Measuring Motor Vehicle Collision Vulnerability in the Greater Toronto Area: Final Report

Zhihao Wang (20695562)

Yu Shi Wang (20575661)

Jiachen Wei (20695056)

Stephanie Wen (2048485)

University of Waterloo

GEOG 481 (Winter 2018)

Dr. Su-Yin Tan

Monday April 9, 2018

Table of Contents

ABSTRACT

Relative recent research has been conducted on accident rates or injury-severity rates. Focusing more on the factors that cause these accidents and injuries. According to the World Health Organization in 2013, road collisions are responsible for 1.24 million deaths worldwide. This study identifies highway segments in the Greater Toronto Area and road segments in the City of Toronto that are most vulnerable to motor vehicle collisions. A multi-criteria analysis was conducted on driver-, weather-, and road-related factors. The results of the multi-criteria analysis were displayed on a web-GIS application for government. All the datasets used were gathered from open data portals, which limited our ability to get detailed data, such as the exact location of the collisions. This introduced a Modifiable Areal Unit Problem. The results of this study determine that more collisions happen on a clear day, regardless of being on a highway or road network. Potential suggestions to decrease car accidents include better signage on highways and to raise awareness for driving safety in busier driving conditions.

Chapter 1 Introduction

1.1 Background

Many studies have been conducted on accident frequencies; however, few studies have examined risk factors for motor vehicle crashes on the 400-series highways in Ontario (Rzeznikiewiz, Tamim, & Macpherson, 2012). Road traffic collisions are responsible for 1.24 million fatalities and up to 50 million non-fatal injuries worldwide, in 2010 (Rothman, et al., 2015). Fatal motor vehicle collisions are the sixth leading cause of death for Canadians. Fatal motor vehicle collisions in 2004 had an average social cost of 15.7 million (Rzeznikiewiz et al., 2012). According to the Federal Highway Administration (2017) in the United States of America, 1.3 million motor vehicle crashes from 2005 to 2014 were caused by weather-related factors. Road-related factors such as the geometry of the road, traffic volume, and the number of lanes will also be considered for this study. Increased accidents are correlated to the number of highway lanes according to Highway Safety Information System for the State of Illinois (Noland & Oh, 2004).

Using these factors, this study will determine collision vulnerability on the provincial-level highways in the Greater Toronto Area (GTA) and road network for the City of Toronto. This project will assess where motor vehicle accidents occur and the possible leading factors behind these collisions to help government facilities identify areas in need. Developing a web map to display the results of this study, using human-related and environmental factors.

1.2 Mission Statement

The mission statement of the project is to showcase collision vulnerability on highway sections within the Greater Toronto Area and road segments within the City of Toronto using a web GIS, as well as indicating the accident-prone areas using different driver-related and environmental factors, to inform drivers and government facilities.

1.3 Objectives

The following is a list of objectives required to reach the project goal:

1. Measure collision vulnerability based on driver-related and environmental factors
2. Create a visual display of the data results in charts and on a web-GIS map
3. Make suggestions based on the results to the municipal or provincial government to potentially implement changes to reduce the risk of motor vehicle accidents

1.4 Report Outline

This report contains eight chapters. Chapter 1 introduces the project motivation, objectives and finally, outlines the project structure. Chapter 2 reviews previous literature on how environmental, human and road-related factors motor vehicle collisions. The research gap is discussed at the end of literature review. Chapter 3 presents the study area in the GTA and the City of Toronto. In Chapter 4, data and software used is discussed, and how it helped to achieve the project objectives. Chapter 5 describes the detailed methods used to process the data, measure the collision vulnerability, and develop the Web-GIS. Chapter 6 demonstrates how the Web-GIS visualizes the results from Multi-Criteria Analysis (MCA) and what important findings are discovered. Chapter 7 discusses the results and limitations existing in the datasets and methods for future improvement. Chapter 8 summarizes the entire project regarding the motivations, objectives, and how these objectives are accomplished, as well as suggestions for future improvement.

Chapter 2 Literature Review

2.1 Driver-related Factors

Many accidents are caused by risk taking, or drivers overestimating their own abilities and limits (Iverson, 2004). Driver-related factors such as age, impaired and distracted driving, and nighttime driving are factors affecting highway accidents that directly impact how the driver evaluates risk. Younger drivers are more inexperienced, leading to more mistakes than a more capable driver would make, while elderly drivers potentially underestimate how much health conditions can affect driving.

There have been few studies conducted for the Ontario's 400-series highways and the risk of dying in a crash (Rzeznikiewiz et al., 2012). The objective of this study is to identify areas prone to highway accidents within the GTA, and road collisions within the City of Toronto. There are many factors as to why motor vehicle collisions occur. A study of Ontario highway crash records showed that drivers under the age of 30 years are more likely to be involved and injured in single vehicle crashes, which is associated with driving faster and more recklessly in quieter highway conditions (Lee & Li, 2014). Such highway conditions can be found at nighttime, where the volume of commuting vehicles is much lighter than the afternoon to evening rush hour where traffic is much heavier and slower.

Highway accidents caused by either impaired or distracted driving are directly related to risk perception, since in these cases, the driver would be operating a vehicle at their own discretion despite facing potential distractions or impairments. Accidents involving alcohol also almost exclusively occurred at night. The combination of sleep loss and alcohol were more likely to cause accidents, even if they consumed below the legal limit (Philip & Åkerstedt, 2006). Drivers who had cell phone conversations while driving, were at the same risks as those who had a blood alcohol concentration level of 0.08. Hands-free cellphone devices are just as detrimental as handheld, as it interferes with cognitive demands, rather than distractions due to manipulations (Charlton, 2009).

2.2 Weather-related Factors

Weather acts through visibility interference, precipitation, and extreme temperature that affect drivers' capabilities, road surface conditions, traffic flow, and relief efficiency. Precipitation and temperature are among the top considerations in current literature (Ahammed & Tighe, 2011; Ivan et al., 2012). According to Federal Highway Administration (2017), nearly 1,300,000 vehicle crashes in the US from 2005 to 2014 were caused by weather-related factors, among which wet pavement accounted for 73% of traffic crashes and precipitation accounted for 46%.

Extreme ambient temperature is a factor that significantly affects driving performance. Researchers who examined regions with similar temperature to the GTA found significant positive association between heat and car collisions during summer in Indiana, USA (Malyshkina, et al, 2009), and during heat wave days in Catalonia, Spain (Basagaña et al., 2015). Heat challenges drivers physically and intellectually (Ramsey, 1995); however, understanding its effect on highway accidents requires adjustment for other factors. For example, Basagaña et al. (2015) performed sensitivity analysis by excluding concurrent rain, vacation, and drugs and alcohol when studying Catalonia, Spain.

Although many studies demonstrated the potential of different probability models to analyze the weather-related factors, they did not consider a sufficient range of weather-related factors. Andreescu and Frost (1998) analyzed the effects of rain, snow, and temperature on traffic accidents in Montreal from 1990 to 1992 using multivariable linear regression. The results demonstrated that snow was a leading variable as accidents sharply increased after snow falls. When Andreescu and Frost (1998) explored the correlation between daily accident frequency and individual variables, it is noted that their simple linear regression model could not fully reflect the complex relationship between accidents and weather conditions (Montgomery et al., 2012). Comparatively, Sherif (2005) explored the relationship between road surface temperature, surface moisture, and accidents from November 2001 to March 2002 in the City of Ottawa. The researcher employed an empirical Bayesian technique—a probabilistic model—to analyze collisions and indicated that wet surfaces were more hazardous when the temperature ranges from -2 to +1 ºC (95% of significant confidence). The results showed the potential of advanced probability models when analyzing complex and nonlinear relationships in the real world, but the research is limited on the diversity of data source as it merely considered temperature and humidity factors. Similarly, when Peng et al. (2017) investigated the impacts of reduced visibility on traffic accidents, they also used a sophisticated probability model (Log-Inverse Gaussian regression) but limited their factors to visibility.

2.3 Road-related Factors

Factors such as road geometry, traffic volume, and the number of lanes on highways are commonly considered in current literature. Sharif Tehrani, Cowe Falls, & Mesher (2017) suggested that car collisions in Alberta are more likely to occur on curves, graded sections of two-lane highways, and wet four-lane highways.

An increased risk in accidents was found to be correlated to an increase in the number of highway lanes in a study done using data from the Highway Safety Information System for the State of Illinois (Noland & Oh, 2004). While lane expansions initially improve road conditions and overall safety by relieving traffic congestion, it increases the risk over time as traffic congestion becomes more frequent (Kononov et al, 2008). This is mainly due to an increase in lane change frequency, and combined with a growth in car volume, this can be associated with the rising rate of multi-vehicle accidents. Noland & Oh (2004) also found that separating the highway into collectors and express lanes with a median or barrier reduced collision opportunities that are generally associated with larger number of highway lanes.

