State-Level Spatial Analysis of Shooting Incidents in the United States

Nam-Anh Tran

Abstract

Shooting incidents in the U.S. significantly impact mortality, morbidity, and mental health. We aim to investigate shooting rates to quantify better the burden of firearm-related injuries and deaths and, thus, shape strategies to reduce these adverse outcomes. We apply statistical spatial models within a Bayesian framework to examine shooting risk factors and evaluate rates across 48 contiguous U.S. states and the District of Columbia in 2018. Our findings reveal a positive association between drug use and shooting incidents. Delaware, the District of Columbia, Louisiana, Mississippi, and Alabama show significantly higher relative shooting risk than expected under a constant-risk assumption, indicating that unmeasured spatial factors and local differences play a crucial role in shooting violence. These results highlight the importance of spatial models in explaining complex risk patterns and provide actionable insights for targeted policy interventions.

1 Introduction

Firearm violence claimed 38,390 lives in the United States, with an average of 105 deaths per day in 2018, leading firearms to become one of the nation's leading causes of injury and mortality. That year also saw 14,414 firearm-related homicides and 323 mass shooting incidents, collectively resulting in 387 deaths and 1,283 injuries. This dramatically impacts communities across the country.[1,2] Notably, firearms have become the leading cause of death among children and adolescents,[3] which indicates a growing public health crisis with ripple impact through families, schools, and entire neighbourhoods; this implies the urgent need for comprehensive intervention.

Addressing this crisis imposed enormous remediation costs, including financing emergency medical treatment, long-term mental health care, criminal justice responses, victim support services, and community recovery efforts. The societal cost of gun violence in 2018 reached $280 billion for direct medical care, criminal justice expenses, lost productivity, and the immeasurable pain and suffering endured by victims and communities.[4] Moreover, survivors of firearm injury often endure lifelong disabilities, post-traumatic stress, depression, and chronic health problems, overwhelming hospitals and mental health systems. Meanwhile, communities experience eroded trust, reduced property values, and stunted economic growth.

It is an urgent necessity to reduce shooting-related incidents due to the heavy human and economic toll. The relevant factors of shootings should be identified to provide critical understanding to make targeted prevention efforts and maximize life-saving impact. Also, shooting incidents are not uniformly distributed across the United States, requiring different prevention approaches across regions. We employ a spatial analysis of shooting incidents to facilitate the determination of geographic "hotspots". This allows policymakers, public health officials, and local stakeholders to implement evidence-based interventions and strategically allocate resources to build safe communities.

We examine why some states experienced significantly more shooting incidents than others in 2018 by compiling a state-level dataset on firearm violence for that year. The response is the total number of reported shootings, involving both fatal and nonfatal, across the 49 U.S. states (including the District of Columbia) extracted from the Gun Violence Archive's verified 2018 incident logs.[2]

The connection between shooting incidents and potential factors relating to socioeconomic conditions, urbanization, gun-related factors, and community and social influences are also evaluated as they play an important role in understanding the causes of violence. Considering these variables in spatial analyses allows for an accurate assessment of their impacts, aiding in developing more effective, evidence-based strategies to reduce shooting incidents.[5–7] We gather state-level factors from multiple sources, reported for 2018.

Statistical models are developed within a Bayesian framework to analyze the spatial dynamics of shooting incidents and their influencing factors. Notable spatial models, such as the Conditional Autoregressive (CAR) model, the Besag-York-Mollié (BYM) model, and its variant (known as BYM 2) effectively capture correlations among adjacent regions while quantifying uncertainty in parameter estimates through posterior distributions.[8] We employ the Bayesian paradigm as it enables external information from previous studies to be incorporated into the model, and this enhances the stability of estimates in regions characterized by limited data.

By fitting spatial models within the Bayesian paradigm, we aim to achieve three key objectives: First, we examine whether shooting incidents in the considered U.S. states have a spatial structure; second, we investigate factors that impact shooting incidents; finally, we assess the relative risk of shooting incidents under the influence of key factors and spatial structure.

