Final Project: Gun Violence in the United States from 2014-2017

Leah Fawzi

12/16/2018

Overview: My project centers around a csv file that I renamed "gun_violence.csv" that I obtained from Kaggle (https://www.kaggle.com/jameslko/gun-violence-data). The file contains data for all recorded gun violence incidents in the U.S. between January 2013 and March 2018. It is important to note that "gun violence incident" refers to any incident in which shots were fired within the United States, and does not necessarily mean anyone was injured or killed.

It contains columns such as: - date (Date of crime) - state (State of crime) - city_or_county (City/County of crime) - n_killed (Number of people killed) - latitude (Location of the incident) - longitude (Location of the incident) - participant_age_group (Age group of participant(s) at the time crime) - incident_characteristics (Characteristics of the Incident)

In the notes about the dataset, it states "The list of incidents from 2013 is not exhaustive; only 279 incidents from that year were catalogued." Additionally, because the data for 2018 is only until March, it is not complete. Thus, I remove 2013 and 2018's data from my data set so that my analysis is accurate.

Main Question: How has the nature of gun violence in the United States changed from 2014 to 2017 and what factors, such as location, date, and motive, affect it?

In addressing the above question, I utilize the libraries listed below:

```r
library(lubridate, quietly=T)
```

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date
```

```r
library(ggplot2, quietly=T)
library(magrittr, quietly=T)
library(dplyr, quietly=T)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:lubridate':
##
##     intersect, setdiff, union

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tidyr, quietly=T)
```

```
##
```

```
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:magrittr':
##
##     extract
```

```
library(ggplot2, quietly=T)
library(maps, quietly=T)
library(reshape, quietly=T)
```

```
##
## Attaching package: 'reshape'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, smiths
```

```
## The following object is masked from 'package:dplyr':
##
##     rename
```

```
## The following object is masked from 'package:lubridate':
##
##     stamp
```

Cleaning the Data Set: In cleaning the data set, I filter to exclude 2013 and 2018's data for reasons I mentioned earlier. Then I convert the date column to the more usable "year", "month", "day", and "weekday" columns using the lubridate package.

```
df <- function(file){
  df <- read.csv(file, header=T, stringsAsFactors = F)
  df_filt <- df[year(df$date)=="2014"|year(df$date)=="2015"|year(df$date)=="2016"| year(df$date)=="2017
  df_filt$year <- year(df_filt$date)
  df_filt$month <- lubridate::month(df_filt$date, label=TRUE)
  df_filt$day <- day(df_filt$date)
  df_filt$weekday <- lubridate::wday(df_filt$date, label=TRUE)
  return(df_filt)
}

print(head(df("gun_violence.csv")$year))
```

```
## [1] 2014 2014 2014 2014 2014 2014
```

```
print(head(df("gun_violence.csv")$month))
```

```
## [1] Jan Jan Jan Jan Jan Jan
## 12 Levels: Jan < Feb < Mar < Apr < May < Jun < Jul < Aug < Sep < ... < Dec
```

```
print(head(df("gun_violence.csv")$day))
```

```
## [1] 1 1 1 1 1 1
```

```
print(head(df("gun_violence.csv")$weekday))
```

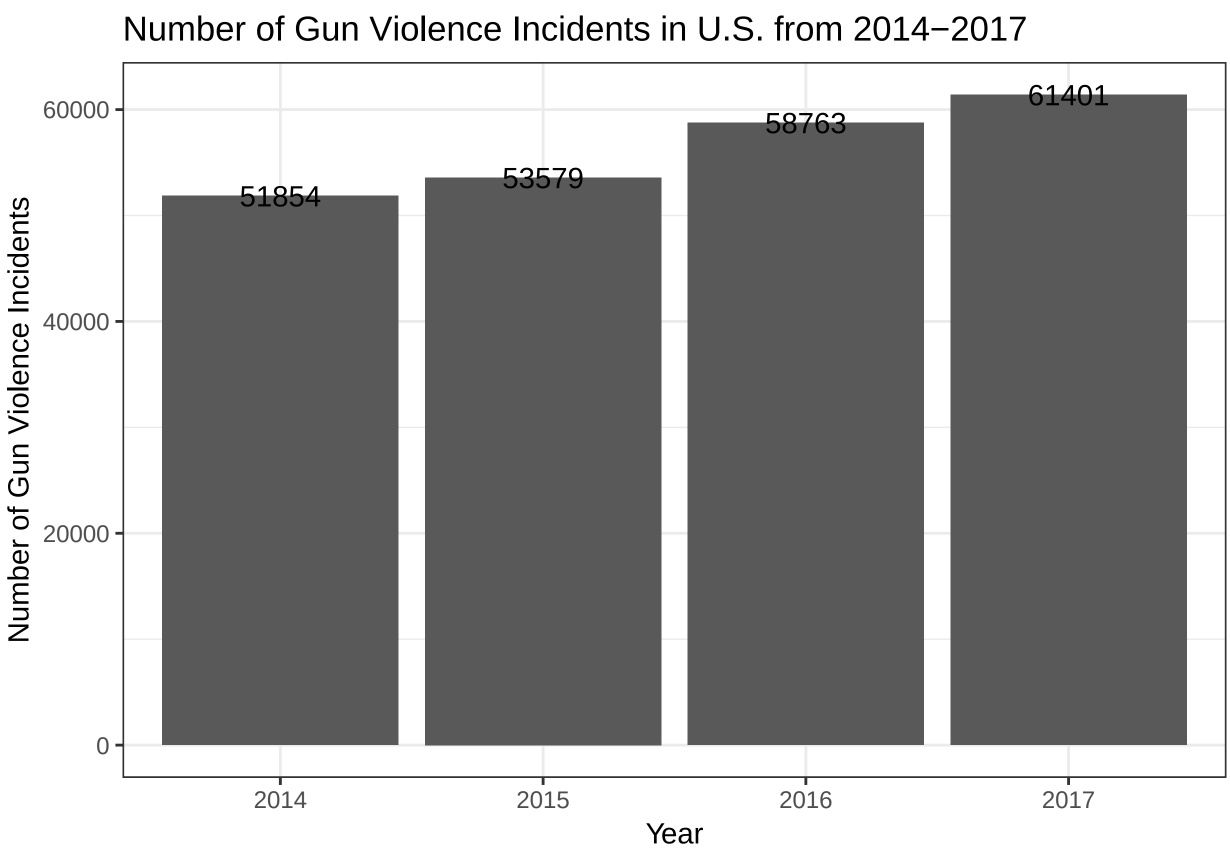
```
## [1] Wed Wed Wed Wed Wed Wed
## Levels: Sun < Mon < Tue < Wed < Thu < Fri < Sat
```

Note that the above calls of the function show the same value 6 times because there were clearly at least 6 gun violence incidents on January 1, 2014. Nevertheless, we have confirmed that this function has cleaned the data set.

General Analysis:

First, I write a function to generate a data frame and a barplot that show the number of gun violence incidents by year, from 2014-2017.

```
shootings_per_year <- function(df){
  df_years <- table(df$year) %>% as.data.frame

  years <- df_years$Var1
  freq <- df_years$Freq
  p <- ggplot(data=df_years, aes(x=years, y=freq))
  p <- p + geom_bar(stat="identity") + xlab("Year") + ylab("Number of Gun Violence Incidents")
  p <- p + ggtitle("Number of Gun Violence Incidents in U.S. from 2014-2017")
  p <- p + theme(plot.title = element_text(hjust = 0.5)) + geom_text(aes(label=Freq)) + theme_bw()
  print(p)

