

Understanding the Risk of Casualties in US School Shootings*

Leveraging Bayesian Analysis to Guide Safety and Prevention Strategies

Yun Chu

December 6, 2024

School shootings in the United States are a persistent and devastating issue. This paper utilizes a Bayesian Logistic Regression model to predict the probability of casualties in school shooting incidents, using historical data from 1999 onward. Key predictors, including shooting type and geographic latitude, are identified and analyzed. The results demonstrate that targeted and indiscriminate shootings significantly increase casualty risks, highlighting the need for targeted safety measures, such as enhanced threat assessment programs and stricter firearm regulations. This analysis provides a framework for implementing data-driven strategies to reduce casualties and improve school safety nationwide.

Table of contents

1	Introduction	3
2	Data	3
2.1	Overview	3
2.2	Measurement	4
2.3	Summary Statistics & Relationship Between Variables	4
2.4	Outcome Variables	5
2.5	Predictor Variables	6
3	Model	7
3.1	Model Set-Up	7
3.1.1	Mathematical Formulation	8
3.1.2	Model Justification	9

*Code and data are available at: <https://github.com/chuyun2024/School-Shooting-Analysis>.

3.2	Model Limitations	9
3.3	Model Validation	10
3.4	Alternative Models Considered	10
4	Results	10
4.1	Overview	10
4.2	Intercept	10
4.3	Shooting Type	11
4.4	Latitude	11
4.5	Predicted Existence of Casualties by State	11
5	Discussion	12
5.1	Policy Implications	12
5.1.1	Geographic Targeting of Resources	12
5.1.2	Focus on High-Risk Shooting Types	12
5.1.3	Strengthening Gun Control Measures	13
5.1.4	Investment in Mental Health Resources	13
5.1.5	Ongoing Policy Evaluation	13
5.2	Limitations of the Analysis	13
5.3	Recommendations for Future Research	14
5.3.1	Expanding the Predictive Framework	14
5.3.2	Evaluating Policy Effectiveness	14
5.3.3	Encouraging Interdisciplinary Collaboration	14
A	Appendix	15
A.1	Analysis of The Washington Post’s Data Collection Approach	15
A.1.1	Data Collection and Triangulation	15
A.1.2	Defining Inclusion and Exclusion Criteria	15
A.1.3	Sampling and Estimation Techniques	15
A.1.4	Reproducibility and Transparency	16
A.1.5	Addressing Potential Biases	16
A.1.6	Weaknesses	16
A.1.7	Conclusion	16
A.2	Model details	17
A.3	Posterior predictive check	17
A.4	Diagnostics	17
	References	18

1 Introduction

School shootings in the United States remain a deeply concerning issue, resulting in devastating consequences for students, educators, and communities. Despite significant public discourse and legislative efforts, effective strategies to predict and prevent such incidents remain underdeveloped (RAND Corporation 2015). Understanding the factors contributing to school shooting casualties is critical for designing evidence-based policies that address the root causes of this violence.

The estimand in this paper is the casualty in a school shooting incident in US. Using data from The Washington Post (The Washington Post 2024b), this paper seeks to address this gap by employing a Bayesian Logistic Regression model to analyze the probability of casualties in school shooting incidents. Using historical data from 1999 onward, the analysis focuses on key predictors, such as shooting type and geographic latitude, to uncover patterns influencing casualty risks. Unlike previous research, this study emphasizes predictive modeling to offer actionable recommendations for policy and intervention strategies.

The findings indicate that targeted and indiscriminate shootings notably raise the probability of casualties, while geographic latitude offers further understanding of regional patterns. This analysis underscores the importance of tailored interventions, including proactive threat assessment programs and stricter firearm regulations, to mitigate the risks associated with high-casualty incidents.

This study is significant because it provides policymakers with a data-driven framework to allocate resources effectively and prioritize safety measures in high-risk areas. By utilizing data analysis to help with policy implementation, the research aims to contribute to the reduction of school shooting casualties and foster safer educational environments.

The structure of this paper is as follows: Section 2 discusses the data, including its sources, measurements, and key variables; Section 3 outlines the model used for prediction, including its justification and limitation, model validation is included in Section A; Section 4 presents the results, including summary statistics and spatial visualizations; and Section 5 concludes with a discussion of the findings, limitations, and recommendations for future research and policy.

2 Data

2.1 Overview

The dataset has 416 entries, with each entry representing a unique school shooting incident. Incidents occurring during after-hours events, accidental gun discharges that only injured the individual handling the firearm, and private suicides that did not endanger other children were excluded from consideration. Additionally, shootings at colleges and universities, which involve

young adults rather than children, were not included in the analysis (The Washington Post 2024b). These entries cover 50 variables that provide information about the schools and its students, locations, date and time of shooting, shooters details, number of killed and injured, and the relationship of the shooter to school, the weapon type and source.

Since the federal government does not consistently track school shootings, this dataset from The Washington Post provides detailed information on these incidents.. Although sources like FBI crime reports and local school incident logs were reviewed, they lack the detail and comprehensive coverage of this dataset. Its unparalleled breadth and depth make it the strongest foundation for predictive modeling and generating actionable insights.

