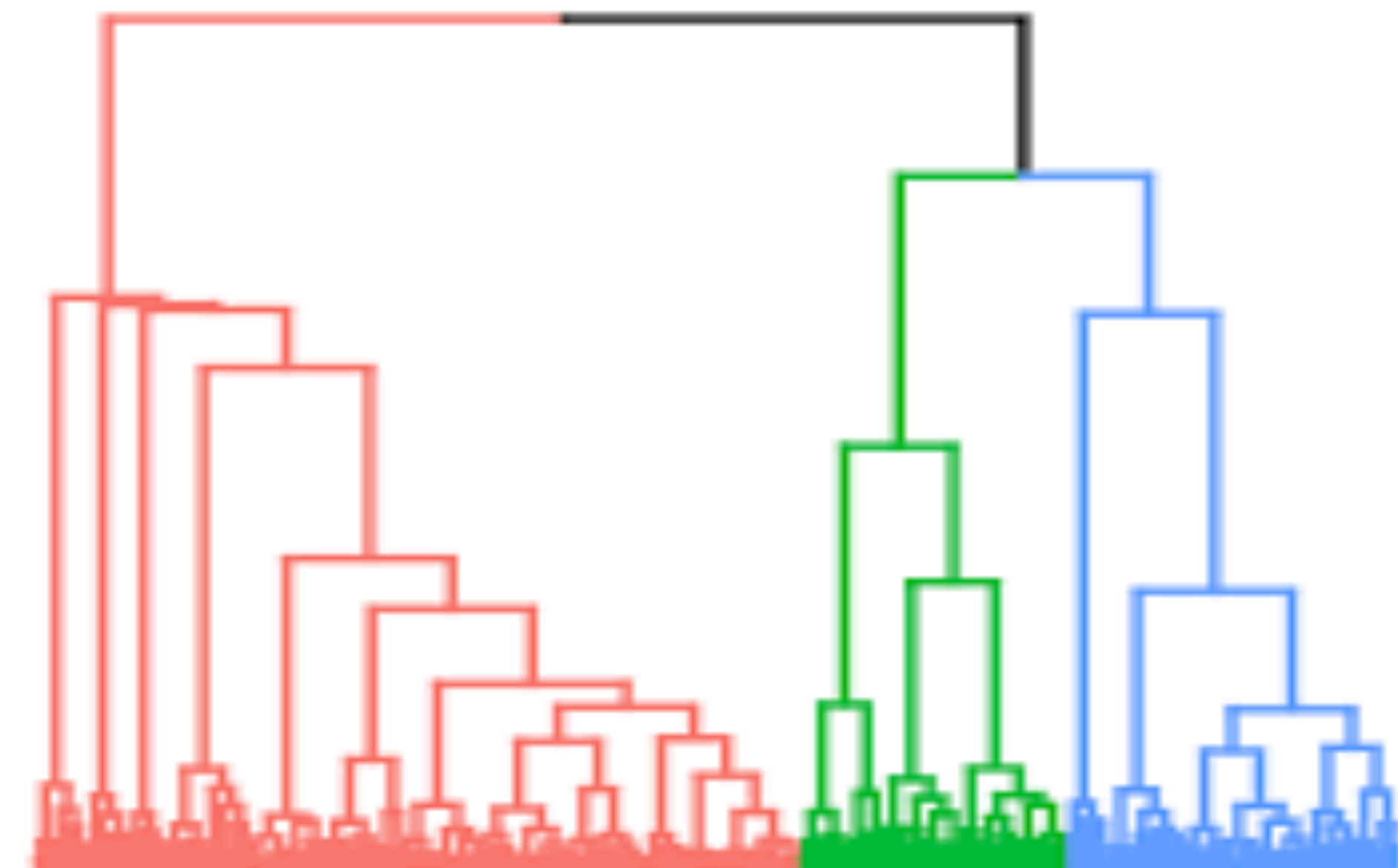
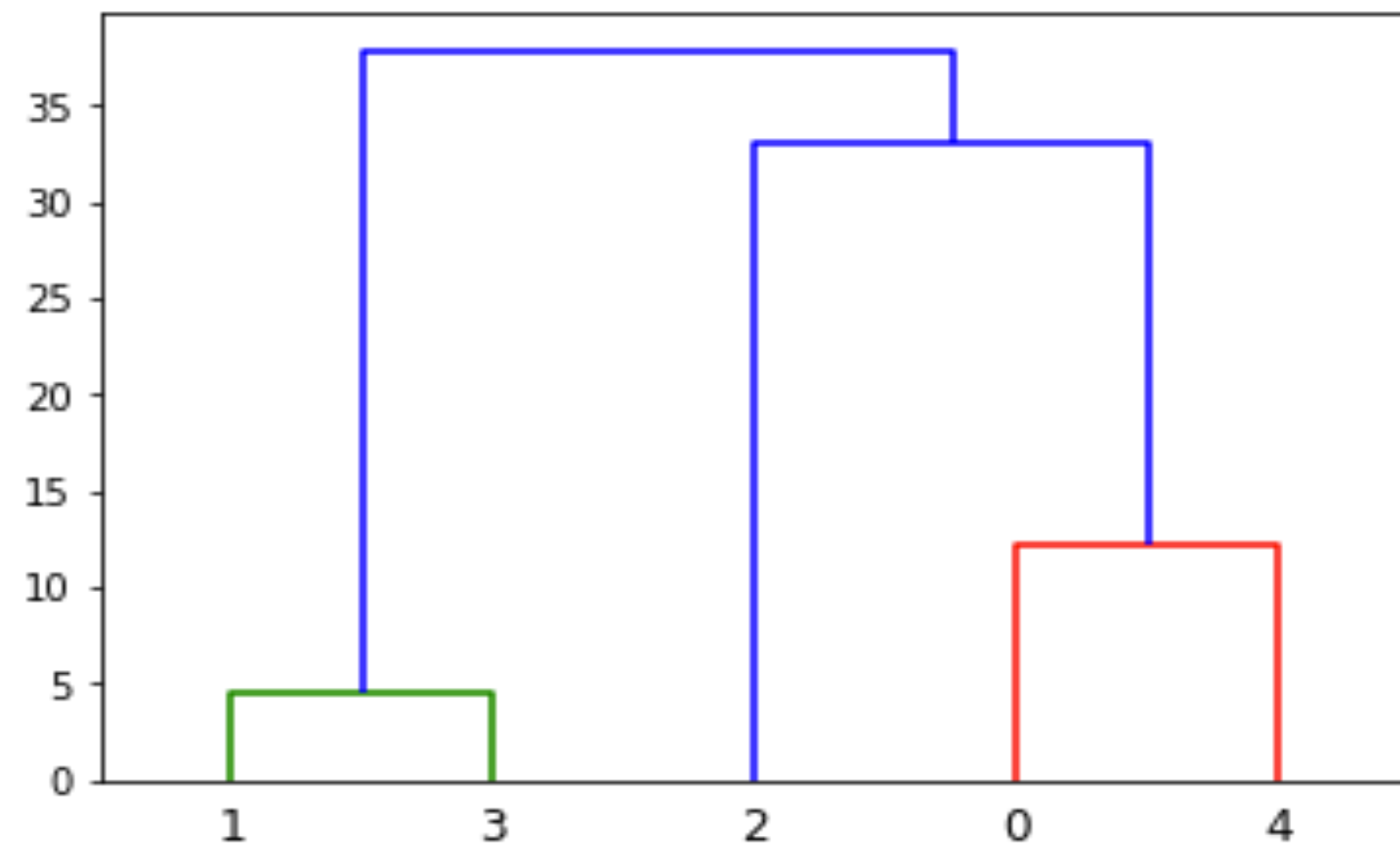


Hierarchical Clustering

Part 2



How does Similarity or Dissimilarity
matrix algorithm work?

