Comparision Results

Parismita Das

5 March 2017

Comparision between rioja and adjClustBand_heap

From the Merge Data Table and Dendrogram of rioja and adjClustBand_heap

Rioja Data Table

	V1	V2	height
1	-30	-31	0.3145317
2	-96	-97	0.6362516
3	-28	-29	1.0764627
4	-127	-128	1.5500365
5	-81	-82	2.2093439
6	-24	-25	2.8824763

From the table we can say that the (30,31) merges at height 0.31, (96,97) at 0.6 and so on.

adjClustBand_heap Data Table

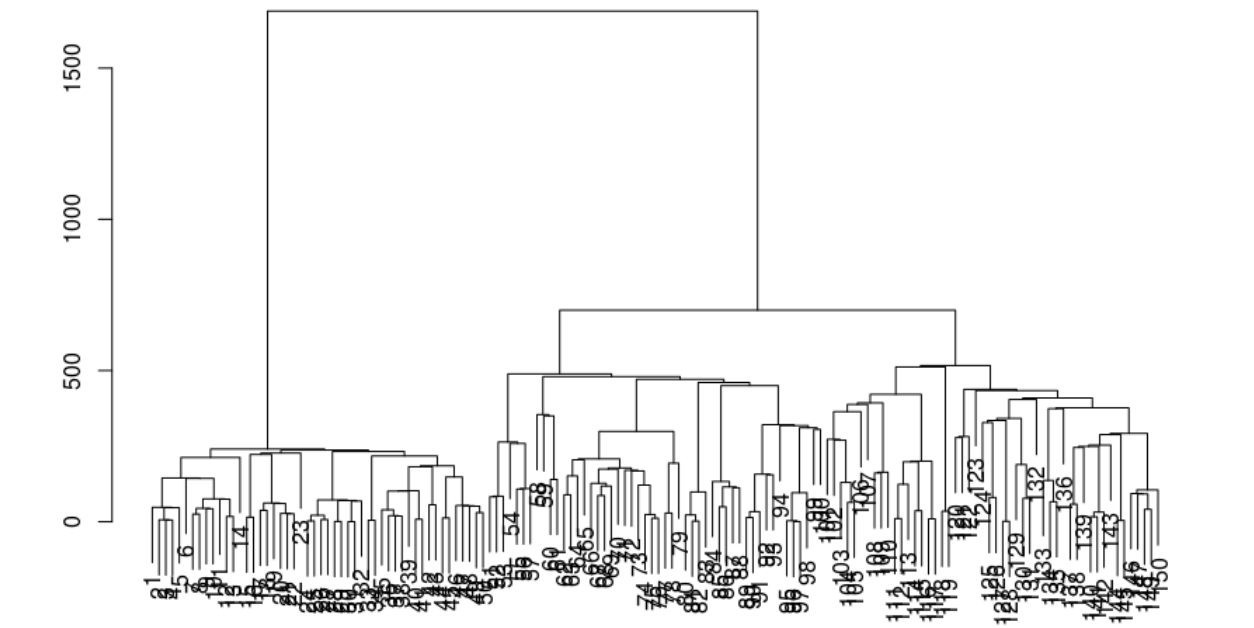
	V1	V2	gain	height
1	-118	-119	-12.452729	0.3145317
2	-131	-132	-8.758996	0.6362516
3	-13	-14	-8.089596	1.0764627
4	-42	-43	-7.771329	1.5500365
5	-22	-23	-7.299329	2.2093439
6	-38	-39	-7.224996	2.8824763

From the table we can say that the (118,119) merges at height 0.31, (131,132) at 0.6 and so on.

