

Segmentação: exemplo snacks (Lawless, 2013)

Adilson dos Anjos

Segmentação

Pacotes utilizados

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

Dados sobre *snacks*

```
snack<-read_excel('snack.xls')  
head(snack)
```

```
# A tibble: 6 x 12
```

	snack	A	B	C	D	E	F	G	
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Peanuts	8	7	9	4	4	3	3	
2	Beer nuts	8	8	8	4	4	4	4	
3	Almonds	7	8	7	3	3	4	3	
4	Popcorn	4	4	3	7	8	7	4	
5	Cheese corn	3	3	4	8	7	9	3	
6	Caramel corn	3	3	4	6	7	8	4	

```
# ... with 1 more variables: K <dbl>
```

- ▶ Vamos utilizar apenas as notas atribuídas a cada marca por cada consumidor;
- ▶ No arquivo, as colunas 2:12 correspondem a essas notas:

```
snack<-snack[,2:12]
```

Obtendo a matriz de distâncias euclidianas

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

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

```
[1] 3.162278 3.316625 10.246951 11.489125 10.344080 10.19
```

Obtenção do agrupamento pelo Método hierárquico:

- Vamos utilizar o método de Ward:

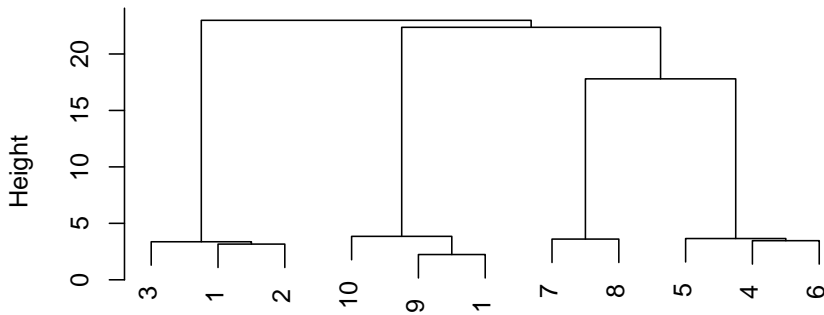
```
hc1 <- hclust(d, method = 'ward.D' ) # Método de Ward
```

Obtenção do Dendrograma:

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

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

Dendrograma – Dados dos snacks

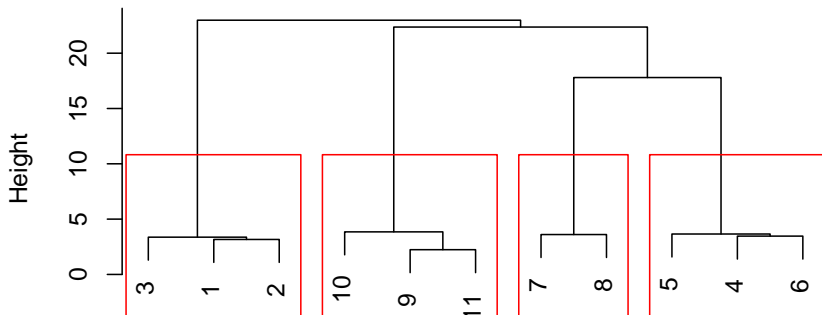


Identificando os grupos formados

- Definimos 3 grupos:

```
plot(hc1, main= "Dendrograma - Dados dos snacks",  
     xlab="Consumidores", sub="Ward - Dist. Euclidiana")  
rect.hclust(hc1, k = 4, border = "red")
```

Dendrograma – Dados dos snacks



Definindo um ponto de corte para formar grupos:

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

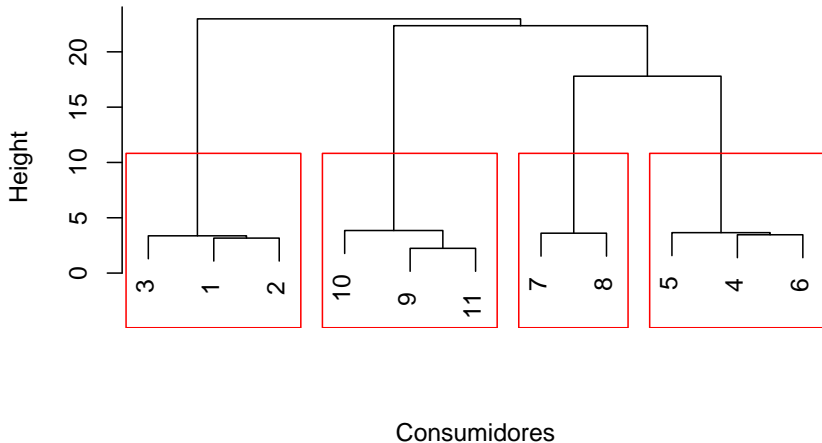
```
groups <- cutree(hc1, k=4)  
groups
```

```
[1] 1 1 1 2 2 2 3 3 4 4 4
```

- Como seria o gráfico com 3 grupos:

```
plot(hc1, main= "Dendrograma - Dados dos snacks", xlab="Consumidores",  
rect.hclust(hc1, k=4, border="red"))
```

Dendrograma - Dados dos snacks



Método K médias

```
cl <- kmeans(snack, 4)
cl$cluster
```

```
[1] 1 1 1 2 2 2 3 3 4 4 4
```

Centro do cluster para cada variável/atributo.

- Observe as médias para cada agrupamento:

```
round(cl$centers,1)
```

	A	B	C	D	E	F	G	H	I	J	K
1	7.7	7.7	8.0	3.7	3.7	3.7	3.3	3.7	4.0	4.0	4.0
2	3.3	3.3	3.7	7.0	7.3	8.0	3.7	4.3	4.0	4.0	4.3
3	5.0	5.0	4.0	4.0	3.5	4.5	9.0	8.0	3.5	3.5	4.5
4	5.0	4.7	3.7	4.0	3.3	5.0	4.0	4.3	8.0	7.7	8.7

Padronização das variáveis (opcional)

- Utilizar quando a escala das variáveis for diferente.

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

k-medias

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

Tamanho do cluster

- ▶ Quantos indivíduos em cada cluster:

```
cl$size
```

```
[1] 5 6
```


Mostra o centróide

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

```
cl$centers
```

	A	B	C	D	E	
1	-0.7094635	-0.6644425	-0.5233594	0.6390058	0.6732108	0
2	0.5912196	0.5537021	0.4361328	-0.5325049	-0.5610090	-0
	G	H	I	J	K	
1	0.5281326	0.5905962	-0.5720776	-0.5772986	-0.4986308	-0
2	-0.4401105	-0.4921635	0.4767313	0.4810821	0.4155257	0

Mostra os grupos

- ▶ Onde cada indivíduo foi localizado:

```
cl$cluster
```

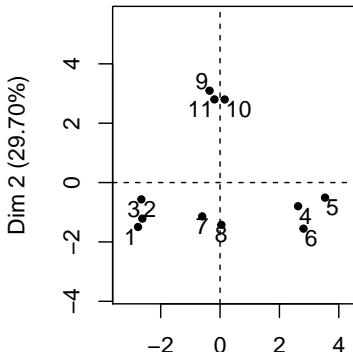
```
[1] 2 2 2 1 1 1 1 1 2 2 2
```

Utilizando o método das Componentes Principais:

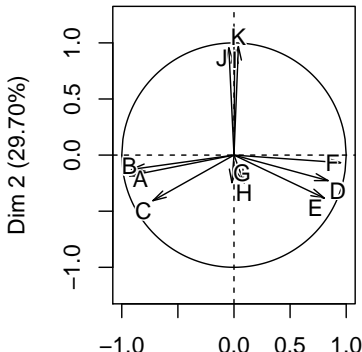
- Qual sua percepção de agrupamento?

