

Analyzing Violent Crime Statistics in Toronto Neighbourhoods*

Exploring how violent crime rates vary in neighbourhoods of a city

Hritik Shukla

January 23, 2024

The City of Toronto has made crime statistics for each of its neighbourhoods publicly available through its Open Data Program. This data will be analyzed to observe how violent crime rates are distributed amongst the neighbourhoods of the city. A trend has been observed where a small number of neighbourhoods exhibit a larger crime rate than the majority of other neighbourhoods. However, these distributions vary from crime to crime.

1 Introduction

While we read about rising and falling crime rates in the cities we live in, we don't know if these crime rates affect every person equally. It could be the case that these crime occurrences are distributed evenly throughout the city, where every person living in different parts of the city has the same probability of being a victim of a crime. However, it could also be the case that not every person will run into these crimes in their day to day life, and this probability will be heavily influenced by where they live. The aim of this paper is to explore if all neighbourhoods of Toronto follow the general trends of rise and fall in violent crimes, or are there variances between neighbourhoods.

In the Data Section [2](#) section of this paper, I will go through the dataset used for this analysis in detail, covering the source of the data, the different features present in the data and the variables of interest to us.

In the Results Section [3](#) section of this paper, I will go through the conclusions I derived, supporting my arguments with plots and data generated from the data acquired above, with the help of various libraries in R.

*Code and data are available at: <https://github.com/hritikshuklas/toronto-neighbourhood-crimes>

2 Data

The analysis for this project was done using R programming language (R Core Team 2022), and multiple libraries were used within R to help with tasks such as plotting graphs, data cleaning and manipulation. In particular, the `opendatatoronto` library (Gelfand 2022) was used for acquiring the dataset off of Open Data Toronto’s servers, `here` library (Müller 2020) was used for loading in data from the local project files, `janitor` (Firke 2023), `tidyverse` (Wickham et al. 2019), `reshape2` (Wickham 2007) and `dplyr` (Wickham et al. 2023) were used for cleaning, manipulating and testing the dataset, and `ggplot2` (Wickham 2016) along with `knitr` (Xie 2023) was used for creating plots and graphs shown in the analysis. The details of how data collection, cleaning and analysis was done can be found below.

2.1 Data Collection

The data used for this paper is published by Toronto Police Services under the Open Government License for the City of Toronto. The data was acquired from the City of Toronto’s open data portal through the `opendatatoronto` library (Gelfand 2022) and saved locally for further processing. .

2.2 Data Features

This data contains crime statistics for all 158 neighborhoods of Toronto. The 158 neighbourhood model for the city was adopted fairly recently, as it was brought into use in April 2022 (*Toronto’s Social Planning Neighbourhoods Changed* 2022), compared to the 140 neighbourhood model, which had been devised in 1990s and brought into use in the early 2000s.

The data includes counts for individual occurrence for each category of crime - assault, auto theft, break and entering, robbery, homicide, theft over (\$5000), shooting and firearm discharges. The data also includes the crime rate for each of these categories, calculated as crimes committed per 100,000 population (following the standard definition of crime rate set by Statistics Canada). These crime rates were calculated using population estimates for each neighborhood in 2023, which were provided by Environics Analytics. Both the individual crime occurrences as well as crime rates have been given for the year 2014 to 2023, where each year gets its own column (*Neighbourhood Crime Rates - City of Toronto Open Data Portal* 2024).

Table 1: Sample of assault count and assault rate arranged according to years

assault_2022	assault_2023	assault_rate_2014	assault_rate_2015
128	101	344.9786	332.1355
130	105	386.9636	449.1277

assault_2022	assault_2023	assault_rate_2014	assault_rate_2015
95	104	412.0070	422.1597

Note that our purposes, we only need to study violent crimes, i.e., assault, robbery, homicide and shootings - other categories of crime as well as population estimates were not used in this paper.