Similarity or dissimilarity matrix

An experiment with 'n' observations

Dissimilarity matrix is a symmetric matrix of order 'n' x 'n'

Similarity or dissimilarity matrix

An experiment with 'n' observations

Dissimilarity matrix is a symmetric matrix of order 'n' x 'n'

Observations	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

5 x 5 symmetric distance matrix

Example: Step 1

Cluster	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

Example: Step 1

Cluster	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

Take the minimum of all distances.

Example: Step 1

Cluster	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

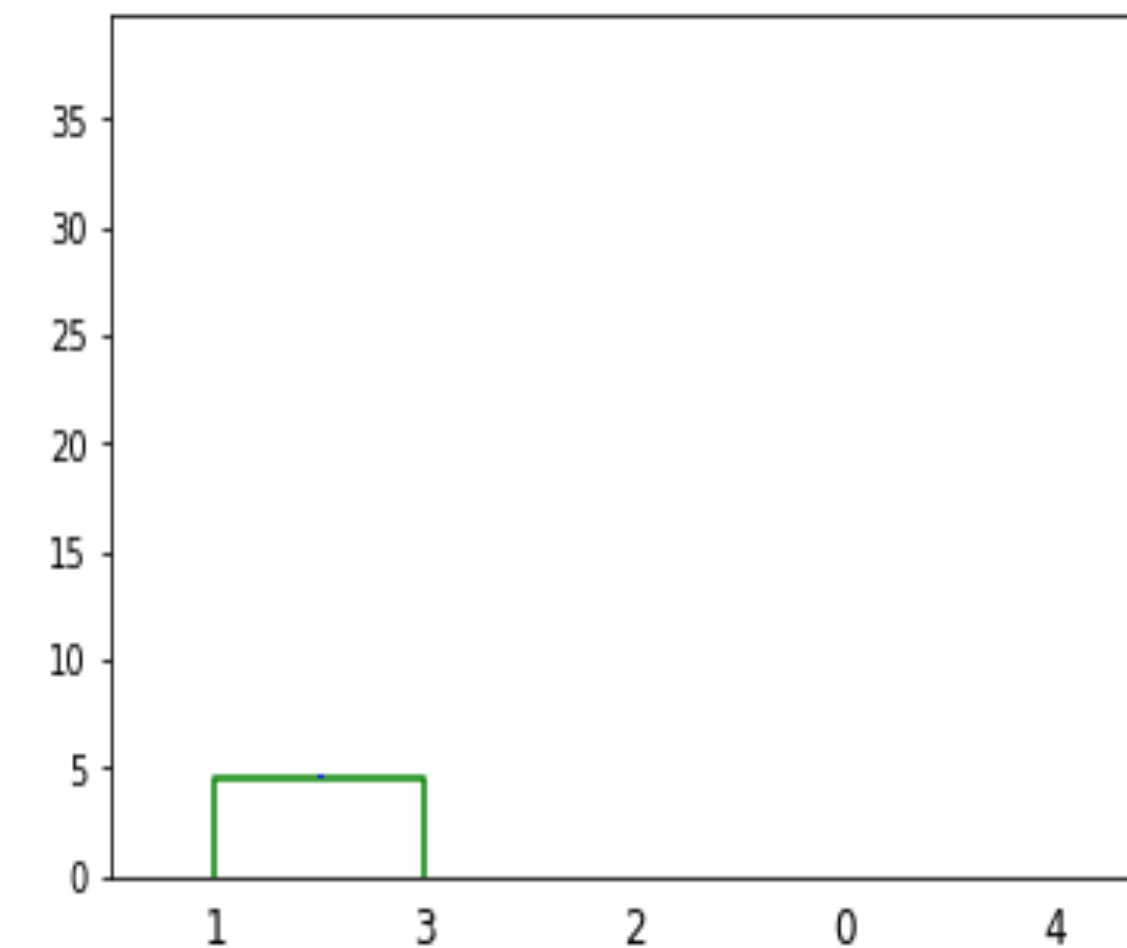
Cluster 1 and Cluster 3 are merged at distance 4.6

Example: Step 1

Cluster	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

Cluster 1 and Cluster 3 are merged at distance 4.6

A new cluster is formed. We will call it cluster (1,3)

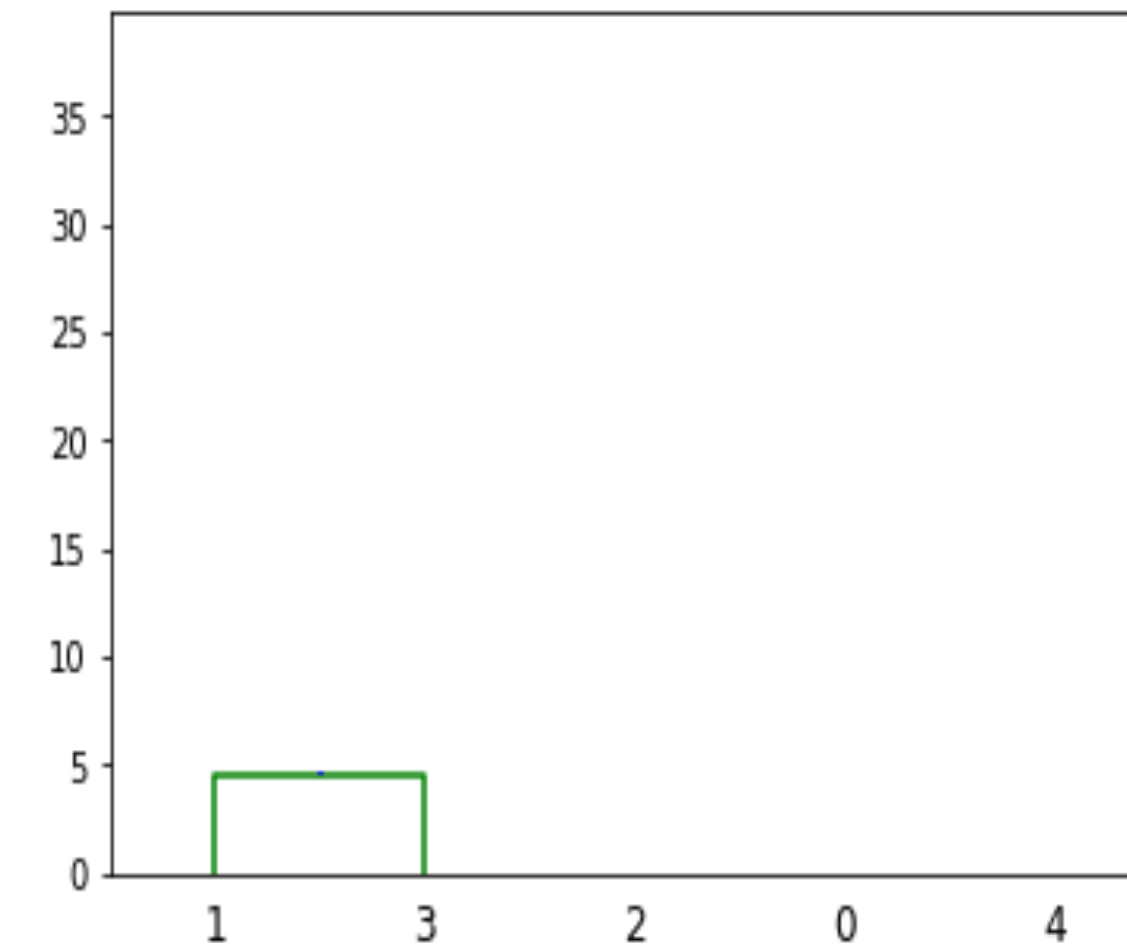


Example: Step 1

Cluster	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

Cluster 1 and Cluster 3 are merged at distance 4.6

A new cluster is formed. We will call it cluster (1,3)



