Deliverable 2

Guillem Valls, Sergio Mazzariol

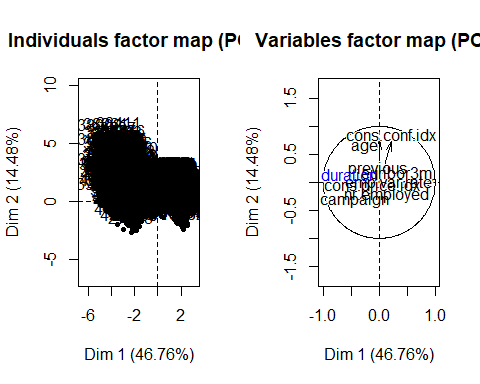
Table of Contents

## Valores propios y ejes dominantes

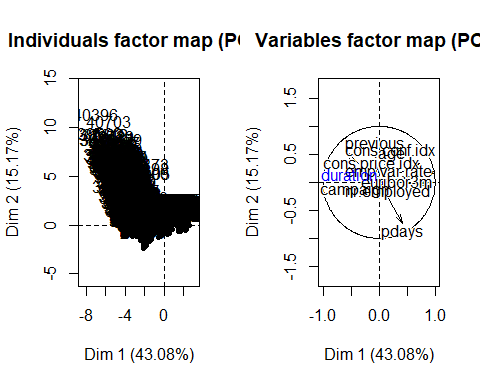
### Eigenvalues and dominant axes analysis. How many axes we have to interpret according to Kaiser and Elbow’s rule?

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

vars\_num\_sin\_pday = vars\_num[-3];  
par(mfrow=c(1,2))  
res2.pca<-PCA(df[,c('duration',vars\_num\_sin\_pday)],quanti.sup=1)



res.pca<-PCA(df[,c('duration',vars\_num)],quanti.sup=1)



par(mfrow=c(1,1))

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 graficamente tiene mayor magnitud ademas que las variables socio economicas, se ven mejor representadas en el primer eje factorial.

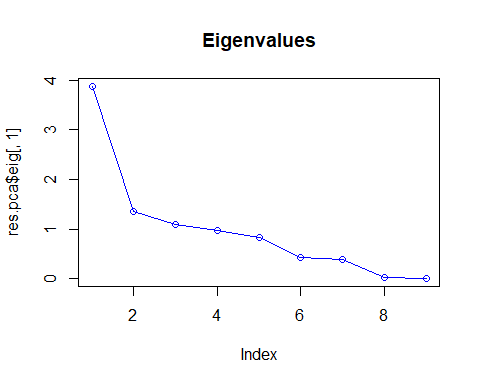
Por la ley de Kaiser, deberiamos utilizar los 3 primeros ejes factoriales, los cuales son mayores a 1. Por la ley de ElBow, al realizar el grafico podemos ver que la grafica empieza a ser plana a partir de la 2da dimension, es decir que se cogen las 2 primeras dimensiones. Si tomamos en cuenta el criterio del 80% se deberian coger las 4 primeras dimensiones.

Para realizar el futuro analisi, combiene utilizar dimensiones pares, por lo que decidimos de 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")

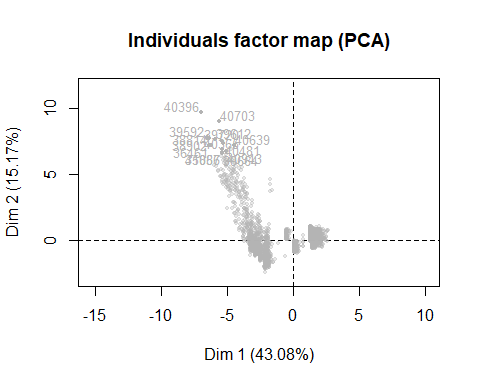


# Individuals point of view: Are they any individuals “too contributive”? To better understand the axes meaning use the extreme individuals. Detection of multivariant outliers and influent data.

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 significante, 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, la duración de la llamada ha sido mayor a los 300s.

Para la dimension 2 podemos ver practicamente las mismas caracteristas menos la duración que ha sido menor. Cabe a destacar que sin embargo 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.

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



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] # ver si hay alguna caracteristica relevnte, ver los 5 mas relevantes de este eje y del eje 2, ver que comparten en común.  
  
# mas\_ctr\_dim1  
# 40396 39592 38902 38814 36461   
#0.2482556 0.2196878 0.2057888 0.2053101 0.1994545   
# mas\_ctr\_dim2  
# 40396 40703 39592 38902 39612   
#1.3837961 1.1876086 0.8973750 0.8642930 0.8630042   
  
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 f.education  
## 40396 y.yes f.job.Admin-Managment f.season.Jun-Aug Professional  
## 39592 y.yes f.job.Not-working f.season.Mar-May High School  
## 38902 y.yes f.job.Entrep-Retired-selfEmpl f.season.Sep-Dec Basic  
## 38814 y.yes f.job.Entrep-Retired-selfEmpl f.season.Sep-Dec Professional  
## 36461 y.no f.job.Entrep-Retired-selfEmpl f.season.Jun-Aug Basic  
## f.age f.duration f.campaign  
## 40396 f.age-(40,50] f.duration-(120,180] f.campaign-(2,20]  
## 39592 f.age-[18,30] f.duration-(180,300] f.campaign-[0,1]  
## 38902 f.age-(50,92] f.duration-(180,300] f.campaign-[0,1]  
## 38814 f.age-(50,92] f.duration-(180,300] f.campaign-[0,1]  
## 36461 f.age-(50,92] f.duration-(300,2.1e+03] f.campaign-(1,2]  
## f.pdays f.previous  
## 40396 f.pdays-[0,22] f.previous-(1,6]  
## 39592 f.pdays-[0,22] f.previous-(1,6]  
## 38902 f.pdays-[0,22] f.previous-(1,6]  
## 38814 f.pdays-[0,22] f.previous-(1,6]  
## 36461 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 f.education  
## 40396 y.yes f.job.Admin-Managment f.season.Jun-Aug Professional  
## 40703 y.no f.job.Entrep-Retired-selfEmpl f.season.Sep-Dec Professional  
## 39592 y.yes f.job.Not-working f.season.Mar-May High School  
## 38902 y.yes f.job.Entrep-Retired-selfEmpl f.season.Sep-Dec Basic  
## 39612 y.yes f.job.Serv-Tech-BlueC f.season.Mar-May Professional  
## f.age f.duration f.campaign f.pdays  
## 40396 f.age-(40,50] f.duration-(120,180] f.campaign-(2,20] f.pdays-[0,22]  
## 40703 f.age-(50,92] f.duration-[5,120] f.campaign-(2,20] f.pdays-[0,22]  
## 39592 f.age-[18,30] f.duration-(180,300] f.campaign-[0,1] f.pdays-[0,22]  
## 38902 f.age-(50,92] f.duration-(180,300] f.campaign-[0,1] f.pdays-[0,22]  
## 39612 f.age-(50,92] f.duration-(180,300] f.campaign-[0,1] f.pdays-[0,22]  
## f.previous  
## 40396 f.previous-(1,6]  
## 40703 f.previous-(1,6]  
## 39592 f.previous-(1,6]  
## 38902 f.previous-(1,6]  
## 39612 f.previous-(1,6]

