Mass Shootings in the US

Table of Contents

Introduction:

A mass shooting is considered as a terrible act committed by individuals or terrorist groups in public or non-public places. In recent times, terrorist groups have used the tactics of mass shootings to fulfill their objectives. Individuals who commit these kinds of activities fall into several categories like students, co-workers, family killers or random strangers. A mass shooting is generally defined as an incident involving four or more victims of guns related violence. When we hear the word mass shootings, United States is the first thing which comes to the mind. The country which has the highest number of mass shootings in the history is the United States. Between the year 1966-2018, the United States has witnessed more than 328 mass shootings that resulted in more than 1477 deaths and 2025 injured. The worst mass shooting in the history of United States is the Las Vegas Shooting which occurred on October 1st, 2017 that resulted in 59 deaths and more than 527 people injured. The number of people injured in this incident is more than the number of people injured in all attacks of 2015 and 2016 combined. On an average, 8 shootings occurred every year in the last 50 years that took 35 lives and 47 injured per year.

Problem Description:

This objective of the report is to find out why the United States has the highest number of mass shootings in the history of all countries and how to put a check to the criminal offenses. Is it because of the higher accessibility and ownership of guns or mental illness of the individuals who commit these activities or failure of the government background checks of citizens and unregulated laws related to guns? This is achieved by exploring and visualizing US Mass shootings data from 1966 to 2018 using Tableau and R. Using the data exploration, the following questions should be answered.

Motivation:

These incidents happen very rarely around the world and whenever I read it on the internet or watch it in the news, my heart gets shattered into pieces. I just wonder why anyone would commit such activities and was curious to know what their objectives were. The Las Vegas Shooting which happened in the year 2017 was a historic tragedy. With a broken heart, I would want to explore the data and find out the main reasons behind these activities and how the government must control such terrible crimes. This can be achieved by finding answers to the following questions.

Questions:

1) How many people were killed and injured per year and month?
2) Which cities and states in the USA are more prone to these attacks?
3) How does the gun ownership in each state affect these shootings?
4) Can we find out any deadly months and dates?
5) Is there any correlation between mental illness of the shooter and cause of the attack?
6) Who are the primary targets in these attacks?
7) Can we find any pattern that can help in the prediction of such attacks in future?
8) Is there any correlation between shooter's gender, race, and age?

Let's dive and explore the data of the US Mass shootings from 1966-2018 to find answers to the above-mentioned questions. Before we start exploring, we need to wrangle the data to get into a formatted mode. Let's discuss the structure of the datasets and the attributes and how can it be cleaned in the next section.

Data Checking and Wrangling:

In total, 7 datasets are used to explore this data. Before exploration, let's clean the data to obtain an enriched analysis. The details of the datasets are mentioned below.

1) **US Mass Shootings(1966-2017):** The data of the US Mass Shootings is obtained from the Kaggle website and the link to the website is https://www.kaggle.com/zusmani/us-mass-shootings-last-50-years. This dataset is in CSV format and contains detailed information of 323 mass shootings occurred in the United States from 1966 to 2018 in a tabular format. The variables in the dataset are S.no, Title of the mass shooting, Location, Date, Summary, Incident area, Open/Close location, Target, Cause of the attack, Fatalities, Injured, Total Victims, Policeman Killed, Age of the shooter/shooters, Mental health condition, Race, Gender, Latitude, and Longitude of the location.
2) **Mother Jones' Investigation_ US Mass Shootings, 1982-2018:** The data is obtained from the website https://data.world/awram/us-mass-shootings. This dataset is xlsx format and contains detailed information of 97 mass shootings occurred in the United States from 1982 to 2018 in a tabular format. The variables in the dataset are Case, Location, Date, Year, Summary, Fatalities, Injured, Total Victims, Venue, Mental health issues and details, Types and details of weapons used by shooter, Race, Gender, Latitude, and Longitude of the location.
3) **US Mass Shootings:** The data is obtained from the website https://data.world/awram/us-mass-shootings. This dataset is in xlsx format and contains detailed information of 71 mass shootings from 1982-2015. The variables in the dataset are similar to the above dataset.
4) **50_us_states_all_data:** The data is obtained from the website https://scottontechnology.com/list-of-50-us-states-in-excel/. This dataset is in CSV format and contains information on 50 state names and state codes.
5) **US_cities_lat_long:** The data is obtained from the website https://simplemaps.com/data/us-cities. This dataset is in CSV format and contains information related to the US cities latitudes and longitudes.

6) **<u>Number of registered weapons in the U.S. in 2017, by state:</u>** The data is obtained from the website https://www.statista.com/statistics/215655/number-of-registered-weapons-in-the-us-by-state/. This data consists of a state-wise number of registered weapons in the United States, which is depicted in a bar graph on the website.
7) **<u>Percentage of households in the United States owning one or more firearms from 1972 to 2017:</u>** The dataset is obtained from the website https://www.statista.com/statistics/249740/percentage-of-households-in-the-united-states-owning-a-firearm/. This data consists of a year-wise percentage of citizens who has guns in the United States, which is depicted in a bar graph in the website.

The first three data sets consist of same variables and hence, they are merged into a single dataset based on the 'Location', and 'Date' using VLOOKUP function in EXCEL. The location name is split into City and State columns, then converted into lowercase. The 'City', 'State' and 'Date' are concatenated and using this as reference column, VLOOKUP function is used to combine three datasets. After merging, the dataset consists of 328 incidents from 1966 to 2018. This merged dataset should be checked for any input errors, missing, and null values. In this dataset, there are few input errors in the 'longitude' section. The longitudes of US cities are negative and few of the values in the data are positive, which are rectified. Other than that, there are no more input errors in the data.

