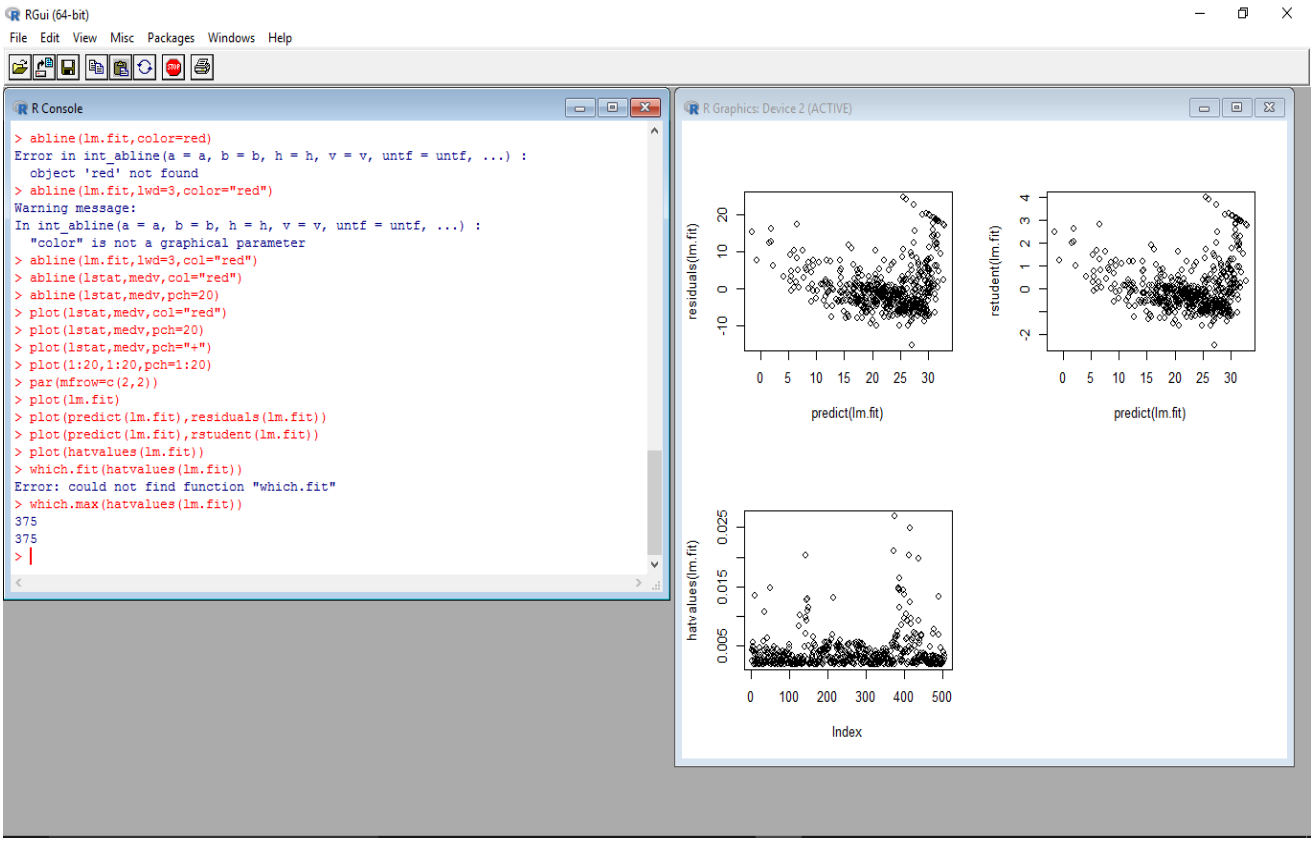
HOMEWORK-2(Big Data)

Lalitha Madhuri Putchala

My Central ID: 700657631

3.6.2 ~ 3.6.6 (5 Captures)

3.6.2 (Capture 1):



3.6.3 (Capture 2):

RGui (64-bit) - [R Console]

File Edit View Misc Packages Windows Help

```
> vif(lm.fit)
Error: could not find function "vif"
> library(car)
> vif(lm.fit)
    crim       zn    indus     chas      nox       rm      age      dis      rad      tax  ptratio    black    lstat
1.792192 2.298758 3.991596 1.073995 4.393720 1.933744 3.100826 3.955945 7.484496 9.008554 1.799084 1.348521 2.941491
> lm.fit1=lm(medv~.,data=Boston)
> summary(lm.fit1)

Call:
lm(formula = medv ~ ., data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-15.595  -2.730  -0.518   1.777  26.199 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.646e+01  5.103e+00   7.144 3.28e-12 ***
crim        -1.080e-01  3.286e-02  -3.287 0.001087 ** 
zn           4.642e-02  1.373e-02   3.382 0.000778 ***
indus        2.056e-02  6.150e-02   0.334 0.738288    
chas         2.687e+00  8.616e-01   3.118 0.001925 ** 
nox         -1.777e+01  3.820e+00  -4.651 4.25e-06 ***
rm           3.810e+00  4.179e-01   9.116  < 2e-16 ***
age          6.922e-04  1.321e-02   0.052 0.958229    
dis         -1.476e+00  1.995e-01  -7.398 6.01e-13 ***
rad          3.060e-01  6.635e-02   4.613 5.07e-06 ***
tax         -1.233e-02  3.760e-03  -3.280 0.001112 ** 
ptratio     -9.527e-01  1.308e-01  -7.283 1.31e-12 ***
black        9.312e-03  2.686e-03   3.467 0.000573 ***
lstat       -5.248e-01  5.072e-02 -10.347  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.745 on 492 degrees of freedom
Multiple R-squared:  0.7406,	Adjusted R-squared:  0.7338 
F-statistic: 108.1 on 13 and 492 DF,  p-value: < 2.2e-16

> lm.fit1=update(lm.fit,~.-age)
>
```

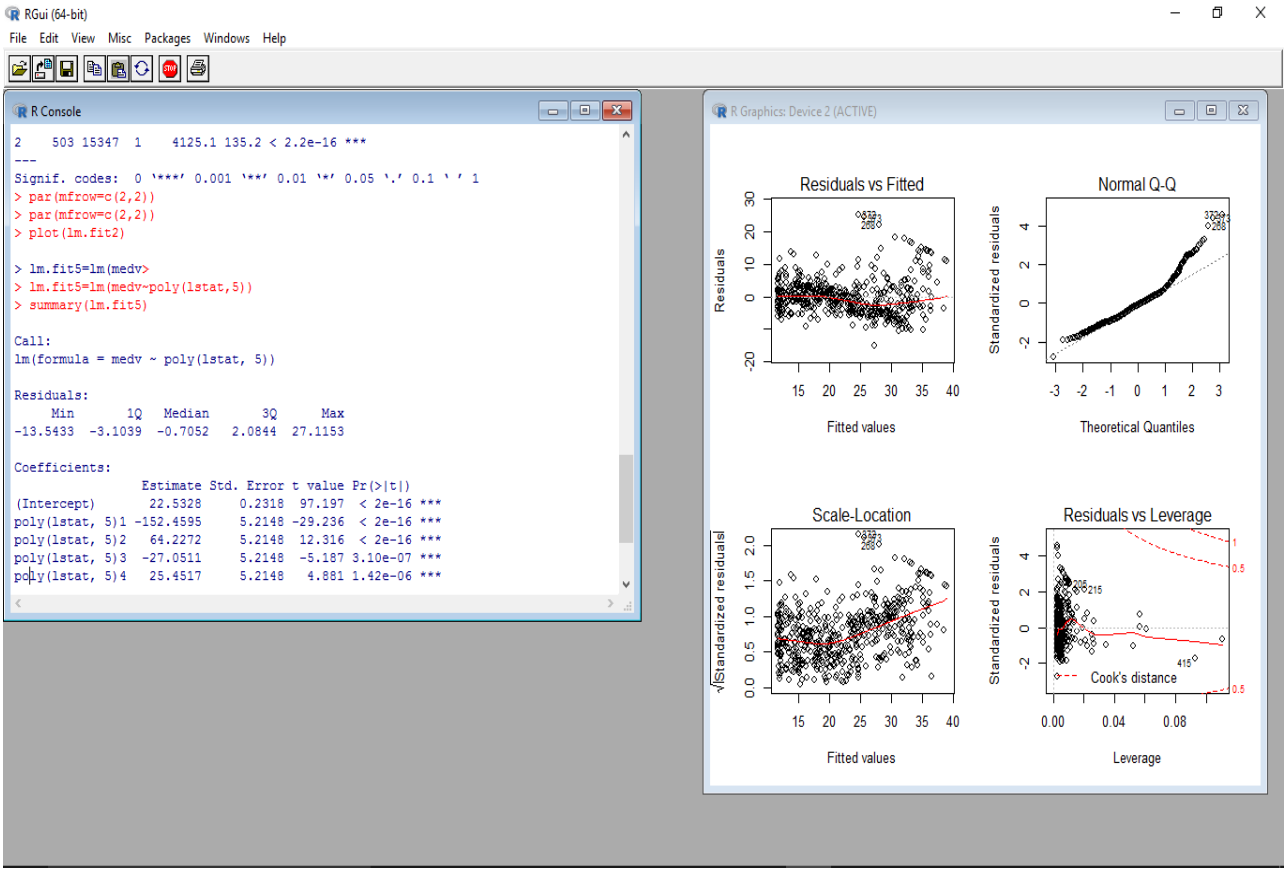
3.6.4 (Capture 3):

RGui (64-bit) - [R Console]

File Edit View Misc Packages Windows Help

```
The downloaded binary packages are in
        C:\Users\lalitha\AppData\Local\Temp\Rtmpeog2BX\downloaded_packages
> utils:::menuInstallPkgs()
trying URL 'https://rweb.crmda.ku.edu/cran/bin/windows/contrib/3.3/ISLR_1.0.zip'
Content type 'application/zip' length 2912971 bytes (2.8 MB)
downloaded 2.8 MB

package 'ISLR' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\lalitha\AppData\Local\Temp\Rtmpeog2BX\downloaded_packages
> library(MASS)
> library(ISLR)
> fix(Boston)
> names(Boston)
 [1] "crim"    "zn"      "indus"   "chas"    "nox"     "rm"      "age"
 [8] "dis"     "rad"     "tax"     "ptratio" "black"   "lstat"   "medv"
> summary(lm(medv~lstat*age,data=Boston))

Call:
lm(formula = medv ~ lstat * age, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max
-15.806  -4.045  -1.333   2.085  27.552

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 36.0885359  1.4698355  24.553  < 2e-16 ***
lstat       -1.3921168  0.1674555  -8.313 8.78e-16 ***
age         -0.0007209  0.0198792  -0.036   0.9711
lstat:age    0.0041560  0.0018518   2.244   0.0252 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.149 on 502 degrees of freedom
Multiple R-squared:  0.5557,    Adjusted R-squared:  0.5531
F-statistic: 209.3 on 3 and 502 DF,  p-value: < 2.2e-16

>
```

3.6.5 (Capture 4):



3.6.6 (Capture 5):

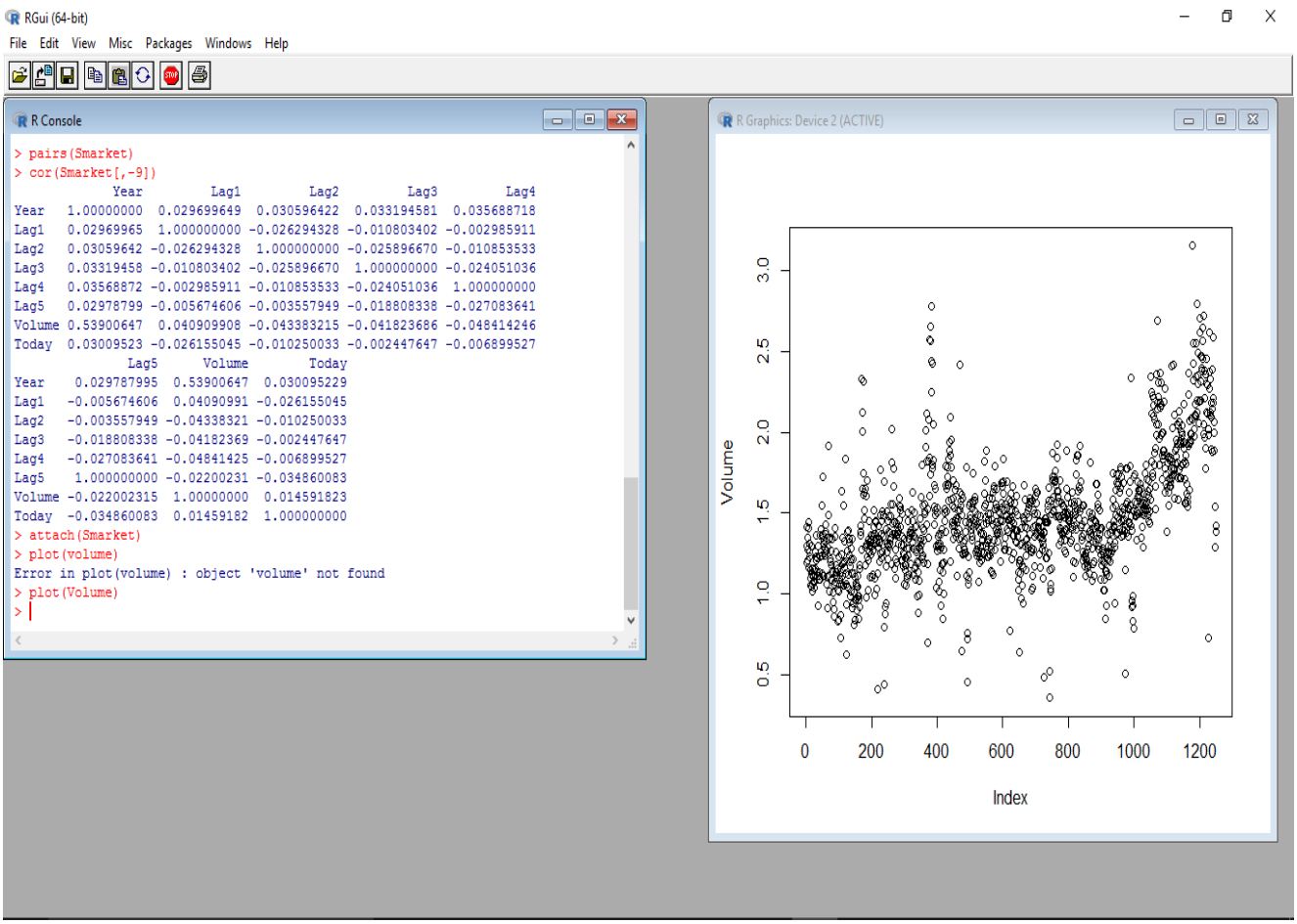
RGui (64-bit) - [R Console]