# which(res.pca$ind$dist>=max(res.pca$ind$dist)) # encontrar el ind con mayor distancia al centro  
  
# pendiente apuntar los indv mas ctr en colores diferente

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

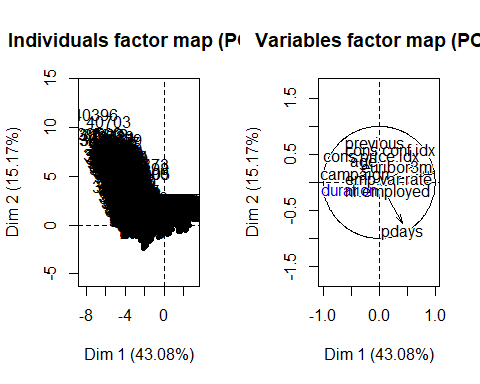
Al hacer el PCA con la variable target duration como suplementaria, podemos ver que su modulo es practicamente nulo, esto quiere decir que la variable no se ve representada en niguno de los ejes factoriles.

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

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

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

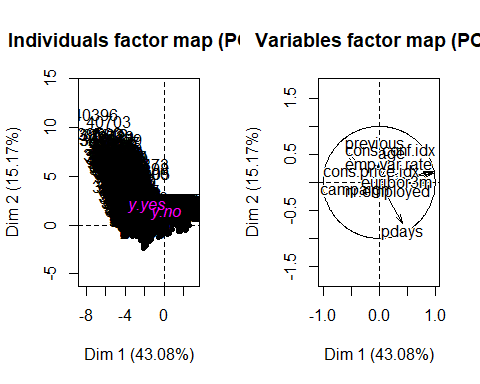
par(mfrow=c(1,2))  
res.pca<-PCA(df[,c('duration',vars\_num)],quanti.sup=1)



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)



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

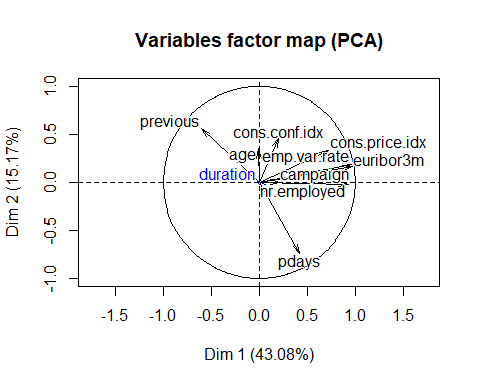
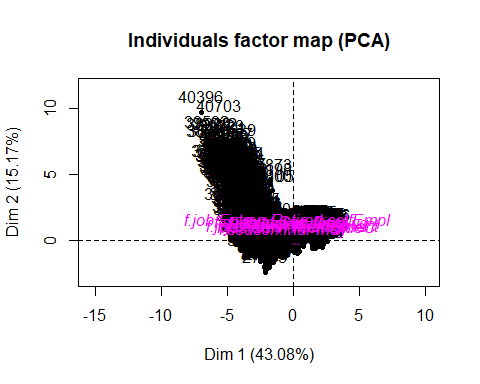
##   
## Call:  
## PCA(X = df[, c("y", vars\_num)], quali.sup = 1)   
##   
##   
## Eigenvalues  
## Dim.1 Dim.2 Dim.3 Dim.4 Dim.5 Dim.6 Dim.7  
## Variance 3.88 1.36 1.10 0.97 0.83 0.43 0.39  
## % of var. 43.08 15.17 12.28 10.74 9.25 4.82 4.29  
## Cumulative % of var. 43.08 58.24 70.52 81.26 90.50 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 |

par(mfrow=c(1,1))

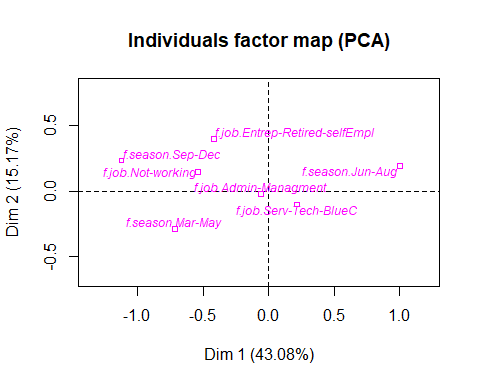
### Perform a PCA taking into account also supplementary variables the supplementary variables can be quantitative and/or categorical

Hemos dividido el plot en diferentes partes, para así poder entender y ver mejor el resultado. Para el primer plot, tenemos f.job y f.season para los cuales podemos ver que

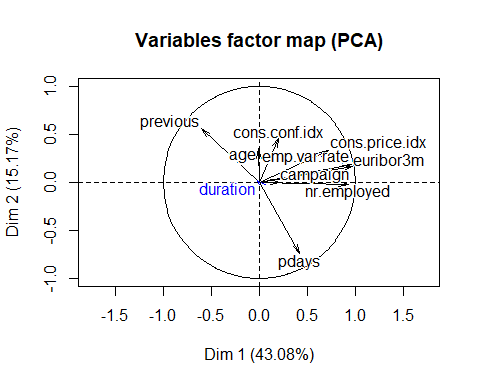
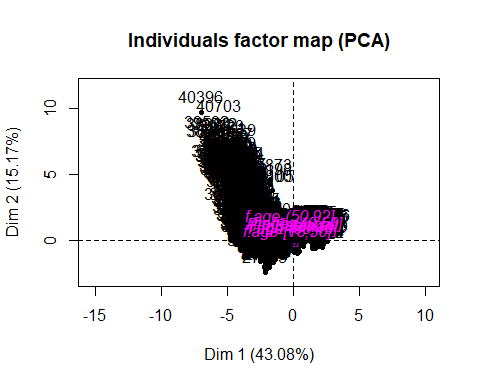
# preguntar y pendiente cambiar de color cada variable con sus factores  
  
 vars\_factorizadas<- c("f.job","f.season","f.education","f.age","f.duration","f.campaign","f.pdays","f.previous","y");  
# vars\_factorizadas<- c("f.job","f.season","f.education","f.age","f.pdays","y");  
  
res.pca<-PCA(df[,c('duration',vars\_num, "f.job","f.season")],quanti.sup=1, quali.sup = c(11:12))



