

shopping_analysis.R

ra016120

Fri Feb 19 17:11:28 2016

```
## Carregando o conjunto de dados
shopping <- read.csv('shopping.csv')
shopping <- shopping[,-1]
```

```
## Análise descritiva
summary(shopping)
```

```
##      Diversão      Orçamento      Comer_Fora      Melhor_Compra
## Min.      :1.000    Min.      :2.000    Min.      :1.000    Min.      :2.000
## 1st Qu.:2.000    1st Qu.:3.000    1st Qu.:2.000    1st Qu.:3.000
## Median :4.000    Median :4.000    Median :4.000    Median :4.000
## Mean   :4.107    Mean   :3.929    Mean   :4.214    Mean   :3.929
## 3rd Qu.:6.000    3rd Qu.:5.000    3rd Qu.:6.000    3rd Qu.:5.000
## Max.   :7.000    Max.   :7.000    Max.   :7.000    Max.   :7.000
## Não_Importa  Compara_Preço      Renda      Visitas
## Min.      :1.000    Min.      :2.00    Min.      :25000    Min.      :0.000
## 1st Qu.:2.000    1st Qu.:3.00    1st Qu.:30000    1st Qu.:2.000
## Median :3.000    Median :4.00    Median :50000    Median :3.000
## Mean   :3.321    Mean   :4.25    Mean   :47857    Mean   :3.179
## 3rd Qu.:4.250    3rd Qu.:4.25    3rd Qu.:60000    3rd Qu.:4.000
## Max.   :7.000    Max.   :7.00    Max.   :80000    Max.   :7.000
```

```
## Obtendo a média por variável
apply(shopping, 2, mean)
```

```
##      Diversão      Orçamento      Comer_Fora      Melhor_Compra      Não_Importa
##      4.107143      3.928571      4.214286      3.928571      3.321429
## Compara_Preço      Renda      Visitas
##      4.250000  47857.142857      3.178571
```

```
## Mudando a escala
pmatrix <- scale(shopping)
pcenter <- attr(pmatrix, "scaled:center")
pscale <- attr(pmatrix, "scaled:scale")
```

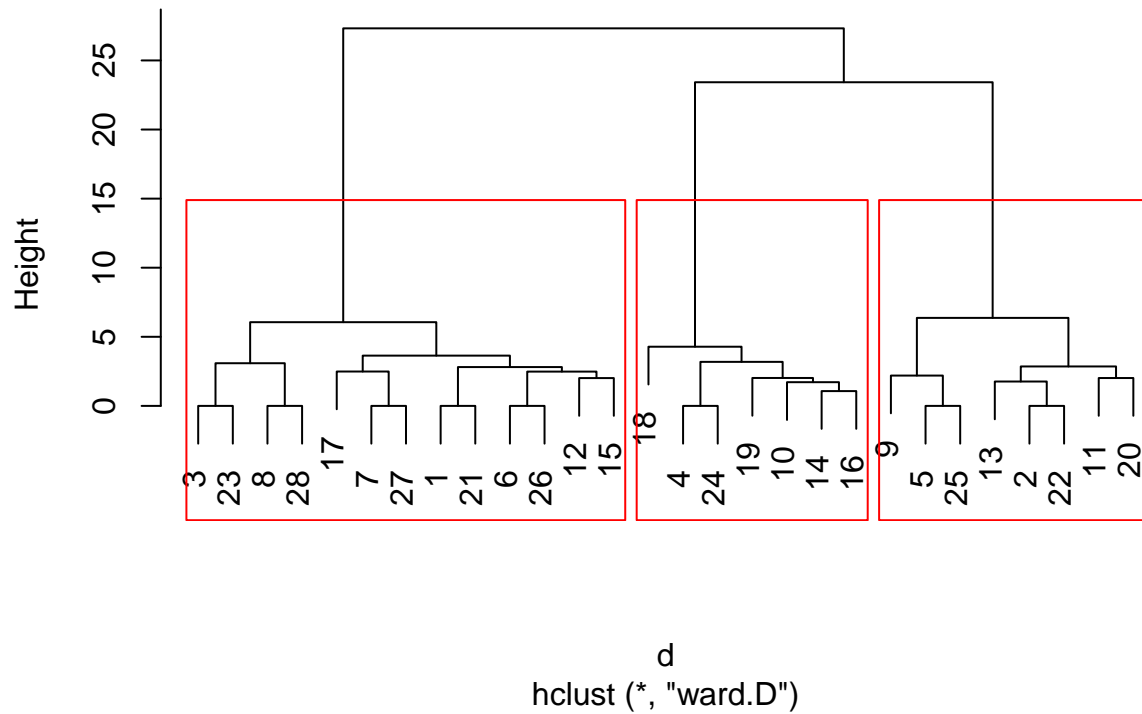
```
## Criando o cluster
d <- dist(pmatrix, method="euclidean")
pfit <- hclust(d, method="ward")
```

```
## The "ward" method has been renamed to "ward.D"; note new "ward.D2"
```