The 'Longitude' column consist of missing values. The dataset 5 which consists of cities latitude and longitude values is helpful in filling these missing values. The 'State' column in this dataset consists of both state names and state codes. We must separate the 'State' column into 'State name' and 'State code'. The 'State' column is split into two columns using 'Text to Columns' function with comma as a separator in EXCEL. Now we have missing values in the split columns. The dataset 4 is used to fill the missing values in both columns of 'State name' and 'State code' using VLOOKUP function in EXCEL. Hence the 'location' column is split into 'city', 'statename', and 'statecode'. The merged dataset is shown in the *Figure1*. We can see that the data still has few missing values and the 'gender' column has irregularities. Let's use python for further cleaning and formatting of the data.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
S.No	Title	Date	City	StateCode	State	Incident Area	Target	Cause	Summary	Fatalities	Injured	Total victims	Age	Mental Health	Mental health - details	Race	Gender	Latitude	Longitude	Type of weapons	Weapon details
1	Texas church mass shooting	11/5/2017	Sutherland Springs	TX	Texas	Church	random	unknown	Devin Patr	26	20	46	26	No	Kelley had a history of d	White	M	29.2733	-98.0567	semiautomatic rif	Ruger AR-556; Ke
2	Walmart shooting in suburban Denver	11/1/2017	Thornton	CO	Colorado	Wal-Mart	random	unknown	Scott Aller	3	0	3	47	No		White	M	39.868	-104.972	semiautomatic handgun	
3	Edgewood businees park shooting	10/18/2017	Edgewood	MD	Maryland	Remodeling Store	coworkers	unknown	Radee Lab	3	3	6	37	No		Black	M	39.4187	-76.2944	handgun	.38-caliber; make
4	Las Vegas Strip mass shooting	10/1/2017	Las Vegas	NV	Nevada	Las Vegas Strip Cor	random	unknown	Stephen C	59	527	585	64	Unclear	Perpetrator's history ur	White	M	36.1813	-115.134	23 firearms, most	AR-15-style and A
5	San Francisco UPS shooting	6/14/2017	San Francisco	CA	California	UPS facility	coworkers		Jimmy Lan	3	2	5	38	Yes	Lam had a history of do	Asian	M	37.7749	-122.419	two handguns	MAC-1-style "assa
6	Pennsylvania supermarket shooting	6/7/2017	Tunkhannock	PA	Pennsylvania	Weis grocery	coworkers	terrorism	Randy Sta	3	0	3	24	Unclear		White	M	41.5387	-75.9466	shotguns	
7	Florida awning manufacturer shooting	6/5/2017	Orlando	FL	Florida	manufacturer Fiam	coworkers	unemploye	John Robe	5	0	5	45	Unclear		Unkno	M	28.5383	-81.3792	semiautomatic handgun	
8	Rural Ohio nursing home shooting	5/12/2017	Kirkersville	OH	Ohio	a nursing home	coworkers		Thomas H	3	0	3	43	Yes	Hartless had a violent c	White	M	39.9595	-82.5957	handgun, shotgun	
9	Fresno downtown shooting	4/18/2017	Fresno	CA	California	a street in downto	random	racism	Kori Ali Mu	3	0	3	39	Unclear		Black	M	36.7378	-119.787	handgun	.357 revolver
10	Fort Lauderdale airport shooting	1/6/2017	Fort Lauderdale	FL	Florida	baggage claim are	random	terrorism	Esteban Sa	5	6	11	26	Yes	Among other signs, San	Latino	M	26.1224	-80.1373	handgun	Walther 9mm sen
11	Cascade Mall shooting	9/23/2016	Burlington	WA	Washington	cosmetics section c	women	terrorism	Arcan Ceti	5	0	5	20	Yes	According to the Cetin's	Unkno	M	48.4676	-122.33	Rifle	
12	Baton Rouge police shooting	7/17/2016	Baton Rouge	LA	Louisiana		police		Gavin Long	3	3	6		Yes	Unclear	Black	M	30.4515	-91.1871	Two semiautomat	IWI Tavor SAR 5.5
13	Dallas police shooting	7/7/2016	Dallas	TX	Texas	at protest	police	racism	Micah Xav	5	11	16	25	Unclear	Unclear	Black	M	32.7767	-96.797	Semiautomatic rif	Izhmash-Saiga 5.4
14	Orlando nightclub massacre	6/12/2016	Orlando	FL	Florida	at nightclub	random		Omar Mat	49	53	102	29	Unclear	Unclear	Other	M	28.5383	-81.3792	Semiautomatic rif	Sig Sauer MCX rifl
15	Ferguson, MO Drive by	4/29/2016	Ferguson	MO	Missouri		random		A group of	0	4	4	20	Unknown		Unkno	Unknow	38.7442	-90.3054		
16	Forestville, Maryland Drive-by	4/26/2016	Forestville	MD	Maryland	in street			Shooter sh	1	4	5		Unknown		Unkno	Unknow	38.8451	-76.875		
17	Halifax County, VA	4/24/2016	Halifax	VA	Virginia	crown	random		Male shoo	0	6	6		Unknown		Black A	Male	36.766	-78.9283		
18	Tire-Slashing revenge escalation	4/21/2016	Baltimore	MD	Maryland	block party	random	frustration	Shooter w	0	4	4		Unknown		Black A	Male	39.2904	-76.6122		
19	Chicago Rap video Shootout	4/19/2016	Chicago	IL	Illinois	in a park	random		Group of y	1	4	5		Unknown		Unkno	Unknow	41.8781	-87.6298		
20	Texas family murder-suicide	4/19/2016	Katy	TX	Texas	Home	Family	domestic d	Man killed	4	0	4		Unknown		White	Male	29.7858	-95.8244		
21	Alabama highway random shooting	4/19/2016	Brooksville	AL	Alabama	along a highway	random		Shooter fir	1	4	4		Yes		White	Male	34.162	-86.4755		
22	Long Beach Street murder	4/18/2016	Signal Hill	CA	California	at street corner	random	terrorism	Group of r	0	3	4		Unknown		Unkno	Unknow	33.7701	-118.194		
23	Albuquerque, NM House party shooting	4/9/2016	Albuquerque	NM	New Mexico	at party	uninvited g	anger	A Man sho	0	4	4		Unknown		White	Male	35.0853	-106.606		
24	Memphis, TN gas station shooting	4/9/2016	Memphis	TN	Tennessee	at gas station	random	anger	Four peop	0	4	4		Unknown		Unkno	Unknow	35.1495	-90.049		
25	Chicago Birthday Party Bus Shooting	4/7/2016	Chicago	IL	Illinois	south shore	birthday party bus		Birthday p	0	5	5		Unknown		Unkno	Unknow	41.8781	-87.6298		
26	Albuquerque, NM Family restaurant shoo	4/1/2016	Albuquerque	NM	New Mexico	restaurant	Family		The shoote	3	1	3		Unknown		Asian A	Male	35.0853	-106.606		
27	Richmond, Virginia	3/31/2016	Richmond	VA	Virginia	bus station	Trooper	frustration	A man fata	2	2	3		Unknown		Black A	Male	37.5407	-77.436		
28	Louisburg, North Carolina	3/26/2016	Louisburg	NC	North Carolina			terrorism	Three peo	3	0	3		Unknown		Black A	Male	36.099	-78.3011		
29	Lawrenceburg, Tennessee	3/25/2016	Lawrenceburg	TN	Tennessee	in home	Family	domestic d	The man w	2	2	3		Unknown		Unkno	Male	35.2423	-87.3347		
30	Greenhill, AL Family murder-suicide	3/25/2016	Normal	AL	Alabama	home in rural Alaba	Family	domestic d	Husband r	2	2	3		Yes		White	Male	34.7593	-86.6025		
31	Sherman, Texas Family Murder-Suicide	3/21/2016	Sherman	TX	Texas		Family	domestic d	A man sho	4	0	3		Unknown		Unkno	Male	33.6357	-96.6089		
32	Louisville, KY Family Murder-Suicide	3/20/2016	Louisville	KY	Kentucky	Home	Family	domestic d	A former s	4	0	3		Yes		White	Male	38.2527	-85.7585		
33	Plantation, Florida	3/19/2016	Plantation	FL	Florida	Home	party guests		A Spring B	1	4	5		Unknown		Unkno	Unknow	26.1276	-80.2331		
34	Wetumpka Drive-by	3/19/2016	Wetumpka	AL	Alabama	drive-by in Wetum	random	terrorism	Shooter fir	2	2	4		Unknown		Black A	Male	32.5437	-86.2119		

Figure 1 US Mass Shootings merged data before formatting

Now the dataset is loaded into python environment for further cleaning and formatting. The 'pandas' library in python is used to load the data into a data frame. The following python code is used to fill the missing values in the other columns.

The merged dataset is loaded into a data frame using pandas library in Python environment. The 'target' column has few nulls and they are imputed with the mode of the values in the column. The column 'Age' has missing values and to maintain data integrity, they are replaced with zeroes since age is a number. The other columns such as 'Cause', 'Mental Health condition', and 'Mental health details', 'Weapon details', and 'Types of weapons' has missing values which are filled with 'Unknown'.

```
import pandas as pd #importing pandas library

#reading csv file into dataframe
us_shootings = pd.read_csv('US_Mass_Shootings.csv',encoding='latin-1')

#mode imputation for Target column
lis = list(us_shootings['Target'])
mode_target = max(set(lis),key=lis.count)
us_shootings['Target'] = us_shootings['Target'].fillna(mode_target)

#replacing missing values with unknown in Cause column
us_shootings['Cause'] = us_shootings['Cause'].fillna('unknown')

#replacing missing values with zeros in Age column
us_shootings['Age'] = us_shootings['Age'].fillna(0)

#formatting gender data
us_shootings['Gender'] = us_shootings['Gender'].replace('M/F','Male/Female')
us_shootings['Gender'] = us_shootings['Gender'].replace('M','Male')

#replacing missing values with unknown in the following columns
us_shootings['Mental health - details'] = us_shootings['Mental health - details'].fillna('Unknown Details')
us_shootings['Type of weapons'] = us_shootings['Type of weapons'].fillna('Unknown')
us_shootings['Weapon details'] = us_shootings['Weapon details'].fillna('Unknown Details')
us_shootings['Incident Area'] = us_shootings['Incident Area'].fillna('Unknown')
us_shootings.set_index('S.No',inplace=True) #setting Serial No as index
#writing to csv format
us_shootings.to_csv('us_shootings_fmtd.csv',sep='\t',encoding='utf-32')
```