2 Data exploration

In our analysis of shooting incidents across 49 U.S. states in 2018, we consider nine state-level covariates: gun policy score, gun ownership, median age, percentage of the population with a bachelor's degree or higher, poverty, unemployment, urbanization, drug use, and per capita personal income. Prior studies indicate that both the regulatory environment and socioeconomic conditions significantly influence firearm violence.[9–11] By examining these factors, we aim to capture the direct regulatory impacts and socioeconomic conditions affecting shooting incidents.

Some of these covariates may have a high correlation as they are interconnected. For example, states with high poverty may have lower educational attainment and higher unemployment. This multicollinearity can complicate model estimation. Thus, only the most significant predictors should be selected to address these issues.

As shown in the Appendix (Figure A.1), the correlation matrix reveals that income and higher educational attainment are highly correlated (0.9). We chose to retain income in our model, as it is considered a more robust indicator of economic stress and inequality linked to higher firearm violence rates. Research by Siegel et al.[10] demonstrates that lower income levels and greater income inequality strongly correlate with increased firearm homicide rates; while educational attainment is important, it serves as a less direct predictor. Despite some remaining high correlations (the highest is 0.83) among other covariates, we retain eight variables to fit a selection model that identifies those significantly contributing to shooting incidents. Excluding the higher education degree reduces the highest correlation from 0.9 to 0.83, helping to mitigate multicollinearity in the variable selection model.

Figure 1: The map shows shooting incidents in U.S States in 2018.

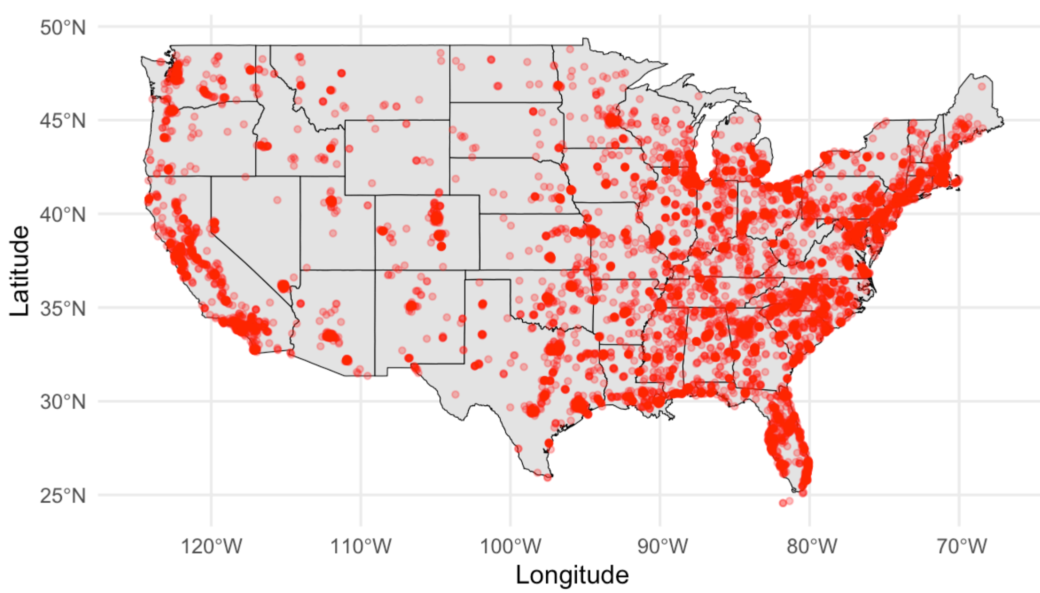


Figure 1 provides a point-based visualization of shooting incidents across considered U.S. states in 2018, with each red dot representing a reported case. At first glance, there appear to be particularly dense clusters of incidents in and around major metropolitan areas, suggesting that urban centers may experience higher concentrations of shootings. While this point map offers valuable insight into the spatial spread of individual cases, a choropleth map could be helpful to show the total number of incidents per state, and hence an aggregated view that more clearly reveals state-level differences and potential regional patterns in the data.

Figure 2: The map shows density of SIR in U.S States in 2018.

