

Course Practical Assignment - 2nd Deliverable (21 d'abril del 2019)

Josep Clotet Ginovart

Eric Martin Obispo

Bank client data



Description of input variables:

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

Loading packages:

Load data from Deliverable 1:

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

load( paste0(dirwd, "/bank-additional/Bank5000_validated.RData") )
summary(df)
```

```

##          age                job                marital
## Min.      :18.00   job-admin.      :1246   marital-divorced: 554
## 1st Qu.:32.00   job-blue-collar:1171   marital-married :3055
## Median :38.00   job-technician : 796   marital-single  :1377
## Mean      :40.07   job-services    : 498
## 3rd Qu.:47.00   job-management  : 411
## Max.      :87.00   job-retired     : 205
##                (Other)      : 659
##                education                default
## education-basic.4y      : 533   default-no      :3954
## education-basic.6y      : 289   default-unknown:1032
## education-basic.9y      : 767
## education-high.school   :1218
## education-professional.course: 615
## education-university.degree :1564
##
##          housing                loan                contact
## housing-no :2261   loan-no :4217   contact-cellular :3122
## housing-yes:2725   loan-yes: 769   contact-telephone:1864
##
##
##
##
##          month                day_of_week                duration
## month-may:1741   day_of_week-1mon:1016   Min.      : 5.0
## month-jul: 829   day_of_week-2tue:1043   1st Qu.: 101.0
## month-aug: 697   day_of_week-3wed: 971   Median : 177.0
## month-jun: 652   day_of_week-4thu:1034   Mean      : 250.6
## month-nov: 507   day_of_week-5fri: 922   3rd Qu.: 316.0
## month-apr: 310                                     Max.      :1580.0
## (Other) : 250
##          campaign                pdays                previous
## Min.      : 1.000   Min.      : 0.00   Min.      :0.0000
## 1st Qu.: 1.000   1st Qu.:19.00   1st Qu.:0.0000
## Median : 2.000   Median :19.00   Median :0.0000
## Mean      : 2.535   Mean      :18.53   Mean      :0.1598
## 3rd Qu.: 3.000   3rd Qu.:19.00   3rd Qu.:0.0000
## Max.      :25.000   Max.      :19.00   Max.      :4.0000
##
##                poutcome                emp.var.rate                cons.price.idx
## poutcome-failure      : 477   Min.      :-3.40000   Min.      :92.20
## poutcome-nonexistent:4353   1st Qu.: -1.80000   1st Qu.:93.08
## poutcome-success      : 156   Median : 1.10000   Median :93.75
##                                     Mean      : 0.06446   Mean      :93.57
##                                     3rd Qu.: 1.40000   3rd Qu.:93.99
##                                     Max.      : 1.40000   Max.      :94.77
##
##          cons.conf.idx                euribor3m                nr.employed                y
## Min.      :-50.80   Min.      :0.635   Min.      :4964   y-no :4429
## 1st Qu.: -42.70   1st Qu.:1.334   1st Qu.:5099   y-yes: 557
## Median : -41.80   Median :4.857   Median :5191
## Mean      : -40.43   Mean      :3.614   Mean      :5166
## 3rd Qu.: -36.40   3rd Qu.:4.961   3rd Qu.:5228

```

```

## Max.      :-26.90    Max.      :5.000    Max.      :5228
##
## num_missings      num_outliers      num_errors
## Min.      :0.0000    Min.      :0.00000    Min.      :0
## 1st Qu.:0.0000    1st Qu.:0.00000    1st Qu.:0
## Median :0.0000    Median :0.00000    Median :0
## Mean    :0.1111    Mean    :0.00361    Mean    :0
## 3rd Qu.:0.0000    3rd Qu.:0.00000    3rd Qu.:0
## Max.     :3.0000    Max.     :2.00000    Max.     :0
##
##                f.season      minutes      f.age
## season-spring      :2117    Min.      : 0.08333    f.age-[18,32]:1352
## season-summer      :2178    1st Qu.: 1.68333    f.age-(32,38]:1205
## season-autumnwinter: 691    Median : 2.95000    f.age-(38,47]:1220
##                      Mean    : 4.17703    f.age-(47,87]:1209
##                      3rd Qu.: 5.26667
##                      Max.     :26.33333
##
##                f.duration      f.campaign
## f.duration-[5,101]      :1252    f.campaign-[0,2] :3392
## f.duration-(101,177]    :1243    f.campaign-(2,5] :1181
## f.duration-(177,316]    :1247    f.campaign-(5,25]: 413
## f.duration-(316,1.58e+03]:1244
##
##
##
##                f.pdays      f.previous
## f.pdays-sometime: 177    f.previous-never:4353
## f.pdays-never    :4809    f.previous-some : 633
##
##
##
##
##                f.emp.var.rate      f.cons.price.idx
## f.emp.var.rate-[-Inf,0]:2086    f.cons.price.idx-[92.2,93.1]:1409
## f.emp.var.rate-(0, Inf]:2900    f.cons.price.idx-(93.1,93.7]:1086
##                                f.cons.price.idx-(93.7,94] :1819
##                                f.cons.price.idx-(94,94.8]  : 672
##
##
##
##                f.cons.conf.idx      f.euribor3m
## f.cons.conf.idx-[-50.8,-42.7]:1856    f.euribor3m-[0.635,1.33]:1254
## f.cons.conf.idx-(-42.7,-41.8]: 967    f.euribor3m-(1.33,4.86] :1466
## f.cons.conf.idx-(-41.8,-36.4]:1231    f.euribor3m-(4.86,4.96] :1130
## f.cons.conf.idx-(-36.4,-26.9]: 932    f.euribor3m-(4.96,5]    :1136
##
##
##
##                f.nr.employed
## f.nr.employed-[4.96e+03,5.1e+03] :1639
## f.nr.employed-(5.1e+03,5.19e+03] :1003
## f.nr.employed-(5.19e+03,5.23e+03]:2344

```



```
##
##
##
##
```

CORRESPONDENCE ANALYSIS (CA)

Realitzarem un analisi amb taules de correspondencia i mapes de factors entre la variable numerica target duracio discretitzada en 4 nivells (corregit de l'entrega 1) i diversos factors que hem trobat que tenen una correspondencia significativa.

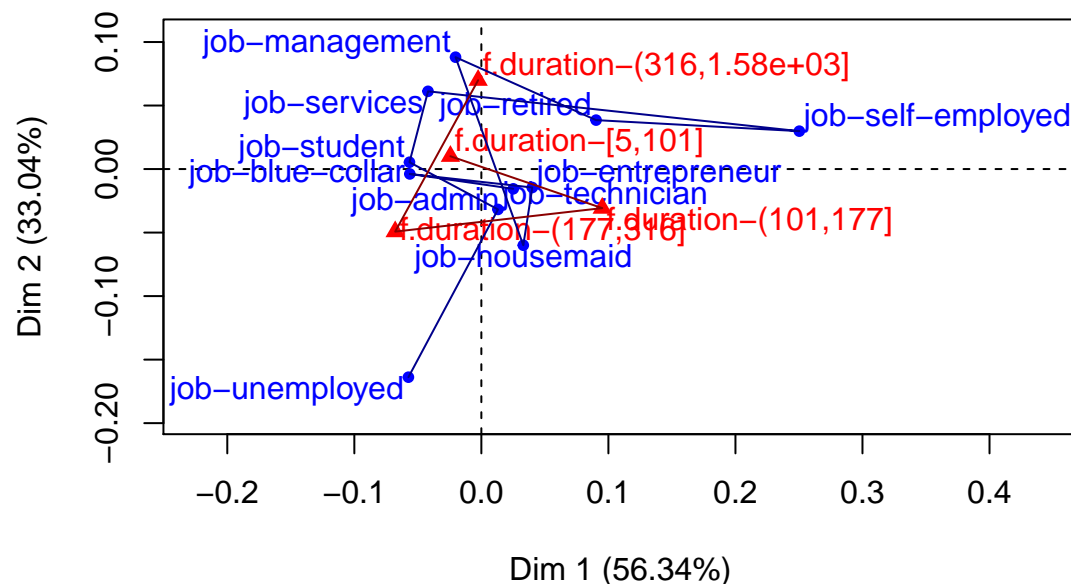
Primer veiem com la duracio de la trucada no te cap relacio amb el job de l'individu, ja que no podem rebutjar la hipotesi H_0 : *f.duration* no te cap relacio amb la variable *job* amb el valor p obtingut en el Chi Square test. Com mes aprop surten al grafic les categories d'ambdues variables analitzades, mes relacionades estan. En aquesta comparacio, com s'acaba de comentar, no es pot extreure res significatiu, mes enlla que potser les categories job-unemployed i job-self-employed van mes per lliure.

```
# H0: f.duration no te cap relacio amb variable job
chisq.test( table( df$job, df$f.duration) )
```

```
##
## Pearson's Chi-squared test
##
## data: table(df$job, df$f.duration)
## X-squared = 31.496, df = 30, p-value = 0.3913
```

```
# CA - f.duration vs variable job
res.ca<-CA( table( df$job, df$f.duration) )
lines(res.ca$row$coord[,1], res.ca$row$coord[,2], col="darkblue")
lines(res.ca$col$coord[,1], res.ca$col$coord[,2], col="darkred")
```

CA factor map



Ara testejarem la mateixa hipotesi i mostrarem el mateix mapa de factors pero amb altres categories factor que si que obtindrem que tenen una relacio significativa. El primer cas es l'epoca de l'any **f.season** en la qual es realitza la trucada (p valor = 2.506e-07). Comparant el profile de la taula de contingencia de proporcions per fila amb el profile marginal de la duracio veiem com hi ha un 28,5% de trucades amb duracions molt curtes a l'estiu respecte un 25% de trucades amb duracions curtes en tot l'any. D'altra banda, hi ha per sobre d'un 27% de trucades amb duracions llargues a la primavera, respecte un 25% de trucades en la mateixa duracio en tot l'any. Si comparem el profile de la taula de contingencia de proporcions per columna amb el profile marginal de la f.season veiem com el 42.5% de trucades es realitzen a la primavera, i en canvi mes d'un 46% de trucades corresponen a la primavera i a duracions llargues. A mes, el 43.7% de trucades es realitzen a l'estiu, i nomes prop d'un 40% de trucades corresponen a l'estiu i a duracions llargues. Aquesta mateixa informacio es pot veure representada en un mapa de factors de dues dimensions. Agafant nomes la primera dimensio ja seria suficient per a representar un 98.9% de la variancia del conjunt de les dades (Kaiser: take as many dimensions as eigenvalue > mean of eigenvalues).

```
chisq.test( table( df$f.season, df$f.duration) )

##
## Pearson's Chi-squared test
##
## data:  table(df$f.season, df$f.duration)
## X-squared = 41.318, df = 6, p-value = 2.506e-07
```

#Row/Column profile

```
prop.table( table(df$f.season, df$f.duration), 1 ) #1->per files
```

	f.duration-[5,101]	f.duration-(101,177]
season-spring	0.2106755	0.2442135
season-summer	0.2855831	0.2534435
season-autumnwinter	0.2662808	0.2518090

	f.duration-(177,316]	f.duration-(316,1.58e+03]
season-spring	0.2735002	0.2716108
season-summer	0.2277319	0.2332415
season-autumnwinter	0.2489146	0.2329957

#Marginal Row/Column profile

```
prop.table( table(df$f.duration)) #1->per files
```

	f.duration-[5,101]	f.duration-(101,177]
	0.2511031	0.2492980

	f.duration-(177,316]	f.duration-(316,1.58e+03]
	0.2501003	0.2494986

#2->per columnes

```
prop.table( table(df$f.season, df$f.duration), 2 ) #2->per columnes
```

	f.duration-[5,101]	f.duration-(101,177]
season-spring	0.3562300	0.4159292
season-summer	0.4968051	0.4440869
season-autumnwinter	0.1469649	0.1399839

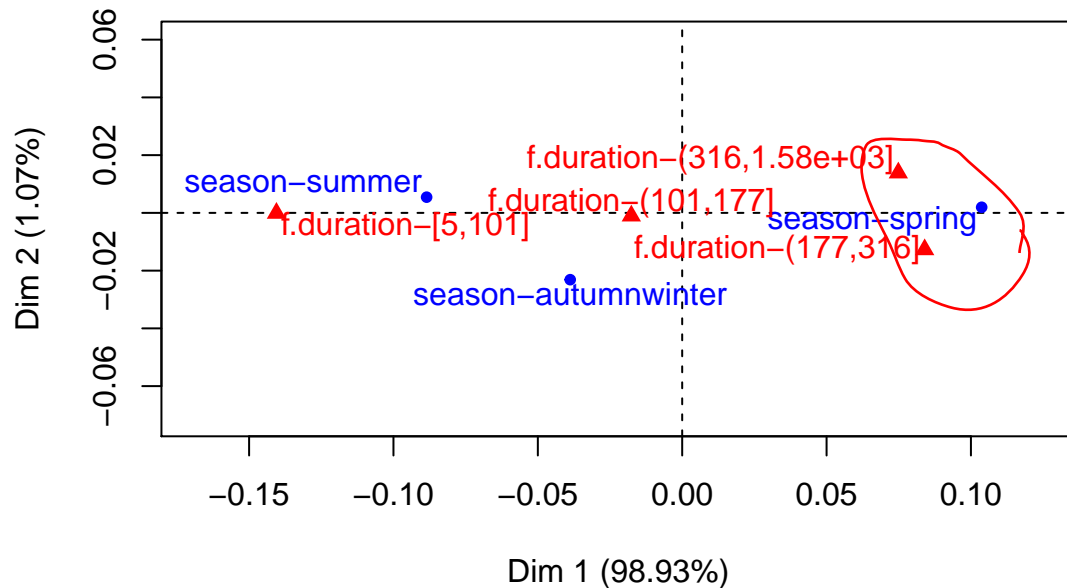
	f.duration-(177,316]	f.duration-(316,1.58e+03]
season-spring	0.4643144	0.4622186
season-summer	0.3977546	0.4083601
season-autumnwinter	0.1379310	0.1294212

```
prop.table( table(df$f.season)) #2->per columnes
```

```
##
##      season-spring      season-summer season-autumnwinter
##      0.4245888      0.4368231      0.1385880
```

```
res.ca<-CA( table( df$f.season, df$f.duration) )
```

CA factor map



```
attributes(res.ca); res.ca$eig #valors eig no normalitzats!
```

```
## $names
## [1] "eig" "call" "row" "col" "svd"
##
## $class
## [1] "CA" "list"
##
##      eigenvalue percentage of variance
## dim 1 8.198140e-03      98.929073
## dim 2 8.874655e-05      1.070927
##      cumulative percentage of variance
## dim 1      98.92907
## dim 2     100.00000
```

```
mean(res.ca$eig[,1]) #Kaiser: take as many dimensions as eigenvalue > mean of eigenvalues
```

```
## [1] 0.004143443
```

```
#En una taula de correspondències simples podem tenir màxim tantes
#dimensions com categories d'una variable menys 1!
#f.season te 3 categories -> -1 -> 2 dimensions!!
```

```
#La inèrcia total ens indica com de relacionades estan les dues variables,
```

```
#com mes proxím el valor a 0, menys relacionades estan!
sum(res.ca$eig[,1])
```

```
## [1] 0.008286887
```

```
#Coordenades:
```

```
#res.ca$row #files son la f.season!
```

```
#res.ca$col #columnes son la duration!
```

El segon cas en el qual obtenim una relacio significativa (p valor = $1.203e-07$) es el valor de l'euribor **f.euribor3m**, el qual es un indicador trimestral. Comparant el profile de la taula de contingencia de proporcions per fila amb el profile marginal de la duracio veiem com hi ha un 30.8% de trucades amb duracions molt curtes quan l'euribor te un valor alt, respecte un 25% de trucades amb duracions sense tenir en compte la fluctuacio de l'indicador. De la mateixa manera, quan el valor de l'euribor es baix, hi ha major % de trucades que acaben amb duracions relativament altes. Si comparem el profile de la taula de contingencia de proporcions per columna amb el profile marginal de f.euribor3m veiem com el 22.8% de trucades es realitzen amb un euribor molt alt, i en canvi un 28.0% de trucades corresponen a un euribor alt i a duracions molt curtes. A mes, mirant el mapa de factors podem veure aquesta mateixa informacio de manera grafica. Si ens centrem en la 1a dimensio del grafic (que representa un 89% de la variancia del conjunt de les dades), podem observar com hi ha una tendencia similar en ambdues variables (tot i que les duracions extremadament llargues trenquen una molt bona correlacio del 1r eix). Tambe veiem com les categories f.euribor3m-(4.96,5] i f.duration-[5,101] estan molt relacionades ja que es troben molt proximes en el mapa de factors; així com tambe les categories f.euribor3m-[0.635,1.33] i f.duration-(177,316]. (Kaiser: take as many dimensions as eigenvalue > mean of eigenvalues, el que equival a agafar 1 sola dimensio).

