Are some women more susceptible to crime than others in Toronto?*

A visualization of crime rates over the years, and the ages they affect most.

Sehar Bajwa

January 26, 2024

Gender-based violence is a grave societal concern, inflicting profound physical, emotional, and psychological harm, and demanding urgent attention and robust intervention strategies. This study endeavours to examine annual trends in the incidence of crimes against females in Toronto, focusing on both age demographics and the nature of the offences. Leveraging data extracted from the Toronto Police Report on Victims of Crimes, the analysis reveals this paper establishes that assault victimization rates have never been higher and that females from ages 25 to 34 are most susceptible to crime. This gender-specific research could serve as a critical foundation for policymakers, furnishing insights that can inform the development of targeted policies tailored to reduce the specific risks faced by females in the city.

Table of contents

1 Introduction

Gender-based violence is perhaps the most widespread and socially tolerated of human rights violations (Johnson 2006).In Toronto and elsewhere worldwide, violence against women impedes women's equality, rights and fundamental freedom. Examining this crisis through analysis like this study supports policymakers in developing and evaluating measures to prevent and eliminate violence against women, and increase public awareness of the nature and extent of violence against women. On a more personal level, the motivation behind this research stems from my own position as a young woman living alone in this city. I wanted to explore how safe Toronto is and how victimization rates have altered, and also how my age impacted my vulnerability. The Victims of Crime Report (City of Toronto, 2014) published by the Toronto Police in the years 2014 through 2022 provides data on the gender and ages of victims subjected to Sexual Violence, Robberies, Assaults and other violations in the City of Toronto. However, the focus of this investigation is to examine only the female victim statistics, and the rationale behind this decision lies in the undeniable differences between male and female victimization. Previous research has consistently revealed that violence against women differs in significant ways from violence against men, such as the types of offences(Johnson 2006), the severity of the crime (Johnson 2005), and the consequence of victimization (Vaillancourt 2008). Infact, in Canada in 2019, females(106 victims per 1000 women) were violently victimized at a rate nearly double that of males(59 victims per 1000 men)(Cotter 2019).

The findings of this study support the gravity of gender-based violence: it was discovered that Assault is at an all time high, and that the reduced numbers in crimes that the Covid-19 pandemic brought have regained their pre-2019 trends. Also consistent with other papers that attribute age as a predictor of victimization(Johnson 2006), this study found that young women between the ages of 25 to 34 are most susceptible to crime. To study Female victimization in Toronto with reference to the ages of the victims and the offences they are subjected to, this paper is organised into the following sections: Data, Visualization, Discussion and References. In the data section, I provide a synopsis of the dataset and the variables of interest. Also included here is a quick review of the data processing, the subtypes of offences analysed, and the creation of a new Summation variable. The visualization section examines two lines of

questioning, the first being: "How have female victimization rates varied over the years?", and the second being: "Are there some age groups that are more susceptible to crime than others?". Each of these sections features relevant tables, graphs and observations. Lastly, the Discussion section is a high-level overview of the limitations of the analysis, and concluding remarks featuring the next steps for future iterations of this study.

2 Data

This section delves into the characteristics of note and outlines the process undertaken to create a ready-to-analyse dataset, including addressing missing values and standardizing variable names.

2.1 Data Overview

Published by the Toronto Police Service on the Toronto Open Data Portal , the Toronto Police Annual Statistical Report on Crime Victims(City of Toronto, 2014) covers all crimes committed against the person, including those deemed unfounded post-investigation. The entries are filtered by the reported year, and each year is associated with the following features: type of crime, age cohort of victim, gender, and counts of the exact crime. The dataset was initiated in 2014 and is marked 'updated annually', but the latest recorded year available is 2022.

With a focus on demographic insights, the dataset stands at 1111 data points across 9 variables. Age cohort is a defining feature of this dataset, and refers to one of 8 age bins that are as follows: "<12", "12 to 17", "18 to 24", "25 to 34","35 to 44","45 to 54","55 to 64","65+".Efforts were made to attribute recognizable labels to these cohorts, but there was admittedly difficulty distinguishing between the middle aged adults that arguably spanned three of the 9 age cohorts. Therefore, the original identifiers were retained.

