DATA 2401

Author: Tuan Pham

Instructor: Dr Katherine Shoemaker

4/25/2023

**The Final Project**

# Slide 1

# Slide 2

The gun violent is always a controversial topic, and it never ceases to be one even though we had to witness the terrific events like mass shooting with a higher density.

During my presentation, I do not try to seek one or many solutions for this matter, but I want to give you through this data an edge of the problems, a tip of the iceberg with my limited skills and knowledge.

I hope my work here will raise your insights about the problems, and then you can draw conclusions for yourself in order to protect you and your family. You can recognize the primitive signs, not to live in skeptics but to improve your awareness.

# Slide 3

# Slide 4 - Data

- To conduct this presentation, I use the data from \*\*Mother Jones website\*\* <https://www.motherjones.com/politics/2012/12/mass-shootings-mother-jones-full-data/>

The data's authors are MARK FOLLMAN, GAVIN ARONSEN, and DEANNA PAN.

I chose this data set because it is related to my subject and more important it's up to date. Everybody likes up-to-date things, right? The time span they collect is from 1982 to present, which included `originally 143 observations and 25 variables`. This is my work's mainframe.

- ...but I want a broader picture for us, so I look up for more data, and found a bigger source of data about this subject at \*\*The Violence Project website\*\*

<https://www.theviolenceproject.org/>

* They did not show any author's names over the website, so I just assume this data set is the collaborating work of the team
* The limitation of my work in term of collecting data is that two data sets I would introduce to you have different structures, so it was impossible for me to use the Join technique to combine data.
* Hence I picked up the `21 observations` from my second data set to plug it manually to the first one. These observations are about 20 years extended from 1960 to 1982. Therefore it makes my complete data set `\*\*164 observations and 26 variables\*\*` from 1960 to 2023.

# Slide 5 - Data (cont.)

The both data set is in form of csv file.

As I mentioned earlier, my framework is on the Mother Jones Database because this data set look more familiar with what we have been learned so far.

Since the second data set has more specific information, so the structure they have build has multiple layers hierarchically nested to the main variable. Some variables are coded by numbers, and you have to look up the including note in different data sheet to decode. They have coded races by number from 1-6 for instance.

I found that appropriate for time-consumed researches in reality, not for a limited-time presentation. By the way, I had read and picked up the interesting variables that we could find it interesting, and put it on my main data set.

# Slide 6

Most of the data variables are pretty straightforward as you can see.

# Slide 7 - Data Key Terms

The packages I use "obviously" the tidyverse.

I use this package for data wrangling and plots.

# Slide 8 –

# Slide 9 – Data Wrangling

- Look up the non-character variables and change it to integer's one. The raw data set I got has all the variables' types as character for some reason.

- Fix up some typos and remove the new lines to prevent it from creating extra categories of same criteria.

# Slide 10 – What’s the Mass Shooting?

# Slide 11 - A First Glance about the Incidents over Years

Quick intuitively look over the graph, we can see three points that stood out

* First, the graph skewed intensively to the right, which means incidents happen more more in the last two decades.
* Second, there’s a sharp drop in 2020, which we can easily explain by the Covid-19 lockdown in the same year.
* Finally, there’s no cases reported as mass shooting in 2002, and a supremely low number of case reported before 2004, the year which marks the Assault Weapons ban was expired. After that year, the incidents surge prolifically.

# Slide 12 – Data Summary

* When I examine this summary table, I spotted a ‘wild’ number with a case of more than 600 victims, so I filter it out to see what that horrific event was, and here I found out.
* Another case that stands out, which caught my attention is the minimum and maximum age of the shooter is 11 and 72 respectively. So I looked up for what’s going on for that case as well.
* These three observations are the outliers for each of the variables that they represent.

# Slide 13 – Data Extremes

# Slide 14 – Who Are They?

There’s some missing observations here, listed as NA, and some unclear details about the race.

When I tried to find out there's any connections about the age and race of the shooter, I expected this might be the least related subject, then the information it's provided gives me some insights.

* 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 4 decades before 2002, and only 2 decades after 2002.

