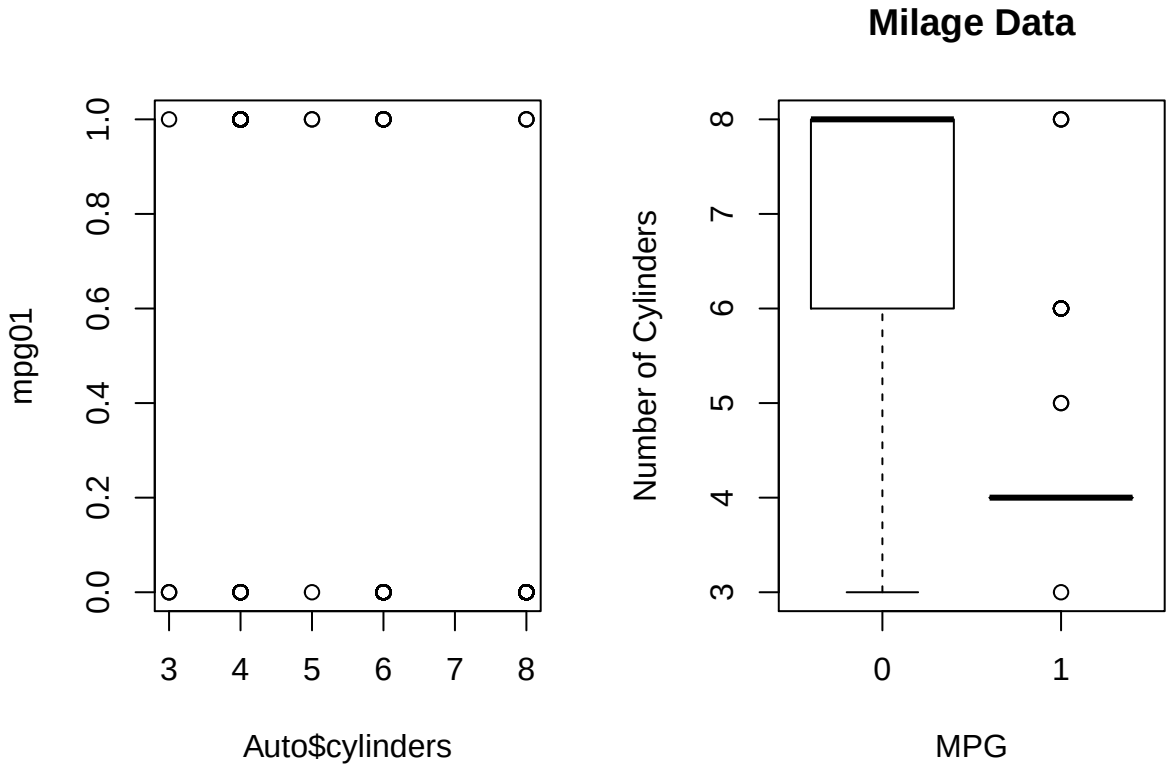
Midterm

Jarod Wright

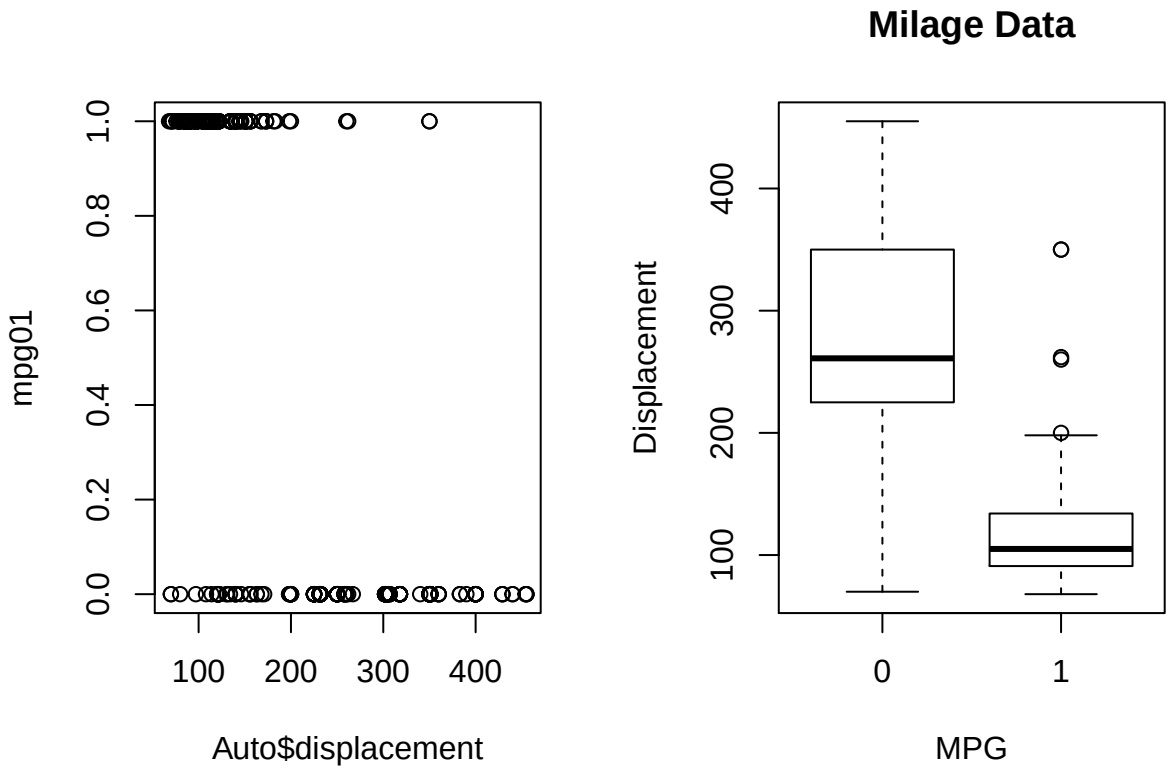
3/14/2021

#1 Cross Validation

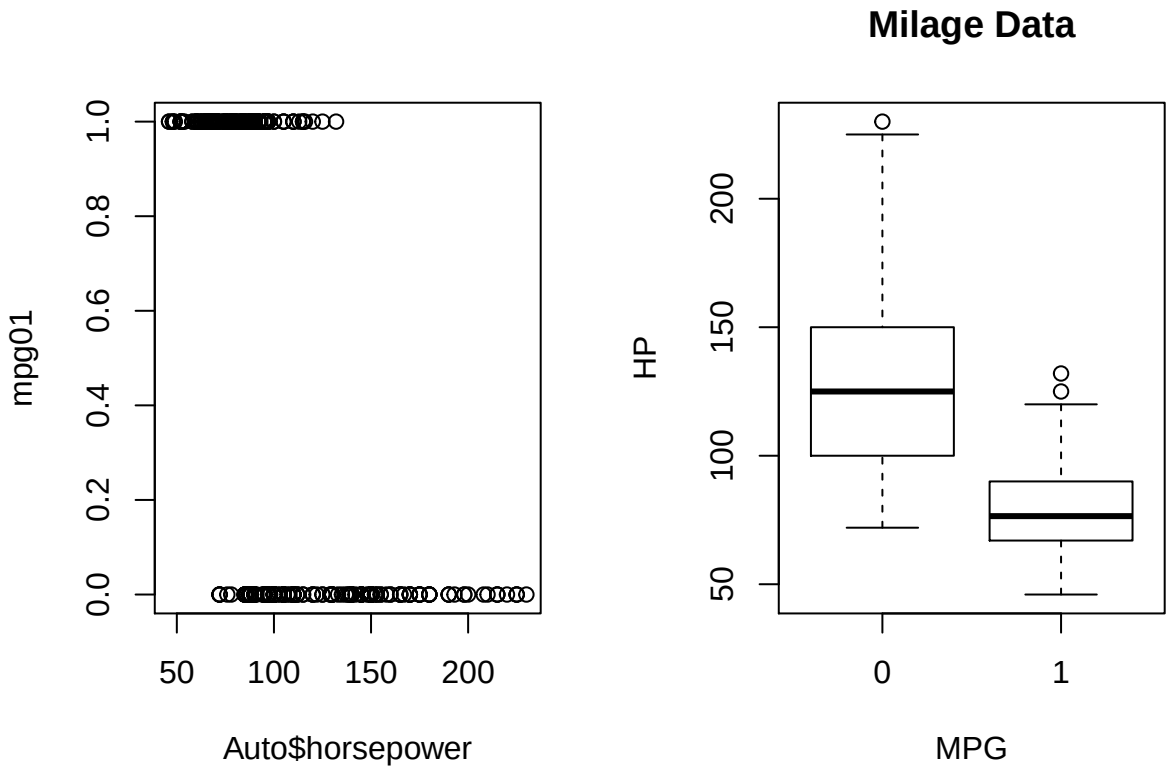
```
#Computational Statistics Exam
# cross validation

set.seed(123)
library(ISLR)
library(tidyverse)
library(caret)
library(boot)
library(caTools)
indx=0
data(Auto)

mpg01=c()
med <- median(Auto$mpg)

for (i in (Auto$mpg))
{ indx=indx+1

  if(i>med){
    mpg01[indx]=1
  }

  else{
    mpg01[indx]=0
  }


}

par(mfrow=c(1,2))
