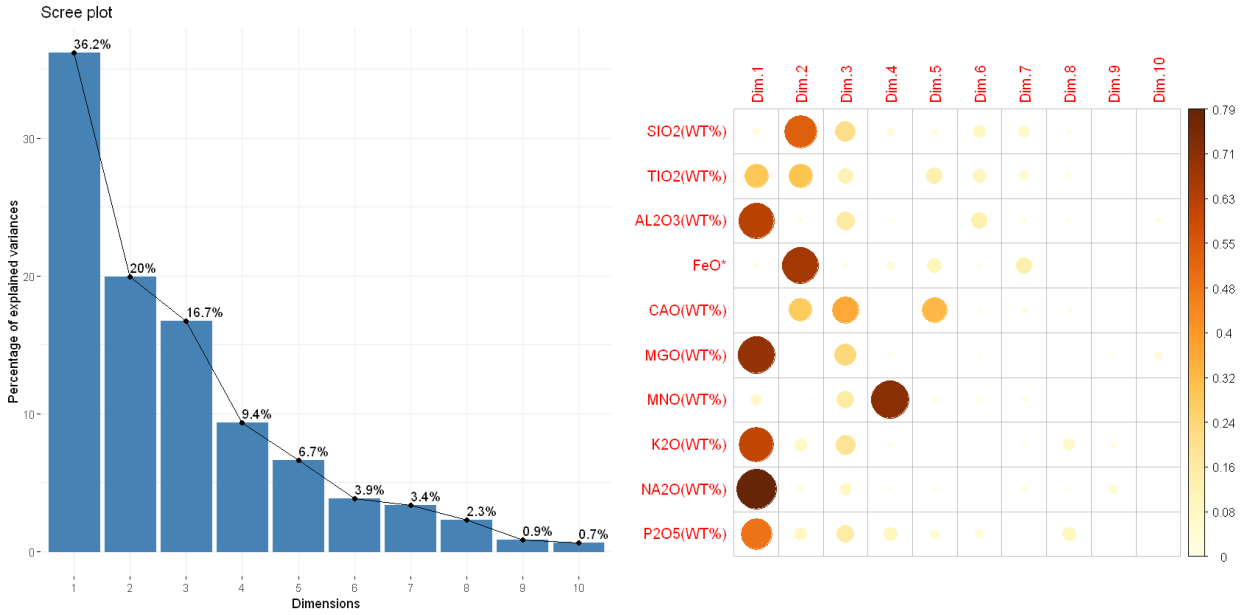
PCA Analysis Report

Analysis was done using the primary elements from the Hawaii dataset. Samples with missing data points for any of these 10 variables was excluded. This reduced the total number of usable samples from 12,995 to 2,866 for the purposes of this analysis.

The used elements were as follows:

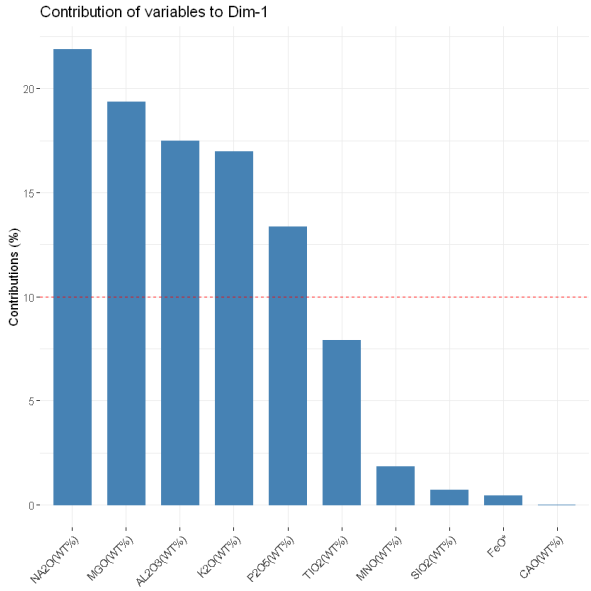
SiO_2	Na_2O
AlO_3	K_2O
FeO*	TiO_2
MgO	MnO
CaO	P_2O_5

