Seattle, WA Collision Trend Analysis

Ali Akbar Ahmadi Ali Soleimani Borujerdi Clayton Malone Ramin Shahbad
Alec Wick

December 14, 2022

Abstract
Car ownership in the United States is rapidly increasing. According to ValuePenguin's article titled "Car Ownership Statistics in the U.S.", "91.55% of households reported have access to at least one vehicle"[1]. This metrics is an 0.73% increase from 2015. Fatalities from car accidents have increased 9.4% during the same time period. Our study aims to analyze vehicle collision trends in Seattle, Washington from 2004 to 2021. The data set being used was published by the City of Seattle's Department of Transportation (SDOT). There are several important fields contained within the data set including accident location, vehicle type, time of day, weather, and light condition. The objective of this study is to determine significant factors that cause collisions and to discover time series trends.

Introduction

"Millions of deaths and injuries occur annually as a result of collisions. Around 1.25 million people die in car accidents each year around the world, and for people aged between 1 and 29, car accidents are the greatest cause of death worldwide" [1] Road accidents are currently the ninth leading cause of death; however, if road safety is not improved, they will undoubtedly rise. Developed countries have successfully decreased collision rates by applying restrictive rules such as seat belts, speed limits and laws prohibiting the consumption of alcohol while driving. For instance, fatalities due to collisions have decreased by almost 25% in the United States from 2005 to 2014 [2]. Collaborating with several sectors can prevent or reduce crash-related fatalities and injuries [3]. For example, data on collisions can aid in developing conceptual frameworks for traffic safety [4]. Even though responsible agencies have been trying to recognize the factors in crashes, the annual number of collisions are still significantly high. There were around 35,000 road deaths in the US in 2015, up 7.2% from the previous year [5]. Hence it is necessary to obtain a greater understanding about why collisions occur. In our research, we were able to determine the leading factors and trends.

About the Data and Methodology

The data used in this report consists of all reported vehicular collisions in the City of Seattle from 2004 to 2021 [6]. The data set is provided by the Seattle Department of Transportation (SDOT) and includes the location, severity, weather and lighting condition at the time of the accident. Other features in the data set include the type of intersection, number of passengers, and whether the driver was speeding. The provided data allows for a thorough analysis of what conditions contribute to sever vehicle collision.

1. Preparing data

The data set used in this study had no major alterations done before the data analysis phase. The data source was downloaded from SDOT's website in CSV (comma-separated-value) file format. To allow for reproducibility, the CSV file used was added to a GitHub repository so it could be easily accessed without local storage.

2. Methods

The given data set was analyzed to evaluate the relationship between several parameters such as location, weather, type and number of collisions overtime. To see what trends are present, several different figures were generated. The obtained result displays to the reader what factors lead to collisions and what areas

of Seattle have the most vehicular collisions. The time series analysis displays the progress of the SDOT in limiting the amount and severity of collisions despite the cities growing population.

Results

As mentioned earlier, the study's primary purpose is to find insights within the data that show us what factors may contribute to the number or severity of vehicle collisions and how they have changed overtime. Therefore, in the first step of the study, we evaluated severity and several types of collisions.

1. Severity and Types of Collision

To understand the severity of collisions in Seattle WA, we used the column *SEVERITYCODE*. This data column helps classify the crash's severity and determine the total number of collisions in each category. This coding standard has five different codes, including 0, 1, 2, 2b and 3, which stand for the following respective groups: unknown, property damage, injury, severe injury and fatality. Figure 1 suggests that around 69% of collisions lead to property damage, and the remaining involves injuries and fatalities.

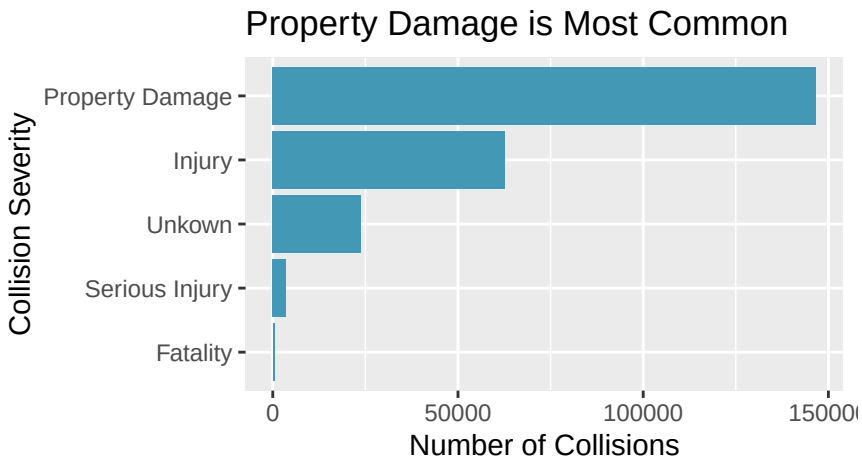


Figure 1: Collision frequency by collision Severity

To have a better view of the number of accidents which have led to fatalities, we considered all fatal accidents as one group and the other remaining categories, including property damage and both injury types, as another group. Figure 2 shows that the total number of fatal and non-fatal accidents. As depicted in this plot, there have only been 407 fatal accidents since 2004 in Seattle.

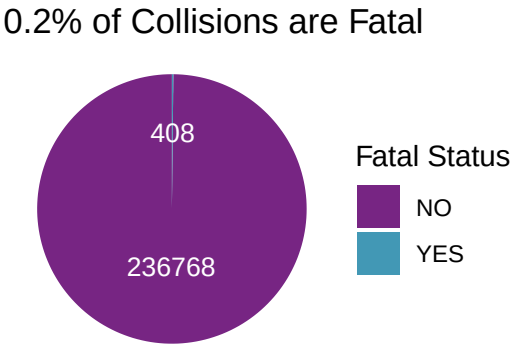


Figure 2: Fatal and Non-Fatal collision frequency

To examine the different types of collisions in Seattle WA, we used the column *COLLISIONTYPE*. This column helped determine the total number of collisions by each type. The horizontal bar chart shown below (Figure 3) displays the number of collisions for each type. As you can see, most of the accidents happened were Parked Car collisions. After Park Car collision, angles contributes to the second highest number of collisions. The lowest occurance of collisions is for head-on collisions. Collisions including cycles and pedestrian also had a relatively high number of collisions.

