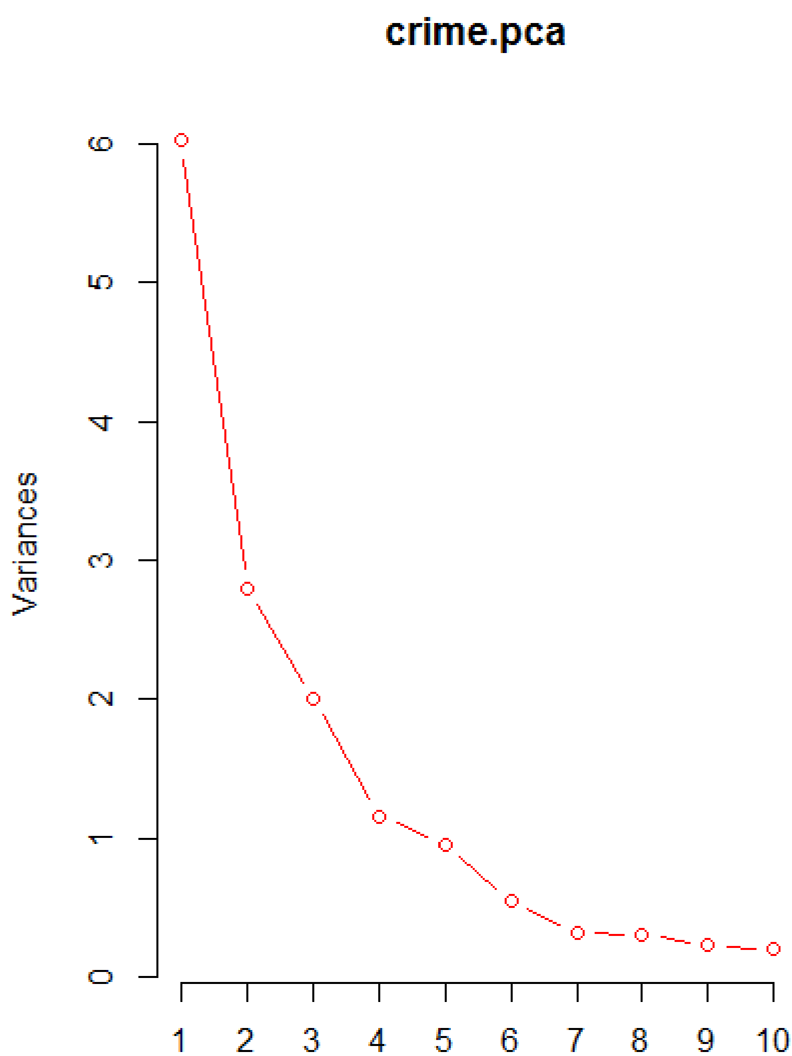
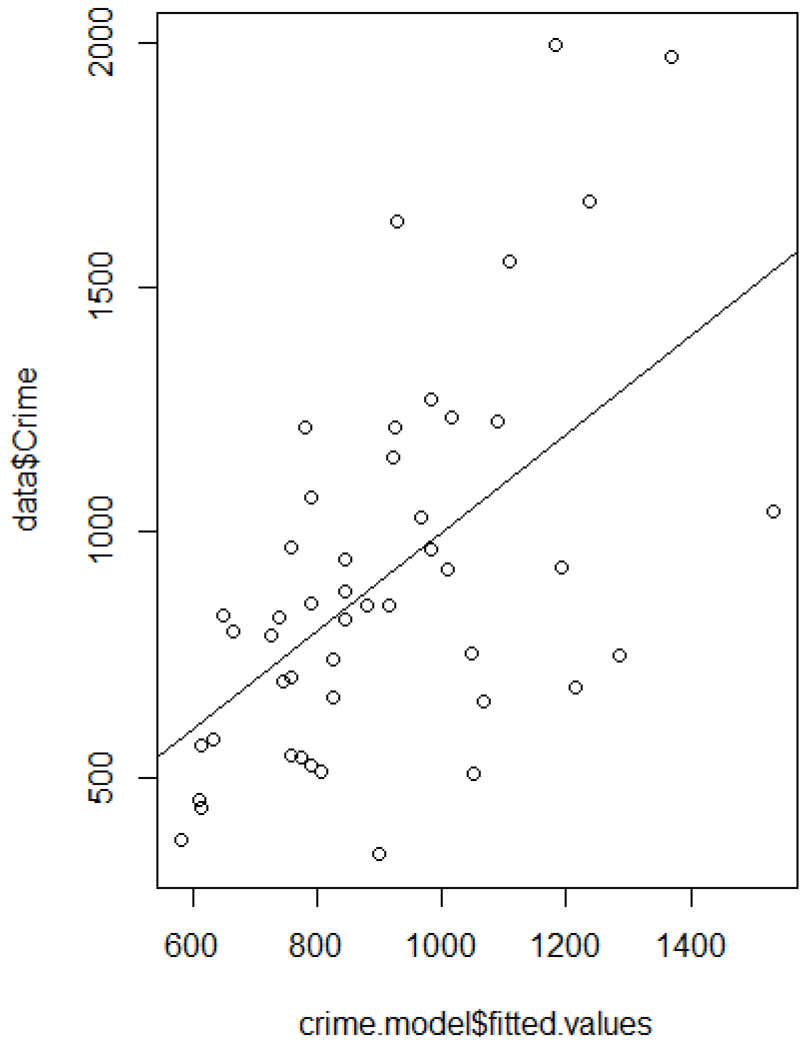
Question 1:-

```
> summary(crime.pca)

Importance of components:

                          PC1    PC2    PC3     PC4     PC5     PC6     PC7
Standard deviation     2.4534 1.6739 1.4160 1.07806 0.97893 0.74377 0.56729
Proportion of Variance 0.4013 0.1868 0.1337 0.07748 0.06389 0.03688 0.02145
Cumulative Proportion  0.4013 0.5880 0.7217 0.79920 0.86308 0.89996 0.92142
                           PC8     PC9    PC10    PC11    PC12    PC13   PC14
Standard deviation     0.55444 0.48493 0.44708 0.41915 0.35804 0.26333 0.2418
Proportion of Variance 0.02049 0.01568 0.01333 0.01171 0.00855 0.00462 0.0039
Cumulative Proportion  0.94191 0.95759 0.97091 0.98263 0.99117 0.99579 0.9997
                          PC15
Standard deviation     0.06793
Proportion of Variance 0.00031
Cumulative Proportion  1.00000
```

