# Load Required Packages: to be increased over the course

# Load Required Packages: to be increased over the course  
options(contrasts=c("contr.treatment","contr.treatment"))  
  
requiredPackages <- c("missMDA","chemometrics","mvoutlier","effects","FactoMineR","car", "factoextra","RColorBrewer","ggplot2","dplyr","ggmap","ggthemes","knitr","ggpubr", "corrplot")  
missingPackages <- requiredPackages[!(requiredPackages %in% installed.packages()[,"Package"])]  
  
if(length(missingPackages)) install.packages(missingPackages)  
lapply(requiredPackages, require, character.only = TRUE)

## Loading required package: missMDA

## Loading required package: chemometrics

## Loading required package: rpart

## Loading required package: mvoutlier

## Loading required package: sgeostat

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

## sROC 0.1-2 loaded

## Loading required package: effects

## Loading required package: carData

## Registered S3 methods overwritten by 'lme4':  
## method from  
## cooks.distance.influence.merMod car   
## influence.merMod car   
## dfbeta.influence.merMod car   
## dfbetas.influence.merMod car

## lattice theme set by effectsTheme()  
## See ?effectsTheme for details.

## Loading required package: FactoMineR

## Loading required package: car

## Loading required package: factoextra

## Loading required package: ggplot2

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

## Loading required package: RColorBrewer

## Loading required package: dplyr

##   
## Attaching package: 'dplyr'

## The following object is masked from 'package:car':  
##   
## recode

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## Loading required package: ggmap

## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.

## Please cite ggmap if you use it! See citation("ggmap") for details.

## Loading required package: ggthemes

## Loading required package: knitr

## Loading required package: ggpubr

## Loading required package: corrplot

## corrplot 0.84 loaded

## [[1]]  
## [1] TRUE  
##   
## [[2]]  
## [1] TRUE  
##   
## [[3]]  
## [1] TRUE  
##   
## [[4]]  
## [1] TRUE  
##   
## [[5]]  
## [1] TRUE  
##   
## [[6]]  
## [1] TRUE  
##   
## [[7]]  
## [1] TRUE  
##   
## [[8]]  
## [1] TRUE  
##   
## [[9]]  
## [1] TRUE  
##   
## [[10]]  
## [1] TRUE  
##   
## [[11]]  
## [1] TRUE  
##   
## [[12]]  
## [1] TRUE  
##   
## [[13]]  
## [1] TRUE  
##   
## [[14]]  
## [1] TRUE  
##   
## [[15]]  
## [1] TRUE

## Load Processed data

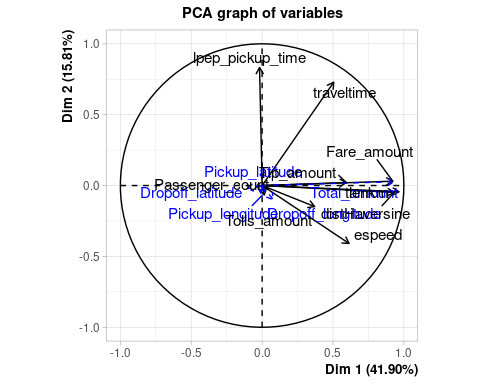
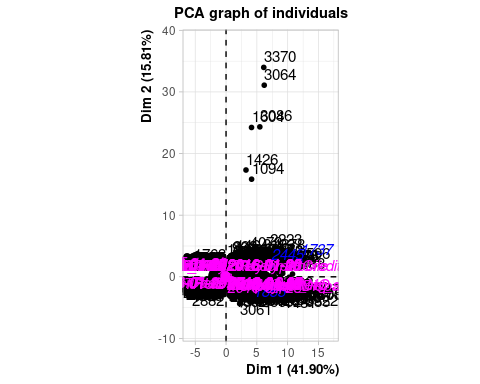
## null device   
## 1

## Modify Data

vars\_dis <- c("VendorID", "Payment\_type", "Store\_and\_fwd\_flag", "RateCodeID", "f.Extra", "f.MTA\_tax", "f.Improvement\_surcharge", "lpep\_pickup\_period", "Trip\_type", "lpep\_pickup\_date", "multiouts", "f.espeed", "f.tlenkm", "f.traveltime", "f.distHaversine", "AnyToll", "f.Fare\_amount", "f.Passenger\_count")  
  
vars\_con <- c( "Passenger\_count", "tlenkm", "Pickup\_longitude", "Pickup\_latitude", "Dropoff\_longitude", "Dropoff\_latitude","Fare\_amount", "espeed", "Tip\_amount", "Tolls\_amount", "lpep\_pickup\_time", "traveltime", "distHaversine")  
for( i in vars\_dis){  
 df[,i] <- as.factor(df[,i])  
}  
  
vars\_res <- c("AnyTip", "Total\_amount")  
  
d1 <- dim(df)[1]  
d2 <- dim(df)[2]  
df[,d2-1] <- as.factor(df[,d2-1])

###### Principal Component Analysis (PCA)

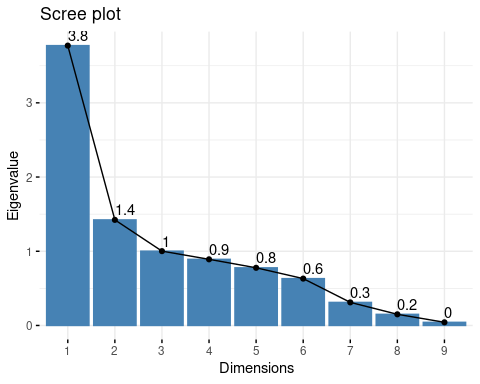
res.pca<-PCA(df[,c(1:10,19:(d2-2),d2)],quali.sup=c(1:10),quanti.sup=c(13:16,24), ncp=4, ind.sup = which(df$multiouts==TRUE))

 Para realizar PCA, hemos decidido decidido dividir el dataset en: -Variables numericas activas: Passenger\_count, tlenkm, Fare\_amount, espeed, Tip\_amount, Tolls\_amount, lpep\_pickup\_time, traveltime, distHaversine. -Variables numericas suplementarias:Pickup\_longitude, Pickup\_latitude, Dropoff\_longitude, Dropoff\_latitude, total\_amount. -Variables categoricas suplementarias: VendorID, Payment\_type, Store\_and\_fwd\_flag, RateCodeID, f.Extra, f.MTA\_tax, f.Improvement\_surcharge, lpep\_pickup\_period, Trip\_type, lpep\_pickup\_date. -Individuos activos: aquellos individuos sin outliers multivariante. -Individuos suplementarios: aquellos individuos con outliers multivariante. En este caso son los individuos cuyos indices en el dataset son 1737 1855 2445. ### Eigenvalues and dominant axes analysis.

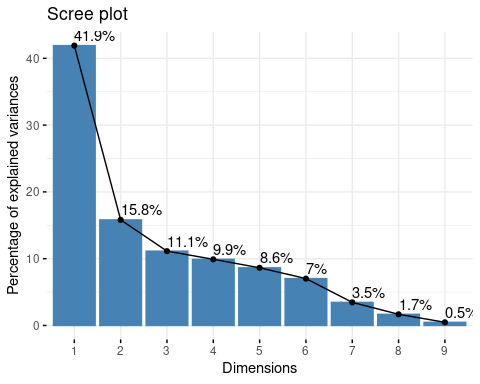
summary(res.pca,nb.dec=2,nbind=1,nbelements=1000,ncp=4)

