# Toronto Neighbourhood Crime Rate Overview*

Huakun Shen

07 February 2022

**Abstract**

The crime rates in different regions of Toronto ranges largely. The crime rate has an increasing trend from 2014 to 2019, and a sudden drop in 2020 probably due to covid 19.

## 1 Introduction

Crime Rate is one of the biggest concerns about peoples live and safety. OpenDataToronto(Gelfand 2020) contains many public available datasets containing data about Toronto. One of the datasets called "Neighbourhood Crime Rates" contains the number of different types of crimes happened in Toronto from 2014 to 2020. The crime rates (out of 100,000) of each type of crime are also reported.

In this paper, I did some basic analysis on the topic of Toronto's crime rate in recent years using (R Core Team 2020). The crime rate in Toronto has a increasing trend starting 2014, and had a sudden drop in 2020. The sudden drop can probably be explained by Covid 19. The hypothesis will be explained in Section 4.

Some other R packages were also used. (Firke 2021) was used to clean up data frame column names. (Wickham et al. 2019), (Wickham and Girlich 2022), (Wickham et al. 2021) were used for data processing. (Kassambara 2020) was used to organize plots (Xie 2021), (Xie 2015), (Xie 2014) were used to generate tables. (Teetor 2011) was used for R syntax.

Section 2 explains where the dataset is taken from, what limitations it has, and how I pre-processed the dataset. Section 3 shows the results (tables and plots) used for analysis. Section 4 contains my understanding to the dataset, analysis, and explains some findings I derived from the data. Furthermore, 4.3 contains some thoughts about what analysis can be done in the future on this topic.

## 2 Data

The dataset was downloaded from OpenDataToronto(Gelfand 2020), "Neighbourhood Crime Rates" dataset. This dataset includes the Crime Data by Neighbourhood, from 2014 to 2020, including 140 neighbourhoods.

No data is 100% accurate and precise. Assuming that there is no human factor affecting data collection, the data collected in this dataset may still be inaccurate due to various reasons. For example, if a race or a group of people doesn't like to deal with police, the region where they are a majority might have a lower crime reporting rate, and the crime rate if that region won't reflect the actual crime rate well. For example, a region full of gangster may have very high crime rate but lower-than-the-actual reporting rate.

The types of crimes include Assault, Auto Theft, Break and Enter, Robbery, Theft Over, Homicide and Shooting & Firearm Discharges. Crime rate per 100,000 is also calculated based on the population estimates provided by Environic Analytics.

---

*Code and data are available at: https://github.com/HuakunShen/Toronto-Crime-Rate-Analysis

A limitation of this dataset is its bad design. There is no "year" column, instead, the year is included in column names. For example, Crime type "Robbery" has the following related columns

- Robbery_2014
- Robbery_2015
- Robbery_2016
- Robbery_2017
- Robbery_2018
- Robbery_2019
- Robbery_2020
- Robbery_Rate2014

- RobberyRate_2015
- Robbery_Rate2016
- Robbery_Rate2017
- Robbery_Rate2018
- Robbery_Rate2019
- Robbery_Rate2020

There are 104 columns in total while there are only 7 crime types. It's very hard to do analysis on years with this dataset structure, for example, I can't do a crime rate by year on this.

Table 1 is a subset preview of the original dataset. It's impossible to show the entire table as there are too many columns.

Table 1: Preview of a subset of the original dataset

| Neighbourhood | Assault_2014 | Assault_2015 | geometry |
|---|---|---|---|
| Yonge-St.Clair | 16 | 25 | POLYGON ((-79.39115 43.6810… |
| York University Heights | 273 | 298 | POLYGON ((-79.50525 43.7598… |
| Lansing-Westgate | 42 | 81 | POLYGON ((-79.43994 43.7615… |
| Yorkdale-Glen Park | 106 | 137 | POLYGON ((-79.43965 43.7056… |
| Stonegate-Queensway | 91 | 74 | POLYGON ((-79.49258 43.6474… |
| Tam O'Shanter-Sullivan | 103 | 88 | POLYGON ((-79.31975 43.7683… |

To solve this problem, I pre-processed the data to make it have a "year" column. The resulting dataset has the structure in table 2.

