Crime Rates in Toronto Neighborhoods*

Meha Grewal

06 February 2022

Abstract

Neighbourhood crime rates in the city of Toronto is essential data to understand illegal activity in the city. We cleaned and analyzed crime statistics in 2020 based on different crimes using graphs and tables. In our findings we concluded that the most prominent crime occurring in Toronto has been assault rates, and the number of assaults has gradually increased over the period of 2014-2020. This data can be utilized to understand what areas need to be focused on by the city and law enforcement agencies, this information can be influential in understanding what neighbourhoods to distribute social services and where to focus on the prevention of crime. (why it matters).

1 Introduction

Crime is an ever present part of society that has an influence on every individual, from cautious walks in the outdoors, to security systems or avoiding certain areas due to reputation. Neighborhoods in Toronto can vary dramatically in terms of crime rates and what crimes are most prominent. Crime counts and rates are an essential metric to be calculated as it allows for comparison between different infractions and how neighborhoods diverge from one another. The first step in reducing crime and ensuring safety is understanding the crime statistics of the region.

The city of Toronto is divided into many neighborhoods and areas that vary in population and experience the fluctuations of different crimes. We are interested in the most notable felony, assaults, which has on average been committed at a higher rate than other crimes recorded by the Toronto Police Service. Additionally, we were motivated by how this offense changed over the years recorded. Our data indicated that between the period of 2014 to 2020 the assault rate increased. However, we examined 10 of the neighborhoods with the highest projected population for 2020 and compared those same regions in 2014 and 2020, there were no considerable differences. But, specific neighborhoods such as Waterfront Communities - The Islands and the Church-Yonge Corridor experience a high number of assaults.

This information emphasizes what areas need to be focused on for the safety of Toronto residents. Moreover, it gives valuable knowledge to law enforcement agencies on how to tackle criminal offenses when they can specify where and what crimes are occurring. Most notable, count of assaults overall are higher than other transgressions. This indicates it needs to be an area of focus for governmental authorities. Furthermore, assaults have increased over recent years, rising population may be a factor. The data indicates the neighborhoods that endure high volumes of assault that should specifically be targeted by the Toronto Police service. These statistics will highlight ket issues in regards to public safety and assist in preventing crime.

---

*Code and data are available at: https://github.com/meha-g/Paper1.

2 Data

2.1 Data Source

Paragraph or two introducing data set broadly. This paper focuses on data about neighborhood crime rates in Toronto from Open Data Toronto https://open.toronto.ca/dataset/neighbourhood-crime-rates/. We obtained our data set using the 'opendatatoronto' package from the City of Toronto's Open Data Portal (Gelfand 2020) and the statistical programming language R (R Core Team 2020). This data set was provided by the Toronto Police Service for the Public Safety Data Portal: Open Data Portal. The data set was obtained in CSV format and was last updated on May 6, 2021.

2.2 Methodology

Our data is about neighborhood crime data from the years 2014 to 2020 as reported by the Toronto Police Service for 140 neighborhoods. This data set includes counts for criminal offenses including assault, auto theft, break and enter, robbery, theft over, homicide, and shootings. A population projection variable for 2020 provided by Environmental Analytics was included in the data, using this variable the crime rates were calculated per 100,000 population. The crime rates were determined by the StatsCan definition, thus allowing for comparison of crimes with different size populations.

