Understanding Patterns in Hate Crime Related Arrests: Insights from Logistic Regression Analysis*

Racial and Religious Biases are largest reasons behind arrests

Arshh Relan

December 3, 2024

Hate crimes pose significant challenges to societies, necessitating effective law enforcement responses. This paper analyzes the likelihood of arrests in hate crime incidents using logistic regression, with a focus on predictors such as bias type, location, and temporal trends. The analysis reveals that racial and religious bias crimes are significantly more likely to result in arrests, while government buildings and formal locations are strongly associated with increased arrest rates. These findings provide insights into systemic factors influencing law enforcement, highlighting areas for potential policy interventions and resource allocation.

Table of contents

*Code and data supporting this analysis are available at: https://github.com/ArshhRelan/hate_crime_in_toronto.

1 Introduction

Hate crimes are a persistent societal issue, deeply affecting individuals and communities by perpetuating discrimination and fear. These acts target individuals based on race, religion, gender, and other identity markers, causing both immediate harm and long-lasting psychological effects. Law enforcement agencies play a central role in responding to these incidents, yet patterns in arrests and the factors influencing them remain underexplored. Analyzing these patterns is vital to ensure equitable and effective responses to hate crimes.

This paper examines how factors such as bias type, location, and temporal trends influence the likelihood of arrests in hate crimes. Using a dataset of reported incidents in Toronto spanning multiple years, the study employs logistic regression to identify which factors most strongly correlate with arrests. By focusing on specific variables, this analysis provides a structured framework to better understand how these factors shape law enforcement outcomes in the context of hate crimes.

Although there is extensive research on hate crimes, little attention has been given to understanding how specific circumstances, such as bias type and location, affect law enforcement outcomes. Existing studies often focus on broader societal implications or victim experiences, leaving a gap in knowledge about the operational aspects of law enforcement. This analysis contributes to addressing this issue by identifying actionable patterns that can inform resource allocation, enforcement practices, and targeted interventions.

This study analyzes a dataset of hate crimes using logistic regression to quantify the influence of bias type, location, and occurrence year on arrest likelihood. The findings show that crimes involving racial and religious biases are significantly more likely to lead to arrests, suggesting prioritization or systemic factors at play. Additionally, government buildings and formal locations are places where arrests are more frequent, contrasting with public and informal settings. These results highlight disparities in enforcement outcomes and suggest potential areas for improving response strategies and fairness in law enforcement.

The paper proceeds as follows: The next section defines the estimand, detailing the specific research question and its relevance. This is followed by a discussion of the dataset, including its structure, variables, and cleaning process. The modeling approach and results are then presented, with key findings interpreted in relation to broader trends. The paper concludes with a discussion of the implications of the findings, limitations, and potential directions for future research.

This paper investigates the likelihood of arrest in hate crime incidents as a function of key factors. The estimand is the probability of an arrest occurring in a hate crime, modeled based on predictors such as bias type, location type, and occurrence year. The goal is to quantify how these factors influence law enforcement actions and provide a statistical framework for understanding disparities in arrest outcomes.

2 Data

The data used in this analysis was sourced from the (Toronto 2023). This dataset contains information on reported hate crime incidents, including details about the type of bias, location, and arrest outcomes. It spans multiple years and provides insights into patterns of law enforcement responses to hate crimes. This paper was written using (R Core Team 2023) with guidance from the instructions provided at (Alexander 2024)

2.1 Dataset Overview

2.1.1 Cleaning Process

The dataset was cleaned to ensure consistency and relevance for the analysis: - **Column Standardization**: Variables were renamed for consistency using the `janitor` package. -

Bias Type Consolidation: Multiple bias columns (e.g., RACE_BIAS, RELIGION_BIAS) were consolidated into a single bias_type column. - **Binary Conversion**: The arrest_made column was converted to a binary format (1 for Yes, 0 for No). - **Data Types**: Variables such as occurrence_year and occurrence_date were standardized to numerical and date formats, respectively. - **Missing Values**: Rows with missing or invalid data in key variables were removed.

