

# Examining the Influence of Premises Type and Time of Day on Violent Crime in Toronto\*

## A Bayesian Approach to Analyzing Contextual and Temporal Factors in Urban Crime Dynamics

Tim Chen

December 3, 2024

This paper analyzes the contextual and temporal factors influencing violent crimes in Toronto using data from the Major Crime Indicators dataset. A Bayesian logistic regression model explains that violent crimes are more likely to occur in transit and outdoor locations compared to residential and educational premises. Crimes are also more likely to be violent during afternoon and evening hours than early mornings. These findings support urban safety planning and targeted crime prevention strategies.

### Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Overview . . . . .	1
1.2	Estimand . . . . .	2
1.3	Results Summary . . . . .	2
1.4	Why This Study Matters . . . . .	2
1.5	Paper Structure . . . . .	3
<b>2</b>	<b>Data</b>	<b>3</b>
2.1	Overview . . . . .	3
2.2	Measurement . . . . .	3
2.3	Data Cleaning . . . . .	4
2.4	Outcome variable . . . . .	5
2.5	Predictor variables . . . . .	6

---

\*Code and data are available at: <https://github.com/timchen0326/crime-analysis-toronto>.

<b>3</b>	<b>Model</b>	<b>8</b>
3.1	Model Set-up . . . . .	8
3.2	Model Justification . . . . .	8
<b>4</b>	<b>Results</b>	<b>9</b>
<b>5</b>	<b>Discussion</b>	<b>11</b>
5.1	Relationship between Premises Type and Violent Crime . . . . .	11
5.2	Relationship between Time of Day and Violent Crime . . . . .	11
5.3	Expanding Implications for Policy and Prevention . . . . .	12
5.4	Limitations . . . . .	12
5.5	Expanding Future Directions . . . . .	13
5.6	Conclusion . . . . .	14
<b>A</b>	<b>Appendix: Survey Design to Investigate Factors Influencing Violent Crimes in Toronto</b>	<b>15</b>
A.1	Introduction . . . . .	15
A.2	Objectives . . . . .	15
A.3	Methodology . . . . .	15
A.4	Budget Allocation . . . . .	16
A.5	Survey Instrument . . . . .	17
A.6	Ethical Considerations . . . . .	22
A.7	Data Collection Procedures . . . . .	22
A.8	Data Management and Analysis . . . . .	23
A.9	Quality Assurance . . . . .	24
A.10	Dissemination of Results . . . . .	24
A.11	Conclusion . . . . .	25
<b>B</b>	<b>Model details</b>	<b>26</b>
B.1	Posterior Predictive Check and Prior-Posterior Comparison . . . . .	26
B.2	Diagnostics . . . . .	26
B.3	90% Credibility Interval . . . . .	27
	<b>References</b>	<b>32</b>

# 1 Introduction

## 1.1 Overview

This study examines the factors influencing the likelihood of violent crimes in Toronto, focusing on two important predictors: the type of premises where the crime occurred and the time of day. Violent crime, encompassing offenses such as assault and robbery, poses significant risks

to community safety and requires evidence-based strategies for prevention. Non-violent crimes, such as theft and property-related offenses, provide a contrasting baseline for analysis.

Using a detailed dataset from the Toronto Police Service, this research aims to uncover patterns and relationships between the environment, timing, and the probability of violent crimes. By leveraging Bayesian logistic regression modeling, the study accounts for uncertainty while using prior knowledge to generate practical findings. The findings aim to inform crime prevention efforts, resource allocation, and urban safety policies by identifying high-risk contexts for violence.

## **1.2 Estimand**

The primary estimand of this study is the probability of a violent crime occurring given specific premises (e.g., residential, commercial, or transit) and the time of day (e.g., early morning, evening). Violent crimes include serious offenses such as assaults and robberies, while non-violent crimes encompass thefts and similar offenses. This research employs a Bayesian logistic regression model to estimate these probabilities, accounting for uncertainty and using prior knowledge.

## **1.3 Results Summary**

The results of this study provide clarity on the relationship between premises type, time of day, and the likelihood of a crime being violent. Premises such as “Transit” and “Outside” exhibit greater odds of violent crimes compared to the reference category (“Apartment”), reflecting their open and public nature, which may increase opportunities for conflicts. Conversely, premises like “Commercial,” “Educational,” and “House” are associated with a lower likelihood of violent crimes, underscoring the role of structural and social controls in reducing violence.

Time of day also plays a significant role, with “Afternoon” and “Evening” periods showing elevated odds of violent crimes compared to the baseline “Early Morning.” These findings align with patterns of increased social interactions and activities during these periods. The credible interval plots further confirm the statistical significance of these predictors, emphasizing their impact on crime likelihood. Overall, the results highlight patterns that can guide focused crime prevention strategies in high-risk locations and time periods.

## **1.4 Why This Study Matters**

Understanding the contextual factors contributing to violent crimes can help city planners and policymakers allocate safety interventions effectively. This research focuses on premises and times of day where the likelihood of violence is highest, enabling tailored prevention strategies. The findings from this study can directly impact resource allocation for policing and urban planning, potentially reducing crime rates and enhancing public safety.

## 1.5 Paper Structure

The remainder of this paper is structured as follows. In Section 2, we present an overview of the dataset used in this study, along with a detailed description of the variables and the data cleaning process. Section 3 outlines the Bayesian logistic regression model applied in our analysis, including the model setup, assumptions, and justification. In Section 4, we summarize the results of our analysis and interpret the findings within the context of existing literature on road safety. Finally, Section 5 discusses the implications of these findings, acknowledges the limitations of the study, and suggests directions for future research.

## 2 Data

### 2.1 Overview

We use the statistical programming language R (R Core Team 2023) to process, clean, and analyze the dataset for this study. Key R packages employed include `tidyverse` for data manipulation and visualization (Wickham et al. 2019), `arrow` for efficient data storage and retrieval (Developers 2024), `here` for managing file paths (Müller 2023), `rstanarm` for Bayesian modeling (Goodrich et al. 2022), `kableExtra` for enhanced table formatting (Zhu 2023), `janitor` for cleaning messy data (Firke 2023), `lubridate` for working with dates and times (Grolemund and Wickham 2023), `marginalEffects` for calculating marginal effects (Arel-Bundock 2024), `plumber` for creating REST APIs (Allen et al. 2023), `testthat` for unit testing (Wickham and Hester 2023), and `opendatatoronto` for accessing Toronto’s open data (Open Data Toronto 2021).

Our data is sourced from the Open Data Toronto portal (Open Data Toronto 2021), specifically from the “Major Crime Indicators” dataset, which records reported crimes across Toronto. Compiled by the Toronto Police Service, the dataset includes details such as crime categories, locations, and times, enabling an analysis of crime trends and patterns.

By examining factors like crime type, neighborhood, and temporal distribution, this study aims to identify areas and times of higher crime occurrence. The findings can support targeted interventions and resource allocation. Following principles outlined in *Telling Stories with Data* (Alexander 2023), this study prioritizes clarity and reproducibility to inform community safety strategies effectively.

### 2.2 Measurement

The dataset used in this study is sourced from the Toronto Police Service and includes records of Major Crime Indicators (MCI) occurrences. These records capture information on five important categories of crime: Assault, Break and Enter, Auto Theft, Robbery, and Theft

Over. Unlike some crime datasets, this dataset excludes occurrences classified as “unfounded,” meaning that only verified incidents are included. The definition of “unfounded,” provided by Statistics Canada, is: “It has been determined through police investigation that the offence reported did not occur, nor was it attempted.”

To align with the study’s focus on temporal and environmental factors influencing violent crime, we selected crimes reported during the month of July. This decision was informed by research highlighting July as a high-risk period for violent crimes due to its status as one of the warmest months globally and its alignment with peak social activity (World Meteorological Organization 2024; Pinkerton Insights 2021). Higher temperatures during July have been linked to increased aggression and lowered inhibition, a phenomenon described by the “Temperature/Aggression” theory (Pinkerton Insights 2021). Moreover, the Routine Activities theory explains that greater public activity in July provides more opportunities for interaction between offenders and victims, potentially elevating crime rates (Pinkerton Insights 2021).

