Analysis of Major Crimes in Toronto And Suggestions on How to Protect Ourselves*

Boxuan Yi

2024-01-20

Safety is one of the most important factors that people consider when deciding where to live and go. This report uses data from Open Data Toronto to analyze the occurrences of various types of crimes in different regions and time of Toronto, as well as the trend over the past ten years. The result shows that Downtown Toronto and evenings & noon have a relatively high occurrences of crimes, which could be used as a suggestion on both individual and government dicision-making.

Table of contents

1 Introduction **1**

2 Data **2**
2.1 Data Source and Software 2
2.2 Data Collection 2
2.3 Data Analysis and Visualizations 3

3 Conclusion **4**

References **6**

1 Introduction

On Dec 25th, 2023, the Christmas Day, a news has raised huge concerns among people — a man was seriously injured after stabbing in downtown Toronto. This was not the first such incident on that block in December; unfortunately, such distressing news has become more and

*Code and data in this analysis is available at: https://github.com/Elaineyi1/Sta302_Major_Crime_Indicators

more prevalent nowadays. As the pandemic comes to an end, there is a noticeable increase in people returning to the office and going out more frequently than in previous years. Naturally, public safety and private security have become particularly crucial, leading to a significant question: How can we proactively prevent such incidents from affecting ourselves and our loved ones?

Major crimes statistics is an important tool for resource allocation and law enforcement strategy. It also provides the public a clear picture of the safety and helps individuals make informed decisions, influencing choices such as when to go out and where to buy a house. For this paper, I will use open-access data, titled 'Major Crime Indicator' from the OpenDataToronto (Gelfand 2022) to analyze the distribution of major crimes—classified as Assault, Break and Enter, Auto Theft, Robbery, and Theft Over—over the past ten years. My analysis will also delve into practical actions ordinary citizens and the government can do to enhance safety and efficiency.

In this paper, I discuss the data in Section 2, including data source and software (Section 2.1), data collection (Section 2.2), data analysis and visualizations (Section 2.3). In Section 3, I conclude that Toronto-East York & evening and noon experience the highest numbers of crimes. Among the five major crime categories as mentioned, 'Assault' records the highest count, with 'Break and Enter' and 'Auto Theft' following closely.

2 Data

2.1 Data Source and Software

The data utilized throughout this paper, in csv format, was obtained from Toronto Police Services on the City of Toronto Open Data Website (Gelfand 2022), with the title "MAJOR CRIME INDICATORS".

The dataset will be processed and analyzed in R (R Core Team 2022) using packages Dplyr (Wickham et al. 2022), Tidyverse (Wickham et al. 2019), Here (Müller 2020), Leaflet (Cheng et al. 2023), and Janitor (Firke 2021). Visualizations including tables and figures will be created using ggplot2 (Wickham 2016) and Knitr(**Knitr?**).

2.2 Data Collection

The dataset contains all Major Crime Indicators occurrences by report data. The categories of major crimes are Assault, Break and Enter, Auto Theft, Robbery and Theft Over. The latest refresh was on Jan 11, 2024.

There are two factors that may influence the reliability of the data. First, this dataset includes all occurrences reported to the Toronto Police Station, except those have been considered as

unfounded. Second, this data is provided at the offence and victim level, so one occurrence number may have a few rows of data with different major crime indicators types used to categorize the occurrence.

2.3 Data Analysis and Visualizations

The Major Crimes in Toronto dataset contains 27 columns and 372899 rows, including the repeated occurrences with different crime categories. All the samples of major crimes were reported from 2014 to 2023. Among 27 columns, the variables I will use are x_id, report_year, occ_year, occ_month, occ_dow (Day of the Week Offence Occurred), occ_hour, mci_category, division. I will create two more variables, division_area, which classifies 18 police divisions (including a NSA) into 6 regions of Toronto (including Other for NSA), and weekend_weekday, which indicates whether the occurrence happen on a weekday or a weekend.

Table 1: The Number of Major Crimes and the Month with the Most Crimes (2014-2023)

Year	Total Number of Crimes	Month with the Most Crimes
2014	32477	October
2015	32938	August
2016	33654	July
2017	35547	July
2018	37545	October
2019	40098	July
2020	35196	July
2021	34777	October
2022	41299	October
2023	47833	August

Table 1 provides a brief summary of the total number of crimes from 2014 to 2023. Over this period, the overall number of crimes shows a consistent increase, except for the years 2020 and 2021, which have a decline probably due to the impact of the pandemic. However, in 2022 and 2023, there is a significant increase by more than 6000. In each year, the month with the highest number of crimes tends to be either July, August, or October, concentrating primarily in the summer-fall period.

Factors contributing to this uptrend are multifaceted, ranging from societal shifts associated with the pandemic to economic uncertainties. In particular, I will discuss the distribution of crimes from three aspects: types of crime, regions and time in a day.

