

---

# Predicting Firearm Mortality

*A Data-Driven Analysis of State-Level Gun Policies and Socioeconomic Factors*

---

Kristin Lloyd  
Joshua Lin  
Allen Wu  
Kelsey Szafranski

May 1, 2025

## Abstract

Firearm mortality is a persistent and complex public health crisis in the United States, with significant variation across states. This project investigates the social, economic, and policy-related predictors of firearm mortality using data from all 50 U.S. states and Washington, D.C. We applied Principal Component Regression (PCR) to uncover latent structural drivers, Hierarchical and Stepwise Linear Regression to model the cumulative influence of different variable groups, and Lasso Logistic Regression to identify which state-level gun policies are most associated with high or low firearm death rates.

Our results show that gun ownership, incarceration rates, educational attainment, and poverty are among the strongest continuous predictors of firearm mortality. In terms of policy, restrictions on weapon purchases, safety regulations like secure storage laws, and prohibitions for high-risk individuals emerged as the most protective factors. Notably, Lasso Logistic Regression revealed that some commonly referenced policies, such as red flag laws and domestic violence prohibitions, may not add unique predictive value when accounting for broader policy domains.

These findings suggest that the most effective strategies to reduce firearm mortality involve comprehensive policy packages that limit access to firearms while also addressing the broader social conditions that foster gun violence. A multi-pronged approach—combining targeted legislation with investments in education, health, and economic opportunity—may offer the greatest potential to save lives.

## 1 Introduction & Literature Review

Gun violence remains one of the most urgent public health issues in the United States. In 2022 alone, over 48,000 Americans died from firearm-related causes, averaging one death every 11 minutes. These fatalities include suicides (55%), homicides (41%), accidental deaths, and law enforcement shootings, reflecting the broad and devastating impact of gun violence on communities nationwide. Beyond deaths, over 200 people are treated each day for nonfatal firearm injuries<sup>1</sup>.

What makes this crisis particularly complex is the wide variation in firearm mortality rates across U.S. states. These differences are shaped by a mix of policy, social, and economic conditions. States with stricter gun safety laws—such as background checks, assault weapon bans, and safe storage requirements—tend to experience lower firearm mortality rates<sup>2</sup>. Meanwhile, states with weaker or fewer laws report much higher rates. Despite this evidence, many high-risk states continue to score

---

<sup>1</sup>Johns Hopkins Center for Gun Violence Solutions, 2023

<sup>2</sup>Everytown for Gun Safety Support Fund, 2025

poorly on gun law strength, suggesting a disconnect between the severity of the crisis and the strength of legislative protections.<sup>3</sup>

Research has also shown that broader social indicators, such as gun ownership rates, educational attainment, incarceration rates, and access to mental health care, influence firearm mortality across states. These factors are often deeply interconnected, making it difficult for traditional statistical approaches that rely on isolated predictors to capture the full picture.<sup>4</sup>

To address this complexity, our project applies a statistical learning framework that uses several regression techniques to identify the most predictive factors of firearm mortality. We incorporate Principal Component Regression (PCR) to reduce dimensionality and uncover underlying societal patterns, Hierarchical and Stepwise Regression to assess the incremental contribution of different factor groups, and Lasso Logistic Regression to isolate the most influential binary gun policies.

Our study focuses on three main research questions:

1. Which continuous state-level factors, such as incarceration rates, education levels, and health metrics, are most predictive of firearm mortality?
2. Which specific gun policy domains, including assault weapon bans, background checks, and red flag laws, are associated with lower odds of high firearm mortality?
3. How do different statistical modeling techniques compare in identifying key predictors of firearm deaths?<sup>5</sup>

By combining these approaches with state-level data, our goal is to produce actionable insights that can inform evidence-based public policy. Ultimately, we hope this work contributes to broader discussions on how to reduce firearm-related harm and improve public safety outcomes.

## 2 Data Collection

### 2.1 Datasets and Variables

Our analysis integrates state-level data from multiple publicly available sources to examine predictors of firearm mortality across the 50 U.S. states and Washington, D.C. The dependent variable, firearm mortality rate per 100,000 residents, was sourced from CDC's 2022 firearm death statistics.<sup>6</sup> Independent variables span four major categories:

- **Gun Policies and Laws:** We included 25 binary indicators (0/1) for specific firearm regulations.<sup>7</sup> These cover background check requirements, red flag laws, assault weapon bans, safe storage laws, minimum age laws, high-capacity magazine bans, prohibitions for high-risk individuals, and restrictions on open or concealed carry. A composite gun law strength score<sup>8</sup> scaled from 0 (weakest) to 100 (strongest), was also included.
- **Socioeconomic and Demographic Factors:** Variables include poverty rates, unemployment rates, income inequality (Gini index), labor force participation rates for women, teen birth rates, and educational attainment levels.<sup>9</sup> Population size and availability of domestic violence shelters were also used.
- **Violence and Criminal Justice Metrics:** We incorporated violent and non-violent crime rates, adult and youth incarceration rates, incarceration racial disparities, domestic violence prevalence, domestic violence spending, school shooting incidents, and child firearm mortality rates.
- **Health and Behavioral Indicators:** Health-related factors include mental health access scores, suicide rates, funding for suicide prevention, alcohol-related and drug overdose death rates, and gun ownership rates.

---

<sup>3</sup>Lloyd, Lin, Wu, & Szafranski, 2025

<sup>4</sup>"Gun Violence in the U.S. – Linear Regression," 2025

<sup>5</sup>Lloyd, Lin, Wu, & Szafranski, 2025

<sup>6</sup>CDC, 2022

<sup>7</sup>Everytown Research & Policy, 2023

<sup>8</sup>Korhonen, 2024

<sup>9</sup>U.S. Census Bureau, BLS, NCES

## 2.2 Data Preprocessing

- **Merging:** Datasets were merged by standardized state names and abbreviations.
- **Missing Data:** Variables with more than 15% missingness were excluded; remaining gaps were imputed with median values.
- **Standardization:** Continuous variables were z-score normalized; policy variables remained binary (0/1).
- **Composite Variables:** Gun law strength and domestic violence shelter accessibility metrics were constructed from multiple related inputs.

This meticulous, multi-sourced integration and preprocessing approach ensures a comprehensive, standardized dataset that captures the wide range of factors influencing firearm mortality. It provides a robust foundation for our visualizations and multivariate modeling techniques applied in subsequent sections.

## 3 Methods

Our study conducted a state-level, cross-sectional analysis of firearm mortality in the United States, combining traditional regression techniques with regularized models to assess both continuous and binary predictors.

Predictor variables were categorized into two groups:

- **Continuous Variables:** Socioeconomic indicators (e.g., poverty rate, Gini index), crime and incarceration rates (e.g., violent crime rate, incarceration disparity), health measures (e.g., suicide rates, drug overdose rates), educational attainment, and gun ownership rates.
- **Binary Policy Indicators:** Presence or absence of specific firearm laws, grouped into six domains:
  - Weapon purchase restrictions
  - High-risk individual prohibitions
  - Domestic violence-related restrictions
  - Location-based bans
  - Child access and safety regulations
  - Risk-based interventions (e.g., red flag laws)

All continuous variables were standardized prior to analysis. For classification modeling, states were grouped based on whether their firearm mortality rate was above or below the national median.

We employed the following analytical strategy:

- **Principal Component Regression (PCR):** Used to address multicollinearity by transforming correlated predictors into orthogonal principal components.
- **Stepwise Regression:** Applied backward elimination to select the most predictive subset of continuous variables.
- **Hierarchical Regression:** Assessed the incremental contribution of conceptual domains by entering variables in theoretically informed blocks (e.g., socioeconomic → crime → health).
- **Logistic Regression (with Lasso and Ridge Regularization):** Evaluated binary firearm policy indicators, with Lasso (L1) regularization improving model interpretability through feature selection.

Model performance was assessed using cross-validation:

- Continuous outcome models were evaluated using cross-validated  $R^2$  scores.
- Classification models were evaluated using cross-validated accuracy rates.

Model parsimony and interpretability were considered in selecting final models.

## 4 Results

### 4.1 Continuous Predictors of Firearm Mortality

#### 4.1.1 Model Performance

Stepwise Regression achieved the highest cross-validated performance ( $CV R^2 = 0.8041$ ), outperforming Principal Component Regression ( $CV R^2 = 0.6773$ ) and Hierarchical Regression ( $CV R^2 = 0.5406$ ) (Figure 1).

| Model Selection |               |             |                      |
|-----------------|---------------|-------------|----------------------|
| Model           | Mean CV $R^2$ | Train $R^2$ | Number of Predictors |
| PCR             | 0.6773        | 0.8179      | 5                    |
| Stepwise        | 0.8041        | 0.8615      | 4                    |
| Hierarchical    | 0.5406        | 0.8893      | 18                   |

Figure 1: Model performance comparison across PCR, Stepwise, and Hierarchical regression models.

