4.

(a)

如果x<0.05,只能抽到[0,x+0.05],機率有(100x+5)%,積分得0.375

同理,如果x>0.95只能抽到[x-0.05,1],機率有(105-100x)%,積分得0.375

如果x介於[0.05,0.95]間,抽到[x-0.05,x+0.05]的機率都為1 ,積分得9

9+0.375+0.375=9.75

9.75%

(b)

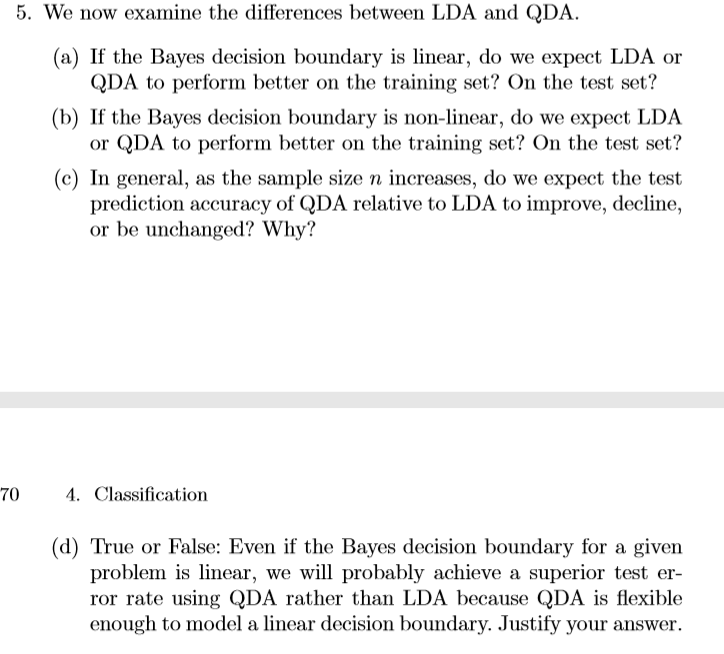
機率為:0.0975^2=0.0950625

(c)

0.0975^100=7.951729e-102,趨近於0

(d)

knn是由x附近的點去估計,由a,b,c可見當維度越高的時候,越沒有附近的樣本點去作估計。



(a)

QDA,LDA , QDA在training set 上表現一定比LDA好,但是在如果真實資料近似線性的話,QDA在testing set上表現會比LDA差 。這種狀況稱為overfitting。

(b)

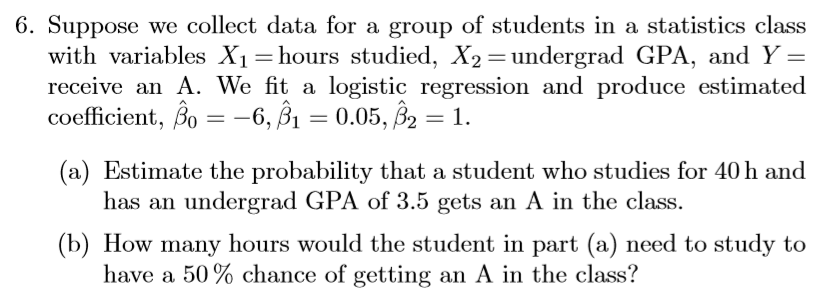
QDA,QDA, QDA在training set 上表現一定比LDA好,且如果如果真實資料是非線性的話,QDA在testing set上表現也會比LDA好。

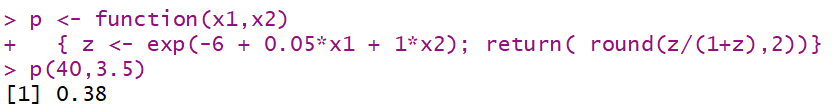
(c)

Improve,因為越多的資料可以避免QDA的overfitting的問題

(d)

