STA130 Final Project
Determining the Most Dangerous Province to Drive In

Alec Sampaleanu and Gaoyi Zhou, TUT0101E4

Introduction

In 2015, Canadian roads were the site of over 116 000 collisions that lead to personal injuries. Almost 2000 people died in traffic accidents. The question that inevitably arises is: "which province is the most dangerous to drive in?"

Using the Geotab hazardous driving data set, we set out to find the most hazardous province in Canada. Geotab has telemetry sensors in hundreds of thousands of cars accross Canada, which allows them to track these following variables for individual $150 m^2$ areas, as well as others:

- ▶ Number of traffic incidents; separated into incidents involving cars, multi-person vehicles, and small, medium, and large trucks
- ▶ Province
- ▶ Severity Score, a score for the area that reflects the traffic flow, incidents, and general hazard of the area

Objectives

Our main objective is to define hazardous driving and determine a way to measure and rank it for all provinces, and so therefore to answer three questions:

1. Define what your group considers to be "hazardous driving" for the purposes of your project. Note that we expect that each group may have a different definition. The analysis your group does will follow from how you choose to define hazardous driving.
2. According to your definition of hazardous driving, which province has the most driving hazards?
3. How does hazardous driving in the province you identified in 2. compare to each of the other provinces?

Our secondary objective was to see which variables were most useful in creating a ranking system for the provinces.

The final objective was to attempt to utilize as much of this semester's curriculum without overanalyzing the data.

Data Summary

The first step was to import and clean the data. We removed several columns that did not provide relevent data or were empty, such as rationale, and filtered the dataset to only include observations within Canada.

```
## Observations: 10,241
## Variables: 15
## $ AvgLatitude     <dbl> 53.57867, 51.03260, 51.00657, 51
## $ AvgLongitude    <dbl> -113.5907, -113.9585, -113.9994,
## $ MinLatitude     <dbl> 53.57758, 51.03149, 51.00540, 51
## $ MinLongitude    <dbl> -113.5918, -113.9598, -113.9996,
## $ MaxLatitude     <dbl> 53.57895, 51.03287, 51.00677, 51
## $ MaxLongitude    <dbl> -113.5904, -113.9584, -113.9983,
## $ province        <fctr> Alberta, Alberta, Alberta, Albe
## $ SeverityScore   <dbl> 0.1155, 0.4326, 0.1731, 0.0428,
## $ NumberIncidents <int> 61, 28, 33, 41, 28, 106, 39, 51
## $ CarIncidents    <int> 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0,
## $ MpvIncidents    <int> 1, 0, 0, 4, 0, 2, 0, 1, 11, 0, 0
```

Exploratory Analysis
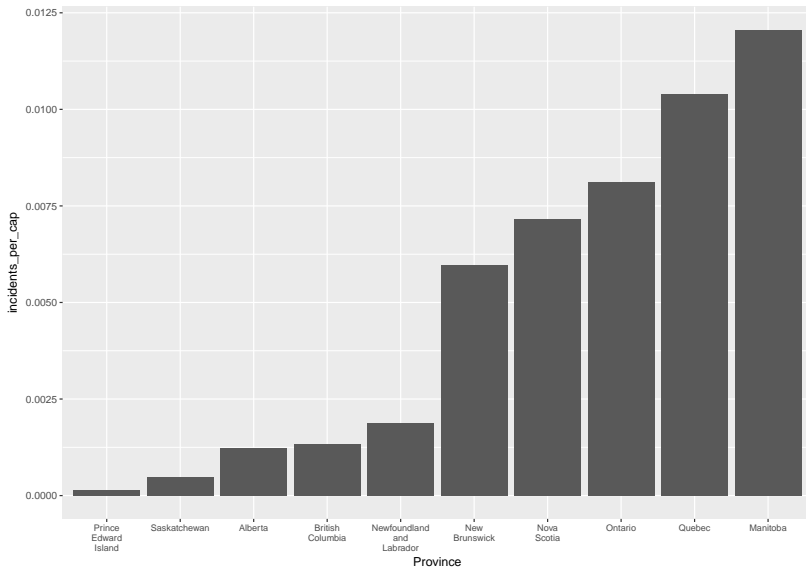
We decided not to use incidents in our ranking because they are already taken into account by the severity score.