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

Individuals factor map (PCA)



Variables factor map (PCA)

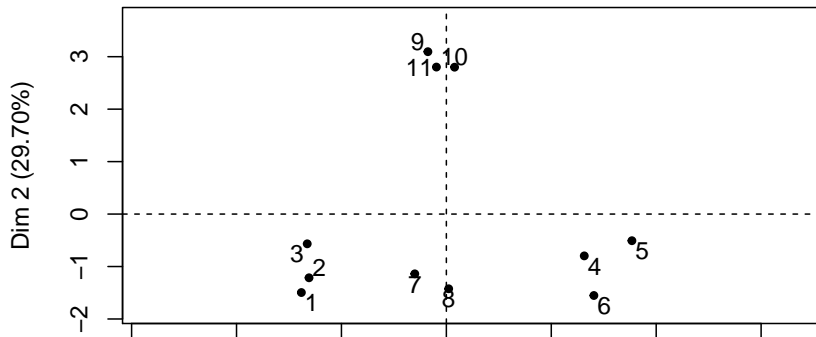


Gerando os gráficos individuais

- Consumidores:

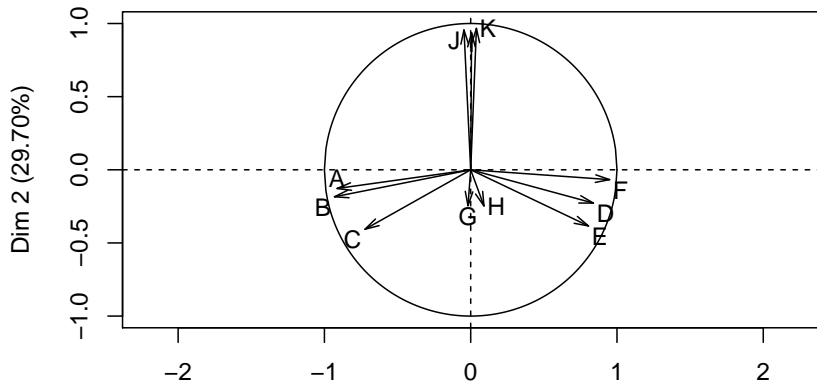
```
par(mfrow=c(1,1))  
plot(snack.pca,choix=c('ind'))
```

Individuals factor map (PCA)



► Snacks

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

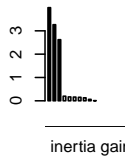
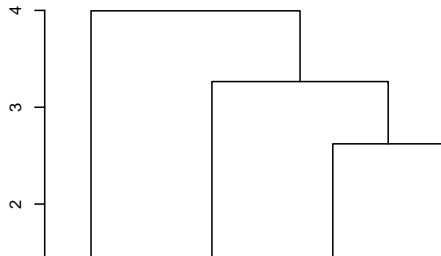
Variables factor map (PCA)

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

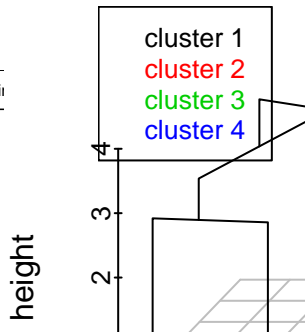
```
par(mfrow=c(1,1))  
snack.hpc<-HCPC(snack.pca,nb.clust=-1)
```

