Analyzing Gun Violence Trends from 2014-2021

W200 - Project 2, Aug 2021

Florencia Froebel, Emily Huang, James Kelly

Github repository: UC-Berkeley-I-School/Project2_Froebal_Huang_Kelly (github.com)
Primary dataset: **U.S. Gun Violence Records 2014-2021 | Kaggle**

Data structure

There are 3230 observations each signifying a reported gun incident. And there are 7 variables, including:

- **Incident_id:** This will help us uniquely identify each incident
- **Incident_date:** Using this variable we can analyze **when** incidents tend to occur
- **State: Using this** variable we can analyze **where** incidents tend to occur
- **City_or_county:** Using this variable we can analyze **where** incidents tend to occur
- **Address:** Using this variable we can analyze **where** incidents tend to occur
- **Killed (Integer):** Using this variable we can differentiate between different gun violence incidents where someone was fatally shot. This is important because fatal gun violence incidents may be less common than gun violence injuries
- **Injured (Integer):** Using this variable we can identify how many people were injured in an incidents

Out of these variables, **incident_date**, **state**, **city_or_county**, **killed**, and **injured** were the most useful for analyzing gun trends (See appendix for information regarding the variables)

Introduction

According to the “Educational Fund to Stop Gun Violence” website, “more than 100 Americans die by gun violence”. We all know how preventable yet rampant gun violence has been over the years and how much it has affected communities all over the states. Because of this, we were interested in exploring the data related to this and see how much gun violence affects the states in the United States. We are specifically looking at a dataset that dates back to 2014 and covers all gun violence incidents to the end September 2021, which gives us around 3k observations. This dataset is from the Gun Violence Archive website where “incidents collected from over 7,500 law enforcement, media, government and commercial sources daily in an effort to provide near-real time data about the results of gun violence”. More details about the source can be read here: About | Gun Violence Archive

Some questions we are looking to explore are:

1. A state’s gun violence vs. National gun violence statistics
2. What are the states with the most incidents (injured and killed) due to gun violence?
    a. Are there states that have more injuries but not deaths? Vice versa?
3. What are the cities with the most incidents (injured and killed) due to gun violence?
    a. Are there cities that have more injuries but not deaths? Vice versa?

    b. How does a city and its state compare in gun violence incidents?
4. What dates/time period do most gun violence (deaths or injuries) occur generally?
    a. How does the date affect gun violence? Does it have any effect in general?
    b. Does the date affect the mortality rate of victims in a gun violence incident (i.e. are victims more likely to be injured than killed if it's around the summer, etc.)?
    c. Are there cities/states that differ from the general time period of frequent gun violence incidents? How so?

Sanity Checking

Before even starting to analyze the data, we need to make sure that the data is valid. Again, we're limiting our sanity checking to variables we're actually using in our analysis. For the variables that we want to analyze, here is what we want to make sure of:

- Incident_date: valid dates and appear between 2014 and 2021 inclusively
- State: make sure all states are actual states
- Killed (Integer): should be a zero or non-negative number
- Injured (Integer): should be a zero or non-negative number

Numeric descriptions for variables 'injured' and 'killed', and unique states in the dataset:

```
count    3228.00000
mean        4.19114
std         8.02234
min         0.00000
25%         3.00000
50%         4.00000
75%         5.00000
max       441.00000
Name: injured, dtype: float64
```

```
count    3230.000000
mean        1.057276
std         2.080927
min         0.000000
25%         0.000000
50%         1.000000
75%         1.000000
max        59.000000
Name: killed, dtype: float64
```

```
array(['Illinois', 'Minnesota', 'New York', 'California', 'Washington',
       'Tennessee', 'South Carolina', 'Pennsylvania', 'North Carolina',
       'District of Columbia', 'Missouri', 'Florida', 'Texas',
       'Louisiana', 'Georgia', 'Maryland', 'Ohio', 'Wisconsin',
       'Michigan', 'Alabama', 'New Jersey', 'Delaware', 'Kansas',
       'Nevada', 'West Virginia', 'Oregon', 'Colorado', 'Indiana',
       'Mississippi', 'Massachusetts', 'Virginia', 'Rhode Island',
       'Kentucky', 'Iowa', 'Arizona', 'New Hampshire', 'Connecticut',
       'Nebraska', 'Alaska', 'Utah', 'Arkansas', 'New Mexico', 'Oklahoma',
       'South Dakota', 'Idaho', 'Maine', 'Montana', 'Wyoming', 'Vermont'],
      dtype=object)
```

Important thing to note is Hawaii and North Dakota do not appear in the data set but the District of Columbia is included.

We also want to check if there are any empty columns in the dataset.

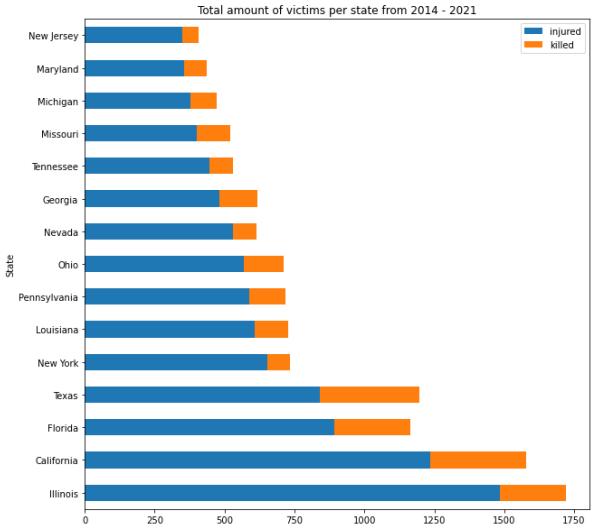
```
incident_date : 0
state : 0
city_or_county : 0
killed : 0
injured : 2
```

	incident_id	incident_date	state	city_or_county	address	killed	injured
51	2115336	2021-09-05	Missouri	Wyatt	0	4	NaN
107	2089557	2021-08-10	Mississippi	Clarksdale	2	2	NaN

Looking at the results for the empty columns check, it seems reasonable to have only these 2 observations have empty values in the injured column since there are incidents where there are no injured victims. Because all empty observations are in the same column, our fix was just to replace all 'Nan" with 0. After fixing the values, we can now proceed into some initial exploration into the data and get a general understanding of the shape of the gun violence incidents.

Initial Exploration

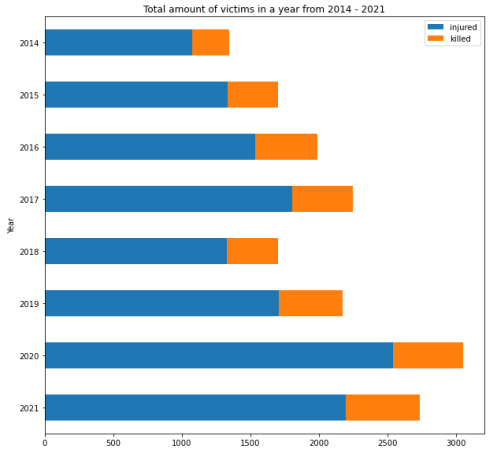
Now that we've checked to ensure the data is valid, we can begin some initial exploration of the data. From the Sanity Checking, out of the 50 states, there are 48 that appear in the dataset plus the District of Columbia. Out of the following states that appear in the dataset, here are the top states with the largest number of victims from gun violence incidents from 2014-2021.



state	injured	killed	total_num_victims
Illinois	1484.0	235	1719.0
California	1235.0	344	1579.0
Florida	892.0	273	1165.0
Texas	840.0	355	1195.0
New York	653.0	82	735.0
Louisiana	608.0	118	726.0
Pennsylvania	588.0	130	718.0
Ohio	568.0	142	710.0
Nevada	531.0	84	615.0
Georgia	482.0	135	617.0
Tennessee	446.0	83	529.0
Missouri	401.0	121	522.0
Michigan	377.0	94	471.0
Maryland	356.0	82	438.0
New Jersey	348.0	59	407.0

As you can see in the chart and the graph, the top 5 states with the most victims from 2014 - 2021 are Illinois, California, Florida, Texas, and New York. This makes sense as they are all in the top 10 states with the highest populations in the US. Interestingly enough, Louisiana appears in the top 10 of states with the largest number of victims in a state despite being only the 25th largest state in the US.

