

# First delivery - ADEI

*Alex Rubio i Josep Bernad*

*1 de març de 2019*

## Contents

<b>1</b>	<b>Presentation</b>	<b>2</b>
<b>2</b>	<b>Bank client data</b>	<b>2</b>
2.1	Description . . . . .	2
<b>3</b>	<b>Loading packages</b>	<b>3</b>
<b>4</b>	<b>Loading data</b>	<b>3</b>
<b>5</b>	<b>Univariate Descriptive Analysis</b>	<b>4</b>
5.1	Transform missing and wrong data to NAs . . . . .	4
5.2	Create new factors corresponding to qualitative concepts. . . . .	6
5.2.1	Month . . . . .	6
5.2.2	Job . . . . .	7
5.2.3	Pdays . . . . .	7
5.2.4	Education . . . . .	8
5.2.5	Extra Factorization . . . . .	9
5.3	Create new factors corresponding to quantitative concepts. . . . .	10
5.3.1	Age discreatization . . . . .	10
<b>6</b>	<b>Exploratory Data Analysis</b>	<b>11</b>
6.1	Age . . . . .	11
6.2	Job . . . . .	12
6.3	Marital . . . . .	14
6.4	Default-Housing-Loan . . . . .	17
6.5	Contact Device . . . . .	19
6.6	Date - Month and season . . . . .	19
6.7	Date - Day of the week . . . . .	21
6.8	Duration . . . . .	23
6.9	Campaign . . . . .	24
6.10	PDays . . . . .	25
6.11	Previously . . . . .	27
6.12	POutcome . . . . .	28
6.13	Y . . . . .	29
<b>7</b>	<b>Data Quality Report</b>	<b>30</b>
7.1	Missing Values . . . . .	30
7.2	Errors . . . . .	31
7.2.1	Job . . . . .	32
7.2.2	Marital . . . . .	32
7.2.3	Education . . . . .	32
7.2.4	Default . . . . .	32
7.2.5	Housing . . . . .	32
7.2.6	Loan . . . . .	32
7.2.7	Contact . . . . .	32
7.2.8	Month . . . . .	33

7.2.9	Day of week . . . . .	33
7.2.10	Poutcome . . . . .	33
7.2.11	Y . . . . .	33
7.3	Outliers . . . . .	33
7.3.1	Age . . . . .	34
7.3.2	duration . . . . .	35
7.3.3	campaign . . . . .	35
7.3.4	pdays . . . . .	36
7.3.5	previous . . . . .	37
<b>8</b>	<b>Rank Variables</b>	<b>39</b>
8.1	Individual . . . . .	40
<b>9</b>	<b>Correlation</b>	<b>40</b>
<b>10</b>	<b>Imputation</b>	<b>45</b>
10.1	Numeric Variables . . . . .	45
10.2	Factors . . . . .	46
<b>11</b>	<b>Profiling</b>	<b>47</b>

## 1 Presentation

We are going to work with dataset bank-additional-full.csv with all examples (41188) and 20 inputs, ordered by date (from May 2008 to November 2010), very close to the data analyzed in [Moro et al., 2014].

Deliverables are concerned with Multivariant Data Analysis and model building for response variables: Y- Duration of the call and binary factor Y (Binary Target) if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

## 2 Bank client data

### 2.1 Description

*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')

### 3 Loading packages

### 4 Loading data

```
#rm(list=ls())
# Load Required Packages: to be increased over the course

#setwd("C:/Users/lmontero/Dropbox/DOCENCIA/FIB-ADEI/PRACTICA/BankMarketing")
#setwd("D:/DOCENCIA/FIB-ADEI/PRACTICA/BankMarketing")

# Josep
#setwd("~/Developer/r-studio/laboratory-adei/data-directory")
#load("~/Developer/r-studio/laboratory-adei/data-directory/5000_samples.RData")

# Alex
setwd("D:/Google Drive/Uni/ADEI/data-directory")
load(path.expand("D:/Google Drive/Uni/ADEI/data-directory/5000_samples.RData"))

summary(df)
```

```
##          age          job          marital
## Min.   :17.00  admin.   :1288  divorced: 546
## 1st Qu.:32.00  blue-collar:1156  married :3029
## Median :38.00  technician : 831  single  :1416
## Mean   :39.97  services  : 471  unknown : 9
## 3rd Qu.:47.00  management : 345
## Max.    :92.00  retired   : 187
##          (Other)   : 722
##          education  default          housing          loan
## university.degree :1431  no      :3939  no      :2226  no      :4138
## high.school        :1169  unknown:1061  unknown: 112  unknown: 112
## basic.9y           : 758  yes      : 0  yes     :2662  yes     : 750
## professional.course: 668
## basic.4y           : 493
## basic.6y           : 272
## (Other)            : 209
##          contact    month    day_of_week    duration
## cellular :3182  may    :1679  fri: 948  Min.    : 4.0
## telephone:1818  jul    : 907  mon:1017  1st Qu.: 104.0
##          aug      : 699  thu:1031  Median  : 181.0
##          jun      : 660  tue:1005  Mean    : 263.7
```

```
##          nov      : 502   wed: 999   3rd Qu.: 328.0
##          apr      : 323               Max.    :3078.0
##          (Other): 230
##      campaign      pdays      previous      poutcome
##  Min.    : 1.000   Min.    : 0.0   Min.    :0.0000   failure    : 493
## 1st Qu.: 1.000   1st Qu.:999.0   1st Qu.:0.0000   nonexistent:4315
## Median : 2.000   Median :999.0   Median :0.0000   success    : 192
## Mean    : 2.647   Mean    :957.9   Mean    :0.1772
## 3rd Qu.: 3.000   3rd Qu.:999.0   3rd Qu.:0.0000
## Max.    :42.000   Max.    :999.0   Max.    :5.0000
##
##      emp.var.rate   cons.price.idx   cons.conf.idx   euribor3m
##  Min.    :-3.4000   Min.    :92.20   Min.    : -50.80   Min.    :0.634
## 1st Qu.: -1.8000   1st Qu.:93.08   1st Qu.: -42.70   1st Qu.:1.344
## Median : 1.1000   Median :93.88   Median : -41.80   Median :4.857
## Mean    : 0.1029   Mean    :93.58   Mean    : -40.59   Mean    :3.641
## 3rd Qu.: 1.4000   3rd Qu.:93.99   3rd Qu.: -36.40   3rd Qu.:4.961
## Max.    : 1.4000   Max.    :94.77   Max.    : -26.90   Max.    :5.045
##
##      nr.employed      y
##  Min.    :4964   no :4416
## 1st Qu.:5099   yes: 584
## Median :5191
## Mean    :5168
## 3rd Qu.:5228
## Max.    :5228
##
```

## 5 Univariate Descriptive Analysis

Creem factors per cada variable posant abans NA a aquells valors erronis o faltants.

### 5.1 Transform missing and wrong data to NAs

```
#Default
sel<-which(df$default=="unknown");length(sel)
```

```
## [1] 1061
```

```
df$default[sel] <- NA
df$default <- factor(df$default)
summary(df$default)
```

```
##      no NA's
## 3939 1061
```

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

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

```
df$marital[sel] <- NA
df$marital <- factor(df$marital)
summary(df$marital)
```

```
## divorced married single NA's
```

```
##          546          3029          1416          9
```

```
#Housing  
sel<-which(df$housing=="unknown");length(sel)
```

```
## [1] 112
```

```
df$housing[sel] <- NA  
df$housing <- factor(df$housing)  
summary(df$housing)
```

```
##    no  yes NA's  
## 2226 2662  112
```

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

```
## [1] 112
```

```
df$loan[sel] <- NA  
df$loan <- factor(df$loan)  
summary(df$loan)
```

```
##    no  yes NA's  
## 4138  750  112
```

```
#Job  
sel<-which(df$job=="unknown");length(sel)
```

```
## [1] 43
```

```
df$job[sel] <- NA  
df$job <- factor(df$job)  
summary(df$job)
```

```
##      admin.    blue-collar  entrepreneur    housemaid    management  
##      1288          1156          181          132          345  
##      retired self-employed      services      student      technician  
##      187          152          471          100          831  
##      unemployed          NA's  
##      114          43
```

```
S
```

```
## function (object, brief, ...)  
## {  
##     UseMethod("S")  
## }  
## <bytecode: 0x0000000019741ed8>  
## <environment: namespace:car>
```

```
#Education  
sel<-which(df$education=="unknown");length(sel)
```

```
## [1] 207
```

```
df$education[sel] <- NA  
df$education <- factor(df$education)  
summary(df$education)
```

```
##          basic.4y          basic.6y          basic.9y
```

```
##           493           272           758
##      high.school      illiterate professional.course
##           1169           2           668
## university.degree      NA's
##           1431           207
```

```
#Pdays
```

```
sel<-which(df$pdays==999);length(sel)
```

```
## [1] 4793
```

```
df$pdays[sel] <- NA
summary(df$pdays)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##    0.000   3.000   6.000   5.792   7.000  18.000  4793
```

```
#Poutcome
```

```
sel<-which(df$poutcome=="nonexistent");length(sel)
```

```
## [1] 4315
```

```
df$poutcome[sel] <- NA
df$poutcome <- factor(df$poutcome)
summary(df$poutcome)
```

```
## failure success   NA's
##     493     192   4315
```

## 5.2 Create new factors corresponding to qualitative concepts.

### 5.2.1 Month

```
#Modify factor levels label
```

```
df$f.month <- factor(df$month, labels=paste("Month", sep="-", levels(df$month)))
table(df$f.month)
```

```
##
## Month-apr Month-aug Month-dec Month-jul Month-jun Month-mar Month-may
##      323      699      19      907      660      66      1679
## Month-nov Month-oct Month-sep
##      502      79      66
```

```
# Define new factor categories: 1-Spring / 2-Summer / 3-Resta
```

```
df$f.season <- 3
```

```
# 1 level - spring
```

```
sel<-which(df$f.month %in% c("Month-mar","Month-apr","Month-may"))
df$f.season[sel] <-1
```

```
# 2 level - Summer
```

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

```
table(df$f.season);summary(df$f.season)
```

```
##
##      1      2      3
```

```
## 2068 2266 666

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00   1.00   2.00   1.72   2.00   3.00

df$f.season<-factor(df$f.season,levels=1:3,labels=c("Spring","Summer","Aut-Win"))
summary(df$f.season)

##   Spring  Summer Aut-Win
##    2068    2266     666
```

## 5.2.2 Job

```
#Modify factor levels label
df$f.job <- factor(df$job, labels=paste("Job", sep="-", levels(df$job)))

table(df$f.job)

##
##      Job-admin.  Job-blue-collar  Job-entrepreneur  Job-housemaid
##           1288           1156           181           132
##      Job-management  Job-retired  Job-self-employed  Job-services
##           345           187           152           471
##      Job-student  Job-technician  Job-unemployed
##           100           831           114

# Define new factor categories: 1-selfemployed / 2-worker / 3-other
df$f.jobsituation<-3

# 1 level - self-employed
sel<-which(df$f.job %in% c("Job-entrepreneur","Job-housemaid","Job-self-employed"))
df$f.jobsituation[sel] <- 1

# 2 level - worker
sel<-which(df$f.job %in% c("Job-admin","Job-blue-collar","Job-management","Job-services","Job-technician"))
df$f.jobsituation[sel] <- 2

table(df$f.jobsituation);summary(df$f.jobsituation)

##
##      1      2      3
##    465 2803 1732

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   2.000   2.000   2.253   3.000   3.000

df$f.jobsituation<-factor(df$f.jobsituation,levels=1:3,labels=c("Self-employed","Worker","Other"))
summary(df$f.jobsituation)

## Self-employed      Worker      Other
##           465           2803           1732
```

## 5.2.3 Pdays

```
table(df$pdays)
```

```
##
```

```
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 15 16 17 18
## 1 5 12 62 17 5 48 13 5 9 7 2 4 8 3 1 4 1

# Define new factor categories: 1-contacted / 2-not contacted
df$f.prev_contacted<-2

# 1 level - contacted
sel<-which(df$pdays %in% c(1:20))
df$f.prev_contacted[sel] <- 1

# 2 level - not contacted
sel<-which(df$pdays %in% c(21:1000))
df$f.prev_contacted[sel] <- 2

table(df$f.prev_contacted);summary(df$f.prev_contacted)

##
##      1      2
## 206 4794

##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      1.000  2.000   2.000   1.959   2.000   2.000

df$f.prev_contacted<-factor(df$f.prev_contacted,levels=1:2,labels=c("Contacted","No-contacted"))
summary(df$pdays)

##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.      NA's
##      0.000  3.000   6.000   5.792   7.000  18.000      4793
```

#### 5.2.4 Education

```
#Modify factor levels label
df$education <- factor(df$education, labels=paste("Edu", sep="-", levels(df$education)))

table(df$education)

##
##      Edu-basic.4y      Edu-basic.6y      Edu-basic.9y
##           493           272           758
##      Edu-high.school      Edu-illiterate      Edu-professional.course
##           1169                2           668
##      Edu-university.degree
##           1431

# Define new factor categories: 1-mandatory / 2-nonmandatory / 3-other
df$f.education<-3

# 1 level - mandatory
sel<-which(df$education %in% c("Edu-basic.4y","Edu-basic.6y", "Edu-basic.9y", "Edu-high.school"))
df$f.education[sel] <- 1

# 2 level - nonmandatory
sel<-which(df$education %in% c("Edu-professional.course","Edu-university.degree"))
df$f.education[sel] <- 2

table(df$f.education);summary(df$f.education)
```



```
##
##      1      2      3
## 2692 2099  209

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   1.000   1.000   1.503   2.000   3.000

df$f.education<-factor(df$f.education,levels=1:3,labels=c("Mandatory","Non-Mandatory","Other"))
summary(df$f.education)

##      Mandatory Non-Mandatory      Other
##           2692           2099           209
```

