Rising Firearm Violence in Toronto: Temporal Trends and Neighborhood Hotspots (2004–2024)*

Exploring Two Decades of Data to Uncover Patterns, Predict Future Incidents, and Guide Interventions

Heyucheng Zhang

December 7, 2024

This paper examines firearm violence in Toronto from 2004 to 2024 using Open Data Toronto to understand its temporal and geographic patterns. Findings show an upward trend in firearm-related injuries and deaths, with incidents disproportionately concentrated in neighborhoods such as Glenfield-Jane Heights and Black Creek. Linear regression models predict continued increases if patterns persist, highlighting seasonal peaks during summer months. These results aim to guide policy and community efforts to address firearm violence through localized interventions and resource allocation.

Table of contents

---

*Code and data are available at: https://github.com/heyuchengzhang/Toronto-Shootings-Firearm. Shiny Web App is available at: https://79muet-heyucheng-zhang.shinyapps.io/Toronto-Shootings-Firearm/.

1 Introduction

In Toronto, firearm violence has steadily increased over the past two decades, significantly affecting neighborhoods and public safety. Between 2004 and 2024, firearm-related injuries and deaths have risen, disrupting lives and presenting ongoing challenges for policymakers and law enforcement. This study analyzes firearm violence in Toronto over this period to identify

temporal patterns, highlight neighborhood-specific concentrations, and guide interventions aimed at addressing these issues.

Using detailed data from the Open Data Toronto and statistical and geospatial tools in the R programming environment, this study examines two decades of shooting and firearm discharge data. The analysis identifies upward trends in injuries and deaths, as well as a consistent concentration of incidents in neighborhoods such as Glenfield-Jane Heights and Black Creek. These findings underscore the importance of tailored interventions that address firearm violence at both the city and neighborhood levels.

The estimand in this study is twofold, focusing on two key questions:

1. **Cumulative Neighborhood Impact**: What is the total number of firearm-related incidents within each Toronto neighborhood from 2004 to 2024?

2. **Temporal Trends in Severity**: How have city-wide annual injuries and deaths from firearm incidents changed over time, and what trajectory is projected if current trends continue?

By focusing on these questions, the study aims to quantify the scope and distribution of firearm violence, providing a foundation for informed decision-making.

Through statistical modeling, geospatial mapping, and trend analysis, this research examines firearm violence in a systematic and actionable manner. It aims to support policymakers, law enforcement, and community stakeholders in developing targeted strategies to reduce firearm violence and its societal impacts.

The paper is organized as follows: Section Section 2 outlines the data sources and methodology employed; Section Section 3 presents the statistical models and their theoretical underpinnings; Section Section 4 summarizes the results, emphasizing spatial and temporal patterns; Section Section 5 discusses findings, limitations, and recommendations for future research and policy; and Section Section 6 provides an appendix addressing observational data, sampling methodologies, and the challenges associated with analyzing crime data. This structure ensures that the findings of this study contribute to efforts to create safer neighborhoods in Toronto.

2 Data

2.1 Methodology

In analyzing firearm violence in Toronto, we utilized the R programming language (R Core Team 2022) alongside essential packages from the tidyverse (Wickham et al. 2019), including dplyr (Wickham et al. 2023) and tibble (Müller and Wickham 2023), to streamline data manipulation. The janitor package (Firke 2023) ensured variable names were clean and duplicates were removed. ggplot2 (Wickham 2016) was employed to create compelling visualizations, and

kableExtra (Zhu 2021) was used to generate formatted tables. The here package (Müller 2020) enabled efficient file path management to ensure reproducibility.

Spatial analysis was conducted using the sf package (Pebesma 2018), integrating geographic data with precision. Data from the Toronto Open Data portal, processed with the opendata-toronto package (Gelfand 2022), provided rich geographic context. Statistical summaries and modeling were facilitated by the modelsummary (Arel-Bundock 2022) and broom (Robinson, Hayes, and Couch 2022) packages. The arrow package (Richardson et al. 2023) ensured efficient data import and handling.

2.2 Dataset Sources

2.2.1 Shooting and Firearm Discharge Data

The Open Data Toronto's Shooting and Firearm Discharge Data (Service 2024) contains all reported shooting-related occurrences. It includes incidents where projectiles were discharged from firearms as defined under the Criminal Code of Canada. Notably, the dataset encompasses two categories:

- **Shooting Event/Occurrence**: Incidents where a projectile discharge injured a person, excluding suicides and police-involved discharges.
- **Firearm Discharge**: Incidents involving discharged projectiles without reported injuries, including accidental discharges and celebratory fire.

Each record includes:

- **Spatial Data**: Neighborhood identifiers (`NEIGHBOURHOOD_158`), geographic coordinates (`LONG_WGS84`, `LAT_WGS84`), and offset locations to the nearest intersection for privacy.
- **Temporal Data**: Dates of incidents (`OCC_DATE`).
- **Impact Data**: Severity metrics, including `DEATH` (fatalities) and `INJURIES` (non-fatal injuries).

2.2.2 Neighbourhoods

Toronto's 158 neighborhoods (Social Development 2024) serve as the geographic framework for this study. Defined based on Statistics Canada census tracts, these neighborhoods provide a consistent boundary for longitudinal analysis of socio-economic and crime-related data. An important tool for spotting trends and organizing interventions, stable borders provide comparability throughout time.

Neighborhood profiles enable targeted views into localized trends of firearm violence, supporting efforts by government and community agencies in service delivery and planning. These

Table 1: Sample of Cleaned Data

OCC_DATE	DEATH	INJURIES	NEIGHBOURHOOD_158	LONG_WGS84	LAT_WGS84
2004-01-01	NA	1	Pelmo Park-Humberlea (23)	-79.53822	43.73374
2004-01-03	NA	NA	NSA	NA	NA
2004-01-04	NA	2	Kensington-Chinatown (78)	-79.39916	43.65612
2004-01-05	NA	1	Kennedy Park (124)	-79.25604	43.73290
2004-01-06	NA	NA	Malvern East (146)	-79.20741	43.80180
2004-01-06	NA	NA	Malvern East (146)	-79.20652	43.80835
2004-01-08	1	NA	Rustic (28)	-79.50727	43.71742
2004-01-08	1	NA	Woburn North (142)	-79.23385	43.78178
2004-01-13	NA	1	Malvern West (145)	-79.22095	43.80564
2004-01-18	NA	1	Newtonbrook East (50)	-79.41962	43.79632

profiles offer a socio-economic lens to interpret variations in firearm-related incidents across the city.

