FIT5147 Data exploration and visualisation S1 2022

Name: Prachi Jaiswal,

Student ID: 32192673,

Tutor: Mohit Gupta and Abishek Sriramulu, Tutorial 01.

Table of Content

Analysis on the Frequency of Road Crash in Australian Capital Territory.

1 . Introduction

In a country as vast as Australia, driving is an essential part of life. Driving here is not a luxury but a necessity. The number of individuals killed or wounded in road accidents in Australia was always a concern for which the government imposed a set of strict road safety rules and regulations. Due to such directives, road accidents have seen quite a noticeable fall in the number of unforeseen contingencies. As a result, the following facts prompted me to learn more and investigate how the crashes have decreased over time. As a responsible resident, I also need to educate myself on road accident data and patterns as they develop and evolve. Therefore, I decided to conduct a study on the frequency of road crashes in order to learn more about the elements and places that contribute to these collisions.

2 . Data Description

The dataset has been downloaded from ATC Government Open Data Portal subcategorised as ACT Road Crash Data. With more than 7000 rows and 14 columns, the following data consists of both tabular and geographic details. It contains the latitude and longitude points of traffic incidents documented in the Australian Capital Territory from 2012 to 2021. The police or the general public have entered these into the “AFP Crash Report Form.” This dataset only includes crashes reported through the “AFP Crash Report Form”; any other accidents are not included (“ACT Road Crash Data”, 202, para.1).

3 . Data Wrangling

When it comes to data analysis, your results are only as good as your data (What Is Data Wrangling & Why Is It Necessary? 2021). And to report accurate results, the data must be cleaned. Likewise, below is the table that shows the raw data converted to a desired dataset for the analysis. The columns in Italics will be removed as they are not needed for analysis.

Few columns from the original dataset were removed and underwent some cleaning for a better and more accurate analysis. For example, columns like Crash Date and Crash Time were broken down into different columns and their datatype was also changed.

3.1 Data Structure

Sr. No.	Column Name	Data Type	Data Description
1	Day	<int> Integer	Date of the month when crash occurred
2	Month	<int> Integer	Month of the year when crash occurred
3	Year	<int> Integer	Year in which crash occurred
4	Hour	<int> Integer	Hour of date when crash occurred
5	Minute	<int> Integer	Minute of the hour when crash happened
6	Sec	<int> Integer	Second of the minute when crash occurred
7	SUBURB_LOCATION	<chr> Character	Suburb where the crash occurred
8	LONGITUDE	<dbl> Double	Longitude coordinate on which crash occurred
9	LATITUDE	<dbl> Double	Latitude coordinate on which crash occurred
10	CRASH_DIRECTION	<chr> Character	Direction of travel at the location of crash
11	CRASH_SEVERITY	<chr> Character	Level of crash severity
12	LIGHTING_CONDITION	<chr> Character	Lighting condition when the crash occurred
13	ROAD_CONDITION	<chr> Character	Road condition where the crash occurred
14	WEATHER_CONDITION	<chr> Character	Weather condition when the crash occurred