```
chisq.test( table( df$f.euribor3m, df$f.duration) )
```

```
##
```

```
## Pearson's Chi-squared test
```

```
##
```

```
## data: table(df$f.euribor3m, df$f.duration)
```

```
## X-squared = 49.745, df = 9, p-value = 1.203e-07
```

```
#Row/Column profile
```

```
prop.table( table(df$f.euribor3m, df$f.duration), 1 ) #1->per files
```

```
##
```

```
## f.duration-[5,101] f.duration-(101,177]
```

```
## f.euribor3m-[0.635,1.33] 0.2129187 0.2432217
```

```
## f.euribor3m-(1.33,4.86] 0.2312415 0.2530696
```

```
## f.euribor3m-(4.86,4.96] 0.2619469 0.2345133
```

```
## f.euribor3m-(4.96,5] 0.3080986 0.2658451
```

```
##
```

```
## f.duration-(177,316] f.duration-(316,1.58e+03]
```

```
## f.euribor3m-[0.635,1.33] 0.2775120 0.2663477
```

```
## f.euribor3m-(1.33,4.86] 0.2694407 0.2462483
```

```
## f.euribor3m-(4.86,4.96] 0.2353982 0.2681416
```

```
## f.euribor3m-(4.96,5] 0.2095070 0.2165493
```

```
#Marginal Row/Column profile
```

```
prop.table( table(df$f.duration)) #1->per files
```

```
##
```

```
## f.duration-[5,101] f.duration-(101,177]
```

```
## 0.2511031 0.2492980
```

```
## f.duration-(177,316] f.duration-(316,1.58e+03]
```

```
## 0.2501003 0.2494986
```

```
prop.table( table(df$f.euribor3m, df$f.duration), 2 ) #2->per columnes
```

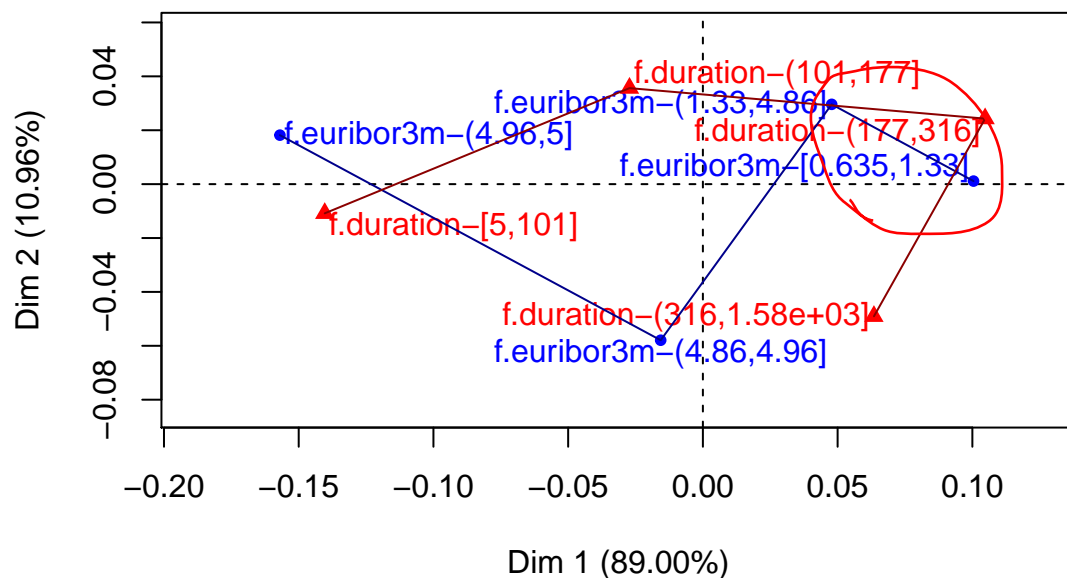
```
##
##               f.duration-[5,101] f.duration-(101,177]
## f.euribor3m-[0.635,1.33]          0.2132588          0.2453741
## f.euribor3m-(1.33,4.86]          0.2707668          0.2984714
## f.euribor3m-(4.86,4.96]          0.2364217          0.2131939
## f.euribor3m-(4.96,5]             0.2795527          0.2429606
##
##               f.duration-(177,316] f.duration-(316,1.58e+03]
## f.euribor3m-[0.635,1.33]          0.2790698          0.2684887
## f.euribor3m-(1.33,4.86]          0.3167602          0.2901929
## f.euribor3m-(4.86,4.96]          0.2133119          0.2435691
## f.euribor3m-(4.96,5]             0.1908581          0.1977492
```

```
prop.table( table(df$f.euribor3m)) #2->per columnes
```

```
##
## f.euribor3m-[0.635,1.33] f.euribor3m-(1.33,4.86] f.euribor3m-(4.86,4.96]
##               0.2515042               0.2940233               0.2266346
## f.euribor3m-(4.96,5]
##               0.2278379
```

```
res.ca<-CA( table( df$f.euribor3m, df$f.duration) )
lines(res.ca$row$coord[,1], res.ca$row$coord[,2], col="darkblue")
lines(res.ca$col$coord[,1], res.ca$col$coord[,2], col="darkred")
```

CA factor map



```
attributes(res.ca); res.ca$eig #valors eig no normalitzats!
```

```
## $names
## [1] "eig" "call" "row" "col" "svd"
```



```
##
## $class
## [1] "CA"    "list"

##          eigenvalue percentage of variance
## dim 1 8.879478e-03      88.99940237
## dim 2 1.093876e-03      10.96396899
## dim 3 3.654443e-06       0.03662864
##          cumulative percentage of variance
## dim 1      88.99940
## dim 2      99.96337
## dim 3     100.00000

mean(res.ca$eig[,1]) #Kaiser: take as many dimensions as eigenvalue > mean of eigenvalues

## [1] 0.00332567

#En una taula de correspondencies simples podem tenir maxims tantes
#dimensions com categories d'una variable menys 1!
#f.euribor3m te 4 categories -> -1 -> 3 dimensions!!

#La inercia total ens indica com de relacionades estan les dues variables,
#com mes proxims el valor a 0, menys relacionades estan!
sum(res.ca$eig[,1])

## [1] 0.009977009

#A vegades va be eliminar algunes categories amb pocs individus d'una variable
#per a poder veure millor les possibles relacions!
```

PRINCIPAL COMPONENT ANALYSIS (PCA)

Primerament realitzem un PCA sobre les variables continues de la nostra mostra de dades, on la variable “duration” ha de ser suplementaria, ja que es tracta de la variable target!

Eigenvalues and dominant axes

El summary ens permet veure els 9 diferents eigenvalues obtinguts amb aquest PCA, amb les seves dades corresponents de percentatges de variancia del conjunt de les dades que representen. Segons el criteri de Kaiser, que diu que s’han de descartar les dimensions amb valors eig normalitzats per sota d’1, hauriem d’agafar les 3 primeres dimensions per a una bona representació del conjunt de dades. Essent flexibles amb el criteri de Kaiser, podríem agafar també la quarta dimensió, la qual té una variancia del 0.9656, amb un valor molt proper a 1. La incorporació d’aquesta nova dimensió ens donaria una variancia acumulada del 81%, obtenint d’aquesta manera una variancia acumulada per sobre el 80%. Per últim, si ens basem en la regla del colze (llegurament subjectiva), i l’apliquem sobre el gràfic dels eigenvalues i %variancies obtingut amb les llibreries *ggplot*, hauríem d’agafar les 3 primeres dimensions.

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

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

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

# PCA:
res.pca<-PCA( df[, vars_con], quanti.sup=2, graph=FALSE) #duration com a suplementaria
```

```
#nb.dec: number of decimal printed
#ncp: number of dimensions printed
summary(res.pca, nb.dec=2, ncp=5, nbind=0)
```

```
##
## Call:
## PCA(X = df[, vars_con], quanti.sup = 2, graph = FALSE)
##
##
## Eigenvalues
##
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7
## Variance	3.90	1.37	1.05	0.97	0.86	0.43	0.39
## % of var.	43.34	15.21	11.67	10.73	9.53	4.75	4.38
## Cumulative % of var.	43.34	58.55	70.22	80.95	90.48	95.24	99.61

```
##
```

	Dim.8	Dim.9
## Variance	0.02	0.01
## % of var.	0.27	0.12
## Cumulative % of var.	99.88	100.00

```
##
## Variables
##
```

	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2
## age	-0.02	0.01	0.00	0.35	8.73	0.12	0.67	42.16	0.44
## campaign	0.20	1.03	0.04	0.01	0.01	0.00	-0.35	11.85	0.12
## pdays	0.43	4.84	0.19	-0.71	37.11	0.51	0.32	9.77	0.10
## previous	-0.59	9.05	0.35	0.55	21.75	0.30	-0.32	9.71	0.10
## emp.var.rate	0.97	23.96	0.93	0.17	2.19	0.03	-0.09	0.75	0.01
## cons.price.idx	0.75	14.49	0.57	0.25	4.49	0.06	-0.25	5.88	0.06
## cons.conf.idx	0.16	0.68	0.03	0.56	23.16	0.32	0.46	19.73	0.21
## euribor3m	0.97	23.91	0.93	0.19	2.54	0.03	-0.01	0.01	0.00
## nr.employed	0.93	22.02	0.86	0.01	0.02	0.00	-0.04	0.14	0.00

```
##
```

	Dim.4	ctr	cos2	Dim.5	ctr	cos2
## age	0.39	15.54	0.15	0.53	33.33	0.29
## campaign	0.89	81.66	0.79	-0.22	5.44	0.05
## pdays	0.08	0.68	0.01	-0.05	0.31	0.00
## previous	-0.05	0.29	0.00	0.14	2.42	0.02
## emp.var.rate	-0.06	0.43	0.00	0.07	0.54	0.00
## cons.price.idx	-0.07	0.52	0.00	0.26	7.62	0.07
## cons.conf.idx	-0.03	0.09	0.00	-0.66	50.14	0.43
## euribor3m	-0.07	0.54	0.01	-0.02	0.05	0.00
## nr.employed	-0.05	0.25	0.00	0.04	0.15	0.00

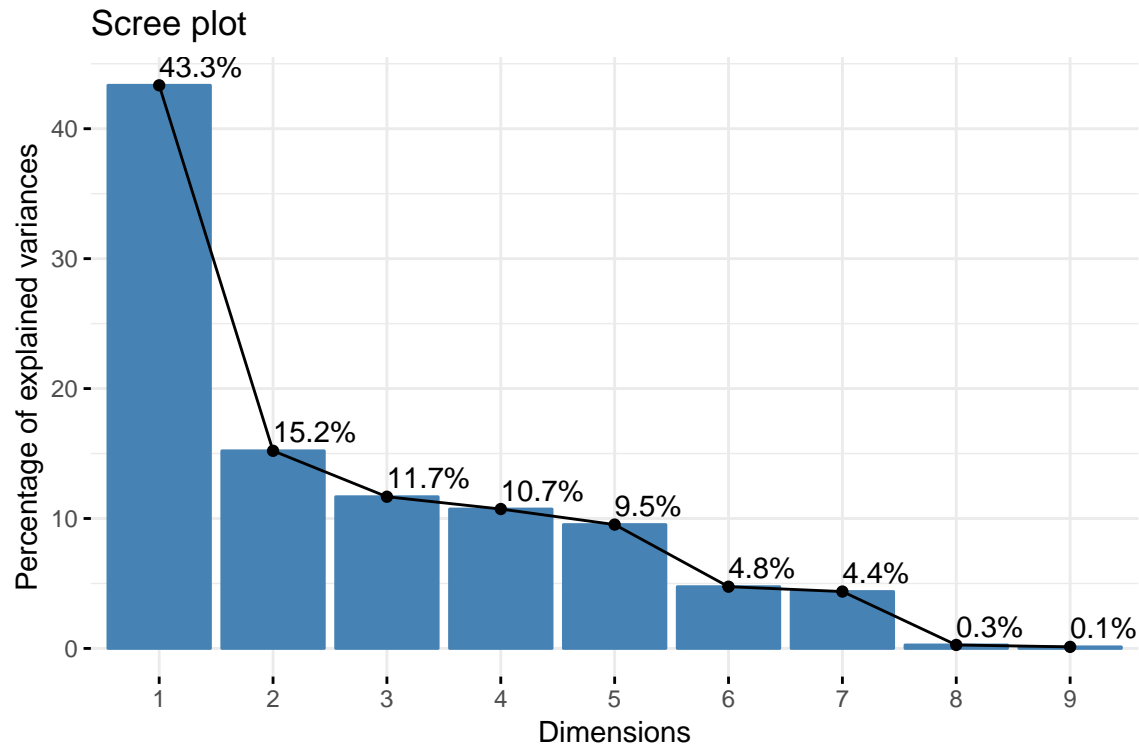
```
##
## Supplementary continuous variable
##
```

	Dim.1	cos2	Dim.2	cos2	Dim.3	cos2	Dim.4	cos2
## duration	-0.03	0.00	0.00	0.00	0.00	0.00	-0.07	0.00

```
##
## duration
```

	Dim.5	cos2
## duration	0.04	0.00

```
#GGPLOT: Use modern ggplot facilities per la regola de l'ultim colze:
#at some point the marginal gain will drop, giving an angle in the graph
fviz_eig(res.pca, addlabels=TRUE)
```



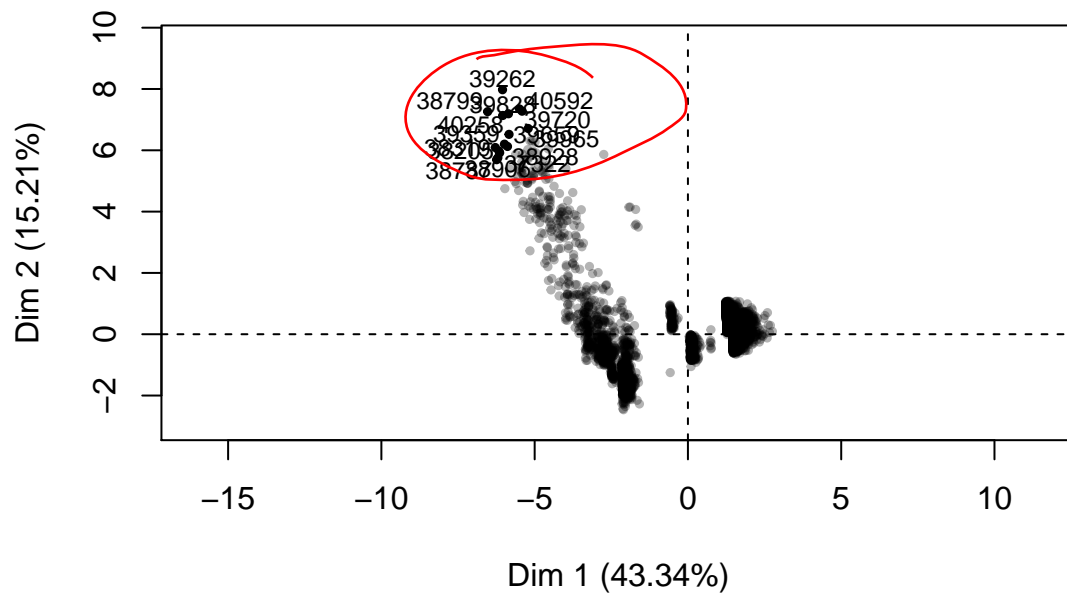
Individuals point of view

Pintem primer el mapa de factors dels individus de les dues primeres dimensions (70% de la variancia del conjunt de dades) i etiquetem amb el numero d'individu els 15 mes contributius. A continuacio mostrem per a cadascun dels dos primers eixos, les coordenades i el registre complet d'aquests 3 individus mes contributius. Es fa exactament el mateix amb els individus mes ben representats (\cos^2) en les dues primeres dimensions.

#nomes pinta les etiquetes dels 15 individus mes contributius!

