IBM Software

Clustering Course
Lab 5: Using clustering methods for asset selection

Contents

Hello everybody and welcome to another lab session!

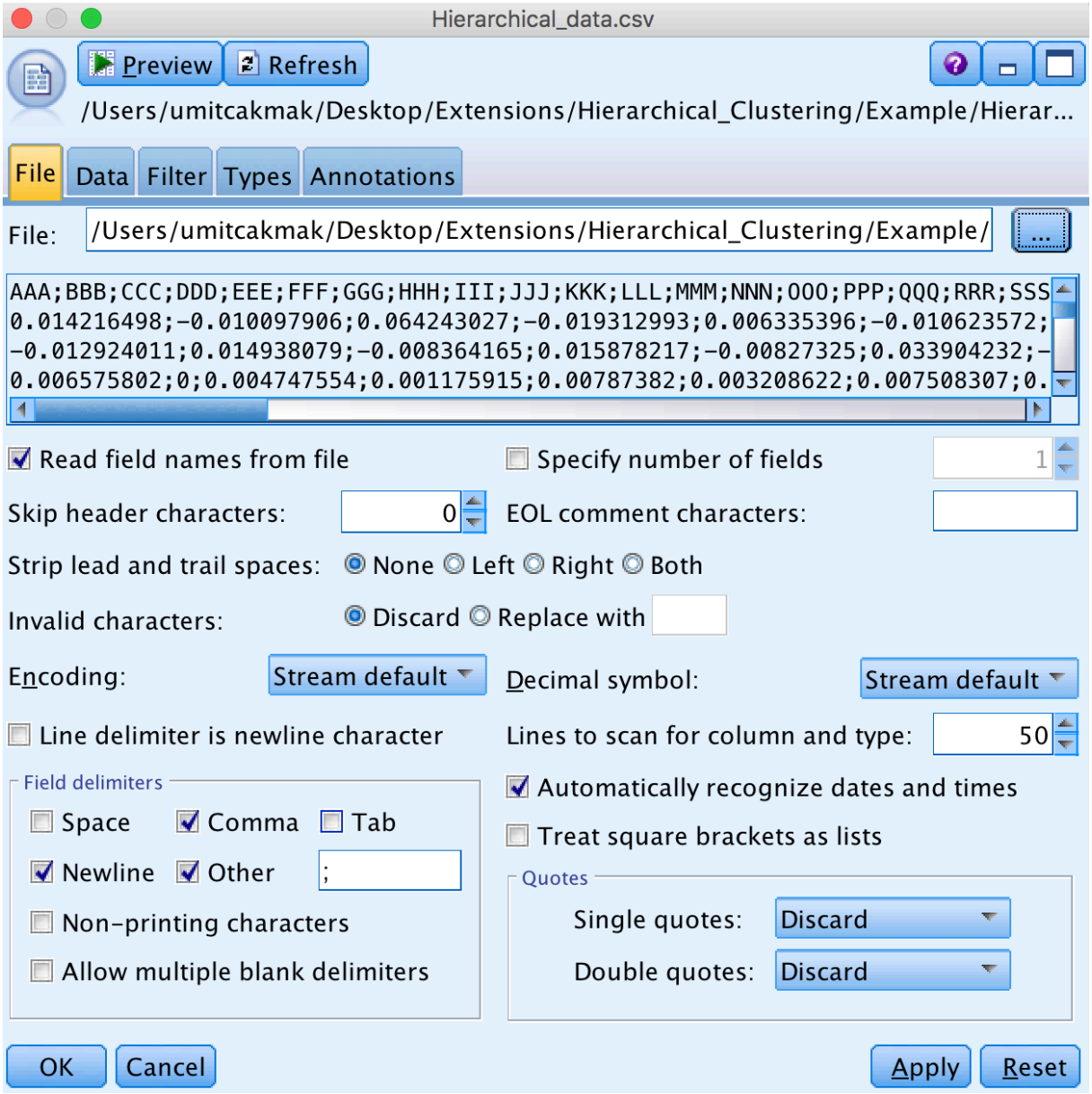
In this lab, we are going to learn how to apply hierarchical clustering by using external SPSS Modeler extension.

Let’s get started!

