

# Análisis de la influencia del consumo de alcohol en los resultados académicos utilizando aprendizaje automático

*José Antonio Yébenes Montenegro*

*Junio 2017*

## Contents

<b>Configuración del entorno de trabajo</b>	<b>3</b>
<b>Preparación de los datos</b>	<b>3</b>
<b>Análisis de la calidad de los datos</b>	<b>10</b>
Detección de nulls y na . . . . .	10
Detección de outliers . . . . .	11
<b>Interpretación de los datos</b>	<b>18</b>
<b>Análisis univariante y bivalente</b>	<b>39</b>
<b>Segmentación</b>	<b>50</b>
Segmentación óptima . . . . .	53
<b>Reglas de asociación</b>	<b>66</b>
<b>Profiling</b>	<b>110</b>
<b>Árboles de decisión</b>	<b>119</b>

El objetivo del trabajo es analizar si existe relación entre el consumo de alcohol y los resultados académicos en estudiantes de educación secundaria. Para ello se trabaja con un conjunto de datos que contiene información social y académica de un grupo de estudiantes de educación secundaria de edades comprendidas entre los 15 y los 22 años de dos escuelas Portuguesas.

El análisis de este conjunto de datos tiene como objetivos:

- **Analizar la calidad de los datos:** es importante asegurar que los datos son completos y coherentes antes de iniciar el estudio. En caso de existir valores faltantes se deberán tomar las medidas necesarias para regularizarlo.
- **Detectar si existen valores outliers:** es importante conocer si existen valores extremos ya que su presencia puede alterar los resultados obtenidos sobre todo si se aplican técnicas basadas en distancias y no se estandarizan los datos.
- **Entender los datos mediante visualizaciones sencillas:** para ver los valores que toma cada una de las variables y su distribución, se mostrarán gráficos de barras que permitirán entender mejor los datos que se están trabajando. Estas visualizaciones permitirán una mejor contextualización a la hora de interpretar los resultados obtenidos.
- **Entender los datos a través del análisis estadístico univariante y bivariante:** para entender los datos y poderlos interpretar correctamente se realizará un análisis univariante sencillo (summary). También se analizará si existen variables altamente correlacionadas y se buscarán relaciones entre las variables mediante visualizaciones sencillas (diagramas de puntos por pares de variables). Dependiendo del tipo de técnica de análisis que se utilice las variables muy correlacionadas habrá que excluirlas, de este modo se puede reducir la dimensionalidad y eliminar redundancias en los datos.
- **Segmentar los datos:** mediante la técnica de aprendizaje automático no supervisado Kmeans, se agruparán los estudiantes en grupos lo más diferentes posible de manera que dentro de cada grupo los estudiantes sean los más parecidos posible. Esto permitirá ver si el consumo de alcohol manifiesta comportamientos distintos en los diferentes grupos de estudiantes o por el contrario es parecido.
- **Detectar reglas de asociación:** a través de esta técnica de aprendizaje automático no supervisado se buscarán relaciones entre las variable y ciertos valores del estilo “cuando pasa A también pasa B”. Este tipo de reglas no denotan una relación de causa-efecto, detectan situaciones que se dan a la vez pero no necesariamente una es la causa de la otra. Interesará detectar relaciones en las que en A aparezcan variables relacionadas con el consumo de alcohol y donde en B aparezca la variable de éxito (superar o no la asignatura).
- **Detectar las variables más influyentes en el hecho de superar o no la asignatura:** usando la técnica de profiling (descripción de características) se estudiará qué variables son las más influyentes sobre la variable de éxito/fracaso (superar o no la asignatura). Esto ayudará a determinar si el consumo del alcohol es la más influyente o existen otros factores más importantes.
- **Construir un modelo predictivo sobre la variable objetivo (superar o no la asignatura):** usando los árboles de decisión (técnica de aprendizaje automático supervisado) se pretende entrenar y evaluar un modelo predictivo. Este modelo alertará cuando un estudiante no va a superar la asignatura, prestando especial atención si uno de los factores influyentes es el consumo de alcohol.

## Configuración del entorno de trabajo

```
# Establecer el directorio de trabajo

# linux
setwd("/home/j82/Dropbox/UOC_Master/TFM/1_R/")

# windows
#setwd("C:/Users/JoséAntonio/Dropbox/UOC_Master/TFM/1_R")
getwd()

## [1] "/home/j82/Dropbox/UOC_Master/TFM/1_R"

#save.image("myspace_TFM.RData")
load("myspace_TFM.RData")
```

## Preparacion de los datos

```
# cargar ficheros
d1<-read.table("student-mat.csv",sep=",",header=TRUE)
d2<-read.table("student-por.csv",sep=",",header=TRUE)

# añadir marca de fichero
d1$file<-"math"
d2$file<-"portuguese"

# añadir marca de asignatura
d1$subject<-"math"
d2$subject<-"portuguese"

# combinar los dos ficheros
data<-rbind(d1,d2)

# verificar si hay duplicados
nrow(data)

## [1] 1044

nrow(unique(data))

## [1] 1044

# no hay registros repetidos

# normalizar nombres de las variables
colnames(data) <- tolower(colnames(data))

# crear un identificador único de estudiante
data$student_key<-paste(data$school, data$sex, data$age, data$address, data$famsize, data$pstatus, data$

# crear una variable binaria para las notas finales
data$bing3 <- ifelse(data$g3>=10, 1, 0)
```

```

# crear una variable que indique si se aprueba una asignatura o la otra
data$result <- paste(iffelse(data$g3>=10, "pass", "fail"), data$subject, sep=" ")

# crear una variable que mida el consumo de alcohol global como media ponderada por el número de días
data$alcohol <- round((data$dalc*5 + data$walc*2)/7,2)

# estudiantes repetidos (están en las dos clases)
duplicated_rows <- duplicated(data$student_key)
data_duplicated <- data[duplicated_rows,]
nrow(data_duplicated)

## [1] 382

# hay 382 estudiantes que están en ambos ficheros tal como indica en la documentación de los ficheros

# función para informar que estos estudiantes están en las dos asignaturas
add_duplicates_subject <- function(duplicated){

  #data$school, data$sex, data$age, data$address, data$famsize, data$pstatus, data$medu, data$fedu, dat

  rows <- data$school == duplicated$school &
  data$sex == duplicated$sex &
  data$age == duplicated$age &
  data$address == duplicated$address &
  data$famsize == duplicated$famsize &
  data$pstatus == duplicated$pstatus &
  data$medu == duplicated$medu &
  data$fedu == duplicated$fedu &
  data$mjob == duplicated$mjob &
  data$fjob == duplicated$fjob &
  data$reason == duplicated$reason &
  data$nursery == duplicated$nursery &
  data$internet == duplicated$internet
  # actualizar asignatura
  data[rows,"subject"] <- "math&portuguese"
}

# aplicar la función a los alumnos repetidos en los dos ficheros
for(i in 1:nrow(data_duplicated)){
  add_duplicates_subject(data_duplicated[i,])
}

# guardamos un backup de los datos
raw_data <- data

# codificación de las variables categóricas
data$school <- factor(data$school, labels=c("Gabriel Pereira", "Mousinho da Silveira"))

data$sex <- factor(data$sex, labels=c("female", "male"))

data$address <- factor(data$address, labels=c("rural", "urban"))

```

```

data$famsize <- factor(data$famsize, labels=c("greater than 3","less or equal to 3"))
data$pstatus <- factor(data$pstatus, labels=c("living apart","living together"))
data$medu <- factor(data$medu, labels=c("none","primary", "5th-9th grade", "secondary", "higher") )
data$fedu <- factor(data$fedu, labels=c("none","primary", "5th-9th grade", "secondary", "higher") )
data$traveltime <- factor(data$traveltime, labels=c("<15 min.,"15 to 30 min.", "30 min to 1 hour", ">1
data$studytime <- factor(data$studytime, labels=c("<2 hours","2 to 5 hours", "5 to 10 hours", ">10 hours
data$famrel <- factor(data$famrel, labels=c("very bad","bad", "fair", "good", "excellent"))
data$freetime <- factor(data$freetime, labels=c("very low", "low", "medium", "high", "very high"))
data$goout <- factor(data$goout, labels=c("very low", "low", "medium", "high", "very high"))
data$dalc <- factor(data$dalc, labels=c("very low", "low", "medium", "high", "very high"))
data$walc <- factor(data$walc, labels=c("very low", "low", "medium", "high", "very high"))
data$health <- factor(data$health, labels=c("very bad","bad", "fair", "good", "very good"))
data$file <- as.factor(data$file)
data$subject <- as.factor(data$subject)
data$bing3 <- factor(data$bing3, labels=c("fail","pass"))
data$result <- as.factor(data$result)

str(data)

```

```

## 'data.frame': 1044 obs. of 39 variables:
## $ school : Factor w/ 2 levels "Gabriel Pereira",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ sex : Factor w/ 2 levels "female","male": 1 1 1 1 1 2 2 1 2 2 ...
## $ age : int 18 17 15 15 16 16 16 17 15 15 ...
## $ address : Factor w/ 2 levels "rural","urban": 2 2 2 2 2 2 2 2 2 2 ...
## $ famsize : Factor w/ 2 levels "greater than 3",...: 1 1 2 1 1 2 2 1 2 1 ...
## $ pstatus : Factor w/ 2 levels "living apart",...: 1 2 2 2 2 2 2 1 1 2 ...
## $ medu : Factor w/ 5 levels "none","primary",...: 5 2 2 5 4 5 3 5 4 4 ...
## $ fedu : Factor w/ 5 levels "none","primary",...: 5 2 2 3 4 4 3 5 3 5 ...
## $ mjob : Factor w/ 5 levels "at_home","health",...: 1 1 1 2 3 4 3 3 4 3 ...
## $ fjob : Factor w/ 5 levels "at_home","health",...: 5 3 3 4 3 3 3 5 3 3 ...
## $ reason : Factor w/ 4 levels "course","home",...: 1 1 3 2 2 4 2 2 2 2 ...
## $ guardian : Factor w/ 3 levels "father","mother",...: 2 1 2 2 1 2 2 2 2 2 ...
## $ traveltime : Factor w/ 4 levels "<15 min.,"15 to 30 min.",...: 2 1 1 1 1 1 1 2 1 1 ...
## $ studytime : Factor w/ 4 levels "<2 hours","2 to 5 hours",...: 2 2 2 3 2 2 2 2 2 2 ...
## $ failures : int 0 0 3 0 0 0 0 0 0 0 ...
## $ schoolsup : Factor w/ 2 levels "no","yes": 2 1 2 1 1 1 1 2 1 1 ...
## $ famsup : Factor w/ 2 levels "no","yes": 1 2 1 2 2 2 1 2 2 2 ...
## $ paid : Factor w/ 2 levels "no","yes": 1 1 2 2 2 2 1 1 2 2 ...

```

```

## $ activities : Factor w/ 2 levels "no","yes": 1 1 1 2 1 2 1 1 1 2 ...
## $ nursery   : Factor w/ 2 levels "no","yes": 2 1 2 2 2 2 2 2 2 2 ...
## $ higher    : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2 2 ...
## $ internet  : Factor w/ 2 levels "no","yes": 1 2 2 2 1 2 2 1 2 2 ...
## $ romantic  : Factor w/ 2 levels "no","yes": 1 1 1 2 1 1 1 1 1 1 ...
## $ famrel    : Factor w/ 5 levels "very bad","bad",...: 4 5 4 3 4 5 4 4 4 5 ...
## $ freetime  : Factor w/ 5 levels "very low","low",...: 3 3 3 2 3 4 4 1 2 5 ...
## $ goout     : Factor w/ 5 levels "very low","low",...: 4 3 2 2 2 2 4 4 2 1 ...
## $ dalc      : Factor w/ 5 levels "very low","low",...: 1 1 2 1 1 1 1 1 1 1 ...
## $ walc      : Factor w/ 5 levels "very low","low",...: 1 1 3 1 2 2 1 1 1 1 ...
## $ health    : Factor w/ 5 levels "very bad","bad",...: 3 3 3 5 5 5 3 1 1 5 ...
## $ absences  : int 6 4 10 2 4 10 0 6 0 0 ...
## $ g1        : int 5 5 7 15 6 15 12 6 16 14 ...
## $ g2        : int 6 5 8 14 10 15 12 5 18 15 ...
## $ g3        : int 6 6 10 15 10 15 11 6 19 15 ...
## $ file      : Factor w/ 2 levels "math","portuguese": 1 1 1 1 1 1 1 1 1 1 ...
## $ subject   : Factor w/ 3 levels "math","math&portuguese",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ student_key: chr "GP-F-18-U-GT3-A-4-4-at_home-teacher-course-yes-no" "GP-F-17-U-GT3-T-1-1-at_hom
## $ bing3     : Factor w/ 2 levels "fail","pass": 1 1 2 2 2 2 1 2 2 ...
## $ result    : Factor w/ 4 levels "fail math","fail portuguese",...: 1 1 3 3 3 3 1 3 3 ...
## $ alcohol   : num 1 1 2.29 1 1.29 1.29 1 1 1 1 ...

```

```
summary(data)
```

```

##          school          sex          age          address
## Gabriel Pereira      :772 female:591   Min.    :15.00   rural:285
## Mousinho da Silveira:272 male   :453   1st Qu.:16.00   urban:759
##                                     Median :17.00
##                                     Mean    :16.73
##                                     3rd Qu.:18.00
##                                     Max.    :22.00
##          famsize          pstatus          medu
## greater than 3      :738   living apart  :121   none          : 9
## less or equal to 3:306   living together:923   primary       :202
##                                     5th-9th grade:289
##                                     secondary     :238
##                                     higher        :306
##
##          fedu          mjob          fjob          reason
## none          : 9   at_home :194   at_home : 62   course       :430
## primary       :256   health  : 82   health  : 41   home         :258
## 5th-9th grade:324   other   :399   other   :584   other        :108
## secondary     :231   services:239   services:292   reputation:248
## higher        :224   teacher :130   teacher : 65
##
##          guardian          traveltime          studytime          failures
## father:243 <15 min.      :623 <2 hours   :317   Min.    :0.0000
## mother:728 15 to 30 min. :320 2 to 5 hours :503   1st Qu.:0.0000
## other : 73 30 min to 1 hour: 77 5 to 10 hours:162   Median :0.0000
##                                     >1 hour   : 24 >10 hours   : 62   Mean    :0.2644
##                                     Max.     :3.0000
## schoolsup famsup   paid   activities nursery   higher   internet
## no :925 no :404 no :824 no :528 no :209 no : 89 no :217
## yes:119 yes:640 yes:220 yes:516 yes:835 yes:955 yes:827

```

```

##
##
##
##
## romantic      famrel      freetime      goout      dalc
## no :673      very bad : 30      very low : 64      very low : 71      very low :727
## yes:371      bad      : 47      low      :171      low      :248      low      :196
##              fair      :169      medium   :408      medium   :335      medium   : 69
##              good      :512      high     :293      high     :227      high     : 26
##              excellent:286      very high:108      very high:163      very high: 26
##
##          walc          health          absences          g1
## very low :398      very bad :137      Min.      : 0.000      Min.      : 0.00
## low      :235      bad      :123      1st Qu.: 0.000      1st Qu.: 9.00
## medium   :200      fair     :215      Median   : 2.000      Median   :11.00
## high     :138      good     :174      Mean     : 4.435      Mean     :11.21
## very high: 73      very good:395      3rd Qu.: 6.000      3rd Qu.:13.00
##              Max.      :75.000      Max.      :19.00
##
##          g2          g3          file          subject
## Min.      : 0.00      Min.      : 0.00      math      :395      math      : 25
## 1st Qu.: 9.00      1st Qu.:10.00      portuguese:649      math&portuguese:751
## Median   :11.00      Median   :11.00      portuguese :268
## Mean     :11.25      Mean     :11.34
## 3rd Qu.:13.00      3rd Qu.:14.00
## Max.     :19.00      Max.     :20.00
##
## student_key      bing3          result          alcohol
## Length:1044      fail:230      fail math      :130      Min.      :1.000
## Class :character      pass:814      fail portuguese:100      1st Qu.:1.000
## Mode  :character      pass math      :265      Median   :1.290
##              pass portuguese:549      Mean     :1.721
##              3rd Qu.:2.178
##              Max.     :5.000

```

```

variables<-data.frame(names(data), "", sapply(data, class), 0, 0, "")
names(variables)<-c("variable", "descripcion", "tipo", "cuenta_null", "cuenta_na", "color")

variables$descripcion<-as.character(variables$descripcion)

variables$descripcion[1]="Student's school"
variables$descripcion[2]="Student's sex"
variables$descripcion[3]="Student's age"
variables$descripcion[4]="Student's home address type"
variables$descripcion[5]="Family size"
variables$descripcion[6]="Parent's cohabitation status"
variables$descripcion[7]="Mother's education"
variables$descripcion[8]="Father's education"
variables$descripcion[9]="Mother's job"
variables$descripcion[10]="Father's job"
variables$descripcion[11]="Reason to choose this school"
variables$descripcion[12]="Student's guardian"
variables$descripcion[13]="Home to school travel time"
variables$descripcion[14]="Weekly study time"
variables$descripcion[15]="Number of past class failures"

```

```

variables$descripcion[16]="Extra educational support"
variables$descripcion[17]="Family educational support"
variables$descripcion[18]="Extra paid classes within the course subject"
variables$descripcion[19]="Extra-curricular activities"
variables$descripcion[20]="Attended nursery school"
variables$descripcion[21]="Wants to take higher education"
variables$descripcion[22]="Internet access at home"
variables$descripcion[23]="With a romantic relationship"
variables$descripcion[24]="Quality of family relationships"
variables$descripcion[25]="Free time after school"
variables$descripcion[26]="Going out with friends"
variables$descripcion[27]="Workday alcohol consumption"
variables$descripcion[28]="Weekend alcohol consumption"
variables$descripcion[29]="Current health status"
variables$descripcion[30]="Number of school absences"
variables$descripcion[31]="First period grade"
variables$descripcion[32]="Second period grade"
variables$descripcion[33]="Final grade"
variables$descripcion[34]="File"
variables$descripcion[35]="Subject"
variables$descripcion[36]="Primary key for student"
variables$descripcion[37]="Binary final grade result"
variables$descripcion[38]="Subject and result"
variables$descripcion[39]="Global alcohol consumption"

```

*# color asignado a cada variable*

```

variables$color<-as.character(variables$color)
variables$color[1]="doscolores"
variables$color[2]="doscolores"
variables$color[3]="rainbow"
variables$color[4]="doscolores"
variables$color[5]="doscolores"
variables$color[6]="doscolores"
variables$color[7]="semaforo"
variables$color[8]="semaforo"
variables$color[9]="rainbow"
variables$color[10]="rainbow"
variables$color[11]="rainbow"
variables$color[12]="rainbow"
variables$color[13]="rainbow"
variables$color[14]="semaforo"
variables$color[15]="semaforo_inverso"
variables$color[16]="noyes"
variables$color[17]="noyes"
variables$color[18]="noyes"
variables$color[19]="noyes"
variables$color[20]="noyes"
variables$color[21]="noyes"
variables$color[22]="noyes"
variables$color[23]="noyes"
variables$color[24]="semaforo"
variables$color[25]="semaforo"

```



```

variables$color[26]="semaforo"
variables$color[27]="semaforo_inverso"
variables$color[28]="semaforo_inverso"
variables$color[29]="semaforo"
variables$color[30]="heat_inverso"
variables$color[31]="heat"
variables$color[32]="heat"
variables$color[33]="heat"
variables$color[34]="doscolores"
variables$color[35]="rainbow"
variables$color[36]=""
variables$color[37]="noyes"
variables$color[38]="semaforo"
variables$color[39]="heat_inverso"

```

```

variables_numericas<-c(as.character(variables[variables$tipo==c("integer"),1]),
as.character(variables[variables$tipo==c("numeric"),1]))
length(variables_numericas)

```

```
## [1] 7
```

```
variables_numericas
```

```
## [1] "age"      "failures" "absences" "g1"      "g2"      "g3"
## [7] "alcohol"
```

```
# variables no caracter
```

```

variables_discretas<-as.character(variables[variables$tipo!=c("character"),1])
variables_discretas_names<-as.character(variables[variables$tipo!=c("character"),2])
variables_discretas_color<-as.character(variables[variables$tipo!=c("character"),6])
length(variables_discretas)

```

```
## [1] 38
```

```
variables_discretas
```

```

## [1] "school"    "sex"      "age"      "address"  "famsize"
## [6] "pstatus"  "medu"     "fedu"     "mjob"     "fjob"
## [11] "reason"   "guardian" "traveltime" "studytime" "failures"
## [16] "schoolsup" "famsup"   "paid"     "activities" "nursery"
## [21] "higher"   "internet" "romantic"  "famrel"   "freetime"
## [26] "goout"    "dalc"     "walc"     "health"   "absences"
## [31] "g1"       "g2"       "g3"       "file"     "subject"
## [36] "bing3"    "result"   "alcohol"

```

```
# se eliminan los objetos que ya no serán necesarios
```

```

rm(d1)
rm(d2)
rm(duplicated_rows)
rm(i)
#write.csv(data, file="data.csv")
save.image("myworkspace_TFM.RData")

```

# Análisis de la calidad de los datos

## Detección de nulls y na

```
for(i in 1:length(variables$variable)){
  var<-paste(variables$variable[i])
  variables$cuenta_null[i]<-sum(sapply(data[var], function(x) sum(is.null(x))))
  variables$cuenta_na[i]<-sum(sapply(data[var], function(x) sum(is.na(x))))
}
```

```
# vemos si hay nullos o na
nrow(variables[variables$cuenta_null>0,])
```

```
## [1] 0
```

```
nrow(variables[variables$cuenta_na>0,])
```

```
## [1] 0
```

```
# De forma alternativa:
sapply (data, function(x) (sum(is.na(x))))
```

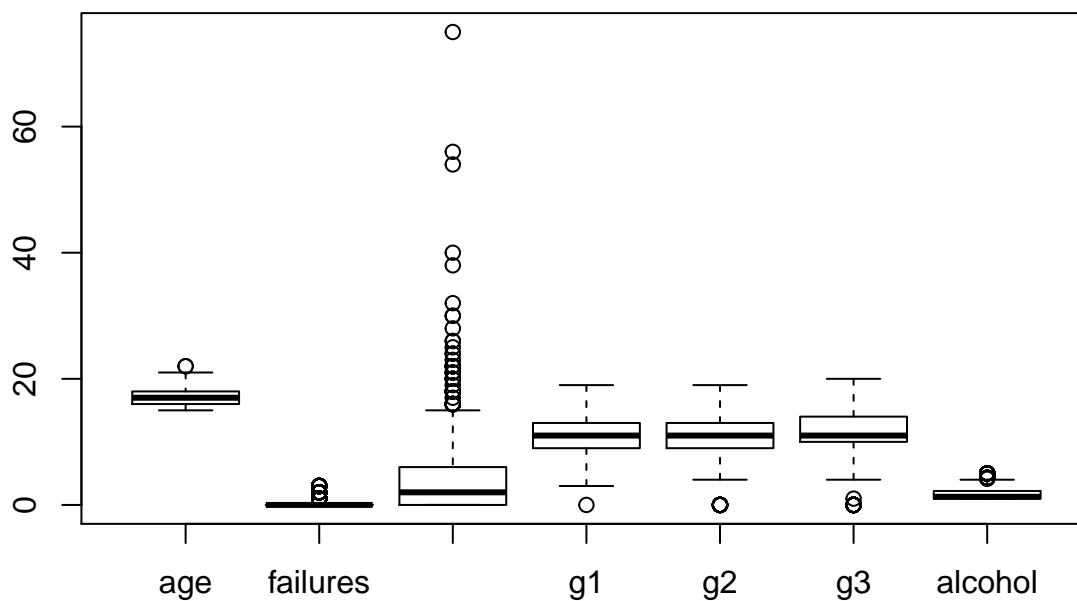
```
##      school      sex      age      address      famsize      pstatus
##      0          0          0          0          0          0
##      medu      fedu      mjob      fjob      reason      guardian
##      0          0          0          0          0          0
##      travelttime  studytime  failures  schoolsup  famsup      paid
##      0          0          0          0          0          0
##      activities  nursery      higher    internet  romantic    famrel
##      0          0          0          0          0          0
##      freetime    goout      dalc      walc      health      absences
##      0          0          0          0          0          0
##      g1          g2          g3          file      subject    student_key
##      0          0          0          0          0          0
##      bing3      result      alcohol
##      0          0          0
```

```
sapply (data, function(x) (sum(is.null(x))))
```

```
##      school      sex      age      address      famsize      pstatus
##      0          0          0          0          0          0
##      medu      fedu      mjob      fjob      reason      guardian
##      0          0          0          0          0          0
##      travelttime  studytime  failures  schoolsup  famsup      paid
##      0          0          0          0          0          0
##      activities  nursery      higher    internet  romantic    famrel
##      0          0          0          0          0          0
##      freetime    goout      dalc      walc      health      absences
##      0          0          0          0          0          0
##      g1          g2          g3          file      subject    student_key
##      0          0          0          0          0          0
##      bing3      result      alcohol
##      0          0          0
```

## Detección de outliers

```
boxplot(data[variables_numericas])
```



```
# hay valores atípicos en todas las variables numéricas
```

```
# se detectan y eliminan los atípicos en un nuevo dataframe
```

```
# primera iteración
```

```
data_clean1 <- data  
dim(data)
```

```
## [1] 1044 39
```

```
eliminados1 <- 0
```

```
for(i in 1:length(variables_numericas)){
```

```
  variable<-data_clean1[variables_numericas][[i]]  
  # usamos boxplot.stats para ver los valores outliers
```

```
  outliers<-boxplot.stats(variable)$out
```

```
  index <- which( variable %in% outliers)
```

```
  eliminados1<-eliminados1+length(index)
```

```
  # eliminamos los registros que contienen outliers
```

```
  if(length(index)>0) {  
    data_clean1<-data_clean1[-index,]
```

```
  }
```

```
}
```

```
eliminados1
```

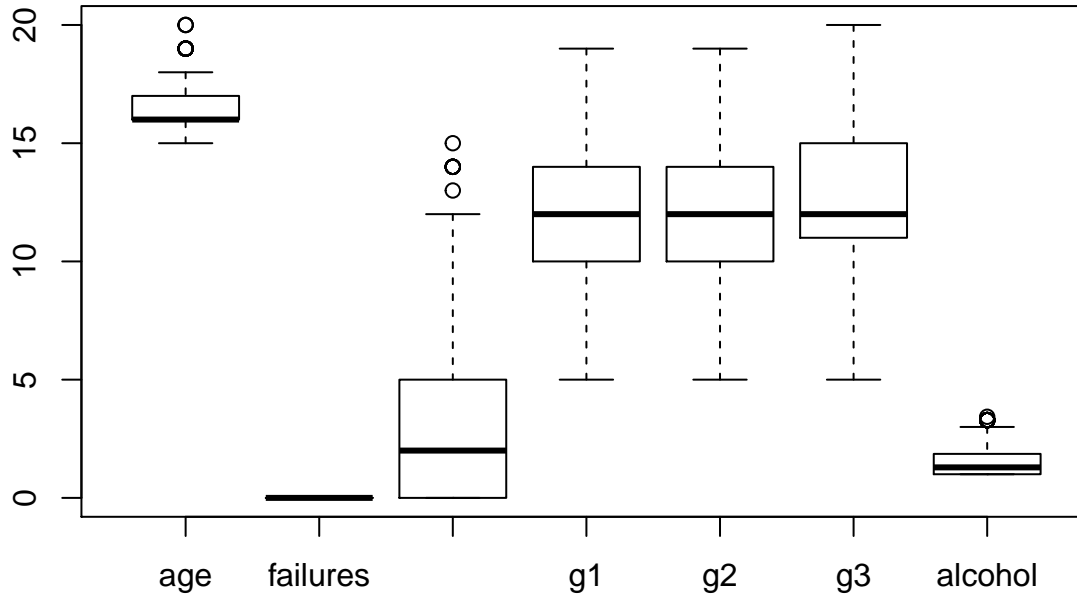
```
## [1] 283
```

```
dim(data_clean1)
```

```
## [1] 761 39
```

```
# vemos si han desaparecido todos los valores atípicos
```

```
boxplot(data_clean1[variables_numericas])
```



```
# aún hay valores atípicos
```

```
# segunda iteración
```

```
data_clean2 <- data_clean1
```

```
eliminados2 <- 0
```

```
for(i in 1:length(variables_numericas)){
```

```
  variable<-data_clean2[variables_numericas][[i]]
```

```
  # usamos boxplot.stats para ver los valores outliers
```

```
  outliers<-boxplot.stats(variable)$out
```

```
  index <- which( variable %in% outliers)
```

```
  eliminados2<-eliminados2+length(index)
```

```
  # eliminamos los registros que contienen outliers
```

```
  if(length(index)>0) {
```

```
    data_clean2<-data_clean2[-index,]
```

```
  }
```

```
}
```

```
eliminados2
```

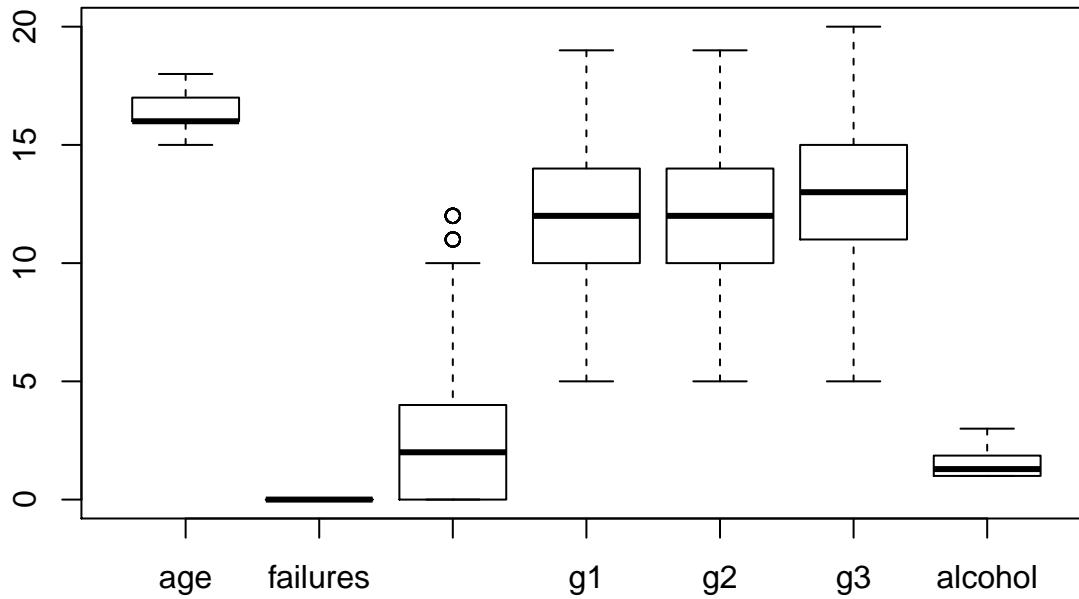
```
## [1] 48
```

```
dim(data_clean2)
```

```
## [1] 713 39
```

```
# vemos si han desaparecido todos los valores atípicos
```

```
boxplot(data_clean2[variables_numericas])
```



```
# aún hay valores atípicos

# tercera iteración
data_clean3 <- data_clean2

eliminados3 <-0

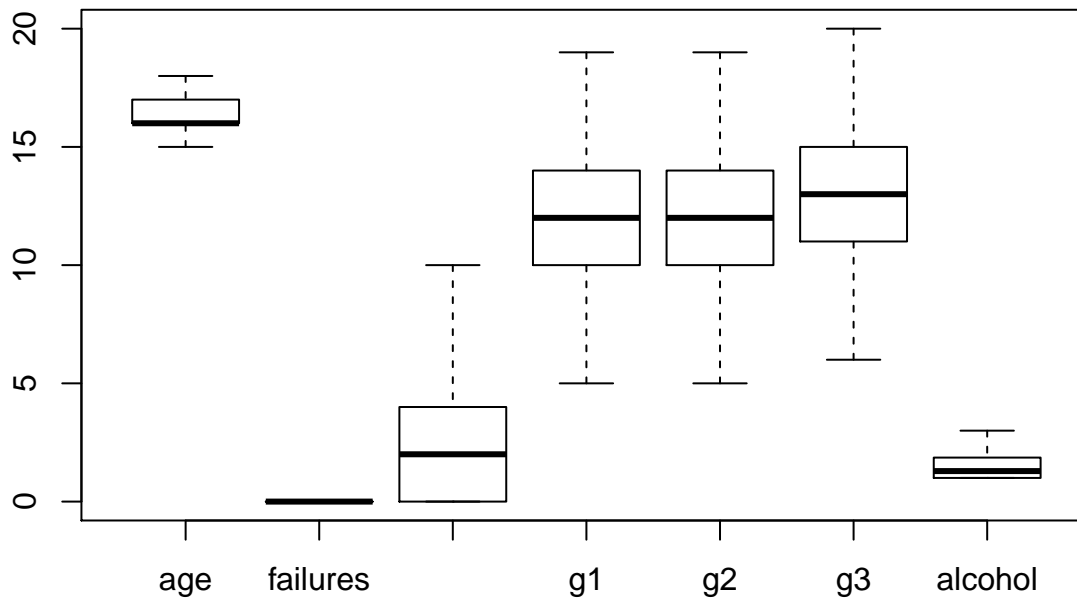
for(i in 1:length(variables_numericas)){

  variable<-data_clean3[variables_numericas][[i]]
  # usamos boxplot.stats para ver los valores outliers
  outliers<-boxplot.stats(variable)$out
  index <- which( variable %in% outliers)
  eliminados3<-eliminados3+length(index)
  # eliminamos los registros que contienen outliers
  if(length(index)>0) {
    data_clean3<-data_clean3[-index,]
  }
}

eliminados3

## [1] 18
dim(data_clean3)

## [1] 695 39
# vemos si han desaparecido todos los valores atípicos
boxplot(data_clean3[variables_numericas])
```



```
eliminados1+eliminados2+eliminados3
```

```
## [1] 349
```

```
# aún hay valores atípicos
```

```
# cuarta iteración
```

```
data_clean4 <- data_clean3
```

```
eliminados4 <-0
```

```
for(i in 1:length(variables_numericas)){
```

```
  variable<-data_clean3[variables_numericas][[i]]
```

```
  # usamos boxplot.stats para ver los valores outliers
```

```
  outliers<-boxplot.stats(variable)$out
```

```
  index <- which( variable %in% outliers)
```

```
  eliminados4<-eliminados4+length(index)
```

```
  # eliminamos los registros que contienen outliers
```

```
  if(length(index)>0) {
    data_clean4<-data_clean4[-index,]
```

```
  }
```

```
}
```

```
eliminados4
```

```
## [1] 0
```

```
dim(data_clean4)
```

```
## [1] 695 39
```

```
# vemos si han desaparecido todos los valores atípicos
```

```
boxplot(data_clean4[variables_numericas])
```

```
# después de 4 iteraciones aún sigue detectando outliers en la variable ausencias
```

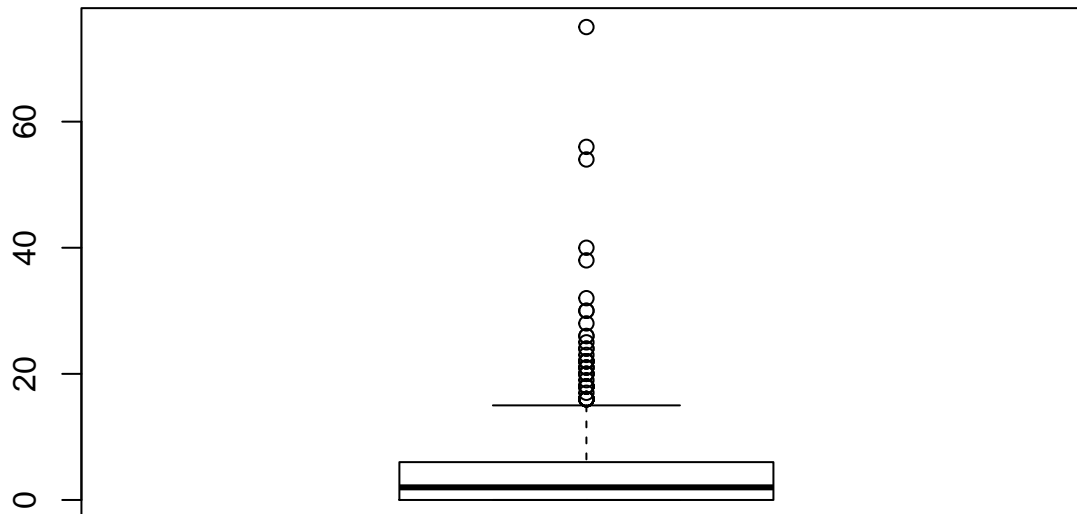
Vamos a tratar sólo los outliers sobre la variable absences.

```
data_clean <- data
```

```
summary(data$absences)
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000  0.000   2.000   4.435  6.000  75.000
```

```
boxplot(data$absences)
```



```
variable<-data_clean$absences
# usamos boxplot.stats para ver los valores outliers
outliers<-boxplot.stats(variable)$out
outliers
```

```
## [1] 16 16 25 54 18 26 20 18 16 16 56 24 18 28 22 16 18 20 16 21 75 22 30
## [24] 19 20 38 18 20 22 40 23 16 17 16 16 24 22 16 32 16 16 30 21 16 18 16
## [47] 26 16 16 22 18 18 16 21
```

```
table(outliers)
```

```
## outliers
## 16 17 18 19 20 21 22 23 24 25 26 28 30 32 38 40 54 56 75
## 17  1  8  1  4  3  5  1  2  1  2  1  2  1  1  1  1  1  1  1
```

```
index <- which( variable %in% outliers)
length(index)
```

```
## [1] 54
```

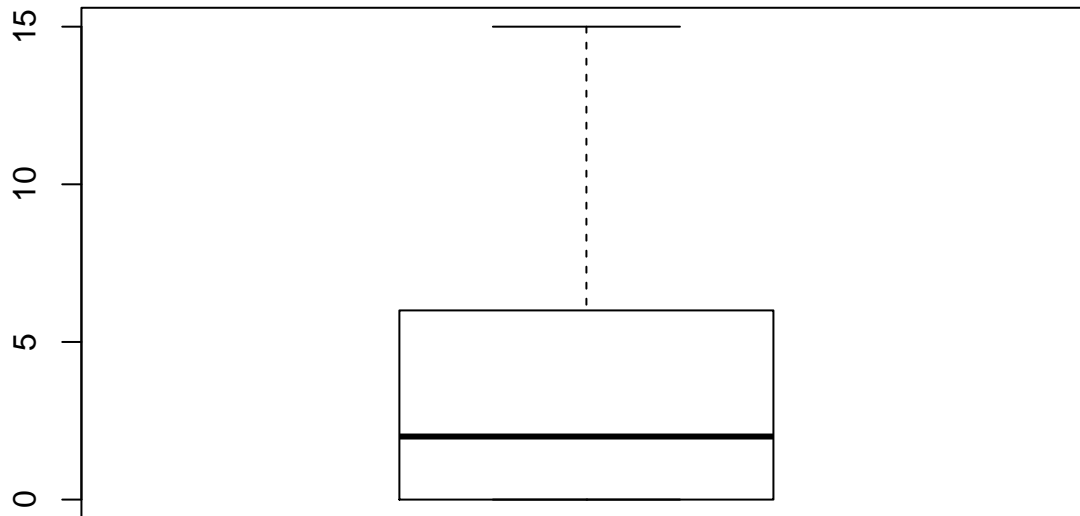
```
# eliminamos los registros que contienen outliers
data_clean<-data_clean[-index,]
```

```
# ha eliminado todos los registros con más de 15 ausencias
```

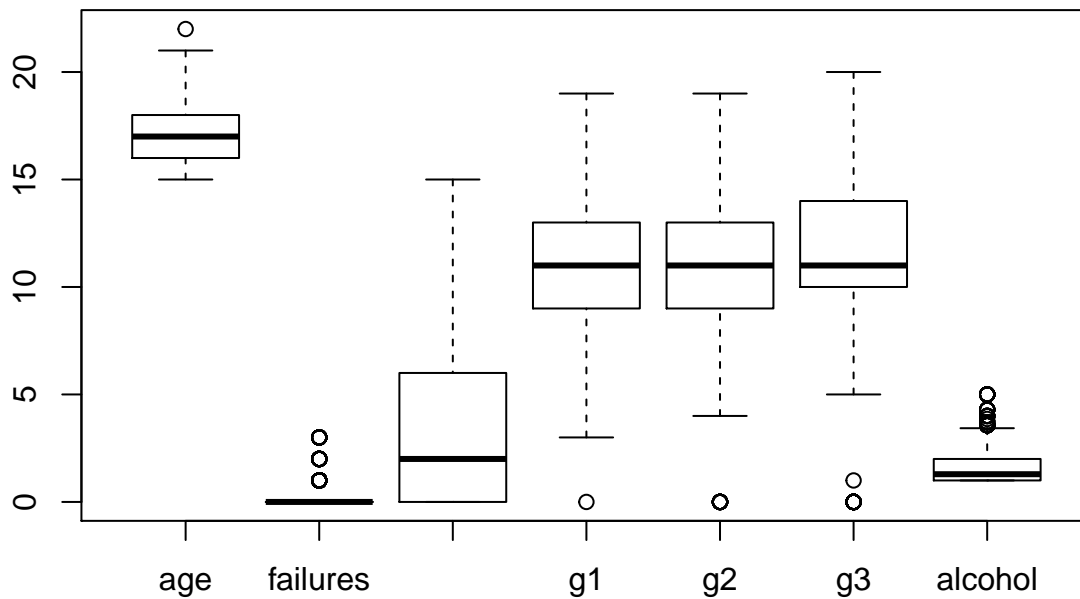
```
dim(data[data$absences>15,])
```

```
## [1] 54 39
```

```
# ya no hay outliers en la variable absences
boxplot(data_clean$absences)
```



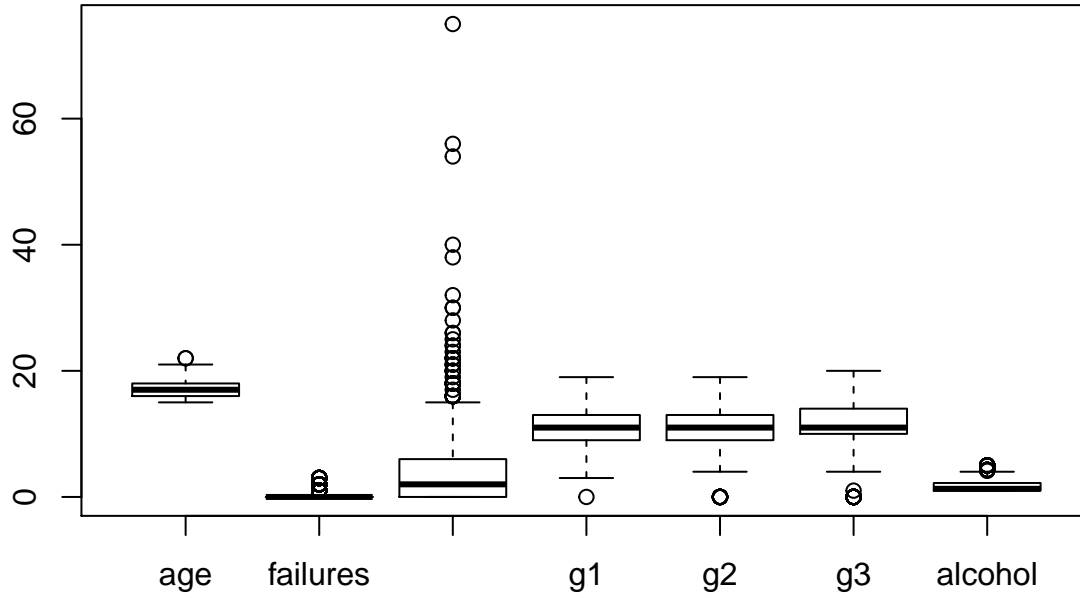
```
boxplot(data_clean[variables_numericas])
```



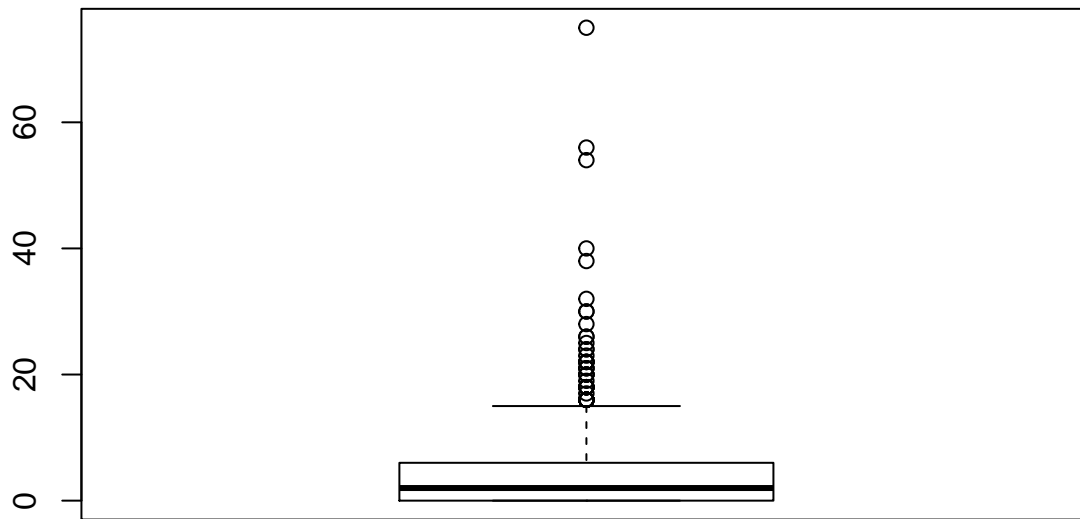
```
rm(data_clean1)
rm(data_clean2)
rm(data_clean3)
rm(data_clean4)
rm(eliminados1)
rm(eliminados2)
rm(eliminados3)
rm(eliminados4)
rm(i)
rm(index)
rm(outliers)
rm(variable)
rm(var)
save.image("myworkspace_TFM.RData")
```



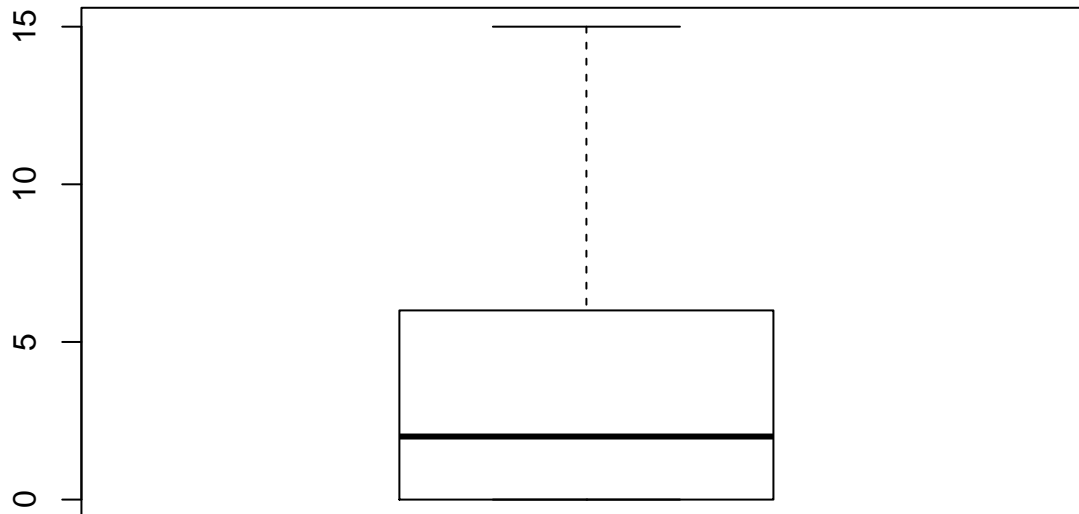
```
boxplot(data[variables_numericas])
```



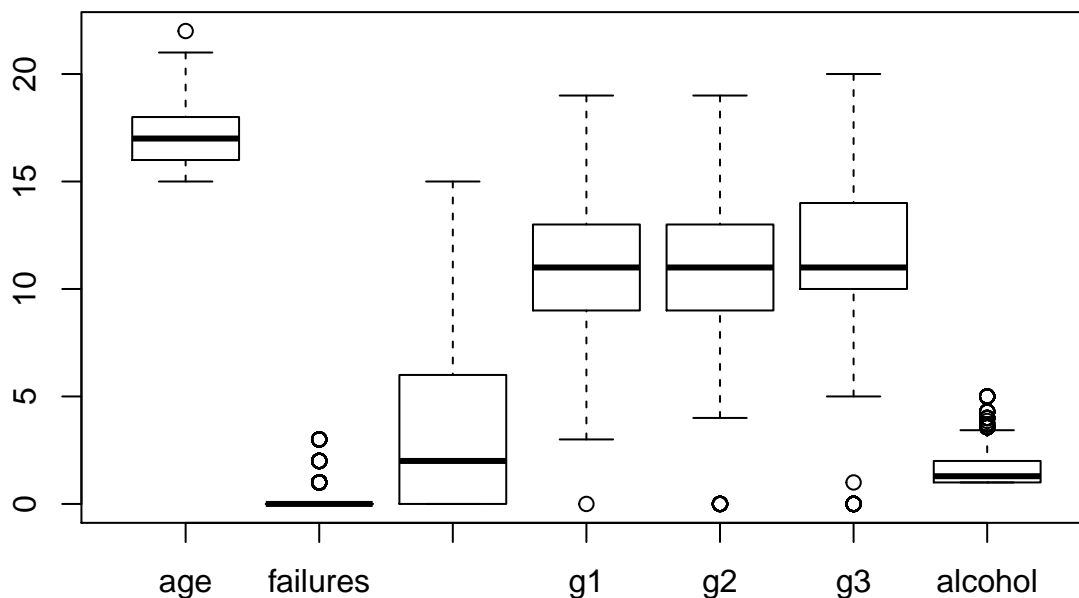
```
boxplot(data$absences)
```



```
boxplot(data_clean$absences)
```



```
boxplot(data_clean[variables_numericas])
```



Con el

boxplot se observan valores atípicos en todas las variables numéricas.

En la variable edad, 22 años se considera outlier pero es correcta. En las variables g1, g2 y g3 considera outliers el valor 0 pero es un valor correcto.

En la variable failures considera outliers los valores 1, 2 y 3 pero también son valores correctos.

Solo se eliminan los outliers en la variable ausencias ya que presentan valores muy extremos y pueden afectar por ejemplo en el análisis cluster. Los datos que quedan después de eliminar los valores extremos se guardan en el data.frame data\_clean.

