**Hierarchical Clustering with hclust**

Hierarchical agglomeration is a “bottom-up” approach to clustering, in which individual observations are sequentially merged into clusters according to a particular combinatorial formula.

**Table 1: Comparison of Hierarchical Clustering Methods**

Each method in **Table 1** represents a special case of the general method of hierarchical agglomerative clustering. All seven methods are available within R function hclust.

The formulas in **Table 1** give the “distance” from an observation (or cluster) “J” to a cluster “M”, which contains sub-clusters “K” and “L”, as defined by each method of agglomeration.

|  |  |
| --- | --- |
| **Method** | **Complete** |
| **Full Name** | Complete Linkage Agglomerative Clustering |
| **Reference** | [(Sørensen, 1948)](https://paperpile.com/c/WH60jr/DZmL) |
| **Equation** |  |
| **Description** | DJM is the *maximum* distance between J and *all* members of M. |
| **Merge Rule** | “You must *all* be X% similar to join” |
| **Comments:** | * Tends to separate-out overlapping groups well * Does poorly if recovering non-standard shapes * Biased towards producing clusters with similar diameters. * Can be severely distorted by moderate outliers: e.g., once an outlier has been joined to a cluster, the threshold to join that cluster may be raised prohibitively high, banning individuals from joining that would be natural members of that cluster by other methods. * Ensures that *all* members of a cluster have something in common. |
| **Method** | **Single** |
| **Full Name** | Single Linkage Agglomerative Clustering |
| **Reference** | [(Legendre & Legendre, 1983)](https://paperpile.com/c/WH60jr/GhbI) |
| **Equation** |  |
| **Description** | DJM is the *minimum* distance between J and *any* member of M. |
| **Merge Rule** | “As long as *any* two of you are X% similar, you can *all* join” |
| **Comments:** | * Any degree of overlap between clusters can cause “incorrect” merging. * Imposes no constraints on the shape of clusters; hence, can detect long and irregular clusters, but tends not to produce compact clusters. * The minimum requirement can lead to problematic “*chaining*”: successive individuals may be joined to a group with some degree of overlap, but (unlike Complete-Linkage) many members of the resulting cluster may have *nothing* in common with the others. |
| **Method** | **Average** |
| **Full Name** | Average Linkage Agglomerative Clustering |
| **Reference** | [(Sneath, Sokal, & Freeman, 1975)](https://paperpile.com/c/WH60jr/MLlW) |
| **Equation** |  |
| **Description** | DJM is the *average* distance between J and *all* *individual* members of M. |
| **Merge Rule** | “You must *all* be, *on average*, X% similar to join” |
| **Comments:** | * Gives a result usually between the extremes offered by Complete- and Single-Linkage clustering. * Tends to join clusters with small variances. * Slightly biased towards producing clusters with the same variance. * Relatively robust to outliers. |
| **Method** | **Centroid** |
| **Full Name** | Centroid Clustering |
| **Reference** | [(Gower, 1967)](https://paperpile.com/c/WH60jr/Ukwq) |
| **Equation** |  |
| **Description** | DJM is the *weighted* average distance between J and *sub-clusters* of M. |
| **Merge Rule** | “You must be, *on average*, X% similar to existing *sub-clusters* to join, bearing in mind their *relative size*” |
| **Comments:** | * Moderately robust to outliers. * Weighting causes clusters containing sub-clusters to be dominated by the largest of those sub-clusters. |
| **Method** | **McQuitty** |
| **Full Name** | McQuitty’s Similarity Analysis |
| **Reference** | [(McQuitty, 1966)](https://paperpile.com/c/WH60jr/eJ8j) |
| **Equation** |  |
| **Description** | DJM is the *average* distance between J and *sub-clusters* of M. |
| **Merge Rule** | “You must be, *on average*, X% similar to existing *sub-clusters* to join (*regardless of* their relative size)” |
| **Comments:** | * Offers few advantages over the other methods. * Tends to give similar results to Median and Complete |
| **Method** | **Median** |
| **Full Name** | Median Clustering |
| **Reference** | [(Gower, 1967)](https://paperpile.com/c/WH60jr/Ukwq) |
| **Equation** |  |
| **Description** | DJM is the *average* distance between J and *sub-clusters* of M, *minus* the *median* distance between those *sub-clusters*. |
| **Merge Rule** | “You must be, *on average*, X% similar to existing *sub-clusters* to join,  *penalizing* for the *distances between* the sub-clusters you may join.” |
| **Comments:** | * Offers few advantages over the other methods. * Assumes that observations can be represented in Euclidean space. * Not biased towards producing clusters of equal size. * Increases the gravity/”pull” of tightly-knit clusters: e.g., An individual *closest* to a *loosely*-knit cluster might instead be assigned to a more distant tightly-knit cluster. |
| **Method** | **Ward** |
| **Full Name** | Ward’s Minimum Variance Method |
| **Reference** | [(Murtagh & Legendre, 2014; Ward, 1963)](https://paperpile.com/c/WH60jr/IXZ5+tqea) |
| **Equation** |  |
| **Description** | DJM is the *sum of squares* between J and *all sub-clusters* of M. |
| **Merge Rule** | “You must cause *only* an X% (i.e., the minimum) increase in the *total within-cluster variance* to join at this stage.” |
| **Comments:** | * Must use the squared Euclidean distances as input into Ward. * Minimization of ESS at each generation is performed with some assumptions: (i) multivariate normal mixture, (ii) equal spherical covariance matrices, (iii) equal sampling probabilities. * Similar to K-means clustering (uses same objective criteria) * Finds compact spherical clusters. * Strongly biased towards producing clusters of similar size (biased against small clusters). * Sensitive to outliers. |

