

# Applied Problem Set 3

Fernanda Sobrino

11/19/2021

Due Friday December 3, midnight Central Time.

Submit by pushing your code [here](#)

Remember to upload your pdf to Canvas.

This submission is my work alone and complies with the 30535 integrity policy.

Add your initials to indicate your agreement: \*\* \_\_ \*\*

Add names of anyone you discussed this problem set with: \*\* \_\_ \*\*

Late coins used this pset: 0. Late coins left after submission: X.

Download the data for this Pset from [here](#). Just click on Download and it should download automatically to your computer. Be careful when pushing back your work into github, this file is over 100MG so you won't be able to push your pset if the data is still there. This data set contains more than 260K gun violence incidents, with detail information about each incident The data is from [gunviolencearchive.org](#).

In the questions where you are asked to produce a graph or a map. Make sure that:

1. you add an informative title
2. your axis have informative names
3. graphs are not the default color (whichever you pick is fine)

Points will be take off if your graphs and maps do not follow this.

## 1 Load the data and first glimpse (5 pts)

1. Load the data and show how many rows and columns it has
2. Explore the variables on your own (we do not want to see any code here) and write a short paragraph of what you find. Pay attention on which characteristics we have for each event.
3. Which variables you might have to format later? (just name them you do not have to do anything else right now)

## 2 Time ralated trends of gun violence (10 points)

1. Make sure the `date` variable is in a Date format. If not transform it.
2. Is the number of incidents increasing by year? Show your result in a graph and write a short paragraph about what you are seeing. Add the count for each year as an annotation label.
3. Is there a particular violent month? Answer this question with a graph. Be careful: should you use the whole data to answer this question? Add the count for each month as an annotation label.
4. Is there a particular violent day of the week? Add the count for each day as an annotation label.
5. Write a short paragraph with your findings on the number of incidents by year, month, and weekday.

### 3 Characteristics of Gun violence incidents (20 points)

**WARNING!!** some variables in this data set need to be manipulated before you try to plot them. It might be a good idea to go back and check your notes on Strings, Tidy, and Joins.

1. What is the average number of guns involved in an incident?
2. Which type of guns are more commonly used? Use a plot to show your answer. Do not show `Unkwon` or missing values on your plot
3. Explore the `incident_characteristics` variable. What is this variable telling us? Why are there multiple characteristics for each incident?
4. Show in a plot which are the most common `incident_characteristics` cap your graph at 20 (i.e just show the 20 most common `incident_characteristics`)

### 4 Suspects Characteristics (30 points)

1. Explore the following variables and write a short paragraph of what they mean and how they are connect to each other: `participant_age`, `participant_gender`, `participant_type`, and `participant_status`. Do you see any technical difficulties for how these variables are coded? If so, explain them.
2. What is the average of suspects and victims per incident?
3. Create a new data frame with just the suspects include the following variables: `incident_id`, `participant_age`, `participant_gender`. Just print the `head()` of this new data frame. (Hints: 1. Google what the function `unnest` does, is part of `tidyr` library, 2. review the tidy and joins materials)
4. Show the distribution of suspects age, crop your plot if you find any suspect over the age of 100
5. What percentage of suspects are male (exclude missing values)?
6. How many different status are there?
7. What percentage of all suspects got arrested? Be careful for some suspects there are more than 1 categories.

### 5 Geographic variation (15 points)

1. What was the state with more incidents in 2017? Use a graph to answer this question.
2. Use your census API to get population by state, remember to also download the `geometry` we will use it later. Re-do your previous plot but adjusting by population, i.e incidents by a 100,000 inhabitants.
3. Show the results from your previous plot in a map

### 6 Mass shootings (20 points)

1. Create a new data frame of mass shootings.
2. Show the top 15 incidents by number of victims in a map as points. The points should be proportional to the number of victims. Follow the map slides to make the map. Here are the coordinates that will give you the continental US out of the world map. Extra credit if you add state lines and change the theme of the map.

```
coord_sf(xlim = c(-124.6813, -67.00742),  
         ylim = c(25.12993, 49.38323), expand = FALSE)
```

3. Use the data set you create about suspects in previous sections to explore if there are differences between the suspects of mass shootings vs other types of incidents. Compare suspects age, gender, percentage of arrested, percentage of killed and number of shooters/suspects per incident. Write a short paragraph with your findings.