

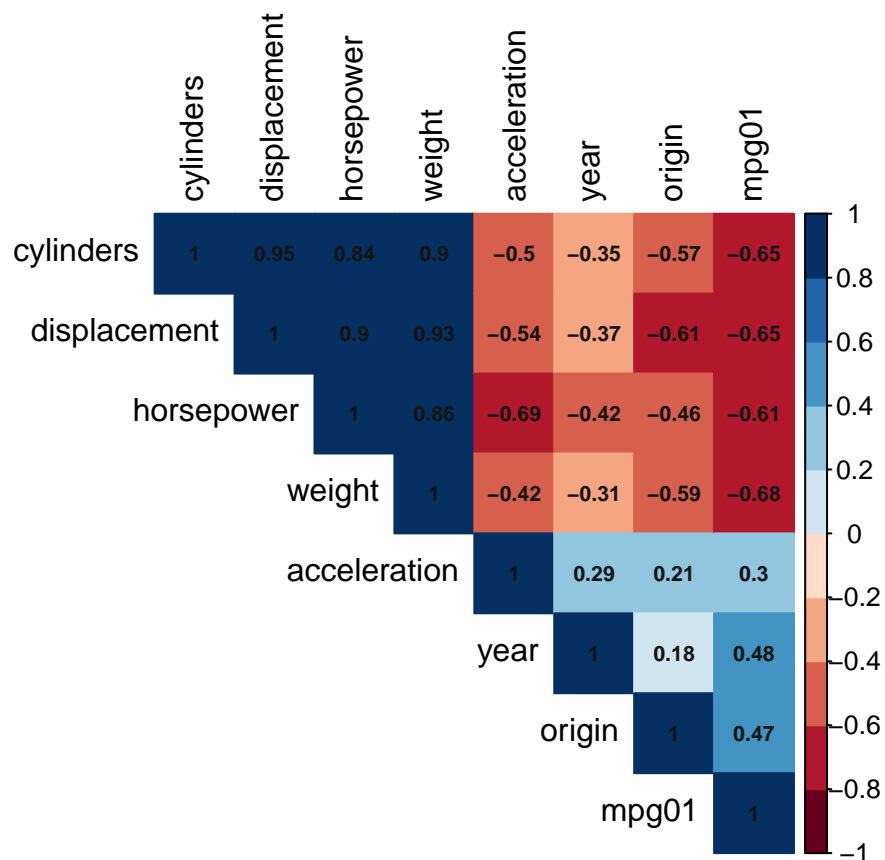
Q2.

(a)

```
library(ISLR)
data("Auto")
mpg01 <- ifelse(Auto$mpg > 25, 1, 0)
Auto <- data.frame(Auto, mpg01)
```

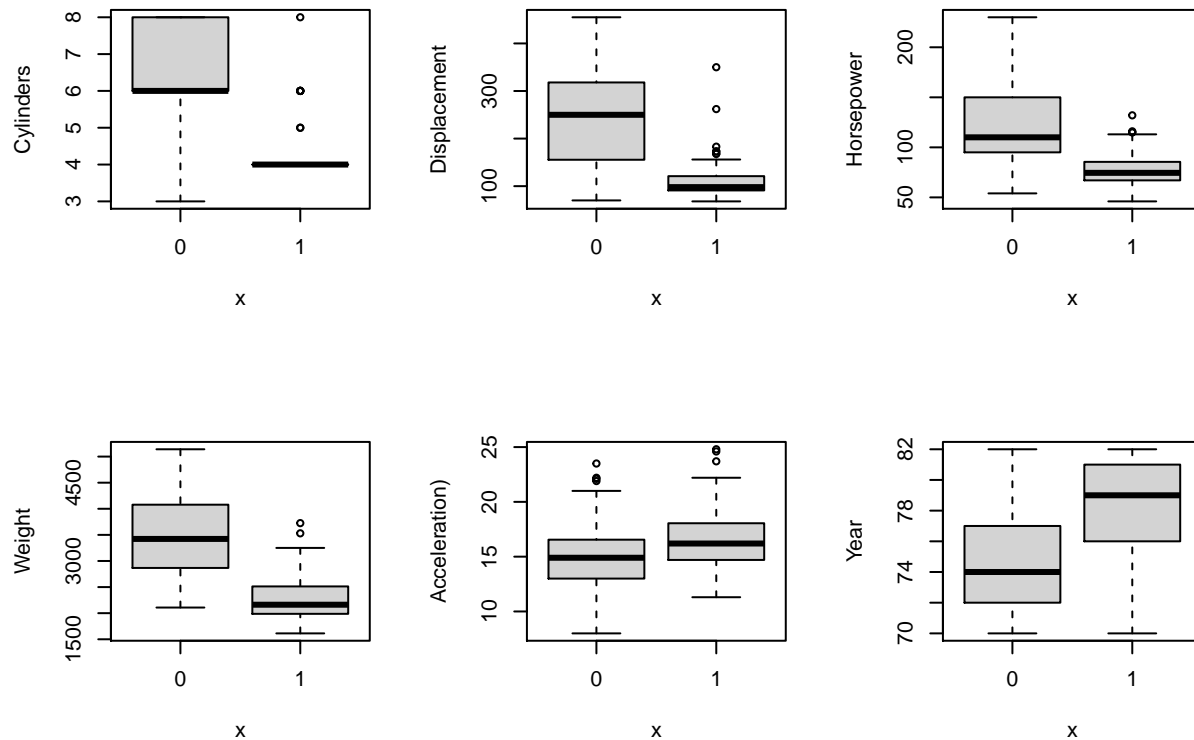
(b)

```
corrplot(cor(Auto[,c(2:8,10)]), method="color", type = "upper", col=brewer.pal(n=10, name="RdBu"),
         tl.col="black",tl.srt=90, addCoef.col = "gray8", diag = T, number.cex = 0.65)
```



```
par(mfrow = c(2, 3))
plot(factor(Auto$mpg01), Auto$cylinders, ylab = "Cylinders")
plot(factor(Auto$mpg01), Auto$displacement, ylab = "Displacement")
plot(factor(Auto$mpg01), Auto$horsepower, ylab = "Horsepower")
plot(factor(Auto$mpg01), Auto$weight, ylab = "Weight")
plot(factor(Auto$mpg01), Auto$acceleration, ylab = "Acceleration")
plot(factor(Auto$mpg01), Auto$year, ylab = "Year")
mtext("Boxplots for cars with above(1) and below(0) median mpg", outer = TRUE, line = -3)
```

