

## HW4\_Sol

Q1.

```
library(ISLR)
data(Auto)
dim(Auto) #392 observations, 9 features

## [1] 392  9

head(Auto) #take a look

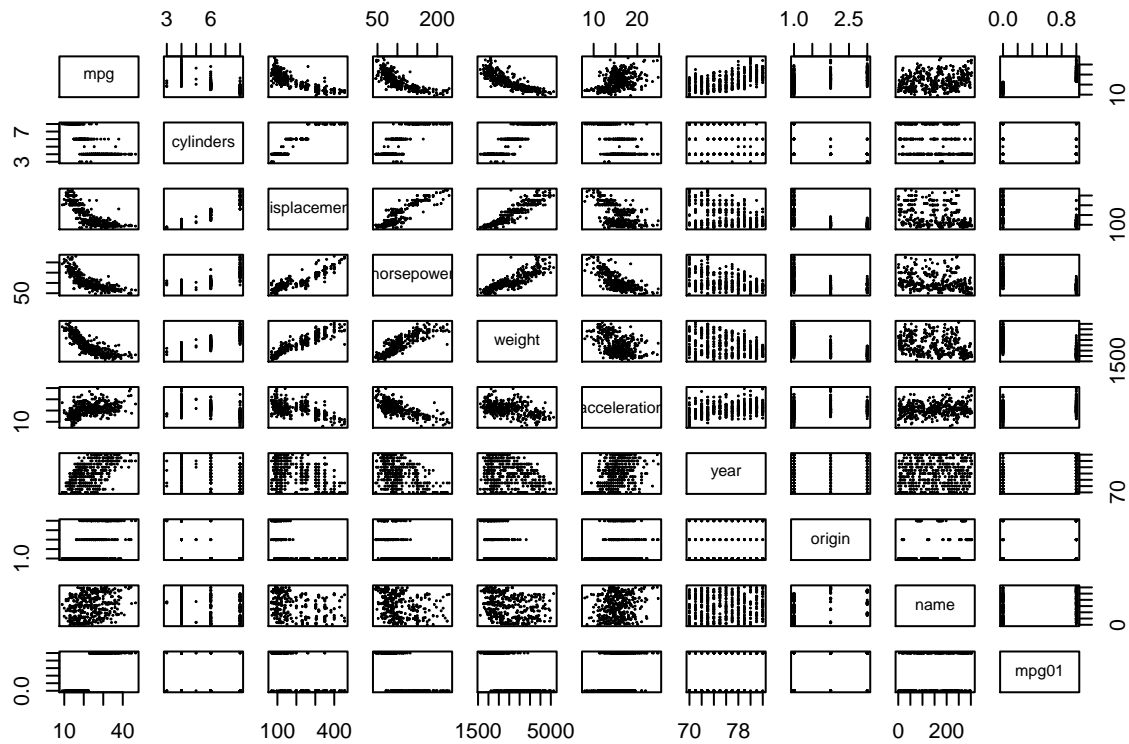
##   mpg cylinders displacement horsepower weight acceleration year origin
## 1   18         8         307         130   3504          12.0    70      1
## 2   15         8         350         165   3693          11.5    70      1
## 3   18         8         318         150   3436          11.0    70      1
## 4   16         8         304         150   3433          12.0    70      1
## 5   17         8         302         140   3449          10.5    70      1
## 6   15         8         429         198   4341          10.0    70      1
##                                name
## 1 chevrolet chevelle malibu
## 2      buick skylark 320
## 3    plymouth satellite
## 4      amc rebel sst
## 5      ford torino
## 6      ford galaxie 500

#we use Auto$mpg01 to add a feature
Auto$mpg01=as.numeric(Auto$mpg>median(Auto$mpg)) #as.numeric() transfers T/F into 0/1
dim(Auto) #now we have 10 features

## [1] 392  10
```