3.1 Data Dimension

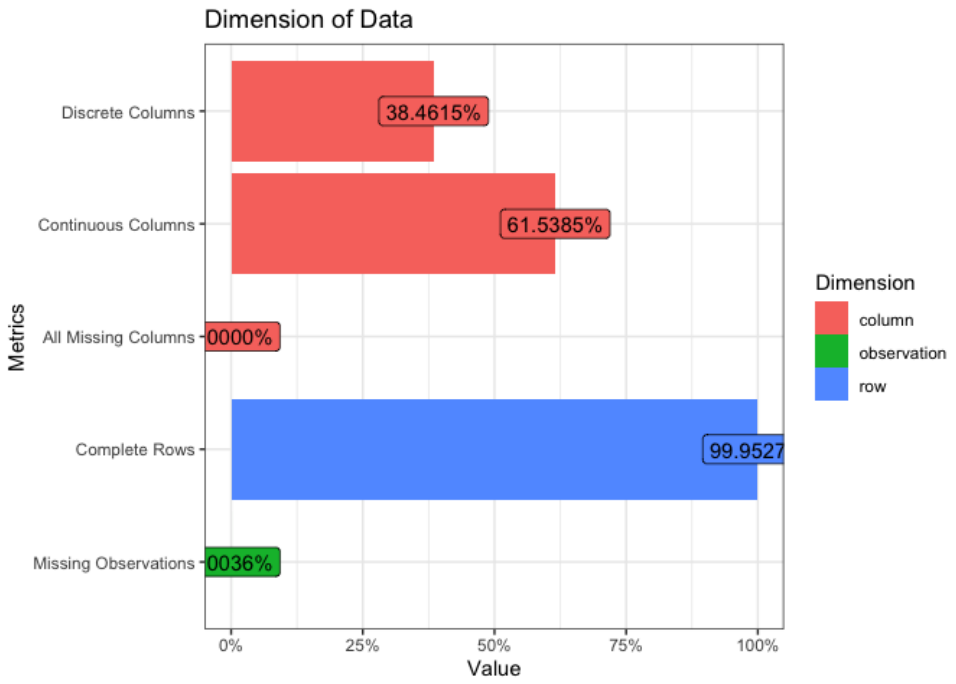

Figure 1

Figure 1 provides basic facts and figures about discrete columns, continuous columns, the total number of null values, and missing observations. The data set has a large number of continuous columns, along with 38% of discrete variables present in the data. The data consist of very less amount of null values which are more elaborated on in the next section.

3.2 Missing Values

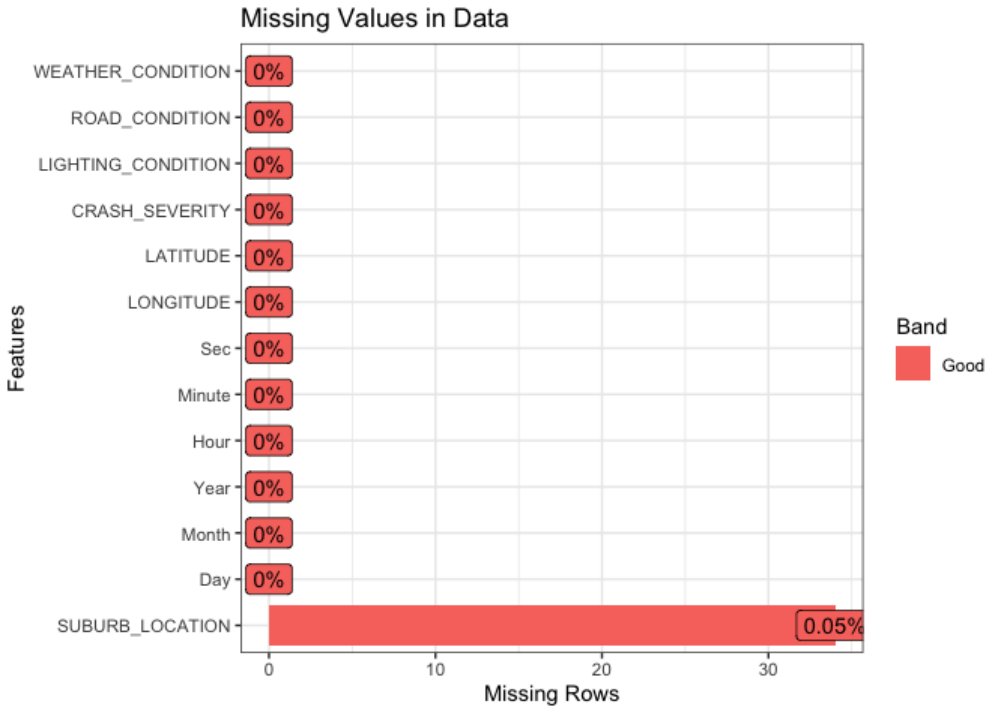

Figure 2

From figure 2 we can notice that the data is all filled with values except it consists of only one column with missing details. The column Suburb_Location has a few missing values which come up to 0.05% of missing data.

3.3 Duplicate Values

The original data had a duplicate column named **Location** which implied the spatial values for the suburbs in the ACT. The data is now cleaned and don't have any duplicate values as such.

