Investigating the change of Toronto crime rates from 2014-2023*

Charles Cai

January 24, 2024

Table of contents

*Code and data are available at: https://github.com/peachvegetable/STA_302_neighbourhood_crime-rate

1 Introduction

Toronto, Canada's largest city, hosts a population exceeding three million. Renowned for its robust economic landscape and cultural diversity, the city has concurrently experienced notable challenges in managing urban crime. In 2022, Toronto recorded a significant 15 percent increase in its crime rate, surpassing the national urban average by more than threefold [@citeCP24].

This paper employs the Neighbourhood Crime Rates dataset, encompassing a period from 2014 to 2023, to analyze the trend of crime in Toronto. The dataset provides detailed information on various crime categories, such as theft, shooting, and assault. Through statistical analysis, including the generation of descriptive tables and graphical representations, this study aims to elaborate the temporal patterns and fluctuations in crime rates.

Section 2 introduces the neighbourhood crime rates dataset obtained from OpenDataToronto. The dataset is cleaned and only a part of the crime types are chosen to be analyzed. In section 2.1, I choose a location to do the analysis, and modify the data frame into a table which is easier to plot the graphs in the coding aspect. Section 2.2 introduces another location to do the comparison with the previous location, and a similar method is applied to the data. Lastly, section 3, the discussion part concludes the overall idea of this paper and point out what could be done to further to investigte in this dataset.

2 Data

To examine the trends of crime in Toronto, I have sourced the "Neighbourhood Crime Rates" dataset from the Toronto Open Data Portal (Gelfand 2022). This dataset, published by the City of Toronto, details crime statistics across various neighborhoods from 2014 to 2023. It has different crime types, including but not limited to assault, shooting, and theft, and presents these data points in the context of annual crime rates per 100,000 population.

Given the dataset's extensive scope, encompassing 158 neighborhoods, a selection approach was necessary to manage the breadth of the analysis. Therefore, a subset of locations was chosen to be analyzed. The dataset consists of 185 variables (columns) and 158 locations. For the purposes of this paper, the focus will be narrowed to the rate of six common crimes (i.e. assault, break enter, homicide, robbery, shooting and bike theft) spanned from 2014 to 2023 (not crime counts) and Hood_id. In particular, the analysis is based on crime rates over crime counts. This decision is driven by the fact that each location within the dataset has varying population sizes. By using crime rates, which are normalized per 100,000 population, we effectively eliminate the bias that could arise from comparing crime counts across differently populated areas. This approach ensures a more accurate comparison of crime data.

The analysis will be using R (R Core Team 2023) and various R packages such as "tidyverse" (Wickham et al. 2019), "ggplot2" (Wickham 2016), "knitr" (Xie 2023), "kableExtra" (Zhu 2021), and "dplyr" (Wickham et al. 2023), the dataset is cleaned and transformed to a table,

Table 1: Sample of example Neighbourhood Crime Rates Dataset

Id	Assault	BreakEnter	Homicide	Robbery	Shooting	BikeTheft
174	459.4	43	NA	13.6	NA	168.3
173	696.4	31	NA	92.9	NA	145.9
172	751.6	46	NA	151.8	7.2	274.6
171	872.7	82	NA	68.6	7.6	209.6
170	3686.1	108	6.8	420.9	13.6	1547.8
169	745.2	61	NA	105.0	5.2	293.9

since the year variable is in the columns, the table only displays the 2023 data for assault, break enter, homicide, robbery, shooting, and bike theft as an example. Moreover, the digits are to the 1 decimal place (see Table 1).

This comprehensive dataset includes a total of nine distinct crime categories: robbery, assault, theftover, homicide, break and enter, bike theft, auto theft, break enter, and theft from motor vehicles, spanning the years 2014 to 2023. Since there are 4 types of theft, I choose the most common one to be in my analysis, i.e. bike theft. Table 1 presents the initial six rows from the Neighbourhood Crime Rates Dataset.

2.1 Wexford crime data

Given the expansive coverage of 158 neighborhoods in the dataset, the scope of this study will be concentrated on two specific neighborhoods for detailed examination. Kensington-Chinatown (HOOD_ID: 78) and Wexford/Maryvale (HOOD_ID: 119) have been strategically chosen based on their geographical differences. Kensington-Chinatown is located in downtown Toronto, while Wexford/Maryvale is in the northeastern part of the city. This geographical contrast generages the hypothesis that crime rates and trends may vary significantly between these two distinct regions.