##   
## Call:  
## PCA(X = df[, c(1:10, 19:(d2 - 2), d2)], ncp = 4, ind.sup = which(df$multiouts ==   
## TRUE), quanti.sup = c(13:16, 24), quali.sup = c(1:10))   
##   
##   
## Eigenvalues  
## Dim.1 Dim.2 Dim.3 Dim.4 Dim.5 Dim.6 Dim.7 Dim.8  
## Variance 3.77 1.42 1.00 0.89 0.78 0.63 0.31 0.15  
## % of var. 41.90 15.81 11.13 9.90 8.63 7.02 3.46 1.67  
## Cumulative % of var. 41.90 57.72 68.84 78.75 87.38 94.40 97.86 99.53  
## Dim.9  
## Variance 0.04  
## % of var. 0.47  
## Cumulative % of var. 100.00  
##   
## Individuals (the 1 first)  
## Dist Dim.1 ctr cos2 Dim.2 ctr cos2   
## 1 | 1.91 | -0.27 0.00 0.02 | -1.07 0.02 0.31 |  
## Dim.3 ctr cos2 Dim.4 ctr cos2   
## 1 -0.40 0.00 0.04 | -0.36 0.00 0.04 |  
##   
## Supplementary individuals (the 1 first)  
## Dist Dim.1 cos2 Dim.2 cos2 Dim.3 cos2   
## 1737 | 16.38 | 11.96 0.53 | 2.57 0.02 | -0.32 0.00 |  
## Dim.4 cos2   
## 1737 -2.58 0.02 |  
##   
## Variables  
## Dim.1 ctr cos2 Dim.2 ctr cos2 Dim.3  
## Passenger\_count | 0.02 0.01 0.00 | -0.07 0.31 0.00 | 0.99  
## tlenkm | 0.97 24.73 0.93 | -0.04 0.14 0.00 | -0.01  
## Fare\_amount | 0.92 22.52 0.85 | 0.03 0.07 0.00 | 0.00  
## espeed | 0.61 9.97 0.38 | -0.41 11.72 0.17 | -0.04  
## Tip\_amount | 0.59 9.35 0.35 | 0.02 0.03 0.00 | -0.04  
## Tolls\_amount | 0.37 3.64 0.14 | -0.15 1.55 0.02 | 0.12  
## lpep\_pickup\_time | -0.02 0.01 0.00 | 0.83 48.67 0.69 | 0.04  
## traveltime | 0.51 6.77 0.26 | 0.73 37.34 0.53 | 0.04  
## distHaversine | 0.93 22.99 0.87 | -0.05 0.17 0.00 | -0.02  
## ctr cos2 Dim.4 ctr cos2   
## Passenger\_count 97.69 0.98 | -0.12 1.62 0.01 |  
## tlenkm 0.02 0.00 | -0.11 1.32 0.01 |  
## Fare\_amount 0.00 0.00 | -0.12 1.64 0.01 |  
## espeed 0.18 0.00 | -0.07 0.63 0.01 |  
## Tip\_amount 0.18 0.00 | 0.08 0.65 0.01 |  
## Tolls\_amount 1.53 0.02 | 0.90 90.83 0.81 |  
## lpep\_pickup\_time 0.17 0.00 | 0.13 2.02 0.02 |  
## traveltime 0.20 0.00 | -0.03 0.13 0.00 |  
## distHaversine 0.03 0.00 | -0.10 1.17 0.01 |  
##   
## Supplementary continuous variables  
## Dim.1 cos2 Dim.2 cos2 Dim.3 cos2 Dim.4  
## Pickup\_longitude | 0.00 0.00 | -0.07 0.00 | 0.03 0.00 | -0.01  
## Pickup\_latitude | -0.05 0.00 | -0.03 0.00 | -0.03 0.00 | 0.06  
## Dropoff\_longitude | 0.08 0.01 | -0.10 0.01 | 0.02 0.00 | -0.02  
## Dropoff\_latitude | -0.11 0.01 | -0.02 0.00 | -0.02 0.00 | 0.07  
## Total\_amount | 0.93 0.87 | 0.03 0.00 | 0.00 0.00 | -0.03  
## cos2   
## Pickup\_longitude 0.00 |  
## Pickup\_latitude 0.00 |  
## Dropoff\_longitude 0.00 |  
## Dropoff\_latitude 0.00 |  
## Total\_amount 0.00 |  
##   
## Supplementary categories  
## Dist Dim.1 cos2 v.test Dim.2 cos2  
## f.Vendor-Mobile | 0.18 | 0.02 0.01 0.35 | 0.02 0.01  
## f.Vendor-VeriFone | 0.05 | -0.01 0.01 -0.35 | -0.01 0.01  
## f.PayType-Cash | 0.68 | -0.45 0.44 -16.47 | -0.03 0.00  
## f.PayType-Credit card | 0.70 | 0.46 0.43 16.49 | 0.03 0.00  
## f.PayType-Others | 0.80 | -0.05 0.00 -0.12 | 0.09 0.01  
## FALSE | 0.00 | 0.00 0.28 0.56 | 0.00 0.11  
## TRUE | 0.50 | -0.26 0.28 -0.56 | -0.16 0.11  
## Others | 2.11 | 1.88 0.79 11.66 | -0.56 0.07  
## Standard rate | 0.06 | -0.05 0.79 -11.66 | 0.02 0.07  
## f.Extra-0 | 0.21 | 0.07 0.10 2.24 | -0.04 0.04  
## f.Extra-0.5 | 0.28 | 0.03 0.01 0.91 | -0.17 0.35  
## f.Extra-1 | 0.68 | -0.27 0.16 -4.26 | 0.52 0.60  
## f.MTA\_tax\_NO | 1.12 | 0.77 0.47 4.60 | -0.43 0.15  
## f.MTA\_tax\_YES | 0.03 | -0.02 0.47 -4.60 | 0.01 0.15  
## f.Improvement\_surcharge\_NO | 1.00 | 0.62 0.39 3.80 | -0.40 0.16  
## f.Improvement\_surcharge\_YES | 0.03 | -0.02 0.39 -3.80 | 0.01 0.16  
## Period afternoon | 0.64 | -0.21 0.10 -3.57 | 0.48 0.56  
## Period morning | 0.62 | 0.05 0.01 0.67 | -0.36 0.34  
## Period night | 0.27 | 0.09 0.12 2.97 | -0.16 0.37  
## Period valley | 0.21 | -0.03 0.02 -0.70 | 0.10 0.23  
## f.TripType-Dispatch | 1.23 | 0.88 0.51 5.02 | -0.45 0.13  
## f.TripType-Street-Hail | 0.03 | -0.02 0.51 -5.02 | 0.01 0.13  
## 2016-01-01 | 0.49 | 0.08 0.03 0.66 | -0.15 0.09  
## 2016-01-02 | 0.31 | 0.14 0.21 1.03 | -0.05 0.03  
## 2016-01-03 | 0.46 | 0.13 0.08 0.81 | -0.37 0.64  
## 2016-01-04 | 0.23 | 0.10 0.18 0.63 | -0.11 0.21  
## 2016-01-05 | 0.33 | -0.28 0.71 -1.70 | 0.10 0.09  
## 2016-01-06 | 0.29 | 0.21 0.52 1.33 | -0.01 0.00  
## 2016-01-07 | 0.31 | 0.21 0.45 1.27 | -0.04 0.02  
## 2016-01-08 | 0.17 | 0.05 0.10 0.37 | 0.10 0.39  
## 2016-01-09 | 0.32 | -0.23 0.49 -1.62 | -0.21 0.43  
## 2016-01-10 | 0.49 | 0.14 0.08 0.92 | -0.42 0.73  
## 2016-01-11 | 0.22 | 0.01 0.00 0.06 | -0.06 0.09  
## 2016-01-12 | 0.28 | -0.17 0.37 -1.10 | -0.04 0.02  
## 2016-01-13 | 0.24 | 0.12 0.23 0.76 | 0.09 0.14  
## 2016-01-14 | 0.28 | -0.18 0.41 -1.21 | 0.08 0.08  
## 2016-01-15 | 0.38 | -0.35 0.84 -2.45 | 0.01 0.00  
## 2016-01-16 | 0.13 | -0.01 0.00 -0.07 | -0.07 0.32  
## 2016-01-17 | 0.41 | 0.19 0.22 1.31 | -0.34 0.67  
## 2016-01-18 | 0.29 | -0.09 0.11 -0.54 | 0.00 0.00  
## 2016-01-19 | 0.29 | -0.22 0.57 -1.35 | 0.08 0.08  
## 2016-01-20 | 0.20 | -0.18 0.83 -1.18 | -0.01 0.00  
## 2016-01-21 | 0.33 | 0.22 0.44 1.50 | 0.13 0.15  
## 2016-01-22 | 0.30 | -0.08 0.08 -0.61 | 0.27 0.81  
## 2016-01-23 | 1.36 | 0.13 0.01 0.44 | -0.77 0.32  
## 2016-01-24 | 0.52 | -0.21 0.16 -0.99 | 0.36 0.49  
## 2016-01-25 | 0.48 | 0.03 0.00 0.22 | 0.27 0.32  
## 2016-01-26 | 0.46 | -0.05 0.01 -0.31 | 0.33 0.50  
## 2016-01-27 | 0.40 | -0.09 0.05 -0.58 | 0.26 0.42  
## 2016-01-28 | 0.25 | -0.10 0.15 -0.59 | 0.14 0.33  
## 2016-01-29 | 0.29 | 0.06 0.04 0.42 | 0.18 0.39  
## 2016-01-30 | 0.25 | 0.00 0.00 0.02 | 0.23 0.86  
## 2016-01-31 | 0.50 | 0.31 0.40 2.25 | -0.33 0.44  
## v.test Dim.3 cos2 v.test Dim.4 cos2  
## f.Vendor-Mobile 0.58 | -0.16 0.82 -5.95 | -0.01 0.00  
## f.Vendor-VeriFone -0.58 | 0.04 0.82 5.95 | 0.00 0.00  
## f.PayType-Cash -1.55 | 0.04 0.00 2.61 | -0.05 0.01  
## f.PayType-Credit card 1.49 | -0.04 0.00 -2.50 | 0.05 0.01  
## f.PayType-Others 0.40 | -0.14 0.03 -0.74 | -0.16 0.04  
## FALSE 0.57 | 0.00 0.29 1.12 | 0.00 0.00  
## TRUE -0.57 | -0.27 0.29 -1.12 | -0.03 0.00  
## Others -5.66 | -0.11 0.00 -1.31 | 0.00 0.00  
## Standard rate 5.66 | 0.00 0.00 1.31 | 0.00 0.00  
## f.Extra-0 -2.39 | -0.05 0.07 -3.57 | 0.01 0.00  
## f.Extra-0.5 -7.76 | 0.03 0.01 1.72 | -0.06 0.05  
## f.Extra-1 13.52 | 0.08 0.02 2.57 | 0.13 0.04  
## f.MTA\_tax\_NO -4.19 | -0.17 0.02 -1.94 | -0.15 0.02  
## f.MTA\_tax\_YES 4.19 | 0.00 0.02 1.94 | 0.00 0.02  
## f.Improvement\_surcharge\_NO -3.93 | -0.15 0.02 -1.76 | -0.15 0.02  
## f.Improvement\_surcharge\_YES 3.93 | 0.00 0.02 1.76 | 0.00 0.02  
## Period afternoon 13.55 | 0.06 0.01 1.91 | 0.08 0.01  
## Period morning -7.96 | -0.11 0.03 -2.95 | -0.04 0.01  
## Period night -8.46 | 0.02 0.01 1.46 | -0.06 0.05  
## Period valley 3.52 | -0.03 0.02 -1.14 | 0.06 0.09  
## f.TripType-Dispatch -4.20 | -0.16 0.02 -1.76 | -0.19 0.02  
## f.TripType-Street-Hail 4.20 | 0.00 0.02 1.76 | 0.00 0.02  
## 2016-01-01 -1.90 | 0.07 0.02 1.10 | -0.16 0.11  
## 2016-01-02 -0.62 | 0.04 0.02 0.62 | -0.04 0.01  
## 2016-01-03 -3.83 | 0.06 0.02 0.71 | -0.12 0.06  
## 2016-01-04 -1.10 | -0.03 0.01 -0.35 | 0.16 0.49  
## 2016-01-05 0.96 | -0.08 0.06 -0.99 | 0.00 0.00  
## 2016-01-06 -0.09 | -0.08 0.07 -0.95 | 0.05 0.03  
## 2016-01-07 -0.41 | -0.13 0.19 -1.59 | 0.14 0.21  
## 2016-01-08 1.18 | -0.03 0.02 -0.34 | 0.02 0.02  
## 2016-01-09 -2.47 | 0.00 0.00 0.04 | 0.02 0.00  
## 2016-01-10 -4.63 | 0.07 0.02 0.86 | -0.08 0.03  
## 2016-01-11 -0.70 | -0.08 0.14 -1.06 | -0.12 0.33  
## 2016-01-12 -0.43 | 0.11 0.15 1.37 | 0.11 0.15  
## 2016-01-13 0.97 | 0.14 0.36 1.83 | 0.01 0.00  
## 2016-01-14 0.85 | 0.04 0.02 0.48 | 0.15 0.30  
## 2016-01-15 0.09 | 0.03 0.01 0.42 | 0.11 0.09  
## 2016-01-16 -0.93 | -0.04 0.12 -0.67 | -0.06 0.21  
## 2016-01-17 -3.71 | 0.01 0.00 0.11 | -0.05 0.02  
## 2016-01-18 0.03 | -0.09 0.09 -0.96 | -0.07 0.07  
## 2016-01-19 0.81 | -0.11 0.15 -1.33 | 0.06 0.04  
## 2016-01-20 -0.10 | -0.02 0.01 -0.23 | 0.06 0.10  
## 2016-01-21 1.44 | 0.04 0.01 0.50 | 0.05 0.02  
## 2016-01-22 3.27 | -0.01 0.00 -0.21 | 0.06 0.03  
## 2016-01-23 -4.15 | 0.04 0.00 0.25 | -0.33 0.06  
## 2016-01-24 2.80 | 0.08 0.03 0.78 | -0.03 0.00  
## 2016-01-25 2.85 | -0.18 0.15 -2.30 | -0.10 0.04  
## 2016-01-26 3.31 | 0.06 0.02 0.73 | 0.00 0.00  
## 2016-01-27 2.83 | -0.01 0.00 -0.07 | -0.08 0.04  
## 2016-01-28 1.43 | 0.03 0.01 0.31 | -0.04 0.02  
## 2016-01-29 2.04 | -0.01 0.00 -0.18 | 0.09 0.10  
## 2016-01-30 3.09 | 0.04 0.03 0.63 | -0.04 0.03  
## 2016-01-31 -3.87 | 0.01 0.00 0.19 | 0.05 0.01  
## v.test   
## f.Vendor-Mobile -0.29 |  
## f.Vendor-VeriFone 0.29 |  
## f.PayType-Cash -3.87 |  
## f.PayType-Credit card 4.01 |  
## f.PayType-Others -0.92 |  
## FALSE 0.15 |  
## TRUE -0.15 |  
## Others -0.05 |  
## Standard rate 0.05 |  
## f.Extra-0 0.60 |  
## f.Extra-0.5 -3.76 |  
## f.Extra-1 4.16 |  
## f.MTA\_tax\_NO -1.87 |  
## f.MTA\_tax\_YES 1.87 |  
## f.Improvement\_surcharge\_NO -1.88 |  
## f.Improvement\_surcharge\_YES 1.88 |  
## Period afternoon 2.76 |  
## Period morning -1.22 |  
## Period night -3.81 |  
## Period valley 2.78 |  
## f.TripType-Dispatch -2.25 |  
## f.TripType-Street-Hail 2.25 |  
## 2016-01-01 -2.61 |  
## 2016-01-02 -0.55 |  
## 2016-01-03 -1.52 |  
## 2016-01-04 2.13 |  
## 2016-01-05 0.03 |  
## 2016-01-06 0.68 |  
## 2016-01-07 1.80 |  
## 2016-01-08 0.31 |  
## 2016-01-09 0.24 |  
## 2016-01-10 -1.18 |  
## 2016-01-11 -1.73 |  
## 2016-01-12 1.42 |  
## 2016-01-13 0.09 |  
## 2016-01-14 2.12 |  
## 2016-01-15 1.63 |  
## 2016-01-16 -0.96 |  
## 2016-01-17 -0.74 |  
## 2016-01-18 -0.89 |  
## 2016-01-19 0.75 |  
## 2016-01-20 0.86 |  
## 2016-01-21 0.72 |  
## 2016-01-22 0.86 |  
## 2016-01-23 -2.25 |  
## 2016-01-24 -0.31 |  
## 2016-01-25 -1.32 |  
## 2016-01-26 0.06 |  
## 2016-01-27 -1.06 |  
## 2016-01-28 -0.46 |  
## 2016-01-29 1.32 |  
## 2016-01-30 -0.68 |  
## 2016-01-31 0.68 |

fviz\_eig(res.pca, choice = "eigenvalue", addlabels = TRUE)