The data collection process primarily relies on police-reported incidents. However, there are certain limitations in the completeness of the dataset. For example, some location coordinates are missing or left blank when the location could not be verified. Additionally, the dataset includes occurrences that are located outside the City of Toronto, which may slightly affect the analysis for Toronto-specific crime trends.

Spatial data in the dataset is provided using both the old 140-neighborhood structure and the newer 158-neighborhood structure for the City of Toronto. This dual structure ensures compatibility across different time periods and administrative updates. Temporal data includes detailed timestamps that allow crimes to be categorized by time of day and season.

The unit of measurement for numerical data, such as timestamps and neighborhood IDs, is standard counts or identifiers with no inherent units. Other features, such as crime type, premises type, and time of day, are categorical variables with no associated units.

Although the data provides a robust framework for understanding crime patterns in Toronto, some aspects are modeled or incomplete. For instance, location data for certain crimes may be estimated, and the dataset does not include crimes that were not reported to the police. Despite these limitations, the dataset remains a useful resource for analyzing environmental and temporal factors influencing violent and non-violent crimes in Toronto.

## 2.3 Data Cleaning

The raw crime data underwent a cleaning process to prepare it for analysis, ensuring the dataset was both relevant and consistent for the study’s objectives. Initially, only key columns were selected from the raw dataset, including `EVENT_UNIQUE_ID` (Unique Identifier), `REPORT_DATE` (Date of Report), `OCC_HOUR` (Hour of Occurrence), `PREMISES_TYPE` (Type of Premises), and `OFFENCE` (Offense Type). This selection ensured that only the most pertinent information was included for examining factors associated with violent crimes.

Rows with missing or irrelevant data in important columns, such as `EVENT_UNIQUE_ID`, `PREMISES_TYPE`, `TIME_OF_DAY`, and `VIOLENT_CRIME`, were removed to maintain data integrity and focus on valid observations. Offenses were categorized as either violent or non-violent to create a binary outcome variable (`VIOLENT_CRIME`), where offenses such as “Assault” and “Robbery” were classified as violent (1), and crimes like “Theft” or “Break and Enter” were classified as non-violent (0). Offenses that did not fit into either category were excluded from the dataset.

To analyze temporal patterns, the `OCC_HOUR` variable was grouped into broader time categories:

- **Early Morning:** 12 AM to 6 AM
- **Morning:** 6 AM to 12 PM
- **Afternoon:** 12 PM to 6 PM
- **Evening:** 6 PM to 12 AM

Additionally, the `PREMISES_TYPE` variable, which describes the type of location where the incident occurred, was trimmed to remove excess whitespace and ensure consistent formatting. This step was to prevent unexpected NA values during analysis.

The cleaning process also included filtering for crimes reported during the summer month of July, aligning the dataset with the study’s focus on temporal and environmental patterns in criminal activity. After grouping, formatting, and filtering the data, all rows containing NA values in important columns were dropped to ensure a clean and analyzable dataset.

Finally, the cleaned dataset was saved as a Parquet file for further analysis. This format ensures efficient storage and retrieval while preserving the structure of the data.

## 2.4 Outcome variable

The primary outcome variable in this study is Violent Crime, which categorizes each incident based on whether the crime is classified as violent or non-violent. Violent crime is coded as a binary variable, where “0” represents non-violent offenses (e.g., theft, break and enter), and “1” denotes violent offenses (e.g., assault, robbery). This classification allows for an assessment of factors contributing to the likelihood of a crime being violent, providing a framework for understanding the underlying conditions associated with violent behavior.

By focusing on crimes involving violence, this outcome variable highlights important cases of interest for public safety and crime prevention efforts, as violent offenses often have severe consequences for victims and communities. This binary categorization simplifies the statistical

analysis, enabling the application of logistic regression models to estimate the probability of violent crimes under varying conditions.

Analyzing violent crime in conjunction with predictor variables such as premises type and time of day helps identify the circumstances under which the risk of violence is heightened. This outcome measure serves as a indicator for assessing and improving community safety policies, as it reflects the most serious consequences of criminal activity and directs attention to factors that could reduce violent incidents.

Table 1: Preview of the cleaned Crime dataset

Violent Crime	Premises Type	Time of Day
1	House	Evening
1	House	Evening
1	House	Evening
1	Apartment	Evening
1	Outside	Early Morning

## 2.5 Predictor variables

Figure 1 illustrates the distribution of violent versus non-violent crimes across different premises types. The results show that the highest number of crimes, both violent and non-violent, occur in “Outside” locations, followed by “Apartment” premises. Notably, violent crimes significantly outnumber non-violent ones in “Outside” locations, while “Apartments” and “Houses” have a more balanced distribution between the two crime types. Premises categorized as “Educational,” “Transit,” and “Other” show relatively low numbers of both violent and non-violent crimes, with “Educational” premises having the fewest incidents overall.

Figure 2 highlights the distribution of violent versus non-violent crimes by the time of day. The data explains that violent crimes are most common during the “Evening,” followed by “Afternoon” and “Early Morning.” In contrast, non-violent crimes are more evenly distributed throughout the day but still peak during the “Evening.” These findings suggest that the time of day plays a important role in influencing the likelihood of violent crimes, with evenings being a particularly high-risk period.

By examining these distributions, we gain intuitions into where and when violent crimes are most likely to occur. Understanding these patterns is essential for developing focused crime prevention strategies, such as increasing law enforcement presence in high-risk areas during evening hours or implementing community-based interventions in “Outside” locations.