2.3 Data Preparation and Selection Justification

The analysis relied on a rigorous data cleaning and preparation process to ensure accuracy and reliability:

1. **Data Cleaning**: Invalid geographic coordinates (e.g., `LONG_WGS84 = 0` or `LAT_WGS84 = 0`) were removed. Neighborhoods identified as 'NSA' (Not Specified Area) due to missing spatial data were flagged for exclusion in geospatial analysis.

2. **Variable Selection**: Key variables, including `OCC_DATE`, `NEIGHBOURHOOD_158`, `LONG_WGS84`, `LAT_WGS84`, `INJURIES`, and `DEATH`, were extracted for focused analysis.

3. **Temporal Aggregation**: Data was aggregated monthly and yearly to uncover trends over time.

4. **Spatial Integration**: Neighborhood boundaries were joined with incident data to visualize shooting hotspots and analyze spatial distributions.

Table 1 provides an overview of the cleaned data used for analysis. The cleaned dataset (`cleaned_data`) forms the basis for statistical and spatial analyses, while aggregated datasets (`yearly_counts`, `monthly_counts`, and `neighbourhood_counts`) provide summaries for trend and neighborhood-specific analyses.

2.4 Measurement

The dataset offers a robust foundation for exploring firearm violence through carefully defined metrics:

1. **Temporal Dimensions**: Shooting dates (`OCC_DATE`) allow for monthly and yearly trend analysis, identifying potential seasonal patterns.

2. **Spatial Dimensions**: Neighborhood identifiers (`NEIGHBOURHOOD_158`) and geographic coordinates (`LONG_WGS84`, `LAT_WGS84`) facilitate the mapping of incidents and identification of hotspots.

3. **Severity Metrics**: Variables like `INJURIES` (non-fatal injuries) and `DEATH` (fatalities) quantify the human impact of firearm violence.

2.4.1 Challenges in Measurement

Several limitations inherent to the dataset must be acknowledged:

- **Offset Locations**: Incident locations are deliberately shifted to the nearest road intersection for privacy, which may slightly affect spatial accuracy.

- **Legacy Data**: Prior to 2014, limited geographic data results in some neighborhoods being labeled as 'NSA.'

- **Data Completeness**: Not all incidents may be reported, and those classified as firearm discharges may include non-criminal events (e.g., celebratory fire).

Despite these challenges, the dataset provides a whole view of firearm violence in Toronto, offering the specificity and breadth needed for this analysis. By carefully preparing the data and addressing potential biases, this study ensures robust and reliable understanding into the temporal and spatial patterns of shootings across the city.

3 Model

3.1 Model Set-up

Through our analysis of firearm-related incidents in Toronto, we identified a significant relationship between time and the number of shootings, injuries, and deaths. To explore these relationships and forecast future trends, we constructed two linear regression models: one to predict injuries (`INJURIES`) and another to predict deaths (`DEATH`) over time.

We formalize the models as follows:

3.1.1 Injury Model

$$Y_i^{(I)} = \beta_0^{(I)} + \beta_1^{(I)} Time\_Index_i \qquad (1)$$

3.1.2 Death Model

$$Y_i^{(D)} = \beta_0^{(D)} + \beta_1^{(D)} Time\_Index_i \qquad (2)$$

Where:

- $Y_i^{(I)}$ represents the number of injuries at time index i in the injury model (Equation 1).
- $Y_i^{(D)}$ represents the number of deaths at time index i in the death model (Equation 2).
- $\beta_0^{(I)}$ and $\beta_0^{(D)}$ are the intercepts, indicating the estimated number of injuries and deaths, respectively, at the starting point of the time index.
- $\beta_1^{(I)}$ and $\beta_1^{(D)}$ are the coefficients for the time index, capturing the rate of change in injuries and deaths over time.
- `Time_Index` is the numerical representation of time, enabling us to examine temporal trends.

The dependent variables (`INJURIES` and `DEATH`) represent the severity of incidents, while the independent variable (`Time_Index`) provides a chronological structure to detect trends and predict future values.

3.1.3 Model Justification

Linear regression was chosen for its simplicity, interpretability, and effectiveness in analyzing trends over time. Preliminary analysis indicated a linear trend in the data, supporting the use of this model. The decision to use two separate models for injuries and deaths reflects the differing dynamics and factors influencing these outcomes, ensuring a more tailored and accurate analysis.

Key justifications for the model include:

1. **Interpretability**: Linear regression provides clear coefficients that quantify the relationship between time and firearm-related outcomes.

2. **Flexibility**: The models allow for easy extension, such as adding covariates (e.g., neighborhood effects) in future analyses.

3. **Initial Simplicity**: As a foundational analysis, linear regression avoids overfitting and serves as a baseline for comparison with more complex models.

Assumptions underlying the models include:

- Linearity: A linear relationship exists between `Time_Index` and the dependent variables.

- Homoscedasticity: The variance of residuals is constant across time.

- Independence: Observations are independent of one another.

- Normality: Residuals are normally distributed.

These assumptions were validated through diagnostic plots, including residual-vs-fitted and Q-Q plots. Limitations, such as the exclusion of external factors like policy changes or socio-economic conditions, are acknowledged and discussed in the paper's conclusion.

3.2 Model Prediction

After validating the models, we extended the `Time_Index` beyond the current dataset to forecast injuries and deaths for the next 12 months. This predictive exercise provides important views into future firearm violence trends and guides proactive measures for public safety.

Using the established models:

3.2.1 Injury Prediction

$$\hat{Y}_i^{(I)} = \beta_0^{(I)} + \beta_1^{(I)} Time\_Index_i \tag{3}$$

3.2.2 Death Prediction

$$\hat{Y}_i^{(D)} = \beta_0^{(D)} + \beta_1^{(D)} Time\_Index_i \tag{4}$$

The predictions (Equation 3 & Equation 4) were calculated for a `Time_Index` incremented by 12 months, representing the forecast horizon.

