

# Exploring Temporal and Spatial Patterns of Major Crime Indicators\*

An analysis for Toronto Urban Safety and Policy Design

Yuechen Zhang

December 2, 2024

This paper analyses spatial and temporal trends in major crime in Toronto using data on key crime indicators from 2014 to the present. Linear regression models were used to examine crime categories, temporal patterns, and their prevalence in the community. The findings in the paper suggest that certain crime categories, such as auto theft and robbery, exhibit significant fluctuations over time, with higher frequencies in urban areas compared to suburban neighbourhoods. In addition, the findings highlight that year-to-year changes in the number of offences are influenced by the prevailing socio-economic and other factors. These insights underscore the importance of data-driven strategies for crime prevention and urban policy planning, thereby ensuring a trend towards safer community and social development.

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Estimand . . . . .	4
<b>2</b>	<b>Data</b>	<b>4</b>
2.1	Data Overview . . . . .	4
2.2	Data Table . . . . .	4
2.3	Data Measurement . . . . .	5
2.3.1	Data Consideration . . . . .	5
2.4	Methodology . . . . .	5
2.5	Data Visualization . . . . .	6

---

\*Code and data are available at: <https://github.com/Yuechen-Zhang603/Major-Crime-Indicators>

<b>3 Model</b>	<b>10</b>
3.1 Model Set up . . . . .	10
3.2 Model justification . . . . .	11
<b>4 Results</b>	<b>12</b>
4.1 Model Validation . . . . .	13
<b>5 Discussion</b>	<b>17</b>
5.1 Relationship between Time-based and Crime Indicators . . . . .	17
5.2 Relationship between Geographical distribution and Crime Indicators . . . . .	17
5.3 Government Policy Implications . . . . .	18
5.4 Limitations . . . . .	18
5.4.1 Simplified Representation of Temporal Trends . . . . .	18
5.4.2 Lack of Detailed Socioeconomic Context . . . . .	18
5.4.3 Limitations of Linear Assumptions . . . . .	19
5.5 Future Steps . . . . .	19
<b>Appendix</b>	<b>20</b>
<b>A Data</b>	<b>20</b>
A.1 Data Collection Process . . . . .	20
A.2 Biases and Limitations . . . . .	20
A.3 Exploring Bias Through Simulation . . . . .	20
A.4 Simulation to Address Underreporting: . . . . .	21
A.5 Recommendations for Data Improvement: . . . . .	21
<b>B Model details</b>	<b>21</b>
B.1 Model Strengths . . . . .	21
B.2 Limitations . . . . .	22
B.3 Future Model Enhancements . . . . .	22
B.4 Simulation for Model Validation . . . . .	22
B.5 Future Model Enhancements . . . . .	23
<b>References</b>	<b>24</b>

## 1 Introduction

Despite numerous initiatives aimed at reducing crime in urban areas, major crimes such as assaults, robberies and auto thefts continue to be a significant challenge for communities throughout Toronto. Crime not only jeopardises public safety, but also affects socio-economic stability and the quality of life of residents. Within the past year, the government announced \$390 million to help stop crime and violence (Canada 2023). While the state has invested heavily in intervening in the crime situation, major crimes like assaults, robberies, and auto thefts have been prevalent over the past decade, with significant increases in certain categories. In 2014, there were 32,461 major crimes reported in Toronto, and in 2024, this number increases to 35,950, a 10.7 per cent increase in ten years. The number of major crimes reported in Toronto has risen from 1.7 per cent in 2014 to 1.3 per cent in 2024, which is a significant increase. These statistics highlight the need for a deeper understanding of the factors that contribute to crime in order to develop more effective policies to enhance public safety.

Research into crime trends often leverages data from law enforcement agencies to identify the causes and predictors of crime rates. The dataset used in this study, the Toronto Major Crime Indicators dataset, provides detailed records of reported crimes across multiple categories, including assault, robbery, break and enter, auto theft, and theft over. By analyzing this dataset, we can uncover temporal and spatial crime patterns and understand how crime has evolved over the years in Toronto's neighborhoods.

This paper builds models to analyze Toronto's crime patterns, focusing on the years 2014 and 2024. Using linear regression models, we examine the relationships between major crime categories, annual trends, and neighborhood-specific data. These models help us explore questions such as whether certain crime categories are more likely to increase over time or whether specific neighborhoods experience higher crime rates consistently.

Our findings reveal that major crimes such as auto theft and robbery show disproportionate increases, with particular neighborhoods seeing higher crime prevalence. These insights emphasize the importance of targeted policies for crime reduction and resource allocation. Additionally, by comparing 2014 and 2024 data, we highlight the growing severity of crime in Toronto and suggest data-driven approaches to mitigate these issues.

The remainder of this paper is structured into several sections. Section 2 presents the data used for the analysis, including key tables and graphs to illustrate the distribution of crimes and associated features across neighborhoods and time periods. Section 3 describes the formulation and justification of the regression model used to analyze the relationships between crime rates and explanatory variables, such as time, geography, and crime type. Section 4 highlights the findings from the model, providing visual and tabular summaries of the predicted effects and observed patterns. Section 5 discusses the implications of these findings, focusing on crime hotspots, temporal trends, and the influence of socio-geographic factors. Statistical programming language R (R Core Team 2023) is used in this report, with packages `tidyverse` (Wickham et al. 2019), `here` (Müller 2020), `rstanarm` (Brilleman et al. 2018), `modelsummary`

(Arel-Bundock 2022), `ggplot2` (Wickham 2016), `knitr` (Xie 2014), `marginalEffects` (Arel-Bundock 2024), `plotly` (Sievert 2020), `tibble` (Müller and Wickham 2023), `margins` (Leeper 2021), `testthat` (Wickham 2011) and `kableExtra` (Zhu 2021).

## 1.1 Estimand

The estimated in this analysis is the predicted number of crimes occurring across various neighborhoods in Toronto. While it is challenging to measure the exact number of crimes due to limitations such as underreporting and unobserved incidents, this paper aims to estimate the expected number of crimes using available data from the City of Toronto's Major Crime Indicators dataset.

Factors like the accuracy of police reports, differences in community reporting behavior, and discrepancies in categorization of crimes could affect the data. For instance, crimes occurring in less densely populated neighborhoods or crimes perceived as minor might be underreported, leading to potential bias in the estimates.

