

# Annexes

Yann Trividic

## Annexe 1

```
# le sous-arbre de la classe 2 est extrait
cars.hc.ward.cut4.tree2 <- cut(as.dendrogram(cars.hc.ward), h = 13)$lower[[3]]

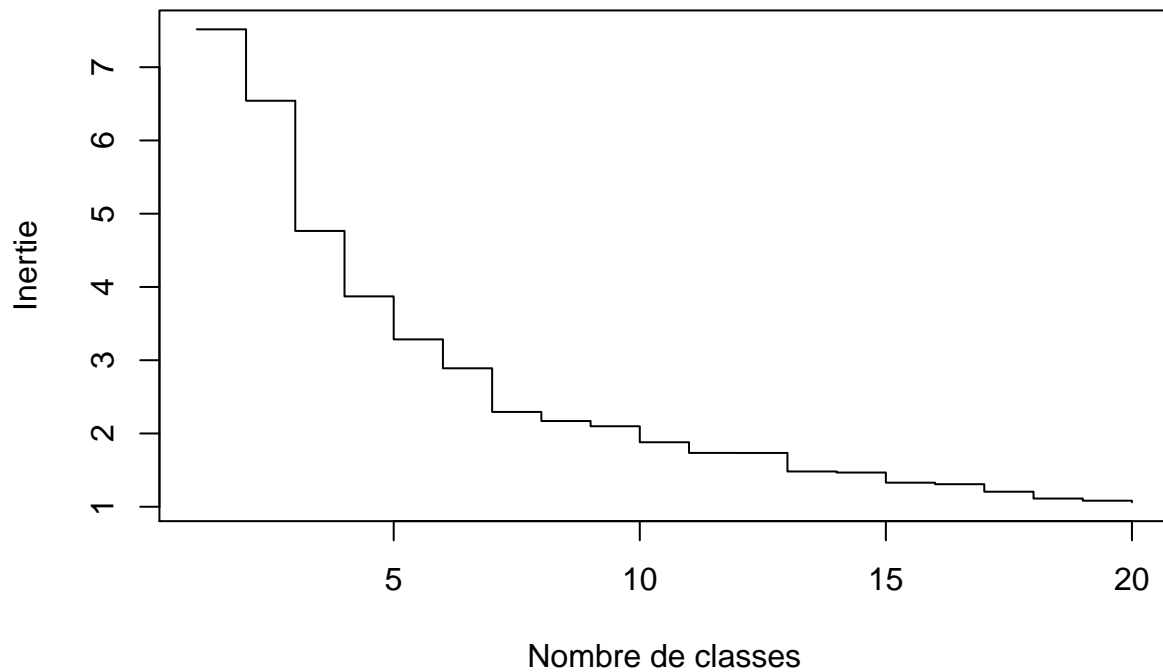
# puis on extrait les indices
library(dendextend)
cars.hc.ward.cut4.tree2.ind <- as.numeric(names(cutree(cars.hc.ward.cut4.tree2, k=1)))

dist = dist(cars.scaled[cars.hc.ward.cut4.tree2.ind, -8])
cah.ward = hclust(dist, method="ward.D2")

inertia = sort(cah.ward$height, decreasing = T)

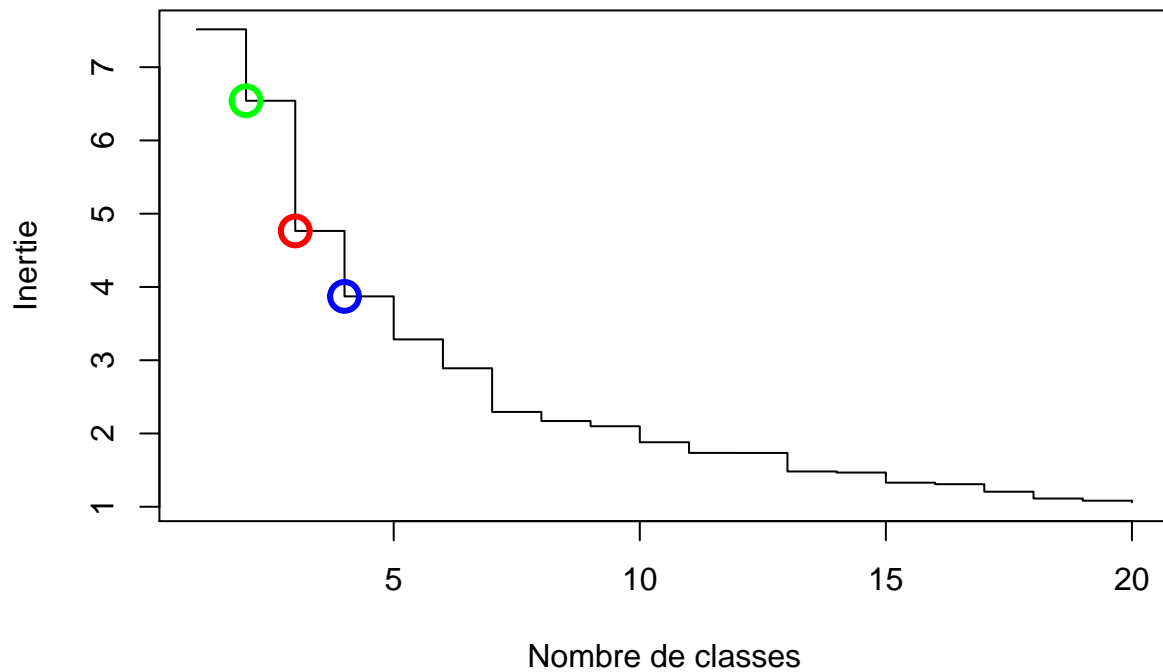
plot(inertia[1:20], type = "s",
      xlab = "Nombre de classes",
      ylab = "Inertie",
      main = "Inertie du dendrogramme en fonction du nombre de classes")
```

