

Extensions

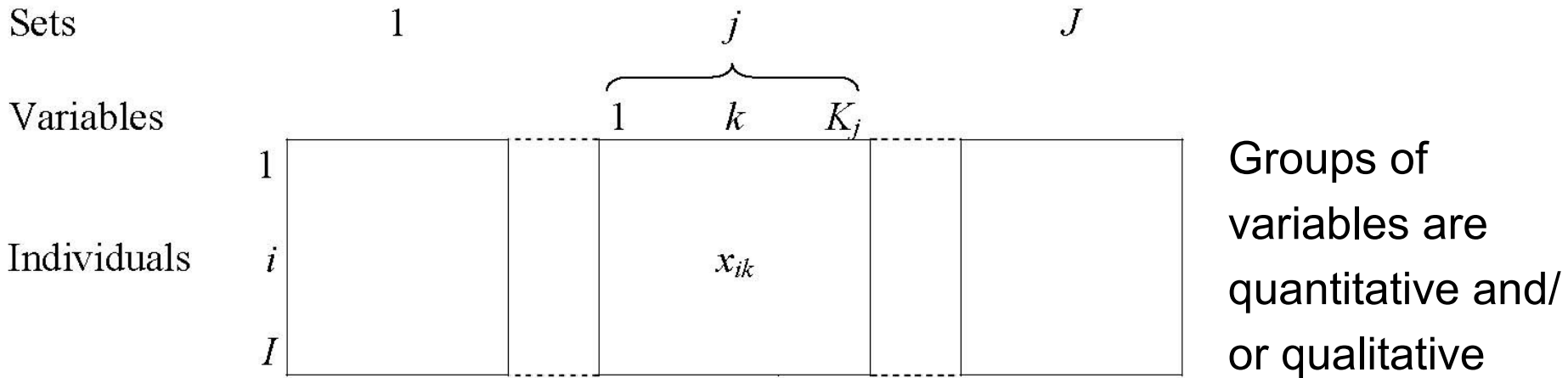
- Different structures on the data
- Mixed data (continuous and categorical variables)
- Missing values
- Graphical User Interface

Structure on the data

Different structures on the data are proposed:

- ❑ a partition on the variables: several sets of variables are simultaneously studied: [Multiple Factor Analysis](#)
- ❑ a hierarchy on the variables: variables are grouped and subgrouped (like in questionnaires structured in topics and subtopics): [Hierarchical Multiple Factor Analysis](#)
- ❑ a partition on the individuals: several sets of individuals described by the same variables: [Dual Multiple Factor Analysis](#)

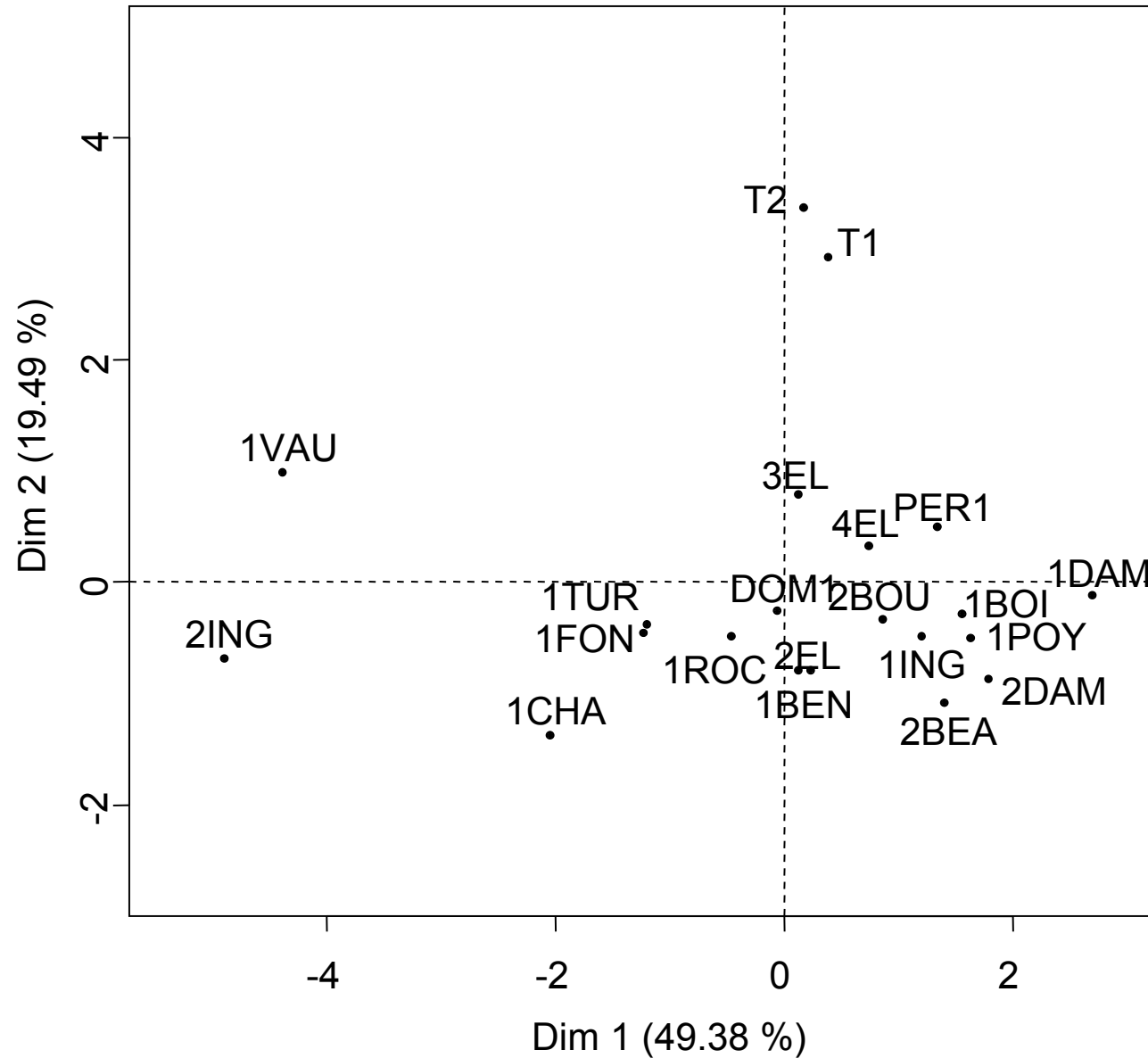
Groups of variables (MFA)