Table 2: Crime Rates by Year and Crime Type for Wexford/Maryvale

Year	ASSAULT	BIKETHEFT	BREAKENTER	HOMICIDE	ROBBERY	SHOOTING
2014	670.403	67.395	439.8411	3.5	180.9024	10.64132
2015	778.4855	35.3857	325.5485	0.0	127.3885	10.61571
2016	711.468	28.17695	267.681	7.0	84.53085	14.08848
2017	710.3559	17.41068	302.9459	0.0	121.8748	20.89282
2018	693.229	24.02279	343.1827	6.9	120.1139	17.15913
2019	821.7039	20.37282	292.0105	3.4	105.2596	10.18641
2020	552.0371	37.25404	270.9385	3.4	98.21519	10.16019

(continued)

Year	ASSAULT	BIKETHEFT	BREAKENTER	HOMICIDE	ROBBERY	SHOOTING
2021	719.0584	33.91785	237.425	3.4	94.96999	13.56714
2022	705.0806	33.47841	253.0194	3.4	131.5701	16.86796
2023	870.4385	90.3917	257.7837	3.4	103.7831	13.39136

Data from the neighborhood with HOOD_ID 119 was filtered and restructured into Table 2, displaying years as rows and crime types as columns. Each cell indicates the crime rate for a specific year and crime type. The Homicide column's NA values were replaced with the mean rate to maintain consistency in trend analysis.

This study aims to identify patterns in crime rates from 2014 to 2023, assessing whether trends are linearly increasing or exhibit fluctuations.

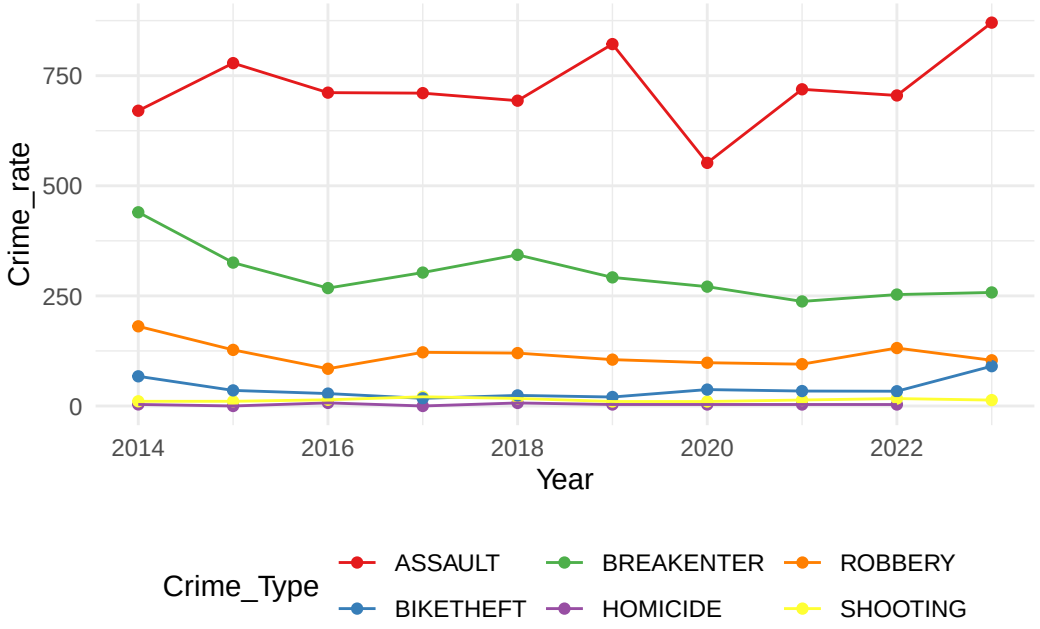


Figure 1: Crime Rates by Year and Crime Type for Wexford/Maryvale

Figure 1 facilitates a comparative analysis of six distinct categories of crime: assault, bike theft, breaking and entering, homicide, robbery, and shooting. From a statistical standpoint, the following observations can be deduced: the assault category shows a volatile pattern, with a notable peak in 2015 and 2019, a subsequent decline in 2020, followed by an increase 2023. The fluctuations suggest that external factors may be influencing the incidence of assault crimes. The bike theft category exibits less variability, bike theft rates appear to maintain a relatively low and steady rate with a slight upward trend towards 2023. Then, the rate of

