SI 618 Final Project Report

Examining the Relationship between Poverty and Gun Violence in the United States

Siyuan Cao (siyuansc), Yilin Li (yilinliz), Ziyiyang Wang (ziyiyanw)

1. Motivation

Gun violence in the United States is a devastating problem resulting in more than 100 fatalities daily. This issue is compounded by the fact that millions of Americans struggle to make ends meet and experience poverty, exacerbating the problem of gun violence. The effects of poverty are far-reaching and can contribute to a wide range of social problems, including crime, violence, and inequality.

To address this issue, our final project seeks to analyze the relationship between poverty and gun violence at the national and state levels. By conducting a comprehensive analysis, we hope to gain new insights into the specific factors contributing to poverty-related gun violence and identify the most affected states. Our analysis will also explore the potential solutions that can effectively reduce both poverty and gun violence and develop evidence-based policy recommendations to guide policymakers.

Through this initiative, we hope to provide a better understanding of the complex relationship between poverty and gun violence in the United States. By identifying effective strategies for addressing these issues, we aim to help reduce the impact of gun violence on individuals and communities across the country. This project has the potential to bring about positive change and improve the lives of millions of Americans who are impacted by these critical social issues.

2. Data Sources

In order to conduct our analysis, we will use two primary datasets: the "Gun Violence Data" and the "Poverty Statistics" datasets.

a. The "Gun Violence Data" Dataset

The "Gun Violence Data" dataset is a comprehensive source of information about gun violence incidents in the United States from 2013 to 2018. This dataset contains a wealth of information about each incident, including the date, time, and location of the incident, the number of individuals who were injured or killed, the types of weapons used, and other important details such as the race and gender of the victims and the location of the incident. The data is stored in a CSV format. Access to this dataset can be obtained through the Kaggle website.With over 239,000 records and a file size of approximately 75 MB, this dataset is quite large and offers a wealth of information that we can use to better understand the relationship between poverty and gun violence.
The URL of "Gun Violence Data" dataset:
https://www.kaggle.com/datasets/jameslko/gun-violence-data)

The "Gun Violence Data" dataset has 29 columns and contains the features shown in the table below.

variable	description	variable	description

incident_id	gunviolencearchive.org ID for incident	location_description	description of location where incident took place
date	date of occurrence	longitude	longitude where incident took place
state	state where incident took place	n_guns_involved	number of guns involved
city_or_county	city or county where incident took place	notes	additional notes about the incident
address	address where incident took place	participant_age	key: participant ID
n_killed	number of people killed	participant_age_group	key: participant ID, value: description of age group, e.g. 'Adult 18+'
n_injured	number of people injured	participant_gender	key: participant ID, value: 'Male' or 'Female'
incident_url	link to gunviolencearchive.org webpage containing details of incident	participant_name	key: participant ID
source_url	link to online news story concerning incident	participant_relationship	key: participant ID, value: relationship of participant to other participants
incident_url_fields_missing	ignore, always False	participant_status	key: participant ID, value: 'Arrested', 'Killed', 'Injured', or 'Unharmed'
congressional_district		participant_type	key: participant ID, value: 'Victim' or 'Subject-Suspect'
gun_stolen	key: gun ID, value: 'Unknown' or 'Stolen'	sources	links to online news stories concerning incident
gun_type	key: gun ID, value: description of gun type	state_house_district	
incident_chara	list of incident characteristics	state_senate_district	

cteristics			
latitude	latitude where incident took place		

b. The "Poverty Statistics" Dataset

The "Poverty Statistics" dataset, compiled by the US Census Bureau, provides information about poverty rates in various American areas and demographic groups, as well as information about poverty criteria and rates of social program participation. To access the "Poverty Statistics" dataset, we will use the Census Bureau's API, which allows us to extract the relevant data and save it into a CSV file. This dataset will allow us to explore the relationship between poverty and gun violence in more detail and to identify which areas and groups are most affected by poverty-related gun violence.
The URL of "Poverty Statistics" dataset: https://www.census.gov/data/developers/data-sets/Poverty-Statistics.html,

By combining these two datasets and conducting a thorough analysis, we hope to gain new insights into the relationship between poverty and gun violence in the United States. We plan to use this information to develop evidence-based policy recommendations that can help address poverty and gun violence and reduce their impact on American communities.

