R Notebook

# Lecture des données

data = read.table("GSE53626\_series\_matrix2.txt", sep='\t', header=T, comment.char="!", row.names=1, na.string="NC")

nationality = as.factor(t(read.table("GSE53626\_series\_matrix2.txt", skip=42, nrows=1))[-1])  
table(nationality)

## nationality  
## nationality: Bulgarian nationality: EastSicilian   
## 18 10   
## nationality: Greek nationality: Moroccan   
## 17 15   
## nationality: Norwegian nationality: Polish   
## 15 11   
## nationality: SouthItalian nationality: Spanish   
## 18 18   
## nationality: Tunisian nationality: UAE   
## 12 14   
## nationality: WestSicilian   
## 10

#names(nationality)  
#names(nationality) = c("Bulgarian", "East-Sicilian", "Greek", "Moroccan", "Norweigian", "Polish","South-Italian", "Spanish", "Tunisian", "UAE", "West-Sicilian")  
length(nationality)

## [1] 158

unlist(nationality)

## [1] nationality: SouthItalian nationality: SouthItalian  
## [3] nationality: SouthItalian nationality: SouthItalian  
## [5] nationality: SouthItalian nationality: SouthItalian  
## [7] nationality: SouthItalian nationality: SouthItalian  
## [9] nationality: SouthItalian nationality: SouthItalian  
## [11] nationality: SouthItalian nationality: SouthItalian  
## [13] nationality: SouthItalian nationality: SouthItalian  
## [15] nationality: SouthItalian nationality: SouthItalian  
## [17] nationality: SouthItalian nationality: SouthItalian  
## [19] nationality: Bulgarian nationality: Bulgarian   
## [21] nationality: Bulgarian nationality: Bulgarian   
## [23] nationality: Bulgarian nationality: Bulgarian   
## [25] nationality: Bulgarian nationality: Bulgarian   
## [27] nationality: Bulgarian nationality: Bulgarian   
## [29] nationality: Bulgarian nationality: Bulgarian   
## [31] nationality: Bulgarian nationality: Bulgarian   
## [33] nationality: Bulgarian nationality: Bulgarian   
## [35] nationality: Bulgarian nationality: Bulgarian   
## [37] nationality: Greek nationality: Greek   
## [39] nationality: Greek nationality: Greek   
## [41] nationality: Greek nationality: Greek   
## [43] nationality: Greek nationality: Greek   
## [45] nationality: Moroccan nationality: Moroccan   
## [47] nationality: Moroccan nationality: Moroccan   
## [49] nationality: Moroccan nationality: Moroccan   
## [51] nationality: Moroccan nationality: Moroccan   
## [53] nationality: Moroccan nationality: Moroccan   
## [55] nationality: Moroccan nationality: Moroccan   
## [57] nationality: Moroccan nationality: Moroccan   
## [59] nationality: Moroccan nationality: Norwegian   
## [61] nationality: Spanish nationality: Norwegian   
## [63] nationality: Norwegian nationality: Norwegian   
## [65] nationality: Norwegian nationality: Norwegian   
## [67] nationality: Norwegian nationality: Norwegian   
## [69] nationality: Norwegian nationality: Norwegian   
## [71] nationality: Norwegian nationality: Norwegian   
## [73] nationality: Norwegian nationality: Spanish   
## [75] nationality: Polish nationality: Polish   
## [77] nationality: Polish nationality: Polish   
## [79] nationality: Polish nationality: Polish   
## [81] nationality: Polish nationality: Polish   
## [83] nationality: Polish nationality: Polish   
## [85] nationality: Polish nationality: Spanish   
## [87] nationality: Spanish nationality: Spanish   
## [89] nationality: Norwegian nationality: Spanish   
## [91] nationality: Norwegian nationality: Spanish   
## [93] nationality: Spanish nationality: Spanish   
## [95] nationality: Spanish nationality: Spanish   
## [97] nationality: Spanish nationality: Spanish   
## [99] nationality: Spanish nationality: Spanish   
## [101] nationality: Spanish nationality: Spanish   
## [103] nationality: Spanish nationality: EastSicilian  
## [105] nationality: EastSicilian nationality: EastSicilian  
## [107] nationality: EastSicilian nationality: EastSicilian  
## [109] nationality: EastSicilian nationality: EastSicilian  
## [111] nationality: EastSicilian nationality: EastSicilian  
## [113] nationality: EastSicilian nationality: WestSicilian  
## [115] nationality: WestSicilian nationality: WestSicilian  
## [117] nationality: WestSicilian nationality: WestSicilian  
## [119] nationality: WestSicilian nationality: WestSicilian  
## [121] nationality: WestSicilian nationality: WestSicilian  
## [123] nationality: WestSicilian nationality: Tunisian   
## [125] nationality: Tunisian nationality: Tunisian   
## [127] nationality: Tunisian nationality: Tunisian   
## [129] nationality: Tunisian nationality: Tunisian   
## [131] nationality: Tunisian nationality: Tunisian   
## [133] nationality: Tunisian nationality: Tunisian   
## [135] nationality: Tunisian nationality: UAE   
## [137] nationality: UAE nationality: UAE   
## [139] nationality: UAE nationality: UAE   
## [141] nationality: UAE nationality: UAE   
## [143] nationality: UAE nationality: UAE   
## [145] nationality: UAE nationality: UAE   
## [147] nationality: UAE nationality: UAE   
## [149] nationality: UAE nationality: Greek   
## [151] nationality: Greek nationality: Greek   
## [153] nationality: Greek nationality: Greek   
## [155] nationality: Greek nationality: Greek   
## [157] nationality: Greek nationality: Greek   
## 11 Levels: nationality: Bulgarian ... nationality: WestSicilian