Figure 2 Python code for formatting data file

The 'gender' column has values M, Male, male etc. which are same. Hence the values in the column are formatted accordingly. Now the formatted data is exported to CSV file 'us_shootings_fmtd'. Now, the data looks cleaned and formatted, which is depicted in *Figure 3*.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
S.No	Title	Date	City	StateCode	State	Incident Area	Target	Cause	Summary	Fatalities	Injured	Total victims	Age	Mental Health	Mental health - details	Race	Gender	Latitude	Longitude	Type of weapons	Weapon details
1	Texas church mass shooting	11/5/2017	Sutherland Springs	TX	Texas	Church	random	unknown	Devin Patr	26	20	46	26	Yes	Kelley had a history of d	White	Male	29.273	-98.0567	semiautomatic rif	Ruger AR-556; Kelley also
2	Walmart shooting in suburban Denver	11/1/2017	Thornton	CO	Colorado	Wal-Mart	random	unknown	Scott Aller	3	0	3	47	No	Unknown Details	White	Male	39.868	-104.972	semiautomatic ha	Unknown Details
3	Edgewood businees park shooting	10/18/2017	Edgewood	MD	Maryland	Remodeling St	coworkers	unknown	Radee Lab	3	3	6	37	No	Unknown Details	Black	Male	39.419	-76.2944	handgun	.38-caliber; make unclear
4	Las Vegas Strip mass shooting	10/1/2017	Las Vegas	NV	Nevada	Las Vegas Strip	random	unknown	Stephen C	59	527	585	64	Unclear	Perpetrator's history un	White	Male	36.181	-115.134	23 firearms, most	AR-15-style and AK-47-sty
5	San Francisco UPS shooting	6/14/2017	San Francisco	CA	California	UPS facility	coworkers	unknown	Jimmy Lan	3	2	5	38	Yes	Lam had a history of do	Asian	Male	37.775	-122.419	two handguns	MAC-1-style assault pisto
6	Pennsylvania supermarket shooting	6/7/2017	Tunkhannock	PA	Pennsylvania	Weis grocery	coworkers	terrorism	Randy Sta	3	0	3	24	Unclear	Unknown Details	White	Male	41.539	-75.9466	shotguns	Unknown Details
7	Florida awning manufacturer shooting	6/5/2017	Orlando	FL	Florida	manufacturer	coworkers	unemployement	John Robe	5	0	5	45	Unclear	Unknown Details	Unkno	Male	28.538	-81.3792	semiautomatic ha	Unknown Details
8	Rural Ohio nursing home shooting	5/12/2017	Kirkersville	OH	Ohio	a nursing hom	coworkers	unknown	Thomas H	3	0	3	43	Yes	Hartless had a violent c	White	Male	39.96	-82.5957	handgun, shotgur	Unknown Details
9	Fresno downtown shooting	4/18/2017	Fresno	CA	California	a street in dov	random	racism	Kori Ali Mu	3	0	3	39	Unclear	Unknown Details	Black	Male	36.738	-119.787	handgun	.357 revolver
10	Fort Lauderdale airport shooting	1/6/2017	Fort Lauderdale	FL	Florida	baggage claim	random	terrorism	Esteban Sa	5	6	11	26	Yes	Among other signs, San	Latino	Male	26.122	-80.1373		Walther 9mm semi-autor
11	Cascade Mall shooting	9/23/2016	Burlington	WA	Washington	cosmetics sect	women	terrorism	Arcan Ceti	5	0	5	20	Yes	According to the Cetin's	Unkno	Male	48.468	-122.33	Rifle	Unknown Details
12	Baton Rouge police shooting	7/17/2016	Baton Rouge	LA	Louisiana	Unknown	police	unknown	Gavin Long	3	3	6	0	Yes	Unclear	Black	Male	30.452	-91.1871	Two semiautomat	IWI Tavor SAR 5.56 calibe
13	Dallas police shooting	7/7/2016	Dallas	TX	Texas	at protest	police	racism	Micah Xav	5	11	16	25	Unclear	Unclear	Black	Male	32.777	-96.797	Semiautomatic rif	Izhmash-Saiga 5.45mm (
14	Orlando nightclub massacre	6/12/2016	Orlando	FL	Florida	at nightclub	random	unknown	Omar Mat	49	53	102	29	Unclear	Unclear	Other	Male	28.538	-81.3792	Semiautomatic rif	Sig Sauer MCX rifle, Glock
15	Ferguson, MO Drive by	4/29/2016	Ferguson	MO	Missouri	Unknown	random	unknown	A group of	0	4	4	20	Unknown	Unknown Details	Unkno	Unknow	38.744	-90.3054	Unknown	Unknown Details
16	Forestville, Maryland Drive-by	4/26/2016	Forestville	MD	Maryland	in street	random	unknown	Shooter sh	1	4	5	0	Unknown	Unknown Details	Unkno	Unknow	38.845	-76.875	Unknown	Unknown Details
17	Halifax County, VA	4/24/2016	Halifax	VA	Virginia	crown	random	unknown	Male shoo	0	6	6	0	Unknown	Unknown Details	Black	Male	36.766	-78.9283	Unknown	Unknown Details
18	Tire-Slashing revenge escalation	4/21/2016	Baltimore	MD	Maryland	block party	random	frustration	Shooter w	0	4	4	0	Unknown	Unknown Details	Black	Male	39.29	-76.6122	Unknown	Unknown Details
19	Chicago Rap video Shootout	4/19/2016	Chicago	IL	Illinois	in a park	random	unknown	Group of y	1	4	5	0	Unknown	Unknown Details	Unkno	Unknow	41.878	-87.6298	Unknown	Unknown Details
20	Texas family murder-suicide	4/19/2016	Katy	TX	Texas	Home	Family	domestic dispute	Man killed	4	0	4	0	Unknown	Unknown Details	White	Male	29.786	-95.8244	Unknown	Unknown Details
21	Alabama highway random shooting	4/19/2016	Brooksville	AL	Alabama	along a highw	random	unknown	Shooter fir	1	4	4	0	Yes	Unknown Details	White	Male	34.162	-86.4755	Unknown	Unknown Details
22	Long Beach Street murder	4/18/2016	Signal Hill	CA	California	at street corn	random	terrorism	Group of r	0	3	4	0	Unknown	Unknown Details	Unkno	Unknow	33.77	-118.194	Unknown	Unknown Details
23	Albuquerque, NM House party shooting	4/9/2016	Albuquerque	NM	New Mexico	at party	uninvited g	anger	A Man sho	0	4	4	0	Unknown	Unknown Details	White	Male	35.085	-106.606	Unknown	Unknown Details
24	Memphis, TN gas station shooting	4/9/2016	Memphis	TN	Tennessee	at gas station	random	anger	Four peop	0	4	4	0	Unknown	Unknown Details	Unkno	Unknow	35.15	-90.049	Unknown	Unknown Details
25	Chicago Birthday Party Bus Shooting	4/7/2016	Chicago	IL	Illinois	south shore	birthday pa	unknown	Birthday p	0	5	5	0	Unknown	Unknown Details	Unkno	Unknow	41.878	-87.6298	Unknown	Unknown Details
26	Albuquerque, NM Family restaurant shooti	4/1/2016	Albuquerque	NM	New Mexico	restaurant	Family	unknown	The shoote	3	1	3	0	Unknown	Unknown Details	Asian	Male	35.085	-106.606	Unknown	Unknown Details
27	Richmond, Virginia	3/31/2016	Richmond	VA	Virginia	bus station	Trooper	frustration	A man fat	2	2	3	0	Unknown	Unknown Details	Black	Male	37.541	-77.436	Unknown	Unknown Details
28	Louisburg, North Carolina	3/26/2016	Louisburg	NC	North Carolina	Unknown	random	terrorism	Three peo	3	0	3	0	Unknown	Unknown Details	Black	Male	36.099	-78.3011	Unknown	Unknown Details
29	Lawrenceburg, Tennessee	3/25/2016	Lawrenceburg	TN	Tennessee	in home	Family	domestic dispute	The man v	2	2	3	0	Unknown	Unknown Details	Unkno	Male	35.242	-87.3347	Unknown	Unknown Details
30	Greenhill, AL Family murder-suicide	3/25/2016	Normal	AL	Alabama	home in rural	Family	domestic dispute	Husband r	2	2	3	0	Yes	Unknown Details	White	Male	34.759	-86.6025	Unknown	Unknown Details
31	Sherman, Texas Family Murder-Suicide	3/21/2016	Sherman	TX	Texas	Unknown	Family	domestic dispute	A man sho	4	0	3	0	Unknown	Unknown Details	Unkno	Male	33.636	-96.6089	Unknown	Unknown Details
32	Louisville, KY Family Murder-Suicide	3/20/2016	Louisville	KY	Kentucky	Home	Family	domestic dispute	A former s	4	0	3	0	Yes	Unknown Details	White	Male	38.253	-85.7585	Unknown	Unknown Details
33	Plantation, Florida	3/19/2016	Plantation	FL	Florida	Home	party gues	unknown	A Spring B	1	4	5	0	Unknown	Unknown Details	Unkno	Unknow	26.128	-80.2331	Unknown	Unknown Details
34	Wetumpka Drive-by	3/19/2016	Wetumpka	AL	Alabama	drive-by in We	random	terrorism	Shooter fir	2	2	4	0	Unknown	Unknown Details	Black	Male	32.544	-86.2119	Unknown	Unknown Details
35	Atlanta Nightclub shooting	3/15/2016	Atlanta	GA	Georgia	outside nightc	random	anger	Two grour	0	4	4	0	Unknown	Unknown Details	Unkno	Unknow	33.749	-84.388	Unknown	Unknown Details

Figure 3 US Mass Shootings merged data after formatting

The datasets 6 and 7 consists data in bar graphs in the websites provided, which is shown in *Figure 4*. The data must be scraped and converted into a tabular format which is a lot easier for analysis. To scrape the data, we need to inspect the elements of the HTML code in the website. The HTML code is shown in *Figure 5*. The data is present in the '<table>' nested element which consists of inner nested elements such as

‘<tbody>’, ‘<tr>’, and ‘<tr>’. We need to use python packages ‘urlopen’, ‘Beautifulsoup’ to extract the data between HTML elements.