plot(Auto$cylinders,mpg01)
boxplot(cylinders~mpg01,data=Auto, main="Milage Data",ylab="Number of Cylinders", xlab = "MPG")
```



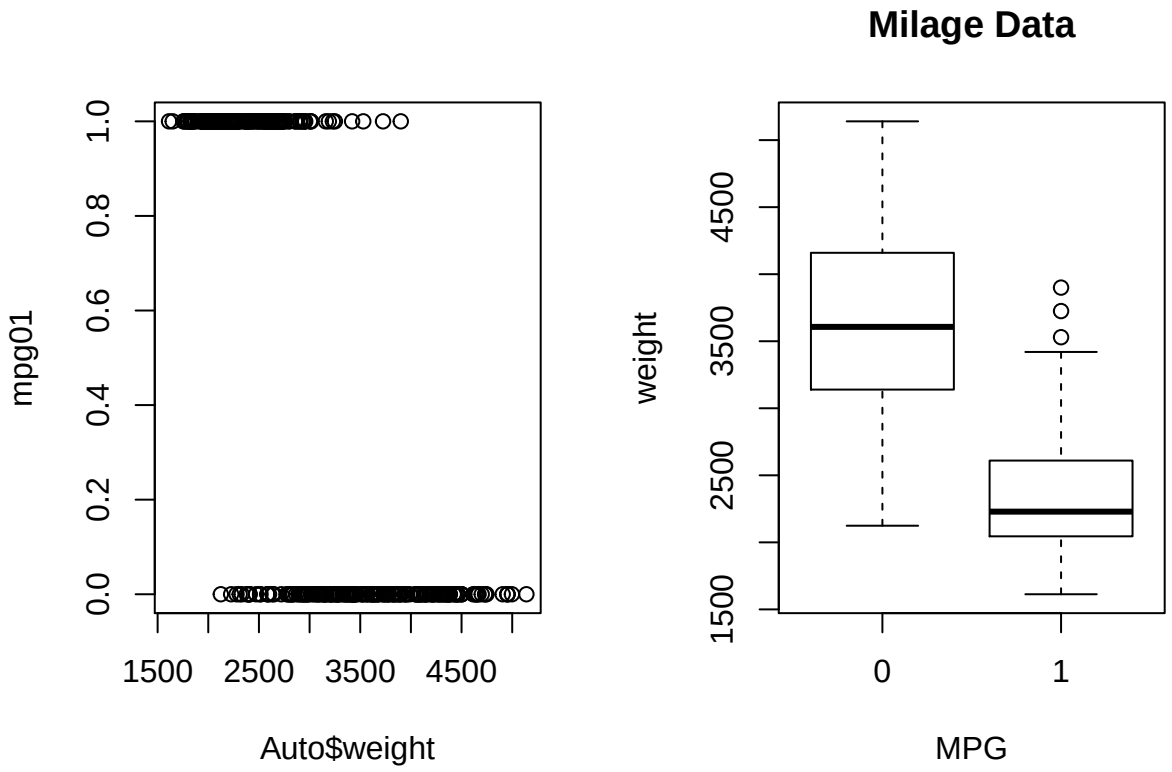

```
par(mfrow=c(1,2))
plot(Auto$displacement,mpg01)
boxplot(displacement~mpg01,data=Auto, main="Milage Data",ylab="Displacement", xlab = "MPG")
```



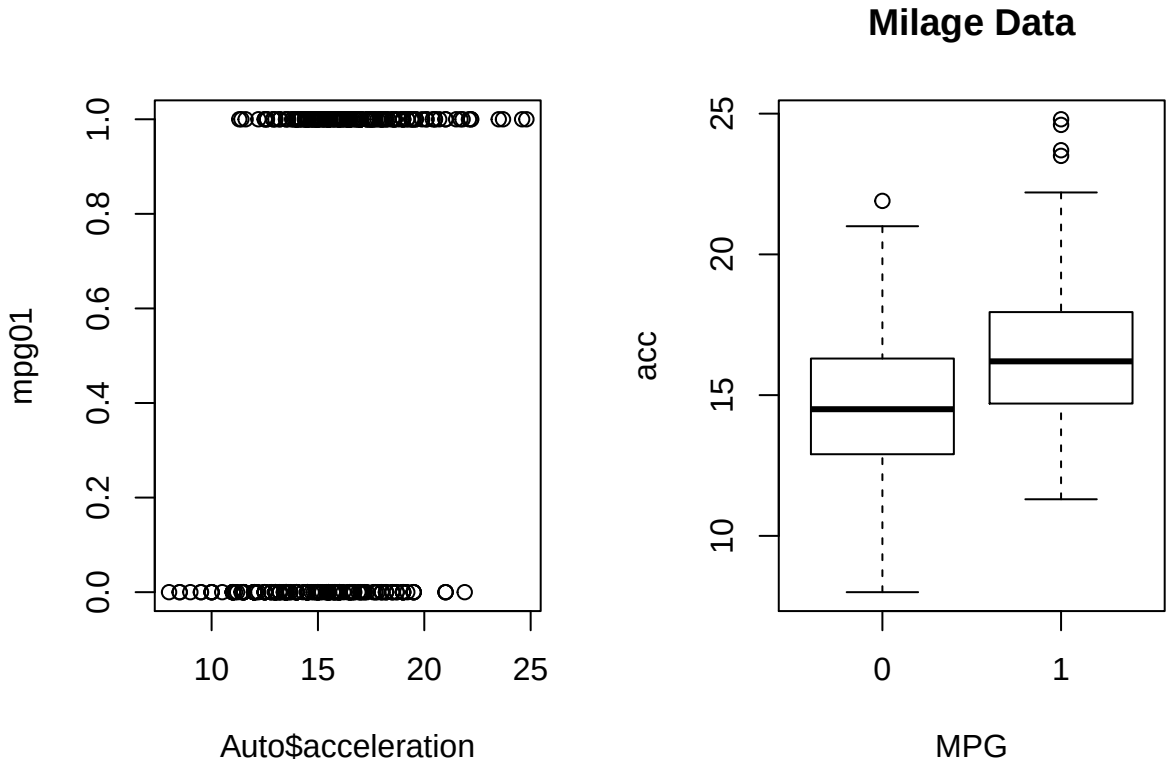
```
par(mfrow=c(1,2))
plot(Auto$horsepower,mpg01)
boxplot(horsepower~mpg01,data=Auto, main="Milage Data",ylab="HP", xlab = "MPG")
```



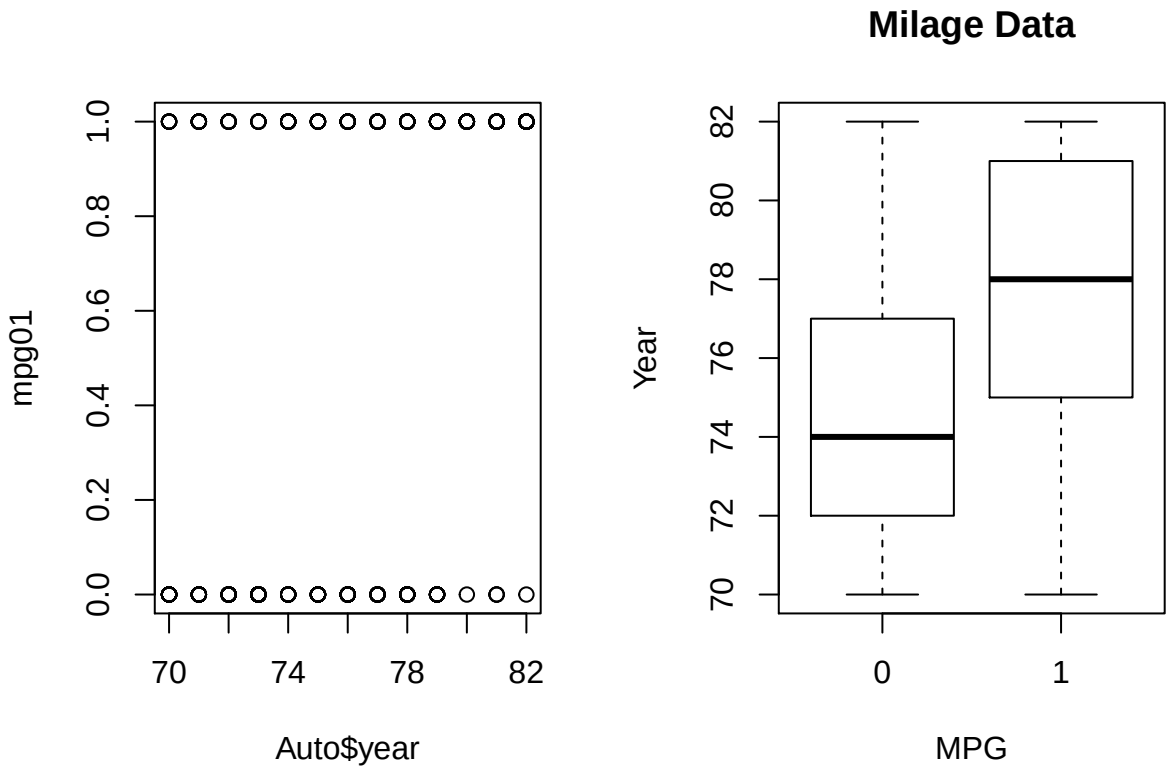

