How the timing of shootings affects death and injury trends*

An in-depth study of gun incidents from 2013 to 2023

Yanyu wu

21 April 2024

Abstract

Using data from Toronto Open Data I explore the landscape of gun violence in Toronto from 2013 to 2023, and the relationship between the timing of these incidents and the resulting casualties. The purpose of this article is to examine the relationship between the timing of shootings, location of shootings, and number of shootings. We find that certain time periods and locations are significantly more prone to shooting incidents, revealing potential temporal, regional patterns in the occurrence and severity of such incidents. This insight into the temporal dynamics of gun violence in Toronto highlights the urgent need for targeted prevention and policy interventions aimed at reducing the incidence of shootings during high-risk periods.

Contents

Introduction 2

Estimand . . . 2

Data 3

Dataset Description and Methodology . . . 3

Data Visualization . . . 5

Model 6

Model Setup: . . . 6

Model Justification . . . 7

Model Results . . . 10

Discussion 11

Analysis within one week . . . 11

Think deeply . . . 11

Weaknesses . . . 12

Next steps . . . 12

References 13

*Code and data are available at: https://github.com/YANYUWU0428/final.git

Introduction

On December 19, 2022, a tragic incident occurred in an apartment in the Toronto suburb of Vaughan. A 73-year-old gunman killed five people and wounded one before being shot dead by police. The attack involved multiple units within the same building. The attacker was reportedly involved in a legal dispute with several members of the condo board.(Sharma 2022)Another shooting in downtown Toronto left two men dead and a woman injured. The victims and suspects were related, adding a personal touch to the tragedy. A suspect was arrested at the scene(Rocca 2024)

The issue of gun violence has remained at the forefront of public discussion in recent years, with the proliferation of social media dramatically increasing engagement with such incidents. In Toronto, a city known for its richness and vitality, the number of shootings has caused heightened concern. The prevalence of shootings often resonates with sensitivities in communities that, when they occur, can lead to a sense that all previous efforts to protect the community have been futile, and this heightened sensitivity reflects a growing desire to mitigate factors that contribute to gun violence. In the global discussion around gun violence and control, Canada stands out for its strict gun control laws, which are designed to reduce unauthorized access to firearms and reduce gun-related crime . This article takes a closer look at the subtle relationship between the timing and incidence of shootings in the city of Toronto, despite Canada's strict gun control laws. It aims to carefully analyze shooting trends spanning the decade from 2013 to 2023, using cleaned shooting data from Toronto Open Data. The core goal is to understand how different times of day affect the likelihood of a shooting. While crime data has historically been available, there remains a gap in nuanced, localized analysis of shootings that takes into account the temporal dimension. Previous studies have tended to focus on broader crime categories or lack the granularity needed to inform targeted interventions. This study seeks to fill this gap by providing detailed insights into shooting incidents to help develop more effective response strategies. We categorized the data into different time intervals to examine correlations between different times of day. and frequency of shootings. This involves not only counting incidents within each time period, but also employing statistical models to determine whether certain time periods exhibit a higher likelihood of violence. The goal is to reveal specific temporal patterns of gun violence, with the ultimate goal of guiding strategic decisions about police patrols to prevent and mitigate the risk of shootings during their most vulnerable times. The model section uses multiple linear regression, and the model is constructed to demonstrate the relationship between time, location, and number of shootings, as well as future predictions. Interpretation of the final model and all findings related to it. This article is structured as follows: After the introduction, Part 2 discusses data sources, sampling, and central data points . Part 3 develops the study of the model, and Part 4 presents the data and model results. The article concludes with a discussion section on the impact of policies on low-income groups and gun control bills introduced in recent years, outlining possible shortcomings of the study and proposing future directions for development.

Estimand

It is expected that evenings will have more fatalities and injuries.

Data

For the analysis, we used a range of R (R Core Team 2023) packages tailored for data analysis and reporting. (Wickham et al. 2019)'s `Tidyverse` is used for data wrangling, (Firke 2023)'s `janitor` package is used for data cleaning operations, and (Xie 2023)'s `knitr` is used for data presentation in data tables. (Wickham 2016)'s "ggplot"Analyze images and tables.The following code snippet is intended to import important packages that are crucial for checking missing values in a data set. We ran the model in R (R Core Team 2023) using (Goodrich et al. 2022)'s `rstanarm` package and `here`package(Müller 2020). We use the default prior of `rstanarm`. To perform comprehensive mixed-effects model analyses, we utilize the `broom.mixed` package (Bolker and Robinson 2022), which extends the `broom` package functionality to mixed models, facilitating the extraction, organization, and representation of model outputs. Additionally, the `modelsummary` package (Arel-Bundock 2022) provides tools for creating customizable summary tables of model results, thereby enhancing the interpretability and dissemination of statistical results.