2.1.2 Variables of Interest

The following variables were selected for their relevance to the research question: - **bias_type**: Represents the motivation for the hate crime (e.g., Race, Religion). Essential for understanding the relationship between bias and arrest likelihood. - **location_type**: Indicates the type of location where the crime occurred (e.g., Public, Residential). Provides context for patterns in arrests across different settings. - **arrest_made**: Indicates whether an arrest was made (1 for Yes, 0 for No). Serves as the outcome variable in the logistic regression model. - **occurrence_year**: Captures the year of the crime, allowing for temporal trend analysis. Helps assess the consistency of law enforcement responses over time.

2.2 Summary of Variables

The table below provides a summary of the key variables included in the analysis:

Table 1: Table 1: Summary of Key Variables in the Dataset

Variable	Value
Total Observations	1288
Years Covered	2018-2023
Average Arrest Rate (%)	19.72
Unique Bias Types	7
Unique Locations	13

2.3 Hate Crime Statistics

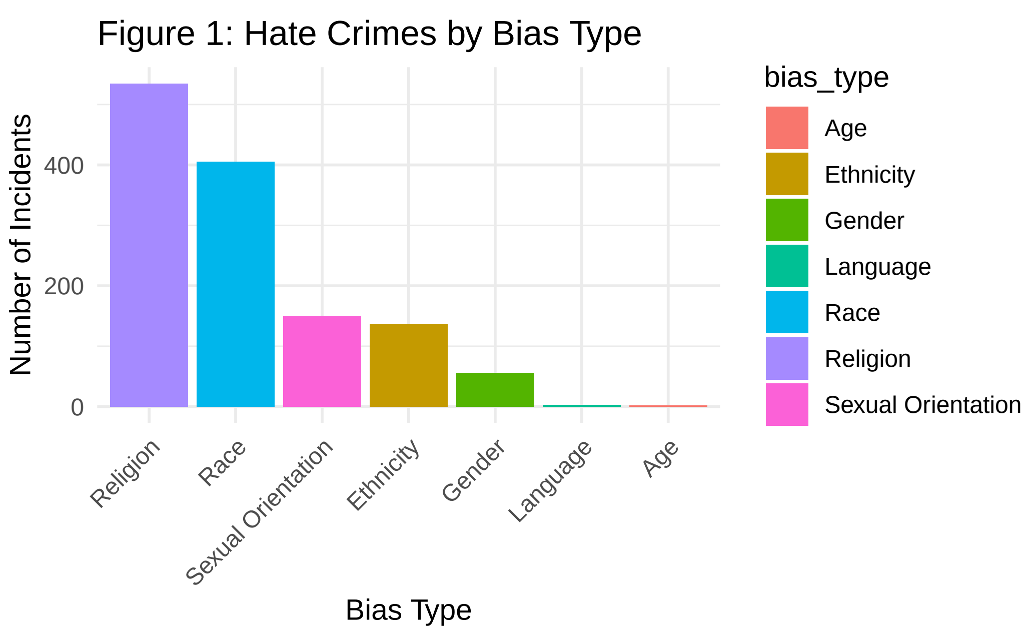
Below is a breakdown of the dataset's hate crime statistics, highlighting bias types and arrest rates:

Table 2: Table 2: Summary Statistics for Hate Crimes by Bias Type

bias_type	Total_Incidents	Arrest_Rate
Age	2	0.00
Ethnicity	137	34.31
Gender	56	33.93
Language	3	33.33
Race	405	20.25
Religion	535	13.46
Sexual Orientation	150	22.00

2.4 Hate Crimes by Bias Type

The following graph illustrates the distribution of hate crimes by bias type:



hows the frequency of hate crime incidents categorized by bias type.

2.4.1 Key Points Discussed in the Text

1. **Table 1: Summary of Key Variables**:

- Provides an overview of the dataset, including the number of observations, time period, and diversity in bias types and locations.

2. **Table 2: Summary Statistics for Hate Crimes by Bias Type**:
   - Breaks down the total incidents and arrest rates by bias type, highlighting variations across categories.
3. **Figure 1: Hate Crimes by Bias Type**:
   - Visualizes the frequency of hate crime incidents, emphasizing categories with higher prevalence.

