

# Predicting Counties with Higher/Lower Risk of a School Shooting Incident

---

Alex Bass, Connie Cui, Tulsi Ratnam

# Motivation and Context Overview

Mass shootings have been a rising concern in the United States, and in particular we are seeing more incidences of school shootings across the country. Using our dataset, we aim to answer the following questions:

- Are there certain characteristics of counties that influence gun violence in schools?
- Are these incidences concentrated in particular regions in America? If so, which?
- How do state gun laws play a role in deterring school shootings?

# Data

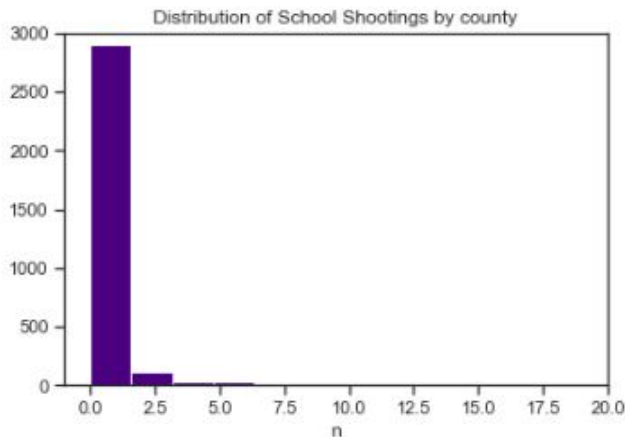
We combined several datasets for our study including:

- A compilation of data with over 2000 school shootings by county over the last 50 years from the Center for Homeland Defense and Security
- Census Data with county level demographics
- A crime dataset with various gun laws by state
- Gun law strictness scores from FiveThirtyEight



# Approach

Because our dependent variable is count data (the number of school shootings in a county), we are creating our regression equation based off of Poisson distributions. We estimate the lambda (our expected mean rate) with an exponential link function. Each of our six models has the equation below as a foundation:



$$\theta = \exp(\beta X)$$

$$Y_{school\_shootings} = Poisson(\theta)$$

# Models

We explored six different models:

- Full Pooled Poisson Regression
- Selective Pooled Poisson Regression
- Full Pooled Zero Inflated Poisson Regression
- Selective Pooled Zero Inflated Poisson Regression
- Hierarchical Poisson Regression
- Hierarchical Zero Inflated Poisson Regression

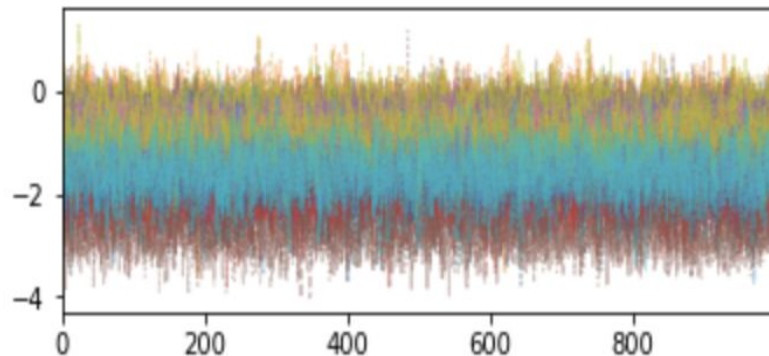
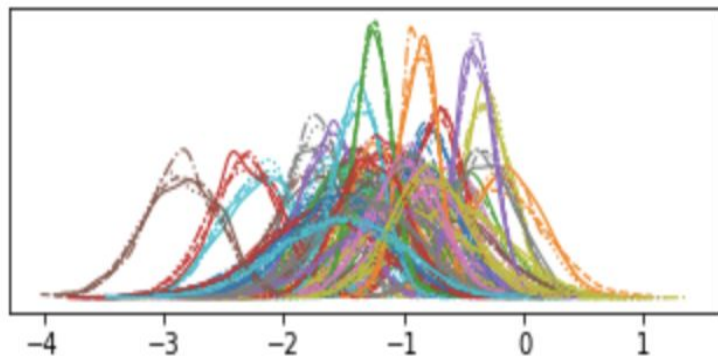
# Model Comparison

	rank	loo	p_loo	d_loo	weight	se	dse	warning	loo_scale
heirarchical_zero_inflated	0	-2225.573280	157.060659	0.000000	4.737757e-01	161.098970	0.000000	True	log
pooled_zero_inflated_all	1	-2337.220527	80.242197	111.647247	1.727807e-01	187.177762	39.590988	True	log
heirarchical_poisson	2	-2428.352785	196.183776	202.779505	2.429195e-01	182.250195	42.404933	True	log
pooled_zero_inflated_selected	3	-2450.220935	78.026997	224.647656	6.276121e-02	179.594327	42.282797	True	log
pooled_all	4	-2640.131213	103.946605	414.557933	6.195499e-15	222.966946	79.446732	True	log
pooled_selected	5	-2816.587878	102.496695	591.014598	4.776295e-02	225.085549	89.752794	True	log

Of the 6 models, the hierarchical zero inflated Poisson regression model performed the best and varies significantly from the other models. The hierarchical poisson and zero-inflated poisson models were more similar in performance, while the other models varied more widely.

