Dataset Presentation

1. Describe your dataset. Explain its attributes.

   The title of the dataset is called “American University Data” IPEDS dataset from Kaggle, containing information about colleges in the US. The original dataset consisted 1534 observations and 145 variables. The dataset after tidying consists of 1534 observations and 27 variables. Variables include enrollment rate, graduation rate, and percentage of first-year students receiving financial aid in 2013, etc. (more details below)

   The following explains the variables in more detail:
   - Name: (chr) name of the college – categorical
   - High.Degree: (chr) the highest degree that can be obtained from that college – ordinal
   - Longitude: (num) longitude location of institution - quantitative (-158.062 - -67.4565)
   - Latitude: (num) latitude location of institution – quantitative (19.70185 - 64.85705)
   - Applicants: (int) number of applicant – quantitative (0 - 72676)
   - Admissions: (int) number of admissions offered – quantitative (0 - 35815)
   - Enrolled: (int) number of enrollment – quantitative (0 - 10241)
   - SAT.Reading.25: (int) SAT critical reading 25th percentile score – quantitative (310 - 720)
   - SAT.Reading.75: (int) SAT critical reading 75th percentile score – quantitative (360 - 800)
   - SAT.Math.25: (int) SAT math 25th percentile score – quantitative (280 - 770)
   - SAT.Math.75: (int) SAT math 75th percentile score – quantitative (410 - 800)

- ACT.25: (int) ACT composite score 25th percentile score – quantitative (13 - 33)
- ACT.75: (int) ACT composite score 75th percentile score – quantitative (14 - 35)
- Percent.Admitted: (int) acceptance rate – quantitative (6 - 100)
- Tuition: (int) tuition and fees – quantitative (1032 - 49138)
- Tuition.Instate: (int) total price for in-state students living on campus – quantitative (9768 - 64988)
- Tuition.Outstate: (int) total price for out-state students living on campus – quantitative (9768 - 64988)
- State: (chr) state the institution is located – categorical
- Region: (chr) region the institution is located – categorical
- Control: (chr) private/public – categorical
- TEnrollment: (int) student population (undergrad) – quantitative (0 - 51333)
- Percent.women: (int) percent of undergraduate enrollment that are women – quantitative (0 - 100)
- Percent.Instate: (int) percentage of in-state freshmen – quantitative (0 - 100)
- Percent.Outstate: (int) percentage of out-state freshmen – quantitative (0 - 100)
- Percent.Foreign: (int) percentage of international student – quantitative (0 - 40)
- GradRate.4yr: (int) graduation rate within 4 years – quantitative (0 - 100)
- Percent.Finan.Aid: (int) percent of freshmen receiving any financial aid – quantitative (18 - 100)

Limitations:

- Missing info
- Since the dataset is from 2013, it is outdated and some of the information extracted will not be applicable today
- Tuition: price will be different due to inflation
- Applicants: as more schools provide test-optional policy and students have more information about college than 9 years ago, the number must have increased

2. List your questions and which attributes you will use to help you answer them. What are you going to visualize?

   Questions:
   - Do colleges that have higher percentage of students coming from out-state have higher ACT/SAT scores?
     - Attributes: SAT.Reading.25, SAT.Reading.75, SAT.Math.25, SAT.Math.75, ACT.25, ACT.75, Percent.Instate, Percent.Outstate
     - Visualization: bar chart
   - Do colleges in certain regions charge more money to students than other regions?
     - Attributes: Region, Tuition, long, lati
     - Visualization: bar chart, scatterplot comparing two regions for every combination possible and faceting them, Mapping locations

   Trying to extract data from existing variables such as sat/act scores, the region they prefer to study in, in-state vs. out-state, and make a visual representation for users to find the appropriate college for them. Let the user specify their preference and output based on the selected/specified criteria.
   - Table
   - Implement a responsive site
   - Mapping locations