A Neighborhood and Temporal Analysis of Toronto's Car Accidents*

My subtitle if needed

Cristina Burca

September 27, 2024

Abstract

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of contents

*Code and data are available at: LINK.

1 Introduction

Toronto is a city infamous for its heavy traffic, recently ranked as the city with the third worst traffic in the world (cite). The traffic is due to several factors. Toronto has experienced one of the fastest-growing populations in North America, with an increase of approximately a million residents in the past 15 years. The city has not kept pace in growth in terms of road infrastructure or public transport. Numerous planned road expansion and maintenance projects have been delayed, resulting in prolonged construction periods and bottlenecks across major routes. Furthermore, with years of under-investment in public transit and only two major transit lines leading into the center of the city, public transport has become overcrowded and unreliable. It is difficult for commuters to choose public transport over driving, and thus a heavy car dependency has been established.With roads heavily congested during peak hours, accidents have become a frequent occurrence. On average, there are more than 4 documented collisions each day.These accidents not only cause significant disruptions to daily commutes but become a saftey concern for commuters.

As accidents have become a common occurrence, it is crucial to analyze their underlying causes in order to implement effective solutions. Understanding where and when these accidents happen can provide valuable insight into how to prevent them. This paper aims to identify accident hotspots across Toronto's neighborhoods, as well as examine the frequency of accidents during specific hours of the day. By analyzing these patterns, we hope to give insight to policy makers on the infrastructure, traffic flow, or road design issues that might be contributing to higher accident rates.

2 Data

2.1 Data source and referencing

The data is sourced from the City of Toronto Open Data, a portal containing Licensed official data of Toronto. Used R to compile this paper as well as packages....

2.2 Introduction to the Data

This dataset consists of 18,763 observations of automobile accidents in the Region of Toronto, recorded from January 1st, 2006, to December 29th, 2023. It includes six key variables of interest:

- `Date`: the date of the accident,
- `Time`: the time of the accident,
- `Street 1` and `Street 2`: the nearest intersection where the accident occurred,
- `Hood`: the neighborhood ID (Toronto is divided into 158 neighborhoods),
- `Injury`: identifying whether injuries occurred and their severity,
- `geometry`: containing the latitude and longitude coordinates of the accident location.

The data was cleaned to ensure that the term "automobile accidents" includes any incidents involving cars, trucks, motorcycles, transit vehicles, or emergency vehicles. Accidents involving pedestrians or cyclists are also included. Local or side roads are excluded form this data, This comprehensive approach allows for an inclusive analysis of all types of collisions on Toronto's streets, providing a holistic view of the traffic and safety challenges.

2.2.1 Missing Data

This data offers a diverse and detailed array of accidents in the greater Toronto area, however due to the nature of accidents it cannot be said to be entirely complete. For example accidents may not always be reported and thus would not be included in this paper. accidents may not be reported due to a variety of reason such as hit and runs or accidents handles without external interaction. In this paper we will focus on areas of high frequency as they are of the most interest to the question of .... (where po0licy makers should focus their efforts. Doing this however does make our analysis miss impiortent context for baseline levels in other less interesting neighborhoods.)

2.3 Measurement?

3 Data Overview

The map in Figure 1 illustrates high-density accident points across Toronto. The darker and larger circles represent intersections or areas with a higher number of accidents. These accident hotspots are concentrated along major roadways and intersections, as expected. The densest regions appear in Downtown Toronto, particularly near major highways, such as the intersections along Highway 401, 427, and 409. Additionally, several clusters appear in densely populated neighborhoods, such as North York and Scarborough.