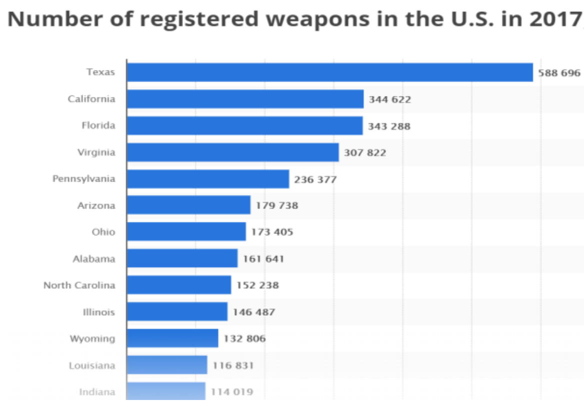

Figure 4

```
▼<table id="statTable--mobile" class="table dataTable">
  ▼<thead>
    ▼<tr role="row">
        <th class="sorting" role="columnheader" tabindex="0" aria-controls="statTable--mobile" rowspan="1" colspan="1"
        aria-label=": activate to sort column ascending"></th>
        <th class="sorting" role="columnheader" tabindex="0" aria-controls="statTable--mobile" rowspan="1" colspan="1"
        aria-label="Number of registered weapons: activate to sort column ascending">Number of registered weapons</th>
      </tr>
    </thead>
  ▼<tbody role="alert" aria-live="polite" aria-relevant="all">
    ▼<tr class="odd">
        <td class=" ">Texas</td>
        <td class=" ">588,696</td>
      </tr>
    ▼<tr class="even">
        <td class=" ">California</td>
        <td class=" ">344,622</td>
      </tr>
    ▶<tr class="odd">…</tr>
    ▶<tr class="even">…</tr>
    ▶<tr class="odd">…</tr>
    ▶<tr class="even">…</tr>
    ▶<tr class="odd">…</tr>
```

Figure 5 Html code of the graphical data

The python code to extract the data between the elements is shown in *Figure 6*. The graphical data is converted into a tabular format which is stored as a CSV file.

```
#importing packages
import csv
from urllib.request import urlopen
from bs4 import BeautifulSoup
#url link is provided
URL = "https://www.statista.com/statistics/215655/number-of-registered-weapons-in-the-us-by-country
url_response = urlopen(URL) #extracting html code from url
parsed_response = BeautifulSoup(url_response, "html.parser") #parsing each line
corresponding_table = parsed_response.find("table", attrs={"id": "statTable"}) #finding the data ta
#extracting data from html elements
all_rows = corresponding_table.findAll("tr")
header = all_rows[0]
rows = all_rows[1:]
state_weapons_rows = []
for row in rows:
    cells = row.findAll("td")
    state = cells[0].text
    weapon_count = cells[1].text
    state_weapons_rows.append([state, weapon_count])
#writing to csv file
with open('us_weapons_count.csv', 'w') as csvfile:
    spamwriter = csv.writer(csvfile, delimiter=",")
    for row in state_weapons_rows:
```

Figure 6 Python code for extracting tabular data from the graph

In the similar fashion, the graphical content in the dataset 6 is also extracted into a CSV file format. Now we have all the datasets which are clean and ready for the exploration.

Data Exploration:

The first step in data analysis is data exploration and in general, it involves summarizing the inferences from the datasets. In this project, it is conducted by using visualization tool Tableau and advanced statistical software R programming. Let's try to analyze the datasets by using Tableau and then we will use R programming to run few statistical tests on data.

The formatted dataset which is shown in *Figure 3* is loaded into Tableau. The 'Date' column in the dataset should be split into 'Day', 'Month', and 'Year' to explore the data year-wise and month-wise. This is achieved by creating a calculated field and using 'datename' function in the tableau. And, the 'age' column in the dataset is grouped into buckets. This is also achieved by creating a calculated field and using if, elif, and else statements in the tableau. Now we have all the required fields in shape for the exploration. Let's begin exploration of the data by analyzing data on various levels.

Mass Shootings in the US by year:

In *Figure 7*, we can see that year is plotted on the x-axis, number of shootings is plotted on left-hand side y-axis, and number of total victims is plotted on the right-hand side of y-axis which is achieved by using dual axis in Tableau. The number of shootings is a bar graph by year and number of total victims is a line graph. A trend line is added for number of total victims.

We can see that the number of shootings increased drastically in the years 2015 and 2016. By looking at the trend line, we can infer that the total number of victims increased over the period along with the number of shootings. In 2017, even though there are lesser shootings compared to the previous years, but the total number of victims are more than 2015 and 2016. This is because of the Las Vegas Shooting happened in the year 2017, which is an outlier. In Figure 8, the bar graph depicts the number of fatalities, injured, and the total number of victims in each year. We can infer that 2015, 2016, and 2017 are the most tragic years in the history of the United States. The number of people injured in 2017 is more than the number of people injured in 2015 and 2016 combined because of the Las Vegas Shooting which occurred in 2017.

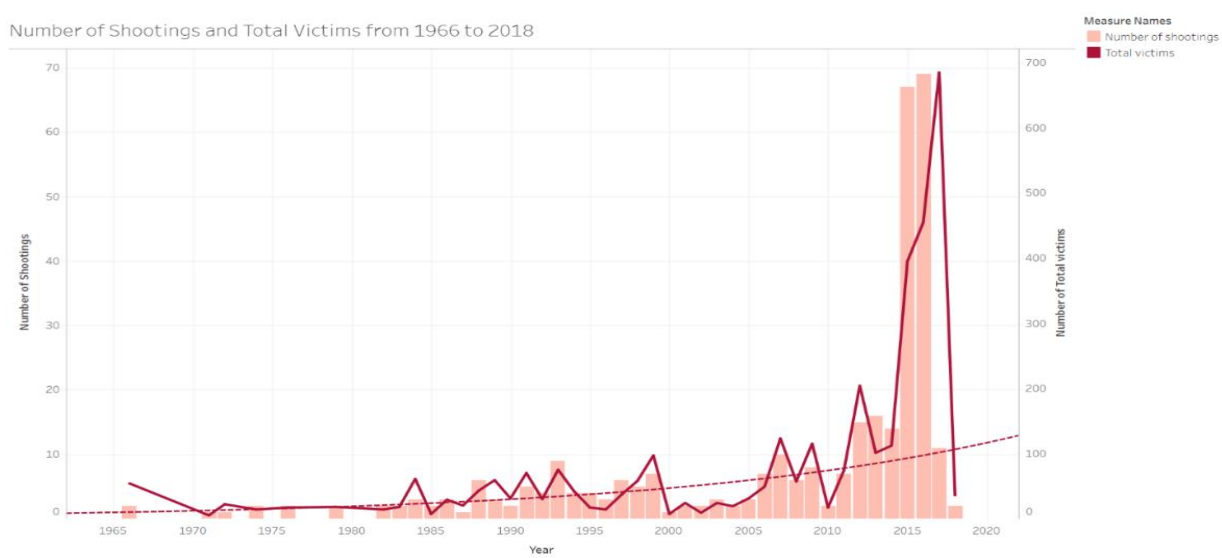

Figure 7 Number of shootings and total victims year-wise

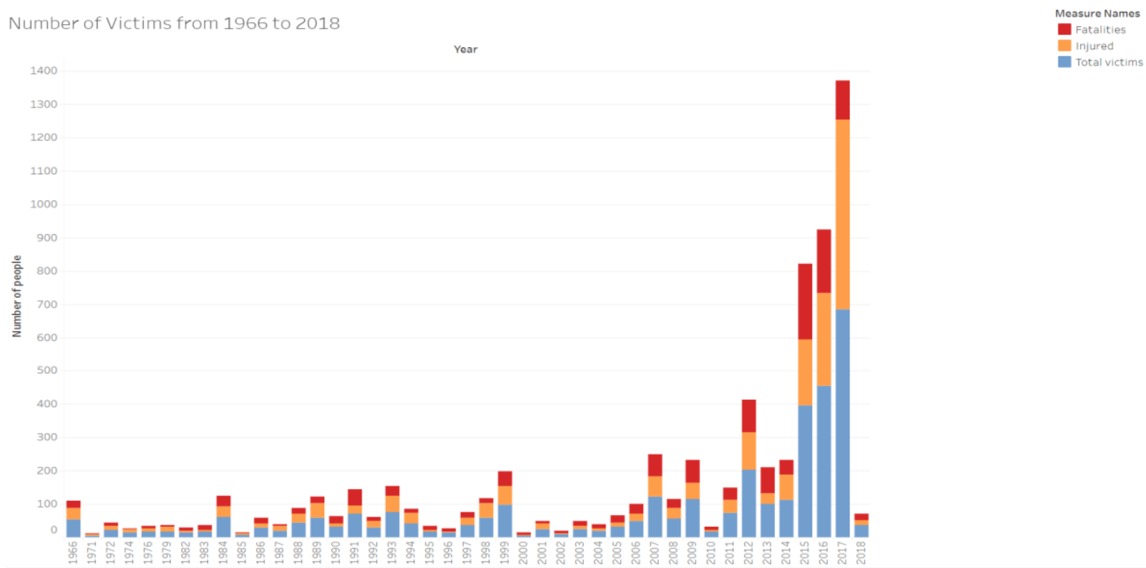

Figure 8 Total Number of victims in each year

Mass Shootings in the US by city and state:

In *Figure 9*, cities are plotted on the x-axis, number of shootings, and total victims are plotted on the dual y-axis. The bar graph depicts the number of shootings whereas the line graph depicts the total number of victims in each city. This graph only consists of top 10 cities with the highest number of shootings. We can infer that number of shootings in the cities do not differ much. Las Vegas and Killeen cities have the highest number of victims.

