

# STAT 213: Project I (80 pts)

Colin Reimer Dawson

Due via Blackboard by 9am Sunday 3/13

**Goals** The goal of this project is to carry out the process of choosing, fitting, assessing, and using a regression model from beginning to end, documenting the process and writing about the results in a reproducible format (i.e., RMarkdown).

There is a dataset on my website containing information about various U.S. liberal arts colleges as of 2014. You may use any quantitative variable (or function of more than one) in this data as the response; similarly for the predictor. If you prefer, you may use another dataset of interest, in which case you need to provide the raw data and a code book, and give background information. You should consider multiple models and explain how you chose among them (but don't write in full detail about toooo many).

Data URL: [http://colinreimerdawson.com/data/colleges\\_2014.csv](http://colinreimerdawson.com/data/colleges_2014.csv)

Code book URL: [http://colinreimerdawson.com/data/colleges\\_2014-codebook.htm](http://colinreimerdawson.com/data/colleges_2014-codebook.htm)

**What to include** Your writeup should contain the same elements as the Simple Linear Regression group lab. See that lab for a summary of the key R commands, and the recommended workflow process for creating a Markdown document. The main differences between this project and that lab (other than the weight for your grade) are that (1) you can choose what response variable to model, as well as what predictor to use, and (2) you should include a bit more background text on the thought process guiding your choice of models. **Your writeup should *not* simply be a collection of code, graphs and brief text snippets, but should contain “glue” text that guides the reader through your modeling process.**

**Honor Code** Unlike the labs, this project should be done individually. You are free to consult any materials you would like, but should not collaborate with other students.

**Grading Rubric** The grade for this project is determined by the following components:

Component	Weight
Completeness (all components listed in Lab 3 are included and explained)	55 pts
Clarity/style	25 pts

**Deadline and Submission Method** Turn in your writeup via Blackboard as you would a lab assignment by **Sunday, 3/13 at 9am**. I will try to grade all of them on Sunday during the day so that you can get them back before the exam. Include a PDF of the compiled report, as well as the source .Rmd. If you used a dataset other than the one provided, include the data and code book as well.