```
par(mfrow=c(1,2))
plot(Auto$weight,mpg01)
boxplot(weight~mpg01,data=Auto, main="Milage Data",ylab="weight", xlab = "MPG")
```



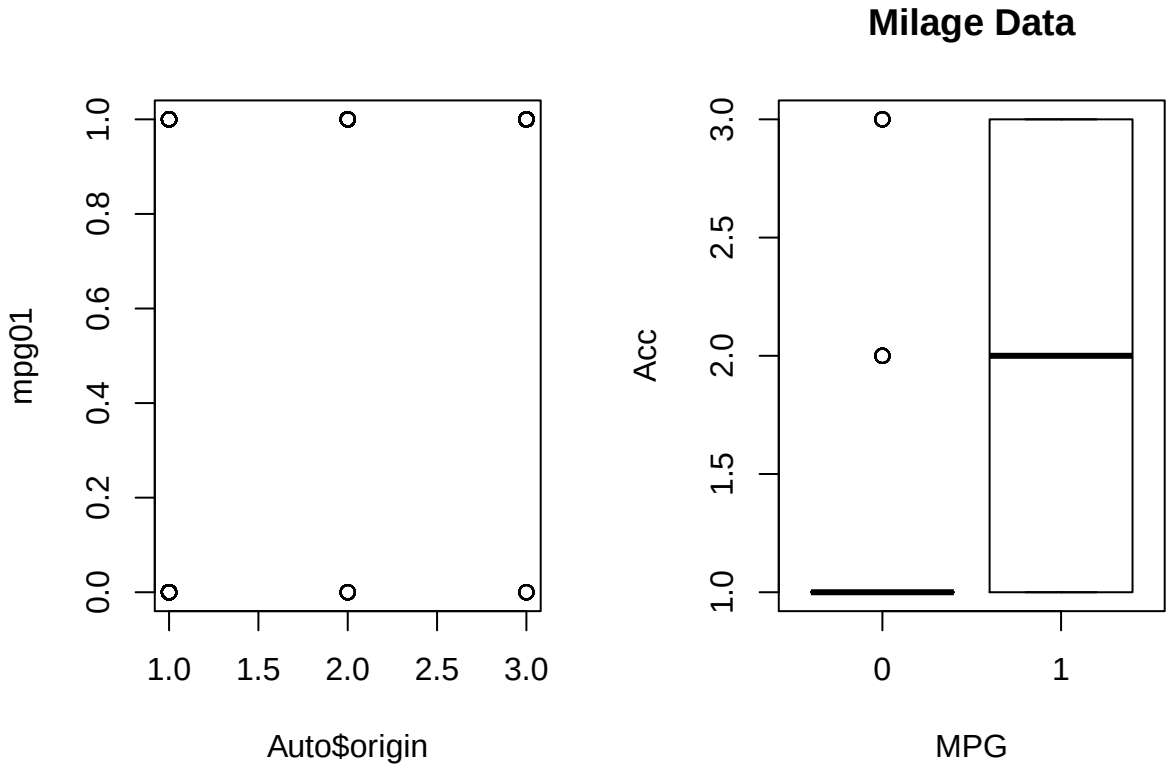

```
par(mfrow=c(1,2))
plot(Auto$acceleration,mpg01)
boxplot(acceleration~mpg01,data=Auto, main="Milage Data",ylab="acc", xlab = "MPG")
```



```
par(mfrow=c(1,2))
plot(Auto$year,mpg01)
boxplot(year~mpg01,data=Auto, main="Milage Data",ylab="Year", xlab = "MPG")
```




```
par(mfrow=c(1,2))
plot(Auto$origin,mpg01)