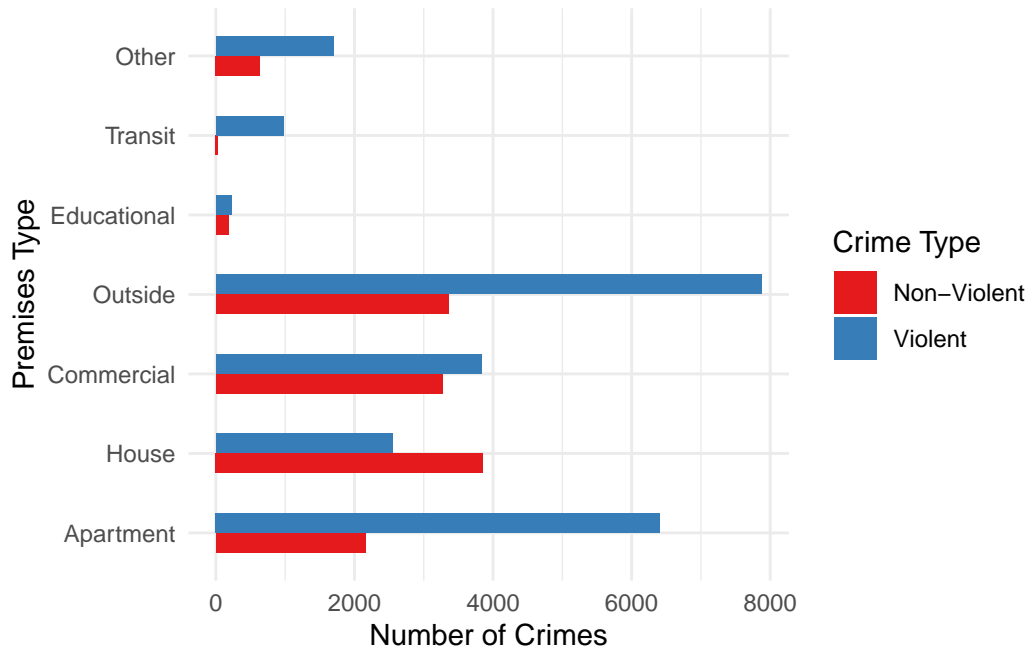


Figure 1: The distribution of violent vs. non-violent crimes by premises type

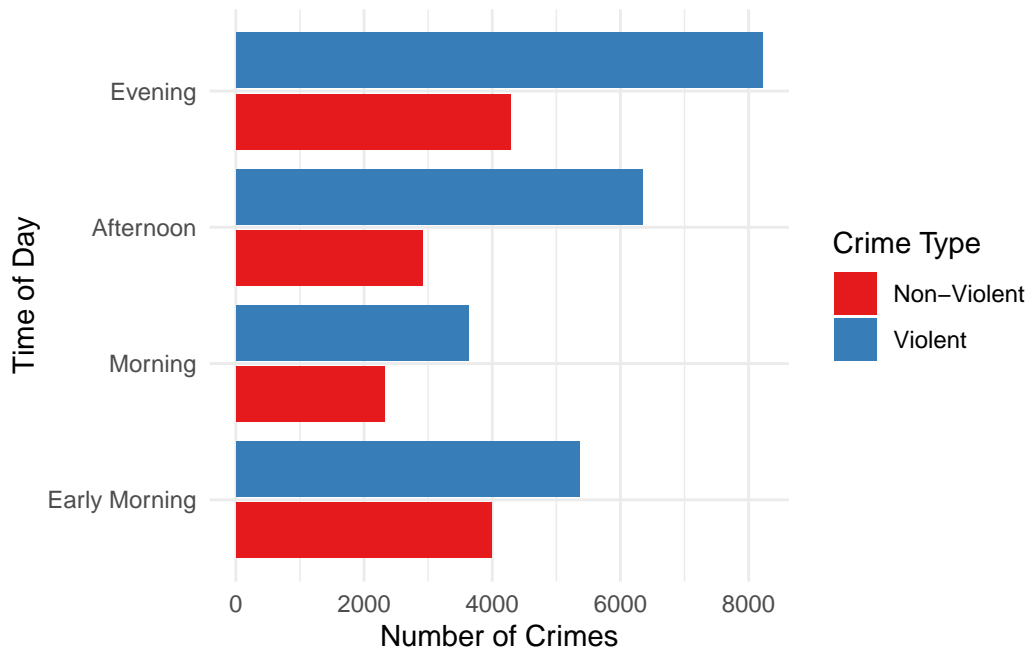


Figure 2: The distribution of violent vs. non-violent crimes by time of day



### 3 Model

In our analysis, we utilized a Bayesian logistic regression model to examine the relationship between violent crime occurrence and two important predictors: premises type and time of day. By using this model, we aimed to understand how the location and timing of incidents influence the likelihood of a crime being classified as violent. Detailed model diagnostics and background are provided in Appendix B.

#### 3.1 Model Set-up

$$y_i | \pi_i \sim \text{Bern}(\pi_i) \tag{1}$$

$$\text{logit}(\pi_i) = \alpha + \beta_1 \times \text{PREMISES\_TYPE}_i + \beta_2 \times \text{TIME\_OF\_DAY}_i \tag{2}$$

$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$

$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{4}$$

$$\beta_2 \sim \text{Normal}(0, 2.5) \tag{5}$$

In this model:

- $\beta_1$  is the coefficient corresponding to the type of premises ( $\text{PREMISES\_TYPE}_i$ ) where the crime occurred.
- $\beta_2$  is the coefficient corresponding to the time of day ( $\text{TIME\_OF\_DAY}_i$ ) when the crime occurred.

The logistic link function ( $\text{logit}(\pi_i)$ ) models the log-odds of violent crime as a linear combination of the predictors. The priors for the intercept ( $\alpha$ ) and the coefficients ( $\beta_1, \beta_2$ ) follow a Normal distribution with mean 0 and standard deviation 2.5. These priors regularize the model, constraining the parameters to plausible values based on prior knowledge.

This Bayesian logistic regression model was implemented using the **rstanarm** package, leveraging Markov Chain Monte Carlo (MCMC) sampling for parameter estimation. Further diagnostics and posterior summaries are discussed in the supplementary materials (see Appendix Section B).

#### 3.2 Model Justification

Each predictor in the model represents a specific characteristic of the crime and its context, with hypothesized effects on the likelihood of violent crime. For **PREMISES\_TYPE**, it is expected that certain locations, such as “Transit” or “Outside,” may have higher odds of violent crime due to increased public interactions or a lack of controlled environments. Conversely, premises like “Educational” or “House” are anticipated to have lower odds of violent crime, reflecting

their relatively private or structured settings. Locations categorized as “Other” may include mixed contexts, resulting in varying effects.

For `TIME_OF_DAY`, it is hypothesized that the “Evening” period may exhibit a higher likelihood of violent crimes, as this time is often associated with increased social activities and potential for interpersonal conflict. “Early Morning” may also show elevated odds due to decreased public surveillance and fewer witnesses. In contrast, “Morning” and “Afternoon” are expected to have lower odds, reflecting typical daytime routines with higher levels of oversight and structured activity.

These predictors, incorporated into a Bayesian logistic regression model, highlight relationships between crime characteristics and violent outcomes.

## 4 Results

Our results are summarized in Table 2. The findings align with our expectations, providing intuitions into how premises type and time of day influence the likelihood of violent crimes. To avoid multicollinearity, the model excludes one level from each categorical predictor as the reference group: “Apartment” for `PREMISES_TYPE` and “Early Morning” for `TIME_OF_DAY`. The intercept represents the estimated log-odds of violent crime occurring when all other predictors are held constant at their reference levels. In this case, the estimated log-odds of a violent crime occurring at an “Apartment” during “Early Morning” is 0.881.

The type of premises strongly influences the likelihood of violent crimes. For instance, crimes occurring in “Transit” premises are far more likely to be violent compared to the reference group. The estimated coefficient for `PREMISES_TYPETransit` is 2.363, indicating a substantial increase in the log-odds of violent crime in this context. On the other hand, crimes occurring in “Commercial” and “Educational” premises are less likely to be violent, with coefficients of -0.925 and -0.957, respectively. Similarly, crimes at “House” premises exhibit the lowest likelihood of violence, as reflected by a coefficient of -1.519. These results highlight that location plays a important role in determining whether a crime is violent or non-violent.

Time of day also significantly impacts the likelihood of violent crimes. Compared to the reference level of “Early Morning,” the coefficients for `TIME_OF_DAYAfternoon` and `TIME_OF_DAYEvening` are 0.413 and 0.325, respectively, suggesting an elevated likelihood of violent crimes during these times. The “Morning” time of day exhibits a negligible increase in violent crime likelihood, as shown by a coefficient of 0.074.

Figure 9 shows the range of coefficient estimates for our model within the 90% probability interval. However, because the credibility intervals for several predictors are relatively narrow, particularly for premises types like “Transit,” it is challenging to observe the trends clearly. To address this, Figure 10 was created with the x-axis limited to a range of -5 to 5. Combining Figure 9 and Figure 10, we observe statistical significance for several predictors, including “Transit,” “Afternoon,” and “Evening,” as their credibility intervals do not cross zero.

Table 2: Explanatory model Injury Severity Prediction (n = 1000)

	Violent Crime
(Intercept)	0.881 (0.030)
PREMISES_TYPECommercial	−0.925 (0.034)
PREMISES_TYPEEducational	−0.957 (0.104)
PREMISES_TYPEHouse	−1.519 (0.034)
PREMISES_TYPEOther	−0.138 (0.053)
PREMISES_TYPEOutside	−0.267 (0.032)
PREMISES_TYPETransit	2.363 (0.190)
TIME_OF_DAYMorning	0.074 (0.035)
TIME_OF_DAYAfternoon	0.413 (0.032)
TIME_OF_DAYEvening	0.325 (0.029)
Num.Obs.	37 061
R <sup>2</sup>	0.090
Log.Lik.	−22 567.707
ELPD	−22 577.6
ELPD s.e.	72.5
LOOIC	45 155.2
LOOIC s.e.	145.0
WAIC	45 155.1
RMSE	0.46

The coefficients are reported in log-odds, where a positive value indicates an increased likelihood of violent crime, and a negative value reflects a decreased likelihood. The results demonstrate clear statistical significance for important predictors such as “Transit,” “Afternoon,” and “Evening,” emphasizing the importance of both location and time of day in shaping violent crime outcomes.