## Interpretación de los datos

```
library("ggplot2")
library(RColorBrewer)
```

```

factor_var_plot <- function(var, var_name, color){
  df_plot<-as.data.frame(table(data[,var]))
  names(df_plot)<-c(var, "count")

  if( nrow(df_plot)>6 ){ lab<-paste0(df_plot$count) }
  else{ lab<-paste0(df_plot$count, " (", round(df_plot$count/1044*100,1), "%)" ) }

  switch(color,
    rainbow={ cols=rainbow(nrow(df_plot), alpha=0.8) },
    heat={
      cols=heat.colors(nrow(df_plot))
    },
    heat_inverso={
      cols=heat.colors(nrow(df_plot)+2)
      cols=cols[length(cols):3]
      #c[length(c):1]
    },
    doscolores={ cols=c("#00FF66CC","#0066FFCC") #cols=c("plum", "dodgerblue") #cols=rainbow(nrow
    },
    noyes={ cols=c("#FF0000CC","#0000FFCC") },
    semaforo={ cols=brewer.pal(nrow(df_plot),"RdYlGn") },
    semaforo_inverso={
      cols=brewer.pal(nrow(df_plot),"RdYlGn")
      cols=cols[length(cols):1]
    },
    { cols=palette() }
  )

  plot<-ggplot(data=df_plot, aes(x=df_plot[,1], y=df_plot[,2], fill=df_plot[,1]) ) +
    geom_bar(stat="identity")+#scale_fill_brewer(palette="RdYlGn")+

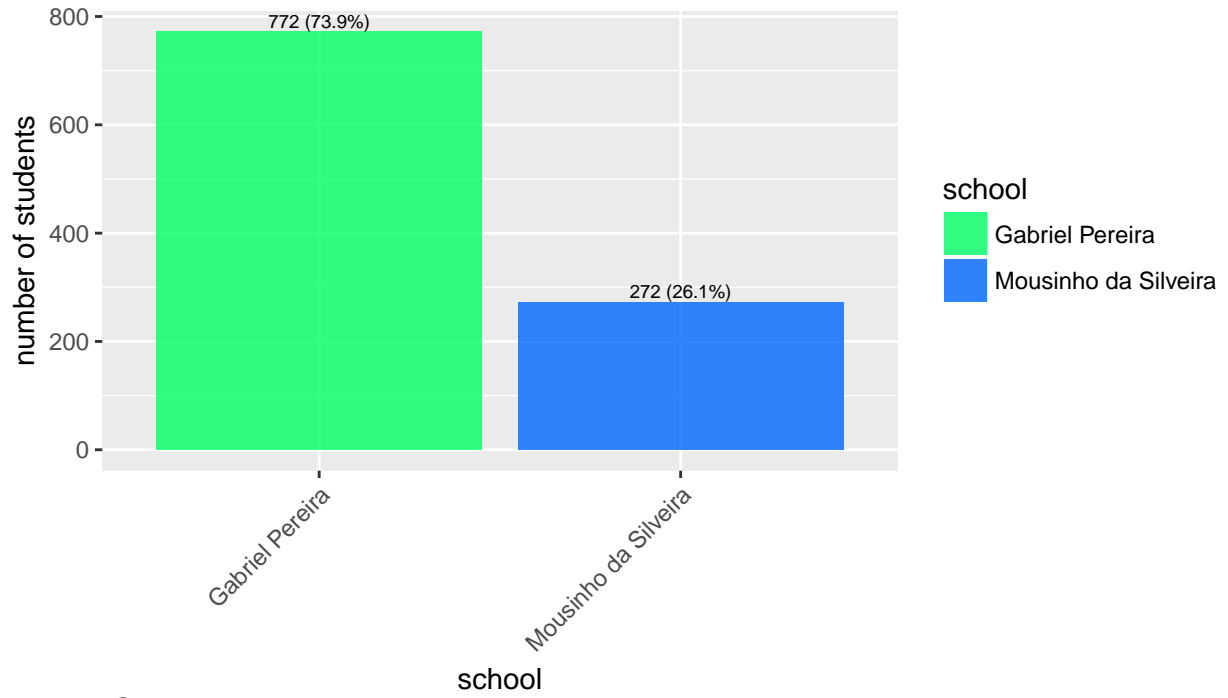
    scale_fill_manual(values=cols)+

    geom_text(aes(label=lab), vjust=-0.3, size=2.5)+
    ggtitle(paste("\n", "\n", var_name, "\n" ) +
    xlab(var) + ylab("number of students") + labs(fill=var)+
    theme(axis.text.x=element_text(angle=45, hjust=1))
  show(plot)
}

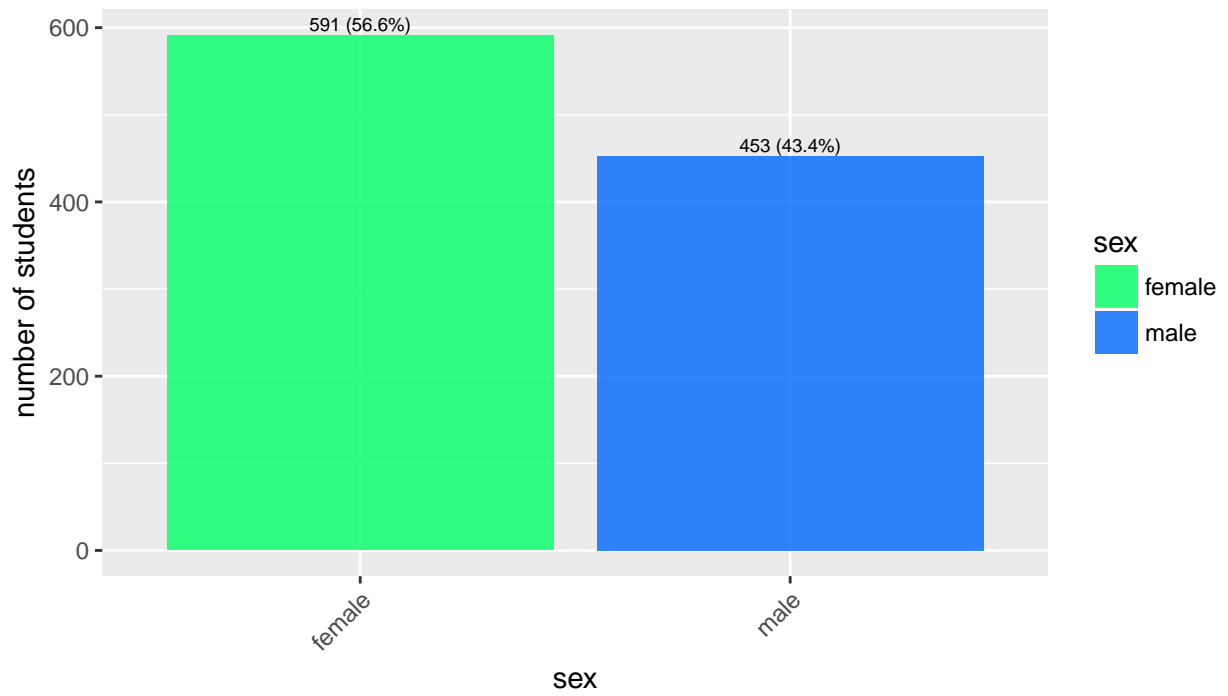
for(i in 1:length(variables_discretas)){
  var<-paste(variables_discretas[i])
  var_name<-paste(variables_discretas_names[i])
  color<-variables_discretas_color[i]
  factor_var_plot(var,var_name,color)
}

```

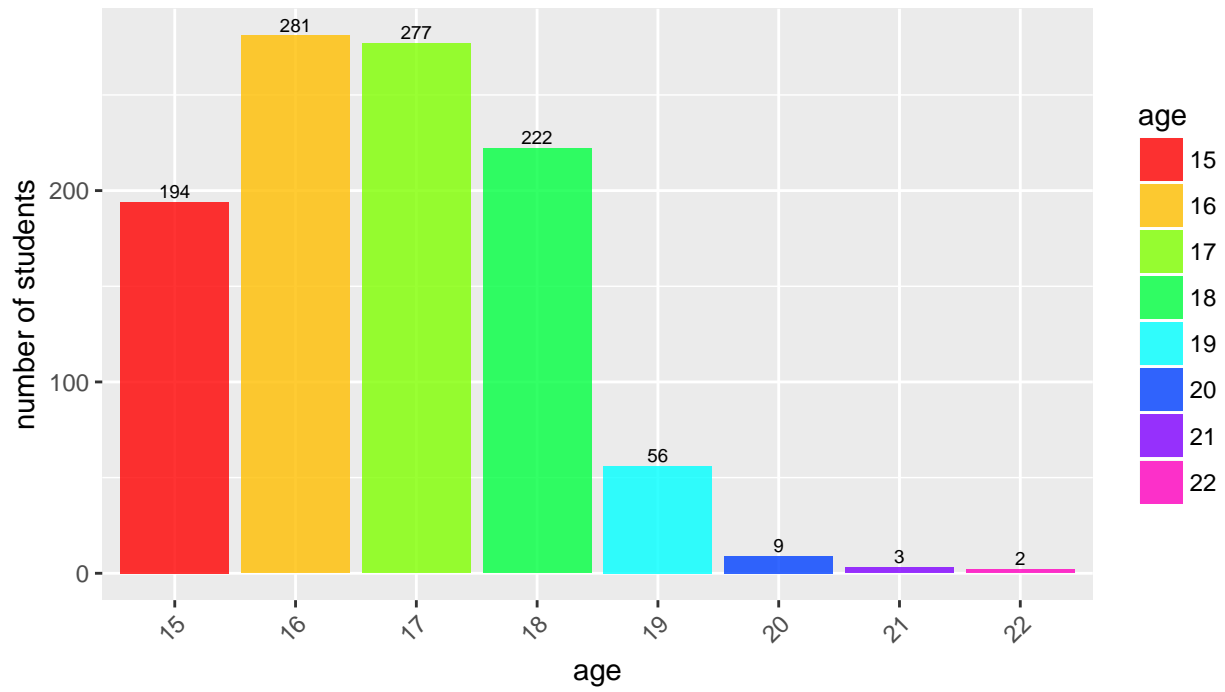
### Student's school



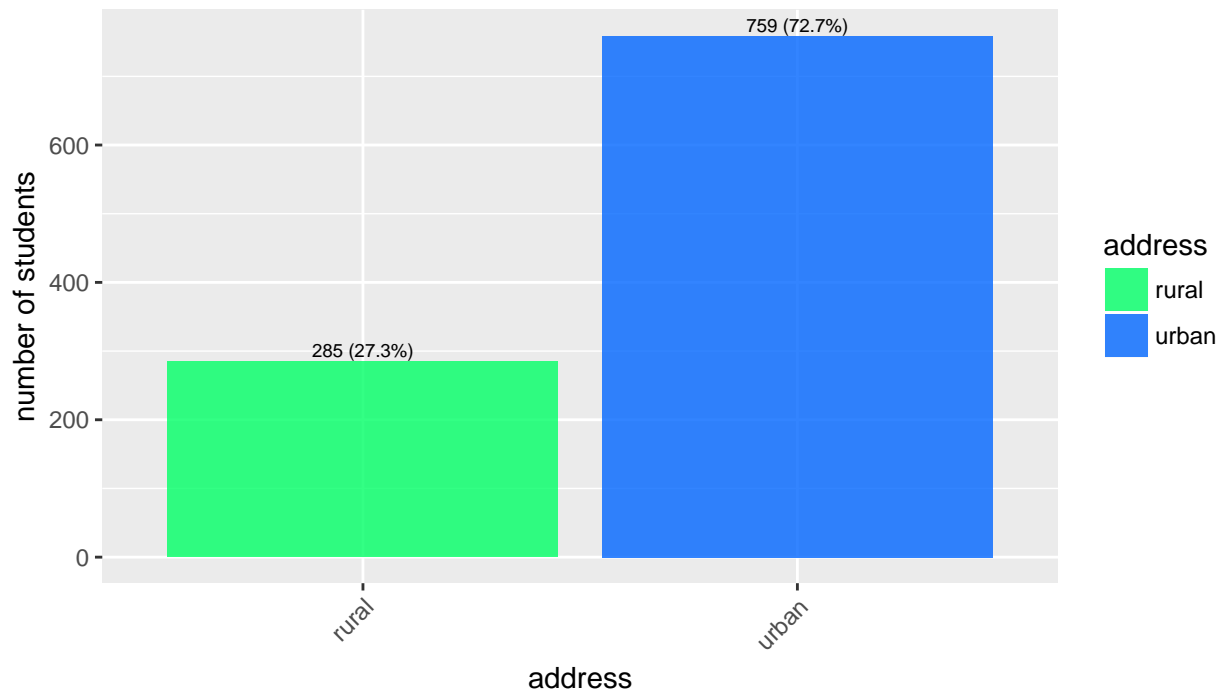
### Student's sex



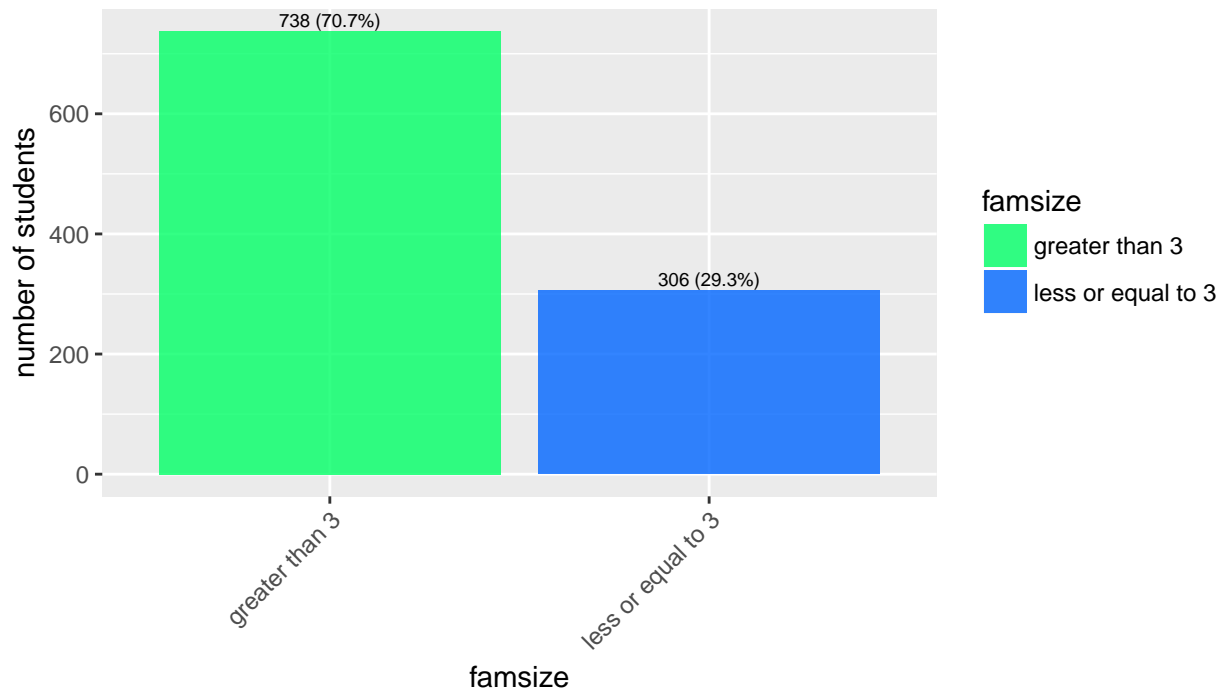
### Student's age



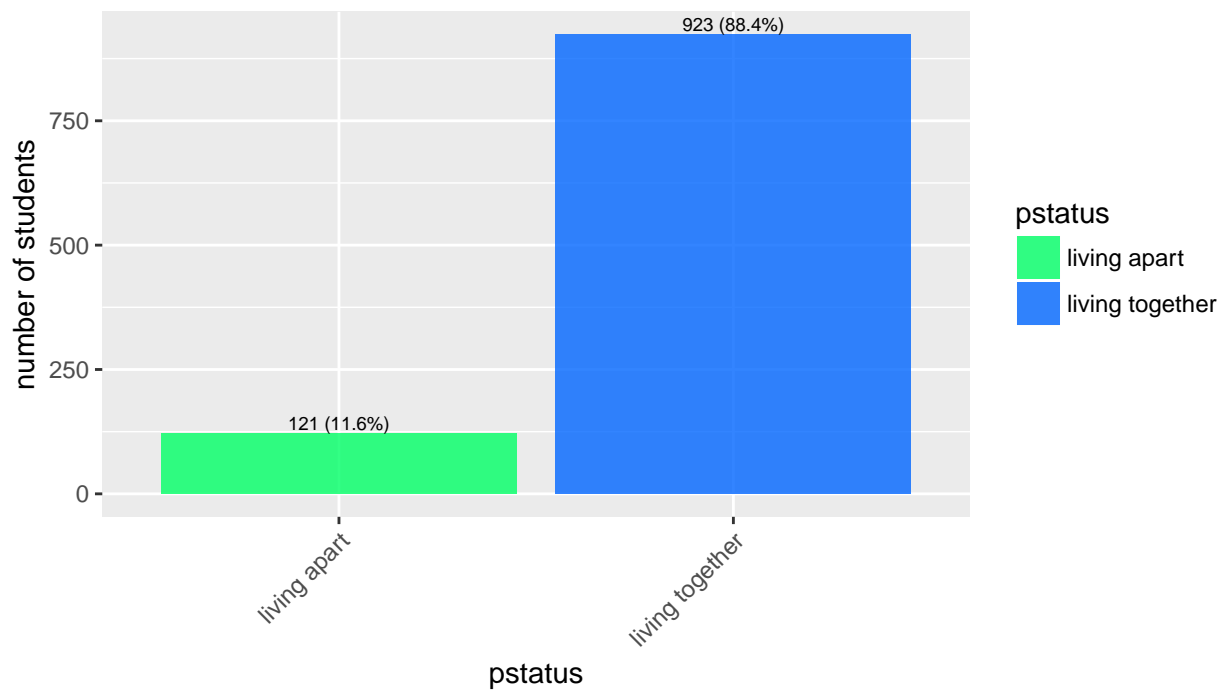
### Student's home address type



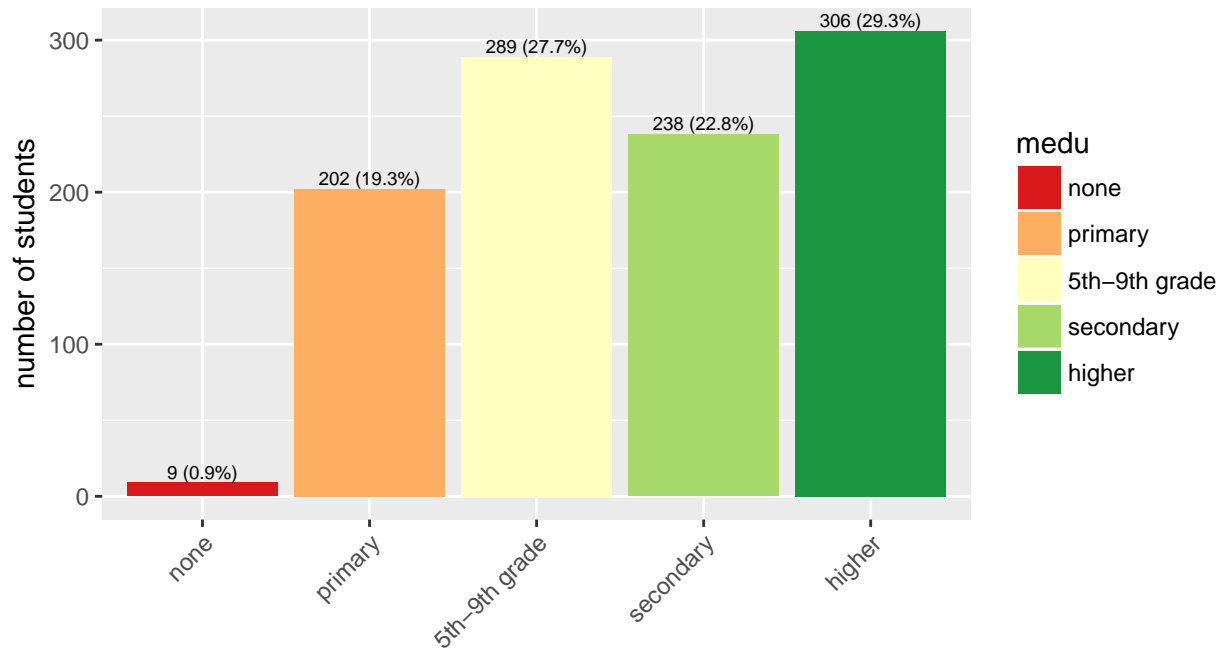
### Family size



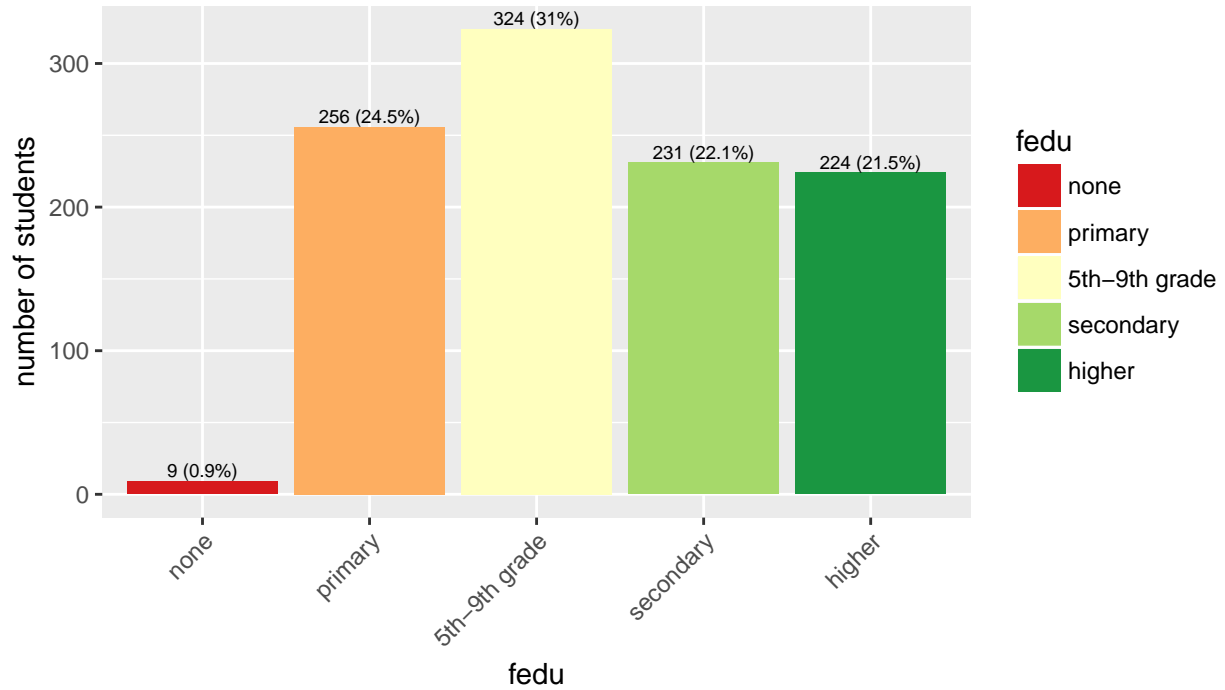
### Parent's cohabitation status



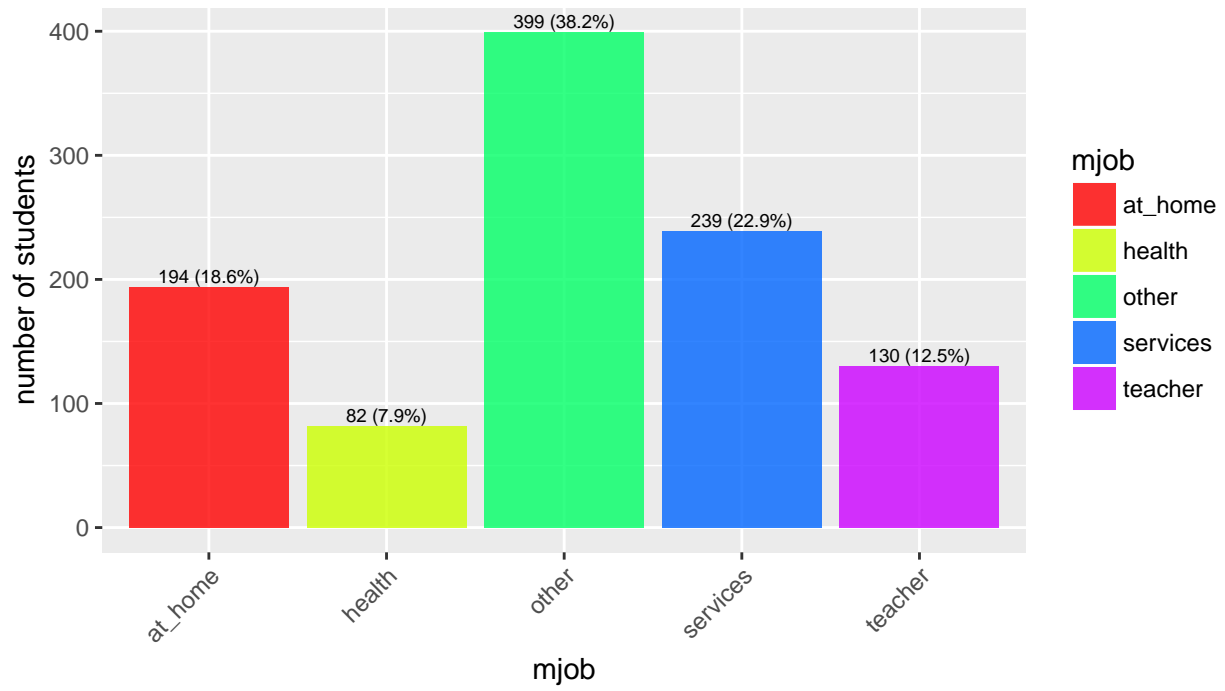
### Mother's education



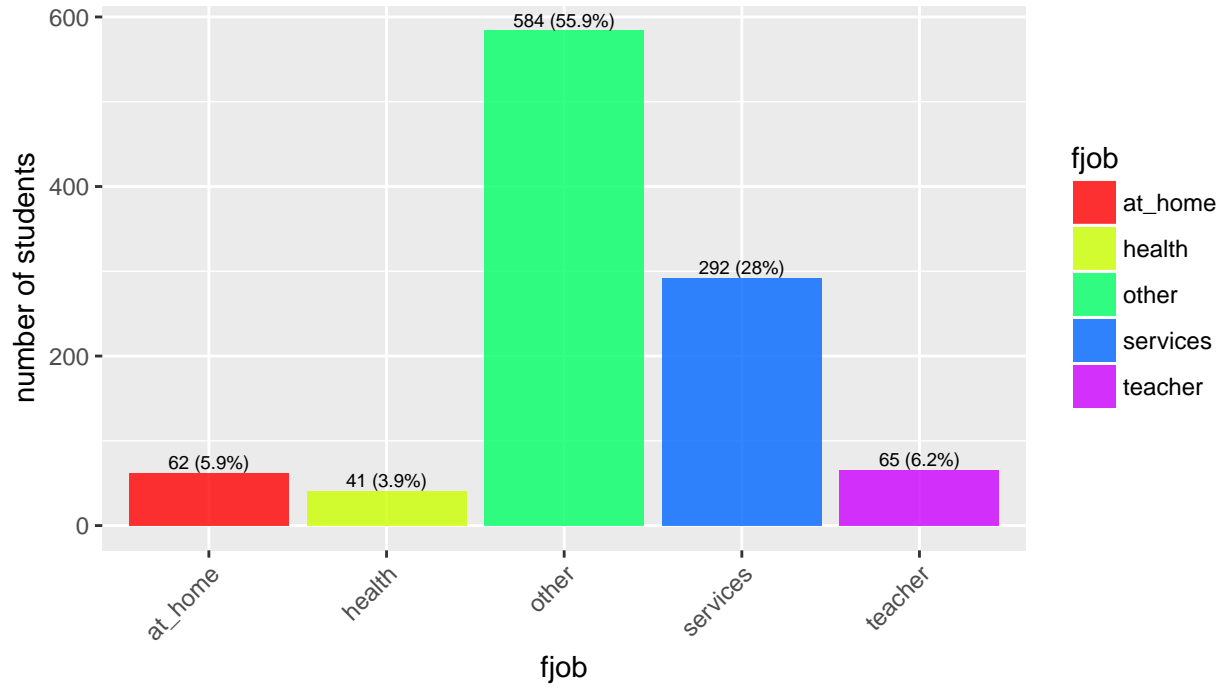
### Father's education



### Mother's job

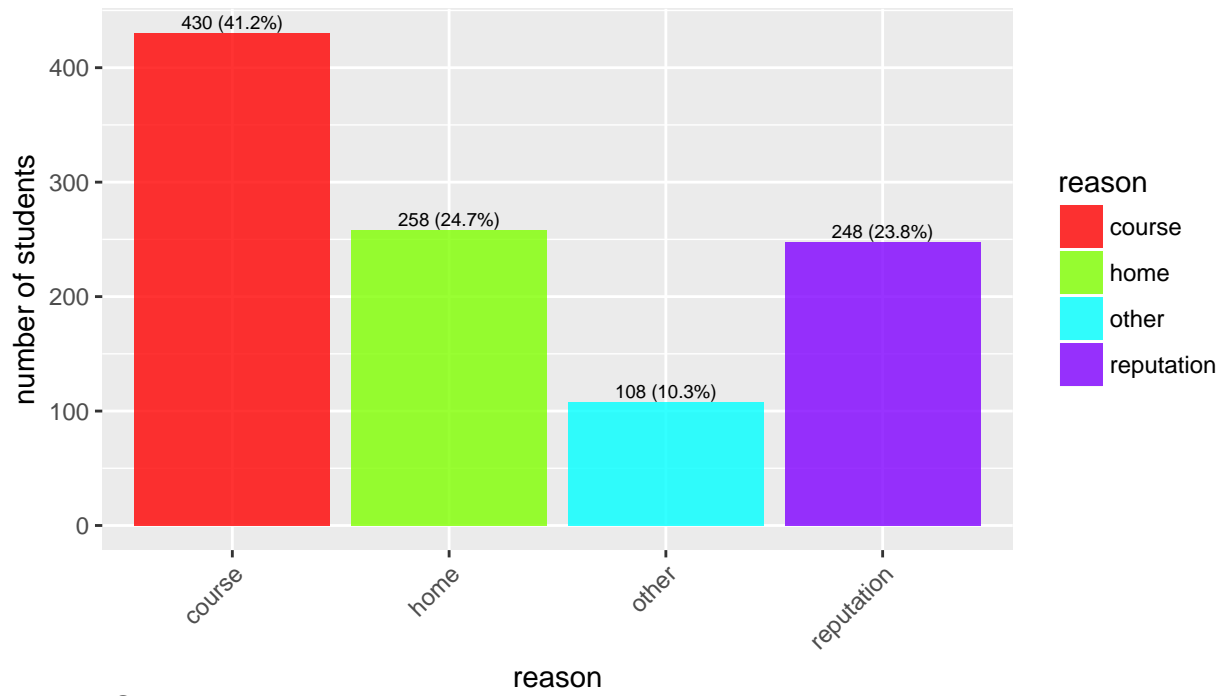


### Father's job

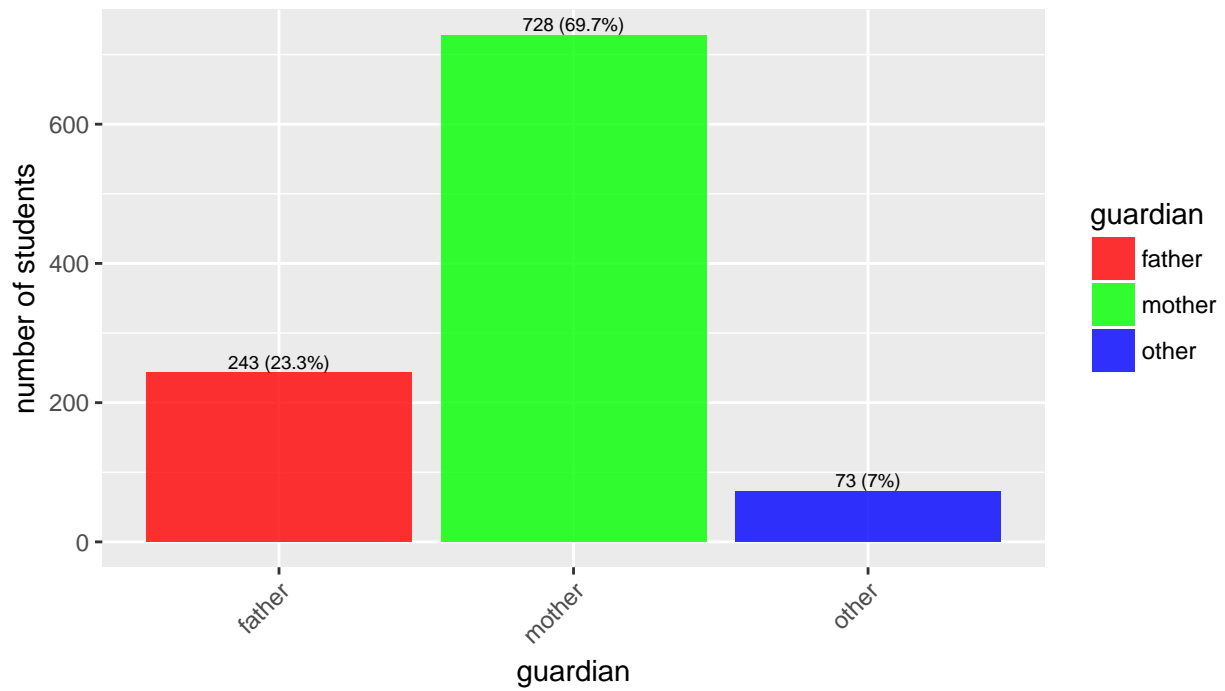




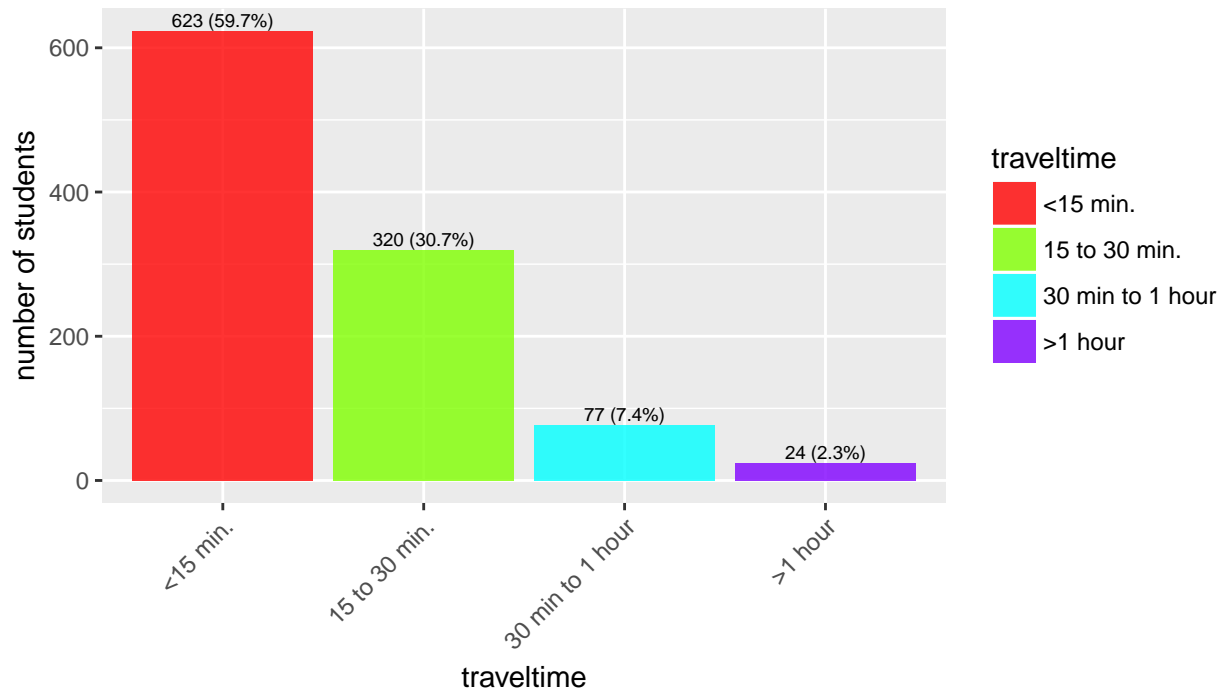
### Reason to choose this school



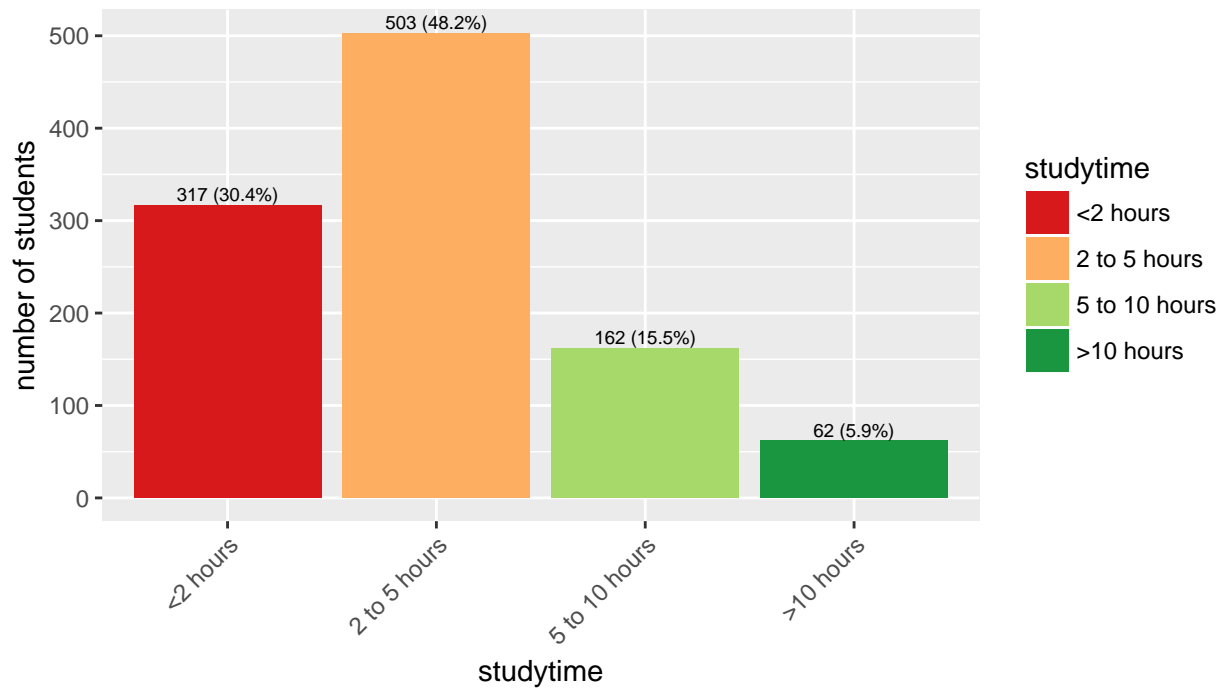
### Student's guardian



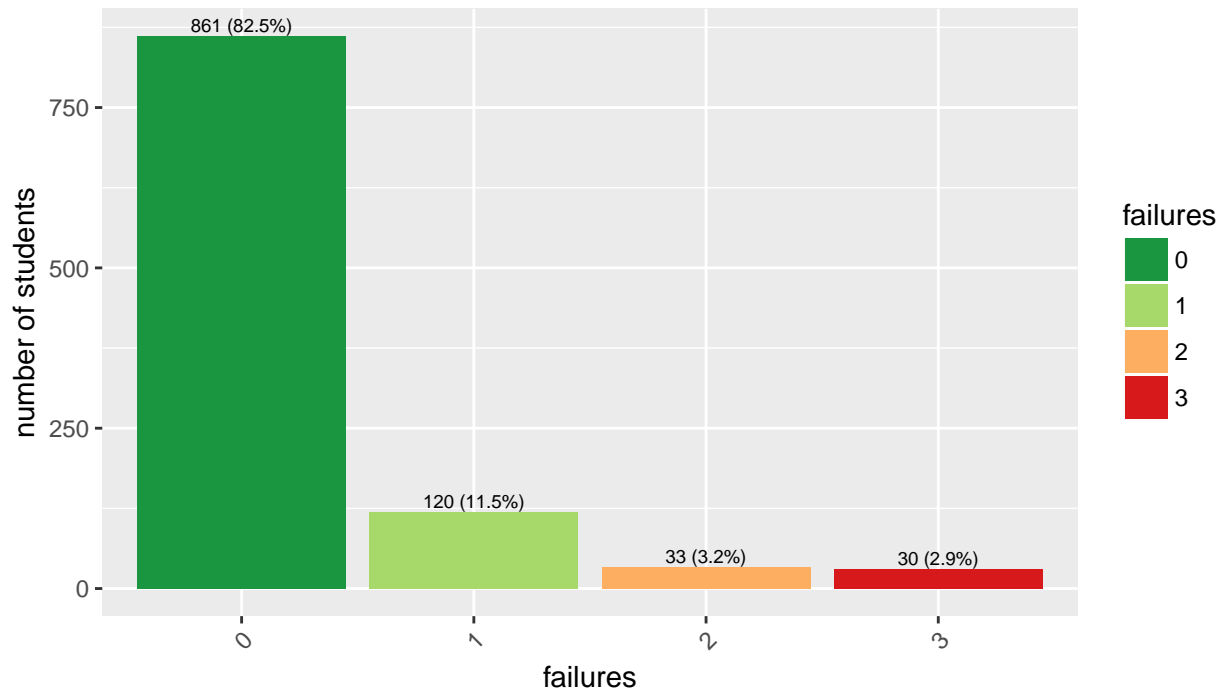
### Home to school travel time



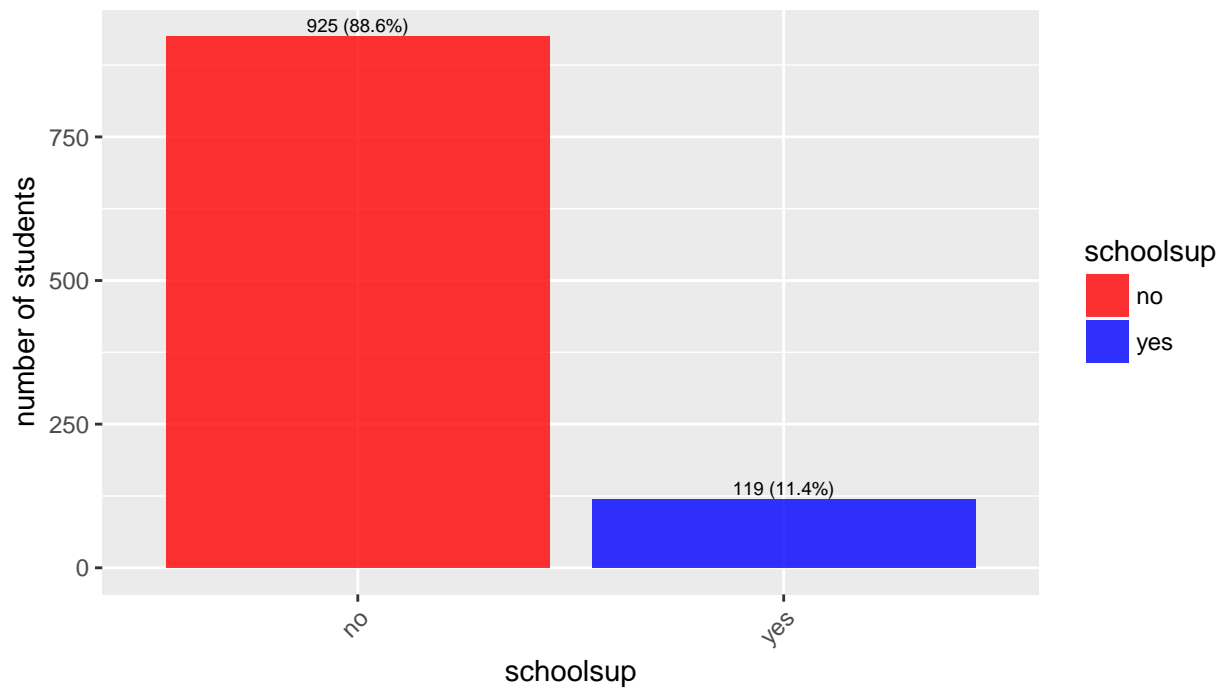
### Weekly study time



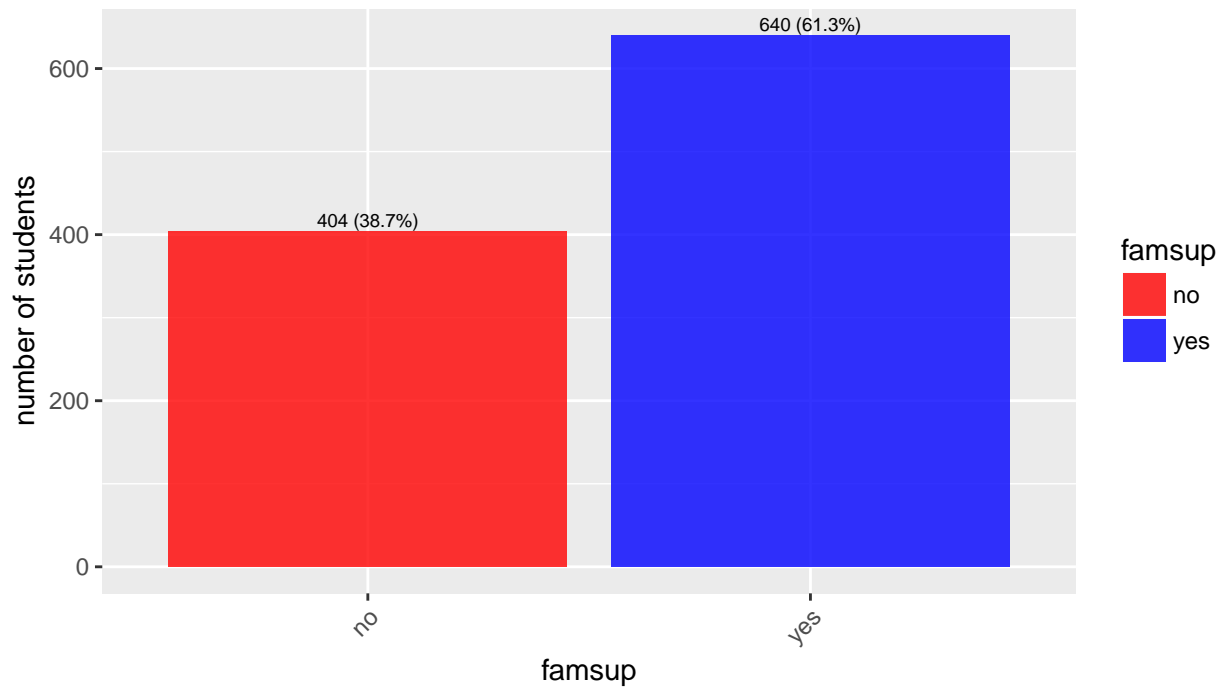
Number of past class failures



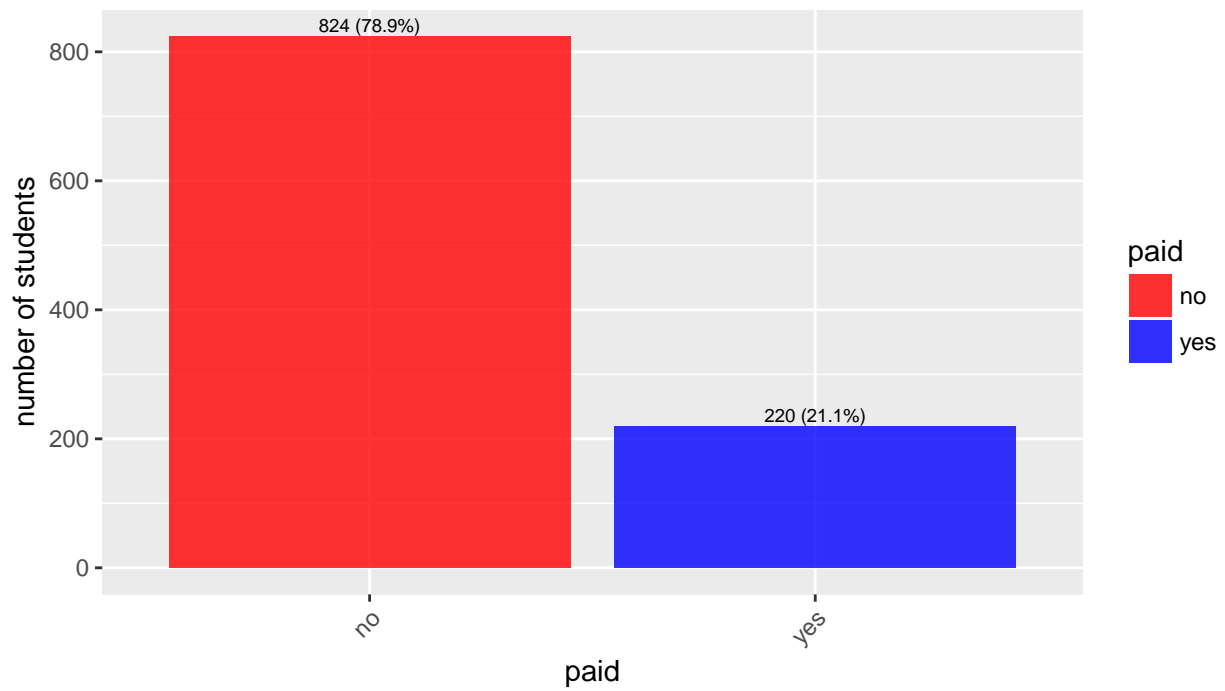
Extra educational support



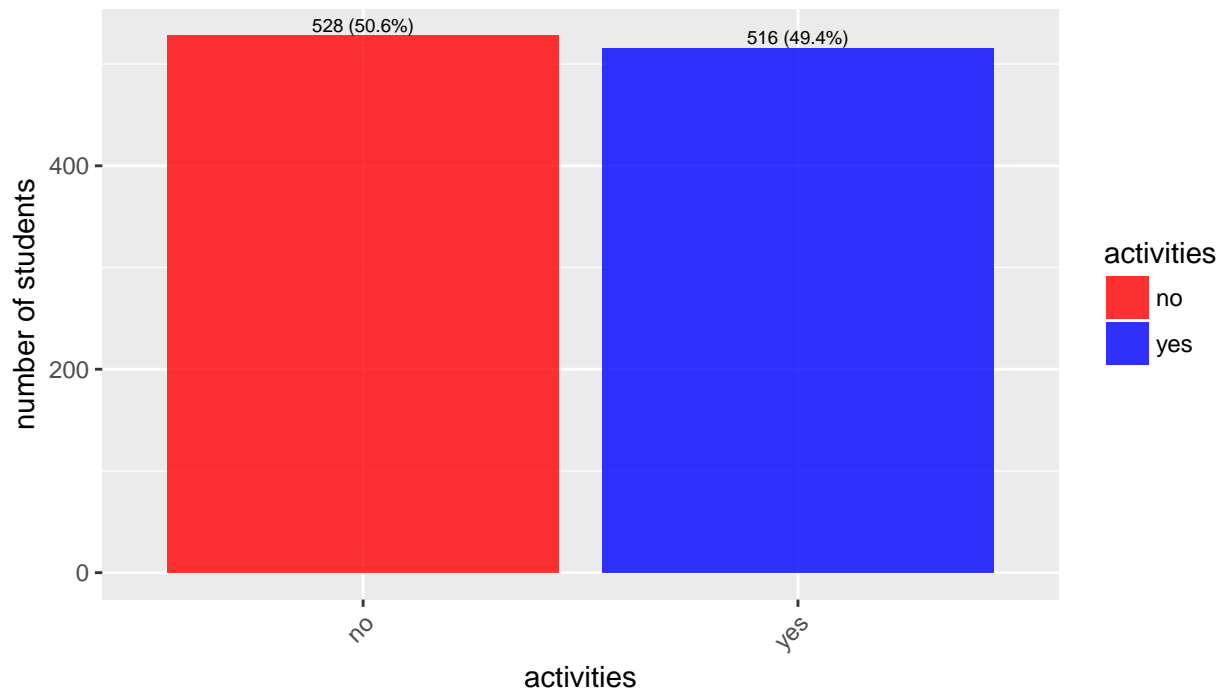
### Family educational support



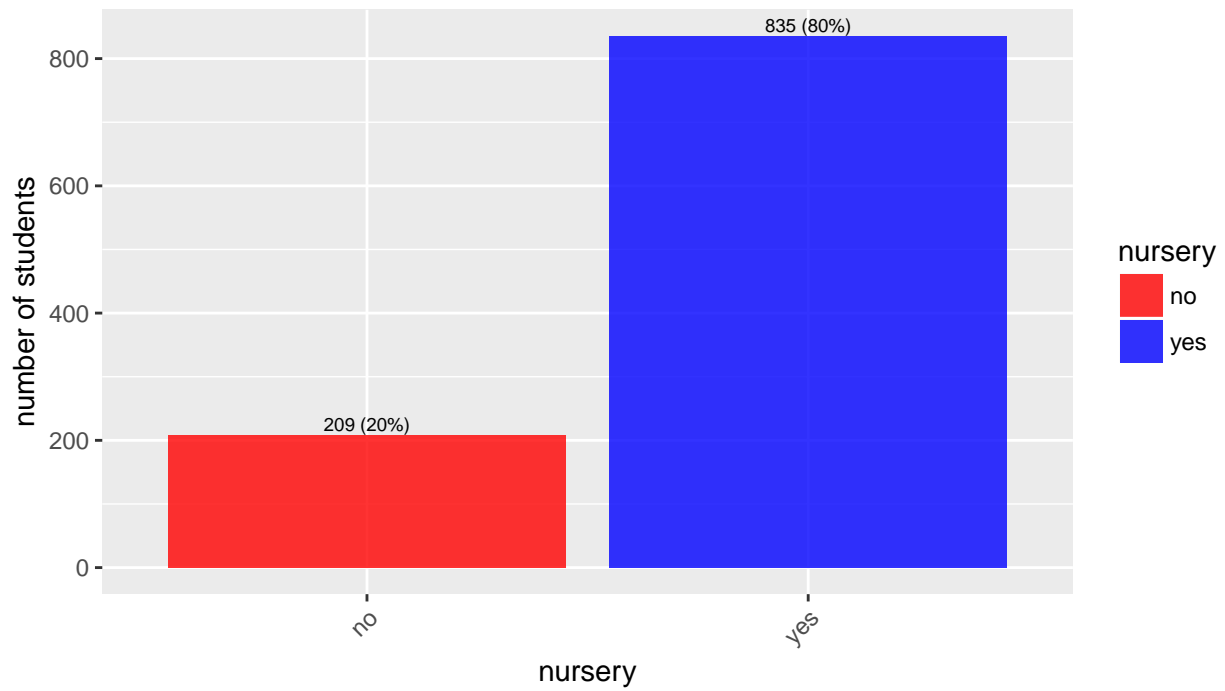
### Extra paid classes within the course subject



