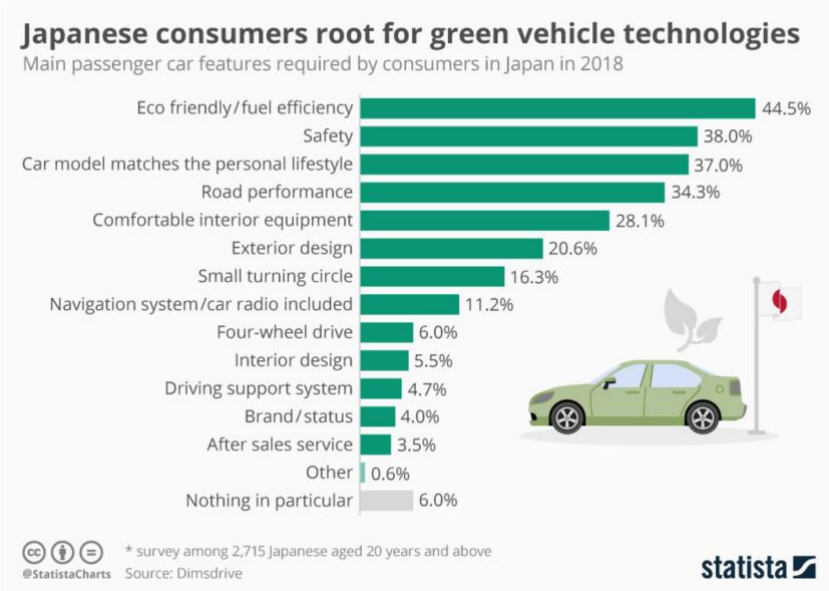
Course: STAT 488D (Statistics)
Semester: Summer 2022
Name: Shanjie,Xu

Introduction:

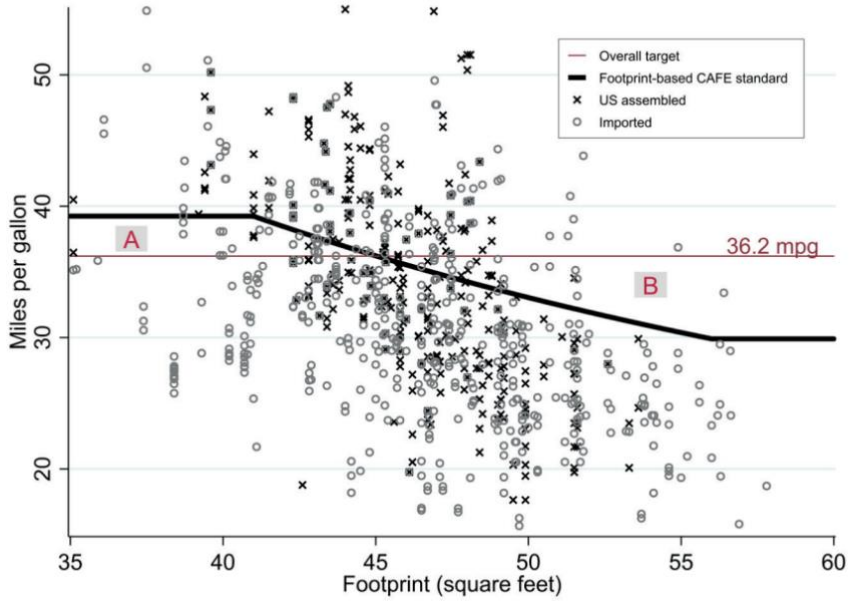
Most people believe there is a linear relationship between fuel consumption in Miles Per Gallon (MPG) and Engine Sizes. However, many other factors and reasonable variables would also change fuel consumptions. As mentioned in the article, "*Miles Per Gallon(MPG) is a unit which we use to evaluate the efficiency of a transporting vehicle in terms of the energy produced".*

"The 'generic' Japanese car ranked highest on the other two factors (fuel economy and reliability)."(Brown,1987). We could also see the graph below fomr statista shows Fuel efficiency is the top of the list.Since Japanese cars have mass production and are well known as the most fuel-efficient car country, we will not take them as experimental variables here.



