



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*



ECFG932- ing3 FOO'DS Techniques avancées en évaluation sensorielle et étude consommateurs (2)

TRAITEMENT DE DONNÉES « TRI LIBRE »

Le Tri libre

Introduction

- Le Tri Libre fait partie des méthodes holistiques.
- Les méthodes holistiques
 - présentent l'ensemble des produits d'intérêt en même temps aux panélistes
 - mesurent les similarités ou dissimilarités entre les produits en fonction de leurs propriétés sensorielles globales
 - Ne nécessitent aucune caractérisation analytique contrairement aux méthodes descriptives.

Principe

- On demande à N participants de proposer une partition de P produits en sous ensembles disjoints.
- Le nombre de partitions sera en général compris entre 2 et $P - 1$.
- Il s'agit d'une technique descriptive et non verbale qui est proposée à des sujets non entraînés.

Avantages

- Lorsque le nombre de produits est relativement important pour éviter la lourdeur cognitive et la fatigue induite des méthodes descriptives.
- Peut capter des informations latentes difficiles à verbaliser par des descripteurs sensoriels analytiques.

Inconvénients

- Dépendant de l'espace produit
- Impossible d'agréger les données de plusieurs études (facilement)
- Besoin d'informations descriptives additionnelles pour comprendre l'espace produit

Du côté de R

Package **FreeSorteR** écrit par P. Courcoux.

On va se baser sur un des exemples de ce package **AromaSort**.

Data AromaSort dans FreeSortR

On va étudier un premier exemple de tri libre de 16 arômes proposés à 31 sujets.

Food Quality and Preference 32 (2014) 107–112



Contents lists available at [SciVerse ScienceDirect](#)

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual



Determination of the consensus partition and cluster analysis of subjects in a free sorting task experiment



Ph. Courcoux *, P. Faye, E.M. Qannari

LUNAM University, ONIRIS, USC Sensometrics and Chemometrics Laboratory, Nantes F-44322, France
INRA, Nantes F-44316, France

Exemple individu 1 :

[1] "{Citron, pamplemousse, poire, noisettes grillées, fraise}{ananas, miel, poivre vert}{beurre, pain grillé}{framboise, cerise, cassis}{fumé, poivre, réglisse}"



Stimuli	Ind1
Citron	5
pamplemousse	5
ananas	4
poire	5
miel	4
beurre	1
pain grillé	1
noisettes grillées	5
fraise	5
framboise	3
cerise	3
cassis	3
poivre vert	4
fumé	2
poivre	2
réglisse	2

Matrice de dissimilarités individuelles

On note $D^{(n)} = (d_{i,j}^{(n)})_{i,j}$ la matrice de dissimilarités obtenue à partir de la partition du n -ième individu. On a

$d_{i,j}^{(n)} = 0$ si i, j sont dans la même partition,

$d_{i,j}^{(n)} = 1$ sinon.

Remarque La matrice $D^{(n)}$ est une matrice symétrique dont la diagonale est nulle.

Exemple individu 1 :

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Citron	0	0	1	0	1	1	1	0	0	1	1	1	1	1	1	1
2. pamplemousse	0	0	1	0	1	1	1	0	0	1	1	1	1	1	1	1
3. ananas	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1
4. poire	0	0	1	0	1	1	1	0	0	1	1	1	1	1	1	1
5. miel	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1
6. beurre	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1
7. pain grillé	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1
8. noisettes grillées	0	0	1	0	1	1	1	0	0	1	1	1	1	1	1	1
9. fraise	0	0	1	0	1	1	1	0	0	1	1	1	1	1	1	1
10. framboise	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1
11. cerise	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1
12. cassis	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1
13. poivre vert	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1
14. fumé	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
15. poivre	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
16. réglisse	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0

Matrice de dissimilarités à l'échelle du panel

On ajoute toutes les matrices de dissimilarités individuelles. On obtient la matrice D de coefficients :

$$\delta_{i,j} = \sum_{n=1}^N d_{i,j}^{(n)}$$

Cette matrice est symétrique et sa diagonale est nulle.

Tous les coefficients de cette matrice sont des entiers inférieurs ou égaux à N .