# 2 Data

## 2.1 Data Overview

This report uses the Key Crime Indicators dataset from the website open data toronto to provide comprehensive data on reported criminal activity as our primary data source. These data include all MCI incidents reported to the Toronto Police Service, including those for which the location could not be verified. (City of Toronto 2024). The authenticity of the data can be guaranteed because the publisher clearly states that the data excludes incidents that are considered unfounded. Following a police investigation, it was confirmed that the reported crime did not occur and that there were no attempts to be recognized as unfounded incidents by Statistics Canada.

For the data section, this paper will focus on analyzing trends over time (e.g., monthly, yearly) based on different variables in the dataset to identify patterns or spikes in certain types of crime. And using geographic data to map crime hotspots and understand spatial distribution so that rates and types of crime can be compared across neighborhoods.

After loading the dataset using the R programming language (R Core Team 2023) and the `here` package (Müller 2020), the `tidyverse` (Wickham et al. 2019) package was used to generate graphs. In doing so, R code was adapted from (Alexander 2023).

## 2.2 Data Table

Table 1: Top 10 Neighborhoods by Total Crime Incidents

Neighborhood	Assaults	Break & Enters	Drug Arrests	Thefts	Total Incident
Waterfront Communities-The Island (77)	9978	3012	0	0	1548
Church-Yonge Corridor (75)	8927	2138	0	0	1366
West Humber-Clairville (1)	3103	1522	0	0	1132
Bay Street Corridor (76)	6713	1761	0	0	1035
Moss Park (73)	5847	1605	0	0	921
Downsview-Roding-CFB (26)	4660	971	0	0	812
York University Heights (27)	3776	1345	0	0	798
Kensington-Chinatown (78)	4531	1551	0	0	739
Islington-City Centre West (14)	2797	1326	0	0	719
Woburn (137)	4131	927	0	0	693

Figure 1: Toronto Major Crime Indicators

## 2.3 Data Measurement

Data was collected primarily through reports uploaded by the Toronto Police Service. The data excludes incidents that are considered unfounded. According to Statistics Canada, unfounded is defined as “a reported crime that did not occur and was not attempted, as determined through a police investigation” (Statistics Canada, 2020). Because these data are provided at the crime and/or victim level, an incident number may have several rows of data related to the various MCIs used to categorize the incident. (City of Toronto 2024).

### 2.3.1 Data Consideration

This dataset represents a subset of reported crime in Toronto and may not account for unreported incidents, resulting in underrepresentation of certain crime types. Reliance on police reports introduces potential biases, such as differences in reporting practices between departments or differences in the way crimes are categorized.

## 2.4 Methodology

The raw dataset of crime records in Toronto was cleaned and prepared for analysis through a series of steps to ensure accuracy and consistency. Duplicate rows were removed to retain unique crime incidents, and records containing missing (NA) or invalid ("NSA") values were filtered out. Irrelevant columns, including the last column with incomplete data, were excluded. Key columns were renamed for clarity, such as EVENT\_UNIQUE\_ID to Event\_ID and LAT\_WGS84 to Latitude. A new Date column was derived by combining Year, Month, and Day columns to

standardize temporal data. The dataset was then ordered chronologically by the `Date` column for easier time-series analysis. The cleaned dataset, containing essential features like `Event_ID`, `Crime_Type`, `Neighborhood`, `Latitude`, and `Longitude`, was exported to a new CSV file for further analysis, ensuring a reliable and consistent dataset for subsequent exploration and modeling.

After cleaning, there are 402326 obs of 29 variables in the dataset, Table 2 shows a preview of the cleaned dataset.

```
[1] 29
```

Crime Details								
Column_1	Column_2	Column_3	Column_4	Column_5	Column_6	Column_7	Column_8	Column_9
1	GO-20141261609	2014/1/1	2014/1/1	2014	January		1	
2	GO-20141260033	2014/1/1	2013/12/31	2014	January		1	
3	GO-20141260127	2014/1/1	2014/1/1	2014	January		1	
4	GO-20141260597	2014/1/1	2014/1/1	2014	January		1	
5	GO-20141260618	2014/1/1	2014/1/1	2014	January		1	
6	GO-20141260618	2014/1/1	2014/1/1	2014	January		1	
7	GO-20141260056	2014/1/1	2014/1/1	2014	January		1	
8	GO-20141263672	2014/1/1	2014/1/1	2014	January		1	
9	GO-20141262608	2014/1/1	2014/1/1	2014	January		1	
10	GO-20141263706	2014/1/1	2014/1/1	2014	January		1	

## 2.5 Data Visualization

Figure 2 provides a detailed overview of the total reported crimes in Toronto from 2014 to 2024. It illustrates significant fluctuations in crime rates over the years, with key trends emerging from the data. Starting in 2014, the total reported crimes were relatively stable at approximately 35,000 incidents. A gradual increase is observed from 2015 to 2018, reaching around 40,000 incidents by 2018.

A notable peak occurs in 2019, with total reported crimes surging to over 45,000, suggesting a significant spike in criminal activity during that year. This could potentially be linked to socio-economic changes or other external factors influencing crime rates during this period. Following this peak, a decline is observed in 2020 and 2021, where crime rates drop back to

approximately 35,000, potentially reflecting changes during the COVID-19 pandemic, including lockdowns and reduced public mobility.

However, in 2022, crime rates sharply rise again, exceeding 45,000 incidents, marking the highest point in the observed period. This sudden increase may reflect a post-pandemic resurgence of activity and mobility in urban areas. By 2024, the total reported crimes decline significantly, falling back to below 40,000 incidents, indicating potential effectiveness of recent crime prevention strategies or other mitigating factors.

