

Gun Violence in the US

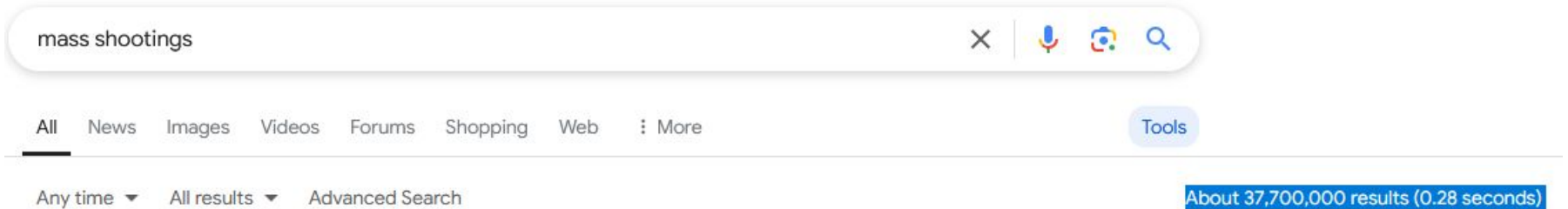
Taking a look at trends from 2014 - 2023

Emily Huang

Motivation for Study

Mass shootings are known to be a phenomenon mainly in the United States. In recent years, it seemed as though there was a rising number of incidents in the news.

Alarmed by this observation, I wanted to explore whether if that were really true.



Topics to explore

In this study, I will be analyzing the Gun Violence Archive dataset to answer the following questions:

1. Are mass shootings really more common in recent years?
2. If so, are there any patterns in the data in regard to the date and/or location of the incident?
3. What underlying factors may have caused these incidents?

About the Dataset

- The original data on mass shootings is from the [Gun Violence Archive](#), and the compiled dataset is from [figshare](#).
- This analysis will be using the compiled dataset which contains additional columns of the separated year, month, and date of the incident.
- The Gun Violence Archive **defines mass shootings** as a minimum of four victims shot, either injured or killed, not including any shooter who may also have been killed or injured in the incident.

Data columns (total 18 columns):

#	Column	Non-Null	Count	Dtype
0	Incident_ID	4683	non-null	int64
1	Incident_Date	4683	non-null	object
2	Incident_Time	3387	non-null	object
3	State_Name	4683	non-null	object
4	City_or_County	4683	non-null	object
5	Business_or_Location_Name	1628	non-null	object
6	Address	4668	non-null	object
7	Latitude	4683	non-null	float64
8	Longitude	4683	non-null	float64
9	Victims_Killed	4683	non-null	int64
10	Victims_Injured	4683	non-null	int64
11	Suspects_Killed	4683	non-null	int64
12	Suspects_Injured	4683	non-null	int64
13	Suspects_Arrested	4683	non-null	int64
14	Incident_Characteristics	4683	non-null	object
15	Year	4683	non-null	int64
16	Month	4683	non-null	int64
17	Day	4683	non-null	int64

dtypes: float64(2), int64(9), object(7)

Summary Statistics

State_Name		City_or_County		Business_or_Location_Name	
Illinois	468	Chicago	340	7-Eleven	7
California	420	Philadelphia	151	Waffle House	7
Texas	335	Baltimore	91	Walmart	6
Florida	268	Houston	81	McDonald's	4
Pennsylvania	225	Washington	76	West Garfield Park	3

Year		Month		Victims_Injured	Victims_Killed	Total_Victims
2021	690	7	596			
2023	656	6	549	count	4679.000000	4679.000000
2022	643	8	483	mean	4.138064	1.011755
2020	609	5	459	std	6.780262	1.857260
2019	413	9	424	min	0.000000	0.000000
2016	383	10	410	25%	3.000000	0.000000
2017	346	4	356	50%	4.000000	1.000000
2018	335	11	337	75%	5.000000	1.000000
2015	332	1	274	max	439.000000	60.000000
2014	272	12	268			
		3	268			
		2	255			

From the get go, there are already some interesting insights:

- The years during and after the pandemic saw a growth in incidents.
- The warmer months have more incidents.
- The top states and cities are those with a higher population and therefore more foot traffic.

Variances

Victims_Injured	45.971956
Victims_Killed	3.449413
Total_Victims	59.990589

Data Cleaning and Feature Engineering

- After exploring the data and analyzing the summary statistics, I performed the data cleaning and feature engineering portion of the process.
- For data cleaning, I performed the following steps:
 - Because the mass shootings are defined by the victims, the **columns for number of suspects injured or killed will be removed** from the dataset.
 - Mass shootings are defined to be minimum four victims (injured or killed), so **rows that had less than four total number of victims will also be removed** from the dataset to better focus the scope.
 - **Only 4 rows were dropped after this**, leaving the dataset with 4,679 incidents.
- For data engineering, I performed the following steps:
 - Because of the **varied answers in the Address, Business or Location Name, and Incident Characteristics columns**, I **removed them** from the main dataset.
 - Because I was **keeping City and State**, I thought it would be interesting to also **create a “US Region” column** to see how each state played a part in their region’s statistics.

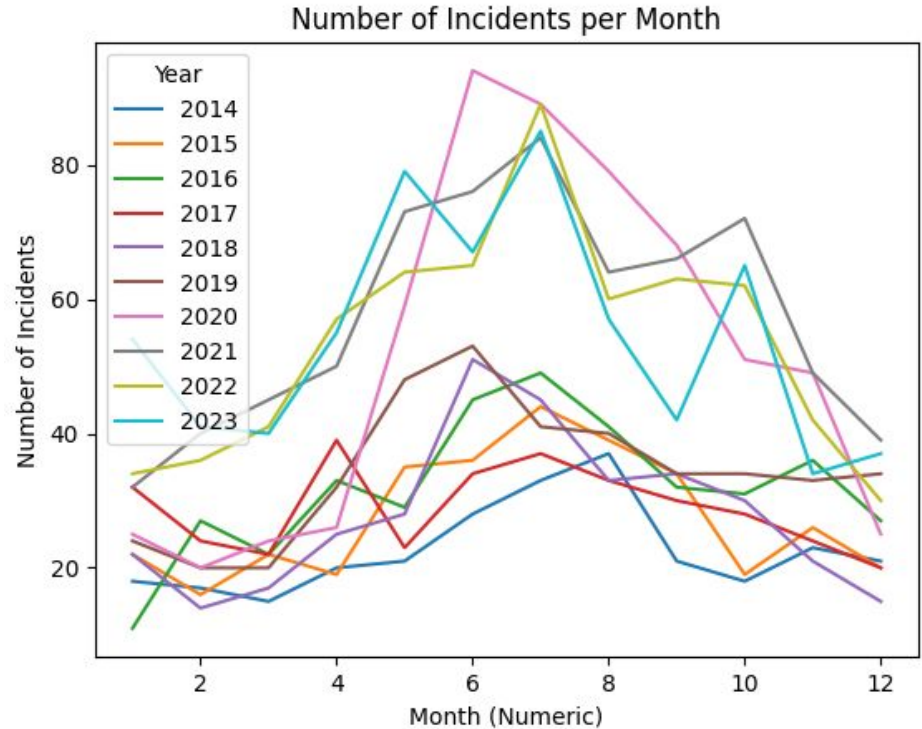
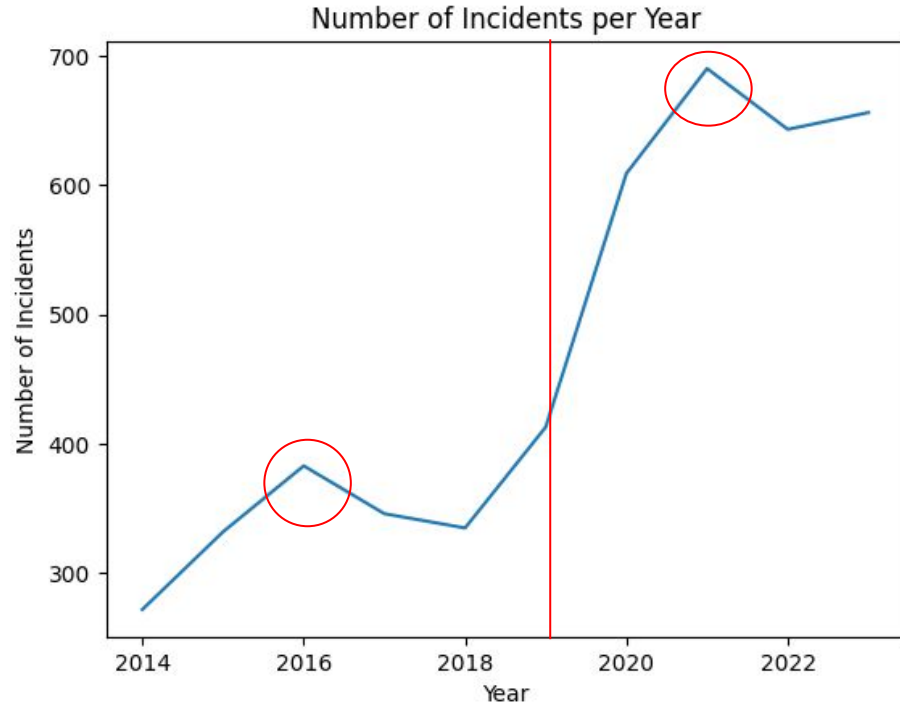
Exploring Trends with Time

Notes about Time

The incident time recorded for each incident had inconsistencies in the time recorded (i.e. missing values). The day of the month of the incident date was perceived as not as significant compared to the Month and Year of the incident.

