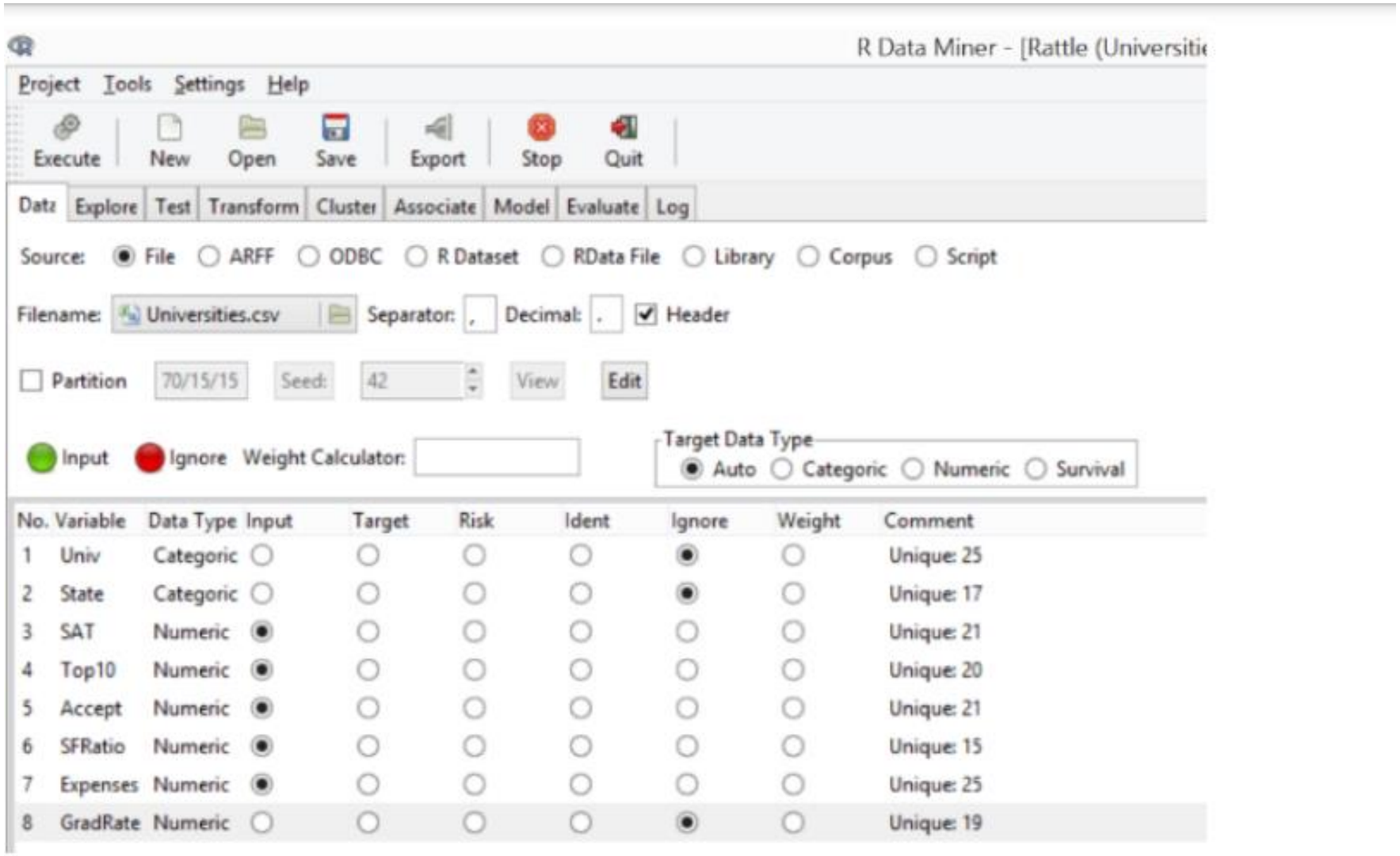
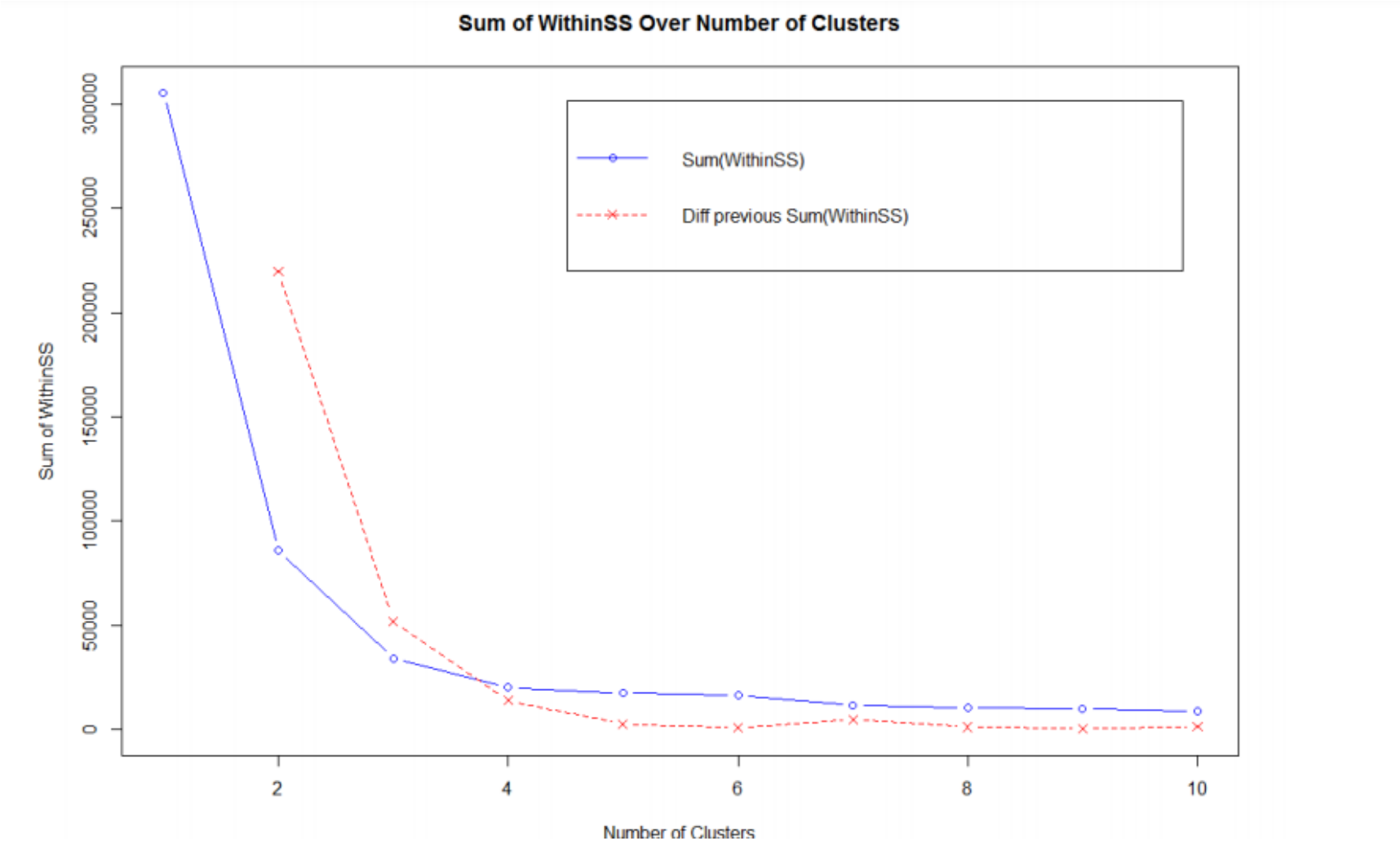
