

A R package for analyzing Free Sorting data.

Philippe Courcoux

ONIRIS, USC "Sensometrics and Chemometrics Laboratory", Nantes, France.

philippe.courcoux@oniris-nantes.fr

OVERVIEW

Free sorting is becoming more and more popular as a tool to investigate the perception of a panel of subjects towards a set of stimuli. In the course of a free sorting experiment, each subject is asked to give a partition of the whole set of objects. This task is very natural and simple and may be used with untrained assessors.

The package FreeSortR gives to the practitioner several tools for analysing data resulting from this task:

- ✓ Factorial representation of the distances between products using Mds techniques,
- ✓ Interpretation of the configuration of stimuli by representation of the words given during a verbalization task,
- ✓ Computation of the consensus partition among the subjects, giving a segmentation of the products.

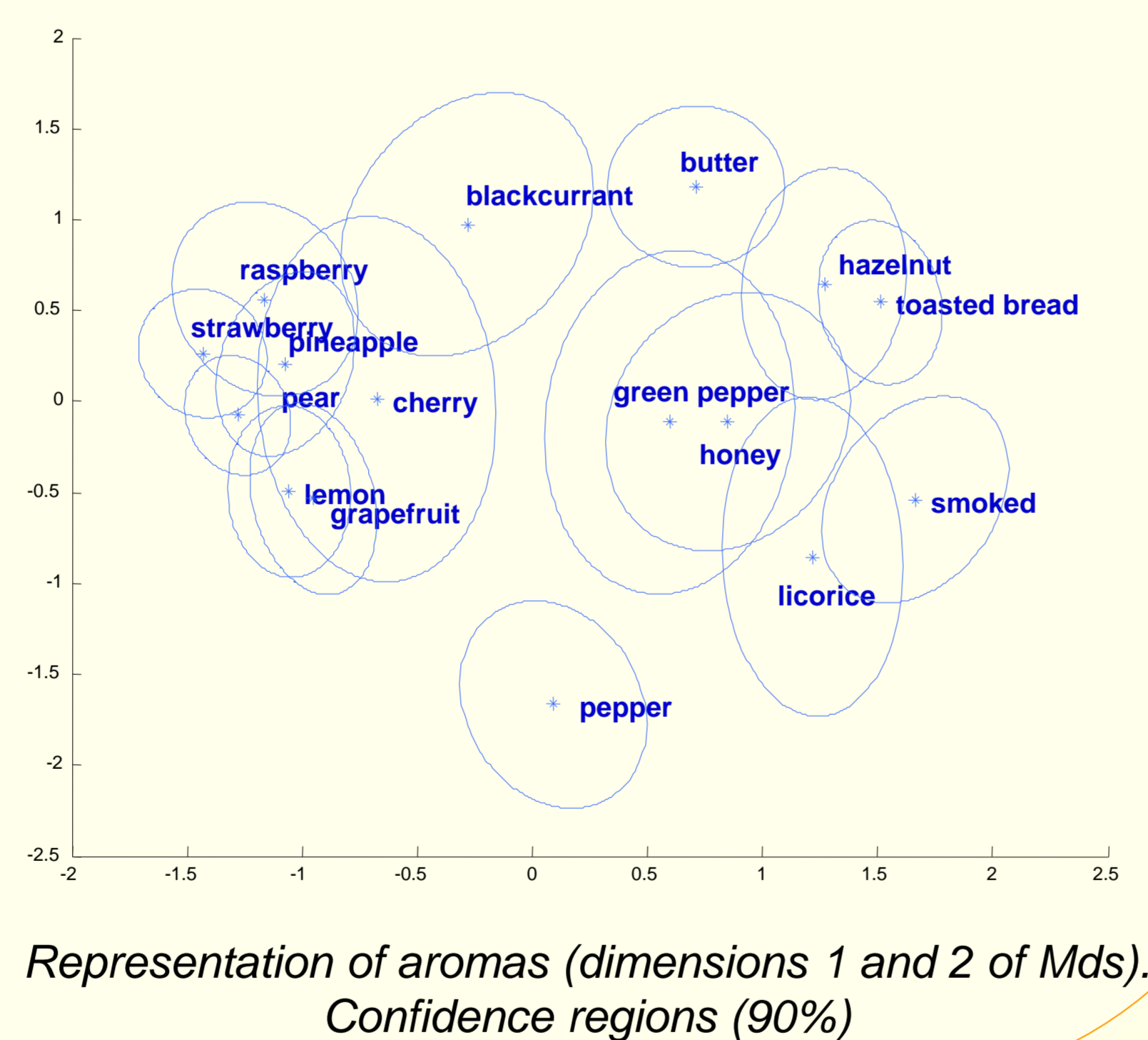
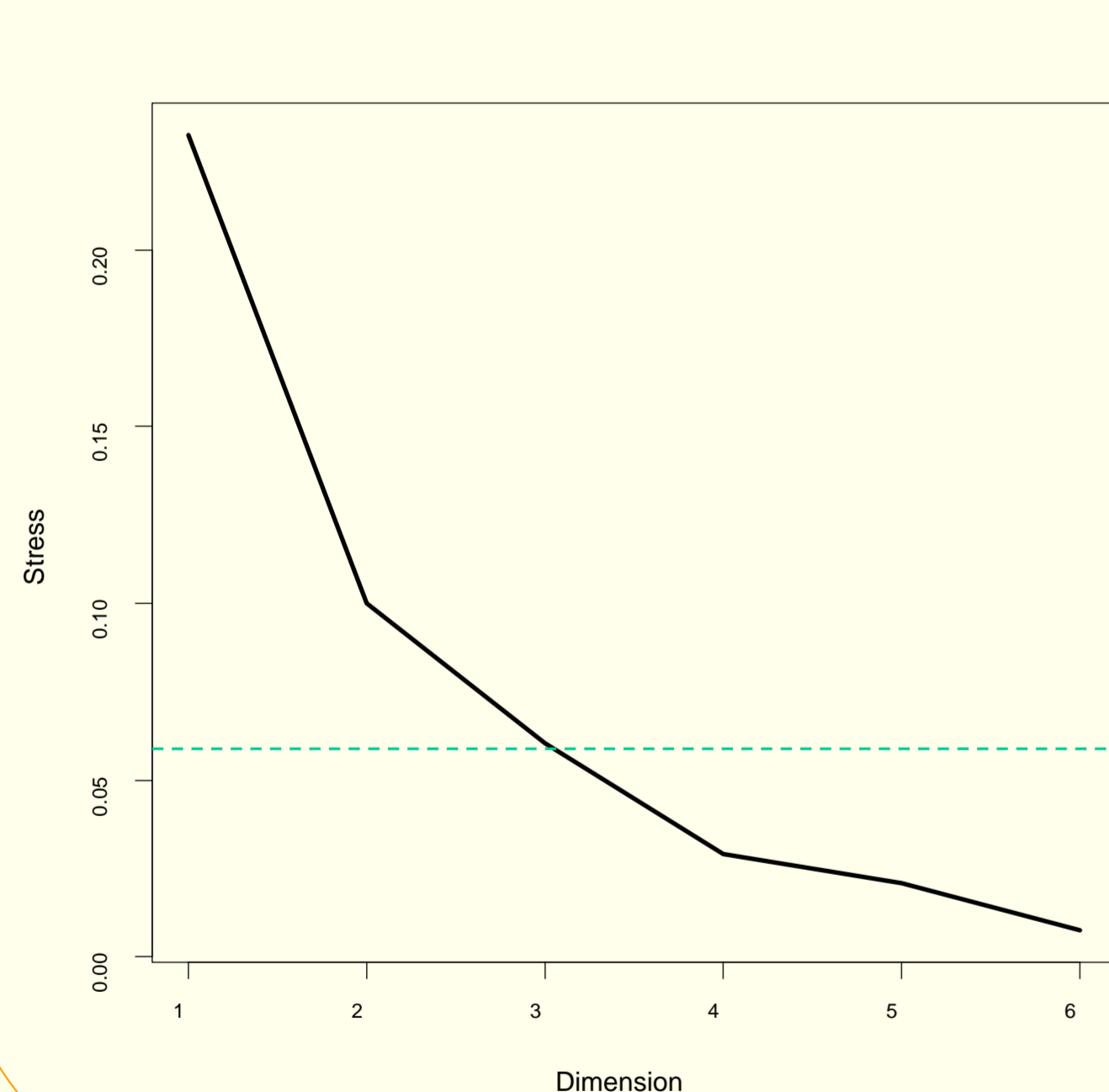
REPRESENTATION OF PRODUCTS