3.2.3 Prediction

Figure 1 and Figure 2, juxtaposed against historical data, highlight projected trends in firearm-related injuries and deaths.

The visualizations demonstrate the models' ability to capture past trends and provide reasonable forecasts. However, caution is warranted, as the models do not account for external factors such as policy changes or significant societal events.
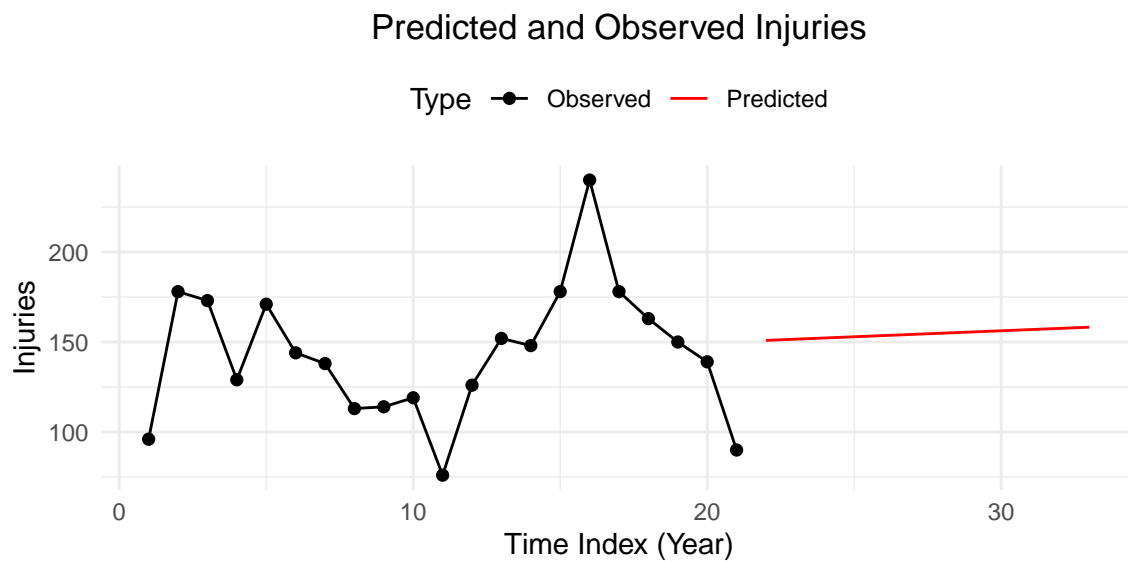
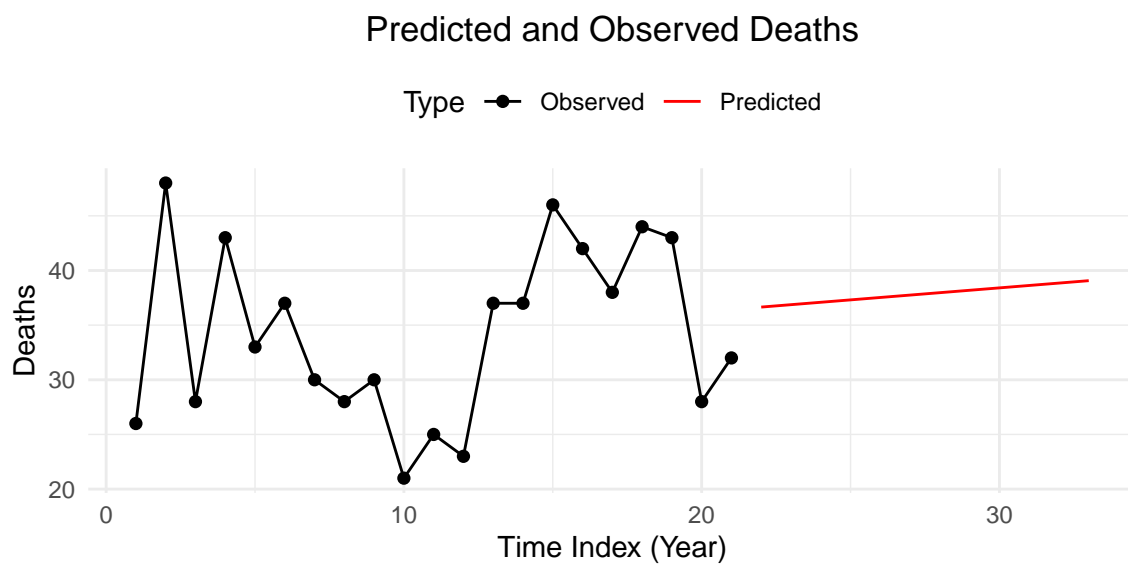Figure 1: Prediction for Injuries

Figure 2: Prediction for Deaths

3.2.4 Implications for Policy and Prevention

The predictive models emphasize the need for proactive strategies to mitigate firearm violence. Policymakers and law enforcement can use these findings to:

- Allocate resources effectively in neighborhoods with projected increases in incidents.

- Develop targeted interventions for high-risk periods and locations.

- Monitor trends to assess the effectiveness of implemented strategies.

Future refinements to the models could integrate additional predictors, such as neighborhood socio-economic factors, enforcement activity, and community programs, to enhance predictive accuracy and policy relevance. These iterative improvements will ensure that the models remain robust tools for combating firearm violence and promoting public safety.

4 Results

We offer significant findings on the temporal and spatial trends of shootings, injuries, and fatalities based on our analysis of firearm-related incidents in Toronto during the last 20 years. The results are presented through a combination of tables, figures, and statistical analyses to offer a detailed understanding of the patterns in firearm violence across the city.

4.1 Neighborhood-Specific Trends

Table 2 highlights the disparities in firearm-related incidents across Toronto's neighborhoods, showcasing the top 10 neighborhoods with the highest recorded counts. Among these, Glenfield-Jane Heights and Black Creek consistently emerge as hotspots for firearm violence. These areas demonstrate significantly higher incident rates compared to other neighborhoods, suggesting the presence of underlying socio-economic, geographic, and systemic factors contributing to these patterns.