Dataset Description and Methodology

The release of the shooting and firearm discharge dataset is made available through the OpenData Toronto platform(Gelfand 2022), which serves as a data distribution platform to provide the public with a transparent and easily accessible channel to enable residents and researchers in the City of Toronto to access this important public safety data. The data is intended to increase public safety awareness and support the development of crime prevention measures and law enforcement strategies. The data set included information on shootings, where a firearm fired a pellet that led to an injury, and firearm misfires, which covered all situations in which there was evidence of bullets being fired from a firearm, such as accidental discharges and celebratory fires. The data set also specifically defined "injured person" as someone who was injured by a bullet as a result of a gunshot, excluding cases of suicide, police action, or the use of a non-real gun(Services 2024). The dataset provides a breakdown of injury levels, distinguishing between events that resulted in death and non-fatal physical injury. The data set is updated quarterly, with the last update date being April 4, 2024, and the data quality is rated "gold", reflecting its high reliability.

Table 1: Table 1: Summary of Top Ten Shootings

Year	dow	Timerange	Deaths	Injuries
2013	Tuesday	Night	0	2
2013	Thursday	Evening	0	1
2013	Friday	Evening	0	0
2013	Friday	Night	0	0
2013	Saturday	Night	0	1
2013	Monday	Evening	0	0
2013	Friday	Night	0	1
2013	Wednesday	Evening	0	1
2013	Friday	Afternoon	1	0
2013	Tuesday	Night	0	1

Table 1 shows the first ten rows of the cleaned dataset, which contains 3983 variables for a total of 5 variables. The target population for the dataset is the combination of those injured or killed in a shooting incident. As you know, the information in the dataset was extracted from Toronto Open Data. The variable "Year", this column records the year in which the shooting occurred. Variable "Week" This column records the week in which the shooting occurred. Variable "Time Range" This column indicates the time period in which the shooting occurred, categorizing the time of day when the shooting occurred into four categories (morning, afternoon, evening, and night). The variable "dow" indicates the total number of fatalities and injuries from shootings that occurred each day of the week. The variable "Deaths" is the number of fatalities resulting from the shooting. By analyzing the time of day (week and time frame) when shootings occur, law enforcement agencies can better allocate resources and personnel. If the data indicates that more incidents occur during certain times of the day, then police forces can increase patrols during these high-risk times.

Data Visualization

In order to further familiarize ourselves with the dataset and estimate possible associations between the number of deaths and injuries and time, exploratory analyses were conducted by performing data visualizations to see if the different temporal patterns conformed to the general expectations of trends in the number of deaths and injuries. First, we will look at overall trends in Toronto shootings.