## Inertie du dendrogramme en fonction du nombre de classes



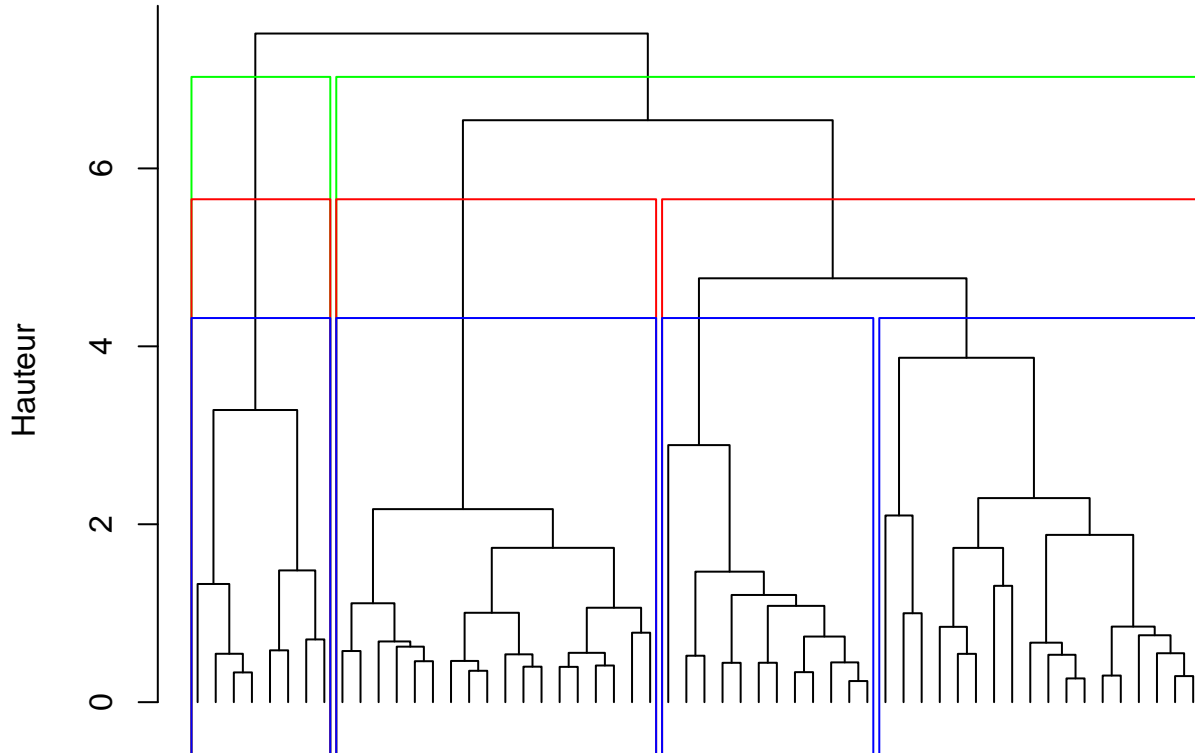
```
# les points d'intérêts sont 2, 3 et 4
plot(inertia[1:20], type = "s",
      xlab = "Nombre de classes",
      ylab = "Inertie",
      main = "Inertie du dendrogramme avec plus grands gains entourés")
points(c(2, 3, 4), inertia[c(2, 3, 4)], col = c("green", "red", "blue"), cex = 2, lwd = 3)
```

