Title

School Shootings: A Geostatistical Analysis

Author

Lopez Robles, Valeria Monserrat

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

School Shootings: A Geostatistical Analysis

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

In Applied Statistics and Data Science

by

Valeria Monserrat Lopez Robles

2024

ABSTRACT OF THE THESIS

School Shootings: A Geostatistical Analysis

by

Valeria Monserrat Lopez Robles

Master of Applied Statistics and Data Science

University of California, Los Angeles, 2024

Professor Frederic R. Paik Schoenberg, Chair

This study presents a novel analysis of K-12 school shootings in the continental United States by examining both the geographic distribution and socio-economic factors that may influence these incidents. Unlike previous studies that have primarily focused on temporal patterns or general causal links, this research employs advanced geostatistical methods, such as kernel smoothing and spatial point process functions (F, G, K, and J), to uncover regional clusters and spatial dependencies of school shootings across the country. This work determined that school shootings are not evenly distributed, and clustering is observed on the coasts and in urban areas. By integrating factors such as mental health statistics, gun ownership data, income levels, and homicide rates, the study provides new insights into the complex interplay of factors contributing to these tragic events. While socioeconomic factors, particularly homicide rates, were associated with school shootings, they did not fully explain the frequency of these incidents. The study also

explored the role of guns per capita. The analysis revealed that there was no direct correlation between high gun ownership rates and school shootings, as seen in Wyoming. These findings challenge existing assumptions and highlight the need for a multifaceted approach. This approach should consider geographic, demographic, and socio-economic influences in order to develop more targeted and effective interventions to address this pressing issue.

The thesis of Valeria Monserrat Lopez Robles is approved.

David Paige
Vivian Lew

Frederic R. Paik Schoenberg, Committee Chair

University of California, Los Angeles

2024

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

Introduction

Gun violence in the United States is a pressing social and public health issue. According to the Pew Research Center, 48,830 Americans died from gun-related injuries, such as gun suicides and gun murders, in 2021 [1]. Although the motivations for someone who chooses to partake in these actions are complex, certain factors can be attested to economic and political conditions, mental health issues, and access to firearms. The immediate environment of a person, along with sociocultural influences, are also factors at play [2]. Furthermore, studies by the American Psychological Association suggest that childhood trauma exposure may contribute to violent behavior, and ultimately, gun violence [3].

The early years of a child can have lifelong physical, social, and emotional impacts. Schools play a pivotal role in the holistic development of children, providing a structured environment that nurtures their intellectual, social, emotional, and physical growth, and equips them with essential skills and knowledge for their future [4]. For the majority of American children, school is the place where they spend most of their childhood outside of home. Moreover, many children perceive school as their "second home". Many even rely on school meals as their primary source of food, and identify school to be their safe place.

But what kind of safe space can a school offer, when the most staggering mass shooting locations are K-12 schools? According to the national nonprofit, Sandy Hook Promise, each day 12 children die from gun violence in America; another 32 are shot and injured. Since the shooting at Columbine High School in 1999, more than 338,000 students in the U.S. have experienced gun violence at school [5].

Although there have been nationwide efforts to mitigate this problem by increasing shooter drills, there is minimal research that confirms the value of these shooter drills either for preventing school shootings or for protecting the school community [6]. With the Courts upholding gun rights, due to the second amendment, and the well-funded National Rifle Association, it has become difficult for the federal, state, and local governments to pass gun restrictions and gun safety measure proposals that could benefit the nation [7].

For the purpose of this analysis, a school shooting is defined as "every instance a gun is brandished, fired, or a bullet hits school property for any reason, regardless of the number of victims, time of day, or day of week" [8]. While school shootings occur nationwide, they are not distributed homogeneously, even when normalized with population size. Thus, there are areas with greater concentrations of school shootings than others. Utilizing data from the Center for Homeland Defense Security (CHDS) School Shooting Safety Compendium database [8] and the Zip Code database compiled by Yin [9], this analysis applies geostatistical methods through a point process model to assess clustering of nationwide school shootings, and conditionally, explore the socioeconomic factors and covariates at play.

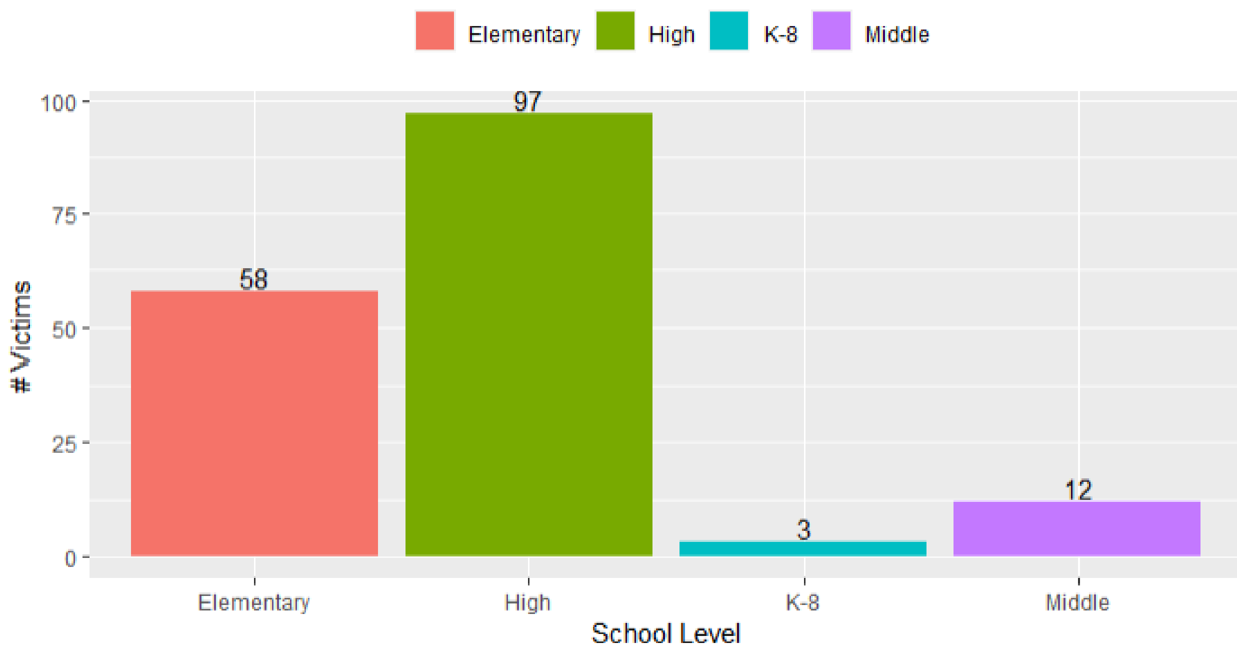
CHAPTER 2

Previous Literature

Schools should be a safe space for students, and one should never have to fear a tragedy of any sort, especially one relating to guns. Exposure to gun violence in a school setting can have a long-lasting, negative and traumatizing effect on the students for years to come. These experiences can lead to mental health issues and gaps in educational development. According to the National Center on Safe Supportive Learning Environments, the levels of crime and violence that a school experiences are strongly correlated to school-wide test scores, graduation rates, and attendance rates [10].

Amidst the increase in violence and shootings, schools have taken on additional security measures and zero tolerance policies to enhance school safety [11]. In 2019–2020, public schools reported controlled access (97%), the use of security cameras (91%), and badges or picture IDs (77%) to promote safety. In addition, high schools (84%), middle schools (81%), and elementary schools (55%) reported the presence of school resource officers (SRO), which are law enforcement officers who work in a school to ensure safety and prevent crime [12].

Despite the increased security, in 2019-2020 there were at least 170 students who were victims of gun violence at school. Another alarming detail is the 30% difference in SROs between elementary schools and high schools, especially when elementary schools are more prone to gun violence than middle schools. Figure 2.1 below is a graphical representation of victims of gun violence at school, per their respective grade level. Please note that K-8 is a separate category as it combines both Elementary and Middle schools grade levels.



<u>Figure 2.1:</u> *National Distribution of School Shooting Victims, per Grade Level (2019-2020)*

Even through the failed security measures, the children and youth continue to look to adults for guidance on how to navigate through these unprecedented times. Yet, it is the adults that are preventing progress in school safety. For example, federal legislation passed in 1996 restricts funding for the National Center for Injury Prevention and Control at the Centers for Disease Control and Prevention (CDC) [11]. Consequently, it is prohibited to use that funding to advocate or promote firearm control and prevention [11]. How is that funding allocated for injury prevention and control cannot be used to prevent/control injury due to school gun violence?

Law officials and school personnel must consider a different approach when implementing school safety measures. What benefit does investing in SROs, controlled access, and security cameras have on school safety if school shootings are not decreasing? There may be better usage in investing that money per the particular needs of each school, rather than a one size fits all mirage of a solution.

CHAPTER 3

Center for Homeland Defense Security (CHDS)

School Shooting Safety Compendium Data

The nationwide data for school shooting incidents in the United States is provided by the Center for Homeland Defense Security (CHDS) School Shooting Safety Compendium Database [8]. It is based on publicly available data for shooting incidents at K-12 schools that occurred from January 1970 through June of 2022.

The data are separated into four subsets: overall incidents, shooter information, victim information, and weapon information. The *Incident* subset holds 2,069 observations, and includes 18 variables: incident identification, city, state, date, school level, and various others. The *Shooter* subset holds 2,294 observations and twelve variables, including incident identification, age, gender, outcome, and various others. The number of observations differs from the *Incident* subset because some school shooting incidents have multiple shooters. The *Victim* subset holds 3,099 observations and includes six variables: incident identification, age, gender, and a couple others. The *Weapon* subset holds 2,071 observations, and includes three variables: incident identification, age, weapon caliber, and weapon type.

The data were filtered to reflect shooting incidents in the continental United States, and omit data relevant to Alaska and Hawaii. After merging the four data subsets, the resulting dataset had 4,153 observations and 36 variables.

CHAPTER 4

ZIP Code, Latitude, Longitude,

City, State, County, Database

The file with all the United States zip codes and their associated latitude, longitude, city, state, and county information was compiled from a city/county/state database and geocoded with Google Maps by Thai Yin [9]. The data holds 42,741 observations and includes six variables: zip code, latitude, longitude, city, state, and county.

Additional data reduction to condense duplicate city-state latitude and longitude coordinates into unique coordinates, and omitting data relevant to Alaska and Hawaii, narrowed down the observations to 29,040 entries. This dataset was the foundation for plotting the shooting incidents against a United States map.

After merging this dataset with the previous CHDS dataset, additional data cleaning, engineering, and feature selection, the resulting dataset had 4,002 observations and 40 variables. This super dataset is used as the basis for all modeling and analysis.

CHAPTER 5

Application of Geostatistical Methods

5.1 Introduction to Point Processes

Point processes are spatial statistical methods that are used to study and understand how events are arranged in space and time. These methods enable us to understand underlying patterns and predict future events, by identifying structure within the data that we can analyze. These insights help us make better informed decisions in a wide range of fields, such as ecology, urban planning, disaster response, and various other industries.

According to Schoenberg, point processes are random collections of points, representing time or location, falling in some space [13]. The most basic models for point processes are the Poisson processes. A Poisson process is a simple point process, where all the points are distinct, with conditional intensity $\boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ does not depend on what points have occurred previously [14]. Complete spatial randomness (CSR) refers to a type of spatial point pattern where events occur entirely randomly and independently. For any space-time location (t,x,y). $\boldsymbol{\lambda}(t,x,y)$ is the limiting expected rate of accumulation of points around (t,x,y), given all points prior to t [14]. The intensity function, $\boldsymbol{\lambda}(x)$ is random, but predictable, and determines the distribution of simple point processes.

` The main types of Poisson processes are categorized as homogenous, where $\boldsymbol{\lambda}$ is constant across R^2, and inhomogenous, where $\boldsymbol{\lambda}$ varies due to space dependencies. Homogeneous point processes are great in theory, as they simplify mathematical and statistical modeling, however, they are not always realistic, due to the simplifications. In real-world applications, one often deals with inhomogeneity because the point patterns are affected by external factors[15]. Therefore, depending on what interactions one is analyzing, covariates and confounding variables should be considered.

The two types of point pattern properties used in this analysis are first-order and second-order. First-order properties treat each point as independent individuals, meaning there is no interaction between points over a specified region. On the other hand, second-order properties measure the interactions between marked sets of points, giving insight into their tendency to cluster or disperse [16]. Both first- and second-order processes produce clustering and smooth variations in density, but the mechanisms involved and the relationships to other variables are very different [17].

5.2 Point Pattern Analysis Tools

5.2.1 Kernel Smoothing

A simple way to start summarizing a spatial point process is by kernel smoothing [13]. Kernel smoothing is a nonparametric density technique used to estimate model structure and probability density from the underlying data [18]. The attractiveness of the Kernel Density Estimation (KDE) model comes from the idea that the more data points in a sample occurring around a location, the higher the likelihood of an observation occurring at that location [18]. Kernel functions, such as Uniform and Gaussian functions, help us compute the probability density given the distance between a fixed location and an observation. The bandwidth controls how smooth the estimated density curve is [18]. Larger bandwidths include points further from our locations, resulting in smoother distributions, but at the cost of data loss information, also known as oversmoothing. Contrary, smaller bandwidths introduce unnecessary randomness into our data, resulting in undersmoothing.

5.2.2 The F Function

The F function, also known as the empty space function, measures the distribution of all distances from an arbitrary location, to its nearest observed point [19]. Let F(r) be the probability that the distance from a randomly chosen location to its nearest point of the process is ≤ r [13], then [20]:

$$\hat{F}(r) = \frac{1}{m}\sum_{j} 1\{d(u_j, x) \leq r\}$$

When considering a homogeneous Poisson process, we model for events assuming they occur uniformly [20]:

$$F_{pois}(r) = 1 - e^{\lambda \pi r^2}$$

The F function is used to determine clustering by comparing the plot of $\widehat{F}$(r) against the plot of $F_{pois}(r)$. Cases where $\widehat{F}$(r) < $F_{pois}(r)$ suggests clustering is present within our point pattern [13].