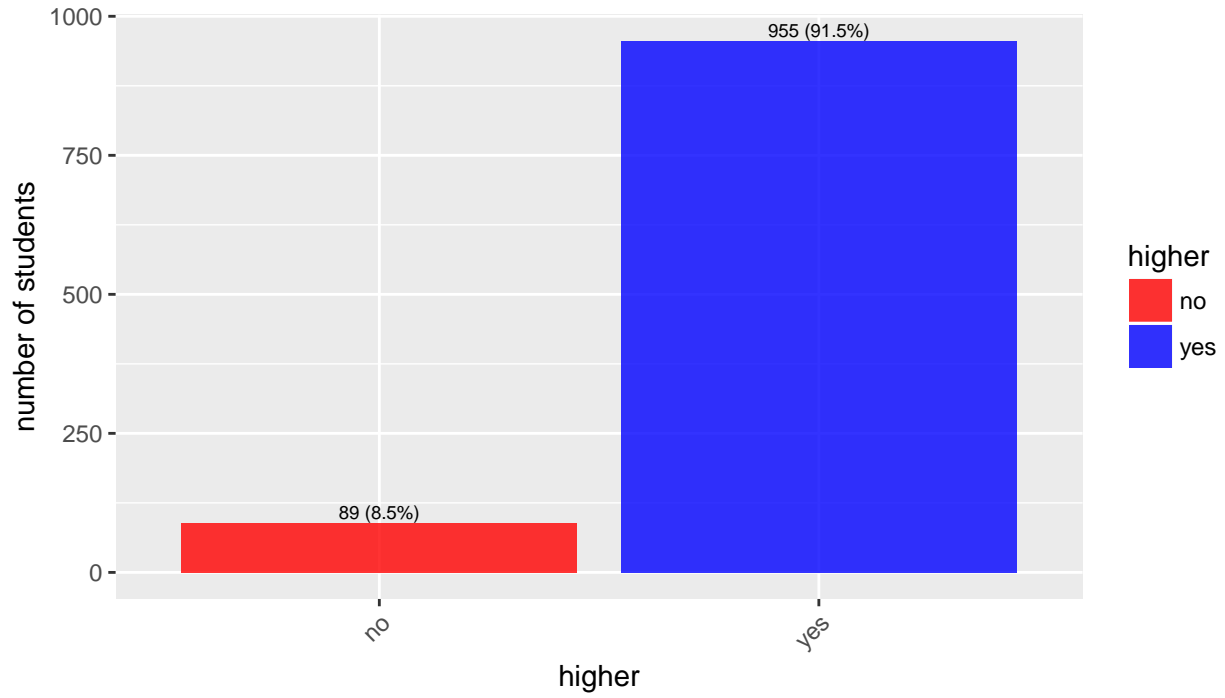
### Extra-curricular activities



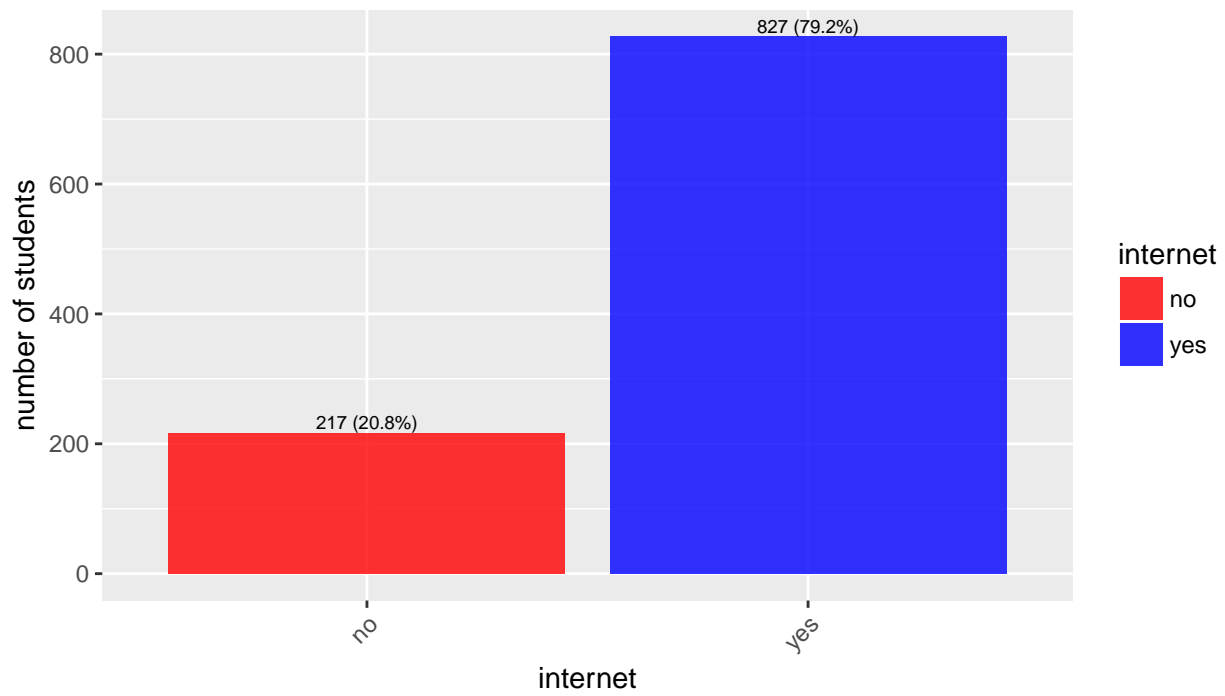
### Attended nursery school



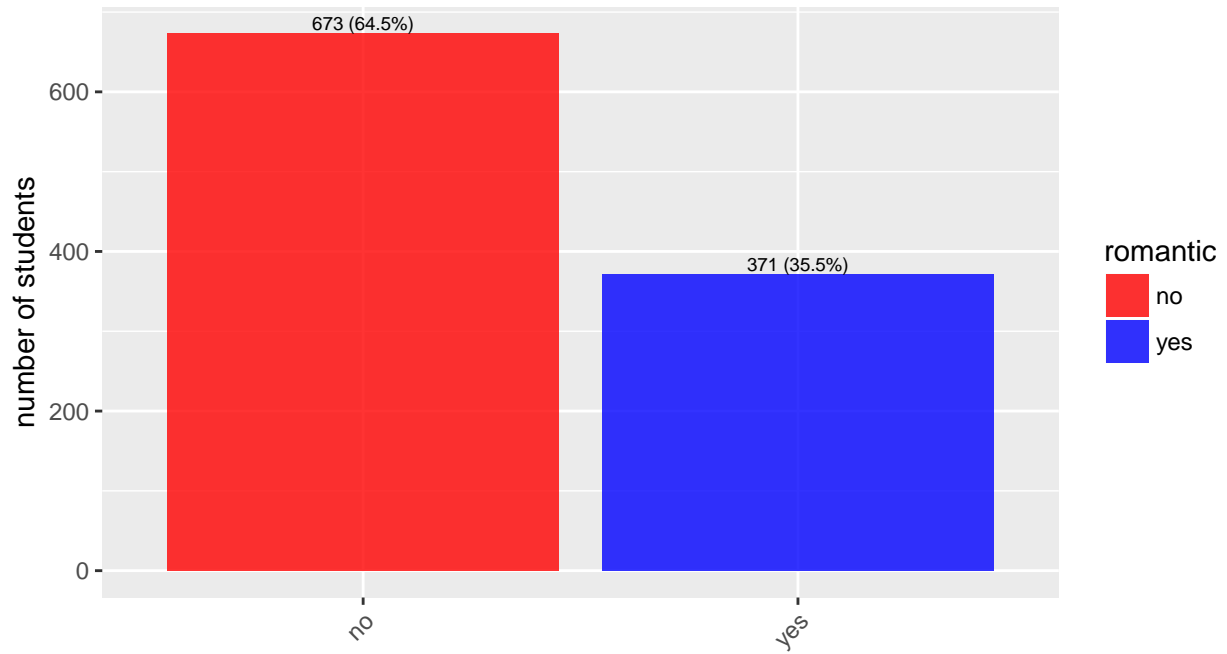
### Wants to take higher education



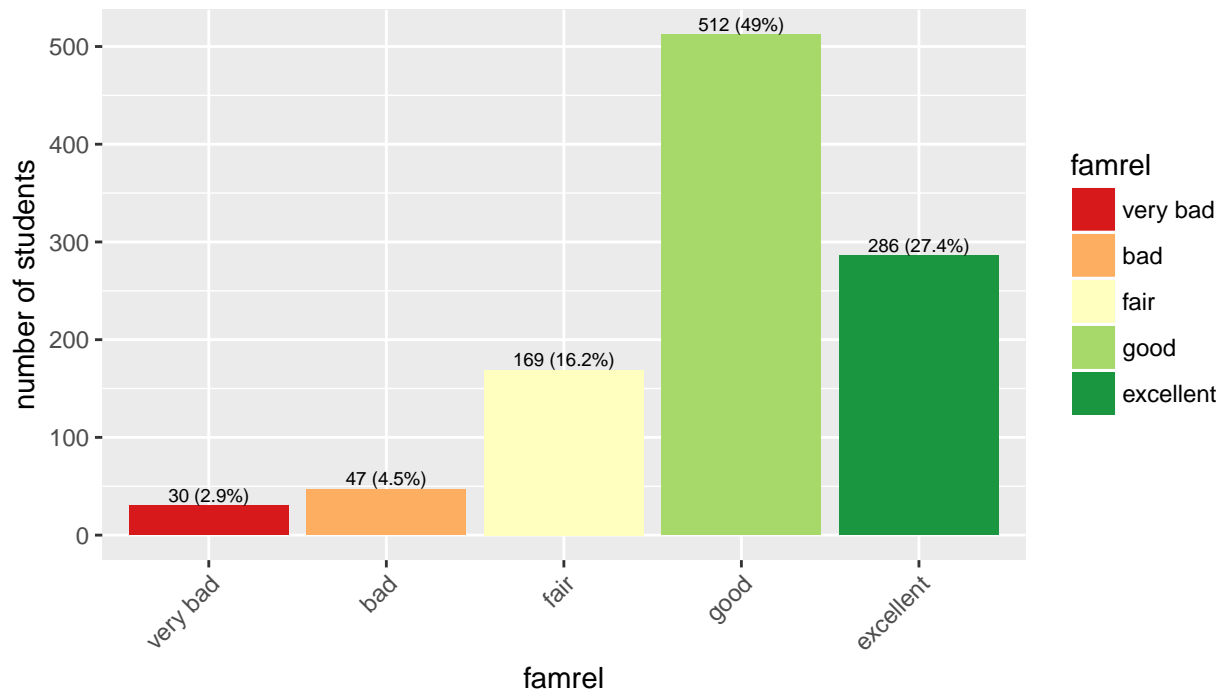
### Internet access at home



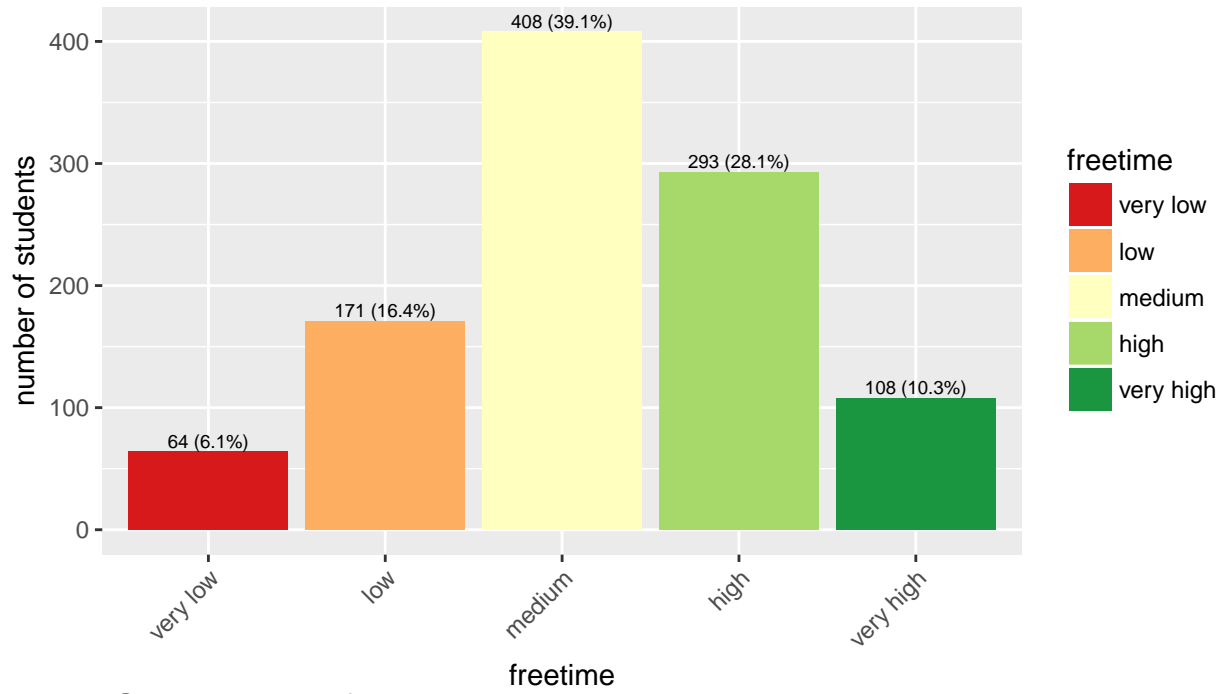
### With a romantic relationship



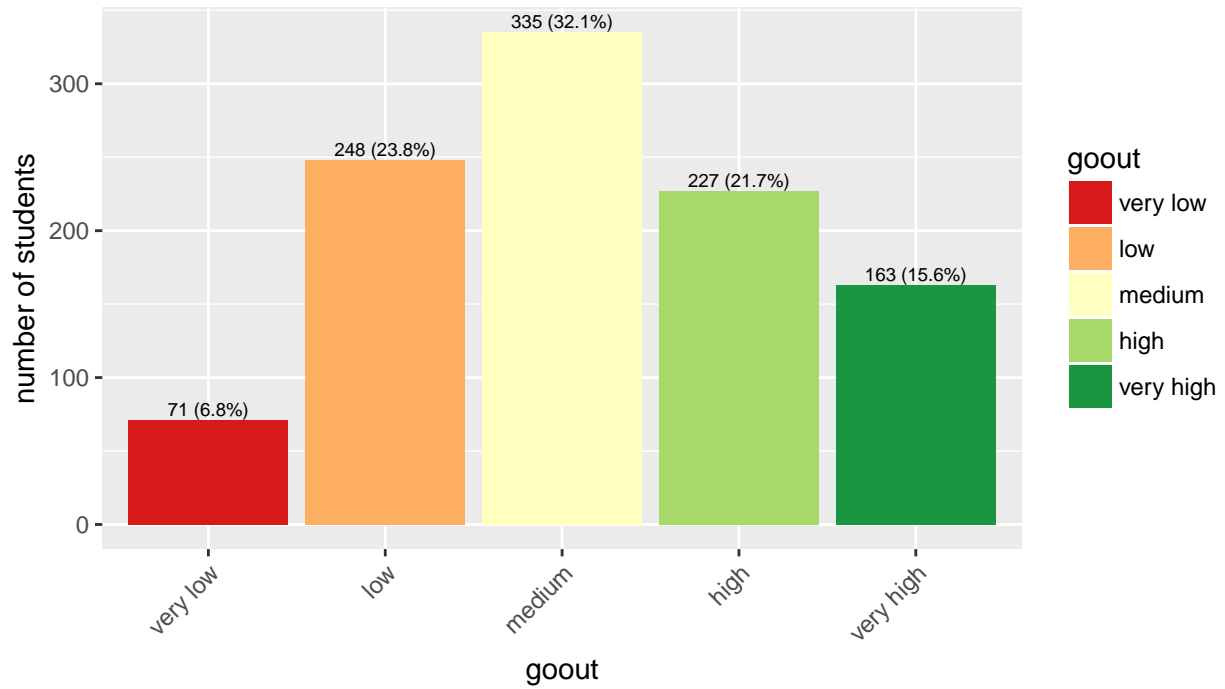
### Quality of family relationships



### Free time after school

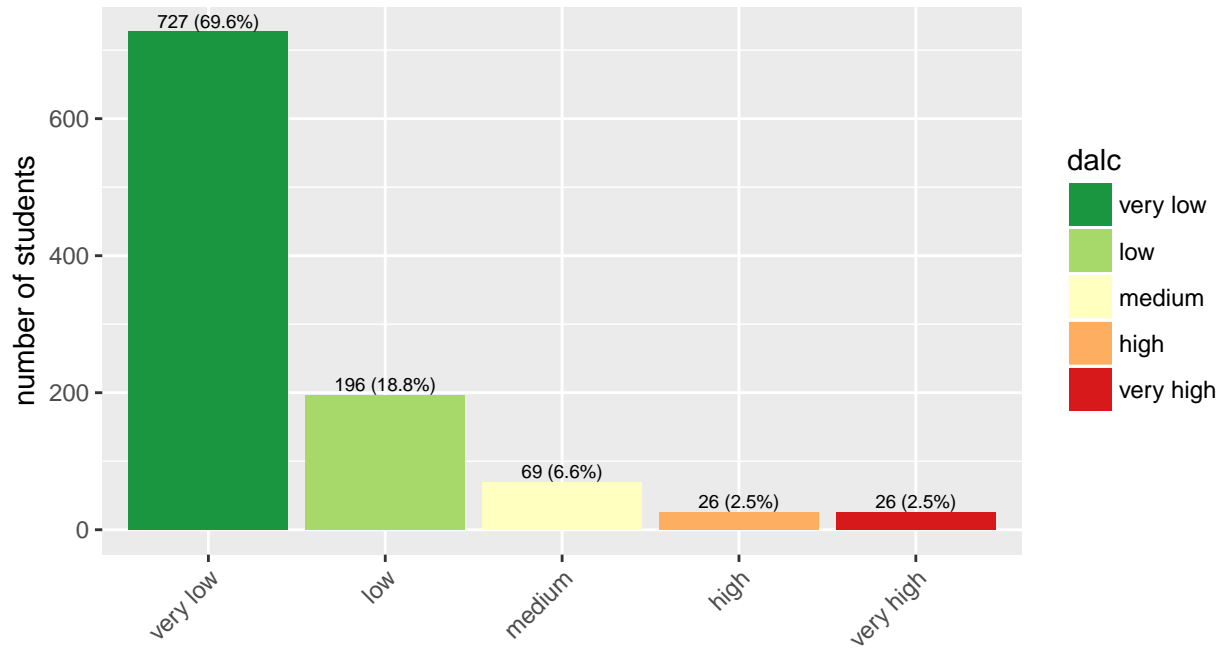


### Going out with friends

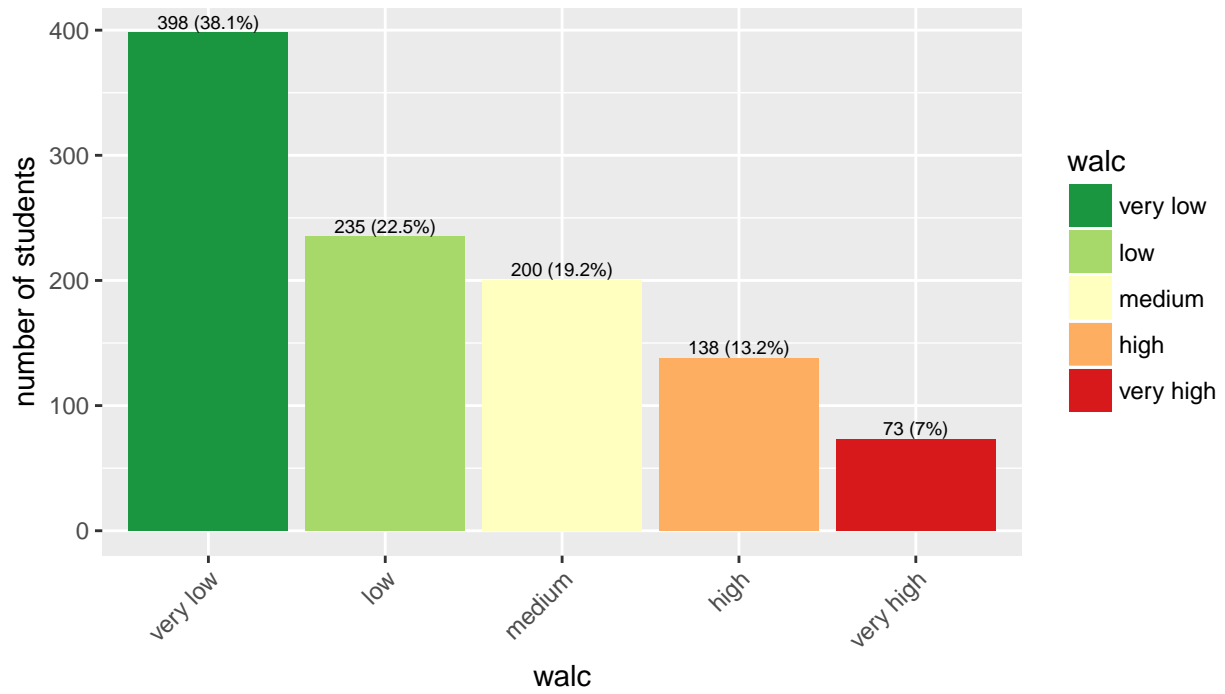




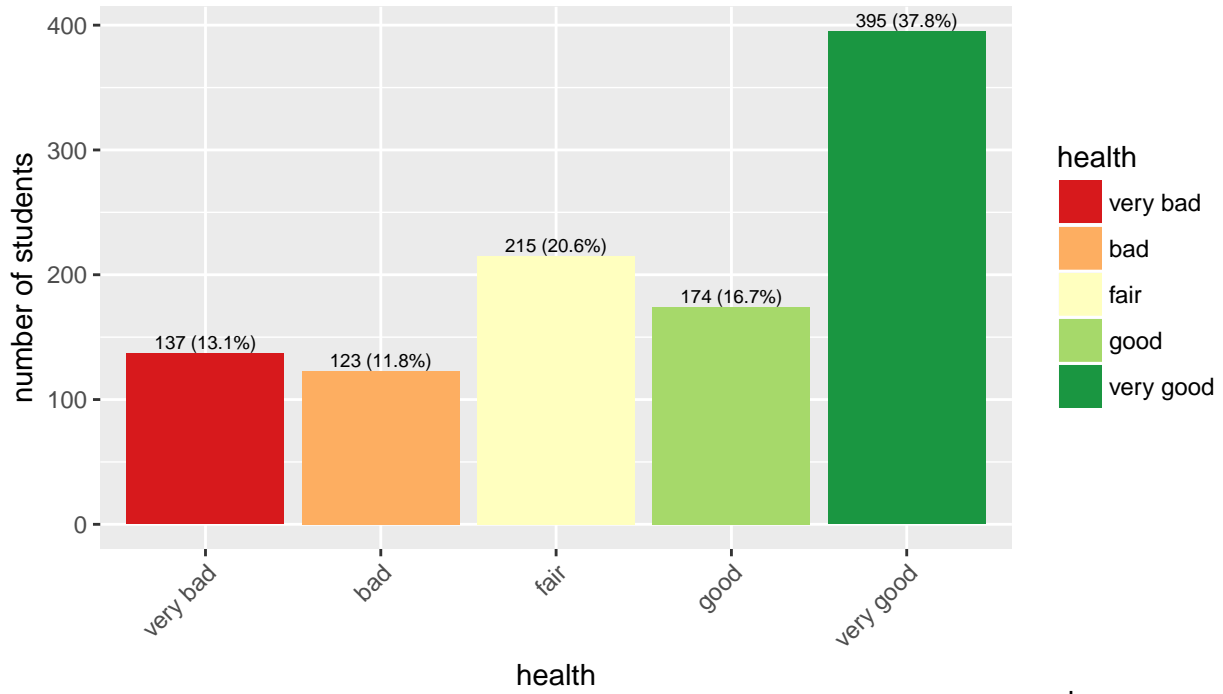
### Workday alcohol consumption



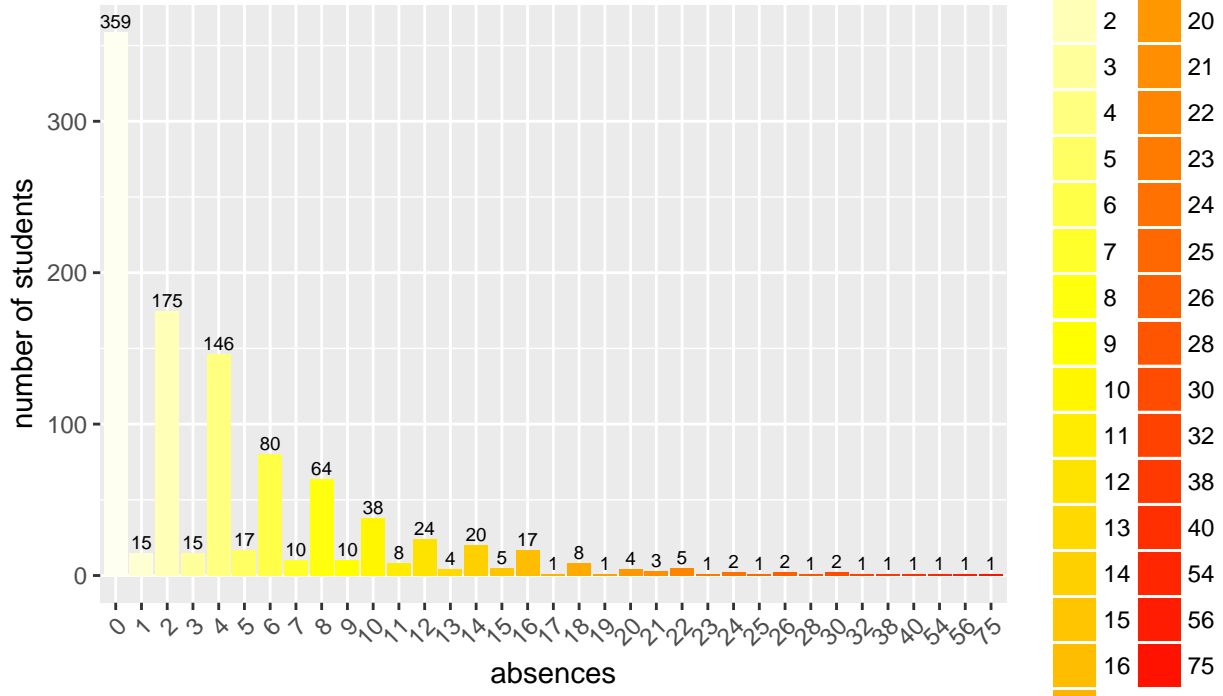
### Weekend alcohol consumption

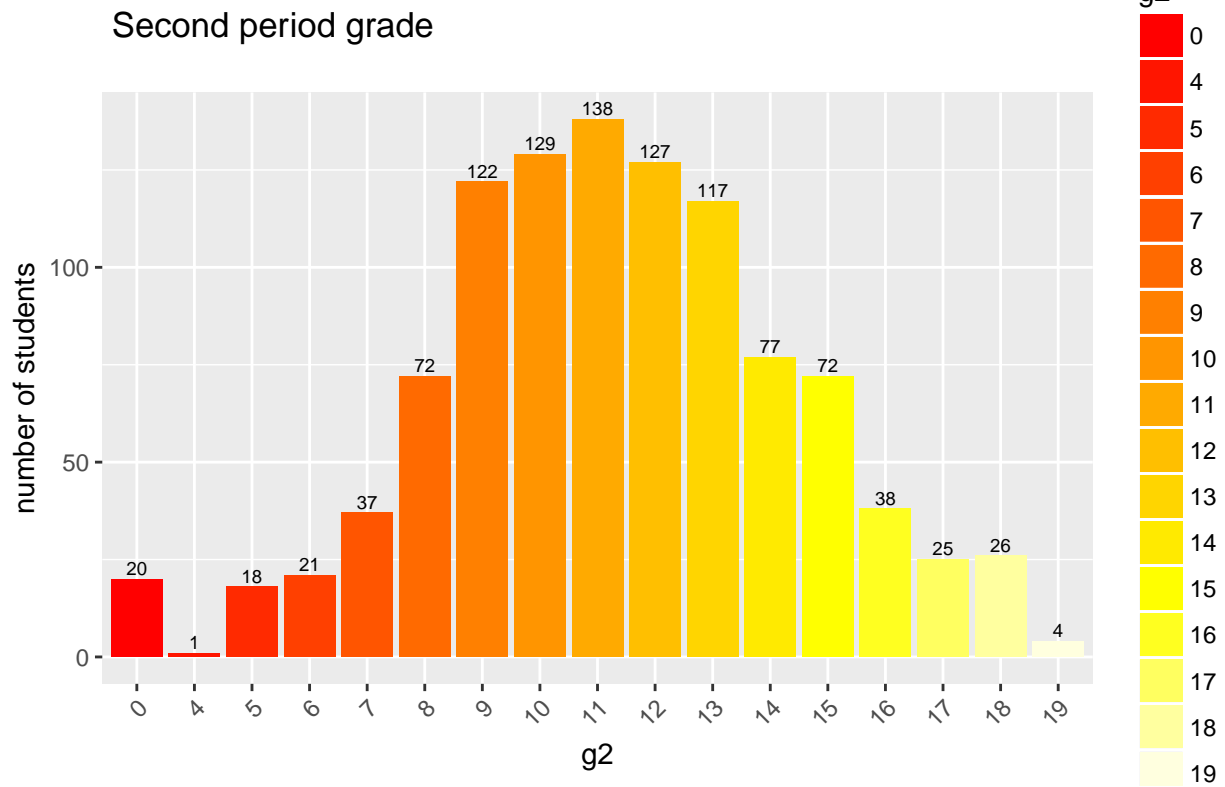
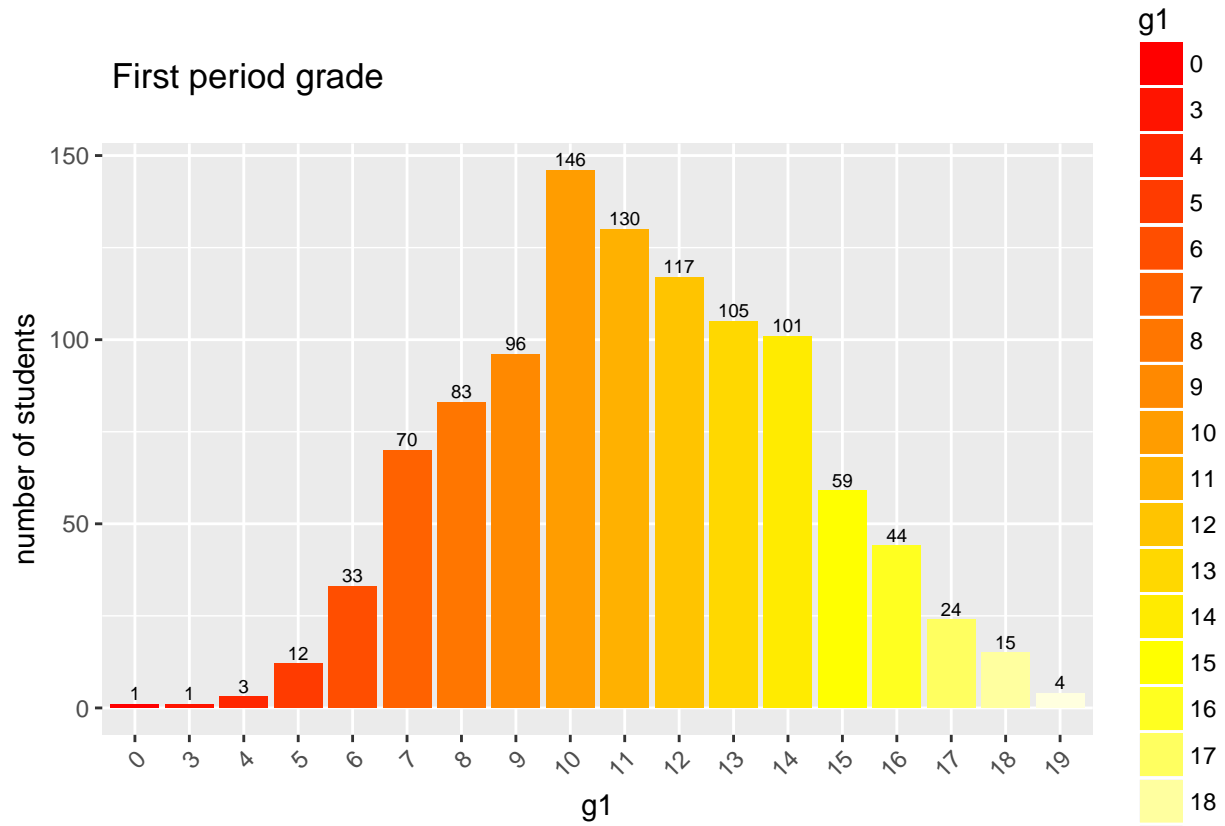


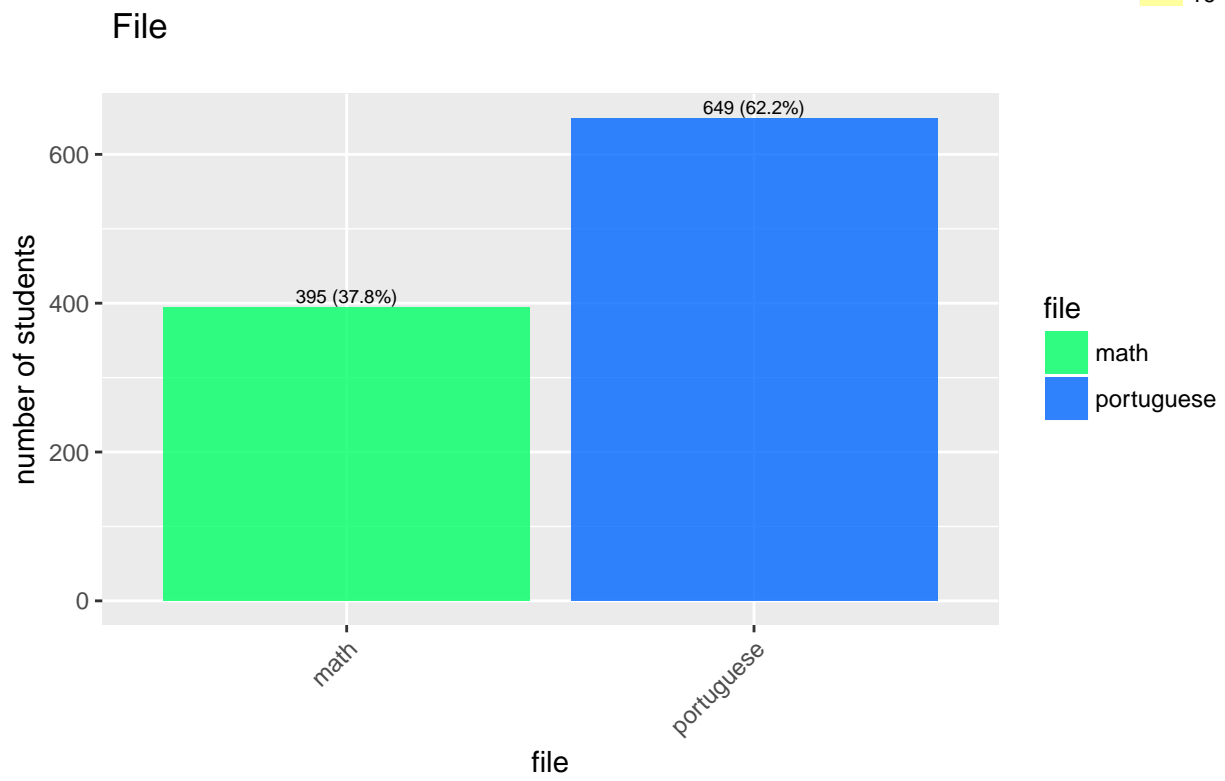
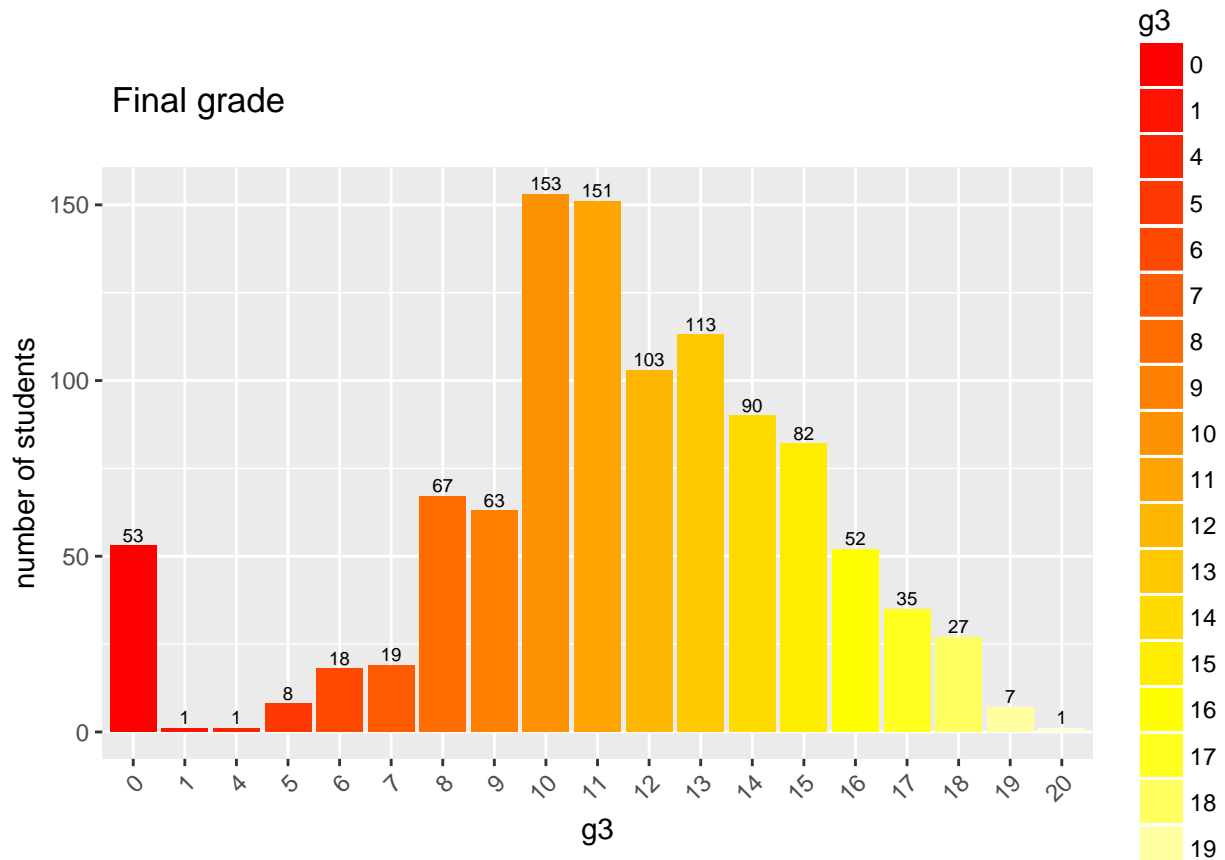
### Current health status



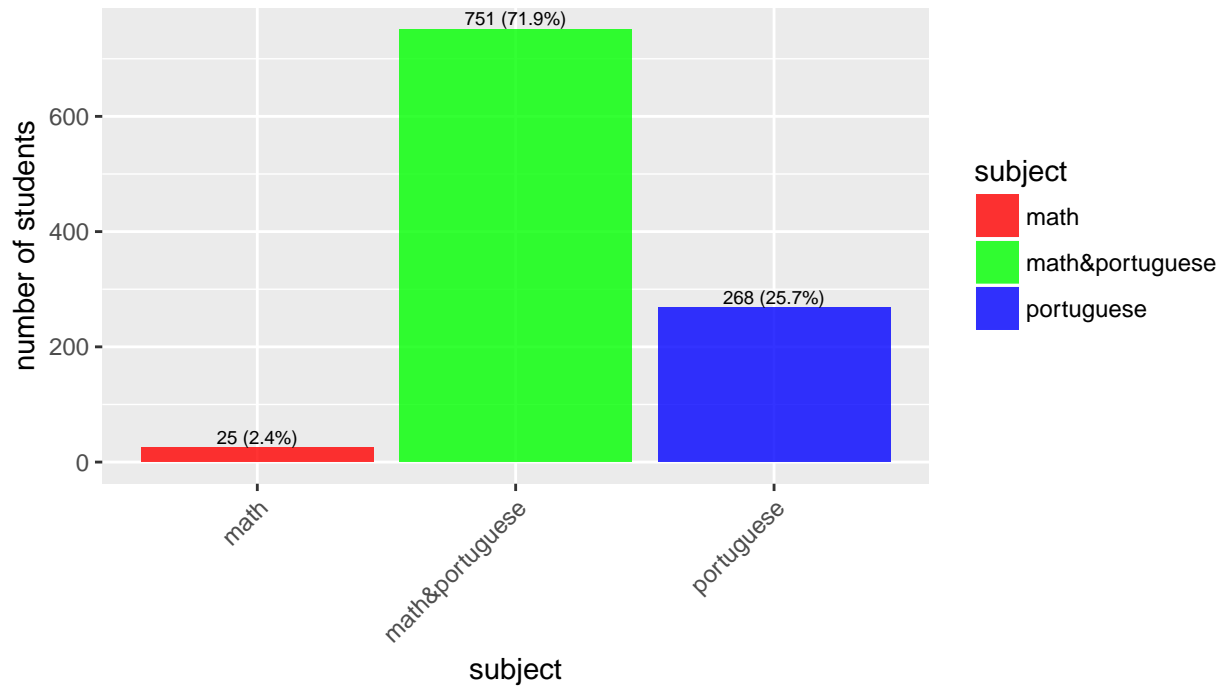
### Number of school absences



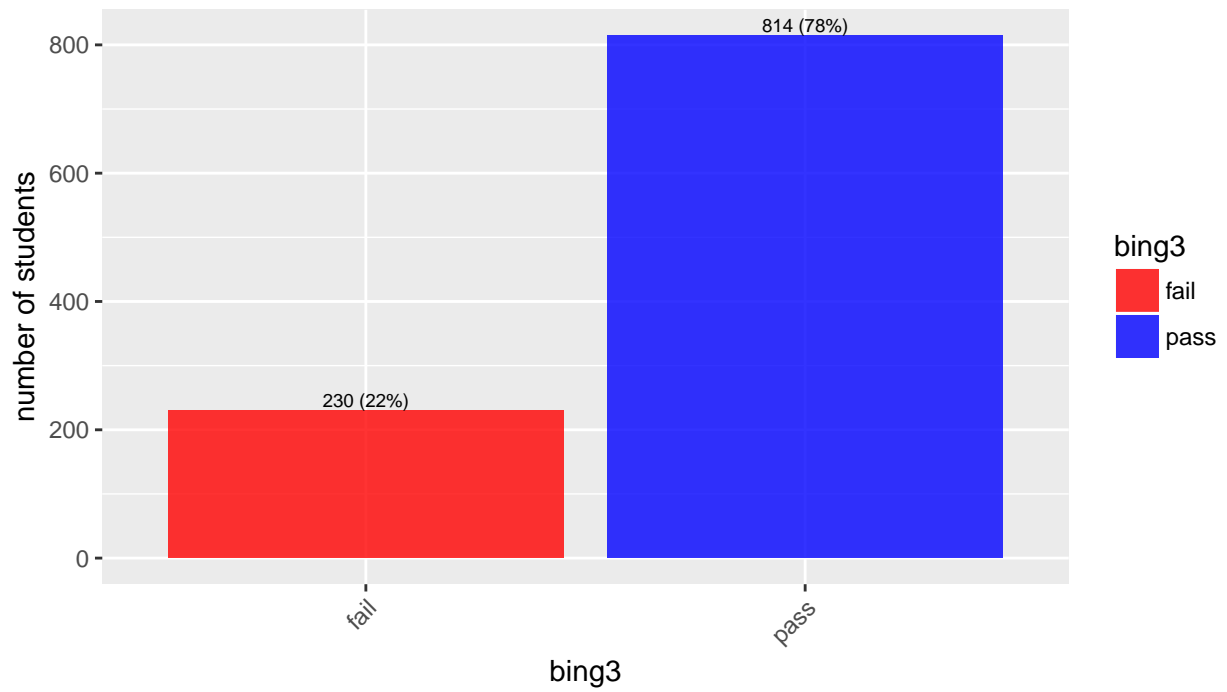




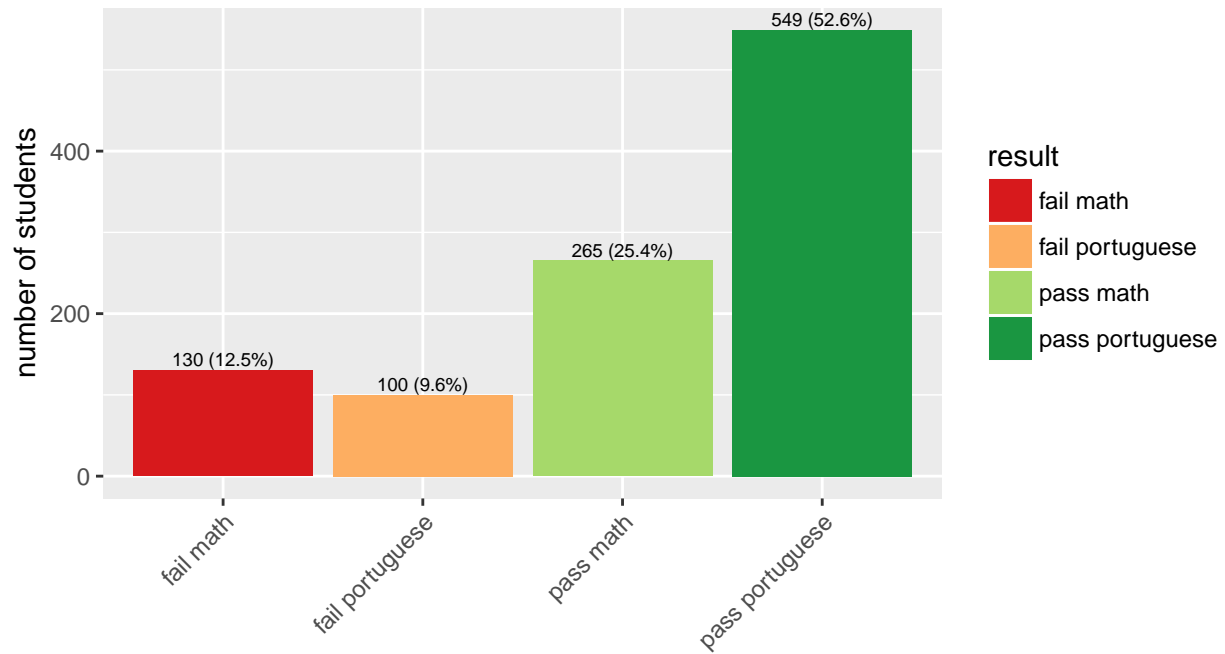
Subject



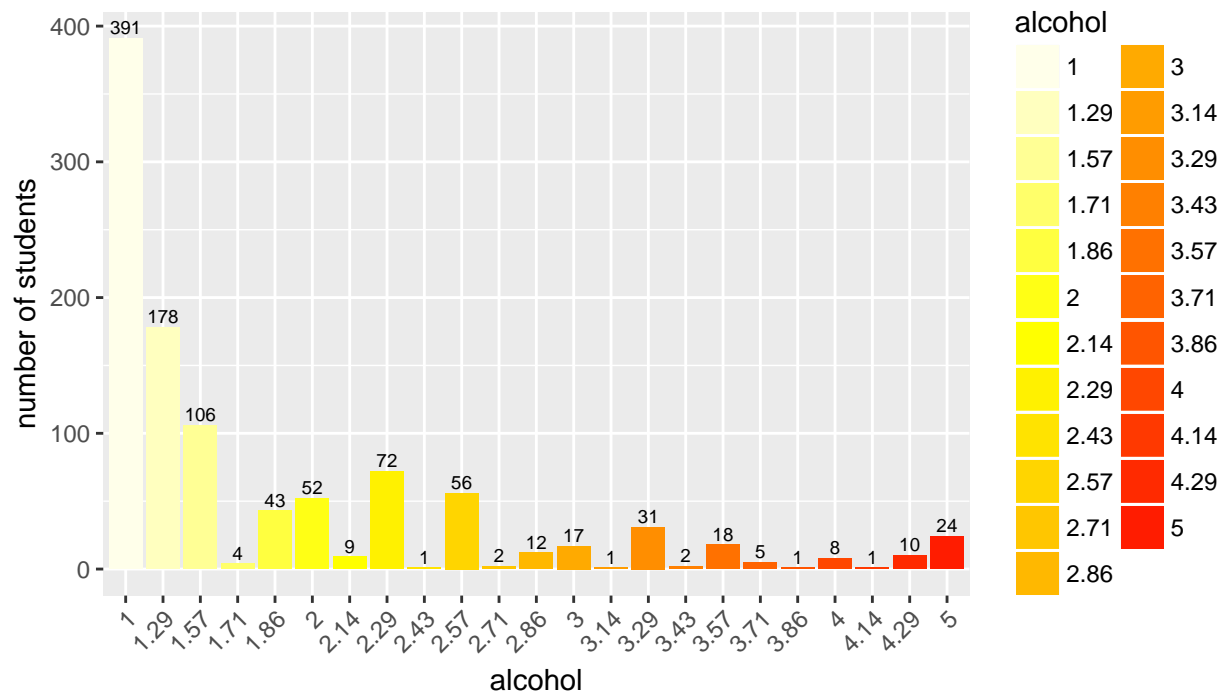
Binary final grade result



## Subject and result



## Global alcohol consumption



```

rm(i)
rm(var)
rm(var_name)
rm(color)
save.image("mywspace_TFM.RData")

```

## Análisis univariante y bivariante

summary(data)

```
##          school          sex          age          address
## Gabriel Pereira      :772  female:591  Min.      :15.00  rural:285
## Mousinho da Silveira:272  male   :453  1st Qu.:16.00  urban:759
##
##                               Median :17.00
##                               Mean    :16.73
##                               3rd Qu.:18.00
##                               Max.    :22.00
##
##          famsize          pstatus          medu
## greater than 3      :738  living apart  :121  none      : 9
## less or equal to 3:306  living together:923  primary   :202
##
##                               5th-9th grade:289
##                               secondary   :238
##                               higher     :306
##
##          fedu          mjob          fjob          reason
## none      : 9  at_home :194  at_home : 62  course   :430
## primary   :256  health  : 82  health  : 41  home     :258
## 5th-9th grade:324  other   :399  other   :584  other    :108
## secondary :231  services:239  services:292  reputation:248
## higher    :224  teacher :130  teacher : 65
##
##          guardian          traveltime          studytime          failures
## father:243 <15 min.      :623 <2 hours      :317  Min.      :0.0000
## mother:728 15 to 30 min. :320 2 to 5 hours :503  1st Qu.:0.0000
## other : 73 30 min to 1 hour: 77 5 to 10 hours:162  Median :0.0000
##
##                               >1 hour      : 24 >10 hours      : 62  Mean      :0.2644
##
##                               3rd Qu.:0.0000
##                               Max.      :3.0000
##
## schoolsup famsup      paid      activities nursery      higher      internet
## no :925  no :404  no :824  no :528  no :209  no : 89  no :217
## yes:119  yes:640  yes:220  yes:516  yes:835  yes:955  yes:827
##
##
##
##          romantic          famrel          freetime          goout          dalc
## no :673  very bad : 30  very low : 64  very low : 71  very low :727
## yes:371  bad      : 47  low      :171  low      :248  low      :196
##
##          fair :169  medium :408  medium :335  medium : 69
##          good :512  high   :293  high   :227  high   : 26
##          excellent:286  very high:108  very high:163  very high: 26
##
##          walc          health          absences          g1
## very low :398  very bad :137  Min.      : 0.000  Min.      : 0.00
## low      :235  bad      :123  1st Qu.: 0.000  1st Qu.: 9.00
## medium   :200  fair     :215  Median : 2.000  Median :11.00
## high     :138  good     :174  Mean    : 4.435  Mean    :11.21
## very high: 73  very good:395  3rd Qu.: 6.000  3rd Qu.:13.00
##
##                               Max.      :75.000  Max.      :19.00
```

```
##          g2          g3          file          subject
## Min.    : 0.00   Min.    : 0.00   math       :395   math       : 25
## 1st Qu.: 9.00   1st Qu.:10.00  portuguese:649  math&portuguese:751
## Median :11.00   Median :11.00                                     portuguese :268
## Mean    :11.25   Mean     :11.34
## 3rd Qu.:13.00   3rd Qu.:14.00
## Max.    :19.00   Max.     :20.00
## student_key      bing3          result          alcohol
## Length:1044      fail:230   fail math    :130   Min.    :1.000
## Class :character pass:814   fail portuguese:100 1st Qu.:1.000
## Mode  :character pass math    :265   Median :1.290
##                                     pass portuguese:549 Mean    :1.721
##                                     3rd Qu.:2.178
##                                     Max.    :5.000
```

```
data_num<- data[variables_numericas]
```

```
summary(data_num)
```

```
##          age          failures          absences          g1
## Min.    :15.00   Min.    :0.0000   Min.    : 0.000   Min.    : 0.00
## 1st Qu.:16.00   1st Qu.:0.0000   1st Qu.: 0.000   1st Qu.: 9.00
## Median :17.00   Median :0.0000   Median : 2.000   Median :11.00
## Mean    :16.73   Mean     :0.2644   Mean     : 4.435   Mean    :11.21
## 3rd Qu.:18.00   3rd Qu.:0.0000   3rd Qu.: 6.000   3rd Qu.:13.00
## Max.    :22.00   Max.     :3.0000   Max.     :75.000   Max.    :19.00
##          g2          g3          alcohol
## Min.    : 0.00   Min.    : 0.00   Min.    :1.000
## 1st Qu.: 9.00   1st Qu.:10.00   1st Qu.:1.000
## Median :11.00   Median :11.00   Median :1.290
## Mean    :11.25   Mean     :11.34   Mean    :1.721
## 3rd Qu.:13.00   3rd Qu.:14.00   3rd Qu.:2.178
## Max.    :19.00   Max.     :20.00   Max.    :5.000
```

```
sapply(data_num, sd)
```

```
##          age failures absences          g1          g2          g3 alcohol
## 1.2399747 0.6561418 6.2100166 2.9833939 3.2850711 3.8647958 0.9269683
```

*# como habíamos visto tanto en el boxplot como en el diagrama de barras la variable con mayor dispersión*

```
correlacion <- cor ( data_num )
```

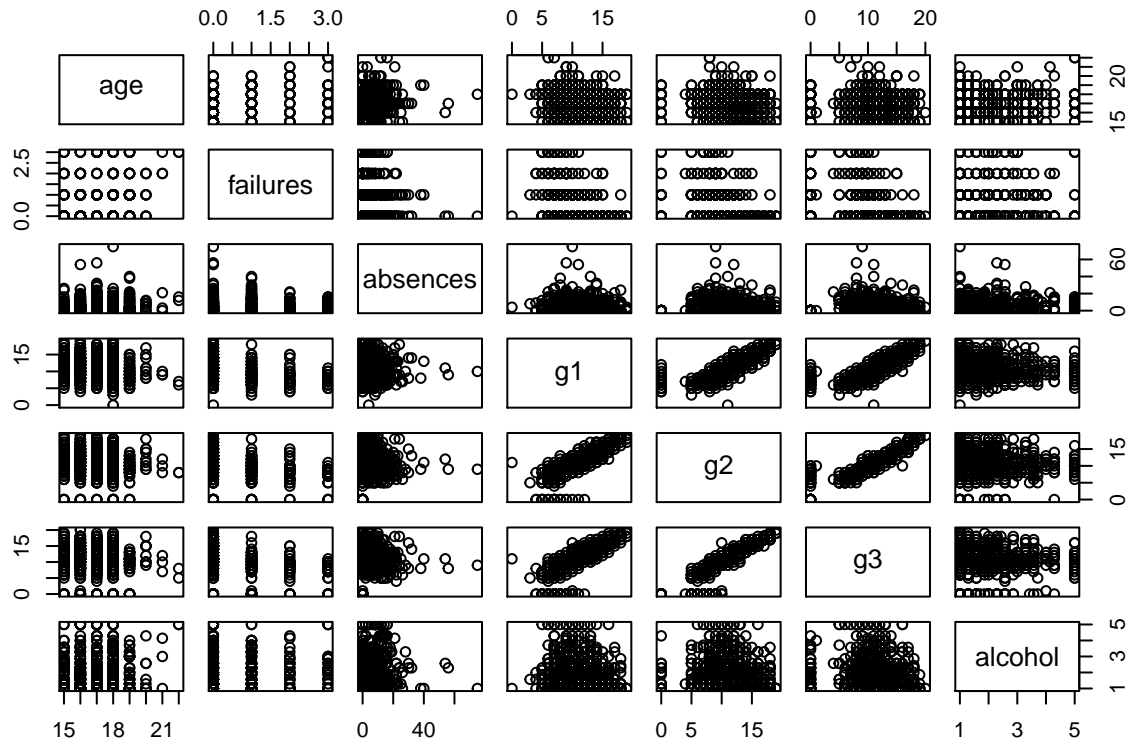
```
correlacion
```

```
##          age          failures          absences          g1          g2
## age      1.0000000 0.28236357 0.15319565 -0.12412125 -0.11947474
## failures 0.2823636 1.00000000 0.09999785 -0.37417487 -0.37717218
## absences 0.1531956 0.09999785 1.00000000 -0.09242463 -0.08933169
## g1       -0.1241212 -0.37417487 -0.09242463 1.00000000 0.85873875
## g2       -0.1194747 -0.37717218 -0.08933169 0.85873875 1.00000000
## g3       -0.1252824 -0.38314528 -0.04567058 0.80914172 0.91074316
## alcohol  0.1327481 0.12410073 0.14877659 -0.16222869 -0.14306963
##          g3          alcohol
## age      -0.12528243 0.1327481
## failures -0.38314528 0.1241007
```