ScreePlot:-

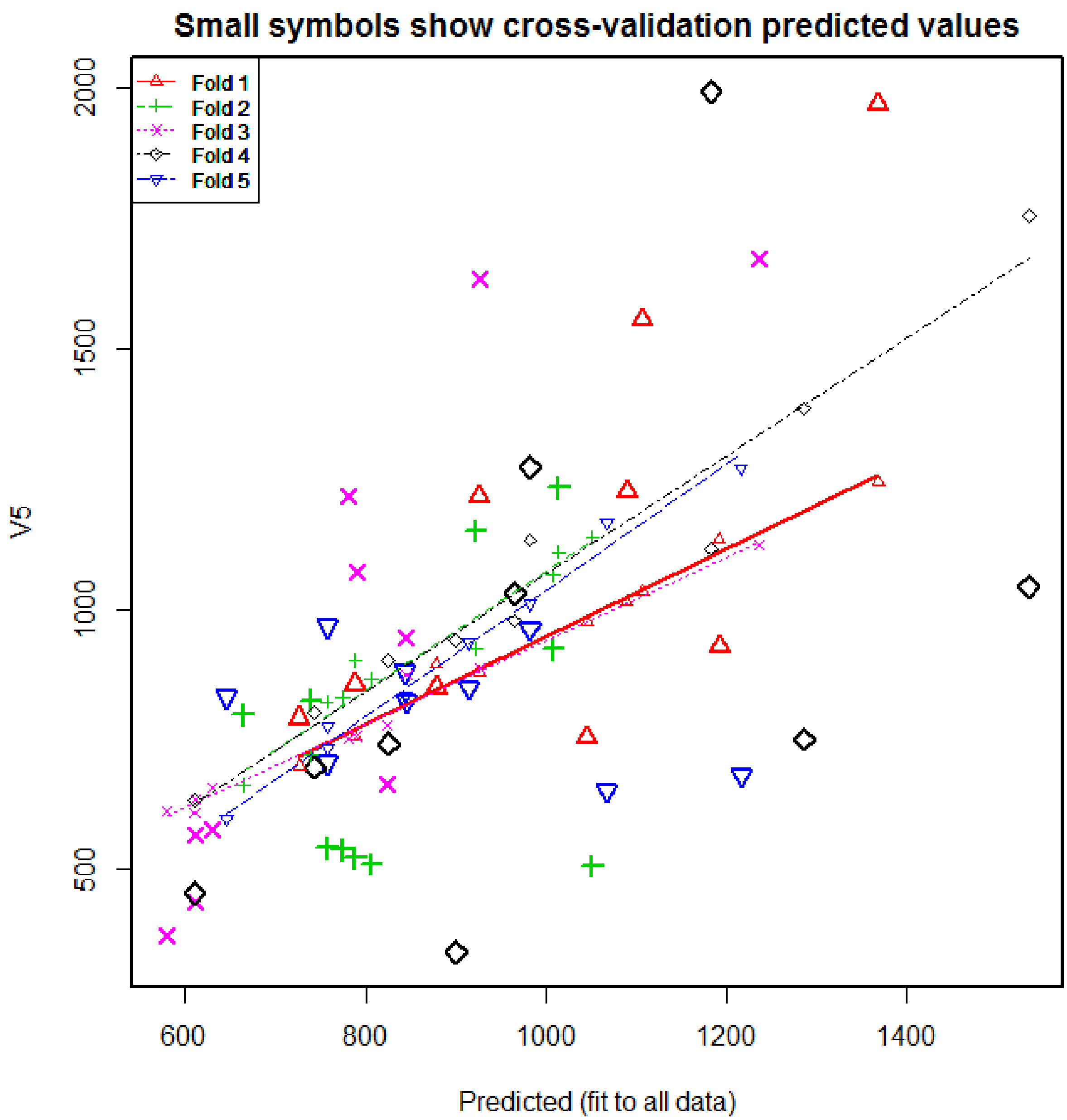





```
(Intercept)       PC1       PC2       PC3
 905.08511   65.21593  -70.08312   25.19408
      PC4
  69.44603
> t(alphas)
                 M        So        Ed       Po1       Po2
[1,] -21.27796 10.22309 14.35261 63.45643 64.55797
                LF       M.F       Pop        NW
[1,] -14.00535 -24.43757 39.83067 15.43455
```

```
              U1         U2    Wealth       Ineq
[1,] -27.22228 1.425902 38.60786 -27.53635
            Prob       Time
[1,] 3.295707 -6.612616
```