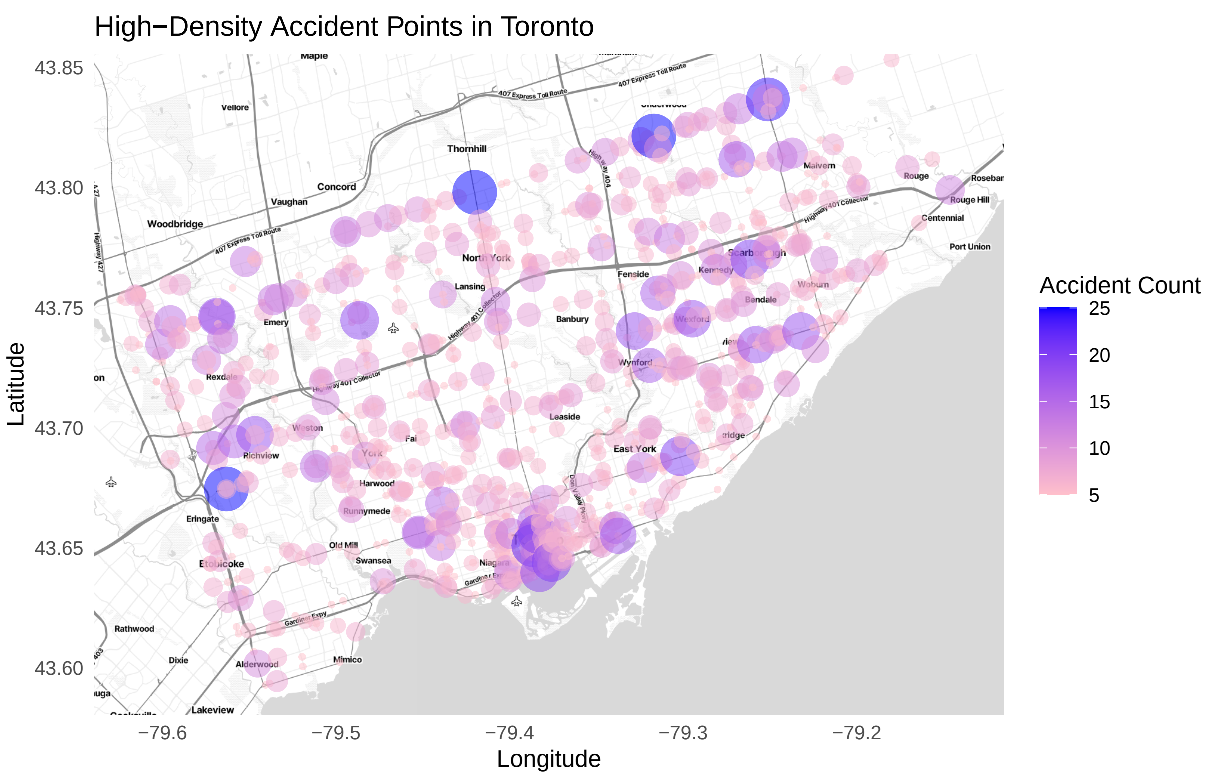


Figure 1: somethinf

In Figure 2, the top five neighborhoods with the highest number of accidents are presented. Neighborhood 1 (West Humber-Clairville, Etobicoke) stands out with the most accidents, followed by neighborhoods like Wexford/Maryvale (119), Yonge-Bay Corridor (170), St Lawrence-East Bayfront The Islands (166) and South Riverdale (70). The number of accidents in each of these neighborhoods points to potential urban planning challenges, such as high traffic density, poorly designed intersections, or insufficient traffic control measures. For example, the West Humber-Clairville area, with its proximity to major highway interchanges, experiences a high volume of vehicular movement, which may contribute to its elevated accident rates. Similarly, the Yonge-Bay Corridor, located in the heart of the city's financial district, may experience heavy traffic due to commuter flows, particularly during peak hours.

put an image of toronto with major highways, roads nad divided by hood change grpah least to greatest