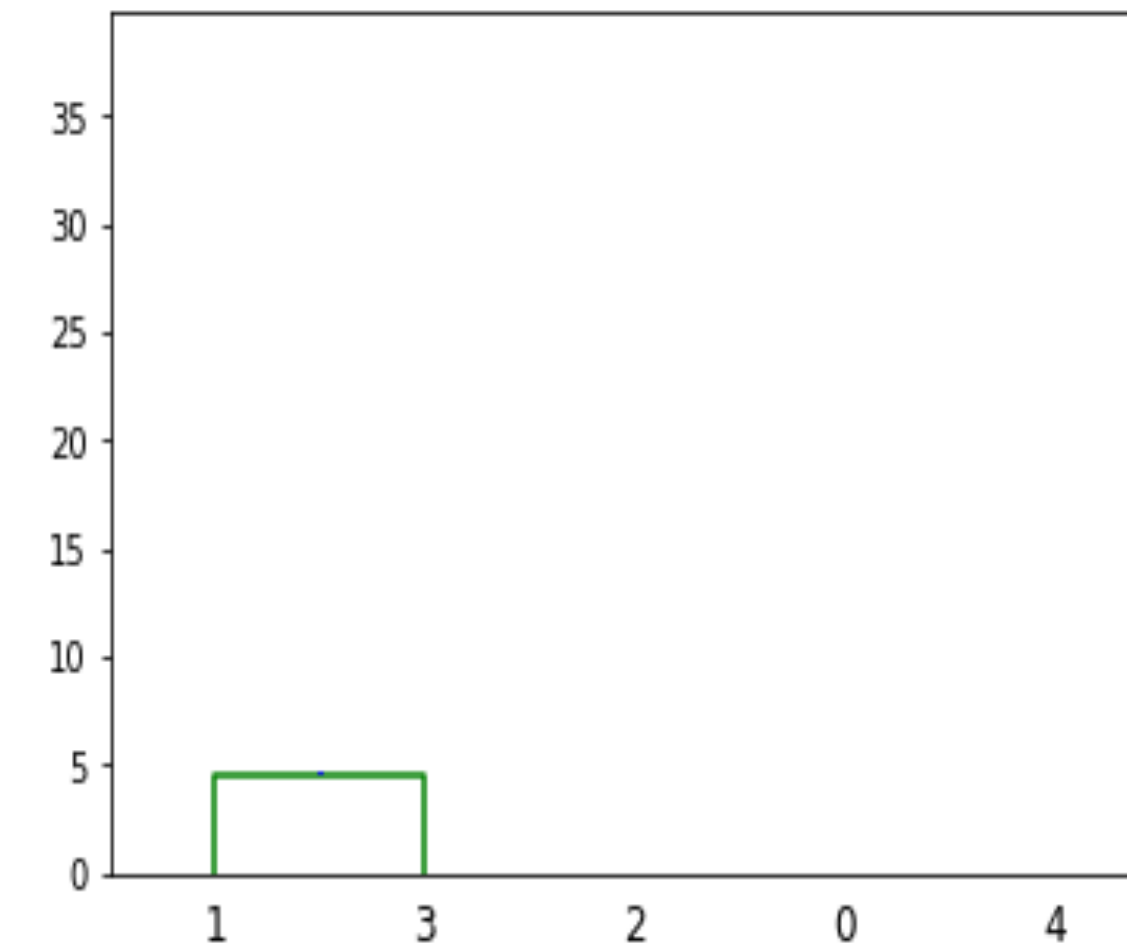
Discard columns and rows corresponding cluster 1 and cluster 3. And add cluster (1,3)

Example: Step 1

Cluster	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

Cluster 1 and Cluster 3 are merged at distance 4.6

A new cluster is formed. We will call it cluster (1,3)



Discard columns and rows corresponding cluster 1 and cluster 3. And add cluster (1,3)

What are distances between existing clusters and newly formed cluster?

Example: Step 1

Cluster	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

Cluster	0	1, 3	2	4
0	0			
1, 3	?	0		
2			0	
4				0

What are distances between existing clusters and newly formed cluster?

Example: Step 1

Cluster	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

Cluster	0	1, 3	2	4
0	0			
1, 3	?	0		
2			0	
4				0

$$\text{dist}(0, (1, 3)) = \min(\text{dist}(0, 1), \text{dist}(0, 3))$$

Example: Step 1

Cluster	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2			0	
4				0

$$\text{dist}(0, (1, 3)) = \min(\text{dist}(0, 1), \text{dist}(0, 3))$$

Example: Step 1

Cluster	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2		?	0	
4		?		0

$$\text{dist}(0, (1, 3)) = \min(\text{dist}(0, 1), \text{dist}(0, 3))$$

Example: Step 1

Cluster	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2		71.1	0	
4		37.9		0

$$\text{dist}(0, (1, 3)) = \min (\text{dist}(0,1), \text{dist}(0,3))$$

Example: Step 1

Cluster	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9		0

$$\text{dist}(0, (1, 3)) = \min (\text{dist}(0,1), \text{dist}(0,3))$$

Example: Step 1

Cluster	0	1	2	3	4
0	0				
1	42.0	0			
2	33.0	75.0	0		
3	38.2	4.6	71.1	0	
4	12.2	42.2	35.7	37.9	0

