

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

November 26, 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 reveals 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 offer critical insights for urban safety planning and targeted crime prevention strategies.

Table of contents

1	Introduction	1
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
2	Data	3
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>.

3	Model	6
3.1	Model Set-up	8
3.2	Model Justification	8
4	Results	9
5	Discussion	11
5.1	Relationship between Premises Type and Violent Crime	11
5.2	Relationship between Time of Day and Violent Crime	11
5.3	Implications for Policy and Prevention	11
5.4	Limitations	12
5.4.1	Data Constraints	12
5.4.2	Simplification of Time Categories	12
5.5	Future Steps	12
A	Appendix: Survey Design to Investigate Factors Influencing Violent Crimes in Toronto	13
A.1	Introduction	13
A.2	Objectives	13
A.2.1	Primary Objective	13
A.2.2	Secondary Objectives	13
A.3	Methodology	13
A.3.1	Study Design	13
A.3.2	Target Population	13
A.3.3	Sampling Frame	14
A.3.4	Sampling Method	14
A.3.5	Inclusion Criteria	14
A.3.6	Exclusion Criteria	14
A.4	Survey Instrument	14
A.4.1	Section 1: Demographics	14
A.4.2	Section 2: Incident Details	15
A.4.3	Section 3: Circumstances of the Incident	15
A.4.4	Section 4: Perceptions and Feelings	15
A.4.5	Section 5: Reporting and Response	15
A.4.6	Section 6: Additional Factors	16
A.5	Ethical Considerations	16
A.5.1	Informed Consent	16
A.5.2	Confidentiality	16
A.5.3	Protection of Participants	17
A.5.4	Ethical Approval	17
A.6	Data Collection Procedures	17
A.6.1	Mode of Data Collection	17
A.6.2	Interviewer Training	17

A.6.3	Timeline	17
A.6.4	Contact Protocol	17
A.7	Data Management and Analysis	18
A.7.1	Data Entry and Storage	18
A.7.2	Data Cleaning	18
A.7.3	Analysis Plan	18
A.8	Quality Assurance	18
A.8.1	Pilot Testing	18
A.8.2	Data Monitoring	18
A.9	Dissemination of Results	19
A.10	Conclusion	19
B	Model details	19
B.1	Posterior Predictive Check and Prior-Posterior Comparison	19
B.2	Diagnostics	19
B.2.1	Markov Chain Monte Carlo Convergence Check	19
B.3	90% Credibility Interval	21
	References	25

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. Our data is sourced from the Open Data Toronto portal (Open Data Toronto 2021), specifically from the “Motor Vehicle Collisions Involving Killed or Seriously Injured Persons” dataset, which captures detailed information on traffic incidents in Toronto where individuals were either killed or seriously injured. This dataset, compiled by the Toronto Police Service, includes variables that record the time, location, environmental conditions, and specific parties involved in each collision, showing factors that may increase the risk of severe outcomes.

The detailed information within the dataset enables a rigorous analysis of patterns in collision severity. By examining factors such as driver condition, road surface, and lighting conditions, we aim to identify high-risk scenarios that may require targeted safety interventions. Following principles outlined in “Telling Stories with Data” (Alexander 2023), this study prioritizes transparency and reproducibility in presenting data-driven findings that can inform road safety policies.

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.”

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 comprehensive 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 crucial 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 crucial 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 reveals 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 insights 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.

