Homicides in Toronto: A Data Analysis*

Joseph Chung

Feb 3, 2023

Abstract

Toronto has experienced a high number of homicides. This paper analyzed opendatatoronto data regarding homicides in Toronto. This paper analyzed the data, including the Occurrence Year, Homicide Type (Shootings, Stabbings, and other types), and Neighborhood, with extra attention. One interesting aspect of the analysis is that the number of gun shootings in Toronto has always been high, and this hasn't changed over the course of 19 years. This paper focuses on neighbourhoods with 20 or more homicides to determine whether or not there is a correlation with the data.

Table of contents

List of Figures

List of Tables

*Code and data are available at: https://github.com/UtopianYoungChung/Homicides-in-Toronto.git

1 Introduction

Canada is regarded as one of the safest and most peaceful places to live in the world. Nonetheless, Toronto's reputation is threatened by the high number of homicides. Although the overall crime rate in Canada has decreased by about 30 percent over the past two decades (Statista Research Department 2021), there is no evidence that homicides have decreased. This fact prompted questions about Toronto and its significance.

This paper used R (R Core Team 2020) to analyze a dataset from opendatatoronto (Gelfand 2020) on homicide occurrences in Toronto to get a better idea of how exactly homicides have differed in Toronto recently and tell a story about these changes. The analyzed dataset contains observations on how many homicides occurred each year in each Neighborhood from 2004-2020. Using tables and useful visualizations like scatter plots and bar graphs, the trends of homicide occurrences in Toronto could be seen. It was discovered that homicide occurrences in Toronto have constantly been high over the years; most of all, gun shootings have been the highest. It was also found that the high number of occurrences of homicide remained true for all areas.

These findings have implications for them. There might be a positive correlation between certain races and income levels, resulting in the concentration of Blacks and other people of colour in low-income neighbourhoods (Gelman, Kiss, and Fagan 2005). This claim might indicates a correlation between income levels and homicide occurrences (Bronner 2020). Nevertheless, the homicides across the city will continuously remain high unless a proper social programs are implemented. The purpose of this paper is to provide the semantic information of the data to help in developing such social programs.

The rest of this paper will be as follows: Section 2 Data goes over the dataset being used and analyzes it with the use of tables and figures to tell a story about homicides in Toronto.

2 Data

The dataset used in this paper was taken from the City of Toronto's open data portal using the open data Toronto package (Gelfand 2020). It is a subset of a larger data collection called The Toronto Police Services Annual Statistical Report. This report contains data based on police-related topics, including various types of crimes and other incidents, and administrative information. The dataset that will be analyzed draws from the "Homicides Occurrence" portion of the annual statistical report. It contains all Toronto homicide incidents between 2004 and 2020. A homicide occurs when a person directly or indirectly, by any means, causes the death of another human being.

The data is collected by Policemen through the Toronto Police system. Any time an officer reports a homicide, the report goes into the system and is added as an observation in the overall dataset. This data is the population of all police-reported homicide incidents by their definition. However, it could also be considered as a sample of all homicide incidents considering there are deaths by criminal negligence, accidental or justifieable homicide. As the dataset is made up of observations that follow a specific definition, it helps to avoid many of the biases we typically see in datasets, as there is less subjectivity. The system Toronto Police allows for this dataset to be very reliable and trustworthy. Toronto Police take high measurement to uphold the integrity of the data. This can be noticed by their Public Data Safety Portal (Toronto Police Services 2022), and thus the dataset this report is working with is reliable.

Despite this, the data does not come without any implications. Considering we are working with crime and police data, there are historically many social and cultural issues that impact our data and make it slightly biased, and thus must be considered from an ethical perspective (Crawford 2021). As most are aware, there have been a lot of biases compiled over the years, such as a claim that low-income neighbourhoods have the highest homicide incidents. Studies have shown that in the USA, social and spatial incarceration of young males has become part of the developmental ecology of adolescence in the nation's poorest neighbourhoods (Fagan, West, and Holland 2004). This fact may be applies in Toronto as well, as the high crime category includes 25% of patrol zones with the highest crime rates; there is a positive relationship between community crime and the number of Special Investigations Unit investigations (Wortley, Laniyonu, and Laming 2020).