For Ameircan car, we won't consider them in this case as America has cheaper gas prices which is even cheaper than Canada. From the model below, we may see how unstable MPG is in USA.

Figure 1. Car models by fuel economy and footprint; model year 2015.



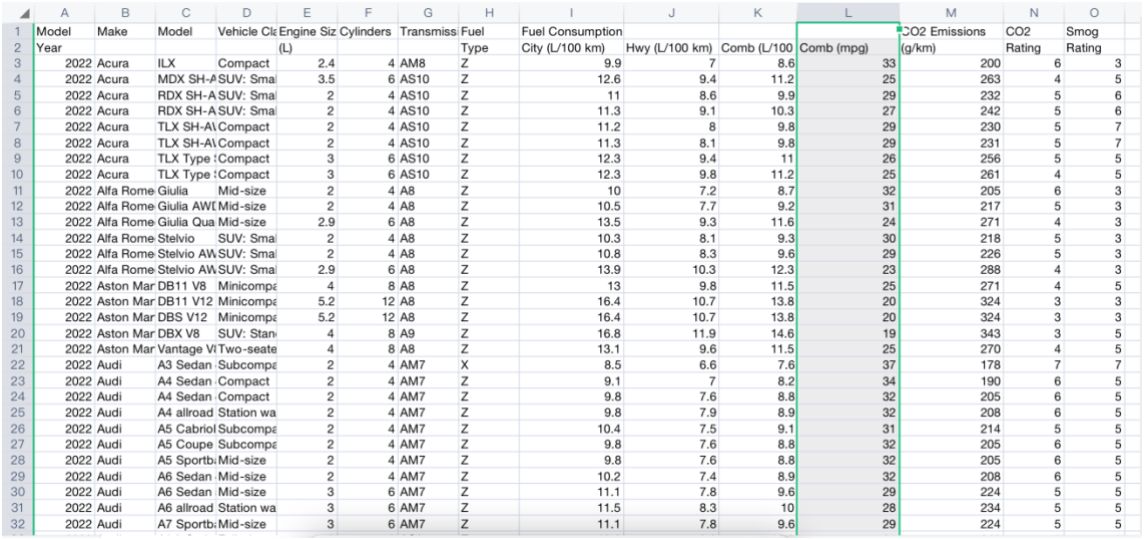
German cars have been designed and used since World War Two. The Global Fuel Economy website indicates:*"Germany is the fourth car manufacturer globally (after China, the United States and Japan) and the main one in Europe. The average fuel economy of newly registered LDVs reached 6.0 Lge/100 km in 2015, which is ahead of the global average of 7.6 Lge/100 km."*

Another reason for German cars is they have huge taxes on pollution. From the model below, the more Carbon Dioxide we emit, the more tax we pay. This could limit the extreme powerful performance cars and get more realistic data from the daily family using purpose.

Brands are also important. *“The rivalry is quite intense, especially among the major players that include Audi, BMW, Mercedes-Benz (and Toyota’s Lexus Motors in the US). These companies not only compete vigorously for market leadership in the major luxury car markets across the world, but also for the recognition of being the leading luxury brand of vehicle from the perception of the consumers.” (Haleder,2018).* This indicates Audi, BMW and Benz are reliable brands and they are not only just focusing on fuel-saving technology but developing in a balanced way.

Original Data:

https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64/resource/87fc1b5e-fafc-4d44-ac52-66656fc2a245



