Rising Gun Violence in Toronto: An Analysis of Shootings and Firearm Discharges*

Jin Zhang

September 24, 2024

In this paper, we study reported shootings and gun incidents in Toronto from 2004 to 2024. Our analysis identifies key trends over time in different neighborhoods, showing higher-risk areas, seasons, and times of day. The data suggest that there may be a relationship between these factors. This study provides important insights that will help the Toronto Police Service improve security measures and effectively target high-risk areas through increased police presence.

Table of contents

*Code and data are available at: https://github.com/KrystalJin1/Shootings-Firearm-Discharges-Research.git

1 Introduction

We used R programming language (R Core Team 2023) to analyze the data and plot the graphs for this paper. The folder structure for this paper follows the startup folder created by (Wickham et al. 2019a). It is very helpful in keeping everything organized. I also referenced many of the examples and techniques provided by Telling Stories with Data (Alexander 2023), which helped me understand how to visualize the data and communicate the findings effectively. We used `dplyr` (Wickham et al. 2023) to process and transform the dataset for this paper efficiently, which made it easier to filter, group, and summarize the data. In addition, the `tidyverse` (Wickham et al. 2019b) packages is essential for simplifying data cleaning and analysis.`styler` (Müller and Walthert 2024) used in arranging code style. These tools help to organize and present data efficiently. I also plotted the graphs with `ggplot2` (Wickham 2016) and read the csv file with `here` (Müller 2020).

Shootings and gun discharges have become a growing concern in Toronto, but the public still has an incomplete understanding of how specific factors, such as time and place, affect gun violence. The increase in gun-related incidents not only threatens public safety but also stresses policymakers, law enforcement, and local communities. Detailed analysis of data related to shooting incidents can show high-risk areas and times and provide solutions to reduce violence. In this paper, I analyze Toronto's reported shooting and gun discharge incidents between 2004 and 2024 and provide the public with relevant information in a brief, easy-to-read chart. We utilize "year", "month" and "hour" variables to reveal trends in the number of victims over time and utilize latitude and longitude data to map the density of victims in order to determine which regions and time periods are at a higher risk of having a shooting incident.

Data on shooting incidents over the past two decades show fluctuations in the number of victims from year to year, but the number of the events has increased in recent years. The frequency of shootings is significantly higher during the summer months, and the incidents tend to occur at night, especially between 9:00 p.m. and 2:00 a.m. The data shows that the most dangerous areas are centered in the downtown area. The data show that the most dangerous areas are concentrated in the city centers. These findings highlight specific times and places that may require increased attention in terms of crime prevention and security measures.

The rest of the paper is structured as follows: in Section 2, we describe the source of the dataset and its context in detail. Next, we show our findings through various types of graphs with visualizations. After each chart, we discuss each graph's formation, the significance of the findings, and finally summarize key observations. Finally, there is an appendix and references.

2 Data

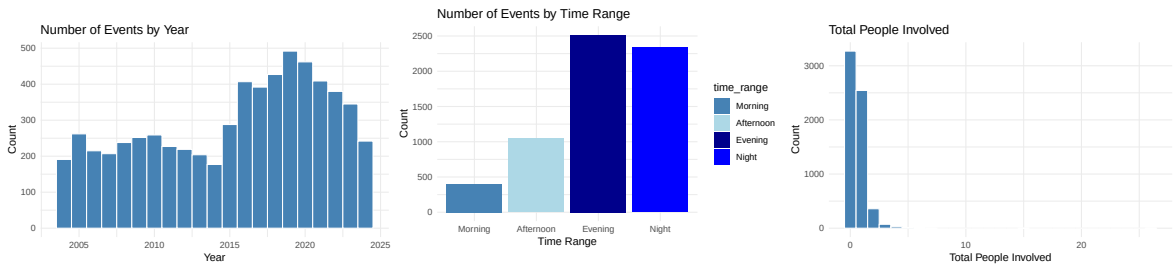
2.1 Source

The dataset analyzed in this study, the "Shootings & Firearm Discharges" dataset, was retrieved from Open Data Toronto (Gelfand 2022). This dataset was chosen because it provides detailed information on shooting-related incidents reported to the Toronto Police Service from 2004 to 2024, covering various details like time, location, and number of victims. The information about time and location is particularly useful for identifying trends and pinpointing high-risk periods and locations within the city.This dataset is excellent to help us achieve our research goals.

