Demographic Insights into Urban Crime: An Analysis of Arrest Data in Toronto

pu yuan

2024-01-21

A Statistics Canada report shows a significant increase in the Violent Crime Index in 2022, reaching its highest point nationally since 2007. This increase is mainly due to a 15% increase in robberies and an 8% increase in homicides. The number and severity of crimes reflect a worrying trend. Therefore, it is important to reduce crime by delving into crime data to understand specific patterns of criminal activity and the demographic characteristics of those involved. The purpose of this paper is to deepen our understanding of this issue by analyzing Toronto's Arrested and Charged Persons dataset. Through this analysis, we want to reveal patterns in the demographic characteristics of persons arrested in recent years, to provide an important basis for the development of targeted law enforcement strategies to help curb rising crime rates in Canadian cities.

Table of contents

1 1 Introduction

Data from Statistics Canada suggests that the crime situation in Canada in 2022 witnessed a disturbing increase in violent activity. There was a significant increase of 6% in violent crime scene investigations, with substantial increases in specific crimes such as robbery (up 15%), extortion (up 39%) and homicide (up 8%) (Fox 2023). In addition, the number of reported violent incidents rose dramatically to 243,2022 in 531 and homicides reached their highest level in 30 years with 874 reported. There is an urgent need for in-depth analysis of crime data to develop effective crime prevention and community safety strategies. (Statistics Canada 2022) While national statistics provide a macro view of crime trends, there is an urgent need to understand the local dynamics of criminal activity. Focusing on the City of Toronto, this paper aims to provide a more nuanced analysis of the complex patterns of crime. Specifically, it examines the demographics of arrested and charged individuals as recorded in the Annual Police Statistical Report - Arrested and Charged Persons.

In this paper, we analyze a dataset from opendatatoronto using R (R Core Team 2020) focusing on arrested and charged persons in Toronto, with the aim of revealing demographic trends in criminal activity. The study provides a comprehensive analysis of the above dataset, focusing on the years 2018 to 2022. Our analysis of the dataset reveals patterns in the age, gender, and crime type of arrestees, providing insight into the demographic groups that are more frequently involved in specific crimes. We used tables, scatterplots, and bar charts for visualization, revealing that certain age groups and genders are predominantly associated with specific crime types, providing insights into which age groups and genders are more frequently involved in certain criminal activities. These findings are critical to reported national crime trends and provide law enforcement with valuable insights that may help develop more targeted and effective policing strategies in Toronto. However, we must confront the limitations of this dataset. We must take into account the potential for bias and the fact that the data may not be fully representative of the entire spectrum of criminal activity due to systemic issues in law enforcement practices.

2 2 Data

The dataset for this analysis is the "Annual Police Statistical Report - Arrested and Charged Persons" (City of Toronto 2023a) from the Toronto Open Data Center. This dataset provides the total number of persons arrested and charged by the Toronto Police, with detailed breakdowns by various demographic and crime-related categories. It contains the following variables. 2.1 Key Variables in the Dataset ARREST_YEAR: The year in which the arrest was made. This variable is useful in understanding temporal trends in crime and arrest patterns. DIVISION: The geographic division where the crime occurred. It is crucial for spatial analysis and understanding the distribution of crime across different areas of the city. HOOD and NEIGHBOURHOOD: The neighborhood where the crime took place. This variable allows

the spatial analysis to be more precise. SEX: The sex of the individual arrested. Analysis of this variable can reveal potential gender patterns in criminal activity. AGE_COHORT and AGE_GROUP: These variables provide information on the age of the arrestee and categorize them into cohorts and broader groups (youth or adult), contributing to a deeper understanding of age-related arrest trends. CATEGORY and SUBTYPE: The crime category and its specific subtype, allow for a detailed analysis of the various types of crimes. ARREST_COUNT: The count of arrests, it is essential for quantifying the data.

## 2.1 2.2 Potential Similar Datasets and Reasons for Not Using Them

There is a similar dataset called "About Police Annual Statistical Report - Calls for Service Attended" (City of Toronto 2023b), also published by the Toronto Police Service on Toronto Open Data. However, in that dataset, geographic location is the dominant factor, with many geographic variables, but crime types and demographic characteristics of offenders are not mentioned. In this paper, we want to find the relationship between offender demographic characteristics and crime, so this dataset is not suitable for us.