## Inertie du dendrogramme avec plus grands gains entourés



```
# affichage du sous-arbre
par(mar=c(1,4.5,2,0))
plot(cah.ward, labels = FALSE, main = "Partitions en 2, 3 ou 4 classes", xlab = "", ylab = "Hauteur", s
rect.hclust(cah.ward, 2, border = "green")
rect.hclust(cah.ward, 3, border = "red")
rect.hclust(cah.ward, 4, border = "blue")
```

## Partitions en 2, 3 ou 4 classes



```
# affichage des tables de contingence
cut_affiche_table_contingence(cah.ward, c(2, 3, 4), cars$brand[cars.hc.ward.cut4.tree2.ind])
```

```
##
## Cut = 2
## Valeurs absolues :
##      1 2
## Europe 7 4
## Japan 18 4
## US     23 0
## tot    48 8
##
## Valeurs relatives :
##      1      2
## Europe 12.50000 7.142857
## Japan 32.14286 7.142857
## US 41.07143 0.000000
## tot 85.71429 14.285714
##
## Cut = 3
## Valeurs absolues :
##      1 2 3
## Europe 4 3 4
## Japan 8 10 4
## US 6 17 0
## tot 18 30 8
```

```
##
## Valeurs relatives :
##      1      2      3
## Europe  7.142857  5.357143  7.142857
## Japan   14.285714 17.857143  7.142857
## US      10.714286 30.357143  0.000000
## tot     32.142857 53.571429 14.285714
##
## Cut = 4
## Valeurs absolues :
##      1  2  3  4
## Europe  4  3  4  0
## Japan   8  8  4  2
## US      6  7  0 10
## tot     18 18  8 12
##
## Valeurs relatives :
##      1      2      3      4
## Europe  7.142857  5.357143  7.142857  0.000000
## Japan   14.285714 14.285714  7.142857  3.571429
## US      10.714286 12.500000  0.000000 17.857143
## tot     32.142857 32.142857 14.285714 21.428571

# on remarque que le partitionnement en deux classes n'est pas suffisant, et que celui à quatre
# classes est trop. On remarque aussi que la distinction europe-japon est faible.

# combinaison des marques pour le partitionnement en 3 classes :
cah.ward.cut3 <- cutree(cah.ward, 3)
afficher_table_contingence_clusters(cars.combined.brand[cars.hc.ward.cut4.tree2.ind], cah.ward.cut3)

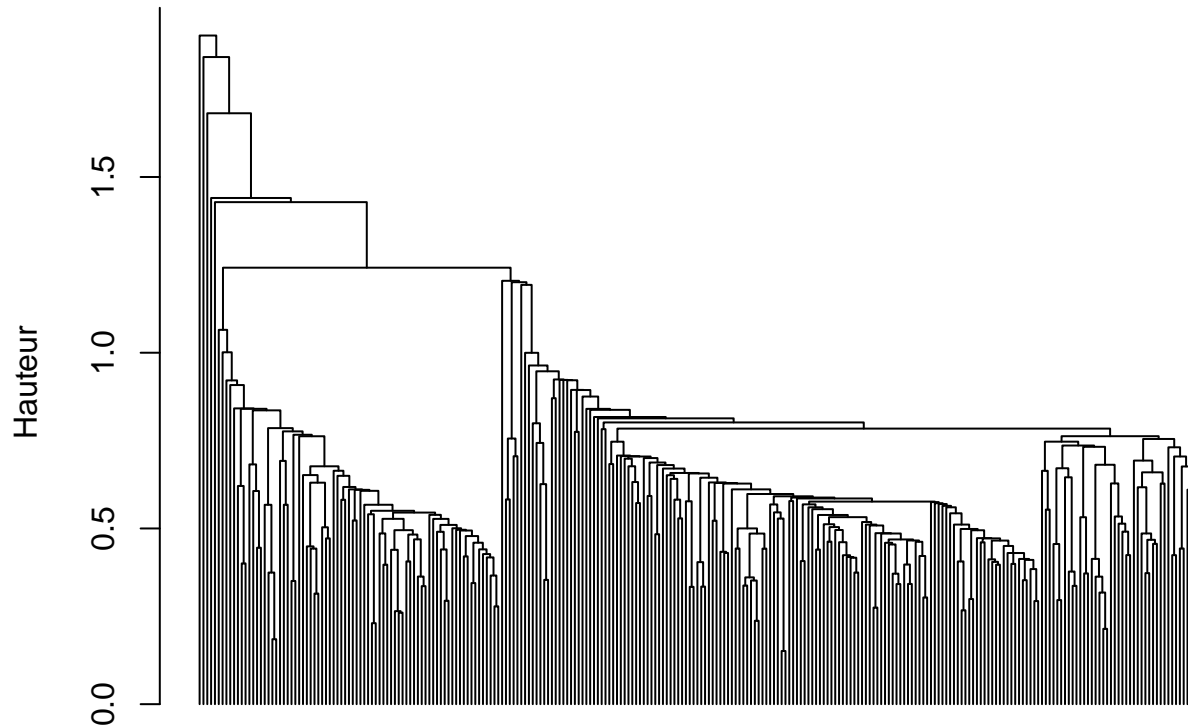
## Valeurs absolues :
##      1  2  3
## Non-US 12 13  8
## US      6 17  0
## tot     18 30  8
##
## Valeurs relatives :
##      1      2      3
## Non-US 21.42857 23.21429 14.28571
## US      10.71429 30.35714  0.00000
## tot     32.14286 53.57143 14.28571

# les résultats obtenus permettent de constater une amélioration du partitionnement,
# mais celui-ci n'est pas assez significatif pour être conservé
```

## Annexe 2

```
cars.hc.min = hclust(cars.dist, method="single")
par(mar=c(1,4.5,2,0))
plot(cars.hc.min, labels = FALSE, main = "Dendrogramme suite à la CAH avec lien minimum", xlab = "", ylab = "")
```

