My title*

My subtitle if needed

Yimiao Yuan

January 22, 2024

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

1 Introduction

You can and should cross-reference sections and sub-sections.

The remainder of this paper is structured as follows. Section 2….

2 Data

This report uses "Police Annual Statistical Report - Arrested and Charged Persons" as the primary dataset, which obtained from the City of Toronto's Open Data Portal ((**CityofToronto?**)). This dataset was accessed through 'R' (R Core Team (2022)) using 'opendatatoronto' package ((**opendatatoronto?**)) and saved using 'readr' ((**readr?**)). The data is published by Toronto Police Services and refreshed on a yearly basis, with last updated on January 14, 2024. This dataset provides a total count of persons who have been arrested and charged over a period of time. The data consists 129374 rows and 11 variables, each indicating a detailed summary of the crime with the year the arrest was made, location of the crime, victim informations and count of arrested and charged person. The following table shows a preview of the primary dataset.

Since it is difficult to observe through 129374 rows with 11 variables, this reports will only observe and analyze through specific aspects. The dataset contains information about victims, so one way is to focus on the number of arrested and charged person through the victims aspects. In addition, it stated the geographic division and the neighbourhood of the crime location, so we can further investigate the dataset on the division perspective. The analysis

---

*Code and data are available at: https://github.com/YimiaoYuan09/Arrested_and_Charged_Persons.git

Table 1: Preview of the arrested and charged persons dataset

X_id	ARREST_YEAR	DIVISION	HOOD_158	NEIGHBOURHOOD_158
1	2019	D14	83	Dufferin Grove (83)
2	2022	D12	30	Brookhaven-Amesbury (30)
3	2018	D14	165	Harbourfront-CityPlace (165)

SEX	AGE_COHORT	AGE_GROUP
Female	25 to 34	Adult
Male	<18	Youth
Male	18 to 24	Adult

CATEGORY	SUBTYPE	ARREST_COUNT
Other Criminal Code Violations	Other	1
Crimes Against the Person	Assaults	2
Other Criminal Code Violations	Other	1

data was cleaned to exclude the "unknown" information on the sex and age_cohort column and "No Specified Address (NSA)" on the division column to have more specific and detailed data. The cleaned data will be analyzed and performed using R (R Core Team (2022)) with 'tidyverse' ((**tidyverse?**)), 'dplyr' ((**dplyr?**)), 'ggplot2' ((**ggplot2?**)) and 'here' ((**here?**)).

The dataset that this report will be using to analyze through the report consists five variables: arrest year, division, victims' age and sex, and the count of the arrested and charged person. Working with these variables, we can generate a great deal of graphs to gain a better understand of how the crime cases changed through time. Figure 1 shows the total number of arrested and charged person each year.

According to this bar chart (Figure 1), we can see that the number of arrested and charged person is steadily decreasing each year. The number of arrested and charged person in 2014 is around 80000, which is the highest, then it is slowly decreasing from 2015 to 2017, and is slightly increasing from 2020 to 2022. For a more detailed analysis, we want to explore how this trend changes in specific divisions. The City of Toronto is divided into different police divisions, so we can take advantage of this to see how the number of arrested and charged person are changing in each division.