  return(df_years)
}

print(shootings_per_year(df("gun_violence.csv")))
```

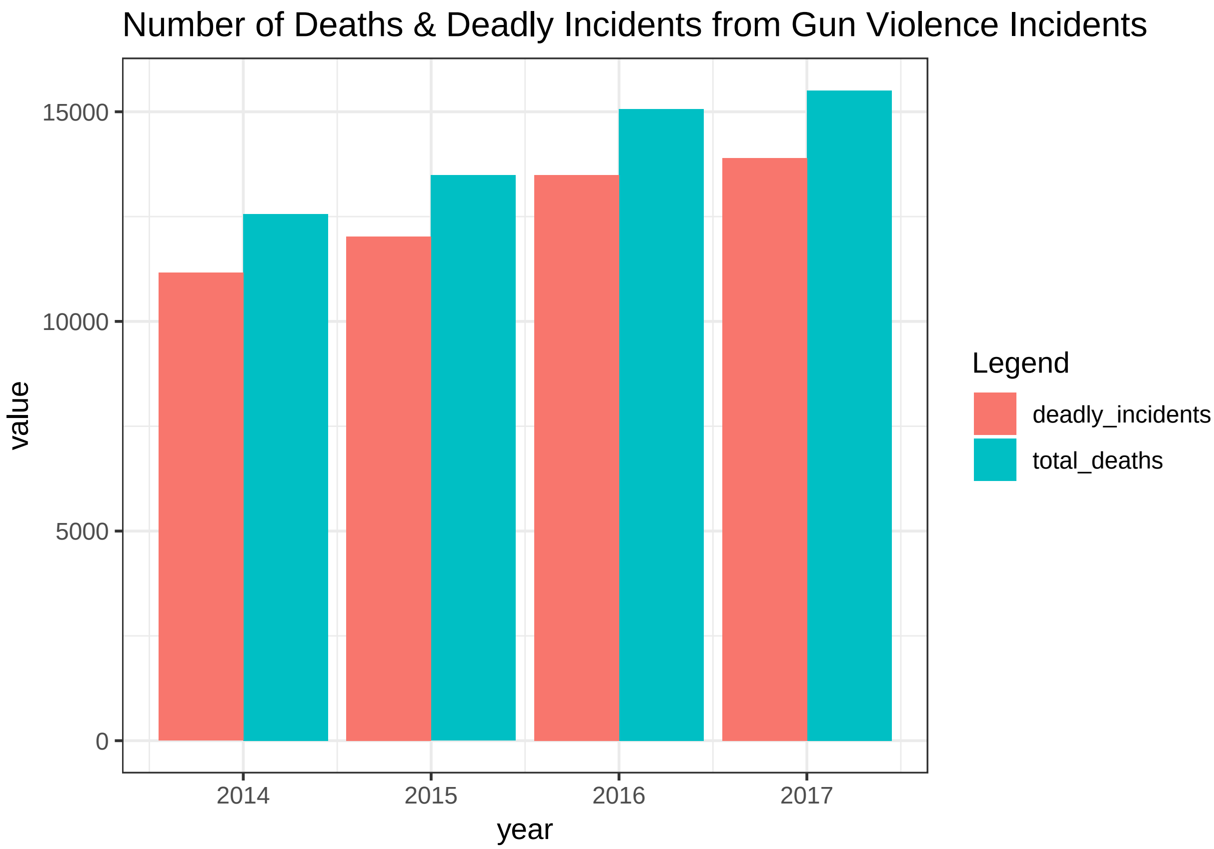



```
##   Var1  Freq
## 1 2014 51854
## 2 2015 53579
## 3 2016 58763
## 4 2017 61401
```

This shows that gun violence has been steadily increasing since 2014 and now is at a high of 61,401 incidents in 2017.

In addition to number of incidents, I want to analyze the change over time in the amount of gun violence incidents that result in death and in the amount of people killed due to gun violence. The distinction between the two is important: in the following function, I define the variable "deadly incidents" to be the amount of gun violence incidents that resulted in at least one death and the variable "total deaths" to be the number of deaths that have resulted from all gun violence incidents. I then plot how the two variables have changed throughout the years.

```r
killed_per_year <-function(df){
  year <- unique(df$year)
  deadly_incidents <- rep(NA, length(year))
  total_deaths <- rep(NA, length(year))
  for(i in 1:length(year)){
    df_filt <- df[df$year==year[i],]
    deadly_incidents[i] <- nrow(df_filt[df_filt$n_killed >0,])
    total_deaths[i] <- sum(df_filt$n_killed)
  }

  df_deaths <- data.frame(year=year, deadly_incidents=deadly_incidents, total_deaths=total_deaths, strin

  #plotting the data
  df_final <- melt(df_deaths, id.vars='year')
  p <- ggplot(df_final, aes(year,value, fill=factor(variable))) + geom_bar(stat="identity",position="do
  print(p)

  return(df_deaths)
}

print(killed_per_year(df("gun_violence.csv")))
```




```
##   year deadly_incidents total_deaths
## 1 2014            11159        12557
## 2 2015            12028        13484
## 3 2016            13494        15066
## 4 2017            13900        15511
```

From the data frame and plot, we can see that the total number of deaths has increased since 2014. The total number of incidents resulting in death has also increased. This follows the overall trend of gun violence incidents increasing since 2014.

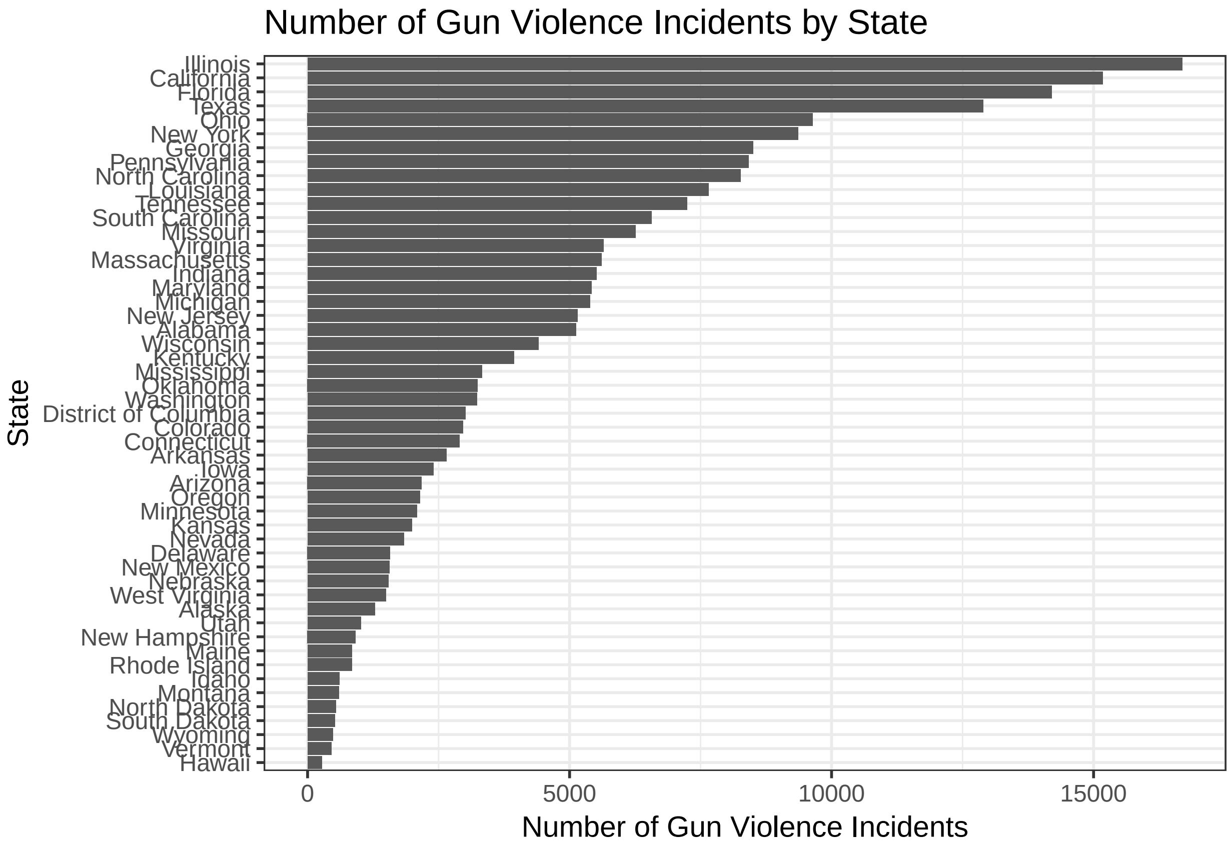
Location Analysis (State/City):

After observing some general trends regarding gun violence, I wanted to see how location impacted gun violence.

First, I created a function to get the total number of gun violence incidents from 2014-2017 in each state and to plot a bar plot illustrating this.