### 5.2.5 Extra Factorization

#### *#Housing*

```
df$f.housing<-factor(df$housing,labels=paste("f",sep=".",levels(df$housing)))
table(df$f.housing);summary(df$f.housing);
```

```
##
## f.no f.yes
## 2226 2662

## f.no f.yes NA's
## 2226 2662  112
```

#### *#Marital*

```
df$f.marital<-factor(df$marital,labels=paste("f",sep=".",levels(df$marital)))
table(df$f.marital);summary(df$f.marital);
```

```
##
## f.divorced f.married f.single
##          546         3029         1416

## f.divorced f.married f.single      NA's
##          546         3029         1416         9
```

#### *#Default*

```
df$f.default<-factor(df$default, labels=paste("f",sep=".",levels(df$default)))
df$f.default <- factor(df$f.default , levels = c(levels(df$f.default), "f.si"))
table(df$f.default);
```

```
##
## f.no f.si
## 3939    0
```

#### *#Loan*

```
df$f.loan<-factor(df$loan,labels=paste("f",sep=".",levels(df$loan)))
table(df$f.loan);summary(df$f.loan)
```

```
##
## f.no f.yes
## 4138  750

## f.no f.yes NA's
## 4138  750  112
```

```
#Contact
df$f.contact<-factor(df$contact,labels=paste("f",sep=".",levels(df$contact)))
table(df$f.contact);summary(df$f.contact)
```

```
##
## f.cellular f.telephone
##      3182      1818
## f.cellular f.telephone
##      3182      1818
```

```
#Day of Week
df$f.day<-factor(df$day_of_week,labels=paste("f.day",sep=".",levels(df$day)))
table(df$f.day);summary(df$f.day)
```

```
##
## f.day.fri f.day.mon f.day.thu f.day.tue f.day.wed
##      948      1017      1031      1005      999
## f.day.fri f.day.mon f.day.thu f.day.tue f.day.wed
##      948      1017      1031      1005      999
```

## 5.3 Create new factors corresponding to quantitative concepts.

### 5.3.1 Age discretization

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

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      17.00  32.00   38.00   39.97  47.00   92.00
```

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

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

```
## varaux
## [17,32] (32,38] (38,47] (47,92]
##      1353      1248      1202      1197
```

```
tapply(df$age,varaux,median)
```

```
## [17,32] (32,38] (38,47] (47,92]
##      29      35      43      53
```

```
varaux<-factor(cut(df$age,breaks=c(17,30,40,50,95),include.lowest=T))
table(varaux)
```

```
## varaux
## [17,30] (30,40] (40,50] (50,95]
##      887      2003      1252      858
```

```
tapply(df$age,varaux,median)
```

```
## [17,30] (30,40] (40,50] (50,95]
##      28      35      45      55
```

```
df$f.age<-factor(cut(df$age,breaks=c(17,30,40,50,95),include.lowest=T))
```

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

```
## [17,30] (30,40] (40,50] (50,95]  
##      887      2003      1252      858
```

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

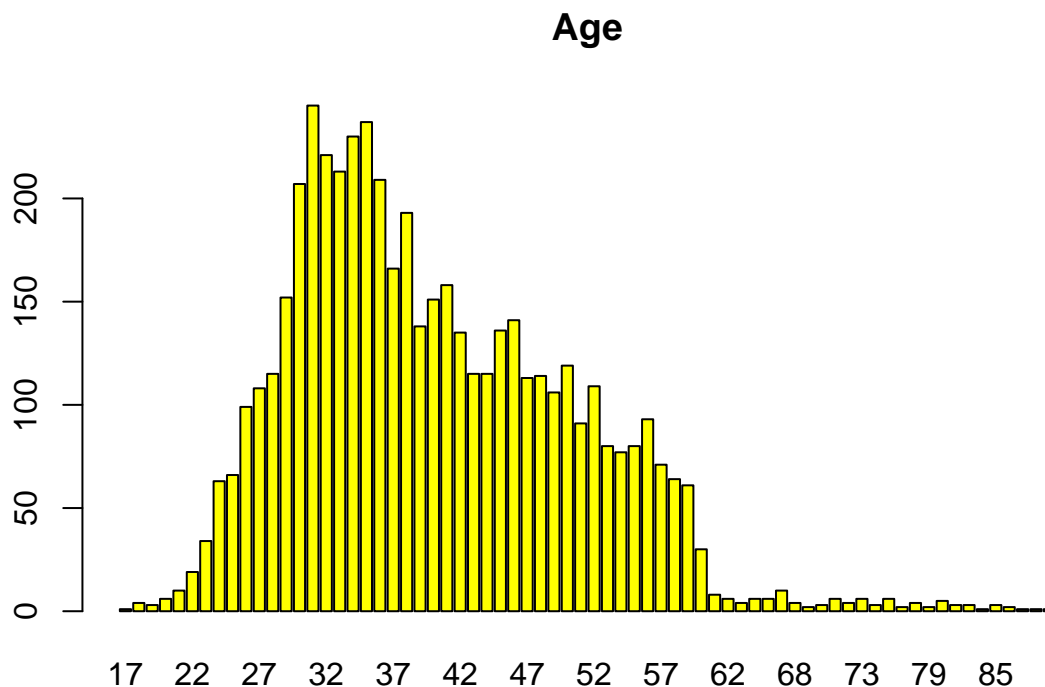
## 6 Exploratory Data Analysis

### 6.1 Age

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

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##    17.00   32.00   38.00   39.97   47.00   92.00
```

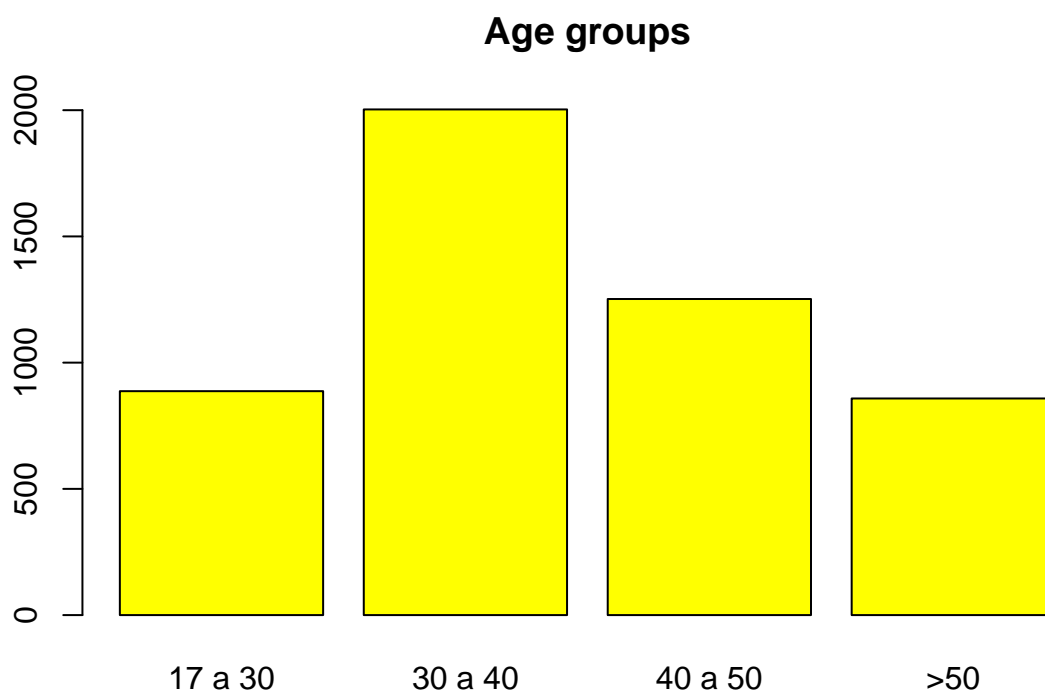
```
barplot(table(df$age), main= "Age",col="yellow")
```



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

```
## f.age-[17,30] f.age-(30,40] f.age-(40,50] f.age-(50,95]  
##           887           2003           1252           858
```

```
barplot(table(df$f.age), main="Age groups",names.arg=c("17 a 30","30 a 40","40 a 50",">50"),col="yellow")
```



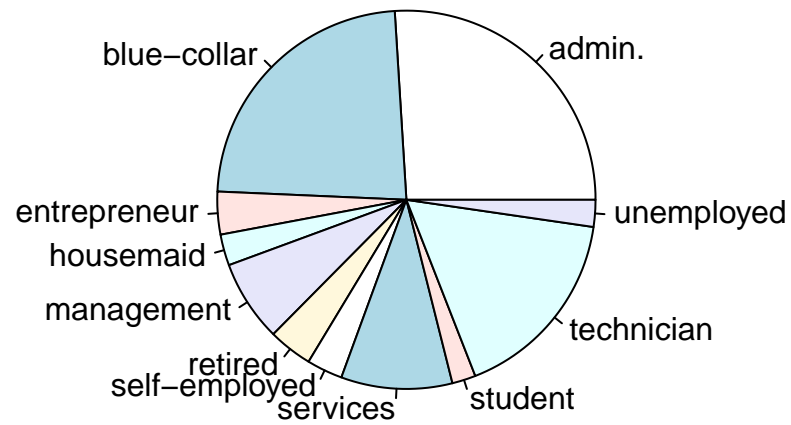
## 6.2 Job

```
table(df$job)
```

```
##
##      admin.  blue-collar entrepreneur  housemaid  management
##      1288      1156         181         132         345
##      retired self-employed      services      student  technician
##      187      152         471         100         831
##      unemployed
##      114
```

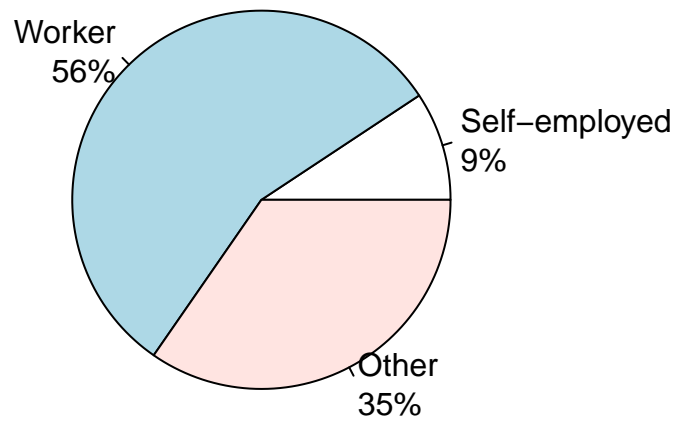
```
pie(table(df$job), main= "Job")
```

## Job



```
aux <- table(df$f.jobssituation)
pct <- round(aux/sum(aux)*100)
lbls <- paste(names(aux), "\n", pct, sep="")
lbls <- paste(lbls,"%",sep="") # add % to labels
pie(aux,labels = lbls, main="Job Situation")
```

## Job Situation



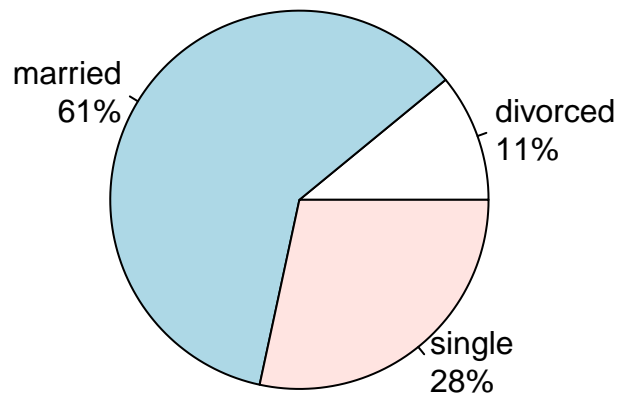
## 6.3 Marital

```
table(df$marital)
```

```
##  
## divorced married single  
##      546      3029      1416
```

```
aux <- table(df$marital)  
pct <- round(aux/sum(aux)*100)  
lbls <- paste(names(aux), "\n", pct, sep="")  
lbls <- paste(lbls,"%",sep="") # add % to labels  
pie(aux,labels = lbls, main="Marital Situation")
```

## Marital Situation



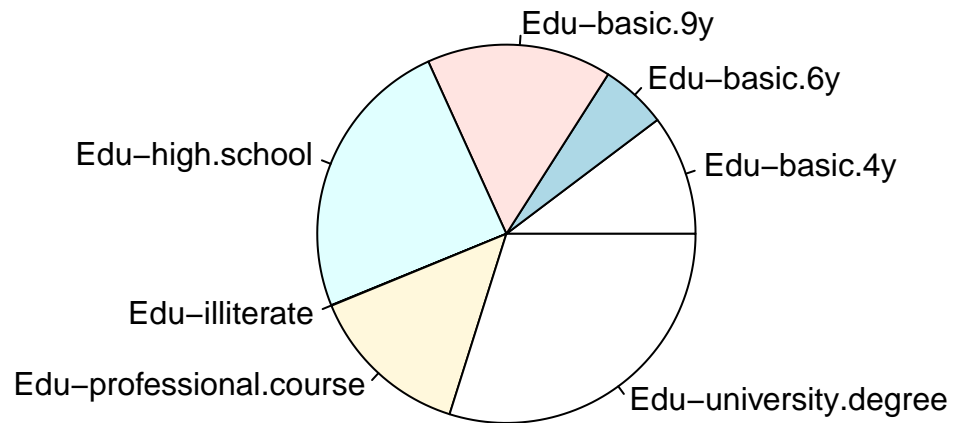
##Education

```
table(df$education)
```

```
##
##      Edu-basic.4y      Edu-basic.6y      Edu-basic.9y
##              493              272              758
##      Edu-high.school      Edu-illiterate      Edu-professional.course
##              1169              2              668
##      Edu-university.degree
##              1431
```

```
pie(table(df$education), main= "Education")
```

