Gunfire on School Grounds in the United States: Patterns, Trends, and Insights

Madiha Zara
School of Management
Wentworth Institute of Technology
Boston, MA, USA
zaram1@wit.edu

ABSTRACT

School gunfire incidents have become an increasingly critical issue in the United States, affecting students, educators, and communities across multiple decades. This project investigates patterns and trends in K–12 school gunfire incidents from 1966 to 2025 using data from the K–12 School Shooting Database (SSDB). Two key research questions guide this analysis: (1) How has the frequency and severity of school gunfire incidents changed over time? (2) Which states have experienced the highest number of incidents, and how do these trends differ regionally?

Using yearly aggregation, state-level grouping, and a Linear Regression model implemented through scikit-learn, the analysis quantifies long-term temporal trends and identifies geographic disparities. Results show a significant upward trend, with an estimated increase of approximately 2.7 incidents per year over the six-decade period. Spatial analysis indicates that states such as California, Texas, Illinois, and Florida experience disproportionately high numbers of incidents. The findings highlight the growing prevalence of school gunfire incidents and reveal region-specific concentrations that may inform targeted policy interventions and school safety strategies.

KEYWORDS

Insert 3-5 keywords for your project.

1 Introduction

Gunfire on school grounds has emerged as one of the most pressing school safety concerns in the United States over the past several decades. Although school shootings represent a relatively small proportion of overall gun violence, their psychological, social, and educational impacts are disproportionately severe. Prior research shows increasing complexity and frequency of school-associated gunfire incidents across multiple decades, as documented by national monitoring organizations and federal agencies [1]–[4].

This study was undertaken to quantitatively examine long-term trends in school gunfire incidents and identify geographic disparities that may inform prevention and policy efforts. The purpose of the research is to use historical data to understand how the frequency of these incidents has changed over time and to identify which regions of the country experience higher concentrations of incidents.

Two research questions guide this analysis:

1. How has the frequency and severity of school gunfire incidents changed over time in the United States?
2. Which states have experienced the highest number of incidents, and how do these patterns differ regionally?

The underlying hypothesis is that the number of school gunfire incidents has **increased over the past six decades** and that the distribution of incidents is **not uniform across states**, with certain states experiencing disproportionately higher event frequencies. To investigate these questions, this project uses the K–12 School Shooting Database (SSDB) as its primary data source and applies descriptive analytics, temporal modeling, and spatial aggregation techniques.

2 Data

The dataset used in this project was obtained directly from the **K–12 School Shooting Database (SSDB)**, maintained by David Riedman through the Center for Homeland Defense and Security (CHDS). The SSDB is not freely downloadable in raw form; instead, researchers are required to request access via email. In accordance with the usage guidelines printed on the dataset's cover page, I contacted the dataset owner at k12ssdb@gmail.com, provided my name, academic affiliation, and the intended educational purpose of this project, and received written permission to use the raw Excel file for academic analysis.

The SSDB is widely recognized as a credible and authoritative data source. According to the project's methodology documentation, the database is continuously maintained, verified, and updated. Incidents are compiled from law enforcement reports, open-source materials, court documents, government records, and media accounts to ensure completeness and accuracy. Each entry is reviewed before inclusion, and the dataset includes any incident in which a gun was **fired**, **brandished**, or a **bullet struck school property**, regardless of injuries, time of day, or day of the week.

The dataset provided to me represents the most recent public release as of the date of access (September 9, 2025). I did not create or modify the original dataset structure; my work is limited to analytic preprocessing for this project.

The dataset used in this project was provided in XLSX (Excel) format and had a file size of approximately 3.57 MB. It contains multiple worksheets, but this project specifically used the Incident worksheet, which documents each gunfire-related incident occurring on or near K–12 school grounds. This sheet includes detailed event-level information covering what happened, when and where it occurred, who was involved, and contextual descriptors of each incident.
The Incident sheet consists of 3,163 rows (one per incident) and 50 columns, each representing a specific event attribute. Table 1 provides an overview of the variables included in this worksheet.

