Data 2401 - Final Report

Tuan Pham

2023-04-26

Why is this subject?

- A controversial topic (Gun Violence)
- Provide an edge of the problems, a tip of the iceberg.
- Raise your insights about the problems, and then you can draw conclusions for yourself in order to protect you and your family.
- Recognize the primitive signs, not to live in skeptics but to enhance your awareness.

> *"If you look at the number of Americans killed since 9/11 by terrorism, it's less than 100. If you look at the number been killed by gun violence, it's in the tens of thousands."*

"The 44th US President - Barack Obama"

Source: NBCNews

DATA

`Words on my data set`

- **164 observations and 26 variables** from 1960 to 2023. (It was 143 observations and 25 variables originally from 1982 to 2022)

- My data set's limitation:
  - Two data sets with two completely different structures.
  - Various manually plug-in data (15%)

Mother Jones

https://www.motherjones.com/politics/2012/12/mass-shootings-mother-jones-full-data/

- The data set's authors are MARK FOLLMAN, GAVIN ARONSEN, and DEANNA PAN.
- The time span they collect is from 1982 to present, which included `originally 143 observations and 25 variables`. This is my work's mainframe.

The Violence Project

https://www.theviolenceproject.org/

- No specific author's names.
- The time span is from 1960 to 2022. This is the supplemental data for my main data frame.

DATA (cont.)

```
my_data <- read.csv('Mother Jones - Mass Shootings Database, 1982 - 2023 - Sheet1.csv', na.strings = "-
my_data2 <- read.csv('Violence Project Mass Shooter Database - Version 6.1 - Full Database.csv', na.str
```

- both in form of `csv file`.

  First set: familiar with what we have been learned so far.

  Second set: highly complicated

- Missing data:

  NA (Not Available) - Missing value

  Unclear - The information has not been revealed by the authority.

  TBD (To Be Determined) - The information has not been confirm yet.

A Glimpse over the Data

```
glimpse(my_data)
```

```
## Rows: 164
## Columns: 26
## $ case                                <chr> "Louisville bank shooting", "Nashvill~
## $ city                                <chr> "Louisville", "Nashville", "East Lans~
## $ state                               <chr> "KY", "TN", "MI", "CA", "CA", "VA", "~
## $ date                                <chr> "4/10/2023", "3/27/2023", "2/13/2023"~
## $ summary                             <chr> "Connor Sturgeon, 25, opened fire ins~
## $ fatalities                          <int> 5, 6, 3, 7, 11, 6, 5, 3, 5, 3, 7, 3, ~
## $ injured                             <int> 8, 6, 5, 1, 10, 6, 25, 2, 2, 2, 46, 0~
## $ total_victims                       <int> 13, 12, 8, 8, 21, 12, 30, 5, 7, 5, 53~
## $ location                            <chr> "workplace", "School", "School", "wor~
## $ age_of_shooter                      <int> 25, 28, 43, 67, 72, 31, 22, 22, 15, 2~
## $ prior_signs_mental_health_issues    <chr> "Yes", NA, NA, NA, "Yes", NA, "Yes", ~
## $ mental_health_details               <chr> NA, NA, NA, NA, "According to the LA ~
## $ weapons_obtained_legally            <chr> "Yes", "Yes", "Yes", NA, NA, NA, NA, ~
## $ where_obtained                      <chr> "gun dealership in Louisville", NA, N~
## $ weapon_type                         <chr> "Semiautomatic Rifle", "One Semiautom~
## $ weapon_details                      <chr> "AR-15 rifle", NA, NA, NA, NA, NA, NA~
## $ race                                <chr> "White", "White", "Black", "Asian", "~
## $ gender                              <chr> "M", "F (\"identifies as transgender\~
## $ sources                             <chr> "https://apnews.com/article/downtown-~
## $ mental_health_sources               <chr> NA, NA, NA, NA, "https://www.latimes.~
## $ sources_additional_age              <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ latitude                            <dbl> NA, NA, NA, NA, NA, 36.77262, 38.8809~
## $ longitude                           <dbl> NA, NA, NA, NA, NA, -76.25128, -104.7~
## $ type                                <chr> "Mass", "Mass", "Mass", "Spree", "Mas~
## $ year                                <int> 2023, 2023, 2023, 2023, 2023, 2022, 2~
## $ day_of_week                         <chr> "Monday", "Monday", "Monday", "Monday~
```

Data Key Terms

Case: Case's name well-known by the media

City/State: Location where the incidents happened

Date/Year/Date of Week: Specific day and year when the incidents occurred

Summary: Summary about the case

Fatalities/Injured/Total Victims: Facts that stand out from the incident

Location: Type of location where the incidents occurred

Age of Shooter/Race/Gender/Prior Sign Mental Issues: The shooter profile

Weapons Obtained Legally/Where Obtained/Weapons Type/Weapons Details: Weapons profile

Sources/Mental Health Sources/Additional Age Source: All sources that were used to conduct this data set.

Longitude/Latitude: GPS coordination of the incident's location

Type: Mass Shooting or Shooting Spree designated to the incident

Packages

- The tidyverse.

- Data wrangling and plots.

```
#install.packages(tidyverse)
library(dplyr)
library(tidyverse)
```

Data Wrangling

- All the data type was character, even with numeric variables' types
- There is a limitation in term of data information about the observations of `location` variable in the raw data. Most of the Nightlife observations like Bar/Club/Restaurants was classified as `Other`, which make the original data had a big chunk of number in `Other` category.

```
# Assign the variables to the data type of my choice.
 my_data$age_of_shooter <- as.integer(my_data$age_of_shooter)
 my_data$fatalities <- as.integer(my_data$fatalities)
 my_data$injured <- as.integer(my_data$injured)
 my_data$total_victims <- as.integer(my_data$total_victims)
 my_data$latitude <- as.numeric(my_data$latitude)
 my_data$longitude <- as.numeric(my_data$longitude)
```

```
# remove newline
my_data$location <- str_replace_all(my_data$location,'[\r\n]','')
my_data$race <- str_replace_all(my_data$race,'[\r\n]','')


# replace a string by another one
my_data$location[my_data$location == 'religious' | my_data$location == 'Religious'] <- 'Religious Place
my_data$location[my_data$location == 'workplace'] <- 'Workplace'

my_data$race[my_data$race == 'unclear'] <- 'Unclear'
my_data$race[my_data$race == 'black'] <- 'Black'
my_data$race[my_data$race == 'white'] <- 'White'

my_data$gender[2] <- 'Trans'
my_data$gender[my_data$gender == 'M'] <- 'Male'
my_data$gender[my_data$gender == 'F'] <- 'Female'
```

What's The Mass Shooting?

> The FBI defines `a mass shooting` as any incidents in which **at least four people** are murdered with a gun.