2.2 Measurement

Our research focuses on analyzing the relationship between key time and geographic factors (specifically year and time of day) and the incidence of shootings and firearm discharges in Toronto. The **Year** variable allows us to track trends in incident frequency over time, while **Time of day** helps to identify periods of increased risk. In addition, we included **location data (latitude and longitude)** to identify high-risk areas, and **total number of people involved (deaths and injuries)** to measure the level of seriousness of incidents. These variables were chosen to provide a deeper understanding of the gun violence incidents. As detailed explained in Appendix A, the data cleaning process involved several steps to prepare the dataset for analysis.

2.3 Graphs and analysis



(a) Number of Events by Year (b) Number of Events by Time Range (c) Total People Involved

Figure 1: Visualization of Events and People Involved in Shootings and Firearm Discharges

Figure 1 provides an overall view of the number of shooting incidents, victims involved, and time distribution in Toronto from 2004 to 2024. Figure 1a shows the number of incidents by year**, using the variable "year" to plot the distribution of incidents over time. The results

suggest that the number of incidents has increased significantly in recent years, possibly as a result of poor regulation.

Figure 1b represents the **number of events by time range**, based on the `time_range` and 'total_people' variable. This plot reveals that nighttime is a particularly dangerous period for shootings and firearm discharges.

Figure 1c describes the **total** number of people involved in each incident using the "total_people" variable. The analysis showed that most incidents resulted in fewer than three victims, suggesting that although shootings are frequent, the number of people involved in each incident is usually small.

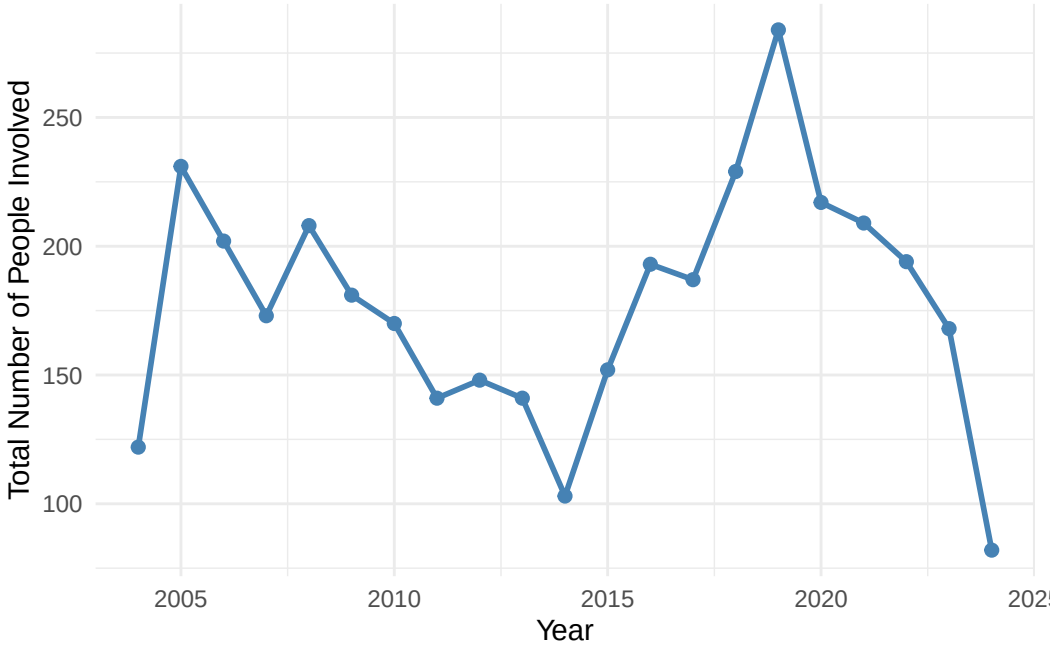


Figure 2: Trend of Deaths and Injuries by year

Figure 2 shows the trend of the total number of people involved in incidents, including deaths and injuries, over the years from 2004 to 2024. In the early 2000s, the number of people involved was quite high, reaching above 250 in 2004. After that, there is a general decrease, especially around 2010, where the number drops to around 150. However, after 2015, the numbers start to rise again, peaking at over 250 in 2019. After 2020, the number of people involved starts to decrease sharply again. This figure shows how the total number of incidents has changed over time, with clear periods of increase and decrease.