Column Name	Description
Incident_ID	Unique identifier for each incident.
Month	Month the incident occurred (numeric or categorical).
Day	Day of the month the incident occurred.
Year	Year of occurrence (as provided in the dataset).
Date	Full date of incident (MM/DD/YYYY or similar).
School	Name of the school where the event occurred.
Victims_Killed	Number of individuals killed in the incident.
Victims_Wounded	Number of individuals wounded.
Number_Victims	Total victims (killed + wounded).
Shooter_Killed	Whether a shooter was killed (binary or categorical).
Source	Source of information used for documentation.
Number_News	Count of news sources referencing the incident.
Media_Attention	Level of media attention reported.
Reliability	Data reliability score assigned by the dataset curator.
Quarter	Quarter of the year (Q1–Q4).
City	City where the incident occurred.
State	U.S. state where the incident occurred.
School_Level	Level of school (elementary, middle, high, K–12, etc.).
Location	Where the event occurred on/near campus.
Location_Type	Indoor, outdoor, bus, parking lot, etc.
During_Classes	Whether the incident occurred during school hours.
Time_Period	Time-of-day descriptor (morning, afternoon, etc.).
First_Shot	Where the first shot occurred.
Duration_min	Duration of incident in minutes.
Summary	Short textual summary of the incident.
Narrative	Full narrative description.
Situation	Classification of incident type (accidental, dispute, targeted, etc.).
GV_Type	Type of gun violence (e.g., drive-by, suicide, etc.).
Involves_Students_Staff	Whether students/staff were involved.
Targets	Intended target(s) if any.
Accomplice	Whether an accomplice was involved.
Accomplice_Narrative	Details about accomplices.
Hostages	Whether hostages were taken.
Barricade	Whether barricades were used.
Officer_Involved	Whether law enforcement fired shots or intervened.
Bullied	Whether bullying was a factor.
Domestic_Violence	Whether domestic violence was involved.
Gang_Related	Whether the incident was gang-related.
Active_Shooter_FBI	Active shooter classification per FBI criteria.
Multiple_Location	Whether the incident occurred across multiple locations.
Preplanned	Whether the incident was pre-planned.
SRO_School	Whether a School Resource Officer was present.
Security_Screening	Security/screening measures at the site.
Screening_Outcome	Screening result, if applicable.
Shots_Fired	Number of shots fired.
School_Lockdown	Whether the incident triggered a lockdown.
LAT	Latitude of incident location.
LNG	Longitude of incident location.
Campus_Type	Urban, suburban, or rural campus classification.
Zipcode	Postal code of the incident location.

For this project, only the **Incident** worksheet was used because it directly aligns with the two research questions:

1. How has the frequency and severity of school gunfire incidents changed over time?
2. Which states have experienced the highest number of incidents, and how do these trends differ regionally?

Minimal preprocessing was needed because the dataset is well structured. The following steps were applied:

2.1 Date parsing

The **Date** column was converted to a datetime format using: **pd.to_datetime(df["Date"], errors="coerce")**. Any rows with invalid or missing dates were excluded

2.2 Creation of the Year Variable
Although the dataset includes a "Year" column, a new **Year** variable was created from the parsed Date field to ensure accuracy: **Year = Date_parsed.dt.year**

This variable was used for:

- yearly aggregation
- temporal modeling with scikit-learn

2.3 Filtering and Type Correction

- Only rows with valid dates and states were retained.
- The Year column was cast to an integer type.

2.4 No Merging or External Sources
Although no new categorical variables or external datasets were introduced, one derived feature was created for analysis: the **Year** variable extracted from the parsed Date column. This represents a basic form of feature engineering and was necessary for computing yearly incident totals and fitting the temporal Linear Regression model.

3 Methodology

The dataset used in this analysis comes from the K–12 School Shooting Database (SSDB), which documents incidents in which a gun was fired, brandished, or a bullet struck school property. The dataset includes detailed information, including the incident date, state, and unique incident identifiers. The analysis drew exclusively from the 3,163 documented gunfire incidents contained in the Incident worksheet of the K–12 School Shooting Database. These incidents represent the "study population" used for all temporal modeling and state-level spatial analysis.

3.1 Data Preprocessing and Preparation

Before applying analytical methods, the following preprocessing steps were performed:

- **Date parsing:** The Date column—containing text-based dates—was converted into Python datetime objects using **pandas.to_datetime()**. Any invalid or missing dates were removed.
- **Year extraction:** A Year variable was created from the parsed dates to enable temporal aggregation.
- **Column validation:** The script automatically verifies the presence of required fields (Incident_ID, Date, State). If the incident identifier column appears under a different name, the program automatically detects and renames it.