In Figure 3, we visualize the top 10 neighborhoods with the highest counts of firearm-related incidents. These results suggest focused community safety initiatives and law enforcement presence in these areas could mitigate the risk of further incidents.

By expanding patrols, utilizing community policing, and allocating resources based on facts, law enforcement can effectively prevent firearm violence in high-risk neighborhoods. Initiatives such as youth engagement programs, job creation schemes, and mental health support in these neighborhoods can address root causes of violence and foster safer communities. By integrating these findings into a practical strategy, policymakers and stakeholders can work toward reducing firearm-related violence in Toronto's most affected neighborhoods, improving the safety and well-being of its residents.

Table 2: Sample of Neighborhood Counts of Firearm-Related Incidents

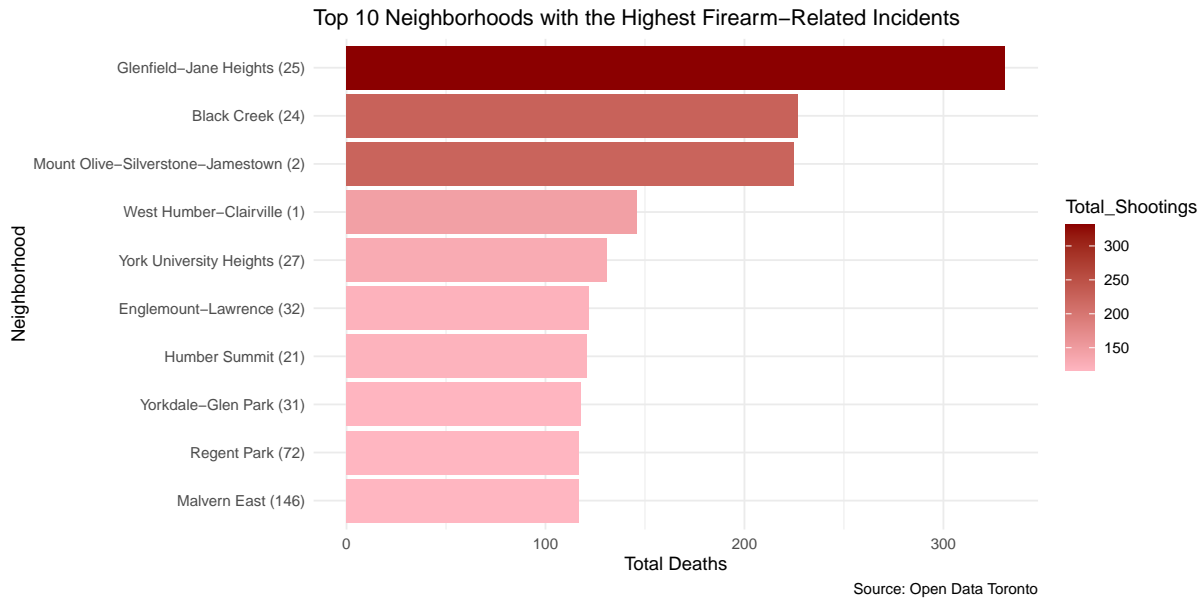| NEIGHBOURHOOD_158	Total_Shootings	Total_Injuries	Total_Deaths
Glenfield-Jane Heights (25)	331	149	21
Black Creek (24)	227	103	27
Mount Olive-Silverstone-Jamestown (2)	225	100	29
West Humber-Clairville (1)	146	79	13
York University Heights (27)	131	67	15
Englemount-Lawrence (32)	122	40	7
Humber Summit (21)	121	41	15
Yorkdale-Glen Park (31)	118	62	13
Malvern East (146)	117	41	10
Regent Park (72)	117	50	18

Figure 3: Top 10 Toronto Neighborhoods with the Highest Firearm-Related Incidents

Table 3: Yearly Counts of Firearm-Related Incidents, Injuries, and Deaths

OCC_YEAR	Total_Shootings	Total_Injuries	Total_Deaths	Time_Index
2004	191	96	26	1
2005	262	178	48	2
2006	215	173	28	3
2007	207	129	43	4
2008	238	171	33	5
2009	252	144	37	6
2010	259	138	30	7
2011	227	113	28	8
2012	219	114	30	9
2013	204	119	21	10
2014	177	76	25	11
2015	288	126	23	12
2016	407	152	37	13
2017	392	148	37	14
2018	427	178	46	15
2019	492	240	42	16
2020	462	178	38	17
2021	409	163	44	18
2022	380	150	43	19
2023	345	139	28	20
2024	356	90	32	21

4.2 Annual Trends in Shootings, Injuries, and Deaths

Table 3 summarizes the yearly progression of shootings, injuries, and deaths in Toronto from 2004 to 2024. A clear upward trend in injuries and deaths highlights the increasing severity of firearm violence over time.

In Figure 4, we visualize the annual counts of shootings, injuries, and deaths. The steady rise in injuries and deaths necessitates policy attention and proactive measures.

4.3 Geospatial Analysis

The spatial distribution of firearm-related incidents is shown in Figure 5, where neighborhoods with high concentrations of shootings are highlighted. This geospatial analysis provides a clear picture of the areas most affected by firearm violence.

