Deliverable 2

Guillem Valls, Sergio Mazzariol

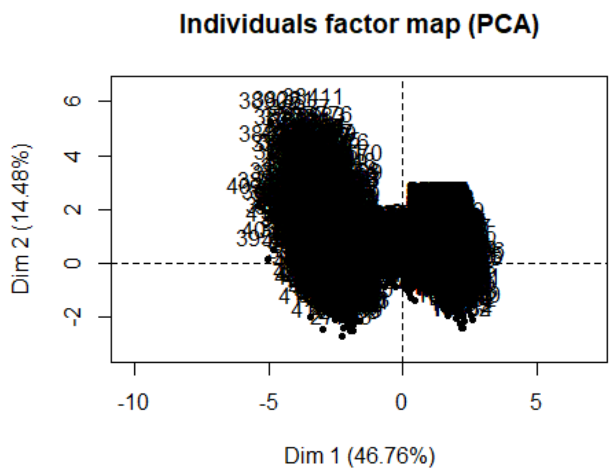
Table of Contents

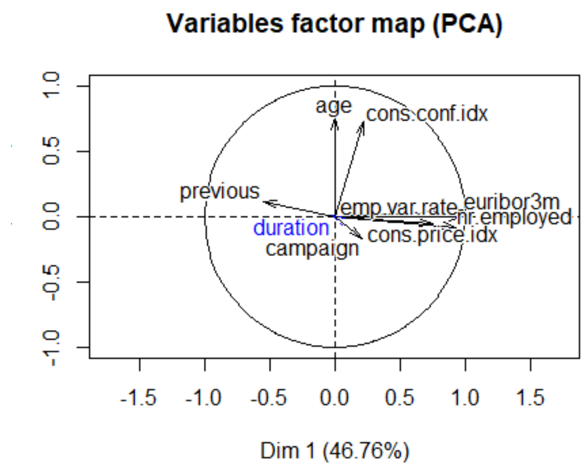
Valores propios y ejes dominantes

Eigenvalues and dominant axes analysis. How many axes we have to interpret according to Kaiser and Elbow's rule?

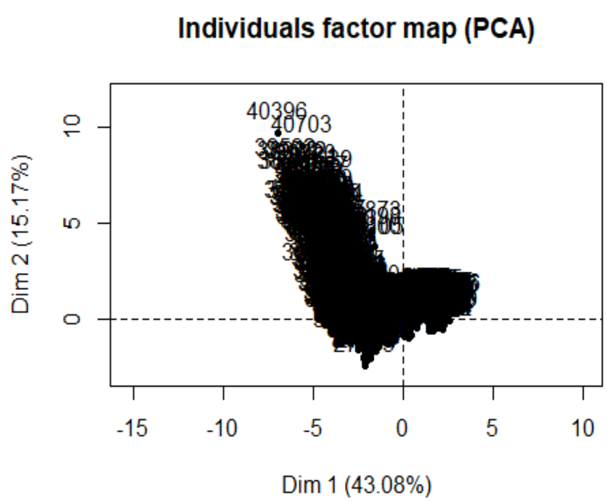
Hemos decidido probar como se ve el PCA sin y con la variable pdays, ya que consideramos que es una variable con bastantes missings, aún así aporta información por lo tanto la vamos a considerar.

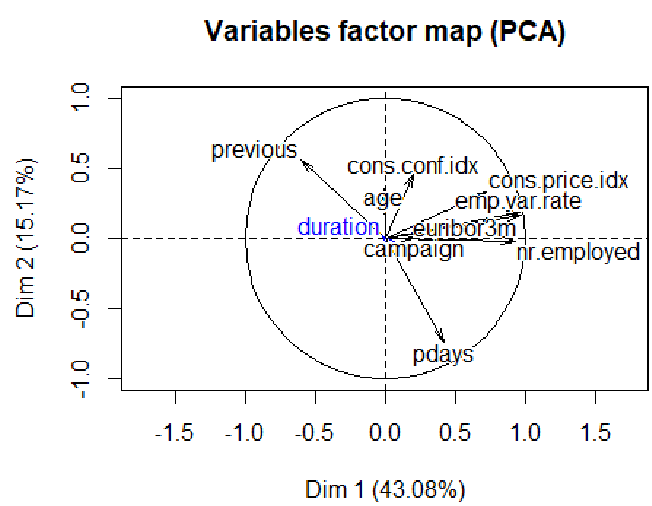
```
vars_num_sin_pday = vars_num[-3];
res2.pca<-PCA(df[,c('duration',vars_num_sin_pday)],quanti.sup=1)
```






```r
res.pca<-PCA(df[,c('duration',vars_num)],quanti.sup=1)
```





Vemos que con pdays existe una relación inversa con previous, respecto a los dos ejes factoriales, sin pdays se puede ver que la contribución de la variable age con el segundo eje factorial es mayor, ya que gráficamente tiene mayor magnitud además que las variables socio económicas, se ven mejor representadas en el primer eje factorial.

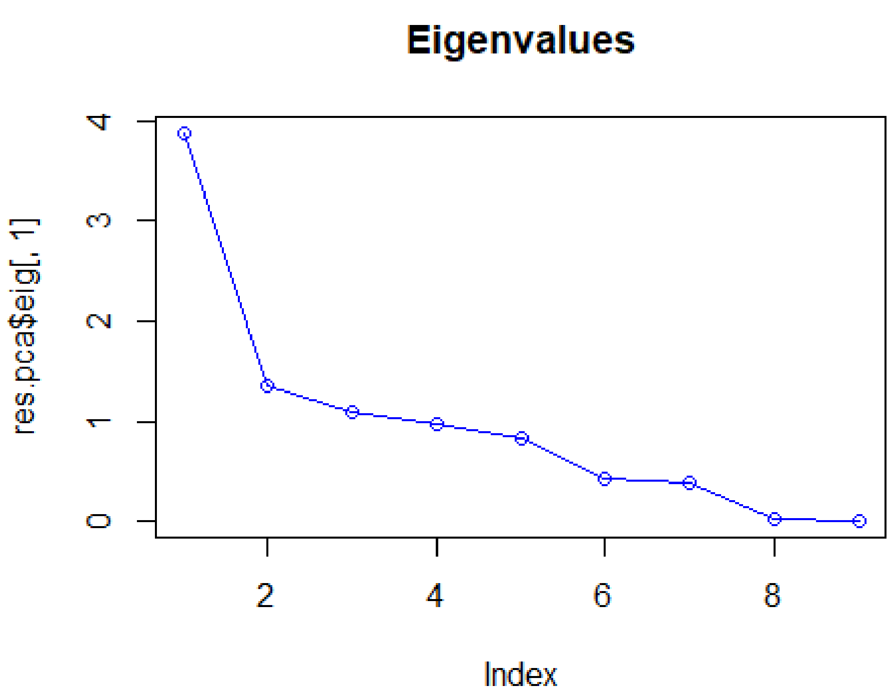
Por la ley de Kaiser, deberíamos utilizar los 3 primeros ejes factoriales, los cuales son mayores a 1. Por la ley de ElBow, al realizar el gráfico podemos ver que la gráfica empieza a ser plana a partir de la 2da dimensión, es decir que se cogen las 2 primeras dimensiones.

Si tomamos en cuenta el criterio del 80% se deberían coger las 4 primeras dimensiones.

Para realizar el futuro análisis, conviene utilizar dimensiones pares, por lo que decidimos solo usar 2.

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

##
## Call:
## PCA(X = df[, c("duration", vars_num)], quanti.sup = 1)
##
##
## Eigenvalues
##                        Dim.1  Dim.2  Dim.3  Dim.4  Dim.5  Dim.6  Dim.7
## Variance                3.88   1.36   1.10   0.97   0.83   0.43   0.39
## % of var.              43.08  15.17  12.28  10.74   9.25   4.82   4.29
## Cumulative % of var.   43.08  58.24  70.52  81.26  90.50  95.33  99.62
##                        Dim.8  Dim.9
## Variance                0.02   0.01
## % of var.               0.27   0.12
## Cumulative % of var.   99.88 100.00
##
## Individuals (the 10 first)
##                    Dist    Dim.1   ctr  cos2   Dim.2   ctr  cos2   Dim.3
## 20               |  1.74 |  1.29  0.01  0.55 |  0.44  0.00  0.06 |  0.41
## 21               |  1.99 |  1.30  0.01  0.42 |  0.18  0.00  0.01 | -0.13
## 30               |  2.24 |  1.28  0.01  0.33 |  0.90  0.01  0.16 |  1.38
## 33               |  1.93 |  1.29  0.01  0.44 |  0.73  0.01  0.14 |  1.02
## 48               |  1.74 |  1.29  0.01  0.55 |  0.47  0.00  0.07 |  0.47
## 56               |  2.24 |  1.28  0.01  0.33 |  0.90  0.01  0.16 |  1.38
## 61               |  1.85 |  1.29  0.01  0.48 |  0.67  0.01  0.13 |  0.90
## 62               |  2.02 |  1.29  0.01  0.41 |  0.78  0.01  0.15 |  1.14
## 65               |  1.89 |  1.29  0.01  0.46 |  0.70  0.01  0.14 |  0.96
## 84               |  1.93 |  1.29  0.01  0.44 |  0.73  0.01  0.14 |  1.02
##                    ctr  cos2   Dim.4   ctr  cos2
## 20                0.00  0.06 | -0.81  0.01  0.22 |
## 21                0.00  0.00 | -1.06  0.02  0.28 |
## 30                0.03  0.38 | -0.38  0.00  0.03 |
## 33                0.02  0.28 | -0.54  0.01  0.08 |
## 48                0.00  0.07 | -0.79  0.01  0.21 |
## 56                0.03  0.38 | -0.38  0.00  0.03 |
## 61                0.01  0.23 | -0.60  0.01  0.10 |
## 62                0.02  0.32 | -0.49  0.00  0.06 |
## 65                0.02  0.26 | -0.57  0.01  0.09 |
## 84                0.02  0.28 | -0.54  0.01  0.08 |
##
## Variables
##                    Dim.1   ctr  cos2   Dim.2   ctr  cos2   Dim.3   ctr  cos2
## age              | -0.01  0.00  0.00 |  0.35  8.93  0.12 |  0.67 40.10  0.44
## campaign         |  0.21  1.13  0.04 |  0.00  0.00  0.00 | -0.23  4.74  0.05
## pdays            |  0.42  4.63  0.18 | -0.74 40.21  0.55 |  0.28  7.10  0.08
## previous         | -0.60  9.37  0.36 |  0.56 22.84  0.31 | -0.30  8.38  0.09
## emp.var.rate     |  0.96 23.97  0.93 |  0.19  2.54  0.03 | -0.11  1.00  0.01
## cons.price.idx   |  0.72 13.43  0.52 |  0.33  8.11  0.11 | -0.30  8.08  0.09
## cons.conf.idx    |  0.20  1.06  0.04 |  0.46 15.36  0.21 |  0.58 30.44  0.34
```

```
## euribor3m       |  0.97 24.20  0.94 |  0.16  1.98  0.03 | -0.01  0.01  0.00
## nr.employed     |  0.93 22.20  0.86 | -0.02  0.03  0.00 | -0.04  0.16  0.00
##                  Dim.4   ctr  cos2  
## age             |  0.28  8.23  0.08 |
## campaign        |  0.93 89.89  0.87 |
## pdays           |  0.04  0.17  0.00 |
## previous        | -0.02  0.05  0.00 |
## emp.var.rate    | -0.06  0.39  0.00 |
## cons.price.idx  | -0.06  0.41  0.00 |
## cons.conf.idx   | -0.04  0.13  0.00 |
## euribor3m       | -0.07  0.51  0.00 |
## nr.employed     | -0.05  0.24  0.00 |
## 
## Supplementary continuous variable
##                  Dim.1  cos2   Dim.2  cos2   Dim.3  cos2   Dim.4  cos2  
## duration        | -0.02  0.00 |  0.02  0.00 |  0.00  0.00 | -0.05  0.00 |

plot(res.pca$eig[,1],main="Eigenvalues",type="o", col="blue")
```

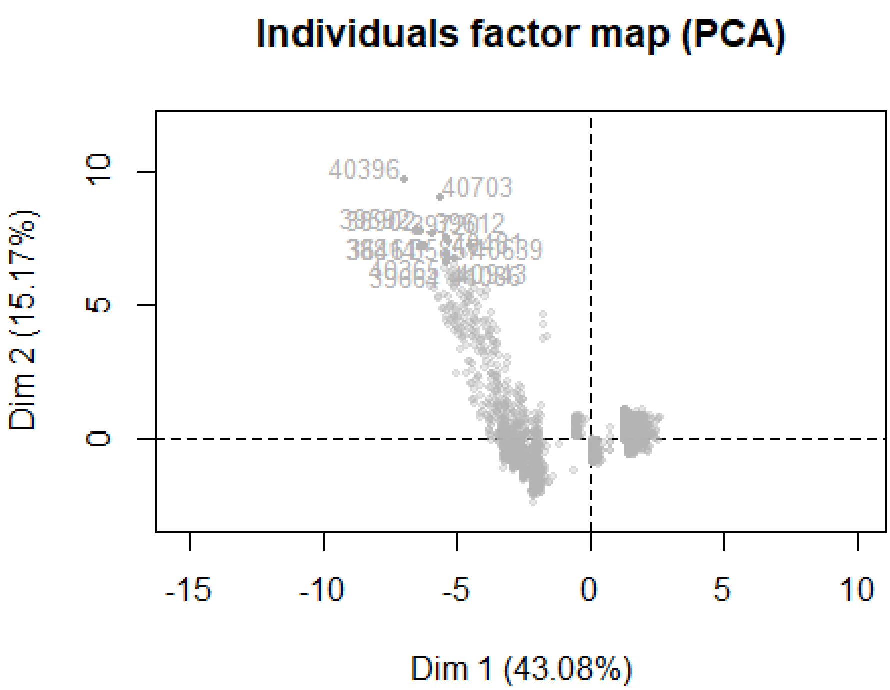


Individuals point of view: Are they any individuals "too contributive"? To better understand the axes meaning use the extreme individuals. Detection of multivariant outliers and influent data.

Primero graficamos en rp los 15 individuos más contributivos en ambos ejes, luego analizamos los 5 individuos más contributivos en la dimensión 1 y 2. Al ver si estos tienen alguna relación significativa, podemos decir que para los 5 individuos de la dimension 1, vemos que principalmente son gente mayor de 45años, todos han comprado el producto, han sido contactados mediante el móvil, han sido contactados previamente, comprado un producto en una campaña anterior y la duración de la llamada ha sido mayor a los 300s.

Para la dimensión 2 podemos ver prácticamente las mismas características menos la duración que ha sido menor. Cabe destacar sin embargo que hay dos individuos que son muy contributivos en ambos ejes, eso hace pensar que pueden ser posibles outliers pero de igual forma los dejamos en los datos.

```r
plot.PCA(res.pca, choix=c("ind"),cex=0.8,col.ind="grey70",select="contrib15",axes=c(1,2))
```




```r
mas_ctr_dim1 <- sort(res.pca$ind$contrib[,1], decreasing = TRUE)[1:5]
mas_ctr_dim2 <- sort(res.pca$ind$contrib[,2], decreasing = TRUE)[1:5]
df[names(mas_ctr_dim1),]
```

```
##       age         job          marital                     education
## 40396  48  job.admin. marital.divorced education.university.degree
## 39592  24 job.student   marital.single       education.high.school
## 38902  83 job.retired marital.divorced          education.basic.4y
## 38814  65 job.retired  marital.married education.university.degree
## 36461  55 job.retired  marital.married          education.basic.4y
##           default     housing    loan          contact     month
## 40396 default.no  housing.no loan.no contact.cellular month.aug
## 39592 default.no housing.yes loan.no contact.cellular month.may
## 38902 default.no  housing.no loan.no contact.cellular month.nov
## 38814 default.no  housing.no loan.no contact.cellular month.nov
```

```
## 36461 default.no  housing.no loan.no contact.cellular month.jun
##           day_of_week duration campaign pdays previous         poutcome
## 40396 day_of_week.thu      172        3     3        6 poutcome.success
## 39592 day_of_week.wed      258        1     3        5 poutcome.success
## 38902 day_of_week.tue      242        1     3        3 poutcome.success
## 38814 day_of_week.fri      226        1     3        3 poutcome.success
## 36461 day_of_week.tue      553        2     3        4 poutcome.failure
##       emp.var.rate cons.price.idx cons.conf.idx euribor3m nr.employed
## 40396         -1.7         94.027         -38.3     0.904      4991.6
## 39592         -1.8         93.876         -40.0     0.672      5008.7
## 38902         -3.4         92.649         -30.1     0.716      5017.5
## 38814         -3.4         92.649         -30.1     0.714      5017.5
## 36461         -2.9         92.963         -40.8     1.262      5076.2
##           y                        f.job         f.season
## 40396 y.yes         f.job.Admin-Managment f.season.Jun-Aug
## 39592 y.yes             f.job.Not-working f.season.Mar-May
## 38902 y.yes f.job.Entrep-Retired-selfEmpl f.season.Sep-Dec
## 38814 y.yes f.job.Entrep-Retired-selfEmpl f.season.Sep-Dec
## 36461  y.no f.job.Entrep-Retired-selfEmpl f.season.Jun-Aug
##                    f.education         f.age               f.duration
## 40396 f.education.Professional f.age-(40,50]     f.duration-(120,180]
## 39592  f.education.High School f.age-[18,30]     f.duration-(180,300]
## 38902        f.education.Basic f.age-(50,92]     f.duration-(180,300]
## 38814 f.education.Professional f.age-(50,92]     f.duration-(180,300]
## 36461        f.education.Basic f.age-(50,92] f.duration-(300,2.1e+03]
##              f.campaign        f.pdays       f.previous
## 40396 f.campaign-(2,20] f.pdays-[0,22] f.previous-(1,6]
## 39592  f.campaign-[0,1] f.pdays-[0,22] f.previous-(1,6]
## 38902  f.campaign-[0,1] f.pdays-[0,22] f.previous-(1,6]
## 38814  f.campaign-[0,1] f.pdays-[0,22] f.previous-(1,6]
## 36461  f.campaign-(1,2] f.pdays-[0,22] f.previous-(1,6]
```