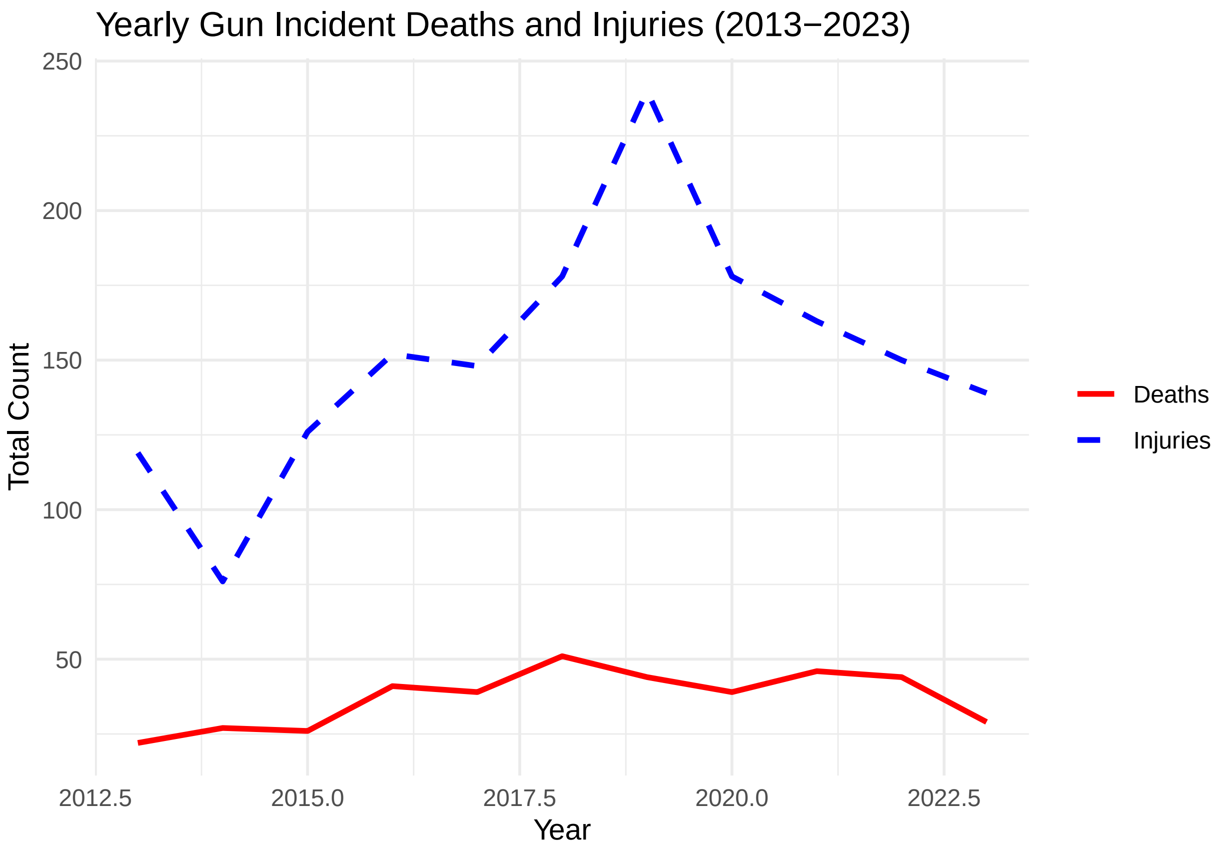
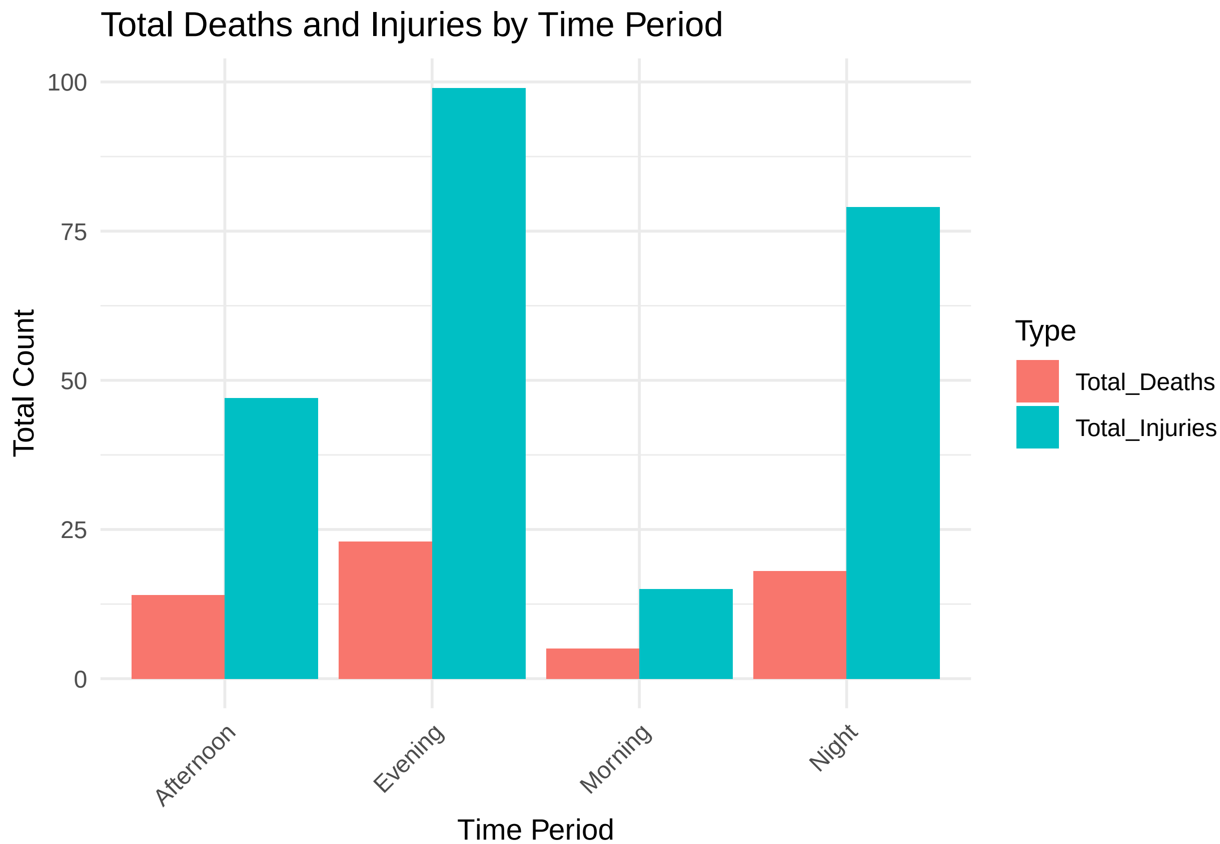