File Edit View Misc Packages Windows Help

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         6.5755654  1.0087470   6.519 2.22e-10 ***
CompPrice           0.0929371  0.0041183  22.567  < 2e-16 ***
Income              0.0108940  0.0026044   4.183 3.57e-05 ***
Advertising         0.0702462  0.0226091   3.107 0.002030 **
Population          0.0001592  0.0003679   0.433 0.665330
Price              -0.1008064  0.0074399 -13.549  < 2e-16 ***
ShelveLocGood       4.8486762  0.1528378  31.724  < 2e-16 ***
ShelveLocMedium     1.9532620  0.1257682  15.531  < 2e-16 ***
Age                -0.0579466  0.0159506  -3.633 0.000318 ***
Education          -0.0208525  0.0196131  -1.063 0.288361
UrbanYes            0.1401597  0.1124019   1.247 0.213171
USYes              -0.1575571  0.1489234  -1.058 0.290729
Income:Advertising  0.0007510  0.0002784   2.698 0.007290 **
Price:Age           0.0001068  0.0001333   0.801 0.423812
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.011 on 386 degrees of freedom
Multiple R-squared:  0.8761,	Adjusted R-squared:  0.8719
F-statistic:   210 on 13 and 386 DF,  p-value: < 2.2e-16

> attach(Carseats)
> contrasts(ShelveLoc)
       Good Medium
Bad       0      0
Good      1      0
Medium    0      1
> LoadLibraries
Error: object 'LoadLibraries' not found
> LoadLibraries()
Error: could not find function "LoadLibraries"
> LoadLibraries=function(){
+ library(ISLR)
+ library(MASS)
+ print("The libraries have been loaded.")
+ }
> LoadLibraries()
[1] "The libraries have been loaded."
> LoadLibraries()
```

4.6.1 ~ 4.6.4 (4 Captures)

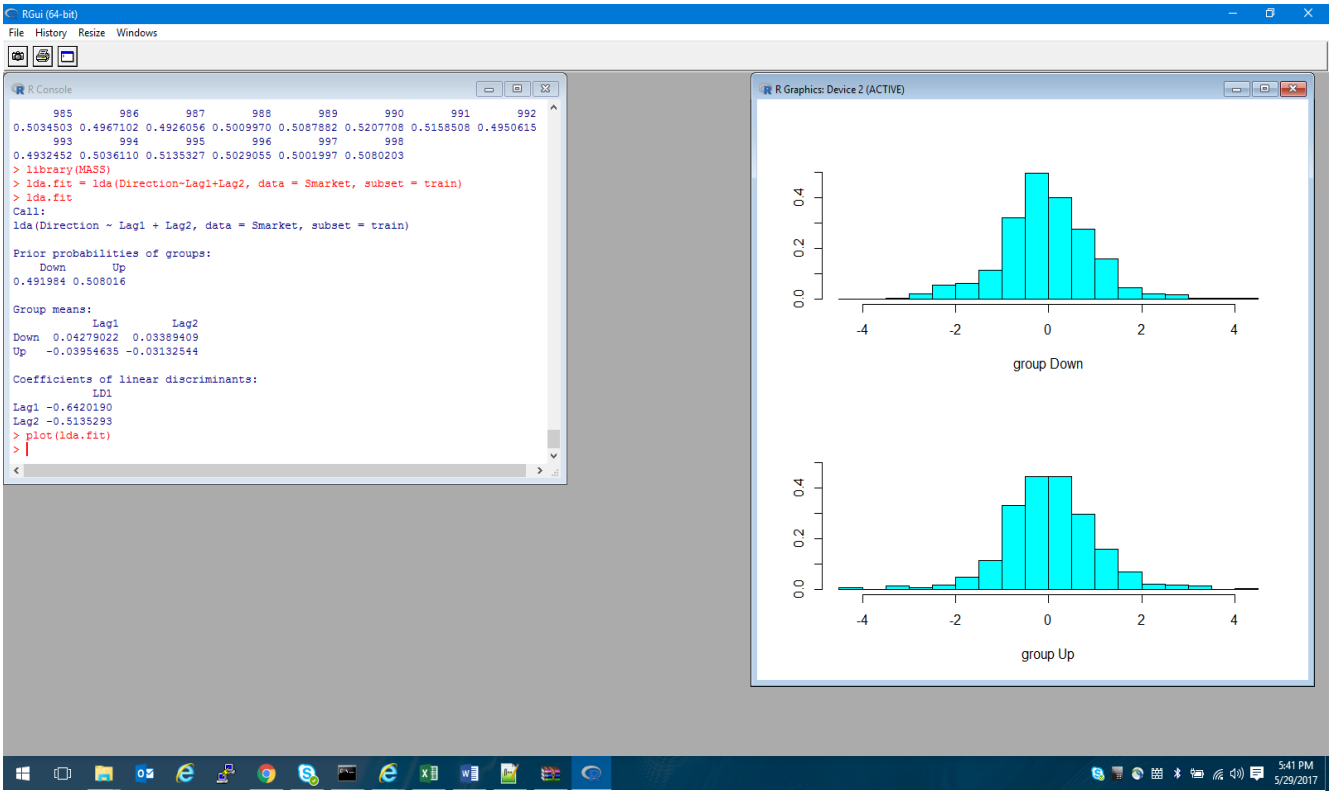
4.6.1 (Capture 1):



4.6.2 (Capture 2):

```
RGui (64-bit) - [R Console]
File  Edit  View  Misc  Packages  Windows  Help

glm.pred Down  Up
    Down  145 141
    Up    457 507
> (507+145)/1250
[1] 0.5216
> mean(glm.pred==Direction)
[1] 0.5216
> train=(Year<2005)
> Smarket.2005=Smarket[!train,]
> dim(Smarket.2005)
[1] 252   9
> Direction.2005=Direction[!train]
> glm.fit=glm(Direction~Lag1+Lag2+Lag3+Lag4+Lag5+Volume,family=binomial,data=Smarket,subset=train)
> glm.probs=predict(glm.fit,Smarket.2005,type="response")
> glm.pred=rep("Down",252)
> glm.pred[glm.probs>.5]="Up"
> table(glm.pred,Direction.2005)
        Direction.2005
glm.pred Down Up
    Down   77 97
    Up     34 44
> mean(glm.pred==Direction.2005)
[1] 0.4801587
> mean(glm.pred!==Direction.2005)
Error: unexpected '=' in "mean(glm.pred!=="
> mean(glm.pred!=Direction.2005)
[1] 0.5198413
> glm.fit=glm(Direction~Lag1+Lag2+Lag3+Lag4+Lag5+Volume,family=binomial,data=Smarket,subset=train)
> glm.probs=predict(glm.fit,Smarket.2005,type="response")
> glm.pred=rep("Down",252)
> glm.pred[glm.probs>.5]="Up"
> table(glm.pred,Direction.2005)
        Direction.2005
glm.pred Down Up
    Down   77 97
    Up     34 44
> mean(glm.pred==Direction.2005)
[1] 0.4801587
> 106/(106+76)
[1] 0.5824176
> predict(glm.fit,newdata=data.frame(Lag1=c(1.2,1.5),Lag2=c(1.1,-0.8)),type="response")
```

4.6.3 (Capture 3):



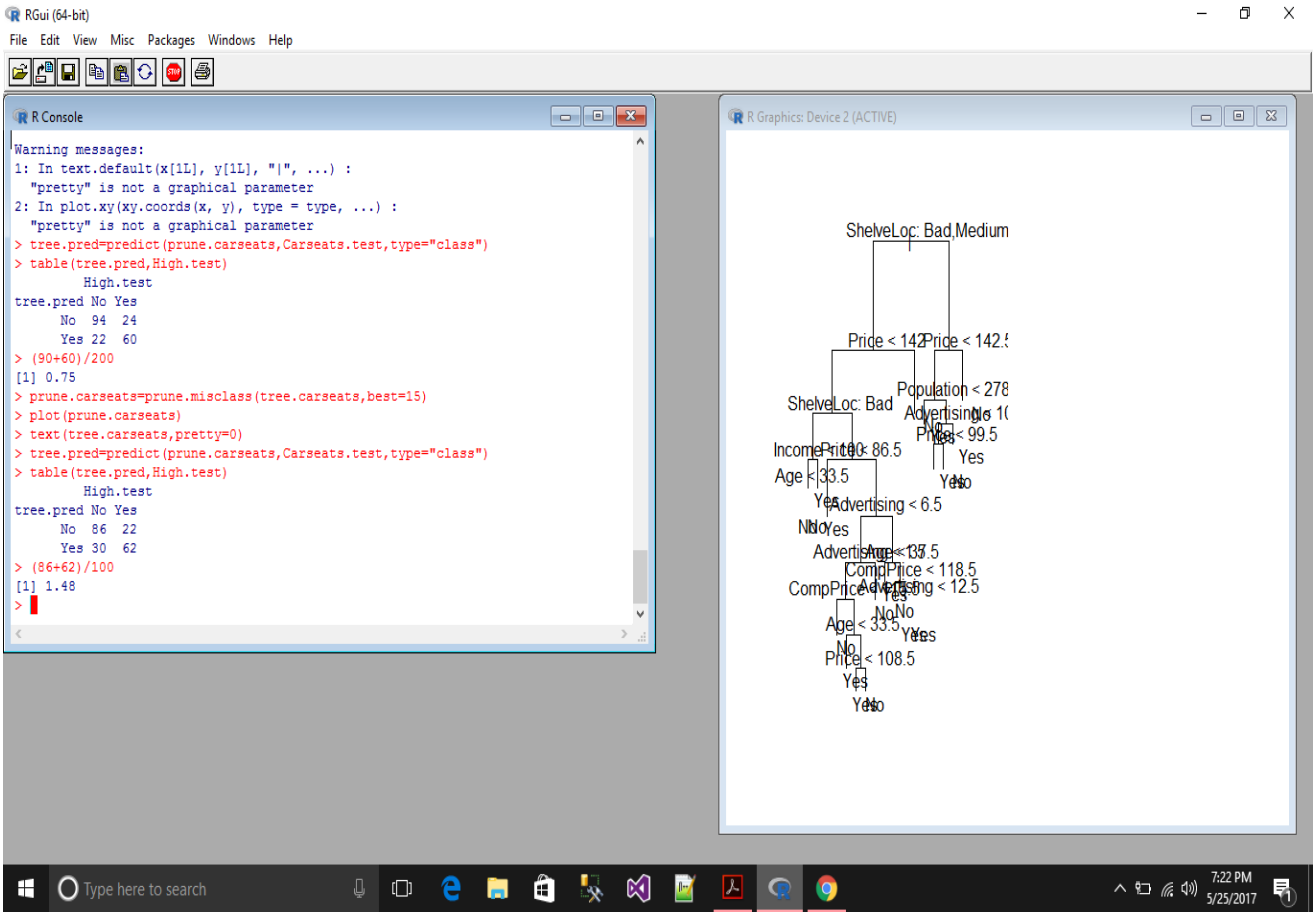