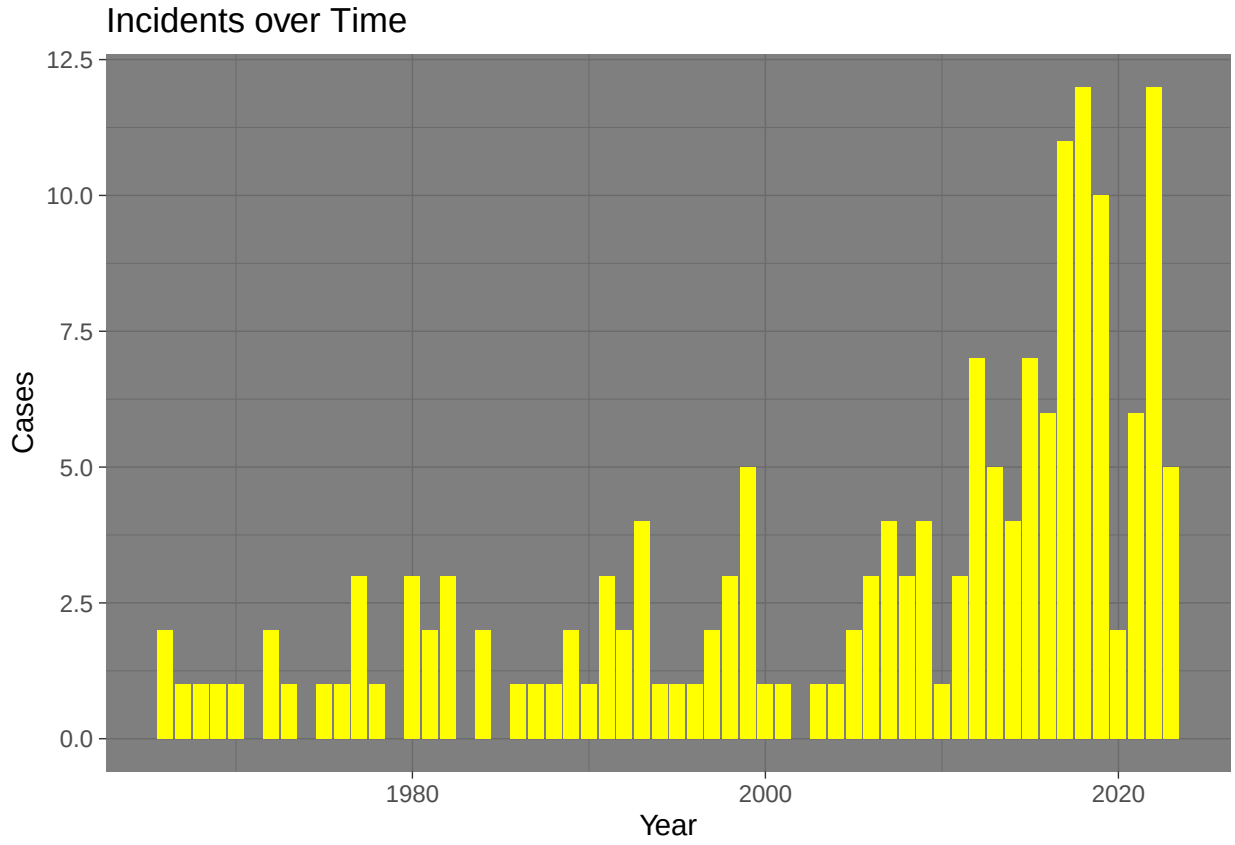
Source: DOJ

What questions should be raised?

1. Will the time frame would say anything about the incidents `in general`?
2. Would the `age, race, and gender` give any insights about the shooter's profile?

3. What would stand out if we cross `the shooter with prior mental health issues` out of the equation?
4. Where are the `locations` that the incidents likely take place?
5. What `types of weapons` the assaiants likely use?
6. What conclusion about the `age of shooter`, `race` and `prior mental health issues` could we draw?
7. What is interesting about the connection between `age of the shooter` over `year`?
8. Will `gender` play any roles in corresponding to `age of the shooter`?
9. How have the incidents distributed `across the America`?

A First Glance about the Incidents over Years



Data Summary

```
##  age_of_shooter   fatalities      injured       total_victims
##  Min.   :11.0   Min.   : 3.0   Min.   :  0.00   Min.   :  3.00
##  1st Qu.:23.0   1st Qu.: 4.0   1st Qu.:  1.00   1st Qu.:  6.00
##  Median :32.0   Median : 6.0   Median :  3.00   Median : 10.00
##  Mean   :33.9   Mean   : 7.5   Mean   : 10.56   Mean   : 18.09
##  3rd Qu.:43.0   3rd Qu.: 8.0   3rd Qu.:  9.50   3rd Qu.: 16.00
##  Max.   :72.0   Max.   :58.0   Max.   :546.00   Max.   :604.00
##                                NA's   :1        NA's   :1
```

1. Las Vegas Strip Massacre: 604 victims

```
##                       Case Fatality Injured Victims      City State      Date
## 1 Las Vegas Strip massacre       58     546     604 Las Vegas    NV 10/1/2017
##   Age  Race Gender
## 1  64 White   Male
##
## 1 23 firearms, mostly rifles; including scopes, and two modified for "fully automatic" firing; two w
```

2. LA Dance Studio Mass Shooting: Oldest age for a mass shooter

3. West Middle School Killings: Youngest age

```
##                             Case Age Fatality Injured Victims          City
## 1   LA dance studio mass shooting  72       11      10      21 Monterey Park
## 2 Westside Middle School killings  11        5      10      15     Jonesboro
##   State      Date  Race Gender
## 1    CA 1/21/2023 Asian   Male
## 2    AR 3/24/1998 White   Male
##                                                                      Weapons
## 1                                                    Semiautomatic Handgun
## 2 Two Rifles, Two Semiautomatic Handguns, Three Revolvers and Two Derringers
```

- Those three cases are the outliers for each of the variables that they represent.

Who are they?

