

# Geographic Hotspots and Incident Repetition Analysis of Gun Violence in the United States

1<sup>st</sup> Mirza Shaheen Iqbal  
M.Sc. Data Science  
22998316  
mirza.shaheen.iqbal@fau.de

**Abstract**—This study investigates the geographic, temporal, and demographic factors influencing gun violence in the United States. Utilizing the *FiveThirtyEight Gun Deaths Dataset* and the *Jamesqo Gun Violence Incident Data*, we identify regions most affected by gun violence, periods of escalation, and demographic disparities. The results indicate that Illinois, California, and Texas are the highest-risk states, with Chicago recording the highest fatality rates. Temporal analysis reveals spikes in gun violence incidents during 2012-2013 and 2014-2015, suggesting that certain periods are prone to escalation. Demographic analysis shows that Black males are disproportionately affected by homicides, while individuals with lower educational attainment, particularly in the Black community, are more vulnerable to gun violence. The findings highlight the need for targeted interventions in high-risk areas, during high-risk periods, and focusing on socio-economic factors contributing to gun violence. Further research is needed to explore additional contributing factors, such as gun control laws and mental health, to develop more comprehensive strategies for reducing gun violence in the U.S.

**Index Terms**—Gun violence, geographic hotspots, temporal patterns, United States, FiveThirtyEight Gun Deaths Dataset, Jamesqo Gun Violence Incident Data, fatalities, injuries, demographic disparities, intervention strategies.

## I. INTRODUCTION

Gun violence remains a significant issue in the United States, affecting public safety and societal well-being. The severity of gun violence varies across regions, with urban areas generally experiencing higher rates compared to rural regions. Identifying **geographic hotspots** is crucial for targeting interventions effectively. Even within urban centers, there are variations in the frequency and severity of incidents, making it essential to focus efforts on the most impacted areas.

An equally important factor is whether gun violence incidents repeat over time. Some areas experience chronic violence, while others see sporadic outbursts. Analyzing patterns of **incident repetition** can identify regions where violence is not only frequent but also escalating. Escalation can manifest as an increase in fatalities, injuries, or the frequency of incidents. Identifying these trends can help predict future outbreaks and enable timely interventions.

This study focuses on addressing two main research questions:

- Which geographic locations in the U.S. experience the highest severity of gun violence?

- Are there repeated incidents over time in these high-risk areas, and do these incidents show patterns of escalation?

### A. Motivation

The motivation behind this analysis lies in the need for data-driven solutions to combat gun violence. By identifying **geographic hotspots** and understanding patterns of escalation, policymakers can prioritize interventions in high-risk areas. Chronic violence and escalating trends require long-term strategies, including law enforcement efforts, public health initiatives, and community programs aimed at prevention.

### B. Approach

To answer the research questions, this study uses two datasets: the **FiveThirtyEight Gun Deaths Dataset** [1] and the **Jamesqo Gun Violence Incident Data** [2]. These datasets provide critical information on gun violence incidents across the U.S., including fatalities, injuries, and geographic locations.

An automated **ETL pipeline** was employed to preprocess the data by cleaning, transforming, and aggregating it. This pipeline ensures that the data is structured for reliable analysis. The focus of this study is on identifying geographic hotspots, tracking repeated incidents, and examining the escalation of violence over time to provide actionable insights for policy decisions.

## II. USED DATA

This study utilizes two primary datasets to analyze gun violence incidents across the United States: the **FiveThirtyEight Gun Deaths Dataset** and the **Jamesqo Gun Violence Incident Data**. These datasets were processed through an automated ETL pipeline to ensure data consistency and reliability.

### A. FiveThirtyEight Gun Deaths Dataset

The **FiveThirtyEight Gun Deaths Dataset** offers an overview of gun-related deaths in the United States, focusing on key demographic information such as age, sex, race, and the intent behind each incident (e.g., homicide, suicide, or accidental). The dataset spans from 2011 to 2014 and provides valuable insights into trends in gun violence fatalities.