The statistical programming language R (R Core Team 2023) is used to download, clean, analyze and model the US School Shooting Data. The US School Shooting dataset is downloaded from The Washington Post (The Washington Post 2024a). The following libraries in R are utilized in this paper: `tidyverse` (Hadley Wickham et al. 2024e), `dplyr` (Hadley Wickham et al. 2024a), `lubridate` (Garrett Golemund and Hadley Wickham 2024), `readr` (Hadley Wickham et al. 2024b), `stringr` (Hadley Wickham et al. 2024c), `arrow` (Neal Richardson et al. 2024), `testthat` (Hadley Wickham et al. 2024d), `modelsummary` (Vincent Arel-Bundock 2023), `ggplot2` (Hadley Wickham 2016), `maps` (Ray Brownrigg et al. 2023), `knitr` (Yihui Xie 2023), `here` (Kirill Müller 2023), `kableExtra` (Hao Zhu 2023), `rstanarm` (Gabriel A. Fonseca, Jonah Gabry, Ben Goodrich, Sam Brilleman, Ben Lambert, and Andrew Gelman 2023), `arrow` (Project 2024) and `patchwork` (Pedersen 2020).

2.2 Measurement

The Washington Post spent a year determining how many children have been affected by school shootings, beyond just those killed or injured. The dataset, compiled by The Washington Post, translates real-world school shooting incidents into structured entries by aggregating information from news articles, open-source databases, law enforcement reports, and direct calls to schools. Only verified incidents, such as shootings during school hours or on school property, were included. Events like after-hours shootings, private suicides, or accidental discharges without other injuries were excluded (The Washington Post 2024b).

2.3 Summary Statistics & Relationship Between Variables

In this dataset, there are three variables that have relationship: $casualties = killed + knjured$.

Table 1 summarizes the mean, median and standard deviation of casualties from school shooting events in US from 1999 till now. The standard deviation of 3.72 indicates that while most incidents have casualties close to the mean, there is a wide range of variability, with some incidents having significantly higher number of casualties.

Table 1: Summary Statistics for Casualties

Mean	Median	Standard Deviation
1.6	1	3.72

Figure 1 illustrates the distribution of casualties across different shooting types. “Targeted” and “Indiscriminate” categories exhibit significantly higher casualty counts compared to other types. As a result, these two types are combined into one class, while the remaining types form a separate class.

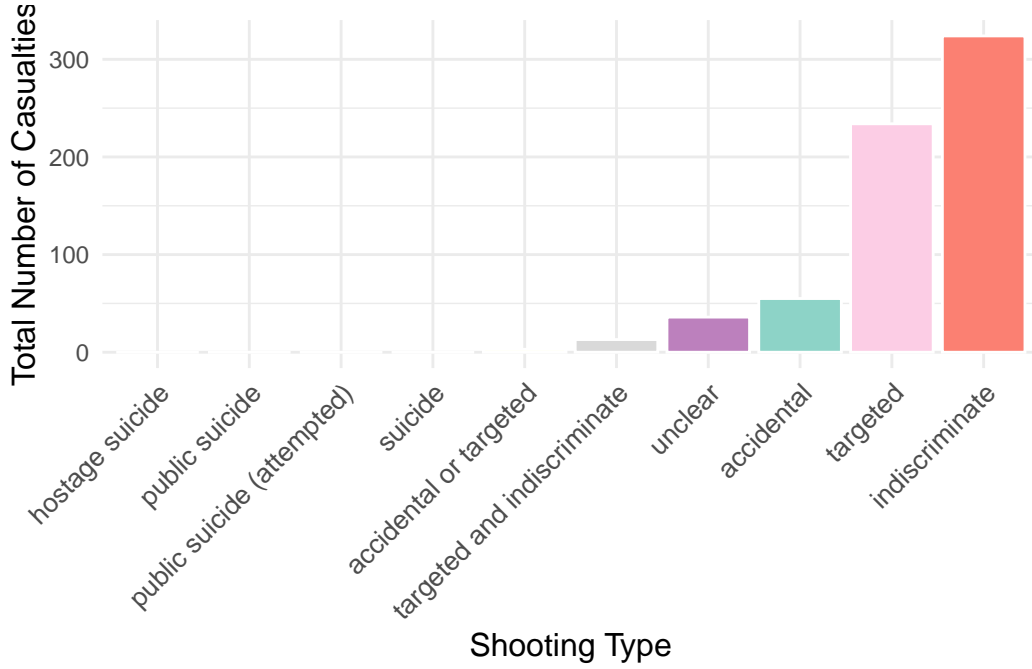


Figure 1: Number of Casualties for Each Shooting Type in US School Shooting Incidents

2.4 Outcome Variables

The outcome variable for this analysis is casualties for a School Shooting Event. In the model, we classify the existence of casualties.

Figure 2 illustrates the distribution of casualties, highlighting a positively skewed pattern. The majority of incidents involve fewer than seven casualties, though a few school shootings account for exceptionally high casualty numbers.