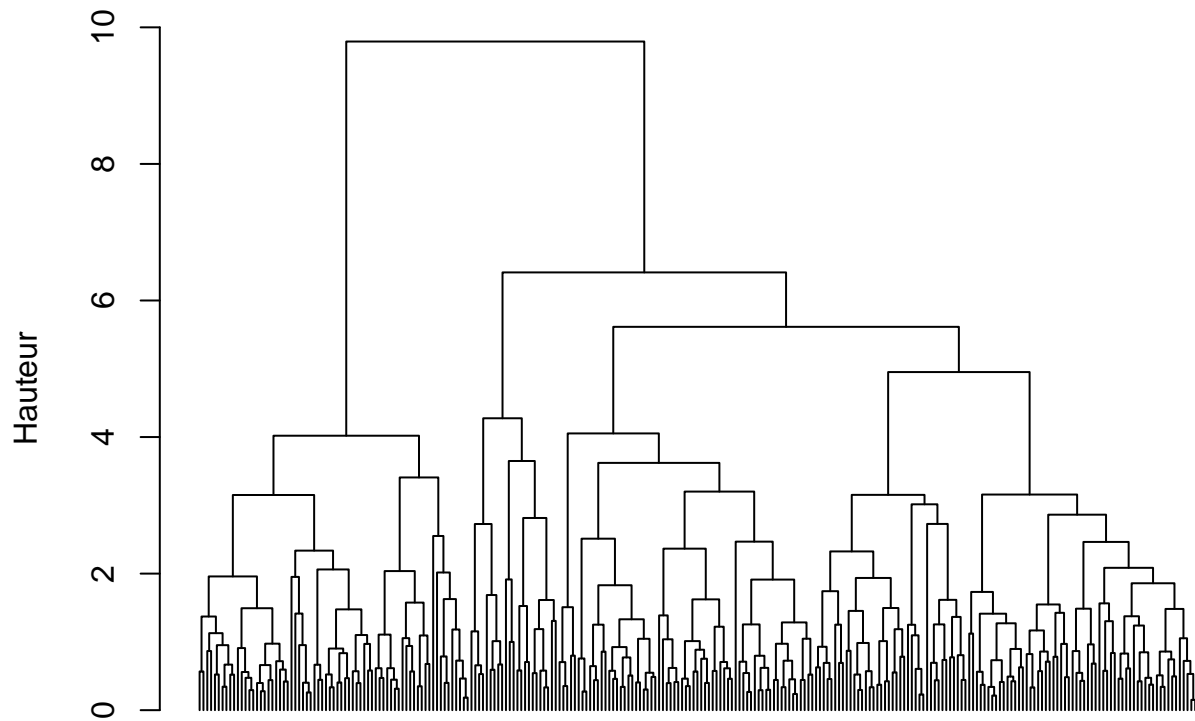
## Dendrogramme suite à la CAH avec lien minimum