Table 2: Total Deaths by Time Period

year	timerange	Total_Deaths	Total_Injuries
2013	Afternoon	7	23
2013	Evening	8	45
2013	Morning	0	4
2013	Night	7	47
2014	Afternoon	4	17

As shown in Figure 1, the number of people killed and injured in shootings between 2013 and 2023 is displayed. Since 2013 to the present, there has been an overall increasing trend in shootings of Canadians. The vertical axis on the graph represents "number of people". The horizontal axis represents "years". There are two lines in the chart, one for deaths (shown in red) and one for injuries (shown as a dotted blue line). The chart shows that the number of injuries is significantly higher than the number of deaths, and that the number of injuries peaked around 2017-2018 and then declined. In contrast, the number of fatalities remained relatively stable over this time period, with slight fluctuations but no significant peaks or troughs.



As shown in Figure 2, this bar graph illustrates the total deaths and injuries that occurred during different time periods of the day. The time periods have been categorized as afternoon, evening, morning, and night. The data is presented as red and blue bars, where red represents the total number of deaths and blue represents the total number of injuries. The height of the bar graphs allows for a comparison of the different time periods: the number of injuries and fatalities is much higher in the evening hours than in the other hours, and the number of injuries, in particular, shows a significant peak in the evening. In the morning hours, on the other hand, the total number of deaths and injuries is the lowest. The bar graph clearly depicts the security situation at different times of the day, suggesting the frequency of shootings and their severity during a given time period.

Model

Model Setup:

In the initial analysis, we considered using the Poisson regression model for our study because of its simplicity and high interpretability. However, a limitation of this model is that it may not adapt to overdispersed data, i.e. when the variance of the observed data exceeds the mean. The presence of overdispersion may reduce the model's predictive accuracy on new data. Therefore, we choose to use the most basic linear model. Linear-based models can provide additional insights when exploring the relationship between total number of shootings and time of day (morning, afternoon, evening, night). Such models help reveal correlations between the independent variable, time period, and the dependent variable, number of shootings. (Goodrich et al. 2022)

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \beta_0 + \beta_1 \times \text{Timerange}_i + \epsilon \tag{2}$$
$$\beta_0 \sim \text{Normal}(\mu_0, \sigma_0) \tag{3}$$
$$\beta_1 \sim \text{Normal}(\mu_1, \sigma_1) \tag{4}$$
$$\sigma \sim \text{Exponential}(\lambda) \tag{5}$$

In Model:

Y_i to be the expected rate of occurrence of deaths or injuries.β_0 is the coefficient for intercept.β_1 is the coefficient for the continuous time range variable.

Model Justification

Based on the models discussed above, determine whether each time period is positively or negatively correlated with the number of deaths and casualties, and in the upcoming model results section, we will elaborate on the association between each model, the number of injuries, the number of deaths and the time of casualties.