Next we wanted to check the total number of victims from gun violence incidents each year from 2014-2021.

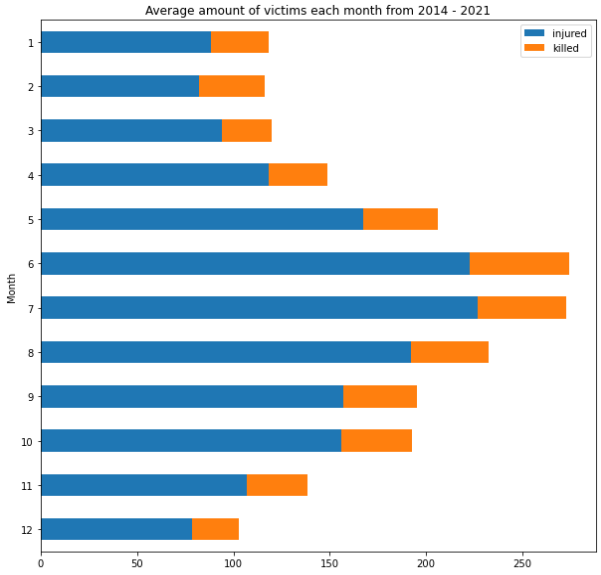


year	injured	killed
2021	2195.0	538
2020	2539.0	513
2019	1709.0	465
2018	1330.0	372
2017	1807.0	441
2016	1536.0	452
2015	1335.0	367
2014	1078.0	267

From the graph and the table, we see that there is a general upward trend of injured and killed victims. We can see that the number of injured victims increased at a faster rate than the amount of killed victims.

Because we can't see the entire picture with just the number of victims in a year, we wanted to explore the breakdown of average number of injured and killed victims by month.

From the graph and table, we see incidents garner the most amount of victims during the months of June, July, and August which coincidentally are during the Summer seasonal period.



month	injured	killed
1	88.250	30.125
2	82.000	34.125
3	94.125	25.750
4	118.375	30.250
5	167.500	38.750
6	222.500	52.000
7	226.875	46.125
8	192.250	40.250
9	157.125	38.000
10	156.250	36.500
11	107.125	31.125
12	78.750	23.875

General Gun Violence Trends in the Nation

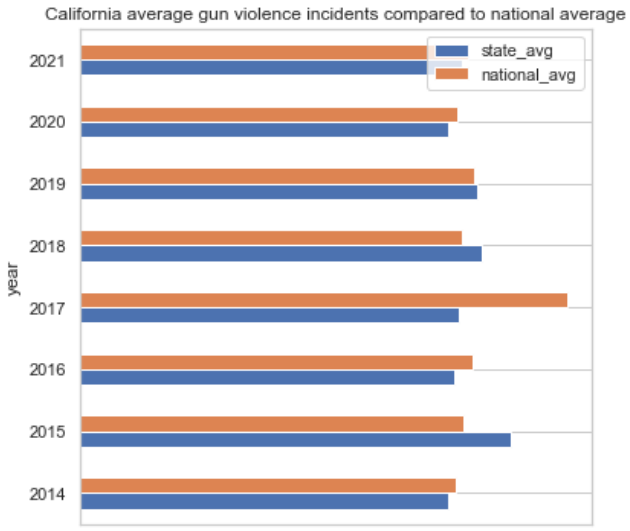
With initial exploration done, we want to start diving deeper into the data and check the statistics for gun violence across the nation before focusing on specific states. We thought it would be good to first establish national statistics like the average number of gun incidents that occur in a year and what is the mortality rate each year in relation to the average number of gun incidents each year.

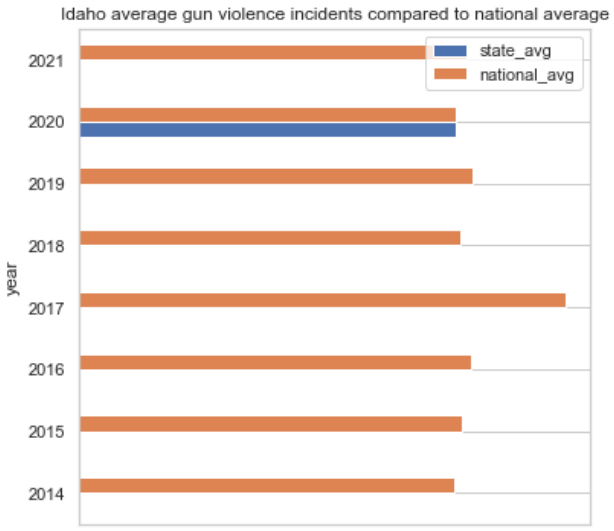
- General data analysis on gun violence trends (like injured vs. killed)

We found that the number of daily incidents of gun violence remains pretty low and constant throughout time. There are exceptional cases where there happen to be a large number of incidents but for the most part the range of occurrences stay constant. The number of deaths due to gun violence are usually much lower than injuries and happen to be 0 much more often. The average number of deaths is lower than the average number of injuries.

- General analysis of state vs national gun violence trends

When comparing different state gun violence averages vs the national average we found it was interesting that for Texas, the gun violence incidents exceeded the national average more often (and more dramatically) than most other states. For example, California and Illinois gun violence incidents tend to stick very close to the national average. Although, certain states such as New Hampshire and Idaho had gun violence averages much lower than the national averages.



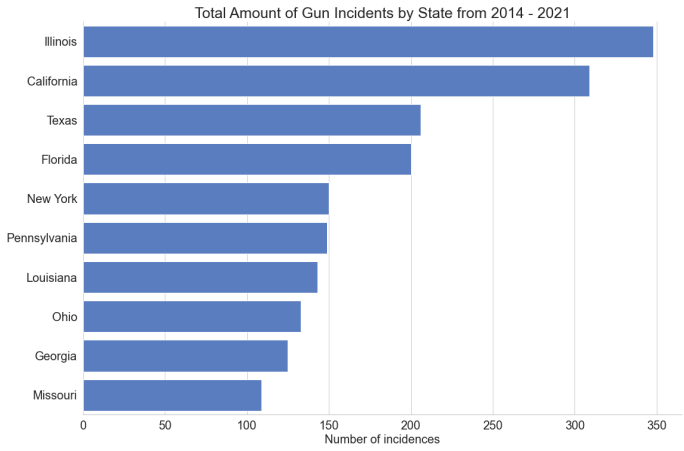


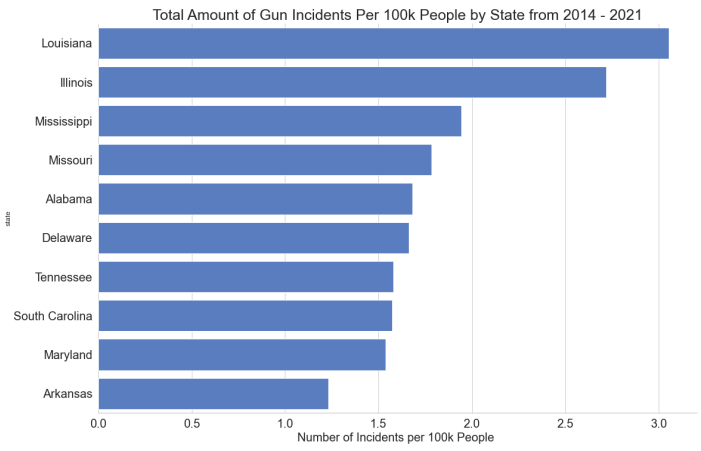
State Gun Violence Trends in the Nation

After establishing the national statistics for gun violence, we wanted to start comparing states that were either way above or way below the national average.