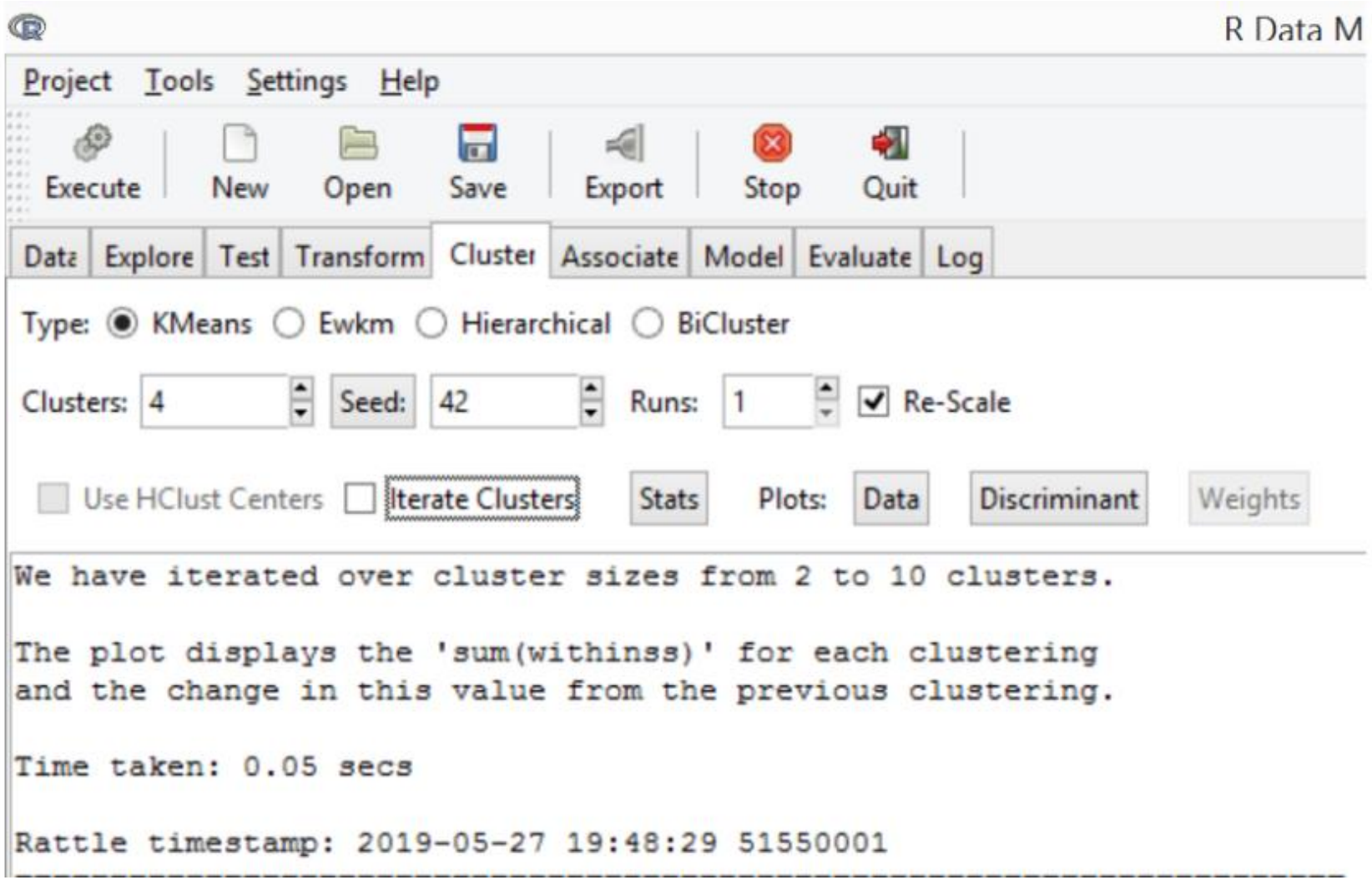


Loading the dataset and selecting