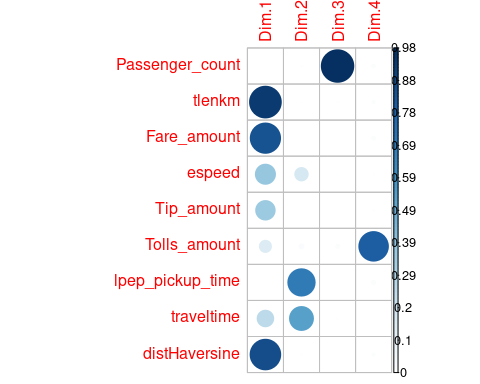


fviz\_eig(res.pca, addlabels = TRUE)

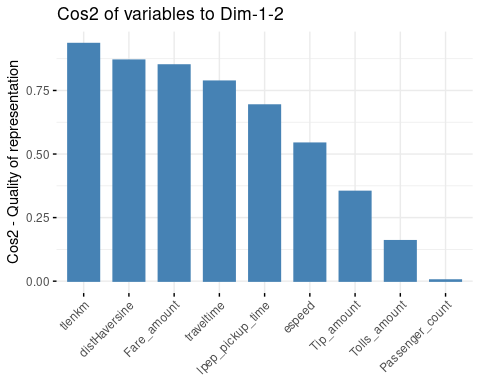
 En esta imagen podemos ver los valores propios. Como podemos ver hasta la tercera dimensión tenemos un valor propio igual a 1. Según el criterio de Kaiser deberíamos eliminar todas las componentes con valor propio por debajo de 1, lo que significa que deberíamos coger hasta la tercera dimension. Segun la regla de Elbow, debemos coger hasta que no haya un descenso significativo, lo que significa que también se debería coger hasta la terecera dimensio. A pesar de todo, hemos decidido incluir hasta la cuarta dimensión, ya que nos facilita el estudio. Como podemos ver a partir del summary, hasta la cuarta dimensión encontramos una varianza acumulada del 78.75%. También podemos admirar como la primera dimensión contribuye mucho en el PCA, explicando un 41.9% de la varianza.

## Quality of representation

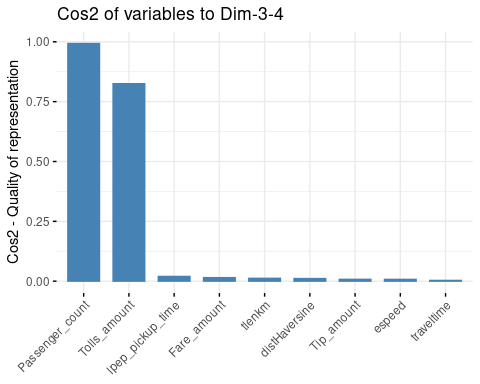
corrplot(res.pca$var$cos2, is.corr=FALSE)



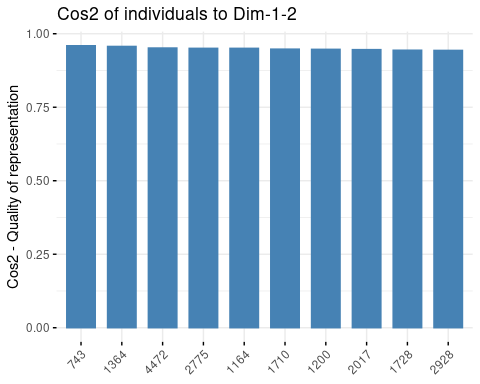
fviz\_cos2(res.pca, choice = "var", axes = 1:2, top = 10)



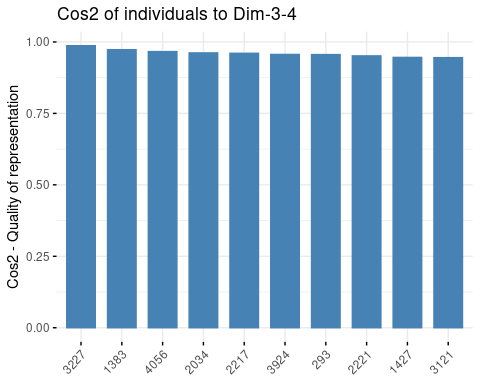
fviz\_cos2(res.pca, choice = "var", axes = 3:4, top = 10)



fviz\_cos2(res.pca, choice = "ind", axes = 1:2, top = 10)

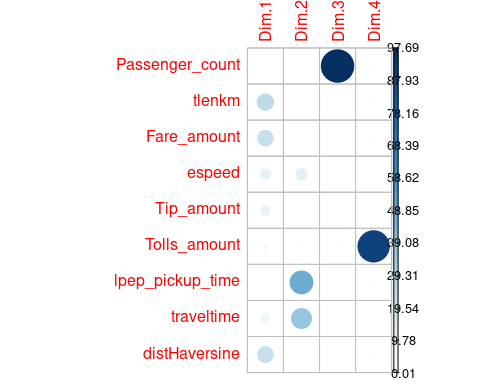


fviz\_cos2(res.pca, choice = "ind", axes = 3:4, top = 10)

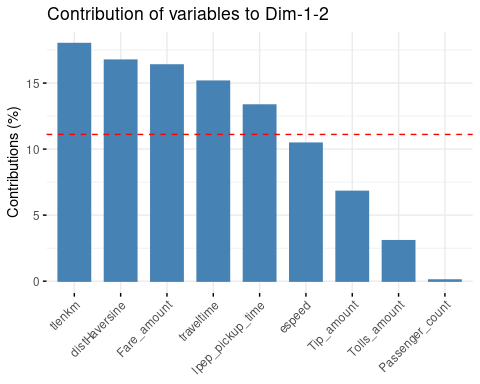
 A partir de la suma del coseno al cuadrado de la primera dimensión más el coseno al cuadrado de la segunda dimensión podemos obtener cualidad de las variables en el primer plano factorial. Como podemos ver tlenkm y distHaversine son las dos variables que mejor se representan el primer plano factorial.

## Contribution

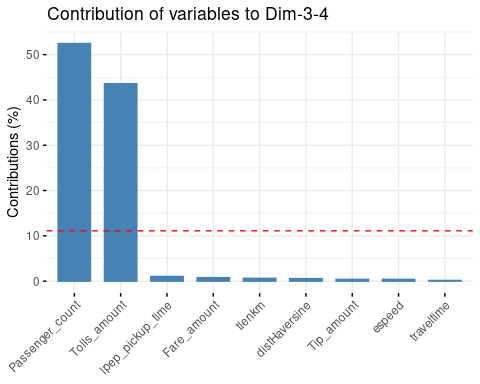
corrplot(res.pca$var$contrib, is.corr=FALSE)



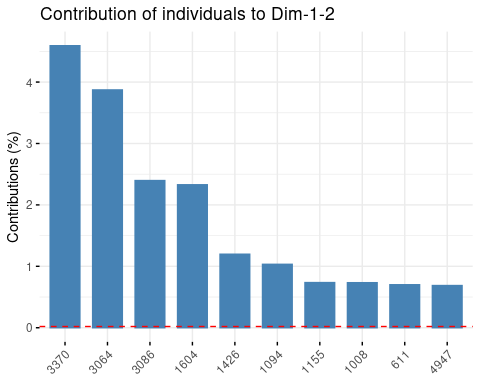
fviz\_contrib(res.pca, choice = "var", axes = 1:2, top = 10)



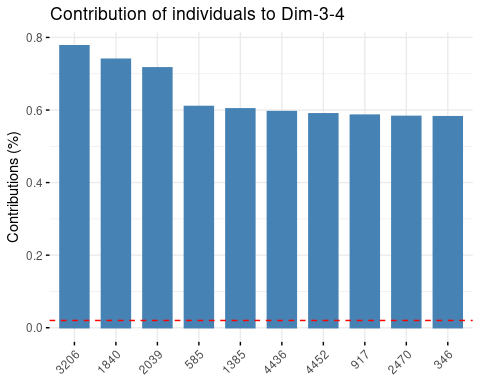
fviz\_contrib(res.pca, choice = "var", axes = 3:4, top = 10)



fviz\_contrib(res.pca, choice = "ind", axes = 1:2, top = 10)



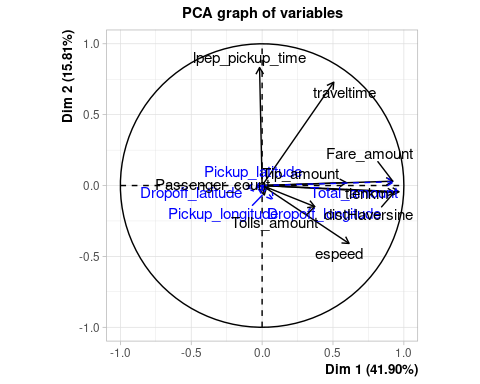
fviz\_contrib(res.pca, choice = "ind", axes = 3:4, top = 10)

 En este plot encontramos las 10 variables que contribuyen mas en el primer plano factorial. Como podemos ver la variable tlenkm es la variable que contribuye más junto con la variable distHaversine. Si vemos el segundo plano factorial, vemos que el numero total de pasajeros y el numero total de pagos por peajes influyen bastante en el segundo plano factorial. En este plot encontramos los 10 individuos que contribuyen más en el primer plano factorial. Como podemos ver el individuo 3370 es el individuo que contribuye más, junto con el individuo 3064 al primer plano factorial. Si vemos el segundo plano factorial, vemos que el individuo 3206 y el individuo 1840, son los individuos que contribuyen más el segundo plano factorial.

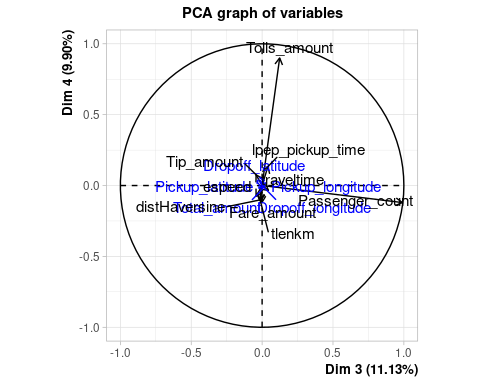
##### Interpreting the axes

### Variables numericas

plot.PCA(res.pca,choix=c("var"), axes=c(1,2))

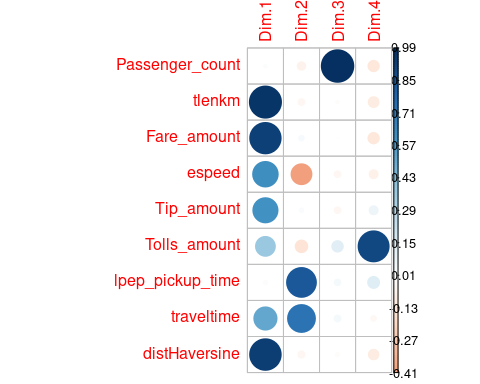


plot.PCA(res.pca,choix=c("var"), axes=c(3,4))

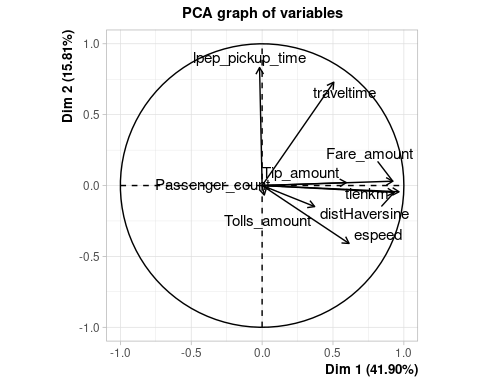


## Variables numericas activas

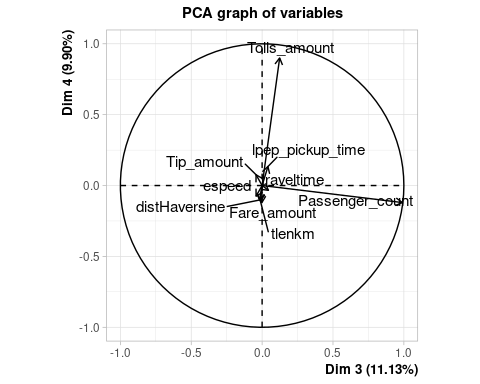
corrplot(res.pca$var$coord, is.corr=FALSE)



plot.PCA(res.pca,choix=c("var"),invisible=c("quanti.sup"), axes=c(1,2))