2.2 2.3 Data Description

For crime incidents, "NSA" was filled in if the location of the occurrence was not identified. No new variables were constructed; the analysis focused on existing variables in the dataset.

2.3 2.4 Data Visualization and Summary

Our analysis begins with the distribution of arrestees by age group. Understanding the age dynamics in crime statistics is critical, as it has the potential to reveal prevalent age groups that are more susceptible to criminal activity. This can be useful for law enforcement agencies in developing preventive measures and allocating resources effectively.

Figure 1: Age Cohort Distribution of Arrests
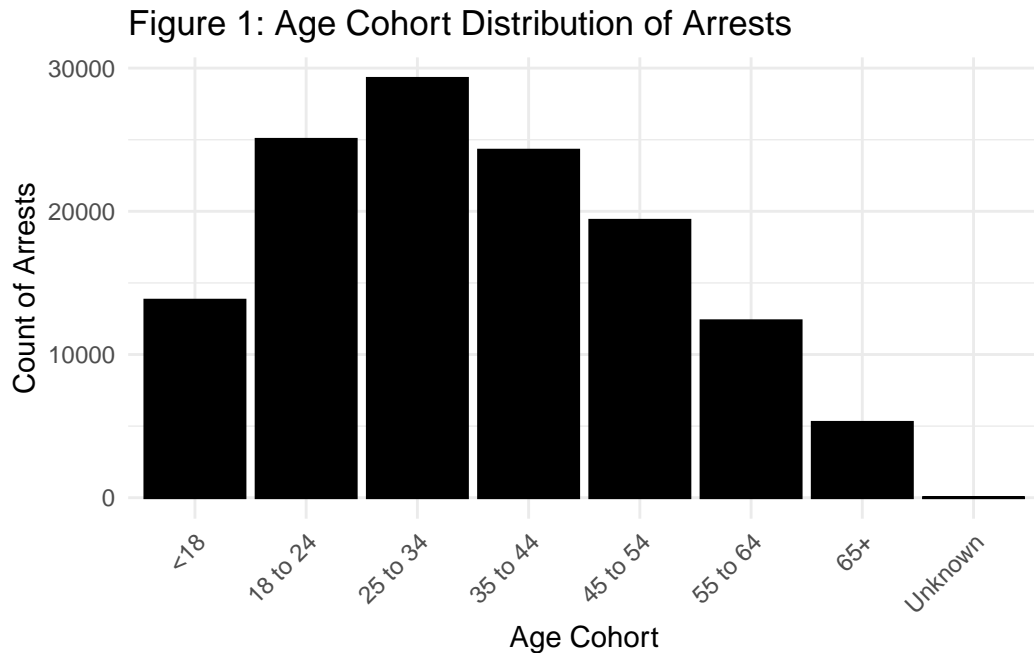
The first graph, Figure 1 in our series is a histogram depicting the age distribution of arrestees. It is generalized using ggplot(Wickham 2016). The histogram categorizes arrestees into age groups such as "18 to 24," "25 to 34," "35 to 44," and so on. The height of each bar in the histogram correlates with the number of arrests in that particular age group. Preliminary observations indicate that the frequency of arrests is higher in the "25 to 34 years" age group, followed by the "18 to 24 years" and "35 to 44 years" age groups. This trend highlights the need to develop targeted policing strategies for these age groups.

However, it is crucial to relate these findings to the broader composition of the population. Without considering the age distribution of the total population, a bar chart perspective alone is limited. Additionally, while the bar chart effectively highlights age groups with higher arrest rates, it does not distinguish between the nature or type of crime involved. This information gap limits the depth of our analysis and necessitates further examination of the specific types of offenses associated with each age group.