Table 2: Cleaned Table (Subset)

| Neighbourhood | year | Shootings_Rate | Robbery_Rate | Homicide_Rate |
|---|---|---|---|---|
| Yonge-St.Clair | 2014 | 0.000000 | 31.753590 | 0 |
| Yonge-St.Clair | 2015 | 0.000000 | 0.000000 | 0 |
| Yonge-St.Clair | 2016 | 7.722008 | 46.332050 | 0 |
| Yonge-St.Clair | 2017 | 0.000000 | 7.567159 | 0 |
| Yonge-St.Clair | 2018 | 0.000000 | 7.403020 | 0 |
| Yonge-St.Clair | 2019 | 0.000000 | 29.006530 | 0 |

# 3 Results

## 3.1 Mean Crime Rate

### 3.1.1 Mean Crime Rate By Neighbourhood

The mean crime rate is calculated for every neighbourhood by grouping-by the "Neighbourhood" column. Here is a preview of the resulting dataset.

Table 3: Average Crime Rate By Neighbourhood (Subset)

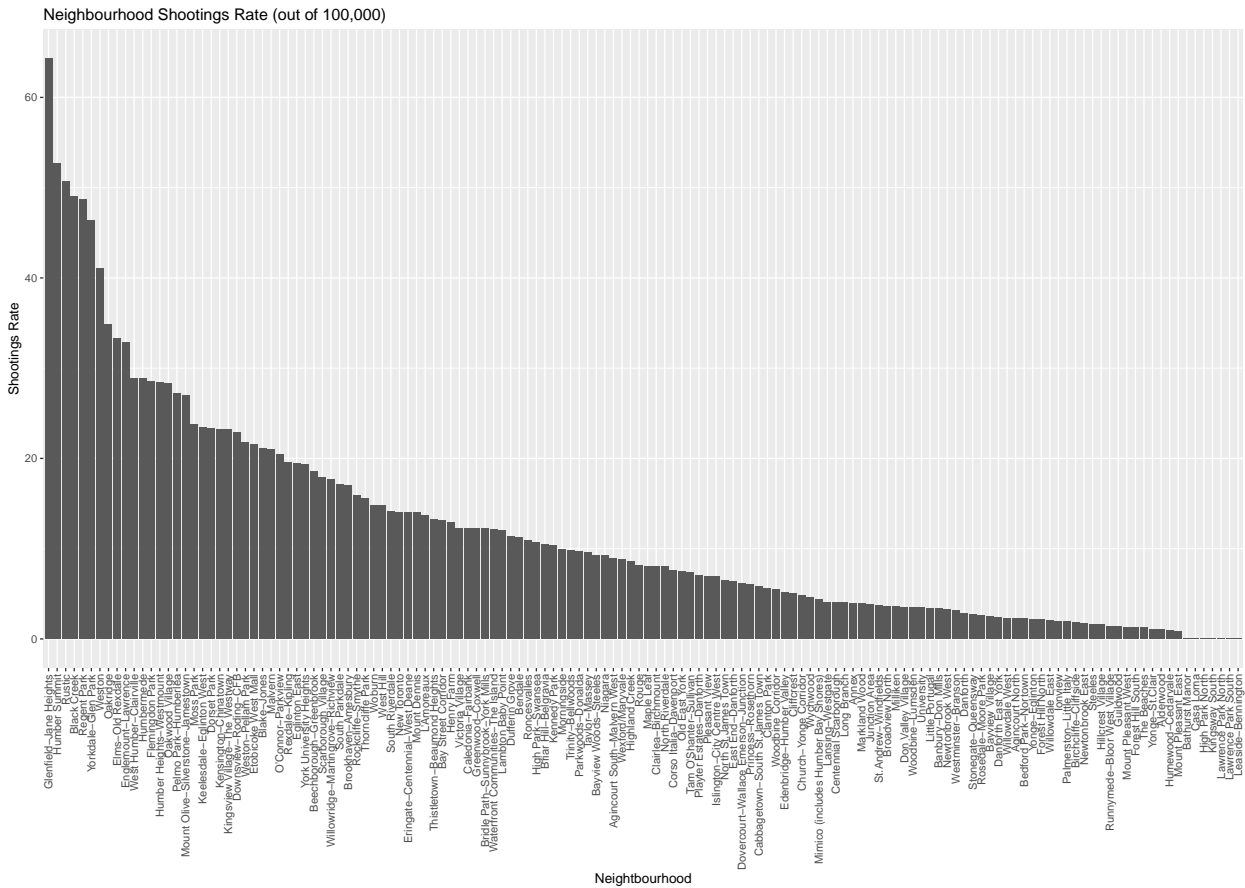| Neighbourhood | Shootings_Rate | Robbery_Rate | Homicide_Rate |
|---|---|---|---|
| Agincourt North | 2.289578 | 77.05904 | 0.0000000 |
| Agincourt South-Malvern West | 8.949272 | 88.98122 | 1.0425246 |
| Alderwood | 1.094607 | 54.84546 | 1.1550545 |
| Annex | 3.977597 | 128.37054 | 0.8518821 |
| Banbury-Don Mills | 3.349794 | 32.21475 | 0.0000000 |
| Bathurst Manor | 0.000000 | 59.30612 | 0.0000000 |