The type of vehicle involved in a car accident can be compared to the type of accident that occurred, which can then be related to road geometry. Side sweeps, which happen on multilane highways with driving recklessly onto neighbouring lanes, are more common for cars than trucks, as found in a study of Ontario provincial highway crash records (Lee & Li, 2014). Ayati & Abbasi (2011) investigated how traffic per lane (passenger, light non-passenger, and heavy) increases frequency and severity of accidents to varying extents in Iran. Another study found that accidents between a smaller vehicle and a larger vehicle, such as a car-truck accident, are more probable to be more severe in terms of fatalities, injuries, and overall damages, rather than the type of vehicle directly affecting the probability of accidents (Lee & Li, 2014).

2.4 Applications of GIS on Analyzing Collisions

Most literature reviews used statistical analysis for these factors, rather than GIS tools. Some of the traditional methods were some type of regression analysis; Poisson, multivariate, or logistic, to name a few, and confidence intervals, and odds ratios. The mores spatial analyses for road collisions often involved a clustering analysis of some sort like Kernel Density Estimation (KDE), Moran's I and Getis-Ord.

The most popular method for a spatial analysis of road collisions are Moran's I statistic and Getis-Ord (Satriaa & Castro, 2016). Moran's I measures the spatial dependency of the accident location; it is an index of values from -1 to +1. It also determines whether the spatial pattern clusters are random or dispersed. A higher positive value represents a greater degree of spatial clustering of similar values. A negative value shows spatial spreading and 0 means there is a randomly distributed pattern of the feature. Getis-Ord is a clustering analysis which represents cold spots and hot spots. Cold spots represent low values of clustering in a location, while hot spots represent high values of clustering (Soltani & Askari, 2017). This indicates how close these collisions were to each other and whether there was spatial correlation. Proximity analysis was also introduced for the case of collision point data. One of the proximity analysis mentioned was Thiessen polygons, are used to create polygons generated from a set of points for inclusion or amalgamation (Lai & Chan, 2004). Another study showed accident black spots by using KDE. They used predefined parameters for their KDE which then outputs a raster map (Hegyi & Attila Borsos, 2017). KDE places a symmetrical surface over each point then determines the distance from the reference location to that incident point. It then sums all the values for all the surfaces for that reference location and repeats, resulting in our density (Anderson, 2009).

For better visualization purposes prior to a hotspot analysis road segments were divided into one kilometre segments, if there is a segment exceeding the threshold it would be considered a hot spot (Gundogdu, 2010). Another study divided their road segments for locational referencing in the database, hence allowing the attributes to be queried their GIS system (Faghri & Raman, 1994). Faghri and Raman (1994) used Boolean overlays to display the accident data based on user’s inquiry. For example, fatalities at intersections, would result in fatalities AND intersections. Associative analysis was often used when more than one map layer is being shown, it is also known as overlay analysis. These overlay operations used map algebra and logical conditions specified in Boolean operations (AND, OR, XOR, NOT, etc.) (Lai & Chan, 2004). This can also be seen as a MCA.

2.5 Research Gaps

Current literatures are limited in data sources and analysis methods. Few studies investigated how accidents were affected by the above-mentioned factors in Ontario. This is probably since the data for these factors were not available, possibly due to privacy concerns. For example, Toronto collision data had six accidents located in one intersection, the date and time were all the same. It is likely that there were six cars involved, however each car is not marked individually. Hence loss of information as to how the accident occurred, whether if it was a rear-ending collision one after another. When age was recorded in a range and not the specific drivers' age. Driver condition was also marked as normal or abnormal, this attribute is very vague as to could the driver had been intoxicated due to prescribed drugs or illicit drugs, alcohol was marked as a yes or no. Additionally, the studies we reviewed dominantly focus on regression analysis.

Chapter 3 Study Area

3.1 Greater Toronto Area

The study area for this research is the GTA, and the City of Toronto. The GTA is located in the Golden Horseshoe, in Southern Ontario (Figure 1). The Golden Horseshoe is home to 8.1 million residents, approximately 25% of Canada's population (Nazzal, Rosen, & Al-Rawabdeh, 2012), and the GTA alone has an estimated population of 5.5 million residents (Wentworth et al., 2015). The GTA includes the City of Toronto, and four surrounding regional municipalities: Durham, Halton, Peel, and York (Mitra, N.Buliung, & Faulkner, 2010). The GTA encompasses 700 km^2 (Wentworth et al., 2015), is the seventh largest urban area and the fifth most populous city, in North America (Nazzal et al., 2012).

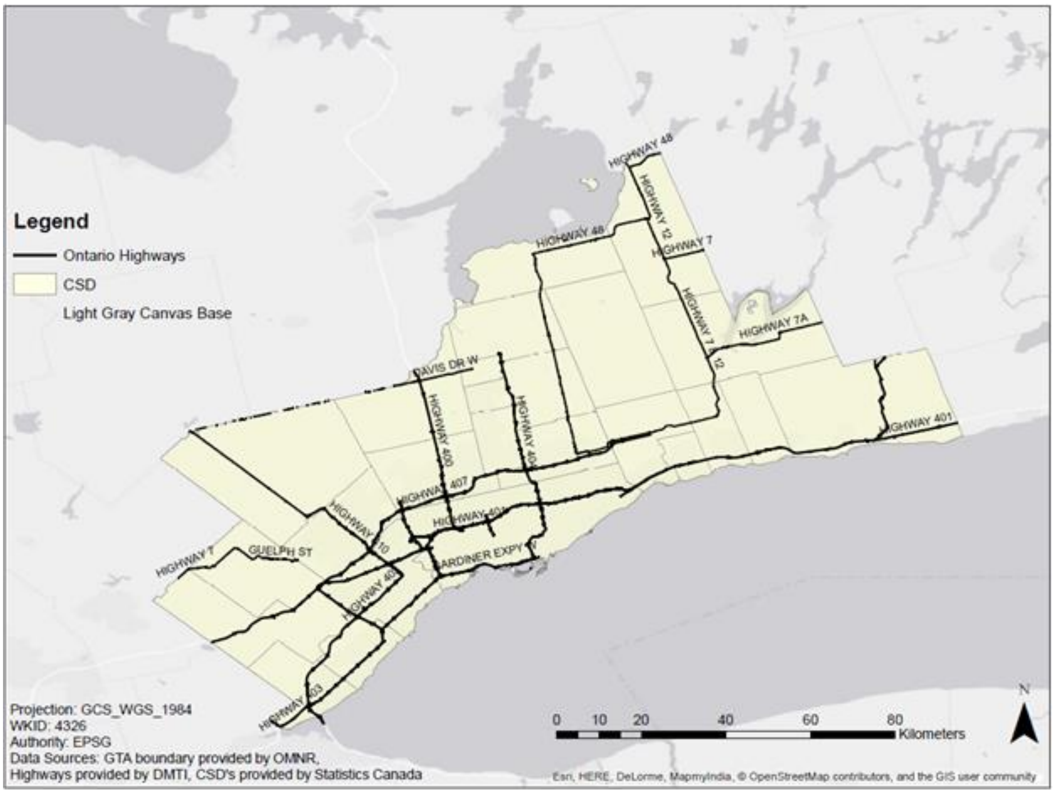

Figure 1: Location of study area, 400 Highway series bounded by the Greater Toronto Area

3.2 Toronto

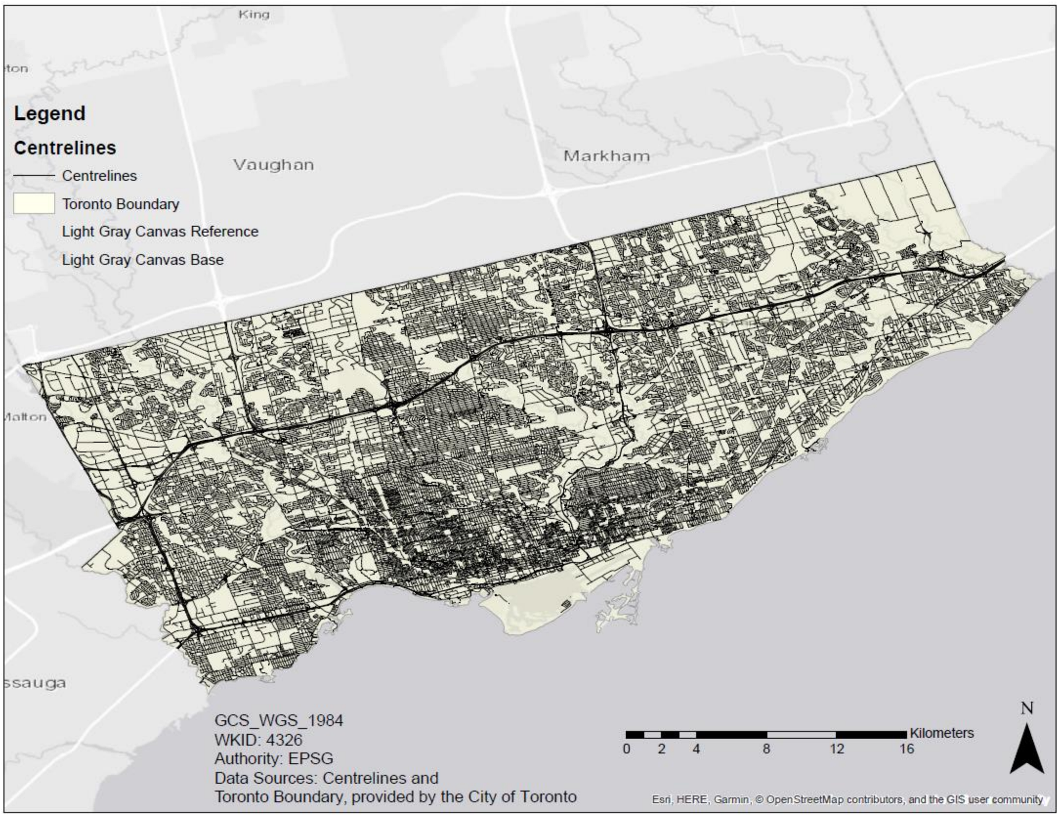

Figure 2: Second study area, centerlines bounded by the City of Toronto

The City of Toronto is the capital of Ontario, Canada, it is located on the shores of Lake Ontario (Figure 2). Toronto has a population of 2.5 million and is the largest urban centre in Canada (Zuo, et al., 2014). Toronto is the centre of the GTA, with Highway 401 and Don Valley Parkway (DVP) passing through the city. Highway 401 is one of the world's busiest highways, the DVP operates beyond its vehicle capacity of 60,000, some sections even carrying 100,000 vehicles a day (Nazzal et al., 2012).