Neighborhoods such as Glenfield-Jane Heights, Black Creek, and Mount Olive-Silverstone-Jamestown exhibit consistently high densities of firearm-related incidents. These areas emerge as important hotspots requiring immediate attention. The map shows clusters of high-density firearm violence, particularly in the northwest and downtown areas of Toronto. This clustering underscores the non-random nature of firearm violence and suggests the influence of localized
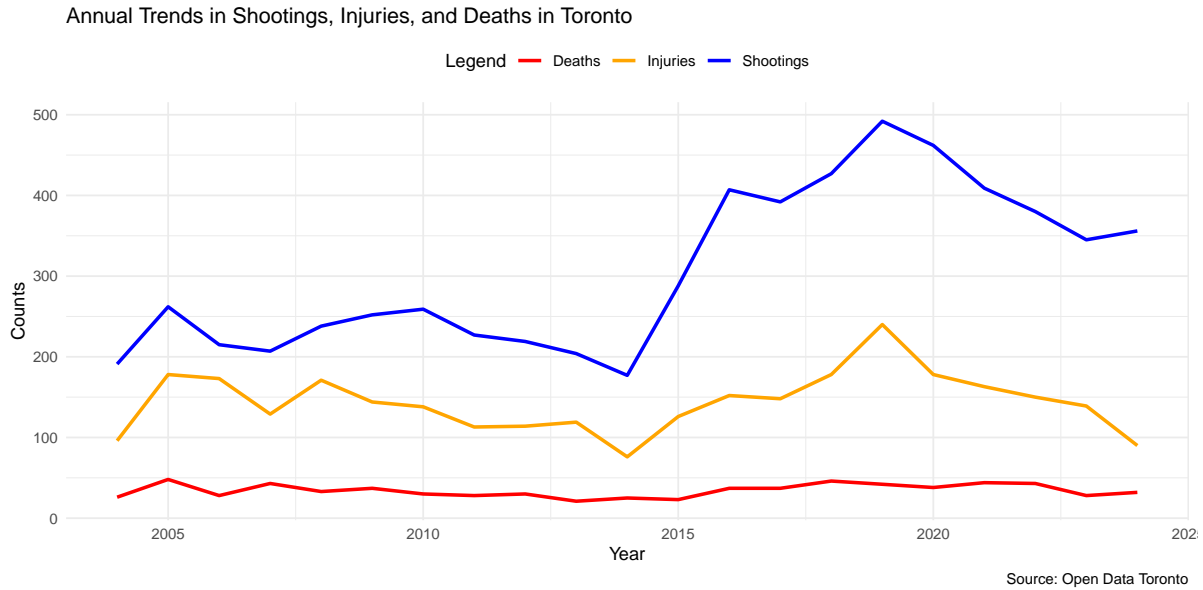
Figure 4: Annual Trends in Shootings, Injuries, and Deaths

socio-economic or environmental factors. The geospatial analysis enables policymakers to allocate law enforcement and community resources more effectively by focusing on neighborhoods with the highest densities of firearm incidents.

4.4 Shiny Web App: Toronto Shooting Incidents Explorer

Shiny Web App is available at: Toronto Shooting Incidents Explorer

To enhance public accessibility and engagement with our analysis, we developed a Shiny web application. This tool allows users to dynamically explore firearm violence data in Toronto from 2004 to 2024, offering an interactive platform to investigate trends and hotspots. The app features two main functionalities:

- **Table View**: Users can select specific years or view aggregated data for all years. The table provides detailed attributes such as the date, neighborhood, and type of incident (injury or fatality). This view enables users to examine the granular details of each firearm-related occurrence.

- **Map View**: Using Leaflet(Cheng, Karambelkar, and Xie 2021), the app displays the geographic distribution of firearm incidents. Each marker corresponds to an incident, plotted based on its longitude and latitude. Users can zoom in on specific neighborhoods to identify high-risk zones, helping to visualize the spatial concentration of firearm violence.
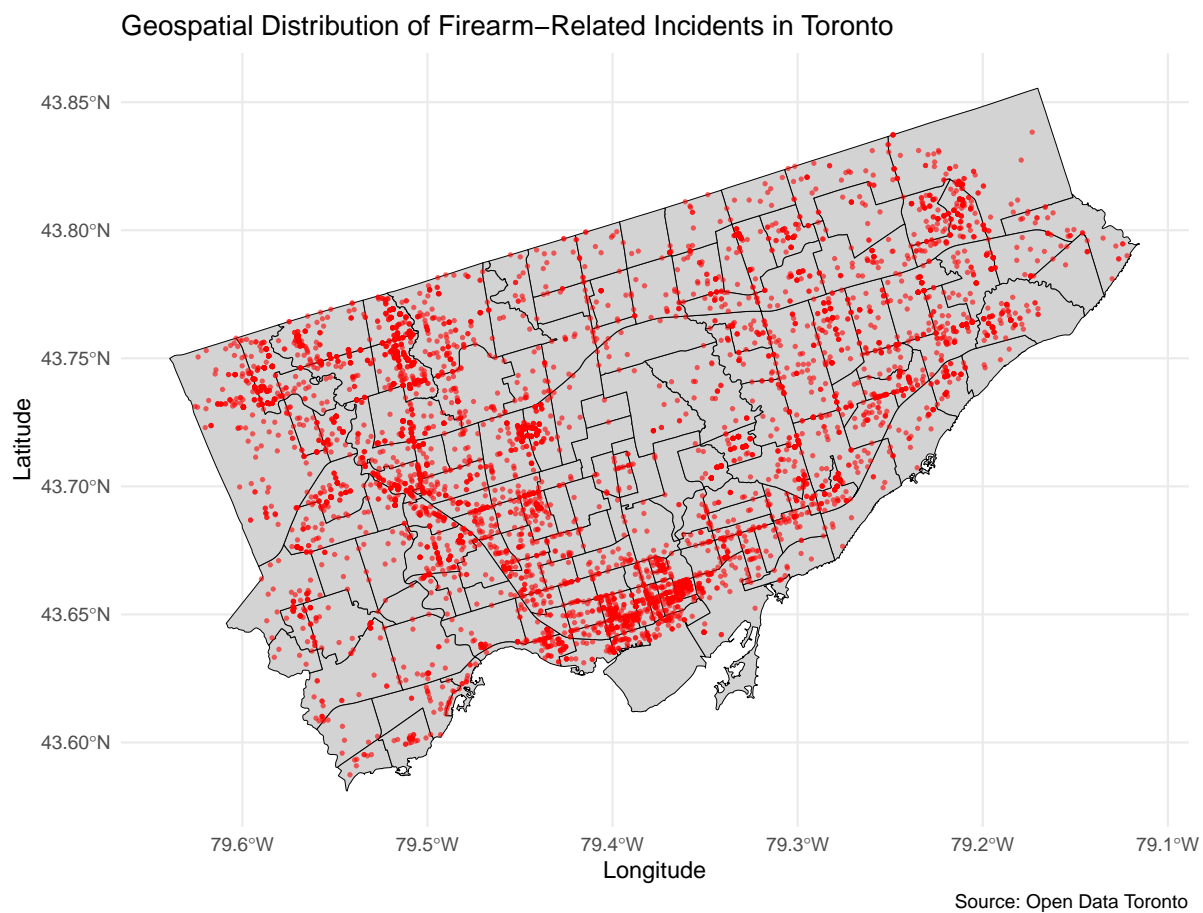
Figure 5: Toronto Firearm-Related Incident Locations

This Shiny app provides a practical tool for researchers, policymakers, and the general public. It empowers users to interact with the dataset, fostering a deeper understanding of the temporal and spatial dynamics of firearm violence in Toronto. By offering a user-friendly interface, the app encourages evidence-based discussions and strategies for crime prevention.

