Cluster\_2023\_1

2023-03-25

library(pvclust)  
library(corrplot)

## corrplot 0.92 loaded

library(car)

## Loading required package: carData

library(factoextra)

## Loading required package: ggplot2

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

library(e1071)  
library(caret)

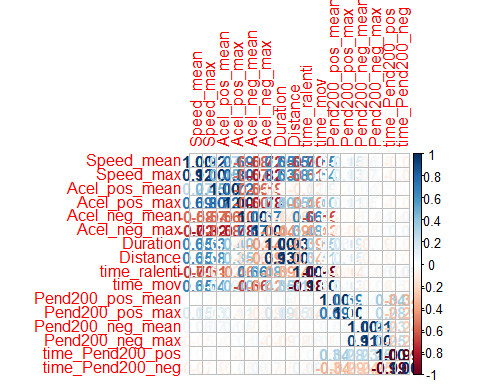
## Loading required package: lattice

datos<-read.csv(file=file.choose(),header=T,sep=';')  
attach(datos)

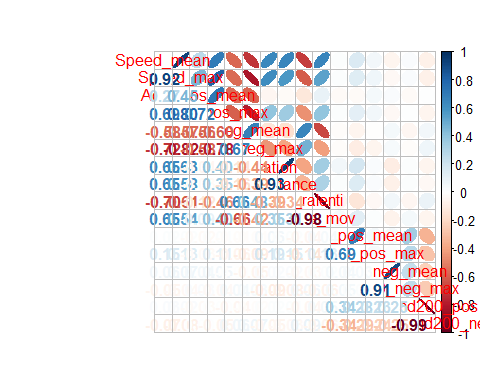
cont<-datos[,c(3:12,18:23)]  
corr\_matrix<-cor(cont)  
corr\_matrix