Severity Scores
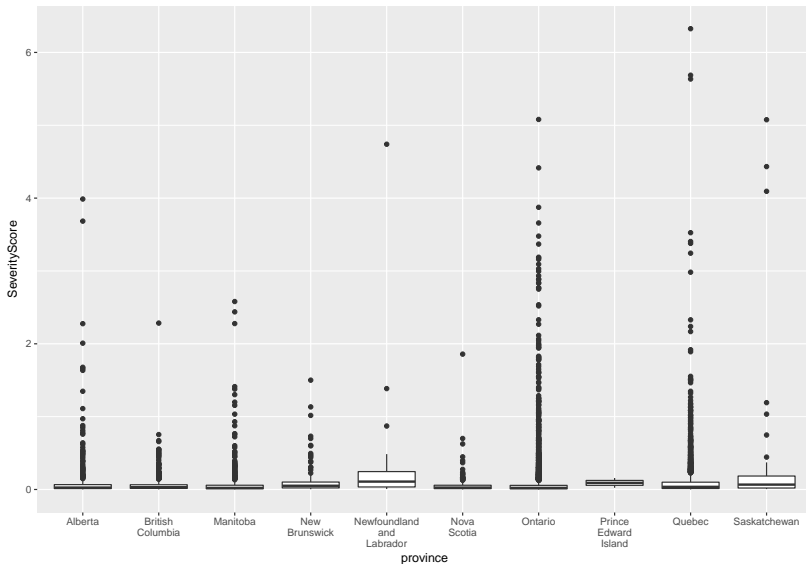
When ranking provinces on this score, we chose to use the mean instead of median, since we wished to account for outliers.

Mean Severity Score

```
## # A tibble: 10 x 2
##    province                  mean_sev
##    <fctr>                       <dbl>
## 1  Saskatchewan                 0.439
## 2  Newfoundland and Labrador    0.272
## 3  New Brunswick                0.130
## 4  Quebec                       0.115
## 5  Alberta                      0.109
## 6  Prince Edward Island         0.0906
## 7  Manitoba                     0.0863
## 8  Ontario                      0.0808
## 9  British Columbia             0.0660
## 10 Nova Scotia                  0.0603
```

Saskatchewan has by far the highest mean severity score, which will
certainly have a negative effect on its ranking.

Areas per Capita

```
## # A tibble: 10 x 2
##    province               areas_per_cap
##    <chr>                          <dbl>
##  1 Manitoba                    0.000470
##  2 Ontario                     0.000439
##  3 Nova Scotia                 0.000263
##  4 Quebec                      0.000261
##  5 New Brunswick               0.000161
##  6 British Columbia            0.000143
##  7 Alberta                     0.000117
##  8 Newfoundland and Labrador   0.0000943
##  9 Saskatchewan                0.0000419
## 10 Prince Edward Island        0.0000140
```

Here we ranked the provinces by hazardous areas per capita.
Manitoba and Ontario are the clear leaders in this category.

Areas per 1000 km of Road

```
## # A tibble: 10 x 2
##    province                 areas_per_tkm_road
##    <chr>                                 <dbl>
##  1 Ontario                                30.9
##  2 Quebec                                 14.7
##  3 British Columbia                        9.37
##  4 Nova Scotia                             8.97
##  5 Manitoba                                6.94
##  6 New Brunswick                           3.81
##  7 Newfoundland and Labrador               2.54
##  8 Alberta                                 2.11
##  9 Prince Edward Island                    0.333
## 10 Saskatchewan                            0.202
```

This is an alternative way to measure the percentage of areas in a
province. By using areas per 100 km of road, we have a better
estimate for low-populated provinces with many roads.

Fatalities per Billion Vehicle-Kilometres

```
## # A tibble: 10 x 2
##    province                 fat_per_bvk
##    <chr>                          <dbl>
##  1 Prince Edward Island            12.3
##  2 Saskatchewan                     8.70
##  3 Newfoundland and Labrador        8.20
##  4 British Columbia                 7.70
##  5 New Brunswick                    6.00
##  6 Alberta                          5.50
##  7 Manitoba                         5.50
##  8 Quebec                           4.90
##  9 Nova Scotia                      4.80
## 10 Ontario                          3.70
```

This is a very important statistic to measure since it is the only one
that directly tracks the human impact of hazardous driving. PEI is
the most dangerous, followed by Saskatchewan.

