Course Practical Assignment - 1st Delivery (17 de marÃ§ del 2019)

Josep Clotet Ginovart

Eric Martin Obispo

Bank client data

Description of input variables:

1. age (numeric)
2. job : type of job (categorical: 'admin', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')
3. marital : marital status (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)
4. education (categorical:'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')
5. default: has credit in default? (categorical: 'no','yes','unknown')
6. housing: has housing loan? (categorical: 'no','yes','unknown')
7. loan: has personal loan? (categorical: 'no','yes','unknown')# related with the last contact of the current campaign:
8. contact: contact communication type (categorical:'cellular','telephone')
9. month: last contact month of year (categorical: 'jan', 'feb', 'mar',. . . , 'nov', 'dec')
10. day_of_week: last contact day of the week (categorical:'mon','tue','wed','thu','fri')
11. duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.
12. campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
13. pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
14. previous: number of contacts performed before this campaign and for this client (numeric)
15. poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')# social and economic context attributes
16. emp.var.rate: employment variation rate - quarterly indicator (numeric)
17. cons.price.idx: consumer price index - monthly indicator (numeric)
18. cons.conf.idx: consumer confidence index - monthly indicator (numeric)
19. euribor3m: euribor 3 month rate - daily indicator (numeric)
20. nr.employed: number of employees - quarterly indicator (numeric)
21. y - has the client subscribed a term deposit? (binary: 'yes','no')

Loading packages:

Loading data:

```
dirwd<-"D:/Users/Usuari/Documents/ADEIpractica"
#dirwd<-"D:/Documents/GitHub/ADEI"
setwd(dirwd)

df<-read.table( paste0(dirwd, "/bank-additional/bank-additional-full.csv"),  header=TRUE, sep=";")

# General description of the bank data
```

```r
#head(df)
nrow(df)
```

```
## [1] 41188
```

```r
ncol(df)
```

```
## [1] 21
```

```r
dim(df)
```

```
## [1] 41188    21
```

```r
# Selection of our 5000 samples with a specific seed value
set.seed(17041998)
llista<-sample(size=5000, x=1:nrow(df), replace=FALSE)
llista<-sort(llista)

# Overwrite the dataframe with our chosen sample and save the RData
df<-df[llista,]
save.image( paste0(dirwd, "/bank-additional/Bank5000_raw.RData") )
```

Our chosen sample:

```r
#load( paste0(dirwd, "/bank-additional/Bank5000_raw.RData") )
summary(df)
```

```
##       age                   job           marital
##  Min.   :18.00   admin.     :1234   divorced: 556
##  1st Qu.:32.00   blue-collar:1154   married :3053
##  Median :38.00   technician : 794   single  :1381
##  Mean   :40.07   services   : 500   unknown :  10
##  3rd Qu.:47.00   management : 413
##  Max.   :87.00   retired    : 205
##                  (Other)    : 700
##                education        default         housing           loan
##  university.degree  :1472   no     :3966   no     :2219   no     :4091
##  high.school        :1171   unknown:1034   unknown: 137   unknown: 137
##  basic.9y           : 716   yes    :   0   yes    :2644   yes    : 772
##  professional.course: 602
##  basic.4y           : 513
##  basic.6y           : 291
##  (Other)            : 235
##       contact          month      day_of_week    duration
##  cellular :3130   may    :1743   fri: 924    Min.   :   1.0
##  telephone:1870   jul    : 831   mon:1018    1st Qu.: 101.0
##                   aug    : 699   thu:1039    Median : 178.0
##                   jun    : 653   tue:1045    Mean   : 254.8
##                   nov    : 509   wed: 974    3rd Qu.: 317.0
##                   apr    : 310               Max.   :3785.0
##                   (Other): 255
##     campaign          pdays          previous            poutcome
##  Min.   : 1.000   Min.   :  0.0   Min.   :0.0000   failure    : 478
##  1st Qu.: 1.000   1st Qu.:999.0   1st Qu.:0.0000   nonexistent:4363
##  Median : 2.000   Median :999.0   Median :0.0000   success    : 159
```

```
##  Mean   : 2.583   Mean   :963.2   Mean   :0.1606
##  3rd Qu.: 3.000   3rd Qu.:999.0   3rd Qu.:0.0000
##  Max.   :33.000   Max.   :999.0   Max.   :4.0000
##
##   emp.var.rate      cons.price.idx  cons.conf.idx     euribor3m
##  Min.   :-3.40000   Min.   :92.20   Min.   :-50.80   Min.   :0.635
##  1st Qu.:-1.80000   1st Qu.:93.08   1st Qu.:-42.70   1st Qu.:1.334
##  Median : 1.10000   Median :93.77   Median :-41.80   Median :4.857
##  Mean   : 0.06326   Mean   :93.57   Mean   :-40.43   Mean   :3.613
##  3rd Qu.: 1.40000   3rd Qu.:93.99   3rd Qu.:-36.40   3rd Qu.:4.961
##  Max.   : 1.40000   Max.   :94.77   Max.   :-26.90   Max.   :5.000
##
##   nr.employed     y
##  Min.   :4964   no :4435
##  1st Qu.:5099   yes: 565
##  Median :5191
##  Mean   :5166
##  3rd Qu.:5228
##  Max.   :5228
##
```

InicialitzaciÃ³ del control d'errors, missings i outliers:

```r
columnes <- names(df) #list of column names

# creem 3 dataframes inicialitzats a 0 d'una fila amb les columnes de la nostra mostra;
# en ells hi posarem el nombre d'errors, missings i outliers per a cada variable
errors <- data.frame(matrix(0, ncol = length(columnes), nrow = 1))
colnames(errors)<-columnes

missings <- data.frame(matrix(0, ncol = length(columnes), nrow = 1))
colnames(missings)<-columnes

outliers <- data.frame(matrix(0, ncol = length(columnes), nrow = 1))
colnames(outliers)<-columnes

# columnes que portaran el control per individu:
df$num_missings <- 0
df$num_outliers <- 0
df$num_errors   <- 0
```

UNIVARIATE DESCRIPTIVE ANALYSIS (to be included for each variable):

Aqui estudiem cada variable buscant missing values, outliers i possibles errors. En el cas que en trobem, els transformem en NAs i procedim a una imputacio manual o els eliminem, o una imputaciÃ³ automatica (en un chunck posterior d'Imputation).

QUALITATIVE VARIABLES:

TambÃ© factoritzem aqui les categories (levels) de les variables qualitatives (discretes).

Job

Jobs “unknown” sÃ³n considerats com a categoria.

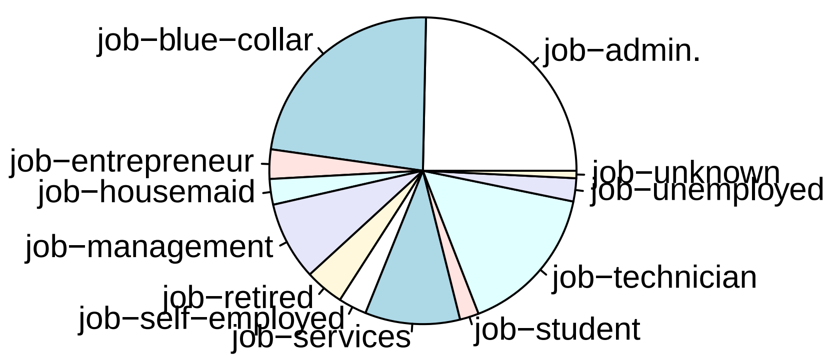
```
# Jobs "unknown" will be considered a category, not a missing value.
table(df$job, useNA="always")
```

```
##
##        admin.   blue-collar  entrepreneur     housemaid    management
##          1234          1154           155           135           413
##       retired self-employed      services       student    technician
##           205           149           500           100           794
##    unemployed       unknown          <NA>
##           122            39             0
```

```
# Missings:
miss<-which(is.na(df$job));
missings$job<-length(miss); length(miss)
```

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

```
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

# Factoritzem les categories (levels) de la columna i afegim l'etiqueta "job-":
df$job<-factor(df$job)
levels(df$job)<-paste0("job-",levels(df$job))

pie(summary(df$job))
```



Marital

Els “unknowns” seran imputats mÃ©s endavant automiticament.

```r
# Marital "unknown" will be a missing value (set to NA):
sel<-which(df$marital=="unknown");length(sel)
```

```
## [1] 10
```

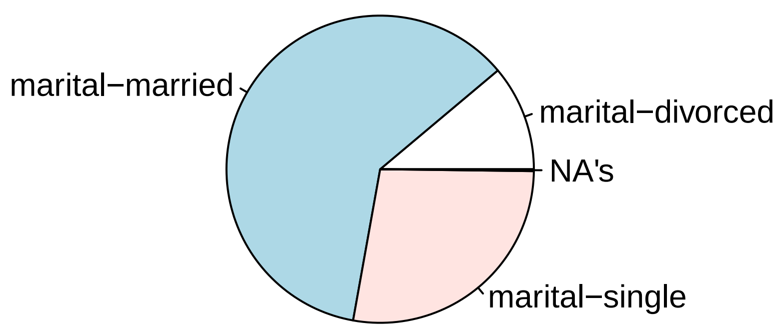
```r
df$marital[sel]<-NA

# Missings:
miss<-which(is.na(df$marital));
missings$marital<-length(miss); length(miss)
```

```
## [1] 10
```

```r
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

# Factoritzem les categories (levels) de la columna i afegim l'etiqueta "marital-":
df$marital<-factor(df$marital)
levels(df$marital)<-paste0("marital-",levels(df$marital))
summary(df$marital)
```

```
## marital-divorced  marital-married   marital-single              NA's
##              556             3053             1381                10
```

```r
pie(summary(df$marital))
```



Education

Education “unknown” Ã©s considerada com a categoria. La categoria “illiterate” Ã©s inclosa manualment a “basic.4y”.

```r
# Education "unknown" will be considered a category, not a missing value.
table(df$education, useNA="always")
```

```
##
##         basic.4y           basic.6y            basic.9y
##              513                291                 716
##      high.school         illiterate professional.course
##             1171                  3                 602
##  university.degree           unknown                <NA>
##             1472                232                   0
```

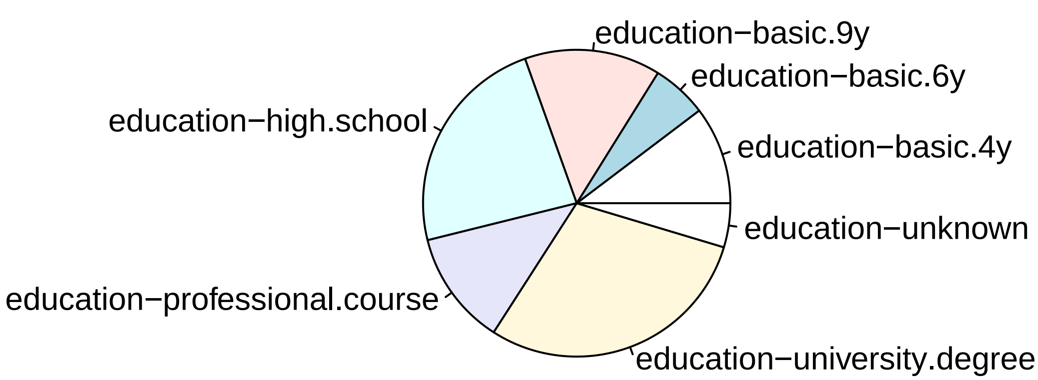
```r
# Illiterates are consired as basic.4y.educated:
sel<-which(df$education=="illiterate");length(sel)
```

```
## [1] 3
```

```r
df[sel, "education"]<-"basic.4y"

# Missings:
miss<-which(is.na(df$education));
missings$education<-length(miss); length(miss)
```

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

```r
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

# Factoritzem les categories (levels) de la columna i afegim l'etiqueta "education-":
df$education<-factor(df$education)
levels(df$education)<-paste0("education-",levels(df$education))

pie(summary(df$education))
```



Default (has credit in default?)

Default (owes credit) “unknown” will be considered a category, not a missing value.

```
table(df$default, useNA="always")
```

```
##
##      no unknown     yes    <NA>
##    3966    1034       0       0
```

```
# Missings:
miss<-which(is.na(df$default));
missings$default<-length(miss); length(miss)
```

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

```
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

# Factoritzem les categories (levels) de la columna i afegim l'etiqueta "default-":
df$default<-factor(df$default)
levels(df$default)<-paste0("default-",levels(df$default))

summary(df$default)
```

```
##      default-no default-unknown
##            3966            1034
```

Housing

Housing “unknown” will be considered a category, not a missing value.

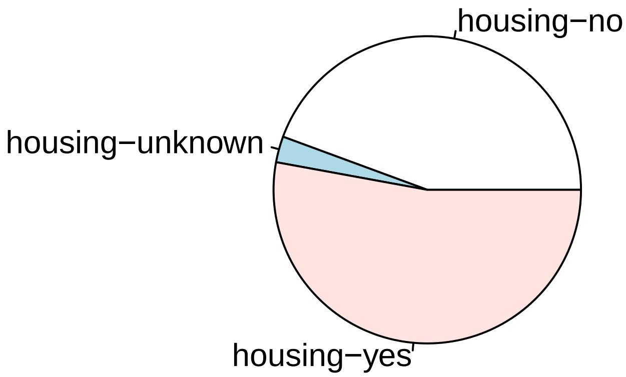
```
table(df$housing, useNA="always")
```

```
##
##      no unknown     yes    <NA>
##    2219     137    2644       0
```

```
# Missings:
miss<-which(is.na(df$housing));
missings$housing<-length(miss); length(miss)
```

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

```
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

# Factoritzem les categories (levels) de la columna i afegim l'etiqueta "housing-":
df$housing<-factor(df$housing)
levels(df$housing)<-paste0("housing-",levels(df$housing))

pie(summary(df$housing))
```



Loan (has personal loan?)

Loan “unknown” will be a missing value (set to NA) i sera imputat mÃ©s endavant automaticament.

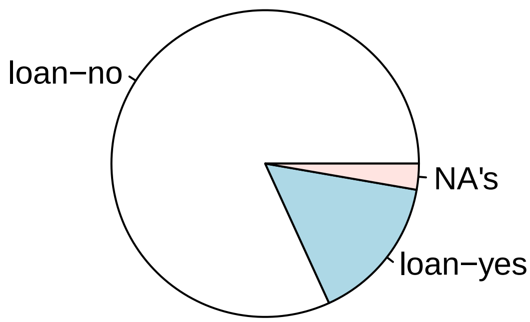
```
sel<-which(df$loan=="unknown");length(sel)
```

```
## [1] 137
```

```
df$loan[sel]<-NA

# Missings:
miss<-which(is.na(df$loan));
missings$loan<-length(miss); length(miss)
```

```
## [1] 137
```

```
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

# Factoritzem les categories (levels) de la columna i afegim l'etiqueta "loan-":
df$loan<-factor(df$loan)
levels(df$loan)<-paste0("loan-",levels(df$loan))

pie(summary(df$loan))
```



Contact

```
summary(df$contact)
```

```
##  cellular telephone
##      3130      1870
```

```
# Missings:
miss<-which(is.na(df$contact));
missings$contact<-length(miss); length(miss)
```

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

```
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

# Factoritzem les categories (levels) de la columna i afegim l'etiqueta "contact-":
df$contact<-factor(df$contact)
levels(df$contact)<-paste0("contact-",levels(df$contact))

summary(df$contact)
```

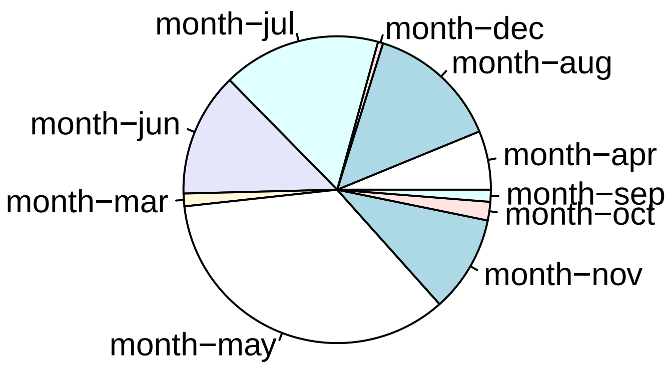
```
##  contact-cellular contact-telephone
##              3130              1870
```

Month

```
miss<-which(is.na(df$month));
missings$month<-length(miss); length(miss)
```

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

```r
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

# Factoritzem les categories (levels) de la columna i afegim l'etiqueta "month-":
df$month<-factor(df$month)
levels(df$month)<-paste0("month-",levels(df$month))

pie(summary(df$month))
```



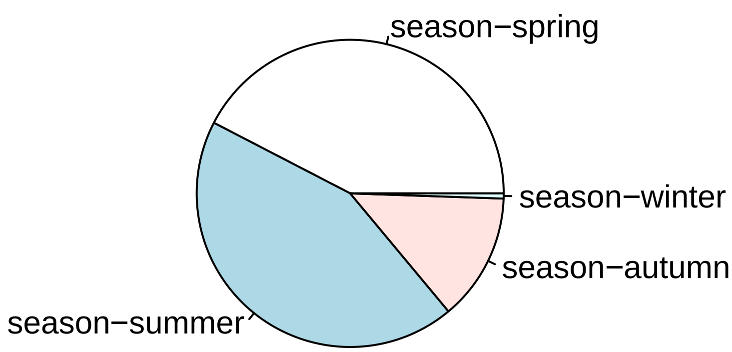
Month -> definim noves factor categories per Season.