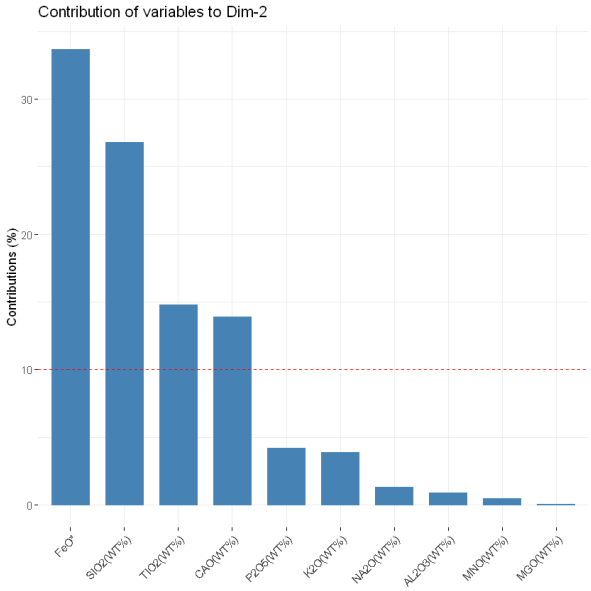
As an experiment, data imputation was first attempted to try and fill in the gaps within missing data. The resulting PCA analysis gave results showing low contribution scores for all the variables in terms of variation. Imputation resulted in most samples having similar features.

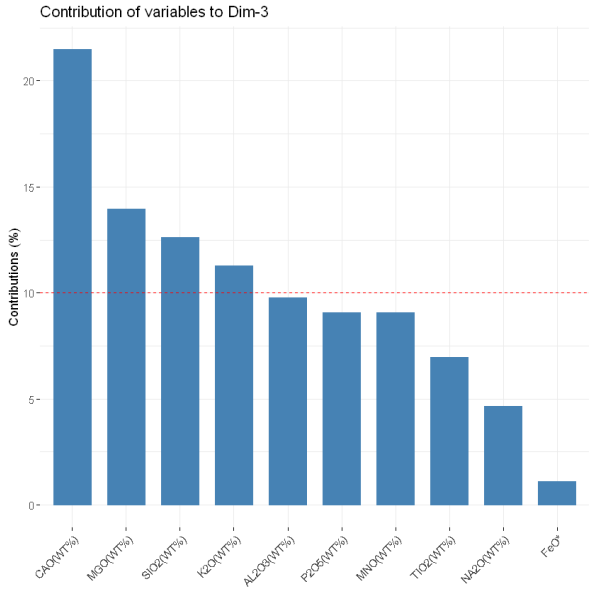


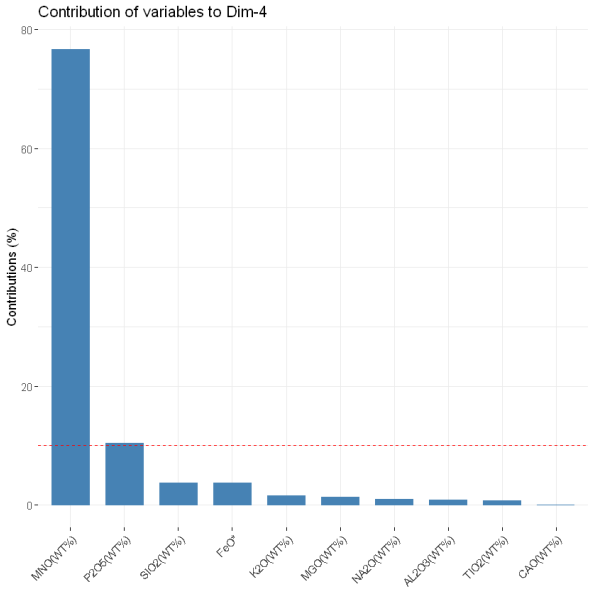
The scree plot to the left shows the percentage of variances explained by each principal component. From this table, only 5 of the principal components are well represented in the data. A look at the correlation plot on the right shows exactly which of the principal components are well represented in which dimensions. Subsequently only the first five dimensions present within the data will be further examined.

