DELIVERABLE 2: BANK MARKETING DATA: CASE STUDY

INDEX

Deliverable2

Montserrrat Martinez i Aleix Costa

21 de abril de 2019

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’)

Principal Component Analysis (PCA)

L'anàlisi de components principals (a partir d'ara PCA) és una tècnica utilitzada per reduir la dimensionalitat d'un conjunt de dades per a poder-les representar gràficament en gràfics de dues o tres dimensions agrupant diverses variables de les dades en factors, o components, compostos per l'agrupació de diverses variables.

Intuïtivament, la tècnica serveix per determinar el nombre de factors explicatius d'un conjunt de dades que determinen en major grau la variabilitat d'aquestes dades. Llavors podrem sintetitzar i visualitzar informació util en un conjunt de dades que contindrà observacions descrites per múltiples variables quantitatives correlacionades.

Com hem pogut observar a la nostra mostra o conjunt de dades, tenim un elevat nombre de variables i això ens dificulta la visualització de la informació que volem tractar en un espai multi-dimensional.

Gràcies al procediment explicat aconseguirem reduir la dimensionalitat de les nostes dades en un baix nombre de components que podrem visualitzar gràficament amb la menor pèrdua de informació i variança possible.

Data format and analysis

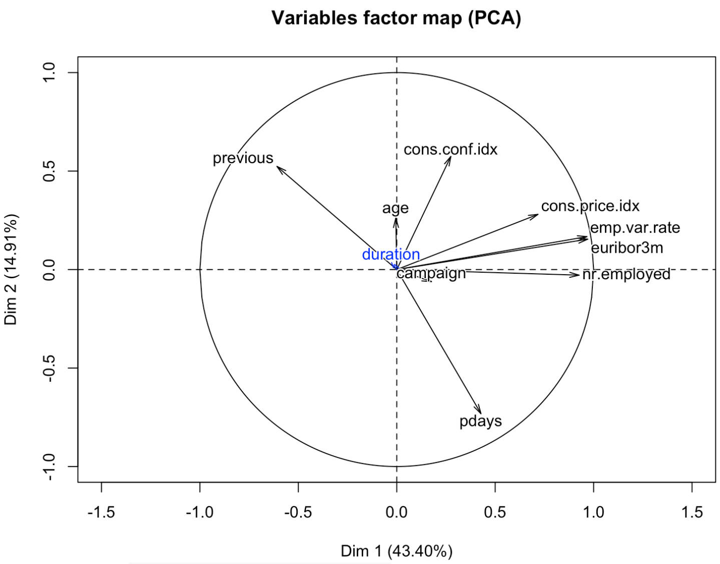
Abans de res, prepararem les dades necessàries per realitzar l'anàlisi de components principals. Escollirem les variables actives que ens permetran realitzar el PCA i també seleccionarem un conjunt de variables suplementàries.

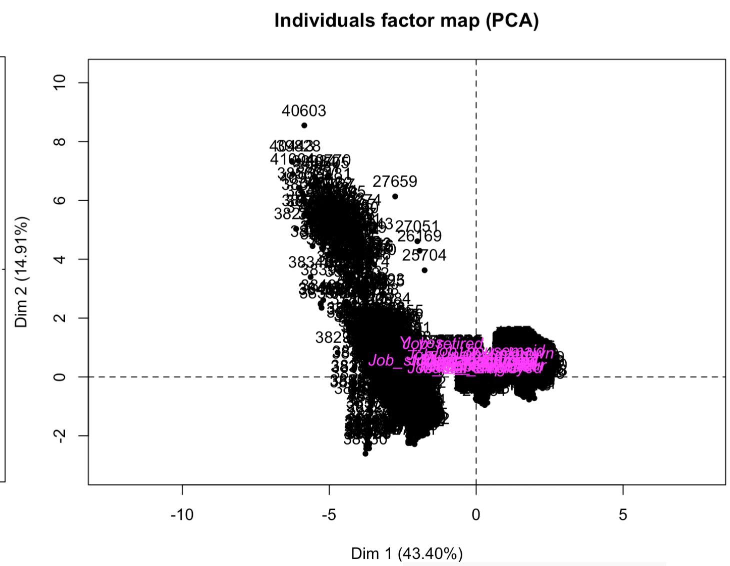
Create PCA

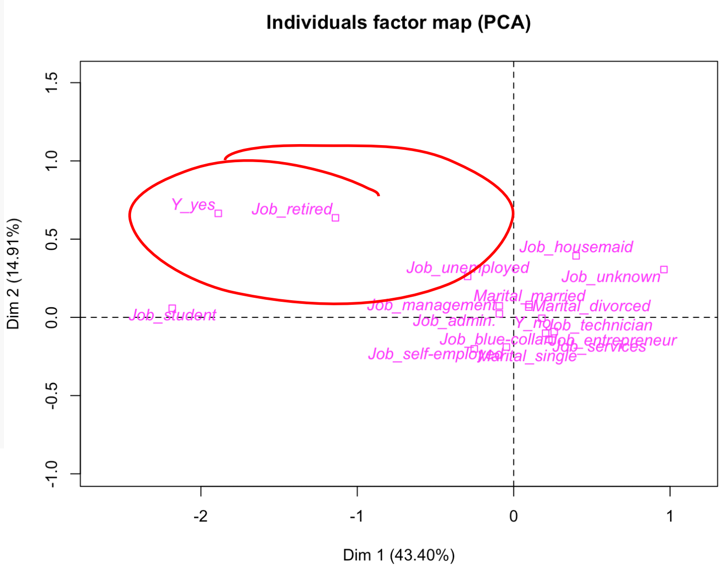
Hem agrupat totes les variables numèriques, les quals utilitzarem com a variables actives menys el target numèric "duration" i com a variables suplementàries tenim "y", "marital" y "job", encara que havíem també seleccionat "education", però la mostra no era del tot concluent.

```
names(df)

##  [1] "age"                    "job"
##  [3] "marital"                "education"
##  [5] "default"                "housing"
##  [7] "loan"                   "contact"
##  [9] "month"                  "day_of_week"
## [11] "duration"               "campaign"
## [13] "pdays"                  "previous"
## [15] "poutcome"               "emp.var.rate"
## [17] "cons.price.idx"         "cons.conf.idx"
## [19] "euribor3m"              "nr.employed"
## [21] "y"                      "missings_indiv"
## [23] "errors_indiv"           "outliers_indiv"
## [25] "season"                 "factor_age"
```

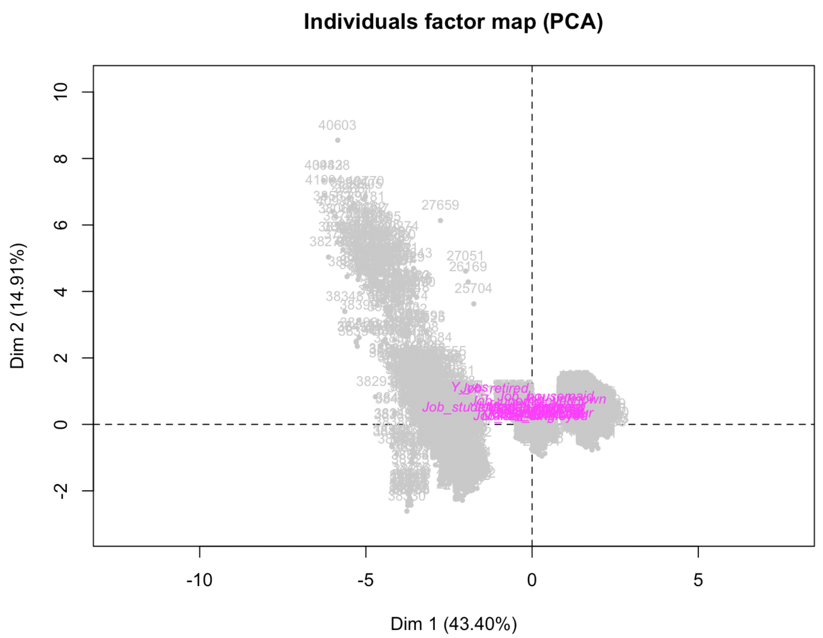
```
## [27] "factor_duration"        "factor_campaign"
## [29] "factor_Pdays"           "factor_Previous"
## [31] "factor_emp.var.rate"    "factor_cons.price.idx"
## [33] "factor_cons.conf.idx"   "factor_euribor3m"
## [35] "factor_nr.employed"

vars_conaux <- names(df)[c(1,12:14,16:20)]
vars_conaux

## [1] "age"              "campaign"        "pdays"           "previous"
## [5] "emp.var.rate"     "cons.price.idx" "cons.conf.idx"   "euribor3m"
## [9] "nr.employed"

res.pca<-PCA(df[,c("duration","y","marital","job",vars_conaux)],quanti.sup =
1,quali.sup = 2:4)
```







```
#LES VARIABLES ACTIVES NO PODEN SER FACTORS!
plot.PCA(res.pca,choix = "ind", invisible = "ind")

plot(res.pca,choix="ind", cex=0.75, col.ind="grey80")
```




```
#par(mfrow=c(1,2)) poner dos graficos juntos!
```

La funció PCA() ha realitzat el PCA del nostre conjunt de dades. Visualitzarem dos gràfics, tenim el "Variables factor map" i el "Individuals factor map" que detallarem amb més profunditat posteriorment.

En el gràfic "Variables factor map" podem observar que les variables "previous"" i "pdays" es troben totalment oposades i també veiem que el nostre target (variable quantitativa suplementària) "duration" no té res a veure amb les variables numèriques ja que la fletxa és molt curta.

Eigenvalues and dominant axes Analysis

En aquest apartat utilitzarem valors propis (Eigenvalues) per determinar quins components principals considerarem per el nostre anàlisi (denominat axes).

Concretament els valors propis mesuren la quantitat de variança proporcionada per cada component principal. A partir d'aquesta informació i les regles de Kaiser i Elbow podrem determinar, com hem dit, els components a considerar i les dimensions necessàries a agafar.

Kaiser Rule

```
res.pca$eig

## eigenvalue percentage of variance cumulative percentage of variance
## comp 1 3.90643762            43.4048625                         43.40486
## comp 2 1.34224472            14.9138303                         58.31869
## comp 3 1.03534030            11.5037811                         69.82247
## comp 4 0.98070837            10.8967597                         80.71923
## comp 5 0.84014761             9.3349735                         90.05421
## comp 6 0.46176101             5.1306779                         95.18488
## comp 7 0.39576928             4.3974364                         99.58232
## comp 8 0.02438733             0.2709704                         99.85329
## comp 9 0.01320375             0.1467083                        100.00000
```

Quan executem aquesta comanda podem visualitzar una taula on observem els valors propis (eigenvalues) de cada component principal.

La primera columna mostra el valor propi per cada component, la suma de tots els valors propis ens dóna una variança de 9. En la segona columna podem observar la proporció de variança de cada component i en la tercera el percentatge acumulat de variança obtingut a partir de la suma dels successius components.

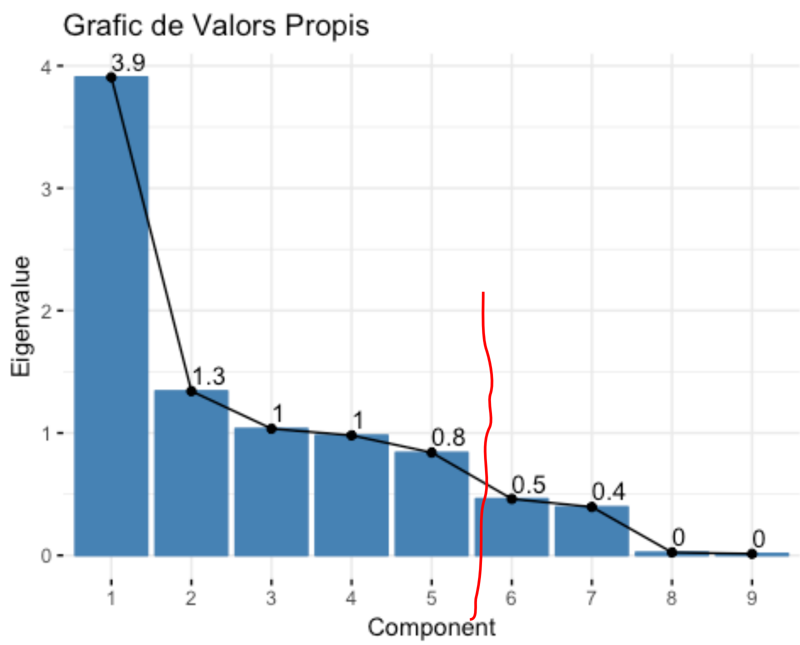
La regla de Kaiser diu que un valor propi (eigenvalue) amb valor superior a 1 indica que les components principals compten amb més variança que una de les variables originals en dades estandaritzades.

Després de l'execució, a partir de la taula de valors propis i seguint la regla de Kaiser hem decidit tenir en compte les 4 primeres components principals. Com podem veure el valor propi de la component numero 4 no supera el valor 1, però el seu valor es de 0.9807 que es molt proxim a 1, llavors també es podria considerar agafar-la. Amb el nostre percentatge de variança (69.822) podem dir que quasi tres quarts (75%) de les nostres dades queden representades amb aquestes 3 components principals i si agaféssim les 4 components seria una mica més de tres quarts de les nostres dades (80.719).

Elbow Rule

També tenim un altre mètode d'interpretació i validació de les nostres components i aquest és el "Elbow Rule", que utilitza un gràfic dels valors propis ordenats de major a menor i determina el nombre de components principals a considerar fins al punt del gràfic en el qual el valor propi és relativament petit.

```
fviz_eig(res.pca, choice = "eigenvalue", addlabels = TRUE, main = "Grafic de Valors
Propis", xlab = "Component", ylab = “Eigenvalue")
```



Com podem observar al gràfic dels valors propis, segons la regla d'elbow hauríem de considerar les 7 primeres components principals. Tot i així, en el nostre cas, decidim considerar les 3 primeres components principals ja que ens proporcionen una variança totalment acceptable (80.71%) i en el cas d'utilitzar les 7 components obtindrem una dimensionalitat massa elevada, fet que no ens interessa molt.

Individuals point of view

En aquest apartat estudiarem diferents aspectes del nostre conjunt de dades i de les nostres components principals a partir del individus de la nostra mostra.

Individuals contribution

Ara el que farem es estudiar les possibles contribucions per part d'alguns individus.

```
#Hacemos esto para poder ver los tres más contributivos al segundo eje de las 4
dimensiones que hemos cogido

sort(res.pca$ind$contrib[,1],decreasing = TRUE)[1:3]

##     40443     41004     38275
## 0.2035832 0.2016805 0.1941485
```

```r
#Se ha de hacer con which
df["40443",]
```

```
##       age        job        marital                    education    default
## 40443  26 Job_admin. Marital_single Education_university.degree Default_no
##          housing    loan          contact     month     day_of_week
## 40443 Housing_no Loan_no Contact_cellular Month_aug Day_of_week_mon
##       duration campaign pdays previous         poutcome emp.var.rate
## 40443      242        1     6        5 Poutcome_success         -1.7
##       cons.price.idx cons.conf.idx euribor3m nr.employed     y
## 40443         94.027         -38.3     0.904      4991.6 Y_yes
##       missings_indiv errors_indiv outliers_indiv season         factor_age
## 40443              0            0              0 Summer factor_age [17,31]
##                 factor_duration       factor_campaign        factor_Pdays
## 40443 factor_duration-(236,329] factor_campaign-[1,2] factor_Pdays-[0,15]
##             factor_Previous             factor_emp.var.rate
## 40443 factor_Previous-(1,5] factor_emp.var.rate-(-1.8,-0.1]
##                 factor_cons.price.idx             factor_cons.conf.idx
## 40443 factor_cons.price.idx-(94,94.8] factor_cons.conf.idx-(-40.3,-36.4]
##                     factor_euribor3m                   factor_nr.employed
## 40443 factor_euribor3m-[0.634,1.266] factor_nr.employed-[4.96e+03,5.1e+03]
```

```r
sort(res.pca$ind$contrib[,2],decreasing = TRUE)[1:3]
```

```
##     40603     39828     40443
## 1.1009452 0.8130194 0.8116665
```

```r
df["40603",]
```

```r
sort(res.pca$ind$contrib[,3],decreasing = TRUE)[1:3]
```

```
##     40930     41004     37819
## 0.7201366 0.5128497 0.4860395
```

```r
df["40930",]
```

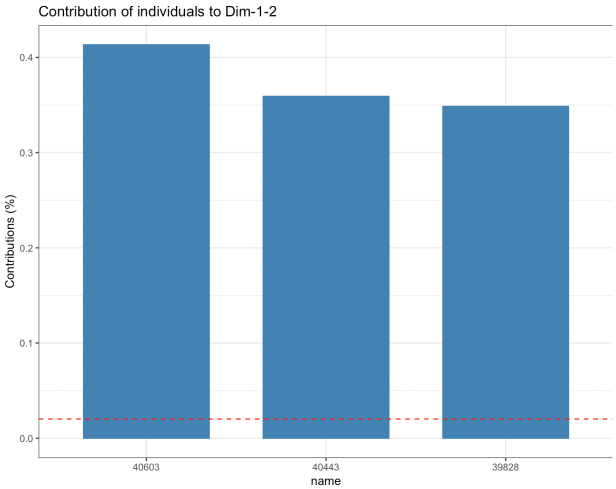
```r
sort(res.pca$ind$contrib[,4],decreasing = TRUE)[1:3]
```

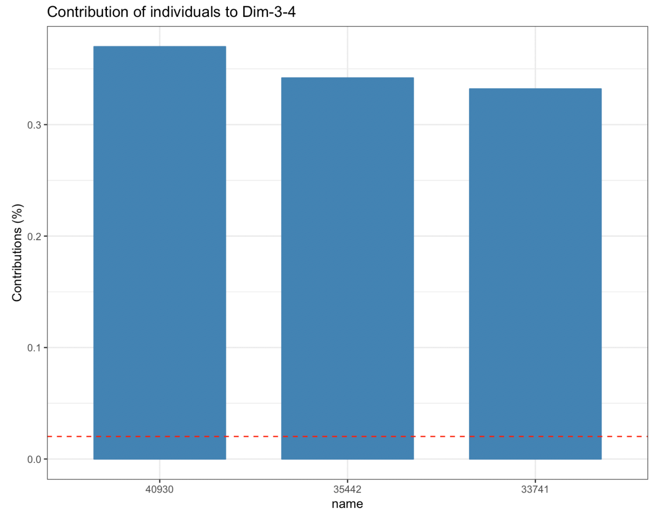
```
##     35442     33741     11630
## 0.6914135 0.6822475 0.6640766
```

```r
df["35442",]
```

```r
#fviz_pca_var(res.pca)
fviz_contrib(res.pca, choice = "ind", axes = 1:2, top = 3)+theme_bw()
```

```r
fviz_contrib(res.pca, choice = "ind", axes = 3:4, top = 3)+theme_bw()
```






```
#Aqui fem el mateix pero separant les
dimensions per fer-ho en un únic gràfic
#fviz_contrib(res.pca, choice = "ind", axes = 1:4, top = 3)+theme_bw()
```

A partir dels dos gràfics anteriors veiem que per cada parell de dimensions hi ha individus determinats que tenen una contribución elevada.