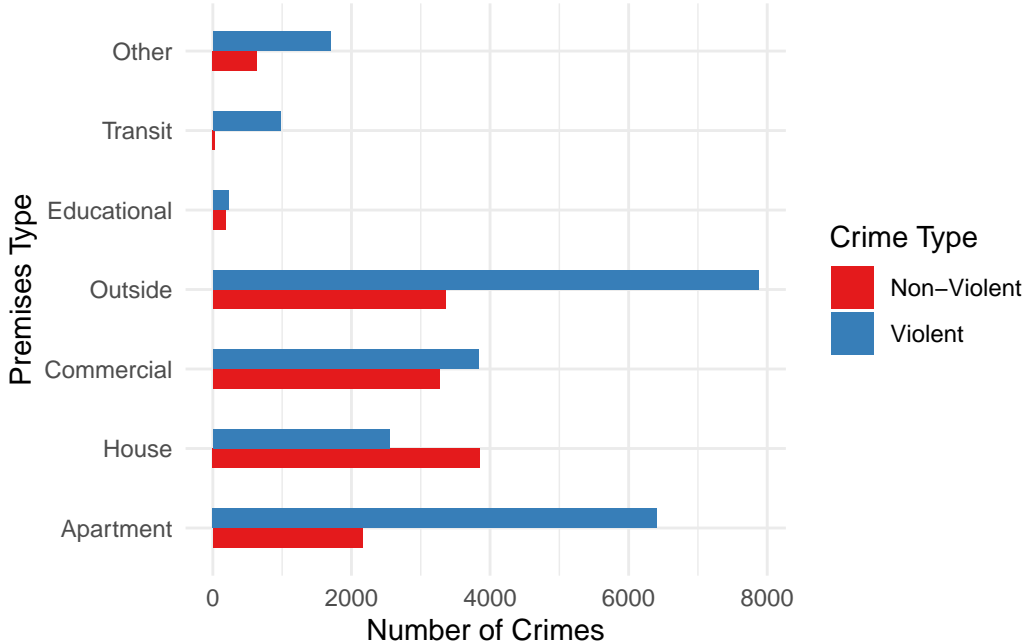


Figure 1: The distribution of violent vs. non-violent crimes by premises type

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](#).

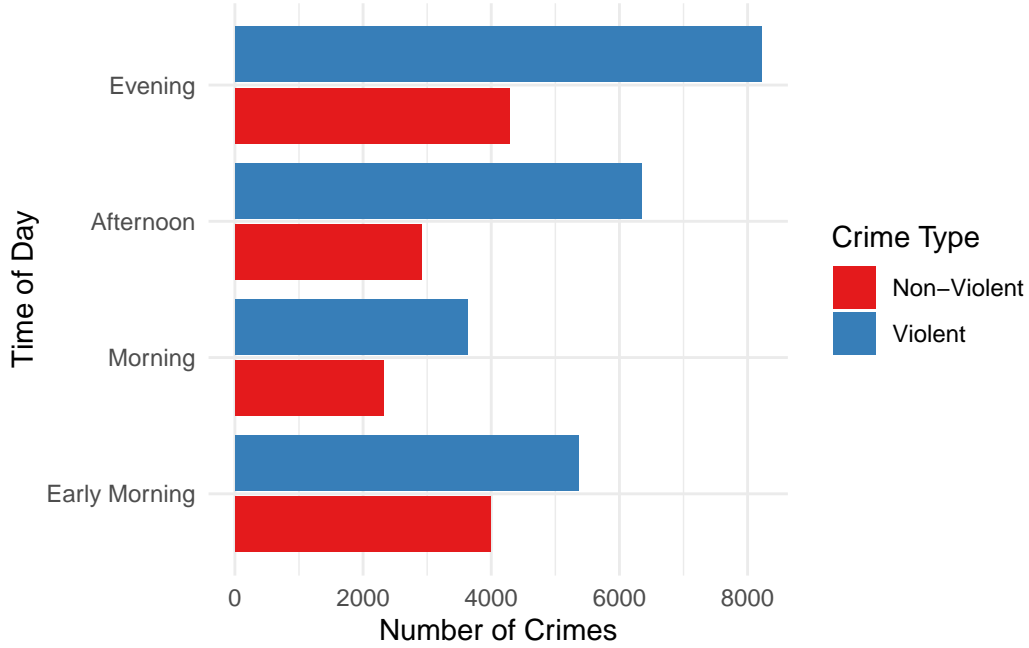


Figure 2: The distribution of violent vs. non-violent crimes by time of day

3.1 Model Set-up

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

$$\text{logit}(\pi_i) = \alpha + \beta_1 \times \text{PREMISES_TYPE}_i + \beta_2 \times \text{TIME_OF_DAY}_i \quad (2)$$

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

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

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

In this model:

- (y_i) represents the binary outcome variable indicating whether a crime is classified as violent ((1) for violent crimes, (0) for non-violent crimes).
- (π_i) is the probability of a violent crime.
- (α) is the intercept, representing the baseline log-odds of a violent crime when all predictors are at their reference levels.
- (β_1) and (β_2) are the coefficients associated with the predictor variables:
 - **PREMISES_TYPE**: the type of premises where the crime occurred.
 - **TIME_OF_DAY**: the time of day the crime occurred.

The logistic link function ($\text{logit}(_i)$) models the log-odds of violent crime as a linear combination of the predictors. The priors for the intercept ($(_)$) and the coefficients ($(_1), (_2)$) 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.

By incorporating these predictors into the Bayesian logistic regression model, this study seeks to identify and quantify the relationships between crime characteristics and the likelihood of violent outcomes, providing practical findings for targeted interventions.