New factors grouping original levels will be considered very positively.

```r
# Define new factor categories: 1- Spring 2-Summer 3-Autumn, 4-Winter
df$f.season <- 4
# 1 level - spring
sel<-which(df$month %in% c("month-mar","month-apr","month-may"))
df$f.season[sel] <-1

# 2 level - summer
sel<-which(df$month %in% c("month-jun","month-jul","month-aug"))
df$f.season[sel] <-2

# 3 level - autumn
sel<-which(df$month %in% c("month-sep","month-oct","month-nov"))
df$f.season[sel] <-3

df$f.season<-factor(df$f.season, levels=1:4, labels=c("season-spring","season-summer",
                                                        "season-autumn", "season-winter"))
```

```r
summary(df$f.season);pie(summary(df$f.season))
```

```
## season-spring season-summer season-autumn season-winter
##          2120          2183           670            27
```



Day_of_week

```r
miss<-which(is.na(df$day_of_week));
missings$day_of_week<-length(miss); length(miss)
```

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

```r
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

# Factoritzem les categories (levels) de la columna i afegim l'etiqueta "day_of_week-":
df$day_of_week<-factor(df$day_of_week)
levels(df$day_of_week)<-paste0("day_of_week-",levels(df$day_of_week))

summary(df$day_of_week)
```

```
## day_of_week-fri day_of_week-mon day_of_week-thu day_of_week-tue
##             924            1018            1039            1045
## day_of_week-wed
##             974
```

Poutcome (outcome of previous marketing campaign)

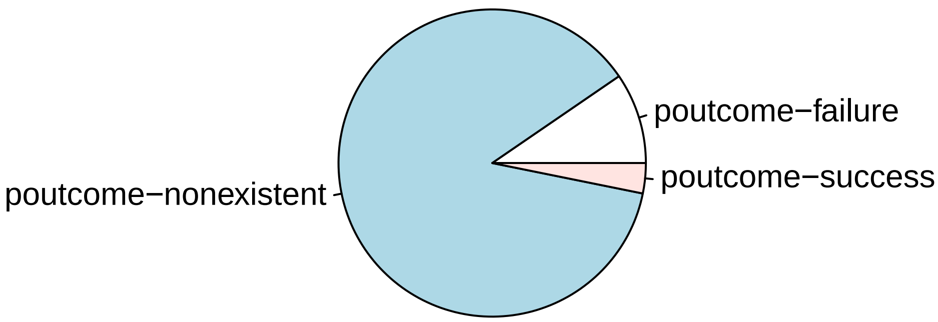
```r
# Poutcome "nonexistent" will be considered a category, not a missing value.
table(df$poutcome, useNA="always")
```

```
##
##     failure nonexistent     success        <NA>
##         478        4363         159           0
```

```r
# All missing data indicated as NA:
miss<-which(is.na(df$poutcome));
missings$poutcome<-length(miss); length(miss)
```

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

```r
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

# Factoritzem les categories (levels) de la columna i afegim l'etiqueta "poutcome-":
df$poutcome<-factor(df$poutcome)
levels(df$poutcome)<-paste0("poutcome-",levels(df$poutcome))

pie(summary(df$poutcome))
```



y (has the client subscribed a term deposit?)

```r
miss<-which(is.na(df$y));
missings$y<-length(miss); length(miss)
```

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

```r
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

# Factoritzem les categories (levels) de la columna i afegim l'etiqueta "y-":
df$y<-factor(df$y)
levels(df$y)<-paste0("y-",levels(df$y))
```

```r
summary(df$y)
```

```
##  y-no y-yes
##  4435   565
```

QUANTITATIVES VARIABLES:

Defining useful function for outliers detection:

```r
calcQ <- function(x){
  s.x <- summary(x)

  iqr <- s.x[5]-s.x[2] # IQR = Q3([5]) -  Q1([2])

  list(souti=s.x[2]-3*iqr, mouti=s.x[2]-1.5*iqr, min=s.x[1], q1=s.x[2],
       q2=s.x[3], q3=s.x[5], max=s.x[6], mouts=s.x[5]+1.5*iqr, souts=s.x[5]+3*iqr)
}
```

Age

```r
summary(df$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   32.00   38.00   40.07   47.00   87.00
```

```r
# No tenim cap missing NA!
miss<-which(is.na(df$age))
missings$age<-length(miss); length(miss)
```

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

```r
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

par(mfrow=c(1,2))
hist(df$age, breaks=10, main="age - histogram")
Boxplot(df$age)
```

```
##  [1] 4570 4634 3623 3628 3631 4755 4612 4734 4740 4512
```

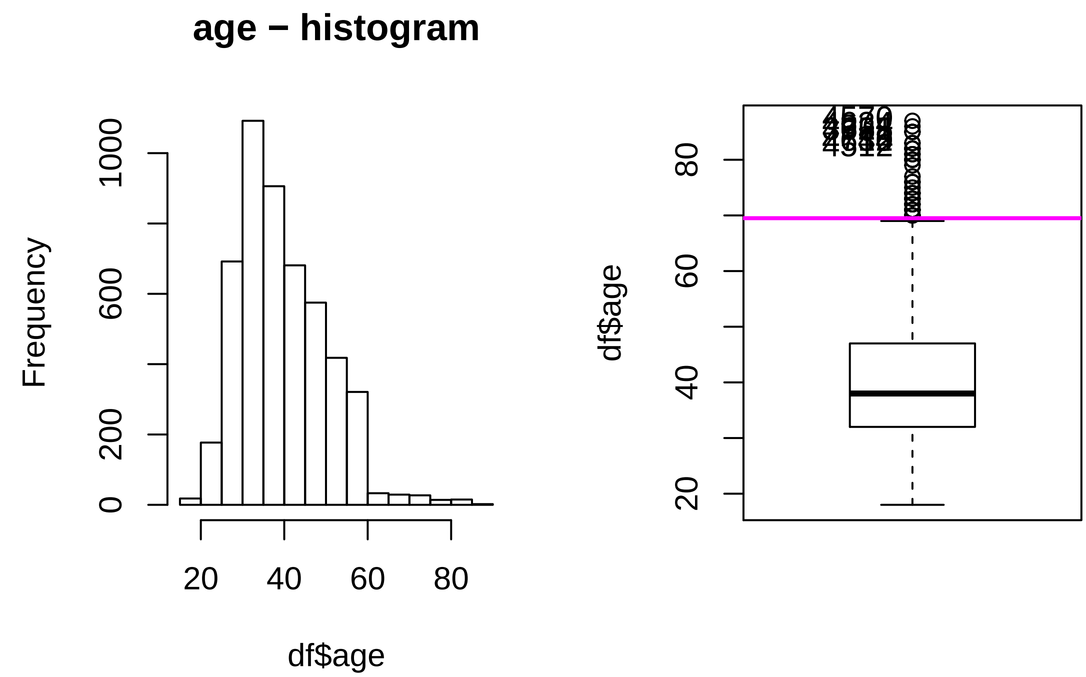
```r
# Errors are under aged people:
err<-which(df$age < 18)
errors$age<-length(err); length(err)
```

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

```r
if(length(err)>0) df<-df[-err,]

# Outliers:
out.var <- calcQ(df$age)
abline(h=out.var[["mouts"]], col="magenta", lwd=2); out.var[["mouts"]]
```

```
## 3rd Qu.
##    69.5
```

```r
# But our outliers will be the ones above 100 years (there is none):
abline(h=100, col="red", lwd=2)
```




```
out<-which(df$age > 100)
outliers$age<-length(out); length(out)
```

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

```
if(length(out)>0) df<-df[-out,]
```

Duration

Els outliers en la variable duraciÃ³ han estat eliminats. Corresponen a duracions per sota els 5 segons (trucada massa curta a un client que potser no podia parlar en aquell moment o penja per error) i per sobre dels 1600 segon s (26 minuts).

```
summary(df$duration)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##     1.0   101.0   178.0   254.8   317.0  3785.0
```

```
# No tenim cap missing NA!
miss<-which(is.na(df$duration));
missings$duration<-length(miss); length(miss)
```

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

```
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

par(mfrow=c(1,2))
hist(df$duration, breaks=20, main="duration - histogram")
Boxplot(df$duration)
```

```
##  [1] 4929 3368 2817 4759 1285 2907 2033 3815 4998 3280
```

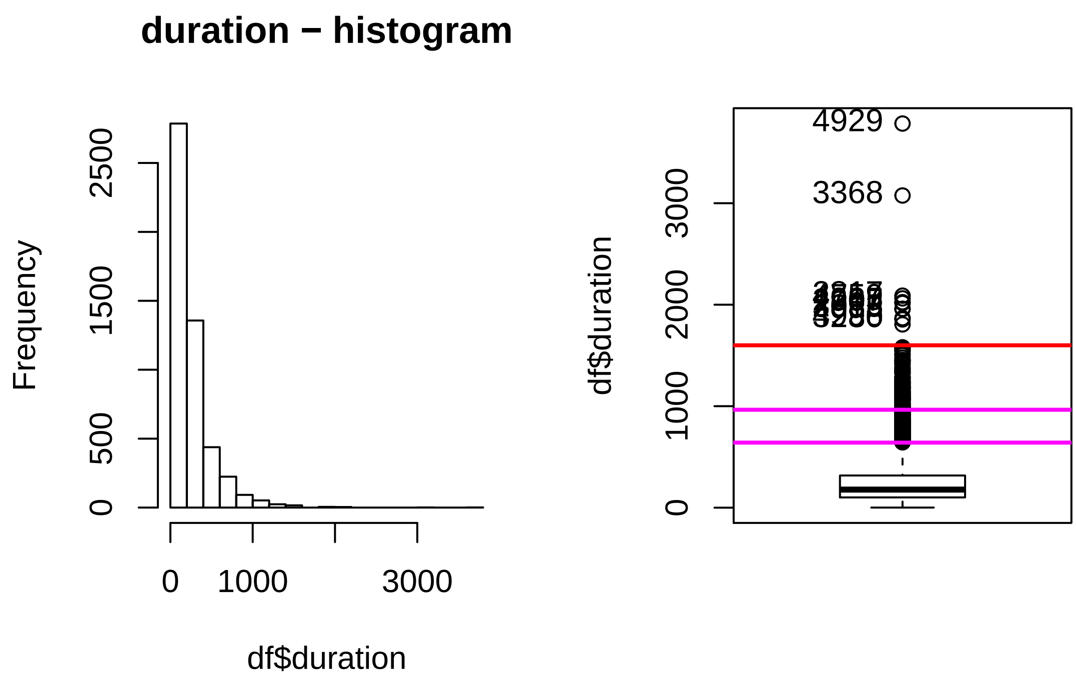
```
# Outliers:
out.var <- calcQ(df$duration)
abline(h=out.var[["mouts"]], col="magenta", lwd=2); out.var[["mouts"]]
```

```
## 3rd Qu.
##     641
```

```
abline(h=out.var[["souts"]], col="magenta", lwd=2); out.var[["souts"]]
```

```
## 3rd Qu.
##     965
```

```
# But our outliers will be the ones above 1600 and below 5 seconds:
abline(h=1600, col="red", lwd=2)
```




```
out<-which( (df$duration < 5) | (df$duration > 1600) )
outliers$duration=length(out); length(out)
```

```
## [1] 14
```

```
df[out, "num_outliers"]<- df[out, "num_outliers"]+1
df[out, "duration"]<-NA

# Eliminem els outliers:
if(length(out)>0) df<-df[-out,]

# Final summary of duration variable:
# par(mfrow=c(1,1))
# summary(df$duration)
# Boxplot(df$duration)
```

Duration -> creem una columna de duracio en minuts:

```
df$minutes<-df$duration/60
summary(df$minutes)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##  0.08333  1.68333  2.95000  4.17703  5.26667 26.33333
```

Campaign

```
# summary(df$campaign)
# No tenim cap missing NA!
miss<-which(is.na(df$campaign));
missings$campaign<-length(miss); length(miss)
```

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

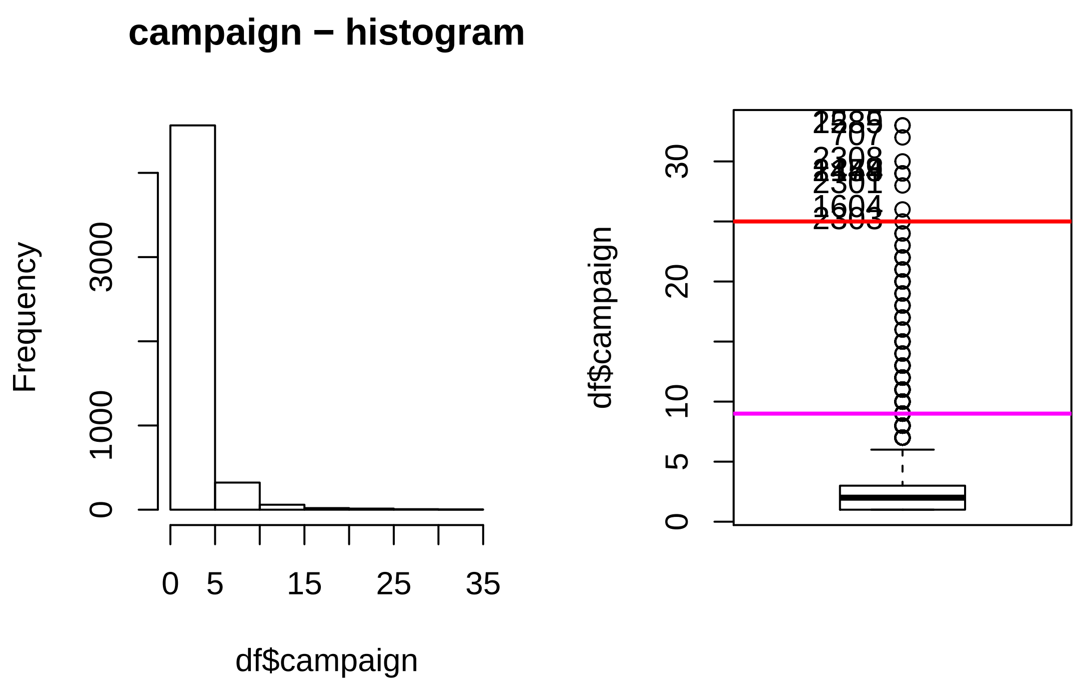
```
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

par(mfrow=c(1,2))
hist(df$campaign, breaks=10, main="campaign - histogram")
Boxplot(df$campaign)
```

```
##  [1] 1589 2285  707 2308 1158 1474 2149 2301 1604 2303
```

```
# Outliers:
out.var <- calcQ(df$campaign)
abline(h=out.var[["souts"]], col="magenta", lwd=2); out.var[["souts"]]
```

```
## 3rd Qu.
##       9
```

```
# But our outliers will be the ones contacted more than 25 times:
abline(h=25, col="red", lwd=2)
```




```
out<-which(df$campaign > 25)
df[out, "num_outliers"]<- df[out, "num_outliers"]+1
outliers$campaign=length(out); length(out)
```

```
## [1] 9
```

```
df[out, "campaign"]<-NA

# Final summary of campaign variable:
# par(mfrow=c(1,1))
# summary(df$campaign)
# Boxplot(df$campaign)
```

Pdays

```
# No tenim cap missing NA!
miss<-which(is.na(df$pdays));
missings$pdays<-length(miss); length(miss)
```

