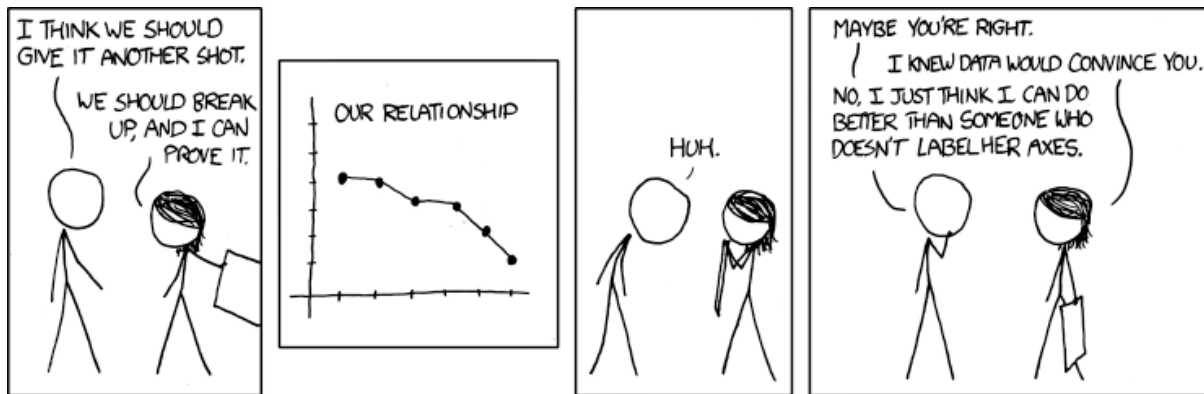# STAT 4010/5010

Statistical Methods and Applications II

Final Project Description

About the Project

You are expected to either (1) conduct a thorough analysis of a data set using statistical modeling techniques, (2) use statistical methodology to gain insights into some theory, or (3) explain and apply some (nontrivial) statistical method that we have not discussed in class (e.g., Bayesian linear regression, kriging). You may work in small groups (2-3 max) if you wish. I hope that you choose a data set or problem based on your interests (e.g., for graduate students, it would be a great idea to choose a data set relevant to your own research interests), rather than just "going through the motions". You will be expected to describe your results in a short **formal paper** and give a short **video presentation** at the end of the semester (due dates to come). The project grade will be determined on the basis of the accuracy of the statistical analysis and the quality of the paper/poster. I expect graduate students to have a more sophisticated understanding of statistical techniques, procedures, etc., and thus, I will hold graduate students to a higher standard. Here's an example of a graduate student project from a past semester: https://bit.ly/2OEg4uj.

Grading Procedure

The project grade will be determined on the basis of the quality of the statistical analysis, paper, and presentation. The project will be graded out of **100 points**. A specific rubric follows:

Ten points will be related to oral communication.

Twenty five points will be related to written communication. You will receive a grade of 0-5 in each of the following categories: (1) context and purpose of writing; (2) content development; (3) genre and disciplinary conventions; (4) sources and evidence; and (5) control of syntax and mechanics.

Sixty five points will be related to problem solving. You will receive a grade between 0-10 on how well you define the problem/question of interest. You will receive a grade between 0-10 on how well you propose solutions to answer the questions. You will receive a grade between 0-25 on implementing solutions to answer the question of interest. This is the proper application of statistical methods to the problem at hand. You will receive a grade between 0-20 on how well you evaluate outcomes, explain conclusions, interpret results, etc.

Sections and Topics for Data Analysis Paper

I suggest that you include the following sections and topics in your paper and presentation. If you are not analyzing a specific data set (e.g., if you are explaining a modeling method), you should be able to adjust these sections accordingly:

1. **Introduction/Background**

   - Why are you interested in this problem?

   - What is the relevant background information for readers to understand your project? Assume that your audience is not an expert in the application field.

   - Is there any prior research on your topic that might be helpful for the audience?

   - From where did the data come? Is this an experiment or observational study? Who collected the data? Why was the data collected (if you weren't the one doing the collecting)?

   - What are the questions of interest that you hope to answer?

2. **Methods/Results (experimental design and data collection)**

   - How did you obtain the data? A research lab? Government website? Web scraping?

   - Describe your exploratory data analysis methods. What needed to be done to the dataset to make it amenable to analysis?

   - What analyses are most appropriate to answer the question of interest?

   - Describe the analyses used. Check your assumptions!

   - Present relevant graphics and interpret results.

   - Explicitly connect your technical (e.g., statistical, mathematical) results to your research questions.

3. **Conclusions**

   - What are your conclusions? What did you learn?

   - How would you extend this research? What future research ideas come to mind based on your results and experience with this analysis?

Proposal

I want you to start thinking of a data set or problem that you would like to work on for your final project relatively early. If you are choosing a data analysis project, here are some of the questions that I want you to consider for your proposal:

1. What are the data that you plan to work with?

2. Where did the data come from? Are they experimental or observational?

3. Why is this data interesting to you? What questions do you hope to answer about it?

4. What are the relationships between the variables? Does theory suggest that they are related in some way?

5. What random components are present (e.g., measurement error)?

6. What prior research on your topic might be helpful to consider?

7. What methods might be useful in analyzing this data?

Please answer these questions (and perhaps others) in a project proposal for your final project. The proposal should be typed, using complete sentences and proper grammar (I expect proposals to be roughly a page in length). Check Canvas for the proposal due date.