The Picture in General

```
## # A tibble: 8 x 3
##   Race            Total Ratio
##   <chr>           <int> <dbl>
## 1 White              96 58.5
## 2 Black              29 17.7
## 3 Latino             13  7.93
## 4 Asian              11  6.71
## 5 Other               8  4.88
## 6 Native American     3  1.83
## 7 Unclear             2  1.22
## 8 <NA>                2  1.22
```

Before 2002

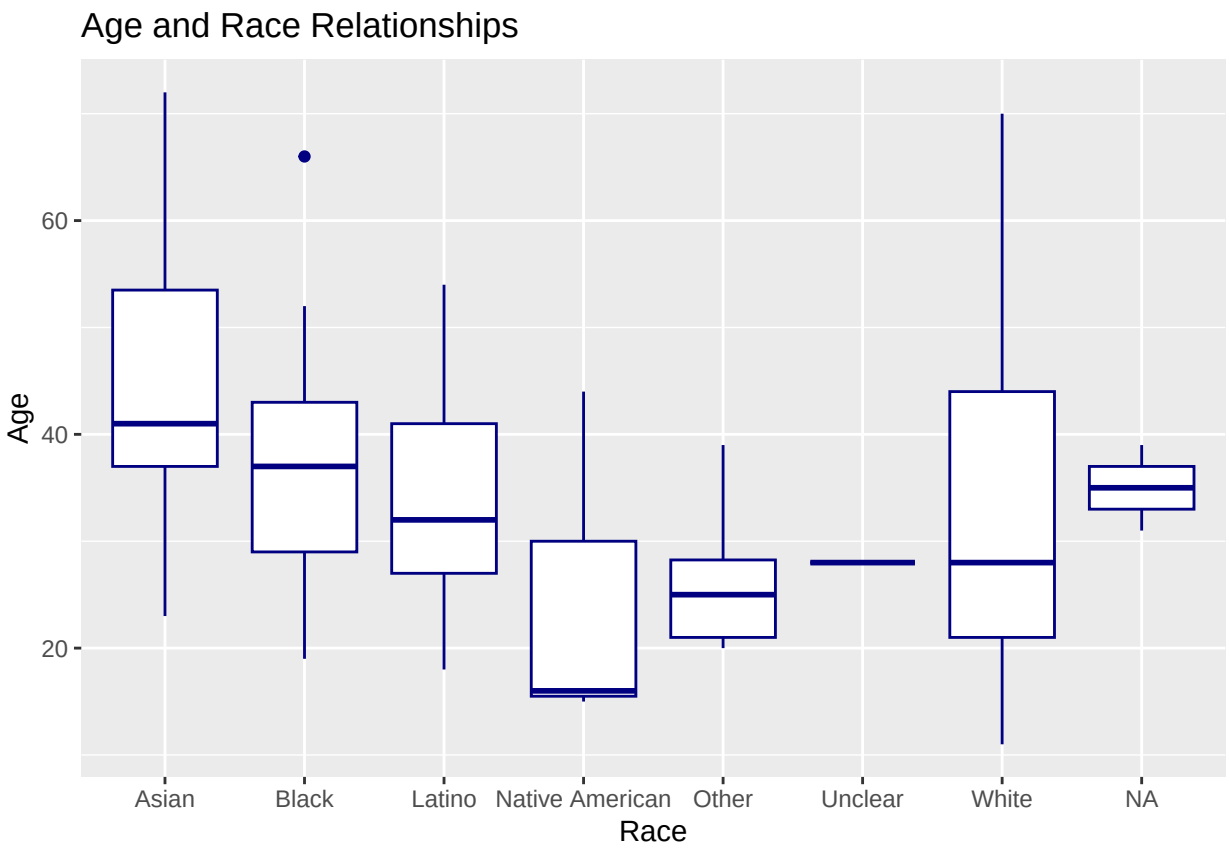
```
## # A tibble: 5 x 3
##   Race    Count Ratio
##   <chr>   <int> <dbl>
## 1 White      39 70.9
## 2 Black      10 18.2
## 3 Asian       3  5.45
## 4 Latino      2  3.64
## 5 Unclear     1  1.82
```

After 2002

```
## # A tibble: 8 x 3
##   Race            Count  Ratio
##   <chr>           <int>  <dbl>
## 1 White              57 52.3
## 2 Black              19 17.4
## 3 Latino             11 10.1
## 4 Asian               8  7.34
## 5 Other               8  7.34
## 6 Native American     3  2.75
## 7 <NA>                2  1.83
## 8 Unclear             1  0.917
```

- Since 2002 is the year without any major incidents about the mass shooting, I chose it as a reference point for my split stats.
- Before 2002, the story seems to be about some certain races, but after 2002, it becomes all the races' story.
- Remember there are four decades before 2002, and only two decades after 2002, but the cases after 2002 shoot up more than double before 2002, 55 versus 109 respectively.

The Average Age of the Shooters among Races



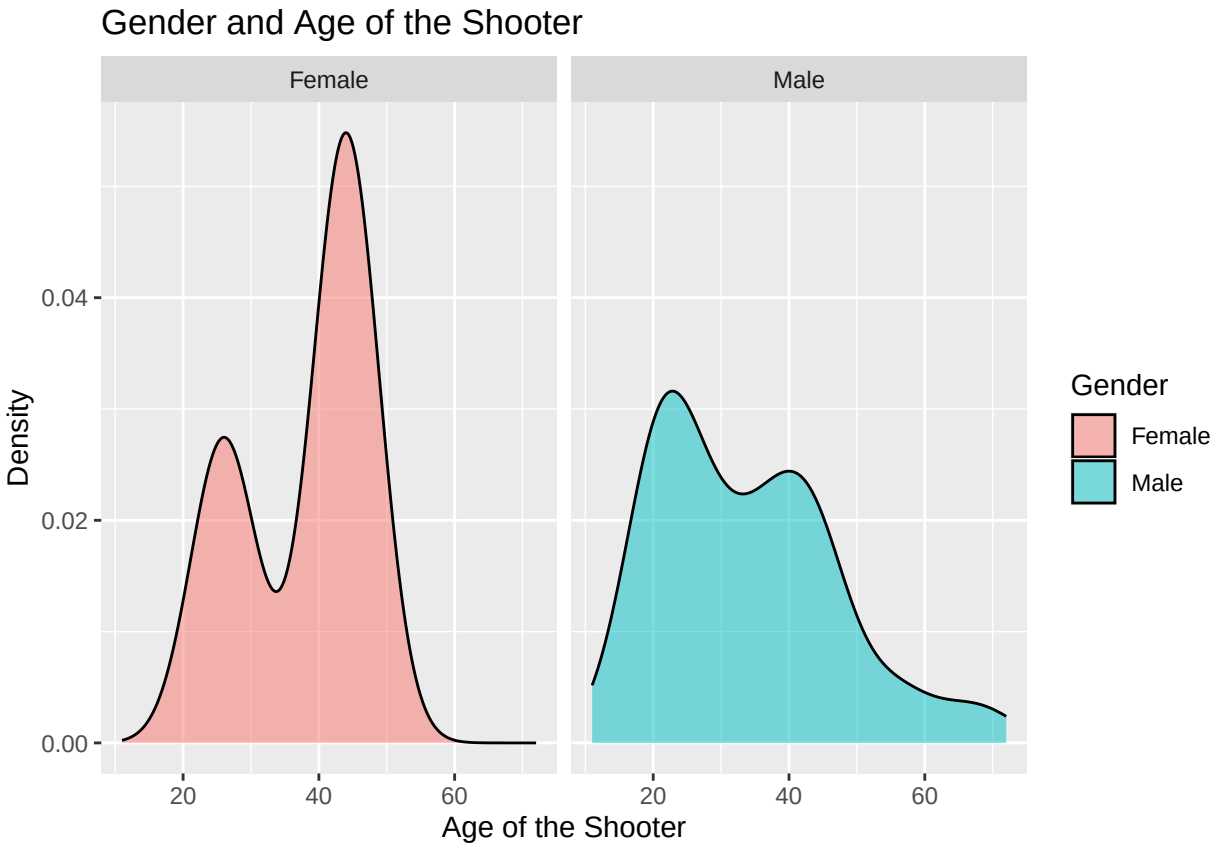
Race	Average Age of the Shooters
White	~ 28-29 years old
Latino	~ 32-33 years old
Black	~ 38-39 years old
Asian	~ 41 years old
Native Am.	~ 18 years old