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Model	Make	Model	Vehicle Cla	Engine Siz	Cylinders	Transmiss	Fuel	Fuel Consumption				CO2 Emissions	CO2	Smog
2	Year				(L)			Type	City (L/100 km)	Hwy (L/100 km)	Comb (L/100	Comb (mpg)	(g/km)	Rating	Rating
3	2022	Acura	ILX	Compact	2.4	4	AM8	Z	9.9	7	8.6	33	200	6	3
4	2022	Acura	MDX SH-A	SUV: Smal	3.5	6	AS10	Z	12.6	9.4	11.2	25	263	4	5
5	2022	Acura	RDX SH-A	SUV: Smal	2	4	AS10	Z	11	8.6	9.9	29	232	5	6
6	2022	Acura	RDX SH-A	SUV: Smal	2	4	AS10	Z	11.3	9.1	10.3	27	242	5	6
7	2022	Acura	TLX SH-A\	Compact	2	4	AS10	Z	11.2	8	9.8	29	230	5	7
8	2022	Acura	TLX SH-A\	Compact	2	4	AS10	Z	11.3	8.1	9.8	29	231	5	7
9	2022	Acura	TLX Type !	Compact	3	6	AS10	Z	12.3	9.4	11	26	256	5	5
10	2022	Acura	TLX Type !	Compact	3	6	AS10	Z	12.3	9.8	11.2	25	261	4	5
11	2022	Alfa Rome	Giulia	Mid-size	2	4	A8	Z	10	7.2	8.7	32	205	6	3
12	2022	Alfa Rome	Giulia AWI	Mid-size	2	4	A8	Z	10.5	7.7	9.2	31	217	5	3
13	2022	Alfa Rome	Giulia Qua	Mid-size	2.9	6	A8	Z	13.5	9.3	11.6	24	271	4	3
14	2022	Alfa Rome	Stelvio	SUV: Smal	2	4	A8	Z	10.3	8.1	9.3	30	218	5	3
15	2022	Alfa Rome	Stelvio AW	SUV: Smal	2	4	A8	Z	10.8	8.3	9.6	29	226	5	3
16	2022	Alfa Rome	Stelvio AW	SUV: Smal	2.9	6	A8	Z	13.9	10.3	12.3	23	288	4	3
17	2022	Aston Mar	DB11 V8	Minicompa	4	8	A8	Z	13	9.8	11.5	25	271	4	5
18	2022	Aston Mar	DB11 V12	Minicompa	5.2	12	A8	Z	16.4	10.7	13.8	20	324	3	3
19	2022	Aston Mar	DBS V12	Minicompa	5.2	12	A8	Z	16.4	10.7	13.8	20	324	3	3
20	2022	Aston Mar	DBX V8	SUV: Stan	4	8	A9	Z	16.8	11.9	14.6	19	343	3	5
21	2022	Aston Mar	Vantage V	Two-seate	4	8	A8	Z	13.1	9.6	11.5	25	270	4	5
22	2022	Audi	A3 Sedan	Subcompa	2	4	AM7	X	8.5	6.6	7.6	37	178	7	7
23	2022	Audi	A4 Sedan	Compact	2	4	AM7	Z	9.1	7	8.2	34	190	6	5
24	2022	Audi	A4 Sedan	Compact	2	4	AM7	Z	9.8	7.6	8.8	32	205	6	5
25	2022	Audi	A4 allroad	Station wa	2	4	AM7	Z	9.8	7.9	8.9	32	208	6	5
26	2022	Audi	A5 Cabriol	Subcompa	2	4	AM7	Z	10.4	7.5	9.1	31	214	5	5
27	2022	Audi	A5 Coupe	Subcompa	2	4	AM7	Z	9.8	7.6	8.8	32	205	6	5
28	2022	Audi	A5 Sportb	Mid-size	2	4	AM7	Z	9.8	7.6	8.8	32	205	6	5
29	2022	Audi	A6 Sedan	Mid-size	2	4	AM7	Z	10.2	7.4	8.9	32	208	6	5
30	2022	Audi	A6 Sedan	Mid-size	3	6	AM7	Z	11.1	7.8	9.6	29	224	5	5
31	2022	Audi	A6 allroad	Station wa	3	6	AM7	Z	11.5	8.3	10	28	234	5	5
32	2022	Audi	A7 Sportb	Mid-size	3	6	AM7	Z	11.1	7.8	9.6	29	224	5	5

Original data contains 12 vairables and 6671 data together. I decide to filter out useful data shown below:

Make	Model	Vehicle.Class	Engine.Size	Fuel.Consumption	MPG	CO2.Emissions	CO2
Audi	A3 Sedan 40 TFSI quattro	Subcompact	2.0	8.5	37	178	7
Audi	A4 Sedan 40 TFSI quattro	Compact	2.0	9.1	34	190	6
Audi	A4 Sedan 45 TFSI quattro	Compact	2.0	9.8	32	205	6
Audi	A4 allroad 45 TFSI quattro	Station wagon: Small	2.0	9.8	32	208	6
Audi	A5 Cabriolet 45 TFSI quattro	Subcompact	2.0	10.4	31	214	5
Audi	A5 Coupe 45 TFSI quattro	Subcompact	2.0	9.8	32	205	6
Audi	A5 Sportback 45 TFSI quattro	Mid-size	2.0	9.8	32	205	6
Audi	A6 Sedan 45 TFSI quattro	Mid-size	2.0	10.2	32	208	6
Audi	A6 Sedan 55 TFSI quattro	Mid-size	3.0	11.1	29	224	5
Audi	A6 allroad 55 TFSI quattro	Station wagon: Mid-size	3.0	11.5	28	234	5
Audi	A7 Sportback 55 TFSI quattro	Mid-size	3.0	11.1	29	224	5
Audi	A8 L Sedan 55 TFSI quattro	Full-size	3.0	12.6	27	248	5
Audi	Q3 40 TFSI quattro	SUV: Small	2.0	10.4	31	215	5
Audi	Q3 45 TFSI quattro	SUV: Small	2.0	11.4	28	233	5
Audi	Q5 40 TFSI quattro	SUV: Small	2.0	10.3	30	217	5
Audi	Q5 45 TFSI quattro	SUV: Small	2.0	10.3	30	220	5
Audi	Q5 Sportback 45 TFSI quattro	SUV: Small	2.0	10.3	30	220	5
Audi	Q7 45 TFSI quattro	SUV: Standard	2.0	12.0	26	252	5
Audi	Q7 55 TFSI quattro	SUV: Standard	3.0	12.8	24	273	4
Audi	Q8 55 TFSI quattro	SUV: Standard	3.0	12.8	24	273	4
Audi	R8 Coupe Performance	Two-seater	5.2	16.7	20	322	3
Audi	R8 Coupe Performance quattro	Two-seater	5.2	17.9	18	356	3
Audi	R8 Spyder Performance	Two-seater	5.2	16.7	20	322	3
Audi	R8 Spyder Performance quattro	Two-seater	5.2	17.9	18	356	3
Audi	RS 5 Coupe quattro	Subcompact	2.9	13.0	25	267	4
Audi	RS 5 Sportback quattro	Mid-size	2.9	13.1	25	268	4

Showing 1 to 26 of 147 entries, 8 total columns

Variables Originally we have:

- Year
- Make (The brand of the car)
- Model (The model name of a car)
- Vehicle Class (body styles)
- Engine Size (from 1.2 to 8 Litres)
- Cylinders
- Transmission
- Fuel type
- Fuel Consumption (City, Highway, KM/H and MPG Combined)
- CO2 Emissions (the tailpipe emissions of carbon dioxide in grams per kilometre for combined city and highway driving
- CO2 (the tailpipe emissions of carbon dioxide rated on a scale from 1 (worst) to 10 (best))
- Smog (the tailpipe emissions of smog-forming pollutants rated on a scale from 1 (worst) to 10 (best))

The original data has some variables we won't use in this project. So here are the *useful* variables:

- Make (The brand of the car)
- Model (The model name of a car)
- Vehicle Class (body styles)
- Engine Size (from 1.2 to 8 Litres)
- Fuel Consumption (MPG Combined)
- CO2 Emissions (the tailpipe emissions of carbon dioxide in grams per kilometre for combined city and highway driving)
- CO2

Notice that CO2 Rating would not be useful for this anylsis. The key variable is the CO2 Emissions.

Data Summerization:

```
          Make                       Model          Vehicle.Class Engine.Size Fuel.Consumption      MPG
Audi         :44   330i xDrive Sedan    :  1   Subcompact   :34    2.0:50      Min.   : 8.50    Min.   :18.0
BMW          :60   430i xDrive Cabriolet:  1   SUV: Small   :24    2.9: 4      1st Qu.:10.80    1st Qu.:22.0
Mercedes-Benz:43   430i xDrive Coupe    :  1   SUV: Standard:23    3.0:55      Median :12.00    Median :27.0
                   530i xDrive Sedan    :  1   Mid-size     :21    4.0: 8      Mean   :12.68    Mean   :26.1
                   540i xDrive Sedan    :  1   Compact      :16    4.4:25      3rd Qu.:14.50    3rd Qu.:30.0
                   750i xDrive Sedan    :  1   Full-size    :10    5.2: 4      Max.   :18.00    Max.   :37.0
                   (Other)              :141   (Other)      :19    6.6: 1
CO2.Emissions   CO2
Min.   :178.0   2: 5
1st Qu.:221.0   3:29
Median :248.0   4:32
Mean   :259.4   5:63
3rd Qu.:293.0   6:17
Max.   :364.0   7: 1
```