4 . Data Exploration

4.1. An examination of road crash severity contributing to car accidents in the ACT.

4.1.1. Accidents occurred from 2012 to 2021.

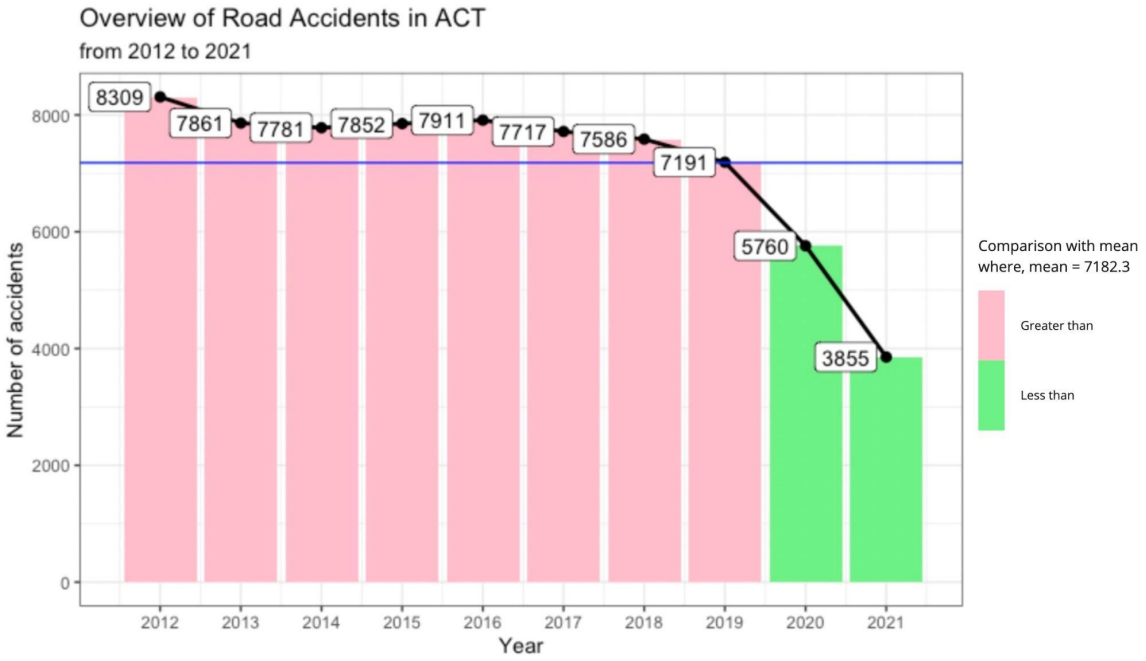

Figure 3

Figure 3 shows a downward trend in the number of car accidents in the past decade. During the earlier years, a minute increase and decrease in accident numbers are observed. The mean for the number of accidents comes to **7182.3** which is indicated by a blue line in Figure 3. In recent years, the green lines indicate a decline in numbers. This is due to the pandemic in 2020 which resulted in lockdowns and less traffic itself. The dataset consists of incomplete data for the year 2021 which resulted in a very low volume of road accidents for the same year.