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

$$\text{dist}(0, (1, 3)) = \min (\text{dist}(0,1), \text{dist}(0,3))$$

Example: Step 1

New symmetric distance matrix with dimension 4 x 4

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

Example: Step 2

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

Example: Step 2

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

We will repeat the process that we did in step 1

Example: Step 2

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

We will repeat the process that we did in step 1

Example: Step 2

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

We will repeat the process that we did in step 1

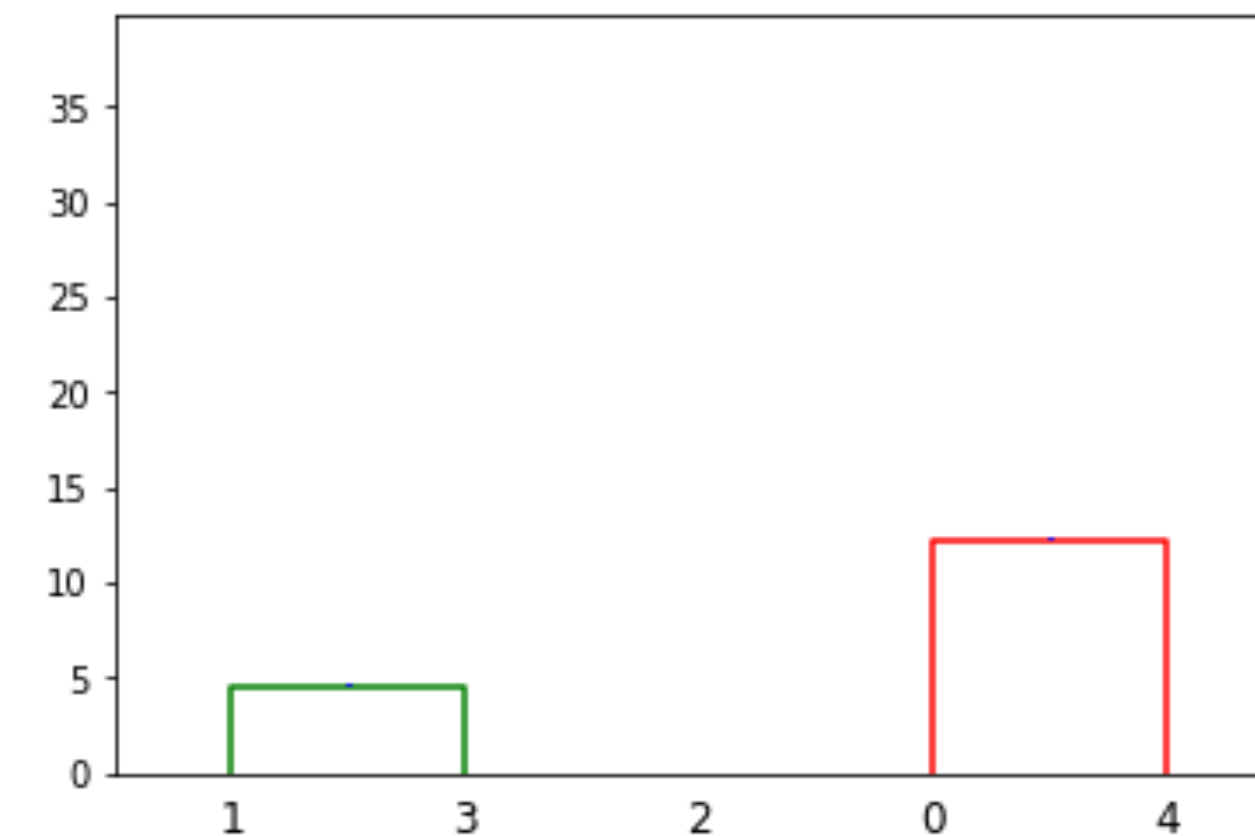
We merge these two clusters at distance 12.2

