The Team   
Jatin Patel  
Nimisha Patel  
Margaret Ulrich

Problem Statement a.k.a Decompose The Ask  
Does ranking necessarily correlate to graduation rate, acceptance rate, tuition fees, SAT/ACT score requirements?

Identify Data Sources  
<https://www.kaggle.com/theriley106/university-statistics>

Data source will be from Kaggle.com. Kaggle.com has a data set of US World and News Report’s college ranking data of about 300 universities taken from U.S. News.

Data is available in CSV and JSON format from Kaggle.com.   
Supplemental data can be found at World Bank.

Define Strategy and Metrics   
Hypothesis 1: There is a higher concentration of engineering schools in SW vs. NE and vice-versa, there is a higher concentration of Business schools in NE vs. SW. To investigate this, we intend to visualize the relationships of Business and Engineering with their locations.  
Null Hypothesis 1: ­There is no evidence of a concentration of engineering schools in the Southwest, as there is no evidence of a higher concentration of Business schools in the Northeast.

Hypothesis 2: There is no correlation between the cost of attending highly ranked college vs. lower ranked college. To investigate this, we hope that a scatter plot of cost of tuition to ranking of school will demonstrate that this is not a relevant correlation.  
Null Hypothesis 2: ­There is a correlation between cost of tuition and ranking.

Hypothesis 3: Higher ranking colleges do not always require higher SAT/ACT scores. To do this, a simple scatter plot of college rank by SAT and college rank by ACT will suffice to demonstrate that the data either does point to a correlation or proves it does not.  
Null Hypothesis 3: There is evidence of college rank and higher SAT/ACT scores.

Hypothesis 4: NE colleges are more expensive than any other region. To show this, a heatmap of tuition-cost by location of college will be sufficient to prove the hypothesis.  
Null Hypothesis 4: College expense is not regionally specific.

Hypothesis 5: Colleges where more students receive aid are less prestigious. The “percent of students receiving aid” vs. the business ranking and the engineering ranking would be enough to demonstrate any such correlation.  
Null Hypothesis 5: Percent receiving aid and ranking in business or engineering are not related.  
(Since a few of our hypotheses involve only highly-ranked schools, it might be a good idea to find a data source of just those top schools, perhaps with more data points.)

Description of Data Analysis Tools You Plan to Use  
We plan to use Python’s PANDAS, MATPLOTLIB, NUMPY, gmap and other modules to extract and analyze the data.  
We will use PowerPoint to present the data findings.

Describe the Data Products Your Project Will Produce  
We will be using various statistical and analytical methods to illustrate our findings. Each of our hypotheses will require at least one visualization, as outlined above.