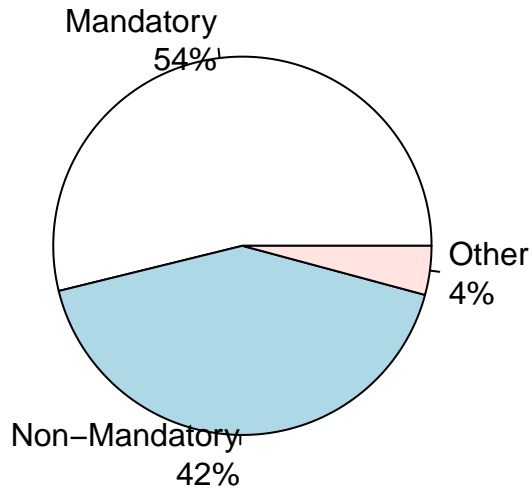
## Education



```
aux <- table(df$f.education)
pct <- round(aux/sum(aux)*100)
lbls <- paste(names(aux), "\n", pct, sep="")
lbls <- paste(lbls,"%",sep="") # add % to labels
pie(aux,labels = lbls, main="Education Level")
```



## Education Level



### 6.4 Default-Housing-Loan

```
table(df$default)
```

```
##  
##  no  
## 3939
```

```
table(df$housing)
```

```
##  
##  no  yes  
## 2226 2662
```

```
table(df$loan)
```

```
##  
##  no  yes  
## 4138  750
```

```
attach(mtcars)
```

```
## The following object is masked from package:ggplot2:
```

```
##
```

```
##  mpg
```

```
par(mfrow=c(1,2))
```

```
aux <- table(df$loan)
```

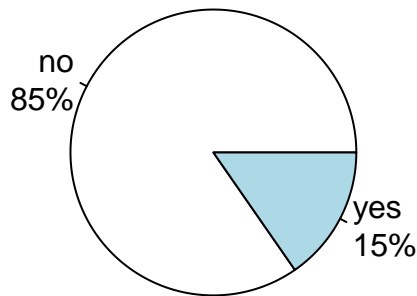
```
pct <- round(aux/sum(aux)*100)
```

```

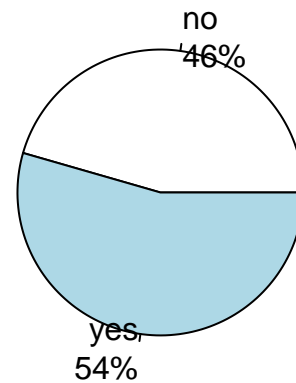
lbls <- paste(names(aux), "\n", pct, sep="")
lbls <- paste(lbls, "%", sep="") # ad % to labels
pie(aux, labels = lbls, main="Personal Loan")
aux <- table(df$housing)
pct <- round(aux/sum(aux)*100)
lbls <- paste(names(aux), "\n", pct, sep="")
lbls <- paste(lbls, "%", sep="") # ad % to labels
pie(aux, labels = lbls, main="Housing Loan")

```

**Personal Loan**



**Housing Loan**



```

# Retornar l'attach a l'estat predeterminat
attach(mtcars)

```

```

## The following objects are masked from mtcars (pos = 3):
##
##      am, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs, wt
##
## The following object is masked from package:ggplot2:
##
##      mpg

```

```

par(mfrow=c(1,1))

```

Com es pot veure no hem el gràfic de deutes, ja que el 100% d'individus que han contestat a l'enquesta no en tenien.

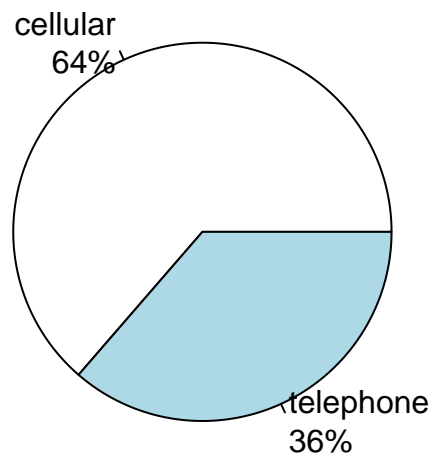
## 6.5 Contact Device

```
table(df$contact)

##
##  cellular telephone
##      3182      1818

aux <- table(df$contact)
pct <- round(aux/sum(aux)*100)
lbls <- paste(names(aux), "\n", pct, sep="")
lbls <- paste(lbls,"%",sep="") # add % to labels
pie(aux,labels = lbls, main="Contact Device")
```

### Contact Device



## 6.6 Date - Month and season

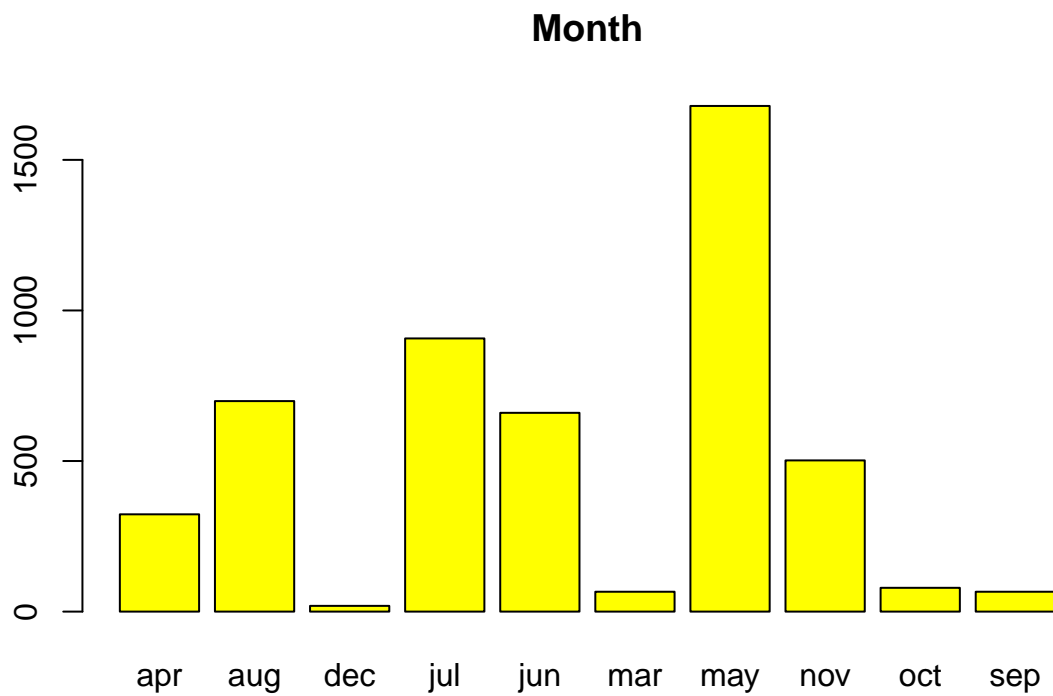
```
table(df$month)

##
##  apr  aug  dec  jul  jun  mar  may  nov  oct  sep
##  323  699   19  907  660   66 1679  502   79   66

table(df$f.season)

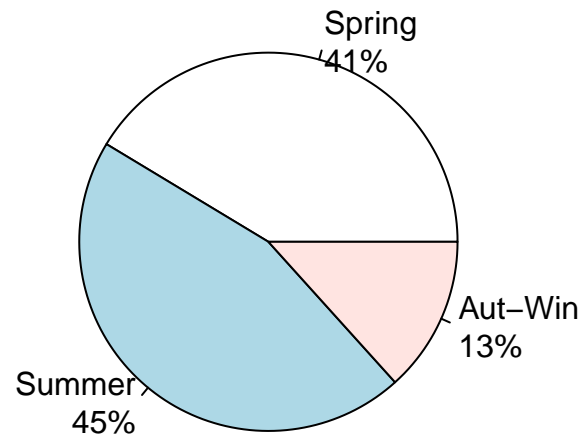
##
##  Spring  Summer  Aut-Win
##   2068    2266    666
```

```
barplot(table(df$month), main= "Month", col="yellow")
```



```
aux <- table(df$f.season)
pct <- round(aux/sum(aux)*100)
lbls <- paste(names(aux), "\n", pct, sep="")
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(aux,labels = lbls,
    main="Season")
```

## Season

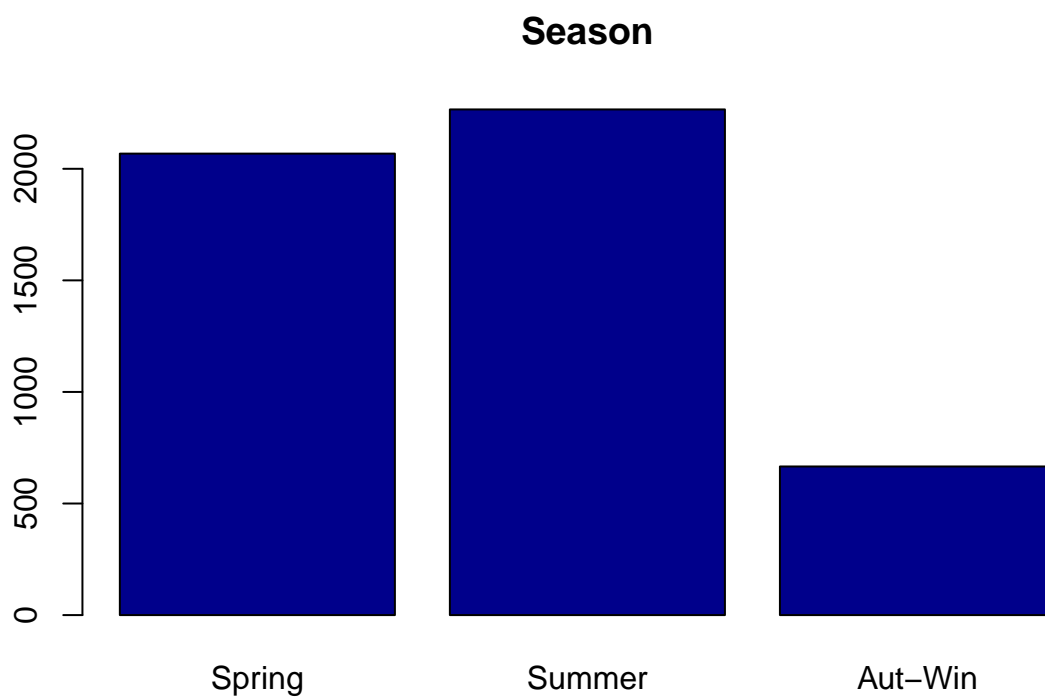


### 6.7 Date - Day of the week

```
table(df$day_of_week)
```

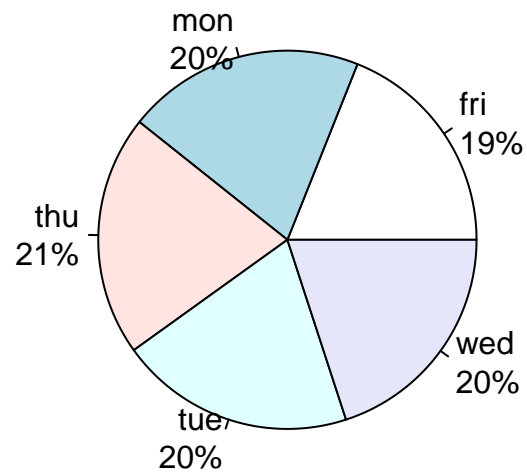
```
##  
##  fri  mon  thu  tue  wed  
##  948 1017 1031 1005  999
```

```
barplot(table(df$f.season), main= "Season", col="darkblue")
```



```
aux <- table(df$day_of_week)
pct <- round(aux/sum(aux)*100)
lbls <- paste(names(aux), "\n", pct, sep="")
lbls <- paste(lbls,"%",sep="") # add % to labels
pie(aux,labels = lbls,
    main="Day of the week")
```

## Day of the week



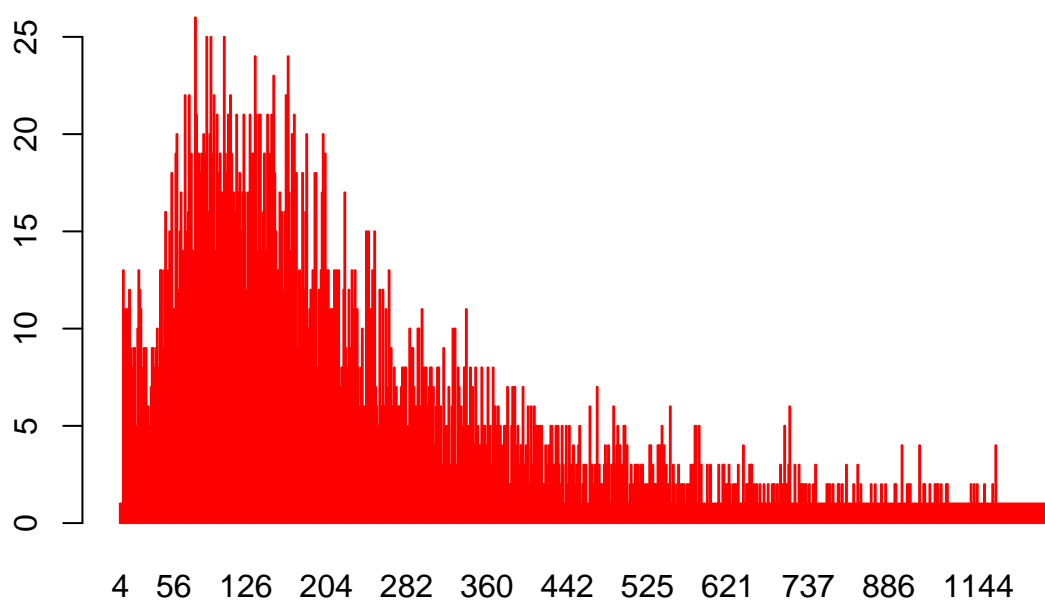
## 6.8 Duration

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

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       4.0   104.0   181.0   263.7   328.0   3078.0
```

