Analysis of Victim Counts*

A Multidimensional Study Based on Age,Gender and Crime Type

Yawen Tan

September 27, 2024

This report offers a comprehensive multidimensional analysis of victim counts from 2014 to 2023, exploring the intricate relationships between age group, gender, and crime type. Through comparative analyses, key trends and disparities in crime patterns are identified across these demographic factors. Our findings indicate that young adults have a higher vulnerability in crime, young women are more likely to suffer sexual violence and young men are more likely to suffer property robbery. Grasping these dynamics is crucial for formulating effective prevention strategies and informed public policies.

Table of contents

*Code and data are available at: https://github.com/YawennnnnnTan/Analysis-of-Crime-count

1 Introduction

Crime victimization is a complex and urgent social issue that impacts not only individual victims but entire communities. Research by Ainsworth (2001) and Wilcox, Land, and Hunt (2003) highlights that factors such as age, gender, and socioeconomic status significantly influence patterns of victimization. In urban environments, distinct differences emerge in the victimization experiences of various age groups and genders and crime types. For instance, studies by Duncan (2017) and Piquero (2018) indicate that young individuals, particularly those aged 18 to 34, are generally at a higher risk of becoming victims of crime. Additionally, specific crime types—such as violent crime, robbery, and sexual assault—exhibit varying patterns of victimization, as demonstrated by the findings of Chappell and Maggard (2004) and Farrall and Gadd (2005). Understanding these dynamics is essential for the development of effective prevention strategies and informed public policies.

Although existing literature has identified the correlation between victimization and demographic factors, it often fails to fully explore the interactive effects of gender and age under different crime types, and this gap urgently needs to be filled. To this end, this study focuses on crime victim data in Toronto from 2014 to 2023, and uses the "Annual Police Statistical Report - Crime Victims" dataset for in-depth analysis. The data was cleaned and visualized using R language and its various data analysis tools, so as to analyze Toronto's crime data and explore the relationship between victim age, gender, crime type and crime volume, as well as their interactive relationship. This study found that young women are more likely to suffer sexual violence, young men are more likely to suffer property robbery, and young adults have a higher vulnerability in crime. These findings emphasize the importance of considering different age groups and genders and crime types and social dynamics when developing targeted crime prevention strategies, thereby providing data support and theoretical basis for policymakers.

The remainder of this paper is structured as follows. Section 2 discusses raw data, analytical data, data measurement and data visualization which includes the plots of each variable and its response, as well as plots of variable interactions. Section 3 shows the relationship between different age groups and genders in crime victimization, and explores specific crime types. Section 4 discusses the results of Section 3 going into detail which highlights the complex relationship between sociocultural background and crime patterns. Section 5 points out the limitations of the data and research, and proposes the focus that future research should focus on to enhance the depth and accuracy of the research.

2 Data

The data for this paper comes from *Open Data Dataset* (n.d.) and was accessed using the opendatatoronto library Gelfand (2022). The dataset selected, "About Police Annual Statistical Report - Victims of Crimes" (n.d.), contains crime counts from 2014 to 2023, is updated

annually, and was last refreshed on August 2, 2024. Sourced and published by Toronto Police Services, this data complies with the Municipal Freedom of Information and Protection of Privacy Act, ensuring privacy protection for individuals involved in the reported incidents. The dataset was chosen for its extensive data and variety of variables recorded over the five-year period, enabling a comprehensive analysis of crime characteristics.

2.1 Raw Data

The raw downloaded data is shown in Table 1 and Table 2, as you can see it contains a total of 9 variables and see the details of these 9 variables in Section 6:

Table 1: First 8 entries of Raw Crime Count Dataset

_id	REPORT_YEAR	CATEGORY	SUBTYPE
1	2019	Crimes Against the Person	Assault
2	2019	Crimes Against the Person	Assault
3	2023	Crimes Against the Person	Assault
4	2023	Crimes Against the Person	Assault
5	2019	Crimes Against the Person	Other
6	2015	Crimes Against the Person	Other
7	2023	Crimes Against the Person	Assault
8	2017	Crimes Against the Person	Sexual Violation

