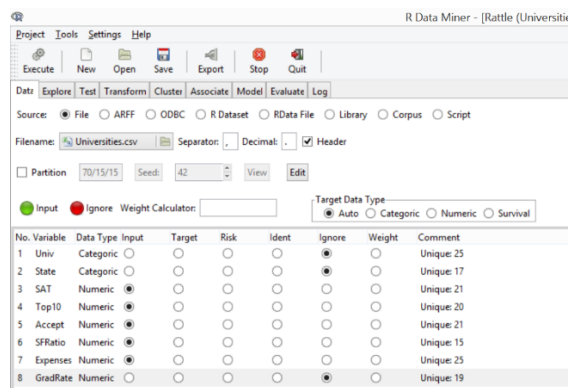


IMPORTANT: We trust that you will not read through this answer key until you have completed your own assignment.

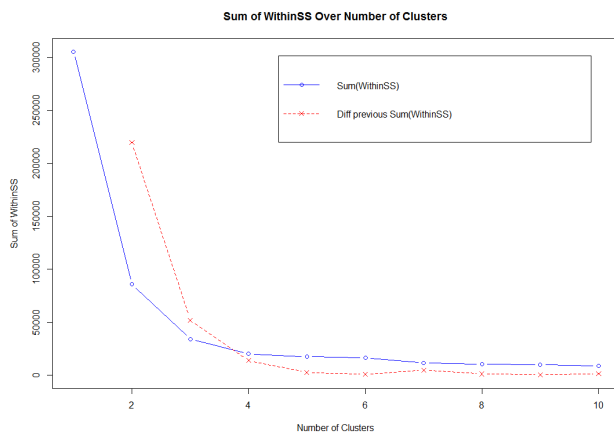
Step 1. Get data in a csv file

Loading of data

Selecting the variables

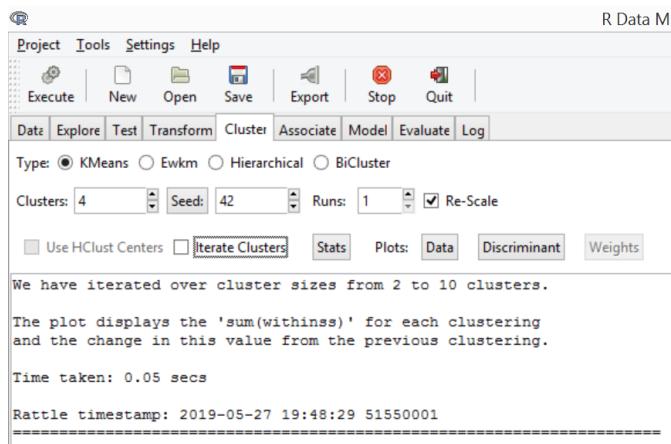


Step 2. Kmeans clustering (iterative up to k=10)