Individuals best representation

Ara veurem els individuals que tenen una millor representació

```
#Millor representats

sort(res.pca$ind$cos2[,1],decreasing = TRUE)[1:3]

##     38571     38490     38345
## 0.8867685 0.8752577 0.8582645

df["38571",]

##       age            job        marital                    education
## 38571  34 Job_technician Marital_single Education_university.degree
##          default    housing    loan          contact     month
## 38571 Default_no Housing_no Loan_no Contact_cellular Month_oct
##           day_of_week duration campaign pdays previous         poutcome
## 38571 Day_of_week_thu      136        1    16        1 Poutcome_failure
##       emp.var.rate cons.price.idx cons.conf.idx euribor3m nr.employed
## 38571         -3.4         92.431     -42.33883     0.722      5017.5
##           y missings_indiv errors_indiv outliers_indiv  season
## 38571 Y_yes              1            0              1 Aut-Win.
```

```
##                factor_age           factor_duration       factor_campaign
## 38571 factor_age (31,36] factor_duration-(104,139] factor_campaign-[1,2]
##                factor_Pdays       factor_Previous
## 38571 factor_Pdays-(15,17] factor_Previous-[0,1]
##                    factor_emp.var.rate           factor_cons.price.idx
## 38571 factor_emp.var.rate-[-3.4,-1.8] factor_cons.price.idx-[92.2,93]
##                    factor_cons.conf.idx               factor_euribor3m
## 38571 factor_cons.conf.idx-(-46.2,-42] factor_euribor3m-[0.634,1.266]
##                           factor_nr.employed
## 38571 factor_nr.employed-[4.96e+03,5.1e+03]
```

```
sort(res.pca$ind$cos2[,2],decreasing = TRUE)[1:3]
```

```
##     40603     39181     39505
## 0.5929517 0.5861391 0.5856818
```

```
df["40603",]
```

```
sort(res.pca$ind$cos2[,3],decreasing = TRUE)[1:3]
```

```
##     37819     27018     26458
## 0.7361513 0.6887437 0.6855514
```

```
df["37819",]
```

```
sort(res.pca$ind$cos2[,4],decreasing = TRUE)[1:3]
```

```
##     26278     16663     12711
## 0.8875421 0.8809677 0.8802130
```

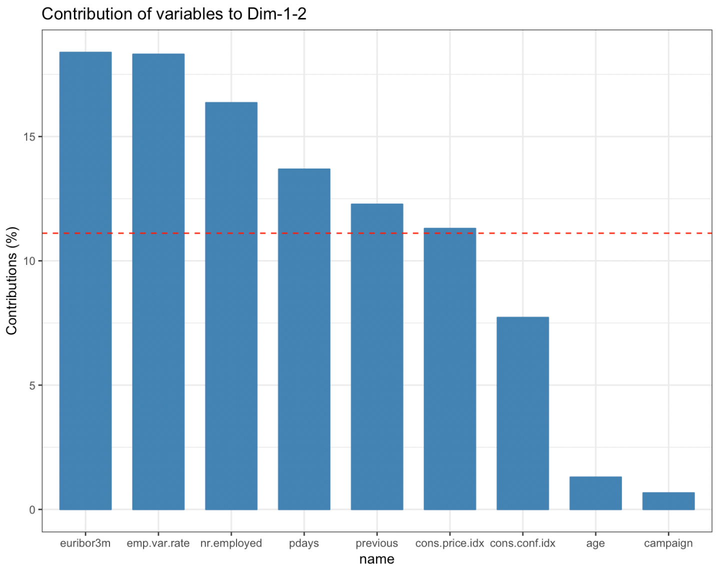
```
df["26278",]
```

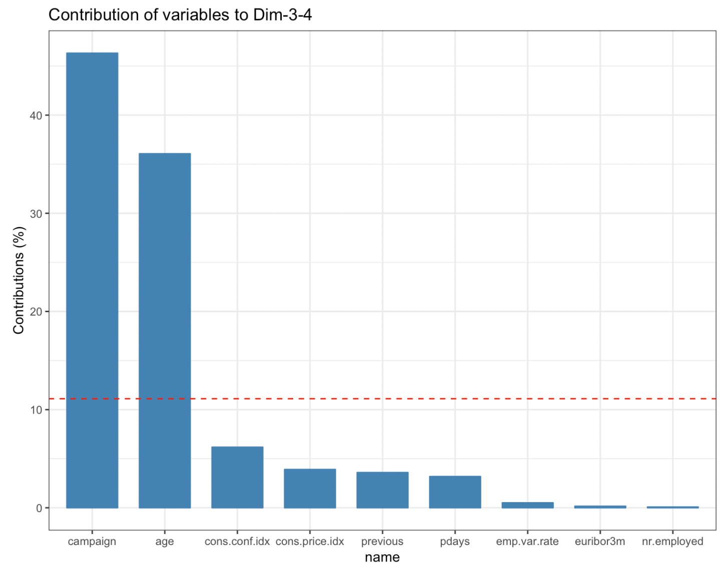
```
# Quality of individuals
head(res.pca$ind$cos2)
```

```
##         Dim.1     Dim.2      Dim.3      Dim.4      Dim.5
## 8  0.5323093 0.1191539 0.03523428 0.22230641 0.02308846
## 16 0.3307334 0.1613837 0.37792288 0.06591536 0.02620388
## 23 0.3130605 0.1604586 0.40085486 0.05815995 0.03190719
## 32 0.2502604 0.1544261 0.47812305 0.03413490 0.05539532
## 39 0.5323093 0.1191539 0.03523428 0.22230641 0.02308846
## 59 0.3130605 0.1604586 0.40085486 0.05815995 0.03190719
```

Variables contribution

```
fviz_contrib(res.pca, choice = "var", axes = 1:2)+theme_bw()

fviz_contrib(res.pca, choice = "var", axes = 3:4)+theme_bw()
```






```
#fviz_contrib(res.pca, choice = "var", axes = 1:4)+theme_bw()
```

Com podem veure en els gràfics que surten despres d'executar les comandes anteriors, podem veure que les variables que tenen més contribució o els individuals més contributius són els corresponents a les variables "euribor3m", "emp.var.rate", i "nr.employed", això pel que fa a la dim 1-2 i a la dim 3-4 tenim les variables "campaign" i "age" com les més destacades.

Interpreting the axes

```
summary(res.pca, nb.dec = 2,ncp = 4)

##
## Call:
## PCA(X = df[, c("duration", "y", "marital", "job", vars_conaux)],
##      quanti.sup = 1, quali.sup = 2:4)
##
##
## Eigenvalues
##                        Dim.1  Dim.2  Dim.3  Dim.4  Dim.5  Dim.6  Dim.7
## Variance                3.91   1.34   1.04   0.98   0.84   0.46   0.40
## % of var.              43.40  14.91  11.50  10.90   9.33   5.13   4.40
## Cumulative % of var.   43.40  58.32  69.82  80.72  90.05  95.18  99.58
##                        Dim.8  Dim.9
## Variance                0.02   0.01
## % of var.               0.27   0.15
## Cumulative % of var.   99.85 100.00
##
```

```
## Individuals (the 10 first)
##                       Dist    Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3
## 8                   |  1.75 |  1.28  0.01  0.53 |  0.60  0.01  0.12 |  0.33
## 16                  |  2.21 |  1.27  0.01  0.33 |  0.89  0.01  0.16 |  1.36
## 23                  |  2.27 |  1.27  0.01  0.31 |  0.91  0.01  0.16 |  1.44
## 32                  |  2.54 |  1.27  0.01  0.25 |  1.00  0.02  0.15 |  1.76
## 39                  |  1.75 |  1.28  0.01  0.53 |  0.60  0.01  0.12 |  0.33
## 59                  |  2.27 |  1.27  0.01  0.31 |  0.91  0.01  0.16 |  1.44
## 75                  |  1.77 |  1.28  0.01  0.52 |  0.52  0.00  0.08 |  0.01
## 80                  |  1.76 |  1.28  0.01  0.53 |  0.63  0.01  0.13 |  0.41
## 82                  |  2.05 |  1.27  0.01  0.39 |  0.82  0.01  0.16 |  1.12
## 85                  |  1.76 |  1.28  0.01  0.53 |  0.54  0.00  0.09 |  0.09
##                        ctr   cos2    Dim.4    ctr   cos2
## 8                     0.00  0.04 | -0.82  0.01  0.22 |
## 16                    0.04  0.38 | -0.57  0.01  0.07 |
## 23                    0.04  0.40 | -0.55  0.01  0.06 |
## 32                    0.06  0.48 | -0.47  0.00  0.03 |
## 39                    0.00  0.04 | -0.82  0.01  0.22 |
## 59                    0.04  0.40 | -0.55  0.01  0.06 |
## 75                    0.00  0.00 | -0.90  0.02  0.26 |
## 80                    0.00  0.05 | -0.80  0.01  0.21 |
## 82                    0.02  0.30 | -0.63  0.01  0.09 |
## 85                    0.00  0.00 | -0.88  0.02  0.25 |
##
## Variables
##                        Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3    ctr
## age                 |  0.00  0.00  0.00 |  0.26  5.07  0.07 |  0.83 66.36
## campaign            |  0.18  0.80  0.03 | -0.06  0.28  0.00 | -0.11  1.17
## pdays               |  0.43  4.68  0.18 | -0.73 39.88  0.54 |  0.25  5.98
## previous            | -0.61  9.47  0.37 |  0.52 20.43  0.27 | -0.26  6.70
## emp.var.rate        |  0.97 23.89  0.93 |  0.17  2.08  0.03 | -0.10  0.94
## cons.price.idx      |  0.72 13.17  0.51 |  0.28  5.86  0.08 | -0.28  7.56
## cons.conf.idx       |  0.27  1.93  0.08 |  0.57 24.57  0.33 |  0.34 11.19
## euribor3m           |  0.97 24.09  0.94 |  0.15  1.77  0.02 | -0.03  0.07
## nr.employed         |  0.93 21.96  0.86 | -0.03  0.06  0.00 | -0.02  0.03
##                       cos2    Dim.4    ctr   cos2
## age                   0.69 |  0.20  4.09  0.04 |
## campaign              0.01 |  0.96 93.98  0.92 |
## pdays                 0.06 | -0.05  0.23  0.00 |
## previous              0.07 |  0.06  0.32  0.00 |
## emp.var.rate          0.01 | -0.02  0.05  0.00 |
## cons.price.idx        0.08 |  0.02  0.05  0.00 |
## cons.conf.idx         0.12 | -0.09  0.88  0.01 |
## euribor3m             0.00 | -0.05  0.26  0.00 |
## nr.employed           0.00 | -0.04  0.14  0.00 |
##
## Supplementary continuous variable
##                        Dim.1   cos2    Dim.2   cos2    Dim.3   cos2    Dim.4   cos2
## duration            | -0.03  0.00 |  0.04  0.00 | -0.03  0.00 | -0.04  0.00 |
```

```
##
## Supplementary categories (the 10 first)
##                        Dist    Dim.1   cos2 v.test    Dim.2   cos2 v.test
## Y_no               |   0.28 |   0.26   0.87  24.90 |  -0.09   0.11 -14.94 |
## Y_yes              |   2.03 |  -1.89   0.87 -24.90 |   0.66   0.11  14.94 |
## Marital_divorced   |   0.44 |   0.10   0.05   1.23 |   0.07   0.02   1.43 |
## Marital_married    |   0.24 |   0.10   0.17   4.39 |   0.08   0.11   6.16 |
## Marital_single     |   0.70 |  -0.25   0.13  -5.57 |  -0.20   0.08  -7.58 |
## Job_admin.         |   0.20 |  -0.09   0.22  -1.95 |   0.02   0.01   0.82 |
## Job_blue-collar    |   0.34 |   0.25   0.53   4.80 |  -0.15   0.18  -4.85 |
## Job_entrepreneur   |   0.37 |   0.20   0.31   1.33 |  -0.10   0.08  -1.12 |
## Job_housemaid      |   0.72 |   0.40   0.31   2.30 |   0.39   0.30   3.86 |
## Job_management     |   0.35 |  -0.09   0.07  -0.90 |   0.07   0.04   1.22 |
##                     Dim.3   cos2 v.test    Dim.4   cos2 v.test
## Y_no                 0.02   0.00   3.56 |   0.00   0.00   0.28 |
## Y_yes               -0.14   0.00  -3.56 |  -0.01   0.00  -0.28 |
## Marital_divorced     0.31   0.51   7.68 |   0.07   0.03   1.89 |
## Marital_married      0.19   0.59  16.14 |   0.04   0.03   3.68 |
## Marital_single      -0.54   0.59 -23.02 |  -0.13   0.03  -5.52 |
## Job_admin.          -0.10   0.29  -4.33 |   0.00   0.00  -0.01 |
## Job_blue-collar     -0.08   0.06  -3.16 |  -0.04   0.01  -1.54 |
## Job_entrepreneur     0.14   0.15   1.76 |  -0.01   0.00  -0.19 |
## Job_housemaid        0.35   0.24   3.95 |   0.09   0.02   1.04 |
## Job_management       0.30   0.74   5.76 |   0.06   0.03   1.11 |

dimdesc(res.pca, axes = 1:4)

## $Dim.1
## $Dim.1$quanti
##                correlation       p.value
## euribor3m       0.97012135  0.000000e+00
## emp.var.rate    0.96596055  0.000000e+00
## nr.employed     0.92622181  0.000000e+00
## cons.price.idx  0.71732355  0.000000e+00
## pdays           0.42778256 2.747395e-219
## cons.conf.idx   0.27475758  2.220057e-86
## campaign        0.17647126  6.925306e-36
## duration       -0.02789008  4.984006e-02
## previous       -0.60838071  0.000000e+00
##
## $Dim.1$quali
##                  R2       p.value
## y       0.125386566 4.727704e-146
## job     0.050547845  1.431720e-48
## marital 0.006555608  4.090296e-07
##
## $Dim.1$category
##                     Estimate       p.value
## Y_no              1.07413334 4.727704e-146
## Job_blue-collar   0.38172472  1.578463e-06
```

```
## Marital_married  0.31200050  1.111008e-05
## Job_unknown      1.09681411  1.361795e-03
## Job_technician   0.31629473  5.350108e-03
## Job_services     0.36150713  9.057754e-03
## Job_housemaid    0.53511934  2.165660e-02
## Marital_single  -0.03928771  2.479330e-08
## Job_retired     -1.00385338  2.180033e-17
## Job_student     -2.04730655  1.106798e-30
## Y_yes           -1.07413334 4.727704e-146
##
##
## $Dim.2
## $Dim.2$quanti
##                correlation      p.value
## cons.conf.idx   0.57422055 0.000000e+00
## previous        0.52363024 0.000000e+00
## cons.price.idx  0.28034870 5.339409e-90
## age             0.26095722 8.309706e-78
## emp.var.rate    0.16716817 2.513370e-32
## euribor3m       0.15421659 1.052205e-27
## duration        0.04037167 4.515730e-03
## nr.employed    -0.02841696 4.567319e-02
## campaign       -0.06123050 1.638747e-05
## pdays          -0.73167488 0.000000e+00
##
## $Dim.2$quali
##                 R2      p.value
## y       0.04515955 1.302046e-51
## job     0.02376411 3.195565e-20
## marital 0.01239773 2.598827e-13
##
## $Dim.2$category
##                       Estimate      p.value
## Y_yes               0.377861724 1.302046e-51
## Job_retired         0.538789877 6.759304e-16
## Marital_single      0.004356898 2.903848e-14
## Marital_married     0.287205205 6.730200e-10
## Job_housemaid       0.295587120 1.140921e-04
## Job_unemployed      0.164414804 1.813003e-02
## Job_self-employed -0.286908713 4.048030e-02
## Job_services      -0.237667993 5.602952e-03
## Job_blue-collar   -0.243007374 1.206646e-06
## Y_no              -0.377861724 1.302046e-51
##
##
## $Dim.3
## $Dim.3$quanti
##                correlation       p.value
## age             0.82888610  0.000000e+00
```

```
## cons.conf.idx   0.34042071 1.951278e-134
## pdays           0.24888801  1.013084e-70
## duration       -0.03380074  1.744413e-02
## emp.var.rate   -0.09882178  3.278995e-12
## campaign       -0.11007792  8.282436e-15
## previous       -0.26339058  2.783976e-79
## cons.price.idx -0.27971856  1.379085e-89
##
## $Dim.3$quali
##                  R2       p.value
## job     0.178569210 2.700620e-201
## marital 0.108636234 7.136894e-123
## y       0.002560617  3.706473e-04
##
## $Dim.3$category
##                     Estimate       p.value
## Job_retired       1.68149671 5.680668e-154
## Marital_married   0.20097034  4.022987e-60
## Marital_divorced  0.32318275  1.363916e-14
## Job_management    0.21386530  7.974613e-09
## Job_housemaid     0.26294711  7.749768e-05
## Y_no              0.07902348  3.706473e-04
## Job_unknown       0.32659224  6.931541e-03
## Job_technician   -0.19259908  2.190366e-03
## Job_blue-collar  -0.17391611  1.561623e-03
## Y_yes            -0.07902348  3.706473e-04
## Job_admin.       -0.19544362  1.450930e-05
## Job_services     -0.28666694  1.024159e-05
## Job_student      -1.31635556  2.904328e-36
## Marital_single   -0.52342349 6.555178e-124
##
##
## $Dim.4
## $Dim.4$quanti
##                correlation      p.value
## campaign        0.96002062 0.000000e+00
## age             0.20031553 6.085712e-46
## previous        0.05584528 8.510511e-05
## duration       -0.03555363 1.239946e-02
## nr.employed    -0.03657601 1.009608e-02
## pdays          -0.04772882 7.858589e-04
## euribor3m      -0.05064684 3.662674e-04
## cons.conf.idx -0.09302262 5.577676e-11
##
## $Dim.4$quali
##                  R2      p.value
## marital 0.006967409 1.511213e-07
## job     0.008422687 1.773990e-05
##
```

```
## $Dim.4$category
##                  Estimate      p.value
## Job_retired      0.3107102 3.526031e-07
## NA               0.5653440 4.474349e-02
## Job_student     -0.3084071 3.956014e-03
## Marital_married -0.1436551 2.306144e-04
## Marital_single  -0.3104477 3.333871e-08
```