Example: Step 2

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

We will repeat the process that we did in step 1

We merge these two clusters at distance 12.2



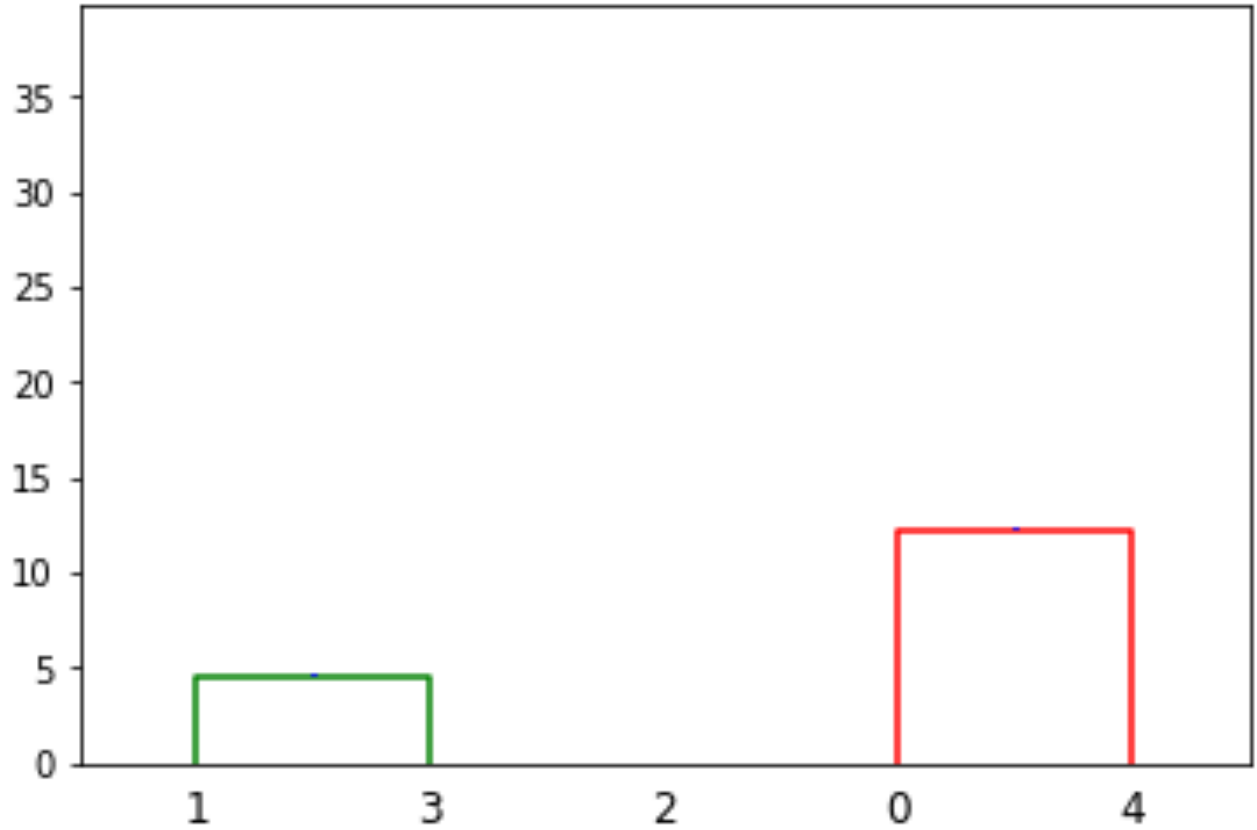
Example: Step 2

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

Cluster	0, 4	1, 3	2
0, 4	0		
1, 3		0	
2			0

We will repeat the process that we did in step 1

We merge these two clusters at distance 12.2



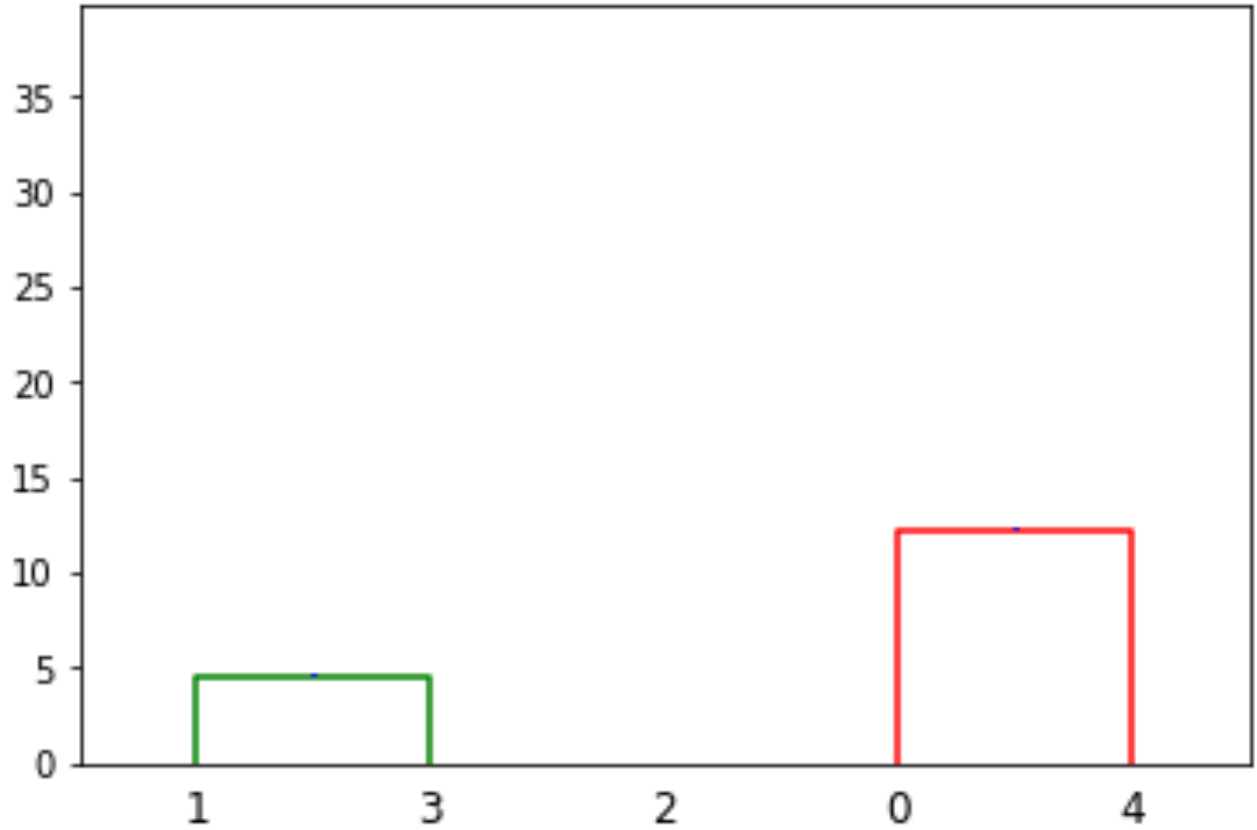
Example: Step 2

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

Cluster	0, 4	1, 3	2
0, 4	0		
1, 3	?	0	
2	?	71.1	0

We will repeat the process that we did in step 1

We merge these two clusters at distance 12.2



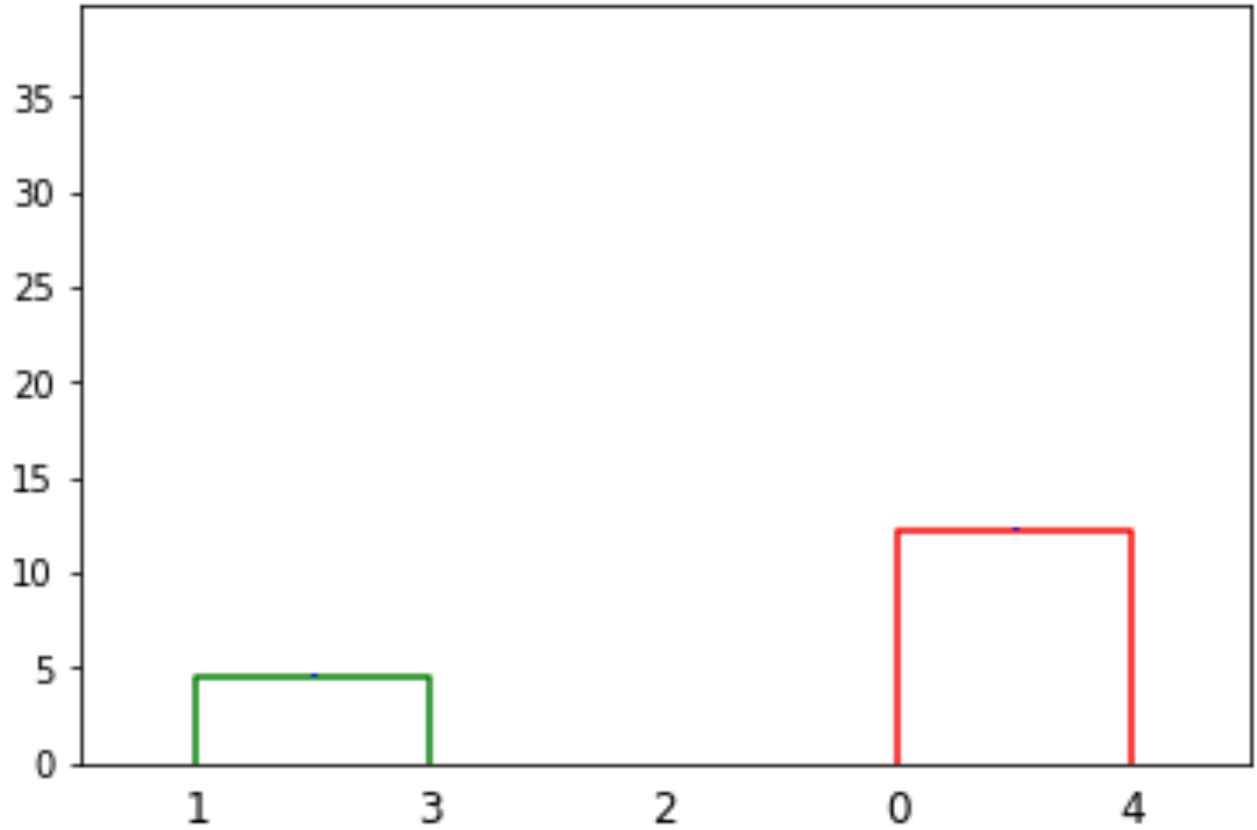
Example: Step 2

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

Cluster	0, 4	1, 3	2
0, 4	0		
1, 3	?	0	
2	?	71.1	0

We will repeat the process that we did in step 1

We merge these two clusters at distance 12.2



$\text{dist}((0, 4), (1, 3)) = \min (\text{dist}(0, (1, 3)), \text{dist}(4, (1, 3)))$