```
shootings_by_state <- function(df){
  df_states <- table(df$state) %>% as.data.frame
  df_states <- df_states[order(df_states$Freq, decreasing=TRUE),]
  states <- factor(df_states$Var1, levels = df_states$Var1[order(df_states$Freq)])
  freq <- df_states$Freq
  p <- ggplot(data=df_states, aes(x=states, y=freq)) + geom_col() + coord_flip()
  p <- p + xlab("State") + ylab("Number of Gun Violence Incidents")
  p <- p + ggtitle("Number of Gun Violence Incidents by State")
  p <- p + theme(plot.title = element_text(hjust = 0.5)) + theme_bw()
  print(p)
```

```r
  return(df_states)
}

print(shootings_by_state(df("gun_violence.csv")))
```



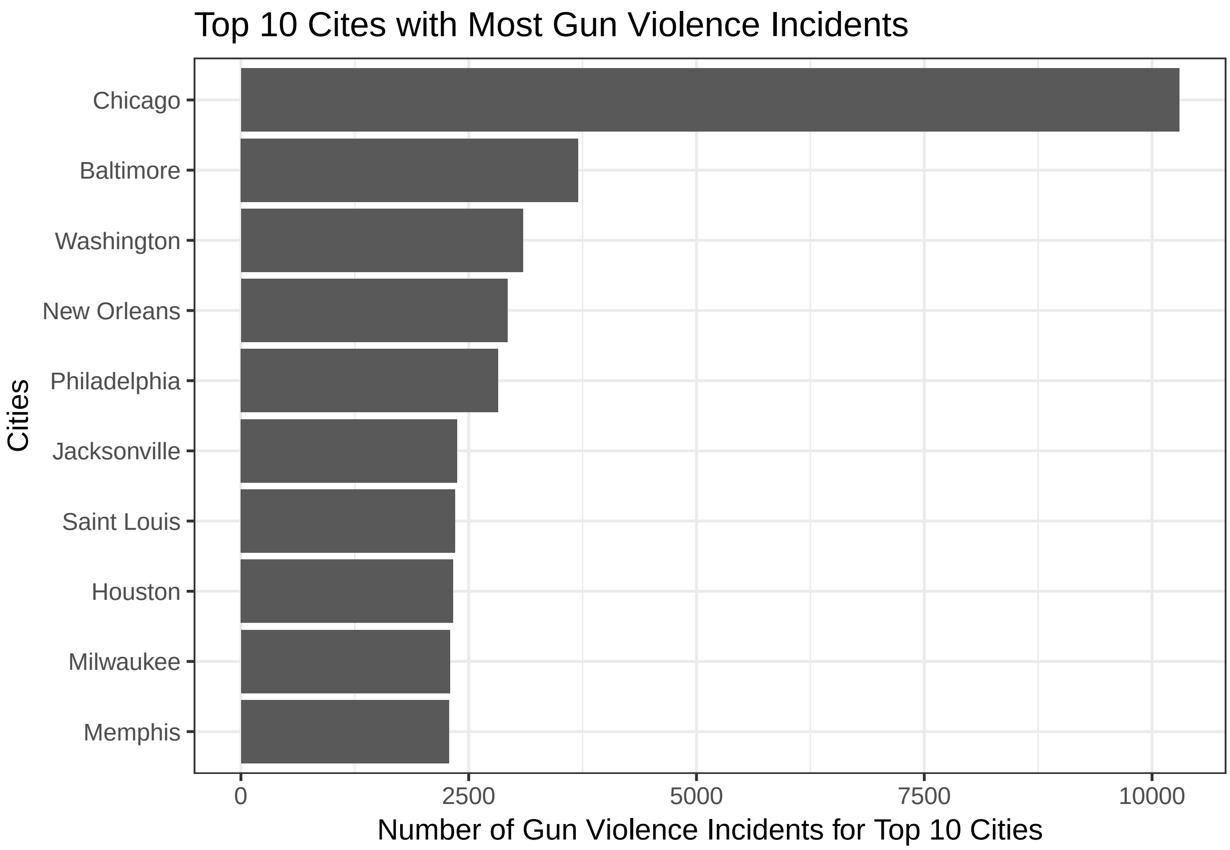

```
##                      Var1  Freq
## 14               Illinois 16695
## 5              California 15171
## 10                Florida 14197
## 44                  Texas 12890
## 36                   Ohio  9643
## 33               New York  9359
## 11                Georgia  8501
## 39           Pennsylvania  8413
## 34         North Carolina  8263
## 19              Louisiana  7655
## 43              Tennessee  7241
## 41         South Carolina  6567
## 26               Missouri  6256
## 47               Virginia  5650
## 22          Massachusetts  5611
## 15                Indiana  5517
## 21               Maryland  5422
## 23               Michigan  5391
## 31             New Jersey  5149
## 1                 Alabama  5122
## 50              Wisconsin  4410
```

```
## 18             Kentucky  3942
## 25          Mississippi  3326
## 37             Oklahoma  3246
## 48           Washington  3230
## 9  District of Columbia  3015
## 6              Colorado  2965
## 7           Connecticut  2903
## 4              Arkansas  2650
## 16                 Iowa  2401
## 3               Arizona  2177
## 38               Oregon  2147
## 24            Minnesota  2087
## 17               Kansas  1992
## 29               Nevada  1840
## 8              Delaware  1576
## 32           New Mexico  1562
## 28             Nebraska  1541
## 49        West Virginia  1493
## 2                Alaska  1288
## 45                 Utah  1015
## 30        New Hampshire   917
## 20                Maine   849
## 40         Rhode Island   843
## 13                Idaho   610
## 27              Montana   597
## 35         North Dakota   542
## 42         South Dakota   518
## 51              Wyoming   480
## 46              Vermont   452
## 12               Hawaii   270
```

Here we see that Illinois, California, and Florida have the highest frequency of shootings from 2014-2017.

Next, I created a function to get the total number of gun violence incidents from 2014-2017 for the 10 cities with the most incidents and to plot a bar plot illustrating this.

```r
shootings_by_city <- function(df){
  df_city <- table(df$city_or_county) %>% as.data.frame
  df_city <- df_city[order(df_city$Freq, decreasing=TRUE),]
  df_city <- df_city[1:10,]
  cities <- factor(df_city$Var1, levels = df_city$Var1[order(df_city$Freq)])
  freq <- df_city$Freq
  p <- ggplot(data=df_city, aes(x=cities, y=freq)) + geom_col() + coord_flip() + xlab("Cities") + ylab(
  print(p)

  return(df_city)
}

print(shootings_by_city(df("gun_violence.csv")))
```




```
##                Var1  Freq
## 1970        Chicago 10298
## 520       Baltimore  3703
## 11869    Washington  3095
## 7866    New Orleans  2926
## 8792   Philadelphia  2824
## 5439   Jacksonville  2372
## 9905    Saint Louis  2351
## 5183        Houston  2331
## 7284      Milwaukee  2294
## 7082        Memphis  2283
```

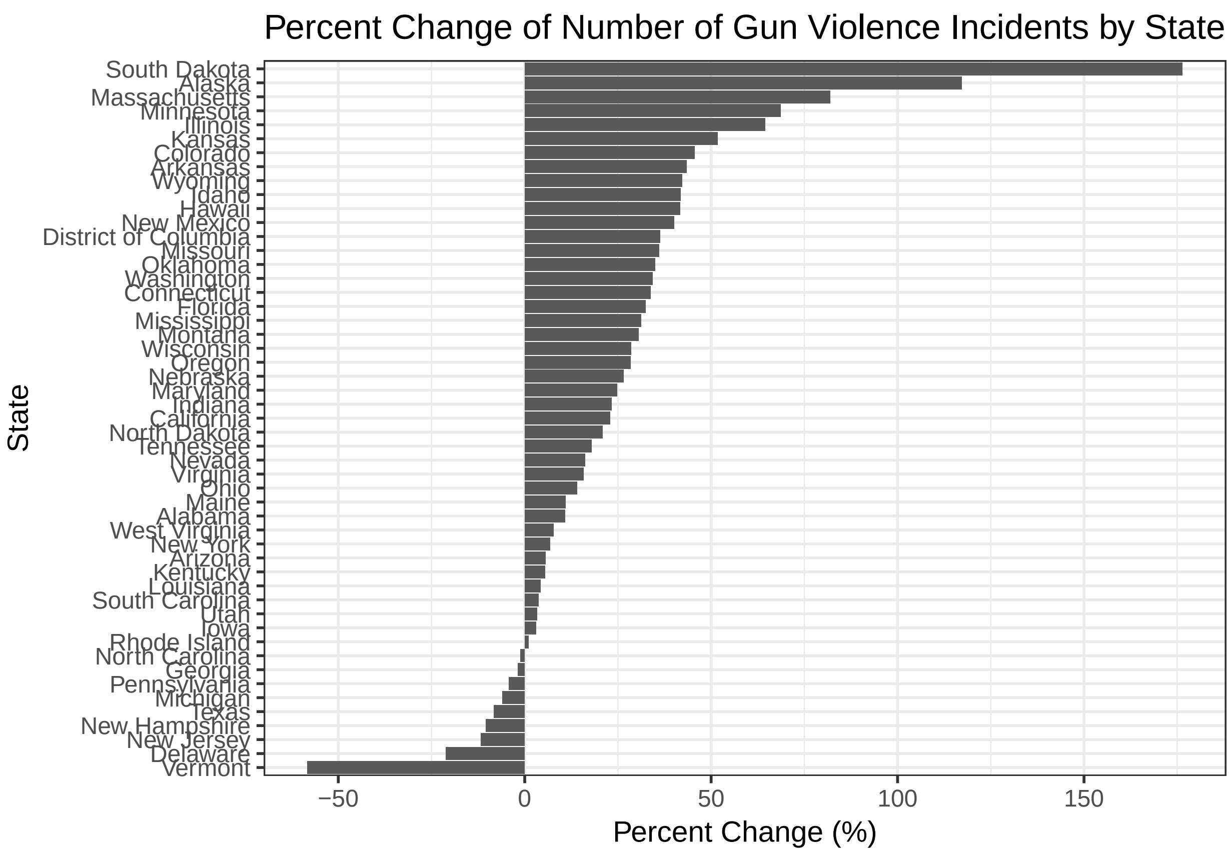
Clearly, Chicago stands out from the rest as the city with the most gun violence by far. Along with it are other major metropolitan cities that one would expect to have more crime on average.

While it is interesting to analyze which places have the most quantity of gun violence, it is important to note that bigger states are naturally going to have more gun violence than small states because there are just more people present to commit these crimes. Also, analyzing the states without looking at the breakdown of the incidents over the years doesn't let us see how states have changed over time.

To be able to see objectively which state has changed the most in terms of gun violence, I create a function to get the percent change from 2014 to 2017 of how many incidents have occured in each state.

