Strong Evidence of Income and Population being Correlated with Crimes Committed in Toronto Neighbourhoods in 2020*

Janel Gilani

January 25, 2024

This study examines the crime landscape in Toronto during the year 2020, focusing on the relationship between crime rates, income levels, and population figures across neighborhoods using a comprehensive dataset from Open Data Toronto. Our findings reveal an inverse correlation between crime rates and income, indicating that neighborhoods with lower income levels tend to experience higher crime rates. Additionally, a positive correlation emerged between population size and crime rates, suggesting that as the population increases, so does the incidence of crime in Toronto. These insights are crucial for informing targeted policies, enhancing law enforcement strategies, and developing community initiatives to address crime and socio-economic disparities in urban settings.

Table of contents

*Code and data are available at: https://github.com/JanelGilani/spread-of-crimes-in-toronto

1 Introduction

Toronto, once hailed for its safety, confronts a contemporary challenge as crime rates in its neighborhoods become a growing concern. Despite the overall crime rates across Canada witnessing a commendable 30% decrease in the past two decades (Statista 2024), Toronto has experienced an unsettling rise in criminal activities. This paper embarks on a comprehensive statistical analysis of crime in Toronto neighborhoods throughout the year 2020, delving into the intricate relationship between crime rates, income levels, and population figures.

The global context of the COVID-19 pandemic in 2020 adds a layer of complexity to the study, prompting an exploration of how crime dynamics unfolded during these unprecedented times. The dataset used for this analysis is meticulously sourced from opendatatoronto (Gelfand 2022), providing a granular account of various crimes committed across Toronto's diverse neighborhoods. Utilizing robust statistical tools, particularly R (R Core Team 2020), our investigation seeks to unearth nuanced patterns and relationships within the data. It is crucial to note that the emphasis of this study lies in analyzing crimes in general, rather than focusing on a specific crime category.

Preliminary findings allude to the possibility that crime rates in Toronto may deviate from the national trend, warranting a more profound exploration into the intricate dynamics of criminal activities. Acknowledging potential limitations within the dataset, including the influence of racial bias as evidenced in previous studies on crime in Toronto (StatCan 2021), underscores the importance of interpreting our findings within a broader socio-economic and cultural context. Understanding these complexities becomes imperative for informed policymaking, strategic law enforcement, and the development of community engagement initiatives.

Beyond statistical analysis, this research delves into the social fabric of Toronto's neighborhoods, offering critical insights into the challenges and opportunities for intervention. Section 2: Data of the paper will meticulously explore the dataset's characteristics, employing a variety of tables and figures to construct a comprehensive narrative around crime occurrences in Toronto during 2020. Section 3: Results will delve into the intricate relationships between crime rates, income levels, and population figures, shedding light on correlations and patterns that emerge from the data. It was found that crimes and income are inversely correlated, indicating that as income levels decrease, crime rates tend to increase. Furthermore, there is a positive correlation between population size and crime rates, suggesting that as the population increases, so

does the incidence of crime. Section 4: Discussion will intricately discuss the implications of these findings within the broader socio-economic context of crime in Toronto, addressing the multifaceted nature of these influences. Finally, the paper will conclude in Section 5: Conclusion, summarizing key discoveries and proposing avenues for future research in understanding and addressing the complex dynamics of crime in Toronto's neighborhoods.

2 Data

This paper utilises datasets retrieved from Open Data Toronto Portal through the library `opendatatoronto` (Gelfand 2022). Two different data sources were leveraged to analyze the relationship of crimes committed across Toronto with income and population: `Neighbourhood Crime Rates` (Toronto 2024) for crime data and `Neighbourhood Profiles, 2021 (158 Model)` (Toronto 2023) for relevant neighbourhood information. Data was simulated, cleaned , tested and analyzed using the open source statistically programming language R (R Core Team 2020), using functionalities from `tidyverse` (Wickham et al. 2019), `ggplot2` (Wickham 2016), `dplyr` (Wickham et al. 2022), `readr` (Wickham, Hester, and Bryan 2022), `tibble` (Müller and Wickham 2022), `janitor` (Firke 2021), `kableExtra` (Zhu 2021) and `knitr` (Xie 2014). Details and examination of the dataset context, variables creation and extraction processes are discussed in the subsections below.