```
plot(res.pca, choix="ind", cex=0.75, col.ind="black", select="contrib 15", title="Factor map - 15 indiv.
```

Factor map – 15 individuos mes contributius



```
#2 individus mes contributius al 1r eix:
```

```
contrib<-sort(res.pca$ind$contrib[,1], decreasing=TRUE)[1:2]; contrib
```

```
##      38799      38319
## 0.2194923 0.2029850
```

```
df[c(names(contrib)), ]
```

```
##      age      job      marital      education
## 38799  62  job-housemaid marital-married education-university.degree
## 38319  37  job-blue-collar marital-married      education-basic.6y
##      default      housing      loan      contact      month
## 38799 default-no housing-yes loan-no contact-cellular month-nov
## 38319 default-no housing-no loan-no contact-cellular month-oct
##      day_of_week duration campaign pdays previous      poutcome
## 38799 day_of_week-4thu      237      1      3      3 poutcome-success
## 38319 day_of_week-4thu      128      2      6      3 poutcome-failure
##      emp.var.rate cons.price.idx cons.conf.idx euribor3m nr.employed
## 38799      -3.4      92.649      -30.1      0.714      5017.5
## 38319      -3.4      92.431      -26.9      0.740      5017.5
##      y num_missings num_outliers num_errors      f.season
## 38799 y-no      0      0      0 season-autumnwinter
## 38319 y-yes      0      0      0 season-autumnwinter
##      minutes      f.age      f.duration      f.campaign
## 38799 3.950000 f.age-(47,87] f.duration-(177,316] f.campaign-[0,2]
## 38319 2.133333 f.age-(32,38] f.duration-(101,177] f.campaign-[0,2]
##      f.pdays      f.previous      f.emp.var.rate
## 38799 f.pdays-sometime f.previous-some f.emp.var.rate-[-Inf,0]
## 38319 f.pdays-sometime f.previous-some f.emp.var.rate-[-Inf,0]
##      f.cons.price.idx      f.cons.conf.idx
```

```
## 38799 f.cons.price.idx-[92.2,93.1] f.cons.conf.idx-(-36.4,-26.9]
## 38319 f.cons.price.idx-[92.2,93.1] f.cons.conf.idx-(-36.4,-26.9]
##          f.euribor3m          f.nr.employed
## 38799 f.euribor3m-[0.635,1.33] f.nr.employed-[4.96e+03,5.1e+03]
## 38319 f.euribor3m-[0.635,1.33] f.nr.employed-[4.96e+03,5.1e+03]
```

```
#2 individus mes contributius al 2n eix:
```

```
contrib<-sort(res.pca$ind$contrib[,2], decreasing=TRUE)[1:2]; contrib
```

```
##      39262      39720
## 0.9322541 0.7912329
```

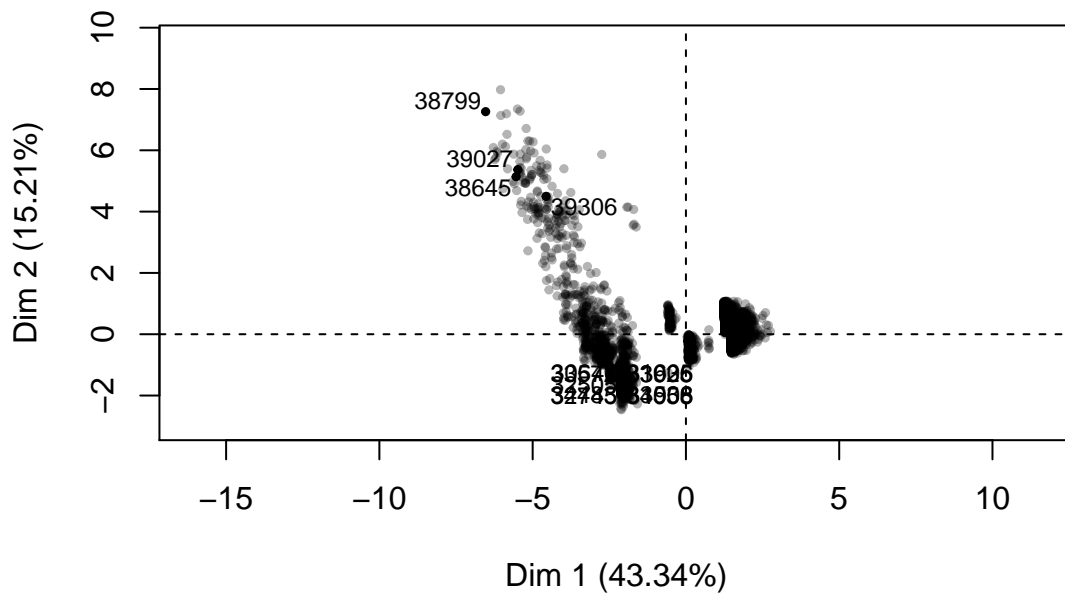
```
df[c(names(contrib)), ]
```

```
##      age      job      marital      education      default
## 39262  80 job-retired marital-married education-basic.4y default-no
## 39720  80 job-retired marital-married education-basic.4y default-no
##      housing      loan      contact      month      day_of_week
## 39262 housing-no loan-no contact-cellular month-mar day_of_week-5fri
## 39720 housing-no loan-no contact-cellular month-may day_of_week-1mon
##      duration campaign pdays previous      poutcome emp.var.rate
## 39262      213      3      6      4 poutcome-success      -1.8
## 39720      382      1      3      3 poutcome-success      -1.8
##      cons.price.idx cons.conf.idx euribor3m nr.employed      y
## 39262      93.369      -34.8      0.649      5008.7 y=yes
## 39720      93.876      -40.0      0.697      5008.7 y=yes
##      num_missings num_outliers num_errors      f.season minutes
## 39262      0      0      0 season-spring 3.550000
## 39720      0      0      0 season-spring 6.366667
##      f.age      f.duration      f.campaign
## 39262 f.age-(47,87] f.duration-(177,316] f.campaign-(2,5]
## 39720 f.age-(47,87] f.duration-(316,1.58e+03] f.campaign-[0,2]
##      f.pdays      f.previous      f.emp.var.rate
## 39262 f.pdays-sometime f.previous-some f.emp.var.rate-[-Inf,0]
## 39720 f.pdays-sometime f.previous-some f.emp.var.rate-[-Inf,0]
##      f.cons.price.idx      f.cons.conf.idx
## 39262 f.cons.price.idx-(93.1,93.7] f.cons.conf.idx-(-36.4,-26.9]
## 39720 f.cons.price.idx-(93.7,94] f.cons.conf.idx-(-41.8,-36.4]
##      f.euribor3m      f.nr.employed
## 39262 f.euribor3m-[0.635,1.33] f.nr.employed-[4.96e+03,5.1e+03]
## 39720 f.euribor3m-[0.635,1.33] f.nr.employed-[4.96e+03,5.1e+03]
```

```
#noms pinta les etiquetes dels 15 individus mes ben representats!
```

```
plot(res.pca, choix="ind", cex=0.75, col.ind="black", select="cos2 15", title="Factor map - 15 individus")
```

Factor map – 15 individus mes ben representats



```
#2 individus mes ben representats al 1r eix:
repr<-sort(res.pca$ind$cos2[,1], decreasing=TRUE)[1:2]; repr
```

```
##      18385      13817
## 0.8444095 0.8442091
```

```
#df[c(names(repr)), ]
#2 individus mes ben representats al 2n eix:
repr<-sort(res.pca$ind$cos2[,2], decreasing=TRUE)[1:2]; repr
```

```
##      39431      39350
## 0.5706147 0.5679583
```

```
#df[c(names(repr)), ]
```

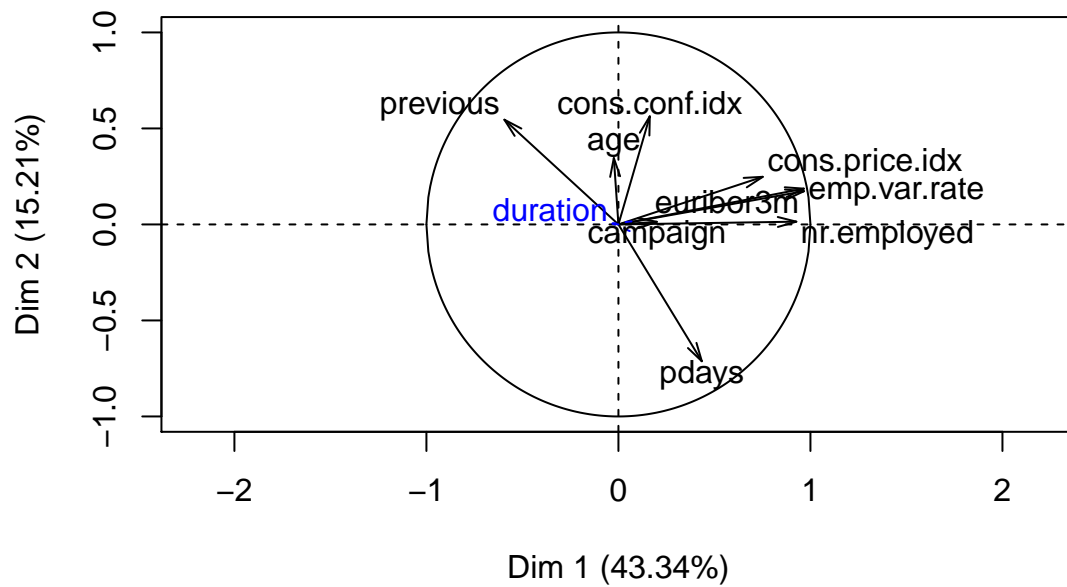
Interpretation of the PCA

En el mapa de factors de les variables (2D, primeres dues dimensions) es pot observar en blau la variable “duration” com a suplementaria, la qual surt quasi centrada, el que vol dir que les variables vars_ con utilitzades en el PCA no ens ajuden a dir res o predir el valors de la variable target. La variable previous (numero de contactes en campanyes antigues) esta relacionada inversament amb pdays (dies que feia que no es trucava el client per altres campanyes), ja que com es pot veure en el grafic, ambdues fletxes apunten oposadament. Sembla ser tambe que tots els indicadors socioeconomics (a excepcio una mica de cons.conf.idx) apunten en la mateixa direccio, el que vol dir que estan relacionats entre ells i contribueixen d’una manera similar als eixos. Es pot veure el % de contribucio de les variables a les tres primeres dimensions mitjancant una eina de la llibreria ggplot, on es veu com basicament els indicadors socioeconomics i les variables “previous”, “campaign” i “age” son les variables numeriques que mes han contribuït. El summary ens permet veure tambe de forma numerica la contribucio (ctr) de cadascuna de les variables en els 4 primers eixos, aixi com la qualitat de la representacio (cos2) de les mateixes en cadascun dels eixos. Si be s’acaben de descriure les variables mes contributives, les que estan millor representades (cos2 mes proxim a 1) en el primer eix son euribor3m, emp.var.rate i nr.employed; i en el segon eix son pdays, previous i cons.conf.idx. Un l’ultim mapa

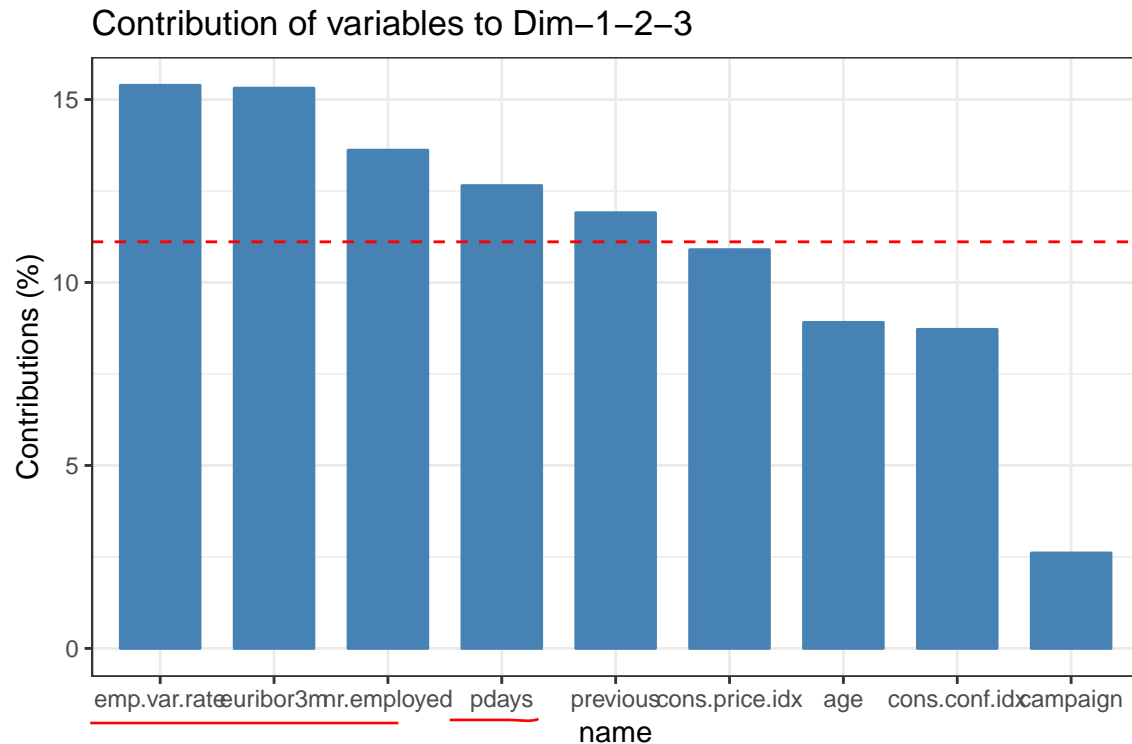
de factors de variables observem els eixos de les dimensions 3 i 4, les quals representen aproximadament un 11% de la variància de les dades cadascuna.

```
# PCA:  
plot.PCA(res.pca, choix = c("var")) #variables factor map
```

Variables factor map (PCA)



```
#GGPLOT contribution of variables  
#fviz_pca_var(res.pca)  
fviz_contrib(res.pca, choice="var", axes=1:3)+theme_bw()
```

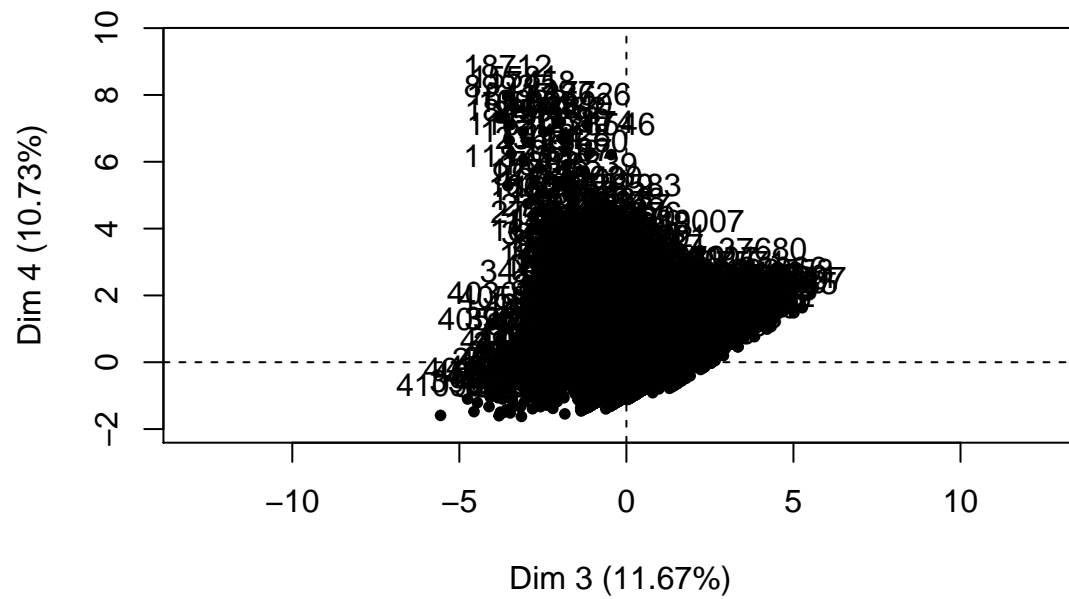


```
summary(res.pca, nb.dec=2, ncp=2, nbind=0)
```

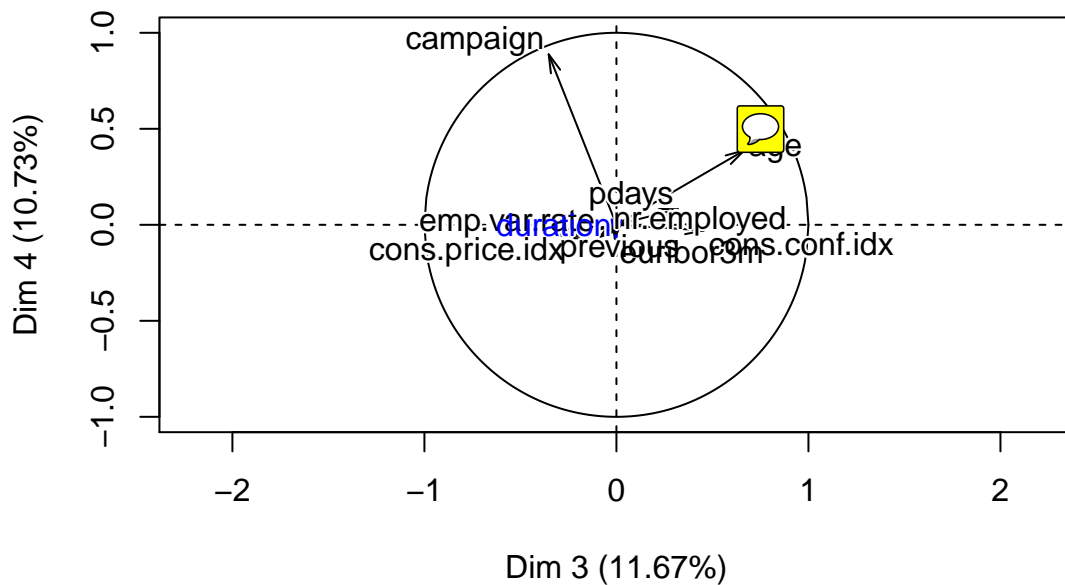
```
##
## Call:
## PCA(X = df[, vars_con], quanti.sup = 2, graph = FALSE)
##
##
## Eigenvalues
##
##          Dim.1  Dim.2  Dim.3  Dim.4  Dim.5  Dim.6  Dim.7
## Variance      3.90   1.37   1.05   0.97   0.86   0.43   0.39
## % of var.     43.34  15.21  11.67  10.73   9.53   4.75   4.38
## Cumulative % of var. 43.34  58.55  70.22  80.95  90.48  95.24  99.61
##
##          Dim.8  Dim.9
## Variance      0.02   0.01
## % of var.      0.27   0.12
## Cumulative % of var. 99.88 100.00
##
## Variables
##
##          Dim.1  ctr  cos2  Dim.2  ctr  cos2
## age          | -0.02 0.01 0.00 | 0.35 8.73 0.12 |
## campaign     | 0.20 1.03 0.04 | 0.01 0.01 0.00 |
## pdays        | 0.43 4.84 0.19 | -0.71 37.11 0.51 |
## previous     | -0.59 9.05 0.35 | 0.55 21.75 0.30 |
## emp.var.rate | 0.97 23.96 0.93 | 0.17 2.19 0.03 |
## cons.price.idx | 0.75 14.49 0.57 | 0.25 4.49 0.06 |
## cons.conf.idx | 0.16 0.68 0.03 | 0.56 23.16 0.32 |
## euribor3m    | 0.97 23.91 0.93 | 0.19 2.54 0.03 |
## nr.employed  | 0.93 22.02 0.86 | 0.01 0.02 0.00 |
##
```