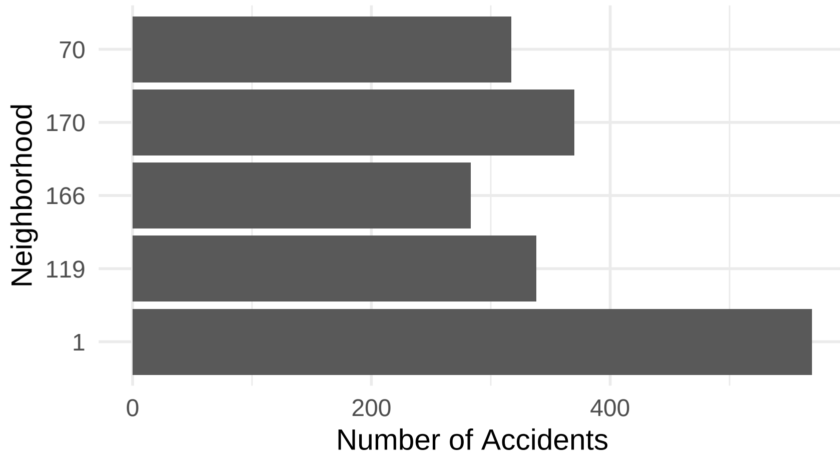


Figure 2: somethinf

4 Neighbourhood Analysis

The map below Figure 3 provides an overview of the top 5 neighborhoods in Toronto with the highest number of accidents from 2006 to 2023. Each point represents a specific location where an accident occurred, with the size and color of the point reflecting the accident count in that area. Darker and larger points indicate areas where accidents are more concentrated.

This high-level visualization helps to identify Toronto's most accident-prone neighborhoods. Notably, these areas include regions such as West Humber-Clairville in the west, the downtown core, and Wexford/Maryvale in the east. As we zoom in on each of these regions, the finer details of accident distribution, high-traffic intersections, and potential infrastructure weaknesses become more apparent.

4.1 South Toronto

In Figure 4, Accidents in South Toronto, we see that accidents are heavily concentrated in the downtown core, particularly around major intersections such as those along Lake Shore Boulevard West, Queen Street, and University Avenue. These areas are likely to experience high traffic volumes due to their proximity to key business, entertainment, and tourist districts. The density of accidents suggests that congestion, pedestrian movement, and possibly outdated traffic management systems contribute to the high accident counts.

Many of these points are located near high-traffic public areas and close to transit hubs, such as Union Station and the Gardiner Expressway. This indicates that improving traffic management and pedestrian crossings in these regions may help reduce accidents.

4.2 West Toronto

Figure 5 provides a closer look at the neighborhoods around West Humber-Clairville. This neighbourhood is located on the West edge of Toronto, consisting of the Clairville, West Humber neighbourhood in Etobicoke-York, and intersections of major highways in Toronto, including the 427, 407, 409, and 401. The accident hotspots are primarily found along major arterial roads and intersections, such as those around the 427 and Highway 401 interchange. These intersections are known for their high traffic volumes, especially during rush hours, and are major connectors for drivers entering and leaving the city.