Cross validation



R squared:-

.894

Comparison:- No major change observed between regular Model and PCA. Rsquare is pretty close

R- Code:-

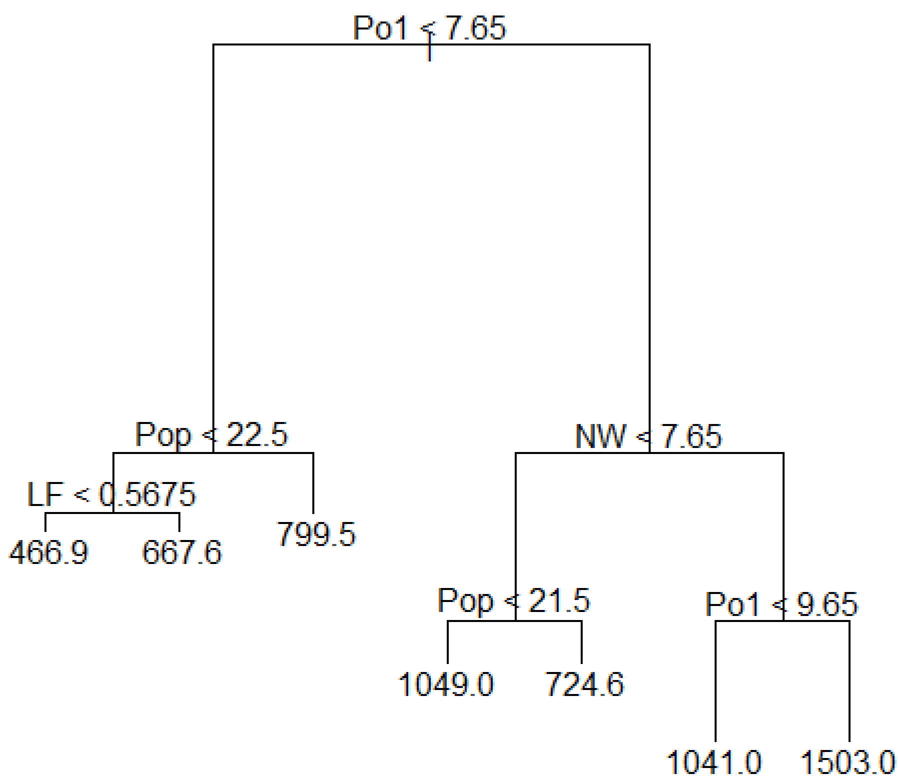
Assignment #1

```
data <-
read.table("http://www.statsci.org/data/general/uscrime.txt",s
ep="\t", header=TRUE)
head(data)
install.packages("CompGLM")
library(CompGLM)
crime.pca <- prcomp(data[,1:15],scale. = TRUE)
summary(crime.pca)
screeplot(crime.pca,type="lines",col="red")
crime.pca
var=crime.pca$sdev^2
propvar = var/sum(var)
plot(cumsum(propvar))
pcs = crime.pca$x[,1:4]
pccrime <- cbind(pcs,data[,"Crime"])
crime.model <- lm(V5~.,data=as.data.frame(pccrime))
summary(crime.model)
plot(crime.model$fitted.values,data$Crime)
abline(0,1)
plot(crime.model$fitted.values,scale(crime.model$residuals))
betas <- crime.model$coefficients[2:5]
alphas <- crime.pca$rotation[,1:4]%*%betas
t(alphas)
install.packages("pls")
library(pls)
pcr.fit = pcr(Crime~.,data=data,scale=TRUE)
summary(pcr.fit)
coef(pcr.fit)
```