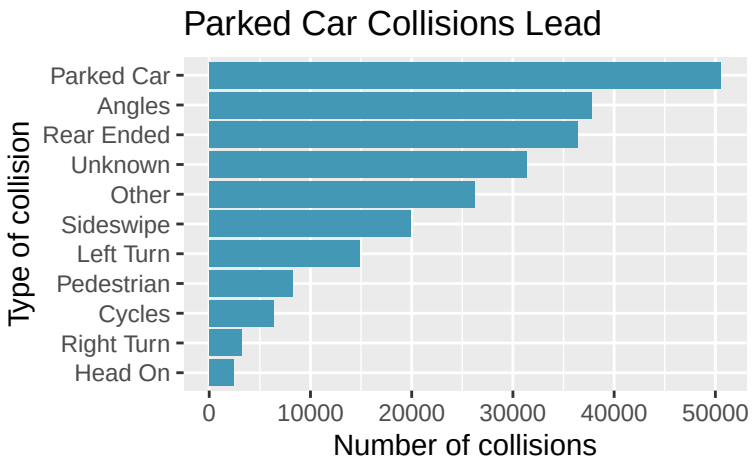


Figure 3: Number of collisions by type of collision

2. Contributing Factors

As part of this study, we would also like to determine how different road factors contribute to collisions. Therefore, the following three factors were researched: *road conditions*, *weather conditions* and *light conditions*. Figure 4 is a horizontal bar chart showing the number of collisions for different weather conditions. It's easy to see that clear weather conditions have more collision than any other weather conditions (even snowing or raining), which is not expected.

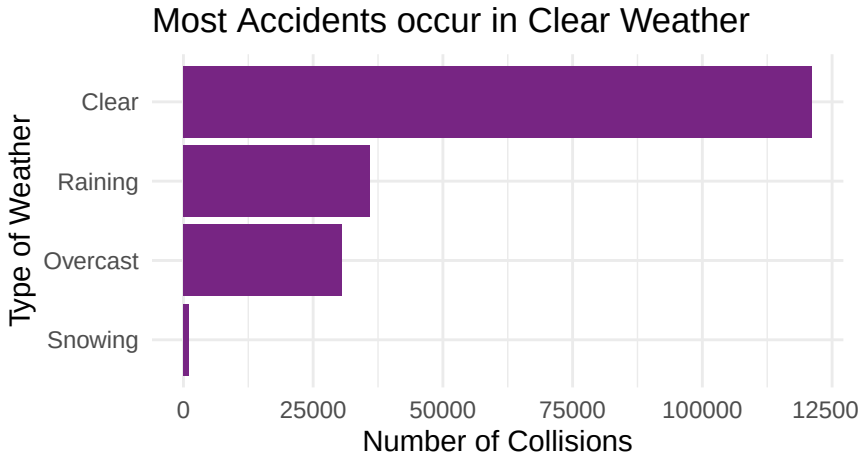


Figure 4: Effect of weather on number of collisions

In Seattle, it rains for nearly half of the year. Before analyzing the data, our team expected that most collisions would occur when the roads are wet or slippery. However, the result shows that the number of

collisions is much higher during clear days. It can be speculated that on rainy or icy days drivers could be more careful while driving. Figure 5 is a horizontal bar chart showing the number of accidents for different road conditions. The number of collisions during dry conditions are much higher than other road conditions, which is not what we initially expected. However, the number of wet collisions are much higher than Ice and Snow/Slush.

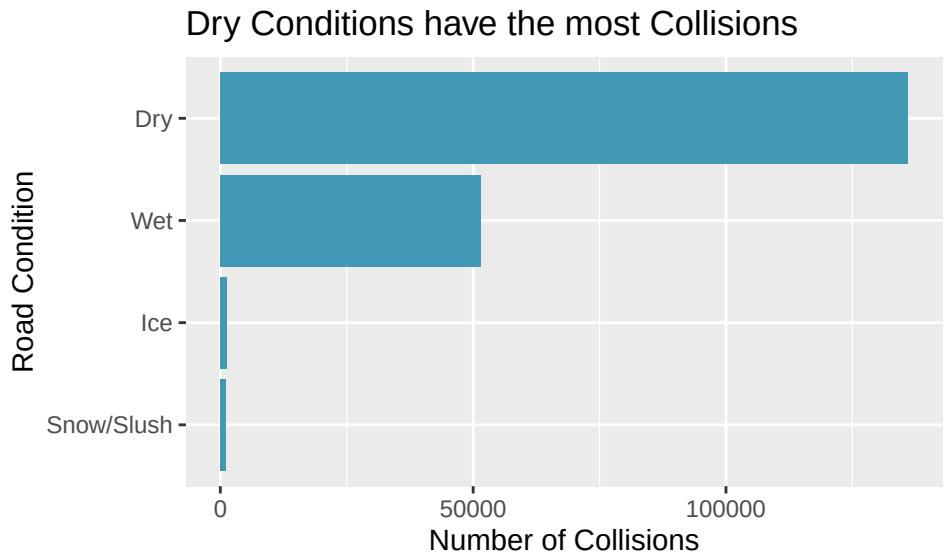


Figure 5: Number of collisions plotted by road condition

The chart below is a horizontal bar chart (Figure 6) which shows the number of collisions for various light conditions. Interestingly, the majority of accidents occurred when there is daylight. Another conclusion is that the number of collisions in dark condition when the streetlights were on was much higher than when either the streetlights were off or when there was no streetlight at all.