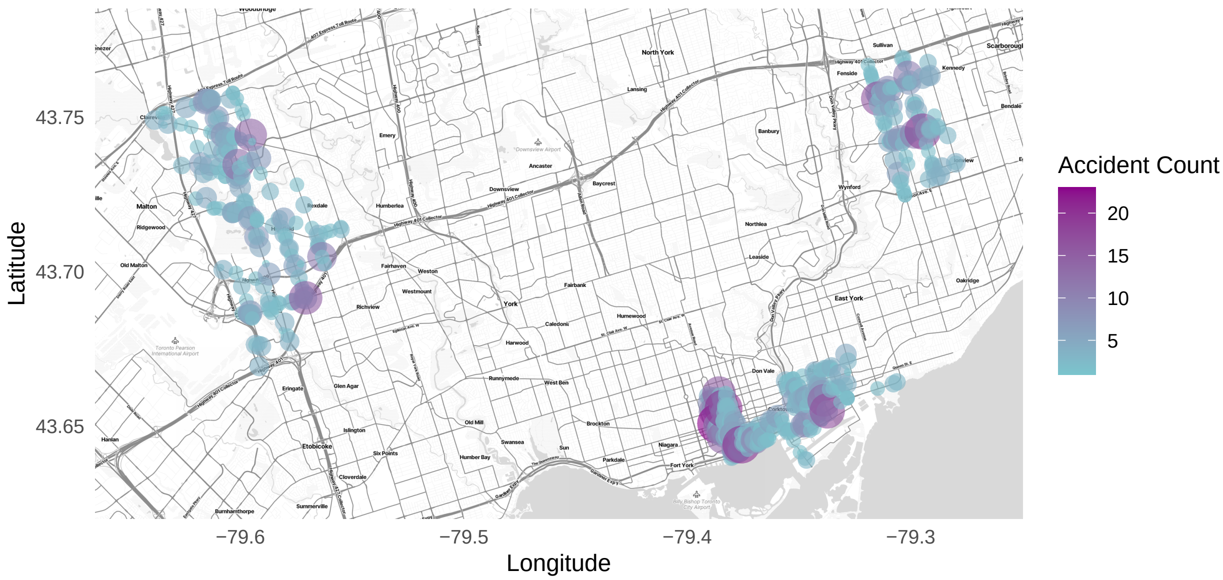


Figure 3: somethinf

Figure 4: SOUTH

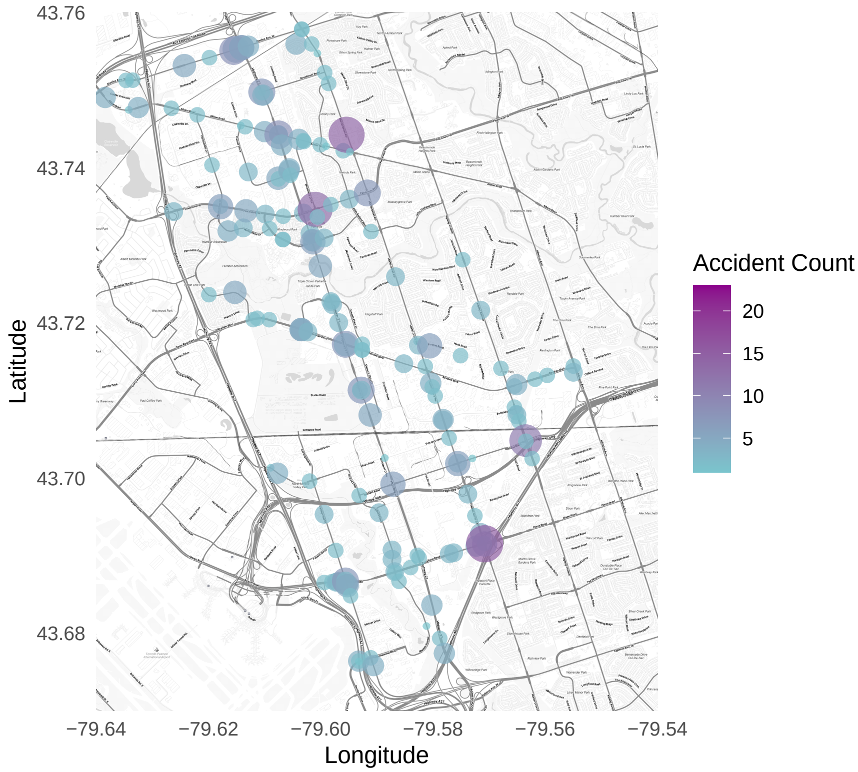


Figure 5: NEIGHBOURHOOD 1

The patterns in this map suggest that large traffic volumes, with high-speed highways and low-speed merging, contribute to the number of accidents. Several high-accident intersections are located near residential and commercial areas, suggesting that drivers are merging from side streets into faster-moving highway traffic, which could increase the risk of collisions.

4.3 East Toronto

In Figure 6, the accidents are concentrated along major streets, including Ellesmere Road, Warden Avenue, and Lawrence Avenue. These streets connect various residential neighborhoods to commercial areas and are likely to experience heavy traffic during commuting hours. The distribution of accidents along these streets suggests that traffic congestion, signal timing, and road design may contribute to the higher accident rates.