5.2.3 The G Function

The G function, also known as the nearest neighbor distribution function, is a more sophisticated approach for evaluating complete spatial randomness and measures the distribution of distances from an arbitrary point to its nearest neighbor [19]. Let G(r) be the probability that the distance from a randomly chosen point to its nearest neighbor is ≤ r, then [13]:

$$\hat{G}(r) = \frac{1}{n(x)} \sum_{i} 1\{t_i \leq r\}$$

For a homogeneous Poisson process in an XY plane, the distribution function becomes [19]:

$$G_{pois}(r) = 1 - e^{-\lambda \pi r^2}$$

If $\widehat{G}$(r) > $G_{pois}(r)$, then there are more nearby points than expected under complete spatial randomness, and therefore, indicates clustering. $\widehat{G}$(r) < $G_{pois}(r)$, suggests that as points lie further apart, they influence each other less [21].

For a Poisson process in R^2, $F_{pois}(r) = G_{pois}(r)$ [22].

5.2.4 The K Function

The K function calculates the number of points expected to be a given distance, *r*, away from a certain point [19]:

$$K(r) = \frac{1}{\lambda} E[N_0(r)]$$

The unbiased estimator is therefore [20]:

$$\hat{K}(r) = \frac{1}{\hat{\lambda}\; area(W)} \sum_i \sum_{i \neq j} 1\{||x_i - x_j||\} e(x_i \cdot x_j)\}$$

For a homogeneous Poisson process [16], the function becomes:

$$K_{pois}(r) = \lambda \pi r^2$$

The K function provides an overview of whether the observed distances between a group of points are clustered or dispersed. For clustered spatial point patterns and small values of *r, K(r)* will be large because points are likely to be surrounded by more points. In comparison, for regular/dispersed spatial point patterns and small values of r, K(r) will be small because points are likely to be surrounded by empty space [23]. To determine the size of *K(r)*, the *observed* pattern is compared to the homogeneous Poisson pattern. Therefore, cases where $\widehat{K}(r) > K_{pois}(r)$ indicates clustering, and $\widehat{K}(r) < K_{pois}(r)$ indicates inhibition [23].

5.2.5 The J Function

The J function combines point-to-point (G function) and space-to-point (F function) measurements [19]:

$$J(r) = \frac{1 - G(r)}{1 - F(r)}$$

As mentioned earlier, for a homogeneous Poisson process:

$$F_{pois}(r) = 1 - e^{\lambda \pi r^2} = G_{pois}(r)$$

And therefore,

$$J_{pois}(r) = 1$$

Hence, clustering is present when *J(r) < 1*.

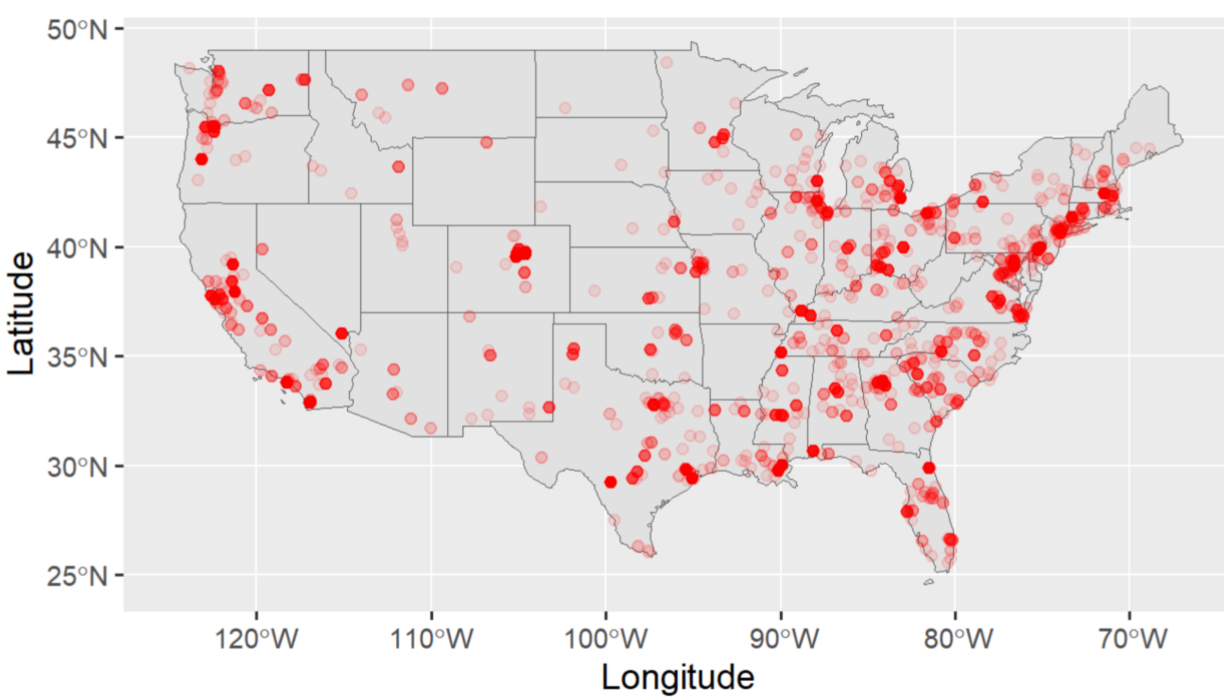
CHAPTER 6

United States:

Data Analysis and Results

6.1 National Overview

Exploratory data analysis is helpful in gaining an overall understanding of the distribution of nationwide K-12 school shootings across the continental United States. Figure 6.1 shows the distribution of school shooting incidents per year, from 1971 through 2022, with each red dot representing a single incident.



<u>Figure 6.1:</u> *National Distribution of School Shootings (12/06/1971 - 06/01/2022)*

Some dots appear to be a darker red, as a gradient, than others. This means that there have been multiple incidents at or near the same location. In 52 years, schools in the continental United States have experienced over 2,000 school shootings.

Figure 6.2 shows the distribution of school shooting victim injuries over the span of 1971 through 2022.

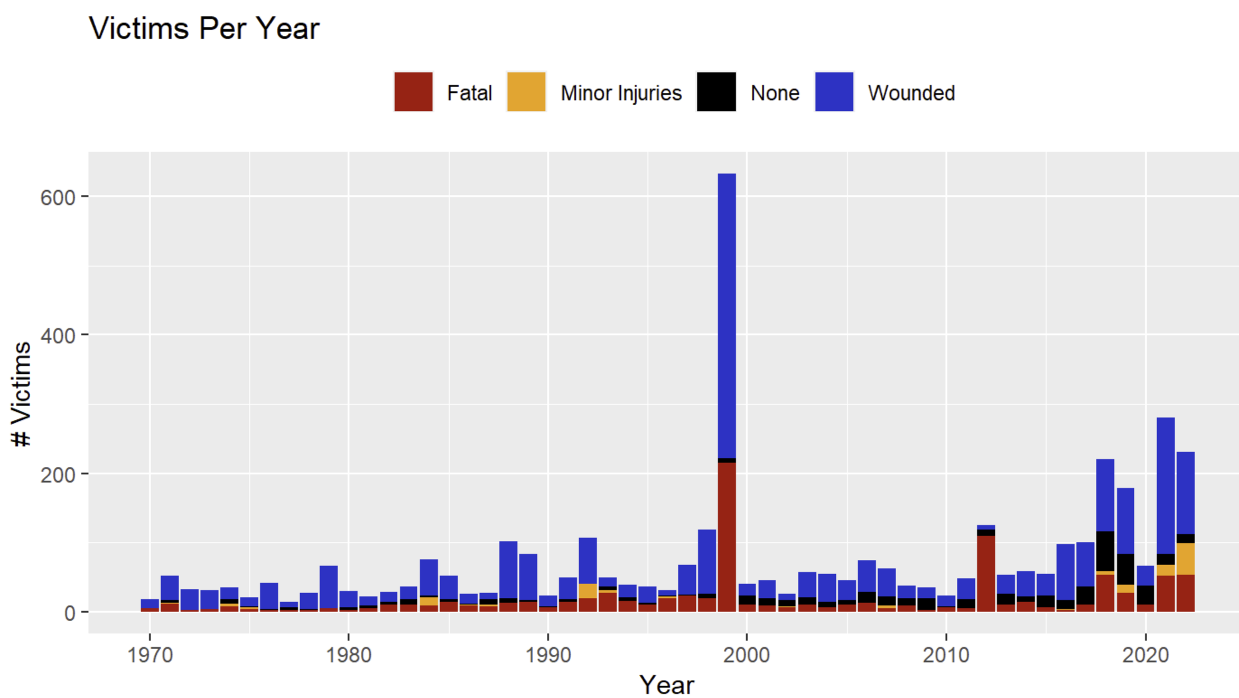


Figure 6.2: *School Shooting Victims Per Year (1971 - 2022)*

In 1999, there were over 600 people that were victims of school shootings. Narrowing down on that year, Figure 6.3 below illustrates that some of the affected states include Georgia and Colorado, notably for the Columbine High School Massacre.

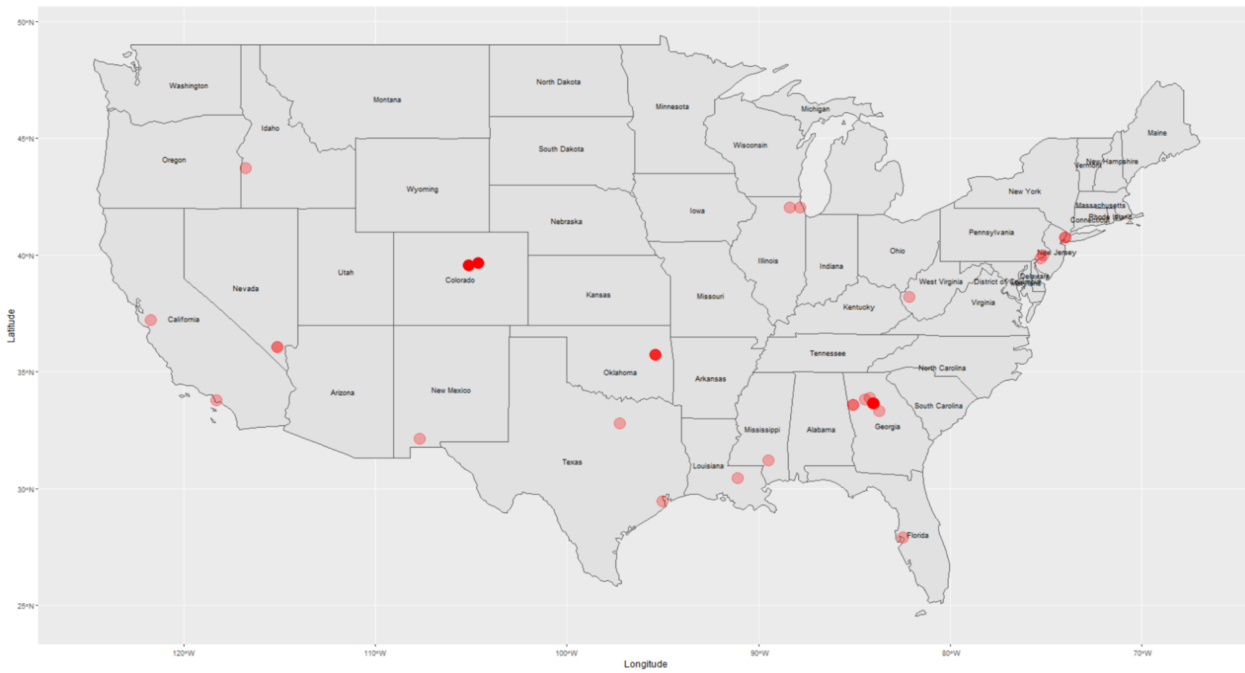


Figure 6.3: *National Distribution of School Shootings (1999)*

Removing the outlier that is 1999, allows for a closer look of Figure 6.2.

