Final Project Guidelines

CISC 5420 - Applied Statistics & Probability

Due Date: December 19th, 2017 - (emailed to me before 5:30 PM)

Late Penalty: For each day the project is late, I will deduct 1 full letter grade. This penalty begins if the project is not emailed to me by 5:30 PM on the due date i.e. a project received at 5:31 PM on the due date will score no higher than a B.

Format: I expect to receive an email containing a compressed folder. When uncompressed, this folder should contain the dataset being analyzed and a Jupyter notebook containing your analysis. I strongly encourage each of you to move large sections of code to separate files and import these into the Jupyter notebook. As I am grading your analysis (not the Python code), I encourage you to move your text descriptions and figures to a separate document that I can easily read. However, you're also allowed to hand in a notebook containing all the figures and analysis, if you prefer not to put together a separate document.

I will be opening these Jupyter notebooks and clicking 'Run all cells'. The notebook should run all the cells without errors. Any errors will result in point deductions. Any figures includes that are not generated by code will result in point deductions.

Plagiarism: Any code I suspect to be plagiarized will be investigated. If it is found that the student has plagiarized, the student will receive an F on the midterm project and will promptly be reported to the the Dean's Department as well as the Chair of the Computer Science Department.

Rubric:
1. Does the student's methodology make sense in terms of answering his/her research question? I.e. plots and analysis should go towards answering the student's question - not toward filling up space in a Jupyter notebook.
2. Has the student applied the methods from ThinkStats chapters 1-13 to answer the research questions? Specific attention will be paid to the methods from chapters 6-13.
3. How well does the student articulate his/her ideas and methodologies? Plots and analyses without clear, well described analysis are of little value.

4. Has the student performed at least one hypothesis test? Ideally, multiple hypothesis tests should be performed to answer the research question.
5. Has the student used linear regression or logistic regression?
6. Does the student come to some conclusion on his/her research question? What is this answer? How well is it supported. Please provide a substantive conclusion.