According to Wortley, community crime level may be an important predictor of police use of force. However, it is also important to note that Special Investigations Unit investigations and use of force cases happen in low-crime and high-crime patrol zones (Wortley, Laniyonu, and Laming 2020). As shown below, the number of Police dispatch is highest in low-income neighbourhoods (Division 31), whereas the number of homicides is spread across the city. This fact raises the question that the police might likely have a slight biased on homicide incidents in these neighbourhoods. This means that our data comes from a biased sample as opposed to a full one, as the police and their biases have affected the observations in our dataset (Bronner 2020). One may argue that this bias is less likely to occur since the data was taken from Toronto.

To understand and to conduct an appropriate approach on the subject, this paper analyzed the data in R (R Core Team 2020), using packages such as Tidyverse (Wickham et al. 2019), Dplyr (Wickham et al. 2021), and Knitr (Xie 2014) to produce tables and figures that tell us important information about the dataset and can be used to tell a story about homicide occurrences in Toronto.

Table 1: 12 observations from dataset of Homicide occurences in Toronto

Year	Type	Neighbourhood
2004-01-03	Other	Yonge-St.Clair (97)
2004-03-03	Shooting	Old East York (58)
2005-02-15	Shooting	Dorset Park (126)
2005-07-06	Shooting	Thistletown-Beaumond Heights (3)
2011-12-02	Other	Weston (113)
2016-09-13	Stabbing	York University Heights (27)
2014-03-19	Shooting	East End-Danforth (62)
2012-11-12	Other	Mount Olive-Silverstone-Jamestown (2)
2006-03-02	Other	Taylor-Massey (61)
2007-06-25	Stabbing	Rouge (131)
2008-10-25	Shooting	Clanton Park (33)
2015-11-10	Other	Clairlea-Birchmount (120)

Table 1 shows an extract of 12 data points from the overall dataset, which contains 1166 observations. Two observations from each year are shown in the extract. The original data had 10 variables. As shown, it was narrowed down to the three main variables, plus one identifier variable was kept for organizational purposes. These three important variables that will be worked with in the paper are 'Year', 'Neighbourhood', 'Homicide types', and 'Number of homicides', which are 'Occurred_year', 'Neighbourhood', and 'Homicide_Type', respectively, in the dataset.

Occurred year is fairly self-explanatory and is the year in which all the homicides in an observation took place. It is a numerical variable. The neighbourhood is a categorical variable representing each neighbourhood's geographically determined area boundaries. The number of homicides is a numerical variable that counts the number of Homicide_Type corresponding to each observation. Figure 1 shows the number of homicide types over 16 years in Toronto.

Plotting every single observation in the dataset gives us a better idea about the trend for Toronto's homicide occurrences and also allows us to compare each homicide type. However, the plot points for each year seem to have no pattern. Therefore, one can conclude that the number of homicides in Toronto would remain high with no signs of decreasing.

Next, this paper projected a bar plot to visualize the relationship between the neighbourhood and the number of homicides Figure 2.

Figure 2 also depicts the number of homicide types in each neighbourhood more evident. Every coloured section representing the neighbourhood contains more information, and it is worth looking at these deeper.