Table 3: Comparison of Mean and Variance of Total Deaths and Injuries in Toronto from time period

Mean_Deaths	Variance_Deaths	Mean_Injuries	Variance_Injuries
9.272727	45.9704	37.93182	626.4371

Preliminary analysis showed that the upward or downward trends in shootings over time were not strictly linear and that the data exhibited overdispersion (the variance of observations exceeded the mean), which justified the use of a linear model. By looking at the coefficients related to the time index Time_range, we can determine whether shootings are increasing or decreasing, and the rate at which these changes occur. Before developing the model, we used historical data to study the distribution and trends of shooting incidents. Our goal is to gain a deeper understanding of shooting trends in Toronto by applying the model. Future research will focus on this understanding.

#Results

This stacked line chart depicts the number of injuries in Toronto at different times of the year from 2012 to approximately 2023. The annual trend in injuries is represented by different colored lines for four time periods: afternoon (red line), evening (blue line), morning (green line) and nighttime (purple line). Injuries peaked in the afternoon and have remained mostly stable during that time. Late-night injuries rose slightly around 2017, then slowly declined. The number of injuries was lowest and most consistent in the early morning hours. Even though the nightly statistics fluctuated, the overall trend didn't change much. Urban planning, the development of safety precautions, and the deployment of emergency services could all benefit from the insights provided by this data, particularly in terms of resource allocation and preventive measures. It is possible to increase safety levels at specific times and gain a better understanding of potential risk factors by looking at the number of injuries that occur during those time periods.

This chart is again a stacked line chart showing the annual number of deaths in Toronto over different time periods from around 2012 to 2023. The number of deaths was relatively highest in the afternoon period, showing slight fluctuations but no obvious upward or downward trend. The evening session has a significant decline in the intervening years and then rebounds. Death tolls remained low and relatively stable during the morning hours. Deaths in the nighttime period fluctuated somewhat throughout the time frame, but began to trend downward in 2017.