2.2 Data Cleanup and Processing

`R` (R Core Team 2022) was the language and environment used for the bulk of this analysis, alongside the `dplyr` (Wickham et al. (2023)),`tidyverse`(Wickham et al. 2019), `janitor`(Firke 2023), `knitr`(Xie 2023) and `readr`(Wickham, Hester, and Bryan 2024) packages. The formatting packages used for the scatter plot was `pheatmap`(Kolde 20219). To enhance analytical readiness, missing values were addressed, variable names are standardized and new summation columns created.

2.2.1 Missing Data Points

The dataset featured incomplete data points, which are discounted at this time by filtering and subsequent removal. Though making analysis remarkably easier, this poses a compromise regarding validity and poses significant implications for subsequent analyses, a concern explored in detail later.

2.2.2 Categorization of Crimes

The dataset features 4 major crime types (Sexual Violation, Robbery, Assault, and Other). There is an additional data point labelled Assault subtype, providing subcategories only for this of the four major crime types. These subtypes include "Assault/Aggravating Peace Officer", "Bodily Harm", and "Resist Arrest". The subtype column has been discounted after deliberation since it skews the dataset in terms of specificity. While the first three labels are forthright and succinct, Other is vague, and trying to find it's intended definition proved a trickier task than expected. According to the glossary accompanying the Annual Police Statistical report(City of Toronto, 2014), Other criminal code violations are defined as Non-traffic Criminal Code violations that are classified as neither violent nor property violations. The label Other is retained, but this definition proves useful to mention to viewers trying to gain an absolute understanding of the data.

2.2.3 Duplicate entries

Removing the crime subtype column leaves multiple different entries for the same age group, crime and year. For example, An assault could have been recorded in 2015 against a 65 + female, but it could have been a resisted arrest, or bodily harm. Summing up the crime counts per category removes duplicate entries with matching data (with the only differentiator being crime counts). A new variable, labelled sum count was initiated for this purpose, and added as a new column.

Report Year	Crime Type	Age Cohort	Sum Count
2015	Assault	25 to 34	4
2015	Assault	35 to 44	7
2015	Assault	45 to 54	2
2015	Assault	65+	2
2015	Assault	25 to 34	1
2015	Assault	18 to 24	10

Figure 1: First 10 entries after the data has been cleaned and summed so that there is a singular entry for every crime per age group

3 Visualization

3.1 Trends In Female Victimization

A dataset featuring years and counts inherently presents the initial statistical inquiry into the temporal dynamics and trends over the observed period. Naturally, this begs the question "How have the counts of different crimes varied over the years?". To begin formulating an answer, one would envision a scatter plot with Years plotted against the number of crimes and 4 lines for each of the Crime types. Using the `group by` function in the `dplyr` package, data was grouped by the report year, and based on the type of crime. The `summarise` function then added all counts for different age groups for the crime type per year, and the resulting table is shown below.

Report Year	Crime Type	Sum Count
2014	Assault	7132
2014	Other	2618
2014	Robbery	904
2014	Sexual Violation	1810
2015	Assault	7346
2015	Other	3008

Figure 2: First 10 entries after the data has been manipulated to create one datapoint per every year for every type of crime.

A scatterplot has been created to visually represents how victimization rates have increased over the years pertaining to every crime type (see Figure 3) .