```
> lda.pred = predict(lda.fit, Smarket.2005)
> names(lda.pred)
[1] "class"     "posterior" "x"
> lda.class = lda.pred$class
> table(lda.class, Direction.2005)
         Direction.2005
lda.class Down  Up
     Down   35  35
     Up     76 106
> mean(lda.class == Direction.2005)
[1] 0.5595238
> sum(lda.pred$posterior[,1]>=.5)
[1] 70
> sum(lda.pred$posterior[,1]<.5)
[1] 182
> lda.pred$posterior[1:20,1]
      999      1000      1001      1002      1003      1004      1005      1006
0.4901792 0.4792185 0.4668185 0.4740011 0.4927877 0.4938562 0.4951016 0.4872861
     1007      1008      1009      1010      1011      1012      1013      1014
0.4907013 0.4844026 0.4906963 0.5119988 0.4895152 0.4706761 0.4744593 0.4799583
     1015      1016      1017      1018
0.4935775 0.5030894 0.4978806 0.4886331
> lda.class[1:20]
 [1] Up   Up   Up   Up   Up   Up   Up   Up   Up   Up   Up   Down Up   Up   Up   Up   Up   Down Up   Up
Levels: Down Up
> sum(lda.pred$posterior[,1]>.9)
[1] 0
```

4.6.4 (Capture 4):

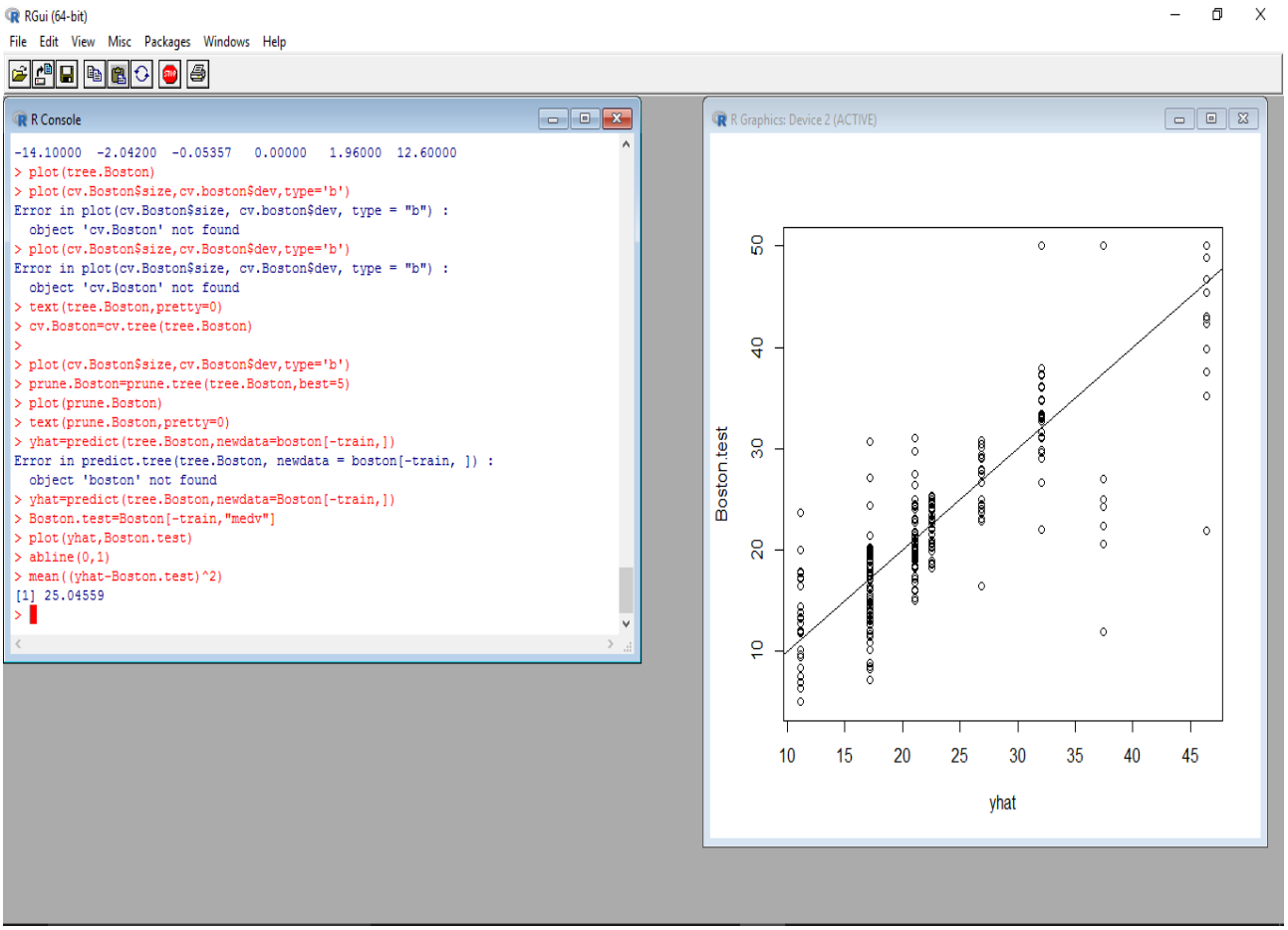
```
> qda.fit = qda(Direction~Lag1+Lag2, data = Smarket, subset = train)
> qda.fit
Call:
qda(Direction ~ Lag1 + Lag2, data = Smarket, subset = train)