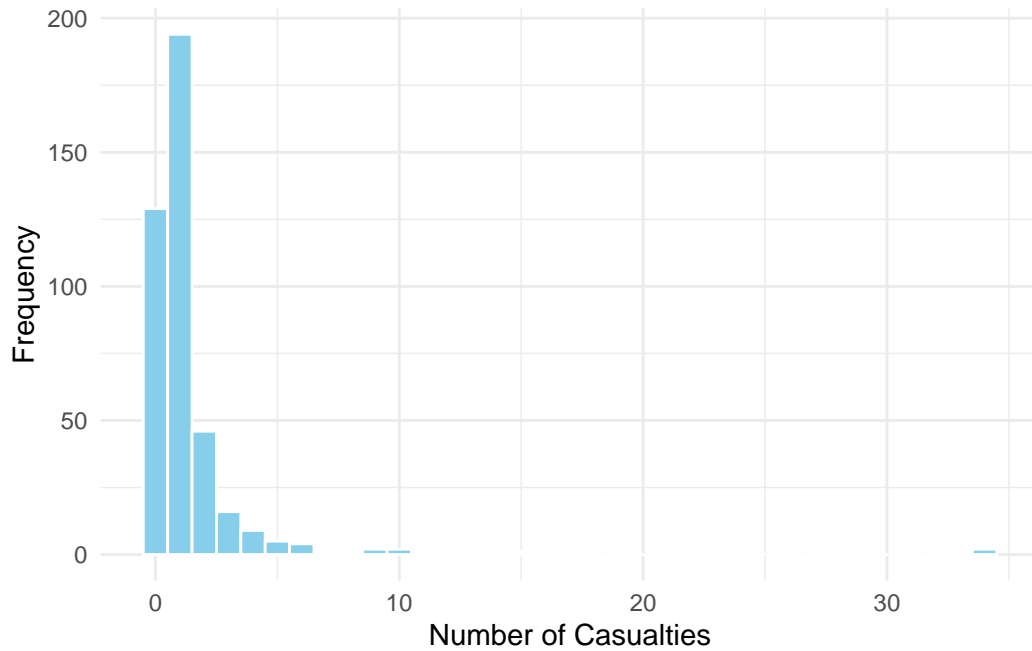


Figure 2: Distribution of Casualties in US School Shooting Incidents from 1999 to 2024

Figure 3 visualizes the number of school shooting incidents in US by state for all data since 1999. California, Texas, Florida and North Carolina all have more than twenty school shooting incidents in the past twenty-five years while other states have less than twenty school shootings.

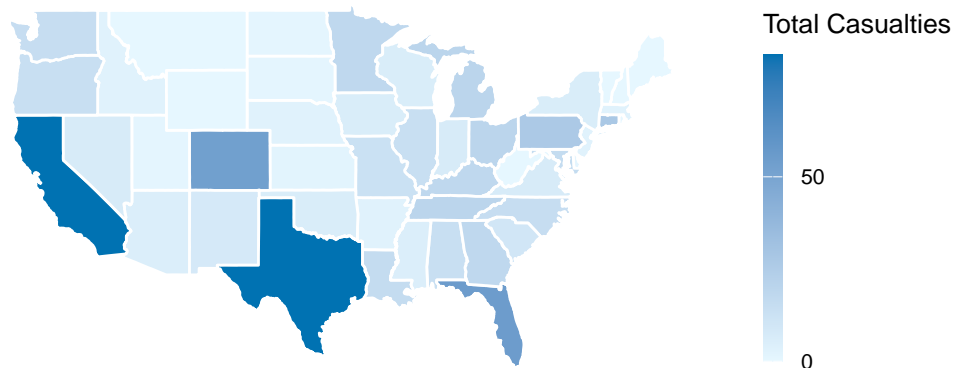


Figure 3: US School Shooting Casualties by State from 1999 to 2024

2.5 Predictor Variables

Predictor variables in this analysis is Shooting Category and Latitude.

- Shooting Category is a binary variable derived from Shooting Type with “Targeted” and “Indiscriminate” as one type while the rest is classified as the second type. Shooting Type represents the type of shootings with 5 categories: accidental, hostage suicide, indiscriminate, public suicide, targeted, and unclear. There are also some rows with a combination of these 5 types, for example, “accidental or targeted” and “targeted and indiscriminate”.
- Latitude represents the latitude of the school.

Figure 4 presents a pie chart depicting the proportions of various shooting types. Targeted shootings account for 56% of all school shooting incidents.