4 Results

Our results are summarized in Table 2. The findings align with our expectations, providing insights 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

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 ²	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

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 7 (see Section B.3) 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 8 was created with the x-axis limited to a range of -5 to 5. Combining Figures 7 and 8, we observe statistical significance for several predictors, including “Transit,” “Afternoon,” and “Evening,” as their credibility intervals do not cross zero.

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 highlights that premises type significantly influences the likelihood of violent crimes. Crimes occurring in “Transit” locations are strongly associated with higher odds of being classified as violent compared to the reference group (“Apartment”). This finding aligns with expectations, as transit environments often involve high volumes of people and limited surveillance, which may increase the potential for interpersonal conflicts. Similarly, “Outside” locations exhibit a higher likelihood of violent crimes, likely due to their open and less controlled nature, which may facilitate confrontations or opportunistic violence. In contrast, premises such as “Commercial,” “Educational,” and “House” are associated with lower odds of violent crimes. These findings suggest that environments with greater structural or social control (e.g., schools, private residences) are less conducive to violent activities. These insights

underscore the importance of targeted safety measures in high-risk locations, such as enhancing surveillance in transit areas or increasing community policing efforts in outdoor public spaces.

5.2 Relationship between Time of Day and Violent Crime

Time of day also plays an important role in determining the likelihood of violent crimes. The model results indicate that “Afternoon” and “Evening” periods are associated with elevated odds of violent crimes compared to the reference period (“Early Morning”). This pattern likely reflects increased social interactions and activities during these times, which may lead to heightened opportunities for conflicts or criminal acts. In contrast, “Morning” hours exhibit a negligible increase in violent crime likelihood, possibly reflecting a time of day characterized by structured routines and reduced social friction. These findings highlight the need for time-sensitive interventions, such as increased police presence during evening hours or community-based programs to address the underlying causes of violent behavior during high-risk times.

5.3 Implications for Policy and Prevention

The results provide practical findings for crime prevention and public safety policies. For example, transit and outdoor areas, identified as high-risk premises, could benefit from increased surveillance, improved lighting, and targeted community engagement programs. Similarly, allocating law enforcement resources more strategically to focus on evening and afternoon periods could help deter violent crimes during these high-risk times. The findings also support urban planning efforts that enhance safety features in public spaces, such as the installation of security cameras and the promotion of neighborhood watch programs. By addressing the specific premises and times associated with violent crimes, policymakers can implement evidence-based strategies to reduce violence and improve community safety.

5.4 Limitations

5.4.1 Data Constraints

One limitation of this study is the exclusion of crimes that could not be definitively classified as violent or non-violent. While necessary to ensure data consistency, this filtering process may omit incidents that could provide additional context for understanding violent crime dynamics. Future research should explore ways to include these ambiguous cases to improve the comprehensiveness of the analysis.

5.4.2 Simplification of Time Categories

Grouping time into four broad categories (Early Morning, Morning, Afternoon, and Evening) simplifies analysis but may obscure more granular patterns in violent crime occurrence. Future research could explore finer temporal granularity, such as analyzing crimes on an hourly basis, to capture additional nuances in time-of-day effects.

5.5 Future Steps

To enhance the utility and scope of this research, future studies should consider incorporating additional predictors, such as socioeconomic factors, neighborhood characteristics, or the presence of law enforcement resources. These variables could provide a more comprehensive understanding of the conditions contributing to violent crimes. Additionally, integrating spatial analysis with precise location data could identify specific hotspots for targeted interventions. Policymakers and city planners could leverage these insights to allocate resources more effectively, reduce violent crimes, and promote public safety in Toronto and beyond.

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 Survey Instrument

The questionnaire is divided into six sections:

A.4.1 Section 1: Demographics

1. What is your age?
2. What is your gender? (Male, Female, Non-binary, Prefer not to say)
3. What is your ethnicity? (Open-ended)
4. What is your highest level of education completed?
5. What is your current occupation?

A.4.2 Section 2: Incident Details

6. On what date did the incident occur?
7. At what time did the incident occur? (Specify hour)
8. Where did the incident occur? (Specific address or nearest intersection)
9. What type of premises was it? (Apartment, House, Commercial, Outside, Transit, Educational, Other)
10. What were you doing just before the incident? (Open-ended)
11. Were you familiar with this area? (Yes/No)

A.4.3 Section 3: Circumstances of the Incident

12. Was the perpetrator known to you? (Yes/No)
13. How many perpetrators were involved?
14. Can you describe the perpetrator(s)? (Open-ended)
15. Were any weapons used? (Yes/No)
 - If Yes: Please specify the type of weapon.
16. Do you know or suspect the motive behind the crime? (Open-ended)

A.4.4 Section 4: Perceptions and Feelings

17. Prior to the incident, how safe did you feel in that area? (Likert scale 1-5)
18. Were you aware of any risks or previous incidents in the area? (Yes/No)
19. How did you feel during the incident? (Open-ended)
20. How have you felt since the incident? (Open-ended)

A.4.5 Section 5: Reporting and Response

21. How was the incident reported? (Self-reported, Witness reported, Other)

22. Approximately how long did it take for authorities to respond? (Specify in minutes)

23. How satisfied are you with the response from authorities? (Likert scale 1-5)

24. Have you received any support services since the incident? (Yes/No)

- If Yes: Please specify.