Ara comentarem a partir de les comandes executades anteriorment quines variables són més explicatives segons cada dimensió:

A la dimensió 1 les variables més explicatives són les que mostren els diferents indicadors relacionats amb l'individu i l'estat de l'economia. També podem veure que la variable previous (número de cops que s'ha contactat amb el client anteriorment) és inversament proporcional.

A la dimensió 2 la variable més clarament explicativa és "cons.conf.idx", que és l'índex de confiança del consumidor.

A la dimensió 3 veiem que "age", "cons.conf.idx" i "pdays" tenen una alta contribució, les dues variables relacionades amb la confiança i amb aspectes específics d'aquest client abans de realitzar l'actual campanya.

Finalment a la dimensió 4, veiem que "campaign" i "age" són les variables més explicatives.

K-Means Classification

Ara farem un nou mètode d'agrupament, que és el clustering i ens permetrà buscar dins de les nostres observacions grups d'individus amb característiques similars.

```
# Fixed number of groups/clusters

dclu<-res.pca$ind$coord[,1:4] # Significant axes
kcla <- kmeans(dclu,7) # No less than 6 groups

#names(kcla)
#summary(kcla)
table(kcla$cluster)

##
##    1    2    3    4    5    6    7
##  312 1744  828  166 1245  376  275

kcla$totss #inercia total

## [1] 35931.36

kcla$betweenss #inercia entre grups

## [1] 28960.27
```

```
kcla$withinss #inercia intra grups

## [1]  809.6068 1638.6338  875.0929  811.3854 1164.1523  822.3318  849.8907

#Set clusters m'expliquen una mica mes d'un 80% de l'informacio, es la qualitat de la representacio
info<-kcla$betweenss/kcla$totss
info

## [1] 0.8059886
```

Sabem que no hi ha una manera del tot correcte per determinar el nombre de clusters, però sabem que no hem d'agafar menys de 6, però sabem que s'han d'agafar un mínim per a que el nombre de clusters sigui més òptim i poder veure una bona representació dels nostres clusters. Podem comprobar que amb set clusters tenim una mica més d'un 80% de qualitat en la representació de l'informació i això ho sabem amb la nostra nova variable creada "info".

