Analysis of victims structure from 2014 to 2023 in Toronto*

Xiyan Chen

September 28, 2024

Through data exploration, cleaning, visualization, and analysis, this research finds that assault is the most common crime in Toronto, with the total number of crimes increasing over the past 10 years. A closer look at assault cases reveals that male victims outnumber female victims. This finding provides a clearer perspective on public safety concerns, helping inform public safety strategies.

1 Introduction

Through the process of data exploration, cleaning, visualization, and analysis, this research identifies key trends in Toronto's crime landscape over the past decade(2014-2023). The research reveals that assault is the most frequently reported crime in the city while the total number of crime cases has shown a general upward trend over the last 10 years, with some fluctuations. This increase in crime rates raises important concerns regarding public safety, as it reflects a growing issue that requires attention from policymakers and law enforcement agencies.

A deeper examination of assault cases reveals a notable gender disparity among victims. The data indicates that male victims are more numerous than female victims, with approximately 93,750 male victims compared to 81,250 female victims over the 2014 to 2023 period. This finding provides important insights into how different demographics are impacted by crime in Toronto and highlights the need for targeted interventions and support services. By uncovering this gender-based distinction, the research offers a more nuanced understanding of victimization patterns, which can help inform public safety strategies and preventative measures.

The structure of this paper begins with an overview of data sources and visualization methods, followed by graphical representations. And then the analysis section provided then the summary about the key outcomes.

*Code and data are available at: https://github.com/siaxiyann/crime.git

2 Data

2.1 Sources

All the data was sourced from the Police Annual Statistical Report - Victims of Crimes dataset, available on open data portal of Toronto@opendatatoronto. The data was cleaned and analyzed using R@citeR, with the assistance of packages such as tidyverse@tidyverse, dplyr@rdplyr and Wickham et al. (2019). The graphs were generated using ggplot2 Wickham (2016).

2.2 Variables and Measurements

There are similar datasets, such as the Police Annual Statistical Report - Crimes, which also provide insights into crime in Toronto. However, this paper focuses specifically on victims, making the Police Annual Statistical Report - Victims of Crimes dataset more relevant for this study.

The original dataset contains 10 variables, four of which were selected for in-depth analysis. After cleaning the missing data and removing unselected columns, the variables chosen were: report year, subtype, sex, and count.

More specifically, the report year represents the year the crime was reported, covering the period from 2014 to 2023. This variable helps examine the overall trend of victimization in Toronto over 10 years (see Figure Figure 1). The subtype refers to the crime category, divided into four types: assault, robbery, sexual violation, and other. This categorization aids in further exploration by focusing on specific types of crimes – assult.

