SSLC (Karnataka) DATA ANALYSIS

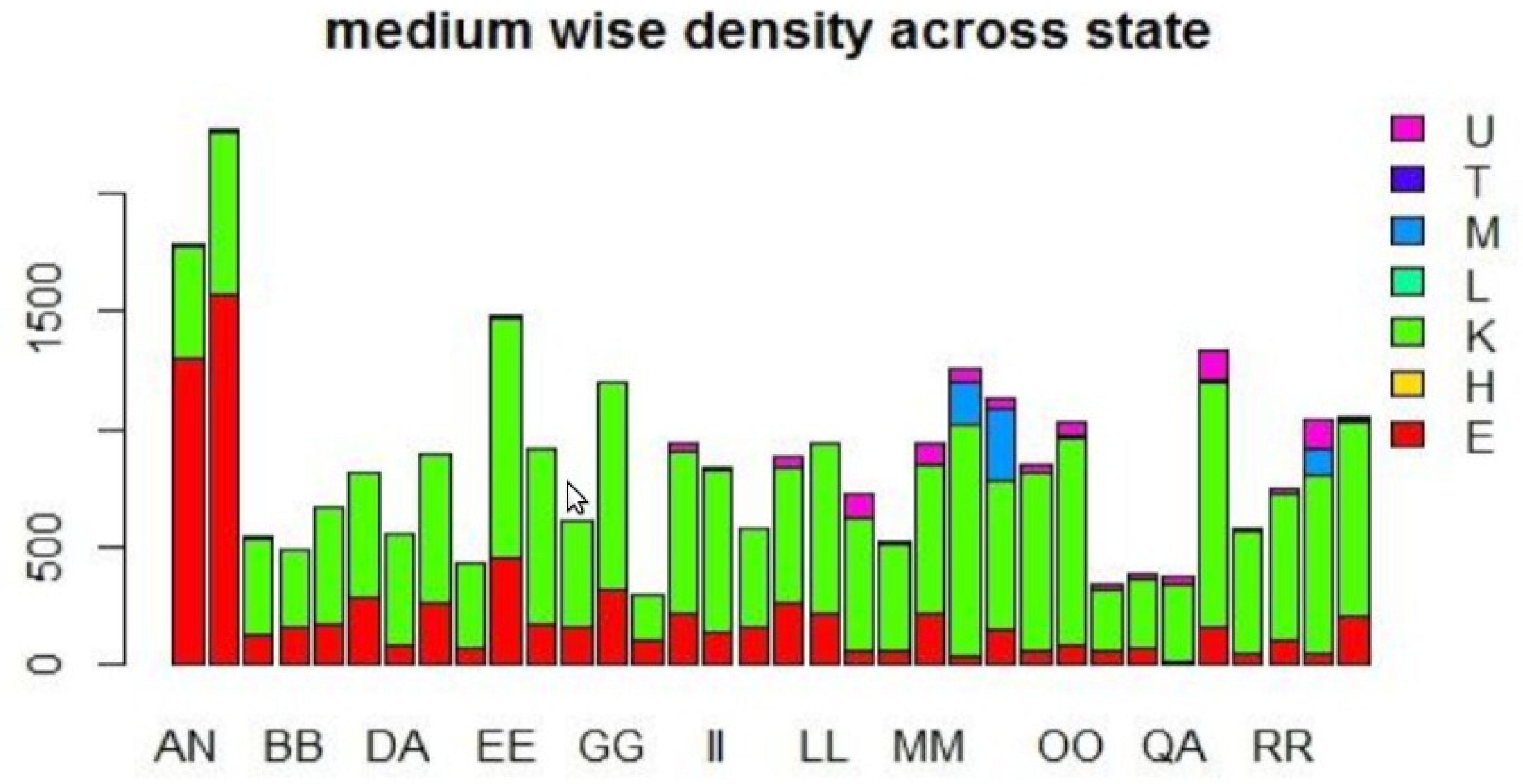
Group 21
Disha Shah(MT201431)
Roopa Gupta(MT2014096)
Shweta Mishra(MT2014116)
Vinita Goyal(MT2014138)

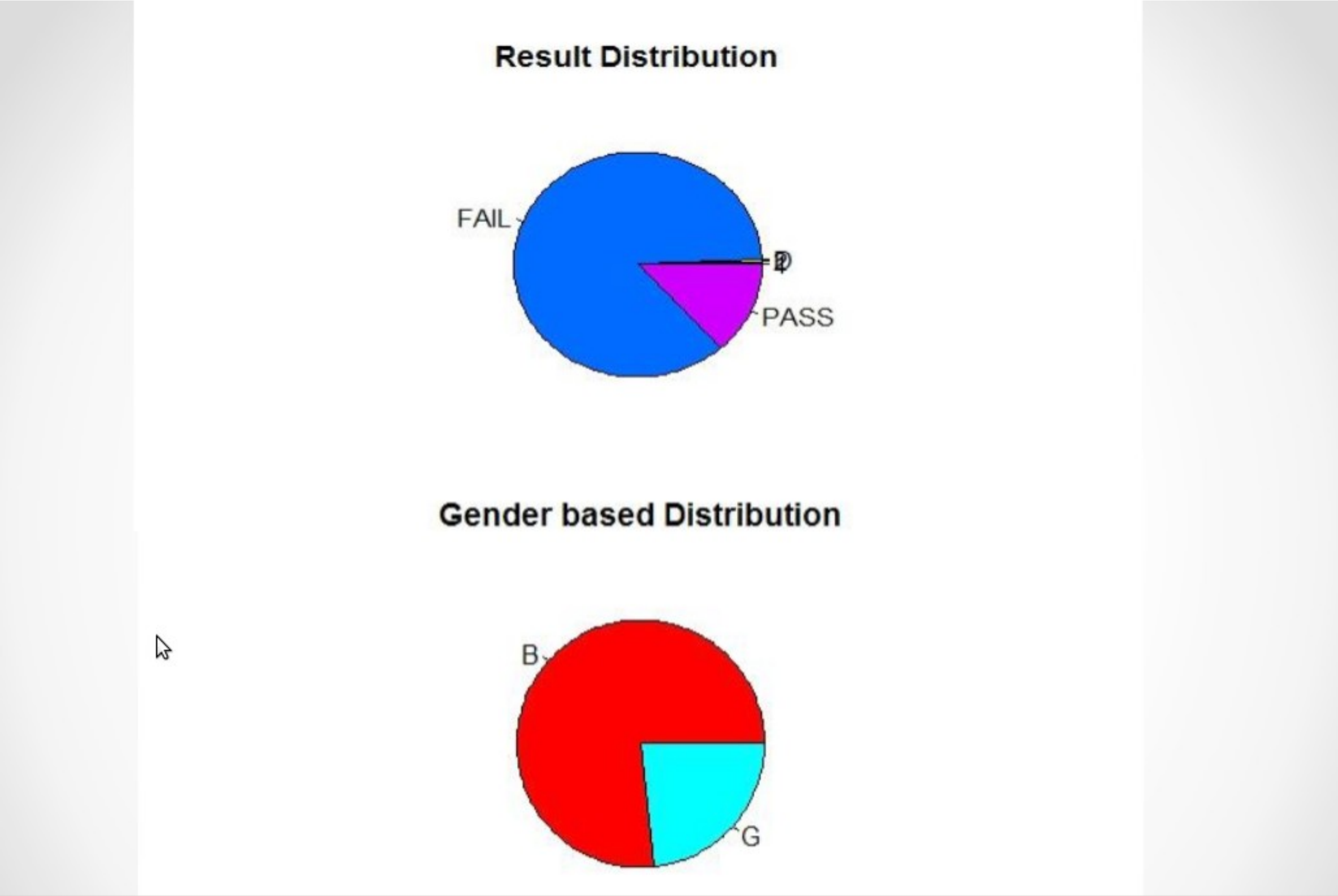
Data Preparation

- Removed * from marks (eg *46 is replaced by 46)
- Performed scaling on L1_MARKS
- Updated TOTAL_MARKS
- Replaced 888 with 0