```
df[names(mas_ctr_dim2),]
```

```
##       age            job          marital                   education
## 40396  48     job.admin. marital.divorced education.university.degree
## 40703  82    job.retired  marital.married education.university.degree
## 39592  24    job.student   marital.single       education.high.school
## 38902  83    job.retired marital.divorced          education.basic.4y
## 39612  52 job.technician  marital.married education.university.degree
##               default     housing    loan          contact     month
## 40396      default.no  housing.no loan.no contact.cellular month.aug
## 40703 default.unknown  housing.no loan.no contact.cellular month.sep
## 39592      default.no housing.yes loan.no contact.cellular month.may
## 38902      default.no  housing.no loan.no contact.cellular month.nov
## 39612      default.no  housing.no loan.no contact.cellular month.may
##           day_of_week duration campaign pdays previous         poutcome
## 40396 day_of_week.thu      172        3     3        6 poutcome.success
## 40703 day_of_week.mon       81        3     3        4 poutcome.success
## 39592 day_of_week.wed      258        1     3        5 poutcome.success
```

```
## 38902 day_of_week.tue      242        1     3        3 poutcome.success
## 39612 day_of_week.thu      211        1     3        4 poutcome.success
##       emp.var.rate cons.price.idx cons.conf.idx euribor3m nr.employed
## 40396         -1.7         94.027         -38.3     0.904      4991.6
## 40703         -1.1         94.199         -37.5     0.879      4963.6
## 39592         -1.8         93.876         -40.0     0.672      5008.7
## 38902         -3.4         92.649         -30.1     0.716      5017.5
## 39612         -1.8         93.876         -40.0     0.677      5008.7
##           y                         f.job         f.season
## 40396 y.yes         f.job.Admin-Managment f.season.Jun-Aug
## 40703  y.no f.job.Entrep-Retired-selfEmpl f.season.Sep-Dec
## 39592 y.yes             f.job.Not-working f.season.Mar-May
## 38902 y.yes f.job.Entrep-Retired-selfEmpl f.season.Sep-Dec
## 39612 y.yes         f.job.Serv-Tech-BlueC f.season.Mar-May
##                    f.education         f.age           f.duration
## 40396 f.education.Professional f.age-(40,50] f.duration-(120,180]
## 40703 f.education.Professional f.age-(50,92]   f.duration-[5,120]
## 39592  f.education.High School f.age-[18,30] f.duration-(180,300]
## 38902        f.education.Basic f.age-(50,92] f.duration-(180,300]
## 39612 f.education.Professional f.age-(50,92] f.duration-(180,300]
##              f.campaign        f.pdays       f.previous
## 40396 f.campaign-(2,20] f.pdays-[0,22] f.previous-(1,6]
## 40703 f.campaign-(2,20] f.pdays-[0,22] f.previous-(1,6]
## 39592  f.campaign-[0,1] f.pdays-[0,22] f.previous-(1,6]
## 38902  f.campaign-[0,1] f.pdays-[0,22] f.previous-(1,6]
## 39612  f.campaign-[0,1] f.pdays-[0,22] f.previous-(1,6]
```

Interpreting the axes: Variables point of view coordinates, quality of representation, contribution of the variables

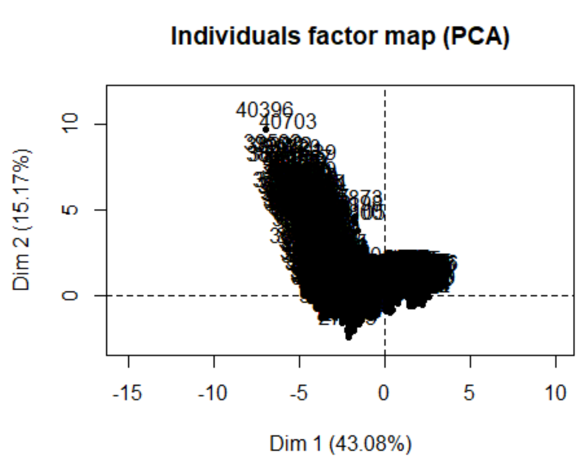
Al hacer el PCA con la variable target duration como suplementaria, podemos ver que su módulo es prácticamente nulo, esto quiere decir que la variable no se ve representada en niguno de los ejes factoriales.

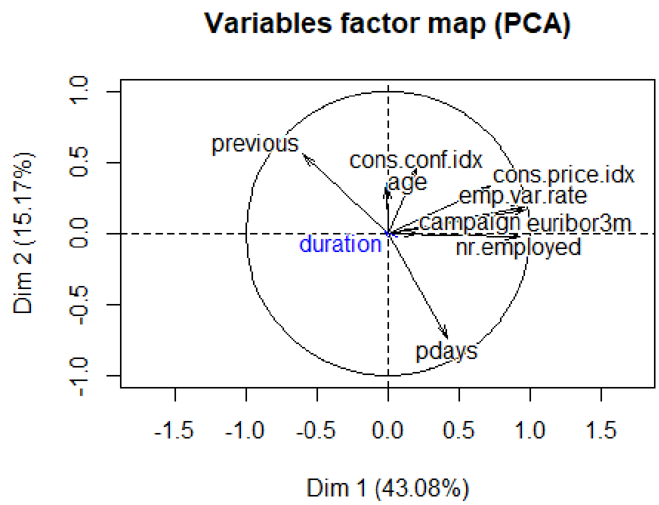
El eje horizontal está muy relacionado con las variables socio económicas, mirando el cos2 del summary podemos ver que las variables que están mejor representadas con la dimensión 1 son: euribor3m, emp.var.rate, nr.emplyed Para el eje vertical: pdays y previous

Para el eje vertical, podemos decir que está relacionado con las campañas previas.

Al hacer el PCA con la variable target Y como suplementaria, podemos ver que en el gráfico de rp, el factor NO, esta muy cerca del centro, por lo que no se ve representada en estos ejes factoriales. En cambio el factor SI, está a una distancia mayor del centro, aunque poco significativa.

```
res.pca<-PCA(df[,c('duration',vars_num)],quanti.sup=1)
```

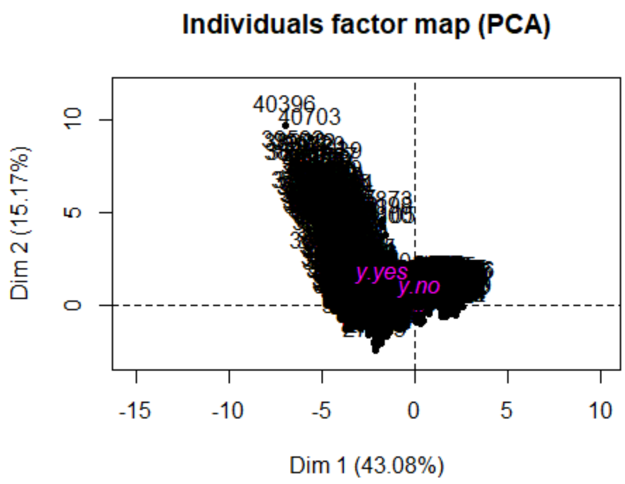


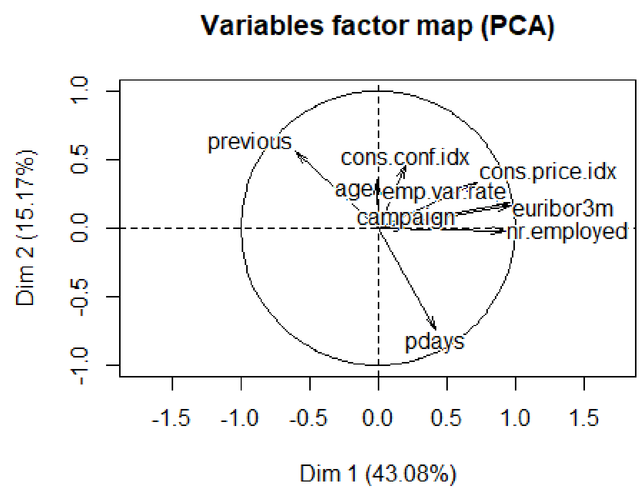



```
summary(res.pca,ncp=4,nb.dec=2)

##
## Call:
## PCA(X = df[, c("duration", vars_num)], quanti.sup = 1)
##
##
## Eigenvalues
##                        Dim.1  Dim.2  Dim.3  Dim.4  Dim.5  Dim.6  Dim.7
## Variance                3.88   1.36   1.10   0.97   0.83   0.43   0.39
## % of var.              43.08  15.17  12.28  10.74   9.25   4.82   4.29
## Cumulative % of var.   43.08  58.24  70.52  81.26  90.50  95.33  99.62
##                        Dim.8  Dim.9
## Variance                0.02   0.01
## % of var.               0.27   0.12
## Cumulative % of var.   99.88 100.00
##
## Individuals (the 10 first)
##                    Dist    Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3
## 20               |  1.74 |  1.29   0.01   0.55 |  0.44   0.00   0.06 |  0.41
## 21               |  1.99 |  1.30   0.01   0.42 |  0.18   0.00   0.01 | -0.13
## 30               |  2.24 |  1.28   0.01   0.33 |  0.90   0.01   0.16 |  1.38
## 33               |  1.93 |  1.29   0.01   0.44 |  0.73   0.01   0.14 |  1.02
## 48               |  1.74 |  1.29   0.01   0.55 |  0.47   0.00   0.07 |  0.47
## 56               |  2.24 |  1.28   0.01   0.33 |  0.90   0.01   0.16 |  1.38
## 61               |  1.85 |  1.29   0.01   0.48 |  0.67   0.01   0.13 |  0.90
## 62               |  2.02 |  1.29   0.01   0.41 |  0.78   0.01   0.15 |  1.14
## 65               |  1.89 |  1.29   0.01   0.46 |  0.70   0.01   0.14 |  0.96
## 84               |  1.93 |  1.29   0.01   0.44 |  0.73   0.01   0.14 |  1.02
##                     ctr   cos2    Dim.4    ctr   cos2
## 20                 0.00   0.06 | -0.81   0.01   0.22 |
## 21                 0.00   0.00 | -1.06   0.02   0.28 |
## 30                 0.03   0.38 | -0.38   0.00   0.03 |
```

```
## 33               0.02  0.28 | -0.54  0.01  0.08 |
## 48               0.00  0.07 | -0.79  0.01  0.21 |
## 56               0.03  0.38 | -0.38  0.00  0.03 |
## 61               0.01  0.23 | -0.60  0.01  0.10 |
## 62               0.02  0.32 | -0.49  0.00  0.06 |
## 65               0.02  0.26 | -0.57  0.01  0.09 |
## 84               0.02  0.28 | -0.54  0.01  0.08 |
##
## Variables
##                   Dim.1    ctr   cos2   Dim.2    ctr   cos2   Dim.3    ctr   cos2
## age             | -0.01  0.00  0.00 |  0.35  8.93  0.12 |  0.67 40.10  0.44
## campaign        |  0.21  1.13  0.04 |  0.00  0.00  0.00 | -0.23  4.74  0.05
## pdays           |  0.42  4.63  0.18 | -0.74 40.21  0.55 |  0.28  7.10  0.08
## previous        | -0.60  9.37  0.36 |  0.56 22.84  0.31 | -0.30  8.38  0.09
## emp.var.rate    |  0.96 23.97  0.93 |  0.19  2.54  0.03 | -0.11  1.00  0.01
## cons.price.idx |  0.72 13.43  0.52 |  0.33  8.11  0.11 | -0.30  8.08  0.09
## cons.conf.idx   |  0.20  1.06  0.04 |  0.46 15.36  0.21 |  0.58 30.44  0.34
## euribor3m       |  0.97 24.20  0.94 |  0.16  1.98  0.03 | -0.01  0.01  0.00
## nr.employed     |  0.93 22.20  0.86 | -0.02  0.03  0.00 | -0.04  0.16  0.00
##                   Dim.4    ctr   cos2
## age             |  0.28  8.23  0.08 |
## campaign        |  0.93 89.89  0.87 |
## pdays           |  0.04  0.17  0.00 |
## previous        | -0.02  0.05  0.00 |
## emp.var.rate    | -0.06  0.39  0.00 |
## cons.price.idx | -0.06  0.41  0.00 |
## cons.conf.idx   | -0.04  0.13  0.00 |
## euribor3m       | -0.07  0.51  0.00 |
## nr.employed     | -0.05  0.24  0.00 |
##
## Supplementary continuous variable
##                   Dim.1   cos2    Dim.2   cos2    Dim.3   cos2    Dim.4   cos2
## duration        | -0.02  0.00 |  0.02  0.00 |  0.00  0.00 | -0.05  0.00 |
```

```
res.pca<-PCA(df[,c('y',vars_num)],quali.sup=1)
```






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

##
## Call:
## PCA(X = df[, c("y", vars_num)], quali.sup = 1)
##
##
## Eigenvalues
##                        Dim.1  Dim.2  Dim.3  Dim.4  Dim.5  Dim.6  Dim.7
## Variance                3.88   1.36   1.10   0.97   0.83   0.43   0.39
## % of var.              43.08  15.17  12.28  10.74   9.25   4.82   4.29
## Cumulative % of var.   43.08  58.24  70.52  81.26  90.50  95.33  99.62
##                        Dim.8  Dim.9
## Variance                0.02   0.01
## % of var.               0.27   0.12
## Cumulative % of var.   99.88 100.00
##
## Individuals (the 10 first)
##                    Dist    Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3
## 20               |  1.74 |  1.29   0.01   0.55 |  0.44   0.00   0.06 |  0.41
## 21               |  1.99 |  1.30   0.01   0.42 |  0.18   0.00   0.01 | -0.13
## 30               |  2.24 |  1.28   0.01   0.33 |  0.90   0.01   0.16 |  1.38
## 33               |  1.93 |  1.29   0.01   0.44 |  0.73   0.01   0.14 |  1.02
## 48               |  1.74 |  1.29   0.01   0.55 |  0.47   0.00   0.07 |  0.47
## 56               |  2.24 |  1.28   0.01   0.33 |  0.90   0.01   0.16 |  1.38
## 61               |  1.85 |  1.29   0.01   0.48 |  0.67   0.01   0.13 |  0.90
## 62               |  2.02 |  1.29   0.01   0.41 |  0.78   0.01   0.15 |  1.14
## 65               |  1.89 |  1.29   0.01   0.46 |  0.70   0.01   0.14 |  0.96
## 84               |  1.93 |  1.29   0.01   0.44 |  0.73   0.01   0.14 |  1.02
##                   ctr   cos2    Dim.4    ctr   cos2
## 20               0.00   0.06 | -0.81   0.01   0.22 |
```

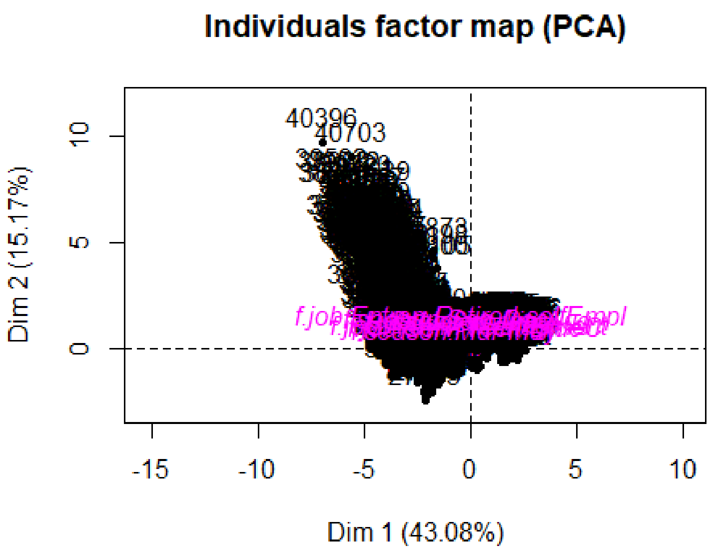
```
## 21               0.00  0.00 | -1.06  0.02  0.28 |
## 30               0.03  0.38 | -0.38  0.00  0.03 |
## 33               0.02  0.28 | -0.54  0.01  0.08 |
## 48               0.00  0.07 | -0.79  0.01  0.21 |
## 56               0.03  0.38 | -0.38  0.00  0.03 |
## 61               0.01  0.23 | -0.60  0.01  0.10 |
## 62               0.02  0.32 | -0.49  0.00  0.06 |
## 65               0.02  0.26 | -0.57  0.01  0.09 |
## 84               0.02  0.28 | -0.54  0.01  0.08 |
##
## Variables
##                   Dim.1   ctr   cos2   Dim.2   ctr   cos2   Dim.3   ctr   cos2
## age             | -0.01  0.00  0.00 |  0.35  8.93  0.12 |  0.67 40.10  0.44
## campaign        |  0.21  1.13  0.04 |  0.00  0.00  0.00 | -0.23  4.74  0.05
## pdays           |  0.42  4.63  0.18 | -0.74 40.21  0.55 |  0.28  7.10  0.08
## previous        | -0.60  9.37  0.36 |  0.56 22.84  0.31 | -0.30  8.38  0.09
## emp.var.rate    |  0.96 23.97  0.93 |  0.19  2.54  0.03 | -0.11  1.00  0.01
## cons.price.idx |  0.72 13.43  0.52 |  0.33  8.11  0.11 | -0.30  8.08  0.09
## cons.conf.idx  |  0.20  1.06  0.04 |  0.46 15.36  0.21 |  0.58 30.44  0.34
## euribor3m       |  0.97 24.20  0.94 |  0.16  1.98  0.03 | -0.01  0.01  0.00
## nr.employed     |  0.93 22.20  0.86 | -0.02  0.03  0.00 | -0.04  0.16  0.00
##                   Dim.4   ctr   cos2
## age             |  0.28  8.23  0.08 |
## campaign        |  0.93 89.89  0.87 |
## pdays           |  0.04  0.17  0.00 |
## previous        | -0.02  0.05  0.00 |
## emp.var.rate    | -0.06  0.39  0.00 |
## cons.price.idx | -0.06  0.41  0.00 |
## cons.conf.idx  | -0.04  0.13  0.00 |
## euribor3m       | -0.07  0.51  0.00 |
## nr.employed     | -0.05  0.24  0.00 |
##
## Supplementary categories
##                     Dist    Dim.1    cos2 v.test    Dim.2    cos2 v.test
## y.no            |    0.23 |   0.21    0.84  21.77 |  -0.08    0.13 -14.15 |
## y.yes           |    1.89 |  -1.74    0.84 -21.77 |   0.67    0.13  14.15 |
##                  Dim.3    cos2 v.test    Dim.4    cos2 v.test
## y.no            -0.01    0.00  -1.74 |   0.01    0.00   1.09 |
## y.yes            0.07    0.00   1.74 |  -0.04    0.00  -1.09 |
```

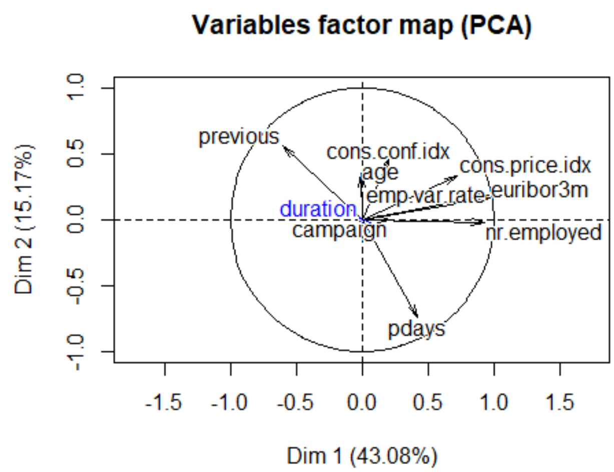
Perform a PCA taking into account also supplementary variables the supplementary variables can be quantitative and/or categorical

Hemos dividido el plot en diferentes partes, para así poder entender y ver mejor el resultado. Para la primera dimensión podemos ver que para los niveles mejor representados son: f.season.jun-aug, f.previous-(0.9,1], y.no, y.yes.

