STATS60: Final Project

The project counts for 40% of your total grade, and has two components: a project proposal (worth 80% of the project grade) and a final project (worth 20% of the project grade).

Project Proposal (20%)

For the project proposal, submit a 1-2 paragraph description of your final project. You should provide details on the problem you wish to explore, any datasets that you will use, and potential visualizations. The proposal is a way for me to make sure that you have enough structure and data to complete the project.

The project proposal should be submitted as a text file (any format will do).

In implementing your final project, you should try to stick as closely to your proposal as possible. If you are unable to do so, please provide a paragraph in your final project to explain the deviations from the proposal.

Final Project (80%)

To demonstrate your understanding of the material we have covered, you will perform a data analysis project in R. **You may choose to do a project on your own or with up to two other teammates.** (If working with a teammate, only one person has to submit the work. The names and SUNet IDs of both teams should be clearly labeled.)

Here are the project minimum requirements:

- .Rmd file, as well as 1 .html file. (Don't worry about what these file formats are and how to create them right now; we'll talk about this later in the course.) Work with a dataset not used in class. Provide the data file and/or a link to it. My advice is to use a dataset that is in .csv or .txt format, as these will be easier to import into R.
- An "Introduction" section describing the problem and the datasets to be used. This should include top-level summaries of the data (e.g. how many observations and features, histogram/scatterplots of values, anything unusual).
- A "Data analysis" section where you have both code and accompanying text describing what the code is doing and interpreting the results. At least 3 data visualizations, each of a

different type. (If working with a teammate, at least 6 data visualizations, with at least 3 different types.)
- A “Conclusion” section to summarize the results of the analysis, as well as any deviations from the project proposal.

While your analysis should not be too skimpy, it does not have to be comprehensive either. Take a look at the links below for examples of what the final project could look like, as well as ideas for datasets which you might want to work on.

Grading and Deadlines

Both the project proposal and final project should be submitted through Canvas.

- The project proposal is due by July 15 (Wed), 5PM Pacific Time and is worth 20%.
- The final project is due by August 11 (Tue), 5PM Pacific Time and is worth 80%.

For each late day, a multiplicative penalty factor of 0.8 will be applied. **The final project cannot be submitted late**. If you foresee that you may have difficulty meeting these deadlines, please come and speak with me **immediately**.

What might the final project look like?

Here are some examples of good student projects from previous runs of this class:

- World Happiness Analysis, Clementine Chou.
- Analysis of Chocolate Bars, William Chow.
- California Election Analysis, Kyle Dixon.
- Street Trees in San Francisco, CA, Jessica Eggers.
- An (Incomprehensive) Analysis of the College Scorecard Data, Alex Hurtado.
- Oregon Timber Harvest, Sierra Killian.
- Wine Project Report, Miranda Li.
- SAT Score Analysis, Angela Luo.
- San Francisco Airbnb Analysis, Jenny Nova.

I don’t really have ideas for what data to use…

The Resources tab has several suggestions for you.