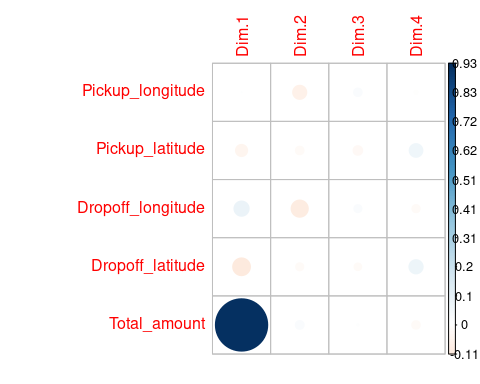


plot.PCA(res.pca,choix=c("var"),invisible=c("quanti.sup"), axes=c(3,4))

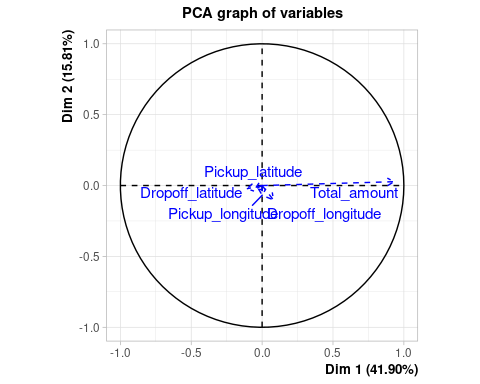


## Variables numericas suplementarias

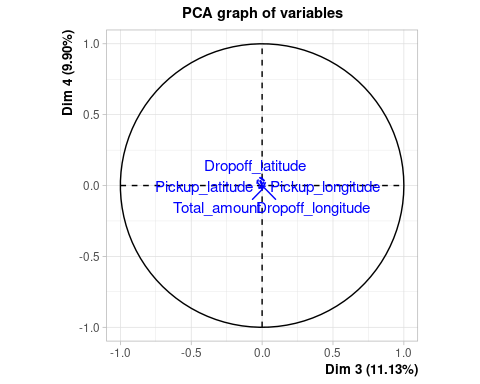
corrplot(res.pca$quanti.sup$coord, is.corr=FALSE)



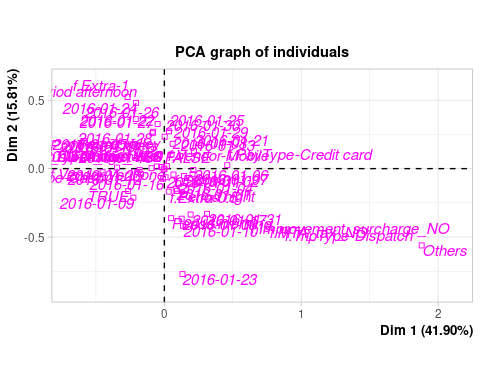
plot.PCA(res.pca,choix=c("var"),invisible=c("var"), axes=c(1,2))



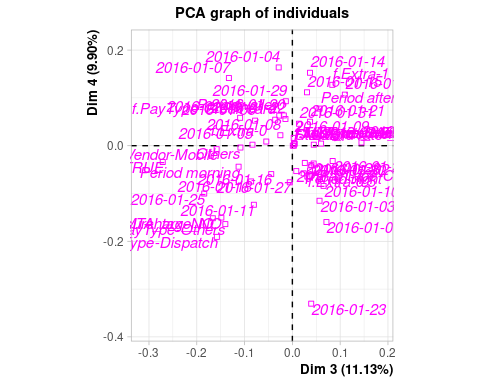
plot.PCA(res.pca,choix=c("var"),invisible=c("var"), axes=c(3,4))

 ### Variables categoricas suplementarias

plot.PCA(res.pca,choix=c("ind"), invisible = c("ind", "ind.sup"), axes=c(1,2))

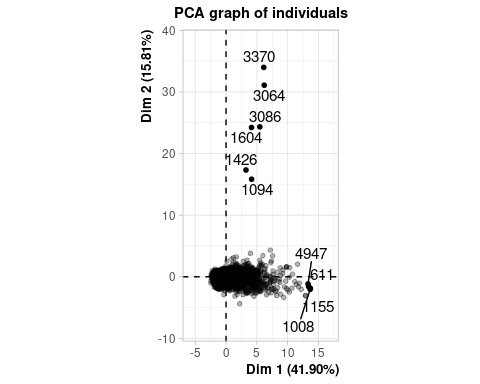


plot.PCA(res.pca,choix=c("ind"), invisible = c("ind", "ind.sup"), axes=c(3,4))

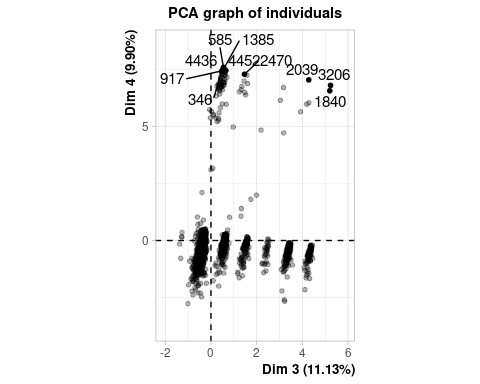


### Individuos

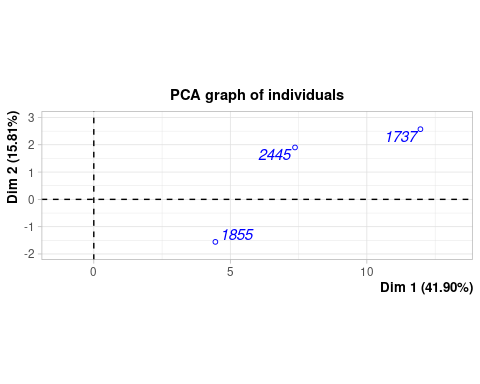
plot(res.pca,choix=c("ind"),invisible = c("ind.sup", "var", "quali"),select="contrib 10", axes=c(1,2))



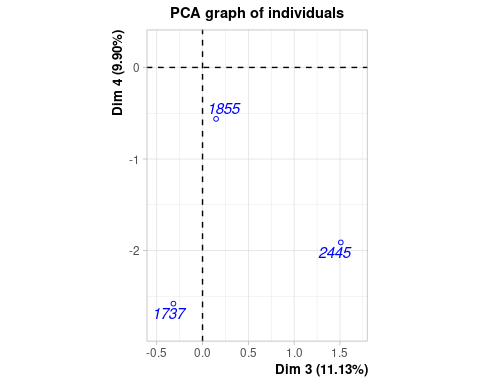
plot(res.pca,choix=c("ind"),invisible = c("ind.sup", "var", "quali"),select="contrib 10", axes=c(3,4))

 ### Individuos suplementarios

plot(res.pca,choix=c("ind"),invisible = c("ind", "var", "quali"), axes=c(1,2))

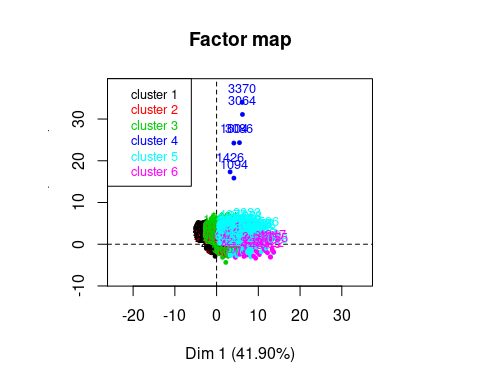
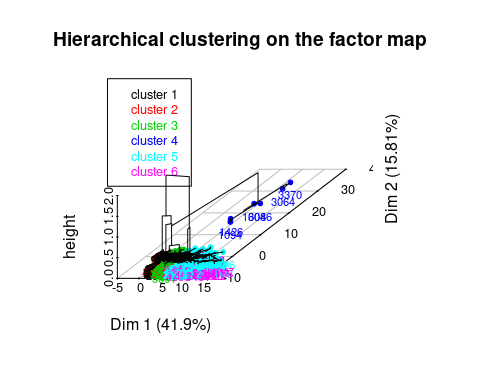
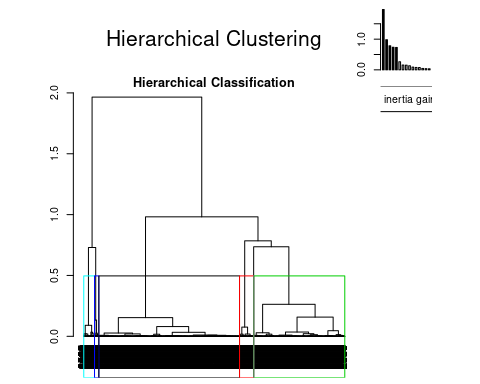


plot(res.pca,choix=c("ind"),invisible = c("ind", "var", "quali"), axes=c(3,4))



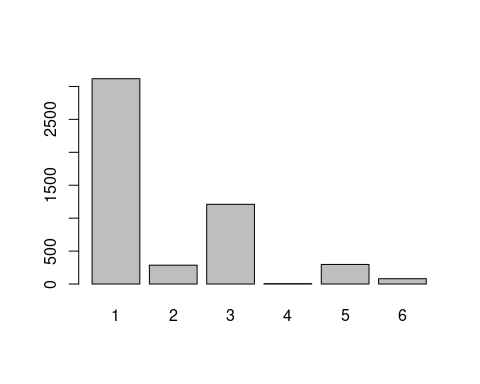
HCPC

res.hcpc<-HCPC(res.pca,nb.clust = -1,order=TRUE)

 Viendo la inertia gain (pérdida importante de ir entre n clusters a n+1 clusters) y aplicando Kaiser Rule podemos ver que el número de clusters óptimo es 6.

###### Interpretar los resultados de la clasificación

barplot(table(res.hcpc$data.clust$clust))



summary(res.hcpc$data.clust$clust)

## 1 2 3 4 5 6   
## 3117 286 1210 6 297 81

### desc.var ###  
### A. The description of the clusters by the variables ###  
names(res.hcpc$desc.var)

## [1] "test.chi2" "category" "quanti.var" "quanti" "call"

### desc.var$test.chi2 ###  
### A.1. The categorical variables which characterizes the clusters ###  
res.hcpc$desc.var$test.chi2

## p.value df  
## Payment\_type 4.677899e-40 10  
## RateCodeID 3.444603e-14 5  
## VendorID 9.148768e-11 5  
## f.Extra 8.851884e-08 10  
## lpep\_pickup\_period 6.244114e-06 15  
## Trip\_type 4.028542e-05 5  
## f.MTA\_tax 1.027535e-04 5  
## f.Improvement\_surcharge 8.409311e-04 5

