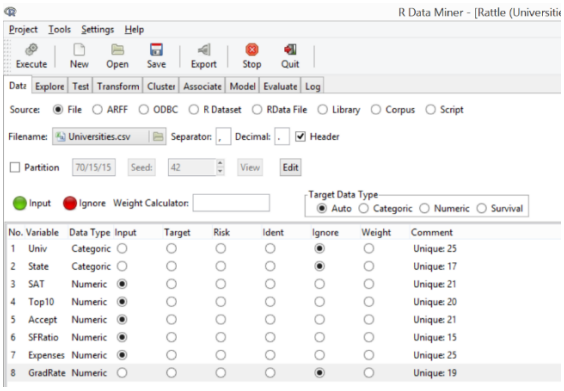
IMPORTANT: We trust that you will not read through this answer key until you have completed your own assignment.

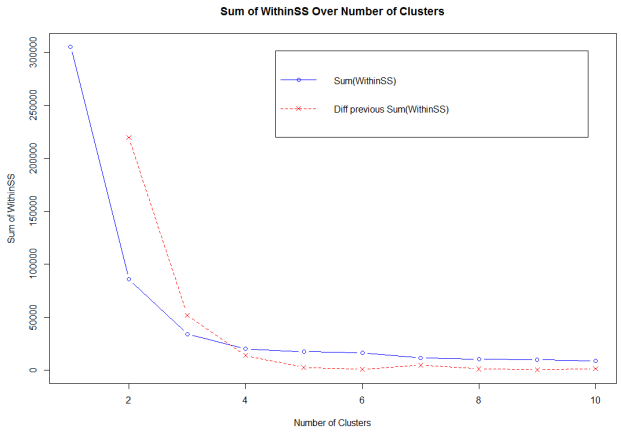
Step 1. Get data in a csv file

Loading of data

Selecting the variables

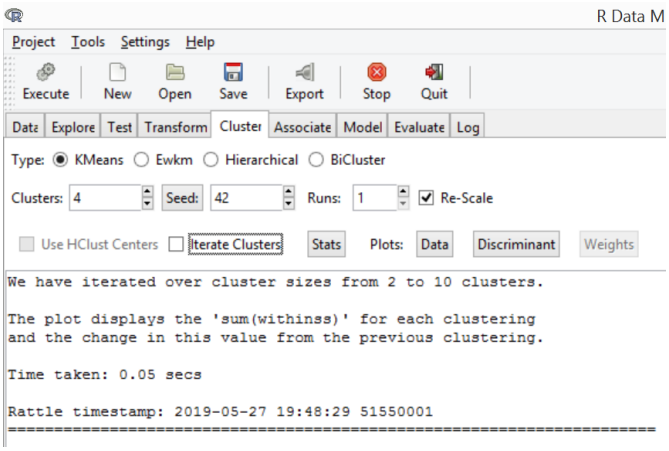


Step 2. Kmeans clustering (iterative up to k=10)