Hierarchical Clustering

Hierarchical Classification



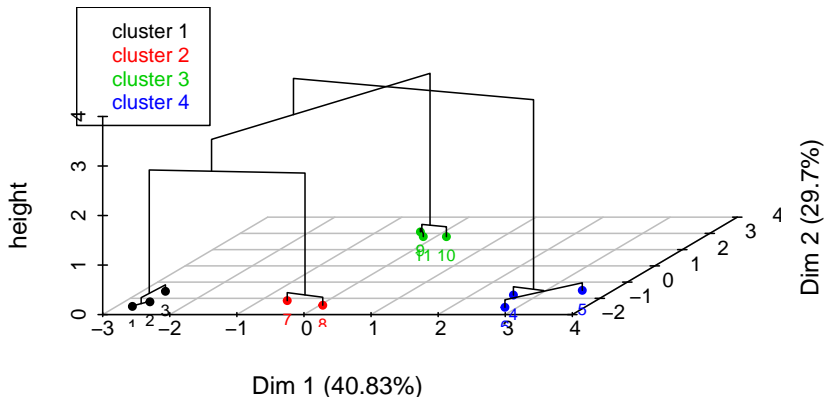
Hierarchical cluster



Outra visualização gráfica:

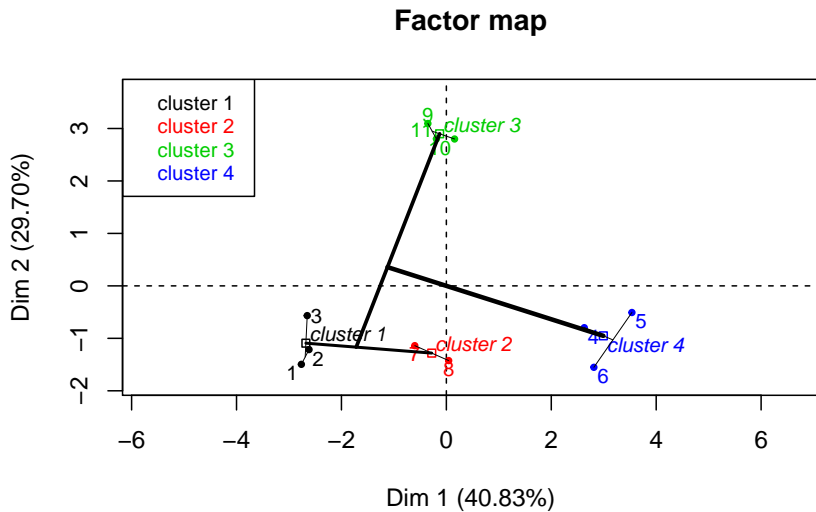
```
plot(snack.hpc)
```

Hierarchical clustering on the factor map



Mais uma visualização gráfica:

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



Número de observações por grupo

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

1 2 3 4

3 2 3 3

Descrição dos grupos

```
snack.hpc$desc.var
```

```
$quanti.var
```

	Eta2	P-value
G	0.9725343	7.908558e-06
K	0.9590108	3.193928e-05
B	0.9367816	1.442175e-04
E	0.9280105	2.264346e-04
H	0.9276316	2.305972e-04
F	0.9083333	5.232198e-04
A	0.8964218	7.983635e-04
C	0.8812416	1.280188e-03
J	0.8600164	2.255378e-03
I	0.8522727	2.713838e-03
D	0.7634409	1.353688e-02

Outras descrições

- ▶ Consumidores típicos de cada cluster:

```
snack.hpc$desc.ind$para
```

Cluster: 1

1	2	3
0.4936791	0.5882471	0.9789369

Cluster: 2

8	7
0.9419796	0.9419796

Cluster: 3

9	11	10
0.3502113	1.1431038	1.2080832

para -> indivíduo típico

- ▶ a marca 1 é típica do cluster 1

- ▶ Marcas mais distantes de cada cluster:

```
snack.hpc$desc.ind$dist
```

Cluster: 1

2	1	3
5.095999	5.031082	4.491306

Cluster: 2

7	8
5.128505	4.726944

Cluster: 3

11	10	9
5.011286	4.986001	4.954981

Cluster: 4

5	4	6
5.255316	4.874167	4.604772

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

- ▶ a marca mais distante do centro do cluster 1 é a 2.