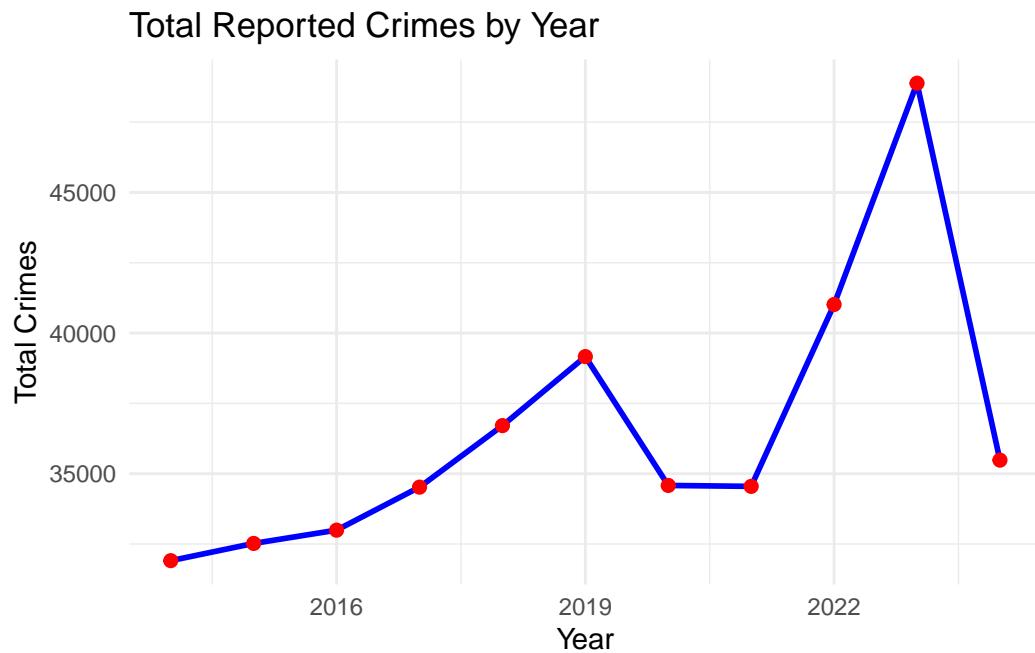


Figure 2: Time-Series Plot of Total Reported Crimes by Year in Toronto

Figure 3 illustrates the distribution of reported crimes in Toronto categorized by major crime types. The visualization highlights the relative prevalence of each crime type, providing insights into which offenses contribute most significantly to the overall crime burden in the city.

From the graph, it is evident that Assault represents the largest category of reported crimes, far surpassing other crime types. This indicates that violent interactions are a key area of concern for law enforcement and policymakers. Following Assault, Break and Enter and Auto Theft are the next most frequent crime categories, reflecting issues related to property security and vehicle safety. Robbery and Theft Over are less frequent but remain notable contributors to the overall crime landscape.

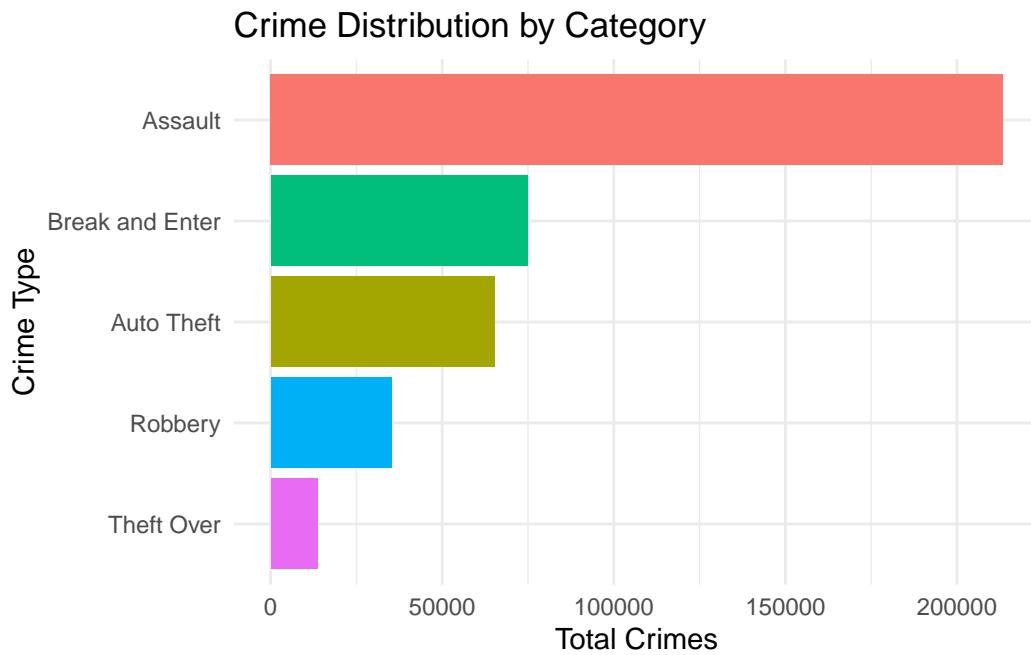


Figure 3: Bar Plot Showing the Distribution of Crimes by Category in Toronto

Figure 4 visualizes the spatial distribution of reported crimes across Toronto. Each point on the map represents a reported crime, plotted using geographic coordinates (latitude and longitude) over the city's boundaries. The visualization provides a comprehensive view of how crime incidents are geographically dispersed throughout Toronto.

From the map, it is evident that crimes are concentrated in specific urban areas, with a higher density in downtown Toronto compared to suburban and outlying neighborhoods. This spatial clustering indicates potential hotspots of criminal activity, often associated with higher population density, commercial zones, or socio-economic disparities. Conversely, areas with fewer points suggest lower reported crime rates, possibly reflecting suburban or less populated regions.

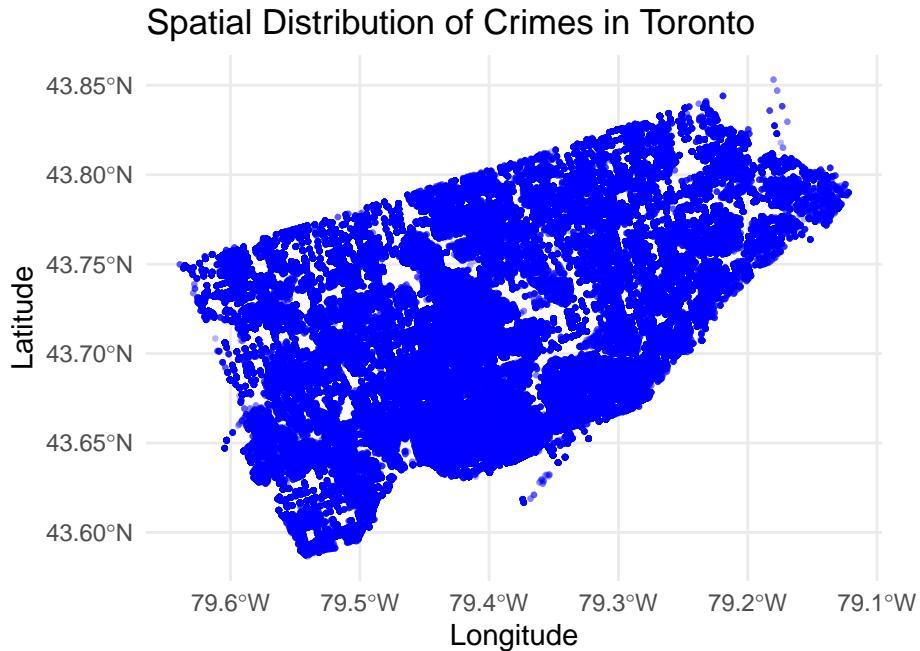


Figure 4: Spatial Distribution of Crimes in Toronto Visualized on a Geographic Map

## 3 Model

In our analysis, we utilized a Linear Regression model to examine the relationship between crime rates and a variety of factors, including temporal indicators (e.g., year, time of day), spatial variables (e.g., geographic coordinates, neighborhood), and crime-specific characteristics (e.g., crime type, division). Linear regression provides a straightforward and interpretable framework to quantify how these predictors influence crime rates, making it ideal for understanding patterns and informing policy decisions.

