

Hierarchical Clustering

Kumar Rahul

6/9/2017

Preparing Data

Read data from a specified location

```
beerdata <- read.csv("/Users/Rahul/Documents/Datasets/Hclust_Beer data.csv",  
header = TRUE, sep = ",", na.strings = c("", " ", "NA"))
```

Summary of the data on which model is built and Standardizing the variables

```
str(beerdata)  
  
## 'data.frame':    20 obs. of  6 variables:  
## $ ID   : int  1 2 3 4 5 6 7 8 9 10 ...  
## $ BEER: Factor w/ 20 levels "Augsburger","Becks",...: 4 19 12 11 9 15 1 20  
##      14 3 ...  
## $ CAL  : int  144 151 157 170 152 145 175 149 99 113 ...  
## $ SOD  : int   15 19 15  7 11 23 24 27 10  8 ...  
## $ ALC  : num   4.7 4.9 4.9 5.2 5 4.6 5.5 4.7 4.3 3.7 ...  
## $ COST: num   0.43 0.43 0.48 0.73 0.77 0.28 0.4 0.42 0.43 0.44 ...  
  
beerdata <- na.omit(beerdata)  
c <- scale(beerdata[,3:6])
```

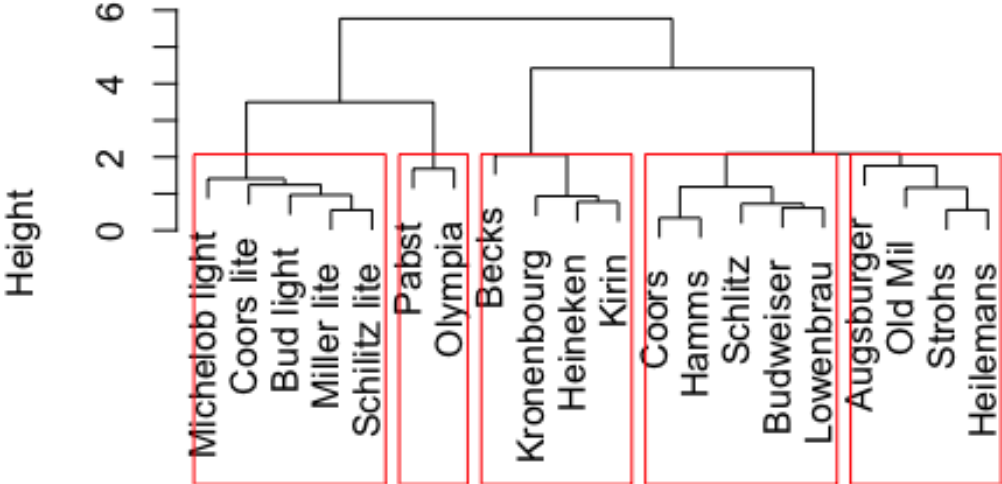
Using euclidean as distance measure for hierarchical clustering

```
d <- dist(c, method = "euclidean") # distance matrix  
hclust.result <- hclust(d, method="complete")  
hclust.result$labels <- beerdata[,2] #add beer names as labels
```

Plot using dendrogram plot to visualize the clusters

```
plot(hclust.result)  
# cut tree into 5 clusters  
groups <- cutree(hclust.result, k=5)  
# draw dendrogram with red borders around the 5 clusters  
rect.hclust(hclust.result, k=5, border="red")
```

Cluster Dendrogram



d
hclust (*, "complete")