Prior probabilities of groups:
    Down       Up
0.491984 0.508016

Group means:
            Lag1        Lag2
Down  0.04279022  0.03389409
Up   -0.03954635 -0.03132544
> qda.class = predict(qda.fit, Smarket.2005)$class
> table(qda.class, Direction.2005)
         Direction.2005
qda.class Down  Up
     Down   30  20
     Up     81 121
> mean(qda.class == Direction.2005)
[1] 0.5992063
```

8.3.1 ~ 8.3.2 (2 Captures)

8.3.1 (Capture 1):

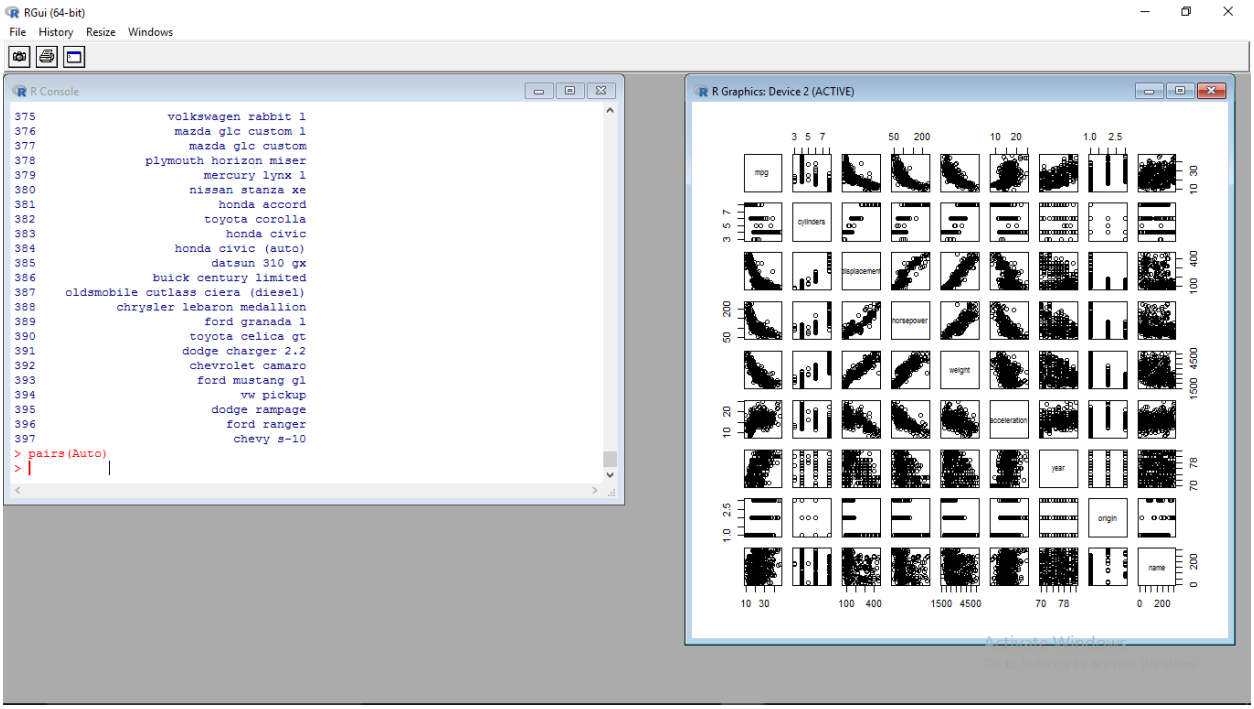


8.3.2 (Capture 2):

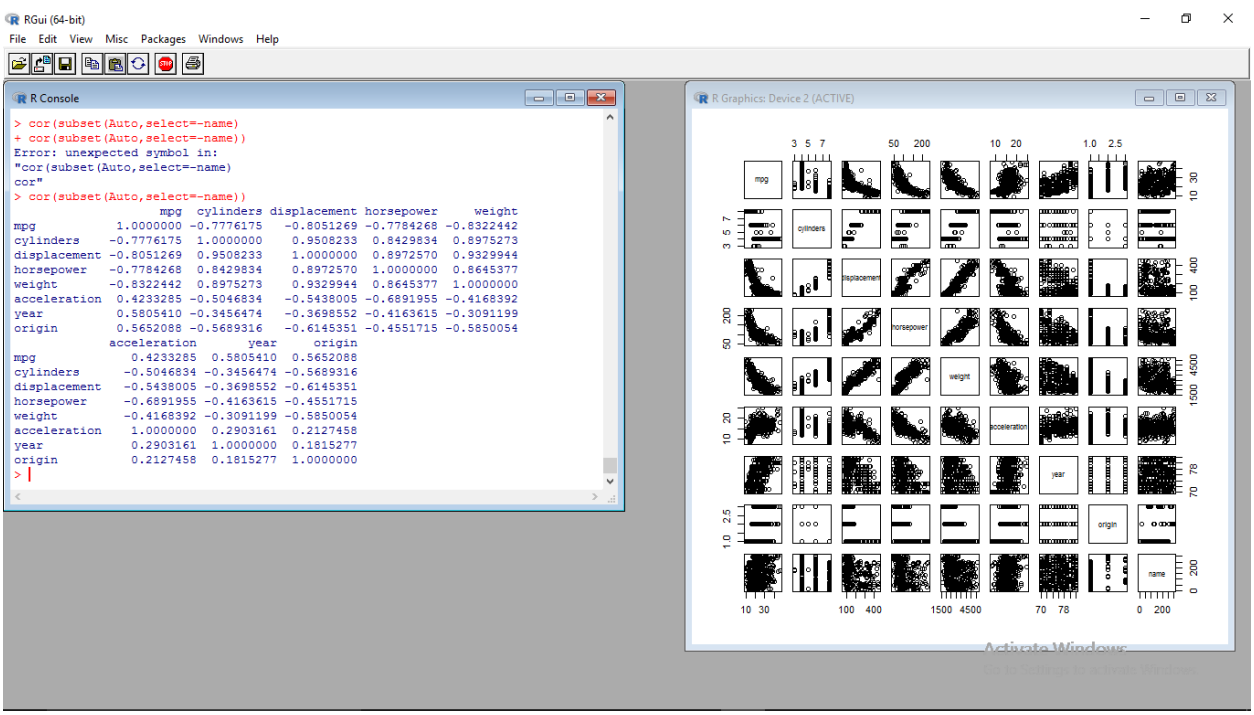


Chapter-3 (Exercise 3.7)

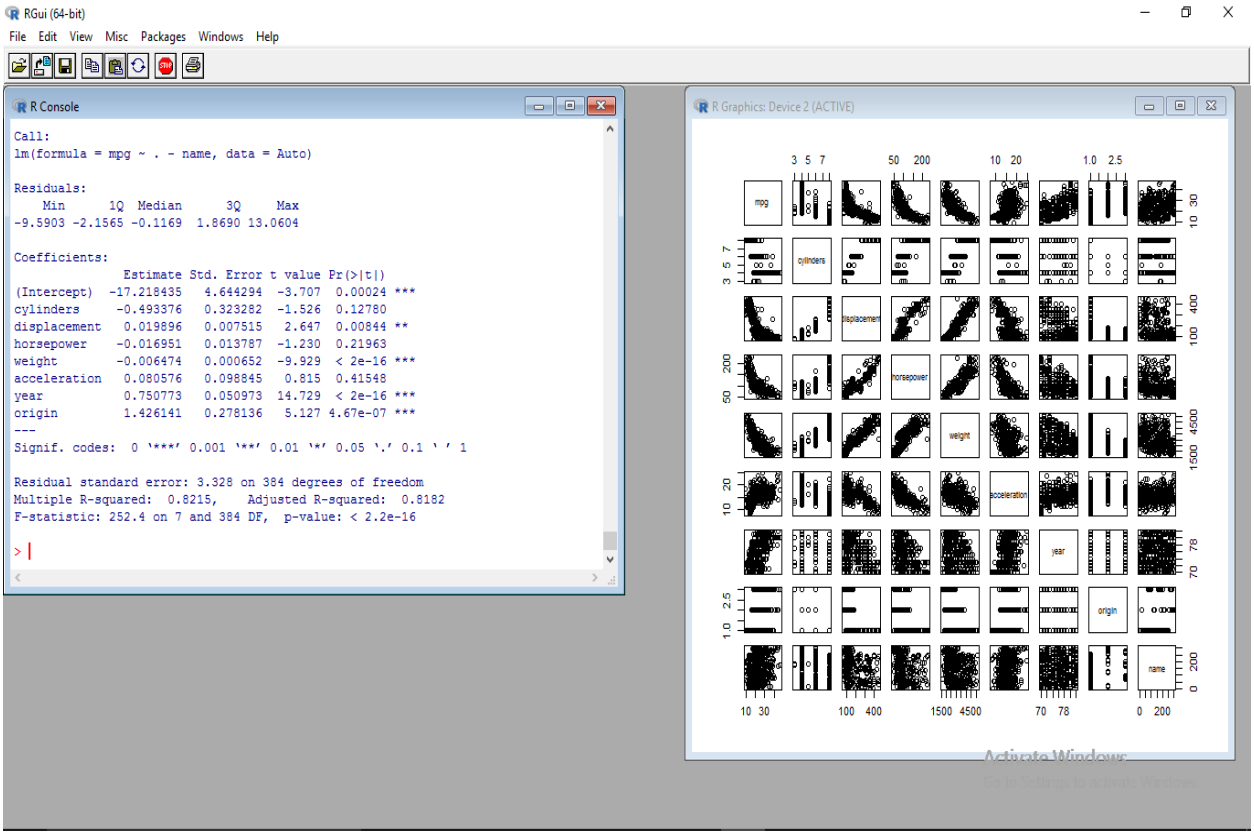
Q9. (a)



Q 9. (b)



Q.9 ©

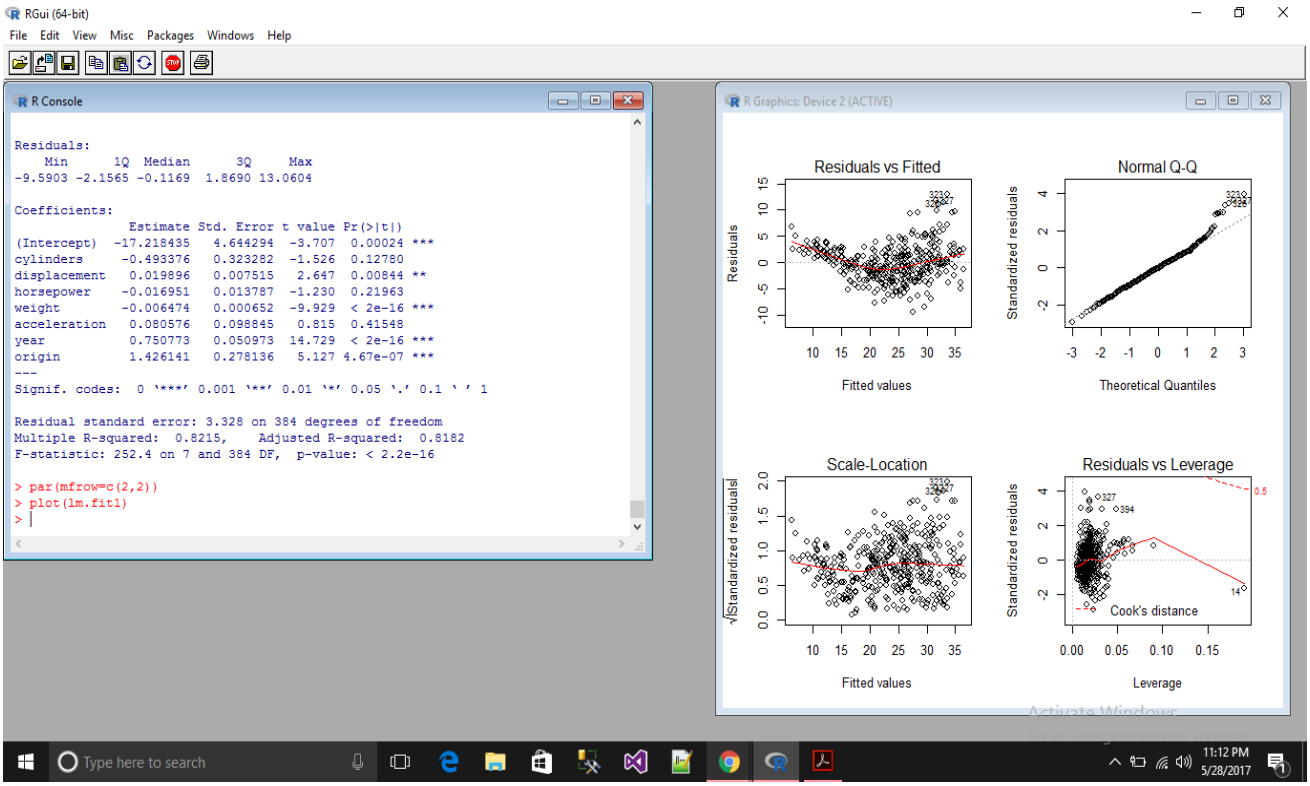


(i) Here all the regression coefficients are zero by testing the null hypothesis. So, we can say that there is a relationship between predictors and response.

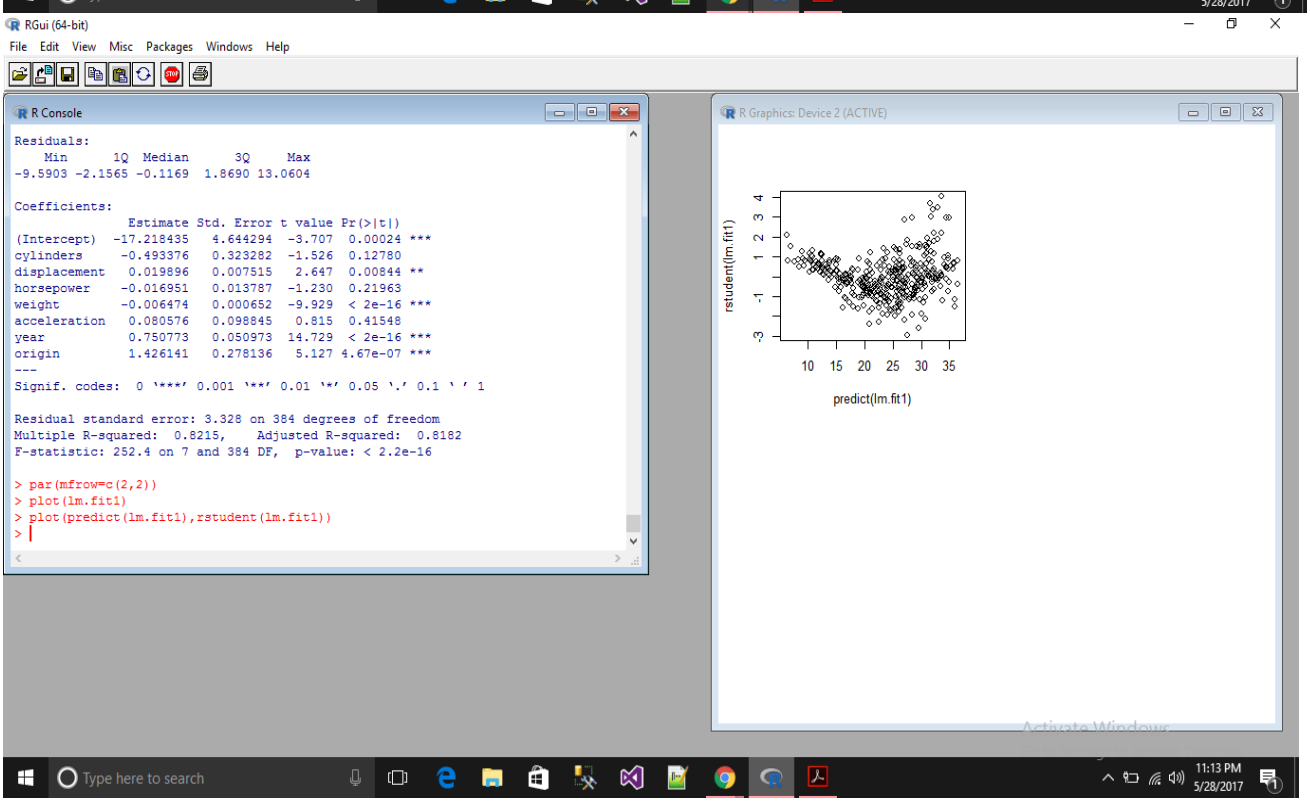
(ii) From the p-values and t-static values, we can see that origin, weight, year and displacement have statistically significant relationship.

(iii) Regression coefficient for the year,0.7508, says that mpg increases by coefficient for every one year.

Q.9(d)


RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console
Residuals:
Min 1Q Median 3Q Max
-9.5903 -2.1565 -0.1169 1.8690 13.0604
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.218435 4.644294 -3.707 0.00024 ***
cylinders -0.493376 0.323282 -1.526 0.12780
displacement 0.019896 0.007515 2.647 0.00844 **
horsepower -0.016951 0.013787 -1.230 0.21963
weight -0.006474 0.000652 -9.929 < 2e-16 ***
acceleration 0.080576 0.098845 0.815 0.41548
year 0.750773 0.050973 14.729 < 2e-16 ***
origin 1.426141 0.278136 5.127 4.67e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 3.328 on 384 degrees of freedom
Multiple R-squared: 0.8215, Adjusted R-squared: 0.8182
F-statistic: 252.4 on 7 and 384 DF, p-value: < 2.2e-16
> par(mfrow=c(2,2))
> plot(lm.fit1)
R Graphics: Device 2 (ACTIVE)
Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
Type here to search
11:12 PM
5/28/2017


RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console
Residuals:
Min 1Q Median 3Q Max
-9.5903 -2.1565 -0.1169 1.8690 13.0604
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.218435 4.644294 -3.707 0.00024 ***
cylinders -0.493376 0.323282 -1.526 0.12780
displacement 0.019896 0.007515 2.647 0.00844 **
horsepower -0.016951 0.013787 -1.230 0.21963
weight -0.006474 0.000652 -9.929 < 2e-16 ***
acceleration 0.080576 0.098845 0.815 0.41548
year 0.750773 0.050973 14.729 < 2e-16 ***
origin 1.426141 0.278136 5.127 4.67e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 3.328 on 384 degrees of freedom
Multiple R-squared: 0.8215, Adjusted R-squared: 0.8182
F-statistic: 252.4 on 7 and 384 DF, p-value: < 2.2e-16
> par(mfrow=c(2,2))
> plot(lm.fit1)
> plot(predict(lm.fit1),rstudent(lm.fit1))
R Graphics: Device 2 (ACTIVE)
rstudent(lm.fit1)
predict(lm.fit1)
Type here to search
11:13 PM
5/28/2017

Q.9(e)

```
RGui (64-bit) - [R Console]
File  Edit  View  Misc  Packages  Windows  Help

386                buick century limited
387    oldsmobile cutlass ciera (diesel)
388           chrysler lebaron medallion
389                       ford granada l
390                     toyota celica gt
391                    dodge charger 2.2
392                    chevrolet camaro
393                     ford mustang gl
394                          vw pickup
395                       dodge rampage
396                         ford ranger
397                          chevy s-10
> lm.fit2 = lm(mpg~cylinders*displacement+displacement*weight)
Error in eval(expr, envir, enclos) : object 'mpg' not found
> lm.fit2 = lm(mpg~cylinders*displacement+displacement*weight,data=Auto)
> summary(lm.fit2)

Call:
lm(formula = mpg ~ cylinders * displacement + displacement *
    weight, data = Auto)

Residuals:
     Min       1Q   Median       3Q      Max
-13.2934  -2.5184  -0.3476   1.8399  17.7723

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          5.262e+01  2.237e+00  23.519  < 2e-16 ***
cylinders            7.606e-01  7.669e-01   0.992    0.322
displacement        -7.351e-02  1.669e-02  -4.403 1.38e-05 ***
weight              -9.888e-03  1.329e-03  -7.438 6.69e-13 ***
cylinders:displacement -2.986e-03  3.426e-03  -0.872    0.384
displacement:weight   2.128e-05  5.002e-06   4.254 2.64e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.103 on 386 degrees of freedom
Multiple R-squared:  0.7272,    Adjusted R-squared:  0.7237
F-statistic: 205.8 on 5 and 386 DF,  p-value: < 2.2e-16

>
```

Q.9(f) (6 captures)