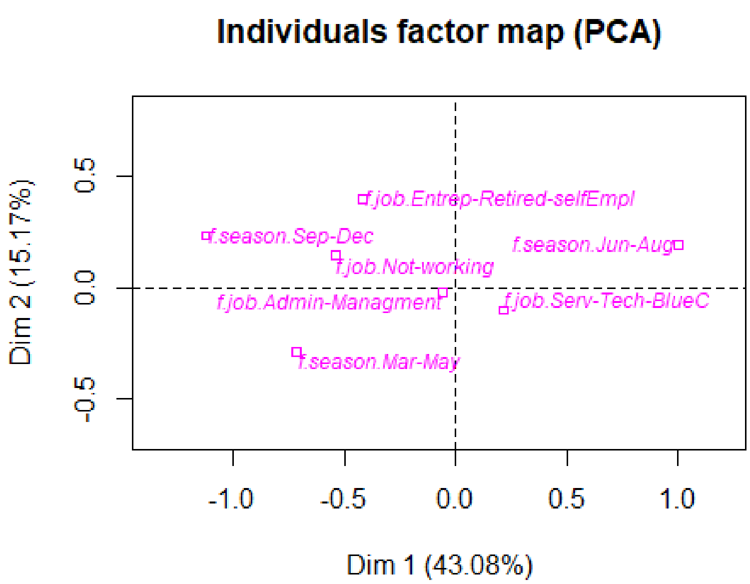
Para la segunda dimsensión, las que se ven mejor representadas son: f.pdays-[0,22],f.pdays-(22,23], f.previous-(1,6]

```
vars_factorizadas<- c("f.job","f.season","f.education","f.age","f.duration","f.c
ampaign","f.pdays","f.previous","y");

res.pca<-PCA(df[,c('duration',vars_num, "f.job","f.season")],quanti.sup=1, quali
.sup = c(11:12))
```

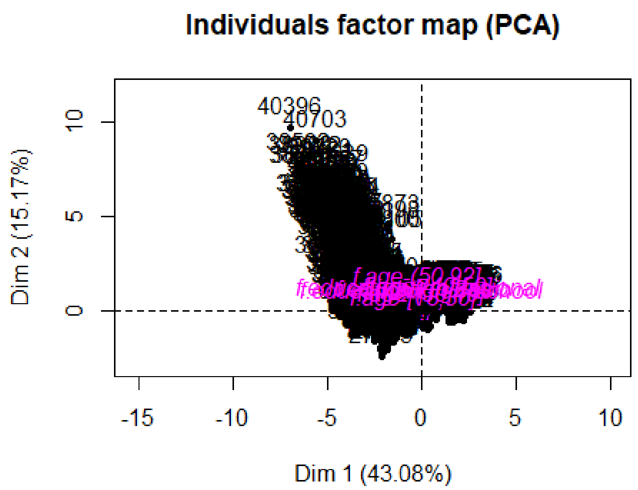


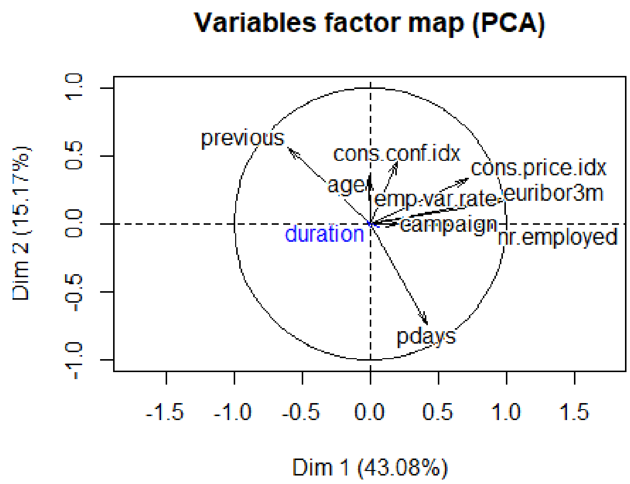



```
plot.PCA(res.pca,choix="ind",invisible="ind",cex=0.75)
```

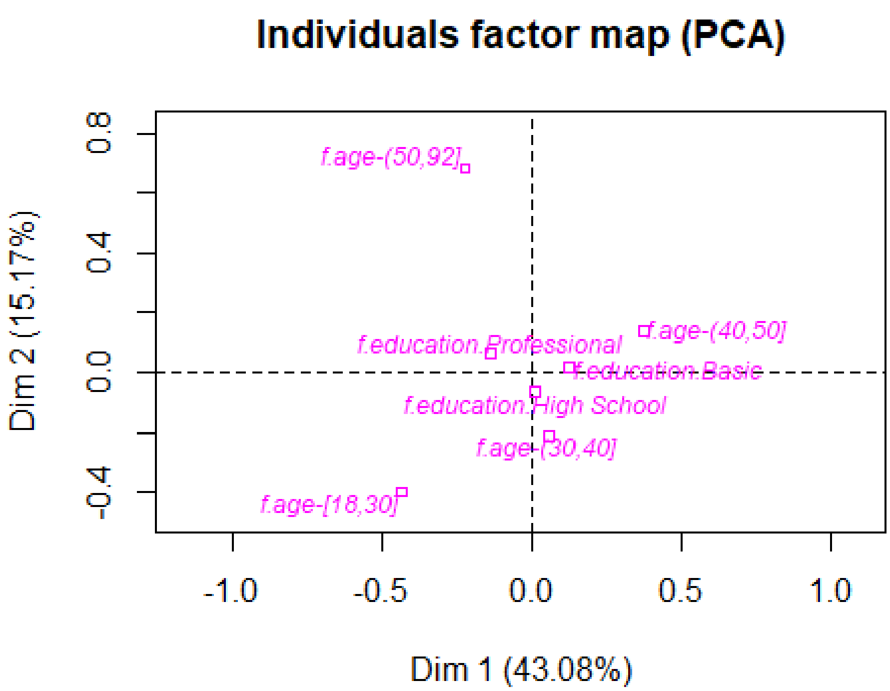



```
res.pca<-PCA(df[,c('duration',vars_num,"f.education","f.age")],quanti.sup=1, qua
li.sup = c(11:12))
```

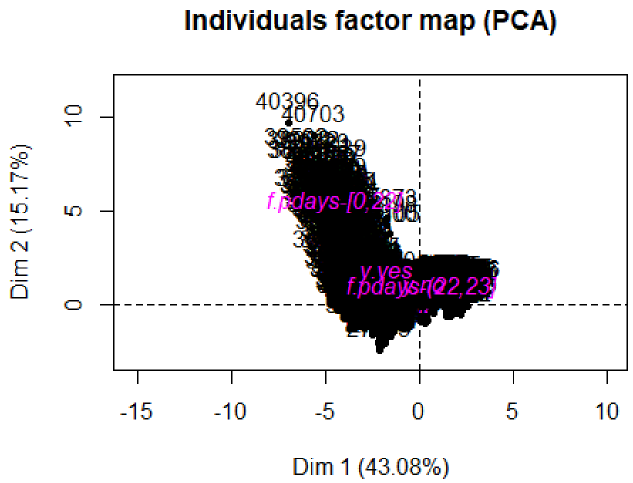


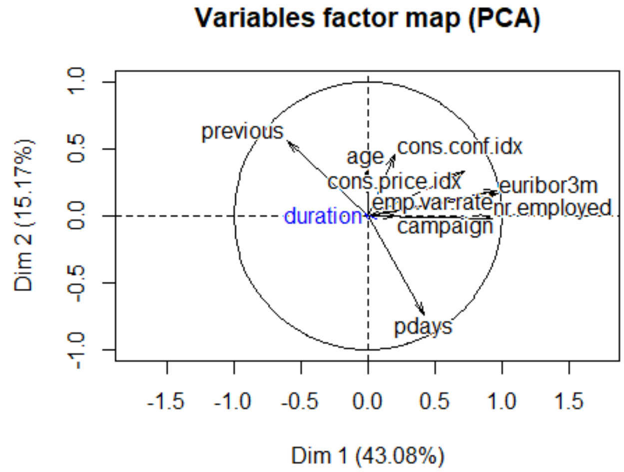



```
plot.PCA(res.pca,choix="ind",invisible="ind",cex=0.75)
```

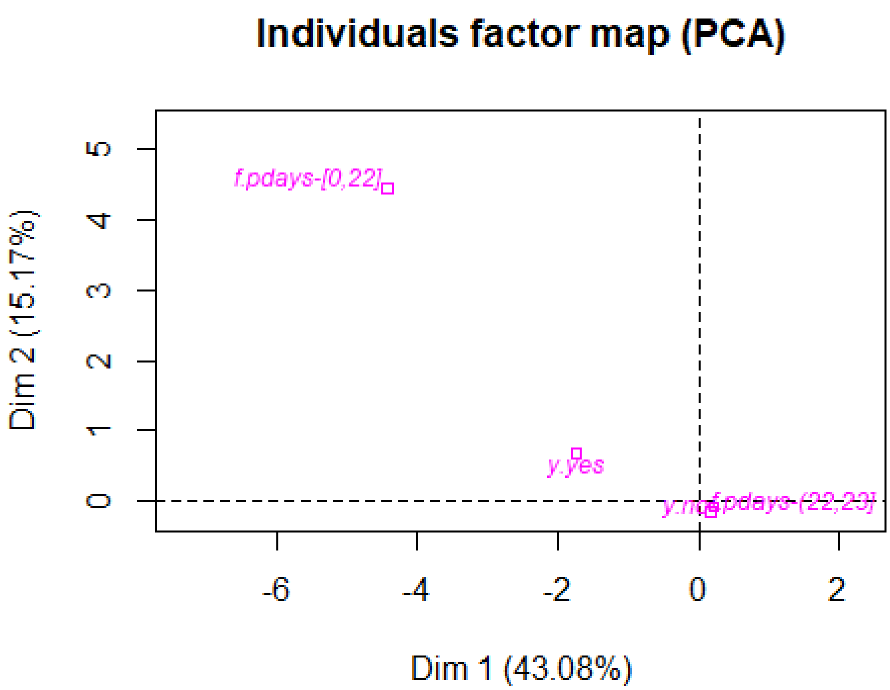


```
res.pca<-PCA(df[,c('duration',vars_num,"f.pdays","y")],quanti.sup=1, quali.sup =
c(11:12))
```

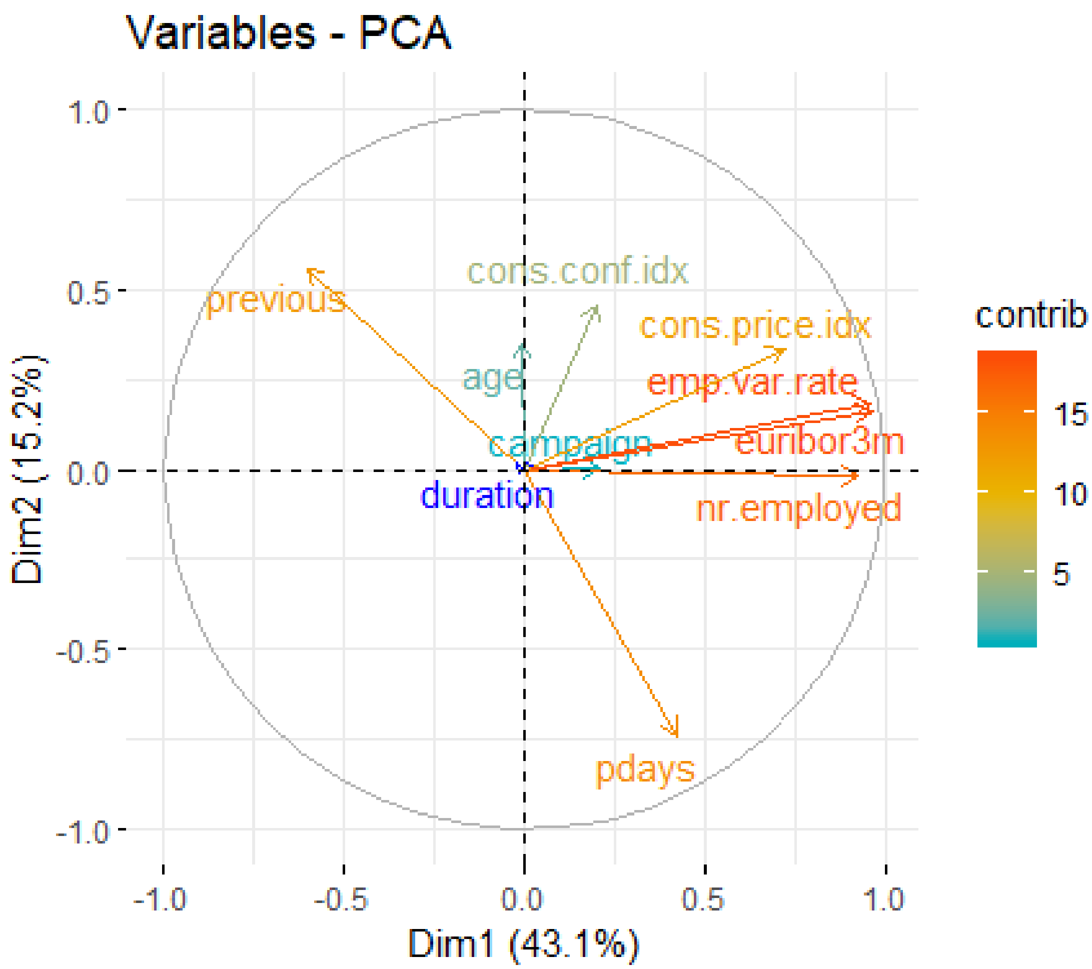





```
plot.PCA(res.pca,choix="ind",invisible="ind",cex=0.75)
```




```
res.pca<-PCA(df[,c('duration',vars_num,vars_factorizadas)],quanti.sup=1, quali.s
up = c(11:19),graph=FALSE)

fviz_pca_var(res.pca, col.var = "contrib",gradient.cols = c("#00AFBB", "#E7B800"
, "#FC4E07"),repel = TRUE)
```




```
summary(res.pca,dig = 2, nbelements= 30, nbind=3, ncp=2)

##
## Call:
## PCA(X = df[, c("duration", vars_num, vars_factorizadas)], quanti.sup = 1,
##      quali.sup = c(11:19), graph = FALSE)
##
##
## Eigenvalues
##                        Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6
## Variance               3.877   1.365   1.105   0.966   0.832   0.434
## % of var.             43.078  15.166  12.276  10.736   9.246   4.823
## Cumulative % of var.  43.078  58.244  70.521  81.257  90.503  95.326
##                        Dim.7   Dim.8   Dim.9
## Variance               0.386   0.024   0.010
## % of var.              4.291   0.266   0.116
## Cumulative % of var.  99.618  99.884 100.000
##
## Individuals (the 3 first)
##                                 Dist    Dim.1    ctr   cos2    Dim.2    ctr
## 20                           | 1.739 | 1.291 0.009 0.551 | 0.441 0.003
```