- What are the states with the most incidents (injured and killed) due to gun violence?

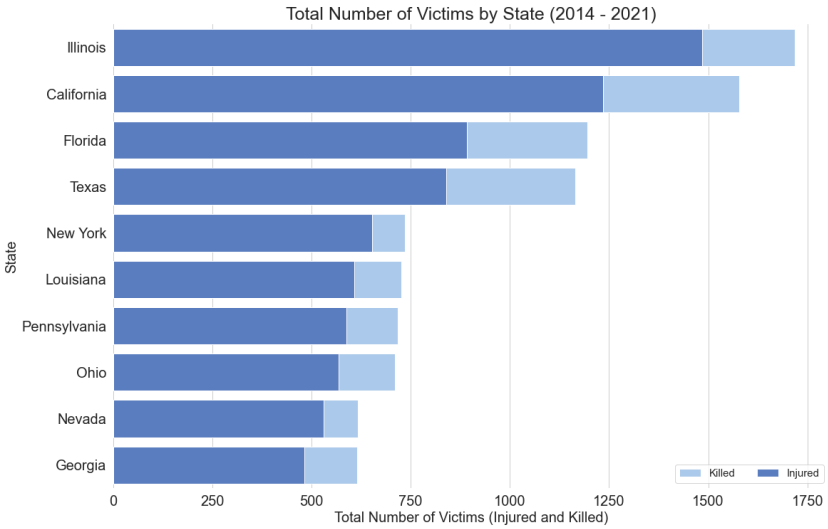
In order to determine the states with the most incidents, we first created a bar chart showing the absolute number of incidents by state. The top states were Illinois, California, and Texas. However, it seemed likely that these states were at the top of the list due to their size. To control for population effects, we scraped state population statistics from Census.gov and re-charted our findings based on state incidents per 100k people. From the updated barchart, we were able to determine that Louisiana, Illinois, and Mississippi actually had the most gun violence incidents per capita.



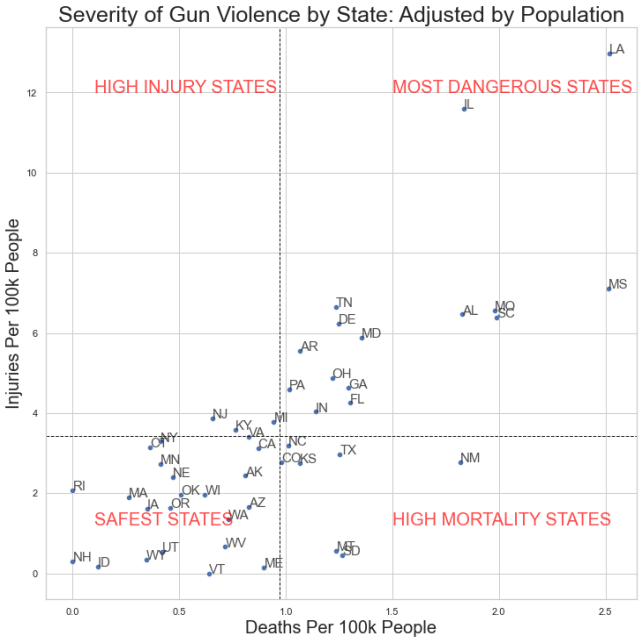


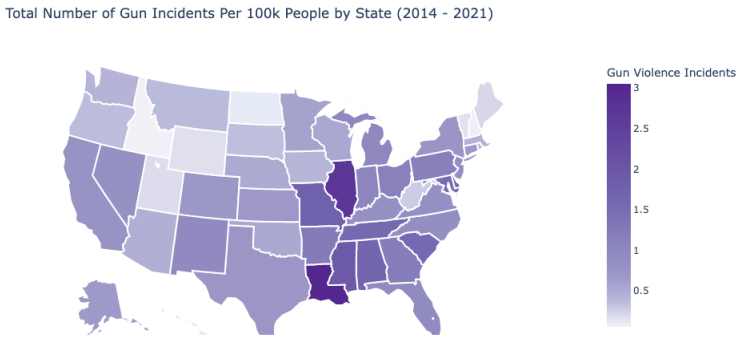
- Are there states that have more injuries but not deaths? Vice versa?

Next we looked deeper into gun violence incidents by creating a barchart of the total number of gun victims by state. We then break apart total victims by total killed and total injured. Overall, Illinois had the largest number of Victims over the 8 year time span. And a vast majority of states had more gun injuries than deaths. The only states, this wasn't true were WV, MT, ME, SD, VT. We postulate that these states had more deaths than injuries due to their small sample sizes of gun incidents.



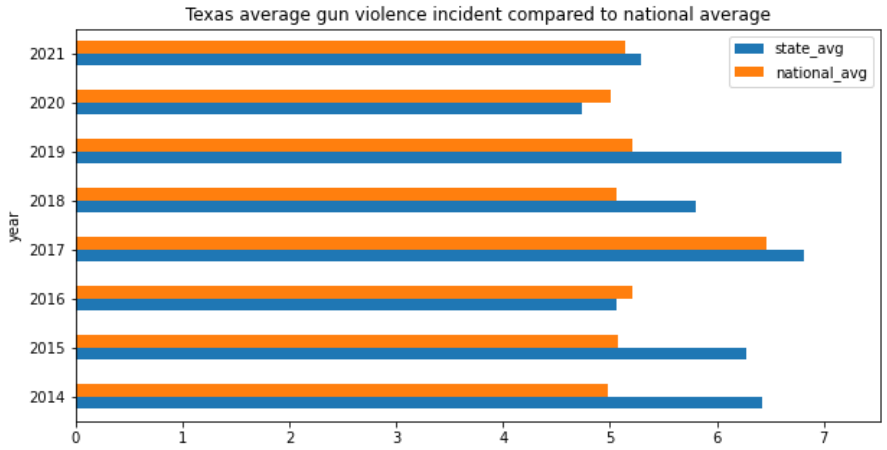
We then decided to further compare gun Deaths vs Injuries to determine the most dangerous states in America, which we defined as states with above average gun injuries, and above average gun deaths per 100k people. The Scatterplot below uses national averages of both injuries and deaths to organize states into four quadrants based on gun-violence severity. Using this ranking system, we discovered that Louisiana and Illinois were the most dangerous states per capita. This was also apparent when we looked at a state map of incidents per 100k people.





City Gun Violence Trends in the Nation

- Are there cities that have more injuries but not deaths? Vice versa?

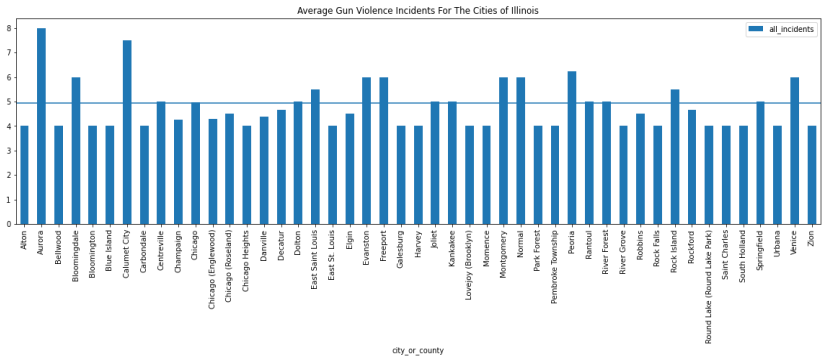


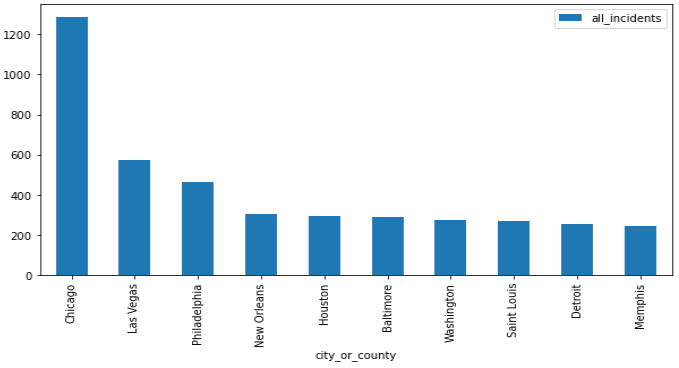
We found that: **686** unique cities or counties have a greater injured count than death count over the years, **153** unique cities or counties have a greater death count than injured count over the years and **62** unique cities or counties have the injured count equal to the death count over the years. Thus, most cities or counties have a greater injury count than death count.

- What cities had the most gun violence (injured and killed)

In descending order: Chicago, Las Vegas, Philadelphia, New Orleans, Houston, Baltimore, Washington, Saint Louis, Detroit, Memphis

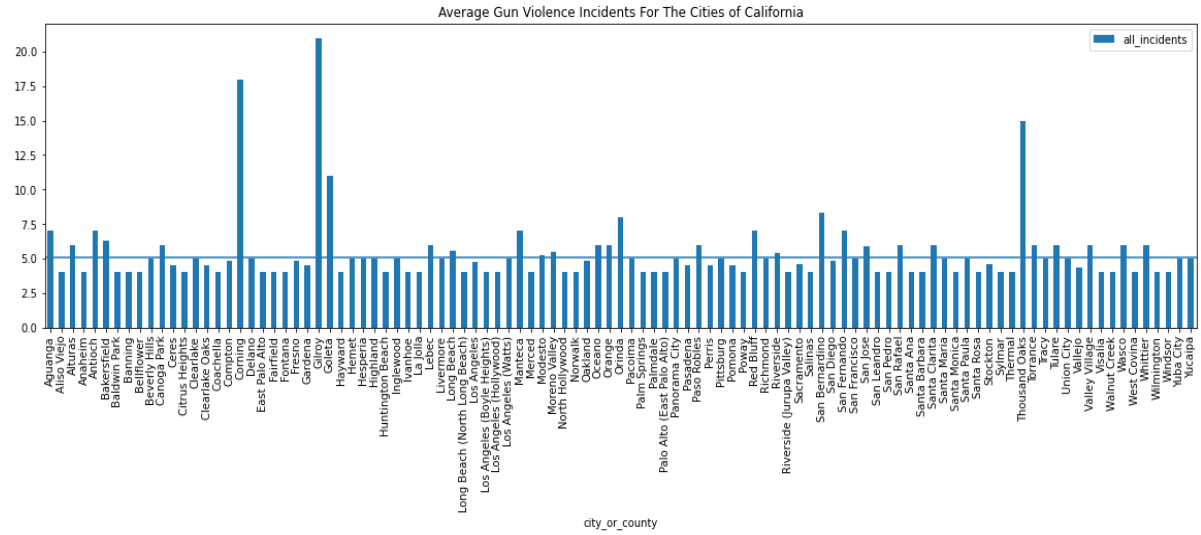
- How does a city and its state compare in gun violence incidents?

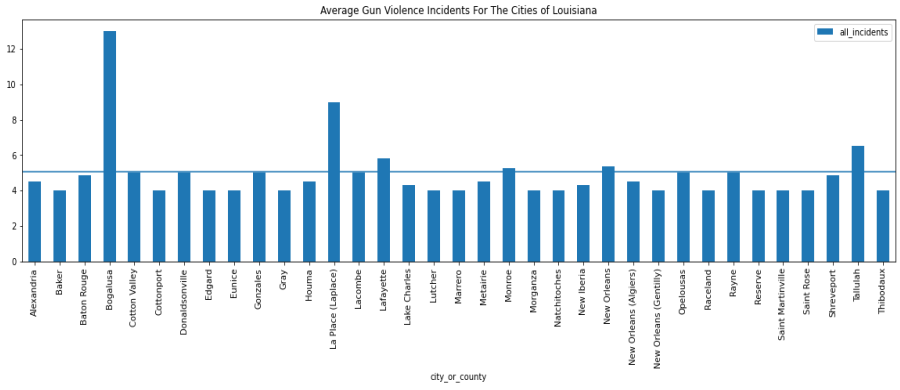


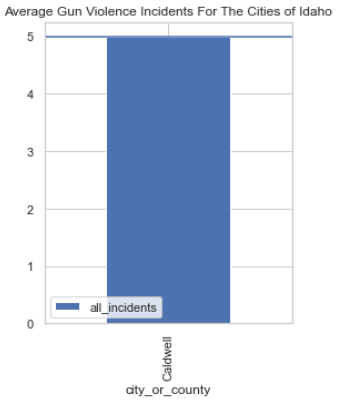


After comparing the average number of incidents for a city to its state's average gun violence incidents, we found that most cities stick close to the average and only a few cities in the state are outliers with a high number of incidents. For example, most cities within Louisiana are close to the average but the city/county of Bogalusa is an outlier since it has a large number of incidents compared to other cities. Similarly, we see this pattern continue in several cities of Illinois and California. Meanwhile, states with low averages of gun violence incidents tend to have fewer cities with reported incidents. For

example, Idaho only has one reported city/county.



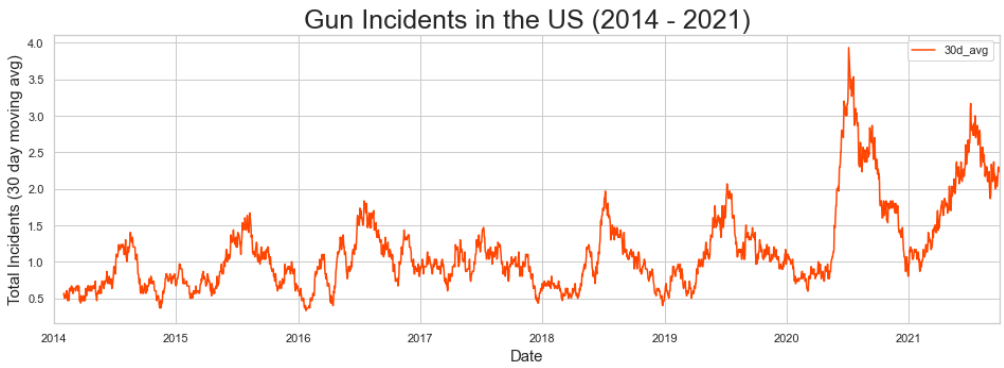




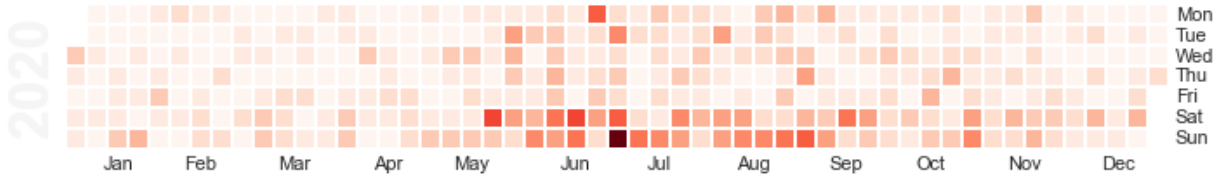
Time Period Gun Violence Trends in the Nation

To begin, we defined time period as calendar dates, days of the week, months, and years because calendar date was the most granular time period in our data set. Next, we created unique variables for all the mentioned time periods by extracting the relevant data from each incident's calendar date. This allowed us to streamline the rest of the time analysis.