1) *Dataset Structure*: The FiveThirtyEight dataset includes the following key columns:

- year, month: Temporal data representing when the incident occurred.
- intent, sex, age, race, place, education, hispanic: categorical variables describing the victim's demographic details and incident type.
- age, hispanic: Continuous variables representing the age and ethnicity of the victims.

2) *Figure 1: Geographic Distribution of Gun Deaths*:

Figure 1 shows the geographic distribution of gun deaths from the FiveThirtyEight dataset, highlighting areas in the U.S. where gun-related deaths are most prevalent.

incident_id	date	state	city_or_county	n_killed	n_injured	gun_stolen	gun_type	incident_characteristics	latitude
461105	2013-01-01 00:00:00	Pennsylvania	Mckeesport	0	4	unknown	unknown	Wounded(Injured)Mass Shooting (4+ vict...	40.3467
460726	2013-01-01 00:00:00	California	Hawthorne	1	3	unknown	unknown	Wounded(Injured)Shot - Dead (murder, a...	33.9090
478855	2013-01-01 00:00:00	Ohio	Lorain	1	3	0:Unknown 1:Unknown	0:Unknown 1:Unknown	Wounded(Injured)Shot - Dead (murder, a...	41.4455

Fig. 1. Geographic Distribution of Gun Deaths (FiveThirtyEight Dataset)

### B. Jamesqo Gun Violence Incident Data

The **Jamesqo Gun Violence Incident Data** provides incident-level details about gun violence in the U.S. from January 2013 to March 2018. The dataset includes the number of fatalities, injuries, and the geographic location of each incident, making it ideal for identifying regional trends in gun violence.

1) *Dataset Structure*: The Jamesqo dataset includes the following key columns:

- incident\_id, date, state, city\_or\_county: Identifiers and geographic location for each incident.
- n\_killed, n\_injured: Continuous variables for the number of fatalities and injuries in each incident.
- gun\_stolen, gun\_type, incident\_characteristics: Categorical variables describing the characteristics of the weapon and the incident type.
- latitude, longitude: Geographic coordinates for incident locations.
- year, month: Temporal data for time-based analysis.

2) *Figure 2: Gun Violence Incident Locations*: Figure 2 shows the locations of gun violence incidents based on the Jamesqo dataset, visualized by latitude and longitude coordinates. This map provides insights into the geographical spread of gun violence incidents.

	year	month	intent	police	sex	age	race	hispanic	place	education
0	2012	1	suicide	0	M	34.0	asian/pacific islander	100	home	ba+
1	2012	1	suicide	0	F	21.0	white	100	street	some college
2	2012	1	suicide	0	M	60.0	white	100	other specified	ba+
3	2012	2	suicide	0	M	64.0	white	100	home	ba+
4	2012	2	suicide	0	M	31.0	white	100	other specified	hs/ged

Fig. 2. Map of Gun Violence Incidents (Jamesqo Dataset)

### C. Data Licensing

Both the **FiveThirtyEight Gun Deaths Dataset** and the **Jamesqo Gun Violence Incident Data** are made available under the **Creative Commons Attribution 4.0 International (CC BY 4.0)** license [3]. This license allows the datasets to be shared, modified, and used for any purpose, as long as proper credit is given to the original authors. To comply with this license, proper attribution will be provided in all uses, including reports, presentations, and publications resulting from this research. Additionally, any modified versions of the datasets will carry the same **CC BY 4.0** license to ensure continued transparency and promote further research use.

## III. ANALYSIS

This section outlines the analysis conducted to answer two key research questions regarding gun violence in the United States. The first question investigates the **geographic locations** that experience the highest severity of gun violence, while the second examines whether **gun violence incidents show patterns** of escalation over time in these high-risk areas. The analysis involves the steps of data collection, cleaning, geographic and temporal analysis, and demographic insights, which are presented in detail below.

### A. Method

To address the two primary research questions, we employed a systematic methodology encompassing data collection, cleaning, transformation, and analysis.

