ASSIGNMENT-4

Packages Used:

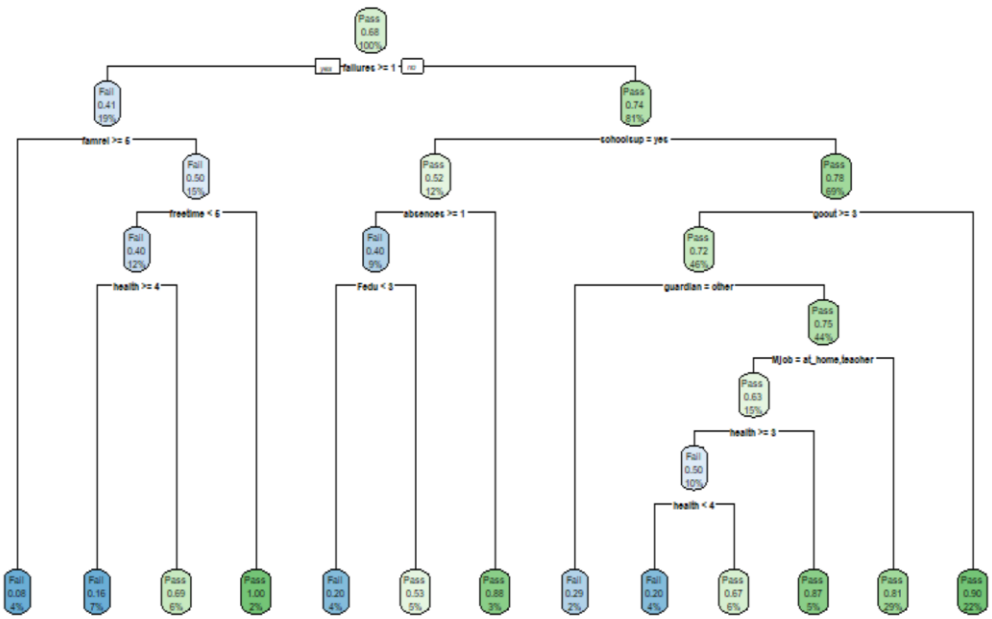
- Caret
- Rpart
- Rpart.plot
- Naïvebayes

Steps and Formula

- Using condition if G3<10; mark as Fail or else Pass and converted G3 column
  - **R Formula:** record$G3<-ifelse(record$G3<10,"Fail","Pass")
  - record$G3<-factor(record$G3)
- Set the random sample seed to 1234
- Divided the entire data with 70% of being the training data and remaining 30% being the testing data
  - **R Formula:** sample(2,nrow(record),replace=TRUE,prob = c(0.7,0.3))->partition
  - record[partition==1,]->training_dataset
  - record[partition==2,]->validation_dataset

DECISION TREE

- The following tree is plot using all variables.
  - **R Formula:** rpart(G3~ .,training_dataset,method = "class")->tree
  - rpart.plot(tree)



- On the basis of the above plot, we have predicted the 30% Test Data Set
    - **R Formula:** predict(tree,validation_dataset,type='class')->test
- Following is the Matrix of the test data based on the Decision Tree Algorithm
    - **R Formula:** (table(validation_dataset$G3,test1)->table_DT)

	Fail	Pass
Fail	15	25
Pass	11	64

- Following are the criteria to validate the model
    - Accuracy: It is the proportion of true positive and true negative over the sum of the matrix. This model has 68.7% accuracy
        - **R Formula:** sum(diag(table_DT))/sum(table1)->accuracy_DT
    - Recall: Recall measures how many truly relevant results are returned. This model has 57.7% recall rate
        - **R Formula:** (table_DT[1,1])/(sum(table_DT[,1]))->recall_DT
    - Precision: This is a measure of result relevancy. This model has 37.5% precision
        - **R Formula:** (table_DT[1,1])/(sum(table_DT[1,]))->precision_DT

NAÏVE BAYES ALGORITHM

- Run Naïve Bayes Algorithm on the Training Data and Plot it
    - **R Formula:** naive_bayes(G3~ .,data=training_dataset)->naive
    - plot(naive)
    - Please note the plots are below
- Following is the Matrix of the test data based on the Naïve Bayes Algorithm
    - **R Formula:** predict(naive,validation_dataset,type='class')->test2
    - (table(validation_dataset$G3,test2)->table_NB)

	Fail	Pass
Fail	20	20
Pass	8	67

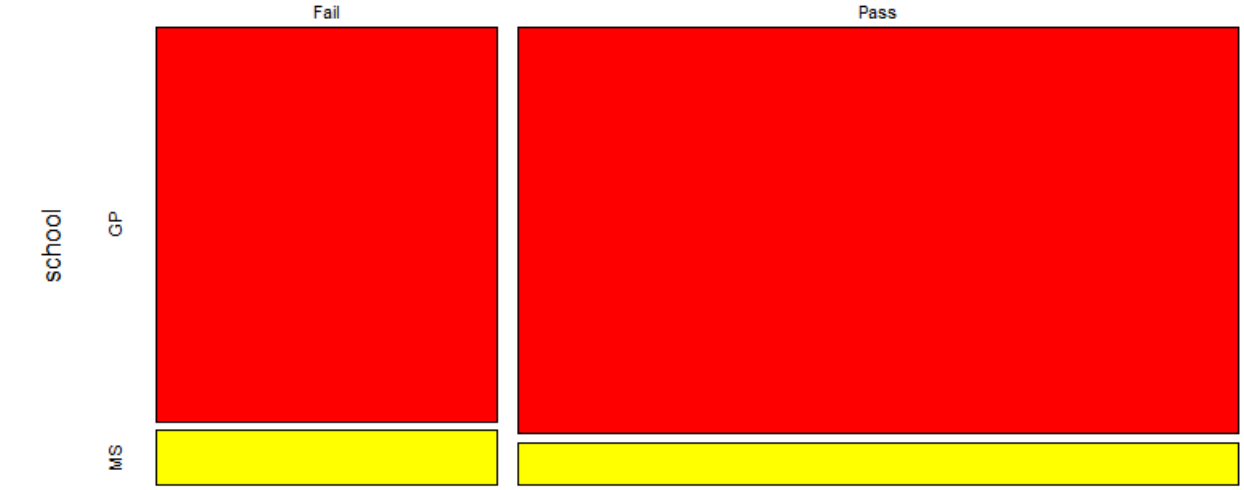
- Following are the criteria to validate the model
    - Accuracy: 75.65%
        - R Formula: sum(diag(table_NB))/sum(table_NB)->accuracy_NB
    - Recall: 71.42%
        - R Formula: (table_NB[1,1])/(sum(table_NB[,1]))->recall_NB
    - Precision: 50%
        - R Formula: (table_NB[1,1])/(sum(table_NB[1,]))->precision_NB

