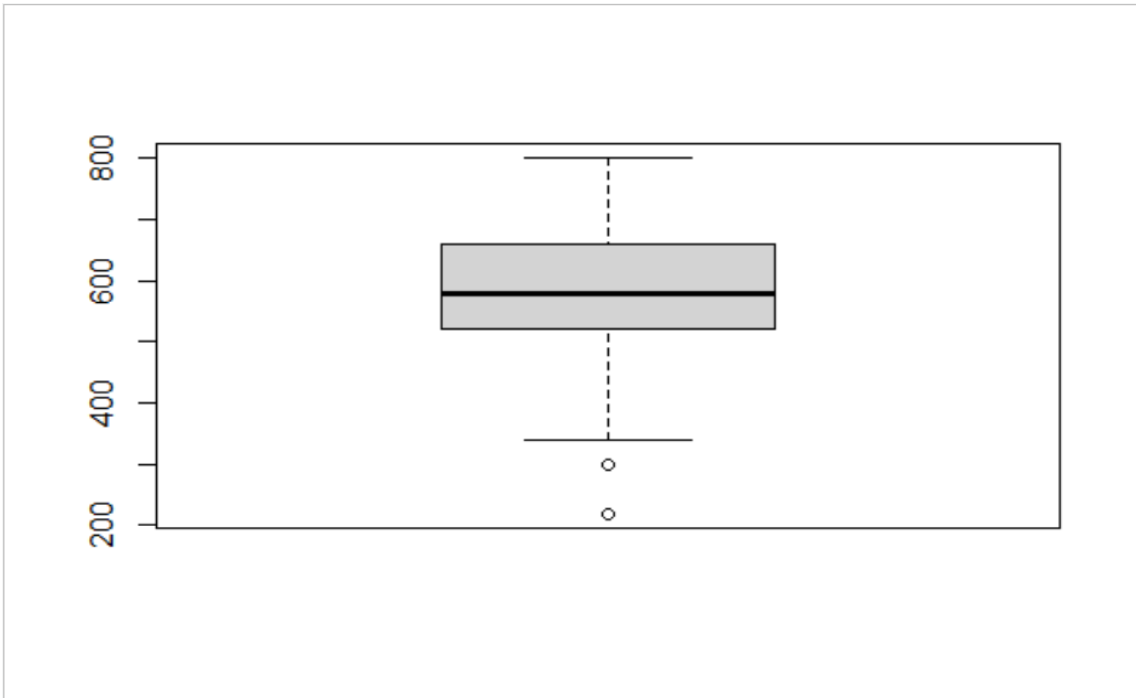
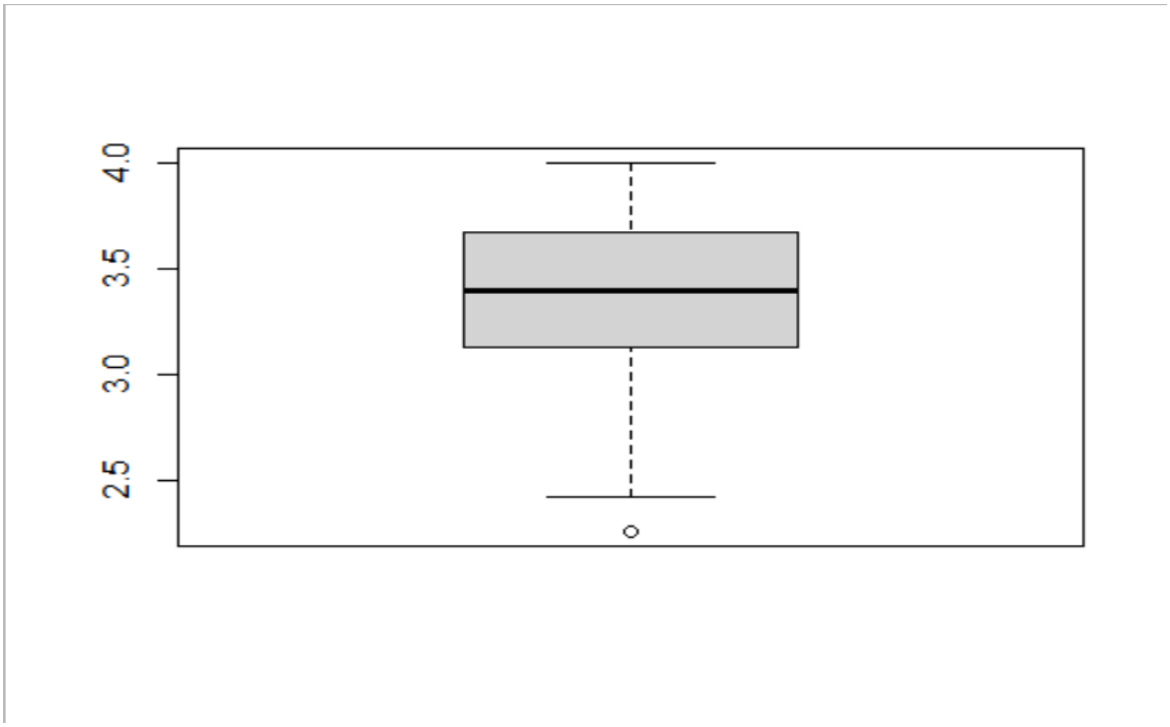
SCREEN SHOTS – COLLEGE ADMISSION PROJECT