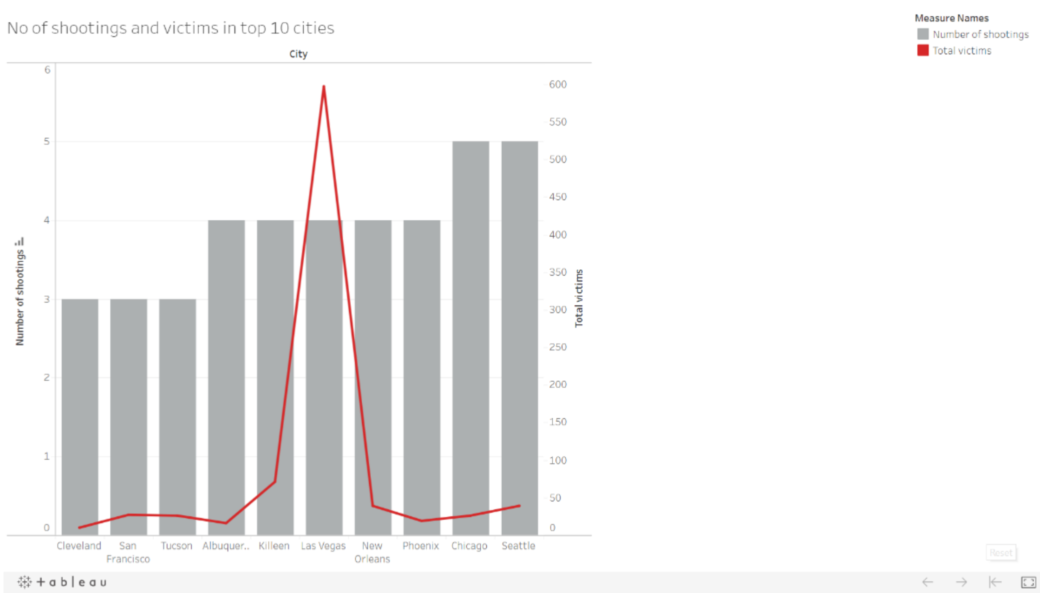

Figure 9 Number of Shootings and Victims by city

In *Figure 10*, states are plotted on the x-axis, number of shootings, and total victims are plotted on the dual y-axis. The bar graph depicts the number of shootings whereas the line graph depicts the total number of victims in each state. *Figure 10* and *Figure 11* depicts the number of shootings and percentage of shootings in each state respectively. *In Figure 10*, the number of shootings is shown in a bar graph and number of victims in a line graph whereas in *Figure 11*, the percentage of shootings is shown on a US map. The states which have the highest number of shootings overall are California, Florida, Texas, Washington, and Georgia. The number of victims in each state is proportional to number of shootings occurred except in the Nevada state because of the Las Vegas attack, which is an outlier.

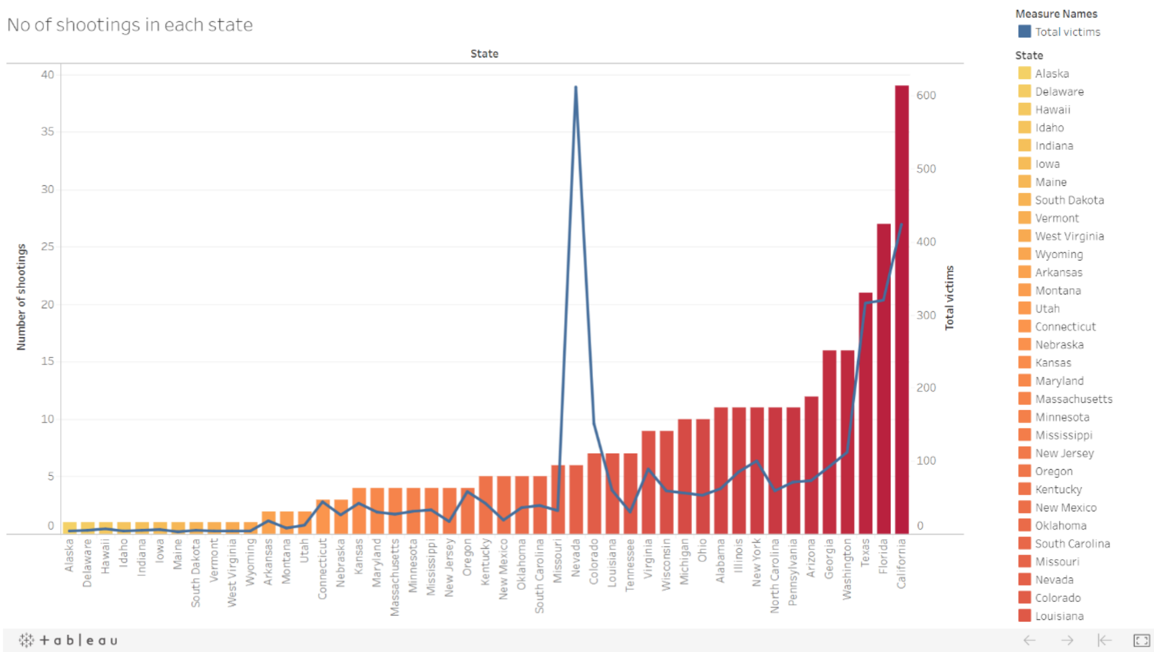

Figure 10 Number of Shootings and Victims by state

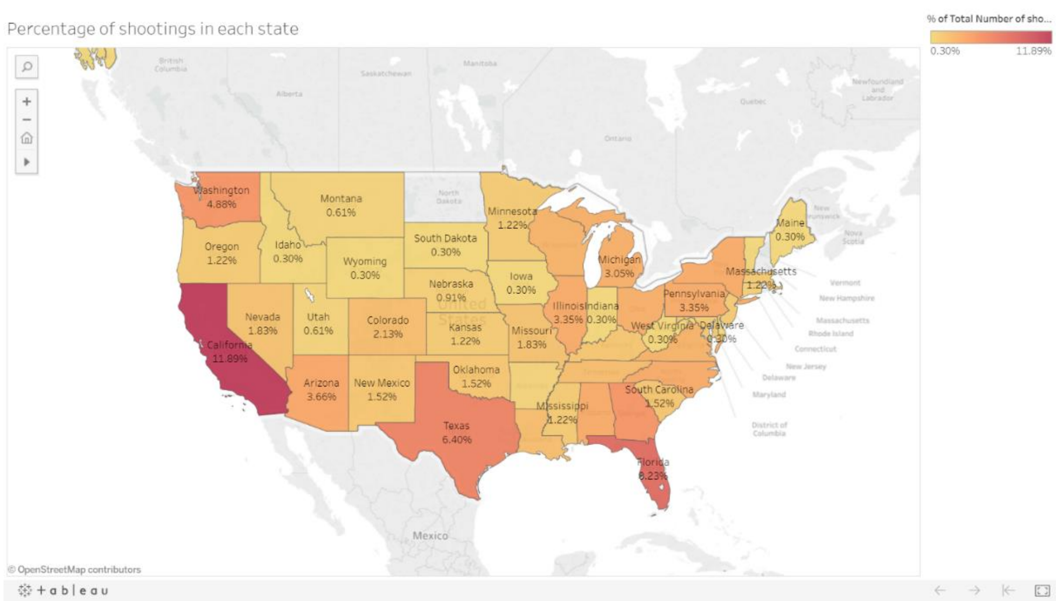

Figure 11 Percentage of Shootings in each state shown on US map

From *Figure 11*, we can infer that mass shootings happened in every state except North Dakota. Based on this we can develop a hypothesis that the number of shootings in a state is dependent on the gun ownership in that state. Now we must find out if there is any correlation between the number of shootings and number of weapons in a state. This can be achieved by calculating the Pearson Correlation Coefficient in R. The dataset 6 contains the data of number of weapons in each state. We need to merge the merged dataset we used for analysis earlier and dataset 6 to calculate the correlation. The total number of shootings and number of weapons are plotted in a graph which is shown in *Figure 12.*

We load the dataset into a 'dataframe' in R and calculate Pearson correlation coefficient using the function 'cor( )'. We get the correlation coefficient as 0.725, from which we can infer that the relationship between the number of weapons and shootings in a state is linearly related with a direct proportionality. Also, from *Figure 12*, we can see that more the number of weapons, there are more shootings in that state for most of the states.