Because of this, the following slides will focus only on the Month and Year of incidents.

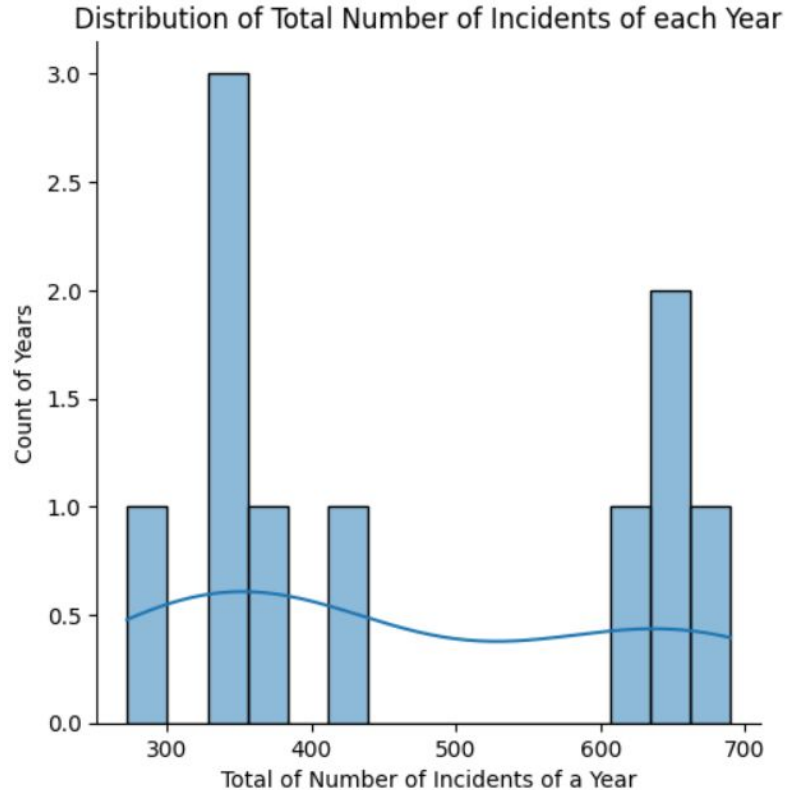
Number of Incidents by the Year



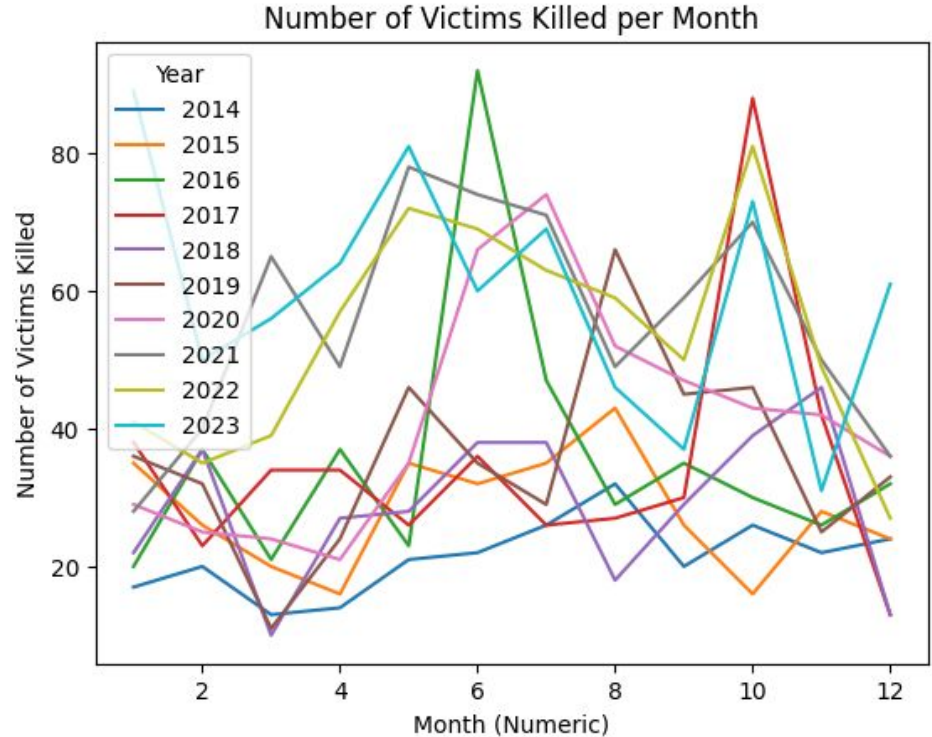
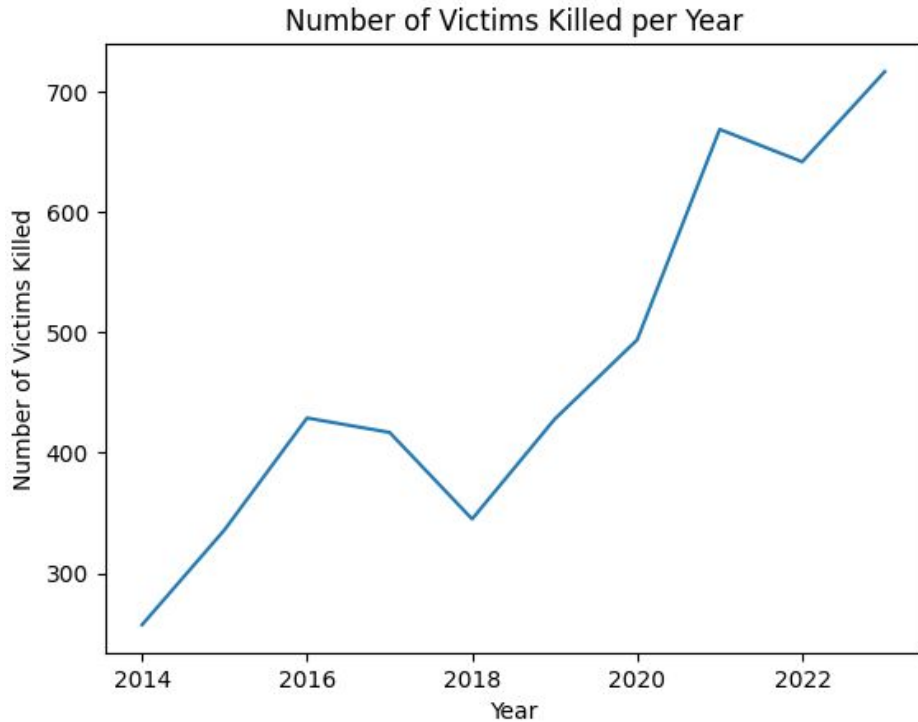
Distribution - Number of Incidents by the Year

This chart tells us there's a bimodal distribution in the number of incidents each year, meaning there are two high points in the data.