The leading neighbourhoods in homicides are apparent now. The bar for Malvern (132) is much higher than for Church-Yonge Corridor (75). According to MLS Toronto, the average household income is $72,330,

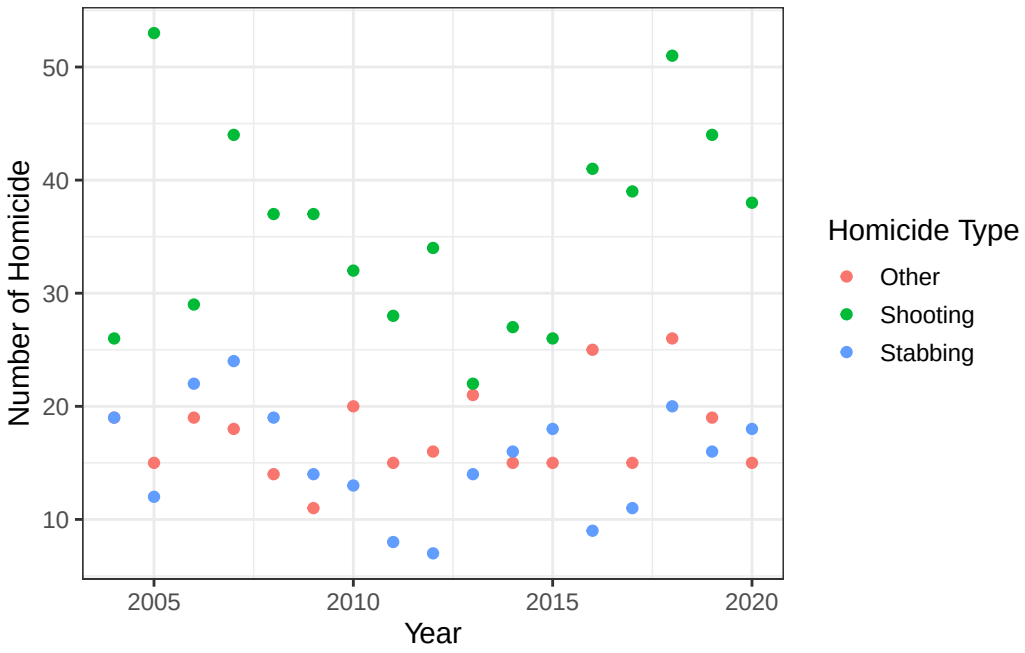

Figure 1: The number of homicides types over 16 years in Toronto

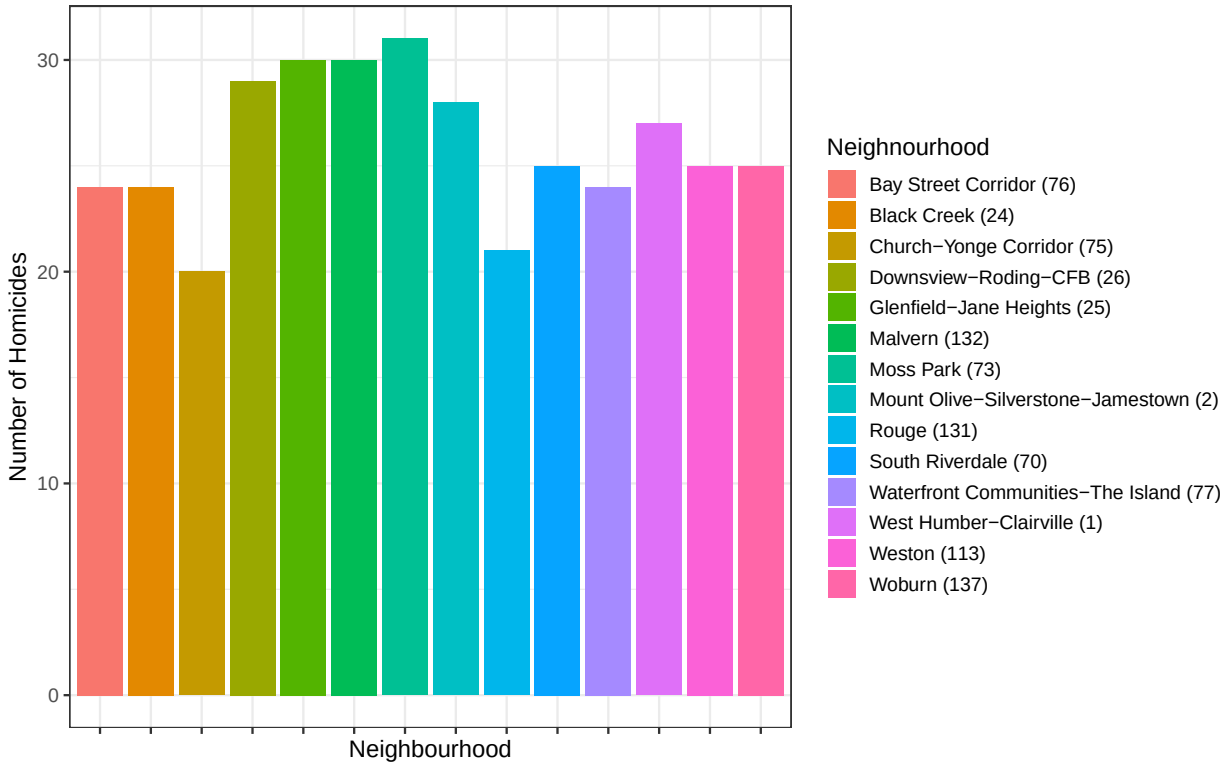

Figure 2: The number of homicides in each neighbourhood

with a median age of 38.9 years in the Malvern neighbourhood ("MLS Toronto" n.d.). In Church-Yonge Corridor, the average household income is $403,902, and the median age group is 43.5 ("MLS Toronto" n.d.). Comparing these two neighbourhoods, it is not a far stretch that the homicide occurrence rates are related to house income. However, is there any other evidence to support the claim of the negative relationship between household income and the rates of homicide occurrence? This paper took one more dataset to look closely at the subject.

As shown in Table 1, each observation tells us the number of homicides that occur in a specific neighbourhood, based on the geographical boundaries described above, on a certain day of the year. For example, the last observation seen in Table 1 tells us that on 2015-11-10, a homicide occurred with an unclassified (other) method of murder in Clairlea-Brichmount. These data points can be analyzed together to tell us a story about homicide incidents in Toronto.

Table 2: Yearly number of each homicide type

No.	Year	Other	Shooting	Stabbing
1	2004	19	26	19
2	2005	15	53	12
3	2006	19	29	22
4	2007	18	44	24
5	2008	14	37	19
6	2009	11	37	14
