Analysis of Motor Vehicle Collision in Toronto*

Youngho Kim

Invalid Date

Car Accidents are fatal yet common. This paper examines Motor Vehicle Collision (MVC) occurrences and their changes from January, 2014 to June, 2022 in Toronto.

1. Introduction

Motor vehicle is one of the essential asset for modern day people. Its capacity and speed allowed faster shipment and displacement in larger scale than anytime before. However, it also created public safety issue in every alley of the city. In this paper, I explore the record of car accidents in the Toronto to learn about some important characteristics of car accidents happening in the Toronto for finding how impactful the car accidents are in our society and how to avoid it.

In this paper, I used R(R Core Team 2022) to analyze data on Motor Vehicle Collision (MVC) from opendatatoronto(Gelfand 2022) using dplyr(Wickham et al. 2022) and tidyverse(Wickham et al. 2019) packages. The purpose of this paper is to get better understanding on how tendency of car accident has changed over a time. The analyzed data contains changes of number of accidents occured within the period of data collection (Figure 2) and what time it has occured(Figure 3). It also looks at the level of the seriousness of the accidents (Figure 1) to evaluate how much impact the car accidents have in our daily lives.

2. Data

2.1. Overview

To begin the research on the car accidnets in the Toronto and the changes over the years, I utilized the Police Annual Statistical Report on Traffic Collisions data (Services 2022) from the

*Code and data are available at: https://github.com/YoungKim164/traffic_collision_analysis

Toronto Open Data portal (Gelfand 2022). The Police Annual Statistical Report is collected and published by the Toronto Police Services and was published on October 11, 2022. The raw data includes 553780 motor vehicel collisions in the Greater Toronto Area, Time and date of the accidents, and the result of the accidents into three categories:Property Damage, Injury, and Fail to Remain. `R` (R Core Team 2022), `tidyverse` (Wickham et al. 2019), `dplyr` (Wickham et al. 2022), `janitor`(Firke 2021), `data.table` (Dowle and Srinivasan 2022), `zoo` (Zeileis and Grothendieck 2005), `chron` (James and Hornik 2022), and `lubridate`(Grolemund and Wickham 2011) I cleaned and extracted the necessary data from the raw data to start anlaysis.

2.2. Analysis

Figure 1: percentage of each categories of accidents

First, I looked at the three types of the result the Toronto Police Services used to categorize the car accidents. I counted individual categories and grouped them together to visually compare what types of accidents occurs in most frequent occasion. For the intuitive visual comparison, I used `tidyverse`(Wickham et al. 2019) to create bar graph comparing each count(Figure 1). Since its purpose is not focused on finding exact number of occurences but to learn the frequancies of each category, the data is represented as percentage instead of actual count.

Figure 1 tells that property damage is close to 3/4 of the total accidents and injury is low-

est among other categories. It indicates the car accidents doesn't always lead to the life threatening situation in Greater Toronto area. Indeed, according to transportation canada on Canadian Motor Vehicle Traffic Collision Statistics: 2020(Canada (2022)) the fatality rate of motovehicle collision has been bellow one percent since 2011 and steadly decreasing over the year. The major reason behind it is the advance of safety technologies on the vehicle. Technologies such as Forward Collision Warning, Lane Departure Warning, Automatic Emergency Braking, and Rearview Video System became common feature in customer vehicles, and more advanced safety systems like vehicle-to-vehicle communication or automated driving are actively researched to ensure roadside safety. Another key reasons of low injury causing accident is the nature of road environment in the Greater Toronto area. Most of the road in the Greater Toronto area has below 80km per hour. With the low speed limit on the side, it is highly unlikely the drivers be able to speed over 80 km per hour since most of the roads are short and narrow urban roads with significant traffics most of the time. A study from IIHS and AAA showed that vehicles crashed at 40mph(64.4kmh) only showed minor intrusions in the cockpit, while vehicles crashed at 50mph(80.5kmh) were resulted noticeable deformation (Szymkowski (2021)). In summary, most accidents in the Greater Toronto area are tends to end up property damages only without life loss due to the modern safety systems installed in the vehicles and the nature of the Toronto roads that prevents the driver to critically speed up.