## Speed\_mean Speed\_max Acel\_pos\_mean Acel\_pos\_max  
## Speed\_mean 1.00000000 0.92339032 0.266168971 0.6933886085  
## Speed\_max 0.92339032 1.00000000 0.452534234 0.8049387910  
## Acel\_pos\_mean 0.26616897 0.45253423 1.000000000 0.7198863630  
## Acel\_pos\_max 0.69338861 0.80493879 0.719886363 1.0000000000  
## Acel\_neg\_mean -0.57956262 -0.56597338 -0.562634318 -0.5973478987  
## Acel\_neg\_max -0.71506734 -0.81911246 -0.584263722 -0.7806670066  
## Duration 0.64935788 0.62903199 -0.041852368 0.3959434780  
## Distance 0.65250664 0.57773603 -0.028185608 0.3520457662  
## time\_ralenti -0.69907412 -0.60526065 -0.124727010 -0.4579862693  
## time\_mov 0.65183122 0.53994745 0.094762228 0.3970370530  
## Pend200\_pos\_mean 0.01961499 0.01848550 -0.021701370 0.0125061662  
## Pend200\_pos\_max 0.15353756 0.13334164 -0.026056862 0.1125076828  
## Pend200\_neg\_mean 0.05720006 0.06551664 0.043970867 0.0516204042  
## Pend200\_neg\_max -0.05071545 -0.03683244 0.028295812 -0.0390272579  
## time\_Pend200\_pos 0.02686537 0.02227195 -0.025000924 0.0009331852  
## time\_Pend200\_neg -0.07400547 -0.07773868 0.003217058 -0.0500546941  
## Acel\_neg\_mean Acel\_neg\_max Duration Distance  
## Speed\_mean -0.579562621 -0.71506734 0.64935788 0.65250664  
## Speed\_max -0.565973381 -0.81911246 0.62903199 0.57773603  
## Acel\_pos\_mean -0.562634318 -0.58426372 -0.04185237 -0.02818561  
## Acel\_pos\_max -0.597347899 -0.78066701 0.39594348 0.35204577  
## Acel\_neg\_mean 1.000000000 0.67379377 -0.13589597 -0.16068995  
## Acel\_neg\_max 0.673793767 1.00000000 -0.44111638 -0.39442165  
## Duration -0.135895968 -0.44111638 1.00000000 0.93221186  
## Distance -0.160689948 -0.39442165 0.93221186 1.00000000  
## time\_ralenti 0.657991598 0.47643054 -0.39048570 -0.33728987  
## time\_mov -0.664316047 -0.41535312 0.34748454 0.31473805  
## Pend200\_pos\_mean -0.012352699 -0.01113436 0.05945340 0.01952358  
## Pend200\_pos\_max -0.058213686 -0.09253833 0.18812378 0.14890490  
## Pend200\_neg\_mean -0.008252028 -0.05442047 0.02066081 0.02184635  
## Pend200\_neg\_max 0.040528905 0.03428801 -0.09476013 -0.07574718  
## time\_Pend200\_pos -0.043906841 -0.01645015 0.01859423 0.01077962  
## time\_Pend200\_neg 0.057233566 0.06600984 -0.05477961 -0.03338687  
## time\_ralenti time\_mov Pend200\_pos\_mean Pend200\_pos\_max  
## Speed\_mean -0.69907412 0.65183122 0.01961499 0.15353756  
## Speed\_max -0.60526065 0.53994745 0.01848550 0.13334164  
## Acel\_pos\_mean -0.12472701 0.09476223 -0.02170137 -0.02605686  
## Acel\_pos\_max -0.45798627 0.39703705 0.01250617 0.11250768  
## Acel\_neg\_mean 0.65799160 -0.66431605 -0.01235270 -0.05821369  
## Acel\_neg\_max 0.47643054 -0.41535312 -0.01113436 -0.09253833  
## Duration -0.39048570 0.34748454 0.05945340 0.18812378  
## Distance -0.33728987 0.31473805 0.01952358 0.14890490  
## time\_ralenti 1.00000000 -0.97591217 -0.04346024 -0.13754526  
## time\_mov -0.97591217 1.00000000 0.03944329 0.12735547  
## Pend200\_pos\_mean -0.04346024 0.03944329 1.00000000 0.68711503  
## Pend200\_pos\_max -0.13754526 0.12735547 0.68711503 1.00000000  
## Pend200\_neg\_mean -0.03245137 0.01927599 0.04302266 0.03261133  
## Pend200\_neg\_max 0.05615067 -0.06105198 0.04587256 0.01084830  
## time\_Pend200\_pos -0.04427825 0.04721246 0.34120259 0.27827483  
## time\_Pend200\_neg 0.08511740 -0.05815343 -0.34171137 -0.28558850  
## Pend200\_neg\_mean Pend200\_neg\_max time\_Pend200\_pos  
## Speed\_mean 0.057200057 -0.05071545 0.0268653718  
## Speed\_max 0.065516638 -0.03683244 0.0222719509  
## Acel\_pos\_mean 0.043970867 0.02829581 -0.0250009242  
## Acel\_pos\_max 0.051620404 -0.03902726 0.0009331852  
## Acel\_neg\_mean -0.008252028 0.04052890 -0.0439068414  
## Acel\_neg\_max -0.054420475 0.03428801 -0.0164501478  
## Duration 0.020660814 -0.09476013 0.0185942291  
## Distance 0.021846346 -0.07574718 0.0107796248  
## time\_ralenti -0.032451375 0.05615067 -0.0442782538  
## time\_mov 0.019275988 -0.06105198 0.0472124572  
## Pend200\_pos\_mean 0.043022661 0.04587256 0.3412025939  
## Pend200\_pos\_max 0.032611330 0.01084830 0.2782748343  
## Pend200\_neg\_mean 1.000000000 0.90767530 0.2322488449  
## Pend200\_neg\_max 0.907675296 1.00000000 0.2530562816  
## time\_Pend200\_pos 0.232248845 0.25305628 1.0000000000  
## time\_Pend200\_neg -0.239011915 -0.25226463 -0.9903407934  
## time\_Pend200\_neg  
## Speed\_mean -0.074005466  
## Speed\_max -0.077738681  
## Acel\_pos\_mean 0.003217058  
## Acel\_pos\_max -0.050054694  
## Acel\_neg\_mean 0.057233566  
## Acel\_neg\_max 0.066009835  
## Duration -0.054779613  
## Distance -0.033386866  
## time\_ralenti 0.085117400  
## time\_mov -0.058153428  
## Pend200\_pos\_mean -0.341711369  
## Pend200\_pos\_max -0.285588497  
## Pend200\_neg\_mean -0.239011915  
## Pend200\_neg\_max -0.252264628  
## time\_Pend200\_pos -0.990340793  
## time\_Pend200\_neg 1.000000000

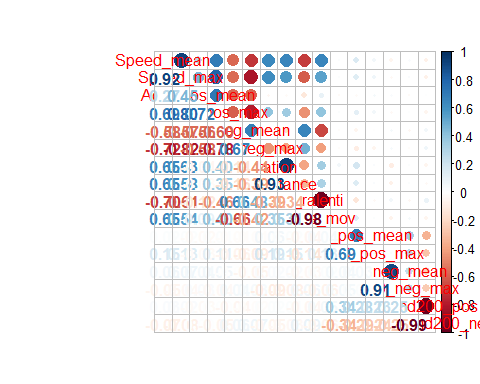
corrplot(corr\_matrix, method = "number")



corrplot.mixed(corr\_matrix, lower="number", upper="ellipse")



corrplot.mixed(corr\_matrix, lower="number", upper="circle")



"Cálculo del PCA"

## [1] "Cálculo del PCA"

res.pca = prcomp(cont, scale = TRUE)# si no quiero estandarizar scale = FALSE  
res.pca