## **5 Discussion**

### **5.1 Relationship between Premises Type and Violent Crime**

The analysis underscores that premises type significantly influences the likelihood of violent crimes. Transit locations, often hubs of activity with limited surveillance, exhibit the highest odds of violent crimes. This aligns with urban safety theories suggesting that environments characterized by anonymity and low surveillance facilitate opportunistic and interpersonal violence. Similarly, outdoor areas, with their open and unregulated nature, remain high-risk zones for confrontations. In contrast, locations with inherent structural or social control, such as residential houses and schools, show markedly reduced odds of violent crimes. These findings emphasize the interplay between spatial control and crime likelihood, suggesting that environments fostering visibility, routine, and structural oversight are critical deterrents for violent activity. Policymakers might explore how urban designs, such as transit hubs with natural surveillance features, could mitigate risks in identified high-crime premises.

### **5.2 Relationship between Time of Day and Violent Crime**

Time of day has emerged as another critical determinant of violent crime patterns, reflecting the temporal rhythms of human activity. The elevated risk of violent crimes during afternoon and evening hours correlates with periods of heightened social interaction and mobility. Beyond the opportunity framework, this trend could also be linked to psychological and behavioral factors, such as stress accumulation or alcohol consumption patterns, which intensify during these hours. Morning hours, often structured by routines and higher institutional oversight, exhibit a stabilizing effect, further reinforcing the connection between predictability in human activity and crime reduction. These findings suggest that temporal interventions, such as staggered policing schedules or temporal zoning laws, could play a key role in mitigating violence during high-risk periods.

## **5.3 Expanding Implications for Policy and Prevention**

### **5.3.1 Integrative Safety Planning**

Building upon the findings, integrating urban design with crime prevention could yield significant dividends. For instance, redesigning transit hubs with open sightlines and enhanced illumination may disrupt anonymity and reduce violent crime opportunities. Similarly, encouraging the adoption of Crime Prevention Through Environmental Design (CPTED) principles in outdoor public spaces could transform potential high-crime locations into safer environments. Measures like pedestrian-oriented lighting, community-centric activities in outdoor spaces, and visible security features would align with these findings to promote safer urban environments.

### **5.3.2 Socioeconomic Dimensions of Intervention**

While the current study primarily addresses spatial and temporal patterns, incorporating socioeconomic factors into crime prevention strategies could provide a more holistic framework. For instance, transit areas may not only be risky due to environmental factors but also because they serve as nexus points for marginalized populations. Acknowledging these dynamics, policymakers could complement structural safety measures with socioeconomic upliftment programs, such as housing initiatives near transit hubs or employment opportunities targeting vulnerable demographics.

### **5.3.3 Strategic Resource Allocation**

The predictive insights offered by this study support law enforcement agencies in adopting data-driven approaches for resource allocation. By utilizing the RESTful API developed from this study, stakeholders can simulate various scenarios to predict high-risk situations and proactively deploy resources to mitigate risks. Expanding the API's capabilities to integrate real-time data inputs, such as weather or crowd density, could further enhance its utility for dynamic and responsive safety planning.

## **5.4 Limitations**

### **5.4.1 Data Constraints**

The study uses police-reported crime data, which excludes unreported incidents. This limitation is particularly important in areas where trust in law enforcement may be lower, potentially leading to underreporting. As a result, the analysis might underestimate the prevalence of violent crime or inaccurately represent its distribution. Future research could address this by

incorporating additional data sources, such as hospital records or victim surveys, to provide a more complete picture.

#### **5.4.2 Simplification of Time Categories**

Dividing time into four broad categories (Early Morning, Morning, Afternoon, and Evening) offers clarity but limits the ability to observe finer patterns. For example, crimes occurring during transitional periods, such as early evening or late night, might follow distinct trends not captured by this framework. Future studies could use more detailed time intervals to better align with changes in social activity and behavior.

#### **5.4.3 Contextual Variability**

Factors such as seasonal events, public holidays, or local enforcement policies may affect crime patterns but were not included in this study. While July was chosen as a representative high-activity month, these findings may not extend to other times of the year. Future work could consider these contextual elements to enhance the understanding of violent crime dynamics.

### **5.5 Expanding Future Directions**

#### **5.5.1 Spatial Granularity**

Future studies could use GIS technology to conduct detailed spatial analyses, identifying crime hotspots within larger categories like transit or outdoor spaces. Such an approach could guide precise interventions at the neighborhood or block level, helping to optimize prevention efforts and resource allocation.

#### **5.5.2 Temporal and Behavioral Factors**

Exploring time with finer granularity, such as hourly trends, could reveal patterns that are currently obscured. Additionally, combining behavioral data—such as crowd size or environmental conditions—with existing predictors may uncover new insights, such as the combined impact of transit activity and evening hours on crime risk.

## 5.6 Conclusion

This study identifies the significant roles that premises type and time of day play in shaping the likelihood of violent crimes in Toronto. High-risk locations, such as transit hubs and outdoor spaces, and peak times, such as afternoons and evenings, highlight areas for focused interventions. These findings can inform practical safety measures, such as improved surveillance and better resource allocation during critical times and in high-risk areas. Although the analysis is limited by its broad categories, it demonstrates the value of understanding crime patterns through spatial and temporal dimensions, laying the groundwork for future research and actionable safety strategies.

## **A Appendix: Survey Design to Investigate Factors Influencing Violent Crimes in Toronto**

### **A.1 Introduction**

This survey is designed to gather detailed information from victims of violent crimes in Toronto during the month of July. The primary goal is to supplement existing police data to better understand the factors influencing the likelihood of violent crimes, with a focus on premises type and time of day. By collecting firsthand accounts from victims, the study aims to uncover additional insights that can inform effective crime prevention strategies and policies.

### **A.2 Objectives**

#### **A.2.1 Primary Objective**

To identify and analyze factors contributing to the occurrence of violent crimes in Toronto by gathering detailed victim accounts.

#### **A.2.2 Secondary Objectives**

- Explore the relationship between victim characteristics and the likelihood of experiencing violent crime.
- Assess environmental and situational factors not captured in police reports.
- Evaluate victims' perceptions and experiences to inform supportive services.

### **A.3 Methodology**

#### **A.3.1 Study Design**

A cross-sectional survey will be conducted using a structured questionnaire administered to victims of violent crimes.

#### **A.3.2 Target Population**

- Individuals aged 18 and above who were victims of reported violent crimes in Toronto during the month of July.



### **A.3.3 Sampling Frame**

Victims listed in the Toronto Police Service records for violent crimes occurring in July.

### **A.3.4 Sampling Method**

- **Census Approach:** Attempt to include all eligible victims.
- **If Not Feasible:** Employ stratified random sampling based on crime type and location to ensure representativeness.

### **A.3.5 Inclusion Criteria**

- Victims of violent crimes (e.g., assault, robbery) reported in July.
- Aged 18 years or older.
- Able to provide informed consent.

### **A.3.6 Exclusion Criteria**

- Victims under the age of 18.
- Incidents classified as unfounded or withdrawn.
- Individuals unable or unwilling to consent.

## **A.4 Budget Allocation**

The total budget for this survey is \$100,000, allocated as follows:

- **Recruitment and Outreach (\$25,000):**
  - Partnering with Toronto Police Services to contact victims via official channels.
  - Direct mailers and digital outreach (e.g., SMS, emails).
- **Incentives for Participation (\$60,000):**
  - \$20 gift card for completing the survey.
  - Estimated 3,000 participants.
- **Survey Design and Implementation (\$10,000):**
  - Professional survey platform subscription.
  - Data cleaning and preprocessing costs.
- **Contingency Funds (\$5,000):**
  - Unanticipated costs related to participant recruitment or data handling.

This allocation prioritizes participant incentives to ensure robust response rates while maintaining sufficient funding for recruitment and operational expenses.