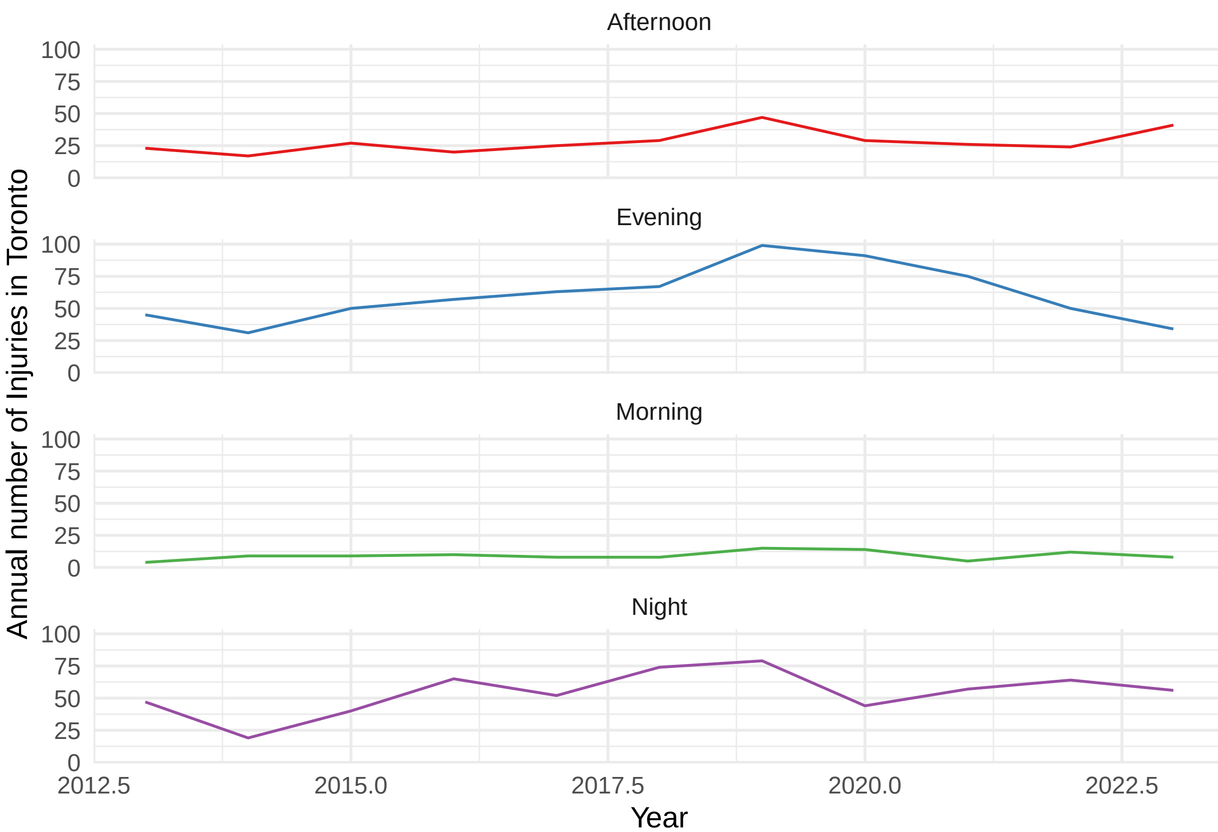


Figure 1: Total Injuries Per Year in Toronto (2013-2023)

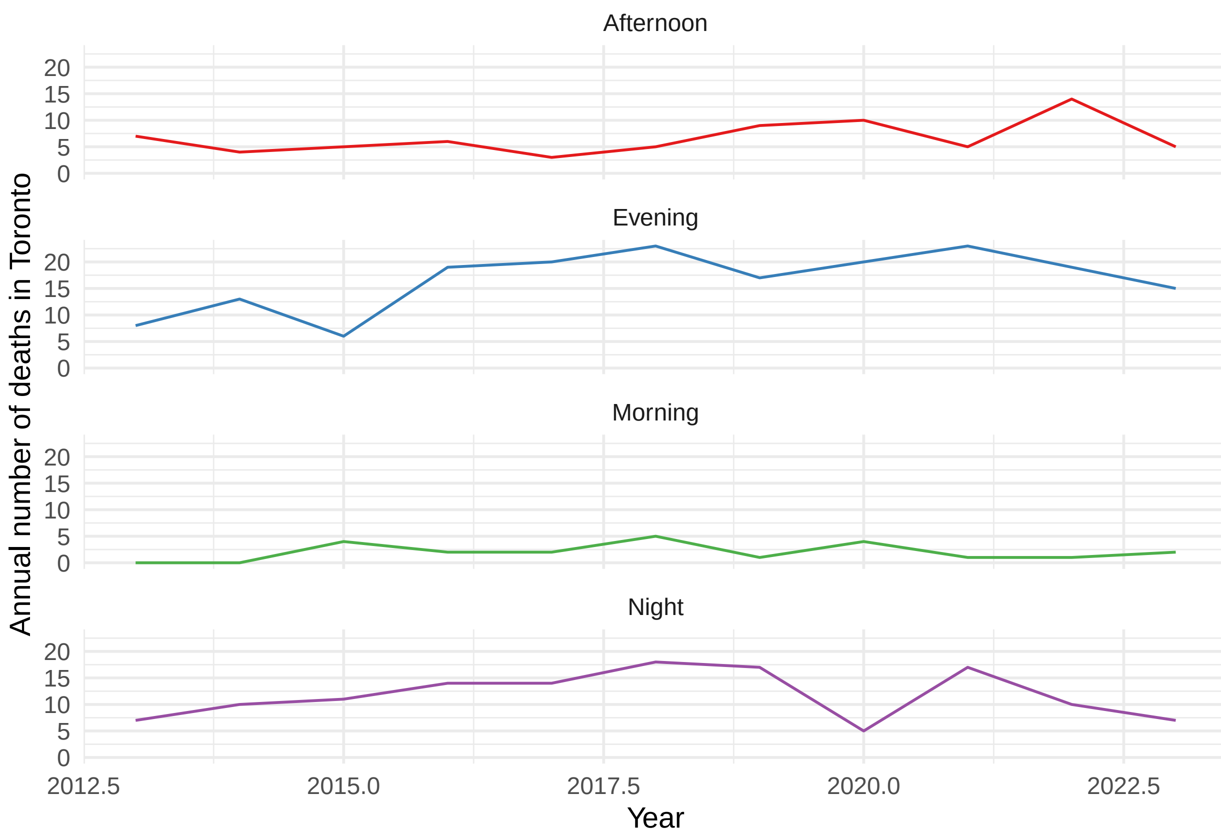


Figure 2: Annual number of deaths for time period in 2023, since 2013, for Toronto

	death_model	injuries_model
(Intercept)	6.636	28.000
	(1.216)	(4.378)
timerangeEvening	10.000	32.182
	(1.720)	(6.192)
timerangeMorning	−4.636	−18.727
	(1.720)	(6.192)
timerangeNight	5.182	26.273
	(1.720)	(6.192)
Num.Obs.	44	44
R2	0.671	0.687
R2 Adj.	0.646	0.663
AIC	253.4	366.1
BIC	262.3	375.0
Log.Lik.	−121.699	−178.062
RMSE	3.85	13.84