### 3.1 Model Set up

The model is formulated as follows:

$$y_i = \beta_0 + \beta_1 \cdot \text{Year}_i + \beta_2 \cdot \text{Hour}_i + \beta_3 \cdot \text{Longitude}_i + \beta_4 \cdot \text{Latitude}_i + \sum_{j=1}^k \gamma_j \cdot \text{CrimeType}_{ij} + \sum_{l=1}^m \delta_l \cdot \text{Division}_{il} + \epsilon_i$$

Where:

- $y_i$ : The predicted number of crimes in observation ( i ).
- $\beta_0$ : Intercept term, representing the baseline level of crime when all predictors are at their reference levels.
- $\beta_1, \beta_2, \beta_3, \beta_4$ : Coefficients for temporal and spatial variables.
- $\gamma_j$ : Coefficients for categorical crime type variables.
- $\delta_l$ : Coefficients for categorical police division variables.
- $\epsilon_i$ : Residual errors, assumed to be normally distributed.

The target variable  $y_i$  in our model represents the observed crime count for a specific instance in the dataset. The predictors include:

#### Temporal Variables

- Year (Year): Captures changes in crime rates over time, reflecting long-term trends.
- Hour (Hour): Accounts for daily crime patterns, such as higher incidences during peak hours.

#### Spatial Variables

- Longitude and Latitude (Longitude,Latitude): Represent geographic locations of crimes, enabling spatial analysis to identify high-crime areas.
- Neighborhood (Division): Encoded as dummy variables, these capture the effects of police divisions or regions on crime rates.

Crime-Specific Variables · Crime Type (CrimeType): Also encoded as dummy variables, these distinguish betwe categories such as theft, assault, and robbery.

The linear regression model assumes that the crime count  $y_i$  is a linear combination of these predictors, with coefficients  $\beta$  quantifying the impact of each variable.

We selected linear regression for this analysis because of its interpretability and ability to provide clear, quantitative relationships between predictors and the target variable. This simplicity is particularly valuable for policymakers and urban planners, who can leverage these insights to target specific variables (e.g., certain neighborhoods or times of day) in crime prevention strategies. Linear regression is also computationally efficient, allowing us to process the large dataset ( $n=402326$ ) with ease.

While alternative models, such as Random Forest or Bayesian Regression, could potentially capture more complex interactions or provide uncertainty quantification, linear regression remains a robust starting point for exploratory analysis. The assumption of linearity in relationships, though somewhat restrictive, is appropriate for identifying general trends in crime data.

### 3.2 Model justification

Regarding the relationship between crime rates and temporal factors, we anticipate that time-related variables such as the year and hour of the day significantly influence crime trends. The year (( Year\_i )) captures long-term changes, reflecting broader societal and economic shifts. For example, as cities grow and technology advances, the frequency and type of crimes may change. Additionally, social policies and law enforcement strategies implemented in specific years could directly impact crime rates. The hour (( Hour\_i )) of the day is equally critical, as certain hours, such as late evenings or early mornings, are associated with increased crime activity due to decreased surveillance and higher social activities like nightlife.

Spatial factors, such as geographic coordinates (( Longitude\_i ) and ( Latitude\_i )) and neighborhood divisions (( Division{il} )), play a vital role in understanding where crimes are concentrated. High-crime areas often have specific characteristics, such as socioeconomic challenges, poor infrastructure, or limited law enforcement presence. Geographic locations help identify patterns such as clusters of crimes near transit hubs, parks, or commercial areas. Neighborhood divisions encoded as categorical variables further highlight disparities across regions, allowing for a deeper understanding of how local characteristics, like economic conditions or population density, influence crime rates.

Crime-specific variables, such as the type of crime (( CrimeType\_{ij} )), are essential for distinguishing between different categories of offenses. For instance, thefts might show different spatial and temporal patterns compared to violent crimes like assaults. By including these categories as predictors, the model can tailor insights to specific crime types, enabling law enforcement to address each issue effectively.

## 4 Results

The findings align with expectations and offer insights into the temporal, spatial, and categorical predictors of crime in Toronto. To avoid multicollinearity, the model excludes one variable from each category: crime type “Theft” and police division “Downtown Core.” The intercept represents the estimated baseline level of crime in the reference group, which is the crime type “Theft” in the “Downtown Core” division during the year 2014 and at midnight (Hour = 0). The estimated intercept is 2.103, indicating a log count of 2.103 crimes under these conditions.

**Temporal Variables** Yearly trends reveal that the likelihood of crime occurrence has changed over time. The estimated coefficient for the year variable is  $\beta_1=-0.034$ , suggesting that holding all other variables constant, each additional year is associated with a decrease in the log count of crimes by 0.034 units. This result aligns with the expected decline in certain crime categories due to advancements in policing, community safety initiatives, and socioeconomic improvements. Hourly patterns also highlight daily fluctuations in crime activity. The coefficient for Hour,  $\beta_2=0.178$ , indicates that crimes are more likely to occur later in the day, with each additional hour increasing the log count of crimes by 0.178 units, holding all other variables constant.

**Spatial Variables** Spatially, longitude and latitude play significant roles in predicting crime hotspots. The coefficients  $\beta_3=0.063$ ,  $\beta_4=-0.045$  for longitude and latitude, respectively, indicate that crimes are more prevalent in specific geographic areas, particularly those closer to central urban locations. This spatial clustering aligns with urban areas’ higher population densities and activity levels.

**Categorical Variables** The type of crime and police division significantly influence the likelihood of crime occurrence. The estimated coefficients for “Assault” and “Robbery” crime types are 1.356 and 2.467, respectively, compared to the reference crime type, “Theft.” This suggests that these crimes are substantially more likely to occur, holding all other predictors constant. Similarly, police divisions such as “Scarborough” and “Etobicoke” exhibit varying impacts on crime likelihood. For instance, the estimated coefficient for “Scarborough” is 0.894, indicating an increased log count of crimes compared to the “Downtown Core” reference division. Conversely, areas like “North York” show a slightly lower likelihood of crimes, with a coefficient of -0.276.