break and entering demonstrates a general downward trend, with occasional year-over-year increases. The overall decline may reflect effective crime prevention strategies or shifts in criminal behavior. Homicide rates are considerably lower compared to other crime types and are characterized by relative stability. Notably, there are years with zero reported incidents. The trend for robbery rates declines over the period, which might suggest improved community safety. Similar to homicide, shooting incidents are less frequent, yet they display a cyclical pattern with periods of increase and decrease.

2.2 Kensington crime data

Table 3: Crime Rates by Year and Crime Type for Kensington-Chinatown

Year	ASSAULT	BIKETHEFT	BREAKENTER	HOMICIDE	ROBBERY	SHOOTING
2014	1764.801	496.519	555.8854	0	426.3587	10.79389
2015	2029.681	567.4378	540.1572	5.456133	294.6312	21.82453
2016	2033.973	703.6447	439.7779	27.48612	368.314	49.47501
2017	2251.045	771.7869	568.1209	0	487.7264	48.23668
2018	2265.861	739.6604	1250.13	15.62663	411.5012	36.46213
2019	1935.287	604.4598	1412.099	0	238.7362	20.31798
2020	1969.606	494.901	869.826	4.999	214.957	29.994
2021	1879.141	529.1267	479.6756	9.890219	123.6277	44.50599
2022	1849.964	419.2378	504.9769	4.855547	203.933	24.27774
2023	2270.479	645.3436	683.028	4.710537	216.6847	23.55269

The data of Kensington-Chinatown is cleaned and transformed to a table in a similar way as what we did to the Wexford data. Since there are no NA values in the Kensington-Chinatown data, we can directly restructure it to a table in the same way. Now, we are going to build the line graph as what we did to the Wexford data.

From Figure 2 we can observe the following patterns: the most prominent trend is the assault rate, which shows a steady increase from 2014 onwards, peaking around 2019 before plateauing. The rate of break and enter fluctuates significantly, with a notable peak around 2017. This could indicate a year-specific factor that temporarily influenced the trend of this crime. Robbery rates demonstrate a general decline over the years, which could be indicative of effective policing strategies or social changes within the community. There is a sharp increase in bike theft around 2017, followed by a sharp decline. The homicide rate remains relatively flat and low in comparison to other crimes, suggesting it is less affected by the factors influencing the rates of other crime types. The shooting rate, while varying, does not exhibit any clear trend, remaining consistently low throughout the observed period.