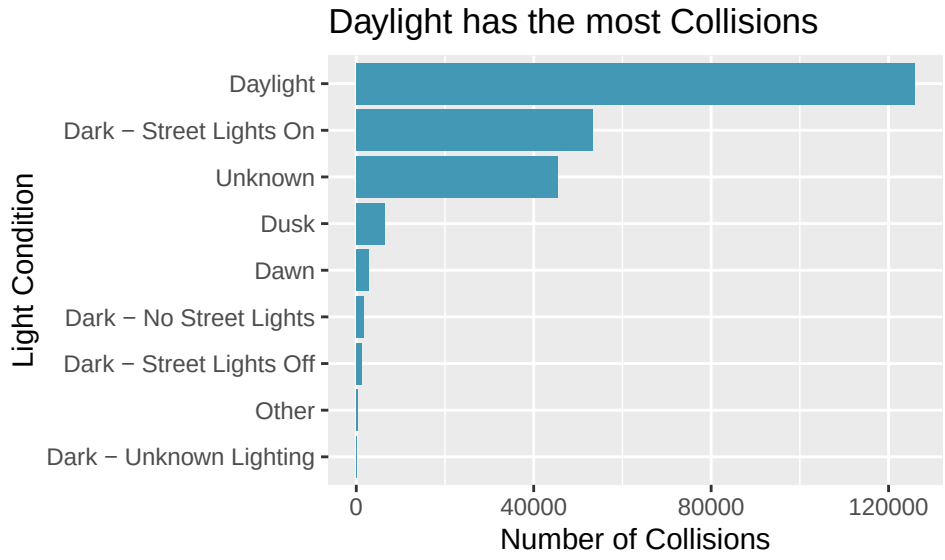


Figure 6: Number of collisions plotted by light condition

Figure 7 and Figure 8 display the number of passengers in cars during collisions and the number of cars involved in collisions. Most accidents include two passengers and two vehicles. This shows that each vehicle is typically occupied by just one passenger. One might expect that more accidents would include three or more passengers due to increased distraction. However, it seems like there are less collisions when there are several passengers in the car.

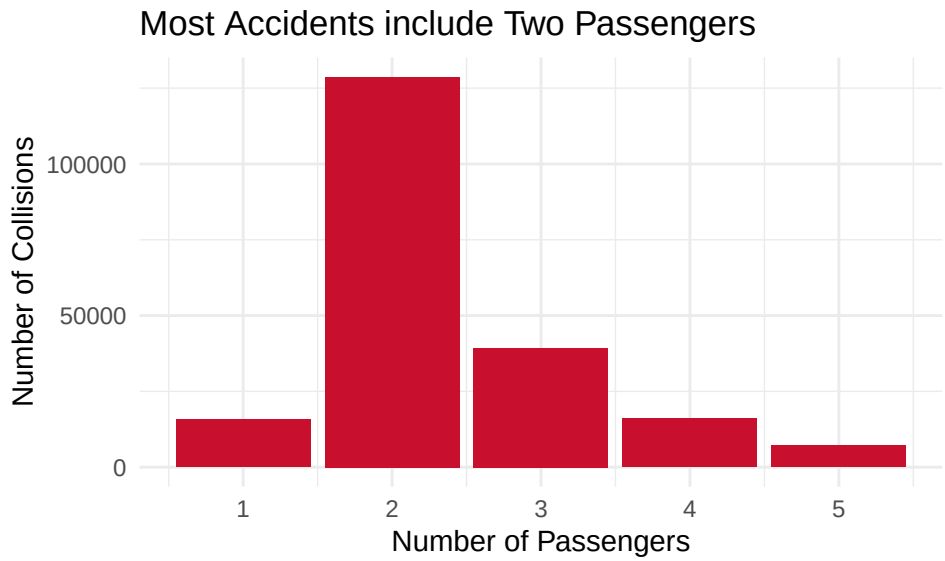


Figure 7: Number of passengers (total) in the vehicles at the accident

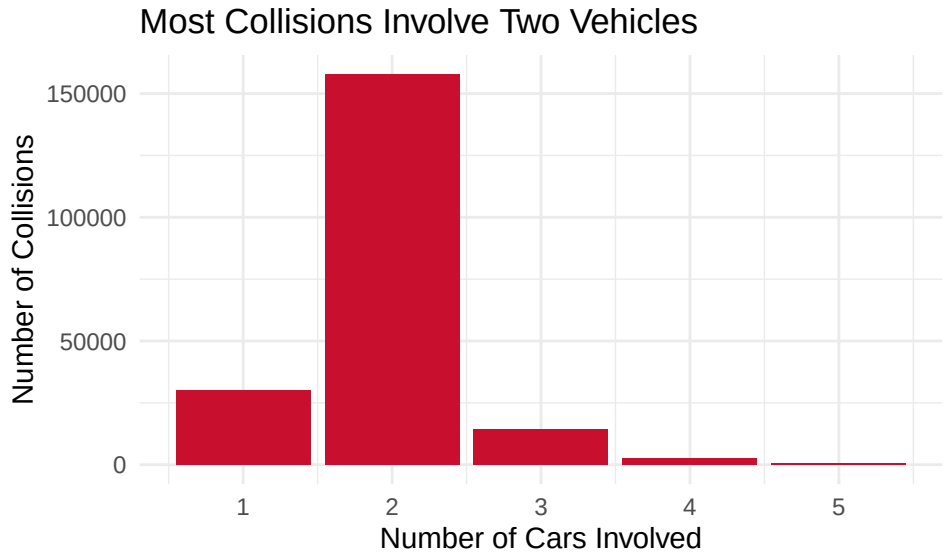


Figure 8: Number of cars in the accident

Figure 9 below shows the frequency of collision descriptions. This figure indicates that the most frequent description type is: Motor vehicle struck another motor vehicle at a front angle. After this type of collision, the motor vehicle struck motor vehicle, rear end motor vehicle, and left sideswipe are the next most frequent types.

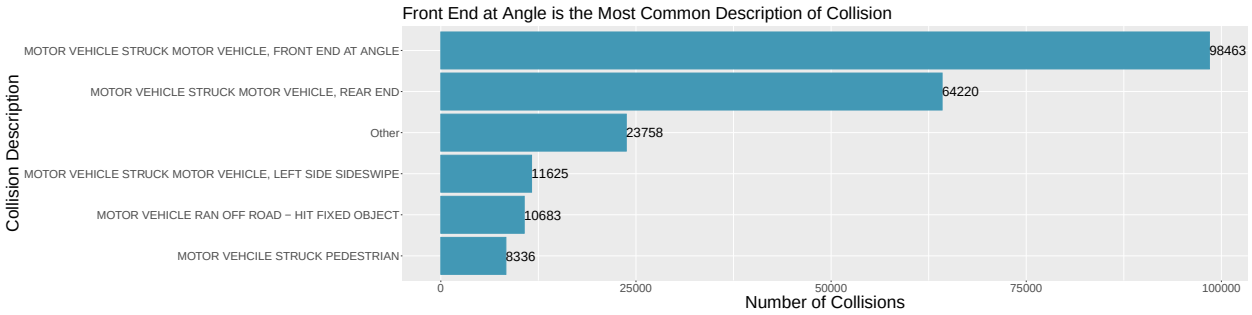


Figure 9: Top 10 most frequent collision types (all data)

Figure 10 below provides a great visualization of the top five most frequent collision types grouped by description. As you can see, front end at angle and rear end are the most frequent collision types for property damage and injury collisions. For serious injuries, the front end at angle is the most frequent followed by the motor vehicle struck pedestrians. For fatal crashes, striking a pedestrian is significantly higher than the other categories. This occurrence is not surprising since pedestrians don't have much protection when a car collides with them.