Chapter 4 Data Sources and Software

4.1 Data Sources

Table 1: Data Documentations

Data Layer	Source	Data Type	Extent/Scale	Major Attributes	Year(s)
Automobile Collisions	Toronto Police Service	Point	Toronto	Visibility, light, Invage, Injury, DrivCond, Speeding	2007-2017
Collision data (geo-tagged)	Ontario Provincial Police	csv	Ontario	accident location (x,y), time, driver information, speed,vehicle type, severity,	1999-2010
Hourly weather (geo-tagged)	EnvCan	csv	Canada	weather station (x,y,z),visibility, humidity,wind (direction, speed),weather types	1999-2016
Toronto Centrelines	City of Toronto	Polyline	Toronto	linear features representing streets, walkways, rivers, railways, highways, etc	Sep-17
Ontario road networks	DMTI	Polyline	Ontario	street names, number of lanes, speed restrictions, toll booths	2009-2012
Highway linear referencing system	Ontario Provincial Police	Point	Ontario	Highway linear referencing points and road segments	2010

Data sources for this project are described in *Table 1*. The project looked at a consistent time-period from 2006 to 2010, for which all the sources can provide data. Two sources of the collision were selected in this project: one is the highway collisions from the

Ontario Provincial Police (OPP) while another is the Toronto collisions from the Toronto Police Service. Both collision datasets contain the locations and other environmental and human information of accidents. Weather-related information such as temperature and visibility provided by Environment Canada was linked to the collision site based on the distance to a weather station. To perform the Multi-Criteria Analysis, both road networks not only served as the base maps, but also provided road segments.

4.2 Software

ArcGIS Pro and ArcMap were used for viewing attributes and shapefiles. ArcMap was used to conduct spatial joins. ArcGIS Pro was used to delete fields, and Modelbuilder was used to automate these processes over multiple feature classes in a geodatabase.

RStudio was used to perform the process of data cleaning and data analysis. After joining the collisions to the nearest segments in ArcMap, the data table was exported as the Comma-separated values (CSV) file. MCA was conducted in R by creating efficient data structure. The final vulnerability values were then exported from R and joined to the shapefile based on the unique road ID.

QGIS was used to hide attributes from the user before creating the web map, these attributes were implemented into pop-ups. The qgis2web plugin was used to convert the final shapefiles to GeoJSON and to create a simple web map to work from. The exported web map uses Leaflet, a JavaScript library used for interactive web mapping. Notepad++ and Adobe Dreamweaver were used to code the web map.

Chapter 5 Methodology

This section describes the detailed approaches used to measure and visualize the collision vulnerability in the study area (Figure 3). Based on the literature, an inclusive range of factors was considered in this project, and the different combinations of factors were examined using MCA. To grasp a better understanding of collisions, the calculated collision vulnerability was visualized in Web-GIS. An overview of the methods is outlined as following: Section 5.1 demonstrates the process of data collection and data cleaning; Section 5.2 introduced the MCA which is the major step used to measure the impacts of different factors on car collisions. Finally, Section 5.3 shows the detailed steps to develop the Web-GIS for data visualization.

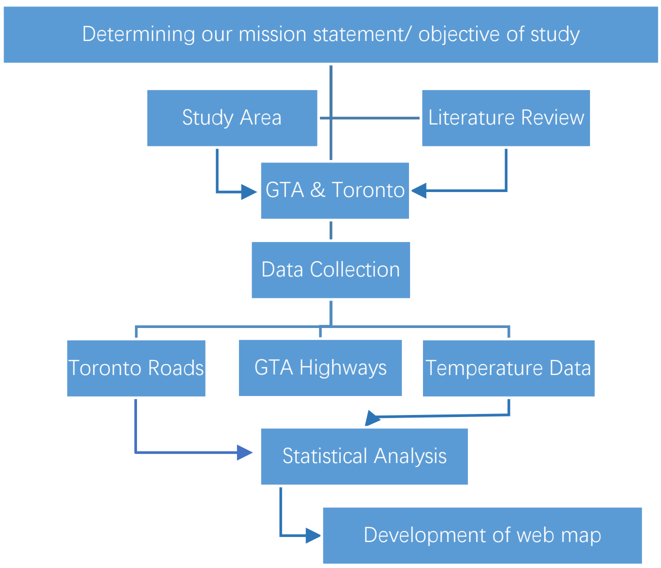

Figure 3: Methodology Outline

5.1 Data Pre-processing

5.1.1 Collisions in Highway

For highway collisions, all data were cleaned before performing any analysis. First, the collision records with missing values in collision ID and highway referencing points were removed. A total of 7,981 collisions were filtered out in this step. Second, collisions located outside of the GTA were removed. A total of 62,017 collisions were filtered out in this section. Lastly, 230 collisions with duplicated collision ID were deleted. After the cleaning process, a total of 61,787 were left for further analysis. Figure 4 shows the cleaned collision from 2006-2010.

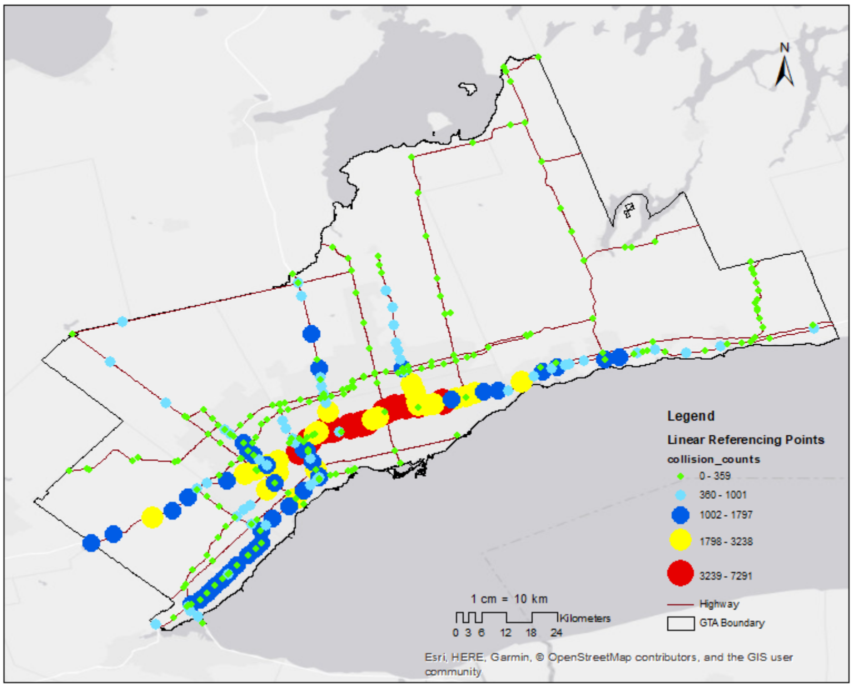

Figure 4: Spatial Distribution of All Highway Collisions

The next step is to select the useful attributes in the collision dataset. There are more than 50 attributes reflecting the detailed information about each collision, such as driver's body conditions, the impact of collisions, weather conditions, etc. Based on the results from literature part, some important attributes (Table 2) are selected for this project, including accident location, road alignment, environment, speeding, driver's age, driver's condition, as well as driver's action.

Table 2: Selected Attributes in Highway Collisions

Accident location	Road alignment	Environment	Speeding	Driver's Age	Driver's Condition	Driver's Action
1 = Non-intersection 2 = Intersection related 3 = In intersection	1 = Straight on level 2 = Straight on hill 3 = Curve on level 4 = Curve on hill	1 = Clear 2 = Rain 3 = Snow 4 = Frizzing Rain 5 = Drifting Snow 6 = Strong Wind 7 = Fog, Smoke, or dust	Yes/No	Detailed ages	1 = Normal 2 = Drinking 3 = Ability Impaired drugs 4 = Fatigue 5 = Inattentive	0 = Unknown 1 = Driving Properly 2 = Improper Turn 3 = Disobey Traffic Controls

The last step is to assign all collisions to the nearest highway segments. In the highway system, each collision is associated to the Liner Reference System (LHRS) points with an offset, so the accurate location of each collision is not stored by longitude and latitude rather than the linear location. Considering the large number of collisions in the study, all collisions were finally joined to the nearest highway segments to perform further analysis.

