Understanding the relationship between crime and population density in Toronto neighbourhoods

An investigation into how crime count obscures the true scope of criminal activities in communites

Ahnaf Alam

January 24, 2024

Abstract

Criminal activities have gone up noticeably in Toronto and there is a prevailing call on elected officials to allocate additional police presence and budgetary resources to areas that have higher population density, as criminal activities tend to be higher in those areas. However, implementing such policies may lead to misallocation of resources. I analyze the relationship between crime rate and population differences across Toronto neighbourhoods and find that it is true that neighbourhoods with larger populations tend to have higher crime activity. However, the crime rate per 100,000 people is identical across Toronto neighbourhoods. This is a policy-relevant finding, as elected officials must take a more targeted approach to policing and not drastically alter the level of policing across neighbourhoods, as that may result in unforeseen repercussions.

Table of contents

1 Introduction [1]

Law enforcement is a prominent and widely debated issue across urban centers in Canada, particularly given the recent surge in population attributed to immigration. Despite studies indicating that immigration tends to lower crime rates (Lee and Martinez 2009), public perception often associates an uptick in criminal activities with the influx of newcomers. The escalation in immigration has led to heightened population density in various neighborhoods across Toronto, prompting calls for increased police presence in areas identified as having elevated criminal activity. Notwithstanding, prior research suggests that a more restrained approach to proactive policing can result in crime reduction (Sullivan and O'Keeffe 2017).

Responding to the concerns voiced by Toronto residents, public officials approve an augmented policing budget for the upcoming fiscal year (Lautenschlager 2023). The effectiveness of this budgetary increase depends on the strategic allocation of the additional resources. Furthermore, there is a mounting demand for public officials to allocate policing resources in accordance with the population density, as crime tends to be more

[1] For replicability and other purposes, please check https://github.com/AhnafAlam1/Analyzing-the-relationship-between-changes-in-population-and-crime-rates-in-Toronto

prevalent in densely populated areas. This study aims to scrutinize the validity of the claim that crime rates are inherently higher in regions with increased population density, exploring potential policy implications for policing decisions.

Research indicates that criminal activities are often influenced by socio-economic disparities (Glaeser and Sacerdote 1999). However, the correlation between these criminal activities and the population density of neighborhoods remains less clear. This paper clarifies this relationship, revealing that while certain criminal activities such as shootings, assault, auto thefts, and robberies are more prevalent in areas with higher population density, the crime rate, measured as crimes committed per 100,000 people, remains consistent across Toronto neighborhoods, regardless of population density. This finding holds significant policy implications, as reallocating resources based solely on higher crime numbers may yield unfavorable community outcomes. An excessive police presence can foster tension and distrust within communities (Rouse 1985). Additionally, a disproportionate increase in police deployment achieved by reallocating resources may inadvertently elevate crime rates in the areas that experience a reduction in policing resources.

In the subsequent sections, this paper will elaborate on the dataset, providing rationale for the inclusion of specific variables in the Data section. The ensuing Discussion and Limitation section will feature graphical representations, along with analysis. Finally, the paper concludes by summarizing the findings and proposing potential modifications for subsequent iterations.

2 Data

The data for this analysis is extracted from Open Data Toronto using the opendatatoronto library (Gelfand 2022). The dataset, Neighbourhood Crime Rates ("Neighbourhood Crime Rates" 2024), is utilized for this project. Furthermore, R (R Core Team 2023), a statistical programming language, is exclusively used in this paper. For the purposes of cleaning and preparing the data, the libraries tidyverse (Wickham et al. 2019), janitor (Firke 2023), knitr (Xie 2023), and kableExtra (Zhu 2021) are used. For the graphs, ggplot2 (Wickham 2016) and ggpubr (Kassambara 2023) is used.

Data used in this paper come from opendatatoronto; however, raw data is compiled by Toronto Police Services. The raw data consists of observations for 158 neighborhoods across Toronto. For each neighborhood, the dataset contains information on its population, geometry, count of assaults, robberies, shootings, auto thefts, breaking and entering, theft from motor vehicles, homicides, bike thefts, and thefts from 2014 to 2023, and the rate of all corresponding crimes. The crime rate is calculated by normalizing each crime committed per 100,000 people. Population numbers reflect only the permanent residents of an area and exclude temporary populations such as commuters, business people, and non-temporary students. Lastly, the variable geometry provides longitude and latitude coordinates for each neighborhood.