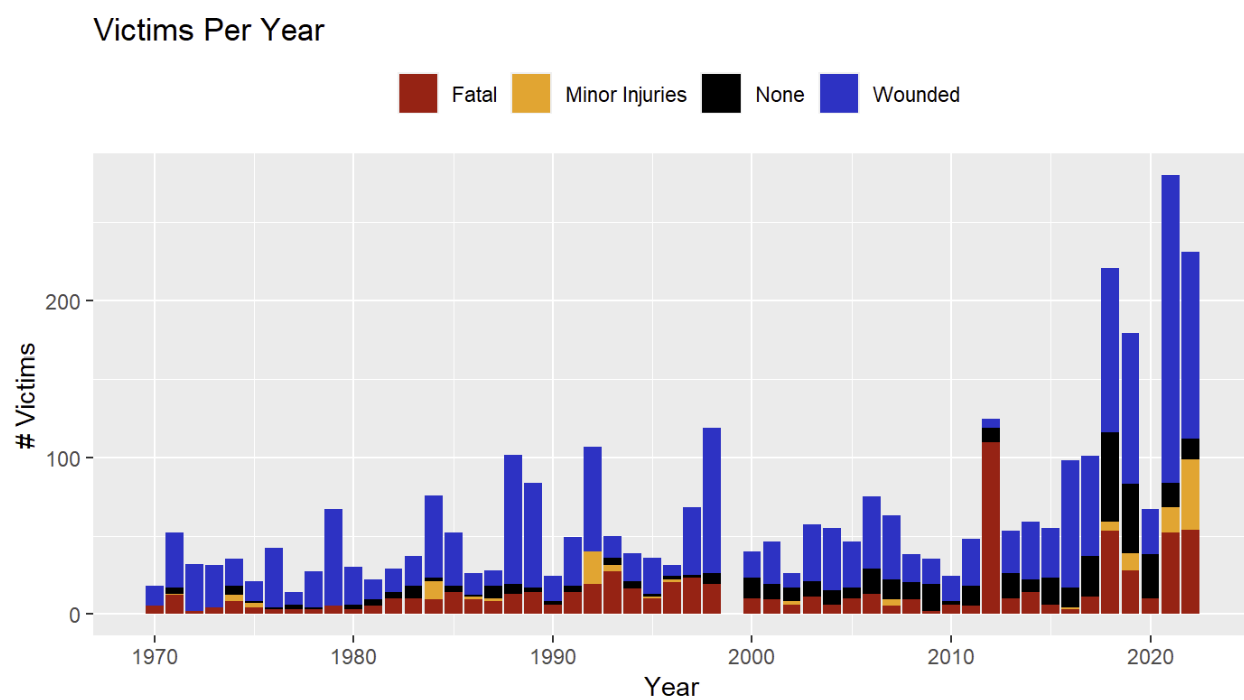
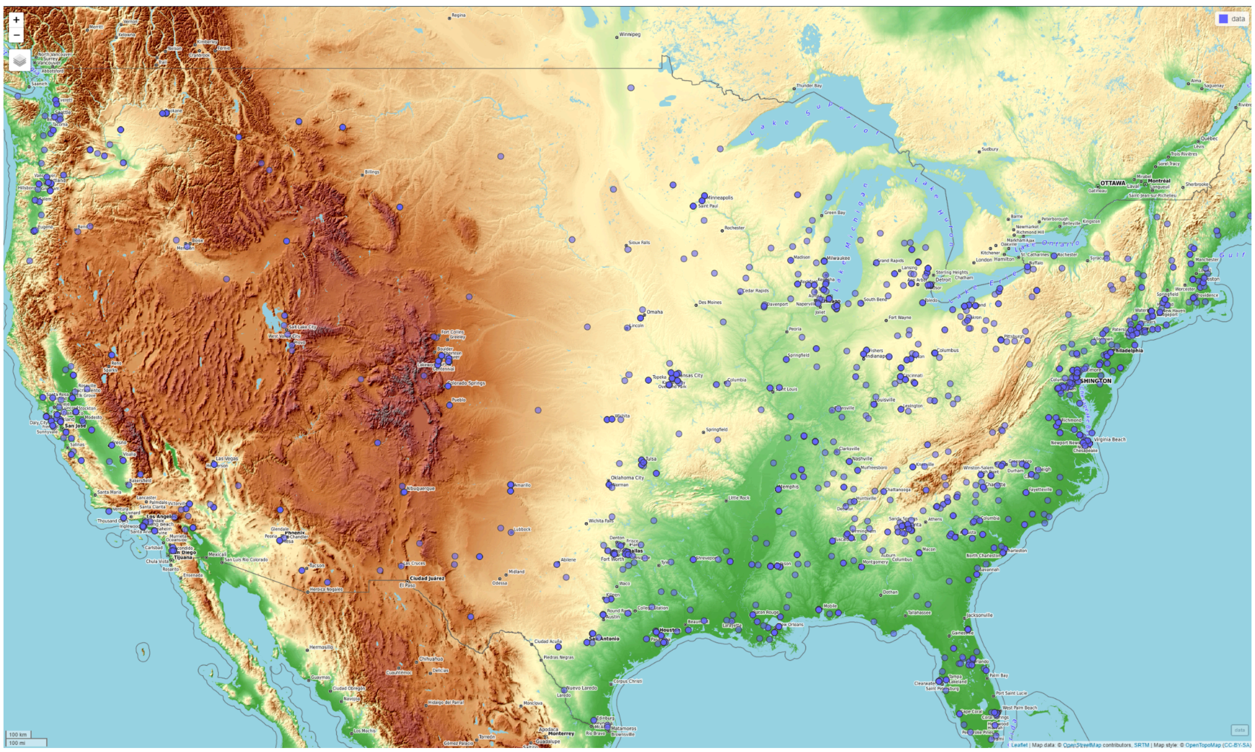


Figure 6.4: *School Shooting Victims Per Year (1971 - 1998, 2000 - 2022)*

Based on these initial observations, it seems that the geography of the school shootings may offer more intel into underlying patterns than time (years). Graphically, it appears that there are denser distributions of school shootings on the outside coasts outlining the United States, as well as the East Coast. In other words, school shootings seem to happen wherever the topography allows it.



<u>Figure 6.5:</u> *Topography Map School Shooting Overlay (1971 - 2022)*

The contour map of the intensity of school shootings, corroborates the above, as the contour lines show that shootings are sparse in the more mountainous areas.

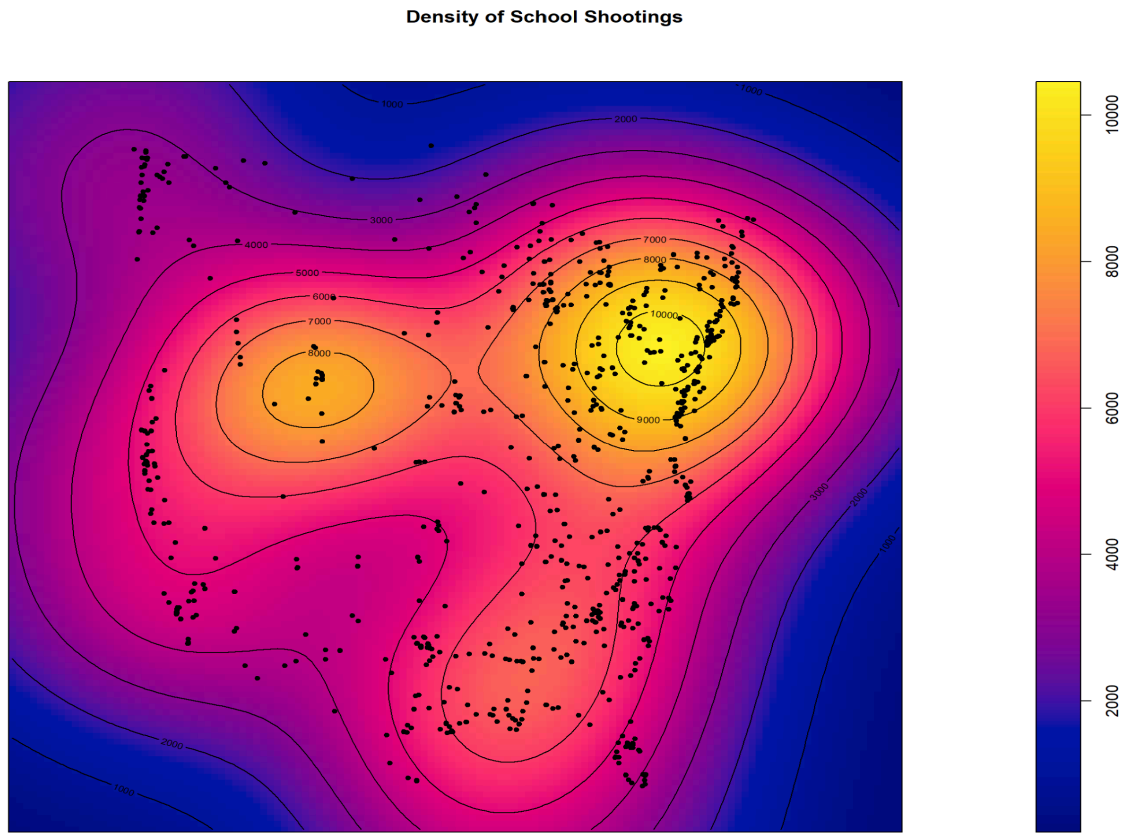


Figure 6.6: *Contour Density Map*

Therefore, this analysis aims to employ geostatistical methods on *spatial* point processes to identify the socioeconomic and geographic characteristics that help inform the distribution of school shooting incidents across the continental United States, rather than entertain *temporal* point processes or prevention methodologies.

6.2 Clustering Analysis

The simplest way to start addressing clustering possibilities within the data set is to perform kernel smoothing on the data. The resulting plot, seen in Figure 6.7, supports our initial observation of denser distributions of school shootings on the outside coasts outlining the United States, as well as the East Coast.

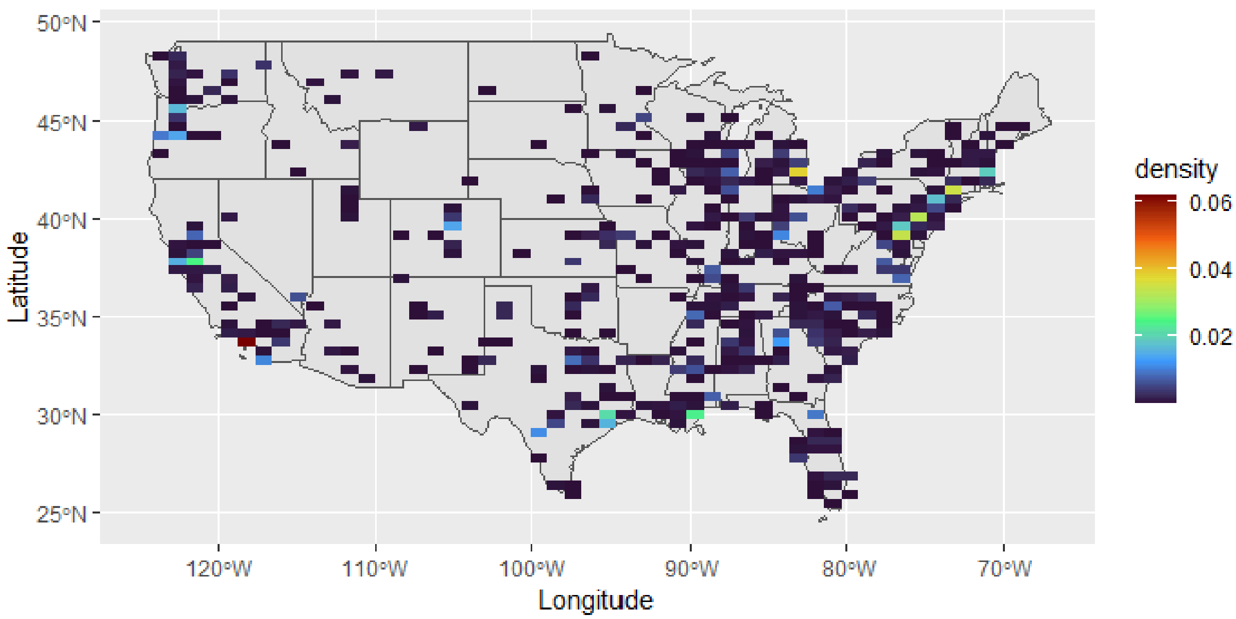


Figure 6.7: *Kernel Distribution of School Shootings (12/06/1971 - 06/01/2022)*

On the West Coast, there are two regions that display kernels of high density: the Los Angeles region in California, which has a red kernel representing the highest density, and the Sacramento region in California. On the East Coast, one area that stands out as a bright kernel is the Detroit region in Michigan, as well as the immediate and surrounding regions near the Tri-State area.

Additionally, plotting F, G, K, and J functions can be used to further measure and characterize the spatial distribution of school shootings. To do so, the data at hand needs to be normalized to fit in a window, as is shown in Figure 6.8 below.

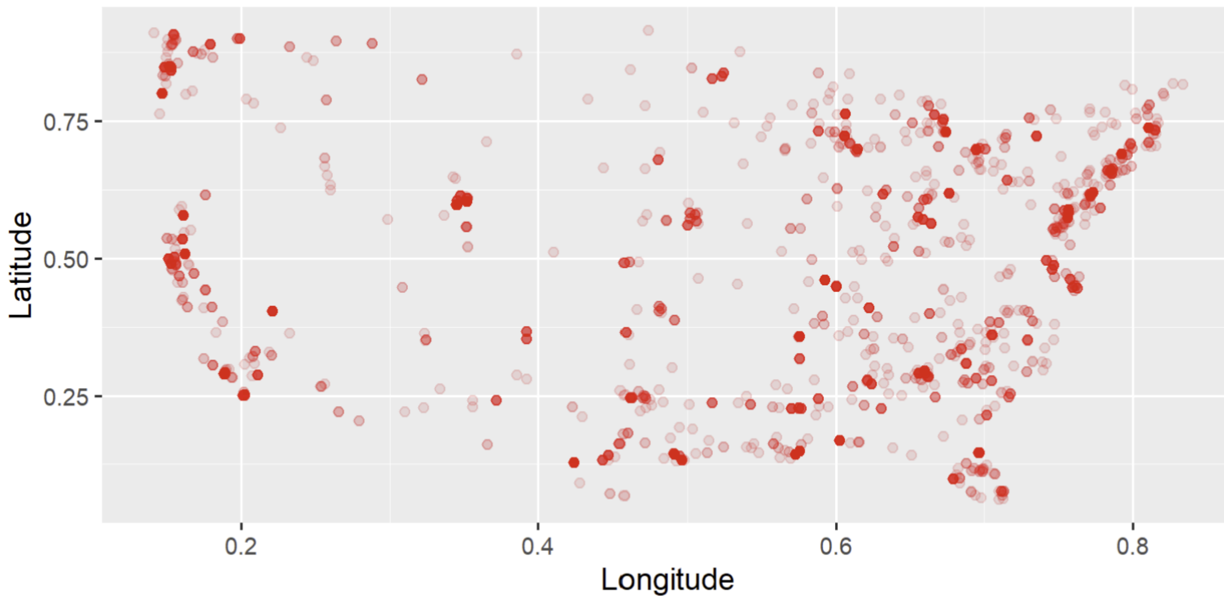


Figure 6.8: *National Distribution of School Shootings (Normalized Lat/Lon)*

The *spatstat* library in R was used to generate the four estimated function curves below, which appear to also support the presence of clustering. The graph includes $F_{pois}(r)$ for homogeneous Poisson processes, the Chiu-Stoyan estimate $F_{cs}(r)$, the reduced sample estimate (border correction) $F_{board}(r)$, and the Baddeley-Gill Kaplan-Meier estimate $F_{km}(r)$ [24]. The three edge corrected estimates are compared to the Poisson function to better understand the results of each individual function [24].

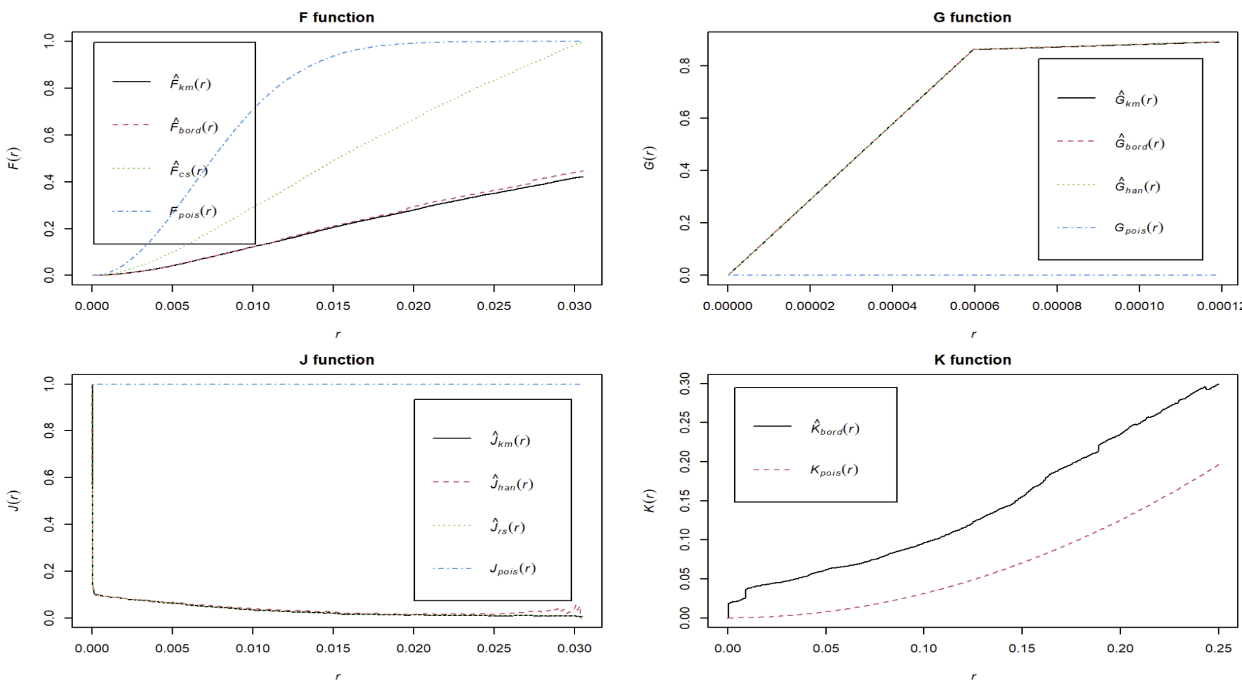


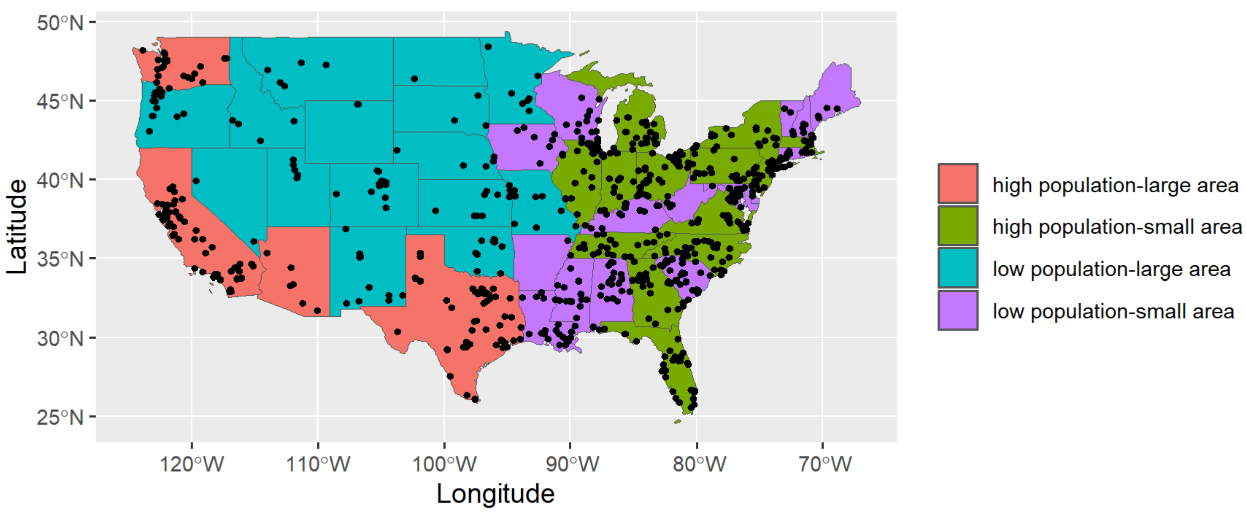
Figure 6.9: *F, G, K, J Plots -- School Shootings*

The top-left corner shows $F_{pois}(r)$ as blue dashes. All three edge correcting estimates fall below the Poisson curve, suggesting that clustering is present within our point pattern. The top-right corner shows $G_{pois}(r)$ as blue dashes. All three edge correcting estimates fall above the Poisson curve, implying that clustering is present within our point pattern. The bottom-left corner shows $J_{pois}(r)$ as blue dashes. All three edge correcting estimates fall below the Poisson curve, alluding that clustering is present within our point pattern. The bottom-right corner shows $K_{pois}(r)$ as orange dashes. The $\widehat{K}(r)$ estimate falls above the Poisson curve, indicating that clustering is present within our point pattern.

6.3 Exploring Covariates

6.3.1 Population

The overall goal of clustering is to identify groups of similar data points within a dataset. Analyzing the locations of the shootings may uncover underlying patterns that explain why these crimes are clustered in specific areas, as it is more likely for clusters to be based on patterns and similarities, rather than randomness. Figure 6.10 below depicts the distribution of school shootings as a function of the population and land area category of each state.



<u>Figure 6.10:</u> *School Shootings by Area Population*

Visually, it appears that regions with higher population per square-kilometer have a higher density of school shootings than areas with a lower population to square-kilometer ratio. Figure 6.11 below shows a supplementary figure, where school shootings by state are normalized to population density, by representing the amount of school shootings per 1 million people.

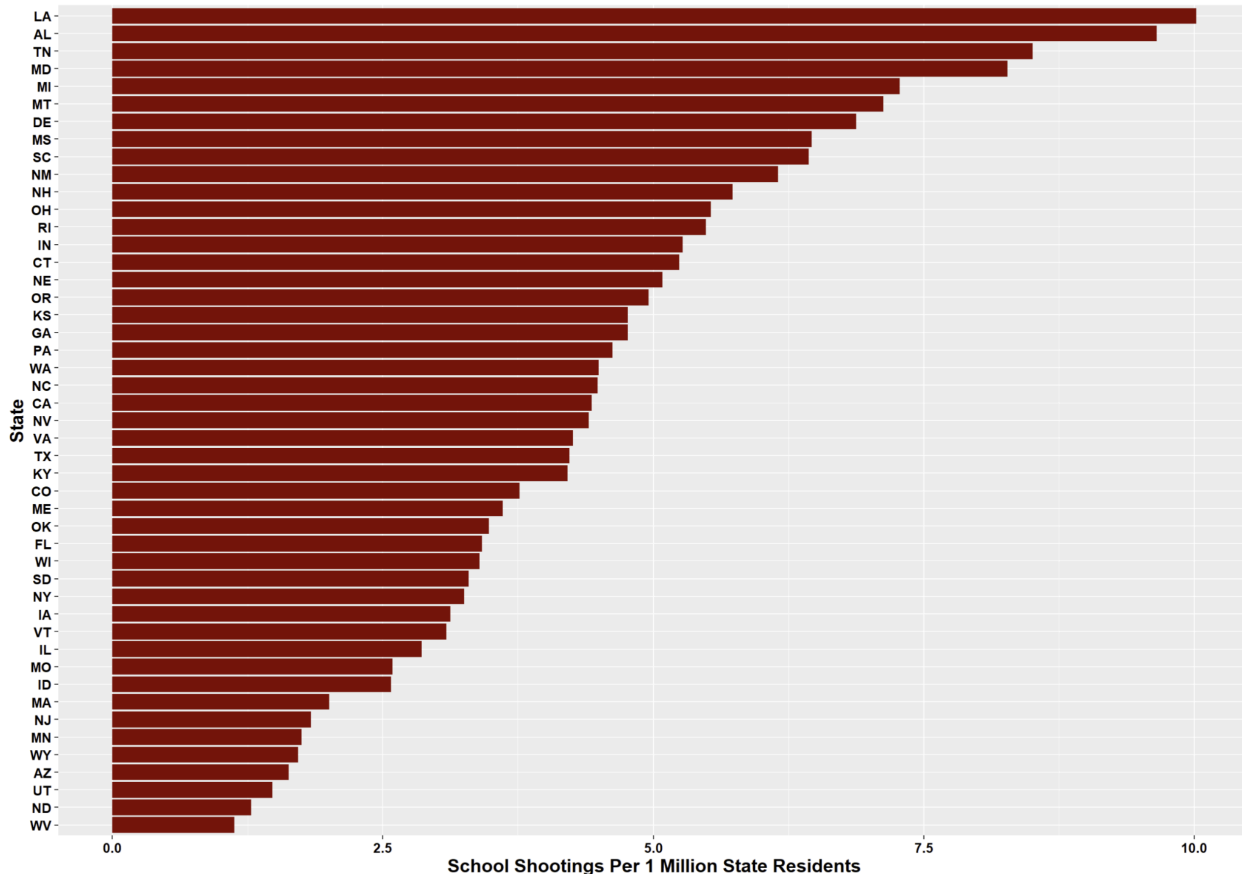


Figure 6.11: *School Shootings and Population, Normalized*

Louisiana and Alabama have the highest number of school shootings per million residents states in the continental United States.

6.3.2 Mental Health

Urbanization refers to the increase in the proportion of people living in towns and cities, and has been reported to influence mental health. Urbanization is linked to increased stressors and factors such as overcrowded and polluted environments, high levels of violence, and reduced social support [25].

To better understand how mental health may influence school shootings, the shooting distribution data were overlaid with data provided by Mental Health America, which shows statistics for the percentage of people, per state, who scored *at risk* for suicide, of those who took an MHA screening [26]. A disclaimer, all of the data presented in the MHA State and County Dashboard are collected through the MHA Online Screening Program, a collection of 11 free, anonymous, confidential, and clinically validated screens that are among the most commonly used mental health screening tools in clinical settings [26]. This means our data inherently has voluntary survey bias and the sample may not be representative of the population as a whole. Additionally, the mental health access disparities between states factor into the results (or lack thereof) publicly available. Having stated that, Figure 6.12, maps out the distribution of school shootings against the data for voluntary screeners that scored at-risk for suicide, per state.

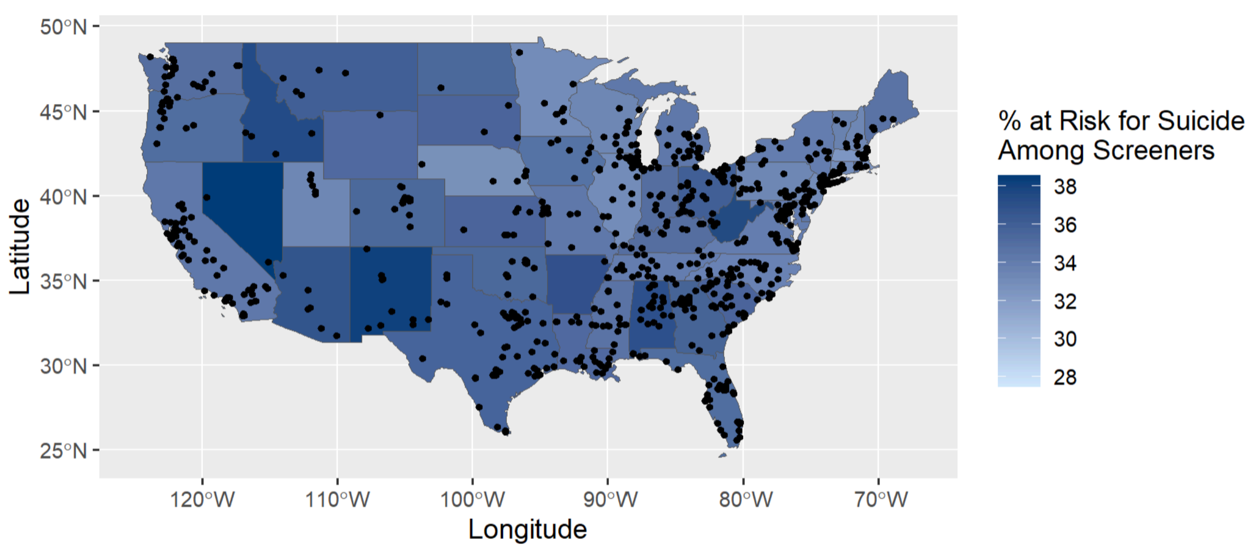


__Figure 6.12:__ *National Distribution of School Shootings and Suicide at-Risks*

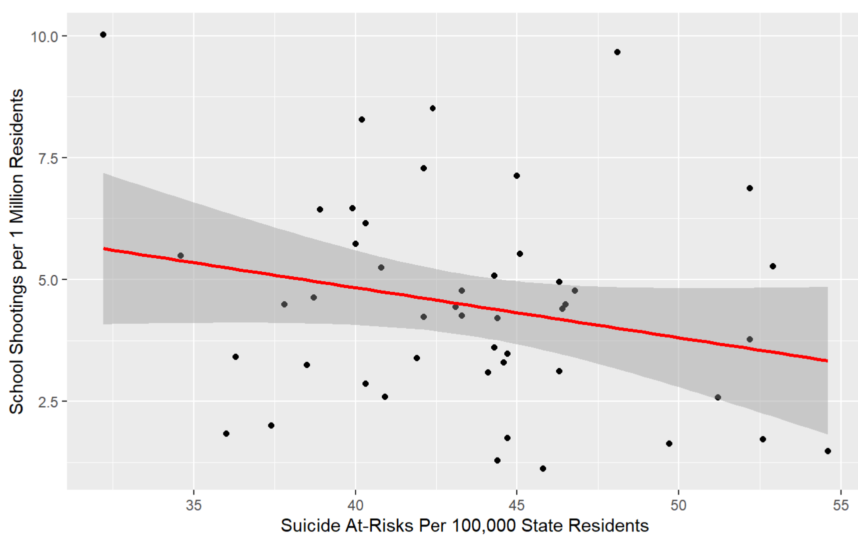
The states with the highest density of school shooters are not necessarily the states experiencing the most suicidal ideations. Below are the results of the linear regression model used to predict school shootings as a function of suicide at-risks per state:

```
Call:
lm(formula = suicide_guns$shooting_per_million ~ suicide_guns$suicide_per_100k)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.4104 -1.5309 -0.1742  1.0569  5.6509 

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)   
(Intercept)                    8.93618    2.72179   3.283  0.00199 **
suicide_guns$suicide_per_100k -0.10251    0.06206  -1.652  0.10553   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.112 on 45 degrees of freedom
Multiple R-squared:  0.05717,   Adjusted R-squared:  0.03622 
F-statistic: 2.729 on 1 and 45 DF,  p-value: 0.1055
```



