Team members
- Florencia Froebel Guillen , Emily Huang , James Kelly

Github repository
- UC-Berkeley-I-School/Project2_Froebal_Huang_Kelly (github.com)

Primary dataset
- **U.S. Gun Violence Records 2014-2021 | Kaggle**

Data structure

There are 3230 observations and 7 variables. The variables are:

- Incident_id
    - EX: 2128023
    - This will help us uniquely identify each incident
- Incident_date
    - EX: July 3 2021
    - Using this variable we can analyze **when** incidents tend to occur the most
- State
    - EX: Illinois
    - Using this variable we can analyze **where** incidents tend to occur the most
- City_or_county
    - EX: Chicago
    - Using this variable we can analyze **where** incidents tend to occur the most
- Address
    - EX: 749 Silver Bluff Rd
- Killed (Integer)
    - EX: 0
    - Using this variable we can differentiate between different gun violence incidents where someone was fatally shot. This is important because fatal gun violence incidents may be less common than gun violence injuries
- Injured (Integer)
    - EX: 2

Out of these variables, we think **incident_date**, **state**, **city_or_county**, **killed**, and **injured** will be interesting for analyzing gun trends.

Initial exploration

Initially we can explore each variable of interest in more depth by determining the mean, median, and mode values of each. We will also plot histograms of the variables to better understand their distribution. We will also want to format all of the data for us to use easily by, for example, formatting all of the dates into date objects for easier manipulation. We will also want to take a look at the unique values under the State variable and confirm they are correct.

What we plan to cover in the final report

We think it's worth exploring gun violence trends from the data set which would include exploring questions like:

1. A state's gun violence vs. National gun violence statistics
2. What are the states with the most incidents (injured and killed) due to gun violence?
   a. Are there states that have more injuries but not deaths? Vice versa?
3. What are the cities with the most incidents (injured and killed) due to gun violence?
   a. Are there cities that have more injuries but not deaths? Vice versa?
   b. How does a city and its state compare in gun violence incidents? What are those factors?
4. What dates/time period do most gun violence (deaths or injuries) occur generally?
   a. How does the date affect a city's or state's gun violence? Does it have any effect in general?
   b. Does the date affect the mortality rate of victims in a gun violence incident (i.e. are victims more likely to be injured than killed if it's around the summer, etc.)?
   c. Are there cities/states that differ from the general time period of frequent gun violence incidents? How so?