A.4.6 Section 6: Additional Factors

25. Had you consumed alcohol or drugs prior to the incident? (Yes/No)

26. Do you believe the perpetrator(s) had consumed alcohol or drugs? (Yes/No/Don't know)

27. What were the lighting conditions at the time? (Well-lit, Dimly lit, Dark)

28. What was the weather like at the time? (Clear, Rainy, Snowy, Other)

29. Were there any security measures present? (Cameras, Security personnel, None, Other)

30. Were there other people around at the time? (Yes/No)

- If Yes: Approximately how many?

A.5 Ethical Considerations

A.5.1 Informed Consent

- Participants will receive detailed information about the study's purpose, procedures, risks, and benefits.
- Consent forms will be signed prior to participation.
- Participants can withdraw at any time without penalty.

A.5.2 Confidentiality

- Personal identifiers will be removed or coded.
- Data will be stored securely with access limited to authorized personnel.
- Results will be reported in aggregate to prevent identification.

A.5.3 Protection of Participants

- Interviews will be conducted sensitively to minimize distress.
- Referral information for support services will be provided.
- Participation is entirely voluntary.

A.5.4 Ethical Approval

The study protocol will be reviewed and approved by an Institutional Review Board (IRB) or Research Ethics Board (REB) prior to commencement.

A.6 Data Collection Procedures

A.6.1 Mode of Data Collection

- **Primary Method:** Telephone interviews conducted by trained researchers.
- **Alternative Methods:** In-person interviews or online questionnaires, depending on participant preference.

A.6.2 Interviewer Training

- Training on trauma-informed interviewing techniques.
- Guidelines on handling disclosures and providing support information.

A.6.3 Timeline

Data collection will occur within two months following the incidents to ensure accurate recall.

A.6.4 Contact Protocol

- Initial contact will be made via mail or email explaining the study and requesting participation.
- Follow-up contact will be made to schedule interviews.

A.7 Data Management and Analysis

A.7.1 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.7.2 Data Cleaning

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

A.7.3 Analysis Plan

- **Descriptive Statistics:** Frequencies, means, and standard deviations for all variables.
- **Inferential Statistics:** Logistic regression to identify factors associated with violent crime likelihood.
- **Software:** Statistical analysis will be conducted using R or SPSS.

A.8 Quality Assurance

A.8.1 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.8.2 Data Monitoring

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

A.9 Dissemination of Results

- Findings will be compiled into a comprehensive 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.10 Conclusion

This survey aims to enhance understanding of violent crimes in Toronto by incorporating victim perspectives and experiences. The comprehensive 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 3 provides an evaluation of how well the Bayesian logistic regression model replicates the observed data. The posterior distribution $((y_{rep}))$ 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 4, 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 less data-driven evidence or smaller effects on the model outcome.