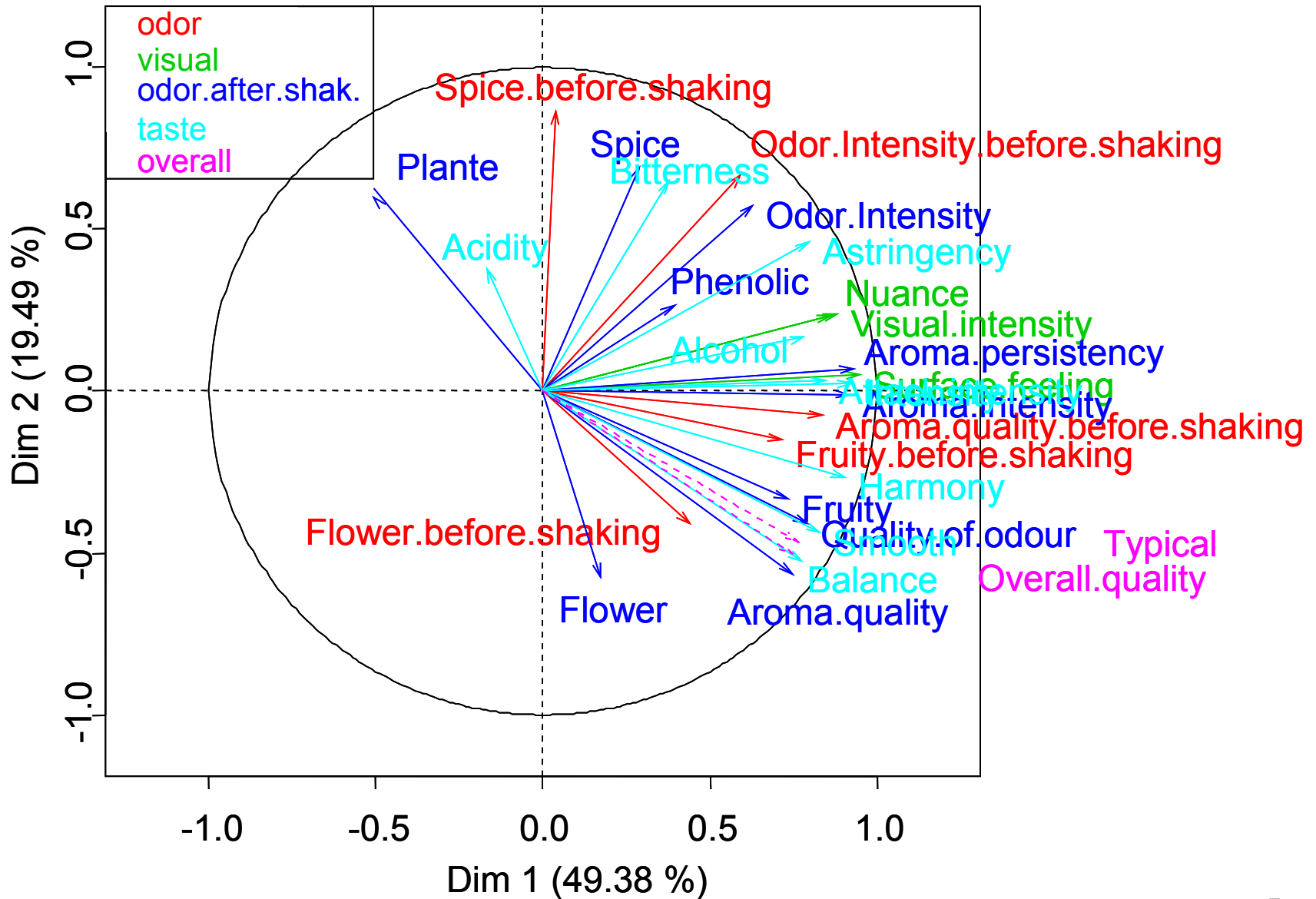
- Objectives:
- study the link between the sets of variables
 - balance the influence of each group of variables
 - give the classical graphs but also specific graphs:
 - groups of variables - partial representation