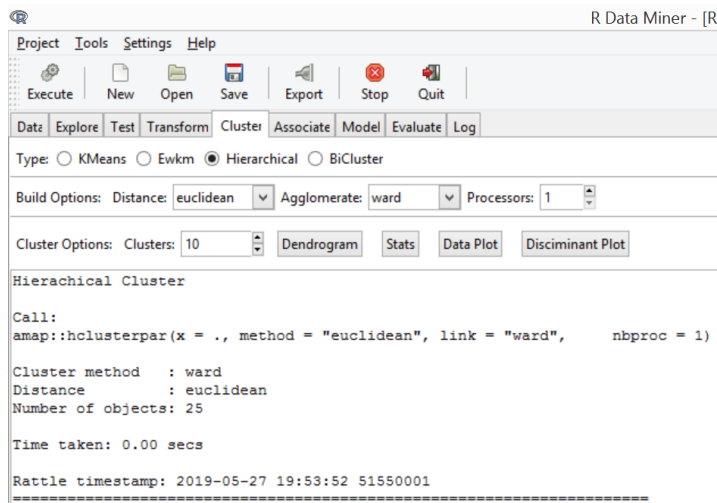
Step 3. Best solution is k=4

Step 4. Running Kmeans for k=4

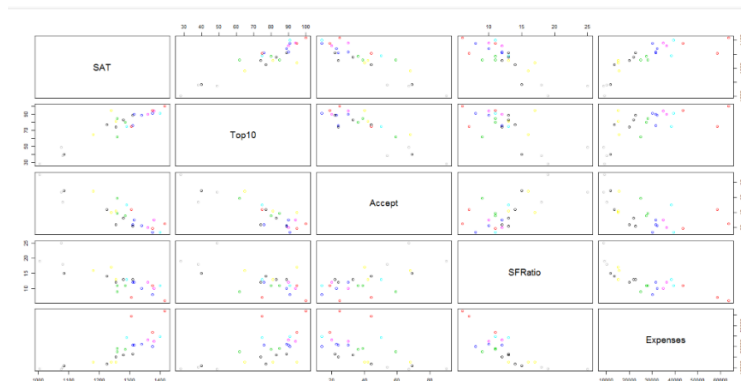


Within cluster sum of squares are: [1] 0.5372319 0.5659548 0.2924521 0.1328684

Step 5. Hierarchical clustering

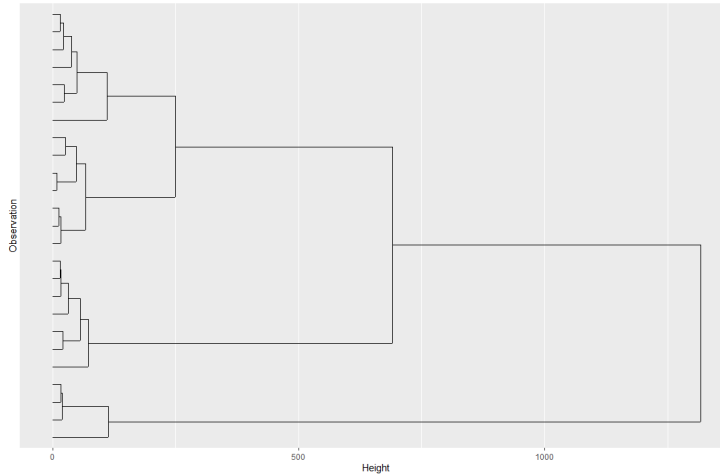


Data Plot: We see we can't even nicely separate 4 clusters.



Dendrogram: Here also we see that we can cut the dendrogram at the red line and consider three clusters. However, please note that there is a lot of subjectivity here and a 4 cluster or even a 2 cluster solution is also equally valid.

Cluster Dendrogram University.csv
Rattle 2019-Jul-08 16:33:31 51550001



Step 6. Run a three cluster Hierarchical clustering again

R Data Miner - [

Project Tools Settings Help

Execute New Open Save Export Stop Quit

Date Explore Test Transform Cluster Associate Model Evaluate Log

Type: ☐ KMeans ☐ Ewkm ☒ Hierarchical ☐ BiCluster

Build Options: Distance: euclidean Agglomerate: ward Processors: 1

Cluster Options: Clusters: 3 Dendrogram Stats Data Plot Discriminant Plot

Hierarchical Cluster

Call:
amap::hclusterpar(x = ., method = "euclidean", link = "ward", nbproc = 1)

Cluster method : ward
Distance : euclidean
Number of objects: 25

Time taken: 0.00 secs

Rattle timestamp: 2019-07-08 16:40:41 51550001

=====
Cluster means:

	S	TI	A	SF	EX
[1,]	1377.857	92.85714	20.71429	9.714286	40.04514
[2,]	1269.286	79.07143	39.64286	12.357143	26.04086
[3,]	1061.500	38.75000	70.00000	19.250000	9.95300

Plot discriminant plot

These four universities (left cluster) make a segment of their own.

Can you identify them?

On what basis are they different from others?