Figure 1 visualizes the annual distribution of major crime types with five categories: Assault, Break and Enter, Auto Theft, Robbery, and Theft Over. Across all years, 'Assault' records

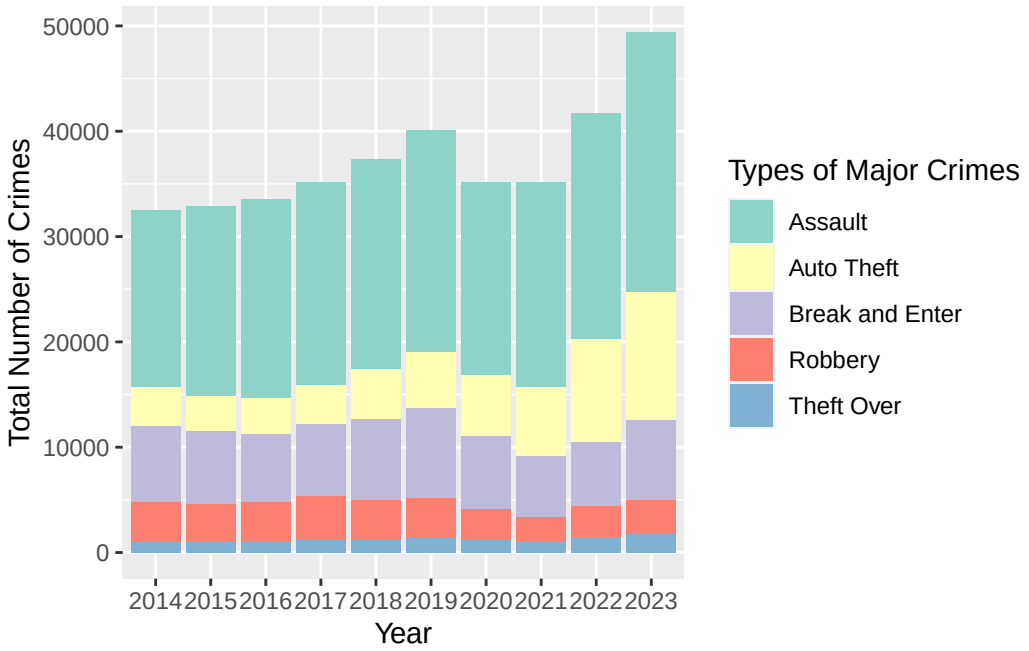

Figure 1: The Number of Crimes Reported, Classified by Crime Types (2014-2023)

the highest count, followed by 'Break and Enter' and 'Auto Theft'. 'Robbery' and 'Theft Over' show comparatively lower numbers. Notably, there has been a slight decrease in "Break and Enter" since the pandemic, and an obvious increase in 'Auto Theft' during 2022 and 2023, leading 'Auto Theft' surpassing the number of 'Break and Enter'.

There are 18 police divisions, which are D11, D12, D13, D14, D22, D23, D31, D32, D33, D41, D42, D43, D51, D52, D53, D54, D55, and NSA. I put them into 6 geographical regions, Etobicoke-York, Etobicoke, North York, Scarborough, Toronto-East York, and other, as shown in Table 2 Based on Figure 2, Toronto-East York reports the highest amount of crimes, and Etobicoke-York has the lowest. North York, Scarborough, and Etobicoke-York have similar numbers, but the rate of increase in North York is higher than that observed in the other two.

Table 2: Police divisions in 6 areas(including other)

division_area	division
Etobicoke	D22, D23
Etobicoke-York	D14, D13, D11, D12
North York	D32, D33, D31
Other	NSA
Scarborough	D43, D42, D41
Toronto-East York	D53, D55, D52, D51, D54

Figure 3 indicates the preferred time to commit crimes. The graph reveals a higher likelihood of criminal behaviours during the evening, with occurrences steadily increasing as the night progresses towards midnight and afterwards, the occurrences decrease until six in the morning. Interestingly, there is a surprising local peak at noon, while 11:00 and 13:00 do not show a high probability.