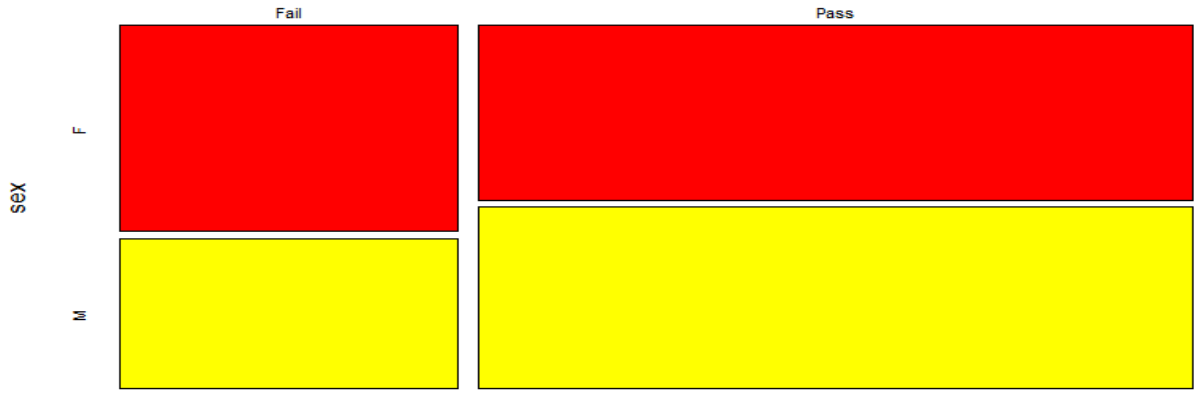
ANALYSIS

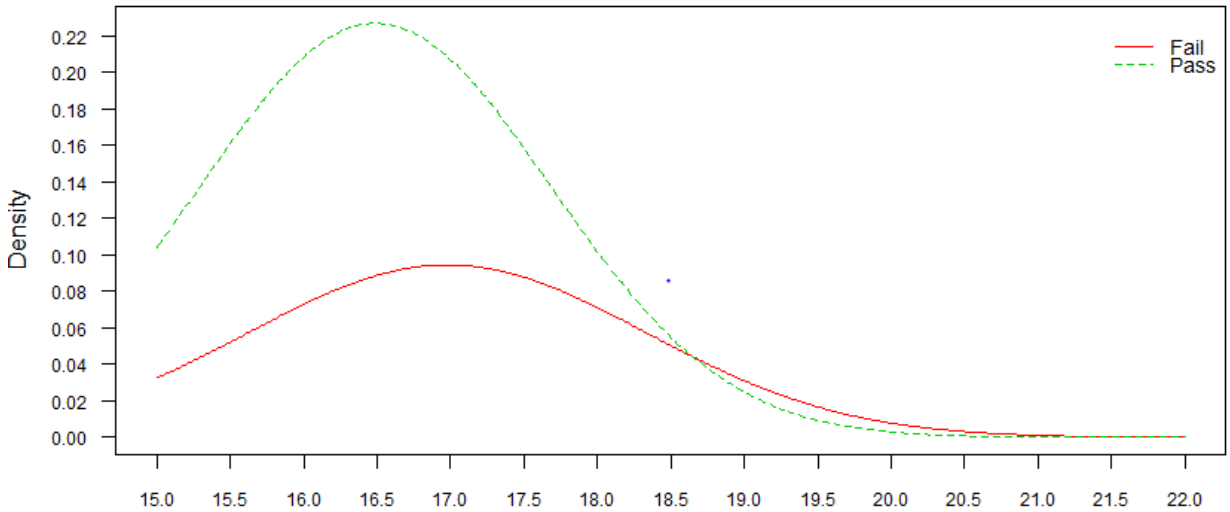
According to Accuracy of both the models , Naïve Bayes algorithm has better accuracy than Decision tree for this particular data(Avoiding Correlation and Feature selection)

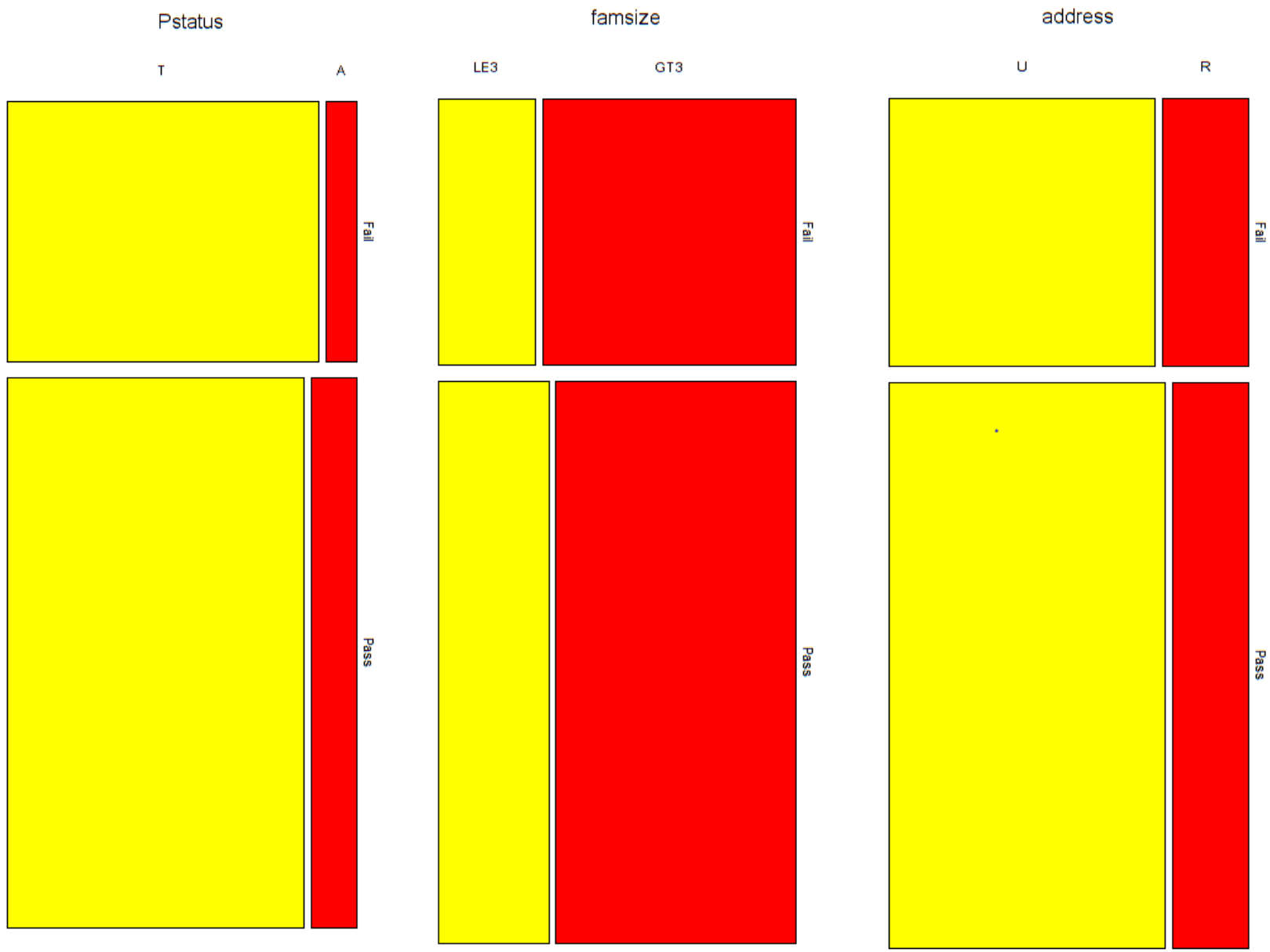
Note:- The Accuracy of data may be change if uses correlation or feature selection.