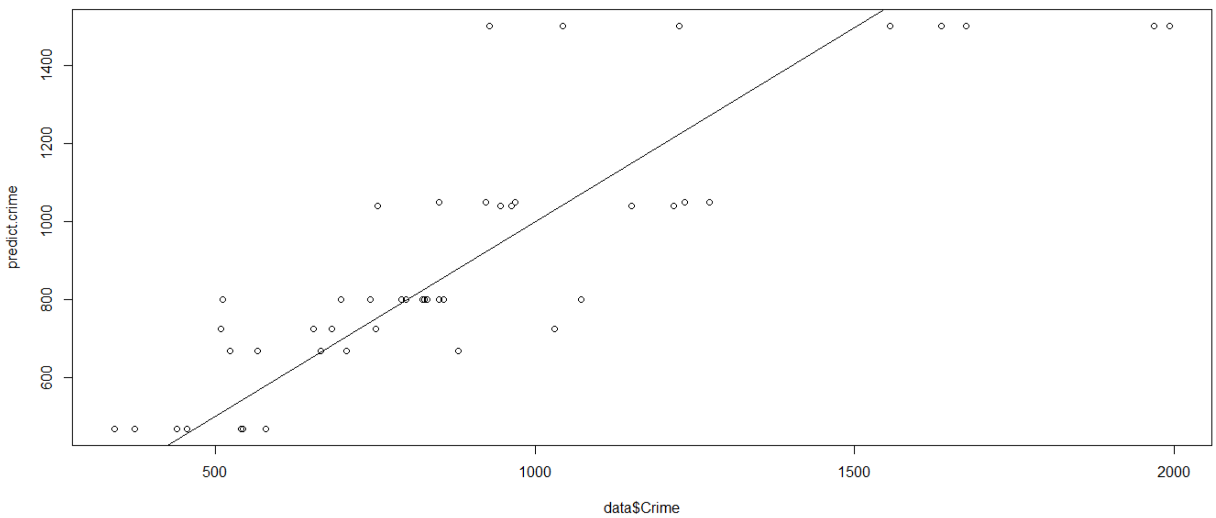
Question2:-

Summary:-

```
Regression tree:
tree(formula = Crime ~ ., data = data)
Variables actually used in tree construction:
[1] "Po1" "Pop" "LF"  "NW"
Number of terminal nodes:  7
Residual mean deviance:  47390 = 1896000 / 40
Distribution of residuals:
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-573.900  -98.300   -1.545    0.000  110.600  490.100
```



Prediction:-




```
> sstot = sum((data$Crime - mean(data$Crime))^2)
> r2 = 1 - ssres/sstot
> r2
[1] 0.7244962
> cv.tree(tree.data)
$size
[1] 7 6 5 4 3 2 1

$dev
[1] 8123895 8141074 8801881 9161379 9561715 8606786 8159704

$k
[1]      -Inf  117534.9  263412.9  355961.8  731412.1 1019362.7 2497521.7

$method
[1] "deviance"

attr(,"class")
[1] "prune"         "tree.sequence"
```

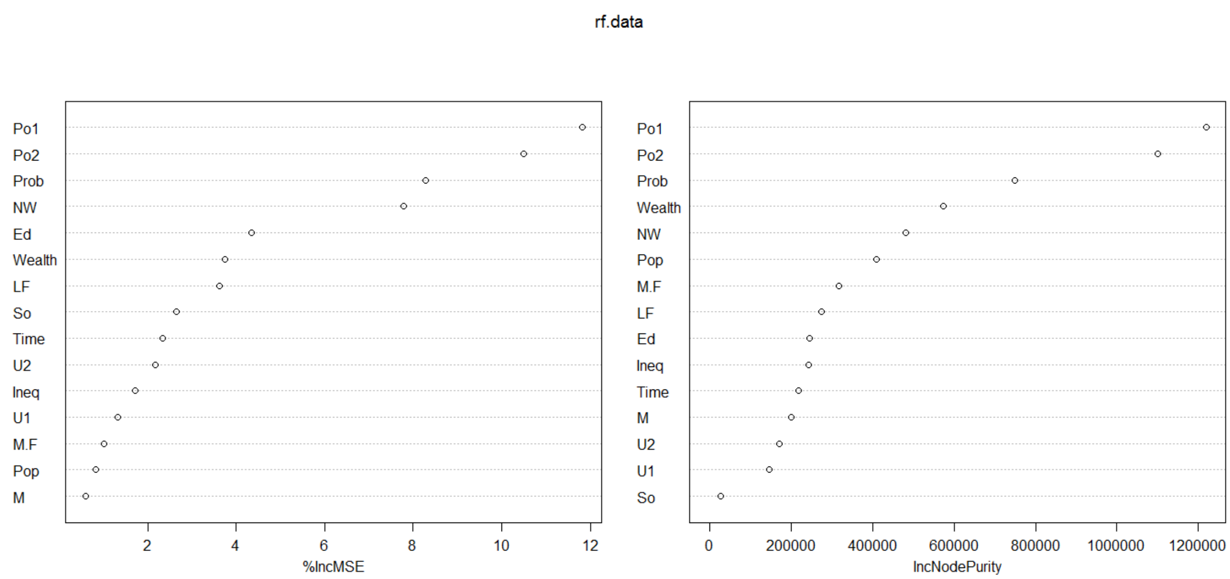
R- Code:

Assignment 2:-

```
install.packages("tree")
install.packages("MASS")
library(tree)
library(MASS)
set.seeed(1)
tree.data <- tree(Crime~.,data=data)
summary(tree.data)
plot(tree.data)
text(tree.data)
cv.tree(tree.data)
termnodes = 5
prune.data <- prune.tree(tree.data,best=termnodes)
predict.crime = predict(tree.data)
ssres = sum((predict.crime-data$Crime)^2)
plot(data$Crime,predict.crime)
abline(0,1)
plot(data$Crime,scale(predict.crime-data$Crime))
abline(0,0)
sstot = sum((data$Crime - mean(data$Crime))^2)
r2 = 1 - ssres/sstot
r2
```

2.2 :- Random Forest

```
> importance(rf.data)
          %IncMSE IncNodePurity
M       0.5949712     200681.22
So      2.6383397      27528.66
Ed      4.3375837     245793.97
Po1    11.8143933    1218574.59
Po2    10.4978019    1100955.83
LF      3.6223058     275606.54
M.F     1.0122219     317880.44
Pop     0.8249897     408418.62
NW      7.7919759     482281.81
U1      1.3302065     146537.98
U2      2.1698443     170303.91
Wealth  3.7358676     574134.71
Ineq    1.7090580     241945.33
Prob    8.2873199     749370.61
Time    2.3432048     218148.78
```