2.1 Number of Crimes Committed in 2020 in Toronto Neighbourhoods

To comprehensively explore crime dynamics in Toronto during the year 2020, the Neighbourhood Crime Rates dataset from the Toronto Open Data portal served as a foundational resource (Gelfand 2022). This dataset, compiled by the Toronto Police Services and last updated on January 11, 2024, encompasses 158 neighborhoods within the Greater Toronto Area. It provides essential information, including the count of various crime types recorded annually between 2014 and 2023 (e.g., Homicide_2020). The crime categories include assault, auto theft, breaking and entering, homicide, robbery, and theft over. However, we only focus on relevant crime data for 2020. The sum of all these crimes committed in 2020 is also included in the cleaned data as the total number of crimes committed in each neighborhood, named `number_of_crimes`.

To ensure a focused examination, the dataset was refined to include specific variables, such as the neighborhood IDs (variable named `hood_id`), neighborhood names (variable named `neighbourhood`), and the total number of crimes for each neighborhood (variable named `number_of_crimes`) in 2020 (see Table 1). This focused approach aimed to elucidate the crime landscape within the selected timeframe. Additionally, considerations were made to enhance data visualization by incorporating neighborhood IDs for readability on plot axes. While the raw data included average per crime and crime rate variables, a decision was made

to prioritize a total count of all crimes. This strategic choice facilitates a more comprehensive assessment, offering a holistic perspective on crime in Toronto neighborhoods during the specified year.

Table 1: Sample of Cleaned 2020 Toronto Neighbourhoods Crime Data

Neighbourhood Name	Hood ID	Number of Crimes Committed
South Eglinton-Davisville	174	257
North Toronto	173	167
Dovercourt Village	172	207
Junction-Wallace Emerson	171	442
Yonge-Bay Corridor	170	797

2.2 Neighbourhood Profiles, 2021 (158 Model)

To capture a comprehensive snapshot of the demographic, social, and economic characteristics of Toronto's neighborhoods, the City of Toronto Neighbourhood Profiles derived from the 2021 Census of Population played a pivotal role in our analysis. Conducted every five years, the Census collects a wide array of data, including age, sex, families, households, language, immigration, ethnocultural diversity, aboriginal peoples, housing, education, income, and labor. For our study, the Neighbourhood Profiles provided essential insights, offering a nuanced understanding of the people and households within each of the 158 neighborhoods in the City of Toronto.

The dataset is downloadable as a Microsoft Excel file and the `hd2021_census_profile` tab within the dataset was utilized to extract pertinent information for the purpose of this paper. Raw data can transposed and cleaned to extract relevant information such as population (variable named `population`) and average total household income (variable named `income`) for each neighbourhood, identified by variable named `hood_id` in 2020 (see Table 2). This information is pivotal in elucidating the economic and demographic landscape of Toronto's neighborhoods, providing a valuable complement to the crime data derived from the Neighbourhood Crime Rates dataset.

Table 2: Sample of Cleaned 2020 Toronto Neighbourhood Profile Data

Hood ID	Popluation	Income
1	33300	104500
2	31345	86200
3	9850	101300
4	10375	90000
5	9355	94600

2.3 Combining Crimes, Income and Population Data

By employing the merge function, we have seamlessly combined cleaned `Neighbourhood Crime Rates` data with cleaned `Toronto Neighbourhood Profiles` data , creating a unified dataset featuring variables including `hood_id`, `neighbourhood`, `number_of_crimes`, `population`, and `income`.

As illustrated in Table 3, our combined dataset comprises several key variables, each providing crucial insights into the socio-economic and crime landscape of Toronto's neighborhoods. The `hood_id` variable serves as a unique identifier for each neighborhood, facilitating a granular examination of the data. The `neighbourhood` variable delineates the specific neighborhood associated with each observation, enhancing the geographical context. `number_of_crimes` is a numerical variable that quantifies the incidence of crimes within each neighborhood, offering a quantitative measure of criminal activity. Furthermore, `population` and `income` variables contribute additional dimensions to our analysis, shedding light on the demographic and economic characteristics of the neighborhoods.

