NORTH AMERICAN UNIVERSITY

COMPUTER SCIENCE DEPARTMENT

CAPSTONE PROJECT

DATA ANALYSIS OF GUN VIOLENCE IN THE UNITED STATES

BY

NADIDE PASALI

ALMAGUL AINABEKOVA

ABSTRACT

Gun violence incidents in the United States significantly increasing year by year. Research has shown that gun homicides climbed up to 30% from 2014 to 2017 and gun violence has reached a peak of 40 years. This study aims to thoroughly scrutinize gun violence incidents and trends.

Based on a dataset of the gun violence incidents, data was collected through media sources and from resources like FBI. Records of incidents are divided into 29 different features in a timely ascending manner and the dataset was analyzed according to two main categories as time and location. Analysis of the study demonstrated that holiday and hot seasons encourage violence rates and larger cities have high crime rates. The results indicate that during vacation seasons people are more exposed to gun related attacks and high crime rated states have short-term and light punishment regulations.

TABLE OF CONTENTS

TABLE OF FIGURES

1. INTRODUCTION

United States is one of the top countries in the world by number of gun ownership, consequently, gun violence is one of the major concerns in the United States. In the Constitution of the United States in Amendment 2 - Right to Bear Arms it is stated:

> "*A well regulated Militia, being necessary to the security of a free State, the right of the people to keep and bear Arms, shall not be infringed*"
>
> (Ratified 12/15/1791) [1].

Therefore, every American constitutionally protected to carry a gun open or concealed which weakens law enforcement as well as boosts the crime rate. Investigation of gun violence incidents is essential for government to regulate the permits, strengthen the usage restrictions and background checks as well as mental competency of a person for obtaining firearms.

1.1. Statement of The Problem

The purpose of this study was to investigate gun violence incidents and trends to find out time and location-based progress.

1.2. Research Objectives

In this research, findings were the significant increase in the number of victims for about 24% in four year range, highest rates were observed in big cities with population over 600,000, the seasonality of incidents was observed, number of males involved are significantly higher than the number of females and age range of involved individuals is between 17-26.

1.3. Structure of The Report

This analysis report elaborates data collection, data manipulation, data cleaning, exploratory data analysis, data visualization, time and location related analyses, time series analysis, gender and age-related analyses, summary of findings and outcomes.

2. DATA

Gun violence incident data was gathered from Gun Violence Archive (GVA). GVA is a non-profit corporation founded in late 2012 to provide free internet access to factual data on gun-related violence in the United States. GVA collects and verifies precision, extensive information on U.S. gun-related violence, and then post it online and publicize it. GVA's gathered data collected from resources like the FBI, CDC, NIH and other organizations with established standards. According to GVA, data is obtained through government and media sources on a daily basis and the slight difference between the FBI and other sources is due to the FBI's and the other sources' records being extrapolated and sampled for aggregated totals in their methodologies [2].

Gun violence data for 2013-2018 years period consists of 6,950,633 records including number of incidents and features. It contains different data types as integer values, objects and floats. Some values are missing, however the part which was used for analysis has less to none missing values. Moreover, four features with URL records were removed, since they contained extra information about incidents, generally, news and TV channel websites. Raw data has features as incident ID, number of killed and injured people, gender and age ranges involved, locations of incidents, suspects and victims, number of guns involved, gun types, incident characteristics, state house and senate house districts and participant relations.

3. METHOD

Mainly, the Python programming language and Tableau were used for this project. Implementation of exploratory data analysis and data preparation part of the gun violence analysis made with Python in Jupyter notebook environment. Numpy, Pandas, Plotly and Folium libraries are utilized. Time and location related analyses and data visualization implemented on Tableau Public environment. For better collaboration Git and GitHub version control system was used and all work done was saved in GitHub.

3.1. Exploratory Data Analysis (EDA)

Exploratory Data Analysis relates to the critical method of conducting original data investigations to detect patterns, detect anomalies, test hypotheses, and use summary statistics and graphical representations to verify assumptions.

EDA of the project includes detailed exploration of data features, descriptive statistics of numeric features, and investigation of missing data. In order to perform gun violence analysis with high accuracy, some data cleaning and preparations were made. Data preprocessed for further functions and visualization.

3.1.1. Exploring Data

After data was imported to Jupyter notebook, first, dimensions of data were assessed with *shape* function in Pandas and listed first elements of data with *head* function. Data frame has shape of 239,677 rows and 29 columns. The following is a screenshot of used functions,

```python
df = pd.read_csv('../../data/gun-violence-data_01-2013_03-2018.csv')

print(df.shape)
df.head()
```

```
(239677, 29)
```

	incident_id	date	state	city_or_county	address	n_killed	n_injured	incident_url	
0	461105	2013-01-01	Pennsylvania	Mckeesport	1506 Versailles Avenue and Coursin Street	0	4	http://www.gunviolencearchive.org/incident/461105	gazette.c
1	460726	2013-01-01	California	Hawthorne	13500 block of Cerise Avenue	1	3	http://www.gunviolencearchive.org/incident/460726	http://www.dailybulletin
2	478855	2013-01-01	Ohio	Lorain	1776 East 28th Street	1	3	http://www.gunviolencearchive.org/incident/478855	http://chronicle.northcoas
3	478925	2013-01-05	Colorado	Aurora	16000 block of East Ithaca Place	4	0	http://www.gunviolencearchive.org/incident/478925	http://www.dailydemocra
4	478959	2013-01-07	North Carolina	Greensboro	307 Mourning Dove Terrace	2	2	http://www.gunviolencearchive.org/incident/478959	http://www.journalnow.cc

5 rows × 29 columns

Figure 1: Reading dataset and showing first five rows of dataset

All columns and their characteristics were listed by *info* functions. Data has 19 object type, 9 numeric type, and 1 Boolean type columns. All columns have different number of records as seen below.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 239677 entries, 0 to 239676
Data columns (total 29 columns):
incident_id                    239677 non-null int64
date                           239677 non-null object
state                          239677 non-null object
city_or_county                 239677 non-null object
address                        223180 non-null object
n_killed                       239677 non-null int64
n_injured                      239677 non-null int64
incident_url                   239677 non-null object
source_url                     239209 non-null object
incident_url_fields_missing    239677 non-null bool
congressional_district         227733 non-null float64
gun_stolen                     140179 non-null object
gun_type                       140226 non-null object
incident_characteristics       239351 non-null object
latitude                       231754 non-null float64
location_description           42089 non-null object
longitude                      231754 non-null float64
n_guns_involved                140226 non-null float64
notes                          158660 non-null object
participant_age                147379 non-null object
participant_age_group          197558 non-null object
participant_gender             203315 non-null object
participant_name               117424 non-null object
participant_relationship       15774 non-null object
participant_status             212051 non-null object
participant_type               214814 non-null object
sources                        239068 non-null object
state_house_district           200905 non-null float64
state_senate_district          207342 non-null float64
dtypes: bool(1), float64(6), int64(3), object(19)
memory usage: 51.4+ MB
```

Figure 2: Summary of the data frame

We eliminated useless columns and calculated rest three numeric columns' descriptive statistics with *describe* function for understanding of distribution.

```
df[['n_killed','n_injured', 'n_guns_involved']].describe()
```

	n_killed	n_injured	n_guns_involved
count	239677.000000	239677.000000	140226.000000
mean	0.252290	0.494007	1.372442
std	0.521779	0.729952	4.678202
min	0.000000	0.000000	1.000000
25%	0.000000	0.000000	1.000000
50%	0.000000	0.000000	1.000000
75%	0.000000	1.000000	1.000000
max	50.000000	53.000000	400.000000

Figure 3: Statistical details of the dataset

We checked for the missing data by percentage. Some columns such as participant relation, name and location description had very few data, even impossible to apply for analysis. However, observed that the number of killed and injured people, date, states and counties columns have zero percent missing data which is fully convenient for applications.

It gave explicit insight that data is very clear and straightforward for location and time-based analysis on number of victims.

```
# checking missing data
total = df.isnull().sum().sort_values(ascending = False)
percent = (df.isnull().sum()/df.isnull().count()*100).sort_values(ascending = False)
missing_gun_data  = pd.concat([total, percent], axis=1, keys=['Total', 'Percent'])
missing_gun_data
```

	Total	Percent
participant_relationship	223903	93.418643
location_description	197588	82.439283
participant_name	122253	51.007397
gun_stolen	99498	41.513370
gun_type	99451	41.493760
n_guns_involved	99451	41.493760
participant_age	92298	38.509327
notes	81017	33.802576
participant_age_group	42119	17.573234
state_house_district	38772	16.176771
participant_gender	36362	15.171251
state_senate_district	32335	13.491073
participant_status	27626	11.526346
participant_type	24863	10.373544
address	16497	6.883013
congressional_district	11944	4.983373
latitude	7923	3.305699
longitude	7923	3.305699
sources	609	0.254092
source_url	468	0.195263
incident_characteristics	326	0.136016
incident_url_fields_missing	0	0.000000
incident_url	0	0.000000
n_injured	0	0.000000
n_killed	0	0.000000
city_or_county	0	0.000000
state	0	0.000000
date	0	0.000000
incident_id	0	0.000000

Figure 4: Checking for missing data

3.1.2. Splitting Date to Day/Month/Year

Data required some pre-modifications for further processes. Date was the first concern. It was observed that date is formatted as ‘yyyy-mm-dd.’ In order to study time-related trend, we decided to split date to *year*, *month* and *monthday*. Moreover, we added *weekday* to investigate from Monday through Sunday.