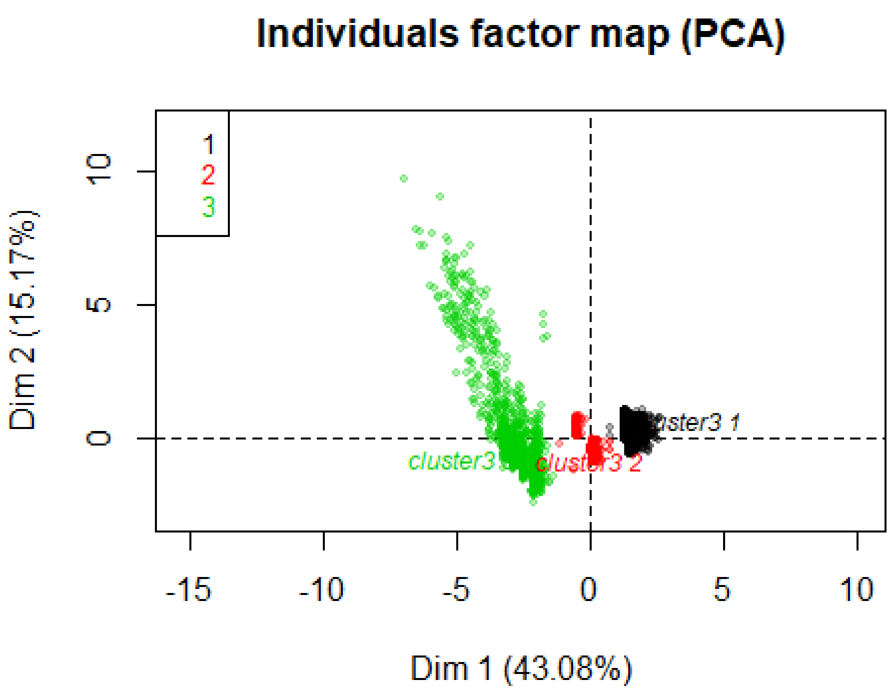
```
## 21                             | 1.990 | 1.295 0.009 0.424 | 0.184 0.000
## 30                             | 2.240 | 1.284 0.009 0.328 | 0.898 0.012
##                                 cos2
## 20                             0.064 |
## 21                             0.009 |
## 30                             0.161 |
##
## Variables
##                                   Dim.1    ctr   cos2    Dim.2    ctr
## age                             | -0.009  0.002  0.000 |  0.349  8.933
## campaign                        |  0.210  1.132  0.044 |  0.003  0.001
## pdays                           |  0.424  4.633  0.180 | -0.741 40.212
## previous                        | -0.603  9.366  0.363 |  0.558 22.836
## emp.var.rate                    |  0.964 23.967  0.929 |  0.186  2.544
## cons.price.idx                  |  0.722 13.431  0.521 |  0.333  8.109
## cons.conf.idx                   |  0.203  1.063  0.041 |  0.458 15.356
## euribor3m                       |  0.969 24.202  0.938 |  0.164  1.978
## nr.employed                     |  0.928 22.205  0.861 | -0.020  0.030
##                                  cos2
## age                             0.122 |
## campaign                        0.000 |
## pdays                           0.549 |
## previous                        0.312 |
## emp.var.rate                    0.035 |
## cons.price.idx                  0.111 |
## cons.conf.idx                   0.210 |
## euribor3m                       0.027 |
## nr.employed                     0.000 |
##
## Supplementary continuous variable
##                                   Dim.1   cos2    Dim.2   cos2
## duration                        | -0.022  0.000 |  0.017  0.000 |
##
## Supplementary categories
##                                     Dist     Dim.1    cos2  v.test
## f.job.Admin-Managment           |   0.125 |  -0.061   0.237  -1.532 |
## f.job.Entrep-Retired-selfEmpl   |   1.020 |  -0.418   0.168  -5.418 |
## f.job.Not-working               |   0.690 |  -0.541   0.615  -5.266 |
## f.job.Serv-Tech-BlueC           |   0.272 |   0.217   0.637   7.570 |
## f.season.Mar-May                |   0.793 |  -0.719   0.821 -21.657 |
## f.season.Jun-Aug                |   1.026 |   1.002   0.954  32.508 |
## f.season.Sep-Dec                |   1.257 |  -1.125   0.801 -16.040 |
## f.education.Basic               |   0.294 |   0.125   0.181   3.071 |
## f.education.High School         |   0.155 |   0.010   0.005   0.285 |
## f.education.Professional        |   0.255 |  -0.140   0.300  -3.384 |
## f.age-[18,30]                   |   1.312 |  -0.436   0.111  -7.188 |
## f.age-(30,40]                   |   0.485 |   0.055   0.013   1.600 |
## f.age-(40,50]                   |   0.606 |   0.373   0.380   7.755 |
## f.age-(50,92]                   |   1.667 |  -0.225   0.018  -3.730 |
## f.duration-[5,120]              |   0.274 |   0.190   0.483   4.596 |
```

```
## f.duration-(120,180]          |   0.138 |   0.076   0.299   1.332 |
## f.duration-(180,300]          |   0.222 |  -0.180   0.658  -3.418 |
## f.duration-(300,2.1e+03]      |   0.166 |  -0.126   0.571  -2.781 |
## f.campaign-[0,1]              |   0.707 |  -0.356   0.254 -10.989 |
## f.campaign-(1,2]              |   0.221 |  -0.006   0.001  -0.127 |
## f.campaign-(2,20]             |   1.083 |   0.474   0.192  11.738 |
## f.pdays-[0,22]                |   6.580 |  -4.417   0.451 -30.206 |
## f.pdays-(22,23]               |   0.239 |   0.161   0.451  30.206 |
## f.previous-[0,0.9]            |   0.543 |   0.487   0.804  43.179 |
## f.previous-(0.9,1]            |   2.787 |  -2.633   0.893 -33.713 |
## f.previous-(1,6]              |   6.104 |  -4.396   0.519 -26.498 |
## y.no                          |   0.231 |   0.212   0.844  21.769 |
## y.yes                         |   1.892 |  -1.738   0.844 -21.769 |
##                                Dim.2    cos2  v.test
## f.job.Admin-Managment         -0.020   0.027  -0.869 |
## f.job.Entrep-Retired-selfEmpl  0.399   0.153   8.730 |
## f.job.Not-working              0.143   0.043   2.353 |
## f.job.Serv-Tech-BlueC         -0.101   0.139  -5.958 |
## f.season.Mar-May              -0.288   0.132 -14.633 |
## f.season.Jun-Aug               0.194   0.036  10.604 |
## f.season.Sep-Dec               0.234   0.035   5.626 |
## f.education.Basic              0.015   0.003   0.613 |
## f.education.High School       -0.066   0.180  -3.025 |
## f.education.Professional       0.062   0.059   2.527 |
## f.age-[18,30]                 -0.400   0.093 -11.108 |
## f.age-(30,40]                 -0.213   0.193 -10.495 |
## f.age-(40,50]                  0.139   0.053   4.874 |
## f.age-(50,92]                  0.682   0.168  19.031 |
## f.duration-[5,120]            -0.094   0.118  -3.822 |
## f.duration-(120,180]           0.014   0.010   0.414 |
## f.duration-(180,300]           0.085   0.147   2.721 |
## f.duration-(300,2.1e+03]       0.029   0.030   1.081 |
## f.campaign-[0,1]               0.003   0.000   0.167 |
## f.campaign-(1,2]               0.014   0.004   0.486 |
## f.campaign-(2,20]             -0.015   0.000  -0.627 |
## f.pdays-[0,22]                 4.435   0.454  51.122 |
## f.pdays-(22,23]               -0.161   0.454 -51.122 |
## f.previous-[0,0.9]            -0.181   0.111 -27.027 |
## f.previous-(0.9,1]             0.442   0.025   9.534 |
## f.previous-(1,6]               3.839   0.396  38.998 |
## y.no                          -0.082   0.126 -14.153 |
## y.yes                          0.671   0.126  14.153 |
```

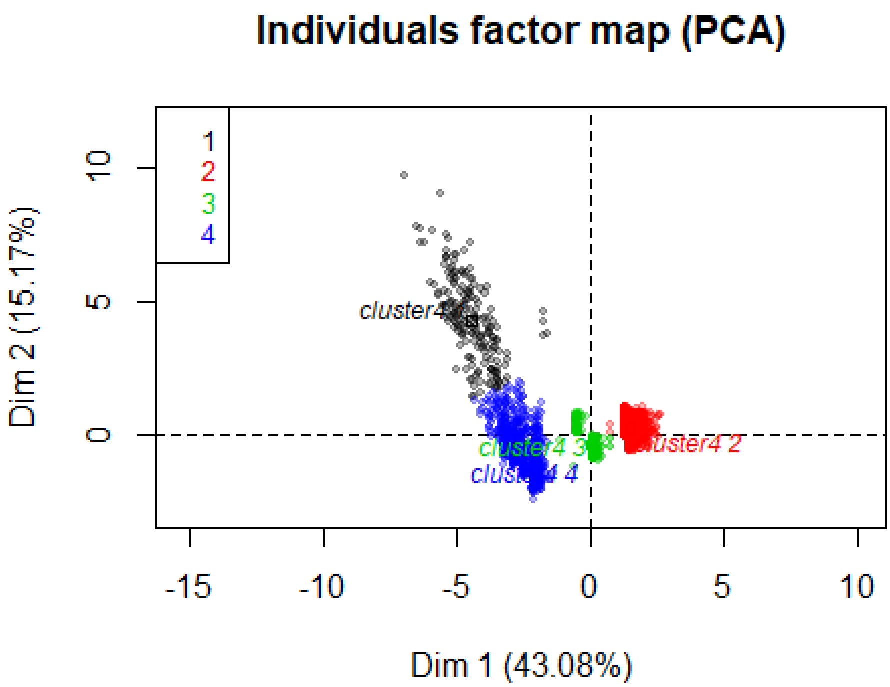
K-Means Classification

Hemos graficado los grupos separados en 3, 4, 5 y 6 clusters, para los cuales nos parece que gráficamente con 4 clusters los grupos están bien definidos, por lo que decidimos usar 4 clusters los cuales analizaremos seguidamente.

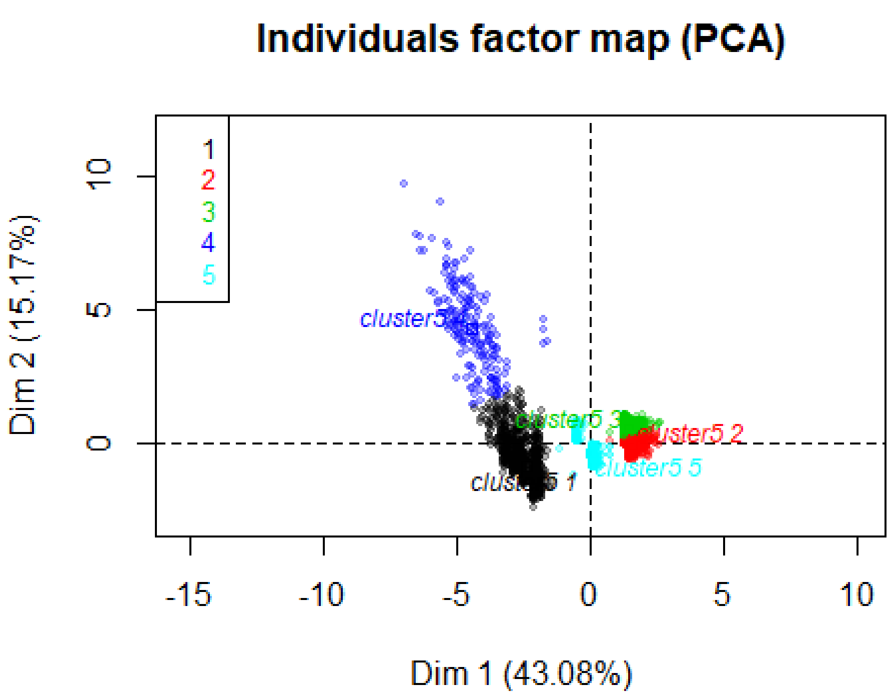
```
dclu<- res.pca$ind$coord[,1:2]; # los dos ejes

kcla<- kmeans(dclu,4);

df$cluster3 = factor(kmeans(dclu,3)$cluster);
df$cluster4 = factor(kmeans(dclu,4)$cluster);
df$cluster5 = factor(kmeans(dclu,5)$cluster);
df$cluster6 = factor(kmeans(dclu,6)$cluster);

res.pca<-PCA(df[,c('duration',vars_num, "cluster3")],quanti.sup=1, quali.sup = 1
1, graph=FALSE)
plot.PCA(res.pca,choix="ind",habillage=11,select=0 ,cex=0.75)
```



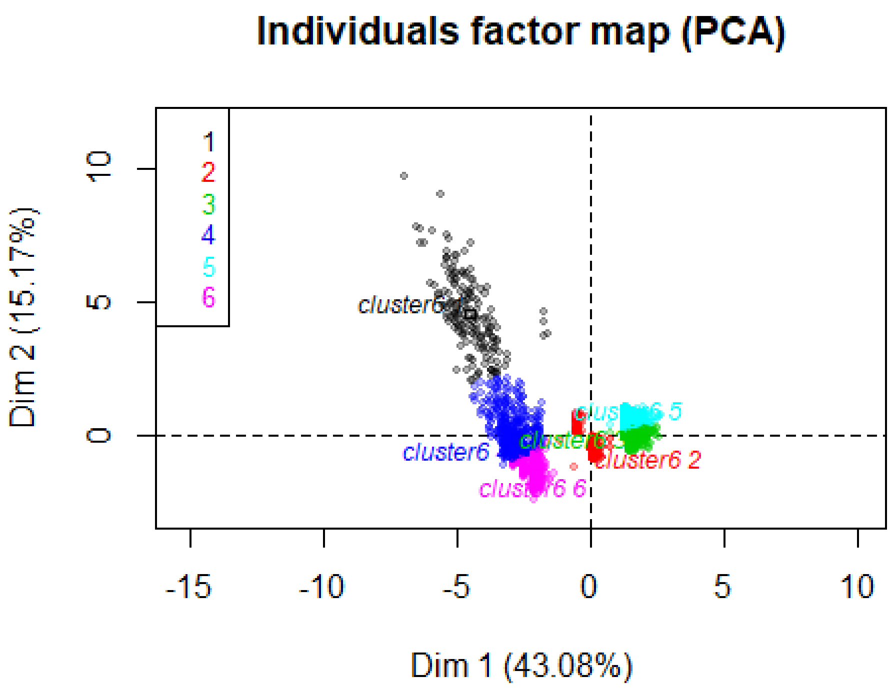

```
res.pca<-PCA(df[,c('duration',vars_num, "cluster4")],quanti.sup=1, quali.sup = 1
1, graph=FALSE)
plot.PCA(res.pca,choix="ind",habillage=11,select=0 ,cex=0.75)
```




```
res.pca<-PCA(df[,c('duration',vars_num, "cluster5")],quanti.sup=1, quali.sup = 1
1, graph=FALSE)
plot.PCA(res.pca,choix="ind",habillage=11,select=0 ,cex=0.75)
```



```
res.pca<-PCA(df[,c('duration',vars_num, "cluster6")],quanti.sup=1, quali.sup = 1
1, graph=FALSE)
plot.PCA(res.pca,choix="ind",habillage=11,select=0 ,cex=0.75)
```




```
df <- df[,c(1:29, 31)] # guardamos la clasificación en 4 clusters
```

Description of clusters

Viendo el chi-square test, podemos saber qué variables se utilizarán para caracterizar nuestros 4 clusters. Viendo las categorías donde el P-value es casi 0 podemos ver que las categorías que cumplen estas caracteristicas son: month (y por extensión también f.season), poutcome, f.pdays, f.previous, contact, y, job(f.job), default, f.age, f.campaign, marital. Ahora veremos que categorías de estas variables son las que caracterizan estos clusters.

Para el catdes del cluster 1, hemos podido ver las categorías que mejor lo definen, la temporada de verano es la que mejor lo caracteriza, f.season.Jun-Aug, tenemos también un poutcome que nos indica que ninguno de los individuos ha sido contactado previamente, podemos ver que y.no representa el 95% de este cluster, han sido contactados en su mayoría por teléfono fijo, también podemos ver que este cluster está ligeramente relacionado con la categoría f.job.Serv-Tech-BlueC. En conclusión podemos decir que este cluster está caracterizado por: - Meses de jun-Ago - No han sido contactados previamente - Han sido contactados más de una vez en la campaña actual - Contactados por teléfono fijo - Trabajo normalmente es, servicio, técnicos o blue collar. - No compraron el producto en su mayoría.

Para el cluster 2, tenemos: - La temporada de mar-may están sobrerepresentadas en este cluster - Han sido contacados en su mayoría por teléfono móvil - No han comprado el producto en campañas anteriores - Han sido contacados en camapañas previas - La categoría student está sobrerepresentada - La aceptación del producto está sobrerepresentada también

Para el cluster 3, tenemos: - Temporada de Sep-Dec - Contactados por móvil en su mayoría - La categoría de job.management está sobrerepresentada - No han adquirido el producto - Una gran cantidad de individuos rechazó el producto (y.no)