These variables collectively enable a comprehensive exploration of the relationships between crime rates, population, and income in Toronto during the year 2020. For example, the first observation in Table 3 indicates that the neighborhood with ID 153, named Avondale, experienced 163 crimes in 2020. Additionally, the population of this neighborhood in 2020 was 13,790, and the average total household income was $102,700. These data points can be analyzed together in order to tell us a compelling story about the relationship between crimes, income and population across Toronto neighbourhoods.

Table 3: Sample of Combined Toronto Neighbourhood Crime, Income and Population Data

Hood ID	Neighbourhood Name	Number of Crimes	Population	Income
153	Avondale	163	13790	102700
154	Oakdale-Beverley Heights	576	21420	89600
155	Downsview	283	17970	97600
156	Bendale-Glen Andrew	256	18685	86400
169	Bay-Cloverhill	230	16670	94000
170	Yonge-Bay Corridor	797	12645	99800
171	Junction-Wallace Emerson	442	23180	103700
172	Dovercourt Village	207	12380	111500
173	North Toronto	167	15885	85400
174	South Eglinton-Davisville	257	22735	103500

3 Results

3.1 Toronto Neighbourhoods Crimes Committed in 2020 Statistics

In the year 2020, the City of Toronto witnessed a total of 42,328 crimes, reflecting the complex dynamics of urban safety. On average, approximately 267 crimes were committed per neighbourhood, demonstrating the variability in crime rates across Toronto's diverse communities. The relatively high standard deviation of 193 emphasizes the dispersion of crime occurrences, highlighting the diverse safety landscapes within the city. Neighbourhoods experiencing the highest number of crimes were Moss Park (Neighbourhood 73) with 1228 incidents, West Humber-Clairville (Neighbourhood 1) with 1175, and Downtown Yonge East (Neighbourhood 168) with 942 crimes. These areas grapple with distinct challenges, contributing to elevated crime rates. Conversely, neighbourhoods with the lowest crime rates include Kingsway South (Neighbourhood 15) with 61 incidents, Lambton Baby Point (Neighbourhood 114) with 71, and Runnymede-Bloor West Village (Neighbourhood 89) with 82 crimes. These insights (see Figure 1) underscore the heterogeneous nature of crime distribution in Toronto, prompting a nuanced exploration of contributing factors and the need for targeted interventions to enhance community safety.

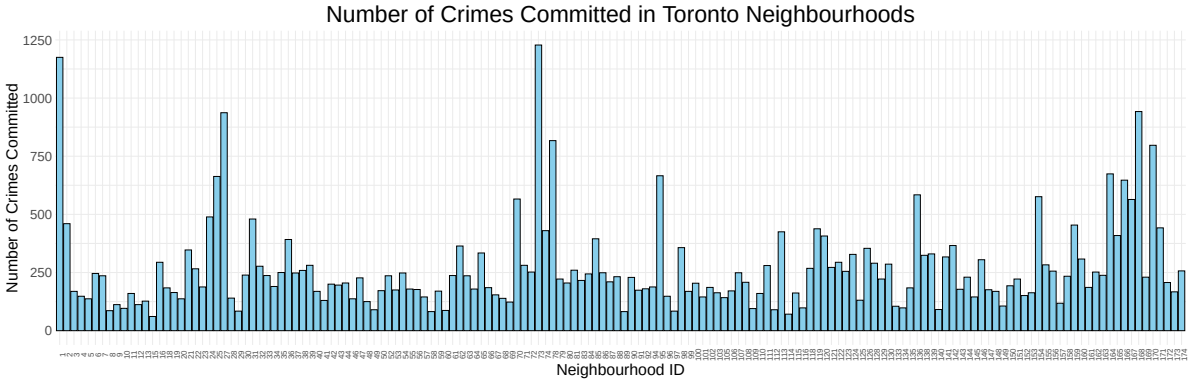

Figure 1: Number of Crimes Committed in Toronto Neighbourhoods

3.2 Toronto Neighbourhood Profile Based on 2021 Census Data

There are 158 neighbourhoods in the City of Toronto. Based 2021 census data, the average population per neighbourhood is 17476, with standard deviation of 6131. The neighbourhoods with the highest population are: West Humber-Clairville (Neighbourhood 1) at 33,300, Mount Olive-Silverstone-Jamestown (Neighbourhood 2) at 31,345, and St Lawrence-East Bayfront-The Islands (Neighbourhood 166) at 33,300. The neighbourhoods with the lowest population are: Beechborough-Greenbrook (Neighbourhood 112) at 6,260, University (Neighbourhood 79) at 6,435, and Blake-Jones (Neighbourhood 69) at 7,475.