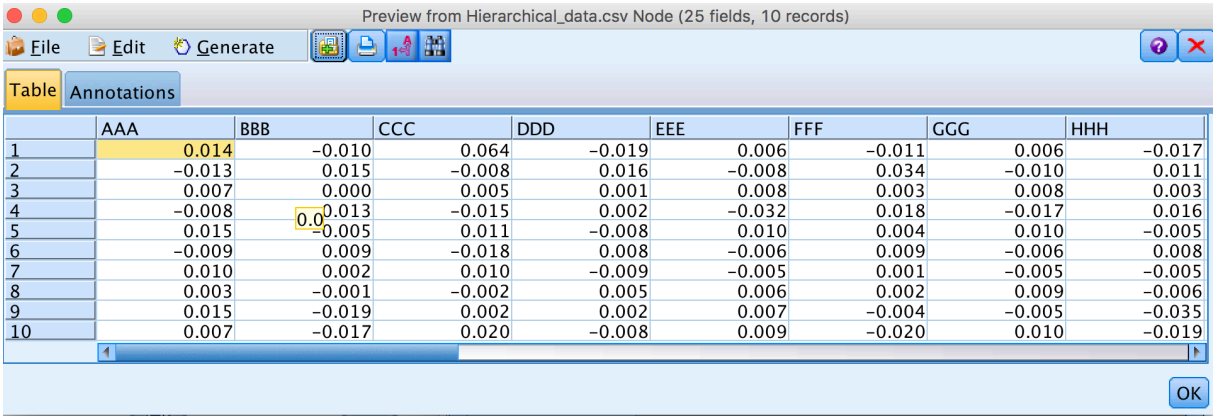
1. Data import

We will start by opening a new stream. I will add “Var. File” node from sources palette.

I will browse file “Hierarchical_data.csv” which includes return series for different assets.
I will set check the box “Other” in “Field delimiters” section to “;” since that’s field separator.

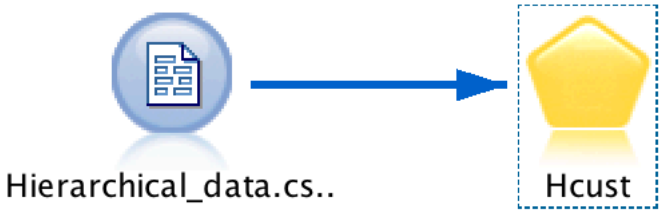


I can click preciew to see preview of the series.

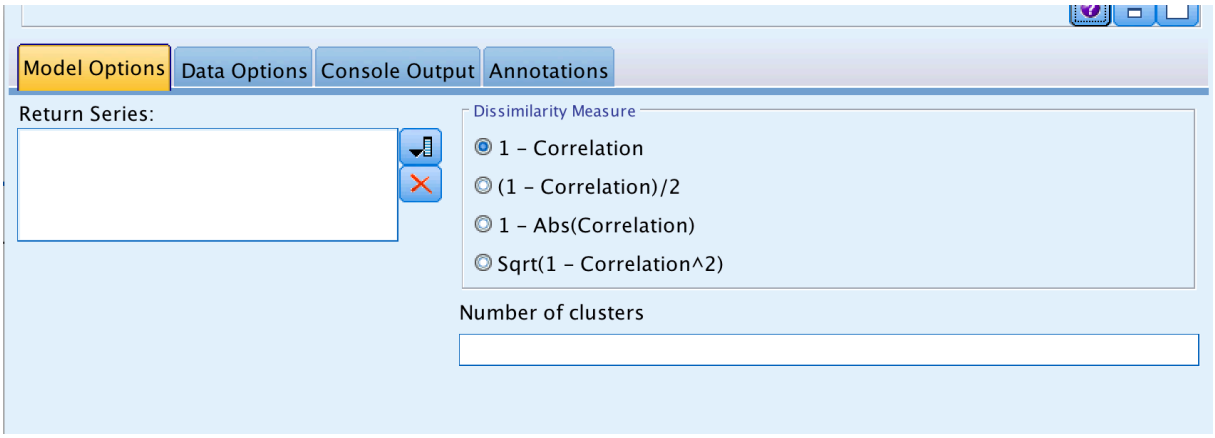


	AAA	BBB	CCC	DDD	EEE	FFF	GGG	HHH
1	0.014	-0.010	0.064	-0.019	0.006	-0.011	0.006	-0.017
2	-0.013	0.015	-0.008	0.016	-0.008	0.034	-0.010	0.011
3	0.007	0.000	0.005	0.001	0.008	0.003	0.008	0.003
4	-0.008	0.013	-0.015	0.002	-0.032	0.018	-0.017	0.016
5	0.015	-0.005	0.011	-0.008	0.010	0.004	0.010	-0.005
6	-0.009	0.009	-0.018	0.008	-0.006	0.009	-0.006	0.008
7	0.010	0.002	0.010	-0.009	-0.005	0.001	-0.005	-0.005
8	0.003	-0.001	-0.002	0.005	0.006	0.002	0.009	-0.006
9	0.015	-0.019	0.002	0.002	0.007	-0.004	-0.005	-0.035
10	0.007	-0.017	0.020	-0.008	0.009	-0.020	0.010	-0.019

Once I'm done with importing data, I can add "Hcust" from "Modeling" palette.