Descripició dels clusters

```
nbcluster <- 7
df$CLUSTER <- nbcluster
df[names(kcla$cluster),"CLUSTER"]<-kcla$cluster

df$f.CLUSTER <- factor(df$CLUSTER, labels =
c("CLUSTER-1","CLUSTER-2","CLUSTER-3","CLUSTER-4","CLUSTER-5","CLUSTER-6","CLUSTER-7")
)

#df$kcla<-factor(kcla$cluster)
#names(df)
#catdes(df,34,prob=0.005)
#res.pca<-PCA(df[,c("duration","y","kcla",vars_con)],quanti.sup=1,quali.sup=2:3,ncp=4)
#plot.PCA(res.pca,choix="ind", habillage=3)


sel <- c(1:21)

vars_km <- names(df[sel])

vars <- c(vars_km,"f.CLUSTER")
targ <- which(vars == "f.CLUSTER")
catdes(df[,vars],targ)

##
## Link between the cluster variable and the categorical variables (chi-square test)
## =================================================================================
##                    p.value df
## month        0.000000e+00 54
## poutcome     0.000000e+00 12
```

```
## y           7.309361e-189  6
## job         2.405672e-165 66
## contact     1.740516e-145  6
## marital      4.107931e-75 18
## default      1.164839e-52  6
## education    1.497215e-20 42
## day_of_week  3.433224e-05 24
##
## Description of each cluster by the categories
## =============================================
## $`CLUSTER-1`
##                                          Cla/Mod    Mod/Cla     Global
## job=Job_retired                        44.660194 29.4871795  4.1649818
## poutcome=Poutcome_failure              21.115538 33.9743590 10.1496159
## y=Y_yes                                17.420436 33.3333333 12.0703599
## contact=Contact_cellular                8.663683 86.8589744 63.2430247
## month=Month_sep                        40.579710  8.9743590  1.3950667
## default=Default_no                      7.332652 91.9871795 79.1346543
## month=Month_dec                        50.000000  3.5256410  0.4448039
## education=Education_basic.4y           12.350598 19.8717949 10.1496159
## month=Month_oct                        24.096386  6.4102564  1.6781237
## month=Month_mar                        25.373134  5.4487179  1.3546300
## marital=Marital_married                 7.566667 72.7564103 60.6550748
## month=Month_apr                        12.455516 11.2179487  5.6813587
## day_of_week=Day_of_week_tue             9.191584 26.6025641 18.2571775
## month=Month_aug                         9.212283 22.1153846 15.1435503
## job=Job_management                      9.565217 10.5769231  6.9753336
## marital=Marital_divorced                8.718861 15.7051282 11.3627173
## education=Education_university.degree  7.402423 35.2564103 30.0444804
## day_of_week=Day_of_week_wed             4.809619 15.3846154 20.1779216
## job=Job_student                         0.952381  0.3205128  2.1229276
## day_of_week=Day_of_week_thu             4.575786 15.3846154 21.2090578
## education=Education_high.school         4.553571 16.3461538 22.6445613
## poutcome=Poutcome_success               1.190476  0.6410256  3.3966842
## job=Job_technician                      3.571429  8.9743590 15.8511929
## education=Education_basic.9y            3.425560  8.3333333 15.3457339
## month=Month_jun                         2.932099  6.0897436 13.1014962
## job=Job_services                        1.902748  2.8846154  9.5632835
## month=Month_may                         3.641975 18.9102564 32.7537404
## job=Job_blue-collar                     2.797203 10.2564103 23.1298019
## default=Default_unknown                 2.422481  8.0128205 20.8653457
## month=Month_jul                         1.791713  5.1282051 18.0549939
## marital=Marital_single                  2.614379 11.5384615 27.8406793
## contact=Contact_telephone               2.255226 13.1410256 36.7569753
## poutcome=Poutcome_nonexistent           4.770814 65.3846154 86.4537000
## y=Y_no                                  4.782709 66.6666667 87.9296401
##                                            p.value     v.test
## job=Job_retired                        2.741695e-59  16.237431
## poutcome=Poutcome_failure              7.003146e-33  11.943706
```

```
## y=Y_yes                              1.093042e-24  10.257677
## contact=Contact_cellular             8.513033e-22   9.593520
## month=Month_sep                      1.333337e-16   8.270558
## default=Default_no                   2.409876e-10   6.332649
## month=Month_dec                      2.114315e-08   5.602377
## education=Education_basic.4y         1.043361e-07   5.319005
## month=Month_oct                      1.381825e-07   5.267648
## month=Month_mar                      5.658811e-07   5.002512
## marital=Marital_married              4.005160e-06   4.611114
## month=Month_apr                      8.740549e-05   3.923131
## day_of_week=Day_of_week_tue          1.658807e-04   3.766005
## month=Month_aug                      7.213479e-04   3.381334
## job=Job_management                   1.486789e-02   2.435581
## marital=Marital_divorced             1.655982e-02   2.396338
## education=Education_university.degree 4.049047e-02   2.048710
## day_of_week=Day_of_week_wed          2.592471e-02  -2.227338
## job=Job_student                      9.145980e-03  -2.606549
## day_of_week=Day_of_week_thu          7.514420e-03  -2.673143
## education=Education_high.school      4.743381e-03  -2.823963
## poutcome=Poutcome_success            1.357037e-03  -3.203637
## job=Job_technician                   2.717358e-04  -3.640853
## education=Education_basic.9y         1.566510e-04  -3.780282
## month=Month_jun                      3.957147e-05  -4.109968
## job=Job_services                     2.723397e-06  -4.690649
## month=Month_may                      2.011190e-08  -5.611036
## job=Job_blue-collar                  1.784928e-09  -6.016260
## default=Default_unknown              2.409876e-10  -6.332649
## month=Month_jul                      4.182234e-12  -6.930882
## marital=Marital_single               7.887592e-13  -7.163095
## contact=Contact_telephone            8.513033e-22  -9.593520
## poutcome=Poutcome_nonexistent        7.915205e-23  -9.835527
## y=Y_no                               1.093042e-24 -10.257677
##
## $`CLUSTER-2`
##                                  Cla/Mod    Mod/Cla     Global
## poutcome=Poutcome_nonexistent   39.663237 97.2477064 86.4537000
## month=Month_jul                 58.230683 29.8165138 18.0549939
## contact=Contact_telephone       46.149615 48.1077982 36.7569753
## marital=Marital_single          46.550472 36.7545872 27.8406793
## y=Y_no                          37.686825 93.9793578 87.9296401
## month=Month_jun                 53.240741 19.7821101 13.1014962
## month=Month_nov                 47.859922 14.1055046 10.3922362
## job=Job_services                43.974630 11.9266055  9.5632835
## education=Education_high.school 39.821429 25.5733945 22.6445613
## day_of_week=Day_of_week_wed     39.979960 22.8784404 20.1779216
## job=Job_technician              40.051020 18.0045872 15.8511929
## education=Education_basic.9y    39.789196 17.3165138 15.3457339
## day_of_week=Day_of_week_thu     38.036225 22.8784404 21.2090578
## marital=NA                       0.000000  0.0000000  0.1415285
```

```
## month=Month_aug                  32.042724 13.7614679 15.1435503
## day_of_week=Day_of_week_mon      32.458698 19.1513761 20.8046907
## job=Job_management               27.246377  5.3899083  6.9753336
## marital=Marital_divorced         28.291815  9.1169725 11.3627173
## month=Month_dec                   0.000000  0.0000000  0.4448039
## job=Job_student                  13.333333  0.8027523  2.1229276
## education=Education_basic.4y     22.709163  6.5366972 10.1496159
## month=Month_oct                   4.819277  0.2293578  1.6781237
## marital=Marital_married          31.466667 54.1284404 60.6550748
## month=Month_mar                   0.000000  0.0000000  1.3546300
## month=Month_sep                   0.000000  0.0000000  1.3950667
## y=Y_yes                          17.587940  6.0206422 12.0703599
## month=Month_may                  24.012346 22.3050459 32.7537404
## poutcome=Poutcome_success         0.000000  0.0000000  3.3966842
## contact=Contact_cellular         28.932225 51.8922018 63.2430247
## job=Job_retired                   1.456311  0.1720183  4.1649818
## poutcome=Poutcome_failure         9.561753  2.7522936 10.1496159
## month=Month_apr                   0.000000  0.0000000  5.6813587
##                                      p.value     v.test
## poutcome=Poutcome_nonexistent   9.983079e-74  18.163820
## month=Month_jul                 2.245179e-54  15.527928
## contact=Contact_telephone       6.123716e-34  12.144659
## marital=Marital_single          1.820325e-24  10.208297
## y=Y_no                          5.068766e-24  10.108435
## month=Month_jun                 9.092724e-24  10.051014
## month=Month_nov                 5.691155e-10   6.198755
## job=Job_services                3.921572e-05   4.112052
## education=Education_high.school 3.071926e-04   3.609158
## day_of_week=Day_of_week_wed     5.232636e-04   3.468556
## job=Job_technician              2.386167e-03   3.037415
## education=Education_basic.9y    4.834718e-03   2.817845
## day_of_week=Day_of_week_thu     3.473033e-02   2.111489
## marital=NA                      4.755159e-02  -1.981354
## month=Month_aug                 4.458324e-02  -2.008565
## day_of_week=Day_of_week_mon     3.396930e-02  -2.120436
## job=Job_management              1.036131e-03  -3.280528
## marital=Marital_divorced        1.985520e-04  -3.720852
## month=Month_dec                 6.832667e-05  -3.982039
## job=Job_student                 3.773708e-07  -5.080032
## education=Education_basic.4y    1.595910e-10  -6.395913
## month=Month_oct                 3.086689e-11  -6.642375
## marital=Marital_married         4.856257e-12  -6.909716
## month=Month_mar                 1.742771e-13  -7.367178
## month=Month_sep                 7.194562e-14  -7.484271
## y=Y_yes                         5.068766e-24 -10.108435
## month=Month_may                 6.732938e-32 -11.754030
## poutcome=Poutcome_success       3.848356e-33 -11.993388
## contact=Contact_cellular        6.123716e-34 -12.144659
## job=Job_retired                 3.356116e-35 -12.379936
```

```
## poutcome=Poutcome_failure       5.197520e-44 -13.914149
## month=Month_apr                 9.136003e-56 -15.731946
##
## $`CLUSTER-3`
##                               Cla/Mod    Mod/Cla     Global       p.value
## month=Month_apr              71.5302491 24.2753623  5.6813587 9.058650e-100
## contact=Contact_cellular     23.8171355 89.9758454 63.2430247  4.766994e-80
## month=Month_may              28.7654321 56.2801932 32.7537404  3.770219e-53
## default=Default_no           18.8298416 89.0096618 79.1346543  4.585452e-16
## job=Job_student              49.5238095  6.2801932  2.1229276  5.110953e-15
## marital=Marital_single       23.3841685 38.8888889 27.8406793  3.735231e-14
## month=Month_mar              52.2388060  4.2270531  1.3546300  2.449652e-11
## y=Y_yes                      23.9530988 17.2705314 12.0703599  1.377677e-06
## job=Job_blue-collar          19.8426573 27.4154589 23.1298019  1.594152e-03
## poutcome=Poutcome_failure    21.7131474 13.1642512 10.1496159  2.243483e-03
## month=Month_oct              30.1204819  3.0193237  1.6781237  2.366941e-03
## day_of_week=Day_of_week_fri 19.3381593 22.5845411 19.5511524  1.728303e-02
## job=Job_unknown               4.6511628  0.2415459  0.8693894  2.069975e-02
## marital=Marital_divorced     13.3451957  9.0579710 11.3627173  1.957459e-02
## education=NA                 10.9523810  2.7777778  4.2458552  1.717205e-02
## job=Job_housemaid             6.3492063  0.9661836  2.5475131  5.317520e-04
## y=Y_no                       15.7507473 82.7294686 87.9296401  1.377677e-06
## job=Job_retired               4.3689320  1.0869565  4.1649818  3.200463e-08
## marital=Marital_married      14.2666667 51.6908213 60.6550748  9.627802e-09
## month=Month_jun               8.3333333  6.5217391 13.1014962  4.302173e-11
## poutcome=Poutcome_success     0.0000000  0.0000000  3.3966842  2.391195e-14
## default=Default_unknown       8.8178295 10.9903382 20.8653457  4.585452e-16
## month=Month_nov               0.9727626  0.6038647 10.3922362  7.650679e-36
## month=Month_aug               1.6021362  1.4492754 15.1435503  2.014127e-47
## month=Month_jul               2.0156775  2.1739130 18.0549939  2.454945e-53
## contact=Contact_telephone     4.5654565 10.0241546 36.7569753  4.766994e-80
##                                 v.test
## month=Month_apr              21.202484
## contact=Contact_cellular     18.945973
## month=Month_may              15.345946
## default=Default_no            8.122005
## job=Job_student               7.824151
## marital=Marital_single        7.569896
## month=Month_mar               6.676351
## y=Y_yes                       4.828207
## job=Job_blue-collar           3.156975
## poutcome=Poutcome_failure     3.055950
## month=Month_oct               3.039852
## day_of_week=Day_of_week_fri   2.380631
## job=Job_unknown              -2.313416
## marital=Marital_divorced     -2.334404
## education=NA                 -2.383003
## job=Job_housemaid            -3.464230
## y=Y_no                       -4.828207
```

```
## job=Job_retired              -5.530101
## marital=Marital_married      -5.737159
## month=Month_jun              -6.593279
## poutcome=Poutcome_success    -7.627624
## default=Default_unknown      -8.122005
## month=Month_nov             -12.498048
## month=Month_aug             -14.465066
## month=Month_jul             -15.373761
## contact=Contact_telephone   -18.945973
##
## $`CLUSTER-4`
##                                           Cla/Mod    Mod/Cla     Global
## poutcome=Poutcome_success              88.0952381 89.1566265  3.3966842
## y=Y_yes                                20.1005025 72.2891566 12.0703599
## month=Month_sep                        31.8840580 13.2530120  1.3950667
## contact=Contact_cellular                4.7953964 90.3614458 63.2430247
## month=Month_oct                        22.8915663 11.4457831  1.6781237
## job=Job_student                        16.1904762 10.2409639  2.1229276
## default=Default_no                      3.9856924 93.9759036 79.1346543
## month=Month_dec                        31.8181818  4.2168675  0.4448039
## month=Month_mar                        13.4328358  5.4216867  1.3546300
## job=Job_retired                         7.7669903  9.6385542  4.1649818
## education=Education_professional.course  5.3781513 19.2771084 12.0299232
## job=Job_admin.                          4.5349731 35.5421687 26.3040841
## education=Education_university.degree    4.3741588 39.1566265 30.0444804
## job=Job_unemployed                      8.4112150  5.4216867  2.1633643
## job=Job_self-employed                   0.6578947  0.6024096  3.0731905
## job=Job_services                        1.6913319  4.8192771  9.5632835
## education=Education_basic.6y            1.1194030  1.8072289  5.4185200
## education=Education_basic.9y            1.5810277  7.2289157 15.3457339
## month=Month_jul                         1.2318029  6.6265060 18.0549939
## job=Job_blue-collar                     1.1363636  7.8313253 23.1298019
## default=Default_unknown                 0.9689922  6.0240964 20.8653457
## month=Month_may                         0.8641975  8.4337349 32.7537404
## contact=Contact_telephone               0.8800880  9.6385542 36.7569753
## y=Y_no                                  1.0577144 27.7108434 87.9296401
## poutcome=Poutcome_nonexistent           0.0000000  0.0000000 86.4537000
##                                              p.value     v.test
## poutcome=Poutcome_success              8.703859e-239  32.997907
## y=Y_yes                                 1.563077e-76  18.514996
## month=Month_sep                         1.485214e-16   8.257688
## contact=Contact_cellular                6.857204e-16   8.073035
## month=Month_oct                         1.540670e-11   6.744017
## job=Job_student                         5.601823e-08   5.431067
## default=Default_no                      7.923004e-08   5.368873
## month=Month_dec                         5.003774e-06   4.564629
## month=Month_mar                         4.144742e-04   3.530692
## job=Job_retired                         1.826039e-03   3.117158
## education=Education_professional.course  6.245827e-03   2.734589
```

```
## job=Job_admin.                             7.620775e-03   2.668425
## education=Education_university.degree    1.092935e-02   2.544950
## job=Job_unemployed                        1.223931e-02   2.505168
## job=Job_self-employed                     3.827209e-02  -2.071929
## job=Job_services                          2.484116e-02  -2.243864
## education=Education_basic.6y              2.234725e-02  -2.284413
## education=Education_basic.9y              1.505217e-03  -3.173676
## month=Month_jul                           1.791942e-05  -4.289353
## job=Job_blue-collar                       1.542134e-07  -5.247457
## default=Default_unknown                   7.923004e-08  -5.368873
## month=Month_may                           5.514070e-14  -7.519133
## contact=Contact_telephone                 6.857204e-16  -8.073035
## y=Y_no                                    1.563077e-76 -18.514996
## poutcome=Poutcome_nonexistent            2.421178e-153 -26.378457
##
## $`CLUSTER-5`
##                                          Cla/Mod     Mod/Cla     Global
## poutcome=Poutcome_nonexistent          28.531338 97.99196787 86.4537000
## default=Default_unknown                40.794574 33.81526104 20.8653457
## marital=Marital_married                30.833333 74.29718876 60.6550748
## month=Month_aug                        42.723632 25.70281124 15.1435503
## contact=Contact_telephone              34.103410 49.79919679 36.7569753
## y=Y_no                                 27.408600 95.74297189 87.9296401
## education=Education_basic.4y           38.247012 15.42168675 10.1496159
## marital=Marital_divorced               34.875445 15.74297189 11.3627173
## month=Month_nov                        34.435798 14.21686747 10.3922362
## job=Job_management                     34.782609  9.63855422  6.9753336
## job=Job_housemaid                      39.682540  4.01606426  2.5475131
## month=Month_may                        28.024691 36.46586345 32.7537404
## job=Job_retired                        34.951456  5.78313253  4.1649818
## job=Job_unknown                        46.511628  1.60642570  0.8693894
## education=Education_university.degree 23.216689 27.71084337 30.0444804
## job=Job_services                       20.084567  7.63052209  9.5632835
## education=Education_high.school        21.785714 19.59839357 22.6445613
## job=Job_admin.                         21.983090 22.97188755 26.3040841
## month=Month_dec                         0.000000  0.00000000  0.4448039
## month=Month_jul                        20.156775 14.45783133 18.0549939
## month=Month_jun                        17.129630  8.91566265 13.1014962
## month=Month_oct                         3.614458  0.24096386  1.6781237
## month=Month_mar                         0.000000  0.00000000  1.3546300
## month=Month_sep                         0.000000  0.00000000  1.3950667
## job=Job_student                         0.952381  0.08032129  2.1229276
## poutcome=Poutcome_success               0.000000  0.00000000  3.3966842
## y=Y_yes                                 8.877722  4.25702811 12.0703599
## contact=Contact_cellular               19.980818 50.20080321 63.2430247
## default=Default_no                     21.052632 66.18473896 79.1346543
## poutcome=Poutcome_failure               4.980080  2.00803213 10.1496159
## month=Month_apr                         0.000000  0.00000000  5.6813587
## marital=Marital_single                  8.932462  9.87951807 27.8406793
```

```
##                                               p.value     v.test
## poutcome=Poutcome_nonexistent         5.574912e-57  15.908020
## default=Default_unknown               5.398614e-36  12.525739
## marital=Marital_married               3.498103e-31  11.614012
## month=Month_aug                       1.826830e-30  11.471863
## contact=Contact_telephone             1.065106e-27  10.907179
## y=Y_no                                9.633353e-27  10.705093
## education=Education_basic.4y          8.065390e-12   6.837381
## marital=Marital_divorced              4.841989e-08   5.457017
## month=Month_nov                       7.075034e-07   4.959293
## job=Job_management                    3.671864e-05   4.127213
## job=Job_housemaid                     2.952524e-04   3.619430
## month=Month_may                       1.339529e-03   3.207374
## job=Job_retired                       1.408222e-03   3.192961
## job=Job_unknown                       2.534928e-03   3.019141
## education=Education_university.degree 3.722773e-02  -2.083258
## job=Job_services                      6.367842e-03  -2.728213
## education=Education_high.school       2.729912e-03  -2.996619
## job=Job_admin.                        1.857985e-03  -3.112041
## month=Month_dec                       1.669487e-03  -3.143486
## month=Month_jul                       1.045039e-04  -3.879889
## month=Month_jun                       1.688691e-07  -5.230700
## month=Month_oct                       1.281132e-07  -5.281525
## month=Month_mar                       3.135831e-09  -5.924325
## month=Month_sep                       1.739524e-09  -6.020432
## job=Job_student                       1.572260e-12  -7.067962
## poutcome=Poutcome_success             2.613913e-22  -9.714554
## y=Y_yes                               9.633353e-27 -10.705093
## contact=Contact_cellular              1.065106e-27 -10.907179
## default=Default_no                    5.398614e-36 -12.525739
## poutcome=Poutcome_failure             3.398883e-36 -12.562395
## month=Month_apr                       2.489983e-37 -12.767508
## marital=Marital_single                6.416022e-69 -17.545698
##
## $`CLUSTER-6`
##                                 Cla/Mod     Mod/Cla    Global      p.value
## poutcome=Poutcome_nonexistent  8.793265 100.0000000 86.453700 1.600430e-25
## month=Month_jul               15.117581  35.9042553 18.054994 8.156780e-18
## contact=Contact_telephone     10.561056  51.0638298 36.756975 4.194641e-09
## y=Y_no                         8.254771  95.4787234 87.929640 2.174139e-07
## month=Month_jun               12.500000  21.5425532 13.101496 2.371417e-06
## default=Default_unknown       10.174419  27.9255319 20.865346 6.895477e-04
## loan=Loan_no                   8.059701  86.1702128 81.277800 9.349510e-03
## day_of_week=Day_of_week_thu    9.246902  25.7978723 21.209058 2.638068e-02
## job=Job_student                2.857143   0.7978723  2.122928 4.699588e-02
## marital=Marital_single         6.390704  23.4042553 27.840679 4.362357e-02
## loan=Loan_yes                  5.625000  11.9680851 16.174687 1.794441e-02
## month=Month_mar                0.000000   0.0000000  1.354630 4.822099e-03
## month=Month_sep                0.000000   0.0000000  1.395067 4.107437e-03
```

```
## month=Month_oct               0.000000   0.0000000  1.678124 1.333768e-03
## default=Default_no            6.923863  72.0744681 79.134654 6.895477e-04
## month=Month_nov               3.307393   4.5212766 10.392236 2.222579e-05
## poutcome=Poutcome_success     0.000000   0.0000000  3.396684 1.341034e-06
## y=Y_yes                       2.847571   4.5212766 12.070360 2.174139e-07
## month=Month_may               4.876543  21.0106383 32.753740 1.827153e-07
## contact=Contact_cellular      5.882353  48.9361702 63.243025 4.194641e-09
## month=Month_apr               0.000000   0.0000000  5.681359 1.135353e-10
## poutcome=Poutcome_failure     0.000000   0.0000000 10.149616 6.087260e-19
##                                 v.test
## poutcome=Poutcome_nonexistent 10.441628
## month=Month_jul               8.597364
## contact=Contact_telephone     5.876329
## y=Y_no                        5.183797
## month=Month_jun               4.718884
## default=Default_unknown       3.393702
## loan=Loan_no                  2.599002
## day_of_week=Day_of_week_thu   2.220561
## job=Job_student              -1.986337
## marital=Marital_single       -2.017690
## loan=Loan_yes                -2.366763
## month=Month_mar              -2.818684
## month=Month_sep              -2.869791
## month=Month_oct              -3.208613
## default=Default_no           -3.393702
## month=Month_nov              -4.241271
## poutcome=Poutcome_success    -4.833574
## y=Y_yes                      -5.183797
## month=Month_may              -5.216114
## contact=Contact_cellular     -5.876329
## month=Month_apr              -6.447733
## poutcome=Poutcome_failure    -8.890430
##
## $`CLUSTER-7`
##                                 Cla/Mod    Mod/Cla    Global
## poutcome=Poutcome_failure    39.0438247 71.2727273 10.149616
## contact=Contact_cellular      7.9283887 90.1818182 63.243025
## month=Month_may               9.8148148 57.8181818 32.753740
## marital=Marital_single        7.9883805 40.0000000 27.840679
## default=Default_no            6.2595810 89.0909091 79.134654
## month=Month_apr              12.0996441 12.3636364  5.681359
## job=Job_student              16.1904762  6.1818182  2.122928
## y=Y_yes                       9.2127303 20.0000000 12.070360
## month=Month_oct              14.4578313  4.3636364  1.678124
## poutcome=Poutcome_success    10.7142857  6.5454545  3.396684
## month=Month_sep              13.0434783  3.2727273  1.395067
## day_of_week=Day_of_week_fri   7.1354705 25.0909091 19.551152
## month=Month_jun               3.0864198  7.2727273 13.101496
## marital=Marital_married       4.5666667 49.8181818 60.655075
```

```
## y=Y_no                          5.0586342 80.0000000 87.929640
## job=Job_retired                 0.4854369  0.3636364  4.164982
## education=Education_basic.4y    1.9920319  3.6363636 10.149616
## default=Default_unknown         2.9069767 10.9090909 20.865346
## month=Month_nov                 1.5564202  2.9090909 10.392236
## month=Month_aug                 1.6021362  4.3636364 15.143550
## month=Month_jul                 1.4557671  4.7272727 18.054994
## contact=Contact_telephone       1.4851485  9.8181818 36.756975
## poutcome=Poutcome_nonexistent  1.4265669 22.1818182 86.453700
##                                     p.value     v.test
## poutcome=Poutcome_failure     6.338765e-145  25.634232
## contact=Contact_cellular       1.723673e-25  10.434584
## month=Month_may                1.748450e-18   8.772434
## marital=Marital_single         7.703282e-06   4.473269
## default=Default_no             8.466879e-06   4.453025
## month=Month_apr                1.392197e-05   4.345088
## job=Job_student                6.909637e-05   3.979376
## y=Y_yes                        1.089450e-04   3.869755
## month=Month_oct                2.487692e-03   3.024835
## poutcome=Poutcome_success      7.460265e-03   2.675568
## month=Month_sep                1.778430e-02   2.370079
## day_of_week=Day_of_week_fri    2.036235e-02   2.319603
## month=Month_jun                1.754083e-03  -3.128990
## marital=Marital_married        1.843993e-04  -3.739483
## y=Y_no                         1.089450e-04  -3.869755
## job=Job_retired                8.641328e-05  -3.925880
## education=Education_basic.4y   4.226314e-05  -4.094746
## default=Default_unknown        8.466879e-06  -4.453025
## month=Month_nov                1.900719e-06  -4.763703
## month=Month_aug                6.200589e-09  -5.811256
## month=Month_jul                1.745944e-11  -6.725830
## contact=Contact_telephone      1.723673e-25 -10.434584
## poutcome=Poutcome_nonexistent 7.515621e-142 -25.357039
##
##
## Link between the cluster variable and the quantitative variables
## ================================================================
##                       Eta2      P-value
## age            0.474040466 0.000000e+00
## campaign       0.558436885 0.000000e+00
## pdays          0.892215906 0.000000e+00
## previous       0.560755628 0.000000e+00
## emp.var.rate   0.894046500 0.000000e+00
## cons.price.idx 0.453861592 0.000000e+00
## cons.conf.idx  0.352386993 0.000000e+00
## euribor3m      0.973955527 0.000000e+00
## nr.employed    0.869891520 0.000000e+00
## duration       0.006155359 3.146859e-05
##
```

```
## Description of each cluster by quantitative variables
## =======================================================
## $`CLUSTER-1`
##                    v.test Mean in category Overall mean sd in category
## age             24.992876       54.1040262   40.0525729     12.9633587
## cons.conf.idx   12.408521      -37.6166907  -40.6182329      6.8636111
## previous         8.392611        0.3942308    0.1708451      0.5787631
## pdays            3.182317       15.9807692   15.6263647      0.1951710
## campaign        -5.053737        1.8397436    2.3891187      1.2785047
## cons.price.idx -24.309345       92.8144647   93.5857345      0.5526930
## euribor3m      -28.297286        0.9678942    3.6487535      0.2725778
## nr.employed    -28.712276     5053.1480769 5167.8073595     40.2045371
## emp.var.rate   -30.784863       -2.5365385    0.1073999      0.7128929
##                Overall sd       p.value
## age            10.2585844 7.307003e-138
## cons.conf.idx   4.4137411  2.349529e-35
## previous        0.4856692  4.754639e-17
## pdays           2.0320681  1.461020e-03
## campaign        1.9835304  4.332492e-07
## cons.price.idx  0.5789159 1.561416e-130
## euribor3m       1.7286683 3.732084e-176
## nr.employed    72.8658491 2.680973e-181
## emp.var.rate    1.5670994 4.178487e-208
##
## $`CLUSTER-2`
##                    v.test Mean in category Overall mean sd in category
## euribor3m       34.964558       4.81339966    3.6487535      0.2864047
## nr.employed     33.750774    5215.19466743 5167.8073595     17.0298403
## emp.var.rate    33.469955       1.11806193    0.1073999      0.5129521
## cons.price.idx  26.054971      93.87637787   93.5857345      0.4030072
## pdays            9.542349      16.00000000   15.6263647      0.0000000
## cons.conf.idx    6.040870     -40.10447248  -40.6182329      2.8899983
## campaign       -13.047872       1.89042626    2.3891187      1.0282696
## previous       -15.315055       0.02752294    0.1708451      0.1636014
## age            -31.388217      33.84805046   40.0525729      5.1452934
##                Overall sd       p.value
## euribor3m       1.7286683 7.781109e-268
## nr.employed    72.8658491 1.041496e-249
## emp.var.rate    1.5670994 1.319276e-245
## cons.price.idx  0.5789159 1.181688e-149
## pdays           2.0320681  1.396325e-21
## cons.conf.idx   4.4137411  1.532856e-09
## campaign        1.9835304  6.534699e-39
## previous        0.4856692  6.066061e-53
## age            10.2585844 2.930161e-216
##
## $`CLUSTER-3`
##                    v.test Mean in category Overall mean sd in category
## pdays            5.797820       16.0000000   15.6263647      0.0000000
```

```
## previous        -2.545245        0.1316425    0.1708451      0.3381017
## campaign        -9.699313        1.7789855    2.3891187      1.1102043
## age            -13.699771       35.5955424   40.0525729      7.7075184
## cons.price.idx -31.605862       93.0054674   93.5857345      0.3555162
## nr.employed    -34.897408     5087.1652174 5167.8073595     31.8350959
## cons.conf.idx  -35.509624      -45.5887066  -40.6182329      3.1682232
## emp.var.rate   -40.362573       -1.8985507    0.1073999      0.3905253
## euribor3m      -43.244710        1.2779831    3.6487535      0.1943923
##                Overall sd       p.value
## pdays           2.0320681  6.718246e-09
## previous        0.4856692  1.092011e-02
## campaign        1.9835304  3.035350e-22
## age            10.2585844  1.018447e-42
## cons.price.idx  0.5789159 3.067183e-219
## nr.employed    72.8658491 8.138904e-267
## cons.conf.idx   4.4137411 3.491769e-276
## emp.var.rate    1.5670994  0.000000e+00
## euribor3m       1.7286683  0.000000e+00
##
## $`CLUSTER-4`
##                    v.test Mean in category Overall mean sd in category
## previous        42.528318        1.7469880    0.1708451      0.9228475
## cons.conf.idx    7.824800      -37.9827704  -40.6182329      6.0515896
## duration         4.547735      351.6385542  262.7672867    274.7841904
## campaign        -4.202564        1.7530120    2.3891187      1.0553178
## cons.price.idx  -5.128640       93.3591687   93.5857345      0.8261510
## emp.var.rate   -18.831759       -2.1445783    0.1073999      0.8798621
## euribor3m      -20.520883        0.9417771    3.6487535      0.5259618
## nr.employed    -26.293197     5021.6084337 5167.8073595     49.4738746
## pdays          -66.391579        5.3313253   15.6263647      3.3588376
##                 Overall sd       p.value
## previous         0.4856692  0.000000e+00
## cons.conf.idx    4.4137411  5.084663e-15
## duration       256.0881160  5.422624e-06
## campaign         1.9835304  2.639083e-05
## cons.price.idx   0.5789159  2.918428e-07
## emp.var.rate     1.5670994  4.147513e-79
## euribor3m        1.7286683  1.401424e-93
## nr.employed     72.8658491 2.293978e-152
## pdays            2.0320681  0.000000e+00
##
## $`CLUSTER-5`
##                    v.test Mean in category Overall mean sd in category
## age             34.592509      48.75341365   40.0525729      6.0606902
## euribor3m       27.183291       4.80089398    3.6487535      0.2850300
## emp.var.rate    25.150085       1.07373494    0.1073999      0.5016688
## nr.employed     23.561540    5209.90128514 5167.8073595     17.9321967
## cons.conf.idx   19.380080     -38.52096386  -40.6182329      2.8913196
## cons.price.idx  13.001943      93.77028514   93.5857345      0.3715335
```

```
## pdays            7.499251      16.00000000   15.6263647      0.0000000
## duration        -2.122793     249.43855422  262.7672867    242.1298277
## campaign        -8.964372       1.95315496    2.3891187      1.0831782
## previous       -12.660989       0.02008032    0.1708451      0.1402751
##                Overall sd       p.value
## age            10.2585844 3.274542e-262
## euribor3m       1.7286683 1.023658e-162
## emp.var.rate    1.5670994 1.410167e-139
## nr.employed    72.8658491 9.560810e-123
## cons.conf.idx   4.4137411  1.136703e-83
## cons.price.idx  0.5789159  1.192733e-38
## pdays           2.0320681  6.418366e-14
## duration      256.0881160  3.377118e-02
## campaign        1.9835304  3.120553e-19
## previous        0.4856692  9.726559e-37
##
## $`CLUSTER-6`
##                   v.test Mean in category Overall mean sd in category
## campaign       50.728690         7.377660    2.3891187      2.2493334
## emp.var.rate   14.853947         1.261436    0.1073999      0.3583756
## euribor3m      14.426375         4.885128    3.6487535      0.2632639
## nr.employed    14.060411      5218.600266 5167.8073595     16.9326584
## cons.price.idx 12.194694        93.935734   93.5857345      0.3601197
## pdays           3.708759        16.000000   15.6263647      0.0000000
## cons.conf.idx   2.580999       -40.053457  -40.6182329      2.9787704
## age             2.265504        41.204787   40.0525729      8.8773006
## previous       -7.095467         0.000000    0.1708451      0.0000000
##                Overall sd      p.value
## campaign        1.9835304 0.000000e+00
## emp.var.rate    1.5670994 6.559004e-50
## euribor3m       1.7286683 3.531615e-47
## nr.employed    72.8658491 6.649921e-45
## cons.price.idx  0.5789159 3.317468e-34
## pdays           2.0320681 2.082779e-04
## cons.conf.idx   4.4137411 9.851477e-03
## age            10.2585844 2.348177e-02
## previous        0.4856692 1.289158e-12
##
## $`CLUSTER-7`
##                   v.test Mean in category Overall mean sd in category
## previous        25.936115        0.9090909    0.1708451      0.6052115
## campaign         9.978225        3.5490909    2.3891187      2.4643548
## duration        -2.048501      232.0218182  262.7672867    238.7423097
## age             -8.119420       35.1709091   40.0525729      7.8079138
## cons.price.idx -11.740027       93.1874073   93.5857345      0.5662071
## cons.conf.idx  -13.746299      -44.1741210  -40.6182329      4.2645317
## emp.var.rate   -21.377494       -1.8560000    0.1073999      0.4394112
## nr.employed    -23.275177     5068.4105455 5167.8073595     48.7286468
## euribor3m      -24.465933        1.1700255    3.6487535      0.2279239
```

```
##                  Overall sd       p.value
## previous          0.4856692 2.608160e-148
## campaign          1.9835304  1.898329e-23
## duration        256.0881160  4.051090e-02
## age              10.2585844  4.684180e-16
## cons.price.idx    0.5789159  7.946097e-32
## cons.conf.idx     4.4137411  5.360204e-43
## emp.var.rate      1.5670994 2.164450e-101
## nr.employed      72.8658491 7.911039e-120
## euribor3m         1.7286683 3.406047e-132
```

