Problem 1

Solution:

a. Quantitative: mpg" "cylinders" "displacement" "horsepower" "weight" "acceleration" "year"

Qualitative: "origin" "name"

b. Range of each quantitative Variables

mpg1	mpg2	cylinders1	cylinders2	displacement1
9.0	46.6	3.0	8.0	68.0

displacement2	horsepower1	horsepower2	weight1	weight2
455.0	46.0	230.0	1613.0	5140.0

acceleration1	acceleration2	year1	year2
8.0	24.8	70.0	82.0

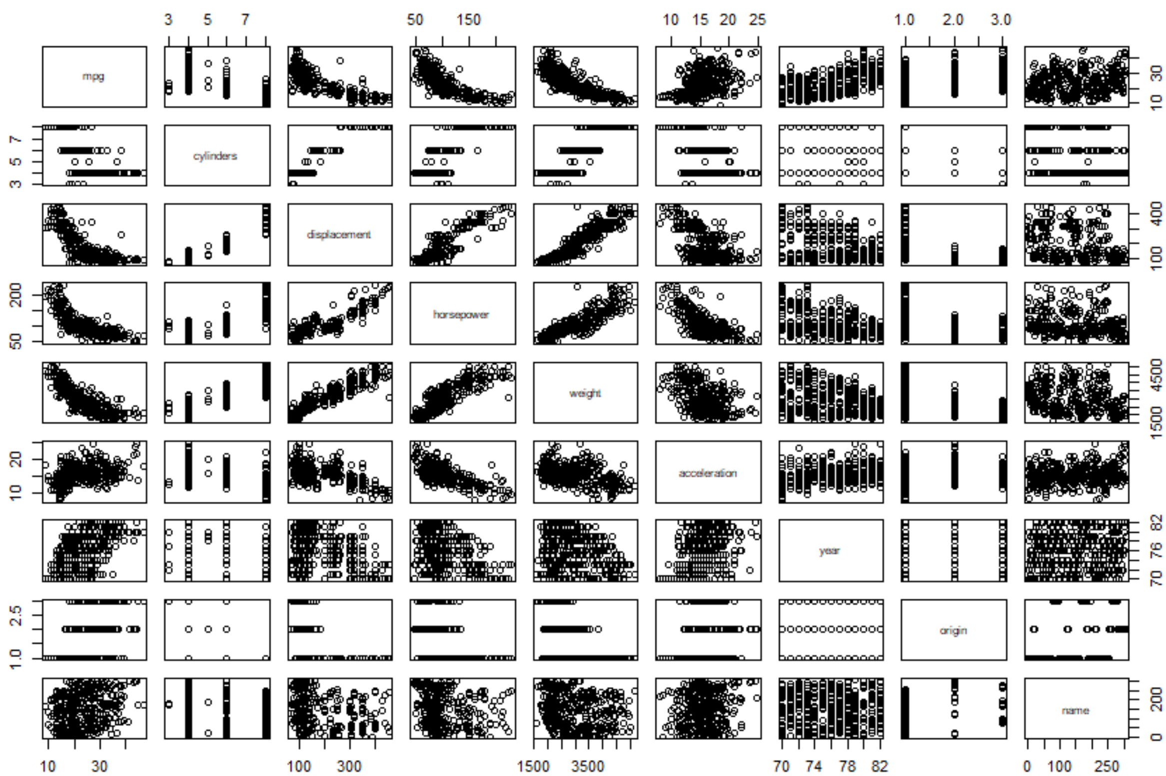
c. Mean and Standard deviation

Variables	Mean	Standard Deviation
mpg	23.445918	7.805007
cylinders	5.471939	1.705783
displacement	194.411990	104.644004
horsepower	104.469388	38.491160
weight	2977.584184	849.402560
acceleration	15.541327	2.758864
year	75.979592	3.683737

d.