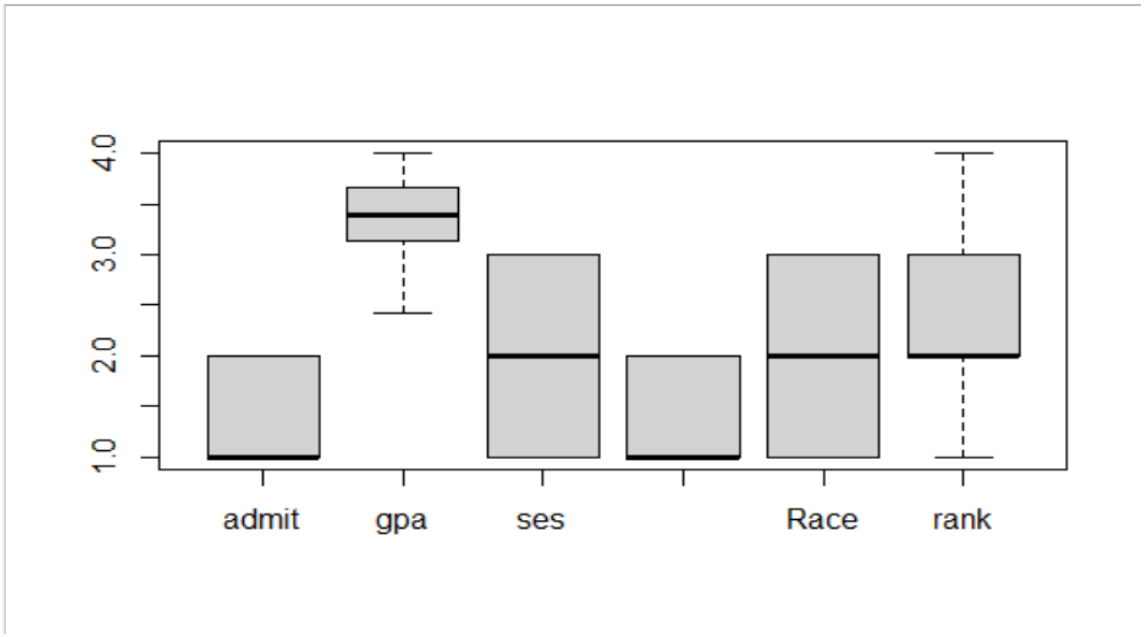
Outlier in GRE:



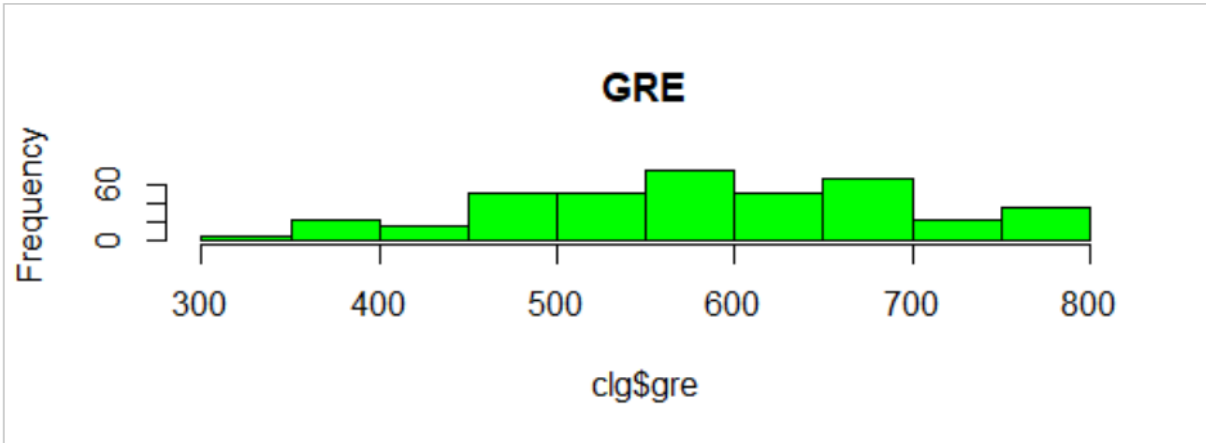
Outlier in GPA:



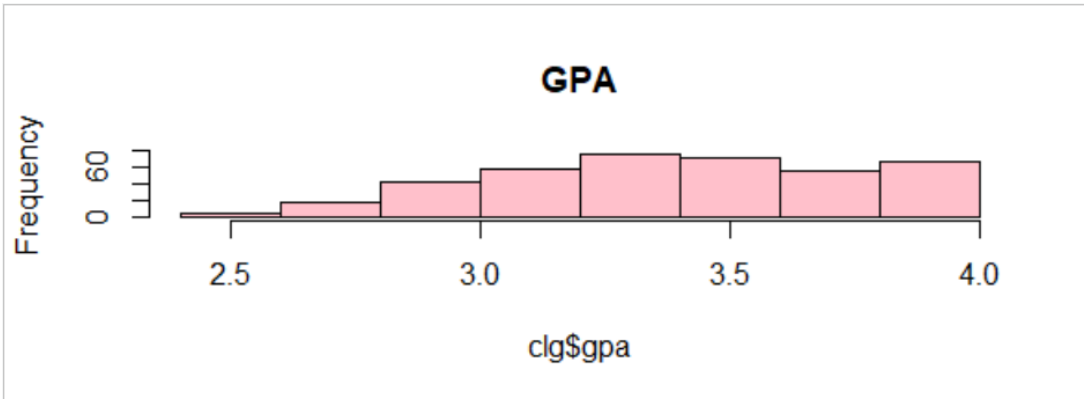
After removing outliers:



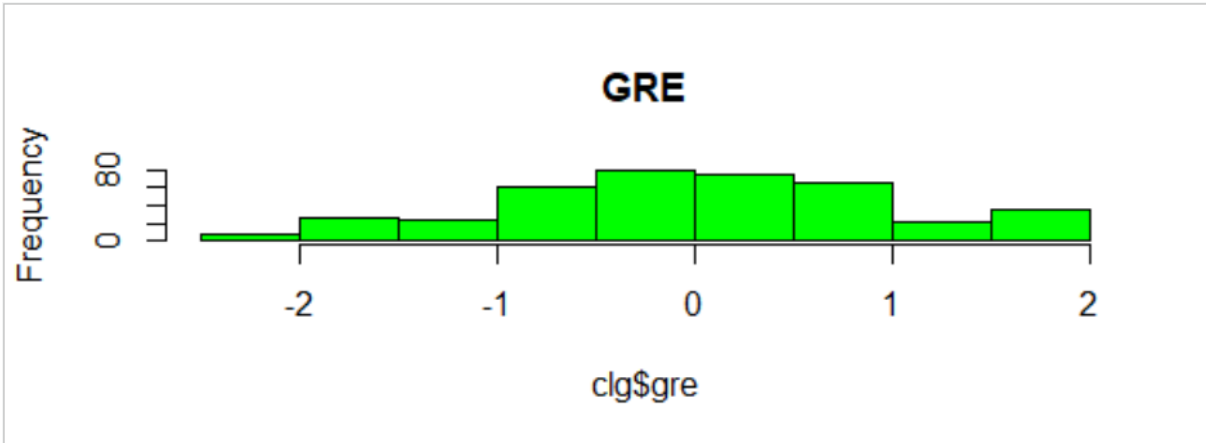
Distribution of GRE:



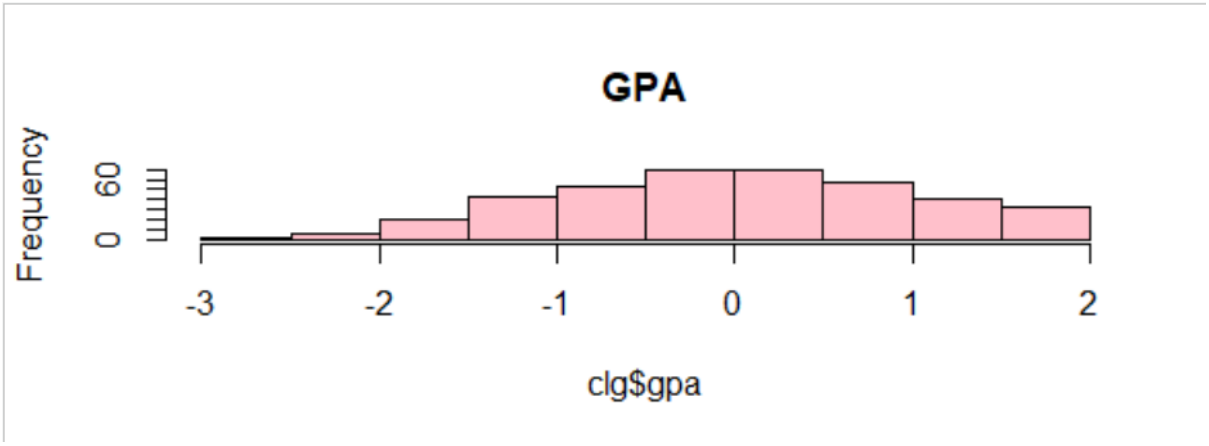
Distribution of GPA:



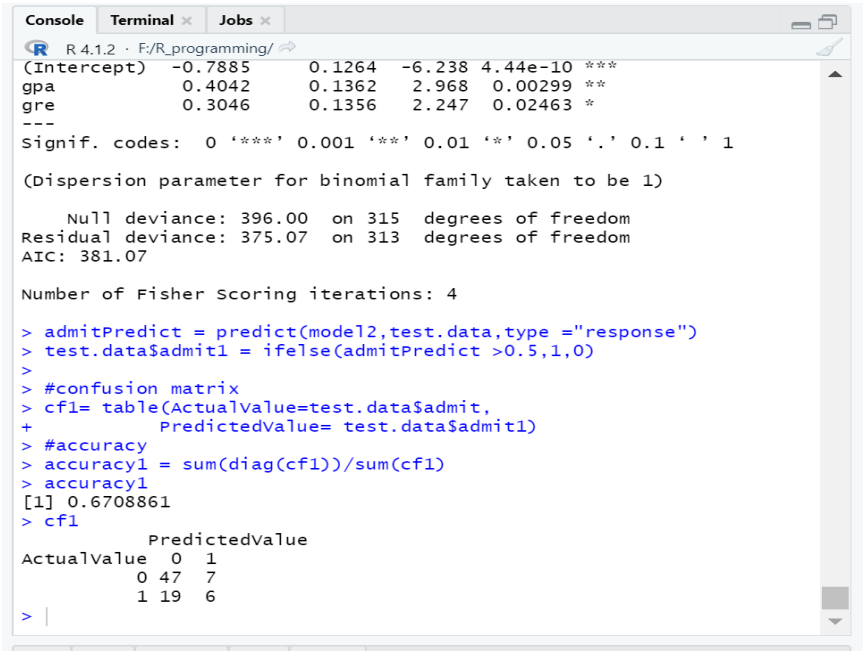
After scaling GRE:



After scaling GPA:

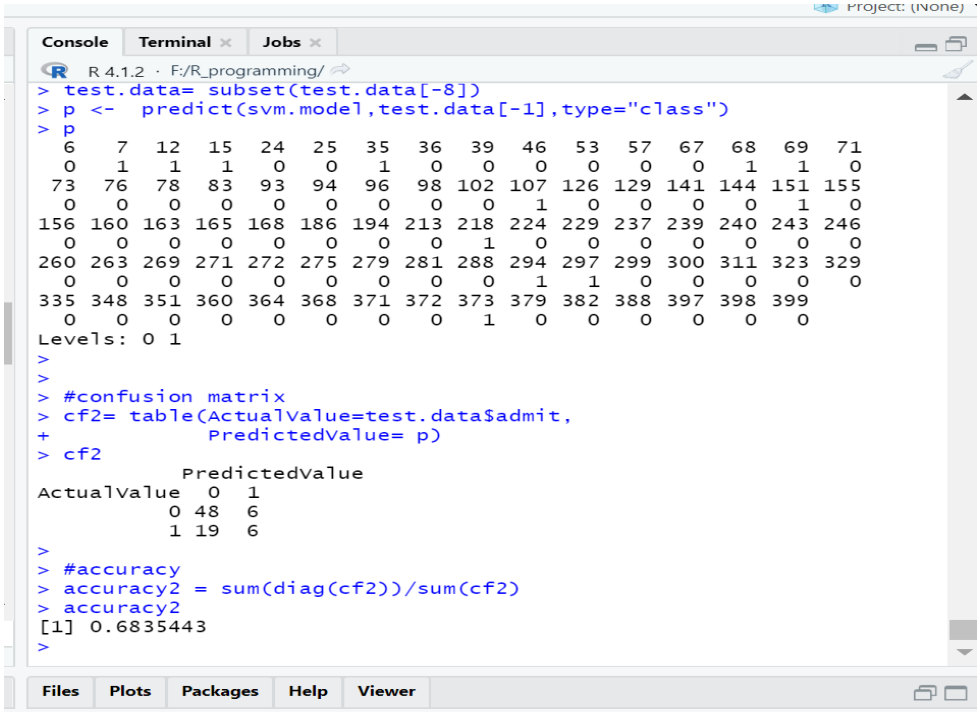


Logistic regression: Accuracy & confusion matrix



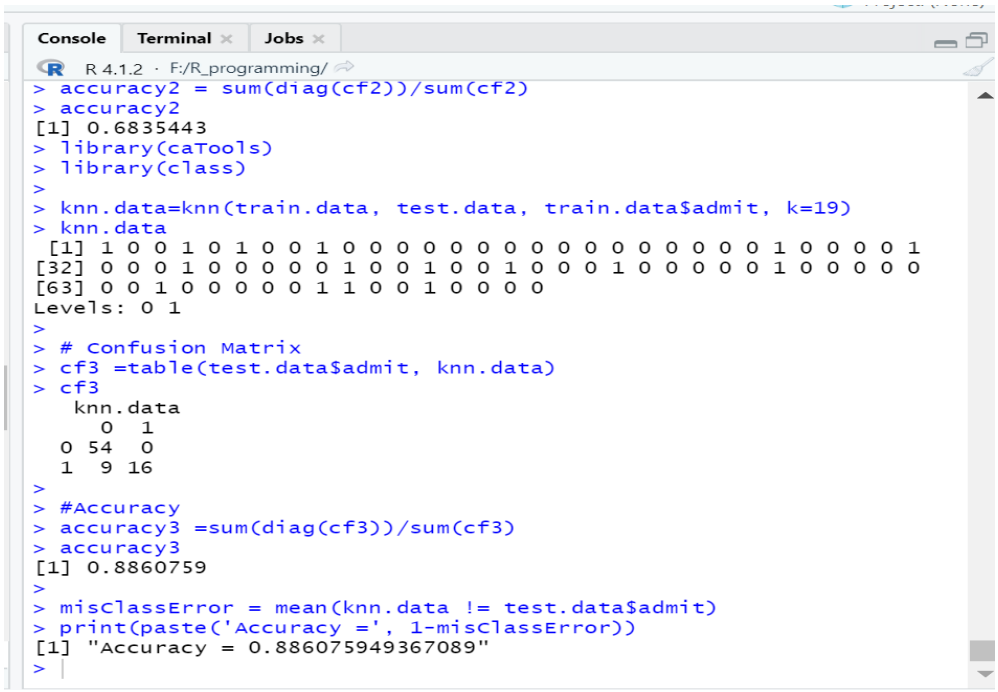
```
Console   Terminal   Jobs
R 4.1.2 · F:/R_programming/
(Intercept)  -0.7885     0.1264  -6.238 4.44e-10 ***
gpa           0.4042     0.1362   2.968  0.00299 **
gre           0.3046     0.1356   2.247  0.02463 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 396.00  on 315  degrees of freedom
Residual deviance: 375.07  on 313  degrees of freedom
AIC: 381.07

Number of Fisher Scoring iterations: 4

> admitPredict = predict(model2,test.data,type ="response")
> test.data$admit1 = ifelse(admitPredict >0.5,1,0)
>
> #confusion matrix
> cf1= table(ActualValue=test.data$admit,
+           PredictedValue= test.data$admit1)
> #accuracy
> accuracy1 = sum(diag(cf1))/sum(cf1)
> accuracy1
[1] 0.6708861
> cf1
           PredictedValue
ActualValue  0  1
          0 47  7
          1 19  6
>
```

