

Rioja Use Cases

Parismita Das

26 February 2017

Using Given Dataset

This is an example to obtain constrained HAC using dataset extracted from dissimilarity.txt. We start with converting the dataset given in format of .txt to distance class.

```
#reading dataset from txt file
data <- read.table("/home/parismita/dissimilarity.txt")
dmat<-as.dist(data)
```

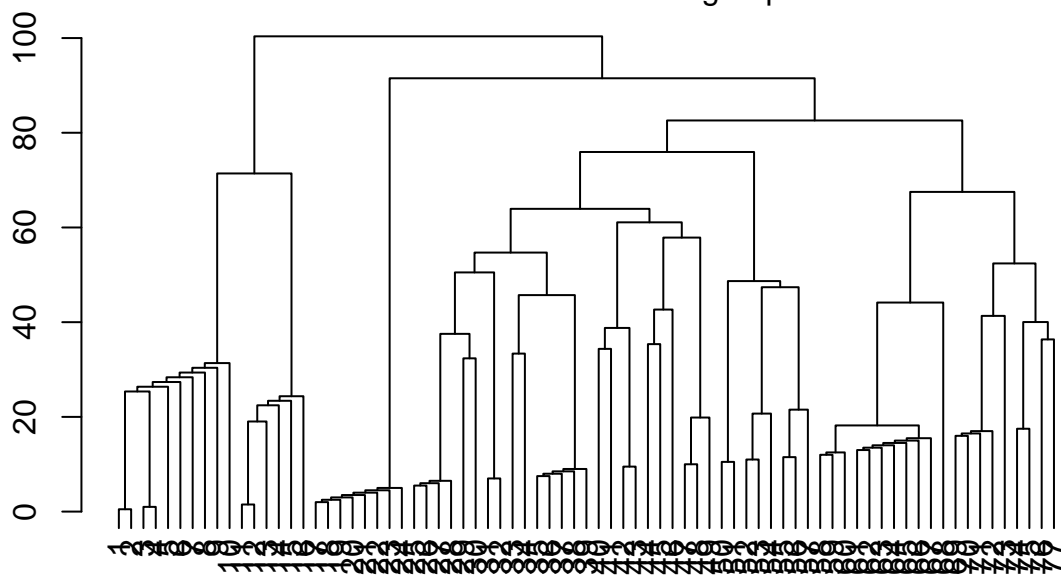
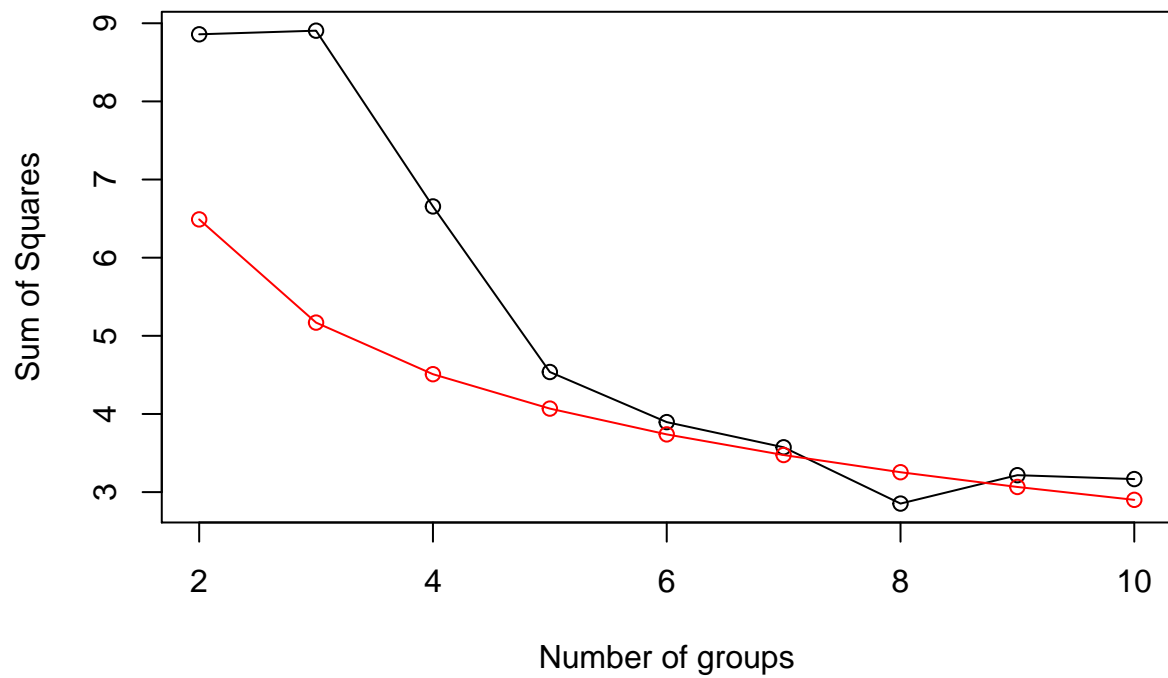
Computing the clusters using chclust. We have two methods in rioja package