<u>Figure 6.13:</u> *Linear Regression Model, School Shootings ~ Suicide At-Risks (top) and plot (bottom)*

Less than 1% of the variability found in the response variable can be explained by the predictor variable. The p-value is 0.11, and thus, suggests that the effect between school shootings per million state residents and the suicidal level of people per state are not statistically significant. Even though the single interaction between our variables is not significant, there may be other confounding variables at play that can have a strong influence on the frequency of school shootings.

6.3.3 Guns

Shifting focus to the elephant in the room, is it the guns that are killing the students? One could assume that the states with the highest rates of school shootings would be the states with the most guns per capita. Figure 6.14, graphically displays that California has the most school shootings, followed by Texas.

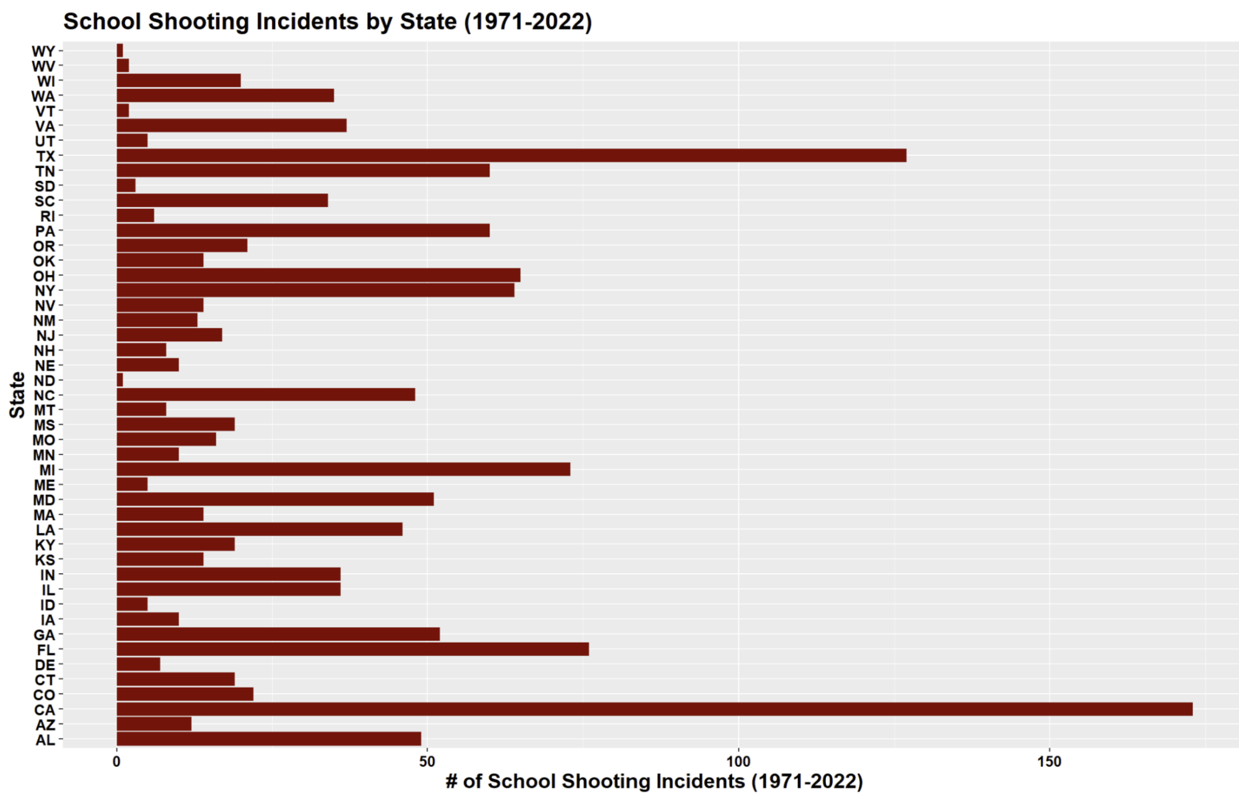


Figure 6.14: *School Shootings Per State (1971-2022)*

Surprisingly, Figure 6.15 shows us that Wyoming is actually the state with the highest guns per capita, not Texas or California. Another thing to note, the *Guns Per Capita* dataset from the World Population Review did not have data for Arkansas, as most firearms in Arkansas do not require registration [27].

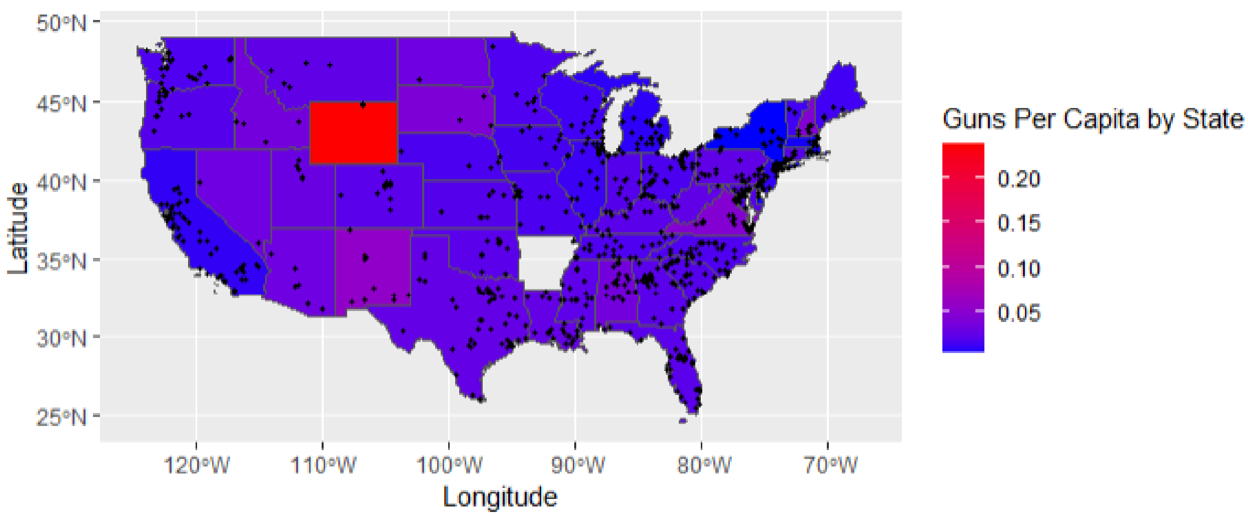


Figure 6.15: *National Distribution of School Shootings and Total Registered Guns per Capita*

Below are the results of the linear regression model used to predict school shootings as a function of guns per capita per state. The first model (Figure 6.16) includes Wyoming, and the second model (Figure 6.17) omits it, as the state is such a huge outlier. Although the regression model changed after omitting Wyoming and normalizing for population changes, observably readjusting the "downward trend", the interpretation is still the same.

Gun ownership accounts for a negligible percentage of the variability found in the response variable. The p-value is greater than 0.05, and thus the effect between school shootings and guns per capita per state are not statistically significant. This conclusion is to be taken with a grain of salt. Just because some results show that more guns does not necessarily lead to increases in school shootings, it does not mean that everyone should be allowed to have guns. There are far more confounding variables behind the scenes.

```
Call:
lm(formula = shooting_per_million ~ GunsPerCapitaRegisteredWeaponsPer1000Residents,
    data = us_guns2)

Residuals:
    Min      1Q  Median      3Q     Max
-3.3596 -1.4510 -0.0629  0.8742  5.5589

Coefficients:
                                                Estimate Std. Error t value
Pr(>|t|)
(Intercept)                                     4.733433   0.403408  11.734
0.000000000000000276
GunsPerCapitaRegisteredWeaponsPer1000Residents -0.009897   0.009504  -1.041
0.303

(Intercept)                                    ***
GunsPerCapitaRegisteredWeaponsPer1000Residents
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.149 on 45 degrees of freedom
Multiple R-squared:  0.02353,   Adjusted R-squared:  0.001832
F-statistic: 1.084 on 1 and 45 DF,  p-value: 0.3033
```

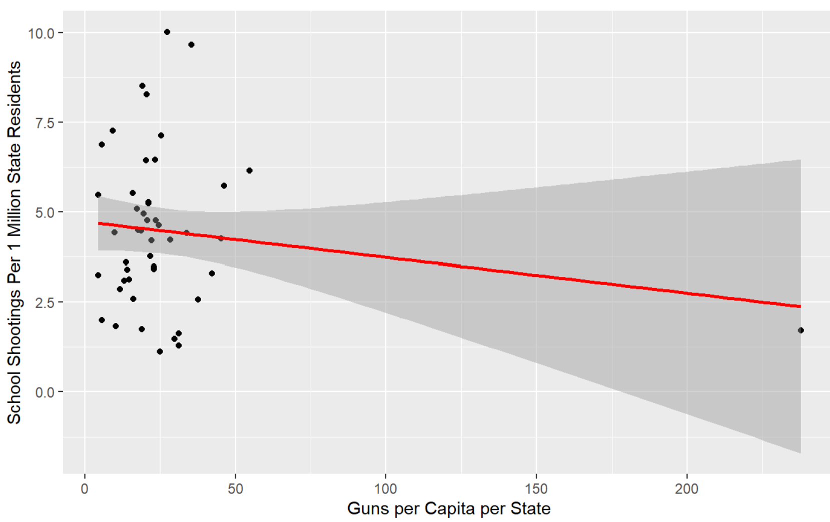


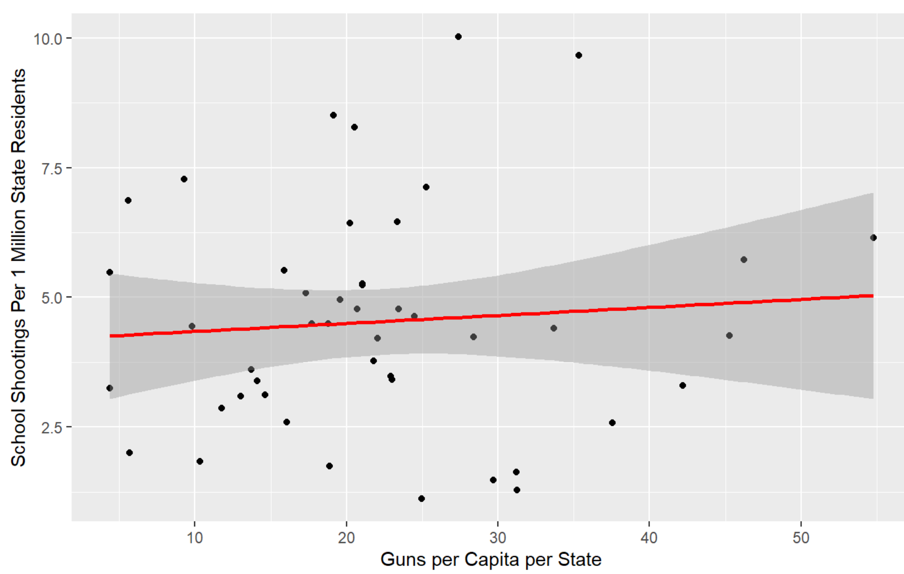
__Figure 6.16:__ Linear Regression Model, School Shootings ~ Guns per Capita per State (top) and plot (bottom)

```
Call:
lm(formula = shooting_per_million ~ GunsPerCapitaRegisteredWeaponsPer1000Residents,
    data = guns_wy_x)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4460 -1.2954 -0.1461  1.0307  5.4113

Coefficients:
                                               Estimate Std. Error t value   Pr(>|t|)
(Intercept)                                     4.18813    0.70964   5.902 0.00000047 ***
GunsPerCapitaRegisteredWeaponsPer1000Residents  0.01540    0.02869   0.537      0.594
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.152 on 44 degrees of freedom
Multiple R-squared:  0.006505,  Adjusted R-squared:  -0.01607
F-statistic: 0.2881 on 1 and 44 DF,  p-value: 0.5941
```



<u>Figure 6.17:</u> *Linear Regression Model Modified (top) and plot (bottom)*

Comparing guns per capita to the actual school shootings per state, normalized by population, paints a different picture. Using the same dataset from above to expand on *Figure 6.14*, the map in Figure 6.18 tells a different story. Here, neither California and Texas, nor Wyoming, are the most affected states.