3 Conclusion

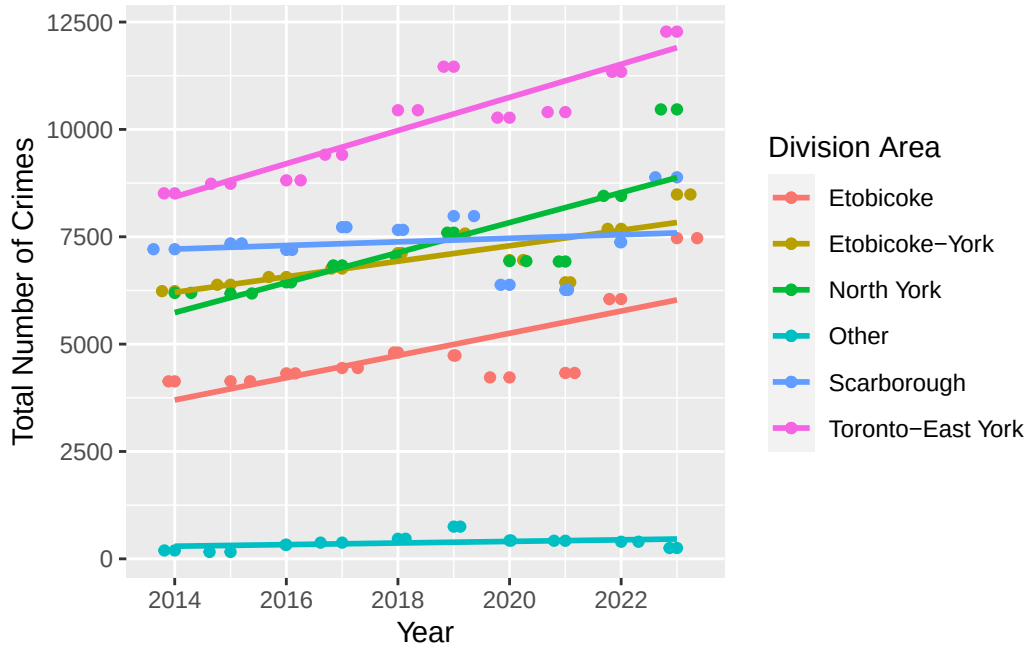

Figure 2: Total Number of Crimes in Different Police Division Areas

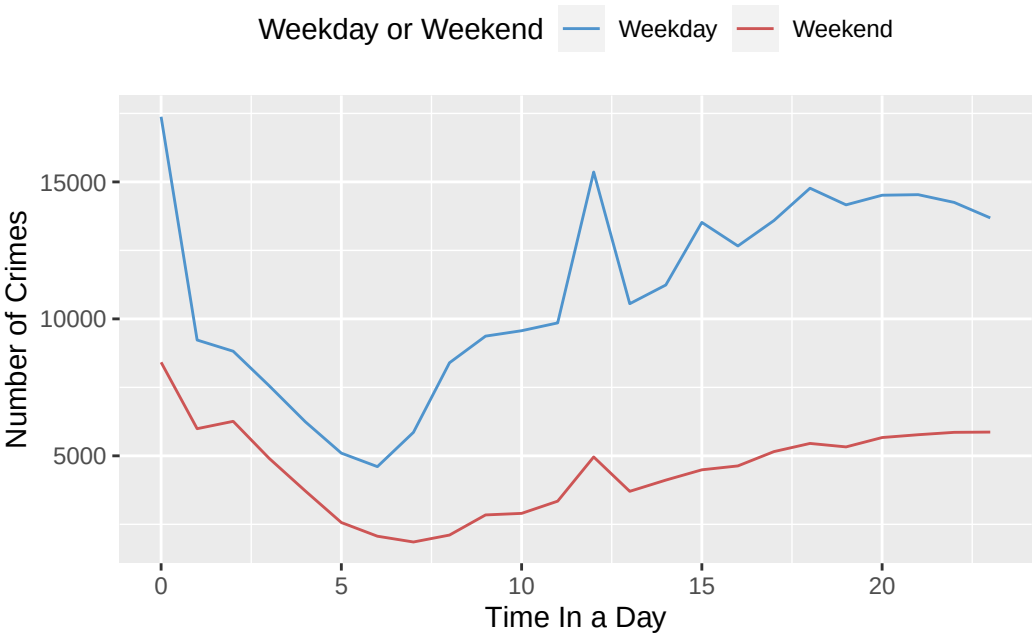

Figure 3: Number of Crimes in a Day

The Toronto-East York region reports the highest number of crimes, with North York, Scarborough, and Etobicoke-York following. Resources and fundings allocation should have an emphasis on Toronto-East York, especially the downtown area, considering the amount of people coming to work despite not residing there.

Focus on safety during midnight and noon hours is important too. The most common time for a crime occurrence include noon and midnight, especially between 6:00 PM to 8:00 PM. In order to prevent such incidents, people who work from home but go out for lunch, as well as those with children returning home for lunch around noon, should keep their ears and eyes open for suspicious surroundings. Evening hours in any region also need more cautions from residents and security from the government.

It is crucial to recognize that, though Toronto-East York records the most crimes, it also has the largest number of residents. Therefore, it could be biased to claim that Toronto-East York has the highest crime rate, or North York is getting dangerous, as North York has witnessed a growing population in recent years. Moreover, when creating this dataset, one occurrence number may have a few rows with different major crime indicators types used to categorize the occurrence. Therefore, some incidents of 'Auto Theft,' 'Robbery,' and 'Theft Over' may be potentially recorded multiple times under 'Break and Enter.' Consequently, the number of crimes classified as just 'Break and Enter' is overestimated, the actual number may be lower than what is shown now in the dataset.

References

Cheng, Joe, Barret Schloerke, Bhaskar Karambelkar, and Yihui Xie. 2023. *Leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. https://rstudio.github.io/leaflet/.

Firke, Sam. 2021. *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://here.r-lib.org/.

R Core Team. 2022. *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, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.