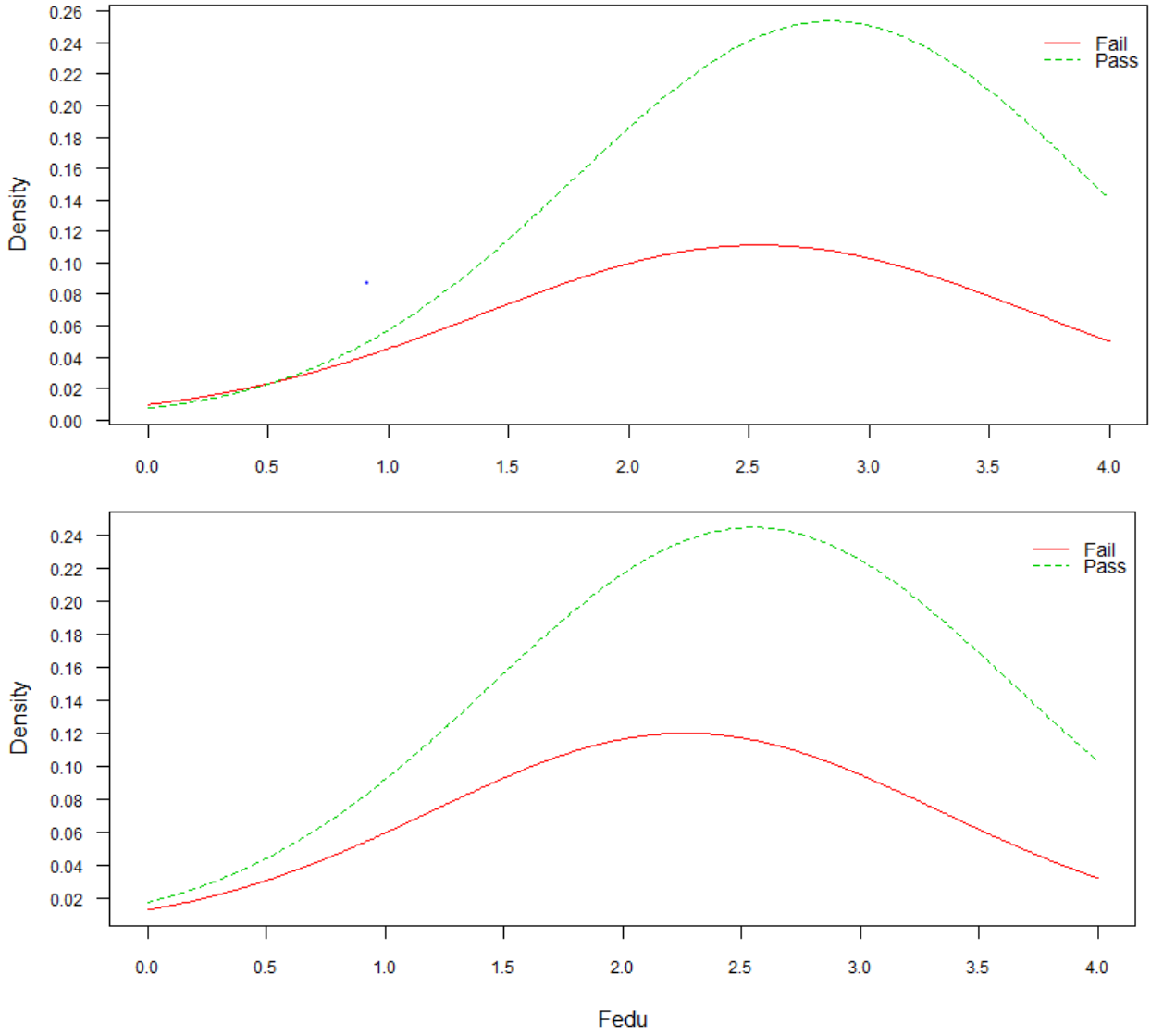
PLOTS OF NAÏVE BAYES



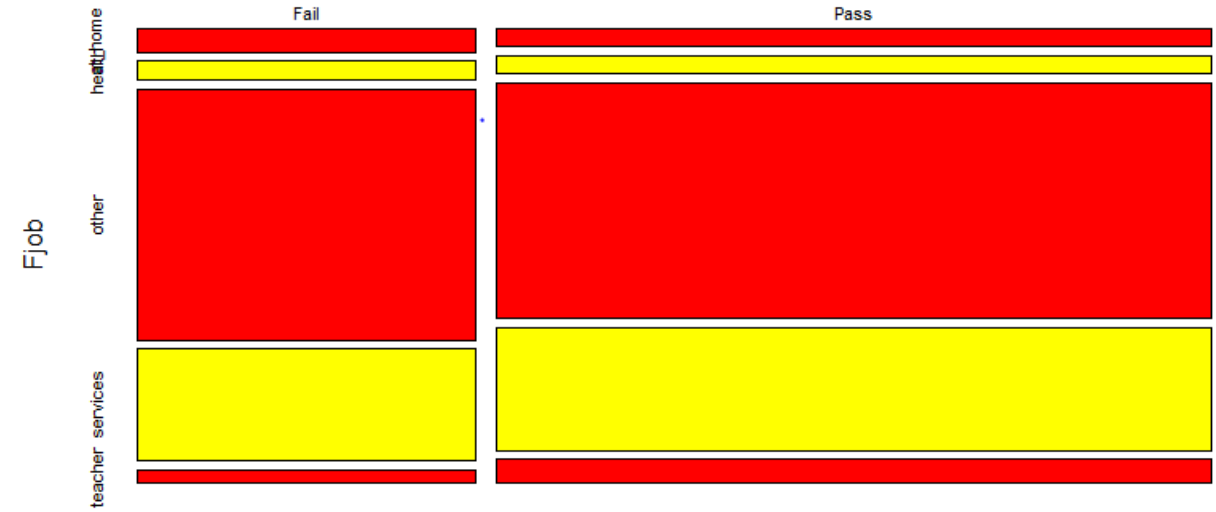




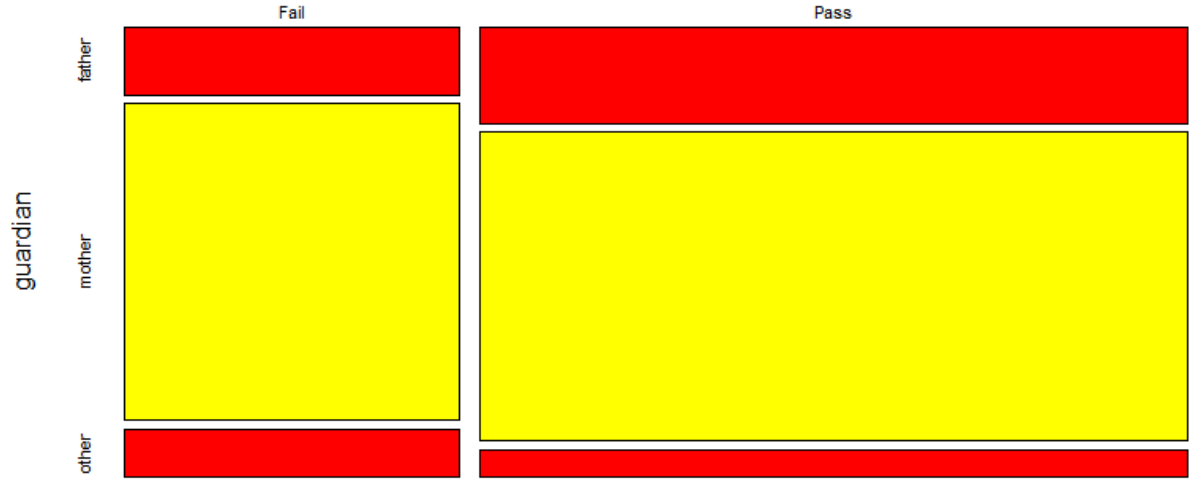

Pstatus
T
A
Fail
Pass
famsize
LE3
GT3
Fail
Pass
address
U
R
Fail
Pass

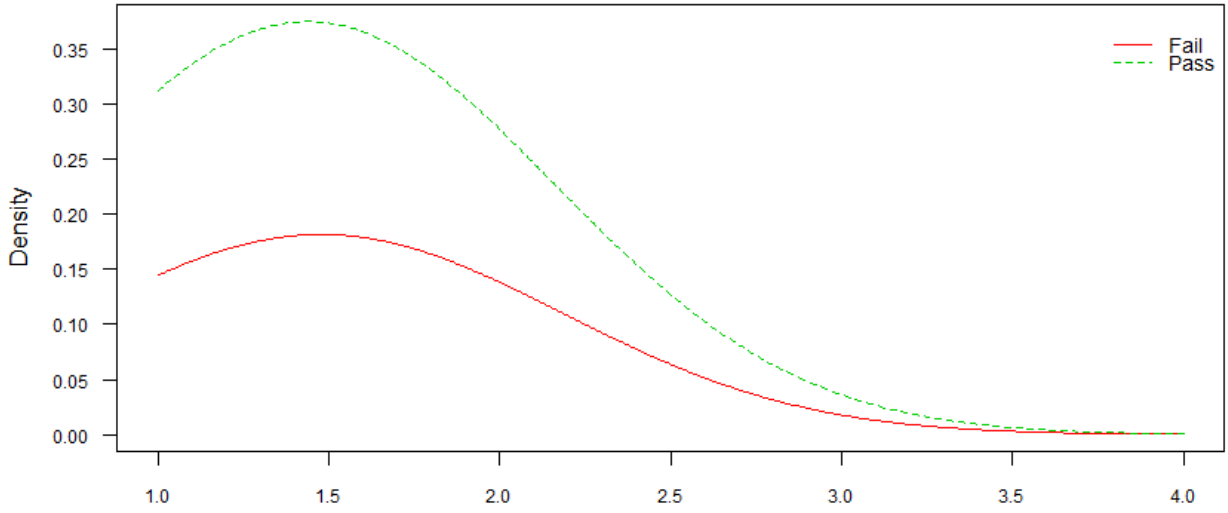

Density
0.26
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Fail
Pass

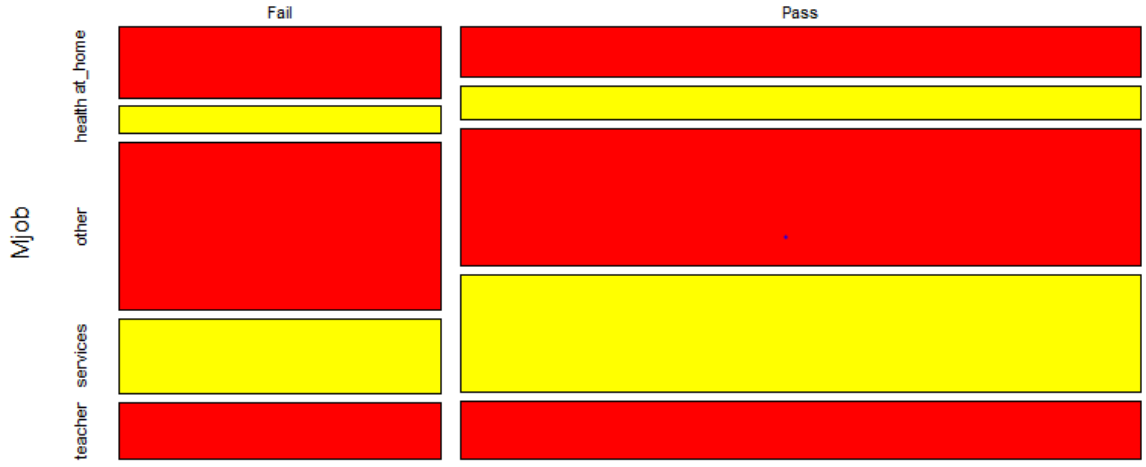
Density
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Fail
Pass
Fedu