library(stringr)  
natio = as.factor(unlist(lapply(nationality, str\_sub, 14)))  
table(natio)

## natio  
## Bulgarian EastSicilian Greek Moroccan Norwegian   
## 18 10 17 15 15   
## Polish SouthItalian Spanish Tunisian UAE   
## 11 18 18 12 14   
## WestSicilian   
## 10

sexe = as.factor(t(read.table("GSE53626\_series\_matrix2.txt", skip=41, nrows=1))[-1])  
table(sexe)

## sexe  
## gender: female gender: male   
## 4 154

length(sexe)

## [1] 158

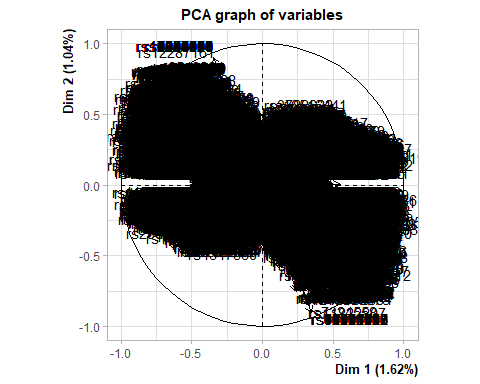
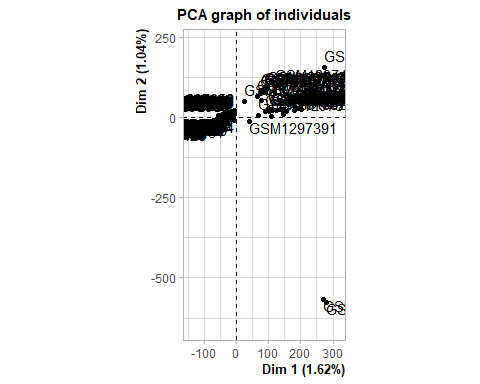
# Gestion des données manquantes

data2 = data[rowSums(is.na(data)) == 0, ]

