Presentation Outline

(One slide for each second level bullet point)

- Introduction
    - Research Question
    - Data
        - Original dataset from Kaggle, webscraped from US News
        - There was a lot of missing and clearly incorrect data - this was corrected by filling in or replacing the incorrect data with CollegeSimply's reported numbers.
            - An example of incorrect data to talk about was that the original webscraping had issues with schools in the same system, like UC or other state schools. It often returned the acceptance rate or average SAT score of a different school in that state's system.
    - Single Variable Relationships
        - Show single variable plots, correlation matrix, pairwise plot
            - sat_avg and hs_gpa_avg roughly normal, enrollment isn't (explain why)
            - sat_avg doesn't exactly have constant variance (explain why)

- Modeling
    - Untransformed full model and interpretation
        - R^2 and p-values show that there is clearly a relationship between our predictor variables and response variable.
        - Point out that enrollment doesn't seem to make much of a difference
    - Transformed full model
        - Transform hs_gpa_avg to (hs_gpa_avg)^5
        - Don't transform sat_avg to log(sat_avg) as it actually makes the model worse (lower R^2, new outliers on diagnostic plots)
        - Ignore enrollment transformation since we're removing it anyways
    - Variable selection
        - Forward step and backward elimination with AIC and BIC all return the same model - include everything except enrollment
    - Final Model and interpretation
        - See .rmd file for model assessment and interpretation notes