the appropriate categories

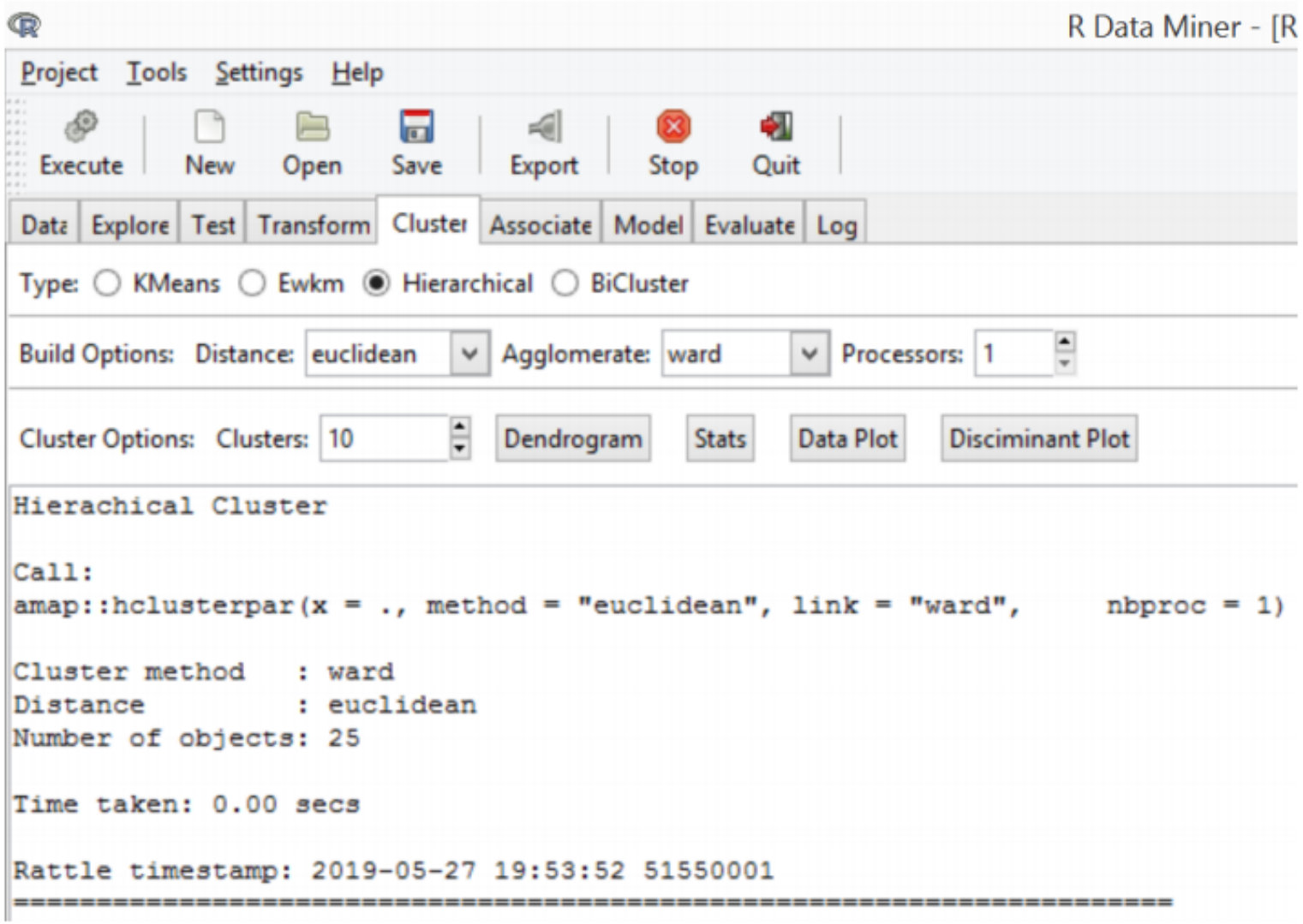