Similar to West Toronto, the larger points appear at major intersections, where traffic volumes are likely highest. The proximity of these accident hotspots to residential areas indicates a potential need for traffic calming measures or infrastructure improvements to reduce accident occurrences.

The neighborhood-level analysis highlights clear accident hotspots across South, West, and East Toronto. In South Toronto, the downtown core sees high accident counts due to dense traffic, pedestrian activity, and congested intersections. West Toronto's accident hotspots are concentrated around major highway interchanges, pointing to the complexities of merging high-speed traffic with local routes. East Toronto experiences clusters of accidents along heavily trafficked residential and commercial connector roads. These insights suggest that a combination of infrastructure improvements, better traffic management, and targeted safety measures could help reduce the frequency of accidents in these high-risk areas.

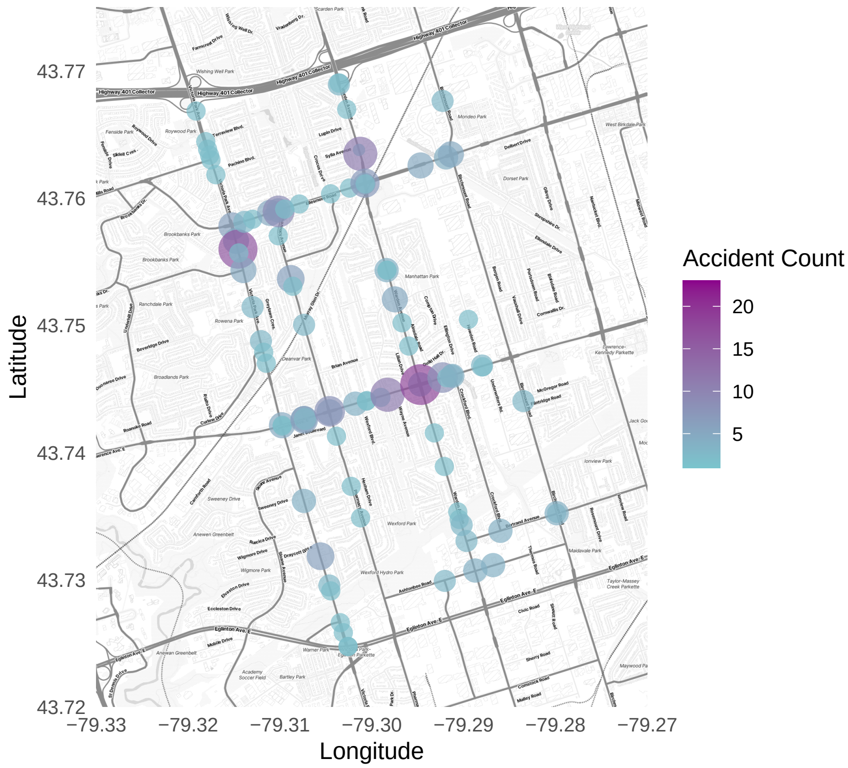


Figure 6: East side Toronto

5 Month, Week and Day Analysis

5.1 Hours

The first graph, Figure 7, a pattern emerges regarding the timing of accidents throughout the day. There is a gradual rise in accidents from the early morning (6:00 AM onwards), peaking sharply between 3:00 PM and 6:00 PM. This time period coincides with the evening rush hour, which is when traffic density is at its highest as people commute home from work or school. The fact that the number of accidents peaks in this window highlights the heightened risk associated with rush-hour driving.

Notably, there's also a significant number of accidents occurring late at night, particularly around 11:00AM-12:00 AM. This could be attributed to factors such as impaired driving, fatigue, or lower visibility at night. The data further reveals that fewer accidents occur in the early morning hours (3:00–5:00 AM), a time when traffic volumes are generally lower.

The steady rise throughout the afternoon suggests that as the day progresses and more vehicles are on the road, the likelihood of accidents increases.