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

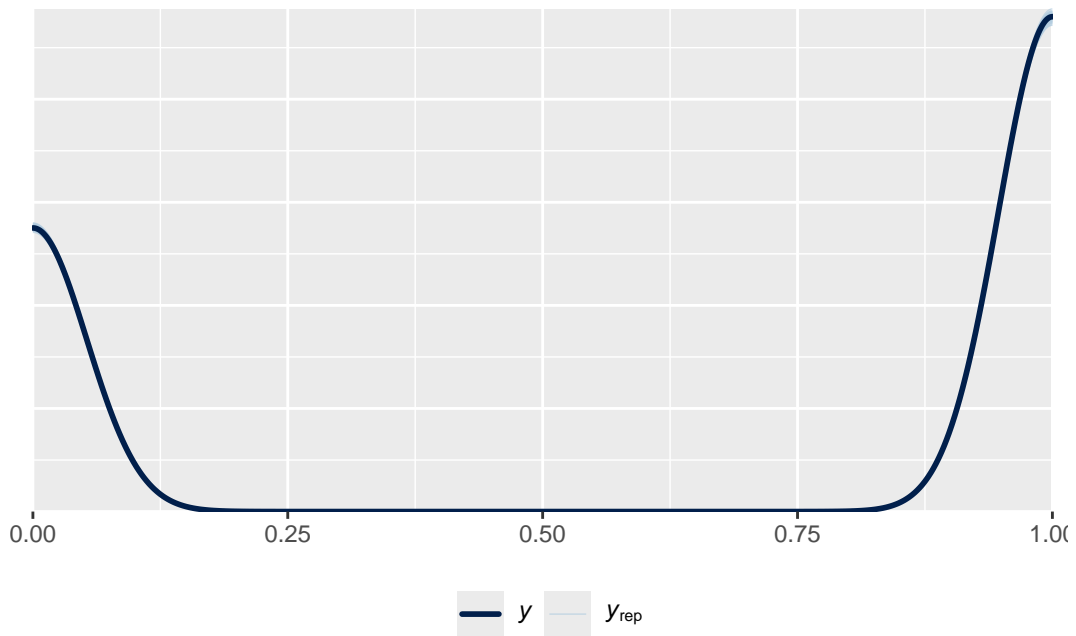


Figure 3: Posterior distribution for logistic regression model

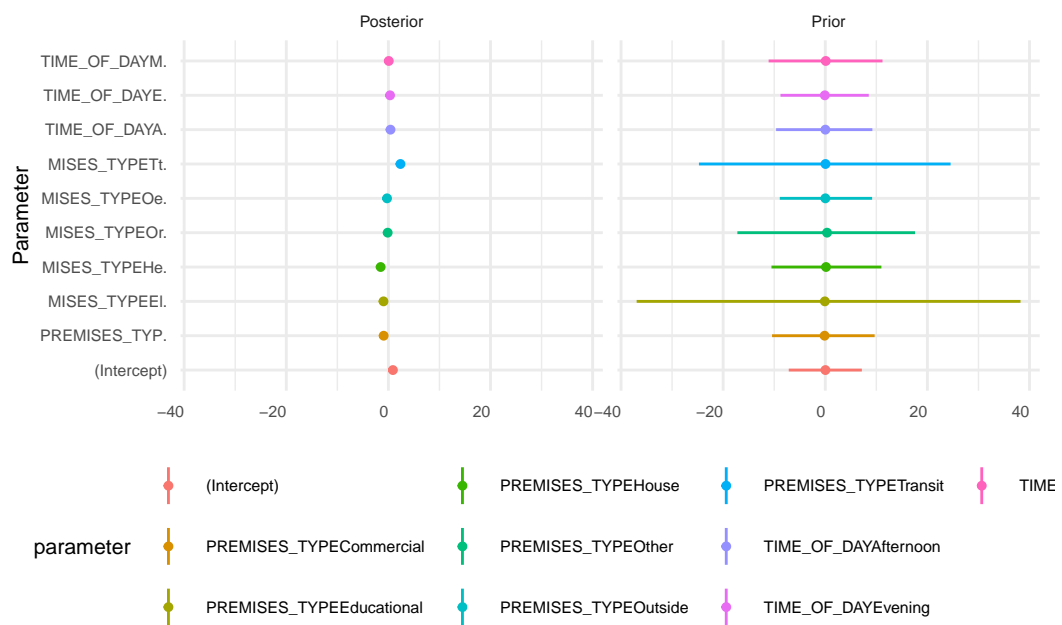


Figure 4: Comparing the posterior with the prior

B.2.1.1 Trace Plots

Figure 5 and Figure 6 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 5, 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 6 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

Figures Figure 8 and Figure 9 visualize the 90% credible intervals for the predictors in the Bayesian logistic regression model. These plots provide insights into the uncertainty surrounding the estimated coefficients for each predictor.

Figure Figure 8 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 Figure 9, 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`).