- Examples:
- Genomic: DNA, protein
 - Sensory analysis: sensorial, physico-chemical
 - Comparison of coding (quantitative / qualitative)

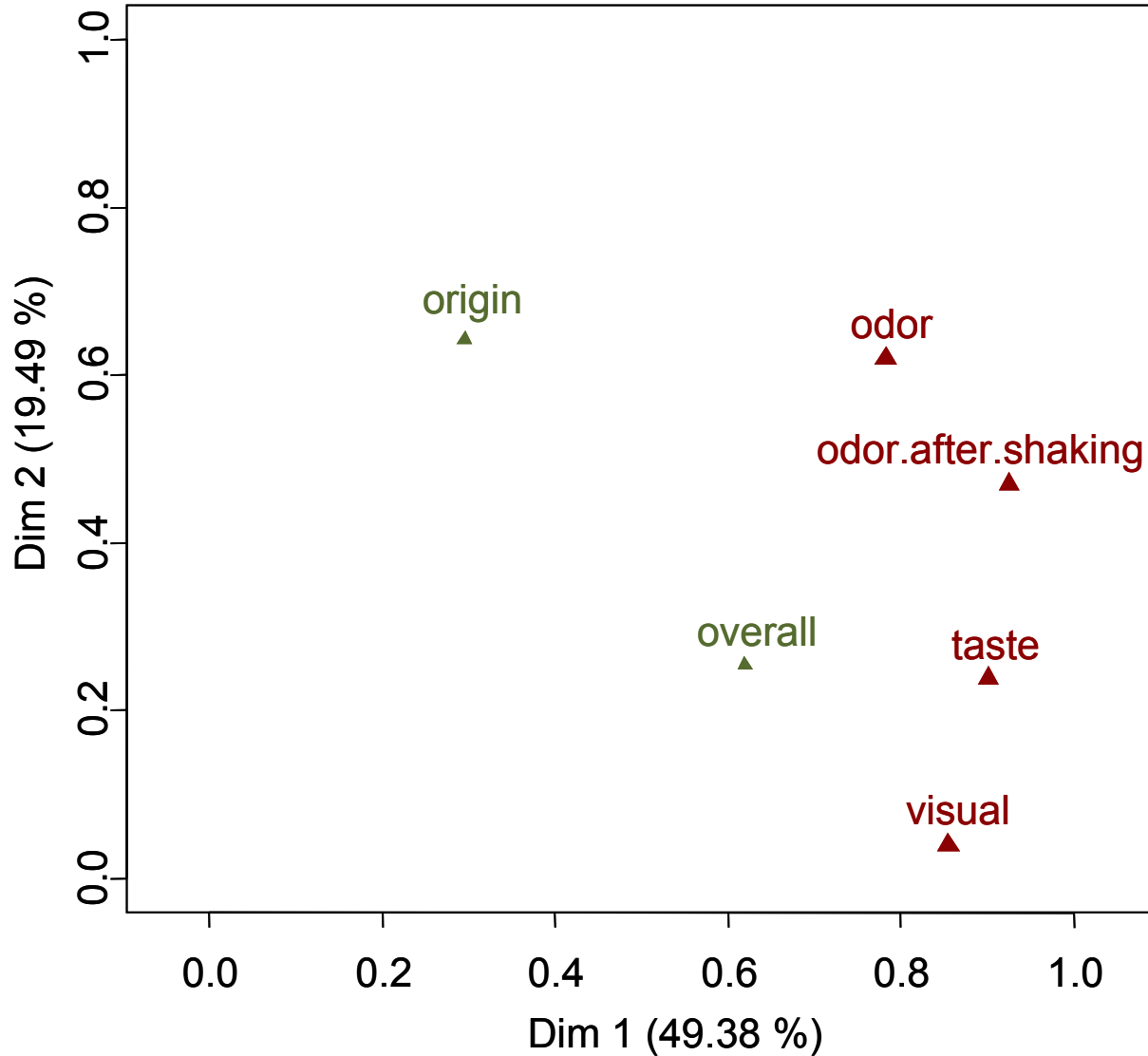
MFA example: representation of the individuals