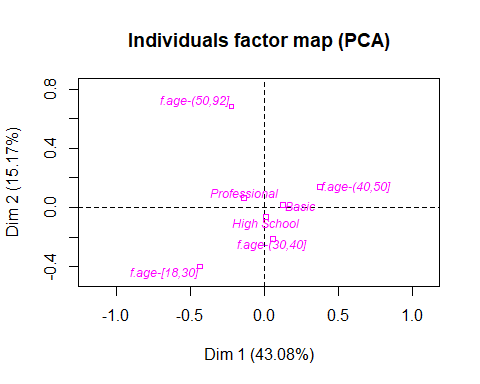
plot.PCA(res.pca,choix="ind",invisible="ind",cex=0.75)



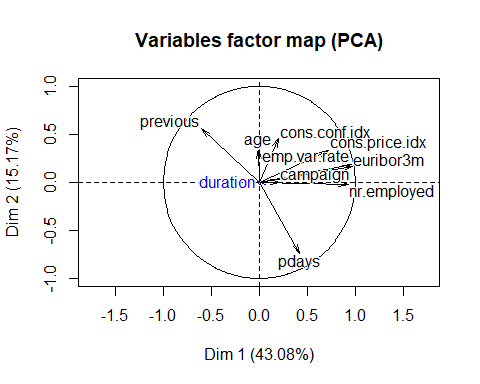
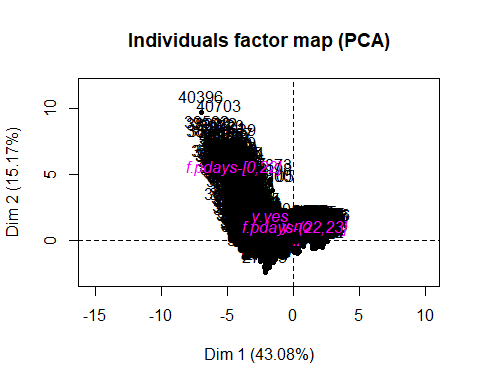
res.pca<-PCA(df[,c('duration',vars\_num,"f.education","f.age")],quanti.sup=1, quali.sup = c(11:12))



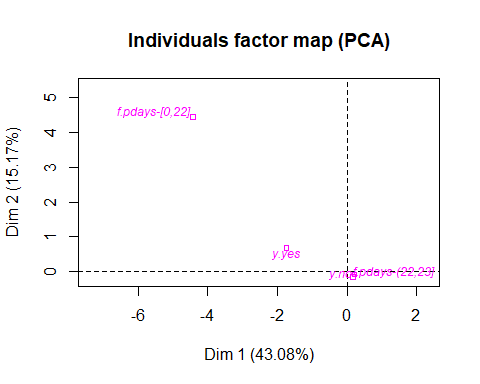
plot.PCA(res.pca,choix="ind",invisible="ind",cex=0.75)



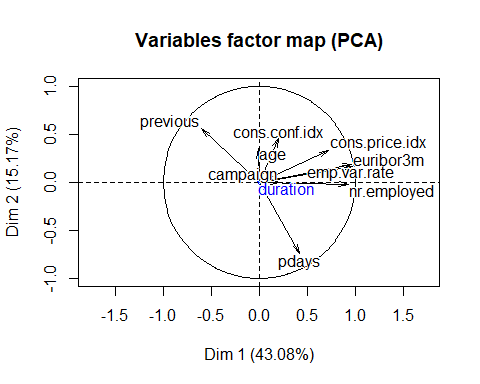
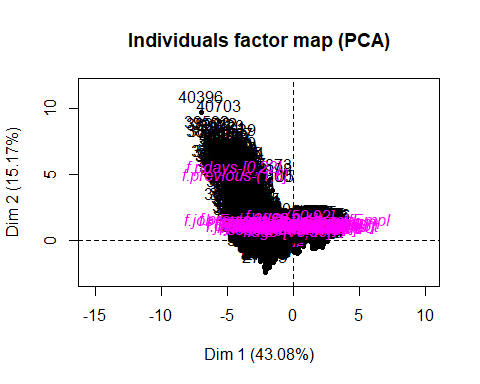
res.pca<-PCA(df[,c('duration',vars\_num,"f.pdays","y")],quanti.sup=1, quali.sup = c(11:12))



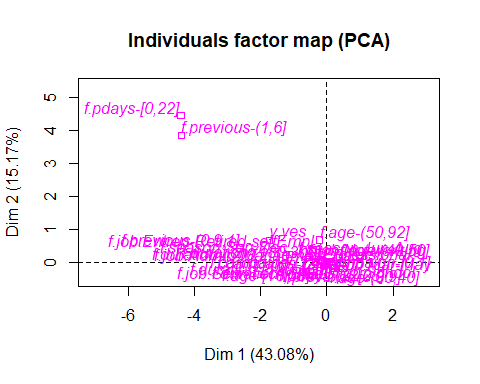
plot.PCA(res.pca,choix="ind",invisible="ind",cex=0.75)



res.pca<-PCA(df[,c('duration',vars\_num,vars\_factorizadas)],quanti.sup=1, quali.sup = c(11:19))



plot.PCA(res.pca,choix="ind",invisible="ind")# no ponerlas todas a la vez, poner por grupos. Pntar grupos de variables, si categorias quedan en el medio no valen de nada, hay que ver si categorias de las variables que estan lejos del centrode gravedad, son las que valen la pena. Fijarse que no cambian los valores de los ejes porque si no no tendremos referencia para comparar



summary(res.pca,dig = 2, nbelements= 50, nbind=3, ncp=4)