4.5 Model Summary and 90% Credibility Interval

Figure 6 encapsulates the results of our linear regression models, displaying the correlation between time and the number of firearm-related injuries and deaths. These models effectively capture temporal trends, providing a foundation for forecasting and policymaking.
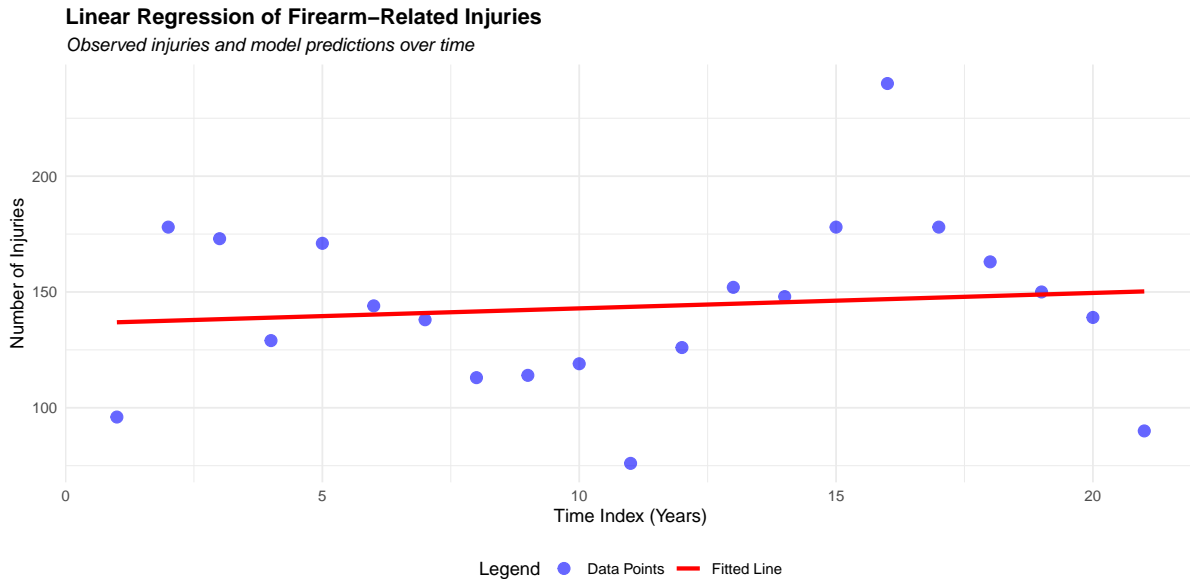

Figure 6: Linear Regression of Firearm-Related Incidents

In addition to the visualizations, Table 4 summarizes the regression coefficients, indicating the intercepts and slopes for injuries and deaths. These models offer broad perspectives into the rising rates of violence involving firearms.

Figure 7 and Figure 8 display the 90% credibility intervals for the coefficients. This visualization underscores the reliability of the estimates and their applicability for future predictions.

Together, the regression models, credibility intervals, and visualizations provide a robust framework for understanding and anticipating trends in firearm violence. Evidence-based initiatives for reducing deaths and injuries in Toronto are guided by these findings, which are a useful tool for stakeholders and policymakers.

Table 4: Regression Model Results for Firearm-Related Injuries and Deaths

	Injury Model	Death Model
(Intercept)	136.24	31.82
	(17.16)	(3.67)
Time_Index	0.67	0.22
	(1.37)	(0.29)
Num.Obs.	21	21
R2	0.012	0.029
R2 Adj.	−0.040	−0.022
AIC	216.2	151.4
BIC	219.3	154.6
Log.Lik.	−105.094	−72.711
RMSE	36.07	7.72

90% Credibility Interval for Injury Model Coefficients
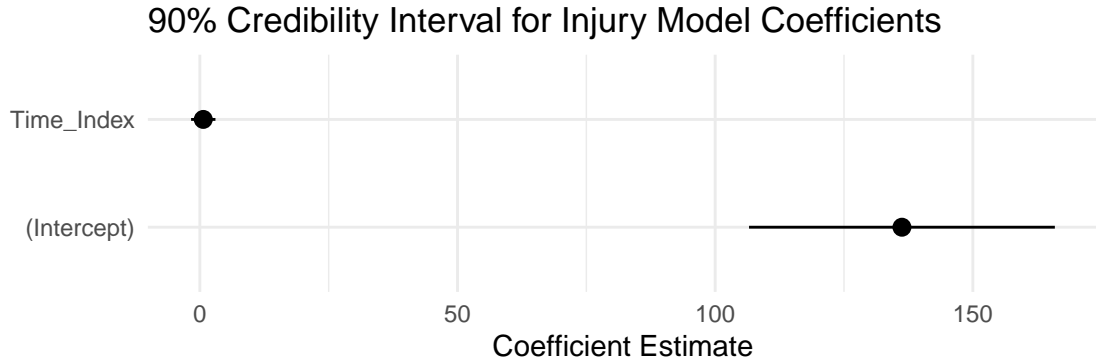
Figure 7: 90% Credibility Interval for Injury Model