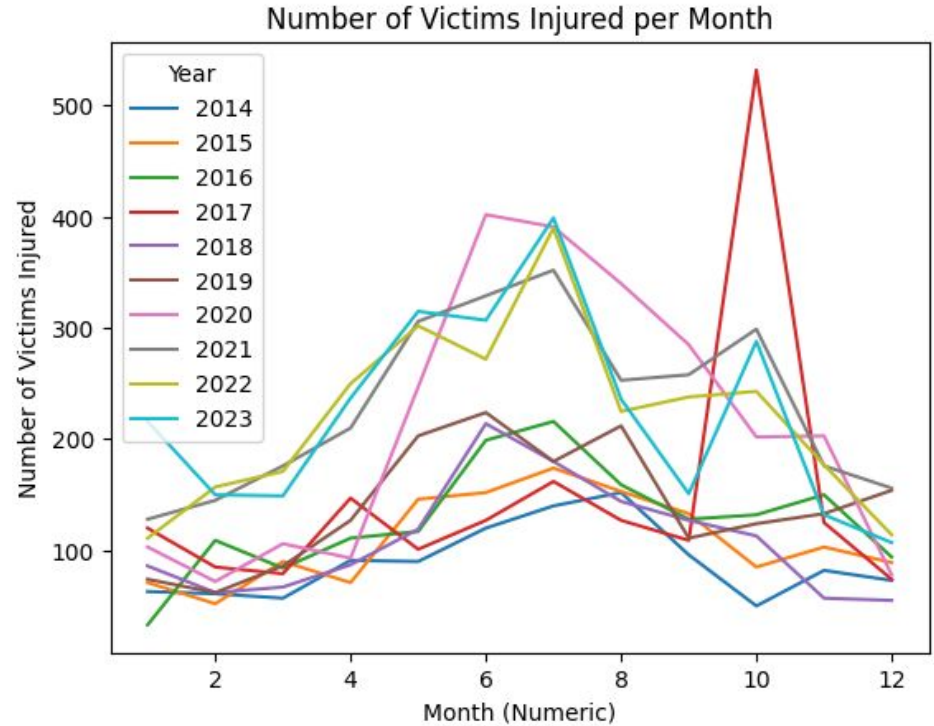
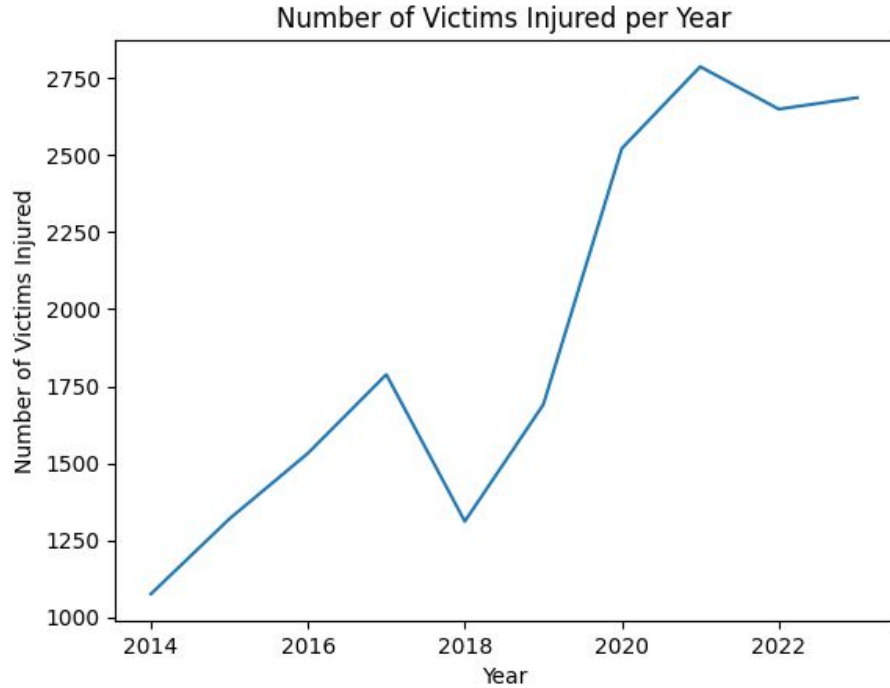
This matches with the previous slides where there were two distinct points before and after 2019.



Number of Victims Killed by the Year

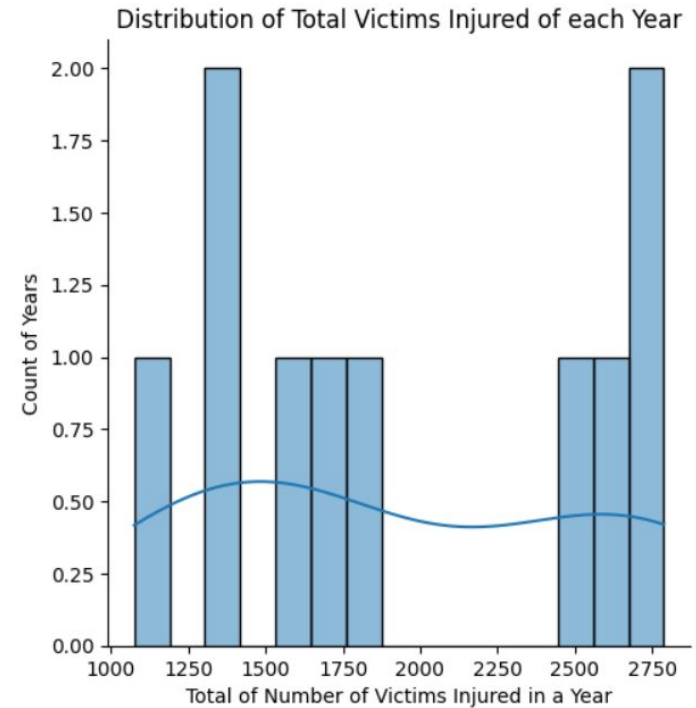
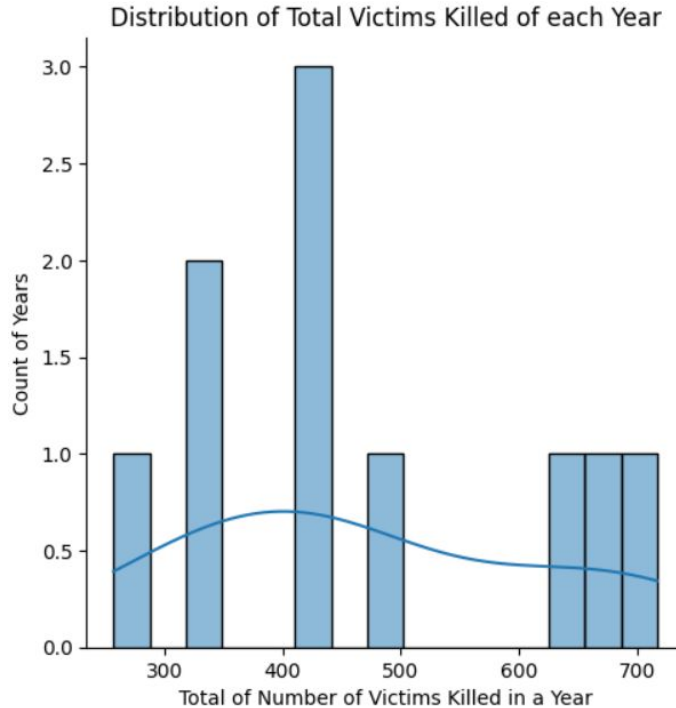


Number of Victims Injured by the Year



Distribution of Number of Victims Killed and Injured by the Year

The charts follow the same trend where the charts show a bimodal distribution.



Insights on Incident Trends based on Time

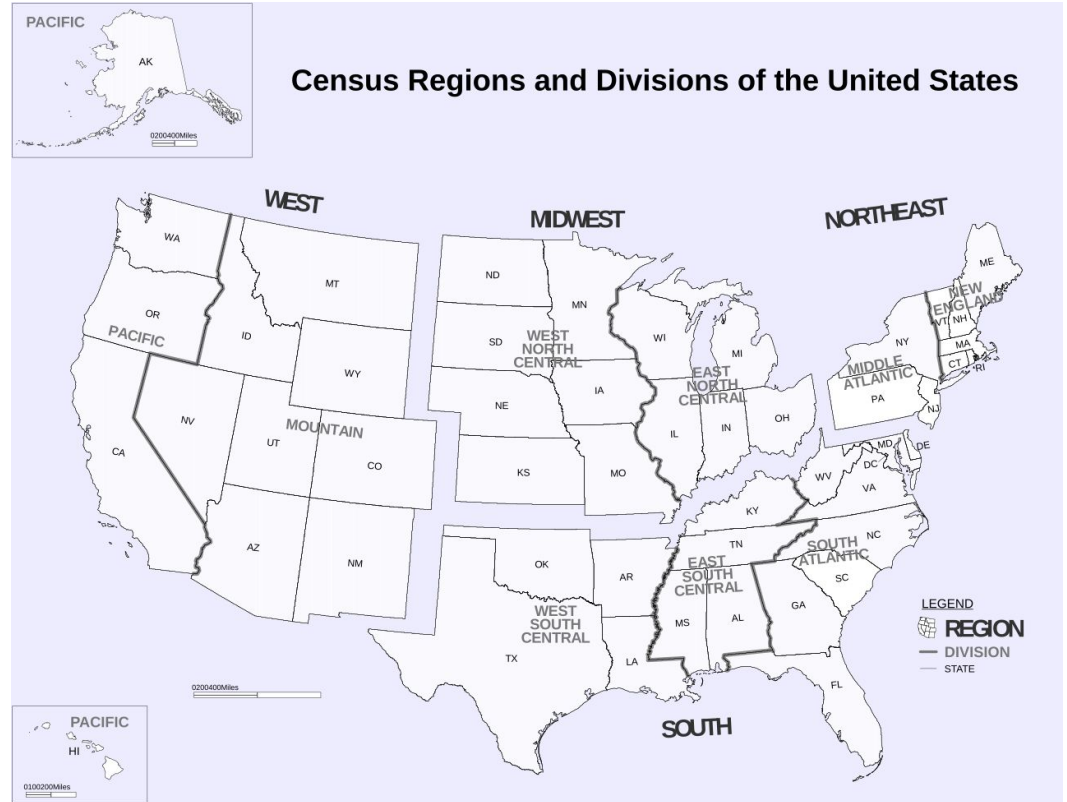
- Confirming what was seen in the summary statistics, there is indeed an increase in the number of incidents, most notably after 2019.
 - Factors like the political climate, unemployment, and racial inequality may have played a part in this rise of incidents.
- Also seen in the summary statistics, summer months like June, July, and August tend to have a higher number of incidents.
 - The summer months tend to be when families are on vacation, meaning there are more people outside exposed to other people.
 - Because of the hot weather during these months in the US, people are also incentivized to go outside and expose themselves to other people.
 - Both previous statements may be especially true after a long period of social isolation caused by COVID.