```
barplot(table(df$duration),col="yellow",border="red", main="Call duration")
```

## Call duration



## 6.9 Campaign

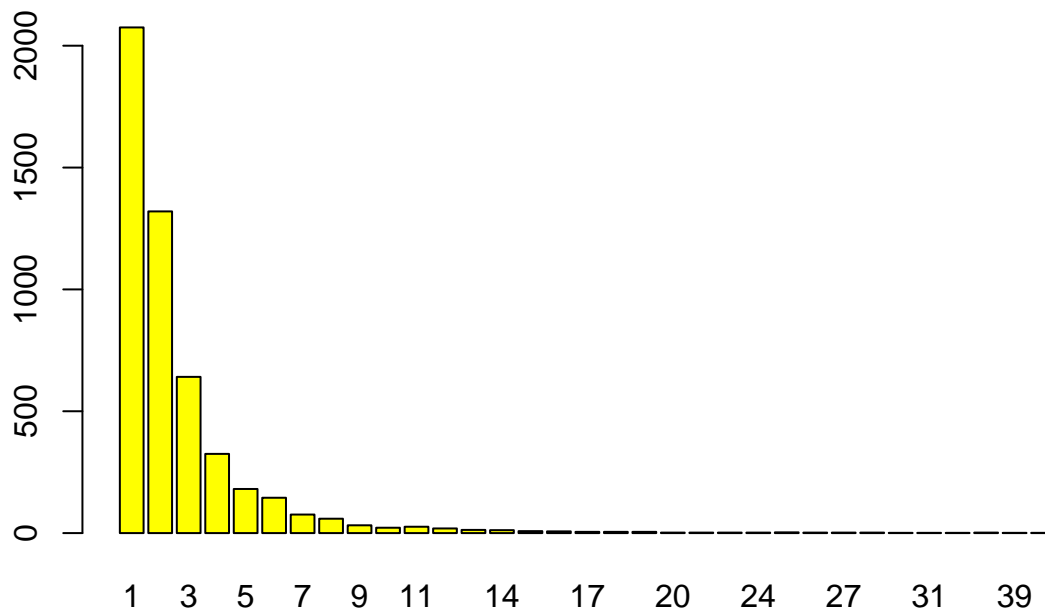
```
summary(df$campaign)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   1.000   2.000   2.647   3.000  42.000
```

```
barplot(table(df$campaign),col="yellow", main="Number of campaigns previously contacted")
```



## Number of campaigns previously contacted



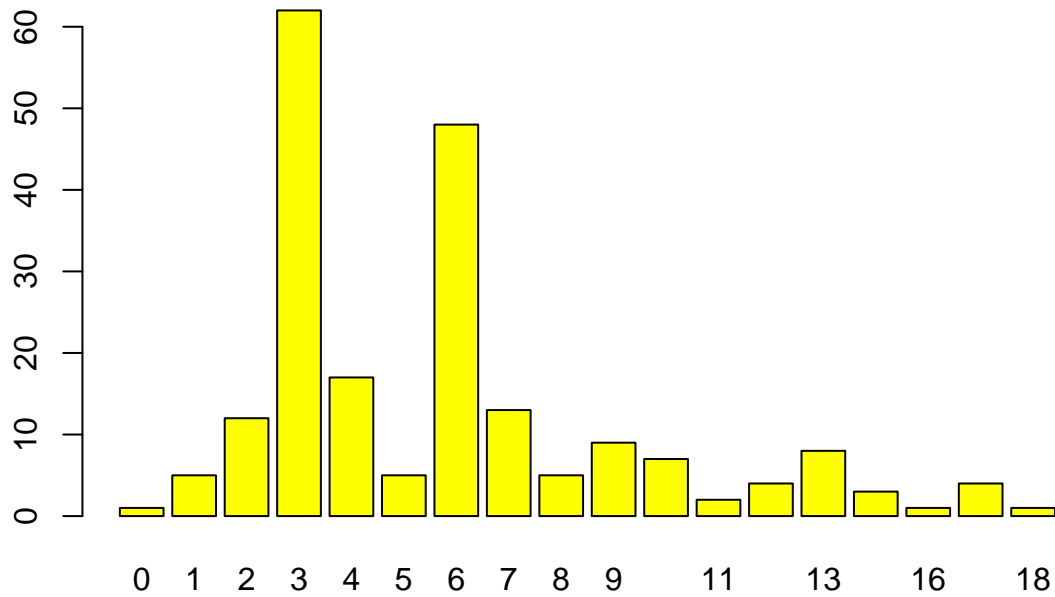
### 6.10 PDays

```
summary(df$pdays)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##      0.000  3.000   6.000   5.792  7.000  18.000  4793
```

```
barplot(table(df$pdays),col="yellow", main="Number of days between the last contact")
```

## Number of days between the last contact

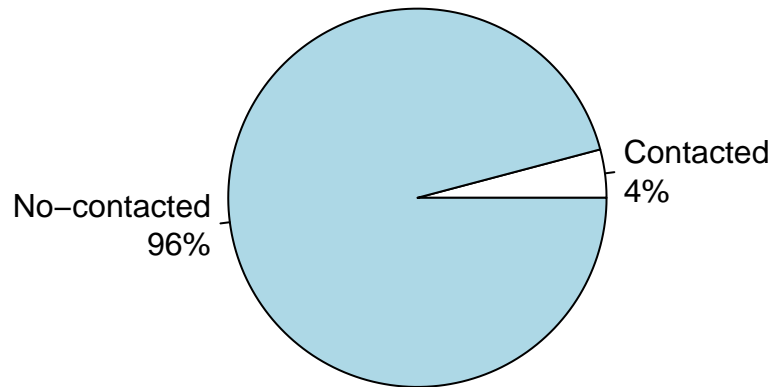


```
table(df$f.prev_contacted)
```

```
##  
##      Contacted No-contacted  
##           206           4794
```

```
aux <- table(df$f.prev_contacted)  
pct <- round(aux/sum(aux)*100)  
lbls <- paste(names(aux), "\n", pct, sep="")  
lbls <- paste(lbls,"%",sep="") # ad % to labels  
pie(aux,labels = lbls,  
     main="Was previously contacted?")
```

## Was previously contacted?



### 6.11 Previously

```
summary(df$previous)
```

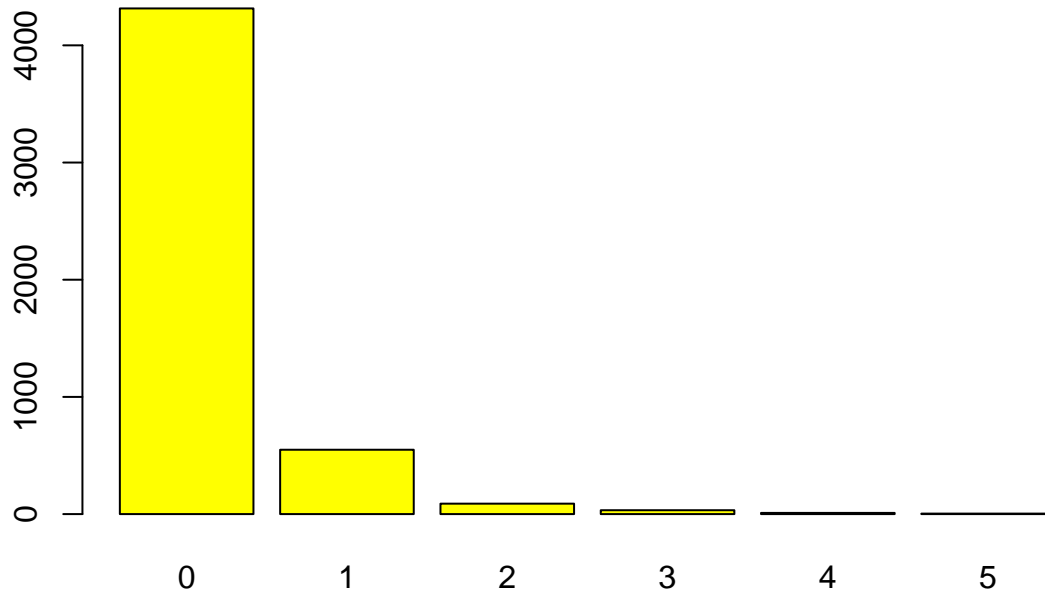
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000 0.0000 0.0000 0.1772 0.0000 5.0000
```

```
table(df$previous)
```

```
##
##      0      1      2      3      4      5
## 4315  549   89   33   10    4
```

```
barplot(table(df$previous),col="yellow", main="Number of contacts before this campaign")
```

## Number of contacts before this campaign

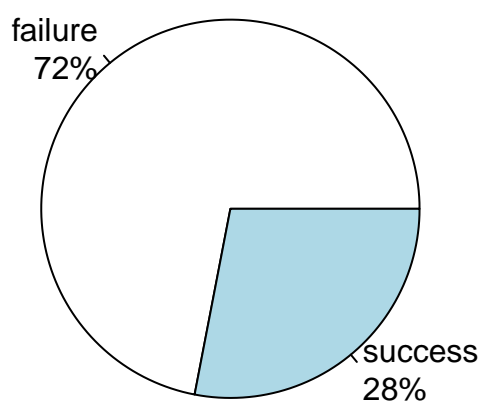


### 6.12 POutcome

```
table(df$poutcome)
```

```
##  
## failure success  
##      493      192  
  
aux <- table(df$poutcome)  
pct <- round(aux/sum(aux)*100)  
lbls <- paste(names(aux), "\n", pct, sep="")  
lbls <- paste(lbls,"%",sep="") # ad % to labels  
pie(aux,labels = lbls,  
     main="Outcome of the previous marketing campaign")
```

## Outcome of the previous marketing campaign



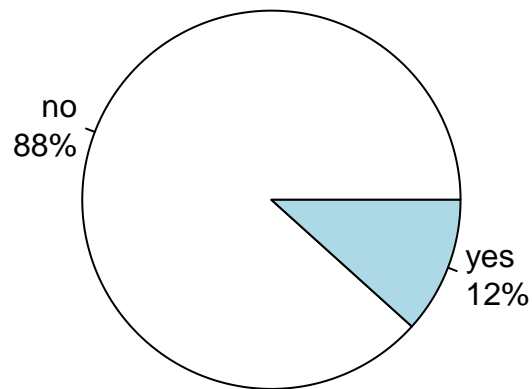
### 6.13 Y

```
table(df$y)
```

```
##  
##   no  yes  
## 4416 584
```

```
aux <- table(df$y)  
pct <- round(aux/sum(aux)*100)  
lbls <- paste(names(aux), "\n", pct, sep="")  
lbls <- paste(lbls,"%",sep="") # add % to labels  
pie(aux,labels = lbls,  
     main="Binary target")
```

## Binary target



## 7 Data Quality Report

### 7.1 Missing Values

```
vmiss<-rep(0,nrow(df))

nInitialVariables<- 21
nmiss<-rep(0,nInitialVariables)

initialVariables <- 0:21
names(nmiss) <- names(df[initialVariables])
names(df[initialVariables])

## [1] "age"          "job"          "marital"      "education"
## [5] "default"      "housing"      "loan"         "contact"
## [9] "month"        "day_of_week"  "duration"     "campaign"
## [13] "pdays"       "previous"     "poutcome"     "emp.var.rate"
## [17] "cons.price.idx" "cons.conf.idx" "euribor3m"    "nr.employed"
## [21] "y"

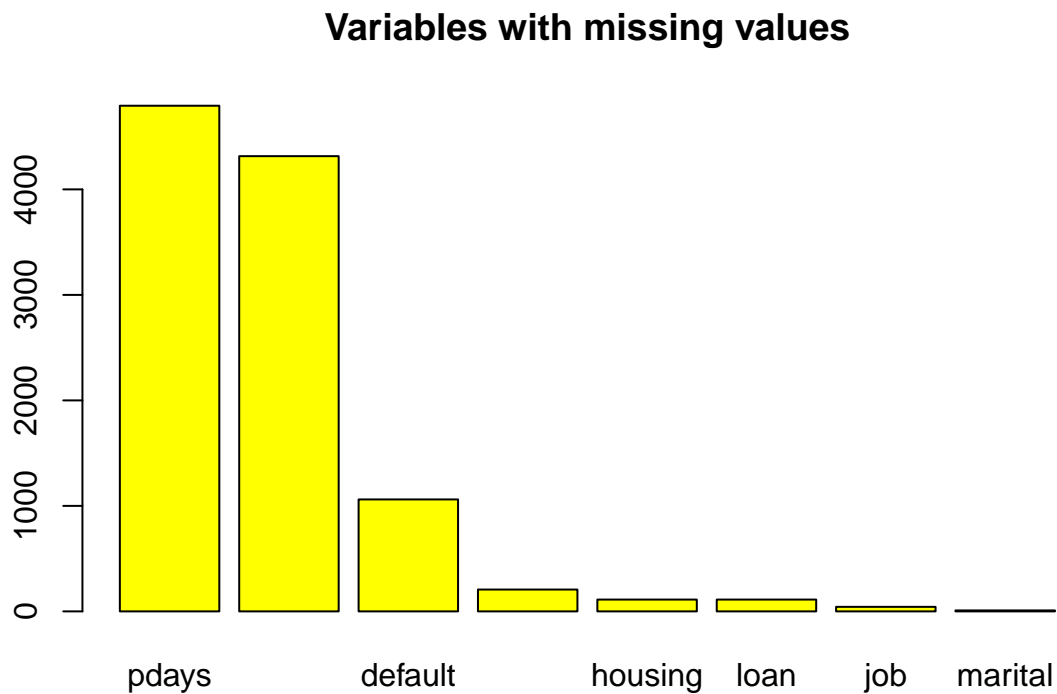
for(j in 1:21) {
  nmiss[j]<-nmiss[j]+sum(is.na(df[,j]))
}

nmiss_aux <- nmiss[ nmiss != 0 ]
nmiss_aux <- sort(nmiss_aux, decreasing = TRUE)
```

```
table(nmiss_aux)
```

```
## nmiss_aux
##      9    43   112   207 1061 4315 4793
##      1     1     2     1     1     1     1
```

```
barplot(nmiss_aux, col="yellow", main="Variables with missing values");
```



Al barplot sols apareixen les variables amb dades mancants.