```
RGui (64-bit) - [R Console]
File  Edit  View  Misc  Packages  Windows  Help

(Intercept)          5.262e+01  2.237e+00  23.519  < 2e-16 ***
cylinders            7.606e-01  7.669e-01   0.992    0.322
displacement        -7.351e-02  1.669e-02  -4.403 1.38e-05 ***
weight              -9.888e-03  1.329e-03  -7.438 6.69e-13 ***
cylinders:displacement -2.986e-03  3.426e-03  -0.872    0.384
displacement:weight   2.128e-05  5.002e-06   4.254 2.64e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.103 on 386 degrees of freedom
Multiple R-squared:  0.7272,    Adjusted R-squared:  0.7237
F-statistic: 205.8 on 5 and 386 DF,  p-value: < 2.2e-16

> lm.fit3 = lm(mpg~log(weight)+sqrt(horsepower)+acceleration+I(acceleration^2))
Error in eval(expr, envir, enclos) : object 'mpg' not found
> lm.fit3 = lm(mpg~log(weight)+sqrt(horsepower)+acceleration+I(acceleration^2),data=Auto)
> summary(lm.fit3)

Call:
lm(formula = mpg ~ log(weight) + sqrt(horsepower) + acceleration +
    I(acceleration^2), data = Auto)

Residuals:
     Min       1Q   Median       3Q      Max
-11.2932  -2.5082  -0.2237   2.0237  15.7650

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      178.30303   10.80451  16.503  < 2e-16 ***
log(weight)      -14.74259    1.73994  -8.473 5.06e-16 ***
sqrt(horsepower)  -1.85192    0.36005  -5.144 4.29e-07 ***
acceleration      -2.19890    0.63903  -3.441 0.000643 ***
I(acceleration^2)  0.06139    0.01857   3.305 0.001037 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.99 on 387 degrees of freedom
Multiple R-squared:  0.7414,    Adjusted R-squared:  0.7387
F-statistic: 277.3 on 4 and 387 DF,  p-value: < 2.2e-16

>
```

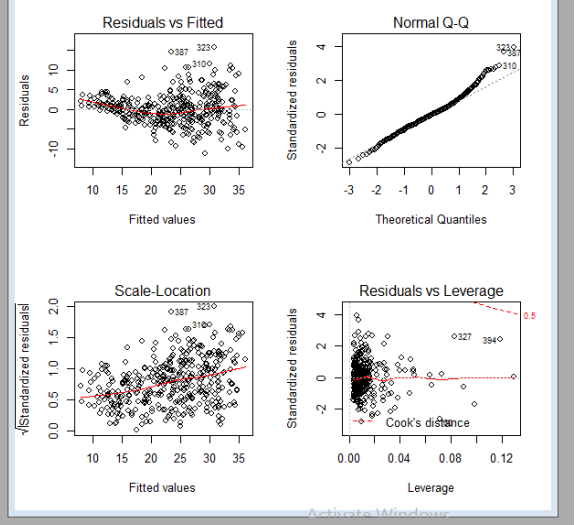
RGui (64-bit)

File Edit View Misc Packages Windows Help

R Console

```
Call:
lm(formula = mpg ~ log(weight) + sqrt(horsepower) + acceleration +
    I(acceleration^2), data = Auto)

Residuals:
     Min       1Q   Median       3Q      Max
-11.2932  -2.5082  -0.2237   2.0237  15.7650

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         178.30303   10.80451  16.503  < 2e-16 ***
log(weight)         -14.74259    1.73994  -8.473 5.06e-16 ***
sqrt(horsepower)     -1.85192    0.36005  -5.144 4.29e-07 ***
acceleration         -2.19890    0.63903  -3.441 0.000643 ***
I(acceleration^2)     0.06139    0.01857   3.305 0.001037 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.99 on 387 degrees of freedom
Multiple R-squared:  0.7414,    Adjusted R-squared:  0.7387
F-statistic: 277.3 on 4 and 387 DF,  p-value: < 2.2e-16

> par(mfrow=c(2,2))
> plot(lm.fit3)
>
```

R Graphics: Device 2 (ACTIVE)



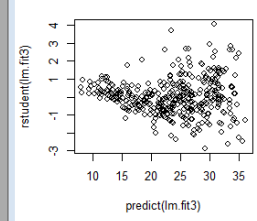
RGui (64-bit)

File Edit View Misc Packages Windows Help

R Console

```
lm(formula = mpg ~ log(weight) + sqrt(horsepower) + acceleration +
    I(acceleration^2), data = Auto)

Residuals:
     Min       1Q   Median       3Q      Max
-11.2932  -2.5082  -0.2237   2.0237  15.7650

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         178.30303   10.80451  16.503  < 2e-16 ***
log(weight)         -14.74259    1.73994  -8.473 5.06e-16 ***
sqrt(horsepower)     -1.85192    0.36005  -5.144 4.29e-07 ***
acceleration         -2.19890    0.63903  -3.441 0.000643 ***
I(acceleration^2)     0.06139    0.01857   3.305 0.001037 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.99 on 387 degrees of freedom
Multiple R-squared:  0.7414,    Adjusted R-squared:  0.7387
F-statistic: 277.3 on 4 and 387 DF,  p-value: < 2.2e-16

> par(mfrow=c(2,2))
> plot(lm.fit3)
> plot(predict(lm.fit3), rstudent(lm.fit3))
>
```

R Graphics: Device 2 (ACTIVE)



Type here to search

11:21 AM
5/29/2017

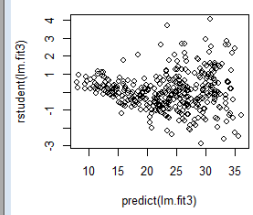
RGui (64-bit)

File Edit View Misc Packages Windows Help

R Console

```
> summary(lm.fit2)

Call:
lm(formula = log(mpg) ~ cylinders + displacement + horsepower +
    weight + acceleration + year + origin, data = Auto)

Residuals:
     Min       1Q   Median       3Q      Max
-0.40955 -0.06533  0.00079  0.06785  0.33925

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.751e+00  1.662e-01  10.533  < 2e-16 ***
cylinders    -2.795e-02  1.157e-02  -2.415  0.01619 *
displacement  6.362e-04  2.690e-04   2.365  0.01852 *
horsepower   -1.475e-03  4.935e-04  -2.989  0.00298 **
weight       -2.551e-04  2.334e-05 -10.931  < 2e-16 ***
acceleration -1.348e-03  3.538e-03  -0.381  0.70339
year          2.958e-02  1.824e-03  16.211  < 2e-16 ***
origin        4.071e-02  9.955e-03   4.089 5.28e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1191 on 384 degrees of freedom
Multiple R-squared:  0.8795,    Adjusted R-squared:  0.8773
```

R Graphics: Device 2 (ACTIVE)



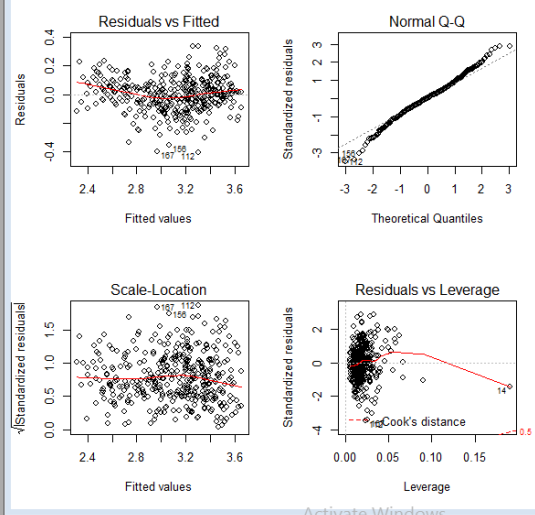
RGui (64-bit)

File Edit View Misc Packages Windows Help

R Console

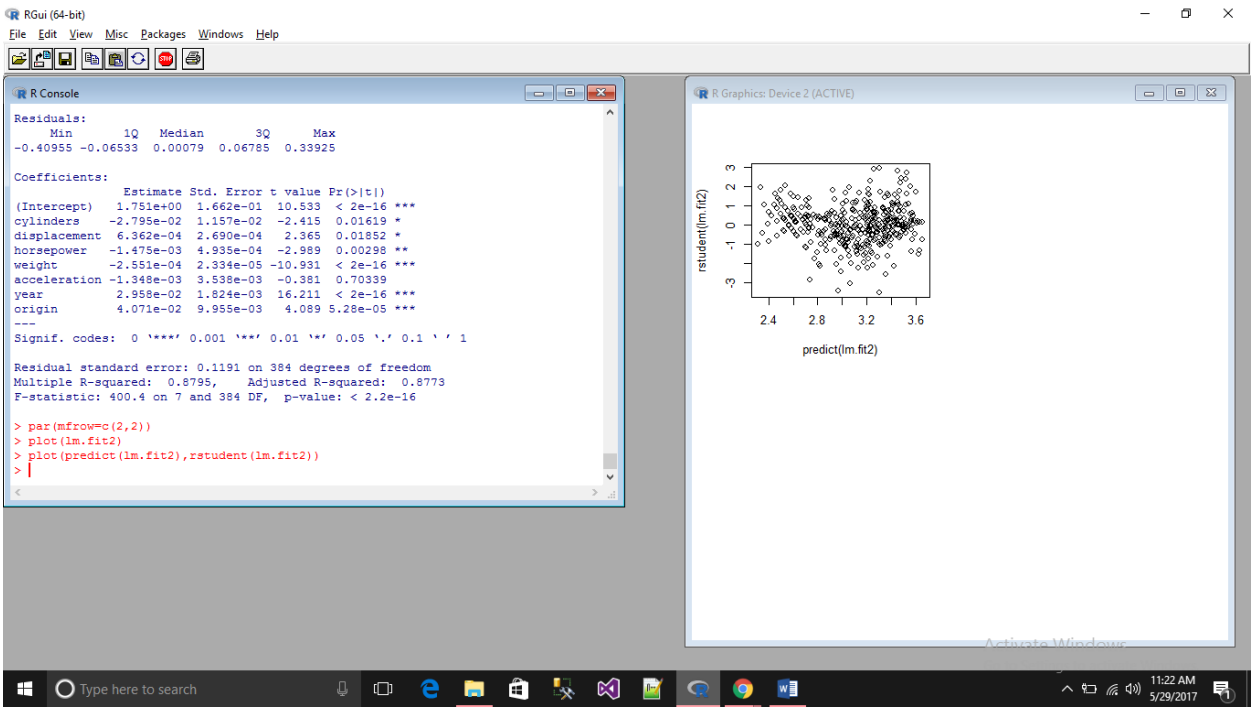
```
Residuals:
     Min       1Q   Median       3Q      Max
-0.40955 -0.06533  0.00079  0.06785  0.33925

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.751e+00  1.662e-01  10.533  < 2e-16 ***
cylinders    -2.795e-02  1.157e-02  -2.415  0.01619 *
displacement  6.362e-04  2.690e-04   2.365  0.01852 *
horsepower   -1.475e-03  4.935e-04  -2.989  0.00298 **
weight       -2.551e-04  2.334e-05 -10.931  < 2e-16 ***
acceleration -1.348e-03  3.538e-03  -0.381  0.70339
year          2.958e-02  1.824e-03  16.211  < 2e-16 ***
origin        4.071e-02  9.955e-03   4.089 5.28e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1191 on 384 degrees of freedom
Multiple R-squared:  0.8795,    Adjusted R-squared:  0.8773
F-statistic: 400.4 on 7 and 384 DF,  p-value: < 2.2e-16