5.2 Week

Figure 8 we observe that accidents are relatively consistent from Monday through Thursday, but there is a notable increase on Fridays. The spike in accidents on Fridays could be related to the increased volume of vehicles as people begin their weekend travels or run errands after work. There may also be a higher incidence of risky driving behaviors, such as speeding or impaired driving, as the weekend approaches.

Interestingly, accident numbers slightly dip on Saturdays and Sundays compared to weekdays, likely due to lower traffic volumes as fewer people commute to work. However, despite the lower traffic on weekends, the

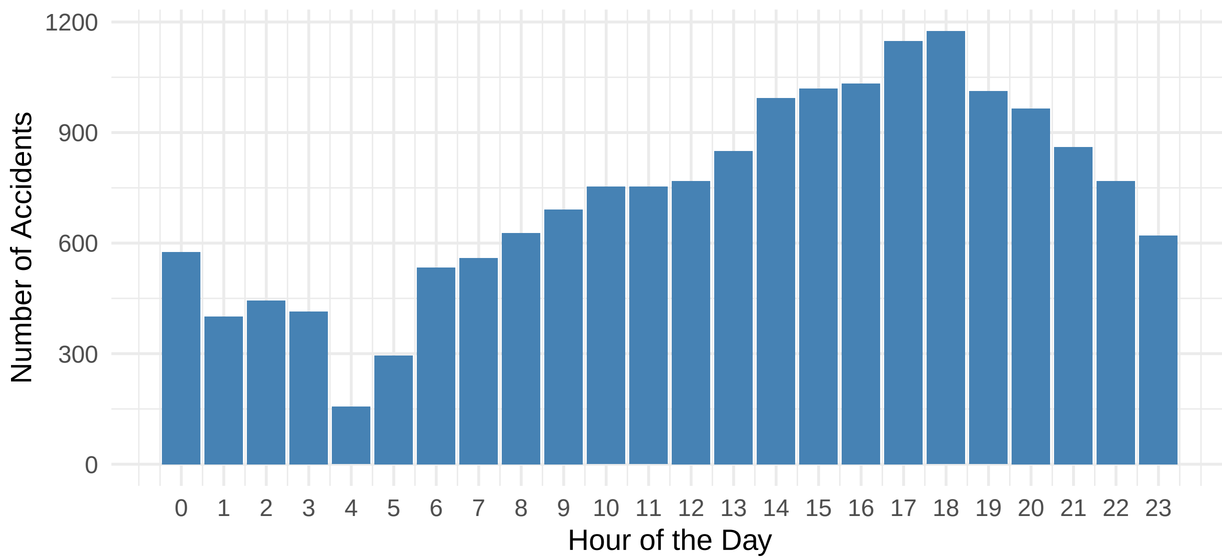


Figure 7: Hours

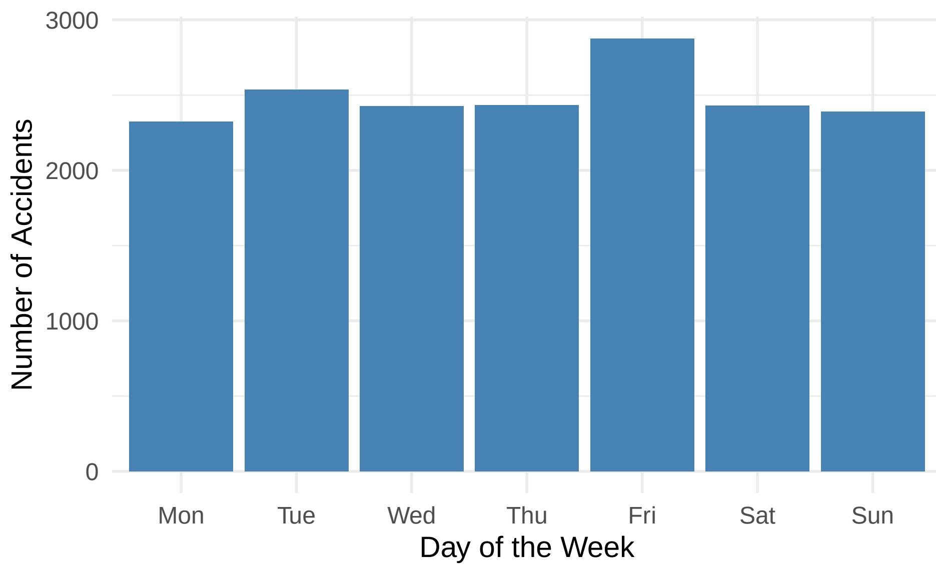


Figure 8: Week

number of accidents remains substantial, suggesting that other factors—such as recreational travel, shopping, and leisure activities—contribute to the accident count.

5.3 Months

Looking at the **?@fig-months**, there are several notable patterns:

Summer Peaks (June–September) The data shows a significant increase in accidents during the summer months, with June, July, August, and September seeing the highest number of accidents. This could be attributed to several factors:

- Increased travel: People tend to travel more during the summer months, whether for vacations, road trips, or simply taking advantage of better weather, leading to increased traffic volume.
- Cyclists and pedestrians: With warmer weather, more cyclists and pedestrians are on the roads, which can contribute to an increased risk of accidents involving these groups.