Boxplots for cars with above(1) and below(0) median mpg



The variables “cylinders”, “displacement”, “horsepower” and “weight” seem to be highly correlated and useful to predict mpg01

(c)

```
set.seed(123)
num_train <- nrow(Auto) * 0.8
inTrain <- sample(nrow(Auto), size = num_train)
training <- Auto[inTrain,]
testing <- Auto[-inTrain,]
```

(d)

```
lda_model <- lda(mpg01 ~ displacement + horsepower + weight + cylinders, data = training)
pred <- predict(lda_model, testing)
table(pred$class, testing$mpg01)
```

```
##
##      0  1
##  0 38  2
##  1 14 25
```

```
1 - mean(pred$class == testing$mpg01)
```

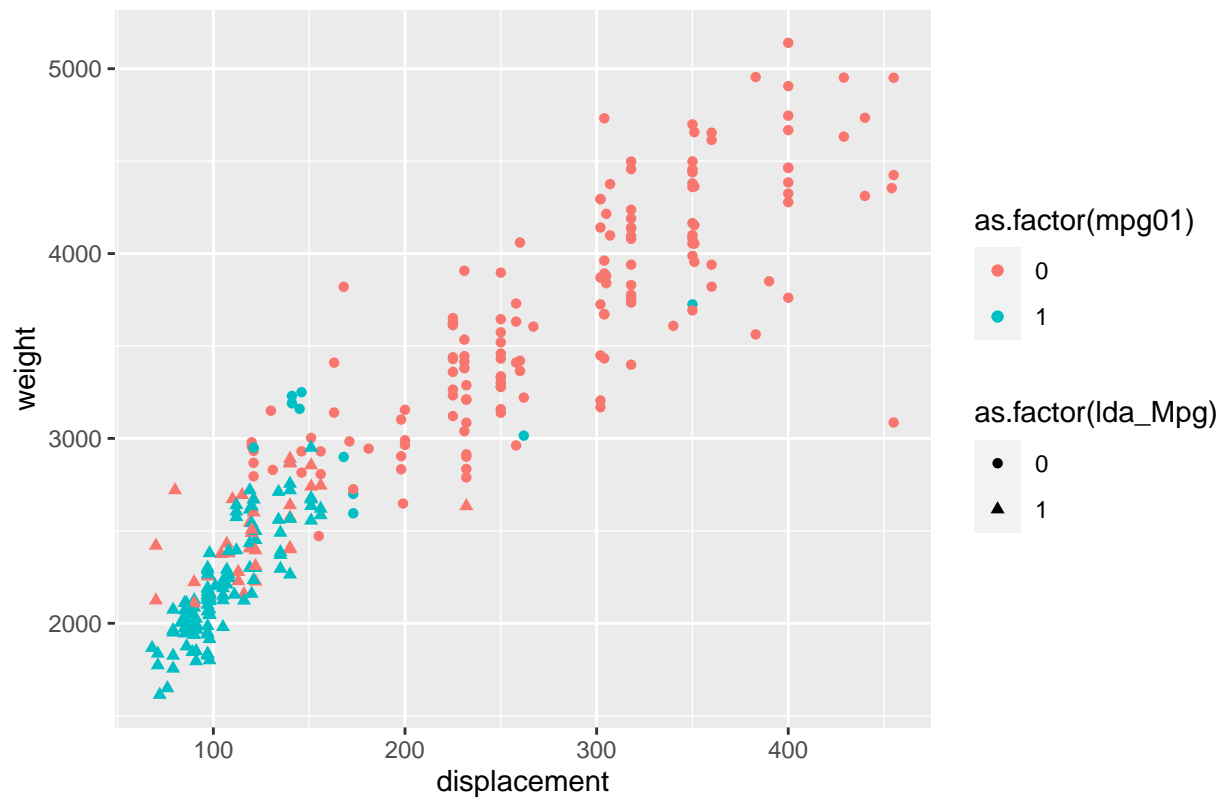
```
## [1] 0.2025316
```

The test error is 0.2025316

```
pred_train <- predict(lda_model, training)
training$lda_Mpg <- pred_train$class
```

```
ggplot(training, aes(x=displacement, y=weight, color = as.factor(mpg01), shape = as.factor(lda_Mpg))) + ge
```

True values vs. Predicted Values of Mpg01 with LDA