Table 2: First 8 entries of Raw Crime Count Dataset

ASSAULT_SUBTYPE	SEX	AGE_GROUP	AGE_COHORT	COUNT_
Assault Peace Officer	F	Adult	45 to 54	4
Assault Resist Arrest	F	Adult	55 to 64	2
Other	U	Adult	35 to 44	7
Other	F	Adult	65+	490
N/A	M	Adult	35 to 44	311
N/A	U	Adult	25 to 34	1
Assault Peace Officer	U	Unknown	Unknown	7
N/A	M	Adult	65+	6

2.2 Cleaned Data(Analysis Data)

The cleaned data is displayed in Table 3, as you can see it contains a total of 4 variables:

Table 3: First 8 entires of Cleaned Crime Count Dataset

crime_type	sex	age_group	crime_count
Assault	Female	45-54	4
Assault	Female	55-64	2
Assault	Female	65+	490
Sexual Violation	Male	65+	6
Assault	Male	45-54	24
Assault	Female	35-44	10
Robbery	Male	25-34	422
Assault	Male	35-44	37

2.3 Data Tools and Measurement

The analyses were conducted in R R Core Team (2023), with support from several packages: tidyverse Wickham et al. (2019), here Müller (2020), dplyr Wickham et al. (2023), janitor Firke (2023), ggplot2, Wickham (2016), and knitr Xie (2014) and opendatatorotno Gelfand (2022).

To analysis crime counts based on gender, crime type and age group, in the original(raw) dataset(Table 1 and Table 2), the variables `id`, `REPORT_YEAR`, and `ASSAULT_SUBTYPE` were not relevant to our analysis. Additionally, since `CATEGORY` contained only one value and `AGE_GROUP` was similar to `AGE_COHORT`, we removed these variables (id, REPORT_YEAR, ASSAULT_SUBTYPE, CATEGORY, AGE_GROUP) to reduce the data size. The remaining variables were renamed for clarity: `SUBTYPE` became `crime_type`, `SEX` was changed to `sex`, `AGE_COHORT` became `age_group`, and `COUNT_` was renamed to `crime_count`. Values such as "other" in `crime_type`, "Unknown" in `sex`, and "Unknown" in `age_group` were also removed as they fell outside the scope of our study and formed the cleaned dataset(Table 3). Finally, the cleaned data was checked to ensure no missing values and that all data points were within their expected value ranges.

2.4 Data Visualization

In this study, we utilize a series of charts to illustrate the relationship between age, gender, and crime type among victims from 2014 to 2023. These visualizations provide clear insights into the trends and characteristics of crime victimization across different demographic groups, offering a deeper understanding of the interplay between these multidimensional factors. In the following sections, each figure will be presented and analyzed in detail to further explore the relevant data.