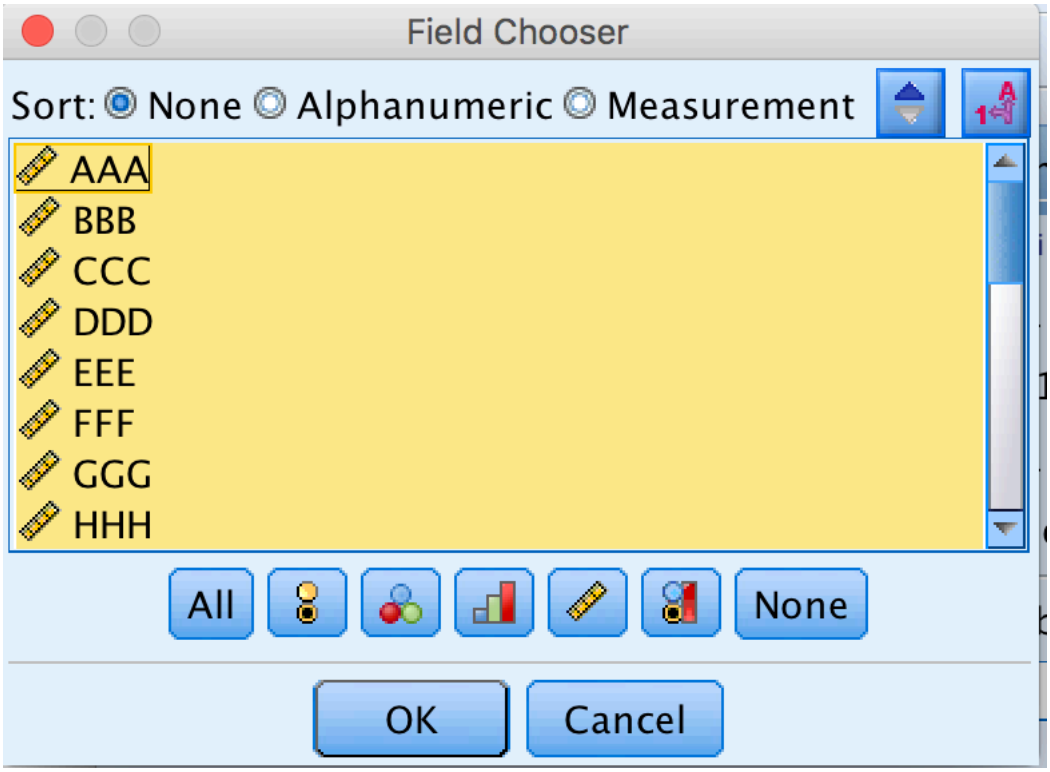


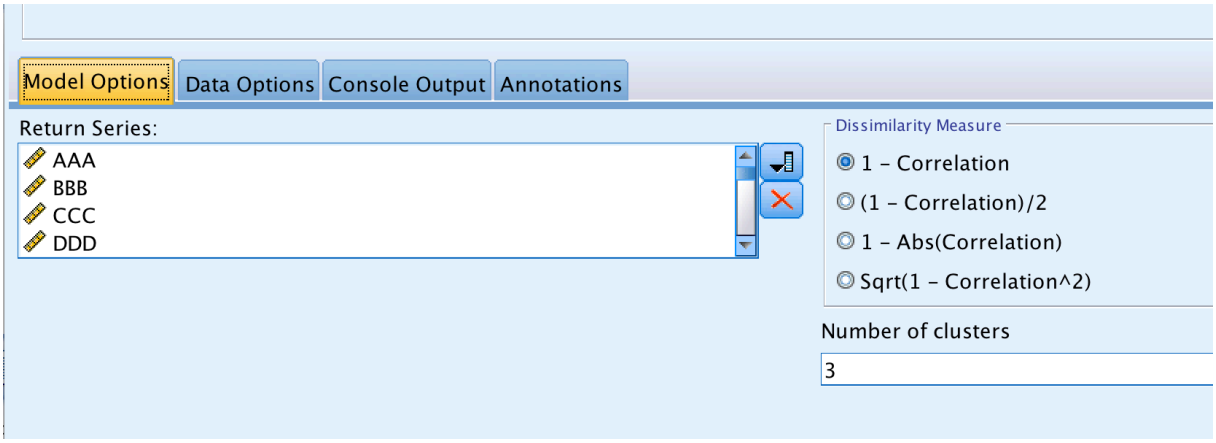
Once I double click "Hcust" node, I see 3 parts.