```
percent_change <- function(df){
  state <- unique(df$state)
  amt_2014 <- rep(NA, length(state))
  amt_2017 <- rep(NA, length(state))
  percent_change <- rep(NA, length(state))
```

```r
  for(i in 1:length(state)){
    df_filt <- df[df$state==state[i],]
    amt_2014[i] <- nrow(df_filt[df_filt$year=="2014",])
    amt_2017[i] <- nrow(df_filt[df_filt$year=="2017",])
    percent_change[i] <- (amt_2017[i]-amt_2014[i])/amt_2014[i] *100
    }
  df_new <- data.frame(state=state,amt_2014=amt_2014, amt_2017=amt_2017, percent_change=percent_change,s
  df_new <- df_new[order(df_new$percent_change, decreasing=TRUE),]

  states <- factor(df_new$state, levels = df_new$state[order(df_new$percent_change)])
  change <- df_new$percent_change
  p <- ggplot(data=df_new, aes(x=states, y=change)) + geom_col() + coord_flip()
  p <- p + xlab("State") + ylab("Percent Change (%)")
  p <- p + ggtitle("Percent Change of Number of Gun Violence Incidents by State")
  p <- p +theme(plot.title = element_text(hjust = 0.5)) + theme_bw()
  print(p)

  return(df_new)
}

print(percent_change(df("gun_violence.csv")))
```



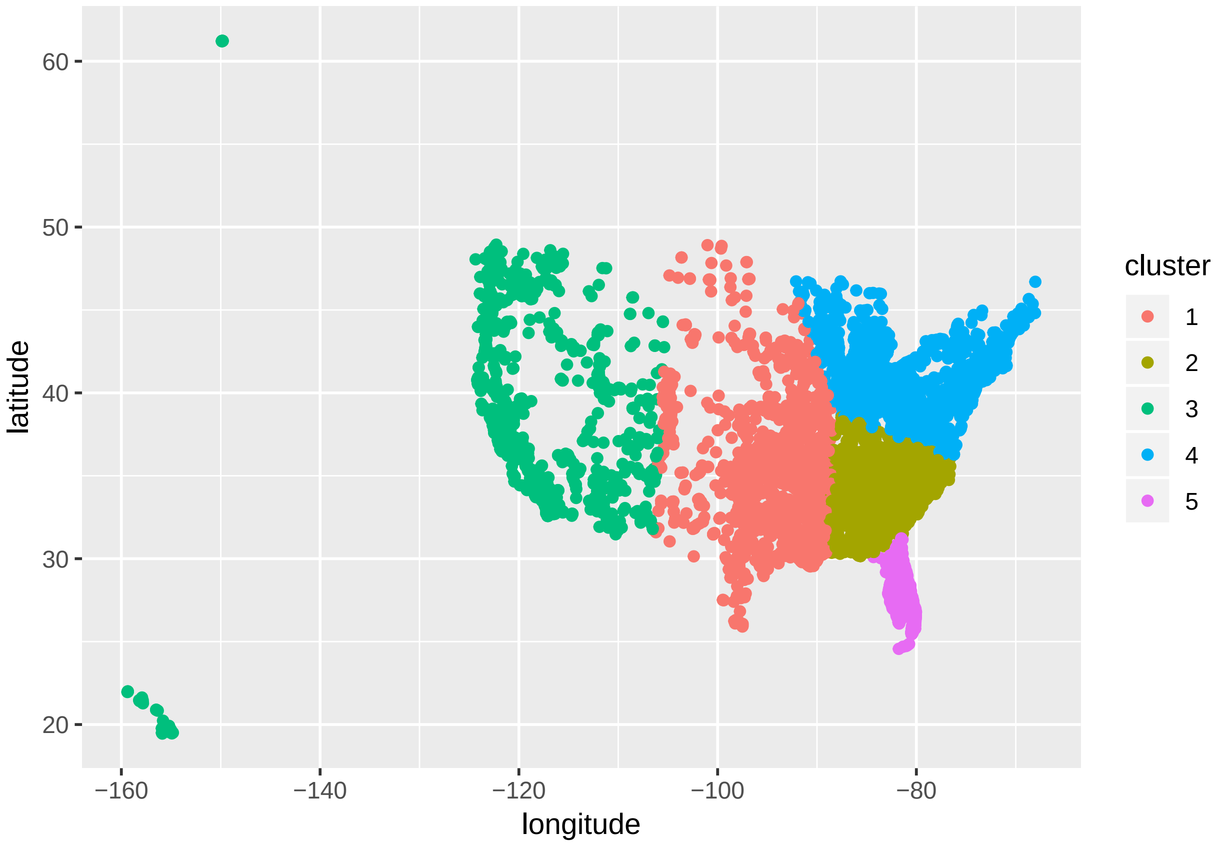

```
##                  state amt_2014 amt_2017 percent_change
## 51        South Dakota       72      199     176.388889
## 38              Alaska      146      317     117.123288
## 20       Massachusetts      968     1761      81.921488
## 42           Minnesota      450      759      68.666667
```

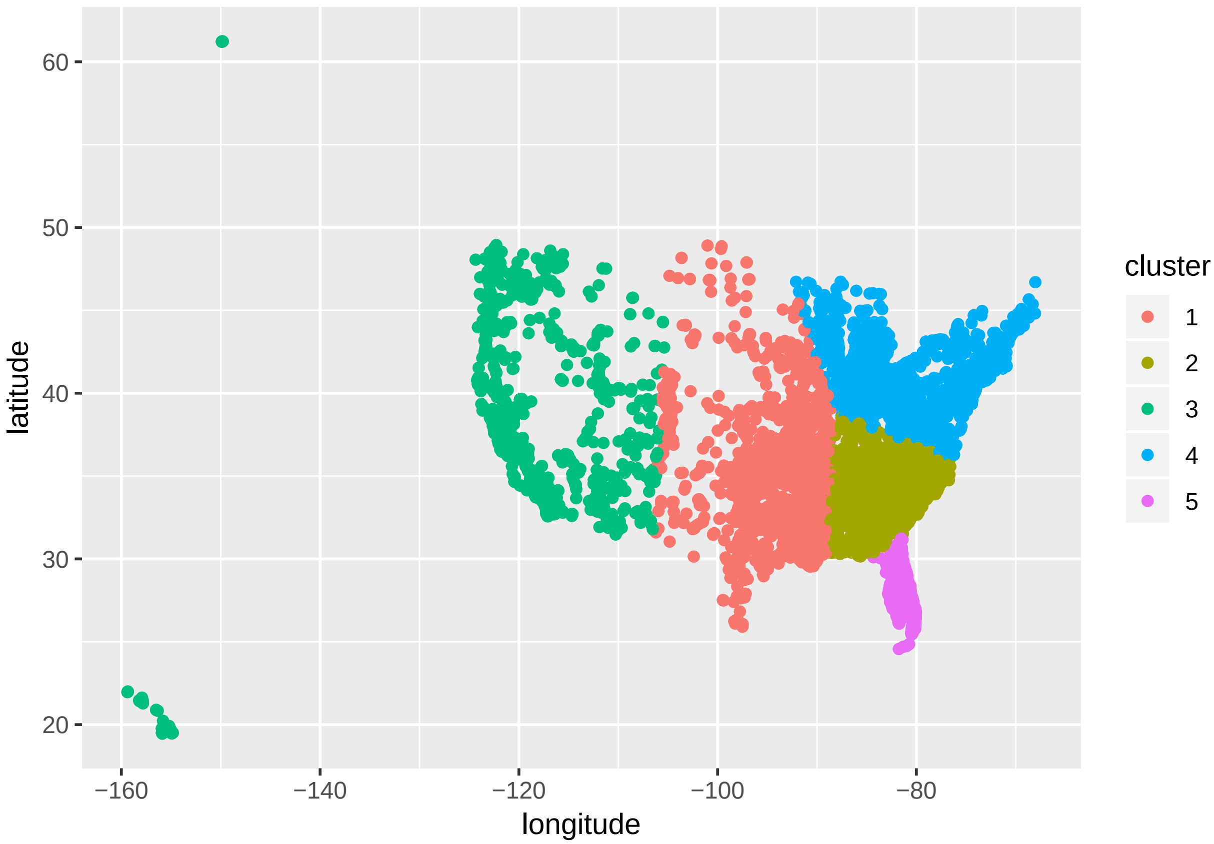
```
## 17             Illinois     3095      5089       64.426494
## 28               Kansas      381       578       51.706037
## 12             Colorado      556       809       45.503597
## 39             Arkansas      572       820       43.356643
## 45              Wyoming       64        91       42.187500
## 43                Idaho      115       163       41.739130
## 25               Hawaii       48        68       41.666667
## 48           New Mexico      320       448       40.000000
## 29 District of Columbia      838      1142       36.276850
## 4              Missouri     1272      1730       36.006289
## 35             Oklahoma      642       866       34.890966
## 26           Washington      656       881       34.298780
## 36          Connecticut      583       780       33.790738
## 21              Florida     3138      4156       32.441045
## 22          Mississippi      784      1029       31.250000
## 49              Montana      128       167       30.468750
## 23            Wisconsin     1065      1369       28.544601
## 44               Oregon      398       511       28.391960
## 19             Nebraska      347       439       26.512968
## 10             Maryland     1266      1579       24.723539
## 16              Indiana     1203      1483       23.275145
## 34           California     3732      4588       22.936763
## 47         North Dakota      110       133       20.909091
## 40            Tennessee     1590      1874       17.861635
## 41               Nevada      376       437       16.223404
## 7              Virginia     1273      1473       15.710919
## 27                 Ohio     2368      2701       14.062500
## 46                Maine      183       203       10.928962
## 13              Alabama     1318      1461       10.849772
## 30        West Virginia      362       390        7.734807
## 3              New York     1903      2031        6.726222
## 24              Arizona      556       587        5.575540
## 6              Kentucky      977      1030        5.424770
## 14            Louisiana     1906      1987        4.249738
## 9        South Carolina     1660      1721        3.674699
## 15                 Utah      217       224        3.225806
## 37                 Iowa      569       586        2.987698
## 32         Rhode Island      189       191        1.058201
## 33       North Carolina     2165      2141       -1.108545
## 5               Georgia     2032      1994       -1.870079
## 8          Pennsylvania     2267      2172       -4.190560
## 1              Michigan     1447      1360       -6.012440
## 18                Texas     3133      2875       -8.234919
## 31        New Hampshire      230       206      -10.434783
## 2            New Jersey     1521      1342      -11.768573
## 11             Delaware      493       389      -21.095335
## 50              Vermont      170        71      -58.235294
```

We see here that South Dakota and Alaska are the states that have increased the most percentage wise in gun violence, but it is important to note that they had pretty low amounts of gun violence to begin with in 2014 so it didn't take that many more incidents to go up as much as they did. It is also motivating to see that some states, such as Vermont, have decreased in incidents of gun violence over the past four years.

Next, I will show visually where deaths from gun violence occur in the nation using the kmeans function.

