

Segmentação: exemplo chocolate

Adilson dos Anjos

Segmentação

Pacotes utilizados

```
library(FactoMineR)
library(readxl)
library(graphics)
library(cluster)
library(fpc)
```

Dados sobre Chocolate

```
choco<-read_excel('chocolate.xls')  
head(choco)
```

```
# A tibble: 6 x 7
```

	Amostra	Marca1A	Marca1L	Marca2A	Marca2L	Marca3A	Marca3L
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	C1	7	9	4	6	5	7
2	C2	7	9	2	6	1	7
3	C3	6	8	3	3	3	5
4	C4	8	9	6	6	4	7
5	C5	4	5	3	3	5	9
6	C6	5	7	6	7	8	9

- ▶ Vamos utilizar apenas as notas atribuídas a cada marca por cada consumidor;
- ▶ No arquivo, as colunas 2:7 correspondem a essas notas:

```
choco<-choco[,2:7]
```

Obtendo a matriz de distâncias euclidianas

- ▶ No R, esse é o tipo de objeto necessário para se obter os agrupamentos:

```
d <- dist(as.matrix(choco))  
head(d)
```

```
[1] 4.472136 4.472136 2.449490 6.244998 5.099020 5.477226
```

Obtenção do agrupamento pelo Método hierárquico:

- ▶ Vamos utilizar o método de Ward:

```
hc1 <- hclust(d, method = 'ward.D' ) # Método de Ward
hc1
```

Call:

```
hclust(d = d, method = "ward.D")
```

Cluster method : ward.D

Distance : euclidean

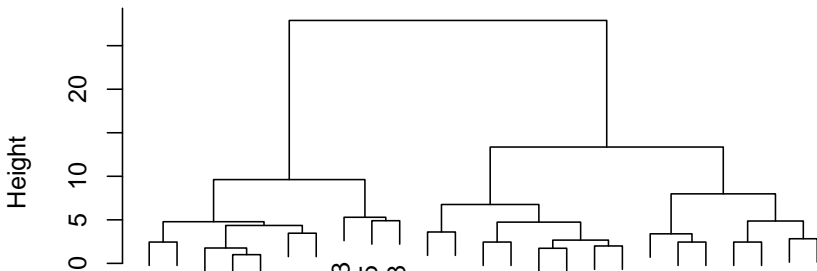
Number of objects: 25

Obtenção do Dendrograma:

- ▶ Graficamente, pode-se obter o resultado do agrupamento por meio de um dendrograma:
- ▶

```
plot(hc1,main= "Dendrograma - Dados dos chocolates",  
     xlab="Consumidores")
```

Dendrograma – Dados dos chocolates

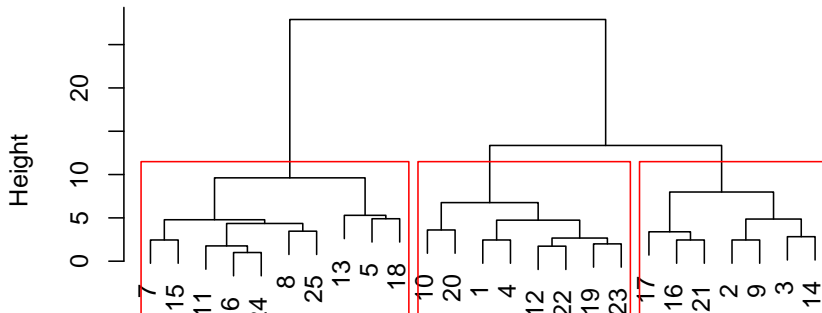


Identificando os grupos formados

- Definimos 3 grupos:

```
plot(hc1, main= "Dendrograma - Dados dos chocolates",
      xlab="consumidores", sub="Ward - Dist. Euclidiana")
rect.hclust(hc1, k = 3, border = "red")
```

Dendrograma – Dados dos chocolates



Definindo um ponto de corte para formar grupos:

- ▶ Aqui é possível obter os consumidores de cada agrupamento:

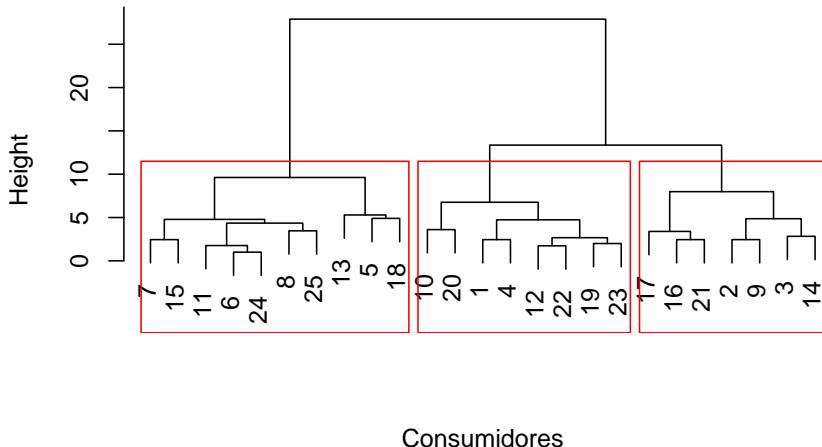
```
groups <- cutree(hc1, k=3)  
groups
```

```
[1] 1 2 2 1 3 3 3 3 2 1 3 1 3 2 3 2 2 3 1 1 2 1 1 3 3
```

- ▶ Como seria o gráfico com 3 grupos:

```
plot(hc1, main= "Dendrograma - Dados dos chocolates", xlab=
rect.hclust(hc1, k=3, border="red")
```

Dendrograma - Dados dos chocolates



Método K médias

```
cl <- kmeans(choco, 3)  
cl$cluster
```

```
[1] 1 1 1 1 2 2 2 2 1 3 2 3 2 1 2 3 1 2 3 3 1 3 3 2 2
```

Centro do cluster para cada variável/atributo.

- ▶ Observe as médias para cada agrupamento:

```
cl$centers
```

	Marca1A	Marca1L	Marca2A	Marca2L	Marca3A	Marca3L
1	7.25	8.875000	3.875000	4.750000	3.625000	6.500000
2	4.90	6.000000	5.400000	5.700000	6.300000	8.700000
3	7.00	8.714286	5.857143	5.857143	6.571429	5.571429

Padronização das variáveis (opcional)

- ▶ Utilizar quando a escala das variáveis for diferente.

```
choco.p<-scale(choco)
```

k-medias

```
cl <- kmeans(choco.p, 2)
```

Tamanho do cluster

- ▶ Quantos indivíduos em cada cluster:

```
cl$size
```

```
[1] 15 10
```


Mostra o centróide

- ▶ A nota média (observe que a nota foi padronizada) de cada agrupamento para cada variável:

```
cl$centers
```

	Marca1A	Marca1L	Marca2A	Marca2L	Marca3A	M
1	-0.2390022	-0.2906867	0.6862981	0.5498491	0.4104465	0
2	0.3585033	0.4360301	-1.0294472	-0.8247736	-0.6156697	-0

Mostra os grupos

- ▶ Onde cada indivíduo foi localizado:

```
cl$cluster
```

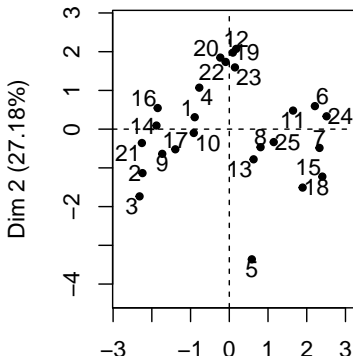
```
[1] 2 2 2 1 2 1 1 1 2 2 1 1 1 2 1 2 2 1 1 1 2 1 1 1 1
```

Utilizando o método das Componentes Principais:

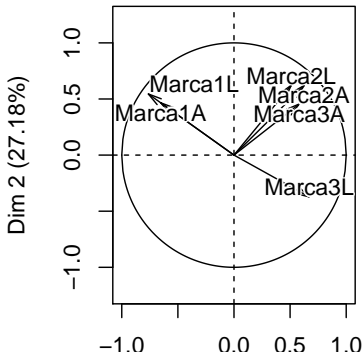
- Qual sua percepção de agrupamento?

```
par(mfrow=c(1,2))
choco.pca<-PCA(choco)
```