# Slide 15 – The Age and Race relationships

# Slide 16 - Where Do The Mass Shootings Likely Occur?

* After filter the data, there are not so many details that been shown, but School, Airport, Workplace, Military, and a “big” category named ‘Other’. So I start doing some researches on cases categorized as ‘Other’ to find out where the incidents actually took place.
* There’s a huge amount of cases in the ‘Other’ category that I called ‘Nightlife’, which included Bar, Club, Restaurant or Music Concert. I do not know if the author of this data did this on purpose but they had classified most of the ‘Nightlife’ as ‘Other’.

# Slide 17 – Location and Frequency

School takes the second place in the list, which means a lot of innocent victims are kids.

Limited information about the observations in `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 'Other' in its categories.

# Slide 18 – What Weapons Were Likely Used by The Assailants?

There’s a little difference between my coding result and the constructed table above because the result from the code just represents the weapons as a combination. For instance, if the shooter carries the combo of one semi-auto handgun and one semi-auto rifle, the percentage of semi-auto handgun in that combo still exist in the semi-auto handgun. If the data was constructed as separated variables, both results would be matched.

# Slide 19 – Race and Mental Health Issues

# Slide 20 – Will the Shooter’s Age Be Related over the Course of 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. Explain

- The density of the right-hand side in the 2nd zone is more condense than the left-hand side, even the area is smaller but the dots seems to be more intense than the other side. Explain

- Take a look back to the whole picture, the right-hand side seems to be outnumbering to the left hand side. Explain

- Why did I choose the three milestones?

By September 13, 2004, the Assault Weapons ban was lifted, which marks 10 years it was enacted. The mass shooting incidents grow prolifically. Coincidence?

Eighteen - It's too old to be considered as "know nothing" or too young to be called "know something”. Moreover, this is the age that you are accountable for your acts as an adult. Even though, this is a really confused time if you do not have a great guidance from the older adults, especially you suffered a longtime abuse since your childhood, an impulse act would be easily triggered, and a human life will change its course forever.

Eighteen to Fifty - This is the current and future human resources for the main labor force of the country. I would split this area to three smaller sub-areas to see if there are any conclusions I can draw out.

# Slide 21 – With Prior Mental Issues

# Slide 22 – Without Mental Health Issues

# Slide 23 – Compare on Real Numbers

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

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.

# Slide 24 – Conclusion

Background check on Mental Health Issues is crucial.

# Slide 25 – Age of the Shooter with prior Health Mental Issues

# Slide 26 – Gender and Age of the Shooter

# Slide 27 - Gender and Age of the Shooter Conclusion

97% of the Mass Shooting incidents carry by male.

# Slide 28 – 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 for the mass shootings.

In my opinion, the edge sides of the country got the population per capita are higher, more people living and bigger cities, so the frequency of the mass shootings is likely higher.

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.

# Slide 29 – Pick Your Day to Go Out

* 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.
* For my opinion, it’s the first day of the week and sometimes it overlapped with the national holiday. The perpetrators have had a long weekend to plan for carrying out the attacks.

# Slide 30 – Conclusion

A Little Thought for This Problem

- Gun control is never an easy topic to debate about, and it draws more and more opposite opinions since incidents keep occurring more frequently nowadays. Some said that gun control is against the second Amendment of US Constitutional. Others believe the freedom of carrying weapons will just make the America wounds more severe. Will the greater scale of recent tragic events ring any bells to the legislators or any law makers?

- Due to the complex characteristics of the firearms’ problems, the Congress still hesitates to make laws to tackle it. Laws will evidently create conflicts to the corporations’ benefits and defy the opposite citizens’ opinions. The more time to waste, the more people’ lives to cost. A gun reform is urgent.

- In my opinion, the Constitution was written hundreds of years ago when firearms were able to shoot once per round. It is obsolete when we talk about the current firearms, which can go for more rounds, create more damage and reach the target more accurately. We will not ban guns, which obviously goes against the Constitution, but we will also not let it circulate uncontrollably. We can meet the controversial idea half-way. Not banned but stricter gun control laws.