Data Exploration

Medium wise distribution across districts

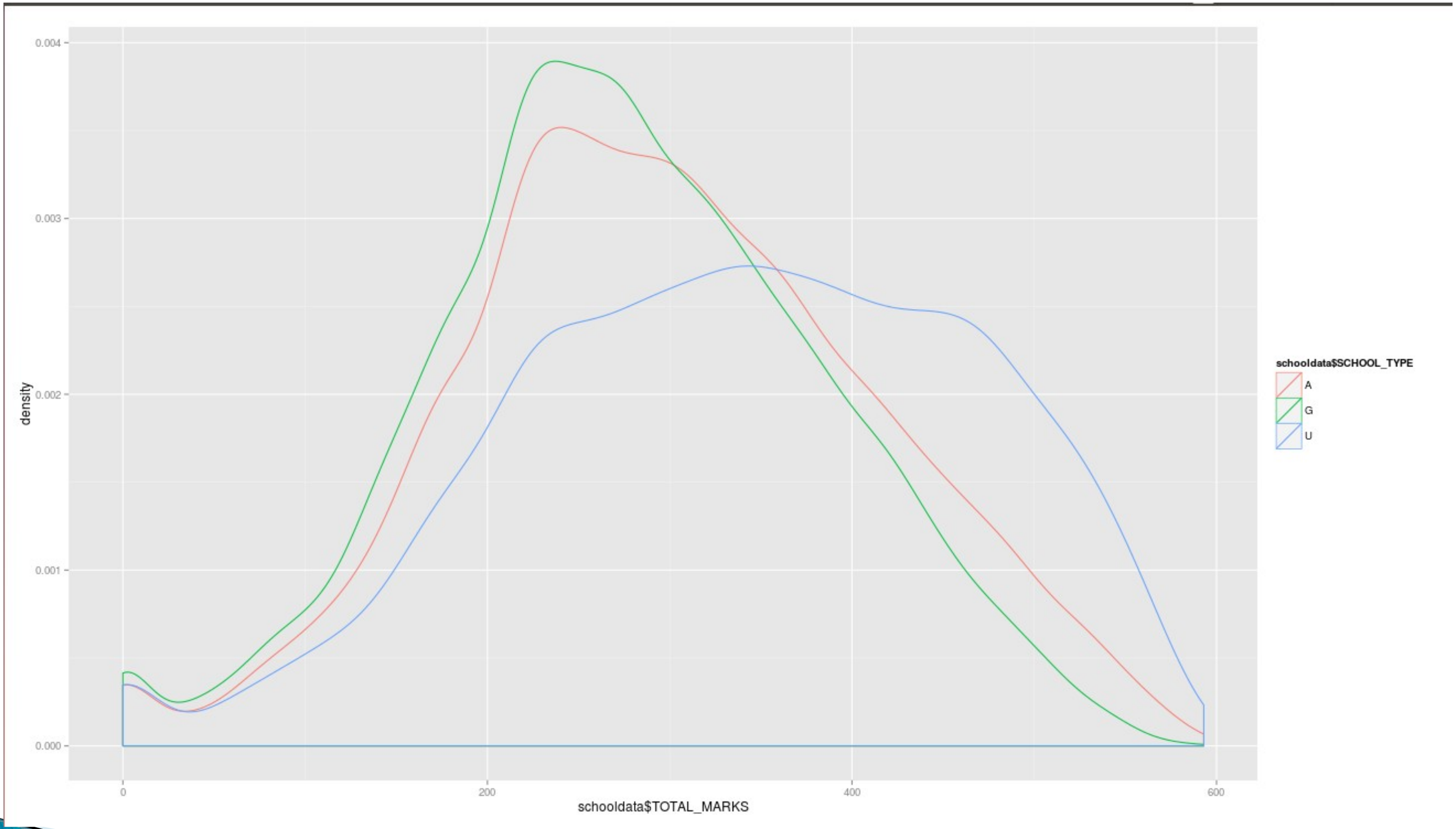




Result Distribution

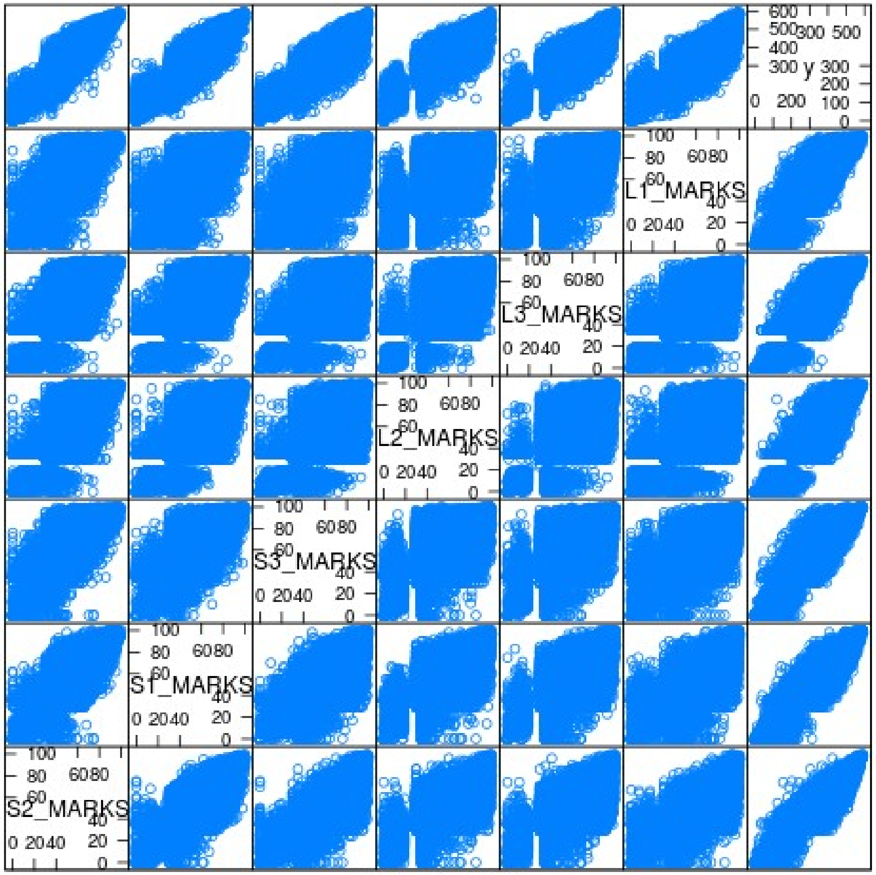
Gender based Distribution

Density plot



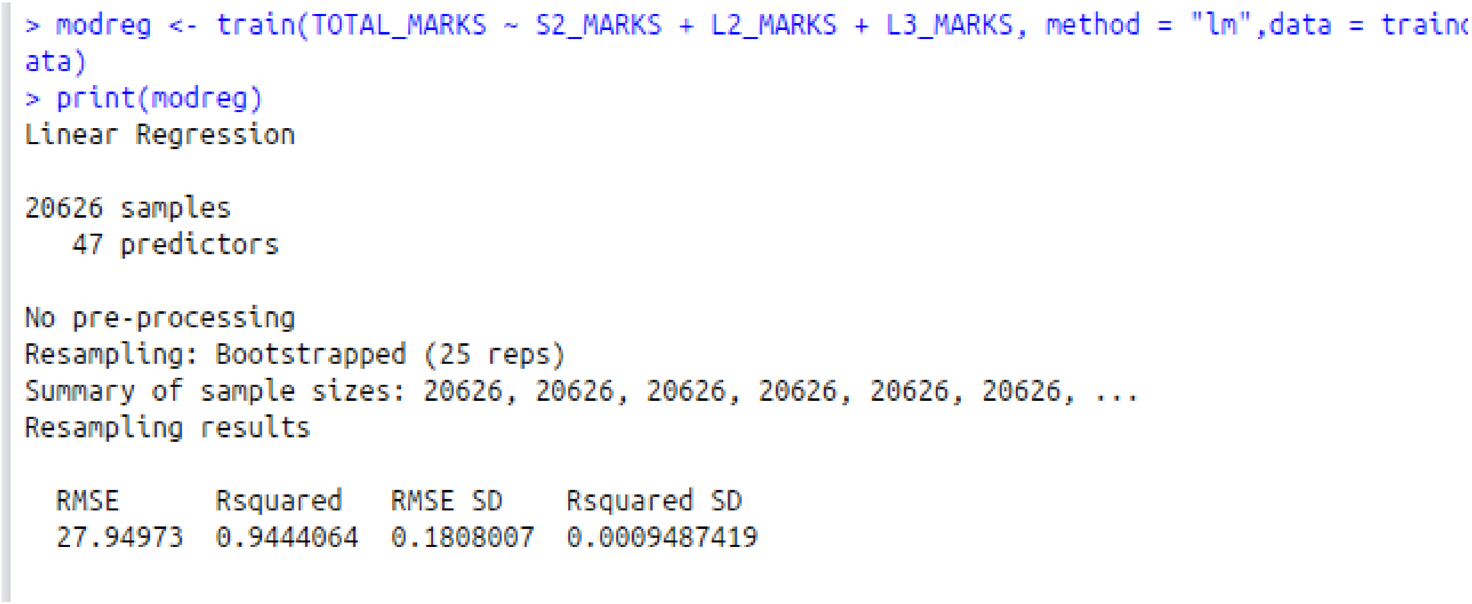
CLASSIFICATION/PREDICTION

Regression



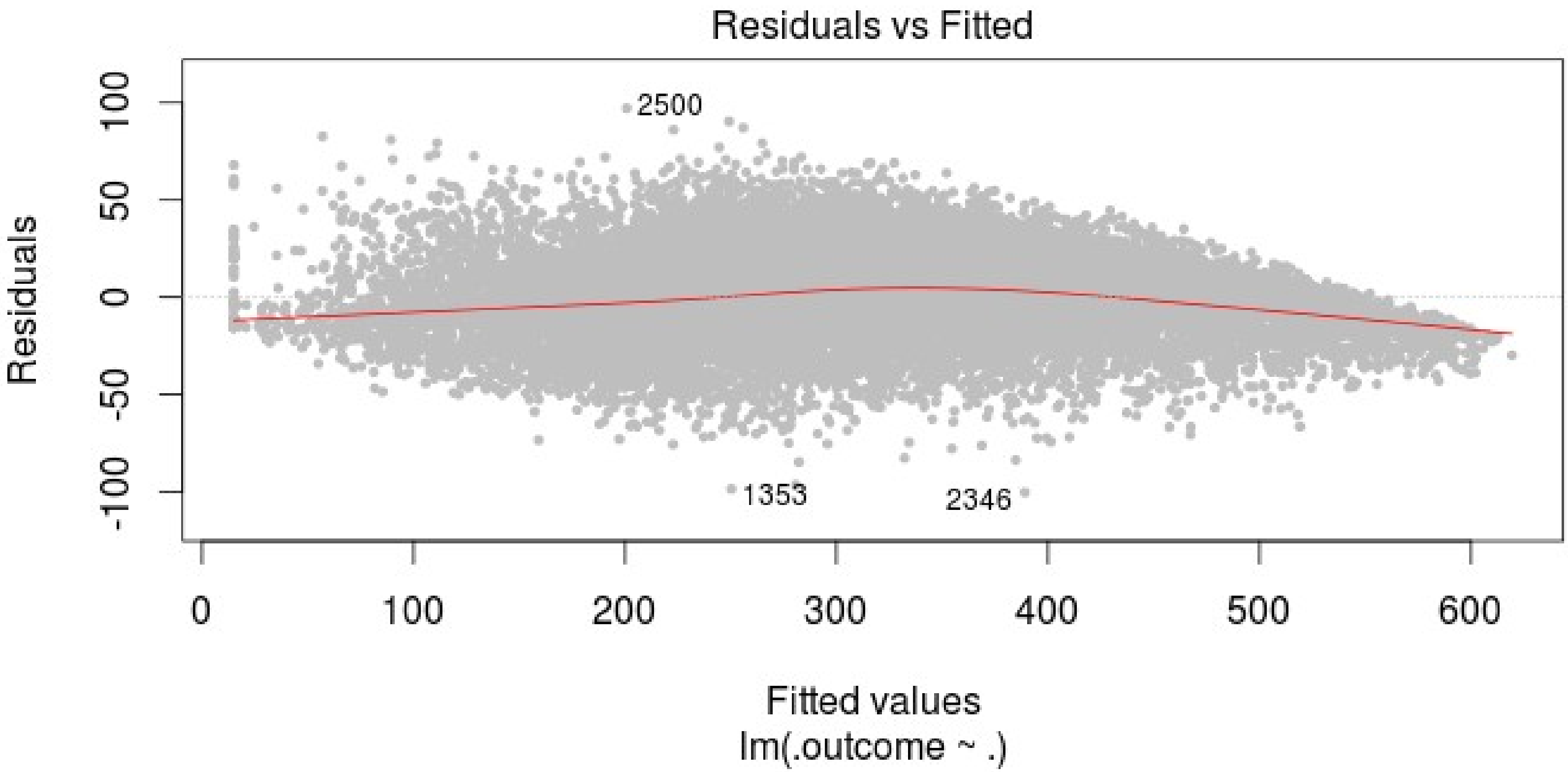
Scatter Plot Matrix

REGRESSION MODEL

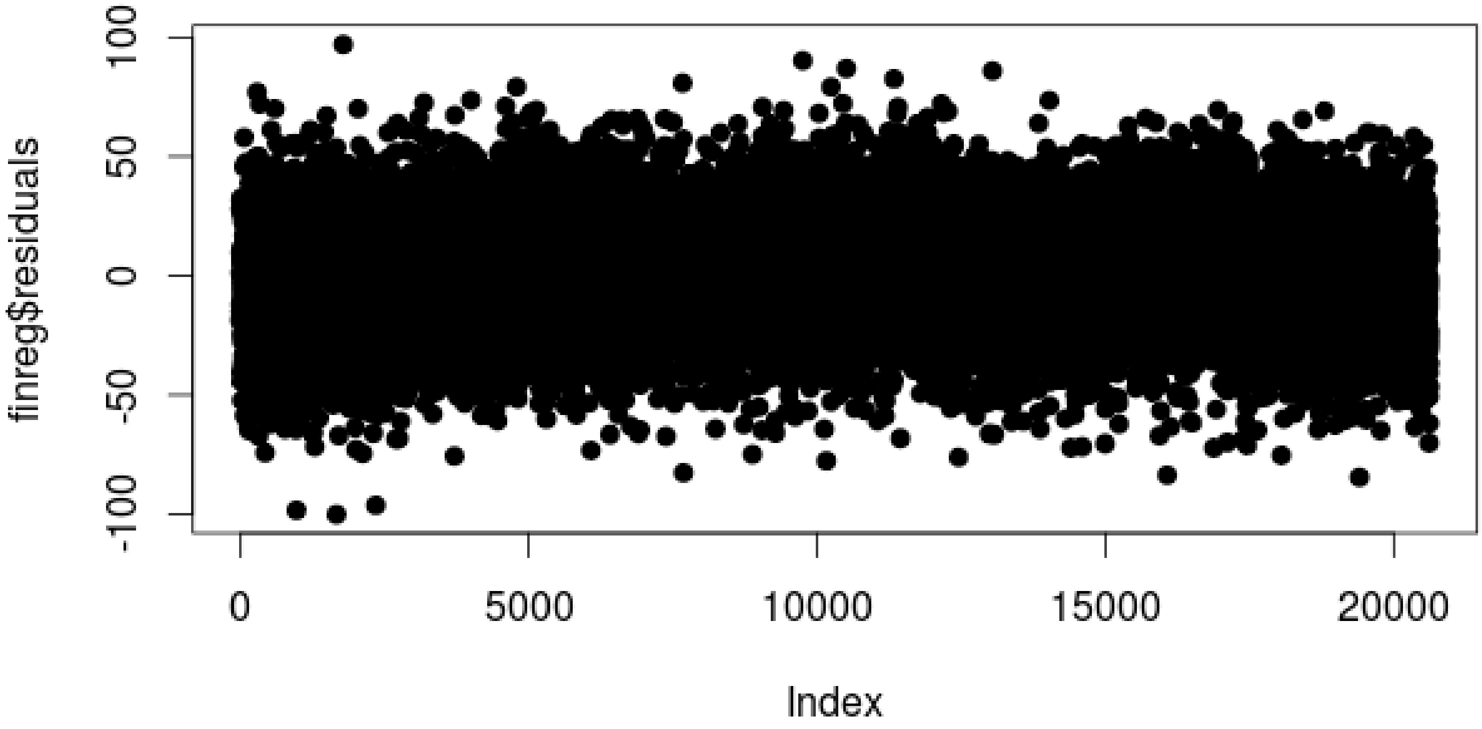


```
> modreg <- train(TOTAL_MARKS ~ S2_MARKS + L2_MARKS + L3_MARKS, method = "lm",data = traind
ata)
> print(modreg)
Linear Regression

20626 samples
   47 predictors

No pre-processing
Resampling: Bootstrapped (25 reps)
Summary of sample sizes: 20626, 20626, 20626, 20626, 20626, 20626, ...
Resampling results

  RMSE      Rsquared   RMSE SD    Rsquared SD
  27.94973  0.9444064  0.1808007  0.0009487419
```

