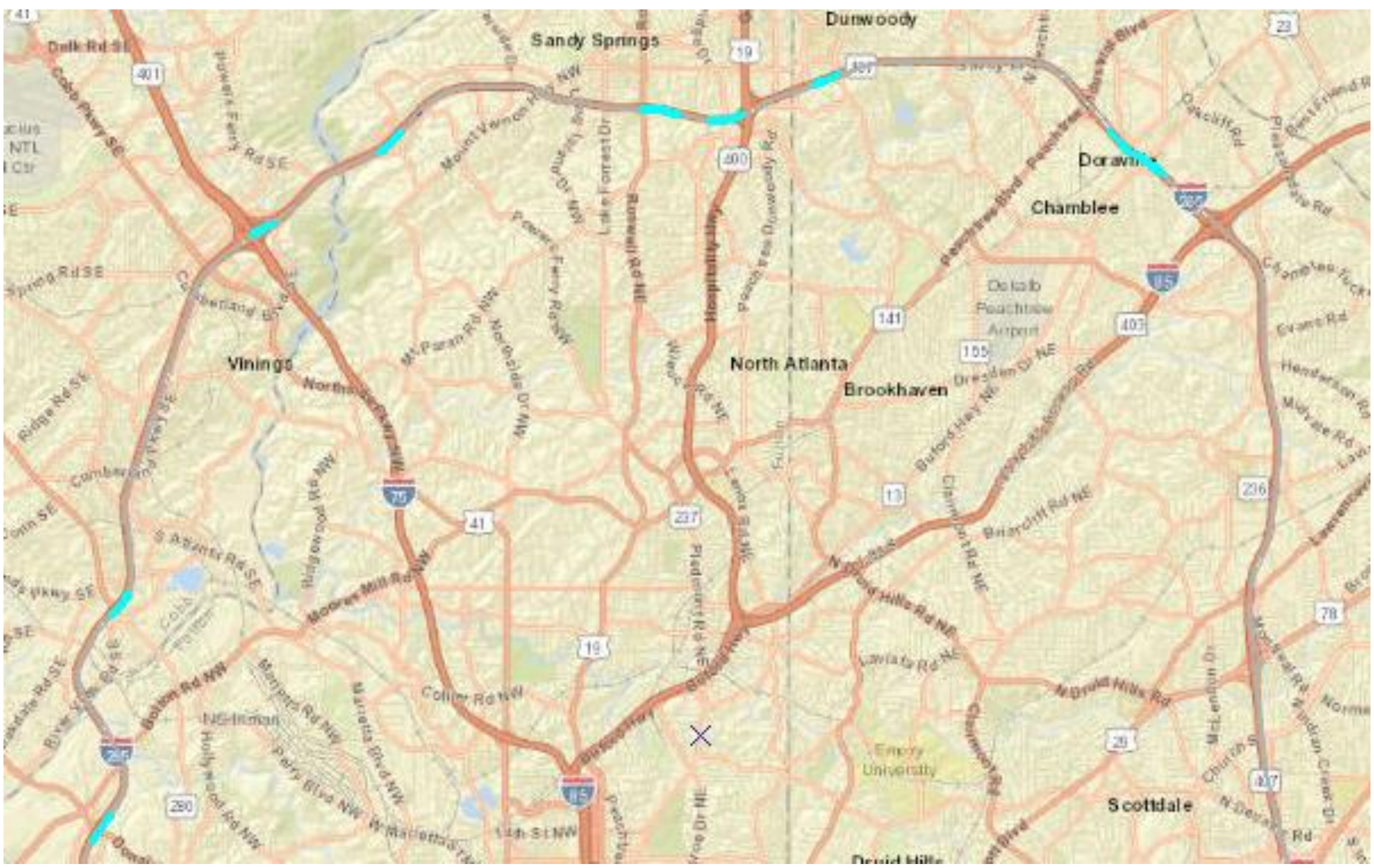


Figure 19. Road segment location of high-high clusters of crashes highlighted in blue.