## **A.5 Survey Instrument**

The questionnaire is designed to collect detailed information from victims of violent crimes. Attention checks are included to ensure data quality. The survey link: [Survey Form](#).

### **A.5.1 Section 1: Demographics**

**1. What is your age?**

*Required*

**2. What is your gender?**

*Required*

- Male
- Female
- Non-binary
- Prefer not to say

**3. What is your ethnicity?**

*Required*

**4. What is your highest level of education completed?**

*Required*

**5. What is your current occupation?**

*Required*

### **A.5.2 Section 2: Incident Details**

**6. On what date did the incident occur?**

*Required (e.g., January 7, 2019)*

**7. At what time did the incident occur?**

*Required*

**8. Where did the incident occur?**

*Required*

**9. What type of premises was it?**

*Required*

- Apartment
- House
- Commercial
- Outside
- Transit
- Educational
- Other: [Open-ended response]

**10. What were you doing just before the incident?**

*Required*

**11. Were you familiar with this area?**

*Required*

- Yes
- No

**12. Select the option that does NOT belong in this list of fruits:**

- Apple
- Banana
- Carrot
- Orange

**A.5.3 Section 3: Circumstances of the Incident**

**13. Was the perpetrator known to you?**

*Required*

- Yes

- No
14. **How many perpetrators were involved?**  
*Required*
15. **Can you describe the perpetrator(s)?**  
*Required*
16. **Were any weapons used?**  
*Required*
- Yes
  - No
  - If Yes: *Please specify the type of weapon*
17. **Do you know or suspect the motive behind the crime?**  
*Required*

#### **A.5.4 Section 4: Perceptions and Feelings**

18. **Prior to the incident, how safe did you feel in that area?**  
*Required (Likert scale: Not Safe at All to Very Safe)*
19. **Were you aware of any risks or previous incidents in the area?**  
*Required*
- Yes
  - No
20. **How did you feel during the incident?**  
*Required*
21. **How have you felt since the incident?**  
*Required*

#### **A.5.5 Section 5: Reporting and Response**

22. **How was the incident reported?**  
*Required*
- Self-reported

- Witness reported
- Other: [Open-ended response]

23. **Approximately how long did it take for authorities to respond?**

*Required*

24. **How satisfied are you with the response from authorities?**

*Required (Likert scale: Not Satisfied at All to Very Satisfied)*

25. **Have you received any support services since the incident?**

*Required*

- Yes
- No
- If Yes: *Please specify*

26. **Which of these options has an even number of letters?**

- Apple
- Banana
- Orange
- Kiwi

#### **A.5.6 Section 6: Additional Factors**

27. **Had you consumed alcohol or drugs prior to the incident?**

*Required*

- Yes
- No

28. **Do you believe the perpetrator(s) had consumed alcohol or drugs?**

*Required*

- Yes
- No

- Don't know

29. **What were the lighting conditions at the time?**

*Required*

- Well-lit
- Dimly lit
- Dark

30. **What was the weather like at the time?**

*Required*

- Clear
- Rainy
- Snowy
- Other: [Open-ended response]

31. **Were there any security measures present?**

*Required*

- Cameras
- Security personnel
- None
- Other: [Open-ended response]

32. **Were there other people around at the time?**

*Required*

- Yes
- No
- If Yes: *Approximately how many?*

## **A.6 Ethical Considerations**

### **A.6.1 Informed Consent**

- Participants will be informed about the purpose, risks, and benefits of the survey. Consent is mandatory.

### **A.6.2 Confidentiality**

- Responses will be anonymized. Data will be securely stored and used only for research purposes.

### **A.6.3 Protection of Participants**

- Sensitive topics will be handled by trained staff. Resources for support services will be provided.

### **A.6.4 Ethical Approval**

- The study protocol will undergo review by a Research Ethics Board (REB).

## **A.7 Data Collection Procedures**

### **A.7.1 Mode of Data Collection**

- **Primary Method:** Online surveys through a secure platform.
- **Alternative Methods:** Telephone interviews for participants with limited internet access.

### **A.7.2 Timeline**

- Data collection will be completed within two months of the survey launch.

### **A.7.3 Contact Protocol**

- Initial invitations via email or mail, followed by reminders to maximize participation.

## A.8 Data Management and Analysis

### A.8.1 Data Entry and Storage

- Data will be entered into a secure, password-protected database.
- Regular backups will be performed.
- Data will be encrypted to protect confidentiality.

### A.8.2 Data Cleaning

- Verification of data entry accuracy.
- Handling of missing data according to predefined protocols.

### A.8.3 Data Simulation

To validate and explore the survey design, a simulated dataset was generated to reflect the variables of interest. This dataset includes:

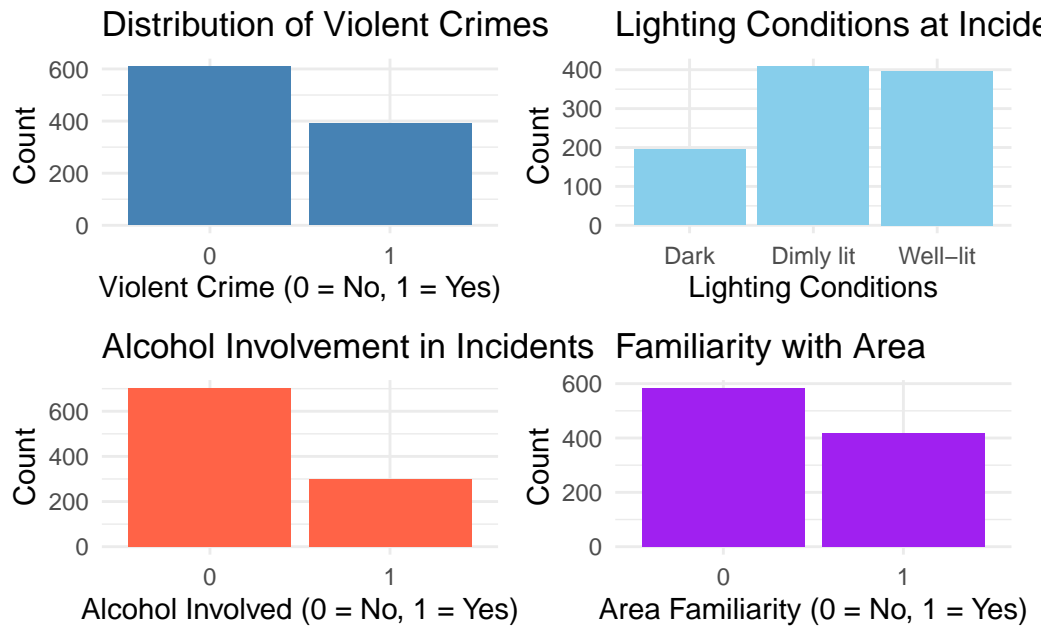
- **VIOLENT\_CRIME**: Binary variable indicating whether a crime was violent (1) or non-violent (0).
- **ALCOHOL\_INVOLVED**: Binary variable indicating whether alcohol was involved in the incident (1 = Yes, 0 = No).
- **AREA\_FAMILIARITY**: Binary variable indicating whether the victim was familiar with the area (1 = Yes, 0 = No).
- **LIGHTING\_CONDITIONS**: Categorical variable describing the lighting conditions at the time of the incident ('Well-lit', 'Dimly lit', 'Dark').

The dataset was generated using R, saved as a CSV file, and subsequently visualized to ensure the data reflects the survey's design specifications.

Table 3: Preview of the Simulated Crime Dataset

Variable	V1	V2	V3	V4	V5
VIOLENT_CRIME	1	1	0	0	1
ALCOHOL_INVOLVED	0	0	0	1	0
AREA_FAMILIARITY	0	1	1	0	0
LIGHTING_CONDITIONS	Dimly lit	Dimly lit	Well-lit	Dark	Well-lit





(a) Proportion of Violent Crimes

Figure 3: Visualizations of Simulated Crime Data Variables

## A.9 Quality Assurance

### A.9.1 Pilot Testing

- The survey will be pilot tested with a small group to identify issues with question clarity or sensitivity.
- Revisions will be made based on feedback.

