

# Crime Rates in Toronto Neighborhoods By Year and Type\*

Yuxuan Yang

04 February 2022

## Abstract

First sentence. Second sentence. Third sentence. Fourth sentence.

## 1 Introduction

First paragraph is going to be motivational and broad.

Second paragraph is about what was done and what was found.

Third paragraph is about implications.

The remainder of this paper is: Section 2 explains the data. Section ?? covers discussions. . . . .

## 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 (Data 2020); the data was last refreshed on May 6, 2021, and I imported the data as csv file directly using the R package `opendatatoronto` (Gelfand, 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 it mentions that the crime rates were calculated using “the standard definition by StatsCan” (cite here). In addition, 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 of number of crime cases on record.

### 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.

---

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

## 2.3 How Respondents Were Found

Respondents in our Neighborhood Crime Rates 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. 2021). 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 et al. 2021).

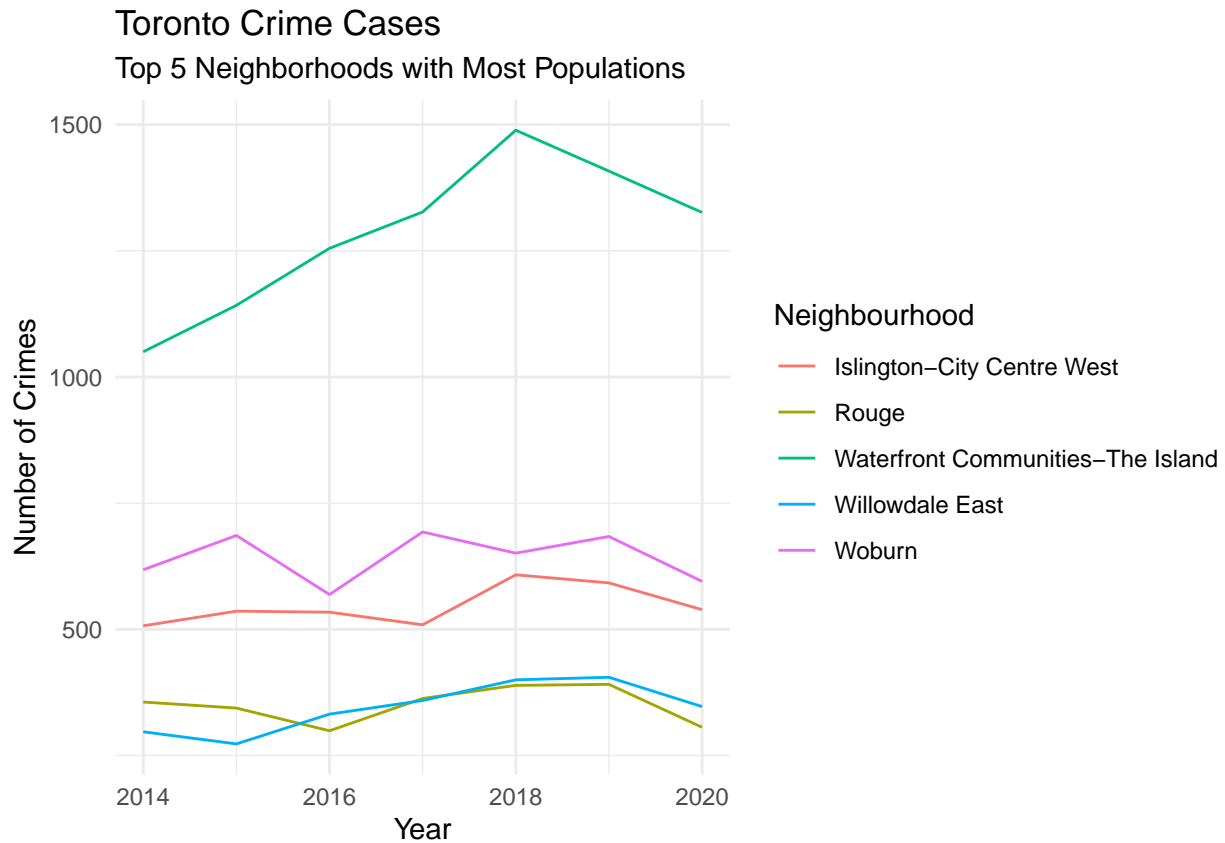
I chose to ignore all the crime rates data and only focus on the number of crime cases because I wanted to see whether crime cases in Toronto have declined or rose from 2014 to 2020. I also ignored Hood\_ID and OBJECTID because they are hard to recognize. What's more, I ignored the specific types of 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 and **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 ??, 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.



From Figure ??, 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.

### 3 Discussion

#### 3.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

#### 3.2 Second discussion point

#### 3.3 Third discussion point

#### 3.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

## Appendix

### A Additional details

## B References