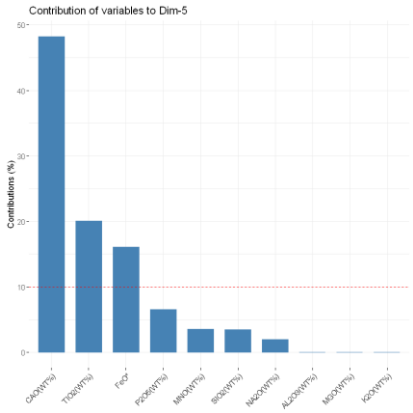
Contributions





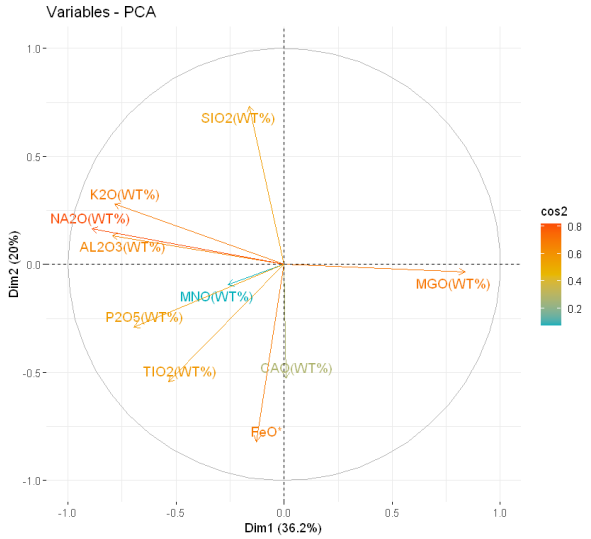




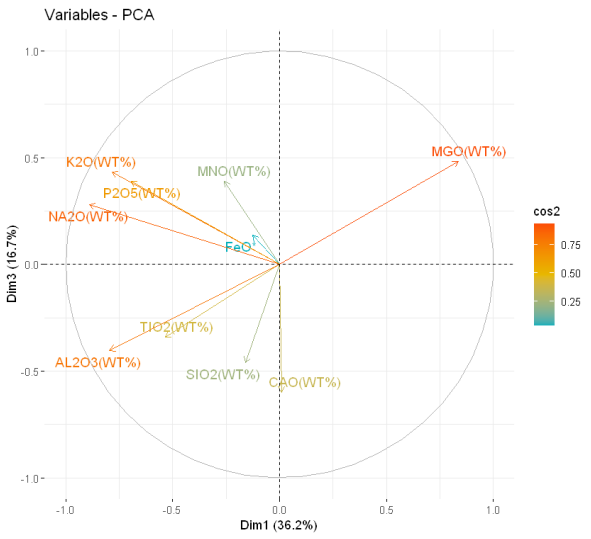


Dimension 1

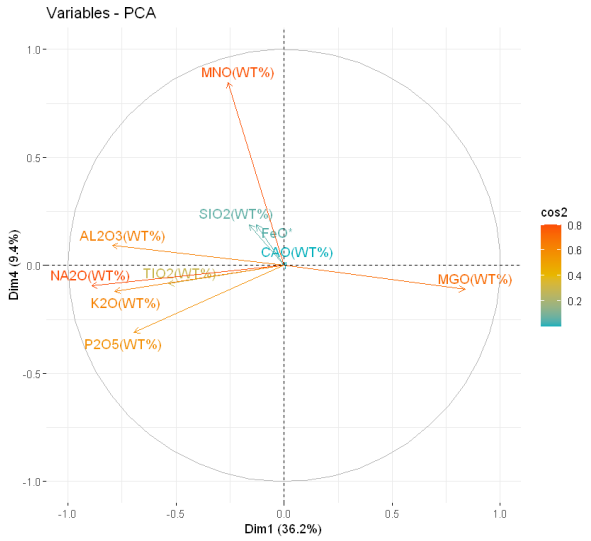
D1 vs D2