The data also contained geographical geometry of each neighbourhood, described with a set of coordinates. This information was not used for our analysis.

2.3 Data Quality

According to the dataset's entry on Open Data Toronto, the data was last updated on 14 January 2024, which means we are working with the latest possible data for crime in the city of Toronto.

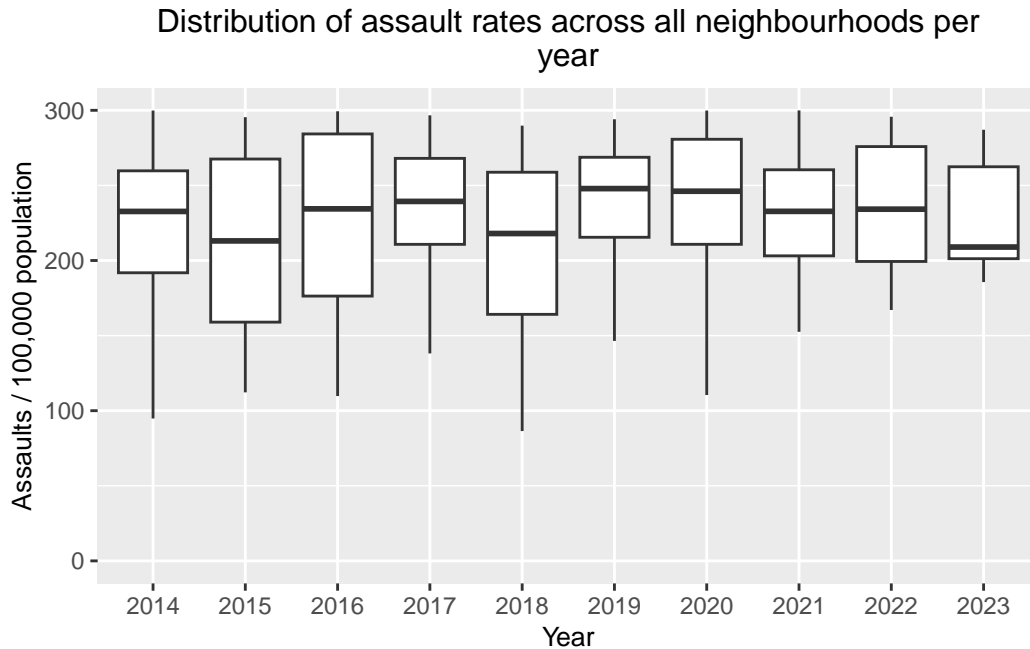
The metadata for the dataset is sufficient - the dataset description on the portal along with the column headers describe the data well.

In terms of completeness, even though the score on the Open Data Portal gives the dataset a 100% score for completeness, there are some missing entries for some statistics for certain neighbourhoods. Such values are treated as zeroes for our purposes. One of the reasons for this missing data might be the age of the new neighbourhood division system, as explained above.

3 Results

For our analysis, we will primarily be using vertical box plots to plot our data, as it gives us a clear view of how different crime rates are distributed amongst the neighbourhoods of Toronto.