Gender of the Shooters

```
## # A tibble: 4 x 3
##   Gender        Count Percentage
##   <chr>         <int>      <dbl>
## 1 Male            158     96.3
## 2 Female            3      1.83
## 3 Male & Female     2      1.22
## 4 Trans             1      0.610
```

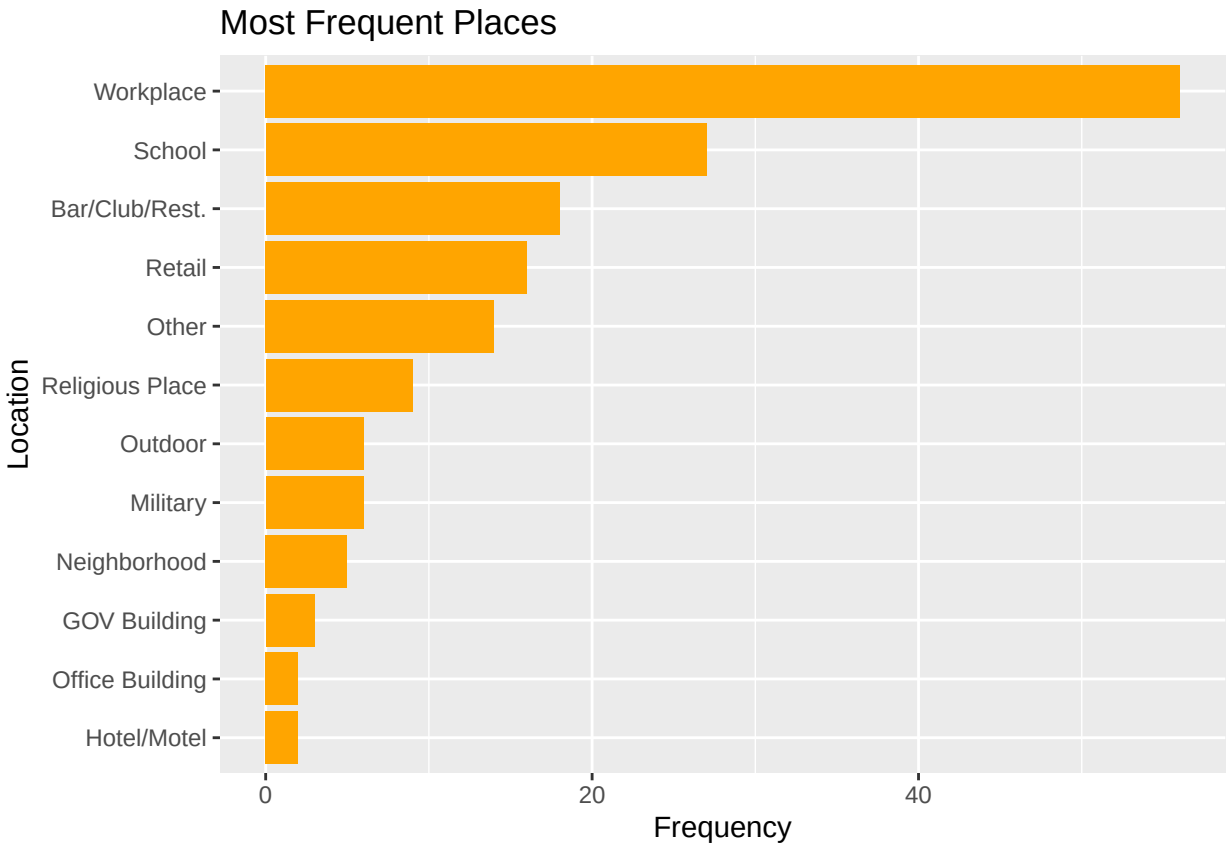
Gender	Percentage
Male	97%
Female	2.5%
Transgender	0.5%

Gender and Age of the Shooters



Gender	Most Likely
Male	Early 20's to Mid 40's
Female	Mid 20's or Mid 40's

Where Do The Mass Shootings Likely Occur?



Location	Frequency
Workplace	~ 34%
School	~ 16%
Bar/Club/Rest.	~ 11%
Retail	~ 10%
Other	~ 9%
Religious Place	~ 6%

It is heartbroken to see School is second rank on the list, which means a lot of innocent kids got their future ahead taken.

What Weapons Were Likely Used by The Assailants?