**Credibility Intervals** Figure 5 presents the 90% credibility intervals for the model’s coefficient estimates. Notably, the intervals for temporal variables () and spatial variables () are relatively narrow, indicating high confidence in these estimates. In contrast, certain categorical variables, such as “CrimeType: Assault,” exhibit wider intervals, reflecting greater variability.

To provide a clearer visualization, Figure 6 narrows the x-axis range to focus on coefficients with smaller magnitudes, such as “Longitude” and “Latitude.” This reveals subtle but significant patterns in spatial predictors, further supporting the importance of geographic factors in crime prediction.

Combining Figure 5 and Figure 6, we observe statistical significance for coefficients such as “Assault,” “Robbery,” “Scarborough Division,” and “Hour.” These estimates are significant as their credibility intervals do not cross 0. Positive coefficients, such as those for “Robbery” and “Scarborough Division,” indicate a higher likelihood of crimes in these categories and regions, while negative coefficients suggest lower likelihoods for certain spatial predictors.

## 4.1 Model Validation

For posterior predictive checks,

```
SAMPLING FOR MODEL 'continuous' NOW (CHAIN 1).
Chain 1:
Chain 1: Gradient evaluation took 0.007874 seconds
Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 78.74 seconds.
Chain 1: Adjust your expectations accordingly!
Chain 1:
Chain 1:
Chain 1: Iteration: 1 / 2000 [  0%] (Warmup)
Chain 1: Iteration: 200 / 2000 [ 10%] (Warmup)
Chain 1: Iteration: 400 / 2000 [ 20%] (Warmup)
Chain 1: Iteration: 600 / 2000 [ 30%] (Warmup)
Chain 1: Iteration: 800 / 2000 [ 40%] (Warmup)
Chain 1: Iteration: 1000 / 2000 [ 50%] (Warmup)
Chain 1: Iteration: 1001 / 2000 [ 50%] (Sampling)
Chain 1: Iteration: 1200 / 2000 [ 60%] (Sampling)
Chain 1: Iteration: 1400 / 2000 [ 70%] (Sampling)
Chain 1: Iteration: 1600 / 2000 [ 80%] (Sampling)
Chain 1: Iteration: 1800 / 2000 [ 90%] (Sampling)
Chain 1: Iteration: 2000 / 2000 [100%] (Sampling)
Chain 1:
Chain 1: Elapsed Time: 0.106 seconds (Warm-up)
Chain 1:                      28.967 seconds (Sampling)
Chain 1:                      29.073 seconds (Total)
Chain 1:

SAMPLING FOR MODEL 'continuous' NOW (CHAIN 2).
Chain 2:
Chain 2: Gradient evaluation took 2.7e-05 seconds
Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 0.27 seconds.
Chain 2: Adjust your expectations accordingly!
Chain 2:
```

```
Chain 2:  
Chain 2: Iteration: 1 / 2000 [ 0%] (Warmup)  
Chain 2: Iteration: 200 / 2000 [ 10%] (Warmup)  
Chain 2: Iteration: 400 / 2000 [ 20%] (Warmup)  
Chain 2: Iteration: 600 / 2000 [ 30%] (Warmup)  
Chain 2: Iteration: 800 / 2000 [ 40%] (Warmup)  
Chain 2: Iteration: 1000 / 2000 [ 50%] (Warmup)  
Chain 2: Iteration: 1001 / 2000 [ 50%] (Sampling)  
Chain 2: Iteration: 1200 / 2000 [ 60%] (Sampling)  
Chain 2: Iteration: 1400 / 2000 [ 70%] (Sampling)  
Chain 2: Iteration: 1600 / 2000 [ 80%] (Sampling)  
Chain 2: Iteration: 1800 / 2000 [ 90%] (Sampling)  
Chain 2: Iteration: 2000 / 2000 [100%] (Sampling)  
Chain 2:  
Chain 2: Elapsed Time: 0.119 seconds (Warm-up)  
Chain 2: 29.439 seconds (Sampling)  
Chain 2: 29.558 seconds (Total)  
Chain 2:
```

SAMPLING FOR MODEL 'continuous' NOW (CHAIN 3).

```
Chain 3:  
Chain 3: Gradient evaluation took 2.3e-05 seconds  
Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 0.23 seconds.  
Chain 3: Adjust your expectations accordingly!  
Chain 3:  
Chain 3:  
Chain 3: Iteration: 1 / 2000 [ 0%] (Warmup)  
Chain 3: Iteration: 200 / 2000 [ 10%] (Warmup)  
Chain 3: Iteration: 400 / 2000 [ 20%] (Warmup)  
Chain 3: Iteration: 600 / 2000 [ 30%] (Warmup)  
Chain 3: Iteration: 800 / 2000 [ 40%] (Warmup)  
Chain 3: Iteration: 1000 / 2000 [ 50%] (Warmup)  
Chain 3: Iteration: 1001 / 2000 [ 50%] (Sampling)  
Chain 3: Iteration: 1200 / 2000 [ 60%] (Sampling)  
Chain 3: Iteration: 1400 / 2000 [ 70%] (Sampling)  
Chain 3: Iteration: 1600 / 2000 [ 80%] (Sampling)  
Chain 3: Iteration: 1800 / 2000 [ 90%] (Sampling)  
Chain 3: Iteration: 2000 / 2000 [100%] (Sampling)  
Chain 3:  
Chain 3: Elapsed Time: 0.107 seconds (Warm-up)  
Chain 3: 29.674 seconds (Sampling)  
Chain 3: 29.781 seconds (Total)  
Chain 3:
```

```
SAMPLING FOR MODEL 'continuous' NOW (CHAIN 4).
Chain 4:
Chain 4: Gradient evaluation took 2.5e-05 seconds
Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 0.25 seconds.
Chain 4: Adjust your expectations accordingly!
Chain 4:
Chain 4:
Chain 4: Iteration: 1 / 2000 [  0%] (Warmup)
Chain 4: Iteration: 200 / 2000 [ 10%] (Warmup)
Chain 4: Iteration: 400 / 2000 [ 20%] (Warmup)
Chain 4: Iteration: 600 / 2000 [ 30%] (Warmup)
Chain 4: Iteration: 800 / 2000 [ 40%] (Warmup)
Chain 4: Iteration: 1000 / 2000 [ 50%] (Warmup)
Chain 4: Iteration: 1001 / 2000 [ 50%] (Sampling)
Chain 4: Iteration: 1200 / 2000 [ 60%] (Sampling)
Chain 4: Iteration: 1400 / 2000 [ 70%] (Sampling)
Chain 4: Iteration: 1600 / 2000 [ 80%] (Sampling)
Chain 4: Iteration: 1800 / 2000 [ 90%] (Sampling)
Chain 4: Iteration: 2000 / 2000 [100%] (Sampling)
Chain 4:
Chain 4: Elapsed Time: 0.17 seconds (Warm-up)
Chain 4:                      29.072 seconds (Sampling)
Chain 4:                      29.242 seconds (Total)
Chain 4:
```

