2025 US School Shooting Predictions and Solutions*

Leveraging Forecasting to Shape Policy and Protect Communities

Yun Chu

November 26, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2….

Paragraph of Estimand!

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 2024). These entries cover 50 variables that provide information about the schools, 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.

As the federal government does not consistently track school shootings, this dataset from *The Washington Post* fills a critical gap. It was carefully assembled using information from diverse sources, including news articles, open-source databases, law enforcement reports, and direct inquiries to schools and police departments. 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.

We use the statistical programming language R (R Core Team 2023) to download, clean, analyze and model the US School Shooting Data. The US School Shooting dataset is downloaded from The Washington Post (**washingtonpost_school_shootings_data?**) . The following libraries are utilized in this paper:

- tidyverse (Wickham et al. 2024e)
- dplyr (Wickham et al. 2024a)
- lubridate (Grolemund and Wickham 2024)
- readr (Wickham et al. 2024b)
- stringr (Wickham et al. 2024c)
- arrow (Richardson et al. 2024)
- testthat (Wickham et al. 2024d)
- randomForest (Liaw and Wiener 2024)
- ggplot2
- maps

2.2 Measurement

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 2024).

2.3 Summary Statistics & Relationship Between Variables

$Causalities = Killed + Injured$

2.4 Outcome variables

The outcome variable for this analysis is number of causalities for each state in 2025.

Figure 1 shows the number of school shooting incidents in US since 1999. The number of school shooting incidents almost doubled since 2020 compared with previous years.

Number of US School Shooting Incidents Over Time

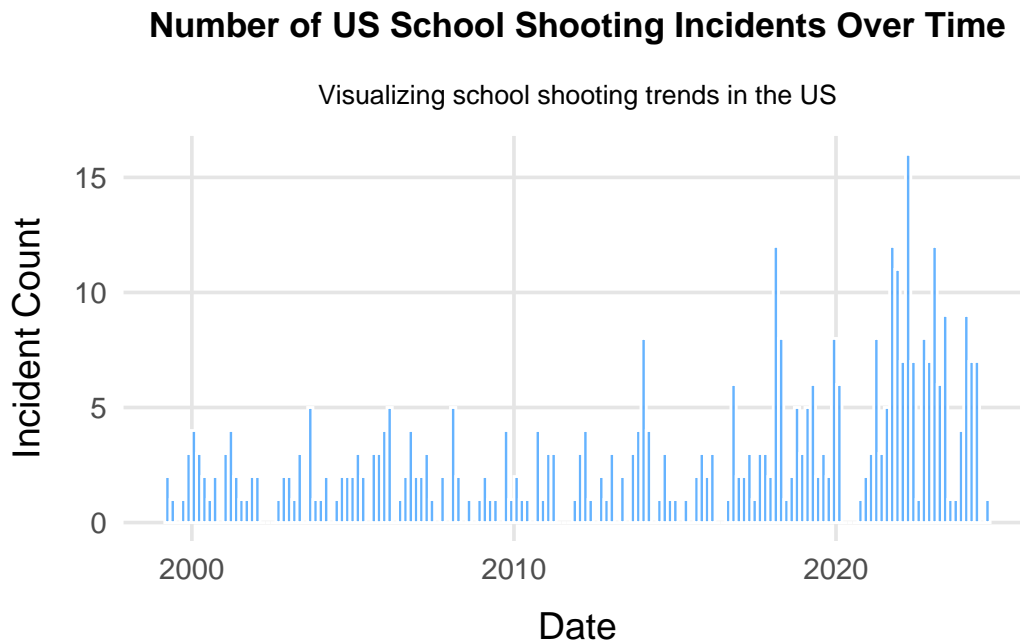Visualizing school shooting trends in the US

Figure 1: Number of US School Shooting Incidents Over Time

Figure 2 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 20 school shooting incidents in the past 25 years while other states has less than 20 school shootings.

2.5 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

3 Model

The goal of our modelling strategy is twofold. Firstly,...

US School Shooting Casualties by State
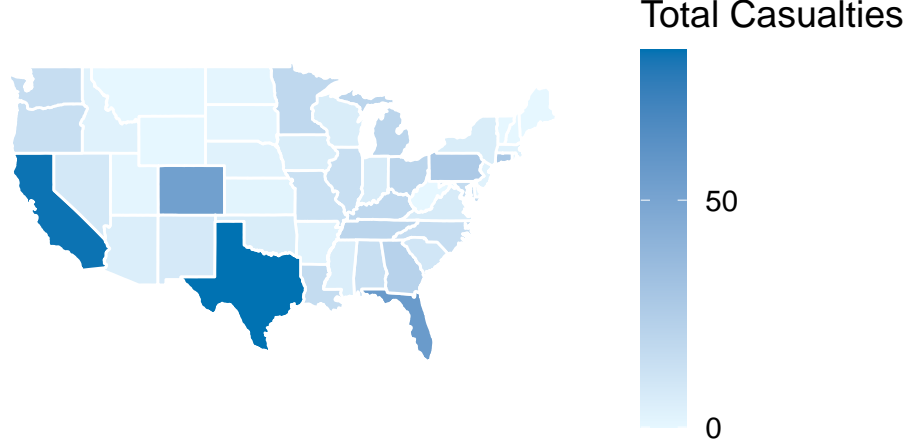
Number of casualties visualized geographically

Figure 2: US School Shooting Casaulties by State

Here we briefly describe the Bayesian analysis model used to investigate… Background details and diagnostics are included in Appendix B.

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\sigma \sim \text{Exponential}(1) \tag{6}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of (**rstanarm?**). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular…

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in Table **??**.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows...

Examining how the model fits, and is affected
by, the data

B.2 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows... This suggests...

?@fig-stanareyouokay-2 is a Rhat plot. It shows... This suggests...

Checking the convergence of the MCMC algo-
rithm

References

Grolemund, Garrett, and Hadley Wickham. 2024. *Lubridate: Make Dealing with Dates a Little Easier.* https://CRAN.R-project.org/package=lubridate.

Liaw, Andy, and Matthew Wiener. 2024. *randomForest: Breiman and Cutler's Random Forests for Classification and Regression.* https://CRAN.R-project.org/package=randomForest.

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

Richardson, Neal et al. 2024. *Arrow: Integration to 'Apache Arrow'.* https://CRAN.R-project.org/package=arrow.

The Washington Post. 2024. "School Shootings Data." https://github.com/washingtonpost/data-school-shootings/blob/master/school-shootings-data.csv.

Wickham, Hadley 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.