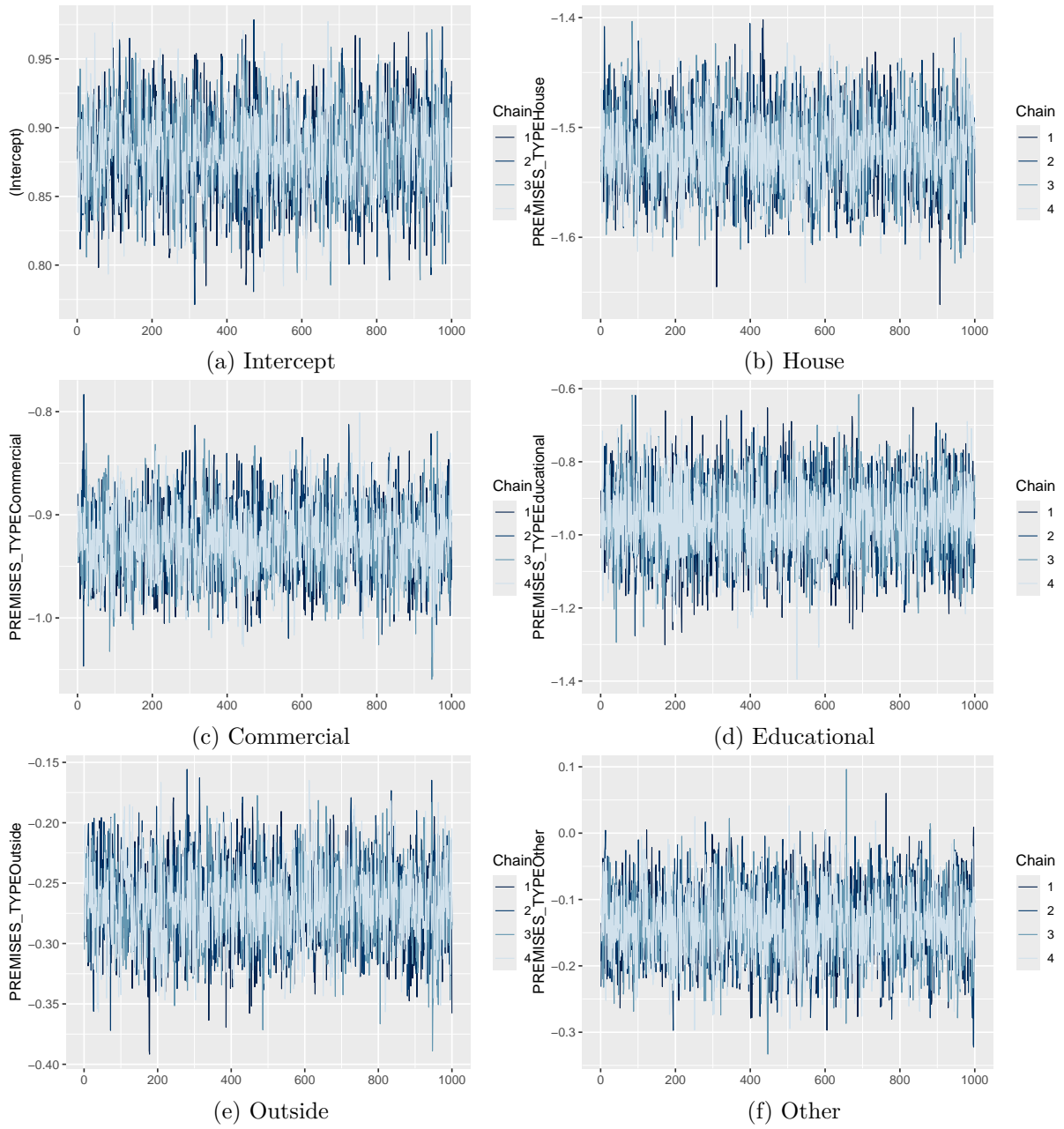


Figure 5: Trace plots of intercept and premises type predictors

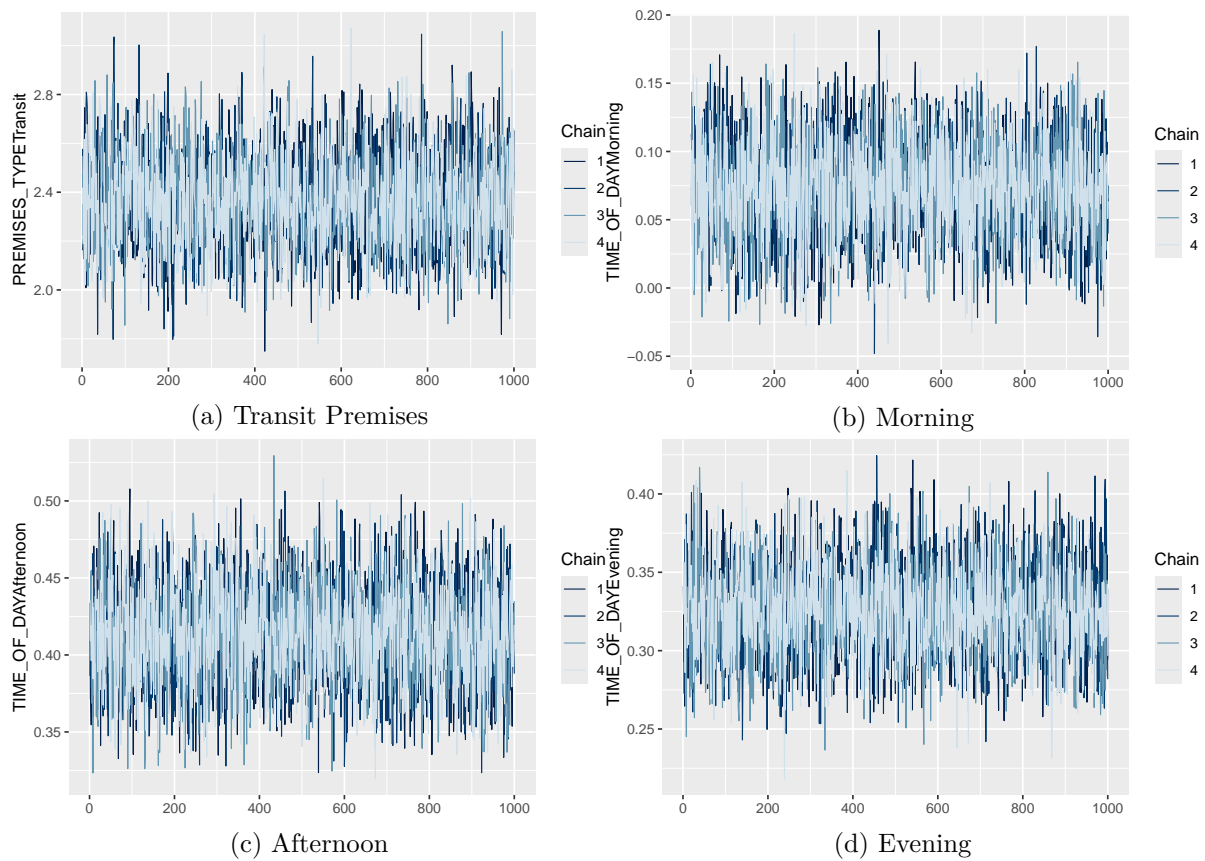


Figure 6: Trace plots of time of day predictors

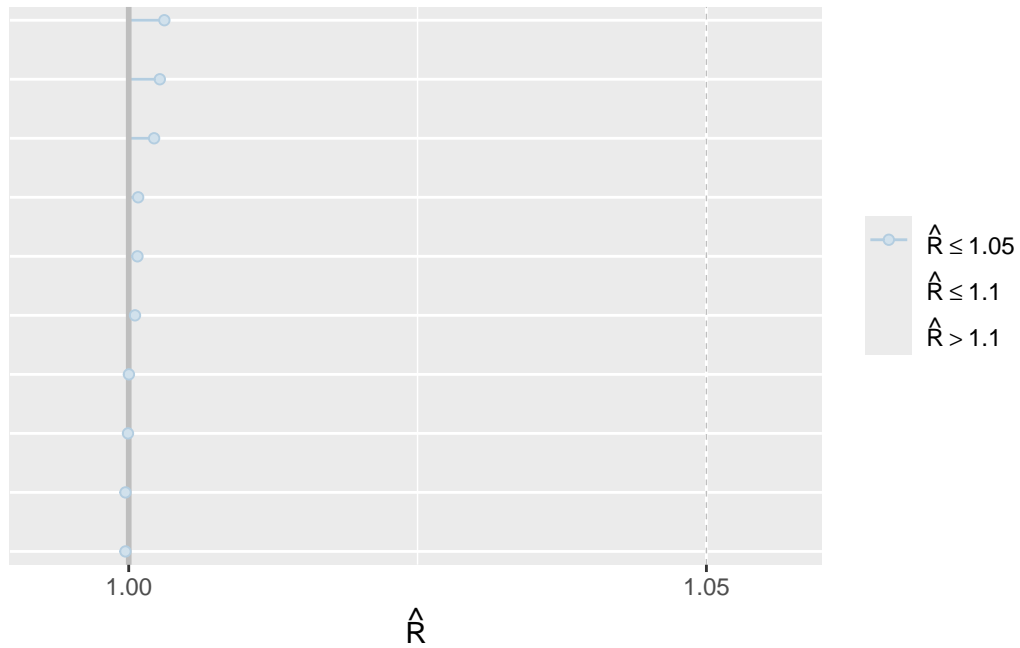


Figure 7: Rhat plot for MCMC convergence