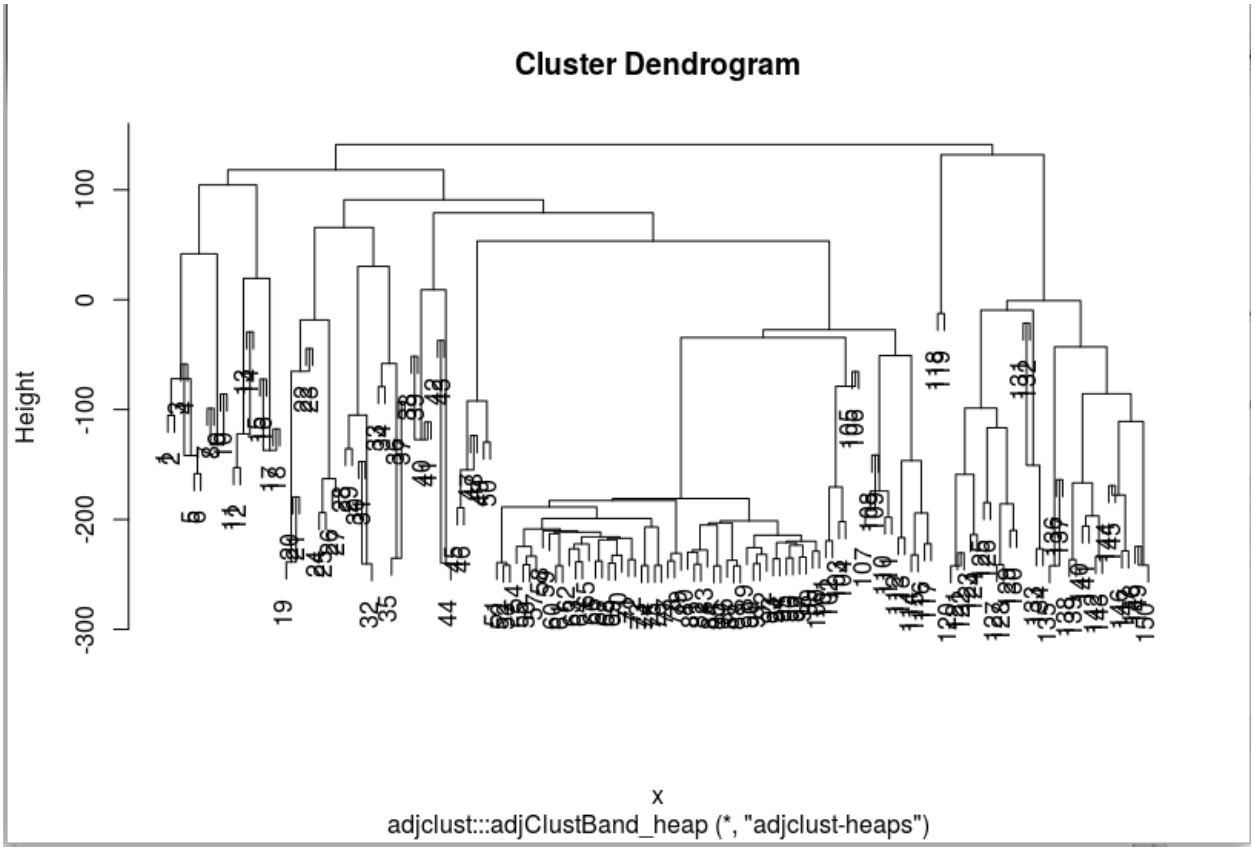
In both the functions we use Ward's method still the cluster output is different because the Distance metric in adjClustBand_heap is calculated considering a section of rectangle of the matrix. And that is minimised.

But in rioja Directly Euclidian distance of the clusters is calculated.