## Standard deviations (1, .., p=16):  
## [1] 2.4184316986 1.6431473614 1.3918231502 1.2836961892 1.1598486736  
## [6] 1.0474923448 0.5899055142 0.5489207325 0.5355524113 0.4240391838  
## [11] 0.3371143160 0.2860026582 0.2486928133 0.1664344125 0.1509866463  
## [16] 0.0009705654  
##   
## Rotation (n x k) = (16 x 16):  
## PC1 PC2 PC3 PC4 PC5  
## Speed\_mean 0.38072213 -0.022742565 -0.07994848 -0.11193101 0.013496444  
## Speed\_max 0.38346553 -0.029807969 0.01598501 -0.06219565 -0.148290818  
## Acel\_pos\_mean 0.19373885 -0.064188761 0.44110529 0.27505918 -0.335160450  
## Acel\_pos\_max 0.33852620 -0.048630349 0.20226805 0.10798772 -0.268711507  
## Acel\_neg\_mean -0.30533692 0.041433747 -0.23893605 -0.24054809 -0.219677621  
## Acel\_neg\_max -0.34645987 0.042084836 -0.16201972 -0.06887546 0.210032399  
## Duration 0.26807441 0.001455934 -0.37282863 -0.33894620 -0.201737718  
## Distance 0.25554367 -0.009928607 -0.35774671 -0.36007243 -0.205711479  
## time\_ralenti -0.31938665 0.004176714 0.05373782 -0.03168012 -0.514212232  
## time\_mov 0.30075503 -0.008935184 -0.05846086 0.04125801 0.562513285  
## Pend200\_pos\_mean 0.03212209 0.349599882 -0.23570865 0.31122774 -0.123222435  
## Pend200\_pos\_max 0.08447286 0.309404196 -0.28770797 0.25395404 -0.129754816  
## Pend200\_neg\_mean 0.02618269 0.337749066 0.34841146 -0.45697001 0.043418075  
## Pend200\_neg\_max -0.02044573 0.349421514 0.37432248 -0.42177924 0.049007009  
## time\_Pend200\_pos 0.03216987 0.517821537 -0.02724308 0.14157317 0.025722517  
## time\_Pend200\_neg -0.05168378 -0.515910147 0.02553131 -0.13533982 -0.009906228  
## PC6 PC7 PC8 PC9  
## Speed\_mean 0.0212420448 0.350812975 0.144261057 -0.057434170  
## Speed\_max 0.0331471657 0.404855664 0.142016483 0.030278576  
## Acel\_pos\_mean -0.0129115257 -0.323442038 -0.256526900 0.311461681  
## Acel\_pos\_max -0.0005667008 0.198712114 -0.236255003 0.360285768  
## Acel\_neg\_mean 0.0227286233 0.504839549 -0.105552211 0.422060335  
## Acel\_neg\_max -0.0313288893 -0.090585196 -0.344697337 0.485999172  
## Duration 0.0637799050 -0.251159846 -0.059118383 0.084840270  
## Distance 0.0843709199 -0.434053459 -0.081494189 0.003124543  
## time\_ralenti 0.0625112178 -0.050198096 0.090475020 -0.205316175  
## time\_mov -0.0711793175 -0.059272141 -0.111630611 0.186119729  
## Pend200\_pos\_mean -0.4537836384 -0.152667405 0.592749875 0.354698112  
## Pend200\_pos\_max -0.4905057899 0.146783266 -0.567998226 -0.379770358  
## Pend200\_neg\_mean -0.2424049511 -0.002705157 -0.008995156 0.003325349  
## Pend200\_neg\_max -0.2289397014 -0.032471773 0.011064703 -0.020974955  
## time\_Pend200\_pos 0.4579577856 -0.026346948 -0.040195418 -0.007289112  
## time\_Pend200\_neg -0.4573948466 -0.041556896 0.029381050 -0.009841597  
## PC10 PC11 PC12 PC13  
## Speed\_mean 0.54987288 -0.032107567 0.070014694 0.266209826  
## Speed\_max 0.19768149 -0.276075334 0.045890706 -0.448397831  
## Acel\_pos\_mean 0.06353379 -0.510654057 0.119805068 0.057239683  
## Acel\_pos\_max -0.12405985 0.689683838 -0.179882517 0.077909743  
## Acel\_neg\_mean -0.31945717 -0.319665780 -0.011345662 0.295740545  
## Acel\_neg\_max 0.58162670 0.116478792 0.056912934 -0.282836532  
## Duration -0.28023860 0.003151090 0.004964643 -0.544122005  
## Distance 0.17321998 -0.013566630 -0.048841979 0.479979937  
## time\_ralenti 0.23293590 0.126305455 0.016397986 -0.020297714  
## time\_mov -0.14309921 -0.109051092 -0.023290412 0.133673216  
## Pend200\_pos\_mean 0.01868752 0.042414590 -0.001858869 0.023539130  
## Pend200\_pos\_max 0.01009436 -0.061651862 0.003851117 -0.006015437  
## Pend200\_neg\_mean -0.08476497 0.147286435 0.679820614 0.035888685  
## Pend200\_neg\_max 0.07776980 -0.122898592 -0.690872700 -0.048969753  
## time\_Pend200\_pos 0.02703346 0.003110489 0.008233742 0.038264653  
## time\_Pend200\_neg 0.03318796 0.010975382 -0.012018538 0.028211954  
## PC14 PC15 PC16  
## Speed\_mean 0.3795787599 0.405657643 -1.049283e-03  
## Speed\_max -0.4283151565 -0.371405969 9.853480e-04  
## Acel\_pos\_mean 0.1154425577 0.111464734 -3.296518e-04  
## Acel\_pos\_max -0.0555985509 -0.023935710 1.484087e-04  
## Acel\_neg\_mean -0.0147591998 0.071670111 -2.305016e-04  
## Acel\_neg\_max -0.0008815751 -0.061600819 1.455258e-04  
## Duration 0.2241171979 0.361721425 5.119915e-04  
## Distance -0.1959494483 -0.362109117 -4.119724e-04  
## time\_ralenti -0.4338875858 0.409562076 3.821106e-01  
## time\_mov -0.4673267697 0.359283068 3.660392e-01  
## Pend200\_pos\_mean -0.0028444832 -0.006253134 -3.400216e-05  
## Pend200\_pos\_max -0.0052728896 -0.007622400 8.322500e-06  
## Pend200\_neg\_mean -0.0341950686 -0.028134635 1.221998e-06  
## Pend200\_neg\_max 0.0334611132 0.032629613 9.759342e-06  
## time\_Pend200\_pos -0.2816521573 0.242814061 -5.972703e-01  
## time\_Pend200\_neg -0.2808897738 0.236364069 -6.027207e-01