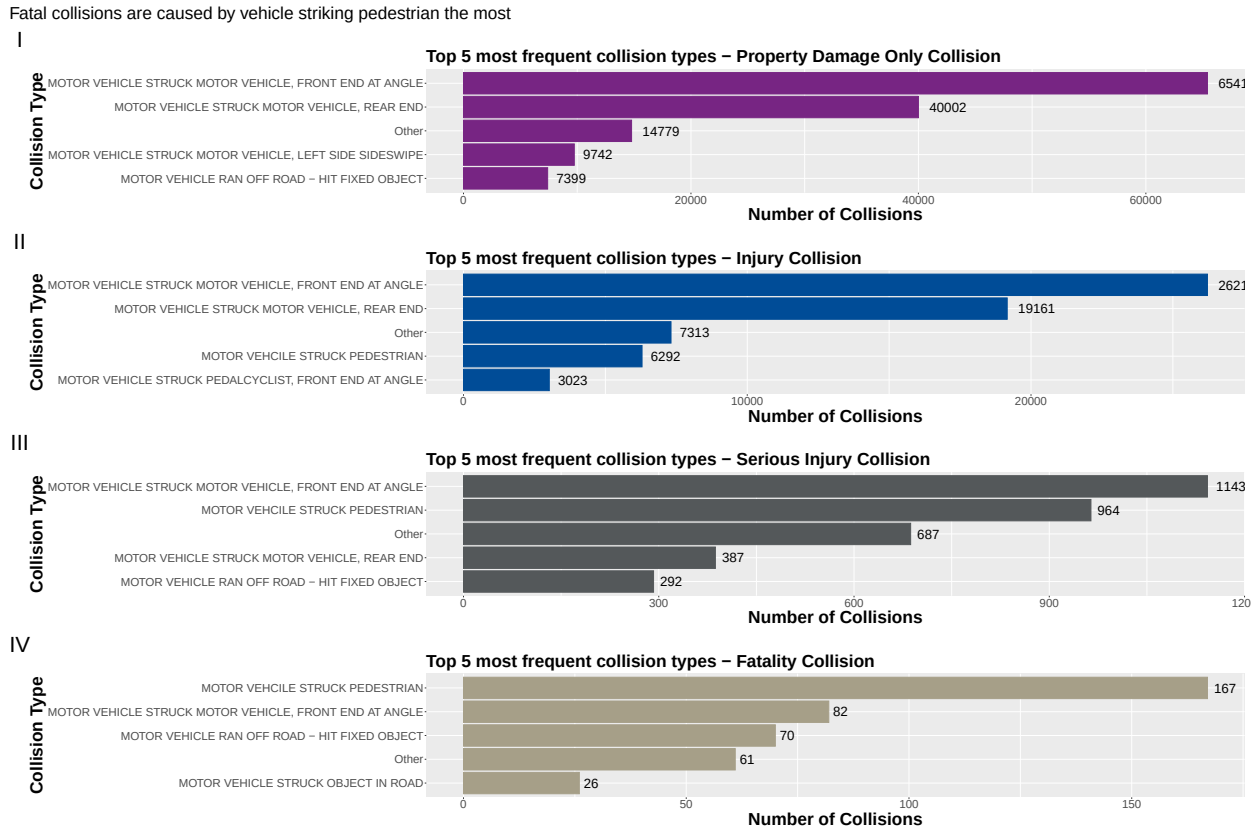


Figure 10: Collision type frequency grouped by description

3. Time Series Analysis

In order to evaluate how collisions have changed over time, we conducted a time-series analysis. As you can see, the number of collisions occurring in Seattle are decreasing. It's interesting to note that between 2010 and 2015, there was a slight increase and then in 2020 there was a sharp decrease. This sharp decrease is due to the COVID-19 pandemic. During this time, the majority of people worked from home and traveled significantly less often.

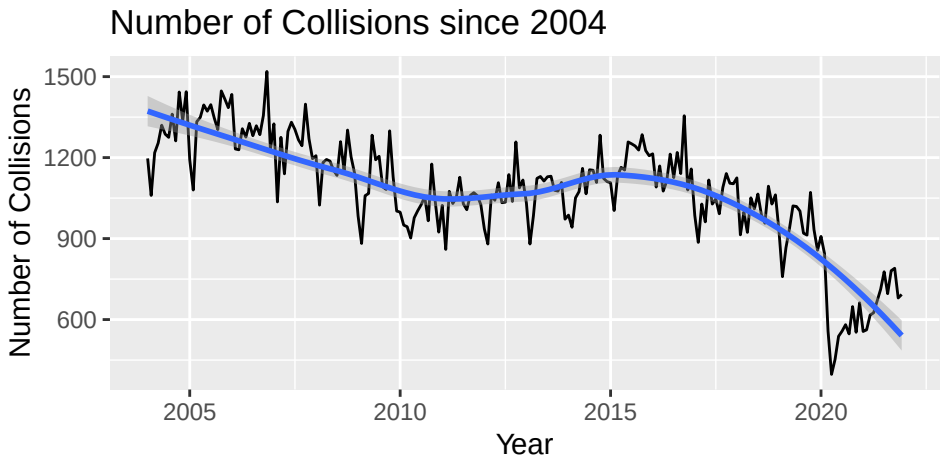


Figure 11: Seattle collision frequency between 2004 and 2021

Our team then calculated the difference between months and years to look for seasonal trends. As you can see from the plots (Figures 12 and 13) below, it appears that there is a large increase in collisions during February/March and also September/October. You can also see that March/April typically see a decline in the number of collisions. From Chart 13, we are able to see large variances during the month of March and April. These are most likely due to the COVID-19 pandemic. It's important to note that the number of collisions between months typically don't differ by more than 300 collisions.