From Figure 2, we can see that the totals for D51 and D52 are significantly higher than the other divisions, with a total of nearly 60000 people, suggesting that more people are committing crimes in these two areas and that it is more dangerous for people living in Toronto. In comparison, D13, D33 and D53 have a more favorable performance, with overall offenses around 20000 over a nine-year period. Overall, the number of arrested and charged person in
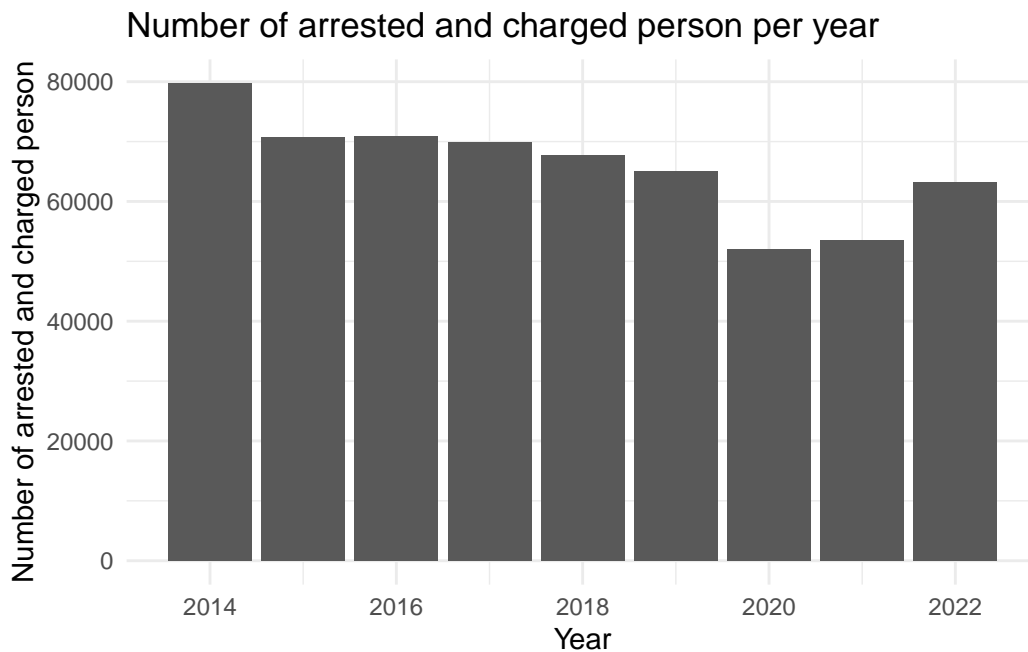
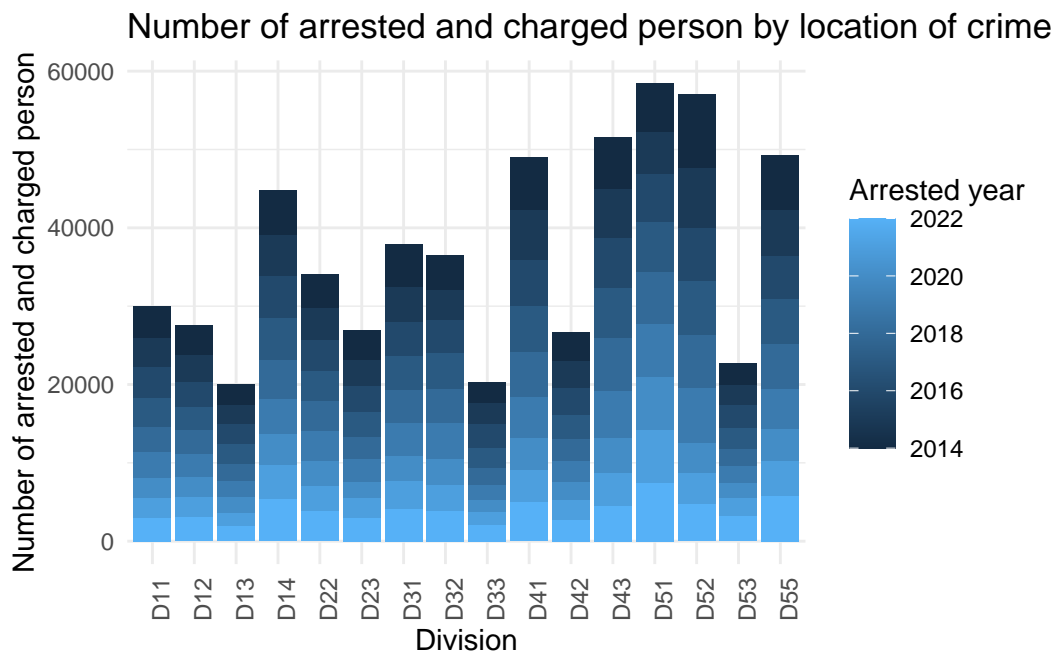Figure 1: Number of arrested and charged person each year

Figure 2: Number of arrested and charged person per division each year

each district shows a downward trend in each year, but in D14, D33, D43 and D53, there is a more significant increase in 2022.

After understanding the data in terms of overall crime, we can also analyze data based on victim information, which shown in Figure 3 and Figure 4.

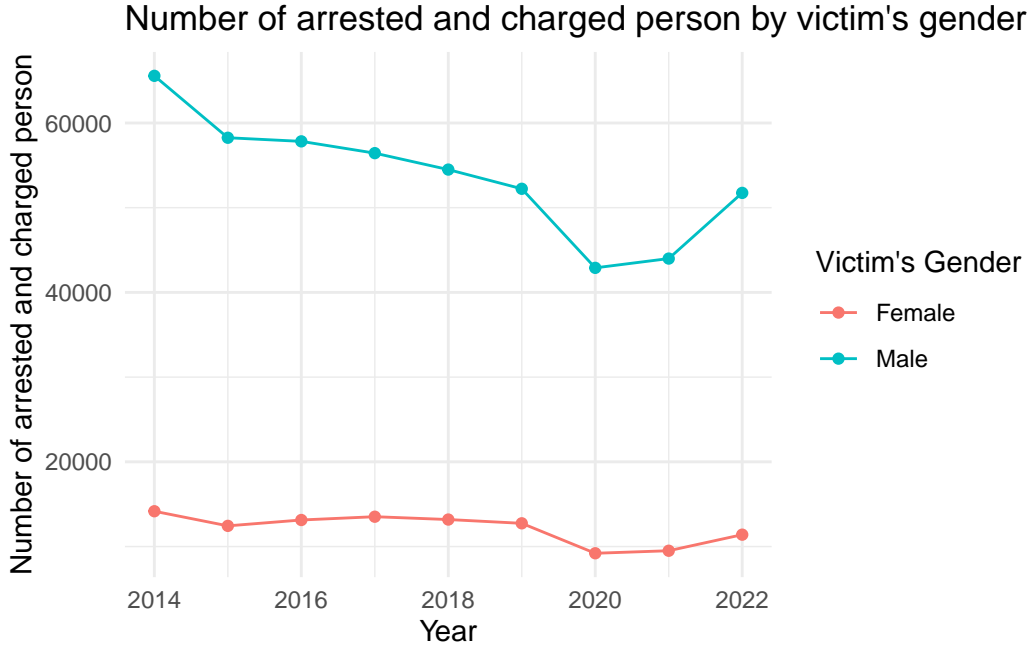## Number of arrested and charged person by victim's gender