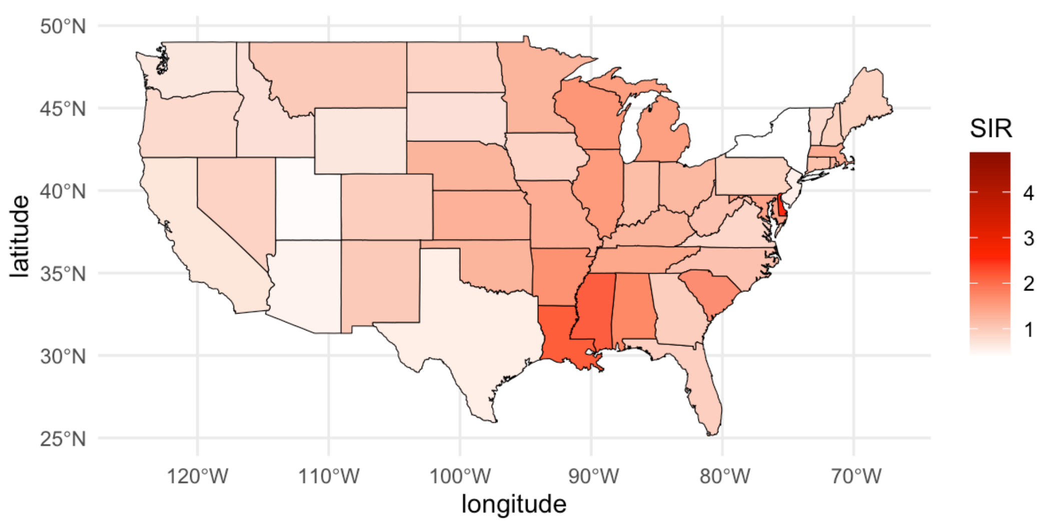


Figure 2 illustrates the distribution of standardized incident ratio (SIR) by state in 2018. We utilize SIR rather than the raw shooting counts since the population across states varies, and a larger state will naturally have more incidents, even if the risk per state is the same. Thus, using the raw counts of shootings can mislead our evaluation of the spatial structure of shooting incidents. In contrast, SIR allows for comparing shooting risks across states on an equal footing, as it represents the number of incidents in a region if its risk were the same as the baseline risk.

The map (Figure 2) highlights apparent regional disparities, with certain states (e.g., in the South and West Coast) showing notably higher counts. This pattern suggests a spatial structure, i.e., some neighbouring states share similarly elevated (or lower) shooting levels. We can employ global and local Moran

Table 1: Result of the global Moran test.

Moran's I	Expected I	Variance	P-value
0.13	-0.02	0.006	0.022

tests to formally investigate and quantify this spatial clustering, which provides statistical evidence of spatial autocorrelation.

The result of the global Moran test shown in Table 1 indicates a statistically significant spatial structure of shooting incidents (p-value = 0.02). However, the relatively low value of the Moran's I test (0.13) implies a weak spatial autocorrelation. Particularly, Figure A.2 in the Appendix shows that only 4 of 49 states have statistical evidence of spatial structure. Thus, modelling shooting incidents is necessitated to better reveal and adjust for local spatial dependencies and covariate effects.

3 Methods

3.1 Variable selection model

We initialize by the variable selection model. We fit the Poisson regression with independent residuals using spike-and-slab priors to assess the level of contribution of covariates to shooting incidents. The model is expressed as follows.

$$\begin{aligned} &y_i \sim \text{Poi}(E_i \lambda_i), \quad i = 1, 2, \ldots, 49 \\ &\ln \lambda_i = \beta_0 + \boldsymbol{x}_i' \boldsymbol{\beta}_1 + \theta_i, \end{aligned} \tag{1}$$

where y_i is the shooting count, E_i is the expected counts, λ_i is the relative risk (RR) corresponding to state i. Also, $\boldsymbol{x}_i$ is the vector of covariates associated with gun policy, gun ownership, median of age, percentage of poverty, unemployment, drug use, urbanization, and income; β_0 and $\boldsymbol{\beta}_1$ are the intercept and vector of coefficients corresponding to covariate vector $\boldsymbol{x}$, respectively; finally, θ_i is the independent error corresponding to state i. The model is fitted using the following spike-and-slab priors.