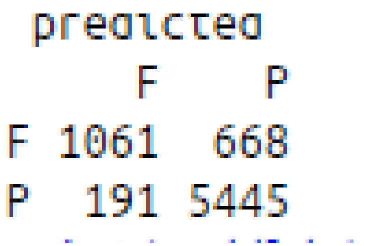
DIAGNOSTICS



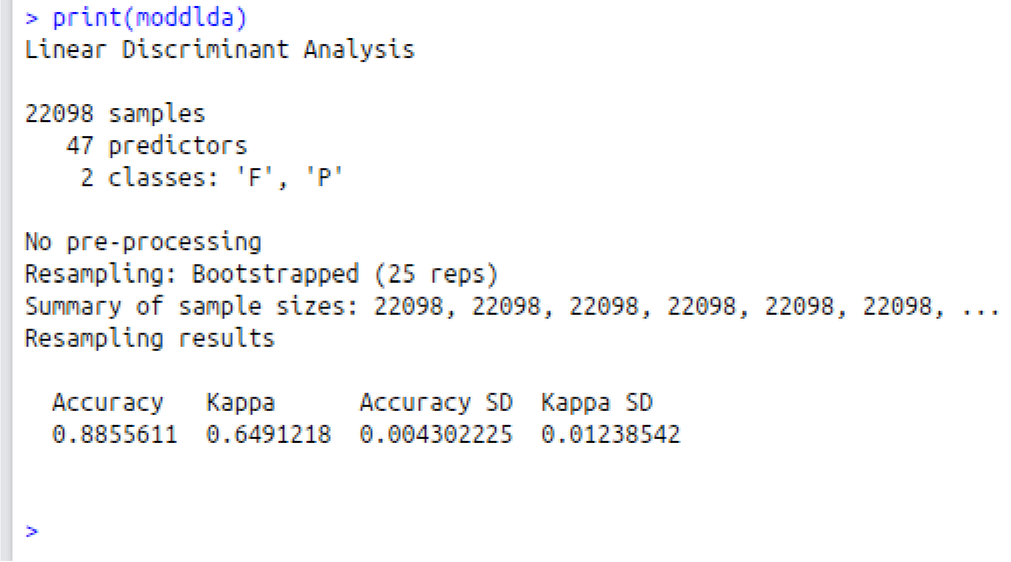
Plot By Index



LDA

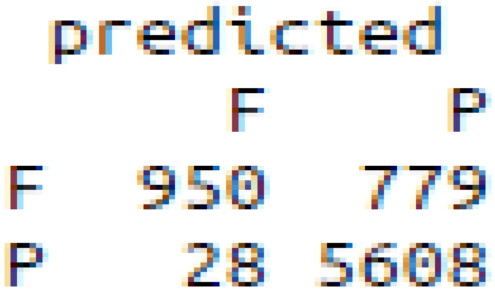


```
   predicted
       F    P
F 1061  668
P  191 5445
```

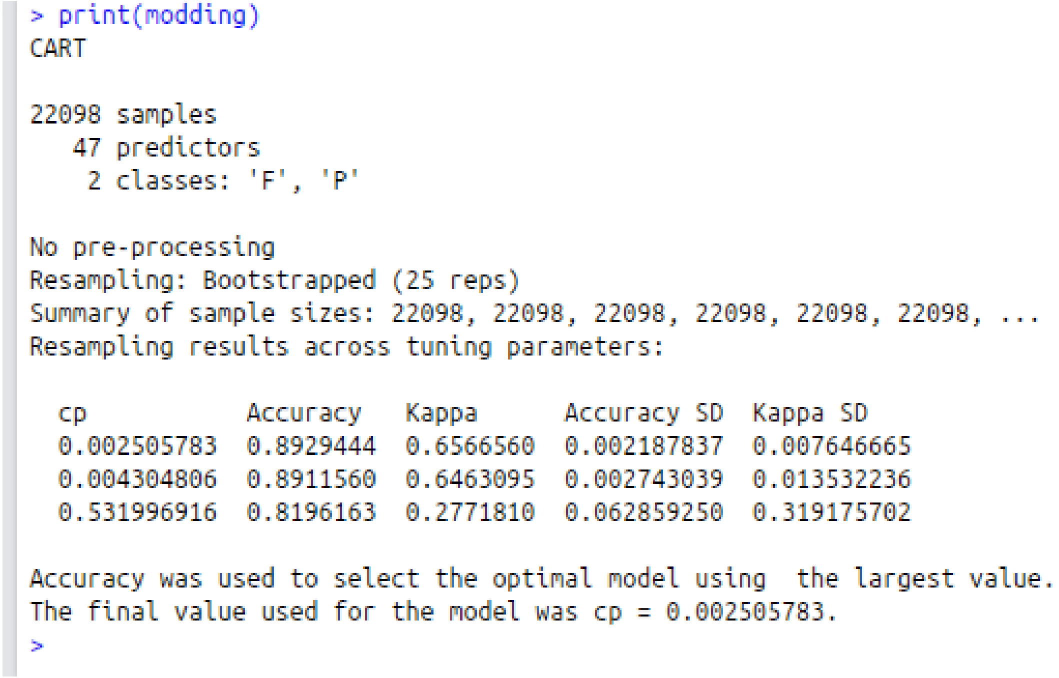


```
> print(moddlda)
Linear Discriminant Analysis

22098 samples
   47 predictors
    2 classes: 'F', 'P'

No pre-processing
Resampling: Bootstrapped (25 reps)
Summary of sample sizes: 22098, 22098, 22098, 22098, 22098, 22098, ...
Resampling results

  Accuracy   Kappa      Accuracy SD  Kappa SD
  0.8855611  0.6491218  0.004302225  0.01238542


>
```

Decision Tree

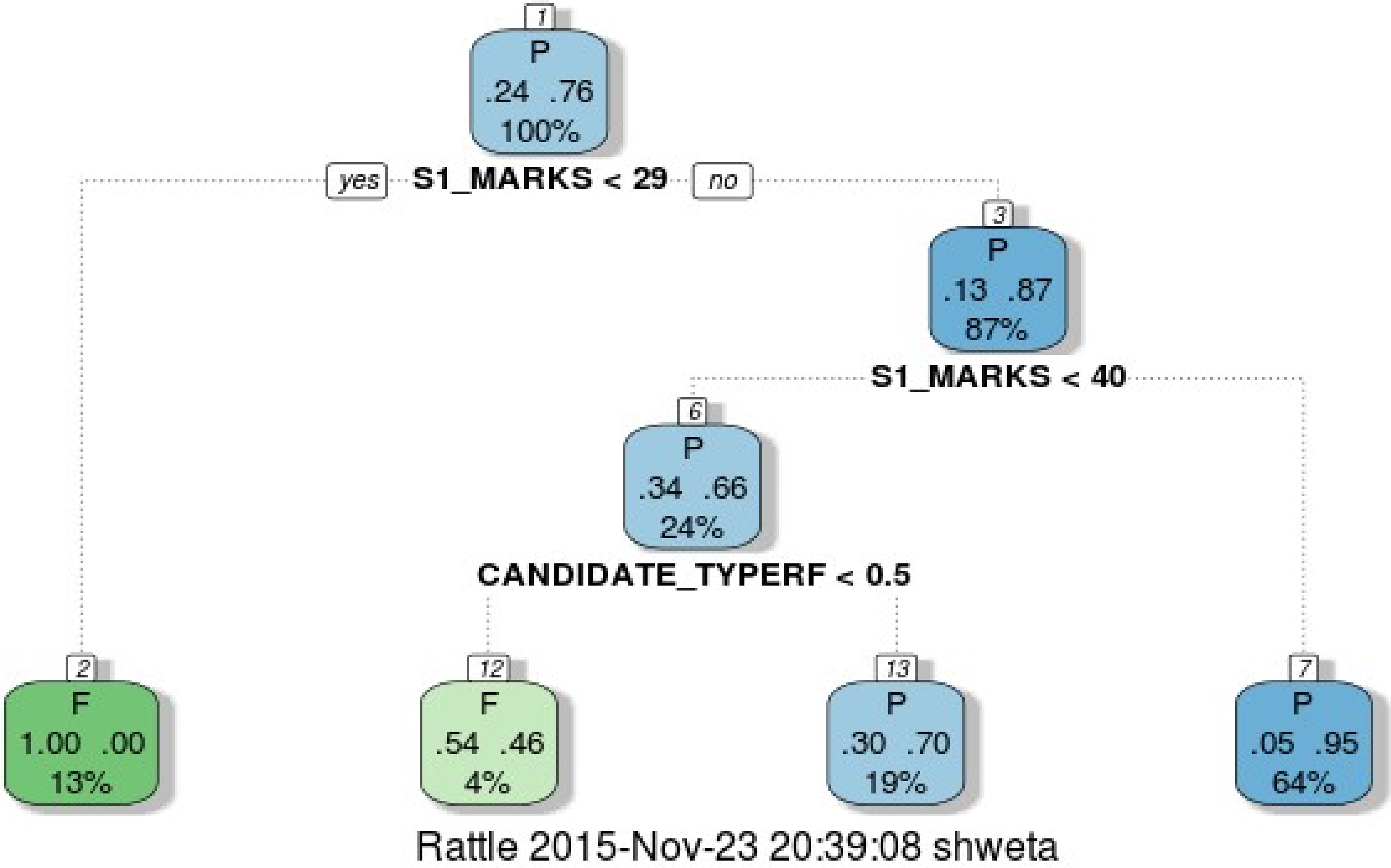


```
 predicted
     F    P
F  950  779
P   28 5608
```



```
> print(modding)
CART

22098 samples
   47 predictors
    2 classes: 'F', 'P'

No pre-processing
Resampling: Bootstrapped (25 reps)
Summary of sample sizes: 22098, 22098, 22098, 22098, 22098, 22098, ...
Resampling results across tuning parameters:

  cp           Accuracy   Kappa      Accuracy SD  Kappa SD
  0.002505783  0.8929444  0.6566560  0.002187837  0.007646665
  0.004304806  0.8911560  0.6463095  0.002743039  0.013532236
  0.531996916  0.8196163  0.2771810  0.062859250  0.319175702

Accuracy was used to select the optimal model using  the largest value.
The final value used for the model was cp = 0.002505783.
>
```