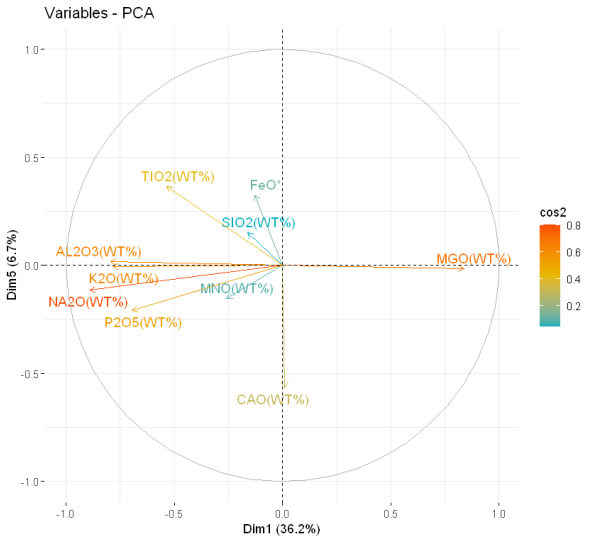
D1 vs D3



D1 vs D4

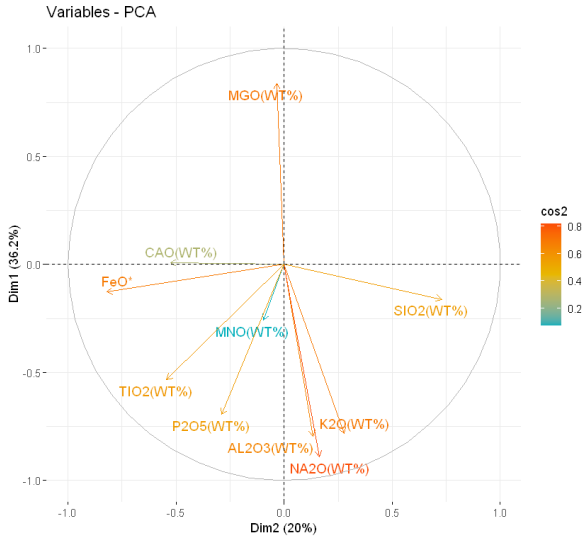


D1 vs D5

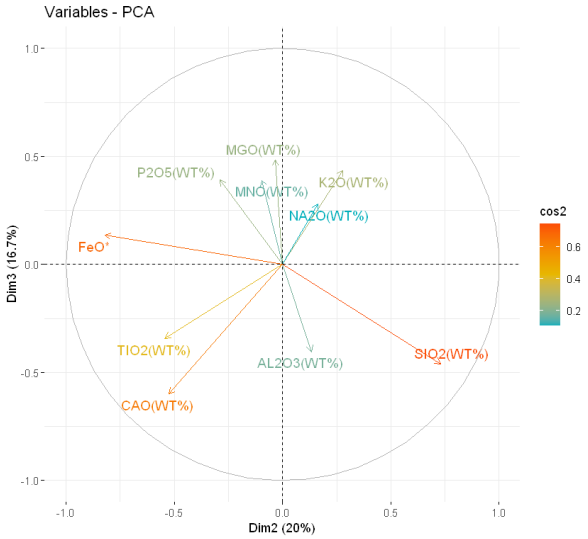


Dimension 2

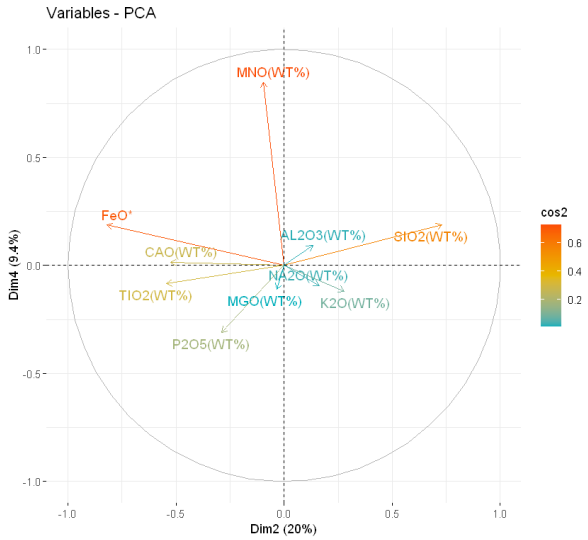
D2 vs D1