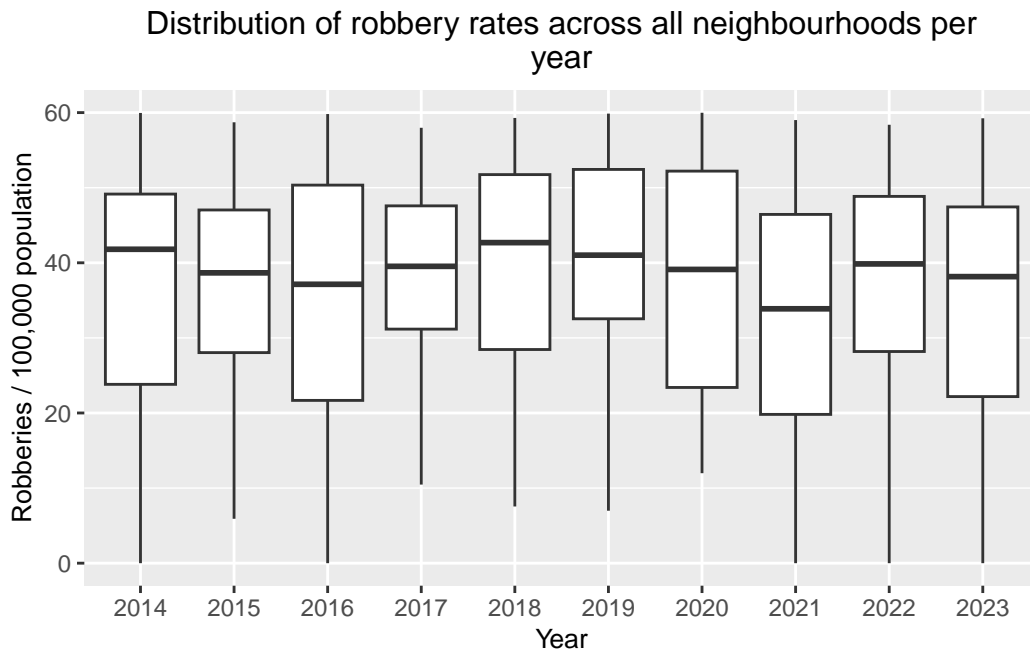
The top whisker (slim black line) of our box plot represents the top 25% of our data, and the bottom whisker represents the bottom 25% of the data. The black line inside the box of the plots represents the median of the data. The top segment of the box represents the top 50% to top 75% of our data, whereas the bottom segment of the box represents the bottom 25% to the bottom 50% of our data. ## Assault Rates {#subsec-results-assault}



From the box plots, we can observe the median and notice a trend line which suggests that assaults were trending up until 2020, after which the median values started dropping. Also notice that up until 2022, the distribution of assault rates have been relatively normal as the box plots have been fairly symmetrical. Also notice the huge variance in all of the years up until 2022, some neighbourhoods of Toronto experience a lot more assaults per capita than others. However, in 2023, we see that this distributions skews heavily to the right, indicating extremely large outliers in the data. Also notice that the bottom whisker is considerably shorter, and that the bottom 50% of the data doesn't deviate much from the median.

From these box plots we can deduce that in 2023, even though the median assault rate had decreased, a lot of neighbourhoods experienced a lot more assaults than they would have in the previous years. This could be explained by the dramatic increase in hate crime, upwards of 42%, in Toronto, as reported by CP24 (Tsekouras 2023).