SVM model : Accuracy and confusion matrix



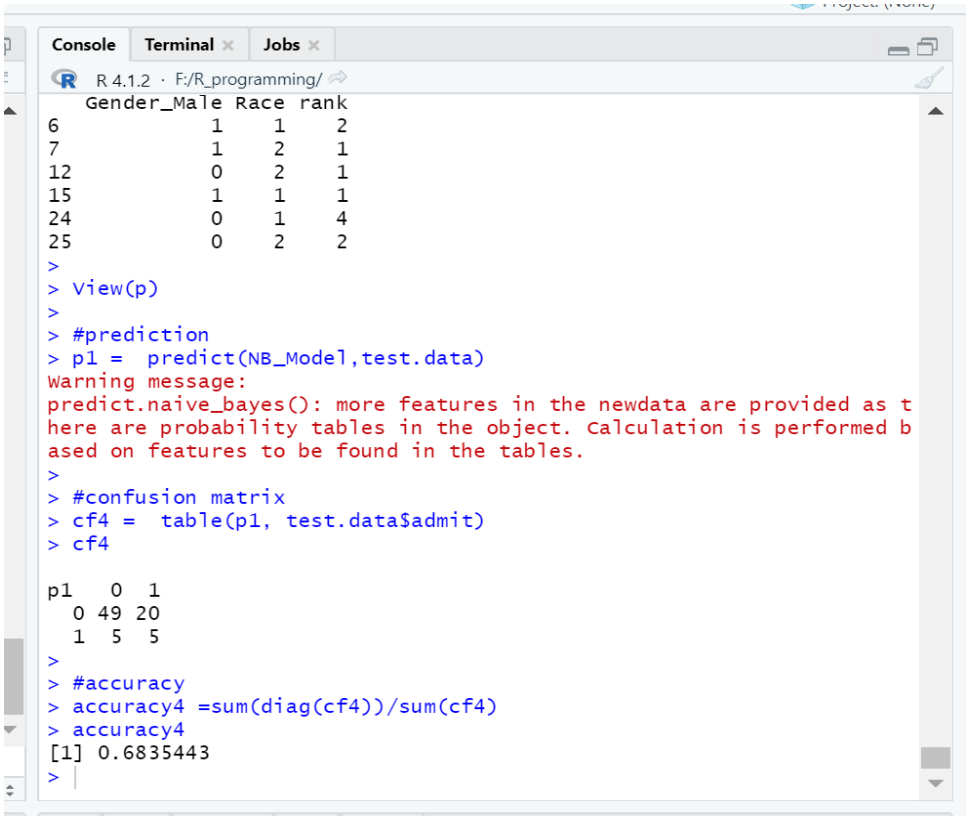
```
> test.data= subset(test.data[-8])
> p <-  predict(svm.model,test.data[-1],type="class")
> p
  6   7  12  15  24  25  35  36  39  46  53  57  67  68  69  71
  0   1   1   1   0   0   1   0   0   0   0   0   0   1   1   0
 73  76  78  83  93  94  96  98 102 107 126 129 141 144 151 155
  0   0   0   0   0   0   0   0   0   1   0   0   0   0   1   0
156 160 163 165 168 186 194 213 218 224 229 237 239 240 243 246
  0   0   0   0   0   0   0   0   1   0   0   0   0   0   0   0
260 263 269 271 272 275 279 281 288 294 297 299 300 311 323 329
  0   0   0   0   0   0   0   0   0   1   1   0   0   0   0   0
335 348 351 360 364 368 371 372 373 379 382 388 397 398 399
  0   0   0   0   0   0   0   0   1   0   0   0   0   0   0
Levels: 0 1
>
>
> #confusion matrix
> cf2= table(ActualValue=test.data$admit,
+            PredictedValue= p)
> cf2
           PredictedValue
ActualValue  0  1
          0 48  6
          1 19  6
>
> #accuracy
> accuracy2 = sum(diag(cf2))/sum(cf2)
> accuracy2
[1] 0.6835443
>
```