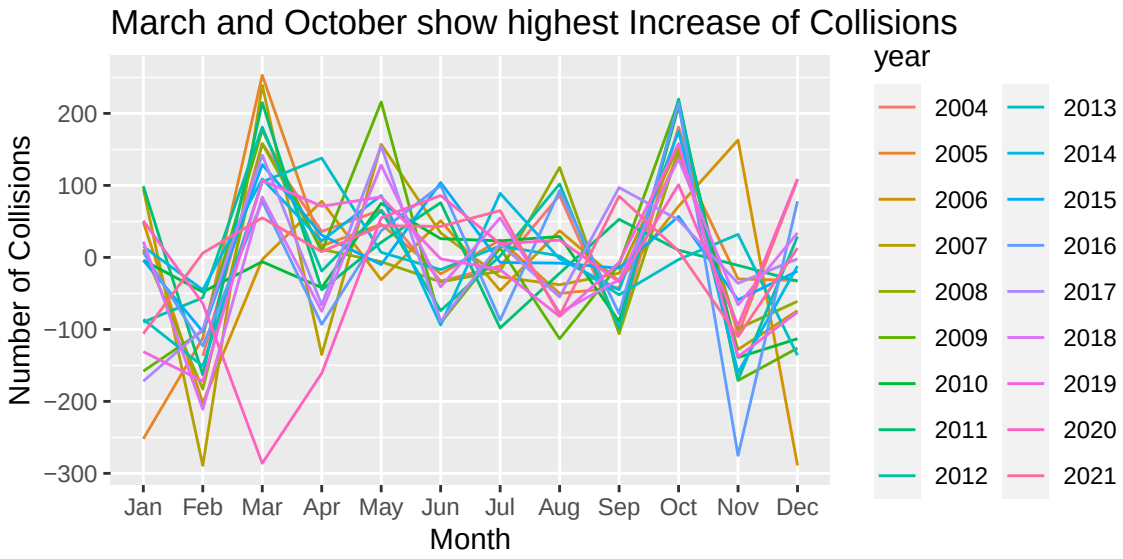


Figure 12: Plot of seasonal difference from 2004-2021

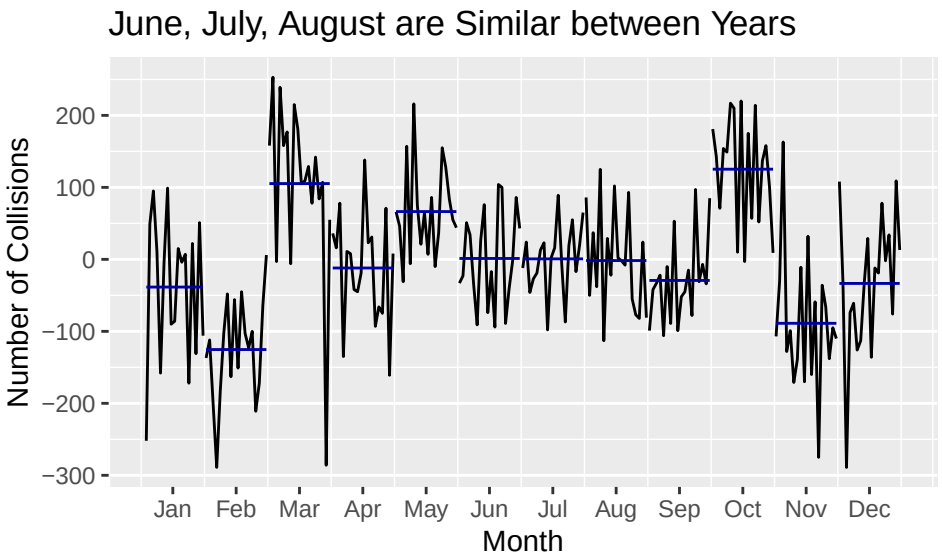


Figure 13: Plot of monthly difference from 2004-2021

We then created an auto ARIMA model to forecast the number of collisions that will occur in the future. This forecast is displayed in Figure 14. The ARIMA model that we used was a Y-Series: ARIMA(0,1,3)(2,0,0)[12]. The plot below displays the model's forecast for the first 8 months of 2022. As you can see, the models predicts that there will be 584 collisions on average for the first eight months of 2022.

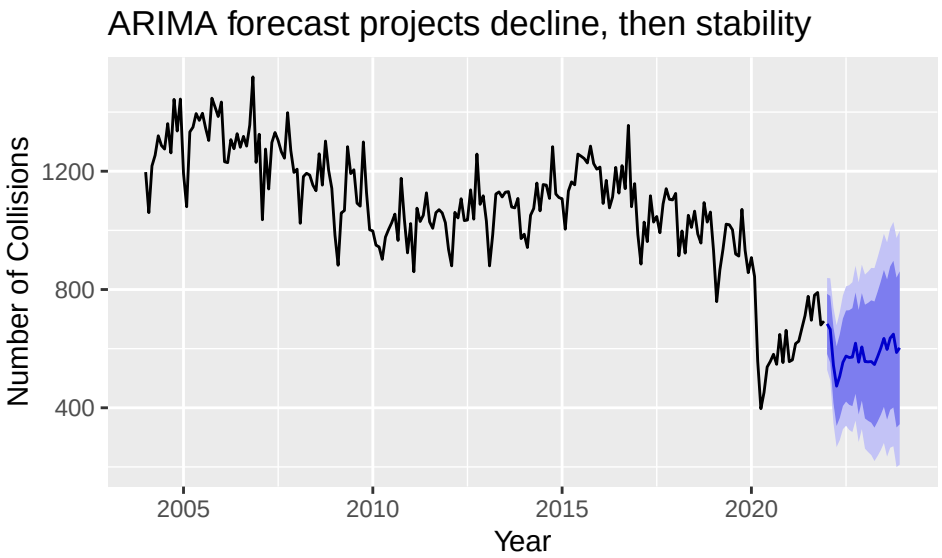


Figure 14: Collision ARIMA Forecast

In addition to the time series analysis, we wanted to determine how different collision variables such as severity has changed over time. Figure 15 below was faceted by the severity description. From the plot we can glean that the annual number of collisions has been generally decreasing, with a noticeable downtrend in the "Property Damage Only" and "Injury Collision" categories. "Property Damage Only" and "Injury Collision" categories are the most frequent severity descriptions, so a decrease over time in these categories has led to increased driver and pedestrian safety in Seattle. Fatal crashes, however, have remained relatively constant with a slight increase during the past few years.

