Annual Report on Homicides in Toronto: An Overview of Police Related Statistics on Homicidal Crimes*

Sarah Lee

January 25, 2024

Recognizing the importance of safety is to keep awareness of the homicide rates in Toronto and its surrounding factors. The dataset retrieved from OpenDataToronto keeps cases of homicide records between 2004 to 2023 covering Toronto. With this dataset, it is shown that there was a major drop of homicide cases through 2010 and 2011, carrying a consistent trend until 2015 and then picking up a considerable increase trend. On the whole, the most noticable factor is shootings being the most common homicide method.

Table of contents

1 Introduction

Violence is exposed all over the world whether you like it or not. As citizens, it is important to understand and recognize the trends and distributions of homicide rates. Studying homicide

*Code and data are available at: https://github.com/sarahhhh02/opendatatoronto.git

rates not only motivates you for your safety and well-being but also delves into all the underlying factors that play into homicides. This paper explores the distribution of homicide rates between 2004 and 2023 along with its predominant methods. As well as recognizing where these distributions were located in Toronto, Canada.

This paper uses the data set from the OpenDataToronto (Gelfand 2022) portal. To be specific, the open data is retrieved from the "Police Annual Statistical Report - Homicides" (Services 2024). This data set records numerous cases of homicides in Toronto with its relevant factors. Factors playing out are the location, year, day, month, day of the week, police division, and its unique ID case. With this data set, I plan to use the R programming language (R Core Team 2022) with its relevant tools like Tidyverse (Wickham et al. 2019), janitor (Firke 2023), knitr (Xie 2014), dplyr (Wickham et al. 2023), and lubridate (Grolemund and Wickham 2011). With the use of this language, I will analyze the types of trends and patterns to how homicide rates play.

The major trends and patterns will be analyzed using the tool from ggplot2 (Wickham 2016) that will graph the needed information. I will also make use of the tool knitr (Xie 2014) to contruct tables to give a generalized view on what we are looking for in this paper.

All relevant materials supporting the code and data are available at Github.

2 Data

The original data set that was retrieved from the Open Data Toronto portal (Gelfand 2022) which includes all cases of homicides from 2004 to 2023. In this data set, there includes the dates consisting of the year, month, day, and the days of the week. Moreover, it contains the details of what type of homicide it is and the location by its nearby neighborhood. Taking this data set, I cleaned the data by using the select() function to keep only the columns of data I was going to use for this paper.

For the general data, we have Table 1 that presents us with the yearly homicide numbers from 2004 to 2023. From this data we are able to notice that the range of homicide counts are 50-100. Which doesn't really tell us much about the trend or pattern. But we will be able to discover some sort of pattern in Section 2.1.

Table 1: Total Homicide Cases per Year

Year	Total Homicides
2004	64
2005	80
2006	70
2007	86
2008	70

Year	Total Homicides
2009	62
2010	65
2011	51
2012	57
2013	57
2014	58
2015	59
2016	75
2017	65
2018	98
2019	79
2020	71
2021	85
2022	71
2023	73

Here now, we are going to look at the statistics behind the type of homicide. By Table 2, we are able to notice a exceeding number of "shooting" with 727 counts. Comparing with the other two types, "Stabbing" with 317 and "Other" with 352, we can notice that majority of homicides are caused from the shooting method.

Table 2: Total Count of Each Homicide Type

Homicide Types	Total
Other	352
Shooting	727
Stabbing	317

2.1 Yearly Homicide Occurences between 2004 to 2023

```
#Line graph of total homicides per year
data <- read.csv(here::here("outputs/data/edited_homicides_data.csv"))

homicides_by_year <- data |>
  group_by(year) |>
  summarise(Homicides = n())

homicides_by_year |>
```

```
ggplot(aes(x = year, y = Homicides)) +
 geom_line(stat = "identity") +
 theme_minimal() +
 labs(x = "Year", y = "Number of Homicides")
```

