Statistical Analysis – Expectations

Goals:

- Provide statistical evidence for claims that you are making about your data
  - Hypothesis tests, confidence intervals
  - Randomization simulations, bootstrapping
- Uncover insights about the data
  - What stories does your data tell?
- Answer (or partially answer) your research questions about your data
- OPTIONAL – use your data to make predictions using machine learning techniques

Expectations:

- Create compelling visuals that show relationships between variables:
  - Continuous variables: scatter plot with trend line/curve
  - Categorical variables: contingency table with a visual (see http://www.sthda.com/english/wiki/chi-square-test-of-independence-in-r for some examples)
- Perform at least three hypothesis tests.
  - Use at least two different types of test; at least one must be a chi-square test
  - Include a confidence interval in your results for at least two of your tests
  - Use appropriate parametric or non-parametric methods depending on the claim you are investigating, variable type (categorical, quantitative), and test assumptions
  - Randomized simulation may be used when appropriate but is not required
  - Give a 1-3 sentence explanation of your results for each test
- OPTIONAL: Use machine learning techniques to make predictions using your data
  - Use appropriate ML methods depending on problem type, data, etc.
- Add any **necessary** text to explain visuals, results, data insights, research question (partial) answers, etc.
  - Text should be limited and in sentence form. No long paragraphs!

Deliverables:

- Code in Python or R.
  - Code should be commented appropriately, readable, and in a format that incorporates code, text, and visualizations (R Markdown or Jupyter Notebook)
- Deliverables should be well-organized and visual appealing
- Post your deliverables on your GitHub page in your capstone project repository.