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

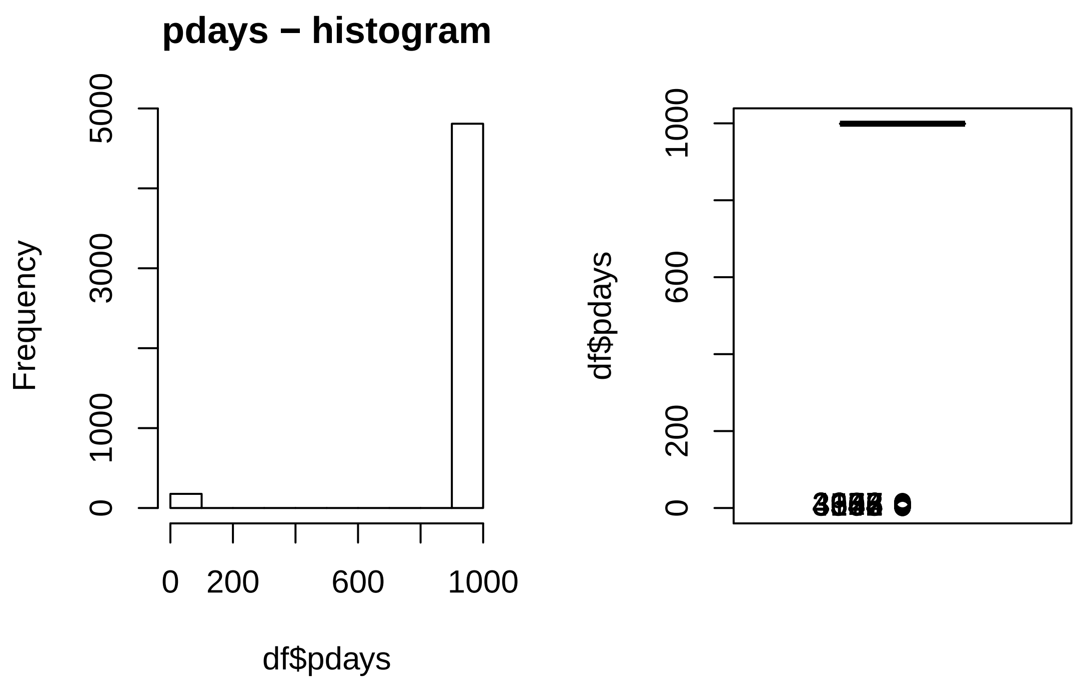
```
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

# Values that are 999 mean never contacted before:
never<-which(df$pdays==999)

# They correspond to this percentage of rows:
length(never)/5000*100
```

```
## [1] 96.18
```

```r
# No outliers!

# Final summary of pdays variable:
summary(df$pdays)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   999.0   999.0   963.7   999.0   999.0
```

```r
par(mfrow=c(1,2))
hist(df$pdays, breaks=10, main="pdays - histogram")
Boxplot(df$pdays)
```




```
##  [1] 3148 4902 3576 4135 4366 3627 3642 3644 3646 4352
```

Previous

```r
# No tenim cap missing NA!
miss<-which(is.na(df$previous));
missings$previous<-length(miss); length(miss)
```

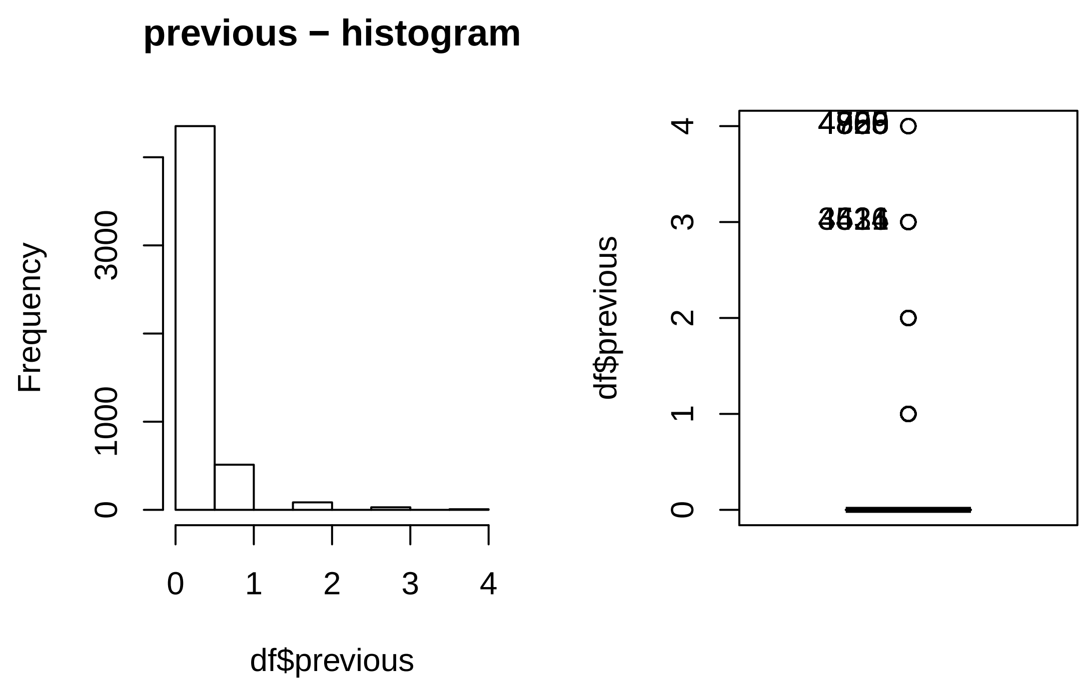
```
## [1] 0
```

```r
df[miss, "num_missings"]<- df[miss, "num_missings"]+1

par(mfrow=c(1,2))
hist(df$previous,  main="previous - histogram")

# Final summary of previous variable:
summary(df$previous)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
## 0.0000 0.0000 0.0000 0.1598 0.0000 4.0000
```

```r
Boxplot(df$previous)
```



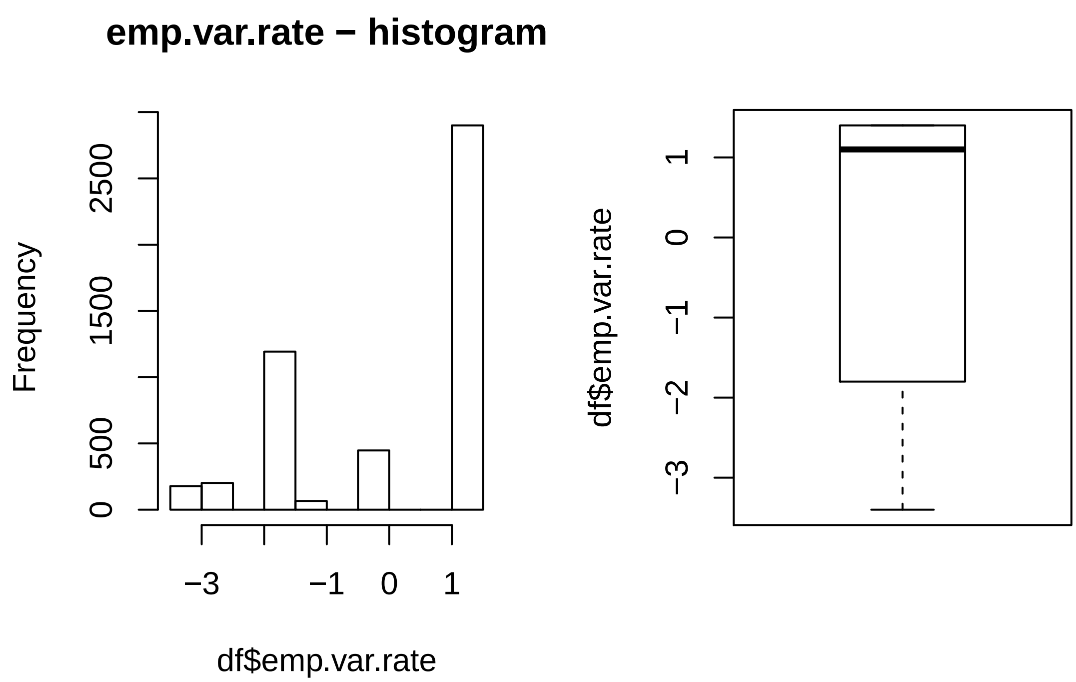

```
## [1] 4769 4786 4805 4826 4850 4888 4925 3431 4516 4624
```

emp.var.rate

```r
# Neither missing, outliers nor error values.
par(mfrow=c(1,2))

hist(df$emp.var.rate,  main="emp.var.rate - histogram")
summary(df$emp.var.rate)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -3.40000 -1.80000  1.10000  0.06446  1.40000  1.40000
```

```r
Boxplot(df$emp.var.rate)
```

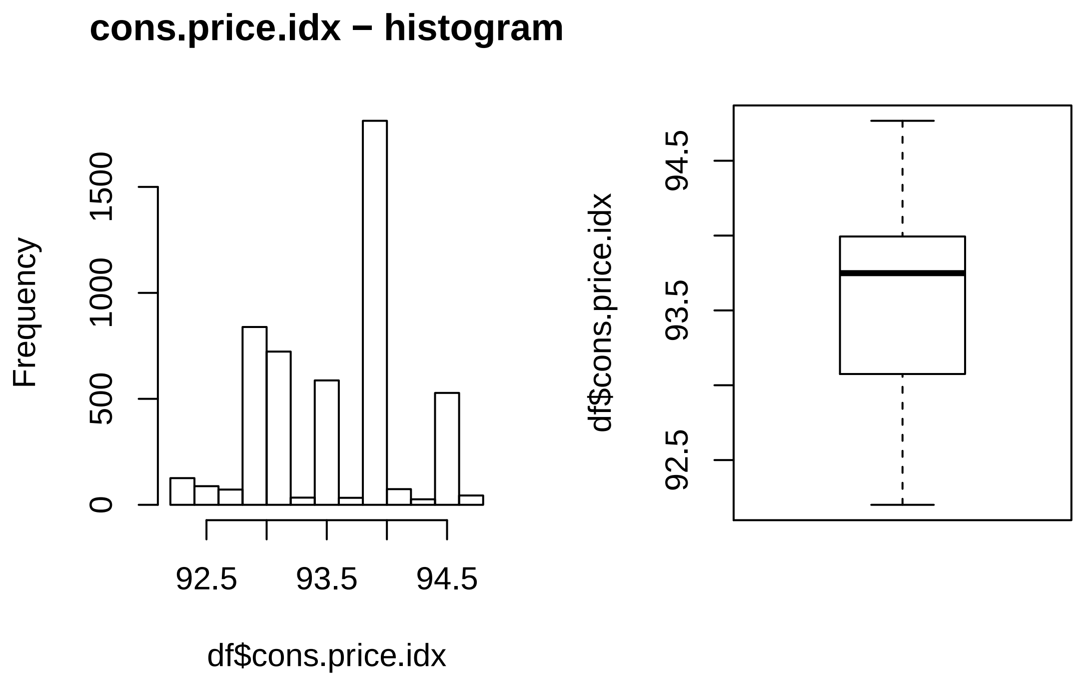


cons.price.idx

```
# Neither missing, outliers nor error values.
par(mfrow=c(1,2))

hist(df$cons.price.idx,  main="cons.price.idx - histogram")
summary(df$cons.price.idx)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   92.20   93.08   93.75   93.57   93.99   94.77
```

```
Boxplot(df$cons.price.idx)
```

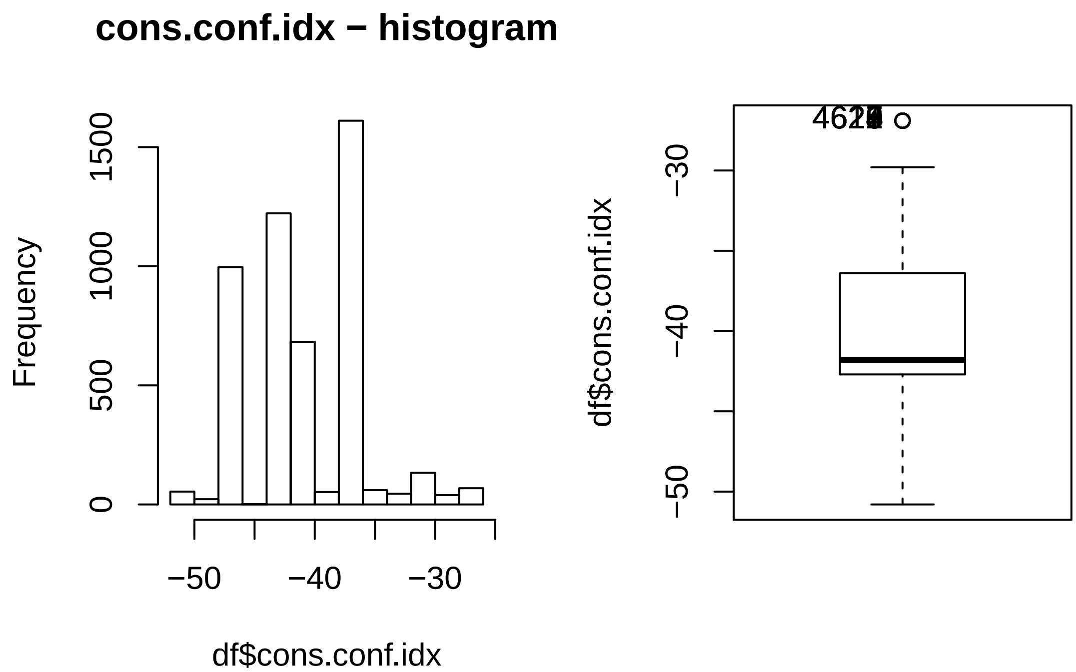


cons.conf.idx

```
# Neither missing, outliers nor error values.
par(mfrow=c(1,2))

hist(df$cons.conf.idx,  main="cons.conf.idx - histogram")
summary(df$cons.conf.idx)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -50.80  -42.70  -41.80  -40.43  -36.40  -26.90
```

```
Boxplot(df$cons.conf.idx)
```




```
## [1] 4617 4618 4619 4620 4621 4622 4623 4624 4625 4626
```

euribor3m

```r
# Neither missing, outliers nor error values.
par(mfrow=c(1,2))

hist(df$euribor3m,  main="euribor3m - histogram")
summary(df$euribor3m)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.635   1.334   4.857   3.614   4.961   5.000
```

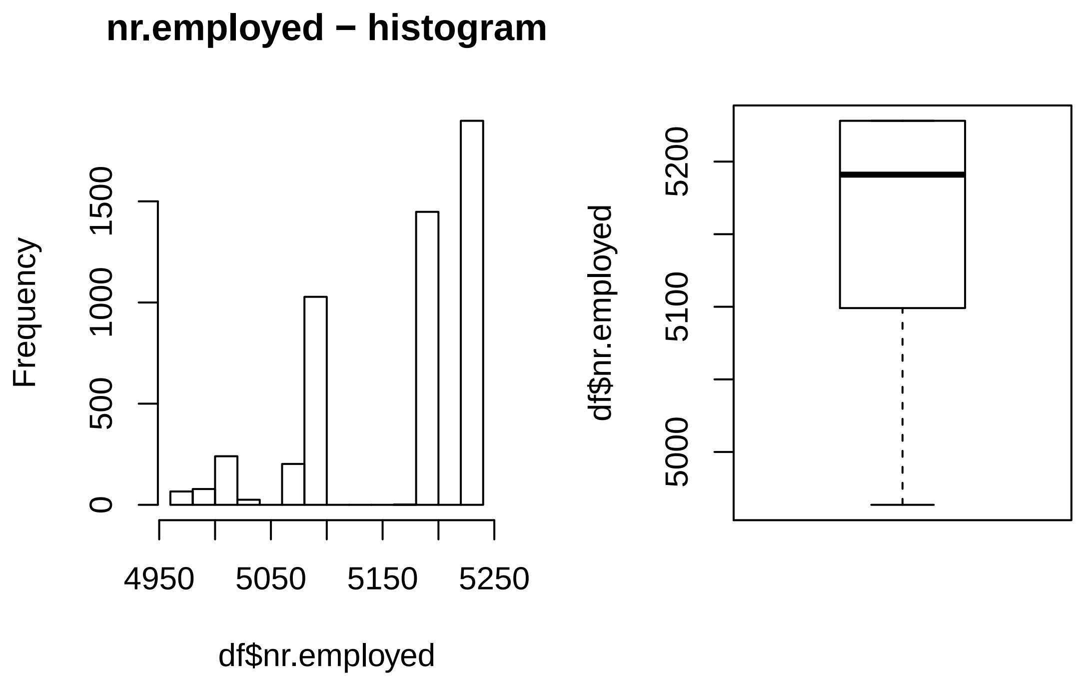
```r
Boxplot(df$euribor3m)
```

nr.employed

```r
# Neither missing, outliers nor error values.
par(mfrow=c(1,2))

hist(df$nr.employed,  main="nr.employed - histogram")
summary(df$nr.employed)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4964    5099    5191    5166    5228    5228
```