boxplot(origin~mpg01,data=Auto, main="Milage Data",ylab="Acc", xlab = "MPG")
```




```
#DESCRIBE FINDINGS
#parts (c),(d)

obs <- data.frame(mpg01,Auto$weight) #observations of weight and mpg01

spl = sample.split(mpg01, SplitRatio = 0.7)

train = subset(obs, spl==TRUE)
test = subset(obs,spl == FALSE)

print(dim(train))
```

```
## [1] 274   2
```

```
print(dim(test))
```

```
## [1] 118   2
```

```
model_glm = glm(mpg01~. , family = "binomial",
                data = train, maxit = 100) #probability score of the target categorical variable as 0 o
predictTest = predict(model_glm, newdata = test,
                      type = "response")
predicted_classes <- as.factor(ifelse(predictTest >= 0.5,
                                      1, 0))
mpg01.factor <- factor(test$mpg01)
err.cv.glm=cv.glm(test, model_glm)$delta
print(err.cv.glm)
```

```
## [1] 0.38966137 0.09036114
```

```
print(confusionMatrix(predicted_classes, mpg01.factor))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 53  9
##          1  6 50
##
##                Accuracy : 0.8729
##                  95% CI : (0.799, 0.9271)
##     No Information Rate : 0.5
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.7458
##
##  Mcnemar's Test P-Value : 0.6056
##
##             Sensitivity : 0.8983
##             Specificity : 0.8475
##          Pos Pred Value : 0.8548
##          Neg Pred Value : 0.8929
##              Prevalence : 0.5000
##          Detection Rate : 0.4492
##    Detection Prevalence : 0.5254
##       Balanced Accuracy : 0.8729
##
##        'Positive' Class : 0
##
```