**Final Project**

**(Preliminary Report due: TUESDAY, APRIL 8!)**

**(Final Report due: TUESDAY, APRIL 22!)**

Your final project will be worth **50% of your final exam grade**. For this project, **you will collect your own data set that is of interest to you**. The data set should have a numerical response variable and at least three to four predictor variables. **(Your dataset cannot be from the textbook.)** You will pose specific research question(s) and use appropriate statistical technique(s) to analyze the data. A report summarizing your results, (including any necessary tables, charts, etc.) and answering the posed research questions will be written to a non-statistical audience. **Your preliminary report will be submitted to the instructor (April 8) who will provide suggestions and edits (by April 15) for your revised and final version of the report. The final report will be due on April 22.**

Your report should include:

* Research questions and the nature of the problem clearly stated.
* Descriptive statistics and charts, and a brief discussion of these charts in context with your research questions.
* Your final regression model including fitted coefficients, R2 value, relevant CIs and/or PIs, etc.
* Appropriate plots and discussion regarding the validity of the assumptions for your model.
* Conclusion – a brief discussion using your regression model to answer the posed research questions.

**Data Set**

Ideally you will find a dataset that is of personal interest to you. If you are having troubles finding a dataset, you may find the link below to be helpful. There is a “Datasets” link at the top of the page.

[Kaggle: Your Machine Learning and Data Science Community (Links to an external site.)](https://www.kaggle.com/).

**Important Dates:**

* March 18 – have your data collected (this is a suggested date).
* April 8 – Submit your preliminary report.
* April 15 – Your preliminary report will be returned to you with suggested edits.
* April 22 – final report is due.