> par(mfrow=c(2,2))
> plot(lm.fit2)
>
```

R Graphics: Device 2 (ACTIVE)



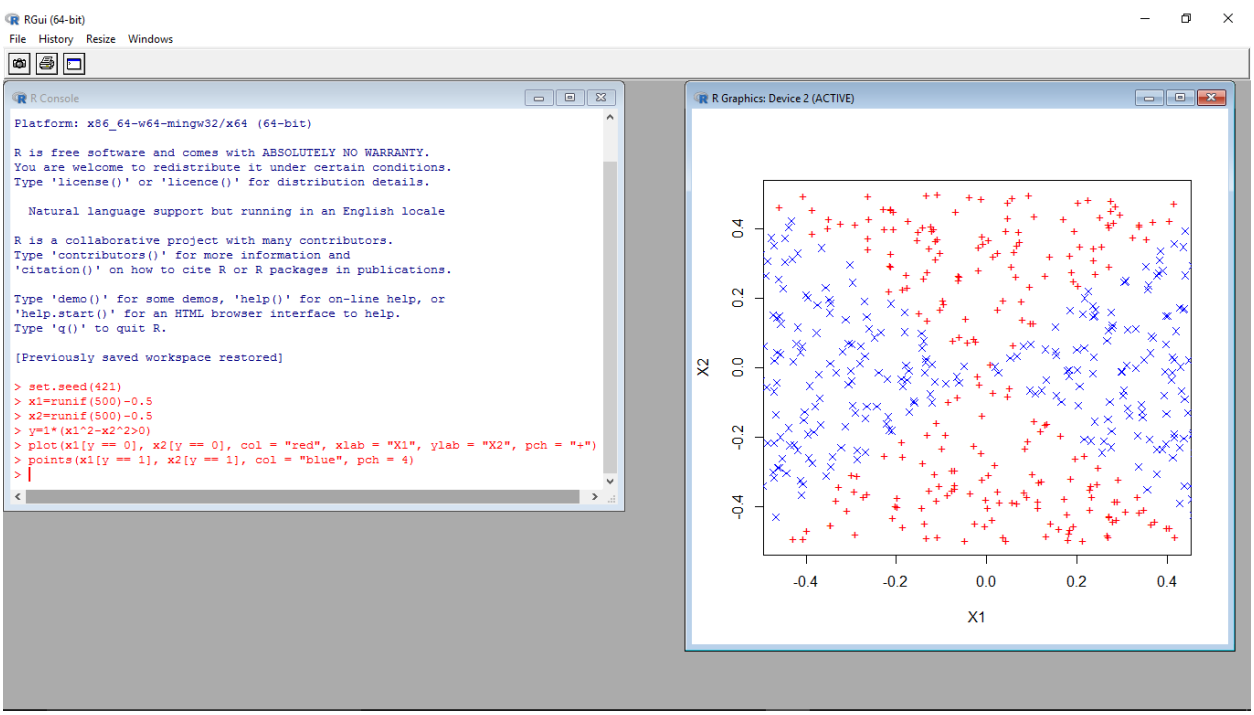
Type here to search

11:22 AM
5/29/2017

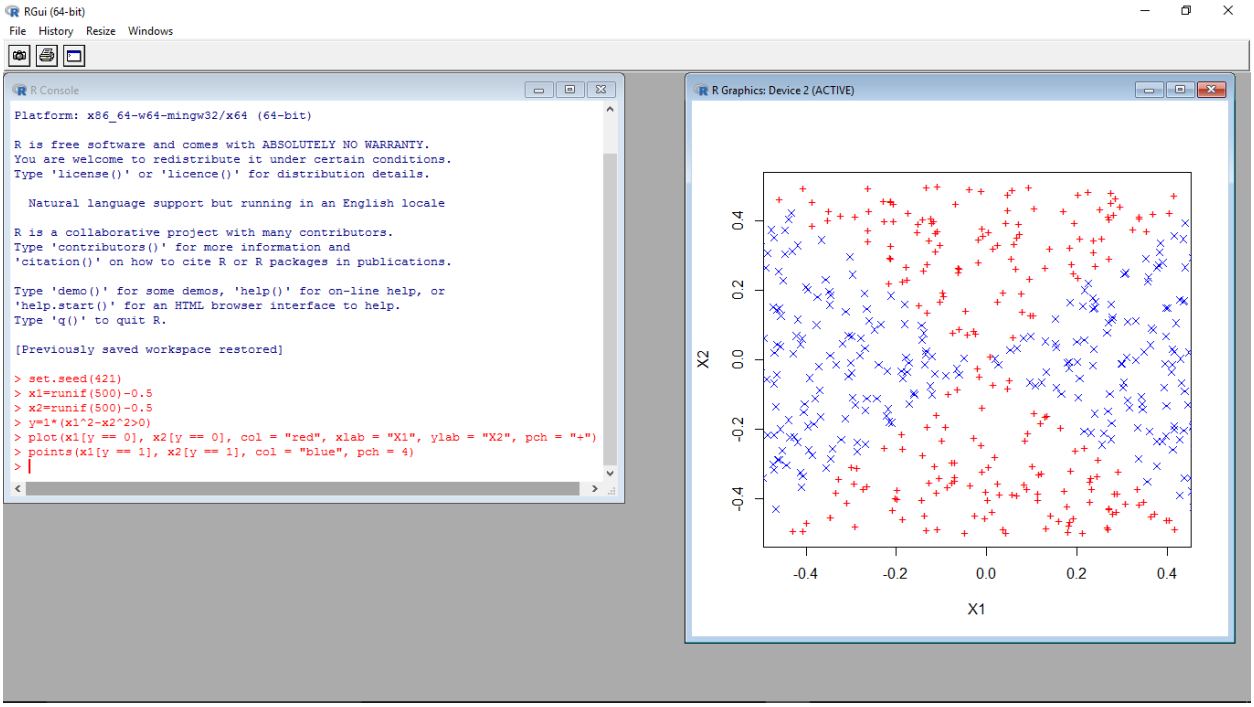


Chapter-9(Exercise 9.7)

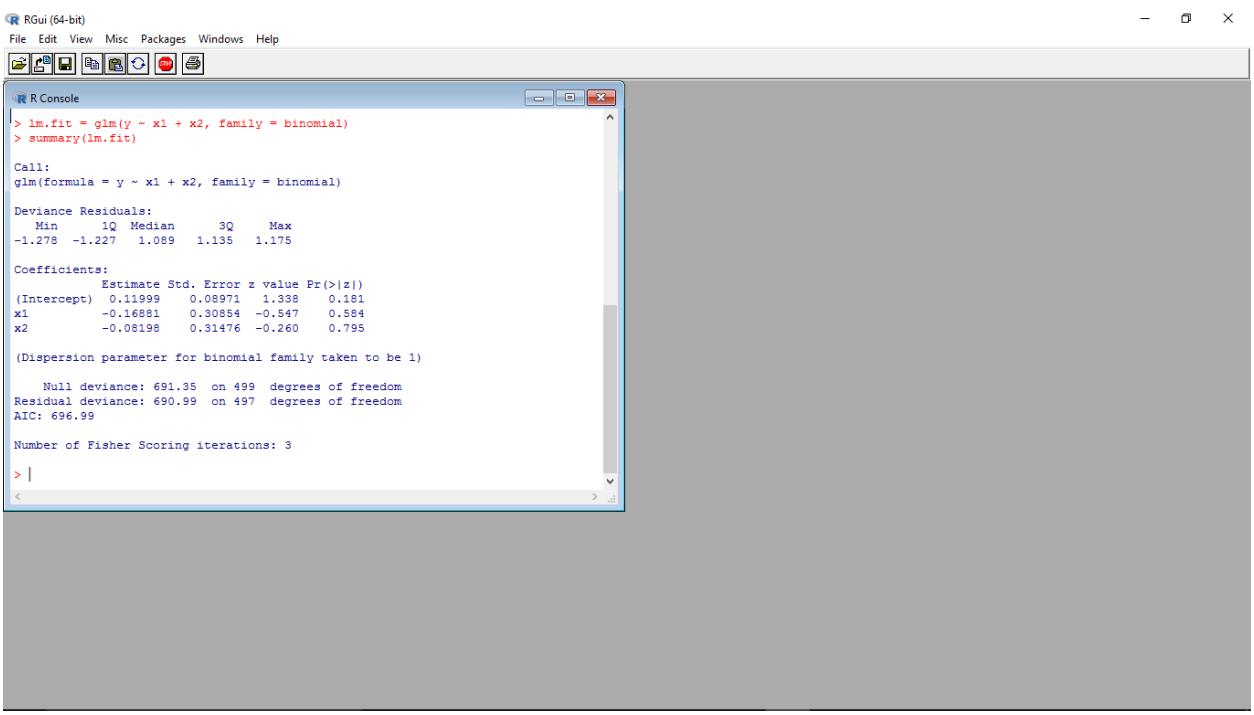
Q.5(a)



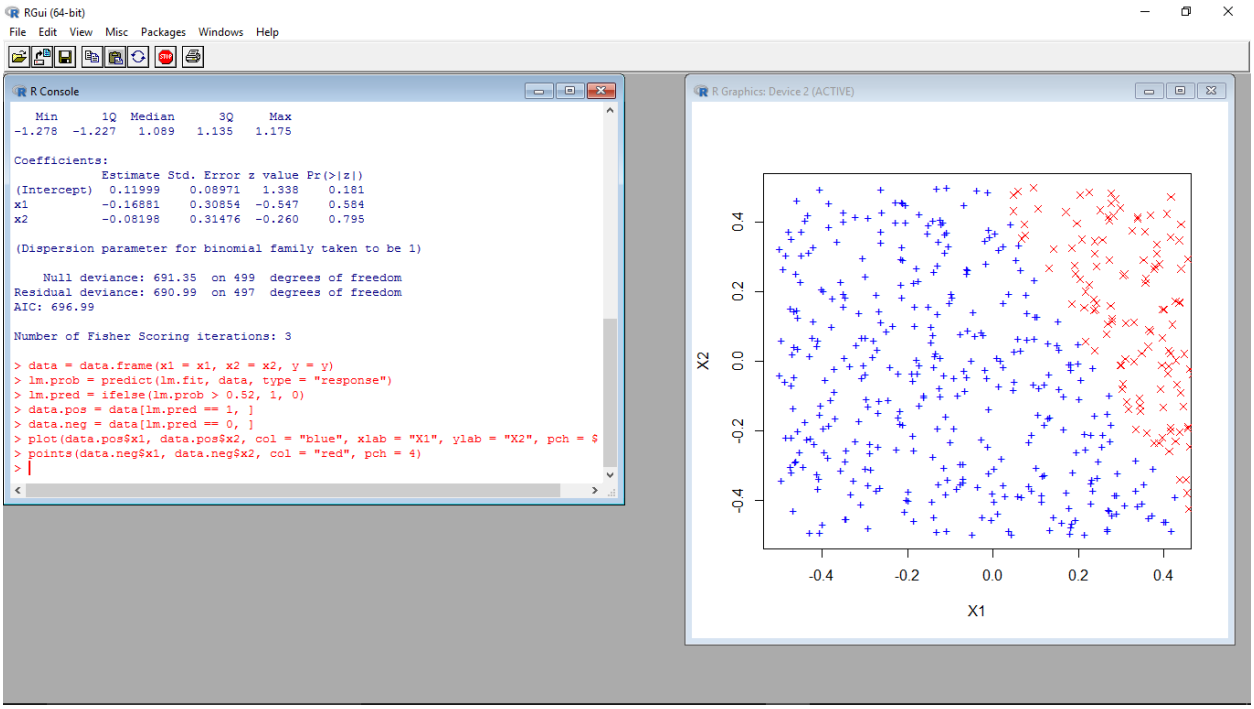
Q.5(b)


RGui (64-bit)
File History Resize Windows
R Console
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Previously saved workspace restored]
> set.seed(421)
> x1=runif(500)-0.5
> x2=runif(500)-0.5
> y=1*(x1^2-x2^2>0)
> plot(x1[y == 0], x2[y == 0], col = "red", xlab = "X1", ylab = "X2", pch = "+")
> points(x1[y == 1], x2[y == 1], col = "blue", pch = 4)
R Graphics: Device 2 (ACTIVE)
X2
X1
-0.4
-0.2
0.0
0.2
0.4

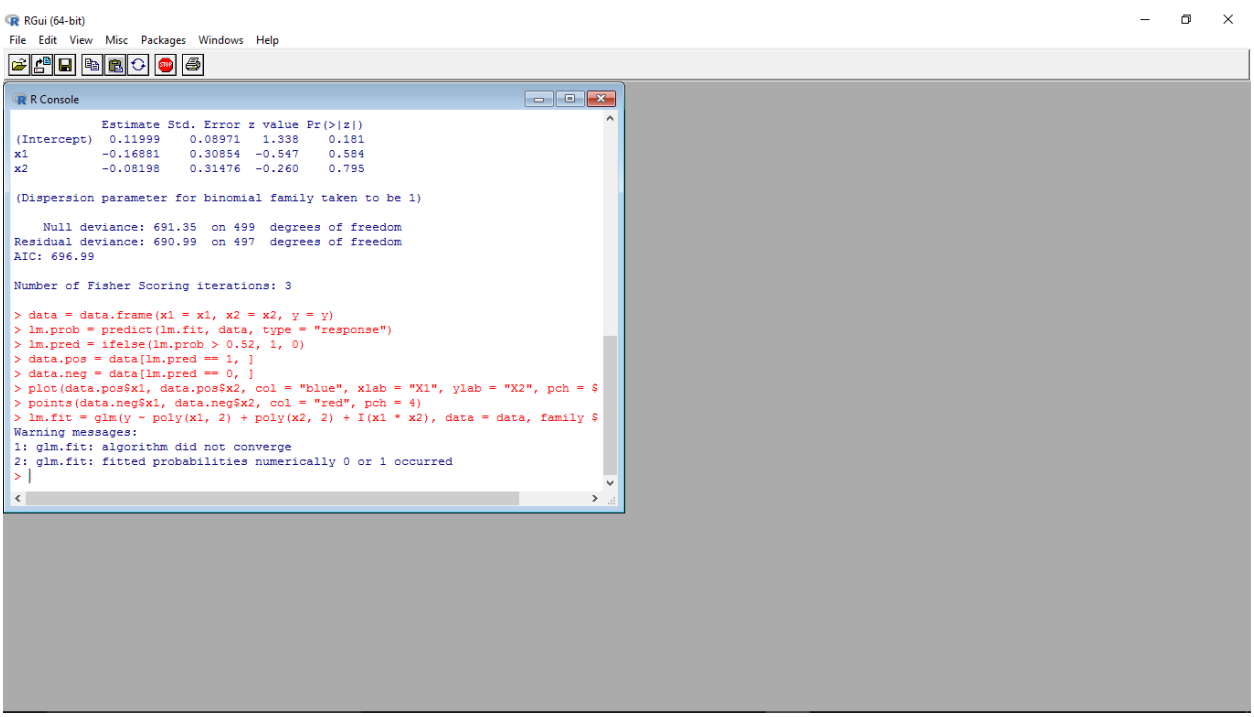
Q.5©


RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console
> lm.fit = glm(y ~ x1 + x2, family = binomial)
> summary(lm.fit)
Call:
glm(formula = y ~ x1 + x2, family = binomial)
Deviance Residuals:
Min 1Q Median 3Q Max
-1.278 -1.227 1.089 1.135 1.175
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.11999 0.08971 1.338 0.181
x1 -0.16881 0.30854 -0.547 0.584
x2 -0.08198 0.31476 -0.260 0.795
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 691.35 on 499 degrees of freedom
Residual deviance: 690.99 on 497 degrees of freedom
AIC: 696.99
Number of Fisher Scoring iterations: 3
>

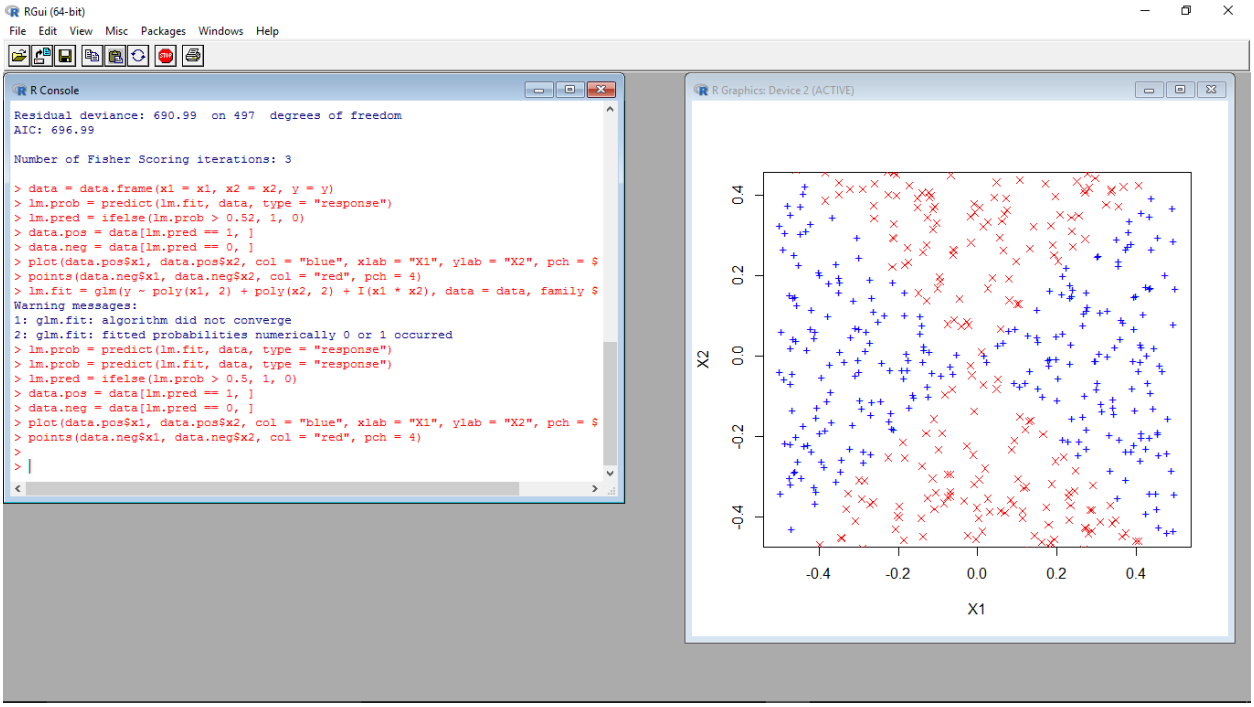
Q.5(d)


RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console
Min 1Q Median 3Q Max
-1.278 -1.227 1.089 1.135 1.175
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.11999 0.08971 1.338 0.181
x1 -0.16881 0.30854 -0.547 0.584
x2 -0.08198 0.31476 -0.260 0.795
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 691.35 on 499 degrees of freedom
Residual deviance: 690.99 on 497 degrees of freedom
AIC: 696.99
Number of Fisher Scoring iterations: 3
> data = data.frame(x1 = x1, x2 = x2, y = y)
> lm.prob = predict(lm.fit, data, type = "response")
> lm.pred = ifelse(lm.prob > 0.52, 1, 0)
> data.pos = data[lm.pred == 1,]
> data.neg = data[lm.pred == 0,]
> plot(data.pos$x1, data.pos$x2, col = "blue", xlab = "X1", ylab = "X2", pch = $
> points(data.neg$x1, data.neg$x2, col = "red", pch = 4)
R Graphics: Device 2 (ACTIVE)
X2
X1
-0.4
-0.2
0.0
0.2
0.4

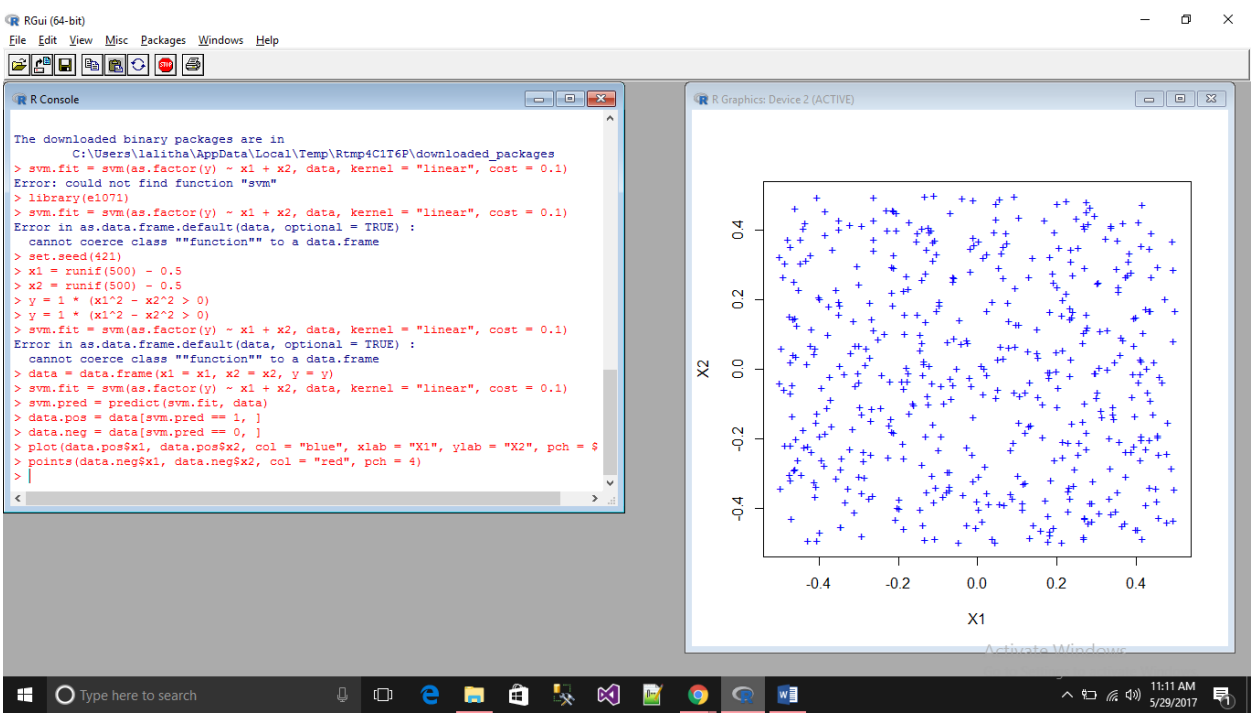
Q.5(e)


RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console
Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.11999 0.08971 1.338 0.181