```
df['date'] = pd.to_datetime(df['date'])
df['year'] = df['date'].dt.year
df['month'] = df['date'].dt.month
df['monthday'] = df['date'].dt.day
df['weekday'] = df['date'].dt.weekday

df.head()
```

...	state_house_district	state_senate_district	n_guns	participant_type_map	participant_age_map	participant_gender_map	year	month	monthday	weekday
...	NaN	NaN	nan	{'0': 'Victim', '1': 'Victim', '2': 'Victim', ...	{'0': '20'}	{'0': 'Male', '1': 'Male', '3': 'Male', '4': '...	2013	1	1	1
...	62.0	35.0	nan	{'0': 'Victim', '1': 'Victim', '2': 'Victim', ...	{'0': '20'}	{'0': 'Male'}	2013	1	1	1
...	56.0	13.0	2.0	{'0': 'Subject-Suspect', '1': 'Subject-Suspect...	{'0': '25', '1': '31', '2': '33', '3': '34', '...	{'0': 'Male', '1': 'Male', '2': 'Male', '3': '...	2013	1	1	1
...	40.0	28.0	nan	{'0': 'Victim', '1': 'Victim', '2': 'Victim', ...	{'0': '29', '1': '33', '2': '56', '3': '33'}	{'0': 'Female', '1': 'Male', '2': 'Male', '3':...	2013	1	5	5
...	62.0	27.0	2.0	{'0': 'Victim', '1': 'Victim', '2': 'Victim', ...	{'0': '18', '1': '46', '2': '14', '3': '47'}	{'0': 'Female', '1': 'Male', '2': 'Male', '3':...	2013	1	7	0

Figure 5: Splitting 'date' column into year, month, day and weekday

3.1.3. Drop Years of 2013 and 2018

After we split year from date, we wanted to count the number of records for each year. With *value_counts* function, we realized that years of 2013 and 2018 have a lot of missing data. These two years with missing data would alter time related results and misguide us, so we dropped all records which happened in 2013 and 2018.

```
df['year'].value_counts()
```

```
2017    61401
2016    58763
2015    53579
2014    51854
2018    13802
2013      278
Name: year, dtype: int64
```

Figure 6: Amount of records for each year

```
df = df.set_index('year')
df = df.drop(2013, axis=0)
df = df.drop(2018, axis=0)
df.reset_index(inplace = True)
df
```

	year	incident_id	date	state	city_or_county	address	n_killed	n_injured	incident_url	
0	2014	95289	2014-01-01	Michigan	Muskegon	300 block of Monroe Avenue	0	0	http://www.gunviolencearchive.org/incident/95289	http:
1	2014	92401	2014-01-01	New Jersey	Newark	Central Avenue	0	0	http://www.gunviolencearchive.org/incident/92401	http
2	2014	92383	2014-01-01	New York	Queens	113th Avenue	1	0	http://www.gunviolencearchive.org/incident/92383	h
3	2014	92142	2014-01-01	New York	Brooklyn	St. Johns Place	0	1	http://www.gunviolencearchive.org/incident/92142	http:
4	2014	95261	2014-01-01	Missouri	Springfield	Beverly Hills and Temple	0	1	http://www.gunviolencearchive.org/incident/95261	h
5	2014	92272	2014-01-01	Georgia	Columbus	1327 23rd Street	0	1	http://www.gunviolencearchive.org/incident/92272	http

Figure 7: Removing years with missing data

3.1.4. User Mapping for Participant Type, Age, and Gender

When we dig into columns, we wanted to reformat some columns which are participant type, age, and gender. For example, *participant_type* column was formatted as "*participant_id*::*participant_type*||...". Gender and age were also formatted in the same manner. Long string of these rules was impractical. Most imitative data saving format like columns was dictionary format of Python. We assumed id as *key*, other parameter as *value*. We defined a function called *get_user_mapping* to map all records by reformatting. Then we added three more columns to store mapping versions.

```
df[['participant_type', 'participant_age', 'participant_gender']].head()
```

	participant_type	participant_age	participant_gender
0	0::Victim||1::Victim||2::Victim||3::Victim||4:...	0::20	0::Male||1::Male||3::Male||4::Female
1	0::Victim||1::Victim||2::Victim||3::Victim||4:...	0::20	0::Male
2	0::Subject-Suspect||1::Subject-Suspect||2::Vic...	0::25||1::31||2::33||3::34||4::33	0::Male||1::Male||2::Male||3::Male||4::Male
3	0::Victim||1::Victim||2::Victim||3::Subject-Su...	0::29||1::33||2::56||3::33	0::Female||1::Male||2::Male||3::Male
4	0::Victim||1::Victim||2::Victim||3::Subject-Su...	0::18||1::46||2::14||3::47	0::Female||1::Male||2::Male||3::Female

```
def get_user_mapping(txt):
    if txt == "NA":
        return {}
    mapping = {}
    for d in txt.split("||"):
        try:
            key = d.split("::")[0]
            val = d.split("::")[1]
            if key not in mapping:
                mapping[key] = val
        except:
            pass
    return mapping

df['participant_type'] = df['participant_type'].fillna("NA")
df['participant_type_map'] = df['participant_type'].apply(lambda x : get_user_mapping(x))
df['participant_age'] = df['participant_age'].fillna("NA")
df['participant_age_map'] = df['participant_age'].apply(lambda x : get_user_mapping(x))
df['participant_gender'] = df['participant_gender'].fillna("NA")
df['participant_gender_map'] = df['participant_gender'].apply(lambda x : get_user_mapping(x))