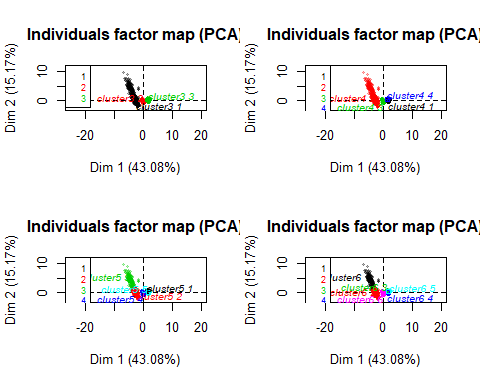
##   
## Call:  
## PCA(X = df[, c("duration", vars\_num, vars\_factorizadas)], quanti.sup = 1,   
## quali.sup = c(11:19))   
##   
##   
## Eigenvalues  
## Dim.1 Dim.2 Dim.3 Dim.4 Dim.5 Dim.6  
## Variance 3.877 1.365 1.105 0.966 0.832 0.434  
## % of var. 43.078 15.166 12.276 10.736 9.246 4.823  
## 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  
## 20 | 1.739 | 1.291 0.009 0.551 | 0.441  
## 21 | 1.990 | 1.295 0.009 0.424 | 0.184  
## 30 | 2.240 | 1.284 0.009 0.328 | 0.898  
## ctr cos2 Dim.3 ctr cos2   
## 20 0.003 0.064 | 0.414 0.003 0.057 |  
## 21 0.000 0.009 | -0.131 0.000 0.004 |  
## 30 0.012 0.161 | 1.383 0.035 0.381 |  
## Dim.4 ctr cos2   
## 20 -0.815 0.014 0.220 |  
## 21 -1.062 0.023 0.285 |  
## 30 -0.376 0.003 0.028 |  
##   
## 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 Dim.3 ctr cos2 Dim.4  
## age 0.122 | 0.666 40.097 0.443 | 0.282  
## campaign 0.000 | -0.229 4.736 0.052 | 0.932  
## pdays 0.549 | 0.280 7.095 0.078 | 0.040  
## previous 0.312 | -0.304 8.383 0.093 | -0.023  
## emp.var.rate 0.035 | -0.105 1.003 0.011 | -0.061  
## cons.price.idx 0.111 | -0.299 8.079 0.089 | -0.063  
## cons.conf.idx 0.210 | 0.580 30.441 0.336 | -0.035  
## euribor3m 0.027 | -0.011 0.010 0.000 | -0.070  
## nr.employed 0.000 | -0.041 0.156 0.002 | -0.048  
## ctr cos2   
## age 8.232 0.080 |  
## campaign 89.885 0.869 |  
## pdays 0.166 0.002 |  
## previous 0.053 0.001 |  
## emp.var.rate 0.386 0.004 |  
## cons.price.idx 0.406 0.004 |  
## cons.conf.idx 0.129 0.001 |  
## euribor3m 0.507 0.005 |  
## nr.employed 0.236 0.002 |  
##   
## Supplementary continuous variable  
## Dim.1 cos2 Dim.2 cos2 Dim.3  
## duration | -0.022 0.000 | 0.017 0.000 | 0.004  
## cos2 Dim.4 cos2   
## duration 0.000 | -0.055 0.003 |  
##   
## 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 |  
## Basic | 0.294 | 0.125 0.181 3.071 |  
## High School | 0.155 | 0.010 0.005 0.285 |  
## 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 Dim.3 cos2  
## f.job.Admin-Managment -0.020 0.027 -0.869 | -0.025 0.038  
## f.job.Entrep-Retired-selfEmpl 0.399 0.153 8.730 | 0.661 0.420  
## f.job.Not-working 0.143 0.043 2.353 | -0.142 0.043  
## f.job.Serv-Tech-BlueC -0.101 0.139 -5.958 | -0.121 0.197  
## f.season.Mar-May -0.288 0.132 -14.633 | -0.046 0.003  
## f.season.Jun-Aug 0.194 0.036 10.604 | -0.065 0.004  
## f.season.Sep-Dec 0.234 0.035 5.626 | 0.353 0.079  
## Basic 0.015 0.003 0.613 | 0.094 0.102  
## High School -0.066 0.180 -3.025 | -0.122 0.619  
## Professional 0.062 0.059 2.527 | 0.047 0.034  
## f.age-[18,30] -0.400 0.093 -11.108 | -0.859 0.429  
## f.age-(30,40] -0.213 0.193 -10.495 | -0.312 0.412  
## f.age-(40,50] 0.139 0.053 4.874 | 0.289 0.228  
## f.age-(50,92] 0.682 0.168 19.031 | 1.148 0.474  
## f.duration-[5,120] -0.094 0.118 -3.822 | -0.014 0.002  
## f.duration-(120,180] 0.014 0.010 0.414 | 0.078 0.318  
## f.duration-(180,300] 0.085 0.147 2.721 | -0.028 0.016  
## f.duration-(300,2.1e+03] 0.029 0.030 1.081 | -0.017 0.011  
## f.campaign-[0,1] 0.003 0.000 0.167 | 0.168 0.056  
## f.campaign-(1,2] 0.014 0.004 0.486 | 0.047 0.045  
## f.campaign-(2,20] -0.015 0.000 -0.627 | -0.258 0.057  
## f.pdays-[0,22] 4.435 0.454 51.122 | -1.540 0.055  
## f.pdays-(22,23] -0.161 0.454 -51.122 | 0.056 0.055  
## f.previous-[0,0.9] -0.181 0.111 -27.027 | 0.105 0.037  
## f.previous-(0.9,1] 0.442 0.025 9.534 | -0.453 0.026  
## f.previous-(1,6] 3.839 0.396 38.998 | -1.424 0.054  
## y.no -0.082 0.126 -14.153 | -0.009 0.002  
## y.yes 0.671 0.126 14.153 | 0.074 0.002  
## v.test Dim.4 cos2 v.test   
## f.job.Admin-Managment -1.156 | -0.004 0.001 -0.182 |  
## f.job.Entrep-Retired-selfEmpl 16.069 | 0.261 0.065 6.781 |  
## f.job.Not-working -2.593 | -0.119 0.030 -2.320 |  
## f.job.Serv-Tech-BlueC -7.883 | -0.043 0.025 -2.995 |  
## f.season.Mar-May -2.593 | 0.025 0.001 1.499 |  
## f.season.Jun-Aug -3.949 | 0.005 0.000 0.336 |  
## f.season.Sep-Dec 9.443 | -0.092 0.005 -2.637 |  
## Basic 4.327 | 0.026 0.008 1.298 |  
## High School -6.235 | -0.044 0.081 -2.411 |  
## Professional 2.125 | 0.025 0.009 1.199 |  
## f.age-[18,30] -26.514 | -0.328 0.063 -10.839 |  
## f.age-(30,40] -17.065 | -0.124 0.065 -7.247 |  
## f.age-(40,50] 11.246 | 0.096 0.025 3.983 |  
## f.age-(50,92] 35.588 | 0.470 0.080 15.598 |  
## f.duration-[5,120] -0.616 | 0.157 0.327 7.577 |  
## f.duration-(120,180] 2.573 | -0.075 0.293 -2.639 |  
## f.duration-(180,300] -1.003 | -0.077 0.121 -2.940 |  
## f.duration-(300,2.1e+03] -0.709 | -0.063 0.145 -2.803 |  
## f.campaign-[0,1] 9.685 | -0.580 0.674 -35.849 |  
## f.campaign-(1,2] 1.837 | -0.206 0.872 -8.607 |  
## f.campaign-(2,20] -11.948 | 0.926 0.730 45.906 |  
## f.pdays-[0,22] -19.728 | -0.215 0.001 -2.947 |  
## f.pdays-(22,23] 19.728 | 0.008 0.001 2.947 |  
## f.previous-[0,0.9] 17.452 | 0.004 0.000 0.797 |  
## f.previous-(0.9,1] -10.856 | 0.008 0.000 0.209 |  
## f.previous-(1,6] -16.077 | -0.174 0.001 -2.101 |  
## y.no -1.738 | 0.005 0.001 1.090 |  
## y.yes 1.738 | -0.043 0.001 -1.090 |

#lines(res.pca$quali.sup$coord[1:2,1:2],col="blue",lwd=2)  
#lines(res.pca$quali.sup$coord[3:6,1:2],col="cyan",lwd=2)

## K-Means Classification

Hemos graficado los grupos separados en 3, 4, 5 y 6 clusters, para los cuales nos parece que graficamente con 4 clusters los grupos están bien definidos, por lo que decidimos usar con 4 clusters los cuals analisaremos 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);  
  
#summary(kcla)  
#table(kcla$cluster)  
#kcla$betweenss/kcla$totss  
  
par(mfrow=c(2,2))  
res.pca<-PCA(df[,c('duration',vars\_num, "cluster3")],quanti.sup=1, quali.sup = 11, 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 = 11, 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 = 11, 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 = 11, graph=FALSE)  
plot.PCA(res.pca,choix="ind",habillage=11,select=0 ,cex=0.75)



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

## Description of clusters

Viendo el chi-square test, podemos saber que variables se utilizarán para caracterizar nuestros 4 clusters. Viendo las categorias donde el P-value es casi 0 podemos ver que las categorias 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 cateogrias de estas variables son las que caracterizan estos clusters.