5.1.2 Collisions in Toronto

There were two main datasets for our analysis of collisions in Toronto. Centrelines from the City of Toronto Open Data Catalogue and Killed or Seriously Injured (KSI) from the Toronto Police Service's, Public Safety Data Portal. Since the highway dataset was for the years 2006-2010, and the KSI data was for the range of 2006-2016. Only the years of 2006-2010 were used from the KSI dataset to keep consistency.

The attributes of the centreline data also included; railway, river, creek, trail, walkway, Hydro Line, and shoreline data. All these attributes were removed (marked in red) from the centreline dataset, resulting in a road network, of municipal highways and local roads (Table 3). The automobile collision data is a subset of the Toronto Police Service's, KSI dataset. The automobile collision data had 50 attributes. The important factors of our study were listed under the following six attributes (Table 1); visibility, light, invage, injury, drivCond, and speeding. Within each attribute we aggregated some of the contents. For example, visibility had the drifting snow and snow, we decided to aggregate these two fields together, to create our "snow" attribute.

Lastly, a spatial join was needed to join the collision point data to the centreline polyline data, for the web GIS. This would give the user a better visual, although the collision data is marked at an intersection, it is not necessarily the exact location as to where the accident happened. This could improve or disprove accuracy to where the accident happened introducing the Modifiable Areal Unit Problem (MAUP).

Table 3: Attribute table for Toronto Centrelines

STREET FEATURE CODES	LINEAR FEATURE DEFINITION
Highway	Highway is designated for fast, long distance travel with restricted access to sustain high speeds.
Highway Ramp	Highway Transfer Ramp provides for transfer between road and highway and also between highway and highway.
Major Arterial Road	Arterial Road is usually under regional jurisdiction and is fed by collector roads and in some cases is connected to other arterial roads or collector roads via Road Ramp.
Major Arterial Road Ramp	Road Ramps (major arterial, minor arterial, collector, other) provides for transfer between two roads.
Minor Arterial Road	
Minor Arterial Road Ramp	
Collector Road	Collector Roads is designated mainly for travel to and from arterial roads with some driveway access. In Metro they are usually under local jurisdiction.
Collector Road Ramp	
Local Road	Local Road is designated to service driveway access and usually connects to collector roads or other local roads.
Other Road	
Other Ramp	
Laneways	Lane is designated mainly for City of Toronto laneways and are usually under local jurisdiction.
Pending	Pending Road is suggested to identify roads with a planned feature code that awaits Council Approval. This is not requested until the road is assumed and may be delayed for 6 years or more.
Busway	Busway is a road dedicated for buses only.

Access Road	Access Road is dedicated to provide access to or within properties such as townhouse complexes, airports etc.
Major Railway	Major Railway is designated for the fast, long distance, inter-provincial movement of cargo or passenger trains.
Minor Railway	Minor Railway is designated for local public transportation and includes above ground rapid transit corridors and subway lines.
Railway under construction/proposed	
River	River is a major waterway.
Creek/Tributary	Creek/Tributary is a minor waterway.
Trail	Trail is a pedestrian way designated for recreational purposes and can include foot-powered vehicles such as bikes or roller-blades etc.
Walkway	Walkway is a designated path primarily for walking.
Hydro Line	Hydro Line is an electricity transportation corridor (high voltage).
Major Shoreline	Major Shoreline is a boundary of a large body of water. E.g. Lake Ontario shoreline.
Minor Shoreline (Land locked)	Minor Shoreline is a boundary of a small body or water such as a pond or reservoir.

5.1.3 Historical Weather Data

As mentioned previously in the Data Sources section, the historical weather station data from Environment Canada is hourly data for each of the following weather stations:

- Billy Bishop Toronto City Airport,
- Lester B. Pearson International Airport,
- Buttonville Municipal Airport,
- John C. Munro Hamilton International Airport,
- Peterborough Airport, and
- Canadian Forces Base Borden.

The chosen weather stations are either within the study area or close to the study area. The data collection excluded weather stations that only had daily historical data to maintain consistency. The datasets originally included several fields, however only the date, time, and temperature in degrees Celsius were kept for further analysis. Any missing values within the temperature field were replaced with the average of the two non-empty cells above and below the missing values on Microsoft Excel.

To obtain temperature data for both highway and Toronto road collisions, the weather station point data was first converted using Thiessen polygons in ArcMap. This creates a new shapefile where the polygons represent the areas closest to the points of which the polygons are based (Create Thiessen Polygons, n.d.). Converting the weather stations from points into polygons allows the accident points to easily be spatially joined to their closest weather station. After the spatial join, the accident points are then joined to the weather data table using the date and time of the accident. Since the times of the accidents in both the highway and Toronto road dataset were exact times, the date and time values were first rounded to the closest hour using Excel.

5.2 Multi-Criteria Analysis

Multi-criteria analysis (MCA) was used to measure the collision vulnerability in highway and Toronto roads. According to Tzeng et al (2005), MCA is a widely used decision-making method which consider a complex problem by using subfactors or criteria. Each factor is ranked by its own influencing level and the weights between factors are then used to integrate all factors.

5.2.1 Selecting Factors

As mentioned, both collision datasets provided a large number attributes, and each attribute had multiple values. For example, the attribute values of the driver action included speeding, and improper turning, passing, or lane changing. The driver condition attribute reported on whether the driver’s ability is impaired by alcohol, drugs, fatigue, or medical defects, etc. This project considers each unique value of the attributes as a factor for car collisions.

Provided with several hundred factors in the datasets, the project needed to select the most important ones and weight them consistently. The project considered the number of accidents associated with a factor to determine the level of importance. As mentioned in Section 5.1, the count of collisions was aggregated by factor in a Python program. Within the time constraint, the project only considered the top nine factors by the number of accidents associated with them.

5.2.2 Classifying Road Segments

The road segments on the highways and in Toronto were classified into six classes for each factor. For each factor, road segments where no collision happened from 2006 to 2010 were assigned a vulnerability score of zero, and the other segments were classified by number of collision incidents into five quantiles. The five quantiles were then assigned vulnerability scores 20 (the least number of collisions happened on these segments), 40, 60, 80, and 100 (the most number of collisions happened on these segments). These values are the vulnerability scores to be displayed on the web application when only one factor is selected by the user.

The thresholds to divide the segments into 5 quantiles were determined in ArcMap using the Quantile classification method. The Natural Breaks classification method was not applied because it simply emphasized differences between classes, causing a larger number of segments to have high vulnerability scores. When the thresholds were determined, the attribute vulnerability score was updated using Field Calculator. This required manual modification of the thresholds for different factors in the Field Calculator Python script.

5.2.3 Calculating Vulnerability Scores for Combinations of Factors

When the user selects more than one factors on the web application, the weighted sum of each selected factors' vulnerability score will be displayed. The vulnerability scores are generated based on quantiles of collision incidents (Section 5.2.2), and the weight (w_i) for each factor *i* in a selection is determined by the equation below.

$$w_i = \frac{CC_i}{\sum_i CC_i}$$

where CC_i is the count of collisions associated with factor *i*.

The weight for the same factor changes when the factor is selected with different factors. The weighted sum of vulnerability scores for each possible selection of factors were computed by an R program (see Appendix A). An alternative approach is to compute upon the user's request query using a web server.

5.3 Web-GIS Development

5.3.1 Further Data Processing

After performing MCA, new attributes were added to the shapefile. These new attributes, represented how vulnerable that area was to that factor. For example, if there was a collision on "Main Street" over the past five years, how many of those collisions occurred on a snowy day. Once the most critical factors were determined, a combination of the top factors were created into attributes as a percentage. These attributes represented the weighted sum of vulnerability scores of the select combination of three factors. At least one factor must be selected for a display to appear on our web application. For example, a selection from an age-group, an environmental factor, and a road-related factor. The selection of these attributes was based on the Boolean operator “OR” factor, not “AND”. This means the selection was “driver age of 25 OR drove on a clear day”, and not “driver age of 25 AND drove on a clear day”. The reason for only using the top factors is to maximize the visualization of the number of collisions on our web map as well as to prevent the web application from experiencing lag.

Since the temperature data was not included in the MCA, it would also be shown separately in the web-GIS results. The accident and temperature data were separated into quantiles and then spatially joined to the road segments, with a count of accidents in each of the temperature quantiles. The count was then converted to a percentage with 100% being the highest count within one quantile. Figures 5 and 6 shows the temperature data quantiles for both the GTA highways and Toronto roads datasets.