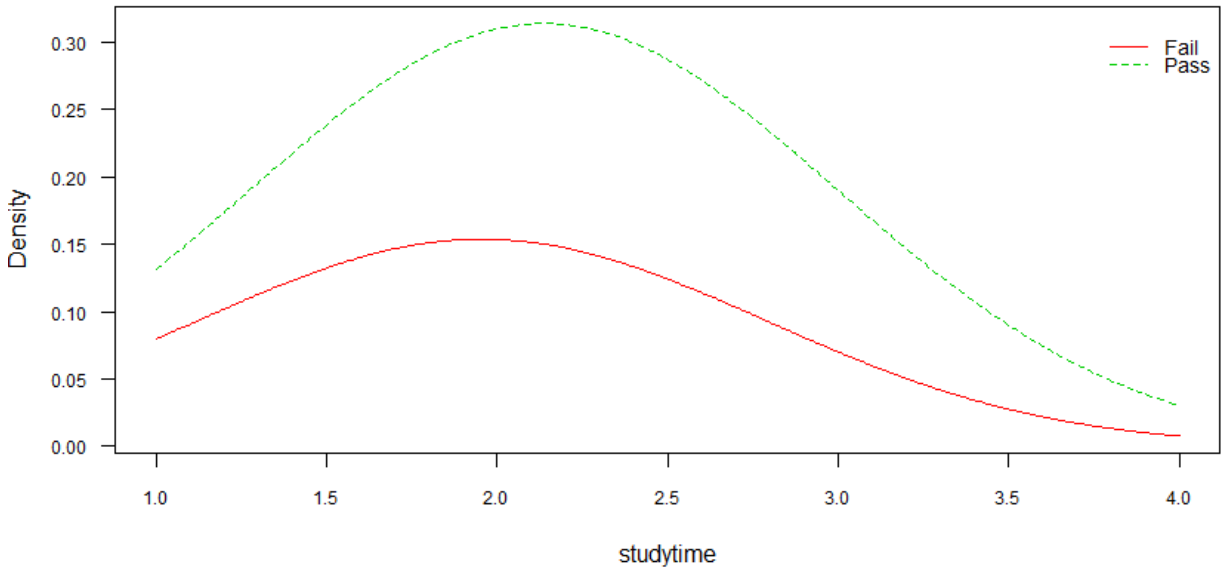

Fail
Pass
Fjob
health
home
other
services
teacher

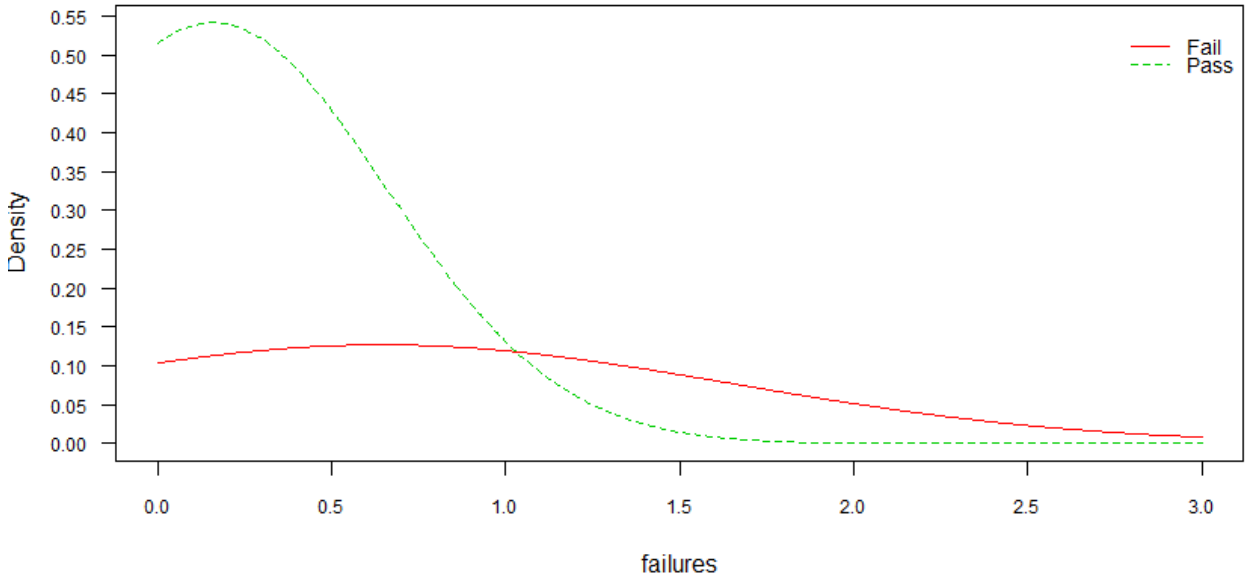

Fail
Pass
reason
course
home
other
reputation

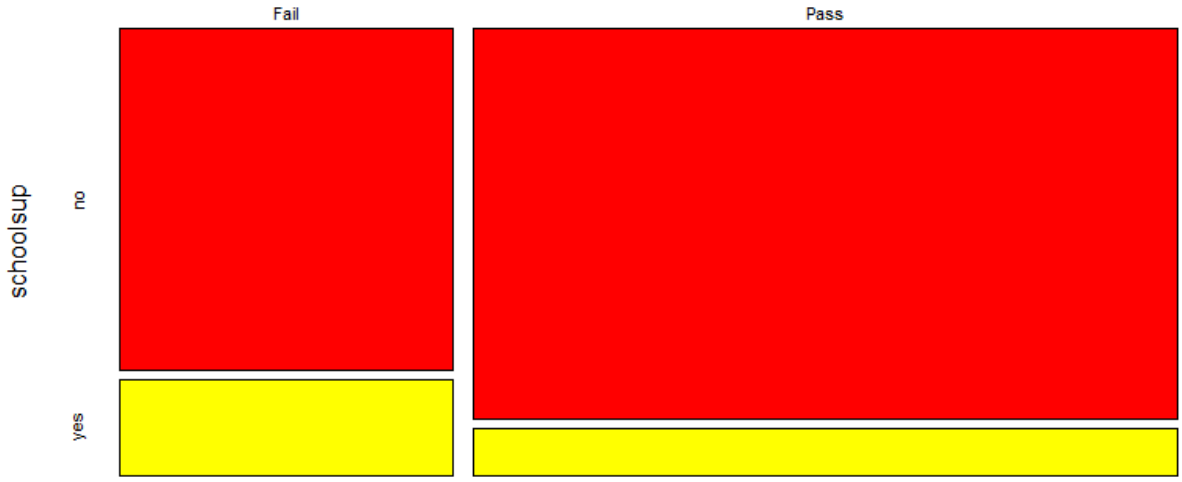

Fail
Pass
guardian
father
mother
other

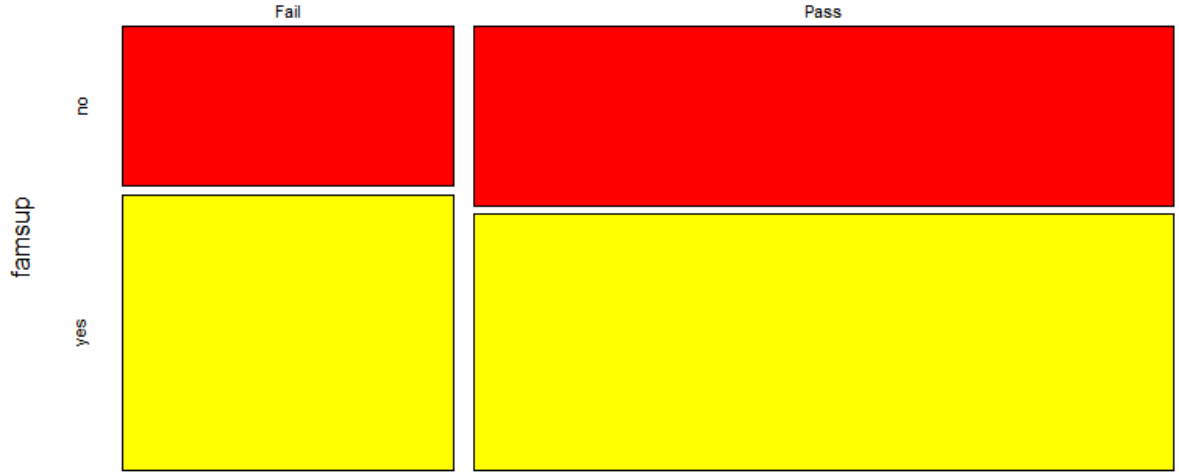

Density
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Fail
Pass

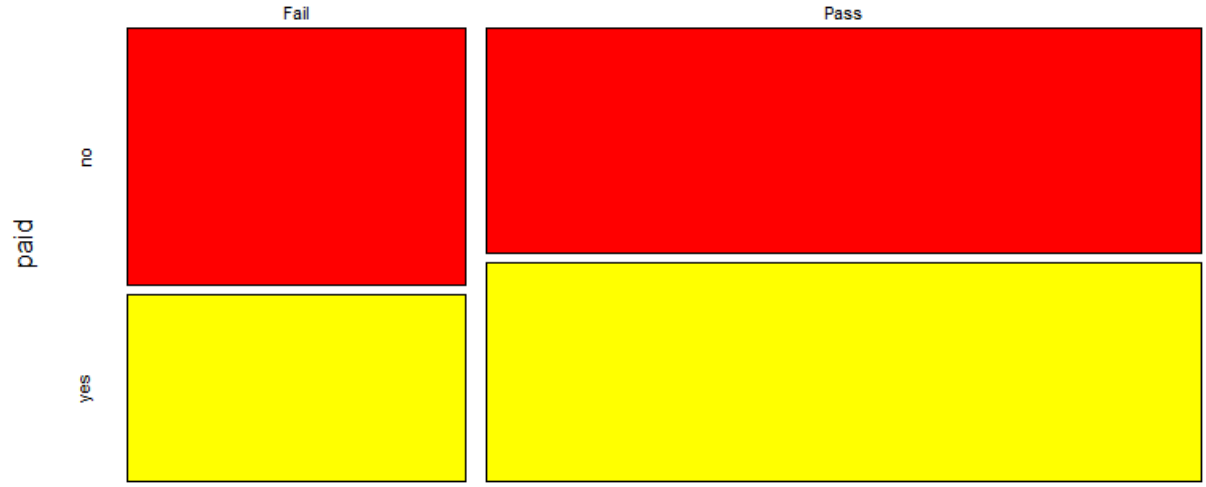

Fail
Pass
Mjob
health
at_home
other
services
teacher

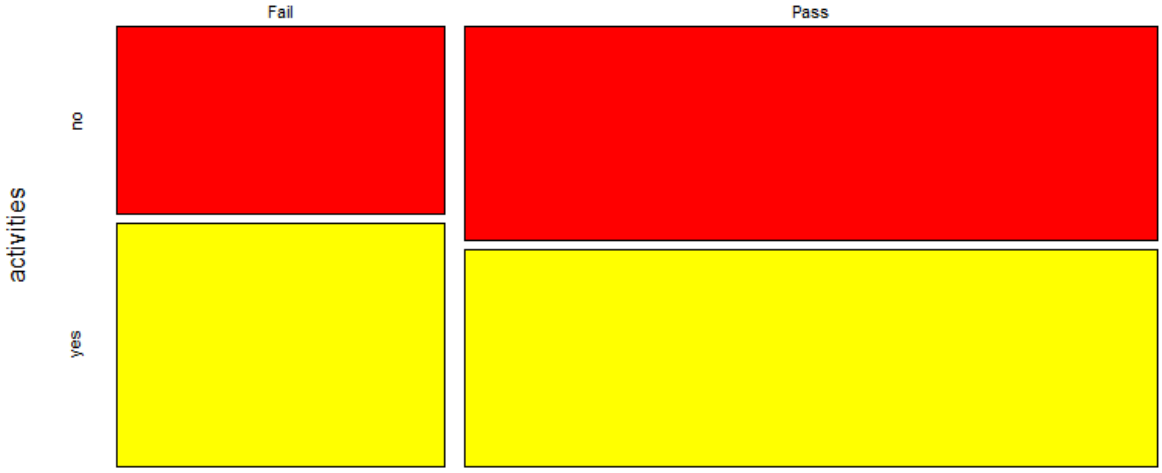

Fail
Pass
Density
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1.0
1.5
2.0
2.5
3.0
3.5
4.0
studytime