(e)

```
qda_model <- qda(mpg01 ~ displacement + horsepower + weight + cylinders, data = training)
pred1 <- predict(qda_model, testing)
table(pred1$class, testing$mpg01)
```

```
##
##      0  1
## 0 41  2
## 1 11 25
```

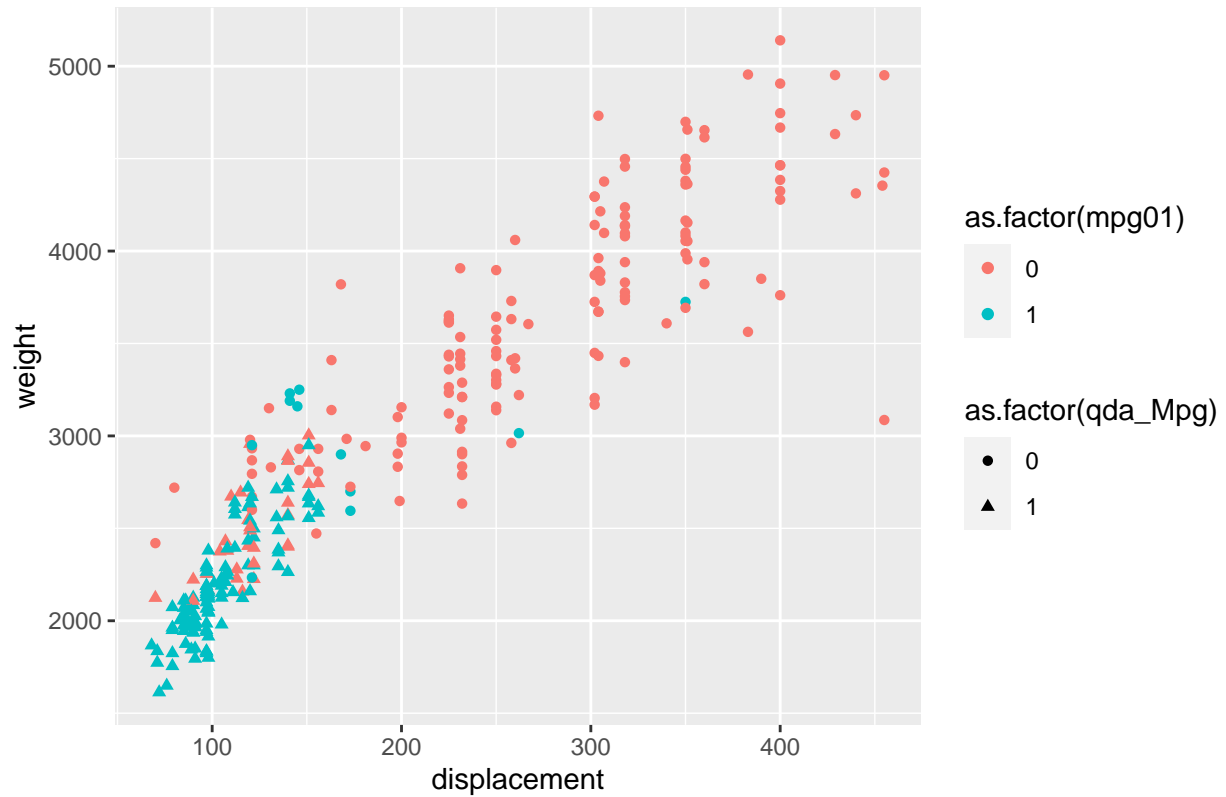
```
1 - mean(pred1$class == testing$mpg01)
```

```
## [1] 0.164557
```

The test error is 0.164557

```
pred1_train <- predict(qda_model, training)
training$qda_Mpg <- pred1_train$class
ggplot(training, aes(x=displacement, y=weight, color = as.factor(mpg01), shape = as.factor(qda_Mpg))) + ge
```

True values vs. Predicted Values of Mpg01 with QDA



(f) The performance of the QDA is better than the performance of LDA in this test. Since we have enough observations to accurately estimate the variances and we have known that the variances are very different between classes, the QDA would perform better as it would take the class-specific covariances into consideration.