



---

# R PROGRAMMING

---

COLLEGE ADMISSION  
PROJECT5



JUNE 4, 2021  
[rayasalehalshammri@gmail.com](mailto:rayasalehalshammri@gmail.com)

There are no missing values:  
> sum(is.na(file))

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
file x f1 x Descriptive x Descriptivefile x
admit gre gpa ses Gender_Male Race rank
Showing 1 to 1 of 400 entries, 7 total columns
Console Terminal x Jobs x
R 4.1.0 - C:/Users/raja_/Documents/Downloads/
> file <- read.csv(file.choose())
> head(file)
  admit gre  gpa ses Gender_Male Race rank
1     0 380 3.61  1         0     3     3
2     1 660 3.67  2         0     2     3
3     1 800 4.00  2         0     2     1
4     1 640 3.19  1         1     2     4
5     0 520 2.93  3         1     2     4
6     1 760 3.00  2         1     1     2
> f1 <- file
> #number of records
> nrow(file)
[1] 400
> #column datatype
> sapply(file, class)
      admit      gre      gpa      ses Gender_Male
"integer" "integer" "numeric" "integer" "integer"
      Race      rank
"integer" "integer"
> #check if there are any missing values
> sum(is.na(file))
[1] 0
> #dataset summary
> summary(file)
      admit      gre      gpa      ses
Min.   :0.0000  Min.   :220.0  Min.   :2.260  Min.   :1.000
1st Qu.:0.0000  1st Qu.:520.0  1st Qu.:3.130  1st Qu.:1.000
Median :0.0000  Median :580.0  Median :3.395  Median :2.000
Mean   :0.3175  Mean   :587.7  Mean   :3.390  Mean   :1.992
3rd Qu.:1.0000  3rd Qu.:660.0  3rd Qu.:3.670  3rd Qu.:3.000
Max.   :1.0000  Max.   :800.0  Max.   :4.000  Max.   :3.000
 Gender_Male      Race      rank
Min.   :0.000  Min.   :1.000  Min.   :1.000
1st Qu.:0.000  1st Qu.:1.000  1st Qu.:2.000
Median :0.000  Median :2.000  Median :2.000
Mean   :0.475  Mean   :1.962  Mean   :2.485
3rd Qu.:1.000  3rd Qu.:3.000  3rd Qu.:3.000
Max.   :1.000  Max.   :3.000  Max.   :4.000
```

## Check the outlier, if any then remove:

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

file f1 Descriptive Descriptivefile

Filter

admit gre gpa ses Gender\_Male Race rank

Showing 1 to 1 of 400 entries, 7 total columns

Console Terminal Jobs

```
R 4.1.0 - C:/Users/raja/Documents/Downloads/
> #outlier
> iqrgre<-IQR(file$gre)
> iqrgre
[1] 140
> quantile(file$gre, na.rm = TRUE)
 0% 25% 50% 75% 100%
220 520 580 660 800
> maxgre<-660+1.5*iqrgre
> maxgre
[1] 870
> mingre<-520-1.5*iqrgre
> mingre
[1] 310
> #points above the upperinner
> print(which(file$gre > maxgre))
integer(0)
> #points below the lowerinner
> print(which(file$gre < mingre))
integer(0)
> #points below the lowerinner
> print(which(file$gre < mingre))
[1] 72 180 305 316
> file<-file[-c(72,180,305,316),]
> nrow(file)
[1] 396
> #outlier for gpa variable
> iqr_gpa<-IQR(file$gpa,na.rm = TRUE)
> iqr_gpa
[1] 0.54
> #outlier for gpa variable
> iqr_gpa<-IQR(file$gpa)
> iqr_gpa
[1] 0.54
> quantile(file$gpa,na.rm = TRUE)
 0% 25% 50% 75% 100%
2.26 3.13 3.40 3.67 4.00
> maxgpa<-3.67+1.5*iqr_gpa
> maxgpa
[1] 4.48
> mingpa<-3.13-1.5*iqr_gpa
> mingpa
[1] 2.32
> #points above the upperinner
> print(which(file$gpa > maxgpa))
integer(0)
```

No outlier  
4 outlier  
<< remove

Console Terminal Jobs

```
R 4.1.0 - C:/Users/raja/Documents/Downloads/
> print(which(file$gpa > maxgpa))
integer(0)
> #points below the lowerinner
> print(which(file$gpa < mingpa))
[1] 288
> file<-file[-c(288),]
> nrow(file)
[1] 395
```

No outlier  
1 outlier  
<< remove

