**Comandos no Rstudio para Cluster Não Hierárquico**

banco2 <- banco [, c(6,12:24,37:41,47:52)] #criação de um banco de dados somente com as perguntas selecionadas

banco2 <- na.omit(banco2) #exclusão dos dados faltantes

pacotes: #carregar pacotes necessários

library(FactoMineR)

library(factoextra)

library(cluster)

k\_means<- kmeans(banco2, 3) #criar um conjunto com o kmeans igual a 3

fviz\_cluster(k\_means, data=banco2) #gerar o gráfico com os agrupamentos gerados

fviz\_nbclust(banco2, kmeans, method ="gap\_stat") #gerar número ótimo de clusters com “gap\_stat”

fviz\_nbclust(banco2, kmeans, method = "silhouette") #gerar número ótimo de clusters com o método silhueta.

fviz\_nbclust(banco2, kmeans, method = "wss") #gerar o número ótimo de clusters com método do cotovelo.

banco3<-merge(banco2,k\_means$cluster,all.x=TRUE) #junto o banco 2 com as perguntas que eu selecionei com os clusters gerados em um novo conjunto de dados. OBS: Está dando erro! As linhas saem de 1865 para 3.478.255 (não era para ficar o mesmo número?)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Cluster Hierárquico**

banco1<- banco[, c(6,12:24,37:41,47:52)] #crio um novo banco com as perguntas selecionadas #

banco1<- na.omit(banco1) #retiro os dados faltantes do novo banco

d1 <- dist(banco1, method = "euclidean") #calculo a distância euclidiana

hc <- hclust(d1, method = "complete" ) #crio os clusters

plot(hc) #gero o dendograma

Últimos comandos usados:

k\_means<- kmeans(banco2, 3)

grupos<-k\_means$cluster

grupos<-data.frame(grupos)

banco3<-merge(banco2,grupos$grupos,by="row.names",all.x=TRUE)

colnames(paises)[1]<-""