Average household income level per neighbourhood is $125,263, with standard deviation of $59,930. The neighbourhoods with the highest household income are: Bridle Path-Sunnybrook-York Mills (Neighbourhood 41) at $519,500, Lawrence Park South (Neighbourhood 103) at $365,600, and Rosedale-Moore Park (Neighbourhood 98) at $334,000. The neighbourhoods with the lowest household income are: Oakridge (Neighbourhood 121) at $72,200, North St.James Town (Neighbourhood 74) at $74,100, and Black Creek (Neighbourhood 24) at $74,700.

3.3 Relationship Between Crimes Committed and Neighbourhood Population

The relationship between crimes committed and neighbourhood population in Toronto is explored in Figure 2. The scatter plot visually represents this correlation, and a linear regression line indicates the direction of the relationship. The correlation coefficient between the number of crimes and population is approximately 0.564, suggesting a moderately positive correlation. This statistical measure quantifies the tendency for neighbourhoods with larger populations to experience higher crime rates. As an example, consider West Humber-Clairville (Neighbourhood 1), which has the highest population at 33,300. This neighbourhood also ranks second in the total number of crimes with 1175 incidents in 2020, supporting the positive correlation hypothesis. The summary statistics reveal that, on average, Toronto neighbourhoods with larger populations tend to have higher crime rates, contributing to a nuanced understanding of urban safety dynamics.

3.4 Relationship Between Crimes Committed and Neighbourhood Income

The statistical exploration of the relationship between crimes committed and neighbourhood income is depicted in Figure 3. The scatter plot visualizes this association, while the linear regression line and a correlation coefficient of -0.217 indicate a weak negative correlation. This suggests that, on average, neighbourhoods with higher household incomes tend to have lower crime rates. The negative sign of the correlation coefficient underscores the inverse relationship between these variables. Bridle Path-Sunnybrook-York Mills (Neighbourhood 41), Lawrence Park South (Neighbourhood 103), and Rosedale-Moore Park (Neighbourhood 98) are examples of neighbourhoods with the highest household incomes, and they also exhibit relatively lower crime rates. In contrast, neighbourhoods with lower household incomes, such as Oakridge (Neighbourhood 121), North St. James Town (Neighbourhood 74), and Black Creek (Neighbourhood 24), tend to experience higher crime rates. While the correlation is statistically weak, the negative coefficient suggests that higher household incomes are associated with lower crime rates, and vice versa.