```
> #convert columns datatype to factor
> clo<-c("admit", "ses", "Gender_Male", "Race", "rank")
> fi[col]<- lapply(fi[col], factor)
Error in [.default](fi, col) : invalid subscript type 'closure'
> fi[clo]<- lapply(fi[clo], factor)
> sapply(fi, class)
      admit      gre      gpa      ses Gender_Male
"factor"  "integer" "numeric" "factor"  "factor"
"factor"  "rank"
"factor"  "factor"
> summary(fi)
admit      gre      gpa      ses      Gender_Male Race
0:273   Min.   :220.0   Min.   :2.260   1:132   0:210   1:143
1:127   1st Qu.:520.0   1st Qu.:3.130   2:139   1:190   2:129
        Median :580.0   Median :3.395   3:129           3:128
        Mean   :587.7   Mean   :3.390
        3rd Qu.:660.0   3rd Qu.:3.670
        Max.   :800.0   Max.   :4.000

rank
1: 61
2:151
3:121
4: 67
```

## Data splitting:

```
Console Terminal Jobs
R 4.1.0 · C:/Users/raja/Documents/Downloads/
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/caTools_1.18.2.zip'
Content type 'application/zip' length 316415 bytes (308 KB)
downloaded 308 KB

package 'bitops' successfully unpacked and MD5 sums checked
package 'caTools' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\raja\AppData\Local\Temp\Rtmp0oacqi\downloaded_packages
> library("caTools")
>
> file[,2:3]<-scale(file[,2:3])
> split<-sample.split(file$admit,SplitRatio = .75)
> train<-subset(file,split==T)
> test<-subset(file,split==F)
```

## Logistic regression:

```
> #logistic regression
> logi<-glm(admit~.,train,family="binomial") #all variable
> summary(logi)

Call:
glm(formula = admit ~ ., family = "binomial", data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6220  -0.8496  -0.5842   1.0101   2.3616

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.6186    0.6196   2.612 0.008988 **
gre           0.2445    0.1495   1.635 0.102014
gpa           0.5449    0.1562   3.488 0.000486 ***
ses          -0.2021    0.1688  -1.197 0.231133
Gender_Male  -0.1075    0.2720  -0.395 0.692718
Race         -0.2273    0.1653  -1.375 0.169075
rank         -0.6493    0.1553  -4.181 2.9e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 370.01  on 295  degrees of freedom
Residual deviance: 323.28  on 289  degrees of freedom
AIC: 337.28

Number of Fisher Scoring iterations: 4
```

## Second logistic model by removing insignificant variables:

```
Console Terminal Jobs
R 4.1.0 · C:/Users/raja/Documents/Downloads/
> #above model gre,ses,gender..etc not significant
> #build a new model with only gpa and rank variables
> logi2<-glm(admit~gpa+rank,train,family="binomial") #all variable
> summary(logi2)

Call:
glm(formula = admit ~ gpa + rank, family = "binomial", data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6576  -0.8952  -0.5888   1.0275   2.3716

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.7757    0.3818   2.032 0.0422 *
gpa           0.6051    0.1445   4.186 2.84e-05 ***
rank         -0.6648    0.1536  -4.329 1.50e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 370.01  on 295  degrees of freedom
Residual deviance: 329.64  on 293  degrees of freedom
AIC: 335.64

Number of Fisher Scoring iterations: 4
```

So, we will use the first model

## Accuracy of logistic model:

```
> predictval1<-predict(logi,test,type = "response")
> test$pred_admit1<- ifelse(predictval1>0.5,1,0)
> #confusion matrix
> confmat1<-table(predicted=test$pred_admit1,actual=test$admit)
> confmat1
      actual
predicted 0 1
0      57 24
1      10  8
> #accuracy
> acc1<-sum(diag(confmat1))/sum(confmat1)
> acc1
[1] 0.6565657
```

## SVM model:

```
Console Terminal Jobs
R 4.1.0 · C:/Users/raja_/Documents/Downloads/

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/proxy_0.4-25.zip'
Content type 'application/zip' length 243972 bytes (238 KB)
downloaded 238 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/e1071_1.7-7.zip'
Content type 'application/zip' length 1022091 bytes (998 KB)
downloaded 998 KB

package 'proxy' successfully unpacked and MD5 sums checked
package 'e1071' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\raja\AppData\Local\Temp\Rtmp0oacqi\downloaded_packages
> library("e1071")
> svmclf=svm(admit~.,train,type='c-classification',kernel='linear')
Error in svm.default(x, y, scale = scale, ..., na.action = na.action) :
  wrong type specification!
> svm_clf = svm(admit ~ .,train, type = 'c-classification', kernel = 'linear')
Error in svm.default(x, y, scale = scale, ..., na.action = na.action) :
  wrong type specification!
> svm_clf = svm(admit ~ ., train, type = "c-classification", kernel = "linear")
> summary(svm_clf)