Individuals factor map (PCA)



Variables factor map (PCA)

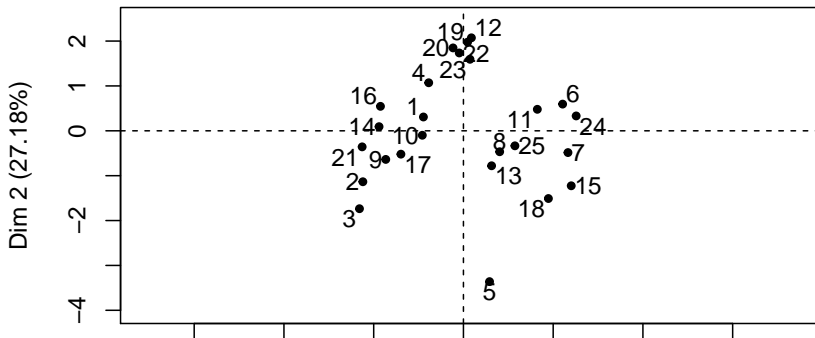


Gerando os gráficos individuais

- ▶ Consumidores:

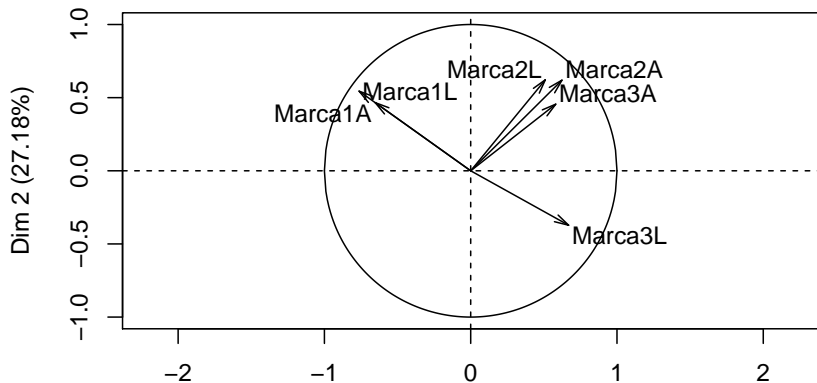
```
par(mfrow=c(1,1))  
plot(choco.pca,choix=c('ind'))
```

Individuals factor map (PCA)



► Marcas

```
par(mfrow=c(1,1))  
plot(choco.pca,choix=c('var'))
```

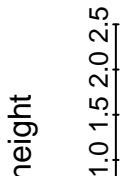
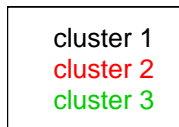
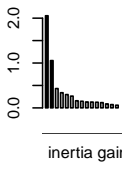
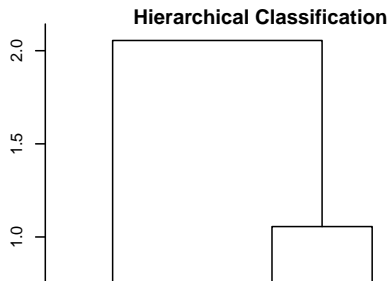
Variables factor map (PCA)

- ▶ O argumento `nb.clust=-1` utiliza o número de cluster sugerido pelo programa:

```
par(mfrow=c(1,1))
choco.hpc<-HCPC(choco.pca,nb.clust=-1)
```

Hierarchical cluster

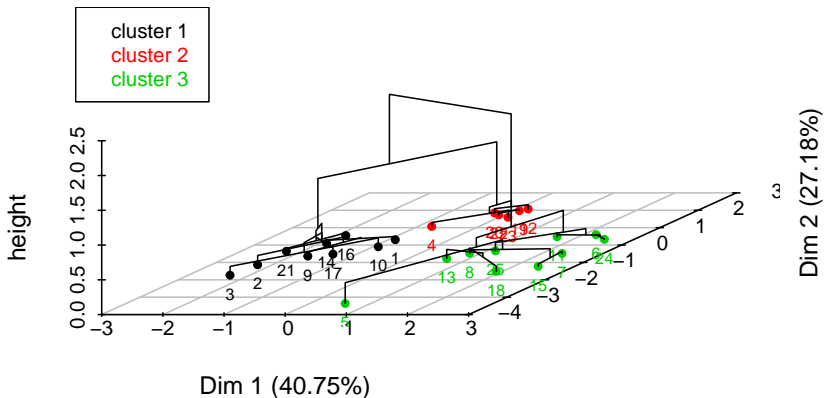
Hierarchical Clustering



Outra visualização gráfica:

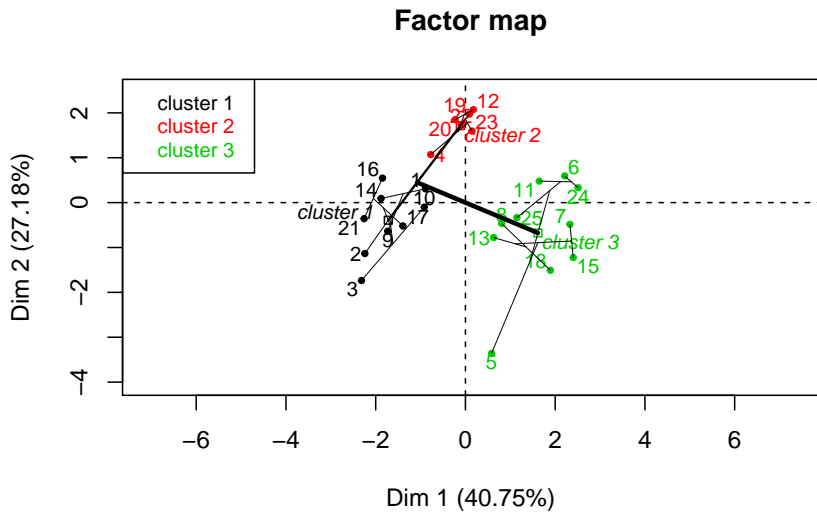
```
plot(choco.hpc)
```

Hierarchical clustering on the factor map



Mais uma visualização gráfica:

```
plot(choco.hpc, choice= "map")
```



Número de observações por grupo

```
table(choco.hpc$data.clust$clus)
```

```
1  2  3  
9  6 10
```

Descrição dos grupos

```
choco.hpc$desc.var
```

```
$quanti.var
```

	Eta2	P-value
Marca1L	0.7189962	8.626491e-07
Marca2A	0.6614140	6.704392e-06
Marca3L	0.6443551	1.151285e-05
Marca1A	0.5112325	3.802989e-04
Marca2L	0.3276846	1.268538e-02
Marca3A	0.2520279	4.099579e-02

```
$quanti
```

```
$quanti$`1`
```

	v.test	Mean in category	Overall mean	sd in category
Marca1L	2.493042	8.777778	7.68	0.4157
Marca1A	2.091269	7.111111	6.24	1.1967

Outras descrições

- ▶ Consumidores típicos de cada cluster:

```
choco.hpc$desc.ind$para
```

```
Cluster: 1
```

```
      14          9          1          21          17  
0.8061739 1.3514587 1.3683779 1.5623851 1.6197491  
-----
```

```
Cluster: 2
```

```
      22          23          19          12          4  
0.4627331 0.8777349 1.0237452 1.1859265 1.4611663  
-----
```

```
Cluster: 3
```

```
      7          8          24          6          15  
0.7945292 1.2349760 1.4506663 1.5046516 1.5099218
```

para -> indivíduo típico

- ▶ o indivíduo 14 é típico do cluster 1

- ▶ Consumidores mais distantes de cada cluster:

```
choco.hpc$desc.ind$dist
```

```
Cluster: 1
```

```
      2      3      21      9      17  
4.282977 4.206401 3.503739 3.077185 2.938673
```

```
-----  
Cluster: 2
```

```
      12      19      20      22      23  
3.418109 3.160583 3.082706 2.854715 2.784526
```

```
-----  
Cluster: 3
```

```
      18      15      5      7      24  
4.006121 4.001038 3.838864 3.306021 3.135607
```

dist -> distância do indivíduo e o centro dos outros clusters

- ▶ o indivíduo mais distante do cluster 1 é o 2