Kmeans clustering graph between Sum and Difference of previous sum

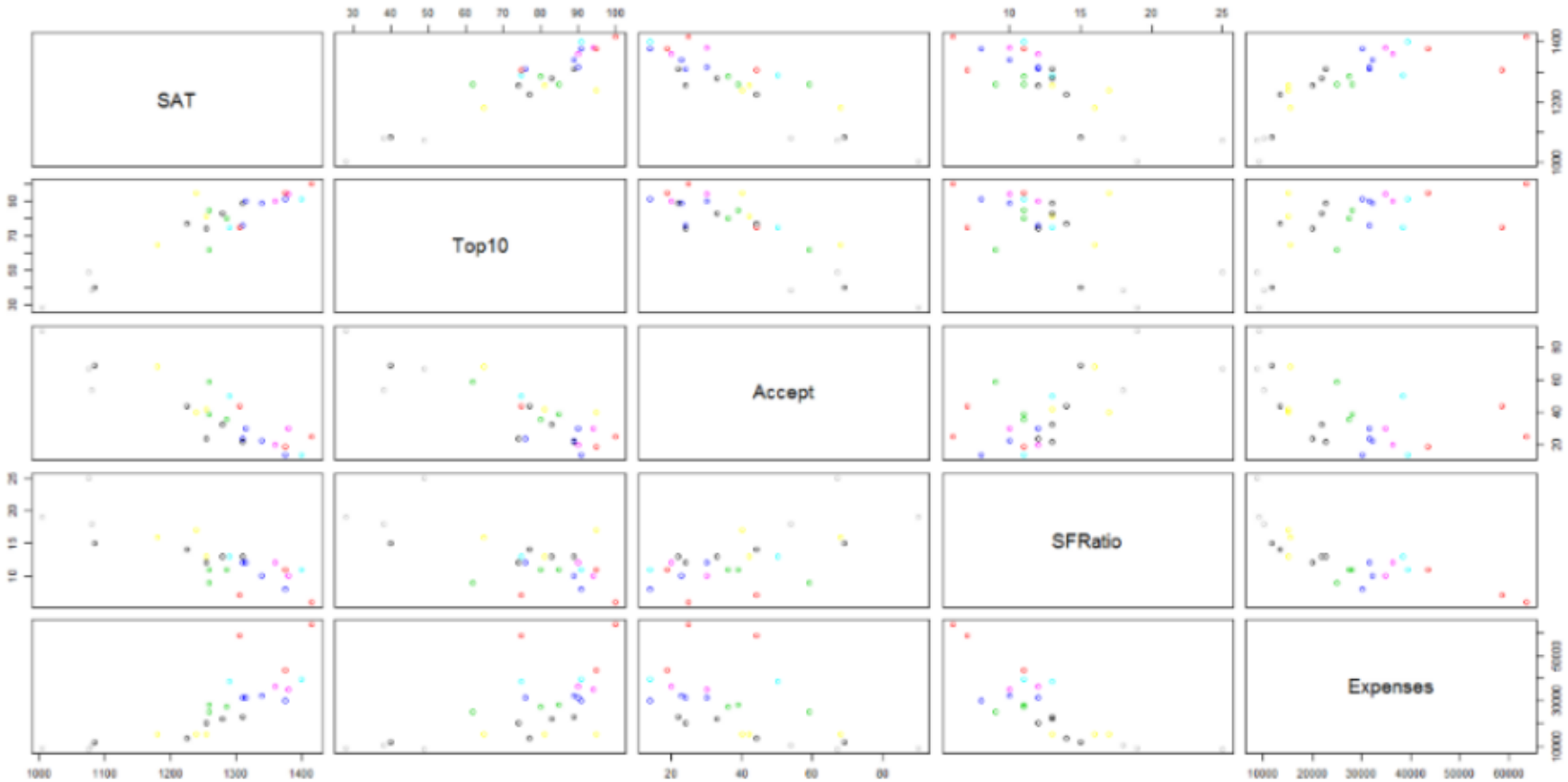


Selecting k = 4 as it is giving the appropriate result.