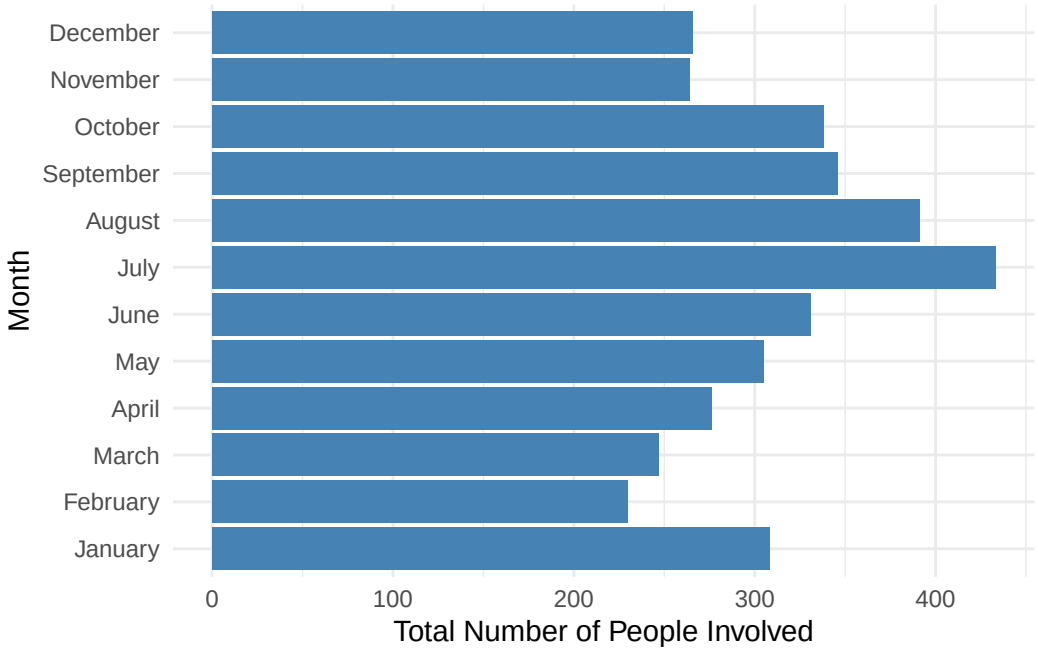


Figure 3: Distribution of Deaths and Injuries by Month

Figure 3 shows the total number of shootings and gun discharges by month. The chart uses the "month" variable to show the distribution of incidents throughout the year. From the graph we can observe that the number of shooting victims peaks in the summer months, especially in July and August. This shows that shootings tend to occur more frequently in the warmer months, which may be due to an increase in outdoor activities. Suggesting that there may be a seasonal pattern to gun violence in Toronto.