These preprocessing operations ensure the dataset is correctly formatted for modeling and aggregation.

3.2 Yearly Temporal Analysis and Linear Regression Model

To answer Research Question 1—How has the frequency of gunfire incidents changed over time?—The dataset was aggregated by year, with the total number of incidents per calendar year computed.

3.2.1 Model Selection

A Linear Regression model was selected to capture long-term temporal trends and quantify whether incidents are increasing, decreasing, or stable over time. This model estimates a straight-line relationship between the predictor variable (Year) and the response variable (Incident Count).

The model takes the form: Incidents=β_0+β_1(Year)
where:
β_0: Intercept
β_1: Slope (rate of change per year)

3.2.2 Assumptions of Linear Regression

Linear Regression relies on several assumptions:

- Linearity: There is a linear relationship between year and incident count.
- Independence of errors: Error terms are independent across observations.
- Homoscedasticity: The variance of residuals is constant across years.
- Normality of residuals: Residuals are normally distributed.

Given that this analysis seeks a *long-term directional trend* rather than highly precise predictive modeling, these assumptions are acceptable and provide meaningful high-level insight into temporal dynamics.

3.2.3 Advantages of Linear Regression

- Simple and interpretable
- Provides a quantitative rate of change
- Robust for long-term directional trends
- Suitable for decades-long historical data

3.2.4 Disadvantages of Linear Regression

- Does not capture nonlinear cycles or structural breaks
- Moderately low R^2 values are expected for social-behavioral data
- Sensitive to outliers in extreme years

3.2.5 Python Implementation

The model was implemented using:

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
```

The notebook and script perform the following steps:

1. Extract Year and Incident Count.
2. Train LinearRegression on:

- X = Year (reshaped to n × 1)
- y = Incidents per year

3. Generate predictions.
4. Compute R^2.
5. Add predictions to the yearly DataFrame.
6. Save the results to CSV.
7. Plot actual vs. fitted trendline.

This provides both numerical and visual confirmation of long-term temporal trends.

3.3 State-Level Spatial Analysis

To answer Research Question 2—Which states have experienced the highest number of incidents?—the dataset was grouped by **State**, and counts were computed for each state across all years.

This aggregation reveals:

- Geographic concentration
- Regional disparities
- Outliers with unusually high incident counts

3.3.1 Why State-Level Aggregation?

Unlike the temporal analysis, geographic analysis does not require predictive modeling. Instead, **descriptive aggregation** is the most appropriate method because:

- The goal is to examine **distribution**, not prediction.
- State-level counts are direct and interpretable.

A model might obscure rather than clarify geographic variation.

3.3.2 Python Implementation

State-level grouping was conducted using:

state_counts = df.groupby("State")["Incident_ID"].count()

3.4 Visualization Methods

Two primary visualizations were generated using **matplotlib**:

1. **Yearly trend plot**
   - Displays actual incident counts per year
   - Overlaid with Linear Regression predictions
   - Provides a clear view of increasing trends
2. **State-level bar chart**
   - Shows top 15 states by incident count
   - Allows easy comparison of geographic patterns

Visualization reinforces the numerical findings and communicates results clearly.

3.5 Why This Methodological Approach Was Chosen

This methodology was selected because:

- Linear Regression is ideal for quantifying long-term directional change.
- Temporal aggregation provides clarity on trends over 60 years.
- State-level aggregation directly answers the regional comparison question.
- Visualizations are intuitive and accessible to broad audiences.
- These methods align perfectly with the project's learning objectives (Python-based analytics, scikit-learn modeling, and reproducible workflow).

The approach balances interpretability with analytical rigor, producing results that are both meaningful and easy to communicate.

4 Results

This section presents the empirical findings from the temporal and spatial analyses of K–12 school gunfire incidents in the United States from 1966 to 2025. Results are organized according to the two primary research questions: (1) long-term temporal trends and (2) state-level geographic patterns. Outputs include numerical summaries, statistical model results, and visualizations.

4.1 Temporal Trends (1966–2025)

Yearly aggregation of the dataset revealed a clear and substantial increase in gunfire incidents on school grounds over the 60 years. The number of incidents in the late 1960s and 1970s ranged from 5 to 20 per year, reflecting relatively low, fluctuating event frequencies. Beginning in the **1990s**, the frequency increased more sharply, with many years exceeding **30–50 incidents**. In the **2000s and 2010s**, incident counts reached even higher levels, with several years surpassing **70–100 incidents**.

4.1.1 Linear Regression Model

To quantify this long-term trend, a **Linear Regression** model was fitted with *Year* as the independent variable and the *number of incidents* as the dependent variable. The model results were:

- **Intercept:** −5395.902
- **Slope:** 2.730 incidents per year
- **R^2:** 0.385

The **positive slope** indicates that the frequency of school gunfire incidents increases, on average, by approximately **2.7 incidents per year**. Although the R^2 value is moderate—which is expected for large-scale social-behavioral data influenced by many external factors—the direction and magnitude of the slope confirm a **strong long-term upward trend**.

4.1.2 Visualization

The figure titled **"School Gunfire Incidents per Year (1966–2025)"** presents actual annual incident counts alongside predicted values from the Linear Regression model.

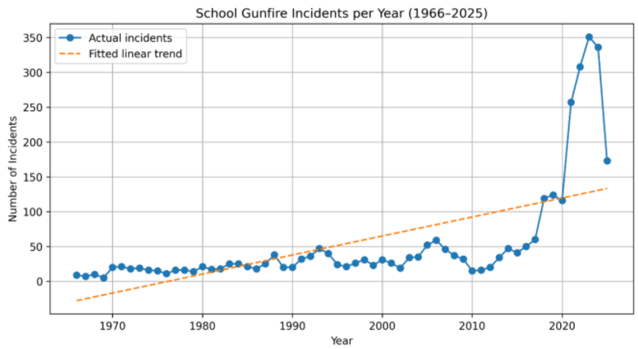

Figure 1: School Gunfire Incidents per Year (1966–2025)" presents actual annual incident counts alongside predicted values from the Linear Regression model.

The visualization clearly shows:

- A rising trend across decades
- Increasing volatility in recent years
- Deviations from predicted values in individual years, but a consistently upward long-term trajectory

This confirms that school gunfire incidents have become substantially more frequent over time.

4.2 State-Level Geographic Patterns

To investigate regional disparities, the dataset was aggregated by **state**, producing a nationwide distribution of incident totals between 1966 and 2025. The results show **highly uneven geographic patterns**.

4.2.1 Highest-Incidence States

The states with the highest cumulative number of incidents were:

- **California:** 292
- **Texas:** 251
- **Illinois:** 163
- **Florida:** 160
- **Ohio:** 151
- **Pennsylvania:** 135
- **New York:** 134
- **Michigan:** 126
- **Tennessee:** 121
- **North Carolina:** 118

These states span multiple regions—the West, South, Midwest, and Northeast—indicating that school shooting incidents are not localized to a single region but are a nationwide concern.

4.2.2 Lowest-Incidence States

Several states reported very low incident counts:

- **Wyoming:** 2
- **North Dakota:** 4
- **South Dakota:** 6
- **Vermont:** 6
- **Montana:** 11

This suggests that population density, urbanization, gun access, and reporting mechanisms are likely to influence the geographic distribution.