x1 -0.16881 0.30854 -0.547 0.584
x2 -0.08198 0.31476 -0.260 0.795
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 691.35 on 499 degrees of freedom
Residual deviance: 690.99 on 497 degrees of freedom
AIC: 696.99
Number of Fisher Scoring iterations: 3
> data = data.frame(x1 = x1, x2 = x2, y = y)
> lm.prob = predict(lm.fit, data, type = "response")
> lm.pred = ifelse(lm.prob > 0.52, 1, 0)
> data.pos = data[lm.pred == 1,]
> data.neg = data[lm.pred == 0,]
> plot(data.pos$x1, data.pos$x2, col = "blue", xlab = "X1", ylab = "X2", pch = $
> points(data.neg$x1, data.neg$x2, col = "red", pch = 4)
> lm.fit = glm(y ~ poly(x1, 2) + poly(x2, 2) + I(x1 * x2), data = data, family $
Warning messages:
1: glm.fit: algorithm did not converge
2: glm.fit: fitted probabilities numerically 0 or 1 occurred

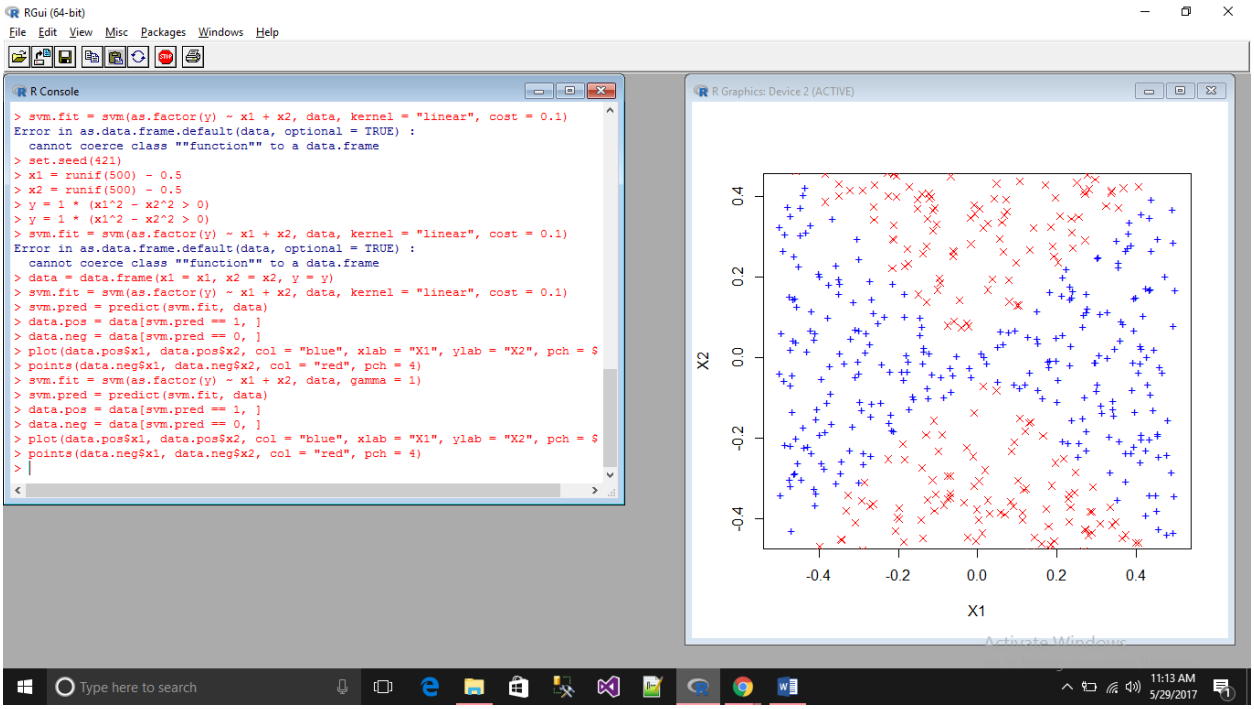
Q.5(f)


RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console
Residual deviance: 690.99 on 497 degrees of freedom
AIC: 696.99
Number of Fisher Scoring iterations: 3
> data = data.frame(x1 = x1, x2 = x2, y = y)
> lm.prob = predict(lm.fit, data, type = "response")
> lm.pred = ifelse(lm.prob > 0.52, 1, 0)
> data.pos = data[lm.pred == 1,]
> data.neg = data[lm.pred == 0,]
> plot(data.pos$x1, data.pos$x2, col = "blue", xlab = "X1", ylab = "X2", pch = $
> points(data.neg$x1, data.neg$x2, col = "red", pch = 4)
> lm.fit = glm(y ~ poly(x1, 2) + poly(x2, 2) + I(x1 * x2), data = data, family $
Warning messages:
1: glm.fit: algorithm did not converge
2: glm.fit: fitted probabilities numerically 0 or 1 occurred
> lm.prob = predict(lm.fit, data, type = "response")
> lm.prob = predict(lm.fit, data, type = "response")
> lm.pred = ifelse(lm.prob > 0.5, 1, 0)
> data.pos = data[lm.pred == 1,]
> data.neg = data[lm.pred == 0,]
> plot(data.pos$x1, data.pos$x2, col = "blue", xlab = "X1", ylab = "X2", pch = $
> points(data.neg$x1, data.neg$x2, col = "red", pch = 4)
R Graphics: Device 2 (ACTIVE)
X2
X1
-0.4
-0.2
0.0
0.2
0.4

Q.5(g)


RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console
The downloaded binary packages are in
C:\Users\lalitha\AppData\Local\Temp\Rtmp4C1T6P\downloaded_packages
> svm.fit = svm(as.factor(y) ~ x1 + x2, data, kernel = "linear", cost = 0.1)
Error: could not find function "svm"
> library(e1071)
> svm.fit = svm(as.factor(y) ~ x1 + x2, data, kernel = "linear", cost = 0.1)
Error in as.data.frame.default(data, optional = TRUE) :
cannot coerce class ""function"" to a data.frame