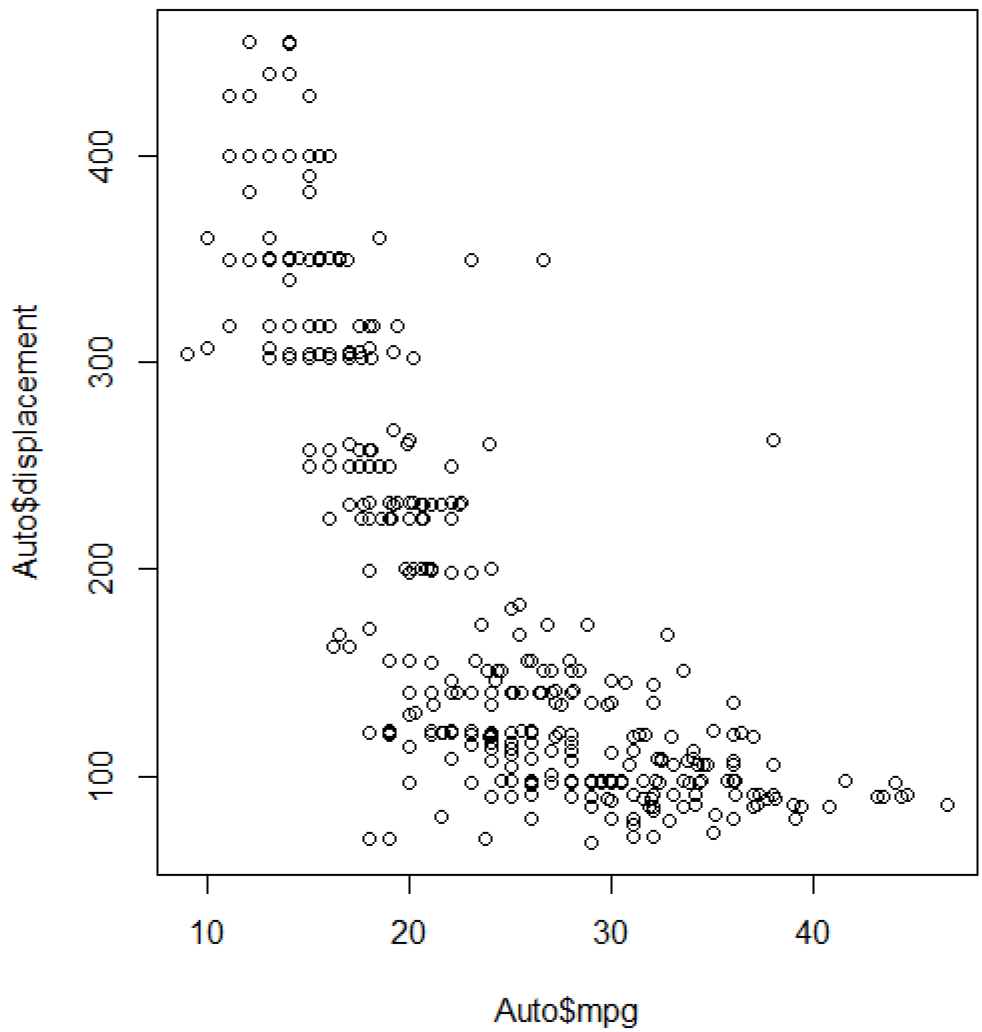
	mpg	cylinders	displacement	horsepower	weight	acceleration	year
min	11.00	3.00	68.00	46.00	1649.00	8.50	70.00
max	46.60	8.00	455.00	230.00	4997.00	24.80	82.00
mean	24.40	5.37	187.24	100.72	2935.97	15.73	77.15
sd	7.87	1.65	99.68	35.71	811.30	2.69	3.11

e.