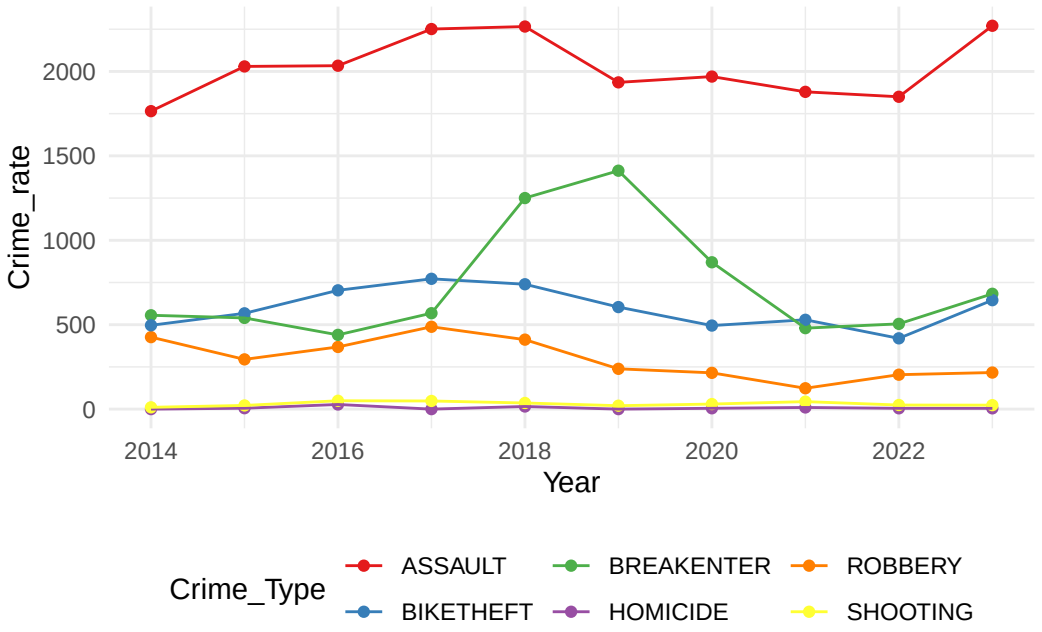
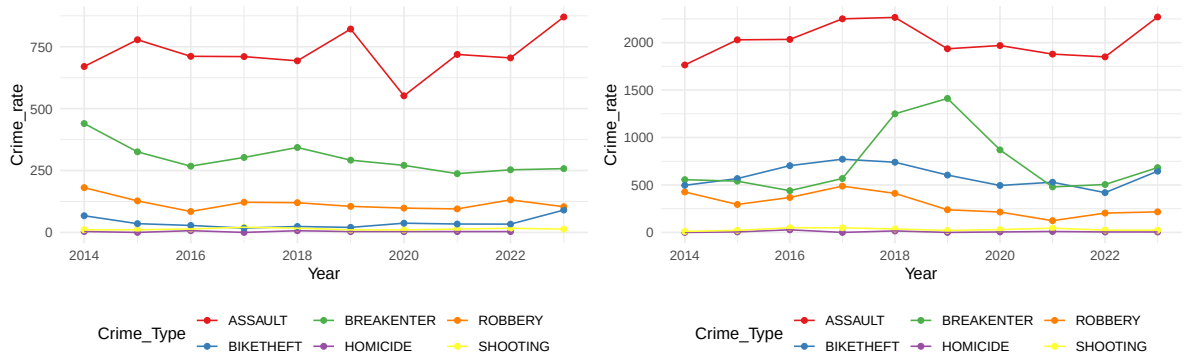


Figure 2: Crime Rates by Year and Crime Type for Kensington-Chinatown

2.3 Comparison of the two locations

In comparing the two line graphs depicting crime rates for Wexford and Kensington-Chinatown, we can make the following comparative analysis: the crime rates in Wexford show variability but less pronounced peaks and troughs compared to Kensington-Chinatown. There's a notable increase in assault rates over time, with a significant drop in 2020, possibly due to external factors such as the COVID-19 pandemic. The Kensington-Chinatown neighborhood experiences more pronounced fluctuations, particularly with break and enter crimes peaking sharply in 2017. It also shows a upward trend in assault rates, which far exceeds those in Wexford. Assault rates in Kensington-Chinatown are consistently higher and show a steeper increase compared to Wexford, suggesting potentially different socio-economic conditions or policing issues. Both neighborhoods exhibit a spike in bike thefts at different times (Wexford in 2023 and Kensington-Chinatown in 2017), indicating that local factors may affect this type of crime at different times. Homicide rates remain low in both areas, with little fluctuation. Kensington-Chinatown shows a slightly more variable rate compared to the almost flat trend in Wexford. Robbery rates are decreasing in both neighborhoods, but the decline is more pronounced in Wexford. Shooting rates are low in both neighborhoods, with no prominent increasing or decreasing trend.

The scale of crime rates is significantly different between the two neighborhoods, with Kensington-Chinatown experiencing higher rates of assault and break and enter crimes. The



(a) Wexford

(b) Kensington

Figure 3: Wexford and Kensington

scope of the crime types is similar across both neighborhoods, but the intensity and variation of crimes are neighborhood-specific.

3 Discussion

The investigation reveals an upward trend in the overall crime rates for both neighbourhoods, with Kensington-Chinatown showing a more noticable increase. Comparative analysis indicates that Kensington-Chinatown's crime rates are not only higher but also more volatile, suggesting the socio-economic and environmental factors may be the reason behind this phenomenom.

The analysis was constrained by the limitations of the crime data, such as potential underreporting and the absence of socio-economic variables, which may provide a more comprehensive understanding of the factors behind crime.

Subsequent research could explore the reasons for the changes in crime rates from 2014 to 2023. Consideration of factors such as the area's income level, police budget, and policy changes could clarify the forces affecting crime patterns.

Reference

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

R Core Team. 2023. *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.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, 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/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.