Immediate observations indicate alarmingly high levels for assault compared to other types of crime, and this could be attributed to assault being a summed group including different assault subtypes. Assault has been growing steadily since 2014 and is at an all-time high in 2022, with a slight decline that can be attributed to Covid 19 restrictions imposed on the city in 2020(Government of Canada 2021). Sexual Violation and 'Other' crime numbers were dented in 2020 and saw a resurgence in the year after, but Robbery decreased in 2021, the second year of the pandemic. A reasonable assumption could be that the first year had stricter stay-at-home policy implementations, resulting in families at home, making it significantly challenging for break-ins to go unnoticed, thus dissuading robbers from attempting them. This dataset has a relatively short, 9 year range, and is perhaps too narrow to predict with accuracy any lasting trends for the decades to follow, but it presents fascinating, magnified data of how a worldwide pandemic impacts the number reported Crime victims.
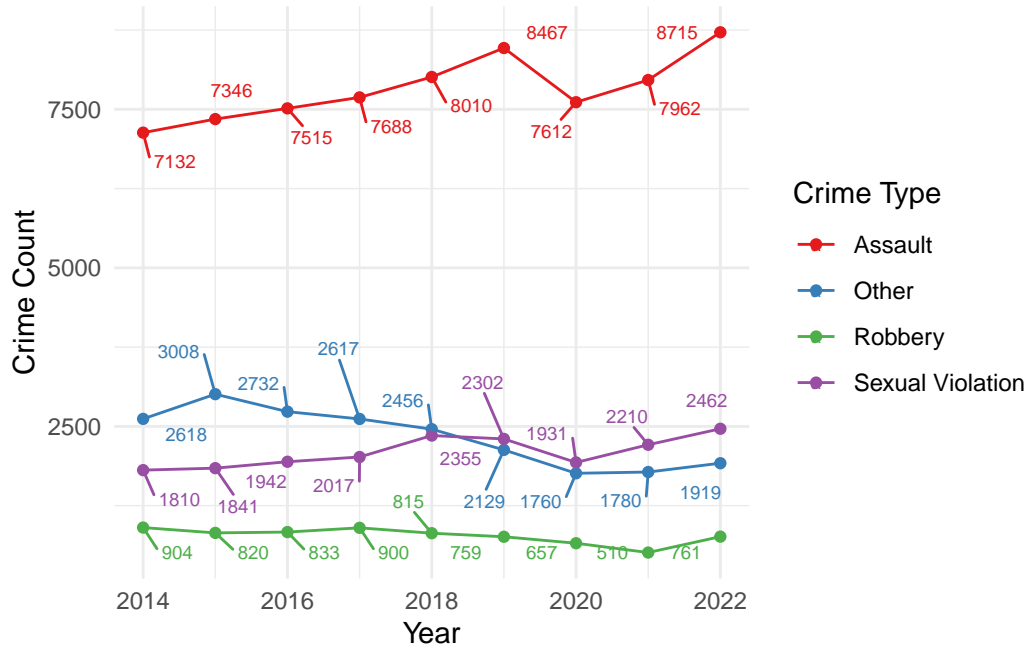
Figure 3: Scatterplot featuring the changing trends in the types of crime over the years

3.2 At-risk Age Groups

The second line of questioning begins with trying to understand which female age groups are most at risk, and resolves the underlying personal motivation outlined before. Using group and summarise again, the data was grouped by report year and based on Age cohort, and summed as before, resulting in the following table.

Report Year	Age Cohort	Sum Count
2014	<12	642
2015	<12	602
2016	<12	533
2017	<12	589
2018	<12	608
2019	<12	617

Figure 4: First 10 entries after the data has been manipulated to sum all crime counts that happen to every age group every year.

Creating a heatmap was an inspired decision for this dataset, with the rationale being a visual that was inherently self-explanatory in how it represented trends and the relationships between the variables. The original dataset was transformed from its long format (1 observation per

row, variables in different columns) to wide format (1 observation per row, all pertaining variables in that same row) by using the pivot_wider() function in the tidyr package. The dataset also needs to be converted to a numeric matrix before the heatmap can be created, and this step utilised the `as.matrix()` function to coerce all data types to numbers.

Age Cohort	2014	2015	2016	2017	2018	2019	2020	2021	2022
<12	642	602	533	589	608	617	466	572	546
12 to 17	1210	1246	1320	1394	1332	1408	913	1057	1439
18 to 24	2661	2663	2688	2617	2749	2525	2170	2133	2368
25 to 34	3174	3459	3380	3422	3523	3505	3330	3526	3813
35 to 44	2055	2144	2139	2200	2357	2433	2258	2284	2547
45 to 54	1562	1638	1593	1611	1627	1580	1439	1424	1530
55 to 64	695	737	824	835	824	933	807	867	942
65+	465	526	545	554	616	656	577	599	672

Figure 5: Pivoted data table showing crime trends in a wide format so that a heatmap can be created.