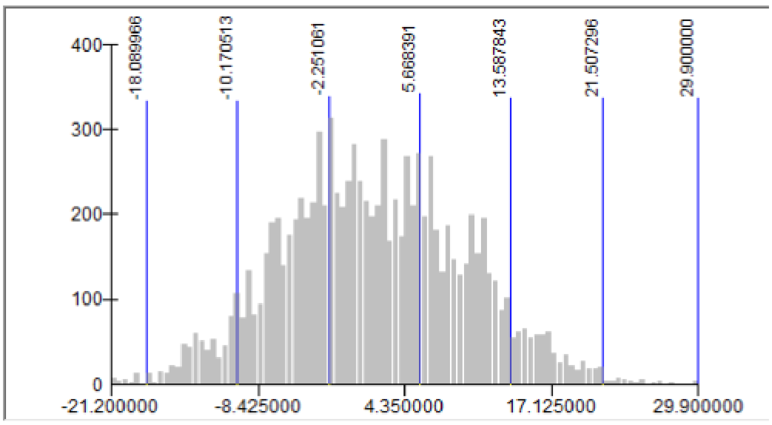

Figure 5: Temperature (°C) Distribution for Accidents in the GTA

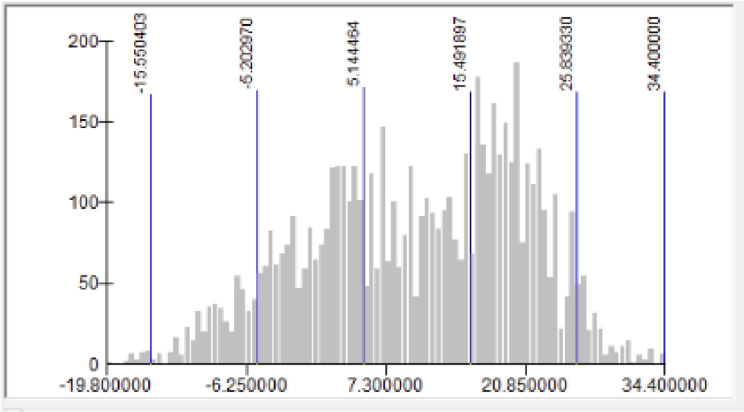

Figure 6: Temperature (°C) Distribution for Accidents in Toronto

After the weather data was processed, the datasets were joined using the road segment ID to the factor dataset so that one dataset contained all the calculated attributes. These attributes were then hidden from the user and only displayed visually or spatially on the map. Finally, we converted the shapefiles into GeoJSON files using QGIS.

5.3.2 Webpage Development

The development of the actual webpage containing the final map was done with multiple map versions in mind. For that reason, a template of how the final product would look like was made first, so that even if the map component of the page changes, updating the website to accommodate to the map changes would be a straightforward process. The web page was made with HTML, JavaScript, as well as CSS to customize formatting further. Figure 7 displays the initial version of the final webpage template. The map component of the web page is embedded in a fixed size container. Bootstrap, an HTML and CSS library, was used to format the buttons on the web map.

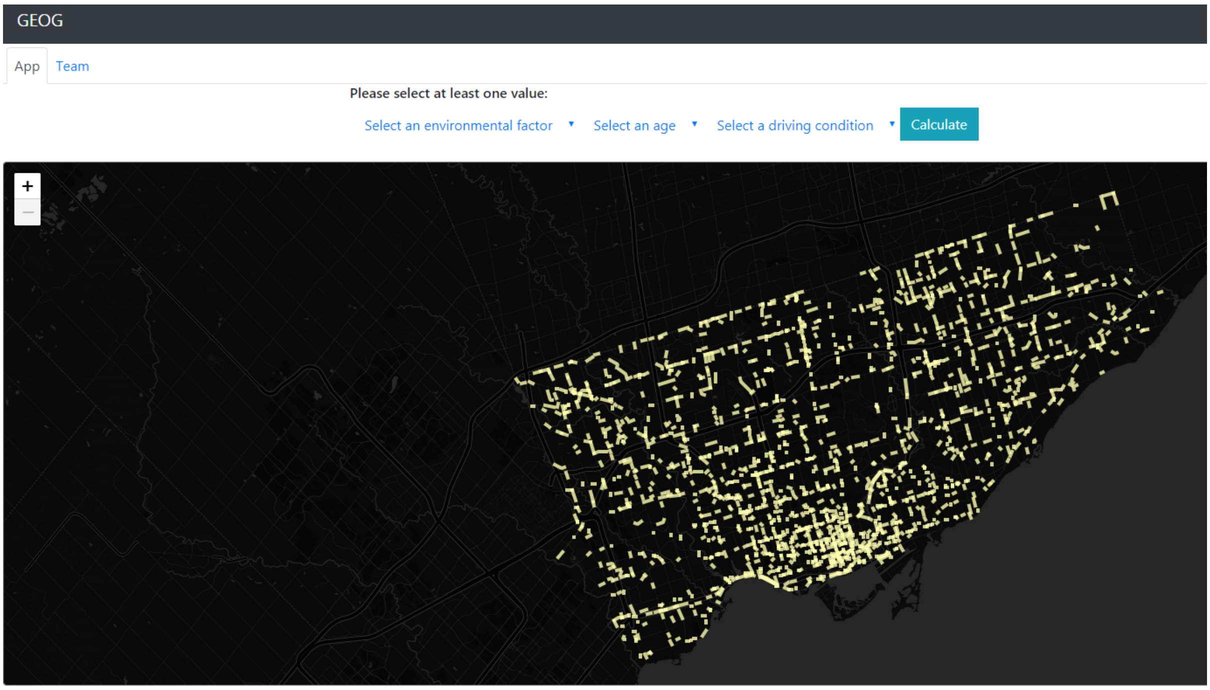

Figure 7: Webpage Template

The first step in creating the web map component was to create a simple map using one of the road segments, specifically the Toronto roads dataset, with QGIS. The symbology on the map was determined from creating ranges for the risk values and temperature quantiles using equal intervals for one of the attributes. All the risk factors and the temperature quantiles are values from 0 to 100, hence the symbology can remain the same even when a different attribute is shown. This ensures that the map display remains consistent when switching viewed layers. The QGIS plugin, qgis2web, was used to create a Leaflet web map based on the existing map.

From there, user selection-enabled buttons and dropdowns for the web map were created. These dropdown options were implemented so that users would be able to toggle which layers to view. This allows the map layout to not appear overcrowded with individual layer selection checkboxes that would traditionally be used on a web map. This user-based dropdown selection works by running a script using JavaScript to determine if the current layer should be changed to a new layer displaying a different field kept within the GeoJSON file. The script only runs if the user clicks on the submit button for the risk factors or if a new temperature quantile is clicked on. From there, the current layer is removed, the style is changed to show a new field, and a new layer is added. This action only refreshes the Leaflet map and not the entire webpage. Figure 8 shows the first version of the web map without formatting.

Figure 8: Web Map First Draft

Once the web map customizations have been properly configured and the map is capable of toggling both the risk factor and temperature values, the source code can be merged with the first map template. As previously mentioned, the web map is embedded within a container and the toggle dropdowns and submit button are placed above the web map.

Legends were also included for both the risk factors and temperature data in the final version of the web map (Figure 9). The left side of the page shows both the risk factor dropdown and the right side shows the temperature dropdown. Below the map container are both legends are placed on their respective sides. A duplicate of this map is made to show the GTA highway dataset, with some of the dropdown values changed according to the dataset. When clicking on any road segment, a pop-up gives the road name as well as a short description of the location. In the GTA highway dataset, this description is the highway section while in the Toronto roads dataset, there is a description of the intersection and neighbourhood of the road segment. The temperature dataset's pop-up shows the percentage of accidents within all the temperature quantiles as well.

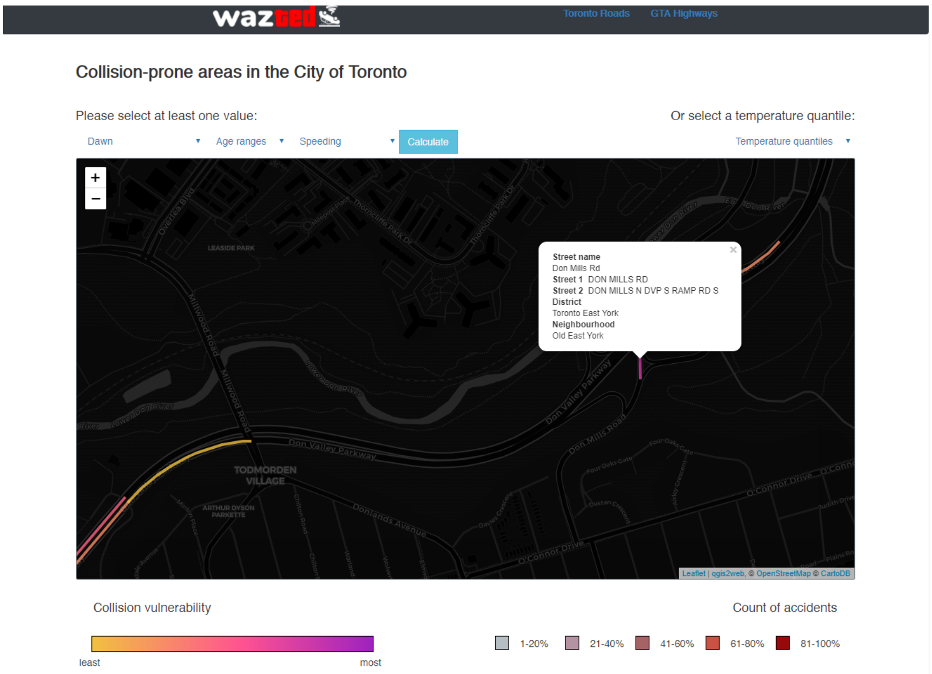

Figure 9: Screen capture of the final web application

Throughout the process of coding and creating a functional web-GIS, the page was inspected to see if there were any specific areas that needed to be changed or added. For further details pertaining to the web map or the web-GIS in general, the HTML source code with comments can be found in Appendix B.

