

# Number of Toronto Crime Cases Is Changing From Rising To Declining\*

Yuxuan Yang

06 February 2022

## Abstract

Neighborhood Crime Rates Data was imported here from the City of Toronto Open Data Portal, and table and line graphs were drawn to investigate the growth trend of Toronto crime cases. A result that Toronto crime cases have increased from 2014 to 2019 but decreased from 2019 to 2020 was found. This result matters because it could motivate Toronto Police Service to investigate causes of this decrease in crime cases and to better prevent crimes.

## 1 Introduction

Fighting against crimes and reducing crime cases were always the core goals of Toronto Police Service. One of the most efficient measure of crime fighting is to be aware of causes of crimes and what could contribute to the decline in crime cases.

In this paper we looked at the growth trend of number of crime cases from 2014 to 2020 in both Toronto and 5 individual neighborhoods with most populations. We created a table and two line graphs to demonstrate our dataset, and found that the number of crime cases in Toronto has experienced a sharp decline from 2019 to 2020 despite the overall increasing trend from 2014 to 2019. Future analysis could focus on finding other data, building regression models, and discovering potential causes of this sharp decline in crime cases.

The remainder of this paper is: Section 2 explains our data, and Section 3 contains discussions about our findings, weaknesses, and next steps.

## 2 Data

### 2.1 Data Source And Collection Methods

This paper uses data on neighborhood crime rates published by the Toronto Police Services on the City of Toronto Open Data Portal (*Neighbourhood Crime Rates* 2021); the data was last refreshed on May 6, 2021. I imported the data as csv file directly using the R package `opendatatoronto` (Gelfand 2020) and the statistical programming language R (R Core Team 2020). This dataset contains information on numbers of different types of crimes in each of the 140 Toronto neighborhoods from the year 2014 to 2020, together with crime rates of different types of crimes in each neighborhood from 2014 to 2020, where the crime rate is calculated as the crime count per 100,000 population\* per year in each neighborhood.

The Toronto Police Service did not specify how exactly the crime cases in this dataset were processed, but we know crimes can be reported to the Police and collected through many channels. For example, citizens can voluntarily report crimes they have experienced or observed. It is also possible that the Police will discover crimes on site through daily patrols. Nevertheless, the fact that crimes can be self-reported indicates the possibility of crimes being under-reported, which could result in potential bias in number of crime cases on record.

---

\*Code and data are available at: [https://github.com/Yuxuan-Yang-Maggie/starter\\_folder](https://github.com/Yuxuan-Yang-Maggie/starter_folder).

## 2.2 Population and Samples

The population of this dataset is the total population in each Toronto neighborhood in each year from 2014 to 2020. Note that the total population in each neighborhood in this dataset is only a projection estimated by Environics Analytics instead of the true value. The sampling frame is all the crime cases reported and recorded in this dataset.

## 2.3 How Respondents Were Found

Respondents in our dataset are the reporters of crime cases in each neighborhood. Since crime cases are mainly collected through either self-reporting or police discoveries, our respondents here are discovered by the Toronto Police Service if evidences of crime are observed. In terms of crime data, the action of non-response will cause crime cases in Toronto to be under-reported and under-represented, leading to significant bias in the results of our crime data analysis.

## 2.4 Key Features, Major Strengths And Weaknesses

The most important feature of Open Data Toronto is that it can be easily accessed and reproduced by the general public. While a typical strength of Open Data Toronto would be that it facilitates the flow of information across different government sections and also between the government and the general public, it also ensures the greatest extent of transparency and authenticity while conveying reliable data to citizens. Nevertheless, Open Data Toronto also has a severe weakness, which is the problem of missing information.

## 2.5 Data characteristics

I am interested in the growth of crimes cases in Toronto neighborhoods from 2014 to 2020. The first thing I did is to write our crime dataset into csv file and read this csv file using the R programming language (R Core Team 2020) and R package `tidyverse` (Wickham et al. 2019). Then I created new variables to record total crimes per neighborhood in each year. I also selected Neighbourhood, population, and all the new variables I just created and transformed them into a new, long dataset using the `gather` function from `tidyr` package (Wickham 2021).

I chose to ignore all the crime rates data and only focus on the number of crime cases because I wanted to see whether crime cases in Toronto have declined or rose from 2014 to 2020. I also ignored `Hood_ID` and `OBJECTID` because they are hard to recognize. What's more, I ignored the specific types of crimes and geometry since I am only interested in total crime cases.

Instead of specific summary statistics, I created a table using the `kable` function from the `knitr` package (Xie 2021) to show the first 10 observations of the long dataset I created (Table 1). This table contains columns recording neighborhood names, population per neighborhood, year, and crime cases per neighborhood per year.