The `pheatmap` package (an appropriate abbreviation for what was indeed a 'pretty heatmap') generated a visually appealing and customisable chart that allowed for clustering and custom labels. (see Figure 6)

Immediately noticeable is that women from the ages of 25 to 34 have been most susceptible to crimes regardless of year, with an all-time high in 2022. The other end of the spectrum is that children have never been safer than they were during COVID-19, staying at home. Moving vertically away from this high-risk group signals a reduced number of crime victims in both directions. This provides a possible inference that converging towards the transitional years that are 25 to 34 years of age, women are most in jeopardy, but is far from conclusive.

4 Discussion

4.1 Limitations

4.1.1 Missing Data

The presence of missing data points in the Age Group column poses a substantial limitation to the analysis's robustness, and so were disregarded in this study. These gaps in the data can lead to biased analyses, as the omitted information may disproportionately affect certain age groups. Consequently, the findings may lack generalizability, inhibiting the ability to draw comprehensive and reliable conclusions regarding trends and risk groups. This could be

Crime Rate Victimzation by age groups

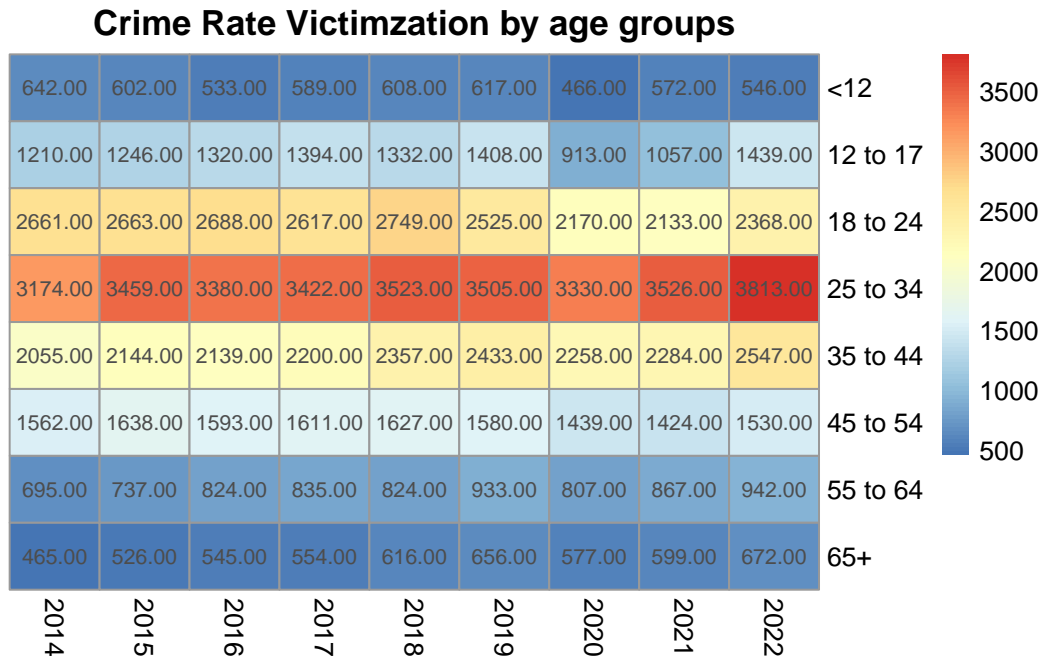|										
642.00	602.00	533.00	589.00	608.00	617.00	466.00	572.00	546.00	<12	
1210.00	1246.00	1320.00	1394.00	1332.00	1408.00	913.00	1057.00	1439.00	12 to 17	
2661.00	2663.00	2688.00	2617.00	2749.00	2525.00	2170.00	2133.00	2368.00	18 to 24	
3174.00	3459.00	3380.00	3422.00	3523.00	3505.00	3330.00	3526.00	3813.00	25 to 34	
2055.00	2144.00	2139.00	2200.00	2357.00	2433.00	2258.00	2284.00	2547.00	35 to 44	
1562.00	1638.00	1593.00	1611.00	1627.00	1580.00	1439.00	1424.00	1530.00	45 to 54	
695.00	737.00	824.00	835.00	824.00	933.00	807.00	867.00	942.00	55 to 64	
465.00	526.00	545.00	554.00	616.00	656.00	577.00	599.00	672.00	65+	
2014	2015	2016	2017	2018	2019	2020	2021	2022		

Color scale: 3500, 3000, 2500, 2000, 1500, 1000, 500

Figure 6: Heatmap using color intensity to visually depict the number of crimes.

addressed by obtaining additional, completed data to take the disregarded entries back into account.

