STAT 201: Practice Midterm

Your name

TYPE THE HONOR CODE PLEDGE IN THE CORRESPONDING AREA:

"I, <your name>, have neither given nor received unauthorized aid on this assignment."

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

We will also be using data about state populations The data are found in `census`. These data come from the U.S. Census. The definition of variables in this file are as follows:

- `state`: state
- `code`: 2-letter state code
- `year`: year
- `population`: population for the state in the given year

Lastly, we have information about the number of gun laws in each state in 2017. These are found in `laws`, with variables as follows:

- `state`: state
- `gun_laws2017`: number of state gun laws in 2017

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, and interpret what you find in a sentence.

Answer:

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.

Now would be a good time to render.

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?

Answer:

Question 4

Combine your `state_incidents` data frame with the other two datasets `census` and `laws` and save the result into a new dataframe called `state_incidents2`. After this step, you should have 6 variables in `state_incidents2`.

Now would be a good time to render.

Question 5

Using your new dataframe, create a new variable called `rate` that is equal to the number of gun violence incidents per 100000 people in a given year. This is calculated as taking the number of incidents divided by number of residents, multiplied by 100000. Save this into the same data frame `state_incidents2`. Then answer the following question using a single table:

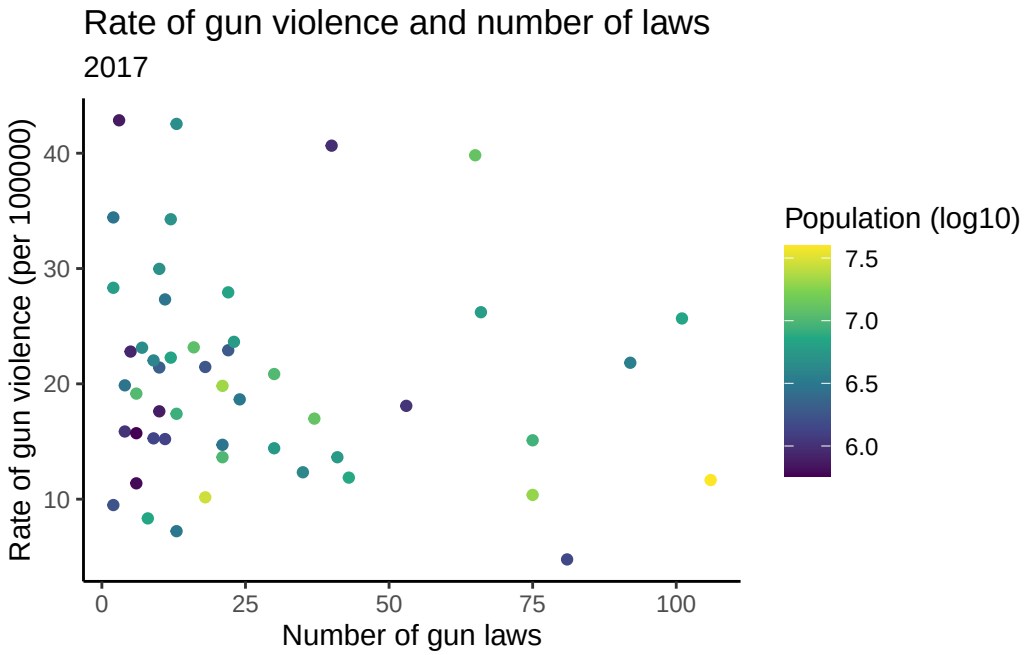
For the years 2014-2017, which three states had the lowest average rate of incidents? Which three states had the highest average rate of incidents? Comment on how your findings here compare to your findings from Question 3.

Answer:

Question 6

Re-create and interpret the following visualization. Tips:

- To take the log base 10 of a number, we can use the function `log10()`.



Answer:

Render and submit the PDF to Gradescope.