Para el cluster 4, podemos ver que aglutina individuos muy bien caracterizados por las siguientes variables: - Han sido contactados previamente f.pdays[0,22] - Han comprado el producto en una campaña previa - Han comprado el producto y.yes - Temporada de Sep-Dec - Han sido contactados por móvil - Una parte importante son job.retired - Una edad de f.age-(50,92]

```
catdes(df, 30, proba = 0.001)
```

```
##
## Link between the cluster variable and the categorical variables (chi-square t
est)
## ============================================================================
====
##                   p.value df
## month        0.000000e+00 27
## poutcome     0.000000e+00  6
## f.season     0.000000e+00  6
## f.pdays      0.000000e+00  3
## f.previous   0.000000e+00  6
## contact     4.750967e-213  3
## y           3.629487e-144  3
## job          4.108692e-67 30
## default      6.986972e-43  3
## f.age        1.030850e-26  9
## f.campaign   6.092760e-24  6
## f.job        1.806095e-19  9
## marital      1.705029e-16  6
## f.duration   1.919242e-10  9
## education    5.661192e-09 18
## f.education  8.669827e-09  6
## housing      1.403613e-07  3
## day_of_week  4.039589e-04 12
##
## Description of each cluster by the categories
## =============================================
## $`1`
##                                        Cla/Mod   Mod/Cla    Global
## f.pdays=f.pdays-[0,22]              96.5714286 88.947368  3.507014
## poutcome=poutcome.success           96.1538462 78.947368  3.126253
## f.previous=f.previous-(1,6]         75.1824818 54.210526  2.745491
## y=y.yes                             21.0332103 60.000000 10.861723
## f.previous=f.previous-(0.9,1]       15.4255319 45.789474 11.302605
## f.season=f.season.Sep-Dec           11.1600587 40.000000 13.647295
```

```
## month=month.oct                       32.5000000 13.684211  1.603206
## contact=contact.cellular               5.4011864 91.052632 64.188377
## default=default.no                     4.6954315 97.368421 78.957916
## job=job.retired                       15.7657658 18.421053  4.448898
## month=month.sep                       25.7142857  9.473684  1.402806
## f.age=f.age-(50,92]                    7.4200913 34.210526 17.555110
## month=month.mar                       20.6349206  6.842105  1.262525
## job=job.student                       15.5963303  8.947368  2.184369
## f.job=f.job.Entrep-Retired-selfEmpl    7.6256499 23.157895 11.563126
## f.duration=f.duration-(180,300]        6.0550459 34.736842 21.843687
## poutcome=poutcome.failure              7.3394495 21.052632 10.921844
## f.education=f.education.Professional   5.4347826 44.736842 31.342685
## education=education.university.degree  5.4347826 44.736842 31.342685
## f.job=f.job.Not-working                8.1871345 14.736842  6.853707
## f.campaign=f.campaign-[0,1]            4.9504950 55.263158 42.505010
## month=month.jun                        1.6233766  5.263158 12.344689
## f.education=f.education.Basic          2.5046963 21.052632 32.004008
## f.season=f.season.Mar-May              2.7131783 29.473684 41.362725
## marital=marital.married                3.0040053 47.368421 60.040080
## f.age=f.age-(40,50]                    1.9952115 13.157895 25.110220
## f.season=f.season.Jun-Aug              2.5835189 30.526316 44.989980
## education=education.basic.9y           1.3580247  5.789474 16.232465
## f.campaign=f.campaign-(2,20]           2.0496894 17.368421 32.264529
## month=month.jul                        1.1560694  5.263158 17.334669
## month=month.may                        1.8452381 16.315789 33.667335
## f.duration=f.duration-[5,120]          1.7341040 14.210526 31.202405
## f.job=f.job.Serv-Tech-BlueC            1.9801980 25.263158 48.577154
## job=job.blue-collar                    0.8650519  5.263158 23.166333
## default=default.unknown                0.4761905  2.631579 21.042084
## contact=contact.telephone              0.9513151  8.947368 35.811623
## y=y.no                                 1.7086331 40.000000 89.138277
## f.previous=f.previous-[0,0.9]          0.0000000  0.000000 85.951904
## poutcome=poutcome.nonexistent          0.0000000  0.000000 85.951904
## f.pdays=f.pdays-(22,23]                0.4361371 11.052632 96.492986
##                                            p.value     v.test
## f.pdays=f.pdays-[0,22]                1.259643e-281  35.860546
## poutcome=poutcome.success             3.796871e-240  33.092604
## f.previous=f.previous-(1,6]           1.553305e-129  24.214815
## y=y.yes                                2.494385e-64  16.934745
## f.previous=f.previous-(0.9,1]          1.156226e-34  12.280262
## f.season=f.season.Sep-Dec              3.391033e-20   9.205810
## month=month.oct                        3.163903e-18   8.705424
## contact=contact.cellular               4.127348e-18   8.675225
## default=default.no                     4.787962e-14   7.537574
## job=job.retired                        2.103831e-13   7.342027
## month=month.sep                        5.701121e-11   6.551371
## f.age=f.age-(50,92]                    1.767166e-08   5.633375
## month=month.mar                        5.214842e-07   5.018235
## job=job.student                        6.469157e-07   4.976658
## f.job=f.job.Entrep-Retired-selfEmpl    4.357820e-06   4.593544
```

```
## f.duration=f.duration-(180,300]        3.266772e-05   4.154025
## poutcome=poutcome.failure             3.513520e-05   4.137342
## f.education=f.education.Professional   8.193961e-05   3.938655
## education=education.university.degree  8.193961e-05   3.938655
## f.job=f.job.Not-working                1.066886e-04   3.874854
## f.campaign=f.campaign-[0,1]            3.261253e-04   3.593612
## month=month.jun                        9.887768e-04  -3.293701
## f.education=f.education.Basic          6.774908e-04  -3.398530
## f.season=f.season.Mar-May              5.767248e-04  -3.442331
## marital=marital.married                3.331185e-04  -3.588083
## f.age=f.age-(40,50]                    4.090062e-05  -4.102332
## f.season=f.season.Jun-Aug              3.480205e-05  -4.139528
## education=education.basic.9y           1.044307e-05  -4.407790
## f.campaign=f.campaign-(2,20]           2.688717e-06  -4.693270
## month=month.jul                        4.872344e-07  -5.031273
## month=month.may                        5.199683e-08  -5.444343
## f.duration=f.duration-[5,120]          4.101428e-08  -5.486428
## f.job=f.job.Serv-Tech-BlueC            2.251008e-11  -6.688740
## job=job.blue-collar                    1.081400e-11  -6.795230
## default=default.unknown                4.787962e-14  -7.537574
## contact=contact.telephone              4.127348e-18  -8.675225
## y=y.no                                 2.494385e-64 -16.934745
## f.previous=f.previous-[0,0.9]         2.196177e-173 -28.071291
## poutcome=poutcome.nonexistent         2.196177e-173 -28.071291
## f.pdays=f.pdays-(22,23]               1.259643e-281 -35.860546
##
## $`2`
##                                          Cla/Mod      Mod/Cla     Global
## f.season=f.season.Jun-Aug             87.0824053  68.47635727 44.9899800
## f.previous=f.previous-[0,0.9]         66.5656330 100.00000000 85.9519038
## poutcome=poutcome.nonexistent         66.5656330 100.00000000 85.9519038
## contact=contact.telephone             86.6256295  54.22066550 35.8116232
## month=month.jul                       93.5260116  28.33625219 17.3346693
## f.pdays=f.pdays-(22,23]               59.2938733 100.00000000 96.4929860
## month=month.aug                       82.3298429  22.03152364 15.3106212
## y=y.no                                61.0161871  95.06129597 89.1382766
## month=month.jun                       83.9285714  18.10858144 12.3446894
## default=default.unknown               75.8095238  27.88091068 21.0420842
## f.campaign=f.campaign-(2,20]          67.0807453  37.82837128 32.2645291
## f.age=f.age-(40,50]                   65.5227454  28.75656743 25.1102204
## f.job=f.job.Serv-Tech-BlueC           62.1699670  52.78458844 48.5771543
## marital=marital.married               60.8144192  63.81786340 60.0400802
## housing=housing.no                    61.6498464  49.21190893 45.6713427
## job=job.technician                    63.7698898  18.24868651 16.3727455
## job=job.blue-collar                   62.2837370  25.21891419 23.1663327
## f.education=f.education.Basic         61.1145899  34.18563923 32.0040080
## job=job.housemaid                     72.2689076   3.01225919  2.3847695
## f.education=f.education.Professional  53.7084399  29.42206655 31.3426854
## education=education.university.degree 53.7084399  29.42206655 31.3426854
## f.job=f.job.Not-working               47.6608187   5.70928196  6.8537074
```

```
## month=month.may                          53.4523810  31.45359019 33.6673347
## f.job=f.job.Entrep-Retired-selfEmpl      48.5268631   9.80735552 11.5631263
## job=job.retired                          41.8918919   3.25744308  4.4488978
## month=month.dec                           0.0000000   0.00000000  0.3607214
## f.age=f.age-[18,30]                      48.6206897  14.81611208 17.4348697
## housing=housing.yes                      53.4857986  50.78809107 54.3286573
## marital=marital.single                   49.8936924  24.65849387 28.2765531
## job=job.student                          14.6788991   0.56042032  2.1843687
## f.campaign=f.campaign-[0,1]              49.6463932  36.88266200 42.5050100
## month=month.mar                           0.0000000   0.00000000  1.2625251
## month=month.oct                           2.5000000   0.07005254  1.6032064
## month=month.sep                           0.0000000   0.00000000  1.4028056
## default=default.no                       52.2588832  72.11908932 78.9579158
## f.previous=f.previous-(1,6]               0.0000000   0.00000000  2.7454910
## y=y.yes                                  26.0147601   4.93870403 10.8617234
## poutcome=poutcome.success                 0.0000000   0.00000000  3.1262525
## f.season=f.season.Mar-May                43.5077519  31.45359019 41.3627255
## f.pdays=f.pdays-[0,22]                    0.0000000   0.00000000  3.5070140
## month=month.apr                           0.0000000   0.00000000  6.4328657
## month=month.nov                           0.0000000   0.00000000 10.2805611
## poutcome=poutcome.failure                 0.0000000   0.00000000 10.9218437
## f.previous=f.previous-(0.9,1]             0.0000000   0.00000000 11.3026052
## contact=contact.cellular                 40.8054948  45.77933450 64.1883768
## f.season=f.season.Sep-Dec                 0.2936858   0.07005254 13.6472946
##                                                p.value     v.test
## f.season=f.season.Jun-Aug                 0.000000e+00        Inf
## f.previous=f.previous-[0,0.9]            3.019215e-293  36.598522
## poutcome=poutcome.nonexistent            3.019215e-293  36.598522
## contact=contact.telephone                1.051844e-235  32.782323
## month=month.jul                          1.306801e-149  26.051114
## f.pdays=f.pdays-(22,23]                   4.293494e-67  17.305259
## month=month.aug                           3.823967e-57  15.931608
## y=y.no                                    8.988442e-55  15.586531
## month=month.jun                           2.951209e-51  15.060372
## default=default.unknown                   8.402225e-45  14.043849
## f.campaign=f.campaign-(2,20]              1.196425e-22   9.793857
## f.age=f.age-(40,50]                       4.665985e-12   6.915384
## f.job=f.job.Serv-Tech-BlueC               5.923144e-12   6.881486
## marital=marital.married                   3.139148e-10   6.291746
## housing=housing.no                        6.199688e-09   5.811280
## job=job.technician                        3.130573e-05   4.163754
## job=job.blue-collar                       6.671792e-05   3.987697
## f.education=f.education.Basic             1.296174e-04   3.827184
## job=job.housemaid                         6.364492e-04   3.415586
## f.education=f.education.Professional      7.361711e-04  -3.375741
## education=education.university.degree    7.361711e-04  -3.375741
## f.job=f.job.Not-working                   2.382353e-04  -3.674586
## month=month.may                           1.335878e-04  -3.819748
## f.job=f.job.Entrep-Retired-selfEmpl       8.231372e-06  -4.459076
## job=job.retired                           2.872930e-06  -4.679701
```

```
## month=month.dec                         2.215519e-07  -5.180281
## f.age=f.age-[18,30]                     2.026128e-08  -5.609755
## housing=housing.yes                     6.199688e-09  -5.811280
## marital=marital.single                  6.116596e-11  -6.540860
## job=job.student                         2.122476e-20  -9.255993
## f.campaign=f.campaign-[0,1]             1.628849e-20  -9.284227
## month=month.mar                         3.477079e-24 -10.145300
## month=month.oct                         8.502567e-27 -10.716652
## month=month.sep                         8.041008e-27 -10.721815
## default=default.no                      8.402225e-45 -14.043849
## f.previous=f.previous-(1,6]             2.340302e-52 -15.227016
## y=y.yes                                 8.988442e-55 -15.586531
## poutcome=poutcome.success               1.057860e-59 -16.295757
## f.season=f.season.Mar-May               8.078887e-61 -16.452266
## f.pdays=f.pdays-[0,22]                  4.293494e-67 -17.305259
## month=month.apr                        1.591937e-125 -23.831058
## month=month.nov                        3.512409e-207 -30.715702
## poutcome=poutcome.failure              2.239752e-221 -31.760980
## f.previous=f.previous-(0.9,1]          6.871049e-230 -32.371692
## contact=contact.cellular               1.051844e-235 -32.782323
## f.season=f.season.Sep-Dec              8.033924e-278 -35.615604
## 
## $`3`
##                                          Cla/Mod    Mod/Cla    Global
## f.season=f.season.Sep-Dec              66.079295 100.000000 13.647295
## month=month.nov                        86.549708  98.666667 10.280561
## contact=contact.cellular               12.425851  88.444444 64.188377
## job=job.management                     17.948718  14.000000  7.034068
## f.pdays=f.pdays-(22,23]                 9.345794 100.000000 96.492986
## default=default.no                     10.076142  88.222222 78.957916
## y=y.no                                  9.622302  95.111111 89.138277
## poutcome=poutcome.failure              14.311927  17.333333 10.921844
## f.campaign=f.campaign-[0,1]            11.032532  52.000000 42.505010
## f.duration=f.duration-[5,120]          11.560694  40.000000 31.202405
## f.previous=f.previous-(0.9,1]          13.829787  17.333333 11.302605
## f.education=f.education.Professional   11.125320  38.666667 31.342685
## education=education.university.degree 11.125320  38.666667 31.342685
## job=job.entrepreneur                   16.402116   6.888889  3.787575
## education=education.basic.4y            5.048544   5.777778 10.320641
## job=job.retired                         2.702703   1.333333  4.448898
## f.age=f.age-[18,30]                     5.747126  11.111111 17.434870
## job=job.student                         0.000000   0.000000  2.184369
## job=job.blue-collar                     5.795848  14.888889 23.166333
## y=y.yes                                 4.059041   4.888889 10.861723
## f.previous=f.previous-(1,6]             0.000000   0.000000  2.745491
## f.education=f.education.Basic           6.199123  22.000000 32.004008
## f.campaign=f.campaign-(2,20]            6.211180  22.222222 32.264529
## poutcome=poutcome.success               0.000000   0.000000  3.126253
## default=default.unknown                 5.047619  11.777778 21.042084
## f.pdays=f.pdays-[0,22]                  0.000000   0.000000  3.507014
```

```
## month=month.apr                         0.000000   0.000000  6.432866
## month=month.jun                         0.000000   0.000000 12.344689
## contact=contact.telephone               2.909905  11.555556 35.811623
## month=month.aug                         0.000000   0.000000 15.310621
## month=month.jul                         0.000000   0.000000 17.334669
## month=month.may                         0.000000   0.000000 33.667335
## f.season=f.season.Mar-May               0.000000   0.000000 41.362725
## f.season=f.season.Jun-Aug               0.000000   0.000000 44.989980
##                                              p.value     v.test
## f.season=f.season.Sep-Dec               0.000000e+00        Inf
## month=month.nov                         0.000000e+00        Inf
## contact=contact.cellular                4.457861e-34  12.170601
## job=job.management                      4.606444e-08   5.465869
## f.pdays=f.pdays-(22,23]                 4.814305e-08   5.458036
## default=default.no                      8.895002e-08   5.347962
## y=y.no                                  3.321039e-06   4.649895
## poutcome=poutcome.failure               1.761402e-05   4.293170
## f.campaign=f.campaign-[0,1]             2.207930e-05   4.242754
## f.duration=f.duration-[5,120]           3.465489e-05   4.140500
## f.previous=f.previous-(0.9,1]           6.332305e-05   4.000073
## f.education=f.education.Professional    5.526671e-04   3.453839
## education=education.university.degree  5.526671e-04   3.453839
## job=job.entrepreneur                    9.628257e-04   3.301170
## education=education.basic.4y            4.140424e-04  -3.530968
## job=job.retired                         1.504256e-04  -3.790366
## f.age=f.age-[18,30]                     1.055513e-04  -3.877463
## job=job.student                         2.982402e-05  -4.174806
## job=job.blue-collar                     5.480188e-06  -4.545512
## y=y.yes                                 3.321039e-06  -4.649895
## f.previous=f.previous-(1,6]             1.972148e-06  -4.756258
## f.education=f.education.Basic           9.408844e-07  -4.903616
## f.campaign=f.campaign-(2,20]            9.107926e-07  -4.909994
## poutcome=poutcome.success               3.093148e-07  -5.117681
## default=default.unknown                 8.895002e-08  -5.347962
## f.pdays=f.pdays-[0,22]                  4.814305e-08  -5.458036
## month=month.apr                         2.294108e-14  -7.632967
## month=month.jun                         8.435071e-28 -10.928369
## contact=contact.telephone               4.457861e-34 -12.170601
## month=month.aug                         6.622682e-35 -12.325263
## month=month.jul                         6.599006e-40 -13.221438
## month=month.may                         8.848336e-86 -19.628385
## f.season=f.season.Mar-May              8.311269e-112 -22.469179
## f.season=f.season.Jun-Aug              2.151192e-125 -23.818443
##
## $`4`
##                                   Cla/Mod    Mod/Cla     Global
## f.season=f.season.Mar-May      53.779070 74.2474916 41.3627255
## month=month.apr                96.261682 20.6688963  6.4328657
## contact=contact.cellular       41.367468 88.6287625 64.1883768
## poutcome=poutcome.failure      78.348624 28.5618729 10.9218437
```

```
## f.previous=f.previous-(0.9,1] 70.744681 26.6889632 11.3026052
## month=month.may                44.702381 50.2341137 33.6673347
## y=y.yes                        48.892989 17.7257525 10.8617234
## default=default.no             32.969543 86.8896321 78.9579158
## f.pdays=f.pdays-(22,23]        30.924195 99.5986622 96.4929860
## job=job.student                69.724771  5.0836120  2.1843687
## month=month.mar                79.365079  3.3444816  1.2625251
## month=month.sep                74.285714  3.4782609  1.4028056
## marital=marital.single         37.774628 35.6521739 28.2765531
## f.age=f.age-[18,30]            40.574713 23.6120401 17.4348697
## f.campaign=f.campaign-[0,1]    34.370580 48.7625418 42.5050100
## month=month.oct                60.000000  3.2107023  1.6032064
## housing=housing.yes            32.312800 58.5953177 54.3286573
## month=month.dec                72.222222  0.8695652  0.3607214
## housing=housing.no             27.161036 41.4046823 45.6713427
## f.season=f.season.Sep-Dec      22.466960 10.2341137 13.6472946
## job=job.technician             22.888617 12.5083612 16.3727455
## marital=marital.married        27.202937 54.5150502 60.0400802
## f.campaign=f.campaign-(2,20]   24.658385 26.5551839 32.2645291
## f.age=f.age-(40,50]            22.984836 19.2642140 25.1102204
## poutcome=poutcome.success       3.846154  0.4013378  3.1262525
## f.pdays=f.pdays-[0,22]          3.428571  0.4013378  3.5070140
## default=default.unknown        18.666667 13.1103679 21.0420842
## month=month.jun                14.448052  5.9531773 12.3446894
## y=y.no                         27.652878 82.2742475 89.1382766
## month=month.aug                12.696335  6.4882943 15.3106212
## month=month.nov                 7.797271  2.6755853 10.2805611
## f.previous=f.previous-[0,0.9]  24.761017 71.0367893 85.9519038
## poutcome=poutcome.nonexistent 24.761017 71.0367893 85.9519038
## month=month.jul                 5.317919  3.0769231 17.3346693
## contact=contact.telephone       9.513151 11.3712375 35.8116232
## f.season=f.season.Jun-Aug      10.334076 15.5183946 44.9899800
##                                      p.value     v.test
## f.season=f.season.Mar-May      1.214684e-211  31.047983
## month=month.apr                2.443356e-153  26.378112
## contact=contact.cellular       7.335999e-138  24.992718
## poutcome=poutcome.failure      5.641457e-137  24.911094
## f.previous=f.previous-(0.9,1] 3.761358e-101  21.351684
## month=month.may                 2.068828e-57  15.969971
## y=y.yes                         9.791179e-23   9.814096
## default=default.no              1.504249e-20   9.292699
## f.pdays=f.pdays-(22,23]         7.889588e-20   9.114690
## job=job.student                 6.601535e-18   8.621617
## month=month.mar                 5.354652e-16   8.103167
## month=month.sep                 1.931681e-14   7.655096
## marital=marital.single          7.928406e-14   7.471503
## f.age=f.age-[18,30]             1.851132e-13   7.359130
## f.campaign=f.campaign-[0,1]     5.524740e-09   5.830543
## month=month.oct                 2.441289e-08   5.577408
## housing=housing.yes             7.419015e-05   3.962432
```

```
## month=month.dec              2.959629e-04   3.618808
## housing=housing.no           7.419015e-05  -3.962432
## f.season=f.season.Sep-Dec    2.698987e-06  -4.692491
## job=job.technician           8.712805e-07  -4.918684
## marital=marital.married      2.090160e-07  -5.191135
## f.campaign=f.campaign-(2,20]  1.226294e-08  -5.696030
## f.age=f.age-(40,50]          2.434489e-10  -6.331081
## poutcome=poutcome.success    4.435265e-17  -8.400779
## f.pdays=f.pdays-[0,22]       7.889588e-20  -9.114690
## default=default.unknown      1.504249e-20  -9.292699
## month=month.jun              1.863347e-21  -9.512388
## y=y.no                       9.791179e-23  -9.814096
## month=month.aug              1.221397e-33 -12.088058
## month=month.nov              8.917776e-38 -12.847202
## f.previous=f.previous-[0,0.9]  4.885281e-80 -18.944683
## poutcome=poutcome.nonexistent  4.885281e-80 -18.944683
## month=month.jul              1.393898e-85 -19.605278
## contact=contact.telephone   7.335999e-138 -24.992718
## f.season=f.season.Jun-Aug   1.401958e-178 -28.493704
##
##
## Link between the cluster variable and the quantitative variables
## ================================================================
##                      Eta2       P-value
## pdays          0.833431759  0.000000e+00
## previous       0.501141537  0.000000e+00
## emp.var.rate   0.949447313  0.000000e+00
## cons.price.idx 0.574427755  0.000000e+00
## euribor3m      0.989874994  0.000000e+00
## nr.employed    0.866883329  0.000000e+00
## cons.conf.idx  0.167721928 3.916150e-198
## campaign       0.027687626  3.757869e-30
## age            0.008632187  2.197904e-09
##
## Description of each cluster by quantitative variables
## =====================================================
## $`1`
##                   v.test Mean in category  Overall mean sd in category
## previous        45.971997        1.8210526    0.17855711      0.9567373
## cons.conf.idx    6.937479      -38.2494737  -40.54192385      7.1162952
## age              5.645279       44.3789474   40.17755511     16.9858634
## campaign        -4.322398        1.7894737    2.51503006      1.2598641
## cons.price.idx  -4.817260       93.3651368   93.56373427      0.8341906
## emp.var.rate   -19.115707       -2.0936842    0.05212425      0.8565304
## euribor3m      -21.054798        0.9672263    3.58457355      0.5121394
## nr.employed    -27.122883     5025.3821053 5165.87569138     51.3340730
## pdays          -64.481704        7.8473684   22.41362725      6.4451425
##                Overall sd       p.value
## previous        0.5020810  0.000000e+00
## cons.conf.idx   4.6436681  3.991593e-12
```

```
## age            10.4585324  1.649135e-08
## campaign        2.3588988  1.543422e-05
## cons.price.idx  0.5793439  1.455430e-06
## emp.var.rate    1.5774788  1.868722e-81
## euribor3m       1.7469207  2.066633e-98
## nr.employed    72.7919889 5.290316e-162
## pdays           3.1744936  0.000000e+00
##
## $`2`
##                   v.test Mean in category  Overall mean sd in category
## emp.var.rate    64.85040         1.304588    0.05212425     0.14415633
## euribor3m       62.64096         4.924314    3.58457355     0.05025276
## nr.employed     56.70185      5216.408091 5165.87569138    17.23254124
## cons.price.idx  52.54683        93.936445   93.56373427     0.32429720
## cons.conf.idx   25.36681       -39.099755  -40.54192385     3.04458233
## pdays           15.08725        23.000000   22.41362725     0.00000000
## campaign        11.60168         2.850088    2.51503006     2.73344964
## previous       -29.04787         0.000000    0.17855711     0.00000000
##                Overall sd       p.value
## emp.var.rate    1.5774788  0.000000e+00
## euribor3m       1.7469207  0.000000e+00
## nr.employed    72.7919889  0.000000e+00
## cons.price.idx  0.5793439  0.000000e+00
## cons.conf.idx   4.6436681 5.863605e-142
## pdays           3.1744936  1.964506e-51
## campaign        2.3588988  4.040627e-31
## previous        0.5020810 1.637199e-185
##
## $`3`
##                    v.test Mean in category Overall mean sd in category
## nr.employed      9.115196      5195.713333  5165.875691     1.29710789
## euribor3m        6.818350         4.120207     3.584574     0.11552217
## pdays            4.107565        23.000000    22.413627     0.00000000
## campaign        -4.708571         2.015556     2.515030     1.58527328
## cons.conf.idx   -6.997289       -42.003111   -40.541924     0.30057591
## cons.price.idx -13.833237        93.203342    93.563734     0.06359003
##                Overall sd      p.value
## nr.employed    72.7919889 7.852866e-20
## euribor3m       1.7469207 9.209193e-12
## pdays           3.1744936 3.998529e-05
## campaign        2.3588988 2.494597e-06
## cons.conf.idx   4.6436681 2.609633e-12
## cons.price.idx  0.5793439 1.606263e-43
##
## $`4`
##                    v.test Mean in category  Overall mean sd in category
## previous        12.312403        0.3123746    0.17855711      0.5101781
## pdays            8.075532       22.9685619   22.41362725      0.5687059
## age             -4.052007       39.2602007   40.17755511     12.0436159
## campaign        -7.780683        2.1177258    2.51503006      1.6827435
```

```
## cons.conf.idx  -25.921260      -43.1475585  -40.54192385      5.9781301
## cons.price.idx -46.092491       92.9856876   93.56373427      0.4315899
## nr.employed    -55.611005     5078.2481605 5165.87569138     38.1020523
## emp.var.rate   -60.712801       -2.0210702    0.05212425      0.5356747
## euribor3m      -63.124684        1.1974843    3.58457355      0.2469009
##                Overall sd       p.value
## previous        0.5020810  7.767580e-35
## pdays           3.1744936  6.718328e-16
## age            10.4585324  5.078016e-05
## campaign        2.3588988  7.213395e-15
## cons.conf.idx   4.6436681 3.835830e-148
## cons.price.idx  0.5793439  0.000000e+00
## nr.employed    72.7919889  0.000000e+00
## emp.var.rate    1.5774788  0.000000e+00
## euribor3m       1.7469207  0.000000e+00
```