After examining age groups, we turned our focus to an in-depth analysis of the gender distribution of arrests. Through pie charts, this portion of the study not only quantifies the number of arrests by gender, but also expresses these numbers in actual numbers and percentages. Pie charts are an intuitive visual tool for comparing the frequency of arrests for males and females, thereby revealing gender differences in criminal behavior.
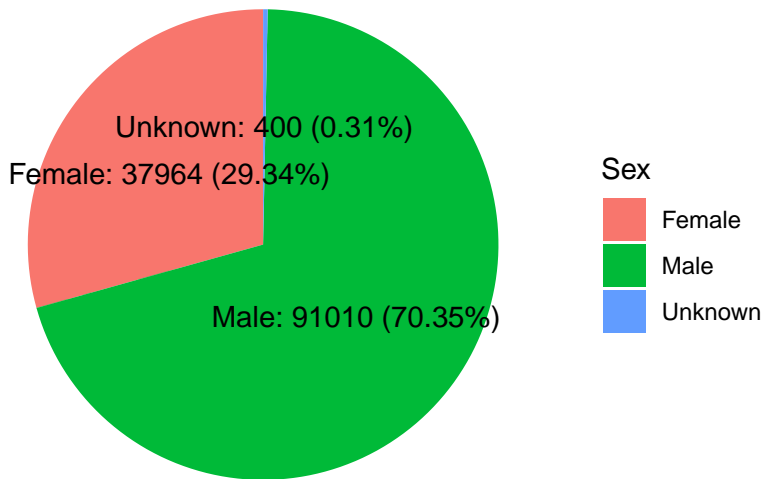
Figure 2: Distribution of Arrests by Sex

Figure 2 is a pie chart made with ggplot(Wickham 2016) and Dplyr(Wickham et al. 2022). As can be seen from the chart, there is a clear dominance of males among those arrested, with males accounting for 70.35% of the 91,010 recorded cases. This preponderance highlights a clear gender imbalance in criminal activity, with male offenders far outnumbering female offenders. The number of arrests of females is about half of the number of arrests of males, which highlights the significant gender disparity in crime. An interesting aspect of the chart is the inclusion of an "unknown" category, which represents 0.31% of arrests. This category indicates that the gender of the offender is either binary or unrecorded. Males are overrepresented in criminal arrests compared to females. This discrepancy calls for gender-specific strategies in crime prevention and policing that may need to be adopted by law enforcement agencies. The overrepresentation of males in criminal activities may reflect underlying social or environmental factors that merit separate research and investigation.

The final figure in our sequence is a stacked bar chart showing the distribution of arrestees across different crime categories, further broken down by gender. This visualization allows us to examine the relationship between crime type and the gender of arrestees. Each bar in the chart represents a particular crime category, and the paragraphs in the bars show the proportion of male and female arrestees in that category.

Figure 3: Distribution of Arrests by Sex and Crime Category
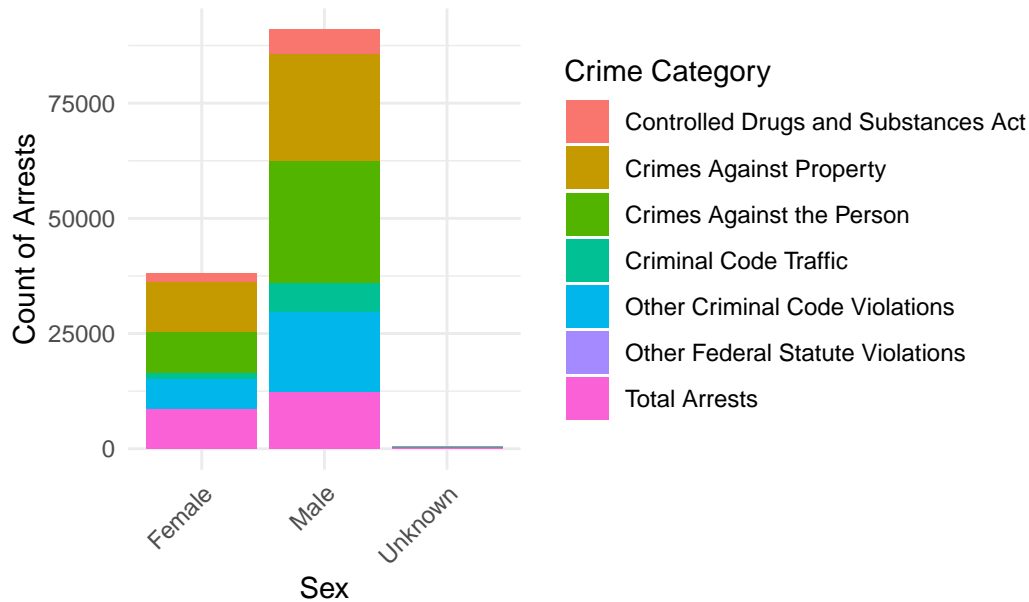
It is clear from figure 3 that males accounted for a significantly higher proportion of "crimes against property" and "crimes against persons" compared with females. This observation suggests that males tend to be more frequently involved in these types of crimes, as they often involve an element of aggression or direct confrontation. However, this interpretation must be treated with caution, as correlation does not imply causation. While the data suggest that males are involved in these crime categories at a higher rate, it should not be generalized to infer an inherent male propensity for aggression. In other crime categories, women were arrested in relatively equal proportions to men. This parity suggests that gender differences in participation in crime may not be as pronounced, with the exception of two specific categories, property crime and personal crime.