KNN model: Accuracy and confusion matrix



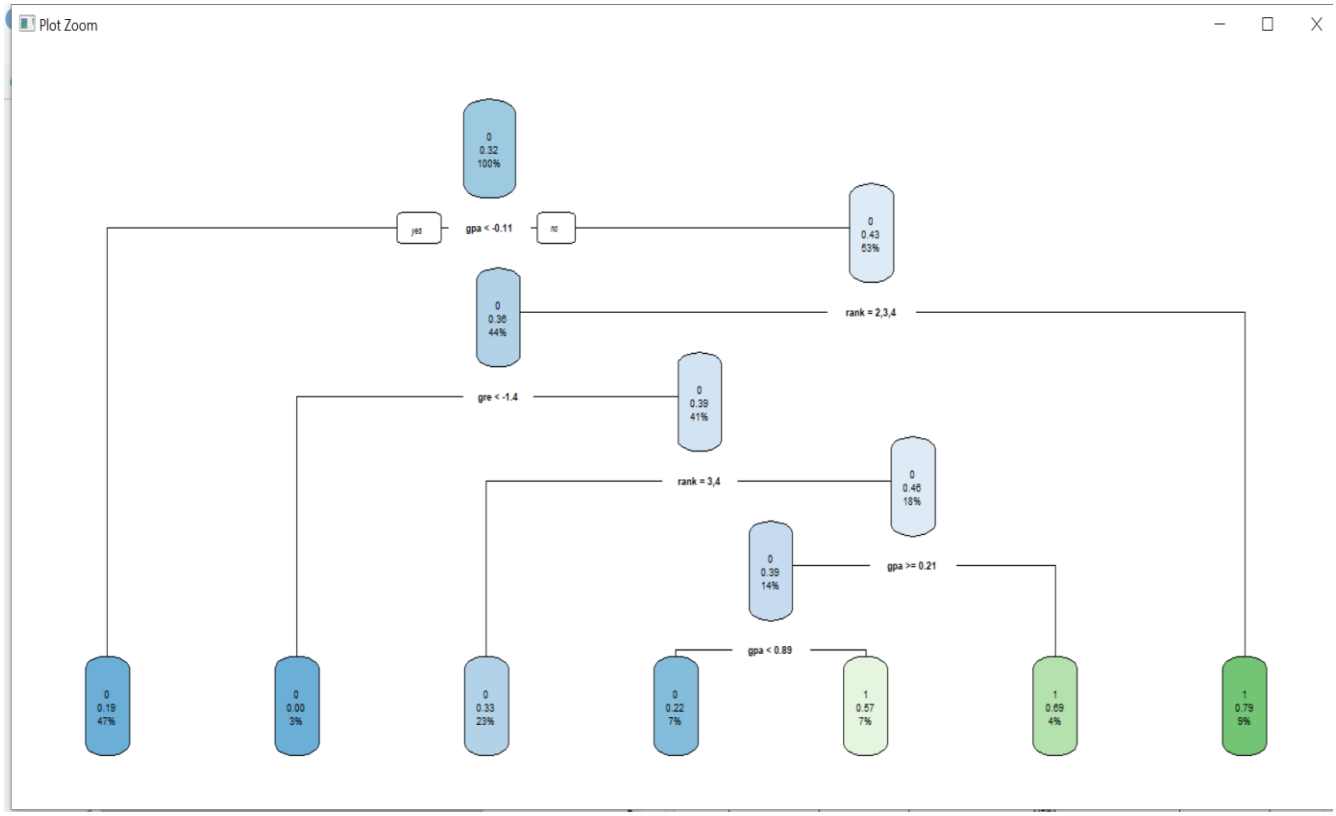
```
> accuracy2 = sum(diag(cf2))/sum(cf2)
> accuracy2
[1] 0.6835443
> library(caTools)
> library(class)
>
> knn.data=knn(train.data, test.data, train.data$admit, k=19)
> knn.data
 [1] 1 0 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1
[32] 0 0 0 1 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0
[63] 0 0 1 0 0 0 0 0 1 1 0 0 1 0 0 0 0
Levels: 0 1
>
> # Confusion Matrix
> cf3 =table(test.data$admit, knn.data)
> cf3
   knn.data
     0  1
  0 54  0
  1  9 16
>
> #Accuracy
> accuracy3 =sum(diag(cf3))/sum(cf3)
> accuracy3
[1] 0.8860759
>
> misClassError = mean(knn.data != test.data$admit)
> print(paste('Accuracy =', 1-misClassError))
[1] "Accuracy = 0.886075949367089"
>
```