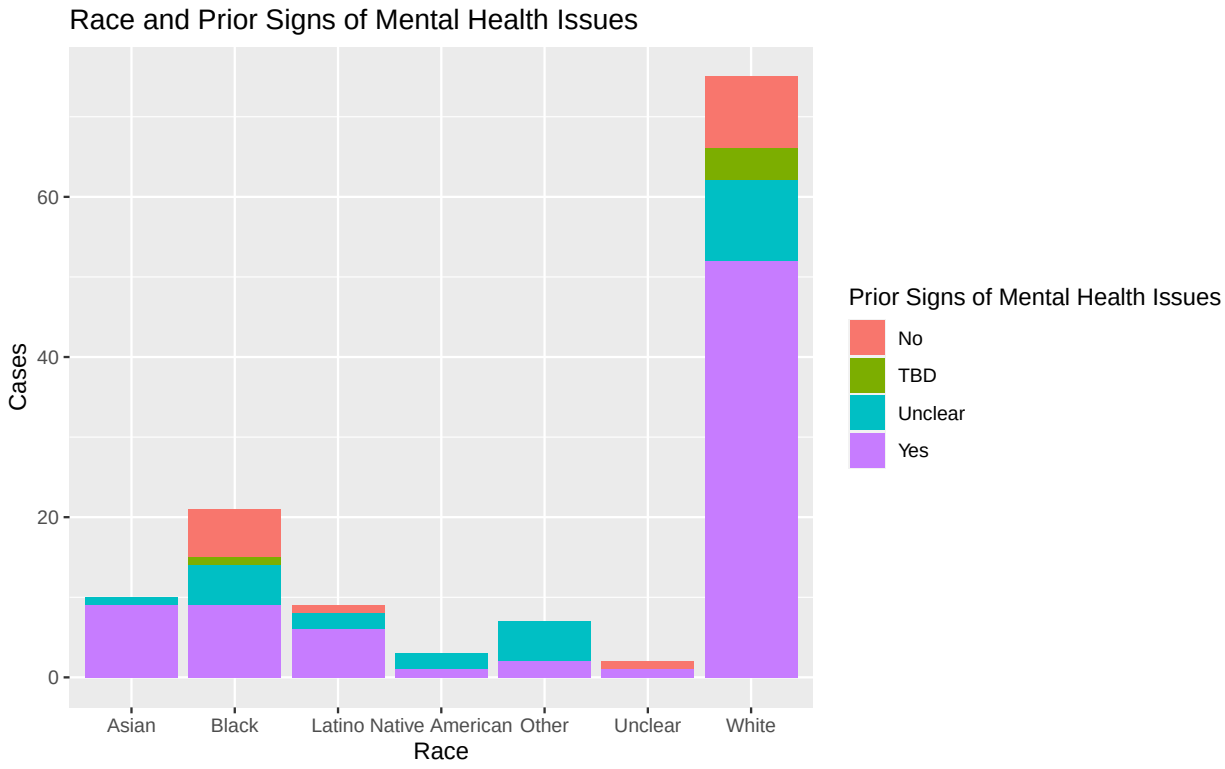
```
## # A tibble: 9 x 3
##   `Weapon Types`                                          Percentage Count
##   <chr>                                                        <dbl> <int>
## 1 Semiautomatic Handgun                                         24.4    40
## 2 Semiautomatic Rifle                                           11.0    18
## 3 One Semiautomatic Rifle and One Semiautomatic Handgun          7.32   12
## 4 Handgun                                                        6.71   11
## 5 Rifle                                                          4.88    8
## 6 Assault Rifle                                                  3.66    6
```

```
## 7 Two Semiautomatic Handguns                              3.66     6
## 8 Two Handguns                                            3.05     5
## 9 One Semiautomatic Handgun and One Revolver              2.44     4
```

Firearm	Percent of Carrying
Semi-Auto Handgun	~ 41%
Semi-Auto Rifle	~ 20%
Handgun(Old Versions)	~ 6.7%
Rifle(Old Version)	~ 5%
Assault Rifle	~ 5%
Shotgun	~ 4%

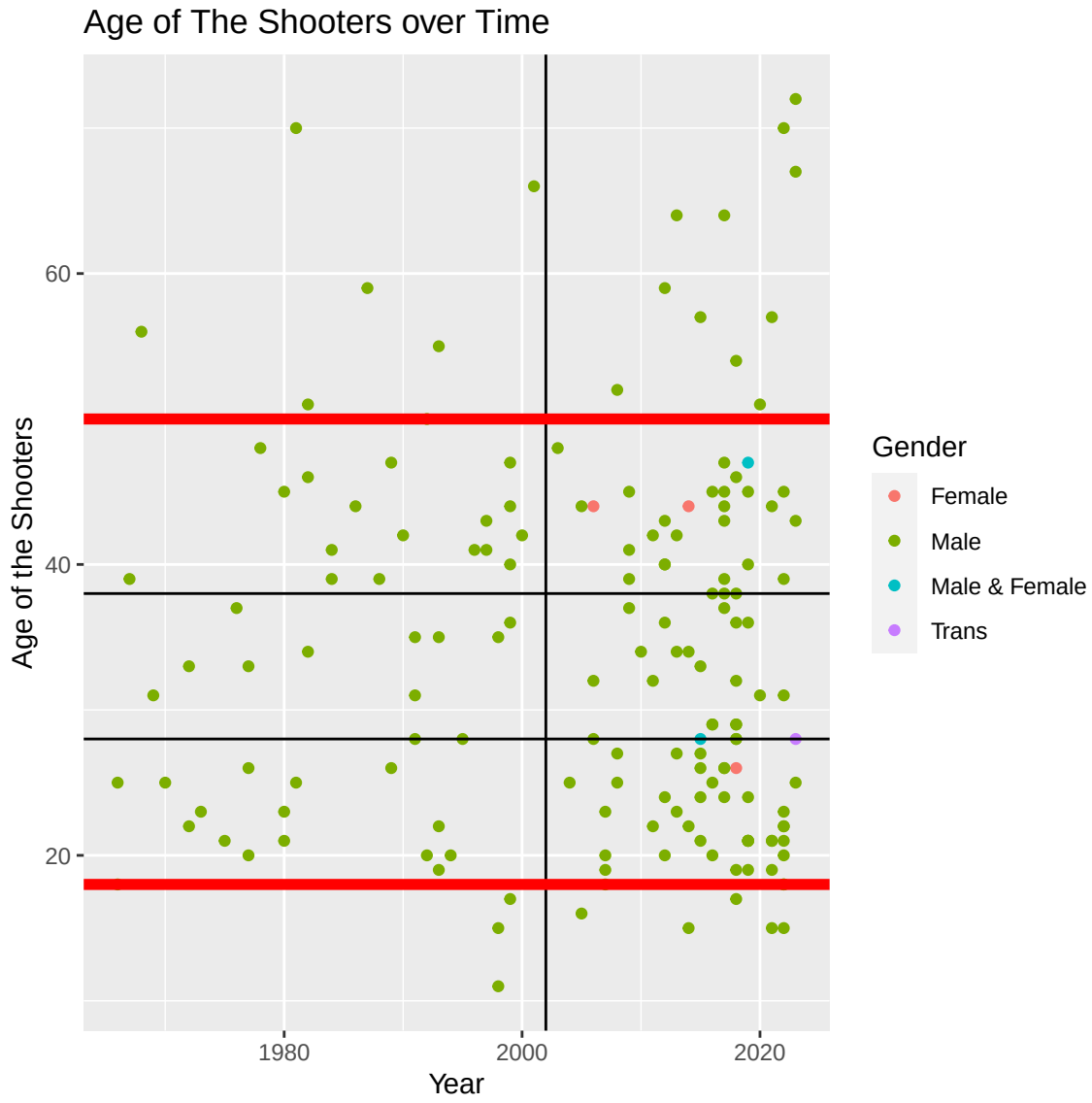
- The preferred weapons was used by the mass shooter are Semi-Auto Handgun, Semi-Auto Rifle or both of them.
- Light weight, high bullet rate, big magazines. It could help them cause mass casualties in a short period of time.