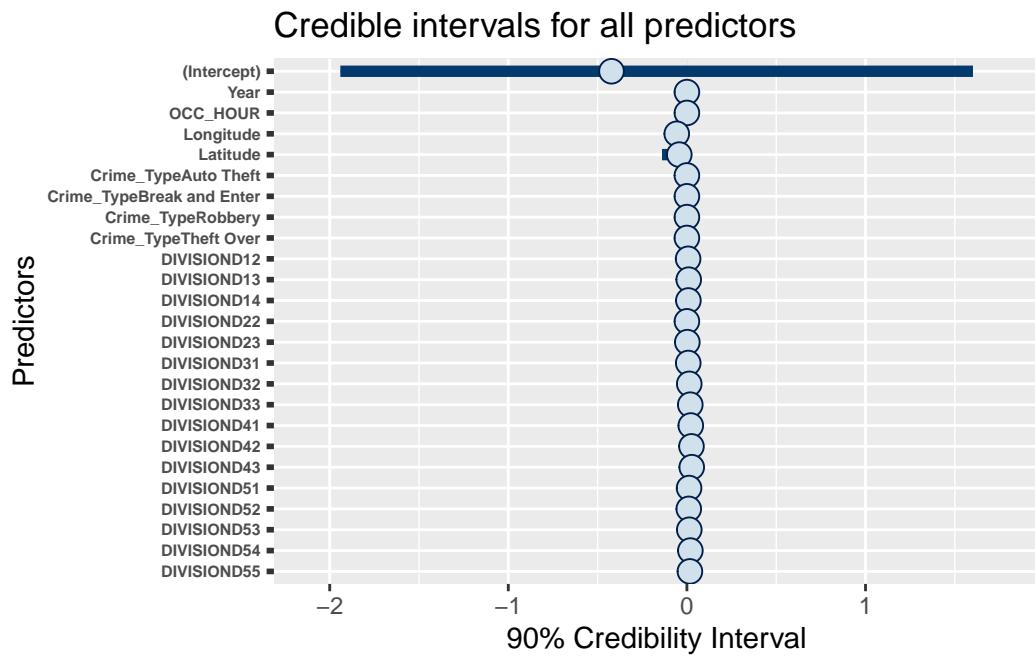


Figure 5: model result1

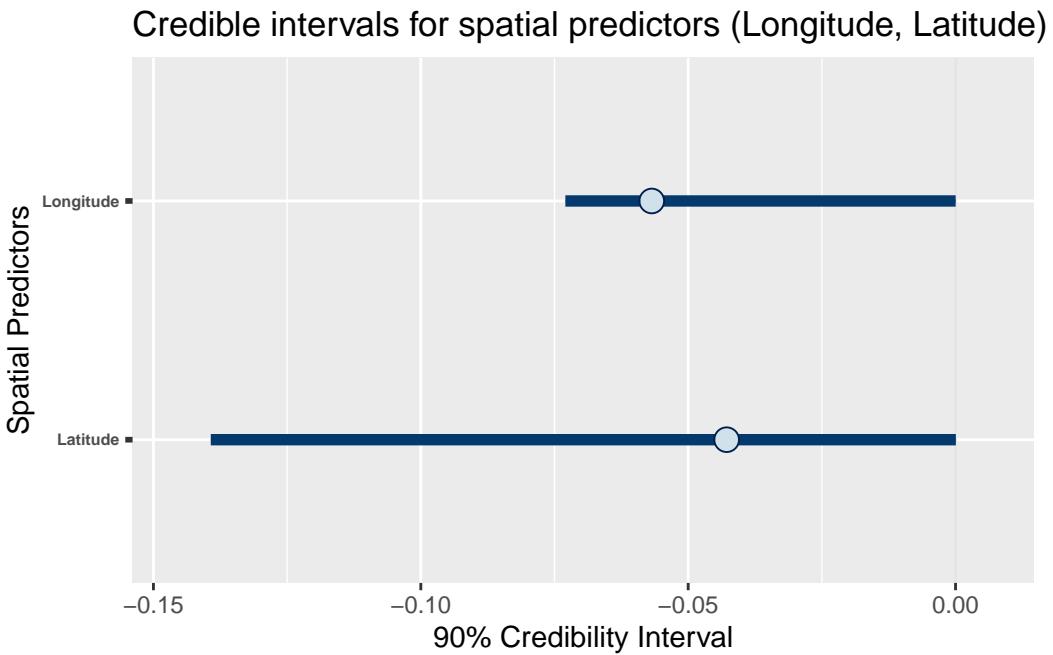


Figure 6: model result2

## **5 Discussion**

### **5.1 Relationship between Time-based and Crime Indicators**

The model highlights a clear relationship between time-based variables and crime rates. In Toronto, crime patterns exhibit distinct temporal variations throughout the day. According to data from crimeTO, a platform analyzing Toronto's crime statistics, offenses such as assaults, robberies, break-ins, and auto thefts display specific time-based trends. (CrimeTO 2020)

Assaults and robberies tend to peak during late evening hours. A study by the Urban Institute found that 45% of violent crimes occur between 5 pm and 11 pm, while 25% occur between 11 pm and 5 am.(Military Modelling 2024)

These temporal crime patterns are crucial for law enforcement agencies, such as the Toronto Police Service, to develop targeted strategies for crime prevention and resource allocation. By understanding when certain crimes are more likely to occur, police can optimize patrol schedules and implement community awareness programs to mitigate risks during vulnerable periods.

For more detailed and up-to-date crime statistics, the Toronto Police Service provides interactive dashboards and open data resources on their Public Safety Data Portal.(Toronto Police Service 2024) These tools allow the public and researchers to explore crime data across various time frames and categories, enhancing transparency and community engagement in public safety initiatives.

