

The task for the PCA lab

The dataset comes from the previous lab about clustering the food matrices. For original data see 10.1016/j.aca.2010.06.017

This lab aims to see how are different nutritional properties influencing the data at hand and use PCA as a data pre-processing and visualisation tool.

Input data: DaSilva.csv after removing missing values and features with near-zero-variance. We can leave the correlated variables in (why?).

1. Carry out PCA. Analyze the results. How many PC's are reasonable?
2. Plot the data for relevant PC's and add the grouping information from lab 3 (k-means and hierarchical clustering).
3. Which properties/features are helping the separation of the groups? Are any redundant?
4. Now, try using PCA scores (relevant number!) as inputs for the clustering (e.g. k-means). Do you note any differences?

You should propose which fruits & vegetables to use for developing the sample preparation. Submit your R code as "*name.R*" file. Keep your code commented. All comments are to be written into the code. Use

```
#comment
```

for this.