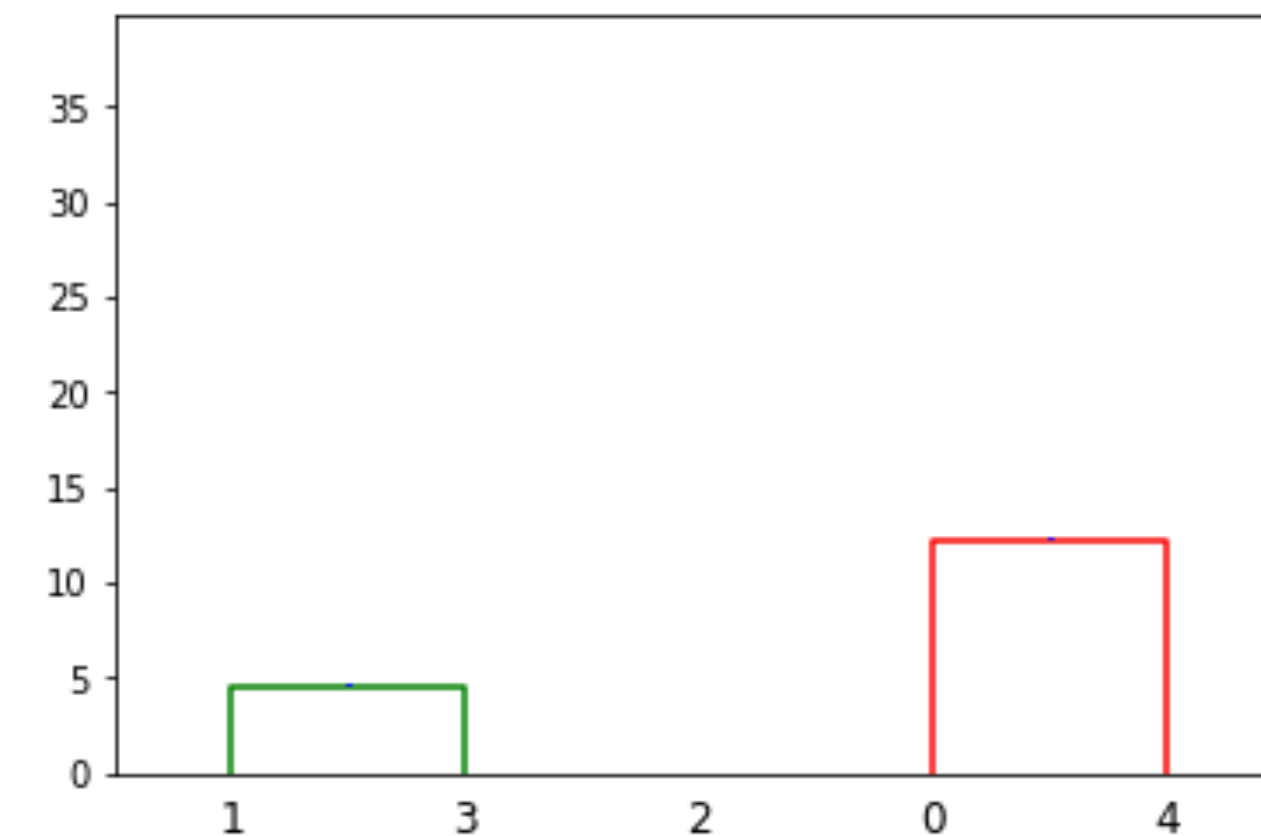
Example: Step 2

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

Cluster	0, 4	1, 3	2
0, 4	0		
1, 3	?	0	
2	?	71.1	0

We will repeat the process that we did in step 1

We merge these two clusters at distance 12.2



$$\begin{aligned}\text{dist}((0, 4), (1, 3)) &= \min (\text{dist}(0, (1, 3)), \text{dist}(4, (1, 3))) \\ &= \min (38.2, 37.9) \\ &= 37.9\end{aligned}$$

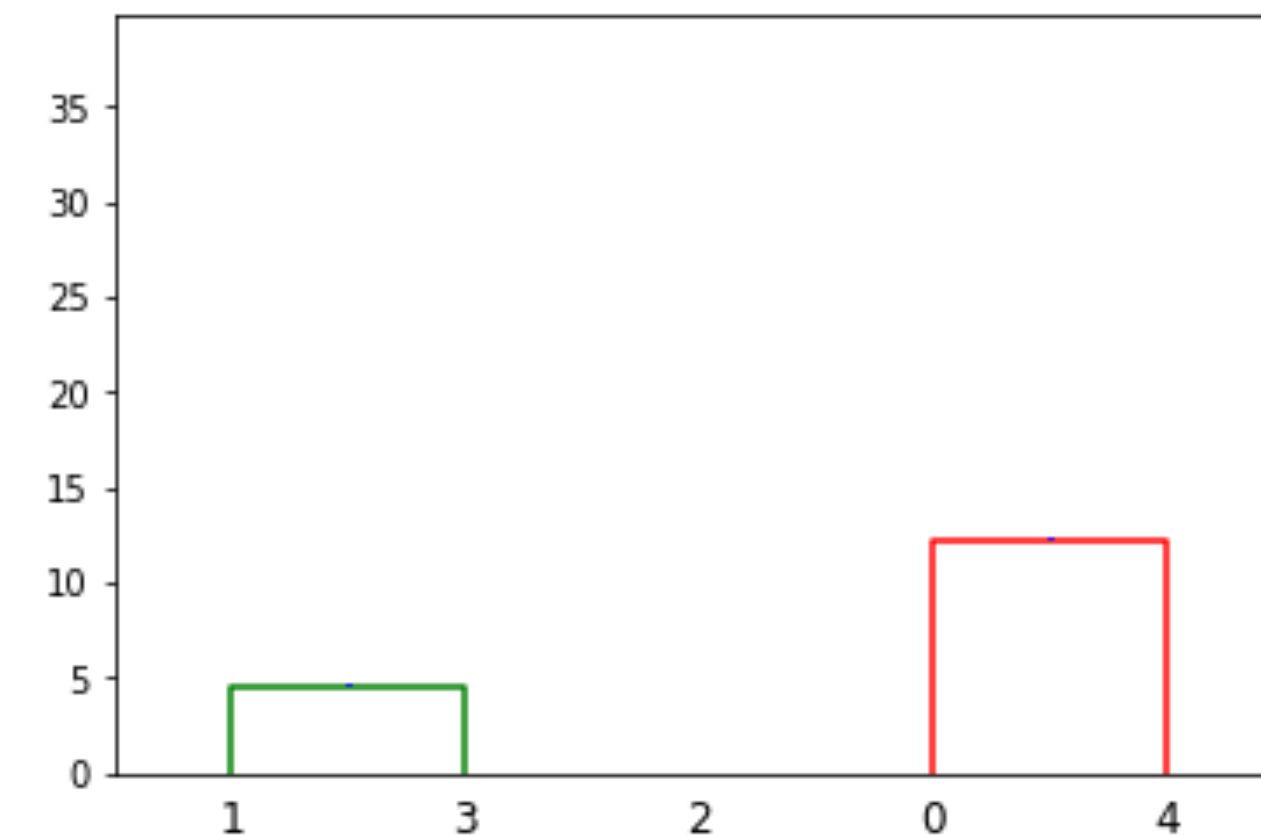
Example: Step 2

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

Cluster	0, 4	1, 3	2
0, 4	0		
1, 3	37.9	0	
2	?	71.1	0

We will repeat the process that we did in step 1

We merge these two clusters at distance 12.2



$$\begin{aligned}\text{dist}((0, 4), (1, 3)) &= \min (\text{dist}(0, (1, 3)), \text{dist}(4, (1, 3))) \\ &= \min (38.2, 37.9) \\ &= 37.9\end{aligned}$$