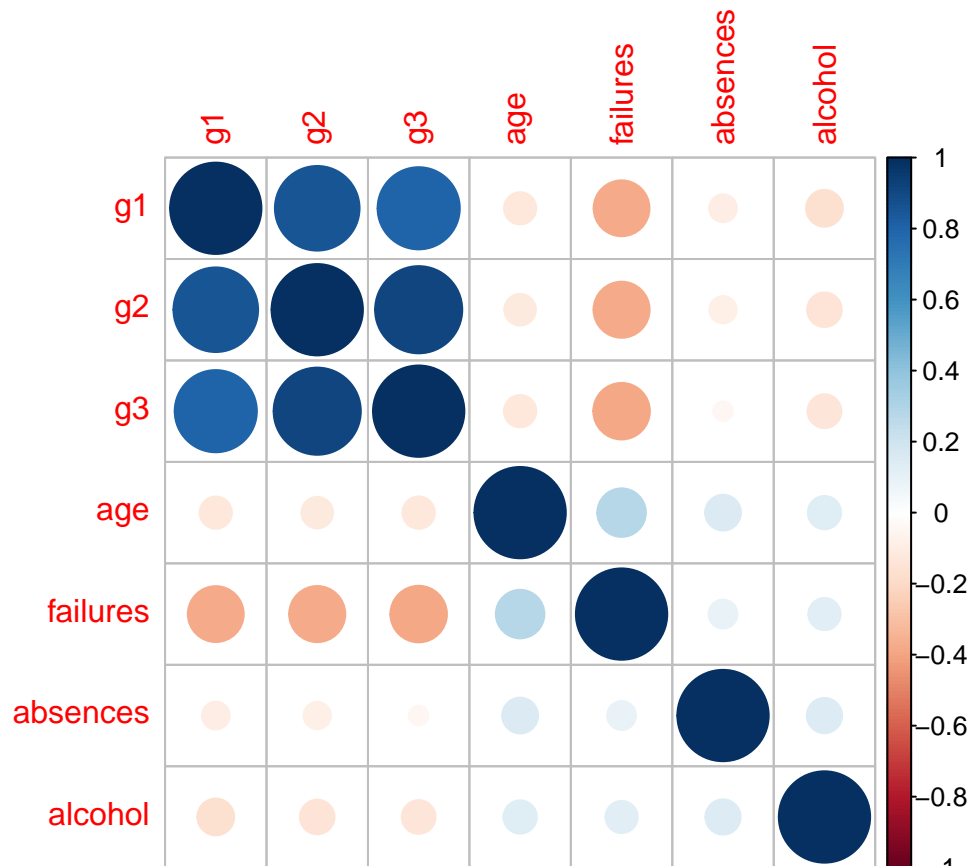


```
## absences -0.04567058  0.1487766
## g1        0.80914172 -0.1622287
## g2        0.91074316 -0.1430696
## g3        1.00000000 -0.1367715
## alcohol  -0.13677155  1.0000000
```

```
plot ( data_num )
```

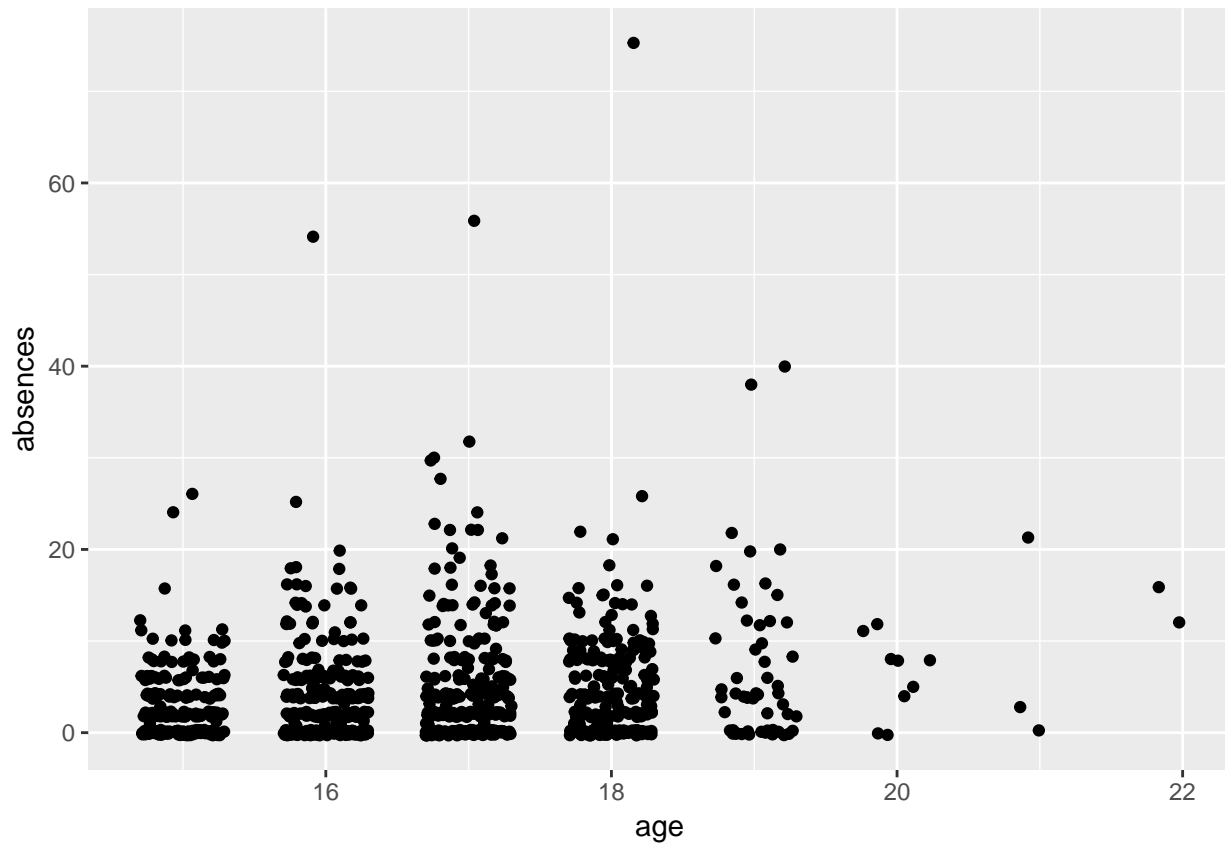


```
library(corrplot)
corrplot ( correlacion ,order="hclust")
```



```
library(ggplot2)

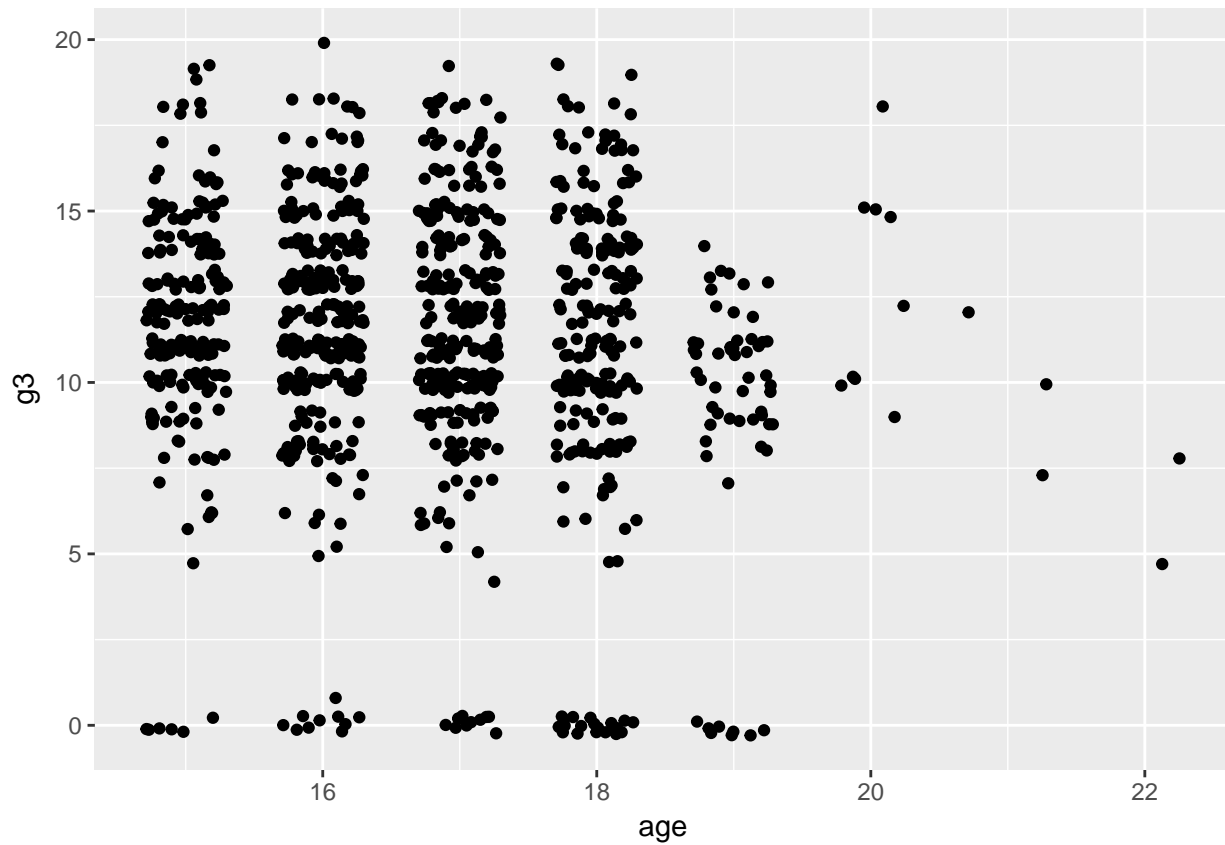
## Analisis age + absences
plot1 <- ggplot (data,aes (x=age,y=absences)) +
geom_point ( position=position_jitter ( w=0.3,h=0.3 ) )
plot1
```



```
round(tapply ( data$absences,data$age,mean ),1)
```

```
##  15  16  17  18  19  20  21  22
##  3.1  3.9  4.9  4.9  6.9  6.2  8.0 14.0
```

```
## Analisis age + g3
plot2 <- ggplot (data,aes (x=age,y=g3)) +
geom_point ( position=position_jitter ( w=0.3,h=0.3 ) )
plot2
```

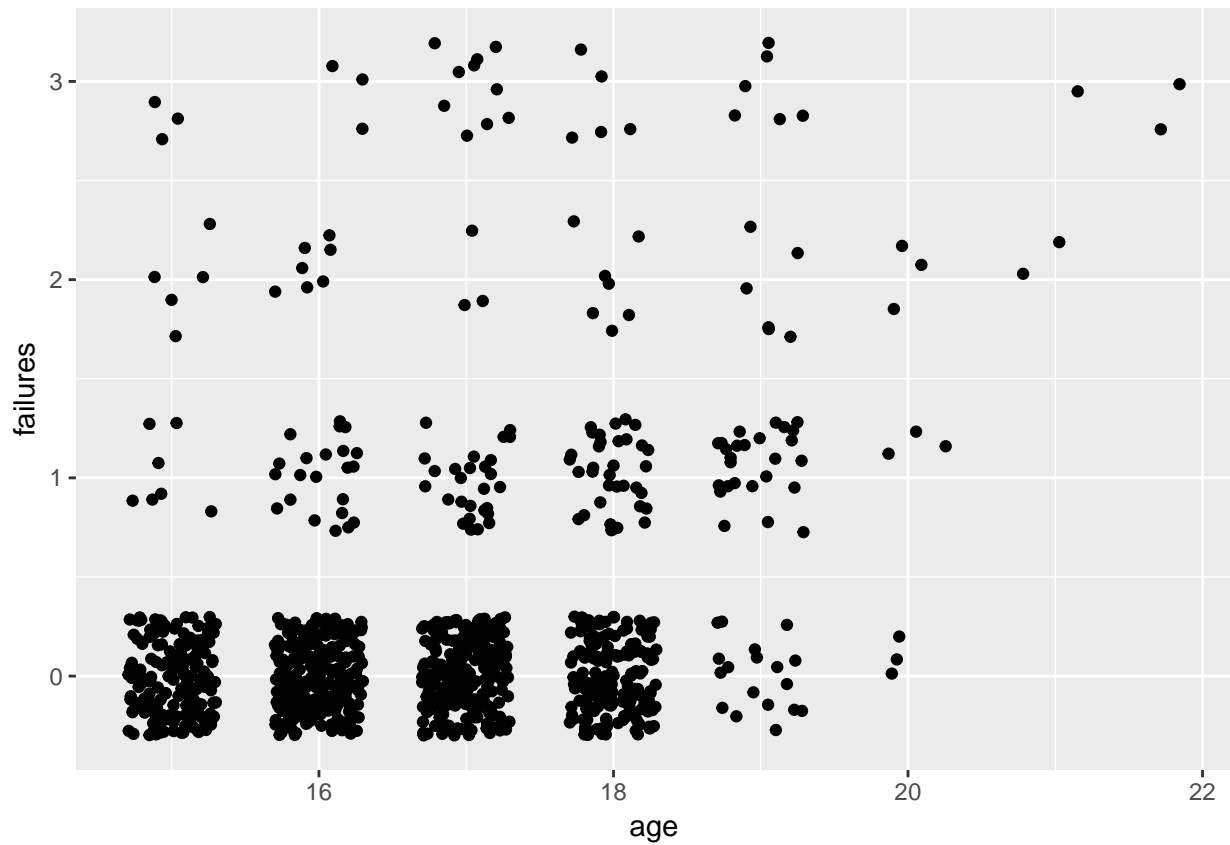


```
round(tapply ( data$g3,data$age,mean ),1)
```

```
## 15 16 17 18 19 20 21 22
## 11.7 11.6 11.6 11.0 9.0 12.7 9.7 6.5
```

```
## Analisis age + failures
```

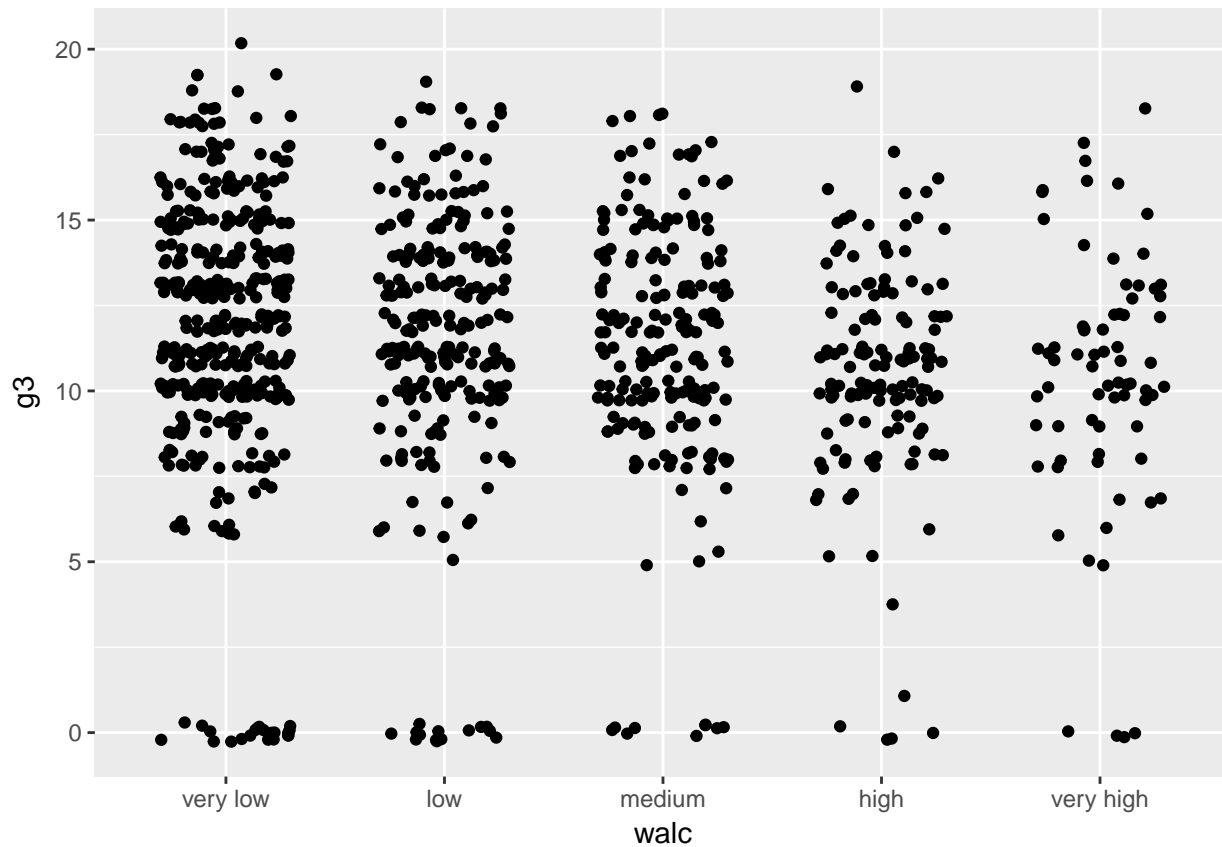
```
plot3 <- ggplot (data,aes (x=age,y=failures)) +
geom_point ( position=position_jitter ( w=0.3,h=0.3 ) )
plot3
```



```
round(tapply ( data$failures,data$age,mean ),1)
```

```
## 15 16 17 18 19 20 21 22
## 0.1 0.2 0.2 0.3 1.0 1.0 2.3 3.0
```

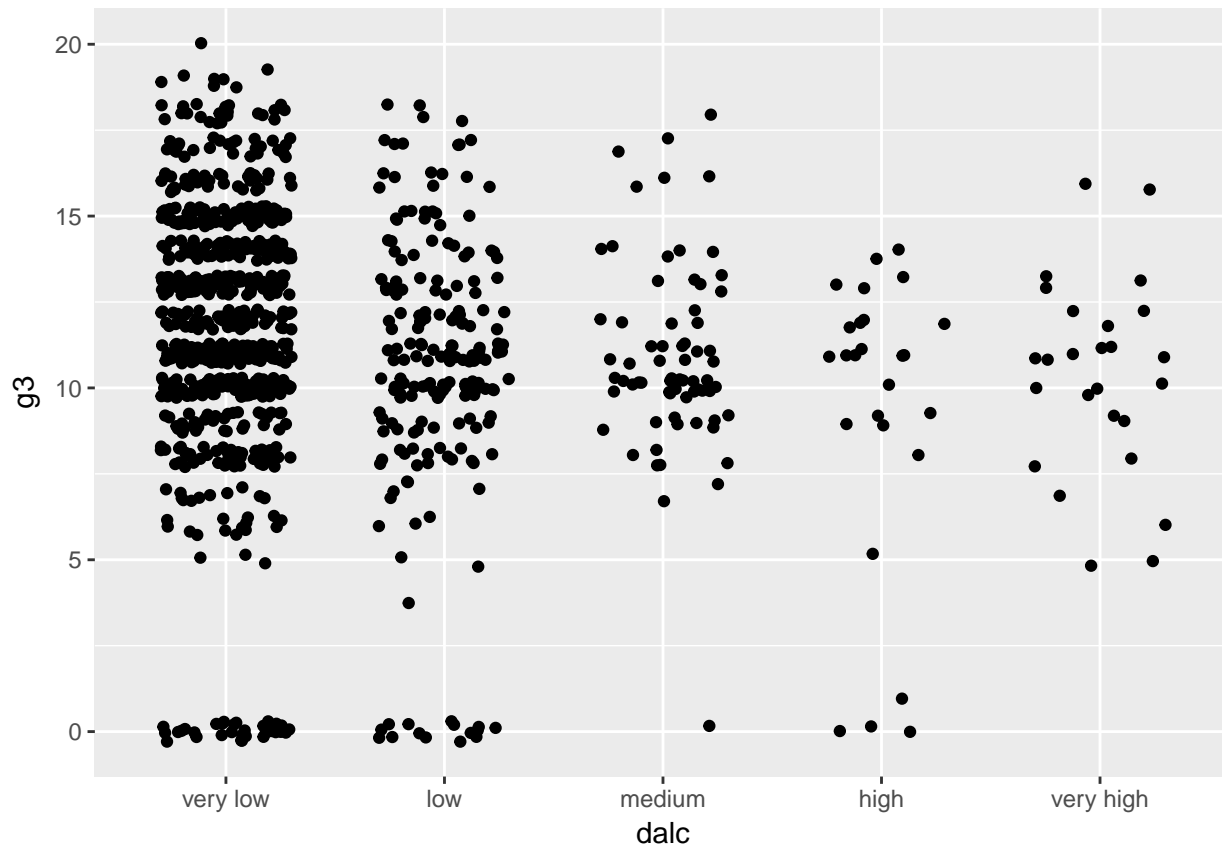
```
## Analisis walc + g3
plot4 <- ggplot (data,aes (x=walc,y=g3)) +
geom_point ( position=position_jitter ( w=0.3,h=0.3 ) )
plot4
```



```
round(tapply ( data$g3,data$walc,mean ),1)
```

```
## very low      low      medium      high very high
##      11.7      11.5      11.3      10.5      10.4
```

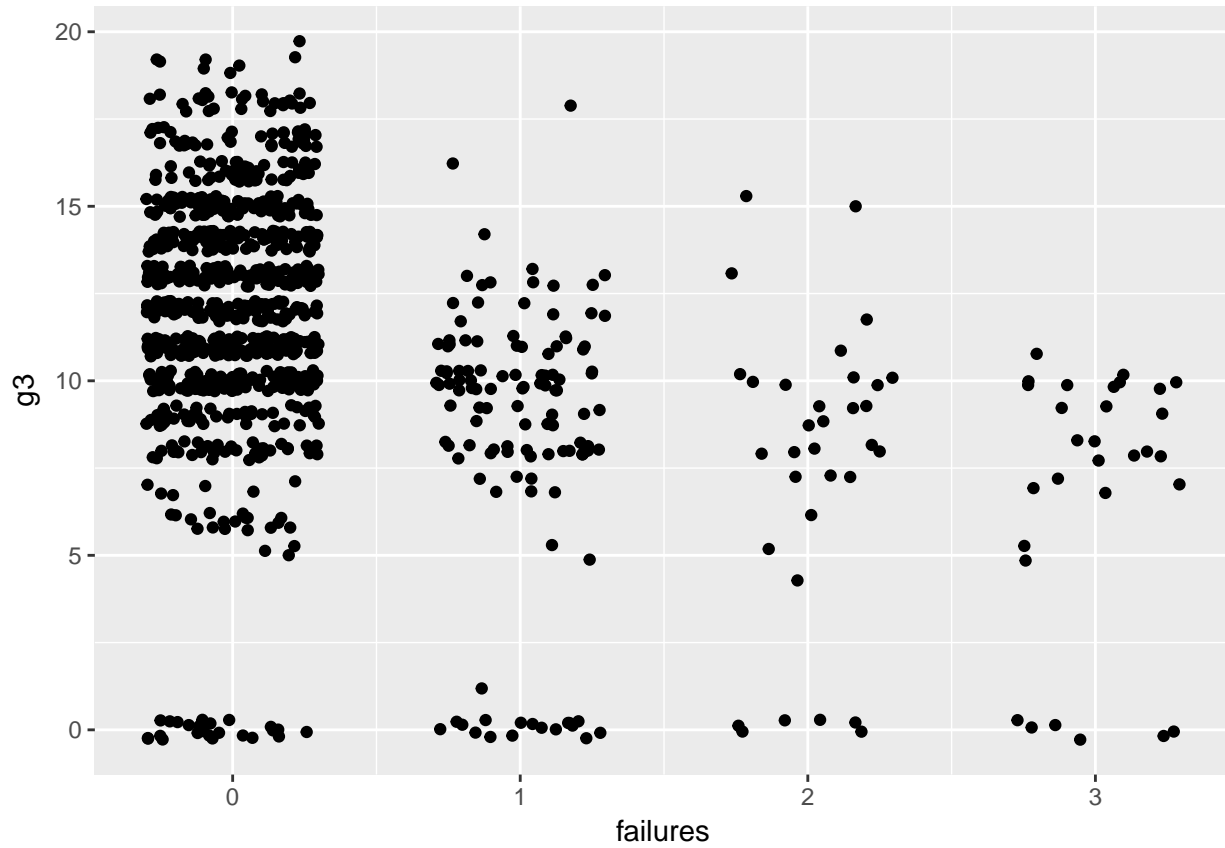
```
## Analisis dalc + g3
plot5 <- ggplot (data,aes (x=dalc,y=g3)) +
geom_point ( position=position_jitter ( w=0.3,h=0.3 ) )
plot5
```



```
round(tapply ( data$g3,data$dalc,mean ),1)
```

```
## very low      low      medium      high very high
##      11.7      10.6      10.9      9.3      10.4
```

```
## Analisis bivariable edad + asignatura no superada
plot7 <- ggplot ( data,aes ( x=failures,y=g3 ) ) +
geom_point ( position=position_jitter ( w=0.3,h=0.3 ) )
plot7
```

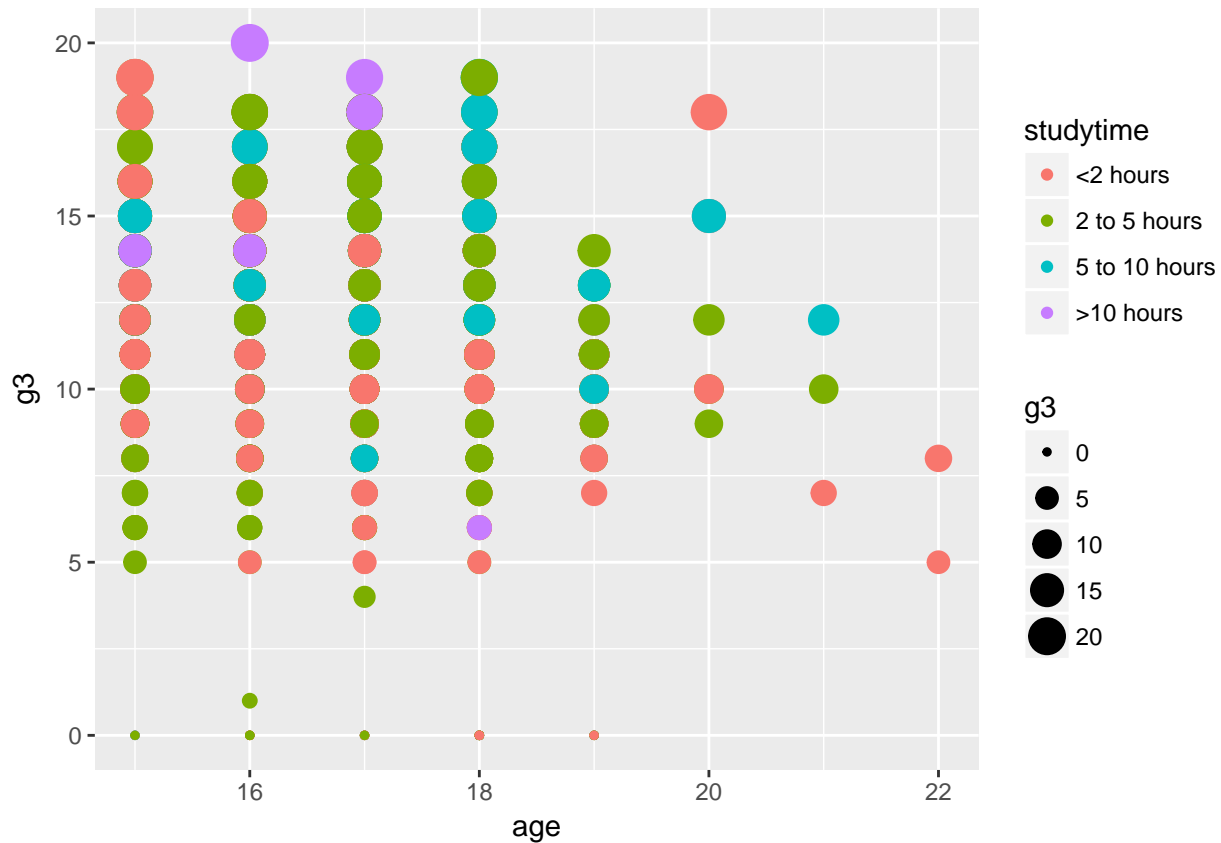


```
round(tapply ( data$g3,data$failures,mean ),1)
```

```
## 0 1 2 3
## 12.1 8.4 7.5 6.8
```

```
## Relacion edad + nota final + horas de estudio
g4 <- ggplot ( data,aes ( x=age,y=g3,color =studytime ) ) + geom_point ( aes ( size=g3 ) )
##+ scale_color_manual ( values=c ( "#33 CCF", "green" ,"yellow" ,"orange" ,"red", "#660066" ))
g4
```





```
## Estudiamos la existencia de diferencias por Fracaso y la media de cada una del resto de variables num
round(tapply ( data$g3,data$age,mean ),1)
```

```
## 15 16 17 18 19 20 21 22
## 11.7 11.6 11.6 11.0 9.0 12.7 9.7 6.5
```

```
tapply ( data$absences,data$g3,mean )
```

```
##      0      1      4      5      6      7      8
## 0.000000 0.000000 22.000000 11.500000 8.277778 6.263158 7.865672
##      9     10     11     12     13     14     15
## 6.841270 4.797386 5.119205 3.786408 3.743363 3.766667 3.012195
##     16     17     18     19     20
## 3.115385 2.685714 3.666667 3.571429 4.000000
```

```
tapply ( data$failures,data$g3,mean )
```

```
##      0      1      4      5      6      7
## 0.88679245 1.00000000 2.00000000 1.25000000 0.11111111 1.26315789
##      8      9     10     11     12     13
## 0.73134328 0.47619048 0.42483660 0.13245033 0.08737864 0.08849558
##     14     15     16     17     18     19
## 0.01111111 0.04878049 0.01923077 0.00000000 0.03703704 0.00000000
##     20
## 0.00000000
```

```
## Tabla de contingencia sexo respecto soportes escolar y familiar
table ( data$sex,data$schoolsup )
```

```
##
```

```
##           no yes
##  female 504  87
##  male   421  32

table ( data$sex,data$famsup )

##
##           no yes
##  female 194 397
##  male   210 243

## Tabla de contingencia sexo respecto supera
X <- table ( data$sex, data$bing3 )
X

##
##           fail pass
##  female 125 466
##  male   105 348

## Calculamos la proporcion
prop.table(X,1)

##
##           fail      pass
##  female 0.2115059 0.7884941
##  male   0.2317881 0.7682119
```

## Segmentación

```
variables_numericas

## [1] "age"      "failures" "absences" "g1"      "g2"      "g3"
## [7] "alcohol"
```

```
variables_cluster<-c("age","failures","absences","g1","g2","g3")

#se usa el dataset en el que se han eliminado los valores atípicos
data_cluster <-data_clean[variables_cluster]

head(data_cluster)

##  age failures absences g1 g2 g3
## 1  18         0         6  5  6  6
## 2  17         0         4  5  5  6
## 3  15         3        10  7  8 10
## 4  15         0         2 15 14 15
## 5  16         0         4  6 10 10
## 6  16         0        10 15 15 15

# data_cluster_norm será el data.frame con los datos normalizados
data_cluster_norm<-data_cluster

# se sustituyen las columnas de data_cluster_norm por las columnas de data_cluster normalizadas:

data_cluster_norm[,c("age")] <- (data_cluster$age-mean(data_cluster$age))/sd(data_cluster$age)
```

```

data_cluster_norm[,c("failures")] <- (data_cluster$failures-mean(data_cluster$failures))/sd(data_clus
data_cluster_norm[,c("absences")] <- (data_cluster$absences-mean(data_cluster$absences))/sd(data_clus
data_cluster_norm[,c("g1")] <- (data_cluster$g1-mean(data_cluster$g1))/sd(data_cluster$g1)
data_cluster_norm[,c("g2")] <- (data_cluster$g2-mean(data_cluster$g2))/sd(data_cluster$g2)
data_cluster_norm[,c("g3")] <- (data_cluster$g3-mean(data_cluster$g3))/sd(data_cluster$g3)

#data_cluster_norm[,c("alcohol")] <- (data_cluster$alcohol-mean(data_cluster$alcohol))/sd(data_cluste

```

Usamos kmeans con k=3 y visualizamos algunos resultados: los centros de grupos (centers), la suma de cuadrados totales (totss), las sumas de cuadrados dentro de cada grupo y para todos de forma conjunta (withinss y tot.withinss) y la suma de cuadrados entre grupos (betweenss).

```

set.seed(123)
clus2_k3<-kmeans(data_cluster_norm,centers=3)
clus2_k3$centers

##          age  failures  absences          g1          g2          g3
## 1  0.7008034  1.6507186  0.05043198 -1.2234992 -1.3724723 -1.5131042
## 2 -0.1492691 -0.3354576 -0.25601065  0.9847198  0.9201951  0.8487943
## 3 -0.1137588 -0.2816977  0.20064553 -0.4179380 -0.3117998 -0.2026738

clus2_k3$totss

## [1] 5934

clus2_k3$withinss

## [1] 1136.7126  945.5809 1221.9351

clus2_k3$tot.withinss

## [1] 3304.229

clus2_k3$betweenss

## [1] 2629.771

```

Para elegir el número de clústers más apropiado buscaremos el k que haga que los individuos pertenecientes a un mismo grupo son lo más homogéneos posible y los individuos pertenecientes a distintos grupos son lo más heterogéneos posible teniendo en cuenta que cuanto más grande sea k más cálculos debe hacer el algoritmo.

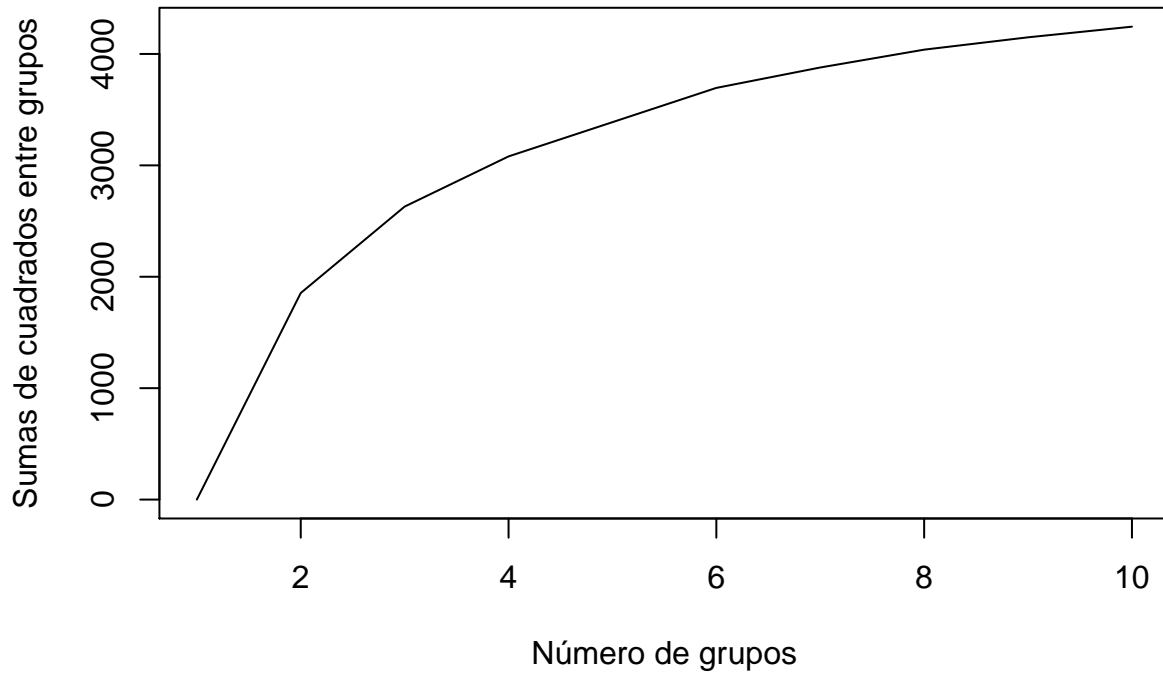
Los parámetros que usamos son las sumas de cuadrados dentro de todos los grupos (tot.withinss) y la suma de cuadrados entre grupos (betweenss).

Queremos maximizar la suma de cuadrados entre grupos y minimizar la suma de cuadrados dentro de los grupos. Veamos si podemos encontrar un k que cumpla con las dos condiciones.

```

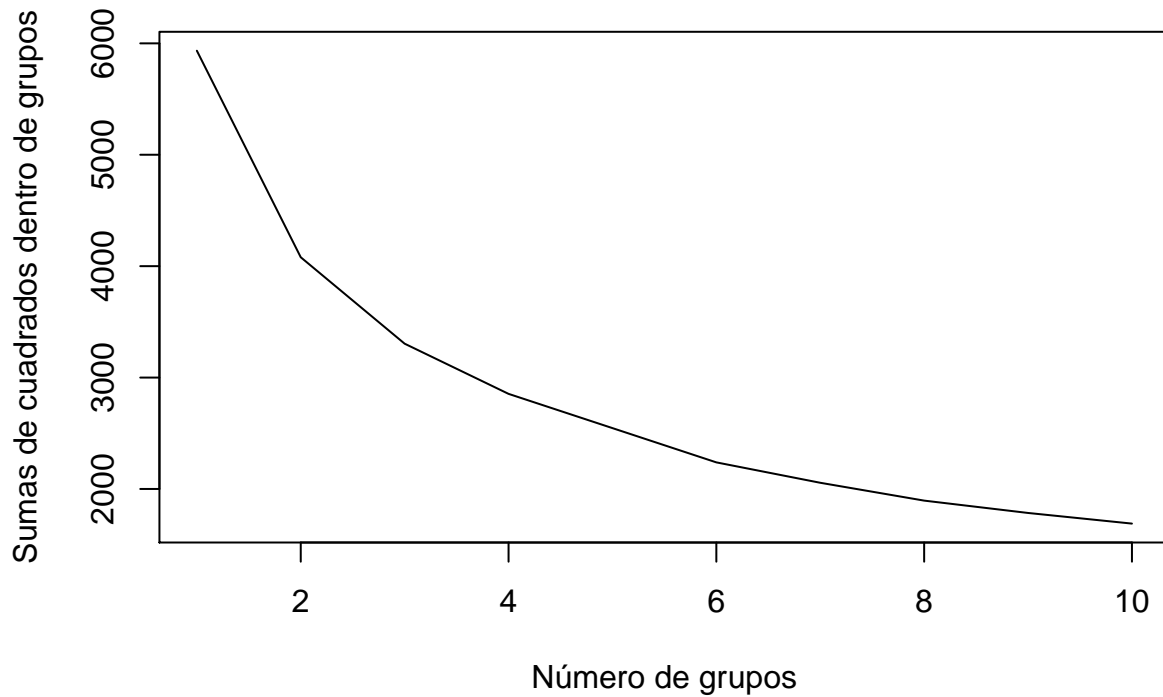
set.seed(123)
bss <- kmeans(data_cluster_norm,centers=1)$betweenss
for (i in 2:10) bss[i] <- kmeans(data_cluster_norm,centers=i)$betweenss
plot(1:10, bss, type="l", xlab="Número de grupos",ylab="Sumas de cuadrados entre grupos")

```



```
set.seed(123)
tw <- kmeans(data_cluster_norm,centers=1)$tot.withinss
for (i in 2:10) tw[i] <- kmeans(data_cluster_norm,centers=i)$tot.withinss

plot(1:10, tw, type="l", xlab="Número de grupos",ylab="Sumas de cuadrados dentro de grupos")
```



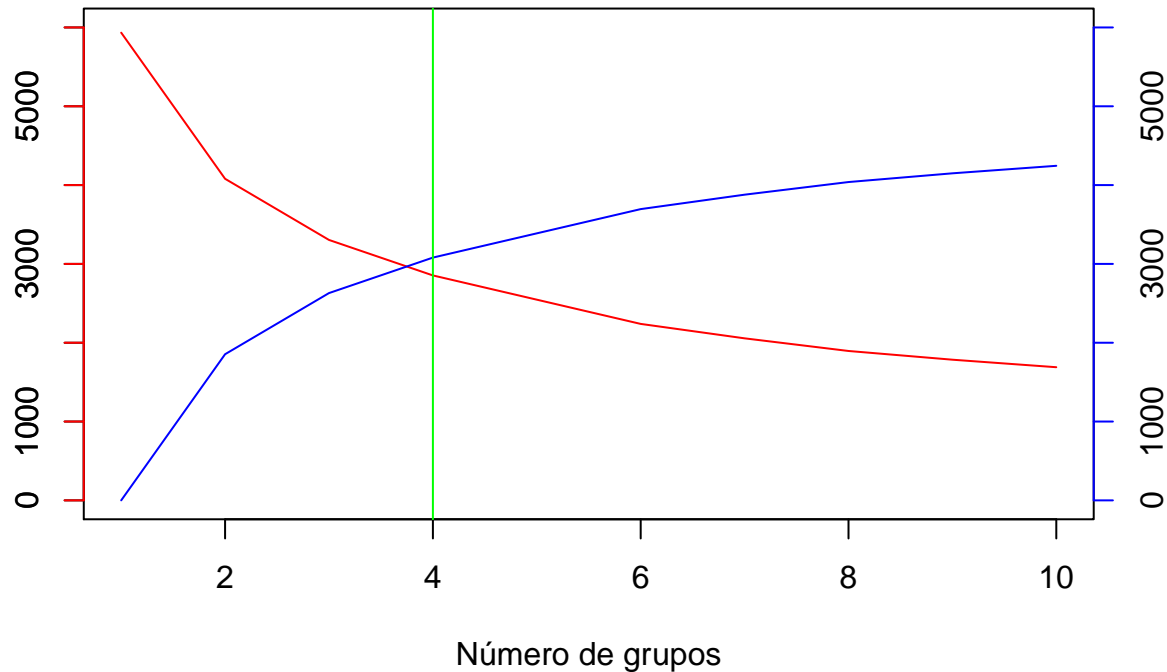
```
# fusionamos los dos gráficos
{
  plot(1:10, tw, type="l",col="red", xlab="",ylab="", ylim=c(0,6000))
}
```

```

axis(side=2, col="red")
par(new=TRUE)
plot(1:10, bss, type="l", col="blue", xlab="Número de grupos", ylab="", ylim=c(0,6000), axes=FALSE)
axis(side=4, col="blue")
title(main="Sumas de cuadrados dentro de grupos (rojo) y entre grupos (azul)")
abline(v=4, col="green")
}

```

## Sumas de cuadrados dentro de grupos (rojo) y entre grupos (azul)



Para  $k=4$  se consigue un punto de equilibrio entre el número de grupos y el objetivo de minimizar la suma de cuadrados dentro de grupos y maximizar la suma de cuadrados entre grupos. A partir de ese valor de  $k$ , la disminución y el aumento respectivamente son muy inferiores a los conseguidos en las transiciones anteriores.

## Segmentación óptima

Se calcula la segmentación para  $k=4$ .

```

set.seed(123)
clus2_k4<-kmeans(data_cluster_norm,centers=4)
clus2_k4$centers

```

```

##          age    failures  absences      g1      g2      g3
## 1  0.61781001  1.767628615 -0.3089366 -1.2447718 -1.4555486 -1.6605118
## 2 -0.08973206 -0.343459054 -0.2660276  1.1078193  1.0426241  0.9529763
## 3 -0.42108419 -0.310166004 -0.3525451 -0.3243741 -0.2315884 -0.1413871
## 4  0.63685887 -0.006868061  1.5078740 -0.3877121 -0.3149020 -0.1955571

```

```
clus2_k4$totss
```

```
## [1] 5934
```

```
clus2_k4$withinss
```

```
## [1] 909.5253 701.8153 683.1203 560.0572
```

```
clus2_k4$tot.withinss
```

```
## [1] 2854.518
```

```
clus2_k4$betweenss
```

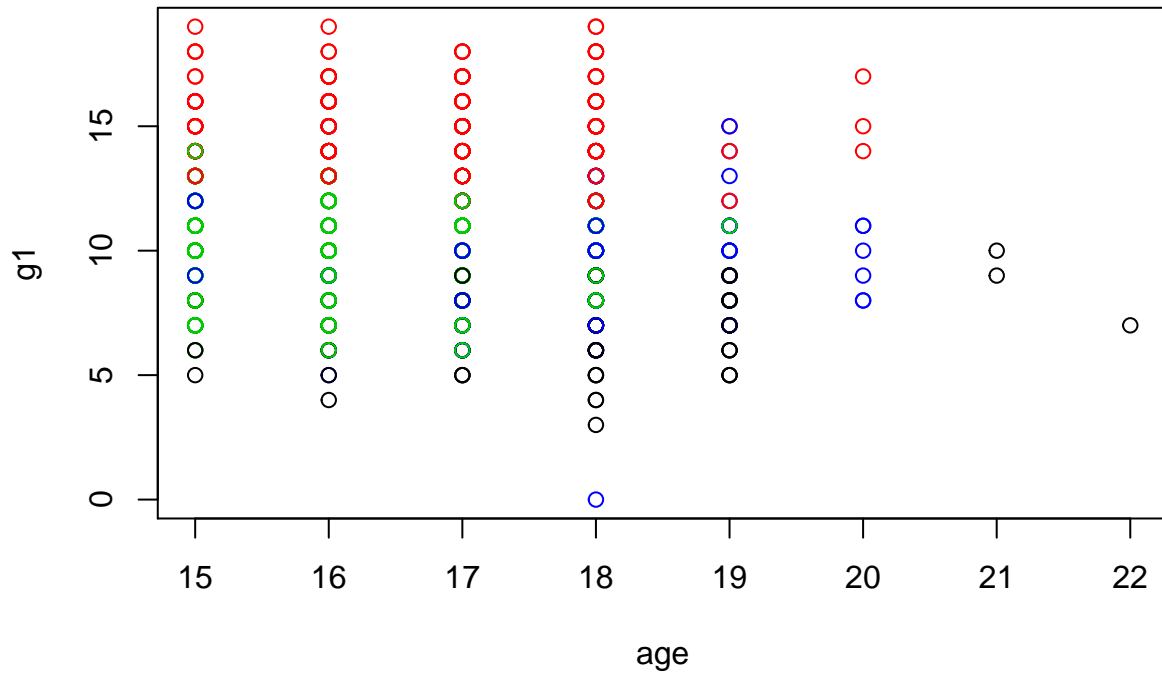
```
## [1] 3079.482
```

```
names(data_cluster_norm)<-c("age_st","failures_st","absences_st", "g1_st", "g2_st", "g3_st")
data_cluster_result<-data.frame(data_cluster, data_cluster_norm, as.factor(clus2_k4$cluster), data_clean)
#names(data_cluster_result)[15]<-"cluster"
names(data_cluster_result)[13]<-"cluster"
names(data_cluster_result)[14]<-"alcohol"
head(data_cluster_result)
```

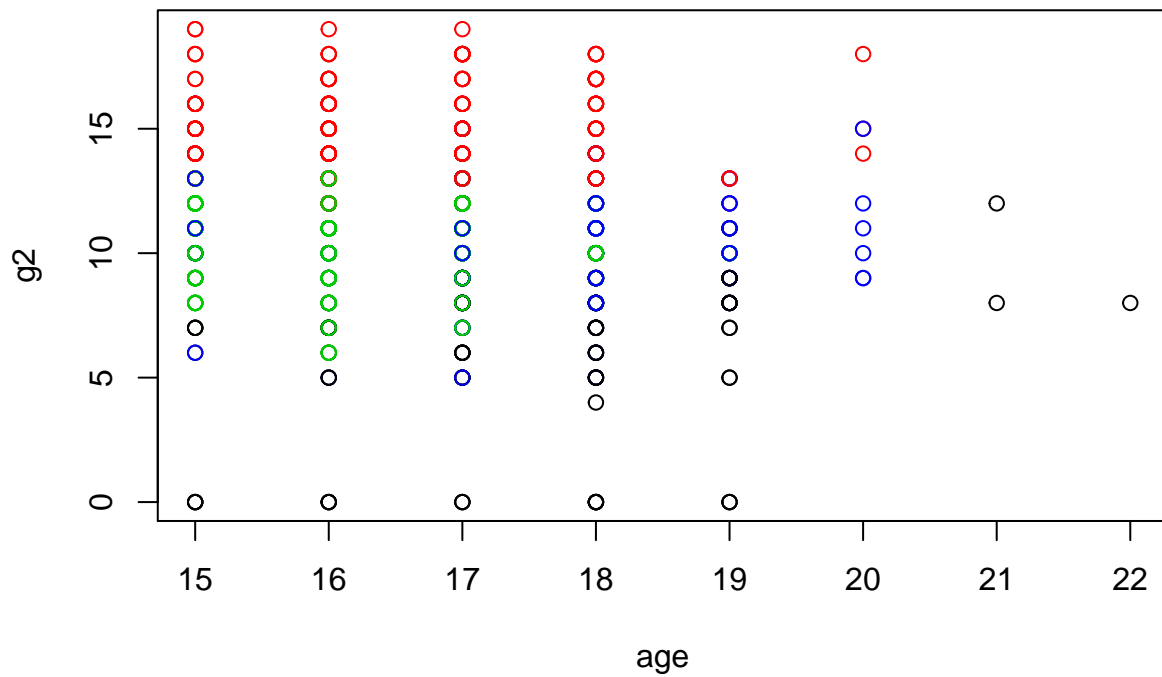
```
##   age failures absences g1 g2 g3   age_st failures_st absences_st
## 1  18         0         6 5 6 6  1.0640333 -0.3876334  0.6787848
## 2  17         0         4 5 5 6  0.2480821 -0.3876334  0.1509003
## 3  15         3        10 7 8 10 -1.3838203  4.2359454  1.7345538
## 4  15         0         2 15 14 15 -1.3838203 -0.3876334 -0.3769842
## 5  16         0         4 6 10 10 -0.5678691 -0.3876334  0.1509003
## 6  16         0        10 15 15 15 -0.5678691 -0.3876334  1.7345538
##      g1_st      g2_st      g3_st cluster alcohol
## 1 -2.088010 -1.6070407 -1.3840741      1      1.00
## 2 -2.088010 -1.9100250 -1.3840741      1      1.00
## 3 -1.419685 -1.0010722 -0.3584503      1      2.29
## 4  1.253616  0.8168333  0.9235794      2      1.00
## 5 -1.753848 -0.3951037 -0.3584503      3      1.29
## 6  1.253616  1.1198176  0.9235794      2      1.29
```

Veamos algunos gráficos:

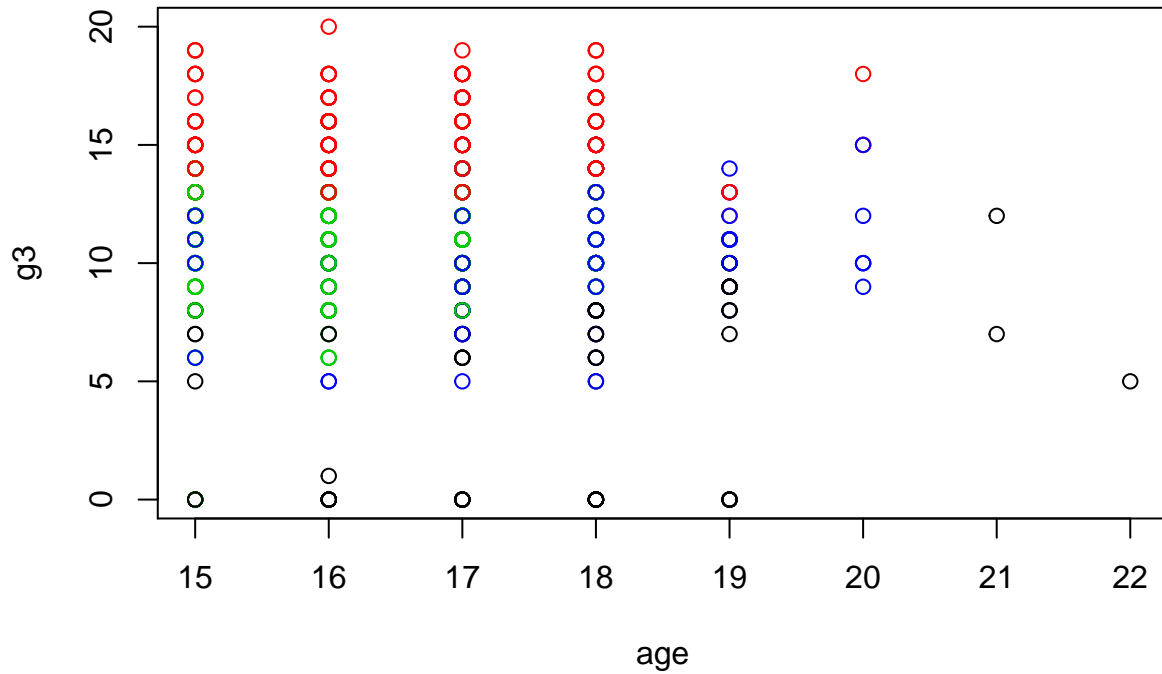
```
plot(data_cluster_result[c("age", "g1")], col=data_cluster_result$cluster)
```



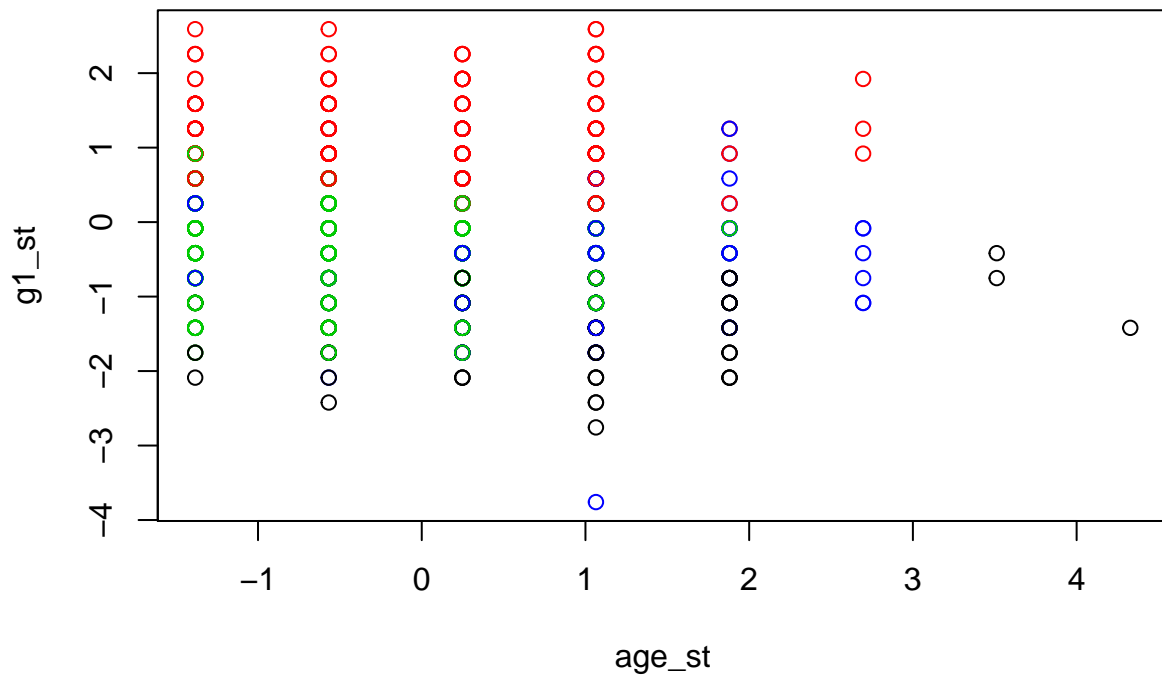
```
plot(data_cluster_result[c("age", "g2")], col=data_cluster_result$cluster)
```



```
plot(data_cluster_result[c("age", "g3")], col=data_cluster_result$cluster)
```