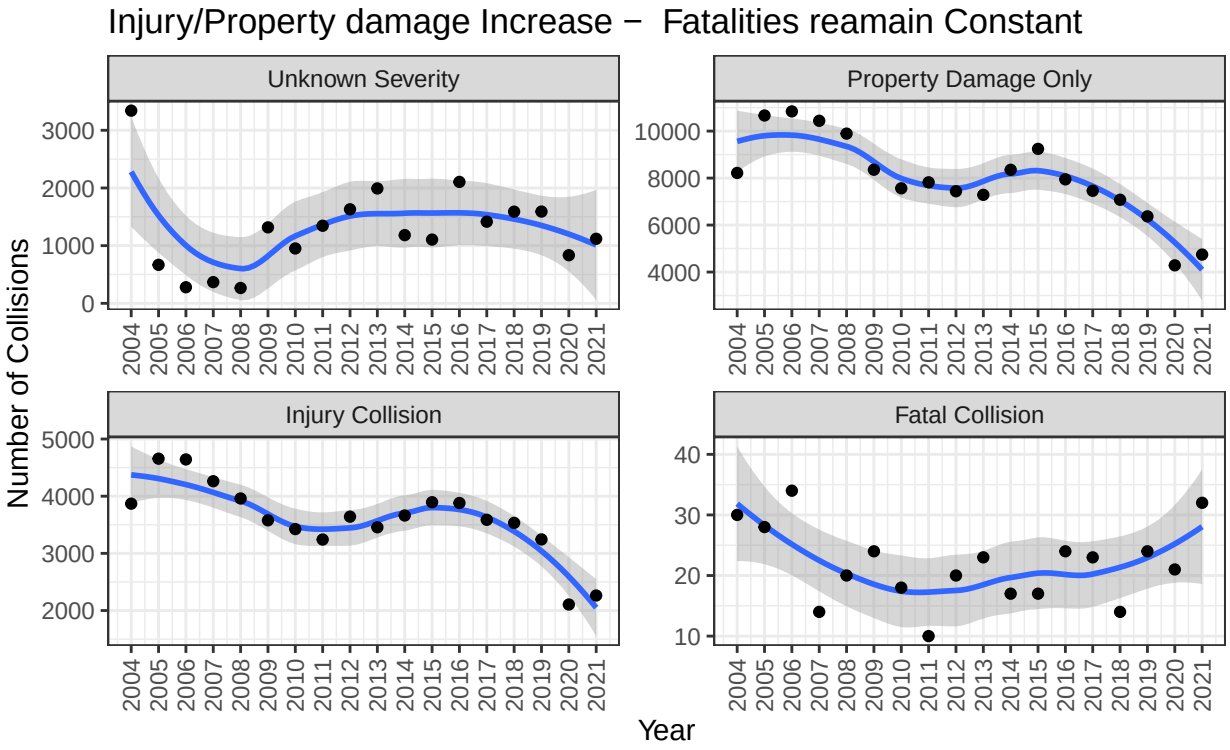


Figure 15: Annual number of collisions grouped by severity

4. Spatial Trend Analysis

Our team then analyzed for spatial trends. From the collision frequency heat map faceted by year shown below (Figure 16), you can see that there are several important characteristics to point out. Since 2004, it's easy to see that the number of crashes in southwest Seattle are becoming more frequent. It's also important to note that the majority of collisions occur in downtown Seattle. Since the 2020 pandemic, collisions in downtown Seattle appear to have decreased.

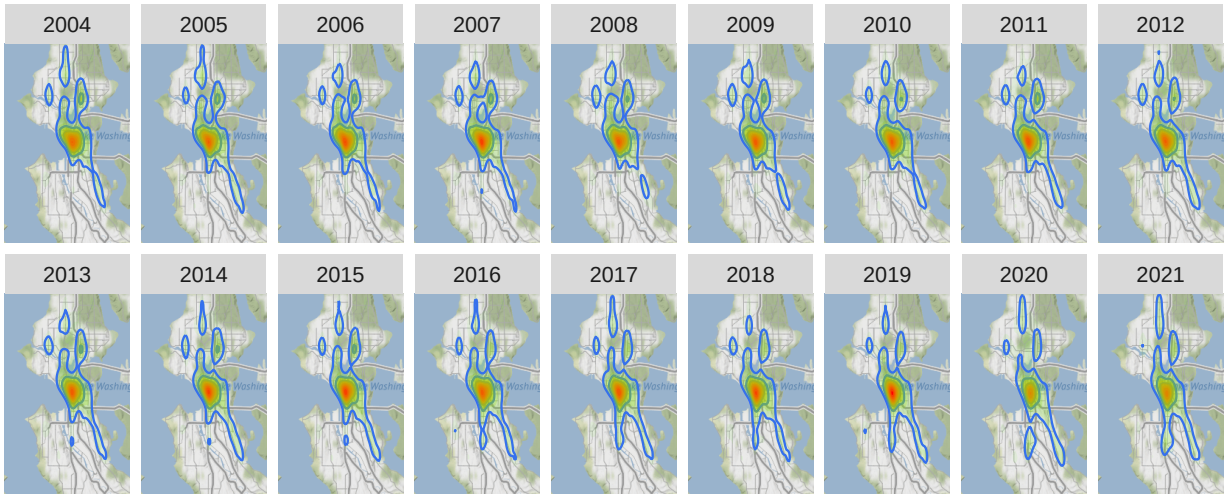


Figure 16: Collision frequency faceted by Year

We then looked at spatial trends faceted by collision type. Collision types include different categories such as Head On, Left Turn, Parked Car, Sideswipe and several others. From the heat map shown below (Figure 17), you can see that most pedestrian, cyclist, sideswipe collisions and parked car collisions occur downtown. Rear end and head on collisions occur frequently through most of Seattle.