Table 3 . Logit Regression Results for all crashes

Logit Regression Results

Dep. Variable:	crash	No. Observations:	1724
Model:	Logit	Df Residuals:	1718
Method:	MLE	Df Model:	5
Date:	Tue, 30 Jul 2019	Pseudo R-squ.:	0.1183
Time:	01:08:29	Log-Likelihood:	-791.54
converged:	True	LL-Null:	-897.79
Covariance Type:	nonrobust	LLR p-value:	6.003e-44

	Coef	std err	z	P>|z|	[0.025	0.975]	Odds ratio
Intercept	-1.5050	0.281	-5.354	0.000	-2.056	-0.954	0.2220
vehCount/lane	0.1258	0.023	5.445	0.000	0.081	0.171	1.1341
lag_speed	-0.0536	0.008	-7.139	0.000	-0.068	-0.039	0.9478
Merging	1.6401	0.170	9.676	0.000	1.308	1.972	5.1557
Diverging	0.9650	0.194	4.974	0.000	0.585	1.345	2.6248
Weaving	1.4577	0.198	7.352	0.000	1.069	1.846	4.2961

The crash prediction model based on traffic density, traffic speed five minutes before the crash, and type of road segment based on exit and/or entrance ramp presence can be seen above in Table 3.

The acceleration of traffic of the road segment five minutes before a crash was not statistically significant. The curvature of the road segment was also not statistically significant.

The vehcount/lane, merging, diverging, and weaving are significant, and all had positive correlations with crash numbers while lag_speed had a negative correlation.

The highest odds ratio was for road segments with merging lanes and was 5.1557. This can be interpreted as the presence of an entrance ramp increases the risk of crashes more than five-fold based on the given data selection. Diverging road segments had an odds ratio of 2.6248 which was around half of merging road segments. Though diverging road segments are less likely to have crashes compared to merging road segments there is still an associated 2.6248 increase of crashes. The odd ratio for weaving road segments was 4.2961 which was between 5.1557 and 2.6248.

The traffic volume, represented by vehicle count divided by number of lanes (vehcount/lane), showed a positive coefficient, which is as expected, since higher volume indicates a higher exposure.

The lag_speed had an odds ratio of 0.9478, so the odds of a crash are 5.22% less for every 1 m/s increase in traffic speed.

Therefore, according to this model, the increase in traffic volume and presence of interchanges were correlated with an increase in crash frequency and an increase in traffic speed five minutes before the crash was correlated with a decrease in the probability of a crash.

Conclusions and Recommendations

There are two main congestion patterns on I-285, separated by weekday and weekends. The congestion 'clouds' during the weekday last several hours and occur in the northern half of the I-285 loop. The congestion periods occur twice a day from around 6:00-11:00am and 1:00-8:00pm. During congestion periods the direction of traffic travels north in the morning and south-bound in the afternoon. The implication is the high congestion periods are from the commuting traffic heading toward Atlanta via Interstates 285 and 85 North in the morning and then returning home in the evening.

The hot-spot analysis shows clusters of road segments with high crash numbers occur in areas that also have high congestion. Further analysis should be focused on the 11 road segments with high accident rates for causation diagnosis. The hotspots appear in the same locations as the identified congestion. This supports the idea that congestion and accidents are positively linked. The hotspots occurred on road segments with interchanges meaning there is evidence of interaction in the relationship between the two variables and crash frequency.

The correlation coefficients obtained from the logistic regression confirm the positive correlation between low speed with crash probability. The probability of a crash is higher if there is already congestion present. This is evidence that congestion has a greater effect on accident frequency than accident frequency does on congestion.

The road segments near or within an interchange are positively correlated with crash frequency. The correlation makes sense as vehicles entering traffic may not be able to merge easily. Diverging road segments had a positive relationship but are less correlated than merging road segments. This may be due to it being easier to exit the interstate than entering as exiting ramps allow a gradual deceleration.

Further research should explore a nested logit model structure to reveal the effect based on different types of crashes. A nested logit model can be used to isolate factors such as crash occurrence by type which could reveal specific factors contributing to different types of accidents.

Research Limitations

There are two main limitations in this study. First, speed data from the I-285 cameras are missing due to periods of malfunctioning of some cameras. The acceleration is calculated over inconsistent time intervals and is not a perfect representation of real time changes. Second, the recorded time of accident from the crash data might not be accurate as it reflects the officer’s arrival time more than the time the accident occurred.

Second, the analysis model used in the study has limitations. Although the logit model provides insight into the influential variables on crash frequency, it did not capture heterogeneity across space, time, and crash type. A multivariate spatio-temporal model should be investigated in future studies to provide a more comprehensive analysis.