Para el catdes del cluster 1, hemos podido ver las categorias 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 mayoria por teléfono local, también podemos ver que este cluster esta ligeramente relacionado con la categoria f.job.Serv-Tech-BlueC. En conclusión podemos decir que este cluster esta 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 es en media servicio, tecnicos o blue collar. - No compraron el producto en su mayoria.

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

Para el cluster 3, tenemos: - Temporada de Sep-Dec - Contactdos por móvil en su mayoria - La categoria de job.management esta sobrerepresentada - No han adquirido el producto - Una gran cantidad de individuos rechazo el productoy y.no

Para el cluster 4, podemos ver que aglutina individuos muy bien caracterizado 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]

# preguntar si hay que analizar Description of each cluster by quantitative variables de catdes.

catdes(df, 30, proba = 0.005)

##   
## Link between the cluster variable and the categorical variables (chi-square test)  
## =================================================================================  
## p.value df  
## month 0.000000e+00 27  
## f.season 0.000000e+00 6  
## contact 2.861952e-283 3  
## poutcome 1.757571e-264 6  
## f.previous 2.446737e-263 6  
## f.age 4.712447e-251 9  
## y 2.426974e-77 3  
## f.pdays 8.592195e-77 3  
## job 8.264472e-72 30  
## default 1.320138e-60 3  
## marital 7.308021e-60 6  
## f.campaign 4.071141e-24 6  
## education 1.202218e-21 18  
## f.job 3.395894e-20 9  
## f.education 2.313466e-18 6  
## housing 9.301649e-08 3  
## day\_of\_week 6.645107e-07 12  
## f.duration 2.006377e-06 9  
##   
## Description of each cluster by the categories  
## =============================================  
## $`1`  
## Cla/Mod Mod/Cla Global  
## f.season=f.season.Jun-Aug 56.9265033 91.35096497 44.9899800  
## month=month.jul 79.8843931 49.39242316 17.3346693  
## f.previous=f.previous-[0,0.9] 32.6183260 100.00000000 85.9519038  
## poutcome=poutcome.nonexistent 32.6183260 100.00000000 85.9519038  
## f.age=f.age-[18,30] 48.6206897 30.23588277 17.4348697  
## f.age=f.age-(30,40] 37.1170266 52.82344532 39.8997996  
## month=month.jun 48.3766234 21.30092924 12.3446894  
## f.pdays=f.pdays-(22,23] 29.0550363 100.00000000 96.4929860  
## marital=marital.single 36.9241673 37.24088635 28.2765531  
## f.campaign=f.campaign-(2,20] 34.9068323 40.17155111 32.2645291  
## y=y.no 29.5863309 94.06719085 89.1382766  
## job=job.technician 37.5764994 21.94424589 16.3727455  
## month=month.aug 37.8272251 20.65761258 15.3106212  
## f.job=f.job.Serv-Tech-BlueC 30.9405941 53.60972123 48.5771543  
## job=job.admin. 32.2530864 29.87848463 25.9719439  
## f.education=High School 30.5631493 39.95711222 36.6533066  
## month=month.dec 0.0000000 0.00000000 0.3607214  
## marital=marital.divorced 22.8130360 9.50679056 11.6833667  
## job=job.management 21.0826211 5.28949249 7.0340681  
## f.job=f.job.Not-working 20.7602339 5.07505361 6.8537074  
## day\_of\_week=day\_of\_week.fri 23.9289446 16.36883488 19.1783567  
## job=job.student 11.0091743 0.85775554 2.1843687  
## f.education=Basic 23.7319975 27.09077913 32.0040080  
## f.campaign=f.campaign-[0,1] 23.7152287 35.95425304 42.5050100  
## marital=marital.married 24.8664887 53.25232309 60.0400802  
## month=month.mar 0.0000000 0.00000000 1.2625251  
## month=month.oct 1.2500000 0.07147963 1.6032064  
## month=month.sep 0.0000000 0.00000000 1.4028056  
## education=education.basic.4y 15.1456311 5.57541101 10.3206413  
## y=y.yes 15.3136531 5.93280915 10.8617234  
## f.job=f.job.Entrep-Retired-selfEmpl 14.9046794 6.14724803 11.5631263  
## f.age=f.age-(40,50] 18.5953711 16.65475340 25.1102204  
## f.previous=f.previous-(1,6] 0.0000000 0.00000000 2.7454910  
## poutcome=poutcome.success 0.0000000 0.00000000 3.1262525  
## job=job.retired 2.7027027 0.42887777 4.4488978  
## 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  
## f.season=f.season.Sep-Dec 0.1468429 0.07147963 13.6472946  
## f.age=f.age-(50,92] 0.4566210 0.28591851 17.5551102  
## month=month.may 7.1428571 8.57755540 33.6673347  
## f.season=f.season.Mar-May 5.8139535 8.57755540 41.3627255  
## p.value v.test  
## f.season=f.season.Jun-Aug 0.000000e+00 Inf  
## month=month.jul 7.361032e-278 35.618058  
## f.previous=f.previous-[0,0.9] 2.827478e-110 22.311973  
## poutcome=poutcome.nonexistent 2.827478e-110 22.311973  
## f.age=f.age-[18,30] 2.444971e-46 14.292273  
## f.age=f.age-(30,40] 6.785349e-31 11.557242  
## month=month.jun 1.589105e-30 11.483919  
## f.pdays=f.pdays-(22,23] 2.910853e-26 10.602189  
## marital=marital.single 5.721658e-18 8.637978  
## f.campaign=f.campaign-(2,20] 1.625967e-13 7.376425  
## y=y.no 1.990017e-13 7.349465  
## job=job.technician 8.852239e-11 6.485356  
## month=month.aug 1.664867e-10 6.389447  
## f.job=f.job.Serv-Tech-BlueC 9.093071e-06 4.437687  
## job=job.admin. 