

# Durham College DATA1204 - Statistical Predictive Modeling

## Assignment 1

Chinedu Onyeka

September 27th, 2021

For the following questions, we will be using the carsDB dataset from Week #2 for this exercise

**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

Load Libraries

```
library(ggplot2)
```

Load the data

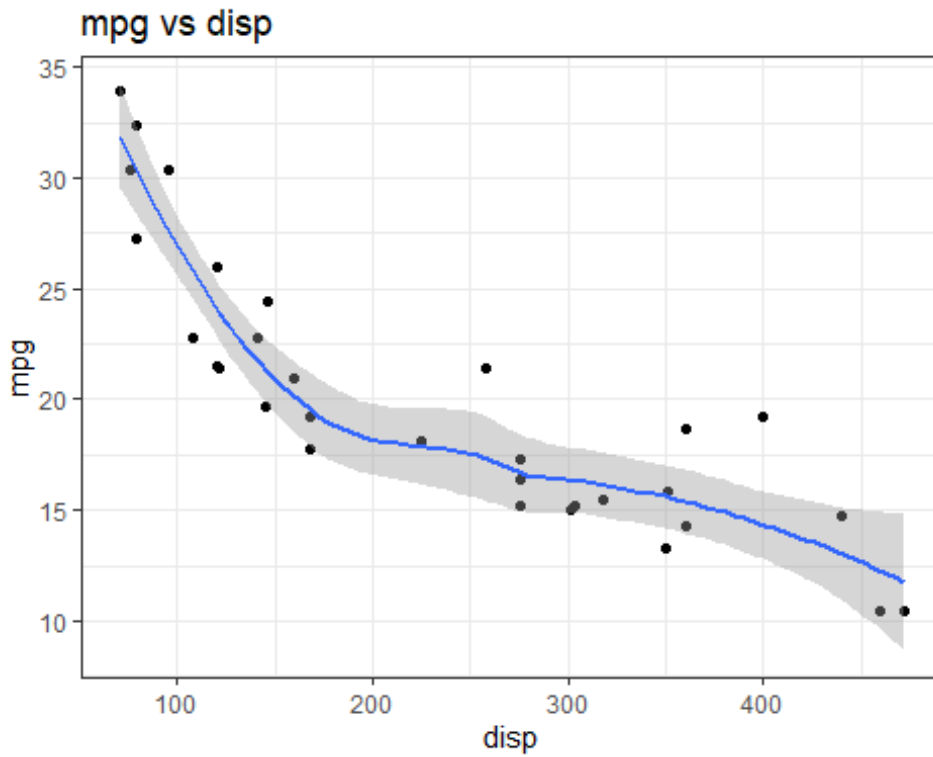
```
data_mtcars <- mtcars
```

```
head(data_mtcars, n = 10)
```

##	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
## Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
## Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
## Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
## Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
## Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
## Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
## Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
## Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
## Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4

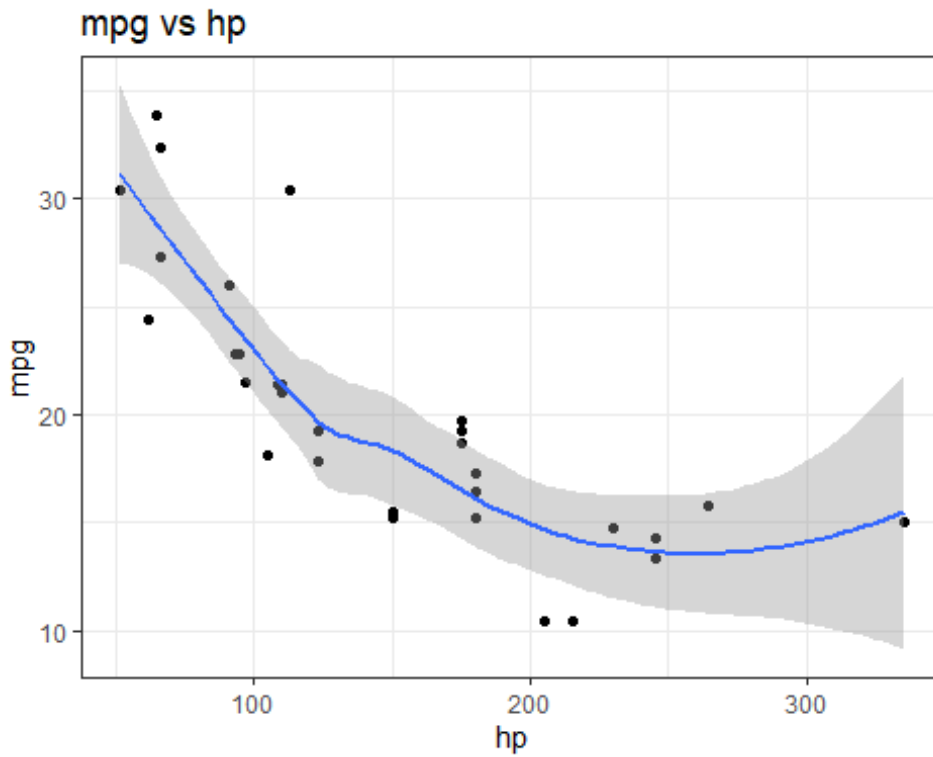
```
g1 <- ggplot(data = data_mtcars, aes(x = disp, y = mpg)) + geom_point() +
  geom_smooth() + labs(title = "mpg vs disp") + theme_bw()
g1
```

```
g1 <- ggplot(data = data_mtcars, aes(x = disp, y = mpg)) + geom_point() +  
  geom_smooth() + labs(title = "mpg vs disp") + theme_bw()  
g1
```