#### 4.1.2 Key Predictors from Stepwise Regression

The Stepwise model retained four predictors: Gun ownership rate, Poverty rate, Violent crime rate, Incarceration rate.

- Among these, **gun ownership rate** emerged as the strongest positive predictor of firearm mortality (Figure 2).

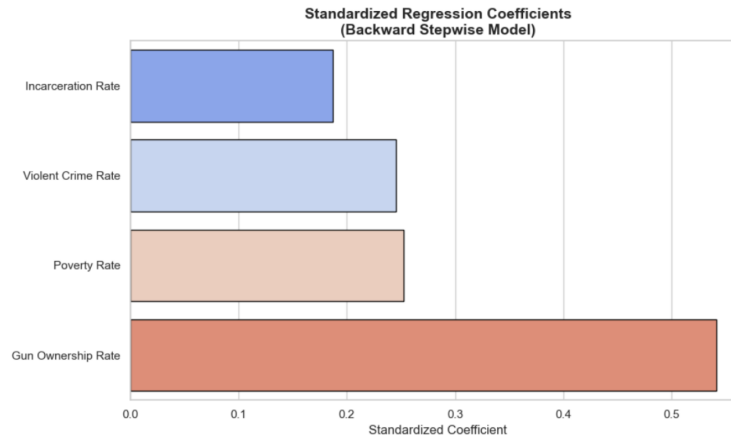


Figure 2: Standardized regression coefficients from the Stepwise model.

#### 4.1.3 Principal Component Regression Insights

Principal Component Regression condensed 27 predictors into five components reflecting latent domains such as gun culture, economic inequality, and health burden.

- Components heavily loaded on **poverty, gun ownership, incarceration, and suicide rates** (Figure 3).

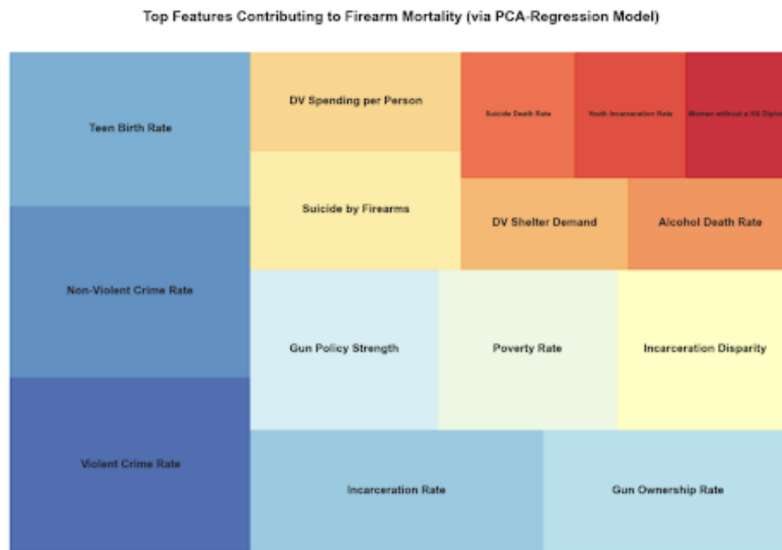


Figure 3: Top features contributing to firearm mortality via PCA-regression.

#### 4.1.4 Hierarchical Regression Results

In the Hierarchical model, variables were entered sequentially by theoretical blocks: Gun Policy and Ownership, Socioeconomic Factors, Crime and Justice System, Health and Substance Abuse, Social Factors and Education.

- The largest increase in explained variance occurred after adding **socioeconomic factors** (Figure 4).

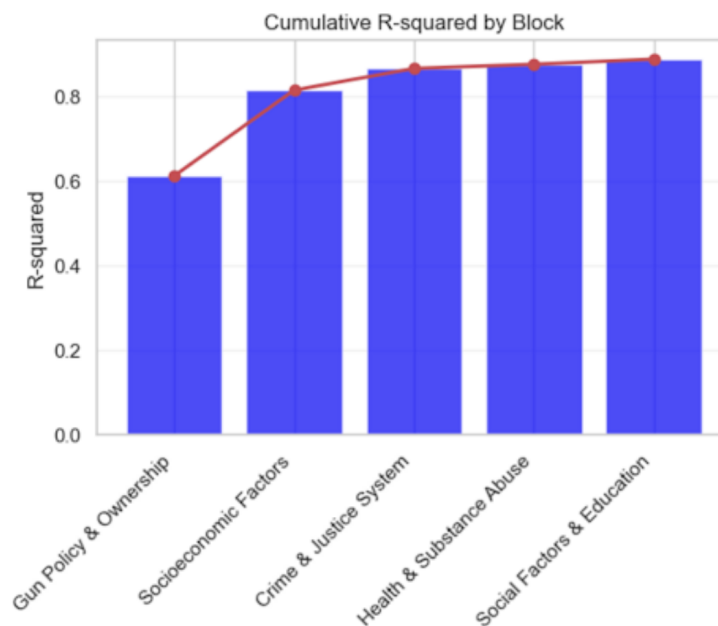


Figure 4: Cumulative  $R^2$  contribution by conceptual block in Hierarchical Regression.