2.4.1 Exploring the Relationship Between Age Group and Victim Counts

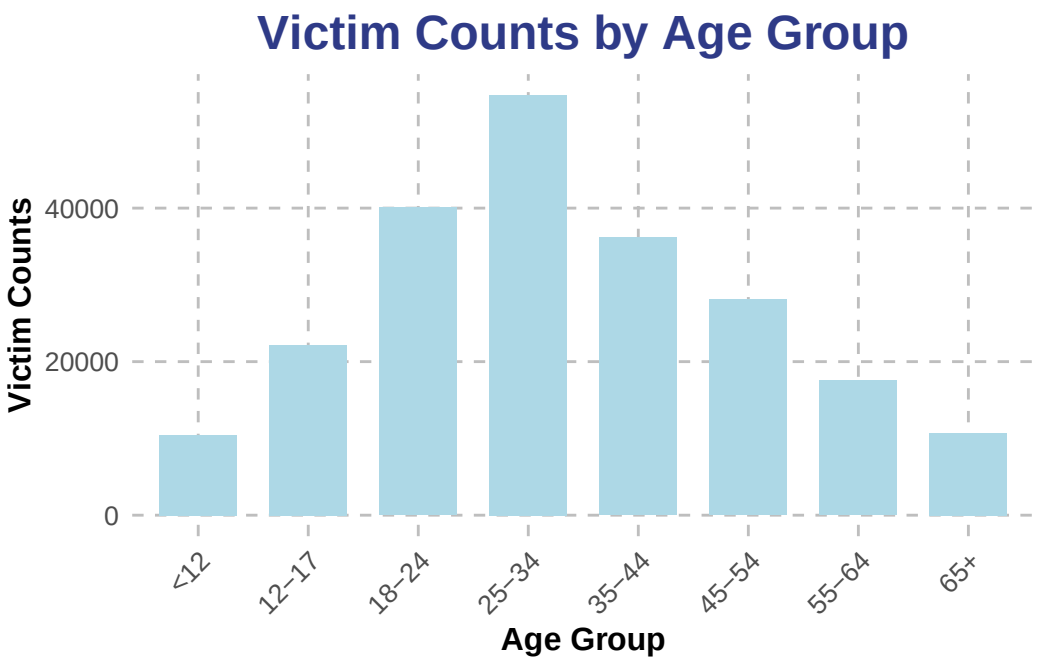

Figure 1: Understanding Victim Numbers:A Breakdown by Age Group

Figure 1 illustrates the total number of victims in each age group from 2014 to 2023 and the age groups are divided into >12, 12-17, 18-24, 25-34, 35-44, 45-54, 55-64 and 65+. Several key insights can be drawn from this visualization. Firstly, the 25-34 age group has the highest number of victims, indicating that individuals in this range are most frequently affected by crime. Additionally, the 18-24 age group also has a significant number of victims, second only to the 25-34 group, suggesting that young adults are at a higher risk of victimization. In contrast, the number of victims in the under-12 and 65+ age groups is relatively low.

2.4.2 Investigating the Relationship Between Gender and Victim Counts

Figure 2 displays and compares the number of crime victims by gender from 2014 to 2023, using green to represent males and red for females. While the number of victims is fairly close between genders, male victims (111,336) slightly outnumber female victims (108,620). The difference, however, is not substantial, with only 2,716 more male victims.