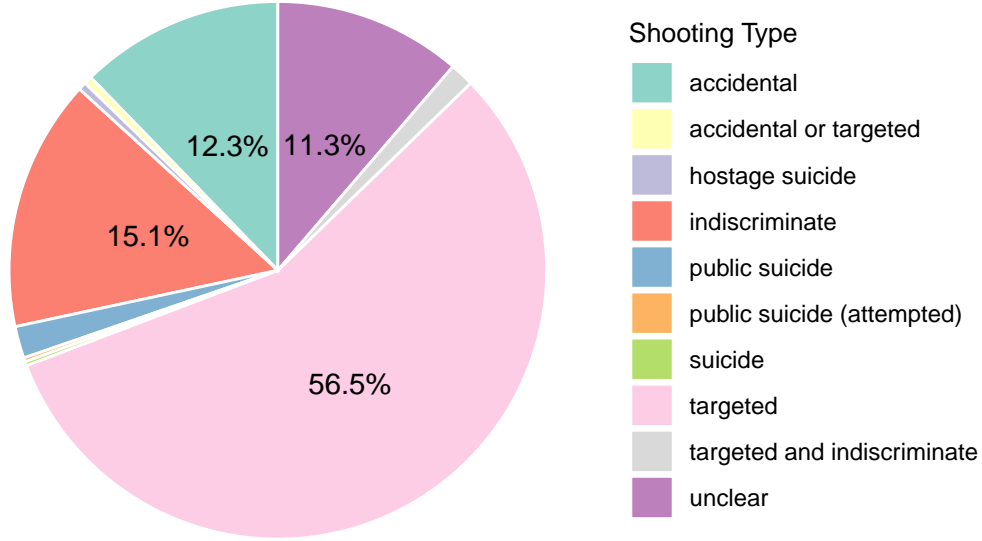


Figure 4: US School Shooting Type Distribution from 1999 to 2024

Figure 5 shows the distribution of latitude values of schools in school shooting events, ranging from approximately twenty to fifty degrees. Most events occur between thirty and forty degrees, with a peak near thirty-five degrees. There are few occurrences at higher latitudes (above forty-five degrees).

3 Model

3.1 Model Set-Up

The analysis employs a Bayesian Logistic Regression Model to predict whether a school shooting incident results in casualties. The binary response variable (`ifcasualty`) is defined as 1 if there are any casualties (injuries or fatalities) and 0 otherwise. The predictors include the **latitude** of the school where the incident occurs and a simplified categorical variable, **shooting**

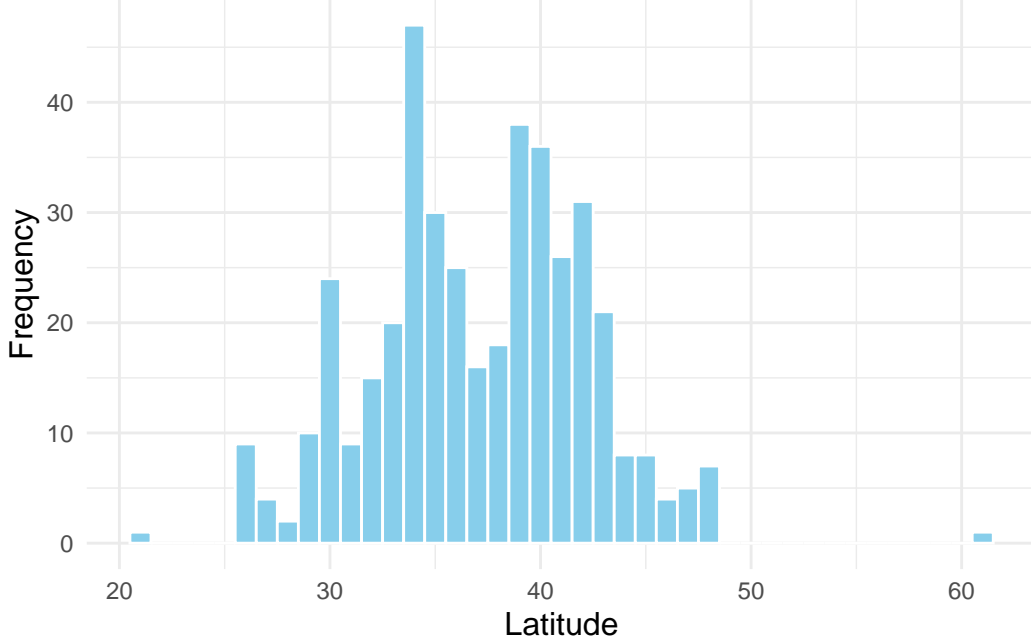


Figure 5: Distribution of Latitude of Schools in School Shooting Incidents in US from 1999 to 2024

category, which distinguishes between “indiscriminate” or “targeted” shootings and “other” types.

3.1.1 Mathematical Formulation

The model estimates the probability of a casualty using the following formulation:

$$P(C = 1 | X) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}$$

$$\beta_0 \sim \text{Normal}(0, 5)$$

$$\beta_1, \beta_2 \sim \text{Normal}(0, 2.5)$$

where:

- $P(C = 1 | X)$: Probability of at least one casualty occurring in an incident.
- β_0 : Intercept term.
- β_1 : Coefficient for the shooting category, capturing the difference between “indiscriminate or targeted” and “other” types.
- β_2 : Coefficient for latitude, accounting for geographic trends.

- X_1 : Shooting category (0 = “indiscriminate or targeted”, 1 = “other”).
- X_2 : Latitude of the school where the incident occurs (continuous).

The priors for the Bayesian logistic regression model are specified as follows:

- **Intercept** (β_0): Reflects uncertainty about the baseline log-odds of casualties.
- **Coefficients** (β_1, β_2): Weakly informative priors that regularize the model and prevent overfitting.

These priors were chosen to represent plausible parameter values while maintaining model stability.