References

Abdulhafedh, A. (2017). A novel hybrid method for measuring the spatial autocorrelation of vehicular crashes: Combining Moran’s Index and Getis-Ord Gi* Statistic. *Open Journal of Civil Engineering, 7*, 208-221. doi: 10.4236/ojce.2017.72013

Aghajani, M.A., Dezfoulian, R.S., Arjroody, A.R., & Rezaei, M. (2016). Applying GIS to identify the spatial and temporal patterns of road accidents using spatial statistics (case study: Ilam province, Iran. *Transportation Research Procedia.* https://doi.org/10.1016/j.trpro.2017.05.409

American Transportation Research Institute (ATRI). (2018). *Top truck bottlenecks.* Retrieved from https://atri-online.org/research/results/congestion_mobility/2018bottleneck/bn001-2018.pdf

Department of Revenue. (2019). *Motor Vehicle Division.* Retrieved August 5, 2019, from https://mvd.dor.ga.gov/motor/stats/renewalsstats.aspx

El-Basyouny, K. & Sayed, T. (2009). Collision prediction models using multivariate Poisson-lognormal regression. *Accident Analysis & Prevention.* https://doi.org/10.1016/j.aap.2009.04.005

Elvik, R., Sagberg, F., & Langeland, P.A. (2019). An analysis of factors influencing accidents on road bridges in Norway. *Accident Analysis & Prevention*. https://doi.org/10.1016/j.aap.2019.05.002.

Environmental Systems Research Institute, Inc. (ESRI). (n.d.). Spatial Autocorrelation (Global Moran's I). *Tool Reference.* Retrieved from https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/spatial-autocorrelation.htm

Environmental Systems Research Institute, Inc. (ESRI). (n.d.). How Cluster and Outlier Analysis (Anselin Local Moran's I) works. *Tool Reference.* Retrieved from https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-cluster-and-outlier-analysis-anselin-local-m.htm

Federal Highway Administration (FHWA). (2019). *FHWA Route Log and Finder List.* Retrieved from https://www.fhwa.dot.gov/planning/national_highway_system/interstate_highway_system/routefinder/table02.cfm

Georgia Department of Transportation (GDOT). (n.d.). Road & traffic data. *Travelers.* Retrieved from http://www.dot.ga.gov/DS/Data

Georgia Department of Transportation (GDOT). (2018). *Georgia's traffic monitoring guide. Georgia's Traffic Monitoring Program. Office of Transportation Data.* Retrieved from http://www.dot.ga.gov/DriveSmart/Data/Documents/Guides/2018_Georgia_Traffic_Monitoring_Program.pdf

Georgia Governor's Office of Highway Safety (n.d.). *Data overview.* Retrieved from https://www.gahighwaysafety.org/statistics/mannerofcollision.pdf

Getis, A., & OrdJ.K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis, 24*, 189-206.

Glennon, J.C., Neuman, T.R., & Leisch, J.E. (1983). Safety and operational considerations for design of rural highway curves. *Engineering Research and Development Bureau.* Retrieved from https://trid.trb.org/view/273415

Green, C.P., Heywood, J. S., & Navarro, M. (2016). Traffic accidents and the London congestion charge, *Journal of Public Economics, 133*, 11-22. https://doi.org/10.1016/j.jpubeco.2015.10.005.

INRIX (2019). INRIX: congestion costs each American 97 hours, $1,348 a year*, Press Releases.* Retrieved from http://inrix.com/press-releases/scorecard-2018-us/

Interstate 285. (2004, May 7). *Interstate 285*. Retrieved August 5, 2019 from http://www.kurumi.com/roads/3di/i285.html

Keum, C. B. (2016). *Analysis of road traffic crashes and injury severity of pedestrian victims in the Gambia.* (Master's thesis, University of Iowa). https://doi.org/10.17077/etd.zqred0ba

Kim, S., Song, T., Rouphail, N.M., Aghdashi, S., Amaro, A., & Gonçalves, G. (2016). Exploring the association of rear-end crash propensity and micro-scale driver behavior, *Safety Science, 89*, 45-54. https://doi.org/10.1016/j.ssci.2016.05.016

Kretlow, B. (2017). *Development of spatiotemporal congestion pattern observation model using historical and near real time data* (Master's thesis, Kennesaw State University).