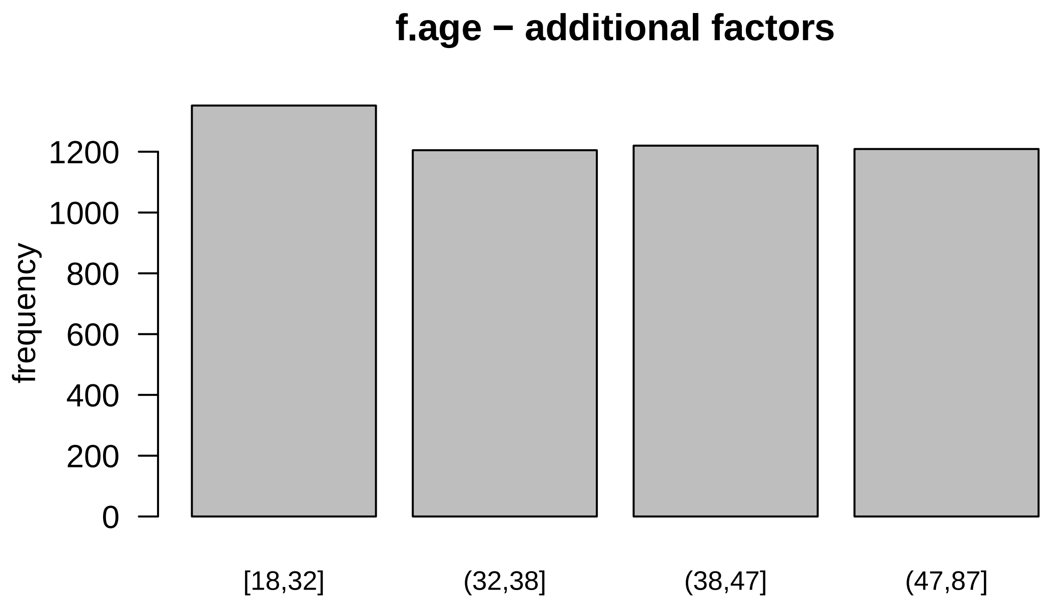
```r
Boxplot(df$nr.employed)
```



DISCRETITZACIO DE VARIABLES NUMERIQUES:

Original numeric variables corresponding to real quantitative concepts are kept as numeric but additional factors should also be created as a discretization of each numeric variable.

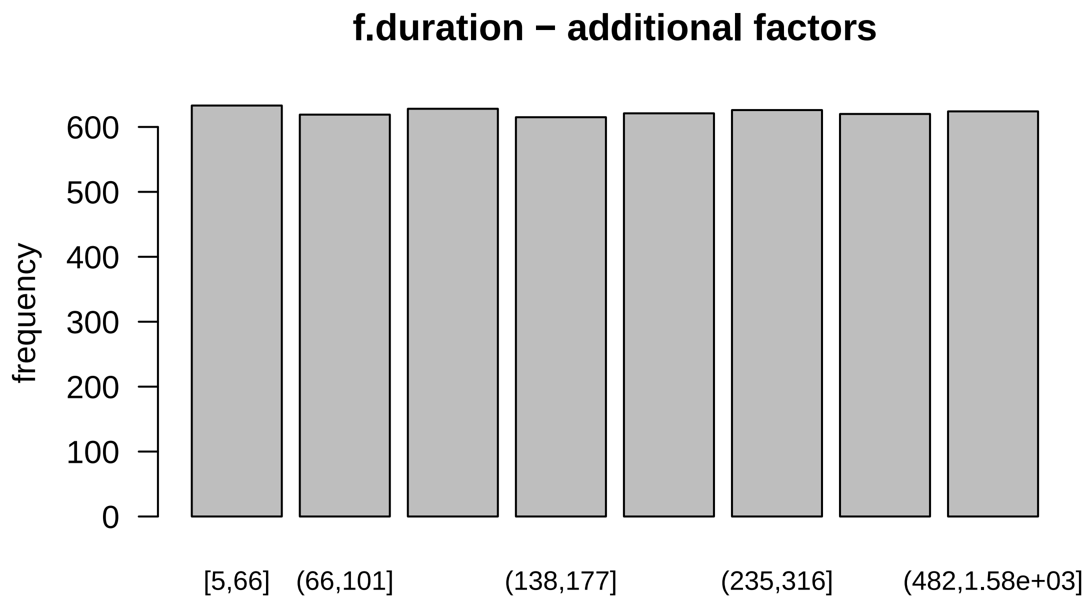
```
par(mfrow=c(1,1))

# AGE
qulist<-quantile(df$age, seq(0,1,0.25), na.rm=TRUE)

df$f.age<-factor( cut(df$age, breaks=qulist, include.lowest=T) )

# Es mostra una distribuciÃ³ d'edats equitativa amb aquesta factoritzaciÃ³:
barplot(table(df$f.age), main="f.age - additional factors", ylab="frequency", las=1, cex.names=0.8)
```




```
summary(df$f.age)
```

```
## [18,32] (32,38] (38,47] (47,87]
##    1352    1205    1220    1209
```

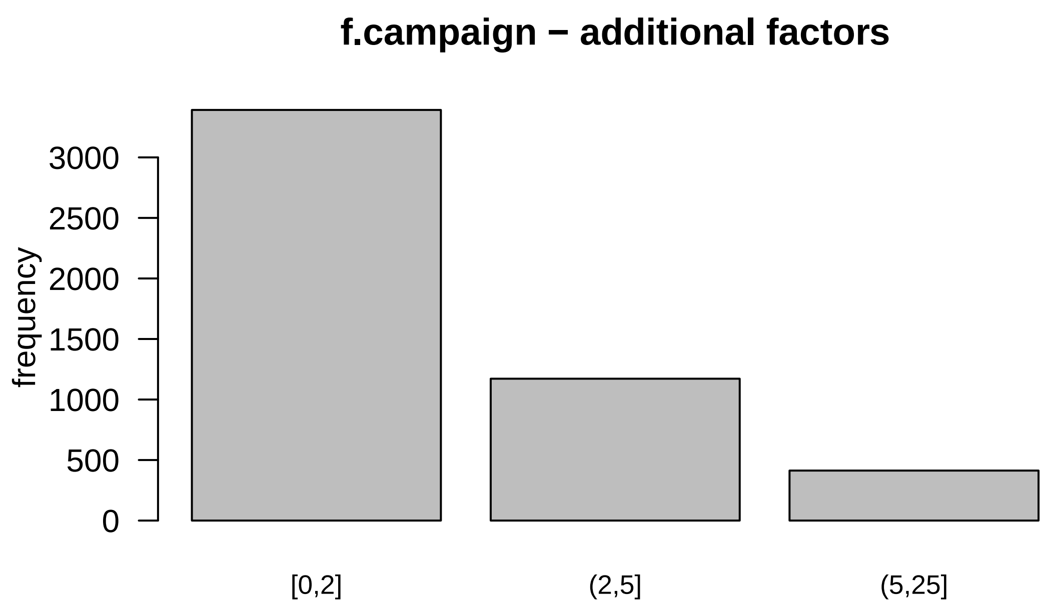
```
levels(df$f.age)<-paste0("f.age-", levels(df$f.age) )


# DURATION
qulist<-quantile(df$duration, seq(0,1,0.125), na.rm=TRUE)

df$f.duration<-factor( cut(df$duration, breaks=qulist, include.lowest=T) )

# Es mostra una distribuciÃ³ de duracions de la trucada equitativa amb aquesta factoritzaciÃ³:
barplot(table(df$f.duration), main="f.duration - additional factors", ylab="frequency", las=1, cex.names
```




```
levels(df$f.duration)<-paste0("f.duration-", levels(df$f.duration) )
summary(df$f.duration)
```

```
##        f.duration-[5,66]       f.duration-(66,101]
##                      633                       619
##      f.duration-(101,138]      f.duration-(138,177]
##                      628                       615
##      f.duration-(177,235]      f.duration-(235,316]
##                      621                       626
##      f.duration-(316,482] f.duration-(482,1.58e+03]
##                      620                       624
```

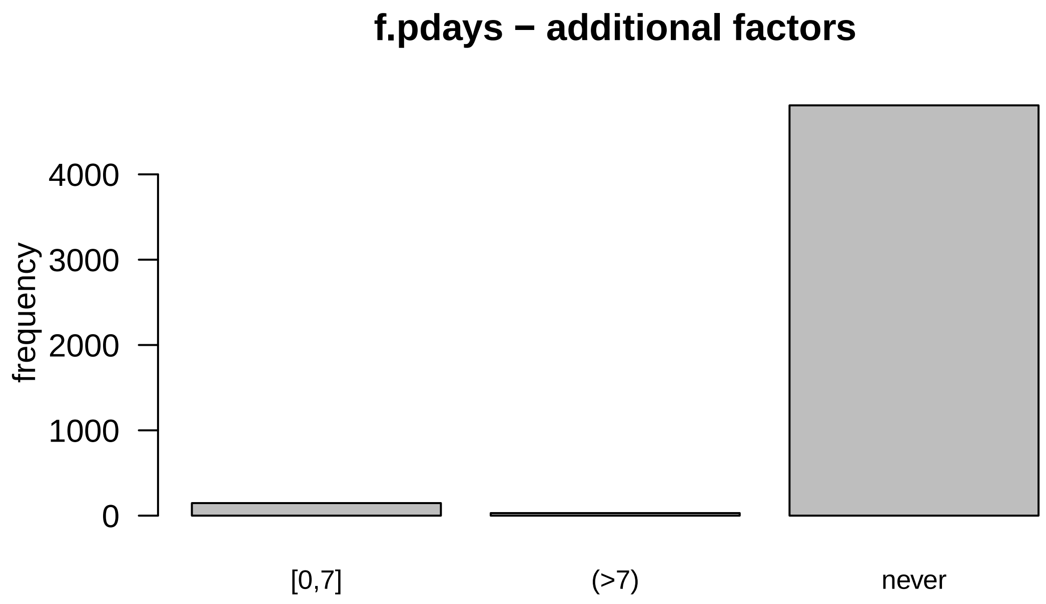
```
# CAMPAIGN
qulist<-quantile(df$campaign, seq(0,1,0.5), na.rm=TRUE)

df$f.campaign<-factor( cut(df$campaign, breaks=c(0,2,5,25), include.lowest=T) )

# Resultat de la factoritzaciÃ³ de cops que s'ha contactat al client en la campanya actual:
barplot(table(df$f.campaign), main="f.campaign - additional factors", ylab="frequency", las=1, cex.name
```




```r
levels(df$f.campaign)<-paste0("f.campaign-", levels(df$f.campaign) )
summary(df$f.campaign)
```

```
##  f.campaign-[0,2]  f.campaign-(2,5] f.campaign-(5,25]              NA's
##              3392              1172               413                 9
```

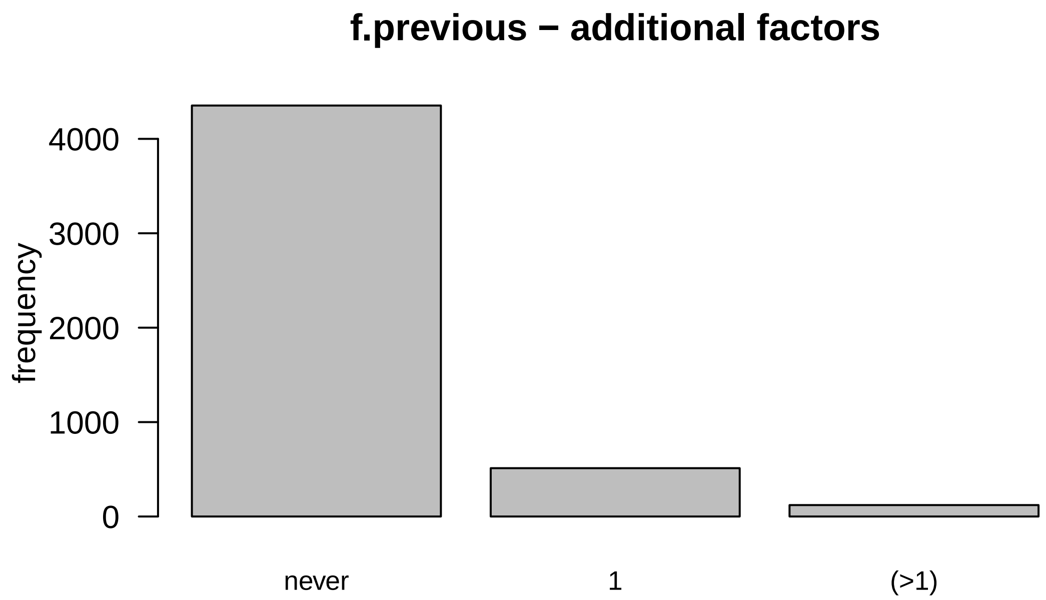
```r
# PDAYS
df$f.pdays<-factor( cut(df$pdays, breaks=c(0, 7, 998, 999), include.lowest=T) )


# Resultat de la factoritzaciÃ³ dels dies que fa
# que s'ha contactat al client en una altra campanya:
levels(df$f.pdays)<-c("[0,7]", "(>7)", "never")
barplot(table(df$f.pdays), main="f.pdays - additional factors", ylab="frequency", las=1, cex.names=0.8)
```




```r
levels(df$f.pdays)<-paste0("f.pdays-", levels(df$f.pdays) )
summary(df$f.pdays)
```

```
## f.pdays-[0,7]  f.pdays-(>7) f.pdays-never
##           147            30          4809
```

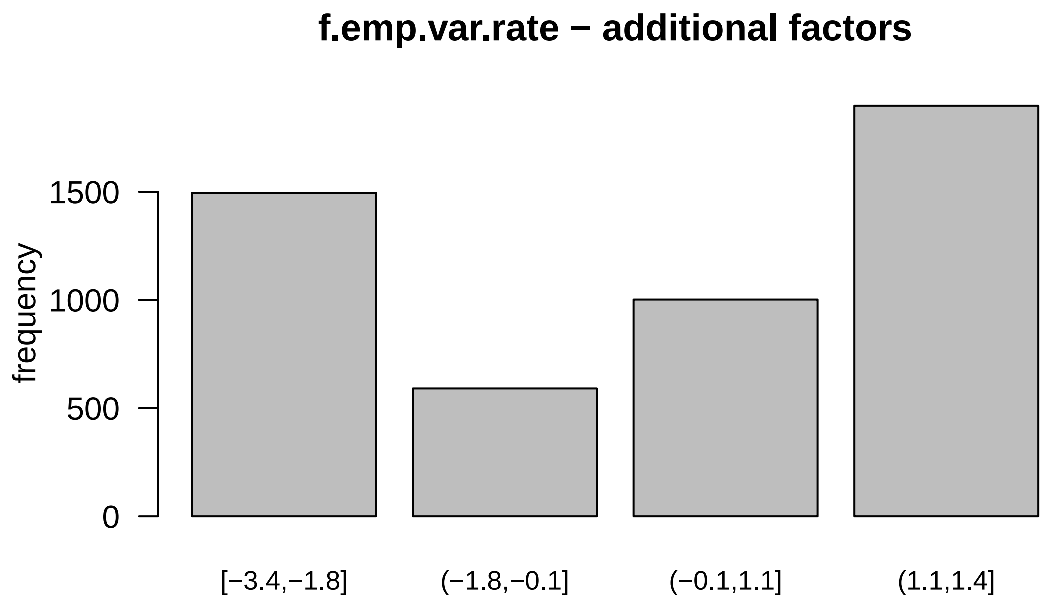
```r
# PREVIOUS
df$f.previous<-factor( cut(df$previous, breaks=c(-Inf, 0, 1, +Inf), include.lowest=T) )

levels(df$f.previous)<-c("never", "1", "(>1)")

# Resultat de la factoritzaciÃ³ de number of contacts performed
# before this campaign and for this client:
barplot(table(df$f.previous), main="f.previous - additional factors", ylab="frequency", las=1, cex.names
```




```
levels(df$f.previous)<-paste0("f.previous-", levels(df$f.previous) )
summary(df$f.previous)
```

```
## f.previous-never     f.previous-1  f.previous-(>1)
##             4353              512              121
```