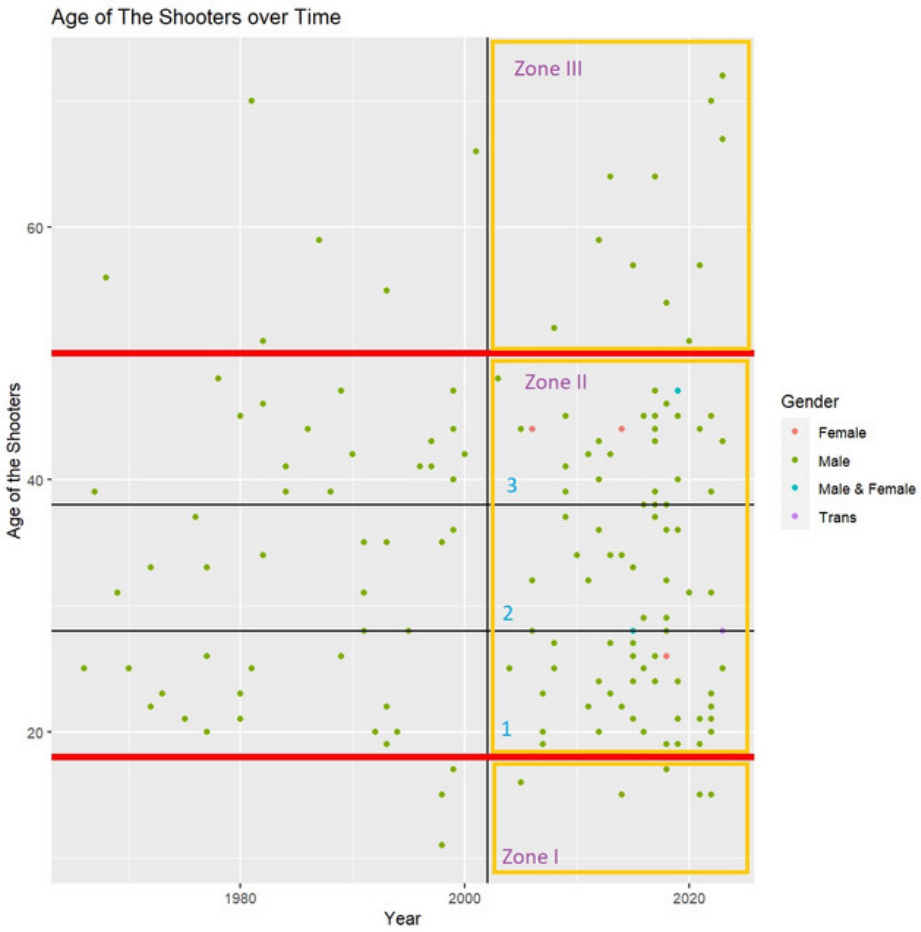
Race and Mental Health Issues



Race	Prior Mental Health Issues
Asian	90%
White	69%
Latino	67%
Black	43%

Will The Shooter's Age Be Related over Time?





+There are 3 things we can spot instinctively that

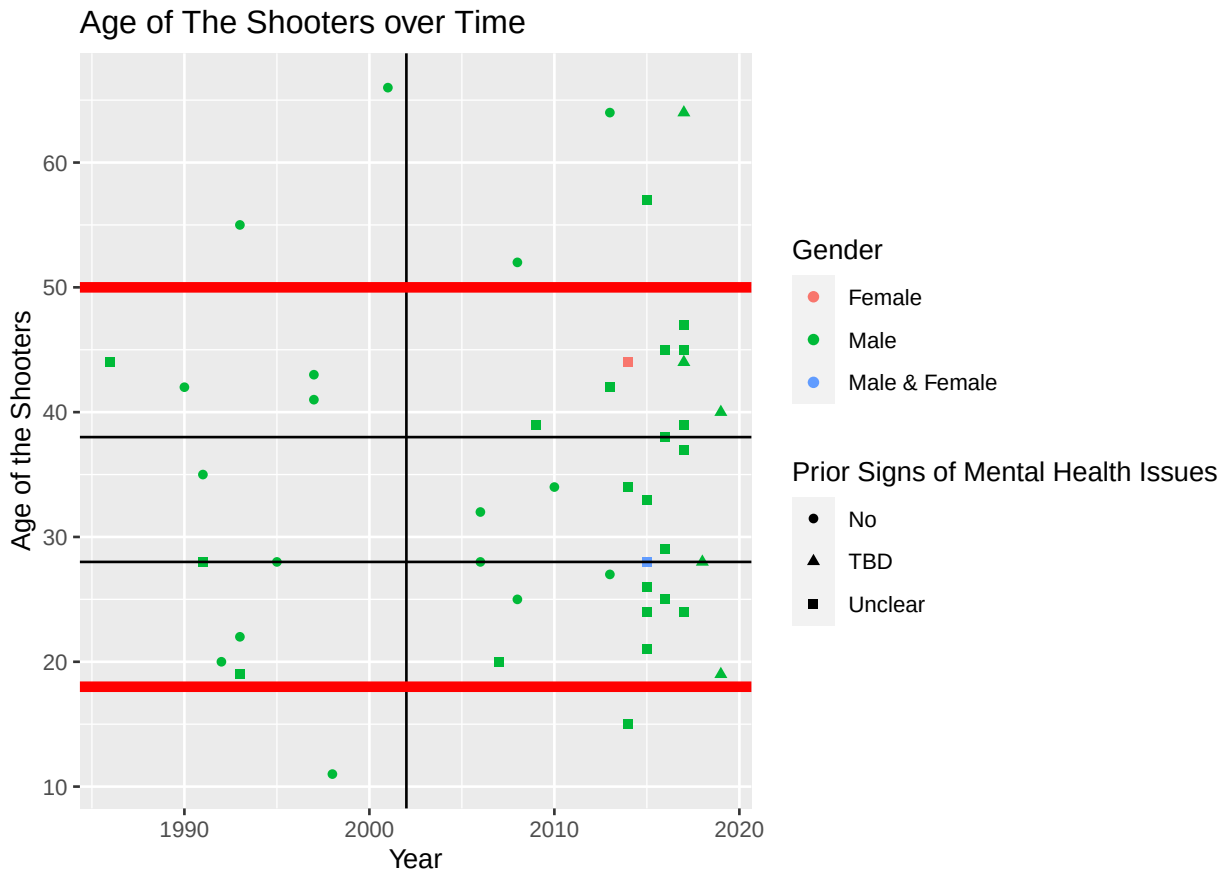
- The right-hand side of the first zone and the third zone yield more dots than the whole bigger left-hand side one. Those zones are the least expected ages for the incidents to happen, but they still have more cases than its left-hand side ones.
- The density of the right-hand side in the second zone is more condense than the left-hand side, even the area is smaller but the dots represent heavier than the other side. Zone II is the main labor force of the market, and it will be divided into three sub-areas.
- Take a look back to the whole picture, the right-hand side seems to be outnumbering to the left hand side, which means the number of cases has rapidly increased for last two decades.
- Take a closer look to the Zone II; I divided it into three sub-zones:
  - Sub-area 1: Young people with no experiences either in life or at work tend to get shocked and get impulses if they got history of abuses. They got full energy to do what they want to do. They will make tremendous mistakes with that source of energy unless they have well-guidance from older adults.
  - Sub-area 2: Most of people in this area got a career to follow, a family to take care of and some life's experiences. Health is at peak. Some of the conflicts at work might trigger their impulses, but in general they can hold it through.
  - Sub-area 3: In this age's range, an adult is mature in life both financially and mentally, even the health is in downturn. Kids got older, career is more stable. Anybody in this age got something to lose, so they normally do not do anything irrationally unless they got some mental health issues. That probably explains why there are no shooters without prior mental health in this range.