3.1 Robbery Rates

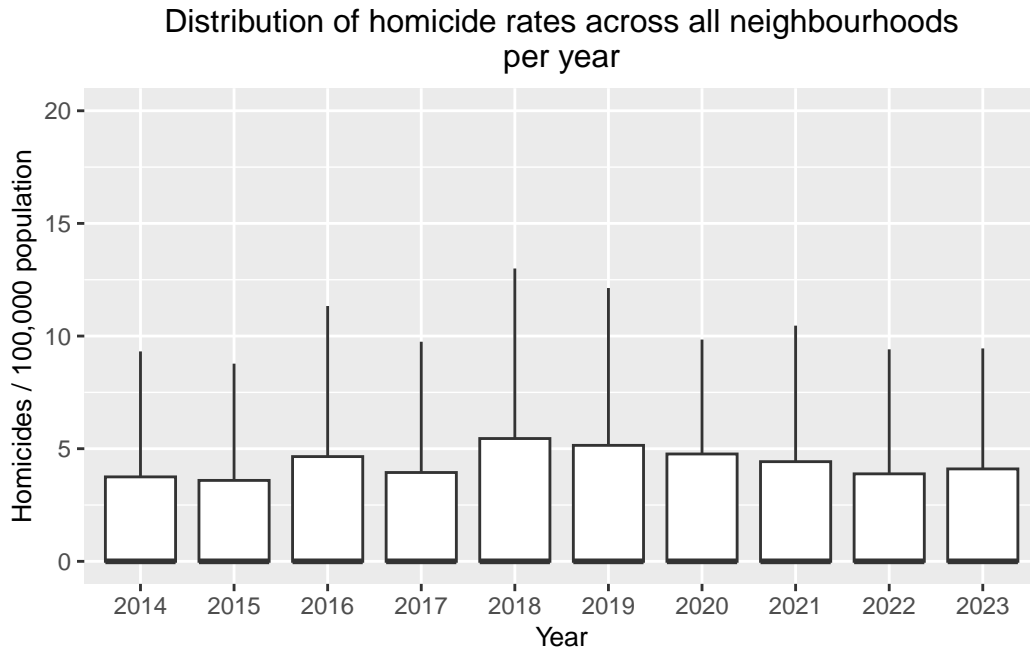


Similar to assaults, we can see that throughout the years, robbery rates are fairly normally distributed and that there is large variance in amongst the neighbourhoods. Top 25% of the neighbourhoods have a robbery rate as high as 50-60%, whereas the bottom 25% go as low as 0 robberies. Such a high variance is indicative of massive differences amongst neighbourhoods of Toronto - not all neighbourhoods experience the same level of robberies. Notice that this is where the similarities end.

Everything is not the same between the two crimes. Notice that the median doesn't show a strong trend in either direction, it hovers around 40 robberies / 100,000 population. Also notice that in 2023, there isn't a huge change in the shape of the boxplot - the distribution stays relatively similar to past years.

This shows that there hasn't been a huge change in robbery rates across neighbourhoods over the years - it has stayed stagnant for about a decade.

