Stat 6021: Project 1

due Mar 23

A client has brought you a dataset, "mileage.txt". The data come from 32 automobiles. The variables are

- y: gas mileage (miles/gallon)
- x_1: Displacement (cubic in.)
- x_2: Horsepower (ft-lb)
- x_3: Torque (ft-lb)
- x_4: Compression ratio
- x_5: Rear axle ratio
- x_6: Carburetor (barrels)
- x_7: No. of transmission speeds
- x_8: Overall length (in.)
- x_9: Width (in.)
- x_{10}: Weight (lb)
- x_{11}: Type of transmission (automatic/manual)

Your client has this main task for you: he wants to find a suitable model that will help him relate the gas mileage of a vehicle to these predictors. The client's main concerns are:

1. the model fits well and
2. simplicity of the model.

Your group is to type up a report for this project. One member of your group is to upload the report and the R script via Assignments on Collab. The report is to include the following sections:

1. **Initial model considered**: In this section, explain how your group decided to start the model building process. What was the first model considered? How was this model evaluated based on the client's main concerns? Be sure to include the relevant R output and graphical summaries.

2. **Other models considered**: Based on the initial model from the previous section, consider at least one other model. Be sure to explain why you considered these models. Like the previous section, be sure to include the relevant R output and graphical summaries. Compare and contrast all of the models you considered. Be sure to explain how you evaluated each model and compared them with one another, bearing in mind the client's main concerns.

3. **Summary of findings**: Write the model(s) that you will recommend. Be sure to explain your recommendations.

4. **Summary of findings for the client**: In this section, your group is to write your recommended model(s) to the client. Be sure to address the client's main concerns. Be sure to explain to the client why you are recommeding a model(s). This section should be written in a way that can be understood by anyone without any statistical background (avoid terms such as p-value, null hypothesis, etc).

Note: Sections 1 to 3 should be written in a way that another classmate in this class can understand what you did. Imagine if the client decided to hire one of your classmates to review your report, so these sections need enough technical detail and clarity. You should not need to perform any calculations by hand; use R for all calculations if needed. The last section is written for the client who has no statistical expertise so he/she will only read this section.

The entire report should be no more than 10 pages double spaced, and needs to include the relevant output from R. As a general rule of thumb, output is relevant when it is referred to in the report. R code does not need to be included in the report; the R code should be supplied as a separate R script (.R file, no .rmd files). **The report and R script is due on Monday, March 23**.

Please also complete the **Project 1 Evaluation Questions**, which is also **due on Monday, March 23**. Please note that this is specifically for the project, and is separate from the module 7 guided question set evaluation that you completed a week prior.