The Bayesian logistic regression model is implemented in R using the `rstanarm` package. The final model is saved as an RDS file for reproducibility and further analysis.

3.1.2 Model Justification

The Bayesian logistic regression model is well-suited for this problem as it is specifically designed to predict binary outcomes, such as whether casualties occur, without requiring transformation of skewed response variables. Additionally, the Bayesian framework provides probabilistic predictions, allows for the incorporation of prior knowledge, and offers a robust quantification of uncertainty in parameter estimates.

The shooting types “targeted” and “indiscriminate” were combined into a binary variable alongside other shooting types, as exploratory data analysis in Figure 1 illustrated that these two categories are strongly associated with a higher number of casualties.

Latitude is included as a predictor to account for geographic variability that might affect casualty likelihood. Longitude was evaluated during model selection but excluded due to its lack of statistical significance.

3.2 Model Limitations

The predictive performance of the Bayesian logistic regression model may be limited under certain circumstances. One key limitation is that the model’s predictions might not generalize well if there are significant changes in societal or policy conditions, as these factors could alter the relationships between predictors and the outcome. Additionally, the model’s reliability is sensitive to the quality of both the prior distributions and the dataset. Poorly chosen priors or biased and incomplete data can lead to inaccurate predictions and reduced robustness of the model.

3.3 Model Validation

The Bayesian logistic regression model is validated through several diagnostics. In Figure 7a, the posterior predictive check shows alignment between the posterior predictive distribution (y_{rep}) and the observed data (y), indicating a good fit. Figure 7b shows that posterior estimates are narrower and shifted compared to priors, reflecting the influence of the data. Additionally, Figure 8a confirms well-mixed and stationary MCMC chains, while Figure 8b shows $\hat{R} \approx 1.00$ for all parameters, verifying convergence and reliability.

3.4 Alternative Models Considered

Several models were evaluated during the development process:

1. **Linear Logistic Regression:**

- **Strengths:** Simplicity and interpretability.
- **Weaknesses:** Cannot incorporate prior knowledge or provide uncertainty quantification.

2. **Random Forest:**

- **Strengths:** Handles complex, non-linear interactions.
- **Weaknesses:** Limited interpretability and probabilistic outputs.

The Bayesian logistic regression model was chosen for its ability to provide interpretable, probabilistic predictions while quantifying uncertainty and incorporating prior information.

4 Results

4.1 Overview

The model results are summarized in Table 2.

4.2 Intercept

The intercept estimate is 2.37 (95% credible interval: 0.82 to 4.02), representing the baseline log-odds of a casualty occurring when predictors are at their reference levels. This corresponds to a high baseline probability, with the credible interval excluding zero, indicating strong evidence for a nonzero probability of casualties under these conditions.

Table 2: Explanatory model of the Existence of Casualty in Each School Shootig Event based on Shooting Type and Latitude

	Casualty Occurrence
(Intercept)	2.37
	(0.82, 4.02)
shooting_categoryother	-0.63
	(-1.09, -0.16)
lat	-0.04
	(-0.08, 0.00)
Num.Obs.	415
ELPD	-255.3
ELPD s.e.	8.1
LOOIC	510.7
LOOIC s.e.	16.2
WAIC	510.7

4.3 Shooting Type

The coefficient for the “other” shooting category is -0.63 (95% credible interval: -1.09 to -0.16), suggesting a lower likelihood of casualties compared to the “indiscriminate” or “targeted” categories.

4.4 Latitude

The latitude coefficient is -0.04 (95% credible interval: -0.08 to 0.00), indicating a slight but significant decrease in the probability of casualties with increasing latitude. This suggests potential geographic trends influencing casualty outcomes, such as differences in population density, response capabilities, or regional policies.

4.5 Predicted Existence of Casualties by State

Using a Bayesian Logistic regression model, predictions for the existence of casualties in each state were generated. The predicted casualties were visualized geographically using a choropleth map in Figure 6 where the states in dark red are classified as casualty exists while the states in grey are classified as no casualty.

Predicted Existence of Casualties Visualization

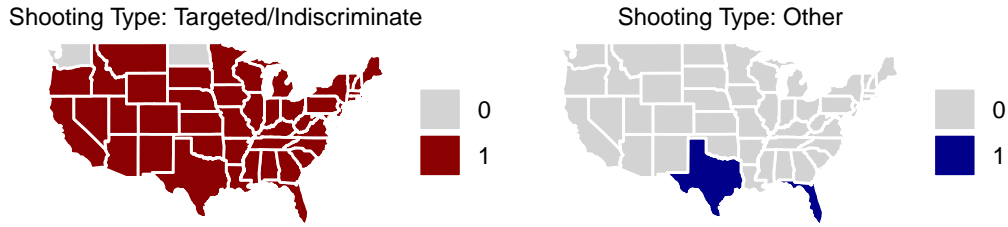


Figure 6: Prediction of the Existence of Casualty in US School Shootings Incidents Based on Bayesian Logistic Regression Model

5 Discussion

This paper utilizes a Bayesian Logistic Regression model to predict the likelihood of casualties in school shootings, focusing on the impact of shooting type and geographic latitude. The findings demonstrate that “targeted” and “indiscriminate” shootings significantly increase casualty probabilities, necessitating their classification into a unified predictive category. This categorization improves the model’s predictive accuracy and highlights the distinct risk these shooting types pose. Furthermore, the model generates spatial predictions at the state level, identifying regions with elevated casualty risks. These predictions provide policymakers with concrete data to prioritize interventions and allocate resources effectively, especially in high-risk areas.