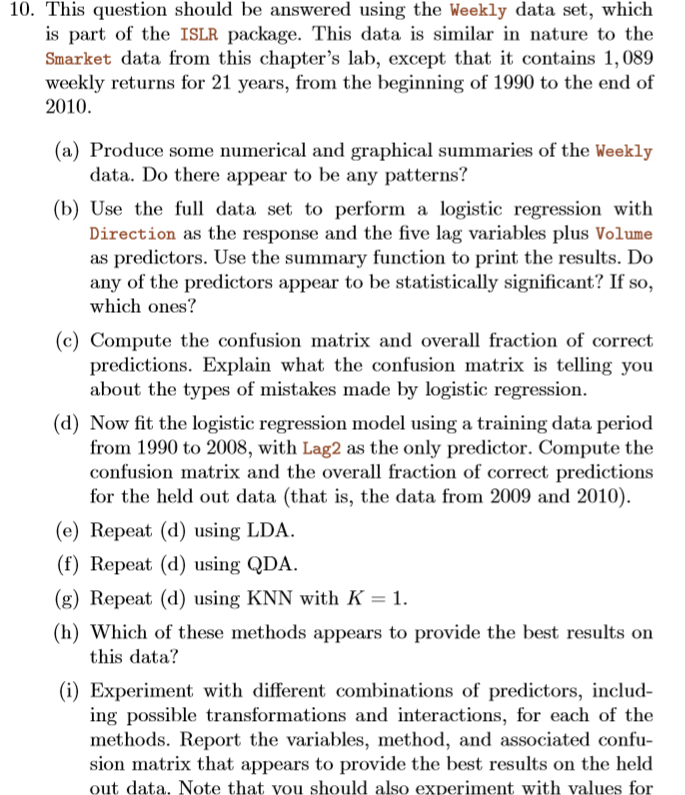
False, QDA在training set 上表現可能比LDA好,但是在如果資料是可線性分割的話,QDA在testing set上表現會比LDA差 。這種狀況稱為overfitting。



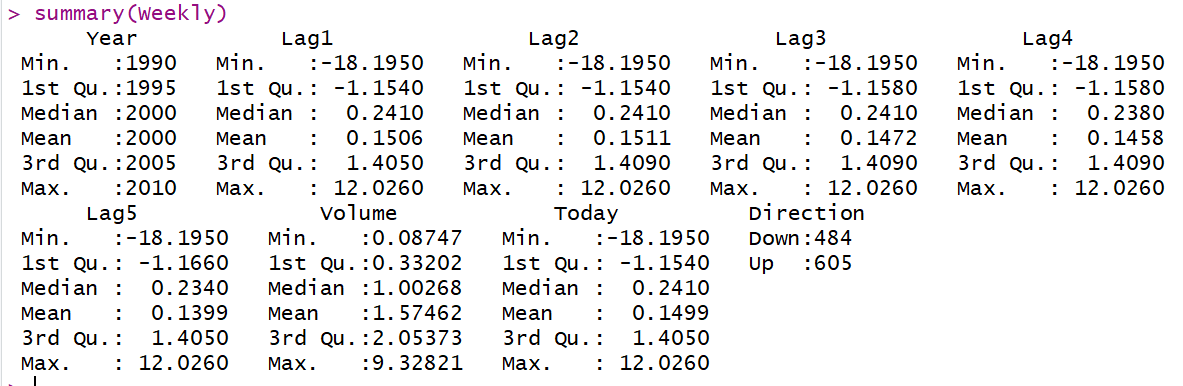
(a)

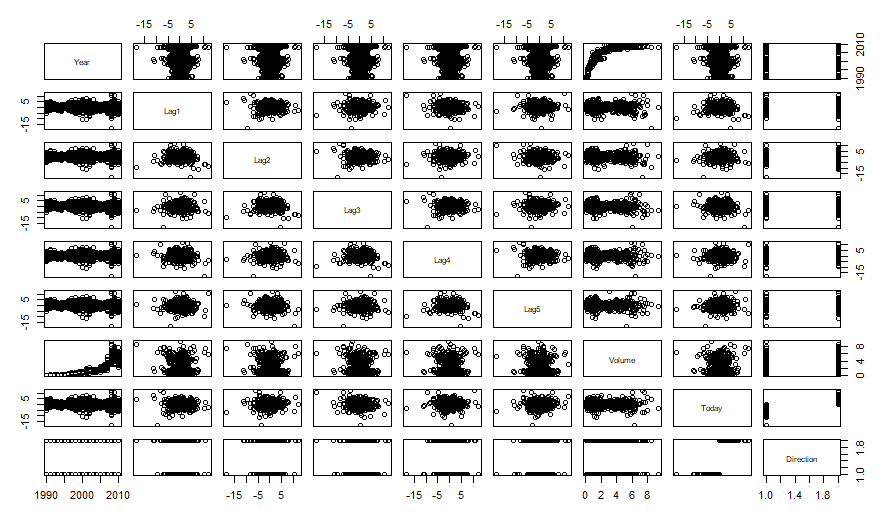
(b)