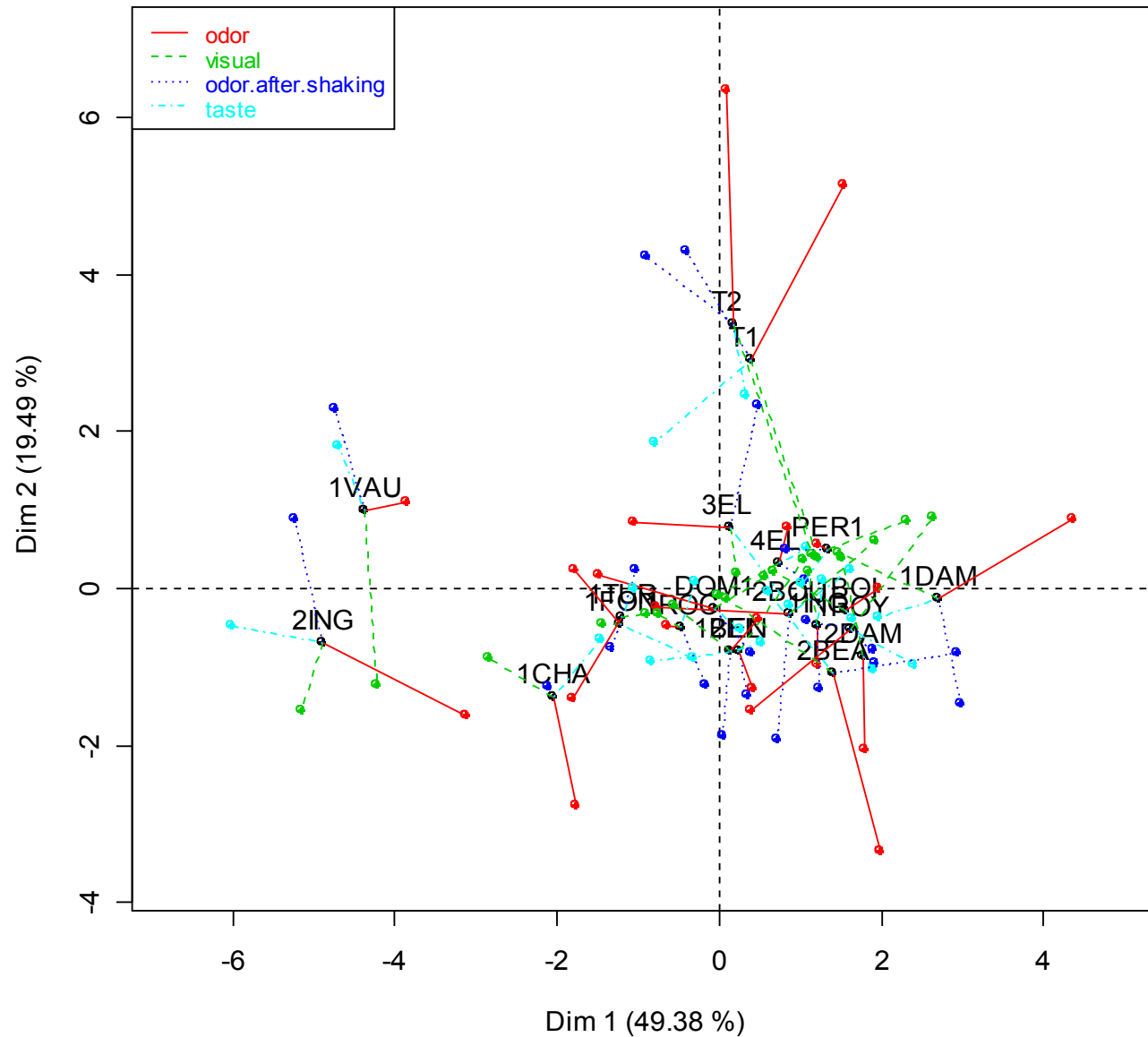
MFA example: representation of the variables