df[['participant_type_map', 'participant_age_map', 'participant_gender_map']].head()
```

	participant_type_map	participant_age_map	participant_gender_map
0	{'0': 'Victim', '1': 'Victim', '2': 'Victim', ...	{'0': '20'}	{'0': 'Male', '1': 'Male', '3': 'Male', '4': '...
1	{'0': 'Victim', '1': 'Victim', '2': 'Victim', ...	{'0': '20'}	{'0': 'Male'}
2	{'0': 'Subject-Suspect', '1': 'Subject-Suspect...	{'0': '25', '1': '31', '2': '33', '3': '34', '...	{'0': 'Male', '1': 'Male', '2': 'Male', '3': '...
3	{'0': 'Victim', '1': 'Victim', '2': 'Victim', ...	{'0': '29', '1': '33', '2': '56', '3': '33'}	{'0': 'Female', '1': 'Male', '2': 'Male', '3':...
4	{'0': 'Victim', '1': 'Victim', '2': 'Victim', ...	{'0': '18', '1': '46', '2': '14', '3': '47'}	{'0': 'Female', '1': 'Male', '2': 'Male', '3':...

Figure 8: Column reformatting

3.2. Detecting Outliers

In this part outliers were detected. Shaded area indicates the standard deviation of incidents between 2014 and 2017, and orange circles show the anomalies of our dataset. Most of those lower part anomalies are for month February, where there is, in general, a low rate of incidents due to shorter, less amount of days and upper part outliers are for months July and May. Here, during July and May there are high rates of incidents due to holidays and longer summer days.

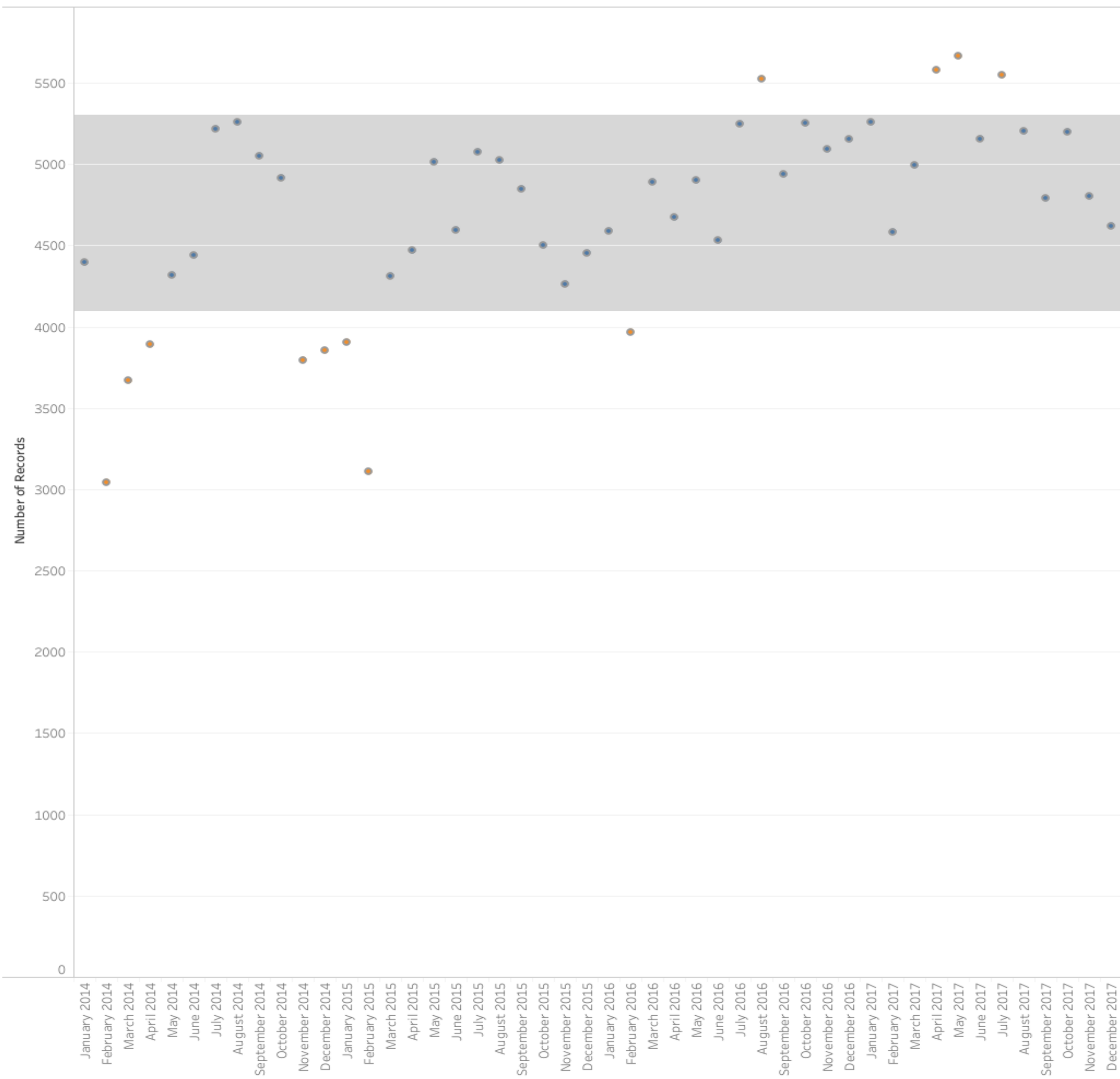

Figure 9: Outliers

4. ANALYSIS

Several analyses have been implemented such as time and location related analysis, gender and age-related analysis, and gun related analysis. The purpose is to get some insights about the data as well as derive some predictions from visual outcomes.

4.1. Time Related Analysis of Gun Violence

Time related analysis is a significant part of research since the gathered data contains a lot of time-based information between 2014 and 2017. Time related analysis helps to analyze past data and gives some insights about future incidents.

4.1.1. Number of Incidents by Year

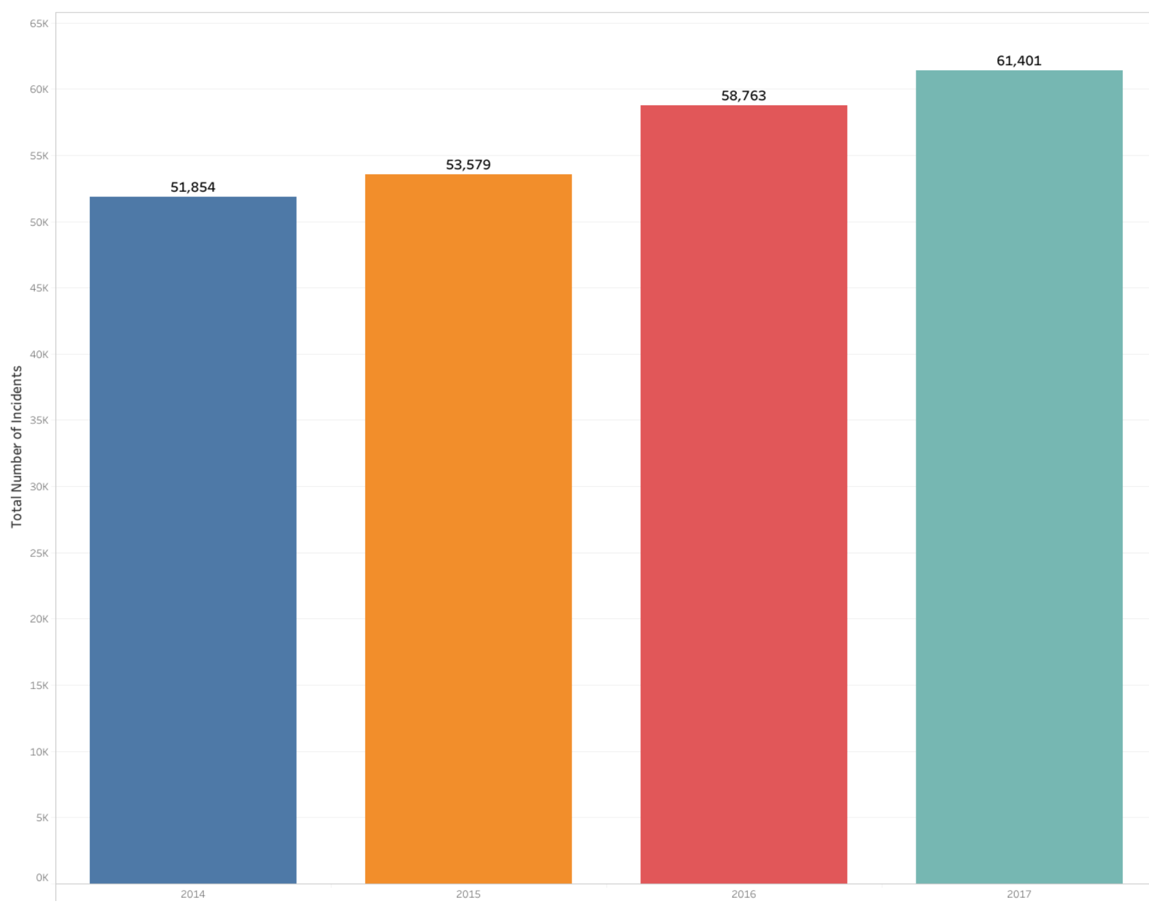

Figure 10: Total number of incidents of each year

Figure 10 shows that the number of gun violence incidents increases each year with an average rate of 6%. There are about 9.7% rise from 2015 to 2016, this is due to a severe mass shooting in one of the nightclubs in Orlando, FL where 49 people were killed and at least 53 were wounded according to CNN (2016) [3].

The following Figure 11 shows the number of victims for years 2014-2017.

Figure 11: Total number of victims for each year

4.1.2. Number of Incidents by Month