### desc.var$category ###  
### A.2. The description of each cluster by the categories ##  
res.hcpc$desc.var$category

## $`1`  
## Cla/Mod Mod/Cla Global p.value  
## Payment\_type=f.PayType-Cash 70.63049 56.785371 50.150090 9.316896e-34  
## RateCodeID=Standard rate 62.95305 98.075072 97.178307 1.435285e-06  
## f.Extra=f.Extra-1 68.17043 17.452679 15.969582 2.034130e-04  
## lpep\_pickup\_period=Period afternoon 67.31813 19.890921 18.431059 5.636393e-04  
## Trip\_type=f.TripType-Street-Hail 62.65381 98.010908 97.578547 1.178483e-02  
## lpep\_pickup\_date=2016-01-05 71.53285 3.144049 2.741645 2.328595e-02  
## lpep\_pickup\_date=2016-01-15 69.39891 4.074431 3.662197 4.416745e-02  
## lpep\_pickup\_date=2016-01-19 70.13889 3.240295 2.881729 4.927721e-02  
## lpep\_pickup\_date=2016-01-31 55.31915 3.336542 3.762257 4.386629e-02  
## lpep\_pickup\_date=2016-01-17 54.16667 2.919474 3.362017 2.726666e-02  
## Trip\_type=f.TripType-Dispatch 51.23967 1.989092 2.421453 1.178483e-02  
## f.Extra=f.Extra-0.5 59.43945 36.060314 37.842706 8.428487e-04  
## RateCodeID=Others 42.55319 1.924928 2.821693 1.435285e-06  
## lpep\_pickup\_period=Period night 58.55897 40.936798 43.606164 9.938570e-07  
## Payment\_type=f.PayType-Credit card 53.95859 42.637151 49.289574 6.570501e-34  
## v.test  
## Payment\_type=f.PayType-Cash 12.110286  
## RateCodeID=Standard rate 4.820041  
## f.Extra=f.Extra-1 3.714740  
## lpep\_pickup\_period=Period afternoon 3.448534  
## Trip\_type=f.TripType-Street-Hail 2.518523  
## lpep\_pickup\_date=2016-01-05 2.268710  
## lpep\_pickup\_date=2016-01-15 2.012498  
## lpep\_pickup\_date=2016-01-19 1.966185  
## lpep\_pickup\_date=2016-01-31 -2.015366  
## lpep\_pickup\_date=2016-01-17 -2.207679  
## Trip\_type=f.TripType-Dispatch -2.518523  
## f.Extra=f.Extra-0.5 -3.338326  
## RateCodeID=Others -4.820041  
## lpep\_pickup\_period=Period night -4.892851  
## Payment\_type=f.PayType-Credit card -12.138898  
##   
## $`2`  
## Cla/Mod Mod/Cla  
## VendorID=f.Vendor-VeriFone 6.957638 94.7552448  
## f.Extra=f.Extra-0.5 6.980434 46.1538462  
## lpep\_pickup\_period=Period night 6.700321 51.0489510  
## f.Improvement\_surcharge=f.Improvement\_surcharge\_YES 5.842419 99.3006993  
## Payment\_type=f.PayType-Cash 6.504389 56.9930070  
## f.MTA\_tax=f.MTA\_tax\_YES 5.837616 99.3006993  
## f.Extra=f.Extra-1 7.393484 20.6293706  
## Trip\_type=f.TripType-Street-Hail 5.824446 99.3006993  
## RateCodeID=Standard rate 5.827842 98.9510490  
## lpep\_pickup\_date=2016-01-07 2.142857 1.0489510  
## RateCodeID=Others 2.127660 1.0489510  
## lpep\_pickup\_period=Period morning 3.960396 8.3916084  
## Trip\_type=f.TripType-Dispatch 1.652893 0.6993007  
## lpep\_pickup\_date=2016-01-25 1.986755 1.0489510  
## Payment\_type=f.PayType-Credit card 4.953309 42.6573427  
## f.MTA\_tax=f.MTA\_tax\_NO 1.515152 0.6993007  
## lpep\_pickup\_period=Period valley 4.415182 19.9300699  
## f.Improvement\_surcharge=f.Improvement\_surcharge\_NO 1.470588 0.6993007  
## f.Extra=f.Extra-0 4.116118 33.2167832  
## VendorID=f.Vendor-Mobile 1.361162 5.2447552  
## Global p.value  
## VendorID=f.Vendor-VeriFone 77.946768 8.868367e-16  
## f.Extra=f.Extra-0.5 37.842706 3.172789e-03  
## lpep\_pickup\_period=Period night 43.606164 9.326632e-03  
## f.Improvement\_surcharge=f.Improvement\_surcharge\_YES 97.278367 1.613422e-02  
## Payment\_type=f.PayType-Cash 50.150090 1.719831e-02  
## f.MTA\_tax=f.MTA\_tax\_YES 97.358415 1.956425e-02  
## f.Extra=f.Extra-1 15.969582 3.139249e-02  
## Trip\_type=f.TripType-Street-Hail 97.578547 3.300006e-02  
## RateCodeID=Standard rate 97.178307 4.514856e-02  
## lpep\_pickup\_date=2016-01-07 2.801681 4.712175e-02  
## RateCodeID=Others 2.821693 4.514856e-02  
## lpep\_pickup\_period=Period morning 12.127276 3.973965e-02  
## Trip\_type=f.TripType-Dispatch 2.421453 3.300006e-02  
## lpep\_pickup\_date=2016-01-25 3.021813 2.926200e-02  
## Payment\_type=f.PayType-Credit card 49.289574 2.088991e-02  
## f.MTA\_tax=f.MTA\_tax\_NO 2.641585 1.956425e-02  
## lpep\_pickup\_period=Period valley 25.835501 1.661855e-02  
## f.Improvement\_surcharge=f.Improvement\_surcharge\_NO 2.721633 1.613422e-02  
## f.Extra=f.Extra-0 46.187713 4.772479e-06  
## VendorID=f.Vendor-Mobile 22.053232 8.868367e-16  
## v.test  
## VendorID=f.Vendor-VeriFone 8.041585  
## f.Extra=f.Extra-0.5 2.950482  
## lpep\_pickup\_period=Period night 2.599843  
## f.Improvement\_surcharge=f.Improvement\_surcharge\_YES 2.405865  
## Payment\_type=f.PayType-Cash 2.382441  
## f.MTA\_tax=f.MTA\_tax\_YES 2.334602  
## f.Extra=f.Extra-1 2.152062  
## Trip\_type=f.TripType-Street-Hail 2.132083  
## RateCodeID=Standard rate 2.003268  
## lpep\_pickup\_date=2016-01-07 -1.985204  
## RateCodeID=Others -2.003268  
## lpep\_pickup\_period=Period morning -2.056445  
## Trip\_type=f.TripType-Dispatch -2.132083  
## lpep\_pickup\_date=2016-01-25 -2.179939  
## Payment\_type=f.PayType-Credit card -2.309968  
## f.MTA\_tax=f.MTA\_tax\_NO -2.334602  
## lpep\_pickup\_period=Period valley -2.395040  
## f.Improvement\_surcharge=f.Improvement\_surcharge\_NO -2.405865  
## f.Extra=f.Extra-0 -4.574549  
## VendorID=f.Vendor-Mobile -8.041585  
##   
## $`3`  
## Cla/Mod Mod/Cla Global p.value  
## Payment\_type=f.PayType-Credit card 29.80106 60.661157 49.289574 8.417416e-20  
## lpep\_pickup\_period=Period night 26.61771 47.933884 43.606164 5.038294e-04  
## f.Extra=f.Extra-0.5 26.22951 40.991736 37.842706 9.723472e-03  
## lpep\_pickup\_date=2016-01-01 29.95595 5.619835 4.542726 4.283924e-02  
## lpep\_pickup\_date=2016-01-19 17.36111 2.066116 2.881729 4.654076e-02  
## lpep\_pickup\_date=2016-01-12 17.10526 2.148760 3.041825 3.330044e-02  
## lpep\_pickup\_date=2016-01-14 16.86747 2.314050 3.321993 2.098555e-02  
## lpep\_pickup\_period=Period afternoon 20.52117 15.619835 18.431059 3.370499e-03  
## f.Extra=f.Extra-1 18.67168 12.314050 15.969582 4.641295e-05  
## Payment\_type=f.PayType-Cash 18.75499 38.842975 50.150090 1.303366e-19  
## v.test  
## Payment\_type=f.PayType-Credit card 9.107665  
## lpep\_pickup\_period=Period night 3.478712  
## f.Extra=f.Extra-0.5 2.585511  
## lpep\_pickup\_date=2016-01-01 2.025274  
## lpep\_pickup\_date=2016-01-19 -1.990456  
## lpep\_pickup\_date=2016-01-12 -2.128442  
## lpep\_pickup\_date=2016-01-14 -2.308244  
## lpep\_pickup\_period=Period afternoon -2.931757  
## f.Extra=f.Extra-1 -4.072992  
## Payment\_type=f.PayType-Cash -9.060095  
##   
## $`4`  
## Cla/Mod Mod/Cla Global p.value  
## lpep\_pickup\_date=2016-01-01 1.3215859 50.00000 4.542726 0.001728979  
## lpep\_pickup\_period=Period night 0.2753557 100.00000 43.606164 0.006848560  
## lpep\_pickup\_date=2016-01-30 0.8403361 33.33333 4.762858 0.031758791  
## v.test  
## lpep\_pickup\_date=2016-01-01 3.133223  
## lpep\_pickup\_period=Period night 2.704120  
## lpep\_pickup\_date=2016-01-30 2.147433  
##   
## $`5`  
## Cla/Mod Mod/Cla  
## Payment\_type=f.PayType-Credit card 8.810394 73.063973  
## RateCodeID=Others 17.021277 8.080808  
## Trip\_type=f.TripType-Dispatch 15.702479 6.397306  
## f.MTA\_tax=f.MTA\_tax\_NO 15.151515 6.734007  
## f.Improvement\_surcharge=f.Improvement\_surcharge\_NO 13.970588 6.397306  
## lpep\_pickup\_date=2016-01-15 2.732240 1.683502  
## f.Extra=f.Extra-1 4.260652 11.447811  
## f.Improvement\_surcharge=f.Improvement\_surcharge\_YES 5.718988 93.602694  
## f.MTA\_tax=f.MTA\_tax\_YES 5.693731 93.265993  
## Trip\_type=f.TripType-Street-Hail 5.701395 93.602694  
## RateCodeID=Standard rate 5.621911 91.919192  
## Payment\_type=f.PayType-Cash 3.072626 25.925926  
## Global p.value  
## Payment\_type=f.PayType-Credit card 49.289574 9.875344e-18  
## RateCodeID=Others 2.821693 2.590335e-06  
## Trip\_type=f.TripType-Dispatch 2.421453 1.008162e-04  
## f.MTA\_tax=f.MTA\_tax\_NO 2.641585 1.108405e-04  
## f.Improvement\_surcharge=f.Improvement\_surcharge\_NO 2.721633 5.039002e-04  
## lpep\_pickup\_date=2016-01-15 3.662197 4.678203e-02  
## f.Extra=f.Extra-1 15.969582 2.396618e-02  
## f.Improvement\_surcharge=f.Improvement\_surcharge\_YES 97.278367 5.039002e-04  
## f.MTA\_tax=f.MTA\_tax\_YES 97.358415 1.108405e-04  
## Trip\_type=f.TripType-Street-Hail 97.578547 1.008162e-04  
## RateCodeID=Standard rate 97.178307 2.590335e-06  
## Payment\_type=f.PayType-Cash 50.150090 1.962636e-18  
## v.test  
## Payment\_type=f.PayType-Credit card 8.575388  
## RateCodeID=Others 4.700887  
## Trip\_type=f.TripType-Dispatch 3.888619  
## f.MTA\_tax=f.MTA\_tax\_NO 3.865547  
## f.Improvement\_surcharge=f.Improvement\_surcharge\_NO 3.478675  
## lpep\_pickup\_date=2016-01-15 -1.988268  
## f.Extra=f.Extra-1 -2.257671  
## f.Improvement\_surcharge=f.Improvement\_surcharge\_YES -3.478675  
## f.MTA\_tax=f.MTA\_tax\_YES -3.865547  
## Trip\_type=f.TripType-Street-Hail -3.888619  
## RateCodeID=Standard rate -4.700887  
## Payment\_type=f.PayType-Cash -8.759417  
##   
## $`6`  
## Cla/Mod Mod/Cla Global p.value  
## Payment\_type=f.PayType-Credit card 2.3954527 72.839506 49.289574 1.592820e-05  
## RateCodeID=Others 7.0921986 12.345679 2.821693 9.249028e-05  
## f.Extra=f.Extra-0 2.1230503 60.493827 46.187713 9.692661e-03  
## lpep\_pickup\_date=2016-01-04 4.0816327 7.407407 2.941765 4.046929e-02  
## lpep\_pickup\_date=2016-01-31 3.7234043 8.641975 3.762257 4.239933e-02  
## lpep\_pickup\_period=Period valley 2.2463207 35.802469 25.835501 4.563025e-02  
## f.Extra=f.Extra-0.5 1.1105235 25.925926 37.842706 2.372100e-02  
## RateCodeID=Standard rate 1.4621087 87.654321 97.178307 9.249028e-05  
## Payment\_type=f.PayType-Cash 0.8778931 27.160494 50.150090 2.489104e-05  
## v.test  
## Payment\_type=f.PayType-Credit card 4.315445  
## RateCodeID=Others 3.909492  
## f.Extra=f.Extra-0 2.586605  
## lpep\_pickup\_date=2016-01-04 2.048927  
## lpep\_pickup\_date=2016-01-31 2.029579  
## lpep\_pickup\_period=Period valley 1.998798  
## f.Extra=f.Extra-0.5 -2.261618  
## RateCodeID=Standard rate -3.909492  
## Payment\_type=f.PayType-Cash -4.215785

#Mod/Cluster Tanto porciento del cluster que forma parte de la modalidad.  
  