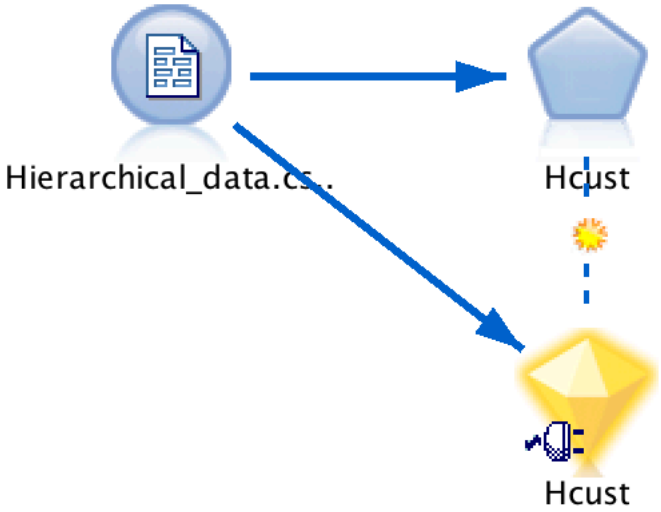
First part is where I select return series to work with, second is selecting dissimilarity measure, we have 4 different dissimilarity measures here and it's good to experiment with it see if result differs. Last and third section is for defining how many clusters I would like to have as a result.

I will select all return series available and I will stick with first dissimilarity measure “1 – Correlation”, and put “3” clusters to “Number of clusters” section.