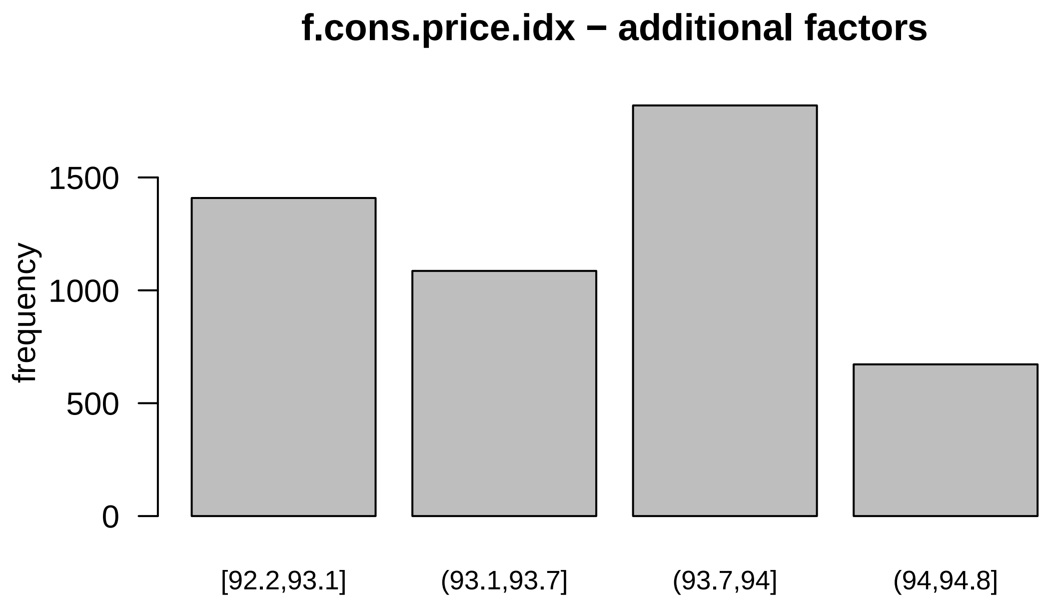
```
# EMP.VAR.RATE
qulist<-quantile(df$emp.var.rate, seq(0,1,0.125), na.rm=TRUE)

df$f.emp.var.rate <-factor( cut(df$emp.var.rate   , breaks=unique(qulist), include.lowest=T) )

barplot(table(df$f.emp.var.rate), main="f.emp.var.rate - additional factors", ylab="frequency", las=1, c
```




```r
levels(df$f.emp.var.rate)<-paste0("f.emp.var.rate-", levels(df$f.emp.var.rate) )
summary(df$f.emp.var.rate)
```

```
## f.emp.var.rate-[-3.4,-1.8] f.emp.var.rate-(-1.8,-0.1]
##                       1495                        591
##  f.emp.var.rate-(-0.1,1.1]   f.emp.var.rate-(1.1,1.4]
##                       1002                       1898
```

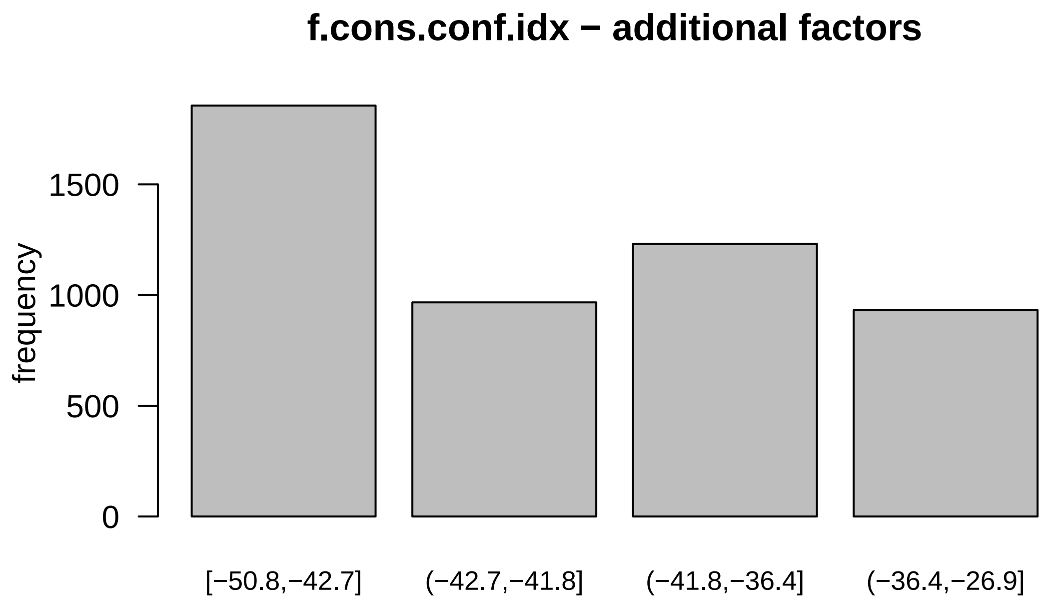
```r
# CONS.PRICE.IDX
qulist<-quantile(df$cons.price.idx, seq(0,1,0.25), na.rm=TRUE)

df$f.cons.price.idx <-factor( cut(df$cons.price.idx  , breaks=unique(qulist), include.lowest=T) )

barplot(table(df$f.cons.price.idx), main="f.cons.price.idx - additional factors", ylab="frequency", las=
```




```r
levels(df$f.cons.price.idx)<-paste0("f.cons.price.idx-", levels(df$f.cons.price.idx) )
summary(df$f.cons.price.idx)
```

```
## f.cons.price.idx-[92.2,93.1] f.cons.price.idx-(93.1,93.7]
##                         1409                         1086
##   f.cons.price.idx-(93.7,94]   f.cons.price.idx-(94,94.8]
##                         1819                          672
```

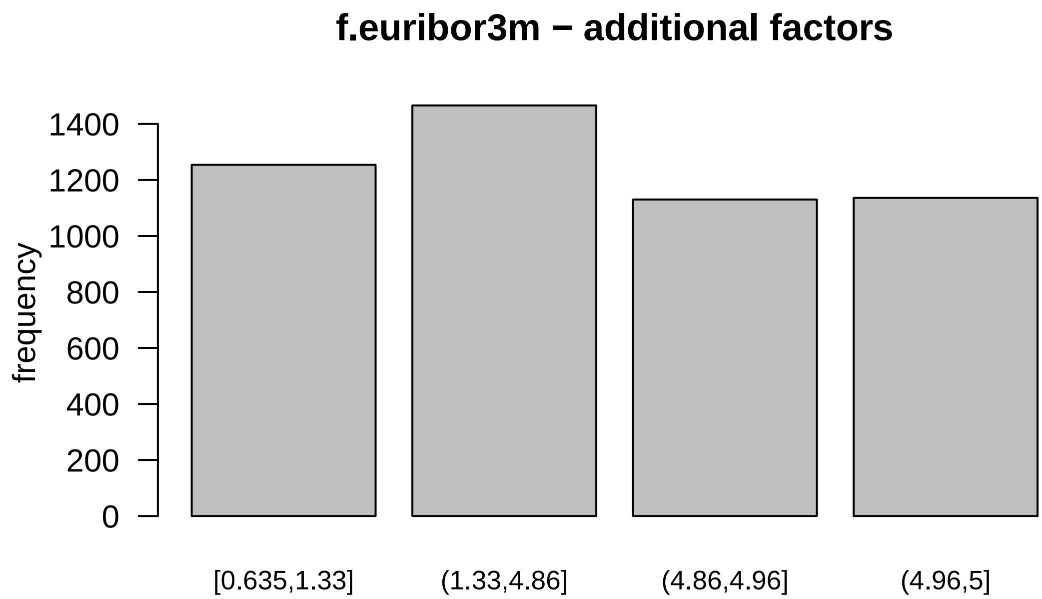
```r
# CONS.CONF.IDX
qulist<-quantile(df$cons.conf.idx, seq(0,1,0.25), na.rm=TRUE)

df$f.cons.conf.idx <-factor( cut(df$cons.conf.idx  , breaks=unique(qulist), include.lowest=T) )

barplot(table(df$f.cons.conf.idx), main="f.cons.conf.idx - additional factors", ylab="frequency", las=1
```




```
levels(df$f.cons.conf.idx)<-paste0("f.cons.conf.idx-", levels(df$f.cons.conf.idx) )
summary(df$f.cons.conf.idx)
```

```
## f.cons.conf.idx-[-50.8,-42.7] f.cons.conf.idx-(-42.7,-41.8]
##                          1856                           967
## f.cons.conf.idx-(-41.8,-36.4] f.cons.conf.idx-(-36.4,-26.9]
##                          1231                           932
```

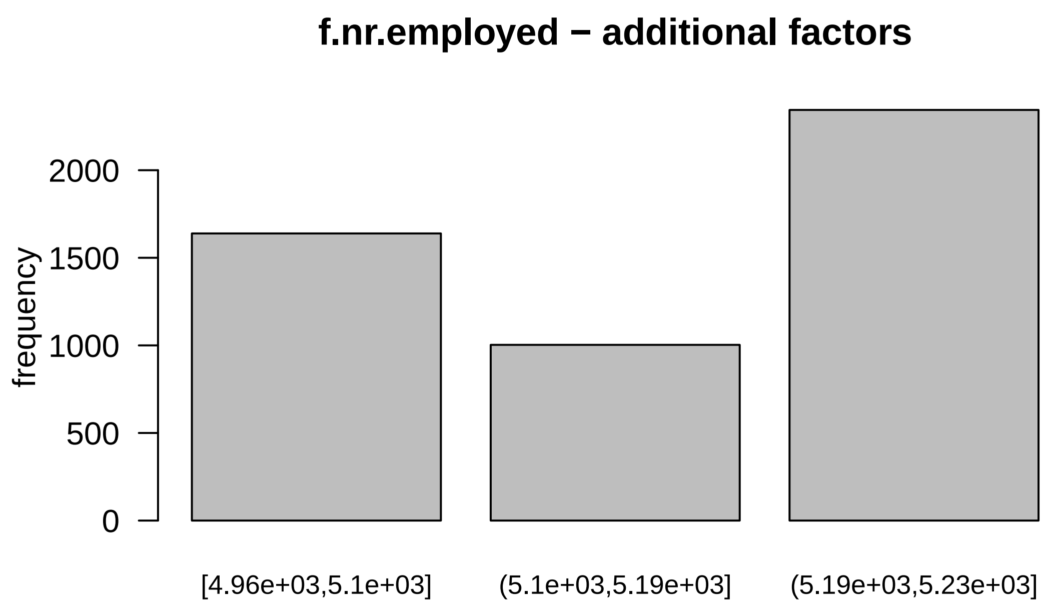
```
# EURIBOR3M
qulist<-quantile(df$euribor3m, seq(0,1,0.25), na.rm=TRUE)

df$f.euribor3m <-factor( cut(df$euribor3m  , breaks=unique(qulist), include.lowest=T) )

barplot(table(df$f.euribor3m), main="f.euribor3m - additional factors", ylab="frequency", las=1, cex.na
```




```
levels(df$f.euribor3m)<-paste0("f.euribor3m-", levels(df$f.euribor3m) )
summary(df$f.euribor3m)
```

```
## f.euribor3m-[0.635,1.33]  f.euribor3m-(1.33,4.86]  f.euribor3m-(4.86,4.96]
##                     1254                     1466                     1130
##     f.euribor3m-(4.96,5]
##                     1136
```

```
# NR.EMPLOYED
qulist<-quantile(df$nr.employed, seq(0,1,0.25), na.rm=TRUE)

df$f.nr.employed <-factor( cut(df$nr.employed  , breaks=unique(qulist), include.lowest=T) )

barplot(table(df$f.nr.employed), main="f.nr.employed - additional factors", ylab="frequency", las=1, ce
```




```r
levels(df$f.nr.employed)<-paste0("f.nr.employed-", levels(df$f.nr.employed) )
summary(df$f.nr.employed)
```

```
##  f.nr.employed-[4.96e+03,5.1e+03]  f.nr.employed-(5.1e+03,5.19e+03]
##                              1639                              1003
## f.nr.employed-(5.19e+03,5.23e+03]
##                              2344
```

Llistat de variables continues i discretes:

```r
vars<-names(df); vars
```

```
##  [1] "age"               "job"               "marital"
##  [4] "education"         "default"           "housing"
##  [7] "loan"              "contact"           "month"
## [10] "day_of_week"       "duration"          "campaign"
## [13] "pdays"             "previous"          "poutcome"
## [16] "emp.var.rate"      "cons.price.idx"    "cons.conf.idx"
## [19] "euribor3m"         "nr.employed"       "y"
## [22] "num_missings"      "num_outliers"      "num_errors"
## [25] "f.season"          "minutes"           "f.age"
## [28] "f.duration"        "f.campaign"        "f.pdays"
## [31] "f.previous"        "f.emp.var.rate"    "f.cons.price.idx"
## [34] "f.cons.conf.idx"   "f.euribor3m"       "f.nr.employed"
```

```r
# Variables continues
vars_con<-names(df)[c(1, 11:14, 16:20)]; vars_con
```

```
##  [1] "age"            "duration"       "campaign"       "pdays"
##  [5] "previous"       "emp.var.rate"   "cons.price.idx" "cons.conf.idx"
```

```
##  [9] "euribor3m"      "nr.employed"
```

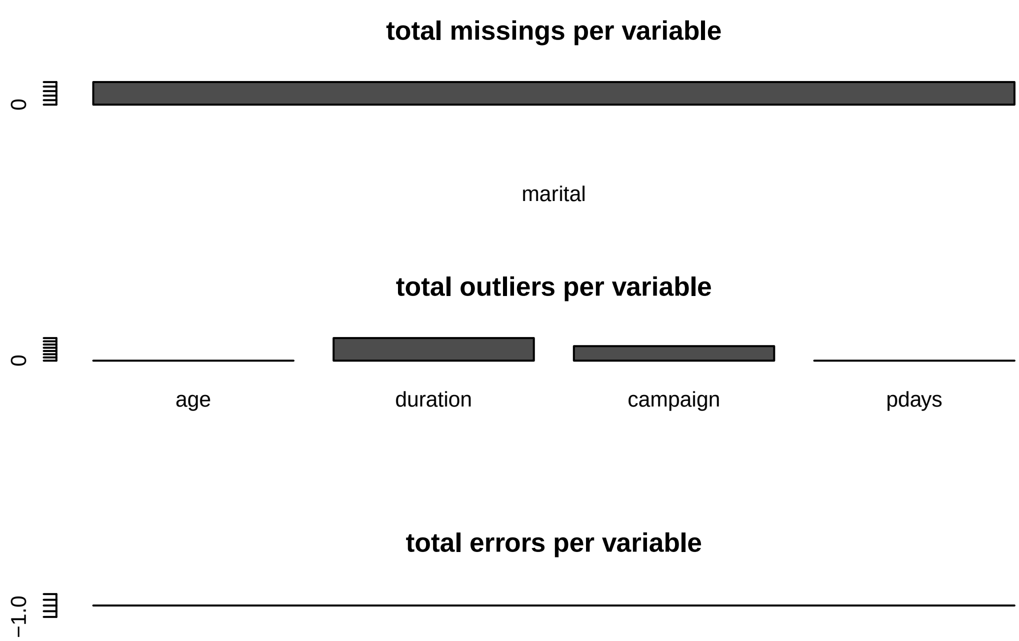
```r
# Variables discretes
vars_dis<-names(df)[c(2:10, 15, 21, 25, 27:36)]; vars_dis
```

```
##  [1] "job"               "marital"           "education"
##  [4] "default"           "housing"           "loan"
##  [7] "contact"           "month"             "day_of_week"
## [10] "poutcome"          "y"                 "f.season"
## [13] "f.age"             "f.duration"        "f.campaign"
## [16] "f.pdays"           "f.previous"        "f.emp.var.rate"
## [19] "f.cons.price.idx" "f.cons.conf.idx"  "f.euribor3m"
## [22] "f.nr.employed"
```

DATA QUALITY REPORT:

Per variable:

```r
par(mfrow=c(3,1))
barplot( t(c(missings[, 3])), main="total missings per variable", xlab="marital")
barplot( t(c(outliers[, c(1, 11, 12, 13)])), main="total outliers per variable")
barplot( t(c(errors[, 13])), main="total errors per variable")
```