The "Poverty Statistics" dataset has 11 columns and contains the features shown in the table below.

variable	description	variable	description
YEAR	the year of the poverty data (2013-2018)	SAEPOVRT0_17_PT	the estimate of the percentage of people in the county under the age of 18 living in poverty
NAME	the name of the county	SAEPOVALL_PT	all ages in Poverty, Count Estimate
SAEMHI_LB90	the lower bound of the 90% confidence interval for the median household income in the county	SAEPOVRTALL_PT	all ages in Poverty, Rate Estimate
SAEMHI_PT	the estimate of the median household income in the county	state_name	the full name of the state
SAEMHI_UB90	the upper bound of the 90% confidence interval for the median household income in	abv	the two-letter abbreviation of the state

	the county		

3. Data Manipulation Methods

a. The "Gun Violence Data" Dataset Preprocessing

- Columns including "incident_id", "address", "incident_url", "source_url", "incident_url_fields_missing", "congressional_district", "gun_stolen", "gun_type", "incident_characteristics", "latitude", "location_description", "longitude", "n_guns_involved", "notes", "participant_age", "participant_name", "participant_relationship", "participant_status", "participant_type", "sources", "state_house_district", "state_senate_district" are meaningless for subsequent analysis, so we removed these features.
- We converted date to datetime object in order to facilitate subsequent data processing.
- We got four features of "year", "month", "monthday", and "weekday" from the original feature "date".
- We got a new feature of "loss" by adding the features "n_killed" and "n_injured".
- We converted the original feature "state" to state abbreviation.
- Some features have special formats. Each list is encoded as a string with separator "||". Each dict is encoded as a string with outer separator "||" and inner separator "::". For the convenience of subsequent processing, we removed special symbols and deduplicated.
- We filled the missing data with "unknown".
- After careful screening of the data, we did not find any obvious outliers.

b. The "Poverty Statistics" Dataset Preprocessing

- We cleaned the dataset and filled NA using KNN.

c. Merge of Two Datasets

- Because we want to analyze the relationship between economic level and incidence of gun violence, we used inner join with the keys “State” and “City_Or_County” to make sure the results are matched in the merged results.

4. Analysis

Four questions were created based on the topic and the two available datasets prior to conducting an in-depth manipulation of the data. These questions include:

- Is the frequency of gun violence related to location?
- What are the temporal trends in the frequency of gun violence incidents?
- How do participants in gun violence incidents differ in terms of age and gender?
- What is the relationship between the frequency of gun violence and the local economic level?

Using these questions as a guide, we conducted a thorough data manipulation process and organized the findings into four categories: geographic differences, temporal differences, age and gender differences, and the correlation between economic status and crime frequency. We will explain each category in detail.

To the first section of the analysis, we focused on examining whether the frequency of gun violence incidents is related to location. To answer this question, we analyzed the dataset on gun violence incidents by state, and created a bar chart that showed the frequency of gun violence incidents in each state.

The bar chart revealed that certain states had much higher frequencies of gun violence incidents compared to others. For example, Illinois, California, Texas, Florida, and Ohio had the highest numbers of incidents. Meanwhile, other states such as Vermont, Rhode Island, Wyoming, and Hawaii had much lower frequencies of gun violence incidents.

This suggests that there is indeed a relationship between the **frequency of gun violence incidents and location**. Certain states may have underlying factors that contribute to higher rates of gun violence, such as a higher prevalence of firearms or social factors such as poverty, crime rates, and gang activity. Understanding these factors can help us develop targeted interventions and policies that are specific to each location, to address the root causes of gun violence and improve public safety.

In the second section of our analysis, we examined the **time trend of the frequency of gun violence incidents**. To answer this question, we analyzed the data set on gun violence incidents by month and year, and created a line chart that showed the trend of gun violence incidents over time.

The line chart revealed that there was a consistent and alarming increase in the frequency of gun violence incidents over time. Specifically, we observe a spike in gun violence in the five selected states, Illinois, California, Texas, Florida, and Ohio, with a high frequency of shootings, particularly from 2016 to 2017.