```r
k_means <- function(df){
df_filt<-na.omit(df[df$n_killed>0,])

m <- dplyr::select(df_filt, longitude, latitude) %>% as.matrix
km_out <- kmeans(m, centers=5)

df_clustered <- data.frame(df_filt, cluster=factor(km_out$cluster))

p <- ggplot()
p <- p + geom_point(mapping=aes(x=longitude, y=latitude, color=cluster), data=df_clustered)
print(p)

}
print(k_means(df("gun_violence.csv")))
```



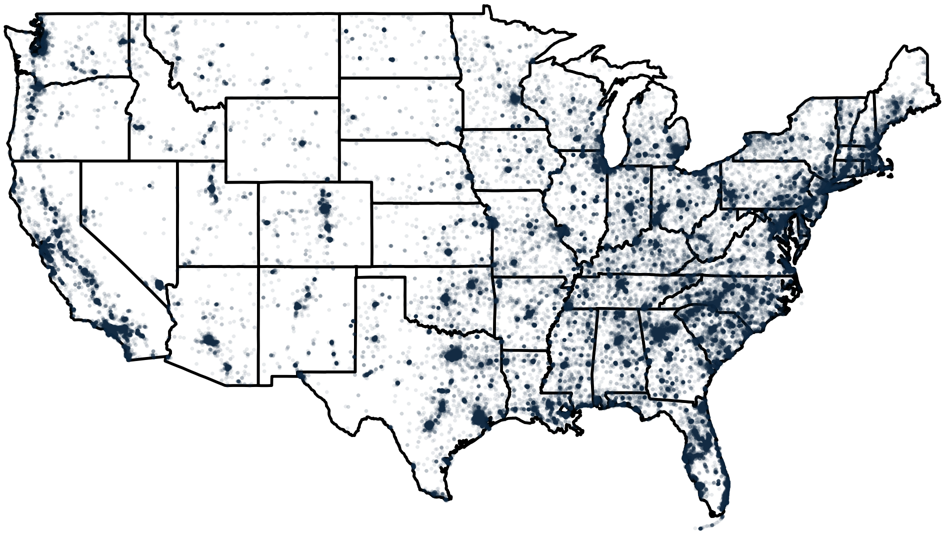


This clearly shows an outline of the United States and specifically the east coast seems the most filled in, representing that it has the most deaths from gun violence.

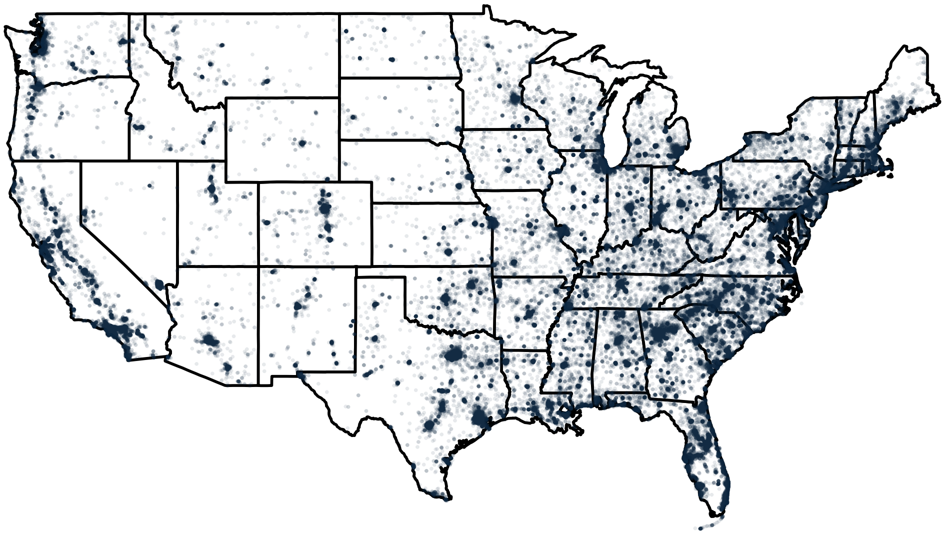
I also discovered the map_data function in the ggplot package that can better visually show where deaths from gun violence take place.

```
map_shooting_deaths <- function(df){
global <- map_data("state")
p <- ggplot(global, aes(x = long, y = lat)) + geom_polygon(aes(group = group), fill = "white", col = "bl
  geom_point(data = df, aes(x = longitude, y = latitude, col = n_killed), size = 0.001, alpha = .1) +
  theme_void() +
  theme(legend.position = "none")
print(p)
}