Model Results

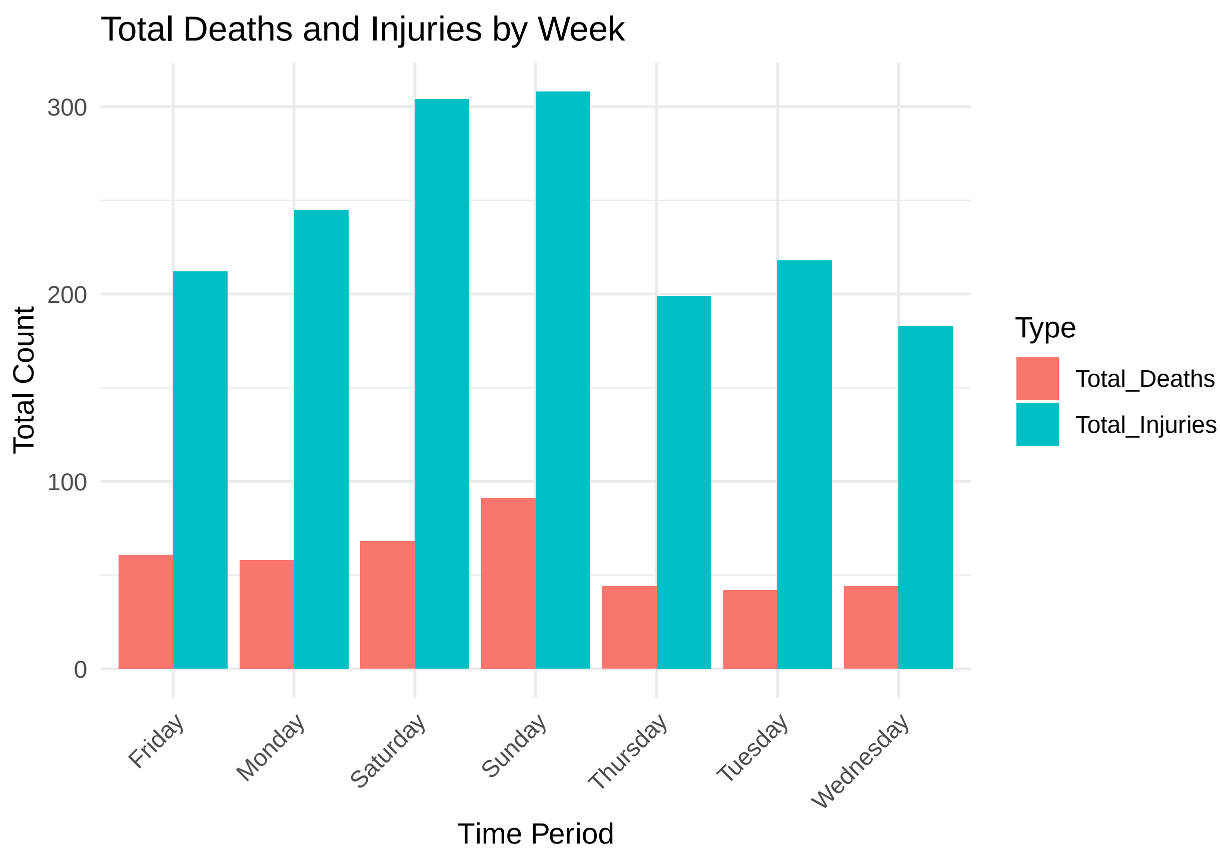
This image shows the output of a statistical model, which contains two models: death_model and injuries_model. It is the output of a regression analysis, typically used to evaluate the impact of one or more predictor variables on the response variable. The predictor variables here are different time periods (Evening, Morning, Night), and the response variables are the number of deaths (death_model) and the number of injuries (injuries_model).