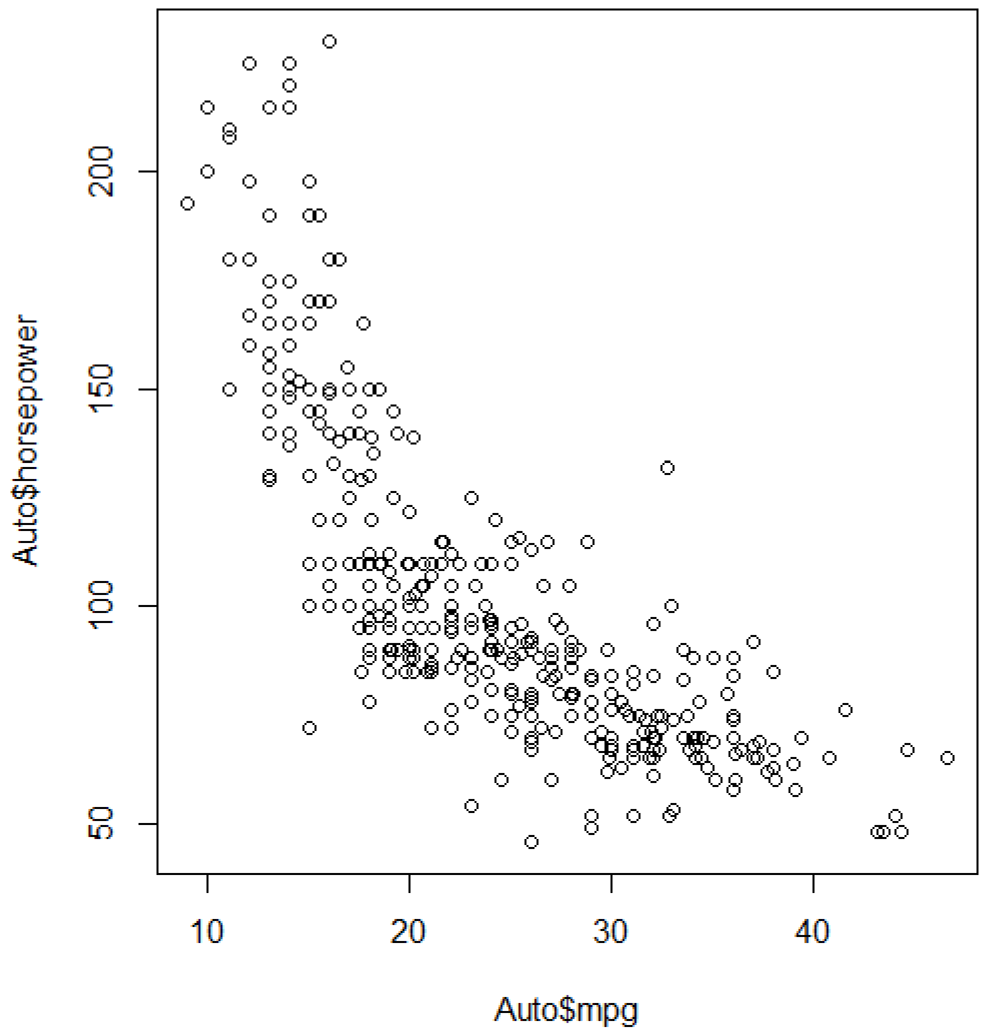
mpg
cylinders
displacement
horsepower
weight
acceleration
year
origin
name