print(map_shooting_deaths(df("gun_violence.csv")))
```

```
## Warning: Removed 6340 rows containing missing values (geom_point).
```



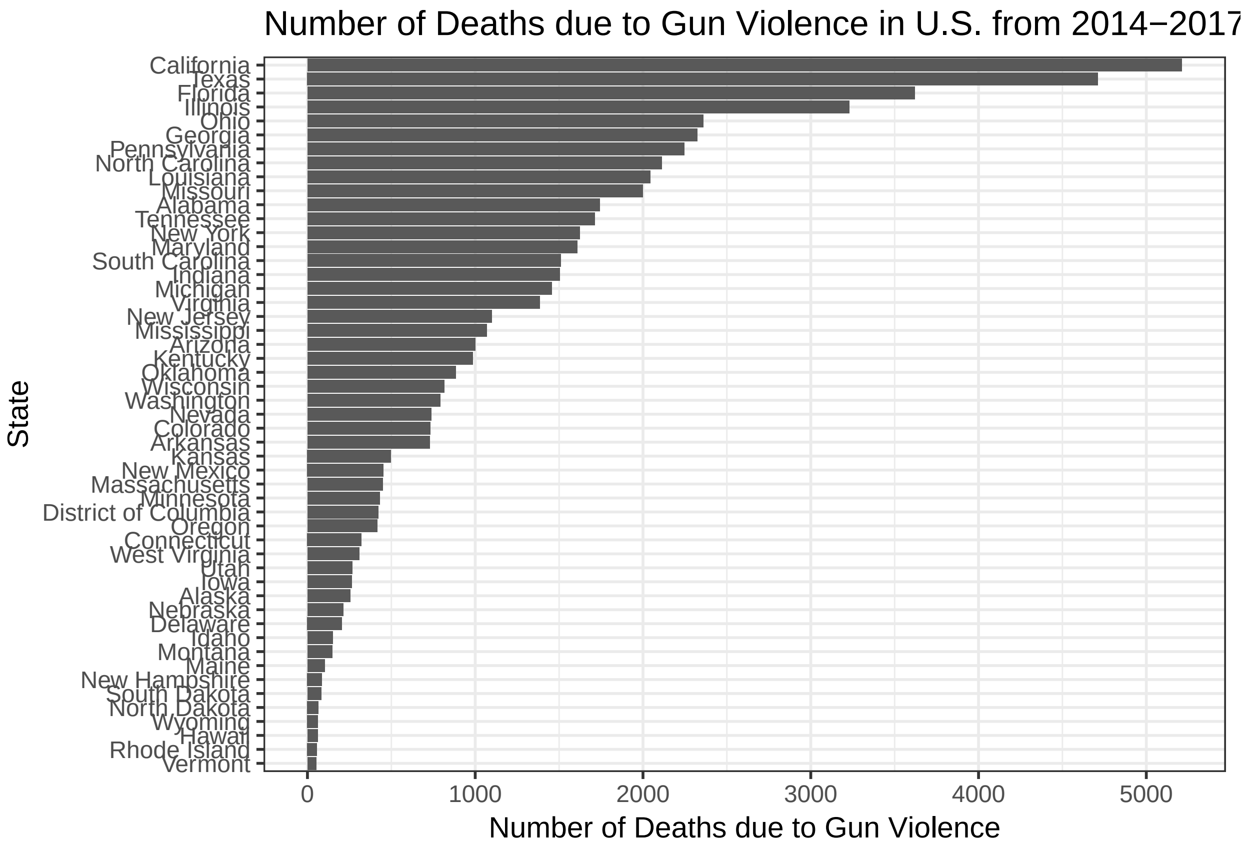
```
## Warning: Removed 6340 rows containing missing values (geom_point).
```



Now, to see the specific states in which deaths are common in gun violence, I create a function to determine the number of "deadly incidents" and "total deaths" per state and plot the total deaths per state.

```
killed_per_state <-function(df){
  df <- df[df$n_killed >0,]
  state <- unique(df$state)
  deadly_incidents <- rep(NA, length(state))
  total_deaths <- rep(NA, length(state))
  for(i in 1:length(state)){
  df_filt <- df[df$state==state[i],]
  deadly_incidents[i] <- nrow(df_filt)
  total_deaths[i] <-sum(df_filt$n_killed)
  }
  df_new <- data.frame(state=state,deadly_incidents=deadly_incidents,totaL_deaths=total_deaths,stringsA
  df_new <- df_new[order(total_deaths, decreasing=TRUE),]

  states <- factor(state, levels = state[order(total_deaths)])
  p <- ggplot(data=df_new, aes(x=states, y=total_deaths)) + geom_col() + coord_flip() + xlab("State") +
```

```
  print(p)

  return(df_new)
}

print(killed_per_state(df("gun_violence.csv")))
```




```
##                 state deadly_incidents totaL_deaths
## 15         California             4680         5212
## 13              Texas             4066         4712
## 10            Florida             3129         3620
## 29           Illinois             3003         3230
## 24               Ohio             2118         2359
## 22            Georgia             2062         2323
## 25       Pennsylvania             2005         2246
## 20     North Carolina             1915         2112
## 6           Louisiana             1851         2042
## 21           Missouri             1804         1999
## 5             Alabama             1581         1742
## 23          Tennessee             1562         1712
## 1            New York             1494         1623
## 12           Maryland             1477         1607
## 2      South Carolina             1368         1510
## 11            Indiana             1336         1503
## 17           Michigan             1313         1457
## 14           Virginia             1227         1384
## 19         New Jersey             1015         1099
```

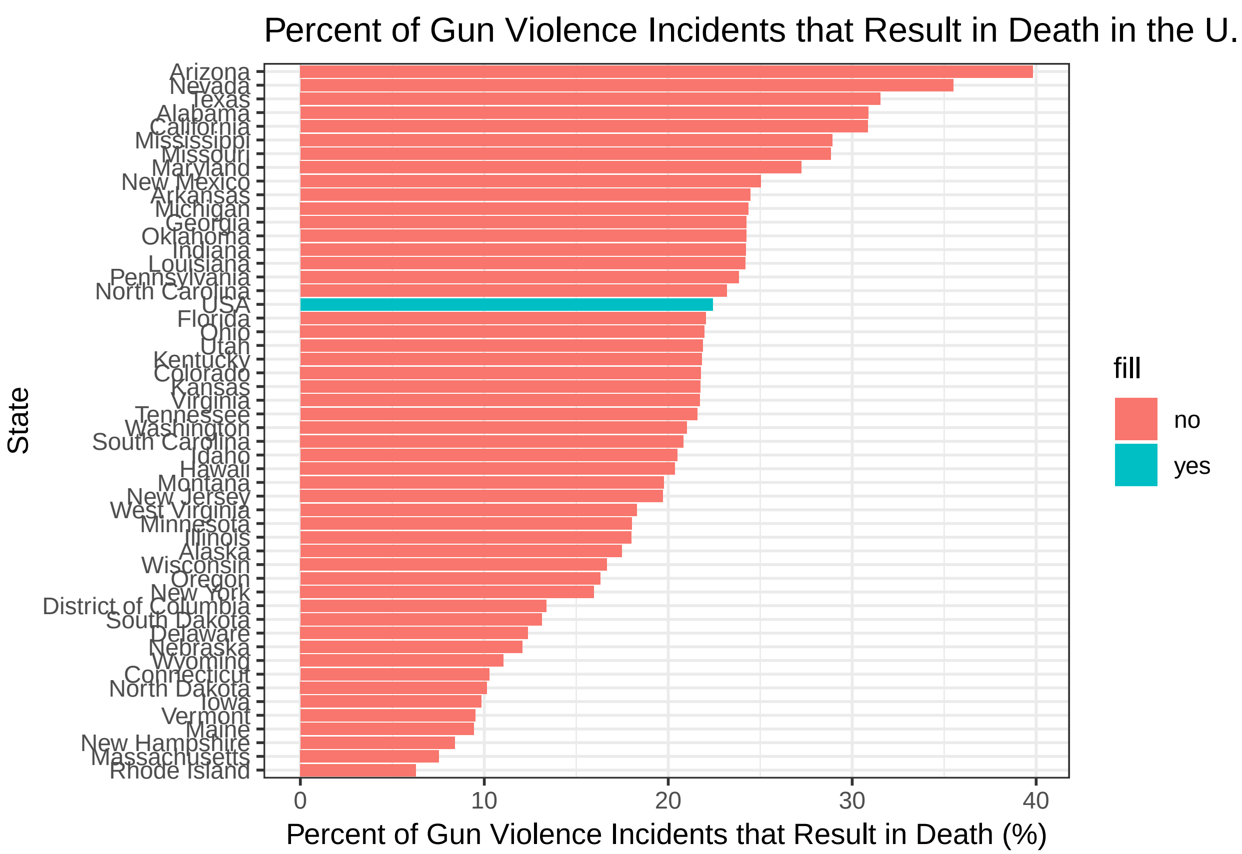
```
## 7           Mississippi              962        1068
## 8               Arizona              867        1001
## 35             Kentucky              860         985
## 18             Oklahoma              787         883
## 33            Wisconsin              735         816
## 30           Washington              679         791
## 28               Nevada              653         738
## 4              Colorado              646         731
## 36             Arkansas              648         730
## 16               Kansas              433         498
## 41           New Mexico              391         453
## 37        Massachusetts              423         450
## 32            Minnesota              376         432
## 45 District of Columbia              403         422
## 27               Oregon              350         415
## 34          Connecticut              298         322
## 26        West Virginia              273         308
## 46                 Utah              222         268
## 44                 Iowa              236         263
## 42               Alaska              225         256
## 39             Nebraska              186         214
## 3              Delaware              195         206
## 49                Idaho              125         152
## 31              Montana              118         149
## 51                Maine               80         104
## 38        New Hampshire               77          87
## 50         South Dakota               68          83
## 43         North Dakota               55          66
## 40              Wyoming               53          63
## 9                Hawaii               55          62
## 48         Rhode Island               53          57
## 47              Vermont               43          53
```

We can see that while Illinois was the top state for overall gun violence incidents, California is top for incidents resulting in death. This could tell us a variety of things, like that shootings in Illinois may not be very violent and could just result in injuries mostly.

Next, I examine the percent, rather than just the amount, of shootings that resulted in death in each state and in the U.S. as a nation to compare states to the nationwide average.