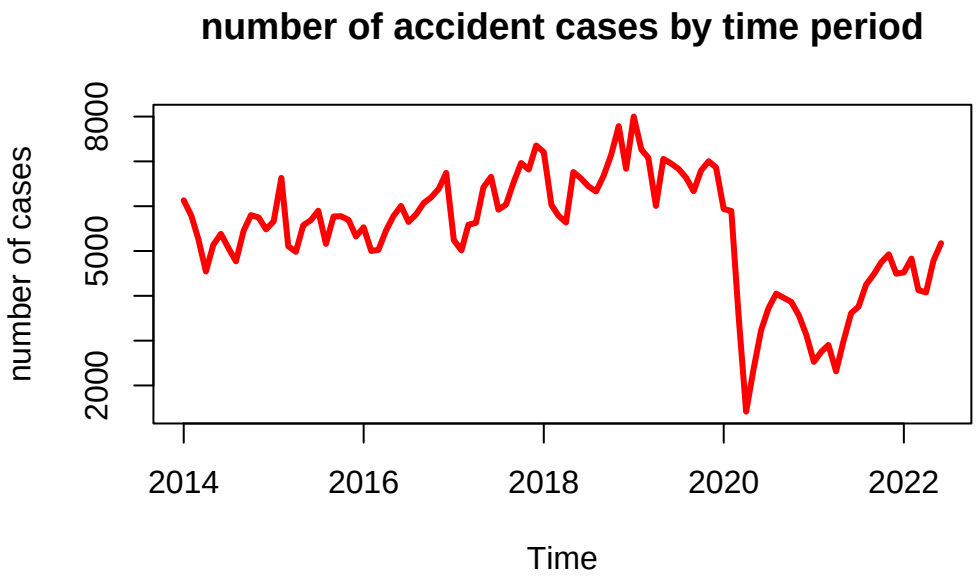

Figure 2: number of accidents overtime

To understand the changes of the trends of accidents occurrence, I reorganized the accidents by the month and year when it was happend. R package `data.table`(Dowle and Srinivasan

2022) and `lubridate`(Grolemund and Wickham 2011) were used for the regrouping data in to chronological order by year and month. The reorganized data is ploted into a line graph to visualize the change overtime (Figure 2).

The most significant sighting here is the large drops on the number of accidents in the begging of 2020. It is the time when the City of Toronto started serious lockdown after the spread of COVID-19. The analysis on traffic during the pandemic by Kouchakzadeh, Mostafa and Bayanouni showed that there was a significant drop on the traffic congestion in GTHA(Greater Toronto hamilton Area) after the COVID-19 lockdown (Kouchakzadeh, Bayanouni, and Roorda (2021)). It was hypothesized that number of accident was decreased as the government lockdown impacted the traffic congestion on the road. As the lockdown was eased, the increase on the number of car accident can be found at the end of 2020.