"solicito las tablas"

## [1] "solicito las tablas"

get\_pca\_var(res.pca)

## Principal Component Analysis Results for variables  
## ===================================================  
## Name Description   
## 1 "$coord" "Coordinates for the variables"   
## 2 "$cor" "Correlations between variables and dimensions"  
## 3 "$cos2" "Cos2 for the variables"   
## 4 "$contrib" "contributions of the variables"

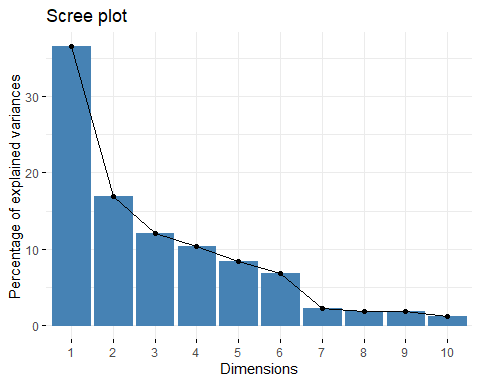
get\_eig(res.pca)#tabla de eigenvalores varianza acumulada

## eigenvalue variance.percent cumulative.variance.percent  
## Dim.1 5.848812e+00 3.655507e+01 36.55507  
## Dim.2 2.699933e+00 1.687458e+01 53.42966  
## Dim.3 1.937172e+00 1.210732e+01 65.53698  
## Dim.4 1.647876e+00 1.029922e+01 75.83620  
## Dim.5 1.345249e+00 8.407806e+00 84.24401  
## Dim.6 1.097240e+00 6.857751e+00 91.10176  
## Dim.7 3.479885e-01 2.174928e+00 93.27669  
## Dim.8 3.013140e-01 1.883212e+00 95.15990  
## Dim.9 2.868164e-01 1.792602e+00 96.95250  
## Dim.10 1.798092e-01 1.123808e+00 98.07631  
## Dim.11 1.136461e-01 7.102879e-01 98.78660  
## Dim.12 8.179752e-02 5.112345e-01 99.29783  
## Dim.13 6.184812e-02 3.865507e-01 99.68439  
## Dim.14 2.770041e-02 1.731276e-01 99.85751  
## Dim.15 2.279697e-02 1.424810e-01 99.99999  
## Dim.16 9.419972e-07 5.887482e-06 100.00000

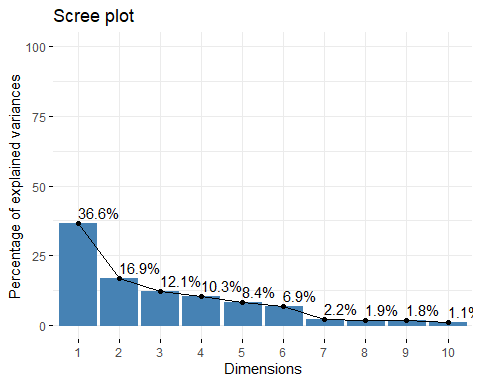
"las gráficas"

## [1] "las gráficas"

fviz\_eig(res.pca)