### A.9.2 Data Monitoring

- Regular checks for data completeness and consistency.
- Interim analyses to identify any emerging issues.

## A.10 Dissemination of Results

- Findings will be compiled into a report.
- Academic papers will be prepared for publication in peer-reviewed journals.
- Results will be shared with the Toronto Police Service, policymakers, and community organizations.
- Presentations will be made at relevant conferences and meetings.

## **A.11 Conclusion**

This survey aims to enhance understanding of violent crimes in Toronto by incorporating victim perspectives and experiences. The data collected will contribute to more effective prevention strategies and support services, ultimately aiming to reduce the incidence of violent crimes and improve community safety.

## B Model details

### B.1 Posterior Predictive Check and Prior-Posterior Comparison

The posterior predictive check in Figure 4 provides an evaluation of how well the Bayesian logistic regression model replicates the observed data. The posterior distribution ( $y_{\text{rep}}$ ) closely aligns with the observed data ( $y$ ), indicating that the model provides a good fit and captures the essential patterns in the data.

In Figure 5, the posterior distributions of the model parameters are compared against their priors. This visual comparison highlights how the observed data influenced the posterior estimates. For instance, predictors such as `PREMISES_TYPETransit` and `TIME_OF_DAYAfternoon` show a notable shift in their posterior distributions compared to the priors, reflecting strong evidence provided by the data. Conversely, predictors with minimal shifts suggest weaker evidence or smaller effects on the model outcome.

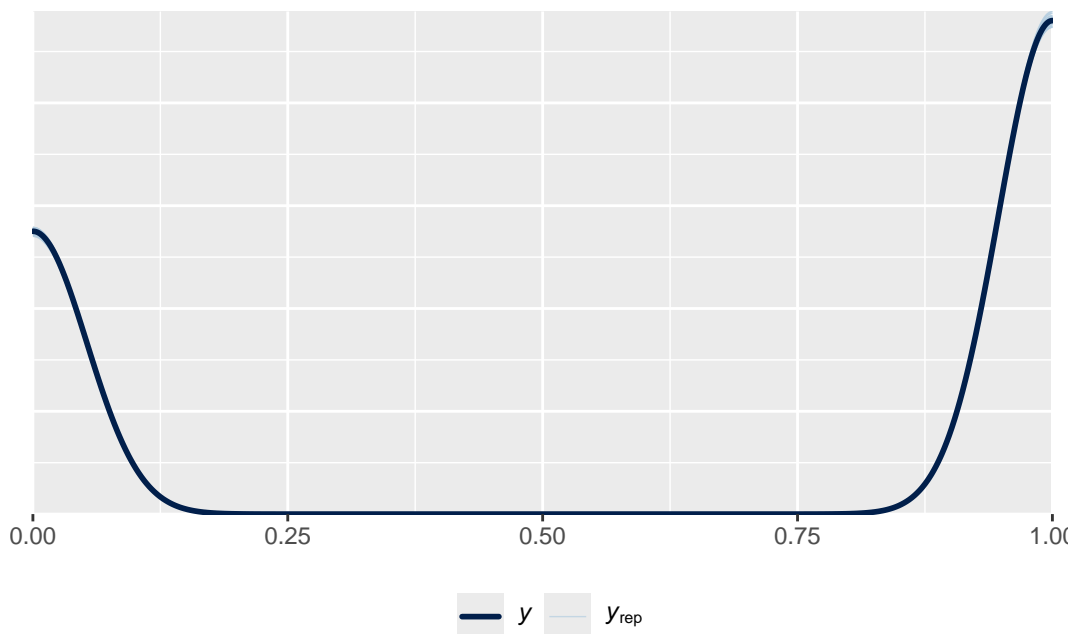


Figure 4: Posterior distribution for logistic regression model

### B.2 Diagnostics

#### B.2.1 Markov Chain Monte Carlo Convergence Check

To assess the convergence of the Markov Chain Monte Carlo (MCMC) sampling for the Bayesian logistic regression model, trace plots and Rhat plots were analyzed. These diag-

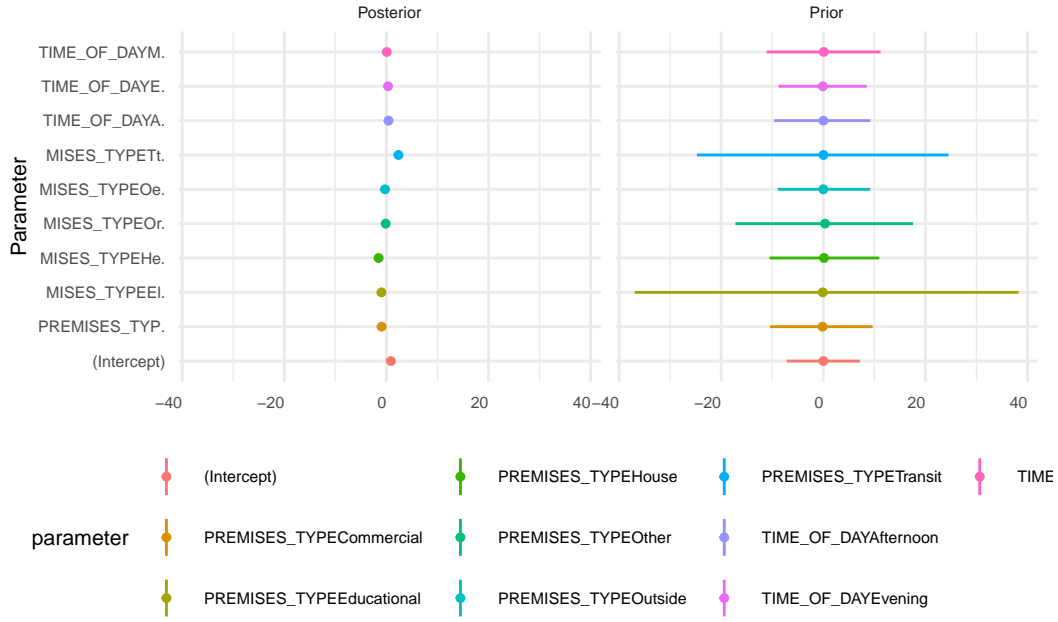


Figure 5: Comparing the posterior with the prior

nostics are to ensure that the chains mix well and converge to a stable distribution, providing reliable parameter estimates.

### B.2.1.1 Trace Plots

Figure 6 and Figure 7 display the trace plots for the model's intercept, premises type predictors, and time of day predictors. Each plot shows the sampled parameter values across iterations for all four chains.

- **Premises Type Predictors:** As seen in Figure 6, the chains for predictors like **House**, **Commercial**, **Educational**, and **Outside** fluctuate around stable mean values, indicating good mixing and convergence.
- **Time of Day Predictors:** Figure 7 illustrates the trace plots for **Morning**, **Afternoon**, and **Evening** time categories, along with **Transit Premises**. The chains for these predictors also show no significant trends or drift, confirming convergence.

## B.3 90% Credibility Interval

Figure 9 and Figure 10 visualize the 90% credible intervals for the predictors in the Bayesian logistic regression model. These plots provide intuitions into the uncertainty surrounding the estimated coefficients for each predictor.