Figure 1: Shootings Rate By Neighbourhood

### 3.1.2 Mean Crime Rate By Year

The mean crime rate is calculated for every year by grouping-by the"year" column.

Here is a preview of the resulting dataset.

Table 4: Average crime Rate By Year (Subset)

| year | Shootings_Rate | Robbery_Rate | Homicide_Rate |
|------|---------------|--------------|---------------|
| 2014 | 0.00000 | 132.7656 | 2.097511 |
| 2015 | 10.75144 | 0.0000 | 2.086116 |
| 2016 | 14.48919 | 128.0834 | 2.601719 |
| 2017 | 13.45811 | 134.1897 | 2.368686 |
| 2018 | 14.21444 | 121.1290 | 2.821659 |
| 2019 | 16.78657 | 116.0732 | 2.775645 |

Figure 2 shows the crime rates of different types from 2014 to 2020.

## 4    Discussion

### 4.1    Crime Rate in different Neighbourhoods

Figure 1 shows the Shootings Rate for different neighbourhoods. The crime rates range greatly across different regions. The region with the highest crime rate has approximately 65 shootings per 100,000 people. The region with the least shootings has close to 0 shooting.

If a person were pick a residence, it's better for them to pick a region with lower crime rate, especially shootings rate.

### 4.2    Mean Crime Rate By Year

Figure 2 shows the crime rate of different types by year. We can observe that, all crime types except for robbery has an increasing rate. Meaning that The overall crime rate has been increasing in Toronto in recent years.

An interesting observation is that all crime types except for Robbery and Auto Theft have a increasing trend if crime rate but a sudden drop in 2020. Given that the data is correct, the sudden drop is a surprise.

Covid 19 is undoubtedly one of the most unexpected variable in year 2020, and it can be used to explain the sudden drop of crime rates in 2020.

The sudden drop of crime rate of Assault, Break-and-Enter, Theft Over, Homicide, Shootings can be explained by the stay-at-home order caused by the pandemic.

Although many people ignored the order, the population going out is still reduced by work-from-home, prohibition to dine-in. As people tend to stay home for longer time, crimes like assualt, theft over, break-and-enter, homicide and shootings which usually happen outside of home tend to occur less frequently.

The exception is Auto Theft crime, the only crime rate increased in 2020. It can also be explained by the same reason. As people stay home for longer, their vehicles are also left alone for longer, giving the thefts more chances to steal a vehicle.