### Annexe 3

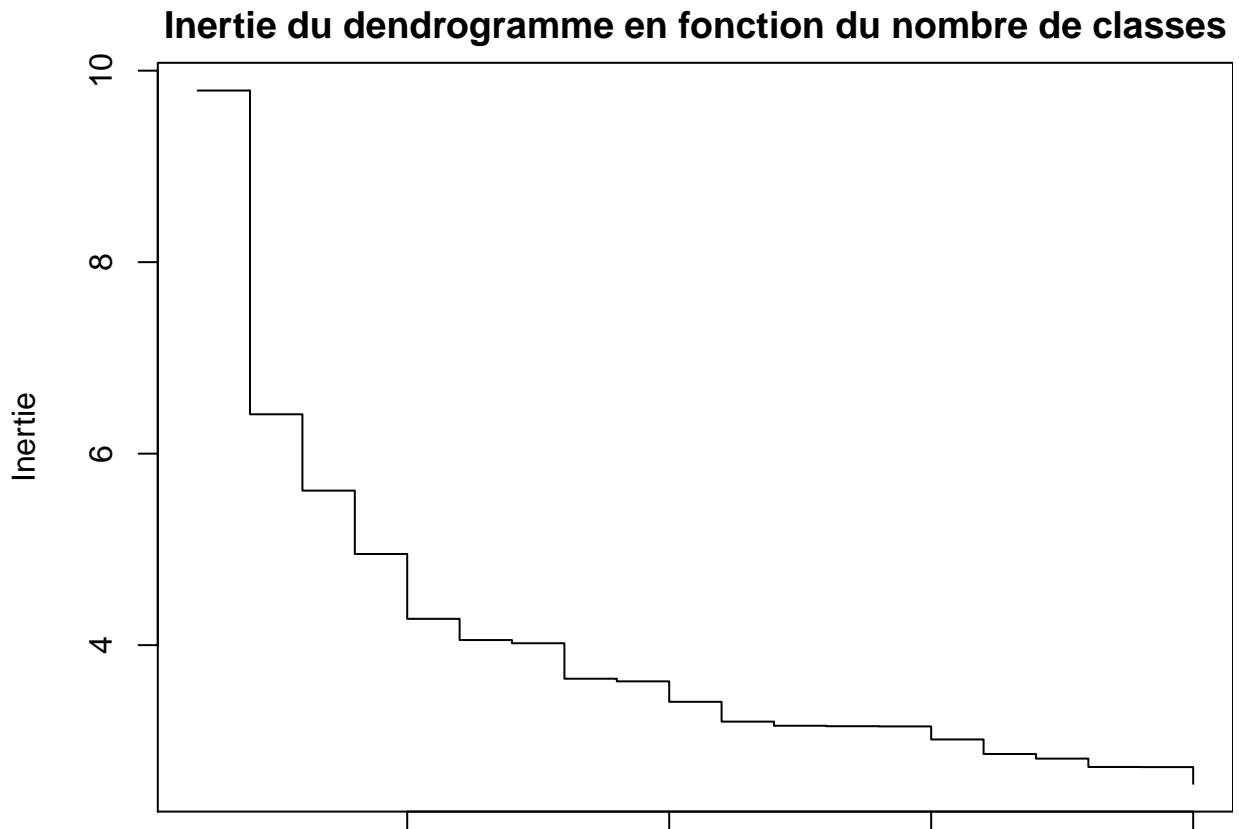
```
cars.hc.max = hclust(cars.dist, method="complete")
par(mar=c(1,4.5,2,0))
plot(cars.hc.max, labels = FALSE, main = "Dendrogramme suite à la CAH avec lien maximum", xlab = "", ylab = "Hauteur")
```

