Unsupervised Learning - Hierarchical Clustering -2

Abhirup Sen

11/06/2021

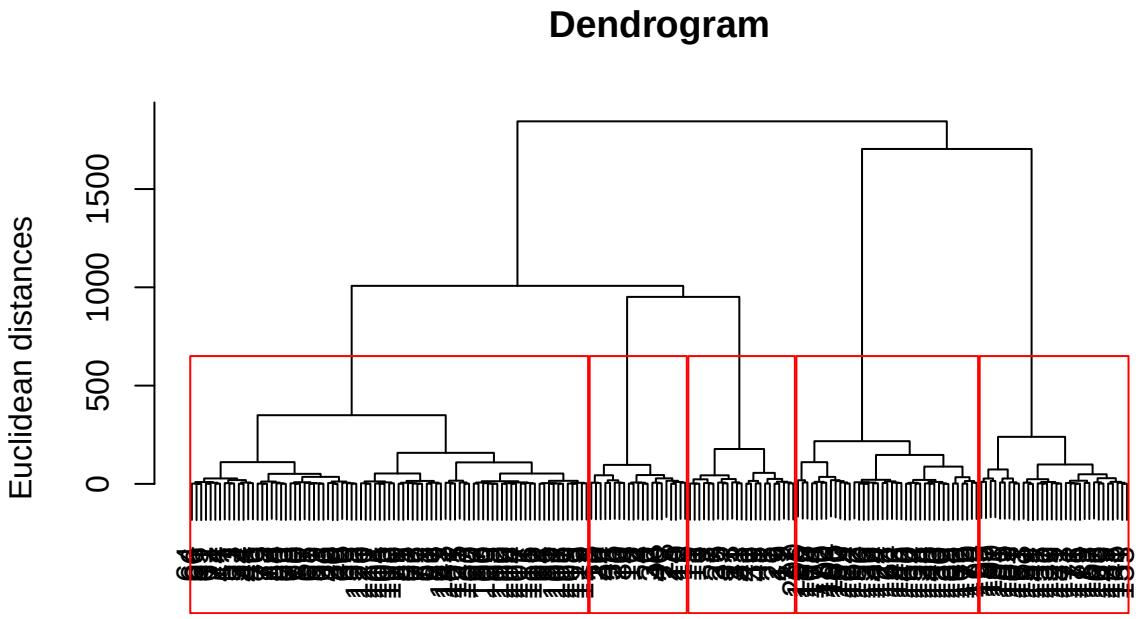
R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

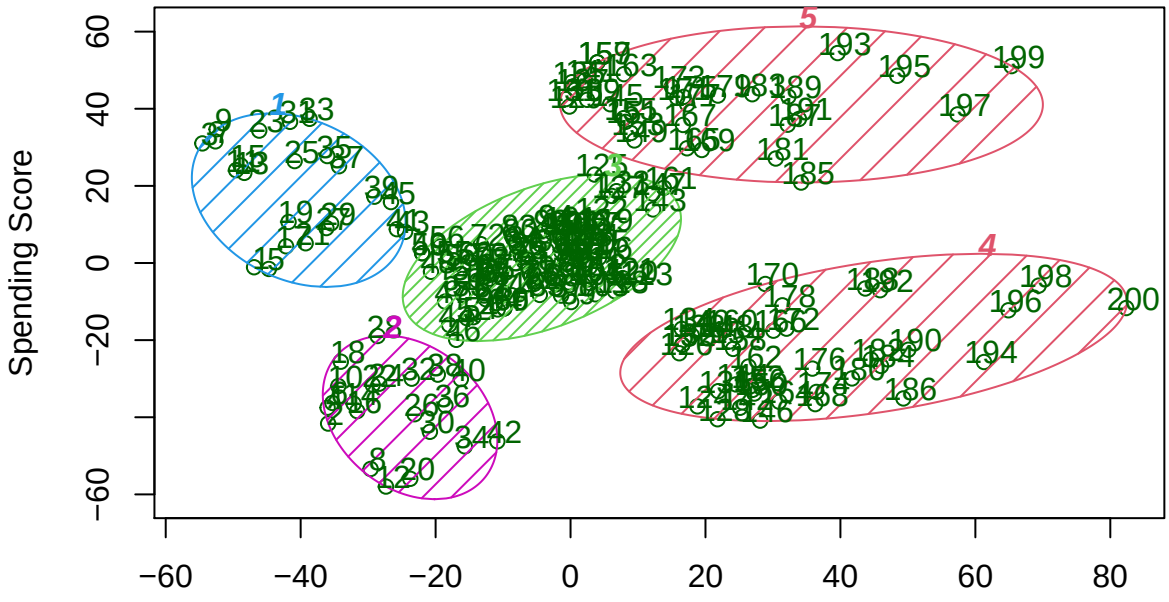
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
# Importing the dataset
dataset = read.csv('Mall_Customers.csv')
dataset = dataset[4:5]
```

```
# Using the dendrogram to find the optimal number of clusters
dendrogram = hclust(d = dist(dataset, method = 'euclidean'), method = 'ward.D')
plot(dendrogram,
     main = paste('Dendrogram'),
     xlab = 'Customers',
     ylab = 'Euclidean distances')
# Fitting Hierarchical Clustering to the dataset
hc = hclust(d = dist(dataset, method = 'euclidean'), method = 'ward.D')
y_hc = cutree(hc, 5)
rect.hclust(hc, k=5, border="red")
```




```
# Visualising the clusters
library(cluster)
clusplot(dataset,
         y_hc,
         lines = 0,
         shade = TRUE,
         color = TRUE,
         labels= 2,
         plotchar = FALSE,
         span = TRUE,
         main = paste('Clusters of customers'),
         xlab = 'Annual Income',
         ylab = 'Spending Score')
```



These two components explain 100 % of the point variability.

```
table(y_hc)
```

```
## y_hc
##  1  2  3  4  5
## 23 21 85 39 32
```