Monthly based graph shows that the highest rates of shooting happens during summer time. According to Giffords Law Center, during warm climate individuals want to get more involved in outdoor activities, also getting a vacation during the summer is more likely as well as there are less summer programs to get youth involved which lead to a spike of mass violence rate during hot summer (2018) [4].

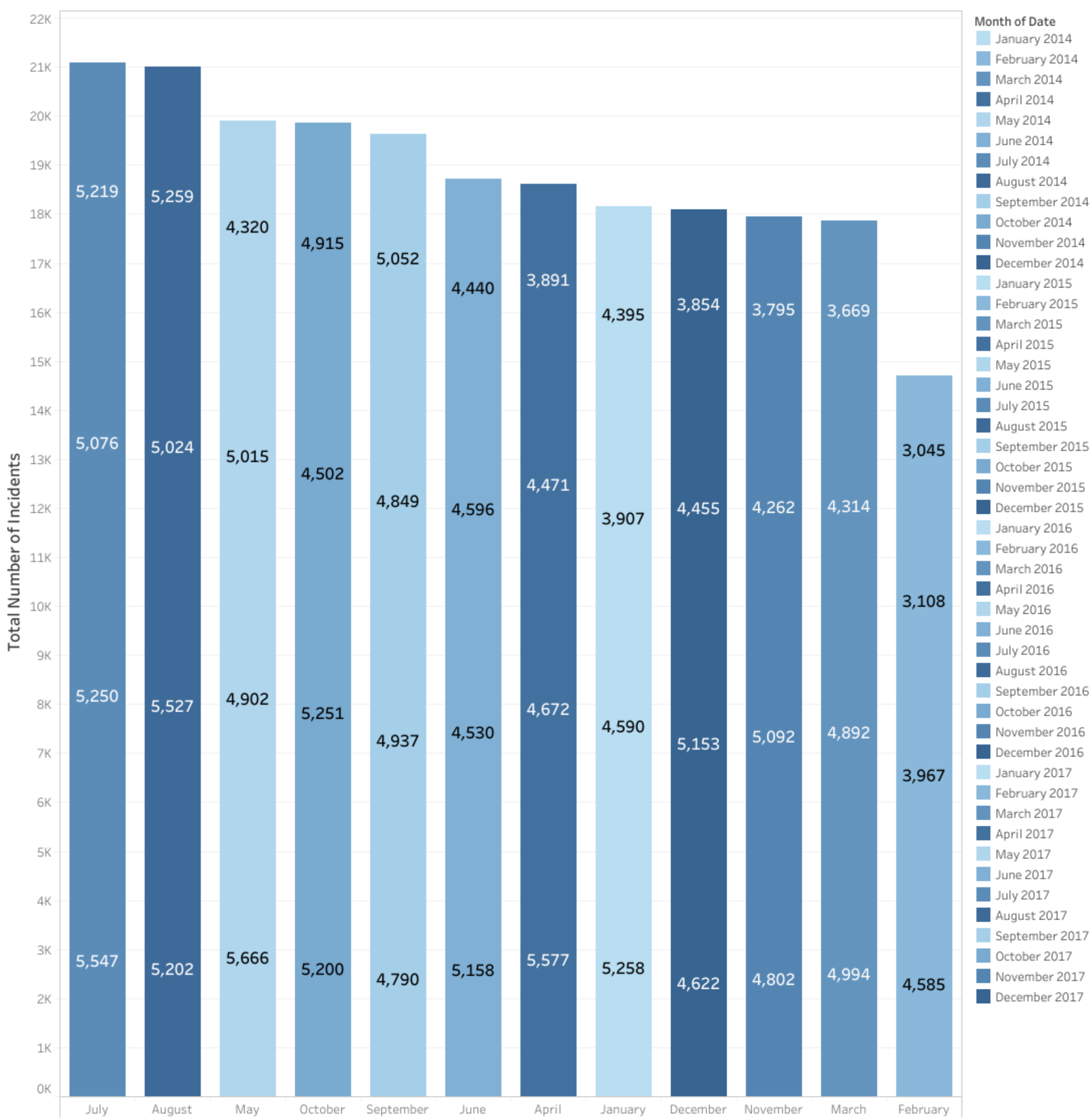

Figure 12: Total number of incidents for each month

The following graph indicates the total number of victims by month where the seasonality of the data can be detected such as February of each year has the lowest rate of loss, whereas July and August rates of each year are at the top. Despite rates are rising each year, from the following Figure 13 we can detect peaks at certain days of the months. These peaks mostly coincide with holiday dates.

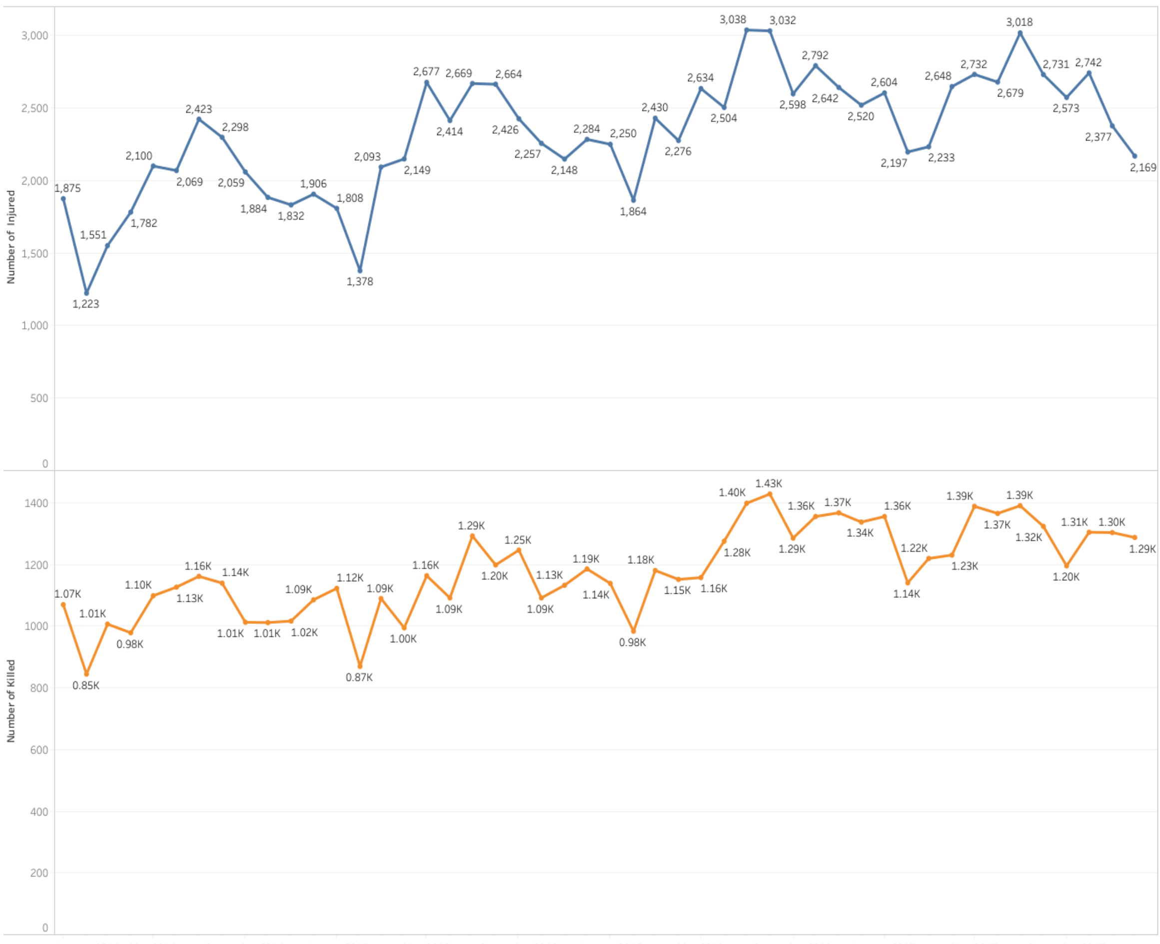

Figure 13: Total number of victims for each month

4.1.3. Number of Victims by Quarter

When observing by quarters of each year, highest rates occur on first month, late spring and mid-summer of every year and the lowest rates occur in late winter and early spring periods. Seasonality of incidents is also seen here, where peaks are in summer.

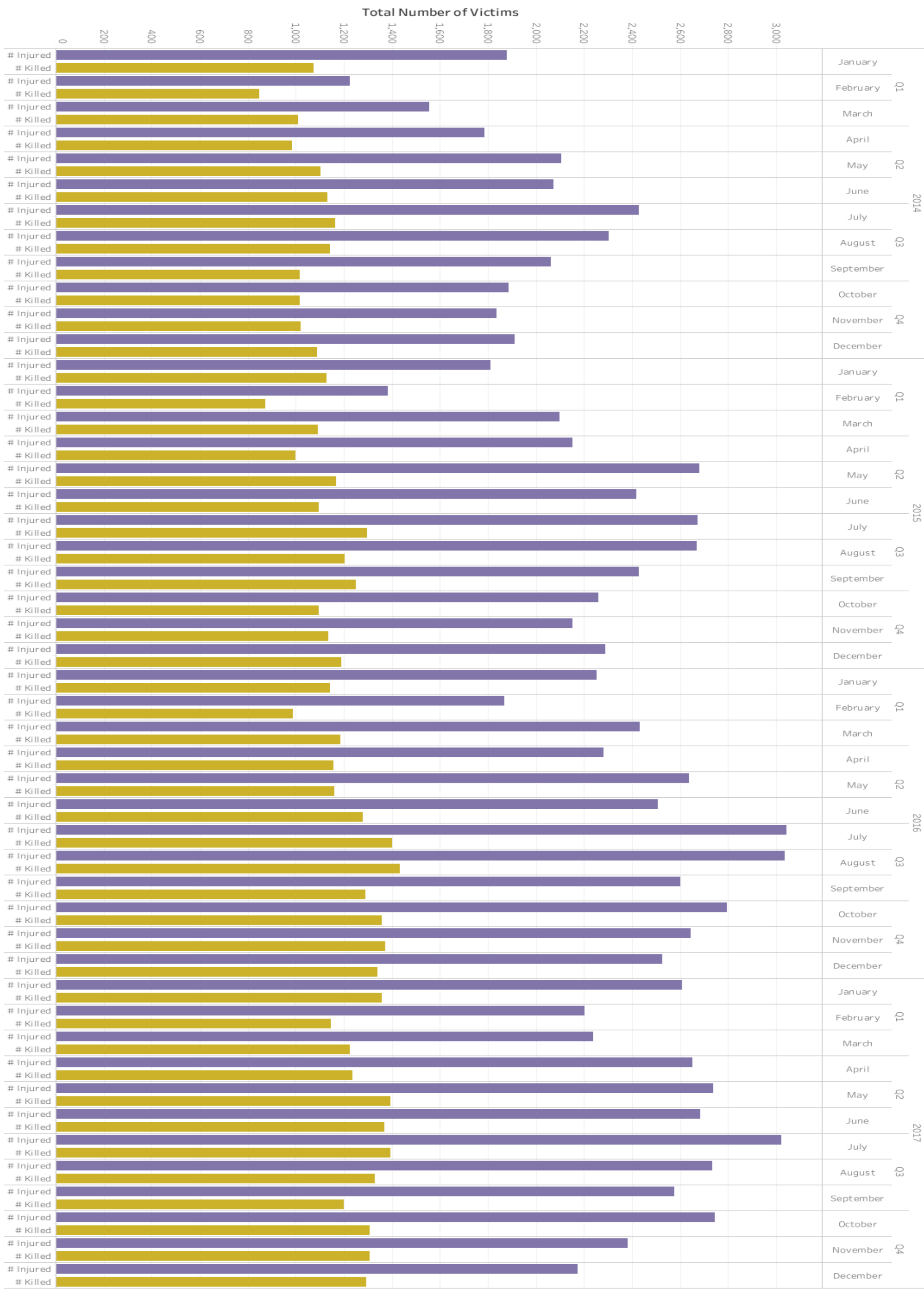

Figure 14: Total number of victims by quarter

4.1.4. Number of Incidents by Weekday

Weekdays graph shows that during weekends violence rates are higher with respect to other days. The reason for that might be easy accessibility to every location, weekend public activities like visiting parks or any other public places or even going to pubs, bars or nightclubs consequently leads to the alcohol consumption which might be the reason for violence afterwards.

Figure 15: Total number of incidents by weekdays

The following figure shows the total number of victims by weekdays for 2014-2017 years. As it can be observed here, highest rates are on weekends followed by Mondays' and Fridays' rates.

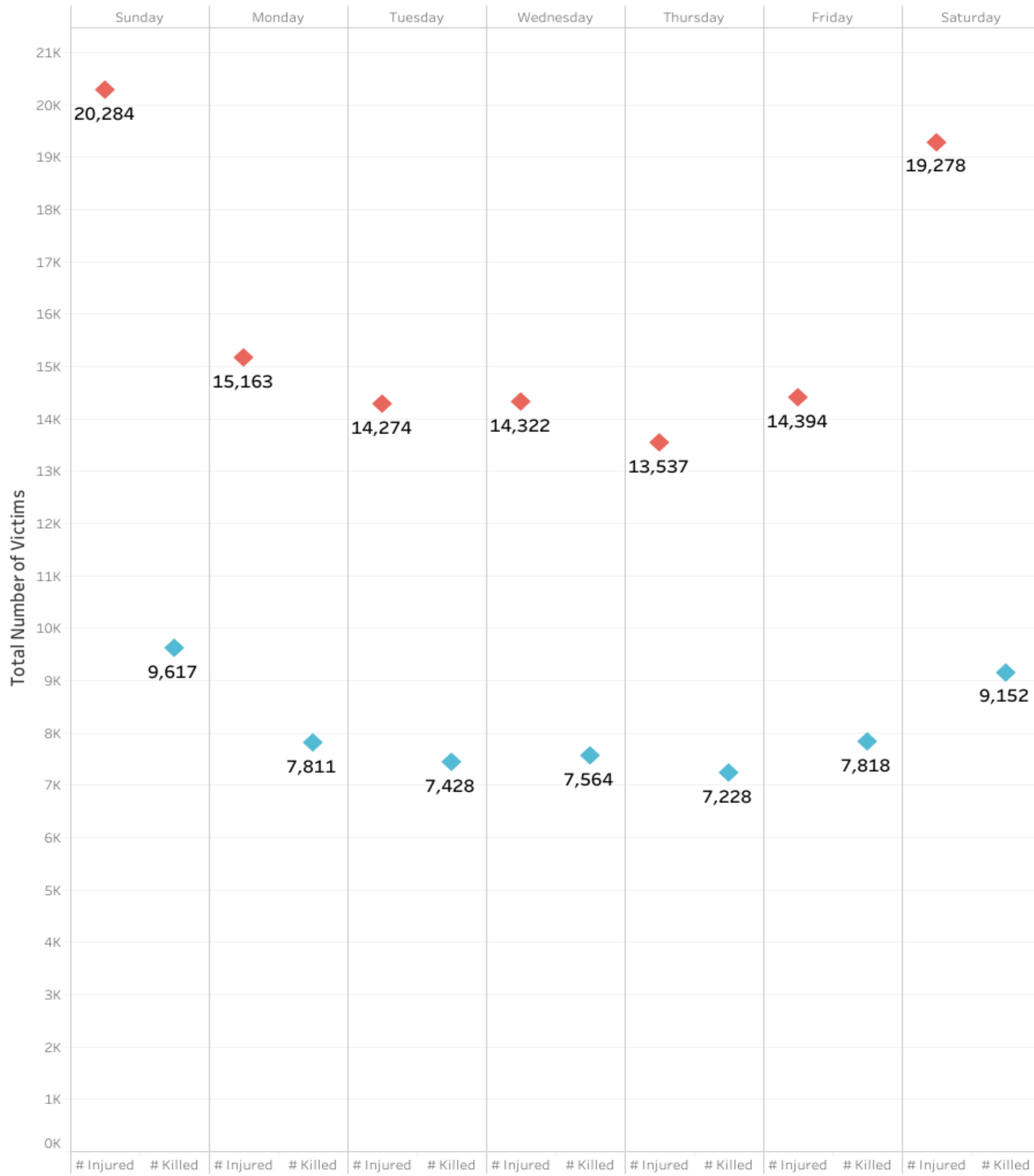

Figure 16: Total number of victims by weekdays

4.2. Number of Incidents During Holidays

Scrutinizing official holidays in the United States and gun violence incidents. As it is shown below, during holidays violence rates spikes significantly. Dr. Sherry Hamby, the founding editor of the Psychological Association journal "*Psychology of Violence*", states: "Crimes spike on days off--so, all year round, crime rates are much higher on Friday and Saturday and is higher at night than in the morning, so any increase during the holidays might be simply due to more idle time and more drinking and other drug use. Moreover, some people experience depression during the holidays because it can highlight feelings of loneliness or create pressure to spend time with family, and for some people that means spending time in high-conflict situations". Shortly, she points out that the holidays season are stressful, thus the stress increases the risk of crime (2017) [5].