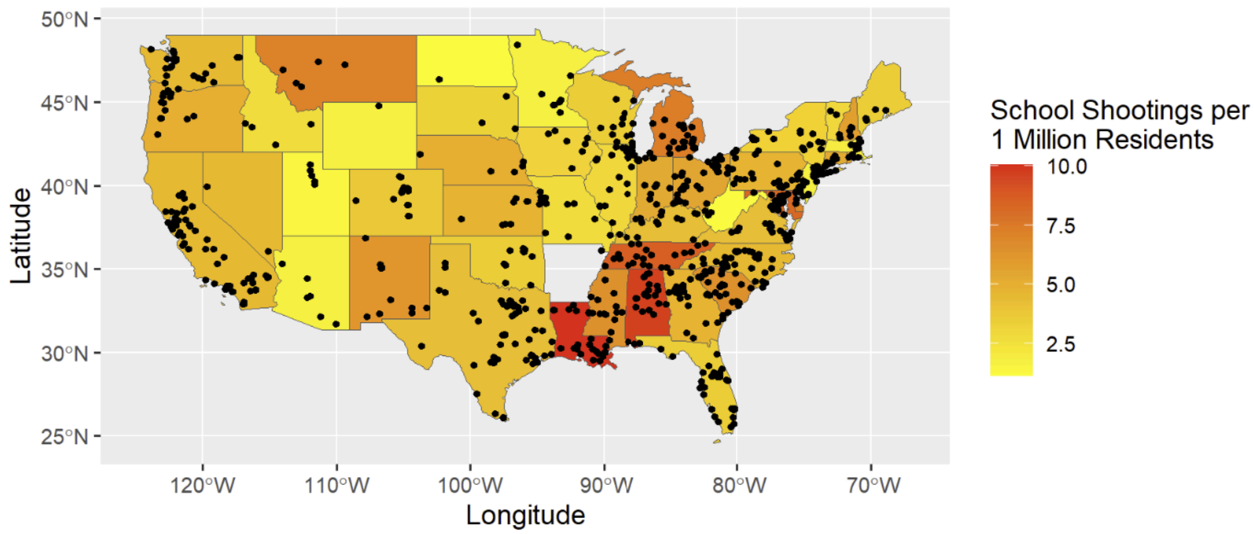
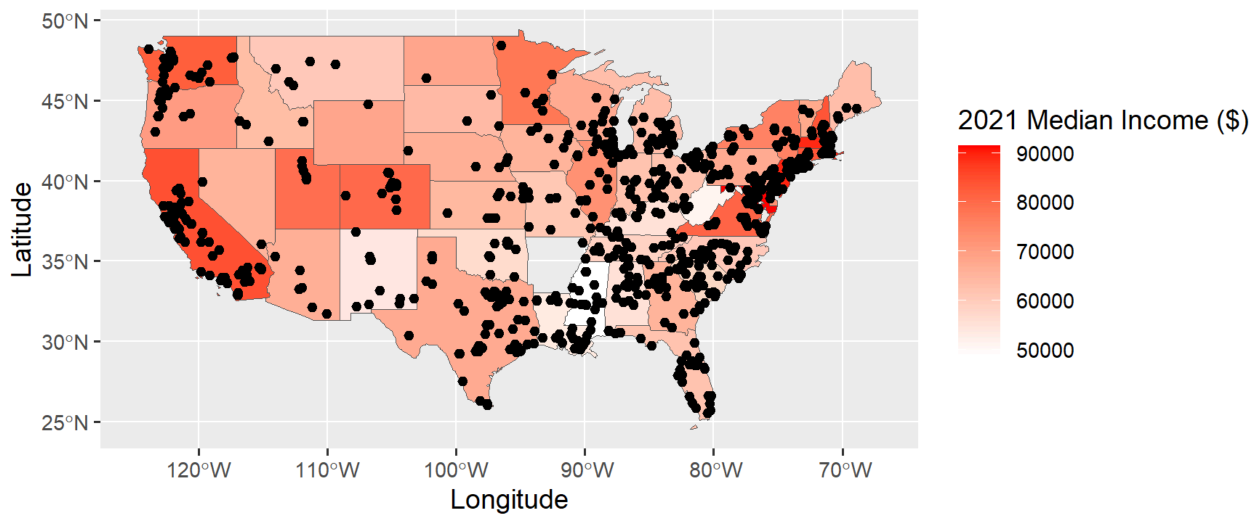


Figure 6.18: *National Distribution of School Shootings per 1 Million State Residents*

6.3.4 Income

The penultimate socioeconomic factor investigated in this analysis is median income per state. The American Community Survey (ACS) data are available through the *tidycensus* package in R, and represents an annual sample of approximately 3 million households [28]. The map below shows the median income per state, as of 2021, overlaying the distribution of school shootings. From this visualization, it is not clear what effect, if any, income has on the frequency of school shootings.



<u>Figure 6.19:</u> *School Shootings and Median Income*

Figure 6.20 below also gives us insight into the same distribution, in a different light. Bigger dots are correlated to a higher median income, and redder marks are correlated to higher school shooting rates. it is interesting to note that the states with the reddest points (most school shootings per million residents) are not the most wealthy states (have smaller points).

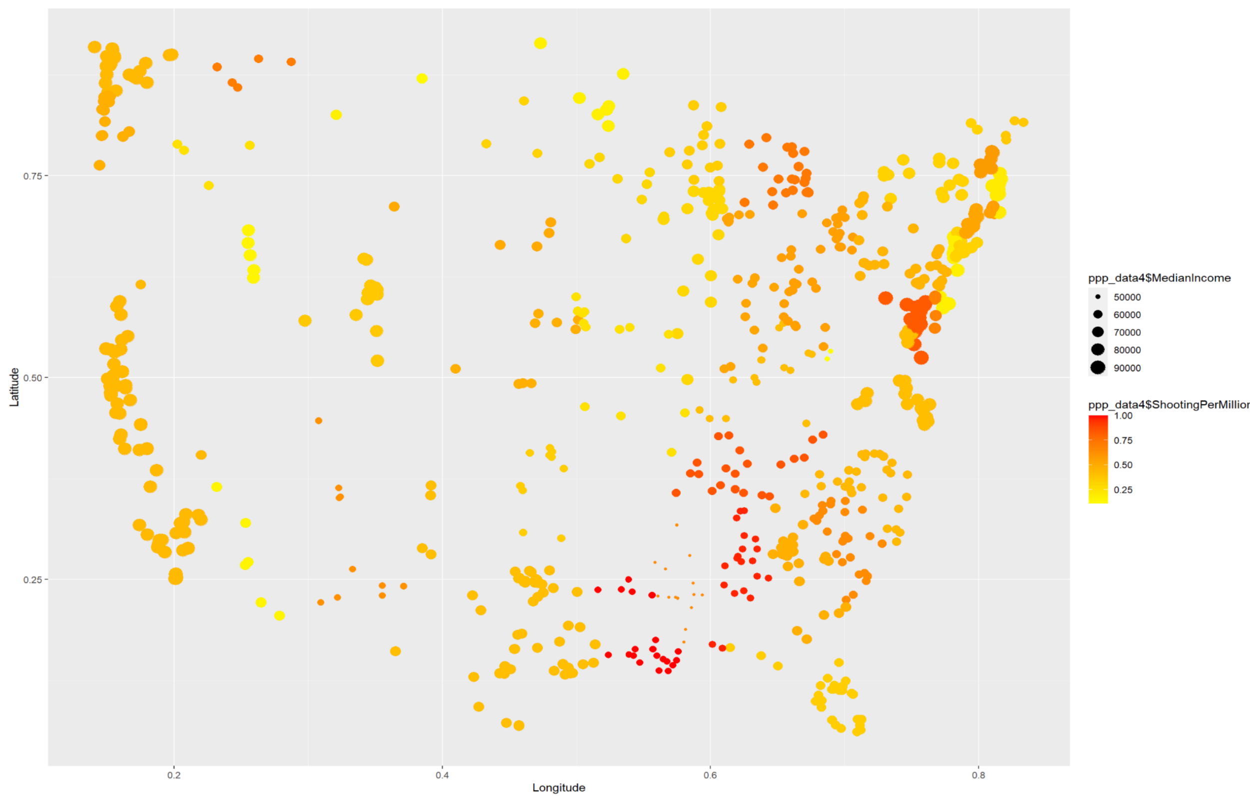


__Figure 6.20:__ School Shootings and Median Income, Visual Effects

The results of the linear regression model below show that income is not a significant variable individually used to cause an effect in school shooting frequency. Even though there appears to be a downward-like trend, where states with lower median income appear to have slightly more school shootings, less than 1% of the variability found in the response variable can be accounted for by the predictor variable. The p-value is 0.105, and thus, suggests that the effect between school shootings per million state residents and the income level of people per state are not statistically significant.

```
Call:
lm(formula = shooting_per_million ~ estimate, data = us_income)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.1776 -1.4201 -0.1118  1.0358  4.9411 

Coefficients:
               Estimate  Std. Error t value Pr(>|t|)    
(Intercept)  7.77948608  2.02384638   3.844 0.000378 ***
estimate    -0.00004864  0.00002939  -1.655 0.104885    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.112 on 45 degrees of freedom
Multiple R-squared:  0.05737,	Adjusted R-squared:  0.03643 
F-statistic: 2.739 on 1 and 45 DF,  p-value: 0.1049
```

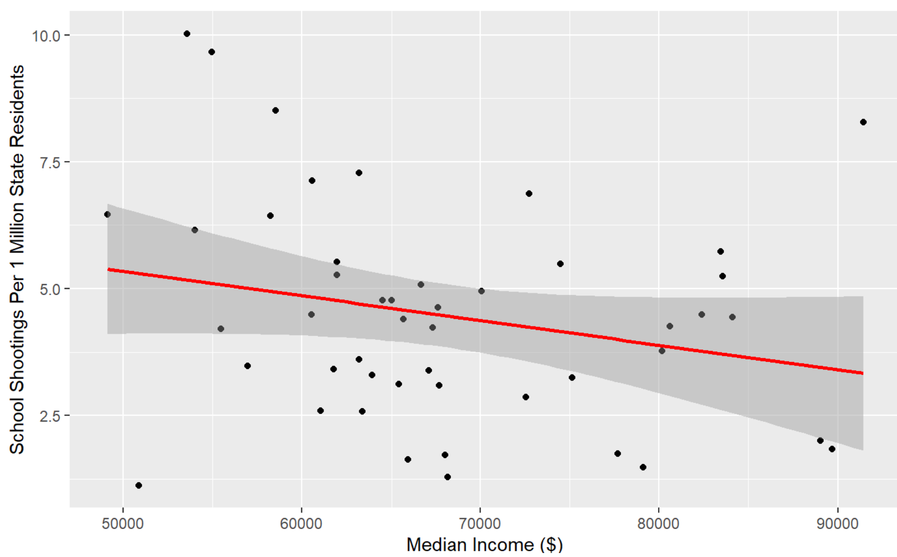


Figure 6.21: *Linear Regression Model, School Shootings ~ Median Income (top) and plot (bottom)*

6.3.5 Homicide

In 2021, 78% of homicides were committed with a firearm which is the highest proportion since 1980 [29]. This highlights the significant role firearms play in violent incidents, and therefore, the final socioeconomic factor explored in this analysis is homicides per capita. The data, acquired through the National Center for Health Statistics, provides information on state homicide rates per 100,000 state residents [30].