## Dendrogramme suite à la CAH avec lien maximum



```
cars.hc.max.inertia = sort(cars.hc.max$height, decreasing = T)

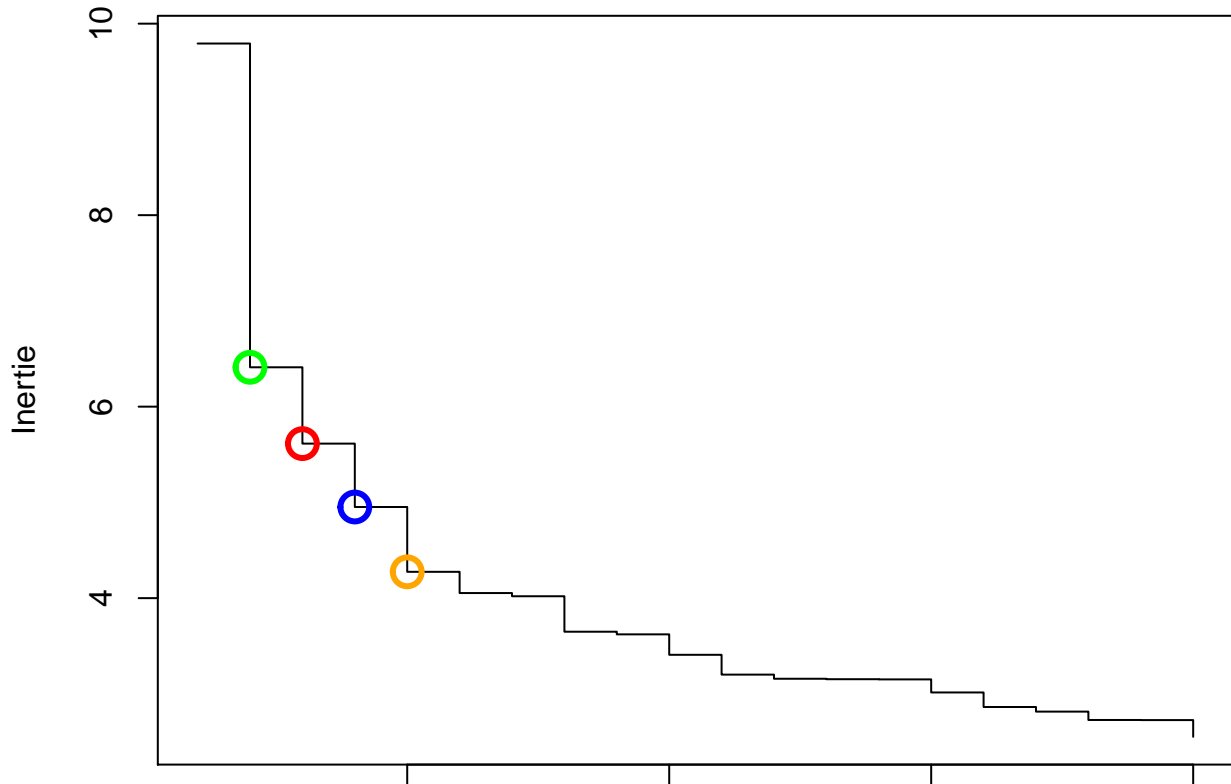
plot(cars.hc.max.inertia[1:20], type = "s",
     xlab = "Nombre de classes",
     ylab = "Inertie",
     main = "Inertie du dendrogramme en fonction du nombre de classes")
```



```
# On remarque des points d'intérêt à 2, 3 et 4
plot(cars.hc.max.inertia[1:20], type = "s",
     xlab = "Nombre de classes",
     ylab = "Inertie",
     main = "Inertie du dendrogramme avec plus grands gains entourés")
points(c(2, 3, 4, 5), cars.hc.max.inertia[c(2, 3, 4, 5)], col = c("green", "red", "blue", "orange"), ce
```