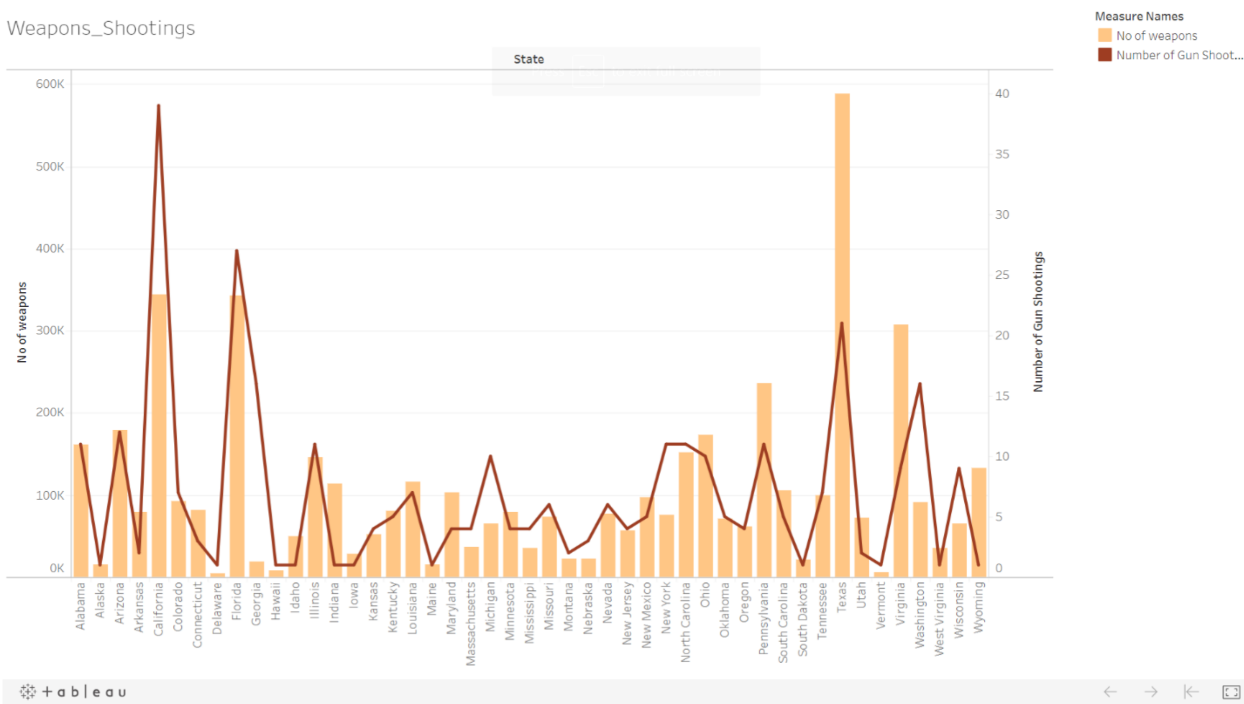

Figure 12Number of gun shootings and weapons by state

Mass Shootings in the US by gender:

The percentage of mass shootings committed by individuals based on their gender is given in the following table. We infer that in around 97% of the shootings, the shooters are males. In few shootings, both males and females are involved.

Percentage of shootings per gender

Gender	
Female	1.63%
Male	96.74%
Male/Female	1.63%

Deadliest Months:

In this section, we will try to identify if there is any pattern of shootings in a specific month. This is achieved by plotting a graph which contains year on the y-axis and month on the x-axis. *In Figure 13*, we can see the plot and the number of shootings are depicted as small circles with varying sizes. The size indicates the number of shootings in each month and year. There is no general pattern of these shootings overall but if we consider from the year 2000, we can infer that January, February, and October are the deadliest months in the history of the United States.

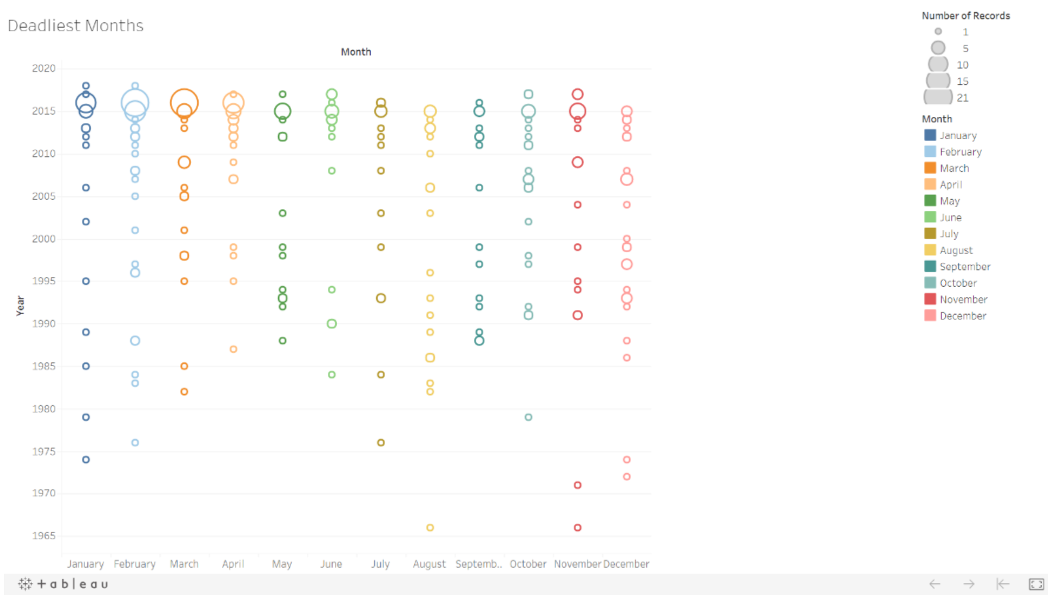

Figure 13 Number of shootings in a year and month

Mental Health issues of shooters:

The individuals who commit this kind of activities, in general, have mental health issues. In *Figure 14*, we can see the percentage of shooters having mental health issues. From the data, we infer that more than 50% of the shooters have mental health issues. To identify what kind of mental health issues they are suffering from, we need to look into the details of the mental health condition which is given by the field 'Mental Health Details' from the data. Text analysis is used to find the details of the mental health issues of shooters. A word cloud is created in R to find out what kind of issues these shooters have. The R code for the word cloud is shown in *Figure 15* and the word cloud in *Figure 14*.

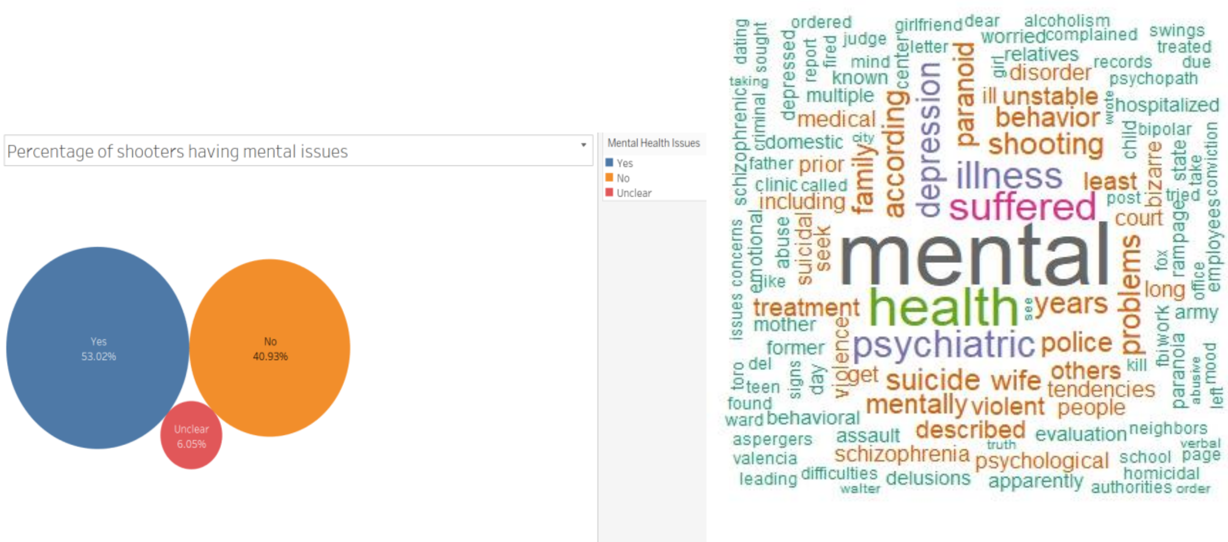

Figure 14 Number of shooters having mental health issues and word cloud for mental health details

From *Figure 14*, we can see that the shooters are suffering from mental health conditions such as depression, psychiatric and psychological problems, paranoid problems, schizophrenia etc.

```r
# Installing packages
install.packages("tm")  # for text mining
install.packages("SnowballC") # for text stemming
install.packages("wordcloud") # word-cloud generator
install.packages("RColorBrewer") # color palettes
# Loading packages
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")

word <- Corpus(VectorSource(shootings$Mental.health...details))

# Convert the text to lower case
word <- tm_map(word, content_transformer(tolower))
# Remove numbers
word <- tm_map(word, removeNumbers)
# Remove english common stopwords
word <- tm_map(word, removeWords, stopwords('english'))
word <- tm_map(word, removeWords, c('unknown','one','two','three','details','history','told','also','said','can','back','unclear'))
# Remove punctuations
word <- tm_map(word, removePunctuation)
# Eliminate extra white spaces
word <- tm_map(word, stripWhitespace)

tdm <- TermDocumentMatrix(word)
mat_word <- as.matrix(tdm)
v_word <- sort(rowSums(mat_word),decreasing=TRUE)
d_word <- data.frame(word = names(v_word),freq=v_word)

set.seed(1234)
wordcloud(words = d_word$word, freq = d_word$freq, min.freq = 1,
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
```

Figure 15 R-code for the creation of word cloud for mental health details

Causes of the shootings and Primary Targets:

The main causes of the shootings are depicted in a pie chart which is shown in *Figure 16*. They are the psychotic behavior of the shooter, terrorism, and anger related issues. Let's identify if there is any relation between the cause of the shooting and mental health condition of the shooter.