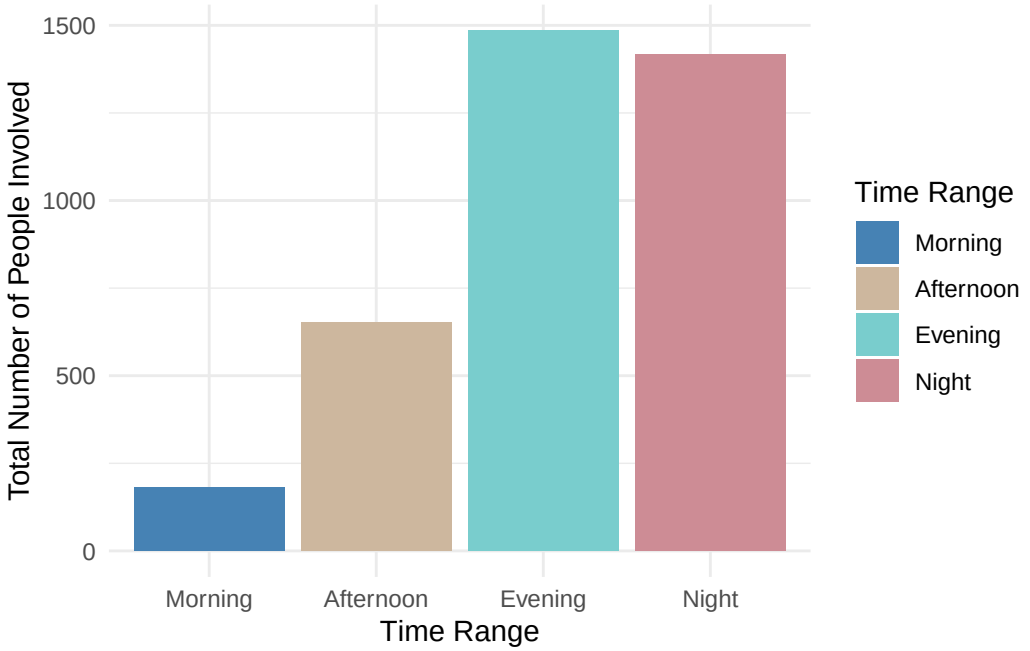


Figure 4: Distribution of Deaths and Injuries by Time Range

Figure 4 shows the total number of people involved in shootings and gun incidents at different times of the day. The chart is divided into four time periods: morning, afternoon, evening, and night. The data shows that the majority of incidents occur in the **Evening** and **Night**, where more than 1,500 people are affected. **Afternoon** shows a moderate incidence of incidents, while **morning** has the lowest number of people affected by incidents. TThis pattern suggests that the risk of shooting increases significantly at night, and that heightened attention may be required during these times.

Figure 5 shows the number of shooting incidents that occurred during different hours of the day. Most shootings happen between **9 PM (21:00)** and **midnight (23:00)**, where the number of incidents is higher than at other times, reaching above 500 incidents at 23:00. The number of incidents decreases in the morning hours, with fewer shootings happening between **6 AM (6:00)** and **12 PM (12:00)**. This graph shows the hours have higher risk for shootings during the day.

Figure 6 shows the geographic distribution of shootings in Toronto, categorized by risk level. Events with more than two deaths or injuries are defined as high risk. The map uses blue dots to indicate low-risk incidents and red dots to indicate high-risk incidents. The map shows that most incidents are categorized as low risk. High-risk incidents tend to occur in the central areas of Toronto, where there is a higher density of dots. This map helps us to understand where the shooting incidents are high and to be safe when traveling.