50hours



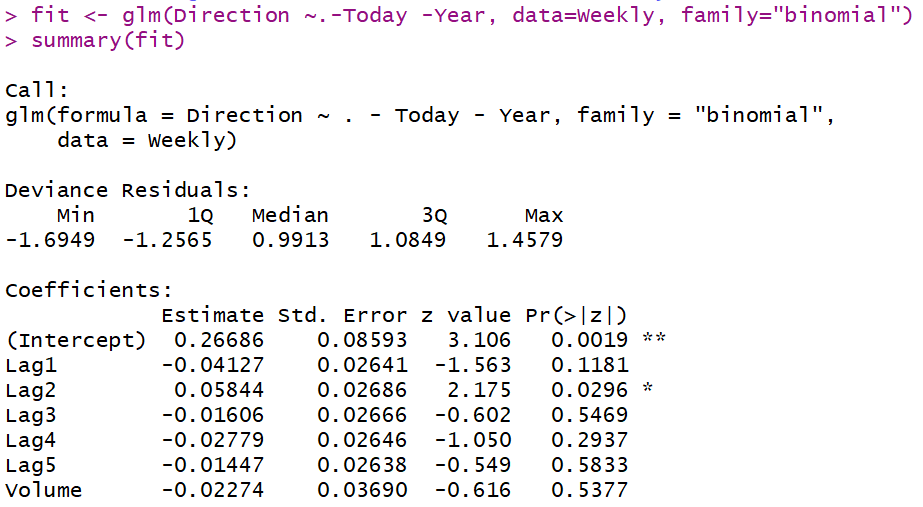
(a)





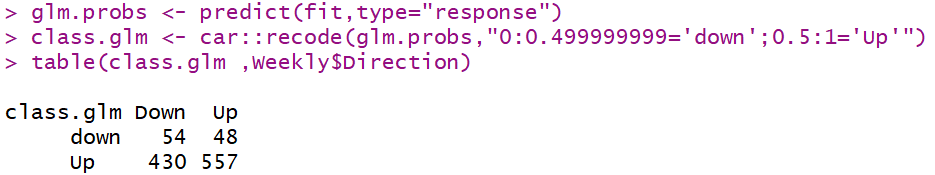
Volume和Year有指數和對數關係。

(b)



Lag2 在alpha=0.05時,此變數有顯著效果

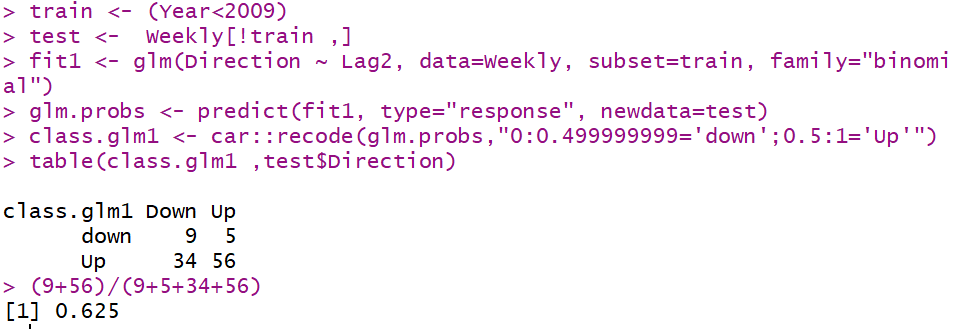
(c)