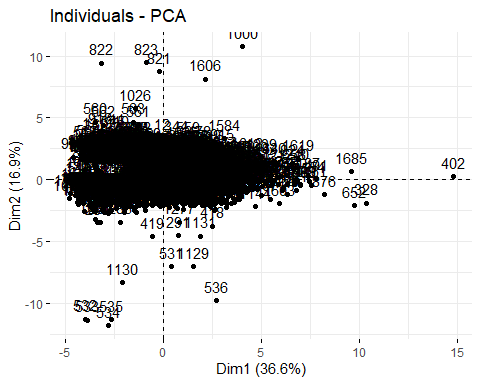


fviz\_screeplot(res.pca, addlabels = TRUE, ylim = c(0, 100))#screeplot

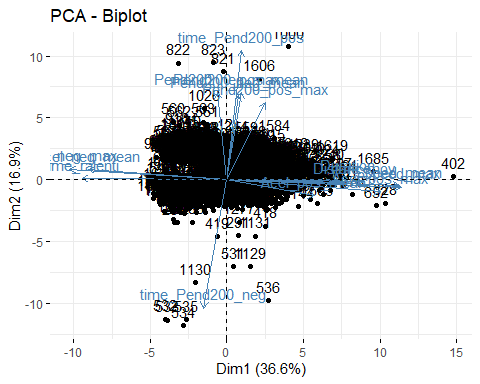


fviz\_pca\_ind(res.pca, repel=F,cex=0.5)#gr?fico individuos y PCA

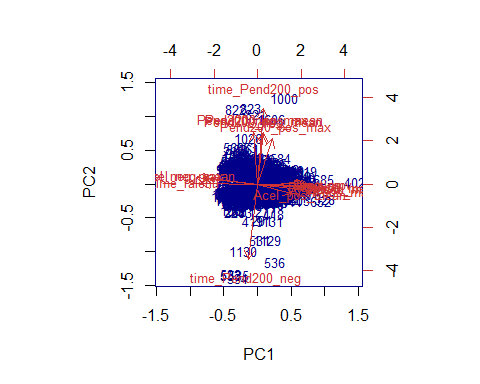
## Warning: Duplicated aesthetics after name standardisation: size



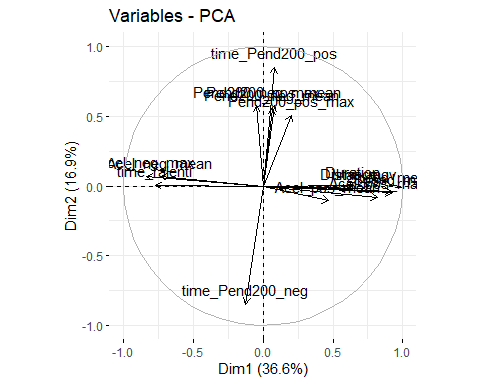
fviz\_pca\_biplot(res.pca) #biplot



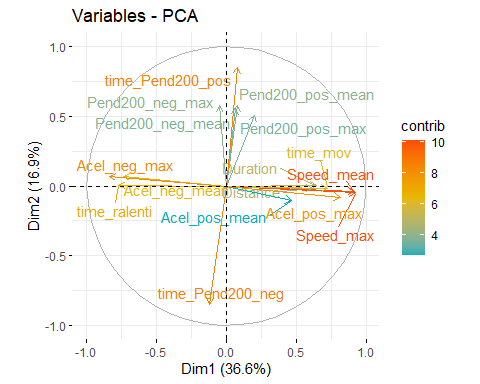
biplot(x = res.pca, scale = 0.5, cex = 0.8,col = c("blue4", "brown3"))



fviz\_pca\_var(res.pca, col.var = "black") #grafico de pesos de los componentes



fviz\_pca\_var(res.pca, col.var="contrib",gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),repel = TRUE)#grafico de pesos de los componentes a color



"Cálculo de los componentes"

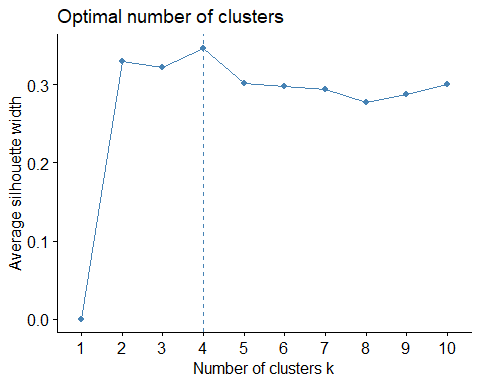
## [1] "Cálculo de los componentes"

pred<-predict(res.pca)  
datos<-cbind(datos,pred)

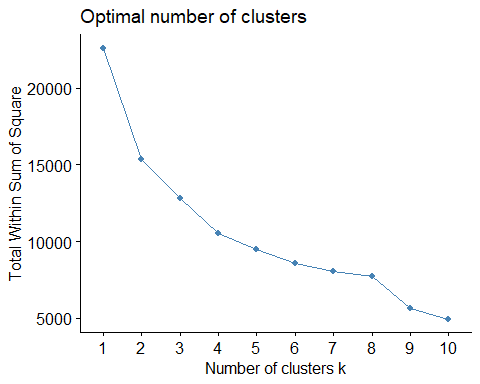
"Para seleccionar el número óptimo de cluster"

## [1] "Para seleccionar el número óptimo de cluster"

comp<-datos[,c(26:29)]  
fviz\_nbclust(comp,kmeans,method = "silhouette")



fviz\_nbclust(comp,kmeans,method = "wss")



"La estimación del cluster jerárquico"