```
#veure eixos 3 i 4:
res.pca<-PCA(df[, vars_con], quanti.sup=2, axes=3:4)
```



Variables factor map (PCA)

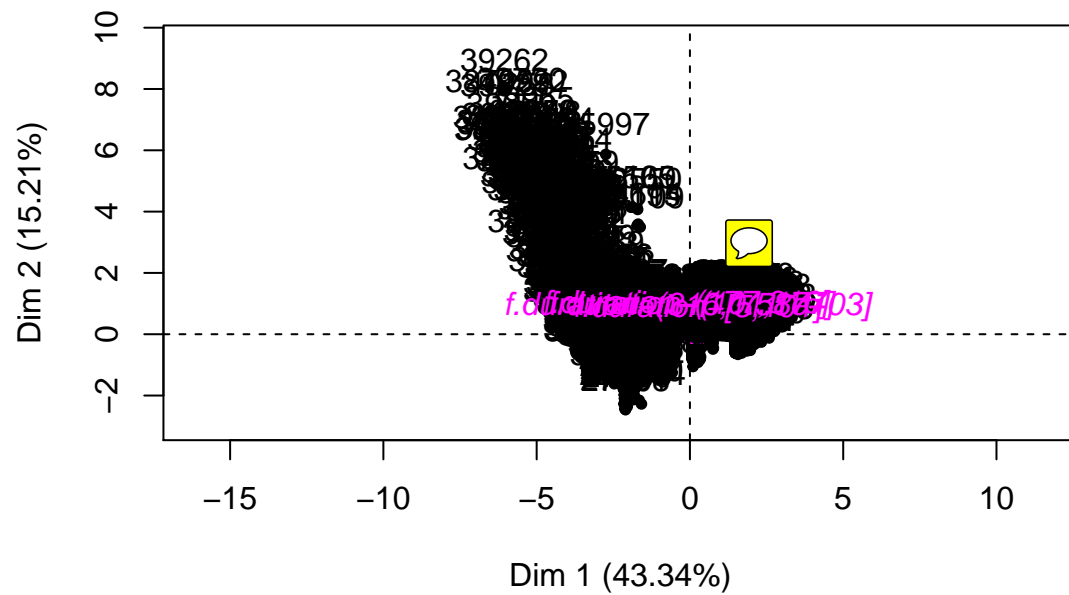


Com que es difícil extreure conclusions a partir dels noms de les variables numèriques, en els gràfics següents es pot observar el posicionament de les dues variables suplementàries “f.duration” i “f.euribor3m” en el mapa de factors dels individus. En el cas de f.duration, si es mira al mapa de factors ampliat on només es representa aquesta variable, sembla ser que té una tendència a créixer cap al 2n i 3r quadrant compost per les dues primeres dimensions. En el cas de f.euribor3m, de la mateixa manera, també es pot trobar una progressió de les dades al llarg del primer eix, ja que el valor de l'euribor3m augmenta amb el valor de la primera dimensió de les dades.

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

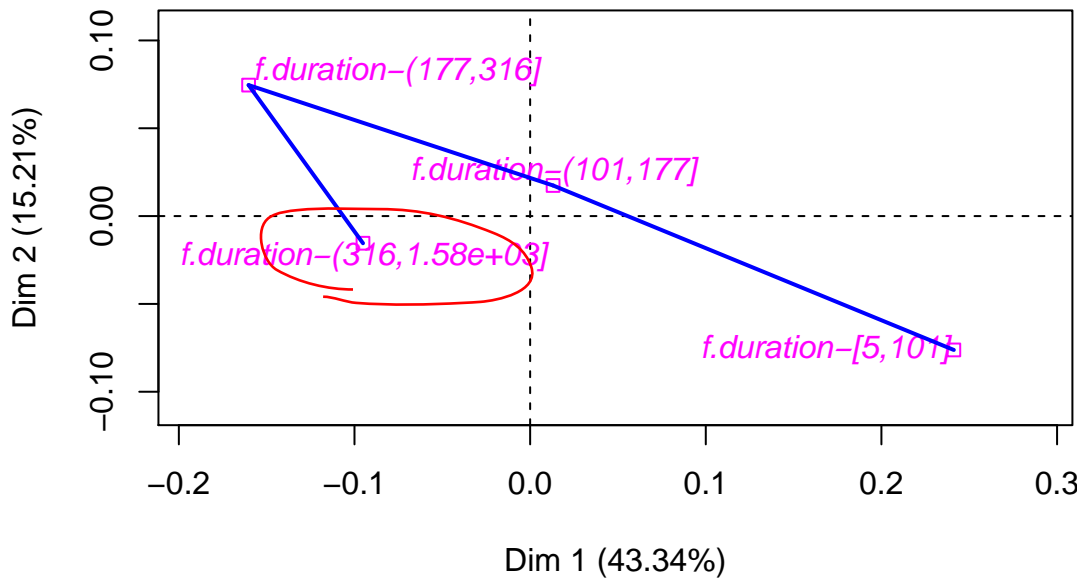
#ara afegim dues variables suplementaries!
res.pca<-PCA( df[, c("f.duration", vars_con) ], quanti.sup=3, quali.sup=1, graph = FALSE )
plot.PCA(res.pca, choix = c("ind") )
```

Individuals factor map (PCA)



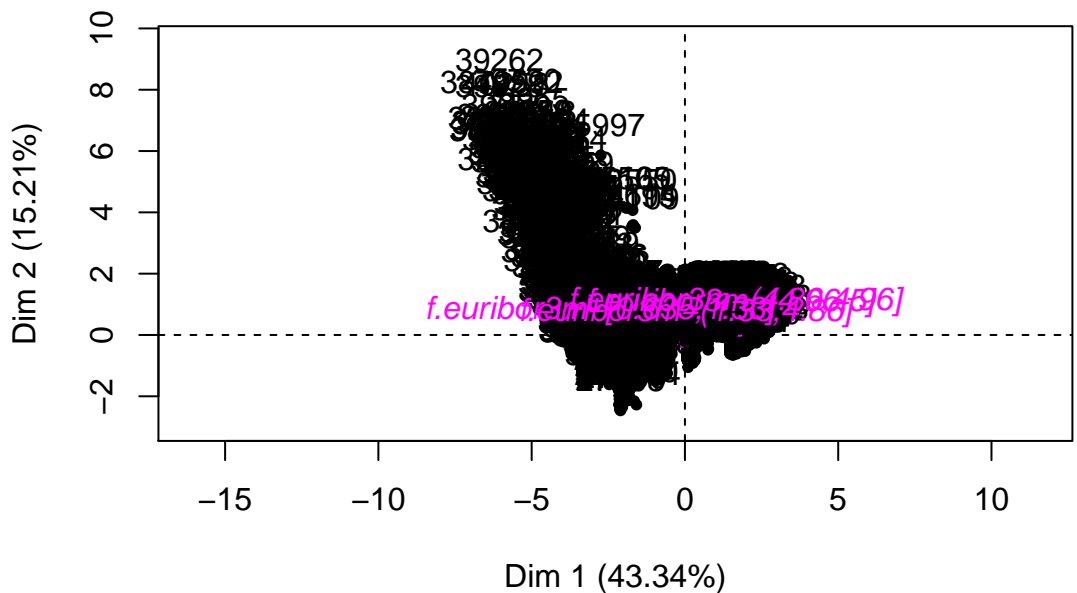
```
#Unir punts de variables suplementaries en el factor map per a una bona representacio:
plot.PCA(res.pca, choix="ind", invisible="ind")
lines(res.pca$quali.sup$coord[1:2, 1:2], col="blue", lwd="2")
lines(res.pca$quali.sup$coord[2:3, 1:2], col="blue", lwd="2")
lines(res.pca$quali.sup$coord[3:4, 1:2], col="blue", lwd="2")
```

Individuals factor map (PCA)



```
res.pca<-PCA( df[, c("f.euribor3m", vars_con) ], quanti.sup=3, quali.sup=1, graph = FALSE )
plot.PCA(res.pca, choix = c("ind") )
```

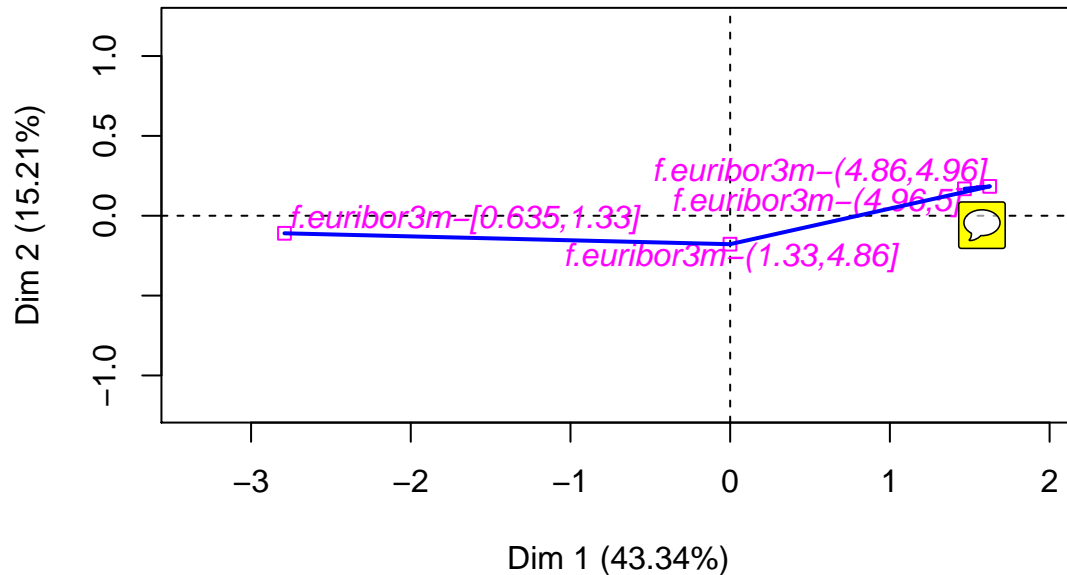
Individuals factor map (PCA)



```
plot.PCA(res.pca, choix="ind", invisible="ind")
lines(res.pca$quali.sup$coord[1:2, 1:2], col="blue", lwd="2")
```

```
lines(res.pca$quali.sup$coord[2:3, 1:2], col="blue", lwd="2")
lines(res.pca$quali.sup$coord[3:4, 1:2], col="blue", lwd="2")
```

Individuals factor map (PCA)



K-Means Classification (Partitioning - Supervised learning)

K-means es un algoritme de clustering que te com a objectiu agrupar les observacions en un determinat nombre de grups o clusters els quals comparteixin característiques similars. Dit d'altra manera, agrupa els individus de manera que els que estan dins un mateix cluster tinguin unes distàncies euclidianes entre ells més petites que respecte els individus d'altres clusters. Cal tenir en compte que com que el paràmetre passat a la crida *kmeans* és el nombre de clusters que s'han d'obtenir (i no el conjunt inicial de centres), es seleccionen aleatoriament un conjunt inicial de *k* centres de cluster. Això vol dir que aquesta selecció aleatòria pot tenir gran influència en el resultat final, el qual serà diferent en cada execució de l'algoritme. Per tal de reduir soroll de les dades innecessari, treballarem només amb les 4 primeres dimensions obtingudes del PCA.

L'objecte retornat per la crida *kmeans* ens permet consultar diferents atributs. L'atribut *withinss* correspon a la suma del quadrat de les distàncies inter-cluster, és a dir, hi ha un valor per a cada cluster. L'atribut *betweenss* és un sol valor mig que correspon a la suma del quadrat de les distàncies inter-cluster. A partir d'aquest valor mig i el total de les distàncies totss podem obtenir el nivell de representació que obtenim només amb els centres de gravetat dels clusters sobre el conjunt de les dades.

```
#only 4 significant axes in order to avoid unnecessary noise
dclu<-res.pca$ind$coord[,1:4]
```

```
#fixed number of clusters (a random set of rows are chosen as the initial centers)
kcla<-kmeans(dclu, 7)
summary(kcla)
```

```
##          Length Class  Mode
## cluster    4986  -none- numeric
```

```
## centers      28  -none- numeric
## totss       1  -none- numeric
## withinss    7  -none- numeric
## tot.withinss 1  -none- numeric
## betweenss   1  -none- numeric
## size        7  -none- numeric
## iter        1  -none- numeric
## ifault      1  -none- numeric
```

```
table(kcla$cluster)
```

```
##
##      1      2      3      4      5      6      7
## 343 1657  883  281 1403  248  171
```

```
attributes(kcla)
```

```
## $names
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
```

```
##
## $class
## [1] "kmeans"
```

```
#INTRA-CLUSTER DISTANCES
```

```
kcla$withinss
```

```
## [1] 483.3356 1482.1199 939.7402 1047.3004 1181.9092 680.7826 845.8692
```

```
#$withinss: is the within cluster sum of squares.
#So it results in a vector with a number for each cluster.
```

```
#INTER-CLUSTER DISTANCES
```

```
kcla$betweenss
```

```
## [1] 29665.91
```

```
#$betweenss: is the between clusters sum of squares.
#In fact it is the mean of distances between cluster centers.
```

```
# Some equalities may help to understand:
```

```
#      $tot.withinss = sum($withinss)
#      $totss = $tot.withinss + $betweenss
```

```
kcla$tot.withinss
```

```
## [1] 6661.057
```

```
kcla$totss
```

```
## [1] 36326.97
```

```
kcla$betweenss/kcla$totss
```

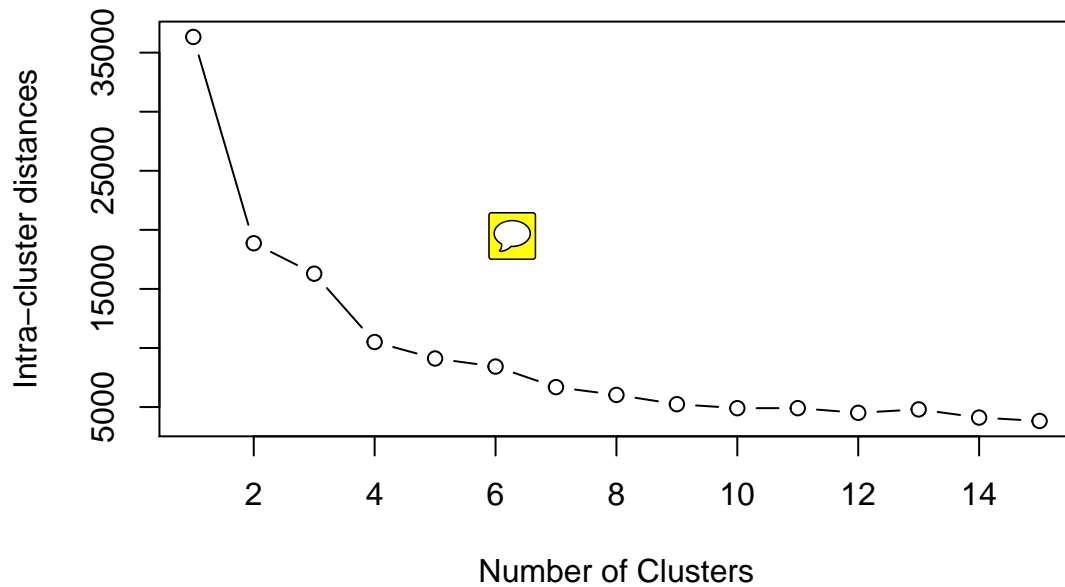
```
## [1] 0.816636
```

```
#noms amb els centres de gravetat estariem obtenint una representacio de les dades del X%
```

Si executem k-means per a diferents valors k podem observar com evolucionen les distancies intra-clusters o el % de representacio obtingut en cada cas. L'execucio d'unes quantes vegades del seguent chunk de codi ens ha ajudat en la seleccio del nombre de clusters k=7.