RSquared is .15

R-Code:-

```
Random Forest:-
install.packages("randomForest")
library(randomForest)
numpred = 4
rf.data =
randomForest(Crime~.,data=data,mtry=numpred,importance=T
RUE)
importance(rf.data)
varImpPlot(rf.data)
for (i in 1:nrow(data)){
rf.x = randomForest(Crime~.,data=data[-
i,],mtry=numpred,imprtance=TRUE)
ssres=ssres + (predict(rf.x,newdata=data[i,]) - data[i,16])^2
}
r2 = 1-ssres/sstot
r2
```

Question 3:-

Reaching to work on time

Response variable:- Am I on time or not

Predictors:-

1. Start time window
2. Traffic in 101 freeway
3. School Traffic
4. Elevator wait time
5. Speed

Question4:-

```
Call:
glm(formula = V21 ~ ., family = binomial(link = "logit"), data = data,
    maxit = 100)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3410  -0.6994  -0.3752   0.7095   2.6116

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.005e-01  1.084e+00   0.369 0.711869
V1A12       -3.749e-01  2.179e-01  -1.720 0.085400 .
V1A13       -9.657e-01  3.692e-01  -2.616 0.008905 **
V1A14       -1.712e+00  2.322e-01  -7.373 1.66e-13 ***
V2           2.786e-02  9.296e-03   2.997 0.002724 **
V3A31        1.434e-01  5.489e-01   0.261 0.793921
V3A32       -5.861e-01  4.305e-01  -1.362 0.173348
V3A33       -8.532e-01  4.717e-01  -1.809 0.070470 .
V3A34       -1.436e+00  4.399e-01  -3.264 0.001099 **
V4A41       -1.666e+00  3.743e-01  -4.452 8.51e-06 ***
V4A410      -1.489e+00  7.764e-01  -1.918 0.055163 .
V4A42       -7.916e-01  2.610e-01  -3.033 0.002421 **
V4A43       -8.916e-01  2.471e-01  -3.609 0.000308 ***
V4A44       -5.228e-01  7.623e-01  -0.686 0.492831
V4A45       -2.164e-01  5.500e-01  -0.393 0.694000
V4A46        3.628e-02  3.965e-01   0.092 0.927082
V4A48       -2.059e+00  1.212e+00  -1.699 0.089297 .
V4A49       -7.401e-01  3.339e-01  -2.216 0.026668 *
V5           1.283e-04  4.444e-05   2.887 0.003894 **
V6A62       -3.577e-01  2.861e-01  -1.250 0.211130
V6A63       -3.761e-01  4.011e-01  -0.938 0.348476
V6A64       -1.339e+00  5.249e-01  -2.551 0.010729 *
V6A65       -9.467e-01  2.625e-01  -3.607 0.000310 ***
V7A72       -6.691e-02  4.270e-01  -0.157 0.875475
V7A73       -1.828e-01  4.105e-01  -0.445 0.656049
```

```
V7A74       -8.310e-01  4.455e-01  -1.866 0.062110 .
V7A75       -2.766e-01  4.134e-01  -0.669 0.503410
V8           3.301e-01  8.828e-02   3.739 0.000185 ***
V9A92       -2.755e-01  3.865e-01  -0.713 0.476040
V9A93       -8.161e-01  3.799e-01  -2.148 0.031718 *
V9A94       -3.671e-01  4.537e-01  -0.809 0.418448
V10A102      4.360e-01  4.101e-01   1.063 0.287700
V10A103     -9.786e-01  4.243e-01  -2.307 0.021072 *
V11          4.776e-03  8.641e-02   0.055 0.955920
V12A122      2.814e-01  2.534e-01   1.111 0.266630
V12A123      1.945e-01  2.360e-01   0.824 0.409743
V12A124      7.304e-01  4.245e-01   1.721 0.085308 .
V13         -1.454e-02  9.222e-03  -1.576 0.114982
V14A142     -1.232e-01  4.119e-01  -0.299 0.764878
V14A143     -6.463e-01  2.391e-01  -2.703 0.006871 **
V15A152     -4.436e-01  2.347e-01  -1.890 0.058715 .
V15A153     -6.839e-01  4.770e-01  -1.434 0.151657
V16          2.721e-01  1.895e-01   1.436 0.151109
V17A172      5.361e-01  6.796e-01   0.789 0.430160
V17A173      5.547e-01  6.549e-01   0.847 0.397015
V17A174      4.795e-01  6.623e-01   0.724 0.469086
V18          2.647e-01  2.492e-01   1.062 0.288249
V19A192     -3.000e-01  2.013e-01  -1.491 0.136060
V20A202     -1.392e+00  6.258e-01  -2.225 0.026095 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1221.73  on 999  degrees of freedom
Residual deviance:  895.82  on 951  degrees of freedom
AIC: 993.82

Number of Fisher Scoring iterations: 5
```