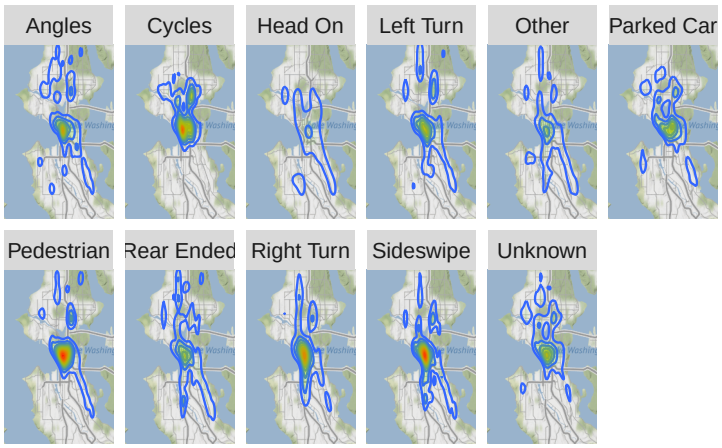


Figure 17: Collision frequency faceted by Type

Our team then determined where the majority of fatal accidents occur. From the fatality frequency heat map (Figure 18) you can see that the number of fatalities vary greatly depending on the year. The years of 2005, 2014, 2015, 2016 and 2018 had the highest numbers of fatality collisions. The locations for these collisions are mostly in the downtown area of Seattle. There are a few outliers like 2011 and 2018 where fatal accidents occurred outside of the downtown area.

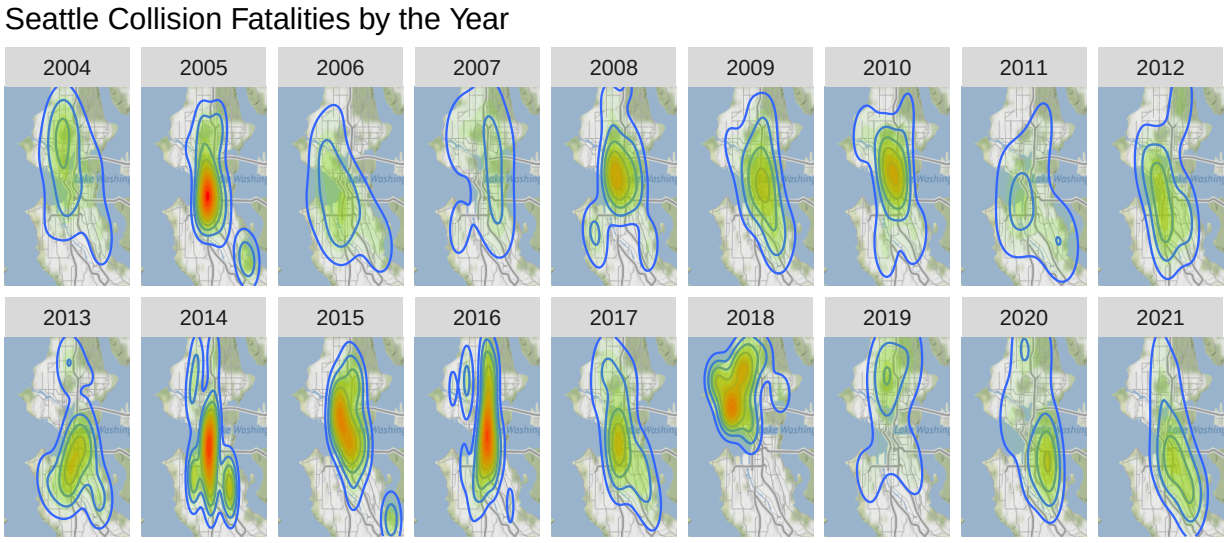


Figure 18: Fatal collision frequency faceted by Year

An additional metric that our team looked at is how speed impacts spatial trends and how speed affects fatality rate. From the heat map shown below (Figure 19) we can see that most speeding collisions occur in south Seattle and the downtown area. It's important to note that the statistical boundary highlighted in blue covers a much larger area. This indicates that speeding collisions also occur within most parts of the city.

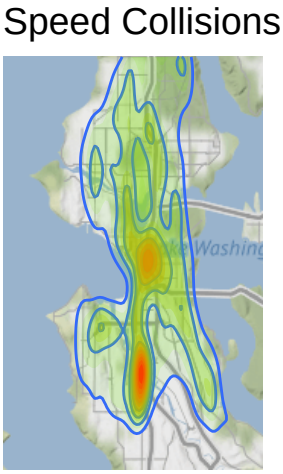


Figure 19: Speed-Collisions frequency shown on the Seattle map

Now let's look at the distribution of non-fatal collisions that involve speeding and the distribution of fatal collisions that involve speeding. As you can see from the plots below (Figure 20), for non-fatal collisions, only 5% were speeding and for fatal collisions 24% were speeding. This shows that collisions that involve speed, typically have higher amounts of fatalities.