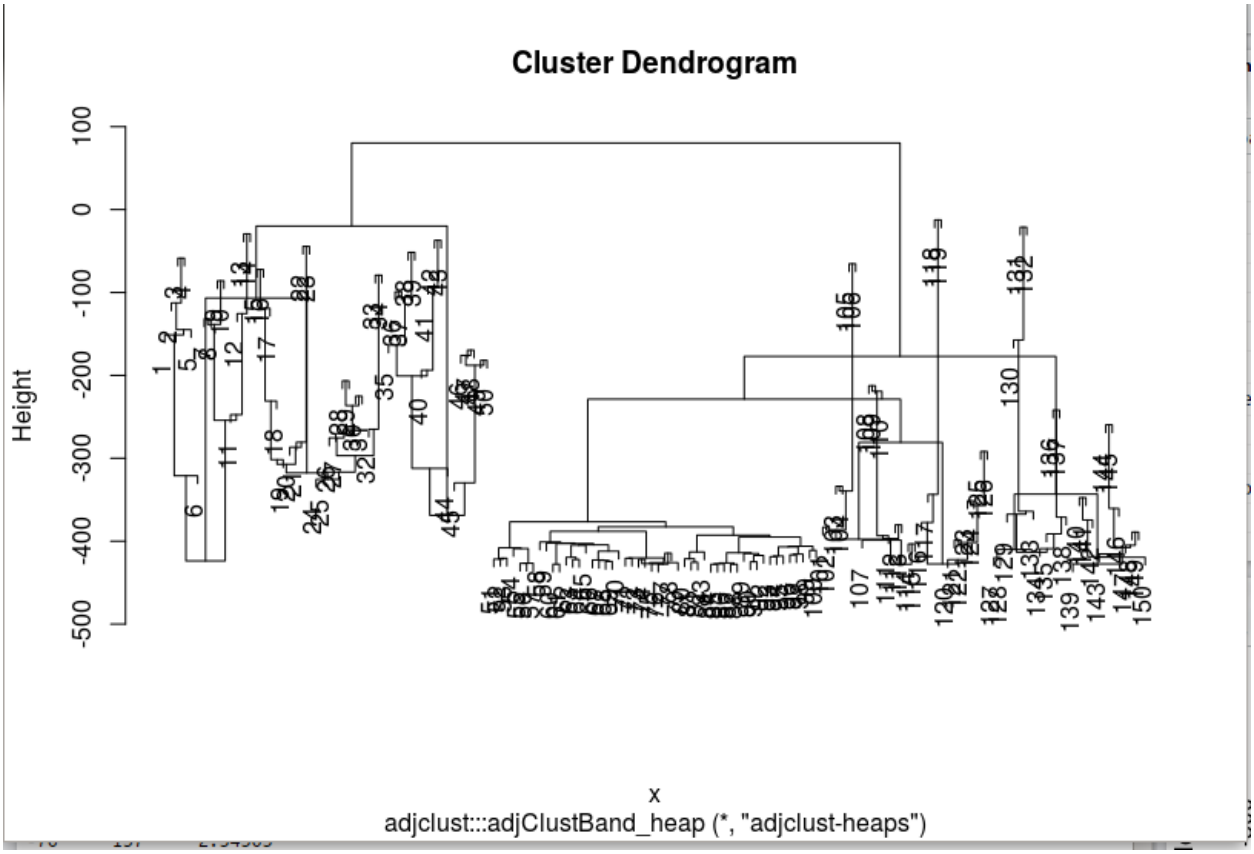
Rioja Dendrogram



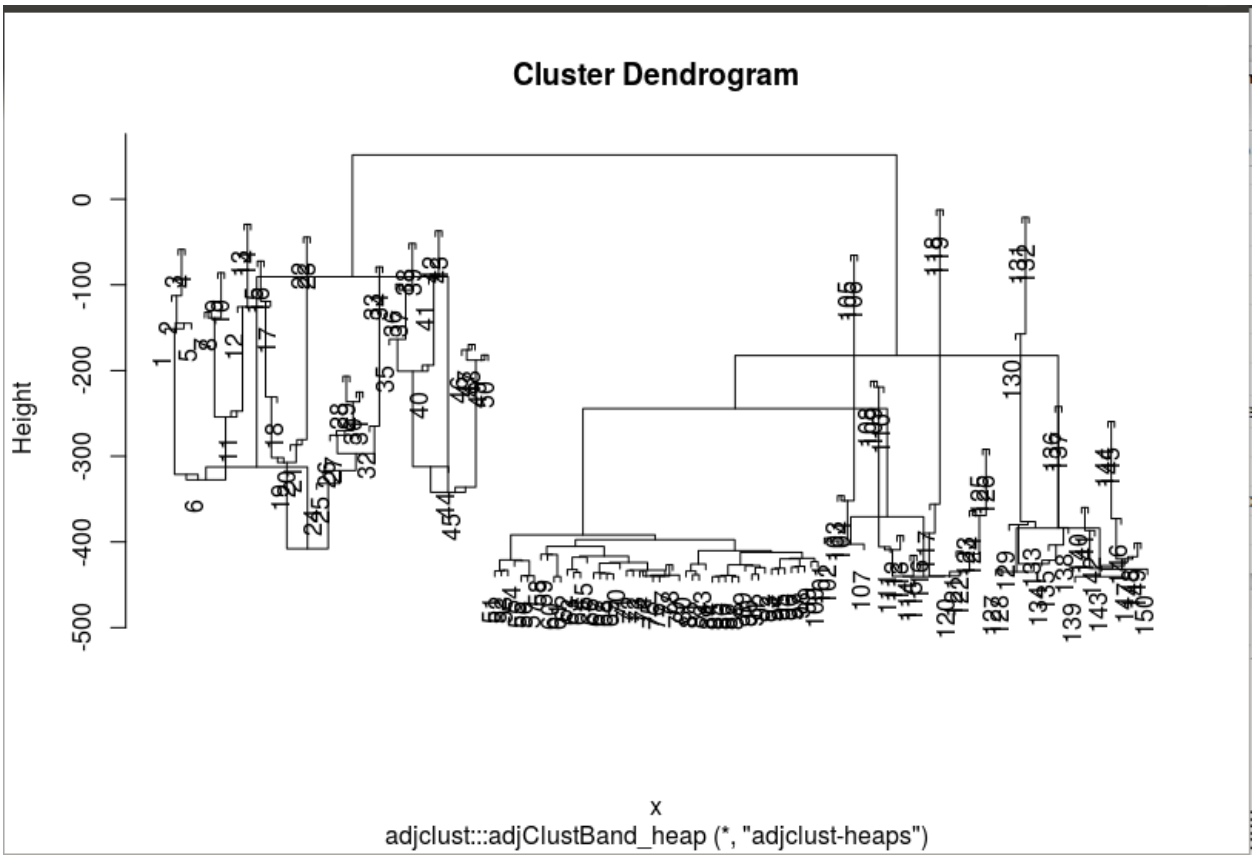
adjClustBand_heap Dendrogram for h = 1



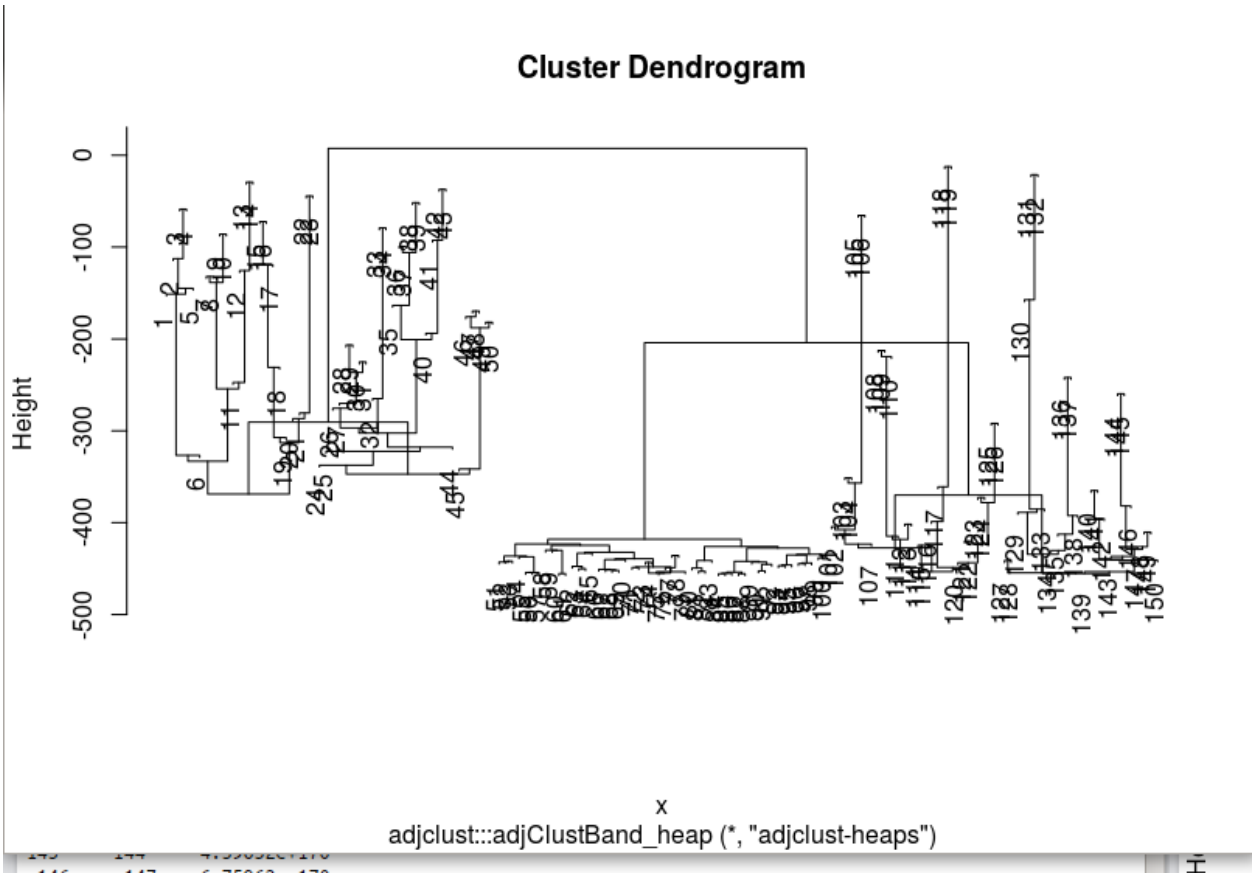