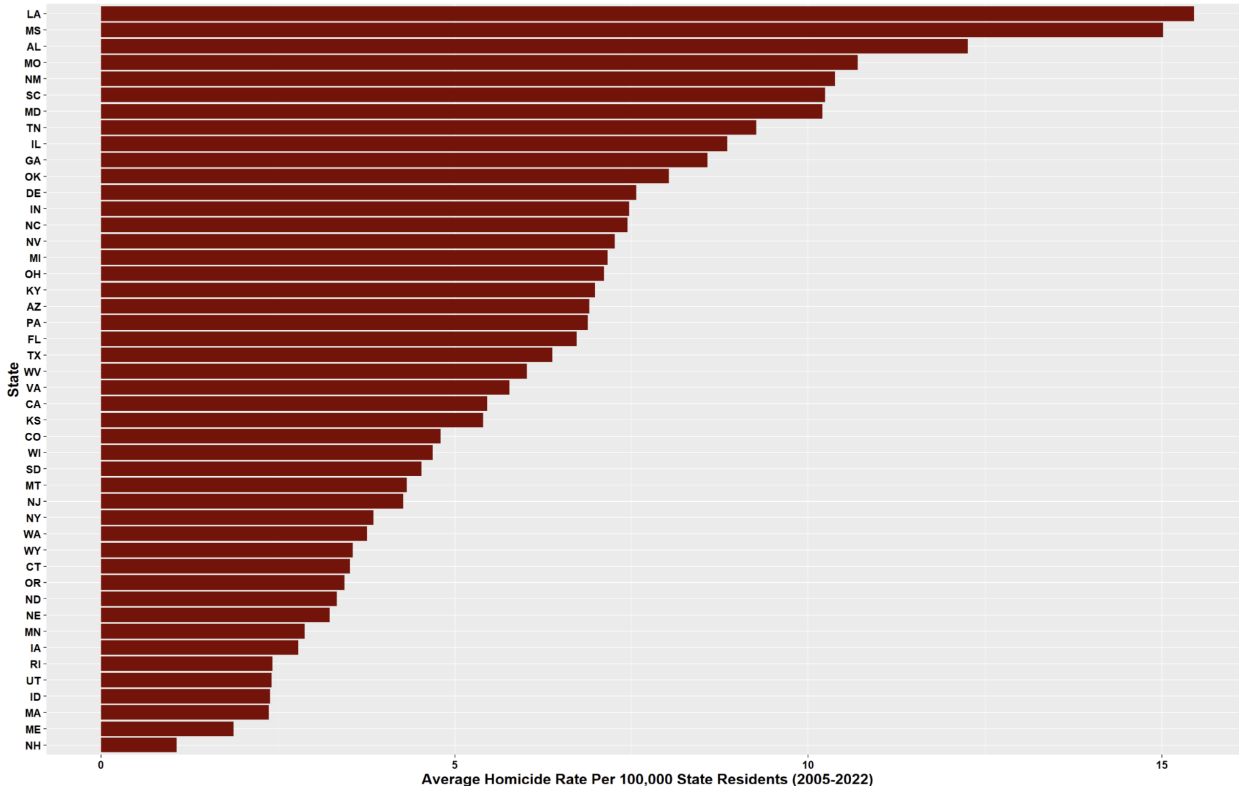
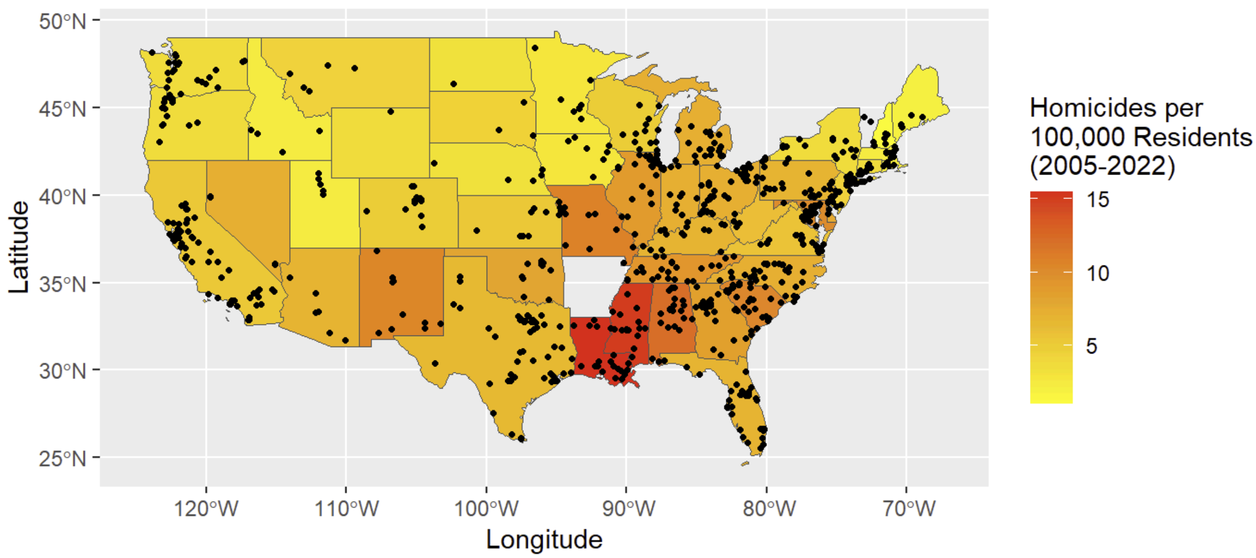


Figure 6.22: *Average Homicide Rate per 100,000 State Residents*

Louisiana and Mississippi are the states with the highest homicide rates. What is really intriguing is the eerie similarity between Figure 6.18 and the map below, which overlays the information from Figure 6.22 and the distribution of school shootings. What is going on in Louisiana that is making it the state with the highest homicide and school shooting rates?

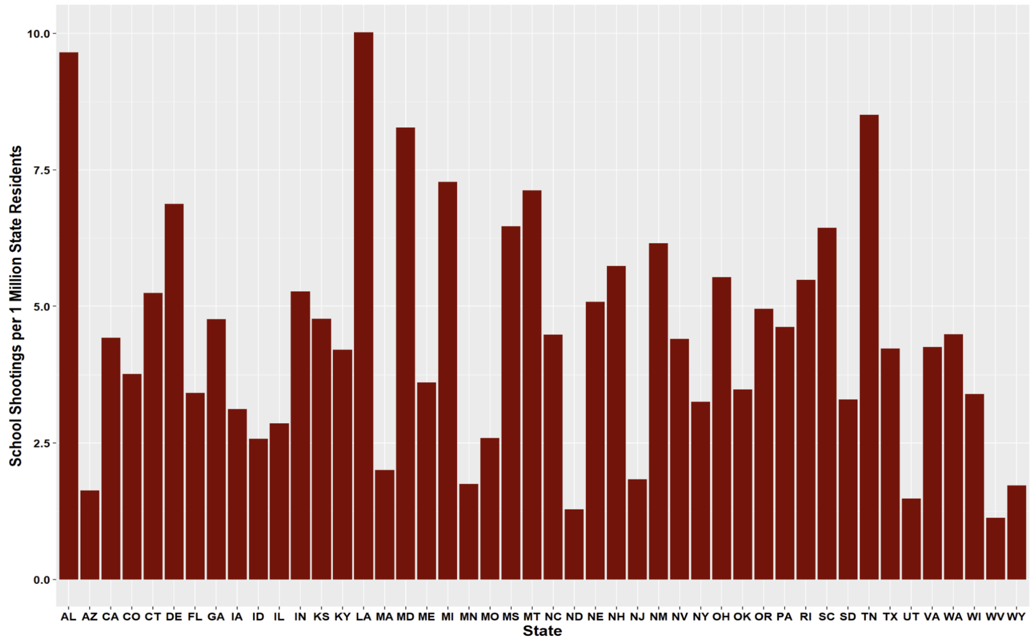


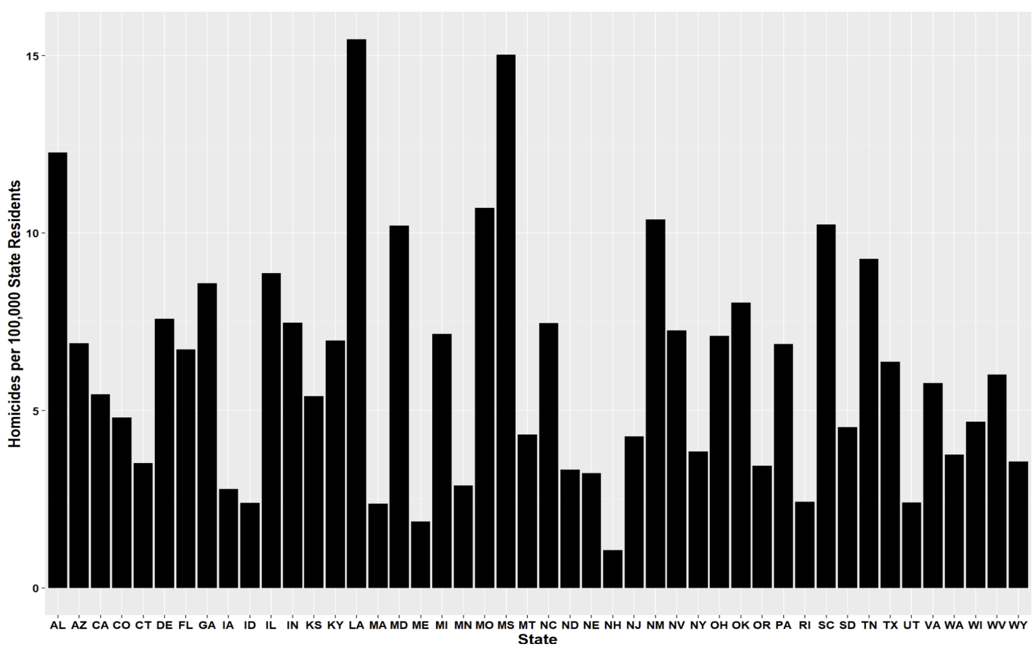
<u>Figure 6.23:</u> *Average Homicide Rate per 100,000 State Residents Map*

The distribution of state homicide and school shooting rates can help further explore this relationship. The histograms below help visualize the overall shape and counts of a distribution, while the density plot compares the overall shape of both distributions.

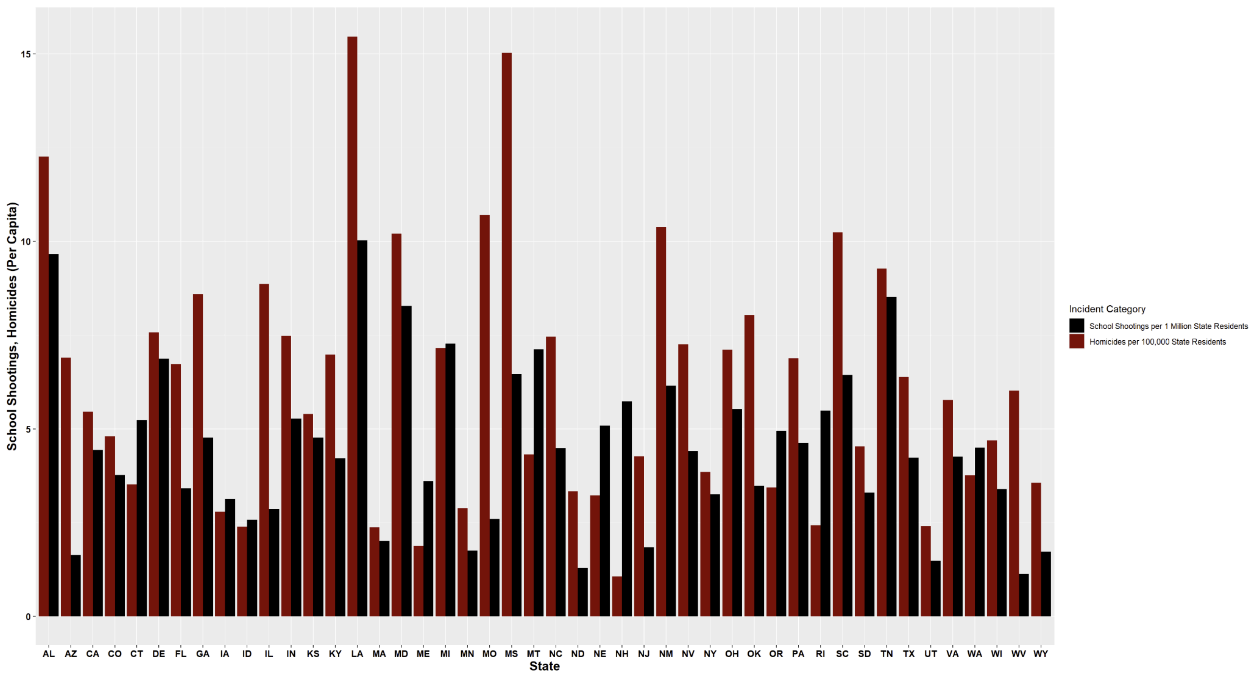
The frequency of incidents per state, as shown in Figure 6.24, seem to follow similar patterns. Additionally, the grouped bar plots in Figure 6.25 seemed to be within relative distance of each other, aside from Louisiana and Mississippi, as mentioned previously.

Furthermore, the density plots in Figure 6.26 are both right-skewed. As a nation, the United States experiences about 4.5 school shootings per million residents and 6.2 homicides per 100,000 residents. The school shooting density plot is mainly unimodal, while the homicide density plot is bimodal at around 3.5 and 6.5, suggesting clustering within the homicide data itself. The spread of both densities, however, are quite similar. Despite differences within the modal shape, the similarities in spreads, tails, and means, can suggest that the two datasets are quite similar in their general distribution characteristics.

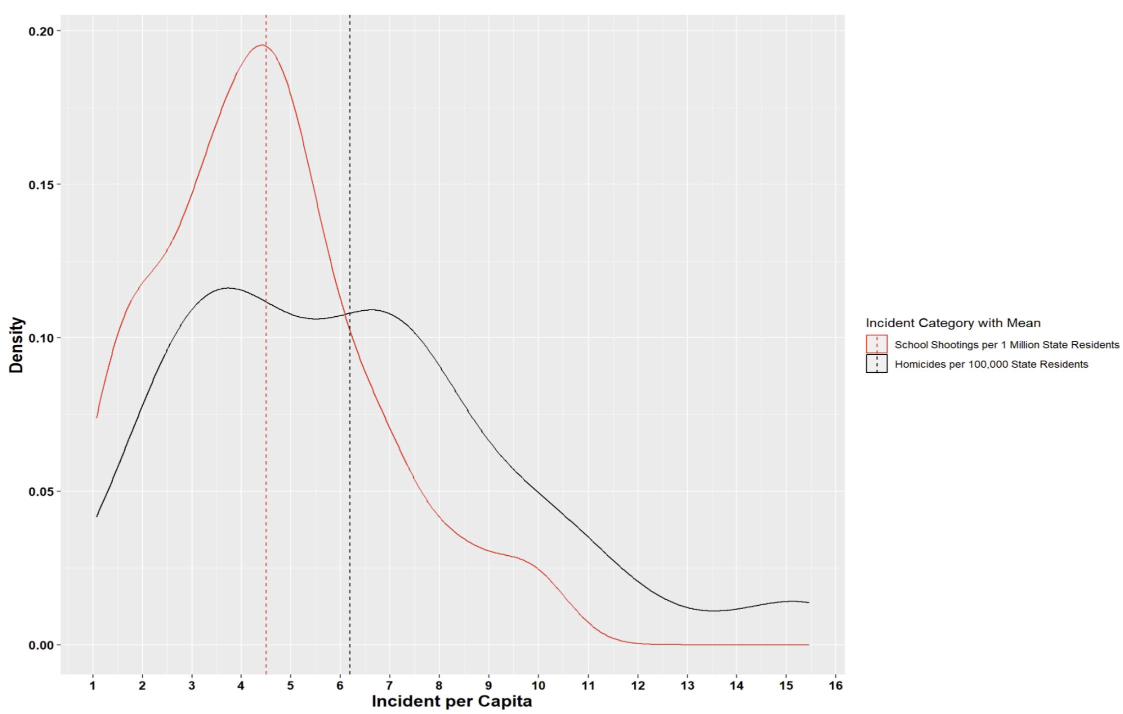




<u>Figure 6.24:</u> *Histograms: School Shootings per Capita (top), Homicides per Capita (bottom)*



<u>Figure 6.25:</u> *Grouped Histogram: School Shootings per Capita, Homicides per Capita*



<u>Figure 6.26:</u> *Density Plot: School Shootings per Capita, Homicides per Capita*

It is important to note that the per capita scale for homicides is 100,000 and the per capita scale for school shootings is 1 million. Therefore, school shootings are scaled to be one-tenth as common as homicides.