We began by cleaning the data set using the 'janitor' package (Firke 2021), this allowed us to check for missing variables, dupes, and convert variable names to lowercase. Then we created a summary of all the crime rates within the data to understand the distributions and specify which crimes are most prominent. We used the 'skimr' package (Waring et al. 2021) to create the summary statistics (Table **??**).

```
##  assault_rate2020 auto_theft_rate2020 break_and_enter_rate2020
##  Min.   : 116.2   Min.   :  41.65    Min.   : 28.09
##  1st Qu.: 320.4   1st Qu.: 102.94    1st Qu.:122.14
##  Median : 484.4   Median : 151.58    Median :191.50
##  Mean   : 563.3   Mean   : 183.20    Mean   :227.14
##  3rd Qu.: 664.9   3rd Qu.: 226.58    3rd Qu.:285.58
##  Max.   :3162.5   Max.   :1066.44    Max.   :840.83
##  robbery_rate2020  theft_over_rate2020 homicide_rate2020 shootings_rate2020
##  Min.   :  7.133   Min.   :  0.00      Min.   : 0.000    Min.   :  0.000
##  1st Qu.: 50.666   1st Qu.: 20.40      1st Qu.: 0.000    1st Qu.:  0.000
##  Median : 76.242   Median : 31.83      Median : 0.000    Median :  7.678
##  Mean   : 87.842   Mean   : 38.65      Mean   : 2.041    Mean   : 14.427
##  3rd Qu.:112.382   3rd Qu.: 48.71      3rd Qu.: 3.414    3rd Qu.: 21.048
##  Max.   :543.819   Max.   :163.47      Max.   :14.972    Max.   :102.934
```

(Table **??**) showcases summaries for all the crime rate variables for the year 2020, this was done to analyze the most recent data. The table indicates that the overall highest crime rate belongs to assaults as seen in the mean of 'assault_rate2020'. Additionally the minimum and maximum assault rate is higher than all other crime rates provided in the data set. Thus we concluded that assaults in Toronto regions is the most common issue currently.

Next, we showcased an extract of the number of assaults for 10 neighborhoods, including the projected population for 2020, (Table 1).

We utilized the package 'knitr' (Xie 2021) to create (Table 1). It displays 10 neighborhoods and the assault rates from the years of 2017 to 2020 because we wanted to showcase the count of assaults in recent years and its trends. Overall there is no clear trend and between this time period the number of assaults in neighborhoods as it fluctuates negatively and positively. Particularly, neighborhoods such as Danforth East York, Thistletown-Beaumond Heights, and especially Yorkdale-Glen Park were able to drastically decrease

Table 1: First ten rows of a dataset of number of homicides between the years of 2017 to 2020.

Neighbourhood	Population	Assault 2017	Assault 2018	Assault 2019	Assault 2020
Yonge-St.Clair	14083	25	28	35	23
York University Heights	30277	351	362	382	341
Lansing-Westgate	18146	84	68	70	97
Yorkdale-Glen Park	17560	163	178	210	156
Stonegate-Queensway	27410	98	86	83	104
Tam O'Shanter-Sullivan	29970	120	131	108	131
The Beaches	23364	114	99	82	84
Thistletown-Beaumond Heights	10948	56	53	75	56
Thorncliffe Park	23518	98	94	119	94
Danforth East York	18427	78	79	85	56

their number of assaults from 2019 to 2020. This emphasizes which neighborhoods safety precautions could potentially be modeled after to reduce the crime rate in other regions of the city.

##Graphs

We are interested in the number of assaults from the years 2014 to 2020 since this showcases the current trends of dangerous crimes being committed and its implications for the safety of neighborhoods in Toronto. By creating histograms of assault rates in the years 2014 and 2020 and comparing them side by side, we could visually see if crime prevention had an overall impact during that time period. We utilized the package 'patchwork' (Pedersen 2020) to join the two graphs together.