Ara procedirem a l'explicació de cada cluster:

Cluster 1: En aquest cluster veiem que el nombre de cops que s'ha contactat anteriorment és superior a la mitjana i també es pot observar que es caracteritza perquè s'ha contactat durant els mesos d'hivern, sobretot desembre.

Cluster 2: En aquest segon cluster veiem que no hi ha hagut cap mena de campanya de marqueting anteriorment i que principalment es caracteritza pels mesos d'estiu, ja que són els que tenen un v.test major, també podem dir que destaquen els individus que estan solters.

Cluster 3: Aquest cluster es caracteritza perquè s'ha contactat durant els mesos de la primavera (abril, maig) a la majoria d'individus i les persones d'aquest cluster són la majoria estudiants.

Cluster 4: Aquest cluster es caracteritza perquè s'ha contactat durant els mesos de setembre i octubre a la majoria d'individus i veiem que hi ha hagut una campanya de marqueting exitosa anteriorment.

Cluster 5: Aquest cluster es caracteritza perquè s'ha contactat durant el mes d'agost principalment i la major part estan casats i a molts els han contactat a traves del mòbil.

Cluster 6: Aquest cluster es caracteritza per un tipus d'individu el qual s'ha contactat a traves del mòbil i el nombre de contactes realitzats durant aquesta campanya i per a aquest individus és superior a la mitjana.

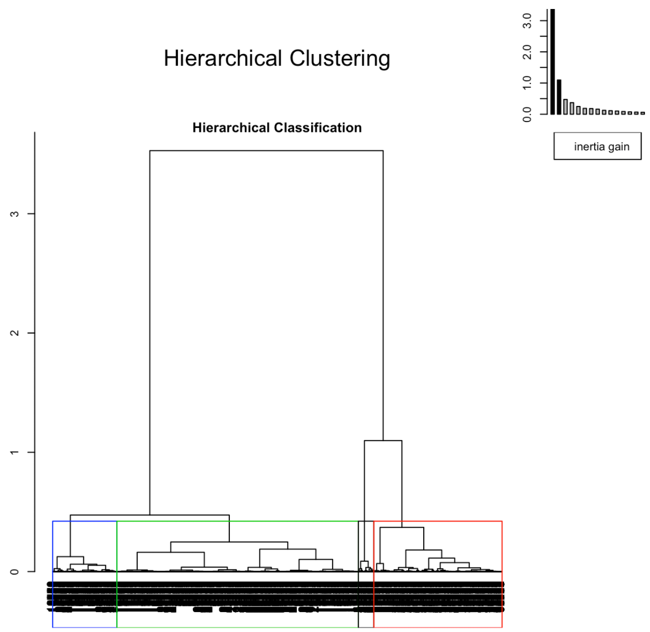
Cluster 7: En aquest cluster veiem que el nombre de cops que s'ha contactat anteriorment és superior a la mitjana i la majoria d'aquest individus estan solters.

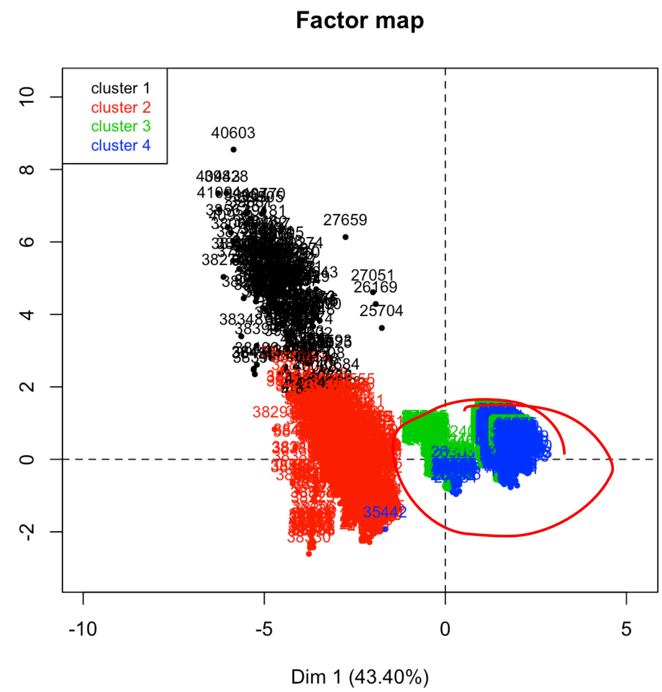
Hierarchical Clustering

Ara el que farem serà aplicar la classificació jeràrquica de clustering.

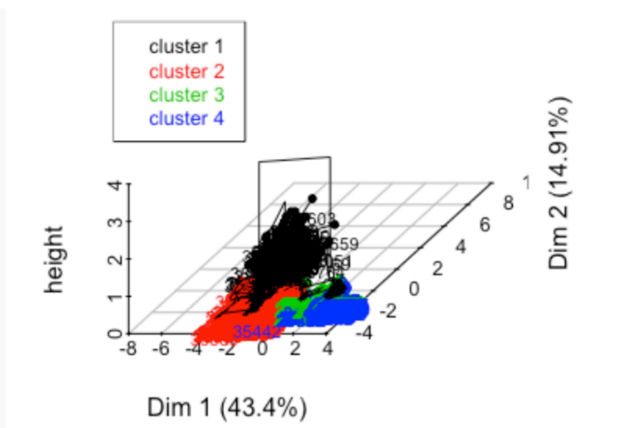
Seguidament executem una comanda específica per poder veure quin és el nombre de clusters més adequat, ja que així podrem veure un gràfic on podrem seleccionar com volem agrupar els clusters.

```
res.hcpc <- HCPC(res.pca, nb.clust = 4, order =TRUE) #Hay que cortar en un punto que no haya muchos saltos apartir de ahi cerca del cero, a primera vista podemos ver que, deberiamos ver grupos uniformes, pero no salen limpias las particiones
```





Hierarchical clustering on the factor map




```r
attributes(res.hcpc) #Tiene esas listas

## $names
## [1] "data.clust" "desc.var"   "desc.axes"  "call"       "desc.ind"
##
## $class
## [1] "HCPC"
```

```r
summary(res.hcpc$data.clust) #Nos dice el tamaño de cada cluster
```

```
##     duration          y                    marital
##  Min.   :   1.0   Y_no :4349   Marital_divorced: 562
##  1st Qu.: 104.0   Y_yes: 597   Marital_married :3000
##  Median : 182.0                Marital_single  :1377
##  Mean   : 262.8                NA's            :   7
##  3rd Qu.: 329.0
##  Max.   :2122.0
##
##               job             age            campaign          pdays
##  Job_admin.     :1301   Min.   :17.00   Min.   : 1.000   Min.   : 1.00
##  Job_blue-collar:1144   1st Qu.:32.00   1st Qu.: 1.000   1st Qu.:16.00
##  Job_technician : 784   Median :38.00   Median : 2.000   Median :16.00
##  Job_services   : 473   Mean   :40.05   Mean   : 2.389   Mean   :15.63
##  Job_management : 345   3rd Qu.:47.00   3rd Qu.: 3.000   3rd Qu.:16.00
##  Job_retired    : 206   Max.   :81.00   Max.   :14.000   Max.   :16.00
##  (Other)        : 693
##     previous       emp.var.rate     cons.price.idx  cons.conf.idx
##  Min.   :0.0000   Min.   :-3.4000   Min.   :92.20   Min.   :-50.80
##  1st Qu.:0.0000   1st Qu.:-1.8000   1st Qu.:93.08   1st Qu.:-42.70
##  Median :0.0000   Median : 1.1000   Median :93.92   Median :-41.80
##  Mean   :0.1708   Mean   : 0.1074   Mean   :93.59   Mean   :-40.62
##  3rd Qu.:0.0000   3rd Qu.: 1.4000   3rd Qu.:93.99   3rd Qu.:-36.40
##  Max.   :5.0000   Max.   : 1.4000   Max.   :94.77   Max.   :-29.80
##
##    euribor3m      nr.employed   clust
##  Min.   :0.634   Min.   :4964   1: 180
##  1st Qu.:1.344   1st Qu.:5099   2:1401
##  Median :4.857   Median :5191   3:2713
##  Mean   :3.649   Mean   :5168   4: 652
##  3rd Qu.:4.961   3rd Qu.:5228
##  Max.   :5.045   Max.   :5228
##
```

```r
# Factors globally related to clustering partition
res.hcpc$desc.var$test.chi2
```

```
##               p.value df
## y       5.654668e-177  3
## job      6.528644e-45 33
## marital  6.394260e-06  9
```

```r
# Numeric variables globally related to clustering partition
res.hcpc$desc.var$quanti.var
```

```
##                       Eta2       P-value
## campaign       0.523277769  0.000000e+00
## pdays          0.844684307  0.000000e+00
## previous       0.483134285  0.000000e+00
```

```
## emp.var.rate   0.886188857  0.000000e+00
## cons.price.idx 0.443420176  0.000000e+00
## euribor3m      0.972728324  0.000000e+00
## nr.employed    0.862075267  0.000000e+00
## cons.conf.idx  0.178928568 6.471519e-211
## duration       0.004360341  7.907623e-05
## age            0.001841458  2.786614e-02
```

```
res.hcpc$desc.var$quanti
```

```
## $`1`
##                    v.test Mean in category Overall mean sd in category
## previous        44.747785        1.7611111    0.1708451      0.9089914
## cons.conf.idx    7.085967      -38.3296660  -40.6182329      6.1337026
## duration         4.545145      347.9388889  262.7672867    273.7414263
## cons.price.idx  -3.937550       93.4189333   93.5857345      0.8322883
## campaign        -4.173711        1.7833333    2.3891187      1.1891874
## emp.var.rate   -19.056038       -2.0777778    0.1073999      0.8795552
## euribor3m      -21.375733        0.9448556    3.6487535      0.5073431
## nr.employed    -27.925157     5018.9133333 5167.8073595     50.1367856
## pdays          -64.626980        6.0166667   15.6263647      4.0599329
##                 Overall sd       p.value
## previous         0.4856692  0.000000e+00
## cons.conf.idx    4.4137411  1.380767e-12
## duration       256.0881160  5.489737e-06
## cons.price.idx   0.5789159  8.231785e-05
## campaign         1.9835304  2.996782e-05
## emp.var.rate     1.5670994  5.854424e-81
## euribor3m        1.7286683 2.247650e-101
## nr.employed     72.8658491 1.320822e-171
## pdays            2.0320681  0.000000e+00
##
## $`2`
##                    v.test Mean in category Overall mean sd in category
## previous        14.010599         0.324768    0.1708451      0.5230112
## pdays            7.336442        15.963597   15.6263647      0.3826571
## age             -2.098021        39.565714   40.0525729     11.9152285
## campaign        -5.633923         2.136331    2.3891187      1.6501597
## cons.conf.idx  -29.535606       -43.567123  -40.6182329      5.4810779
## cons.price.idx -45.687783        92.987431   93.5857345      0.4524177
## nr.employed    -55.147107      5076.909707 5167.8073595     39.0431169
## emp.var.rate   -60.532277        -2.038401    0.1073999      0.5550071
## euribor3m      -62.859907         1.190700    3.6487535      0.2529861
##                Overall sd       p.value
## previous        0.4856692  1.342670e-44
## pdays           2.0320681  2.193465e-13
## age            10.2585844  3.590331e-02
## campaign        1.9835304  1.761553e-08
## cons.conf.idx   4.4137411 1.005224e-191
## cons.price.idx  0.5789159  0.000000e+00
```

```
## nr.employed    72.8658491  0.000000e+00
## emp.var.rate    1.5670994  0.000000e+00
## euribor3m       1.7286683  0.000000e+00
##
## $`3`
##                   v.test Mean in category Overall mean sd in category
## euribor3m       51.50287       4.79738150    3.6487535      0.2961639
## emp.var.rate    48.14380       1.08075931    0.1073999      0.5264601
## nr.employed     47.78941    5212.73276815 5167.8073595     17.5946210
## cons.price.idx  32.15316      93.82588058   93.5857345      0.3977156
## cons.conf.idx   19.13856     -39.52841872  -40.6182329      2.9848752
## pdays           14.25192      16.00000000   15.6263647      0.0000000
## previous       -22.97193       0.02690748    0.1708451      0.1618131
## campaign       -27.58456       1.68322202    2.3891187      0.7827272
##                Overall sd       p.value
## euribor3m       1.7286683  0.000000e+00
## emp.var.rate    1.5670994  0.000000e+00
## nr.employed    72.8658491  0.000000e+00
## cons.price.idx  0.5789159 7.978769e-227
## cons.conf.idx   4.4137411  1.205556e-81
## pdays           2.0320681  4.361442e-46
## previous        0.4856692 8.897223e-117
## campaign        1.9835304 1.704849e-167
##
## $`4`
##                   v.test Mean in category Overall mean sd in category
## campaign       50.391249         6.036810    2.3891187      2.3318009
## emp.var.rate   20.351770         1.271319    0.1073999      0.3024787
## euribor3m      19.794810         4.897537    3.6487535      0.2113150
## nr.employed    18.610074      5217.294939 5167.8073595     17.2495005
## cons.price.idx 15.733815        93.918144   93.5857345      0.3546711
## cons.conf.idx   7.263154       -39.448313  -40.6182329      3.0538591
## pdays           5.038317        16.000000   15.6263647      0.0000000
## previous       -9.639132         0.000000    0.1708451      0.0000000
##                Overall sd      p.value
## campaign        1.9835304 0.000000e+00
## emp.var.rate    1.5670994 4.478120e-92
## euribor3m       1.7286683 3.299949e-87
## nr.employed    72.8658491 2.662485e-77
## cons.price.idx  0.5789159 8.870154e-56
## cons.conf.idx   4.4137411 3.781682e-13
## pdays           2.0320681 4.696423e-07
## previous        0.4856692 5.464823e-22
```