- How does the date affect gun violence? Does it have any effect in general?

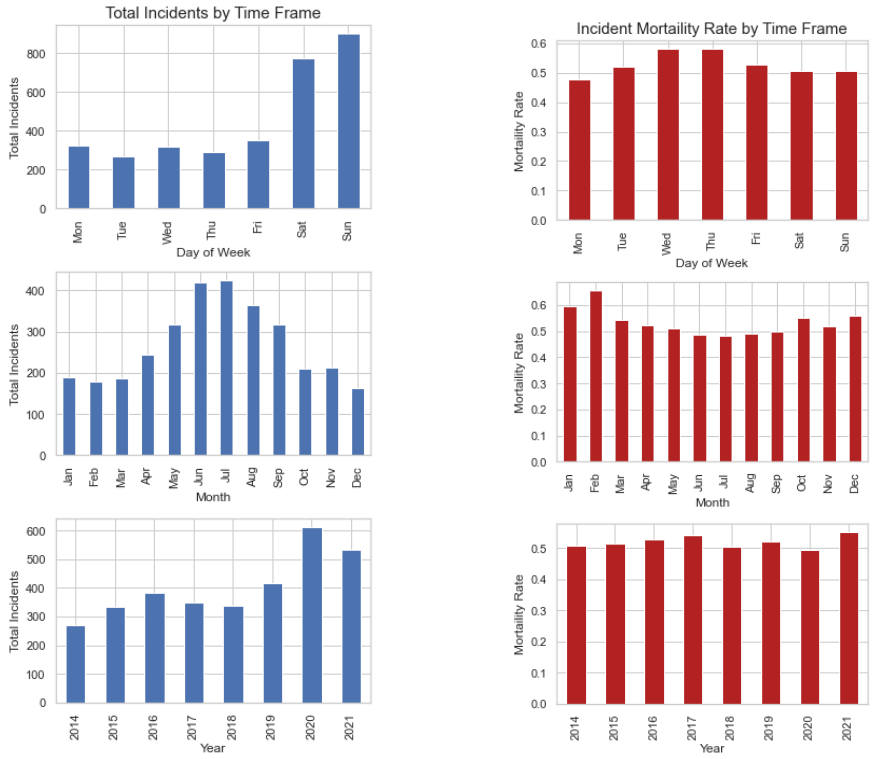


We began by visualizing a time series analysis of total gun incidents by date. Because the gun incident data was extremely noisy due to the small number of incidents, we decided to use a 30 day moving average. This showed the overall growth of gun incidents. We found that gun incidents increased dramatically at the beginning of the Pandemic. And incidents have continued to be elevated since the beginning of 2020.



Next we created a calendar heatmap to look at seasonal gun incident trends throughout the dataset (see Appendix for the entire figure). We found that gun incidents consistently peaked on July 4th every year. For instance, in 2020 there were 5.7 times more incidents on July 4th than an average day.

- Does the date affect the mortality rate of victims in a gun violence incident (i.e. are victims more likely to be injured than killed if it's around the summer, etc.)?

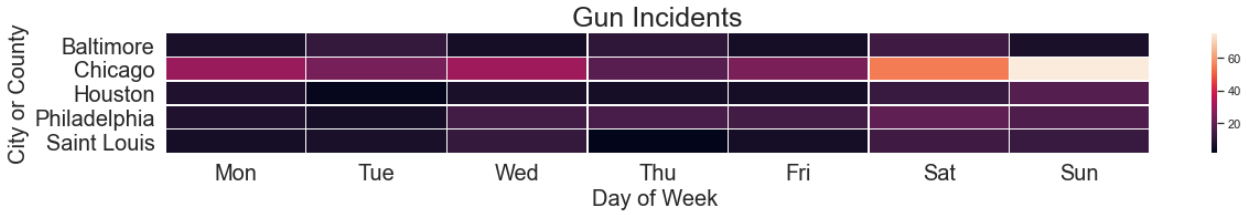


Next we analyzed day-of-week, month, and year effects. Starting with the figure on the left, we looked at the number of gun incidents by time period. We found that gun incidents peaked on

the weekend (i.e. Saturday and Sunday), in the summer (i.e. June and July), and in 2020. We compared those findings to gun incident mortality rates, defined as the percentage of gun incidents with at least 1 fatality, and we found the opposite trend. Mortality rates peaked in the middle of the week (i.e., Wednesday and Thursday), In the Winter (February), and in 2021. The difference between these two trends shows how many gun injuries occur during peak incident periods.

- Are there cities/states that differ from the general time period of frequent gun violence incidents? How so?

Finally we looked for state and city outliers by creating heatmaps of gun incidents grouped by time period (i.e. day-of-week, month, year). We found that among all states in the data set, Louisiana differs the most from national trends because gun incidents occurred at a more consistent rate, with minimal variation between time periods. Similarly, Chicago also stood out from the rest of the cities due to constantly high gun incidents across all time periods. Below is a sample heatmap showing the top 5 cities by weekday (see appendix for additional heatmaps)



Conclusion

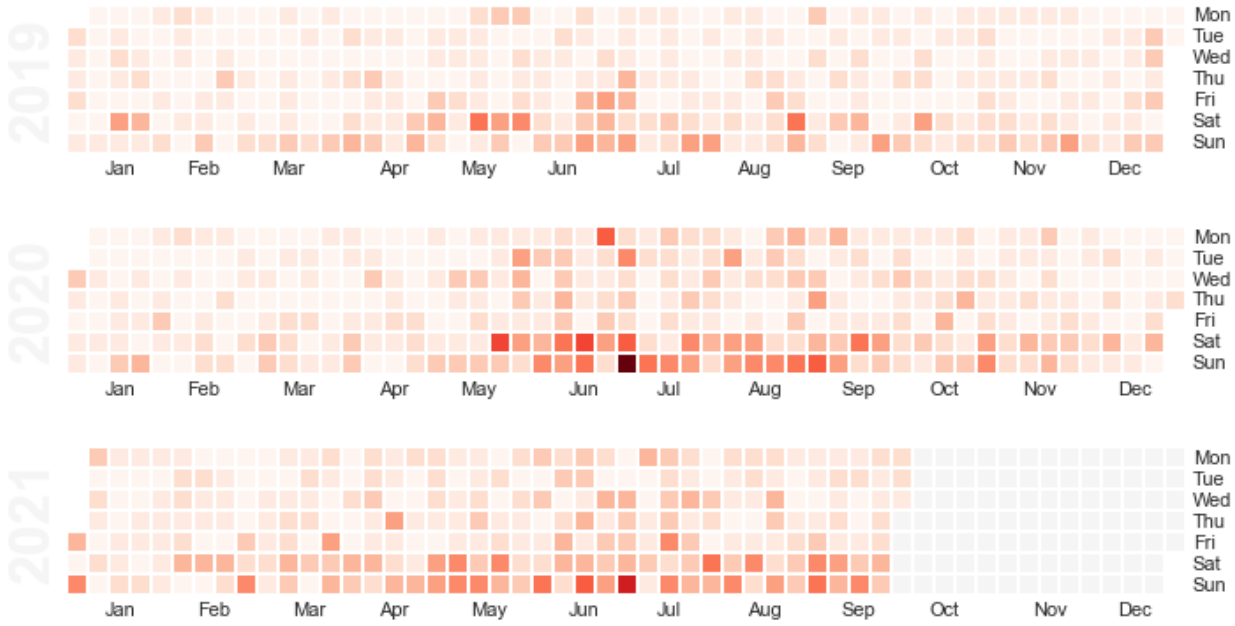
Through our analysis we concluded that there is a general rise in gun violence incidents in the US. Historically, those Incidents garner more injured victims than killed victims. Additionally, States with a higher population tend to have more victims from gun violence incidents. However, when adjusted for population, gun violence incidents seem more prominent in the east side of the country, specifically in the Midwest + South. Furthermore, A vast majority of cities have a lower rate of gun incidents compared to the state they are in. Looking across time periods, Summer is when gun violence incidents occur the most, most specifically July 4th. However, mortality rates peak in February. Top cities & states tend to be outliers because gun incidents consistently occur across the month and day of week