$$\begin{aligned}
&\beta_0 \sim N(0,5), \quad \beta_{1k} \sim \delta_k N(0,5) + (1-\delta_k) N(0, 0.001), \\
&\delta_k \sim Be(\pi_k), \quad \pi_k \sim U(0,1), \quad k = 1, 2, \ldots, 5. \\
&\theta_i \sim N(0, \sigma_\theta^2), \quad 1/\sigma_\theta^2 \sim Gam(1, 0.026)
\end{aligned} \tag{2}$$

The model with the priors defined in Equation 2 is known as a fully continuous spike-and-slab model as the point-mass of the spike component is replaced by a continuous distribution that is heavily concentrated about zero, and $P(\delta_k = 1|\boldsymbol{y})$ is thresholding the posterior inclusion probabilities.[12] We employ the "median thresholding" rule, i.e., $P(\delta_k = 1|\boldsymbol{y}) > 0.5$,[13] to rank predictors. It is worth noting that we use non-pooling inclusion probability, i.e., we relax the condition that all covariates have an analogous chance of being included; this allows the model to adapt to heterogeneity by learning different inclusion probabilities for each predictor, enhancing flexibility.

3.2 Analysis models

We fit three models, CAR, BYM and BYM 2. The formulae of three models are similar to Equation 1, albeit the linear term $\ln \lambda_i$.

$$\begin{aligned}
&\ln \lambda_i = \beta_0 + \boldsymbol{x}_i' \boldsymbol{\beta}_1 + b_i, \quad \text{where} \\
&\text{CAR:} \quad b_i = \phi_i \\
&\text{BYM:} \quad b_i = \phi_i + \theta_i \\
&\text{BYM 2:} \quad b_i = \sigma(\sqrt{1-\rho}\,\theta_i + \sqrt{\rho}\,\phi_i^*),
\end{aligned} \tag{3}$$

where $\phi_i|\boldsymbol{\phi}_{-i} \sim N(\bar{\phi}_{\delta_i}, \frac{\sigma_\phi^2}{n_{\delta_i}})$, where $\bar{\phi}_{\delta_i} = n_{\delta_i}^{-1} \sum_{j \in \delta_i} \phi_j$, with δ_i and n_{δ_i} representing the set of neighbours and the number of neighbours of state i, respectively. Intercept, slopes and the unstructured component are defined similarly in the variable selection model, albeit $\beta_{1k} \sim N(0,5)$. For the extra parameters in

BYM 2, we set $\sigma \sim N^+(0,1)$, $\rho \sim Beta(0.5, 0.5)$; ϕ_i^* is a scaled spatially structured component. More details were discussed by Morris et al.[8] These models assume that shooting incidents are Poisson counts, and a linearly additive effect of the predictors and random effect on the logarithmic relative risk.

It is essential to acknowledge that E_i represents the expected number of shooting incidents if all states had a constant underlying risk. This reflects a hypothetical scenario where the incident rate of each state is uniformly distributed, ignoring local risk factors or spatial structure. E_i is later used as a comparator to evaluate shooting incidents under the influence of covariates and spatial structure. We also calculate the Watanabe-Akaike Information Criterion (WAIC), a fully Bayesian measure used for model comparison, to select the final analysis model from which we draw inferences and answer our research questions. Based on this final analysis model, we estimate state-level relative risk (RR) and exceedance probability at 1.5 to identify states with a high probability that the RR is 50% higher than the expected number of shootings at baseline.