Chapter 6 Results

6.1 Statistic Results

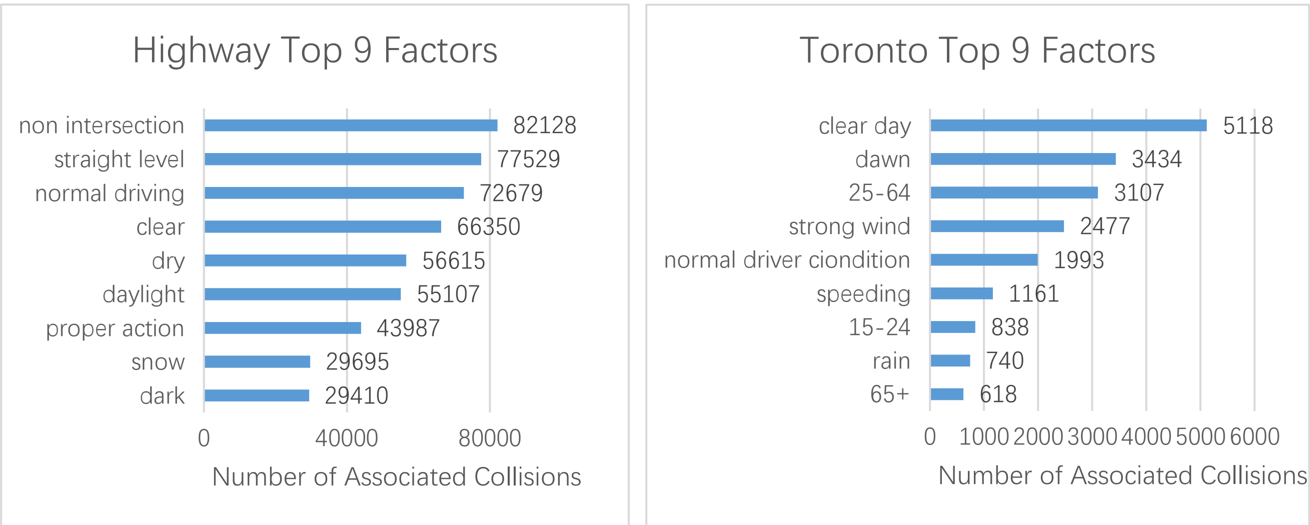

Figure 10: Selected Factors

Of all the factors recorded in the datasets, nine factors associated with the largest numbers of collisions were selected. Figure 10 shows factors selected in each collision dataset. Substantially more car collisions happened on the highways than in Toronto.

Highway	environmental	road geometry	driving
	clear	non intersection	normal driver condition
	daylight	straight level	proper driving action
	snow		
	dark		
	dry road		

Toronto	environmental	age	driving condition
	clear	25-64	normal driver condition
	dawn	15-24	speeding
	rain	65+	

Figure 11: Categories of Factors for Selection

Referring to Figure 11, the selected factors fell into three categories. Both datasets have environmental factors and driving conditions, but road geometry was selected for highways while drivers' age was considered for Toronto. On the web application, the user can select one factor from each category, and the weighted sum of is displayed as collision vulnerability. However, the factors in one category are not mutually excluded. For example, a highway collision can happen on a dry road when there is daylight on a clear day. This classification of factors was considered to limit the number of possible selections.

The user may select 55 scenarios (combinations of factors) for highway and 47 scenarios for Toronto. If there are k number of categories of factors, for category j containing n_j factors, the equation below calculates the number of scenarios. There are n_j+1 options for category j as the user can choose nothing under this category. However, the user must choose at least one factor each time, which is why one scenario is deducted where the user chooses nothing for all categories.

Number of scenarios:

$$\text{Highway: } C_6^1 * C_3^1 * C_3^1 - 1 = 55$$

$$\text{Toronto: } C_4^1 * C_4^1 * C_3^1 - 1 = 47$$

Without access to a server for real-time calculation, the project calculated the weighted sum for each scenario in advance. A Python program was written to calculate the weighted sum in each scenario. The program computes the weights for each selected factor, and the weighted sum of the vulnerability scores of the selected factors.

If the project is to consider over 50 factors, there is likely to be thousands of scenarios. The Python program can calculate the weighted sum for such a large number of scenarios but classifying the road segments needs to be automated or a web server should be used to compute on demand.

6.2 Web-GIS

A simple description of the webpage's features can be found in Section 5.3.2, Webpage Development. The following is a link to the web-GIS:

https://yswang96.github.io/wazted/

6.3 Vulnerability Results

This section presents the results of a few selections of interesting factors on the web application. It provides recommendation for further analysis as the web application is simply a platform to visualize the vulnerability levels.

In Toronto, clear weather is associated with the largest number of collisions. This correlation does not necessarily suggest that clear weather alone is the cause of most collisions. According to Environment Canada (2018), 305 days (86.56% of the year) was clear annually from 1981 to 2010. Car collisions on a clear day is likely to be caused by other factors.

Figure 12: Collision Vulnerability to Each Environmental Factor

Figure 12 shows the results of selecting one environmental factor each time on the web application. Clear weather and dawn were associated with the largest and second largest numbers of collisions in Toronto (Figure 10). Visually, while segments show high vulnerability scores under clear weather, their vulnerability to dawn is less intense. The other environmental factor, rain, shows a very dispersed pattern of high vulnerability segments.

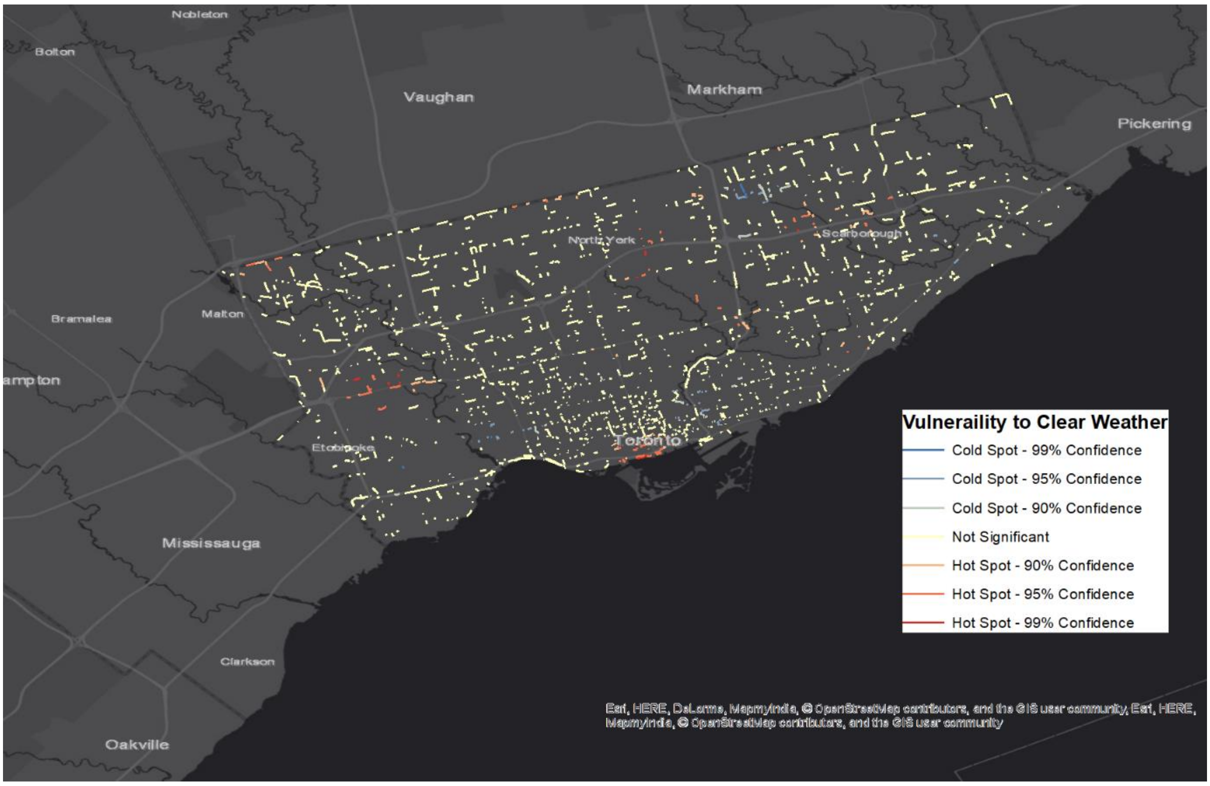

Figure 13: Hot Spot Analysis for Clear Weather Vulnerability

The pattern can be examined statistically. Moran's I statistic reports that the distribution segments that are highly vulnerable to clear weather are not significantly different from random (z-score 0.596). Referring to Figure 13, Getis-Ord Gi* identified significant clusters of segments with high (hot spots) or low (cold spots) vulnerability levels.