```
murder_percent <- function(df){
  state <- unique(df$state)
  murders <- rep(NA, length(state))
  shootings <- rep(NA, length(state))
  percent <- rep(NA, length(state))
  for(i in 1:length(state)){
    df_filt <- df[df$state==state[i],]
    murders[i] <- nrow(df_filt[df_filt$n_killed>0,])
    shootings[i] <- nrow(df_filt)
    percent[i] <- murders[i]/shootings[i] *100
  }
  df_new <- data.frame(state=state, percent_of_shootings = percent,fill=rep("no",51))

  #comparing with the nation in general
  all_murders <- nrow(df[df$n_killed >0,])
```

```r
  all_shootings <- nrow(df)
  all_percent <- all_murders/all_shootings * 100
  df_usa <- data.frame(state="USA",percent_of_shootings=all_percent,fill="yes")

  #combine the states df and the nation df
  df_final <- rbind(df_new,df_usa)
  df_final <- df_final[order(df_final$percent_of_shootings, decreasing = TRUE),]

  states <- factor(df_final$state, levels = df_final$state[order(df_final$percent_of_shootings)])
  percent <- df_final$percent_of_shootings

  p <- ggplot(data=df_final, aes(x=states, y=percent, fill=fill))
  p <- p + geom_bar(stat="identity") + coord_flip() + xlab("State")
  p <- p + ylab("Percent of Gun Violence Incidents that Result in Death (%)")
  p <- p + ggtitle("Percent of Gun Violence Incidents that Result in Death in the U.S. from 2014-2017 b
  p <- p +theme(plot.title = element_text(hjust = 0.5)) + theme_bw()
  print(p)

  return(df_final)
}

print(murder_percent(df("gun_violence.csv")))
```




```
##                      state percent_of_shootings fill
## 24                 Arizona            39.825448   no
## 41                  Nevada            35.489130   no
## 18                   Texas            31.543832   no
```

```
## 13              Alabama           30.866849  no
## 34           California           30.848329  no
## 22          Mississippi           28.923632  no
## 4              Missouri           28.836317  no
## 10             Maryland           27.240871  no
## 48           New Mexico           25.032010  no
## 39             Arkansas           24.452830  no
## 1              Michigan           24.355407  no
## 5               Georgia           24.255970  no
## 35             Oklahoma           24.245225  no
## 16              Indiana           24.216059  no
## 14            Louisiana           24.180274  no
## 8          Pennsylvania           23.832165  no
## 33       North Carolina           23.175602  no
## 52                  USA           22.420954 yes
## 21              Florida           22.039868  no
## 27                 Ohio           21.964119  no
## 15                 Utah           21.871921  no
## 6              Kentucky           21.816337  no
## 12             Colorado           21.787521  no
## 28               Kansas           21.736948  no
## 7              Virginia           21.716814  no
## 40            Tennessee           21.571606  no
## 26           Washington           21.021672  no
## 9        South Carolina           20.831430  no
## 43                Idaho           20.491803  no
## 25               Hawaii           20.370370  no
## 49              Montana           19.765494  no
## 2            New Jersey           19.712566  no
## 30        West Virginia           18.285332  no
## 42            Minnesota           18.016291  no
## 17             Illinois           17.987421  no
## 38               Alaska           17.468944  no
## 23            Wisconsin           16.666667  no
## 44               Oregon           16.301816  no
## 3              New York           15.963244  no
## 29 District of Columbia           13.366501  no
## 51         South Dakota           13.127413  no
## 11             Delaware           12.373096  no
## 19             Nebraska           12.070084  no
## 45              Wyoming           11.041667  no
## 36          Connecticut           10.265243  no
## 47         North Dakota           10.147601  no
## 37                 Iowa            9.829238  no
## 50              Vermont            9.513274  no
## 46                Maine            9.422850  no
## 31        New Hampshire            8.396947  no
## 20        Massachusetts            7.538763  no
## 32         Rhode Island            6.287070  no
```

In the U.S. in general, 22% of shootings result in death. Arizona comes at the top with 40% of their shootings resulting in death. At the opposite end of the spectrum, Rhode Island's shootings only result in death about 6% of the time.

Finally, in examining gun violence by location, I make a function that allows me to get important stats about shooting in a particular city and state or just in a particular state.

```
stats_by_place <- function(df, state, city){
  if(city=="all"){
  df_filt <- df[df$state==state,]
  city <- "all"
  }
  else{
  df_filt <- df[df$state==state & df$city==city,]
  }
  total_shootings <- nrow(df_filt)
  shootings_2014 <- nrow(df_filt[df_filt$year=="2014",])
  shootings_2015 <- nrow(df_filt[df_filt$year=="2015",])
  shootings_2016 <- nrow(df_filt[df_filt$year=="2016",])
  shootings_2017 <- nrow(df_filt[df_filt$year=="2017",])
  percent_change <- (shootings_2017 - shootings_2014)/shootings_2014 * 100
  total_deaths <- sum(df_filt$n_killed)
  deadly_incidents <- nrow(df_filt[df_filt$n_killed>0,])
  percent_deadly_incidents <- total_deaths/total_shootings * 100
  city <- city

  df_new <- data.frame(state=state,city=city, total_shootings=total_shootings,shootings_2014=shootings_2

  return(df_new)
}
print(stats_by_place(df("gun_violence.csv"),"Illinois","Chicago"))
```

```
##      state    city total_shootings shootings_2014 shootings_2015
## 1 Illinois Chicago           10298           2042           2369
##   shootings_2016 shootings_2017 percent_change_num_shootings total_deaths
## 1           3075           2812                     37.70813         2011
##   deadly_incidents percent_deadly_incidents
## 1             1901                 19.52806
```

```
print(stats_by_place(df("gun_violence.csv"),"Illinois","all"))
```

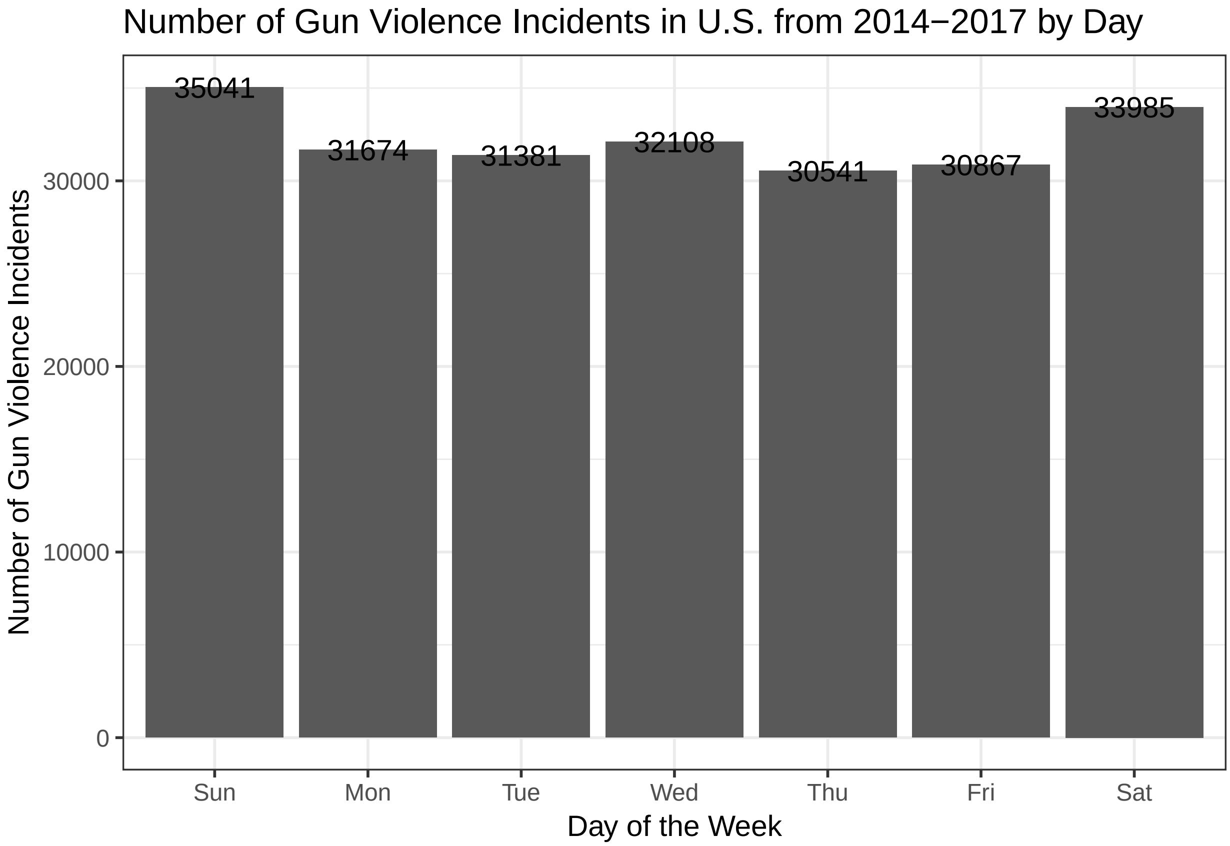
```
##      state city total_shootings shootings_2014 shootings_2015
## 1 Illinois  all           16695           3095           3456
##   shootings_2016 shootings_2017 percent_change_num_shootings total_deaths
## 1           5055           5089                     64.42649         3230
##   deadly_incidents percent_deadly_incidents
## 1             3003                 19.34711
```

Analysis by Time (Day of the Week & Month) I will now do a brief analysis of the factor of time on gun violence in the U.S.

First, I will analyze gun violence by the day of the week to see if there is a correlation.