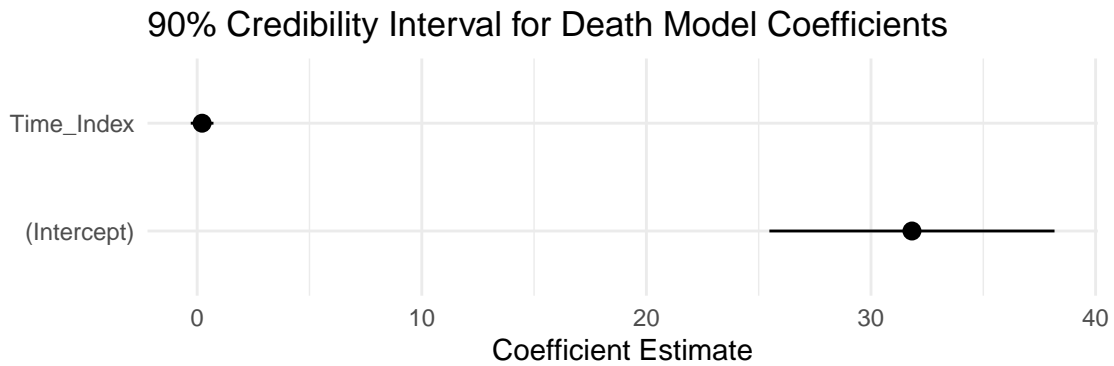### 90% Credibility Interval for Death Model Coefficients

Figure 8: 90% Credibility Interval for Death Model

5 Discussion

5.1 Seasonal patterns in Firearm Violence

Our study highlights a concerning seasonal pattern in firearm-related incidents in Toronto, with a noticeable surge during the summer months compared to the winter. To think deeper, Figure 9, labeled "Monthly Firearm-Related Incidents in 2022 and 2023," focuses on the monthly distribution of these incidents. This detailed visualization shows that the summer months consistently exhibit higher incident counts.

Monthly Firearm–Related Incidents in 2022 and 2023

Comparative monthly counts of incidents in 2022 and 2023.
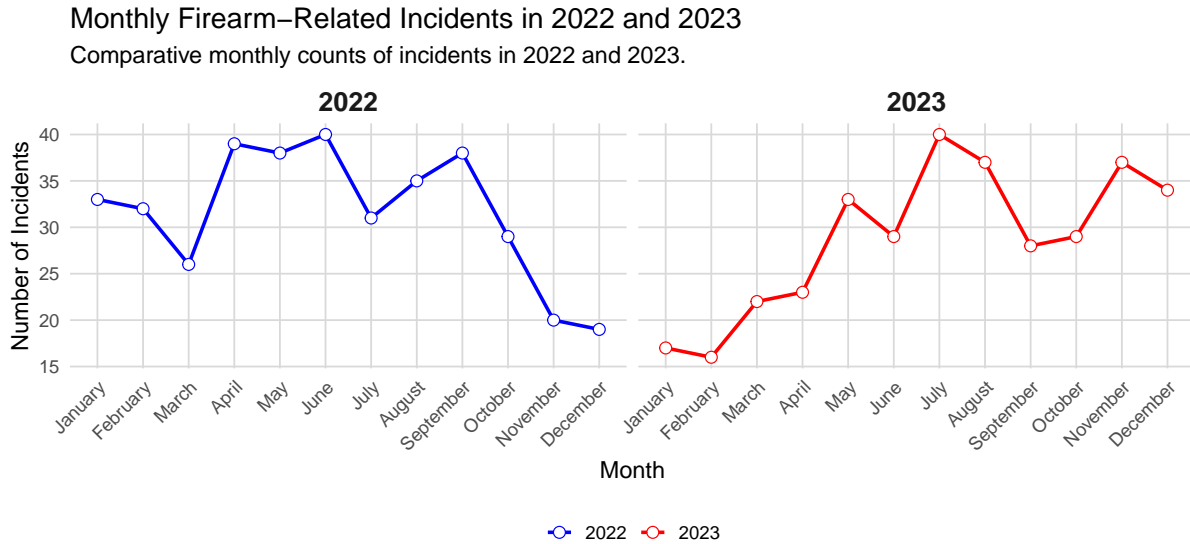

Source: Open Data Toronto

Figure 9: Monthly Firearm-Related Incidents in 2022 and 2023

This trend underscores the vital need for targeted interventions during the summer months to address the seasonal spike in firearm violence. The visualization highlights the importance of proactive measures and evidence-based policies tailored to address the underlying factors contributing to this seasonal variation in firearm violence.

5.2 Spatial Dynamics and Hotspot Analysis

Our spatial analysis shows that firearm violence is not randomly distributed across Toronto but is concentrated in specific neighborhoods. Hotspots such as Glenfield-Jane Heights and Black Creek emerge as epicenters of firearm violence. These areas exhibit persistently high incident counts, signaling the need for focused law enforcement efforts and community interventions.

17

The geospatial distribution of incidents, visualized in Figure 5, reinforces the important role of place-based strategies in addressing urban crime. Factors such as population density, socio-economic conditions, and urban design may contribute to the clustering of firearm violence, necessitating localized and multi-faceted responses.

5.3 Predictive Insights and Future Trends

Our linear regression models provide a robust framework for understanding the temporal trends in firearm-related injuries and deaths. The upward trajectory in these metrics, if unaddressed, suggests a continued escalation in firearm violence. The predictive models underscore the importance of proactive strategies, enabling policymakers to allocate resources effectively and develop targeted interventions.

While the models demonstrate strong predictive capabilities, they do not account for external variables such as economic conditions, changes in policing strategies, or broader societal shifts. Incorporating these factors in future models could enhance predictive accuracy and provide a more detailed understanding of the drivers of firearm violence.

5.4 Limitations and Weaknesses

Despite the strengths of our study, certain limitations warrant acknowledgment:

- **Data Completeness**: Our reliance on officially reported data excludes unreported incidents, potentially underestimating the true scale of firearm violence.

- **Geographic Privacy Measures**: The offsetting of incident locations for privacy reasons may affect the spatial accuracy of our analyses.

- **Model Constraints**: While the linear models effectively capture temporal trends, they lack the granularity to incorporate additional covariates such as socio-economic indicators or enforcement activities.

These limitations highlight the need for ongoing data collection and the integration of more diverse datasets to enhance the robustness of future analyses.