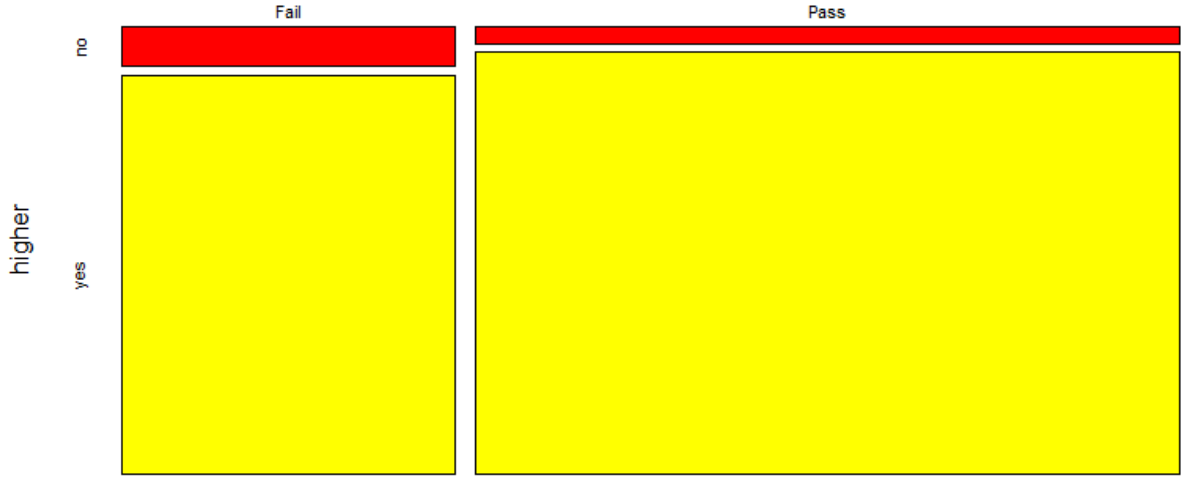

Fail
Pass
Density
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.0
0.5
1.0
1.5
2.0
2.5
3.0
failures

