MJ DonldsonMGMT6233|Module2

Question 1:

```
> summary(auto)
       V1              V2              V3              V4            V5             V6              V7
 Min.   : 9.00   Min.   :3.000   Min.   : 68.0   90     : 20   Min.   :1613   Min.   : 9.50   Min.   :70.00
 1st Qu.:18.00   1st Qu.:4.000   1st Qu.: 98.0   150    : 19   1st Qu.:2217   1st Qu.:14.00   1st Qu.:73.00
 Median :23.00   Median :4.000   Median :140.0   88     : 19   Median :2740   Median :15.50   Median :76.00
 Mean   :23.82   Mean   :5.366   Mean   :186.7   110    : 18   Mean   :2941   Mean   :15.76   Mean   :76.21
 3rd Qu.:29.50   3rd Qu.:6.000   3rd Qu.:250.0   100    : 17   3rd Qu.:3532   3rd Qu.:17.30   3rd Qu.:79.00
 Max.   :46.60   Max.   :8.000   Max.   :455.0   75     : 14   Max.   :5140   Max.   :24.80   Max.   :82.00
                                                 (Other):276
       V8                      V9
 Min.   :1.000   ford pinto    :  6
 1st Qu.:1.000   amc matador   :  5
 Median :1.000   ford maverick :  5
 Mean   :1.595   toyota corolla:  5
 3rd Qu.:2.000   amc gremlin   :  4
 Max.   :3.000   amc hornet    :  4
                 (Other)       :354
```

Which variables are quantitative, and which are qualitative?
V1 ,V2, V3,V5,V6, V7 and V8 are quantitative. V4 and V9 are qualitative.

Question 2:

What is the range of each quantitative variable? You can answer this using the range() function.
v1-(9.0:46.6)
v2- (3:8)
v3- (68:455)
v5- (1613:5140)
v6- (9.50:24.80)
v7- (70.00:82.00)
v8- (1:3)

Question 3:

What is the mean and standard deviation of each quantitative predictor?

Standard Deviation
v1- 7.800271
v2- 1.660147
v3- 99.21695
v5- 845.4093

v6- 2.574206
v7- 3.570694
v8- 0.8094202

<u>Question 4</u>

Remove the 10th through 85ths observations. What is the range, mean, and standard deviation of each variable in this subset of the data?
autoSubset<-auto[-c(10:85),]
mean(autoSubset[,2])

Range: range(autoSubset[,x])

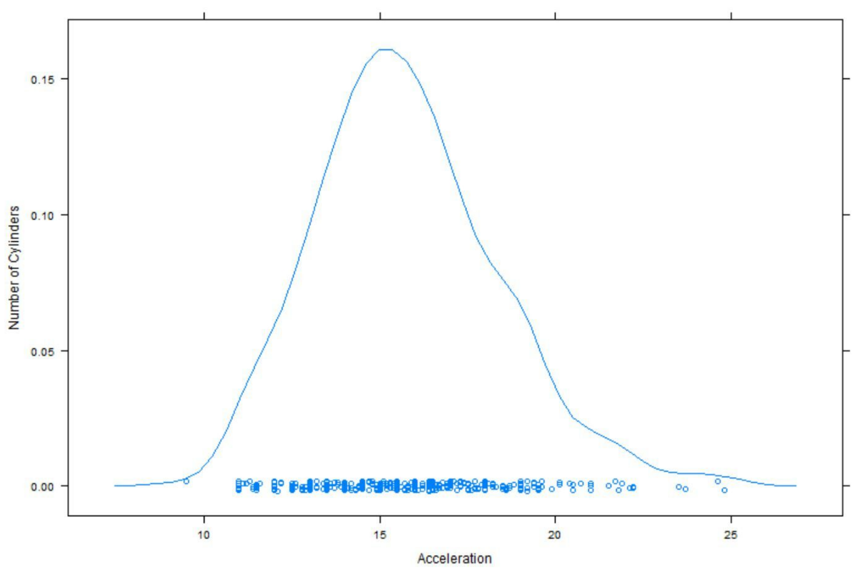
v1- 11.0:46.6
v2- 3:8
v3- 68:400
v5- 1649:4997
v6- 9.5:24.8
v7- 70:82
v8-1:3

Mean: mean(autoSubset[,x])
v1- 25.1557
v2- 5.166124
v3- 172.8241
v5- 2838.313
v6- 16.01336
v7- 77.34202
v8- 1.651466

Standard Deviation:sd(autoSubset[,x])
v1- 7.581004
v2- 1.542987
v3- 88.36264
v5- 761.1744
v6- 2.485458
v7- 3.049209
v8- 0.8278384