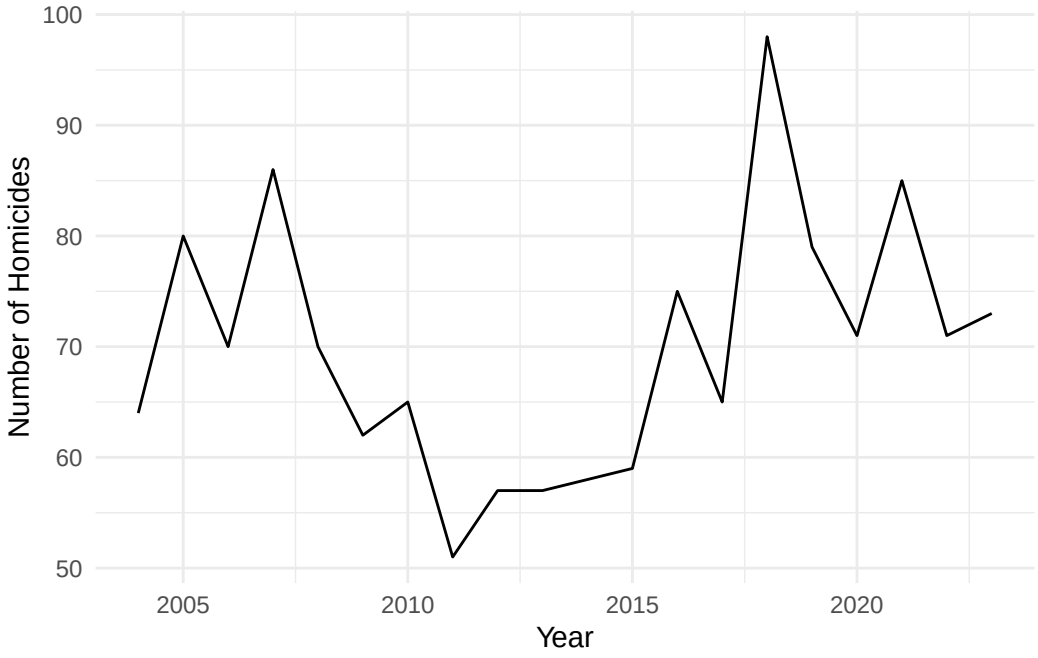

Figure 1: Homicide Occurences from 2004-2023 Yearly

Taking a look at Figure 1, we can notice a major drop of homicide counts near the 2010 and 2011 mark. Then from 2011, it continues to carry a consistent trend until 2015. Then once again striking an increase of cases and continuing with inconsistent trends until the end of the graph. This graph does not tell us a significant trend, but it can tell us what kind of pattern it carries.

2.2 Homicides by Neighbourhood

```
data <- read.csv(here::here("outputs/data/edited_homicides_data.csv"))

homicides_by_location <- data |>
  count(hood_158)
```

```r
homicides_by_location|>
  ggplot(aes(x = hood_158, y = n)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(x = "Neighbourhoods", y = "Number of Homicides")
```

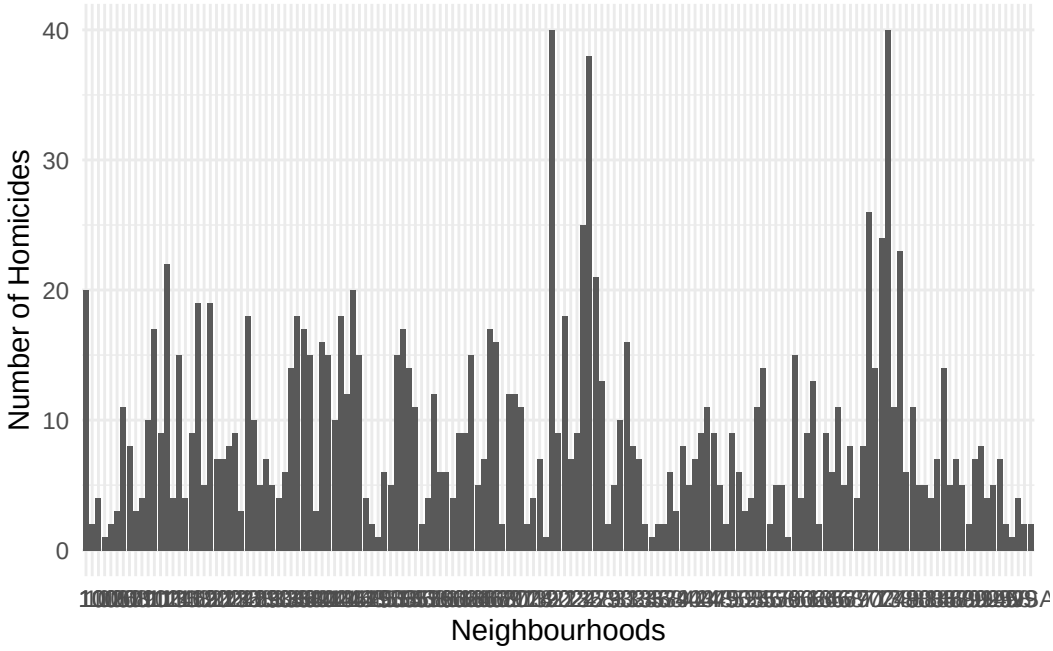

Figure 2: Homicides by Neighbourhood in their Divison Number

References

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://sharlagelfand.github.io/opendatatoronto/.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Services, Toronto Police. 2024. *Police Annual Statistical Report - Homicides*. https://open.toronto.ca/dataset/police-annual-statistical-report-homicide/.

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://dplyr.tidyverse.org.

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.