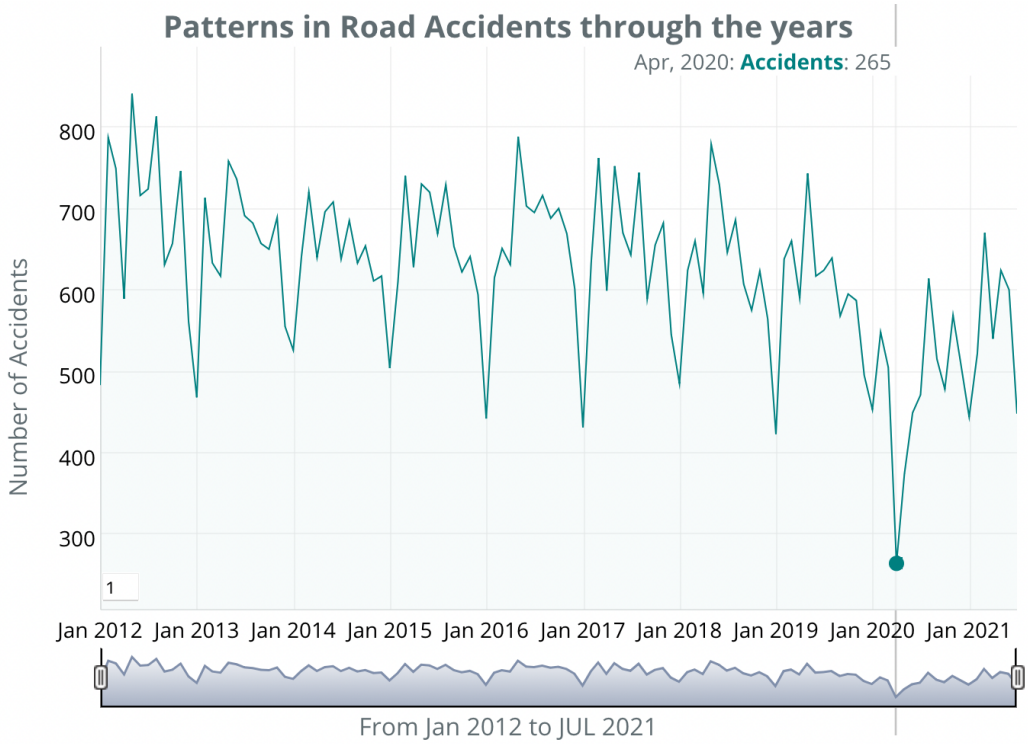

Figure 4

Figure 4 is a series graph that shows the trend of peaks during May or the rainy season. There is also a series of troughs that can be observed every January because that is around the holiday season. There is a sudden drop in accident numbers during Jan 2020, due to the pandemic. This could be because of the lockdowns and a general drop in traffic on the road during those times. But in 2021, as the rules for the lockdown have relaxed and traffic increased on the roads, the number of accidents started to rise again.