## 4.2 Policy Predictors: Binary Logistic Models

### 4.2.1 Model Performance

Among the classification models, **Lasso Logistic Regression** achieved the highest cross-validation accuracy (**78.18%**), outperforming both Ridge and Standard Logistic Regression (Figure 5).

| Model Selection                |          |
|--------------------------------|----------|
| Cross Validation               | Accuracy |
| Standard Logistic Regression   | 0.7618   |
| Lasso (L1) Logistic Regression | 0.7818   |
| Ridge (L2) Logistic Regression | 0.7618   |

Figure 5: Classification accuracy comparison across logistic regression models.

### 4.2.2 Key Protective Policy Domains

The Lasso model identified five major protective policy domains associated with reduced odds of high firearm mortality:

- Weapon purchase restrictions
- Safety regulations (e.g., secure storage laws)
- High-risk individual prohibitions
- Risk-based interventions (e.g., extreme risk protection orders)
- Domestic violence prevention policies

Notably, **location-based restrictions** showed a slight positive association with firearm mortality risk, potentially reflecting policy adoption patterns in high-risk areas (Figure 6).

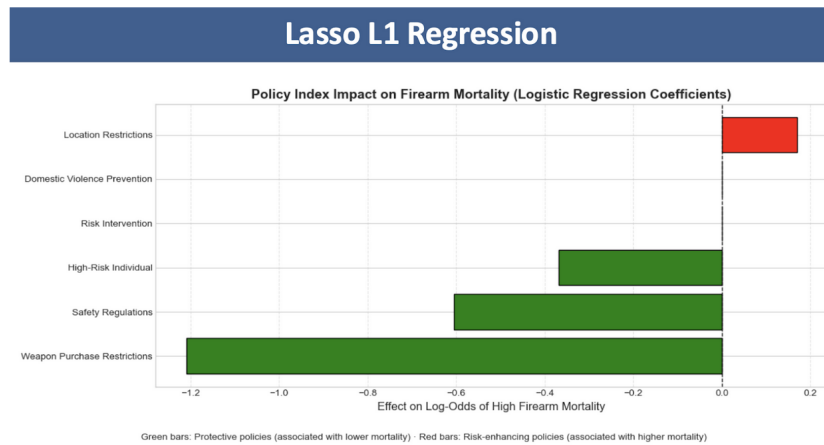


Figure 6: Lasso logistic regression coefficients (green = protective, red = risk-enhancing).

### 4.2.3 Real-World Policy Example

States with an **assault weapons prohibition** had a substantially lower proportion of high firearm mortality outcomes compared to states without such bans (Figure 7).

## Example: Assault Weapons Prohibited:

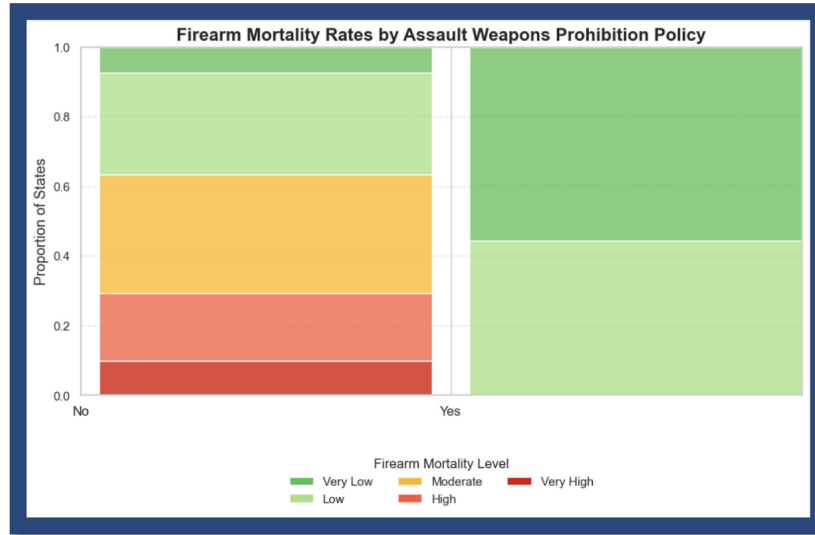


Figure 7: Firearm mortality distribution by assault weapon prohibition status.