2.Choosing the best Model:-

```
> step(model)
Start:  AIC=993.82
V21 ~ V1 + V2 + V3 + V4 + V5 + V6 + V7 + V8 + V9 + V10 + V11 +
    V12 + V13 + V14 + V15 + V16 + V17 + V18 + V19 + V20

       Df Deviance     AIC
- V17   3   896.56  988.56
- V12   3   899.08  991.08
- V11   1   895.82  991.82
- V18   1   896.94  992.94
<none>      895.82  993.82
- V16   1   897.89  993.89
- V7    4   904.03  994.03
- V15   2   900.05  994.05
- V19   1   898.06  994.06
- V13   1   898.34  994.34
- V9    3   905.15  997.15
- V10   2   903.24  997.24
- V20   1   901.88  997.88
- V14   2   903.98  997.98
- V5    1   904.28 1000.28
- V2    1   904.87 1000.87
- V6    4   915.63 1005.63
- V8    1   910.27 1006.27
- V3    4   917.62 1007.62
- V4    9   931.12 1011.12
- V1    3   962.05 1054.05

Step:  AIC=988.56
V21 ~ V1 + V2 + V3 + V4 + V5 + V6 + V7 + V8 + V9 + V10 + V11 +
    V12 + V13 + V14 + V15 + V16 + V18 + V19 + V20

       Df Deviance     AIC
- V12   3   899.79  985.79
- V11   1   896.57  986.57
- V18   1   897.67  987.67
```

```
- V7    4   904.32  988.32
- V16   1   898.47  988.47
<none>      896.56  988.56
- V15   2   900.60  988.60
- V19   1   899.13  989.13
- V13   1   899.19  989.19
- V9    3   905.83  991.83
- V10   2   903.87  991.87
- V20   1   902.67  992.67
- V14   2   904.95  992.95
- V5    1   905.31  995.31
- V2    1   905.85  995.85
- V6    4   917.02 1001.02
- V8    1   911.45 1001.45
- V3    4   918.12 1002.12
- V4    9   931.82 1005.82
- V1    3   962.35 1048.35

Step:  AIC=985.79
V21 ~ V1 + V2 + V3 + V4 + V5 + V6 + V7 + V8 + V9 + V10 + V11 +
    V13 + V14 + V15 + V16 + V18 + V19 + V20

       Df Deviance     AIC
- V11   1   899.81  983.81
- V18   1   900.79  984.79
- V15   2   903.47  985.47
- V16   1   901.49  985.49
<none>      899.79  985.79
- V19   1   901.81  985.81
- V7    4   907.85  985.85
- V13   1   902.52  986.52
- V9    3   908.67  988.67
- V20   1   905.83  989.83
- V10   2   908.05  990.05
- V14   2   908.87  990.87
- V5    1   909.80  993.80
```

```
- V2    1   909.99  993.99
- V6    4   919.78  997.78
- V8    1   915.56  999.56
- V3    4   921.66  999.66
- V4    9   936.35 1004.35
- V1    3   967.78 1047.78

Step:  AIC=983.81
V21 ~ V1 + V2 + V3 + V4 + V5 + V6 + V7 + V8 + V9 + V10 + V13 +
    V14 + V15 + V16 + V18 + V19 + V20

       Df Deviance     AIC
- V18   1   900.81  982.81
- V16   1   901.53  983.53
- V19   1   901.81  983.81
<none>      899.81  983.81
- V7    4   907.86  983.86
- V15   2   903.95  983.95
- V13   1   902.53  984.53
- V9    3   908.69  986.69
- V20   1   905.88  987.88
- V10   2   908.08  988.08
- V14   2   908.87  988.87
- V5    1   909.80  991.80
- V2    1   910.05  992.05
- V6    4   919.78  995.78
- V8    1   915.59  997.59
- V3    4   921.66  997.66
- V4    9   936.35 1002.35
- V1    3   968.09 1046.09

Step:  AIC=982.81
V21 ~ V1 + V2 + V3 + V4 + V5 + V6 + V7 + V8 + V9 + V10 + V13 +
    V14 + V15 + V16 + V19 + V20

       Df Deviance     AIC
```

```
- V16   1   902.80  982.80
<none>      900.81  982.81
- V7    4   908.82  982.82
- V19   1   902.90  982.90
- V15   2   905.01  983.01
- V13   1   903.38  983.38
- V9    3   908.75  984.75
- V20   1   906.79  986.79
- V10   2   908.83  986.83
- V14   2   909.90  987.90
- V5    1   910.50  990.50
- V2    1   910.95  990.95
- V6    4   920.53  994.53
- V8    1   915.95  995.95
- V3    4   923.28  997.28
- V4    9   937.61 1001.61
- V1    3   969.35 1045.35

Step:  AIC=982.8
V21 ~ V1 + V2 + V3 + V4 + V5 + V6 + V7 + V8 + V9 + V10 + V13 +
    V14 + V15 + V19 + V20

       Df Deviance     AIC
- V7    4   910.50  982.50
- V19   1   904.72  982.72
<none>      902.80  982.80
- V15   2   907.17  983.17
- V13   1   905.21  983.21
- V9    3   910.47  984.47
- V10   2   910.88  986.88
- V20   1   909.14  987.14
- V14   2   912.57  988.57
- V2    1   912.50  990.50
- V5    1   912.54  990.54
- V6    4   922.72  994.72
- V3    4   923.43  995.43
```

```
- V8    1   917.65  995.65
- V4    9   940.20 1002.20
- V1    3   971.04 1045.04

Step:  AIC=982.5
V21 ~ V1 + V2 + V3 + V4 + V5 + V6 + V8 + V9 + V10 + V13 + V14 +
    V15 + V19 + V20

       Df Deviance     AIC
<none>      910.50  982.50
- V19   1   912.82  982.82
- V13   1   912.96  982.96
- V15   2   915.01  983.01
- V20   1   916.63  986.63
- V9    3   920.98  986.98
- V10   2   919.41  987.41
- V14   2   920.76  988.76
- V2    1   918.79  988.79
- V5    1   920.07  990.07
- V6    4   931.53  995.53
- V8    1   925.92  995.92
- V3    4   932.30  996.30
- V4    9   947.78 1001.78
- V1    3   979.75 1045.75

Call:  glm(formula = V21 ~ V1 + V2 + V3 + V4 + V5 + V6 + V8 + V9 + V10 +
    V13 + V14 + V15 + V19 + V20, family = binomial(link = "logit"),
    data = data, maxit = 100)

Coefficients:
(Intercept)       V1A12       V1A13       V1A14          V2       V3A31
  1.7495411  -0.3900152  -1.0240813  -1.7177165   0.0256787  -0.1187724
      V3A32       V3A33       V3A34       V4A41      V4A410       V4A42
 -0.8303101  -0.9097304  -1.4917085  -1.6072585  -1.4349203  -0.7404978
      V4A43       V4A44       V4A45       V4A46       V4A48       V4A49
 -0.9194787  -0.5250945  -0.1424475   0.1435655  -2.1643060  -0.7826591
```

```
       V5      V6A62      V6A63      V6A64      V6A65         V8
0.0001294 -0.3282182 -0.4303584 -1.2894345 -0.9628458  0.3299308
    V9A92      V9A93      V9A94    V10A102    V10A103        V13
-0.2872096 -0.8227885 -0.4169133  0.4874391 -1.0404263 -0.0130933
  V14A142    V14A143    V15A152    V15A153    V19A192    V20A202
-0.0786395 -0.6994941 -0.4415029 -0.1496754 -0.2794111 -1.3824572

Degrees of Freedom: 999 Total (i.e. Null);  964 Residual
Null Deviance:      1222
Residual Deviance: 910.5        AIC: 982.5
```