1) *Data Collection and Cleaning*: The data collection and cleaning process was automated through an ETL (Extract, Transform, Load) pipeline to ensure that the datasets were properly structured for analysis. The pipeline follows these key steps:

- **Data Collection**: The raw data was collected from two main sources: the **FiveThirtyEight Gun Deaths Dataset** and the **Jamesqo Gun Violence Incident Data**. Both datasets were sourced from publicly available repositories. The collection process included downloading, unzipping, and importing the files into a structured environment.
- **Data Cleaning**: The data was cleaned to handle missing values, remove irrelevant columns, and standardize formats. Categorical variables with missing values were imputed with "unknown", while numerical variables were imputed using the median value. Irrelevant columns were removed, and date columns were normalized to a standard format (YYYY-MM-DD).
- **Data Transformation**: After cleaning, the data was aggregated by state and city, focusing on fatalities and injuries, and normalized to ensure comparability across regions.
- **Data Saving**: The cleaned and transformed data was saved in both SQLite and CSV formats to facilitate efficient database operations and straightforward data access.

This structured approach ensured the datasets were accurate and ready for analysis.

2) **Geographic Analysis:** In the geographic analysis, we focused on identifying regions with the highest severity of gun violence based on fatalities and injuries. The data was aggregated by state and city. We used bar charts and heatmaps to visualize the distribution of gun violence incidents, which helped in identifying geographic hotspots.

3) **Temporal Analysis:** A 3-month moving average was applied to the gun violence data to smooth out fluctuations and identify long-term trends. This method allowed us to observe whether gun violence incidents show patterns of escalation over time, especially during high-risk years.

4) **Demographic Analysis:** For demographic insights, the analysis focused on race, education level, and gender of the victims and perpetrators. We used stacked bar charts to examine the relationship between these factors and the frequency of gun violence incidents.

## B. Result

Let us dive into the result and see what the outcome is of the analysis.

1) **Geographic Analysis:** The geographic analysis revealed that **Illinois**, **California**, and **Texas** were the states most severely impacted by gun violence:

- **Illinois:** 13,514 injuries and 3,409 fatalities.
- **California:** 7,644 injuries and 5,562 fatalities.
- **Texas:** 6,106 injuries and 5,046 fatalities.

At the city level, **Chicago, Illinois** recorded over 2,000 fatalities, making it the city with the highest fatality count. Other cities like **Houston, Texas** and **Baltimore, Maryland** reported approximately 1,200 fatalities each.

Figure 3 illustrates the total fatalities and injuries by state, with Illinois, California, and Texas identified as the highest-risk states.

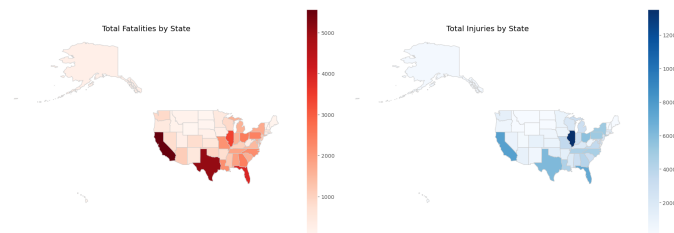


Fig. 3. Total Fatalities and Injuries by State. Illinois, California, and Texas report the highest number of gun violence incidents.

Figure 4 highlights the top 10 cities with the highest gun violence fatalities, with Chicago as the leader.

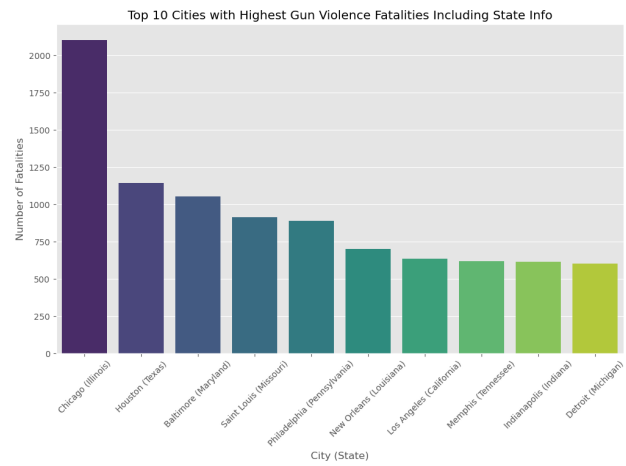


Fig. 4. Total Fatalities and Injuries by Cities report the highest number of gun violence incidents.