- Recreational driving: Summer also sees an uptick in recreational driving and holiday travel, which may contribute to the rise in accidents.

Interestingly, while one might expect the winter months (January, February, December) to have the highest accident rates due to icy and snowy conditions, these months have relatively fewer accidents compared to the summer peak. However, they still have substantial numbers, possibly due to:

- Hazardous driving conditions: Snow, ice, and poor visibility may contribute to the number of accidents, though this is offset by potentially lower traffic volumes in winter months as people avoid unnecessary travel.
- Holiday travel: The spike in December could be associated with increased travel during the holiday season, when more people are on the roads visiting family or shopping.

This analysis of accident trends by month provides a clearer picture of how seasonal factors influence road safety in Toronto. The data suggests that targeted measures during peak accident months (such as improving safety for cyclists and pedestrians in the summer) could help reduce accident rates.

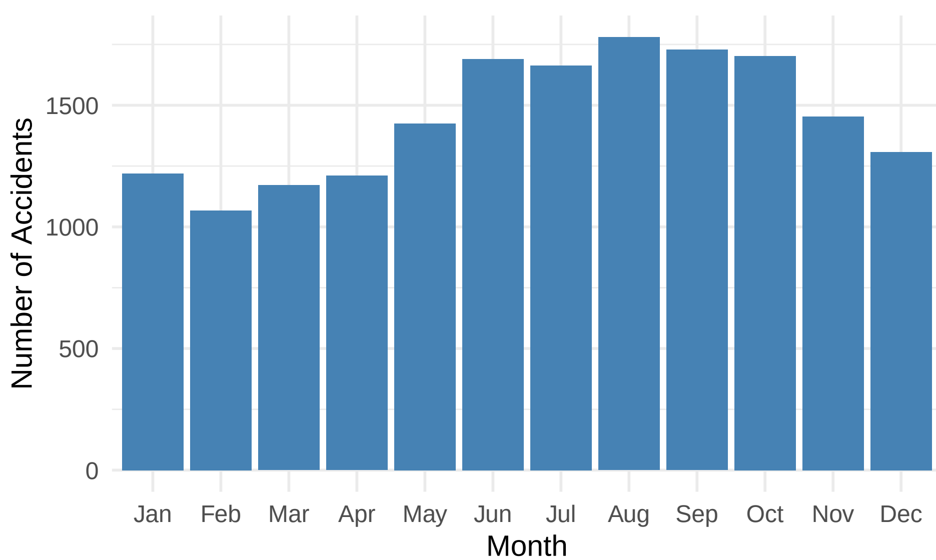


Figure 9: Month

```r
rush_hour_data <- clean_data %>%
  filter(hour >= 15 & hour <= 18)

# Summarize accidents by neighborhood and hour
top_neighborhoods_by_hour <- rush_hour_data %>%
  filter(HOOD_158 %in% c(1, 70, 119, 166, 170)) %>%
  group_by(HOOD_158, hour) %>%
  summarise(total_accidents = n()) %>%
  ungroup()
```