## 7.2 Errors

```
verrs<-rep(0, nrow(df))
```

```
nInitialVariables<- 21
```

```
nerrs<-rep(0, nInitialVariables)
```

```
initialVariables <- 0:21
```

```
names(nerrs) <- names(df[initialVariables])
```

```
names(df[initialVariables])
```

```
## [1] "age"           "job"           "marital"       "education"
## [5] "default"       "housing"       "loan"          "contact"
## [9] "month"        "day_of_week"   "duration"      "campaign"
## [13] "pdays"       "previous"      "poutcome"      "emp.var.rate"
## [17] "cons.price.idx" "cons.conf.idx" "euribor3m"     "nr.employed"
## [21] "y"
```

### 7.2.1 Job

```
v<-c("admin.", "blue-collar", "entrepreneur", "housemaid", "management", "retired", "self-employed", "se  
llista<-which(!is.element(df[2], v));  
verrs[llista]<-verrs[llista]+1  
nerrs[2]<-nerrs[2]+sum(!is.element(df[,2], v))
```

### 7.2.2 Marital

```
v<-c("divorced", "married", "single", NA)  
llista<-which(!is.element(df[3], v));  
verrs[llista]<-verrs[llista]+1  
nerrs[3]<-nerrs[3]+sum(!is.element(df[,3], v))
```

### 7.2.3 Education

```
v<-c("Edu-basic.4y", "Edu-basic.6y", "Edu-basic.9y", "Edu-high.school", "Edu-illiterate", "Edu-professi  
llista<-which(!is.element(df[4], v));  
verrs[llista]<-verrs[llista]+1  
nerrs[4]<-nerrs[4]+sum(!is.element(df[,4], v))
```

### 7.2.4 Default

```
v<-c("no", "yes", NA)  
llista<-which(!is.element(df[5], v));  
verrs[llista]<-verrs[llista]+1  
nerrs[5]<-nerrs[5]+sum(!is.element(df[,5], v))
```

### 7.2.5 Housing

```
v<-c("no", "yes", NA)  
llista<-which(!is.element(df[6], v));  
verrs[llista]<-verrs[llista]+1  
nerrs[6]<-nerrs[6]+sum(!is.element(df[,6], v))
```

### 7.2.6 Loan

```
v<-c("no", "yes", NA)  
llista<-which(!is.element(df[7], v));  
verrs[llista]<-verrs[llista]+1  
nerrs[7]<-nerrs[7]+sum(!is.element(df[,7], v))
```

### 7.2.7 Contact

```
v<-c("cellular", "telephone", NA)  
llista<-which(!is.element(df[8], v));  
verrs[llista]<-verrs[llista]+1  
nerrs[8]<-nerrs[8]+sum(!is.element(df[,8], v))
```



### 7.2.8 Month

```
v<-c("apr", "aug", "dec", "jul", "jun", "mar", "may", "nov", "oct", "sep", "jan", "feb", NA)
llista<-which(!is.element(df[9], v));
verrs[llista]<-verrs[llista]+1
nerrs[9]<-nerrs[9]+sum(!is.element(df[,9], v))
```

### 7.2.9 Day of week

```
v<-c("mon", "tue", "wed", "thu", "fri", NA)
llista<-which(!is.element(df[10], v));
verrs[llista]<-verrs[llista]+1
nerrs[10]<-nerrs[10]+sum(!is.element(df[,10], v))
```

### 7.2.10 Poutcome

```
v<-c("failure", "success", NA)
llista<-which(!is.element(df[,15], v));
verrs[llista]<-verrs[llista]+1
nerrs[15]<-nerrs[15]+sum(!is.element(df[,15], v))
```

### 7.2.11 Y

```
v<-c("yes", "no", NA)
llista<-which(!is.element(df[21], v));
verrs[llista]<-verrs[llista]+1
nerrs[21]<-nerrs[21]+sum(!is.element(df[,21], v))
```

Així els errors queden:

```
nerrs
```

```
##          age          job          marital          education          default
##           0           0           0           0           0
##      housing          loan          contact           month      day_of_week
##           0           0           0           0           0
##      duration      campaign          pdays          previous          poutcome
##           0           0           0           0           0
## emp.var.rate cons.price.idx cons.conf.idx      euribor3m      nr.employed
##           0           0           0           0           0
##           y
##           0
```

## 7.3 Outliers

```
vout<-rep(0,nrow(df))

nInitialVariables<- 21
nout<-rep(0,nInitialVariables)

initialVariables <- 0:21
names(nout) <- names(df[initialVariables])
names(df[initialVariables])
```

```
## [1] "age"          "job"          "marital"      "education"
## [5] "default"      "housing"      "loan"         "contact"
## [9] "month"        "day_of_week"  "duration"     "campaign"
## [13] "pdays"       "previous"     "poutcome"     "emp.var.rate"
## [17] "cons.price.idx" "cons.conf.idx" "euribor3m"    "nr.employed"
## [21] "y"
```

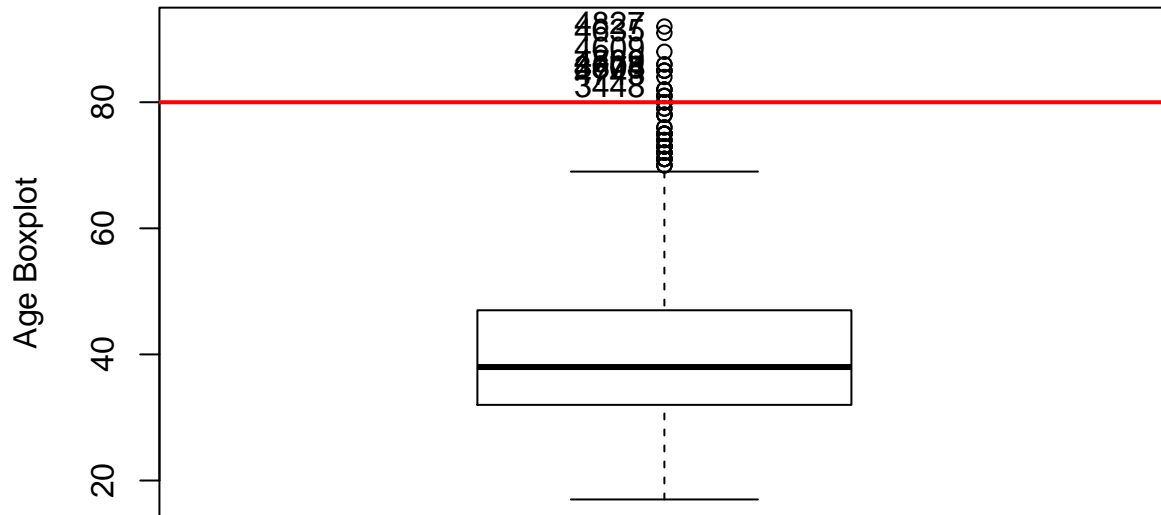
### 7.3.1 Age

```
Boxplot(df$age, ylab = "Age Boxplot")
```

```
## [1] 4827 4635 4609 4732 4869 3675 4803 4804 4743 3448
```

```
sout <- 80
```

```
abline(h=sout,col="red",lwd=2)
```



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

```
## [1] 15
```

```
df$age[outliers] <- NA;
if(length(outliers)>0){
vout[outliers]<-vout[outliers]+1
nout["age"]<-length(outliers)}
```

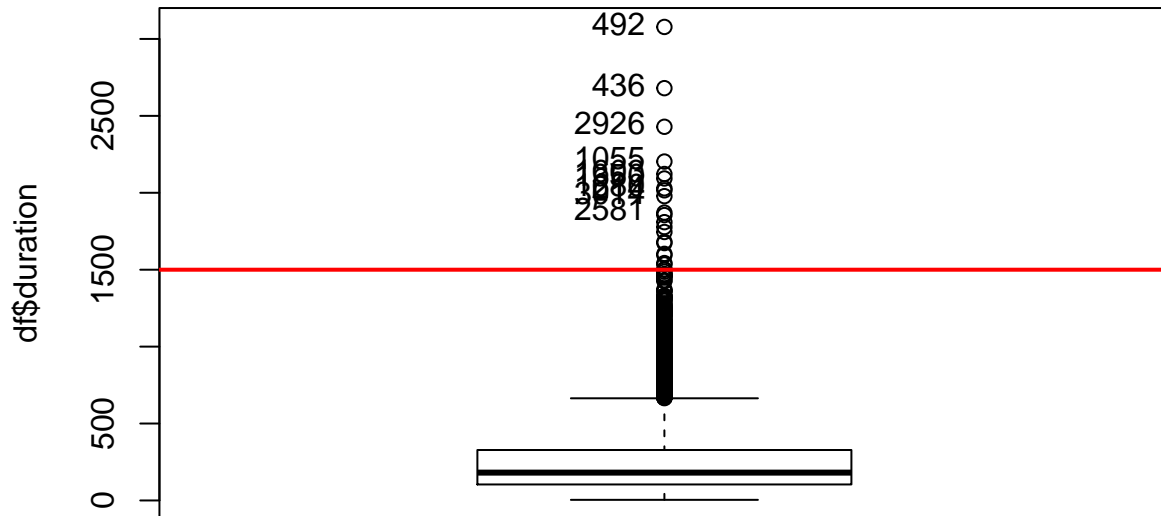
### 7.3.2 duration

```
Boxplot(df$duration)
```

```
## [1] 492 436 2926 1055 1603 1350 1680 214 3014 2581
```

```
sout <- 1500
```

```
abline(h=sout,col="red",lwd=2)
```



```
outliers<-which(df$duration>sout);length(outliers);
```

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

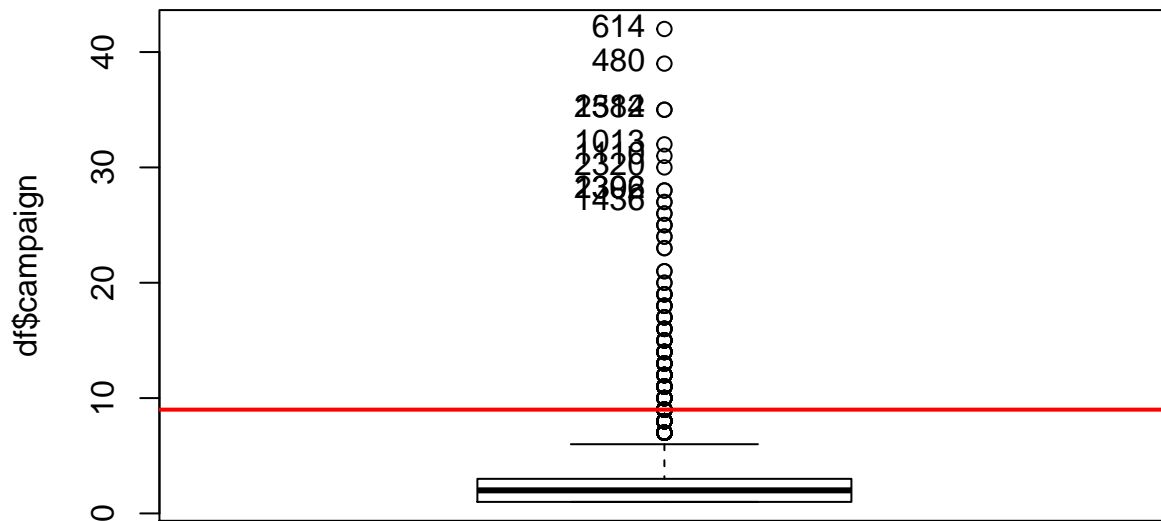
```
if(length(outliers)>0){  
  vout[outliers]<-vout[outliers]+1  
  nout["duration"]<-length(outliers)}
```

### 7.3.3 campaign

```
Boxplot(df$campaign)
```

```
## [1] 614 480 1584 2312 1013 1110 2320 1392 2306 1436
```

```
sout <- 9
abline(h=sout,col="red",lwd=2)
```



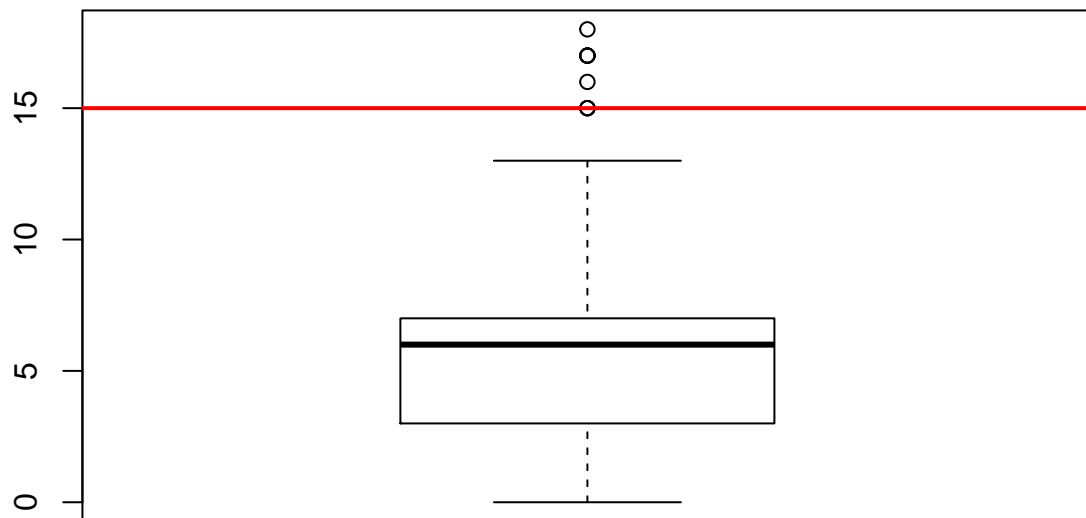
```
outliers<-which(df$campaign>sout);length(outliers);
```

```
## [1] 146
```

```
df$campaign[outliers] <- NA;
if(length(outliers)>0){
  vout[outliers]<-vout[outliers]+1
  nout["campaign"]<-length(outliers)}
```

#### 7.3.4 pdays

```
boxplot(df$pdays);
sout <- 15;
abline(h=sout,col="red",lwd=2);
```



```
outliers<-which(df$pdays> sout); length(outliers);
```

```
## [1] 6
```