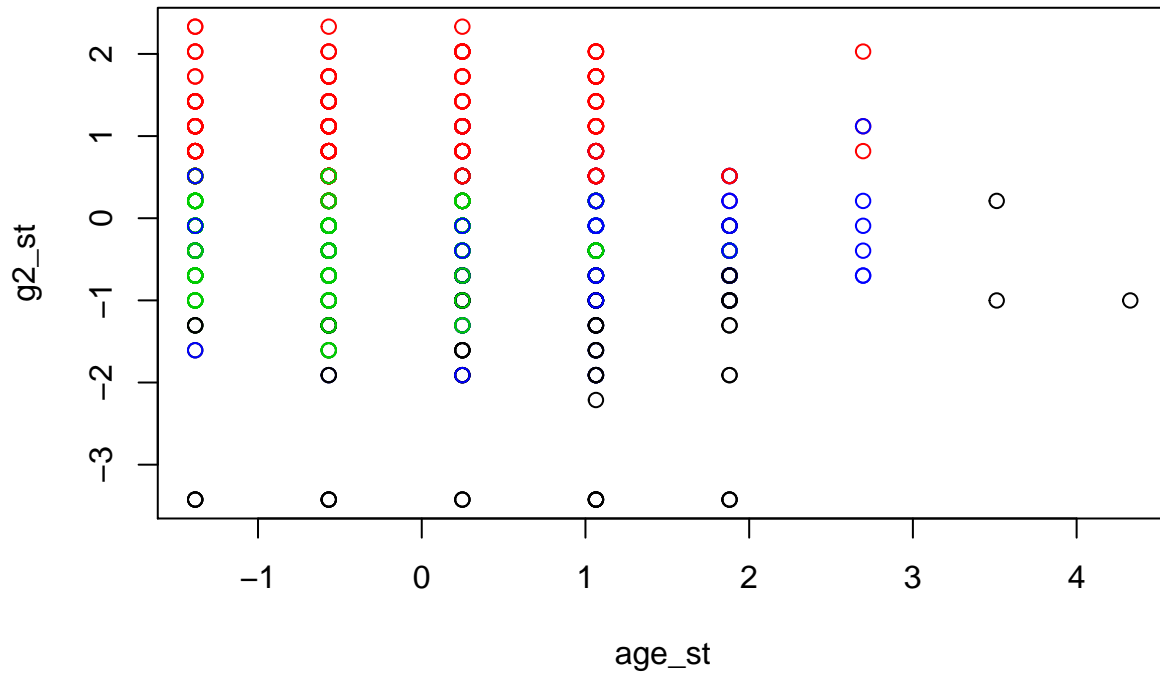


```
plot(data_cluster_result[c("age_st", "g1_st")], col=data_cluster_result$cluster)
```

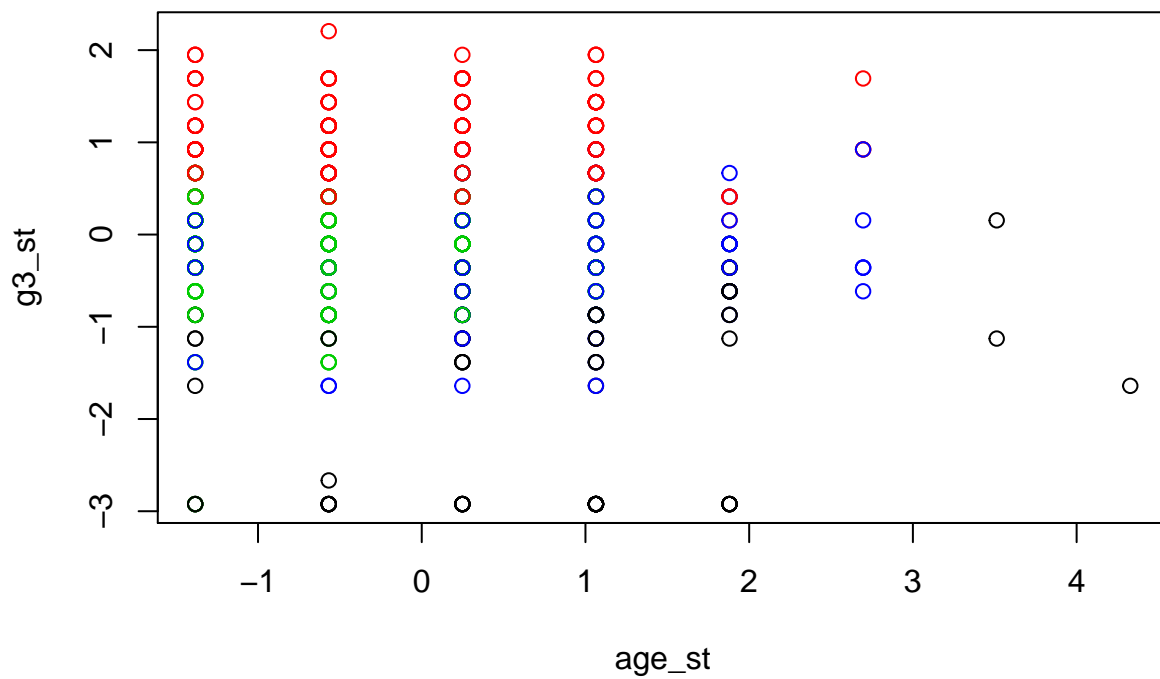


```
plot(data_cluster_result[c("age_st", "g2_st")], col=data_cluster_result$cluster)
```

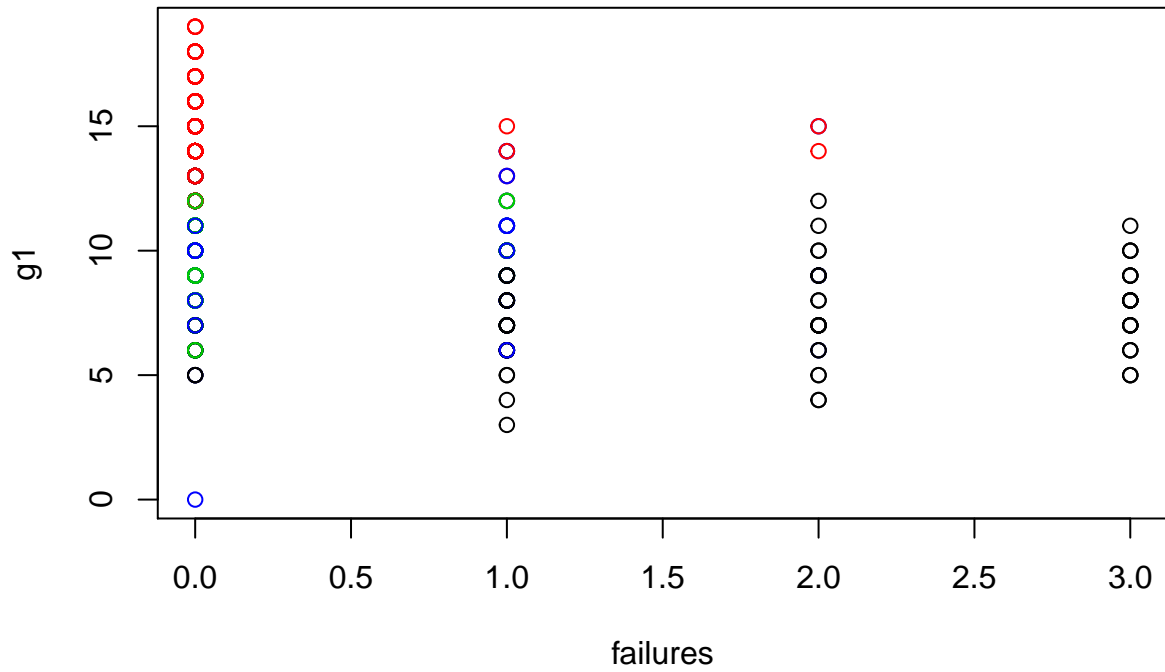




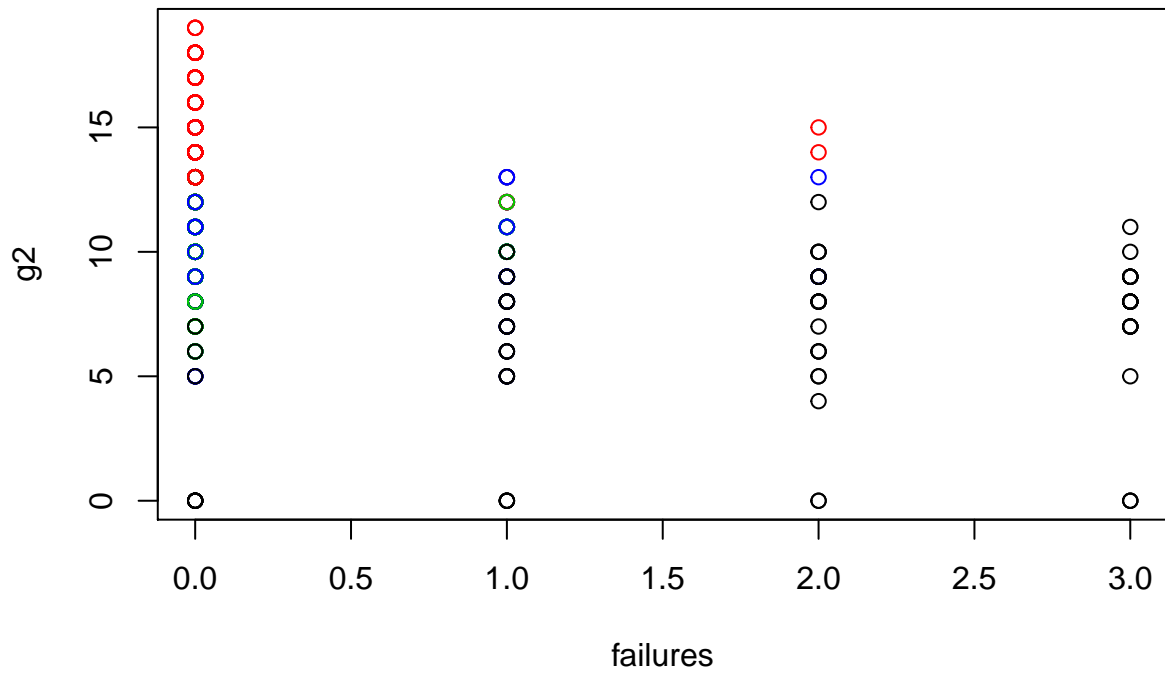
```
plot(data_cluster_result[c("age_st", "g3_st")], col=data_cluster_result$cluster)
```



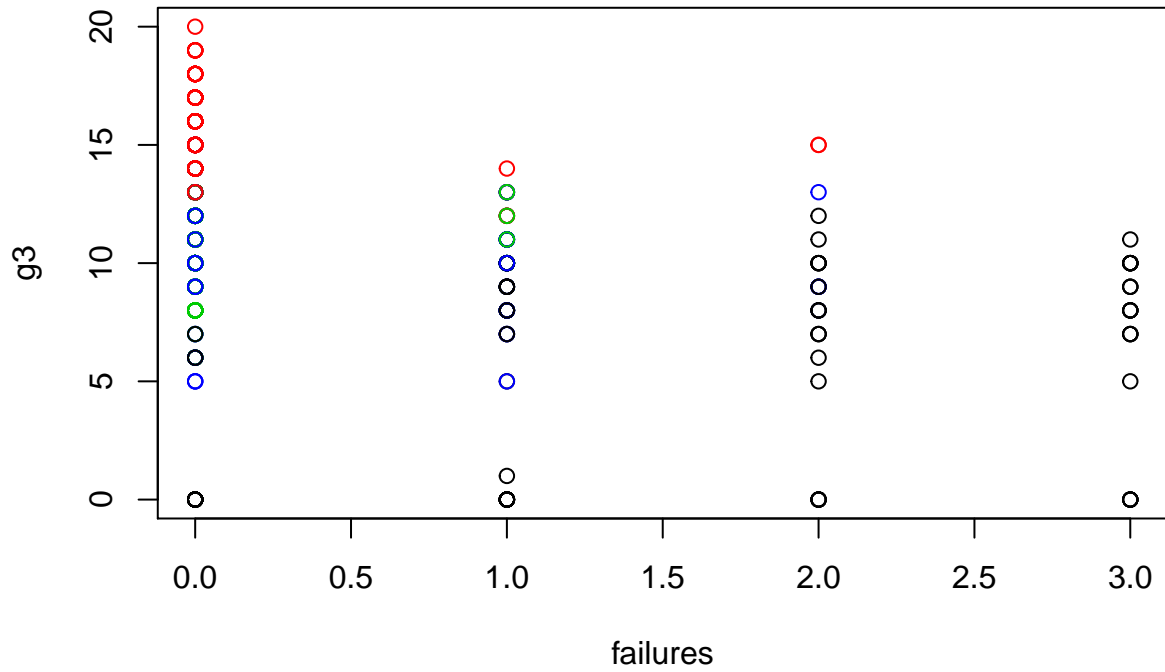
```
plot(data_cluster_result[c("failures", "g1"), col=data_cluster_result$cluster)
```



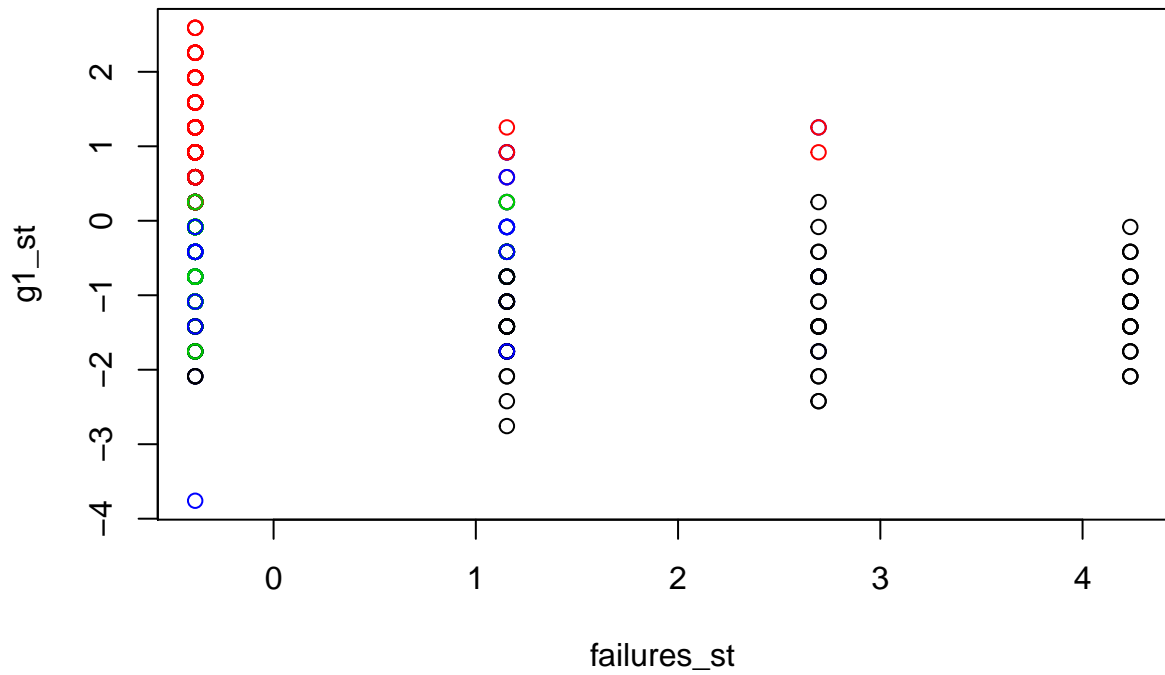
```
plot(data_cluster_result[c("failures", "g2"), col=data_cluster_result$cluster)
```



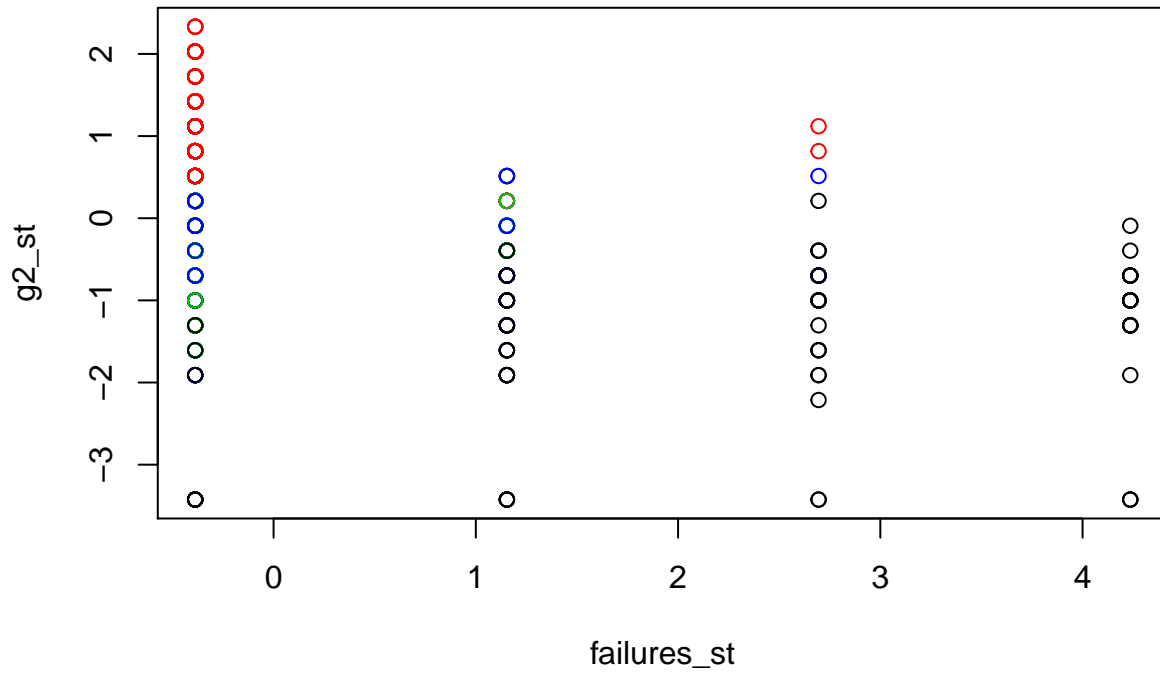
```
plot(data_cluster_result[c("failures", "g3"), col=data_cluster_result$cluster)
```



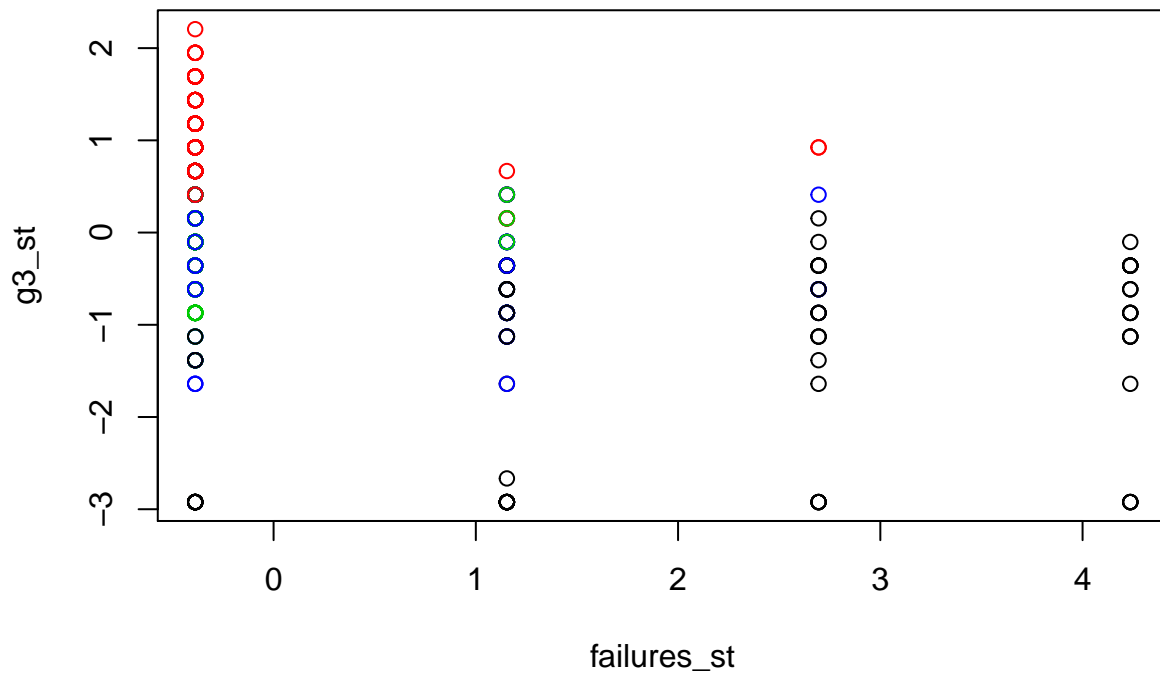
```
plot(data_cluster_result[c("failures_st", "g1_st")], col=data_cluster_result$cluster)
```



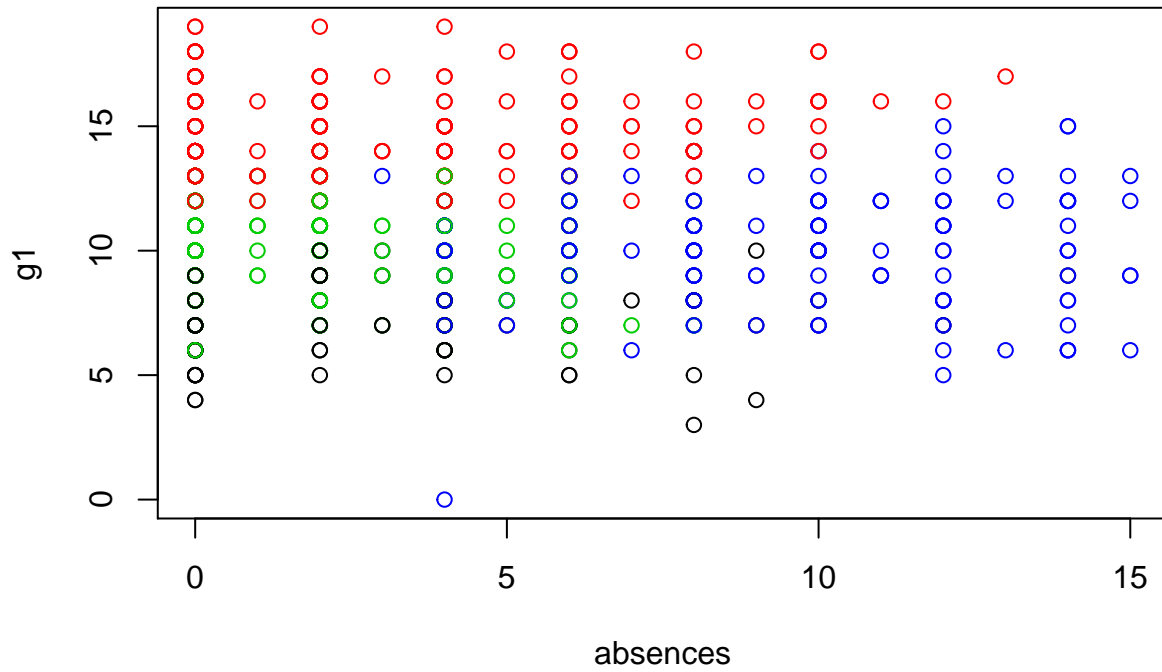
```
plot(data_cluster_result[c("failures_st", "g2_st")], col=data_cluster_result$cluster)
```



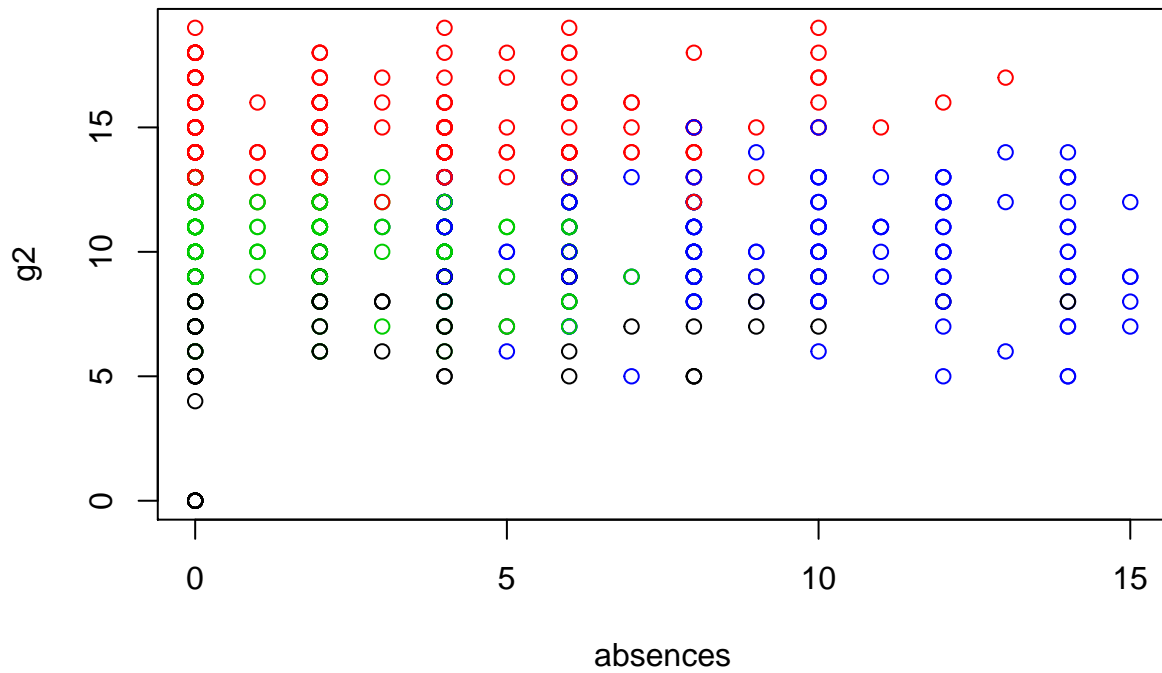
```
plot(data_cluster_result[c("failures_st", "g3_st")], col=data_cluster_result$cluster)
```



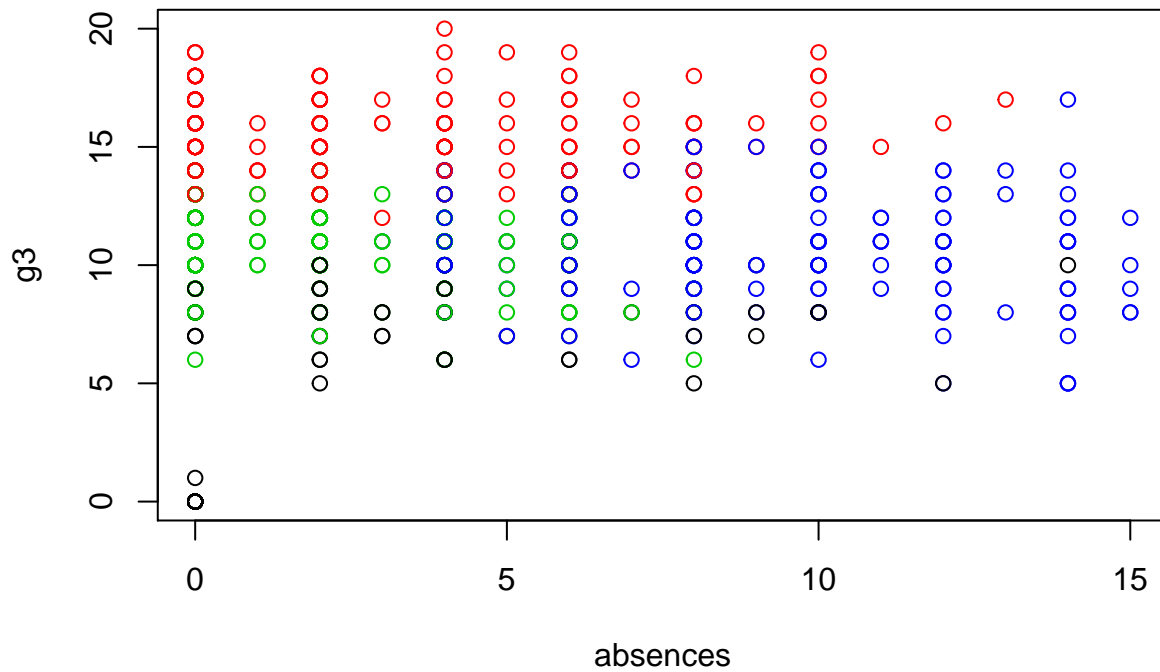
```
plot(data_cluster_result[c("absences", "g1"), col=data_cluster_result$cluster)
```



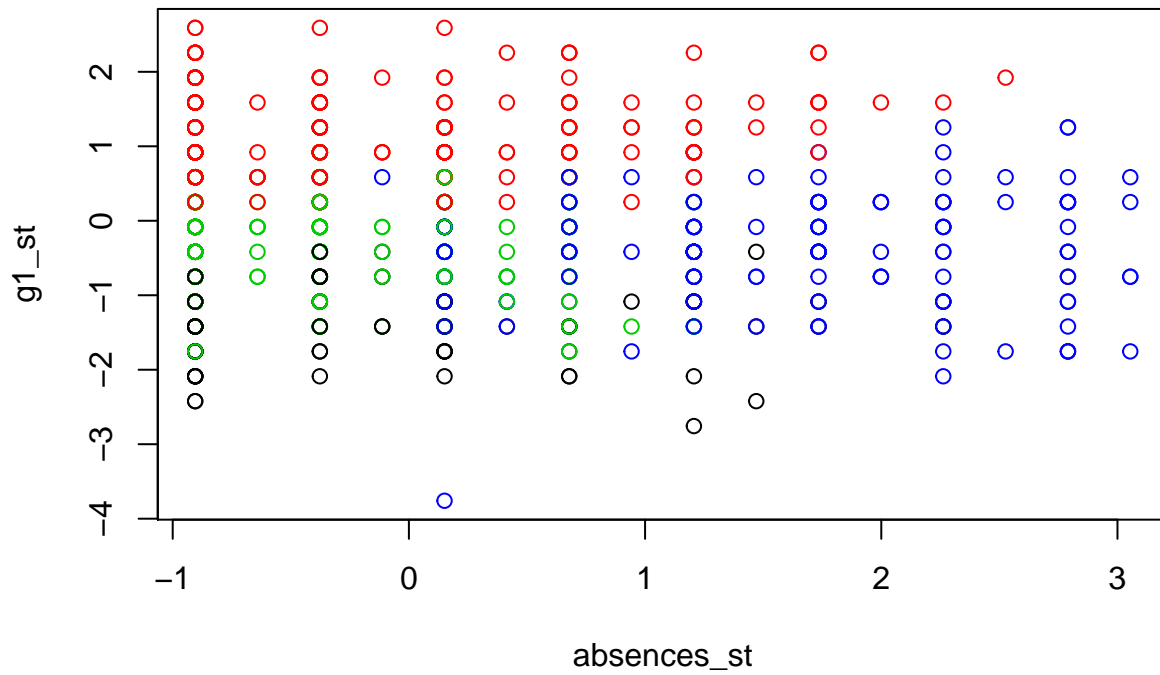
```
plot(data_cluster_result[c("absences", "g2"), col=data_cluster_result$cluster)
```



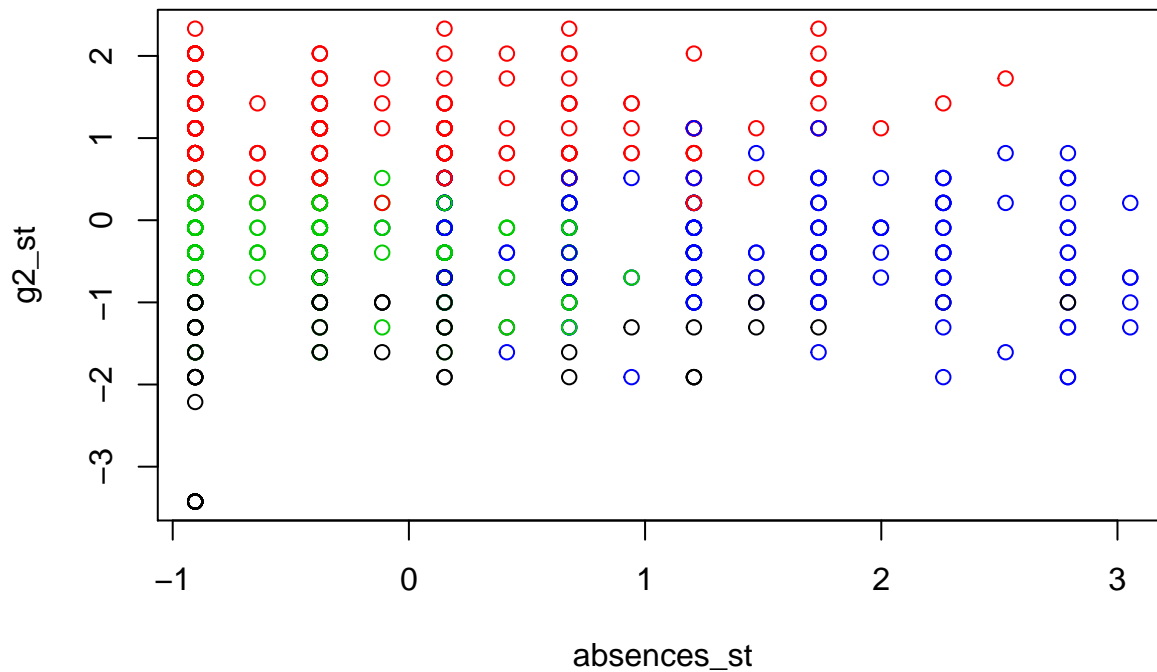
```
plot(data_cluster_result[c("absences", "g3"), col=data_cluster_result$cluster)
```



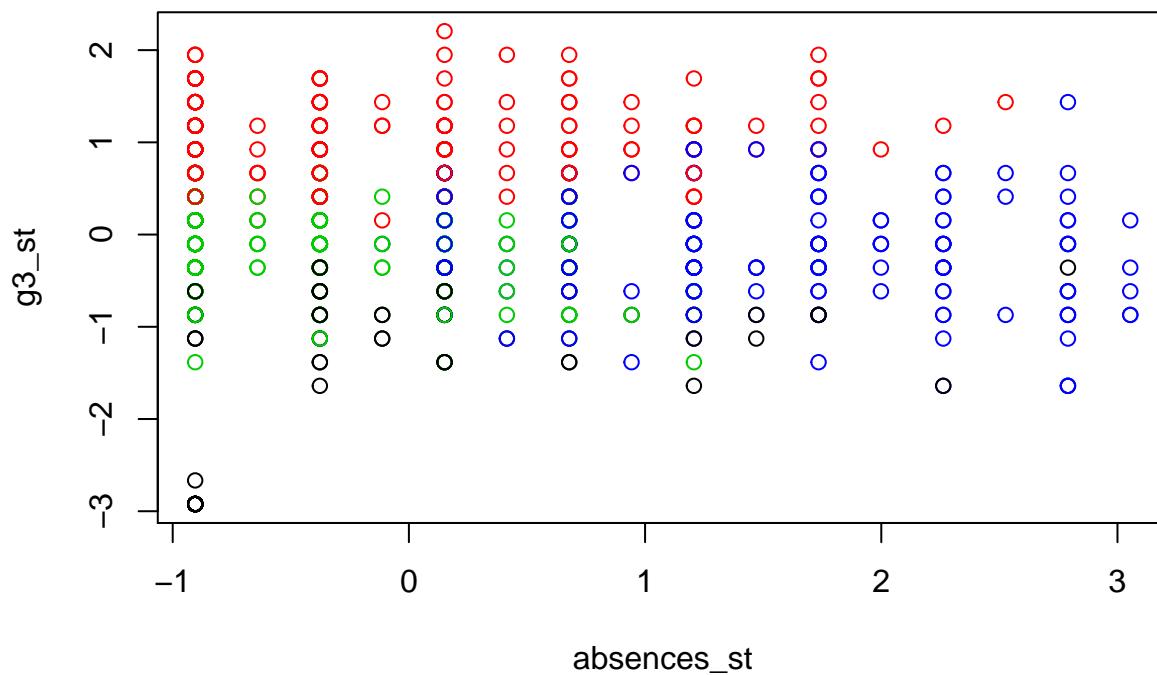
```
plot(data_cluster_result[c("absences_st", "g1_st")], col=data_cluster_result$cluster)
```



```
plot(data_cluster_result[c("absences_st", "g2_st")], col=data_cluster_result$cluster)
```



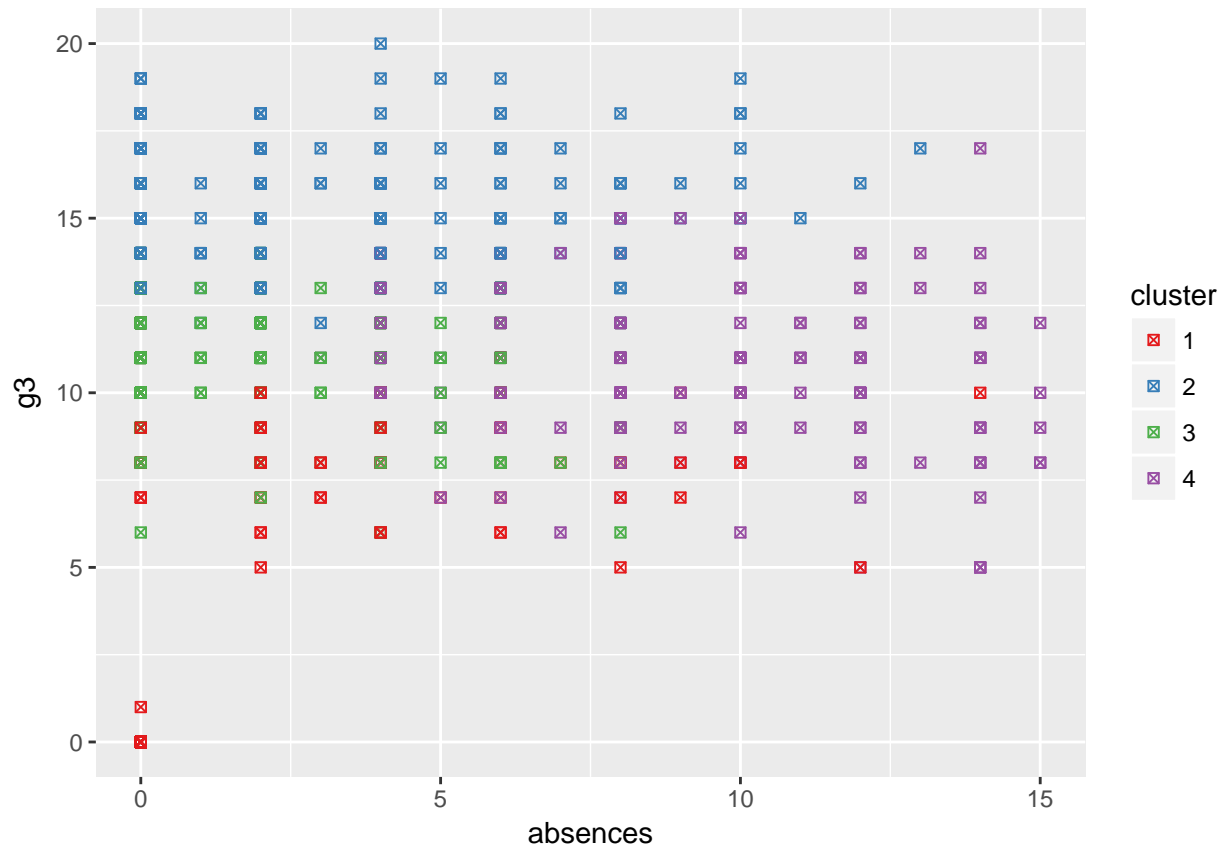
```
plot(data_cluster_result[c("absences_st", "g3_st")], col=data_cluster_result$cluster)
```



```
library("ggplot2")
# grafico_cluster<-
# ggplot(data_cluster_result, aes(x=absences, y=g3, color=cluster)) +
# geom_point(aes(size=g3))
# grafico_cluster

# ggplot(data_cluster_result, aes(x=absences, y=g3, color=cluster)) + geom_point(shape=7)+scale_color_b
```

```
grafico_cluster<-ggplot(data_cluster_result, aes(x=absences, y=g3, color=cluster)) + geom_point(shape=7)
grafico_cluster
```



```
rm(i)
rm(clus2_k3)
save.image("mywspace_TFM.RData")
```

*# los centroides obtenidos no se pueden intepretar ya que los datos estan normalizados*  
clus2\_k4\$centers

```
##      age  failures  absences      g1      g2      g3
## 1  0.61781001  1.767628615 -0.3089366 -1.2447718 -1.4555486 -1.6605118
## 2 -0.08973206 -0.343459054 -0.2660276  1.1078193  1.0426241  0.9529763
## 3 -0.42108419 -0.310166004 -0.3525451 -0.3243741 -0.2315884 -0.1413871
## 4  0.63685887 -0.006868061  1.5078740 -0.3877121 -0.3149020 -0.1955571
```

```
library(sqldf)
```

```
## Loading required package: gsubfn
## Loading required package: proto
## Loading required package: RSQLite
```

```
head(data_cluster_result)
```

```
##  age failures absences g1 g2 g3  age_st failures_st absences_st
## 1  18      0         6  5  6  6  1.0640333 -0.3876334  0.6787848
## 2  17      0         4  5  5  6  0.2480821 -0.3876334  0.1509003
## 3  15      3        10  7  8 10 -1.3838203  4.2359454  1.7345538
```



```
## 4 15      0      2 15 14 15 -1.3838203 -0.3876334 -0.3769842
## 5 16      0      4 6 10 10 -0.5678691 -0.3876334 0.1509003
## 6 16      0     10 15 15 15 -0.5678691 -0.3876334 1.7345538
##      g1_st      g2_st      g3_st cluster alcohol
## 1 -2.088010 -1.6070407 -1.3840741      1      1.00
## 2 -2.088010 -1.9100250 -1.3840741      1      1.00
## 3 -1.419685 -1.0010722 -0.3584503      1      2.29
## 4  1.253616  0.8168333  0.9235794      2      1.00
## 5 -1.753848 -0.3951037 -0.3584503      3      1.29
## 6  1.253616  1.1198176  0.9235794      2      1.29
```

```
sql1 <- sqldf('
select
  cluster
  , avg(age) as age_center
  , avg(failures) as failures_center
  , avg(absences) as absences_center
  , avg(g1) as g1_center
  , avg(g2) as g2_center
  , avg(g3) as g3_center
  , count(*) as n
from data_cluster_result
group by cluster
;')
```

```
## Loading required package: tcltk
```

```
## Warning: Quoted identifiers should have class SQL, use DBI::SQL() if the
## caller performs the quoting.
```

```
nrow(data_cluster_result)
```

```
## [1] 990
```

```
sum(sql1$n)
```

```
## [1] 990
```

```
as.data.frame.matrix(sql1)
```

```
##   cluster age_center failures_center absences_center g1_center g2_center
## 1      1  17.45312      1.39843750      2.257812  7.523438  6.50000
## 2      2  16.58599      0.02866242      2.420382 14.563694 14.74522
## 3      3  16.17989      0.05026455      2.092593 10.277778 10.53968
## 4      4  17.47647      0.24705882      9.141176 10.088235 10.26471
##   g3_center  n
## 1  4.921875 128
## 2 15.114650 314
## 3 10.846561 378
## 4 10.635294 170
```

```
sql2 <- sqldf('
select
  cluster,
  avg(alcohol) alcohol_avg,
  count(*) as n
from data_cluster_result
group by cluster
```

```
;)

sql2

## cluster alcohol_avg n
## 1 1 1.846797 128
## 2 2 1.470191 314
## 3 3 1.698175 378
## 4 4 2.018882 170

tapply ( data_cluster_result$alcohol,data_cluster_result$cluster,mean )

## 1 2 3 4
## 1.846797 1.470191 1.698175 2.018882

save.image("myspace_TFM.RData")
```

## Reglas de asociación

```
library(arules)

## Loading required package: Matrix
##
## Attaching package: 'arules'
## The following objects are masked from 'package:base':
##
## abbreviate, write

# seleccionar las variables que se pueden usar para la búsqueda de asociaciones
variables_aso <- c("school", "sex", "address", "famsize", "pstatus", "medu", "fedu", "mjob", "fjob", "r

data_aso <- data[,variables_aso]

ap_all<-apriori(data_aso)

## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
## 0.8 0.1 1 none FALSE TRUE 5 0.1 1
## maxlen target ext
## 10 rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE
##
## Absolute minimum support count: 104
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[93 item(s), 1044 transaction(s)] done [0.00s].
```

```

## sorting and recoding items ... [74 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 7 8 9 10

## Warning in apriori(data_aso): Mining stopped (maxlen reached). Only
## patterns up to a length of 10 returned!

## done [1.39s].
## writing ... [456831 rule(s)] done [20.50s].
## creating S4 object ... done [0.51s].

# inspect(ap_all)
# sin ninguna restricción aparecen 456.831 reglas

rm(ap1)
ap1 <- apriori(data_aso, parameter=list ( supp=0.30,conf= 0.95,target="rules"))

```

```

## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##      0.95    0.1    1 none FALSE                TRUE      5    0.3    1
## maxlen target  ext
##      10  rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 313
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[93 item(s), 1044 transaction(s)] done [0.00s].
## sorting and recoding items ... [33 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 7 done [0.05s].
## writing ... [537 rule(s)] done [0.02s].
## creating S4 object ... done [0.00s].

```

```
inspect(ap1)
```

	lhs	rhs	support	confidence	lift
## [1]	{walc=very low}	=> {dalc=very low}	0.3745211	0.9824121	1.410782
## [2]	{walc=very low, bing3=pass}	=> {dalc=very low}	0.3026820	0.9875000	1.418088
## [3]	{paid=no, walc=very low}	=> {dalc=very low}	0.3045977	0.9784615	1.405108
## [4]	{nursery=yes, walc=very low}	=> {dalc=very low}	0.3122605	0.9848943	1.414346
## [5]	{pstatus=living together, walc=very low}	=> {dalc=very low}	0.3189655	0.9823009	1.410622
## [6]	{schoolsup=no, walc=very low}	=> {dalc=very low}	0.3084291	0.9787234	1.405485
## [7]	{higher=yes, walc=very low}	=> {dalc=very low}	0.3486590	0.9864499	1.416580
## [8]	{studytime=2 to 5 hours,				

##	dalc=very low}	=> {higher=yes}	0.3342912	0.9509537	1.039577
## [9]	{guardian=mother,				
##	studytime=2 to 5 hours}	=> {higher=yes}	0.3247126	0.9603399	1.049838
## [10]	{address=urban,				
##	studytime=2 to 5 hours}	=> {higher=yes}	0.3591954	0.9640103	1.053850
## [11]	{school=Gabriel Pereira,				
##	studytime=2 to 5 hours}	=> {higher=yes}	0.3477011	0.9502618	1.038820
## [12]	{studytime=2 to 5 hours,				
##	bing3=pass}	=> {higher=yes}	0.3601533	0.9591837	1.048574
## [13]	{studytime=2 to 5 hours,				
##	internet=yes}	=> {higher=yes}	0.3591954	0.9517766	1.040476
## [14]	{studytime=2 to 5 hours,				
##	nursery=yes}	=> {higher=yes}	0.3678161	0.9504950	1.039075
## [15]	{school=Gabriel Pereira,				
##	activities=yes}	=> {higher=yes}	0.3716475	0.9603960	1.049899
## [16]	{activities=yes,				
##	bing3=pass}	=> {higher=yes}	0.3716475	0.9509804	1.039606
## [17]	{sex=female,				
##	traveltime=<15 min.}	=> {higher=yes}	0.3285441	0.9607843	1.050323
## [18]	{sex=female,				
##	address=urban}	=> {higher=yes}	0.3888889	0.9508197	1.039430
## [19]	{school=Gabriel Pereira,				
##	sex=female}	=> {higher=yes}	0.3869732	0.9619048	1.051548
## [20]	{sex=female,				
##	bing3=pass}	=> {higher=yes}	0.4243295	0.9506438	1.039238
## [21]	{sex=female,				
##	internet=yes}	=> {higher=yes}	0.4147510	0.9516484	1.040336
## [22]	{traveltime=<15 min.,				
##	famsup=yes}	=> {higher=yes}	0.3611111	0.9520202	1.040743
## [23]	{traveltime=<15 min.,				
##	dalc=very low}	=> {higher=yes}	0.4090038	0.9595506	1.048975
## [24]	{school=Gabriel Pereira,				
##	traveltime=<15 min.}	=> {higher=yes}	0.4808429	0.9543726	1.043314
## [25]	{traveltime=<15 min.,				
##	bing3=pass}	=> {higher=yes}	0.4588123	0.9735772	1.064309
## [26]	{famsup=yes,				
##	romantic=no}	=> {higher=yes}	0.3802682	0.9566265	1.045778
## [27]	{guardian=mother,				
##	famsup=yes}	=> {higher=yes}	0.4003831	0.9500000	1.038534
## [28]	{school=Gabriel Pereira,				
##	famsup=yes}	=> {higher=yes}	0.4521073	0.9574037	1.046628
## [29]	{famsup=yes,				
##	bing3=pass}	=> {higher=yes}	0.4540230	0.9518072	1.040510
## [30]	{romantic=no,				
##	dalc=very low}	=> {higher=yes}	0.4300766	0.9532909	1.042132
## [31]	{romantic=no,				
##	bing3=pass}	=> {higher=yes}	0.4961686	0.9574861	1.046718
## [32]	{school=Gabriel Pereira,				
##	dalc=very low}	=> {higher=yes}	0.5019157	0.9544627	1.043413
## [33]	{dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.5363985	0.9556314	1.044690
## [34]	{internet=yes,				
##	dalc=very low}	=> {higher=yes}	0.5201149	0.9526316	1.041411
## [35]	{guardian=mother,				

##	bing3=pass}	=> {higher=yes}	0.5191571	0.9609929	1.050551
## [36]	{address=urban, bing3=pass}	=> {higher=yes}	0.5565134	0.9587459	1.048095
## [37]	{school=Gabriel Pereira, bing3=pass}	=> {higher=yes}	0.5785441	0.9633174	1.053093
## [38]	{internet=yes, bing3=pass}	=> {higher=yes}	0.5986590	0.9541985	1.043124
## [39]	{nursery=yes, bing3=pass}	=> {higher=yes}	0.5938697	0.9538462	1.042739
## [40]	{school=Gabriel Pereira, address=urban, studytime=2 to 5 hours}	=> {higher=yes}	0.3036398	0.9694190	1.059763
## [41]	{address=urban, studytime=2 to 5 hours, internet=yes}	=> {higher=yes}	0.3017241	0.9692308	1.059557
## [42]	{address=urban, pstatus=living together, studytime=2 to 5 hours}	=> {higher=yes}	0.3026820	0.9575758	1.046816
## [43]	{address=urban, studytime=2 to 5 hours, schoolsup=no}	=> {higher=yes}	0.3045977	0.9578313	1.047095
## [44]	{pstatus=living together, studytime=2 to 5 hours, bing3=pass}	=> {higher=yes}	0.3055556	0.9550898	1.044098
## [45]	{studytime=2 to 5 hours, schoolsup=no, bing3=pass}	=> {higher=yes}	0.3141762	0.9562682	1.045386
## [46]	{school=Gabriel Pereira, pstatus=living together, famrel=good}	=> {higher=yes}	0.3141762	0.9507246	1.039326
## [47]	{pstatus=living together, activities=yes, dalc=very low}	=> {higher=yes}	0.3036398	0.9519520	1.040668
## [48]	{school=Gabriel Pereira, activities=yes, bing3=pass}	=> {higher=yes}	0.3103448	0.9729730	1.063648
## [49]	{school=Gabriel Pereira, activities=yes, internet=yes}	=> {higher=yes}	0.3227969	0.9683908	1.058639
## [50]	{school=Gabriel Pereira, activities=yes, nursery=yes}	=> {higher=yes}	0.3055556	0.9666667	1.056754
## [51]	{school=Gabriel Pereira, pstatus=living together, activities=yes}	=> {higher=yes}	0.3390805	0.9593496	1.048755
## [52]	{school=Gabriel Pereira, schoolsup=no, activities=yes}	=> {higher=yes}	0.3189655	0.9596542	1.049088
## [53]	{activities=yes, internet=yes, bing3=pass}	=> {higher=yes}	0.3141762	0.9618768	1.051518
## [54]	{activities=yes, nursery=yes, bing3=pass}	=> {higher=yes}	0.3026820	0.9546828	1.043653

```

## [55] {pstatus=living together,
##       activities=yes,
##       bing3=pass}          => {higher=yes}          0.3381226  0.9514825  1.040155
## [56] {schoolsup=no,
##       activities=yes,
##       bing3=pass}          => {higher=yes}          0.3314176  0.9505495  1.039135
## [57] {school=Gabriel Pereira,
##       fjob=other,
##       dalc=very low}      => {higher=yes}          0.3036398  0.9606061  1.050129
## [58] {fjob=other,
##       dalc=very low,
##       bing3=pass}          => {higher=yes}          0.3103448  0.9557522  1.044822
## [59] {school=Gabriel Pereira,
##       fjob=other,
##       guardian=mother}    => {higher=yes}          0.3036398  0.9519520  1.040668
## [60] {fjob=other,
##       guardian=mother,
##       bing3=pass}          => {higher=yes}          0.3132184  0.9589443  1.048312
## [61] {school=Gabriel Pereira,
##       fjob=other,
##       bing3=pass}          => {higher=yes}          0.3381226  0.9566396  1.045792
## [62] {sex=female,
##       pstatus=living together,
##       famsup=yes}          => {higher=yes}          0.3227969  0.9546742  1.043644
## [63] {school=Gabriel Pereira,
##       sex=female,
##       dalc=very low}      => {higher=yes}          0.3065134  0.9609610  1.050516
## [64] {sex=female,
##       guardian=mother,
##       bing3=pass}          => {higher=yes}          0.3045977  0.9665653  1.056643
## [65] {sex=female,
##       guardian=mother,
##       nursery=yes}         => {higher=yes}          0.3208812  0.9517045  1.040397
## [66] {sex=female,
##       pstatus=living together,
##       guardian=mother}    => {higher=yes}          0.3285441  0.9581006  1.047390
## [67] {school=Gabriel Pereira,
##       sex=female,
##       address=urban}       => {higher=yes}          0.3237548  0.9712644  1.061780
## [68] {sex=female,
##       address=urban,
##       bing3=pass}          => {higher=yes}          0.3141762  0.9590643  1.048443
## [69] {sex=female,
##       address=urban,
##       internet=yes}        => {higher=yes}          0.3180077  0.9707602  1.061229
## [70] {sex=female,
##       address=urban,
##       nursery=yes}         => {higher=yes}          0.3180077  0.9512894  1.039944
## [71] {sex=female,
##       address=urban,
##       pstatus=living together} => {higher=yes}          0.3352490  0.9589041  1.048268
## [72] {school=Gabriel Pereira,
##       sex=female,
##       bing3=pass}          => {higher=yes}          0.3189655  0.9680233  1.058237