Amb la comanda del "chi2" podem observar que les variables "y", "job" i "marital" són les que més caracteritzen la partició en els quatre clusters que utilitzarem en el nostre anàlisi i també es podria fer amb 5 clusters, pero com no canviava molt hem vist mes convenient agafar o fer la partició en 4 clusters pel nostre estudi.

Descripció dels clusters

```
# Categories over/under represented in each cluster
res.hcpc$desc.var$category

## $`1`
##                           Cla/Mod    Mod/Cla    Global      p.value
## y=Y_yes                20.9380235 69.4444444 12.070360 2.066147e-76
## job=Job_student        17.1428571 10.0000000  2.122928 3.138518e-08
## job=Job_retired         9.2233010 10.5555556  4.164982 1.951846e-04
## job=Job_admin.          4.9961568 36.1111111 26.304084 3.190110e-03
## job=Job_unemployed      9.3457944  5.5555556  2.163364 6.838252e-03
## job=Job_self-employed   0.6578947  0.5555556  3.073190 2.588876e-02
## job=Job_services        1.6913319  4.4444444  9.563283 1.061796e-02
## job=Job_blue-collar     1.1363636  7.2222222 23.129802 9.936978e-09
## y=Y_no                  1.2646585 30.5555556 87.929640 2.066147e-76
##                            v.test
## y=Y_yes                 18.499963
## job=Job_student          5.533529
## job=Job_retired          3.725169
## job=Job_admin.           2.948799
## job=Job_unemployed       2.704620
## job=Job_self-employed   -2.227876
## job=Job_services        -2.555027
## job=Job_blue-collar     -5.731801
## y=Y_no                 -18.499963
##
## $`2`
##                           Cla/Mod    Mod/Cla     Global      p.value
## y=Y_yes                  49.74874 21.1991435 12.0703599 2.352298e-32
## job=Job_student          65.71429  4.9250535  2.1229276 1.147770e-15
## job=Job_retired          48.05825  7.0663812  4.1649818 9.670365e-10
## marital=Marital_single   33.69644 33.1192006 27.8406793 2.569686e-07
## job=Job_unknown          11.62791  0.3568879  0.8693894 1.005915e-02
## job=Job_housemaid        17.46032  1.5703069  2.5475131 4.501339e-03
## job=Job_technician       23.85204 13.3476089 15.8511929 2.166634e-03
## marital=Marital_married  26.13333 55.9600286 60.6550748 2.309388e-05
## y=Y_no                   25.38515 78.8008565 87.9296401 2.352298e-32
##                             v.test
## y=Y_yes                  11.842536
## job=Job_student           8.009926
## job=Job_retired           6.114758
## marital=Marital_single    5.152550
## job=Job_unknown          -2.573790
## job=Job_housemaid        -2.840709
## job=Job_technician       -3.066386
## marital=Marital_married  -4.232665
## y=Y_no                  -11.842536
```

```
##
## $`3`
##                           Cla/Mod     Mod/Cla     Global      p.value
## y=Y_no                   59.14003 94.80280133 87.9296401 1.640791e-61
## marital=Marital_married 56.56667 62.55068190 60.6550748 2.650603e-03
## job=Job_entrepreneur    64.37500  3.79653520  3.2349373 1.346392e-02
## job=Job_services        59.83087 10.43125691  9.5632835 2.190374e-02
## job=Job_blue-collar     57.69231 24.32731294 23.1298019 2.760138e-02
## job=Job_technician      58.29082 16.84482123 15.8511929 3.477799e-02
## job=Job_unknown         69.76744  1.10578695  0.8693894 4.818052e-02
## marital=NA              14.28571  0.03685957  0.1415285 4.004320e-02
## marital=Marital_single  50.10893 25.43309989 27.8406793 3.229271e-05
## job=Job_retired         32.52427  2.46959086  4.1649818 4.817695e-11
## job=Job_student         12.38095  0.47917435  2.1229276 5.264065e-20
## y=Y_yes                 23.61809  5.19719867 12.0703599 1.640791e-61
##                            v.test
## y=Y_no                  16.548523
## marital=Marital_married   3.005597
## job=Job_entrepreneur      2.471257
## job=Job_services          2.292033
## job=Job_blue-collar       2.202906
## job=Job_technician        2.110934
## job=Job_unknown           1.975773
## marital=NA               -2.053303
## marital=Marital_single   -4.156665
## job=Job_retired          -6.576463
## job=Job_student          -9.158465
## y=Y_yes                 -16.548523
##
## $`4`
##                   Cla/Mod    Mod/Cla    Global      p.value    v.test
## y=Y_no          14.210163 94.7852761 87.929640 3.150217e-10  6.291200
## job=Job_student  4.761905  0.7668712  2.122928 4.802179e-03 -2.820012
## y=Y_yes          5.695142  5.2147239 12.070360 3.150217e-10 -6.291200
```

Cluster 1: Els individus que pertanyen al cluster número 1 es detaquen perquè tenen la variable “y = yes”, per tant, això vol dir que són individus que SI que contracten el producte i a més també podem observar que la majoria d’aquests individus son estudiants.

Cluster 2: Els individus que pertanyen al cluster número 2 es detaquen perquè tenen la variable “y = yes”, per tant, això vol dir que són individus que SI que contracten el producte i a més també podem observar que la majoria d’aquests individus són estudiants i estan solters.

Cluster 3: Els individus que pertanyen al cluster número 3 es detaquen perquè tenen la variable “y = no”, per tant, això vol dir que són individus que NO contracten el producte i a més tambeé podem observar que la majoria d’aquests individus treballen com empresaris o en el sector de serveis i que estan casats.

Cluster 4: Els individus que pertanyen al cluster número 4 es detaquen perquè tenen la variable “y = no”, per tant, això vol dir que són individus que NO contracten el producte i a més també podem observar que la majoria d’aquests individus són estudiants.

```
### The description of the clusters by the individuals
names(res.hcpc$desc.ind)

## [1] "para" "dist"

res.hcpc$desc.ind$para  #Close to center of gravity

## Cluster: 1
##     36910     40420     40457     40031     39208
## 0.8996255 0.9520736 1.0182792 1.0842884 1.1687768
## --------------------------------------------------------
## Cluster: 2
##     34135     31328     31002     32850     32962
## 0.7368927 0.7400291 0.7406566 0.7427179 0.7427179
## --------------------------------------------------------
## Cluster: 3
##     24034      4467      4473       726      5358
## 0.6391974 0.6502367 0.6502367 0.6503246 0.6503246
## --------------------------------------------------------
## Cluster: 4
##      5296      7006      3322      6693      1049
## 0.6445766 0.6572942 0.6627406 0.6627473 0.6627498

res.hcpc$desc.ind$dist

## Cluster: 1
##    41004    40603    40930    40443    39828
## 11.14194 10.75528 10.61921 10.42103 10.07574
## --------------------------------------------------------
## Cluster: 2
##    37819    38061    38985    38677    38583
## 6.455196 6.447230 6.406478 6.351079 6.344856
## --------------------------------------------------------
## Cluster: 3
##    18895    23309    22214    14894    19305
## 3.303387 3.303373 3.303371 3.265879 3.249192
## --------------------------------------------------------
## Cluster: 4
##    18491    11713    11630    23559    35442
## 6.349686 6.335066 6.315248 6.301241 6.048853

# NO ES NECESSARI!

#### Characteristic individuals
para1<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[1]]))
para2<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[2]]))
para3<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[3]]))
para4<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[4]]))
# to be completed... as many as cluster you choose

dist1<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[1]]))
```

```
dist2<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[2]]))
dist3<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[3]]))
dist4<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[4]]))
```

Correspondence Analysis (CA)

En la part final del nostre estudi el que farem serà l'anàlisi de correspondències simples (CA) per poder analitzar les relacions entre 2 factors de les dades de la nostra mostra.

Per fer l'anàlisi de correspondencies simples utilitzarem com a target el factor discretitzat "factor_duration" i realitzarem dues taules de contingència per comparar aquest target amb 2 variables qualitatives més. Aquestes dues variables serán "job" i "factor_age".

Factor_age i Factor_duration

```
# Contingency tables - Complex : solo cuentan con los target discretizados
names(df)

# Target factor_duration vs job
# Podemos elegir la variable que queramos con la de f_duration y en este caso hemos
elegido job para este ejemplo

table(df$factor_age, df$factor_duration)

##
##                        factor_duration-[1,68] factor_duration-(68,104]
##   factor_age [17,31]                      129                      127
##   factor_age (31,36]                      155                      137
##   factor_age (36,41]                      104                      112
##   factor_age (41,49]                      119                      108
##   factor_age (49,81]                      122                      139
##
##                        factor_duration-(104,139] factor_duration-(139,182]
##   factor_age [17,31]                         143                       140
##   factor_age (31,36]                         125                       123
##   factor_age (36,41]                         101                       105
##   factor_age (41,49]                         124                       117
##   factor_age (49,81]                         119                       135
##
##                        factor_duration-(182,236] factor_duration-(236,329]
##   factor_age [17,31]                         135                       135
##   factor_age (31,36]                         126                       139
##   factor_age (36,41]                         101                       110
##   factor_age (41,49]                         126                       119
##   factor_age (49,81]                         120                       116
##
##                        factor_duration-(329,504]
##   factor_age [17,31]                         148
##   factor_age (31,36]                         127
```

```
##   factor_age (36,41]                         114
##   factor_age (41,49]                         110
##   factor_age (49,81]                         119
##
##                       factor_duration-(504,2.12e+03]
##   factor_age [17,31]                              156
##   factor_age (31,36]                              130
##   factor_age (36,41]                               83
##   factor_age (41,49]                              130
##   factor_age (49,81]                              118
```

```
#Le digo que calcule unas probabilidades en la dimension 1, calculo los perfiles por fila que tenemos
#Calculo los perfiles de fila y la suma tendria que dar mas o menos 1 y tenemos que ver si es equivalente al perfil marginal fila
prop.table(table(df$factor_age, df$factor_duration), 1) # Por filas
```

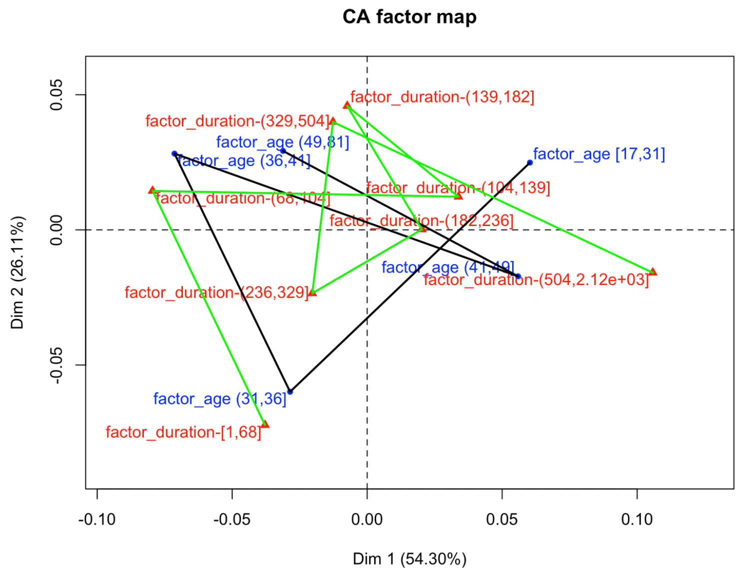
```
##
##                       factor_duration-[1,68] factor_duration-(68,104]
##   factor_age [17,31]              0.1159030                0.1141060
##   factor_age (31,36]              0.1459510                0.1290019
##   factor_age (36,41]              0.1253012                0.1349398
##   factor_age (41,49]              0.1248688                0.1133263
##   factor_age (49,81]              0.1234818                0.1406883
##
##                       factor_duration-(104,139] factor_duration-(139,182]
##   factor_age [17,31]                 0.1284816                 0.1257862
##   factor_age (31,36]                 0.1177024                 0.1158192
##   factor_age (36,41]                 0.1216867                 0.1265060
##   factor_age (41,49]                 0.1301154                 0.1227702
##   factor_age (49,81]                 0.1204453                 0.1366397
##
##                       factor_duration-(182,236] factor_duration-(236,329]
##   factor_age [17,31]                 0.1212938                 0.1212938
##   factor_age (31,36]                 0.1186441                 0.1308851
##   factor_age (36,41]                 0.1216867                 0.1325301
##   factor_age (41,49]                 0.1322141                 0.1248688
##   factor_age (49,81]                 0.1214575                 0.1174089
##
##                       factor_duration-(329,504]
##   factor_age [17,31]                 0.1329739
##   factor_age (31,36]                 0.1195857
##   factor_age (36,41]                 0.1373494
##   factor_age (41,49]                 0.1154250
##   factor_age (49,81]                 0.1204453
##
##                       factor_duration-(504,2.12e+03]
##   factor_age [17,31]                      0.1401617
##   factor_age (31,36]                      0.1224105
##   factor_age (36,41]                      0.1000000
```

```
##   factor_age (41,49]                        0.1364113
##   factor_age (49,81]                        0.1194332

#Marginal row profile
prop.table(table(df$factor_duration))

##
##        factor_duration-[1,68]       factor_duration-(68,104]
##                     0.1271735                      0.1259604
##      factor_duration-(104,139]      factor_duration-(139,182]
##                     0.1237364                      0.1253538
##      factor_duration-(182,236]      factor_duration-(236,329]
##                     0.1229276                      0.1251516
##      factor_duration-(329,504] factor_duration-(504,2.12e+03]
##                     0.1249495                      0.1247473

#Esta proporcion se mantiene en cualquiera de los colectivos mirados anteriormente? Se
tiene que hacer la comparacion

#Podemos comprobar ahora los perfiles columna
#Column profile
prop.table(table(df$factor_age, df$factor_duration), 2) # dim 2

##
##                        factor_duration-[1,68] factor_duration-(68,104]
##   factor_age [17,31]              0.2050874                0.2038523
##   factor_age (31,36]              0.2464229                0.2199037
##   factor_age (36,41]              0.1653418                0.1797753
##   factor_age (41,49]              0.1891892                0.1733547
##   factor_age (49,81]              0.1939587                0.2231140
##
##                        factor_duration-(104,139] factor_duration-(139,182]
##   factor_age [17,31]                 0.2336601                 0.2258065
##   factor_age (31,36]                 0.2042484                 0.1983871
##   factor_age (36,41]                 0.1650327                 0.1693548
##   factor_age (41,49]                 0.2026144                 0.1887097
##   factor_age (49,81]                 0.1944444                 0.2177419
##
##                        factor_duration-(182,236] factor_duration-(236,329]
##   factor_age [17,31]                 0.2220395                 0.2180937
##   factor_age (31,36]                 0.2072368                 0.2245557
##   factor_age (36,41]                 0.1661184                 0.1777060
##   factor_age (41,49]                 0.2072368                 0.1922456
##   factor_age (49,81]                 0.1973684                 0.1873990
##
##                        factor_duration-(329,504]
##   factor_age [17,31]                 0.2394822
##   factor_age (31,36]                 0.2055016
##   factor_age (36,41]                 0.1844660
##   factor_age (41,49]                 0.1779935
```

```
##   factor_age (49,81]                 0.1925566
##
##                       factor_duration-(504,2.12e+03]
##   factor_age [17,31]                          0.2528363
##   factor_age (31,36]                          0.2106969
##   factor_age (36,41]                          0.1345219
##   factor_age (41,49]                          0.2106969
##   factor_age (49,81]                          0.1912480

#Marginal colum profile
prop.table(table(df$factor_age))

##
## factor_age [17,31] factor_age (31,36] factor_age (36,41]
##          0.2250303          0.2147190          0.1678124
## factor_age (41,49] factor_age (49,81]
##          0.1926810          0.1997574

#El perfil columna de les diferents columnes es pot considerar diferent que el
marginal? Evidentment SI

# HO: factor_duration -factor_age independency
chisq.test(table(df$factor_age, df$factor_duration))

##
##  Pearson's Chi-squared test
##
## data:  table(df$factor_age, df$factor_duration)
## X-squared = 24.084, df = 28, p-value = 0.6771

# Accepto la hipotesi nula porque el pvalor es 0.6771
```