MPG Calculation:

```
> mean(newData$MPG)
[1] 26.09524
> median(newData$MPG)
[1] 27
> sd(newData$MPG)
[1] 4.52744
> ##Variance
> var(newData$MPG)
[1] 20.49772
> ## IQR
> quantile(newData$MPG)
  0%  25%  50%  75% 100%
  18   22   27   30   37
```

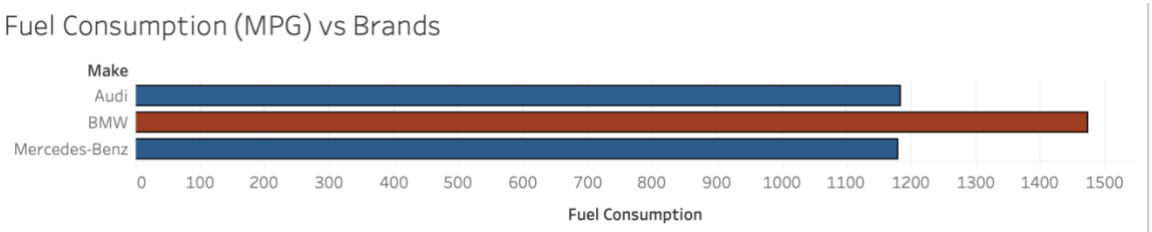
Possible Problem(s):

1. What is data Looks like?

```
> str(newData)
'data.frame':   147 obs. of  8 variables:
 $ Make            : Factor w/ 3 levels "Audi","BMW","Mercedes-Benz": 1 1 1 1 1 1 1 1 1 1 ...
 $ Model           : Factor w/ 147 levels "330i xDrive Sedan",..: 10 12 13 11 14 15 16 18 19 17 ...
 $ Vehicle.Class   : Factor w/ 10 levels "Compact","Full-size",..: 7 1 1 6 7 7 3 3 3 5 ...
 $ Engine.Size     : Factor w/ 7 levels "2.0","2.9","3.0",..: 1 1 1 1 1 1 1 1 3 3 ...
 $ Fuel.Consumption: num  8.5 9.1 9.8 9.8 10.4 9.8 9.8 10.2 11.1 11.5 ...
 $ MPG             : int  37 34 32 32 31 32 32 32 29 28 ...
 $ CO2.Emissions   : int  178 190 205 208 214 205 205 208 224 234 ...
 $ CO2             : Factor w/ 6 levels "2","3","4","5",..: 6 5 5 5 4 5 5 5 4 4 ...
```

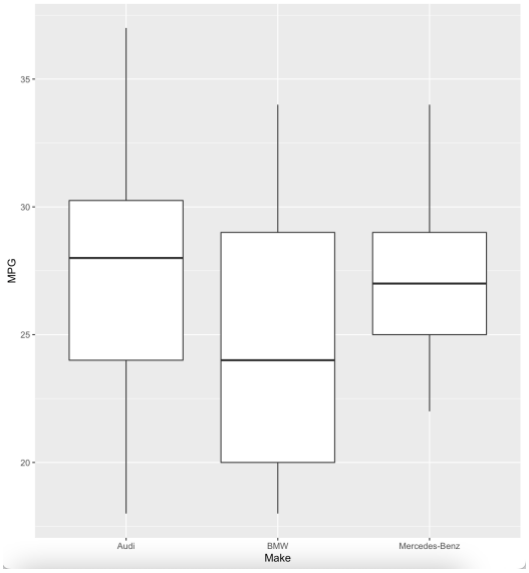
Notice that Only Fuel.Consumption, MPG and CO2.Emissions are numerical data. Others are categorical data. Also be ware that we will be using 'MPG' variable not 'Fuel. Consumption' variable.