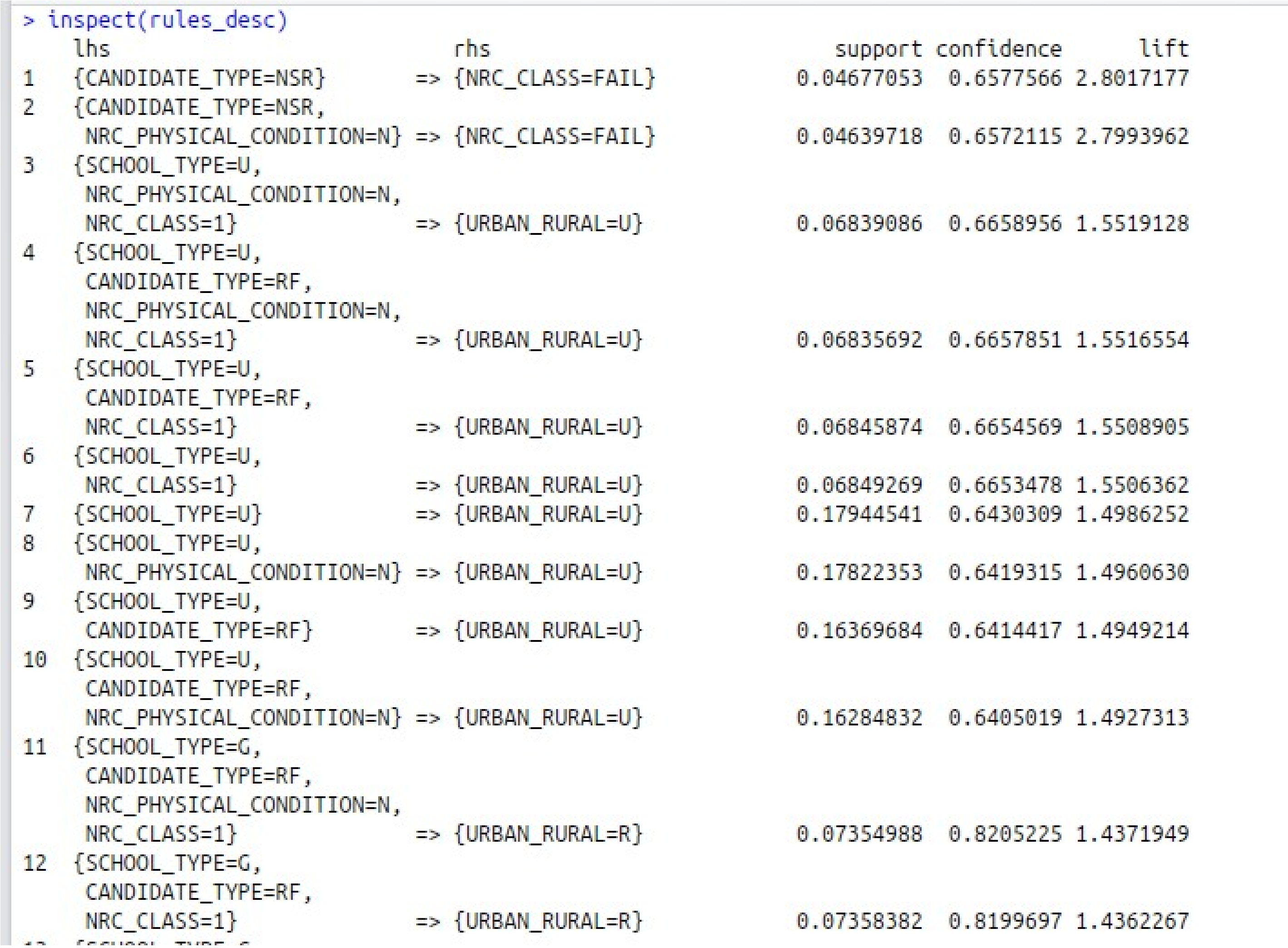
Decision Tree



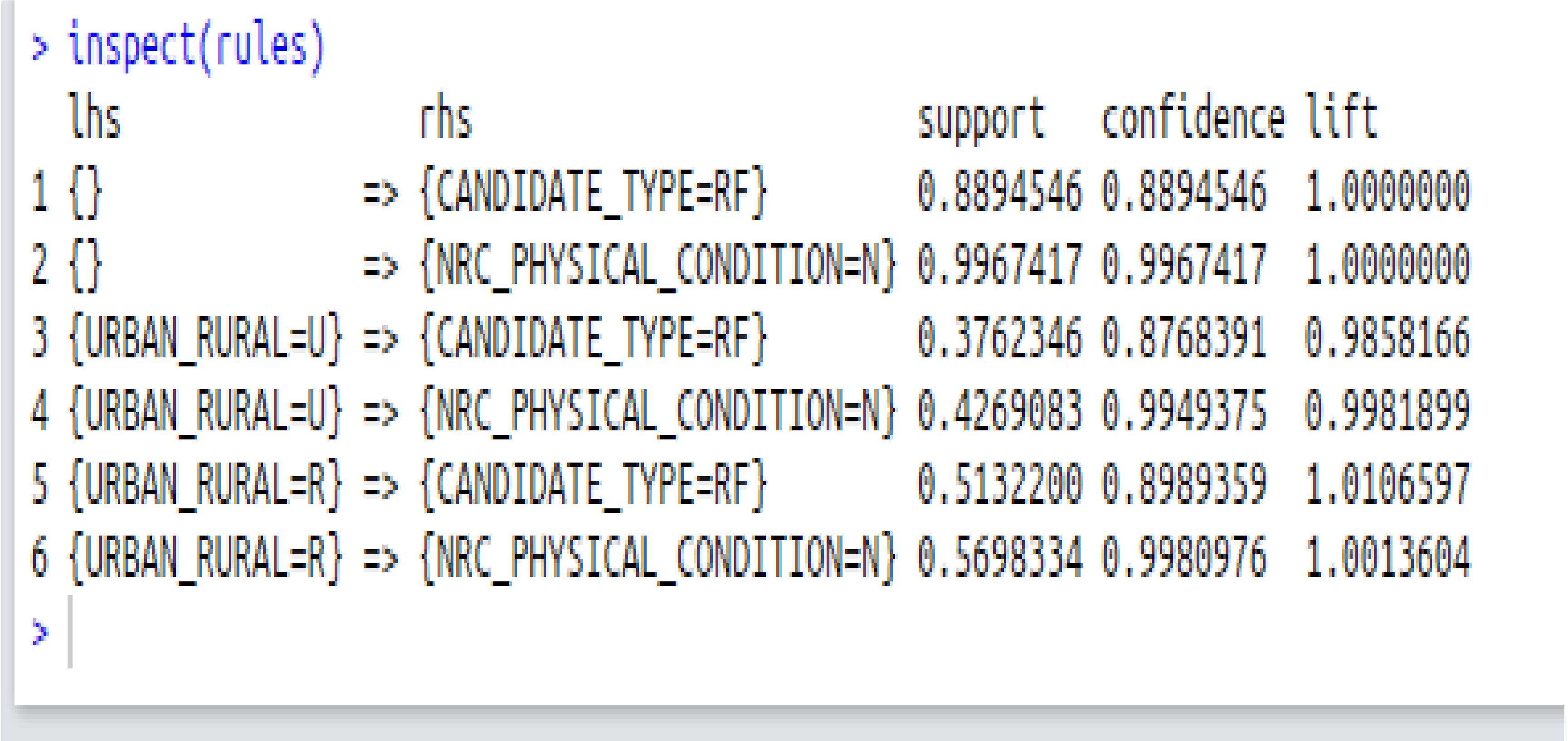
ASSOCIATION

Association

```
> library(arules)
> datasub<-
schooldata[,c("SCHOOL_TYPE","CANDIDATE_TYPE","NRC_PHYSICAL_CONDITION","
URBAN_RURAL","NRC_CLASS")]
> rules<-apriori(datasub)
> inspect(rules)
> rules_desc<-sort(rules,by="lift")
```



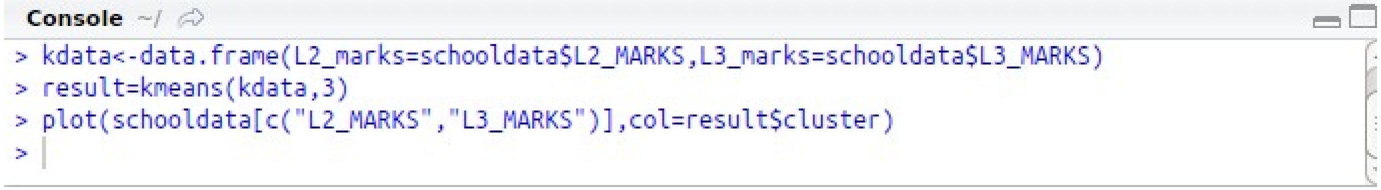
```
> inspect(rules_desc)
    lhs                             rhs                      support confidence      lift
1   {CANDIDATE_TYPE=NSR}         => {NRC_CLASS=FAIL}      0.04677053  0.6577566 2.8017177
2   {CANDIDATE_TYPE=NSR,
     NRC_PHYSICAL_CONDITION=N} => {NRC_CLASS=FAIL}      0.04639718  0.6572115 2.7993962
3   {SCHOOL_TYPE=U,
     NRC_PHYSICAL_CONDITION=N,
     NRC_CLASS=1}               => {URBAN_RURAL=U}       0.06839086  0.6658956 1.5519128
4   {SCHOOL_TYPE=U,
     CANDIDATE_TYPE=RF,
     NRC_PHYSICAL_CONDITION=N,
     NRC_CLASS=1}               => {URBAN_RURAL=U}       0.06835692  0.6657851 1.5516554
5   {SCHOOL_TYPE=U,
     CANDIDATE_TYPE=RF,
     NRC_CLASS=1}               => {URBAN_RURAL=U}       0.06845874  0.6654569 1.5508905
6   {SCHOOL_TYPE=U,
     NRC_CLASS=1}               => {URBAN_RURAL=U}       0.06849269  0.6653478 1.5506362
7   {SCHOOL_TYPE=U}              => {URBAN_RURAL=U}       0.17944541  0.6430309 1.4986252
8   {SCHOOL_TYPE=U,
     NRC_PHYSICAL_CONDITION=N} => {URBAN_RURAL=U}       0.17822353  0.6419315 1.4960630
