

RDoc_pca_report

Ilaria Bonavita

25 May 2018

PCA on MARS variables assigned to RDoc domains

In this analysis we aim at assessing the validity of assignment of MARS variable to RDoC domains (constructs and subconstructs) through PCA. The original datasets contains 1418 subjects and 149 features (phenotypes). However, due to the high number of missing phenotypes, in order not to reduce the sample size dramatically, we included only variables belonging to the SCL-90-R, tpq and life event (el) scales. Moreover, we summed all the el variables in a single more informative feature and we removed the SCL-90-R and tpq variables that represent subscale, retaining only the single item variables. The final datatest has 292 samples and 65 variables:

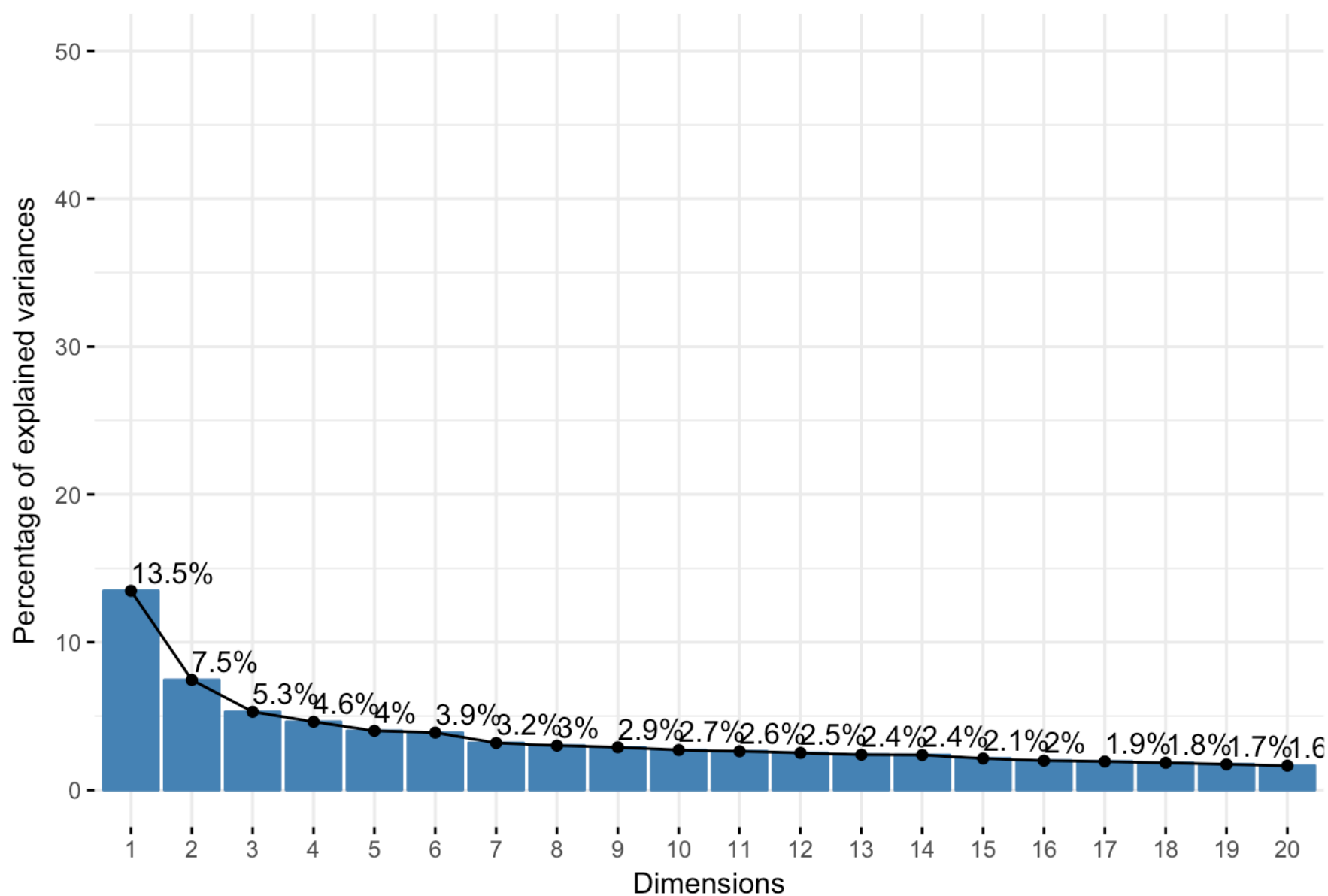
The number of variables assigned to negative valence, positive valence, system for social processes, cognitive system and arousal is the following:

##					
##	neg	cogn	arou	pos	social
##	20	7	2	32	4

Variance explained by the PCs

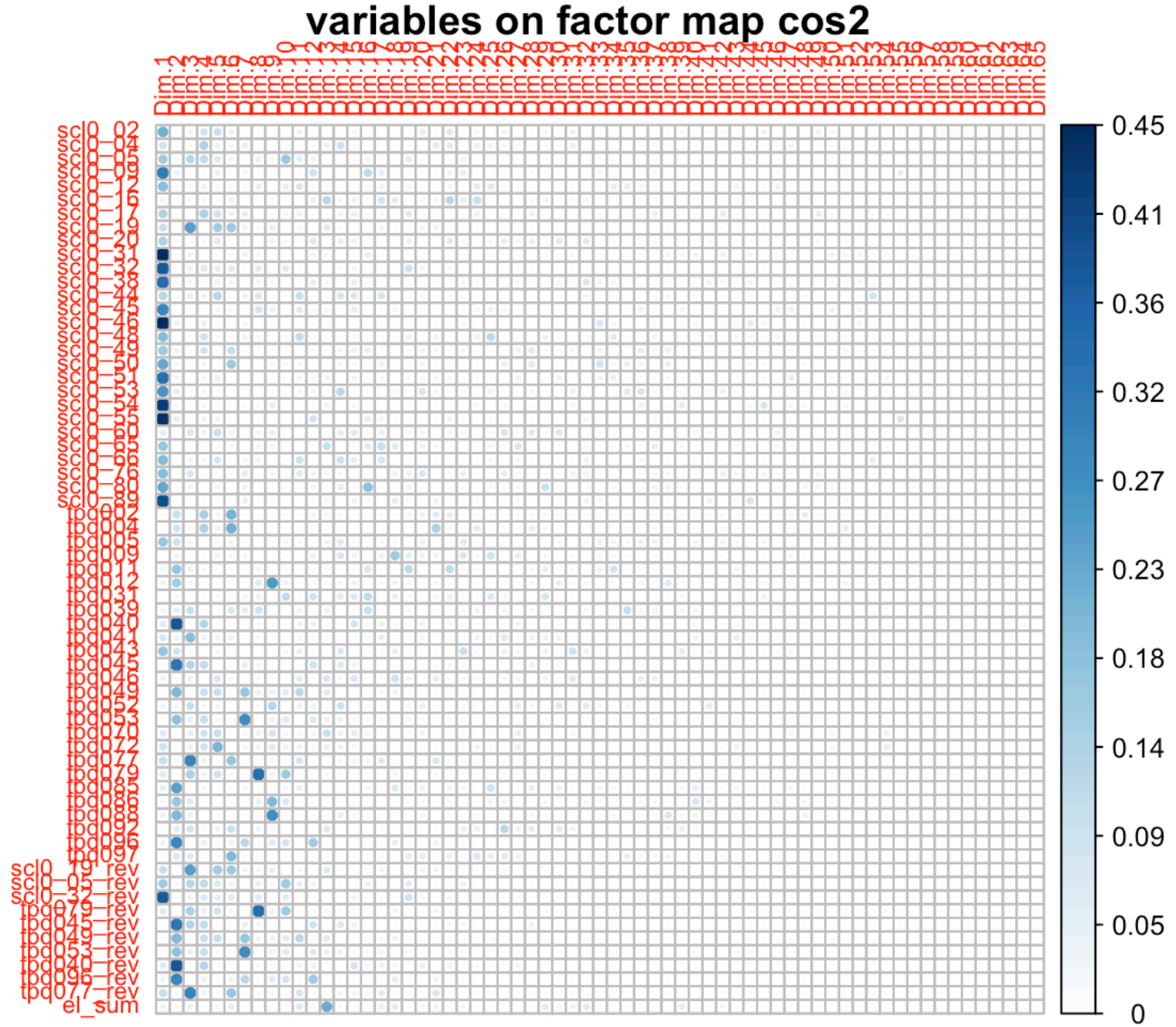
At first, we can do a PCA on all the variables:

Scree plot



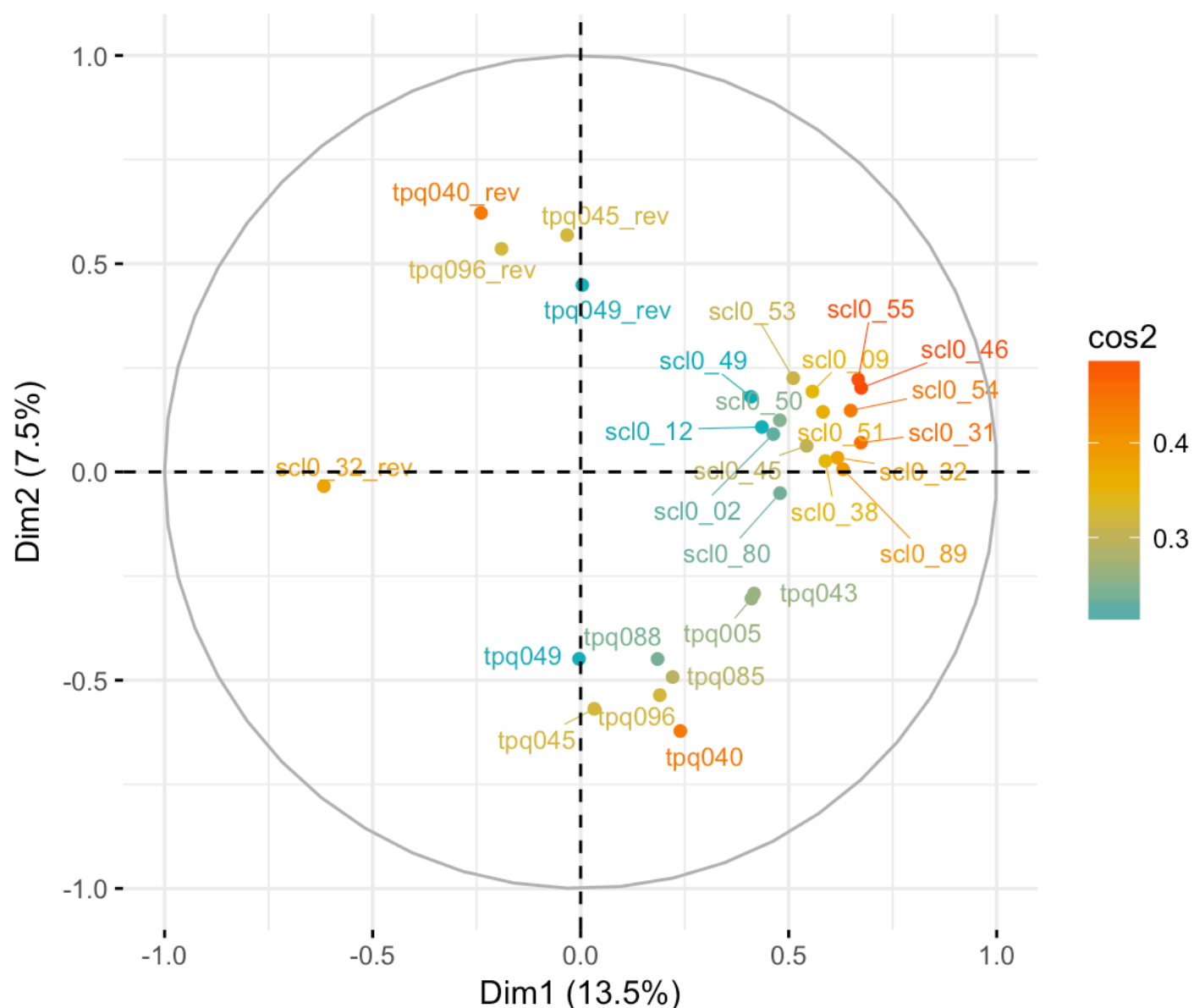
Quality of the representation

The quality of representation of the variables on factor map is indicated by \cos^2 (square cosine, squared coordinates) We can visualize the \cos^2 of variables on all the dimensions:



The correlation between a variable and a principal component (PC) is used as the coordinates of the variable on the PC. Here we can see the quality of the factor map by coloring the degree

Variables - PCA

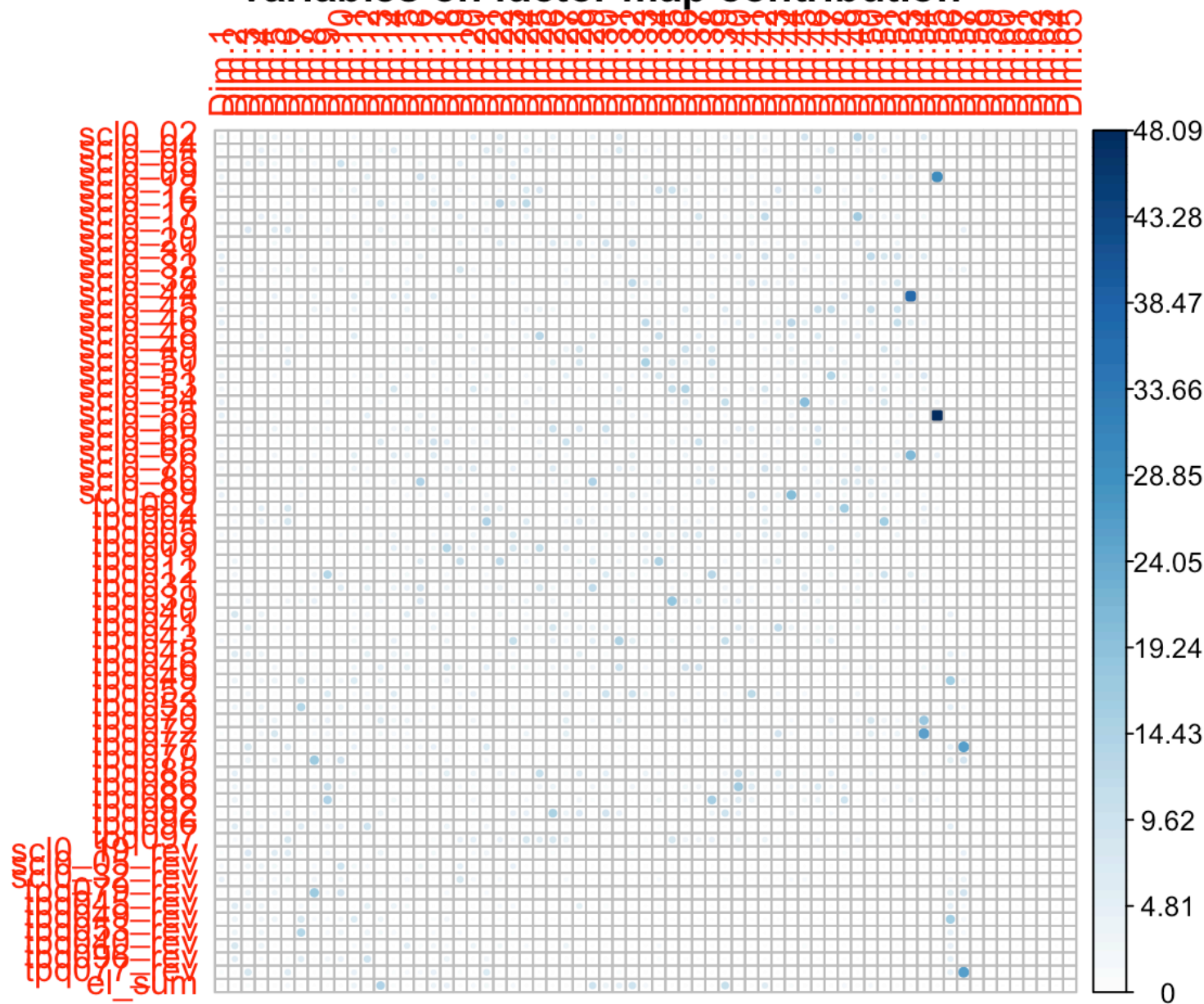


Note that: - Positively correlated variables are grouped together. - Negatively correlated variables are positioned on opposite sides of the plot origin (opposed quadrants). - The distance between variables and the origin measures the quality of the variables on the factor map. Variables that are away from the origin are well represented on the factor map. - A high cos2 indicates a good representation of the variable on the principal component. - A low cos2 indicates that the variable is not perfectly represented by the PCs.

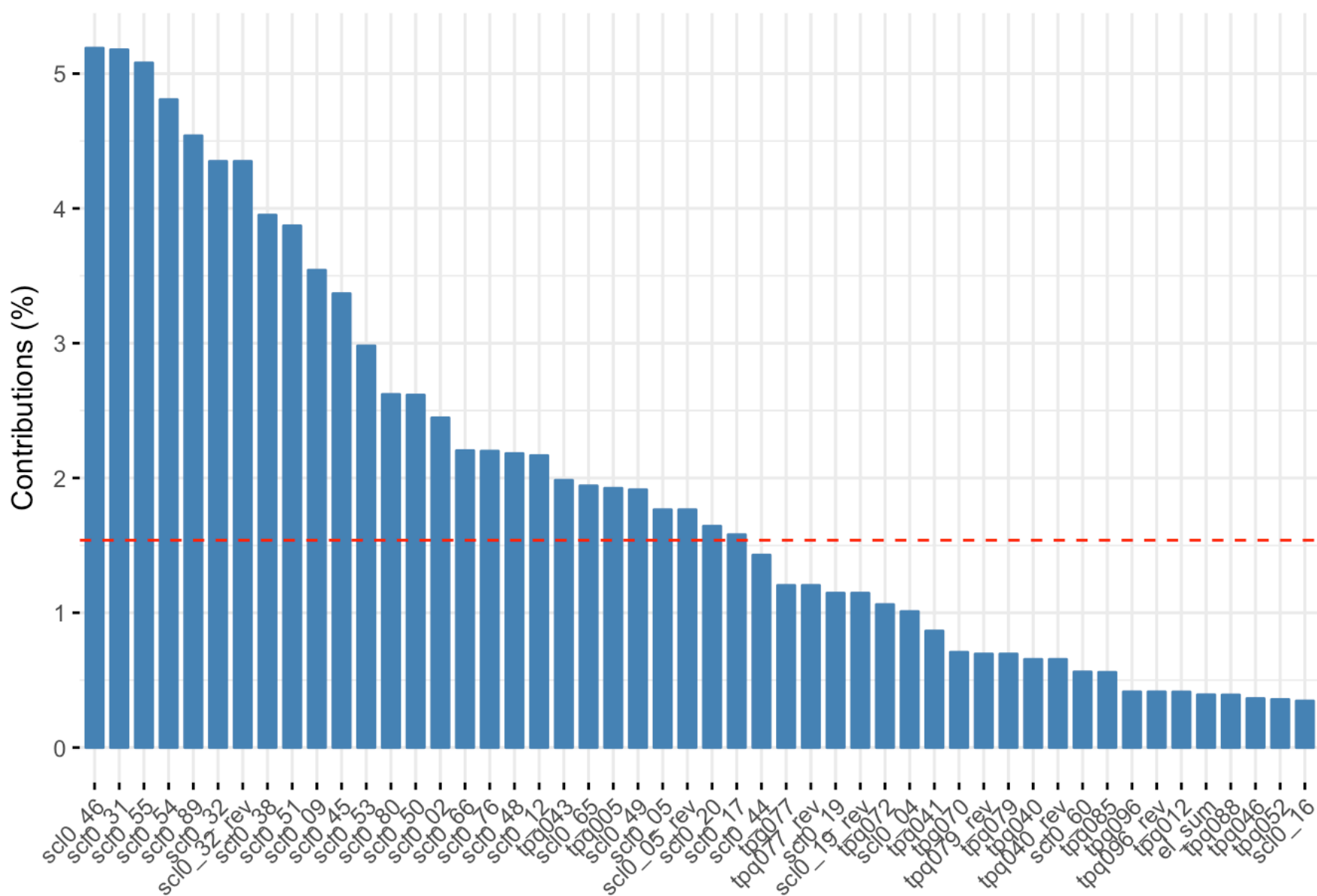
Contributions of variables to PC1 and PC2

The contributions of variables in accounting for the variability in a given principal component are expressed in percentage. - Variables that are correlated with PC1 (i.e., Dim.1) and PC2 (i.e., Dim.2) are the most important in explaining the variability in the data set. - Variables that do not correlated with any PC or correlated with the last dimensions are variables with low contribution and might be removed to simplify the overall analysis.

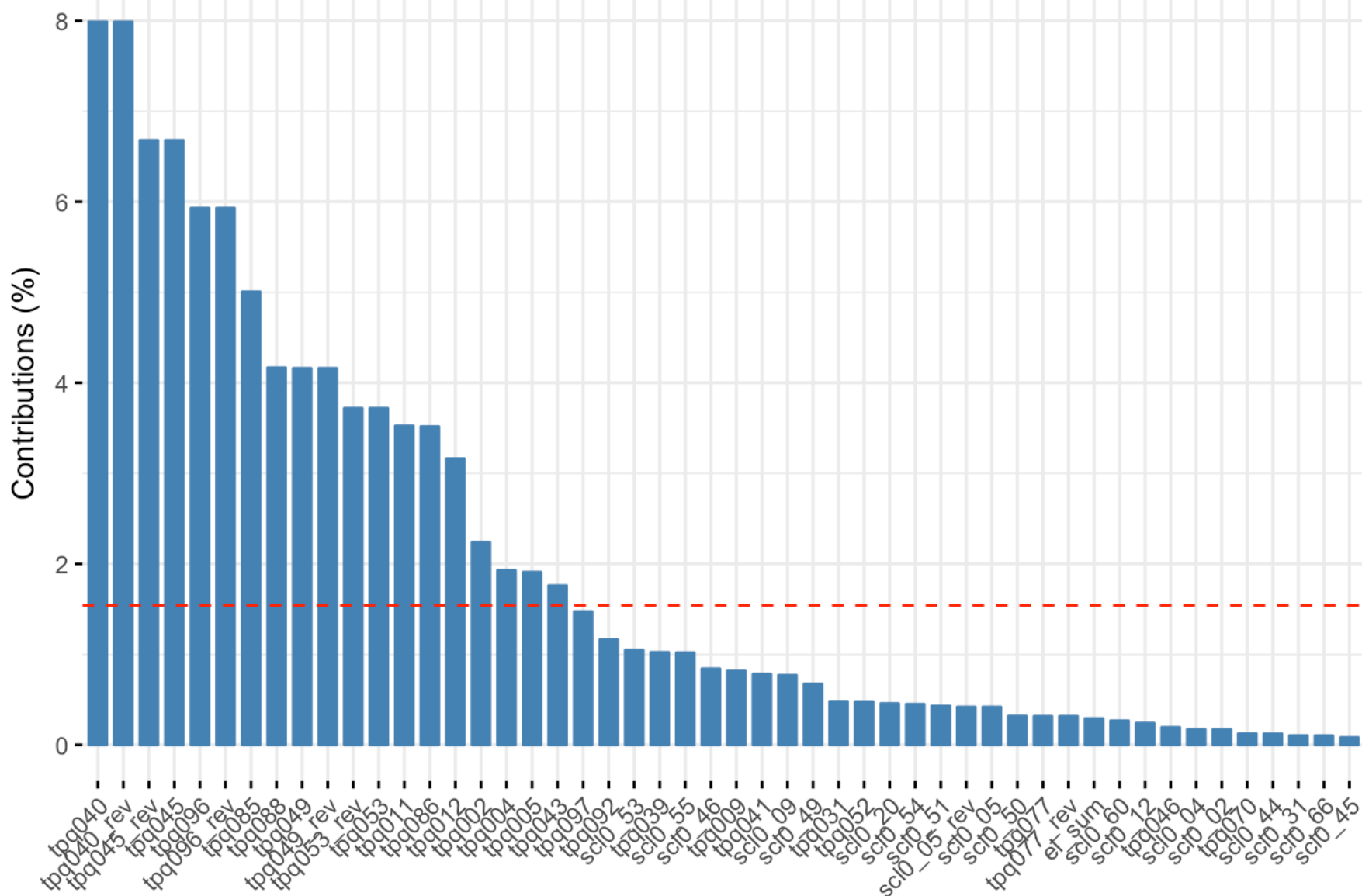
variables on factor map contribution



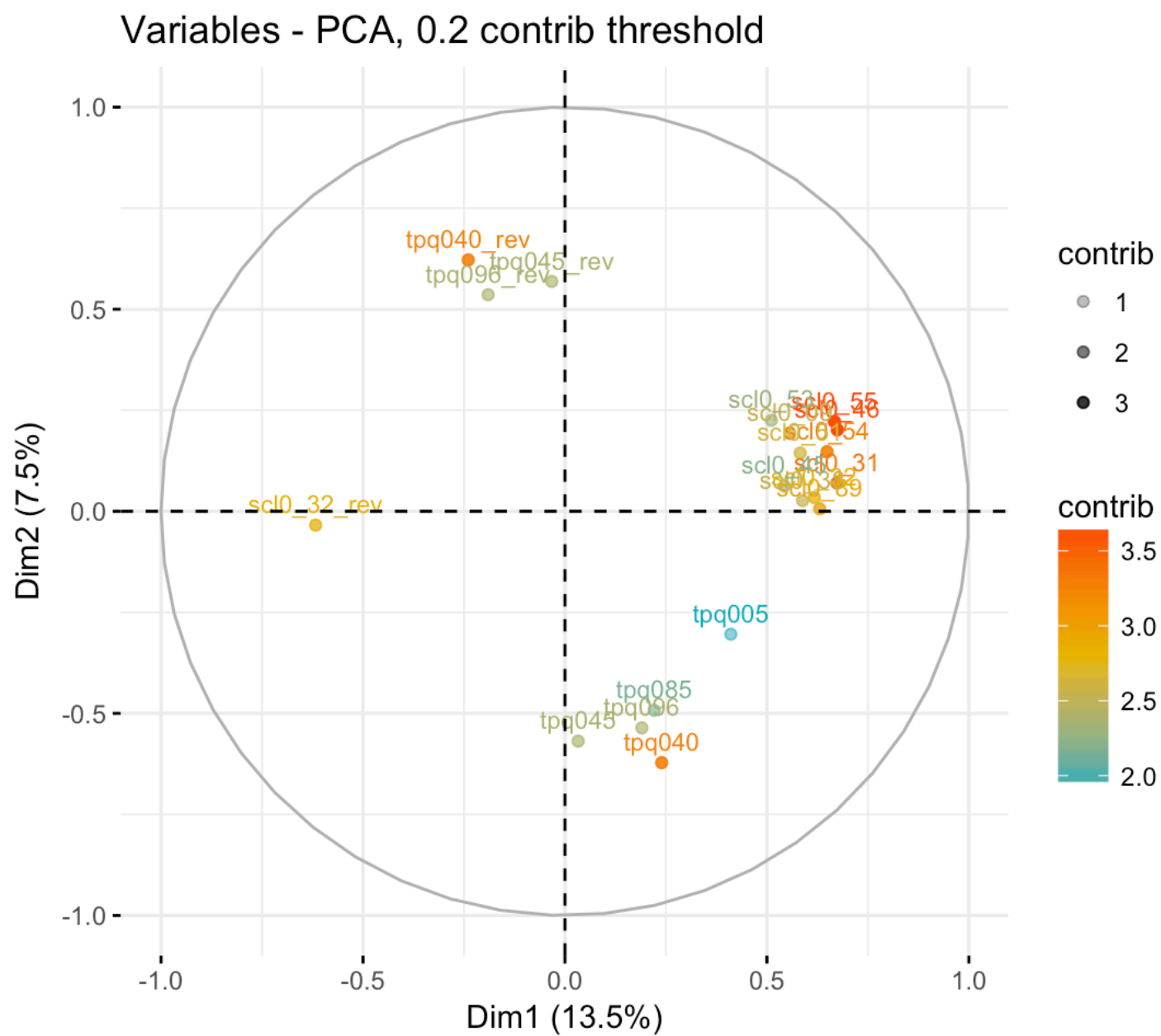
Contribution of variables to Dim-1



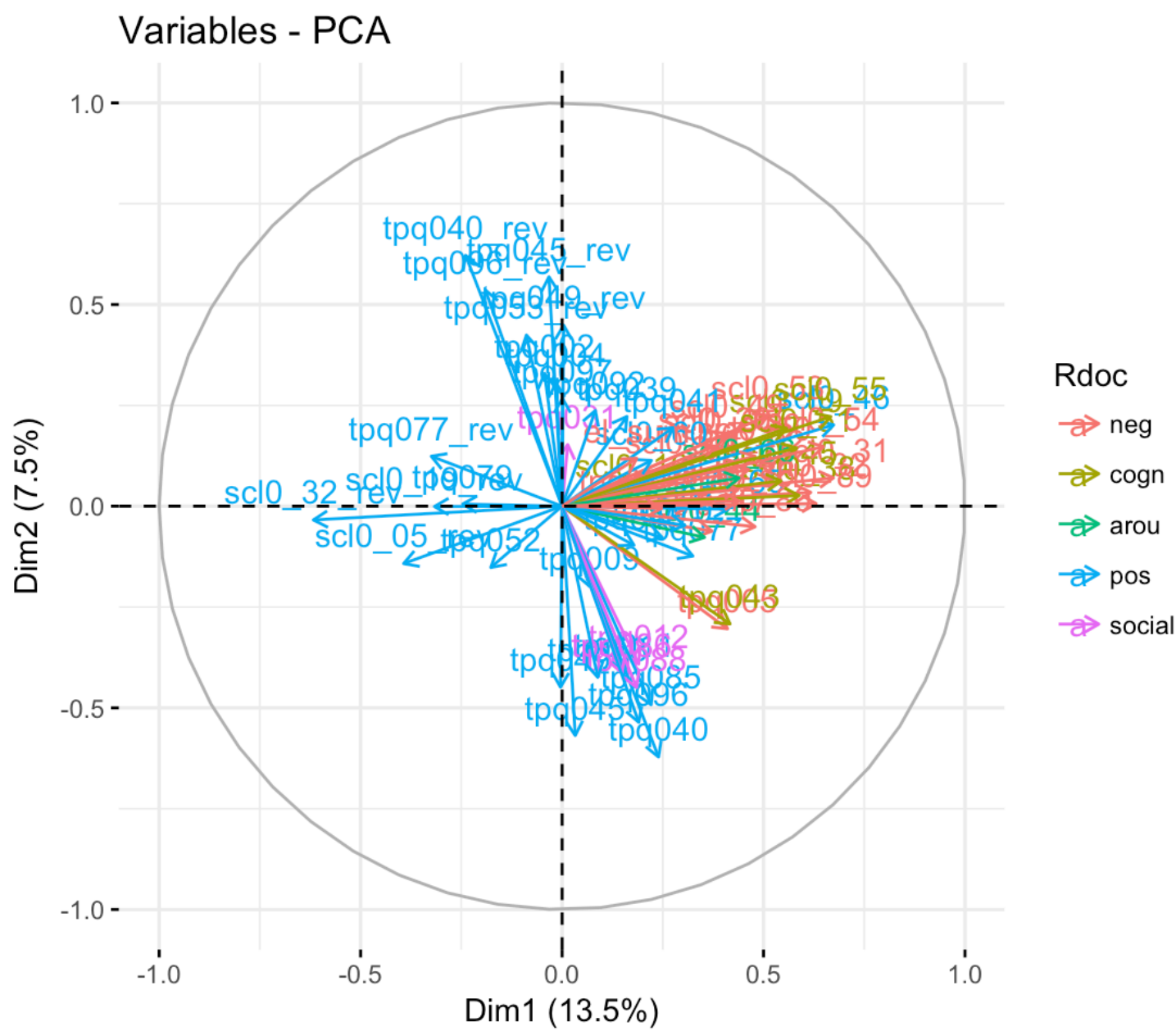
Contribution of variables to Dim-2



The 20 most important contributing variables are highlighten in the plot:



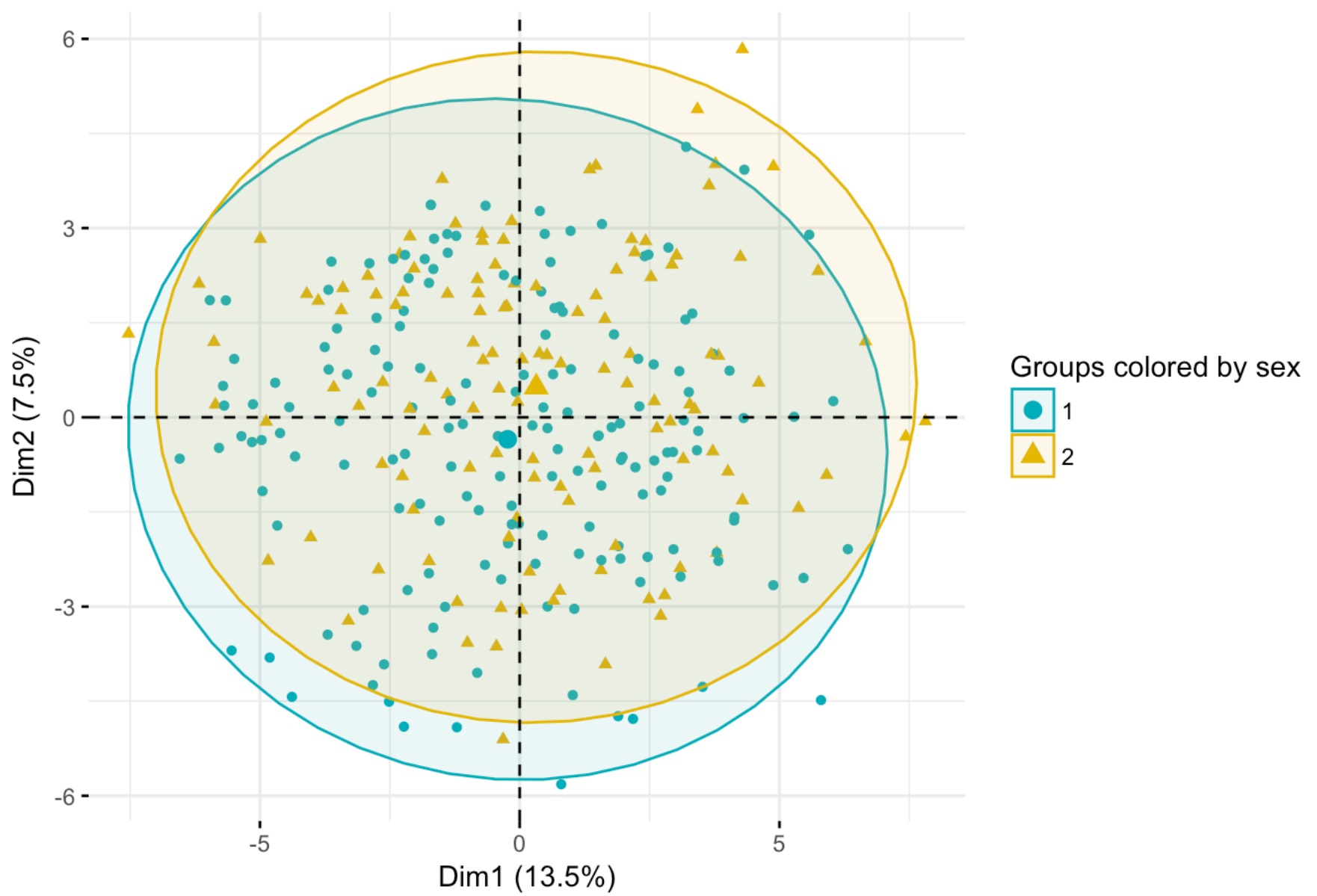
Grouping variables according to RDoC assignment



Graph of individuals

We can check the effect of the sex in the PCA results:

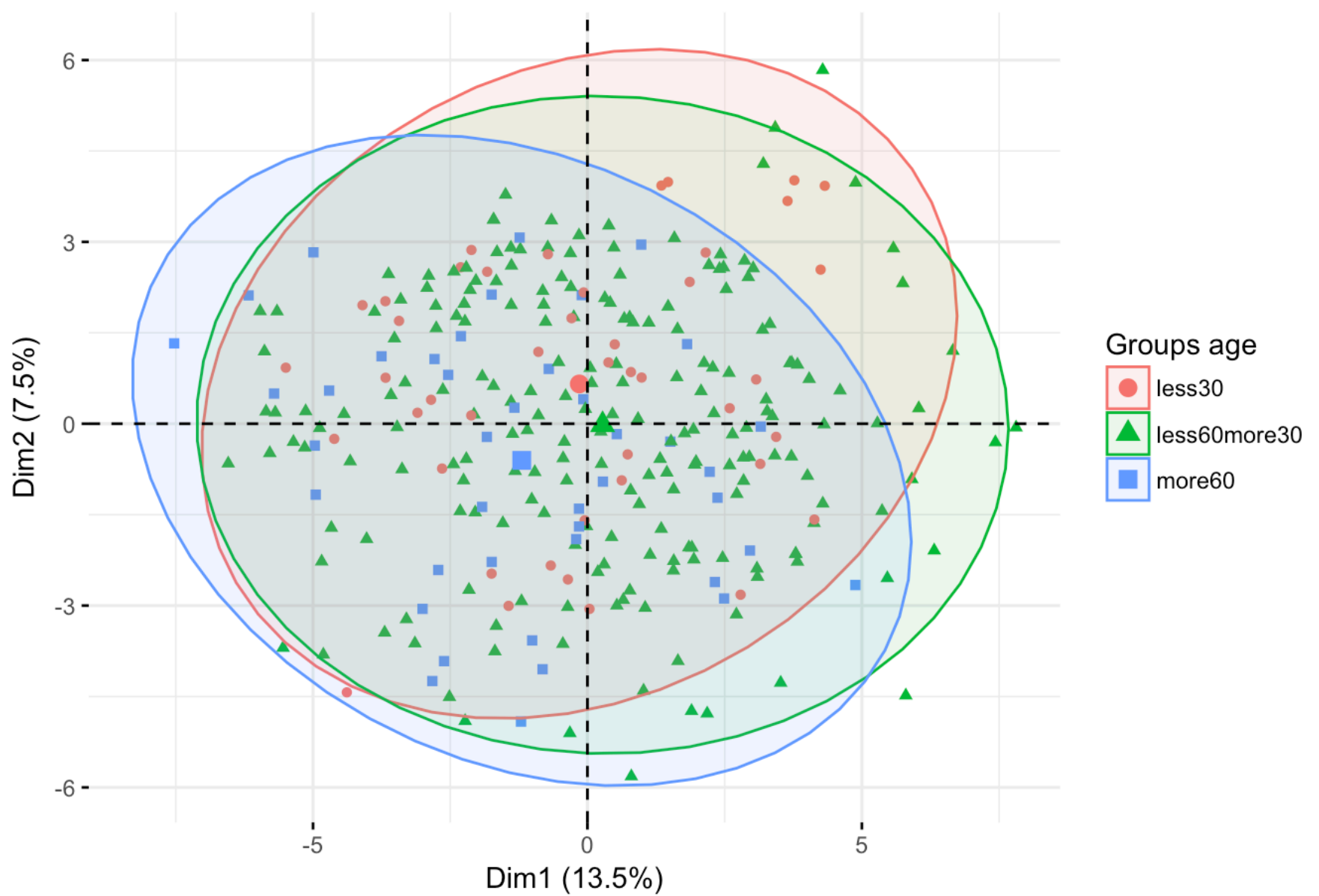
Individuals - PCA



And the effect of age:

##	age
##	less30 less60more30 more60
##	45 206 42

Individuals - PCA

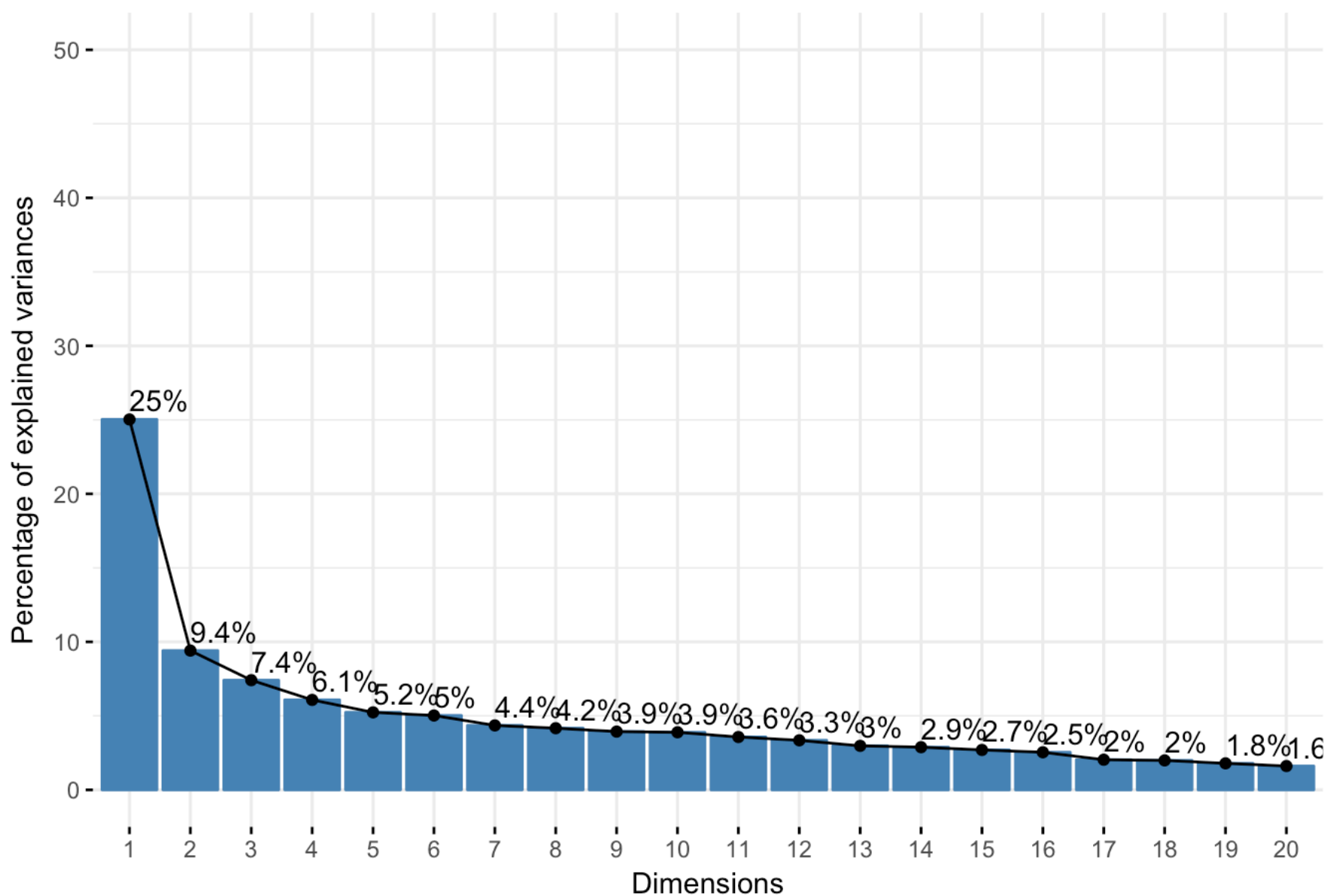


PCA on each RDoC-assigned variables

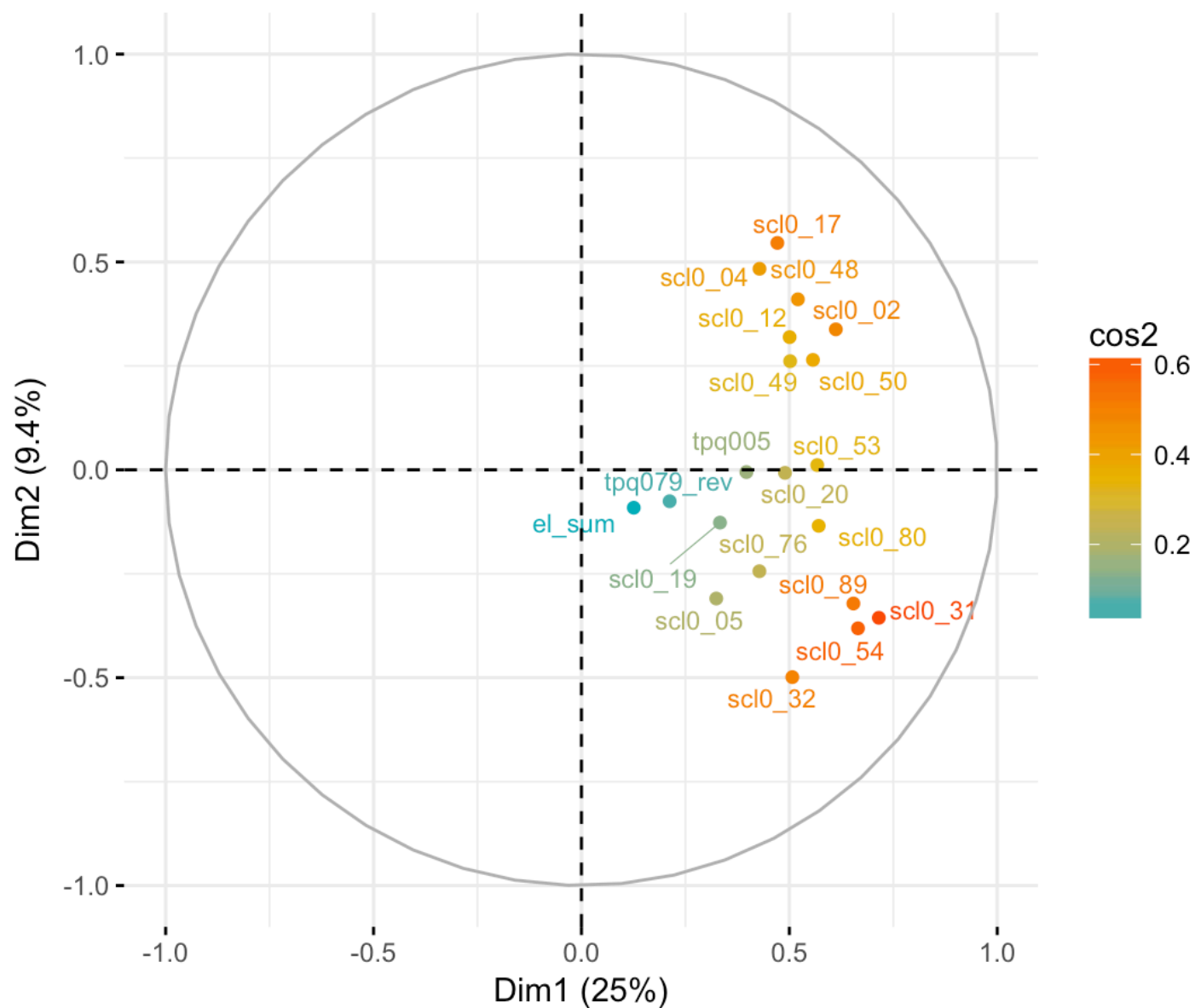
Negative valence

```
## [1] 292 20
```

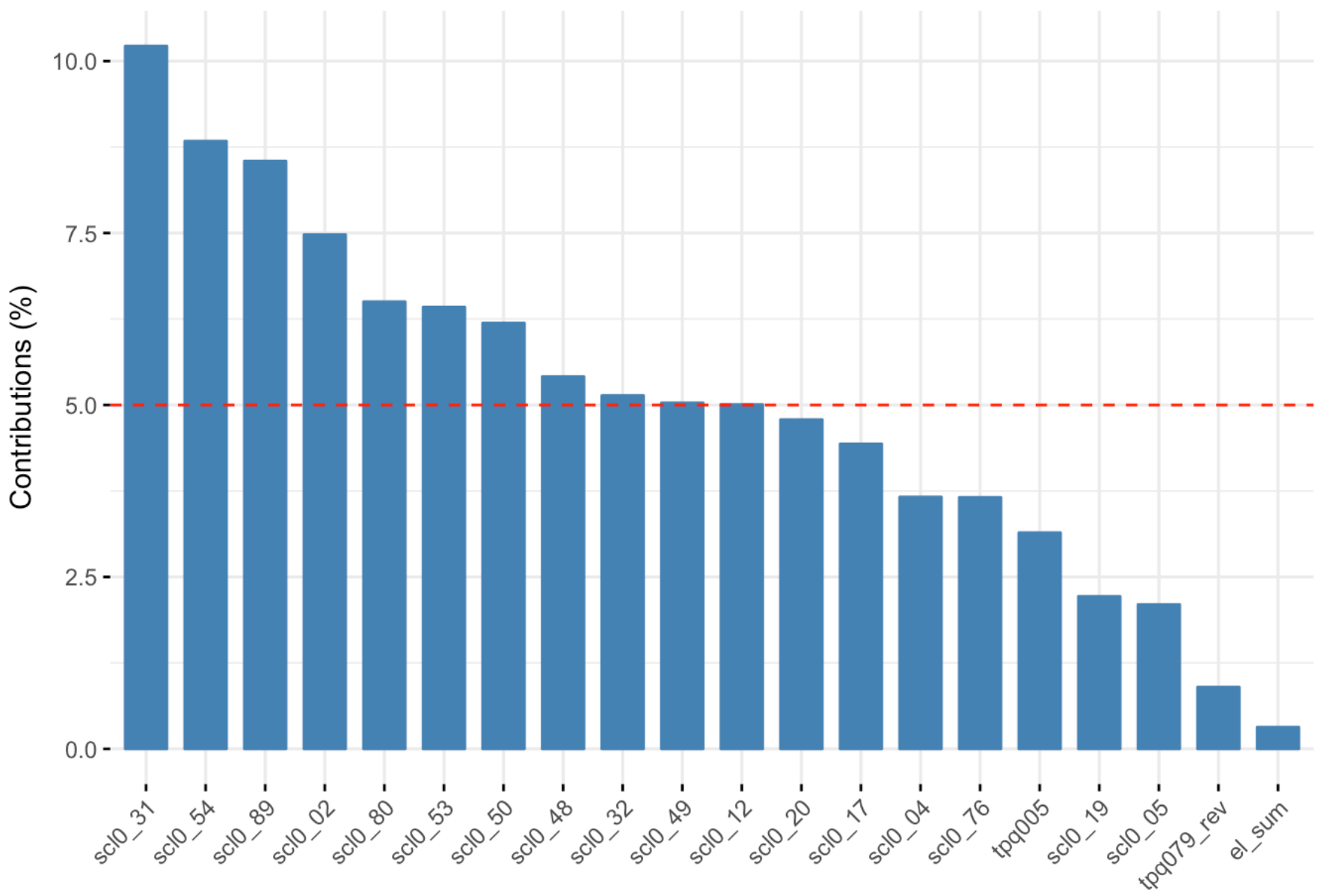

Variance explained - Negative valence



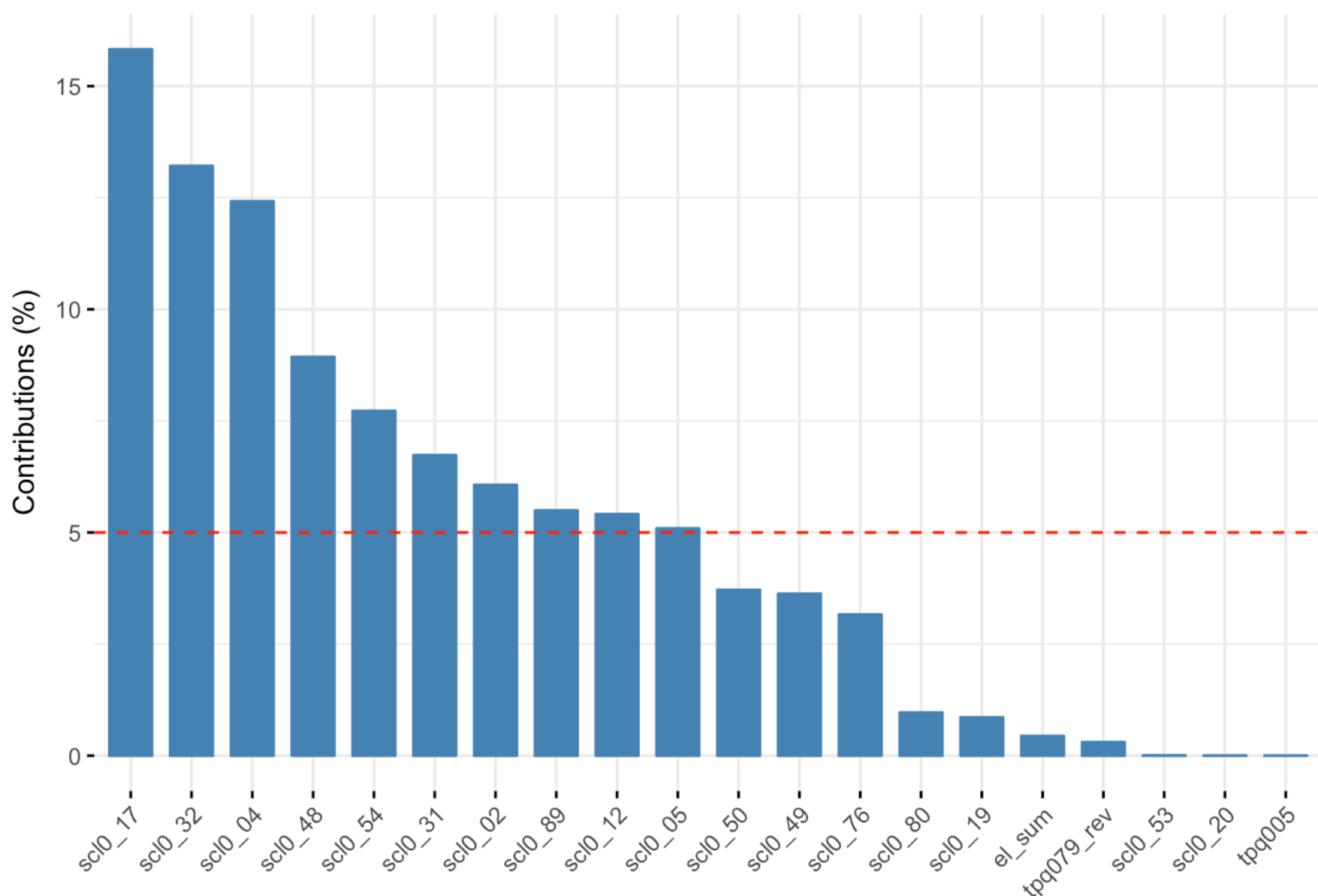
Variables - PCA, Negative Valence



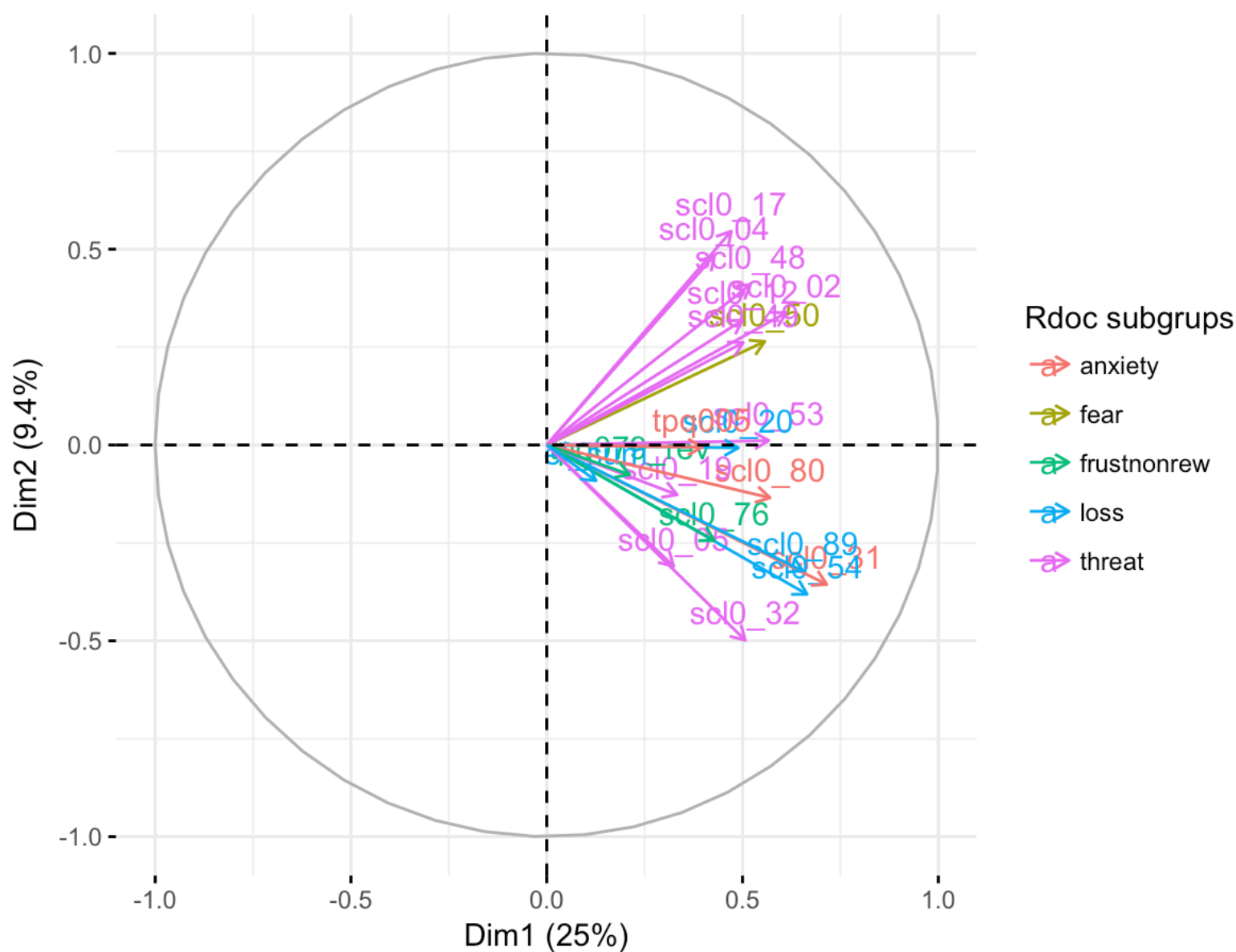
Contribution of variables to Dim-1



Contribution of variables to Dim-2



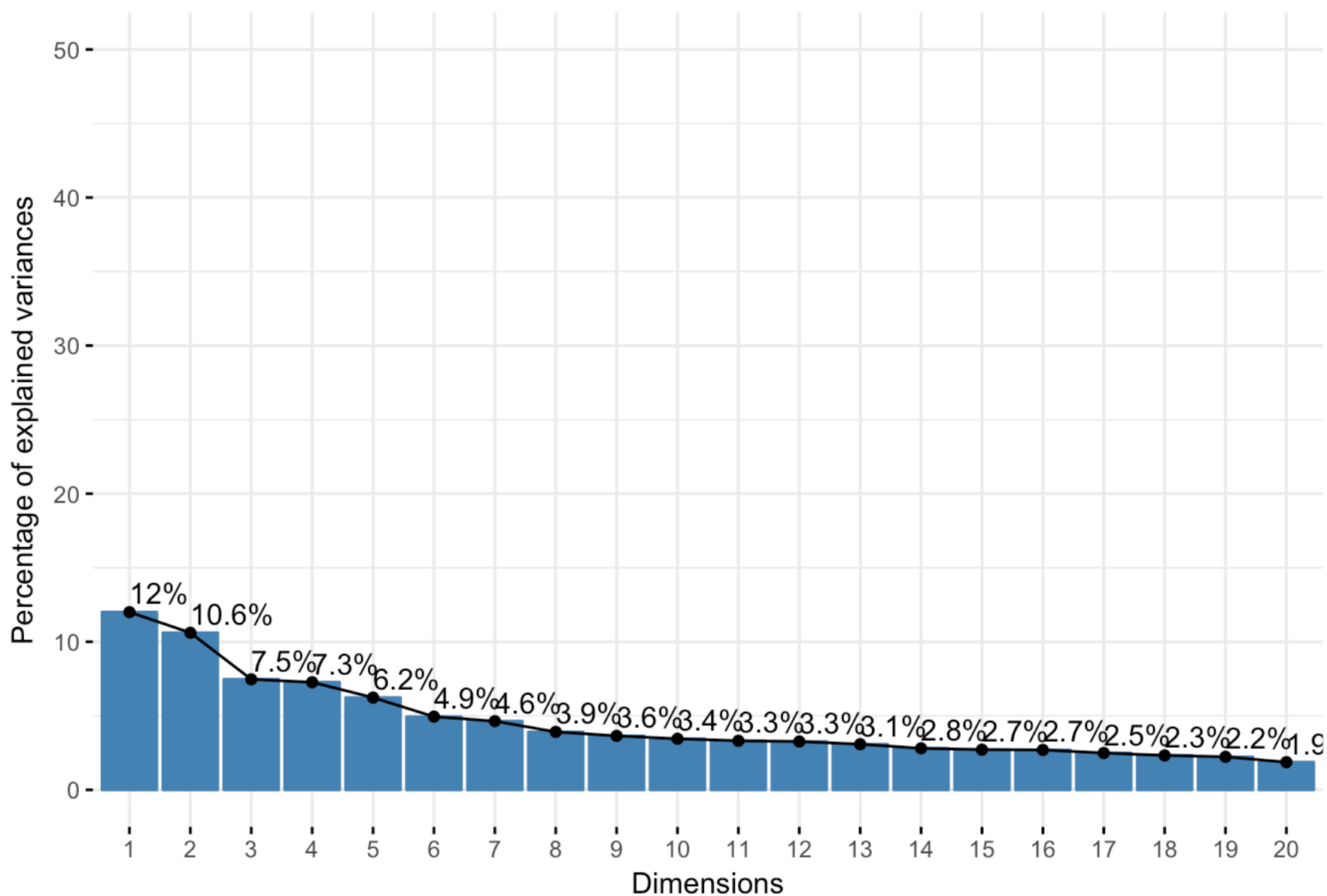
Variables - PCA - Negative valence subgroups