Hierarchical Clustering

Al hacer el HCPC podemos ver que el gráfico de ganancia de inercia nos da la mayoría en dos variables, y luego dos picos más pequeños.

Ahora vemos el clustering no supervisado. Vamos a clasificar estos clusters.

Para el cluster 1:
- Está caracterizado por personas que han sido contactados previamente
- Han aceptado el producto
- Se han contactado en f.season.Sep-Dec
- Tienen una sobrerepresentación de f.job.Entrep-Retired-selfEmpl
- Llamadas de duración mayor a 3min

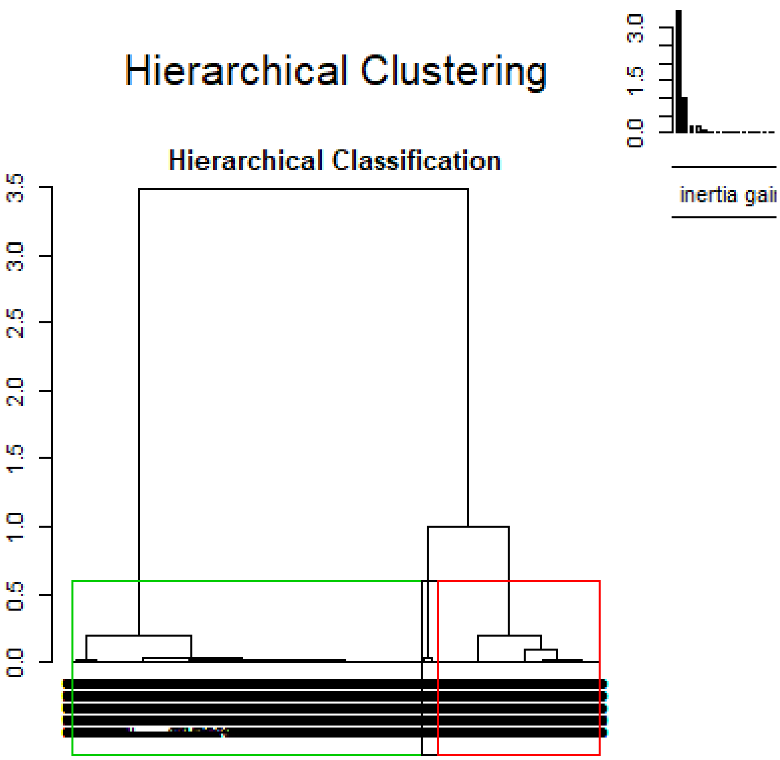
Para el cluster 2:
- Han sido contactados f.season.Mar-May
- Han sido contactados en campañas previas
- Han aceptado el producto (y.yes)

Para el cluster 3:
- Han sido contactados previamente
- No han sido contactados en campañas previas
- Han sido contactados en la temporada de f.season.Jun-Aug
- Han rechazado el producto
- Tienen una leve representación de f.job.Serv-Tech-BlueC

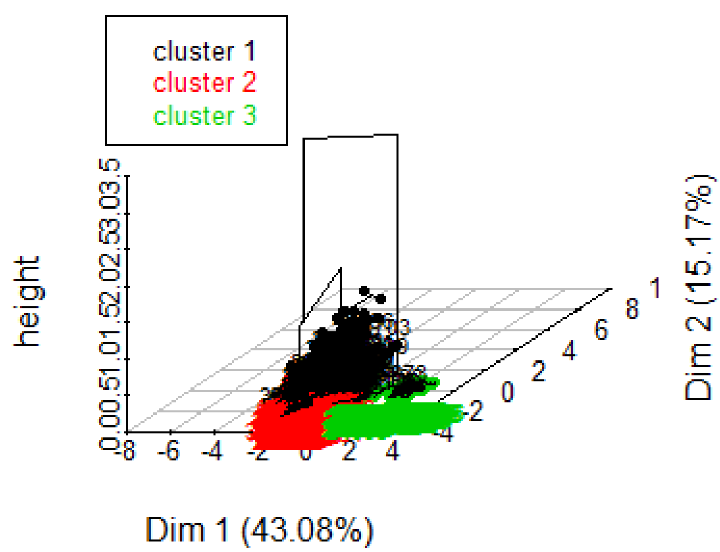
Además vemos que los parengones tienen mucho parecido a lo descrito para cada cluster.

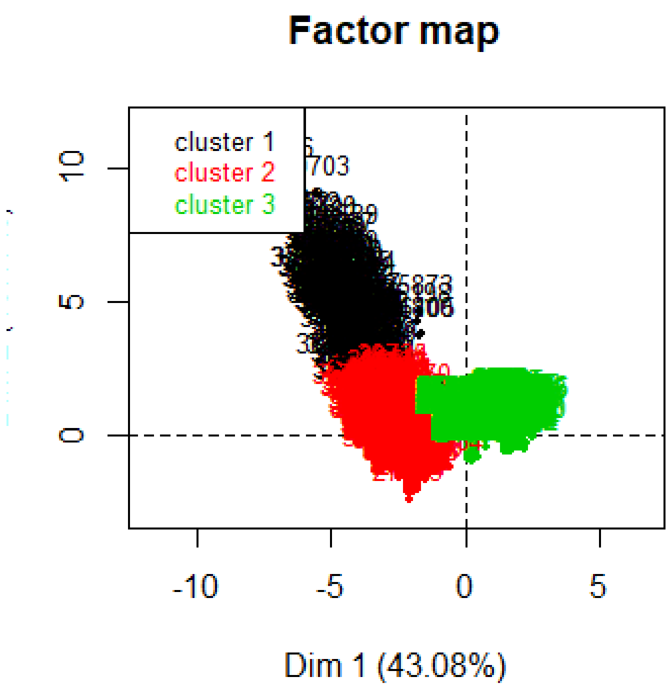
Utilizamos el atributo nb.clust=3, después de haber visto que era el mejor corte.

```
res.pca<-PCA(df[,c('duration',vars_num,vars_factorizadas)],quanti.sup=1, quali.s
up = c(11:19), ncp=2, graph=FALSE)
res.hcpc<-HCPC(res.pca,order=TRUE, nb.clust = 3)
```



Hierarchical clustering on the factor map






```r
attributes(res.hcpc)

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

summary(res.hcpc$data.clust)

##     duration           age           campaign          pdays
##  Min.   :   5.0   Min.   :18.00   Min.   : 1.000   Min.   : 0.00
##  1st Qu.: 103.0   1st Qu.:32.00   1st Qu.: 1.000   1st Qu.:23.00
##  Median : 178.5   Median :38.00   Median : 2.000   Median :23.00
##  Mean   : 259.9   Mean   :40.18   Mean   : 2.515   Mean   :22.41
##  3rd Qu.: 321.0   3rd Qu.:47.00   3rd Qu.: 3.000   3rd Qu.:23.00
##  Max.   :2078.0   Max.   :92.00   Max.   :20.000   Max.   :23.00
##     previous       emp.var.rate       cons.price.idx  cons.conf.idx
##  Min.   :0.0000   Min.   :-3.40000   Min.   :92.20   Min.   :-50.80
##  1st Qu.:0.0000   1st Qu.:-1.80000   1st Qu.:93.08   1st Qu.:-42.70
##  Median :0.0000   Median : 1.10000   Median :93.44   Median :-41.80
##  Mean   :0.1786   Mean   : 0.05212   Mean   :93.56   Mean   :-40.54
##  3rd Qu.:0.0000   3rd Qu.: 1.40000   3rd Qu.:93.99   3rd Qu.:-36.40
##  Max.   :6.0000   Max.   : 1.40000   Max.   :94.77   Max.   :-26.90
##    euribor3m      nr.employed                                f.job
##  Min.   :0.634   Min.   :4964   f.job.Admin-Managment        :1647
##  1st Qu.:1.334   1st Qu.:5099   f.job.Entrep-Retired-selfEmpl: 577
```

```
##  Median :4.857   Median :5191   f.job.Not-working            : 342
##  Mean   :3.585   Mean   :5166   f.job.Serv-Tech-BlueC        :2424
##  3rd Qu.:4.961   3rd Qu.:5228
##  Max.   :5.045   Max.   :5228
##             f.season                         f.education
##  f.season.Mar-May:2064   f.education.Basic       :1597
##  f.season.Jun-Aug:2245   f.education.High School :1829
##  f.season.Sep-Dec: 681   f.education.Professional:1564
##
##
##
##           f.age                         f.duration
##  f.age-[18,30]: 870   f.duration-[5,120]      :1557
##  f.age-(30,40]:1991   f.duration-(120,180]    : 966
##  f.age-(40,50]:1253   f.duration-(180,300]    :1090
##  f.age-(50,92]: 876   f.duration-(300,2.1e+03]:1377
##
##
##             f.campaign               f.pdays                  f.previous
##  f.campaign-[0,1] :2121   f.pdays-[0,22] : 175   f.previous-[0,0.9]:4289
##  f.campaign-(1,2] :1259   f.pdays-(22,23]:4815   f.previous-(0.9,1]: 564
##  f.campaign-(2,20]:1610                          f.previous-(1,6]  : 137
##
##
##
##      y        clust
##  y.no :4448   1: 190
##  y.yes: 542   2:1497
##               3:3303
##
##
##

attributes(res.hcpc$desc.var)

## $names
## [1] "test.chi2"  "category"   "quanti.var" "quanti"     "call"
##
## $class
## [1] "catdes" "list "

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

##                   p.value df
## f.pdays      0.000000e+00  2
## f.previous   0.000000e+00  4
## f.season    3.418306e-239  4
## y           2.373371e-145  2
## f.age        4.054913e-27  6
## f.job        5.294443e-17  6
```

```
## f.campaign   6.236440e-15  4
## f.duration   2.126018e-09  6
## f.education  7.523481e-04  4
```