9   {SCHOOL_TYPE=U,
     CANDIDATE_TYPE=RF}         => {URBAN_RURAL=U}       0.16369684  0.6414417 1.4949214
10  {SCHOOL_TYPE=U,
     CANDIDATE_TYPE=RF,
     NRC_PHYSICAL_CONDITION=N} => {URBAN_RURAL=U}       0.16284832  0.6405019 1.4927313
11  {SCHOOL_TYPE=G,
     CANDIDATE_TYPE=RF,
     NRC_PHYSICAL_CONDITION=N,
     NRC_CLASS=1}               => {URBAN_RURAL=R}       0.07354988  0.8205225 1.4371949
12  {SCHOOL_TYPE=G,
     CANDIDATE_TYPE=RF,
     NRC_CLASS=1}               => {URBAN_RURAL=R}       0.07358382  0.8199697 1.4362267
13  {SCHOOL_TYPE=G,
> rules <- apriori(datasub, control = list(verbose=T), parameter =
list(supp=0.004,conf=0.6), appearance = list(lhs=c("URBAN_RURAL=R",
"URBAN_RURAL=U"),default="rhs"))
```



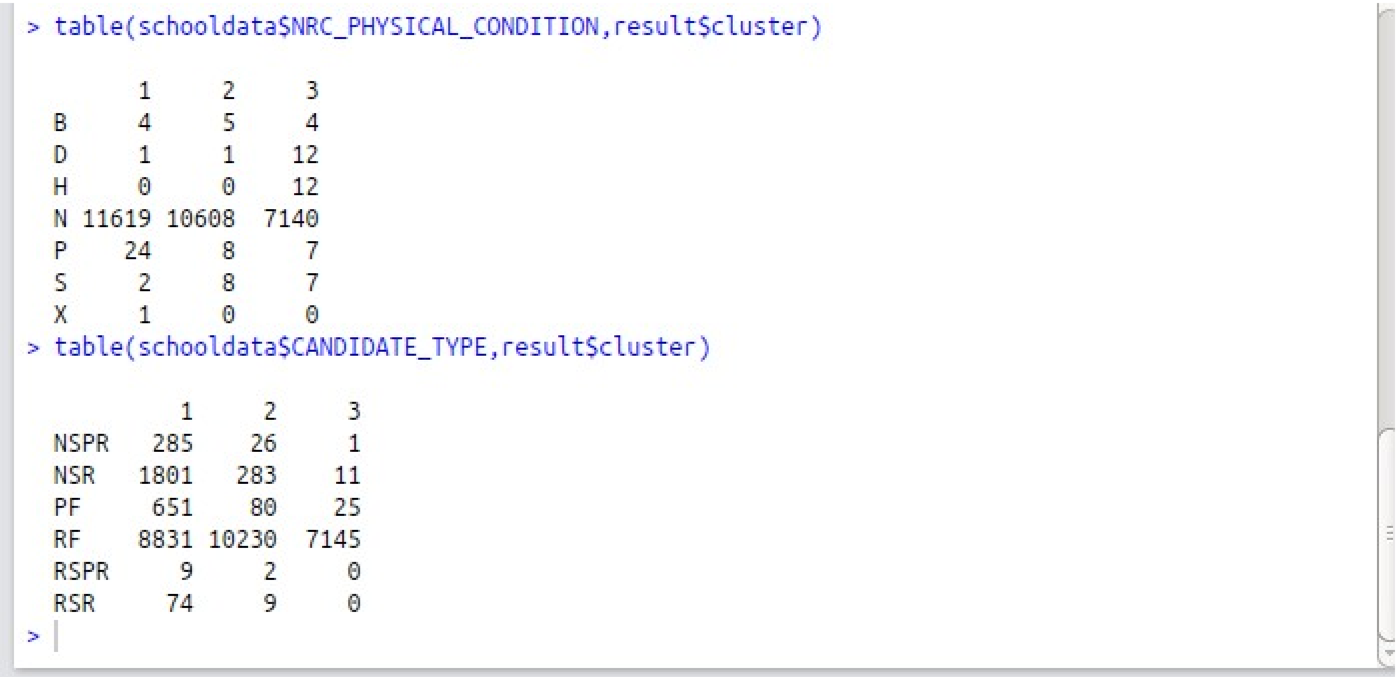
```
> inspect(rules)
  lhs                rhs                           support   confidence lift
1 {}              => {CANDIDATE_TYPE=RF}           0.8894546 0.8894546  1.0000000
2 {}              => {NRC_PHYSICAL_CONDITION=N}    0.9967417 0.9967417  1.0000000
3 {URBAN_RURAL=U} => {CANDIDATE_TYPE=RF}           0.3762346 0.8768391  0.9858166
4 {URBAN_RURAL=U} => {NRC_PHYSICAL_CONDITION=N}    0.4269083 0.9949375  0.9981899
5 {URBAN_RURAL=R} => {CANDIDATE_TYPE=RF}           0.5132200 0.8989359  1.0106597
6 {URBAN_RURAL=R} => {NRC_PHYSICAL_CONDITION=N}    0.5698334 0.9980976  1.0013604
>
```