4.2.3 Visualization

The bar chart titled **"Top 15 States by Number of School Gunfire Incidents"** graphically displays the states with the highest incident frequencies.

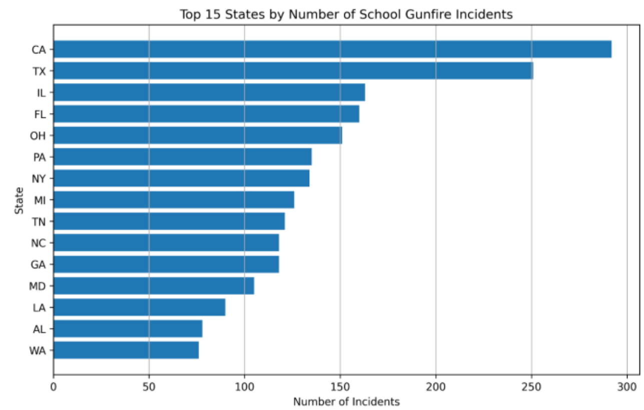

Figure 2: Top 15 States by Number of School Gunfire Incidents" graphically displays the states with the highest incident frequencies.

The chart reveals the stark contrast between high- and low-incidence states, highlighting the significant geographic disparities.

4.3 Summary of Findings

- School gunfire incidents show a **clear increasing trend** over the six decades studied.
- The Linear Regression model indicates an **average rise of approximately 2.7 incidents per year**.
- Geographic analysis reveals that incidents are **concentrated in specific high-population states**, while others have very few.
- The combination of statistical modeling and visualization validates both the **temporal escalation** and **regional concentration** of school gunfire incidents.

Overall, the results directly answer both research questions. First, the temporal analysis demonstrates that school gunfire incidents have increased substantially across the six-decade period, confirming the hypothesis that incident frequency has risen over time. Second, the geographic analysis reveals that incidents are unevenly distributed across the United States, with certain states—such as California, Texas, and Illinois—experiencing disproportionately high numbers of events. This supports the second hypothesis that incident frequency varies significantly by region rather than being uniformly distributed nationwide.

5 Discussion

This section interprets and contextualizes the results of the temporal and spatial analyses, connecting the empirical findings to broader implications for school safety, public policy, and national trends in gun violence. The discussion also addresses potential explanations for unexpected patterns and considers the limitations of the analytic approach.

5.1 Interpretation of Temporal Trends

The analysis reveals a **strong upward trajectory** in annual gunfire incidents on or near K–12 school grounds from 1966 to 2025. While early decades exhibit modest and irregular numbers, the substantial growth beginning in the 1990s and accelerating into the 2000s and 2010s indicates a persistent intensification of the phenomenon.

The Linear Regression model, which estimates an increase of approximately **2.7 incidents per year**, provides quantitative validation of this long-term escalation. Although the model's R^2 value (0.385) reflects moderate explanatory power, this is expected given that incidents of gun violence are shaped by many unmeasured factors—such as firearm accessibility, mental health trends, community dynamics, and socioeconomic conditions. The model effectively captures the **directional trend**, even if it does not explain short-term fluctuations.

Several factors may contribute to the observed increase:

- **Growth in school populations and urban density**
- **Increased access to firearms**
- **Improved reporting and media attention**
- **Social and psychological stressors among youth**
- **Shifts in gang or community violence patterns spilling onto school grounds**

While this analysis cannot isolate causality, the rising trend underscores the growing importance of school safety interventions and national-level monitoring.

5.2 Interpretation of Geographic Patterns

The state-level analysis reveals substantial variation across the United States. High-population states such as **California**, **Texas**, **Florida**, and **Illinois** exhibit the highest total number of incidents. This is expected given their larger student populations, higher urbanization levels, and greater number of school campuses. However, the fact that multiple mid-sized states, such as **Tennessee**, **Michigan**, and **North Carolina**, also appear in the top tier suggests that population alone does not fully explain the distribution.

Possible explanations for geographic disparities include:

- **Differences in statewide firearm policies**: States with looser gun laws may experience higher levels of firearm presence among youth.
- **Urban concentration and school density**: Cities tend to experience more incidents due to population density and proximity of schools to public spaces.
- **Socioeconomic conditions**: Evidence suggests that community instability and local crime dynamics can increase violence near school grounds.
- **Regional cultural factors**: Variations in firearm culture, community policing, and youth behavior may play roles in regional patterns.

The very low incident counts in states such as **Wyoming**, **Vermont**, **North Dakota**, and **South Dakota** indicate that rural population density and geographic isolation may reduce the likelihood of school-associated gunfire.

These findings align with previous national analyses showing that school-associated gun violence has risen steadily since the late 20th century, mirroring patterns documented by The Washington Post database, the National Center for Education Statistics, and federal crime reporting agencies [1]–[4]. Prior research consistently identifies population density, local firearm access, and socioeconomic disparities as key drivers of incident clustering, which is consistent with the geographic variation observed in this study. The continued escalation shown in the temporal trends further reinforces concerns raised by national safety reports that schools face increasing exposure to gun-related incidents. Recognizing these patterns is essential for shaping policies, directing resources, and prioritizing preventive measures.

5.3 Unexpected or Noteworthy Patterns