2. Which brand has more fuel consumption?

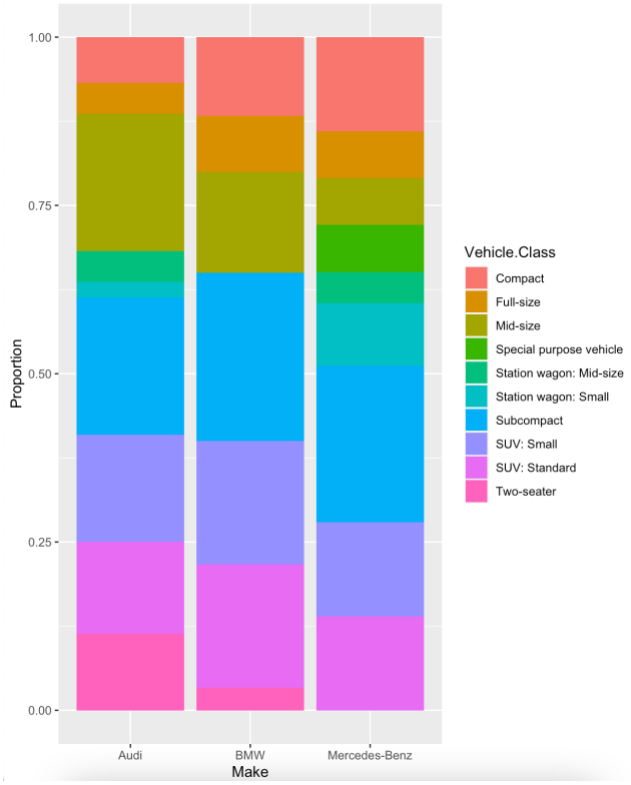


This beautiful graph generated from Tableau which shows clearly the BMW has more fuel consumption than Audi and Benz.

3. How MPG varies across Brands?

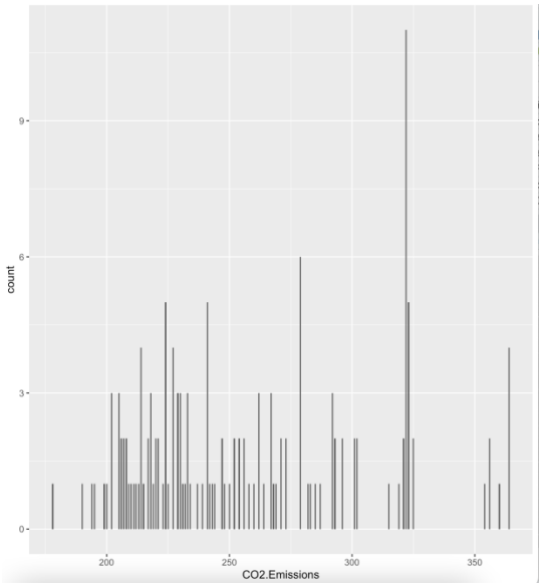


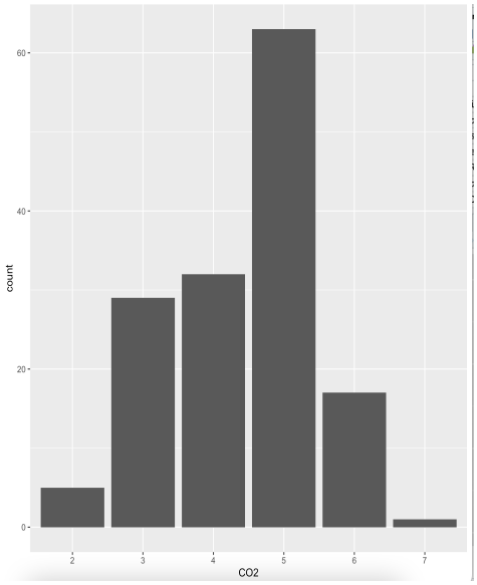
4. What is the proportion for each vehicle class among three brands?



We could see seems like all the model has blue Subcompact more than other class. For the same colour, the proportion varies differently in these three brands. For example:" THe BMW has more Standard SUV than other two brands.