Some additional plots
Mpg vs displacement



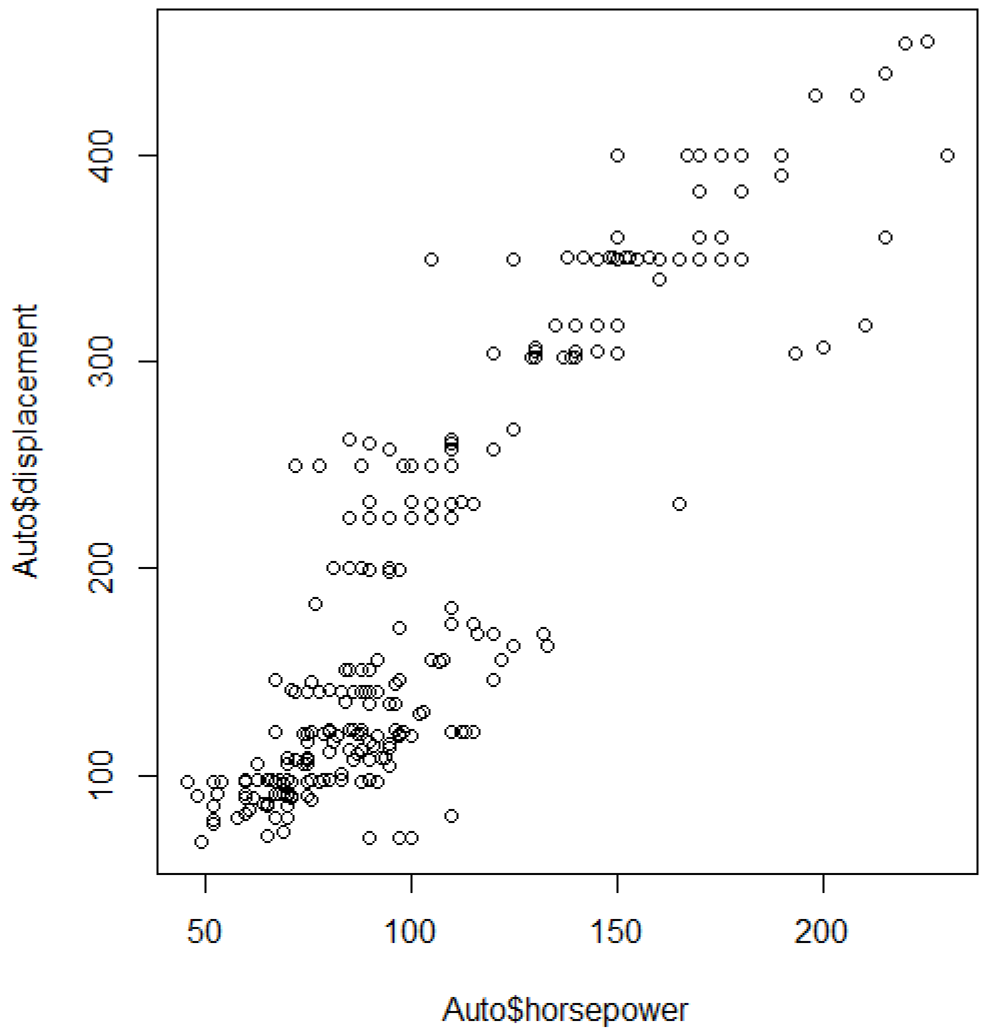
Showing decreasing trend which means negatively correlated.

Mpg vs horsepower



Showing decreasing trend which means negatively correlated.