En aquesta part de la nostra investigació podem veure que la hipòtesi nula s'accepta perquè el pvalor es 0.6771, es més gran que un 5%. Llavors, podem dir que la durada de la trucada no depèn de l'edat del nostre individu.

```
# CA - factor_duration vs factor_age
res.ca <- CA(table(df$factor_age, df$factor_duration))
# Interpretacio numerica: Com mes lluny estigui la rodona (blau) hace referencia al
factor que esta en las filas y el rojo a las columnas, entonces como mas lejos este
del centro de gravedad, quiere decir que es mas remarcables, es decir, mas raro es,
los que estan mas cerca no me aporta nada

#Link levels in row
#plot.CA(res.ca)
lines(res.ca$row$coord[,1], res.ca$row$coord[,2],lwd=2)
#No tenemos que ver nada porque hemos visto que no tienen nada que ver

#Link levels in columns
lines(res.ca$col$coord[,1], res.ca$col$coord[,2],lwd=2, col = “green")
```



Com podem veure a l'hora de l'execució tenim que el factor_duration-(182,236] és el que més destaca en que no ens aporta cap mena d'informacio ja que es troba més a prop del centre de gravetat, A partir de les taules de contingència i els seus diferents perfils intentem observar si hi pot haver alguna relació de dependència entre els dos factors, tot i així visualment ens resulta complicat.

Eigenvalues and dominant axes analysis

En aquest subapartat realitzarem un estudi dels valors propis i dels eixos dominants per tal de determinar quantes dimensions tindrem en compte.

```
#attributes(res.ca)
res.ca$eig

##         eigenvalue percentage of variance
## dim 1 0.0026443419              54.304636
## dim 2 0.0012712615              26.106835
## dim 3 0.0006783276              13.930247
## dim 4 0.0002755277               5.658282
##       cumulative percentage of variance
## dim 1                          54.30464
## dim 2                          80.41147
## dim 3                          94.34172
## dim 4                         100.00000

#No es extraño que los eigenvalues sean pequeños, cojemos tantas dimensiones como las
que tengan un valor propio > mitjana de este valor
mean(res.ca$eig[,1]) #Mean eigenvalue

## [1] 0.001217365
```

```
#KAISER: take as many as dimensions as eigenvalue > mean eig
sum(res.ca$eig[,1]) #Total inertia, contra mas grande hay mas realcion entre las
variables

## [1] 0.004869459

#Rows
res.ca$row

## $coord
##                         Dim 1       Dim 2        Dim 3        Dim 4
## factor_age [17,31]  0.06028947  0.02489821  0.016824123 -0.017911266
## factor_age (31,36] -0.02855663 -0.05990896  0.001849387 -0.011976354
## factor_age (36,41] -0.07137954  0.02821861  0.035431074  0.012368421
## factor_age (41,49]  0.05590097 -0.01718799 -0.006454706  0.027347518
## factor_age (49,81] -0.03117779  0.02922096 -0.044479508 -0.003718468
##
## $contrib
##                       Dim 1     Dim 2      Dim 3     Dim 4
## factor_age [17,31] 30.931887 10.973430  9.3900174 26.201633
## factor_age (31,36]  6.621655 60.620433  0.1082646 11.177750
## factor_age (36,41] 32.333583 10.511397 31.0565438  9.317238
## factor_age (41,49] 22.769835  4.477691  1.1834564 52.300922
## factor_age (49,81]  7.343039 13.417048 58.2617178  1.002457
##
## $cos2
##                       Dim 1      Dim 2        Dim 3       Dim 4
## factor_age [17,31] 0.7481200 0.12759233 0.0582576789 0.066029941
## factor_age (31,36] 0.1791707 0.78856386 0.0007514654 0.031513932
## factor_age (36,41] 0.6979823 0.10908582 0.1719751003 0.020956822
## factor_age (41,49] 0.7422793 0.07017444 0.0098965007 0.177649715
## factor_age (49,81] 0.2545863 0.22363176 0.5181605763 0.003621366
##
## $inertia
## [1] 0.0010933337 0.0009772756 0.0012249745 0.0008111667 0.0007627082

#Tenemos las coordenadas, las contribuciones, el cos2 (indica la calidad de la
representacion de cada una de las categorias en el eje que toca), inertia

#Cols
res.ca$col

## $coord
##                                    Dim 1         Dim 2        Dim 3
## factor_duration-[1,68]        -0.037838157 -0.0722213624 -0.003223726
## factor_duration-(68,104]      -0.079484346  0.0144258906 -0.032136966
## factor_duration-(104,139]      0.033810408  0.0122767778  0.007661304
## factor_duration-(139,182]     -0.007382708  0.0458581152 -0.028290026
## factor_duration-(182,236]      0.020271631  0.0001678354 -0.004295253
## factor_duration-(236,329]     -0.020308835 -0.0234604264  0.030890479
```

```
## factor_duration-(329,504]      -0.012654910  0.0399387089  0.047274023
## factor_duration-(504,2.12e+03]  0.105787697 -0.0158309148 -0.017544387
##                                       Dim 4
## factor_duration-[1,68]         -0.007649896
## factor_duration-(68,104]       -0.009047632
## factor_duration-(104,139]       0.013727248
## factor_duration-(139,182]       0.001523774
## factor_duration-(182,236]       0.031880042
## factor_duration-(236,329]       0.009810846
## factor_duration-(329,504]      -0.019119842
## factor_duration-(504,2.12e+03] -0.020319730
##
## $contrib
##                                     Dim 1        Dim 2      Dim 3
## factor_duration-[1,68]          6.8855540 5.217867e+01  0.1948379
## factor_duration-(68,104]       30.0939744 2.061979e+00 19.1780380
## factor_duration-(104,139]       5.3490977 1.467004e+00  1.0706886
## factor_duration-(139,182]       0.2583755 2.073648e+01 14.7898845
## factor_duration-(182,236]       1.9103338 2.723842e-04  0.3343393
## factor_duration-(236,329]       1.9520412 5.418430e+00 17.6054166
## factor_duration-(329,504]       0.7567194 1.567789e+01 41.1661239
## factor_duration-(504,2.12e+03] 52.7939041 2.459281e+00  5.6606712
##                                     Dim 4
## factor_duration-[1,68]          2.7011103
## factor_duration-(68,104]        3.7422993
## factor_duration-(104,139]       8.4625062
## factor_duration-(139,182]       0.1056364
## factor_duration-(182,236]      45.3442221
## factor_duration-(236,329]       4.3720403
## factor_duration-(329,504]      16.5782111
## factor_duration-(504,2.12e+03] 18.6939743
##
## $cos2
##                                    Dim 1        Dim 2       Dim 3
## factor_duration-[1,68]         0.2131635 7.765763e-01 0.001547281
## factor_duration-(68,104]       0.8268767 2.723722e-02 0.135172172
## factor_duration-(104,139]      0.7418214 9.780641e-02 0.038089380
## factor_duration-(139,182]      0.0184129 7.104333e-01 0.270369437
## factor_duration-(182,236]      0.2842387 1.948377e-05 0.012760955
## factor_duration-(236,329]      0.2048606 2.733758e-01 0.473955531
## factor_duration-(329,504]      0.0367676 3.662142e-01 0.513088482
## factor_duration-(504,2.12e+03] 0.9201376 2.060604e-02 0.025308058
##                                       Dim 4
## factor_duration-[1,68]         0.0087129229
## factor_duration-(68,104]       0.0107138956
## factor_duration-(104,139]      0.1222828325
## factor_duration-(139,182]      0.0007843904
## factor_duration-(182,236]      0.7029808969
## factor_duration-(236,329]      0.0478080704
```

```
## factor_duration-(329,504]      0.0839297108
## factor_duration-(504,2.12e+03] 0.0339483217
##
## $inertia
## [1] 0.0008541689 0.0009624017 0.0001906772 0.0003710622 0.0001777230
## [6] 0.0002519696 0.0005442359 0.0015172202

#Durada mes curta es la que te mes contribucio!

#Phi2 = Intensity of the association Chisq/nobservations
sum(res.ca$eig[,1]) #Total inertia = Phi2

## [1] 0.004869459

chisq.test(table(df$factor_age, df$factor_duration))

##
##  Pearson's Chi-squared test
##
## data:  table(df$factor_age, df$factor_duration)
## X-squared = 24.084, df = 28, p-value = 0.6771

#24.084/4946 porque son las observaciones
```

Job i Factor_duration

```
# Contingency tables - Complex : solo cuentan con los target discretizados

# Target factor_duration vs job
# Podemos elegir la variable que queramos con la de f_duration y en este caso hemos
elegido job para este ejemplo

table(df$job, df$factor_duration)

##
##                    factor_duration-[1,68] factor_duration-(68,104]
##   Job_admin.                          162                      169
##   Job_blue-collar                     131                      141
##   Job_entrepreneur                     18                       17
##   Job_housemaid                        14                       14
##   Job_management                       47                       35
##   Job_retired                          18                       29
##   Job_self-employed                    20                       25
##   Job_services                         75                       61
##   Job_student                           8                       17
##   Job_technician                      109                       96
##   Job_unemployed                       20                       14
##   Job_unknown                           7                        5
##
##                    factor_duration-(104,139] factor_duration-(139,182]
```

```
##   Job_admin.                          164                     167
##   Job_blue-collar                     133                     135
##   Job_entrepreneur                     12                      18
##   Job_housemaid                        22                      17
##   Job_management                       39                      47
##   Job_retired                          24                      33
##   Job_self-employed                    23                      20
##   Job_services                         52                      52
##   Job_student                          10                       7
##   Job_technician                      116                     105
##   Job_unemployed                       10                      16
##   Job_unknown                           7                       3
##
##                    factor_duration-(182,236] factor_duration-(236,329]
##   Job_admin.                          150                     157
##   Job_blue-collar                     137                     157
##   Job_entrepreneur                     24                      21
##   Job_housemaid                        16                      19
##   Job_management                       53                      45
##   Job_retired                          21                      28
##   Job_self-employed                    12                      13
##   Job_services                         54                      57
##   Job_student                          13                      19
##   Job_technician                      111                      85
##   Job_unemployed                       10                      15
##   Job_unknown                           7                       3
##
##                    factor_duration-(329,504]
##   Job_admin.                          167
##   Job_blue-collar                     165
##   Job_entrepreneur                     18
##   Job_housemaid                        10
##   Job_management                       43
##   Job_retired                          29
##   Job_self-employed                    17
##   Job_services                         64
##   Job_student                          14
##   Job_technician                       82
##   Job_unemployed                        5
##   Job_unknown                           4
##
##                    factor_duration-(504,2.12e+03]
##   Job_admin.                               165
##   Job_blue-collar                          145
##   Job_entrepreneur                          32
##   Job_housemaid                             14
##   Job_management                            36
##   Job_retired                               24
##   Job_self-employed                         22
```

```
## Job_services                                  58
## Job_student                                   17
## Job_technician                                80
## Job_unemployed                                17
## Job_unknown                                    7
```

```r
#Le digo que calcule unas probabilidades en la dimension 1, calculo los perfiles por
fila que tenemos
#Calculo los perfiles de fila y la suma tendria que dar mas o menos 1 y tenemos que
ver si es equivalente al perfil marginal fila
prop.table(table(df$job, df$factor_duration), 1) # Por filas
```

```
##
##                    factor_duration-[1,68] factor_duration-(68,104]
##   Job_admin.                   0.12451960               0.12990008
##   Job_blue-collar              0.11451049               0.12325175
##   Job_entrepreneur             0.11250000               0.10625000
##   Job_housemaid                0.11111111               0.11111111
##   Job_management               0.13623188               0.10144928
##   Job_retired                  0.08737864               0.14077670
##   Job_self-employed            0.13157895               0.16447368
##   Job_services                 0.15856237               0.12896406
##   Job_student                  0.07619048               0.16190476
##   Job_technician               0.13903061               0.12244898
##   Job_unemployed               0.18691589               0.13084112
##   Job_unknown                  0.16279070               0.11627907
##
##                    factor_duration-(104,139] factor_duration-(139,182]
##   Job_admin.                      0.12605688                0.12836280
##   Job_blue-collar                 0.11625874                0.11800699
##   Job_entrepreneur                0.07500000                0.11250000
##   Job_housemaid                   0.17460317                0.13492063
##   Job_management                  0.11304348                0.13623188
##   Job_retired                     0.11650485                0.16019417
##   Job_self-employed               0.15131579                0.13157895
##   Job_services                    0.10993658                0.10993658
##   Job_student                     0.09523810                0.06666667
##   Job_technician                  0.14795918                0.13392857
##   Job_unemployed                  0.09345794                0.14953271
##   Job_unknown                     0.16279070                0.06976744
##
##                    factor_duration-(182,236] factor_duration-(236,329]
##   Job_admin.                      0.11529593                0.12067640
##   Job_blue-collar                 0.11975524                0.13723776
##   Job_entrepreneur                0.15000000                0.13125000
##   Job_housemaid                   0.12698413                0.15079365
##   Job_management                  0.15362319                0.13043478
##   Job_retired                     0.10194175                0.13592233
##   Job_self-employed               0.07894737                0.08552632
##   Job_services                    0.11416490                0.12050740
```

```
##   Job_student                      0.12380952                0.18095238
##   Job_technician                   0.14158163                0.10841837
##   Job_unemployed                   0.09345794                0.14018692
##   Job_unknown                      0.16279070                0.06976744
##
##                      factor_duration-(329,504]
##   Job_admin.                       0.12836280
##   Job_blue-collar                  0.14423077
##   Job_entrepreneur                 0.11250000
##   Job_housemaid                    0.07936508
##   Job_management                   0.12463768
##   Job_retired                      0.14077670
##   Job_self-employed                0.11184211
##   Job_services                     0.13530655
##   Job_student                      0.13333333
##   Job_technician                   0.10459184
##   Job_unemployed                   0.04672897
##   Job_unknown                      0.09302326
##
##                      factor_duration-(504,2.12e+03]
##   Job_admin.                            0.12682552
##   Job_blue-collar                       0.12674825
##   Job_entrepreneur                      0.20000000
##   Job_housemaid                         0.11111111
##   Job_management                        0.10434783
##   Job_retired                           0.11650485
##   Job_self-employed                     0.14473684
##   Job_services                          0.12262156
##   Job_student                           0.16190476
##   Job_technician                        0.10204082
##   Job_unemployed                        0.15887850
##   Job_unknown                           0.16279070

#Marginal row profile
prop.table(table(df$factor_duration))

##
##        factor_duration-[1,68]       factor_duration-(68,104]
##                     0.1271735                      0.1259604
##      factor_duration-(104,139]      factor_duration-(139,182]
##                     0.1237364                      0.1253538
##      factor_duration-(182,236]      factor_duration-(236,329]
##                     0.1229276                      0.1251516
##      factor_duration-(329,504] factor_duration-(504,2.12e+03]
##                     0.1249495                      0.1247473

#Esta proporcion se mantiene en cualquiera de los colectivos mirados anteriormente? Se
tiene que hacer la comparacion

#Podemos comprobar ahora los perfiles columna
```

```r
#Column profile
prop.table(table(df$job, df$factor_duration), 2) # dim 2
```

```
##
##                     factor_duration-[1,68] factor_duration-(68,104]
##   Job_admin.                   0.257551669              0.271268058
##   Job_blue-collar              0.208267091              0.226324238
##   Job_entrepreneur             0.028616852              0.027287319
##   Job_housemaid                0.022257552              0.022471910
##   Job_management               0.074721781              0.056179775
##   Job_retired                  0.028616852              0.046548957
##   Job_self-employed            0.031796502              0.040128411
##   Job_services                 0.119236884              0.097913323
##   Job_student                  0.012718601              0.027287319
##   Job_technician               0.173290938              0.154093098
##   Job_unemployed               0.031796502              0.022471910
##   Job_unknown                  0.011128776              0.008025682
##
##                     factor_duration-(104,139] factor_duration-(139,182]
##   Job_admin.                      0.267973856               0.269354839
##   Job_blue-collar                 0.217320261               0.217741935
##   Job_entrepreneur                0.019607843               0.029032258
##   Job_housemaid                   0.035947712               0.027419355
##   Job_management                  0.063725490               0.075806452
##   Job_retired                     0.039215686               0.053225806
##   Job_self-employed               0.037581699               0.032258065
##   Job_services                    0.084967320               0.083870968
##   Job_student                     0.016339869               0.011290323
##   Job_technician                  0.189542484               0.169354839
##   Job_unemployed                  0.016339869               0.025806452
##   Job_unknown                     0.011437908               0.004838710
##
##                     factor_duration-(182,236] factor_duration-(236,329]
##   Job_admin.                      0.246710526               0.253634895
##   Job_blue-collar                 0.225328947               0.253634895
##   Job_entrepreneur                0.039473684               0.033925687
##   Job_housemaid                   0.026315789               0.030694669
##   Job_management                  0.087171053               0.072697900
##   Job_retired                     0.034539474               0.045234249
##   Job_self-employed               0.019736842               0.021001616
##   Job_services                    0.088815789               0.092084006
##   Job_student                     0.021381579               0.030694669
##   Job_technician                  0.182565789               0.137318255
##   Job_unemployed                  0.016447368               0.024232633
##   Job_unknown                     0.011513158               0.004846527
##
##                     factor_duration-(329,504]
##   Job_admin.                      0.270226537
##   Job_blue-collar                 0.266990291
##   Job_entrepreneur                0.029126214
```

```
##   Job_housemaid                    0.016181230
##   Job_management                   0.069579288
##   Job_retired                      0.046925566
##   Job_self-employed                0.027508091
##   Job_services                     0.103559871
##   Job_student                      0.022653722
##   Job_technician                   0.132686084
##   Job_unemployed                   0.008090615
##   Job_unknown                      0.006472492
##
##                      factor_duration-(504,2.12e+03]
##   Job_admin.                            0.267423015
##   Job_blue-collar                       0.235008104
##   Job_entrepreneur                      0.051863857
##   Job_housemaid                         0.022690438
##   Job_management                        0.058346840
##   Job_retired                           0.038897893
##   Job_self-employed                     0.035656402
##   Job_services                          0.094003241
##   Job_student                           0.027552674
##   Job_technician                        0.129659643
##   Job_unemployed                        0.027552674
##   Job_unknown                           0.011345219
```

```r
#Marginal colum profile
prop.table(table(df$job))
```

```
##
##        Job_admin.   Job_blue-collar  Job_entrepreneur     Job_housemaid
##       0.263040841       0.231298019       0.032349373       0.025475131
##    Job_management       Job_retired Job_self-employed      Job_services
##       0.069753336       0.041649818       0.030731905       0.095632835
##       Job_student    Job_technician    Job_unemployed       Job_unknown
##       0.021229276       0.158511929       0.021633643       0.008693894
```