```

## [73]	{school=Gabriel Pereira, sex=female, internet=yes}	=> {higher=yes}	0.3208812	0.9626437	1.052356
## [74]	{school=Gabriel Pereira, sex=female, nursery=yes}	=> {higher=yes}	0.3141762	0.9675516	1.057721
## [75]	{school=Gabriel Pereira, sex=female, pstatus=living together}	=> {higher=yes}	0.3381226	0.9697802	1.060158
## [76]	{school=Gabriel Pereira, sex=female, schoolsup=no}	=> {higher=yes}	0.3141762	0.9562682	1.045386
## [77]	{sex=female, internet=yes, bing3=pass}	=> {higher=yes}	0.3342912	0.9587912	1.048145
## [78]	{sex=female, nursery=yes, bing3=pass}	=> {higher=yes}	0.3448276	0.9549072	1.043899
## [79]	{sex=female, pstatus=living together, bing3=pass}	=> {higher=yes}	0.3745211	0.9583333	1.047644
## [80]	{sex=female, nursery=yes, internet=yes}	=> {higher=yes}	0.3371648	0.9539295	1.042830
## [81]	{sex=female, pstatus=living together, internet=yes}	=> {higher=yes}	0.3687739	0.9577114	1.046964
## [82]	{address=urban, traveltime=<15 min., famsup=yes}	=> {higher=yes}	0.3093870	0.9500000	1.038534
## [83]	{school=Gabriel Pereira, traveltime=<15 min., famsup=yes}	=> {higher=yes}	0.3160920	0.9677419	1.057929
## [84]	{traveltime=<15 min., famsup=yes, internet=yes}	=> {higher=yes}	0.3093870	0.9584570	1.047779
## [85]	{traveltime=<15 min., famsup=yes, nursery=yes}	=> {higher=yes}	0.3007663	0.9602446	1.049733
## [86]	{pstatus=living together, traveltime=<15 min., famsup=yes}	=> {higher=yes}	0.3170498	0.9538905	1.042787
## [87]	{school=Gabriel Pereira, traveltime=<15 min., romantic=no}	=> {higher=yes}	0.3189655	0.9596542	1.049088
## [88]	{traveltime=<15 min., romantic=no, bing3=pass}	=> {higher=yes}	0.3007663	0.9691358	1.059453
## [89]	{traveltime=<15 min., internet=yes, romantic=no}	=> {higher=yes}	0.3113027	0.9502924	1.038854
## [90]	{address=urban, traveltime=<15 min., dalc=very low}	=> {higher=yes}	0.3544061	0.9560724	1.045172

## [91]	{school=Gabriel Pereira, traveltime=<15 min., dalc=very low}	=> {higher=yes}	0.3524904	0.9735450	1.064273
## [92]	{traveltime=<15 min., dalc=very low, bing3=pass}	=> {higher=yes}	0.3429119	0.9754768	1.066385
## [93]	{traveltime=<15 min., internet=yes, dalc=very low}	=> {higher=yes}	0.3477011	0.9680000	1.058212
## [94]	{traveltime=<15 min., nursery=yes, dalc=very low}	=> {higher=yes}	0.3419540	0.9622642	1.051941
## [95]	{pstatus=living together, traveltime=<15 min., dalc=very low}	=> {higher=yes}	0.3534483	0.9609375	1.050491
## [96]	{traveltime=<15 min., schoolsup=no, dalc=very low}	=> {higher=yes}	0.3524904	0.9533679	1.042216
## [97]	{school=Gabriel Pereira, guardian=mother, traveltime=<15 min.}	=> {higher=yes}	0.3524904	0.9633508	1.053129
## [98]	{guardian=mother, traveltime=<15 min., bing3=pass}	=> {higher=yes}	0.3285441	0.9800000	1.071330
## [99]	{famsize=greater than 3, traveltime=<15 min., bing3=pass}	=> {higher=yes}	0.3237548	0.9684814	1.058738
## [100]	{school=Gabriel Pereira, address=urban, traveltime=<15 min.}	=> {higher=yes}	0.4214559	0.9503240	1.038888
## [101]	{address=urban, traveltime=<15 min., bing3=pass}	=> {higher=yes}	0.3898467	0.9713604	1.061885
## [102]	{school=Gabriel Pereira, traveltime=<15 min., bing3=pass}	=> {higher=yes}	0.4080460	0.9793103	1.070576
## [103]	{school=Gabriel Pereira, traveltime=<15 min., internet=yes}	=> {higher=yes}	0.4147510	0.9516484	1.040336
## [104]	{school=Gabriel Pereira, traveltime=<15 min., nursery=yes}	=> {higher=yes}	0.3908046	0.9645390	1.054428
## [105]	{school=Gabriel Pereira, pstatus=living together, traveltime=<15 min.}	=> {higher=yes}	0.4204981	0.9502165	1.038771
## [106]	{traveltime=<15 min., paid=no, bing3=pass}	=> {higher=yes}	0.3572797	0.9663212	1.056376
## [107]	{traveltime=<15 min., internet=yes, bing3=pass}	=> {higher=yes}	0.3850575	0.9733656	1.064077
## [108]	{traveltime=<15 min., nursery=yes, bing3=pass}	=> {higher=yes}	0.3745211	0.9750623	1.065932



```

## [109] {pstatus=living together,
##       traveltime=<15 min.,
##       bing3=pass} => {higher=yes} 0.3965517 0.9718310 1.062400
## [110] {traveltime=<15 min.,
##       schoolsup=no,
##       bing3=pass} => {higher=yes} 0.4032567 0.9722864 1.062897
## [111] {school=Gabriel Pereira,
##       famsup=yes,
##       romantic=no} => {higher=yes} 0.3045977 0.9724771 1.063106
## [112] {famsup=yes,
##       romantic=no,
##       bing3=pass} => {higher=yes} 0.3093870 0.9670659 1.057190
## [113] {famsup=yes,
##       internet=yes,
##       romantic=no} => {higher=yes} 0.3084291 0.9728097 1.063469
## [114] {famsup=yes,
##       nursery=yes,
##       romantic=no} => {higher=yes} 0.3189655 0.9680233 1.058237
## [115] {pstatus=living together,
##       famsup=yes,
##       romantic=no} => {higher=yes} 0.3400383 0.9620596 1.051718
## [116] {schoolsup=no,
##       famsup=yes,
##       romantic=no} => {higher=yes} 0.3208812 0.9571429 1.046343
## [117] {school=Gabriel Pereira,
##       famsup=yes,
##       dalc=very low} => {higher=yes} 0.3285441 0.9554318 1.044472
## [118] {famsup=yes,
##       dalc=very low,
##       bing3=pass} => {higher=yes} 0.3295019 0.9529086 1.041714
## [119] {famsup=yes,
##       internet=yes,
##       dalc=very low} => {higher=yes} 0.3352490 0.9589041 1.048268
## [120] {school=Gabriel Pereira,
##       guardian=mother,
##       famsup=yes} => {higher=yes} 0.3208812 0.9598854 1.049341
## [121] {guardian=mother,
##       famsup=yes,
##       bing3=pass} => {higher=yes} 0.3160920 0.9677419 1.057929
## [122] {guardian=mother,
##       famsup=yes,
##       internet=yes} => {higher=yes} 0.3285441 0.9581006 1.047390
## [123] {guardian=mother,
##       famsup=yes,
##       nursery=yes} => {higher=yes} 0.3362069 0.9512195 1.039867
## [124] {pstatus=living together,
##       guardian=mother,
##       famsup=yes} => {higher=yes} 0.3496169 0.9530026 1.041816
## [125] {school=Gabriel Pereira,
##       famsize=greater than 3,
##       famsup=yes} => {higher=yes} 0.3314176 0.9531680 1.041997
## [126] {famsize=greater than 3,
##       famsup=yes,
##       bing3=pass} => {higher=yes} 0.3285441 0.9527778 1.041571

```

## [127]	{school=Gabriel Pereira, address=urban, famsup=yes}	=> {higher=yes}	0.3687739	0.9600998	1.049575
## [128]	{address=urban, famsup=yes, bing3=pass}	=> {higher=yes}	0.3400383	0.9646739	1.054575
## [129]	{address=urban, famsup=yes, internet=yes}	=> {higher=yes}	0.3620690	0.9569620	1.046145
## [130]	{address=urban, famsup=yes, nursery=yes}	=> {higher=yes}	0.3496169	0.9530026	1.041816
## [131]	{school=Gabriel Pereira, famsup=yes, bing3=pass}	=> {higher=yes}	0.3697318	0.9674185	1.057576
## [132]	{school=Gabriel Pereira, famsup=yes, internet=yes}	=> {higher=yes}	0.3917625	0.9600939	1.049569
## [133]	{school=Gabriel Pereira, famsup=yes, nursery=yes}	=> {higher=yes}	0.3716475	0.9651741	1.055122
## [134]	{school=Gabriel Pereira, pstatus=living together, famsup=yes}	=> {higher=yes}	0.3975096	0.9562212	1.045335
## [135]	{school=Gabriel Pereira, schoolsup=no, famsup=yes}	=> {higher=yes}	0.3802682	0.9566265	1.045778
## [136]	{famsup=yes, internet=yes, bing3=pass}	=> {higher=yes}	0.3764368	0.9562044	1.045317
## [137]	{famsup=yes, nursery=yes, bing3=pass}	=> {higher=yes}	0.3659004	0.9526185	1.041397
## [138]	{pstatus=living together, famsup=yes, bing3=pass}	=> {higher=yes}	0.4042146	0.9525959	1.041372
## [139]	{schoolsup=no, famsup=yes, bing3=pass}	=> {higher=yes}	0.3965517	0.9517241	1.040419
## [140]	{guardian=mother, romantic=no, dalc=very low}	=> {higher=yes}	0.3170498	0.9538905	1.042787
## [141]	{famsize=greater than 3, romantic=no, dalc=very low}	=> {higher=yes}	0.3189655	0.9541547	1.043076
## [142]	{address=urban, romantic=no, dalc=very low}	=> {higher=yes}	0.3275862	0.9500000	1.038534
## [143]	{school=Gabriel Pereira, romantic=no, dalc=very low}	=> {higher=yes}	0.3409962	0.9647696	1.054680
## [144]	{romantic=no, dalc=very low, bing3=pass}	=> {higher=yes}	0.3591954	0.9664948	1.056566

## [145]	{internet=yes, romantic=no, dalc=very low}	=> {higher=yes}	0.3362069	0.9696133	1.059975
## [146]	{nursery=yes, romantic=no, dalc=very low}	=> {higher=yes}	0.3524904	0.9608355	1.050379
## [147]	{pstatus=living together, romantic=no, dalc=very low}	=> {higher=yes}	0.3879310	0.9551887	1.044206
## [148]	{schoolsup=no, romantic=no, dalc=very low}	=> {higher=yes}	0.3639847	0.9500000	1.038534
## [149]	{school=Gabriel Pereira, guardian=mother, romantic=no}	=> {higher=yes}	0.3429119	0.9623656	1.052052
## [150]	{guardian=mother, romantic=no, bing3=pass}	=> {higher=yes}	0.3553640	0.9661458	1.056185
## [151]	{guardian=mother, internet=yes, romantic=no}	=> {higher=yes}	0.3371648	0.9565217	1.045664
## [152]	{guardian=mother, nursery=yes, romantic=no}	=> {higher=yes}	0.3553640	0.9586563	1.047997
## [153]	{famsize=greater than 3, romantic=no, bing3=pass}	=> {higher=yes}	0.3429119	0.9521277	1.040860
## [154]	{famsize=greater than 3, internet=yes, romantic=no}	=> {pstatus=living together}	0.3342912	0.9587912	1.084483
## [155]	{famsize=greater than 3, schoolsup=no, romantic=no}	=> {pstatus=living together}	0.3678161	0.9576060	1.083143
## [156]	{address=urban, romantic=no, bing3=pass}	=> {higher=yes}	0.3678161	0.9576060	1.046849
## [157]	{school=Gabriel Pereira, romantic=no, bing3=pass}	=> {higher=yes}	0.3965517	0.9650350	1.054970
## [158]	{school=Gabriel Pereira, internet=yes, romantic=no}	=> {higher=yes}	0.3860153	0.9504717	1.039050
## [159]	{school=Gabriel Pereira, nursery=yes, romantic=no}	=> {higher=yes}	0.3812261	0.9636804	1.053489
## [160]	{internet=yes, romantic=no, bing3=pass}	=> {higher=yes}	0.3965517	0.9650350	1.054970
## [161]	{nursery=yes, romantic=no, bing3=pass}	=> {higher=yes}	0.4003831	0.9675926	1.057766
## [162]	{pstatus=living together, romantic=no, bing3=pass}	=> {higher=yes}	0.4425287	0.9565217	1.045664

## [163]	{schoolsup=no, romantic=no, bing3=pass}	=> {higher=yes}	0.4348659	0.9578059	1.047067
## [164]	{school=Gabriel Pereira, guardian=mother, dalc=very low}	=> {higher=yes}	0.3687739	0.9529703	1.041781
## [165]	{guardian=mother, dalc=very low, bing3=pass}	=> {higher=yes}	0.3831418	0.9592326	1.048627
## [166]	{guardian=mother, internet=yes, dalc=very low}	=> {higher=yes}	0.3726054	0.9534314	1.042285
## [167]	{school=Gabriel Pereira, famsize=greater than 3, dalc=very low}	=> {higher=yes}	0.3745211	0.9513382	1.039997
## [168]	{famsize=greater than 3, dalc=very low, bing3=pass}	=> {higher=yes}	0.3898467	0.9509346	1.039556
## [169]	{school=Gabriel Pereira, address=urban, dalc=very low}	=> {higher=yes}	0.4204981	0.9627193	1.052439
## [170]	{address=urban, dalc=very low, bing3=pass}	=> {higher=yes}	0.4204981	0.9606127	1.050136
## [171]	{address=urban, internet=yes, dalc=very low}	=> {higher=yes}	0.4204981	0.9606127	1.050136
## [172]	{address=urban, nursery=yes, dalc=very low}	=> {higher=yes}	0.4070881	0.9529148	1.041720
## [173]	{school=Gabriel Pereira, dalc=very low, bing3=pass}	=> {higher=yes}	0.4252874	0.9673203	1.057468
## [174]	{school=Gabriel Pereira, internet=yes, dalc=very low}	=> {higher=yes}	0.4233716	0.9608696	1.050417
## [175]	{school=Gabriel Pereira, nursery=yes, dalc=very low}	=> {higher=yes}	0.4118774	0.9641256	1.053976
## [176]	{school=Gabriel Pereira, pstatus=living together, dalc=very low}	=> {higher=yes}	0.4406130	0.9563410	1.045466
## [177]	{school=Gabriel Pereira, schoolsup=no, dalc=very low}	=> {higher=yes}	0.4272031	0.9509595	1.039583
## [178]	{internet=yes, dalc=very low, bing3=pass}	=> {higher=yes}	0.4310345	0.9635974	1.053399
## [179]	{nursery=yes, dalc=very low, bing3=pass}	=> {higher=yes}	0.4425287	0.9645094	1.054396
## [180]	{pstatus=living together, dalc=very low, bing3=pass}	=> {higher=yes}	0.4693487	0.9589041	1.048268

## [181]	{schoolsup=no, dalc=very low, bing3=pass}	=> {higher=yes}	0.4741379	0.9537572	1.042641
## [182]	{nursery=yes, internet=yes, dalc=very low}	=> {higher=yes}	0.4233716	0.9546436	1.043610
## [183]	{pstatus=living together, internet=yes, dalc=very low}	=> {higher=yes}	0.4616858	0.9544554	1.043405
## [184]	{schoolsup=no, internet=yes, dalc=very low}	=> {higher=yes}	0.4597701	0.9504950	1.039075
## [185]	{famsize=greater than 3, guardian=mother, bing3=pass}	=> {higher=yes}	0.3544061	0.9585492	1.047880
## [186]	{address=urban, guardian=mother, bing3=pass}	=> {higher=yes}	0.3898467	0.9667458	1.056840
## [187]	{school=Gabriel Pereira, guardian=mother, bing3=pass}	=> {higher=yes}	0.4157088	0.9730942	1.063780
## [188]	{school=Gabriel Pereira, guardian=mother, nursery=yes}	=> {higher=yes}	0.4157088	0.9559471	1.045035
## [189]	{guardian=mother, paid=no, bing3=pass}	=> {higher=yes}	0.4080460	0.9508929	1.039510
## [190]	{guardian=mother, internet=yes, bing3=pass}	=> {higher=yes}	0.4176245	0.9688889	1.059183
## [191]	{guardian=mother, nursery=yes, bing3=pass}	=> {higher=yes}	0.4319923	0.9719828	1.062565
## [192]	{pstatus=living together, guardian=mother, bing3=pass}	=> {higher=yes}	0.4454023	0.9627329	1.052454
## [193]	{guardian=mother, schoolsup=no, bing3=pass}	=> {higher=yes}	0.4645594	0.9603960	1.049899
## [194]	{address=urban, famsize=greater than 3, bing3=pass}	=> {higher=yes}	0.3754789	0.9537713	1.042657
## [195]	{school=Gabriel Pereira, famsize=greater than 3, bing3=pass}	=> {higher=yes}	0.4090038	0.9617117	1.051337
## [196]	{famsize=greater than 3, internet=yes, bing3=pass}	=> {higher=yes}	0.4118774	0.9513274	1.039985
## [197]	{school=Gabriel Pereira, address=urban, bing3=pass}	=> {higher=yes}	0.4789272	0.9671180	1.057247
## [198]	{school=Gabriel Pereira, address=urban, nursery=yes}	=> {higher=yes}	0.4626437	0.9526627	1.041445

## [199]	{address=urban, internet=yes, bing3=pass}	=> {higher=yes}	0.4798851	0.9653179	1.055279
## [200]	{address=urban, nursery=yes, bing3=pass}	=> {higher=yes}	0.4549808	0.9634888	1.053280
## [201]	{address=urban, pstatus=living together, bing3=pass}	=> {higher=yes}	0.4827586	0.9600000	1.049466
## [202]	{address=urban, schoolsup=no, bing3=pass}	=> {higher=yes}	0.4942529	0.9573284	1.046545
## [203]	{school=Gabriel Pereira, paid=no, bing3=pass}	=> {higher=yes}	0.4482759	0.9551020	1.044112
## [204]	{school=Gabriel Pereira, internet=yes, bing3=pass}	=> {higher=yes}	0.4913793	0.9642857	1.054151
## [205]	{school=Gabriel Pereira, nursery=yes, bing3=pass}	=> {higher=yes}	0.4712644	0.9704142	1.060851
## [206]	{school=Gabriel Pereira, pstatus=living together, bing3=pass}	=> {higher=yes}	0.5067050	0.9653285	1.055291
## [207]	{school=Gabriel Pereira, schoolsup=no, bing3=pass}	=> {higher=yes}	0.5057471	0.9635036	1.053296
## [208]	{nursery=yes, internet=yes, bing3=pass}	=> {higher=yes}	0.4808429	0.9543726	1.043314
## [209]	{pstatus=living together, internet=yes, bing3=pass}	=> {higher=yes}	0.5296935	0.9550950	1.044104
## [210]	{schoolsup=no, internet=yes, bing3=pass}	=> {higher=yes}	0.5373563	0.9540816	1.042996
## [211]	{pstatus=living together, nursery=yes, bing3=pass}	=> {higher=yes}	0.5134100	0.9503546	1.038922
## [212]	{schoolsup=no, nursery=yes, bing3=pass}	=> {higher=yes}	0.5277778	0.9532872	1.042128
## [213]	{pstatus=living together, schoolsup=no, activities=yes, bing3=pass}	=> {higher=yes}	0.3026820	0.9518072	1.040510
## [214]	{sex=female, pstatus=living together, paid=no, bing3=pass}	=> {higher=yes}	0.3007663	0.9515152	1.040190
## [215]	{sex=female, pstatus=living together, schoolsup=no, bing3=pass}	=> {higher=yes}	0.3199234	0.9542857	1.043219

```

## [216] {sex=female,
##       pstatus=living together,
##       schoolsup=no,
##       internet=yes}      => {higher=yes}      0.3160920  0.9537572  1.042641
## [217] {school=Gabriel Pereira,
##       address=urban,
##       traveltime=<15 min.,
##       dalc=very low}    => {higher=yes}    0.3160920  0.9705882  1.061041
## [218] {address=urban,
##       traveltime=<15 min.,
##       dalc=very low,
##       bing3=pass}      => {higher=yes}    0.3055556  0.9725610  1.063198
## [219] {address=urban,
##       traveltime=<15 min.,
##       internet=yes,
##       dalc=very low}   => {higher=yes}   0.3113027  0.9672619  1.057405
## [220] {address=urban,
##       pstatus=living together,
##       traveltime=<15 min.,
##       dalc=very low}   => {higher=yes}   0.3026820  0.9575758  1.046816
## [221] {address=urban,
##       traveltime=<15 min.,
##       schoolsup=no,
##       dalc=very low}   => {higher=yes}   0.3093870  0.9500000  1.038534
## [222] {school=Gabriel Pereira,
##       traveltime=<15 min.,
##       dalc=very low,
##       bing3=pass}      => {higher=yes}    0.3045977  0.9814815  1.072949
## [223] {school=Gabriel Pereira,
##       traveltime=<15 min.,
##       internet=yes,
##       dalc=very low}   => {higher=yes}   0.3084291  0.9757576  1.066692
## [224] {school=Gabriel Pereira,
##       pstatus=living together,
##       traveltime=<15 min.,
##       dalc=very low}   => {higher=yes}   0.3055556  0.9725610  1.063198
## [225] {pstatus=living together,
##       traveltime=<15 min.,
##       internet=yes,
##       dalc=very low}   => {higher=yes}   0.3026820  0.9663609  1.056420
## [226] {traveltime=<15 min.,
##       schoolsup=no,
##       internet=yes,
##       dalc=very low}   => {higher=yes}   0.3036398  0.9635258  1.053320
## [227] {pstatus=living together,
##       traveltime=<15 min.,
##       schoolsup=no,
##       dalc=very low}   => {higher=yes}   0.3084291  0.9554896  1.044535
## [228] {school=Gabriel Pereira,
##       address=urban,
##       guardian=mother,
##       traveltime=<15 min.} => {higher=yes}    0.3045977  0.9578313  1.047095
## [229] {school=Gabriel Pereira,
##       guardian=mother,

```

##	traveltime=<15 min.,				
##	internet=yes}	=> {higher=yes}	0.3036398	0.9577039	1.046956
## [230]	{school=Gabriel Pereira,				
##	pstatus=living together,				
##	guardian=mother,				
##	traveltime=<15 min.}	=> {higher=yes}	0.3045977	0.9607251	1.050259
## [231]	{school=Gabriel Pereira,				
##	guardian=mother,				
##	traveltime=<15 min.,				
##	schoolsup=no}	=> {higher=yes}	0.3045977	0.9578313	1.047095
## [232]	{school=Gabriel Pereira,				
##	address=urban,				
##	traveltime=<15 min.,				
##	bing3=pass}	=> {higher=yes}	0.3591954	0.9791123	1.070359
## [233]	{school=Gabriel Pereira,				
##	address=urban,				
##	traveltime=<15 min.,				
##	nursery=yes}	=> {higher=yes}	0.3419540	0.9622642	1.051941
## [234]	{address=urban,				
##	traveltime=<15 min.,				
##	internet=yes,				
##	bing3=pass}	=> {higher=yes}	0.3419540	0.9727520	1.063406
## [235]	{address=urban,				
##	traveltime=<15 min.,				
##	nursery=yes,				
##	bing3=pass}	=> {higher=yes}	0.3208812	0.9738372	1.064593
## [236]	{address=urban,				
##	pstatus=living together,				
##	traveltime=<15 min.,				
##	bing3=pass}	=> {higher=yes}	0.3333333	0.9693593	1.059698
## [237]	{address=urban,				
##	traveltime=<15 min.,				
##	schoolsup=no,				
##	bing3=pass}	=> {higher=yes}	0.3448276	0.9677419	1.057929
## [238]	{school=Gabriel Pereira,				
##	traveltime=<15 min.,				
##	paid=no,				
##	bing3=pass}	=> {higher=yes}	0.3132184	0.9732143	1.063912
## [239]	{school=Gabriel Pereira,				
##	traveltime=<15 min.,				
##	internet=yes,				
##	bing3=pass}	=> {higher=yes}	0.3515326	0.9786667	1.069872
## [240]	{school=Gabriel Pereira,				
##	traveltime=<15 min.,				
##	nursery=yes,				
##	bing3=pass}	=> {higher=yes}	0.3333333	0.9830508	1.074665
## [241]	{school=Gabriel Pereira,				
##	pstatus=living together,				
##	traveltime=<15 min.,				
##	bing3=pass}	=> {higher=yes}	0.3534483	0.9787798	1.069996
## [242]	{school=Gabriel Pereira,				
##	traveltime=<15 min.,				
##	schoolsup=no,				
##	bing3=pass}	=> {higher=yes}	0.3563218	0.9789474	1.070179



```

## [243] {school=Gabriel Pereira,
##       traveltime=<15 min.,
##       nursery=yes,
##       internet=yes}           => {higher=yes}           0.3390805  0.9593496  1.048755
## [244] {school=Gabriel Pereira,
##       pstatus=living together,
##       traveltime=<15 min.,
##       nursery=yes}           => {higher=yes}           0.3371648  0.9591281  1.048513
## [245] {school=Gabriel Pereira,
##       traveltime=<15 min.,
##       schoolsup=no,
##       nursery=yes}           => {higher=yes}           0.3314176  0.9611111  1.050681
## [246] {pstatus=living together,
##       traveltime=<15 min.,
##       paid=no,
##       bing3=pass}            => {higher=yes}           0.3074713  0.9639640  1.053799
## [247] {traveltime=<15 min.,
##       schoolsup=no,
##       paid=no,
##       bing3=pass}            => {higher=yes}           0.3122605  0.9644970  1.054382
## [248] {traveltime=<15 min.,
##       nursery=yes,
##       internet=yes,
##       bing3=pass}            => {higher=yes}           0.3151341  0.9733728  1.064085
## [249] {pstatus=living together,
##       traveltime=<15 min.,
##       internet=yes,
##       bing3=pass}            => {higher=yes}           0.3342912  0.9721448  1.062743
## [250] {traveltime=<15 min.,
##       schoolsup=no,
##       internet=yes,
##       bing3=pass}            => {higher=yes}           0.3419540  0.9727520  1.063406
## [251] {pstatus=living together,
##       traveltime=<15 min.,
##       nursery=yes,
##       bing3=pass}            => {higher=yes}           0.3180077  0.9707602  1.061229
## [252] {traveltime=<15 min.,
##       schoolsup=no,
##       nursery=yes,
##       bing3=pass}            => {higher=yes}           0.3275862  0.9743590  1.065163
## [253] {pstatus=living together,
##       traveltime=<15 min.,
##       schoolsup=no,
##       bing3=pass}            => {higher=yes}           0.3515326  0.9708995  1.061381
## [254] {pstatus=living together,
##       famsup=yes,
##       nursery=yes,
##       dalc=very low}         => {higher=yes}           0.3007663  0.9515152  1.040190
## [255] {pstatus=living together,
##       guardian=mother,
##       schoolsup=no,
##       famsup=yes}            => {higher=yes}           0.3036398  0.9519520  1.040668
## [256] {school=Gabriel Pereira,
##       address=urban,

```

```

##      famsup=yes,
##      bing3=pass}          => {higher=yes}          0.3017241  0.9722222  1.062827
## [257] {school=Gabriel Pereira,
##      address=urban,
##      famsup=yes,
##      internet=yes}        => {higher=yes}          0.3266284  0.9660057  1.056031
## [258] {school=Gabriel Pereira,
##      address=urban,
##      famsup=yes,
##      nursery=yes}         => {higher=yes}          0.3026820  0.9723077  1.062921
## [259] {school=Gabriel Pereira,
##      address=urban,
##      pstatus=living together,
##      famsup=yes}          => {higher=yes}          0.3227969  0.9601140  1.049591
## [260] {school=Gabriel Pereira,
##      address=urban,
##      schoolsup=no,
##      famsup=yes}          => {higher=yes}          0.3141762  0.9590643  1.048443
## [261] {address=urban,
##      pstatus=living together,
##      famsup=yes,
##      internet=yes}        => {higher=yes}          0.3189655  0.9568966  1.046073
## [262] {address=urban,
##      schoolsup=no,
##      famsup=yes,
##      internet=yes}        => {higher=yes}          0.3151341  0.9563953  1.045525
## [263] {address=urban,
##      pstatus=living together,
##      famsup=yes,
##      nursery=yes}         => {higher=yes}          0.3074713  0.9553571  1.044390
## [264] {address=urban,
##      schoolsup=no,
##      famsup=yes,
##      nursery=yes}         => {higher=yes}          0.3026820  0.9518072  1.040510
## [265] {school=Gabriel Pereira,
##      famsup=yes,
##      internet=yes,
##      bing3=pass}          => {higher=yes}          0.3180077  0.9679300  1.058135
## [266] {school=Gabriel Pereira,
##      famsup=yes,
##      nursery=yes,
##      bing3=pass}          => {higher=yes}          0.3017241  0.9722222  1.062827
## [267] {school=Gabriel Pereira,
##      pstatus=living together,
##      famsup=yes,
##      bing3=pass}          => {higher=yes}          0.3237548  0.9684814  1.058738
## [268] {school=Gabriel Pereira,
##      schoolsup=no,
##      famsup=yes,
##      bing3=pass}          => {higher=yes}          0.3180077  0.9707602  1.061229
## [269] {school=Gabriel Pereira,
##      famsup=yes,
##      nursery=yes,
##      internet=yes}        => {higher=yes}          0.3218391  0.9600000  1.049466

```

```

## [270] {school=Gabriel Pereira,
##       pstatus=living together,
##       famsup=yes,
##       internet=yes} => {higher=yes} 0.3419540 0.9596774 1.049113
## [271] {school=Gabriel Pereira,
##       schoolsup=no,
##       famsup=yes,
##       internet=yes} => {higher=yes} 0.3323755 0.9612188 1.050798
## [272] {school=Gabriel Pereira,
##       pstatus=living together,
##       famsup=yes,
##       nursery=yes} => {higher=yes} 0.3256705 0.9631728 1.052934
## [273] {school=Gabriel Pereira,
##       schoolsup=no,
##       famsup=yes,
##       nursery=yes} => {higher=yes} 0.3132184 0.9674556 1.057616
## [274] {school=Gabriel Pereira,
##       pstatus=living together,
##       schoolsup=no,
##       famsup=yes} => {higher=yes} 0.3342912 0.9561644 1.045273
## [275] {famsup=yes,
##       nursery=yes,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.3026820 0.9518072 1.040510
## [276] {pstatus=living together,
##       famsup=yes,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.3314176 0.9558011 1.044876
## [277] {schoolsup=no,
##       famsup=yes,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.3323755 0.9585635 1.047896
## [278] {schoolsup=no,
##       famsup=yes,
##       nursery=yes,
##       bing3=pass} => {higher=yes} 0.3218391 0.9545455 1.043503
## [279] {pstatus=living together,
##       schoolsup=no,
##       famsup=yes,
##       bing3=pass} => {higher=yes} 0.3534483 0.9534884 1.042347
## [280] {school=Gabriel Pereira,
##       pstatus=living together,
##       romantic=no,
##       dalc=very low} => {higher=yes} 0.3045977 0.9636364 1.053441
## [281] {pstatus=living together,
##       romantic=no,
##       dalc=very low,
##       bing3=pass} => {higher=yes} 0.3218391 0.9655172 1.055497
## [282] {schoolsup=no,
##       romantic=no,
##       dalc=very low,
##       bing3=pass} => {higher=yes} 0.3093870 0.9670659 1.057190
## [283] {pstatus=living together,
##       internet=yes,

```

```

##      romantic=no,
##      dalc=very low}      => {higher=yes}      0.3065134  0.9696970  1.060067
## [284] {pstatus=living together,
##      nursery=yes,
##      romantic=no,
##      dalc=very low}      => {higher=yes}      0.3103448  0.9614243  1.051023
## [285] {pstatus=living together,
##      schoolsup=no,
##      romantic=no,
##      dalc=very low}      => {higher=yes}      0.3333333  0.9534247  1.042278
## [286] {pstatus=living together,
##      guardian=mother,
##      romantic=no,
##      bing3=pass}          => {higher=yes}      0.3093870  0.9641791  1.054035
## [287] {guardian=mother,
##      schoolsup=no,
##      romantic=no,
##      bing3=pass}          => {higher=yes}      0.3132184  0.9674556  1.057616
## [288] {pstatus=living together,
##      guardian=mother,
##      nursery=yes,
##      romantic=no}          => {higher=yes}      0.3045977  0.9578313  1.047095
## [289] {guardian=mother,
##      schoolsup=no,
##      nursery=yes,
##      romantic=no}          => {higher=yes}      0.3036398  0.9577039  1.046956
## [290] {famsize=greater than 3,
##      higher=yes,
##      internet=yes,
##      romantic=no}          => {pstatus=living together} 0.3122605  0.9560117  1.081339
## [291] {famsize=greater than 3,
##      schoolsup=no,
##      higher=yes,
##      romantic=no}          => {pstatus=living together} 0.3390805  0.9541779  1.079265
## [292] {school=Gabriel Pereira,
##      address=urban,
##      romantic=no,
##      bing3=pass}          => {higher=yes}      0.3247126  0.9630682  1.052820
## [293] {school=Gabriel Pereira,
##      address=urban,
##      nursery=yes,
##      romantic=no}          => {higher=yes}      0.3141762  0.9590643  1.048443
## [294] {address=urban,
##      internet=yes,
##      romantic=no,
##      bing3=pass}          => {higher=yes}      0.3160920  0.9649123  1.054836
## [295] {address=urban,
##      nursery=yes,
##      romantic=no,
##      bing3=pass}          => {higher=yes}      0.3007663  0.9661538  1.056193
## [296] {address=urban,
##      pstatus=living together,
##      romantic=no,
##      bing3=pass}          => {higher=yes}      0.3227969  0.9573864  1.046609

```

```

## [297] {address=urban,
##       schoolsup=no,
##       romantic=no,
##       bing3=pass}          => {higher=yes}          0.3237548  0.9575071  1.046741
## [298] {address=urban,
##       nursery=yes,
##       internet=yes,
##       romantic=no}        => {higher=yes}          0.3026820  0.9518072  1.040510
## [299] {school=Gabriel Pereira,
##       paid=no,
##       romantic=no,
##       bing3=pass}          => {higher=yes}          0.3055556  0.9550898  1.044098
## [300] {school=Gabriel Pereira,
##       internet=yes,
##       romantic=no,
##       bing3=pass}          => {higher=yes}          0.3295019  0.9662921  1.056344
## [301] {school=Gabriel Pereira,
##       nursery=yes,
##       romantic=no,
##       bing3=pass}          => {higher=yes}          0.3227969  0.9768116  1.067844
## [302] {school=Gabriel Pereira,
##       pstatus=living together,
##       romantic=no,
##       bing3=pass}          => {higher=yes}          0.3486590  0.9655172  1.055497
## [303] {school=Gabriel Pereira,
##       schoolsup=no,
##       romantic=no,
##       bing3=pass}          => {higher=yes}          0.3429119  0.9675676  1.057739
## [304] {school=Gabriel Pereira,
##       nursery=yes,
##       internet=yes,
##       romantic=no}        => {higher=yes}          0.3141762  0.9562682  1.045386
## [305] {school=Gabriel Pereira,
##       pstatus=living together,
##       internet=yes,
##       romantic=no}        => {higher=yes}          0.3467433  0.9501312  1.038677
## [306] {school=Gabriel Pereira,
##       schoolsup=no,
##       internet=yes,
##       romantic=no}        => {higher=yes}          0.3285441  0.9501385  1.038685
## [307] {school=Gabriel Pereira,
##       pstatus=living together,
##       nursery=yes,
##       romantic=no}        => {higher=yes}          0.3275862  0.9606742  1.050203
## [308] {school=Gabriel Pereira,
##       schoolsup=no,
##       nursery=yes,
##       romantic=no}        => {higher=yes}          0.3160920  0.9649123  1.054836
## [309] {paid=no,
##       internet=yes,
##       romantic=no,
##       bing3=pass}          => {higher=yes}          0.3084291  0.9554896  1.044535
## [310] {paid=no,
##       nursery=yes,

```

##	romantic=no,				
##	bing3=pass}	=> {higher=yes}	0.3170498	0.9594203	1.048832
##	[311] {nursery=yes,				
##	internet=yes,				
##	romantic=no,				
##	bing3=pass}	=> {higher=yes}	0.3189655	0.9680233	1.058237
##	[312] {pstatus=living together,				
##	internet=yes,				
##	romantic=no,				
##	bing3=pass}	=> {higher=yes}	0.3553640	0.9661458	1.056185
##	[313] {schoolsup=no,				
##	internet=yes,				
##	romantic=no,				
##	bing3=pass}	=> {higher=yes}	0.3515326	0.9683377	1.058581
##	[314] {pstatus=living together,				
##	nursery=yes,				
##	romantic=no,				
##	bing3=pass}	=> {higher=yes}	0.3467433	0.9653333	1.055296
##	[315] {schoolsup=no,				
##	nursery=yes,				
##	romantic=no,				
##	bing3=pass}	=> {higher=yes}	0.3467433	0.9705094	1.060955
##	[316] {pstatus=living together,				
##	schoolsup=no,				
##	romantic=no,				
##	bing3=pass}	=> {higher=yes}	0.3908046	0.9577465	1.047002
##	[317] {schoolsup=no,				
##	nursery=yes,				
##	internet=yes,				
##	romantic=no}	=> {higher=yes}	0.3295019	0.9502762	1.038836
##	[318] {school=Gabriel Pereira,				
##	address=urban,				
##	guardian=mother,				
##	dalc=very low}	=> {higher=yes}	0.3017241	0.9633028	1.053077
##	[319] {school=Gabriel Pereira,				
##	guardian=mother,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3122605	0.9673591	1.057511
##	[320] {school=Gabriel Pereira,				
##	guardian=mother,				
##	internet=yes,				
##	dalc=very low}	=> {higher=yes}	0.3065134	0.9580838	1.047371
##	[321] {school=Gabriel Pereira,				
##	guardian=mother,				
##	nursery=yes,				
##	dalc=very low}	=> {higher=yes}	0.3065134	0.9638554	1.053681
##	[322] {school=Gabriel Pereira,				
##	pstatus=living together,				
##	guardian=mother,				
##	dalc=very low}	=> {higher=yes}	0.3227969	0.9573864	1.046609
##	[323] {school=Gabriel Pereira,				
##	guardian=mother,				
##	schoolsup=no,				
##	dalc=very low}	=> {higher=yes}	0.3141762	0.9507246	1.039326

```

## [324] {guardian=mother,
##       internet=yes,
##       dalc=very low,
##       bing3=pass}           => {higher=yes}           0.3055556  0.9666667  1.056754
## [325] {guardian=mother,
##       nursery=yes,
##       dalc=very low,
##       bing3=pass}           => {higher=yes}           0.3218391  0.9710983  1.061599
## [326] {pstatus=living together,
##       guardian=mother,
##       dalc=very low,
##       bing3=pass}           => {higher=yes}           0.3275862  0.9633803  1.053161
## [327] {guardian=mother,
##       schoolsup=no,
##       dalc=very low,
##       bing3=pass}           => {higher=yes}           0.3400383  0.9594595  1.048875
## [328] {guardian=mother,
##       nursery=yes,
##       internet=yes,
##       dalc=very low}        => {higher=yes}           0.3132184  0.9589443  1.048312
## [329] {pstatus=living together,
##       guardian=mother,
##       internet=yes,
##       dalc=very low}        => {higher=yes}           0.3256705  0.9577465  1.047002
## [330] {guardian=mother,
##       schoolsup=no,
##       internet=yes,
##       dalc=very low}        => {higher=yes}           0.3323755  0.9532967  1.042138
## [331] {school=Gabriel Pereira,
##       address=urban,
##       famsize=greater than 3,
##       dalc=very low}        => {higher=yes}           0.3074713  0.9582090  1.047508
## [332] {school=Gabriel Pereira,
##       famsize=greater than 3,
##       dalc=very low,
##       bing3=pass}           => {higher=yes}           0.3141762  0.9647059  1.054610
## [333] {school=Gabriel Pereira,
##       famsize=greater than 3,
##       internet=yes,
##       dalc=very low}        => {higher=yes}           0.3084291  0.9554896  1.044535
## [334] {school=Gabriel Pereira,
##       famsize=greater than 3,
##       nursery=yes,
##       dalc=very low}        => {higher=yes}           0.3007663  0.9602446  1.049733
## [335] {school=Gabriel Pereira,
##       famsize=greater than 3,
##       pstatus=living together,
##       dalc=very low}        => {higher=yes}           0.3409962  0.9518717  1.040580
## [336] {famsize=greater than 3,
##       internet=yes,
##       dalc=very low,
##       bing3=pass}           => {higher=yes}           0.3093870  0.9584570  1.047779
## [337] {famsize=greater than 3,
##       nursery=yes,

```

##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3113027	0.9530792	1.041900
##	[338] {school=Gabriel Pereira,				
##	address=urban,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3639847	0.9718670	1.062439
##	[339] {school=Gabriel Pereira,				
##	address=urban,				
##	paid=no,				
##	dalc=very low}	=> {higher=yes}	0.3151341	0.9508671	1.039482
##	[340] {school=Gabriel Pereira,				
##	address=urban,				
##	internet=yes,				
##	dalc=very low}	=> {higher=yes}	0.3726054	0.9676617	1.057842
##	[341] {school=Gabriel Pereira,				
##	address=urban,				
##	nursery=yes,				
##	dalc=very low}	=> {higher=yes}	0.3429119	0.9701897	1.060605
##	[342] {school=Gabriel Pereira,				
##	address=urban,				
##	pstatus=living together,				
##	dalc=very low}	=> {higher=yes}	0.3630268	0.9619289	1.051575
##	[343] {school=Gabriel Pereira,				
##	address=urban,				
##	schoolsup=no,				
##	dalc=very low}	=> {higher=yes}	0.3639847	0.9620253	1.051680
##	[344] {address=urban,				
##	paid=no,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3323755	0.9506849	1.039283
##	[345] {address=urban,				
##	internet=yes,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3639847	0.9693878	1.059729
##	[346] {address=urban,				
##	nursery=yes,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3505747	0.9656992	1.055696
##	[347] {address=urban,				
##	pstatus=living together,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3611111	0.9617347	1.051362
##	[348] {address=urban,				
##	schoolsup=no,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3745211	0.9606880	1.050218
##	[349] {address=urban,				
##	nursery=yes,				
##	internet=yes,				
##	dalc=very low}	=> {higher=yes}	0.3409962	0.9595687	1.048995
##	[350] {address=urban,				
##	pstatus=living together,				
##	internet=yes,				
##	dalc=very low}	=> {higher=yes}	0.3668582	0.9598997	1.049356



```

## [351] {address=urban,
##       schoolsup=no,
##       internet=yes,
##       dalc=very low}      => {higher=yes}      0.3745211  0.9606880  1.050218
## [352] {address=urban,
##       pstatus=living together,
##       nursery=yes,
##       dalc=very low}      => {higher=yes}      0.3438697  0.9522546  1.040999
## [353] {address=urban,
##       schoolsup=no,
##       nursery=yes,
##       dalc=very low}      => {higher=yes}      0.3572797  0.9515306  1.040207
## [354] {school=Gabriel Pereira,
##       paid=no,
##       dalc=very low,
##       bing3=pass}         => {higher=yes}      0.3323755  0.9612188  1.050798
## [355] {school=Gabriel Pereira,
##       internet=yes,
##       dalc=very low,
##       bing3=pass}         => {higher=yes}      0.3591954  0.9689922  1.059296
## [356] {school=Gabriel Pereira,
##       nursery=yes,
##       dalc=very low,
##       bing3=pass}         => {higher=yes}      0.3524904  0.9761273  1.067096
## [357] {school=Gabriel Pereira,
##       pstatus=living together,
##       dalc=very low,
##       bing3=pass}         => {higher=yes}      0.3716475  0.9700000  1.060398
## [358] {school=Gabriel Pereira,
##       schoolsup=no,
##       dalc=very low,
##       bing3=pass}         => {higher=yes}      0.3697318  0.9674185  1.057576
## [359] {school=Gabriel Pereira,
##       paid=no,
##       internet=yes,
##       dalc=very low}      => {higher=yes}      0.3151341  0.9508671  1.039482
## [360] {school=Gabriel Pereira,
##       paid=no,
##       nursery=yes,
##       dalc=very low}      => {higher=yes}      0.3093870  0.9556213  1.044679
## [361] {school=Gabriel Pereira,
##       nursery=yes,
##       internet=yes,
##       dalc=very low}      => {higher=yes}      0.3486590  0.9629630  1.052705
## [362] {school=Gabriel Pereira,
##       pstatus=living together,
##       internet=yes,
##       dalc=very low}      => {higher=yes}      0.3726054  0.9652605  1.055217
## [363] {school=Gabriel Pereira,
##       schoolsup=no,
##       internet=yes,
##       dalc=very low}      => {higher=yes}      0.3668582  0.9598997  1.049356
## [364] {school=Gabriel Pereira,
##       pstatus=living together,

```

##	nursery=yes,				
##	dalc=very low}	=> {higher=yes}	0.3563218	0.9612403	1.050822
##	[365] {school=Gabriel Pereira,				
##	schoolsup=no,				
##	nursery=yes,				
##	dalc=very low}	=> {higher=yes}	0.3496169	0.9630607	1.052812
##	[366] {school=Gabriel Pereira,				
##	pstatus=living together,				
##	schoolsup=no,				
##	dalc=very low}	=> {higher=yes}	0.3802682	0.9543269	1.043264
##	[367] {paid=no,				
##	internet=yes,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3381226	0.9566396	1.045792
##	[368] {paid=no,				
##	nursery=yes,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3563218	0.9587629	1.048114
##	[369] {pstatus=living together,				
##	paid=no,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3735632	0.9512195	1.039867
##	[370] {nursery=yes,				
##	internet=yes,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3544061	0.9660574	1.056088
##	[371] {pstatus=living together,				
##	internet=yes,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3802682	0.9659367	1.055956
##	[372] {schoolsup=no,				
##	internet=yes,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3869732	0.9642005	1.054058
##	[373] {pstatus=living together,				
##	nursery=yes,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3793103	0.9611650	1.050740
##	[374] {schoolsup=no,				
##	nursery=yes,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.3917625	0.9646226	1.054519
##	[375] {pstatus=living together,				
##	schoolsup=no,				
##	dalc=very low,				
##	bing3=pass}	=> {higher=yes}	0.4195402	0.9584245	1.047744
##	[376] {pstatus=living together,				
##	nursery=yes,				
##	internet=yes,				
##	dalc=very low}	=> {higher=yes}	0.3706897	0.9508600	1.039474
##	[377] {schoolsup=no,				
##	nursery=yes,				
##	internet=yes,				
##	dalc=very low}	=> {higher=yes}	0.3745211	0.9536585	1.042534

```

## [378] {pstatus=living together,
##       schoolsup=no,
##       internet=yes,
##       dalc=very low}      => {higher=yes}      0.4109195  0.9533333  1.042178
## [379] {famsize=greater than 3,
##       pstatus=living together,
##       guardian=mother,
##       bing3=pass}        => {higher=yes}      0.3266284  0.9578652  1.047132
## [380] {famsize=greater than 3,
##       guardian=mother,
##       schoolsup=no,
##       bing3=pass}        => {higher=yes}      0.3122605  0.9588235  1.048180
## [381] {school=Gabriel Pereira,
##       address=urban,
##       guardian=mother,
##       bing3=pass}        => {higher=yes}      0.3409962  0.9753425  1.066238
## [382] {school=Gabriel Pereira,
##       address=urban,
##       guardian=mother,
##       nursery=yes}      => {higher=yes}      0.3371648  0.9617486  1.051378
## [383] {address=urban,
##       guardian=mother,
##       paid=no,
##       bing3=pass}        => {higher=yes}      0.3026820  0.9575758  1.046816
## [384] {address=urban,
##       guardian=mother,
##       internet=yes,
##       bing3=pass}        => {higher=yes}      0.3352490  0.9722222  1.062827
## [385] {address=urban,
##       guardian=mother,
##       nursery=yes,
##       bing3=pass}        => {higher=yes}      0.3333333  0.9747899  1.065634
## [386] {address=urban,
##       pstatus=living together,
##       guardian=mother,
##       bing3=pass}        => {higher=yes}      0.3295019  0.9690141  1.059320
## [387] {address=urban,
##       guardian=mother,
##       schoolsup=no,
##       bing3=pass}        => {higher=yes}      0.3486590  0.9680851  1.058305
## [388] {school=Gabriel Pereira,
##       guardian=mother,
##       paid=no,
##       bing3=pass}        => {higher=yes}      0.3180077  0.9651163  1.055059
## [389] {school=Gabriel Pereira,
##       guardian=mother,
##       internet=yes,
##       bing3=pass}        => {higher=yes}      0.3486590  0.9732620  1.063964
## [390] {school=Gabriel Pereira,
##       guardian=mother,
##       nursery=yes,
##       bing3=pass}        => {higher=yes}      0.3486590  0.9837838  1.075466
## [391] {school=Gabriel Pereira,
##       pstatus=living together,

```

##	guardian=mother,				
##	bing3=pass}	=> {higher=yes}	0.3572797	0.9764398	1.067438
##	[392] {school=Gabriel Pereira,				
##	guardian=mother,				
##	schoolsup=no,				
##	bing3=pass}	=> {higher=yes}	0.3649425	0.9744246	1.065235
##	[393] {school=Gabriel Pereira,				
##	guardian=mother,				
##	nursery=yes,				
##	internet=yes}	=> {higher=yes}	0.3534483	0.9534884	1.042347
##	[394] {school=Gabriel Pereira,				
##	pstatus=living together,				
##	guardian=mother,				
##	nursery=yes}	=> {higher=yes}	0.3553640	0.9512821	1.039936
##	[395] {school=Gabriel Pereira,				
##	guardian=mother,				
##	schoolsup=no,				
##	nursery=yes}	=> {higher=yes}	0.3534483	0.9534884	1.042347
##	[396] {guardian=mother,				
##	paid=no,				
##	internet=yes,				
##	bing3=pass}	=> {higher=yes}	0.3218391	0.9600000	1.049466
##	[397] {guardian=mother,				
##	paid=no,				
##	nursery=yes,				
##	bing3=pass}	=> {higher=yes}	0.3371648	0.9643836	1.054258
##	[398] {pstatus=living together,				
##	guardian=mother,				
##	paid=no,				
##	bing3=pass}	=> {higher=yes}	0.3477011	0.9527559	1.041547
##	[399] {guardian=mother,				
##	schoolsup=no,				
##	paid=no,				
##	bing3=pass}	=> {higher=yes}	0.3678161	0.9504950	1.039075
##	[400] {guardian=mother,				
##	nursery=yes,				
##	internet=yes,				
##	bing3=pass}	=> {higher=yes}	0.3515326	0.9760638	1.067027
##	[401] {pstatus=living together,				
##	guardian=mother,				
##	internet=yes,				
##	bing3=pass}	=> {higher=yes}	0.3572797	0.9713542	1.061878
##	[402] {guardian=mother,				
##	schoolsup=no,				
##	internet=yes,				
##	bing3=pass}	=> {higher=yes}	0.3773946	0.9704433	1.060883
##	[403] {pstatus=living together,				
##	guardian=mother,				
##	nursery=yes,				
##	bing3=pass}	=> {higher=yes}	0.3630268	0.9693095	1.059643
##	[404] {guardian=mother,				
##	schoolsup=no,				
##	nursery=yes,				
##	bing3=pass}	=> {higher=yes}	0.3840996	0.9733010	1.064007

```

## [405] {pstatus=living together,
##       guardian=mother,
##       schoolsup=no,
##       bing3=pass} => {higher=yes} 0.4013410 0.9632184 1.052984
## [406] {school=Gabriel Pereira,
##       address=urban,
##       famsize=greater than 3,
##       bing3=pass} => {higher=yes} 0.3323755 0.9665738 1.056652
## [407] {address=urban,
##       famsize=greater than 3,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.3218391 0.9627507 1.052473
## [408] {address=urban,
##       famsize=greater than 3,
##       schoolsup=no,
##       bing3=pass} => {higher=yes} 0.3314176 0.9531680 1.041997
## [409] {school=Gabriel Pereira,
##       famsize=greater than 3,
##       paid=no,
##       bing3=pass} => {higher=yes} 0.3170498 0.9538905 1.042787
## [410] {school=Gabriel Pereira,
##       famsize=greater than 3,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.3419540 0.9648649 1.054784
## [411] {school=Gabriel Pereira,
##       famsize=greater than 3,
##       nursery=yes,
##       bing3=pass} => {higher=yes} 0.3227969 0.9656160 1.055605
## [412] {school=Gabriel Pereira,
##       famsize=greater than 3,
##       pstatus=living together,
##       bing3=pass} => {higher=yes} 0.3745211 0.9606880 1.050218
## [413] {school=Gabriel Pereira,
##       famsize=greater than 3,
##       schoolsup=no,
##       bing3=pass} => {higher=yes} 0.3544061 0.9635417 1.053338
## [414] {famsize=greater than 3,
##       schoolsup=no,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.3678161 0.9528536 1.041654
## [415] {school=Gabriel Pereira,
##       address=urban,
##       paid=no,
##       bing3=pass} => {higher=yes} 0.3716475 0.9580247 1.047307
## [416] {school=Gabriel Pereira,
##       address=urban,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.4252874 0.9694323 1.059777
## [417] {school=Gabriel Pereira,
##       address=urban,
##       nursery=yes,
##       bing3=pass} => {higher=yes} 0.3927203 0.9738717 1.064630
## [418] {school=Gabriel Pereira,
##       address=urban,