# ACP

library(FactoMineR)

res.pca = PCA(X = t(data2))



library(factoextra)

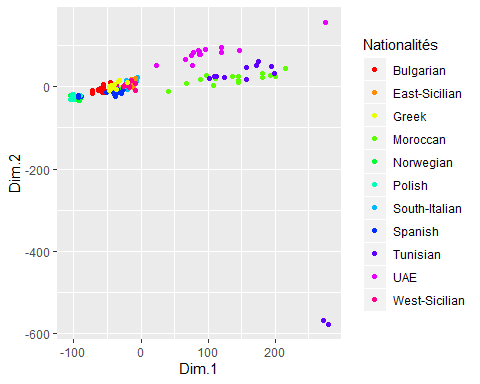
## Loading required package: ggplot2

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

ind = get\_pca\_ind(res.pca)

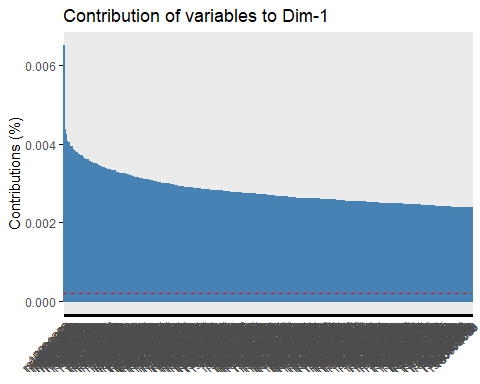
data\_ind = ind$coord[,1:2]  
data\_ind = cbind(data\_ind, natio)

library(ggplot2)  
  
data\_ind = as.data.frame(data\_ind)  
ggplot(data\_ind, aes(x = data\_ind[,1],y = data\_ind[,2], colour = factor(data\_ind[,3]))) + geom\_point() + scale\_color\_manual(name="Nationalités", labels = c("Bulgarian", "East-Sicilian", "Greek", "Moroccan", "Norwegian", "Polish", "South-Italian", "Spanish", "Tunisian", "UAE", "West-Sicilian"), values = rainbow(11)) + xlab("Dim.1") + ylab("Dim.2")

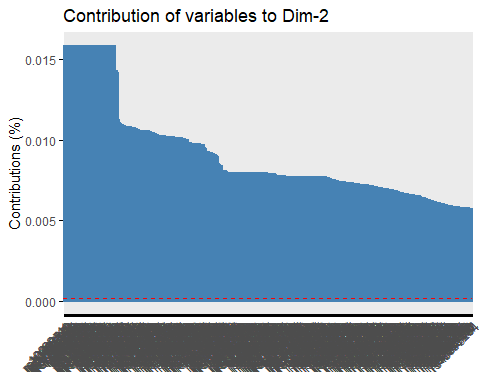


var <- get\_pca\_var(res.pca)

# Contributions des variables à PC1  
fviz\_contrib(res.pca, choice = "var", axes = 1, top = 500)



# Contributions des variables à PC2  
fviz\_contrib(res.pca, choice = "var", axes = 2, top = 500)



# Clustering: k-means

res.kmeans = kmeans(x = t(data2), centers = 11)

library(cowplot)

##   
## \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

## Note: As of version 1.0.0, cowplot does not change the

## default ggplot2 theme anymore. To recover the previous

## behavior, execute:  
## theme\_set(theme\_cowplot())

## \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

plot\_grid(  
ggplot(data\_ind, aes(x = data\_ind[,1],y = data\_ind[,2])) + geom\_point(colour = factor(data\_ind[,3])) + xlab("Dim.1") + ylab("Dim.2") + ggtitle("Groupes nationality"),  
ggplot(data\_ind, aes(x = data\_ind[,1],y = data\_ind[,2])) + geom\_point(colour = factor(res.kmeans$cluster)) + xlab("Dim.1") + ylab("Dim.2") + ggtitle("Groupes kmeans"))