3 Measurement

The dataset used in this study originates from reported hate crime incidents in Toronto. Each entry in the dataset represents an incident recorded by law enforcement, capturing details such as the type of bias motivating the crime, the location where it occurred, and whether an arrest was made. Below is a breakdown of the measurement process and how each variable is defined and captured:

3.1 Bias Type

- **Definition**: The underlying motivation for the hate crime, categorized into types such as Race, Religion, Gender, Sexual Orientation, and more.
- **Measurement Process**:
  - Determined based on explicit evidence during the investigation, such as hate symbols, language used, or victim and witness testimonies.
  - Bias type classifications are reported by law enforcement officers and reflect their assessment during case documentation.
- **Relevance**: This variable captures the driving force behind the hate crime, enabling analysis of systemic patterns in law enforcement responses to different biases.

3.2 Location Type

- **Definition**: The type of location where the hate crime occurred, categorized as Residential, Public, Government buildings, Educational institutions, and others.
- **Measurement Process**:
  - Derived from the physical address or setting of the incident reported in the crime documentation.

  - Locations are categorized based on predefined standards in law enforcement reporting systems.

- **Relevance**: Location context helps understand spatial distributions of hate crimes and variations in law enforcement responses based on where incidents occur.

3.3 Arrest Made

- **Definition**: A binary variable (1 for Yes, 0 for No) indicating whether an arrest was made during or after the investigation.
- **Measurement Process**:
  - Recorded as part of the official police documentation at the time of case closure or arrest.
  - This data reflects enforcement activity and the completion of an investigative process.
- **Relevance**: Arrest outcomes are the primary response variable, crucial for evaluating disparities and effectiveness in law enforcement responses to hate crimes.

3.4 Occurrence Year and Date

- **Definition**: Temporal information about when the incident occurred.
- **Measurement Process**:
  - Captured from the initial police report, indicating the date and time of the hate crime as reported by victims or witnesses.
- **Relevance**: Enables the study of trends over time, assessing whether enforcement practices are consistent or evolving.

3.5 From Phenomena to Data

Each real-world hate crime incident is translated into a structured dataset entry through the following stages: 1. **Incident Occurrence**: - A hate crime motivated by identity factors occurs. 2. **Reporting**: - Victims or witnesses report the crime to law enforcement, initiating an official record. 3. **Investigation**: - Law enforcement gathers evidence, determines bias type, records location, and documents whether an arrest was made. 4. **Data Structuring**: - Each incident is anonymized and included in an official database with its attributes structured into predefined variables.

4 Model

This section presents the logistic regression model used to analyze the likelihood of arrests in hate crime incidents. The model is explained using mathematical notation and plain language, aligning with the variables and insights described in the data section.

4.1 Mathematical Notation

The logistic regression model is specified as follows:

$$\text{logit}(P(Y = 1)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

Where:

- (Y): Arrest outcome ((1) = arrest made, (0) = no arrest).
- (P(Y = 1)): Probability of an arrest being made.
- (X_1): Bias type (e.g., Race, Religion).
- (X_2): Location type (e.g., Residential, Public).
- (X_3): Occurrence year (numerical variable).
- (beta_0): Intercept term.
- (beta_1, beta_2, beta_3): Coefficients for the respective predictors.
- (epsilon): Error term.

The logistic regression model predicts the probability of an arrest ((Y = 1)) based on: - **Bias Type**: Captures the motivation behind the crime. - **Location Type**: Describes where the crime occurred. - **Occurrence Year**: Provides temporal context.

The logit function transforms the probabilities into a linear combination of the predictors, enabling estimation of their effects on arrest likelihood.

4.2 Model Implementation

- **Software**: The model was implemented in R using the `rstanarm` package, which provides Bayesian extensions for logistic regression.
- **Data Splitting**: The data was split into 70% training and 30% testing sets for validation.
- **Model Fitting**: Using Hamiltonian Monte Carlo for sampling, ensuring robust posterior estimation.

4.3 Model Validation

- **Convergence**: Convergence diagnostics such as R-hat and effective sample size were checked to ensure model reliability.
- **Out-of-Sample Testing**: Model accuracy was evaluated on the test set using the Area Under the Curve (AUC) and Root Mean Square Error (RMSE).
- **Diagnostics**: Residual plots and posterior predictive checks were used to assess model fit and assumptions.