CLUSTERING

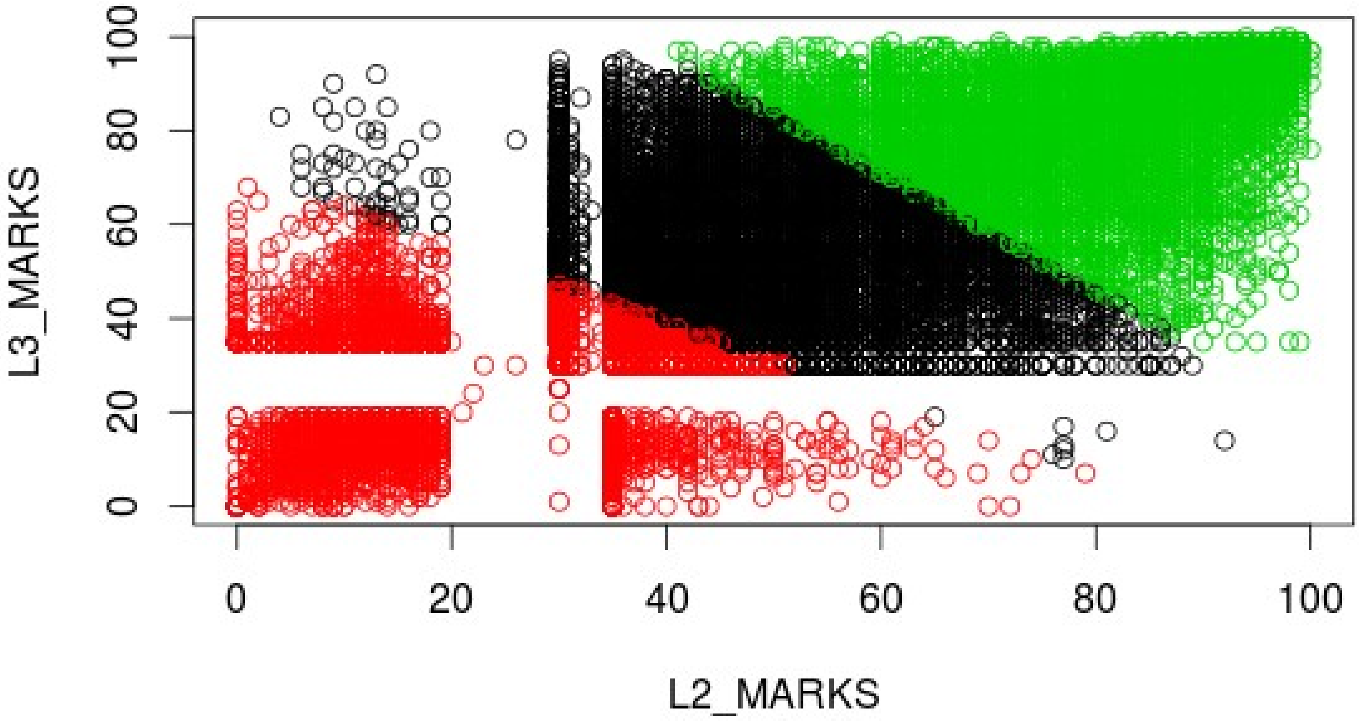
K-Means



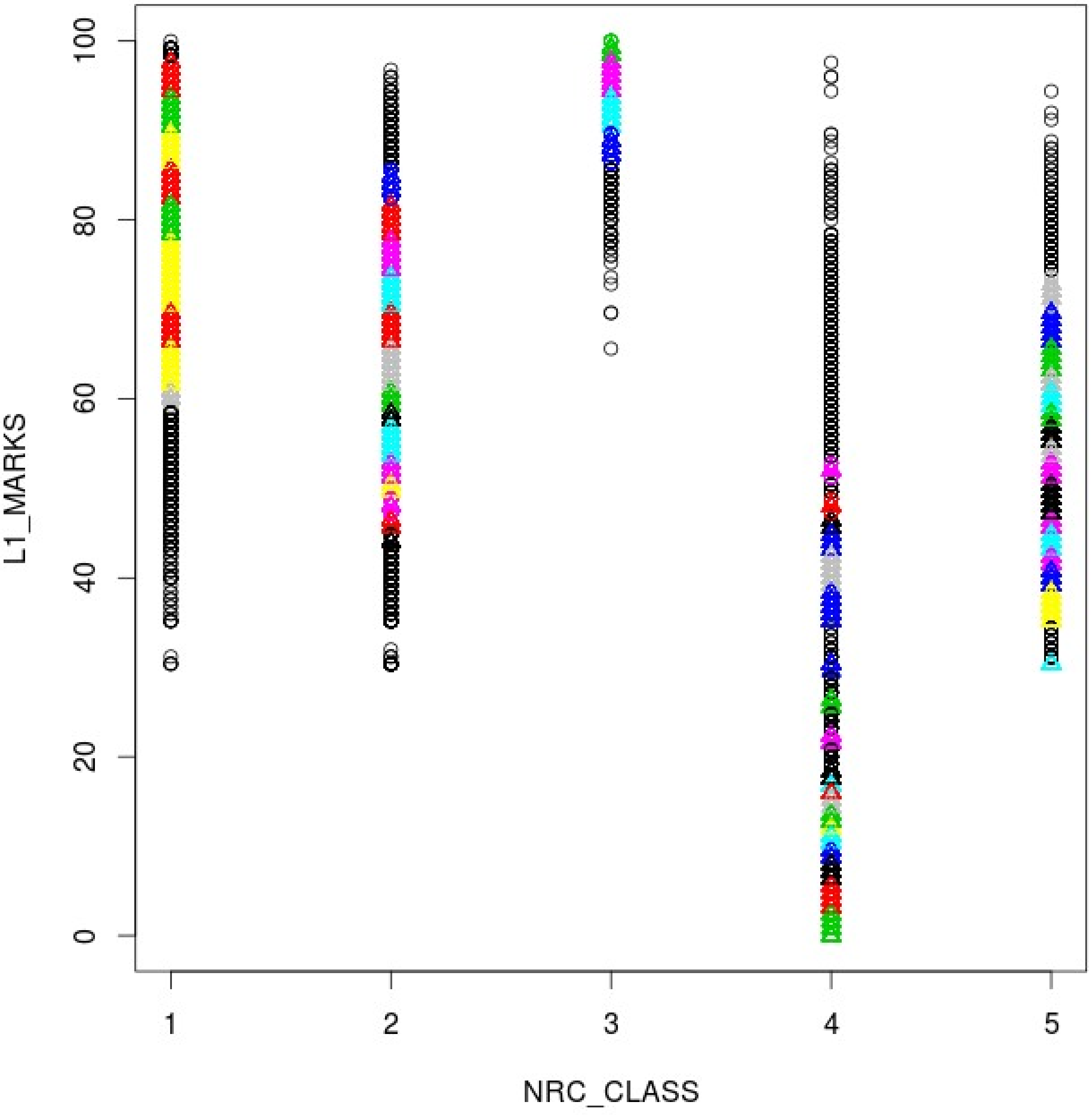
```
> kdata<-data.frame(L2_marks=schooldata$L2_MARKS,L3_marks=schooldata$L3_MARKS)
> result=kmeans(kdata,3)
> plot(schooldata[c("L2_MARKS","L3_MARKS")],col=result$cluster)
> 
```