```

##	pstatus=living together, bing3=pass}	=> {higher=yes}	0.4157088	0.9687500	1.059031
## [419]	{school=Gabriel Pereira, address=urban, schoolsup=no, bing3=pass}	=> {higher=yes}	0.4195402	0.9668874	1.056995
## [420]	{address=urban, paid=no, internet=yes, bing3=pass}	=> {higher=yes}	0.3706897	0.9555556	1.044607
## [421]	{address=urban, paid=no, nursery=yes, bing3=pass}	=> {higher=yes}	0.3572797	0.9539642	1.042868
## [422]	{address=urban, nursery=yes, internet=yes, bing3=pass}	=> {higher=yes}	0.3898467	0.9667458	1.056840
## [423]	{address=urban, pstatus=living together, internet=yes, bing3=pass}	=> {higher=yes}	0.4204981	0.9669604	1.057075
## [424]	{address=urban, schoolsup=no, internet=yes, bing3=pass}	=> {higher=yes}	0.4310345	0.9656652	1.055659
## [425]	{address=urban, pstatus=living together, nursery=yes, bing3=pass}	=> {higher=yes}	0.3879310	0.9619952	1.051647
## [426]	{address=urban, schoolsup=no, nursery=yes, bing3=pass}	=> {higher=yes}	0.4032567	0.9633867	1.053168
## [427]	{address=urban, pstatus=living together, schoolsup=no, bing3=pass}	=> {higher=yes}	0.4329502	0.9596603	1.049095
## [428]	{school=Gabriel Pereira, paid=no, internet=yes, bing3=pass}	=> {higher=yes}	0.3735632	0.9558824	1.044965
## [429]	{school=Gabriel Pereira, paid=no, nursery=yes, bing3=pass}	=> {higher=yes}	0.3620690	0.9642857	1.054151
## [430]	{school=Gabriel Pereira, pstatus=living together, paid=no, bing3=pass}	=> {higher=yes}	0.3908046	0.9577465	1.047002
## [431]	{school=Gabriel Pereira, schoolsup=no, paid=no, bing3=pass}	=> {higher=yes}	0.3898467	0.9553991	1.044436

```

## [432] {school=Gabriel Pereira,
##       nursery=yes,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.4022989 0.9677419 1.057929
## [433] {school=Gabriel Pereira,
##       pstatus=living together,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.4300766 0.9676724 1.057853
## [434] {school=Gabriel Pereira,
##       schoolsup=no,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.4339080 0.9658849 1.055899
## [435] {school=Gabriel Pereira,
##       pstatus=living together,
##       nursery=yes,
##       bing3=pass} => {higher=yes} 0.4051724 0.9701835 1.060598
## [436] {school=Gabriel Pereira,
##       schoolsup=no,
##       nursery=yes,
##       bing3=pass} => {higher=yes} 0.4118774 0.9728507 1.063514
## [437] {school=Gabriel Pereira,
##       pstatus=living together,
##       schoolsup=no,
##       bing3=pass} => {higher=yes} 0.4454023 0.9667360 1.056830
## [438] {pstatus=living together,
##       nursery=yes,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.4185824 0.9520697 1.040797
## [439] {schoolsup=no,
##       nursery=yes,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.4329502 0.9556025 1.044659
## [440] {pstatus=living together,
##       schoolsup=no,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.4770115 0.9558541 1.044934
## [441] {pstatus=living together,
##       schoolsup=no,
##       nursery=yes,
##       bing3=pass} => {higher=yes} 0.4597701 0.9504950 1.039075
## [442] {school=Gabriel Pereira,
##       address=urban,
##       traveltime=<15 min.,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.3208812 0.9795322 1.070818
## [443] {school=Gabriel Pereira,
##       address=urban,
##       pstatus=living together,
##       traveltime=<15 min.,
##       bing3=pass} => {higher=yes} 0.3074713 0.9786585 1.069863
## [444] {school=Gabriel Pereira,
##       address=urban,
##       traveltime=<15 min.,
##       schoolsup=no,

```

```

##      bing3=pass}          => {higher=yes}          0.3160920  0.9763314  1.067319
## [445] {school=Gabriel Pereira,
##      address=urban,
##      traveltime=<15 min.,
##      nursery=yes,
##      internet=yes}      => {higher=yes}          0.3055556  0.9579580  1.047234
## [446] {address=urban,
##      traveltime=<15 min.,
##      schoolsup=no,
##      internet=yes,
##      bing3=pass}        => {higher=yes}          0.3055556  0.9696049  1.059966
## [447] {school=Gabriel Pereira,
##      pstatus=living together,
##      traveltime=<15 min.,
##      internet=yes,
##      bing3=pass}        => {higher=yes}          0.3045977  0.9784615  1.069648
## [448] {school=Gabriel Pereira,
##      traveltime=<15 min.,
##      schoolsup=no,
##      internet=yes,
##      bing3=pass}        => {higher=yes}          0.3113027  0.9789157  1.070144
## [449] {school=Gabriel Pereira,
##      pstatus=living together,
##      traveltime=<15 min.,
##      schoolsup=no,
##      bing3=pass}        => {higher=yes}          0.3113027  0.9789157  1.070144
## [450] {school=Gabriel Pereira,
##      pstatus=living together,
##      schoolsup=no,
##      romantic=no,
##      bing3=pass}        => {higher=yes}          0.3036398  0.9694190  1.059763
## [451] {pstatus=living together,
##      schoolsup=no,
##      internet=yes,
##      romantic=no,
##      bing3=pass}        => {higher=yes}          0.3160920  0.9705882  1.061041
## [452] {pstatus=living together,
##      schoolsup=no,
##      nursery=yes,
##      romantic=no,
##      bing3=pass}        => {higher=yes}          0.3026820  0.9693252  1.059660
## [453] {school=Gabriel Pereira,
##      address=urban,
##      internet=yes,
##      dalc=very low,
##      bing3=pass}        => {higher=yes}          0.3237548  0.9740634  1.064840
## [454] {school=Gabriel Pereira,
##      address=urban,
##      nursery=yes,
##      dalc=very low,
##      bing3=pass}        => {higher=yes}          0.3026820  0.9783282  1.069502
## [455] {school=Gabriel Pereira,
##      address=urban,
##      pstatus=living together,

```



```

##      dalc=very low,
##      bing3=pass} => {higher=yes}      0.3132184  0.9732143  1.063912
## [456] {school=Gabriel Pereira,
##      address=urban,
##      schoolsup=no,
##      dalc=very low,
##      bing3=pass} => {higher=yes}      0.3208812  0.9738372  1.064593
## [457] {school=Gabriel Pereira,
##      address=urban,
##      nursery=yes,
##      internet=yes,
##      dalc=very low} => {higher=yes}    0.3026820  0.9663609  1.056420
## [458] {school=Gabriel Pereira,
##      address=urban,
##      pstatus=living together,
##      internet=yes,
##      dalc=very low} => {higher=yes}    0.3247126  0.9685714  1.058836
## [459] {school=Gabriel Pereira,
##      address=urban,
##      schoolsup=no,
##      internet=yes,
##      dalc=very low} => {higher=yes}    0.3295019  0.9690141  1.059320
## [460] {school=Gabriel Pereira,
##      address=urban,
##      pstatus=living together,
##      schoolsup=no,
##      dalc=very low} => {higher=yes}    0.3208812  0.9626437  1.052356
## [461] {address=urban,
##      nursery=yes,
##      internet=yes,
##      dalc=very low,
##      bing3=pass} => {higher=yes}      0.3007663  0.9691358  1.059453
## [462] {address=urban,
##      pstatus=living together,
##      internet=yes,
##      dalc=very low,
##      bing3=pass} => {higher=yes}      0.3160920  0.9705882  1.061041
## [463] {address=urban,
##      schoolsup=no,
##      internet=yes,
##      dalc=very low,
##      bing3=pass} => {higher=yes}      0.3304598  0.9718310  1.062400
## [464] {address=urban,
##      schoolsup=no,
##      nursery=yes,
##      dalc=very low,
##      bing3=pass} => {higher=yes}      0.3132184  0.9674556  1.057616
## [465] {address=urban,
##      pstatus=living together,
##      schoolsup=no,
##      dalc=very low,
##      bing3=pass} => {higher=yes}      0.3275862  0.9633803  1.053161
## [466] {address=urban,
##      schoolsup=no,

```

```

##      nursery=yes,
##      internet=yes,
##      dalc=very low}      => {higher=yes}      0.3036398  0.9606061  1.050129
## [467] {address=urban,
##      pstatus=living together,
##      schoolsup=no,
##      internet=yes,
##      dalc=very low}      => {higher=yes}      0.3304598  0.9610028  1.050562
## [468] {address=urban,
##      pstatus=living together,
##      schoolsup=no,
##      nursery=yes,
##      dalc=very low}      => {higher=yes}      0.3093870  0.9528024  1.041598
## [469] {school=Gabriel Pereira,
##      pstatus=living together,
##      internet=yes,
##      dalc=very low,
##      bing3=pass}        => {higher=yes}      0.3141762  0.9732938  1.063999
## [470] {school=Gabriel Pereira,
##      schoolsup=no,
##      internet=yes,
##      dalc=very low,
##      bing3=pass}        => {higher=yes}      0.3180077  0.9707602  1.061229
## [471] {school=Gabriel Pereira,
##      pstatus=living together,
##      nursery=yes,
##      dalc=very low,
##      bing3=pass}        => {higher=yes}      0.3017241  0.9752322  1.066118
## [472] {school=Gabriel Pereira,
##      schoolsup=no,
##      nursery=yes,
##      dalc=very low,
##      bing3=pass}        => {higher=yes}      0.3074713  0.9786585  1.069863
## [473] {school=Gabriel Pereira,
##      pstatus=living together,
##      schoolsup=no,
##      dalc=very low,
##      bing3=pass}        => {higher=yes}      0.3275862  0.9715909  1.062137
## [474] {school=Gabriel Pereira,
##      pstatus=living together,
##      nursery=yes,
##      internet=yes,
##      dalc=very low}      => {higher=yes}      0.3036398  0.9606061  1.050129
## [475] {school=Gabriel Pereira,
##      schoolsup=no,
##      nursery=yes,
##      internet=yes,
##      dalc=very low}      => {higher=yes}      0.3026820  0.9634146  1.053199
## [476] {school=Gabriel Pereira,
##      pstatus=living together,
##      schoolsup=no,
##      internet=yes,
##      dalc=very low}      => {higher=yes}      0.3247126  0.9658120  1.055820
## [477] {school=Gabriel Pereira,

```

```

##      pstatus=living together,
##      schoolsup=no,
##      nursery=yes,
##      dalc=very low}          => {higher=yes}          0.3084291  0.9611940  1.050771
## [478] {schoolsup=no,
##      paid=no,
##      internet=yes,
##      dalc=very low,
##      bing3=pass}            => {higher=yes}          0.3026820  0.9575758  1.046816
## [479] {pstatus=living together,
##      paid=no,
##      nursery=yes,
##      dalc=very low,
##      bing3=pass}            => {higher=yes}          0.3036398  0.9548193  1.043802
## [480] {schoolsup=no,
##      paid=no,
##      nursery=yes,
##      dalc=very low,
##      bing3=pass}            => {higher=yes}          0.3141762  0.9590643  1.048443
## [481] {pstatus=living together,
##      schoolsup=no,
##      paid=no,
##      dalc=very low,
##      bing3=pass}            => {higher=yes}          0.3323755  0.9506849  1.039283
## [482] {pstatus=living together,
##      nursery=yes,
##      internet=yes,
##      dalc=very low,
##      bing3=pass}            => {higher=yes}          0.3065134  0.9638554  1.053681
## [483] {schoolsup=no,
##      nursery=yes,
##      internet=yes,
##      dalc=very low,
##      bing3=pass}            => {higher=yes}          0.3199234  0.9681159  1.058338
## [484] {pstatus=living together,
##      schoolsup=no,
##      internet=yes,
##      dalc=very low,
##      bing3=pass}            => {higher=yes}          0.3448276  0.9677419  1.057929
## [485] {pstatus=living together,
##      schoolsup=no,
##      nursery=yes,
##      dalc=very low,
##      bing3=pass}            => {higher=yes}          0.3409962  0.9621622  1.051830
## [486] {pstatus=living together,
##      schoolsup=no,
##      nursery=yes,
##      internet=yes,
##      dalc=very low}          => {higher=yes}          0.3314176  0.9505495  1.039135
## [487] {school=Gabriel Pereira,
##      address=urban,
##      guardian=mother,
##      internet=yes,
##      bing3=pass}            => {higher=yes}          0.3036398  0.9753846  1.066284

```

```

## [488] {school=Gabriel Pereira,
##       address=urban,
##       guardian=mother,
##       schoolsup=no,
##       bing3=pass} => {higher=yes} 0.3017241 0.9782609 1.069429
## [489] {school=Gabriel Pereira,
##       address=urban,
##       guardian=mother,
##       nursery=yes,
##       internet=yes} => {higher=yes} 0.3007663 0.9573171 1.046533
## [490] {address=urban,
##       guardian=mother,
##       schoolsup=no,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.3045977 0.9754601 1.066367
## [491] {school=Gabriel Pereira,
##       guardian=mother,
##       schoolsup=no,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.3103448 0.9759036 1.066852
## [492] {school=Gabriel Pereira,
##       guardian=mother,
##       schoolsup=no,
##       nursery=yes,
##       bing3=pass} => {higher=yes} 0.3036398 0.9875389 1.079571
## [493] {school=Gabriel Pereira,
##       pstatus=living together,
##       guardian=mother,
##       schoolsup=no,
##       bing3=pass} => {higher=yes} 0.3160920 0.9792285 1.070486
## [494] {school=Gabriel Pereira,
##       guardian=mother,
##       schoolsup=no,
##       nursery=yes,
##       internet=yes} => {higher=yes} 0.3055556 0.9522388 1.040981
## [495] {guardian=mother,
##       schoolsup=no,
##       paid=no,
##       nursery=yes,
##       bing3=pass} => {higher=yes} 0.3017241 0.9662577 1.056307
## [496] {pstatus=living together,
##       guardian=mother,
##       schoolsup=no,
##       paid=no,
##       bing3=pass} => {higher=yes} 0.3151341 0.9536232 1.042495
## [497] {guardian=mother,
##       schoolsup=no,
##       nursery=yes,
##       internet=yes,
##       bing3=pass} => {higher=yes} 0.3160920 0.9792285 1.070486
## [498] {pstatus=living together,
##       guardian=mother,
##       schoolsup=no,
##       internet=yes,

```

```

##      bing3=pass}          => {higher=yes}          0.3256705  0.9742120  1.065002
## [499] {pstatus=living together,
##      guardian=mother,
##      schoolsup=no,
##      nursery=yes,
##      bing3=pass}          => {higher=yes}          0.3256705  0.9714286  1.061960
## [500] {school=Gabriel Pereira,
##      address=urban,
##      famsize=greater than 3,
##      pstatus=living together,
##      bing3=pass}          => {higher=yes}          0.3026820  0.9634146  1.053199
## [501] {school=Gabriel Pereira,
##      famsize=greater than 3,
##      pstatus=living together,
##      internet=yes,
##      bing3=pass}          => {higher=yes}          0.3151341  0.9648094  1.054724
## [502] {school=Gabriel Pereira,
##      famsize=greater than 3,
##      schoolsup=no,
##      internet=yes,
##      bing3=pass}          => {higher=yes}          0.3017241  0.9692308  1.059557
## [503] {school=Gabriel Pereira,
##      famsize=greater than 3,
##      pstatus=living together,
##      schoolsup=no,
##      bing3=pass}          => {higher=yes}          0.3256705  0.9631728  1.052934
## [504] {famsize=greater than 3,
##      pstatus=living together,
##      schoolsup=no,
##      internet=yes,
##      bing3=pass}          => {higher=yes}          0.3400383  0.9517426  1.040439
## [505] {school=Gabriel Pereira,
##      address=urban,
##      paid=no,
##      internet=yes,
##      bing3=pass}          => {higher=yes}          0.3247126  0.9603399  1.049838
## [506] {school=Gabriel Pereira,
##      address=urban,
##      paid=no,
##      nursery=yes,
##      bing3=pass}          => {higher=yes}          0.3026820  0.9663609  1.056420
## [507] {school=Gabriel Pereira,
##      address=urban,
##      pstatus=living together,
##      paid=no,
##      bing3=pass}          => {higher=yes}          0.3189655  0.9596542  1.049088
## [508] {school=Gabriel Pereira,
##      address=urban,
##      schoolsup=no,
##      paid=no,
##      bing3=pass}          => {higher=yes}          0.3218391  0.9572650  1.046476
## [509] {school=Gabriel Pereira,
##      address=urban,
##      nursery=yes,

```

##	internet=yes,				
##	bing3=pass}	=> {higher=yes}	0.3477011	0.9731903	1.063886
##	[510] {school=Gabriel Pereira,				
##	address=urban,				
##	pstatus=living together,				
##	internet=yes,				
##	bing3=pass}	=> {higher=yes}	0.3716475	0.9724311	1.063056
##	[511] {school=Gabriel Pereira,				
##	address=urban,				
##	schoolsup=no,				
##	internet=yes,				
##	bing3=pass}	=> {higher=yes}	0.3783525	0.9705160	1.060962
##	[512] {school=Gabriel Pereira,				
##	address=urban,				
##	pstatus=living together,				
##	nursery=yes,				
##	bing3=pass}	=> {higher=yes}	0.3352490	0.9749304	1.065788
##	[513] {school=Gabriel Pereira,				
##	address=urban,				
##	schoolsup=no,				
##	nursery=yes,				
##	bing3=pass}	=> {higher=yes}	0.3438697	0.9755435	1.066458
##	[514] {school=Gabriel Pereira,				
##	address=urban,				
##	pstatus=living together,				
##	schoolsup=no,				
##	bing3=pass}	=> {higher=yes}	0.3678161	0.9696970	1.060067
##	[515] {address=urban,				
##	pstatus=living together,				
##	paid=no,				
##	internet=yes,				
##	bing3=pass}	=> {higher=yes}	0.3218391	0.9572650	1.046476
##	[516] {address=urban,				
##	schoolsup=no,				
##	paid=no,				
##	internet=yes,				
##	bing3=pass}	=> {higher=yes}	0.3304598	0.9556787	1.044742
##	[517] {address=urban,				
##	pstatus=living together,				
##	paid=no,				
##	nursery=yes,				
##	bing3=pass}	=> {higher=yes}	0.3007663	0.9515152	1.040190
##	[518] {address=urban,				
##	schoolsup=no,				
##	paid=no,				
##	nursery=yes,				
##	bing3=pass}	=> {higher=yes}	0.3132184	0.9533528	1.042199
##	[519] {address=urban,				
##	pstatus=living together,				
##	nursery=yes,				
##	internet=yes,				
##	bing3=pass}	=> {higher=yes}	0.3371648	0.9670330	1.057154
##	[520] {address=urban,				
##	schoolsup=no,				

```

##      nursery=yes,
##      internet=yes,
##      bing3=pass} => {higher=yes}          0.3496169  0.9681698  1.058397
## [521] {address=urban,
##      pstatus=living together,
##      schoolsup=no,
##      internet=yes,
##      bing3=pass} => {higher=yes}          0.3802682  0.9682927  1.058531
## [522] {address=urban,
##      pstatus=living together,
##      schoolsup=no,
##      nursery=yes,
##      bing3=pass} => {higher=yes}          0.3486590  0.9629630  1.052705
## [523] {school=Gabriel Pereira,
##      paid=no,
##      nursery=yes,
##      internet=yes,
##      bing3=pass} => {higher=yes}          0.3026820  0.9604863  1.049998
## [524] {school=Gabriel Pereira,
##      pstatus=living together,
##      paid=no,
##      internet=yes,
##      bing3=pass} => {higher=yes}          0.3256705  0.9604520  1.049960
## [525] {school=Gabriel Pereira,
##      schoolsup=no,
##      paid=no,
##      internet=yes,
##      bing3=pass} => {higher=yes}          0.3285441  0.9581006  1.047390
## [526] {school=Gabriel Pereira,
##      pstatus=living together,
##      paid=no,
##      nursery=yes,
##      bing3=pass} => {higher=yes}          0.3103448  0.9642857  1.054151
## [527] {school=Gabriel Pereira,
##      schoolsup=no,
##      paid=no,
##      nursery=yes,
##      bing3=pass} => {higher=yes}          0.3141762  0.9675516  1.057721
## [528] {school=Gabriel Pereira,
##      pstatus=living together,
##      schoolsup=no,
##      paid=no,
##      bing3=pass} => {higher=yes}          0.3409962  0.9595687  1.048995
## [529] {school=Gabriel Pereira,
##      pstatus=living together,
##      nursery=yes,
##      internet=yes,
##      bing3=pass} => {higher=yes}          0.3467433  0.9679144  1.058118
## [530] {school=Gabriel Pereira,
##      schoolsup=no,
##      nursery=yes,
##      internet=yes,
##      bing3=pass} => {higher=yes}          0.3563218  0.9712794  1.061797
## [531] {school=Gabriel Pereira,

```

```

##      pstatus=living together,
##      schoolsup=no,
##      internet=yes,
##      bing3=pass}          => {higher=yes}          0.3802682  0.9706601  1.061120
## [532] {school=Gabriel Pereira,
##      pstatus=living together,
##      schoolsup=no,
##      nursery=yes,
##      bing3=pass}          => {higher=yes}          0.3572797  0.9738903  1.064651
## [533] {pstatus=living together,
##      schoolsup=no,
##      nursery=yes,
##      internet=yes,
##      bing3=pass}          => {higher=yes}          0.3793103  0.9542169  1.043144
## [534] {school=Gabriel Pereira,
##      address=urban,
##      schoolsup=no,
##      nursery=yes,
##      internet=yes,
##      bing3=pass}          => {higher=yes}          0.3093870  0.9758308  1.066772
## [535] {school=Gabriel Pereira,
##      address=urban,
##      pstatus=living together,
##      schoolsup=no,
##      internet=yes,
##      bing3=pass}          => {higher=yes}          0.3323755  0.9747191  1.065557
## [536] {address=urban,
##      pstatus=living together,
##      schoolsup=no,
##      nursery=yes,
##      internet=yes,
##      bing3=pass}          => {higher=yes}          0.3055556  0.9696049  1.059966
## [537] {school=Gabriel Pereira,
##      pstatus=living together,
##      schoolsup=no,
##      nursery=yes,
##      internet=yes,
##      bing3=pass}          => {higher=yes}          0.3084291  0.9728097  1.063469

```

```

rm(ap2)
ap2<-apriori(data_aso,parameter=list ( supp=0.25,conf= 0.90,target="rules"), appearance=list(rhs=c("bing

```

```

## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##      0.9      0.1      1 none FALSE              TRUE      5      0.25      1
## maxlen target  ext
##      10 rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##      0.1 TRUE TRUE FALSE TRUE      2      TRUE
##
## Absolute minimum support count: 261

```



```
##
## set item appearances ...[2 item(s)] done [0.00s].
## set transactions ...[93 item(s), 1044 transaction(s)] done [0.00s].
## sorting and recoding items ... [41 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 7 8 done [0.09s].
## writing ... [18 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
```

```
inspect(ap2)
```

	lhs	rhs	support	confidence	lift
## [1]	{school=Gabriel Pereira, address=urban, guardian=mother, paid=no, higher=yes}	=> {bing3=pass}	0.2595785	0.9063545	1.162450
## [2]	{address=urban, guardian=mother, paid=no, higher=yes, internet=yes}	=> {bing3=pass}	0.2557471	0.9050847	1.160821
## [3]	{school=Gabriel Pereira, address=urban, paid=no, higher=yes, internet=yes}	=> {bing3=pass}	0.3247126	0.9015957	1.156346
## [4]	{school=Gabriel Pereira, address=urban, paid=no, nursery=yes, higher=yes}	=> {bing3=pass}	0.3026820	0.9028571	1.157964
## [5]	{school=Gabriel Pereira, address=urban, schoolsup=no, paid=no, higher=yes}	=> {bing3=pass}	0.3218391	0.9008043	1.155331
## [6]	{address=urban, paid=no, nursery=yes, higher=yes, internet=yes}	=> {bing3=pass}	0.2988506	0.9017341	1.156524
## [7]	{address=urban, schoolsup=no, paid=no, higher=yes, internet=yes}	=> {bing3=pass}	0.3304598	0.9007833	1.155304
## [8]	{school=Gabriel Pereira, address=urban, schoolsup=no, nursery=yes, higher=yes, dalc=very low}	=> {bing3=pass}	0.2681992	0.9032258	1.158437
## [9]	{address=urban, schoolsup=no,				

```

##      paid=no,
##      higher=yes,
##      internet=yes,
##      dalc=very low}          => {bing3=pass} 0.2538314 0.9106529 1.167963
## [10] {address=urban,
##      schoolsup=no,
##      nursery=yes,
##      higher=yes,
##      internet=yes,
##      dalc=very low}          => {bing3=pass} 0.2739464 0.9022082 1.157132
## [11] {school=Gabriel Pereira,
##      address=urban,
##      paid=no,
##      nursery=yes,
##      higher=yes,
##      internet=yes}          => {bing3=pass} 0.2624521 0.9194631 1.179262
## [12] {school=Gabriel Pereira,
##      address=urban,
##      schoolsup=no,
##      paid=no,
##      higher=yes,
##      internet=yes}          => {bing3=pass} 0.2863985 0.9171779 1.176331
## [13] {school=Gabriel Pereira,
##      address=urban,
##      pstatus=living together,
##      paid=no,
##      nursery=yes,
##      higher=yes}            => {bing3=pass} 0.2557471 0.9020270 1.156900
## [14] {school=Gabriel Pereira,
##      address=urban,
##      schoolsup=no,
##      paid=no,
##      nursery=yes,
##      higher=yes}            => {bing3=pass} 0.2614943 0.9222973 1.182897
## [15] {school=Gabriel Pereira,
##      address=urban,
##      pstatus=living together,
##      schoolsup=no,
##      paid=no,
##      higher=yes}            => {bing3=pass} 0.2777778 0.9062500 1.162316
## [16] {address=urban,
##      schoolsup=no,
##      paid=no,
##      nursery=yes,
##      higher=yes,
##      internet=yes}          => {bing3=pass} 0.2653257 0.9172185 1.176383
## [17] {address=urban,
##      pstatus=living together,
##      schoolsup=no,
##      paid=no,
##      higher=yes,
##      internet=yes}          => {bing3=pass} 0.2873563 0.9009009 1.155455
## [18] {school=Gabriel Pereira,
##      schoolsup=no,

```

```

##      paid=no,
##      nursery=yes,
##      higher=yes,
##      internet=yes}          => {bing3=pass} 0.2662835 0.9055375 1.161402
# quitamos la variable higher ya que en el 91.5% de los casos tiene el valor sí
variables_aso <- c("school", "sex", "address", "famsize", "pstatus", "medu", "fedu", "mjob", "fjob", "r

data_aso <- data[,variables_aso]

rm(ap1)
ap1 <- apriori(data_aso, parameter=list ( supp=0.40,conf= 0.90,target="rules"))

## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##      0.9      0.1      1 none FALSE              TRUE          5      0.4      1
## maxlen target  ext
##      10 rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##      0.1 TRUE TRUE  FALSE TRUE      2      TRUE
##
## Absolute minimum support count: 417
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[91 item(s), 1044 transaction(s)] done [0.00s].
## sorting and recoding items ... [22 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 done [0.01s].
## writing ... [31 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].

inspect(ap1)

##      lhs                                rhs                                support confidence lift
## [1] {sex=male}                          => {schoolsup=no}                    0.4032567 0.9293598 1.048921
## [2] {activities=yes}                     => {pstatus=living together} 0.4530651 0.9166667 1.036836
## [3] {famsize=greater than 3}             => {pstatus=living together} 0.6551724 0.9268293 1.048331
## [4] {schoolsup=no,                       => {pstatus=living together} 0.4022989 0.9190372 1.039518
##      activities=yes}
## [5] {famsize=greater than 3,             => {pstatus=living together} 0.4176245 0.9316239 1.053754
##      famsup=yes}
## [6] {romantic=no,                       => {pstatus=living together} 0.4061303 0.9002123 1.018225
##      dalc=very low}
## [7] {famsize=greater than 3,             => {pstatus=living together} 0.4291188 0.9451477 1.069051
##      romantic=no}
## [8] {internet=yes,                       => {pstatus=living together} 0.4549808 0.9082218 1.027284
##      romantic=no}
## [9] {schoolsup=no,                       => {pstatus=living together} 0.5019157 0.9003436 1.018374
##      romantic=no}

```

```

## [10] {famsize=greater than 3,
##       dalc=very low} => {pstatus=living together} 0.4703065 0.9160448 1.036133
## [11] {famsize=greater than 3,
##       guardian=mother} => {pstatus=living together} 0.4521073 0.9254902 1.046817
## [12] {guardian=mother,
##       paid=no} => {schoolsup=no} 0.4913793 0.9047619 1.021158
## [13] {address=urban,
##       famsize=greater than 3} => {pstatus=living together} 0.4607280 0.9161905 1.036298
## [14] {school=Gabriel Pereira,
##       famsize=greater than 3} => {pstatus=living together} 0.4865900 0.9186257 1.039052
## [15] {famsize=greater than 3,
##       bing3=pass} => {pstatus=living together} 0.5019157 0.9257951 1.047161
## [16] {famsize=greater than 3,
##       paid=no} => {pstatus=living together} 0.5086207 0.9202773 1.040920
## [17] {famsize=greater than 3,
##       internet=yes} => {pstatus=living together} 0.5210728 0.9331046 1.055429
## [18] {famsize=greater than 3,
##       nursery=yes} => {pstatus=living together} 0.5057471 0.9246935 1.045916
## [19] {famsize=greater than 3,
##       schoolsup=no} => {pstatus=living together} 0.5766284 0.9304482 1.052425
## [20] {famsize=greater than 3,
##       schoolsup=no,
##       dalc=very low} => {pstatus=living together} 0.4080460 0.9220779 1.042957
## [21] {pstatus=living together,
##       guardian=mother,
##       bing3=pass} => {schoolsup=no} 0.4166667 0.9006211 1.016485
## [22] {pstatus=living together,
##       guardian=mother,
##       paid=no} => {schoolsup=no} 0.4262452 0.9100204 1.027093
## [23] {address=urban,
##       famsize=greater than 3,
##       schoolsup=no} => {pstatus=living together} 0.4070881 0.9239130 1.045033
## [24] {school=Gabriel Pereira,
##       famsize=greater than 3,
##       internet=yes} => {pstatus=living together} 0.4109195 0.9265659 1.048033
## [25] {school=Gabriel Pereira,
##       famsize=greater than 3,
##       schoolsup=no} => {pstatus=living together} 0.4166667 0.9216102 1.042428
## [26] {famsize=greater than 3,
##       internet=yes,
##       bing3=pass} => {pstatus=living together} 0.4013410 0.9269912 1.048514
## [27] {famsize=greater than 3,
##       schoolsup=no,
##       bing3=pass} => {pstatus=living together} 0.4444444 0.9298597 1.051759
## [28] {famsize=greater than 3,
##       schoolsup=no,
##       paid=no} => {pstatus=living together} 0.4492337 0.9250493 1.046318
## [29] {famsize=greater than 3,
##       nursery=yes,
##       internet=yes} => {pstatus=living together} 0.4013410 0.9331849 1.055520
## [30] {famsize=greater than 3,
##       schoolsup=no,
##       internet=yes} => {pstatus=living together} 0.4588123 0.9300971 1.052027
## [31] {famsize=greater than 3,

```

```

##      schoolsup=no,
##      nursery=yes}          => {pstatus=living together} 0.4454023  0.9318637 1.054026
ap2<-apriori(data_aso,parameter=list ( supp=0.25,conf= 0.90,target="rules"), appearance=list(rhs=c("bing
## Apriori
##
## Parameter specification:
## confidence minval smax arem  aval originalSupport maxtime support minlen
##          0.9   0.1   1 none FALSE          TRUE        5   0.25   1
## maxlen target  ext
##          10  rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 261
##
## set item appearances ...[2 item(s)] done [0.00s].
## set transactions ...[91 item(s), 1044 transaction(s)] done [0.00s].
## sorting and recoding items ... [40 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 7 done [0.06s].
## writing ... [0 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
inspect(ap2)

ap3<-apriori(data_aso,parameter=list ( supp=0.1,conf= 0.45,target="rules"), appearance=list(rhs=c("bing
## Apriori
##
## Parameter specification:
## confidence minval smax arem  aval originalSupport maxtime support minlen
##          0.45   0.1   1 none FALSE          TRUE        5   0.1   1
## maxlen target  ext
##          10  rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 104
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[91 item(s), 1044 transaction(s)] done [0.00s].
## sorting and recoding items ... [73 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 7 8 9 10
##
## Warning in apriori(data_aso, parameter = list(supp = 0.1, conf = 0.45,
## target = "rules"), : Mining stopped (maxlen reached). Only patterns up to a
## length of 10 returned!

```

```
## done [0.73s].
## writing ... [0 rule(s)] done [0.01s].
## creating S4 object ... done [0.01s].
```

```
inspect(ap3)

#str(data)
#names(data)

save.image("mywspace_TFM.RData")
```

## Profiling

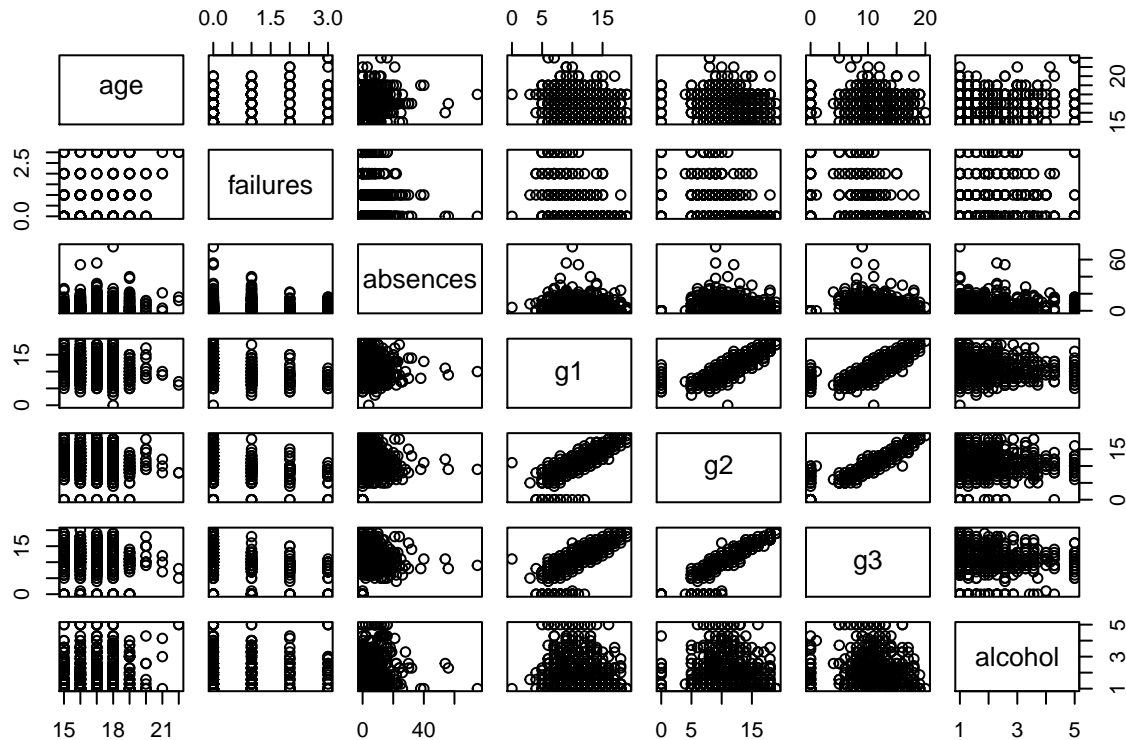
Se usa como variable de éxito *nbing3* .

```
data_profiling<- data
str(data_profiling)

## 'data.frame': 1044 obs. of 39 variables:
## $ school : Factor w/ 2 levels "Gabriel Pereira",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ sex : Factor w/ 2 levels "female","male": 1 1 1 1 1 2 2 1 2 2 ...
## $ age : int 18 17 15 15 16 16 16 17 15 15 ...
## $ address : Factor w/ 2 levels "rural","urban": 2 2 2 2 2 2 2 2 2 2 ...
## $ famsize : Factor w/ 2 levels "greater than 3",...: 1 1 2 1 1 2 2 1 2 1 ...
## $ pstatus : Factor w/ 2 levels "living apart",...: 1 2 2 2 2 2 2 1 1 2 ...
## $ medu : Factor w/ 5 levels "none","primary",...: 5 2 2 5 4 5 3 5 4 4 ...
## $ fedu : Factor w/ 5 levels "none","primary",...: 5 2 2 3 4 4 3 5 3 5 ...
## $ mjob : Factor w/ 5 levels "at_home","health",...: 1 1 1 2 3 4 3 3 4 3 ...
## $ fjob : Factor w/ 5 levels "at_home","health",...: 5 3 3 4 3 3 3 5 3 3 ...
## $ reason : Factor w/ 4 levels "course","home",...: 1 1 3 2 2 4 2 2 2 2 ...
## $ guardian : Factor w/ 3 levels "father","mother",...: 2 1 2 2 1 2 2 2 2 2 ...
## $ traveltime : Factor w/ 4 levels "<15 min.,"15 to 30 min.",...: 2 1 1 1 1 1 1 1 2 1 ...
## $ studytime : Factor w/ 4 levels "<2 hours","2 to 5 hours",...: 2 2 2 3 2 2 2 2 2 2 ...
## $ failures : int 0 0 3 0 0 0 0 0 0 0 ...
## $ schoolsup : Factor w/ 2 levels "no","yes": 2 1 2 1 1 1 1 2 1 1 ...
## $ famsup : Factor w/ 2 levels "no","yes": 1 2 1 2 2 2 1 2 2 2 ...
## $ paid : Factor w/ 2 levels "no","yes": 1 1 2 2 2 2 1 1 2 2 ...
## $ activities : Factor w/ 2 levels "no","yes": 1 1 1 2 1 2 1 1 1 2 ...
## $ nursery : Factor w/ 2 levels "no","yes": 2 1 2 2 2 2 2 2 2 2 ...
## $ higher : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2 2 ...
## $ internet : Factor w/ 2 levels "no","yes": 1 2 2 2 1 2 2 1 2 2 ...
## $ romantic : Factor w/ 2 levels "no","yes": 1 1 1 2 1 1 1 1 1 1 ...
## $ famrel : Factor w/ 5 levels "very bad","bad",...: 4 5 4 3 4 5 4 4 4 5 ...
## $ freetime : Factor w/ 5 levels "very low","low",...: 3 3 3 2 3 4 4 1 2 5 ...
## $ goout : Factor w/ 5 levels "very low","low",...: 4 3 2 2 2 2 4 4 2 1 ...
## $ dalc : Factor w/ 5 levels "very low","low",...: 1 1 2 1 1 1 1 1 1 1 ...
## $ walc : Factor w/ 5 levels "very low","low",...: 1 1 3 1 2 2 1 1 1 1 ...
## $ health : Factor w/ 5 levels "very bad","bad",...: 3 3 3 5 5 5 3 1 1 5 ...
## $ absences : int 6 4 10 2 4 10 0 6 0 0 ...
## $ g1 : int 5 5 7 15 6 15 12 6 16 14 ...
## $ g2 : int 6 5 8 14 10 15 12 5 18 15 ...
## $ g3 : int 6 6 10 15 10 15 11 6 19 15 ...
## $ file : Factor w/ 2 levels "math","portuguese": 1 1 1 1 1 1 1 1 1 1 ...
```

```
## $ subject      : Factor w/ 3 levels "math","math&portuguese",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ student_key: chr  "GP-F-18-U-GT3-A-4-4-at_home-teacher-course-yes-no" "GP-F-17-U-GT3-T-1-1-at_hom
## $ bing3        : Factor w/ 2 levels "fail","pass": 1 1 2 2 2 2 1 2 2 ...
## $ result       : Factor w/ 4 levels "fail math","fail portuguese",...: 1 1 3 3 3 3 3 1 3 3 ...
## $ alcohol      : num  1 1 2.29 1 1.29 1.29 1 1 1 1 ...
```

```
plot(data_profiling[variables_numericas])
```



```
cor(data_profiling[variables_numericas])
```

```
##          age      failures      absences      g1      g2
## age      1.0000000  0.28236357  0.15319565 -0.12412125 -0.11947474
## failures 0.2823636  1.00000000  0.09999785 -0.37417487 -0.37717218
## absences 0.1531956  0.09999785  1.00000000 -0.09242463 -0.08933169
## g1      -0.1241212 -0.37417487 -0.09242463  1.00000000  0.85873875
## g2      -0.1194747 -0.37717218 -0.08933169  0.85873875  1.00000000
## g3      -0.1252824 -0.38314528 -0.04567058  0.80914172  0.91074316
## alcohol  0.1327481  0.12410073  0.14877659 -0.16222869 -0.14306963
##          g3      alcohol
## age      -0.12528243  0.1327481
## failures -0.38314528  0.1241007
## absences -0.04567058  0.1487766
## g1       0.80914172 -0.1622287
## g2       0.91074316 -0.1430696
## g3       1.00000000 -0.1367715
## alcohol -0.13677155  1.0000000
```

```
# se observan correlaciones altas entre las notas g1, g2 y g3.
```

```
# se define la variable éxito / fracaso nbing3 como numérica
data_profiling$nbing3 <- ifelse(data_profiling$g3>=10, 1, 0)
```

```
# se eliminan aquellas variables que no deben participar
data_profiling <- subset(data_profiling, select = -c(g1,g2,g3,file,subject,student_key,result,bing3) )
```

```
head(data_profiling)
```

```
##          school    sex age address          famsize          pstatus
## 1 Gabriel Pereira female 18  urban    greater than 3  living apart
## 2 Gabriel Pereira female 17  urban    greater than 3  living together
## 3 Gabriel Pereira female 15  urban less or equal to 3 living together
## 4 Gabriel Pereira female 15  urban    greater than 3  living together
## 5 Gabriel Pereira female 16  urban    greater than 3  living together
## 6 Gabriel Pereira  male 16  urban less or equal to 3 living together
##          medu          fedu      mjob      fjob      reason guardian
## 1  higher          higher at_home teacher    course  mother
## 2  primary          primary at_home  other    course  father
## 3  primary          primary at_home  other    other   mother
## 4  higher 5th-9th grade health services    home   mother
## 5 secondary          secondary other    other    home   father
## 6  higher          secondary services other reputation mother
##          traveltime      studytime failures schoolsup famsup paid activities
## 1 15 to 30 min.  2 to 5 hours      0      yes    no    no    no
## 2   <15 min.  2 to 5 hours      0      no    yes    no    no
## 3   <15 min.  2 to 5 hours      3      yes    no    yes    no
## 4   <15 min.  5 to 10 hours     0      no    yes    yes    yes
## 5   <15 min.  2 to 5 hours      0      no    yes    yes    no
## 6   <15 min.  2 to 5 hours      0      no    yes    yes    yes
##          nursery higher internet romantic      famrel freetime goout      dalc
## 1  yes  yes      no      no      good    medium  high very low
## 2  no  yes      yes      no excellent  medium  medium very low
## 3  yes  yes      yes      no      good    medium  low   low
## 4  yes  yes      yes      yes    fair     low    low very low
## 5  yes  yes      no      no      good    medium  low very low
## 6  yes  yes      yes      no excellent  high   low very low
##          walc      health absences alcohol nbing3
## 1 very low    fair      6      1.00    0
## 2 very low    fair      4      1.00    0
## 3  medium     fair     10     2.29    1
## 4 very low very good      2      1.00    1
## 5  low very good      4      1.29    1
## 6  low very good     10     1.29    1
```

```
names(data_profiling)
```

```
## [1] "school"      "sex"         "age"         "address"     "famsize"
## [6] "pstatus"     "medu"        "fedu"        "mjob"        "fjob"
## [11] "reason"      "guardian"    "traveltime"  "studytime"   "failures"
## [16] "schoolsup"  "famsup"      "paid"        "activities"  "nursery"
## [21] "higher"     "internet"    "romantic"    "famrel"      "freetime"
## [26] "goout"      "dalc"        "walc"        "health"      "absences"
## [31] "alcohol"    "nbing3"
```

```
# la variable éxito / fracaso es data_profiling$nbing3
```

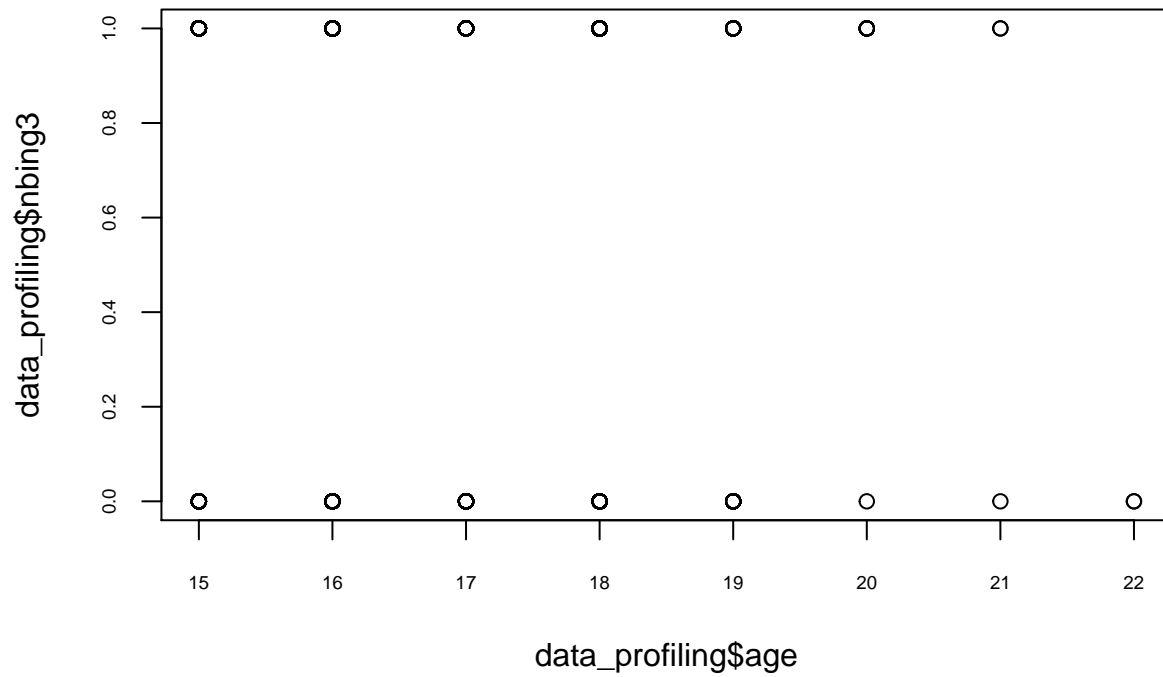
```
dim(data_profiling)
```



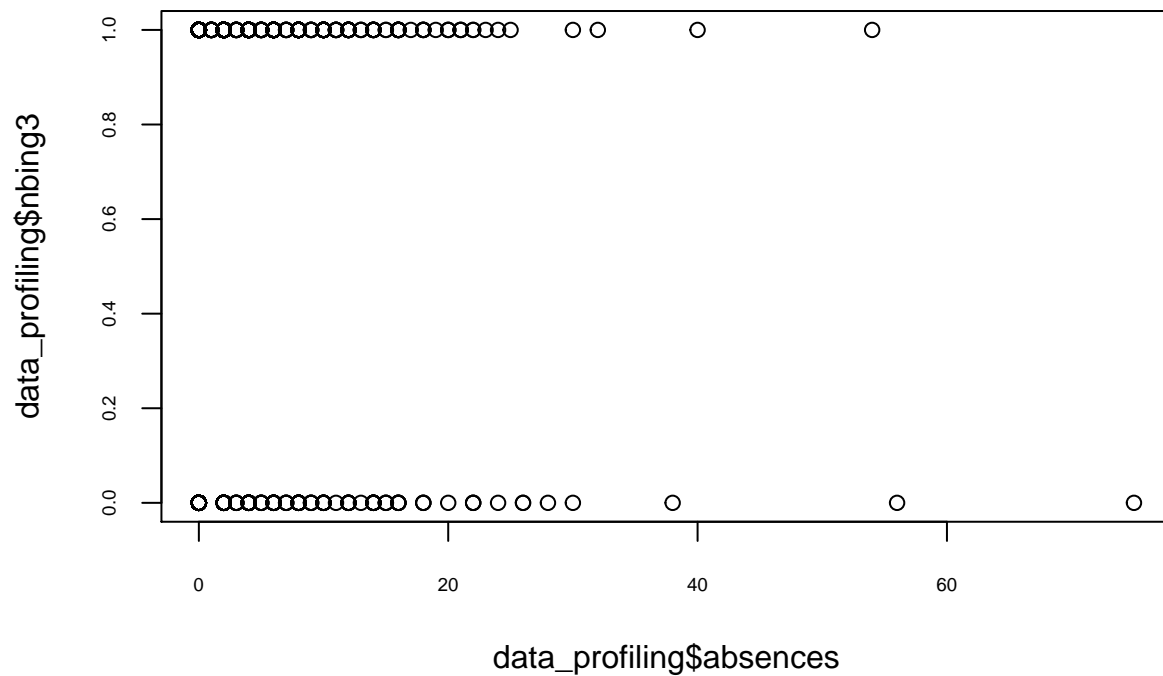
```
## [1] 1044 32
```

```
# se observa relación?
```

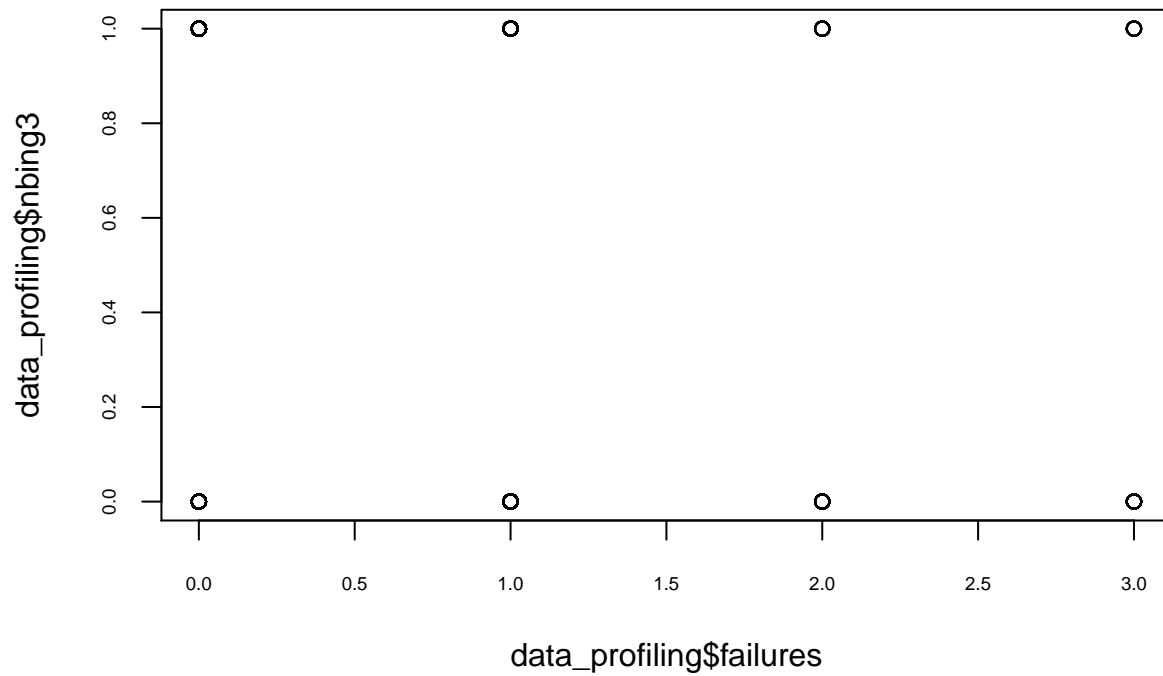
```
plot(data_profiling$age,data_profiling$nbing3, cex.axis=0.6)
```



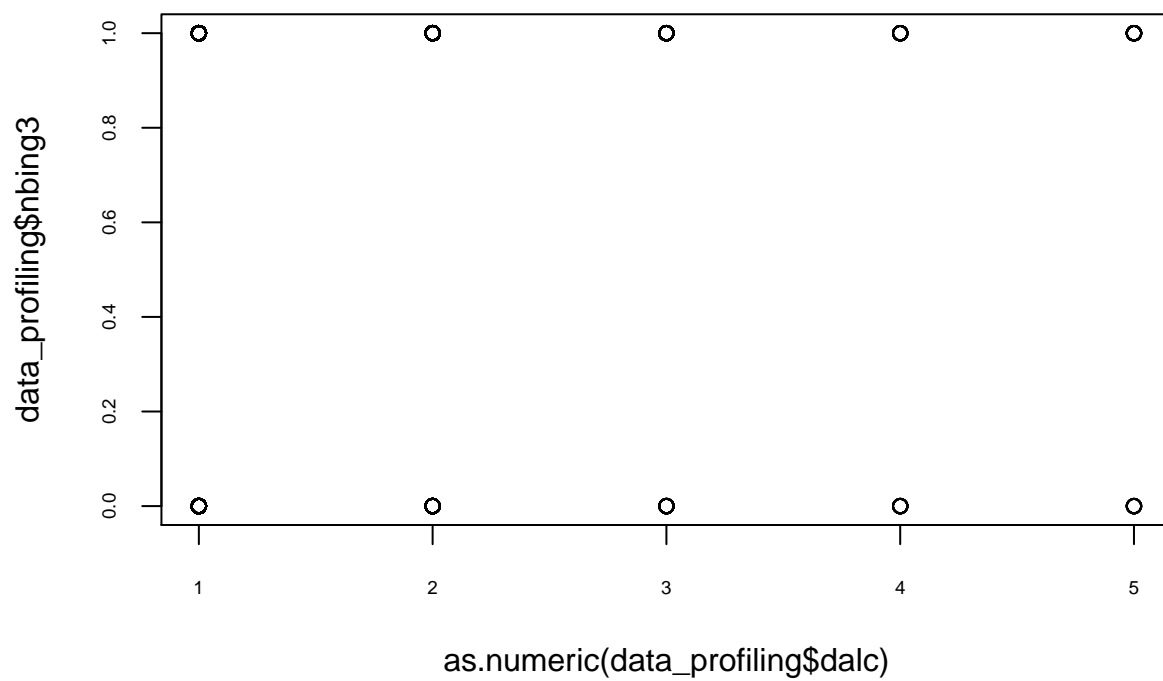
```
plot(data_profiling$absences,data_profiling$nbing3, cex.axis=0.6)
```



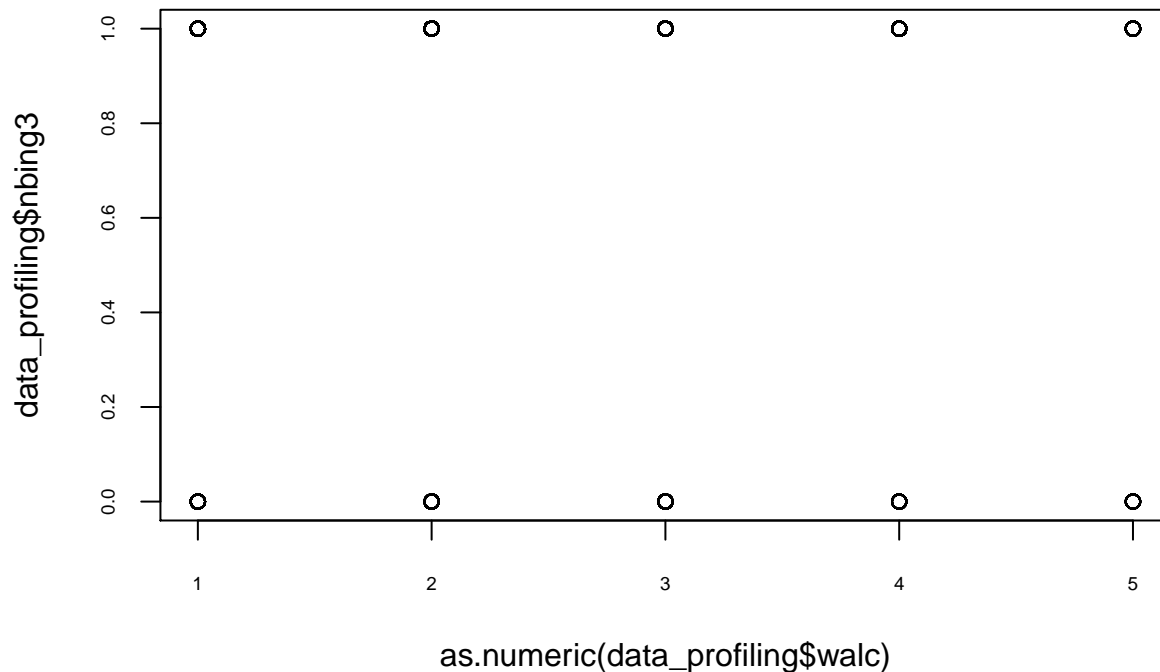
```
plot(data_profiling$failures,data_profiling$nbing3, cex.axis=0.6)
```



```
plot(as.numeric(data_profiling$dalc),data_profiling$nbings3, cex.axis=0.6)
```



```
plot(as.numeric(data_profiling$walc),data_profiling$nbings3, cex.axis=0.6)
```



Análisis con la función `catdes` de `FactoMineR`

Usamos el paquete **FactoMineR** para hacer el análisis completo con todas las variables categóricas y numéricas que pueden explicar el éxito.