9.903158e-05 3.892952  
## f.education=High School 2.594668e-03 3.012077  
## month=month.dec 2.647678e-03 -3.005932  
## marital=marital.divorced 2.422968e-03 -3.032799  
## job=job.management 2.107768e-03 -3.074612  
## f.job=f.job.Not-working 1.486893e-03 -3.177230  
## day\_of\_week=day\_of\_week.fri 1.476947e-03 -3.179175  
## job=job.student 1.591560e-05 -4.315619  
## f.education=Basic 2.828798e-06 -4.682874  
## f.campaign=f.campaign-[0,1] 4.359075e-09 -5.869957  
## marital=marital.married 1.197337e-09 -6.080600  
## month=month.mar 8.537841e-10 -6.134591  
## month=month.oct 9.782546e-11 -6.470274  
## month=month.sep 8.226884e-11 -6.496393  
## education=education.basic.4y 4.316243e-13 -7.245256  
## y=y.yes 1.990017e-13 -7.349465  
## f.job=f.job.Entrep-Retired-selfEmpl 3.453023e-15 -7.873338  
## f.age=f.age-(40,50] 9.680559e-19 -8.838738  
## f.previous=f.previous-(1,6] 1.264163e-20 -9.311183  
## poutcome=poutcome.success 1.947667e-23 -9.975678  
## job=job.retired 2.335349e-24 -10.184090  
## f.pdays=f.pdays-[0,22] 2.910853e-26 -10.602189  
## month=month.apr 1.985219e-48 -14.623632  
## month=month.nov 7.000612e-79 -18.804018  
## poutcome=poutcome.failure 4.131893e-84 -19.432090  
## f.previous=f.previous-(0.9,1] 3.088544e-87 -19.798146  
## f.season=f.season.Sep-Dec 2.442748e-104 -21.691974  
## f.age=f.age-(50,92] 8.185941e-132 -24.430126  
## month=month.may 5.261949e-141 -25.280294  
## f.season=f.season.Mar-May 3.483865e-218 -31.528962  
##   
## $`2`  
## Cla/Mod Mod/Cla Global  
## f.season=f.season.Mar-May 56.492248 69.1988131 41.3627255  
## f.previous=f.previous-(0.9,1] 86.170213 28.8427300 11.3026052  
## contact=contact.cellular 46.768654 88.9020772 64.1883768  
## month=month.apr 100.000000 19.0504451 6.4328657  
## poutcome=poutcome.failure 85.688073 27.7151335 10.9218437  
## f.pdays=f.pdays-[0,22] 100.000000 10.3857567 3.5070140  
## poutcome=poutcome.success 100.000000 9.2581602 3.1262525  
## y=y.yes 69.926199 22.4925816 10.8617234  
## f.previous=f.previous-(1,6] 100.000000 8.1305638 2.7454910  
## month=month.may 46.547619 46.4094955 33.6673347  
## month=month.sep 100.000000 4.1543027 1.4028056  
## default=default.no 37.664975 88.0712166 78.9579158  
## month=month.mar 100.000000 3.7388724 1.2625251  
## job=job.student 85.321101 5.5192878 2.1843687  
## month=month.oct 92.500000 4.3916914 1.6032064  
## marital=marital.single 42.523033 35.6083086 28.2765531  
## f.age=f.age-[18,30] 45.632184 23.5608309 17.4348697  
## f.campaign=f.campaign-[0,1] 39.321075 49.4955490 42.5050100  
## job=job.retired 55.405405 7.2997033 4.4488978  
## month=month.dec 88.888889 0.9495549 0.3607214  
## f.job=f.job.Not-working 45.029240 9.1394659 6.8537074  
## housing=housing.yes 36.370343 58.5163205 54.3286573  
## f.job=f.job.Entrep-Retired-selfEmpl 41.074523 14.0652819 11.5631263  
## f.duration=f.duration-(300,2.1e+03] 36.964415 30.2077151 27.5951904  
## job=job.housemaid 21.848739 1.5430267 2.3847695  
## f.duration=f.duration-[5,120] 29.672447 27.4183976 31.2024048  
## housing=housing.no 30.671347 41.4836795 45.6713427  
## job=job.technician 26.805386 12.9970326 16.3727455  
## f.job=f.job.Serv-Tech-BlueC 30.115512 43.3234421 48.5771543  
## marital=marital.married 30.206943 53.7091988 60.0400802  
## f.campaign=f.campaign-(2,20] 26.708075 25.5192878 32.2645291  
## f.age=f.age-(40,50] 24.980048 18.5756677 25.1102204  
## month=month.jun 16.071429 5.8753709 12.3446894  
## month=month.aug 17.670157 8.0118694 15.3106212  
## month=month.nov 13.450292 4.0949555 10.2805611  
## default=default.unknown 19.142857 11.9287834 21.0420842  
## y=y.no 29.361511 77.5074184 89.1382766  
## f.pdays=f.pdays-(22,23] 31.360332 89.6142433 96.4929860  
## month=month.jul 6.473988 3.3234421 17.3346693  
## contact=contact.telephone 10.464466 11.0979228 35.8116232  
## f.season=f.season.Jun-Aug 12.917595 17.2106825 44.9899800  
## f.previous=f.previous-[0,0.9] 24.761017 63.0267062 85.9519038  
## poutcome=poutcome.nonexistent 24.761017 63.0267062 85.9519038  
## p.value v.test  
## f.season=f.season.Mar-May 2.785537e-180 28.630731  
## f.previous=f.previous-(0.9,1] 1.710280e-166 27.500950  
## contact=contact.cellular 1.781833e-166 27.499462  
## month=month.apr 1.122106e-161 27.095184  
## poutcome=poutcome.failure 5.163924e-157 26.696533  
## f.pdays=f.pdays-[0,22] 5.769361e-86 19.650105  
## poutcome=poutcome.success 1.961538e-76 18.502763  
## y=y.yes 1.822374e-74 18.256935  
## f.previous=f.previous-(1,6] 5.675297e-67 17.289181  
## month=month.may 2.431397e-41 13.467443  
## month=month.sep 3.757836e-34 12.184537  
## default=default.no 1.458380e-31 11.688560  
## month=month.mar 9.045618e-31 11.532520  
## job=job.student 3.245927e-29 11.220184  
## month=month.oct 1.302973e-28 11.096607  
## marital=marital.single 4.418991e-16 8.126490  
## f.age=f.age-[18,30] 1.244669e-15 7.999953  
## f.campaign=f.campaign-[0,1] 1.127439e-12 7.113979  
## job=job.retired 1.511189e-11 6.746822  
## month=month.dec 2.065692e-06 4.746889  
## f.job=f.job.Not-working 7.828016e-06 4.469834  
## housing=housing.yes 2.185588e-05 4.245035  
## f.job=f.job.Entrep-Retired-selfEmpl 9.849594e-05 3.894267  
## f.duration=f.duration-(300,2.1e+03] 3.336199e-03 2.934933  
## job=job.housemaid 4.279022e-03 -2.856825  
## f.duration=f.duration-[5,120] 3.458368e-05 -4.140972  
## housing=housing.no 2.185588e-05 -4.245035  
## job=job.technician 3.051917e-06 -4.667295  
## f.job=f.job.Serv-Tech-BlueC 1.122693e-07 -5.305655  
## marital=marital.married 8.242978e-11 -6.496098  
## f.campaign=f.campaign-(2,20] 1.932831e-13 -7.353362  
## f.age=f.age-(40,50] 1.089573e-14 -7.728341  
## month=month.jun 1.288967e-25 -10.462150  
## month=month.aug 1.120542e-26 -10.691084  