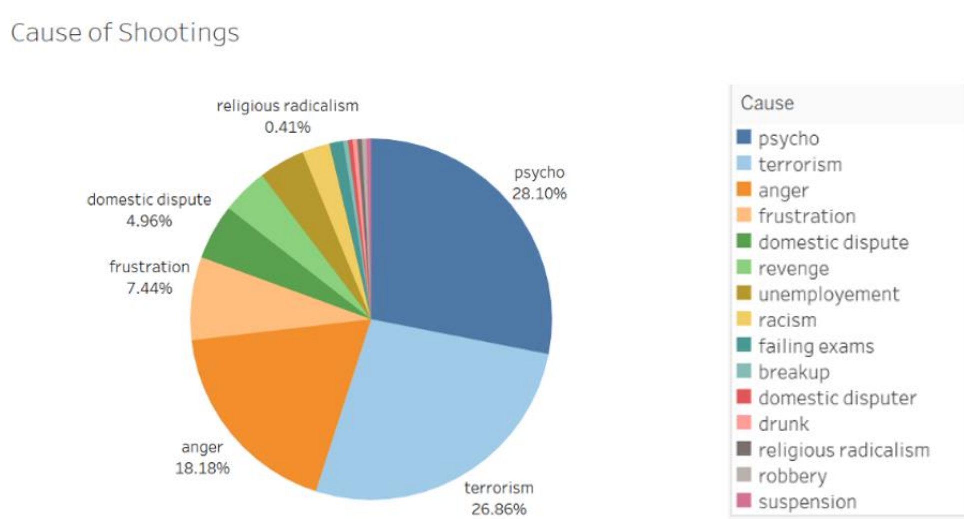

Figure 16 Pie-chart showing cause of the shootings

In Figure 17, we can see that the more than 60% of the shooters with causes of revenge, terrorism, domestic dispute, racism, psychotic behavior etc. are suffering from mental health conditions. The top primary targets of the shooters are random people, family, co-workers, students, and teachers. This can be inferred from Figure 18.

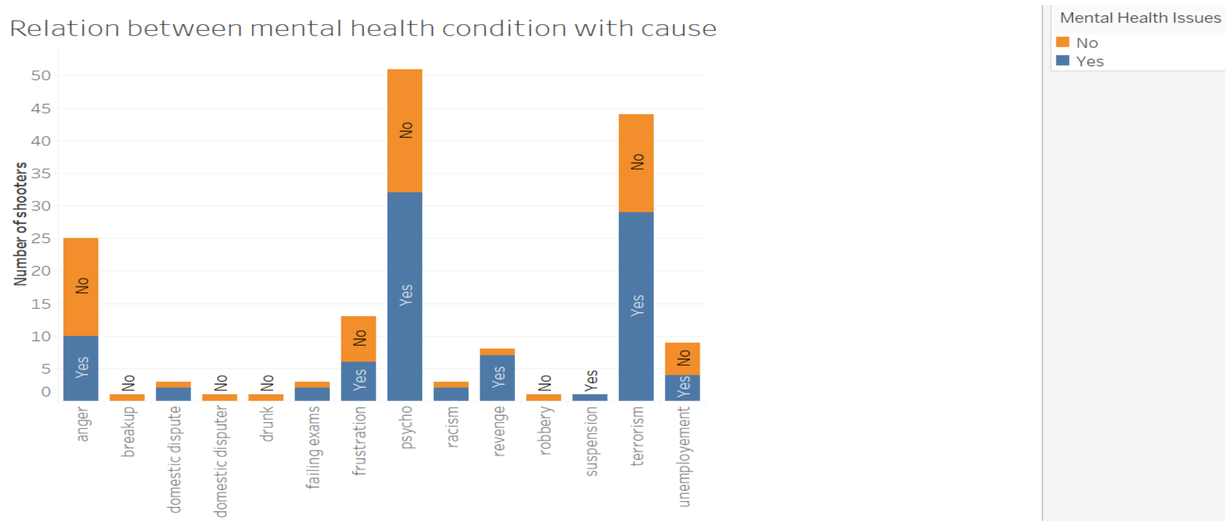

Figure 17 Relation between mental health condition of shooter and cause of the shooting

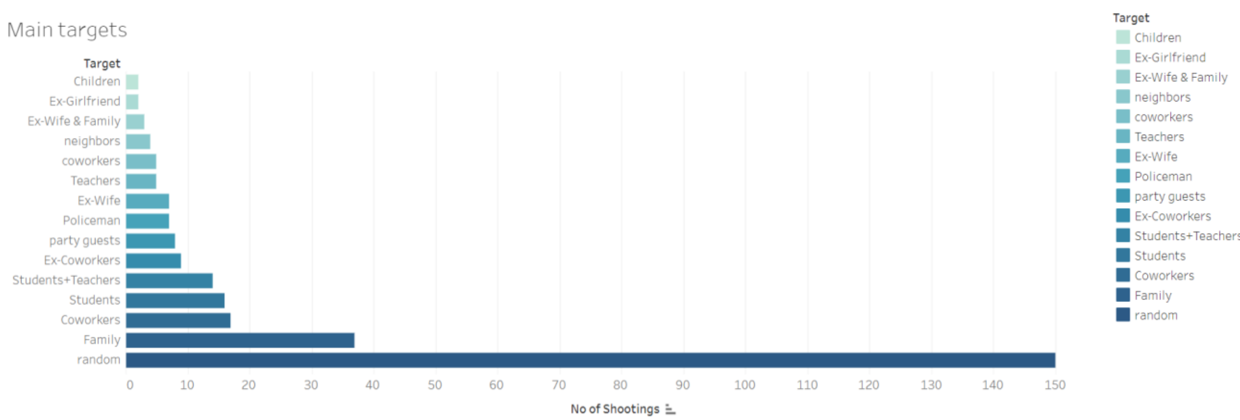

Figure 18 Main Targets of the Shooters

Number of shootings by race and gender:

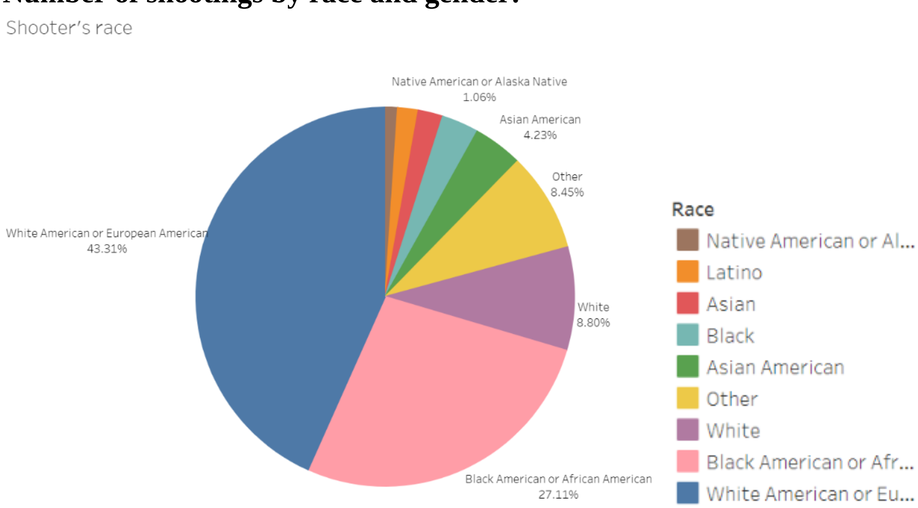

Figure 19 Number of shootings by race

The percentage of shootings by race is shown in *Figure 19*. More than 70% of the shooters are American natives. The highest number of shootings are committed by the White Americans or European Americans and Black Americans or African Americans. In *Figure 20*, we can see that the highest number of shootings are committed by the shooters of age between 10 and 20. As the age increases, the number of shootings committed decreases. We can infer that in all the age groups, the highest number of shootings done by the individuals belong to White American or European American race.