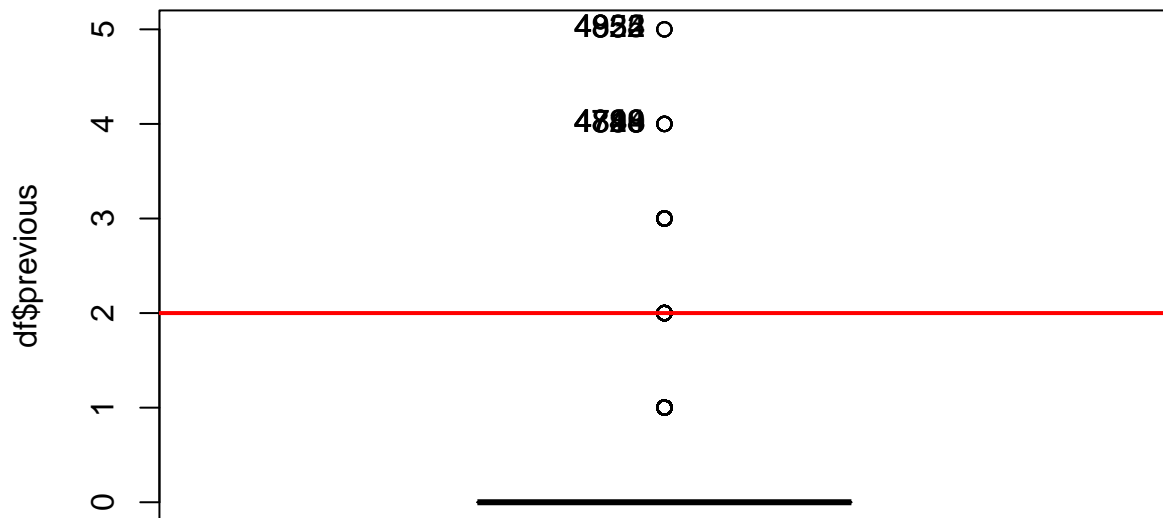
```
df$pdays[outliers] <- NA;
if(length(outliers)>0){
  vout[outliers]<-vout[outliers]+1
  nout["pdays"]<-length(outliers)}
```

### 7.3.5 previous

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

```
## [1] 4822 4835 4952 4954 4719 4783 4790 4828 4844 4848
```

```
sout <- 2
abline(h=sout,col="red",lwd=2)
```



```
outliers<-which(df$previous> sout);
df$previous[outliers] <- NA;
length(outliers);
```

```
## [1] 47
```

```
if(length(outliers)>0){
vout[outliers]<-vout[outliers]+1
nout["previous"]<-length(outliers)}
```

Així els outliers queden:

```
nout
```

```
##      age      job      marital      education      default
##      15       0       0          0              0
##      housing    loan      contact      month      day_of_week
##      0         0       0          0              0
##      duration   campaign    pdays      previous      poutcome
##      21        146        6          47           0
##      emp.var.rate cons.price.idx cons.conf.idx euribor3m nr.employed
##      0           0           0          0           0
##      y
##      0
```

## 8 Rank Variables

```
miss <- sort(nmiss, decreasing = TRUE)
miss
```

```
##      pdays      poutcome      default      education      housing
##      4793         4315         1061         207         112
##      loan        job        marital        age        contact
##      112         43         9         0         0
##      month      day_of_week      duration      campaign      previous
##      0         0         0         0         0
##      emp.var.rate cons.price.idx cons.conf.idx euribor3m nr.employed
##      0         0         0         0         0
##      y
##      0
```

```
err <- sort(nerrs, decreasing = TRUE)
err
```

```
##      age        job        marital      education      default
##      0         0         0         0         0
##      housing      loan      contact      month      day_of_week
##      0         0         0         0         0
##      duration      campaign      pdays      previous      poutcome
##      0         0         0         0         0
##      emp.var.rate cons.price.idx cons.conf.idx euribor3m nr.employed
##      0         0         0         0         0
##      y
##      0
```

```
miss <- sort(nmiss, decreasing = TRUE)
miss
```

```
##      pdays      poutcome      default      education      housing
##      4793         4315         1061         207         112
##      loan        job        marital        age        contact
##      112         43         9         0         0
##      month      day_of_week      duration      campaign      previous
##      0         0         0         0         0
##      emp.var.rate cons.price.idx cons.conf.idx euribor3m nr.employed
##      0         0         0         0         0
##      y
##      0
```

```
out <- sort(nout, decreasing = TRUE)
out
```

```
##      campaign      previous      duration      age      pdays
##      146         47         21         15         6
##      job        marital      education      default      housing
##      0         0         0         0         0
##      loan      contact      month      day_of_week      poutcome
##      0         0         0         0         0
##      emp.var.rate cons.price.idx cons.conf.idx euribor3m nr.employed
##      0         0         0         0         0
##      y
##      0
```

```
ranking <- nmiss + nerrs + nout;
ranking <- sort(ranking, decreasing = TRUE);
ranking
```

```
##      pdays      poutcome      default      education      campaign
##      4799        4315        1061        207        146
##      housing      loan      previous      job      duration
##      112         112         47        43        21
##      age      marital      contact      month      day_of_week
##      15          9          0          0          0
##      emp.var.rate cons.price.idx cons.conf.idx euribor3m nr.employed
##      0           0           0           0           0
##      y
##      0
```

## 8.1 Individual

```
vmis<-rep(0, nrow(df))
nmis<-rep(0, ncol(df))
for(i in 1:nrow(df)) {
  vmis[i]<-vmis[i]+sum(is.na(df[i,]))
}
### Create variable adding the total number missing values, outliers and errors
df$outliers<-vout
df$errors<-verrs
df$missings<-vmis
```

## 9 Correlation

```
##Outliers
condes(df, num.var = 35)
```

```
## $quanti
##      correlation      p.value
## cons.price.idx  0.09277707 4.935483e-11
## duration        0.09217400 6.578690e-11
## missings         0.08750383 5.725813e-10
## cons.conf.idx    0.04206987 2.926427e-03
## campaign         -0.03485868 1.515141e-02
## nr.employed      -0.06723109 1.953740e-06
##
## $quali
##      R2      p.value
## f.prev_contacted 0.0227976441 6.937572e-27
## month            0.0177019926 2.370453e-15
## f.month          0.0177019926 2.370453e-15
## poutcome         0.0126995605 1.354103e-14
## y                0.0058927611 5.488243e-08
## job              0.0078449930 4.672357e-05
## f.job            0.0078449930 4.672357e-05
## f.jobsituation   0.0036459260 1.088007e-04
## f.season         0.0026809265 1.221971e-03
```



## education	0.0035745603	1.249920e-02
## contact	0.0007892906	4.698187e-02
## f.contact	0.0007892906	4.698187e-02
## loan	0.0012042854	4.925618e-02
## f.loan	0.0012042854	4.925618e-02
##		
## \$category		
##	Estimate	p.value
## Contacted	0.081812419	6.937572e-27
## success	0.072116782	5.134633e-14
## Month-sep	0.132112071	3.375808e-10
## sep	0.132112071	3.375808e-10
## yes	0.025739217	5.488243e-08
## Other	0.014226777	4.381288e-05
## Job-retired	0.049221530	1.038034e-04
## retired	0.049221530	1.038034e-04
## Month-oct	0.059231366	1.236697e-04
## oct	0.059231366	1.236697e-04
## Job-student	0.062269658	6.151821e-04
## student	0.062269658	6.151821e-04
## Summer	0.009445982	1.934820e-03
## Edu-basic.4y	0.027994349	1.707508e-02
## f.no	0.016447061	1.881531e-02
## no	0.016447061	1.881531e-02
## f.single	0.020057898	3.525089e-02
## single	0.020057898	3.525089e-02
## f.married	0.004782537	3.900623e-02
## married	0.004782537	3.900623e-02
## f.age-(50,95]	0.011668659	4.206059e-02
## f.telephone	0.006289677	4.698187e-02
## telephone	0.006289677	4.698187e-02
## f.cellular	-0.006289677	4.698187e-02
## cellular	-0.006289677	4.698187e-02
## failure	-0.022404431	3.040848e-02
## Edu-basic.6y	-0.022588815	2.423028e-02
## Job-services	-0.032252635	2.269021e-02
## services	-0.032252635	2.269021e-02
## Month-jul	-0.017164598	1.433353e-02
## jul	-0.017164598	1.433353e-02
## Month-jun	-0.013342474	1.180245e-02
## jun	-0.013342474	1.180245e-02
## Job-blue-collar	-0.024858369	1.096798e-02
## blue-collar	-0.024858369	1.096798e-02
## Edu-basic.9y	-0.013266682	7.404904e-03
## Month-nov	-0.064072886	6.542548e-04
## nov	-0.064072886	6.542548e-04
## Month-may	-0.048442732	3.132088e-04
## may	-0.048442732	3.132088e-04
## Spring	-0.014074703	2.863111e-04
## Worker	-0.013828160	4.717187e-05
## no	-0.025739217	5.488243e-08
## NA	-0.049712351	1.179041e-09
## No-contacted	-0.081812419	6.937572e-27

```
##Errors
#condes(df, num.var = 36) ##NO FUNCIONA, NO HI HA ERRORS
##Missings
condes(df, num.var = 37)
```

```
## $quanti
##          correlation      p.value
## emp.var.rate    0.32059379 6.915468e-120
## euribor3m       0.31925495 7.539951e-119
## nr.employed     0.31676931 6.159986e-117
## cons.price.idx  0.25210903 2.394819e-73
## age             0.13225565 6.805599e-21
## outliers        0.08750383 5.725813e-10
## cons.conf.idx   0.04099752 3.738104e-03
## campaign        0.03483158 1.523054e-02
## previous        -0.42315303 2.030358e-214
##
## $quali
##          R2      p.value
## default    0.535780204 0.000000e+00
## f.default  0.535780204 0.000000e+00
## loan        0.241643238 7.313059e-301
## f.loan      0.241643238 7.313059e-301
## housing     0.241517653 1.106051e-300
## f.housing   0.241517653 1.106051e-300
## poutcome    0.204236107 1.286462e-248
## f.prev_contacted 0.113007932 2.380717e-132
## education   0.080884953 6.578439e-87
## job         0.072449068 9.771474e-74
## f.job       0.072449068 9.771474e-74
## f.education 0.061370385 1.891864e-69
## contact     0.046652810 7.324880e-54
## f.contact   0.046652810 7.324880e-54
## month       0.049126070 4.716899e-49
## f.month     0.049126070 4.716899e-49
## y           0.029025116 7.096717e-34
## f.season    0.028568002 3.547259e-32
## f.age       0.022838134 7.424390e-25
## marital     0.022022645 5.860022e-24
## f.marital   0.022022645 5.860022e-24
## f.jobsituation 0.007767691 3.455343e-09
##
## $category
##          Estimate      p.value
## NA          1.087223059 0.000000e+00
## NA          1.087223059 0.000000e+00
## NA          2.698104360 2.243802e-302
## NA          2.688966220 2.243802e-302
## NA          2.698104360 2.243802e-302
## NA          2.688966220 2.243802e-302
## NA          1.158397313 1.944798e-227
## No-contacted 1.027183048 2.380717e-132
## NA          1.151292413 5.804433e-57
## Other        0.903536808 4.449317e-56
```