The sex variable denotes the gender of the identified victims, categorized as female, male, or unknown. The "unknown" category represents cases where the victim's gender could not be determined. Lastly, count reflects the number of reported cases. Note that the count includes instances where the same person may be victimized multiple times for the same offense within the same period. To enhance the analysis, a new variable called total_count was created to reflect the total number of crimes each victim was involved in from 2014 to 2023. This was used in Figure Figure 2 to analyze the distribution of victims by gender.

```
analysis_data <- read_csv("../data/raw_data/analysis_data.csv")
```

```
Rows: 1244 Columns: 4
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (2): SUBTYPE, SEX
dbl (2): REPORT_YEAR, COUNT_
```

```
i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

3 Analysis

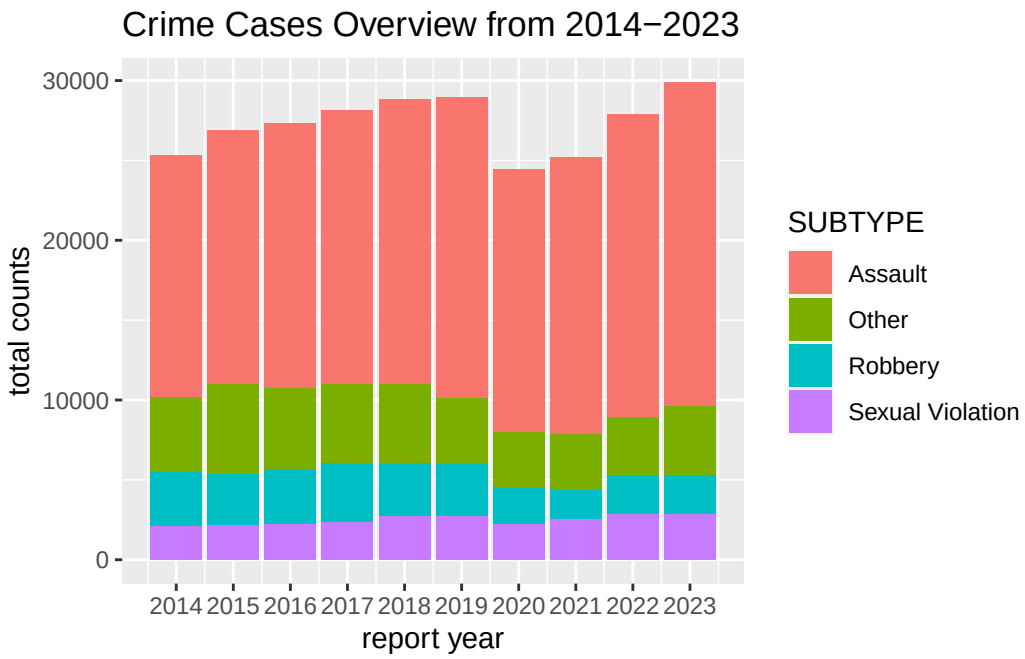

Figure 1: Distribution of crime types by year

The bar plot Figure 1 provides an overview of the number of victims in each crime case and the distribution of crime types. There was a slight increase in the total number of victims from 2014 to 2019, followed by a small decline in 2020. However, a significant upward trend is observed from 2020 to 2023. Notably, assault (represented by the red bar) is consistently the most frequent crime in Toronto each year, with the "other" crime category ranking second. Robbery and sexual violations are relatively equally distributed in terms of frequency.

Focusing on assault, which is the most common crime in Toronto, Figure Figure 2 illustrates the distribution of assault victims by gender from 2014 to 2023. Male victims outnumber female victims, with approximately 93,750 male victims compared to 81,250 female victims during this period.

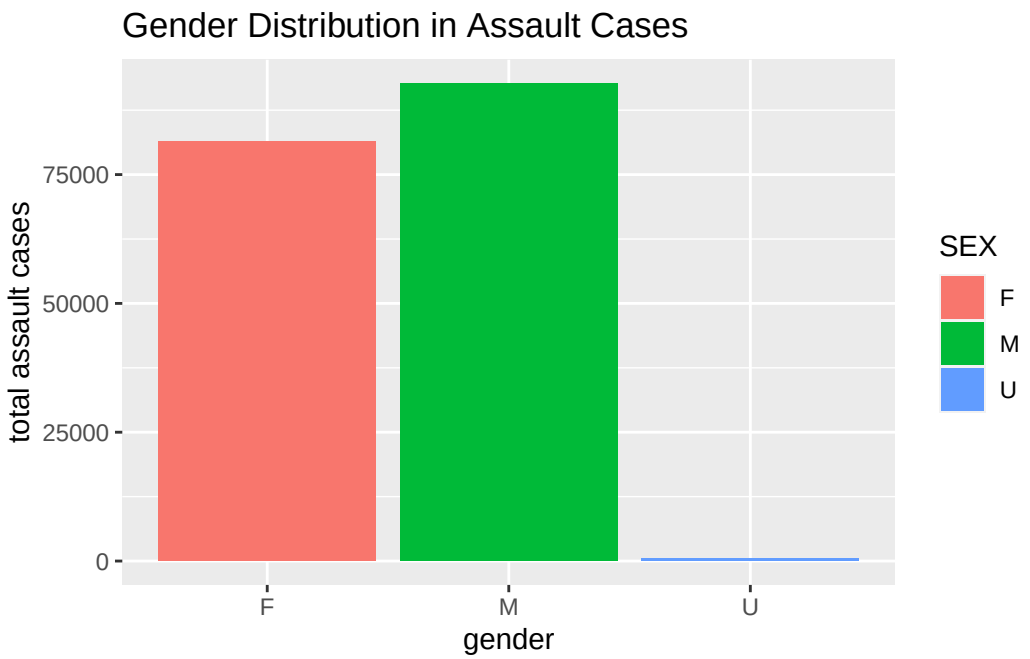

Figure 2: examing the number of crime case in 4 types from 2014-2024

References

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.