3.2 Homicide Rates

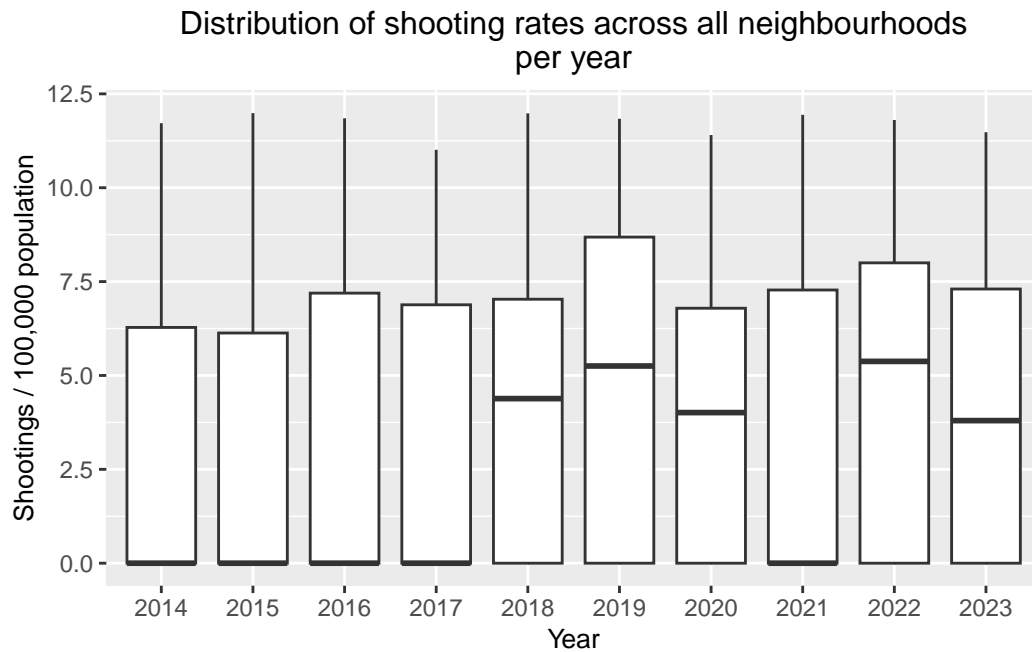


Homicides have historically been low in Toronto, as we can see from this graph, with the median homicides sitting at approximately 0 per 100,000 population. From the box plots we can see that about 50% of Toronto's neighbourhoods have experienced little to no homicides for the last decade. Even in 2018, when Toronto broke a homicide record of 27 years with 96 homicides (Goodfield 2023), we see no deviation from the median, only an increase in the maximum (as indicated by the top whisker), indicating that majority of these homicides happen in a small number of neighbourhoods.

Supporting this idea, notice that the top 25% of homicide rates sit extremely far from this median, ranging from 5 to 10 homicides for the last 10 years. This is a significant portion of the Toronto's neighbourhoods which experience very high rates of homicide every year, and there isn't any indication of this changing for the better or for the worse.

From this, we can conclude that for the last ten years, a small portion, about 25%, of Toronto's neighbourhoods experience a majority of the homicides reported in the city. The rest experience little to no homicides at all.

3.3 Shooting Rates



An interesting observation we can make is that up until 2017, the shooting rate distributions heavily resemble the distributions of homicide rates in the city. We can observe from the plots that up until 2017, about 50% of the neighbourhoods experienced little to no shootings at all and those which did showed huge variances in shooting rates, indicating large outliers. Notice that the median value spikes suddenly in 2018, which was Toronto's deadliest year in recorded history (Goodfield 2023). From this point on, the median returns to historical lows only once, in 2021, presumably from the COVID lockdowns that were put in place, which kept most people in their homes - it hovers near the high of 2018 for the rest of the years. Moreover, right after 2021, the median jumps back to the highs of pre-COVID years of 2018 and 2019. Notice that we do observe a noticeable drop in 2023, however, the shooting rate is still far away from what it used to be a decade ago. Also notice that nothing changes for the top 25% of neighbourhoods before or after 2017, they show the same trends for the entire decade.

From these distributions, we can conclude that up until 2017, about 50% of the neighbourhoods of Toronto experienced little to no shootings, and a large number of shootings occurred in only a few neighbourhoods. Most of the city was largely sheltered from gun violence up until 2017, after which only 25% of the neighbourhoods remained sheltered. More of the city was exposed to gun violence, as the neighbourhoods which stand in the first quartile (bottom 25% to 50%) are now experience some levels of gun violence which they didn't before. Meanwhile, the neighbourhoods which experience the most gun violence haven't shown any positive or negative trends - they've been experiencing the same levels of violence for over a decade now, with no signs of improvement in the future.

References

- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- Goodfield, Kayla. 2023. “Hate Crimes in Toronto Increased by 42 Per Cent in 2023, Most Calls about Antisemitism: Police.” <https://toronto.ctvnews.ca/toronto-s-most-deadly-year-a-look-at-the-96-homicide-victims-of-2018-1.4232568>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- Neighbourhood Crime Rates - City of Toronto Open Data Portal*. 2024. Toronto Police Services. <https://open.toronto.ca/dataset/neighbourhood-crime-rates/>.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Toronto’s Social Planning Neighbourhoods Changed*. 2022. City of Toronto. <https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/about-toronto-neighbourhoods/>.
- Tsekouras, Phil. 2023. “Hate Crimes in Toronto Increased by 42 Per Cent in 2023, Most Calls about Antisemitism: Police.” <https://www.cp24.com/news/hate-crimes-in-toronto-increased-by-42-per-cent-in-2023-most-calls-about-antisemitism-police-1.6721710>.
- Wickham, Hadley. 2007. “Reshaping Data with the reshape Package.” *Journal of Statistical Software* 21 (12): 1–20. <http://www.jstatsoft.org/v21/i12/>.
- . 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- 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>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.