Per individu:

Cap individu en tÃ© mÃ©s d'un. Es mostra en format taula el nÃºmero d'individus que tenen 0 i/o 1(o mÃ©s) missings, errors i outliers. Per Ãºltim, es mostren alguns dels individus que han tingut algun outlier i que aquest ha estat imputat.

```r
par(mfrow=c(1,1))
table(df$num_missings)
```

```
##
##    0    1
## 4839  147
```

```r
table(df$num_errors)
```

```
##
##    0
## 4986
```

```r
table(df$num_outliers)
```

```
##
##    0    1
## 4977    9
```

```r
head(df[which(df$num_outliers>0), ], 2) #individus amb algun outlier
```

```
##      age             job         marital                    education
## 5565  39      job-admin. marital-married education-university.degree
## 9014  30 job-blue-collar marital-married          education-basic.9y
##         default     housing    loan           contact     month
## 5565 default-no housing-yes loan-no contact-telephone month-may
## 9014 default-no  housing-no loan-no contact-telephone month-jun
##          day_of_week duration campaign pdays previous             poutcome
## 5565 day_of_week-mon       14       NA   999        0 poutcome-nonexistent
## 9014 day_of_week-thu       53       NA   999        0 poutcome-nonexistent
##      emp.var.rate cons.price.idx cons.conf.idx euribor3m nr.employed    y
## 5565          1.1         93.994         -36.4     4.857      5191.0 y-no
## 9014          1.4         94.465         -41.8     4.866      5228.1 y-no
##      num_missings num_outliers num_errors      f.season   minutes
## 5565            0            1          0 season-spring 0.2333333
## 9014            0            1          0 season-summer 0.8833333
##               f.age        f.duration f.campaign       f.pdays
## 5565 f.age-(38,47] f.duration-[5,66]       <NA> f.pdays-never
## 9014 f.age-[18,32] f.duration-[5,66]       <NA> f.pdays-never
##            f.previous            f.emp.var.rate           f.cons.price.idx
## 5565 f.previous-never f.emp.var.rate-(-0.1,1.1] f.cons.price.idx-(93.7,94]
## 9014 f.previous-never  f.emp.var.rate-(1.1,1.4] f.cons.price.idx-(94,94.8]
##                    f.cons.conf.idx             f.euribor3m
## 5565 f.cons.conf.idx-(-41.8,-36.4] f.euribor3m-(1.33,4.86]
## 9014 f.cons.conf.idx-(-42.7,-41.8] f.euribor3m-(4.86,4.96]
##                          f.nr.employed
## 5565  f.nr.employed-(5.1e+03,5.19e+03]
## 9014 f.nr.employed-(5.19e+03,5.23e+03]
```

Outliers Multivariants:

No hem aconseguit trobar una configuraciÃ³ del aq.plot que ens doni una bona grafica per a veure les distancies de Mahalanobis i detecter outliers multivariants.

```r
# Consider subset of numeric variables:
# summary(df[,vars_con])
vars_con_sub<-vars_con[c(1,2,3,6:10)]
```

```
x<-df[,vars_con_sub]
# aq.plot(x, delta=qchisq(0.995, df=ncol(x)) )
```

IMPUTATION:

Factors:

De totes les variables discretes que hem analitzat, hem vist que el "marital" status es podria imputar facilment amb imputeMCA(), ja que els unknown (passats previament a NA) corresponen nomes una petita part de la mostra. El mateix fem amb la variable "loan". Com hem vist previament, els unknowns han estat considerats categoria propia en altres variables.

```
res.impf<-imputeMCA(df[,vars_dis], ncp=10)

# Original:
summary(df$marital)
```

```
## marital-divorced  marital-married   marital-single             NA's
##              554             3046             1376               10
```

```
summary(df$loan)
```

```
##  loan-no loan-yes     NA's
##     4080      769      137
```

```
# Amb dades imputades:
summary(res.impf$completeObs$marital)
```

```
## marital-divorced  marital-married   marital-single
##              554             3055             1377
```

```
summary(res.impf$completeObs$loan)
```

```
##  loan-no loan-yes
##     4217      769
```

```
# Acceptem la imputaciÃ³:
df$loan<-res.impf$completeObs[,"marital"]
df$loan<-res.impf$completeObs[,"loan"]
#summary(df[,vars_dis])
```

Numeric Variables:

La variable numÃ¨rica campaign tÃ© certs individus que han estat considerats outliers prÃ¨viament. Aqui els imputem mitjaÃ§ant la imputaciÃ³ automatica imputePCA().

```
res.imp<-imputePCA(df[,vars_con], ncp=8)

# Original:
summary(df$campaign)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   1.000   2.000   2.535   3.000  25.000       9
```

```
# Amb dades imputades:

# Acceptem la imputaciÃ³:
```

```r
df$campaign<-res.imp$completeObs[,"campaign"]
#summary(df[,vars_con])
```

PROFILING:

CONTINOUS DESCRIPTION - Numeric Target (Duration):

La funciÃ³ d'R "condes" ens descriu la variable continua "duration" a partir d'altres variables quantitatives o de les variables categÃ²riques. AixÃ² ho fa mitjanÃ§ant els tres outputs diferenciats mÃ©s avall; etiquetats con a "*quanti*", "quali" i "$category".

El primer dels quals ($quanti) ens mostra la correlaciÃ³ de la variable estudiada "duration" amb altres variables numÃ¨riques, mostrant nomÃ©s les correlacions que tenen un p-value per sota del llindar o nivell de significaciÃ³ del 5% (en aquest cas). Com mÃ©s petit Ã©s el p-valor, menys evidÃ¨ncia hi ha de que la hipÃ²tesi nul.la sigui certa i mÃ©s segurs estem del rebuig de la hipÃ²tesi nul.la. Aquesta hipÃ²tesi nul.la H0 afirma que la correlaciÃ³ o resultat obtingut Ã©s fruit d'una aleatorietat de les dades i no pot ser atribuible a una causa especifica. Per tant, a partir d'ara, direm que quan el p-valor esta per sota del nivell de significaciÃ³ establert, els resultats sÃ³n significatius.

Comentar que ens apareix el valor NA perÃ² no tenim cap valor en la nostra mostra (ho vam estar mirant a classe), tot i aixi no afecta al resultat obtingut, simplement l'obviem. De la mateixa manera obviem la correlaciÃ³ d'1 entre la duraciÃ³ de la trucada en segons i en minuts, ja que Ã©s una correlaciÃ³ perfecta deguda a una conversiÃ³ d'unitats. Dit aixÃ², observem lleugeres correlacions negatives significatives (ordenades de mÃ©s correlaciÃ³ positiva a no correlaciÃ³ i desprÃ©s a mÃ©s correlaciÃ³ negativa) entre la duraciÃ³ de la trucada i la variable pdays, euribor3m, nr.employed i campaign. Es pot veure com la duraciÃ³ de la trucada augmenta com menys cops s'ha contactat al client en aquesta campanya (campaign), el quals Ã©s lÃ²gic perquÃ¨ un client molt contactat estara cansat ja de rebre trucades. TambÃ© es pot veure com la duraciÃ³ de la trucada augmenta com menys dies fa que s'ha contactat a un client en relaciÃ³ a una campanya anterior (pdays), el que pot estar relacionat amb l'interÃ¨s del client per les diferents campanyes actuals que se li han exposat. Finalment tenim dos indicados socioecÃ²nomics que tenen una lleugera correlaciÃ³ negativa amb la duraciÃ³ de la trucada.

El segon output ($quali) ens mostra els factors (variables categÃ²riques) que estan mÃ©s relacionades amb la variable target "duration". Ens mostra els resultats significatius ordenats per factors de mÃ©s a menys relacionats la duraciÃ³. Obviant la discretitzaciÃ³ de la duraciÃ³ (f.duration) que Ã²bviament esta molt relacionada, observem com la la decisiÃ³ final (y) del client a contractar un servei esta forÃ§a relacionada amb la duraciÃ³ d'una trucada. Molt menys relacionades (perÃ² lleugurament) ho estan les variables "f.campaign", "month", aixi com altres indicadors socioeconÃ²mics.

El tercer output ($category) ens indica una estimaciÃ³ de les unitats que la durada de la trucada esta per sobre (+) o per sota (-) de la mitja global quan el registre pertany a la categoria en qÃ¼estiÃ³; ordenades per p-valor. Deixant de banda les categories de f.duration que sÃ³n fruit de la discretitzaciÃ³, pot veure com quan el producte Ã©s contractat (y-yes), la duraciÃ³ de la trucada esta 148 segons per sobre, com era d'esperar en una contractaciÃ³ per telÃ¨fon. Altres resultats obtinguts interessants sÃ³n que la duraciÃ³ de la trucada esta 72 segons per sobre quan s'ha contactat amb el client en aquesta campanya 1 o 2 cops (f-campaign-[0,2]) i que tambÃ© augmenta en 38 segons quan el resultat de la campanya anterior va ser positiu pel mateix client (poutcome-success). TambÃ© podem destacar el mes d'abril (month-apr), en el qual les duracions de les trucades estan 28 segons per sobre de la mitja, o la primavera (season-spring) amb 18 segons per sobre de la mitja. D'altra banda podem veure com en el mes d'agost (month-aug) la duraciÃ³ de les trucades esta 28 segons per sota la mitja, en el novembre (month-nov) 20 segons per sota, i que els clients que mai han estat contactats abans (f.pdays-never) estan 28 segons menys al telÃ¨fon que la mitja.

El oneway.test d'R ens compara si dues o mÃ©s mostres de variables amb distribuciÃ³ normal tenen o no la mateixa mitjana (no cal assumir igualtat de variancies pels grups implicats que es comparen). En aquest cas ens permet concluore que la mitjana de la durada de la trucada en els casos que s'ha contractat el servei Ã©s significativament diferent a la dels casos en els quals no s'ha contractat el servei. L'estadadistic de contrast segueix una distribuciÃ³ F de Fisher i pren el valor 447.7, que Ã©s molt significatiu (p-value < 1e-16).

```
pos_duration<-which(names(df)=="duration"); pos_duration
```

```
## [1] 11
```

```
condes(df, num.var=pos_duration, proba = 0.05)
```

```
## $quanti
##              correlation      p.value
## <NA>                  NA           NA
## minutes       1.00000000 0.000000e+00
## pdays        -0.03478274 1.404179e-02
## euribor3m    -0.03512962 1.311237e-02
## num_outliers -0.04065979 4.085021e-03
## nr.employed  -0.04831097 6.438109e-04
## campaign     -0.07479201 1.241577e-07
##
## $quali
##                           R2       p.value
## f.duration       0.855794028  0.000000e+00
## y                0.164777620 3.759496e-197
## f.campaign       0.006187857  8.807648e-07
## f.cons.conf.idx  0.004067507  1.465565e-04
## f.nr.employed    0.002912867  6.975062e-04
## f.cons.price.idx 0.003246051  1.031905e-03
## month            0.005064462  2.674014e-03
## f.euribor3m      0.002462249  6.473152e-03
## f.season         0.002391458  7.627865e-03
## poutcome         0.001851161  9.887924e-03
## day_of_week      0.002352912  1.942616e-02
## f.pdays          0.001214169  4.846375e-02
## f.emp.var.rate   0.001574759  4.916221e-02
##
## $category
##                                     Estimate       p.value
## f.duration-(482,1.58e+03]         493.613665  0.000000e+00
## y-yes                             148.441504 3.759496e-197
## f.duration-(316,482]              134.394010  8.476109e-56
## f.campaign-(5,25]                  14.794426  2.638343e-06
## season-spring                      17.952283  5.877554e-04
## poutcome-success                   38.359032  5.480212e-03
## f.campaign-[0,2]                   71.765001  7.136472e-03
## f.nr.employed-[4.96e+03,5.1e+03]    9.017147  8.355482e-03
## f.duration-(235,316]               22.169724  9.317648e-03
## f.cons.conf.idx-[-50.8,-42.7]      14.076002  1.238528e-02
## NA                                132.886872  1.491425e-02
## month-may                           9.867780  1.599295e-02
## f.cons.price.idx-(93.7,94]         11.621760  2.081111e-02
## f.pdays-[0,7]                      16.460640  2.262020e-02
## f.cons.conf.idx-(-41.8,-36.4]      16.349262  2.392080e-02
## month-apr                          27.731238  2.403940e-02
## education-high.school               9.358222  4.228302e-02
## day_of_week-wed                    13.376659  4.495212e-02
```

```
## month-nov                          -20.376410  4.421467e-02
## education-university.degree        -14.109465  2.294239e-02
## f.emp.var.rate-(1.1,1.4]           -10.129703  2.036833e-02
## day_of_week-mon                    -15.133836  1.838350e-02
## season-summer                       -3.899443  1.752241e-02
## f.pdays-never                      -27.755294  1.396985e-02
## f.cons.conf.idx-(-36.4,-26.9]      -14.862166  7.024095e-03
## f.cons.conf.idx-(-42.7,-41.8]      -15.563098  4.192506e-03
## NA                                -154.540521  4.085021e-03
## f.euribor3m-(4.96,5]               -19.423787  1.079935e-03
## month-aug                          -28.383026  6.707022e-04
## f.nr.employed-(5.19e+03,5.23e+03]  -16.466612  1.395228e-04
## f.cons.price.idx-(93.1,93.7]       -22.699701  8.027710e-05
## f.duration-(177,235]               -47.149040  5.572506e-08
## f.duration-(138,177]               -94.204089  1.668437e-27
## f.duration-(101,138]              -131.656740  5.328783e-54
## f.duration-(66,101]               -167.038569  1.102835e-85
## f.duration-[5,66]                 -210.128961 1.924209e-141
## y-no                              -148.441504 3.759496e-197
```

```r
# mitjana de la duraciÃ³ per categoria de la duraciÃ³
# tapply(df$duration, df$f.duration, mean)

# duraciÃ³ global
summary(df$duration)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     5.0   101.0   177.0   250.6   316.0  1580.0
```

```r
# mitjana de la duraciÃ³ per categoria de la y
tapply(df$duration, df$y, mean)
```

```
##     y-no    y-yes
## 217.4563 514.3393
```

```r
oneway.test(df$duration~df$y)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  df$duration and df$y
## F = 447.7, num df = 1.00, denom df = 605.83, p-value < 2.2e-16
```

CATEGORICAL DESCRIPTION - Factor (Y, Final Decision):

La funciÃ³ d'R "catdes" ens descriu la variable categÃ²rica "y" a partir d'altres variables categÃ²riques o de les variables quantitatives. AixÃ² ho fa mitjanÃ§ant outputs diferenciats mÃ©s avall. Notem que el nostre llindar de significaciÃ³ en aquest cas Ã©s del 0.025 per tal de limitar una mica la gran quantitat de resultats mostrats.

L'apartat "Link between the cluster variable and the categorical variables (chi-square test)" ens mostra les variables categÃ²riques que han caracteritzat al factor "y" ordenades de mÃ©s a menys caracteritzaciÃ³ del factor (de menys a mÃ©s p-value). La columna "df" mostra els Degrees of Freedom, que corresponen amb el nombre de categories del factor menys 1. Les variables categÃ²riques que han influenciat mÃ©s en la decisiÃ³ final (y) sÃ³n la f.duration (perÃ² Ã©s una dada que s'obtÃ© a posteriori de la trucada, no ens serveix per a generar un perfil de client), f.pdays (nombre de dies des de l'Ãºltim contacte), poutcome (si la Ãºltima campanya va ser acceptada per aquest client o no), el mes (month), previous (si havia estat contactat o no abans d'aquesta campanya), diferents indicadors socioeconÃ²mics, contact (via de contacte), el job (feina), etc.

L'apartat "Description of each cluster by the categories" ens mostra per a cada categoria de la "y" (y-yes, y-no), una descripciÃ³ de les variables categÃ²riques per tal de poder estudiar-ne el seu enllaÃ§. La primera columna, Cla/Mod, en mostra el tant percent d'individus corresponents a la categoria de la fila que pertanyen tambÃ© al cluster. Per altra banda, la segona columna (Mod/Cla) mostra la operaciÃ³ inversa, es a dir, dels individus que pertanyen al cluster quin tant percent pertany tambÃ© a la categoria de la fila en questio. La columna Global ens mostra quin tant percent de la poblaciÃ³ total posseix la caracteristica de la fila. Per acabar, v.test ens indica si la categoria de la fila es troba sobre representada (v.test>0) o menys representada (v.test<0) entre els individus que conformen el cluster. Per al cluster "y-no" podem deduir que el fet de la haver realitzat un contacte previ i que es realitzin menys de 2 contactes en l'actual campanya tÃ© un gran pes en la decisiÃ³ final. A mÃ©s, per a valors positius del euribor (desfavorables per al client) hi a una tendencia a l'alÃ§a a refusar el producte. El mateix succeix amb valor elevats del IPC i quan es redueix la taxa d'ocupaciÃ³. En relaciÃ³ al cluster "y-yes" les presones que van ser contactades fa menys de 7 dies per altres campanyes i aquelles que el resultat en la campanya anterior va ser exitÃ³s tendeixen a donar un si com a resposta. TambÃ© quan els valors socioeconomics es mostren favorables al client aquest tendeix a adquirir el producte amb una major probabilitat.