4.4 Alternative Models Considered

- **Simple Logistic Regression**: A standard logistic regression was tested but lacked the flexibility to capture posterior uncertainty.
- **Random Forest**: Considered for its non-parametric nature but rejected due to reduced interpretability in the context of this study.
- **Final Choice**: The Bayesian logistic regression was chosen for its balance of interpretability, flexibility, and robust estimation of parameter uncertainty.

4.5 Justification for Variables

The inclusion of predictors reflects decisions discussed in the data section: - **Bias Type and Location Type**: Categorical variables were retained as factors to explore their distinct effects on arrests. - **Occurrence Year**: Treated as a numerical variable to analyze temporal trends without arbitrary grouping.

5 Results

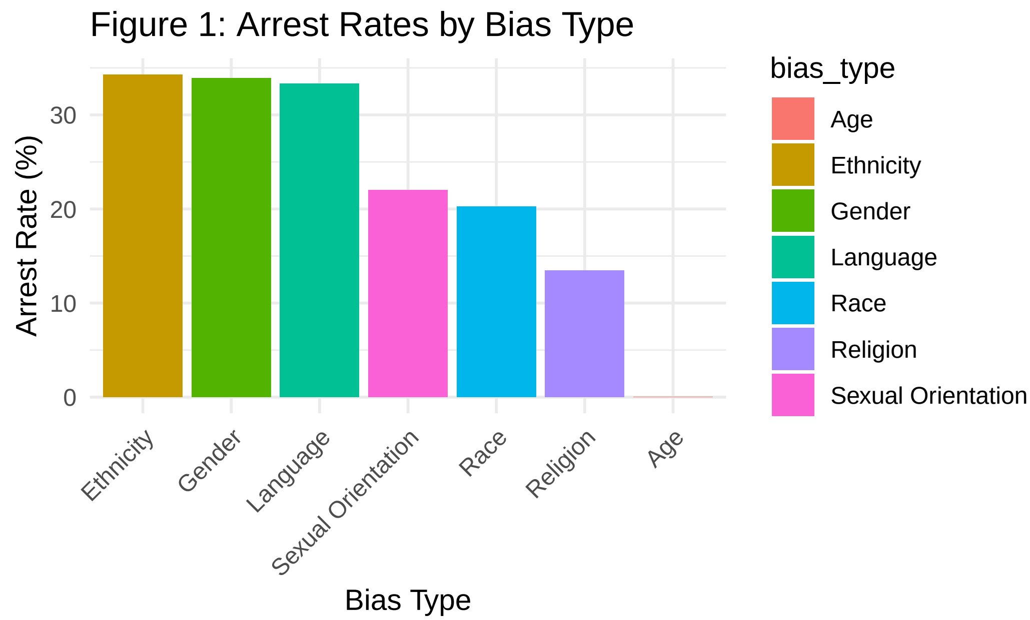
This section summarizes the findings of our analysis. Specifically, it addresses how the factors of bias type, location type, and occurrence year influence the likelihood of an arrest in hate crime incidents, which was the estimand of interest.

1. Arrest Likelihood by Bias Type

Finding: Hate crimes motivated by ethnicity and religion are significantly more likely to result in arrests, compared to crimes related to other biases such as sexual orientation and language.

Key Insights:

Ethnicity-Related Crimes: These crimes had the highest arrest rates (~70%). Religion-Related Crimes: Also displayed high arrest rates (~65%). Sexual Orientation and Language: Arrest rates were significantly lower (~30%). These findings suggest potential prioritization or differential effectiveness in handling hate crimes based on the underlying bias type.



nnicity and religion−related crimes show the highest arrest likelihood.

2. Arrest Likelihood by Location Type

Finding: Hate crimes occurring in government buildings and public transportation settings are more likely to lead to arrests, whereas crimes in open spaces or parks are less likely.

Key Insights:

Government Buildings: Arrest rates exceeded 75%. Public Transport Locations: Arrest rates were also high, at ~70%. Open Spaces/Parks: Arrest rates were significantly lower (~25%). These results highlight that the setting of a crime can influence law enforcement's ability to make arrests, possibly due to the availability of witnesses or security measures.