4.1.2. Accidents Severity

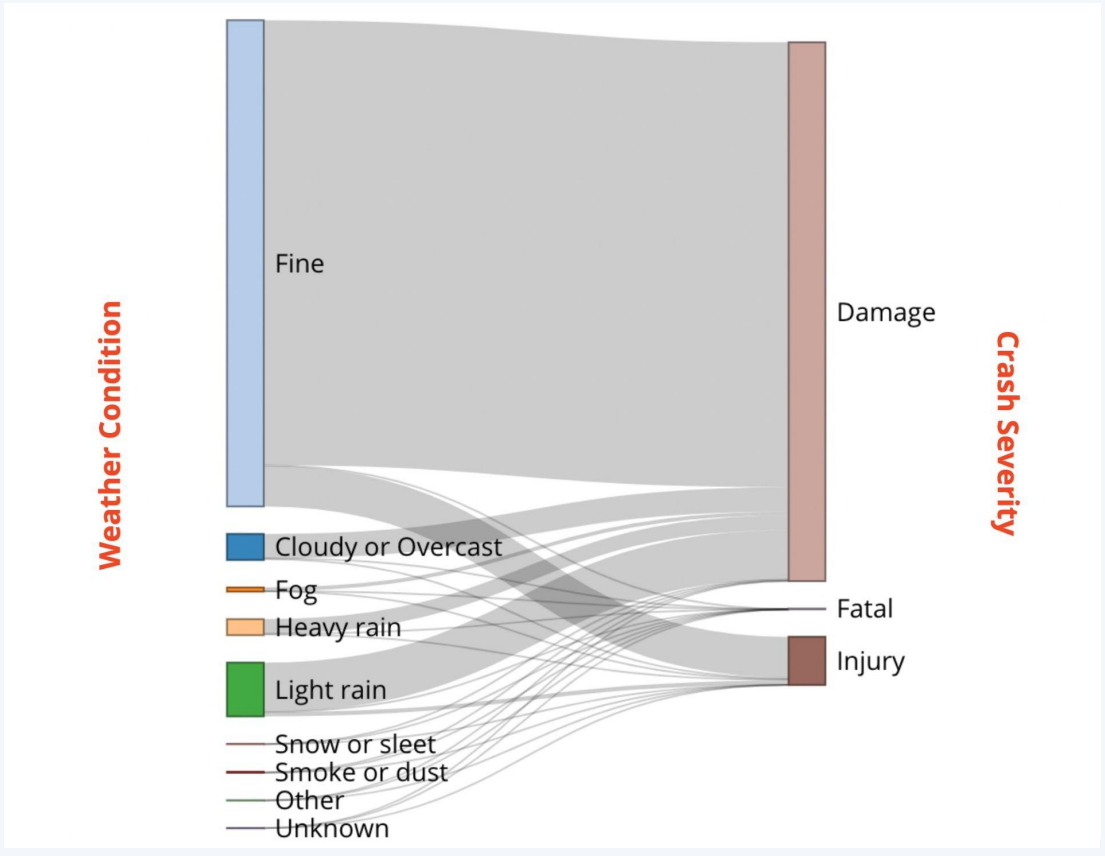

Figure 5

The above graph (Figure5) explores the weather conditions and the relative consequences of road accidents. Fine weather seems to be the widest parameter and during this time, it resulted mostly in property damages. A large part of property Damages also occurred during rainy weather. Injuries mostly occurred during fine weather barring the few that occurred during rainy or snowy weather. But fatalities mostly occurred due to rainy weather light or heavy. From the previous graph, it was evident that the accident trends indicate that rainy weather had the highest numbers.

4.2. Identifying the most accident-prone suburbs/areas, as well as the factors that contribute to the collisions.

4.2.1. Accident-prone Suburbs

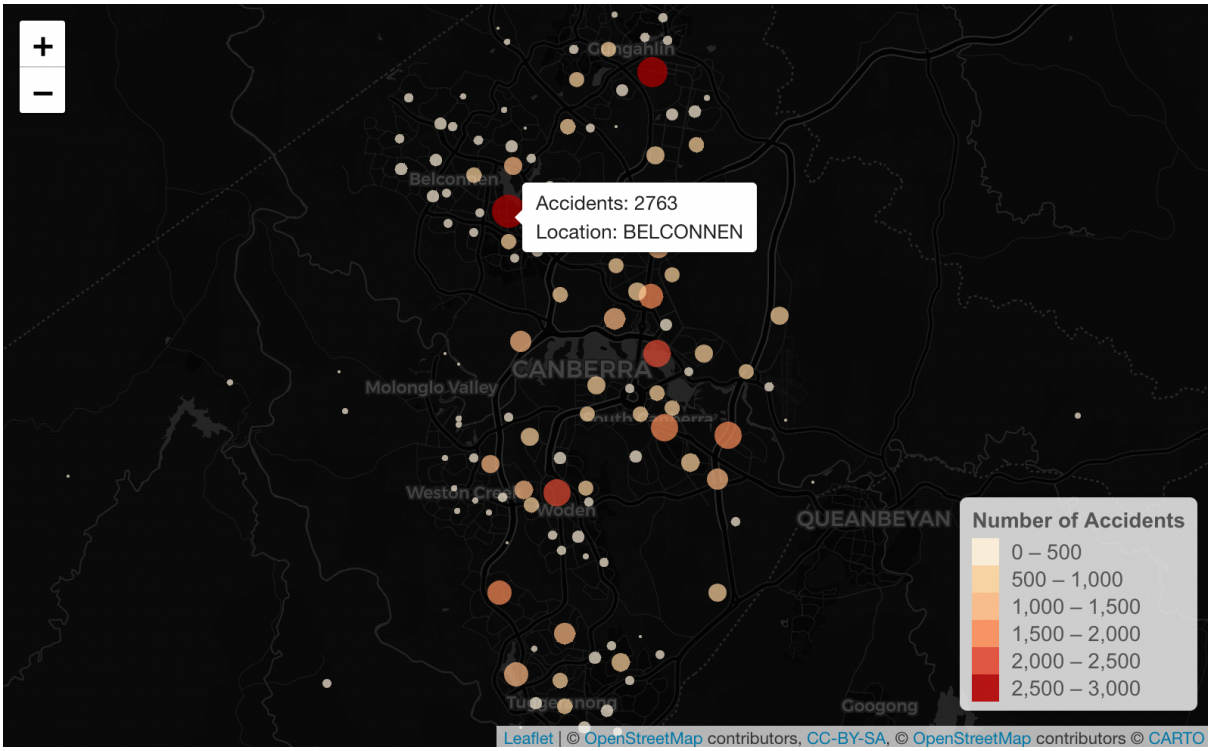

Figure 6

The figure above (Figure 6) depicts the accident-prone suburbs, with the radius representing the total number of accidents in that area. The most accidents occurred in BELCONNEN (2763), followed by GUNGAHLI (2728). The least number of accidents are seen mostly in the outskirts of the ACT. The lowest number of accidents took place in suburb Kenny with only a count of 2.