> set.seed(421)
> x1 = runif(500) - 0.5
> x2 = runif(500) - 0.5
> y = 1 * (x1^2 - x2^2 > 0)
> y = 1 * (x1^2 - x2^2 > 0)
> svm.fit = svm(as.factor(y) ~ x1 + x2, data, kernel = "linear", cost = 0.1)
Error in as.data.frame.default(data, optional = TRUE) :
cannot coerce class ""function"" to a data.frame
> data = data.frame(x1 = x1, x2 = x2, y = y)
> svm.fit = svm(as.factor(y) ~ x1 + x2, data, kernel = "linear", cost = 0.1)
> svm.pred = predict(svm.fit, data)
> data.pos = data[svm.pred == 1,]
> data.neg = data[svm.pred == 0,]
> plot(data.pos$x1, data.pos$x2, col = "blue", xlab = "X1", ylab = "X2", pch = $
> points(data.neg$x1, data.neg$x2, col = "red", pch = 4)
R Graphics: Device 2 (ACTIVE)
X2
X1
-0.4
-0.2
0.0
0.2
0.4
Type here to search
11:11 AM
5/29/2017

Q.5(h)



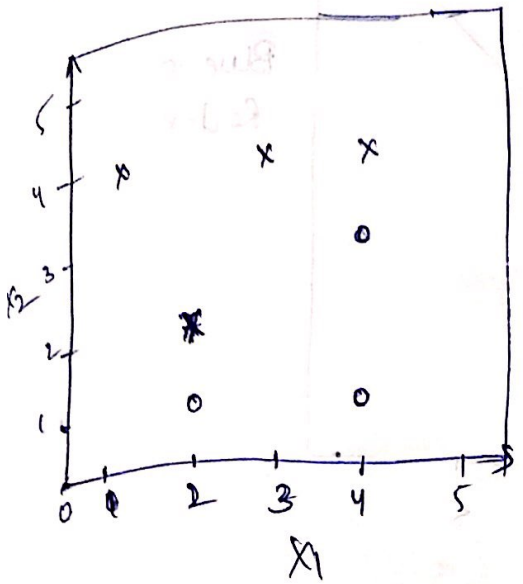
Q.5(i)

From the above results, it can be inferred that the Support Vector machines when used with non-linear kernel have greater impact in finding the non-linear boundary. However, both SVMs when used with kernels and the logistic regressions when used with non-interactions, are not successful in finding the decision boundary. If we add the interaction terms to logistic regression, it will have same impact as the radial-basis kernels. But, this requires some tuning and manual effort. This may not be fruitful if there are large number of features. On the other hand, radial basis kernels may need only tuning through cross validation of one parameter i.e gamma

chapter 9 - (Excercise 9.7)

Q.3(a) $n=7$ $p=2$

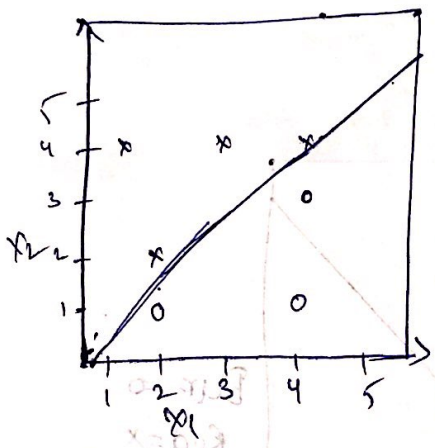
obs	X_1	X_2	Y
1	3	4	Red
2	2	2	Red
3	4	4	Red
4	1	4	Red
5	2	1	Blue
6	4	3	Blue
7	4	1	Blue



Red = X

Blue = O

b)



Blue = O

Red = X

(Maximal Margin classifier has to be between observation 2,3 & 5,6

So (2,2), (4,4)

(2,1), (4,3)

$\Rightarrow (2, 1.5)\ (4, 3.5)$

$b = \frac{(3.5-1.5)}{4-2}$

$= 1$

$a = X_2 - X_1 = 1.5 - 2$

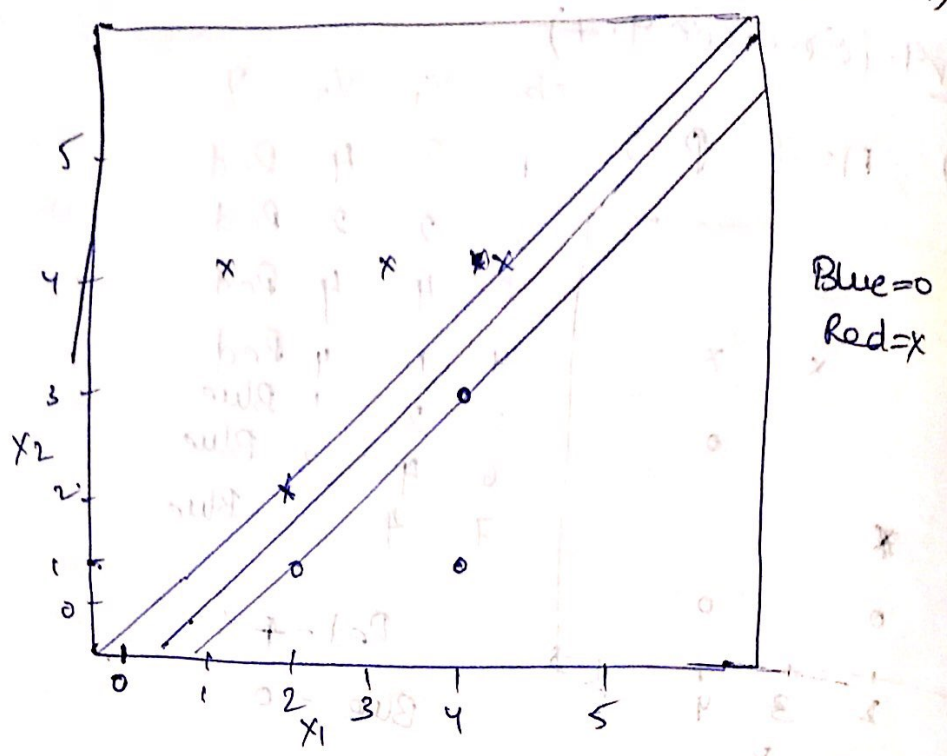
$= 0.5$

$\therefore$ abline$(-0.5, 1)$

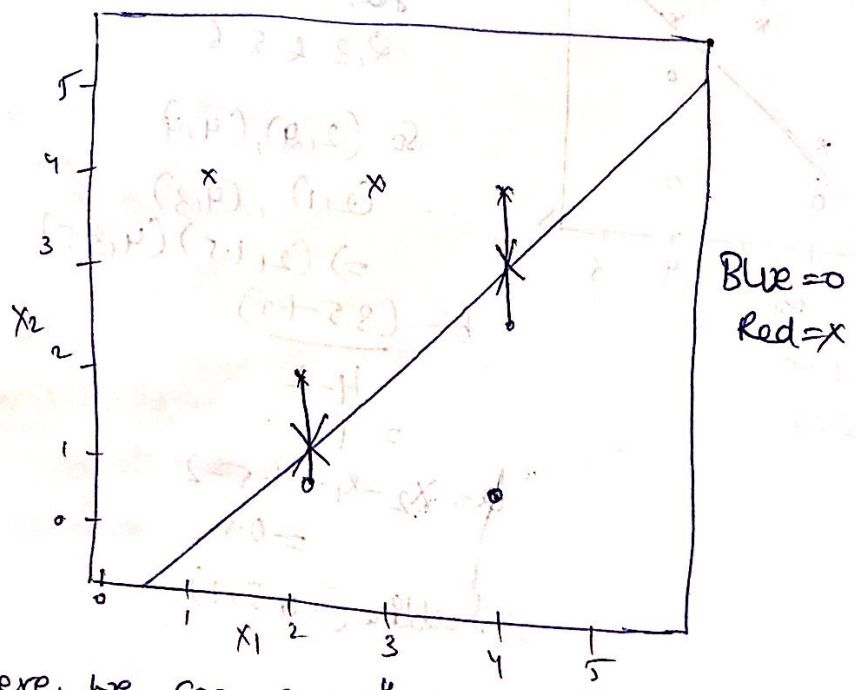
c) $0.5 - X_1 + X_2 > 0$

abline(-0.5,1)
abline(-1,1)
abline(0,1)

d)



e)

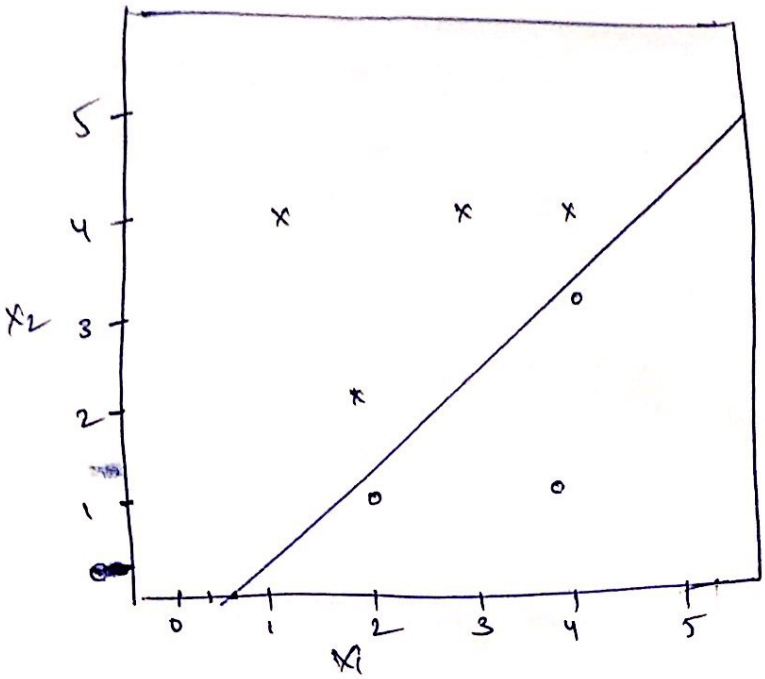


f) Here, we can see that (4,1) is outside the margin, so it would not have effect of maximal hyper plane. So, the observation 7 has no effect on maximal Margin Hyperplane

abluni (-0.8, 1)

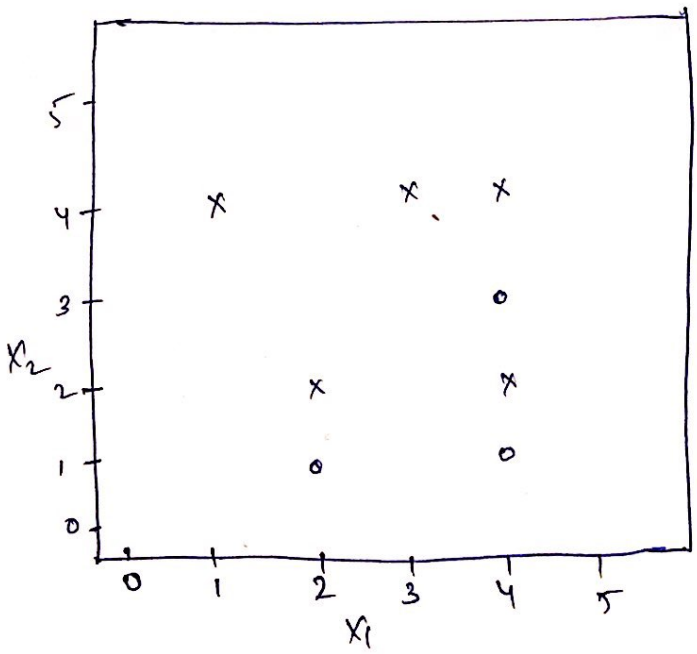
$-0.9 - x_1 + x_2 > 0$

g)



Blue = o
Red = x

h)



Blue = o
Red = x

Point (4, 2)
col = Red