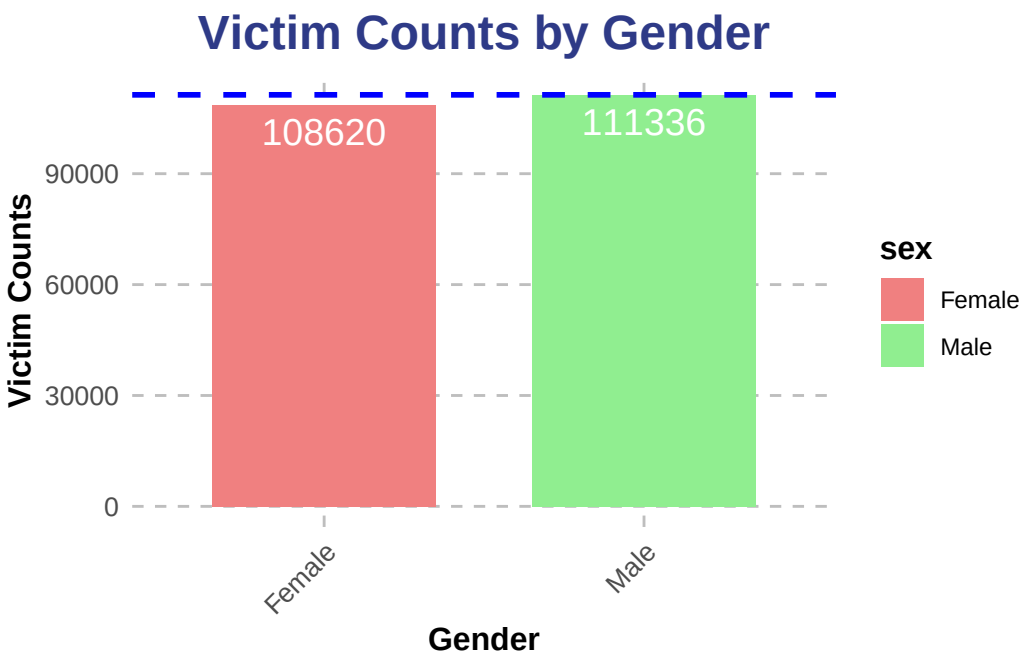

Figure 2: Gender-Based Victim Counts:A Statistical Overview

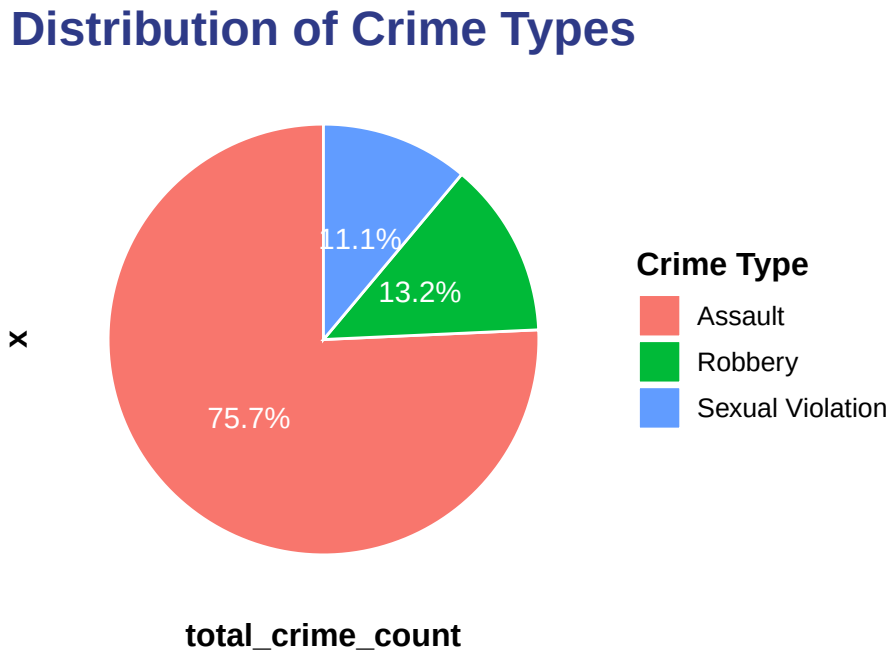

Figure 3: Distribution of Crime Types:A Comparative View

2.4.3 Examining the Link Between Crime Types and Victim Counts

Figure 3 indicates the distribution of three major crime categories—Assault, Robbery, and Sexual Violation—over the period from 2014 to 2023.Assault is represented by red, robbery is represented by green, and sexual violation is represented by blue. Assault dominates, making up approximately 75.7% of all crimes, while Robbery accounts for 13.2%, and Sexual Violation represents 11.1%. The distribution clearly indicates that assault is by far the most prevalent crime, comprising three-quarters of all recorded incidents. In contrast, robbery and sexual violation occur much less frequently.