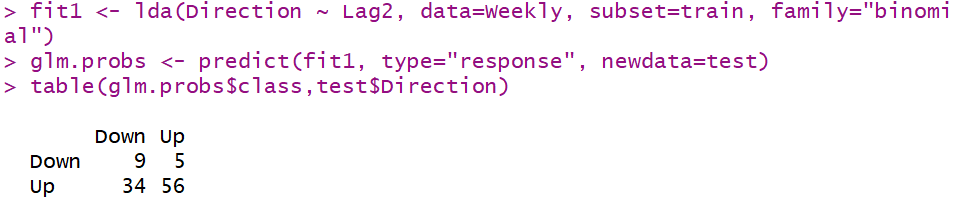
48和430是預測錯誤的次數,48是預測down結果是up,430是預測up結果是down

準確率=54+557/(54+48+430+557)=54.511

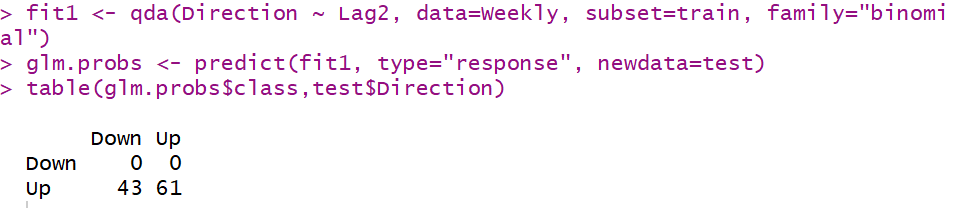
(D)



(e)

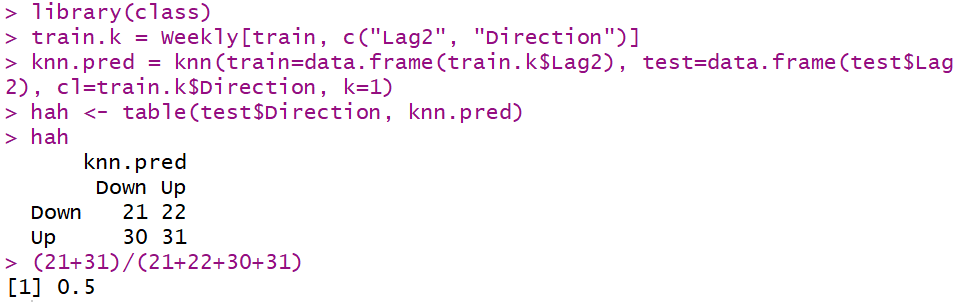


(f)

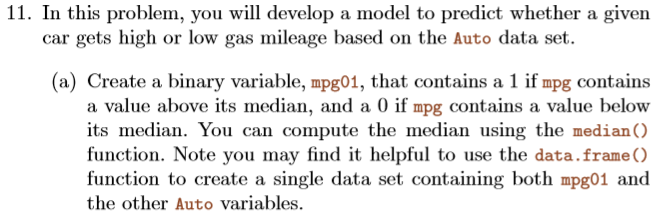


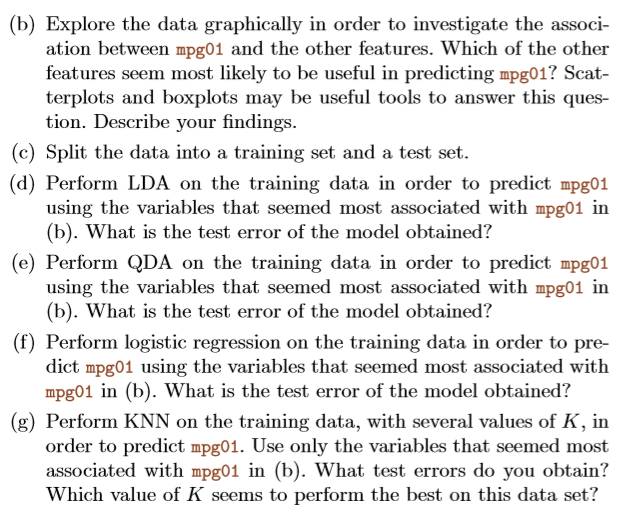
61/(43+61)=0.5865

(g)