L'apartat "Link between the cluster variable and the quantitative variables" ens mostra les variables quantitatives que han caracteritzat al factor "y" ordenades de mÃ©s a menys caracteritzaciÃ³ del factor (de menys a mÃ©s correlaciÃ³). Les variables quantitatives que han influenciat mÃ©s en la decisiÃ³ final (y) sÃ³n la duration i minutes (perÃ² son dades que s'obtenen a posteriori de la trucada, no ens serveixen per a generar un perfil de client), pdays (nombre de dies des de l'Ãºltim contacte), previous (si havia estat contactat o no abans d'aquesta campanya), diferents indicadors socioeconÃ²mics, etc.

L'apartat "Description of each cluster by quantitative variables". D'aquesta part de l'analisi no en podem extreure informaciÃ³ dels individus que conformen el cluster "y-no", donat que els valors que es presenten de les categories dins el cluster i de manera general no presenten una diversificaciÃ³ notable. Per altra banda del cluster "y-yes" si que en poden extreure informaciÃ³, podem veure que la mitjana de la duracio de les trucades dels individus del cluster duplica la mitjana global (tot i que la duraciÃ³ no ens pot servir com a indicador, donat que aquelles persones no interesades en el producte refusen la trucada mÃ©s rapidament o el fet de contractar el prouncte pot repercutir en un increment de la durada). Altres factors con l'euribor o la taxa de variacio de la ocupaciÃ³ tambÃ© tenen un impacte en la decisiÃ³ final.

```
pos_y<-which(names(df)=="y"); pos_y
```

```
## [1] 21
```

```
catdes(df, num.var=pos_y, proba = 0.025)
```

```
##
## Link between the cluster variable and the categorical variables (chi-square test)
## =================================================================================
##                        p.value df
## f.duration       2.794524e-159  7
## f.pdays          9.362887e-100  2
## poutcome          3.053387e-95  2
## f.nr.employed     1.703080e-89  2
## f.euribor3m       5.470503e-79  3
## month             1.690776e-65  9
## f.emp.var.rate    7.969229e-62  3
## f.previous        5.590487e-45  2
## f.cons.price.idx  5.572278e-38  3
## f.cons.conf.idx   4.786677e-23  3
## contact           2.110136e-21  1
## job               8.420857e-16 11
```

```
## default          9.768051e-13  1
## f.season         1.176664e-10  3
## f.age            7.936723e-09  3
## education        6.361426e-06  6
## marital          1.452705e-04  3
## f.campaign       1.037416e-03  3
##
## Description of each cluster by the categories
## =============================================
## $`y-no`
##                                                   Cla/Mod    Mod/Cla
## f.pdays=f.pdays-never                            90.64255 98.4195078
## f.nr.employed=f.nr.employed-(5.19e+03,5.23e+03] 94.70990 50.1241815
## f.previous=f.previous-never                      91.01769 89.4558591
## poutcome=poutcome-nonexistent                    91.01769 89.4558591
## f.duration=f.duration-[5,66]                     99.52607 14.2244299
## f.emp.var.rate=f.emp.var.rate-(1.1,1.4]          94.52055 40.5057575
## contact=contact-telephone                        94.31330 39.6929329
## f.duration=f.duration-(66,101]                   98.38449 13.7502822
## f.cons.price.idx=f.cons.price.idx-(93.7,94]      94.11765 38.6543238
## f.nr.employed=f.nr.employed-(5.1e+03,5.19e+03]   96.11167 21.7656356
## f.emp.var.rate=f.emp.var.rate-(-0.1,1.1]         96.10778 21.7430571
## f.cons.conf.idx=f.cons.conf.idx-(-42.7,-41.8]    96.07032 20.9753895
## default=default-unknown                          95.05814 22.1494694
## month=month-may                                  93.33716 36.6899977
## f.euribor3m=f.euribor3m-(4.86,4.96]              94.51327 24.1137954
## f.euribor3m=f.euribor3m-(4.96,5]                 94.36620 24.2041093
## f.duration=f.duration-(101,138]                  96.01911 13.6148115
## job=job-blue-collar                              93.74457 24.3621585
## f.euribor3m=f.euribor3m-(1.33,4.86]              92.70123 30.6841273
## f.duration=f.duration-(138,177]                  94.79675 13.1632423
## f.cons.price.idx=f.cons.price.idx-(93.1,93.7]    92.90976 22.7816663
## f.age=f.age-(38,47]                              92.54098 25.4910815
## f.campaign=f.campaign-(5,25]                     94.18886  8.7830210
## education=education-basic.9y                     92.72727 14.9695191
## marital=marital-married                          89.92121 61.8424023
## month=month-jul                                  91.31484 17.0918943
## education=education-basic.6y                     93.07958  6.0736058
## f.season=season-spring                           90.08030 43.0571235
## f.age=f.age-(32,38]                              90.62241 24.6556785
## poutcome=poutcome-failure                        85.53459  9.2120117
## education=education-unknown                      82.68398  4.3124859
## f.cons.price.idx=f.cons.price.idx-(94,94.8]      85.41667 12.9600361
## f.campaign=f.campaign-[0,2]                      87.94222 67.3515466
## f.season=season-winter                           65.38462  0.3838338
## month=month-dec                                  65.38462  0.3838338
## education=education-university.degree            86.51226 28.6746444
## f.emp.var.rate=f.emp.var.rate-(-1.8,-0.1]        84.09475 11.2214947
## f.duration=f.duration-(316,482]                  83.87097 11.7407993
## job=job-retired                                  78.92157  3.6351321
## marital=marital-single                           85.68314 26.6200045
## f.age=f.age-[18,32]                              85.35503 26.0555430
## f.pdays=f.pdays-(>7)                             53.33333  0.3612554
## job=job-student                                  70.00000  1.5804922
```

```
## month=month-apr                                  78.70968  5.5091443
## f.season=season-autumn                           82.25564 12.3504177
## month=month-sep                                  57.37705  0.7902461
## month=month-mar                                  57.57576  0.8579815
## f.cons.conf.idx=f.cons.conf.idx-(-36.4,-26.9]    81.22318 17.0918943
## default=default-no                               87.20283 77.8505306
## f.previous=f.previous-1                          77.53906  8.9636487
## month=month-oct                                  54.63918  1.1966584
## f.previous=f.previous-(>1)                       57.85124  1.5804922
## contact=contact-cellular                         85.55413 60.3070671
## f.cons.price.idx=f.cons.price.idx-[92.2,93.1]    80.48261 25.6039738
## f.emp.var.rate=f.emp.var.rate-[-3.4,-1.8]        78.59532 26.5296907
## f.pdays=f.pdays-[0,7]                            36.73469  1.2192368
## poutcome=poutcome-success                        37.82051  1.3321291
## f.euribor3m=f.euribor3m-[0.635,1.33]             74.16268 20.9979679
## f.nr.employed=f.nr.employed-[4.96e+03,5.1e+03]   75.96095 28.1101829
## f.duration=f.duration-(482,1.58e+03]             59.13462  8.3314518
##                                                      Global      p.value
## f.pdays=f.pdays-never                            96.4500602  2.410684e-59
## f.nr.employed=f.nr.employed-(5.19e+03,5.23e+03] 47.0116326  2.158488e-37
## f.previous=f.previous-never                      87.3044525  1.438650e-30
## poutcome=poutcome-nonexistent                    87.3044525  1.438650e-30
## f.duration=f.duration-[5,66]                     12.6955475  1.487124e-30
## f.emp.var.rate=f.emp.var.rate-(1.1,1.4]          38.0665864  1.340920e-25
## contact=contact-telephone                        37.3846771  3.447929e-23
## f.duration=f.duration-(66,101]                   12.4147613  7.696941e-22
## f.cons.price.idx=f.cons.price.idx-(93.7,94]      36.4821500  7.057265e-21
## f.nr.employed=f.nr.employed-(5.1e+03,5.19e+03]   20.1163257  1.424235e-19
## f.emp.var.rate=f.emp.var.rate-(-0.1,1.1]         20.0962696  1.574618e-19
## f.cons.conf.idx=f.cons.conf.idx-(-42.7,-41.8]    19.3943041  1.401017e-18
## default=default-unknown                          20.6979543  1.230324e-14
## month=month-may                                  34.9177698  1.726364e-14
## f.euribor3m=f.euribor3m-(4.86,4.96]              22.6634577  1.693548e-13
## f.euribor3m=f.euribor3m-(4.96,5]                 22.7837946  6.639818e-13
## f.duration=f.duration-(101,138]                  12.5952667  1.010774e-11
## job=job-blue-collar                              23.0846370  1.884818e-10
## f.euribor3m=f.euribor3m-(1.33,4.86]              29.4023265  6.796806e-09
## f.duration=f.duration-(138,177]                  12.3345367  5.342775e-08
## f.cons.price.idx=f.cons.price.idx-(93.1,93.7]    21.7809868  4.701642e-07
## f.age=f.age-(38,47]                              24.4685118  9.135370e-07
## f.campaign=f.campaign-(5,25]                      8.2831929  1.084374e-04
## education=education-basic.9y                     14.3401524  1.876745e-04
## marital=marital-married                          61.0910550  2.314946e-03
## month=month-jul                                  16.6265544  1.093857e-02
## education=education-basic.6y                      5.7962294  1.335614e-02
## f.season=season-spring                           42.4588849  1.562952e-02
## f.age=f.age-(32,38]                              24.1676695  2.153346e-02
## poutcome=poutcome-failure                         9.5667870  1.986516e-02
## education=education-unknown                       4.6329723  4.270710e-03
## f.cons.price.idx=f.cons.price.idx-(94,94.8]      13.4777377  3.445794e-03
## f.campaign=f.campaign-[0,2]                      68.0304854  3.359672e-03
## f.season=season-winter                            0.5214601  1.657365e-03
## month=month-dec                                   0.5214601  1.657365e-03
## education=education-university.degree            29.4424388  9.565525e-04
```

```
## f.emp.var.rate=f.emp.var.rate-(-1.8,-0.1]       11.8531889  1.984797e-04
## f.duration=f.duration-(316,482]                 12.4348175  6.392065e-05
## job=job-retired                                  4.0914561  2.982842e-05
## marital=marital-single                          27.5972724  2.055013e-05
## f.age=f.age-[18,32]                             27.1159246  3.567657e-06
## f.pdays=f.pdays-(>7)                             0.6016847  1.202754e-06
## job=job-student                                  2.0056157  2.508620e-07
## month=month-apr                                  6.2174087  1.047741e-07
## f.season=season-autumn                          13.3373446  5.062563e-08
## month=month-sep                                  1.2234256  3.276634e-10
## month=month-mar                                  1.3237064  7.597160e-11
## f.cons.conf.idx=f.cons.conf.idx-(-36.4,-26.9]   18.6923385  1.352020e-14
## default=default-no                              79.3020457  1.230324e-14
## f.previous=f.previous-1                         10.2687525  7.464256e-15
## month=month-oct                                  1.9454473  8.959508e-18
## f.previous=f.previous-(>1)                       2.4267950  1.002106e-18
## contact=contact-cellular                        62.6153229  3.447929e-23
## f.cons.price.idx=f.cons.price.idx-[92.2,93.1]   28.2591256  3.335427e-29
## f.emp.var.rate=f.emp.var.rate-[-3.4,-1.8]       29.9839551  1.289177e-46
## f.pdays=f.pdays-[0,7]                            2.9482551  6.682675e-54
## poutcome=poutcome-success                        3.1287605  2.946325e-55
## f.euribor3m=f.euribor3m-[0.635,1.33]            25.1504212  3.042037e-70
## f.nr.employed=f.nr.employed-[4.96e+03,5.1e+03]  32.8720417  1.759629e-84
## f.duration=f.duration-(482,1.58e+03]            12.5150421 4.894928e-100
##                                                     v.test
## f.pdays=f.pdays-never                            16.245323
## f.nr.employed=f.nr.employed-(5.19e+03,5.23e+03]  12.778626
## f.previous=f.previous-never                      11.492513
## poutcome=poutcome-nonexistent                    11.492513
## f.duration=f.duration-[5,66]                     11.489650
## f.emp.var.rate=f.emp.var.rate-(1.1,1.4]          10.458406
## contact=contact-telephone                         9.918824
## f.duration=f.duration-(66,101]                    9.603908
## f.cons.price.idx=f.cons.price.idx-(93.7,94]       9.372891
## f.nr.employed=f.nr.employed-(5.1e+03,5.19e+03]    9.050417
## f.emp.var.rate=f.emp.var.rate-(-0.1,1.1]          9.039450
## f.cons.conf.idx=f.cons.conf.idx-(-42.7,-41.8]     8.797336
## default=default-unknown                           7.712857
## month=month-may                                   7.669524
## f.euribor3m=f.euribor3m-(4.86,4.96]               7.370998
## f.euribor3m=f.euribor3m-(4.96,5]                  7.186654
## f.duration=f.duration-(101,138]                   6.804960
## job=job-blue-collar                               6.370444
## f.euribor3m=f.euribor3m-(1.33,4.86]               5.795870
## f.duration=f.duration-(138,177]                   5.439509
## f.cons.price.idx=f.cons.price.idx-(93.1,93.7]     5.038105
## f.age=f.age-(38,47]                               4.909404
## f.campaign=f.campaign-(5,25]                      3.870893
## education=education-basic.9y                      3.735055
## marital=marital-married                           3.046536
## month=month-jul                                   2.544655
## education=education-basic.6y                      2.474129
## f.season=season-spring                            2.417454
## f.age=f.age-(32,38]                               2.298498
```

```
## poutcome=poutcome-failure                         -2.328885
## education=education-unknown                       -2.857442
## f.cons.price.idx=f.cons.price.idx-(94,94.8]       -2.924889
## f.campaign=f.campaign-[0,2]                       -2.932757
## f.season=season-winter                            -3.145618
## month=month-dec                                   -3.145618
## education=education-university.degree             -3.303003
## f.emp.var.rate=f.emp.var.rate-(-1.8,-0.1]         -3.720944
## f.duration=f.duration-(316,482]                   -3.997849
## job=job-retired                                   -4.174772
## marital=marital-single                            -4.258828
## f.age=f.age-[18,32]                               -4.635100
## f.pdays=f.pdays-(>7)                              -4.855183
## job=job-student                                   -5.157057
## month=month-apr                                   -5.318243
## f.season=season-autumn                            -5.449099
## month=month-sep                                   -6.285090
## month=month-mar                                   -6.508368
## f.cons.conf.idx=f.cons.conf.idx-(-36.4,-26.9]     -7.700814
## default=default-no                                -7.712857
## f.previous=f.previous-1                           -7.776358
## month=month-oct                                   -8.586582
## f.previous=f.previous-(>1)                        -8.834875
## contact=contact-cellular                          -9.918824
## f.cons.price.idx=f.cons.price.idx-[92.2,93.1]    -11.217779
## f.emp.var.rate=f.emp.var.rate-[-3.4,-1.8]        -14.336770
## f.pdays=f.pdays-[0,7]                            -15.457815
## poutcome=poutcome-success                        -15.657639
## f.euribor3m=f.euribor3m-[0.635,1.33]             -17.718064
## f.nr.employed=f.nr.employed-[4.96e+03,5.1e+03]  -19.475855
## f.duration=f.duration-(482,1.58e+03]             -21.231431
##
## $`y-yes`
##                                                    Cla/Mod    Mod/Cla
## f.duration=f.duration-(482,1.58e+03]             40.8653846 45.7809695
## f.nr.employed=f.nr.employed-[4.96e+03,5.1e+03]  24.0390482 70.7360862
## f.euribor3m=f.euribor3m-[0.635,1.33]             25.8373206 58.1687612
## poutcome=poutcome-success                        62.1794872 17.4147217
## f.pdays=f.pdays-[0,7]                            63.2653061 16.6965889
## f.emp.var.rate=f.emp.var.rate-[-3.4,-1.8]        21.4046823 57.4506284
## f.cons.price.idx=f.cons.price.idx-[92.2,93.1]    19.5173882 49.3716338
## contact=contact-cellular                         14.4458680 80.9694794
## f.previous=f.previous-(>1)                       42.1487603  9.1561939
## month=month-oct                                  45.3608247  7.8994614
## f.previous=f.previous-1                          22.4609375 20.6463196
## default=default-no                               12.7971674 90.8438061
## f.cons.conf.idx=f.cons.conf.idx-(-36.4,-26.9]    18.7768240 31.4183124
## month=month-mar                                  42.4242424  5.0269300
## month=month-sep                                  42.6229508  4.6678636
## f.season=season-autumn                           17.7443609 21.1849192
## month=month-apr                                  21.2903226 11.8491921
## job=job-student                                  30.0000000  5.3859964
## f.pdays=f.pdays-(>7)                             46.6666667  2.5134650
## f.age=f.age-[18,32]                              14.6449704 35.5475763
```