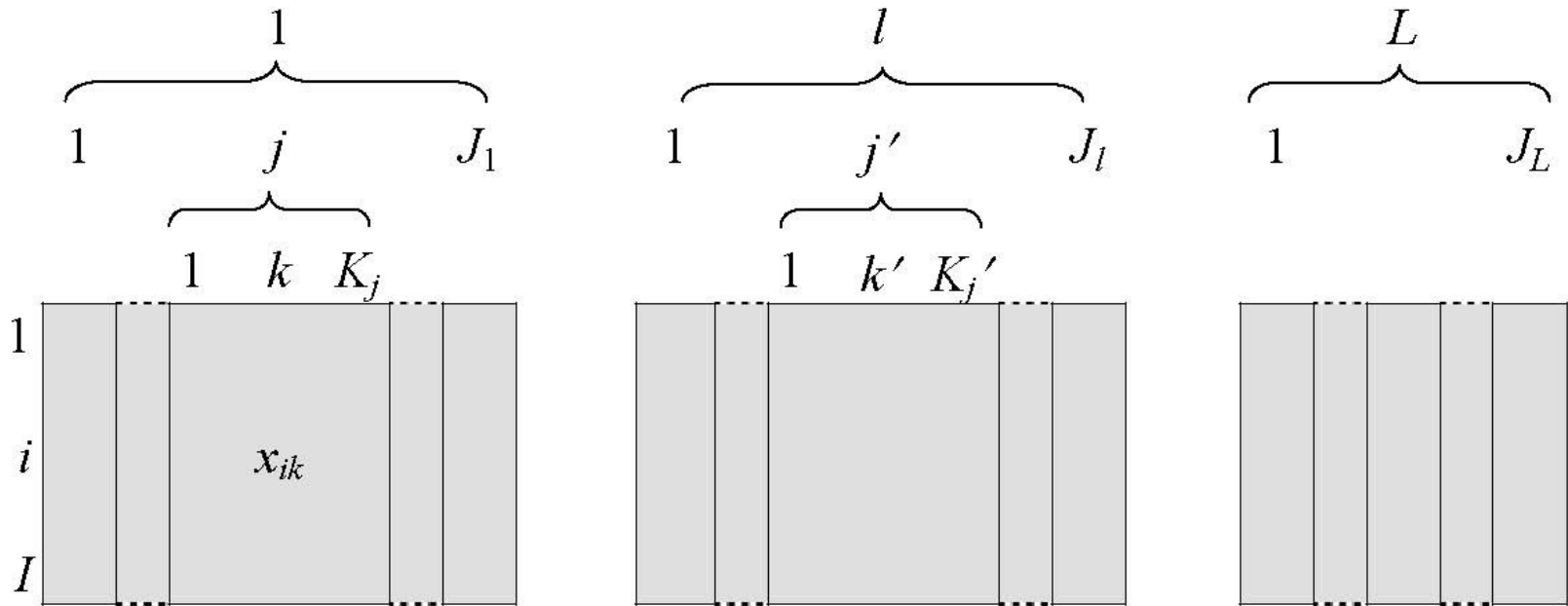
MFA example: representation of the groups