4. Rerunning the model that is the best

```
Call:
glm(formula = V21 ~ V1 + V2 + V3 + V4 + V5 + V6 + V6 + V8 + V9 +
    V10 + V13 + V14 + V18 + V20, family = binomial(link = "logit"),
    data = data, maxit = 100)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2194  -0.7062  -0.3810   0.7679   2.8151

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.122e+00  7.671e-01   1.463 0.143482
V1A12       -4.105e-01  2.107e-01  -1.948 0.051365 .
V1A13       -1.035e+00  3.617e-01  -2.863 0.004202 **
V1A14       -1.774e+00  2.273e-01  -7.804 5.98e-15 ***
V2           2.685e-02  8.813e-03   3.047 0.002315 **
V3A31       -1.460e-01  5.201e-01  -0.281 0.779003
V3A32       -8.610e-01  4.057e-01  -2.122 0.033832 *
V3A33       -9.597e-01  4.606e-01  -2.084 0.037203 *
V3A34       -1.544e+00  4.280e-01  -3.608 0.000308 ***
V4A41       -1.552e+00  3.615e-01  -4.292 1.77e-05 ***
V4A410      -1.558e+00  7.527e-01  -2.070 0.038412 *
V4A42       -7.000e-01  2.525e-01  -2.772 0.005570 **
V4A43       -9.193e-01  2.424e-01  -3.792 0.000150 ***
```

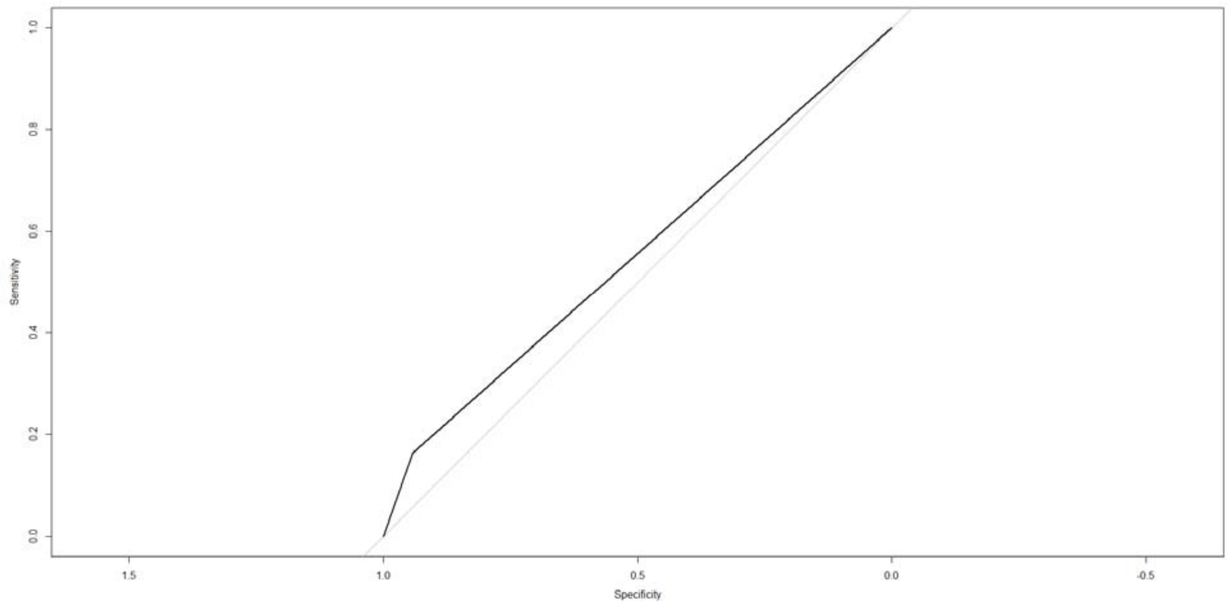
```
V4A44       -5.399e-01  7.375e-01  -0.732 0.464146
V4A45       -1.347e-01  5.356e-01  -0.252 0.801420
V4A46        1.858e-01  3.880e-01   0.479 0.632129
V4A48       -2.125e+00  1.201e+00  -1.769 0.076830 .
V4A49       -8.264e-01  3.223e-01  -2.565 0.010331 *
V5           1.180e-04  4.065e-05   2.902 0.003711 **
V6A62       -3.040e-01  2.750e-01  -1.106 0.268863
V6A63       -4.019e-01  3.965e-01  -1.014 0.310788
V6A64       -1.338e+00  5.108e-01  -2.619 0.008821 **
V6A65       -9.567e-01  2.554e-01  -3.746 0.000179 ***
V8           3.251e-01  8.495e-02   3.827 0.000130 ***
V9A92       -2.082e-01  3.693e-01  -0.564 0.573012
V9A93       -8.617e-01  3.643e-01  -2.365 0.018017 *
V9A94       -3.898e-01  4.392e-01  -0.888 0.374787
V10A102      5.587e-01  3.967e-01   1.408 0.159001
V10A103     -1.019e+00  4.134e-01  -2.465 0.013700 *
V13         -1.576e-02  7.979e-03  -1.976 0.048172 *
V14A142     -1.457e-01  4.024e-01  -0.362 0.717335
V14A143     -6.784e-01  2.341e-01  -2.898 0.003759 **
V18          2.891e-01  2.434e-01   1.188 0.234873
V20A202     -1.325e+00  6.206e-01  -2.135 0.032765 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1221.73  on 999  degrees of freedom
Residual deviance:  915.85  on 966  degrees of freedom
AIC: 983.85

Number of Fisher Scoring iterations: 5
```