- 1) slink
- 2) ward's

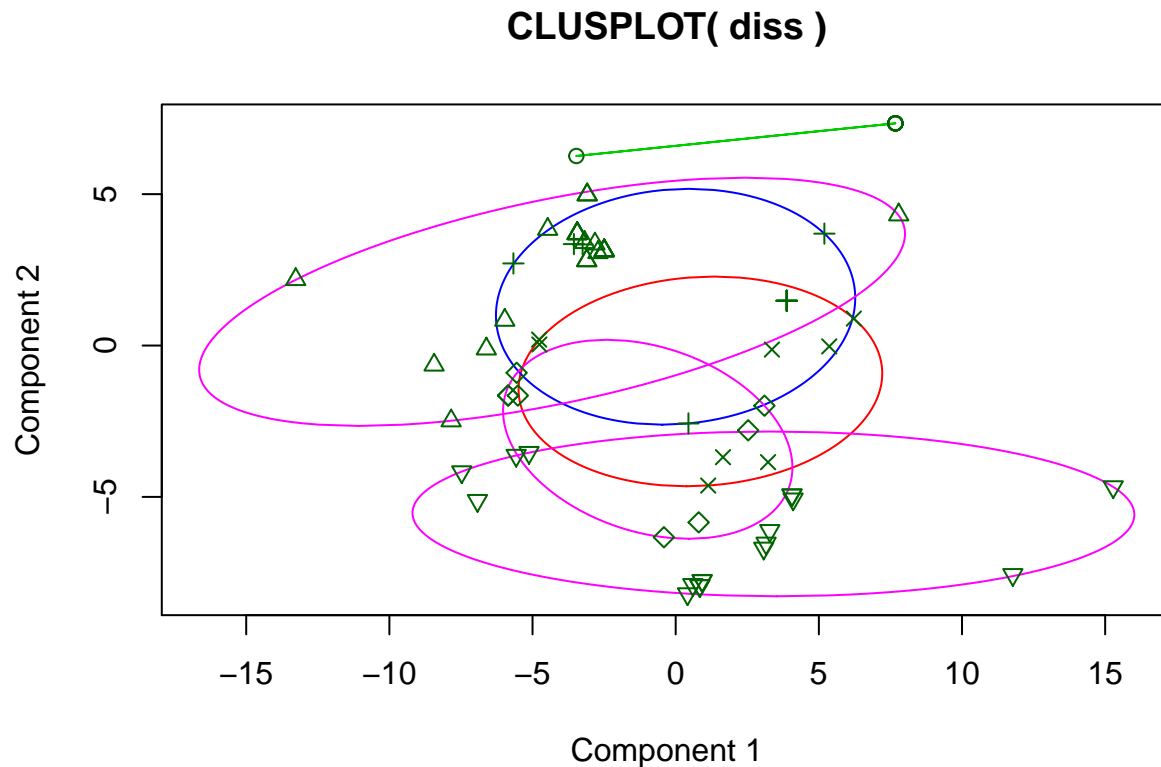
```
#chclust function for Constrained hierarchical clustering method
clust <- chclust(diss)                #using ward's
clust <- chclust(dmat,method = "conslink")  #using slink
```

Plotting the Bstick representation of the data for ward's method

```
## This is rioja 0.9-9
```



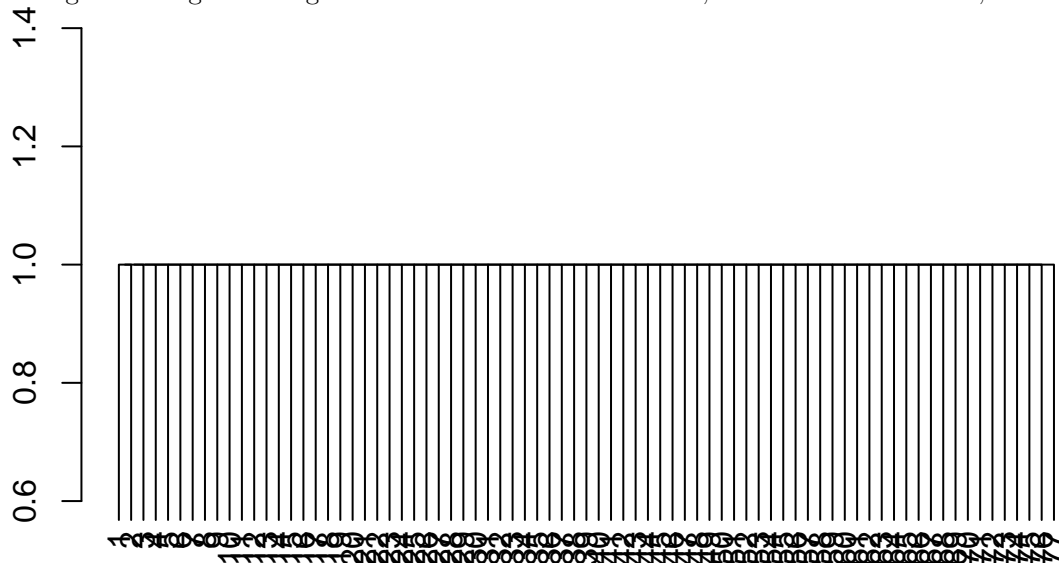
Plotting the dendrogram diagram which we obtain after clustering, according to its distance.



These two components explain 62.93 % of the point variability.

Showing the Clusters we get using using the function hclust, Using Ward's Method.

Using the Single Linkage Criterion on the same data, and distance metric, We get the plot as:



Hence this shows the Hierarchical clustering using our own data.

Problems with Non-Euclidian Data

With the given data, problems found due to non-euclidean metric are:

For ward's linkage: It can yield non-increasing merges if the dissimilarity is non-euclidean.

For Single linkage: The data have no variance hence cant separate into clusters as all clusters are at same height so even if we cut the dendrogram to obtain clusters, it gives same number of elements