During holidays it is easy to access the crowd or break into someone's property while everyone is in a holiday bustle not paying enough attention for own safety. For instance, disastrous epidemic of gun violence is in Chicago. Once, shootings in the summer months at the Windy City reached dramatic spikes during weekends on holiday — Memorial Day, Independence Day, and Labor Day (2016) [4]. From here, we can say that holidays are target days for gun violence incidents where people are more stressed with holiday season bustle.

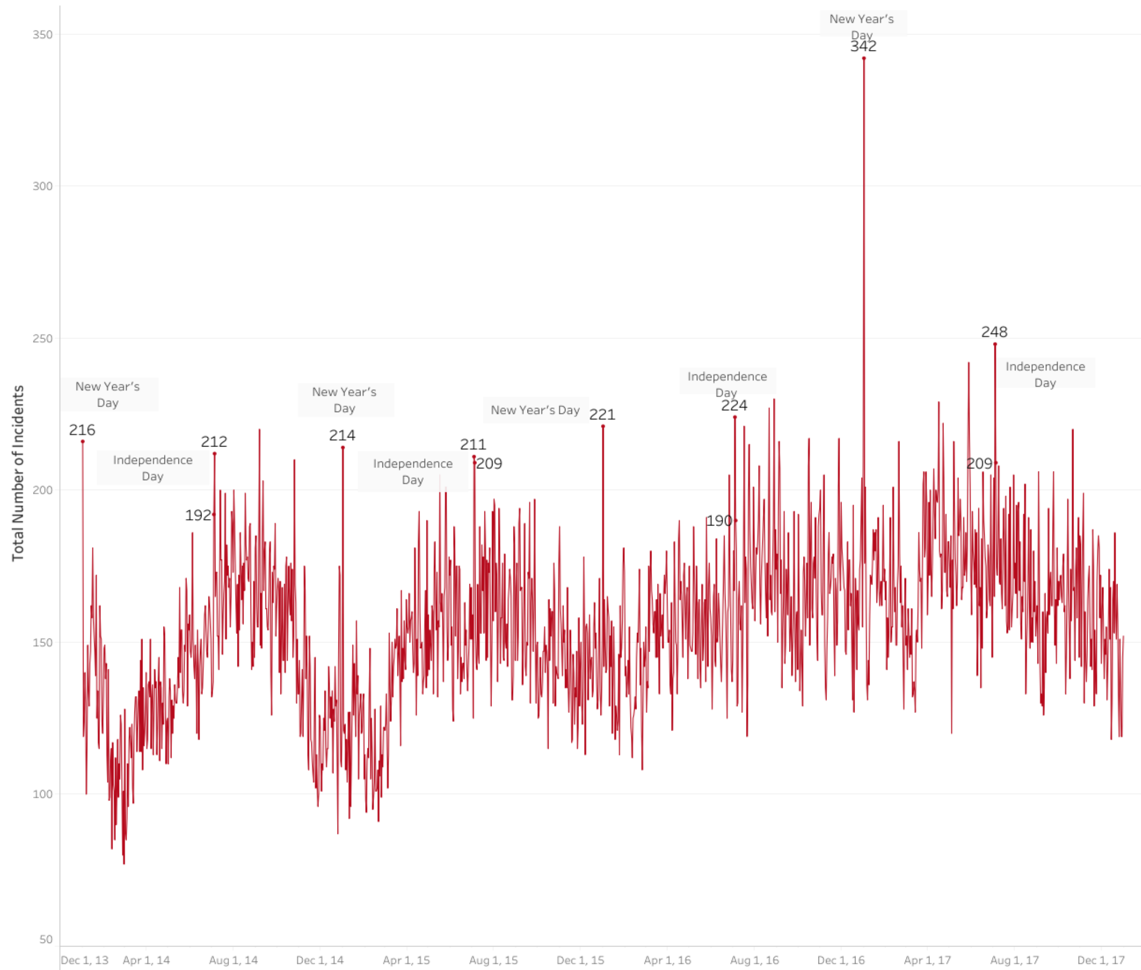

Figure 17: Gun violence incidents during holidays

4.3. Location Related Analysis of Gun Violence

4.3.1. Number of Incidents by State

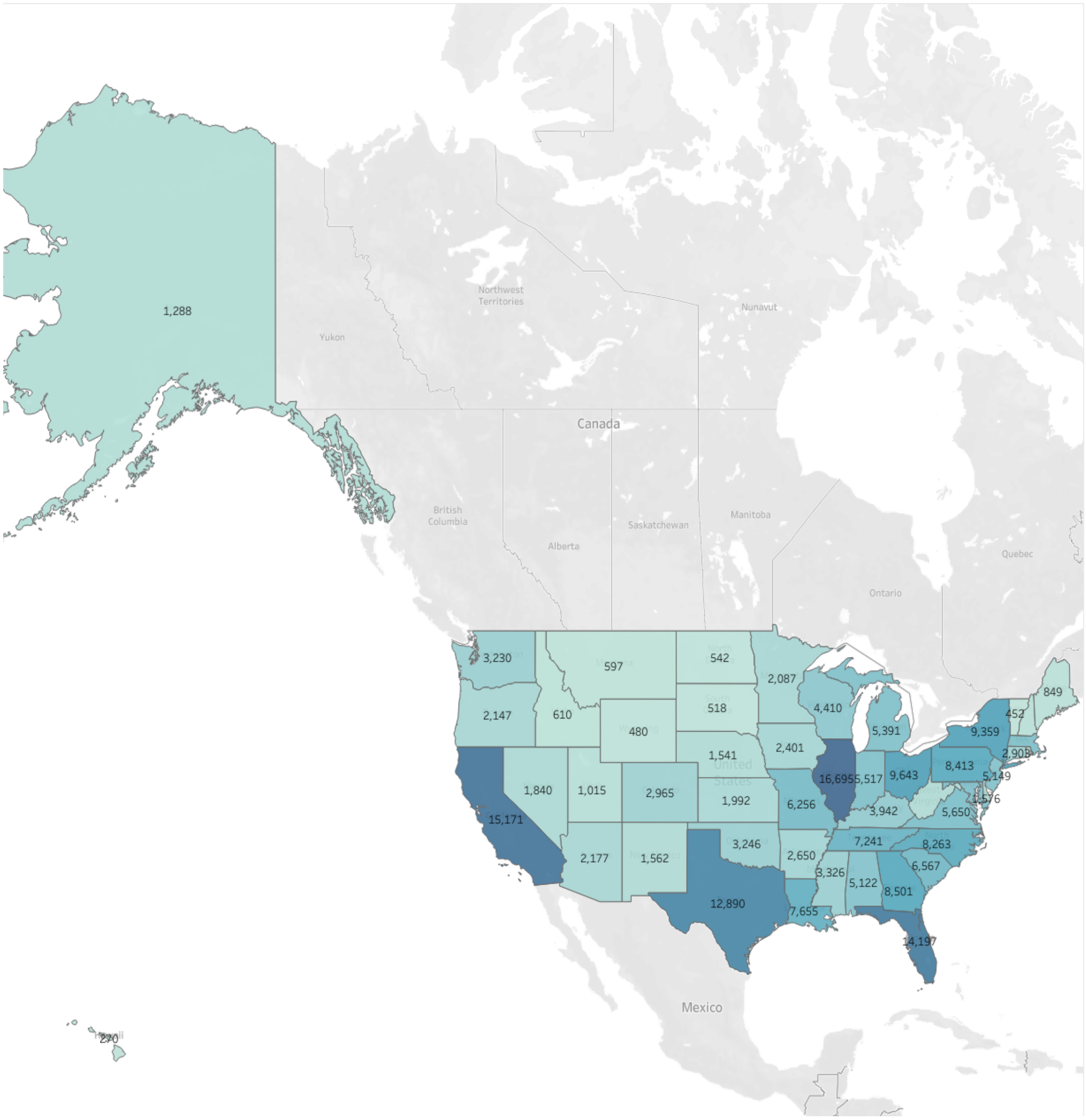

Figure 18: Number of incidents by States

The map colors States according to number of gun violence rates. The darker part of the map, which is south-east, demonstrates that the United States have high gun violence rate. From the above map, State Illinois was recorded as the state with the largest number of occurrences of gun violence in these three years, the amount being 16,600. California follows with approximately 15,000 reported incidents, Florida with 14,000 incidents, and Texas with 12,800 incidents of gun violence. Following bar chart demonstrates gun violence numbers for all states more clearly.

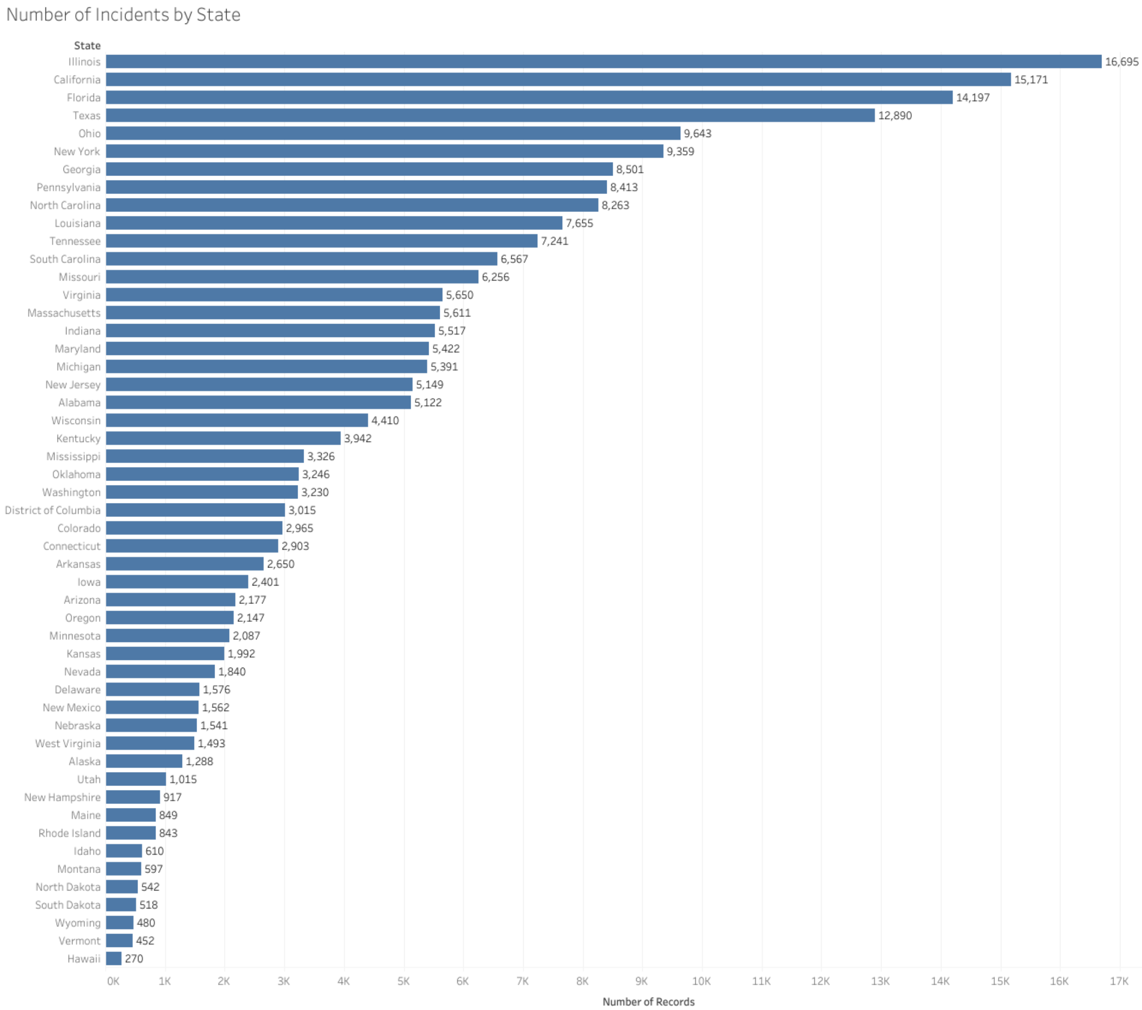

Figure 19: Number of incidents by States (from highest to lowest)

Top states with the highest numbers of incidents such as California, Texas, New York are bigger and more crowded states as expected. Here, Illinois State is an exception. In the following part of the report we will investigate more about the reasons behind how Illinois is, specifically Chicago, addressing gun violence. Besides that, top states show that the southern part of the country have more incidents. We may generalize that people who live in hotter places are inclined to commit an offense with gun.

4.3.2. Top 10 States and Cities with Highest Number of Incidents

Top ten states and cities are demonstrated with separate charts below. In top cities, Chicago takes the lead. This encouraged us to research more about it.

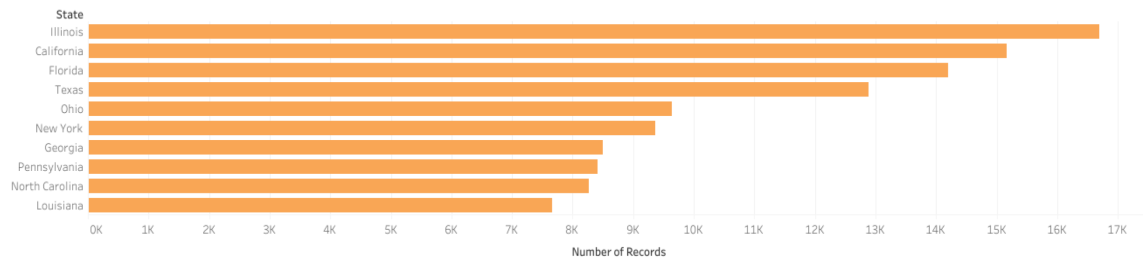