Horsepower vs displacement

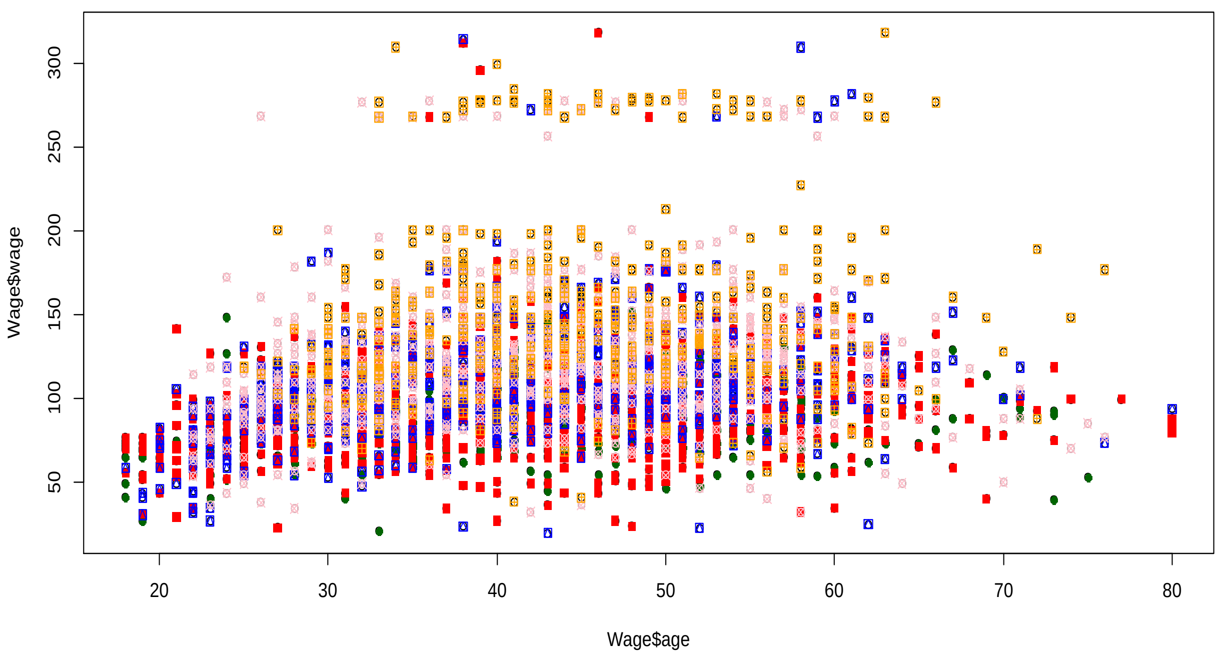


From this plot we can see a strong linear increasing trend in between these two predictors. Positively correlated.

f. Yes, year, acceleration, and origin would be good predictors of mpg.

Problem 2

Solution: Scatter plot



Problem 3

Solution:

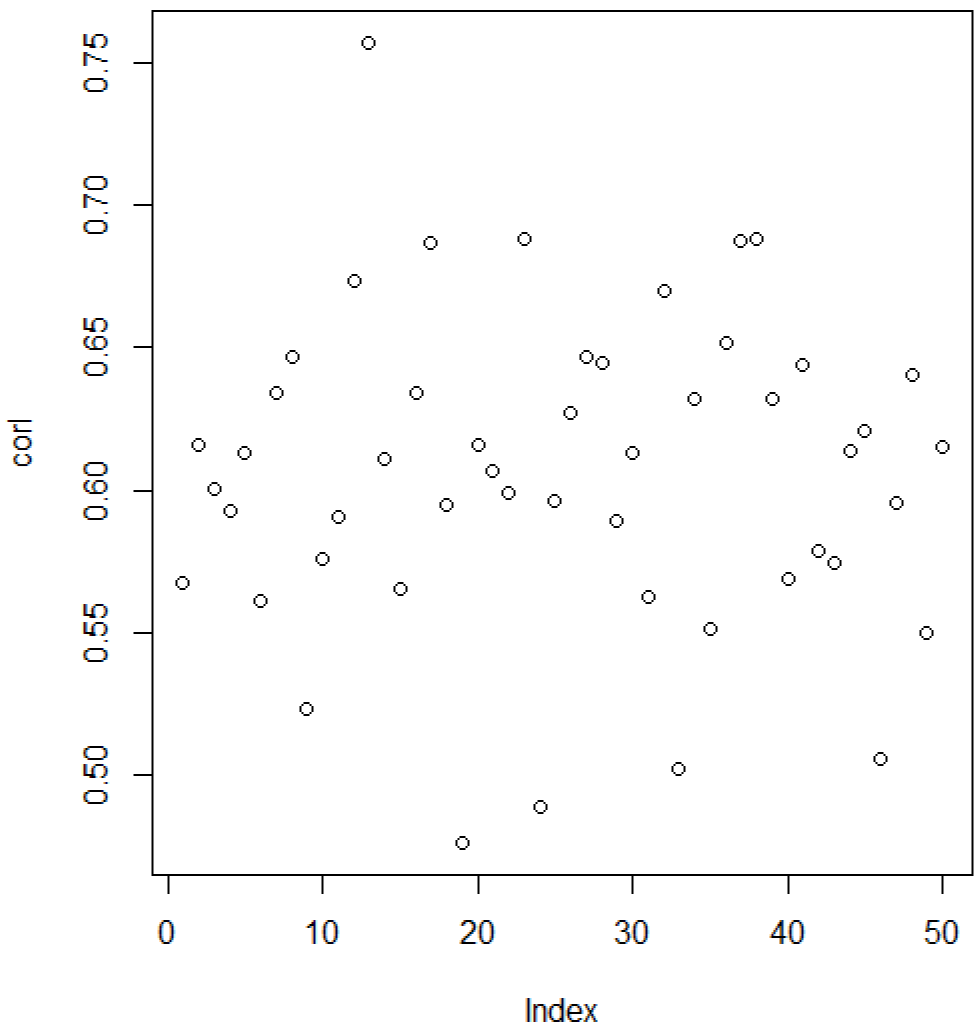
	1. < HS Grad	2. HS Grad	3. Some College	4. College Grad	5. Advanced Degree
1. Industrial	**190**	**636**	**342**	**274**	**102**
2. Information	**78**	**335**	**308**	**411**	**324**