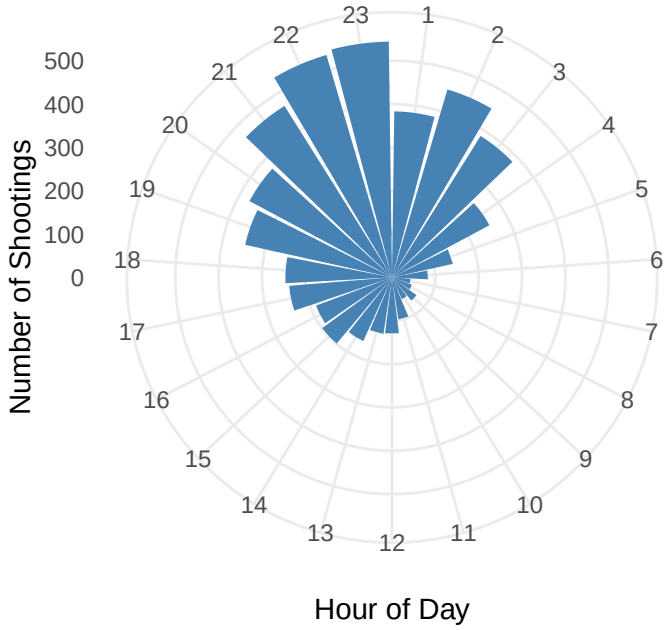


Figure 5: Shootings Distribution by Hour of Day

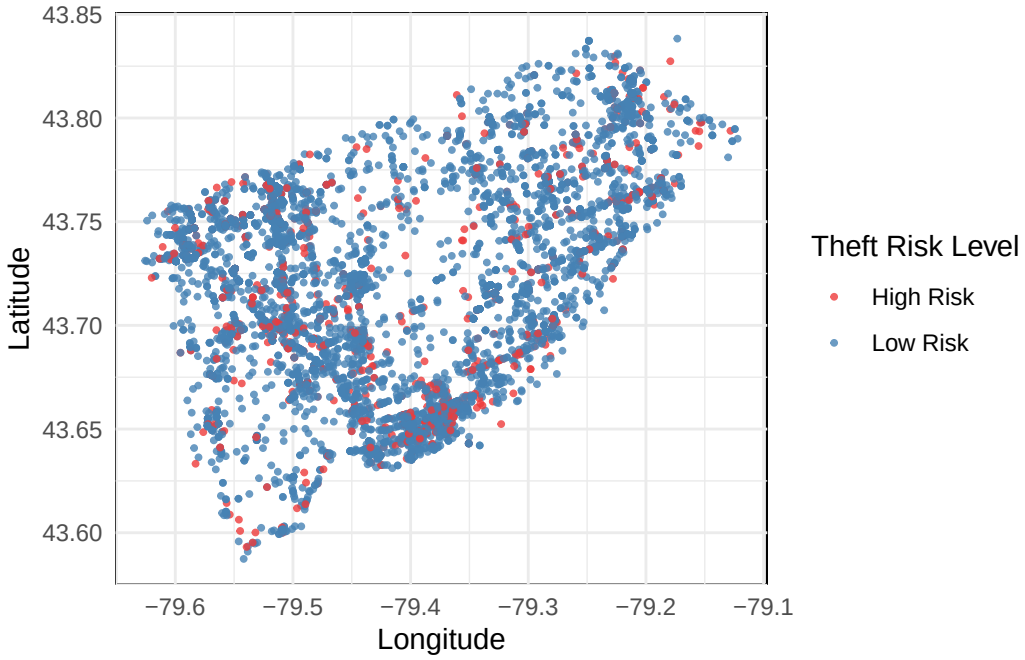


Figure 6: Map of Shootings Highlighted by Risk Level

3 Discussion

In summary, our analysis of Toronto shootings and gun discharges from 2004 to 2024 shows that gun violence is more frequent during the summer months and at night, especially between 9 p.m. and 2 a.m., and downtown areas are at higher risk. R Core Team (2023), Wickham et al. (2023), and Wickham (2016) helped simplify the data analysis and visualization process, enabling us to present these findings clearly. This data-driven approach emphasizes the necessity of targeted safety measures in high-risk areas and time periods, providing valuable insights for policymakers to reduce urban gun violence.

A Appendix: Data Cleaning Process

A.1 Initial Data Cleaning

The first step involves loading the original dataset, normalizing and simplifying the column names. This makes the data easier to analyze. Only the most important variables were kept. Here is the explanation of the variables we have selected:

year: Represents the year in which the shooting incident occurred. This variable is crucial for identifying yearly trends and analyzing changes in gun violence over time.

month: Indicates the specific month when the incident took place. It useful in seasonal analysis.

hour: This variable captures the exact hour of the day when the shooting occurred. It is useful for find the high-risk times of the day and understanding the daily patterns of incidents.

time_range: Groups the time of the day into broader categories such as morning, afternoon, evening, and night. This is for a more generalized view of when shootings are most likely to occur.

death: The number of deaths resulting from the shooting incident. This variable is important for understanding the fatality rates across incidents.

injuries: The count of people injured during the shooting.

geometry: This variable contains geographic coordinates (longitude and latitude) that represent the exact location where each shooting or firearm discharge incident occurred.

A.2 Renaming Variables

Some variables were renamed to make the dataset easier to interpret. For example, `occ_date` was renamed to `date` and `occ_year` was renamed to `year`. Renaming is an important step in ensuring that we label variables clearly.

A.3 Creating the total_people Variable

A new variable "total_people" was created by adding the values from the `deaths` and `injuries` columns. If one of the columns had missing data (i.e., an NA value), it was replaced with a zero. This adjustment provides a more complete and accurate view of the total number of people affected in each incident.

A.4 Splitting the Geometry Column

The `geometry` column containing the geographic coordinates is split into two columns: `longitude` and `latitude`. This conversion makes it easier to analyze and plot the

References

Alexander, Rohan. 2023. *Telling Stories with Data: With Applications in r*. Chapman; Hall/CRC.

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.

Müller, Kirill, and Lorenz Walthert. 2024. *Styler: Non-Invasive Pretty Printing of r Code*. https://CRAN.R-project.org/package=styler.

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. 2019a. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

———, et al. 2019b. "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.