Figure 20: Top 10 States with highest number of gun violence

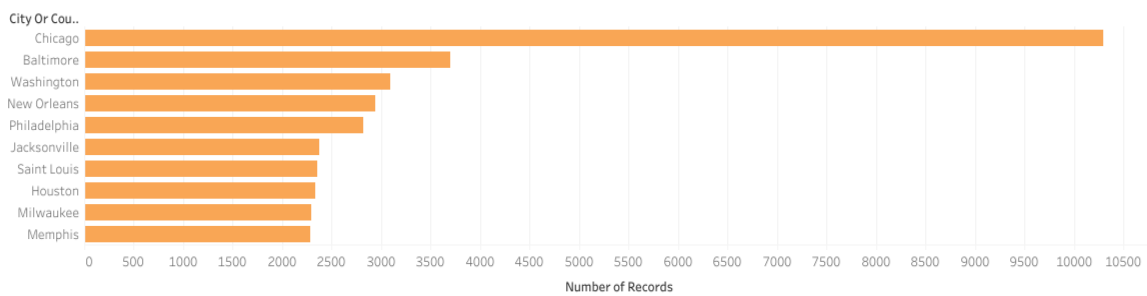

Figure 21: Top 10 cities with highest number of gun violence

The murder clearance rate fell from over 70% in 1991 to below 34% in 2011, according to the Chicago Police Department's 2011 Homicide Report. Former Chicago Police Supt. Garry McCarthy said a pervasive "no-snitch code" on the street remains the biggest reason why more assassinations are not solved in Chicago, adding, "We're not doing well because we're not getting cooperation" [6].

Chicago's murder rates differ widely depending on the neighborhood involved. Many South Side neighborhoods are poor, lacking instructional resources, predominantly African American, and street gangs infested. For instance, Englewood's South Side neighborhoods and Austin's West have murder rates ten times greater than other areas of the town.

For those illegally discovered in possession of a firearm, Chicago was criticized for relatively light sentencing guidelines. Most individuals convicted of illegal weapon ownership receive a minimum sentence, one year, an assessment discovered in the Chicago Sun-Times, and serve less than half of the sentence for excellent conduct and pre-trial imprisonment. For violent criminals discovered in possession of a gun, the minimum sentence is two years. Those accused of easy weapon possession had four previous arrests on average. Those accused of gun ownership by a felon had ten previous arrests on average. Next chapters studied the states and cities with the largest occurrences of gun violence per 100,000 individuals [7].

4.3.3. Number of Incidents Per 100,000 People by State and City

We had to look for a good source to be able to relate the numbers above to the sizes of the population. On the U.S. Census Bureau website, we found what we needed. After, created calculated field for population, we made bar charts for state and city wise.

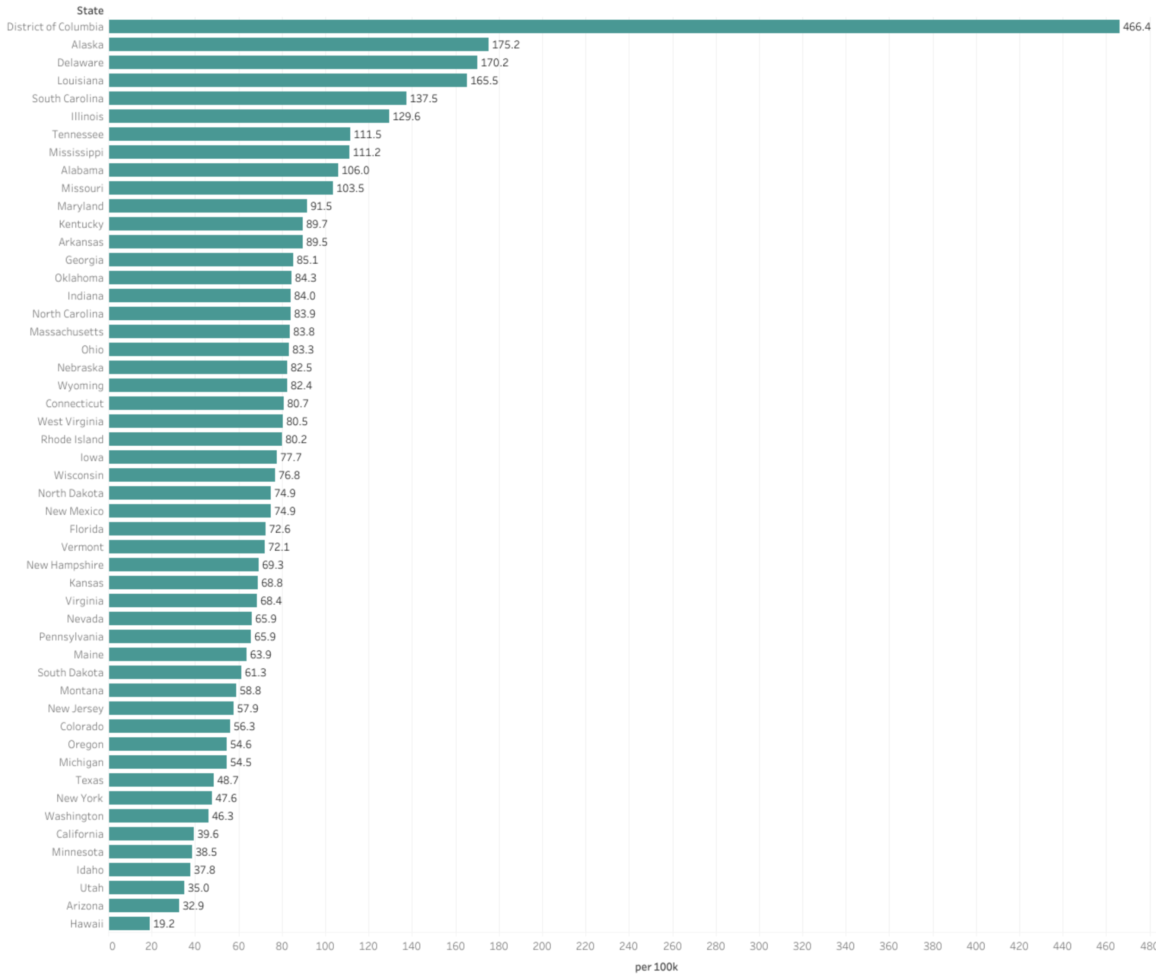

Figure 22: Number of incidents per 100K people by State

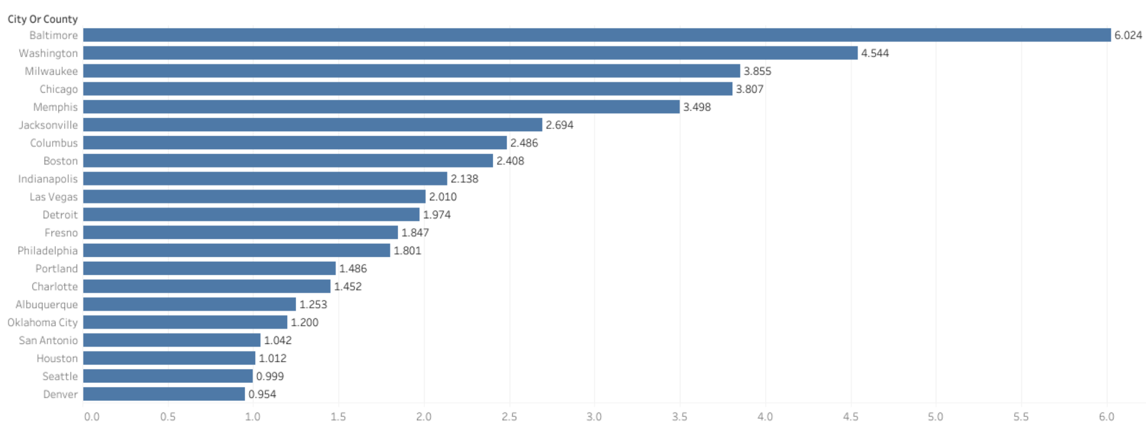

Figure 23: Number of incidents per 100K people by city

As indicated, crowded states and cities have more gun violence records. Even though Illinois state had most numbered record, when we proportioned with population

Illinois step down to sixth place. District of Columbia has very high rate of gun violence by a wide margin to other states. Nearly, 0.005% of people are involved to gun violence. Surprisingly, Alaska and Delaware are following the District of Columbia.

