Is Our Community Getting Safer in Recent Years: A Data-Driven Study of Crime Victims in Toronto*

Huayan Yu

September 24, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

In the body of this paper, we will first discuss the data source and our computational environment, along with some basic facts and significances about the data set in Section 2. We then aim to uncover the dynamics and insights from the data through diagram and table visualizations in Section 3 to address our primary research question for the trend of public safety in recent years in the City of Toronto.

2 Data

2.1 Data Source and Computational Environment

The dataset used in this study, "Police Annual Statistical Report: Victims of Crime," is provided by the City of Toronto through its open data initiative. [1] It reflects crime incident reports, focusing on the experiences of victims from 2019 to 2023. The data is publicly accessible, ensuring transparency in city governance and fostering community engagement in crime monitoring. Ethically, the data adheres to the Open Government License, which mandates privacy protections for individuals involved, promoting responsible use in academic and policy

---

*Code and data are available at: https://github.com/huayan1998/toronto_crime_analysis-main

1

analysis. [2] Statistically, the dataset offers insights into crime trends, aiding in evidence-based policymaking. The data set also provides rich context for statisticians to extract useful insights by analyzing victim components and trends over time.

All data analysis and visualization in this paper were performed using the R programming language (R Core Team 2023) along with the following packages: `tidyverse` (Wickham et al. 2019) which includes the `dplyr` component (Hadley Wickham 2023) for data frame manipulation and `ggplot2` (Wickham 2016) for visualizations, as well as `knitr` (Xie 2023) for table formatting.

2.2 Data Description

The dataset consists of 1244 records from 2014 to 2023, with each column representing critical information as follows:

- `REPORT_YEAR`: The year in which the crime was reported.
- `CATEGORY`: The general classification of the crime, such as "Crimes Against the Person."
- `SUBTYPE`: A more detailed breakdown of the crime, including types like "Assault" or "Robbery."
- `SEX`: Gender of the victim, represented as Male (M), Female (F), or Unknown (U).
- `AGE_GROUP`: The age range of the victim, including "Child", "Adult", "Youth", and "Unknown."
- `AGE_COHORT`: Specific age groupings (e.g., 25 to 34) for more detailed demographic analysis.
- `COUNT_`: The number of occurrences for each record.

In similar studies, other datasets like overall crime reports which usually focus more on offenders could have been used. However, these datasets were not appropriate for our study because this dataset uniquely focuses on victims, providing insights specific to victim demographics and experiences that are important for understanding crime impacts within Toronto local communities.

Table 1: Number of Records in Each Categorical Variable

CATEGORY	SUBTYPE	SEX	AGE_COHORT
Crimes Against the Person:1244	Assault :617 Other :222 Robbery :201 Sexual Violation:204	F:518 M:570 U:156	Unknown :193 25 to 34:165 35 to 44:154 45 to 54:148 18 to 24:143 55 to 64:128 (Other) :313

Table 1 above gives an overview of the distribution of key variables, showing that all crimes recorded in the dataset fall under the "Crimes Against the Person" categories. "Assault" is the most common subtype (617), with a fairly even gender distribution (518 females, 570 males). Age cohorts show a diverse range, with "Unknown" being frequent, and many victims in the 25 to 34 age range (165).

The dataset has no missing values in each of its column, and we will not be analyzing the column `ASSAULT_SUBTYPE` because limited information is found in the Open Data Toronto data description page. [1] Hence no further data cleaning is needed other than removing the `ASSAULT_SUBTYPE` column. This comprehensive dataset facilitates focused analysis on victim profiles and crime trends, making it valuable for understanding community safety.

3 Results

3.1 Discovering the victim patterns for different crime subtypes

We first attempt to discover trends in crime over time across different subtypes and identify specific areas that might require targeted interventions. We first consider Figure 1 below which focuses on analyzing on crime dynamics in Toronto from 2014 to 2023. Assault has a prominent increasing trend, starting around 15000 victims in 2014 and rising to over 20000 by 2023, indicating a substantial escalation in assaults over the past decade except the sudden drop in 2019. On the other hand, crimes categorized as "Other" display a gradual decline after a peak in 2015, stabilizing in recent years, but this subtype is also witnessing a reclimb starting from 2022. Robbery and sexual violations show generally more stable, moderate patterns with minor fluctuations.