Problem 4

Solution:

Correlation: 0.594998

After repeating 50 times



Theoretical value should be approximately 0.60.

R-Code

```
############Problem 1#############
Auto=read.csv ("Auto.csv")
fix(Auto)



Auto=read.csv ("Auto.csv", header =T,na.strings ="?")
fix(Auto)
dim(Auto)
Auto=na.omit(Auto)
dim(Auto)
names(Auto)
head(Auto)

attach(Auto)
sapply(Auto, class)
summary(Auto)
Auto$origin <- factor(Auto$origin, levels=1:3, labels=c("U.S.", "Europe", "Japan"))
sapply(Auto, class)
quantitative <- sapply(Auto, is.numeric)
quantitative




x=c("mpg"=range(mpg),"cylinders"=range(cylinders),"displacement"=range(displacement)
,"horsepower"=range(horsepower)
,"weight"=range(weight)
,"acceleration"=range(acceleration)
,"year"=range(Auto[,7]))
```

```
summary(Auto)

mpgms=c(mean=mean(mpg), standarddeviation=sd(mpg))
cylindersms=c(mean=mean(cylinders), standarddeviation=sd(cylinders))
displacementms=c(mean=mean(displacement), standarddeviation=sd(displacement))
horsepowerms=c(mean=mean(horsepower), standarddeviation=sd(horsepower))
weightms= c(mean=mean(weight), standarddeviation=sd(weight))
accelarationms=c(mean=mean(acceleration), standarddeviation=sd(acceleration))
yearms= c(mean=mean(Auto[,7]), standarddeviation=sd(Auto[,7]))

MeanandStad=c(mpgms,cylindersms,displacementms,horsepowerms,weightms,accelarationms,yearms)
MeanandStad

newmsd <- sapply(Auto[-10:-85, quantitative], function(x) round(c(range(x), mean(x), sd(x)), 2))
rownames(newmsd) <- c("min", "max", "mean", "sd")
newmsd

pairs(Auto)
plot(Auto$mpg ,Auto$displacement)
plot(Auto$mpg ,Auto$horsepower)
plot(Auto$horsepower ,Auto$displacement)
```

```r
#############Problem 2#########
require(ISLR)
  data(Wage)
    head(Wage)
  plot(Wage$age, Wage$wage)
levels(Wage$education)

 gp1 = (Wage$education == "1. < HS Grad")
      points(Wage[gp1,2], Wage[gp1,11], pch = 16, col="darkgreen")

gp2 = (Wage$education ==  "2. HS Grad")
points(Wage[gp2,2], Wage[gp2,11], pch = 15, col="red")

gp3 = (Wage$education ==  "3. Some College")
 points(Wage[gp3,2], Wage[gp3,11], pch = 14, col="blue")

gp4 = (Wage$education ==  "4. College Grad" )
 points(Wage[gp4,2], Wage[gp4,11], pch = 13, col="pink")

gp5 = (Wage$education == "5. Advanced Degree")
 points(Wage[gp5,2], Wage[gp5,11], pch = 12, col="orange")
```

```
########Problem 3#########

mytable=table(Wage$jobclass, Wage$education)

ftable(mytable)




#####Problem 4#########

X=rnorm(100)

Y=rnorm(100)

var1= 2*X+Y

var2= 2*X-Y

corr=cor(var1,var2)

corr

###########2nd Part###

nrep=50

for(i in 1:nrep){

X=rnorm(100)

Y=rnorm(100)

var1= 2*X+Y

var2= 2*X-Y

corl[i]=cor(var1,var2)

}

plot(corl)
```