Appendix

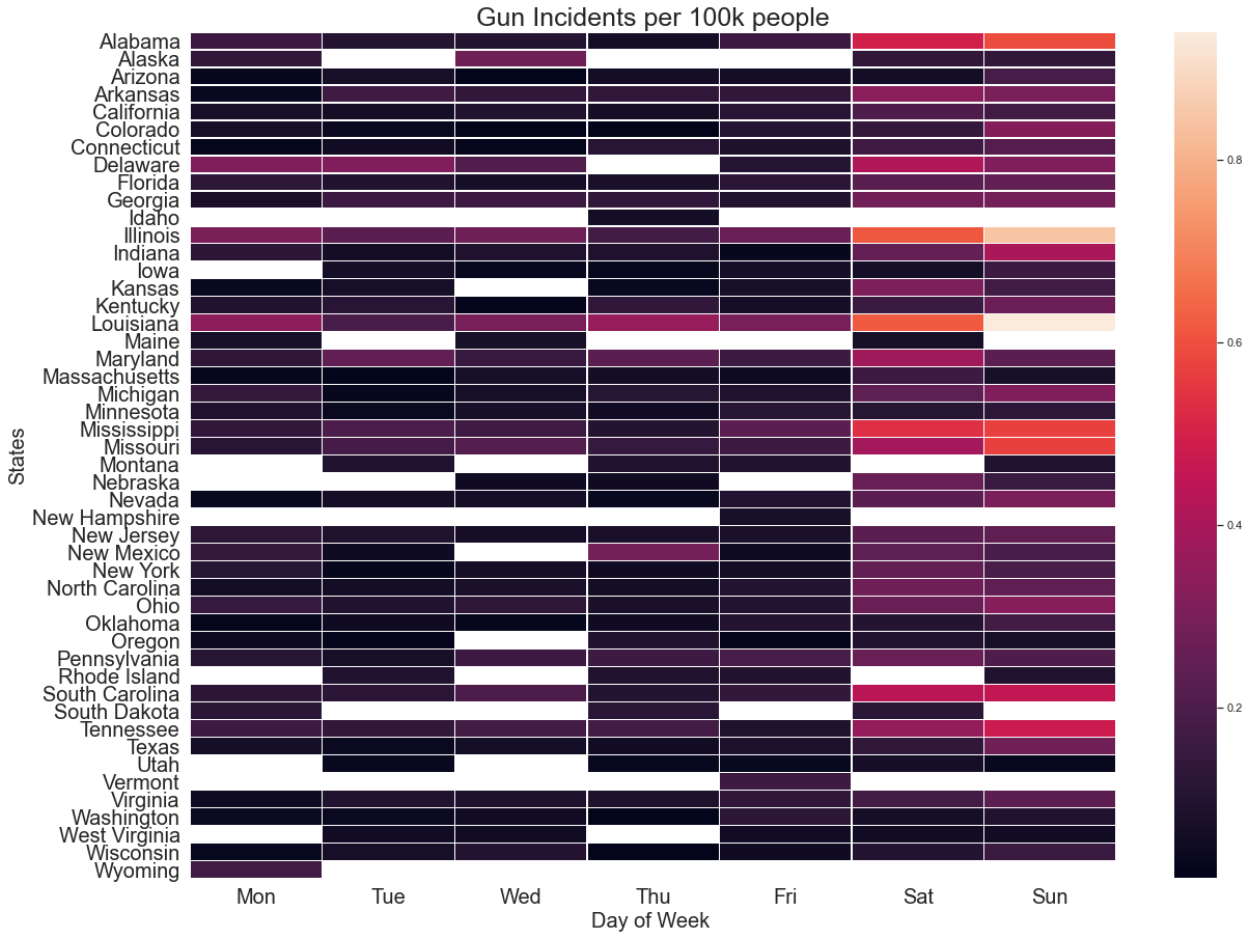
Data structure

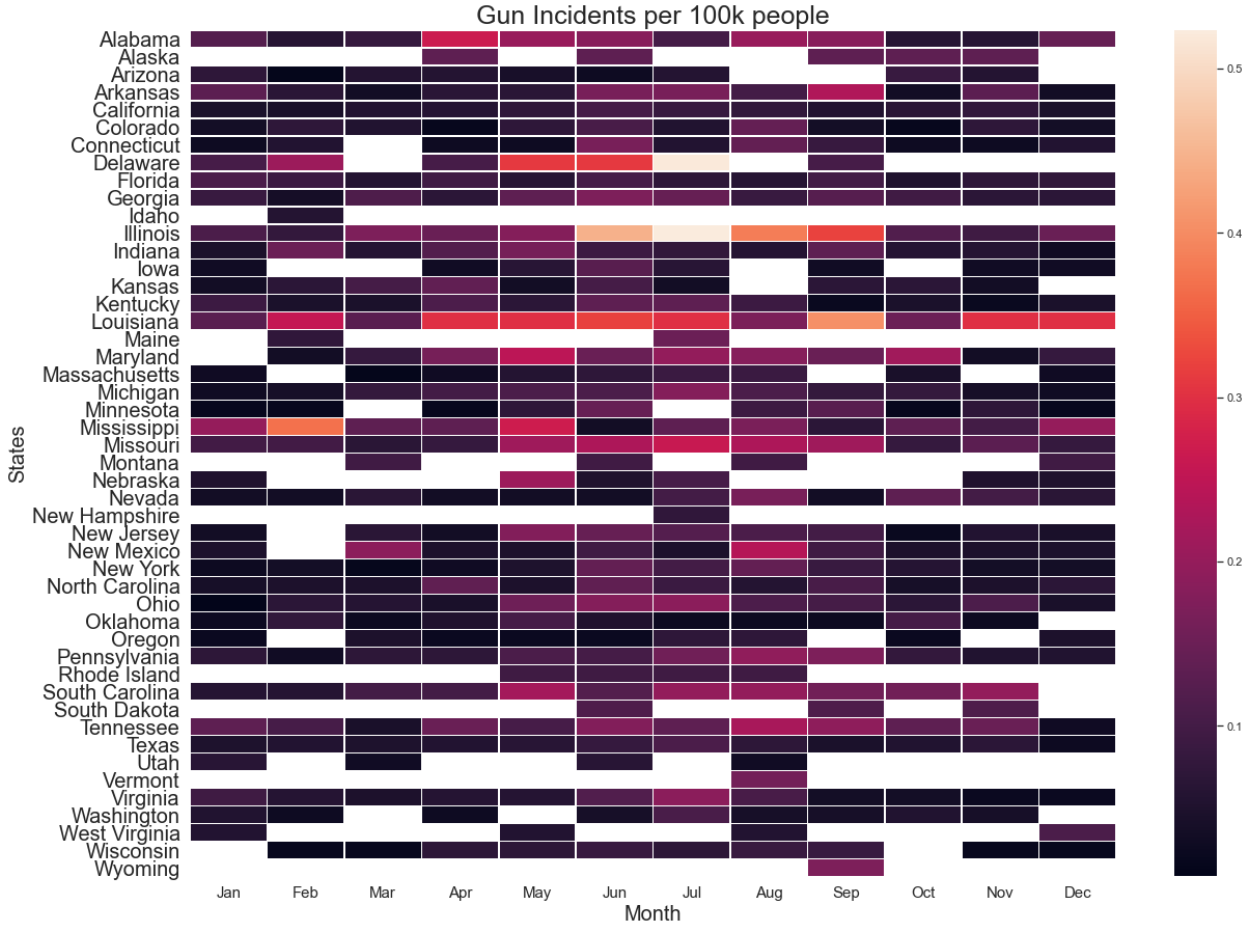
- Incident_id
    - EX: 2128023
    - This will help us uniquely identify each incident
- Incident_date
    - EX: July 3 2021
    - Using this variable we can analyze **when** incidents tend to occur the most
- State
    - EX: Illinois
    - Using this variable we can analyze **where** incidents tend to occur the most
- City_or_county
    - EX: Chicago
    - Using this variable we can analyze **where** incidents tend to occur the most
- Address
    - EX: 749 Silver Bluff Rd
- Killed (Integer)
    - EX: 0
    - Using this variable we can differentiate between different gun violence incidents where someone was fatally shot. This is important because fatal gun violence incidents may be less common than gun violence injuries
- Injured (Integer)
    - EX: 2