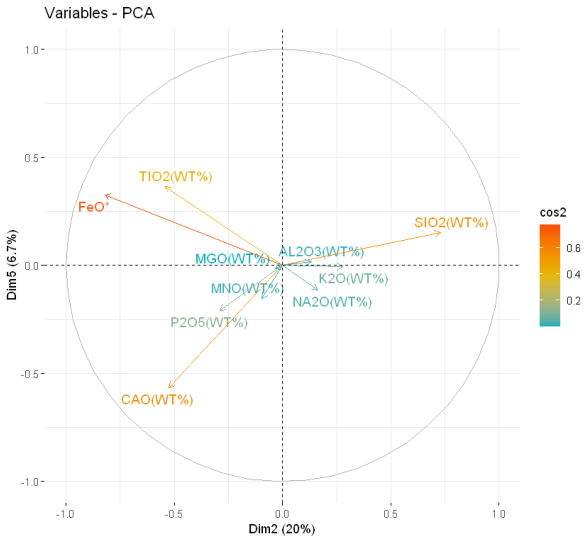
D2 vs D3



D2 vs D4

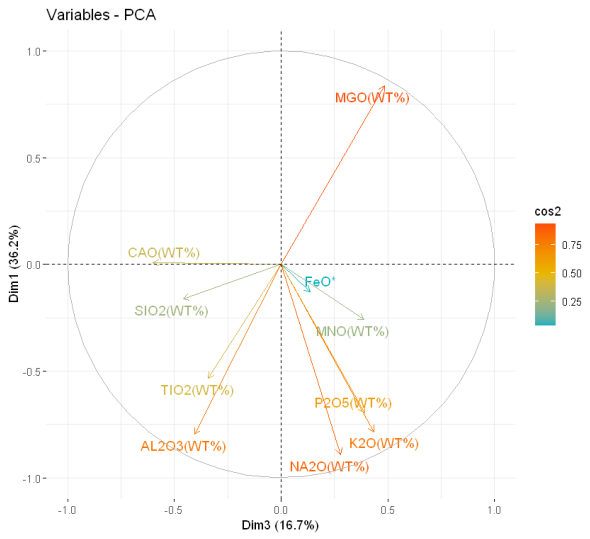


D2 vs D5

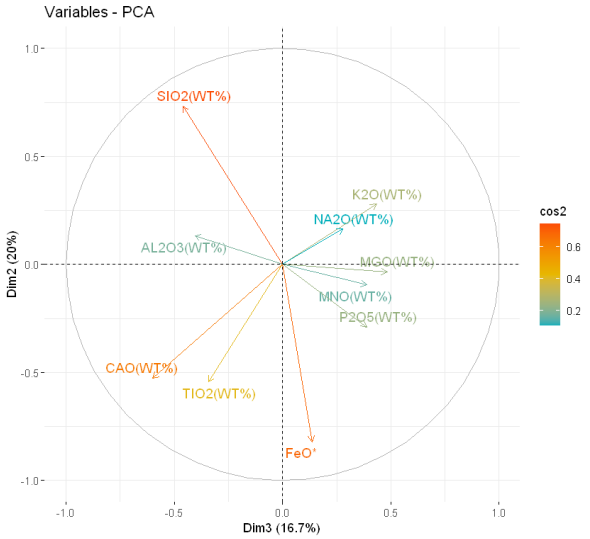


Dimension 3

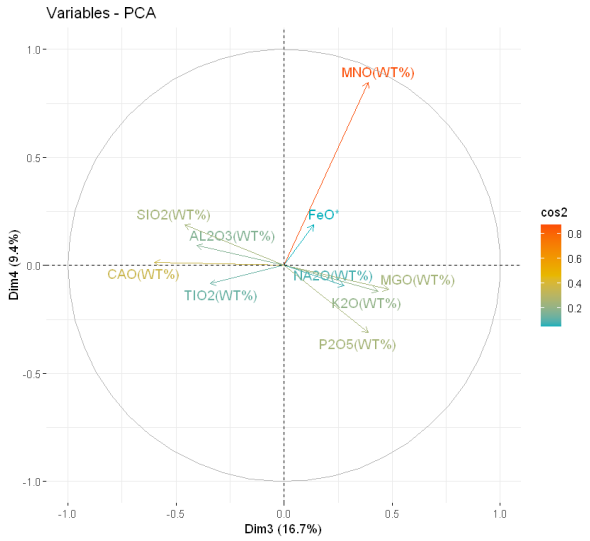
D3 vs D1



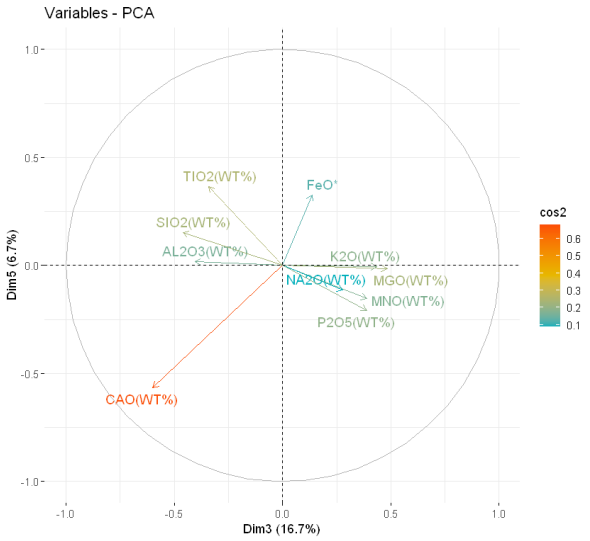