## f.telephone	0.272689164	7.324880e-54
## telephone	0.272689164	7.324880e-54
## NA	2.521893692	1.678143e-49
## NA	2.521893692	1.678143e-49
## no	0.322138053	7.096717e-34
## Job-blue-collar	0.098621784	1.020782e-18
## blue-collar	0.098621784	1.020782e-18
## Summer	0.274890250	5.082947e-18
## Edu-basic.4y	0.249890173	1.760364e-14
## NA	2.130389105	5.531075e-12
## NA	2.130389105	5.531075e-12
## f.age-(40,50]	0.184697514	9.683462e-11
## Month-jun	0.626592566	1.109488e-10
## jun	0.626592566	1.109488e-10
## f.age-(50,95]	0.197797300	5.304722e-08
## Worker	0.084409151	8.550677e-08
## Edu-basic.6y	0.209956096	5.955419e-07
## Month-jul	0.515360065	2.256065e-06
## jul	0.515360065	2.256065e-06
## Job-housemaid	0.177813354	6.666548e-04
## housemaid	0.177813354	6.666548e-04
## Month-may	0.413964198	3.280969e-03
## may	0.413964198	3.280969e-03
## Job-management	-0.307496264	3.868777e-02
## management	-0.307496264	3.868777e-02
## Job-technician	-0.266257317	2.039955e-02
## technician	-0.266257317	2.039955e-02
## Month-dec	-0.308176174	1.922055e-02
## dec	-0.308176174	1.922055e-02
## f.yes	-1.368490878	2.567887e-03
## yes	-1.368490878	2.567887e-03
## f.age-(30,40]	-0.087930286	1.231375e-04
## Edu-high.school	-0.270472752	7.333131e-05
## Mandatory	-0.310407397	1.913837e-05
## f.no	-1.341862621	4.952234e-06
## no	-1.341862621	4.952234e-06
## Job-student	-0.747641192	2.037418e-06
## student	-0.747641192	2.037418e-06
## Month-apr	-0.014058527	4.532814e-08
## apr	-0.014058527	4.532814e-08
## f.yes	-1.347103599	7.963467e-09
## yes	-1.347103599	7.963467e-09
## Month-oct	-0.439422010	7.506767e-09
## oct	-0.439422010	7.506767e-09
## Month-mar	-0.521892282	5.489624e-09
## mar	-0.521892282	5.489624e-09
## Other	-0.143651719	4.699821e-10
## Month-sep	-0.643104404	2.884680e-11
## sep	-0.643104404	2.884680e-11
## Month-nov	-0.008008424	7.921110e-12
## nov	-0.008008424	7.921110e-12
## f.married	-0.550730444	3.578250e-13
## married	-0.550730444	3.578250e-13
## f.single	-0.855957411	5.880094e-14

```
## single -0.855957411 5.880094e-14
## f.age-[17,30] -0.294564528 6.798179e-15
## Job-admin. -0.404908272 4.641734e-15
## admin. -0.404908272 4.641734e-15
## Edu-university.degree -0.426412445 4.461532e-25
## Non-Mandatory -0.593129411 1.085083e-26
## f.no -1.329613482 2.633646e-27
## no -1.329613482 2.633646e-27
## Aut-Win -0.361408450 1.598668e-27
## yes -0.322138053 7.096717e-34
## f.cellular -0.272689164 7.324880e-54
## cellular -0.272689164 7.324880e-54
## failure -0.104236689 2.835968e-95
## success -1.054160624 5.664133e-127
## Contacted -1.027183048 2.380717e-132
## f.no -1.087223059 0.000000e+00
## no -1.087223059 0.000000e+00
```

```
aggregate(df$missings, by=list(df$f.age), FUN=mean)
```

```
##          Group.1          x
## 1 f.age-[17,30] 2.153326
## 2 f.age-(30,40] 2.359960
## 3 f.age-(40,50] 2.632588
## 4 f.age-(50,95] 2.645688
```

```
aggregate(df$outliers, by=list(df$f.age), FUN=mean)
```

```
##          Group.1          x
## 1 f.age-[17,30] 0.04847802
## 2 f.age-(30,40] 0.04193709
## 3 f.age-(40,50] 0.04472843
## 4 f.age-(50,95] 0.06060606
```

```
aggregate(df$missings, by=list(df$f.jobssituation), FUN=mean)
```

```
##          Group.1          x
## 1 Self-employed 2.496774
## 2          Worker 2.521941
## 3          Other 2.293880
```

```
aggregate(df$outliers, by=list(df$f.jobssituation), FUN=mean)
```

```
##          Group.1          x
## 1 Self-employed 0.04946237
## 2          Worker 0.03603282
## 3          Other 0.06408776
```

```
aggregate(df$missings, by=list(df$f.education), FUN=mean)
```

```
##          Group.1          x
## 1      Mandatory 2.508544
## 2 Non-Mandatory 2.225822
## 3          Other 3.722488
```

```
aggregate(df$outliers, by=list(df$f.education), FUN=mean)
```

```
##          Group.1          x
```

```
## 1      Mandatory 0.04309064
## 2 Non-Mandatory 0.05050024
## 3           Other 0.06220096
```

```
aggregate(df$missings, by=list(df$f.marital), FUN=mean)
```

```
##      Group.1      x
## 1 f.divorced 2.368132
## 2 f.married 2.541103
## 3 f.single 2.235876
```

```
aggregate(df$outliers, by=list(df$f.marital), FUN=mean)
```

```
##      Group.1      x
## 1 f.divorced 0.04945055
## 2 f.married 0.04192803
## 3 f.single 0.05720339
```

## 10 Imputation

### 10.1 Numeric Variables

```
#Outliers -> missings
```

```
#Delete duration outliers
```

```
outliers<-which(df$duration>1500);length(outliers);
```

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

```
df <- df[-outliers, ]
```

```
var_num <-names(df)[c(1, 12:14)] ## age,campaign,pdays,previous
length(var_num)
```

```
## [1] 4
```

```
summary(df[,var_num])
```

```
##      age      campaign      pdays      previous
##  Min.   :17.00   Min.   :1.000   Min.   : 0.000   Min.   :0.000
## 1st Qu.:32.00   1st Qu.:1.000   1st Qu.: 3.000   1st Qu.:0.000
##  Median :38.00   Median :2.000   Median : 5.000   Median :0.000
##  Mean   :39.83   Mean   :2.269   Mean   : 5.458   Mean   :0.147
## 3rd Qu.:47.00   3rd Qu.:3.000   3rd Qu.: 7.000   3rd Qu.:0.000
##  Max.   :80.00   Max.   :9.000   Max.   :15.000   Max.   :2.000
##  NA's   :15     NA's   :145   NA's   :4778   NA's   :47
```

```
res <- imputePCA(df[,var_num],ncp=2)
summary(res$completeObs)
```

```
##      age      campaign      pdays      previous
##  Min.   :17.00   Min.   :1.000   Min.   : 0.000   Min.   :-0.007828
## 1st Qu.:32.00   1st Qu.:1.000   1st Qu.: 5.338   1st Qu.: 0.000000
##  Median :38.00   Median :2.000   Median : 5.394   Median : 0.000000
##  Mean   :39.83   Mean   :2.271   Mean   : 5.389   Mean   : 0.146942
## 3rd Qu.:47.00   3rd Qu.:3.000   3rd Qu.: 5.433   3rd Qu.: 0.000000
##  Max.   :80.00   Max.   :9.000   Max.   :15.000   Max.   : 2.000000
```

```
#S'han imputat valors negatius a previous, els posem a 0
```

```
sel <- which(res$completeObs[, "previous"] < 0)
res$completeObs[sel, "previous"] <- 0
```

```
df$age <- res$completeObs[, "age"]
df$campaign <- res$completeObs[, "campaign"]
df$pdays <- res$completeObs[, "pdays"]
df$previous <- res$completeObs[, "previous"]
```

## 10.2 Factors

```
# TODO: ADD f.yes in f.loan with value 0
```

```
factors <- names(df)[c(24, 28, 29, 31)]; # f.job, f.housing, f.marital, f.loan
summary(df[, factors])
```

```
##           f.job      f.housing      f.marital      f.loan
## Job-admin.      :1283   f.no :2220   f.divorced: 545   f.no :4120
## Job-blue-collar:1154   f.yes:2647   f.married :3018   f.yes: 747
## Job-technician : 829   NA's : 112   f.single  :1407   NA's : 112
## Job-services   : 469                      NA's      : 9
## Job-management : 343
## (Other)        : 860
## NA's           : 41
```

```
resfact <- imputeMCA(df[, factors], ncp=3);
summary(resfact$completeObs)
```

```
##           f.job      f.housing      f.marital      f.loan
## Job-admin.      :1303   f.no :2223   f.divorced: 545   f.no :4232
## Job-blue-collar:1175   f.yes:2756   f.married :3027   f.yes: 747
## Job-technician : 829                      f.single  :1407
## Job-services   : 469
## Job-management : 343
## Job-retired    : 186
## (Other)        : 674
```

```
df$f.housing <- resfact$completeObs[, "f.housing"]
df$f.marital <- resfact$completeObs[, "f.marital"]
df$f.loan <- resfact$completeObs[, "f.loan"]
df$f.job <- resfact$completeObs[, "f.job"]
```

```
#Imputem manualment poutcome ja que pensem que els que no han respós a la pregunta molt probablement ta
```

```
sel <- which(is.na(df$poutcome))
```

```
df$poutcome <- factor(df$poutcome, labels=paste("Pout", sep="-", levels(df$poutcome)))
```

```
table(df$poutcome)
```

```
##
## Pout-failure Pout-success
##           491           192
```

```

df$f.poutcome<-2

# 1 level - failure
sel<-which(df$poutcome %in% c("Pout-failure"))
df$f.poutcome[sel] <- 1
sel<- which(is.na(df$poutcome))
df$f.poutcome[sel] <- 1
table(df$f.poutcome)

##
##      1      2
## 4787  192

# 2 level - success
sel<-which(df$poutcome %in% c("Pout-success"))
df$f.poutcome[sel] <- 2
summary(df$f.education)

##      Mandatory Non-Mandatory      Other
##      2685      2086      208

df$f.poutcome<-factor(df$f.poutcome,levels=1:2,labels=c("f.Pout-failure","f.Pout-success"))
summary(df$f.poutcome)

## f.Pout-failure f.Pout-success
##      4787      192

#Imputem manualment default ja que pensem que els que no han respós a la pregunta no poden ser imputats

sel <- which(is.na(df$default))
df$f.default[sel] <- "f.no"
table(df$f.default)

##
## f.no f.si
## 4979    0

```

## 11 Profiling

```

condes(df[c(1:29, 31:34,38)],11)

## $quanti
##      correlation      p.value
## pdays      0.02993732 0.0346537787
## nr.employed -0.03189122 0.0244288564
## campaign    -0.04723473 0.0008560601
##
## $quali
##      R2      p.value
## y      0.160738690 1.169792e-191
## month   0.006371350 2.170869e-04
## f.month  0.006371350 2.170869e-04
## day_of_week 0.002933693 5.566255e-03
## f.day     0.002933693 5.566255e-03
## f.prev_contacted 0.001373939 8.903355e-03

```

```
## f.poutcome      0.001278653  1.162430e-02
## contact         0.001010183  2.491598e-02
## f.contact       0.001010183  2.491598e-02
## f.housing       0.001007355  2.511947e-02
## poutcome        0.001405185  3.024097e-02
##
## $category
##              Estimate      p.value
## yes          148.5619189 1.169792e-191
## f.day.wed     23.8425938  3.704865e-04
## wed          23.8425938  3.704865e-04
## Contacted     21.9755934  8.903355e-03
## Month-dec     122.9070700  9.727753e-03
## dec          122.9070700  9.727753e-03
## Job-self-employed 45.0705179 1.128480e-02
## self-employed 42.4765578 1.128480e-02
## f.Pout-success 21.9270496 1.162430e-02
## Pout-success  31.6018377 1.162430e-02
## Month-jul      0.4019466 1.503004e-02
## jul           0.4019466 1.503004e-02
## no            10.6116297 2.191781e-02
## f.cellular     7.8044172 2.491598e-02
## cellular       7.8044172 2.491598e-02
## f.no           7.5382090 2.511947e-02
## Mandatory     11.4647925 2.667720e-02
## f.day.mon     -13.0962186 4.765549e-02
## mon          -13.0962186 4.765549e-02
## yes           -4.7751417 2.952173e-02
## Summer        -9.4533836 2.921958e-02
## f.yes         -7.5382090 2.511947e-02
## f.telephone   -7.8044172 2.491598e-02
## telephone     -7.8044172 2.491598e-02
## Job-housemaid -48.9550116 2.360381e-02
## housemaid     -51.5489717 2.360381e-02
## f.Pout-failure -21.9270496 1.162430e-02
## No-contacted  -21.9755934 8.903355e-03
## Month-aug     -40.1682151 4.976285e-03
## aug          -40.1682151 4.976285e-03
## Month-jun     -43.5063607 1.930997e-03
## jun          -43.5063607 1.930997e-03
## no           -148.5619189 1.169792e-191
```

```
catdes(df, num.var = 21)
```

```
##
## Link between the cluster variable and the categorical variables (chi-square test)
## =====
##              p.value df
## f.default      0.000000e+00 1
## f.prev_contacted 1.746438e-113 1
## poutcome       1.256455e-110 2
## f.poutcome      6.599570e-109 1
## month          2.092803e-78 9
## f.month         2.092803e-78 9
## f.job           2.831986e-27 10
```



```

## job                6.520196e-27 11
## contact            7.944988e-25 1
## f.contact          7.944988e-25 1
## default            1.313876e-11 1
## f.jobssituation    3.313476e-08 2
## f.age              4.789647e-08 3
## f.season           5.088671e-08 2
## f.marital          1.549949e-05 2
## marital            3.916274e-05 3
## education          8.492460e-05 7
## f.education        7.801545e-03 2
##
## Description of each cluster by the categories
## =====
## $no
##
## Cla/Mod    Mod/Cla    Global
## f.prev_contacted=No-contacted 90.65577 98.1624319 95.8626230
## f.poutcome=f.Pout-failure 90.53687 98.3212341 96.1438040
## poutcome=NA 91.13128 88.8157895 86.2823860
## f.contact=f.telephone 94.68733 38.8157895 36.2924282
## contact=telephone 94.68733 38.8157895 36.2924282
## f.month=Month-may 93.12612 35.3448276 33.6011247
## month=may 93.12612 35.3448276 33.6011247
## default=NA 94.41816 22.6406534 21.2291625
## job=blue-collar 94.02080 24.6143376 23.1773448
## f.job=Job-blue-collar 93.95745 25.0453721 23.5991163
## f.jobssituation=Worker 90.30411 57.2595281 56.1357702
## f.marital=f.married 89.85795 61.7059891 60.7953404
## marital=married 89.82770 61.5018149 60.6145812
## f.age=f.age-(30,40] 90.27569 40.8575318 40.0682868
## education=Edu-basic.9y 91.80978 15.7667877 15.2038562
## f.age=f.age-(40,50] 90.93098 25.7032668 25.0251054
## f.education=Mandatory 89.68343 54.6279492 53.9264913
## f.month=Month-nov 91.58317 10.3675136 10.0220928
## month=nov 91.58317 10.3675136 10.0220928
## f.season=Summer 89.63226 45.8938294 45.3303876
## f.month=Month-jun 90.85366 13.5208711 13.1753364
## month=jun 90.85366 13.5208711 13.1753364
## f.jobssituation=Self-employed 91.32321 9.5508167 9.2588873
## f.job=Job-management 85.13120 6.6243194 6.8889335
## job=management 85.13120 6.6243194 6.8889335
## poutcome=Pout-failure 85.33605 9.5054446 9.8614180
## f.month=Month-dec 63.15789 0.2722323 0.3816027
## month=dec 63.15789 0.2722323 0.3816027
## f.age=f.age-(50,95] 85.12881 16.4927405 17.1520386
## education=Edu-university.degree 86.07595 27.7676951 28.5599518
## f.age=f.age-[17,30] 84.50226 16.9464610 17.7545692
## f.job=Job-retired 76.88172 3.2441016 3.7356899
## job=retired 76.88172 3.2441016 3.7356899
## f.marital=f.single 85.14570 27.1778584 28.2586865
## marital=single 85.14570 27.1778584 28.2586865
## f.month=Month-apr 78.57143 5.7395644 6.4671621
## month=apr 78.57143 5.7395644 6.4671621
## f.season=Aut-Win 81.84569 12.2731397 13.2757582