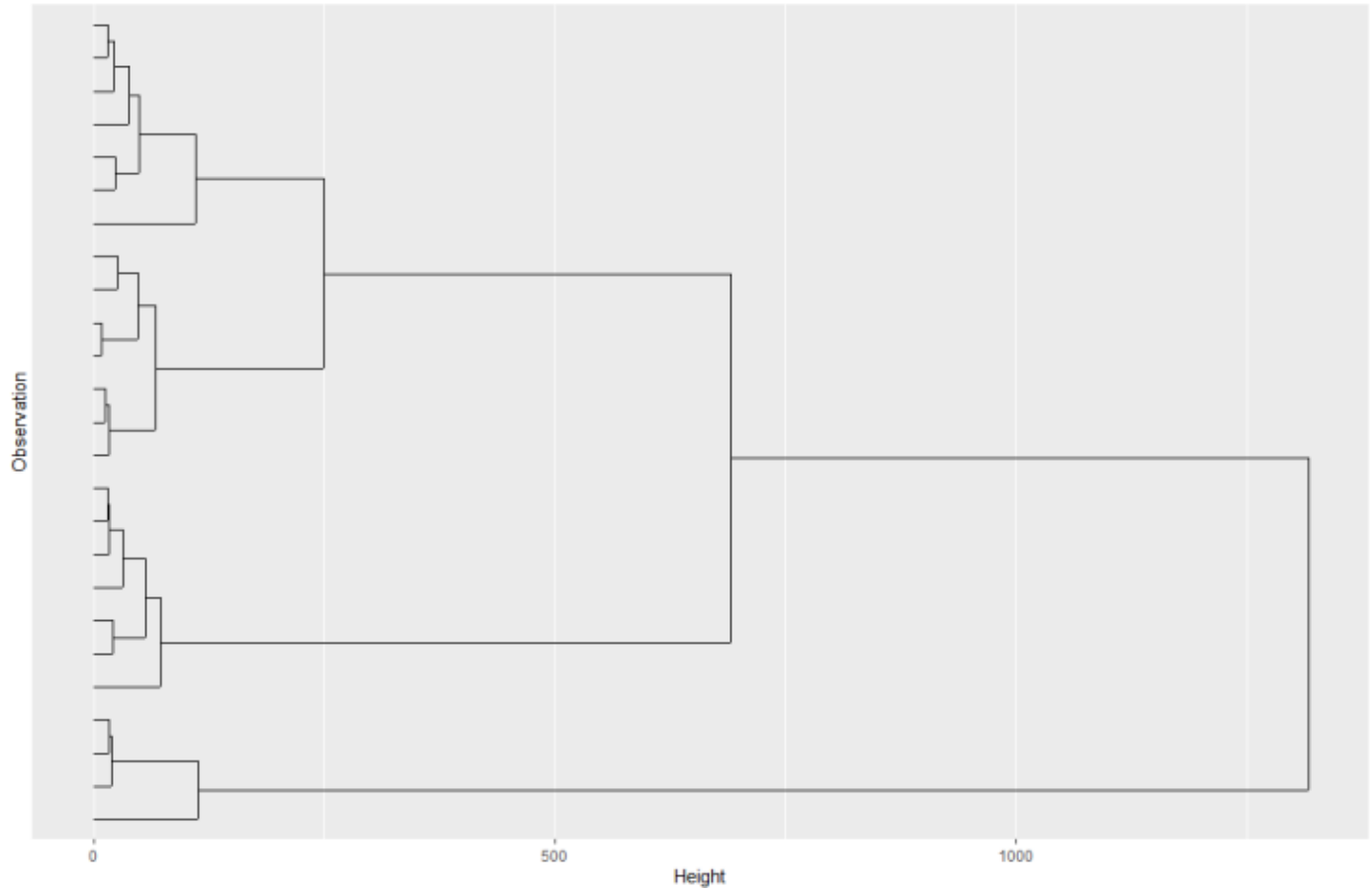


Running the algorithm for k = 4 as selecting the required parameters.