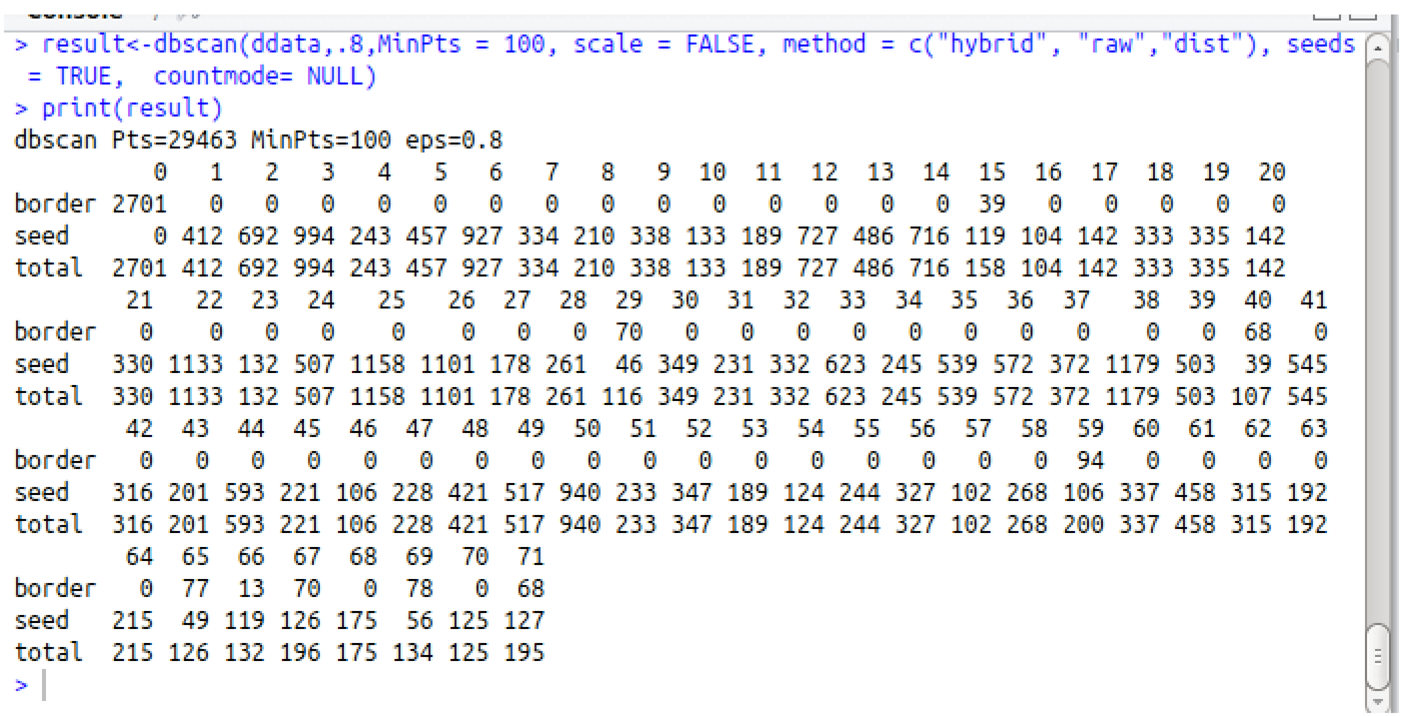


```
> table(schooldata$NRC_PHYSICAL_CONDITION,result$cluster)

         1     2     3
  B      4     5     4
  D      1     1    12
  H      0     0    12
  N  11619 10608  7140
  P     24     8     7
  S      2     8     7
  X      1     0     0
> table(schooldata$CANDIDATE_TYPE,result$cluster)

           1     2     3
  NSPR   285    26     1
  NSR   1801   283    11
  PF     651    80    25
  RF    8831 10230  7145
  RSPR     9     2     0
  RSR     74     9     0
> 
```

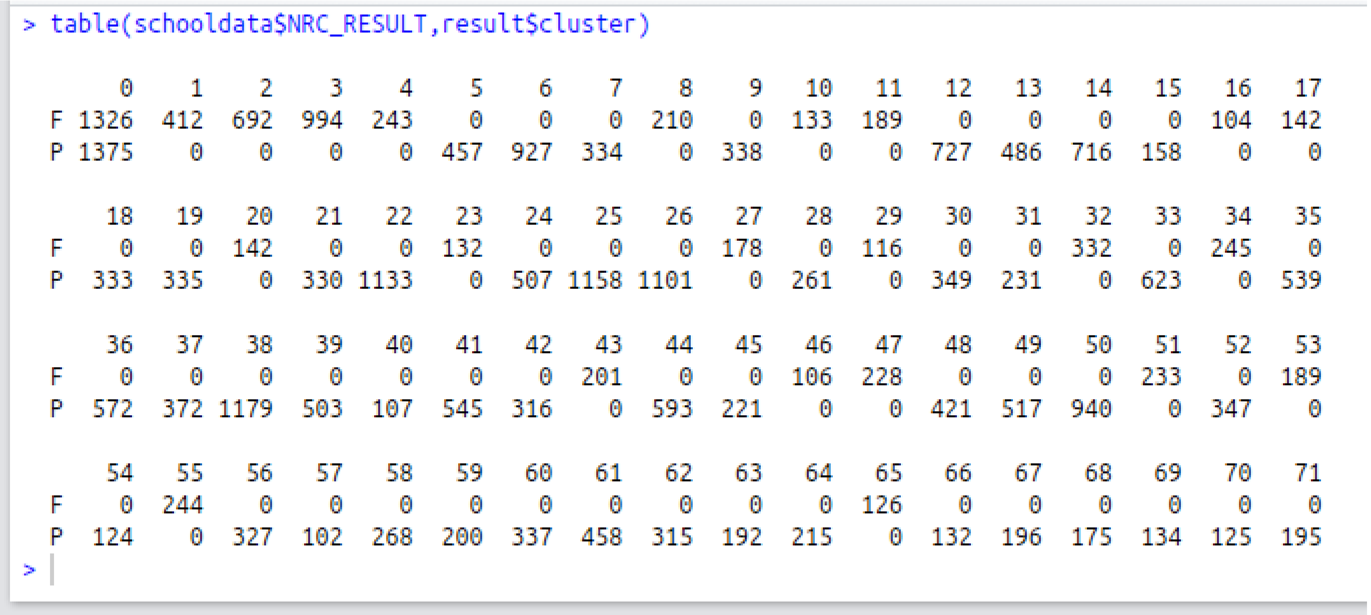

L3_MARKS
0
20
40
60
80
100
L2_MARKS
0
20
40
60
80
100

DBSCAN





```
> result<-dbscan(ddata,.8,MinPts = 100, scale = FALSE, method = c("hybrid", "raw","dist"), seeds
 = TRUE,  countmode= NULL)
> print(result)
dbscan Pts=29463 MinPts=100 eps=0.8
          0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20
border 2701   0   0   0   0   0   0   0   0   0   0   0   0   0   0  39   0   0   0   0   0
seed      0 412 692 994 243 457 927 334 210 338 133 189 727 486 716 119 104 142 333 335 142
total  2701 412 692 994 243 457 927 334 210 338 133 189 727 486 716 158 104 142 333 335 142
        21   22  23  24   25   26  27  28  29  30  31  32  33  34  35  36  37   38  39  40  41
border   0    0   0   0    0    0   0   0  70   0   0   0   0   0   0   0   0    0   0  68   0
seed   330 1133 132 507 1158 1101 178 261  46 349 231 332 623 245 539 572 372 1179 503  39 545
total  330 1133 132 507 1158 1101 178 261 116 349 231 332 623 245 539 572 372 1179 503 107 545
        42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60  61  62  63
border   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  94   0   0   0   0
seed   316 201 593 221 106 228 421 517 940 233 347 189 124 244 327 102 268 106 337 458 315 192
total  316 201 593 221 106 228 421 517 940 233 347 189 124 244 327 102 268 200 337 458 315 192
        64  65  66  67  68  69  70  71
border   0  77  13  70   0  78   0  68
seed   215  49 119 126 175  56 125 127
total  215 126 132 196 175 134 125 195
> 
```



```
> table(schooldata$NRC_RESULT,result$cluster)

       0    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17
  F 1326  412  692  994  243    0    0    0  210    0  133  189    0    0    0    0  104  142
  P 1375    0    0    0    0  457  927  334    0  338    0    0  727  486  716  158    0    0

      18   19   20   21   22   23   24   25   26   27   28   29   30   31   32   33   34   35
  F    0    0  142    0    0  132    0    0    0  178    0  116    0    0  332    0  245    0
  P  333  335    0  330 1133    0  507 1158 1101    0  261    0  349  231    0  623    0  539

      36   37   38   39   40   41   42   43   44   45   46   47   48   49   50   51   52   53
  F    0    0    0    0    0    0    0  201    0    0  106  228    0    0    0  233    0  189
  P  572  372 1179  503  107  545  316    0  593  221    0    0  421  517  940    0  347    0

      54   55   56   57   58   59   60   61   62   63   64   65   66   67   68   69   70   71
  F    0  244    0    0    0    0    0    0    0    0    0  126    0    0    0    0    0    0
  P  124    0  327  102  268  200  337  458  315  192  215    0  132  196  175  134  125  195
> 
```

Thank You