Exploring Trends with Location

Notes about Location

As mentioned in the “Data Cleaning and Feature Engineering” slide, a new column was created based on the US Census Bureau.

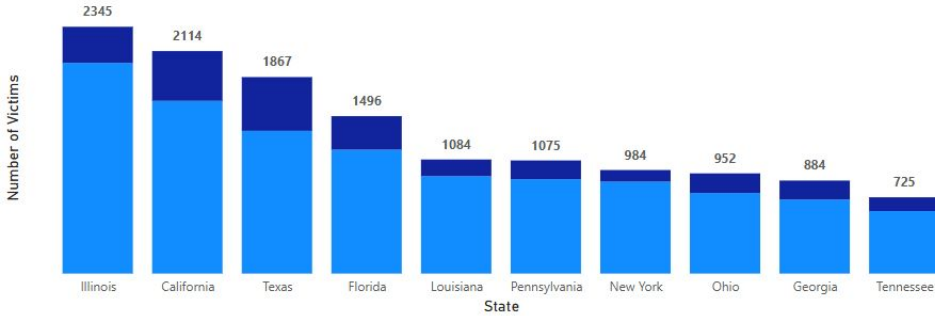
This section will focus on the new “US Region” column as well as the city and state column.



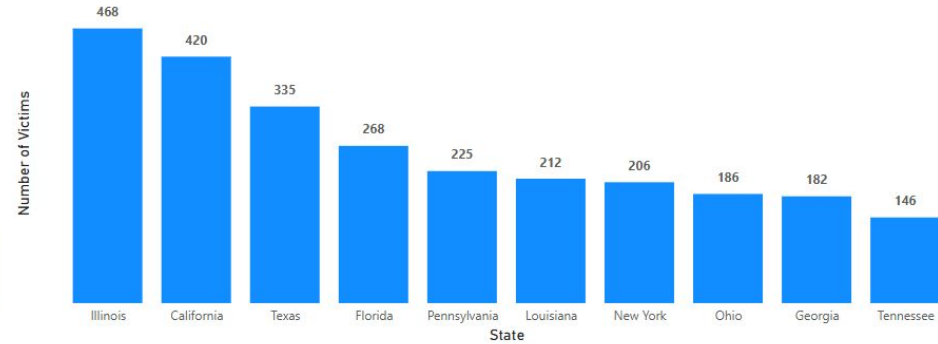
Top 10 States - # of Incidents, Total # of Victims

Total Injured and Total Killed by State

● Total Injured ● Total Killed

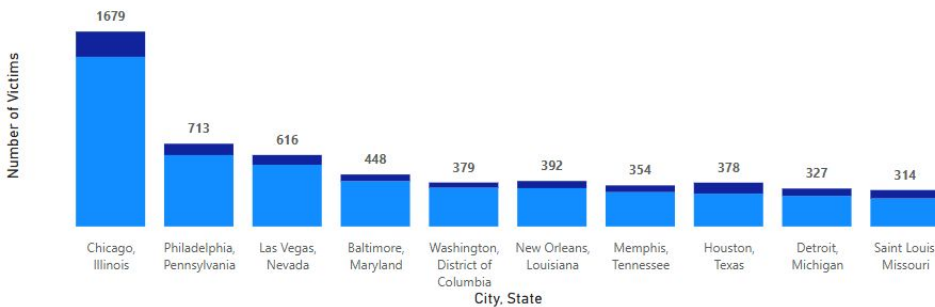


of Incidents by State

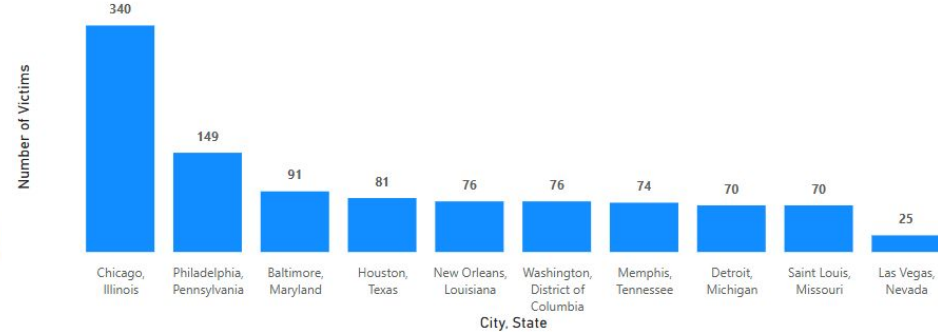


Total Injured and Total Killed by City, State

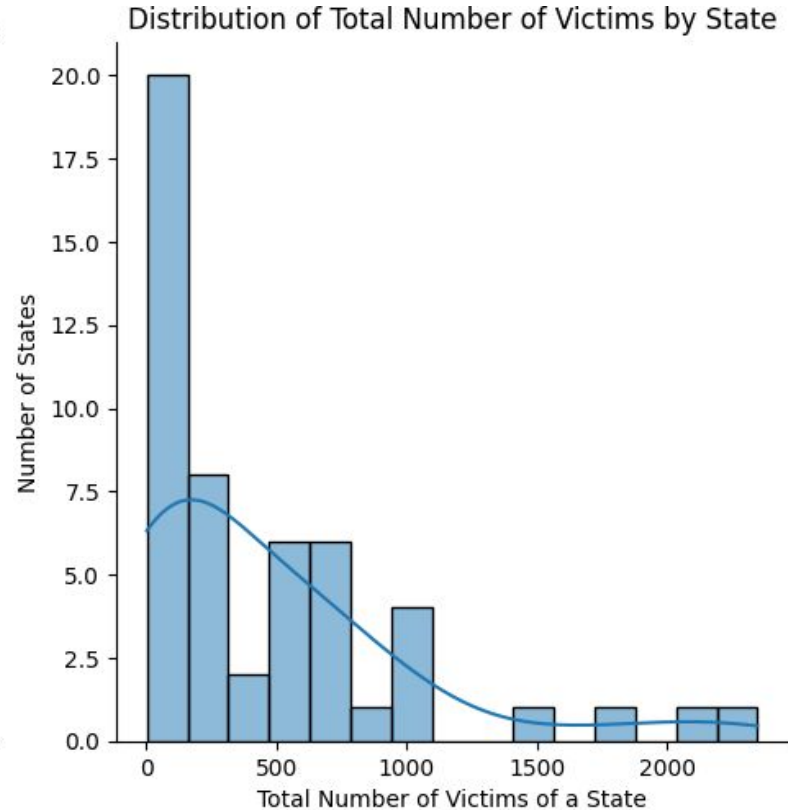
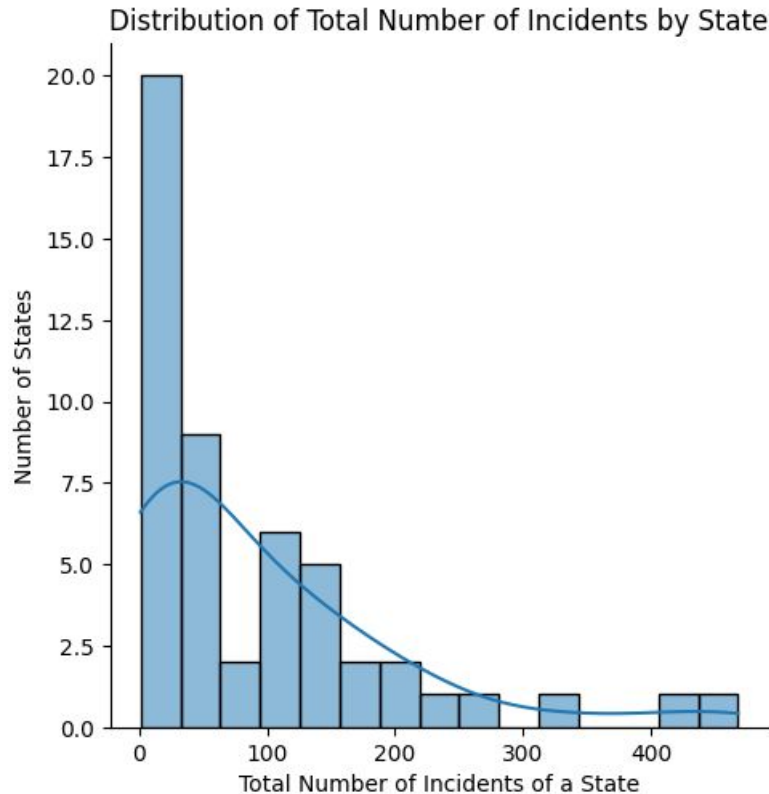
● Total Injured ● Total Killed