Call:
svm(formula = admit ~ ., data = train, type = "c-classification", kernel = "linear")

Parameters:
  SVM-Type:  c-classification
 SVM-Kernel: linear
      cost:  1

Number of Support Vectors: 190

( 94 96 )

Number of Classes: 2

Levels:
0 1
```

## Accuracy of SVM:

```
> #accuracy of svm
> predictval12<-predict(svm_clf,test[-1])
> predictval12
 1 11 14 16 26 29 31 35 37 38 52 60 61 63 68 70 74 82 94 95 102
0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0
104 109 113 114 116 118 121 126 127 137 138 139 144 150 151 159 161 172 176 179 192
0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0
197 198 202 203 205 206 212 214 215 216 223 233 236 238 243 247 248 251 253 256 260
0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
266 269 270 274 276 277 279 285 286 288 294 296 304 312 315 317 320 321 324 325 332
0 0 0 1 0 0 0 0 0 0 1 0 1 1 0 0 0 0 0 0 0
339 342 358 359 369 370 373 375 382 385 386 391 394 396 400
0 0 0 0 1 1 0 0 0 0 0 0 0 1 0
Levels: 0 1
> #confusion matrix
> confmat2<-table(predicted=predictval12,actual=test$admit)
> confmat2
      actual
predicted 0 1
      0 58 26
      1 9 6
> #accuracy
> acc2<-sum(diag(confmat2))/sum(confmat2)
> acc2
[1] 0.6464646
```

## Decision tree model:

```

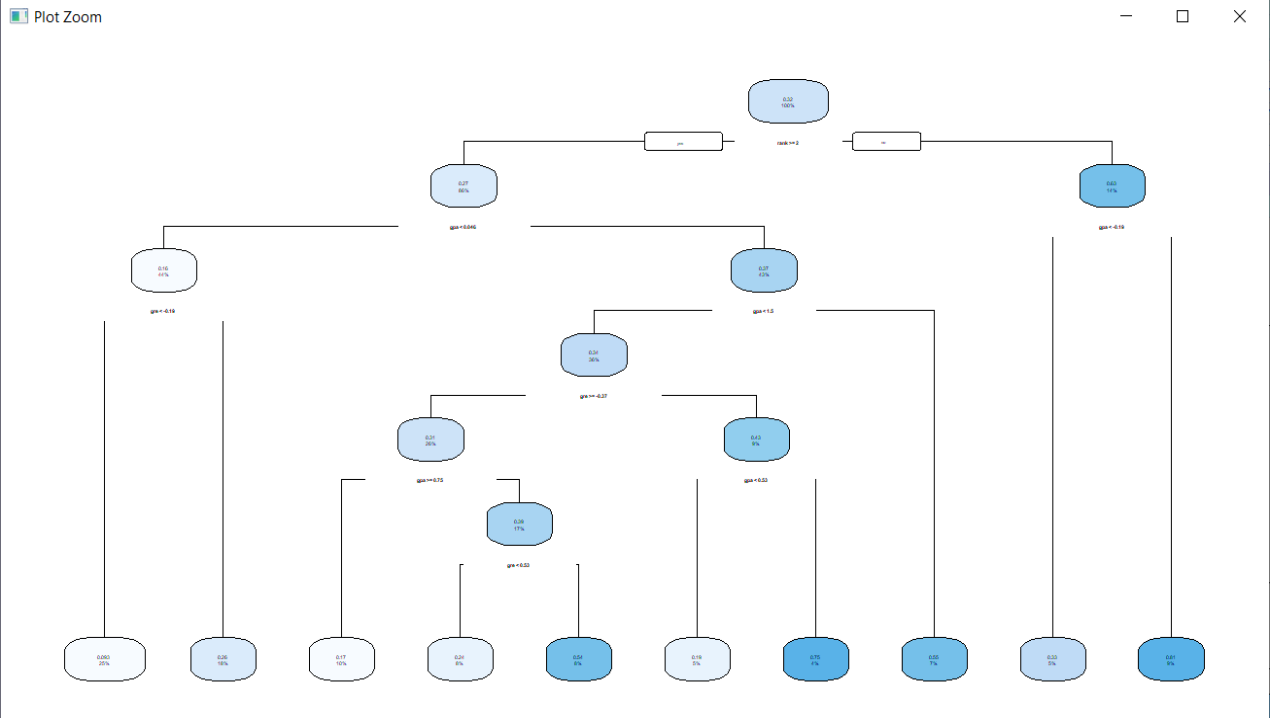
The downloaded binary packages are in
  C:\Users\raya\AppData\Local\Temp\Rtmp0oacqi\downloaded_packages
> library("rpart.plot")
> nrow(train)
[1] 296
> nrow(test)
[1] 99
> 0.03*nrow(train)
[1] 8.88
> 0.03*nrow(train)*3
[1] 26.64
> r.cntrl<-rpart.control(minsplit = 26,minbucket = 9,xval = 5)
Error: object 'r.cntrl' not found
> r.cntrl<-rpart.control(minsplit = 26,minbucket = 9,xval = 5)
> declf<-rpart(admit~.,control = r.cntrl,data = train)
> rpart.plot(declf)
> summary(declf)
Call:
rpart(formula = admit ~ ., data = train, control = r.cntrl)
n= 296