5.5 Future Exploration and Recommendations

Building on the current study, future research should aim to:

1. **Incorporate Additional Variables**: Expanding the scope of the analysis to include socio-economic factors, policing strategies, and community programs will provide a more detailed understanding of firearm violence.

2. **Explore Causal Relationships**: Moving beyond correlation, future studies could investigate causal links between economic conditions, urban design, and firearm violence.

3. **Develop Predictive Models**: Integrating machine learning techniques could enhance the predictive capabilities of models, accommodating non-linear relationships and a broader range of predictors.

Investigating these topics will help us better understand firearm violence and guide the creation of focused, research-based public safety initiatives.

6 Appendix: Surveys, Sampling, and Observational Data

6.1 Overview of Observational Data in Crime Analysis

The study of firearm-related violence relies heavily on observational data collected through official reporting mechanisms such as police records, community surveys, and open data platforms. Observational data, unlike experimental data, is inherently non-random and reflects the real-world phenomena being studied. This appendix explores the role of observational data in understanding firearm violence trends, the sampling methodologies used, and the limitations and biases inherent in such data. Additionally, it discusses potential remedies to improve data reliability and applicability for policy-making.

6.2 Source of Observational Data

In this study, the primary observational dataset is the Toronto Police Service's Shooting and Firearm Discharge Data. This dataset includes all reported incidents involving firearm discharges, whether resulting in injuries, fatalities, or no physical harm. Each incident is recorded with the following attributes:

- **Temporal Variables**: Incident dates allow for the analysis of monthly and yearly trends.

- **Spatial Variables**: Geographical coordinates and neighborhood identifiers provide views into the spatial distribution of firearm incidents.

- **Severity Variables**: Metrics such as injuries and deaths quantify the human impact of each incident.

The dataset was supplemented with neighborhood profiles from the City of Toronto, enabling the integration of socio-economic and demographic variables to contextualize the spatial patterns observed.

6.3 Sampling Design and Representativeness

The observational data used in this study is collected through a passive reporting system where police records serve as the primary source of information. The dataset represents a form of convenience sampling, as it only includes incidents reported to or recorded by law enforcement. This sampling method has several implications:

1. **Non-Random Nature of Data**: The dataset does not represent a random sample of all firearm-related incidents in Toronto. Factors such as neighborhood-level trust in law enforcement, the visibility of incidents, and individual reporting behavior can influence the likelihood of an incident being recorded.

2. **Underreporting Bias**: Certain firearm-related incidents, particularly those involving no injuries or fatalities, may go unreported. This introduces an underreporting bias, leading to an underestimation of the true prevalence of firearm violence.

3. **Geographic Bias**: Data completeness may vary by neighborhood. Areas with higher police presence or community surveillance may exhibit higher reporting rates, skewing spatial analyses.

4. **Temporal Consistency**: Reporting practices may have evolved over the two decades covered by this study. Changes in data collection protocols or public awareness campaigns could influence the comparability of data across years.

6.4 Statistical Challenges in Observational Data Analysis

1. **Confounding Variables**: Observational data is subject to confounding, where variables not included in the dataset may influence the observed relationships. For example, socio-economic factors, changes in law enforcement strategies, or policy interventions could affect both firearm violence rates and reporting behaviors.

2. **Simpson's Paradox**: Aggregating data without accounting for neighborhood-level differences or temporal trends may lead to misleading conclusions. For example, a decrease in firearm violence may mask increasing trends in specific neighborhoods.

3. **Spatial Autocorrelation**: Firearm incidents are likely spatially correlated, with neighboring areas experiencing similar levels of violence due to shared socio-economic and environmental factors. Ignoring spatial dependence in statistical models can lead to biased estimates and invalid inference.

6.5 Simulation Study: The Impact of Underreporting

To assess the potential impact of underreporting bias, a simulation was conducted. The true incident count in each neighborhood was assumed to follow a Poisson distribution with a mean based on observed counts. Underreporting was simulated by randomly excluding 20-50% of incidents, reflecting varying levels of trust in law enforcement across neighborhoods. The results indicated that:

- High levels of underreporting led to significant underestimation of incident rates in marginalized neighborhoods.

- The spatial distribution of incidents appeared more concentrated in well-surveyed neighborhoods, masking true hotspots in underreported areas.

- Temporal trends were relatively robust to underreporting but showed greater variability in neighborhoods with low reporting rates.

6.6 Improving Observational Data for Crime Analysis

1. **Incorporating Survey Data**: Supplementing police records with community surveys can provide views into unreported incidents and public perceptions of firearm violence. Randomized household surveys could help estimate the true prevalence of firearm-related incidents.

2. **Improving Spatial Precision**: While privacy concerns necessitate offsetting geographic coordinates, the use of aggregated spatial units (e.g., census tracts) with socio-economic data can enhance analysis without compromising confidentiality.

3. **Adjusting for Underreporting**: Statistical techniques such as capture-recapture methods or Bayesian hierarchical models can adjust for underreporting by incorporating external information on reporting behaviors.

4. **Enhancing Data Transparency**: Open data platforms should provide metadata detailing changes in reporting practices, data collection methods, and potential biases to improve interpretability and reproducibility.

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.

Cheng, Joe, Bhaskar Karambelkar, and Yihui Xie. 2021. *leaflet: Create Interactive Web Maps with the JavaScript "Leaflet" Library.* https://CRAN.R-project.org/package=leaflet.

Firke, Sam. 2023. *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://CRAN.R-project.org/package=here.

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames.* https://CRAN.R-project.org/package=tibble.

Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10 (1): 439–46. https://doi.org/10.32614/RJ-2018-009.

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

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, and Apache Arrow. 2023. *arrow: Integration to Apache Arrow.* https://CRAN.R-project.org/package=arrow.

Robinson, David, Alex Hayes, and Simon Couch. 2022. *broom: Convert Statistical Objects into Tidy Tibbles.* https://CRAN.R-project.org/package=broom.

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

Social Development, Finance & Administration. 2024. *Neighbourhoods.* Open Data Toronto. https://open.toronto.ca/dataset/neighbourhoods/.

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. 2023. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with "kable" and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.