7	2010	20	32	13
8	2011	15	28	8
9	2012	16	34	7
10	2013	21	22	14
11	2014	15	27	16
12	2015	15	26	18
13	2016	25	41	9
14	2017	15	39	11
15	2018	26	51	20
16	2019	19	44	16
17	2020	15	38	18

By grouping the data by year, Table 2 shows how many homicides have occurred each year in Toronto. It can be seen that homicides with gun shootings are dominantly higher than other methods. For the most part, shootings are always leading the numbers each year. This fact should be taken into consideration when developing related social programs accordingly.

As shown in Figure 3, the numbers of police dispatch are measured per division. Interestingly, these numbers are much higher than the actual homicides across the divisions. In this graph, Division 31 leads the number and D51 the second. Interestingly, as shown below, the leading homicide occurrences are in Division 51 ('Cabbagetown-South St.James Town', 'Church-Yonge Corridor', 'Moss Park') and Division 31 ('Black Creek', 'Downsview-Roding-CFB', 'Glenfield-Jane Heights'), respectively. The discrepancy between these two data raises an interesting question.

3 Conclusion

Is there a correlation between poor neighbourhoods and homicide occurrence? The answer is maybe not. As shown in Figure 4, Division51, whose jurisdiction is in the rich neighbourhoods - 'Bay Street Corridor', 'Church-Yonge Corridor', 'Moss Park', and 'Regent Park' - is leading the homicide occurrence in Toronto. This finding strongly suggests that there may be a positive relationship between income level and homicide. This finding contradicts the paper's hypothesis.