of Incidents by City, State

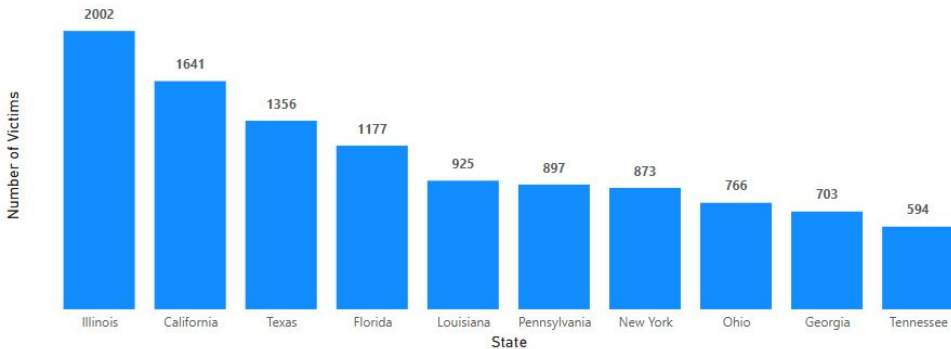


Distribution by State - # of Incidents, Total # of Victims

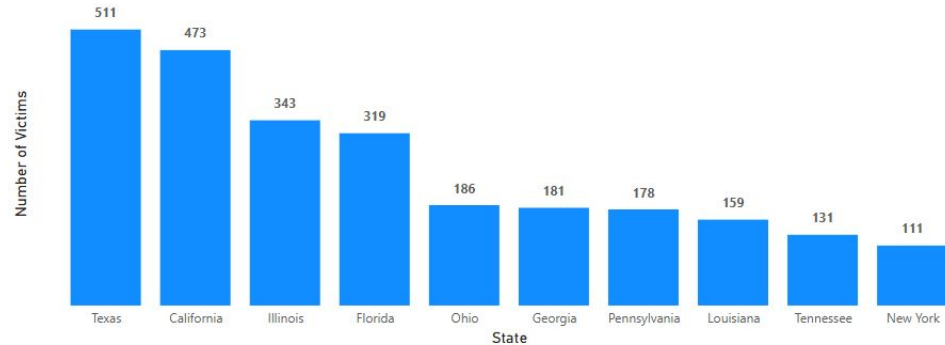


Top 10 States - # of Victims Injured and Killed

of Victims Injured by State



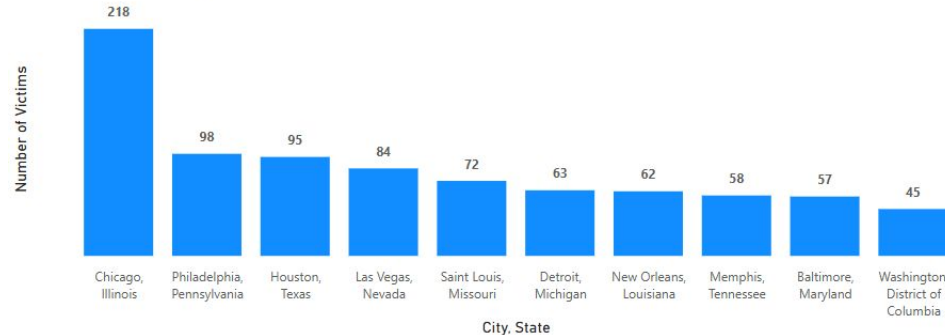
of Victims Killed by State



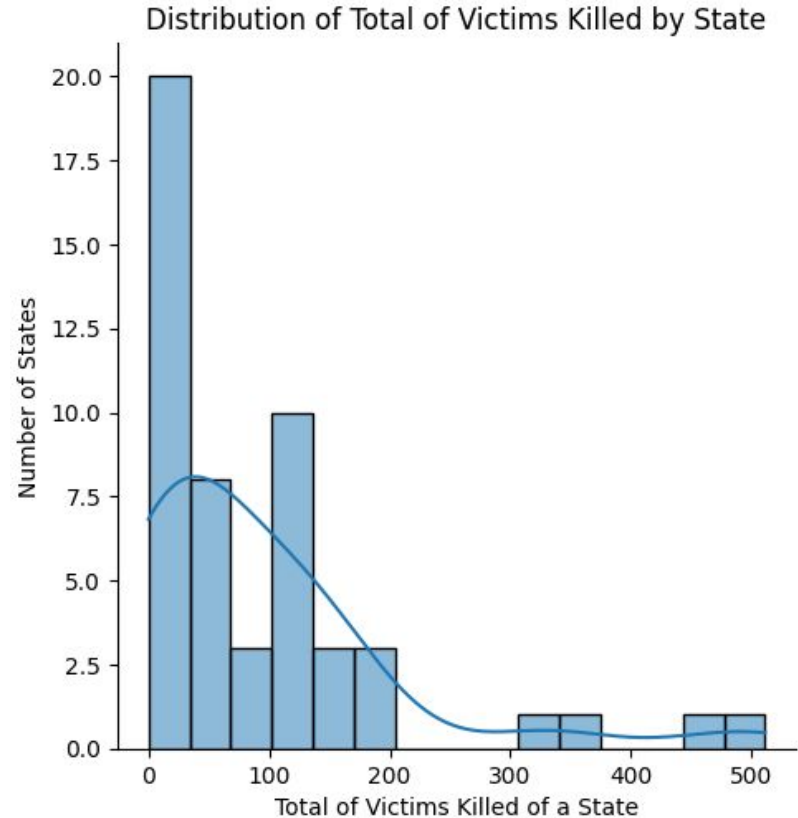
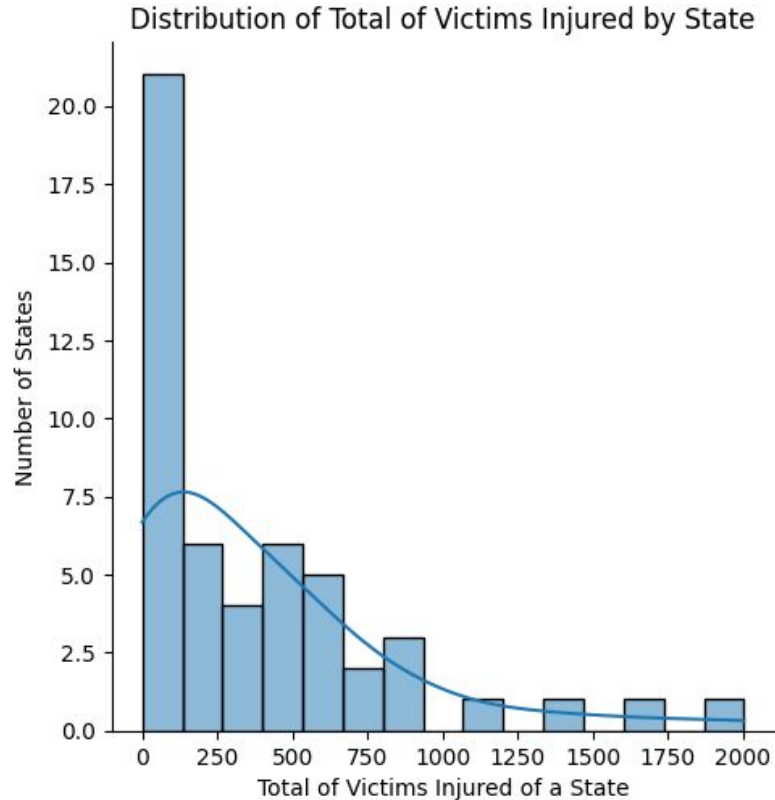
of Victims Injured by City, State