(h) logistic 和lda 準確率是比較高的

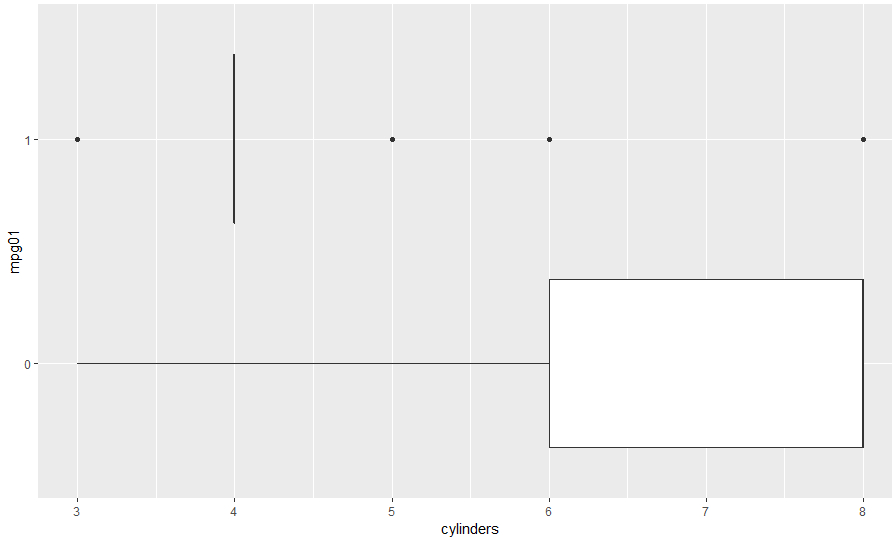


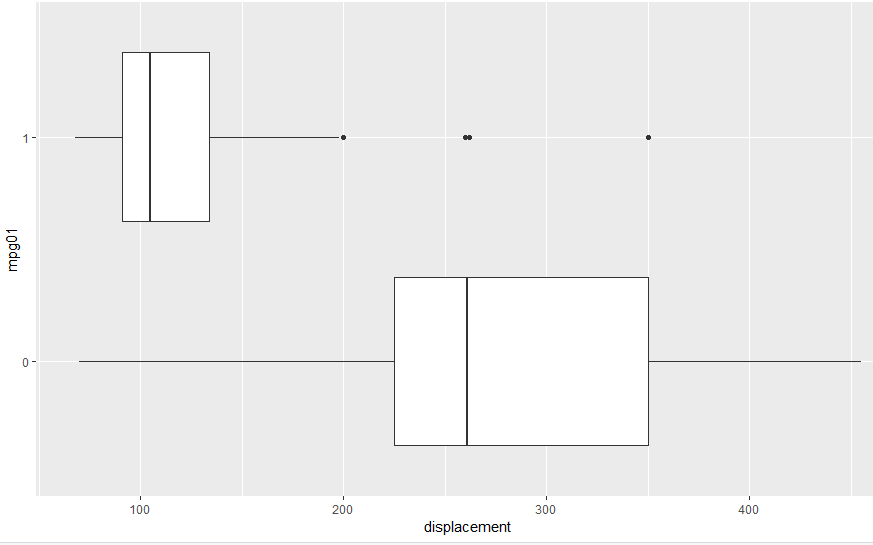


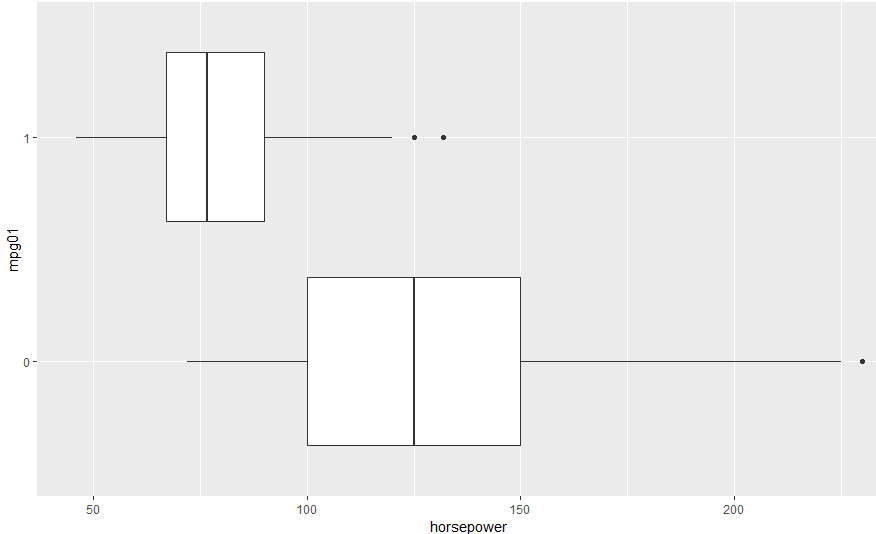
(a)