## [1] "La estimación del cluster jerárquico"

res\_HC\_comp<-eclust(comp,"hclust",k=4,graph = TRUE)

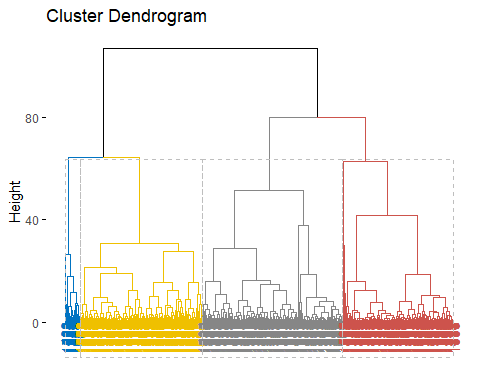
## Registered S3 method overwritten by 'dendextend':  
## method from   
## text.pvclust pvclust

## Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none" instead as  
## of ggplot2 3.3.4.  
## ℹ The deprecated feature was likely used in the factoextra package.  
## Please report the issue at <https://github.com/kassambara/factoextra/issues>.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

"El dendograma"

## [1] "El dendograma"

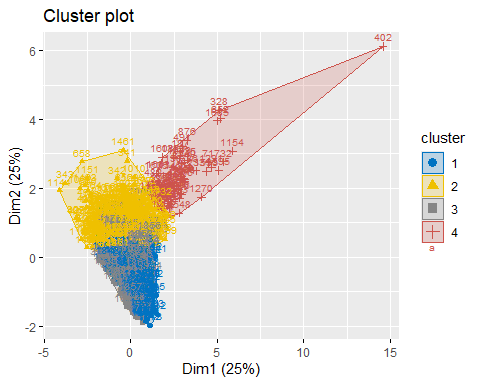
fviz\_dend(res\_HC\_comp,kmeans,palette="jco",rect = TRUE,show\_labels = TRUE)



"los individuos"

## [1] "los individuos"

fviz\_cluster(res\_HC\_comp,ellipse.type = "convex",palette="jco",labelsize = 8)



"para guardar los cluster"

## [1] "para guardar los cluster"

HC4\_comp<-cutree(res\_HC\_comp,k=4)#HAGO K=3 GRUPOS  
datos\_cluster<-cbind(datos,HC4\_comp)  
write.csv2(datos\_cluster, file="datos\_HC4.csv",row.names=T)  
  
datos\_c1 <- datos\_cluster[datos\_cluster$HC4\_comp == "1", c(2:12,24,25)]  
datos\_c2 <- datos\_cluster[datos\_cluster$HC4\_comp == "2", c(2:12,24,25)]  
datos\_c3 <- datos\_cluster[datos\_cluster$HC4\_comp == "3", c(2:12,24,25)]  
datos\_c4 <- datos\_cluster[datos\_cluster$HC4\_comp == "3", c(2:12,24,25)]  
  
resumenC1<-as.table(summary(datos\_c1))   
resumenC2<-as.table(summary(datos\_c2))  
resumenC3<-as.table(summary(datos\_c3))  
resumenC4<-as.table(summary(datos\_c4))  
  
write.table(resumenC1,file="resumenC1.csv")  
write.table(resumenC2,file="resumenC2.csv")  
write.table(resumenC3,file="resumenC3.csv")  
write.table(resumenC4,file="resumenC4.csv")

#ggplot(data = datos, aes(x = t\_ralenti, y = t\_mov, color = as.factor(HC4\_comp))) +  
# geom\_point(size = 6) +  
# theme\_bw() +  
# theme(legend.position = "none")  
  
HC4\_comp <- as.factor(HC4\_comp)  
  
"Dividimos los datos"

## [1] "Dividimos los datos"

attach(datos\_cluster)

## The following object is masked \_by\_ .GlobalEnv:  
##   
## HC4\_comp

## The following objects are masked from datos:  
##   
## Acel\_neg\_max, Acel\_neg\_mean, Acel\_pos\_max, Acel\_pos\_mean, Distance,  
## Driver, Driver\_age, Duration, micro\_ciclo, Pend160\_neg\_max,  
## Pend160\_neg\_mean, Pend160\_pos\_max, Pend200\_neg\_max,  
## Pend200\_neg\_mean, Pend200\_pos\_max, Pend200\_pos\_mean,  
## Social\_stratum, Speed\_max, Speed\_mean, time\_mov, time\_Pend160\_neg,  
## time\_Pend160\_pos, time\_Pend200\_neg, time\_Pend200\_pos, time\_ralenti

head(datos\_cluster)