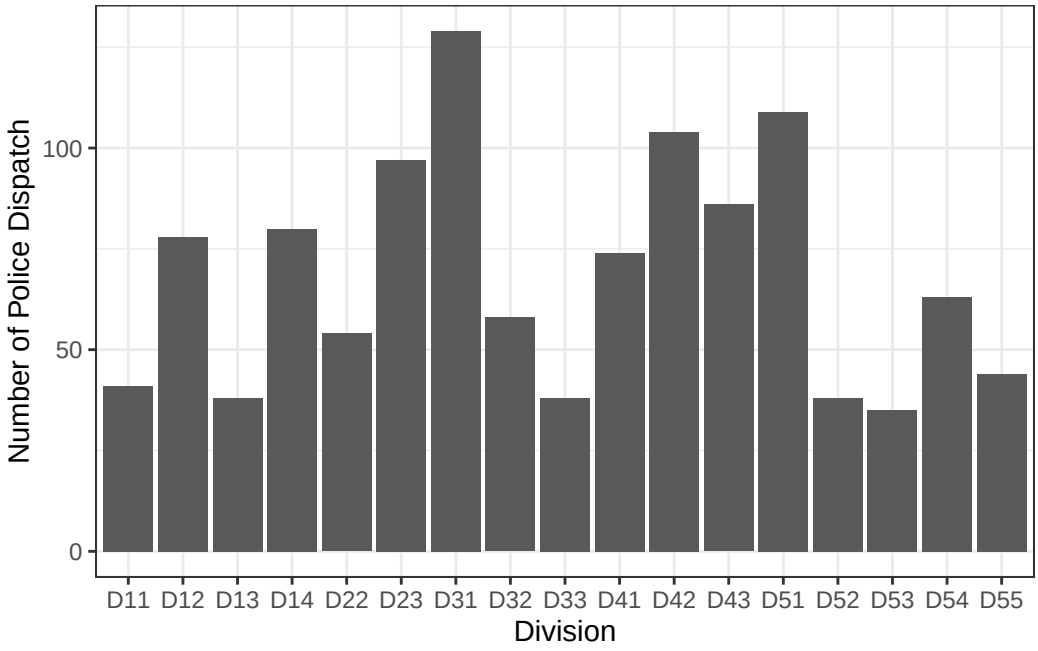

Figure 3: The number of Police dispatch across Divions

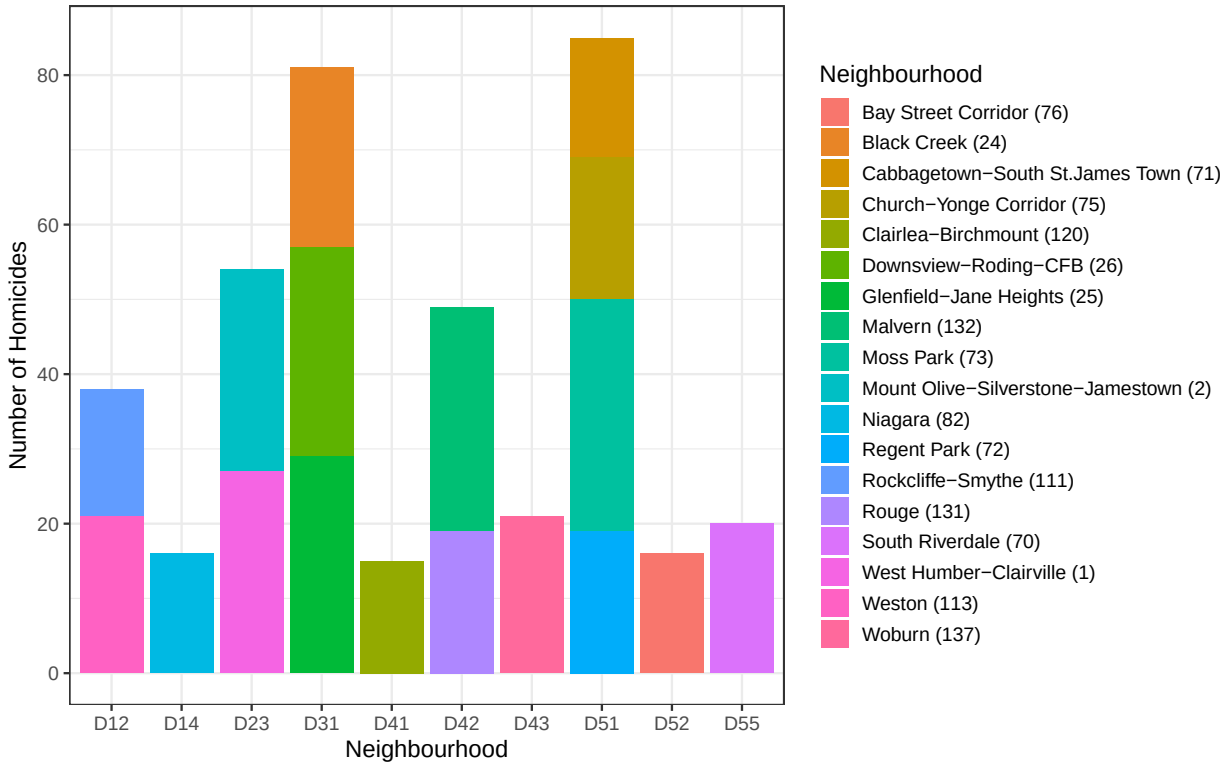

Figure 4: The number of homicides in each Division

This paper started with the hypothesis: 'Low-income neighbourhood leads the homicide occurrence in Toronto and, therefore, there is a negative relationship between the two'. However, based on the analysis of the findings in this paper, there is enough evidence to reject the hypothesis. Therefore, the bias that homicide occurrence in poor neighbourhoods is not true, and it has no relationship to homicide. Furthermore, this paper revealed an interesting aspect that is worth mentioning. This paper concludes with a remark: "there might be a positive relationship between income level and homicides - as income level goes up, the homicide rate also goes up."

Bibliography

Bronner, Laura. 2020. *Why Statistics Don't Capture the Full Extent of the Systemic Bias in Policing*. FiveThirtyEight. https://fivethirtyeight.com/features/why-statistics-dont-capture-the-full-extent-of-the-systemic-bias-in-policing/.

Crawford, Kate. 2021. *Atlas of AI*. Yale University Press.