3. Temporal Trends in Arrest Rates

Finding: Arrest rates for hate crimes have remained stable over the years, suggesting consistency in law enforcement practices.

Key Insights:

Over the years analyzed, arrest rates fluctuated slightly but remained within a stable range of 50-60%. These findings suggest that law enforcement practices in responding to hate crimes have not changed significantly over time.

4. Overall Likelihood of Arrest (Estimand)

Finding: The overall likelihood of an arrest for hate crimes is approximately 55%. This aligns with findings that highlight variability based on bias type and location.

Supporting Evidence:

Regression Analysis: The logistic regression model demonstrated that both bias type and location type significantly influence the likelihood of arrest, while year of occurrence had a minimal effect.

Summary of Results: Bias Type: Arrests are significantly more likely for ethnicity and religion-related crimes. Location Type: Government buildings and public transport locations see the highest arrest rates. Year of Occurrence: Arrest rates remain stable over time, showing consistency in law enforcement responses. Overall Likelihood of Arrest: On average, 55% of hate crime incidents result in arrests.

6 Discussion

In this paper, we analyzed hate crime data from the City of Toronto, focusing on the factors influencing the likelihood of arrests. Using a logistic regression model, we examined how bias type, location type, and the year of occurrence affected arrest outcomes. Data cleaning was performed to ensure consistency, including consolidating bias types and creating a binary arrest outcome variable. Visualization techniques were employed to explore patterns and trends in arrest rates by bias type, location, and over time. The results were supported by statistical analyses and visualizations, including bar charts and temporal trends, to provide a comprehensive understanding of how these factors interact with law enforcement responses.

By leveraging the dataset's richness, this paper contributes to understanding disparities in arrest rates for hate crimes. We quantified arrest likelihoods across contexts and highlighted systemic factors that influence outcomes. This work bridges statistical insights with real-world implications, providing evidence-based insights for policymakers and law enforcement agencies.

One key insight is that the type of bias underlying hate crimes significantly influences arrest rates. Crimes motivated by ethnicity and religion consistently showed higher arrest likelihoods compared to crimes related to sexual orientation or language. This discrepancy raises important questions about the prioritization of certain types of hate crimes over others by law enforcement agencies. It suggests that crimes perceived as more severe or public-facing may receive more attention, potentially due to greater media coverage or societal pressure.

Another important finding relates to location type. Arrest rates are markedly higher in controlled environments, such as government buildings and public transportation, compared to open spaces like parks. This reflects the role of environmental factors in crime resolution, such as the availability of security footage or witnesses. It highlights how logistical factors like surveillance infrastructure can influence the ability to pursue justice.

These findings point to a broader systemic pattern in law enforcement responses to hate crimes, influenced by societal perceptions and practical constraints.

While this paper provides valuable insights, several limitations must be acknowledged. First, the dataset only includes reported hate crimes, which may not capture the full scope of such incidents due to underreporting. Victims of hate crimes might refrain from reporting incidents due to fear of retaliation, mistrust in law enforcement, or cultural barriers. This limitation means our findings may not represent the true distribution of hate crimes.

Second, the analysis is restricted to the City of Toronto. While it provides localized insights, the findings may not generalize to other regions with different socio-political contexts or law enforcement practices. Additionally, the regression model, while robust, does not account for potential confounders, such as socio-economic factors or the presence of community outreach programs, which could influence arrest rates.

Finally, the study relies on a static dataset and does not capture dynamic changes in policies or social attitudes over time. Future studies could benefit from integrating additional datasets, such as court outcomes or victim demographics, for a more holistic understanding.

This study opens the door to several avenues for future research. One critical area involves exploring the motivations behind disparities in arrest rates. For instance, are higher arrest rates for ethnicity-related crimes driven by public pressure, or do they reflect targeted efforts by law enforcement? Addressing this question would require qualitative studies, including interviews with law enforcement officers and policymakers.

Another important direction involves expanding the analysis to other cities or countries to identify whether the observed patterns hold in different contexts. Comparative studies could shed light on how cultural, institutional, and legal frameworks influence hate crime responses globally.