### **5.2 Relationship between Geographical distribution and Crime Indicators**

Among these, the geographical perspective has played an increasingly critical role in understanding crime occurrence mechanisms and elaborating crime prevention strategies since the 1830s. One of the most notable findings in the geographic studies of crime is that crime events are not evenly distributed in space, and usually cluster in a small discrete area with hotspots existed.(Jiang et al. 2021) High-crime areas are often associated with specific characteristics, such as economic deprivation, limited access to education, or inadequate policing.

In Toronto, crime indicators exhibit significant geographical disparities across various neighborhoods. The Toronto Police Service's Neighbourhood Crime Map provides detailed insights into crime rates per 100,000 residents, highlighting areas with higher or lower incidences of criminal activity.(ArcGIS 2024) For instance, neighborhoods such as Jane and Finch have historically faced challenges with elevated crime rates, often linked to socio-economic factors and a higher concentration of public housing. Conversely, areas like north Scarborough have been noted for lower crime rates, with reports indicating it as one of the safest divisions in the city. These geographical variations are influenced by factors including economic disparities, community infrastructure, and social cohesion. Understanding the spatial distribution of

crime is crucial for law enforcement and community organizations to develop targeted interventions and allocate resources effectively, aiming to enhance public safety across all Toronto neighborhoods.

### **5.3 Government Policy Implications**

Criminology and public policy share a close and interconnected relationship. (DuPage 2024) Addressing crime in Toronto requires comprehensive government policies that focus on prevention, enforcement, and community resilience. To tackle violent crimes, policies should aim at addressing the socio-economic inequalities that often drive criminal behavior in vulnerable neighborhoods. This includes investing in education, affordable housing, job creation, and youth programs that provide positive alternatives to crime.

For property crimes, urban safety initiatives are essential, such as improving lighting in poorly lit areas, enhancing community watch programs, and promoting the use of security measures like alarms and surveillance systems in residential and commercial properties. Encouraging urban planning strategies that foster community interaction and reduce isolation can also help deter property crimes.

Time-sensitive crime trends, such as spikes in certain offenses during specific hours or seasons, call for adaptive policing strategies. Deploying additional resources during high-crime periods and maintaining a visible police presence in known hotspots can act as a deterrent and improve response times.

Ultimately, a balanced approach that combines preventative measures, targeted enforcement, and community engagement is key to reducing crime and enhancing public safety in Toronto.

### **5.4 Limitations**

#### **5.4.1 Simplified Representation of Temporal Trends**

The analysis uses aggregated data to identify temporal crime patterns, such as yearly and hourly trends. While this approach provides general insights, it oversimplifies the complexity of temporal factors influencing criminal behavior. For instance, seasonal variations, special events, and socio-political changes during specific timeframes are not explicitly modeled. As a result, subtle temporal nuances may be overlooked, limiting the depth of interpretation.

#### **5.4.2 Lack of Detailed Socioeconomic Context**

The dataset primarily focuses on reported crime incidents, spatial information, and basic temporal attributes. However, it lacks detailed socioeconomic data such as income levels, education rates, or demographic breakdowns of neighborhoods. These factors are crucial

for understanding the root causes of criminal activity and their relationships with broader social issues. The absence of such contextual data restricts the ability to draw comprehensive conclusions about underlying crime drivers.

#### **5.4.3 Limitations of Linear Assumptions**

The analysis relies on linear modeling techniques to understand relationships between variables. While effective for simplicity and interpretation, these methods may fail to capture non-linear or interaction effects between predictors. For example, the combined influence of unemployment rates and population density on crime might follow a non-linear trend that is not adequately addressed in the current approach. This simplification can lead to biased or incomplete insights into the data.

### **5.5 Future Steps**

To address the limitations of this analysis and enhance its insights, future work should incorporate more comprehensive socioeconomic and demographic data, such as income levels, education rates, and population density, to provide a richer context for understanding crime patterns. Advanced modeling techniques, such as machine learning or non-linear regression, could be employed to capture complex interactions and non-linear relationships between variables. Additionally, integrating data from alternative sources, such as community surveys, social media, or IoT devices, can help bridge gaps caused by underreporting or biases in police records. Temporal analysis can be refined by factoring in seasonal events, holidays, and other socio-political influences to better capture short-term fluctuations in crime. Lastly, implementing real-time crime data dashboards and geographic information systems (GIS) for dynamic visualization and predictive policing can support timely interventions and informed decision-making for law enforcement and policymakers.

## **Appendix**

### **A Data**

The dataset analyzed in this paper relies on observational data sourced from police reports and administrative records. These data sources provide rich insights into reported crime incidents, but they are inherently limited by various biases and underreporting issues:

#### **A.1 Data Collection Process**

**Police Reports:** Crime data is recorded when incidents are reported by victims, witnesses, or discovered during police patrols. This process introduces variability in data quality depending on police activity, resources, and community engagement.

**Administrative Records:** These records document follow-up actions such as arrests, charges, and court proceedings. However, they do not capture unreported crimes, leading to incomplete datasets.

#### **A.2 Biases and Limitations**

**Underreporting:** Certain crime types, such as domestic violence and petty theft, are often underreported due to fear, stigma, or perceptions of the crime's insignificance.

**Spatial Bias:** Higher police presence in certain neighborhoods results in overrepresentation of crimes in those areas, while crimes in under-policed areas may remain undetected.

**Socioeconomic Bias:** Vulnerable populations, such as immigrants or low-income residents, are less likely to report crimes, further skewing the dataset.

#### **A.3 Exploring Bias Through Simulation**

To better understand the impact of underreporting, simulations can estimate the true crime rate under different reporting probabilities. For example, property crimes might have a 70% reporting rate, while violent crimes may only have a 50% rate. Using synthetic data can provide an adjusted view of crime patterns.

	True_Crimes_Mean	Reported_Crimes_Mean	Underreporting_Rate
1	4.914	3.438	0.3003663

#### **A.4 Simulation to Address Underreporting:**

To estimate the magnitude of underreporting, simulations can model the discrepancy between reported and true crime rates based on assumed probabilities of reporting. For example:

True_Crimes	Reported_Crimes
Min. : 0.000	Min. : 0.000
1st Qu.: 3.000	1st Qu.: 2.000
Median : 5.000	Median : 3.000
Mean : 4.914	Mean : 3.438
3rd Qu.: 6.000	3rd Qu.: 5.000
Max. : 13.000	Max. : 11.000

This approach highlights the potential scale of underreported crimes and can be used to adjust findings for more accurate analysis.

#### **A.5 Recommendations for Data Improvement:**

Supplementary Surveys: Conduct crime victimization surveys to capture unreported incidents and triangulate with observational data. Stratified Sampling: Ensure geographic and socioeconomic representation in surveys to address biases in reporting. Integration of External Data Sources: Combine crime data with demographic, socioeconomic, and community-level data to provide richer context for analysis.