With Prior Mental Health Issues

The incidents in which the shooter had prior mental health issues have plotted as the plus sign (+) on the plot above.

Without Prior Mental Health Issues Plot

Now we take it off the plot to see how the original plot looks like.



Compare to the original plot, we can see intuitively the dots' density was reduced significantly. Hence, we are going to find the difference between with and without prior mental health issues by numbers.

Realize Intuitive Consideration by Numbers

We can draw some remarks by spotting the plots. Now we are going to consider some numbers from the data to see how much the `mental health issues` contribute to the problem.

```
my_data %>%
  filter(prior_signs_mental_health_issues == "Yes") %>%
  group_by(year) %>%
  nrow()
```

```
## [1] 80
```

If we filter out the cases with the prior mental health issues, there are eighty cases was off the chart, which is **almost half of cases of mass shooting in the US since 1960**.

```
my_data %>%
  filter(weapons_obtained_legally == "No") %>%
  group_by(year) %>%
  nrow()
```

```
## [1] 16
```

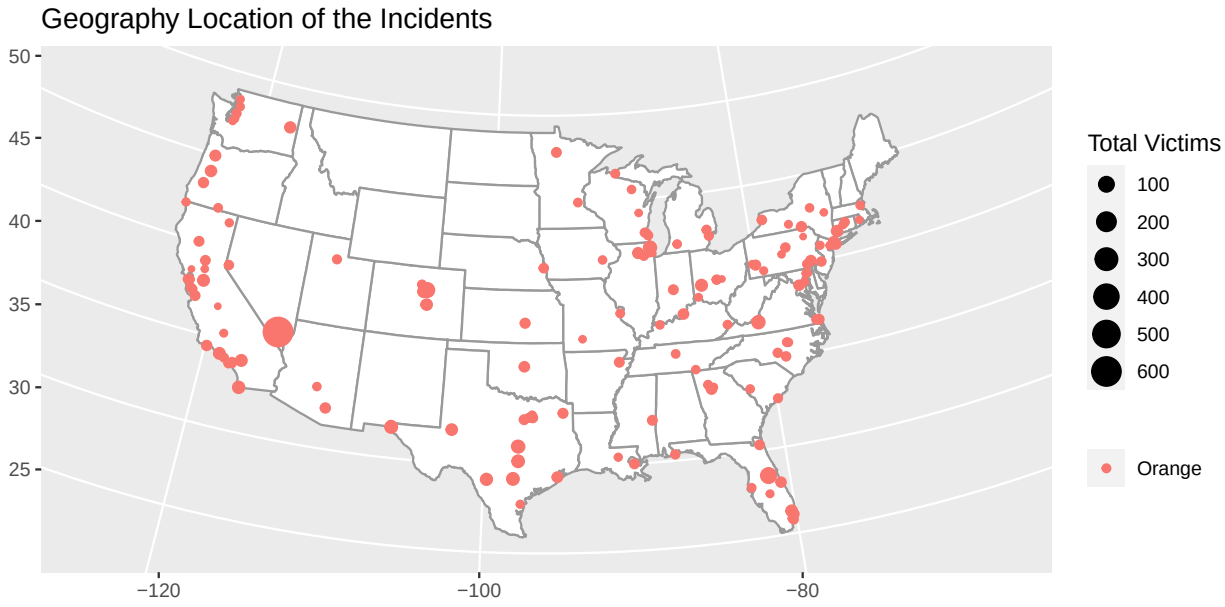
In a different case, I cross off the `legal weapons obtained`, only 16 cases was off the chart, which roughly 10% of all of the cases.

Conclusion:

People always argued about either we should do the background check or adjust the law over gun control. Now we can state that background check is more important than gun control, especially background check on mental health issues is crucial. Decreasing the cases down to fifty percent is ideal, but twenty or thirty percent down is sufficient to save many lives.

Prior Signs of Mental Health Issues	Age of the Shooter
Yes	around 23 and 40
No	around 30

Geography Graph of The Events



The graph shows us an idea that the incidents most likely occurs over the East and West side of the country, and the Mid-west is least likely to happen the mass shootings.