2.4.4 Comparative Analysis of Victim Counts by Gender and Type

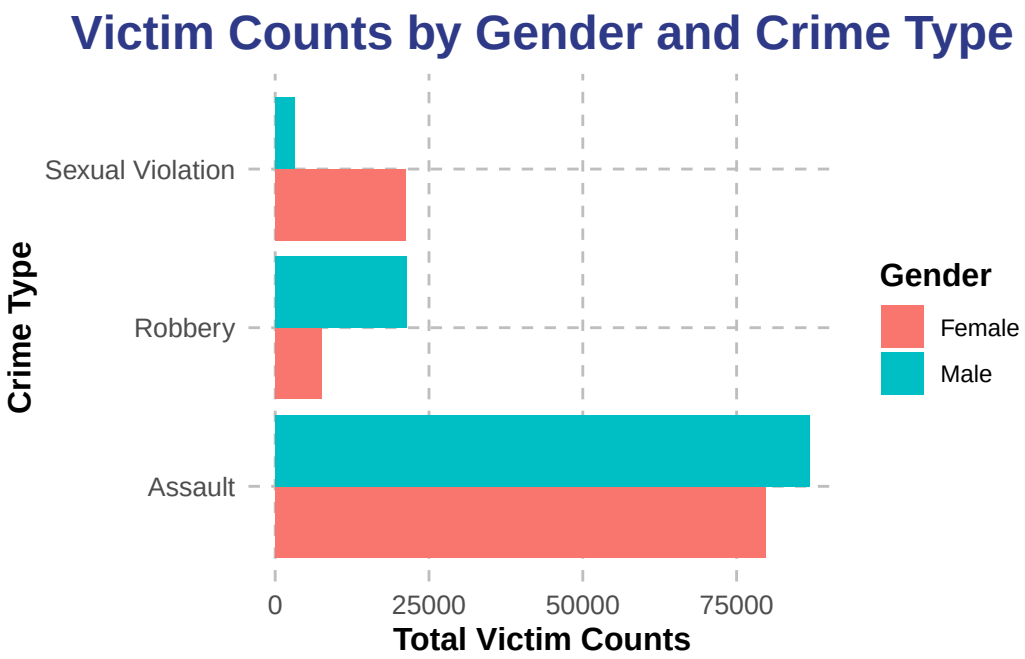

Figure 4: Comparative Analysis of Victim Counts by Gender and Type

Figure 4 shows the number of victims of different genders and crime types from 2014 to 2023. Gender is divided into two categories: Male (blue) and Female (red), and the crime types include: Assault, Robbery, and Sexual Violation. It can be clearly seen from Figure 4 that the number of women suffering from sexual violation is significantly higher than that of men. On the other hand, the number of men suffering from robbery is significantly higher than that of women. In the assault data, although the number of male victims is slightly higher than that of females, the gap between the two is not significant.

2.4.5 Comparative Analysis of Victim Counts by Age and Crime Type

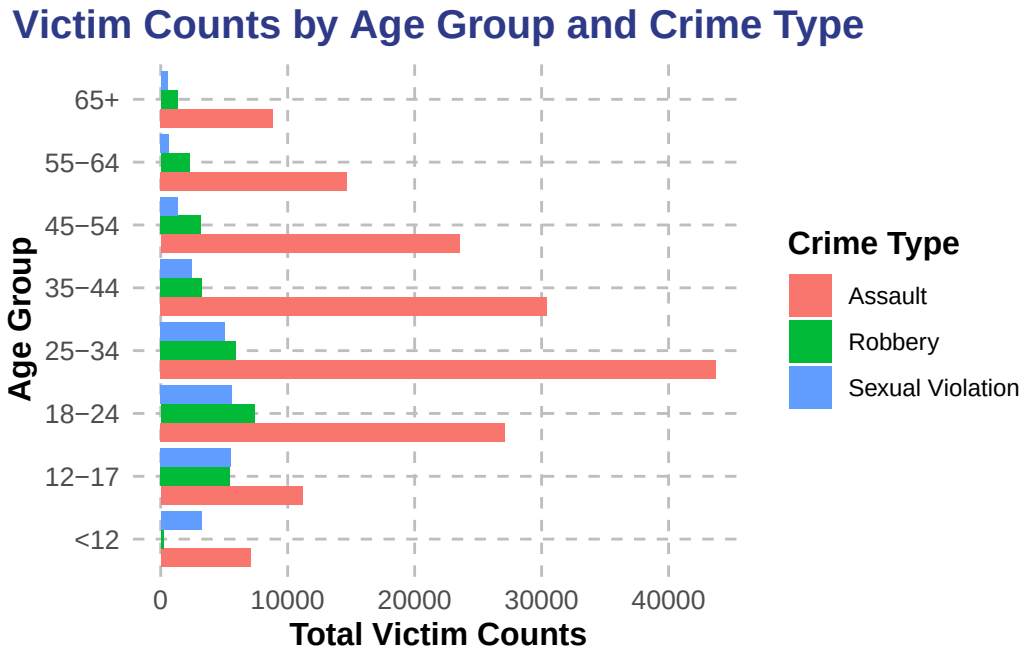

Figure 5: Comparative Analysis of Victim Counts by Age and Crime Type

Figure 5 shows the number of victims by crime type and age group from 2014 to 2023. Crime types are divided into three categories: Assault, Robbery, and Sexual Violation, represented by red, green, and blue, respectively. Age groups include >12, 12-17, 18-24, 25-34, 35-44, 45-54, 55-64, and 65+. Firstly, Figure 5 shows that the number of victims who experienced assault was particularly high in each age group. Secondly, except for the age groups <12 and 12-17, the number of Robbery is always higher than Sexual Violation.