### 4.2.4 Summary of Key Trends

Across all models:

- **Gun ownership, poverty, and violent crime** consistently predicted higher firearm mortality.
- **Secure storage laws, waiting periods, and assault weapon bans** emerged as significant protective policies.
- Broader **social disadvantage** and **health burdens** remain critical factors shaping firearm death rates across states.

## 5 Discussion & Unique Findings

Our analysis shows that firearm mortality is strongly linked to gun ownership rates, socioeconomic disadvantage (poverty, incarceration, education), and the presence or absence of specific gun safety laws. Policies like secure storage laws, waiting periods, and assault weapon bans were consistently associated with lower firearm death rates.

- **Unexpected finding:** Location-based firearm restrictions were linked to slightly higher mortality risk, likely reflecting reactive policy adoption in already high-violence states rather than a causal relationship.
- **Limitations:** The analysis is cross-sectional in nature, limited by a small sample size ( $n = 51$ ), and lacks insight into local cultural or enforcement differences.
- **Regional variation:** Southeastern and Mountain West states exhibited higher firearm mortality rates, while Northeastern and West Coast states showed lower rates—mirroring underlying differences in social conditions and policy environments.
- **Policy implication:** Addressing firearm mortality will likely require a dual approach: targeted firearm policy interventions and broader investments in education, health, and economic opportunity.

## 6 Conclusion

This study highlights that firearm mortality rates across U.S. states are shaped by a combination of social conditions and firearm policy environments. Higher gun ownership, poverty, violent crime, and incarceration rates were strong predictors of elevated mortality, while certain policies such as secure storage laws, waiting periods, and assault weapon bans were consistently associated with lower firearm death rates.

Our findings suggest that no single policy is likely to eliminate firearm violence. Instead, the most effective strategies combine targeted firearm regulations with broader efforts to address economic, educational, and health disparities that create environments of risk.

Future research could extend this work by examining longitudinal data to assess causality over time, investigating city-level variations within states, and exploring the real-world enforcement of firearm policies. A deeper understanding of these dynamics is essential for designing interventions that are both effective and equitable.

## 7 References

- **All Project Variables Gathered:** [https://docs.google.com/document/d/1kxcreCVOP9vRMUYf0UWS0ry4L2pTq9NuusrQo3-j5Ic/edit?usp=drive\\_link](https://docs.google.com/document/d/1kxcreCVOP9vRMUYf0UWS0ry4L2pTq9NuusrQo3-j5Ic/edit?usp=drive_link)
- Johns Hopkins Center for Gun Violence Solutions. (2023). *Gun violence in the United States*. Retrieved from: <https://publichealth.jhu.edu/center-for-gun-violence-solutions/research-reports/gun-violence-in-the-united-states>
- Everytown for Gun Safety Support Fund. (2025). *Gun law rankings: Gun safety laws save lives*. Retrieved from: <https://everytownresearch.org/rankings/>
- Lloyd, K., Lin, J., Wu, A., & Szafranski, K. (2025). *Gun Violence in the U.S*. Retrieved from: [https://kristinmlloyd.github.io/Website-Portfolio/projects/gun-violence/website/\\_site/](https://kristinmlloyd.github.io/Website-Portfolio/projects/gun-violence/website/_site/)
- *Gun Violence in the U.S. – Linear Regression*. (2025). Retrieved from: [https://kristinmlloyd.github.io/Website-Portfolio/projects/gun-violence/website/\\_site/linear\\_regression.html](https://kristinmlloyd.github.io/Website-Portfolio/projects/gun-violence/website/_site/linear_regression.html)

## 8 Project Details

All project code, data processing scripts, and modeling notebooks are available at the following repository:

- **GitHub Repository:** <https://github.com/dsan-5300/project-3.git>
- **Website:** [https://kristinmlloyd.github.io/Website-Portfolio/projects/gun-violence/website/\\_site/](https://kristinmlloyd.github.io/Website-Portfolio/projects/gun-violence/website/_site/)

The repository includes:

- Data cleaning and preprocessing scripts
- Feature engineering workflows
- Principal Component Regression, Stepwise Regression, Hierarchical Regression, and Logistic Regression modeling pipelines
- Instructions for environment setup and reproducibility

For any questions or technical details, please refer to the repository's README file.