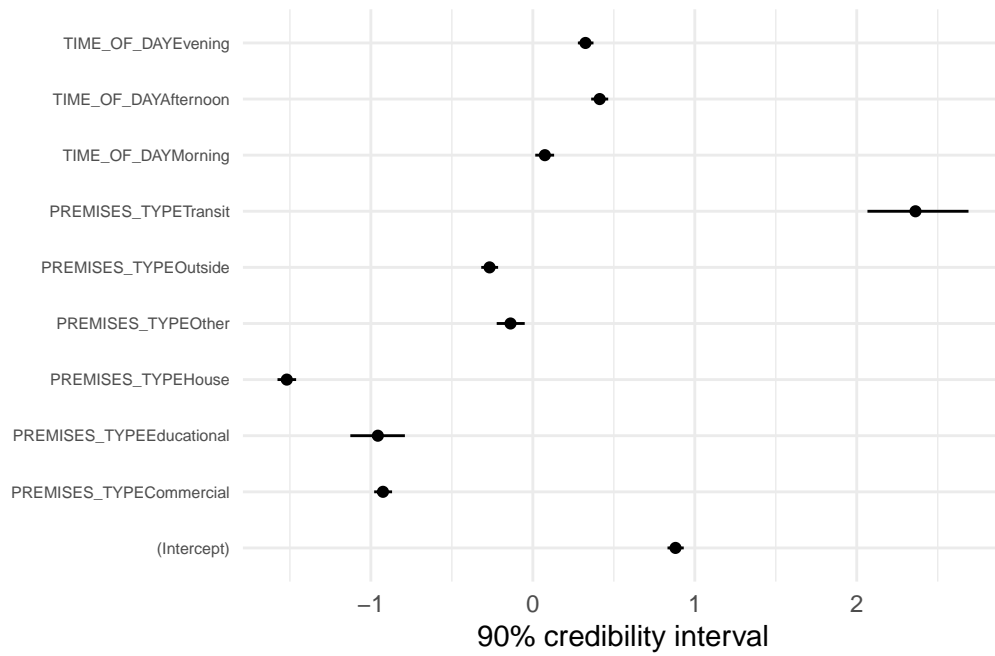


Figure 8: 90% credible intervals for predictors of violent crime likelihood

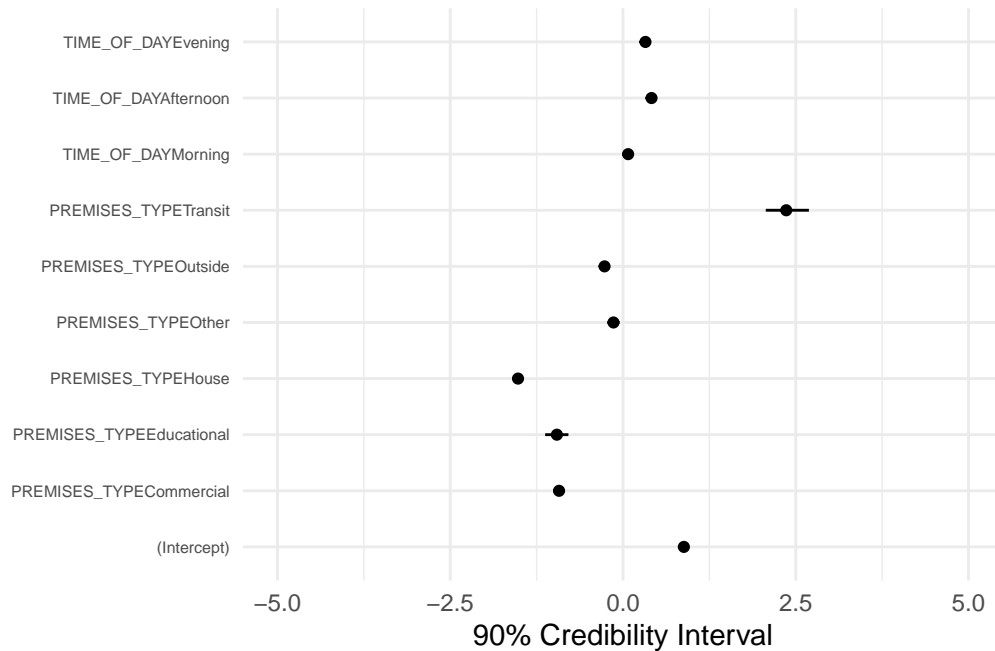


Figure 9: 90% credible intervals for predictors of violent crime likelihood with restricted x-axis limits

References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. <https://tellingstorieswithdata.com/>.
- Open Data Toronto. 2021. “Motor Vehicle Collisions Involving Killed or Seriously Injured Persons.” <https://open.toronto.ca/dataset/motor-vehicle-collisions-involving-killed-or-seriously-injured-persons/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.