```

## f.jobsituation=Other	84.91004	33.1896552	34.6053424
## default=no	86.94544	77.3593466	78.7708375
## f.job=Job-student	59.59596	1.3384755	1.9883511
## job=student	59.59596	1.3384755	1.9883511
## f.month=Month-sep	50.00000	0.7486388	1.3255674
## f.month=Month-mar	50.00000	0.7486388	1.3255674
## month=sep	50.00000	0.7486388	1.3255674
## month=mar	50.00000	0.7486388	1.3255674
## f.month=Month-oct	50.64935	0.8847550	1.5464953
## month=oct	50.64935	0.8847550	1.5464953
## f.contact=f.cellular	85.02522	61.1842105	63.7075718
## contact=cellular	85.02522	61.1842105	63.7075718
## f.poutcome=f.Pout-success	38.54167	1.6787659	3.8561960
## poutcome=Pout-success	38.54167	1.6787659	3.8561960
## f.prev_contacted=Contacted	39.32039	1.8375681	4.1373770
##		p.value	v.test
## f.prev_contacted=No-contacted	1.227915e-68		17.508783
## f.poutcome=f.Pout-failure	1.666964e-65		17.093224
## poutcome=NA	5.763783e-38		12.880929
## f.contact=f.telephone	2.440539e-27		10.831526
## contact=telephone	2.440539e-27		10.831526
## f.month=Month-may	6.034473e-14		7.507332
## month=may	6.034473e-14		7.507332
## default=NA	4.251760e-13		7.247295
## job=blue-collar	1.334762e-12		7.090658
## f.job=Job-blue-collar	1.422485e-12		7.081844
## f.jobsituation=Worker	9.925677e-06		4.418786
## f.marital=f.married	2.858318e-04		3.627813
## marital=married	4.111959e-04		3.532792
## f.age=f.age-(30,40]	1.478138e-03		3.178942
## education=Edu-basic.9y	1.493774e-03		3.175890
## f.age=f.age-(40,50]	1.761035e-03		3.127827
## f.education=Mandatory	5.915694e-03		2.752418
## f.month=Month-nov	2.035659e-02		2.319710
## month=nov	2.035659e-02		2.319710
## f.season=Summer	2.628277e-02		2.222008
## f.month=Month-jun	4.135002e-02		2.040003
## month=jun	4.135002e-02		2.040003
## f.jobsituation=Self-employed	4.326162e-02		2.021175
## f.job=Job-management	4.686185e-02		-1.987547
## job=management	4.686185e-02		-1.987547
## poutcome=Pout-failure	2.293362e-02		-2.274539
## f.month=Month-dec	4.353220e-03		-2.851363
## month=dec	4.353220e-03		-2.851363
## f.age=f.age-(50,95]	8.677261e-04		-3.330235
## education=Edu-university.degree	7.167173e-04		-3.383103
## f.age=f.age-[17,30]	6.154673e-05		-4.006801
## f.job=Job-retired	4.756560e-06		-4.575248
## job=retired	4.756560e-06		-4.575248
## f.marital=f.single	4.173805e-06		-4.602534
## marital=single	4.173805e-06		-4.602534
## f.month=Month-apr	1.145794e-07		-5.301939
## month=apr	1.145794e-07		-5.301939
## f.season=Aut-Win	4.630214e-08		-5.464956

## f.jobsituation=Other	9.853123e-09	-5.733238	
## default=no	4.251760e-13	-7.247295	
## f.job=Job-student	9.483114e-14	-7.447909	
## job=student	9.483114e-14	-7.447909	
## f.month=Month-sep	8.048922e-15	-7.766807	
## f.month=Month-mar	8.048922e-15	-7.766807	
## month=sep	8.048922e-15	-7.766807	
## month=mar	8.048922e-15	-7.766807	
## f.month=Month-oct	1.184930e-16	-8.284614	
## month=oct	1.184930e-16	-8.284614	
## f.contact=f.cellular	2.440539e-27	-10.831526	
## contact=cellular	2.440539e-27	-10.831526	
## f.poutcome=f.Pout-success	1.666964e-65	-17.093224	
## poutcome=Pout-success	1.666964e-65	-17.093224	
## f.prev_contacted=Contacted	1.227915e-68	-17.508783	
##			
## \$yes			
##	Cla/Mod	Mod/Cla	Global
## f.prev_contacted=Contacted	60.679612	21.891419	4.1373770
## f.poutcome=f.Pout-success	61.458333	20.665499	3.8561960
## poutcome=Pout-success	61.458333	20.665499	3.8561960
## f.contact=f.cellular	14.974779	83.187391	63.7075718
## contact=cellular	14.974779	83.187391	63.7075718
## f.month=Month-oct	49.350649	6.654991	1.5464953
## month=oct	49.350649	6.654991	1.5464953
## f.month=Month-sep	50.000000	5.779335	1.3255674
## f.month=Month-mar	50.000000	5.779335	1.3255674
## month=sep	50.000000	5.779335	1.3255674
## month=mar	50.000000	5.779335	1.3255674
## f.job=Job-student	40.404040	7.005254	1.9883511
## job=student	40.404040	7.005254	1.9883511
## default=no	13.054564	89.667250	78.7708375
## f.jobsituation=Other	15.089959	45.534151	34.6053424
## f.season=Aut-Win	18.154312	21.015762	13.2757582
## f.month=Month-apr	21.428571	12.084063	6.4671621
## month=apr	21.428571	12.084063	6.4671621
## f.marital=f.single	14.854300	36.602452	28.2586865
## marital=single	14.854300	36.602452	28.2586865
## f.job=Job-retired	23.118280	7.530648	3.7356899
## job=retired	23.118280	7.530648	3.7356899
## f.age=f.age-[17,30]	15.497738	23.992995	17.7545692
## education=Edu-university.degree	13.924051	34.676007	28.5599518
## f.age=f.age-(50,95]	14.871194	22.241681	17.1520386
## f.month=Month-dec	36.842105	1.225919	0.3816027
## month=dec	36.842105	1.225919	0.3816027
## poutcome=Pout-failure	14.663951	12.609457	9.8614180
## f.job=Job-management	14.868805	8.931699	6.8889335
## job=management	14.868805	8.931699	6.8889335
## f.jobsituation=Self-employed	8.676790	7.005254	9.2588873
## f.month=Month-jun	9.146341	10.507881	13.1753364
## month=jun	9.146341	10.507881	13.1753364
## f.season=Summer	10.367745	40.980736	45.3303876
## f.month=Month-nov	8.416834	7.355517	10.0220928
## month=nov	8.416834	7.355517	10.0220928

## f.education=Mandatory	10.316574	48.511384	53.9264913
## f.age=f.age-(40,50]	9.069021	19.789842	25.0251054
## education=Edu-basic.9y	8.190225	10.858144	15.2038562
## f.age=f.age-(30,40]	9.724311	33.975482	40.0682868
## marital=married	10.172300	53.765324	60.6145812
## f.marital=f.married	10.142055	53.765324	60.7953404
## f.job=situation=Worker	9.695886	47.460595	56.1357702
## f.job=Job-blue-collar	6.042553	12.434326	23.5991163
## job=blue-collar	5.979203	12.084063	23.1773448
## default=NA	5.581835	10.332750	21.2291625
## f.month=Month-may	6.873879	20.140105	33.6011247
## month=may	6.873879	20.140105	33.6011247
## f.contact=f.telephone	5.312673	16.812609	36.2924282
## contact=telephone	5.312673	16.812609	36.2924282
## poutcome=NA	8.868715	66.725044	86.2823860
## f.poutcome=f.Pout-failure	9.463129	79.334501	96.1438040
## f.prev_contacted=No-contacted	9.344228	78.108581	95.8626230
##	p.value	v.test	
## f.prev_contacted=Contacted	1.227915e-68	17.508783	
## f.poutcome=f.Pout-success	1.666964e-65	17.093224	
## poutcome=Pout-success	1.666964e-65	17.093224	
## f.contact=f.cellular	2.440539e-27	10.831526	
## contact=cellular	2.440539e-27	10.831526	
## f.month=Month-oct	1.184930e-16	8.284614	
## month=oct	1.184930e-16	8.284614	
## f.month=Month-sep	8.048922e-15	7.766807	
## f.month=Month-mar	8.048922e-15	7.766807	
## month=sep	8.048922e-15	7.766807	
## month=mar	8.048922e-15	7.766807	
## f.job=Job-student	9.483114e-14	7.447909	
## job=student	9.483114e-14	7.447909	
## default=no	4.251760e-13	7.247295	
## f.job=situation=Other	9.853123e-09	5.733238	
## f.season=Aut-Win	4.630214e-08	5.464956	
## f.month=Month-apr	1.145794e-07	5.301939	
## month=apr	1.145794e-07	5.301939	
## f.marital=f.single	4.173805e-06	4.602534	
## marital=single	4.173805e-06	4.602534	
## f.job=Job-retired	4.756560e-06	4.575248	
## job=retired	4.756560e-06	4.575248	
## f.age=f.age-[17,30]	6.154673e-05	4.006801	
## education=Edu-university.degree	7.167173e-04	3.383103	
## f.age=f.age-(50,95]	8.677261e-04	3.330235	
## f.month=Month-dec	4.353220e-03	2.851363	
## month=dec	4.353220e-03	2.851363	
## poutcome=Pout-failure	2.293362e-02	2.274539	
## f.job=Job-management	4.686185e-02	1.987547	
## job=management	4.686185e-02	1.987547	
## f.job=situation=Self-employed	4.326162e-02	-2.021175	
## f.month=Month-jun	4.135002e-02	-2.040003	
## month=jun	4.135002e-02	-2.040003	
## f.season=Summer	2.628277e-02	-2.222008	
## f.month=Month-nov	2.035659e-02	-2.319710	
## month=nov	2.035659e-02	-2.319710	

```

## f.education=Mandatory      5.915694e-03 -2.752418
## f.age=f.age-(40,50]       1.761035e-03 -3.127827
## education=Edu-basic.9y    1.493774e-03 -3.175890
## f.age=f.age-(30,40]       1.478138e-03 -3.178942
## marital=married           4.111959e-04 -3.532792
## f.marital=f.married       2.858318e-04 -3.627813
## f.job=situation=Worker     9.925677e-06 -4.418786
## f.job=Job-blue-collar     1.422485e-12 -7.081844
## job=blue-collar           1.334762e-12 -7.090658
## default=NA                4.251760e-13 -7.247295
## f.month=Month-may         6.034473e-14 -7.507332
## month=may                 6.034473e-14 -7.507332
## f.contact=f.telephone     2.440539e-27 -10.831526
## contact=telephone         2.440539e-27 -10.831526
## poutcome=NA              5.763783e-38 -12.880929
## f.poutcome=f.Pout-failure 1.666964e-65 -17.093224
## f.prev_contacted=No-contacted 1.227915e-68 -17.508783
##
##
## Link between the cluster variable and the quantitative variables
## =====
##
##              Eta2      P-value
## duration      0.160738690 1.169792e-191
## nr.employed   0.120745600 2.760101e-141
## euribor3m     0.087576045 3.394474e-101
## emp.var.rate  0.081696867 3.070308e-94
## previous      0.042285141 1.108497e-48
## missings      0.030232449 4.294241e-35
## cons.price.idx 0.017977169 2.070129e-21
## cons.conf.idx 0.008055104 2.236211e-10
## campaign      0.005633367 1.143924e-07
## outliers      0.002285655 7.393671e-04
##
## Description of each cluster by quantitative variables
## =====
## $no
##
##          v.test Mean in category Overall mean sd in category
## nr.employed 24.516762 5177.09015426 5168.16794537 63.4164811
## euribor3m   20.879501 3.82261162 3.63896766 1.6252539
## emp.var.rate 20.166482 0.26197822 0.10236995 1.4608943
## missings    12.267727 2.51610708 2.44004820 1.1967189
## cons.price.idx 9.459934 93.61071461 93.58315164 0.5520673
## campaign     5.295555 2.31459362 2.27063047 1.6629733
## outliers    -3.373128 0.03924682 0.04277967 0.1976554
## cons.conf.idx -6.332323 -40.74344374 -40.59754971 4.2611089
## previous    -14.508461 0.11734053 0.14694311 0.3499570
## duration    -28.287050 222.56442831 256.63908415 194.8113004
##
##          Overall sd      p.value
## nr.employed 71.3410455 9.788565e-133
## euribor3m   1.7241963 8.224659e-97
## emp.var.rate 1.5515130 1.928802e-90
## missings    1.2153909 1.349895e-34
## cons.price.idx 0.5711736 3.081386e-21
## campaign    1.6274493 1.186556e-07

```

```

## outliers      0.2053159  7.431944e-04
## cons.conf.idx 4.5165276  2.414979e-10
## previous      0.3999801  1.070970e-47
## duration      236.1424174 4.987506e-176
##
## $yes
##              v.test Mean in category Overall mean sd in category
## duration      28.287050      519.68826620 256.63908415 339.2762889
## previous      14.508461       0.37546879  0.14694311  0.6249775
## cons.conf.idx  6.332323     -39.47127846 -40.59754971  6.0227719
## outliers      3.373128       0.07005254  0.04277967  0.2552355
## campaign      -5.295555       1.93124423  2.27063047  1.2712707
## cons.price.idx -9.459934      93.37037128 93.58315164  0.6639079
## missings     -12.267727       1.85288967  2.44004820  1.1981032
## emp.var.rate  -20.166482      -1.12977233  0.10236995  1.6732331
## euribor3m     -20.879501       2.22127496  3.63896766  1.8058259
## nr.employed   -24.516762     5099.29036778 5168.16794537 89.3017729
##              Overall sd      p.value
## duration      236.1424174 4.987506e-176
## previous      0.3999801  1.070970e-47
## cons.conf.idx  4.5165276  2.414979e-10
## outliers      0.2053159  7.431944e-04
## campaign      1.6274493  1.186556e-07
## cons.price.idx 0.5711736  3.081386e-21
## missings      1.2153909  1.349895e-34
## emp.var.rate   1.5515130  1.928802e-90
## euribor3m      1.7241963  8.224659e-97
## nr.employed    71.3410455 9.788565e-133

```