These findings suggest that the time trend of gun violence incidents is an important factor to consider when developing policies and interventions to reduce gun violence. For example, policymakers could consider increasing funding for violence prevention programs during peak years. Additionally, understanding the reasons behind the peaks in gun violence incidents can help us develop targeted solutions to reduce gun violence in those specific time periods.

To the third section of our analysis, we focused on exploring the **differences in age and gender of gun violence participants**. We used various data visualization techniques such as histograms, bar charts, and scatterplots to gain insights into the characteristics of those involved in gun violence incidents. Most of the crimes involving gun violence are committed by males, almost five times more than females. This suggests that special attention needs to be paid to the male population in terms of crime prevention and control, and that measures need to be taken to prevent and reduce the likelihood of male involvement in gun violence incidents. Besides, the majority of those involved in gun violence crimes are adults. This suggests that incidents of gun violence often involve social, psychological, and economic factors for adults, and that targeted prevention and countermeasures for adults need to be developed.

In the fourth block of our analysis, we explored the relationship between the **frequency of gun violence incidents and the local economic level**. To do so, we examined the unemployment rate and median household income in the areas where the gun violence incidents occurred.

We found that areas with lower median household incomes also tended to have higher frequencies of gun

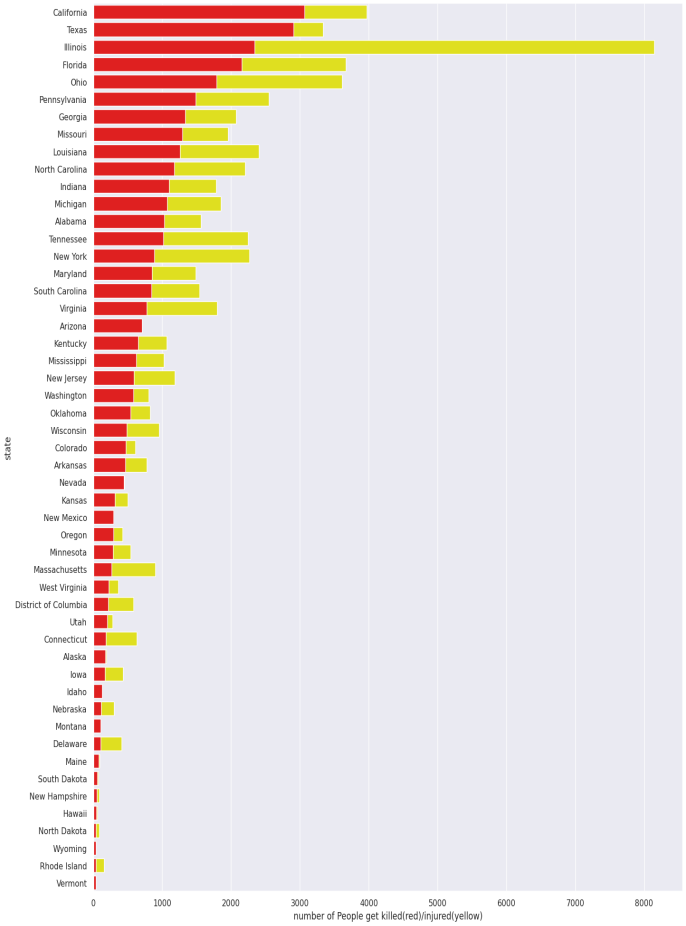
violence incidents, further supporting this correlation.

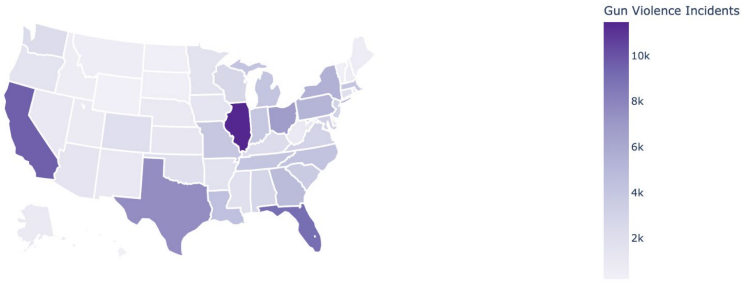
These findings suggest that addressing economic disparities may be an important step in reducing the incidence of gun violence. By focusing on job creation and economic development in areas that are most impacted by gun violence, we may be able to reduce the occurrence of these incidents and improve the overall safety and well-being of these communities.