Lave, C. A. (1985). Speeding, coordination, and the 55 MPH limit. *The American Economic Review, 75*(5),1159-1164.

Liu, C. & Sharma, A. (2017). Exploring spatio-temporal effects in traffic crash trend analysis. *Analytic Methods in Accident Research*. Retrieved from https://www.sciencedirect.com/science/article/pii/S2213665717300210

Liu, C. & Sharma, A. (2018). Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity. *Analytic Methods in Accident Research.* Retrieved from https://www.sciencedirect.com/science/article/pii/S2213665718300162?via=ihub

Lord, D., Manar, A. & Vizioli, A. (2005). Modeling crash-flow-density and crashflow-V/C ratio relationships for rural and urban freeway segments. *Accident Analysis & Prevention, 37*(1), 185-199.

Ma, X., Chen, S., & Chen, F. (2017). Multivariate space-time modeling of crash frequencies by injury severity levels. *Analytic Methods in Accident Research.* Retrieved from https://www.sciencedirect.com/science/article/pii/S221366571730009X

Malin, F., Norros, I., & Innamaa, S. (2019). Accident risk of road and weather conditions on different road types, *Accident Analysis & Prevention, 122*, 181-188. https://doi.org/10.1016/j.aap.2018.10.014

National Center for Statistics and Analysis. (2019). Early estimate of motor vehicle traffic fatalities for 2018. *Crash Stats Brief Statistical Summary. Report No. DOT HS 812 749.* Washington, DC: National Highway Traffic Safety Administration.

National Safety Council (NSC) (n.d.). *Preliminary Estimates*. Retrieved from https://injuryfacts.nsc.org/motor-vehicle/overview/preliminary-estimates/

National Safety Council (2017). *Injury facts chart. Injury Facts® 2017 Edition, pages 40-43*. Retrieved from https://www.nsc.org/work-safety/tools-resources/injury-facts/chart

Office of Highway Policy Information (OHPI). (2018). Motor vehicle registrations, by states, 1900-1995. *Highway Statistics Summary To 1995*. Retrieved from https://www.fhwa.dot.gov/ohim/summary95/mv201.pdf

Pasidis, I. (2019). Congestion by accident? A two-way relationship for highways in England, *Journal of Transport Geography, 76*, 301-314. https://doi.org/10.1016/j.jtrangeo.2017.10.006

Plug, C., Xia, J., & Caulfield, C. (2011). Spatial and temporal visualisation techniques for crash analysis. *Accident Analysis & Prevention.* https://doi.org/10.1016/j.aap.2011.05.007

Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. (2011). Spatio-temporal clustering of road accidents: GIS based analysis and assessment. *Procedia - Social and Behavioral Sciences*. Retrieved from https://www.sciencedirect.com/science/article/pii/S2095756417301988

Shafabakhsh, G., Famili, A., & Bahadori, M. (2017). GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran. *Journal of Traffic and Transportation Engineering (English Edition).* Retrieved from https://www.sciencedirect.com/science/article/pii/S2095756417301988

Vogt, A. & Bared, J. (1998). Accident models for two-lane rural segments and intersections. *Transportation Research Record, 1635*(-1), 18-29.

Wang, C. (2010). *The relationship between traffic congestion and road accidents: an econometric approach using GIS* (Master's thesis). London, UK: Kingston University. doi: 10.13140/RG.2.1.3180.0087.

World Health Organization (WHO). (2018). Road traffic injuries: Key facts. *Newsroom Fact sheets.* Retrieved from https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Xu, C., Bao, J., Wang, C., & Liu, P. (2018). Association rule analysis of factors contributing to extraordinarily severe traffic crashes in China, *Journal of Safety Research, Volume 67*, 65-75. https://doi.org/10.1016/j.jsr.2018.09.013.

Zhan, X., Aziz, H., Abdul M., & Ukkusuri, S. V. (2015). An efficient parallel sampling technique for Multivariate Poisson-Lognormal model: Analysis with two crash count datasets. *Analytic Methods in Accident Research, 8,* 45-60.

Zhang, Y., Li, Z., Liu, P., & Zha, L. (2011). Exploring contributing factors to crash injury severity at freeway diverge areas using ordered probit model. *Procedia Engineering, 21*, 178 – 185.