Chapter 7 Discussions

7.1 Data Limitations

Data limitations come from two different sources of the collision. The highway collisions are obtained from Ontario Provincial Police while collisions in Toronto are downloaded from Toronto Police Service. These two data sources have different attributes, indicating there are two different measurements of a collision. The number of collisions in the two datasets also have great difference: there is a total of 61,787 collisions recorded in the highway datasets while the Toronto dataset only has 6,100 collisions from 2006 to 2010. In addition, the highway dataset and Toronto dataset also include some attributes that other one does not have. For instance, the highway dataset has the information describing the road conditions while Toronto dataset recorded the detailed injury levels. Thus, it is not feasible to merge these two different datasets together. Future improvement should try to use the dataset with the same standard of measurement.

Additionally, the MAUP is introduced in the project. MAUP is a problem in geography and spatial analysis, which refers to an artificial spatial pattern caused by subjective classification of research units on continuous geographic phenomena (Fotheringham, 1991). When pre-processing the collision datasets, all records are aggregated to the nearest road segments. Specifically, the Toronto collision dataset is aggregated to the road central lines while the highway collisions are assigned to the highway linear referencing segments. Although both two sources of road segments are separated already, the MAUP still exists.

The road segments may be separated in the digitizing process while highway segments are separated by the official Highway Linear Referencing points. Since the MCA requires a basic research unit to conduct the analysis, future study can attempt other methods such as analyzing point patterns and spatial weights.

The historical weather data is also affected by MAUP. The method to create Thiessen polygons is simple and time efficient, however it does not account for vertical distance, only horizontal distance. There could be a difference in temperature based on elevation, which can affect the overall accuracy of weather and accident data spatial join. In addition, as the analysis focuses on hourly data, a few weather stations within the study area had to be excluded due to differing time frames. Some weather stations that had hourly temperature data were also excluded since they did not fit the full-time frame of the accident datasets.

7.2 Limitations in Multi-Criteria Analysis and Future Improvements

The limitations exist in the steps of the MCA. First, the criteria were selected based on the number of collisions. While this selection can reflect the collisions associated with the largest counts, some severe collisions such as fatal crashes may be neglected due to their small number. Second, limitations exist in the process of classifying road segments. While a quantile-based method can separate road, segments based on the nature distribution of data, the assigned vulnerability values (0, 20, 40, 60, 80, and 100) are discrete. One issue is that the scores may exaggerate the difference between two segments with similar collision counts. For example, if one road segment has the lowest number of collisions in the second quantile while the other one is the largest one in the first quantile, the former will be assigned 40 as its vulnerability score but the latter will be assigned 20. This difference may influence the interpretation of the results when more factors are considered together. Lastly, the method to determine weight is not accurate only using the number of collisions. Considering the large number of possible scenarios in Web-GIS and limited time for the project, this simplified process of deciding weights can help save more time. However, this method may be not accurate for a complex problem. For example, some collisions with severe impacts tend to have a relative small count, but its weight is assigned a small number.

Future studies should focus on the improvement of MCA. In the criteria selection part, an influencing variable such as the injury level can be used to measure the importance of criteria. The vulnerability score can be assigned using a linear or fuzzy classification to reduce the difference between to similar road segments. The weights can also be decided more accurately. An alternative to determine weights is to use the Analytic Hierarchy Process, but this method is more subjective. Initially, the weights were expected to obtain from the regression model. However, there are two problems: the coefficients of model and the statistical p-value cannot be considered as the weights because these numbers only reflect the fitness of the selected factor and cannot be used to compared with any other factors; second, it is difficult to perform a regression model using the attributes merely using the collision data. For example, most attributes in the collision datasets are recorded as text such as “clear”, “dawn”, etc. When a regression model is performed, this text is first converted to the pre-designed number (1 indicating “clear”, 2 indicating “dawn”). The overall accuracy is low, no more than 0.1. Collecting external data also requires much time for the large number of collisions in this project. Future studies should consider a more suitable model to measure the influencing level of many factors.

To enrich the datasets, the latest Toronto collision dataset in 2017 are available but the highway collision is only from 1999 to 2010. Other datasets such as census data were also considered in this project but the population density along the highway does not have a logical relationship with the number of collisions. A better improvement may concentrate on the severe factors first, and comprehensively analyze the point patterns in terms of spatial-temporal distributions and nearby building types.

7.3 Web-GIS Limitations and Future Improvements

Currently, the web application only displays the weighted sum of the vulnerability scores for the selected factors. The user can visually examine how the pattern changes across different combinations of factors. An improvement of the web application is to inform the user whether the spatial pattern is random, and to identify significant clusters of high and low vulnerability scores.

A limitation that had been mentioned earlier is that the two road datasets cannot be shown in the same map. Since the datasets contain different vulnerability factor fields, the user would not be able to compare the two sets of results with one another. Due to this data limitation, it would be difficult to draw conclusions using both datasets.

A future improvement, given access to a larger dataset, would be to include a temporal factor to the final map display. The user would be able to filter and look at the results based on date and time, so that seasonal or daily patterns could be easier to notice. With the current web map, this feature was not included as this project only investigated only five years and some of the Toronto road segments do not have enough data to show meaningful temporal data.

Chapter 8 Conclusions

Conclusions can be drawn from the statistical results. Section 6.1, Figure 9 shows the factors associated with the most collisions. Within these results, it can be noted that for GTA highways, the top three factors were non-intersection, straight level, and normal driving. Non-intersection related accidents are accidents that occurred away from ramps and straight level roads are any road sections that do not have a curve or slope to them. The top factor in the Toronto dataset is clear weather, followed by dawn and the 25-64 age range. All three of the top highway factors as well as Toronto's clear factor could be explained with careless driving, as there are no hazardous road or weather conditions. Within the Toronto dataset, accidents during dawn can also be related to the fatigue with an early commute for smaller vehicles or a long shift for truck drivers. Driving while drowsy or sleepy is equally as dangerous as driving with Blood Alcohol Concentrations (BAC) at levels deemed dangerous by legislations (MacLean, Davies, & Thiele, 2003). The third factor in the Toronto roads dataset is an age range, also the largest age range in the dataset.

In terms of spatial results, the web map showed that downtown Toronto, particularly around Gardiner Expressway and Queen’s Quay, had the most collisions for any of the factor combinations, which agrees with Figure 13’s hotspot analysis. These areas are busier and more vehicle-dense. The highway dataset highlighted vulnerable sections of highway within the City of Toronto and in Peel and York Regions as well. Durham Region has less overall vulnerability compared to other areas, which can be explained by the fewer number of highways. Overall, the results tend to focus more on areas with more interchanges or expected traffic volume.

From this study, it can be concluded that most accidents do occur from inattentive or reckless driving rather than poor road conditions. In addition, densely populated areas were also shown to be more vulnerable. One potential solution to this is to raise further awareness to general driver safety. For example, there is signage on highways that can be customized to show a safety message. One suggestion is to change the text to reflect a more general warning regarding reckless or inattentive driving rather than impaired driving, especially since impaired driving rates have gone down over the years (Perreault, 2016). Even though it may seem obvious, raising awareness of driving safety and defensive driving in dense areas is the most important step in reducing the overall number of accidents, as most accidents happen in busy areas, and drivers are also more likely to become involved in an accident in such areas, even if they are not the ones at fault.

References

Akerstedt, T., & Kecklund, G. (2001). Age, gender and early morning highway accidents. *J. Sleep Res.*, 105-110.

Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis and Prevention*, 359–364.

Chang, L.-Y. (2017). Exploring contributory factors to highway accidents: A nonparametric multivariate adaptive regression spline approach. *Journal of Transportation Safety & Security*, 419–438.

Charlton, S. G. (2009). Driving while conversing: Cell phones that distract and passengers who react. *Accident Analysis and Prevention*, 160–173.

Create Thiessen Polygons. (n.d.). Retrieved from ArcGIS Pro: http://pro.arcgis.com/en/pro-app/tool-reference/analysis/create-thiessen-polygons.htm

Faghri, A., & Raman, N. (1994). A GIS-Based Traffic Accident Information System. *Journal of Advanced Transportation*, 321-334.

Fotheringham, A. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and planning A*, 23(7), 1025-1044.

Gundogdu, I. B. (2010). Applying linear analysis methods to GIS-supported procedures for preventing traffic accidents: Case study of Konya. *Safety Science*, 763–769.

Hegyi, P., & Attila Borsos, C. K. (2017). Serching possible accident black spot locations with accident analysis and GIS software based on GPS coordinates. *An International Journal for Engineering and Information Sciences*, 129-140.

Huang, D., Kapur, A. K., Ling, P., Purssell, R., Henneberry, R. J., & Champagne, C. R. (2010). CAEP position statement on cellphone use while driving. *Canadian Journal of Emergency Medicine*, 365-376.

Hunton, J., & Rose, J. M. (2005). Cellular Telephones and Driving Performance: The Effects of Attentional Demands on Motor Vehicle Crash Risk. *Risk Analysis*, 855-866.