5. Visualization

a. Geographic differences

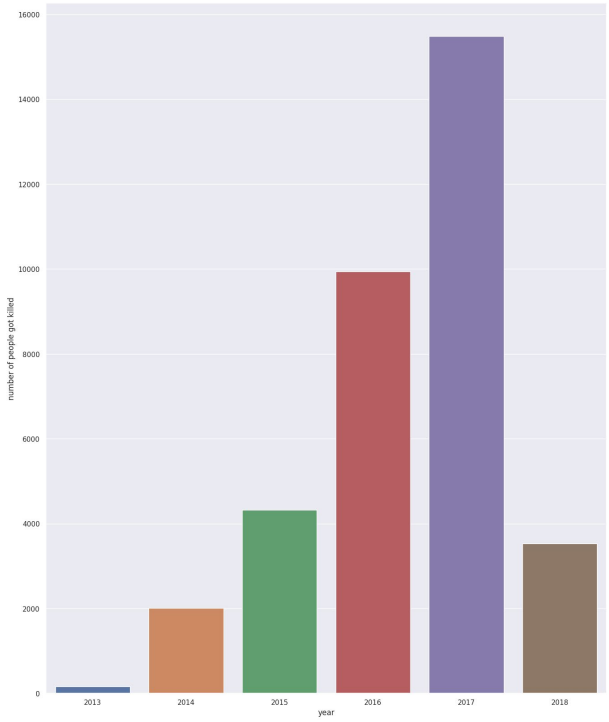
Visualize the number of people getting killed or getting injured during gun violence incidents by different states. Based on the chart, it appears that there is a wide variation in the number of people killed and injured in gun violence incidents across different states. California, Texas, and Florida have some of the highest numbers of both injuries and fatalities, which is expected given their large populations and high crime rates. However, there are also some states with relatively low populations that have high numbers of injuries and fatalities, such as Alaska and New Mexico. This could be due to factors such as high rates of gun ownership or a prevalence of gun violence related to drug trafficking or gang activity.





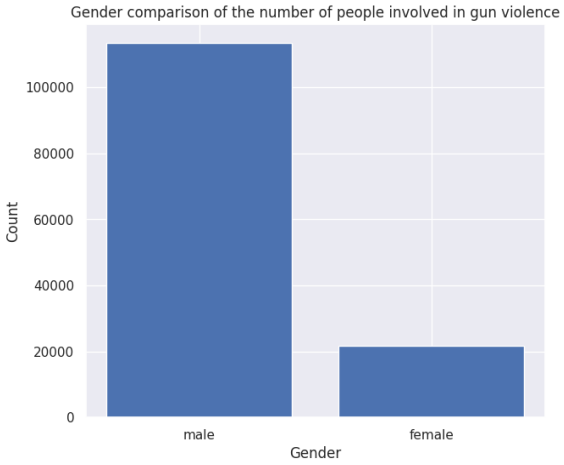
b. Temporal differences

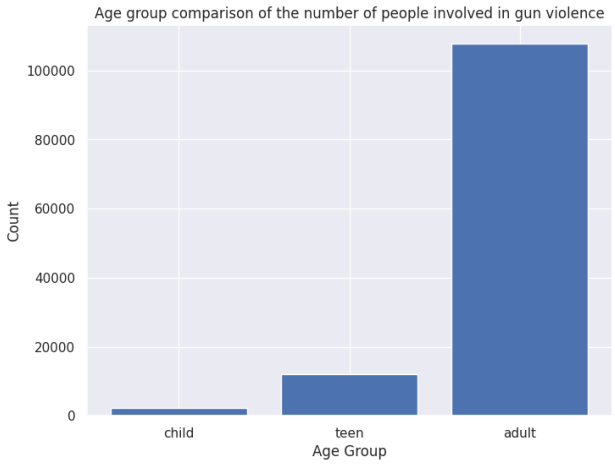
Visualize how many people were killed per year in all states. Based on the chart, it appears that there has been a general upward trend in the number of people killed and injured in gun violence incidents in the United States over the past decade. The number of people killed and injured reached a peak in 2017, with over 39,000 people killed and nearly 81,000 people injured in gun violence incidents that year. While there was a slight decrease in the number of people killed and injured in 2018 and 2019, the numbers remained high compared to earlier years.



