STAT 201: Midterm 1 Extra Coding Practice 1

Your name

Load your packages necessary for wrangling code, plotting, and creating beautiful tables:

The data containing the records of gun violence incidents are found in `gun_violence`. The data are obtained from Kaggle, and have been slightly modified. The variable definitions are as follows:

- `incident_id`: ID number
- `date`: date of incident
- `state`: state in which incident occurred
- `city_or_county`: city or county in which incident occurred
- `address`: address where incident occurred
- `n_killed`: number of people killed
- `n_injured`: number of people injured, but not killed
- `n_guns_involved`: number of guns involved
- `doy`: day of year (month/day) of incident
- `year`: year
- `month`: month

Question 1

In which five cities or counties did the most incidents occur? What are the probabilities that a randomly selected incident in our dataset occurred in each of these cities/counties? Answer this by creating a beautiful table that contains this information.

Question 2

While there are a variety of definitions of mass shootings, one definition is "any incidents in which four or more people were shot, whether injured or killed".

Based on this definition, create a beautiful table that displays the top six incidents that meet this definition of mass shooting in descending order of number of people shot. Only display the date, state, city or county, and number of people.

Question 3

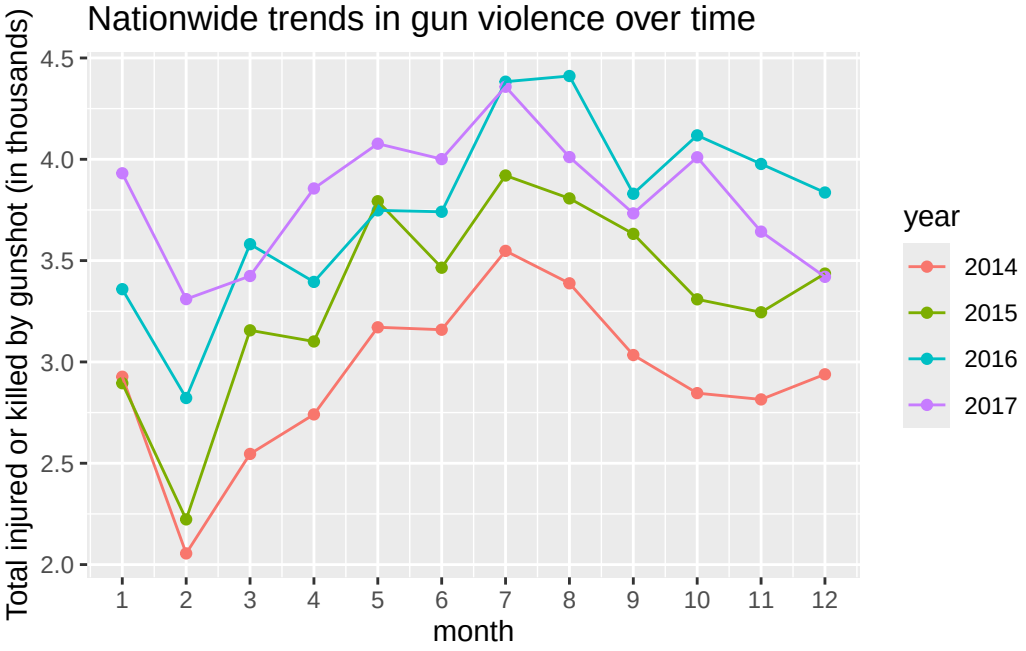
Create a data frame called `state_incidents` where each row is an observation of a state and its total number of incidents in a given year.

Then using your new data frame, answer the following using a single table: For the combined years 2014-2017, which three states had the lowest average number of incidents? Which three states had the highest average number of incidents?

Question 4

Re-create the following plot to the best of your abilities. The following code will be helpful to you:

- If you'd like to turn a numeric variable into a character variable, you can use the `as.character(<variable>)` function. If you'd like to turn a character variable into a numeric variable, you can use the `as.numeric(<variable>)` function.
- `scale_x_continuous(breaks = 1:12)` makes a numeric x-axis have tick marks at the values 1-12.



Then interpret what you see.