Statistical and Predictive Modeling I (DATA 1204)
Assignment #1 – Distribution Analysis (10% of Total Grade)
Professor: Ritwick Dutta

For the following questions, we will be using the **carsDB dataset from Week #2** for this exercise. The dataset contains 32 observations on 11 (numeric) variables.

[, 1]	mpg	Miles/(US) gallon
[, 2]	cyl	Number of cylinders
[, 3]	disp	Displacement (cu.in.)
[, 4]	hp	Gross horsepower
[, 5]	drat	Rear axle ratio
[, 6]	wt	Weight (1000 lbs)
[, 7]	qsec	1/4 mile time
[, 8]	vs	Engine (0 = V-shaped, 1 = straight)
[, 9]	am	Transmission (0 = automatic, 1 = manual)
[,10]	gear	Number of forward gears
[,11]	carb	Number of carburetors

Questions (*please use ggplot2 to construct all of your graphs. Post all graphs and answers in a word document in doc or docx format*)**:**

1. Construct a scatterplot with smoothing for mpg vs. disp
2. Construct a scatterplot with smoothing for mpg vs. hp
3. Construct a scatterplot with smoothing for mpg vs. wt
4. Please explain the following:
   a) Identifiable Trends for each of the three (3) scatterplots
   b) Any anomalies that you see for each of the three (3) scatterplots

Hint: Leverage the Week #2 – Advanced Graphing Tutorial

Please post your <u>**Word Document (.doc or .docx) including your R script**</u> **via assignments under Assignment #1 by 11:59 pm on Wednesday, Jan. 27th, 2021.**

Remember that DC Connect doesn't accept R script files. All you need to do is to copy the code from your top left panel in R and paste it at the bottom of your word document.

RUBRIC

Question	Exemplary (5)	Proficient (3-4)	Incomplete (1-2)	Incorrect or Unacceptable (0)
1	A labelled complete scatterplot of the variables mentioned is shown.	Scatterplot is incomplete or missing some observations or labels.	Scatterplot does not include both the appropriate variables but includes one appropriate variable.	Scatterplot is absent or does not include both the appropriate variables.
2	A labelled complete scatterplot of the variables mentioned is shown.	Scatterplot is incomplete or missing some observations or labels.	Scatterplot does not include both the appropriate variables but includes one appropriate variable.	Scatterplot is absent or does not include both the appropriate variables.
3	A labelled complete scatterplot of the variables mentioned is shown.	Scatterplot is incomplete or missing some observations or labels.	Scatterplot does not include both the appropriate variables but includes one appropriate variable.	Scatterplot is absent or does not include both the appropriate variables.
4	A proper explanation of the distribution and anomalies.	The explanation of the distribution and anomalies is incomplete.	An incorrect explanation of distribution and anomalies.	No explanation is given.