Population of Neighborhoods Has a Limited Impact on Crime Rates*

An Analysis of Toronto Crime Data by Neighborhood

Mengze Zhang

4 February 2023

The report analyzed crime rate data in Toronto neighborhoods obtained from the Toronto Police Service Annual Statistical Report between 2017 and 2021. The findings showed that crime rates in different categories showed varying trends over the years, but crime rates varied by neighborhood size without a direct relationship to population. This report concludes that the population of neighborhoods has a limited impact on crime rates and highlights the importance of considering multiple factors, such as socio-economic status and policing practices, in understanding crime patterns. This study provides valuable insights for policymakers in allocating policing resources and designing crime prevention initiatives, thereby improving the safety and well-being of communities.

1. Introduction

The relationship between population and crime rates has been an area of interest for scholars and policymakers for many years. One popular argument is that population density and population size can have an impact on crime rates, with higher population density and larger populations often associated with higher crime rates (Karstedt 2013).However, the relationship between population and crime is complex and multi-faceted, with many other factors, such as socio-economic status, cultural values, and policing practices, also playing important roles (Wortley and Mazerolle 2008).

In recent years, there has been a growing interest in using crime data and analysis to inform policy and practice. Crime data is increasingly being made available to the public through open data initiatives, and there is a growing body of literature on the use of crime data for policy purposes. In this context, the relationship between population and crime is of particular

*Code and data are available at github.com/MissyZhang/Crime_Data_by_Neighbourhood.

interest, as it has important implications for policy and practice, including the allocation of policing resources and the design of crime prevention initiatives.

To explore this relationship, this report analyses crime rate data in Toronto neighborhoods obtained from the Toronto Police Service Annual Statistical Report (ASR) (Toronto Police Service 2022). The dataset analyzed in this report contains crime rate data for 8 different types of crime between the years 2017 and 2021 for each neighborhood in Toronto.

This report will first consider the year-on-year changes in crime rates for different crime categories. The findings showed that the crime rate in different categories showed varying trends over the years. Assault had the highest rate while homicide had the lowest and remained stable. Theft of motorcycles and breaking and entering showed an initial increase followed by a decrease, while auto theft showed a consistent increase. The second aspect of the data analyzed was the relationship between population and crime rate. Neighborhoods were categorized based on their population, and the findings showed that crime rates varied by neighborhood size but did not have a direct relationship with population. Finally, it will discuss the limits of the crime rate data, in terms of bias and in terms of measuring the societal value of the data.

2. Data

2.1 Data Source

This report utilizes data on crime rates in Toronto neighborhoods obtained from the Toronto Police Service Annual Statistical Report (ASR) (Toronto Police Service 2022). The ASR is a comprehensive overview of police related statistics that has been openly available to the public since 2019 (Toronto Police Service 2022). The neighborhood crime rate dataset analyzed in this report was obtained in CSV format from the City of Toronto Open Data Portal using the R package `opendatatoronto` (Gelfand 2020). The dataset was last updated on May 18th, 2022.

2.2 Data Characteristics

The Neighborhood Crime Rates dataset contains aggregated data on 8 types of crime in each Toronto neighborhood between the years 2014 and 2020. There are 16920 observations in the dataset and 7 attributes: index, object ID, hood name, hood ID, 2021 population projection, count of each crime type in each year, crime rate of each crime type in each year. The first 4 attributes are identifiers that were removed prior to analysis. During analysis, an additional attribute was created to categorize neighborhoods based on population by dividing the population into quartiles and labeling each neighborhood accordingly (e.g. “small”, “medium”, “large”, “giant”). In the dataset, crime counts are aggregated data for Assault, Auto Theft, Break and Enter, Robbery, Theft Over, Homicide, Shootings and Theft from Motor Vehicle of each neighborhood in the time period, which I deleted since I am interested in the crime rate. Finally, the crime rate is calculated as the crime count per 100,000 people by neighborhood based on each year's Projected Population, which I selected data from 2017 to 2021 to cut the data short and clean while also remaining rigorous for reasonable analysis.

It is worth noting that in the data collection process by Toronto Police Service, the location of crime occurrences has been deliberately offset to the nearest road intersection node to protect the privacy of parties involved in the occurrence (Toronto Police Service 2022). Due to the offset of occurrence location, the numbers by neighborhood may not reflect the exact count of occurrences reported within these geographies, which may lead to incorrect conclusions based on the crime data and makes it difficult to establish meaningful relationships between crime rate and population density.

2.3 Data Analysis

Analysis for this project uses the R statistical programming language (R Core Team 2022), as well as `tidyverse` (Wickham et al. 2019), `tidyr` (Wickham 2022) and `dplyr` (Wickham et al. 2021) programming packages. Because the data is managed using R Projects, `here` is used to reference file locations (Müller 2020). Figures and tables are created with `ggplot2` (Wickham 2016) and `kableextra` (Zhu 2020), and a new data frame is created with `tibble` (Müller and Wickham 2022). The package `knitr` (Xie 2021) is used to generate a PDF report. The package `palettetown` (Lucas 2016) is used to create color palettes for figures, and the packages `patchwork` (Pedersen 2022) and `ggpubr` (Kassambara 2022) are used to patch graphs together.