of Victims Killed by City, State



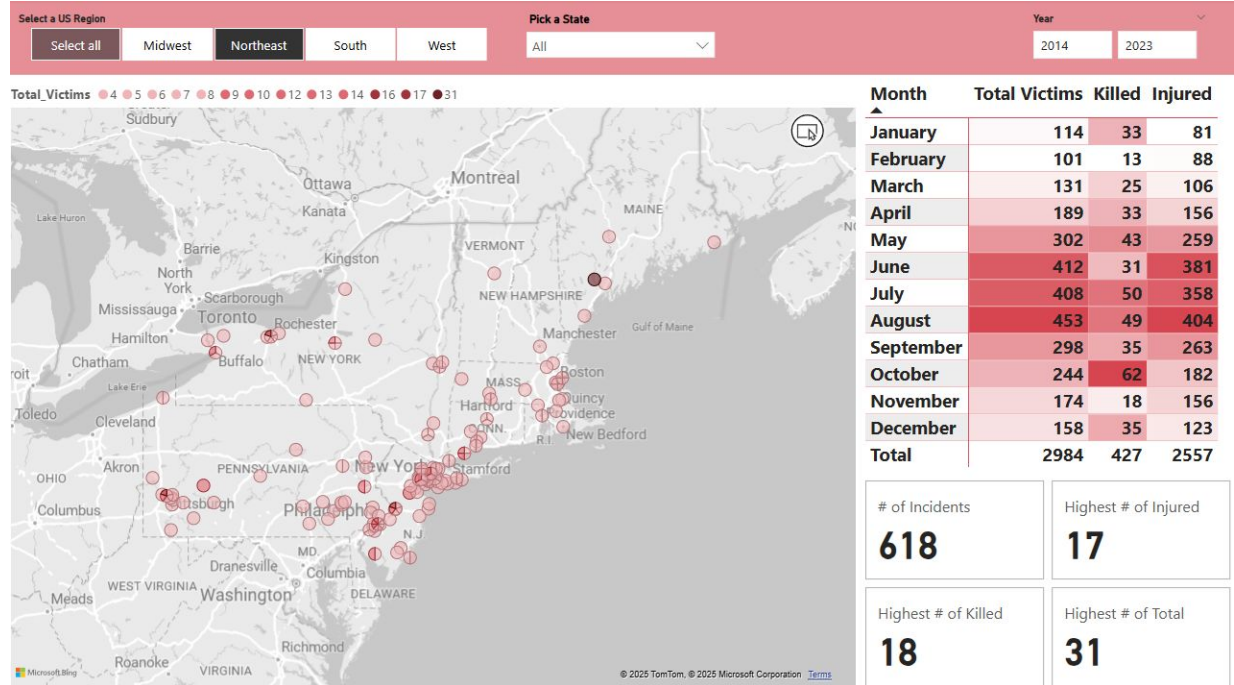
Distribution by State - # of Victims Injured and Killed



Incidents from 2014 - 2023 in the Northeast

The Northeast accounts for:

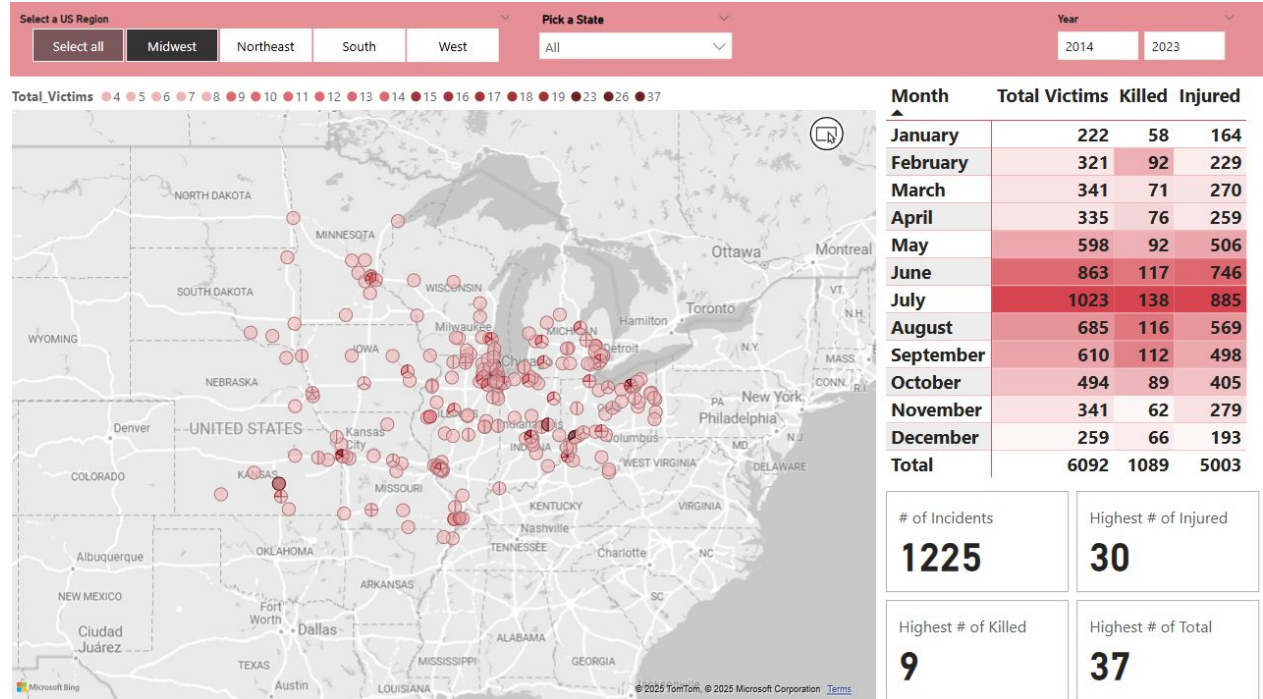
- 13.21% of all unique reported incidents
- 12.38% of all victims involved in any incident
- 13.21% of all injured victims involved in any reported incident
- 9.02% of all killed victims involved in any reported incident



Incidents from 2014 - 2023 in the Midwest

The Midwest accounts for:

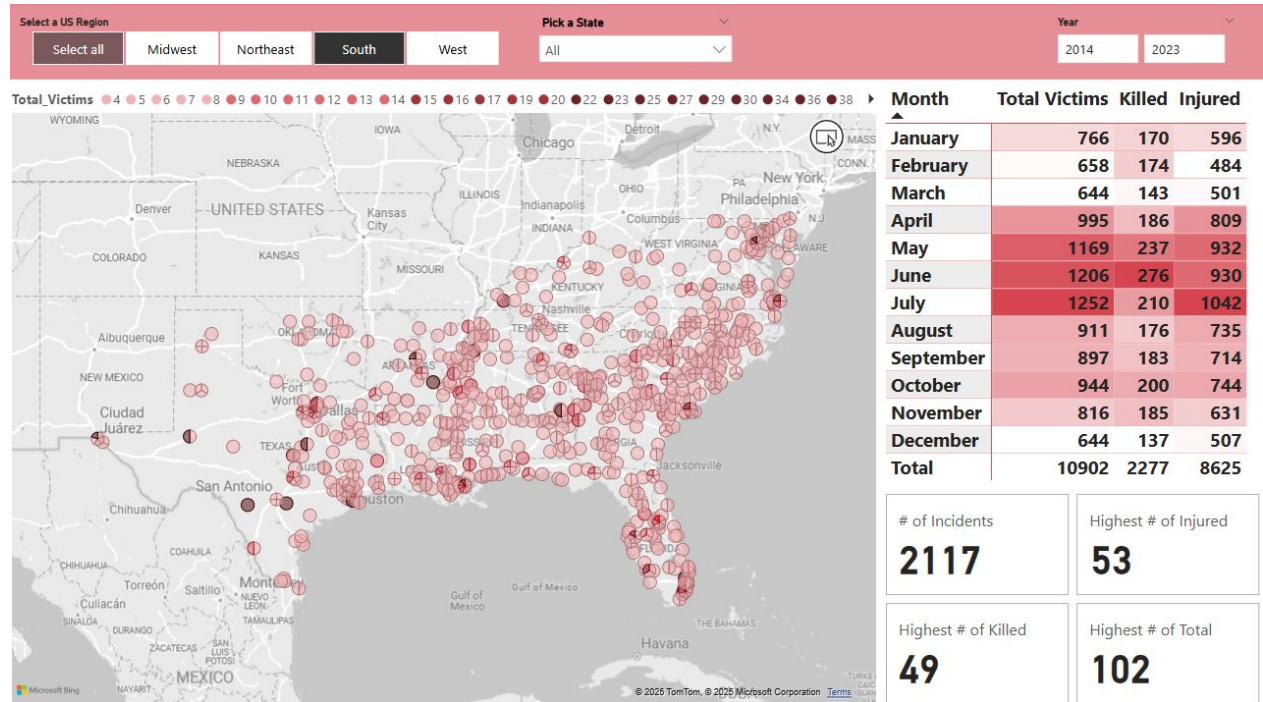
- 26.18% of all unique reported incidents
- 25.28% of all victims involved in any incident
- 25.84% of all injured victims involved in any reported incident
- 23.00% of all killed victims involved in any reported incident