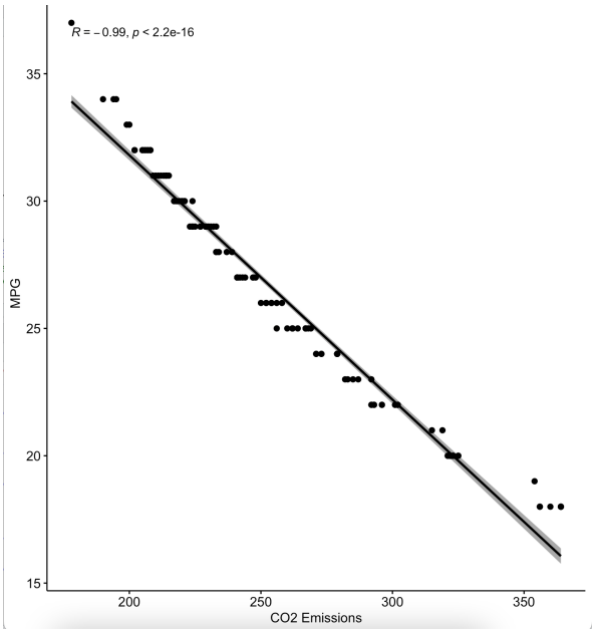
5. Should we use CO2 or CO2.Emissions?





CO2Emissions have more data and better analysing skills than just CO2 Rating.

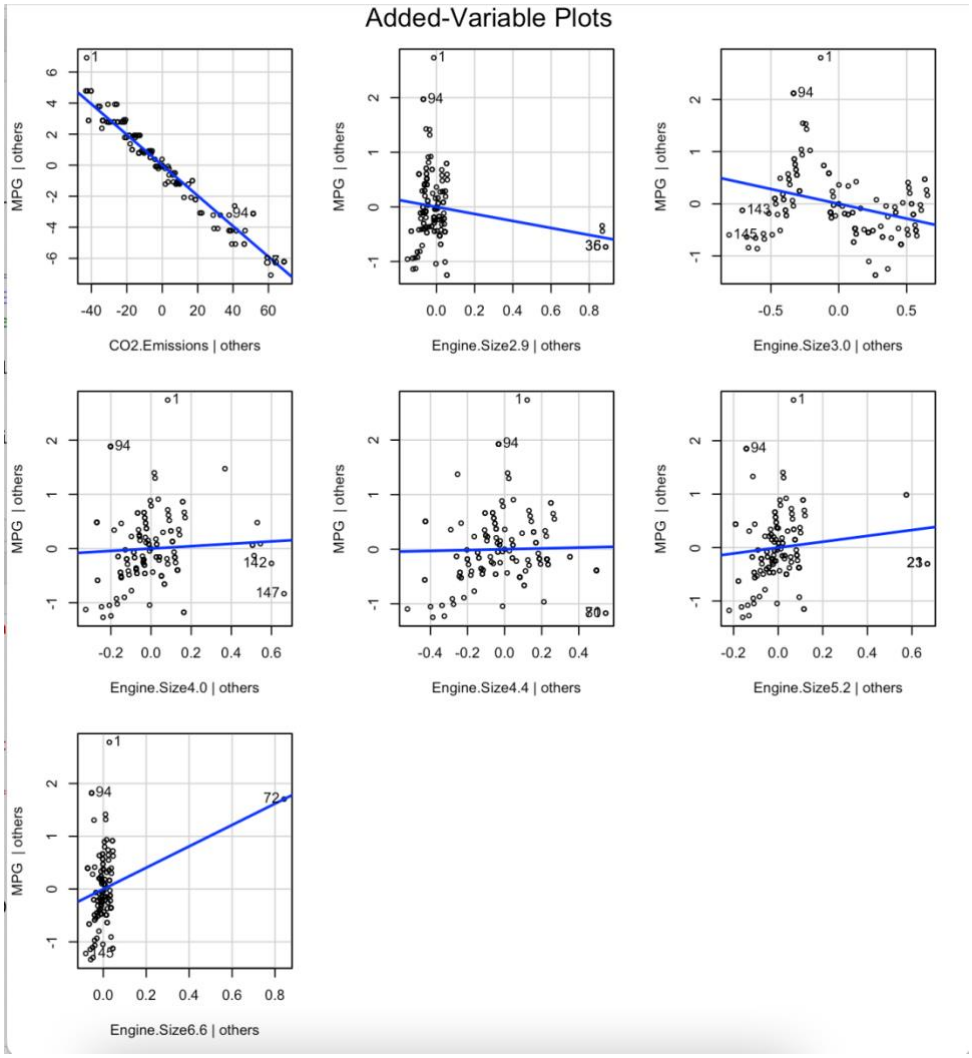
6. Would fuel consumption(MPG) has any correlation with CO2.Emissions?




```
> corr = cor(newData$MPG,newData$CO2.Emissions)
> print(corr)
[1] -0.9864085
```