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

```
## $`1`
##                                       Cla/Mod  Mod/Cla    Global
## f.pdays=f.pdays-[0,22]             96.5714286 88.94737  3.507014
## f.previous=f.previous-(1,6]        75.1824818 54.21053  2.745491
## y=y.yes                            21.0332103 60.00000 10.861723
## f.previous=f.previous-(0.9,1]      15.4255319 45.78947 11.302605
## f.season=f.season.Sep-Dec          11.1600587 40.00000 13.647295
## f.age=f.age-(50,92]                 7.4200913 34.21053 17.555110
## f.job=f.job.Entrep-Retired-selfEmpl   7.6256499 23.15789 11.563126
## f.duration=f.duration-(180,300]     6.0550459 34.73684 21.843687
## f.education=f.education.Professional  5.4347826 44.73684 31.342685
## f.job=f.job.Not-working             8.1871345 14.73684  6.853707
## f.campaign=f.campaign-[0,1]         4.9504950 55.26316 42.505010
## f.duration=f.duration-(300,2.1e+03]   4.9382716 35.78947 27.595190
## f.age=f.age-[18,30]                 5.0574713 23.15789 17.434870
## f.age=f.age-(30,40]                 2.8126570 29.47368 39.899800
## f.education=f.education.Basic       2.5046963 21.05263 32.004008
## f.season=f.season.Mar-May           2.7131783 29.47368 41.362725
## f.age=f.age-(40,50]                 1.9952115 13.15789 25.110220
## f.season=f.season.Jun-Aug           2.5835189 30.52632 44.989980
## f.campaign=f.campaign-(2,20]        2.0496894 17.36842 32.264529
## f.duration=f.duration-[5,120]       1.7341040 14.21053 31.202405
## f.job=f.job.Serv-Tech-BlueC         1.9801980 25.26316 48.577154
## y=y.no                              1.7086331 40.00000 89.138277
## f.previous=f.previous-[0,0.9]       0.0000000  0.00000 85.951904
## f.pdays=f.pdays-(22,23]             0.4361371 11.05263 96.492986
##                                          p.value     v.test
## f.pdays=f.pdays-[0,22]             1.259643e-281  35.860546
## f.previous=f.previous-(1,6]        1.553305e-129  24.214815
## y=y.yes                             2.494385e-64  16.934745
## f.previous=f.previous-(0.9,1]       1.156226e-34  12.280262
## f.season=f.season.Sep-Dec           3.391033e-20   9.205810
## f.age=f.age-(50,92]                 1.767166e-08   5.633375
## f.job=f.job.Entrep-Retired-selfEmpl   4.357820e-06   4.593544
## f.duration=f.duration-(180,300]     3.266772e-05   4.154025
## f.education=f.education.Professional  8.193961e-05   3.938655
## f.job=f.job.Not-working             1.066886e-04   3.874854
## f.campaign=f.campaign-[0,1]         3.261253e-04   3.593612
## f.duration=f.duration-(300,2.1e+03]   1.193267e-02   2.514129
## f.age=f.age-[18,30]                 4.007463e-02   2.052979
## f.age=f.age-(30,40]                 2.422784e-03  -3.032822
## f.education=f.education.Basic       6.774908e-04  -3.398530
## f.season=f.season.Mar-May           5.767248e-04  -3.442331
```

```
## f.age=f.age-(40,50]                4.090062e-05  -4.102332
## f.season=f.season.Jun-Aug          3.480205e-05  -4.139528
## f.campaign=f.campaign-(2,20]       2.688717e-06  -4.693270
## f.duration=f.duration-[5,120]      4.101428e-08  -5.486428
## f.job=f.job.Serv-Tech-BlueC        2.251008e-11  -6.688740
## y=y.no                             2.494385e-64 -16.934745
## f.previous=f.previous-[0,0.9]     2.196177e-173 -28.071291
## f.pdays=f.pdays-(22,23]           1.259643e-281 -35.860546
##
## $`2`
##                                Cla/Mod    Mod/Cla    Global       p.value
## f.season=f.season.Mar-May    53.779070 74.1482966 41.362725 9.639926e-211
## f.previous=f.previous-(0.9,1] 70.921986 26.7201069 11.302605 8.620072e-102
## y=y.yes                      48.892989 17.7020708 10.861723  1.228275e-22
## f.pdays=f.pdays-(22,23]      30.965732 99.5991984 96.492986  7.200421e-20
## f.age=f.age-[18,30]          40.574713 23.5804943 17.434870  2.310583e-13
## f.campaign=f.campaign-[0,1]  34.464875 48.8309953 42.505010  3.637684e-09
## f.job=f.job.Not-working      37.134503  8.4836339  6.853707  3.388459e-03
## f.duration=f.duration-[5,120] 28.066795 29.1917168 31.202405  4.431130e-02
## f.job=f.job.Serv-Tech-BlueC  28.135314 45.5577822 48.577154  5.209983e-03
## f.season=f.season.Sep-Dec    22.760646 10.3540414 13.647295  5.972406e-06
## f.campaign=f.campaign-(2,20]  24.658385 26.5197061 32.264529  9.621448e-09
## f.age=f.age-(40,50]          22.984836 19.2384770 25.110220  1.946669e-10
## f.pdays=f.pdays-[0,22]        3.428571  0.4008016  3.507014  7.200421e-20
## y=y.no                       27.697842 82.2979292 89.138277  1.228275e-22
## f.previous=f.previous-[0,0.9] 24.784332 71.0086840 85.951904  1.676131e-80
## f.season=f.season.Jun-Aug    10.334076 15.4976620 44.989980 3.526427e-179
##                                   v.test
## f.season=f.season.Mar-May      30.981263
## f.previous=f.previous-(0.9,1]  21.420424
## y=y.yes                         9.791202
## f.pdays=f.pdays-(22,23]         9.124596
## f.age=f.age-[18,30]             7.329474
## f.campaign=f.campaign-[0,1]     5.899879
## f.job=f.job.Not-working         2.930106
## f.duration=f.duration-[5,120]  -2.011134
## f.job=f.job.Serv-Tech-BlueC    -2.793756
## f.season=f.season.Sep-Dec      -4.527364
## f.campaign=f.campaign-(2,20]   -5.737271
## f.age=f.age-(40,50]            -6.365490
## f.pdays=f.pdays-[0,22]         -9.124596
## y=y.no                         -9.791202
## f.previous=f.previous-[0,0.9] -19.000910
## f.season=f.season.Jun-Aug     -28.542041
##
## $`3`
##                                     Cla/Mod    Mod/Cla    Global
## f.previous=f.previous-[0,0.9]      75.21567  97.668786 85.951904
## f.season=f.season.Jun-Aug          87.08241  59.188616 44.989980
## f.pdays=f.pdays-(22,23]            68.59813 100.000000 96.492986
```

```
## y=y.no                              70.59353  95.065092 89.138277
## f.age=f.age-(40,50]                 75.01995  28.458977 25.110220
## f.campaign=f.campaign-(2,20]        73.29193  35.725098 32.264529
## f.job=f.job.Serv-Tech-BlueC         69.88449  51.286709 48.577154
## f.duration=f.duration-[5,120]       70.19910  33.091129 31.202405
## f.duration=f.duration-(180,300]     62.66055  20.678171 21.843687
## f.duration=f.duration-(300,2.1e+03] 63.03558  26.279140 27.595190
## f.job=f.job.Entrep-Retired-selfEmpl 58.92548  10.293672 11.563126
## f.job=f.job.Not-working             54.67836   5.661520  6.853707
## f.campaign=f.campaign-[0,1]         60.58463  38.904027 42.505010
## f.age=f.age-[18,30]                 54.36782  14.320315 17.434870
## f.previous=f.previous-(1,6]          0.00000   0.000000  2.745491
## y=y.yes                             30.07380   4.934908 10.861723
## f.pdays=f.pdays-[0,22]               0.00000   0.000000  3.507014
## f.previous=f.previous-(0.9,1]       13.65248   2.331214 11.302605
## f.season=f.season.Mar-May           43.50775  27.187405 41.362725
##                                           p.value     v.test
## f.previous=f.previous-[0,0.9]       6.011990e-240  33.078726
## f.season=f.season.Jun-Aug           3.447463e-187  29.180312
## f.pdays=f.pdays-(22,23]              7.182640e-86  19.638980
## y=y.no                               2.742424e-74  18.234602
## f.age=f.age-(40,50]                  8.378911e-15   7.761714
## f.campaign=f.campaign-(2,20]         1.432305e-13   7.393298
## f.job=f.job.Serv-Tech-BlueC          8.211297e-08   5.362423
## f.duration=f.duration-[5,120]        5.127384e-05   4.049743
## f.duration=f.duration-(180,300]      5.558862e-03  -2.772728
## f.duration=f.duration-(300,2.1e+03]  3.761527e-03  -2.897497
## f.job=f.job.Entrep-Retired-selfEmpl  1.078588e-04  -3.872197
## f.job=f.job.Not-working              4.940469e-06  -4.567300
## f.campaign=f.campaign-[0,1]          6.903561e-13  -7.181332
## f.age=f.age-[18,30]                  1.568440e-15  -7.971434
## f.previous=f.previous-(1,6]          6.723232e-67 -17.279410
## y=y.yes                              2.742424e-74 -18.234602
## f.pdays=f.pdays-[0,22]               7.182640e-86 -19.638980
## f.previous=f.previous-(0.9,1]       1.915585e-167 -27.580337
## f.season=f.season.Mar-May           1.553439e-179 -28.570715
```

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

```
##                       Eta2       P-value
## pdays          0.833431720  0.000000e+00
## previous       0.492162105  0.000000e+00
## emp.var.rate   0.887391268  0.000000e+00
## cons.price.idx 0.451020756  0.000000e+00
## euribor3m      0.972756054  0.000000e+00
## nr.employed    0.859965803  0.000000e+00
## cons.conf.idx  0.137891619 2.106281e-161
## campaign       0.018064870  1.813245e-20
## age            0.008646255  3.945874e-10
```

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

## $`1`
##                     v.test Mean in category  Overall mean sd in category
## previous         45.971997        1.8210526    0.17855711       0.9567373
## cons.conf.idx     6.937479      -38.2494737  -40.54192385       7.1162952
## age               5.645279       44.3789474   40.17755511      16.9858634
## duration          2.110371      297.6368421  259.85110220     204.0561577
## campaign         -4.322398        1.7894737    2.51503006       1.2598641
## cons.price.idx   -4.817260       93.3651368   93.56373427       0.8341906
## emp.var.rate    -19.115707       -2.0936842    0.05212425       0.8565304
## euribor3m       -21.054798        0.9672263    3.58457355       0.5121394
## nr.employed     -27.122883     5025.3821053 5165.87569138      51.3340730
## pdays           -64.481704        7.8473684   22.41362725       6.4451425
##                  Overall sd       p.value
## previous          0.5020810  0.000000e+00
## cons.conf.idx     4.6436681  3.991593e-12
## age              10.4585324  1.649135e-08
## duration        251.6124483  3.482644e-02
## campaign          2.3588988  1.543422e-05
## cons.price.idx    0.5793439  1.455430e-06
## emp.var.rate      1.5774788  1.868722e-81
## euribor3m         1.7469207  2.066633e-98
## nr.employed      72.7919889 5.290316e-162
## pdays             3.1744936  0.000000e+00
##
## $`2`
##                     v.test Mean in category  Overall mean sd in category
## previous         12.347248        0.3126253    0.17855711       0.5102106
## pdays             8.083856       22.9686039   22.41362725       0.5683270
## age              -4.063321       39.2585170   40.17755511      12.0357116
## campaign         -7.817387        2.1162325    2.51503006       1.6821146
## cons.conf.idx   -25.982635      -43.1512358  -40.54192385       5.9749813
## cons.price.idx  -46.161007       92.9853808   93.56373427       0.4313831
## nr.employed     -55.580906     5078.3791583 5165.87569138      38.2446628
## emp.var.rate    -60.699472       -2.0186373    0.05212425       0.5394336
## euribor3m       -63.098140        1.2007649    3.58457355       0.2625532
##                 Overall sd       p.value
## previous         0.5020810  5.040624e-35
## pdays            3.1744936  6.275018e-16
## age             10.4585324  4.837945e-05
## campaign         2.3588988  5.393129e-15
## cons.conf.idx    4.6436681 7.782048e-149
## cons.price.idx   0.5793439  0.000000e+00
## nr.employed     72.7919889  0.000000e+00
## emp.var.rate     1.5774788  0.000000e+00
## euribor3m        1.7469207  0.000000e+00
##
## $`3`
##                     v.test Mean in category  Overall mean sd in category
```

```
## euribor3m       69.642425       4.81553255    3.58457355       0.2815584
## emp.var.rate    66.534303       1.11407811    0.05212425       0.4992563
## nr.employed     64.815247    5213.61292764 5165.87569138      17.5063561
## cons.price.idx  46.665983      93.83728217   93.56373427       0.3924268
## cons.conf.idx   22.363260     -39.49118983  -40.54192385       2.9986228
## pdays           18.255882      23.00000000   22.41362725       0.0000000
## campaign         9.321541       2.73751135    2.51503006       2.6231360
## previous       -30.559574       0.02331214    0.17855711       0.1508930
##                Overall sd       p.value
## euribor3m       1.7469207  0.000000e+00
## emp.var.rate    1.5774788  0.000000e+00
## nr.employed    72.7919889  0.000000e+00
## cons.price.idx  0.5793439  0.000000e+00
## cons.conf.idx   4.6436681 8.971345e-111
## pdays           3.1744936  1.857861e-74
## campaign        2.3588988  1.146631e-20
## previous        0.5020810 4.219155e-205

### desc.ind ###
### C. 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
##      36296      36721      40892      41007      36907
## 0.03946875 0.05735028 0.13237750 0.21737224 0.22738201
## ---------------------------------------------------------
## Cluster: 2
##      30951      36346      36347      36427      36864
## 0.06134805 0.09336824 0.09336824 0.09552183 0.09624810
## ---------------------------------------------------------
## Cluster: 3
##       1467       1752        331       6185      18926
## 0.01424926 0.01424926 0.01478019 0.01478019 0.01491243

res.hcpc$desc.ind$dist

## Cluster: 1
##     40396     40703     39592     38902     39612
## 11.543929 10.404680  9.652239  9.431592  9.270355
## ---------------------------------------------------------
## Cluster: 2
##    37956    38026    38148    37904    38051
## 4.742178 4.688806 4.646346 4.645320 4.644823
## ---------------------------------------------------------
## Cluster: 3
##    11696     8484    11485    11056    10761
## 5.265453 5.246245 5.197173 5.138562 5.009706
```

CA analysis for your data should contain your factor version of the numeric target (duration) in K= 7 (maximum 10) levels and 2 factors:

Eigenvalues and dominant axes analysis. How many axes we have to consider are there any row categories that can be combined/avoided to explain Duration target.

Para experimentar y para que tenga más sentido el análisis de correspondencias, refactorizaremos a 8 niveles la variable duration.

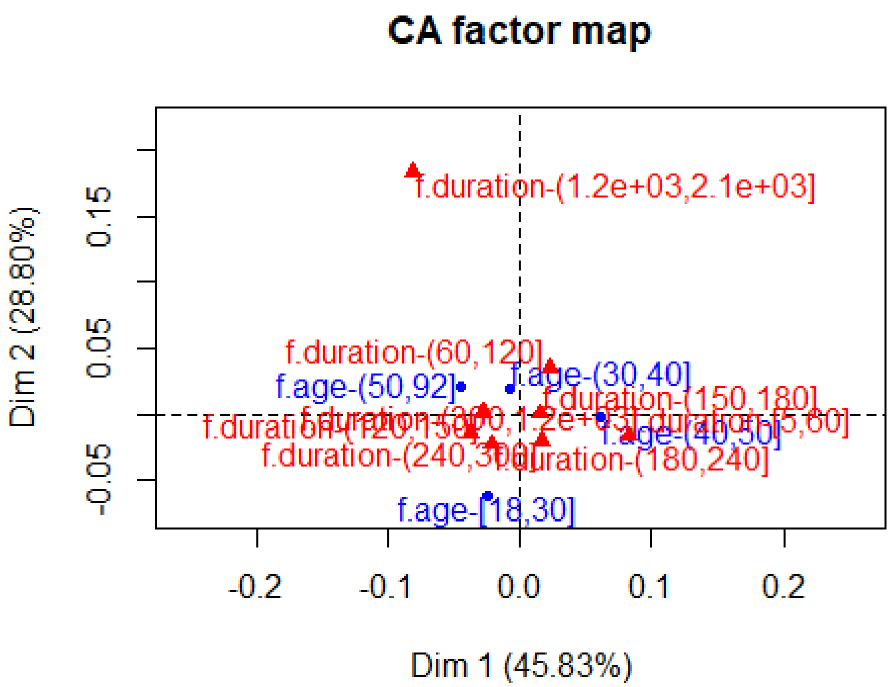
Ahora hacemos el análisis de correspondencias entre nuestra nueva duration factorizada y f.age. Para saber cuantas dimensiones debemos considerar, obtenemos la media de los eigenvalues. Vemos que solamente tiene sentido considerar el primer eje, ya que este es el único valor mayor a la media (kaiser).

Al graficar el CA, podemos ver que los 2 niveles con menores edades, son los que menos representados en ese eje. Para duration, los niveles mejor representados en el eje son los de mayor y menor duración.

Al ejecutar la función del chisq.test podemos ver que el pvalue es muy grande, lo que nos puede decir que la probabilidad de que no tengan relación es muy grande.

CA - duration vs f.age

```
# Para duration
aux2<-c(5,60,120,150,180,240,300,1200,2100) # Niveles "naturales"
duration_k8<-factor(cut(df$duration,breaks=aux2,include.lowest=T))
table(duration_k8)

## duration_k8
##           [5,60]          (60,120]         (120,150]         (150,180]
##              490              1067               496               470
##        (180,240]         (240,300]     (300,1.2e+03] (1.2e+03,2.1e+03]
##              606               484              1311                66

levels(duration_k8)<-paste0("f.duration-",levels(duration_k8)) # Hacemos las eti
quetas más informativas
summary(duration_k8)

##           f.duration-[5,60]          f.duration-(60,120]
##                         490                         1067
##        f.duration-(120,150]         f.duration-(150,180]
##                         496                          470
##        f.duration-(180,240]         f.duration-(240,300]
##                         606                          484
##     f.duration-(300,1.2e+03] f.duration-(1.2e+03,2.1e+03]
##                        1311                           66

res.ca<-CA(table(df$f.age,duration_k8))
```

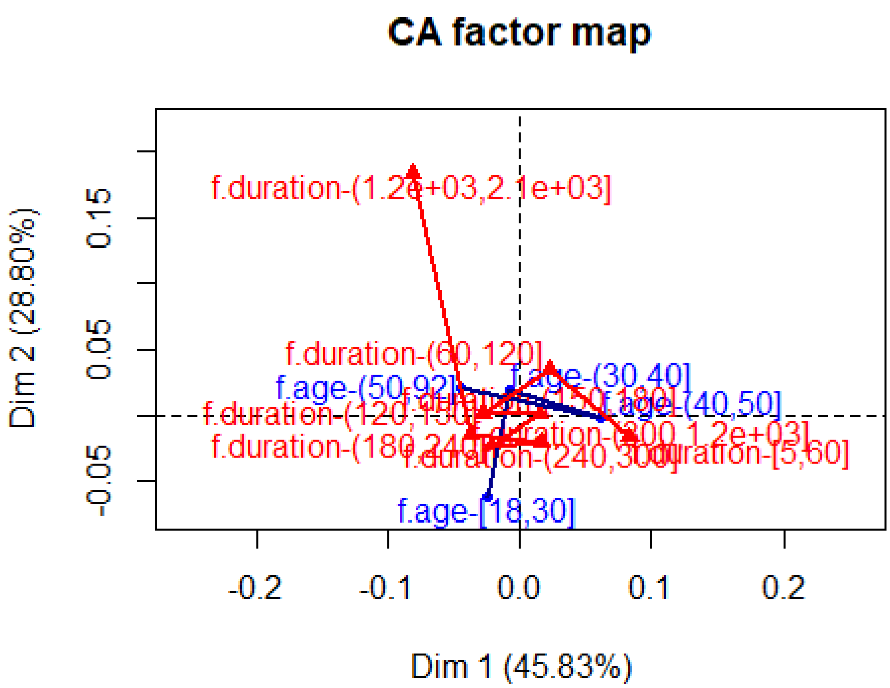



```r
attributes(res.ca)

