

Table of Contents

<i>Cluster analysis (hierarchical agglomerative classification)</i>	1
R functions	1
Note about Ward's hierarchical clustering algorithm	1

Section: [Numerical classification](#)

Cluster analysis (hierarchical agglomerative classification)

Theory **R functions** Examples Exercise 

R functions

- **hclust** - calculates hierarchical cluster analysis. Requires at least two arguments: `d` for distance matrix, and method for agglomerative algorithm, one of `ward.D`, `ward.D2`, `single`, `complete`, `average` (= UPGMA), `mcquitty` (= WPGMA), `median` (= WPGMC) or `centroid` (= UPGMC). Has it's own `plot` function.
- **rect.hclust** - divides dendrogram into given number of groups (argument `k`) and draws rectangles around samples in these groups (argument `border` specifies the color of the rectangle).
- **cutree** - cuts the tree (dendrogram) into given number of clusters (argument `k`) or according to given level of similarity (argument `h`). Returns vector with assignment of samples into groups.
- **agnes (library cluster)** - contains six agglomerative algorithms, some not included in `hclust`. Has it's own `plot` method.
- **library (dendextend)** - contains several functions improving representation of the dendrogram (e.g. plotting dendrogram with branches of different colour)

Note about Ward's hierarchical clustering algorithm

[Murtagh & Legendre \(2014\)](#) have shown that what literature refers to as Ward's clustering algorithm are in fact two slightly different methods, while only one of them is identical with the algorithm originally described by Ward. Both functions `hclust` and `agnes` have the `method = 'ward'`, but with different default. While `hclust` function implements both Ward's algorithms (the genuine one, named `ward.D2`, as well as the second one, called `ward.D`), the `agnes` function implements only the genuine one. For historical reason, the argument `method = 'ward'` in `hclust` calls the `ward.D` algorithm instead of `ward.D2` one. This means that `hclust` and `agnes` function, if both to set to `method = 'ward'`, return slightly different results. To calculate "genuine" Ward's algorithm in both methods, you need to set up `method = 'ward.D2'` in `hclust` (and `method = 'ward'` in `agnes`, but there is no other option for Ward algorithm anyway).

From:

<https://anadat-r.davidzeleny.net/> - **Analysis of community ecology data in R**

Permanent link:

https://anadat-r.davidzeleny.net/doku.php/en:hier-agglom_r

Last update: **2019/03/22 22:37**