So far in this course, emphasis has been placed in gaining technical competence in fitting and interpreting statistical models. In most cases, a clean data set and clear analysis objectives were provided. In practice, a data scientist often self-navigates all steps of the statistical modeling process (discussed in Lesson 1):

1. Understand the problem,
2. Plan and properly collect relevant data,
3. Explore data,
4. Postulate a model,
5. Fit the model,
6. Check the model,
7. Iterate if necessary, and
8. Use the model.

This process includes preparing data, which typically do not come neatly packaged. A data scientist must identify what data are relevant to solving the problem, and prepare them for statistical modeling. In fact, the first three steps in the modeling process are often the most time consuming.

The data analysis project's goal is to help you:

1. Become skilled in all steps of the modeling process,
2. Synthesize the concepts taught in this course,
3. Practice the techniques taught in the course, and
4. Learn to effectively communicate results.

To complete this assignment, you must select a problem that is interesting to you, identify and collect relevant data, proceed through the modeling process, and communicate your findings in a brief report (four pages maximum). To receive credit, you must also serve as a peer-reviewer and grader for two other reports.

Review criteria

This assignment will be graded by your peers. They will not have access to your data or code, and thus cannot verify that the results are technically correct. However, they will be evaluating how well you navigate the statistical modeling process and whether you effectively and efficiently communicate your results. The grading rubric will include questions about whether your data address the question at hand, whether the chosen statistical model is appropriate for the data, whether you have justified your modeling decisions, and whether there is sufficient detail and clarity that the reviewers could reproduce the results if given the data and your report alone.