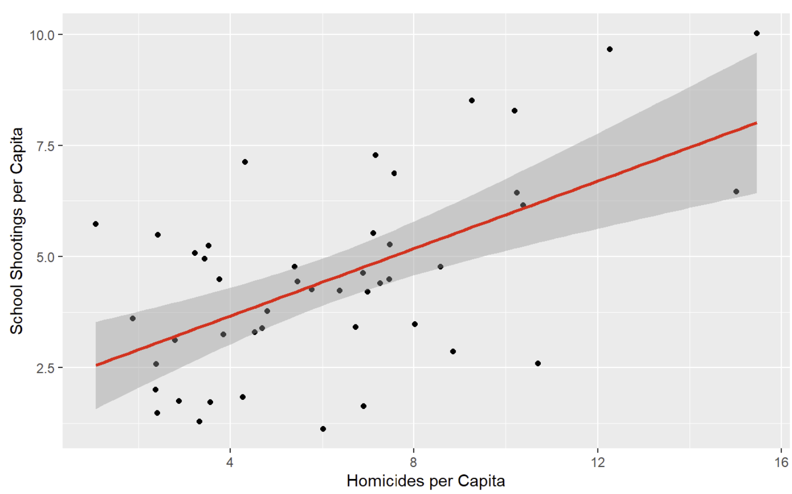
Running a regression model to continue exploring this relationship confirms the significant causal effect between both variables. Homicides account for over 30% of the variability seen in school shootings per capita, and the positive linear-like relationship suggests that places with more homicides tend to have more school shootings

```
Call:
lm(formula = shooting_per_million ~ HomicideRate, data = us_homicides)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6170 -1.2215 -0.1676  1.3542  3.3371 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.14927    0.56006   3.838 0.000393 ***
HomicideRate  0.37921    0.07984   4.749 0.000022 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.78 on 44 degrees of freedom
Multiple R-squared:  0.3389,	Adjusted R-squared:  0.3239 
F-statistic: 22.56 on 1 and 44 DF,  p-value: 0.00002195
```



<u>Figure 6.27:</u> *Linear Regression Model: School Shootings per Capita ~ Homicides per Capita*

The Generalized Additive Model (GAM) below is used to account for non-linearities between the response and predictor variables.

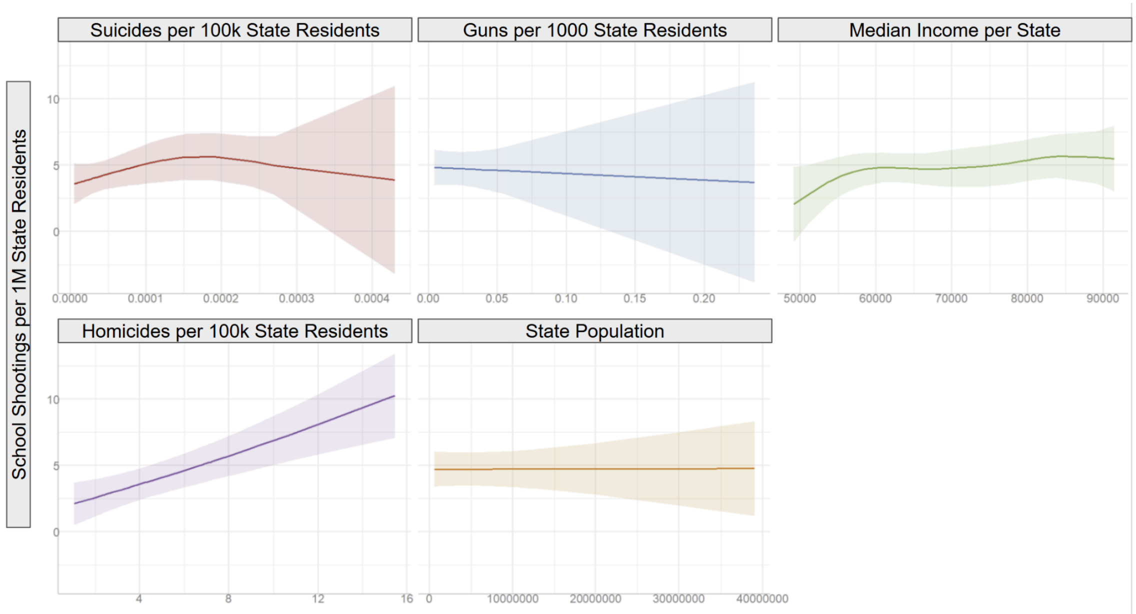


Figure 6.28: *Generalized Additive Model, School Shootings ~ All Variables*

The GAM model shown below illustrates that homicide rates are the most indicative factor for school shootings, when compared to the other socioeconomic factors used in this study.

```
Family: gaussian
Link function: identity

Formula:
shooting_per_million ~ s(ratio_suicide) + s(guns2pop) + s(estimate) +
    s(HomicideRate) + s(TotalPop)

Parametric coefficients:
            Estimate Std. Error t value            Pr(>|t|)
(Intercept)   4.4990     0.2523   17.83 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                   edf Ref.df      F  p-value
s(ratio_suicide) 1.978  2.509  1.076 0.351801
s(guns2pop)      1.000  1.000  0.075 0.785664
s(estimate)      3.725  4.583  0.877 0.450097
s(HomicideRate)  1.318  1.558 13.760 0.000428 ***
s(TotalPop)      1.000  1.000  0.001 0.978701
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.375   Deviance explained = 50.1%
GCV = 3.7437  Scale est. = 2.9281    n = 46
```

Figure 6.29: *Generalized Additive Model, School Shootings ~ All Variables*

CHAPTER 7

Conclusion and Further Research

This analysis explored patterns and potential causes behind school shootings across the continental United States. Through multiple statistical analyses, we observed clustering patterns on the coasts and urban areas which confirmed that shootings are not evenly distributed across the country. A range of factors that might influence where and how often these shootings occur were also explored. Homicide rates showed a meaningful relationship with school shootings, but did not explain everything. Other factors, like mental health, income, and population density may also play a role, though they were not as strongly linked to the frequency of school shootings.

Given the role of guns in school shootings, the rates of guns per capita per state also emerged as an interesting point of discussion. While states like California and Texas had the most school shootings, they did not have the highest rates of guns per capita. In fact, Wyoming, with the highest guns per capita, did not see a corresponding increase in school shootings, suggesting that gun availability alone is not the main driver.

The analysis of homicide rates alongside school shooting rates reveals a significant relationship between the two variables. States with higher homicide rates, particularly Louisiana, also experience a greater frequency of school shootings, suggesting that areas with higher violence levels are more likely to see increased school shooting incidents.

Overall, the findings highlight that school shootings are influenced by a combination of factors, including geography and potential socioeconomic influences. Addressing this complex issue will require a multifaceted approach that takes into account the interplay of these various elements.

REFERENCES

[1] Gramlich, John. “What the Data Says about Gun Deaths in the U.S.” Pew Research Center, Pew Research Center, 26 Apr. 2023, www.pewresearch.org/short-reads/2023/04/26/what-the-data-says-about-gun-deaths-in-the-u-s/.

[2] Soken-Huberty, Emmaline. “15 Root Causes of Gun Violence.” Human Rights Careers, 12 Nov. 2023, www.humanrightscareers.com/issues/root-causes-of-gun-violence/.

[3] Wamser-Nanney, Rachel, et al. “Childhood trauma exposure and gun violence risk factors among victims of gun violence.” Psychological Trauma: Theory, Research, Practice, and Policy, vol. 11, no. 1, Jan. 2019, pp. 99–106, https://doi.org/10.1037/tra0000410.

[4] Chloe. “The Crucial Role of School in Child Development.” LPE Ho Chi Minh, 29 Mar. 2024, www.lpehochiminh.com/en/role-of-school-in-child-development/.

[5] “Child-Related Gun Violence and School Shooting Facts.” Sandy Hook Promise, https://www.sandyhookpromise.org/blog/gun-violence/facts-about-gun-violence-and-school-shootings/.

[6] Rodrigues, Isidoro. “95% of Public Schools Conduct Active Shooter Drills. Are Students Safer?” YouthToday, 2023,

https://youthtoday.org/2023/12/95-of-public-schools-conduct-active-shooter-drills-are-students-safer/#:~:text=%E2%80%9CWe're%20just%20traumatizing%20another,for%20protecting%20the%20school%20community.%E2%80%9D.

[7] Kryzanek, Michael. "Guns, Violence and the Second Amendment." Bridgewater State University, 2023, https://www.bridgew.edu/stories/2023/guns-violence-and-second-amendment.

[8] "Shooting Incidents at K-12 Schools (Jan 1970-Jun 2022)." CHDS School Shooting Safety Compendium, CHDS, 2 Mar. 2023, www.chds.us/sssc/data-map/

[9] Yin, Thai. "Download: ZIP Code Latitude Longitude City State County CSV." Download: Zip Code Latitude Longitude City State County CSV, Gaslamp Media, 19 Jan. 2020, https://docs.gaslamp.media/download-zip-code-latitude-longitude-city-state-county-csv/

[10] "Safety." National Center on Safe Supportive Learning Environments (NCSSLE), safesupportivelearning.ed.gov/topic-research/safety. Accessed 3 Sep. 2024.

[11] Katsiyannis, A., Rapa, L.J., Whitford, D.K. et al. An Examination of US School Mass Shootings, 2017–2022: Findings and Implications. Adv Neurodev Disord 7, 66–76 (2023). https://doi.org/10.1007/s41252-022-00277-3

[12] Irwin, V., Wang, K., Cui, J., & Thompson, A. (2022). Report on indicators of school crime and safety: 2021 (NCES 2022–092/NCJ 304625). National Center for Education Statistics, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC. Retrieved July 21, 2022 from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2022092

[13] Schoenberg, Frederic Paik. Introduction to Point Processes., www.stat.ucla.edu/~frederic/papers/encycpiece.

[14] Daley, Daryl J., and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer, 2005.

[15] González, Jonatan A., et al. "Spatio-Temporal Point Process Statistics: A Review." Spatial Statistics, vol. 18, Nov. 2016, pp. 505–544, https://doi.org/10.1016/j.spasta.2016.10.002.

[16] Rossiter, D G. "Tutorial: Spatial Point Pattern Analysis." Cornell, Mar. 2024, www.css.cornell.edu/faculty/dgr2/_static/files/R_PDF/exPPA.pdf.

[17] "First- and Second-Order Processes." Spatial Analysis Online, spatialanalysisonline.com/HTML/first-_and_second-order_proces.htm.

[18] Eric. "The Fundamentals of Kernel Density Estimation." APTECH, 2023,

https://www.aptech.com/blog/the-fundamentals-of-kernel-density-estimation/.

[19] Rossiter, D G. “Point-Pattern Analysis.” Cornell, 2024, www.css.cornell.edu/faculty/dgr2/_static/files/ov/ov_PPA_Handout.pdf.

[20] Hartmann, K., Krois, J., Waske, B. (2018): E-Learning Project SOGA: Statistics and Geospatial Data Analysis. Department of Earth Sciences, Freie Universitaet Berlin.

[21] Van Lieshout, M.N.M. Theory of Spatial Statistics: A Concise Introduction. CRC Press, 2019.

[22] Matern, B. (1971) Doubly stochastic Poisson processes in the plane. In Statistical Ecology, ed. G. P. Patil, E. C. Pielou and W. E. Waters, Pennsylvania State University Press, University Park, PA

[23] Moraga, Paula. “Spatial Statistics for Data Science: Theory and Practice with R.” Chapter 22 The K-Function, www.paulamoraga.com/book-spatial/the-k-function.html. Accessed 8 Oct. 2024.

[24] “Interactions in Point Pattern Analysis.” SOGA-R, 18 Oct. 2017, www.geo.fu-berlin.de/en/v/soga-r/Advances-statistics/Spatial-Point-Patterns/Analysis-of-Spatial-Point-Patterns/Interactions-in-Point-Pattern-Analysis/index.html.

[25] Srivastava K. Urbanization and mental health. Ind Psychiatry J. 2009 Jul;18(2):75-6. doi: 10.4103/0972-6748.64028. PMID: 21180479; PMCID: PMC2996208.

[26] "County and State Data Map: Defining Mental Health Across Communities." Mental Health America, mhanational.org/mhamapping/mha-state-county-data.

[27] "Guns per Capita 2024." Guns per Capita 2024, worldpopulationreview.com/state-rankings/guns-per-capita.

[28] Walker, Kyle, and Matt Hernan. "Basic Usage of Tidycensus." Tidycensus, walker-data.com/tidycensus/articles/basic-usage.html.

[29] "Trends in Homicide: What You Need to Know." *Council on Criminal Justice*, 21 Nov. 2024, counciloncj.org/homicide-trends-report/.

[30] "Homicide Mortality by State." National Center for Health Statistics, U.S. Department of Health & Human Services, 2 Mar. 2022, www.cdc.gov/nchs/pressroom/sosmap/homicide_mortality/homicide.htm.