## $names
## [1] "eig"  "call" "row"  "col"  "svd"
##
## $class
## [1] "CA"   "list"

res.ca$eig

##         eigenvalue percentage of variance
## dim 1 0.0014130984               45.82949
## dim 2 0.0008880983               28.80273
## dim 3 0.0007821856               25.36778
##       cumulative percentage of variance
## dim 1                          45.82949
## dim 2                          74.63222
## dim 3                         100.00000

mean(res.ca$eig[,1])  # Mean of eigenvalues

## [1] 0.001027794

sum(res.ca$eig[,1])  # Total inertia

## [1] 0.003083382
```

```
# Rows
res.ca$row

## $coord
##                      Dim 1        Dim 2       Dim 3
## f.age-[18,30] -0.024672376 -0.061587071  0.00515479
## f.age-(30,40] -0.007883325  0.019128850  0.02866207
## f.age-(40,50]  0.060905585 -0.002487498 -0.01655528
## f.age-(50,92] -0.044696383  0.021246627 -0.04658344
##
## $contrib
##                   Dim 1      Dim 2      Dim 3
## f.age-[18,30]  7.510490 74.4623554  0.5922851
## f.age-(30,40]  1.754758 16.4394554 41.9059874
## f.age-(40,50] 65.916230  0.1749504  8.7985988
## f.age-(50,92] 24.818522  8.9232388 48.7031287
##
## $cos2
##                    Dim 1       Dim 2       Dim 3
## f.age-[18,30] 0.13746379 0.856535696 0.006000512
## f.age-(30,40] 0.04973439 0.292830101 0.657435510
## f.age-(40,50] 0.92975389 0.001550884 0.068695230
## f.age-(50,92] 0.43249165 0.097726640 0.469781706
##
## $inertia
## [1] 0.0007720623 0.0004985776 0.0010018363 0.0008109061

# Columns: the same
res.ca$col

## $coord
##                                   Dim 1         Dim 2         Dim 3
## f.duration-[5,60]            0.08319396 -0.0161245945  2.458294e-02
## f.duration-(60,120]          0.02249525  0.0348536919 -2.577791e-02
## f.duration-(120,150]        -0.02735408  0.0013898644  1.373241e-02
## f.duration-(150,180]         0.01565127  0.0006091358  6.835979e-05
## f.duration-(180,240]        -0.02116464 -0.0227191287  4.705510e-02
## f.duration-(240,300]         0.01669583 -0.0202658085  3.700409e-04
## f.duration-(300,1.2e+03]    -0.03694459 -0.0143453208 -2.112106e-02
## f.duration-(1.2e+03,2.1e+03] -0.08146358  0.1836311251  1.153212e-01
##
## $contrib
##                                 Dim 1       Dim 2        Dim 3
## f.duration-[5,60]           48.095757  2.87482957 7.586705e+00
## f.duration-(60,120]          7.657254 29.24828562 1.816559e+01
## f.duration-(120,150]         5.263239  0.02162046 2.396431e+00
## f.duration-(150,180]         1.632767  0.00393518 5.627155e-05
## f.duration-(180,240]         3.849653  7.05820696 3.437761e+01
## f.duration-(240,300]         1.913324  4.48550815 1.697988e-03
## f.duration-(300,1.2e+03]    25.376484  6.08781771 1.498386e+01
## f.duration-(1.2e+03,2.1e+03]  6.211521 50.21979635 2.248806e+01
```

```
##
## $cos2
##                                   Dim 1       Dim 2        Dim 3
## f.duration-[5,60]             0.8889838 0.033395490 7.762072e-02
## f.duration-(60,120]           0.2121463 0.509274053 2.785796e-01
## f.duration-(120,150]          0.7970605 0.002057746 2.008818e-01
## f.duration-(150,180]          0.9984686 0.001512388 1.904743e-05
## f.duration-(180,240]          0.1409384 0.162401758 6.966599e-01
## f.duration-(240,300]          0.4042264 0.595574996 1.985675e-04
## f.duration-(300,1.2e+03]      0.6767699 0.102037507 2.211926e-01
## f.duration-(1.2e+03,2.1e+03] 0.1236833 0.628458833 2.478578e-01
##
## $inertia
## [1] 7.645138e-04 5.100466e-04 9.331130e-05 2.310800e-05 3.859799e-04
## [6] 6.688615e-05 5.298621e-04 7.096744e-04
```

```
# Link levels in rows
plot.CA(res.ca)
lines(res.ca$row$coord[,1],res.ca$row$coord[,2],lwd=2,col="darkblue")
lines(res.ca$col$coord[,1],res.ca$col$coord[,2],lwd=2,col="red")
```




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

## [1] 0.003083382

# H0: f.duration - f.age independency
chisq.test(table(df$f.age,duration_k8))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(df$f.age, duration_k8)
## X-squared = 15.386, df = 21, p-value = 0.8031
```

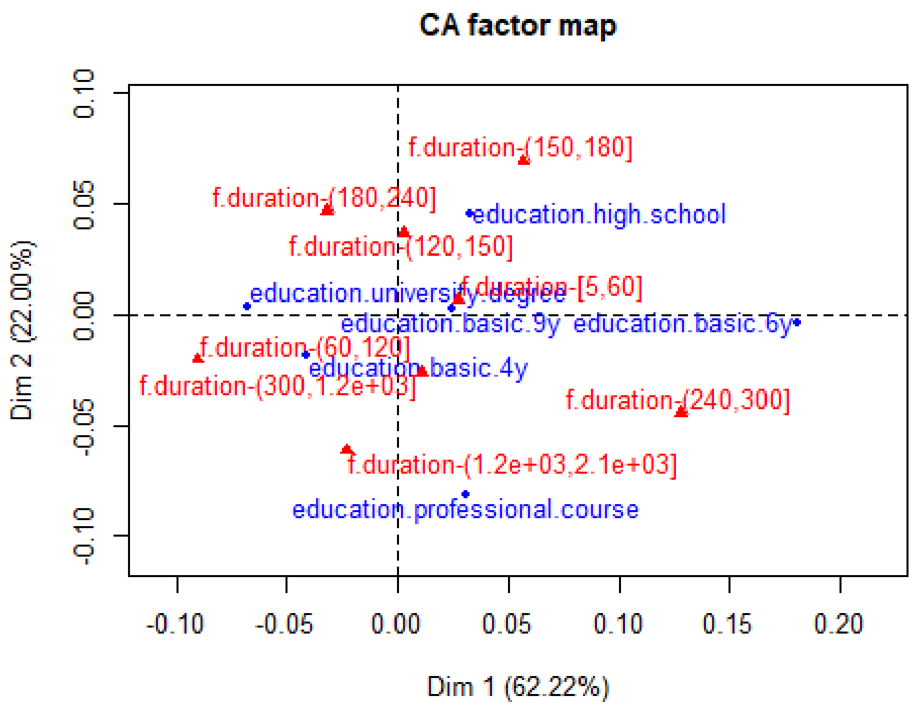
CA - Education vs f.duration

Para la segunda prueba decidimos utilizar duration junto con education. Para education usaremos la variable original con todos sus niveles menos el nivel illiterate el cual nos puede causar inconvenientes.

Por kaiser vemos que las primeras dos dimensiones están por encima de la media, por lo que son las que cogemos.

Podemos ver que para la primera dimensión los valores más lejanos del centro son los niveles de education.basic_6y, education.university.degree, para la primera dimensión. Para la segunda tenemos, f.duration(150,180], education.professioal.course, esto nos puede decir qué niveles se ven mejor representados en las dimensiones.

```
#Education
table(df$education)

##
##            education.basic.4y            education.basic.6y
##                           515                           271
##            education.basic.9y          education.high.school
##                           810                          1196
##          education.illiterate education.professional.course
##                             1                           633
##   education.university.degree
##                          1564

education_k6<-df$education
education_k6[which(education_k6=="education.illiterate")]<-"education.basic.4y"
education_k6=factor(education_k6)

par(cex=0.8)
res.ca<-CA(table(education_k6,duration_k8))
```




```r
res.ca$eig
```

```
##         eigenvalue percentage of variance
## dim 1 3.865234e-03             62.2179391
## dim 2 1.366656e-03             21.9987982
## dim 3 7.963697e-04             12.8190136
## dim 4 1.318112e-04              2.1217398
## dim 5 5.234014e-05              0.8425093
##       cumulative percentage of variance
## dim 1                          62.21794
## dim 2                          84.21674
## dim 3                          97.03575
## dim 4                          99.15749
## dim 5                         100.00000
```

```r
mean(res.ca$eig[,1])
```

```
## [1] 0.001242482
```

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

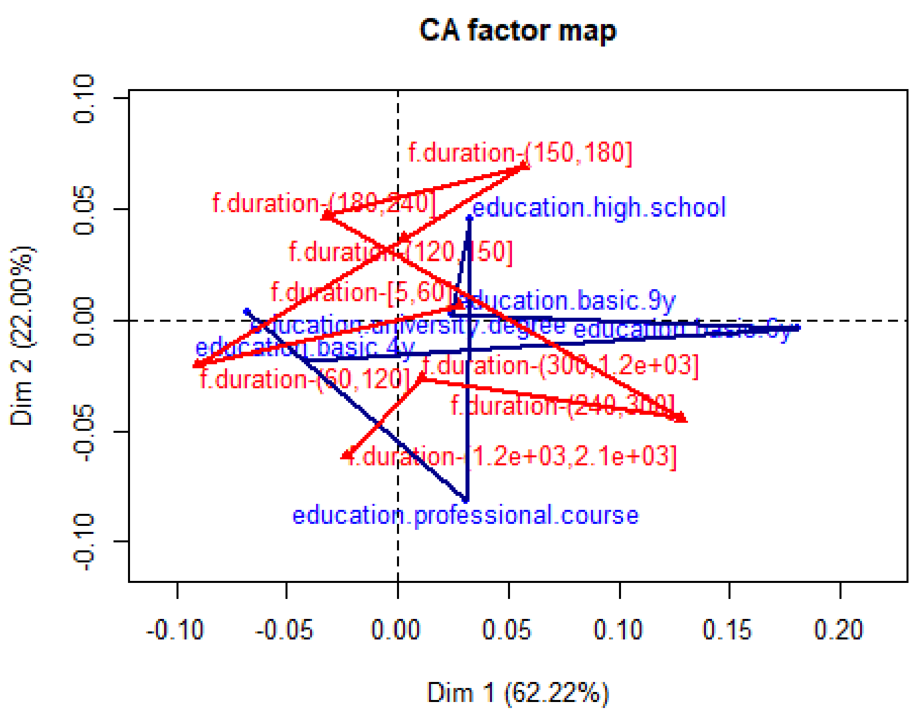
```
## $coord
##                                    Dim 1        Dim 2        Dim 3
## education.basic.4y           -0.04133205 -0.018121606 -0.058412332
## education.basic.6y            0.18102401 -0.004110569  0.042729415
## education.basic.9y            0.02483545  0.002685184 -0.027046691
## education.high.school         0.03234069  0.045050136 -0.004194745
```

```
## education.professional.course  0.03069724 -0.081359937  0.007155999
## education.university.degree   -0.06774789  0.003779150  0.026186751
##                                       Dim 4        Dim 5
## education.basic.4y             0.010678244 -0.012189412
## education.basic.6y            -0.003590873 -0.018490687
## education.basic.9y            -0.022417259  0.003721054
## education.high.school          0.010664921  0.005274952
## education.professional.course  0.008039829  0.008059669
## education.university.degree   -0.002700335 -0.001997413
##
## $contrib
##                                   Dim 1       Dim 2      Dim 3      Dim 4
## education.basic.4y             4.570327  2.48475408 44.3040567  8.9453332
## education.basic.6y            46.043183  0.06714497 12.4511048  0.5312717
## education.basic.9y             2.590316  0.08563935 14.9106984 61.8866872
## education.high.school          6.485648 35.59293113  0.5295743 20.6820512
## education.professional.course  3.092616 61.44198967  0.8156972  6.2207761
## education.university.degree   37.217909  0.32754080 26.9888687  1.7338807
##                                   Dim 5
## education.basic.4y            29.354847
## education.basic.6y            35.476433
## education.basic.9y             4.294194
## education.high.school         12.741859
## education.professional.course 15.743551
## education.university.degree    2.389116
##
## $cos2
##                                   Dim 1       Dim 2       Dim 3
## education.basic.4y            0.2991135 0.057498362 0.597408243
## education.basic.6y            0.9371554 0.000483218 0.052214742
## education.basic.9y            0.3295023 0.003851787 0.390789233
## education.high.school         0.3233541 0.627439873 0.005439900
## education.professional.course 0.1217065 0.854941382 0.006613871
## education.university.degree   0.8658182 0.002694161 0.129359520
##                                      Dim 4        Dim 5
## education.basic.4y            0.0199646527 0.0260152290
## education.basic.6y            0.0003687563 0.0097778938
## education.basic.9y            0.2684598606 0.0073968422
## education.high.school         0.0351637465 0.0086023464
## education.professional.course 0.0083485069 0.0083897596
## education.university.degree   0.0013755312 0.0007526117
##
## $inertia
## [1] 0.0005905913 0.0018990198 0.0003038576 0.0007752660 0.0009821731
## [6] 0.0016615026

# Columns: the same
res.ca$col
```

```
## $coord
##                                    Dim 1        Dim 2        Dim 3
## f.duration-[5,60]             0.027029115  0.006192711  0.035619391
## f.duration-(60,120]          -0.090112862 -0.020538411  0.017086167
## f.duration-(120,150]          0.002502632  0.036564696 -0.049483431
## f.duration-(150,180]          0.056880902  0.069080817  0.004342747
## f.duration-(180,240]         -0.032420109  0.046836380  0.019636428
## f.duration-(240,300]          0.127883956 -0.044401648  0.030160547
## f.duration-(300,1.2e+03]      0.010833855 -0.026353481 -0.028797208
## f.duration-(1.2e+03,2.1e+03] -0.023054621 -0.061622025 -0.029181692
##                                     Dim 4        Dim 5
## f.duration-[5,60]             0.0049402640  0.001195881
## f.duration-(60,120]          -0.0054537007  0.003195489
## f.duration-(120,150]         -0.0047413221  0.015186085
## f.duration-(150,180]         -0.0004840916 -0.004166702
## f.duration-(180,240]          0.0067808780 -0.005449376
## f.duration-(240,300]         -0.0044409552  0.006544906
## f.duration-(300,1.2e+03]     -0.0015812648 -0.007781878
## f.duration-(1.2e+03,2.1e+03]  0.0922854018  0.011622914
##
## $contrib
##                                   Dim 1      Dim 2      Dim 3       Dim 4
## f.duration-[5,60]             1.85602333  0.2755485 15.6442142  1.81820805
## f.duration-(60,120]          44.92226773  6.5999164  7.8385985  4.82496535
## f.duration-(120,150]          0.01610644  9.7240230 30.5622963  1.69522671
## f.duration-(150,180]          7.88414319 32.8891869  0.2230549  0.01674558
## f.duration-(180,240]          3.30236637 19.4930424  5.8800674  4.23634841
## f.duration-(240,300]         41.03941655 13.9921319 11.0792023  1.45126036
## f.duration-(300,1.2e+03]      0.79779697 13.3511603 27.3582403  0.49837825
## f.duration-(1.2e+03,2.1e+03]  0.18187944  3.6749905  1.4143262 85.45886729
##                                  Dim 5
## f.duration-[5,60]             0.2683099
## f.duration-(60,120]           4.1716088
## f.duration-(120,150]         43.7963518
## f.duration-(150,180]          3.1242607
## f.duration-(180,240]          6.8901823
## f.duration-(240,300]          7.9381048
## f.duration-(300,1.2e+03]     30.3973777
## f.duration-(1.2e+03,2.1e+03]  3.4138039
##
## $cos2
##                                   Dim 1      Dim 2       Dim 3
## f.duration-[5,60]            0.354045576 0.01858477 0.614849014
## f.duration-(60,120]          0.915064957 0.04753484 0.032897861
## f.duration-(120,150]         0.001548393 0.33053014 0.605350244
## f.duration-(150,180]         0.402214863 0.59325319 0.002344522
## f.duration-(180,240]         0.283613187 0.59192152 0.104045298
## f.duration-(240,300]         0.847459925 0.10216102 0.047137383
## f.duration-(300,1.2e+03]     0.068871807 0.40752236 0.486604635
## f.duration-(1.2e+03,2.1e+03] 0.038426381 0.27452723 0.061565082
```

```
##                                    Dim 4        Dim 5
## f.duration-[5,60]           1.182758e-02 0.0006930612
## f.duration-(60,120]         3.351668e-03 0.0011506756
## f.duration-(120,150]        5.557584e-03 0.0570136422
## f.duration-(150,180]        2.913267e-05 0.0021582909
## f.duration-(180,240]        1.240707e-02 0.0080129231
## f.duration-(240,300]        1.021974e-03 0.0022196981
## f.duration-(300,1.2e+03]    1.467184e-03 0.0355340175
## f.duration-(1.2e+03,2.1e+03] 6.157147e-01 0.0097666035
##
## $inertia
## [1] 0.0002026283 0.0018975163 0.0004020629 0.0007576562 0.0004500643
## [6] 0.0018717928 0.0004477408 0.0001829489
```

```
# Link levels in rows
plot.CA(res.ca)
lines(res.ca$row$coord[,1],res.ca$row$coord[,2],lwd=2,col="darkblue")
lines(res.ca$col$coord[,1],res.ca$col$coord[,2],lwd=2,col="red")
```




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

## [1] 0.00621241

chisq.test(table(education_k6,duration_k8))

## Warning in chisq.test(table(education_k6, duration_k8)): Chi-squared
## approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(education_k6, duration_k8)
## X-squared = 31, df = 35, p-value = 0.6617
```

```r
# Traditional analysis
table(df$y,duration_k8)
```

```
##        duration_k8
##         f.duration-[5,60] f.duration-(60,120] f.duration-(120,150]
##   y.no                490                1047                  481
##   y.yes                 0                  20                   15
##        duration_k8
##         f.duration-(150,180] f.duration-(180,240] f.duration-(240,300]
##   y.no                   441                  550                  431
##   y.yes                   29                   56                   53
##        duration_k8
##         f.duration-(300,1.2e+03] f.duration-(1.2e+03,2.1e+03]
##   y.no                       983                           25
##   y.yes                      328                           41
```

```r
chisq.test(table(df$y,duration_k8))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(df$y, duration_k8)
## X-squared = 643.03, df = 7, p-value < 2.2e-16
```