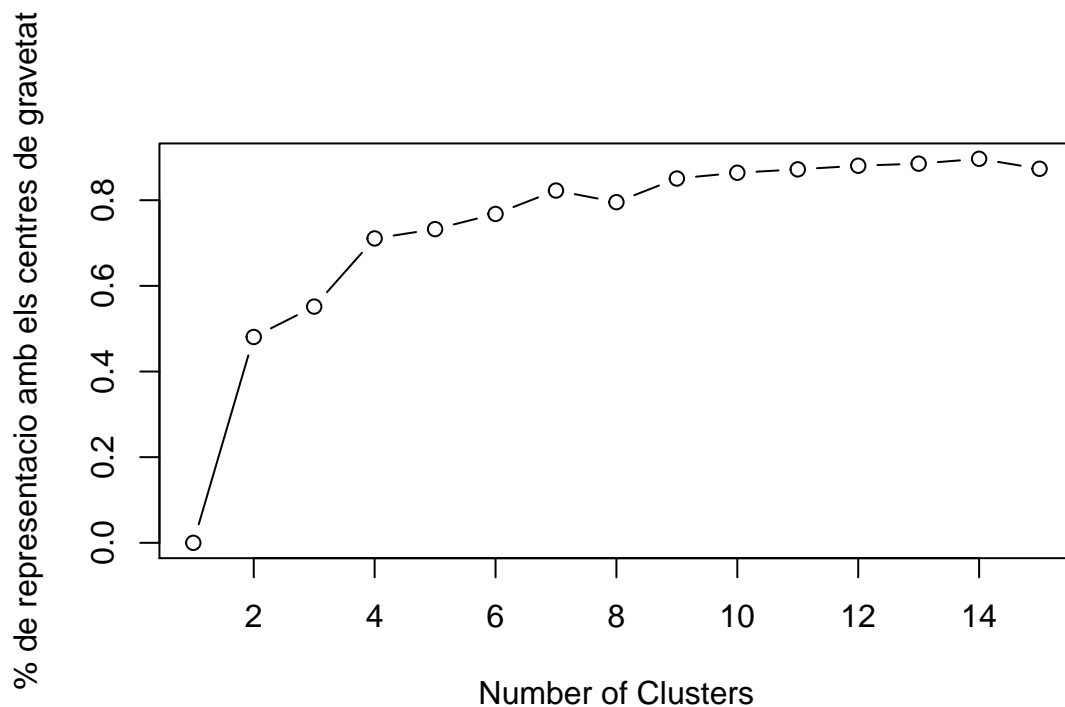
```
wss <- sum(kmeans(dclu,1)$withinss) #k=1
for (i in 2:15) wss[i] <- sum(kmeans(dclu,i)$withinss) #k=2 to 15

plot(1:15, wss, type="b", xlab="Number of Clusters", ylab="Intra-cluster distances")
```



```
km<-kmeans(dclu,1)
repr <- km$betweenss/km$totss #k=1
for (i in 2:15){
  km<-kmeans(dclu,i)
  repr[i] <- km$betweenss/km$totss
}

plot(1:15, repr, type="b", xlab="Number of Clusters", ylab="% de representacio amb els centres de grave")
```



Hierarchical Clustering (Unsupervised learning)

Aquest punt compren la realització d'una clusterització aglomerativa jerarquitzada dels individus. HCPC de la llibreria FactoMineR utilitza les distàncies de entre clusters diferents, per tal de minimitzar la inèrcia inter-cluster.

La primera de les comandes HCPC està comentada ja que requereix l'interacció de l'usuari per a triar per on tallar l'arbre de clusters que es mostra a l'usuari. Després d'estar interactuant amb diferents opcions de nombre de clusters, s'ha decidit finalment agafar `nb.clust = 7`, tal i com es pot veure a continuació. Es mostren els gràfics corresponents a l'arbre el qual s'ha tallat a l'altura de 7 clusters, així com el mapa de factors de les dues primeres dimensions del PCA.

#PCA calculat amb 4 significant axes:

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

```
## [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", "loan", "month", "job", "poutcome", "education", "housing", vars_
```

#res.hcpc<-HCPC(res.pca, order=TRUE)

```
res.hcpc<-HCPC(res.pca, nb.clust=7, order=TRUE, graph=FALSE); res.hcpc
```

```
## **Results for the Hierarchical Clustering on Principal Components**
```

```
##      name
```

```
## 1  "$data.clust"
```

```
## 2  "$desc.var"
```

```
## 3  "$desc.var$quanti.var"
```



```

## 4 "$desc.var$quanti"
## 5 "$desc.var$test.chi2"
## 6 "$desc.axes$category"
## 7 "$desc.axes"
## 8 "$desc.axes$quanti.var"
## 9 "$desc.axes$quanti"
## 10 "$desc.ind"
## 11 "$desc.ind$para"
## 12 "$desc.ind$dist"
## 13 "$call"
## 14 "$call$t"
##   description
## 1  "dataset with the cluster of the individuals"
## 2  "description of the clusters by the variables"
## 3  "description of the cluster var. by the continuous var."
## 4  "description of the clusters by the continuous var."
## 5  "description of the cluster var. by the categorical var."
## 6  "description of the clusters by the categories."
## 7  "description of the clusters by the dimensions"
## 8  "description of the cluster var. by the axes"
## 9  "description of the clusters by the axes"
## 10 "description of the clusters by the individuals"
## 11 "parangons of each clusters"
## 12 "specific individuals"
## 13 "summary statistics"
## 14 "description of the tree"

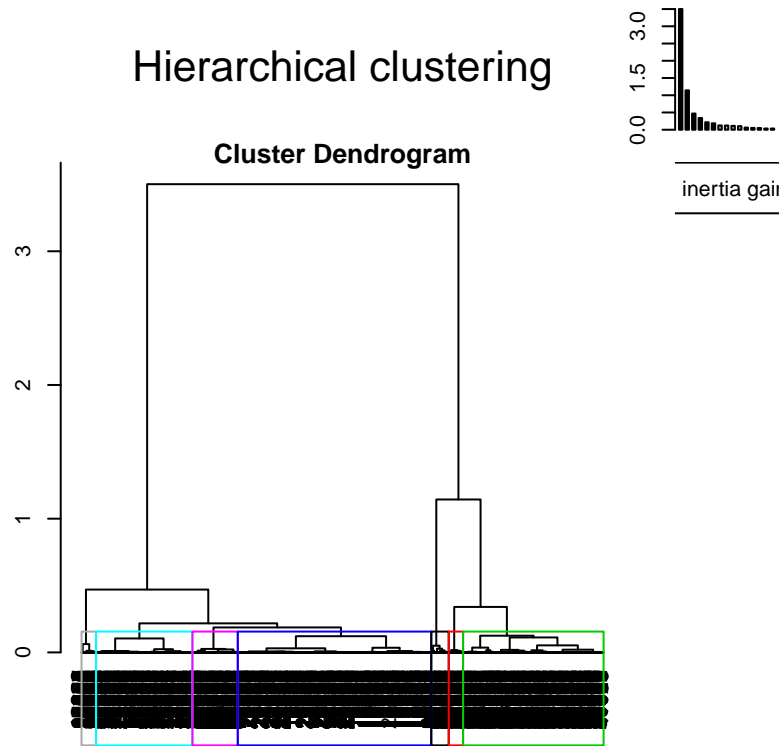
attributes(res.hcpc)

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

plot.HCPC(res.hcpc, choice="tree")

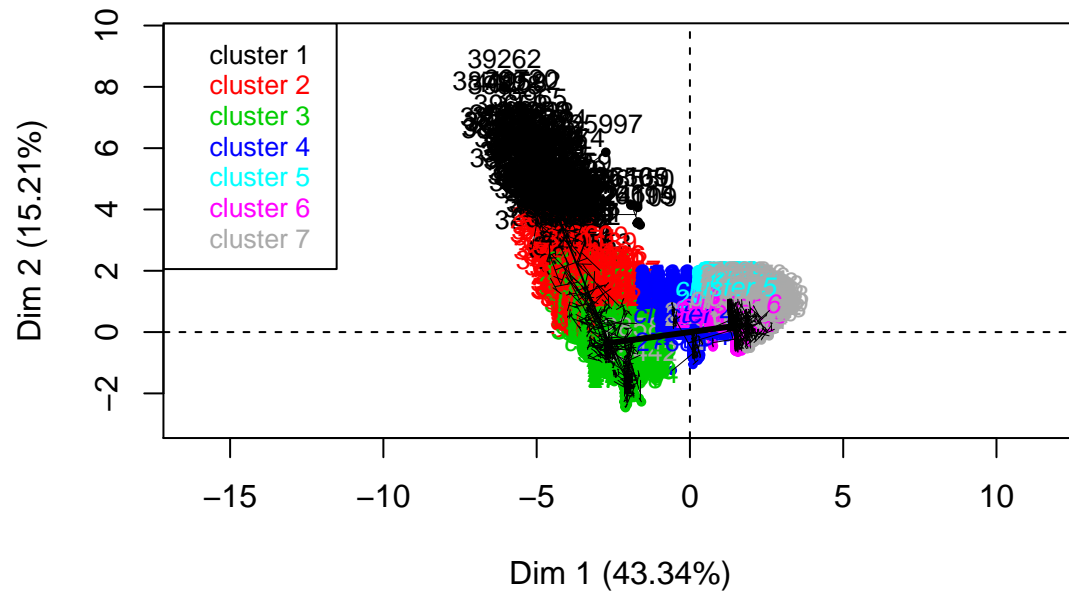
```

Hierarchical clustering

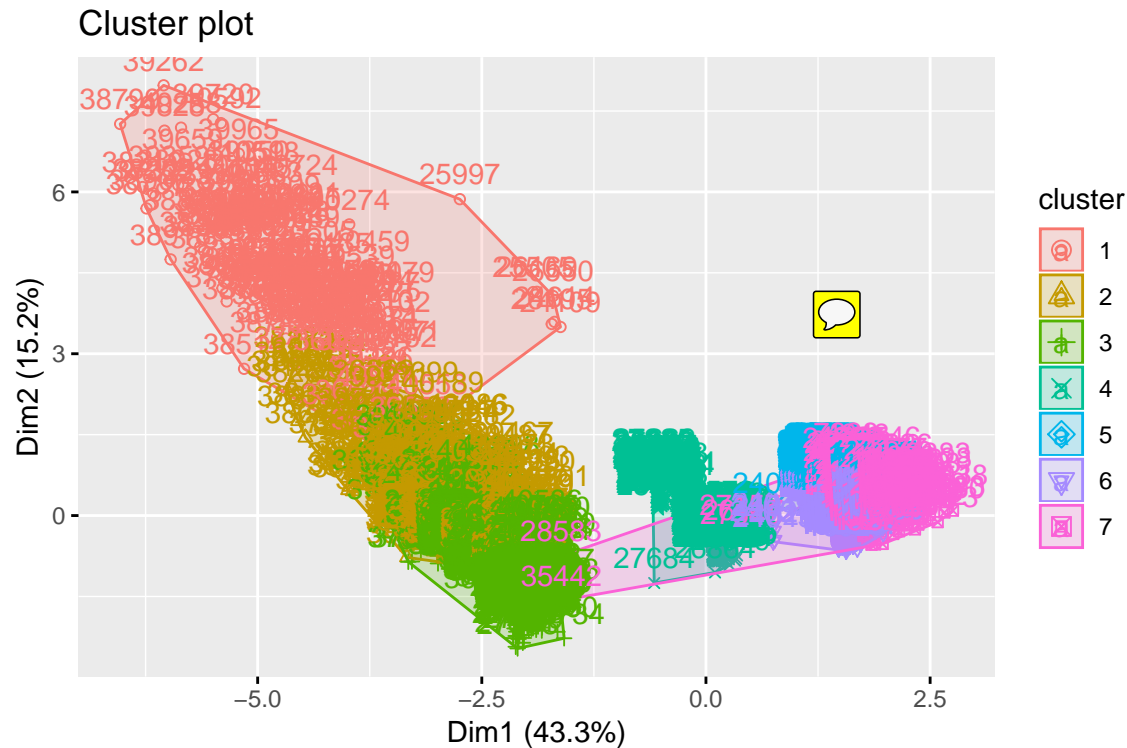


```
plot.HCPC(res.hcpc, choice="map")
```

Factor map



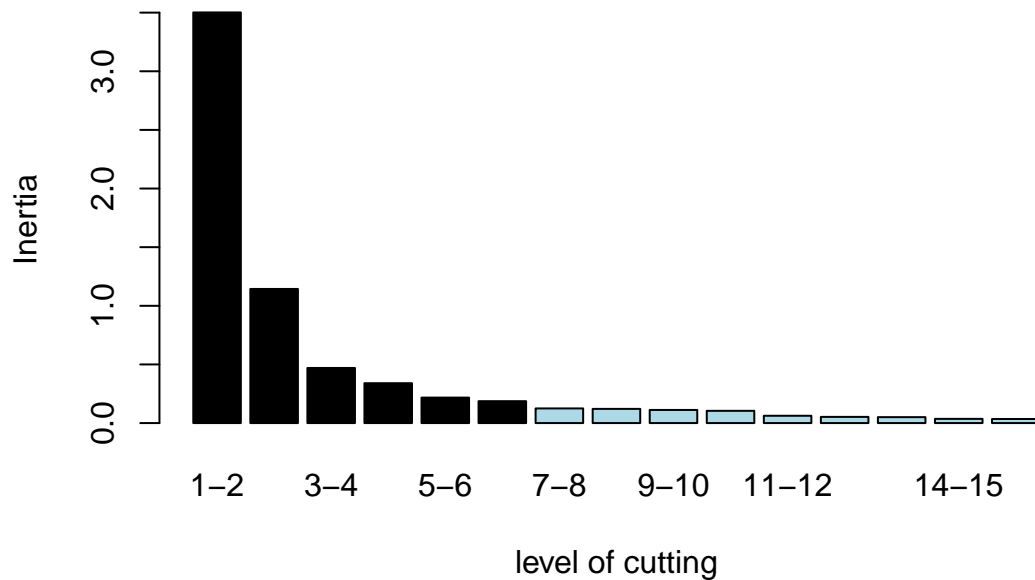
```
fviz_cluster(res.hcpc)
```



El guany (d'1 a 2 clusters, de 2 a 3, de 3 a 4, etc.) d'inèrcia es poden observar en negre a la figura inferior, on també es pot veure la relació entre altres nivells de tall i els seus guanys. Els valors *within* ens mostren la inèrcia dins de cada cluster, que si fos un sol cluster (1) seria la total del conjunt de dades, la qual equival a la suma de tots els guanys d'inèrcies *inert.gain*. Amb la selecció de 6 clusters que hem agafat estariem obtenint una representació de les dades del 80.4%, la qual és molt bona! Finalment, també podem veure la distribució en nombre d'individus en els diferents clusters.

```
plot.HCPC(res.hcpc, choice="bar")
```

Inter-cluster inertia gains



```
res.hcpc$call$t$inert.gain[1:6]
```

```
## [1] 3.5016630 1.1436272 0.4699006 0.3400760 0.2171351 0.1867086
```

```
sum(res.hcpc$call$t$inert.gain)
```

```
## [1] 7.285793
```

```
res.hcpc$call$t$within[1:7]
```

```
## [1] 7.285793 3.784130 2.640503 2.170602 1.830526 1.613391 1.426683
```

```
#X clusters corresponen al %? d'inercia
```

```
sum(res.hcpc$call$t$inert.gain[1:6]) / sum(res.hcpc$call$t$inert.gain) * 100
```

```
## [1] 80.41829
```

```
table(res.hcpc$data.clust$clust) #nombre d'individus en cada cluster
```

```
##
```

```
##      1      2      3      4      5      6      7
## 176  313 1155  430 1268 1441  203
```

Sobre les variables categoriques que han estat incloses com a suplementaries

Aquí analitzarem la representació i relació de les variables categoriques que s'han inclòs com a suplementaries en el PCA destinat a la clusterització jeràrquica. Aquestes variables han estat les següents: “y”, “loan”, “month”, “job”, “poutcome”, “education” i “housing”.

En un primer output basat en el *test chi2* veiem ràpidament quines d'aquestes variables discretes suplementaries estan significativament ($p\text{-value} < 0.05$) relacionades globalment amb la partició de clusters establerta. Les variables “month” i “poutcome” tenen un p-valor tan baix que R no el pot representar i ens el marca com a

0; de la mateixa manera, les variables “y(target)”, “job”, “education” i “housing” estan també relacionades amb el particionament seleccionat.

