Introduction:

Throughout the years, the United States of America has seen an increase in violence through mass shootings. Since the Sandy Hook Elementary School shooting, we have seen the numbers nearly triple in size. While, yes, the country should look at better gun control policies to solve this issue, as data scientists we can look at the data that we have collected from these shootings over the past 10 years to try to prevent these from happening in the future. We can look at different factors such as demographics, socioeconomic status, and the effect that mental illness can have on these shootings. This is important for individuals to be interested in because it is something that can potentially help reduce the number of shootings or hopefully eradicate them completely at some point in the future. With the amount of data that has been compiled since this great increase 10 years ago, it can now be a problem addressed through the help of data science.

Research Questions:

1. What is the choice of weapon for these shootings?
2. Do the shooters have a criminal background or are they first time offenders?
3. Which gender is more prone to commit this type of crime?
4. Where do most shootings take place? What is the most frequent place?
5. Does race play a role in mass shootings?
6. Does mental illness play a role in mass shootings?

Approach:

This approach will not fully address the problem but rather will serve as a means of attempting to predict criminal behavior based on past experiences. By gathering and compiling the data of previous mass shootings, we will aim to create a guide of what to look for or what to flag as suspicious activity. This will assist with helping to provide enough police presence in areas that are more prone to attacks. It will help create a log for legally obtained weapons and who purchased them in order to better track these. The biggest thing this approach will aim to accomplish is determine the correlation (or lack thereof) between mass shootings and individuals with mental illness in order to persuade and implement a program of mental health evaluation prior to firearm sales.

Data:

1. <https://www.kaggle.com/datasets/zusmani/us-mass-shootings-last-50-years>

This dataset looks at mass shooting attacks in the United States of America between 1966 and 2017. The dataset contains Serial No, Title, Location, Date, Summary, Fatalities, Injured, Total Victims, Mental Health Issue, Race, Gender, and Latitude and Longitude information. The data set was posted to Kaggle in 2022 and has been modified multiple times to add new variables, add new data, and add missing data in order to help create visualizations and extract patterns. Similar to my project, the data set aims to aid in predictions for future events and prevention of such.

1. <https://www.kaggle.com/datasets/carlosparadis/stanford-msa>

This dataset attempts to facilitate research on gun violence in the US by making raw data more accessible. This dataset is comparable to the previous one except it has a bit more detail in the mental illness variable which can be very useful with the project.

1. <https://www.kaggle.com/datasets/twinkle0705/mental-health-and-suicide-rates>

The goal of using this dataset is to determine if a correlation exists between mental health and suicide rates. The majority of mass shootings end with the perpetrator being shot down by the police. Because of this, it stands to reason that mass shooters go into the situation not expecting to come out alive (i.e. go in as a suicide attempt while taking people out with them). This dataset compares countries, their resources for mental health, their mental health facilities as well as their suicide rates.

Required Packages:

From what I know so far, the packages required to load and compare these datasets are as follows:

library(ggplot2)

library(dplyr)

library(readr)

library(tidyr)

library(stringr)

library(reshape2)

library(ggthemes)

library(lattice)

library(knitr)

Plots and Table Needs:

Scatterplots, histograms, and boxplots. I would also like to incorporate a map graph to depict the rates per state and an area chart to show the change over time of mass shootings. As far as tables go, I would like to use the simple XY table as well as column tables and grouped table to go a bit more in depth for comparison purposes. A multiple variables table might also prove to be beneficial for displaying data and making it easier to understand for the reader.

Questions for Future Steps:

The main thing that I do not right now that I will need in order to analyze and interpret the data is how to create the correct tables and plots as described above. I am sure that there will need to be other things to learn but right now, I do not know that that is. For the most part, I feel confident in the datasets I have collected as well as with the learnings we have covered so far to be able to compare the datasets and come to a conclusion that could possibly assist with predicting the who, what, where, how, and why of mass shootings in the United States.