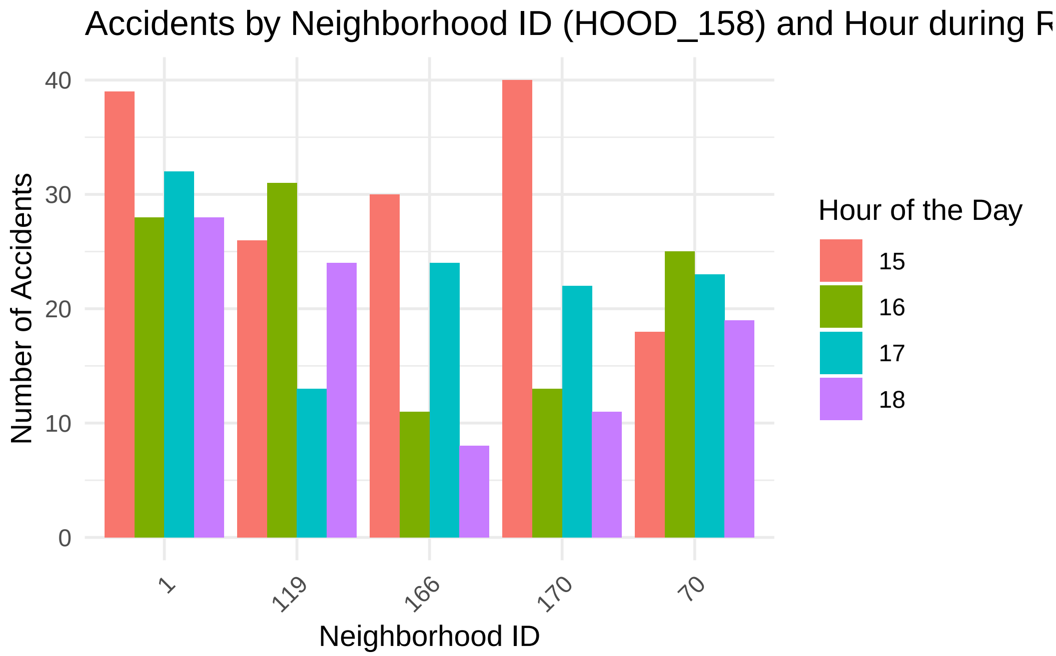
```
`summarise()` has grouped output by 'HOOD_158'. You can override using the
`.groups` argument.
```

```r
avg_accidents_by_hour <- rush_hour_data %>%
  filter(HOOD_158 %in% c(1, 70, 119, 166)) %>%
  group_by(hour) %>%
  summarise(average_accidents = mean(n())) %>%
  ungroup()

avg_accidents_by_hour %>%
  arrange(hour) %>%  # Order by neighborhood and then by hour
  kable(col.names = c("Hour of Day", "Average Accidents"))
```

Hour of Day	Average Accidents
15	113
16	95
17	92
18	79

```r
# Create the bar plot for the specified neighborhoods
top_neighborhoods_by_hour %>%
  ggplot(aes(x = as.factor(HOOD_158), y = total_accidents, fill = as.factor(hour))) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Accidents by Neighborhood ID (HOOD_158) and Hour during Rush Hour",
       x = "Neighborhood ID", y = "Number of Accidents", fill = "Hour of the Day") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



8. Clustering of Accidents by Time and Location Using a combination of time and location data, you could identify specific intersections or streets that experience frequent accidents during certain hours (e.g., a particular road might be especially dangerous during evening rush hour). This would allow for a more targeted analysis of dangerous locations at peak times.

6 Discussion

7 Conclusion