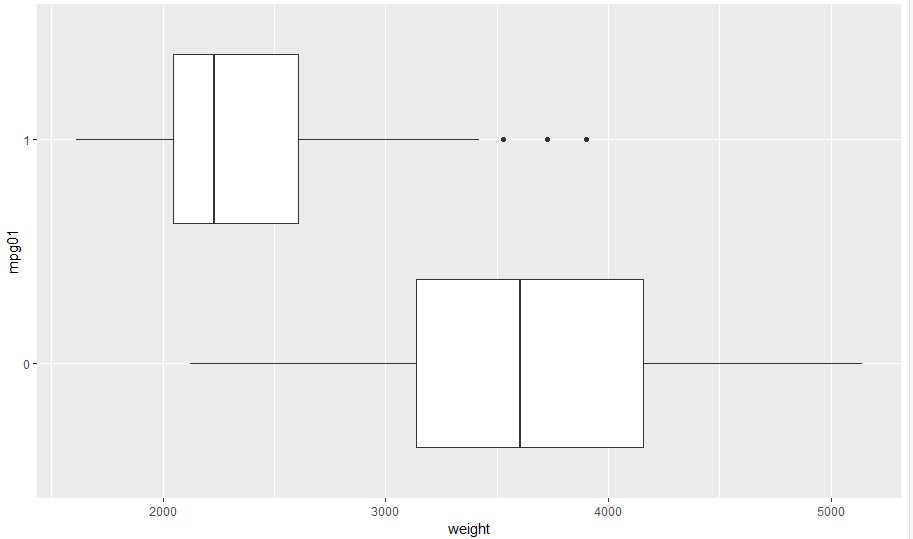


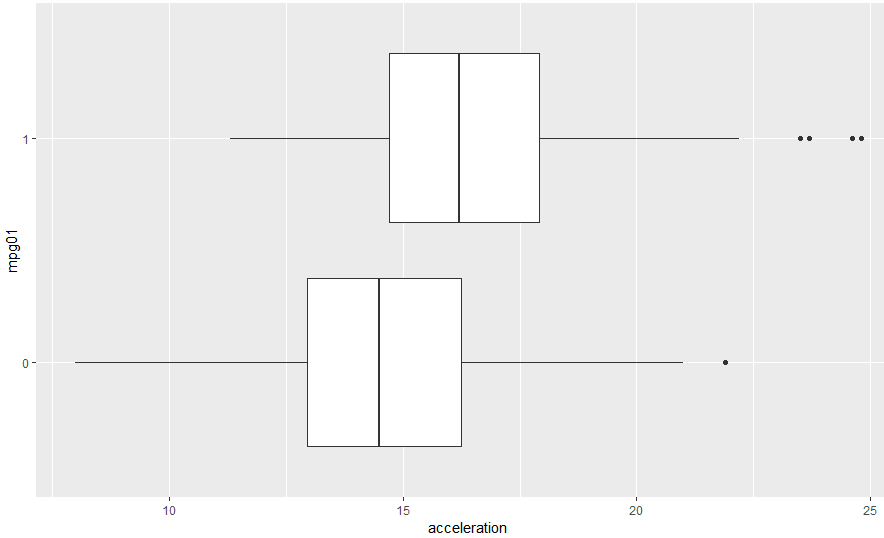
(b)