Table 1: First 10 rows of our long dataset showing crimes per year per neighborhood

Neighbourhood	F2020_Population_Projection	num_crime	year
Yonge-St.Clair	14083	62	2014
York University Heights	30277	578	2014
Lansing-Westgate	18146	109	2014
Yorkdale-Glen Park	17560	270	2014
Stonegate-Queensway	27410	226	2014
Tam O'Shanter-Sullivan	29970	205	2014
The Beaches	23364	167	2014
Thistletown-Beaumont Heights	10948	113	2014
Thorncliffe Park	23518	161	2014
Danforth East York	18427	108	2014

From Table 1, we observed number of total crimes from 10 different neighborhoods in 2014. We then created a line graph using `ggplot2` package (Wickham 2016) to get an insight on how crime cases developed over the course of 7 years. Here we chose the top 5 neighborhoods with most populations as examples to show the trend of crime growth.

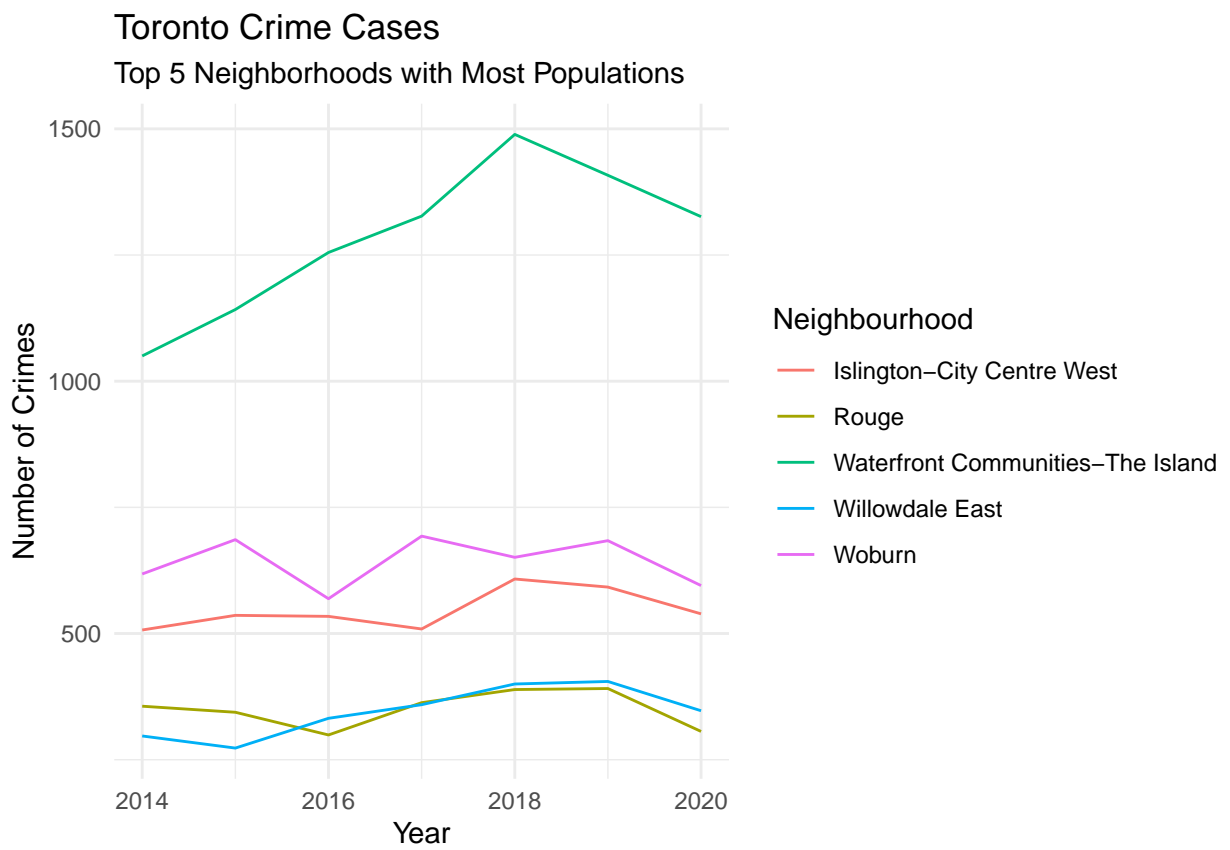


Figure 1: Crime growth of top 5 neighborhoods with most population from 2014 to 2020

From Figure 1, we observed a steady increase in crime cases in Waterfront Communities-The Island from 2014 to 2018 and decrease from 2018 to 2020. For the other 4 neighborhoods, the number of crime cases are relatively steady from 2014 to 2020. This brought us into the next step, which is to look at how number of crime cases developed for the whole Greater Toronto Area. So we will look at the total crime cases in Toronto from 2014 to 2020.