D3 vs D2



D3 vs D4

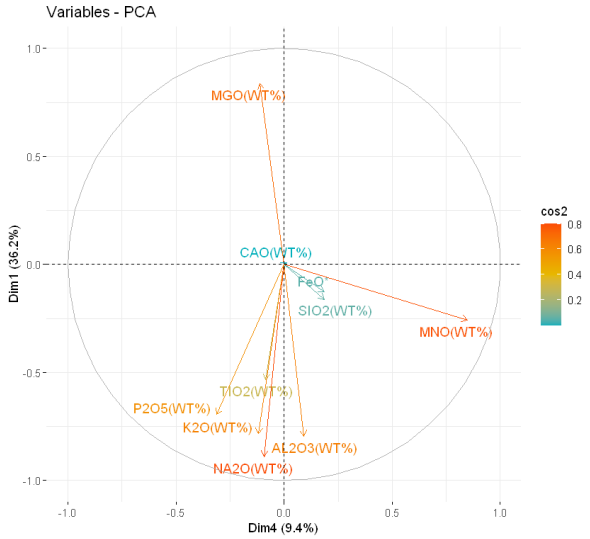


D3 vs D5

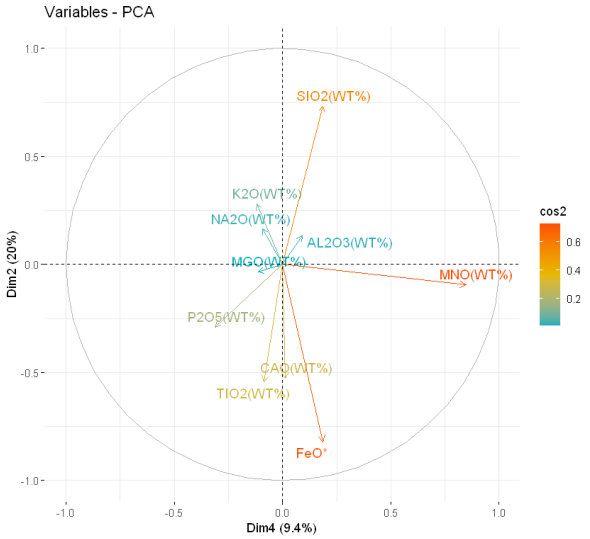


Dimension 4

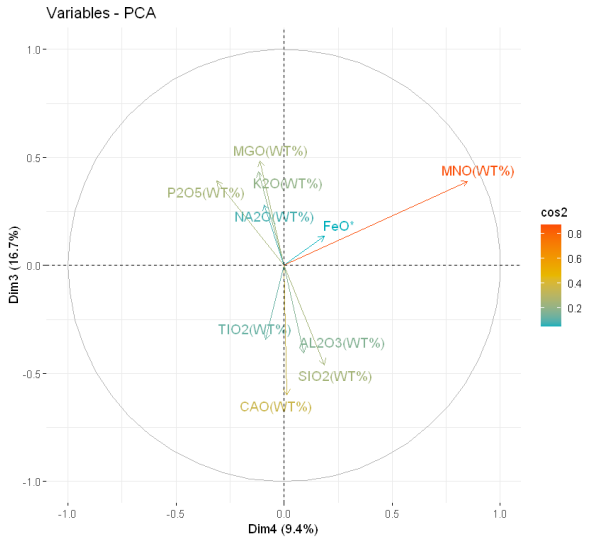
D4 vs D1



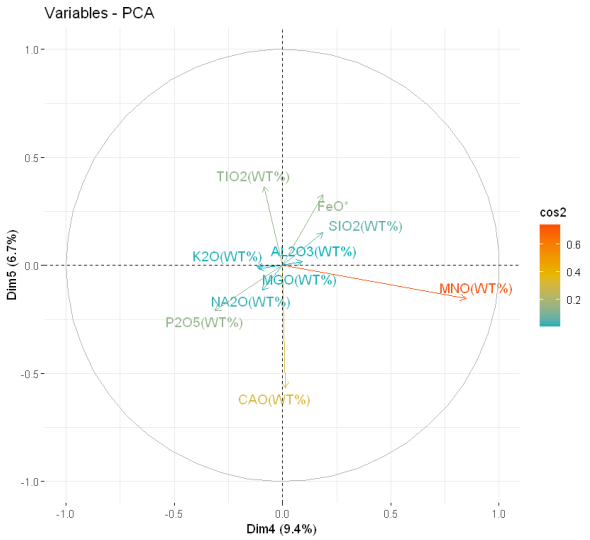
