HOMEWORK MULTIPLE LINEAR REGRESSION

**Gas Mileage**  A researcher is interested in developing a model that describes the gas mileage, measured in miles per gallon (mpg), of automobiles. Based on input from an engineer, she decides that the explanatory variables might be engine size (liters), curb weight (pounds) and horsepower. From a random sample of 13 automobiles, she obtains the following data.

(On the last page of this homework assignment is the direct data table input that you can copy and paste directly a new r script page)

**Engine Size** **Curb Weight** **Horsepower Miles per Gallon,**

2.4 3289 177 24

2.4 3263 158 25

2.5 3230 170 24

3.5 3580 272 22

2.8 3175 255 18

3.5 3643 263 22

3.5 3497 306 20

3.0 3340 230 21

3.6 3861 263 19

2.4 3287 173 24

3.3 3629 234 21

2.5 3270 170 22

3.5 3292 270 22

1) Use and show R code that will produce a correlation matrix.

2) Indicate if the correlation matrix shows multicollinearity problems with pairs of

explanatory variables. The variable Miles per Gallon is the response variable

3) Indicate the explanatory variable(s) that the response variable is strongly correlated with.

4) Use and show R code that will produce a full multiple regression model, that uses all

explanatory variables. Write the full model that shows the coefficients for all explanatory

variables and the intercept.

5) Use and show R code that will produce all important summary statistics for your model.

6) Which explanatory variable(s) is/are significant at the level of .05?

7) Use your book or an internet source and describe the difference between R2 and adjusted R2. (Three or four sentences)

8) What is the proportion of the variability in Milespergallon that is explained by the full model regression equation?

9) What is the proportion of the variability in Milespergallon that is explained by the full model regression equation, only using explanatory variables that have significant impact on the dependent variable.

10) Use and show R code that produces a model without the explanatory variable with the highest p value. Write the new model that indicates the response variable, the explanatory coefficients and the intercept.

11) Explain why the null hypothesis that the population coefficient for Engine Size is equal to zero is not to be rejected.

12) Which model, the one with three explanatory variables or the one with two explanatory variables is the better model for predicting Miles per gallon? Justify your answer in three of four sentences.

**R input data table** (The table may not be aligned perfectly after you copy and paste into r, but it does not have to. You can carry on with your processing as it appears)

library(tidyverse)

tribble(~EngineSize, ~CurbWeight, ~Horsepower, ~MilesperGallon,

2.4, 3289, 177, 24,

2.4, 3263, 158, 25,

2.5, 3230, 170, 24,

3.5, 3580, 272, 22,

2.8, 3175, 255, 18,

3.5, 3643, 263, 22,

3.5, 3497, 306, 20,

3.0, 3340, 230, 21,

3.6, 3861, 263, 19,

2.4, 3287, 173, 24,

3.3, 3629, 234, 21,

2.5, 3270, 170, 22,

3.5, 3292, 270, 22

) -> V

V