Lai, P. C., & Chan, W. Y. (2004). GIS for Road Accident Analysis in Hong Kong. *Geographic lnfortiiution Sciences*, 58-67.

Levy, D. T. (1988). The Effects of Driving Age, Driver Education, and Curfew Laws on Traffic Fatalities of 15-17 Year Olds. *Risk Analysis*, 569-574.

Li, G., Baker, S. P., Langlois, J. A., & Kelen, G. D. (1998). Are Female Drivers Safer? An

Application of the Decomposition Method. *Epidemiology*, 379-384.

Perreault, S. (2016). Impaired Driving in Canada, 2015. *Juristat: Canadian Centre for Justice Statistics*, 1.

Rothman, L., Macpherson, A., Buliung, R., Macarthur, C., To, T., Larsen, K., & Howard, A. (2015). Installation of speed humps and pedestrian-motor vehicle collisions in Toronto, Canada: a quasi-experimental study. *BMC Public Health*, 1-7.

Rzeznikiewiz, D., Tamim, H., & Macpherson, A. K. (2012). Risk of death in crashes on Ontario's highways. *BMC Public Health*, 12:1125.

Satriaa, R., & Castro, M. (2016). GIS TOOLS FOR ANALYZING ACCIDENTS AND ROAD DESIGN: A REVIEW. *Transportation Research Procedia 18*, 242 – 247.

Soltani, A., & Askari, S. (2017). Exploring spatial autocorrelation of traffic crashes based on severity. *Injury*, 637–647.

Szumilas, M. (2010). Explaining Odds Ratios. *Information Management for the Busy Practitioner*, 227-229.

Turner, C., & McClure, R. (2003). Age and gender differences in risk-taking behaviour as an explanation for high incidence of motor vehicle crashes as a driver in young males. *Injury Control and Safety Promotion*, 123–130.

Tzeng, G. H. (2005). Multi-criteria analysis of alternative-fuel buses for public transportation. *Energy Policy, 33(11)*, 1373-1383.

Vanlaara, W., Robertsona, R., Marcouxa, K., Mayhewa, D., Browna, S., & Boase, P. (2012). Trends in alcohol-impaired driving in Canada. *Accident Analysis and Prevention*, 297–302.

Vemulapalli, S. S., Ulak, M. B., Ozguven, E. E., Sando, T., Horner, M. W., Abdelrazig, Y., & Moses, R. (2017). GIS-based Spatial and Temporal Analysis of Aging-Involved Accidents: a Case Study of Three Counties in Florida. *Appl. Spatial Analysis*, 537–563.

Wentworth, G., Murphy, J., & Sills, D. (2015). Impact of lake breezes on ozone and nitrogen oxides in the Greater Toronto Area. *Atmospheric Environment*, 52-60.

Williams, A. F., & Karpf, R. S. (1983). Deaths of teenagers as passengers in motor vehicles. *Accident Anaylsis*, 49-54.

Zhang, J., Fraser, S., Lindsay, J., Clarke, K., & Mao, Y. (1998). Age-specific patterns of factors related to fatal motor vehicle traffic crashes: focus on young and elderly drivers. *Public Health*, 289-295.

Zuo, F., Li, Y., Johnson, S., Johnson, J., Varughese, S., Copes, R., . . . Chen, H. (2014). Temporal and spatial variability of traffic-related noise in the City of Toronto, Canada. *Science of the Total Environment*, 1100–1107.

Appendix A Statistical Analysis

```
# trt
# load collision data
trt_weights <- read_csv("trt
weights.csv",
                    col_types =
cols(Date = col_character()))
trt_scores <- read_csv("trt
scores.csv",
                   col_types =
cols(Date = col_character()))
# create columns
for (i in 0:3){
  for (j in 0:3){
    for(k in 0:2){
      # set the first tab
      if(i == 1){
        tmp1 <- trt_scores$c4_score
        tmp1_weights <-
trt_weights$c4
      } else if (i == 2){
        tmp1 <- trt_scores$c7_score
        tmp1_weights <-
trt_weights$c7
      } else if (i == 3){
        tmp1 <- trt_scores$c2_score
        tmp1_weights <-
trt_weights$c2
      } else {
        tmp1 <- trt_scores$FID
        tmp1_weights <- 0
      }
      # set the second tab
      if(j == 1){
        tmp2 <- trt_scores$c15_score
        tmp2_weights <-
trt_weights$c15
      } else if (j == 2){
        tmp2 <- trt_scores$c14_score
        tmp2_weights <-
trt_weights$c14
      } else if (j == 3){
        tmp2 <- trt_scores$c16_score
        tmp2_weights <-
trt_weights$c16
      } else {
        tmp2 <- trt_scores$FID
        tmp2_weights <- 0
      }
      # set the third
      if(k == 1){
        tmp3 <- trt_scores$c20_score
        tmp3_weights <-
trt_weights$c20
      } else if (i == 2){
        tmp3 <- trt_scores$c21_score
        tmp3_weights <-
trt_weights$c21
      } else {
        tmp3 <- trt_scores$FID
        tmp3_weights <- 0
      }
      # weighted sum
      tmp <- tmp3*tmp3_weights +
tmp2*tmp2_weights +
tmp1*tmp1_weights
      if (i == 0 && j ==0 && k== 0)
{
        weightedSum = 1
      } else{
        weightedSum = tmp3_weights +
tmp2_weights + tmp1_weights
      }
      tmp <- tmp /weightedSum
      tmp <- as.integer(tmp)
      # add columns
      tmp <- as.data.frame(tmp)
      colna <- sprintf("T%d%d%d", i,
j, k)
      names(tmp)[1] <- colna
```

```
     trt_scores <-
cbind(trt_scores, tmp)
     cat(sprintf("The T%d%d%d
column is done!\n", i, j, k))
    }
  }
}
# output new files
trt_senarios <-
as.data.frame(trt_scores$LONUML)
names(trt_senarios)[1] <- "roadID"
trt_senarios <- cbind(trt_senarios,
trt_scores[,127:174])
write.csv(trt_senarios,
"trt_senarios.csv")

# hwy
# hwy set missing value to 0 first
# load collision data
hwy_weights <- read_csv("hwy
weights.csv",
                    col_types =
cols(Date = col_character()))
hwy_scores <- read_csv("hwy
scores.csv",
                   col_types =
cols(Date = col_character()))
hwy_scores[is.na(hwy_scores)] <- 0
# create columns
for (i in 0:5){
  for (j in 0:2){
    for(k in 0:2){
     # set the first tab
     if(i == 1){
       tmp1 <- hwy_scores$c8_score
       tmp1_weights <-
hwy_weights$c8
     } else if (i == 2){
       tmp1 <- hwy_scores$c21_score
       tmp1_weights <-
hwy_weights$C21
     } else if (i == 3){
       tmp1 <- hwy_scores$c10_score
       tmp1_weights <-
hwy_weights$c10
     } else if (i == 4){
       tmp1 <- hwy_scores$c22_score
       tmp1_weights <-
hwy_weights$C22
     } else if (i == 5){
       tmp1 <- hwy_scores$c23_score
       tmp1_weights <-
hwy_weights$C23
     } else {
       tmp1 <- hwy_scores$OBJECTID_1
       tmp1_weights <- 0
     }
     # set the second tab
     if(i == 1){
       tmp2 <- hwy_scores$c1_score
       tmp2_weights <-
hwy_weights$c1
     } else if (i == 2){
       tmp2 <- hwy_scores$c4_score
       tmp2_weights <-
hwy_weights$c4
     }
     else {
       tmp2 <- hwy_scores$OBJECTID_1
       tmp2_weights <- 0
     }
     # set the third
     if(k == 1){
       tmp3 <- hwy_scores$c15_score
       tmp3_weights <-
hwy_weights$C15
     } else if (i == 2){
       tmp3 <- hwy_scores$c19_score
       tmp3_weights <-
hwy_weights$C19
     } else {
       tmp3 <- hwy_scores$OBJECTID
       tmp3_weights <- 0
     }
```

```
      # weighted sum
      tmp <- tmp3*tmp3_weights +
tmp2*tmp2_weights +
tmp1*tmp1_weights
      weightedSum <- tmp3_weights +
tmp2_weights + tmp1_weights
      if (weightedSum == 0) {
        weightedSum <- 1
      }
      tmp <- tmp /weightedSum
      tmp <- as.integer(tmp)
      # add columns
      tmp <- as.data.frame(tmp)
      colna <- sprintf("T%d%d%d", i,
j, k)
      names(tmp)[1] <- colna
      hwy_scores <-
cbind(hwy_scores, tmp)
      cat(sprintf("The T%d%d%d
column is done!\n", i, j, k))
    }
  }
}
# output new files
hwy_senarios <-
as.data.frame(hwy_scores$REFERENCE_
)
names(hwy_senarios)[1] <-
"reference"
hwy_senarios <- cbind(hwy_senarios,
hwy_scores[,64:117])
write.csv(hwy_senarios,
"hwy_senarios.csv")
```

Appendix B Web-GIS Development

A link to the source code can be found here:

https://github.com/yswang96/wazted