## month=month.nov 4.313433e-28 -10.989071  
## default=default.unknown 1.458380e-31 -11.688560  
## y=y.no 1.822374e-74 -18.256935  
## f.pdays=f.pdays-(22,23] 5.769361e-86 -19.650105  
## month=month.jul 9.828418e-96 -20.760623  
## contact=contact.telephone 1.781833e-166 -27.499462  
## f.season=f.season.Jun-Aug 1.418547e-186 -29.131852  
## f.previous=f.previous-[0,0.9] 6.089086e-239 -33.008722  
## poutcome=poutcome.nonexistent 6.089086e-239 -33.008722  
##   
## $`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=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  
## marital=marital.divorced 12.692967 16.444444 11.683367  
## f.job=f.job.Admin-Managment 10.807529 39.555556 33.006012  
## housing=housing.yes 10.143858 61.111111 54.328657  
## day\_of\_week=day\_of\_week.mon 6.805293 16.000000 21.202405  
## month=month.mar 0.000000 0.000000 1.262525  
## housing=housing.no 7.678806 38.888889 45.671343  
## f.job=f.job.Serv-Tech-BlueC 7.714521 41.555556 48.577154  
## month=month.sep 0.000000 0.000000 1.402806  
## 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=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=Professional 5.526671e-04 3.453839  
## education=education.university.degree 5.526671e-04 3.453839  
## job=job.entrepreneur 9.628257e-04 3.301170  
## marital=marital.divorced 1.607675e-03 3.154511  
## f.job=f.job.Admin-Managment 2.234359e-03 3.057172  
## housing=housing.yes 2.381775e-03 3.037970  
## day\_of\_week=day\_of\_week.mon 3.715571e-03 -2.901351  
## month=month.mar 2.495672e-03 -3.023866  
## housing=housing.no 2.381775e-03 -3.037970  
## f.job=f.job.Serv-Tech-BlueC 1.758698e-03 -3.128218  
## month=month.sep 1.275948e-03 -3.221335  
## 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=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  
## contact=contact.telephone 57.0789032 70.05494505 35.8116232  
## f.previous=f.previous-[0,0.9] 33.9473071 100.00000000 85.9519038  
## poutcome=poutcome.nonexistent 33.9473071 100.00000000 85.9519038  
## month=month.may 46.3095238 53.43406593 33.6673347  
## f.age=f.age-(50,92] 54.4520548 32.76098901 17.5551102  
## f.age=f.age-(40,50] 46.9273743 40.38461538 25.1102204  
## default=default.unknown 48.6666667 35.09615385 21.0420842  
## marital=marital.married 35.9479306 73.96978022 60.0400802  
## f.season=f.season.Mar-May 37.6937984 53.43406593 41.3627255  
## y=y.no 31.4298561 96.01648352 89.1382766  
## f.pdays=f.pdays-(22,23] 30.2388370 100.00000000 96.4929860  
## month=month.aug 44.5026178 23.35164835 15.3106212  
## f.education=Basic 37.3825924 41.00274725 32.0040080  
## education=education.basic.4y 44.4660194 15.72802198 10.3206413  
## job=job.blue-collar 35.7266436 28.36538462 23.1663327  
## job=job.housemaid 49.5798319 4.05219780 2.3847695  
## housing=housing.no 31.9877139 50.06868132 45.6713427  
## month=month.jun 35.5519481 15.04120879 12.3446894  
## education=education.basic.9y 34.0740741 18.95604396 16.2324649  
## job=job.retired 39.1891892 5.97527473 4.4488978  
## f.campaign=f.campaign-(2,20] 32.1739130 35.57692308 32.2645291  
## day\_of\_week=day\_of\_week.fri 33.4378265 21.97802198 19.1783567  
## f.job=f.job.Serv-Tech-BlueC 31.2293729 51.99175824 48.5771543  
## month=month.dec 0.0000000 0.00000000 0.3607214  
## f.education=High School 26.2438491 32.96703297 36.6533066  
## education=education.high.school 25.0836120 20.60439560 23.9679359  
## day\_of\_week=day\_of\_week.thu 24.3781095 16.82692308 20.1402806  
## housing=housing.yes 26.8166728 49.93131868 54.3286573  
## f.campaign=f.campaign-[0,1] 25.9311645 37.77472527 42.5050100  
## f.job=f.job.Admin-Managment 25.0758956 28.36538462 33.0060120  
## f.education=Professional 24.2327366 26.03021978 31.3426854  
## education=education.university.degree 24.2327366 26.03021978 31.3426854  
## job=job.admin. 23.0709877 20.53571429 25.9719439  
## month=month.mar 0.0000000 0.00000000 1.2625251  
## month=month.oct 1.2500000 0.06868132 1.6032064  
## month=month.sep 0.0000000 0.00000000 1.4028056  
## job=job.student 3.6697248 0.27472527 2.1843687  
## f.previous=f.previous-(1,6] 0.0000000 0.00000000 2.7454910  
## poutcome=poutcome.success 0.0000000 0.00000000 3.1262525  
## f.pdays=f.pdays-[0,22] 0.0000000 0.00000000 3.5070140  
## y=y.yes 10.7011070 3.98351648 10.8617234  
## month=month.jul 13.6416185 8.10439560 17.3346693  
## f.age=f.age-(30,40] 19.6383727 26.85439560 39.8997996  
## month=month.apr 0.0000000 0.00000000 6.4328657  
## default=default.no 23.9847716 64.90384615 78.9579158  
## marital=marital.single 12.9695252 12.56868132 28.2765531  
## 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  
## f.season=f.season.Sep-Dec 0.1468429 0.06868132 13.6472946  
## f.age=f.age-[18,30] 0.0000000 0.00000000 17.4348697  
## contact=contact.cellular 13.6122385 29.94505495 64.1883768  
## p.value v.test  
## contact=contact.telephone 6.266590e-227 32.160662  
## f.previous=f.previous-[0,0.9] 1.066402e-115 22.863759  
## poutcome=poutcome.nonexistent 1.066402e-115 22.863759  
## month=month.may 8.291566e-78 18.672478  
## f.age=f.age-(50,92] 8.778509e-68 17.396445  
## f.age=f.age-(40,50] 1.545486e-54 15.551861  
## default=default.unknown 9.267066e-52 15.136758  
## marital=marital.married 2.486895e-39 13.121283  
## f.season=f.season.Mar-May 1.974279e-28 11.059394  
## y=y.no 1.553570e-27 10.872801  