### desc.var$quanti.var ###  
### A.3. The quantitative variables which characterizes the clusters ###  
res.hcpc$desc.var$quanti.var

## Eta2 P-value  
## Passenger\_count 0.799291791 0.000000e+00  
## tlenkm 0.727229335 0.000000e+00  
## Fare\_amount 0.670797203 0.000000e+00  
## espeed 0.314290109 0.000000e+00  
## Tip\_amount 0.272032915 0.000000e+00  
## Tolls\_amount 0.990427584 0.000000e+00  
## traveltime 0.777994735 0.000000e+00  
## distHaversine 0.681381026 0.000000e+00  
## Total\_amount 0.678643949 0.000000e+00  
## lpep\_pickup\_time 0.241595171 2.193042e-296  
## Dropoff\_longitude 0.015907729 8.161495e-16  
## Dropoff\_latitude 0.012727075 1.852248e-12  
## Pickup\_latitude 0.010436964 4.507653e-10  
## Pickup\_longitude 0.003898793 1.548666e-03

### desc.var$quanti ###  
### A.4. The description of each cluster by the quantitative variables ###  
res.hcpc$desc.var$quanti

## $`1`  
## v.test Mean in category Overall mean sd in category  
## Dropoff\_latitude 6.608694 40.7487059 40.74452020 0.05661432  
## Pickup\_latitude 5.269799 40.7497613 40.74647599 0.05595800  
## Tolls\_amount -11.895382 0.0000000 0.09184573 0.00000000  
## Passenger\_count -22.239291 1.1154957 1.37422453 0.38266142  
## traveltime -26.743408 7.9959217 12.97009394 4.41713814  
## Tip\_amount -27.546096 0.6343773 1.22829421 0.96136398  
## espeed -30.729194 17.8137784 20.95844647 5.47483395  
## Total\_amount -45.784392 9.3692360 14.51814561 3.42583130  
## distHaversine -45.791318 1.6900619 3.21073979 0.89187850  
## Fare\_amount -46.528084 7.5410027 11.97982650 2.84532120  
## tlenkm -46.851295 2.2499347 4.55938154 1.12112368  
## Overall sd p.value  
## Dropoff\_latitude 0.05764356 3.877246e-11  
## Pickup\_latitude 0.05673910 1.365732e-07  
## Tolls\_amount 0.70271829 1.250789e-32  
## Passenger\_count 1.05882609 1.432004e-109  
## traveltime 16.92796531 1.473215e-157  
## Tip\_amount 1.96230427 4.928480e-167  
## espeed 9.31374480 2.319563e-207  
## Total\_amount 10.23525129 0.000000e+00  
## distHaversine 3.02241973 0.000000e+00  
## Fare\_amount 8.68267276 0.000000e+00  
## tlenkm 4.48628707 0.000000e+00  
##   
## $`2`  
## v.test Mean in category Overall mean sd in category  
## Passenger\_count 63.029866 5.2062937 1.37422453 0.59961722  
## Tolls\_amount -2.276226 0.0000000 0.09184573 0.00000000  
## Pickup\_latitude -2.280332 40.7390468 40.74647599 0.04718369  
## Total\_amount -2.287423 13.1738112 14.51814561 7.61900952  
## tlenkm -2.379404 3.9464421 4.55938154 3.11639707  
## distHaversine -2.469995 2.7820796 3.21073979 2.00794746  
## Tip\_amount -2.832638 0.9091259 1.22829421 1.42164079  
## Overall sd p.value  
## Passenger\_count 1.0588261 0.000000000  
## Tolls\_amount 0.7027183 0.022832482  
## Pickup\_latitude 0.0567391 0.022588017  
## Total\_amount 10.2352513 0.022171144  
## tlenkm 4.4862871 0.017340660  
## distHaversine 3.0224197 0.013511489  
## Tip\_amount 1.9623043 0.004616559  
##   
## $`3`  
## v.test Mean in category Overall mean sd in category  
## Fare\_amount 24.561972 17.317603306 11.97982650 5.64152513  
## distHaversine 23.041016 4.953750900 3.21073979 1.75697148  
## Total\_amount 22.941876 20.395355372 14.51814561 6.23907479  
## tlenkm 22.902459 7.131037812 4.55938154 2.26747907  
## espeed 15.718177 24.622571293 20.95844647 9.41802065  
## traveltime 14.244813 19.005482909 12.97009394 8.34358268  
## Tip\_amount 13.119575 1.872656132 1.22829421 2.02064294  
## Pickup\_longitude -3.862126 -73.939944898 -73.93583169 0.04269785  
## Pickup\_latitude -4.027712 40.740756140 40.74647599 0.05808518  
## Dropoff\_latitude -4.833180 40.737547075 40.74452020 0.05780033  
## Tolls\_amount -4.955566 0.004685526 0.09184573 0.10454202  
## Dropoff\_longitude -6.815254 -73.943508255 -73.93487331 0.05229435  
## Passenger\_count -8.538827 1.147933884 1.37422453 0.43457953  
## Overall sd p.value  
## Fare\_amount 8.68267276 3.221839e-133  
## distHaversine 3.02241973 1.809957e-117  
## Total\_amount 10.23525129 1.776040e-116  
## tlenkm 4.48628707 4.391325e-116  
## espeed 9.31374480 1.135440e-55  
## traveltime 16.92796531 4.828388e-46  
## Tip\_amount 1.96230427 2.543573e-39  
## Pickup\_longitude 0.04255109 1.124044e-04  
## Pickup\_latitude 0.05673910 5.632221e-05  
## Dropoff\_latitude 0.05764356 1.343691e-06  
## Tolls\_amount 0.70271829 7.212020e-07  
## Dropoff\_longitude 0.05062133 9.409739e-12  
## Passenger\_count 1.05882609 1.355910e-17  
##   
## $`4`  
## v.test Mean in category Overall mean sd in category  
## traveltime 55.553018 396.69444 12.97009 104.81931794  
## lpep\_pickup\_time 34.465708 123.54051 13.63317 31.04231660  
## Dropoff\_latitude 2.302874 40.79869 40.74452 0.04491003  
## espeed -5.514512 0.00100 20.95845 0.00000000  
## Overall sd p.value  
## traveltime 16.92796531 0.000000e+00  
## lpep\_pickup\_time 7.81506427 2.619603e-260  
## Dropoff\_latitude 0.05764356 2.128592e-02  
## espeed 9.31374480 3.497490e-08  
##   
## $`5`  
## v.test Mean in category Overall mean sd in category  
## tlenkm 46.548461 16.312466 4.559382 5.60582728  
## distHaversine 44.471177 10.775460 3.210740 3.72617779  
## Fare\_amount 43.676267 33.322973 11.979827 10.41356589  
## Total\_amount 42.827386 39.188738 14.518146 12.81978817  
## espeed 29.268841 36.300707 20.958446 14.60529934  
## Tip\_amount 28.475817 4.373159 1.228294 3.85440534  
## traveltime 19.128415 31.194075 12.970094 16.68078478  
## Dropoff\_longitude 5.933924 -73.917968 -73.934873 0.09098760  
## lpep\_pickup\_time -3.165715 12.240767 13.633166 6.97638989  
## Pickup\_latitude -3.292156 40.735963 40.746476 0.06486814  
## Dropoff\_latitude -4.256127 40.730712 40.744520 0.06880903  
## Overall sd p.value  
## tlenkm 4.48628707 0.000000e+00  
## distHaversine 3.02241973 0.000000e+00  
## Fare\_amount 8.68267276 0.000000e+00  
## Total\_amount 10.23525129 0.000000e+00  
## espeed 9.31374480 2.585711e-188  
## Tip\_amount 1.96230427 2.334972e-178  
## traveltime 16.92796531 1.464587e-81  
## Dropoff\_longitude 0.05062133 2.957795e-09  
## lpep\_pickup\_time 7.81506427 1.547024e-03  
## Pickup\_latitude 0.05673910 9.942254e-04  
## Dropoff\_latitude 0.05764356 2.079985e-05  
##   
## $`6`  
## v.test Mean in category Overall mean sd in category  
## Tolls\_amount 70.342197 5.540000 0.09184573 0.00000000  
## Total\_amount 21.921660 39.248148 14.51814561 15.31579586  
## tlenkm 19.302781 14.104013 4.55938154 7.96807771  
## distHaversine 18.782508 9.467659 3.21073979 6.11971380  
## Fare\_amount 16.929420 28.181047 11.97982650 12.83914942  
## espeed 13.885132 35.212112 20.95844647 14.02499409  
## Tip\_amount 13.241123 4.092099 1.22829421 3.83249289  
## traveltime 5.988931 24.144033 12.97009394 9.11695699  
## Pickup\_latitude 3.273578 40.766948 40.74647599 0.04679217  
## Dropoff\_longitude 2.598702 -73.920374 -73.93487331 0.06862988  
## lpep\_pickup\_time -2.211353 11.728395 13.63316621 5.85422153  
## Overall sd p.value  
## Tolls\_amount 0.70271829 0.000000e+00  
## Total\_amount 10.23525129 1.614610e-106  
## tlenkm 4.48628707 5.089531e-83  
## distHaversine 3.02241973 1.050011e-78  
## Fare\_amount 8.68267276 2.730574e-64  
## espeed 9.31374480 7.795656e-44  
## Tip\_amount 1.96230427 5.078388e-40  
## traveltime 16.92796531 2.112245e-09  
## Pickup\_latitude 0.05673910 1.061950e-03  
## Dropoff\_longitude 0.05062133 9.357705e-03  
## lpep\_pickup\_time 7.81506427 2.701142e-02

### desc.axes ###  
### B. The description of the clusters by the axes ###  
res.hcpc$desc.axes

##   
## Link between the cluster variable and the quantitative variables  
## ================================================================  
## Eta2 P-value  
## Dim.1 0.7885045 0  
## Dim.2 0.5302744 0  
## Dim.3 0.7982828 0  
## Dim.4 0.9208969 0  
##   
## Description of each cluster by quantitative variables  
## =====================================================  
## $`1`  
## v.test Mean in category Overall mean sd in category Overall sd  
## Dim.4 9.783385 0.1014888 -1.503838e-14 0.1732069 0.9441258  
## Dim.3 -20.526302 -0.2257144 1.334529e-13 0.3614196 1.0008046  
## Dim.1 -48.862098 -1.0426045 8.620370e-15 0.5198133 1.9419954  
## p.value  
## Dim.4 1.326971e-22  
## Dim.3 1.253601e-93  
## Dim.1 0.000000e+00  
##   
## $`2`  
## v.test Mean in category Overall mean sd in category Overall sd  
## Dim.3 62.088520 3.5679841 1.334529e-13 0.5649030 1.0008046  
## Dim.1 -2.519380 -0.2809339 8.620370e-15 1.3020644 1.9419954  
## Dim.2 -4.044535 -0.2770348 7.781184e-15 0.8175058 1.1928981  
## Dim.4 -10.295932 -0.5581589 -1.503838e-14 0.3300727 0.9441258  
## p.value  
## Dim.3 0.000000e+00  
## Dim.1 1.175618e-02  
## Dim.2 5.242705e-05  
## Dim.4 7.350400e-25  
##   
## $`3`  
## v.test Mean in category Overall mean sd in category Overall sd  
## Dim.1 23.573239 1.14580628 8.620370e-15 0.8106372 1.9419954  
## Dim.2 1.994082 0.05953744 7.781184e-15 0.8783582 1.1928981  
## Dim.3 -10.472846 -0.26233591 1.334529e-13 0.4135612 1.0008046  
## Dim.4 -13.591523 -0.32117501 -1.503838e-14 0.3001165 0.9441258  
## p.value  
## Dim.1 7.253458e-123  
## Dim.2 4.614309e-02  
## Dim.3 1.151286e-25  
## Dim.4 4.496260e-42  
##   
## $`4`  
## v.test Mean in category Overall mean sd in category Overall sd  
## Dim.2 50.263792 24.466145 7.781184e-15 6.5802620 1.1928981  
## Dim.1 6.166622 4.886551 8.620370e-15 1.1128938 1.9419954  
## Dim.4 3.339284 1.286442 -1.503838e-14 0.5109452 0.9441258  
## Dim.3 3.240573 1.323360 1.334529e-13 0.4531454 1.0008046  
## p.value  
## Dim.2 0.000000e+00  
## Dim.1 6.976429e-10  
## Dim.4 8.399473e-04  
## Dim.3 1.192895e-03  
##   
## $`5`  
## v.test Mean in category Overall mean sd in category Overall sd  
## Dim.1 45.199565 4.9401704 8.620370e-15 1.8674677 1.9419954  
## Dim.3 -5.159056 -0.2905889 1.334529e-13 0.7971951 1.0008046  
## Dim.4 -19.640487 -1.0436181 -1.503838e-14 0.5109345 0.9441258  
## p.value  
## Dim.1 0.000000e+00  
## Dim.3 2.481977e-07  
## Dim.4 6.972697e-86  
##   
## $`6`  
## v.test Mean in category Overall mean sd in category Overall sd  
## Dim.4 63.37186 6.594450 -1.503838e-14 0.6990761 0.9441258  
## Dim.1 25.79213 5.520616 8.620370e-15 3.1822877 1.9419954  
## Dim.3 8.83056 0.974069 1.334529e-13 1.1876405 1.0008046  
## Dim.2 -10.30421 -1.354784 7.781184e-15 0.8459960 1.1928981  
## p.value  
## Dim.4 0.000000e+00  
## Dim.1 1.086752e-146  
## Dim.3 1.041531e-18  
## Dim.2 6.744500e-25

names(res.hcpc$desc.axes)

