And Again:
The Hierarchical Clustering Algorithm

✓ Start with *n* clusters (record = cluster)

✓ Step 1: two closest records are merged into one cluster

At every step, pair of clusters with *smallest distance* are merged.

At this point the distance matrix is re-computed:

- **Two rows+columns are merged into single row+column**
- **Distances to the newly merged cluster are recalculated**

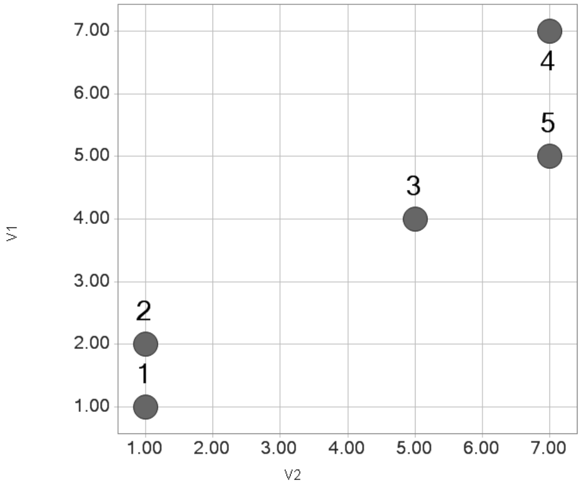
Repeat the last step until a single cluster is formed

The clustering process: example

(from http://obelia.jde.aca.mmu.ac.uk/multivar/dend.htm - no longer)

Two variables, n=5 items:

item	**v1**	**v2**
1	**1**	**1**
2	**2**	**1**
3	**4**	**5**
4	**7**	**7**
5	**5**	**7**



Euclidean distance matrix

	1	2	3	4	5
1	0.0				
2	1.0	0.0			
3	5.0	4.5	0.0		
4	8.5	7.8	3.6	0.0	
5	7.2	6.7	2.2	2.0	0.0

What happens next?

- Merge 1&2 into cluster A
- Use single linkage to compute distances from cluster A:

	1	2	3	4	5
1	0.0				
2	1.0	0.0			
3	5.0	4.5	0.0		
4	8.5	7.8	3.6	0.0	
5	7.2	6.7	2.2	2.0	0.0

→

	A	3	4	5
A	0.0			
3	4.5	0.0		
4	7.8	3.6	0.0	
5	6.7	2.2	2.0	0.0

What happens next?

Merge 4&5 (cluster B)

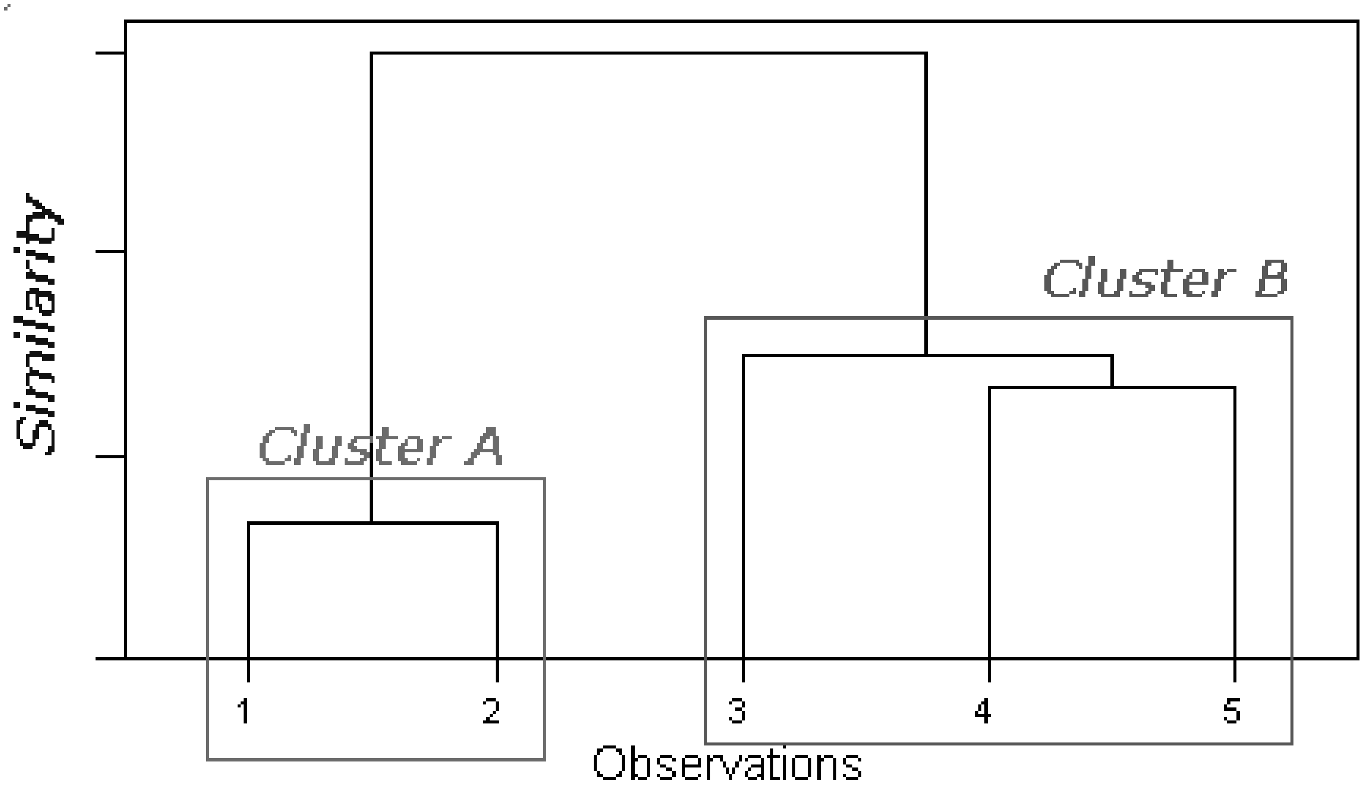
	A	3	B
A	0.0		
3	4.5	0.0	
B	6.7	2.2	0.0

→

Merge 3 & B

	A	B
A	0.0	
B	4.5	0.0

Finally: Summarize process in a **Dendrogram**


Similarity
Cluster A
Cluster B
1
2
3
4
5
Observations