A table of data is provided below (see Table 1). The raw data is cleaned using the clean() function, and only 11 columns are selected going forward using select(). For each of the 158 Toronto neighborhoods, the summary statistics table includes information on their corresponding ID, population in 2023, count of assaults, auto thefts, shootings, robberies, and their rates per 100,000 people. The decision to omit the rest of the crime-related variables was taken as these crimes are often defined to be petty crimes, and petty crimes like bike theft and general theft are severely under-reported (Tarling and Morris 2010); as a result, the number provided to the police may not reflect the actual count. However, the criminal activities selected in summary statistics are defined as serious crimes by the Canadian justice system ("Appendix d - Most Serious Charges List_2022" 2022), and serious crimes are reported more accurately to the police.

The variable "assault" is defined by Toronto Police Services as the act, attempt, or threat of applying force to another person (Services 2019). "Robbery" is defined to be the act of taking property from another person by force or intimidation. Similarly, "auto theft" is the process of taking possession of another person's vehicle. Lastly, "shooting" identifies incidents where an individual is shot or shot at.

An alternative approach for this paper could have involved utilizing a comparable dataset, comprising neighborhoods with aggregate crime counts and crime rates, without the disaggregation of crime types. However, this methodology is not used due to its potential lack of fairness in reflecting the true extent of crime numbers and rates. As previously highlighted, certain crimes are prone to under-reporting, introducing a risk of the statistical values being underestimated compared to the actual occurrences.

Additionally, the summary table intentionally excludes crime data from prior years due to the absence of corresponding population statistics for those years, rendering such data redundant for the purposes of our analysis. The data cleaning process is finalized through the renaming of variables to enhance readability. Following these adjustments, "Pop2023" denotes the population of each neighborhood in 2023, "Assault" signifies the count of assault crimes in the year 2023, and "A_Rate" represents the assault rate for the same year. This renaming convention extends to all other crime-related variables. It is important to note that Table 1 comprises observations for only five randomly selected neighborhoods out of the total 158. This deliberate choice for Table 1 is aimed at offering a snapshot of the data post-cleaning, with the inclusion of information on all 158 neighborhoods deemed unnecessary for the intended purpose.

3 Discussion and Limitations

Figure 1 and Figure 2 consist of four grids, each representing a specific relationship between population and the variable of interest. The x-axis

Table 1: Table displaying count of crime and rates of crimes across select five Toronto neighbourhoods for 2023

ID	Neighbourhood	Pop2023	Assault	A_Rate	Autotheft	AT_Rate	Shooting	S_Rate	Robbery	R_Rate
3	Dovercourt Village	13837	104	751.6	28	202.4	1	7.2	21	151.8
4	Junction-Wallace Emerson	26240	229	872.7	34	129.6	2	7.6	18	68.6
5	Yonge-Bay Corridor	14731	543	3686.1	46	312.3	2	13.6	62	420.9
6	Bay-Cloverhill	19055	142	745.2	17	89.2	1	5.2	20	105.0
7	Bendale-Glen Andrew	20773	237	1140.9	144	693.2	1	4.8	55	264.8

of all the graphs represents the population count of Toronto's neighborhoods. In Figure 1, the y-axis represents the number of criminal activities: shooting, assault, auto theft, and robbery. In contrast, in Figure 2, the y-axis represents crime rates.

Both graphs feature a least squares line, which serves as the optimal fit for a given set of dependent and independent values ("Linear Regression Calculator" 2024). In the context of this paper, the independent variable is the 2023 population statistics, while the dependent variables differ across various crimes and their respective rates. The primary purpose of the line is to highlight potential trends within the data that might not be apparent.

In Figure 1, the least squares line has a positive slope, which applies to all four grids, suggesting that as the population increases on the x-axis, we see an increase in the count of crimes applicable to shooting, assault, robbery, and auto theft. This confirms the hypothesis that a neighborhood with a higher population tends to see more crime. However, if we analyze Figure 2, crime rates are constant in all four grids. This proves that the count of criminal activity may be higher; however, the crime rate committed per 100,000 Toronto residents is comparatively similar, regardless of whether a neighborhood is large, like Junction-Wallace Emerson, or relatively small, like Dovercourt Village.

