Regression Project
Due June 4

Description of Dataset

In the society of gorillas, a favorite sport is to play tug-of-war against humans. Individual gorillas try to beat as many humans at once as they can. Your goal is to build a regression model to determine how many humans a particular gorilla can beat based on the historical scores of other gorillas. You are given the following information:

- HMNS: Our target variable — the number of humans that the gorilla managed to beat in tug-of-war.
- WHT: The gorilla's weight.
- AGE: The gorilla's age.
- FRM: The circumference of the gorillia's forearm
- DSI: Days since the gorilla last had a significant health issue.
- SUS: The gorilla's sub-species, either "Western Lowland", "Mountain", "Grauer's", or "Cross River".
- GND: The gorilla's gender.

1. **Single Variable Regression:**
   Use single-variable regression to predict HMNS from each of:

   (a) WHT
   (b) AGE
   (c) DSI
   (d) SUS

   Comment on how well each model fits the training and test data. What is the training R-squared? Draw scatter plots with a line to show the fit. [20]

2. **Multiple Variable Regression:**
   Use the techniques of multiple regression to build models using combinations of the variables above. Which models work best? Try different combinations. Explain which ones performed best and how you selected them. Show plots of errors. [20]

3. **Model Comparison:**
   For each model you fit, discuss whether it improves on the previous one and justify why. [10]

4. **Interpretability:**
   Discuss whether the models you chose make sense in the context of gorilla tug-of-war. [15]

5. **Model Evaluation:**
   Use appropriate tools to evaluate your models (scatter plots, AIC/BIC, cross-validation, etc). [15]

6. **Feature Engineering:**
   Try to engineer new features that result in better-performing models. Explain why you chose them. [10]