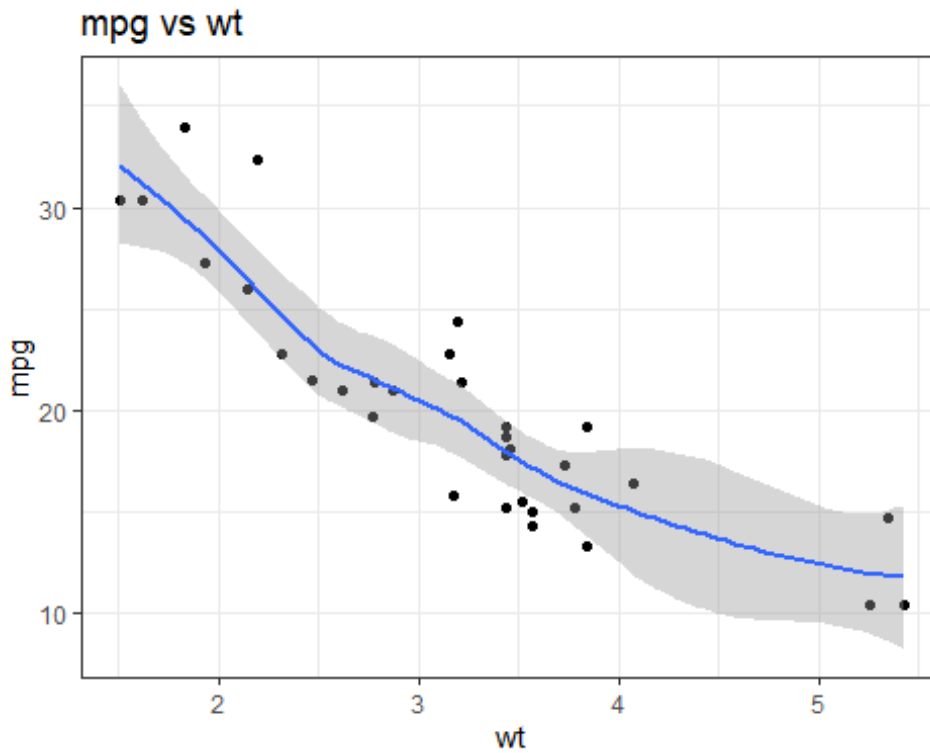
**Question 2 : Construct a scatterplot with smoothing for mpg vs. hp**

```
g2 <- ggplot(data = data_mtcars, aes(x = hp, y = mpg)) + geom_point() +  
  geom_smooth() + labs(title = "mpg vs hp") + theme_bw()  
g2
```



**Question 3 : Construct a scatter plot with smoothing for mpg vs. wt**

```
g3 <- ggplot(data = data_mtcars, aes(x = wt, y = mpg)) + geom_point() +  
  geom_smooth() + labs(title = "mpg vs wt") + theme_bw()  
g3
```



**Question 4 : Please explain the following:**

- a) Identifiable Trends for each of the three (3) scatter plots

Answer: From the plot above, it can be seen that there is a downward trend (negative correlation for each of the three scatterplots). This is inline with what one would think for example in plot 2 and 3 above.

For the third plot, the mpg (miles per gallon) for a vehicle would be negatively correlated with the weight of the vehicle which means that heavier vehicles would tend to perform poorly on gas and therefore have low mpg and vice - versa.

Also, for the second plot, vehicles with high hp(Gross horsepower) then to do poorly with gas as well and have low mpg. They produce more power and consume more gas or fuel, hence low on mpg.

- b) Any anomalies that you see for each of the three (3) scatter plots

Answer: We can see from the graph that all points do not fall perfectly on the trend line. Some points even fall outside of the margin of error which goes to show that the inverse relationship is not perfect and some vehicles will tend to deviate from this relationship.