

# STAT 302 Final Project

Sheridan Grant

Report due on Canvas **Wednesday, June 10th at 11:59am—zero exceptions** (I often let late homework slide, but not this). Verbal assessments will be during the final exam period **2:30-4:20 Thursday, June 11th**, with exceptions made for students in other time zones. Scheduling TBD.

## Submission Instructions

You will write your report via R Markdown, so you will submit a zipped folder with the .RMD file, knitted HTML, and any other files needed to compile your HTML **up to 100 MB total**. If your data is huge, provide instructions for accessing it online in the .RMD file.

The writing and code should flow naturally with one another. You can “hide” code chunks that don’t demonstrate anything useful.

## Grading Breakdown

1. 20% Verbal Assessment
  - (a) 5-6 min if alone, 7-9 min with partner
  - (b) Be able to defend what you did in your report. I’m not going to ask you to produce any new math or analysis. I may ask you general questions about how you might change or extend something.
2. 30% Modeling and Hypothesis Testing
  - (a) Used 1 type of regression model from class (linear model, linear with transformed covariates or outcome, logistic regression) and 1 new type (Random Forest, Neural Network, Poisson GLM, ANOVA, time series AR, etc.).
  - (b) Stated model/test assumptions and addressed their plausibility.
  - (c) Interpreted models correctly, assessed predictive performance.

- (d) Scientific hypotheses are substantive, practical, and well-formulated. “Are any of these variables correlated with the outcome” earns less points than “is there an interaction effect between these two covariates in explaining the outcome.”
3. 15% Visualization and exploratory analysis
- (a) Plots and tables are clearly labeled and formatted well.
  - (b) Plots and tables convey the crucial information about relationships in the data to the reader.
4. 20% Writing
- (a) English writing ability will not be graded beyond the standard of conveying information unambiguously.
  - (b) 500-700 words if alone, 700-1000 if working with a partner. “Words” applies to all text written outside of .RMD code chunks. Within the bounds of word limits, better writing is better than more words. 525 words clearly written is better than 650 words written less well.
  - (c) Writing should address the scientific issues and modeling primarily, and the practical importance of your work briefly.
5. 15% Code
- (a) Correctness.
  - (b) Readability (as I grade your work, I should be able to fairly quickly read through your code and see what you’re doing, even if some of the details are harder to understand).
  - (c) Formatting and commenting. If you write your own function, or do something complicated, write a code comment to briefly explain.