This figure reveals that while certain violent crimes, such as assault, have increased, others have remained relatively consistent, highlighting distinct trajectories in various crime types over time. However, all subtypes exhibit upward trends again in or after 2021, adding new uncertainties and concerns about public safety. Considering the fact that assault and sexual violation crimes are usually associated with more physical and psychological injuries, the local community urges some effective policy changes to control escalating crimes.

As we look closer into the largest crime subtype of assaults in Figure 2, we discover that majority of victims are adults, regardless of gender, with over 70000 adult victims in both male and female categories. It is worth noting that the victims of unknown age group are at comparative figures as child and youth victims, also posting concerns for the data quality being recorded in the police system. Additionally, while youth and child victims appear to be fewer, they still require further attentions especially after we adjust for the population age distribution.
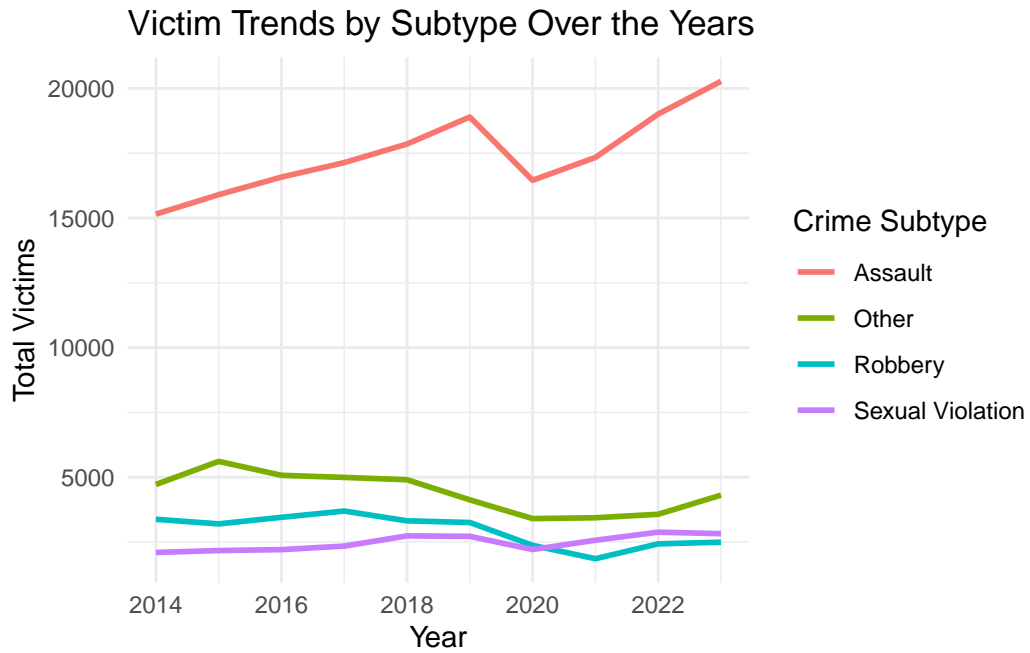
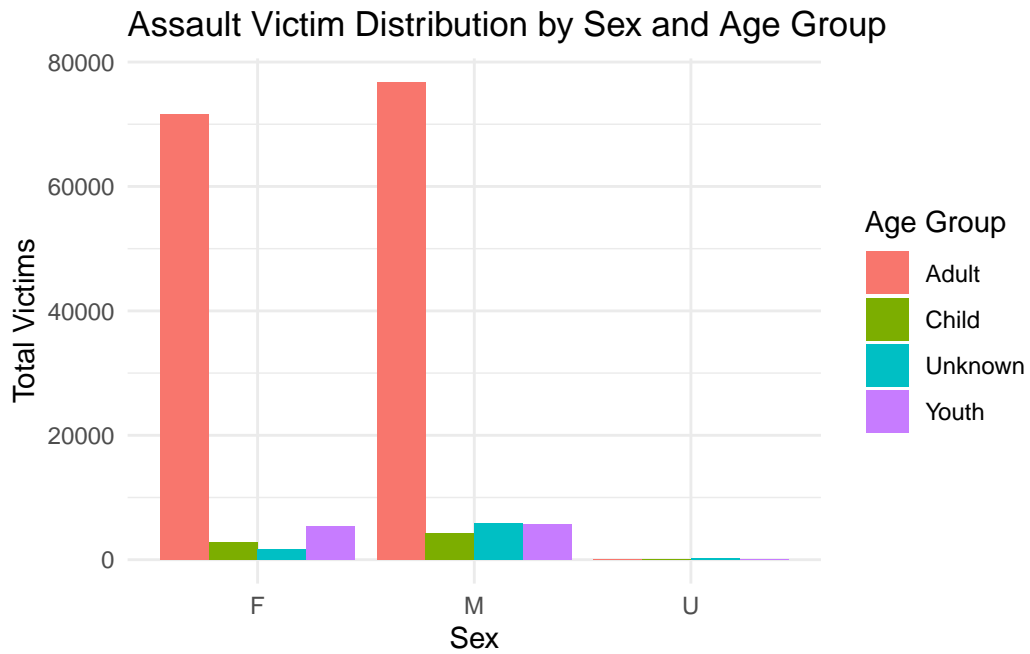Figure 1: Yearly victim trends by subtype from 2014 to 2023