2.4.6 Comparative Analysis of Victim Counts by Gender and Age Group

Figure 6 is a line chart showing the number of victims of different genders and age groups from 2014 to 2023. Gender is divided into two categories: male (blue) and female (red). Age groups include: <12, 12-17, 18-24, 25-34, 35-44, 45-54, 55-64, and 65+. Figure 6 shows that the number of female victims is significantly higher than that of male victims between the ages of 18-44. It is particularly noteworthy that the number of female victims rises sharply between the ages of 12-17 and 18-24. At the same time, after reaching a peak at the age of 25-34, the number of female victims begins to decline significantly from the age of 35-44, indicating that the number of female victims decreases sharply with age. In contrast, the number of male victims fluctuates relatively less in these age groups, showing a more stable trend.

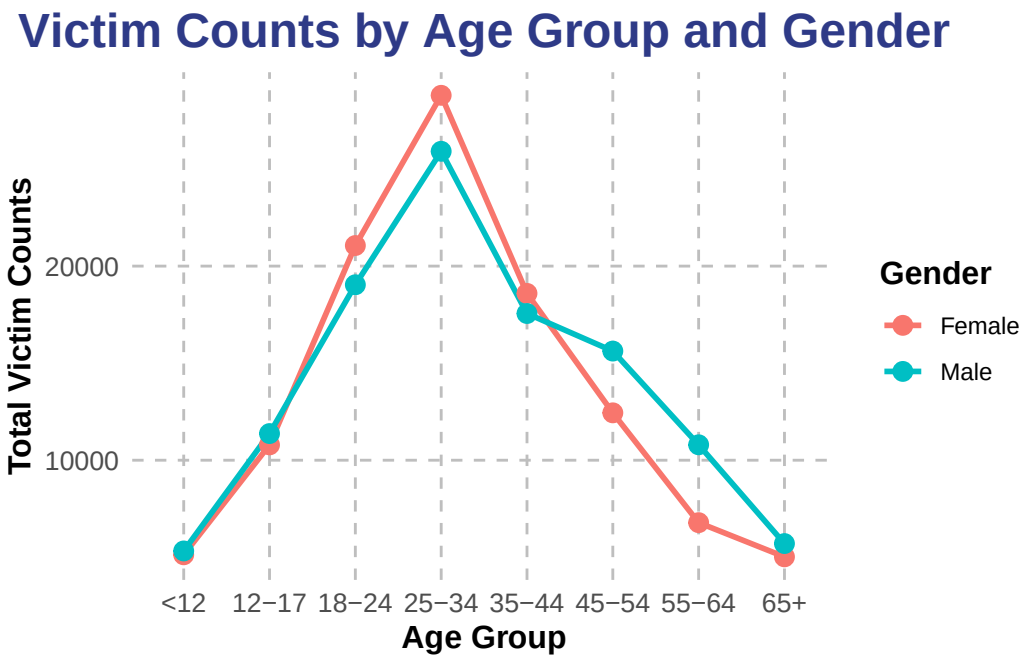

Figure 6: Comparative Analysis of Victim Counts by Gender and Age Group

3 Result

This study undertook a comprehensive analysis of crime victim data spanning from 2014 to 2023, uncovering the intricate relationship between age, gender, and crime type. The findings indicated that individuals in the 25-34 age group represented the highest number of victims, followed closely by those aged 18-24, suggesting heightened vulnerability to crime among young adults. In contrast, the victim counts for those under 12 and over 65 were significantly lower, potentially reflecting the effectiveness of protective measures and lifestyle factors during this period (Figure 1). The analysis of various crime types revealed that assaults constituted the overwhelming majority of incidents (Figure 3). Furthermore, while the number of male and female victims was relatively similar overall, notable disparities emerged across specific crime types (Figure 2 and Figure 4). In conclusion, young women were disproportionately affected by sexual assault cases, whereas young men were more frequently victims of robbery. These findings underscore the importance of considering the complex social dynamics and crime patterns when formulating targeted crime prevention strategies for different genders and age groups, thereby enhancing efforts to reduce victimization.