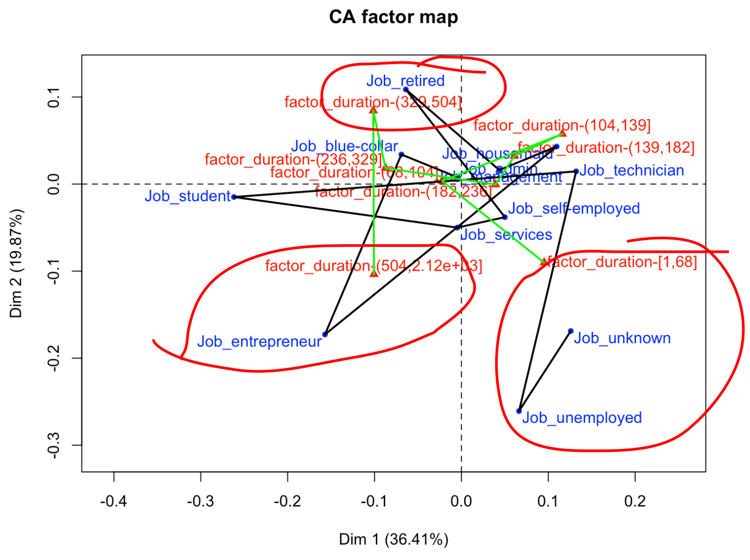
```r
#El perfil columna de les diferents columnes es pot considerar diferent que el
marginal? Evidentment SI

# HO: factor_duration -factor_age independency
chisq.test(table(df$job, df$factor_duration))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(df$job, df$factor_duration)
## X-squared = 95.774, df = 77, p-value = 0.07247
```

```r
# Accepto la hipotesi nula porque el pvalor es 0.07247
```

En aquesta part de la nostra investigacio podem veure que la hipotesi nula s'accepta perquè el pvalor és 0.07247, es mes gran que un 5%.Encara que es molt petit i si que es podria arribar a pensar que es pogués rebutjar. Llavors, podem dir que la durada de la trucada no depèn del treball o a que es dediqui el nostre individu.

```
# CA - factor_duration vs factor_age
res.ca <- CA(table(df$job, df$factor_duration))

#Link levels in row
lines(res.ca$row$coord[,1], res.ca$row$coord[,2],lwd=2)
#No tenemos que ver nada porque hemos visto que no tienen nada que ver
#Link levels in columns
lines(res.ca$col$coord[,1], res.ca$col$coord[,2],lwd=2, col = "green")
```



Com podem veure a l'hora de l'execució tenim que el Job_admin, Job_management i factor_duration-(68,104] són els que més destaquen en que no ens aporta cap mena d'informació ja que es troba més a prop del centre de gravetat, A partir de les taules de contingència i els seus diferents perfils intentem observar si hi pot haver alguna relació de dependència entre ~~els dos factors, tot i així visualment ens resulta complicat.~~

Eigenvalues and dominant axes analysis

En aquest subapartat realitzarem un estudi dels valors propis i dels eixos dominants per tal de determinar quantes dimensions tindrem en compte.

```
res.ca$eig
```

```
##        eigenvalue percentage of variance cumulative percentage of variance
## dim 1 0.007050333              36.409534                          36.40953
## dim 2 0.003847258              19.868124                          56.27766
## dim 3 0.003249026              16.778713                          73.05637
## dim 4 0.002161419              11.162064                          84.21844
## dim 5 0.001718252               8.873445                          93.09188
## dim 6 0.001041702               5.379590                          98.47147
## dim 7 0.000295984               1.528529                         100.00000
```

```
#No es extraño que los eigenvalues sean pequeños, cojemos tantas dimensiones como las
que tengan un valor propio > mitjana de este valor
mean(res.ca$eig[,1]) #Mean eigenvalue
```

```
## [1] 0.002766282
```

```
#KAISER: take as many as dimensions as eigenvalue > mean eig
sum(res.ca$eig[,1]) #Total inertia, contra mas grande hay mas realcion entre las
variables
```

```
## [1] 0.01936397
```

```
#Rows
res.ca$row
```

```
## $coord
##                          Dim 1        Dim 2        Dim 3        Dim 4
## Job_admin.        -0.004186212  0.007701902 -0.028673409 -0.002854957
## Job_blue-collar   -0.069096677  0.033980063  0.012824475 -0.012061634
## Job_entrepreneur  -0.157069320 -0.172777310  0.095319234  0.073797366
## Job_housemaid      0.109639754  0.043116843  0.011457101  0.159644423
## Job_management     0.043596504  0.016129874  0.118496133 -0.013785385
## Job_retired       -0.064052069  0.108786121 -0.065965837  0.053120289
## Job_self-employed  0.050005969 -0.037996065 -0.208308255  0.003606401
## Job_services      -0.004690390 -0.049982823 -0.007662952 -0.096663163
## Job_student       -0.261854900 -0.014852099  0.006879286  0.082853328
## Job_technician     0.132101028  0.014640019  0.023566514  0.012316894
## Job_unemployed     0.066308261 -0.260754891 -0.065825305  0.068094425
## Job_unknown        0.125895275 -0.168760946  0.041355718 -0.024777522
##                          Dim 5
## Job_admin.         0.000286884
## Job_blue-collar    0.001586189
## Job_entrepreneur   0.021690042
## Job_housemaid      0.017201430
## Job_management    -0.044076180
## Job_retired       -0.077827955
## Job_self-employed  0.049353299
## Job_services      -0.007967764
## Job_student        0.094554874
## Job_technician     0.019815606
## Job_unemployed    -0.149509980
```

```
## Job_unknown        0.237538144
##
## $contrib
##                        Dim 1      Dim 2       Dim 3       Dim 4
## Job_admin.         0.06538166  0.4055718  6.65623555  0.09919351
## Job_blue-collar   15.66306072  6.9417492  1.17084112  1.55684451
## Job_entrepreneur  11.31980610 25.1008208  9.04635891  8.15095688
## Job_housemaid      4.34353103  1.2310026  0.10292308 30.03896341
## Job_management     1.88043659  0.4717106 30.14534050  0.61328707
## Job_retired        2.42364957 12.8117579  5.57825188  5.43744611
## Job_self-employed  1.08999261  1.1532285 41.04396339  0.01849262
## Job_services       0.02984114  6.2100805  0.17284073 41.34186280
## Job_student       20.64652601  0.1217193  0.03092208  6.74242469
## Job_technician    39.23419417  0.8830675  2.70956455  1.11256497
## Job_unemployed     1.34913477 38.2334270  2.88510940  4.64102364
## Job_unknown        1.95444563  6.4358644  0.45764881  0.24693979
##                         Dim 5
## Job_admin.         0.001259937
## Job_blue-collar    0.033868409
## Job_entrepreneur   0.885727032
## Job_housemaid      0.438691050
## Job_management     7.886532880
## Job_retired       14.682417960
## Job_self-employed  4.356473625
## Job_services       0.353340345
## Job_student       11.046285370
## Job_technician     3.622345822
## Job_unemployed    28.143834431
## Job_unknown       28.549223139
##
## $cos2
##                        Dim 1       Dim 2        Dim 3        Dim 4
## Job_admin.        0.016594367 0.056171271 0.7785329044 0.0077182245
## Job_blue-collar   0.739280414 0.178790006 0.0254667819 0.0225271988
## Job_entrepreneur  0.315537130 0.381804603 0.1162060750 0.0696545455
## Job_housemaid     0.245681834 0.037995385 0.0026827883 0.5208881086
## Job_management    0.102696948 0.014057762 0.7586868817 0.0102681461
## Job_retired       0.130823433 0.377368975 0.1387577892 0.0899787729
## Job_self-employed 0.048798014 0.028173169 0.8467816202 0.0002538087
## Job_services      0.001699359 0.192978417 0.0045358573 0.7217539703
## Job_student       0.677204963 0.002178584 0.0004673965 0.0677982733
## Job_technician    0.916061394 0.011251113 0.0291543176 0.0079636948
## Job_unemployed    0.040898751 0.632469657 0.0403051494 0.0431318294
## Job_unknown       0.151106062 0.271523191 0.0163055017 0.0058530031
##                         Dim 5
## Job_admin.        7.793471e-05
## Job_blue-collar   3.895871e-04
## Job_entrepreneur  6.017118e-03
## Job_housemaid     6.047363e-03
```

```
## Job_management    1.049693e-01
## Job_retired       1.931481e-01
## Job_self-employed 4.753252e-02
## Job_services      4.903883e-03
## Job_student       8.830119e-02
## Job_technician    2.061232e-02
## Job_unemployed    2.079290e-01
## Job_unknown       5.379349e-01
##
## $inertia
##  [1] 0.0002777825 0.0014937470 0.0025292871 0.0012464633 0.0012909540
##  [6] 0.0013061525 0.0015748203 0.0012380548 0.0021494951 0.0030196024
## [11] 0.0023257064 0.0009119086
```

```
#Tenemos las coordenadas, las contribuciones, el cos2 (indica la calidad de la
representacion de cada una de las categorias en el eje que toca), inertia

#Cols
res.ca$col
```

```
## $coord
##                                      Dim 1         Dim 2        Dim 3
## factor_duration-[1,68]          0.09531274 -0.0902001831  0.004152345
## factor_duration-(68,104]       -0.02226995  0.0035790782 -0.085408752
## factor_duration-(104,139]       0.11616899  0.0575727078 -0.037797854
## factor_duration-(139,182]       0.06103176  0.0322551944 -0.020524250
## factor_duration-(182,236]       0.03929675 -0.0004978809  0.122488603
## factor_duration-(236,329]      -0.08742549  0.0184521454  0.038039160
## factor_duration-(329,504]      -0.10141238  0.0842994678  0.004928502
## factor_duration-(504,2.12e+03] -0.10067398 -0.1036350174 -0.023679053
##                                        Dim 4        Dim 5
## factor_duration-[1,68]         -0.0736844595 -0.023341450
## factor_duration-(68,104]       -0.0007433834  0.016809274
## factor_duration-(104,139]       0.0310640066  0.056862266
## factor_duration-(139,182]       0.0338717427 -0.080127776
## factor_duration-(182,236]       0.0143242316  0.033492169
## factor_duration-(236,329]       0.0421650872 -0.036944879
## factor_duration-(329,504]      -0.0803395322  0.007811121
## factor_duration-(504,2.12e+03]  0.0350721387  0.027175814
##
## $contrib
##                                    Dim 1        Dim 2       Dim 3
## factor_duration-[1,68]         16.386601 2.689429e+01  0.06748859
## factor_duration-(68,104]        0.886059 4.193966e-02 28.28039945
## factor_duration-(104,139]      23.684714 1.066054e+01  5.44099668
## factor_duration-(139,182]       6.622772 3.389889e+00  1.62524571
## factor_duration-(182,236]       2.692484 7.920435e-04 56.76592066
## factor_duration-(236,329]      13.567602 1.107590e+00  5.57372088
## factor_duration-(329,504]      18.226645 2.307983e+01  0.09341381
## factor_duration-(504,2.12e+03] 17.933124 3.482513e+01  2.15281423
```

```
##                                     Dim 4      Dim 5
## factor_duration-[1,68]         31.94547451  4.0324168
## factor_duration-(68,104]        0.00322048  2.0713099
## factor_duration-(104,139]       5.52424913 23.2840689
## factor_duration-(139,182]       6.65386009 46.8400112
## factor_duration-(182,236]       1.16695240  8.0250777
## factor_duration-(236,329]      10.29445944  9.9416456
## factor_duration-(329,504]      37.31246714  0.4436845
## factor_duration-(504,2.12e+03]  7.09931680  5.3617852
##
## $cos2
##                                     Dim 1        Dim 2        Dim 3
## factor_duration-[1,68]         0.38193087 3.420563e-01 0.0007248861
## factor_duration-(68,104]       0.05170872 1.335574e-03 0.7605543115
## factor_duration-(104,139]      0.58668609 1.440982e-01 0.0621097313
## factor_duration-(139,182]      0.25497352 7.121682e-02 0.0288348615
## factor_duration-(182,236]      0.08321936 1.335863e-05 0.8085417008
## factor_duration-(236,329]      0.46736782 2.081979e-02 0.0884798704
## factor_duration-(329,504]      0.42099659 2.909017e-01 0.0009943206
## factor_duration-(504,2.12e+03] 0.41139957 4.359557e-01 0.0227592517
##                                       Dim 4       Dim 5
## factor_duration-[1,68]         2.282625e-01 0.022905431
## factor_duration-(68,104]       5.761707e-05 0.029459366
## factor_duration-(104,139]      4.195080e-02 0.140563863
## factor_duration-(139,182]      7.853412e-02 0.439490460
## factor_duration-(182,236]      1.105742e-02 0.060450181
## factor_duration-(236,329]      1.087148e-01 0.083462456
## factor_duration-(329,504]      2.642136e-01 0.002497602
## factor_duration-(504,2.12e+03] 4.992911e-02 0.029977434
##
## $inertia
## [1] 0.003024919 0.001208115 0.002846243 0.001831278 0.002281069 0.002046699
## [7] 0.003052374 0.003073277
```

```
#Durada mes curta es la que te mes contribucio!

#Phi2 = Intensity of the association Chisq/nobservations
sum(res.ca$eig[,1]) #Total inertia = Phi2
```

```
## [1] 0.01936397
```

```
chisq.test(table(df$job, df$factor_duration))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(df$job, df$factor_duration)
## X-squared = 95.774, df = 77, p-value = 0.07247
```

```
#95.774/4946 porque son las observaciones
```