Figure 3: Number of arrested and charged person by victim's gender each year

Based on Figure 3, among all offenders, the number of people who choose male as victims is shockingly much higher than the number of people who choose female as victims. Victims of different genders show different trends. The number of people who choose female as victims drops from 2014 to 2015, remains almost constant from 2015 to 2019, then decreases sharply from 2019 to 2020, then starts to rise slightly from 2020 to 2022. The number of people who choose male as victims shows a continuous downward trend from 2014 to 2020, and then increases dramatically from 2020 to 2022. It is worth noting that for all genders, the number of victims decreases between 2019 to 2020 and then both increases from 2020 to 2022.

Besides victim's gender, we can also analyze data through victim's age.

From Figure 4, we can see that most of the victims are centered between the ages of 18 to 44, which is in the young and middle-aged demographic. In contrast, victims over the age of 65 are the least. In general, most people who are arrested and charged choose adults and middle-aged people to commit crimes, they don't bother with the elderly and minors.

As can be seen from the four graphs, the overall number of crimes has been decreasing from year to year, but the number of crimes in several divisions has remained high. The police in
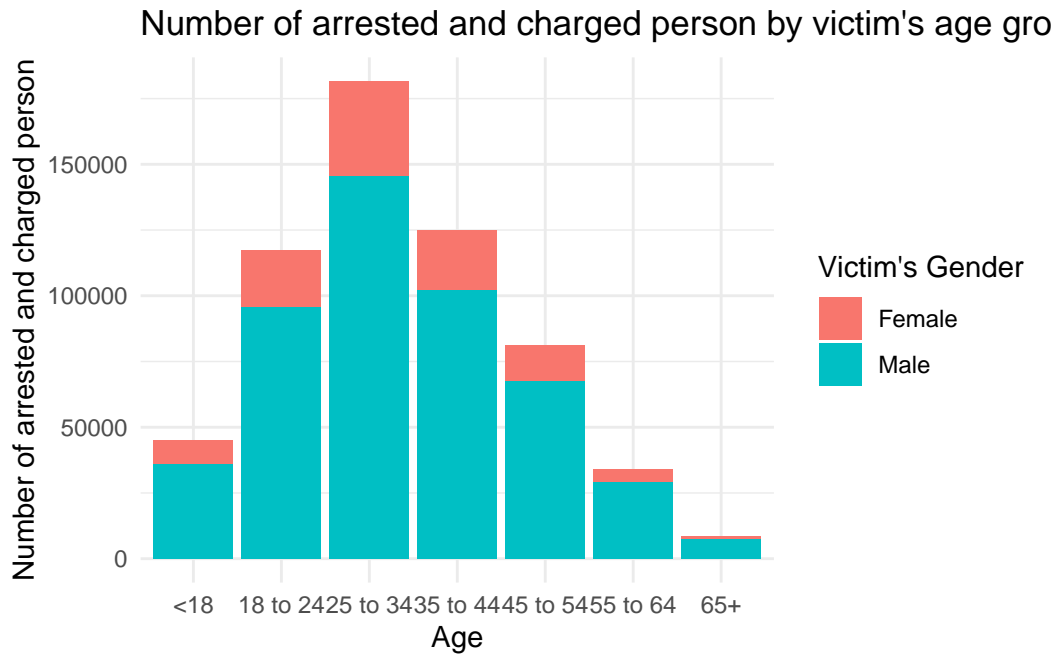
Figure 4: Number of arrested and charged person by victim's age group each year

these areas should strengthen their management or try a variety of preventive measures and warning notices to improve the current situation. It can also be seen that the victims tend to be young and middle-aged, perhaps because that they often leave early and return home late, and go to various crowded places. For further investigation, the data can be improved by looking at the time of the crime, the cause of the crime, or other factors to get more results and more specific conclusions.

References

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