<https://cran.r-project.org/web/packages/FactoMineR/FactoMineR.pdf> <http://factominer.free.fr/factosbest/categories-description.html>

```
#install.packages("FactoMineR")
#install.CRAN("FactoMineR")

library(FactoMineR)

data_profiling<-transform(data_profiling,nbing3=as.factor(nbing3))

# la variable a categorizar es la última por eso se indica como dim(data_profiling)[2]
rm(catdesI)
catdesI<-catdes(data_profiling,num.var=dim(data_profiling)[2],proba = 0.05)
catdesI
```

```
## $test.chi2
##           p.value df
## higher  1.693028e-12  1
## school  1.987478e-05  1
## goout   2.761440e-04  4
## fedu    4.500810e-04  4
## medu    1.292969e-03  4
## studytime 4.155114e-03  3
## guardian 6.064421e-03  2
## romantic 1.115311e-02  1
## dalc    1.469728e-02  4
## reason  1.686899e-02  3
```

```

## address 1.720238e-02 1
## paid 3.471492e-02 1
##
## $category
## $category$`0`
## Cla/Mod Mod/Cla Global p.value
## higher=no 51.68539 20.00000 8.524904 1.278896e-10
## school=Mousinho da Silveira 31.25000 36.95652 26.053640 3.313029e-05
## fedu=primary 31.25000 34.78261 24.521073 6.872376e-05
## goout=very high 32.51534 23.04348 15.613027 7.204126e-04
## medu=primary 30.69307 26.95652 19.348659 1.339017e-03
## studytime=<2 hours 27.44479 37.82609 30.363985 6.054407e-03
## romantic=yes 26.41509 42.60870 35.536398 1.202256e-02
## guardian=other 34.24658 10.86957 6.992337 1.299985e-02
## reason=course 25.81395 48.26087 41.187739 1.421787e-02
## address=rural 27.01754 33.47826 27.298851 1.898112e-02
## dalc=low 28.06122 23.91304 18.773946 2.699476e-02
## paid=yes 27.27273 26.08696 21.072797 3.810241e-02
## paid=no 20.63107 73.91304 78.927203 3.810241e-02
## fedu=higher 16.96429 16.52174 21.455939 3.647090e-02
## freetime=low 15.78947 11.73913 16.379310 2.810519e-02
## famrel=good 19.14062 42.60870 49.042146 2.730465e-02
## goout=medium 17.91045 26.08696 32.088123 2.611151e-02
## guardian=father 16.87243 17.82609 23.275862 2.473245e-02
## address=urban 20.15810 66.52174 72.701149 1.898112e-02
## romantic=no 19.61367 57.39130 64.463602 1.202256e-02
## goout=low 16.12903 17.39130 23.754789 8.938933e-03
## studytime=5 to 10 hours 14.19753 10.00000 15.517241 6.993573e-03
## reason=reputation 15.72581 16.95652 23.754789 5.120815e-03
## dalc=very low 19.39477 61.30435 69.636015 2.213793e-03
## medu=higher 15.35948 20.43478 29.310345 6.329086e-04
## school=Gabriel Pereira 18.78238 63.04348 73.946360 3.313029e-05
## higher=yes 19.26702 80.00000 91.475096 1.278896e-10
## v.test
## higher=no 6.429660
## school=Mousinho da Silveira 4.150809
## fedu=primary 3.980661
## goout=very high 3.381691
## medu=primary 3.207484
## studytime=<2 hours 2.744820
## romantic=yes 2.511482
## guardian=other 2.483773
## reason=course 2.451711
## address=rural 2.345902
## dalc=low 2.211594
## paid=yes 2.073751
## paid=no -2.073751
## fedu=higher -2.091638
## freetime=low -2.195815
## famrel=good -2.207135
## goout=medium -2.224549
## guardian=father -2.245556
## address=urban -2.345902
## romantic=no -2.511482

```

```

## goout=low -2.614381
## studytime=5 to 10 hours -2.697150
## reason=reputation -2.799334
## dalc=very low -3.059943
## medu=higher -3.417105
## school=Gabriel Pereira -4.150809
## higher=yes -6.429660
##
## $category$`1`
## Cla/Mod Mod/Cla Global p.value
## higher=yes 80.73298 94.717445 91.475096 1.278896e-10
## school=Gabriel Pereira 81.21762 77.027027 73.946360 3.313029e-05
## medu=higher 84.64052 31.818182 29.310345 6.329086e-04
## dalc=very low 80.60523 71.990172 69.636015 2.213793e-03
## reason=reputation 84.27419 25.675676 23.754789 5.120815e-03
## studytime=5 to 10 hours 85.80247 17.076167 15.517241 6.993573e-03
## goout=low 83.87097 25.552826 23.754789 8.938933e-03
## romantic=no 80.38633 66.461916 64.463602 1.202256e-02
## address=urban 79.84190 74.447174 72.701149 1.898112e-02
## guardian=father 83.12757 24.815725 23.275862 2.473245e-02
## goout=medium 82.08955 33.783784 32.088123 2.611151e-02
## famrel=good 80.85938 50.859951 49.042146 2.730465e-02
## freetime=low 84.21053 17.690418 16.379310 2.810519e-02
## fedu=higher 83.03571 22.850123 21.455939 3.647090e-02
## paid=no 79.36893 80.343980 78.927203 3.810241e-02
## paid=yes 72.72727 19.656020 21.072797 3.810241e-02
## dalc=low 71.93878 17.321867 18.773946 2.699476e-02
## address=rural 72.98246 25.552826 27.298851 1.898112e-02
## reason=course 74.18605 39.189189 41.187739 1.421787e-02
## guardian=other 65.75342 5.896806 6.992337 1.299985e-02
## romantic=yes 73.58491 33.538084 35.536398 1.202256e-02
## studytime=<2 hours 72.55521 28.255528 30.363985 6.054407e-03
## medu=primary 69.30693 17.199017 19.348659 1.339017e-03
## goout=very high 67.48466 13.513514 15.613027 7.204126e-04
## fedu=primary 68.75000 21.621622 24.521073 6.872376e-05
## school=Mousinho da Silveira 68.75000 22.972973 26.053640 3.313029e-05
## higher=no 48.31461 5.282555 8.524904 1.278896e-10
## v.test
## higher=yes 6.429660
## school=Gabriel Pereira 4.150809
## medu=higher 3.417105
## dalc=very low 3.059943
## reason=reputation 2.799334
## studytime=5 to 10 hours 2.697150
## goout=low 2.614381
## romantic=no 2.511482
## address=urban 2.345902
## guardian=father 2.245556
## goout=medium 2.224549
## famrel=good 2.207135
## freetime=low 2.195815
## fedu=higher 2.091638
## paid=no 2.073751
## paid=yes -2.073751

```

```

## dalc=low -2.211594
## address=rural -2.345902
## reason=course -2.451711
## guardian=other -2.483773
## romantic=yes -2.511482
## studytime=<2 hours -2.744820
## medu=primary -3.207484
## goout=very high -3.381691
## fedu=primary -3.980661
## school=Mousinho da Silveira -4.150809
## higher=no -6.429660
##
##
## $quanti.var
## Eta2 P-value
## failures 0.134823525 1.140091e-34
## age 0.018030000 1.342167e-05
## absences 0.014194706 1.139821e-04
## alcohol 0.008673156 2.594825e-03
##
## $quanti
## $quanti$`0`
## v.test Mean in category Overall mean sd in category Overall sd
## failures 11.858370 0.7173913 0.2643678 0.9705006 0.6558275
## age 4.336507 17.0391304 16.7260536 1.3329402 1.2393807
## absences 3.847737 5.8260870 4.4348659 8.7000858 6.2070417
## alcohol 3.007674 1.8833913 1.7210632 1.0154608 0.9265243
## p.value
## failures 1.947294e-32
## age 1.447650e-05
## absences 1.192141e-04
## alcohol 2.632555e-03
##
## $quanti$`1`
## v.test Mean in category Overall mean sd in category
## alcohol -3.007674 1.6751966 1.7210632 0.8944814
## absences -3.847737 4.0417690 4.4348659 5.2273355
## age -4.336507 16.6375921 16.7260536 1.1968887
## failures -11.858370 0.1363636 0.2643678 0.4594939
## Overall sd p.value
## alcohol 0.9265243 2.632555e-03
## absences 6.2070417 1.192141e-04
## age 1.2393807 1.447650e-05
## failures 0.6558275 1.947294e-32
##
##
## attr("class")
## [1] "catdes" "list "
plot(catdesI)
plot(data_profiling$nbing3, data_profiling$higher, xlab="grupo", ylab="higher", cex.axis=0.6, main="Infl")
plot(data_profiling$nbing3,data_profiling$school, xlab="grupo", ylab="school", cex.axis=0.6, main="Infl")

```

```

plot(data_profiling$nbing3,data_profiling$goout, xlab="grupo", ylab="goout", cex.axis=0.6, main="Influencia de goout")
plot(data_profiling$nbing3,data_profiling$medu, xlab="grupo", ylab="medu", cex.axis=0.6, main="Influencia de medu")
plot(data_profiling$nbing3,data_profiling$fedu, xlab="grupo", ylab="fedu", cex.axis=0.6, main="Influencia de fedu")
save.image("/home/j82/Dropbox/UOC_Master/TFM/1_R/myworkspace_TFM.RData")

```

## Árboles de decisión

```

# Carga C5 .0
library(C50)

data_arbol<-data_profiling
names(data_arbol)

## [1] "school"      "sex"         "age"         "address"     "famsize"
## [6] "pstatus"     "medu"        "fedu"        "mjob"        "fjob"
## [11] "reason"      "guardian"    "traveltime" "studytime"   "failures"
## [16] "schoolsup"   "famsup"      "paid"        "activities"  "nursery"
## [21] "higher"     "internet"    "romantic"    "famrel"      "freetime"
## [26] "goout"      "dalc"        "walc"        "health"      "absences"
## [31] "alcohol"    "nbing3"

# construimos un primer árbol con las variables numericas y las variables binarias
str(data_arbol)

## 'data.frame':  1044 obs. of  32 variables:
## $ school      : Factor w/ 2 levels "Gabriel Pereira",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ sex         : Factor w/ 2 levels "female","male": 1 1 1 1 1 2 2 1 2 2 ...
## $ age         : int  18 17 15 15 16 16 16 17 15 15 ...
## $ address     : Factor w/ 2 levels "rural","urban": 2 2 2 2 2 2 2 2 2 2 ...
## $ famsize     : Factor w/ 2 levels "greater than 3",...: 1 1 2 1 1 2 2 1 2 1 ...
## $ pstatus     : Factor w/ 2 levels "living apart",...: 1 2 2 2 2 2 2 1 1 2 ...
## $ medu        : Factor w/ 5 levels "none","primary",...: 5 2 2 5 4 5 3 5 4 4 ...
## $ fedu        : Factor w/ 5 levels "none","primary",...: 5 2 2 3 4 4 3 5 3 5 ...
## $ mjob        : Factor w/ 5 levels "at_home","health",...: 1 1 1 2 3 4 3 3 4 3 ...
## $ fjob        : Factor w/ 5 levels "at_home","health",...: 5 3 3 4 3 3 3 5 3 3 ...
## $ reason      : Factor w/ 4 levels "course","home",...: 1 1 3 2 2 4 2 2 2 2 ...
## $ guardian    : Factor w/ 3 levels "father","mother",...: 2 1 2 2 1 2 2 2 2 2 ...
## $ traveltime  : Factor w/ 4 levels "<15 min.,"15 to 30 min.",...: 2 1 1 1 1 1 1 2 1 1 ...
## $ studytime   : Factor w/ 4 levels "<2 hours","2 to 5 hours",...: 2 2 2 3 2 2 2 2 2 2 ...
## $ failures    : int  0 0 3 0 0 0 0 0 0 0 ...
## $ schoolsup   : Factor w/ 2 levels "no","yes": 2 1 2 1 1 1 1 2 1 1 ...
## $ famsup      : Factor w/ 2 levels "no","yes": 1 2 1 2 2 2 1 2 2 2 ...
## $ paid        : Factor w/ 2 levels "no","yes": 1 1 2 2 2 2 1 1 2 2 ...
## $ activities  : Factor w/ 2 levels "no","yes": 1 1 1 2 1 2 1 1 1 2 ...
## $ nursery     : Factor w/ 2 levels "no","yes": 2 1 2 2 2 2 2 2 2 2 ...
## $ higher      : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2 2 ...
## $ internet    : Factor w/ 2 levels "no","yes": 1 2 2 2 1 2 2 1 2 2 ...
## $ romantic    : Factor w/ 2 levels "no","yes": 1 1 1 2 1 1 1 1 1 1 ...

```

```
## $ famrel      : Factor w/ 5 levels "very bad","bad",...: 4 5 4 3 4 5 4 4 4 5 ...
## $ freetime   : Factor w/ 5 levels "very low","low",...: 3 3 3 2 3 4 4 1 2 5 ...
## $ goout      : Factor w/ 5 levels "very low","low",...: 4 3 2 2 2 4 4 2 1 ...
## $ dalc       : Factor w/ 5 levels "very low","low",...: 1 1 2 1 1 1 1 1 1 ...
## $ walc       : Factor w/ 5 levels "very low","low",...: 1 1 3 1 2 2 1 1 1 1 ...
## $ health     : Factor w/ 5 levels "very bad","bad",...: 3 3 3 5 5 5 3 1 1 5 ...
## $ absences   : int  6 4 10 2 4 10 0 6 0 0 ...
## $ alcohol    : num  1 1 2.29 1 1.29 1.29 1 1 1 1 ...
## $ nbing3     : Factor w/ 2 levels "0","1": 1 1 2 2 2 2 1 2 2 ...
```

```
names(data_arbol)
```

```
## [1] "school"      "sex"          "age"          "address"      "famsize"
## [6] "pstatus"     "medu"         "fedu"         "mjob"         "fjob"
## [11] "reason"      "guardian"     "traveltime"  "studytime"   "failures"
## [16] "schoolsup"  "famsup"       "paid"        "activities"   "nursery"
## [21] "higher"     "internet"     "romantic"    "famrel"       "freetime"
## [26] "goout"      "dalc"         "walc"        "health"       "absences"
## [31] "alcohol"    "nbing3"
```

```
length(data_arbol)
```

```
## [1] 32
```

```
# Dividimos el fichero en 70% entreno y 30% validación #
```

```
set.seed(1234)
```

```
ind <- sample(2, nrow(data_arbol), replace=TRUE, prob=c(0.7, 0.3))
```

```
# Asignamos a X las variables para construir el arbol
```

```
trainX <- data_arbol[ind==1,c(1:31)]
```

```
testX <- data_arbol[ind==2,c(1:31)]
```

```
train <- data_arbol[ind==1,c(1:32)]
```

```
test <- data_arbol[ind==2,c(1:32)]
```

```
# verificamos las variables implicadas y las dimensiones de los objetos
```

```
names(trainX)
```

```
## [1] "school"      "sex"          "age"          "address"      "famsize"
## [6] "pstatus"     "medu"         "fedu"         "mjob"         "fjob"
## [11] "reason"      "guardian"     "traveltime"  "studytime"   "failures"
## [16] "schoolsup"  "famsup"       "paid"        "activities"   "nursery"
## [21] "higher"     "internet"     "romantic"    "famrel"       "freetime"
## [26] "goout"      "dalc"         "walc"        "health"       "absences"
## [31] "alcohol"
```

```
names(testX)
```

```
## [1] "school"      "sex"          "age"          "address"      "famsize"
## [6] "pstatus"     "medu"         "fedu"         "mjob"         "fjob"
## [11] "reason"      "guardian"     "traveltime"  "studytime"   "failures"
## [16] "schoolsup"  "famsup"       "paid"        "activities"   "nursery"
## [21] "higher"     "internet"     "romantic"    "famrel"       "freetime"
## [26] "goout"      "dalc"         "walc"        "health"       "absences"
```



```

## [31] "alcohol"
names(train)

## [1] "school"      "sex"         "age"         "address"     "famsize"
## [6] "pstatus"    "medu"        "fedu"        "mjob"        "fjob"
## [11] "reason"     "guardian"    "traveltime"  "studytime"   "failures"
## [16] "schoolsup"  "famsup"      "paid"        "activities"  "nursery"
## [21] "higher"     "internet"    "romantic"    "famrel"      "freetime"
## [26] "goout"      "dalc"        "walc"        "health"      "absences"
## [31] "alcohol"    "nbing3"

names(test)

## [1] "school"      "sex"         "age"         "address"     "famsize"
## [6] "pstatus"    "medu"        "fedu"        "mjob"        "fjob"
## [11] "reason"     "guardian"    "traveltime"  "studytime"   "failures"
## [16] "schoolsup"  "famsup"      "paid"        "activities"  "nursery"
## [21] "higher"     "internet"    "romantic"    "famrel"      "freetime"
## [26] "goout"      "dalc"        "walc"        "health"      "absences"
## [31] "alcohol"    "nbing3"

dim(trainX)

## [1] 729 31

dim(testX)

## [1] 315 31

dim(train)

## [1] 729 32

dim(test)

## [1] 315 32

# Asignamos a Y la variable target

trainY <- data_arbol[ind==1,32]
testY <- data_arbol[ind==2,32]

# comprobamos que la proporción de las clases es similar en los dos grupos
table(trainY)/length(trainY)

## trainY
##      0      1
## 0.2208505 0.7791495

table(testY)/length(testY)

## testY
##      0      1
## 0.2190476 0.7809524

# Generamos el modelo y observamos el resultado
modell <- C50::C5.0(trainX,trainY)
summary ( modell )

##

```

```

## Call:
## C5.0.default(x = trainX, y = trainY)
##
##
## C5.0 [Release 2.07 GPL Edition]      Sat Jun 17 13:53:35 2017
## -----
##
## Class specified by attribute `outcome'
##
## Read 729 cases (32 attributes) from undefined.data
##
## Decision tree:
##
## failures <= 0:
## :...higher = no:
## :   :...nursery = no: 1 (6)
## :   :   nursery = yes:
## :   :     :...goout in {low,high}: 0 (6)
## :   :     :   goout in {very low,medium,very high}:
## :   :     :     :...guardian in {father,other}: 1 (11)
## :   :     :       guardian = mother:
## :   :     :         :...alcohol <= 1.29: 1 (3)
## :   :     :           alcohol > 1.29: 0 (5)
## :   higher = yes:
## :     :...paid = no: 1 (436/49)
## :     :   paid = yes:
## :     :     :...guardian = other: 0 (5/1)
## :     :     :   guardian in {father,mother}:
## :     :     :     :...schoolsup = no: 1 (111/21)
## :     :     :       schoolsup = yes:
## :     :     :         :...famsup = no: 1 (2)
## :     :     :           famsup = yes: 0 (16/7)
## failures > 0:
## :...famrel = very bad: 0 (3)
##   famrel = bad:
##     :...health in {very bad,bad,fair,good}: 0 (8/1)
##     :   health = very good: 1 (3)
##   famrel = excellent:
##     :...schoolsup = yes: 1 (2)
##     :   schoolsup = no:
##     :     :...fjob = at_home: 1 (2)
##     :     :   fjob in {health,other,services,teacher}: 0 (29/7)
##   famrel = fair:
##     :...reason = home: 1 (5)
##     :   reason in {course,other,reputation}:
##     :     :...studytime = >10 hours: 0 (0)
##     :     :   studytime = 5 to 10 hours: 1 (2)
##     :     :   studytime in {<2 hours,2 to 5 hours}:
##     :     :     :...health in {very bad,bad,good,very good}: 0 (13/1)
##     :     :     :   health = fair: 1 (2)
##   famrel = good:
##     :...freetime in {very low,low,very high}: 1 (18/3)
##     :   freetime = medium:
##     :     :...famsize = greater than 3: 0 (9)

```

```

##      :  famsize = less or equal to 3: 1 (6/2)
##      freetime = high:
##      :...sex = female: 1 (5)
##      sex = male:
##      :...fjob = at_home: 0 (2)
##      fjob in {health,services,teacher}: 1 (2)
##      fjob = other:
##      :...fedu = primary: 1 (11/2)
##      fedu in {none,5th-9th grade,secondary,higher}: 0 (6/1)
##
##

```

```
## Evaluation on training data (729 cases):
```

```

##      Decision Tree
##      -----
##      Size      Errors
##
##      28   95(13.0%)  <<
##
##      (a)  (b)  <-classified as
##      ----  ----
##      84   77   (a): class 0
##      18  550   (b): class 1
##

```

```
## Attribute usage:
```

```

## 100.00% failures
## 82.44% higher
## 78.19% paid
## 22.22% schoolsup
## 20.99% guardian
## 17.56% famrel
## 8.09% freetime
## 7.13% fjob
## 4.25% nursery
## 3.57% sex
## 3.57% health
## 3.43% goout
## 3.02% reason
## 2.47% famsup
## 2.33% fedu
## 2.33% studytime
## 2.06% famsize
## 1.10% alcohol
##
##

```

```
## Time: 0.0 secs
```

```
C5imp(model1)
```

```

##      Overall
## failures 100.00
## higher   82.44

```

```

## paid          78.19
## schoolsup     22.22
## guardian     20.99
## famrel       17.56
## freetime      8.09
## fjob         7.13
## nursery      4.25
## sex          3.57
## health       3.57
## goout        3.43
## reason       3.02
## famsup       2.47
## fedu         2.33
## studytime    2.33
## famsize      2.06
## alcohol      1.10
## school       0.00
## age          0.00
## address      0.00
## pstatus     0.00
## medu        0.00
## mjob        0.00
## traveltime  0.00
## activities   0.00
## internet    0.00
## romantic    0.00
## dalc        0.00
## walc        0.00
## absences    0.00

```

```
C5imp(modell1, metric = "splits")
```

```

##          Overall
## fjob      9.090909
## guardian  9.090909
## health    9.090909
## schoolsup  9.090909
## alcohol   4.545455
## failures  4.545455
## famrel    4.545455
## famsize   4.545455
## famsup    4.545455
## fedu      4.545455
## freetime  4.545455
## goout     4.545455
## higher    4.545455
## nursery   4.545455
## paid      4.545455
## reason    4.545455
## sex       4.545455
## studytime 4.545455
## school    0.000000
## age       0.000000
## address   0.000000
## pstatus   0.000000

```

```
## medu      0.000000
## mjob      0.000000
## traveltime 0.000000
## activities 0.000000
## internet  0.000000
## romantic  0.000000
## dalc      0.000000
## walc      0.000000
## absences  0.000000

# Calculo de la capacidad predictiva del modelo a partir del conjunto de test
p1 <- predict (modell1, testX, type = "class")
acc1 <- sum ( p1 == testY ) / length ( p1 )
acc1*100
```

```
## [1] 79.36508
```

```
confusion_matrix_1 <- table(Real=testY, "Predicción"=p1)
print(confusion_matrix_1)
```

```
##      Predicción
## Real   0   1
##      0  23  46
##      1  19 227
```

```
# este modelo tiene un acierto global del 79,36%
```

```
# veamos si con boosting mejora
```

```
# Generamos el modelo y observamos el resultado
model2 <- C5.0::C5.0(trainX,trainY, trials=10)
summary ( model2 )
```

```
##
## Call:
## C5.0.default(x = trainX, y = trainY, trials = 10)
##
##
## C5.0 [Release 2.07 GPL Edition]      Sat Jun 17 13:53:35 2017
## -----
##
## Class specified by attribute `outcome'
##
## Read 729 cases (32 attributes) from undefined.data
##
## ----- Trial 0: -----
##
## Decision tree:
##
## failures <= 0:
## :...higher = no:
## :   :...nursery = no: 1 (6)
## :   :   nursery = yes:
## :   :   :...goout in {low,high}: 0 (6)
## :   :   :   goout in {very low,medium,very high}:
```

```

## : : :...guardian in {father,other}: 1 (11)
## : : guardian = mother:
## : : :...alcohol <= 1.29: 1 (3)
## : : alcohol > 1.29: 0 (5)
## : higher = yes:
## : :...paid = no: 1 (436/49)
## : paid = yes:
## : :...guardian = other: 0 (5/1)
## : guardian in {father,mother}:
## : :...schoolsup = no: 1 (111/21)
## : schoolsup = yes:
## : :...famsup = no: 1 (2)
## : famsup = yes: 0 (16/7)
## failures > 0:
## :...famrel = very bad: 0 (3)
## famrel = bad:
## :...health in {very bad,bad,fair,good}: 0 (8/1)
## : health = very good: 1 (3)
## famrel = excellent:
## :...schoolsup = yes: 1 (2)
## : schoolsup = no:
## : :...fjob = at_home: 1 (2)
## : fjob in {health,other,services,teacher}: 0 (29/7)
## famrel = fair:
## :...reason = home: 1 (5)
## : reason in {course,other,reputation}:
## : :...studytime = >10 hours: 0 (0)
## : studytime = 5 to 10 hours: 1 (2)
## : studytime in {<2 hours,2 to 5 hours}:
## : :...health in {very bad,bad,good,very good}: 0 (13/1)
## : health = fair: 1 (2)
## famrel = good:
## :...freetime in {very low,low,very high}: 1 (18/3)
## freetime = medium:
## :...famsize = greater than 3: 0 (9)
## : famsize = less or equal to 3: 1 (6/2)
## freetime = high:
## :...sex = female: 1 (5)
## sex = male:
## :...fjob = at_home: 0 (2)
## fjob in {health,services,teacher}: 1 (2)
## fjob = other:
## :...fedu = primary: 1 (11/2)
## fedu in {none,5th-9th grade,secondary,higher}: 0 (6/1)
##
## ----- Trial 1: -----
##
## Decision tree:
##
## higher = no:
## :...failures <= 2: 1 (51.4/22.9)
## : failures > 2: 0 (10.3/0.8)
## higher = yes:
## :...absences > 13:

```

```

##      :...sex = female: 1 (33.4/14.4)
##      :   sex = male: 0 (23.2/7.1)
##      absences <= 13:
##      :...school = Gabriel Pereira: 1 (447.9/120.5)
##      school = Mousinho da Silveira:
##      :...studytime = 5 to 10 hours: 1 (15.8/0.8)
##      studytime in {<2 hours,2 to 5 hours,>10 hours}:
##      :...alcohol <= 1.29:
##      :...absences <= 4: 1 (49.8/8.8)
##      :   absences > 4: 0 (12.8/3.9)
##      alcohol > 1.29:
##      :...famsize = greater than 3: 0 (62.3/20.5)
##      famsize = less or equal to 3: 1 (22.2/8.8)
##
## ----- Trial 2: -----
##
## Decision tree:
##
## higher = no: 0 (69.5/25.5)
## higher = yes:
## :...failures > 1: 0 (32.4/10.8)
##   failures <= 1:
##     :...fedu = none: 1 (4.7)
##     fedu = primary:
##       :...studytime in {5 to 10 hours,>10 hours}: 1 (23.4/2.6)
##       :   studytime in {<2 hours,2 to 5 hours}:
##       :     :...pstatus = living apart: 1 (9.4/2.1)
##       :     pstatus = living together:
##       :       :...mjob in {health,teacher}: 0 (14.3/1.3)
##       :       mjob in {at_home,other,services}:
##       :         :...goout = high: 1 (17.6/2.7)
##       :         goout in {very low,low,medium,very high}:
##       :           :...famrel in {very bad,fair,excellent}: 0 (42.5/10.7)
##       :           famrel = bad: 1 (2.7)
##       :           famrel = good:
##       :             :...guardian = father: 1 (6)
##       :             guardian = other: 0 (1.3)
##       :             guardian = mother:
##       :               :...internet = no: 1 (7.3/1.3)
##       :               internet = yes: 0 (29.9/10.6)
##     fedu in {5th-9th grade,secondary,higher}:
##     :...paid = yes:
##       :...health = very bad: 1 (12.7)
##       :   health in {bad,fair,good,very good}:
##       :     :...failures > 0: 1 (8.6/1.3)
##       :     failures <= 0:
##       :       :...pstatus = living apart: 1 (8.1/1.3)
##       :       pstatus = living together:
##       :         :...freetime in {very low,medium,high}: 0 (94/35)
##       :         freetime in {low,very high}: 1 (18.1/2.7)
##     paid = no:
##     :...schoolsup = yes:
##     :   :...address = rural: 1 (5.3)
##     :   address = urban:

```

```

##           :   :...medu = none: 0 (0)
##           :       medu = secondary: 1 (7.4)
##           :       medu in {primary,5th-9th grade,higher}:
##           :           :...absences <= 3: 1 (20.2/8.8)
##           :           absences > 3: 0 (19.4/1.3)
## schoolsup = no:
## :...failures > 0: 1 (33/14.7)
##     failures <= 0:
##     :...goout in {very low,low}:
##     :       :...absences <= 19: 1 (59.5/2.1)
##     :       :   absences > 19: 0 (3.6/0.7)
##     goout in {medium,high,very high}:
##     :...activities = no: 1 (75.9/13.8)
##     activities = yes:
##     :...pstatus = living apart: 1 (4.7)
##     pstatus = living together:
##     :...mjob = health: 1 (10.6)
##     mjob in {at_home,other,services,teacher}:
##     :...walc = low: 1 (8)
##     walc in {very low,medium,high,
##     :         very high}:
##     :...address = rural: 0 (25.2/9.3)
##     address = urban:
##     :...absences > 40: 0 (2.9)
##     absences <= 40: [S1]
##
## SubTree [S1]
##
## traveltime in {15 to 30 min.,30 min to 1 hour,>1 hour}: 1 (8)
## traveltime = <15 min.:
## :...nursery = no: 0 (17.1/5.4)
##     nursery = yes: 1 (25.7/5)
##
## ----- Trial 3: -----
##
## Decision tree:
##
## failures > 0:
## :...dalc in {medium,very high}: 1 (23/5.1)
## :   dalc = high: 0 (10.6/2.3)
## :   dalc = low:
## :     :...age <= 18: 0 (33.3/7.7)
## :     :   age > 18: 1 (2.9)
## :     dalc = very low:
## :       :...goout = very high: 0 (8.9/1.1)
## :       goout in {very low,low,medium,high}:
## :         :...reason = home: 1 (7)
## :         reason in {course,other,reputation}:
## :           :...age > 19: 1 (4.4)
## :           age <= 19:
## :             :...nursery = no: 1 (21.2/8)
## :             nursery = yes: 0 (42.2/14.5)
## failures <= 0:
## :...schoolsup = yes:

```



```

##      :...sex = female: 1 (57.5/20.5)
##      :   sex = male: 0 (22/5.7)
##      schoolsup = no:
##      :...guardian in {father,other}: 1 (131.4/28.8)
##      guardian = mother:
##      :...absences > 16: 0 (18.6/5.2)
##      absences <= 16:
##      :...pstatus = living apart: 1 (33.4/5.6)
##      pstatus = living together:
##      :...school = Mousinho da Silveira:
##      :...studytime in {5 to 10 hours,>10 hours}: 1 (7.9)
##      :   studytime in {<2 hours,2 to 5 hours}:
##      :     :...higher = no: 0 (6.6/1.2)
##      :     higher = yes:
##      :       :...sex = male: 0 (29.8/10.5)
##      :       sex = female:
##      :         :...absences <= 1: 1 (22.3/2.4)
##      :         absences > 1: 0 (24/7.2)
##      school = Gabriel Pereira:
##      :...freetime in {very low,very high}: 1 (12.9)
##      freetime in {low,medium,high}:
##      :...sex = male: 1 (91.6/19.1)
##      sex = female:
##      :...reason = other: 1 (7.5)
##      reason in {course,home,reputation}:
##      :...paid = no: 1 (68.2/22.6)
##      paid = yes:
##      :...nursery = no: 1 (4.8)
##      nursery = yes: 0 (37/12.7)
##
## ----- Trial 4: -----
##
## Decision tree:
##
## failures > 0:
## :...health = very bad: 0 (11.3/0.9)
## :   health in {bad,fair,good,very good}:
## :     :...freetime in {very low,low,very high}: 1 (42.8/14)
## :     freetime = high:
## :       :...failures <= 2: 1 (48.2/16.5)
## :       failures > 2: 0 (8.4/2.2)
## :     freetime = medium:
## :       :...age > 19: 1 (4.5)
## :       age <= 19:
## :         :...schoolsup = no: 0 (35/6.8)
## :         schoolsup = yes: 1 (9.1/2.9)
## failures <= 0:
## :...schoolsup = yes:
## :   :...absences <= 1: 1 (22.4/5.3)
## :   absences > 1:
## :     :...medu in {none,higher}: 0 (16.6/1.7)
## :     medu in {primary,5th-9th grade,secondary}:
## :       :...guardian = father: 0 (11.5/2.4)
## :       guardian in {mother,other}: 1 (33/12.9)

```

```

## schoolsup = no:
## ...famrel = fair:
##   ...fjob in {at_home,teacher}: 0 (17.6/4.2)
##   :   fjob in {health,services}: 1 (24.6/7.3)
##   :   fjob = other:
##   :   ...higher = no: 1 (3)
##   :       higher = yes:
##   :       ...age <= 15: 1 (3.2)
##   :       age > 15:
##   :       ...famsize = greater than 3: 0 (33/11.6)
##   :       famsize = less or equal to 3: 1 (6.6/1.4)
## famrel in {very bad,bad,good,excellent}:
## ...fjob = teacher: 1 (11.5)
##   fjob in {at_home,health,other,services}:
##   ...goout = low: 1 (81.6/13.7)
##   goout in {very low,medium,high,very high}:
##   ...freetime in {very low,low,very high}: 1 (78.1/24.2)
##   freetime = medium:
##   ...goout in {very low,high,very high}: 1 (59.6/20.9)
##   :   goout = medium:
##   :   ...romantic = no: 1 (46.1/6.9)
##   :       romantic = yes: 0 (34.8/14.3)
##   freetime = high:
##   ...dalc in {medium,very high}: 1 (9)
##   dalc in {very low,low,high}:
##   ...health = very bad: 1 (6.5)
##   health in {bad,fair,good,very good}:
##   ...paid = yes: 0 (19.6/3.5)
##   paid = no:
##   ...reason in {course,home,other}: 0 (41.7/18.4)
##   reason = reputation: 1 (9.4)
##
## ----- Trial 5: -----
##
## Decision tree:
##
## failures > 0:
## ...reason in {course,other}: 0 (111.1/43.1)
## :   reason in {home,reputation}: 1 (48.7/16.8)
## failures <= 0:
## ...schoolsup = yes:
##   ...studytime = <2 hours: 1 (12.1/2.5)
##   :   studytime in {2 to 5 hours,5 to 10 hours,>10 hours}: 0 (69.5/28.9)
## schoolsup = no:
## ...fedu = none: 1 (3.5)
##   fedu in {5th-9th grade,higher}:
##   ...health in {very bad,fair,good,very good}: 1 (210.5/49.3)
##   :   health = bad:
##   :   ...school = Gabriel Pereira: 0 (30.5/9.9)
##   :       school = Mousinho da Silveira: 1 (6.7)
##   fedu in {primary,secondary}:
##   ...paid = yes:
##   :   ...freetime in {very low,low}: 1 (8.8)
##   :       freetime in {medium,high,very high}: 0 (42.7/13.2)

```

```

##           paid = no:
##           :...mjob = health: 0 (19.1/4.5)
##           mjob in {at_home,other,services,teacher}:
##           :...famrel = very bad: 0 (9.7/2.3)
##           famrel = bad: 1 (4.7)
##           famrel in {fair,good,excellent}:
##           :...pstatus = living apart: 1 (7.6)
##           pstatus = living together:
##           :...dalc = very high: 0 (4.1)
##           dalc in {very low,low,medium,high}:
##           :...activities = no: 1 (60.6/12.1)
##           activities = yes:
##           :...fjob in {at_home,health,
##           :           teacher}: 1 (4.6)
##           fjob in {other,services}:
##           :...address = rural: 0 (32.6/12.8)
##           address = urban: 1 (41.8/14.3)
##
## ----- Trial 6: -----
##
## Decision tree:
##
## higher = no:
## :...walc = very high: 0 (9.8)
## :   walc in {very low,low,medium,high}:
## :   :...traveltime in {30 min to 1 hour,>1 hour}: 1 (10.2)
## :   traveltime in {<15 min.,15 to 30 min.}:
## :   :...schoolsup = yes: 1 (2.5)
## :   schoolsup = no:
## :   :...failures > 2: 0 (6.2)
## :   failures <= 2:
## :   :...guardian in {father,mother}: 0 (37.7/14.6)
## :   guardian = other: 1 (13.3/1.3)
## higher = yes:
## :...failures > 0:
## :   :...mjob in {at_home,health,services}: 1 (59.6/19.9)
## :   mjob in {other,teacher}: 0 (64.4/24.3)
## failures <= 0:
## :...absences <= 6: 1 (400.4/128.1)
## absences > 6:
## :...paid = yes: 0 (45.8/15.4)
## paid = no:
## :...traveltime in {15 to 30 min.,30 min to 1 hour}: 1 (17.1)
## traveltime in {<15 min.,>1 hour}:
## :...age > 17: 1 (6.1)
## age <= 17:
## :...activities = no: 1 (26.1/7.8)
## activities = yes: 0 (28.8/7.1)
##
## ----- Trial 7: -----
##
## Decision tree:
##
## school = Mousinho da Silveira:

```

```

## :...dalc = high: 0 (13.7/1.5)
## :   dalc in {very low,low,medium,very high}:
## :   :...traveltime = <15 min.:
## :       :...paid = no: 0 (66.2/23.1)
## :       :   paid = yes: 1 (10.2/3.1)
## :       traveltime in {15 to 30 min.,30 min to 1 hour,>1 hour}:
## :       :...medu in {none,secondary,higher}: 1 (33.8/3.3)
## :       medu in {primary,5th-9th grade}:
## :       :...failures > 1: 0 (5.1)
## :       failures <= 1:
## :       :...guardian = father: 1 (16.9)
## :       guardian in {mother,other}:
## :       :...pstatus = living apart: 1 (4)
## :       pstatus = living together:
## :       :...freetime in {very low,high,very high}: 1 (39.3/14.3)
## :       freetime in {low,medium}: 0 (25.2/4.7)
## school = Gabriel Pereira:
## :...failures > 1:
## :   :...famsize = greater than 3: 0 (30/8.7)
## :   :   famsize = less or equal to 3: 1 (14.3/2.8)
## :   failures <= 1:
## :   :...schoolsup = yes:
## :       :...absences <= 1: 1 (21.2/3.6)
## :       :   absences > 1:
## :       :   :...nursery = no: 1 (10.9/3.6)
## :       :       nursery = yes:
## :       :       :...fedu in {none,secondary}: 1 (16.4/3.5)
## :       :       fedu in {primary,5th-9th grade,higher}: 0 (50/14.7)
## :       schoolsup = no:
## :       :...paid = yes:
## :           :...address = rural: 1 (13.5/0.6)
## :           :   address = urban:
## :           :   :...goout in {very low,high}: 0 (26.4/5.2)
## :           :       goout in {low,medium,very high}:
## :           :       :...famsup = no: 1 (7.9)
## :           :           famsup = yes:
## :           :           :...freetime in {very low,low,very high}: 1 (11)
## :           :           freetime in {medium,high}:
## :           :           :...alcohol > 2.71: 1 (5.5)
## :           :           alcohol <= 2.71:
## :           :           :...mjob in {at_home,other,
## :           :               :   teacher}: 0 (46.5/17.9)
## :           :           mjob in {health,services}: 1 (17.3/2.9)
## :       paid = no:
## :       :...age > 18: 1 (14.8)
## :       age <= 18:
## :       :...failures > 0:
## :           :...pstatus = living apart: 1 (7.1/1.5)
## :           :   pstatus = living together: 0 (31.9/12.5)
## :       failures <= 0:
## :       :...absences > 21: 0 (9.2/2.2)
## :       absences <= 21:
## :       :...mjob = services: 1 (25.2)
## :       mjob in {at_home,health,other,teacher}:

```

```

##             :...activities = no: 1 (64.5/8.7)
##             activities = yes:
##             :...pstatus = living apart: 1 (4.3)
##             pstatus = living together:
##             :...guardian = father: 1 (7.9)
##             guardian in {mother,other}:
##             :...address = rural: 0 (26.7/9.2)
##             address = urban:
##             :...nursery = no: 0 (25.1/10.9)
##             nursery = yes: 1 (19.8)
##
## ----- Trial 8: -----
##
## Decision tree:
##
## higher = no:
## :...walc = very high: 0 (11.5)
## :   walc in {very low,low,medium,high}:
## :   :...schoolsup = yes: 1 (2.7)
## :   :   schoolsup = no:
## :   :   :...failures <= 2: 1 (63/22.5)
## :   :   :   failures > 2: 0 (4.5)
## higher = yes:
## :...traveltime = 30 min to 1 hour:
## :   :...failures > 1: 0 (5.2)
## :   :   failures <= 1:
## :   :   :...famsup = no: 1 (7.3/0.5)
## :   :   :   famsup = yes: 0 (28.7/10.9)
## :   traveltime in {<15 min.,15 to 30 min.,>1 hour}:
## :   :...fedu in {none,higher}: 1 (113/25.1)
## :   :   fedu = secondary:
## :   :   :...reason in {other,reputation}: 1 (47.6/5.5)
## :   :   :   reason in {course,home}:
## :   :   :   :...internet = no: 0 (13.1/3.8)
## :   :   :   :   internet = yes:
## :   :   :   :   :...failures <= 1: 1 (70.6/26)
## :   :   :   :   :   failures > 1: 0 (3.8)
## :   fedu = primary:
## :   :...sex = male:
## :   :   :...reason = other: 0 (7.8)
## :   :   :   :   reason in {course,home,reputation}:
## :   :   :   :   :...fjob in {health,teacher}: 1 (0)
## :   :   :   :   :   fjob = at_home: 0 (3.1)
## :   :   :   :   :   fjob in {other,services}:
## :   :   :   :   :   :...famrel = very bad: 1 (0)
## :   :   :   :   :   :   famrel = excellent: 0 (15.8/3.1)
## :   :   :   :   :   :   famrel in {bad,fair,good}:
## :   :   :   :   :   :   :...absences <= 12: 1 (36.6/8.4)
## :   :   :   :   :   :   :   absences > 12: 0 (3.4)
## :   :   sex = female:
## :   :   :...mjob = teacher: 1 (0)
## :   :   :   mjob in {health,services}: 0 (16.2/6.4)
## :   :   :   mjob in {at_home,other}:
## :   :   :   :...paid = yes: 1 (18.2)

```

```

##      :           paid = no:
##      :           :...studytime in {<2 hours,>10 hours}: 1 (15.1)
##      :           studytime in {2 to 5 hours,5 to 10 hours}:
##      :           :...alcohol > 2.71: 0 (3.1)
##      :           alcohol <= 2.71:
##      :           :...internet = no: 0 (14.1/5.5)
##      :           internet = yes: 1 (31.7/8.8)
##      fedu = 5th-9th grade:
##      :...nursery = no: 1 (38.2/3.8)
##      nursery = yes:
##      :...dalc = high: 0 (4.4)
##      dalc = very high: 1 (4.2)
##      dalc in {very low,low,medium}:
##      :...pstatus = living apart: 1 (13.8/1.7)
##      pstatus = living together:
##      :...studytime in {<2 hours,>10 hours}:
##      :...reason in {course,other}: 0 (34.5/6.8)
##      :   reason in {home,reputation}: 1 (10.3/1.5)
##      studytime in {2 to 5 hours,5 to 10 hours}:
##      :...internet = no: 1 (13.1)
##      internet = yes:
##      :...medu in {none,higher}: 1 (13.6)
##      medu = primary: 0 (6.2/0.6)
##      medu in {5th-9th grade,secondary}:
##      :...paid = no: 1 (23.7/4.7)
##      paid = yes: 0 (20.1/8)
##
## ----- Trial 9: -----
##
## Decision tree:
##
## higher = no:
## :...walc = very high: 0 (9.7)
## :   walc in {very low,low,medium,high}:
## :   :...alcohol > 2.86: 1 (7.1)
## :   alcohol <= 2.86:
## :   :...traveltime = >1 hour: 0 (0)
## :   traveltime = 30 min to 1 hour: 1 (5)
## :   traveltime in {<15 min.,15 to 30 min.}:
## :   :...nursery = no: 1 (18.8/6.8)
## :   nursery = yes: 0 (49/12.4)
## higher = yes:
## :...failures > 0:
## :   :...famsup = no: 1 (48/16.6)
## :   famsup = yes:
## :   :...pstatus = living apart: 0 (13.6/1.3)
## :   pstatus = living together:
## :   :...paid = no: 0 (56.3/20.4)
## :   paid = yes: 1 (14.8/3.2)
## failures <= 0:
## :...paid = yes:
## :   :...guardian = other: 0 (10.1/0.6)
## :   guardian in {father,mother}:
## :   :...nursery = no: 1 (14.5)

```

```

##      :      nursery = yes:
##      :      :...health = very bad: 1 (15.8)
##      :      health in {bad,fair,good,very good}:
##      :      :...fedu in {none,primary,5th-9th grade,
##      :      :      higher}: 1 (88.4/29.5)
##      :      fedu = secondary:
##      :      :...fjob in {at_home,health,other,teacher}: 0 (33.1/9.2)
##      :      fjob = services: 1 (3.7)
## paid = no:
##      :...studytime in {5 to 10 hours,>10 hours}: 1 (59.8)
##      studytime in {<2 hours,2 to 5 hours}:
##      :...traveltime = 30 min to 1 hour: 1 (20.3)
##      traveltime in {<15 min.,15 to 30 min.,>1 hour}:
##      :...absences <= 3: 1 (118.3/15.5)
##      absences > 3:
##      :...fjob = teacher: 0 (8.7/0.4)
##      fjob in {at_home,health,other,services}:
##      :...mjob in {at_home,health,teacher}: 1 (32.2/7.8)
##      mjob in {other,services}:
##      :...alcohol <= 2.14: 1 (48/11.2)
##      alcohol > 2.14: 0 (26.7/8.5)
##
##
## Evaluation on training data (729 cases):
##
## Trial      Decision Tree
## -----
##      Size      Errors
##
##      0      28  95(13.0%)
##      1      10 151(20.7%)
##      2      34 177(24.3%)
##      3      25 135(18.5%)
##      4      28 163(22.4%)
##      5      19 181(24.8%)
##      6      14 136(18.7%)
##      7      33 154(21.1%)
##      8      34 148(20.3%)
##      9      22 134(18.4%)
## boost      52( 7.1%)  <<
##
##
##      (a)  (b)  <-classified as
##      ----  ----
##      112  49  (a): class 0
##      3    565 (b): class 1
##
##
## Attribute usage:
##
## 100.00% school
## 100.00% failures
## 100.00% higher
## 99.04% traveltime

```

```

## 97.53% schoolsup
## 95.47% fedu
## 94.10% absences
## 93.96% paid
## 91.77% famrel
## 84.22% pstatus
## 83.26% fjob
## 83.26% guardian
## 82.44% goout
## 78.88% studytime
## 78.88% freetime
## 73.25% mjob
## 69.68% dalc
## 69.14% health
## 64.75% sex
## 62.55% nursery
## 62.00% reason
## 58.02% age
## 51.17% activities
## 46.91% alcohol
## 41.98% address
## 33.33% internet
## 32.24% medu
## 28.40% famsup
## 21.12% walc
## 20.30% famsize
## 11.25% romantic
##
##
## Time: 0.1 secs

```

```

p2 <- predict (model2, testX, type = "class" )
acc2 <- sum ( p2 == testY ) / length ( p2 )
acc2*100

```

```
## [1] 80.31746
```

```

confusion_matrix_2 <- table(Real=testY, "Predicción"=p2)
print(confusion_matrix_2)

```

```

##      Predicción
## Real  0  1
##      0 17 52
##      1 10 236

```

```
C5imp(model2)
```

```

##          Overall
## school      100.00
## failures    100.00
## higher      100.00
## traveltime  99.04
## schoolsup   97.53
## fedu        95.47
## absences    94.10
## paid        93.96

```



```

## famrel      91.77
## pstatus    84.22
## fjob       83.26
## guardian   83.26
## goout      82.44
## studytime  78.88
## freetime   78.88
## mjob       73.25
## dalc       69.68
## health     69.14
## sex        64.75
## nursery    62.55
## reason     62.00
## age        58.02
## activities  51.17
## alcohol    46.91
## address    41.98
## internet   33.33
## medu       32.24
## famsup     28.40
## walc       21.12
## famsize    20.30
## romantic   11.25

```

```
C5imp(model2, metric = "splits")
```

```

## Overall
## failures  8.2568807
## absences  5.9633028
## paid      5.5045872
## pstatus   4.5871560
## schoolsup  4.5871560
## guardian  4.1284404
## nursery   4.1284404
## fjob      3.6697248
## freetime  3.6697248
## higher    3.6697248
## mjob      3.6697248
## reason    3.6697248
## studytime 3.6697248
## goout     3.2110092
## health    3.2110092
## traveltime 3.2110092
## age       2.7522936
## alcohol   2.7522936
## fedu      2.7522936
## sex       2.7522936
## address   2.2935780
## dalc      2.2935780
## famrel    2.2935780
## activities 1.8348624
## famsize   1.8348624
## famsup    1.8348624
## internet  1.8348624
## medu      1.8348624

```

```

## school      1.8348624
## walc       1.8348624
## romantic   0.4587156
# Generalizando, guardamos los valores leyendo directamente de la matriz

TP <- confusion_matrix_2[2,2]
TP
## [1] 236

FP <- confusion_matrix_2[1,2]
FP
## [1] 52

FN <- confusion_matrix_2[2,1]
FN
## [1] 10

TN <- confusion_matrix_2[1,1]
TN
## [1] 17

total <- TP+FP+FN+TN
total
## [1] 315

# Accuracy: Overall, how often is the classifier correct?
#      (TP+TN)/total

accuracy <- (TP+TN)/total
accuracy
## [1] 0.8031746

# Misclassification Rate: Overall, how often is it wrong?
#      (FP+FN)/total
#      equivalent to 1 minus Accuracy
#      also known as "Error Rate"

misclassification_rate <- (FP+FN)/total
misclassification_rate
## [1] 0.1968254

# necesitamos algunas variables más...

confusion_matrix_2

##      Predicción
## Real   0   1
##      0  17  52
##      1  10 236

actual_yes <- sum(confusion_matrix_2[2,])#suma de la segunda fila
actual_yes
## [1] 246

```

```

actual_no <- sum(confusion_matrix_2[1,]) #suma de la primera fila
actual_no

## [1] 69

# True Positive Rate: When it's actually yes, how often does it predict yes?
#     TP/actual yes
#     also known as "Sensitivity" or "Recall"

TP_rate <- TP/actual_yes
TP_rate

## [1] 0.9593496

# False Positive Rate: When it's actually no, how often does it predict yes?
#     FP/actual no

FP_rate <- FP/actual_no
FP_rate

## [1] 0.7536232

# Specificity: When it's actually no, how often does it predict no?
#     TN/actual no
#     equivalent to 1 minus False Positive Rate

specificity <- TN/actual_no
specificity

## [1] 0.2463768

# Precision: When it predicts yes, how often is it correct?
#     TP/predicted yes = 100/110 = 0.91

confusion_matrix_2

##      Predicción
## Real   0   1
##      0  17  52
##      1  10 236

predicted_yes <- sum(confusion_matrix_2[,2]) #suma de la segunda columna
predicted_yes

## [1] 288

precision <- TP/predicted_yes
precision

## [1] 0.8194444

# Prevalence: How often does the yes condition actually occur in our sample?
#     actual yes/total = 105/165 = 0.64

prevalence <- actual_yes/total
prevalence

## [1] 0.7809524

```

```

# validemos ahora con k-fold cross-validation

library(C50)
library(plyr)

form <- "nbing3 ~ school+sex+age+address+famsize+pstatus+medu+fedu+mjob+fjob+reason+guardian+traveltime

folds <- split(data_arbol, cut(sample(1:nrow(data_arbol)),10))
acc_c50 <- rep(NA, length(folds))

for (i in 1:length(folds)) {
  test <- ldply(folds[i], data.frame)
  train <- ldply(folds[-i], data.frame)
  tmp_model <- C5.0(as.formula(form), train, trials=10)
  tmp_predict <- predict(tmp_model, newdata=test)
  conf_mat <- table(test$nbing3, tmp_predict)
  acc_c50[i] <- sum(diag(conf_mat))/sum(conf_mat)
}

print(sprintf("average accuracy using k-fold cross validation and C5.0 decision tree algorithm: %.3f pe

## [1] "average accuracy using k-fold cross validation and C5.0 decision tree algorithm: 78.436 percent

save.image("myworkspace_TFM.RData")

```