A category labeled "total number of arrests" exists in the dataset. However, this category is minimally represented in the charts and the dataset does not provide a clear definition of what is included in this category. The "total arrests" category was not considered further in this focused analysis due to its ambiguity and lack of detailed explanation in the raw data.

3 3 Conclusion

The in-depth analysis of Toronto's arrest data in this paper, focusing on age and gender distributions as well as crime categories, provides important insights into patterns of criminal behavior in the city.

The preponderance of younger age groups in arrest records underscores the urgent need for community engagement and intervention programs targeting these populations. Gender anal-

ysis revealed a higher rate of arrests among males, especially in certain types of crime, which calls for a gender-sensitive approach in crime prevention and law enforcement.

As cities like Toronto continue to grow and grapple with the complexities of urban crime, studies such as this one provides an effective aid to more targeted and efficient crime prevention and law enforcement approaches.

4 4 Reference

Gelfand, S. (2020). Opendatatoronto: Access the City of Toronto Open Data Portal. Wickham, H. (2016). Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org Wickham, H., François, R., Henry, L., & Müller, K. (2022). Dplyr: A Grammar of Data Manipulation. Retrieved from https://CRAN.R-project.org/package=dplyr R Core Team. (2020). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/ City of Toronto. (2023). Police Annual Statistical Report - Calls for Service Attended. Retrieved from https://open.toronto.ca/dataset/police-annual-statistical-report-calls-for-service-attended/ City of Toronto. (2023). Police Annual Statistical Report - Arrested and Charged Persons. Retrieved from https://open.toronto.ca/dataset/police-annual-statistical-report-arrested-and-charged-persons/ Statistics Canada. (2022). Police-reported crime statistics in Canada, 2022. Retrieved from https://www150.statcan.gc.ca/n1/daily-quotidien/230727/dq230727b-eng.htm Fox, C. (2023, July 28). Toronto saw 15 per cent increase in its violent crime index in 2022, more than three times the jump seen nationally: StatsCan. CP24. Retrieved from https://www.cp24.com/news/toronto-saw-15-per-cent-increase-in-its-violent-crime-index-in-2022-more-than-three-times-the-jump-seen-nationally-statscan-1.6498269

City of Toronto. 2023a. "Police Annual Statistical Report - Arrested and Charged Persons." https://open.toronto.ca/dataset/police-annual-statistical-report-arrested-and-charged-persons/.

———. 2023b. "Police Annual Statistical Report - Calls for Service Attended." https://open.toronto.ca/dataset/police-annual-statistical-report-calls-for-service-attended/.

Fox, Chris. 2023. "Toronto Saw 15 Per Cent Increase in Its Violent Crime Index in 2022, More Than Three Times the Jump Seen Nationally: StatsCan." *CP24*, July. https://www.cp24.com/news/toronto-saw-15-per-cent-increase-in-its-violent-crime-index-in-2022-more-than-three-times-the-jump-seen-nationally-statscan-1.6498269.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Statistics Canada. 2022. "Police-Reported Crime Statistics in Canada, 2022." https://www150.statcan.gc.ca/n1/daily-quotidien/230727/dq230727b-eng.htm.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.