

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 results.....

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 (cite opendatatoronto here).

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.

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

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

```
## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v dplyr  1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.1.1      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

## Rows: 140 Columns: 104

## -- Column specification -----
## Delimiter: ","
## chr  (3): Neighbourhood, Hood_ID, geometry
## dbl (101): _id, OBJECTID, F2020_Population_Projection, Assault_2014, Assault...

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

## Warning: package 'knitr' was built under R version 4.1.2
```

```
# Plot number of crimes on y-axis and year on x-axis for 5 neighborhoods
# with highest populations in 2020
new_crime_hood %>%
  filter(Neighbourhood %in% c("Waterfront Communities-The Island", "Woburn",
                              "Willowdale East", "Islington-City Centre West",
                              "Rouge")) %>%
  ggplot(aes(x=year, y=num_crime, group=Neighbourhood, color=Neighbourhood)) +
  geom_line() +
  labs(title = "Toronto Crime Cases",
        subtitle = "Top 5 Neighborhoods with Most Populations",
        x = "Year",
        y = "Number of Crimes") +
  theme_minimal()
```

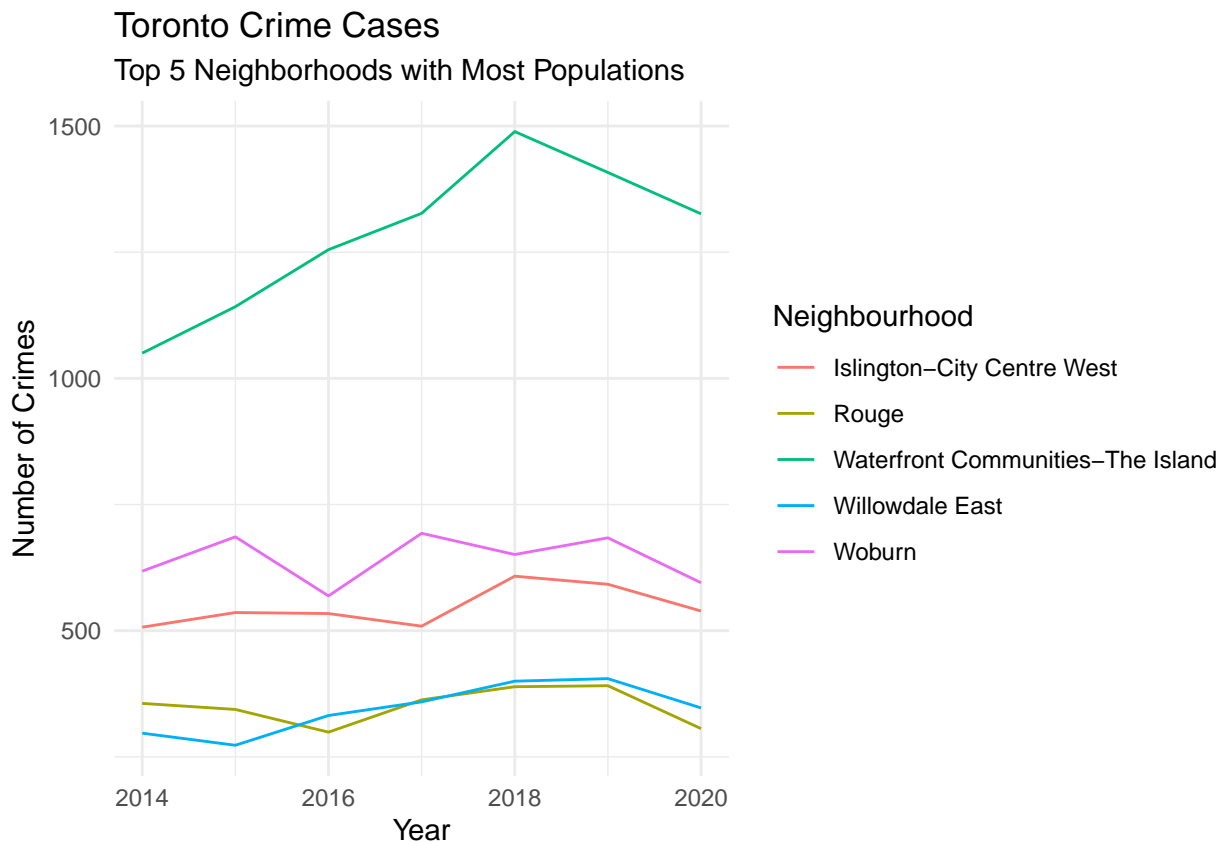


Figure 1: Crime growth of top 5 neighborhoods with most population from 2014 to 2020, echo = FALSE

Our data is of penguins (Figure ??).

Talk more about it.

Also bills and their average (Figure ??). (Notice how you can change the height and width so they don't take the whole page?)

Talk way more about it.

3 Discussion

3.1 First discussion point

If my paper were 10 pages, then should 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