4 Discussion

Based on the information laid out in Section 2, in particular through Figure 1,it highlights that younger (12-17) and older (65+) individuals experience lower crime victimization rates, while young adults aged 18-44 appear to face a higher risk.. This suggests that young adults face a higher risk of victimization. This vulnerability may be influenced by multiple factors, such as inexperience, peer pressure and lifestyle choices, including nightlife activities, which are often associated with higher crime rates. The lower number of victims in the under-12 age group may reflect parental monitoring and protective measures that help protect younger children from potential threats. For the age group over 65, the lower number of victims may indicate a more stable lifestyle, and that older people tend to have fewer opportunities to participate in high-risk activities, thereby reducing the likelihood of victimization.

Figure 2 which indicates a near parity in victimization by gender, with males accounting for a marginally higher number of victims. It suggests a relatively balanced distribution of crime victims between men and women over the period. The small gap could indicate that certain crime types affect both genders similarly, or that criminal behavior during this time was largely gender-neutral.

Figure 3 finds that assault is by far the most prevalent crime, comprising three-quarters of all recorded incidents and robbery and sexual violation occur much less frequently. These findings can inform policy decisions, particularly by highlighting the need for more focused interventions and resource allocation toward addressing assault, given its significant share of overall crime.

Figure 4 clearly presents that the number of women suffering from sexual violation is significantly higher than that of men. This trend reflects that the problem of sexual violence against women in society is still serious. This may be related to cultural background, gender discrimination, and social expectations of women's gender roles, which make women more likely to become targets of sexual assault in many cases. And higher number of robbery of male shows the high risk of men in property crimes. This may be related to the fact that men usually participate in high-risk social activities and travel at night more frequently, which increases their likelihood of being robbed. Figure 4 also indicates that within the assault data, the number of male victims is slightly higher than that of females, the gap between the two is not significant. This may indicate that assault behavior is not completely affected by gender, but is closely related to factors such as specific environment, social interaction and violent culture. This finding deserves attention because it reminds us that when designing prevention measures for victims of different genders, the complexity of the social environment must be taken into account.

Figure 5 shows the number of victims who experienced assault was particularly high in each age group which may reflect the high overall incidence of this crime type or may be due to the relatively broad recording criteria for assault cases in statistical reports. Thus, this type of case dominates in all age groups. Besides, except for the age groups <12 and 12-17, the

number of Robbery is always higher than Sexual Violation. This trend shows that minors are more likely to become victims of sexual violence than robbery, while adult victims are more likely to face property violations in their daily lives. This may be related to the more vulnerable social status of minors and the specific context of violent crimes.

Figure 6 illustrates that the number of female victims is significantly higher than that of male victims between the ages of 18-44 and the number of female victims rises sharply between the ages of 12-17 and 18-24, reflecting that women in this age group are more likely to become victims. The number of female victims begins to decline significantly from the age of 35-44, indicating that the number of female victims decreases sharply with age and the number of male victims fluctuates relatively less in these age groups, showing a more stable trend. This may be related to the lifestyle, social circle, and concern for personal safety of women in this age group, indicating that they may take more precautions to protect themselves. In contrast, the number of male victims fluctuates relatively less in these age groups, showing a more stable trend. This may indicate that certain types of violence or crimes tend to target young women, while the age distribution of male victims is more balanced.

Overall, this analysis highlights the intricate dynamics of crime victimization across various age groups and genders. The data indicates that young adults, particularly those aged 18 to 44, are at a heightened risk of victimization, likely influenced by lifestyle choices and societal pressures. In contrast, the under-12 and over-65 age groups benefit from protective factors such as parental supervision and stable lifestyles, leading to lower rates of victimization. Regarding gender, crime impacts both men and women similarly; however, certain types of crime, notably sexual violence, disproportionately affect women. This disparity reflects ongoing societal issues such as gender discrimination and entrenched cultural norms. Moving forward, it is essential to consider how shifts in social attitudes and law enforcement practices may further influence these trends. By integrating a temporal analysis that accounts for such factors, we can develop more nuanced preventive strategies tailored to the needs of different demographic groups. This approach can ultimately contribute to a safer community environment by fostering targeted interventions aimed at reducing victimization rates across the spectrum of crime.