When we looked at the cities Baltimore, Washington, Milwaukee, and Chicago are leading the list. We may assert that the cities with the highest number of records do not change a lot when its proportioned to per k population.

4.3.4. Total Number of Victim

We would like to analyze how many people are victims who are injured or killed in the incident. As it is shown, top four states with highest number of victims are Illinois has 16,094, California has 12,285, Texas has 10,403 and Florida has 10,277. From previous analysis, it can be said that these states are also densely populated and have a higher crime rate with respect to other states, however, Illinois is an exception in this case by being the least populated state among those four states.

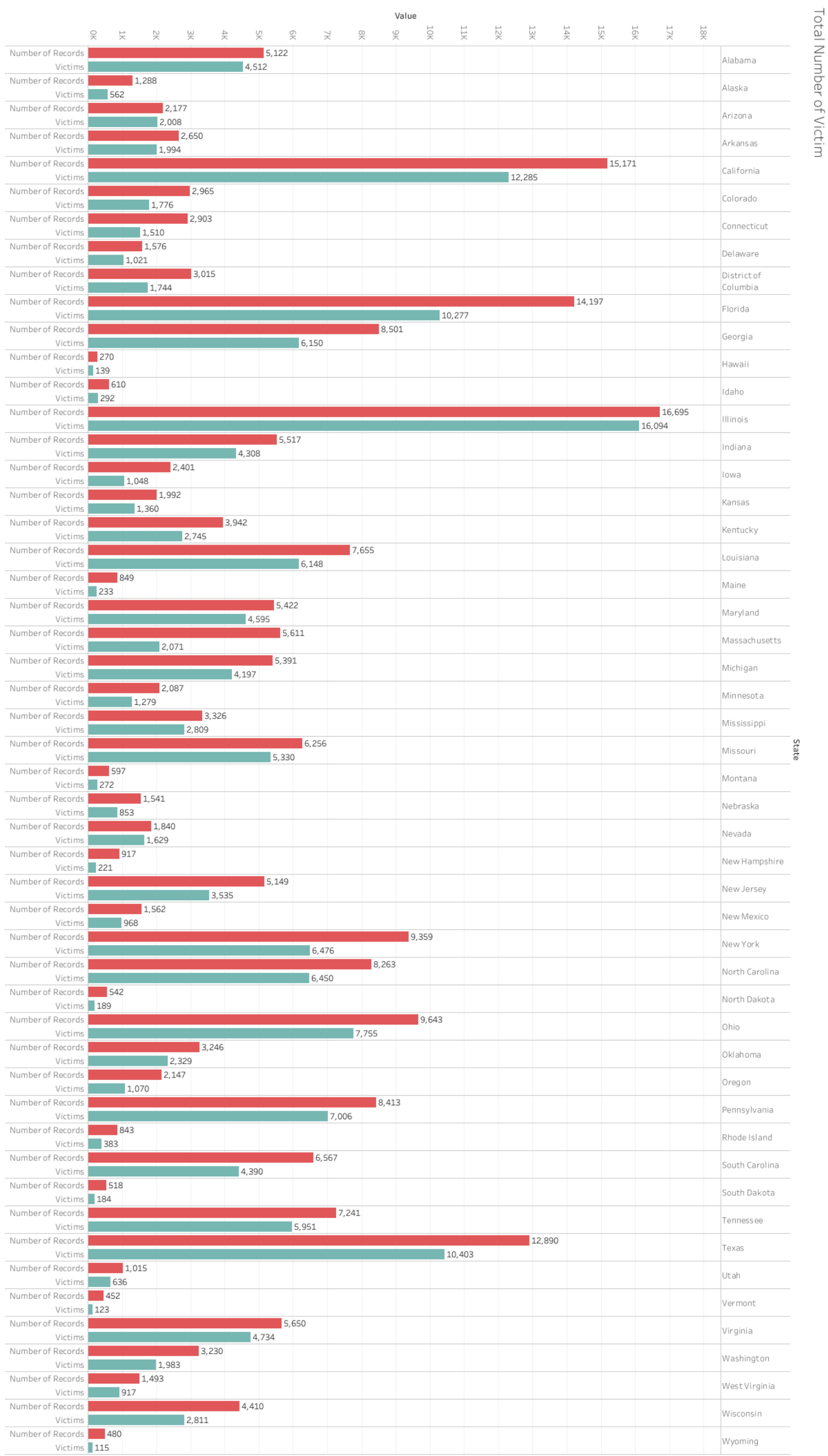

Figure 24: Total number of victims by States

4.4. Number of Guns Involved in Gun Violence Incidents

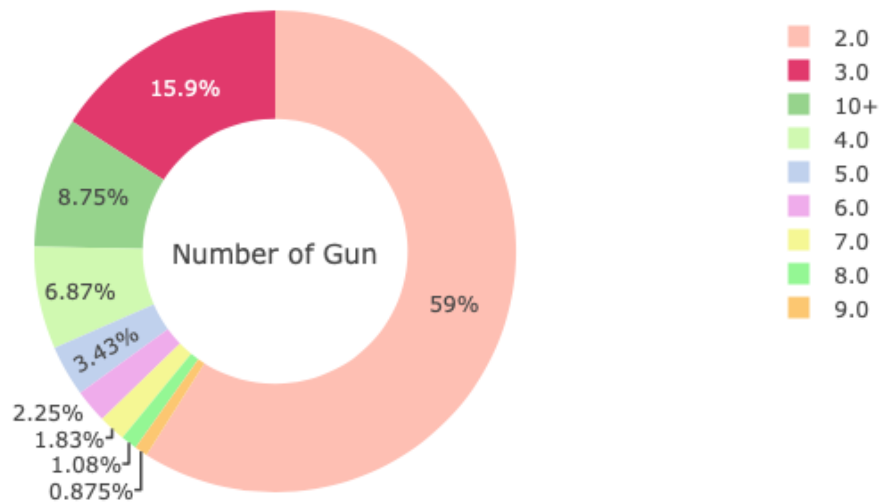

Figure 25: Number of guns involved

This chart shows the number of involved guns, where the amount is greater than 1.

4.5. Gender Related Analysis of Gun Violence

From below chart, it can be said that males are prone to incur to gun violence incidents more than females. More than twice as many females were shot and killed by their spouse or intimate acquaintance than were murdered by strangers using guns (1992) [8].

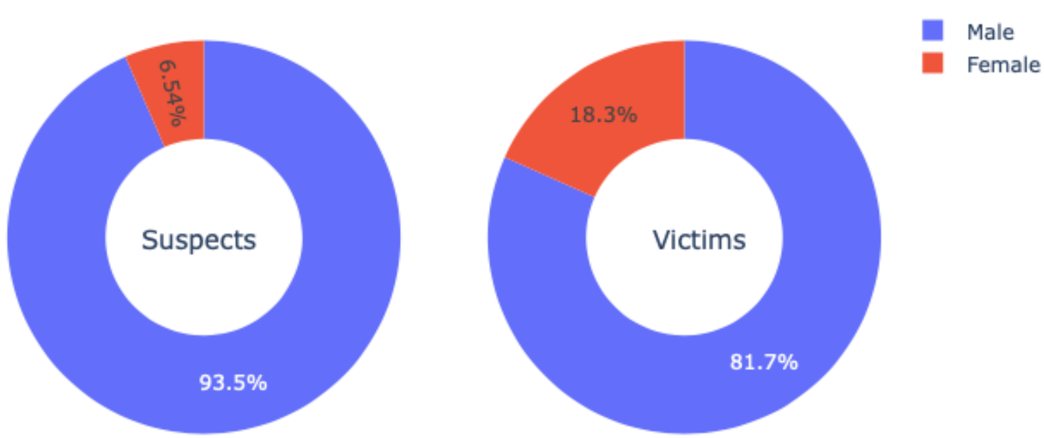

Figure 26: Gender related analysis

4.6. Age Related Analysis of Gun Violence

4.6.1. Suspects' Age Distribution

Figure 27 shows that the age range of suspects are between 17-26 which is above 4000. One the reasons of this age group being involved might be gang participations, showing off to peers, and other drug and gun related dealerships, where youth is more vulnerable to such involvements than other age groups.

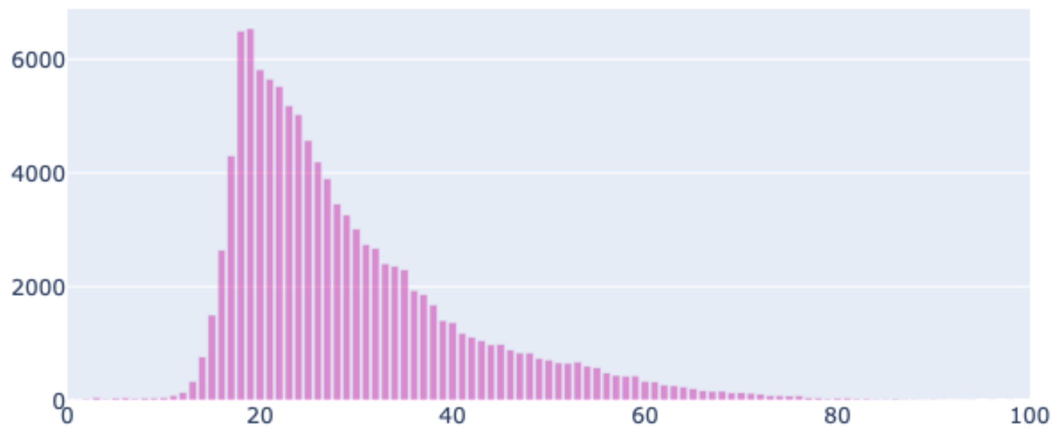

Figure 27: Age distribution of suspects

4.6.2. Victims' Age Distribution

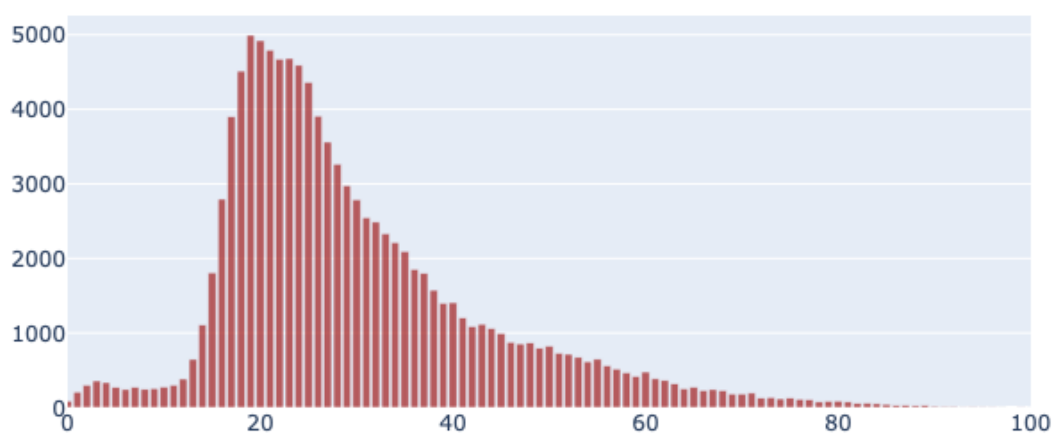

Figure 28: Age distribution of victims

Similar to the previous analysis, almost same age group of 17-28.

4.7. Deadliest Gun Violence Incidents

The top ten deadliest gun violence incidents for the period of 2014-2017. As it is shown below, the top one happened in Las Vegas state Nevada with a total amount of 548 victims which we found out lately during our research and added to the dataset. However, we decided to move on without using this record in our dataset since it would be an outlier and will not significantly affect overall outcome of our analysis. This incident was one of the deadliest mass shootings in the latest 10 years.