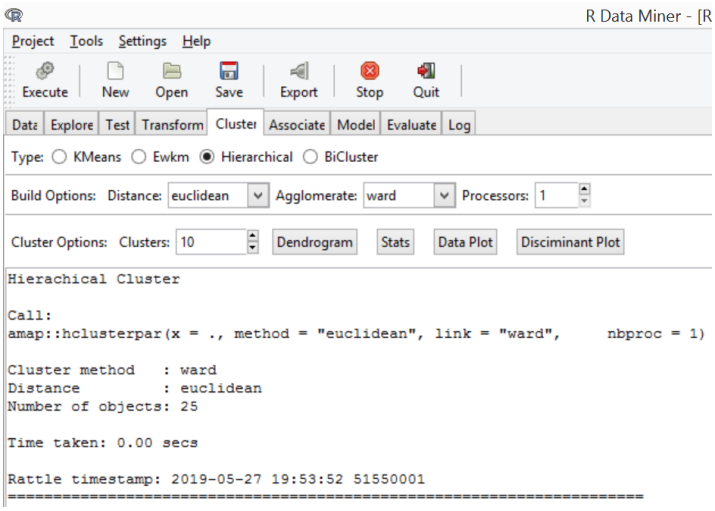
Step 3. Best solution is k=4

Step 4. Running Kmeans for k=4

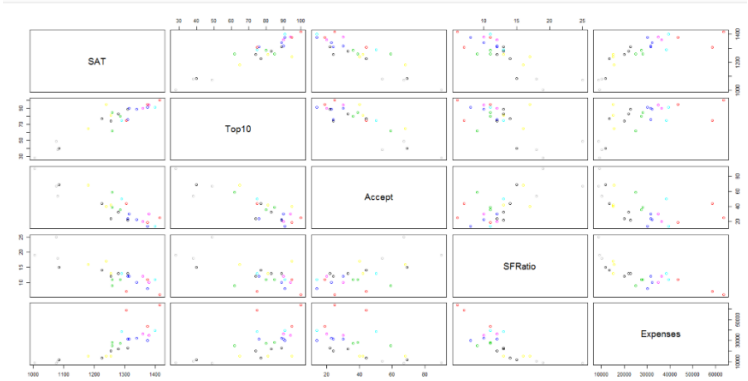


Within cluster sum of squares are: [1] 0.5372319 0.5659548 0.2924521 0.1328684

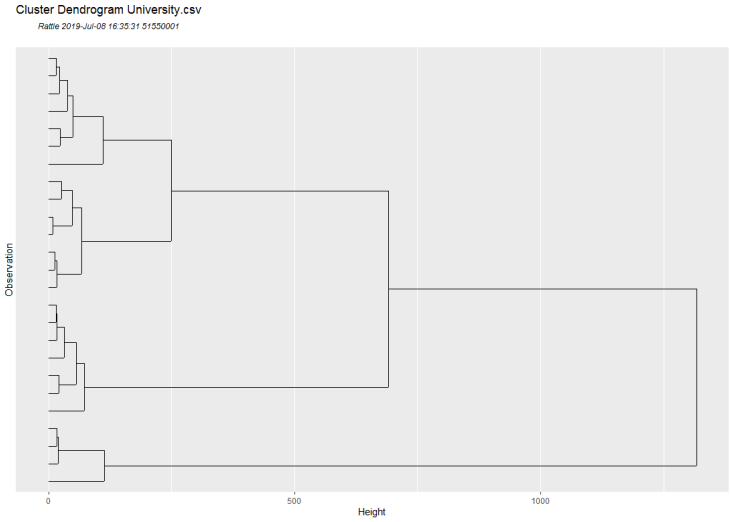
Step 5. Hierarchical clustering



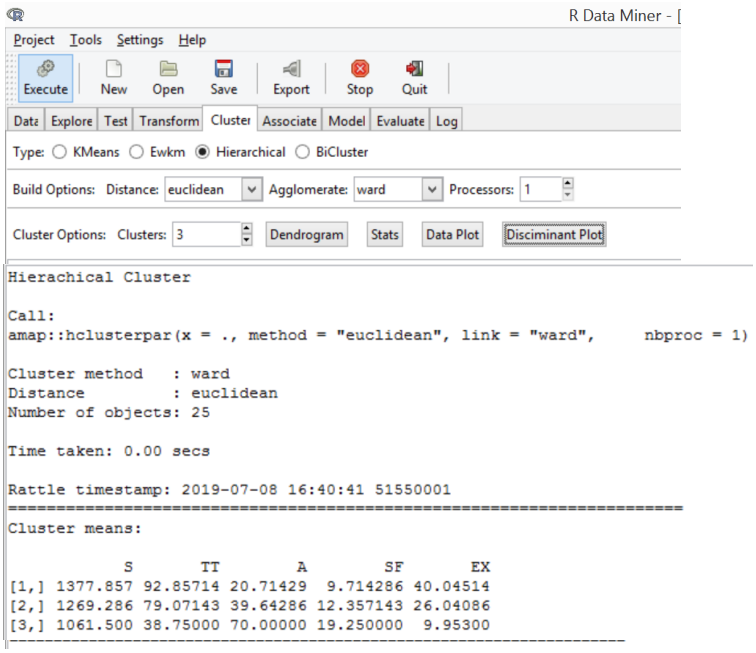
Data Plot: We see we can't even nicely separate 4 clusters.



Dendogram: Here also we see that we can cut the dendogram 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.



Step 6. Run a three cluster Hierarchical clustering again



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?