## micro\_ciclo Driver Speed\_mean Speed\_max Acel\_pos\_mean Acel\_pos\_max  
## 1 1 8 7.053932 18.61637 0.2904402 0.6541200  
## 2 2 8 8.211783 23.51068 0.8515276 1.3009388  
## 3 3 8 21.719606 36.71838 0.2643472 1.3345324  
## 4 4 8 17.959151 34.67187 0.2599726 0.9172304  
## 5 5 8 1.263695 11.10092 0.5285996 0.9130672  
## 6 6 8 14.951410 27.07257 0.3528592 0.7796241  
## Acel\_neg\_mean Acel\_neg\_max Duration Distance time\_ralenti time\_mov  
## 1 -0.30990109 -0.9672719 34 68.57989 38.235294 64.70588  
## 2 -0.39813184 -1.0831680 24 57.02627 45.833333 58.33333  
## 3 -0.29765839 -0.8713056 150 911.01682 4.666667 96.00000  
## 4 -0.26695644 -1.2334983 457 2284.80310 2.188184 98.03063  
## 5 -0.07229099 -0.8906366 49 17.55132 83.673469 18.36735  
## 6 -0.44147062 -1.1471983 71 299.02819 9.859155 91.54930  
## Pend160\_pos\_max Pend160\_neg\_mean Pend160\_neg\_max time\_Pend160\_pos  
## 1 0.00000000 -0.07129972 -0.07129972 0.000000  
## 2 0.00000000 -0.07129972 -0.07129972 0.000000  
## 3 0.00000000 -0.04400980 -0.08555009 0.000000  
## 4 0.02609249 -0.10039211 -0.23371783 6.126915  
## 5 0.00000000 -0.05401604 -0.05401604 0.000000  
## 6 0.03533124 -0.10242632 -0.12336049 26.760563  
## time\_Pend160\_neg Pend200\_pos\_mean Pend200\_pos\_max Pend200\_neg\_mean  
## 1 102.94118 0.00000000 0.00000000 -0.07652013  
## 2 104.16667 0.00000000 0.00000000 -0.07652013  
## 3 100.66667 0.00000000 0.00000000 -0.05241816  
## 4 94.09190 0.01181084 0.01181084 -0.10216823  
## 5 102.04082 0.00000000 0.00000000 -0.06076205  
## 6 74.64789 0.02868728 0.02868728 -0.07375637  
## Pend200\_neg\_max time\_Pend200\_pos time\_Pend200\_neg Social\_stratum Driver\_age  
## 1 -0.07652013 0.000000 102.94118 2 44  
## 2 -0.07652013 0.000000 104.16667 2 44  
## 3 -0.09856547 0.000000 100.66667 2 44  
## 4 -0.20132168 7.877462 92.34136 2 44  
## 5 -0.06076205 0.000000 102.04082 2 44  
## 6 -0.07924286 11.267606 90.14085 2 44  
## PC1 PC2 PC3 PC4 PC5 PC6  
## 1 -1.4221793 -1.596445 0.004848844 -0.2574781 0.6171496 -0.52885561  
## 2 -0.3903591 -1.862065 1.542940300 0.6611131 -0.7740676 -0.53925589  
## 3 0.7298992 -1.665256 -0.488006821 -0.7150379 1.3185021 -0.54910138  
## 4 1.3208531 -1.783771 -2.139106031 -0.9701187 0.7979778 -0.03673671  
## 5 -2.8478848 -1.479174 0.539083509 -0.4255277 -1.9288617 -0.36014854  
## 6 0.2318644 -1.140530 -0.121792313 0.2050089 1.4480430 -0.75211648  
## PC7 PC8 PC9 PC10 PC11 PC12  
## 1 -0.1531494 0.24988665 -0.5755563 -0.07678635 0.009078538 -0.10007269  
## 2 -0.8081998 -0.43855619 0.2100252 0.17461278 -0.345915995 0.01047513  
## 3 0.5104766 -0.12247914 0.4485188 0.02122446 0.412172143 0.01046013  
## 4 -0.4251789 0.07121361 0.1829421 -1.02887672 -0.180582625 0.30351564  
## 5 -0.0775419 -0.08483073 -0.1266192 -0.10144274 -0.079052051 -0.02553035  
## 6 -0.2959691 0.35990597 -0.3840516 -0.06920159 -0.155543776 -0.06309417  
## PC13 PC14 PC15 PC16 HC4\_comp  
## 1 -0.03285959 -0.02429338 -0.14735114 1.033460e-04 1  
## 2 0.04609591 0.11969568 0.05112135 -3.789501e-04 1  
## 3 -0.04519979 0.06111498 0.01196464 1.134422e-04 1  
## 4 -0.64803843 0.34076367 0.45998664 8.522560e-04 2  
## 5 0.08347628 0.15969948 0.06270244 -2.609122e-04 1  
## 6 -0.01667671 0.09369599 -0.05589394 -6.843817e-05 2

set.seed(101)   
sample <- sample.int(n = nrow(datos\_cluster), size = floor(.8\*nrow(datos\_cluster)), replace = F)  
training<- datos\_cluster[sample, ]  
testing <- datos\_cluster[-sample, ]  
  
"Con las variables originales"