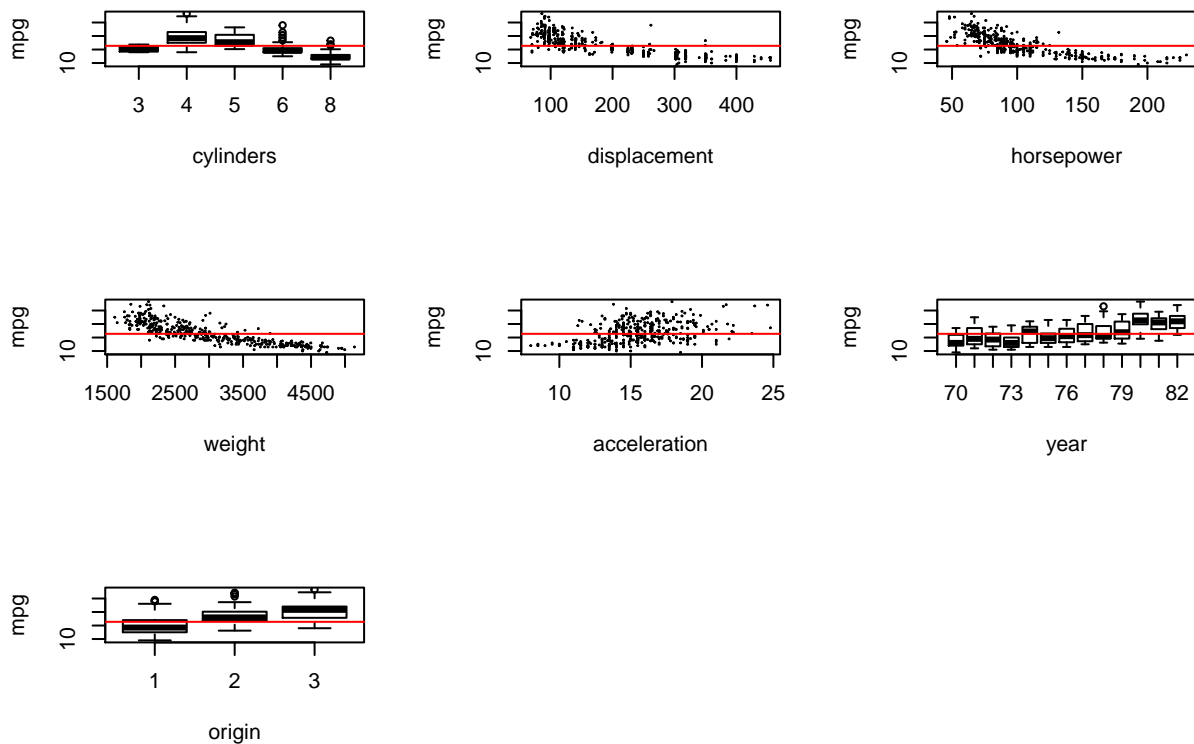
Q2.

```
pairs(Auto,cex=0.1) #cex controls the size of dots
```



*#seams every feature EXCEPT "name" is associated with mpg01*

```
par(mfrow=c(3,3))
boxplot(mpg ~ cylinders,data = Auto)
abline(h = median(Auto$mpg),col="red")
plot(mpg ~ displacement, data = Auto,cex=0.1)
abline(h = median(Auto$mpg),col="red")
plot(mpg ~ horsepower, data = Auto,cex=0.1)
abline(h = median(Auto$mpg),col="red")
plot(mpg ~ weight, data = Auto,cex=0.1)
abline(h = median(Auto$mpg),col="red")
plot(mpg ~ acceleration, data = Auto,cex=0.1)
abline(h = median(Auto$mpg),col="red")
boxplot(mpg ~ year, data = Auto)
abline(h = median(Auto$mpg),col="red")
boxplot(mpg ~ origin, data = Auto)
abline(h = median(Auto$mpg),col="red")
par(mfrow=c(1,1))
```



*#We should only choose numerical features and avoid categorical features  
#I choose horsepower and weight. You may choose other reasonable features*

### Q3.

```
train_index=sample(1:392,300,replace=FALSE) #choose training set randomly
data_train=Auto[train_index,]
data_test=Auto[-train_index,]
```

### Q4

```
library(MASS)
lda.fit=lda(mpg01~horsepower+weight,data=data_train)
lda.pred=predict(lda.fit, data_test)
lda.class=lda.pred$class
true_value=data_test$mpg01
table(lda.class,true_value)
```

```
##           true_value
## lda.class  0  1
##           0 37  2
##           1 10 43
```

```
test_error_LDA=mean(lda.class!=true_value)
test_error_LDA
```

```
## [1] 0.1304348
```

## Q5

```
qda.fit=qda(mpg01~horsepower+weight,data=data_train)
qda.pred=predict(qda.fit, data_test)
qda.class=qda.pred$class
table(qda.class,true_value)
```

```
##           true_value
## qda.class  0  1
##           0 39  2
##           1  8 43
```

```
test_error_QDA=mean(qda.class!=true_value)
test_error_QDA
```

```
## [1] 0.1086957
```