adjClustBand_heap Dendrogram for h = 10



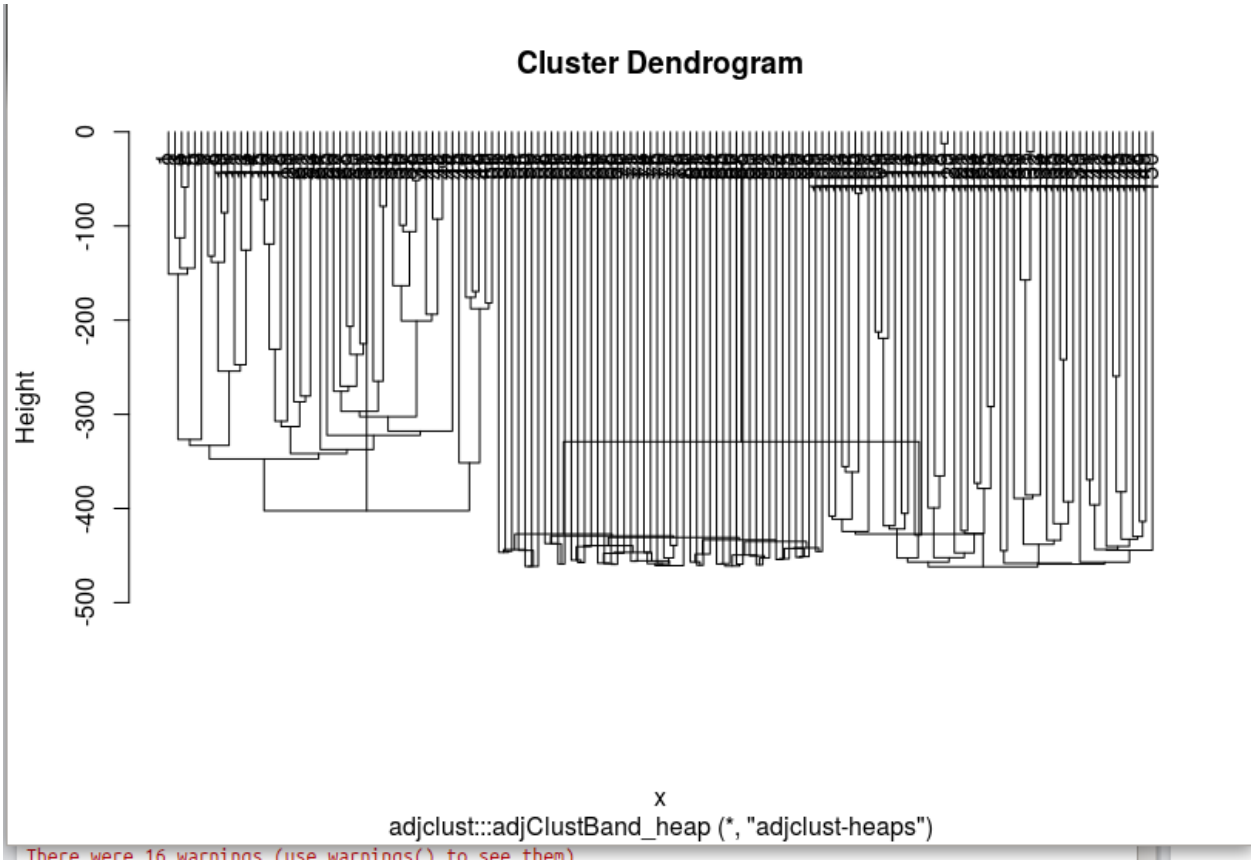
adjClustBand_heap Dendrogram for h = 15



adjClustBand_heap Dendrogram for h = 25



adjClustBand_heap Dendrogram for h = 45



We can say that as the width of diagonal increases, we see a bottomup fashion of clustering. Hence we can say that most of the importance clustering information is found near the diagonal.