5 Weaknesses and next steps

The report has several weaknesses that may compromise the depth and validity of its findings. Firstly, the data utilized lacks specified addresses, resulting in the inclusion of occurrences reported outside the City of Toronto limits, as well as incidents without verified locations. This absence of geographical specificity may impede the ability to draw accurate conclusions regarding crime patterns within the city itself. Furthermore, although the analysis spans multiple years, it does not adequately account for temporal factors that could influence crime dynamics. Specifically, the study overlooks potential changes in crime patterns, shifts in social attitudes, and variations in law enforcement practices over time, all of which may significantly affect victimization rates. Incorporating a temporal analysis would provide a more nuanced

understanding of the evolving nature of crime and victimization, thereby enhancing the overall rigor of the study's conclusions.

Next, we can consider adding geographic identifiers to the data to remove incidents without specific addresses as much as possible. This can be achieved by working with local law enforcement agencies or other data providers to obtain more detailed geographic information. In addition, time series analysis methods can be introduced into the analysis to explore changes between different years, especially factors related to changes in social attitudes and law enforcement practices. Consider using ARIMA models or other time series analysis methods to assess the impact of time on crime trends which is mensioned in Box and Jenkins (1976).

6 Appendix

- **id**: Unique row identifier for Open Data database
- **REPORT_YEAR** : Year crime was reported from 2014 to 2023
- **CATEGORY** : Crime category:
  - Crimes Against the Person
- **SUBTYPE** : Crime category subtype:
  - Assault
  - Sexual Violation
  - Robbery
  - OtherAssault Peace Officer
- **ASSAULT_SUBTYPE** : Subtypes of assault (see **?@sec-appendixA** for details)
  - Assault Peace Officer
  - Other
  - N/A: Not recorded
  - Aggravated Peace Officer
  - Assault Peace Officer Weapon/Bodily Harm
  - Assault Resist Arrest
- **SEX** : Sex of identified victim
  - Female
  - Male
- **AGE_GROUP** : Age group of identified victim
  - Adult
  - Youth
  - Child
  - Unknown
- **AGE_COHORT** : Age cohort of identified victim (see **?@sec-appendixA** for details)
  - <12: the age of victim less than 12 years old
  - 12 to 17: the age of victim lbetween 12 and 17 years old
  - 18 to 24: the age of victim lbetween 18 and 24 years old
  - 25 to 34: the age of victim lbetween 25 and 34 years old
  - 35 to 44: the age of victim lbetween 35 and 44 years old
  - 45 to 54: the age of victim lbetween 45 and 54 years old
  - 55 to 64: the age of victim lbetween 55 and 64 years old
  - 65+: the age of victim larger than 65
  - Unknown
- **COUNT_** : Count of identified victims

References

Ainsworth, P. 2001. *Victimology: Research, Policy and Activism*. Springer. https://doi.org/10.1007/978-1-4020-8923-3.

Box, George E. P., and Gwilym M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day.

Chappell, D., and S. Maggard. 2004. "Gender Differences in Crime Victimization: A Review of the Literature." *Journal of Criminal Justice* 32 (5): 391–404. https://doi.org/10.1016/j.jcrimjus.2004.08.003.

Duncan, M. 2017. "Young People and Crime: A Critical Analysis of Victimization in Urban Contexts." *Journal of Youth Studies* 20 (5): 688–702. https://doi.org/10.1080/13676261.2017.1349712.

Farrall, S., and D. Gadd. 2005. "The Meaning of Victimization: A Qualitative Study of the Experiences of Victims." *British Journal of Criminology* 45 (1): 1–21. https://doi.org/10.1093/bjc/azh071.

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.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Open Data Dataset. n.d. City of Toronto. https://open.toronto.ca/catalogue/.

Piquero, A. R. 2018. "Crime and Age: What the Data Tell Us." *Journal of Criminal Justice* 56: 1–9. https://doi.org/10.1016/j.jcrimjus.2018.04.001.

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. 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://CRAN.R-project.org/package=dplyr.

Wilcox, P., K. C. Land, and K. Hunt. 2003. "Explaining the Relationship Between Crime and Victimization: A Cross-National Analysis." *Crime and Delinquency* 49 (4): 465–90. https://doi.org/10.1007/s10940-003-0561-7.

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.