Stats II Project: Pretty Privilege

Team members: Amanda, Sara, Viv

1. **The source and inspiration for selecting this particular data set.**
   https://www.openintro.org/data/index.php?data=evals
   Open intro is an independently operated nonprofit publisher that hosts a variety of educational material on their website. We choose this dataset because it meets all the project requirements, is clean, and is also funny.

2. **The size of the data (# of observations and # of variables).**
   464 observations, where each row is an instance of a class, for 94 individual professors from the University of Texas at Austin. There are 23 variables

3. **Description of all variables**

   course_id
   Variable identifying the course (out of 463 courses).

   prof_id
   Variable identifying the professor who taught the course (out of 94 professors).

   score
   Average professor evaluation score: (1) very unsatisfactory - (5) excellent. Continuous quantitative. Can be transformed into a categorical variable (unsatisfactory/satisfactory) is needed.

   rank
   Rank of professor: teaching, tenure track, tenured. Categorical ordinal.

   ethnicity
   Ethnicity of professor: not minority, minority. Categorical nominal.

   gender
   Gender of professor: female, male. Categorical nominal.

   language
   Language of school where professor received education: English or non-English. Categorical nominal (binary).

   age
   Age of professor. Quantitative discrete (measured by year).

   cls_perc_eval
   Percent of students in class who completed evaluation. Quantitative continuous.

   cls_did_eval
   Number of students in class who completed evaluation. Quantitative discrete.

   cls_students
   Total number of students in class. Quantitative discrete.

   cls_level
   Class level: lower, upper. Categorical.

   cls_profs
   Number of professors teaching sections in course in sample: single, multiple. Categorical.

   cls_credits
   Number of credits of class: one credit (lab, PE, etc.), multi credit. Categorical

   bty_f1lower
   Beauty rating of professor from lower level female: (1) lowest - (10) highest. Quant discrete.

   bty_f1upper
   Beauty rating of professor from upper level female: (1) lowest - (10) highest. Quant discrete.

   bty_f2upper
   Beauty rating of professor from second level female: (1) lowest - (10) highest. Quant discrete.

   bty_m1lower
   Beauty rating of professor from lower level male: (1) lowest - (10) highest. Quant discrete.

   bty_m1upper
   Beauty rating of professor from upper level male: (1) lowest - (10) highest. Quant discrete.

   bty_m2upper
   Beauty rating of professor from second upper level male: (1) lowest - (10) highest. Quant discrete.

   bty_avg
   Average beauty rating of professor. Quantitative continuous.

   pic_outfit
   Outfit of professor in picture: not formal, formal. Categorical.

   pic_color
   Color of professor's picture: color, black & white. Categorical.

DATASET MINIMUM REQUIREMENTS: Make sure that your data set:

- ✓ Contains a total of at least 10 variables and 50 observations. ✓
- ✓ Is NOT a time series. ✓
- ✓ Has several numerical variables, one of which is expected to act as a response variable in linear regression (your main variable of interest). ✓
- ✓ Contains at least two categorical variables, one of which can potentially act as a response, and another one as a predictor. Variables such as unique ID's or unique names don't count as categorical variables. ✓
- ✓ **NOTE**: Keep in mind that (eventually) you will be expected to fit **at least one linear regression model**, and **at least one logistic regression model**. ✓