The intercept for the death_model is 6.636. When all other model variables are set to zero, this would be the baseline forecast for deaths.The baseline forecast for injuries in the injuries_model is 28.000, which is the intercept.We may get a sense of the precision or variability of the estimated coefficients by looking at the values in parenthesis under the intercepts and the coefficients. More accurate estimations are suggested by smaller standard errors in relation to the coefficients.According to the death_model, the coefficient is 10.000, meaning that when the time range is evening, the expected deaths are expected to rise by 10 in comparison to the baseline category, which may be daytime or another period not specified.The injuries_model's coefficient is 32.182, meaning that, in comparison to the baseline, there will be a 32.182 increase in the number of injuries expected during the nighttime hours.The death_model coefficient is -4.636, indicating a 4.636 drop in expected deaths from the baseline in the morning.The injuries_model's coefficient, which stands at -18.727, shows that, in comparison to the baseline, fewer injuries are expected to occur in the early hours.The death_model coefficient, which is 5.182, shows that there is a rise in anticipated deaths at night in comparison to the baseline.The injuries_model's coefficient, which stands at 26.273, indicates that there is a higher likelihood of injuries at night than there was during the baseline.

The intercepts imply that, if all other variables are left out and there is a time range that isn't displayed that acts as a reference category, there would be an estimated 6.636 fatalities and 28 injuries at that reference period. These baseline forecasts are then modified based on the coefficients for each time range and the time of day. For example, it is expected that evenings will have more fatalities and injuries, mornings will have fewer, and nights will have more deaths and injuries again, albeit not as high as evenings.

Discussion

Analysis within one week



As shown in Figure 3, in this bar graph we see the total number of deaths and injuries that occurred each day of the week. It shows the data in detail for each day from Monday through Sunday. The horizontal axis represents the seven days of the week, while the vertical axis represents the number of people. The graph uses red and blue bars to represent two different data types: the red bar represents the total number of deaths per day and the blue bar represents the total number of injuries per day. As can be seen from the bar graphs, Saturdays appear to be the day with the highest number of deaths and injuries in this particular time period, while Fridays are the day with the lowest number of deaths. This data may provide important information for studying safety on different days of the week or for planning emergency services. For example, high incidence Saturdays may require more police deployment.

Think deeply

Through time analysis, we learned that the shootings were not randomly distributed, but concentrated in a specific period of time. One of the main findings of our study was to identify times when shootings are most common, such as late at night or on weekends. This finding points to the need for increased police patrols and security measures during these times. By targeting these high-risk periods, law enforcement can more effectively deploy resources to prevent future shootings and protect public safety. Furthermore, the prediction of an increase in shootings over time by linear regression models raises concerns that these trends are likely to continue if interventions are not implemented effectively. Using these findings will also help increase the alertness of residents so they can better understand when the risk is high and take appropriate preventive measures.

Weaknesses

Toronto's shooting and firearm discharge dataset, while providing critical public safety information, has several limitations. First, in order to protect privacy, the event location is shifted to the nearest road intersection, which, while necessary, may reduce the accuracy and usefulness of the data in spatial analysis. Second, the data set is updated quarterly, which means that the latest events may not be reflected in the data immediately, affecting the application of real-time data and the development of emergency response strategies. In addition, the definition of "injured person" in the data set excludes suicides, police actions, or the use of non-real guns, which may lead to a lack of comprehensive coverage of gun-related injuries. Therefore, these factors need to be taken into account when using these data so as not to draw misleading conclusions when conducting research.

Next steps

Understanding the use of firearms in suicides and police operations can significantly increase the value and usefulness of data. First, it can provide a more comprehensive perspective to analyze the problem of gun violence, allowing us to more accurately assess and respond to gun-related threats. For example, data on gun use in suicides can help understand the link between mental health issues and gun violence, while data on gun use in police operations can shed light on potential risks in law enforcement.

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bolker, Ben, and David Robinson. 2022. *Broom.mixed: Tidying Methods for Mixed Models*.

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

Gelfand, Sharla. 2022. *opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Müller, Kirill. 2020. *here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rocca, Ryan. 2024. *Victims, suspect in Toronto shooting that killed 2 and left 1 injured were related: police*. https://globalnews.ca/news/10354253/daylight-shooting-parliament-dundas-toronto/.

Services, Toronto Police. 2024. *About Shootings & Firearm Discharges*. https://open.toronto.ca/dataset/shootings-firearm-discharges/.

Sharma, Shweta. 2022. *Toronto shooting – live: Five killed in "horrendous" Canada condo attack*. https://news.yahoo.com/toronto-shooting-live-five-shot-053839838.html?guccounter=1.

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.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.