En un segon output desglosat per numero de cluster veiem la sobrerrepresentació i infrarepresentació de diferents categories en els diferents clusters, ajudant així a caracteritzar els diferents grups formats. La presentació de les dades és la mateixa que la mostrada en el catdes però amb el numero de cluster com a variable fixada, per tant seran interpretades de la mateixa manera descrita en el primer deliverable. En una primera **descripció general**, podem veure com els clusters 1 i 2, i en menor mesura el 3, tenen una sobrerrepresentació d'acceptació de producte financer. Pel que fa a la resta de clusters, tenen una infrarepresentació del mateix. El 71% d'acceptacions del producte financer es troben als 3 primers clusters, mentre que el 29% restant es troben als altres 4. Cada cluster a més, té un lleuger esbiaix en certes categories professionals i d'educació. El **primer cluster** es troba caracteritzat per una immensa sobrerrepresentació dels clients amb acceptació d'una campanya anterior (poutcome-success és un 86% dins el cluster 1 respecte un 3% global); això es tradueix també en que el 97% d'individus que havien acceptat anteriorment un producte estan situats en aquest cluster. El **segon cluster** es troba sobrerrepresentat pels individus retirats (27% respecte un 4% global), el 41% dels quals pertanyen a aquest cluster. També trobem que més de la meitat de trucades realitzades als mesos d'octubre, setembre i desembre són classificades dins del cluster numero 2. A més a més, el segon i en especial el **tercer cluster** (59% dels poutcome-failure pertanyen a aquest) es troben força sobrerrepresentats per individus que anteriorment no van acceptar una campanya (però aquesta sí). També inclouen una sobrerrepresentació d'individus housing=yes. Aquest tercer cluster compren a més una sobrerrepresentació dels mesos d'abril, maig i març. El 99.7% d'individus del **quart cluster** van ser contactats el mes de novembre, el que es tradueix en una immensa sobrerrepresentació en aquest grup, que a més no va acceptar el producte financer (97.0%). El **cinque cluster**, **sise cluster** i **sete cluster** estan formats íntegrament per individus dels quals no es té informació d'èxit o fracàs en contactes de campanyes anteriors. El cinquè cluster té una sobrerrepresentació dels mesos agost i maig; mentre que el sise i sete clusters la tenen dels mesos juliol i juny. Aquests dos últims clusters tenen a més una sobrerrepresentació housing=no (i infra de housing=yes), essent el sete cluster lleugerament diferent del sise pel fet de contenir també una sobrerrepresentació de l'agost i una major infrarepresentació d'acceptació del producte financer.

```
# Factors globally related to clustering partition:
```

```
res.hcpc$desc.var$test.chi2
```

```
##                p.value df
## month          0.000000e+00 54
## poutcome       0.000000e+00 12
## y              4.530266e-154  6
## job            1.045918e-146 60
## education      2.331835e-28  30
## housing        2.178875e-07   6
```

```
# Categories over/under represented in each cluster:
```

```
res.hcpc$desc.var$category
```

```
## $`1`
##                Cla/Mod  Mod/Cla    Global
## poutcome=poutcome-success  96.794872  85.795455  3.1287605
## y=y=yes                    18.850987  59.659091 11.1712796
## month=month-oct            19.587629  10.795455  1.9454473
## month=month-mar            22.727273   8.522727  1.3237064
## month=month-sep            18.032787   6.250000  1.2234256
## month=month-dec            26.923077   3.977273  0.5214601
## job=job-retired             9.268293  10.795455  4.1115122
## education=education-university.degree  5.051151  44.886364 31.3678299
## job=job-student            10.000000   5.681818  2.0056157
## month=month-nov             5.719921  16.477273 10.1684717
## job=job-housemaid           8.148148   6.250000  2.7075812
```

## job=job-admin.	4.494382	31.818182	24.9899719
## poutcome=poutcome-failure	5.241090	14.204545	9.5667870
## education=education-basic.9y	2.086050	9.090909	15.3830726
## job=job-services	1.405622	3.977273	9.9879663
## job=job-blue-collar	1.964133	13.068182	23.4857601
## month=month-jul	1.326900	6.250000	16.6265544
## month=month-may	1.378518	13.636364	34.9177698
## y=y-no	1.603071	40.340909	88.8287204
## poutcome=poutcome-nonexistent	0.000000	0.000000	87.3044525
##	p.value	v.test	
## poutcome=poutcome-success	3.881038e-254	34.051314	
## y=y-yes	2.249261e-57	15.964754	
## month=month-oct	6.983202e-10	6.166468	
## month=month-mar	5.595931e-09	5.828406	
## month=month-sep	7.963183e-06	4.466171	
## month=month-dec	2.444489e-05	4.219865	
## job=job-retired	1.204719e-04	3.845164	
## education=education-university.degree	1.343088e-04	3.818421	
## job=job-student	3.281252e-03	2.940082	
## month=month-nov	8.467352e-03	2.632843	
## job=job-housemaid	1.043595e-02	2.561039	
## job=job-admin.	3.744524e-02	2.080876	
## poutcome=poutcome-failure	4.315053e-02	2.022250	
## education=education-basic.9y	1.353542e-02	-2.469362	
## job=job-services	3.081976e-03	-2.959441	
## job=job-blue-collar	4.706686e-04	-3.496917	
## month=month-jul	3.415301e-05	-4.143845	
## month=month-may	1.098069e-10	-6.452794	
## y=y-no	2.249261e-57	-15.964754	
## poutcome=poutcome-nonexistent	7.847689e-169	-27.695786	
##			
## \$`2`			
##	Cla/Mod	Mod/Cla	Global
## job=job-retired	40.975610	26.8370607	4.1115122
## month=month-oct	58.762887	18.2108626	1.9454473
## y=y-yes	18.312388	32.5878594	11.1712796
## month=month-sep	55.737705	10.8626198	1.2234256
## poutcome=poutcome-failure	18.867925	28.7539936	9.5667870
## month=month-dec	61.538462	5.1118211	0.5214601
## month=month-aug	10.329986	23.0031949	13.9791416
## education=education-basic.4y	10.318949	17.5718850	10.6899318
## education=education-university.degree	8.056266	40.2555911	31.3678299
## month=month-mar	16.666667	3.5143770	1.3237064
## housing=housing-yes	7.192661	62.6198083	54.6530285
## poutcome=poutcome-success	1.923077	0.9584665	3.1287605
## housing=housing-no	5.174701	37.3801917	45.3469715
## job=job-entrepreneur	1.298701	0.6389776	3.0886482
## education=education-high.school	4.515599	17.5718850	24.4283995
## month=month-jun	3.374233	7.0287540	13.0766145
## job=job-services	2.208835	3.5143770	9.9879663
## education=education-basic.9y	2.346806	5.7507987	15.3830726
## month=month-jul	2.412545	6.3897764	16.6265544
## poutcome=poutcome-nonexistent	5.053986	70.2875399	87.3044525
## job=job-blue-collar	1.707942	6.3897764	23.4857601

```

## y=y-no 4.764055 67.4121406 88.8287204
## month=month-may 1.608271 8.9456869 34.9177698
## p.value v.test
## job=job-retired 3.206853e-50 14.901828
## month=month-oct 6.590944e-45 14.061041
## y=y-yes 3.834595e-26 10.576387
## month=month-sep 8.371333e-26 10.502960
## poutcome=poutcome-failure 9.880353e-24 10.042825
## month=month-dec 1.246107e-13 7.411785
## month=month-aug 8.737157e-06 4.446276
## education=education-basic.4y 1.483350e-04 3.793840
## education=education-university.degree 6.031404e-04 3.430198
## month=month-mar 3.078719e-03 2.959767
## housing=housing-yes 3.315110e-03 2.936900
## poutcome=poutcome-success 1.178823e-02 -2.518421
## housing=housing-no 3.315110e-03 -2.936900
## job=job-entrepreneur 3.124217e-03 -2.955244
## education=education-high.school 2.715285e-03 -2.998257
## month=month-jun 4.715598e-04 -3.496412
## job=job-services 1.072126e-05 -4.402091
## education=education-basic.9y 6.730357e-08 -5.398225
## month=month-jul 3.048594e-08 -5.538623
## poutcome=poutcome-nonexistent 1.400973e-16 -8.264657
## job=job-blue-collar 1.157386e-16 -8.287413
## y=y-no 3.834595e-26 -10.576387
## month=month-may 5.102772e-28 -10.973892
##
## $`3`
## Cla/Mod Mod/Cla Global
## month=month-apr 87.741935 23.5497835 6.2174087
## month=month-may 39.402642 59.3939394 34.9177698
## poutcome=poutcome-failure 59.329140 24.5021645 9.5667870
## job=job-student 60.000000 5.1948052 2.0056157
## month=month-mar 60.606061 3.4632035 1.3237064
## y=y-yes 33.572711 16.1904762 11.1712796
## housing=housing-yes 25.614679 60.4329004 54.6530285
## job=job-blue-collar 26.985482 27.3593074 23.4857601
## education=education-high.school 26.026273 27.4458874 24.4283995
## education=education-basic.9y 26.075619 17.3160173 15.3830726
## month=month-dec 7.692308 0.1731602 0.5214601
## month=month-oct 14.432990 1.2121212 1.9454473
## job=job-management 18.734793 6.6666667 8.2430806
## job=job-housemaid 12.592593 1.4718615 2.7075812
## job=job-technician 18.718593 12.9004329 15.9647012
## education=education-basic.4y 16.510319 7.6190476 10.6899318
## housing=housing-no 20.212295 39.5670996 45.3469715
## y=y-no 21.855949 83.8095238 88.8287204
## job=job-retired 6.341463 1.1255411 4.1115122
## month=month-jun 12.269939 6.9264069 13.0766145
## poutcome=poutcome-success 1.282051 0.1731602 3.1287605
## month=month-nov 3.353057 1.4718615 10.1684717
## poutcome=poutcome-nonexistent 19.986216 75.3246753 87.3044525
## month=month-aug 1.865136 1.1255411 13.9791416
## month=month-jul 1.809409 1.2987013 16.6265544

```

	p.value	v.test
## month=month-apr	8.449977e-140	25.170408
## month=month-may	6.491023e-85	19.526860
## poutcome=poutcome-failure	2.400092e-72	17.988443
## job=job-student	1.919893e-15	7.946413
## month=month-mar	7.101962e-11	6.518488
## y=y-yes	2.708602e-09	5.948349
## housing=housing-yes	6.372241e-06	4.513646
## job=job-blue-collar	4.638064e-04	3.500833
## education=education-high.school	6.941146e-03	2.699655
## education=education-basic.9y	3.965890e-02	2.057284
## month=month-dec	4.946472e-02	-1.964564
## month=month-oct	3.354840e-02	-2.125458
## job=job-management	2.394486e-02	-2.258013
## job=job-housemaid	1.872587e-03	-3.109730
## job=job-technician	9.690944e-04	-3.299349
## education=education-basic.4y	7.133660e-05	-3.971783
## housing=housing-no	6.372241e-06	-4.513646
## y=y-no	2.708602e-09	-5.948349
## job=job-retired	5.601283e-11	-6.554008
## month=month-jun	6.684182e-14	-7.493929
## poutcome=poutcome-success	8.577891e-16	-8.045664
## month=month-nov	7.092964e-39	-13.041624
## poutcome=poutcome-nonexistent	4.123377e-39	-13.082912
## month=month-aug	3.750561e-66	-17.179978
## month=month-jul	1.381625e-80	-19.011049

\$`4`

	Cla/Mod	Mod/Cla	Global
## month=month-nov	84.615385	99.7674419	10.168472
## y=y-no	9.415218	96.9767442	88.828720
## job=job-management	17.761557	16.9767442	8.243081
## poutcome=poutcome-failure	16.561845	18.3720930	9.566787
## job=job-entrepreneur	23.376623	8.3720930	3.088648
## education=education-university.degree	11.700767	42.5581395	31.367830
## job=job-unemployed	16.393443	4.6511628	2.446851
## job=job-self-employed	14.189189	4.8837209	2.968311
## job=job-admin.	7.223114	20.9302326	24.989972
## job=job-student	3.000000	0.6976744	2.005616
## education=education-basic.9y	6.258149	11.1627907	15.383073
## month=month-sep	0.000000	0.0000000	1.223426
## month=month-mar	0.000000	0.0000000	1.323706
## job=job-blue-collar	6.233988	16.9767442	23.485760
## poutcome=poutcome-nonexistent	8.063405	81.6279070	87.304452
## month=month-oct	0.000000	0.0000000	1.945447
## poutcome=poutcome-success	0.000000	0.0000000	3.128761
## y=y-yes	2.333932	3.0232558	11.171280
## month=month-apr	0.000000	0.0000000	6.217409
## month=month-jun	0.000000	0.0000000	13.076615
## month=month-aug	0.000000	0.0000000	13.979142
## month=month-jul	0.000000	0.0000000	16.626554
## month=month-may	0.000000	0.0000000	34.917770

	p.value	v.test
## month=month-nov	0.000000e+00	Inf

## y=y-no	1.445137e-10	6.411058	
## job=job-management	6.094995e-10	6.187954	
## poutcome=poutcome-failure	3.411201e-09	5.910476	
## job=job-entrepreneur	1.906432e-08	5.620285	
## education=education-university.degree	3.261080e-07	5.107698	
## job=job-unemployed	5.040211e-03	2.804453	
## job=job-self-employed	2.256934e-02	2.280647	
## job=job-admin.	3.956647e-02	-2.058247	
## job=job-student	2.833524e-02	-2.192613	
## education=education-basic.9y	8.935086e-03	-2.614528	
## month=month-sep	3.940443e-03	-2.882891	
## month=month-mar	2.495043e-03	-3.023942	
## job=job-blue-collar	6.095792e-04	-3.427316	
## poutcome=poutcome-nonexistent	4.218766e-04	-3.526007	
## month=month-oct	1.451605e-04	-3.799206	
## poutcome=poutcome-success	6.138981e-07	-4.986793	
## y=y-yes	1.445137e-10	-6.411058	
## month=month-apr	2.792982e-13	-7.304018	
## month=month-jun	3.466586e-28	-11.008782	
## month=month-aug	3.067865e-30	-11.426923	
## month=month-jul	2.141900e-36	-12.598870	
## month=month-may	1.367297e-85	-19.606258	
##			
## \$`5`			
##	Cla/Mod	Mod/Cla Global	
## poutcome=poutcome-nonexistent	29.129336	100.00000000	87.3044525
## month=month-aug	52.367288	28.78548896	13.9791416
## month=month-may	34.462952	47.31861199	34.9177698
## y=y-no	27.387672	95.66246057	88.8287204
## education=education-basic.4y	38.836773	16.32492114	10.6899318
## education=education-professional.course	30.406504	14.74763407	12.3345367
## loan=loan-no	26.179749	87.06624606	84.5768151
## job=job-retired	33.658537	5.44164038	4.1115122
## housing=housing-no	26.890756	47.94952681	45.3469715
## job=job-housemaid	33.333333	3.54889590	2.7075812
## job=job-blue-collar	27.754056	25.63091483	23.4857601
## housing=housing-yes	24.220183	52.05047319	54.6530285
## job=job-unemployed	17.213115	1.65615142	2.4468512
## job=job-admin.	22.953451	22.55520505	24.9899719
## education=education-high.school	22.495895	21.60883281	24.4283995
## loan=loan-yes	21.326398	12.93375394	15.4231849
## month=month-dec	0.000000	0.00000000	0.5214601
## education=education-university.degree	21.419437	26.41955836	31.3678299
## month=month-sep	0.000000	0.00000000	1.2234256
## month=month-mar	0.000000	0.00000000	1.3237064
## month=month-jun	15.644172	8.04416404	13.0766145
## job=job-student	2.000000	0.15772871	2.0056157
## month=month-oct	1.030928	0.07886435	1.9454473
## poutcome=poutcome-success	0.000000	0.00000000	3.1287605
## y=y-yes	9.874327	4.33753943	11.1712796
## month=month-apr	0.000000	0.00000000	6.2174087
## poutcome=poutcome-failure	0.000000	0.00000000	9.5667870
## month=month-nov	0.000000	0.00000000	10.1684717
##	p.value	v.test	

```

## poutcome=poutcome-nonexistent      5.289007e-88  19.886855
## month=month-aug                     1.191159e-61  16.567794
## month=month-may                     3.493765e-26  10.585108
## y=y-no                             2.522644e-22   9.718174
## education=education-basic.4y        5.108301e-13   7.222387
## education=education-professional.course 2.901818e-03   2.977952
## loan=loan-no                       3.988941e-03   2.879035
## job=job-retired                    7.243663e-03   2.685430
## housing=housing-no                 3.131061e-02   2.153103
## job=job-housemaid                  3.740547e-02   2.081311
## job=job-blue-collar                3.801325e-02   2.074712
## housing=housing-yes                 3.131061e-02  -2.153103
## job=job-unemployed                 3.030370e-02  -2.166098
## job=job-admin.                     1.971720e-02  -2.331686
## education=education-high.school     6.416463e-03  -2.725703
## loan=loan-yes                      3.988941e-03  -2.879035
## month=month-dec                     4.751074e-04  -3.494412
## education=education-university.degree 9.020632e-06  -4.439409
## month=month-sep                     1.482866e-08  -5.663534
## month=month-mar                     3.344948e-09  -5.913706
## month=month-jun                     1.357270e-10  -6.420613
## job=job-student                    8.674956e-11  -6.488405
## month=month-oct                     1.130837e-11  -6.788783
## poutcome=poutcome-success           5.633481e-21  -9.396637
## y=y-yes                             2.522644e-22  -9.718174
## month=month-apr                     9.931295e-42 -13.533406
## poutcome=poutcome-failure           3.651539e-65 -17.047444
## month=month-nov                     1.753726e-69 -17.619231
##
## $`6`
##
## Cla/Mod      Mod/Cla      Global
## month=month-jul      62.243667  35.8084663  16.6265544
## poutcome=poutcome-nonexistent 33.103607 100.0000000  87.3044525
## month=month-jun      57.055215  25.8154060  13.0766145
## y=y-no                30.571235  93.9625260  88.8287204
## job=job-services      36.746988  12.6995142   9.9879663
## housing=housing-no    31.357806  49.2019431  45.3469715
## education=education-basic.9y 33.898305  18.0430257  15.3830726
## education=education-high.school 32.266010  27.2727273  24.4283995
## job=job-technician    32.035176  17.6960444  15.9647012
## job=job-blue-collar   31.255337  25.3990285  23.4857601
## job=job-management    22.627737   6.4538515   8.2430806
## housing=housing-yes   26.862385  50.7980569  54.6530285
## job=job-student      14.000000   0.9715475   2.0056157
## month=month-dec       0.000000   0.0000000   0.5214601
## month=month-oct       6.185567   0.4163775   1.9454473
## education=education-basic.4y 18.011257   6.6620402  10.6899318
## month=month-sep       0.000000   0.0000000   1.2234256
## month=month-mar       0.000000   0.0000000   1.3237064
## y=y-yes               15.619390   6.0374740  11.1712796
## month=month-may       20.964963  25.3296322  34.9177698
## poutcome=poutcome-success 0.000000   0.0000000   3.1287605
## job=job-retired       0.000000   0.0000000   4.1115122
## month=month-apr       0.000000   0.0000000   6.2174087