4.2.2. Factors that contribute to collisions

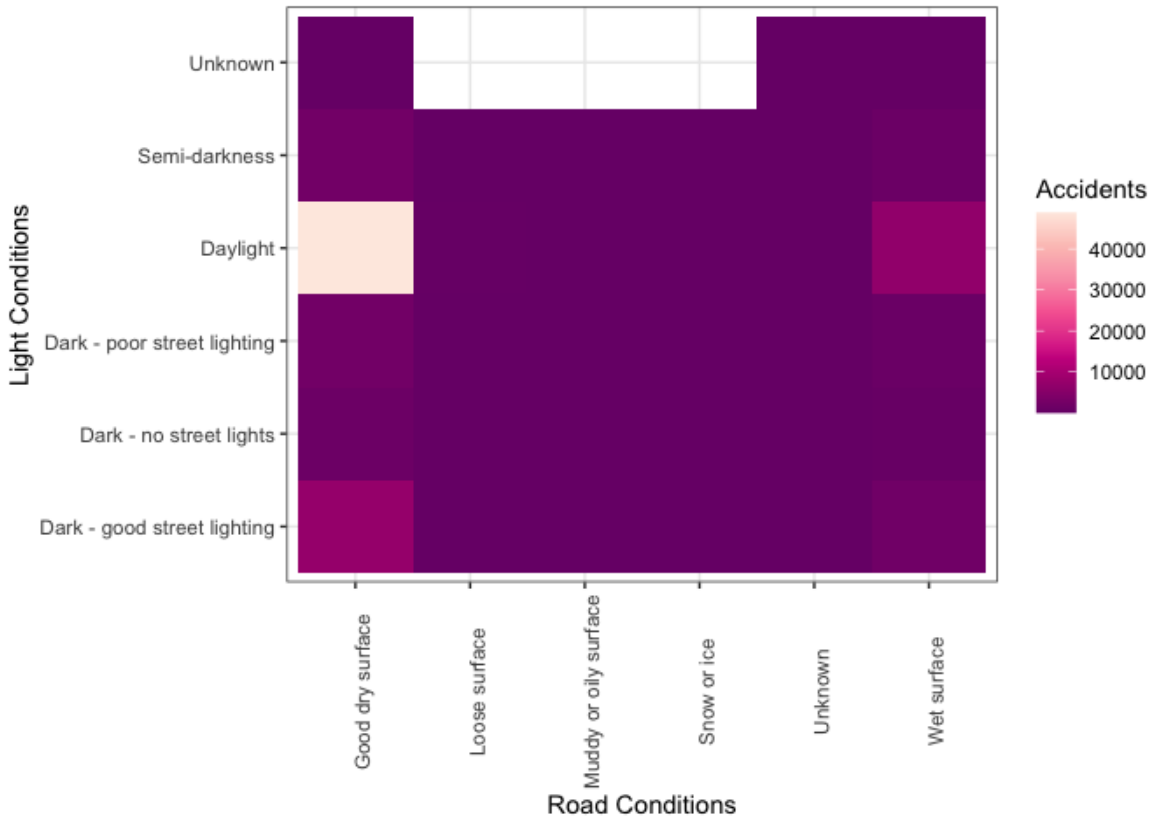

Figure 7

Figure7 indicates that most accidents occur in broad daylight and on a good dry road surface. Dim lights on the roads also contribute to car crashes but comparatively in a less number which means that the external lighting conditions do not play a major role in the number of accidents. The figure also shows the wet surface of roads also is one of the key aspects of road accidents during daylight.