Analyzing Figure 1, we find a graph featuring a neighborhood outlier—a statistical observation significantly deviating from others ("7.1.6.what

Are Outliers in the Data?" 2023). Specifically, the auto theft graph highlights a neighborhood with close to 40,000 people experiencing auto thefts surpassing 400, reaching nearly 800, while every other neighborhood had fewer than 300 car thefts in 2023. This outlier may indicate policing differences or neighborhood-specific issues facilitating car theft.

Eliminating the outlier could enhance our understanding of typical car theft patterns in Toronto neighborhoods. However, lacking sufficient information, we refrain from dismissing it in this research. The next iteration will further investigate these obscure neighborhoods to understand why certain areas exhibit exceptionally high crime rates. Importantly, the current graphs don't reveal whether the auto theft outlier neighborhood is atypical for all other crimes or if there's an underlying factor making it prone to criminal activity- a consideration for future iterations.

This finding holds significant policy implications for the Toronto Police, as an ill-considered allocation of resources based solely on apparent crime counts may inadvertently erode community trust, contrary to the objectives of law enforcement (Rouse 1985). If the police opt to redistribute resources from areas with lower reported crime to those with higher figures, the potential consequence may be an increase in criminal activity in less densely populated regions, stemming from a diminished law enforcement presence in those areas. Striking a balance between addressing areas with high crime counts and maintaining a consistent law enforcement presence in less-densely populated communities is essential for achieving effective and community-oriented policing.

It is crucial to acknowledge a primary limitation in this discussion, stemming from the absence of policing information in the incorporated dataset. The paper assumes an equitable distribution of policing budget and resources across all neighborhoods due to the lack of specific policing data. In reality, variations may exist in the utilization of police resources in different neighborhoods. To enhance the robustness of future iterations, researchers should incorporate comprehensive policing data, allowing for the control of differences in policing measures across diverse neighborhoods.

This will provide a more accurate understanding of resource allocation and help in formulating policy recommendations for law enforcement strategies.

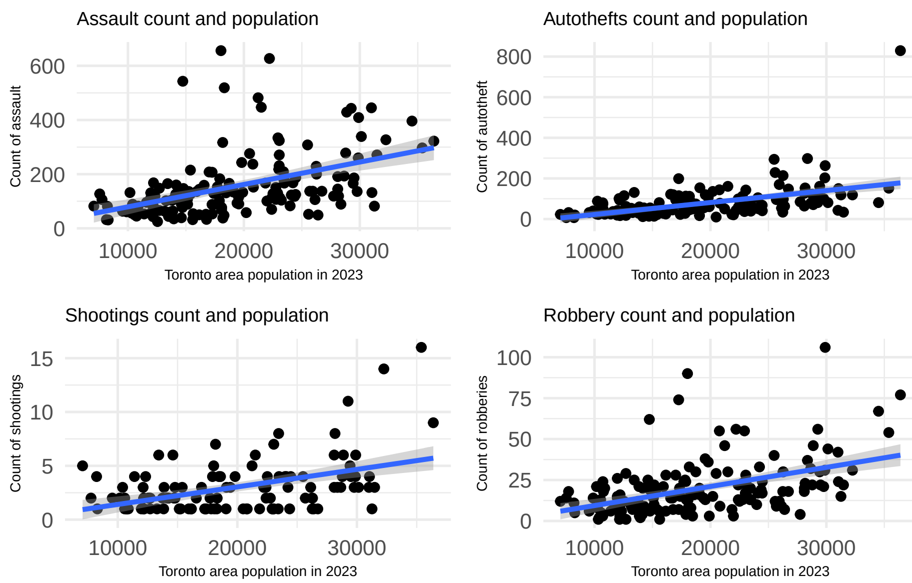

Figure 1: Graph displaying relationship between crime count and population

4 Conclusion

Crime has increased steadily in Toronto over the years, prompting people to ask whether the police should allocate additional resources to areas of higher population density, as they have experienced a surge in criminal activities. However, adding additional resources to a neighborhood that already experiences high crime numbers could have the opposite effect, as researchers have found that sometimes, over-policing can lead to distrust