6. Pick up the column after resetting the data

```
Call:
glm(formula = V21 ~ V1A13 + V4A41 + V4A42 + V4A43 + V5 + V6A64 +
    V6A65 + V14A143, family = binomial(link = "logit"), data = data,
```

```
    maxit = 100)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7457  -0.8780  -0.6603   1.1784   2.3350

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.3126188  0.1999268  -1.564 0.117896
V1A13       -0.3372613  0.3214658  -1.049 0.294116
V4A41       -1.3197124  0.3014160  -4.378 1.20e-05 ***
V4A42       -0.2359357  0.1942581  -1.215 0.224539
V4A43       -0.6332838  0.1812329  -3.494 0.000475 ***
V5           0.0001362  0.0000267   5.102 3.37e-07 ***
V6A64       -1.2339374  0.4509007  -2.737 0.006208 **
V6A65       -0.9791016  0.2203000  -4.444 8.81e-06 ***
V14A143     -0.5693045  0.1771638  -3.213 0.001312 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1221.7  on 999  degrees of freedom
Residual deviance: 1123.6  on 991  degrees of freedom
AIC: 1141.6
```

ROC Curve



Number of Fisher Scoring iterations: 4

7. **Confusion matrix**

```
ypred   0   1
    0 660 251
    1   40  49
```

R- Code:-

```
Assigniment 4:-

data <- read.table("https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/german/german.data",sep=" ")

head(data)

data$V21[data$V21==1] <- 0

data$V21[data$V21==2] <- 1

data = as.data.frame(data)

smp_size <- floor(0.75 * nrow(css))

set.seed(123)

train_ind <- sample(seq_len(nrow(css)),size=smp_size)

data.train <- data[train_ind,]

data.test <- data[-train_ind,]

model = glm(formula=V21~.,family=binomial(link = "logit"),data=data,maxit=100)

model = glm(formula= V21~ V1 + V2 + V3 + V4 + V5 +V6 + V6 + V8 + V9 + V10 +

V13 + V14 + V18 + V20,family=binomial(link="logit"),data=data,maxit=100)
```

```
summary(model)

head(data$V21)

step(model)

conversion:-

data$V1A14[data$V1 == "A14"] <- 1

data$V1A14[data$V1 != "A14"] <- 0

data$V1A13[data$V1 == "A13"] <- 1

data$V1A13[data$V1 != "A13"] <- 0

data$V3A34[data$V3 == "A34"] <- 1

data$V3A34[data$V3 != "A34"] <- 0

data$V4A41[data$V4 == "A41"] <- 1

data$V4A41[data$V4 != "A41"] <- 0

data$V4A42[data$V4 == "A42"] <- 1

data$V4A42[data$V4 != "A42"] <- 0

data$V4A43[data$V4 == "A43"] <- 1

data$V4A43[data$V4 != "A43"] <- 0

data$V6A64[data$V6 == "A64"] <- 1

data$V6A64[data$V6 != "A64"] <- 0

data$V6A65[data$V6 == "A65"] <- 1

data$V6A65[data$V6 != "A65"] <- 0

data$V14A143[data$V14 == "A143"] <- 1

data$V14A143[data$V14 != "A143"] <- 0


model = glm(formula= V21~ V1A13 + V4A41 +  V4A42 + V4A43 + V5 +V6A64 + V6A65 +
V14A143,family=binomial(link="logit"),data=data,maxit=100)

summary(model)


yhat = predict(model,data,type="response")

ypred = as.integer(yhat > 0.5)

head(ypred)


install.packages("pROC")

library(pROC)

roc(data$V21,ypred)

plot(roc(data$V21,ypred))

table <- as.matrix(table(ypred,data$V21))
```