c. Age and gender differences

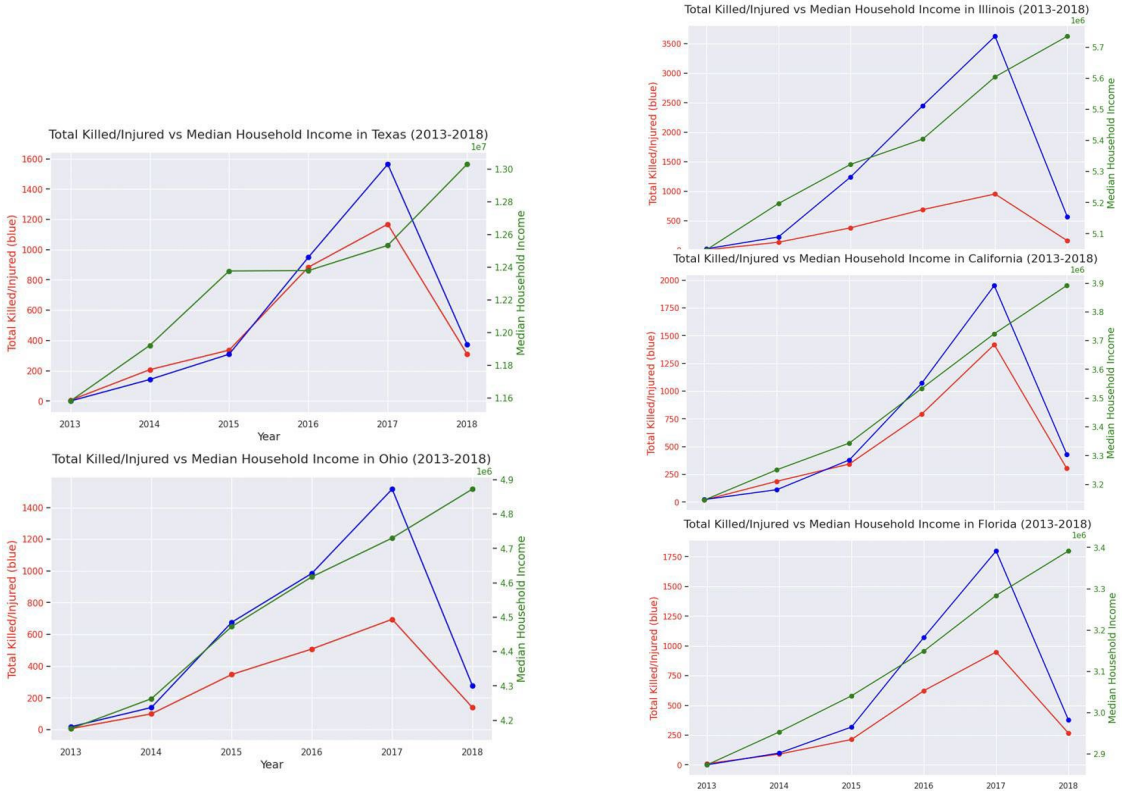
Visualize gender and age group comparison of the number of people involved in gun violence.