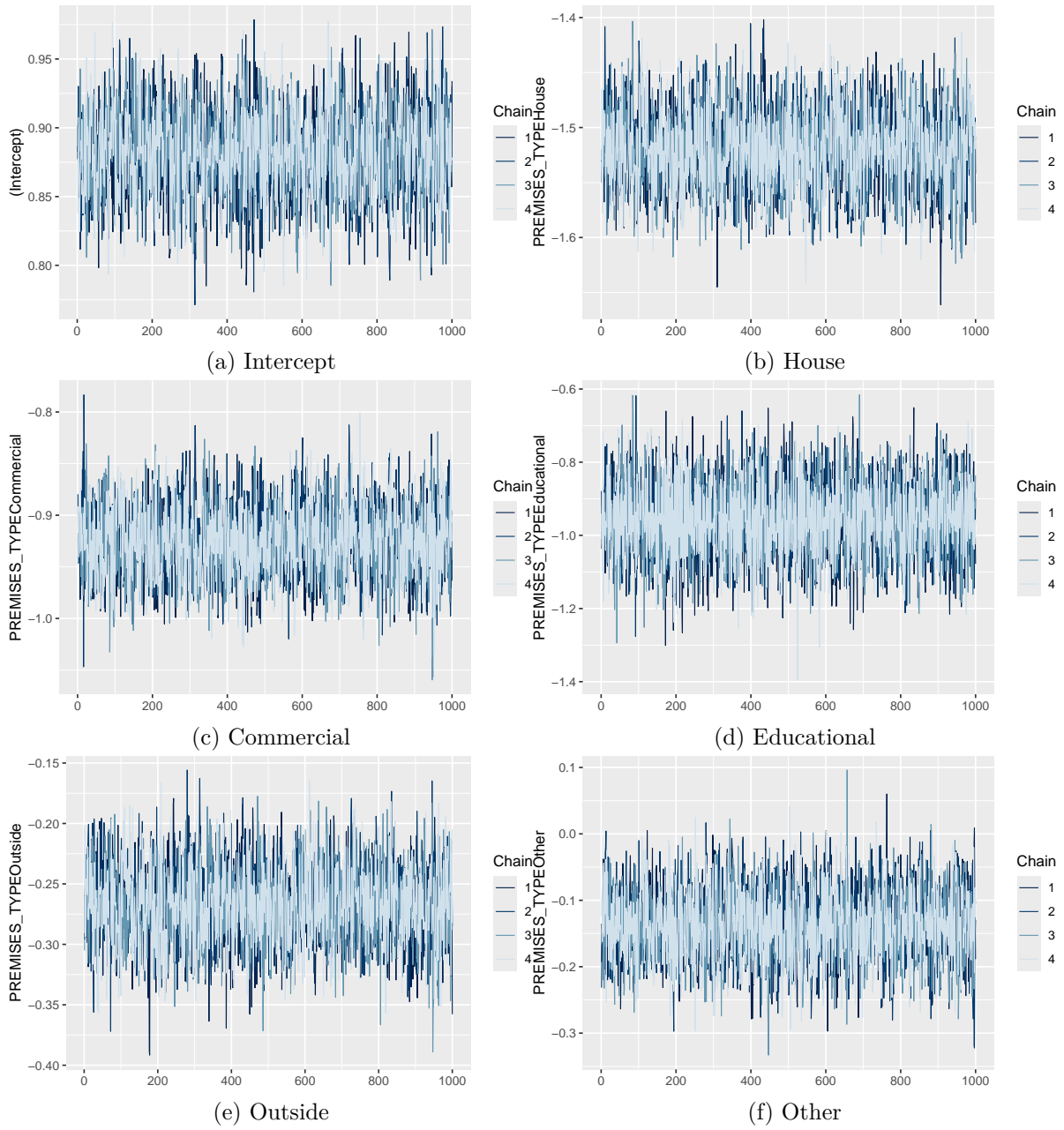


Figure 6: Trace plots of intercept and premises type predictors

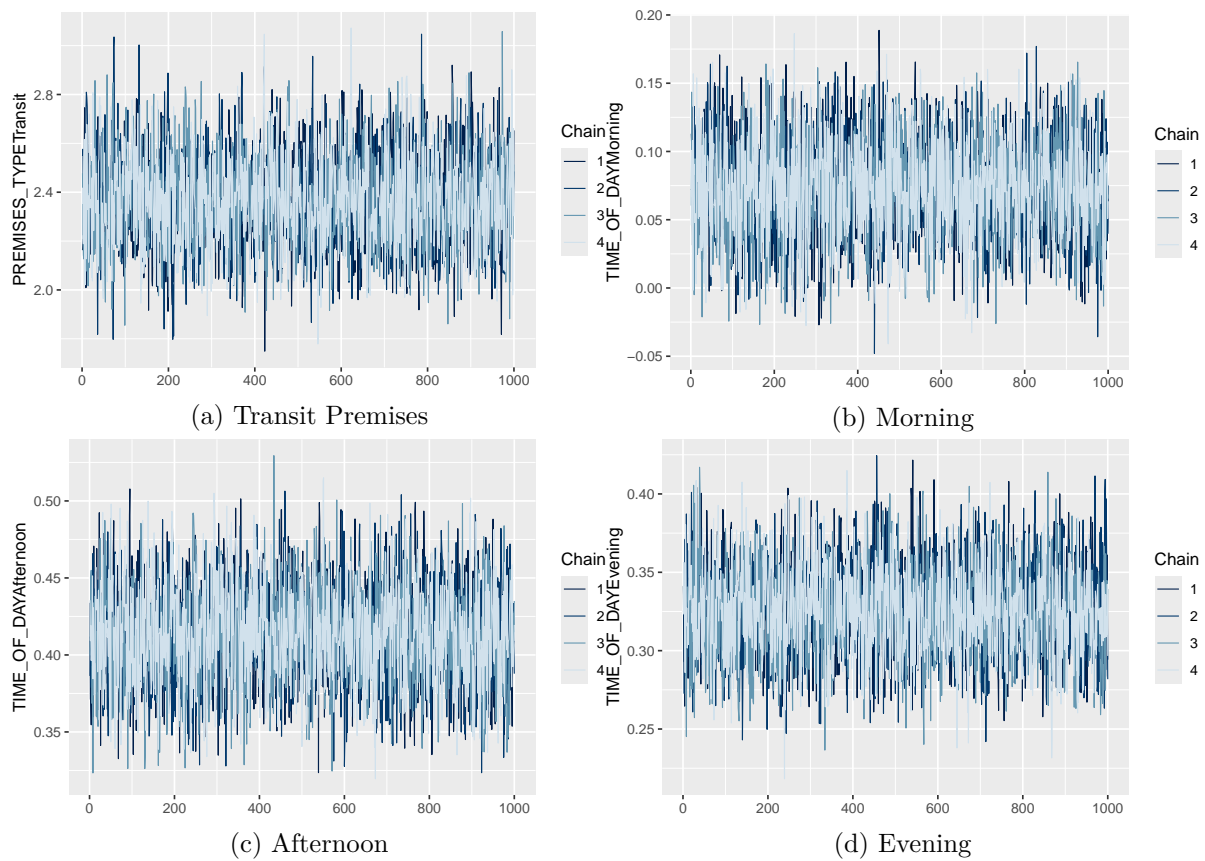


Figure 7: Trace plots of time of day predictors

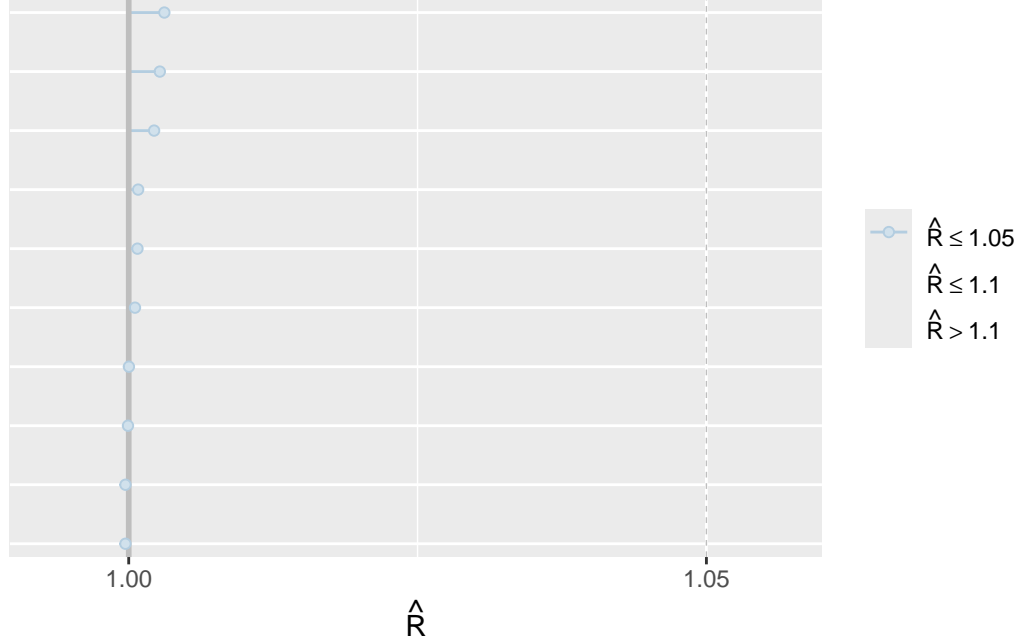


Figure 8: Rhat plot for MCMC convergence

Figure 9 displays the 90% credible intervals for all predictors without restrictions on the x-axis. This plot shows how the predictors such as `PREMISES_TYPE` and `TIME_OF_DAY` influence the likelihood of violent crimes. For example, `PREMISES_TYPETransit` has a high positive credible interval, indicating a significant increase in the likelihood of violent crime in transit areas. Conversely, `PREMISES_TYPECommercial` and `PREMISES_TYPEEducational` have negative credible intervals, indicating a lower likelihood of violent crimes in these premises types.

In Figure 10, the x-axis has been restricted to a range of -5 to 5 to better visualize predictors with smaller intervals. This refined view helps highlight subtle but meaningful differences between categories like `TIME_OF_DAYAfternoon` and `TIME_OF_DAYEvening`, which show a noticeable increase in the likelihood of violent crimes compared to the reference category (`TIME_OF_DAYEarly Morning`).

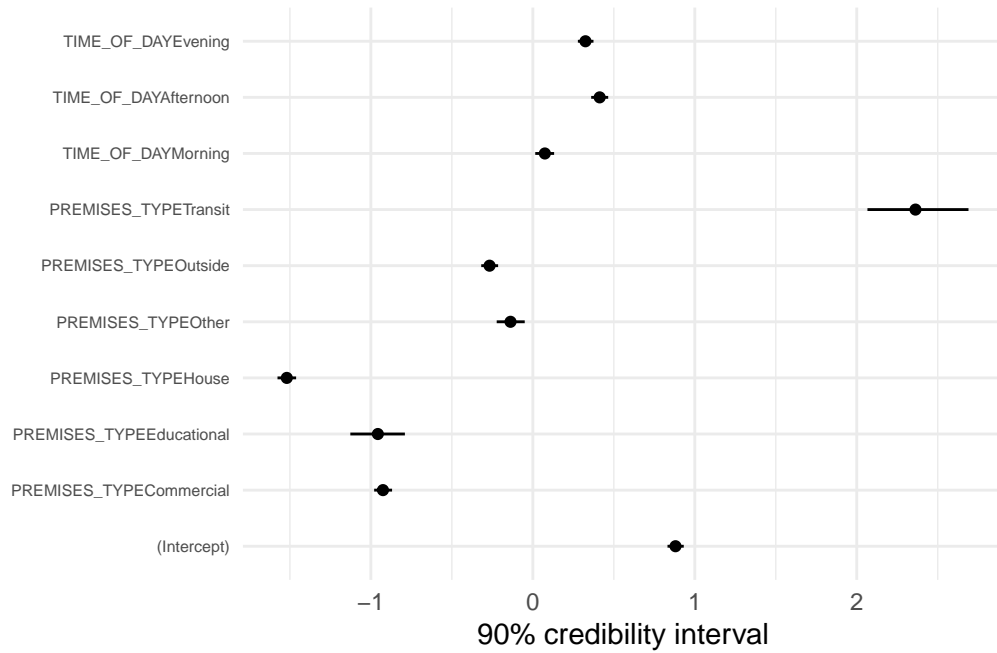


Figure 9: 90% credible intervals for predictors of violent crime likelihood

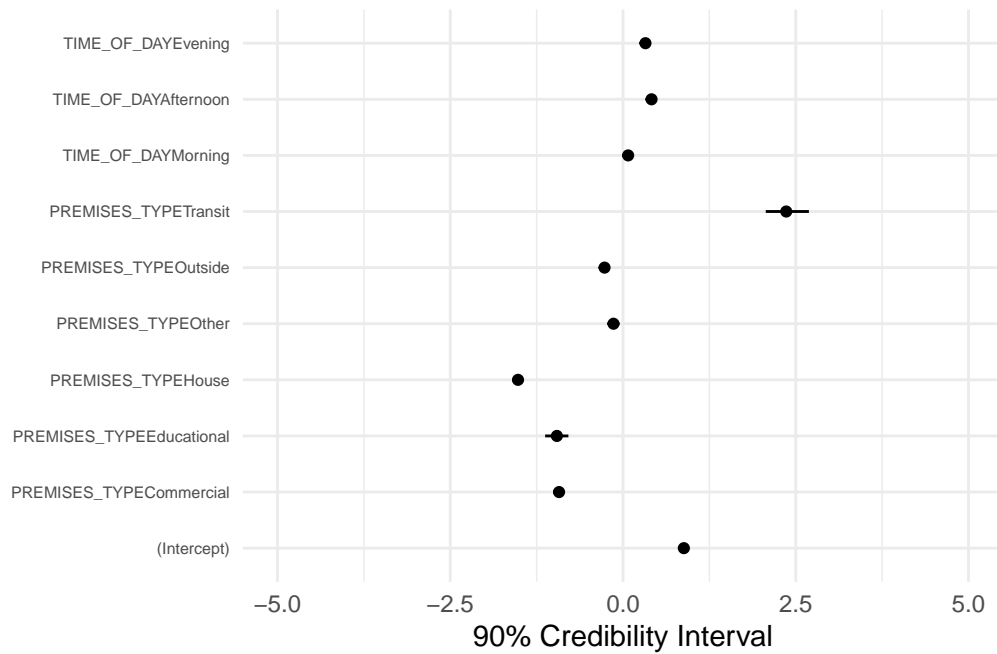


Figure 10: 90% credible intervals for predictors of violent crime likelihood with restricted x-axis limits