Incidents from 2014 - 2023 in the South

The South accounts for:

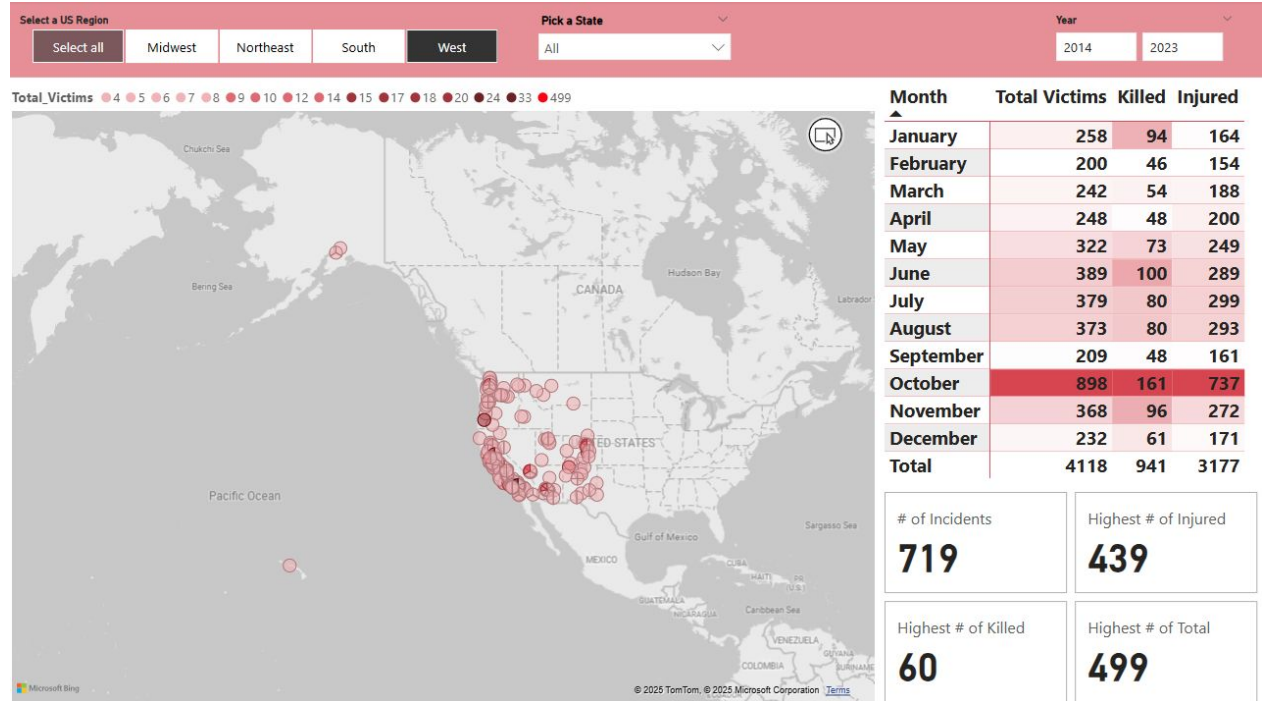
- 45.24% of all unique reported incidents
- 45.24% of all victims involved in any incident
- 44.54% of all injured victims involved in any reported incident
- 48.09% of all killed victims involved in any reported incident



Incidents from 2014 - 2023 in the West

The West accounts for:

- 15.37% of all unique reported incidents
- 17.09% of all victims involved in any incident
- 16.41% of all injured victims involved in any reported incident
- 19.88% of all killed victims involved in any reported incident



Top 30 States - Incident Rate per 100,000 Residents

- Using the US Census Bureau dataset, the number of incidents can be put into perspective with the state population.
- Incident rate was calculated by getting the average number of incidents from 2014 - 2023, dividing it by the average population from 2014 - 2023, and multiplying by 100000.
- Unsurprisingly, states from the South and Midwest regions appear in the top 30.

	NAME	REGION	NUM_INCIDENT	AVG_POPULATION	INCIDENT_RATE
1	District of Columbia	South	7	681554	1.02706461997142
2	Louisiana	South	21	4640666	0.452521254492351
3	Illinois	Midwest	46	12735055	0.36120770581674
4	Mississippi	South	10	2969916	0.336709859807483
5	Maryland	South	14	6074448	0.230473616697353
6	Missouri	Midwest	14	6127893	0.228463519190038
7	Alabama	South	11	4948763	0.222277769212225
8	South Carolina	South	11	5091017	0.216066848725903
9	Tennessee	South	14	6815327	0.205419343782037
10	Pennsylvania	Northeast	22	12868883	0.170955008294038
11	Georgia	South	18	10555165	0.170532625496617
12	Arkansas	South	5	3012116	0.165996263092125
13	Ohio	Midwest	18	11698883	0.153860842953981
14	Indiana	Midwest	10	6721953	0.148766288606897
15	Michigan	Midwest	14	9993546	0.140090414353424
16	North Carolina	South	14	10380962	0.134862260356988
17	Colorado	West	7	5671731	0.123419111378872
18	Florida	South	26	21263235	0.122276784318096
19	Virginia	South	10	8527263	0.117270922686447
20	Texas	South	33	28759741	0.114743731523869
21	Kentucky	South	5	4471307	0.111824126592068
22	New Jersey	Northeast	10	9035026	0.110680367715599
23	California	West	42	39166851	0.107233537871094
24	Minnesota	Midwest	6	5614872	0.106859069984142
25	Wisconsin	Midwest	6	5828329	0.102945458295165
26	Delaware	South	1	976681	0.102387575881992
27	New York	Northeast	20	19671627	0.10166927219594
28	Nevada	West	3	3031360	0.0989654808402...
29	New Mexico	West	2	2101512	0.0951695731454...
30	Washington	West	6	7513856	0.0798524752137...

Top 30 Cities - Incident Rate per 1,000 Residents

- Like the previous slide, the incident rate was also calculated for each city.
- Incident rate for cities was calculated similarly with first averaging the number of incidents from 2014 - 2023, dividing by the average population from 2014 - 2023, and multiplying by 1000.
- The top 30 cities primarily were from the South.