2) **Temporal Analysis:** The temporal analysis using the 3-month moving average revealed significant spikes in gun violence incidents during 2012–2013 and 2014–2015. These peaks indicated periods of escalation, suggesting that external factors may have contributed to these surges. The overall trend showed that gun violence incidents not only repeated over time but also exhibited increased frequency during these periods.

Figure 5 shows the smoothed time series of gun violence incidents, highlighting these spikes during 2012-2013 and 2014-2015.

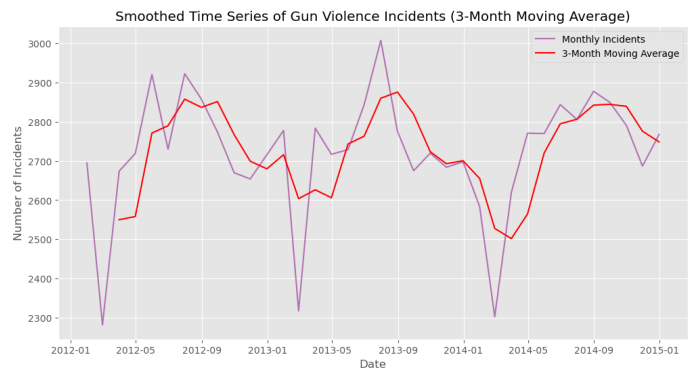


Fig. 5. Smoothed Time Series of Gun Violence Incidents

3) **Demographic Analysis:** The demographic analysis revealed significant disparities in gun violence victimization: - **Males**, particularly **Black males**, are disproportionately affected by homicides. - **White individuals** are more affected by suicides. - Victims with lower educational levels, particularly in the **Black community**, are disproportionately affected by gun violence, indicating a connection between education and vulnerability to gun violence.

Figure 6 visualizes the victims by education level and race, and

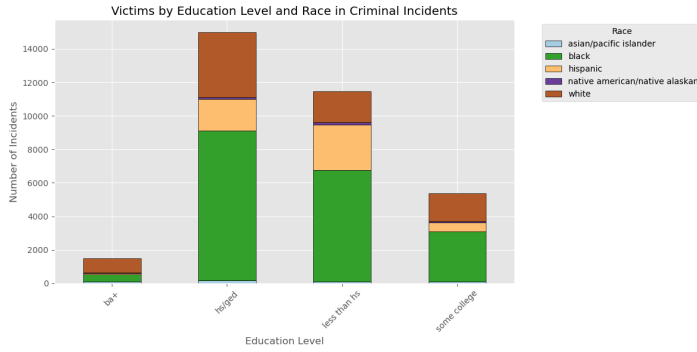


Fig. 6. Victims by Education Level and Race in Criminal Incidents. The chart reveals disparities in gun violence victims across education levels and racial groups.

Figure 7 shows the gender and race breakdown for the perpetrators.

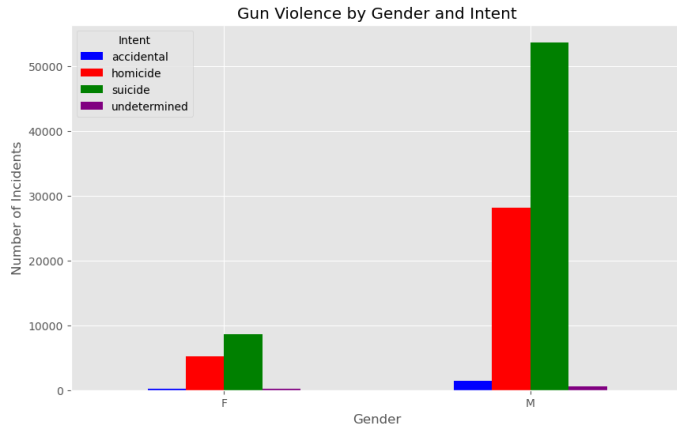


Fig. 7. Criminal's Gender and Race. The chart reveals disparities in gun violence criminals across gender and racial groups.