5.1 Policy Implications

5.1.1 Geographic Targeting of Resources

The geographic analysis in this study shows significant disparities in casualty risks, with southern states showing higher probabilities of school shootings resulting in casualties. This finding emphasizes the need for tailored interventions in these regions. For example, policymakers can allocate funding to implement enhanced physical security measures in schools, such as metal detectors, reinforced entry points, and surveillance systems. Additionally, improving emergency response systems, including faster deployment of law enforcement and medical personnel, can mitigate the impact of shootings in high-risk areas.

5.1.2 Focus on High-Risk Shooting Types

The model’s findings confirm that “targeted” and “indiscriminate” shootings are strongly associated with higher casualty rates. Policymakers can design prevention strategies specifically

aimed at these shooting types. For “targeted” shootings, schools can adopt proactive threat assessment programs to identify and provide support to at-risk individuals. For “indiscriminate” shootings, community-level initiatives addressing factors such as mental health, social isolation, and access to firearms can reduce the likelihood of such incidents.

5.1.3 Strengthening Gun Control Measures

Given the role of firearms in school shootings, this study highlights the importance of stricter gun control measures. Policies such as universal background checks, red flag laws, and safe firearm storage requirements can reduce unauthorized access to firearms and lower casualty risks. These measures can be particularly impactful in regions identified as having elevated risks, complementing other targeted interventions.

5.1.4 Investment in Mental Health Resources

The findings also point to the importance of addressing underlying causes, such as untreated mental health conditions, that may contribute to school shootings. Policymakers can expand access to mental health services by increasing funding for school counselors and psychologists, creating community-based support programs, and implementing mental health education initiatives. These efforts can reduce risk factors while fostering a supportive environment for students.

5.1.5 Ongoing Policy Evaluation

Effective policymaking requires regular assessment of interventions. This study provides a framework for evaluating the impact of existing measures, such as increased school security or stricter firearm regulations. Policymakers can use predictive models to monitor changes in casualty risks and adapt strategies based on emerging data. For example, periodic updates to the model with new data can refine predictions and improve policy responsiveness.

5.2 Limitations of the Analysis

The reliability of the model depends on the availability and quality of the underlying data and priors. Changes in societal conditions or policy environments may reduce the model’s predictive power over time. Additionally, the analysis does not account for several important contextual factors, such as individual school characteristics, socio-economic disparities, and differences in firearm accessibility across states. Including these factors could improve the model’s ability to capture the complexity of school shooting dynamics.

5.3 Recommendations for Future Research

5.3.1 Expanding the Predictive Framework

Future studies should incorporate additional predictors to enhance the model's scope and accuracy. Factors such as school size, local economic conditions, and access to firearms could provide a more detailed understanding of the drivers of casualties. Incorporating longitudinal data could also help capture temporal trends and the impact of policy changes over time.

5.3.2 Evaluating Policy Effectiveness

To ensure that interventions are impactful, future research should evaluate the effectiveness of existing policies. For example, studies could assess whether stricter gun control laws or increased mental health funding significantly reduce the likelihood or severity of school shootings. Policymakers could use these insights to refine strategies and develop evidence-based policies.

5.3.3 Encouraging Interdisciplinary Collaboration

Addressing the complex issue of school shootings requires collaboration across disciplines. Future research should bring together experts from education, mental health, public policy, and law enforcement to design holistic interventions. Collaborative efforts can also facilitate data sharing, improve predictive modeling, and create a more comprehensive understanding of how to reduce casualties effectively.

A Appendix

A.1 Analysis of The Washington Post’s Data Collection Approach

The Washington Post’s investigation into school shootings demonstrates a rigorous approach to data collection and analysis, adhering to established standards in data journalism and social science research.

A.1.1 Data Collection and Triangulation

The Post utilized multiple data sources—including Nexis, news articles, open-source databases, law enforcement reports, and direct communications with schools and police departments—to compile a comprehensive list of school shootings. This method reflects the principle of data triangulation, which enhances the validity of findings by cross-verifying information from diverse sources (Denzin 1978). Such triangulation is crucial in data journalism to ensure accuracy and comprehensiveness.

A.1.2 Defining Inclusion and Exclusion Criteria

By establishing clear criteria for what constitutes a school shooting—excluding incidents like after-hours events, accidental discharges without injury, and private suicides—the Post ensured a focused analysis. This practice aligns with the concept of construct validity, ensuring that the study accurately reflects the phenomenon under investigation (Cronbach and Meehl 1955).