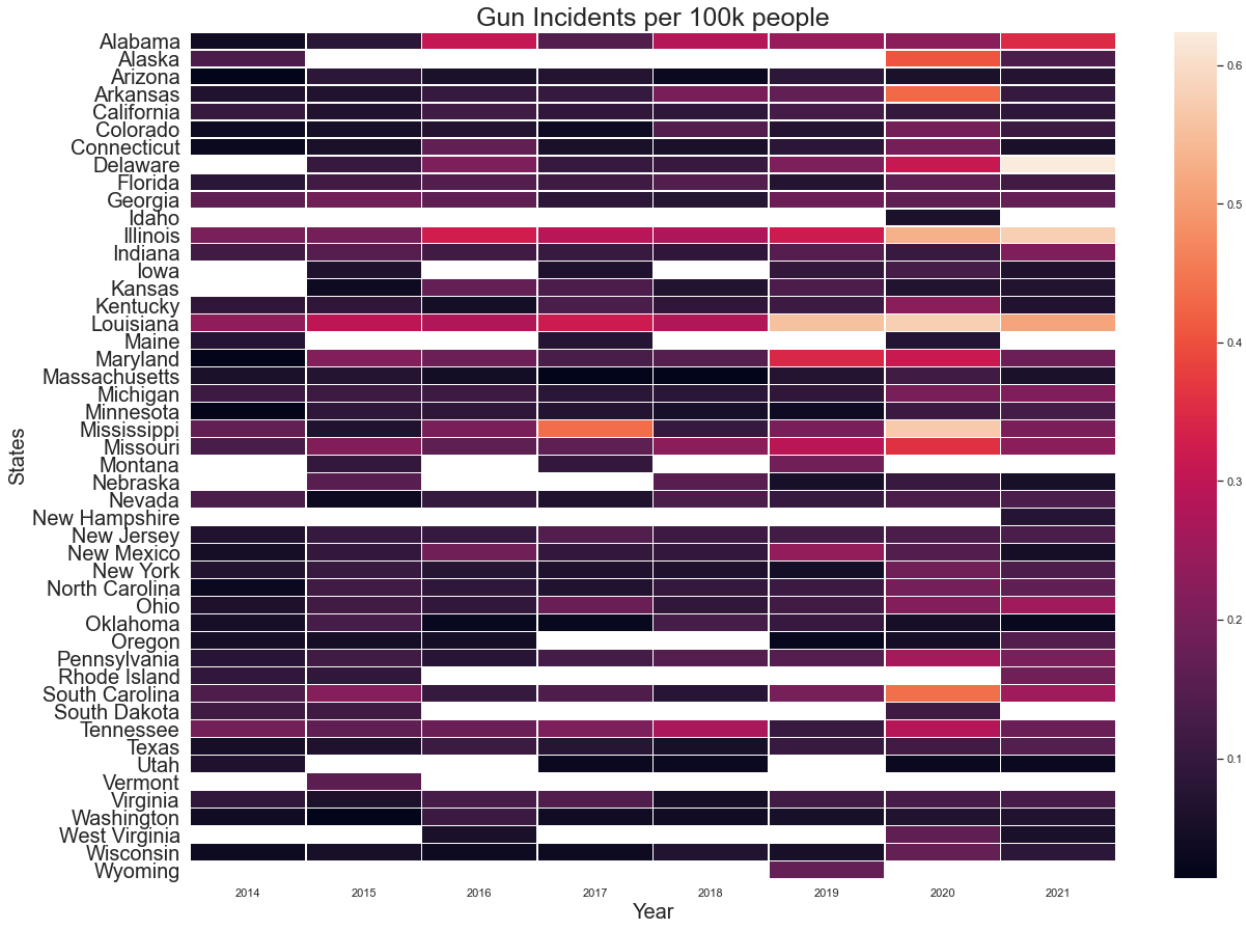
Calendar heatmap

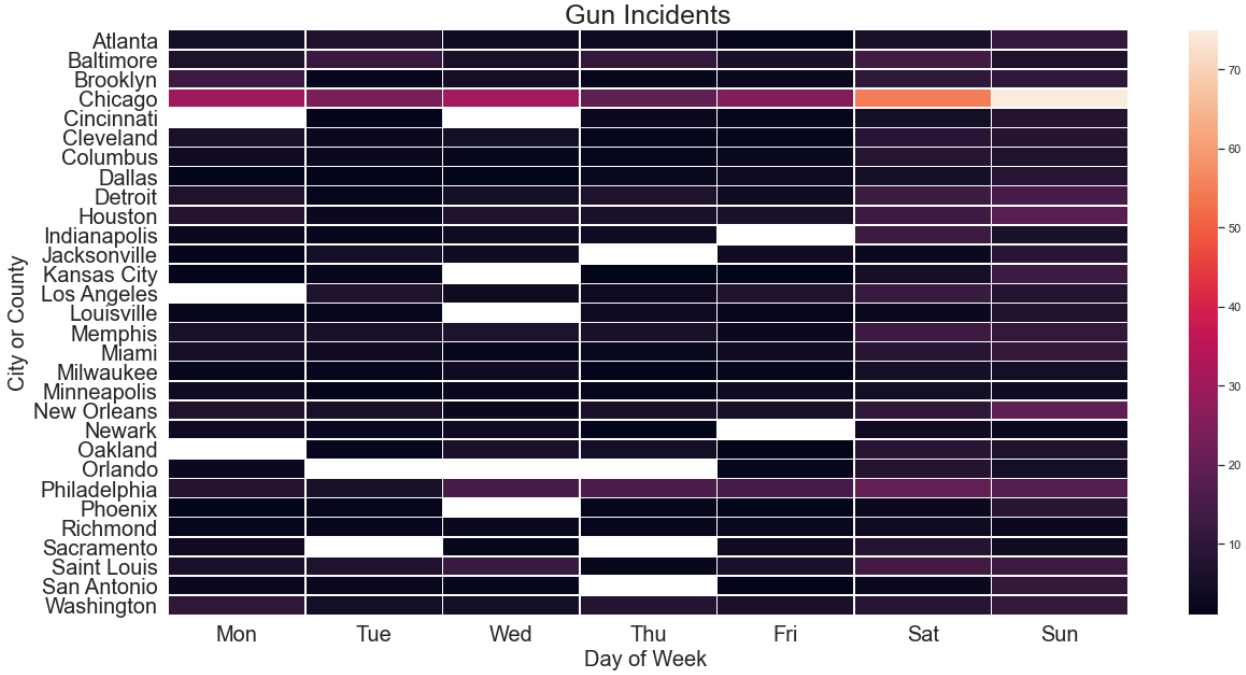


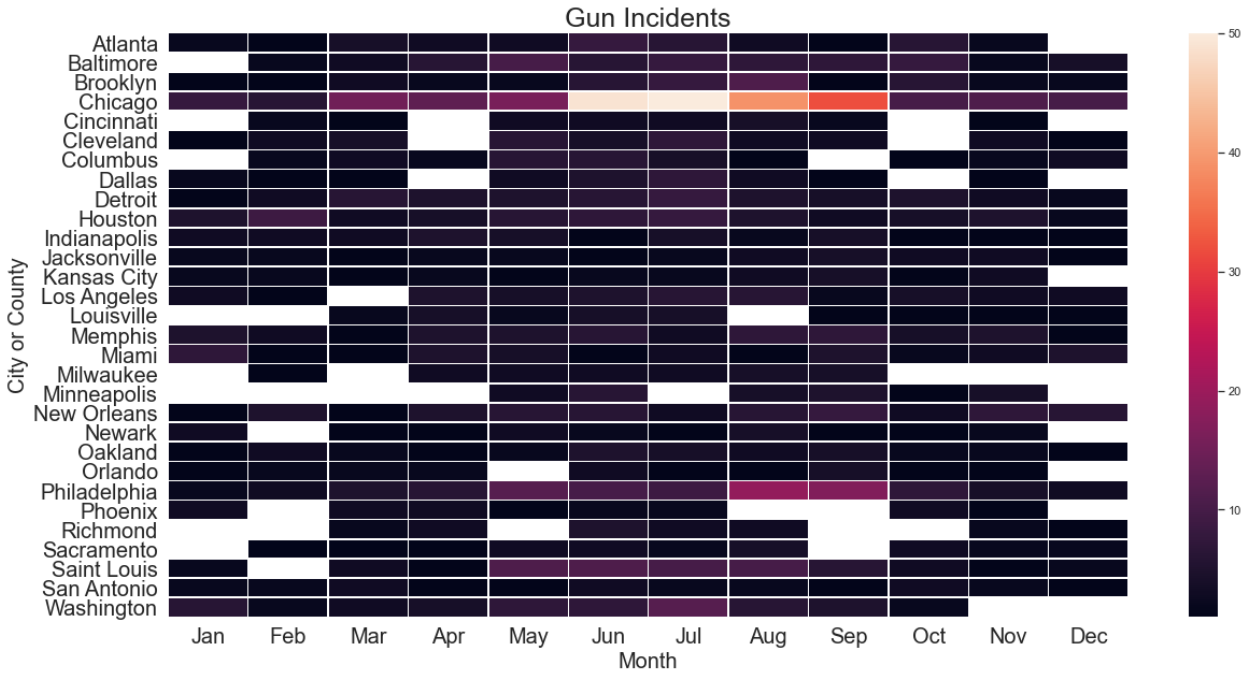
City / State Time Period Outlier Heat maps:

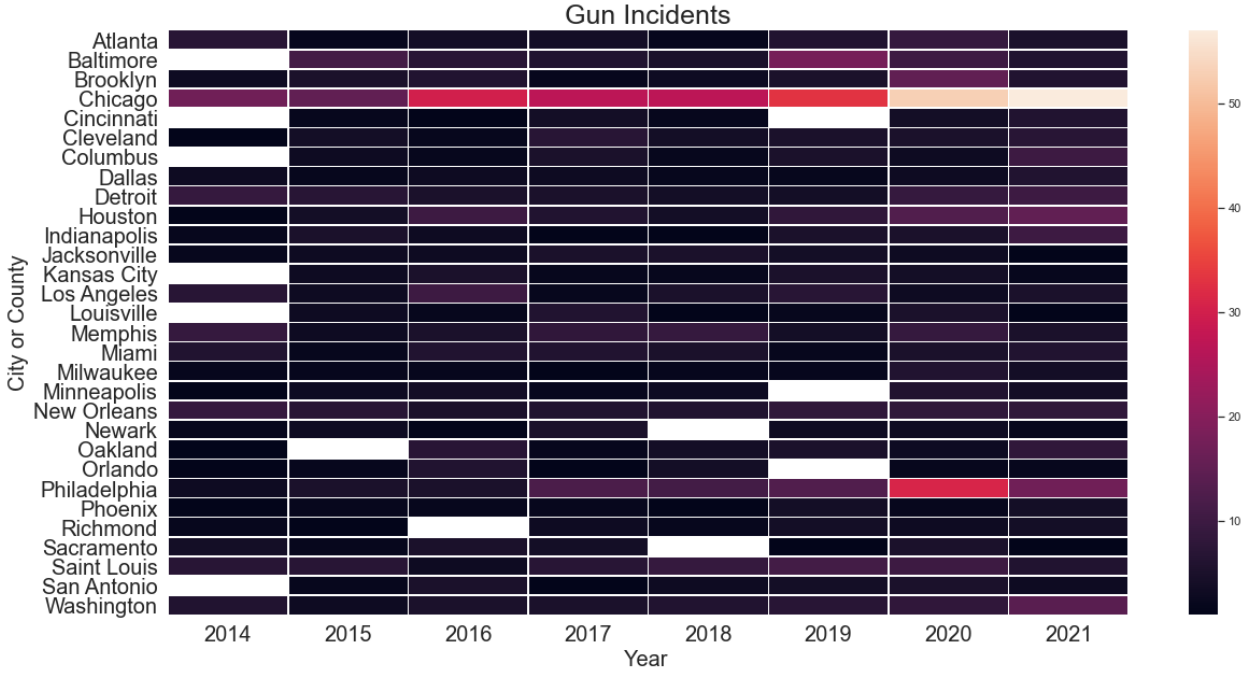



Gun Incidents per 100k people
States
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
Florida
Georgia
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
0.5
0.4
0.3
0.2
0.1


Gun Incidents per 100k people
States
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
Florida
Georgia
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
2014
2015
2016
2017
2018
2019
2020
2021
Year
0.6
0.5
0.4
0.3
0.2
0.1


Gun Incidents
City or County
Atlanta
Baltimore
Brooklyn
Chicago
Cincinnati
Cleveland
Columbus
Dallas
Detroit
Houston
Indianapolis
Jacksonville
Kansas City
Los Angeles
Louisville
Memphis
Miami
Milwaukee
Minneapolis
New Orleans
Newark
Oakland
Orlando
Philadelphia
Phoenix
Richmond
Sacramento
Saint Louis
San Antonio
Washington
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Day of Week
70
60
50
40
30
20
10


Gun Incidents
Atlanta
Baltimore
Brooklyn
Chicago
Cincinnati
Cleveland
Columbus
Dallas
Detroit
Houston
Indianapolis
Jacksonville
Kansas City
Los Angeles
Louisville
Memphis
Miami
Milwaukee
Minneapolis
New Orleans
Newark
Oakland
Orlando
Philadelphia
Phoenix
Richmond
Sacramento
Saint Louis
San Antonio
Washington
City or County
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
50
40
30
20
10


Gun Incidents
Atlanta
Baltimore
Brooklyn
Chicago
Cincinnati
Cleveland
Columbus
Dallas
Detroit
Houston
Indianapolis
Jacksonville
Kansas City
Los Angeles
Louisville
Memphis
Miami
Milwaukee
Minneapolis
New Orleans
Newark
Oakland
Orlando
Philadelphia
Phoenix
Richmond
Sacramento
Saint Louis
San Antonio
Washington
City or County
2014
2015
2016
2017
2018
2019
2020
2021
Year
50
40
30
20
10