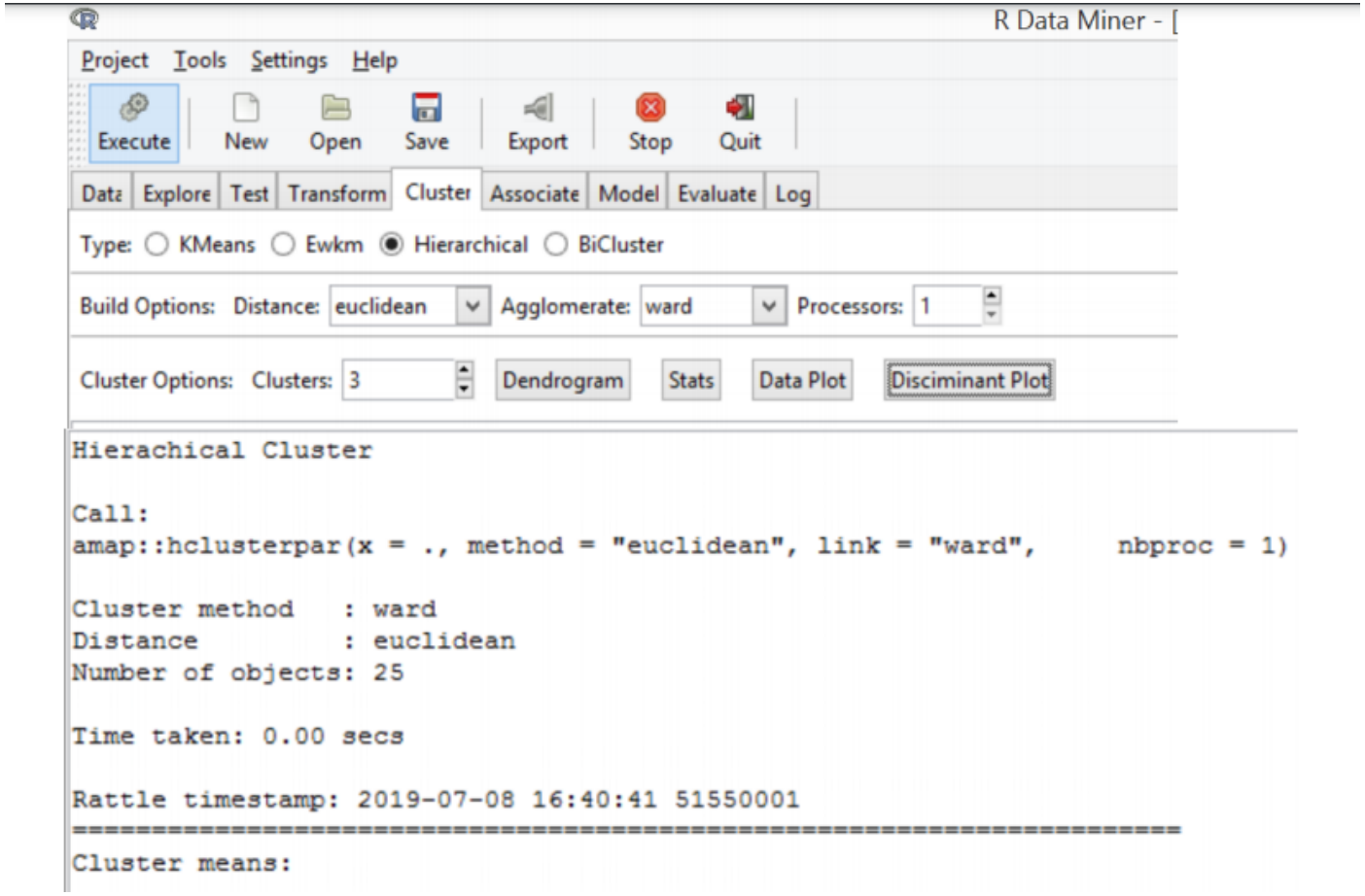



SAT
Top10
Accept
SFRatio
Expenses

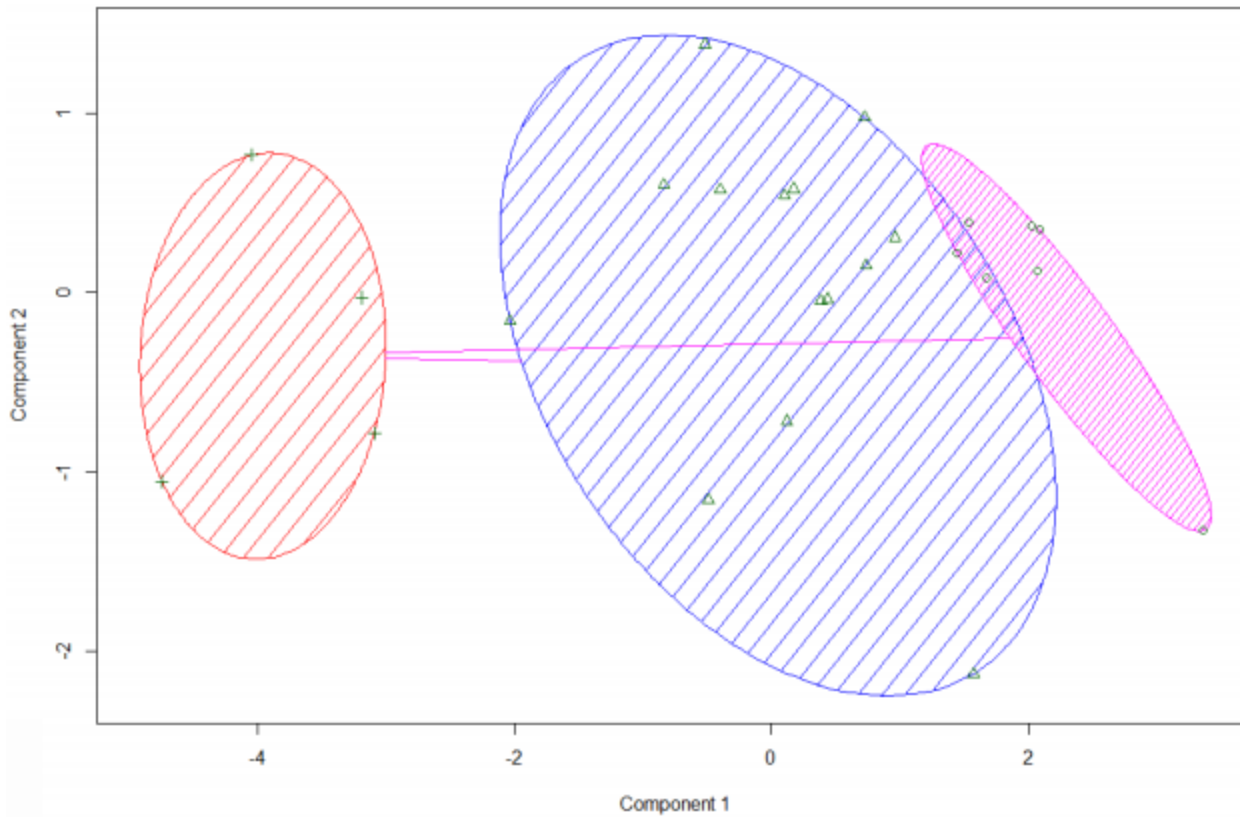
Dendograms



Selecting 3 hierarchical clustering



Discriminant Plot



- Deliverables
- Follow the instruction and provide screenshots of each step and the final results (including KNN, Hierarchical Clustering with data and discriminant plots, and dendogram)
- Briefly explain what you have done in each step and interpret your final result.
- Talk about the challenges you might face or elaborate your thoughts about your findings.

Findings from the project

- First of all I have read the dataset and selecting the appropriate parameters to plot the graph which is shown up in the R studio after analysing the plots we end up selecting the best k value as 4 after iterating through 10 steps, then running k means for k = 4 we obtain the following cluster sum of squares are: 0.3320184, 0.6804389, 0.2820749, and 0.4011498,
- Then we perform Hierarchical clustering by selecting the cluster tab giving 10 clusters and and distance as Euclidean. We can then see the algorithm was not able to distinguish the 4 clusters as seen from the plot.
- Then we obtain the plot for dendogram and see that a 4 cluster or even a 2 cluster solution is also equally valid after running hierarchical clustering again and finally we plot the discriminant plot by selecting the appropriate tab in which component 1 and component 2 explains 92.36% of variability.