## f.pdays=f.pdays-(22,23] 1.649166e-27 10.867353  
## month=month.aug 1.044971e-22 9.807528  
## f.education=Basic 5.410510e-18 8.644366  
## education=education.basic.4y 7.073062e-15 7.783168  
## job=job.blue-collar 3.535086e-08 5.512631  
## job=job.housemaid 2.365182e-06 4.719420  
## housing=housing.no 6.415740e-05 3.996974  
## month=month.jun 2.547657e-04 3.657421  
## education=education.basic.9y 9.338200e-04 3.309742  
## job=job.retired 1.083191e-03 3.267977  
## f.campaign=f.campaign-(2,20] 1.388516e-03 3.197028  
## day\_of\_week=day\_of\_week.fri 1.408127e-03 3.192980  
## f.job=f.job.Serv-Tech-BlueC 1.961552e-03 3.095993  
## month=month.dec 1.983747e-03 -3.092655  
## f.education=High School 4.983556e-04 -3.481639  
## education=education.high.school 3.128488e-04 -3.604422  
## day\_of\_week=day\_of\_week.thu 1.506415e-04 -3.790010  
## housing=housing.yes 6.415740e-05 -3.996974  
## f.campaign=f.campaign-[0,1] 1.356650e-05 -4.350762  
## f.job=f.job.Admin-Managment 6.606569e-06 -4.505985  
## f.education=Professional 1.589796e-07 -5.241844  
## education=education.university.degree 1.589796e-07 -5.241844  
## job=job.admin. 1.180067e-08 -5.702581  
## month=month.mar 3.088269e-10 -6.294282  
## month=month.oct 2.827025e-11 -6.655311  
## month=month.sep 2.654863e-11 -6.664545  
## job=job.student 6.098761e-12 -6.877323  
## f.previous=f.previous-(1,6] 1.352568e-21 -9.545650  
## poutcome=poutcome.success 1.517748e-24 -10.225923  
## f.pdays=f.pdays-[0,22] 1.649166e-27 -10.867353  
## y=y.yes 1.553570e-27 -10.872801  
## month=month.jul 1.071390e-31 -11.714724  
## f.age=f.age-(30,40] 1.269060e-34 -12.272727  
## month=month.apr 9.132574e-51 -14.985507  
## default=default.no 9.267066e-52 -15.136758  
## marital=marital.single 3.430613e-62 -16.642488  
## month=month.nov 9.922675e-83 -19.268254  
## poutcome=poutcome.failure 3.210256e-88 -19.911882  
## f.previous=f.previous-(0.9,1] 1.674179e-91 -20.287027  
## f.season=f.season.Sep-Dec 1.454029e-109 -22.238606  
## f.age=f.age-[18,30] 6.689248e-147 -25.810906  
## contact=contact.cellular 6.266590e-227 -32.160662  
##   
##   
## Link between the cluster variable and the quantitative variables  
## ================================================================  
## Eta2 P-value  
## emp.var.rate 0.95041557 0.000000e+00  
## cons.price.idx 0.56000681 0.000000e+00  
## euribor3m 0.98938151 0.000000e+00  
## nr.employed 0.85653677 0.000000e+00  
## previous 0.19611335 1.117307e-235  
## cons.conf.idx 0.18773544 1.836147e-224  
## age 0.17581151 1.066562e-208  
## pdays 0.06692220 1.479426e-74  
## campaign 0.02988623 1.381451e-32  
##   
## Description of each cluster by quantitative variables  
## =====================================================  
## $`1`  
## v.test Mean in category Overall mean sd in category  
## emp.var.rate 36.920749 1.373195 5.212425e-02 0.09278788  
## nr.employed 35.745040 5224.894639 5.165876e+03 10.41795235  
## euribor3m 34.385554 4.947089 3.584574e+00 0.03647388  
## cons.price.idx 28.863692 93.943032 9.356373e+01 0.33170185  
## campaign 9.437923 3.020014 2.515030e+00 2.95153122  
## pdays 8.143427 23.000000 2.241363e+01 0.00000000  
## previous -15.678746 0.000000 1.785571e-01 0.00000000  
## age -24.175634 34.442459 4.017756e+01 6.30378887  
## Overall sd p.value  
## emp.var.rate 1.5774788 2.147327e-298  
## nr.employed 72.7919889 7.900415e-280  
## euribor3m 1.7469207 4.145840e-259  
## cons.price.idx 0.5793439 3.411466e-183  
## campaign 2.3588988 3.802420e-21  
## pdays 3.1744936 3.842429e-16  
## previous 0.5020810 2.113880e-55  
## age 10.4585324 4.014848e-129  
##   
## $`2`  
## v.test Mean in category Overall mean sd in category  
## previous 30.530126 0.4824926 0.17855711 0.7495730  
## campaign -9.285801 2.0807122 2.51503006 1.6438038  
## pdays -18.272243 21.2635015 22.41362725 5.2769588  
## cons.conf.idx -22.300685 -42.5952522 -40.54192385 6.3102113  
## cons.price.idx -46.596079 93.0284742 93.56373427 0.5080721  
## nr.employed -64.842649 5072.2870030 5165.87569138 43.1835150  
## emp.var.rate -66.544064 -2.0292582 0.05212425 0.5812437  
## euribor3m -69.665001 1.1715205 3.58457355 0.2982714  
## Overall sd p.value  
## previous 0.5020810 1.038189e-204  
## campaign 2.3588988 1.604947e-20  
## pdays 3.1744936 1.376747e-74  
## cons.conf.idx 4.6436681 3.638838e-110  
## 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  
## 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  
## euribor3m 34.201843 4.902431 3.58457355 0.05192614  
## emp.var.rate 34.101559 1.238668 0.05212425 0.15364843  
## cons.price.idx 28.671578 93.930116 93.56373427 0.31689058  
## cons.conf.idx 28.180840 -37.655495 -40.54192385 2.46137972  
## nr.employed 26.394408 5208.253777 5165.87569138 18.50177893  
## age 25.335242 46.021978 40.17755511 7.76550632  
## pdays 8.374396 23.000000 22.41362725 0.00000000  
## campaign 3.301607 2.686813 2.51503006 2.49521339  
## previous -16.123435 0.000000 0.17855711 0.00000000  
## Overall sd p.value  
## euribor3m 1.7469207 2.270116e-256  
## emp.var.rate 1.5774788 6.993849e-255  
## cons.price.idx 0.5793439 8.630373e-181  
## cons.conf.idx 4.6436681 1.004266e-174  
## nr.employed 72.7919889 1.588465e-153  
## age 10.4585324 1.306974e-141  
## pdays 3.1744936 5.550785e-17  
## campaign 2.3588988 9.613268e-04  
## previous 0.5020810 1.746270e-58