```
# Add missing data manually
missing_row = ['sban_1', '2017-10-01', 'Nevada', 'Las Vegas', 'Mandalay Bay 3950 Blvd S', 59, 489, 'https://en.wikipedi
               '-115.171667', 47, 'Route 91 Harvest Festiva; concert, open fire from 32nd floor. 47 guns seized; TOTAL:
df.loc[len(df)] = missing_row

# Add total loss column
df['loss'] = df['n_killed'] + df['n_injured']

#List top 10 serious gun violence incidents
df1 = df.sort_values(['loss'], ascending=[False])
df1[['date', 'state', 'city_or_county', 'address', 'n_killed', 'n_injured']].head(10)
```

	date	state	city_or_county	address	n_killed	n_injured
239677	2017-10-01	Nevada	Las Vegas	Mandalay Bay 3950 Blvd S	59	489
130448	2016-06-12	Florida	Orlando	1912 S Orange Avenue	50	53
217151	2017-11-05	Texas	Sutherland Springs	216 4th St	27	20
101531	2015-12-02	California	San Bernardino	1365 South Waterman Avenue	16	19
232745	2018-02-14	Florida	Pompano Beach (Parkland)	5901 Pine Island Rd	17	17
70511	2015-05-17	Texas	Waco	4671 S Jack Kultgen Fwy	9	18
195845	2017-07-01	Arkansas	Little Rock	220 W 6th St	0	25
137328	2016-07-25	Florida	Fort Myers	3580 Evans Ave	2	19
11566	2014-04-02	Texas	Fort Hood	Motor Pool Road and Tank Destroyer Boulevard	4	16
92624	2015-10-01	Oregon	Roseburg	1140 Umpqua College Rd	10	9

Figure 29: Top 10 deadliest gun violence incidents

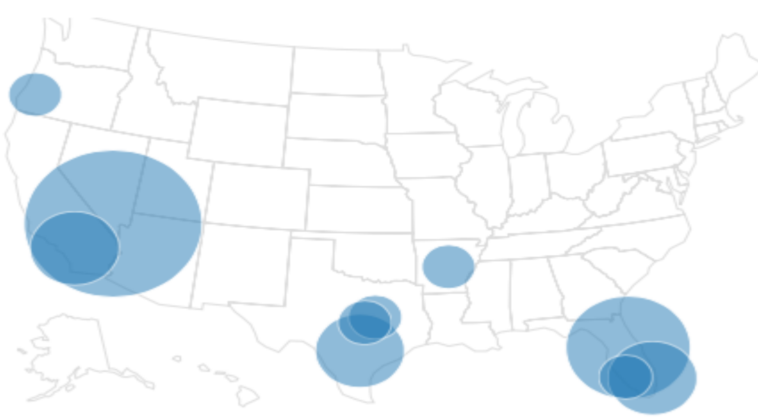

Figure 30: Top 10 deadliest gun violence incidents by city

5. RESULTS, CONCLUSIONS AND RECOMMENDADTIONS

The results of this study have shown that the rates of violence increase year by year with an average rate of 6%, not including the outliers, and reach peaks on months of July, August and May (hot weather periods) as well as official holidays. Top rated states are Illinois, California and Florida and top cities are Chicago, Baltimore and Washington. Most involved age group is between 16-28 and most involved gender is male almost 87.6%.

On the basis of the findings, several conclusions concerning the gun violence incidents can be drawn. The findings of this research have shown that holidays are hotspots for gun violence, larger the cities harder to regulate it, warmer and longer summer days higher the risk of gun attacks.

Some recommendations that might help to prevent gun-related violence are thorough background checks and tough permit/license laws till purchase, inspection for gun storages safety locks at home while purchasing, group violence intervention programs identify the most probable people to engage in gun violence and provide possibilities and resources to break the cycle of violence. In combination with smart gun laws, Richmond, CA, introduced this approach to cut shootings by an amazing 66 percent in just seven years (n.d.) [4].

REFERENCES

[1] U.S. Const. amend. II (ratified 1791).
[2] Gun Violence Archive. (n.d.). Retrieved from https://www.gunviolencearchive.org/ https://www2.census.gov/programs-surveys/popest/datasets/2010-2017/
[3] Fantz, A., Karimi, F., McLaughlin, E. C., & Ellis, R. (2016, June 13). Orlando shooting: 49 killed, shooter pledged ISIS allegiance. Retrieved from https://www.cnn.com/2016/06/12/us/orlando-nightclub-shooting/index.html
[4] Shootings Spike in Summer Months. (n.d.). Retrieved from https://lawcenter.giffords.org/resources/publications/shootings-spike-in-summer-months/
[5] Tron, G. (2017, November 17). It's the Most Dangerous Time of The Year: Why do Crimes Increase Around the Holidays? Retrieved from https://www.oxygen.com/homicide-for-the-holidays/blogs/its-the-most-dangerous-time-of-the-year-why-do-crimes-increase
[6] "Chicago Murder Clearance Rate Worst in More Than 2 Decades - Chicago". Dnainfo.com. 2013-01-04. Archived from the original on 2015-12-22. Retrieved 2015-12-17.
[7] Chicago Sun-Times. (2014). Gun shy: Lighter sentences in Cook County fuel lock 'em up debate. [online] Available at: https://chicago.suntimes.com/2014/11/29/18467208/gun-shy-lighter-sentences-in-cook-county-fuel-lock-em-up-debate [Accessed 5 Aug. 2019].
[8] Kellermann, A. L., & Mercy, J. A. (1992). Men, women, and murder: Gender-specific differences in rates of fatal violence and victimization. The Journal of Trauma: Injury, Infection, and Critical Care, 33(1), 1-5. doi:10.1097/00005373-199207000-00001