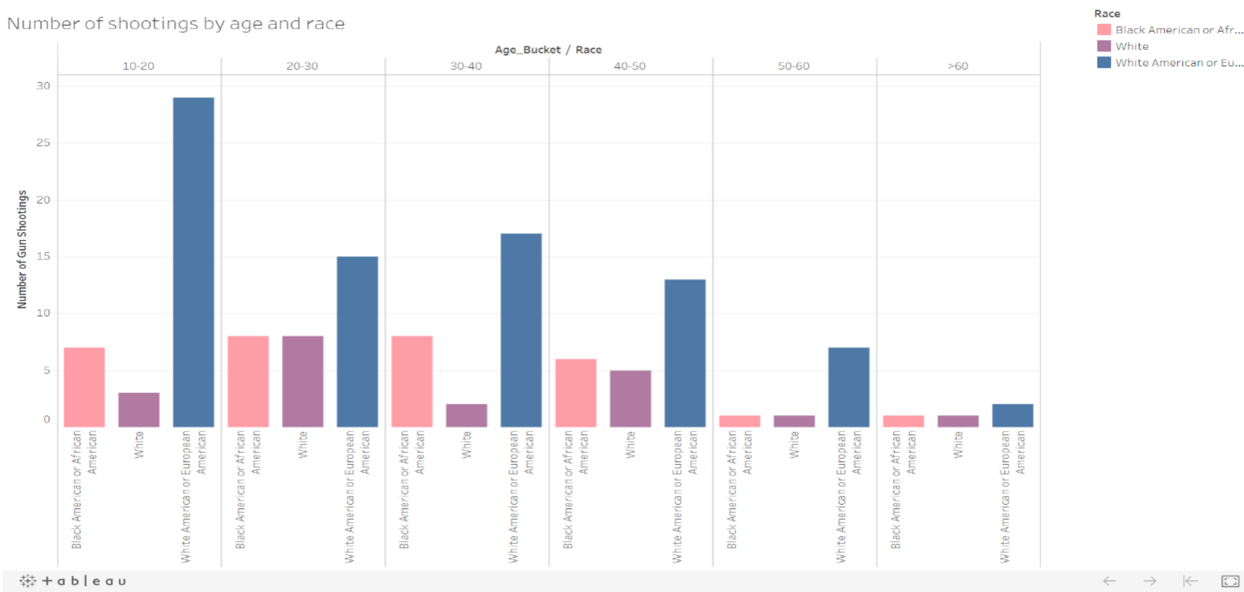

Figure 20 Number of shootings by age and race

Relation between the percentage of guns and shootings by year:

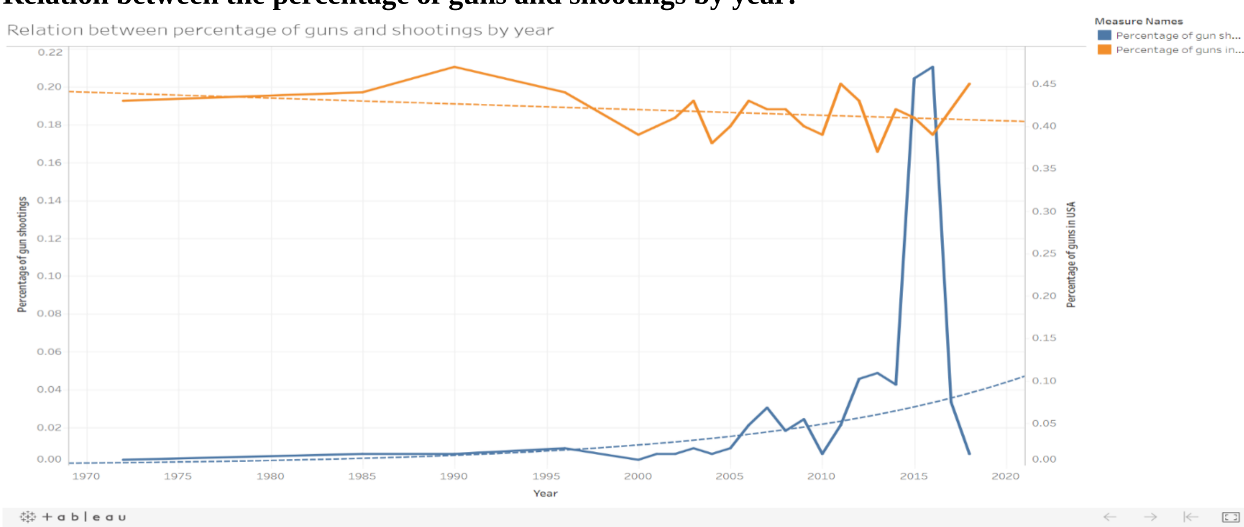

Figure 21 Percentage of Gun Shootings and Percentage of Number of guns for each year

The percentage of gun shootings and the percentage of guns owned by individuals is depicted in *Figure 21*. It is common to assume that when the percentage of guns possessed by individuals over the time decreases, the percentage of shootings decrease. But from the above graph, we can observe that even though the number of guns owned by people has decreased in years the number of mass shootings has increased. The Pearson Correlation coefficient is calculated for the percentage of shootings and percentage of guns by year. It turns out that the correlation coefficient is -0.14, which tells us that the both these factors are inversely correlated. This was an interesting finding from the above exploration as it contradicts the common assumption of ours.

Summary:

Although the number of guns in America decreased over the time, the frequency of shootings has not decreased. Surprisingly, the frequency of mass shootings is tripled in the last few years. On an average, 8 shootings occurred every year in the last 50 years that took 35 lives and 47 injured per year. The states which have the highest number of mass shootings are California, Florida, and Texas whereas the states with the least number of mass shootings are Alaska, Delaware, and Hawaii. The number of mass shootings in a state is directly proportional to the number of weapons owned by individuals in that state. This trend is similar for most of the states in the country. By exploring the data, we inferred that the months of January, February, and October are the deadliest months in the history of the United States. We also observed that more than 50% of the shooters are suffering from mental health conditions and the most causes of the shootings are psychotic behavior, revenge, and terrorism. The primary targets in these attacks are random strangers, co-workers, family members, students, and teachers. More than 70% of the shooters are American natives and these attacks are committed by the individuals of the age group between 10 and 20. The median age of the shooter is 15 years, which infers that individuals at a young age are turning into killers. The insights from the data were able to answer most of the questions mentioned above except the prediction of these attacks in future. There is no specific pattern to predict such kind of attacks in future. To put a check to such kind of attacks in future, the government of the United States must reform the gun laws like Australia did in 1996 and monitor constant background checks of the citizens in the country.

Reflection:

I was really motivated to do this project. Now I know where to find the datasets as I have explored a lot of websites in finding a good dataset and topic I am interested in. This project has enhanced my technical aspects in exploring and visualizing the data. I have improved my technical skills in using Python, R-programming, Tableau, and I also learned how to scrape data from a website. Data wrangling is a laborious process but cleaning the datasets is very crucial to achieve insights from the data. I am overall satisfied with the analysis I have performed but it would have been better if I had used a few more statistical tests to validate the hypothesis I assumed. I am looking forward to work on the Visualization project as I am excited to work on R-shiny and D-3 JS.

References:

1) US Mass Shootings | Kaggle. (2018). Kaggle.com. Retrieved 20 March 2018, from https://www.kaggle.com/zusmani/us-mass-shootings-last-50-years
2) Mass shootings occurring in the US from 1982-2018 (2018). Data.world. Retrieved 1 April 2018, from https://data.world/awram/us-mass-shootings.
3) Number of registered weapons in the U.S. in 2017, s. (2018). U.S. - number of registered weapons by state 2017 | Statistic. Statista. Retrieved 3 April 2018, from https://www.statista.com/statistics/215655/number-of-registered-weapons-in-the-us-by-country/
4) Gun ownership in the U.S. 1972-2017 | Statistic. (2018). Statista. Retrieved 6 April 2018, from https://www.statista.com/statistics/249740/percentage-of-households-in-the-united-states-owning-a-firearm/
5) US Cities Database | Simplemaps.com. (2018). Simplemaps.com. Retrieved 15 April 2018, from https://simplemaps.com/data/us-cities
6) List of 50 US States in Excel - Scott on Technology. (2018). Scott on Technology. Retrieved 16 April 2018, from https://scottontechnology.com/list-of-50-us-states-in-excel/
7) Mass shootings in the United States. (2018). En.wikipedia.org. Retrieved 23 April 2018, from https://en.wikipedia.org/wiki/Mass_shootings_in_the_United_States
8) Gun laws stopped mass shootings in Australia. (2018). The University of Sydney. Retrieved 29 April 2018, from https://sydney.edu.au/news-opinion/news/2018/03/13/gun-laws-stopped-mass-shootings-in-australia.html