MFA example: representation of the partial points



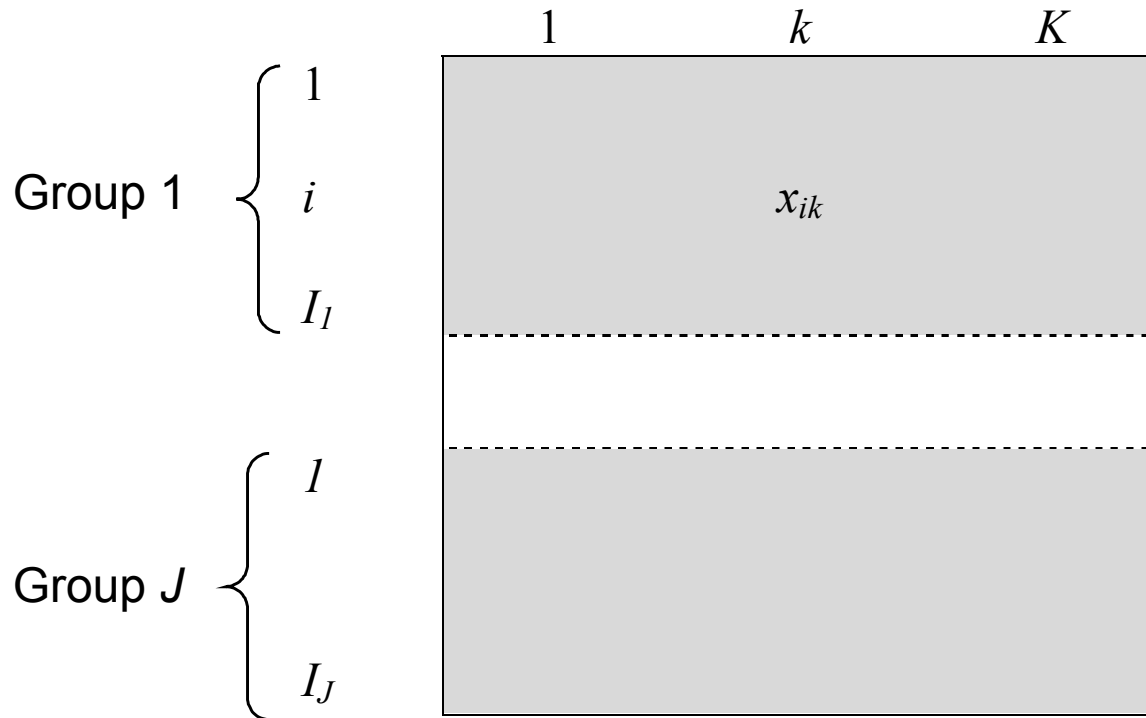
Hierarchy on the variables (HMFA)



Two levels for the hierarchy: the first one contains L groups, each l group contains J_l subgroups, and each subgroup have K_j variables

Objective: to balance the groups and the subgroups of variables

Partition on the individuals (DMFA)



Objective: to compare the covariance matrices

Mixed data analysis

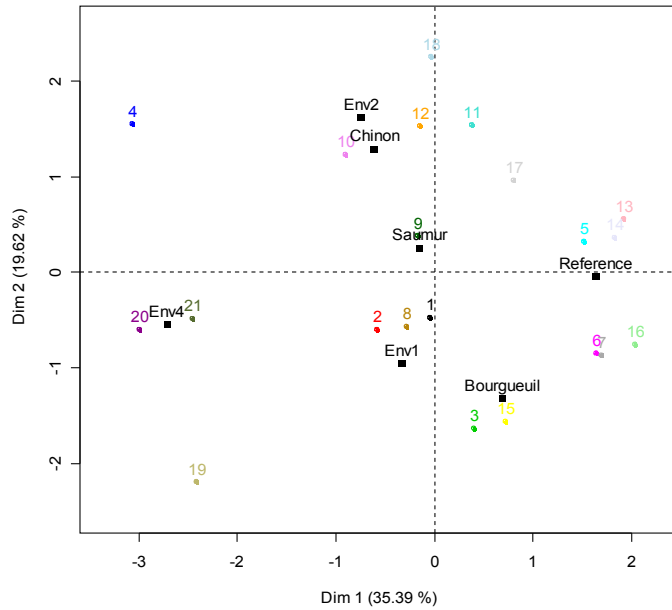
Objectives: - study simultaneously continuous and categorical variables
- balance the influence of each variable

Outputs: - representation of the individuals and the categories
- representation of the correlation circle graph
- representation of all the variables

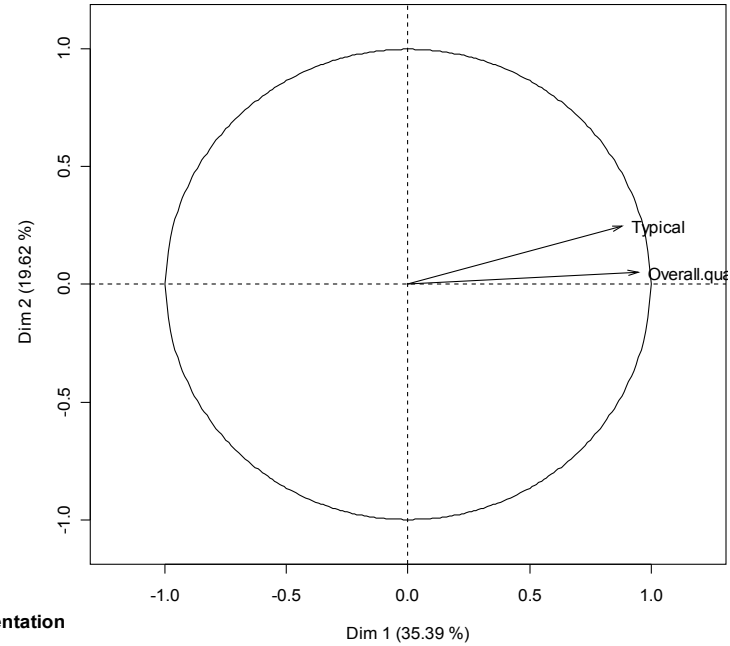
Remark: this method gives the same results than PCA for continuous variables and the same results than MCA for categorical variables