## [1] "quanti.var" "quanti" "call"

res.hcpc$desc.axes$quanti.var

## Eta2 P-value  
## Dim.1 0.7885045 0  
## Dim.2 0.5302744 0  
## Dim.3 0.7982828 0  
## Dim.4 0.9208969 0

res.hcpc$desc.axes$quanti

## $`1`  
## v.test Mean in category Overall mean sd in category Overall sd  
## Dim.4 9.783385 0.1014888 -1.503838e-14 0.1732069 0.9441258  
## Dim.3 -20.526302 -0.2257144 1.334529e-13 0.3614196 1.0008046  
## Dim.1 -48.862098 -1.0426045 8.620370e-15 0.5198133 1.9419954  
## p.value  
## Dim.4 1.326971e-22  
## Dim.3 1.253601e-93  
## Dim.1 0.000000e+00  
##   
## $`2`  
## v.test Mean in category Overall mean sd in category Overall sd  
## Dim.3 62.088520 3.5679841 1.334529e-13 0.5649030 1.0008046  
## Dim.1 -2.519380 -0.2809339 8.620370e-15 1.3020644 1.9419954  
## Dim.2 -4.044535 -0.2770348 7.781184e-15 0.8175058 1.1928981  
## Dim.4 -10.295932 -0.5581589 -1.503838e-14 0.3300727 0.9441258  
## p.value  
## Dim.3 0.000000e+00  
## Dim.1 1.175618e-02  
## Dim.2 5.242705e-05  
## Dim.4 7.350400e-25  
##   
## $`3`  
## v.test Mean in category Overall mean sd in category Overall sd  
## Dim.1 23.573239 1.14580628 8.620370e-15 0.8106372 1.9419954  
## Dim.2 1.994082 0.05953744 7.781184e-15 0.8783582 1.1928981  
## Dim.3 -10.472846 -0.26233591 1.334529e-13 0.4135612 1.0008046  
## Dim.4 -13.591523 -0.32117501 -1.503838e-14 0.3001165 0.9441258  
## p.value  
## Dim.1 7.253458e-123  
## Dim.2 4.614309e-02  
## Dim.3 1.151286e-25  
## Dim.4 4.496260e-42  
##   
## $`4`  
## v.test Mean in category Overall mean sd in category Overall sd  
## Dim.2 50.263792 24.466145 7.781184e-15 6.5802620 1.1928981  
## Dim.1 6.166622 4.886551 8.620370e-15 1.1128938 1.9419954  
## Dim.4 3.339284 1.286442 -1.503838e-14 0.5109452 0.9441258  
## Dim.3 3.240573 1.323360 1.334529e-13 0.4531454 1.0008046  
## p.value  
## Dim.2 0.000000e+00  
## Dim.1 6.976429e-10  
## Dim.4 8.399473e-04  
## Dim.3 1.192895e-03  
##   
## $`5`  
## v.test Mean in category Overall mean sd in category Overall sd  
## Dim.1 45.199565 4.9401704 8.620370e-15 1.8674677 1.9419954  
## Dim.3 -5.159056 -0.2905889 1.334529e-13 0.7971951 1.0008046  
## Dim.4 -19.640487 -1.0436181 -1.503838e-14 0.5109345 0.9441258  
## p.value  
## Dim.1 0.000000e+00  
## Dim.3 2.481977e-07  
## Dim.4 6.972697e-86  
##   
## $`6`  
## v.test Mean in category Overall mean sd in category Overall sd  
## Dim.4 63.37186 6.594450 -1.503838e-14 0.6990761 0.9441258  
## Dim.1 25.79213 5.520616 8.620370e-15 3.1822877 1.9419954  
## Dim.3 8.83056 0.974069 1.334529e-13 1.1876405 1.0008046  
## Dim.2 -10.30421 -1.354784 7.781184e-15 0.8459960 1.1928981  
## p.value  
## Dim.4 0.000000e+00  
## Dim.1 1.086752e-146  
## Dim.3 1.041531e-18  
## Dim.2 6.744500e-25

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

## [1] "para" "dist"

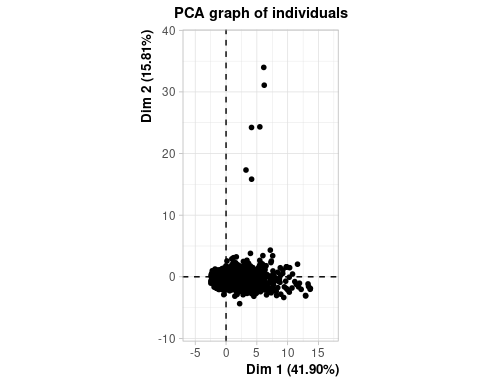
res.hcpc$desc.ind$para

## Cluster: 1  
## 614 2365 309 1177 2370   
## 0.1053532 0.1103592 0.1110806 0.1122301 0.1165857   
## ------------------------------------------------------------   
## Cluster: 2  
## 1925 489 4637 2338 3400   
## 0.2743467 0.3001034 0.3985722 0.4332463 0.4342648   
## ------------------------------------------------------------   
## Cluster: 3  
## 571 3195 4327 1174 335   
## 0.1522525 0.1728705 0.1766572 0.1895828 0.2003418   
## ------------------------------------------------------------   
## Cluster: 4  
## 3086 1604 3064 1426 1094   
## 0.6517680 0.7957736 6.7860419 7.3487257 8.7179857   
## ------------------------------------------------------------   
## Cluster: 5  
## 4296 1642 3773 2980 958   
## 0.3967582 0.4328118 0.4335242 0.4342343 0.4480911   
## ------------------------------------------------------------   
## Cluster: 6  
## 2124 3567 4308 4912 1362   
## 0.6683019 0.6800916 0.6882831 0.7807697 0.8491133

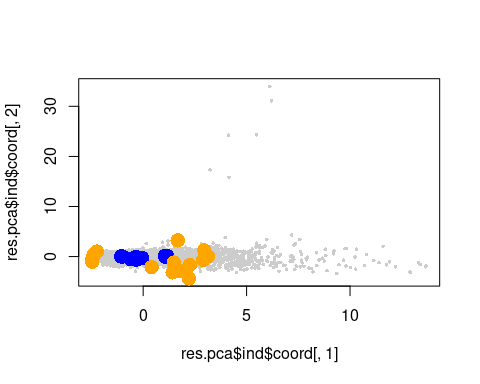
res.hcpc$desc.ind$dist

## Cluster: 1  
## 1492 61 2218 892 4793   
## 3.816327 3.731034 3.634026 3.614683 3.601428   
## ------------------------------------------------------------   
## Cluster: 2  
## 4761 1491 2326 264 2141   
## 5.056433 5.046111 5.017642 5.000313 4.989005   
## ------------------------------------------------------------   
## Cluster: 3  
## 3061 1464 4414 3588 1832   
## 5.080308 4.219700 4.168121 4.110150 4.108611   
## ------------------------------------------------------------   
## Cluster: 4  
## 3370 3064 3086 1604 1426   
## 34.32624 31.43010 24.59804 24.46856 17.47935   
## ------------------------------------------------------------   
## Cluster: 5  
## 1008 1092 3596 3547 3688   
## 12.572148 10.840811 10.762154 10.238636 9.777473   
## ------------------------------------------------------------   
## Cluster: 6  
## 1155 611 4947 3322 985   
## 11.20383 10.87733 10.68739 10.62744 10.50843

# Examinar los valores de los individuos que caracterizan a las clases  
#### Characteristic individuals  
para1<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[1]]))  
para2<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[2]]))  
para3<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$para[[3]]))  
# to be completed...with as many clusters as you have selected  
  
dist1<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[1]]))  
dist2<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[2]]))  
dist3<-which(rownames(res.pca$ind$coord)%in%names(res.hcpc$desc.ind$dist[[3]]))  
# to be completed...with as many clusters as you have selected  
  
plot(res.pca,label="none",invisible=c("quali","ind.sup")) # Potser no us va



plot(res.pca$ind$coord[,1],res.pca$ind$coord[,2],col="grey80",cex=0.5,pch=16)  
points(res.pca$ind$coord[para1,1],res.pca$ind$coord[para1,2],col="blue",cex=2,pch=16)  
points(res.pca$ind$coord[dist1,1],res.pca$ind$coord[dist1,2],col="orange",cex=2,pch=16)  
points(res.pca$ind$coord[para2,1],res.pca$ind$coord[para2,2],col="blue",cex=2,pch=16)  
points(res.pca$ind$coord[dist2,1],res.pca$ind$coord[dist2,2],col="orange",cex=2,pch=16)  
points(res.pca$ind$coord[para3,1],res.pca$ind$coord[para3,2],col="blue",cex=2,pch=16)  
points(res.pca$ind$coord[dist3,1],res.pca$ind$coord[dist3,2],col="orange",cex=2,pch=16)



# to be completed...  
  
res.hcpc$data.clust[which(rownames(res.hcpc$data.clust)%in%names(res.hcpc$desc.ind$para[[1]])),]

## VendorID Payment\_type Store\_and\_fwd\_flag RateCodeID  
## 309 f.Vendor-VeriFone f.PayType-Cash FALSE Standard rate  
## 614 f.Vendor-VeriFone f.PayType-Cash FALSE Standard rate  
## 1177 f.Vendor-VeriFone f.PayType-Cash FALSE Standard rate  
## 2365 f.Vendor-VeriFone f.PayType-Cash FALSE Standard rate  
## 2370 f.Vendor-VeriFone f.PayType-Cash FALSE Standard rate  
## f.Extra f.MTA\_tax f.Improvement\_surcharge lpep\_pickup\_period  
## 309 f.Extra-0 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period night  
## 614 f.Extra-0 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period valley  
## 1177 f.Extra-0 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period night  
## 2365 f.Extra-0 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period valley  
## 2370 f.Extra-0 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period valley  
## Trip\_type lpep\_pickup\_date Passenger\_count tlenkm  
## 309 f.TripType-Street-Hail 2016-01-01 1 2.446203  
## 614 f.TripType-Street-Hail 2016-01-04 1 2.574950  
## 1177 f.TripType-Street-Hail 2016-01-01 1 2.333549  
## 2365 f.TripType-Street-Hail 2016-01-15 1 2.639324  
## 2370 f.TripType-Street-Hail 2016-01-15 1 2.880726  
## Pickup\_longitude Pickup\_latitude Dropoff\_longitude Dropoff\_latitude  
## 309 -73.95614 40.69470 -73.97961 40.69614  
## 614 -73.95305 40.80684 -73.93969 40.82020  
## 1177 -73.92566 40.77220 -73.90671 40.76989  
## 2365 -73.94530 40.79366 -73.94392 40.77645  
## 2370 -73.94859 40.80878 -73.94301 40.79098  
## Fare\_amount espeed Tip\_amount Tolls\_amount lpep\_pickup\_time traveltime  
## 309 8.0 17.07633 0 0 13.77690 8.879814  
## 614 8.0 17.86093 0 0 14.00000 8.650000  
## 1177 9.0 19.13190 0 0 13.93455 8.766875  
## 2365 8.5 15.10583 0 0 12.00000 10.483333  
## 2370 7.5 17.37121 0 0 13.00000 9.950000  
## distHaversine Total\_amount clust  
## 309 1.987182 8.8 1  
## 614 1.864689 8.8 1  
## 1177 1.618292 9.8 1  
## 2365 1.919977 9.3 1  
## 2370 2.036805 8.3 1

res.hcpc$data.clust[which(rownames(res.hcpc$data.clust)%in%names(res.hcpc$desc.ind$dist[[1]])),]