**Table 2: Combinatorial Formulas for Hierarchical Clustering Methods**

(Just a subset of Table 1, in case it’s useful to have the equations on their own…).

|  |
| --- |
| **Complete** |
| (Complete Linkage Agglomerative Clustering) |
| [(Sørensen, 1948)](https://paperpile.com/c/WH60jr/DZmL) |
|  |
| **Single** |
| (Single Linkage Agglomerative Clustering) |
| [(Legendre & Legendre, 1983)](https://paperpile.com/c/WH60jr/GhbI) |
|  |
| **Average** |
| (Average Linkage Agglomerative Clustering) |
| [(Sneath, Sokal, & Freeman, 1975)](https://paperpile.com/c/WH60jr/MLlW) |
|  |
| **Centroid** |
| (Centroid Clustering) |
| [(Gower, 1967)](https://paperpile.com/c/WH60jr/Ukwq) |
|  |
| **McQuitty** |
| (McQuitty’s Similarity Analysis) |
| [(McQuitty, 1966)](https://paperpile.com/c/WH60jr/eJ8j) |
|  |
| **Median** |
| (Median Clustering) |
| [(Gower, 1967)](https://paperpile.com/c/WH60jr/Ukwq) |
|  |
| **Ward** |
| (Ward’s Minimum Variance Method) |
| [(Murtagh & Legendre, 2014; Ward, 1963)](https://paperpile.com/c/WH60jr/IXZ5+tqea) |
|  |

**Example: Maltese Data**

* ????
* Not sure how to interpret Sara’s plots…

**References:**

[Gower, J. C. (1967). A comparison of some methods of cluster analysis. *Biometrics*, *23*(4), 623–637.](http://paperpile.com/b/WH60jr/Ukwq)

[Legendre, P., & Legendre, L. F. J. (1983). *Numerical Ecology* (2nd ed., Vol. 24). Elsevier.](http://paperpile.com/b/WH60jr/GhbI)

[McQuitty, L. L. (1966). Single and multiple hierarchical classification by reciprocal pairs and rank order types. *Educational and Psychological Measurement*. http://doi.org/](http://paperpile.com/b/WH60jr/eJ8j)[10.1177/001316446602600201](http://dx.doi.org/10.1177/001316446602600201)

[Murtagh, F., & Legendre, P. (2014). Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion? *Journal of Classification*, *31*(3), 274–295.](http://paperpile.com/b/WH60jr/tqea)

[Sneath, P. H. A., Sokal, R. R., & Freeman, W. H. (1975). Numerical Taxonomy. The Principles and Practice of Numerical Classification. *Systematic Zoology*, *24*(2), 263–268.](http://paperpile.com/b/WH60jr/MLlW)

[Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskabs Biologiske Skrifter*, *5*, 1–34.](http://paperpile.com/b/WH60jr/DZmL)

[Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244.](http://paperpile.com/b/WH60jr/IXZ5)