```

```

## poutcome=poutcome-failure      0.000000  0.0000000  9.5667870
## month=month-nov                 0.000000  0.0000000 10.1684717
##                                p.value    v.test
## month=month-jul                 2.193896e-108 22.116481
## poutcome=poutcome-nonexistent  2.146626e-102 21.485087
## month=month-jun                 5.177232e-59 16.198381
## y=y-no                         1.177463e-14  7.718457
## job=job-services                6.653178e-05  3.988360
## housing=housing-no              4.994477e-04  3.481052
## education=education-basic.9y    1.044519e-03  3.278253
## education=education-high.school 3.082735e-03  2.959365
## job=job-technician              3.462568e-02  2.112709
## job=job-blue-collar             4.313066e-02  2.022442
## job=job-management              2.864908e-03 -2.981873
## housing=housing-yes             4.994477e-04 -3.481052
## job=job-student                4.497429e-04 -3.509031
## month=month-dec                 1.370593e-04 -3.813416
## month=month-oct                 2.045152e-08 -5.608138
## education=education-basic.4y    1.058277e-09 -6.100363
## month=month-sep                 7.911717e-10 -6.146689
## month=month-mar                 1.400443e-10 -6.415845
## y=y-yes                         1.177463e-14 -7.718457
## month=month-may                 3.650636e-20 -9.197885
## poutcome=poutcome-success       2.829400e-24 -10.165406
## job=job-retired                 7.352963e-32 -11.746586
## month=month-apr                 1.961207e-48 -14.624460
## poutcome=poutcome-failure       9.590135e-76 -18.417043
## month=month-nov                 8.816429e-81 -19.034599
##
## $`7`
##                                Cla/Mod    Mod/Cla    Global    p.value
## poutcome=poutcome-nonexistent  4.6634505 100.0000000 87.304452 5.792869e-13
## month=month-jul                 8.0820265 33.0049261 16.626554 5.404009e-09
## month=month-jun                 8.2822086 26.6009852 13.076615 1.418693e-07
## y=y-no                         4.4027997 96.0591133 88.828720 2.087025e-04
## month=month-aug                 5.7388809 19.7044335 13.979142 2.141058e-02
## housing=housing-no              4.6881911 52.2167488 45.346972 4.567393e-02
## housing=housing-yes             3.5596330 47.7832512 54.653028 4.567393e-02
## month=month-oct                 0.0000000 0.0000000 1.945447 1.704213e-02
## poutcome=poutcome-success       0.0000000 0.0000000 3.128761 1.374795e-03
## y=y-yes                         1.4362657 3.9408867 11.171280 2.087025e-04
## month=month-apr                 0.3225806 0.4926108 6.217409 2.660092e-05
## month=month-nov                 0.5917160 1.4778325 10.168472 6.310181e-07
## month=month-may                 2.1826536 18.7192118 34.917770 2.409376e-07
## poutcome=poutcome-failure       0.0000000 0.0000000 9.566787 8.711382e-10
##                                v.test
## poutcome=poutcome-nonexistent  7.205271
## month=month-jul                 5.834228
## month=month-jun                 5.262811
## y=y-no                         3.708243
## month=month-aug                 2.300665
## housing=housing-no              1.998394
## housing=housing-yes             -1.998394
## month=month-oct                 -2.385798

```

```
## poutcome=poutcome-success      -3.199891
## y=y-yes                          -3.708243
## month=month-apr                  -4.200770
## month=month-nov                  -4.981474
## month=month-may                  -5.164612
## poutcome=poutcome-failure       -6.131391
```

Sobre les variables quantitatives (numeriques)

A continuació es realitzara una altra descripció dels diferents clusters formats però basant-nos en les variables quantitatives (numeriques), les quals han estat utilitzades en el PCA. En el primer output es poden veure les que han estat “p-provades” com a globalment relacionades amb la clusterització, mentre que el detall es pot veure en el segon output.

El **primer cluster** està caracteritzat per individus (sempre respecte la mitja global) més contactats en campanyes anteriors (previous) i menys en l'actual (campaign); així com individus que han estat contactats molt més recentment (pdays). La duració de les trucades en aquest cluster està 35 segons per sobre la mitja global, així com també podem veure lleugers esbiaixos en els indicadors socioeconòmics (valors baixos de euribor3m, emp.var.rate, nr.employed i cons.conf.idx). El **segon cluster** té una mitjana d'edat 12 anys per sobre la mitjana global, si bé també una major desviació estàndard dins la categoria; a part d'un esbiaix similar a l'anterior amb els indicadors socioeconòmics. El **tercer cluster** és similar al cluster 2 però amb una mitja d'edat 4 anys per sota de la global; a més en aquest grup la duració de les trucades està per sobre la mitja per 14 segons. Els quatre altres clusters contenen només individus els quals no han estat contactats abans per cap altra campanya (mitja de pdays=19.0 amb un sd de 0) i els seus indicadors socioeconòmics tenen un comportament similar en l'esbiaix. El **quart cluster** però, a diferència dels altres tres, no mostra una mitja de 0.0 en la variable “previous”, el que fa pensar que aquests individus han estat contactats abans però no necessàriament per a una campanya d'un producte. La duració en aquest cluster és 30 segons per sota la mitja global. Entre els **cinque cluster** i **sise cluster** la diferència principal està en l'edat, que que el cinquè agrupa individus per sobre la mitja i el sise per sota. Per últim, el **sete cluster** destaca per a una mitja de duració de les trucades extremadament curta (65 segons per sota la mitja global), així com per un nombre de vegades que l'individu ha estat contactat en l'actual campanya molt per sobre de la mitja global (11.3 respecte 2.5).

Numeric (quantitative) variables globally related to clustering partition:

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

```
##              Eta2      P-value
## age          0.335988921 0.00000e+00
## campaign     0.540040103 0.00000e+00
## pdays        0.925018800 0.00000e+00
## previous     0.465947411 0.00000e+00
## emp.var.rate 0.960040097 0.00000e+00
## cons.price.idx 0.614485994 0.00000e+00
## cons.conf.idx 0.503365350 0.00000e+00
## euribor3m    0.988716657 0.00000e+00
## nr.employed  0.879442070 0.00000e+00
## duration     0.006588486 1.08361e-05
```

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

```
## $`1`
##              v.test Mean in category Overall mean sd in category
## previous     43.973874          1.687500      0.15984757      0.8848873
## cons.conf.idx  8.805587        -37.359091     -40.42591256      6.5879649
## age           3.296336          42.619318      40.06799037     14.8938468
## duration      2.067896          285.897727     250.62194144     213.7620565
```

```

## campaign      -3.562842      1.880682      2.53512993      1.3282215
## cons.price.idx -7.388255      93.253477      93.57245006      0.7521195
## emp.var.rate   -19.367388      -2.208523      0.06446049      0.8800737
## euribor3m      -20.279982      1.006216      3.61448034      0.6758827
## nr.employed    -25.213543      5032.493750 5166.47621340      52.1226287
## pdays          -67.905435      5.738636      18.52647413      3.5659192
##               Overall sd      p.value
## previous        0.4691873 0.000000e+00
## cons.conf.idx    4.7037753 1.301691e-18
## age             10.4532458 9.795477e-04
## duration         230.3904064 3.864981e-02
## campaign         2.4808187 3.668615e-04
## cons.price.idx    0.5830800 1.487681e-13
## emp.var.rate      1.5850448 1.454512e-83
## euribor3m        1.7370025 1.932036e-91
## nr.employed       71.7679377 2.845704e-140
## pdays            2.5433666 0.000000e+00
##
## `$2`
##               v.test Mean in category Overall mean sd in category
## cons.conf.idx    25.044818      -33.9789137 -40.42591256      6.3573009
## age              19.567738      51.2619808  40.06799037      16.7801531
## previous         7.959341       0.3642173   0.15984757      0.6047713
## pdays            3.149566      18.9648562  18.52647413      0.4099927
## campaign        -5.588859       1.7763578   2.53512993      1.2567327
## cons.price.idx   -26.189093      92.7367668  93.57245006      0.5664489
## euribor3m       -28.684949       0.8877157   3.61448034      0.2248025
## nr.employed     -31.489513      5042.7990415 5166.47621340      36.8870017
## emp.var.rate    -32.543552      -2.7584665   0.06446049      0.6816163
##               Overall sd      p.value
## cons.conf.idx    4.7037753 1.988210e-138
## age             10.4532458 2.913265e-85
## previous         0.4691873 1.729578e-15
## pdays            2.5433666 1.635132e-03
## campaign         2.4808187 2.285660e-08
## cons.price.idx    0.5830800 3.537697e-151
## euribor3m        1.7370025 5.878957e-181
## nr.employed       71.7679377 1.209022e-217
## emp.var.rate      1.5850448 2.583254e-232
##
## `$3`
##               v.test Mean in category Overall mean sd in category
## previous         8.754589       0.2658009   0.15984757      0.4829555
## pdays            7.006604      18.9861472  18.52647413      0.3326118
## duration         2.386888      264.8069264 250.62194144      234.8283557
## campaign        -5.291156       2.1965368   2.53512993      1.6640810
## age            -13.956513      36.3047619  40.06799037      8.5250191
## cons.price.idx   -35.103266      93.0444814  93.57245006      0.3872219
## cons.conf.idx    -42.294031      -45.5575758 -40.42591256      3.1552802
## nr.employed     -43.023029      5086.8302165 5166.47621340      33.6303225
## emp.var.rate    -46.855120      -1.8512554   0.06446049      0.3273476
## euribor3m       -52.114578       1.2794519   3.61448034      0.1767441
##               Overall sd      p.value
## previous        0.4691873 2.048491e-18

```

```

## pdays          2.5433666 2.441716e-12
## duration        230.3904064 1.699168e-02
## campaign        2.4808187 1.215453e-07
## age             10.4532458 2.871432e-44
## cons.price.idx  0.5830800 6.010017e-270
## cons.conf.idx   4.7037753 0.000000e+00
## nr.employed     71.7679377 0.000000e+00
## emp.var.rate    1.5850448 0.000000e+00
## euribor3m       1.7370025 0.000000e+00
##
## $`4`
##               v.test Mean in category Overall mean sd in category
## nr.employed    8.848963    5195.7546512 5166.47621340 9.392794e-01
## euribor3m      6.251387      4.1150930   3.61448034 7.151549e-02
## pdays         4.038403     19.0000000   18.52647413 0.000000e+00
## emp.var.rate   -2.253765     -0.1002326   0.06446049 4.816817e-03
## duration       -2.828692    220.5767442 250.62194144 2.118224e+02
## campaign       -5.593856      1.8953488   2.53512993 1.308550e+00
## cons.conf.idx  -7.300497    -42.0090698 -40.42591256 1.878559e-01
## cons.price.idx -13.893678     93.1989674   93.57245006 2.138667e-02
##               Overall sd      p.value
## nr.employed    71.767938 8.833617e-19
## euribor3m      1.737002 4.068245e-10
## pdays         2.543367 5.381623e-05
## emp.var.rate    1.585045 2.421094e-02
## duration       230.390406 4.673870e-03
## campaign       2.480819 2.220806e-08
## cons.conf.idx   4.703775 2.867066e-13
## cons.price.idx  0.583080 6.918945e-44
##
## $`5`
##               v.test Mean in category Overall mean sd in category
## emp.var.rate    31.018314      1.256861   0.06446049 0.15450722
## euribor3m      30.778624      4.911099   3.61448034 0.05291853
## age            28.431726     47.276025  40.06799037 6.83638157
## nr.employed    25.303757    5210.519322 5166.47621340 18.52135643
## cons.conf.idx  23.501417    -37.744874 -40.42591256 2.61453060
## cons.price.idx 20.434823     93.861426   93.57245006 0.29713094
## pdays         7.676659     19.000000   18.52647413 0.00000000
## campaign       -3.366631      2.332570   2.53512993 1.51977446
## previous       -14.047446      0.000000   0.15984757 0.00000000
##               Overall sd      p.value
## emp.var.rate    1.5850448 3.053044e-211
## euribor3m      1.7370025 5.064155e-208
## age            10.4532458 8.199565e-178
## nr.employed    71.7679377 2.904135e-141
## cons.conf.idx   4.7037753 3.944955e-122
## cons.price.idx  0.5830800 8.198980e-93
## pdays         2.5433666 1.632909e-14
## campaign       2.4808187 7.609255e-04
## previous       0.4691873 7.986314e-45
##
## $`6`
##               v.test Mean in category Overall mean sd in category

```

```
## emp.var.rate      35.593579      1.317765      0.06446049      0.15938206
## cons.price.idx    34.404615      94.018094      93.57245006      0.31164559
## euribor3m         34.019646      4.927205      3.61448034      0.04863405
## nr.employed       32.673604     5218.568217  5166.47621340     16.19038972
## pdays             8.380911      19.000000      18.52647413      0.00000000
## cons.conf.idx     3.801019     -40.028730    -40.42591256      2.94116617
## campaign          -6.468091      2.178667      2.53512993      1.37338503
## previous          -15.336148      0.000000      0.15984757      0.00000000
## age              -27.777266      33.617627     40.06799037      5.41378149
##
## Overall sd      p.value
## emp.var.rate    1.5850448 1.761028e-277
## cons.price.idx  0.5830800 2.151036e-259
## euribor3m       1.7370025 1.141427e-253
## nr.employed     71.7679377 3.704005e-234
## pdays           2.5433666 5.251979e-17
## cons.conf.idx   4.7037753 1.441025e-04
## campaign        2.4808187 9.924849e-11
## previous        0.4691873 4.384517e-53
## age            10.4532458 8.165488e-170
##
## $`7`
##
## v.test Mean in category Overall mean sd in category
## campaign      51.681982      11.349754 2.535130e+00      4.2854821
## emp.var.rate   11.261387      1.291626 6.446049e-02      0.3725545
## nr.employed    10.764761     5219.589655 5.166476e+03     18.9160146
## euribor3m      10.662441      4.887768 3.614480e+00      0.3732829
## cons.price.idx  9.771322      93.964148 9.357245e+01      0.3729365
## pdays          2.708103      19.000000 1.852647e+01      0.0000000
## duration       -4.128695     185.226601 2.506219e+02     230.3643674
## previous       -4.955532      0.000000 1.598476e-01      0.0000000
##
## Overall sd      p.value
## campaign        2.4808187 0.000000e+00
## emp.var.rate     1.5850448 2.035309e-29
## nr.employed      71.7679377 5.049245e-27
## euribor3m        1.7370025 1.525396e-26
## cons.price.idx   0.5830800 1.494883e-22
## pdays            2.5433666 6.766903e-03
## duration         230.3904064 3.648276e-05
## previous         0.4691873 7.213279e-07
```

Descripció dels clusters mitjançant individus

Aquí descrivim els clusters mitjançant els individus, concretament ens centrarem en els individus que estan al centre de gravetat del cluster (para-parangons) i els que estan més allunyats de la resta de clusters, es a dir, els més propis del cluster en qüestió i menys dels altres (dist-específics). En les següents gràfiques es poden veure representats en blau els parangons i en taronja els individus específics per a cada cluster de la classificació.

```
# Description of the clusters by individuals:
```

```
names(res.hcpc$desc.ind)
```

```
## [1] "para" "dist"
```

```
res.hcpc$desc.ind$para #parangons of each clusters
```

```
## Cluster: 1
```

```

##      31670      30284      37727      36469      30158
## 0.3892465 0.6099896 0.6405242 0.6517560 0.6658754
## -----
## Cluster: 2
##      38732      38503      37268      39000      37397
## 0.4162434 0.5086237 0.5839632 0.5993900 0.6629911
## -----
## Cluster: 3
##      36703      36516      36745      36444      36765
## 0.2333947 0.2626528 0.2918447 0.3074016 0.4287422
## -----
## Cluster: 4
##      27423      26065      25081      25324      26601
## 0.2429945 0.2473604 0.2495809 0.2495809 0.2519679
## -----
## Cluster: 5
##      19520      19421      20018      23464      23491
## 0.1425541 0.1427061 0.1428585 0.1431648 0.1434728
## -----
## Cluster: 6
##      12276      14034      16531      16228      18537
## 0.2192499 0.2196483 0.2196483 0.2199155 0.2247631
## -----
## Cluster: 7
##      20528      15744      21650      23413      21920
## 0.1715338 0.1947761 0.2990218 0.3655228 0.3672151
## -----
res.hcpc$desc.ind$dist #specific individuals

## Cluster: 1
##      40258      39828      40592      39359      39659
## 10.171491 10.130157 9.325015 9.035001 9.033797
## -----
## Cluster: 2
##      38207      38192      38185      38558      38253
## 6.441959 6.210520 6.144412 6.115588 6.069564
## -----
## Cluster: 3
##      27734      34291      40877      27890      33321
## 4.536960 4.490433 4.483955 4.461203 4.400579
## -----
## Cluster: 4
##      26974      26571      27451      25845      26975
## 2.669911 2.636325 2.431859 2.426201 2.423020
## -----
## Cluster: 5
##      22846      743      22712      22178      20336
## 2.849191 2.722699 2.706876 2.651221 2.651213
## -----
## Cluster: 6
##      18411      11796      8740      8085      11253
## 2.757309 2.704960 2.704805 2.689915 2.651787
## -----
## Cluster: 7
##      18712      8993      15581      17458      18417

