Project Proposal Guidelines

Due: Wednesday, November 26, 2025 at 11:59pm

For your project you have a several options on what you can do. For example,

1. You could find a suitable data set to which you can apply a statistical learning method (but not just fitting GLMs) to analyze it and answer a "scientific question". You may consider methods we have covered in the course, methods that we have not covered in the class (e.g. from your presentation paper, or extensions of methods we covered, etc.).
2. You could fit different methods on a data set to see how they perform and compare the methods.
3. You could simulate data under various scenarios to better understand the performance of one or more methods.[-4ex]
4. You may want to replicate an analysis done in a paper or book chapter. If you do this, it cannot be an analysis for which the code to analyze it is already available in the text, on a website or elsewhere.
5. You may have your own idea for your project. Come talk to me.

Your proposal need not be more than 2 typed pages (1 double-sided sheet or 2 single-sided sheets, font no smaller than 11pt). The proposal should be roughly separated into sections as follows: Introduction, Motivation, and Methods.
The introduction and motivation may be combined, but should address the following: introduce your data and the scientific problem you will look at. You may wish to include a series of "Questions of Interest" that your project aims to answer or examine. The motivation should explain why you chose this as your project. Is it because there is a personal interest in the subject matter, or there is a statistical question you are interested in? Or are you motivated by the method(s) you want to implement in your project? (To pass the course is not acceptable as a reason here). The methods section doesn't need to delve into the background of the statistical method per se, but rather explain how you intend to do your project. What data will you use and how many observations? How many variables and what are the variable types (categorical, continuous, binary, etc.)? What software or packages will you need? How will you proceed to address your questions of interest? What will be involved in completing the project? Finally, you should try to bring your discussion back to the big picture to reiterate the importance of the proposed project.

Here are some websites that you may find useful while seeking data:

- `http://archive.ics.uci.edu/ml/`
- `http://www.mvstats.com/Resources/page3_datasets_greatideas.htm`
- `http://help.sentiment140.com/for-students/`
- `http://blog.thedataincubator.com/2014/10/data-sources-for-cool-data-science-projects-part-2/`
- `http://www.models.life.ku.dk/datasets`

Several `R` packages also have datasets that are interesting and appropriate.