# Final Model

Our final model ended up being a hierarchical zero-inflated poisson model. This is a trace plot of the state-level intercepts.



A combination of both of these models better fits our dataset and can give us more accurate predictions than our other models.

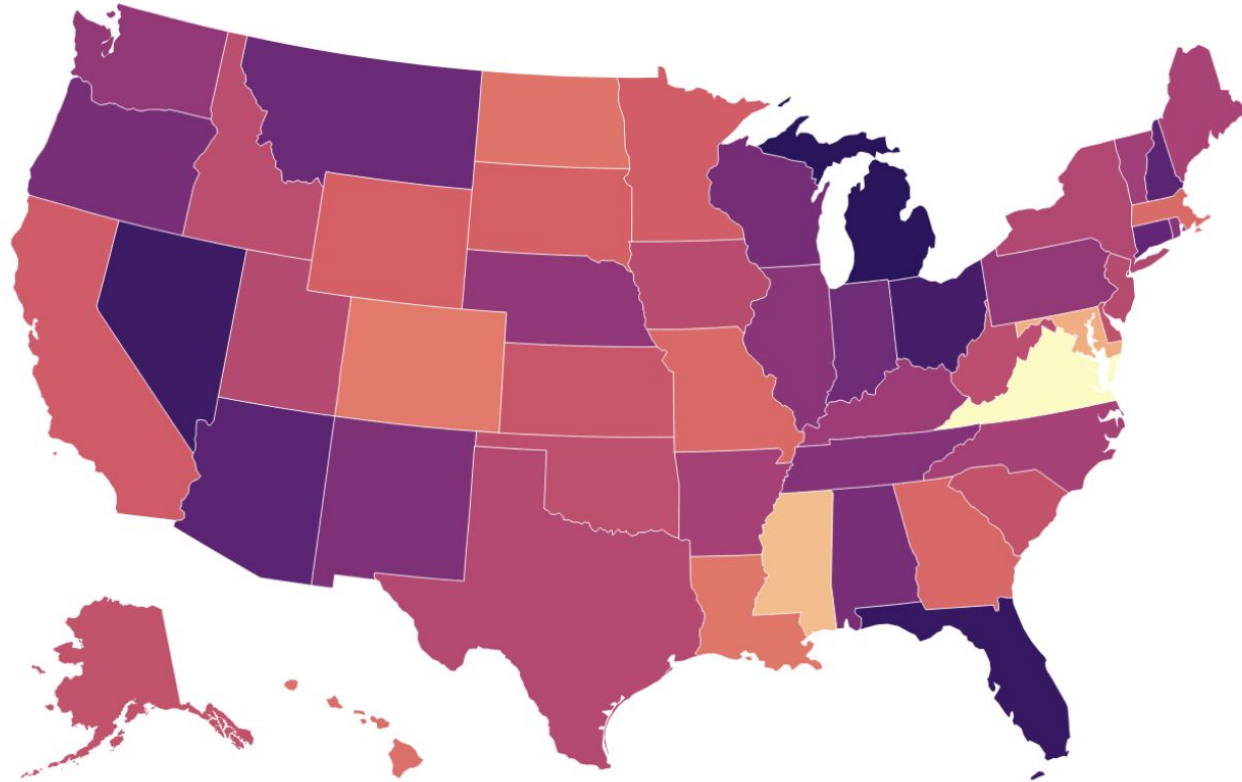
# Findings

	mean	sd	hdi_3%	hdi_97%	r_hat
background_checks	0.509	0.258	0.031	0.997	1
gun_permit_law	-0.662	0.322	-1.274	-0.062	1.01
White	0.229	0.136	-0.029	0.478	1
Black	0.663	0.122	0.443	0.902	1

1. Our model predicts that a one standard deviation increase in Percent of Blacks has the largest impact on predicting school shootings compared to other coefficients.
2. When controlling the other predictors in the model, gun permit laws seem to be more effective at curbing gun violence than background check laws.
3. Different states are predicted to have more shootings than others (see next)



# State Intercepts For School Shootings



# Conclusion

Based on the selection of our model and results, we were able to identify the most significant predictors of school shootings with our available data.

To further this study, we would take our best model and run posterior predictive checks to validate the efficacy of our model's predictions by comparing simulated data with real, observed data. Additionally, we might consider reducing our final model even to only include variables that have a credible interval that does not contain zero or including omitted variables such as party breakdown or the race of shooter and race of victims.

Thank you!

---