```



```
## 9.054553 8.685034 8.669008 8.303776 8.273290

#Characteristic individuals - as many as clusters
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]]))
para5<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[5]]))
para6<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[6]]))
para7<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[7]]))

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]]))
dist5<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[5]]))
dist6<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[6]]))
dist7<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[7]]))

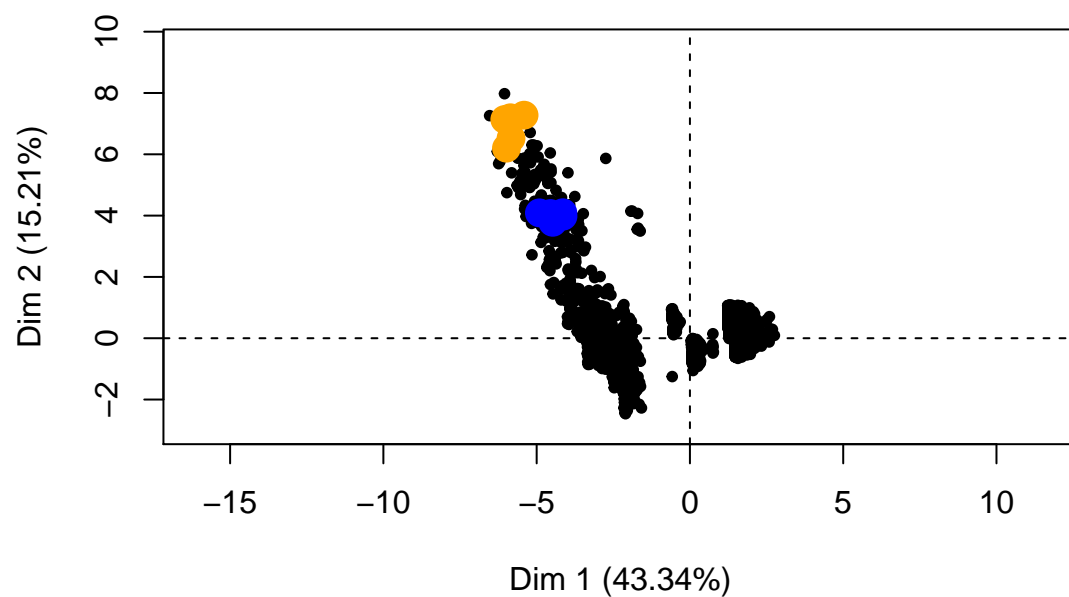
df$clust<-factor(res.hcpc$data.clust$clust)

#cluster as variable suplementaria
res.pca<-PCA(df[,c("duration", "clust", vars_con)], quanti.sup=1,quali.sup=2,ncp=4, graph=FALSE)

#? habillage
#color the individuals among a categorical variable (give the number/name of the categorical variable)

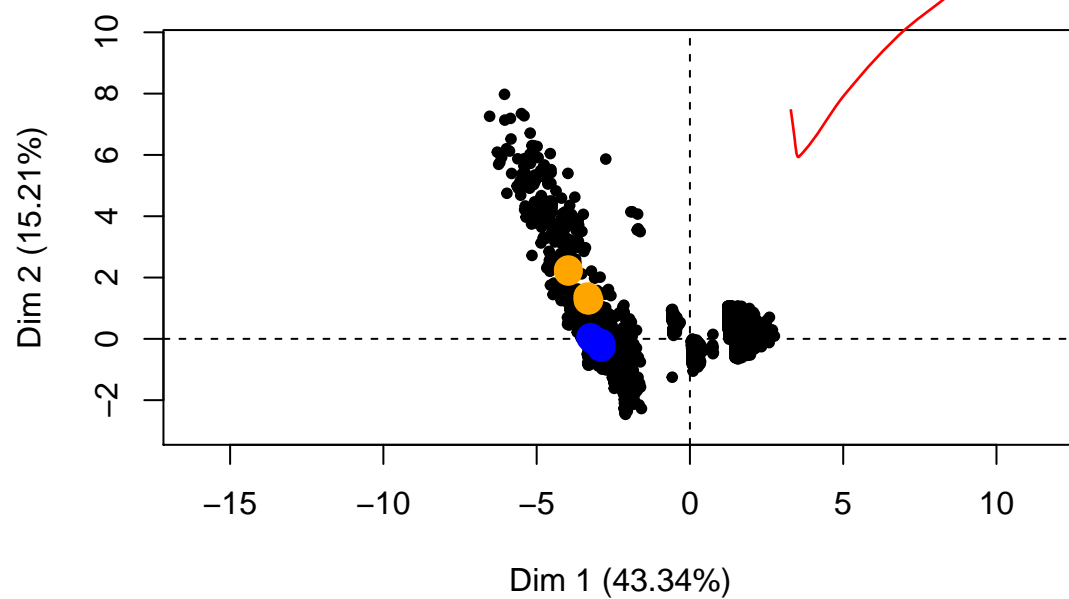
plot.PCA(res.pca, label="none", invisible="quali", title="Characteristic individuals - Cluster 1")
#pintar "para" del cluster 1
points(res.pca$ind$coord[para1,1], res.pca$ind$coord[para1,2], col="blue", cex=2, pch=16)
#pintar "dist" del cluster 1
points(res.pca$ind$coord[dist1,1], res.pca$ind$coord[dist1,2], col="orange", cex=2, pch=16)
```

Characteristic individuals – Cluster 1

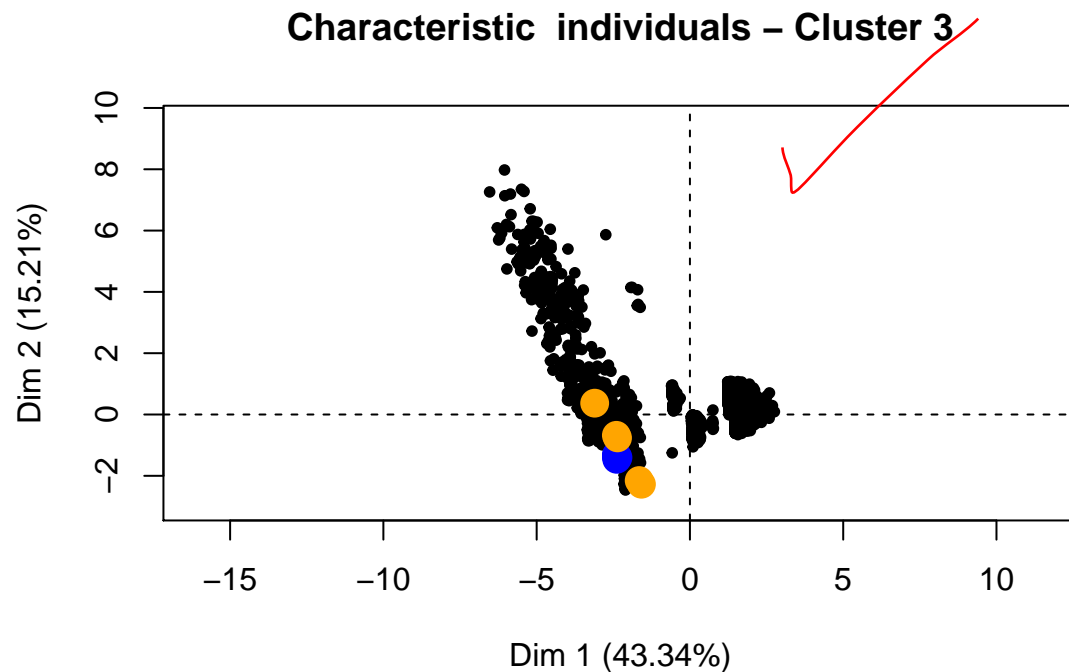


```
plot.PCA(res.pca, label="none", invisible="quali", title="Characteristic individuals - Cluster 2")
points(res.pca$ind$coord[para2,1], res.pca$ind$coord[para2,2], col="blue", cex=2, pch=16)
points(res.pca$ind$coord[dist2,1], res.pca$ind$coord[dist2,2], col="orange", cex=2, pch=16)
```

Characteristic individuals – Cluster 2

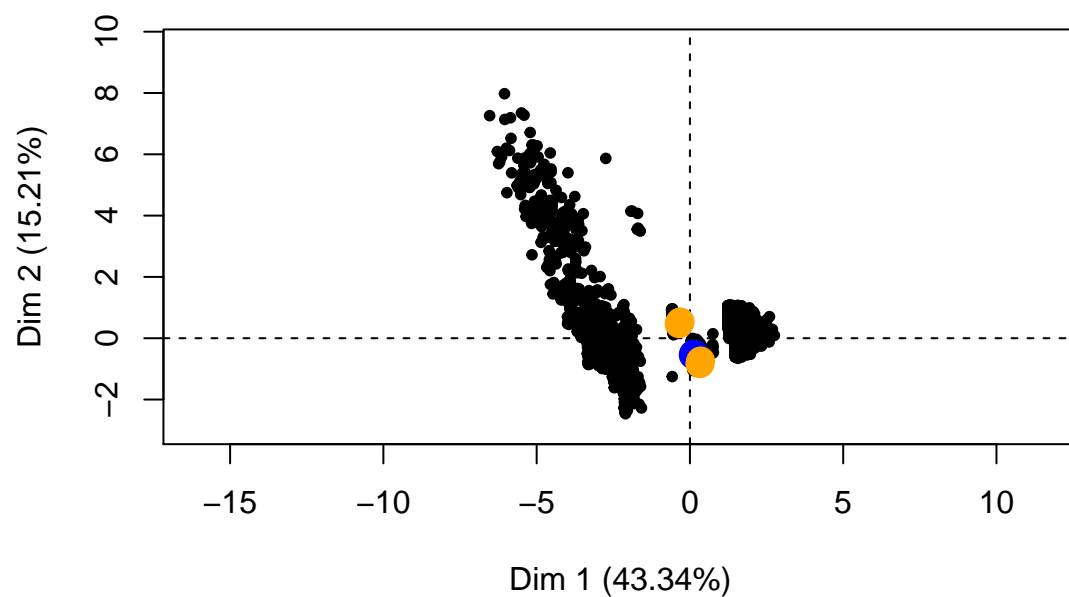


```
plot.PCA(res.pca, label="none", invisible="quali", title="Characteristic individuals - Cluster 3")
points(res.pca$ind$coord[para3,1], res.pca$ind$coord[para3,2], col="blue", cex=2, pch=16)
points(res.pca$ind$coord[dist3,1], res.pca$ind$coord[dist3,2], col="orange", cex=2, pch=16)
```



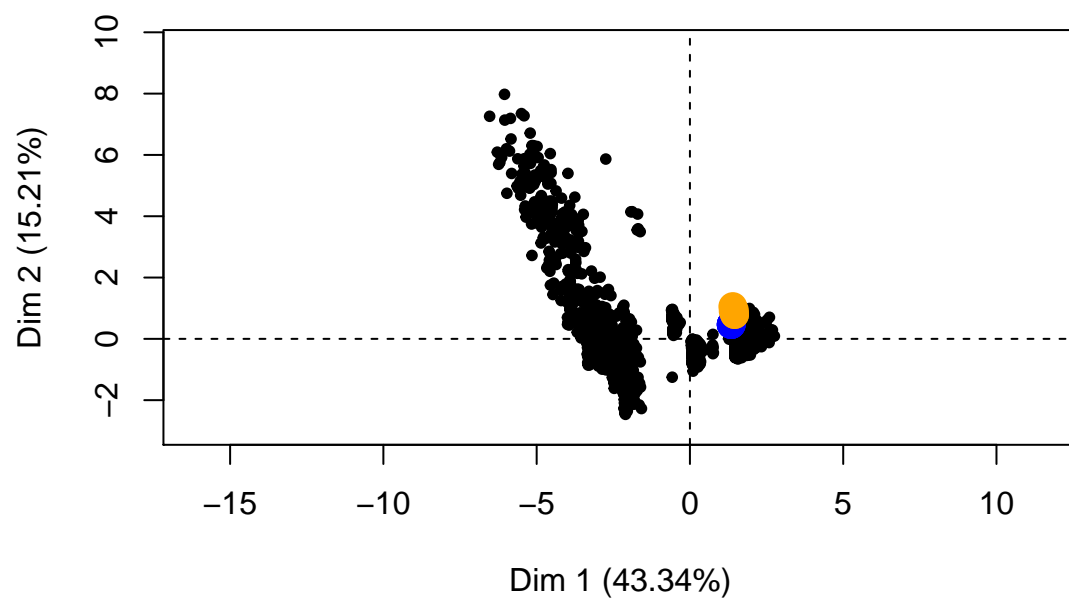
```
plot.PCA(res.pca, label="none", invisible="quali", title="Characteristic individuals - Cluster 4")
points(res.pca$ind$coord[para4,1], res.pca$ind$coord[para4,2], col="blue", cex=2, pch=16)
points(res.pca$ind$coord[dist4,1], res.pca$ind$coord[dist4,2], col="orange", cex=2, pch=16)
```

Characteristic individuals – Cluster 4



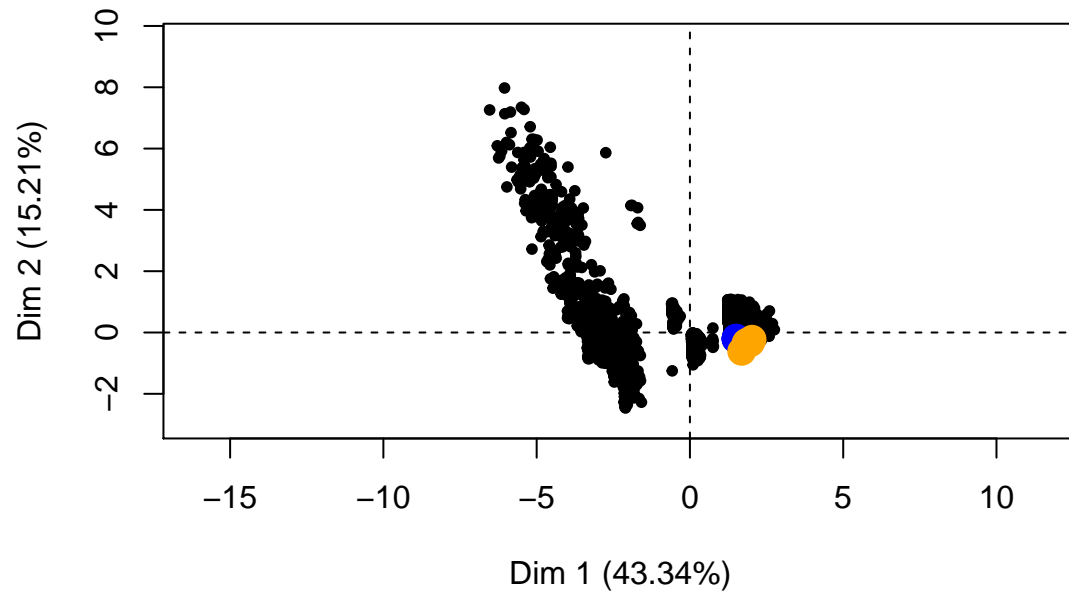
```
plot.PCA(res.pca, label="none", invisible="quali", title="Characteristic individuals - Cluster 5")
points(res.pca$ind$coord[para5,1], res.pca$ind$coord[para5,2], col="blue", cex=2, pch=16)
points(res.pca$ind$coord[dist5,1], res.pca$ind$coord[dist5,2], col="orange", cex=2, pch=16)
```

Characteristic individuals – Cluster 5



```
plot.PCA(res.pca, label="none", invisible="quali", title="Characteristic individuals - Cluster 6")
points(res.pca$ind$coord[para6,1], res.pca$ind$coord[para6,2], col="blue", cex=2, pch=16)
points(res.pca$ind$coord[dist6,1], res.pca$ind$coord[dist6,2], col="orange", cex=2, pch=16)
```

Characteristic individuals – Cluster 6



```
plot.PCA(res.pca, label="none", invisible="quali", title="Characteristic individuals - Cluster 7")
points(res.pca$ind$coord[para7,1], res.pca$ind$coord[para7,2], col="blue", cex=2, pch=16)
points(res.pca$ind$coord[dist7,1], res.pca$ind$coord[dist7,2], col="orange", cex=2, pch=16)
```

Characteristic individuals – Cluster 7