A.1.3 Sampling and Estimation Techniques

To estimate the number of children exposed to gunfire, the Post used enrollment figures from the U.S. Education Department and adjusted for average daily absenteeism. Specifically, they deducted 7% from enrollment totals to account for average absenteeism and reduced enrollment by 50% for shootings occurring just before or after school hours. This approach mirrors sampling techniques discussed in “Telling Stories with Data”, which emphasizes the importance of identifying the target population, understanding the context-specific variables (such as absenteeism rates), and applying appropriate adjustments to ensure the representativeness and reliability of the results (Alexander 2023). For example, the text highlights the necessity of incorporating real-world constraints—such as data availability or measurement gaps—into the design of sampling frameworks to improve the accuracy of population estimates.

A.1.4 Reproducibility and Transparency

The Post’s transparency in detailing their methodology allows for reproducibility and critical assessment, key components in data journalism. As highlighted in *Telling Stories with Data*, reproducible workflows involve sharing data, code, and analytical steps, enabling others to replicate and validate findings (Alexander 2023).

A.1.5 Addressing Potential Biases

The Post acknowledged that their database might not include every qualifying incident due to the absence of a federal tracking system. This recognition of potential data availability bias is essential, as discussed in the literature on observational data, which emphasizes the importance of considering who is included in the dataset and who is systematically excluded (Kaplan et al. 2014).

A.1.6 Weaknesses

Despite its strengths, The Washington Post’s methodology has notable limitations. Reliance on media reports and open-source databases may introduce reporting bias, potentially under-representing incidents in areas with limited media coverage (Kaplan et al. 2014). The exclusion of certain incidents, such as private suicides, could understate the broader psychological impact of gun violence. Additionally, adjustments for absenteeism rates assume consistent patterns across all schools, which may overlook localized variations and specific community contexts (Alexander 2023). The use of proprietary tools like Nexis and reliance on non-public datasets also pose challenges for reproducibility, limiting the ability of others to independently verify findings. Finally, the absence of a centralized federal tracking system for school shootings creates a systemic data gap that even the Post’s robust methodology cannot fully address (Kaplan et al. 2014).

A.1.7 Conclusion

The Washington Post’s methodology in investigating school shootings demonstrates a rigorous application of data collection and analysis principles, aligning with best practices in data journalism and social science research. By employing data triangulation (Denzin 1978), defining clear inclusion criteria (Cronbach and Meehl 1955), adjusting for sampling considerations (Alexander 2023), ensuring reproducibility (Alexander 2023), and acknowledging potential biases (Kaplan et al. 2014), the Post provides a robust framework for understanding the impact of school shootings on children. Nonetheless, the identified weaknesses underscore the importance of continued refinement and broader systemic solutions for data collection.

A.2 Model details

A.3 Posterior predictive check

In Figure 7a, we implement a posterior predictive check. This shows that the posterior predictive distribution (y_{rep}) aligns well with the observed data (y), indicating that the model provides a reasonable fit to the data and captures its underlying structure effectively.

In Figure 7b, we compare the posterior with the prior distributions. This shows that the posterior estimates for all parameters (*Intercept*, *lat*, and *shooting_categoryother*) are informed by the data, as the posterior distributions are narrower and shifted relative to the priors, reflecting updated beliefs based on the observed evidence.

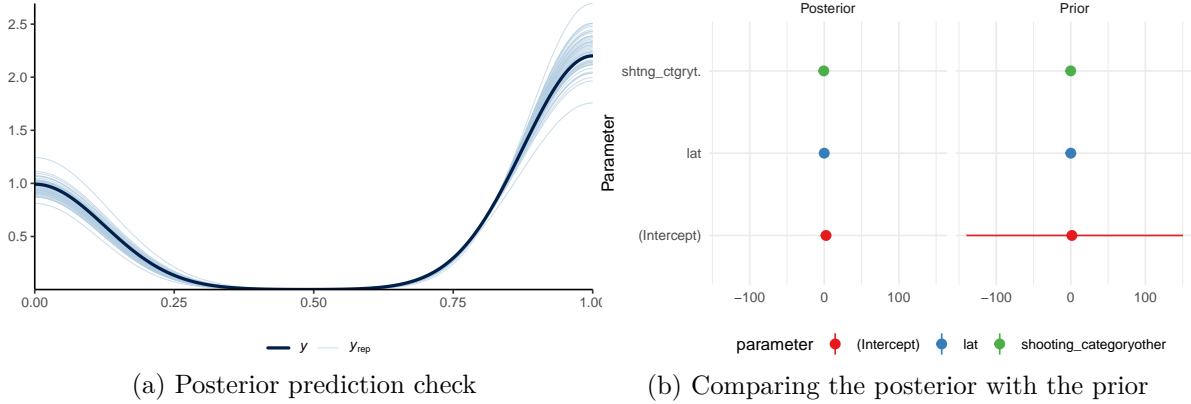


Figure 7: Examining how the model fits, and is affected by, the data

A.4 Diagnostics

Figure 8a is a trace plot. It shows the sampling behavior of the MCMC algorithm for the model parameters (*Intercept*, *shooting_categoryother*, and *lat*) across four chains. The chains are well-mixed, stationary, and overlap significantly, indicating good convergence and effective exploration of the posterior distributions. It suggests that the model's parameters were sampled effectively, with no immediate signs of convergence issues.