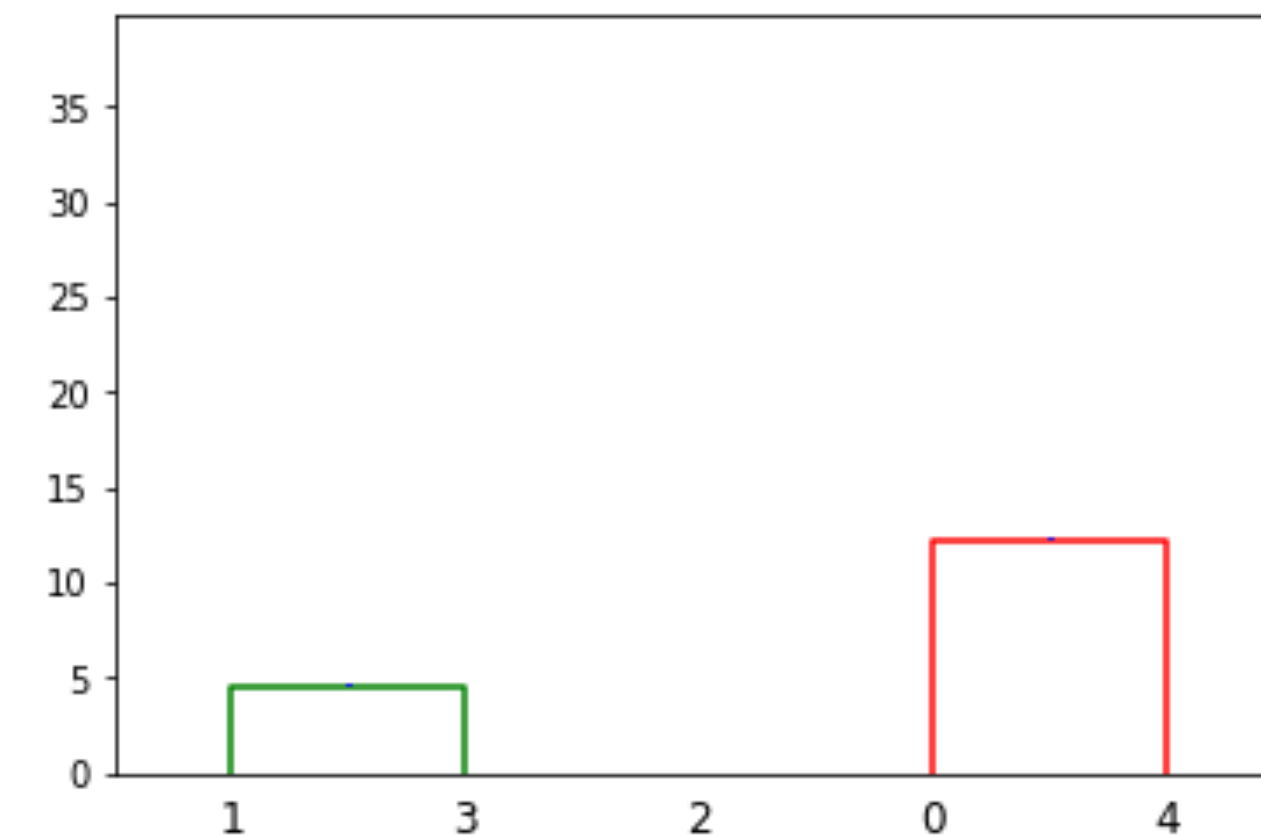
Example: Step 2

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

Cluster	0, 4	1, 3	2
0, 4	0		
1, 3	37.9	0	
2	33.0	71.1	0

We will repeat the process that we did in step 1

We merge these two clusters at distance 12.2



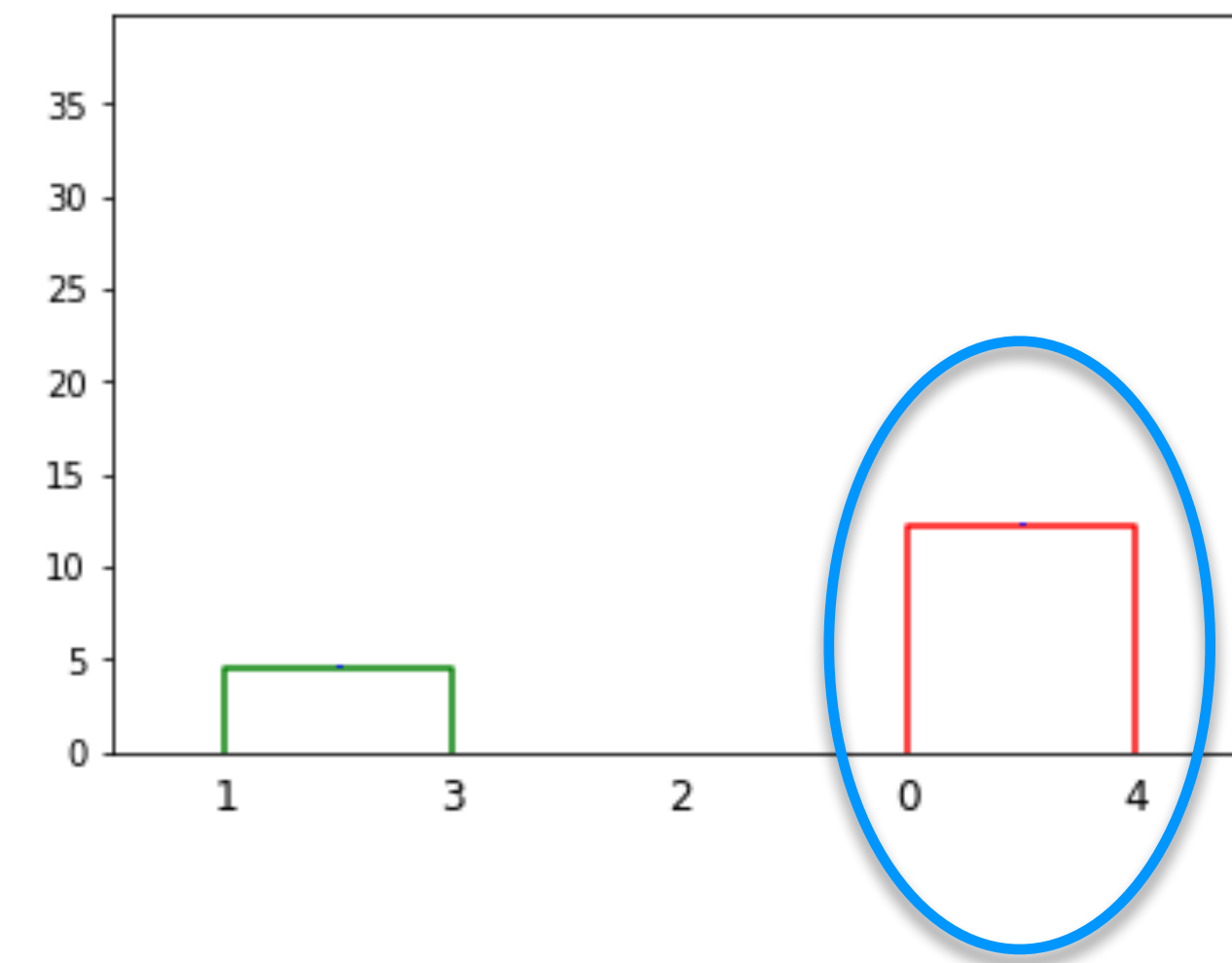
Example: Step 2

Cluster	0	1, 3	2	4
0	0			
1, 3	38.2	0		
2	33.0	71.1	0	
4	12.2	37.9	35.7	0

Cluster	0, 4	1, 3	2
0, 4	0		
1, 3	37.9	0	
2	33.0	71.1	0

We will repeat the process that we did in step 1

We merge these two clusters at distance 12.2



Example: Step 2

New symmetric distance matrix with dimension 3 x 3

Cluster	0, 4	1, 3	2
0, 4	0		
1, 3	37.9	0	
2	33.0	71.1	0

Example: Step 3