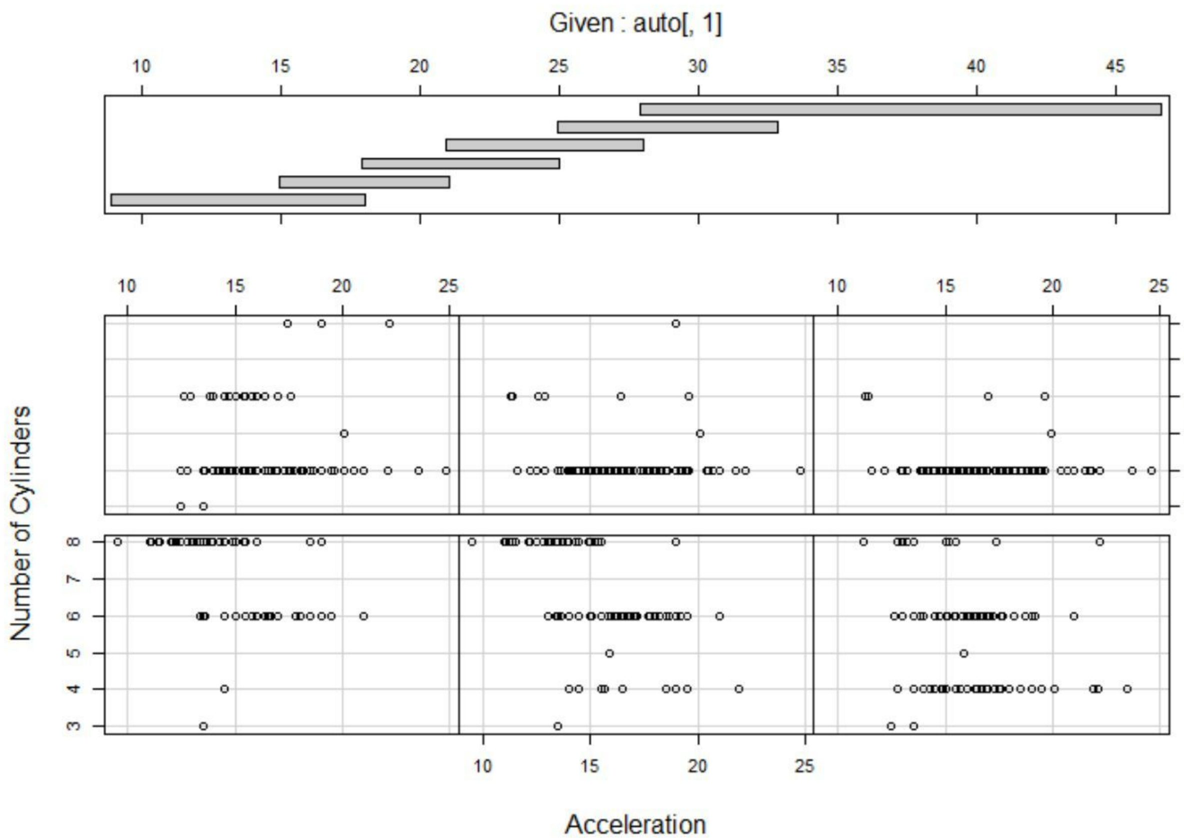
<u>Question 5:</u>Using the full data set again, investigate the variables graphically using scatterplots or other visualizations of your choice. Create some plots highlighting the relationships among the predictors. Write a first blog post about your findings: post at least one figure and comment your findings (aim for about 1-2 paragraphs). You can also post the R code you used to create the plot. You can use Pretty R to nicely format yourR code. Hint: You can start explore the ggplot2 homepage for an overview of different visualizations you might want to use.

Graph 1:
densityplot(auto[,2]~auto[,6], xlab="Acceleration", ylab="Number of Cylinders")



Here, interestingly, Acceleration increases as the number of cylinders increases and then drops off. Suggesting another variable may come into play, the plot itself is not accurate, or I don’t know as much about cars as I thought I did.

Graph 2:
coplot(auto[,2]~auto[,6]|auto[,1],xlab="Acceleration",ylab="Number of Cylinders",main="MPG Compared to NUmber of Cylinders and Acceleration ")

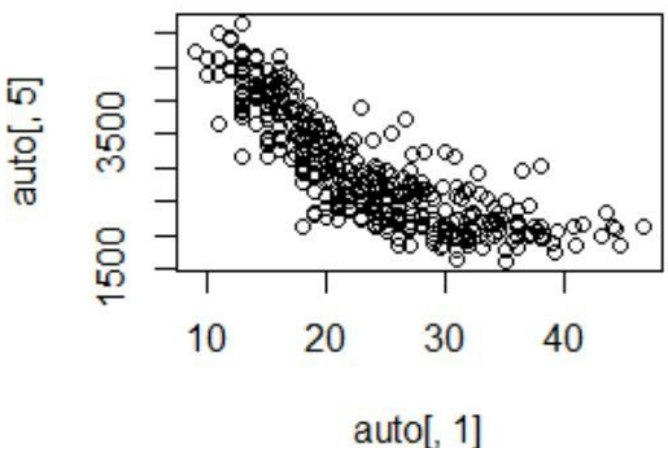


This plot gives a bivariate plot (x,y) for each interval of z. In this case, z is Given:auto[,1] which represents MPG. MPG is divided into 6 intervals, and Number of Cylinders (x) and Acceleration (y) are plotted for both. If we are looking for a car that gets an MPG >22, has a high Acceleration and fewest # of cylinders, we can start by looking at the bottom three graphs. In the last graph we find the highest values for acceleration with the lowest number of cylinders (4) offering the highest MPG. *Not sure if high acceleration = to speed? Might be an inverse relationship. If so, we may have found the slowest but most efficient car in this example

Question 6:

Suppose that we wish to predict gas mileage (mpg) on the basis of other variables. You your plots suggest that any of the other variables might be useful in predicting mpg? Justify your answer. Add this discussion to your blog post.

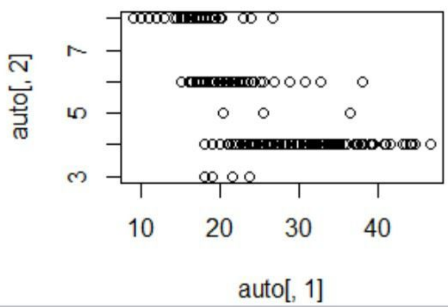
Graph 1:
plot(auto[,1],auto[,5])



Here MPG (X) is plotted against Weight (Y). As expected, an inverse relationship exists between the two; as Weight goes up, MPG goes down.

Graph 2:

plot(auto[,1],auto[,2])



Here again, it is possible to observe an inverse relationship. This time, MPG has been plotted against the number of cylinders. As the number of cylinders goes up, MPG goes down. It is possible to see a fair degree of overlap around ~15-22 mpg.