4.1.2 Uneven Bin Widths for Age groups

Bin width sizes for the age cohorts were another noticeable issue. The range was lower for the adolescent (12 to 17) and young adult(18 to 24) groups, being just 7 years. It was 10 years for all four of the middle-aged groups. As for the fringe ages that were unbounded, <12 and >65, there is no discrete way of making sure what ages within those ranges were accounted for. The absence of bin width standardisation would mean bar charts with ages on the horizontal axis had differing bar widths, and that resulting inferences would not be truly representative of the data.

4.1.3 Unreported Crime cases

Lastly, one must note that this dataset only considers victims who have come forward to the Toronto Police to file official complaints, not the actual number of crimes committed in the city. The General Social Survey(Canada" 2019) conducted by Statistics Canada enables self-reporting of crime through a series of independent, annual, voluntary, cross-sectional surveys.

According to the Victimization cycle of the GSS conducted in 2009, 69% of violent victimizations and 71% of robberies are not reported to the police. This means that this Victims of Crime dataset published by the Toronto Police only reflects a fraction of the crimes committed. It is also prudent to mention that the GSS has the limitation of reliance upon respondents for accurately recalling and reporting the events, but in conjunction with the Annual Statistical report, could provide a more authentic count of the crimes in Toronto. As an aside, the GSS confers with the results of this study posits that being young remains a strong predictor of victimization against women, and that younger females are more vulnerable to violent crime.

4.2 Next steps

This study provides a robust starting point for analysing female victimization in Toronto, and there are several apparent improvements that can be made and considered for future iterations. In addition to resolving the limitations above through complete data, collection socio-demographic information about these victims will provide additional frames for analysis. It would also be a worthwhile investigation to compare this report with the Major Crime Indicators report, and hone in on which neighbourhoods are more at risk, and the months/times of day that record the highest criminal activity in the city.

References

Canada", "Statistics. 2019. *General Social Survey - Canadians' Safety (GSS)*. https://www2
  3.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=1235019.

City of Toronto,. 2014. "Open Data Dataset." *City of Toronto Open Data Portal*. Open data
  catalogue. https://open.toronto.ca/dataset/police-annual-statistical-report-victims-of-
  crime/.

Cotter, Adam. 2019. "Criminal Victimization in Canada, 2019." *Criminal Victimization in
  Canada, 2019.* https://www150.statcan.gc.ca/n1/pub/85-002-x/2021001/article/00014-
  eng.htm.

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

Government of Canada, Statistics Canada. 2021. "After Five Years of Increases,police-
  Reported Crime in Canada Was down in 2020, but Incidents of Hate Crime Increased
  Sharply." *The Daily* -. https://www150.statcan.gc.ca/n1/daily-quotidien/210727/dq210
  727a-eng.htm.

Johnson, Holly. 2005. "Assessing the Prevalence of Violence Against Women in Canada."
  Statistical Journal of the United Nations Economic Commission for Europe 22 (January):
  225–38. https://doi.org/10.3233/SJU-2005-223-404.

———. 2006. *Measuring Violence Against Women: Statistical Trends 2006.* https://www150
  .statcan.gc.ca/n1/pub/85-570-x/85-570-x2006001-eng.htm.

Kolde, Raivo. 20219. *Pheatmap: Pretty Heatmaps.* https://cran.r-project.org/web/packages
  /pheatmap/index.html.

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

Vaillancourt, Roxan. 2008. "Gender Differences in Police-Reported Violent Crime in Canada,
  2008." *Canadian Centre for Justice Statistics Profile Series*, January, 26. https://doi.org/
  10.3233/SJU-2005-223-404.

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.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data.*
  https://readr.tidyverse.org.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.*
  https://cran.r-project.org/web/packages/knitr/index.html.