Statistical Methods

The key to our ranking methodology was that we wanted to ensure that no variable was more important than the other. We felt that by placing equal importance on each variable, we would ensure that our ranking was as fair as possible.

As such, the key statistical method we used is feature scaling. Feature scaling is used to bring all values of a set of numbers into the range [0, 1). The formula for feature scaling is:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

Where X is the variable to be scaled, and X' is the scaled variable.

Statistical Methods Continued

Since we wanted to include both the number of areas per capita and the number of areas per thousand km of road, we decided to weigh each variable by 0.5. As such, the formula for each provinces score is:

$$S = \frac{\text{NMS} + \text{NF} + 0.5 * \text{NAC} + 0.5 * \text{NAR}}{3}$$

Where *NMS* is normalized mean severity score, *NF* is normallized fatalities per billion vehicle-kilometres, *NAC* is normalized areas per capita, and *NAR* is normalized areas per 1000 km of road.
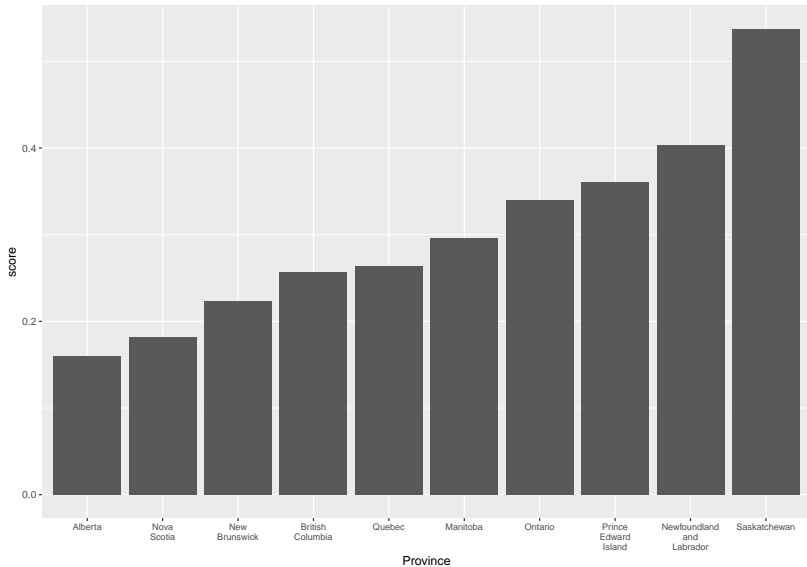
Results

```
## # A tibble: 10 x 3
##    province               score rank
##    <chr>                  <dbl> <fctr>
##  1 Saskatchewan           0.537 1
##  2 Newfoundland and Labrador 0.403 2
##  3 Prince Edward Island   0.361 3
##  4 Ontario                0.340 4
##  5 Manitoba               0.296 5
##  6 Quebec                 0.264 6
##  7 British Columbia       0.257 7
##  8 New Brunswick          0.224 8
##  9 Nova Scotia            0.181 9
## 10 Alberta                0.160 10
```

Saskatchewan is the most dangerous province, while Alberta is the
least dangerous.

Results Continued

Conclusion

In conclusion, we have answered the three questions in the following ways:

1. Provincial driving hazard is measured using three components: the mean severity of the hazardous areas within the province, the number of driving fatalities per billion vehicle-kilometres in the province, and the average of the number of areas per capita in the province and areas per 1000 km of road in the province.
2. The most dangerous province is Saskatchewan.
3. Th least dangerous province is Alberta, and Nova Scotia, the province highlighted in the assignment, is second safest. Ontario is 4th, and very close behind 3rd place PEI.

Acknowledgements

References:

Canada. Statistics Canada. (2017). *Population and Dwelling Count Highlight Tables, 2016 Census: Population and dwelling counts, for Canada, provinces and territories*.

Canada. Transport Canada. (2010). *Road Transportation: Length of Public Road Network in Canada*.

Canada. Transport Canada. (2016). *Canadian Motor Vehicle Traffic Collision Statistics: 2015*.