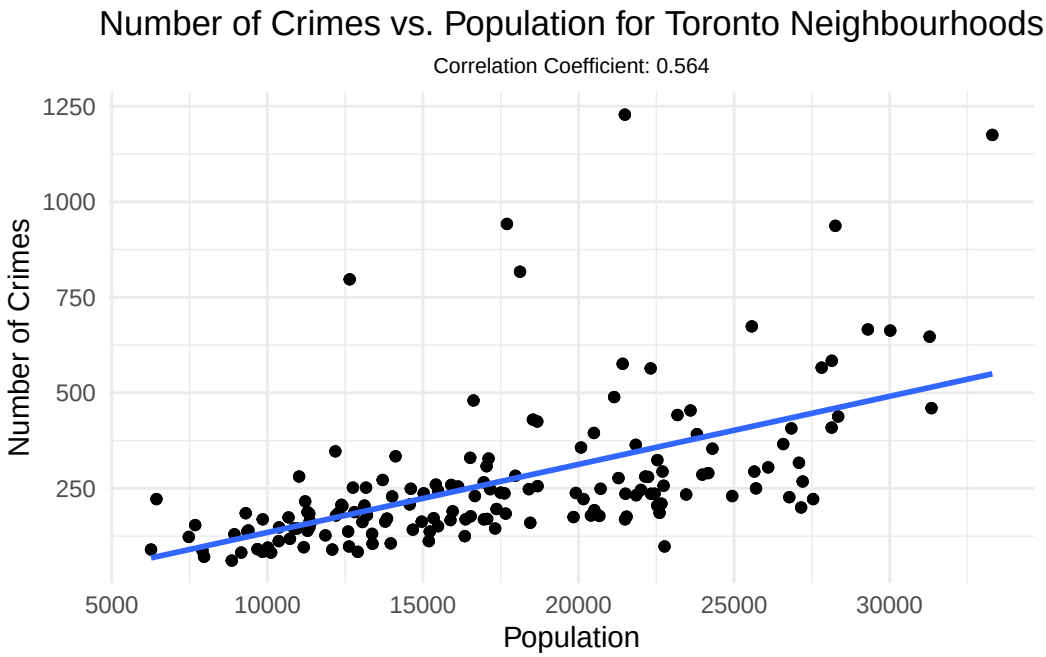

Figure 2: Correlation between Number of Crimes and Population

4 Discussion

5 Conclusion

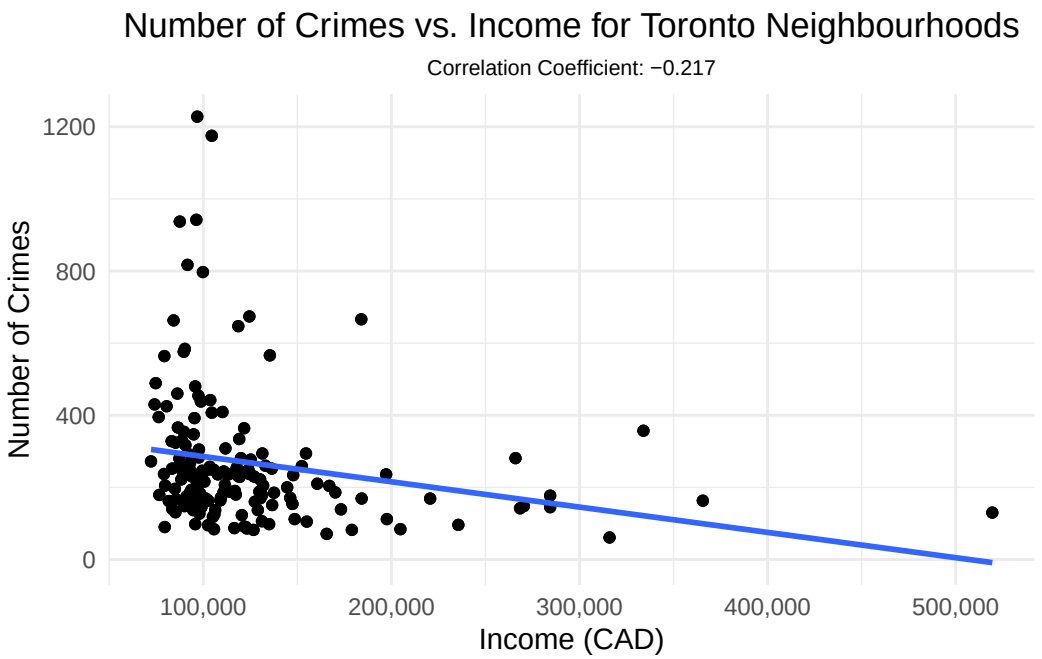

Figure 3: Correlation between Number of Crimes and Income

References

Firke, Sam. 2021. *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.

Müller, Kirill, and Hadley Wickham. 2022. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

StatCan. 2021. "Guide to the Census of Population, 2021 / Chapter 9 – Data Quality Evaluation." https://www12.statcan.gc.ca/census-recensement/2021/ref/98-304/2021001/chap9-eng.cfm.

Statista. 2024. "Crime in Canada." https://www.statista.com/topics/2814/crime-in-canada/#topicOverview.

Toronto, Open Data. 2023. "Neighbourhood Profiles." https://open.toronto.ca/dataset/neighbourhood-profiles/.

———. 2024. "Neighbourhood Crime Rates." https://open.toronto.ca/dataset/neighbourhood-crime-rates/.

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, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2022. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

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