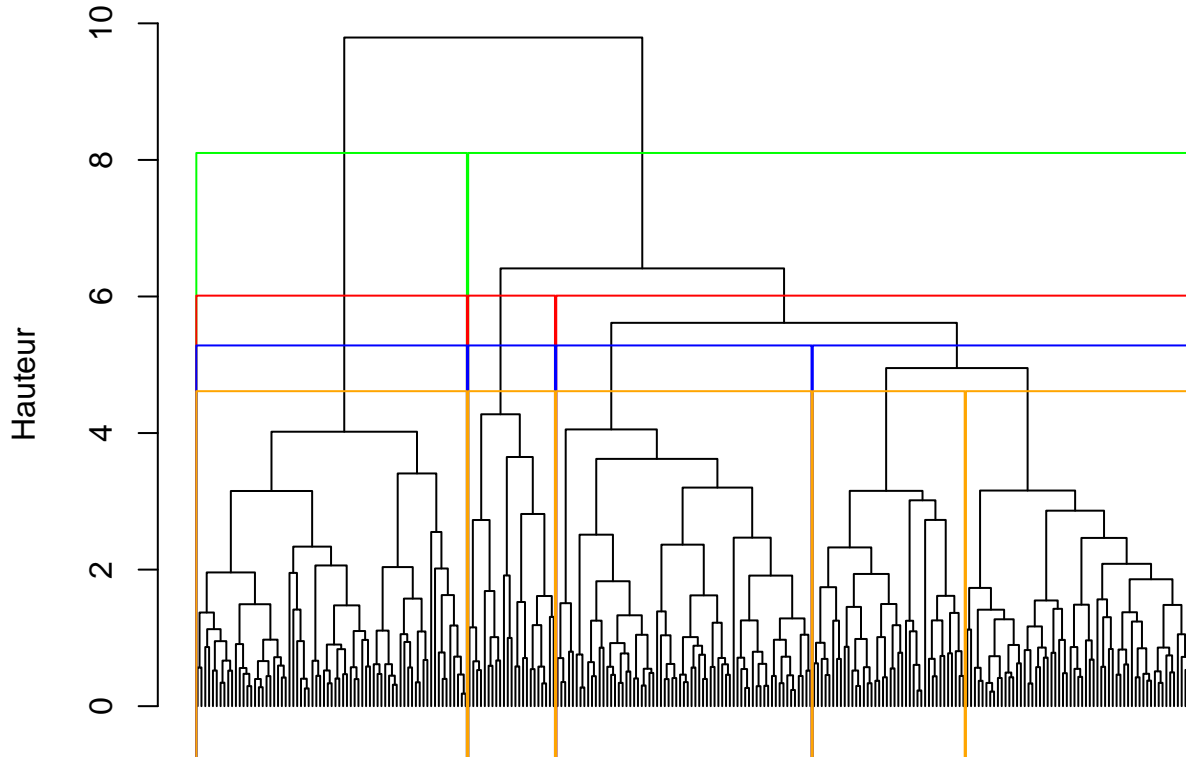


## Inertie du dendrogramme avec plus grands gains entourés



```
# affichage
par(mar=c(1,4.5,2,0))
plot(cars.hc.max, labels = FALSE, main = "Partitions en 2, 3, 4 ou 5 classes", xlab = "", ylab = "Hauteur")
rect.hclust(cars.hc.max, 2, border = "green")
rect.hclust(cars.hc.max, 3, border = "red")
rect.hclust(cars.hc.max, 4, border = "blue")
rect.hclust(cars.hc.max, 5, border = "orange")
```

## Partitions en 2, 3, 4 ou 5 classes



*# affichage des tables de contingence*

```
cut_affiche_table_contingence(cars.hc.max, c(2, 3, 4, 5), cars$brand)
```

```
##
## Cut = 2
## Valeurs absolues :
##      1  2
## Europe 0 48
## Japan   0 51
## US      71 91
## tot     71 190
##
## Valeurs relatives :
##      1      2
## Europe 0.00000 18.39080
## Japan   0.00000 19.54023
## US      27.20307 34.86590
## tot     27.20307 72.79693
##
## Cut = 3
## Valeurs absolues :
##      1  2  3
## Europe 0 35 13
## Japan   0 44  7
## US      71 88  3
## tot     71 167 23
```

```

##
## Valeurs relatives :
##      1      2      3
## Europe 0.00000 13.40996 4.980843
## Japan  0.00000 16.85824 2.681992
## US     27.20307 33.71648 1.149425
## tot    27.20307 63.98467 8.812261
##
## Cut = 4
## Valeurs absolues :
##      1  2  3  4
## Europe 0 11 24 13
## Japan  0 28 16  7
## US     71 28 60  3
## tot    71 67 100 23
##
## Valeurs relatives :
##      1      2      3      4
## Europe 0.00000 4.214559 9.195402 4.980843
## Japan  0.00000 10.727969 6.130268 2.681992
## US     27.20307 10.727969 22.988506 1.149425
## tot    27.20307 25.670498 38.314176 8.812261
##
## Cut = 5
## Valeurs absolues :
##      1  2  3  4  5
## Europe 0 11  3 13 21
## Japan  0 28  2  7 14
## US     71 28 35  3 25
## tot    71 67 40 23 60
##
## Valeurs relatives :
##      1      2      3      4      5
## Europe 0.00000 4.214559 1.1494253 4.980843 8.045977
## Japan  0.00000 10.727969 0.7662835 2.681992 5.363985
## US     27.20307 10.727969 13.4099617 1.149425 9.578544
## tot    27.20307 25.670498 15.3256705 8.812261 22.988506

```