

## Chapter III-7 — Analysis

Column 0 and Column 1 contain the node numbers that are combined to form the node represented by a given row in the dendrogram wave. If a node number is negative, it represents the row in the original vector data. If a node number is positive, it represents the row number within the dendrogram wave of a combined node. Since negative numbers cannot represent zero-based indices, the row numbers are 1-based. If the value in column 0 or 1 is A, then the zero-based row number is  $\text{abs}(A)-1$ .

Column 2 contains the dissimilarity value between the nodes represented by columns 0 and 1. If you are drawing a dendrogram tree, a tie bar should be drawn in a location proportional to this dissimilarity.

Column 3 contains ordering information for the original data. This allows you to draw lines that don't cross. Labels for vector rows should also use this ordering. These numbers are zero-based indices into the original data. If you have a text wave with labels for each of the rows in your vector data, you can use this column to access the correct label for a given row in the dendrogram. If you include a false-color image (heat map) of the dissimilarity matrix or the vector data, the rows of the heat map data should be re-ordered based on column 3.

A data set with N vector rows has N-1 dissimilarity values. The information for column 3 has N values, one for each of the vector rows, so column 0 and column 1 contain NaN in row N to show that those cells are not used. The last row in column 2 has zero, which is a nonsense dissimilarity value. Only column 3 has useful data in the last row.

### Dendrogram Example

Create a matrix wave containing fake data representing XY pairs with the X values in column 0 of the matrix and the Y values in column 1:

```
Make/N=(10,2)/0 vectors
vectors[0][0]= {6.04379,-1.40976,5.32518,-0.140695,1.94087,0.971393,5.29093,-
                0.953138,5.58752,4.16877}
vectors[0][1]= {6.2448,-0.914027,4.36285,1.45288,-
                1.21538,1.21118,5.86274,1.37278,4.43314,4.40538}
```

Each row in vectors, that is, each XY pair, is one input vector to HCluster.

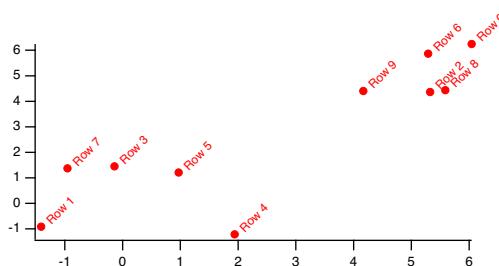
Create a text wave with labels for each of the vector rows:

```
Make/N=10/T/0 labels
labels[0]= {"Row0", "Row1", "Row2", "Row3", "Row4", "Row5", "Row6", "Row7", "Row8", "Row9"}
```

Make a graph of the XY pairs:

```
Display vectors[*][1] vs vectors[*][0]           // For displaying markers
AppendToGraph vectors[*][1] vs vectors[*][0] // For displaying labels
ModifyGraph margin(top)=27
ModifyGraph mode=3
ModifyGraph marker(vectors)=19
ModifyGraph textMarker(vectors#1)={labels,"default",0,45,5,15.00,12.00}
```

Since this is fake data, we have manipulated it so that there are two clear clusters. We used text markers to show which vector row each XY pair represents:



Invoke HCluster to analyze the vectors and create a dendrogram output wave:

HCluster/OTYP=Both vectors

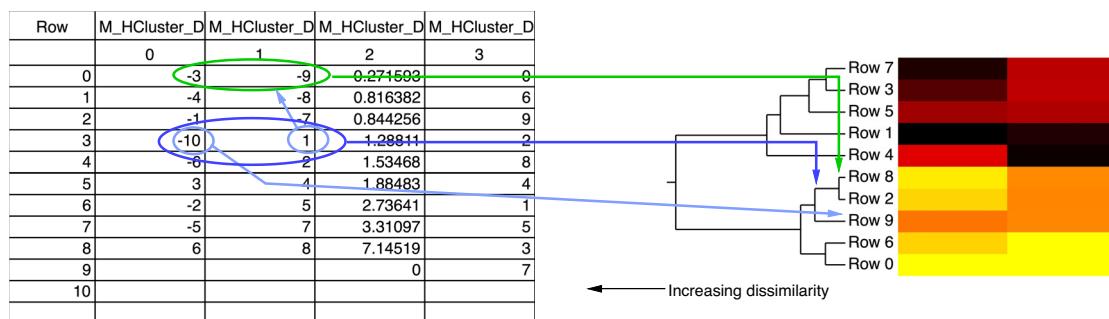
The resulting dendrogram wave, with the default name M\_HCluster\_Dendrogram, has these contents:

Row	M_HCluster_D	M_HCluster_D	M_HCluster_D	M_HCluster_D
	0	1	2	3
0	-3	-9	0.271593	0
1	-4	-8	0.816382	6
2	-1	-7	0.844256	9
3	-10	1	1.28811	2
4	-6	2	1.53468	8
5	3	4	1.88483	4
6	-2	5	2.73641	1
7	-5	7	3.31097	5
8	6	8	7.14519	3

The first row represents the smallest dissimilarity (0.2716). The first two columns indicate that a node has been formed by combining original vector rows 2 and 8 (that is  $\text{abs}(-3)-1$  and  $\text{abs}(-9)-1$ ). If we are drawing a dendrogram, we would position rows 2 and 8 at a position 4 and 5 up from the bottom (or down from the top). We see that from the fact that column 3 has row 2 and row 8 listed in those positions. We would draw a tie line between 2 and 8 at a position on the dendrogram proportional to a distance of 0.2716.

The first three rows have negative numbers in columns 0 and 1, showing that they all combine nodes representing the original vectors. If we have text labels for the rows, these nodes would all tie directly to those labels.

Row 3 lists nodes -10 and 1, indicating that it combines an original vector row 9 with the already combined node from row 0 of the dendrogram wave. The relationships between a dendrogram with a heat map of the original vector data, and the dendrogram wave is illustrated here:



Here we have ordered the vector rows from bottom to top following the ordering given in column 3 of the dendrogram wave. Rows 2 and 8 are tied together with the tie line least distant from the labels. Then Rows 3 and 7 and Rows 0 and 6 are each a bit more distant. Row 3 of the dendrogram wave indicates the tie between the label "Row 9" and the combined node from Row 2 and Row 8.

The two clusters are clearly indicated by the large distance between the most extreme tie line and the next most distant tie line. We might add a "cut line" in that space, and group the two clusters. To show it more clearly, we could color the two clusters differently, possibly like this:

