

Cluster tool instructions

This tool can be used to:

- Perform Cluster analysis.
- Visualize clusters using a tree plot (dendrogram).
- Highlight important clusters.

Instructions:

1) Prepare dataset in the following format.

Column 1: unique IDs
(no spaces or hyphens)

Columns with one or more
linguistic variables
(relative frequencies)

1	ID	A	B	C	D	E	F	G	H	I	J	K	L
2	S_monologue_priv	PAST	PERF	PRES	PLACE	TIME	1PRON	2PRON	3PRON	IT	DEMPRON	INDPRON	
3	S_dialogue_priv	5.0752	0.3505	8.9916	0.2738	0.4968	5.423	2.7447	4.2475	2.6959	0.888	0.0473	
4	S_monologue_unscripted	2.767143	0.448571	7.509429	0.498	0.762571	2.689	1.793857	2.537714	1.965	0.816143	0.034143	
5	S_monologue_scripted	1.156	0.7286	5.9382	0.4424	0.7078	1.4026	0.4792	2.3212	1.1802	0.4058	0.0256	
6	W_student_writing	3.00	0.2529	5.146471	0.238235	0.375294	0.788824	0.144706	2.692353	1.097647	0.653529	0.030588	
7	W_letters					0.9394	4.375455	1.599697	1.476364	1.048788	0.383636	0.023636	
8	W_academic					0.3365	0.51125	0.06425	0.9335	0.7785	0.3975	0.019	
9	W_popular					0.66875	0.5755	0.11025	1.53675	0.99225	0.3205	0.03	
10	W_reportage					0.731	0.5055	0.089	2.1685	0.997	0.2275	0.0255	
11	W_instructional					0.2655	0.6205	1.6925	1.0135	0.831	0.449	0.027	
12	W_persuasive	1.897	1.152	5.309	0.256	0.376	0.595	0.014	1.861	1.471	0.438	0.075	
13	W_creative	8.416	0.7715	3.9185	0.6725	0.478	2.8785	1.3235	7.009	1.5285	0.5465	0.062	
14													

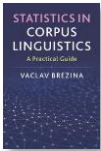
Row 1: header with
variable names (no
spaces or hyphens)

2) Copy and Paste the data into the Cluster tool.

Correlation calculator Clusters MD analysis Videos

1. Paste tab delimited data including header row and id column. For help click [here](#).

ID	PAST	PERF	PRES	PLACE	TIME	1PRON	2PRON	3PRON	IT	DEMPRON	INDPRON	DO	NOMZ	NN
PASS	BYPASS	BE	EXIST	CAUS	CONC	COND	OSUB	PP	JJATR	JJPRED	ADV	TTR	MwL	
CONJ	DwNT	HdG	AMP	EMPH	DEMO	POSS	NECESS	PRED	PUBV	PRIV	SUAV	SMAP	CONT	
SYNEG	ANNEG													
S_dialogue_priv	5.0752	0.3505	8.9916	0.2738	0.4968	5.423	2.7447	4.2475	2.6959	0.888	0.0473	0.4089		
0.3955	14.9268	0.56	0.0122	3.2432	0.4298	0.2339	0.1089	0.3518	0.1356	6.0483	5.0543	0.8469	5.4594	
164.21	3.7473	0.041	0.1693	0.2813	0.1667	1.815	1.351	0.6582	0.1716	0.9636	0.7357	3.0324	0.1711	
0.0408	5.0874	0.1822	1.8282											
S_dialogue_pub	3.406625		0.511	8.287375		0.283125		0.53875	4.061	2.366875			2.54125	
2.07675	1.0015	0.045875		0.320875		1.707375		16.34525	0.84025	0.0495	2.76075	0.480125		
0.267625		0.065125		0.458625		0.137375		8.426	5.1215	0.7165	4.817125	175.3375		
4.111125		0.087625		0.128	0.169875		0.33075	1.064875		1.8195	0.723625		0.174375	
0.83525	0.813625		2.234375		0.333625		0.085125		3.7435	0.165625		1.48525		
S_monologue_unscripted	2.767142857		0.448571429		7.509428571		0.498		0.762571429		2.689		1.793857143	



3) Select the appropriate options or leave the defaults.

2. Select parameters.

Transform data to z-scores

Useful if linguistic variables measured on different scales (default)

Distance measure: Manhattan distance ▾

Clustering method: Ward's method

Different distance measures

3. Select highlight.

Different amalgamation options

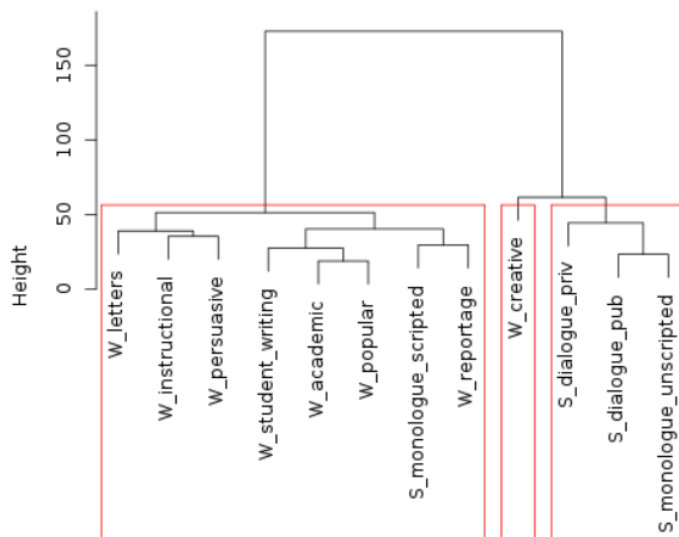
How many cluster groups do you want to highlight? 5

Option to highlight clusters in the output


4) Click 'Analyze'.

5) The output

Tree plot (dendrogram)
3 cluster groups were highlighted.



Read from the bottom. Important clusters highlighted

 R code that performs the analysis can be viewed and copied when going with the mouse pointer to [R code](#)