Figure 2: Count of victims of assault crimes by sex and age groups

3.2 Discovering the trends of changes in victim counts after 2020

As we discovered in Section 3.1, all crime subtypes seem to regain the growth trend starting in or after 2020. This demands a closer examination of the dynamics within the victim data. Table 2 below reveals that the gender distribution of crime victims has remained relatively balanced over the years. This stability in gender distribution suggests a consistent pattern in the way crimes affect male and female populations. More importantly, the two genders both demonstrate a strictly increasing trend in the total number of victims with a growth rate of over 10% from 2021 to 2022. This observation could be evidence for the residual effects of the COVID-19 pandemic and the accompanying economic challenges. The pandemic's disruptions could have contributed to a rise in both opportunity-driven and stress-induced crimes, placing general public mental health condition at another challenge.

In contrast, the unknown category sees a sharp decline, with the total victim count decreasing by 35.23% between 2022 and 2023, indicating improved data collection or identification efforts.

Table 2: Proportion and Growth of Victims by Year and Sex

Year	Sex	Total	Proportion	Pct Change Total
2020	F	12174	0.498	0.00
2021	F	12675	0.503	4.12
2022	F	14096	0.506	11.21
2023	F	14794	0.495	4.95
2020	M	12153	0.497	0.00
2021	M	12395	0.492	1.99
2022	M	13699	0.491	10.52
2023	M	15041	0.503	9.80
2020	U	109	0.004	0.00
2021	U	122	0.005	11.93
2022	U	88	0.003	-27.87
2023	U	57	0.002	-35.23

As we breakdown the age groups into finer ranges, Figure 3 shows that there are consistently a number of outliers in the adult victim category for each year, suggesting that while most adult incidents fall within a consistent range, certain periods saw spikes in crime involving adults, possibly due to heightened vulnerability during specific events. This illustrates the needs for stricter regulations on public events.

Moreover, a slight upward trend in the number of youth and adult victims is observed after 2020, which is likely connected to broader societal impacts, such as increased stress and economic instability following the COVID-19 pandemic. This shift may indicate rising safety