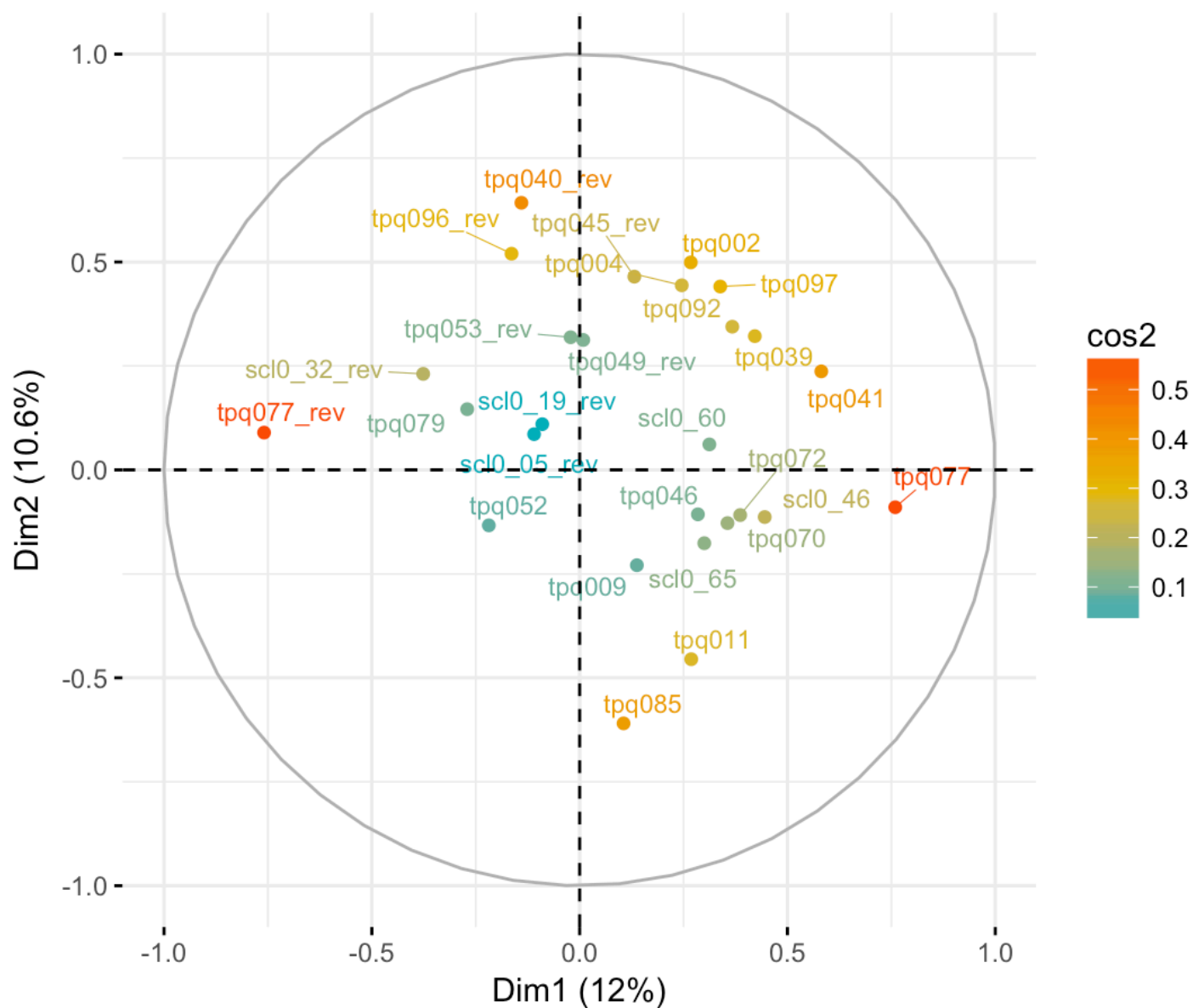
Positive valence

##	[1]	292	32
----	-------	-----	----

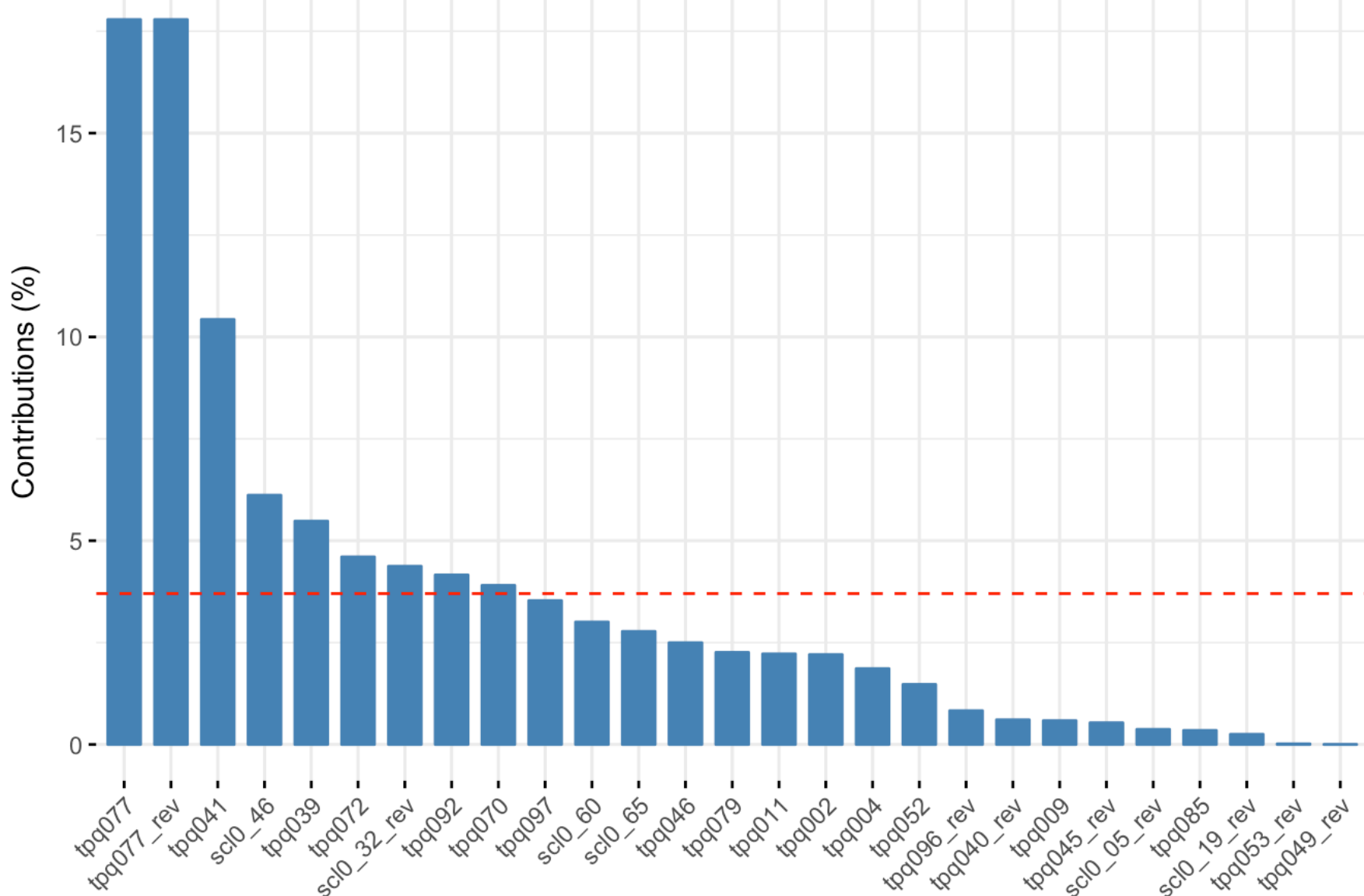
Variance explained - Positive valence



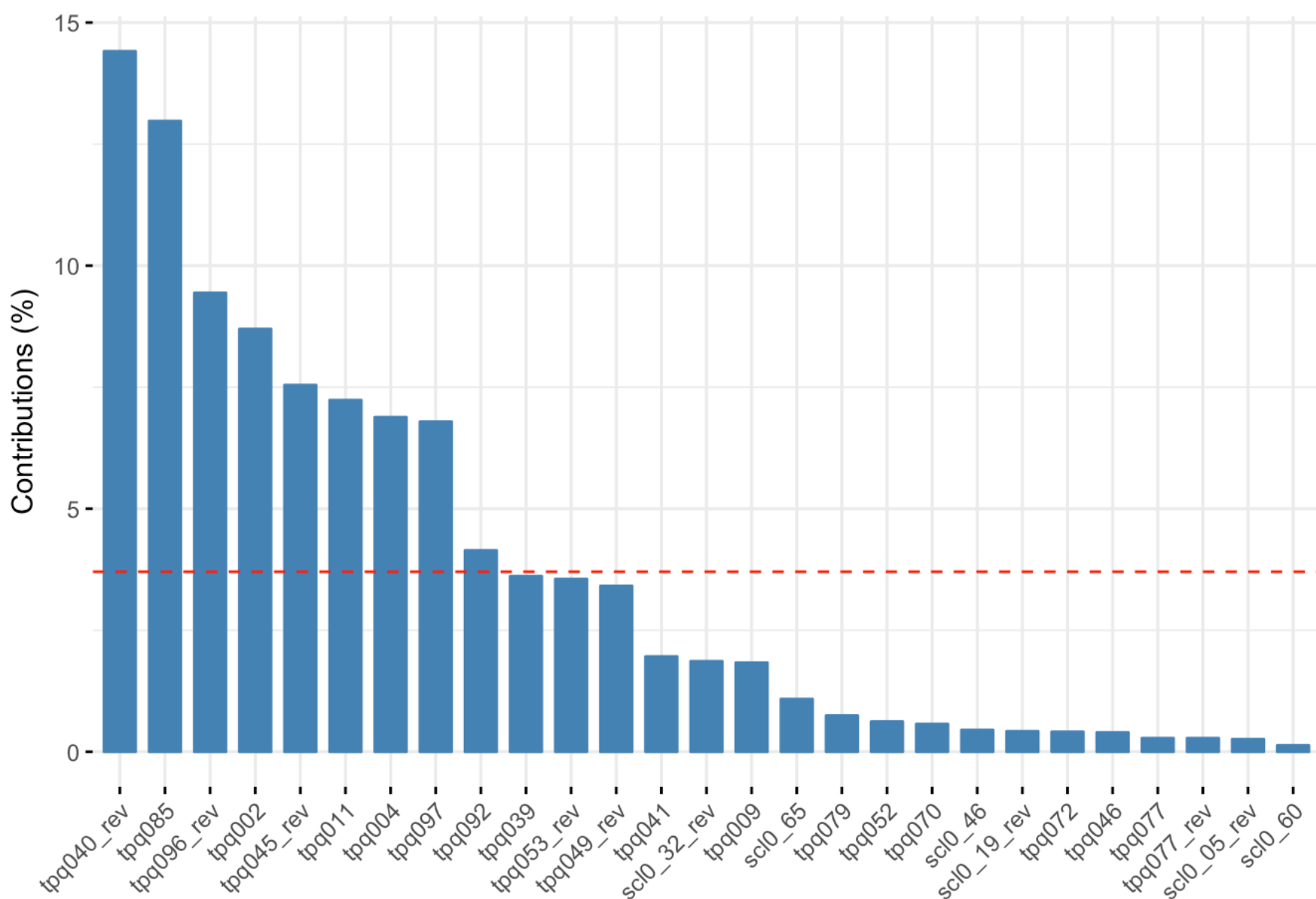
Variables - PCA, Positive Valence



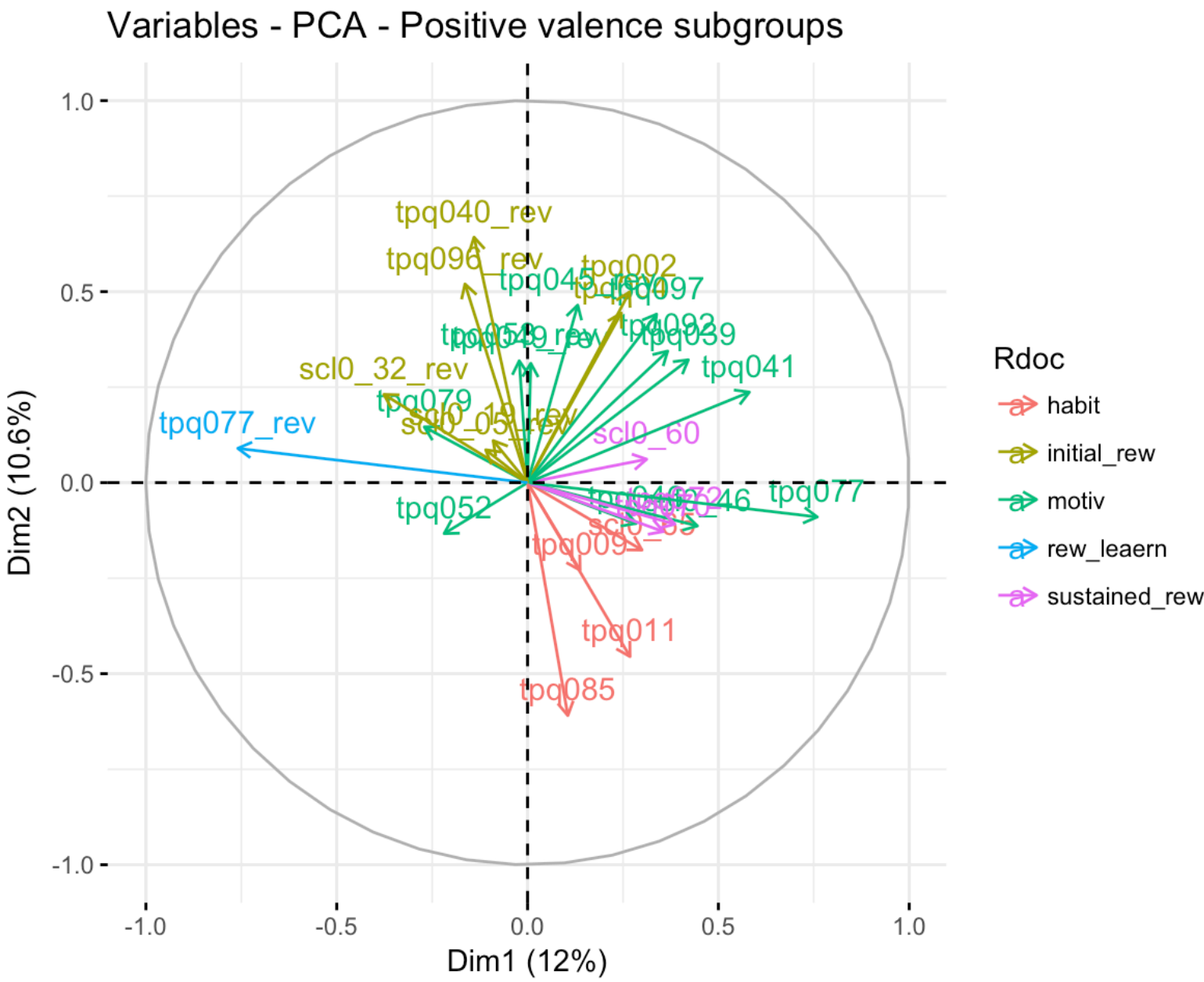
Contribution of variables to Dim-1



Contribution of variables to Dim-2

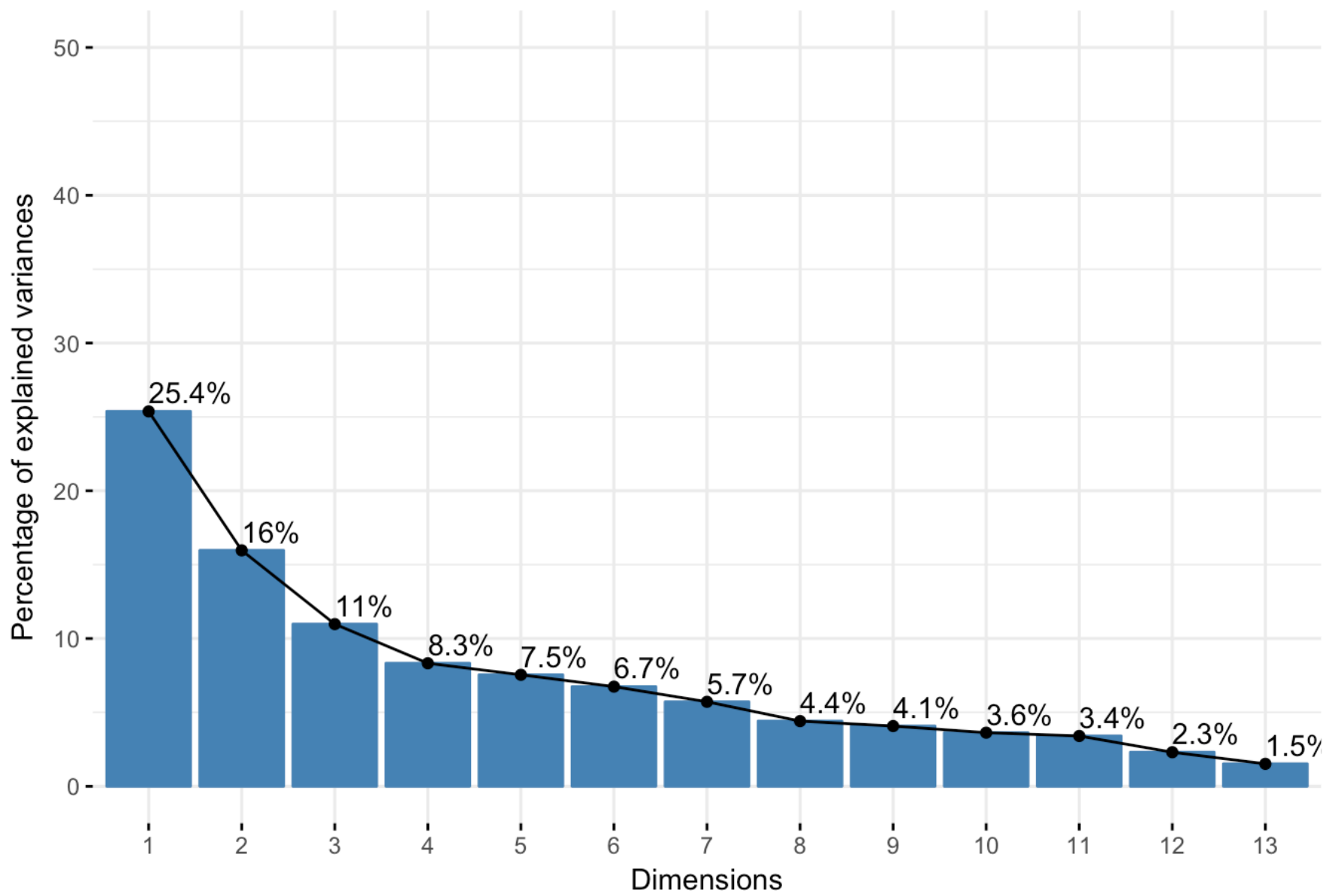