There is a negative almost perfect linear correlation between MPG and CO2 Emissions.

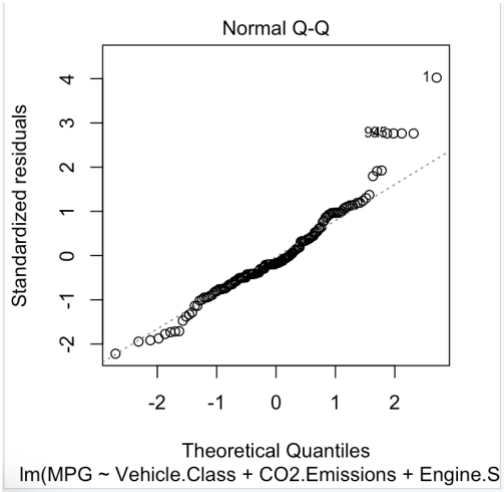
7. Combine both CO2.Emissions and Engine Size first. Any correlation between them and fuel consumption(MPG)?

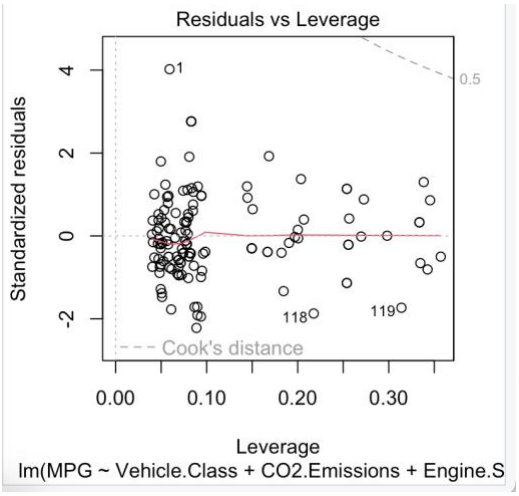


the relationship between the response variable(MPG) and one predictor variable

The x-axis displays a single predictor variable and the y-axis displays the response variable.
The blue line shows the association between the predictor variable and the response variable, while holding the value of all other predictor variables constant.

Check Normality





We could see the line is nearly straight. Considering these three factors are all related with MPG.

Reference

Anable, J., Lane, B., & Banks, N. (2008). Car buyer survey: From ‘mpg paradox’to ‘mpg mirage’. *Research conducted on behalf of the Low Carbon Vehicle Partnership*.

Brown, J. J., Light, C. D., & Gazda, G. M. (1987). Attitudes towards European, Japanese and US cars. European Journal of Marketing.

Engelmann, J., & Richter, F. (2018, September 18). Infographic: Japanese consumers root for Green Vehicle Technologies. Statista Infographics. Retrieved June 9, 2022, from https://www.statista.com/chart/15483/the-main-passenger-car-features-required-by-consumers-in-japan/

Europe Germany - global fuel economy. (2017). Retrieved June 9, 2022, from https://www.globalfueleconomy.org/media/461045/europe_germany.pdf

Halder, A., & Chakraborty, A. (2018). Tata motors’ acquisition of Jaguar and Land rover: A research study.

Levinson, A. (2017). 7 The hidden American tax on imported cars: Fuel economy standards instead of tariffs. Economics and Policy in the Age of Trump, 79.

Shirbhayye, V., Kurmi, D., Dyavanapalli, S., Prasad, A. S. H., & Lal, N. (2020, January). An accurate prediction of MPG (Miles per Gallon) using linear regression model of machine learning. In *2020 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-5). IEEE.