## [1] "Con las variables originales"

modelo\_svm\_puros <- svm(formula = HC4\_comp ~ Speed\_mean+Speed\_max+Acel\_pos\_mean+Acel\_pos\_max+Acel\_neg\_mean+ Acel\_neg\_max+Duration+Distance+time\_ralenti+time\_mov+Pend200\_pos\_mean+Pend200\_pos\_max+Pend200\_neg\_mean+ Pend200\_neg\_max+time\_Pend200\_pos+time\_Pend200\_neg  
, data = datos, kernel = "linear", cost = 10, scale = TRUE,datos=training)  
summary(modelo\_svm\_puros)

##   
## Call:  
## svm(formula = HC4\_comp ~ Speed\_mean + Speed\_max + Acel\_pos\_mean +   
## Acel\_pos\_max + Acel\_neg\_mean + Acel\_neg\_max + Duration + Distance +   
## time\_ralenti + time\_mov + Pend200\_pos\_mean + Pend200\_pos\_max +   
## Pend200\_neg\_mean + Pend200\_neg\_max + time\_Pend200\_pos + time\_Pend200\_neg,   
## data = datos, kernel = "linear", cost = 10, datos = training,   
## scale = TRUE)  
##   
##   
## Parameters:  
## SVM-Type: C-classification   
## SVM-Kernel: linear   
## cost: 10   
##   
## Number of Support Vectors: 300  
##   
## ( 45 139 108 8 )  
##   
##   
## Number of Classes: 4   
##   
## Levels:   
## 1 2 3 4

"Con los PCA"

## [1] "Con los PCA"

modelo\_svm\_PCA <- svm(formula = HC4\_comp ~ PC1+PC2+PC3+PC4, data = datos, kernel = "linear", cost = 10, scale = TRUE,datos=training)  
summary(modelo\_svm\_PCA)

##   
## Call:  
## svm(formula = HC4\_comp ~ PC1 + PC2 + PC3 + PC4, data = datos, kernel = "linear",   
## cost = 10, datos = training, scale = TRUE)  
##   
##   
## Parameters:  
## SVM-Type: C-classification   
## SVM-Kernel: linear   
## cost: 10   
##   
## Number of Support Vectors: 325  
##   
## ( 52 146 120 7 )  
##   
##   
## Number of Classes: 4   
##   
## Levels:   
## 1 2 3 4

"Los errores"

## [1] "Los errores"

newdata<-testing  
predicciones\_puros <- predict(modelo\_svm\_puros,newdata)  
pred <- predict(modelo\_svm\_puros, newdata, decision.values = TRUE)  
predpuros<-as.data.frame(predicciones\_puros)   
  
predicciones\_pca <- predict(modelo\_svm\_PCA,newdata, decision.values = TRUE)  
predpca<-as.data.frame(predicciones\_pca)   
  
HC4\_comp\_test<-as.factor(testing$HC4\_comp)  
  
  
pred\_puros<-as.factor(predicciones\_puros)  
pred\_pca<-as.factor(predicciones\_pca)  
matriz\_conf\_puros<-confusionMatrix(HC4\_comp\_test, pred\_puros)  
matriz\_conf\_pca<-confusionMatrix(HC4\_comp\_test, pred\_pca)  
matriz\_conf\_puros

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 1 2 3 4  
## 1 106 1 0 0  
## 2 5 112 10 0  
## 3 0 5 121 0  
## 4 0 0 0 13  
##   
## Overall Statistics  
##   
## Accuracy : 0.9437   
## 95% CI : (0.9152, 0.9648)  
## No Information Rate : 0.3512   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.9181   
##   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: 1 Class: 2 Class: 3 Class: 4  
## Sensitivity 0.9550 0.9492 0.9237 1.00000  
## Specificity 0.9962 0.9412 0.9793 1.00000  
## Pos Pred Value 0.9907 0.8819 0.9603 1.00000  
## Neg Pred Value 0.9812 0.9756 0.9595 1.00000  
## Prevalence 0.2976 0.3164 0.3512 0.03485  
## Detection Rate 0.2842 0.3003 0.3244 0.03485  
## Detection Prevalence 0.2869 0.3405 0.3378 0.03485  
## Balanced Accuracy 0.9756 0.9452 0.9515 1.00000

matriz\_conf\_pca

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 1 2 3 4  
## 1 106 1 0 0  
## 2 5 109 12 1  
## 3 0 7 119 0  
## 4 0 0 0 13  
##   
## Overall Statistics  
##   
## Accuracy : 0.9303   
## 95% CI : (0.8995, 0.954)  
## No Information Rate : 0.3512   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.8987   
##   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: 1 Class: 2 Class: 3 Class: 4  
## Sensitivity 0.9550 0.9316 0.9084 0.92857  
## Specificity 0.9962 0.9297 0.9711 1.00000  
## Pos Pred Value 0.9907 0.8583 0.9444 1.00000  
## Neg Pred Value 0.9812 0.9675 0.9514 0.99722  
## Prevalence 0.2976 0.3137 0.3512 0.03753  
## Detection Rate 0.2842 0.2922 0.3190 0.03485  
## Detection Prevalence 0.2869 0.3405 0.3378 0.03485  
## Balanced Accuracy 0.9756 0.9307 0.9397 0.96429