STAT 201: Practice Midterm

Your name

TYPE THE HONOR CODE PLEDGE BELOW

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

```
library(readr)
```

The data containing the records of gun violence incidents are found in the `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 and gun laws. The data are found in the `census`. These data come from the U.S. Census and gun laws. The definition of variables in this file are as follows:

- `state`: state
- `code`: 2-letter state code
- `2014`: estimated state population in 2014
- `2015`: estimated state population in 2015
- `2016`: estimated state population in 2016
- `2017`: estimated state population in 2017
- `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 some beautiful tables, and 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.

Question 3

Let's examine which days of the year are most common for gun violence. Display a table of the top three days of the year with highest number of total incidents, displayed in descending order. Comment on what you notice.

Answer:

Question 4

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. For the 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 5

Combine the three datasets into one and save the result into a new dataframe called `state_incidents2`. After this step, you should have 6 variables in `state_incidents2`.

Question 6

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:

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 Exercise 4.

Answer:

Question 7

Create groupings for the state populations for easy visualization. Add a new variable called `pop_class` that groups the states into the following categories. Don't forget to store your results.

```
- "Extra-small" = populations less than 1.5 million residents
- "Small" = population of at least 1.5 million, but less than 5 million residents
- "Medium" = population of at least 5 million, but less than 10 million residents
- "Large" = population of least 10 million residents
```

Question 8

Create a visualization that displays the distribution of the rate of gun violence incidents for each year and by the population class. Describe any trends you notice.