## VendorID Payment\_type Store\_and\_fwd\_flag RateCodeID  
## 61 f.Vendor-VeriFone f.PayType-Cash FALSE Standard rate  
## 892 f.Vendor-VeriFone f.PayType-Cash FALSE Standard rate  
## 1492 f.Vendor-Mobile f.PayType-Cash FALSE Others  
## 2218 f.Vendor-VeriFone f.PayType-Cash FALSE Standard rate  
## 4793 f.Vendor-VeriFone f.PayType-Cash FALSE Standard rate  
## f.Extra f.MTA\_tax f.Improvement\_surcharge lpep\_pickup\_period  
## 61 f.Extra-0.5 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period night  
## 892 f.Extra-0 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period valley  
## 1492 f.Extra-0 f.MTA\_tax\_NO f.Improvement\_surcharge\_NO Period night  
## 2218 f.Extra-1 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period valley  
## 4793 f.Extra-0.5 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period night  
## Trip\_type lpep\_pickup\_date Passenger\_count tlenkm  
## 61 f.TripType-Street-Hail 2016-01-01 1 0.04828032  
## 892 f.TripType-Street-Hail 2016-01-06 1 0.06437376  
## 1492 f.TripType-Dispatch 2016-01-10 2 0.16093440  
## 2218 f.TripType-Street-Hail 2016-01-14 1 0.09656064  
## 4793 f.TripType-Street-Hail 2016-01-30 1 0.19312128  
## Pickup\_longitude Pickup\_latitude Dropoff\_longitude Dropoff\_latitude  
## 61 -73.88078 40.82846 -73.88282 40.82843  
## 892 -73.86945 40.73414 -73.86925 40.73481  
## 1492 -73.90463 40.85099 -73.90364 40.85223  
## 2218 -73.95387 40.81134 -73.95112 40.81020  
## 4793 -73.95201 40.80895 -73.94799 40.81446  
## Fare\_amount espeed Tip\_amount Tolls\_amount lpep\_pickup\_time traveltime  
## 61 2.5 4.042073 0 0 3 0.7166667  
## 892 2.5 7.022592 0 0 14 0.5500000  
## 1492 0.0 7.333719 0 0 1 1.3166667  
## 2218 2.5 5.993419 0 0 16 0.9666667  
## 4793 3.0 7.556920 0 0 23 1.5333333  
## distHaversine Total\_amount clust  
## 61 0.1716187 3.8 1  
## 892 0.0767317 3.3 1  
## 1492 0.1605865 0.0 1  
## 2218 0.2641138 4.3 1  
## 4793 0.7012892 4.3 1

res.hcpc$data.clust[which(rownames(res.hcpc$data.clust)%in%names(res.hcpc$desc.ind$para[[2]])),]

## VendorID Payment\_type Store\_and\_fwd\_flag RateCodeID  
## 489 f.Vendor-VeriFone f.PayType-Cash FALSE Standard rate  
## 1925 f.Vendor-VeriFone f.PayType-Cash FALSE Standard rate  
## 2338 f.Vendor-VeriFone f.PayType-Cash FALSE Standard rate  
## 3400 f.Vendor-VeriFone f.PayType-Credit card FALSE Standard rate  
## 4637 f.Vendor-VeriFone f.PayType-Credit card FALSE Standard rate  
## f.Extra f.MTA\_tax f.Improvement\_surcharge lpep\_pickup\_period  
## 489 f.Extra-0 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period afternoon  
## 1925 f.Extra-1 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period afternoon  
## 2338 f.Extra-0 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period morning  
## 3400 f.Extra-1 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period afternoon  
## 4637 f.Extra-0 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period night  
## Trip\_type lpep\_pickup\_date Passenger\_count tlenkm  
## 489 f.TripType-Street-Hail 2016-01-03 5 4.876312  
## 1925 f.TripType-Street-Hail 2016-01-12 5 3.701491  
## 2338 f.TripType-Street-Hail 2016-01-15 5 3.653211  
## 3400 f.TripType-Street-Hail 2016-01-21 5 2.414016  
## 4637 f.TripType-Street-Hail 2016-01-30 5 3.669304  
## Pickup\_longitude Pickup\_latitude Dropoff\_longitude Dropoff\_latitude  
## 489 -73.89082 40.74651 -73.87373 40.76523  
## 1925 -73.94909 40.79419 -73.94024 40.82050  
## 2338 -73.96690 40.80415 -73.93933 40.80517  
## 3400 -73.94346 40.74673 -73.95760 40.73350  
## 4637 -73.94351 40.70168 -73.95590 40.73005  
## Fare\_amount espeed Tip\_amount Tolls\_amount lpep\_pickup\_time traveltime  
## 489 11.5 29.20919 0.00 0 17 10.016667  
## 1925 9.0 26.97443 0.00 0 19 8.233333  
## 2338 10.5 17.12443 0.00 0 9 12.800000  
## 3400 6.5 30.81723 1.66 0 17 4.700000  
## 4637 10.5 16.78462 1.00 0 7 13.116667  
## distHaversine Total\_amount clust  
## 489 2.533183 12.30 2  
## 1925 3.021841 10.80 2  
## 2338 2.326080 11.30 2  
## 3400 1.895628 9.96 2  
## 4637 3.326656 12.30 2

res.hcpc$data.clust[which(rownames(res.hcpc$data.clust)%in%names(res.hcpc$desc.ind$dist[[2]])),]

## VendorID Payment\_type Store\_and\_fwd\_flag RateCodeID  
## 264 f.Vendor-VeriFone f.PayType-Credit card FALSE Standard rate  
## 1491 f.Vendor-VeriFone f.PayType-Credit card FALSE Standard rate  
## 2141 f.Vendor-VeriFone f.PayType-Credit card FALSE Standard rate  
## 2326 f.Vendor-VeriFone f.PayType-Cash FALSE Standard rate  
## 4761 f.Vendor-VeriFone f.PayType-Credit card FALSE Standard rate  
## f.Extra f.MTA\_tax f.Improvement\_surcharge lpep\_pickup\_period  
## 264 f.Extra-0 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period valley  
## 1491 f.Extra-0.5 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period night  
## 2141 f.Extra-0.5 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period night  
## 2326 f.Extra-0 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period morning  
## 4761 f.Extra-0.5 f.MTA\_tax\_YES f.Improvement\_surcharge\_YES Period night  
## Trip\_type lpep\_pickup\_date Passenger\_count tlenkm  
## 264 f.TripType-Street-Hail 2016-01-02 6 13.148340  
## 1491 f.TripType-Street-Hail 2016-01-10 6 10.991820  
## 2141 f.TripType-Street-Hail 2016-01-14 6 14.725498  
## 2326 f.TripType-Street-Hail 2016-01-15 6 3.958986  
## 4761 f.TripType-Street-Hail 2016-01-30 6 9.881372  
## Pickup\_longitude Pickup\_latitude Dropoff\_longitude Dropoff\_latitude  
## 264 -73.97984 40.67737 -73.97252 40.74634  
## 1491 -73.90288 40.77547 -73.82043 40.78169  
## 2141 -73.88847 40.74707 -73.86604 40.68958  
## 2326 -73.83913 40.72534 -73.84893 40.75195  
## 4761 -73.89146 40.74667 -73.98812 40.76459  
## Fare\_amount espeed Tip\_amount Tolls\_amount lpep\_pickup\_time traveltime  
## 264 25.5 37.15387 0.00 0 13 21.233333  
## 1491 22.5 33.93701 0.00 0 1 19.433333  
## 2141 33.5 19.54712 0.00 0 4 45.200000  
## 2326 8.5 50.18433 0.00 0 8 4.733333  
## 4761 25.0 17.83999 5.26 0 21 33.233333  
## distHaversine Total\_amount clust  
## 264 7.702457 26.30 2  
## 1491 6.984618 23.80 2  
## 2141 6.674442 34.80 2  
## 2326 3.076015 9.30 2  
## 4761 8.391677 31.56 2

# to be completed...  
   
# Partition quality  
#(res.hcpc$call$t$within[1]-res.hcpc$call$t$within['nb.clusters'])/res.hcpc$call$t$within[1]  
(res.hcpc$call$t$within[1]-res.hcpc$call$t$within[6])/res.hcpc$call$t$within[1]

## [1] 0.7332846

### Results for the hierarchical tree ###  
names(res.hcpc$call$t)

## [1] "res" "tree" "nb.clust" "within" "inert.gain"  
## [6] "quot"

### The suggested level to cut the tree ###  
res.hcpc$call$t$nb.clust

## [1] 6

### Within inertias ###  
res.hcpc$call$t$within[1:6]

## [1] 7.087335 5.121886 4.140307 3.355544 2.619776 1.890301

### Ratio between within inertias ###  
res.hcpc$call$t$quot[1:6]

## [1] 0.8083560 0.8104579 0.7807305 0.7215508 0.8607475 0.9030008

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

## [1] 1.9654497 0.9815786 0.7847626 0.7357684 0.7294745 0.2632292

#df$hcpck<-res.hcpc$data.clust$clust  
  
####  
#### THE END  
####

## K-Means

res.pca<-PCA(df[,c(1:10,19:(d2-2),d2)],quali.sup=c(1:10),quanti.sup=c(13:16,24), ncp=5, graph= FALSE)   
ppcc<-res.pca$ind$coord[,1:5]  
dim(ppcc)

## [1] 5000 5

library("factoextra")  
#fviz\_nbclust(ppcc, kmeans, method = "gap\_stat")  
library("NbClust") # It takes a lot ....  
set.seed(123)  
#res.nbclust <- NbClust(ppcc, distance = "euclidean", min.nc = 2, max.nc = 10, method = "complete", index ="all") # Time consuming  
  
dist<-dist(ppcc)  
kc<-kmeans(dist,6, iter.max = 30, trace=T)

## KMNS(\*, k=6): iter= 1, indx=1  
## QTRAN(): istep=5000, icoun=9  
## QTRAN(): istep=10000, icoun=2  
## QTRAN(): istep=15000, icoun=93  
## QTRAN(): istep=20000, icoun=55  
## QTRAN(): istep=25000, icoun=113  
## QTRAN(): istep=30000, icoun=2  
## QTRAN(): istep=35000, icoun=560  
## QTRAN(): istep=40000, icoun=1195  
## QTRAN(): istep=45000, icoun=1424  
## KMNS(\*, k=6): iter= 2, indx=2  
## QTRAN(): istep=5000, icoun=6  
## QTRAN(): istep=10000, icoun=17  
## QTRAN(): istep=15000, icoun=49  
## QTRAN(): istep=20000, icoun=51  
## QTRAN(): istep=25000, icoun=812  
## QTRAN(): istep=30000, icoun=116  
## QTRAN(): istep=35000, icoun=316  
## QTRAN(): istep=40000, icoun=362  
## QTRAN(): istep=45000, icoun=1992  
## QTRAN(): istep=50000, icoun=3142  
## QTRAN(): istep=55000, icoun=3840  
## QTRAN(): istep=60000, icoun=4321  
## KMNS(\*, k=6): iter= 3, indx=26  
## QTRAN(): istep=5000, icoun=17  
## QTRAN(): istep=10000, icoun=2664  
## KMNS(\*, k=6): iter= 4, indx=5000

df$claKM<-0  
df[names(kc$cluster),"claKM"]<-kc$cluster  
df$claKM<-factor(df$claKM)  
kc$betweenss/kc$totss

## [1] 0.7801615

#names(df)  
#catdes(df,38)  
  
# Confusion table  
  
#table(df$hcpck,df$claKM)  
  
#df$hcpck<-factor(df$hcpck,labels=paste("kHP-",1:6))  
#df$claKM<-factor(df$claKM,levels=c(3,6,2,1,5,4),labels=c("kKM-3","kKM-6","kKM-2","kKM-1","kKM-5","kKM-4"))  
#tt<-table(df$hcpck,df$claKM)  
#tt  
#sum(diag(tt)/sum(tt))