	CITY_NAME	STATE_NAME	REGION	NUM_INCIDENT	AVG_POPULATION	INCIDENT_RATE
1	Macon	Georgia	South	1	12781	0.0782411391909866
2	Chicago	Illinois	Midwest	34	1352956	0.0251301594434704
3	Pine Bluff	Arkansas	South	1	42074	0.0237676474782526
4	New Orleans	Louisiana	South	7	383280	0.0182634105614694
5	Wilmington	Delaware	South	1	70935	0.0140974131246916
6	Gary	Indiana	Midwest	1	73059	0.0136875675823649
7	Paterson	New Jersey	Northeast	2	150477	0.0132910677379267
8	Baltimore	Maryland	South	9	716133	0.0125674979368358
9	Trenton	New Jersey	Northeast	1	86097	0.0116148065553968
10	Flint	Michigan	Midwest	1	90267	0.011078245649019
11	Memphis	Tennessee	South	7	642189	0.0109002178486396
12	Detroit	Michigan	Midwest	7	659342	0.0106166450794883
13	Shreveport	Louisiana	South	2	188658	0.01060119369441
14	Cleveland	Ohio	Midwest	4	377984	0.0105824585167626
15	Washington	District of Columbia	South	7	681554	0.0102706461997142
16	Atlanta	Georgia	South	5	490747	0.0101885492932203
17	Birmingham	Alabama	South	2	205909	0.00971302857087354
18	Minneapolis	Minnesota	Midwest	4	421424	0.00949162838376552
19	Rochester	New York	Northeast	1	107673	0.0092873793801603
20	Baton Rouge	Louisiana	South	2	223970	0.00892976737955976
21	Miami	Florida	South	4	448703	0.00891458269724071
22	Philadelphia	Pennsylvania	Northeast	14	1576851	0.00887845459082691
23	North Charleston	South Carolina	South	1	113755	0.00879082238143378
24	Richmond	Virginia	South	1	117572	0.00850542646208281
25	Saginaw	Michigan	Midwest	1	118880	0.00841184387617766
26	Buffalo	New York	Northeast	2	264759	0.00755403971158676
27	Columbia	South Carolina	South	1	135269	0.00739267681434771
28	Oakland	California	West	3	428739	0.00699726406974873
29	Houston	Texas	South	8	1161320	0.0068887128440051
30	Newark	New Jersey	Northeast	2	291177	0.006868674380188

Insights on Incident Trends based on Location

- Places with a higher population tend to have a higher number of incidents.
 - This is intuitive since **higher density regions tend to have a higher crime rate, meaning residents have a higher likelihood of being victims of a shooting.**
 - Though a higher density population implies a higher crime rate, **cities and states from the South had a higher incident rate on average** compared to cities and states from other regions.
- From these observations in the data, **there may be some pattern occurring where locations in the South have a higher rate of mass shootings compared to other locations.**

Understanding the Underlying Reason

Does political leaning impact gun violence?

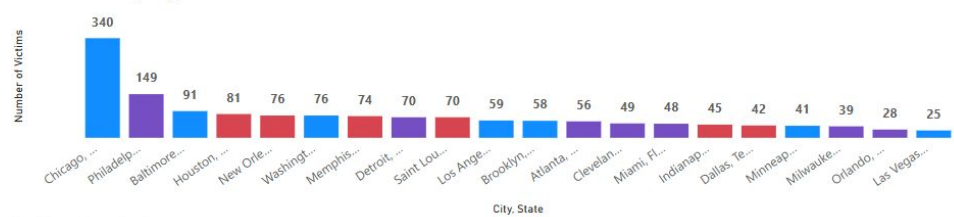
Looking at the same chart but now with the associated party color, there is an **equal amount of red and blue states** when it comes to the number of incidents in a state and in a specific city.

However, **when we take a look at incident rates for each state, we can see Red states are by far more prevalent than the other states.**

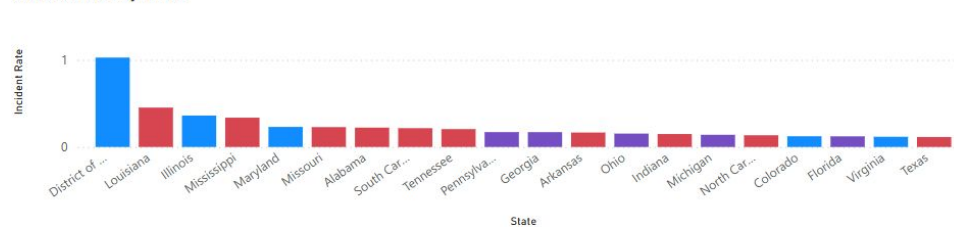
of Incidents by State



of Incidents by City, State



Incident Rate by State



Why do Red states have a higher incident rate?

From this article, Red states have historically been pro-Second amendment, making their lawmakers more likely to make guns more accessible.

However, we can see from the chart in the article that looser guns laws are associated with higher gun deaths.

In Red States, 'Gun Reform' Means Making It Easier To Buy And Carry Guns

By Monica Potts

Graphics by Humera Lodhi

MAY 16, 2023, AT 2:36 PM

States with loosest gun laws have most gun deaths

Gun death rate per 100,000 people and Giffords Law Center's gun law grade, by state



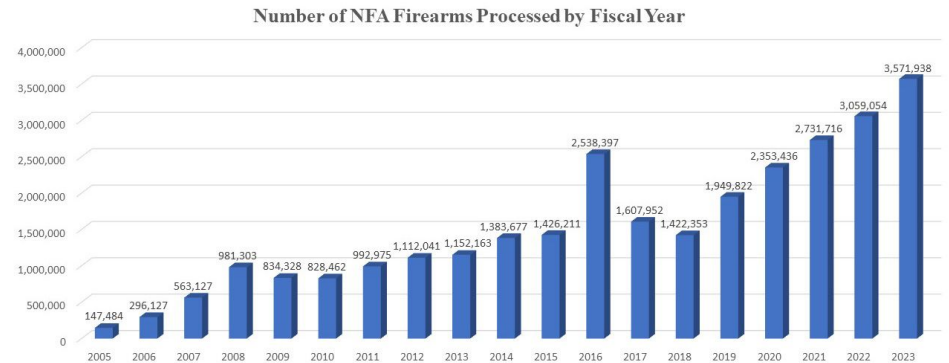
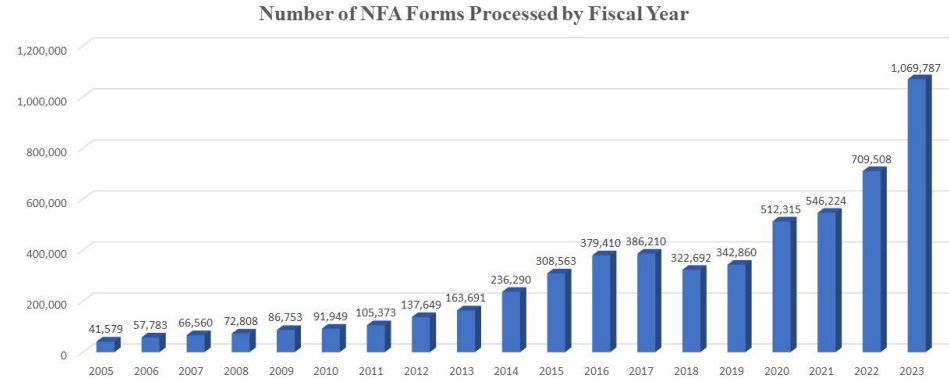
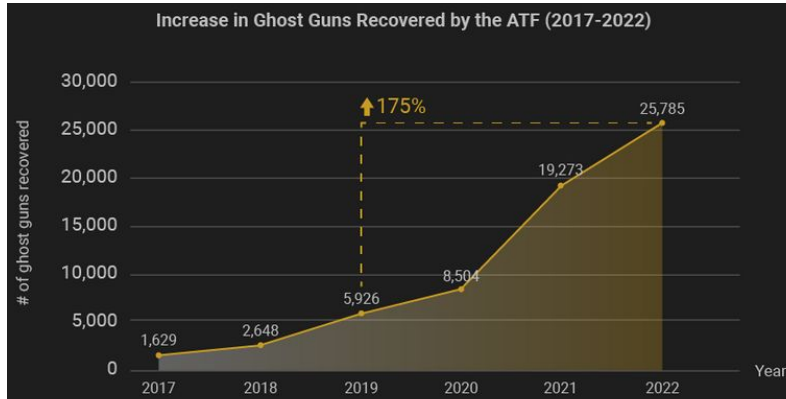
FiveThirtyEight

SOURCE: GIFFORDS LAW CENTER, CENTERS FOR DISEASE CONTROL AND PREVENTION

More guns in the US

Though Red states have more access to guns, Americans as a whole are acquiring more guns.

Charts from the Bureau of Alcohol, Tobacco, Firearms and Explosives show a similar pattern to the “Number of Incidents each Year” chart where there is a sudden spike after 2019.



Takeaways

From this study, we were able to observe that mass shootings were in fact more frequent as of recent years. Incidents were likely to occur during the summer months when more traveling was done, but incidents were notably more frequent in Red states where gun laws are more lenient compared to Purple and Blue states.

However, as a whole, the United States is acquiring more guns, and more access to guns means more opportunities for mass shootings to occur regardless of location.

Sources

Articles:

- <https://ammo.com/articles/how-many-guns-in-the-us>
- <https://www.atf.gov/resource-center/data-statistics>
- <https://fivethirtyeight.com/features/republican-states-expanding-gun-rights-mass-shootings/>

Assisting Data Sources:

- Harvard Dataverse: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/42MVDX>
- US Census Bureau: <https://www.census.gov/data/datasets.html>