Cluster	0, 4	1, 3	2
0, 4	0.0		
1, 3	37.9	0.0	
2	33.0	71.1	0.0

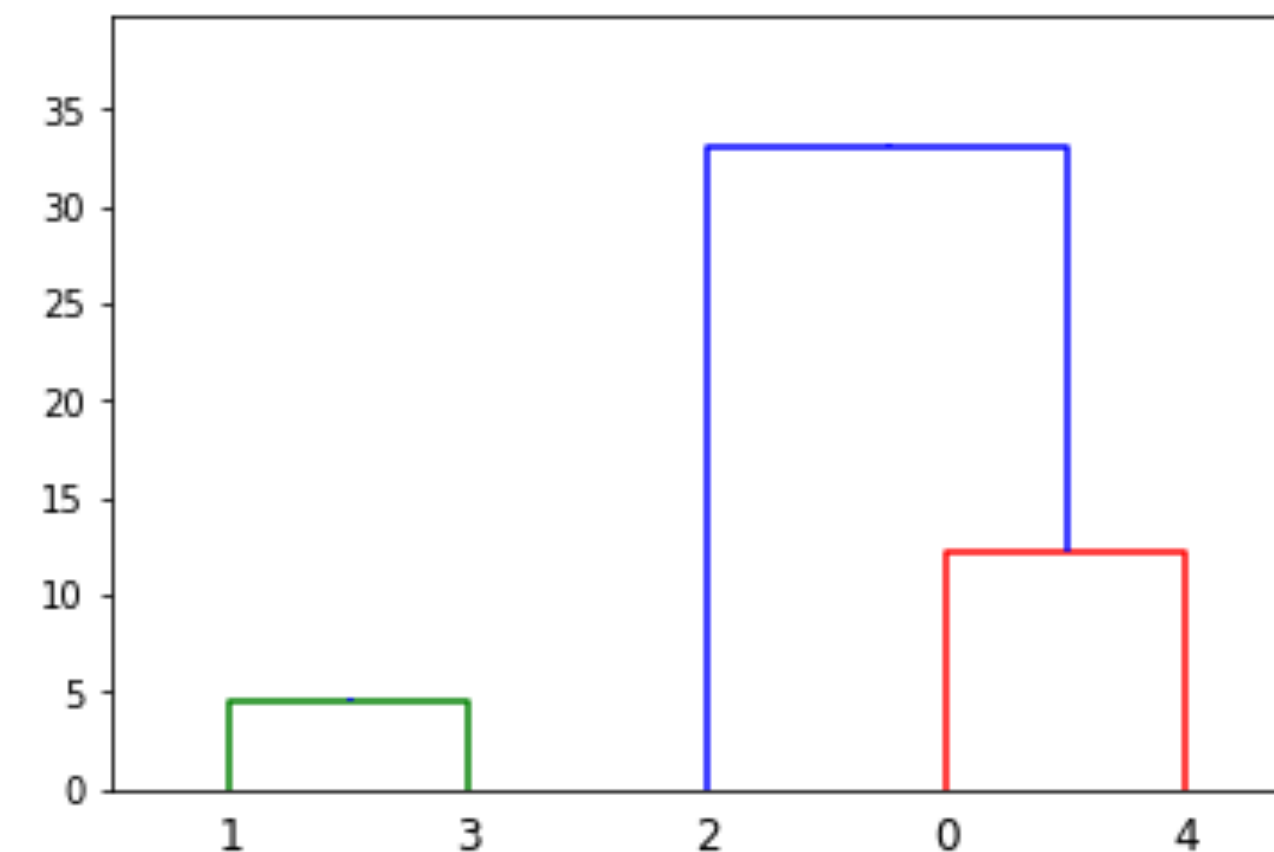
Cluster 2 and Cluster (0,4) are merged at distance 33.0

They formed a new cluster (2, 0, 4)

Discard columns and rows corresponding cluster 2 and cluster (0,4)

Compute distances between existing clusters and newly formed cluster.

Cluster	2, 0, 4	1, 3
2, 0, 4	0.0	
1, 3	37.9	0.0

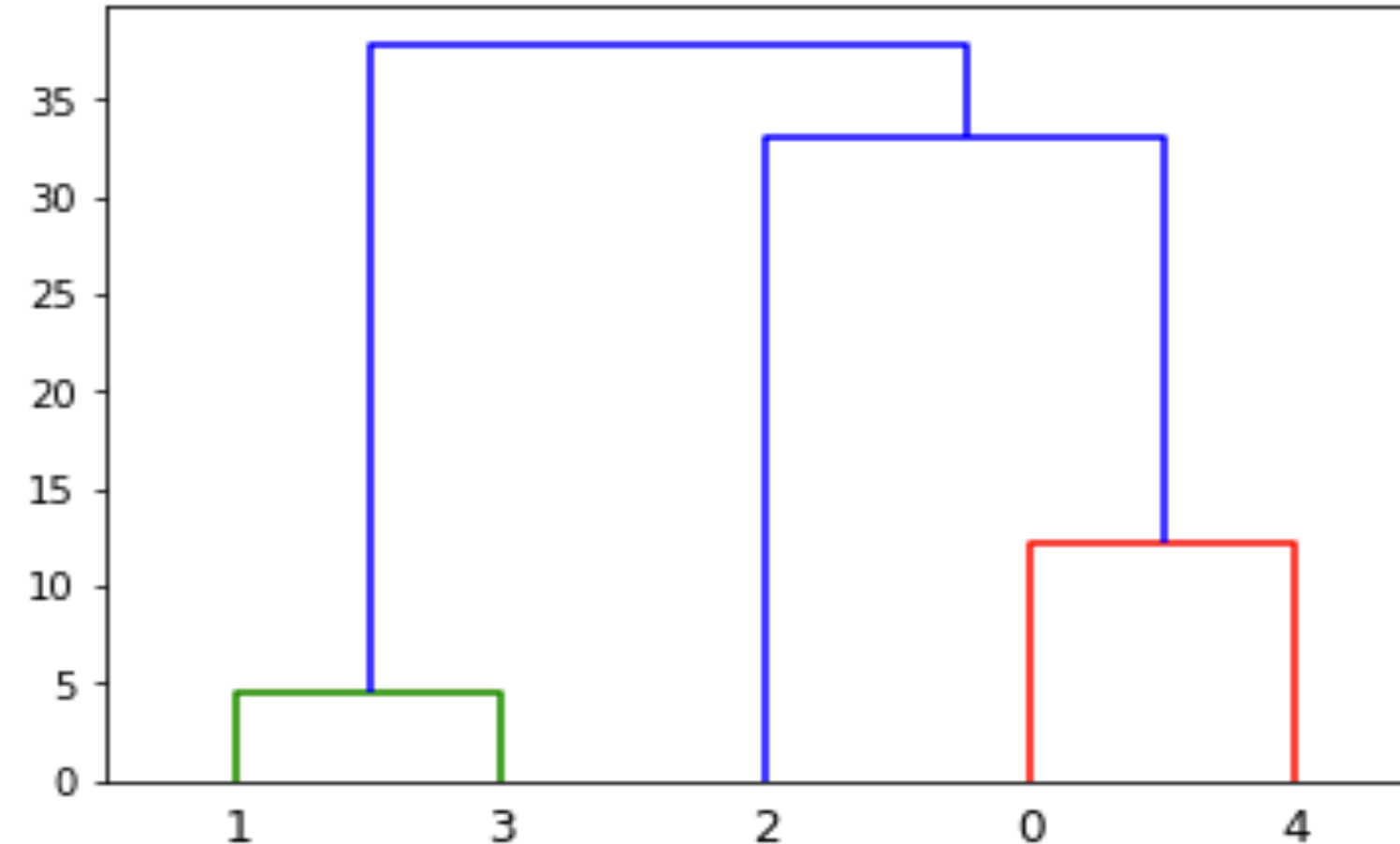


Example: Step 4

Cluster	2, 0, 4	1, 3
2, 0, 4	0.0	
1, 3	37.9	0.0

Cluster (1, 3) and Cluster (2, 0, 4) are merged at distance 37.9

Final cluster consists of all observations (1, 3, 2, 0, 4)



That's all for today. This will be followed by several linkage methods.