

```

# LEITURA da BASE de DADOS
Pontos_geog<- read.csv(".....Seu Caminho...../101pontos.csv", sep = ";", header = T)
str(Pontos_geog)

# ----- Agrupamento com k-MÉDIAS (K- MEANS) -----
k<-5 # Montagem de 5 Clusters
Mod_clusters = kmeans(Pontos_geog[,2:3],k) # Está usando apenas as Colunas 2 e 3 da Base de Dados (corresponde a "X" e "Y")

# Mostra todos os Resultados (Componentes) disponíveis da Clusterização
Mod_clusters

# Mostra todos os Resultados (Componentes) Individuais
Mod_clusters$cluster # ACESSO ao Número do CLUSTER de cada PONTO
Mod_clusters$centers # ACESSO aos CENTROIDES dos CLUSTERS
Mod_clusters$totss # ACESSO à Soma de Quadrados (TSS...SS=Square Sum)
Mod_clusters$withinss # ACESSO às Somas de Quadrados das Variações Internas nos Clusters (WSS...SS=Square Sum)
Mod_clusters$betweenss # ACESSO às Somas de Quadrados das Variações entre os Clusters (BSS...SS=Square Sum)
Mod_clusters$size # ACESSO aos tamanhos dos Clusters
Mod_clusters$iter # ACESSO ao No. de Iterações desenvolvidas até os Resultados
Mod_clusters$ifault # ACESSO ao

# EXPORTA Nos dos CLUSTERS para BASE DE DADOS
# Cria novo campo na Base de Pontos Geográficos com o No. do Cluster
Pontos_geog$cluster <- Mod_clusters$cluster

str(Pontos_geog) # Estrutura da BASE de DADOS com Nos. dos CLUSTERS de CADA PONTO
dim(Pontos_geog) # Dimensão da BASE de DADOS com Nos. dos CLUSTERS de CADA PONTO
View(Pontos_geog) # Visualiza BASE de DADOS com Nos. dos CLUSTERS de CADA PONTO

# Grava em arquivo nova Base de Dados, com o No. do Cluster
write.csv(Pontos_geog, "C:/Users/aaval/OneDrive/Documentos/SCRIPTS_R/Ciencia de Dados/Scripts - Clusters/101pontos_C.csv")
Pontos_geog_C<- read.csv("C:/Users/aaval/OneDrive/Documentos/SCRIPTS_R/Ciencia de Dados/Scripts - Clusters/101pontos_C.csv", sep = ",", header = T)
str(Pontos_geog_C)
# OBS: Note que foi criada uma coluna a mais de ID... fica sem uso
# Já há a coluna 2, "P", para Identificar o No. do Ponto ( de 0 a 100)
head(Pontos_geog_C)
View(Pontos_geog_C)

# Plota os Pontos com cores diferentes para os Clusters
# OBS: Parâmetro col=color...muda de cor, conforme o No. do Cluster
plot(Pontos_geog_C[,2:3], col=Pontos_geog_C[,6],
     xlab="--- X ---",ylab="--- Y ---")

# Plota os Centroides
plot(Mod_clusters$centers)

# Plota os Pontos de cada Cluster
plot(Pontos_geog_C[,6], Pontos_geog_C[,2],
     xlab="--- 3 do Cluster ---",ylab="--- # do Ponto ---")

# ----- Determinação do No. "Ideal" de Clusters (valor de K) -----
# ----- O Valor "Ideal" de K é baseado em "WSS" -----
# WSS = Somas dos Quadrados das Variações Internas nos Clusters

# ***** GRAFICO em COTOVELO (ELBOW) - MODELO 1 *****
# Monta Clusters para varios valores de K (K de 1 a Kmax)
MeusDados <- Pontos_geog
# Soma das Variâncias das Colunas x (n-1)
WSS <- (nrow(MeusDados)-1)*sum(apply(MeusDados,2,var))
WSS

# Detalhamento do Cálculo de WSS.....
nrow(MeusDados)-1 # No. de linhas do dataset menos 1 (n-1)
apply(MeusDados,2,var) # 1 indicates Rows, 2 indicates Columns / var = Variance / Monta variancias das Colunas
sum(apply(MeusDados,2,var)) # Soma variancias das Colunas
?apply
?var

# roda k-means com K=1..Kmax e guarda "withinss"
Kmax=20
for (i in 2:Kmax) WSS[i] <- sum(kmeans(MeusDados, centers=i)$withinss)

# Apresenta o gráfico de WSS x No. de Clusters
# O No. "ideal" de Clusters se situa no ponto em que a curva começa a se estabilizar
# A estabilização se dá no "cotovelo" = "elbow"
plot(1:Kmax, WSS, type="b", xlab="--- # de Clusters ---",ylab="--- WSS ---")
# Note que na função <plot()> faz-se "1:Kmax", K variando de 1 a Kmax
# No eixo horizontal aparecerão valores varuando entre 1 e Kmax

```