## References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. <https://tellingstorieswithdata.com/>.
- Allen, Jeff, Barret Schloerke, Joe Cheng, and Doug Ashton. 2023. *plumber: Turn your R code into a web API*. <https://www.rplumber.io/>.
- Arel-Bundock, Vincent. 2024. *marginaleffects: Marginal Effects, Marginal Means, Predictions, and Contrasts*. <https://vincentarelbundock.github.io/marginaleffects/>.
- Developers, Apache Arrow. 2024. *arrow: Integration to Apache Arrow*. <https://arrow.apache.org/>.
- Firke, Sam. 2023. *janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “rstanarm: Bayesian applied regression modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- Grolemund, Garrett, and Hadley Wickham. 2023. *lubridate: Make Dealing with Dates a Little Easier*. <https://CRAN.R-project.org/package=lubridate>.
- Müller, Kirill. 2023. *here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- Open Data Toronto. 2021. “Major Crime Indicators.” <https://open.toronto.ca/dataset/major-crime-indicators/>.
- Pinkerton Insights. 2021. “The Seasonality of Crime.” *Perspectives in Crime Blog*. <https://pinkerton.com/our-insights/blog/the-seasonality-of-crime>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley, 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, and Jim Hester. 2023. *testthat: Unit Testing for R*. <https://CRAN.R-project.org/package=testthat>.
- World Meteorological Organization. 2024. “July Sets New Temperature Records.” *World Meteorological Organization News*. <https://wmo.int/media/news/july-sets-new-temperature-records>.
- Zhu, Hao. 2023. *kableExtra: Construct Complex Table with 'kable()'*. <https://CRAN.R-project.org/package=kableExtra>.