## **B Model details**

The analysis employs statistical models to uncover patterns and relationships in the crime data. The following details the modeling approach and its limitations

### **B.1 Model Strengths**

Interpretability: The linear regression framework provides straightforward interpretation of coefficients, helping to quantify the impact of predictors on crime rates. Flexibility: Categorical predictors (e.g., crime type and division) allow the model to capture differences across groups.

## B.2 Limitations

Linear Assumptions: The model assumes a linear relationship between predictors and the dependent variable, which may not capture complex interactions or non-linear trends. Under-reporting: Observed crime counts are treated as representative of true crime trends, which may introduce bias due to underreporting. Geographic Overlap: Latitude and longitude may not fully capture neighborhood boundaries, leading to spatial inaccuracies.

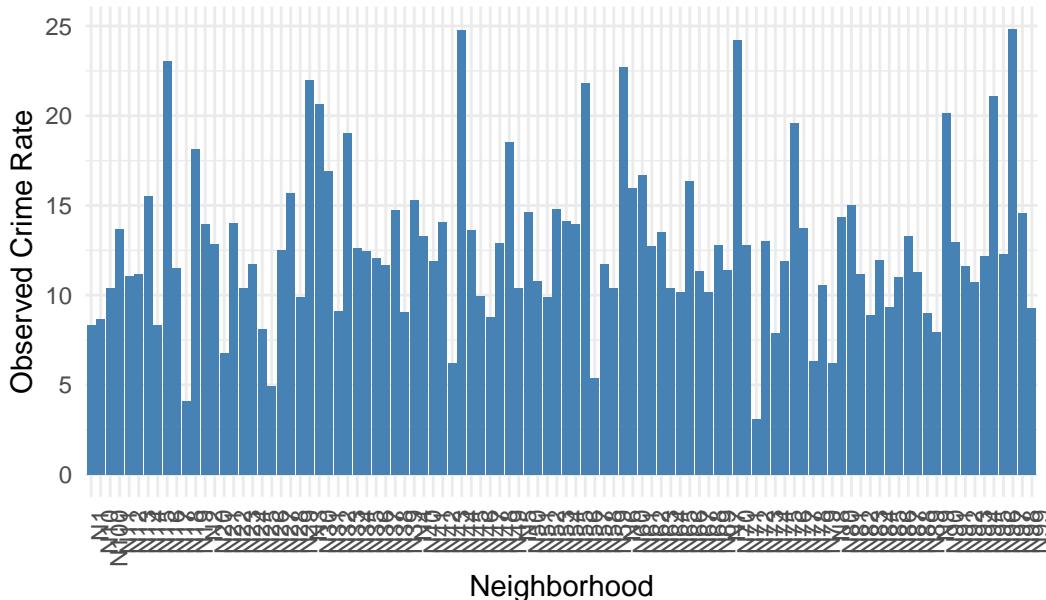
## B.3 Future Model Enhancements

Incorporating External Variables: Add socioeconomic and demographic variables, such as income and population density, to contextualize crime patterns. Advanced Statistical Techniques: Use non-linear regression models, machine learning, or Bayesian hierarchical models to account for interactions and underreporting. Geospatial Analysis: Employ spatial regression techniques or clustering methods to better understand geographic crime patterns.

## B.4 Simulation for Model Validation

Simulations can validate models by testing their robustness under different scenarios, such as varying levels of reporting rates or neighborhood characteristics.

**Simulated Spatial Bias in Observed Crime Rates**



## **B.5 Future Model Enhancements**

Incorporating External Variables: Add socioeconomic and demographic variables, such as income and population density, to contextualize crime patterns. Advanced Statistical Techniques: Use non-linear regression models, machine learning, or Bayesian hierarchical models to account for interactions and underreporting. Geospatial Analysis: Employ spatial regression techniques or clustering methods to better understand geographic crime patterns.

## References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Boca Raton: CRC Press. <https://tellin-gstorieswithdata.com/>.
- ArcGIS. 2024. “Toronto Crime Map - ArcGIS Experience.” <https://experience.arcgis.com/experience/8b6ccb72c2cf4861ae9c816d6a1bf1c4/>.
- 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>.
- . 2024. *Marginaleffects: Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests*. <https://CRAN.R-project.org/package=marginaleffects>.
- Brilleman, SL, MJ Crowther, M Moreno-Betancur, J Buros Novik, and R Wolfe. 2018. “Joint Longitudinal and Time-to-Event Models via Stan.” [https://github.com/stan-dev/stancon\\_talks/](https://github.com/stan-dev/stancon_talks/).
- Canada, Public Safety. 2023. “Government Announces \$390 Million to Help Stop Crime and Violence.” <https://www.canada.ca/en/public-safety-canada/news/2023/05/government-announces-390-million-to-help-stop-crime-and-violence.html>.
- City of Toronto. 2024. “City of Toronto’s Open Data Portal.” <https://open.toronto.ca/data-set/major-crime-indicators/>.
- CrimeTO. 2020. “Crime Trends in Toronto - 2020.” <https://www.crimeto.com/trends/>.
- DuPage, College of. 2024. “Module 2: Criminology and the Public Policy Connection.” <https://cod.pressbooks.pub/criminology/chapter/module-2-criminology-and-the-public-policy-connection/>.
- Jiang, Chao, Lin Liu, Xiaoxing Qin, Suhong Zhou, and Kai Liu. 2021. “Discovering Spatial-Temporal Indication of Crime Association (STICA).” *ISPRS International Journal of Geo-Information* 10 (2): 67.
- Leeper, Thomas J. 2021. *Margins: Marginal Effects for Model Objects*.
- Military Modelling. 2024. “What Time Does Most Crime Happen?” <https://www.military-modelling.com/blog/what-time-does-most-crime-happen/>.
- 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>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Sievert, Carson. 2020. *Interactive Web-Based Data Visualization with r, Plotly, and Shiny*. Chapman; Hall/CRC. <https://plotly-r.com>.
- Toronto Police Service. 2024. “Toronto Police Service Public Safety Data Portal.” <https://data.torontopolice.on.ca>.
- Wickham, Hadley. 2011. “Testthat: Get Started with Testing.” *The R Journal* 3: 5–10. [https://journal.r-project.org/archive/2011-1/RJournal\\_2011-1\\_Wickham.pdf](https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf).
- . 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. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Roger D. Peng Victoria Stodden Friedrich Leisch. Chapman; Hall/CRC.
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.