Figure (Figure **??**) displays the pattern of homicides in Toronto between the years of 2014-2020
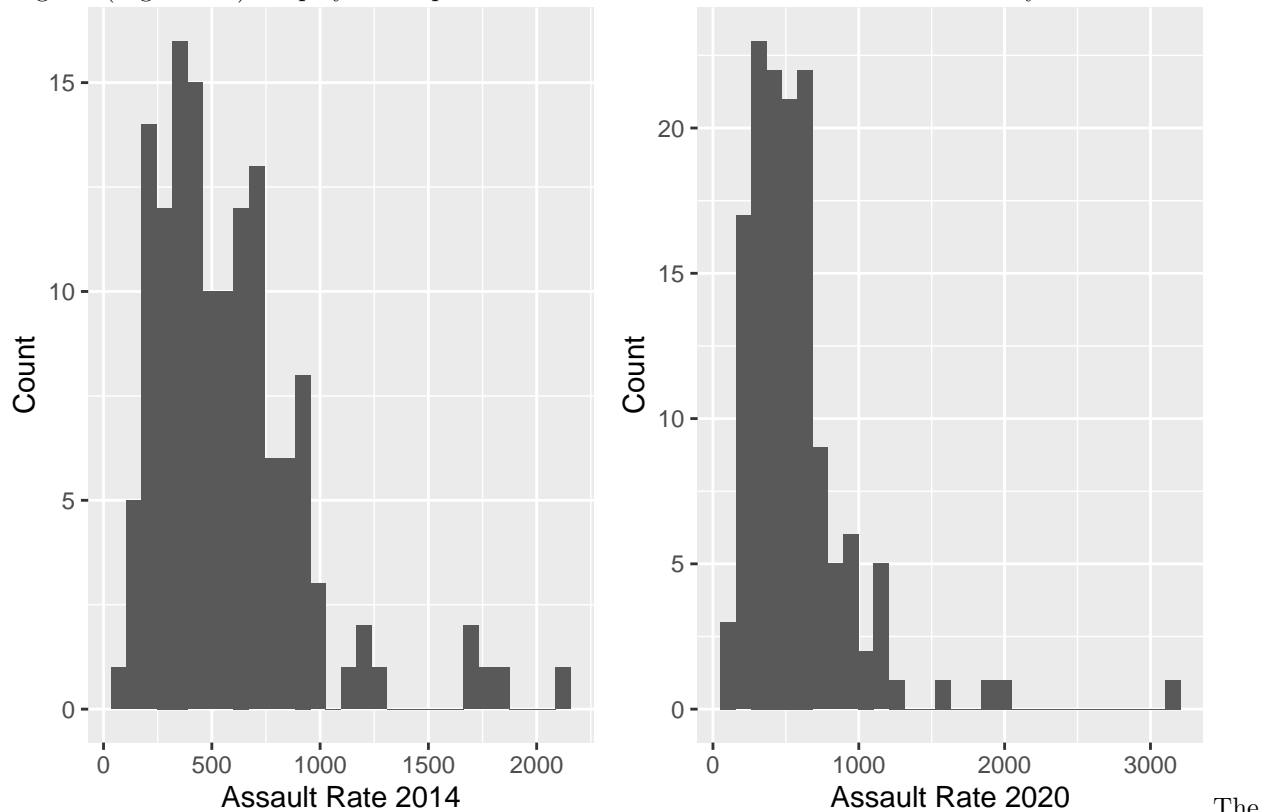
The distributions of assault rates (Figure **??**), in those two years have remained the same. As both have a heavy right skew with most data between the interval 0 and 1000 as the assault rate. However, the assault rate in

2020 has a higher concentration of values and notably larger value that have an assault rate of more than 3000. Meaning the maximum value in 2020 is much higher than the maximum value for 2014.

Furthermore, we (Figure **??**)

However there are limitations to this data set. The population variable only represents those that reside in the region and do not include temporary populations, like commuters. Additionally, not every crime is provided within this data set by the Toronto Police Service, thus we cannot know for sure whether assaults are truly the most common crime. Furthermore, this does not accurately describe crime rates in Toronto neighborhoods as there are still unreported crimes that occur. We do not have access to the method the Toronto Police Service used to acquire this data.

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.

Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots.* https://CRAN.R-project.org/package=patchwork.

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

Waring, Elin, Michael Quinn, Amelia McNamara, Eduardo Arino de la Rubia, Hao Zhu, and Shannon Ellis. 2021. *Skimr: Compact and Flexible Summaries of Data.* https://CRAN.R-project.org/package=skimr.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in R.* https://yihui.org/knitr/.