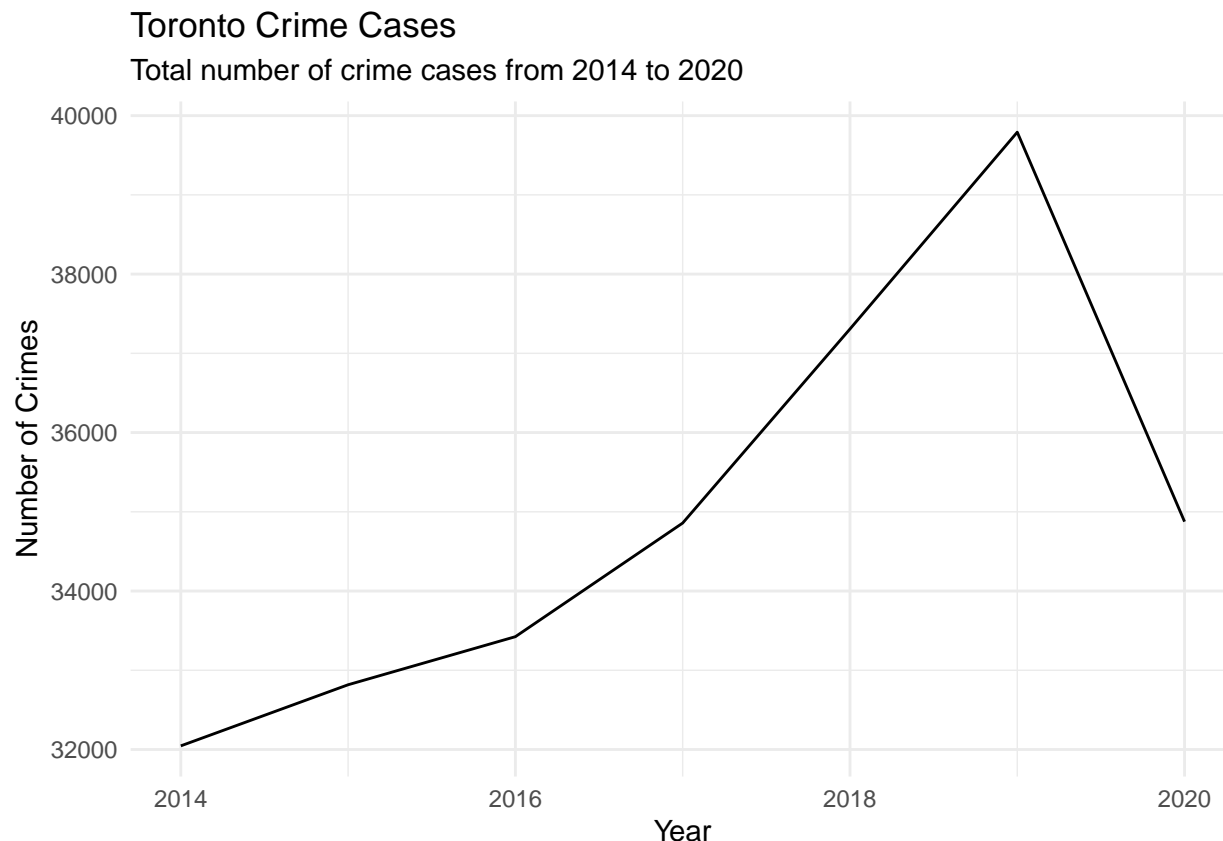


Figure 2: Crime growth in Toronto from 2014 to 2020

From Figure 2, we observed that Toronto experienced a steady increase in total crime cases from 2014 to 2019, but a sharp decline from 2019 to 2020. Since we also observed that all of the 5 neighborhoods we examined before experienced a decline in crime cases from 2019 to 2020, this brought us into a further question of what has happened in 2019 and 2020 that could cause this sharp decline in crime cases.

### 3 Discussion

In this paper we briefly looked at the crime cases over the course of 7 years from 2014 to 2020 in each of the 5 example neighborhoods, together with the total crime cases in Toronto. We determined that the growth trend of Toronto Crime Cases have changed from positive to negative, and that there might be some external factors that led to an overall decline in the number of crime cases from 2019 to 2020. As we researched on possible variables relating to crimes, we found that unemployment rate, poverty rate, population density, and gun ownership could all be potential causes of changing crime cases (Bhattacharya 2020). However, due to the weakness that we have no other information in this dataset except pure crime cases and crime rates, we will need to combine this dataset with other data in the next step in order to perform further analysis on what potential factors could influence the growth trend of crime cases.

## References

- Bhattacharya, D. A. 2020. “Analysis of the Factors Affecting Violent Crime Rates in the US.” *International Journal of Engineering and Management Research*, Volume-10, Issue-5. <https://ssrn.com/abstract=3732028>.
- Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*.
- Neighbourhood Crime Rates. 2021. Toronto Police Services. <https://open.toronto.ca/dataset/neighbourhood-crime-rates/>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2021. *Tidyr: Tidy Messy Data*. <https://CRAN.R-project.org/package=tidyr>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.