##	grp					
##		habit	initial_rew	motiv	rew_leaern	sustained_rew
##		4	7	12	1	3

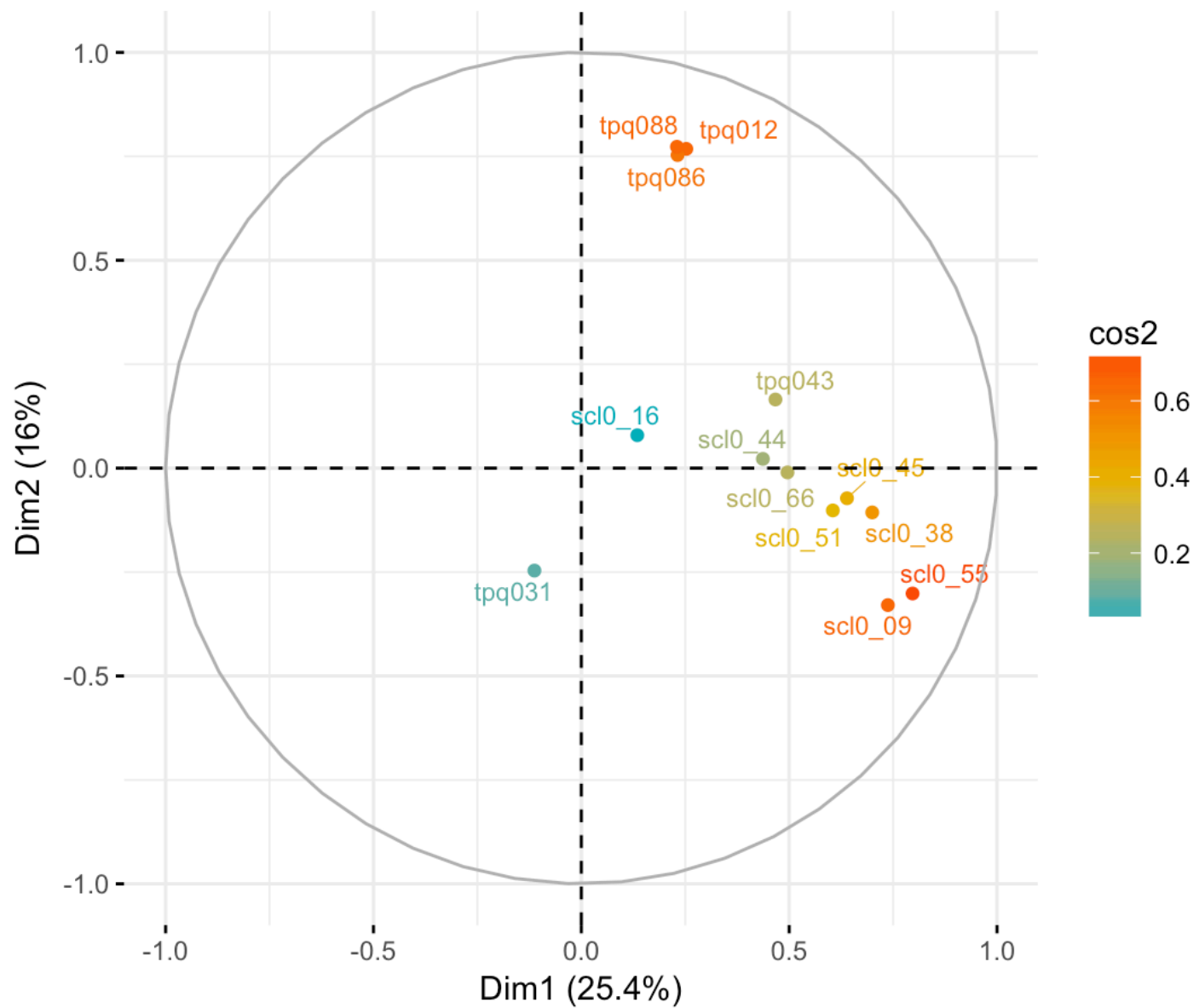


Cognitive System, Systems for Social Process and Arousal

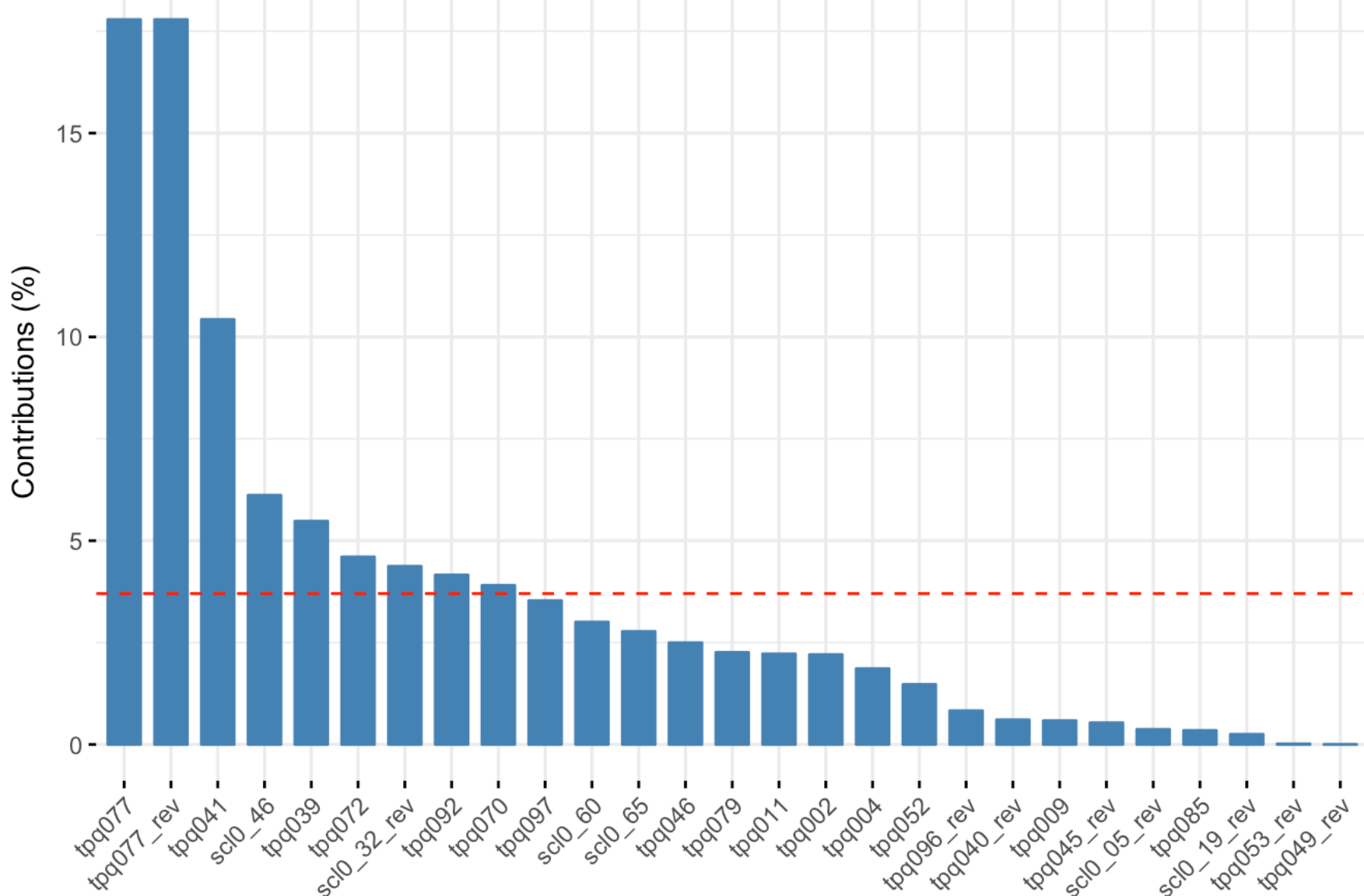
Variance explained - Cognitive, Social, Arousal



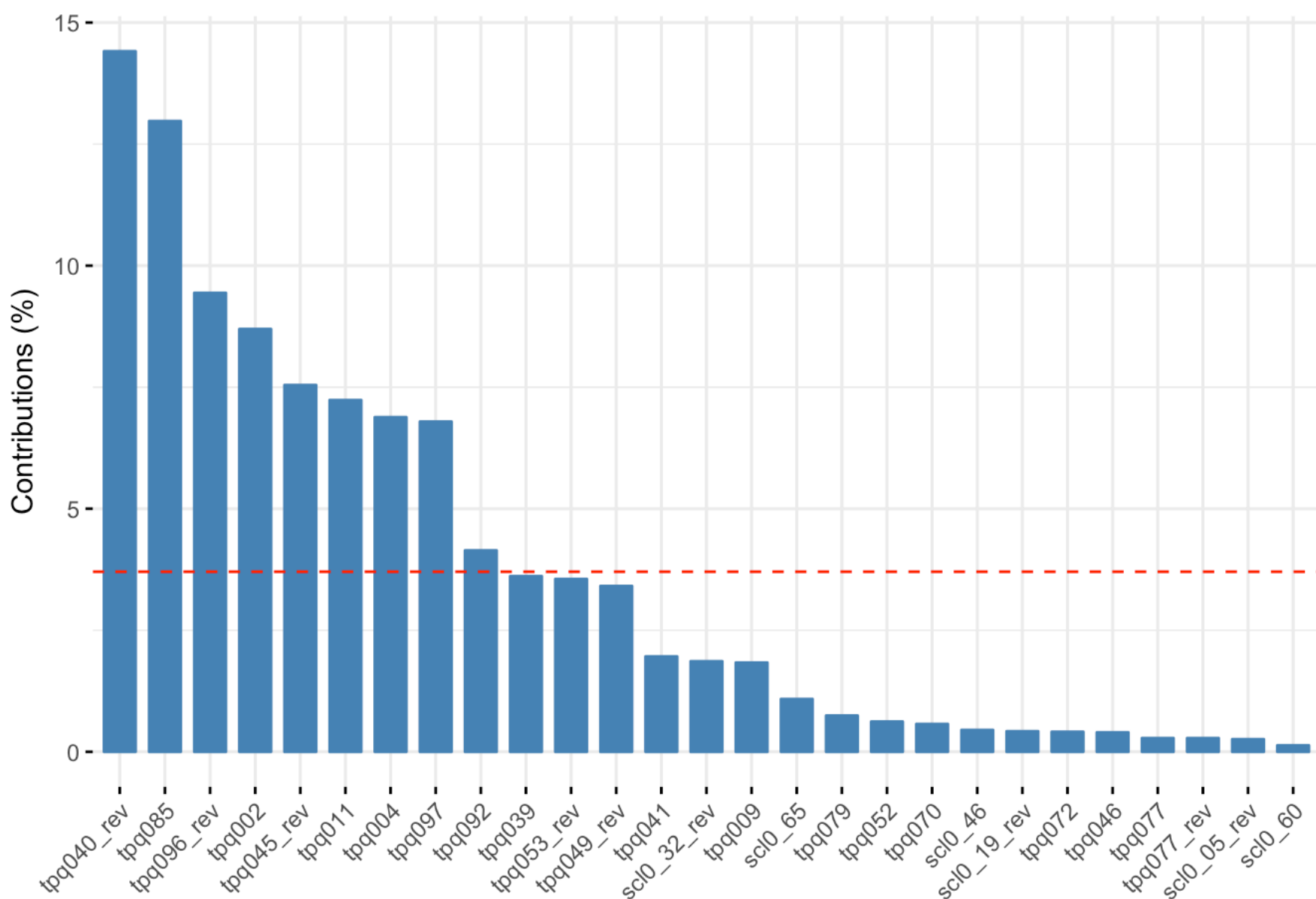
Variables - PCA, Cognitive, Social, Arousal



Contribution of variables to Dim-1



Contribution of variables to Dim-2



##	grp					
##	general	lang	mem	percep	sleep	social
##	3	2	1	1	2	4