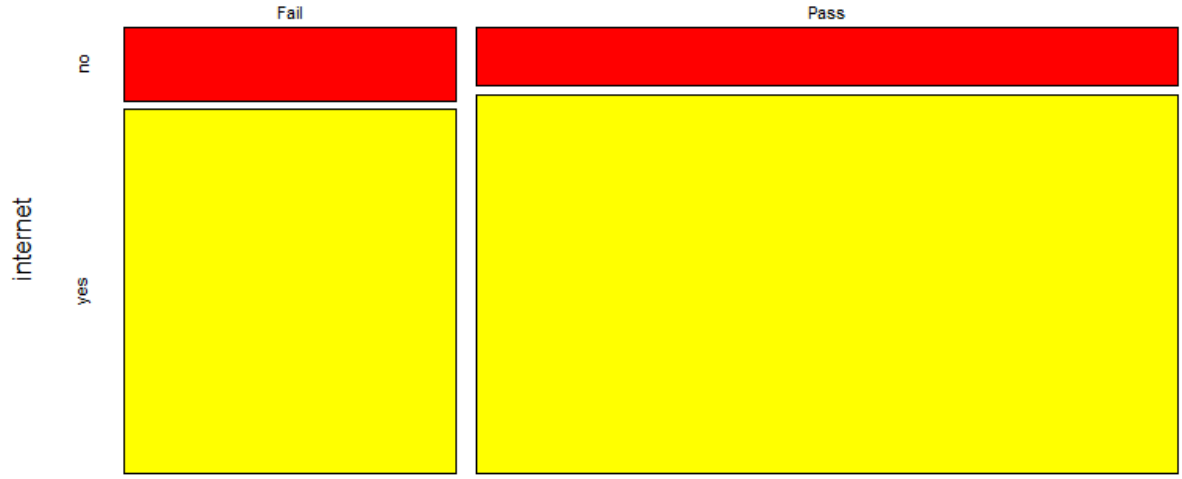

Fail
Pass
schoolsup
no
yes


Fail
Pass
famsup
no
yes

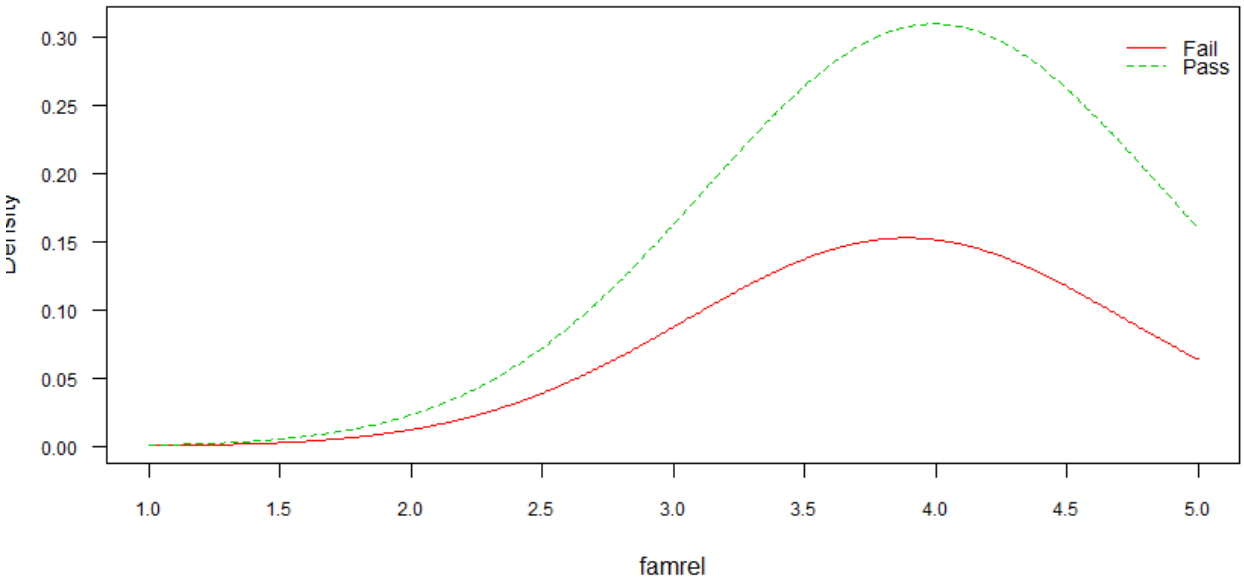

Fail
Pass
paid
no
yes

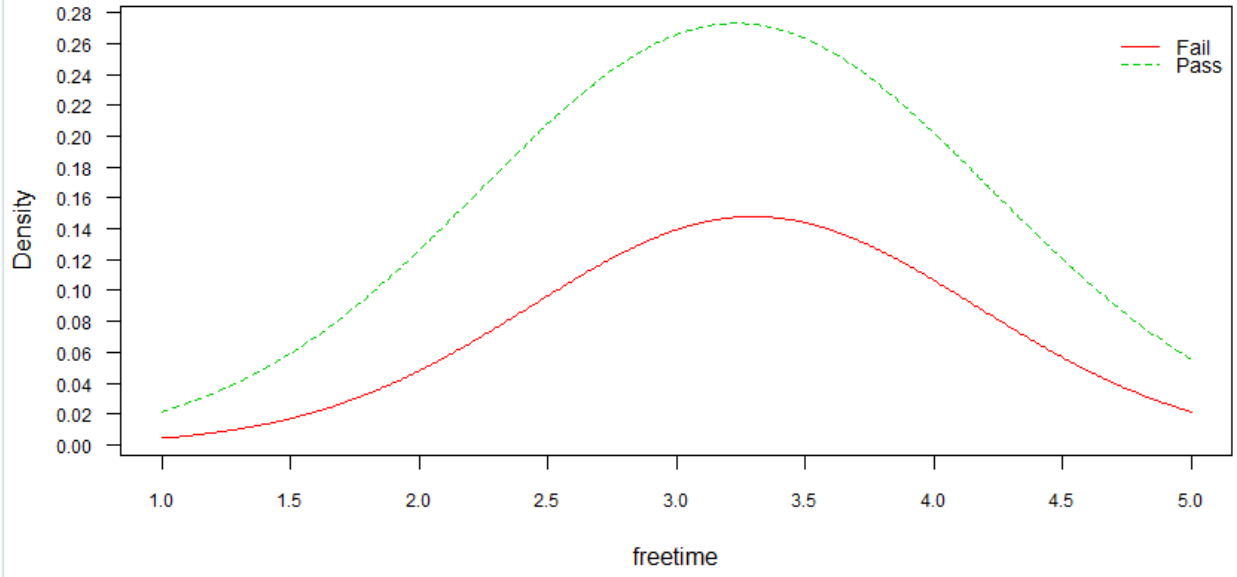

Fail
Pass
activities
no
yes

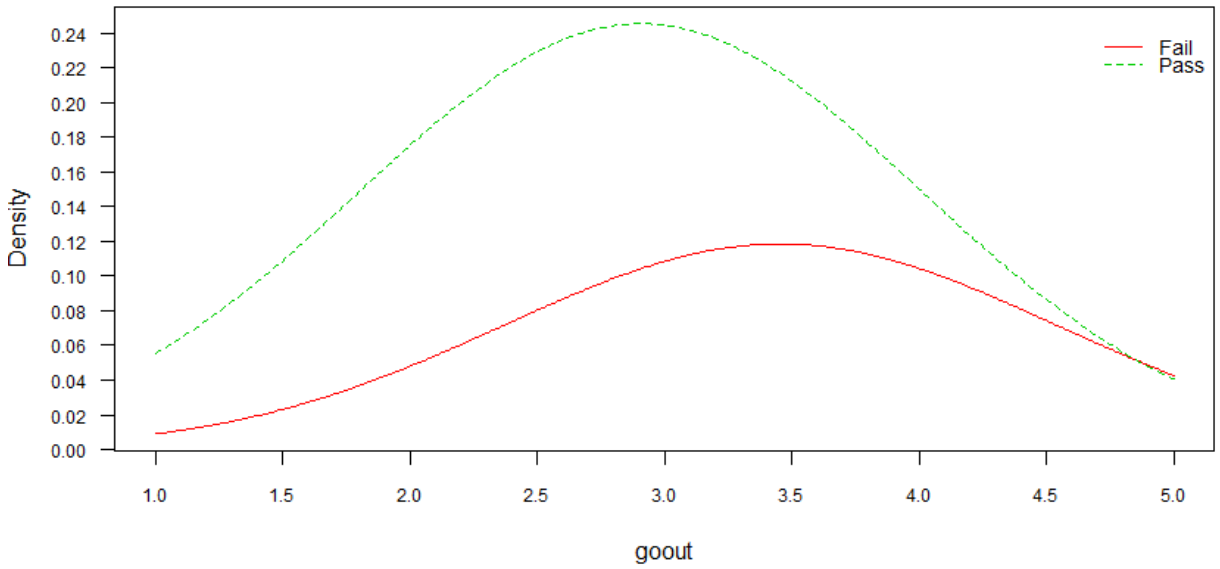

Fail
Pass
higher
no
yes

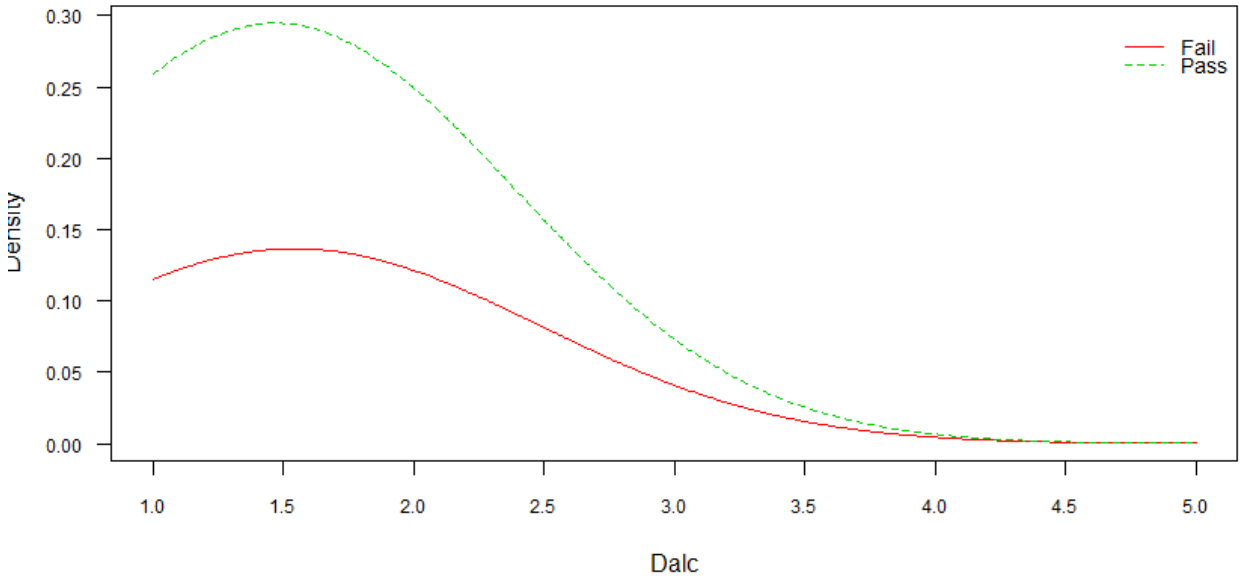

Fail
Pass
internet
no
yes

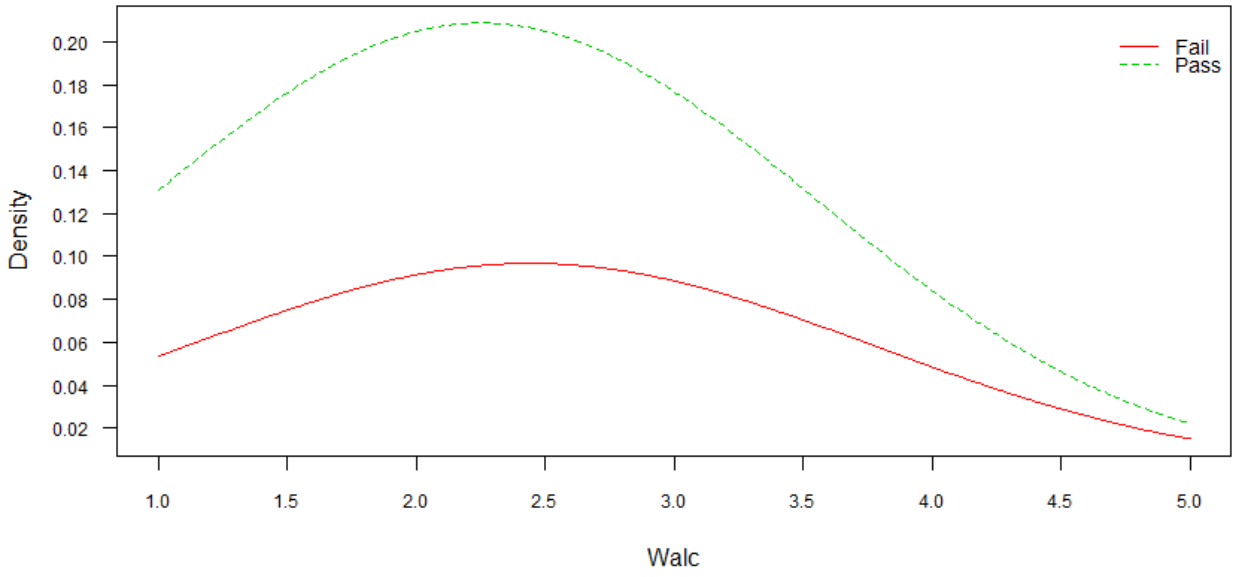