D4 vs D2



D4 vs D3

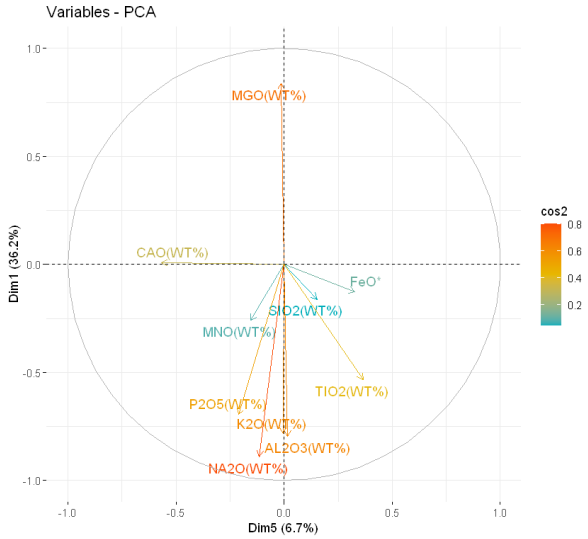


D4 vs D5

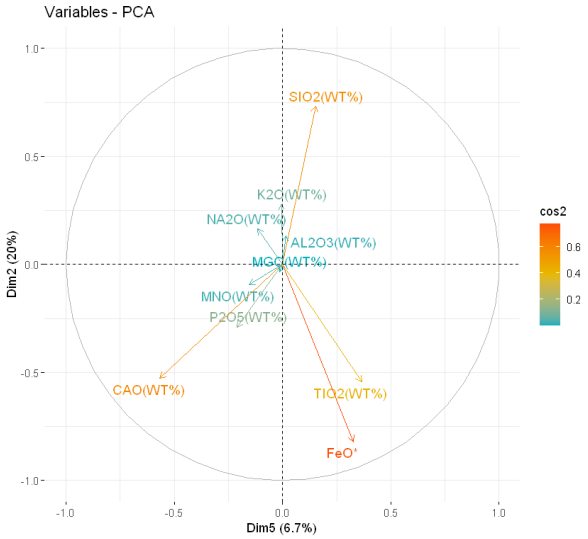


Dimension 5

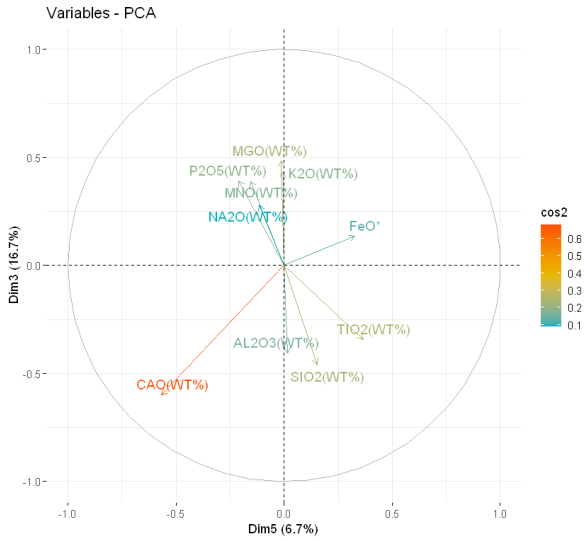
D5 vs D1



D5 vs D2



D5 vs D3



D5 vs D4