Naïve bayes: Accuracy and confusion matrix

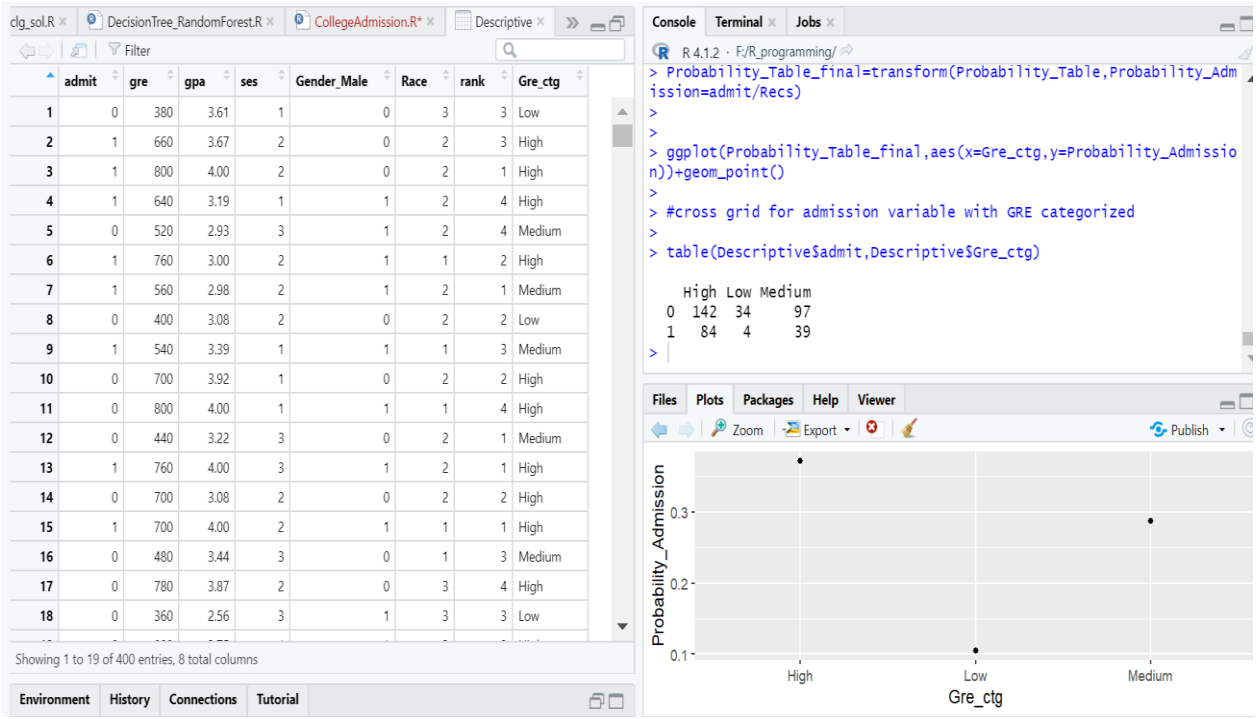


```
Console  Terminal  Jobs
R 4.1.2 · F:/R_programming/
   Gender_Male Race rank
6            1    1    2
7            1    2    1
12           0    2    1
15           1    1    1
24           0    1    4
25           0    2    2
>
> View(p)
>
> #prediction
> p1 =  predict(NB_Model,test.data)
Warning message:
predict.naive_bayes(): more features in the newdata are provided as t
here are probability tables in the object. Calculation is performed b
ased on features to be found in the tables.
>
> #confusion matrix
> cf4 =  table(p1, test.data$admit)
> cf4

p1   0  1
  0 49 20
  1  5  5
>
> #accuracy
> accuracy4 =sum(diag(cf4))/sum(cf4)
> accuracy4
[1] 0.6835443
>
```

Decision Tree:



Categorisation of data:



clg_sol.R | DecisionTree_RandomForest.R | CollegeAdmission.R* | Descriptive

Filter

	admit	gre	gpa	ses	Gender_Male	Race	rank	Gre_ctg
1	0	380	3.61	1	0	3	3	Low
2	1	660	3.67	2	0	2	3	High
3	1	800	4.00	2	0	2	1	High
4	1	640	3.19	1	1	2	4	High
5	0	520	2.93	3	1	2	4	Medium
6	1	760	3.00	2	1	1	2	High
7	1	560	2.98	2	1	2	1	Medium
8	0	400	3.08	2	0	2	2	Low
9	1	540	3.39	1	1	1	3	Medium
10	0	700	3.92	1	0	2	2	High
11	0	800	4.00	1	1	1	4	High
12	0	440	3.22	3	0	2	1	Medium
13	1	760	4.00	3	1	2	1	High
14	0	700	3.08	2	0	2	2	High
15	1	700	4.00	2	1	1	1	High
16	0	480	3.44	3	0	1	3	Medium
17	0	780	3.87	2	0	3	4	High
18	0	360	2.56	3	1	3	3	Low

Showing 1 to 19 of 400 entries, 8 total columns

Environment | History | Connections | Tutorial

Console | Terminal | Jobs

R 4.1.2 · F:/R_programming/

```
> Probability_Table_final=transform(Probability_Table,Probability_Admission=admit/Recs)
>
>
> ggplot(Probability_Table_final,aes(x=Gre_ctg,y=Probability_Admission))+geom_point()
>
> #cross grid for admission variable with GRE categorized
>
> table(Descriptive$admit,Descriptive$Gre_ctg)

   High Low Medium
  0  142  34     97
  1   84   4     39
>
```

Files | Plots | Packages | Help | Viewer

Zoom | Export | Publish