      CP nsplit rel error   xerror   xstd
1 0.07435522      0 1.0000000 1.0165291 0.04649306
2 0.04391390      1 0.9256448 0.9755459 0.05621646
3 0.03336626      2 0.8817309 0.9663791 0.05830602
4 0.01670791      3 0.8483646 0.9926498 0.06142646
5 0.01527016      6 0.7982409 1.0623064 0.06757967
6 0.01347430      8 0.7677006 1.0747693 0.07028684
7 0.01000000      9 0.7542263 1.0755132 0.07082941

Variable importance
      gpa      rank      gre Gender_Male      Race      ses
      45       27       21          3         2         2

Node number 1: 296 observations,      complexity param=0.07435522
mean=0.3175676, MSE=0.2167184
left son=2 (255 obs) right son=3 (41 obs)
Primary splits:
      rank < 1.5      to the right, improve=0.074355220, (0 missing)

```



### Accuracy of decision tree:

```

> predicted_val3 <- predict(dec_c1f.test[-1], type="class")
> predicted_val3
1 11 14 16 26 29 31 35 37 38 52 60 61 63 68 70 74
1 1 0 0 1 0 1 0 0 0 0 0 0 0 0 0 1 1
82 94 95 102 104 109 113 114 116 118 121 126 127 137 138 139 144
0 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 0 0
150 151 159 161 172 176 179 192 197 198 202 203 205 206 212 214 215
0 1 1 0 0 0 0 1 0 0 0 1 1 0 0 0 0 0
216 223 233 236 238 243 247 248 251 253 256 260 266 269 270 274 276
0 0 0 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0
277 279 285 286 288 294 296 304 312 315 317 320 321 324 325 332 339
1 0 0 0 0 1 0 1 1 0 0 0 0 0 0 0 0
342 358 359 369 370 373 375 382 385 386 391 394 396 400
0 0 0 1 0 0 0 0 0 0 0 0 0 1 0
Levels: 0 1
>
> #confusion matrix
> conf_mat3<-table(predicted=predicted_val3,actaul=test$admit)
> conf_mat3
      actaul
predicted 0 1
0 50 22
1 17 10
> #accuracy
> accuracy3<-sum(diag(conf_mat3))/sum(conf_mat3)
> accuracy3
[1] 0.6060606

```

## Knn and accuracy:

```
Console Terminal Jobs
R 4.1.0 · C:/Users/raja_/Documents/Downloads/

> #knn
> library("class")
> knn = knn(train, test[-1], train$admit, k=19)
> knn
[1] 0 1 1 0 1 1 0 0 0 0 0 0 1 0 0 1 0 1 1 1 0 0 0 0 0 0 1 0 0 0 0
[31] 0 1 1 1 1 1 0 1 1 1 1 1 0 0 1 0 0 1 1 1 1 1 0 0 1 0 1 1 0 1
[61] 0 0 0 0 1 0 1 0 0 1 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1 0 0 1
[91] 0 0 1 0 0 1 0 0 0
Levels: 0 1
> #conf matrix
> confmat4<-table(predicted=knn,actual=test$admit)
> confmat4
      actual
predicted 0  1
      0 43 15
      1 24 17
> #accuracy
> acc4<-sum(diag(confmat4))/sum(confmat4)
> acc4
[1] 0.6060606
```

## Naïve Bayes:

```
Console Terminal Jobs
R 4.1.0 · C:/Users/raja_/Documents/Downloads/

> #naive bayes
> nb<-naiveBayes(admit~.,data=train)
> nb

Naive Bayes Classifier for Discrete Predictors

call:
naiveBayes.default(x = x, y = y, laplace = laplace)

A-priori probabilities:
Y
      0      1
0.6824324 0.3175676

Conditional probabilities:
gre
Y      [,1]      [,2]
0 -0.1406362 1.0036519
1  0.2862818 0.9187608

gpa
Y      [,1]      [,2]
0 -0.1601170 0.9929025
1  0.3759513 0.8978631

ses
Y      [,1]      [,2]
0  2.054455 0.7992565
1  1.914894 0.8251370

Gender_Male
Y      [,1]      [,2]
0  0.4702970 0.5003570
1  0.4680851 0.5016559

Race
Y      [,1]      [,2]
0  2.009901 0.8103196
1  1.893617 0.8356066

rank
Y      [,1]      [,2]
0  2.688119 0.8959597
1  2.159574 0.9424652
```

## Accuracy of Naïve Bayes:

```
Console Terminal Jobs
R 4.1.0 · C:/Users/raja_/Documents/Downloads/

> predictval15<-predict(nb,test[-1], type = "class")
> predictval15
[1] 0 1 0 0 1 1 0 0 1 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1
[31] 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
[61] 1 0 1 0 0 0 1 0 0 0 0 0 0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 1 1
[91] 0 0 0 0 0 0 1 1 0
Levels: 0 1
> #conf matrix
> confmat5<-table(predicted=predictval15,actual=test$admit)
> confmat5
      actual
predicted 0  1
      0 53 23
      1 14  9
> #accuracy
> acc5<-sum(diag(confmat5))/sum(confmat5)
> acc5
[1] 0.6262626
```

## Categorize the average of grade point into High, Medium, and Low

file x f1 x Descriptive x

Filter

	admit	gre	gpa	ses	Gender_Male	Race	rank	GreLevels
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

Showing 1 to 12 of 400 entries, 8 total columns

Console Terminal x Jobs x

```
R 4.1.0 · C:/Users/raja/Documents/Downloads/
> #Categorize the average of grade point into High, Medium, and Low
> Descriptive = transform(f1,GreLevels=ifelse(gre<440,"Low",ifelse(gre<580,"Medium","High"))
> view(Descriptive)
> sum_desc= aggregate(admit~GreLevels,Descriptive,FUN = sum)
> length_desc=aggregate(admit~GreLevels,Descriptive,FUN=length)
> Probability_Table = cbind(Sum_Desc,Recs=length_desc[,2])
> Probability_Table_final = transform(Probability_Table,Probability_Admission =
+ admit/Recs)
> Probability_Table_final
  GreLevels admit Recs Probability_Admission
1      High    84  226          0.3716814
2       Low     4   38          0.1052632
3    Medium    39  136          0.2867647
> library("ggplot2")
Registered 53 methods overwritten by 'ggplot2':
method      from
[.quosures  rlang
c.quosures  rlang
print.quosures rlang
> ggplot(Probability_Table_final,aes(x=GreLevels,y=Probability_Admission))+geom_point()
> table(Descriptive$admit,Descriptive$GreLevels)

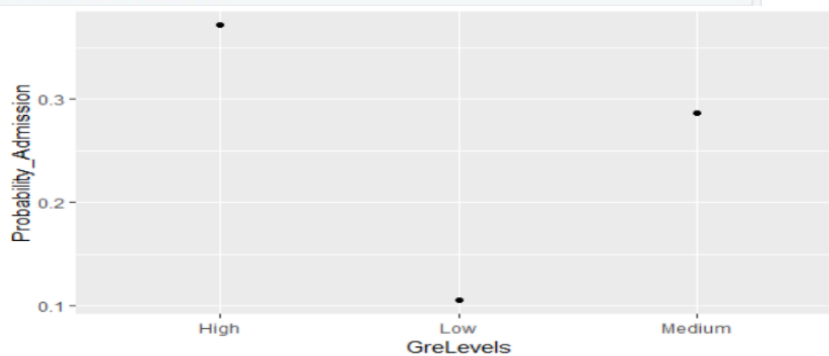
      High Low Medium
0    142  34    97
1     84   4    39
```

file x f1 x Descriptive x

Filter

	admit	gre	gpa	ses	Gender_Male	Race	rank	GreLevels
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
19	0	800	3.75	1	1	3	2	High
20	1	540	3.81	1	0	3	1	Medium
21	0	500	3.17	3	0	2	3	Medium
22	1	660	3.63	1	0	1	2	High
23	0	600	2.82	1	0	3	4	High

Showing 1 to 23 of 400 entries, 8 total columns





### Source code:

```
file <- read.csv(file.choose())
head(file)
f1 <- file
nrow(file)
sapply(file, class)
sum(is.na(file))
summary(file)
iqrgre<-IQR(file$gre)
iqrgre
quantile(file$gre, na.rm = TRUE)
maxgre<-660+1.5*iqrgre
mingre<-520-1.5*iqrgre
mingre
print(which(file$gre > maxgre))
print(which(file$gre > maxgre))
print(which(file$gre < mingre))
file<-file[-c(72,180,305,316),]
nrow(file)
iqrgpa<-IQR(file$gpa,na.rm = TRUE)
iqrgpa
iqrgpa<-IQR(file$gpa)
iqrgpa
quantile(file$gpa,na.rm = TRUE)
maxgpa<-3.67+1.5*iqrgpa
maxgpa
mingpa<-3.13-1.5*iqrgpa
mingpa
print(which(file$gpa > maxgpa))
print(which(file$gpa < mingpa))
file<-file[-c(288),]
nrow(file)
clo<-c("admit", "ses", "Gender_Male", "Race", "rank")
f1[clo]<- lapply(f1[clo], factor)
sapply(f1, class)
summary(f1)
set.seed(0)
install.packages('caTools')
library("caTools")
file[,2:3]<-scale(file[,2:3])
split<-sample.split(file$admit,SplitRatio = .75)
train<-subset(file,split==T)
test<-subset(file,split==F)
```

```

logi<-glm(admit~.,train,family="binomial") #all variable
summary(logi)
logi2<-glm(admit~gpa+rank,train,family="binomial") #all variable
summary(logi2)
predictval1<-predict(logi,test,type = "response")
test$pred_admit1<- ifelse(predictval1>0.5,1,0)
confmat1<-table(predicted=test$pred_admit1,actual=test$admit)
confmat1
acc1<-sum(diag(confmat1))/sum(confmat1)
acc1
install.packages("e1071")
library("e1071")
svm_clf = svm(admit ~ ., train, type = "C-classification", kernel = "linear")
summary(svm_clf)
predictval12<-predict(svm_clf,test[-1])
predictval12
confmat2<-table(predicted=predictval12,actual=test$admit)
confmat2
acc2<-sum(diag(confmat2))/sum(confmat2)
acc2
install.packages("rpart.plot")
library("rpart.plot")
nrow(train)
nrow(test)
0.03*nrow(train)
0.03*nrow(train)*3
r.cntrl<-rpart.control(minsplit = 26,minbucket = 9,xval = 5)
declf<-rpart(admit~.,control = r.cntrl,data = train)
rpart.plot(declf)
summary(declf)
predictval13 <- predict(declf, test[-1], type = "class")
predictval13
confmat3<-table(predicted=predictval12,actual=test$admit)
confmat3
acc3<-sum(diag(confmat3))/sum(confmat3)
acc3
library("class")
knn = knn(train, test[-1], train$admit, k=19)
knn
confmat4<-table(predicted=knn,actual=test$admit)
confmat4
acc4<-sum(diag(confmat4))/sum(confmat4)
acc4
nb<-naiveBayes(admit~.,data=train)

```

```

nb
predictval15<-predict(nb,test[-1], type = "class")
predictval15
confmat5<-table(predicted=predictval15,actual=test$admit)
confmat5
acc5<-sum(diag(confmat5))/sum(confmat5)
acc5
Descriptive =
transform(f1,Grelevels=ifelse(gre<440,"Low",ifelse(gre<580,"Medium","High")))
View(Descriptive)
sum_desc= aggregate(admit~Grelevels,Descriptive,FUN = sum)
install.packages("ggplot2")
length_desc= aggregate(admit~Grelevels,Descriptivefile,FUN = length)
probabilitytab= cbind(sum_desc,Recs=length_desc[,2])
probabilitytabfinal= transform(probabilitytab,probabilityadmission= + admit/Recs)
probabilitytabfinal
library("ggplot2")
ggplot(probabilitytabfinal,aes(x=Grelevels,y= probabilityadmission))+geom_point()
table(Descriptive$admit, Descriptive$Grelevels)

```