### C. Interpretation

Our analysis reveals significant geographic, temporal, and demographic disparities in gun violence across the United States. Our analysis yields several key observations:

1) **Geographic Hotspots:** **Illinois, California, and Texas** emerge as the states most affected by gun violence, with **Chicago** standing out as the city with the highest fatalities. This suggests that targeted interventions in these regions, particularly in Chicago, are critical for reducing the incidence of gun violence.

The results of the geographic analysis (Figure 3 and Figure 4) clearly highlight the need for focused efforts in these high-risk areas to prevent and reduce gun violence.

2) **Repeated Incidents and Escalation:** Gun violence exhibits cyclical patterns, with notable spikes in incidents during 2012–2013 and 2014–2015. These patterns suggest that certain areas are prone to periods of escalation. Proactive interventions should be planned during these periods to mitigate the effects of these surges.

Figure 5 reinforces this finding, showing the smoothed time series and the peaks during the identified periods of escalation.

### 3) Demographic Disparities:

- **Males**, particularly **Black males**, are disproportionately affected by homicides, while **White individuals** are more affected by suicides. This highlights the need for gender and race-specific interventions to address these disparities in gun violence.
- **Education** plays a significant role in vulnerability, with individuals from lower educational backgrounds, particularly in the **Black community**, being more likely to fall victim to gun violence. Socioeconomic factors must be considered in designing preventive measures.

Figure 6 and Figure 7 clearly demonstrate the significant disparities across gender, race, and education levels, underscoring the importance of addressing these factors in gun violence prevention efforts.

In summary, the results provide insights into the geographic hotspots of gun violence, the escalation patterns over time, and the demographic disparities that exist in gun violence victimization and perpetration. These findings answer the key research questions by highlighting high-risk areas, identifying periods of escalation, and revealing the underlying demographic factors that should be addressed in future interventions.

## IV. CONCLUSION

This study examined the geographic, temporal, and demographic factors contributing to gun violence in the United States. The analysis revealed that **Illinois, California, and Texas** are the states most severely impacted, with **Chicago** showing the highest fatality rates. Significant spikes in gun violence were observed during 2012–2013 and 2014–2015, suggesting periodic escalations. Furthermore, demographic analysis highlighted disparities, with **Black males** being disproportionately affected by homicides and individuals with lower education levels, particularly in the **Black community**, being more vulnerable to gun violence.

These findings underscore the need for targeted interventions in high-risk regions and during high-risk periods. Policy recommendations include focusing on geographic hotspots, proactively addressing periods of escalation, and tackling the socioeconomic factors that increase vulnerability to gun violence. Further research should consider additional factors such as gun control laws and mental health issues to provide a more comprehensive understanding of the issue. Ultimately, this study provides a foundation for data-driven strategies to reduce gun violence and its impact on society.

## REFERENCES

- [1] FiveThirtyEight, "Gun Deaths Dataset," available at: [https://raw.githubusercontent.com/fivethirtyeight/guns-data/master/full\\_data.csv](https://raw.githubusercontent.com/fivethirtyeight/guns-data/master/full_data.csv), accessed on Nov. 26, 2024.
- [2] Jamesqo, "Gun Violence Incident Data," available at: [https://github.com/jamesqo/gun-violence-data/raw/master/DATA\\_01-2013\\_03-2018.tar.gz](https://github.com/jamesqo/gun-violence-data/raw/master/DATA_01-2013_03-2018.tar.gz), accessed on Nov. 26, 2024.
- [3] Creative Commons, "Creative Commons Attribution 4.0 International (CC BY 4.0) License," available at: <https://creativecommons.org/licenses/by/4.0/>, accessed on Nov. 26, 2024.