```
shootings_by_day <- function(df){
  df_days <- table(df$weekday) %>% as.data.frame
  day <- df_days$Var1
  num_shootings <- df_days$Freq
  p <- ggplot(data=df_days, aes(x=day, y=num_shootings)) + geom_bar(stat="identity") + xlab("Day of the
  print(p)

  return(df_days)
}
```

```r
print(shootings_by_day(df("gun_violence.csv")))
```

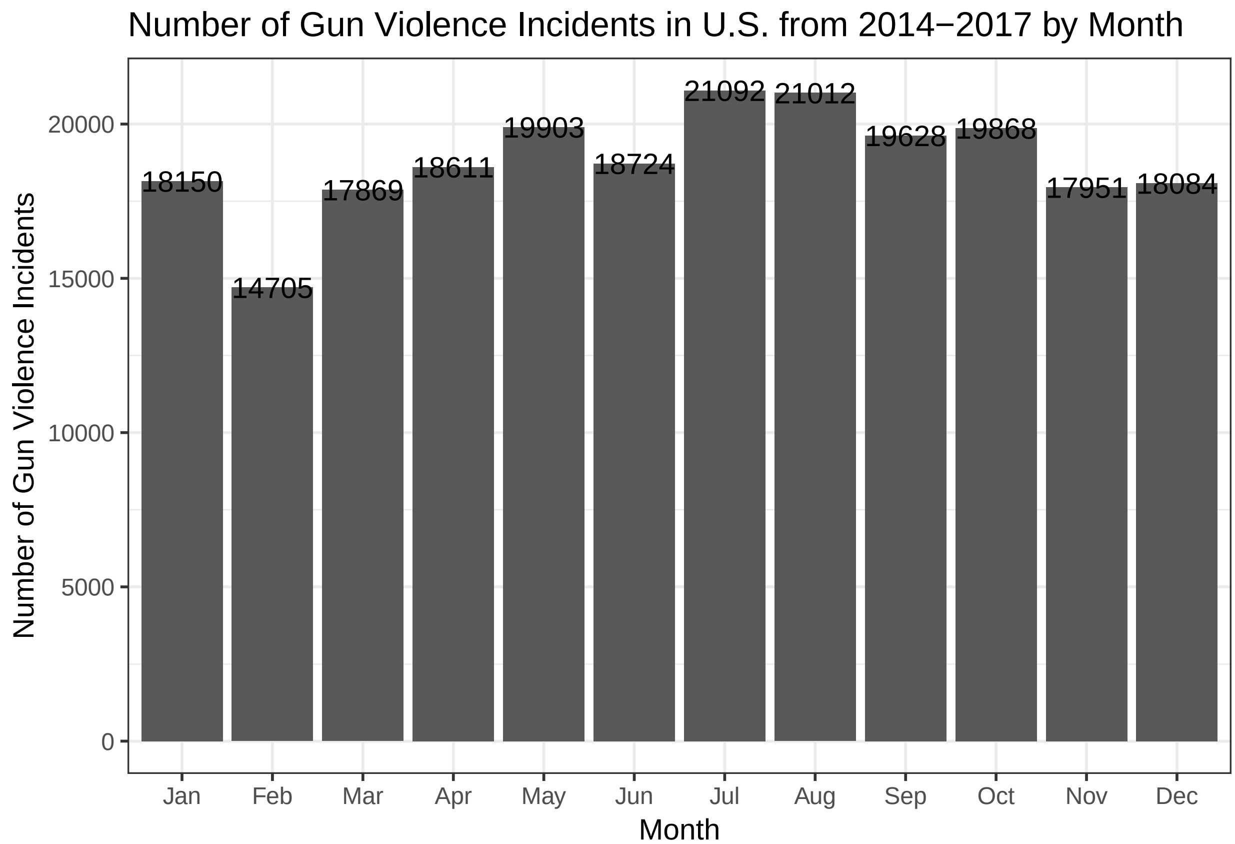



```
##   Var1  Freq
## 1  Sun 35041
## 2  Mon 31674
## 3  Tue 31381
## 4  Wed 32108
## 5  Thu 30541
## 6  Fri 30867
## 7  Sat 33985
```

Sunday has the most shootings followed by Saturday, which makes sense because people have less commitments and more time to get into trouble.

Next, I will analyze gun violence by the month to see if there is a correlation.

```r
shootings_by_month <- function(df){
  df_month <- table(df$month) %>% as.data.frame
  df_month <- df_month[order(df_month$Freq, decreasing=TRUE),]
  month <- df_month$Var1
  num_shootings <- df_month$Freq
  p <- ggplot(data=df_month, aes(x=month, y=num_shootings)) + geom_bar(stat="identity") + xlab("Month")
  print(p)

  return(df_month)
}
print(shootings_by_month(df("gun_violence.csv")))
```




```
##     Var1  Freq
## 7    Jul 21092
## 8    Aug 21012
## 5    May 19903
## 10   Oct 19868
## 9    Sep 19628
## 6    Jun 18724
## 4    Apr 18611
## 1    Jan 18150
## 12   Dec 18084
## 11   Nov 17951
## 3    Mar 17869
## 2    Feb 14705
```

Shootings seem related to the warmer months- they are at an all time high in July and August, warm months, and low in February and March, the coldest months. This could be because people want to go outside and do things that could get you into trouble more during the summer and less during the winter.

Analysis by Characteristics of the Incident Finally I will briefly look at some of the characteristics of shootings as possible factors.

When examining the incident characteristics column of the data frame, I realized that the column described briefly what the shooting was about and Iwanted to look into three categories that popped up alot: "Gang involvement", "Drug involvement" and "Domestic Violence".

```
reason_for_shooting <- function(df){
  year <- unique(df$year)
  gang_involvement <- rep(NA, length(year))
  drug_involvement <- rep(NA, length(year))
```

```r
  domestic_violence <- rep(NA, length(year))
  for(i in 1:length(year)){
    df_filt <- df[df$year==year[i],]
    gang_involvement[i] <- nrow(df_filt[grepl("Gang involvement",df_filt$incident_characteristics),])
    drug_involvement[i] <- nrow(df_filt[grepl("Drug involvement",df_filt$incident_characteristics),])
    domestic_violence[i] <- nrow(df_filt[grepl("Domestic Violence",df_filt$incident_characteristics),])
  }
  df_new <- data.frame(year=year,gang_involvement=gang_involvement, drug_involvement=drug_involvement, d
  return(df_new)
}

print(reason_for_shooting(df("gun_violence.csv")))
```

```
##   year gang_involvement drug_involvement domestic_violence
## 1 2014             1072             2037              1759
## 2 2015             1578             3889              2259
## 3 2016             1647             4736              2884
## 4 2017             1098             5114              3137
```

From this analysis, we can see that gang involvement shootings seem to be staying steady, drug involvement shootings have gone way up, and domestic violence shootings have also gone up. However, a lot of the time the people entering the data may not have known if a certain shooting was related to these three categories so this data could be skewed.

Lastly, I want to see how many children/teens are involved in shootings (whether that be killed, injured,involved).

```r
kid_shootings <- function(df){
  kids_in_shootings <- length(grep("Child 0-11",df$participant_age_group))
  teens_in_shootings <- length(grep("Teen 12-17",df$participant_age_group))
  df_new <- data.frame(kids_in_shootings=kids_in_shootings,teens_in_shootings=teens_in_shootings,string

return(df_new)
}

print(kid_shootings(df("gun_violence.csv")))
```

```
##   kids_in_shootings teens_in_shootings
## 1              3506              18110
```

Luckily, not many kids were involved in shootings, but the number goes way up for teenagers because many probably got more involved in trouble and could actually be the suspect instead of just a young kid caught up in a shooting.

Big Takeaways: 1. Gun violence incidents and the amount of deaths from gun violence have been steadily increasing in the United States from 2014 to 2017 on the average, and in the vast majority of states. 2. Gun violence is more likely to occur in certain states, such as Illinois and California, and is more likely to happen on the weekend and in warmer summer months.