After a free sorting task, data for each subject are often expressed in a symmetric (products by products) matrix depicting the dissimilarity between stimuli (1 for two items in different groups and 0 if not). At the panel level, the individual dissimilarity matrices can be averaged across subjects yielding a global dissimilarity matrix between products.

The package FreeSortR gives several Multidimensional Scaling (MDS) solutions for representing the global dissimilarity matrix: metric and non-metric Mds based on the Smacof algorithm.

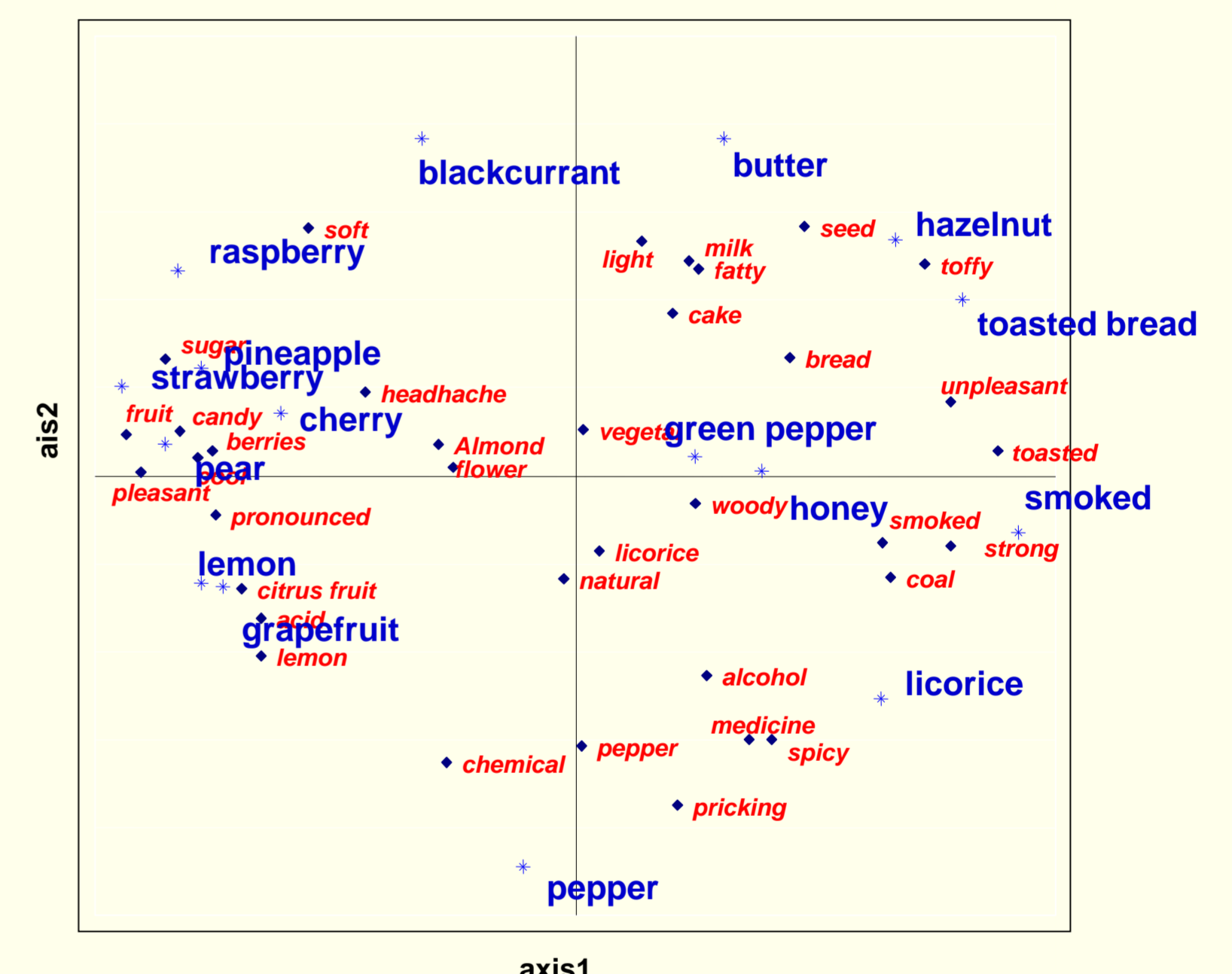
Confidence regions around products may be drawn using a bootstrapping on subjects.

Application: 16 aromas sorted by 31 subjects.



VERBALIZATION TASK

Usually, a verbalization task is performed by the subjects after the sorting task : subjects are asked to freely describe with their own words and expressions the groups they formed. These verbatims are used for interpret the products configuration..



CLUSTERING OF SUBJECTS

The similarity between two subjects can be assessed by computing the adjusted Rand Index between the two partitions given by these subjects. A strategy for segmenting the panel may be based on a clustering the subjects on the basis of the matrix of adjusted Rand Index between subjects.

CONSENSUS PARTITION

It may also be of interest to find a central partition that stands as a group average partition to a group of subjects. The Rand Index is used to determine a central (or consensus) partition to the set of partitions given by a group of subjects. This partition is assumed to be the closest partition to the partitions under consideration.

In the case of the aroma, the consensus partition has 7 classes:

Lemon, Grapefruit, Pineapple, Pear, Strawberry, Raspberry Honey, Toasted bread, Hazelnut
Cherry Blackcurrant, Butter Pepper Licorice, Smoked Green pepper

References

- Gordon A.D. & Vichi M. (1998) Partitions of partitions. *Journal of Classification*, 15:265-285.
Guénoche A. (2011) Consensus of partitions : a constructive approach. *Adv Data Anal Classif*, 5:215-229.
Hubert L. & Arabie L. (1985) Comparing partitions. *Journal of Classification*, 2:193-218.
Krieger A.M. & Green P.E. (1999) A generalized Rand-Index method for consensus clustering of separate partitions of the same data base. *Journal of Classification*, 16:63-89.
Rand W.M. (1971) Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66:326.