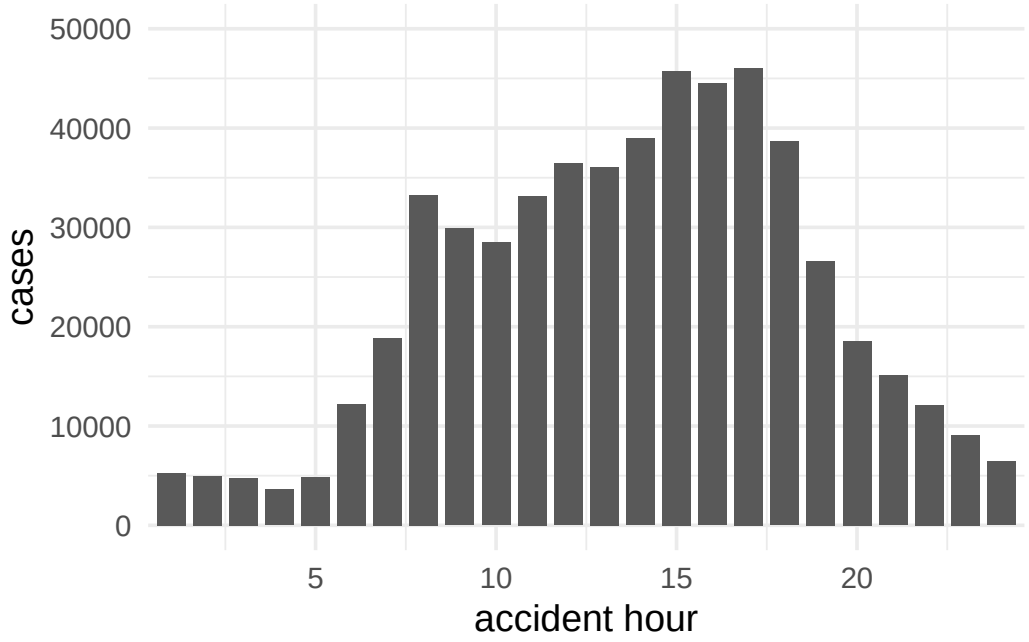

Figure 3: number of accident by time

To have the deeper understanding on when exactly the car accidents happens. I analyzed the accidents by the time it was occured. using 24 hour system, the accidents were grouped by the hour it was happend using the `hour` variables on the raw data. the cleaned data is represented as the bar graph to have simple visual representation of when the accident happened (Figure 3).

The most significant pick is from 1500 to 1700. It was easy to hypothesize that the reason of this pick was because of the rush hour. However, it is noticiable that the number of accidents from 0700 to 0900, which commonly considered as rush hour as well, are lower than or similar to the number of accidents during the regular day hour. Many factors would have effected this

result but it is suspected that the most important factor is the geography of the Toronto. The Toronto has more complex and dense road plan as it gets to the center of the city. However, outer bound has much simpler and wider roads. This makes getting into the city much harder than getting out from the city. As the result the traffic congestion must be worse when people move out from the city than coming in, which cause more accidents.

3. Conclusion

The data showed that the critical car accident which cause serious impact on our lives are very unlike to happen especially in the urban area like the Greater Toronto area where vehicles move relatively slowly on most of the roads. It seems the amount of traffic has the most roads are even safer after the COVID-19 but the accidents count is rising back to the previous level. To minimize the risk of accident, the best practice is to avoid traffic. Most accidents were happened During the rush hour.

Even though the traffic is the most important factor, it is hard to say that it is the only important factor. Further research on other data sources like driving habit, damage level of the accidents, weather and other environmental effect, and/or road condition could show us more deeper understanding and different perspectives on this case.

4. References

Canada, Transport. 2022. "Canadian Motor Vehicle Traffic Collision Statistics: 2020." *Transport Canada. / Gouvernement du Canada*. https://tc.canada.ca/en/road-transportation/statistics-data/canadian-motor-vehicle-traffic-collision-statistics-2020.

Dowle, Matt, and Arun Srinivasan. 2022. *Data.table: Extension of 'Data.frame'*. https://CRAN.R-project.org/package=data.table.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

James, David, and Kurt Hornik. 2022. *Chron: Chronological Objects Which Can Handle Dates and Times*. https://CRAN.R-project.org/package=chron.

Kouchakzadeh, Mostafa, Hasan Bayanouni, and Matthew J. Roorda. 2021. *Analyzing Impact of the COVID-19 Pandemic on GTHA Traffic Congestion Using Travel Speed Data*, November. https://doi.org/https://smartfreightcentre.ca/wp-content/uploads/2021/11/Kouchakzadeh_Bayanouni_Roorda_Covid19_GTHA_Traffic_Final.pdf.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Services, Toronto Police. 2022. "Police Annual Statistical Report - Traffic Collisions," October. https://open.toronto.ca/dataset/police-annual-statistical-report-traffic-collisions/.

Szymkowski, Sean. 2021. "Modern Cars Are Safe, but Higher-Speed Crashes Reveal Serious Weaknesses." *CNET*. CNET. https://www.cnet.com/roadshow/news/cars-safe-speed-crashes-iihs-aaa-study/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Zeileis, Achim, and Gabor Grothendieck. 2005. "Zoo: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software* 14 (6): 1–27. https://doi.org/10.18637/jss.v014.i06.