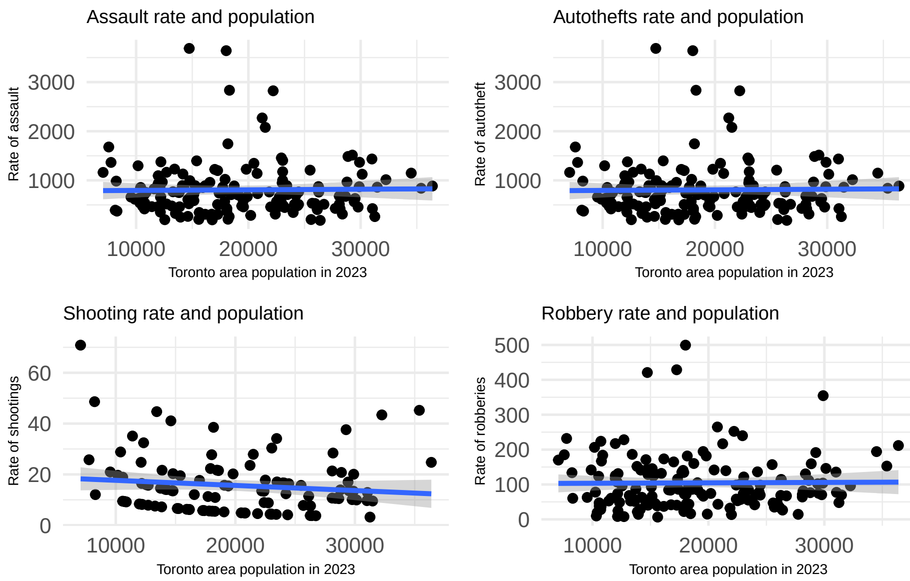

Figure 2: Graph displaying relationship between crime rates and population

in the community between residents and authorities. Therefore, lawmakers must understand whether policing resources need reallocation before implementing a policy.

In this paper, I examine whether the claim that areas with high population density see greater crime holds true. The paper asserts that while the accurate crime count tends to be higher in areas where the population tends to be larger, the crime rates, a variable normalized to crimes committed per 100,000, are constant across all neighborhoods for all types of serious crimes. This suggests that while the crime count may be high in some areas, adding additional police resources to those places may or may not be productive, as the crime rate is the same for almost all communities. More policing could lead to unintended consequences.

The data, however, is limited in that it does not contain information regarding the level of policing, and that stops us from analyzing whether some communities already have a higher-than-expected number of police presence. The next iteration of research should look at the data available regarding policing numbers in the community and incorporate that with the existing data.

Bibliography

"7.1.6.what Are Outliers in the Data?" 2023. *7.1.6. What Are Outliers in the Data?* https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm.

"Appendix d - Most Serious Charges List_2022." 2022. *Appendix D - Most Serious Charges List*. https://www.justice.gc.ca/eng/rp-pr/csj-sjc/crime/rr06_fv3-rr06_vf3/d.html.

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.

Glaeser, Edward L, and Bruce Sacerdote. 1999. "Why Is There More Crime in Cities?" *Journal of Political Economy* 107 (S6): S225–58.

Kassambara, Alboukadel. 2023. *Ggpubr: 'Ggplot2' Based Publication Ready Plots*. https://rpkgs.datanovia.com/ggpubr/.

Lautenschlager, Tyson. 2023. *CBCnews*. CBC/Radio Canada. https://www.cbc.ca/news/canada/toronto/toronto-police-budget-increase-2024-1.7064608.

Lee, Matthew T, and Ramiro Martinez. 2009. "Immigration Reduces Crime: An Emerging Scholarly Consensus." *Immigration, Crime and Justice*, 3–16.

"Linear Regression Calculator." 2024. *Social Science Statistics*. https://www.socscistatistics.com/tests/regression/default.aspx.

"Neighbourhood Crime Rates." 2024. *NEIGHBOURHOOD CRIME RATES*. https://open.toronto.ca/dataset/neighbourhood-crime-rates/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rouse, John Jay. 1985. "The Relationship Between Police Presence and Crime Deterrence." *The Police Journal* 58 (2): 118–31.

Services, Toronto Police. 2019. "Data Definitions and Counting Methodology." 2019. https://data.torontopolice.on.ca/pages/glossary.

Sullivan, Christopher M, and Zachary P O'Keeffe. 2017. "Evidence That Curtailing Proactive Policing Can Reduce Major Crime." *Nature Human Behaviour* 1 (10): 730–37.

Tarling, Roger, and Katie Morris. 2010. "Reporting Crime to the Police." *The British Journal of Criminology* 50 (3): 474–90.

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.

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.