Fail
Pass
romantic
no
yes

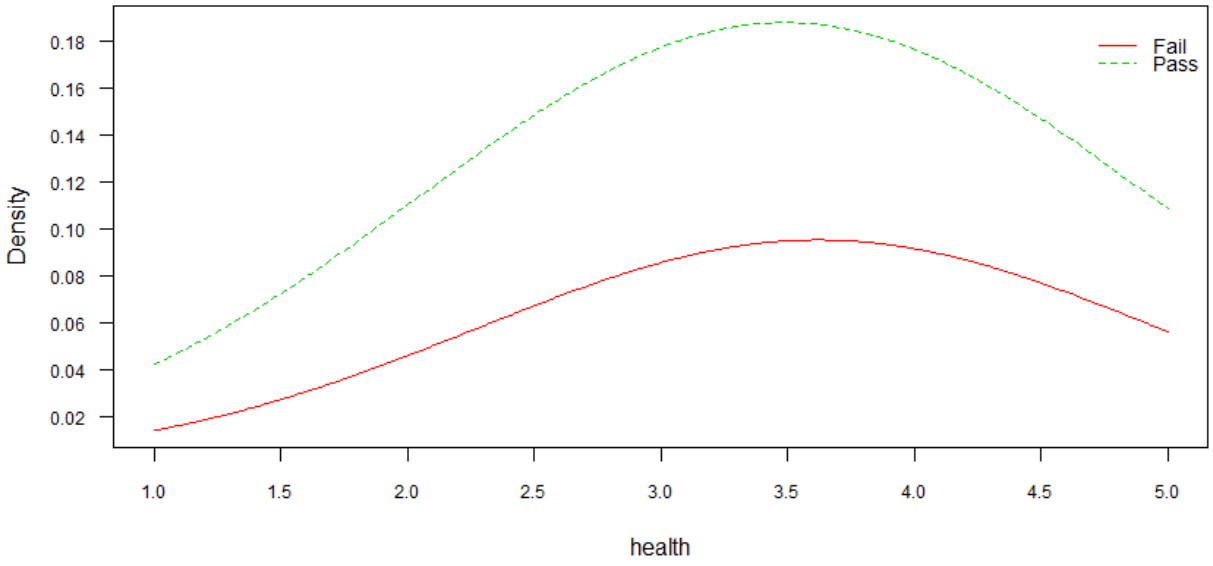

Density
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
famrel
Fail
Pass

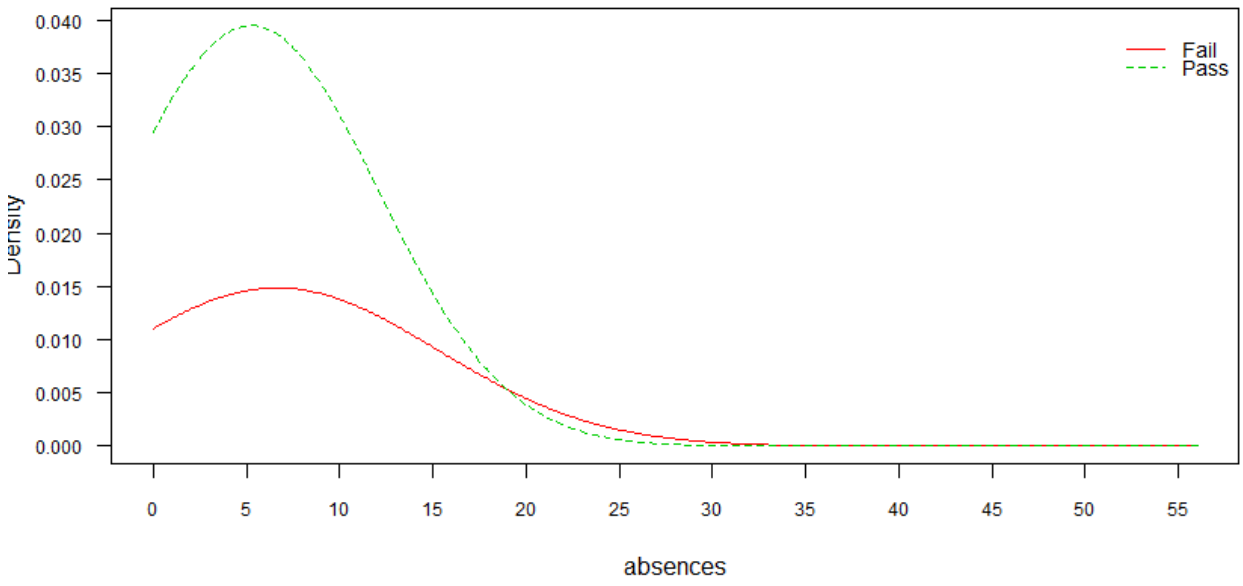

Density
0.28
0.26
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
freetime
Fail
Pass

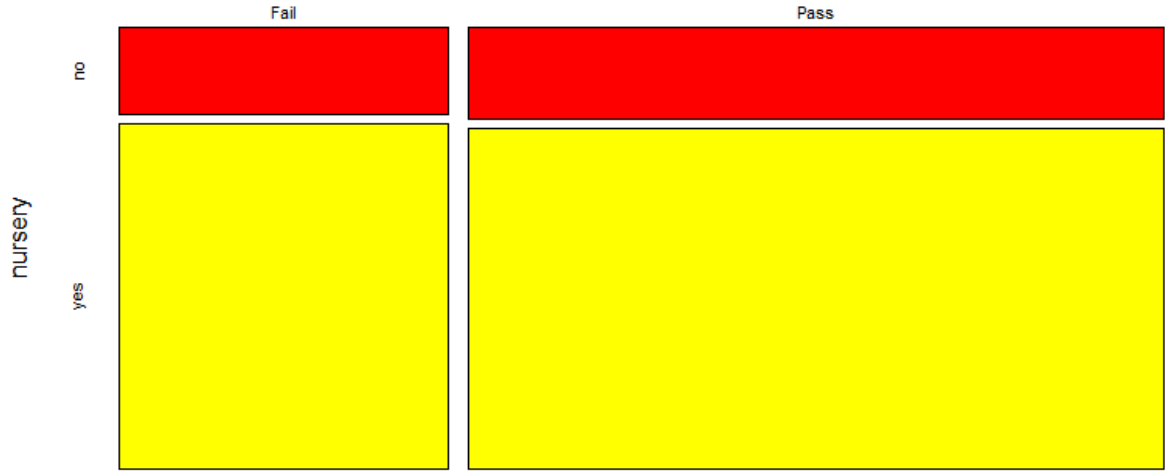

Density
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
goout
Fail
Pass


Density
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Dalc
Fail
Pass


Density
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Walc
Fail
Pass


Density
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
health
Fail
Pass


Density
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
Fail
Pass
0
5
10
15
20
25
30
35
40
45
50
55
absences


Fail
Pass
nursery
no
yes