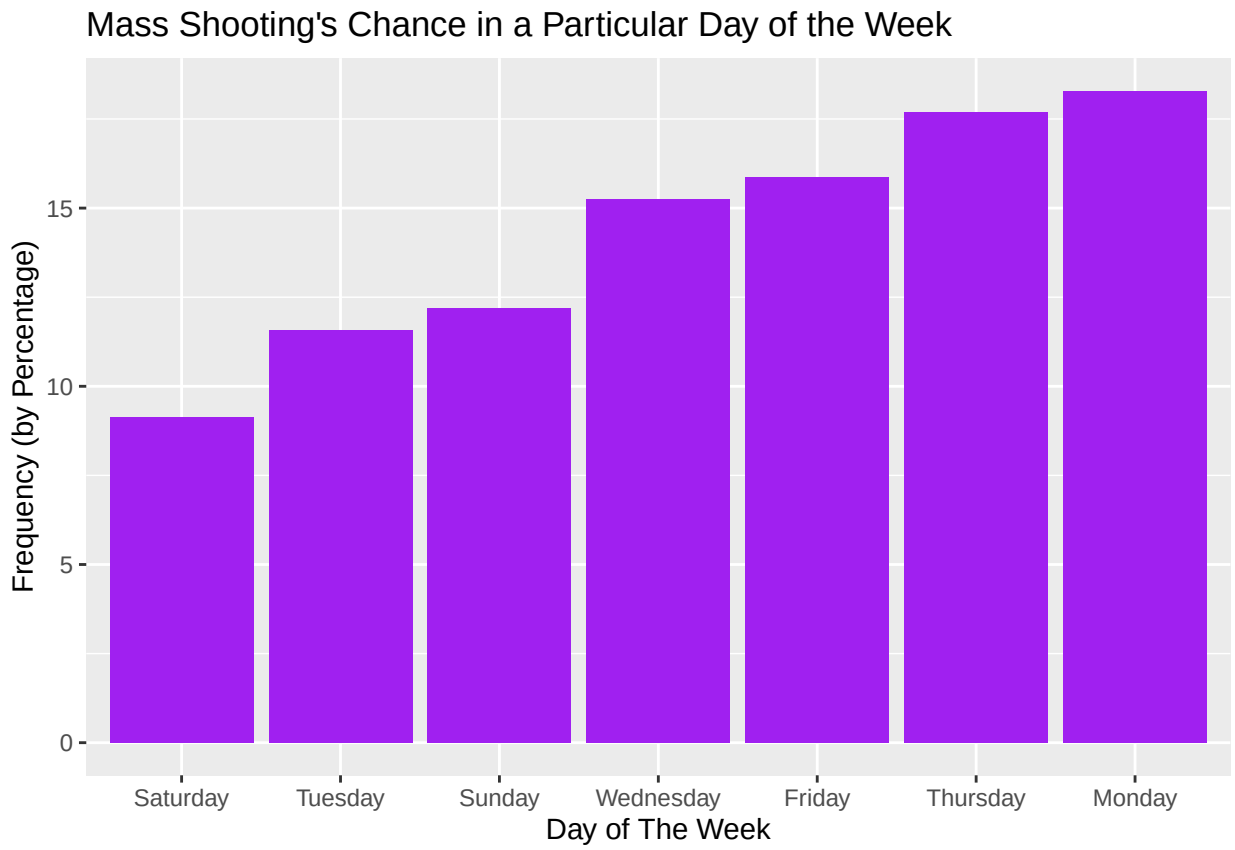
```
## # A tibble: 40 x 3
##    State Cases Percentage
##    <chr> <int>      <dbl>
##  1 CA       28      17.1
##  2 TX       16       9.76
##  3 FL       12       7.32
##  4 PA        8       4.88
##  5 CO        7       4.27
##  6 NY        7       4.27
```

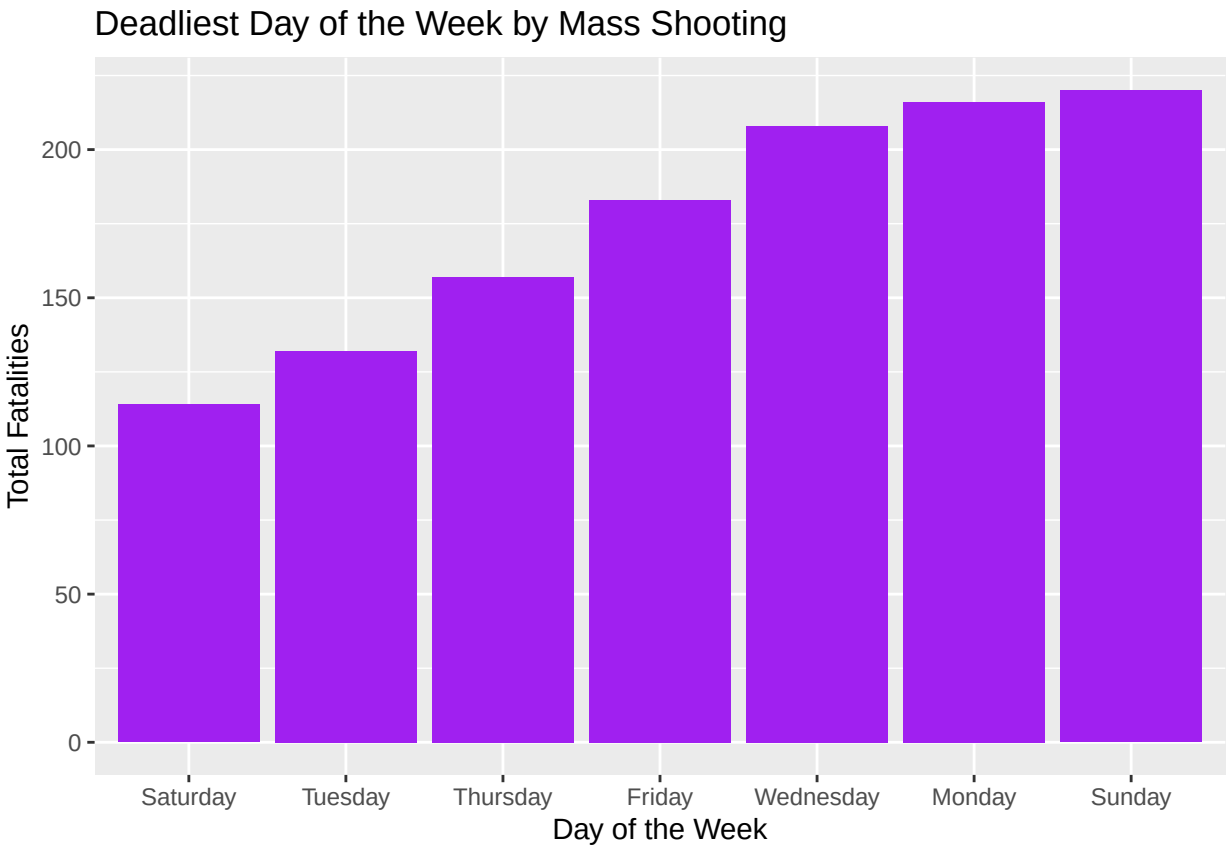
```
##  7 WA        7       4.27
##  8 IL        5       3.05
##  9 MI        5       3.05
## 10 WI        5       3.05
## # i 30 more rows
```

Colorado is the state in top 5 rating of mass shooting even the population rank is not in top 20 nationwide. Massachusetts surprisingly has no records on mass shooting even the population is in top 16 nationwide. State Population Source: https://www.statsamerica.org/sip/rank_list.aspx?rank_label=pop1

Pick Your Day to Go Out.

```
## # A tibble: 7 x 3
##   `Day of the Week` Count Percentage
##   <chr>             <int>      <dbl>
## 1 Monday               30      18.3
## 2 Thursday             29      17.7
## 3 Friday               26      15.9
## 4 Wednesday            25      15.2
## 5 Sunday               20      12.2
## 6 Tuesday              19      11.6
## 7 Saturday             15       9.15
```





Sunday is the deadliest day of the week in term of Mass Shooting but Monday is the most likely day for the Mass Shooter plan to act.