The first aspect of the data I investigated was the year-on-year changes in crime rates for different crime categories. Figure 1 shows changes in average crime rates for 8 different crime types from 2017 to 2021. It is clear that assault has the highest crime rate each year, fluctuating around 600, while homicide has the lowest and remains stable around 10. Both the motorcycle theft rate and break and enter rate have a trend of rising first and then falling, with turning points in 2020 and 2019 respectively. The robbery rate decreases year by year from 2017 to 2021, while auto theft rate sees an increase and even exceeded break and enter rate in 2021. Similar to homicide rate, both theft rate and shooting rate remain relatively low and stable over the time period, with theft rate (around 50) slightly higher than shooting rate (around 30) each year.

In addition to general trends, it is also possible to observe some specific patterns in the data. For example, auto theft has seen a sharp increase in the mean crime rate, with a 33% increase from 2017 to 2021. This suggests a growing problem with vehicle theft and highlights the need for increased attention and resources to address this issue. On the other hand, break & enter has seen a significant decrease in the mean crime rate from 2019 to 2021, with a 21% decrease.

This could indicate that efforts to prevent and address break-ins have been successful, or that crime has shifted to other areas or types of crime.

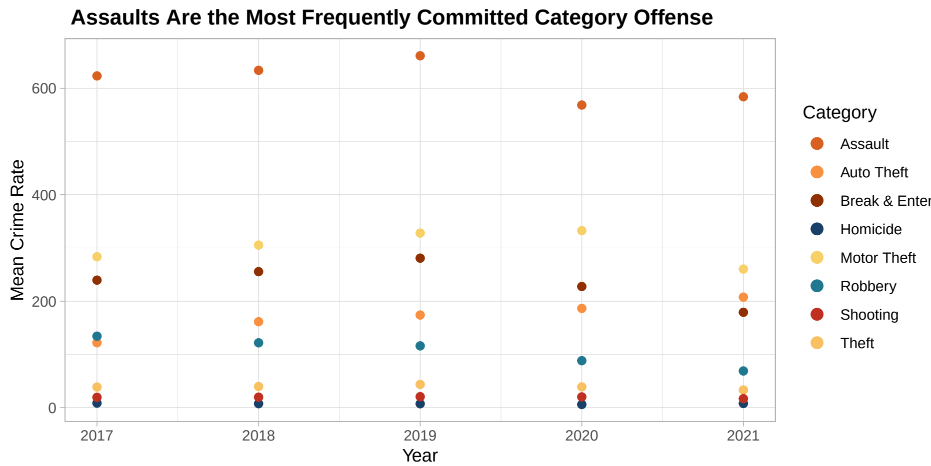

Figure 1: Crime rate in Toronto by category and year

The next aspect of the data I analyzed is the relationship between population and crime rate. I categorized different neighborhoods based on their population for basic analysis. Four categories of neighborhoods are listed as follows:

- Small neighborhood: Has a population of fewer than 10,000
- Medium neighborhood: Has a population between 10,001 and 15,000
- Large neighborhood: Has a population between 15,001 and 2,0000
- Giant neighborhood: Has a population of more than 20,000

Table 1 shows the mean crime rate for each crime category by neighborhood size in 2021. The highest and lowest mean crime rates for each category in the table seem to have a relationship with neighborhood size. The crime rates for assault, break & enter, robbery, and theft in medium-sized neighborhoods are the lowest among 4 neighborhood types, which are 449.48, 163.58, 49.31, and 30.84 respectively. However, smaller neighborhoods have the highest crime rates for assault, break & enter, robbery, homicide, and shooting. In giant-sized neighborhoods, crime rates for theft and motor theft are the highest while the homicide rate and shooting rate are the lowest.

It is also noteworthy that crime rates in different neighborhoods vary significantly depending on the type of crime. For example, the crime rates for theft, homicide, shooting, and motor theft are relatively constant across all neighborhood sizes. In contrast, the crime rates for assault and motor theft vary widely between neighborhood sizes, with a difference between the highest and lowest values of 102.45 and 88.88 respectively.

Table 1: Mean crime rate by neighborhood size in 2021

Neighborhood Size	Small	Medium	Large	Giant
Crime Category				
Assault	641.53	499.48	514.11	666.62
Auto Theft	136.01	224.89	201.09	208.64
Break & Enter	223.34	163.58	173.26	186.77
Robbery	119.00	49.31	65.14	77.99
Theft	31.37	30.84	30.87	35.77
Homicide	27.69	11.80	7.35	5.91
Shooting	31.09	20.30	13.69	16.11
Motor Theft	238.48	247.92	264.42	268.32

Based on the results of Table 1, we could also observe that there is a strict linear relationship between neighborhood size and crime rates of homicide, shooting, and motor theft respectively. Therefore, we used complete data of 2021 population and crime rates of the three above to examine if there is a linear relationship.