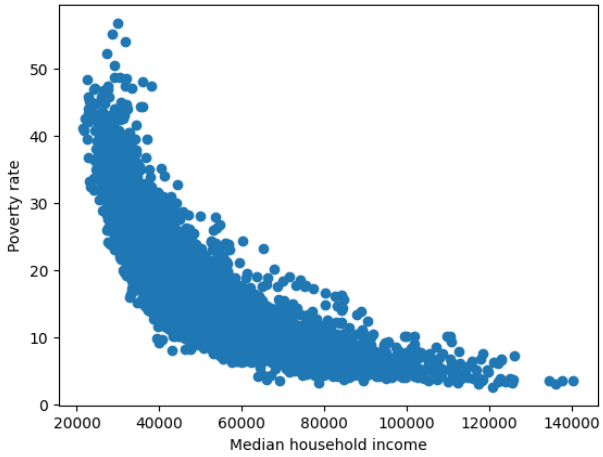
d. The correlation between economic status and crime frequency

Visualize the relationship between gun violence incidents and median household income in five states which occurred between 2013 and 2018. As median household income increases, the number of gun violence incidents also tends to increase. This trend is most apparent in the states of California and Ohio, where there is a clear trend of increasing gun violence incidents as median household income increases.

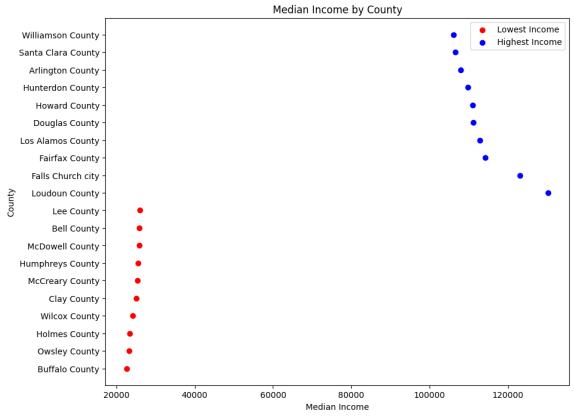


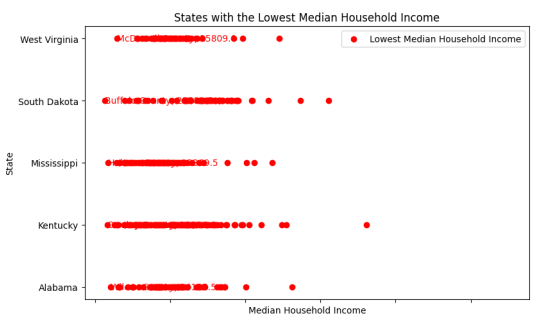
Visualize the relationship between poverty rate and median household income using a scatter plot. Based on the chart, it appears that there is a negative correlation between median household income and poverty rate. In other words, as median household income increases, the poverty rate decreases. This relationship is

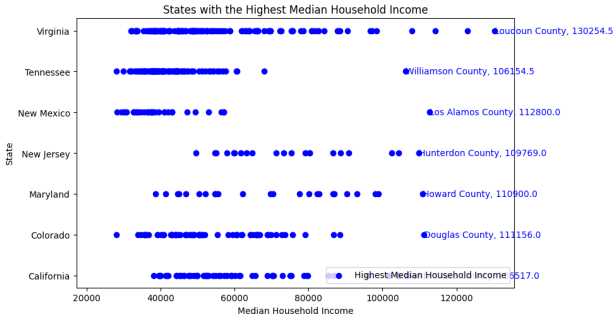
expected since higher household incomes generally mean more financial resources available to meet basic needs and invest in education, health, and other factors that can reduce the likelihood of living in poverty.



Visualize the lowest and the highest median household income in the county. In both charts, it is clear that there is a significant variation in median household income between states. The states with the highest median household income are generally located on the East and West coasts, while the states with the lowest median household income are mostly located in the South and Appalachia regions. The charts also show that there is a general trend of higher median household income in more urban and densely populated states.







6. Statement of Work

Our team collaborated closely throughout the project, with each member taking on specific roles and responsibilities. For this final report, Siyuan Cao was responsible for writing the project motivation and data sources part, Yilin Li focused on the data manipulation and analysis part, and Ziyiyang Wang handled visualization and other aspects of the report. To the code we provided, we worked together on creating a Google Drive file where we collected, organized, and analyzed the data. We held regular check-ins and meetings to discuss progress and plan the next steps. Overall, we had a positive and productive collaboration, and we believe that our teamwork was key to the success of the project.