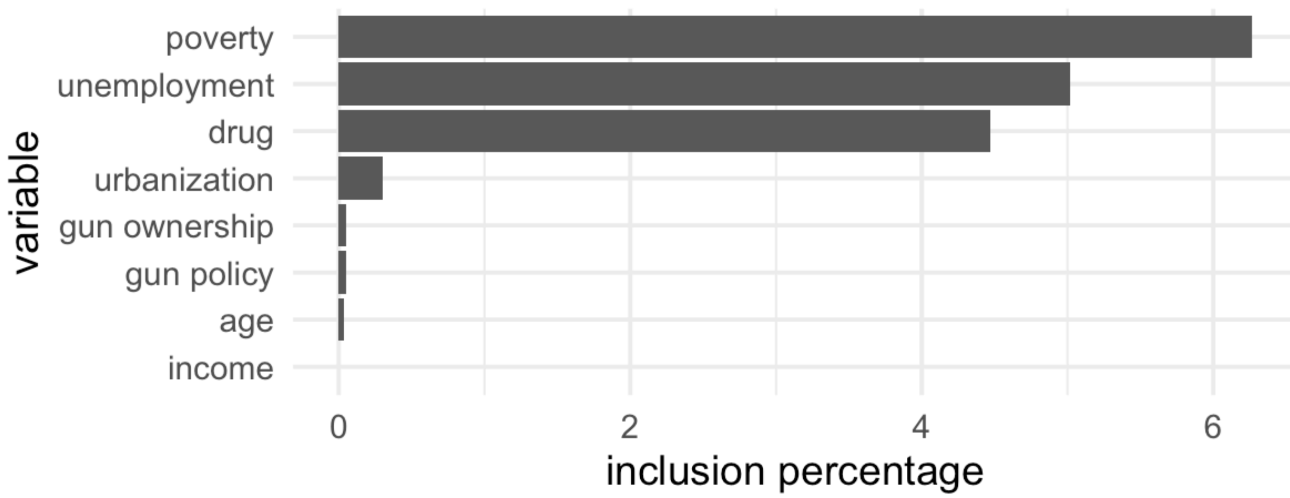
All models, including the variable selection model, are fitted using `stan` in the R environment. `stan` employs full Bayesian inference via Hamiltonian Monte Carlo. We generate 4,000 samples following 10,000 warm-up iterations. Model convergence is assessed using $\hat{R}$, with values below 1.1 indicating successful convergence.

4 Results

4.1 Variable selection model

The bar plot (Figure 3) shows the inclusion percentage of covariates obtained by fitting the Poisson regression with independent residuals using spike-and-slab priors. We exclude income and age as they contribute the least to shooting incidents. While gun ownership and gun policy are critical factors influencing shooting rates, with analogous inclusion percentages, they are highly correlated (-0.81). Given this

Figure 3: The bar plot shows the inclusion percentage of factors, obtained by fitting the Poisson regession with independent residuals.



high collinearity, we retain only one covariate to capture the overall effect of these intertwined factors. Gun policy is selected because its measures are actionable and can be directly targeted by legislative and enforcement interventions. In contrast, gun ownership is deeply rooted in historical, cultural, and socioeconomic factors that are not easily modified in the short term.[14] Focusing on gun policy provides clear, actionable insights that guide policy reforms to decrease shooting incidents. Thus, the analysis models are fitted with five covariates: poverty, unemployment, drug use, urbanization, and gun policy.

4.2 Model fitting

Table 2 shows the WAIC estimates of three analysis models: CAR, BYM, and BYM 2, along with the summary (mean and 95% HDI) of $\rho \in [0, 1]$, which indicates the spatial nature of the residual variation. We found evidence, supported by a 95% HDI of [0.81,1], of the spatial structure of shooting incidents across the states. Thus, all three analysis models are valid as they incorporate this spatial structure, with BYM being the best fit for the data due to its lowest WAIC. However, BYM 2, with a WAIC of 4.21, is not far behind the BYM model. Therefore, we select BYM 2 as the final analysis model to draw inferences; another reason for selecting BYM 2 is its flexibility as it allows the data to determine the magnitude of

Table 2: WAIC of three analysis models: CAR, BYM, and BYM 2, and the summary (mean and 95% HDI) of $\rho \in [0, 1]$, which indicates how sptatial the residual variation is.

	CAR	BYM	BYM 2
$\widehat{WAIC}$ (SE)	421.6 (8)	420.6 (7.9)	421 (8)
$\hat{\rho}$ (95% HDI)			0.96 [0.81, 1]

Table 3: Table shows the summary (mean and 95% HDI) of exponentiated coefficients reflecting the effect of predictor gun policy, poverty, unemployment, urbanization and drug use.