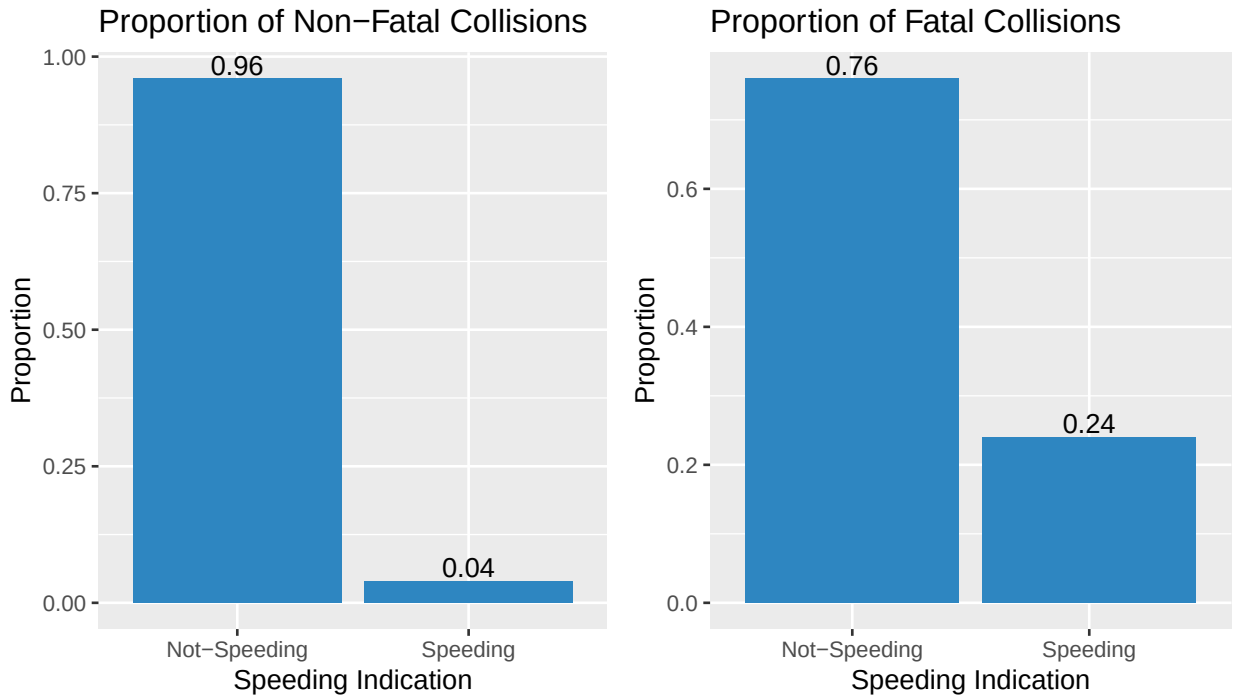
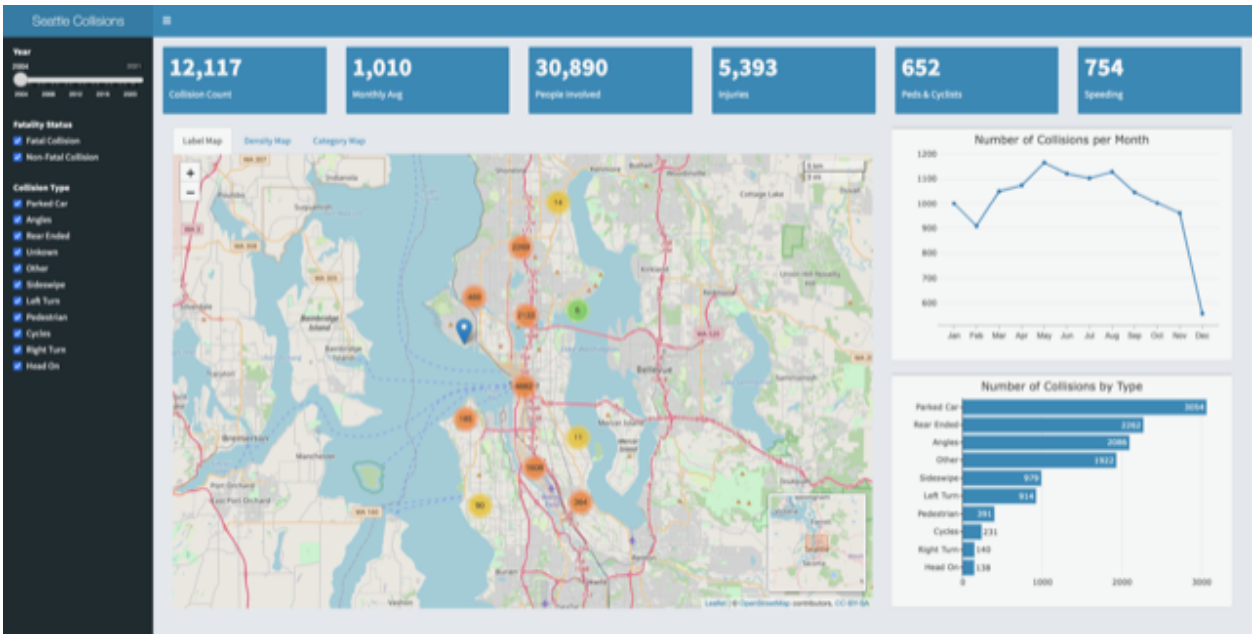


Figure 20: Proportion of Collisions Involving Speeding in (a) Fatal and (B) Non-Fatal groups

5. Seattle RShiny Dashboard

Creating a dashboard is an excellent and interactive way of dynamically visualizing data. We created a RShiny dashboard and hosted it in RStudio Connect. This dashboard is designed to visually demonstrate Seattle collision trends. This dashboard has several important and helpful features. For example, there is a filter pane that allows the end user to filter by year, fatality status and collision type. There is also a label map, density map, and dot map that displays collision data. In addition, there is a line chart, bar chart and tiles to show important metrics.

To access the application, please visit this link: https://beta.rstudioconnect.com/content/2b5fef76-e01f-469e-96ad-3976076c704e/



You may also view our interactive Quarto Presentation by using this link: https://awick10.quarto.pub/seattle-collisions/#/title-slide

Conclusions

Vehicular collision data in the city of Seattle from 2004 to present was utilized in an analysis of what conditions contribute to vehicle collisions and their severity. It was found that the number of total collisions has had a decreasing trend during the given time range, with the largest decrease being seen among property damage and injury collisions, which are the two most common collision types.

Looking at the type of collisions, it was found that the majority of collisions occur when vehicles are traveling at angles or during a left turn. These two collision types account for over 65% of all collisions. Of the fatal accidents that occurred, the majority were attributed to a vehicle striking a pedestrian. These occurrences account for approximately half of all fatalities, or the same amount of all types of vehicle-vehicle fatal collisions combined.

When analyzing the relationship of weather conditions to number of collisions, it was found that the majority of collisions occurred during clear conditions. Clear conditions accounted for 64% of total collisions, whereas only 20% occurred during the rain. However, considering it rains 10% of the time in the city of Seattle, it seems collisions are more probable during rain than may be initially perceived.

The location data of the collisions were used to see what area of Seattle most collisions occur. It was found that year over year, the downtown area of the city had the highest amount of collisions. Specifically looking at the type of collisions that occur in each part of the city, it was found that pedestrian accidents and sideswipes were most common in this high density area.

Collisions in Seattle have been consistently decreasing since 2004. This is expected as SDOT has taken action to increase the safety of drivers while vehicle safety technology has also progressed. This study shows that while the amount of collisions is decreasing, there are still areas where collision prevention and safety can be bettered. The given results suggest changes in vehicle traffic laws in high collision locations to further increase safety for those in the city of Seattle, WA.

Document Reproducibility

This work is conducted employing the R package Knitr and can be compiled and edited for further research and analysis.

References

[1] WHO (2015) WHO Global Report on Road Safety 2015.

[2] NCSA (2015) NHTSA-National Center for Statistics and Analysis.

[3] Mohan, D. (2002) Road Safety in Less-Motorized Environments: Future Concerns. International Journal of Epidemiology, 31, 527-532.

[4] Elvik, R. (2006) Laws of Accident Causation. Accident Analysis and Prevention, 38, 742-747.

[5] NCSA (2016) NHTSA-National Center for Statistics and Analysis.

[6] SDOT (2022) SDOT Collisions - All Years, https://catalog.data.gov/dataset/collisions-all-years-be528.