concerns faced by the younger populations over time, and calls for a more robust satefy awareness education strategy.
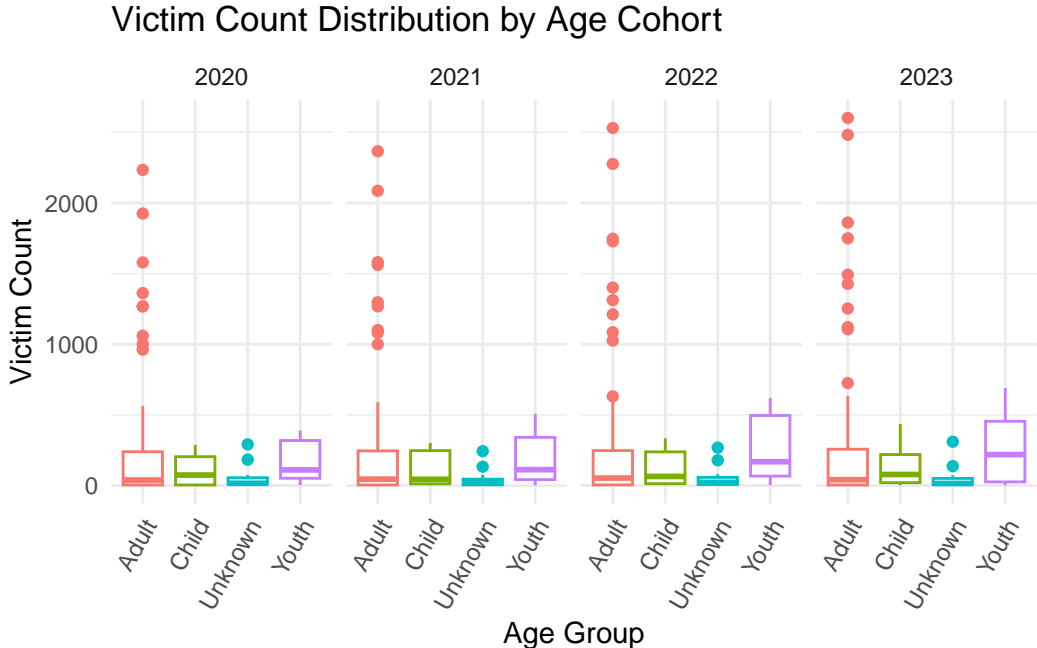
Victim Count Distribution by Age Cohort

Figure 3: Victim count distribution by age cohort

4 Discussion

4.1 Conclusion

In this study, we start from introducing the open-source dataset that describes crime victims in Toronto in the last ten years, and then proceed with figure and table visualizations to address the research question of whether our community is getting safer in recent years. From our analysis in Section 3, we discover that assault cases represent the largest proportion of crimes, showing consistent growth over the years. The gender distribution of victims remains balanced, with no significant differences between male and female victimization rates.

However, crime rates across all subtypes began to rise again after 2020 as seen by the increasing numbers of victims across all subtypes, age groups, and genders. The number of youth and child victims also demonstrate upward trends in the new decade. Even though recent improvements in data collection methods show positive progress in identifying victims more accurately, public safety is still facing an increasing number of challenges. Based on the data-driven results, we are centainly not confident enough to claim significant progress for the public safety in the Toronto communities.

4.2 Limitations and Next Steps

There are still a few limitations that should be acknowledged in this study. The dataset offers a broad overview without including more specific details on the impacts or contexts of individual crimes, limiting our ability to analyze causes and consequences of crime activities in greater depth. Additionally, the analysis was based on visualization without incorporating advanced statistical models, such as regression analysis or time series forecasting, to better quantify trends or correlations. For instance, cross-factor interactions—like how age and sex may jointly affect crime victimization—remain unexplored.

Future work could address these gaps by obtaining more detailed data on crime contexts, such as socio-economic background or psychological impact assessments, and incorporating insights from professionals in relevant public sectors and social services to provide practical and grounded interpretations. Deploying advanced statistical techniques, such as regression analysis, could further enhance understanding by quantifying correlations and interactions between different factors. This would allow for a deeper and more nuanced exploration of crime trends and help guide policy interventions more effectively.

5 References

[1]. About Police Annual Statistical Report - Victims of Crimes (https://open.toronto.ca/dataset/police-annual-statistical-report-victims-of-crime/) [2]. Open Government Licence – Ontario (https://www.ontario.ca/page/open-government-licence-ontario) [3]. R [4]. tidyverse [5]. dplyr [6]. ggplot2 [7]. knitr

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