Mixed data analysis

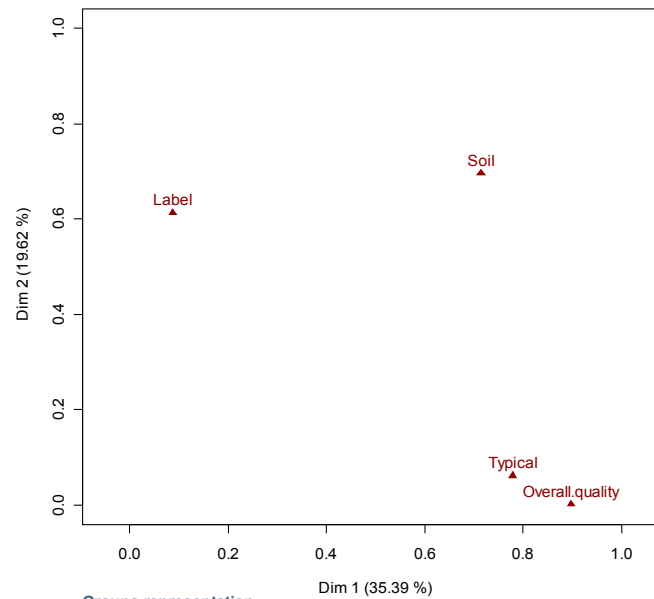
Individual factor map



Correlation circle



Variables representation



Groups representation

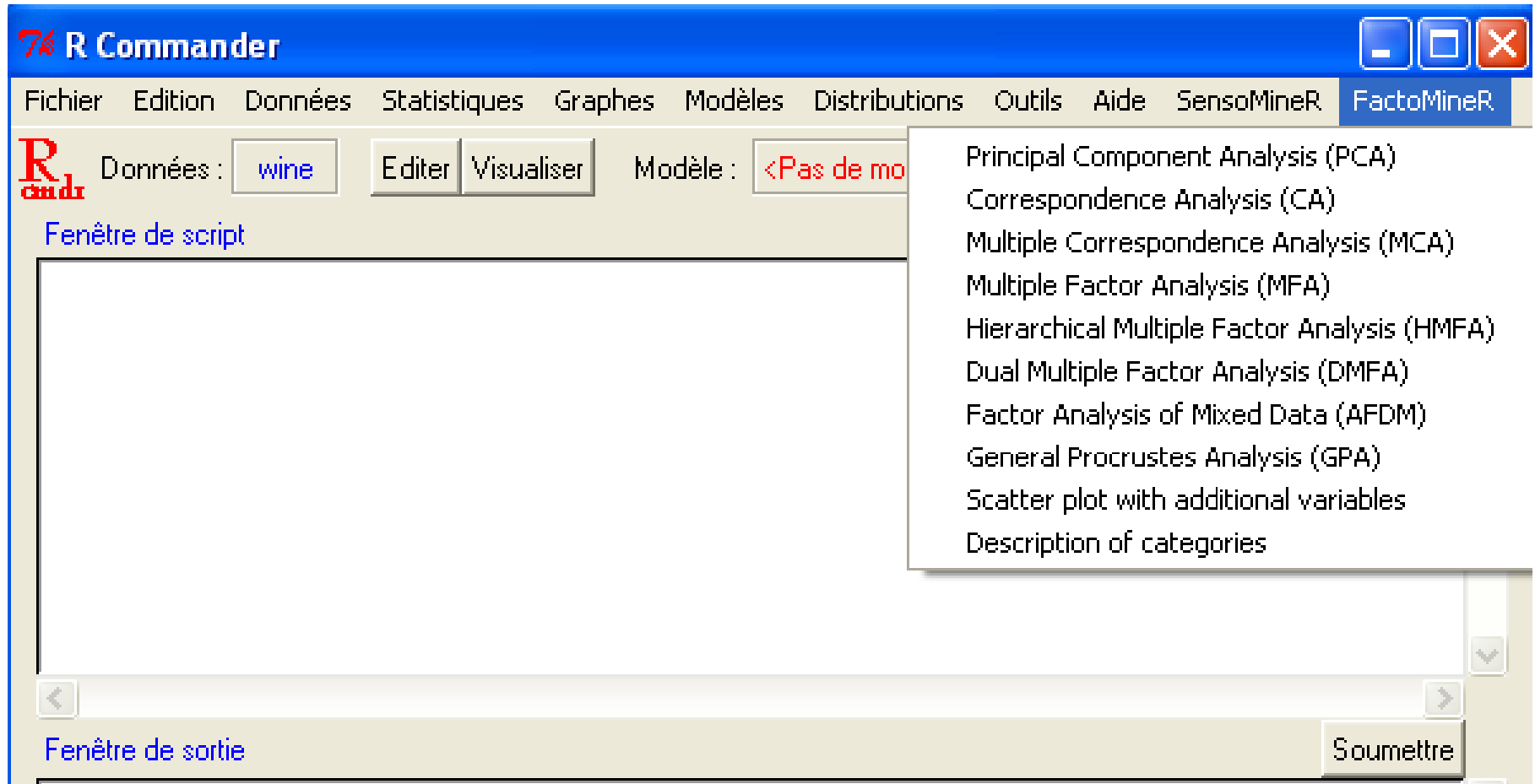
Missing values

- Delete individuals with missing values (not a good idea!)
- Replace missing values by the mean of the variable (for PCA)
- Use imputation method
- Use specific algorithms:
 - nipals, EM algorithm
 - the principle of the EM algorithm is:
 - make factorial analysis
 - complete the data using the factorial results
 - do this to steps alternatively until convergence

Graphical User Interface

The GUI can be simply loaded:

```
source(http://factominer.free.fr/install-facto.r)
```



Menu of the FactoMineR GUI

Graphical User Interface

Main window
of the PCA

PCA

Principal Components Analysis (PCA)

Select active variables (by default all the variables are active)

- X100m
- Long.jump
- Shot.put
- High.jump
- X400m
- X110m.hurdle
- Discus
- Pole.vault
- Javeline
- X1500m

Modify supplementary factors Modify supplementary variables Select supplementary individuals

Graphical options Outputs Restart

Main options

Name of the result object:

Number of dimensions:

Scaled the variables of the group:

Graphical output : select the dimensions :

Apply

OK Annuler Aide

Graphical User Interface

Graphical options

74 Graphical options

Plot individuals graph

Title of the graph

Hide some elements:

ind ind sup quali

Label for the active individuals

Label for the supplementary factor

Color of the active individuals

Color for factors

Coloring for individuals

by.individual

Competition

x limits of the graph:

y limits of the graph:

Plot variables graph

Title of the graph

Draw variables with a $\cos^2 >$:

Labels for the active variables