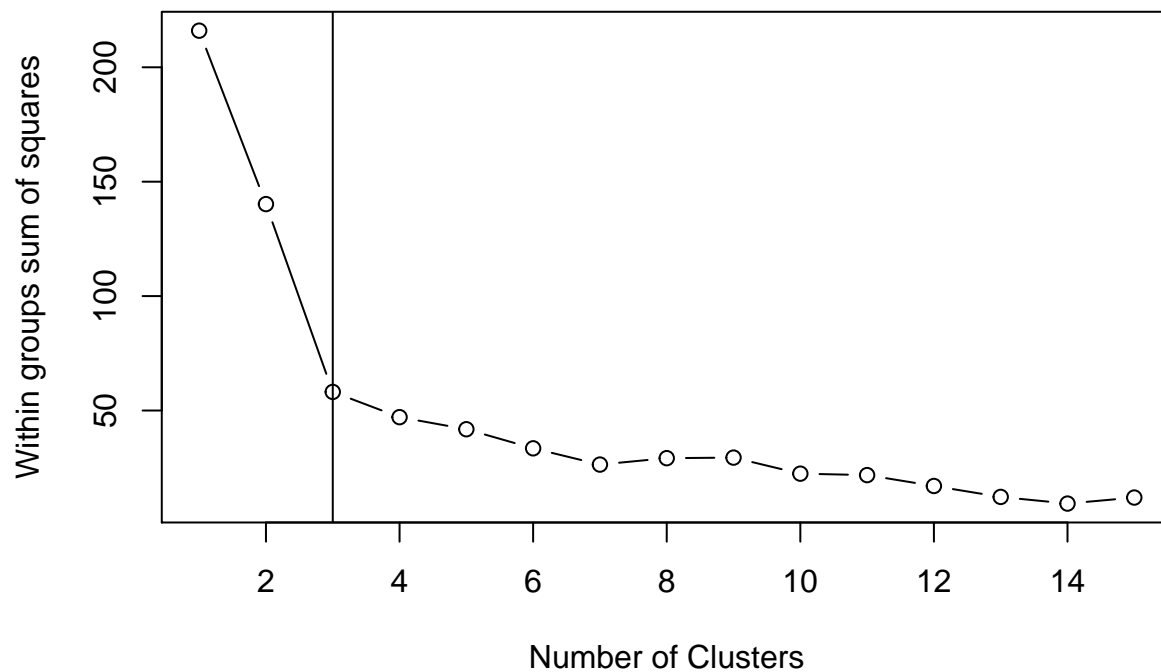
```
plot(pfit)
```

```
## Desenhando a separação dos clusters
rect.hclust(pfit, k=3)
```

Cluster Dendrogram



```
## K-means pelo método do cotovelo
## da proteína
mydata <- pmatrix
wss <- (nrow(mydata)-1)*sum(apply(mydata,2,var))
for (i in 2:15) wss[i] <- sum(kmeans(mydata,
                                   centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Number of Clusters",
     ylab="Within groups sum of squares")
abline(v=3)
```



```
## Pegando os cluster do kmeans
clusters <- kmeans(pmatrix, centers = 3)

## Juntando a informação do cluster
shopping$grupos <- factor(clusters$cluster)

## Perfil das variáveis por grupo
shopping_segments <- split(shopping[, -9], shopping$grupos)
sapply(shopping_segments, function(x) apply(x, 2, mean))
```

##	1	2	3
## Diversão	5.923077	1.625	3.571429
## Orçamento	3.461538	3.000	5.857143
## Comer_Fora	6.153846	1.750	3.428571
## Melhor_Compra	3.230769	3.375	5.857143
## Não_Importa	1.923077	5.500	3.428571
## Compara_Preço	3.769231	3.500	6.000000
## Renda	60000.000000	43125.000	30714.285714
## Visitas	3.153846	1.000	5.714286