		95 percent HDI	
variable	mean	lower	upper
gun policy	1.00	0.99	1.01
poverty	1.01	0.96	1.07
unemployment	1.11	0.87	1.34
urbanization	0.99	0.98	1.00
drug use	1.08	1.01	1.15

the spatial and independent random effects incorporated into the model.

Table 3 shows the summary (mean and 95% HDI) of exponentiated coefficients reflecting the effect of covariates. We found evidence that a one percent increase in drug use leads to an 8% increase in the relative risk (RR) of shooting incidents, supported by a 95% HDI of [1.01, 1.15], after adjusting for other covariates and incorporating spatial structure. Although the association between urbanization and shooting incidents is not evident, urbanization may have a slightly negative association with shooting incidents. Further investigation is therefore required to confirm this link.

We also found evidence that 20 of 49 states have higher RR than the expected number of shooting incidents at baseline. The mean and 95% HDI of state-level RR are shown in the forest plot in the appendix (Figure A.3). Further, we calculated the probability that RR 50% exceeds the expected number of shooting incidents at baseline. Table A.1 shows 11 states with the highest probability that RR is 50% higher than the expected number of shooting incidents at baseline. Five states with an exceedance probability of 1 are Delaware, the District of Columbia, Louisiana, Mississippi, and Alabama.

5 Discussion

In this study, we assessed shooting incidents across U.S. states in 2018. We found evidence of an association between drug use and shooting incidents, indicating that an increase in drug use is linked to a rise in shooting incidents. Furthermore, we identified a higher relative risk (RR) of shootings in 20 states compared to the expected number at baseline, with five states exhibiting a 50% higher RR, supported by a probability of 1. Additionally, the existence of a spatial structure was confirmed based on the mixing parameter, measuring the proportion of the marginal variance explained by the spatial component in the BYM 2 model.

One key limitation of this study is that the shooting counts and predictors are extracted from multiple sources. Thus, there is a risk that covariates and shooting counts are aggregated at different geographic scales or use slightly different boundary definitions. This can lead to measurement errors when matching them. Further, the quality and completeness of data can vary across sources. While some datasets are rigorously validated, others might suffer from under-, over-reporting or reporting biases.

This study focuses exclusively on shooting incidents in 2018. This does not allow the study to capture temporal trends and fluctuations, which may impact shooting rates. Future research should integrate a temporal component to comprehend the trends and determinants of shooting rate well. This approach would enable policymakers and researchers to grasp the dynamics of firearm violence better and evaluate the effectiveness of time-sensitive interventions.

6 Conclusion

Although there is no confirmed causality between drug use and shooting rates, shooting prevention campaigns should consider drug use as a key factor, as their association is statistically confirmed in this study. States with a high-risk ratio must adopt more effective approaches in the future to reduce shooting rates.

References

[1] CDC WONDER. https://wonder.cdc.gov/. Accessed March 27, 2025

[2] Gun Violence Archive. https://www.gunviolencearchive.org/. Accessed March 27, 2025

[3] Kravitz-Wirtz N, Bruns A, Aubel AJ, Zhang X, Buggs SA. Inequities in community exposure to deadly gun violence by race/ethnicity, poverty, and neighborhood disadvantage among youth in large US cities. *Journal of urban health*. 2022;99(4):610-625.

[4] News A. Gun violence cost America $280 billion in 2018: Report — abcnews.go.com. Published online.

[5] Hepburn LM, Hemenway D. Firearm availability and homicide: A review of the literature. *Aggression and Violent behavior*. 2004;9(4):417-440.

[6] Cook PJ, Ludwig J. The social costs of gun ownership. Published online 2004.

[7] Braga AA, Weisburd DL. The effects of focused deterrence strategies on crime: A systematic review and meta-analysis of the empirical evidence. *Journal of research in crime and delinquency*. 2012;49(3):323-358.