```
## marital=marital-single                        14.3168605 35.3680431
## job=job-retired                               21.0784314  7.7199282
## f.duration=f.duration-(316,482]               16.1290323 17.9533214
## f.emp.var.rate=f.emp.var.rate-(-1.8,-0.1]     15.9052453 16.8761221
## education=education-university.degree         13.4877384 35.5475763
## f.season=season-winter                        34.6153846  1.6157989
## month=month-dec                               34.6153846  1.6157989
## f.campaign=f.campaign-[0,2]                   12.0577830 73.4290844
## f.cons.price.idx=f.cons.price.idx-(94,94.8]   14.5833333 17.5942549
## education=education-unknown                   17.3160173  7.1813285
## poutcome=poutcome-failure                     14.4654088 12.3877917
## f.age=f.age-(32,38]                            9.3775934 20.2872531
## f.season=season-spring                         9.9196977 37.7019749
## education=education-basic.6y                   6.9204152  3.5906643
## month=month-jul                                8.6851628 12.9263914
## marital=marital-married                       10.0787919 55.1166966
## education=education-basic.9y                   7.2727273  9.3357271
## f.campaign=f.campaign-(5,25]                   5.8111380  4.3087971
## f.age=f.age-(38,47]                            7.4590164 16.3375224
## f.cons.price.idx=f.cons.price.idx-(93.1,93.7]  7.0902394 13.8240575
## f.duration=f.duration-(138,177]                5.2032520  5.7450628
## f.euribor3m=f.euribor3m-(1.33,4.86]            7.2987722 19.2100539
## job=job-blue-collar                            6.2554301 12.9263914
## f.duration=f.duration-(101,138]                3.9808917  4.4883303
## f.euribor3m=f.euribor3m-(4.96,5]               5.6338028 11.4901257
## f.euribor3m=f.euribor3m-(4.86,4.96]            5.4867257 11.1310592
## month=month-may                                6.6628374 20.8258528
## default=default-unknown                        4.9418605  9.1561939
## f.cons.conf.idx=f.cons.conf.idx-(-42.7,-41.8]  3.9296794  6.8222621
## f.emp.var.rate=f.emp.var.rate-(-0.1,1.1]       3.8922156  7.0017953
## f.nr.employed=f.nr.employed-(5.1e+03,5.19e+03]  3.8883350  7.0017953
## f.cons.price.idx=f.cons.price.idx-(93.7,94]    5.8823529 19.2100539
## f.duration=f.duration-(66,101]                 1.6155089  1.7953321
## contact=contact-telephone                      5.6866953 19.0305206
## f.emp.var.rate=f.emp.var.rate-(1.1,1.4]        5.4794521 18.6714542
## f.duration=f.duration-[5,66]                   0.4739336  0.5385996
## f.previous=f.previous-never                    8.9823110 70.1974865
## poutcome=poutcome-nonexistent                  8.9823110 70.1974865
## f.nr.employed=f.nr.employed-(5.19e+03,5.23e+03]  5.2901024 22.2621185
## f.pdays=f.pdays-never                          9.3574548 80.7899461
##                                                   Global      p.value
## f.duration=f.duration-(482,1.58e+03]          12.5150421 4.894928e-100
## f.nr.employed=f.nr.employed-[4.96e+03,5.1e+03]  32.8720417  1.759629e-84
## f.euribor3m=f.euribor3m-[0.635,1.33]          25.1504212  3.042037e-70
## poutcome=poutcome-success                      3.1287605  2.946325e-55
## f.pdays=f.pdays-[0,7]                          2.9482551  6.682675e-54
## f.emp.var.rate=f.emp.var.rate-[-3.4,-1.8]     29.9839551  1.289177e-46
## f.cons.price.idx=f.cons.price.idx-[92.2,93.1]  28.2591256  3.335427e-29
## contact=contact-cellular                      62.6153229  3.447929e-23
## f.previous=f.previous-(>1)                     2.4267950  1.002106e-18
## month=month-oct                                1.9454473  8.959508e-18
## f.previous=f.previous-1                       10.2687525  7.464256e-15
## default=default-no                            79.3020457  1.230324e-14
## f.cons.conf.idx=f.cons.conf.idx-(-36.4,-26.9]  18.6923385  1.352020e-14
```

```
## month=month-mar                                   1.3237064  7.597160e-11
## month=month-sep                                   1.2234256  3.276634e-10
## f.season=season-autumn                           13.3373446  5.062563e-08
## month=month-apr                                   6.2174087  1.047741e-07
## job=job-student                                   2.0056157  2.508620e-07
## f.pdays=f.pdays-(>7)                              0.6016847  1.202754e-06
## f.age=f.age-[18,32]                              27.1159246  3.567657e-06
## marital=marital-single                           27.5972724  2.055013e-05
## job=job-retired                                   4.0914561  2.982842e-05
## f.duration=f.duration-(316,482]                  12.4348175  6.392065e-05
## f.emp.var.rate=f.emp.var.rate-(-1.8,-0.1]        11.8531889  1.984797e-04
## education=education-university.degree            29.4424388  9.565525e-04
## f.season=season-winter                            0.5214601  1.657365e-03
## month=month-dec                                   0.5214601  1.657365e-03
## f.campaign=f.campaign-[0,2]                      68.0304854  3.359672e-03
## f.cons.price.idx=f.cons.price.idx-(94,94.8]      13.4777377  3.445794e-03
## education=education-unknown                       4.6329723  4.270710e-03
## poutcome=poutcome-failure                         9.5667870  1.986516e-02
## f.age=f.age-(32,38]                              24.1676695  2.153346e-02
## f.season=season-spring                           42.4588849  1.562952e-02
## education=education-basic.6y                      5.7962294  1.335614e-02
## month=month-jul                                  16.6265544  1.093857e-02
## marital=marital-married                          61.0910550  2.314946e-03
## education=education-basic.9y                     14.3401524  1.876745e-04
## f.campaign=f.campaign-(5,25]                      8.2831929  1.084374e-04
## f.age=f.age-(38,47]                              24.4685118  9.135370e-07
## f.cons.price.idx=f.cons.price.idx-(93.1,93.7]    21.7809868  4.701642e-07
## f.duration=f.duration-(138,177]                  12.3345367  5.342775e-08
## f.euribor3m=f.euribor3m-(1.33,4.86]              29.4023265  6.796806e-09
## job=job-blue-collar                              23.0846370  1.884818e-10
## f.duration=f.duration-(101,138]                  12.5952667  1.010774e-11
## f.euribor3m=f.euribor3m-(4.96,5]                 22.7837946  6.639818e-13
## f.euribor3m=f.euribor3m-(4.86,4.96]              22.6634577  1.693548e-13
## month=month-may                                  34.9177698  1.726364e-14
## default=default-unknown                          20.6979543  1.230324e-14
## f.cons.conf.idx=f.cons.conf.idx-(-42.7,-41.8]    19.3943041  1.401017e-18
## f.emp.var.rate=f.emp.var.rate-(-0.1,1.1]         20.0962696  1.574618e-19
## f.nr.employed=f.nr.employed-(5.1e+03,5.19e+03]   20.1163257  1.424235e-19
## f.cons.price.idx=f.cons.price.idx-(93.7,94]      36.4821500  7.057265e-21
## f.duration=f.duration-(66,101]                   12.4147613  7.696941e-22
## contact=contact-telephone                        37.3846771  3.447929e-23
## f.emp.var.rate=f.emp.var.rate-(1.1,1.4]          38.0665864  1.340920e-25
## f.duration=f.duration-[5,66]                     12.6955475  1.487124e-30
## f.previous=f.previous-never                      87.3044525  1.438650e-30
## poutcome=poutcome-nonexistent                    87.3044525  1.438650e-30
## f.nr.employed=f.nr.employed-(5.19e+03,5.23e+03] 47.0116326  2.158488e-37
## f.pdays=f.pdays-never                            96.4500602  2.410684e-59
##                                                     v.test
## f.duration=f.duration-(482,1.58e+03]              21.231431
## f.nr.employed=f.nr.employed-[4.96e+03,5.1e+03]   19.475855
## f.euribor3m=f.euribor3m-[0.635,1.33]              17.718064
## poutcome=poutcome-success                         15.657639
## f.pdays=f.pdays-[0,7]                             15.457815
## f.emp.var.rate=f.emp.var.rate-[-3.4,-1.8]         14.336770
```

```
## f.cons.price.idx=f.cons.price.idx-[92.2,93.1]    11.217779
## contact=contact-cellular                           9.918824
## f.previous=f.previous-(>1)                         8.834875
## month=month-oct                                    8.586582
## f.previous=f.previous-1                            7.776358
## default=default-no                                 7.712857
## f.cons.conf.idx=f.cons.conf.idx-(-36.4,-26.9]      7.700814
## month=month-mar                                    6.508368
## month=month-sep                                    6.285090
## f.season=season-autumn                             5.449099
## month=month-apr                                    5.318243
## job=job-student                                    5.157057
## f.pdays=f.pdays-(>7)                               4.855183
## f.age=f.age-[18,32]                                4.635100
## marital=marital-single                             4.258828
## job=job-retired                                    4.174772
## f.duration=f.duration-(316,482]                    3.997849
## f.emp.var.rate=f.emp.var.rate-(-1.8,-0.1]          3.720944
## education=education-university.degree             3.303003
## f.season=season-winter                             3.145618
## month=month-dec                                    3.145618
## f.campaign=f.campaign-[0,2]                        2.932757
## f.cons.price.idx=f.cons.price.idx-(94,94.8]        2.924889
## education=education-unknown                        2.857442
## poutcome=poutcome-failure                          2.328885
## f.age=f.age-(32,38]                               -2.298498
## f.season=season-spring                            -2.417454
## education=education-basic.6y                      -2.474129
## month=month-jul                                   -2.544655
## marital=marital-married                           -3.046536
## education=education-basic.9y                      -3.735055
## f.campaign=f.campaign-(5,25]                      -3.870893
## f.age=f.age-(38,47]                               -4.909404
## f.cons.price.idx=f.cons.price.idx-(93.1,93.7]     -5.038105
## f.duration=f.duration-(138,177]                   -5.439509
## f.euribor3m=f.euribor3m-(1.33,4.86]               -5.795870
## job=job-blue-collar                               -6.370444
## f.duration=f.duration-(101,138]                   -6.804960
## f.euribor3m=f.euribor3m-(4.96,5]                  -7.186654
## f.euribor3m=f.euribor3m-(4.86,4.96]               -7.370998
## month=month-may                                   -7.669524
## default=default-unknown                           -7.712857
## f.cons.conf.idx=f.cons.conf.idx-(-42.7,-41.8]     -8.797336
## f.emp.var.rate=f.emp.var.rate-(-0.1,1.1]          -9.039450
## f.nr.employed=f.nr.employed-(5.1e+03,5.19e+03]    -9.050417
## f.cons.price.idx=f.cons.price.idx-(93.7,94]       -9.372891
## f.duration=f.duration-(66,101]                    -9.603908
## contact=contact-telephone                         -9.918824
## f.emp.var.rate=f.emp.var.rate-(1.1,1.4]          -10.458406
## f.duration=f.duration-[5,66]                     -11.489650
## f.previous=f.previous-never                      -11.492513
## poutcome=poutcome-nonexistent                    -11.492513
## f.nr.employed=f.nr.employed-(5.19e+03,5.23e+03] -12.778626
## f.pdays=f.pdays-never                            -16.245323
```

```
##
##
## Link between the cluster variable and the quantitative variables
## ================================================================
##                      Eta2       P-value
## duration       0.164777620 3.759496e-197
## minutes        0.164777620 3.759496e-197
## nr.employed    0.121012601 8.238443e-142
## pdays          0.090100788 2.433135e-104
## euribor3m      0.090010720 3.115343e-104
## emp.var.rate   0.085417483  8.992557e-99
## previous       0.042523921  5.101307e-49
## cons.price.idx 0.018386453  6.794885e-22
## cons.conf.idx  0.004669195  1.369222e-06
## campaign       0.004489049  2.189052e-06
## <NA>                    NA            NA
##
## Description of each cluster by quantitative variables
## =====================================================
## $`y-no`
##                    v.test Mean in category  Overall mean sd in category
## nr.employed     24.561104     5175.3298261 5166.47621340     64.3842715
## pdays           21.193217      983.3030029  963.73706378    123.8692868
## euribor3m       21.182621        3.7992890    3.61448034      1.6425449
## emp.var.rate    20.635071        0.2287424    0.06446049      1.4946001
## cons.price.idx   9.573739       93.6004884   93.57245006      0.5619158
## campaign         4.730529        2.5940750    2.53512998      2.5654605
## cons.conf.idx   -4.824514      -40.5398961  -40.42591256      4.4454152
## previous       -14.559593        0.1255362    0.15984757      0.4004406
## duration       -28.660364      217.4563107  250.62194144    191.6321071
## minutes        -28.660364        3.6242718    4.17703236      3.1938685
##                Overall sd       p.value
## nr.employed     71.7679377 3.291367e-133
## pdays          183.8068310  1.102990e-99
## euribor3m        1.7370025  1.381286e-99
## emp.var.rate     1.5850448  1.329502e-94
## cons.price.idx   0.5830800  1.031083e-21
## campaign         2.4808187  2.239350e-06
## cons.conf.idx    4.7037753  1.403451e-06
## previous         0.4691873  5.075919e-48
## duration       230.3904064 1.190744e-180
## minutes          3.8398401 1.190744e-180
##
## $`y-yes`
##                    v.test Mean in category  Overall mean sd in category
## minutes         28.660364         8.572322    4.17703236      5.3967235
## duration        28.660364       514.339318  250.62194144    323.8034093
## previous        14.559593         0.432675    0.15984757      0.7821222
## cons.conf.idx    4.824514       -39.519569  -40.42591256      6.3242738
## campaign        -4.730529         2.066427    2.53512998      1.5845655
## cons.price.idx  -9.573739        93.349503   93.57245006      0.6904449
## emp.var.rate   -20.635071        -1.241831    0.06446049      1.6751620
## euribor3m      -21.182621         2.144969    3.61448034      1.7676126
## pdays          -21.193217       808.157989  963.73706378    391.3731388
```

```
## nr.employed    -24.561104      5096.076481 5166.47621340     86.9764988
##                 Overall sd       p.value
## minutes          3.8398401 1.190744e-180
## duration       230.3904064 1.190744e-180
## previous         0.4691873  5.075919e-48
## cons.conf.idx    4.7037753  1.403451e-06
## campaign         2.4808187  2.239350e-06
## cons.price.idx   0.5830800  1.031083e-21
## emp.var.rate     1.5850448  1.329502e-94
## euribor3m        1.7370025  1.381286e-99
## pdays          183.8068310  1.102990e-99
## nr.employed     71.7679377 3.291367e-133
```

```
# $`y-yes`
#                                          Cla/Mod    Mod/Cla Global      p.value     v.test
# f.duration=f.duration-(483,1.58e+03]  40.8064516 44.7787611  12.40 2.180784e-97  20.942837
# poutcome=poutcome-success             62.2641509 17.5221239   3.18 5.331532e-56  15.766007
# f.pdays=f.pdays-[0,6]                 62.2222222 14.8672566   2.70 2.653287e-47  14.446089
# contact=contact-cellular              14.5686901 80.7079646  62.60 6.688527e-23   9.852462


# Dins el cluster que s'ha acceptat el producte financer, la "durada(483 a 1580]" Ã©s el 44,778% dels v
```