Figure 8b is a Rhat plot. This \hat{R} plot shows that all parameters have $\hat{R} \approx 1.00$, indicating excellent convergence of the MCMC chains. The results suggest that the chains are well-mixed, and the posterior distributions can be considered reliable for interpretation.

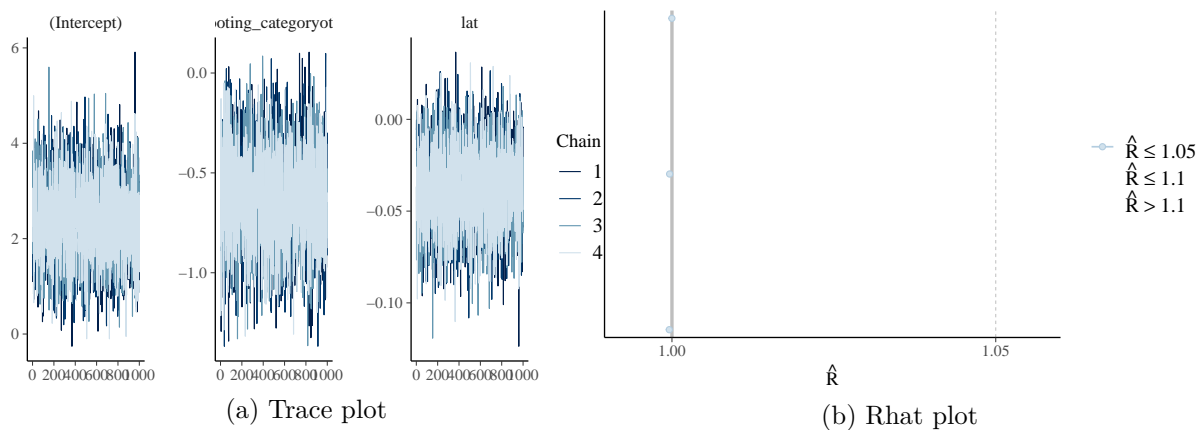


Figure 8: Checking the convergence of the MCMC algorithm

References

- Alexander, R. 2023. *Telling Stories with Data*. Chapman and Hall/CRC.
- Cronbach, L. J., and P. E. Meehl. 1955. “Construct Validity in Psychological Tests.” *Psychological Bulletin* 52: 281–302. <https://doi.org/10.1037/h0040957>.
- Denzin, N. K. 1978. *The Research Act: A Theoretical Introduction to Sociological Methods*. McGraw-Hill.
- Gabriel A. Fonseca, Jonah Gabry, Ben Goodrich, Sam Brilleman, Ben Lambert, and Andrew Gelman. 2023. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. <https://mc-stan.org/rstanarm/>.
- Garrett Golemund, and Hadley Wickham. 2024. *Lubridate: Make Dealing with Dates a Little Easier*. <https://CRAN.R-project.org/package=lubridate>.
- Hadley Wickham. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org/>.
- Hadley Wickham et al. 2024a. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- et al. 2024b. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.
- et al. 2024c. *Stringr: Simple, Consistent Wrappers for Common String Operations*. <https://CRAN.R-project.org/package=stringr>.
- et al. 2024d. *Testthat: Unit Testing for r*. <https://CRAN.R-project.org/package=testthat>.
- et al. 2024e. *The Tidyverse: Easily Install and Load the 'Tidyverse'*. <https://CRAN.R-project.org/package=tidyverse>.
- Hao Zhu. 2023. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.
- Kaplan, S. et al. 2014. “Observational Bias in Epidemiological Studies.” *Journal of Epidemi-*

- ology and Community Health* 68 (5): 400–405. <https://doi.org/10.1136/jech-2013-203497>.
- Kirill Müller. 2023. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- Neal Richardson et al. 2024. *Arrow: Integration to 'Apache Arrow'*. <https://CRAN.R-project.org/package=arrow>.
- Pedersen, Thomas Lin. 2020. “Patchwork: The Composer of Plots.” *R Package Version 1.1.1*. <https://patchwork.data-imaginist.com/>.
- Project, Apache Arrow. 2024. *Arrow: Integration to 'Apache Arrow'*. <https://arrow.apache.org/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- RAND Corporation. 2015. “Can Technology Make Schools Safer?” https://www.rand.org/pubs/research_briefs/RB9922.html.
- Ray Brownrigg, Thomas P Minka, Robert A Becker, Allan R Wilks, Original S code by Richard A Becker, and Allan R Wilks. 2023. *Maps: Draw Geographical Maps*. <https://CRAN.R-project.org/package=maps>.
- The Washington Post. 2024a. “School Shootings Data.” <https://github.com/washingtonpost/data-school-shootings/blob/master/school-shootings-data.csv>.
- . 2024b. “School Shootings Database.” 2024. <https://www.washingtonpost.com/education/interactive/school-shootings-database/>.
- Vincent Arel-Bundock. 2023. *Modelsummary: Summary Tables and Plots for Statistical Models and Data*. <https://vincentarelbundock.github.io/modelsummary/>.
- Yihui Xie. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. Yihui Xie. <https://yihui.org/knitr/>.