Labels for the supplementary variables

Color for active variables

Color for supplementary variables

Bibliography

Greenacre M (1984). Theory and Applications of Correspondence Analysis. Academic Press, London, England.

Greenacre M (2007). Correspondence Analysis in Practice. Chapman & Hall/CRC, Boca Raton, FL, 2nd edition

Greenacre M, Blasius J (2006). Multiple Correspondence Analysis and Related Methods. Chapman & Hall/CRC, Boca Raton, FL.

Escofier B and Pagès J (1988, 1990, 1993, 1998) Analyses factorielles simples et multiples. Dunod.

Escofier B, Pagès J (1994). Multiple Factor Analysis (AFMULT package). Computational statistics & data analysis 18 121-140

Jambu M (1991) Exploratory and Multivariate Data. Elsevier.

Jolliffe I Principal Component Analysis Springer. 2nd edition 2002

Tenenhaus M, Young FW (1985). "An Analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis and Other Methods for Quantifying Categorical Multivariate Data." Psychometrika, 50(1), 91-119.

Bécue M & Pagès J (2003) A principal axes method for comparing contingency tables: MFACT. Computational Statistics and Data Analysis. (45) 3 481-503

Bibliography

Packages

Chessel D, Dufour AB, Thioulouse J (2004). "The ade4 package I: One-table methods." R News, 4/1, 5-10.

de Leeuw J, Mair P (2008). "Homogeneity Analysis in R: The package homals." Journal of Statistical Software.

Dray S, Dufour AB, Chessel D (2007). "The ade4 package II: Two-table and K-table methods." R News, 7/2, 47-52

Lê S, Josse J & Husson F (2008). "FactoMineR: An R package for multivariate." Journal of Statistical Software, 25(1), 1-18.

Nenadi O, Greenacre M (2006). "Correspondence analysis in R, with two- and three-dimensional Graphics: The ca package." Journal of Statistical Software, 20(3), 1-13.