Figure 2 presents the correlation between population size and the crime rates of homicide, shooting, and motor theft that occurred in 2021. The graph showcases the use of statistical methods to understand the relationship between these two variables.

The p-value, which measures the probability of observing a relationship as strong or stronger than the one observed if the null hypothesis is true (no relationship between population size and crime rate), is used to determine the statistical significance of the relationship. A p-value of less than 0.05 is considered statistically significant, meaning that the relationship is unlikely to be due to chance. The R-value, also known as Pearson's correlation coefficient, measures the strength and direction of the relationship between two variables.

For the relationship between population size and homicide rate, the p-value is less than 0.05, which means that the relationship is statistically significant. The negative value of r, which is the correlation coefficient, is -0.44. This implies that as population size increases, the homicide rate decreases. However, the strength of this relationship is not very strong as the value of r is close to 0.

On the other hand, the relationship between population size and shooting rate is not statistically significant, as the p-value is 0.141 which is greater than 0.05. This suggests that the observed relationship between population size and shooting rate could be due to random chance and not a true relationship. Similarly, the p-value of 0.732 for the relationship between population size and motor theft rate is also greater than 0.05, which implies that population size does not have a significant effect on the motor theft rate.

However, the results are based on data from one single year with many missing values, so it is possible that the results may not accurately reflect the true relationship between population and crime rates. In addition, as stated before, the offset of crime locations can also have a significant impact on the results. For example, if two areas have the same population but one is closer to a road intersection, it will appear to have a higher crime rate due to the increased reporting of incidents in that area. This can lead to misleading conclusions and result in biased or unethical decisions and policies being made based on faulty data.

In summary, there is no clear relationship between population size and crime rates, and further research is required to establish a clearer understanding of the relationship.

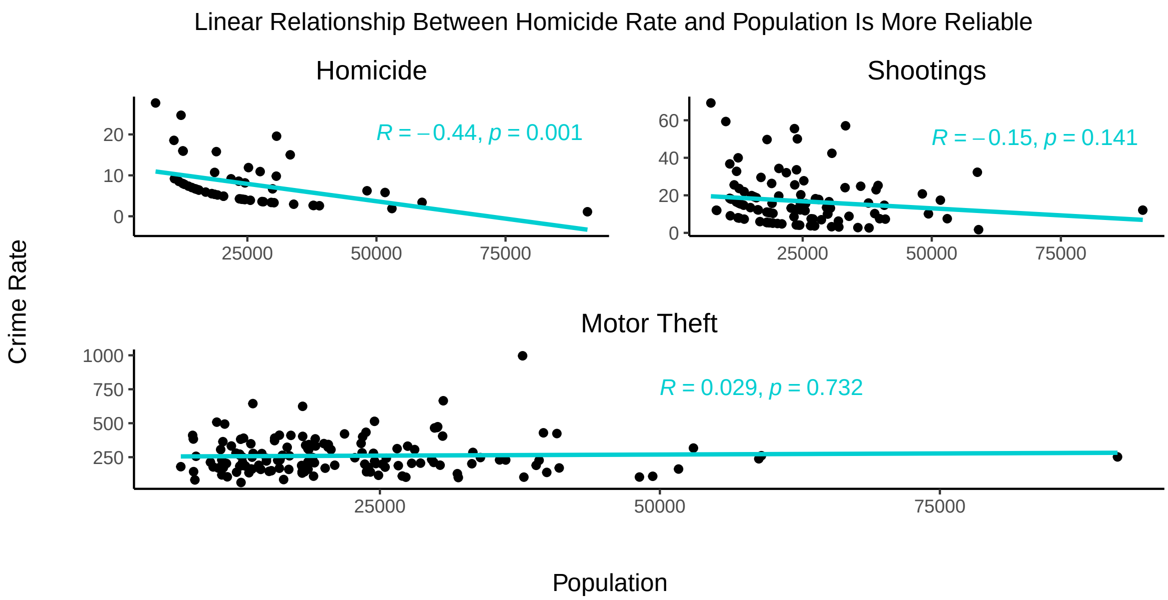

Figure 2: Relationship between crime rate and population

Reference

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Karstedt, Susanne. 2013. *Crime and Globalization*.

Kassambara, Alboukadel. 2022. *Ggpubr: 'Ggplot2' Based Publication Ready Plots*. https://CRAN.R-project.org/package=ggpubr.

Lucas, Tim. 2016. *Palettetown: Use Pokemon Inspired Colour Palettes*. https://CRAN.R-project.org/package=palettetown.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

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

Pedersen, Thomas Lin. 2022. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

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 Police Service. 2022. *Toronto Police Service Public Safety Data Portal*. https://data.torontopolice.on.ca/pages/annualstatisticalreport.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2022. *Tidyr: Tidy Messy Data*. https://tidyr.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. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wortley, Richard, and Lorraine Green Mazerolle. 2008. *Environmental Criminology and Crime Analysis*.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2020. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.