Robbery is the only crime that has a constant dropping rate. 2020 has a exceptionally large drop compared to previous years, which could be explained by the same reason.
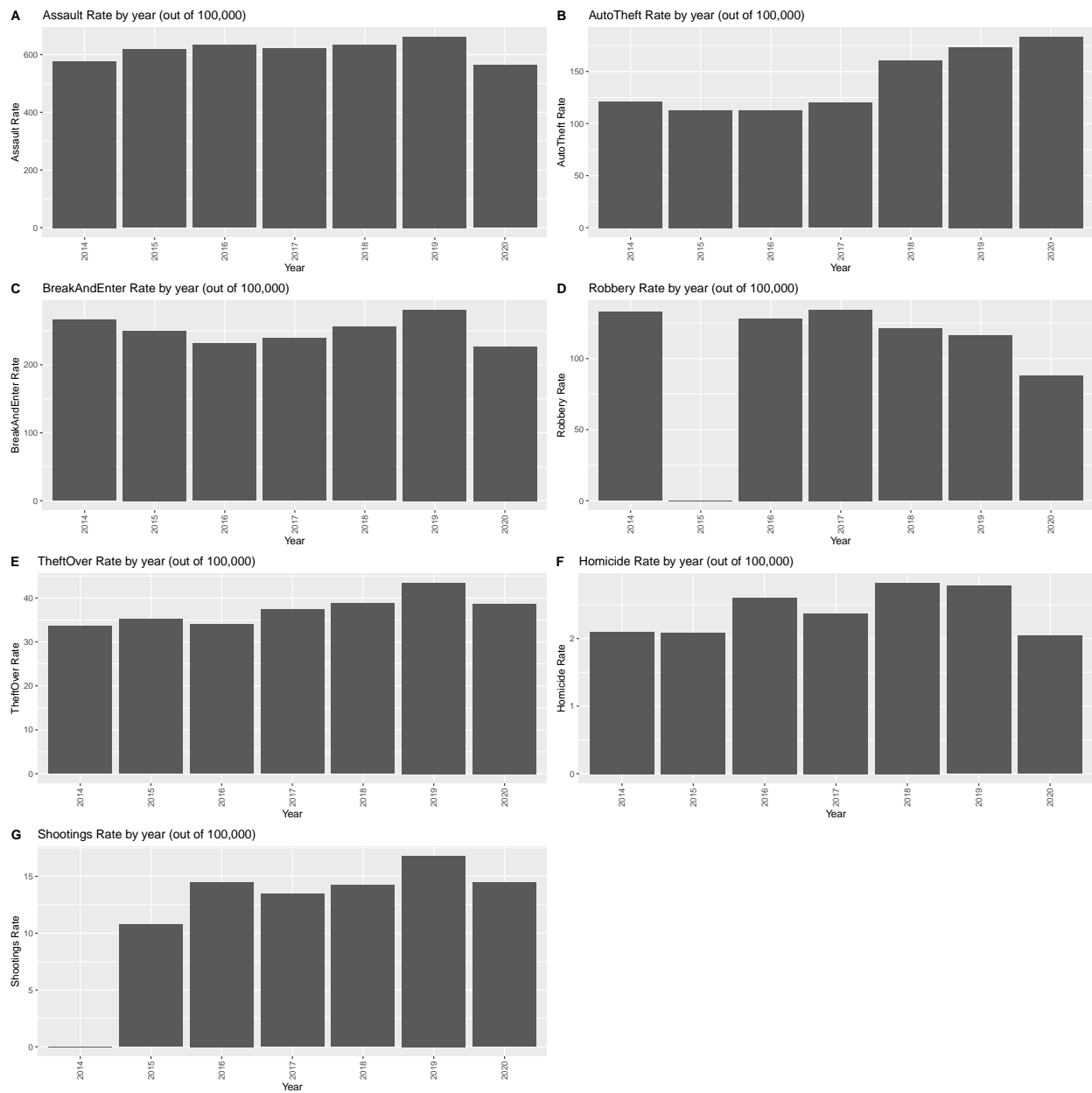
Figure 2: Crime Rate By Year

## 4.3   Weaknesses and next steps

### 4.3.1   Neighbourhood Safety

I only generated the shootings rate histogram for neighbourhoods (which is the most critical type of crime). To better understand the safety in different neighbourhoods, I will need to generate a histogram for each crime type. As there are 140 neighbourhoods, plotting a histogram isn't intuitive enough to visualize the crime rate in different regions. The next step could be plotting a heat map, which is a more intuitive visualization.

It's also interesting that the crime rate ranges largely in different regions. The next step could be finding out what are the causes of the high crime rate in some regions. For example, housing price, rental price, race of residents, education level of residents, population density, income level of residents, etc. The dataset used in this paper is not enough, more data will be required to do such analysis.

### 4.3.2   Cause of Crime Rate Drop In 2020

Based on data in the dataset used in this paper, it's not rigorous enough to conclude that Covid 19 reduces crime rate. The crime rates were reported annually, it should be more accurate if we can use monthly reported crime data and covid counts.

# References

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

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal.* https://CRAN.R-project.org/package=opendatatoronto.

Kassambara, Alboukadel. 2020. *Ggpubr: 'Ggplot2' Based Publication Ready Plots.* https://CRAN.R-project.org/package=ggpubr.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Teetor, Paul. 2011. *R Cookbook.* Beijing: O'Reilly. http://www.amazon.com/Cookbook-OReilly-Cookbooks-Paul-Teetor/dp/0596809158.

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. 2021. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Maximilian Girlich. 2022. *Tidyr: Tidy Messy Data.* https://CRAN.R-project.org/package=tidyr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

———. 2015. *Dynamic Documents with R and Knitr.* 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. https://yihui.org/knitr/.

———. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.