4.3. What time of the day affects the frequency of accidents in the ACT?

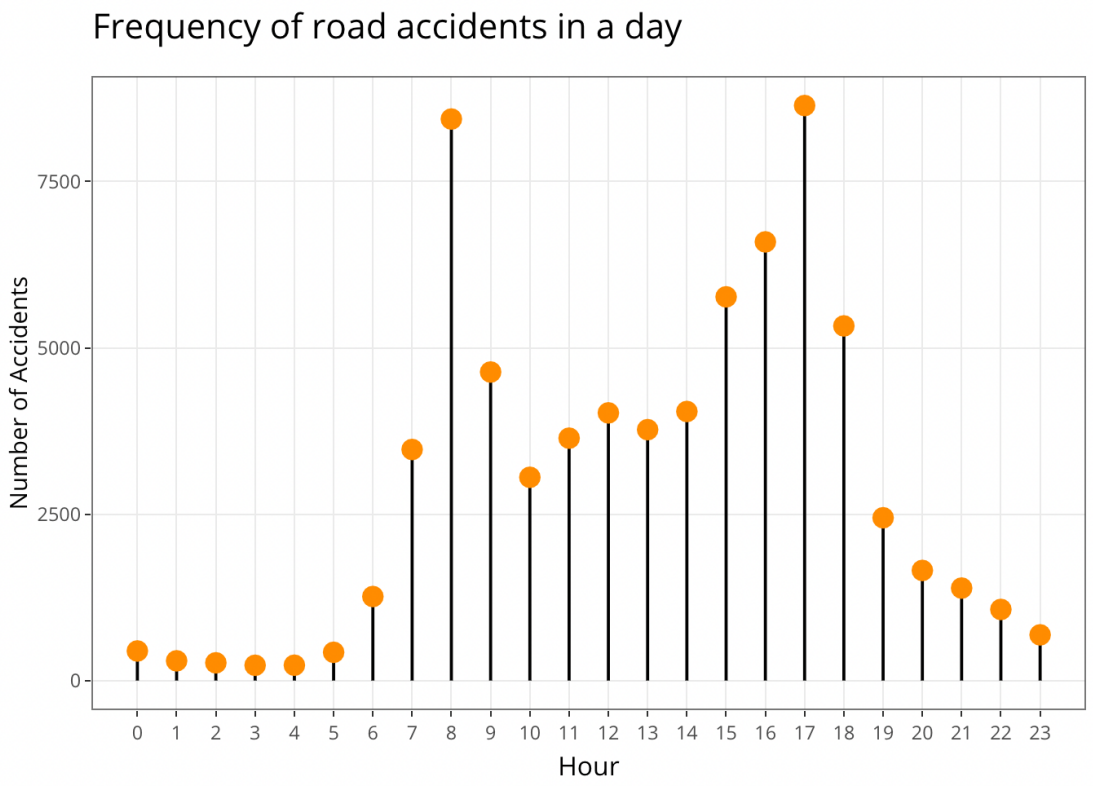

Figure 8

Figure 8, indicates that road accidents in a day are higher during office hours. Peaking at 8:00 AM, dropping a little during the rest of the day and peaking again at 5:00 PM and declining. Accidents in the ACT area are the lowest between 12:00 AM and 5:00 AM.

5. Conclusion

Over the years there were not many noteworthy differences in the number of accidents until the pandemic hit in 2020. Contrary to the belief that night driving leads to road accidents, it was observed that most accidents happen during the day and on good surface roads as none of the extreme weather conditions plays a significant influence on Canberra road crashes even when the weather is fine during the day. But accidents do occur mostly during the rainy season which is in the month of May.

During the day, accidents are high in number which is the standard office hours and more cars can be seen on the roads. Moreover, high severe consequence occurs in the suburbs like —---

6. Reference

1. Road Safety. Retrieved 7 August 2021, https://www.infrastructure.gov.au/roads/safety/
2. Connelly, & Supangan. (2016, November 6). Science Science Direct. Retrieved March 20, 2022,https://www.sciencedirect.com/science/article/abs/pii/S0001457506000649
3. ACT Road Crash Data. (2021). Retrieved 10 August 2021, fromhttps://www.data.act.gov.au/Transport/ACT-Road-Crash-Data/6jn4-m8rx
4. Car Accident Statistics 2020 | Car Research & Statistics — Budget DirectTM. (2020). Retrieved 7 August 2021, from https://www.budgetdirect.com.au/car-insurance/research/car-accident-statistics.html
5. plot_intro function - RDocumentation. (2022). Retrieved 12 April 2022, from https://www.rdocumentation.org/packages/DataExplorer/versions/0.8.2/topics/plot_intro
6. What Is Data Wrangling & Why Is It Necessary? (2021, May 14). MonkeyLearn Blog. https://monkeylearn.com/blog/data-wrangling/
7. GGPLOT Point Shapes Best Tips - Datanovia. (2022). Retrieved 14 April 2022, from https://www.datanovia.com/en/blog/ggplot-point-shapes-best-tips/
8. Holtz, Y. (2022). Interactive area chart with R and plotly. Retrieved 14 April 2022, from https://r-graph-gallery.com/163-interactive-area-chart-plotly.html