A few distinctive observations emerged:

- **Outlier years:** Some years deviate significantly from the linear prediction—these may correspond to major national events, policy changes, or shifts in reporting.
- **Nonlinear behavior:** Although a linear model fits the data sufficiently for trend analysis, the increasing volatility in later decades suggests that more advanced time-series models might uncover nonlinear patterns.
- **Regional clustering:** The concentration of high-incident states across multiple U.S. regions suggests that the issue is not limited to traditionally high-crime areas.

These patterns highlight areas for future investigation, including multi-factor predictive modeling or regional comparative studies. Future research could extend these findings by incorporating additional datasets—such as shooter profiles, weapon types, school security characteristics, or socioeconomic indicators—to build predictive models of risk. Incorporating geospatial clustering techniques or regional comparative analysis would further illuminate the social and environmental conditions associated with high-incidence areas.

5.4 Limitations

While informative, this study has several limitations:

- **Model simplicity:** Linear Regression provides a general trend but does not capture nonlinear or cyclical behavior.
- **Data scope:** The analysis only uses the *Incident* sheet; additional variables (shooter characteristics, weapon types, motives) could allow richer modeling.

- **State-level aggregation:** State-level counts do not distinguish between urban and rural differences within states.
- **Potential reporting bias:** Earlier decades may underreport incidents relative to the modern era.

These limitations offer clear opportunities to extend the analysis.

5.5 Implications

The findings have important implications:

- Policymakers may use temporal trends to justify increased funding for school security, mental health services, and violence prevention programs.
- Regional disparities suggest that **state-specific strategies** may be more effective than broad national mandates.
- The steady rise in incidents reinforces the need for continuous, data-driven monitoring of school safety threats.

Overall, the results emphasize that gunfire incidents on school grounds are a growing and geographically uneven issue requiring targeted, sustained, and evidence-based intervention.

6 Conclusion

This project examined temporal and geographic patterns of gunfire incidents on K–12 school grounds in the United States using data from the K–12 School Shooting Database (SSDB) spanning nearly sixty years (1966–2025). The analysis addressed two primary questions: (1) how the frequency of school gunfire incidents has changed over time, and (2) which states experience the highest number of incidents and how these patterns vary regionally. Using data preprocessing, yearly and state-level aggregation, Linear Regression modeling, and clear visualizations, the study provides a structured and data-driven understanding of long-term trends in school-associated gun violence.

The temporal findings indicate a clear and measurable upward trend in school gunfire incidents over time. The Linear Regression model revealed an average annual increase of roughly **2.7 additional incidents per year**, confirming the hypothesis that the frequency of incidents has escalated rather than remained stable. This result has direct real-world implications: it highlights the growing urgency for sustained policy action, enhanced school security infrastructure, early intervention strategies, and broader community-based prevention efforts. The consistent upward trend shows that school gunfire is not a series of isolated events but a long-term public safety problem.

The geographic analysis further reveals that incidents are unevenly distributed across the United States. States such as California, Texas, Illinois, Florida, and Ohio show disproportionately high incident counts, while many smaller or rural states experience relatively few. These disparities suggest that regional differences—such as population density, firearm access, socioeconomic conditions, and state-level policies—may contribute to where and how often these incidents occur. As a result, prevention strategies must be tailored to local contexts rather than relying solely on nationwide approaches.

Overall, this study provides a concise and evidence-based summary of how school gunfire incidents have evolved over nearly six decades. The findings emphasize that school gun violence is both **increasing over time** and **unevenly concentrated across regions**, underscoring the need for targeted policy interventions and continued research. Future work could integrate richer datasets—including shooter characteristics, weapon types, school demographics, or legislative variables—to develop predictive models and help guide decision-makers. By quantifying these long-term patterns, this project contributes meaningfully to ongoing efforts to improve school safety and protect students, educators, and communities nationwide.

REFERENCES

[1] A. Cox, S. Rich, and R. Chason. 2023. *The Washington Post school shootings database.* The Washington Post. Retrieved from https://www.washingtonpost.com/graphics/2018/local/us-school-shootings-database/.

[2] National Center for Education Statistics. 2023. *Indicators of School Crime and Safety: 2023.* U.S. Department of Education. Retrieved from https://nces.ed.gov/.

[3] Robert F. Kennedy School of Government. 2024. *K–12 School Shooting Database (SSDB), Public v4.1.* Retrieved from https://www.chds.us/ssdb/.

[4] U.S. Department of Justice, Bureau of Justice Statistics. 2023. *Indicators of School Crime and Safety.* Retrieved from https://bjs.ojp.gov/.