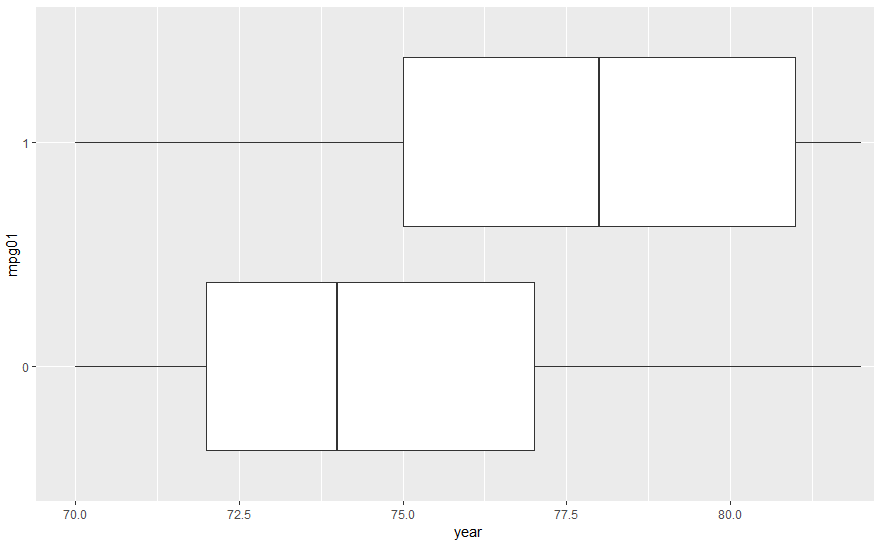




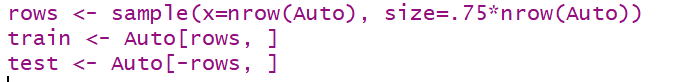




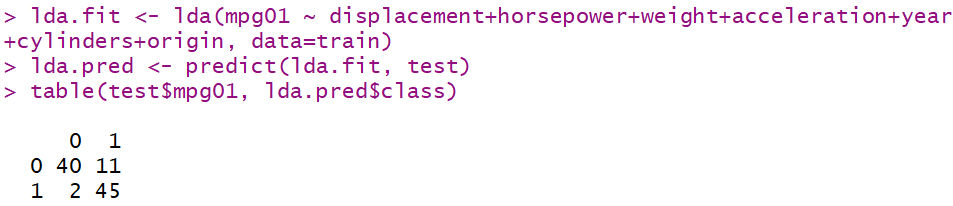




(C)

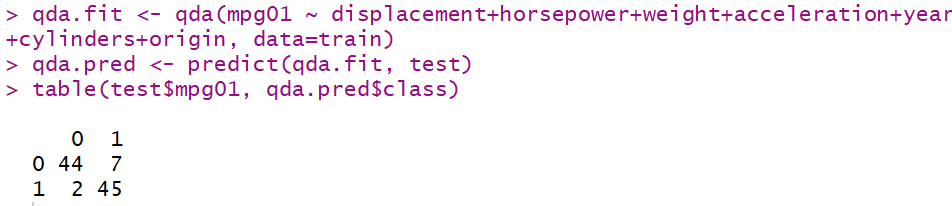


(d)



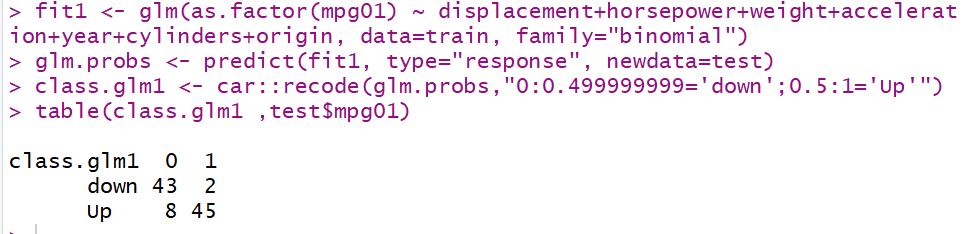
Test error:(11+2)/(40+45+11+2)=0.132

(E)



Test error: 9/(9+44+45)=0.09

(f)



Test error=10/(10+43+45)=0.0102