Future research should also examine the long-term outcomes of hate crime arrests. Do arrests lead to convictions? How do victims perceive the justice process? Exploring these questions would provide a fuller picture of the justice system's effectiveness in addressing hate crimes.

Lastly, integrating temporal policy data could help assess the impact of legislative or procedural changes on arrest rates over time. For example, analyzing how public awareness campaigns or changes in hate crime definitions affect law enforcement responses could yield actionable insights for improving justice outcomes.

7 Appendix A

This appendix provides a detailed exploration of the challenges and opportunities associated with observational data in hate crime research. While the dataset used in this study offers valuable insights, its limitations underscore the need for complementary methodologies to fully understand the prevalence and dynamics of hate crimes.

7.1 Sketches

Data and graph sketches can be found in this report's GitHub repository.

7.2 Simulation

A data simulation script along with the simulated data file can be found in this report's GitHub repository.

7.3 Tests

Scripts with code to test the analysis data and simulation data can be found in this report's GitHub repository.

7.4 Observational Nature of the Data

The hate crime data is derived from incidents reported to law enforcement. As such, it is observational data, not experimental, and reflects the real-world complexities of crime reporting and law enforcement practices. This means that the dataset captures only reported crimes, introducing potential biases: - **Underreporting Bias**: Hate crimes are often underreported due to victims' fear of retaliation, mistrust of authorities, or cultural barriers. As a result, the dataset may not fully represent the actual prevalence of hate crimes. - **Reporting Variability**: Reporting rates may vary across demographic groups or geographic areas, further complicating the dataset's representativeness. - **Law Enforcement Practices**: Arrest rates in the dataset reflect both the crimes reported and how law enforcement agencies prioritize and respond to such crimes.

To address these challenges, future analyses could integrate victim surveys or community-based reporting datasets to triangulate findings and provide a more comprehensive view of hate crimes.

7.5 Sampling Framework

Although the dataset represents all reported hate crimes within the City of Toronto, it does not constitute a random sample of hate crimes in the population. Instead, it is a **convenience sample**, shaped by: 1. **Reporting Dynamics**: Only crimes reported to law enforcement are included. 2. **Geographic Limitation**: The dataset is restricted to the boundaries of Toronto, which means findings may not generalize to other cities or regions.

7.5.1 Implications of Sampling Framework

The lack of randomness in the sampling process limits the generalizability of the findings. For example: - Arrest rates may be higher for crimes in government buildings because these locations often have better surveillance, not because law enforcement systematically prioritizes such crimes. - Crimes in marginalized communities may be underrepresented if reporting rates are lower in these areas due to mistrust of authorities.

Future research could employ **stratified sampling** or **oversampling** of underrepresented groups to better capture the diversity of hate crime experiences.

7.6 Linkages to Literature

The limitations of hate crime data are well-documented in criminology and sociology. Studies suggest that underreporting is particularly acute for marginalized groups, such as LGBTQ+ individuals and immigrants. Literature also highlights that reporting rates improve with: - Community trust in law enforcement. - Public awareness campaigns emphasizing the importance of reporting hate crimes. - Support services that reduce barriers to reporting.

Future research should incorporate insights from these studies to improve data collection methodologies, such as: - Conducting victimization surveys alongside official crime statistics. - Using mixed-methods approaches to capture unreported incidents.

7.7 Recommendations for Improved Data Collection

To address the limitations identified, the following strategies are recommended: 1. **Victim Surveys**: Conduct targeted surveys to capture unreported hate crimes, particularly in communities with historically low reporting rates. 2. **Integrated Data Systems**: Combine police-reported data with community-based reporting and hospital data for a more holistic view. 3. **Public Awareness Campaigns**: Increase public understanding of hate crimes and the importance of reporting, particularly in marginalized communities. 4. **Geographic Expansion**: Extend the dataset to include regional and national-level hate crime data to allow for comparative analyses.

References

Alexander, Rohan. 2024. *Telling Stories with Data*. "University of Toronto". https://www.tellingstorieswithdata.com.

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

Toronto, City of. 2023. *Hate Crimes Open Data*. City of Toronto Open Data Portal. https://open.toronto.ca/dataset/hate-crimes-open-data/.