Fagan, Jeffery, Valerie West, and Jan Holland. 2004. *Neighborhood, Crime, and Incarceration in New York City*. Columbia Public Law Research Paper. https://scholarship.law.columbia.edu/faculty_scholarship/2202/.

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*.

Gelman, Andrew, Alex Kiss, and Jeffrey Fagan. 2005. *An Analysis of the NYPD's Stop-and-Frisk Policy in the Context of Claims of Racial Bias, Columbia Public Law Research Paper*. Columbia Public Law Research Paper. https://scholarship.law.columbia.edu/faculty_scholarship/1390.

"MLS Toronto." n.d. Accessed February 3, 2023. https://www.realtor.ca.

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

Statista Research Department. 2021. *Crime in Canada - Statistics & Facts*. Statista. https://www.statista.com/topics/2814/crime-in-canada/#dossierKeyfigures.

Toronto Police Services. 2022. *TPS Public Safety Data Portal*. https://data.torontopolice.on.ca/.

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, Scot, Ayobami Laniyonu, and Erick Laming. 2020. *Use of Force by the Toronto Police Service*. Ontario Human Rights Commission. https://www.ohrc.on.ca/sites/default/files/Use%20of%20force%20by%20the%20Toronto%20Police%20Service%20Final%20report.pdf.

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.