[8] Morris M, Wheeler-Martin K, Simpson D, Mooney SJ, Gelman A, DiMaggio C. Bayesian hierarchical spatial models: Implementing the besag york mollié model in stan. *Spatial and spatio-temporal epidemiology*. 2019;31:100301.

[9] Wennerstrom A, Rush CH. The terminology of community health workers. *American journal of public health*. 2016;106(8):e10.

[10] Siegel M, Ross CS, King III C. The relationship between gun ownership and firearm homicide rates in the united states, 1981–2010. *American journal of public health*. 2013;103(11):2098-2105.

[11] Glaeser EL, Sacerdote B, Scheinkman JA. Crime and social interactions. *The Quarterly journal of economics*. 1996;111(2):507-548.

[12] Bai R, Ročková V, George EI. Spike-and-slab meets LASSO: A review of the spike-and-slab LASSO. *Handbook of Bayesian variable selection*. Published online 2021:81-108.

[13] Barbieri MM, Berger JO. Optimal predictive model selection. Published online 2004.

[14] Hemenway D. *Private Guns, Public Health*. University of Michigan Press; 2017.

Appendix

Figure A.1: Correlation matrix.

Figure A.2: Local Moran test.

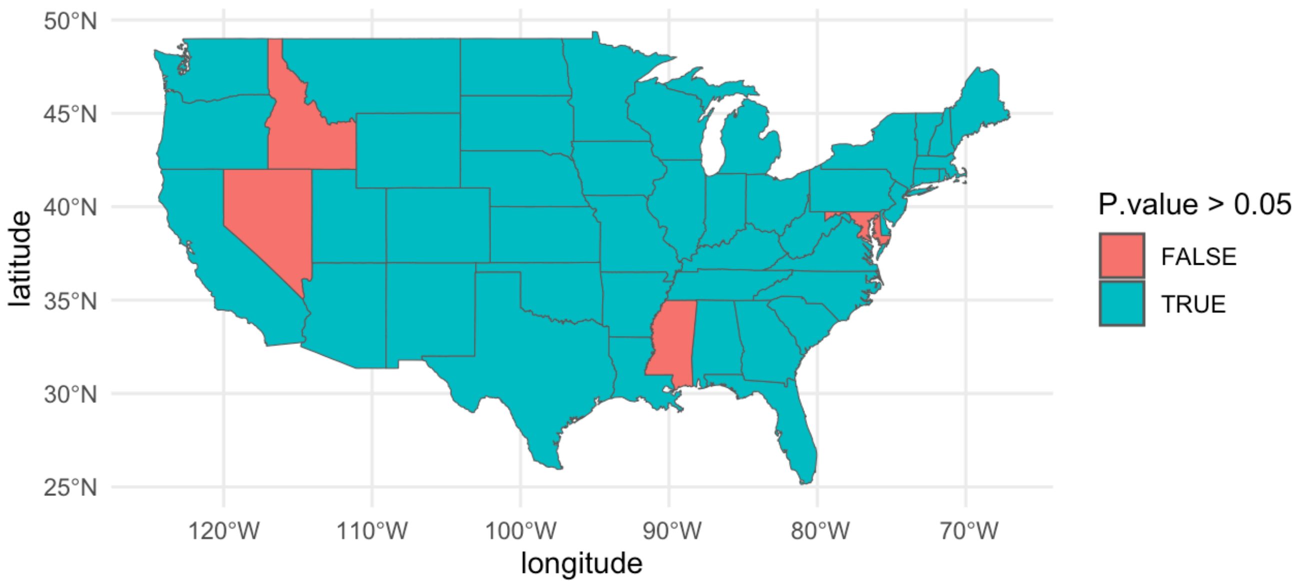


Table A.1: Table shows 11 states that have highest probability that the RR is 50% higher than the expected number of shooting incidents at baseline.

state	exceedance probability
delaware	1.00
district of columbia	1.00
louisiana	1.00
mississippi	1.00
alabama	0.99
south carolina	0.95
arkansas	0.81
wisconsin	0.79
illinois	0.62
michigan	0.54
maryland	0.17

Figure A.3: Forest plot of state-level relative risk.