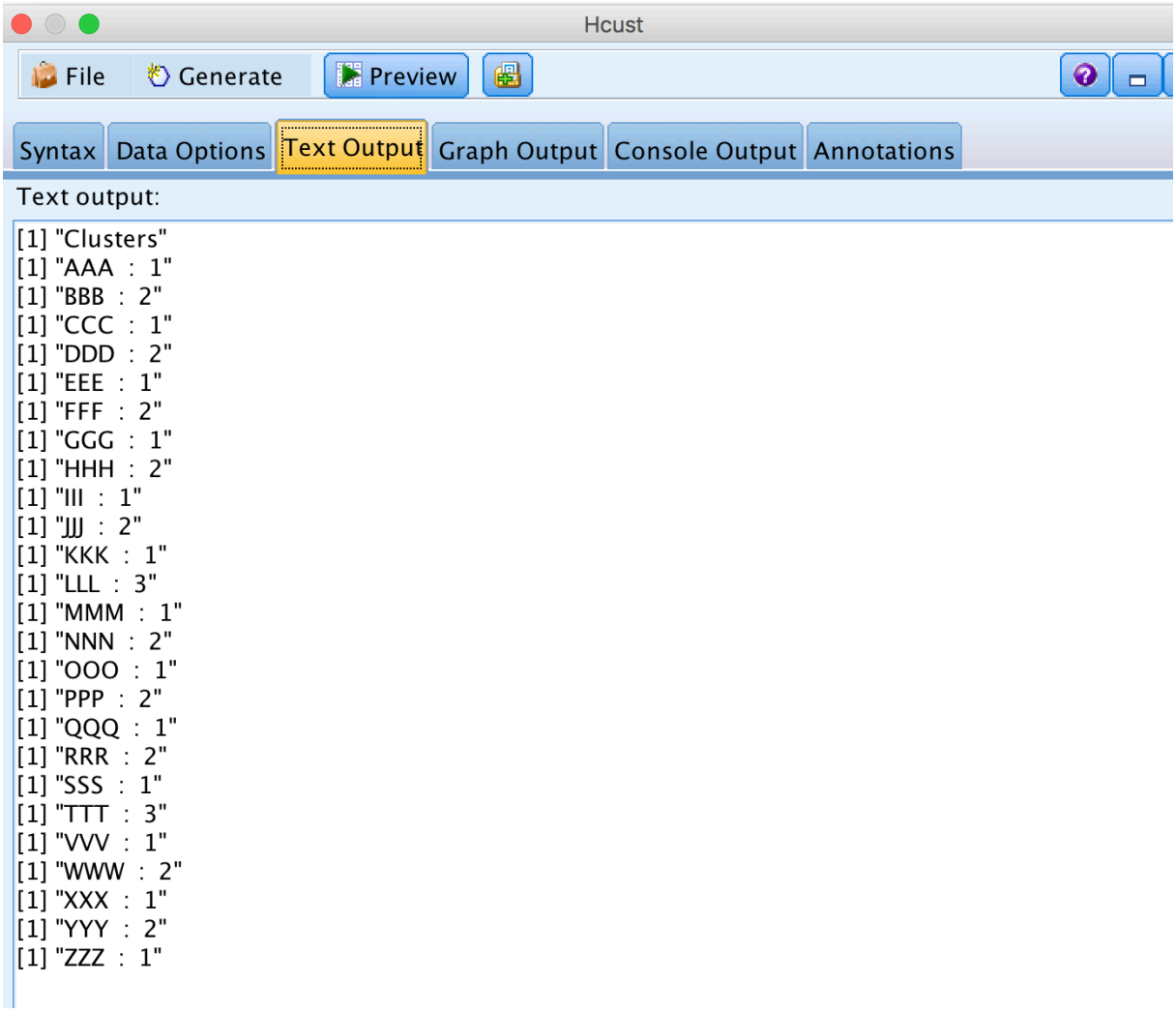


I will click to run the stream.

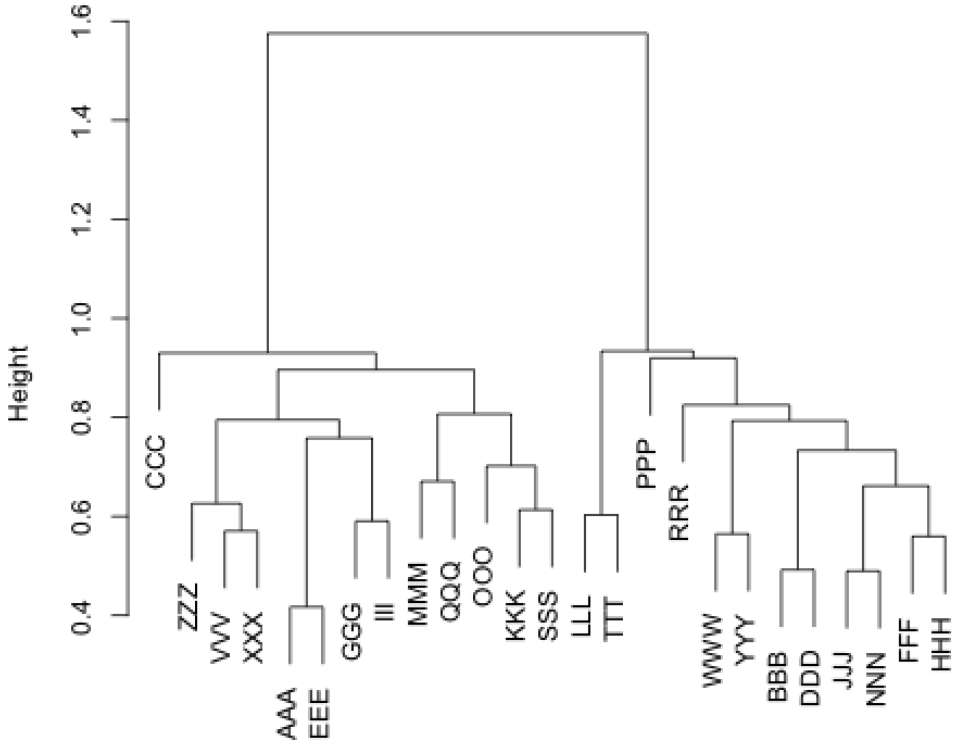


Resulting node is our model nugget where we have outcome. Let's open that and see hierarchical clustering done based on our dissimilarity measure.

In "Text Output" section, you will see cluster memberships for each records



We can also navigate to “Graph Output” tab to see dendogram.



We can visualize clusters by using dendogram.

We can already see logical groupings here starting from 2 big clusters.

Summary

As you can see hierarchical clustering helps us to logically group assets based on their correlations. This could definitely gives us a new and improved perspective in asset selection rather than assuming that grouping based on companies principal business is giving us correct picture.

I hope you enjoyed the course as well as the labs. I suggest you keep experimenting with SPSS Modeler and try and learn new features which will definitely help you in your journey. Thank you very much for your time and hope to see you next time.