Retour sur l'exemple

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Citron	0	9	21	15	26	30	31	29	20	21	25	25	29	31	24	30
2. pamplemousse	9	0	18	17	26	29	31	29	20	24	26	25	27	31	24	29
3. ananas	21	18	0	15	27	28	30	30	14	20	22	24	25	31	29	31
4. poire	15	17	15	0	28	30	31	30	11	14	23	24	28	31	27	30
5. miel	26	26	27	28	0	23	22	22	28	30	28	25	24	24	26	23
6. beurre	30	29	28	30	23	0	22	22	28	27	27	21	28	28	30	27
7. pain grillé	31	31	30	31	22	22	0	12	31	30	28	27	24	23	30	24
8. noisettes grillées	29	29	30	30	22	22	12	0	30	29	28	26	26	24	29	27
9. fraise	20	20	14	11	28	28	31	30	0	14	22	26	30	31	30	31
10. framboise	21	24	20	14	30	27	30	29	14	0	24	22	27	30	31	28
11. cerise	25	26	22	23	28	27	28	28	22	24	0	26	29	30	26	28
12. cassis	25	25	24	24	25	21	27	26	26	22	26	0	27	30	29	28
13. poivre vert	29	27	25	28	24	28	24	26	30	27	29	27	0	26	27	26
14. fumé	31	31	31	31	24	28	23	24	31	30	30	30	26	0	28	22
15. poivre	24	24	29	27	26	30	30	29	30	31	26	29	27	28	0	24
16. réglisse	30	29	31	30	23	27	24	27	31	28	28	28	26	22	24	0

MDS (Multi Dimensional Scaling)

On cherche à représenter la matrice D qui est symétrique de dimension P dans un espace de dimension $k < P$. Cela suppose :

- L'existence de variables latentes (ie qui permettent de résumer des dissimilarités entre produits)
- Le choix d'un indice de qualité de la projection des dissimilarités dans l'espace latent (\leadsto **Stress**).

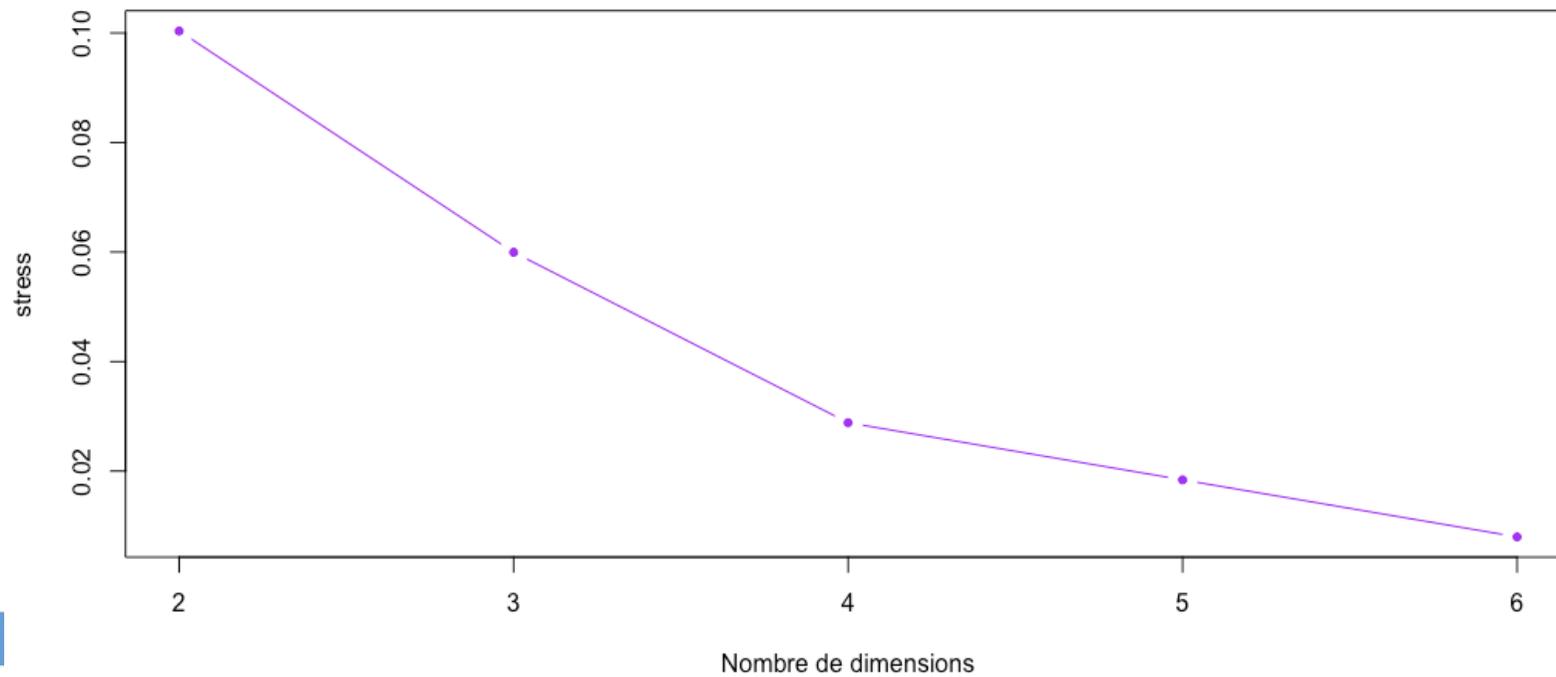
Le stress

Soit $D' = (\hat{\delta}_{i,j})_{i,j}$ la matrice des dissimilarités restituées dans l'espace latent on définit

- le stress $S = \sum_{i < j} (\delta_{i,j} - \hat{\delta}_{i,j})^2$

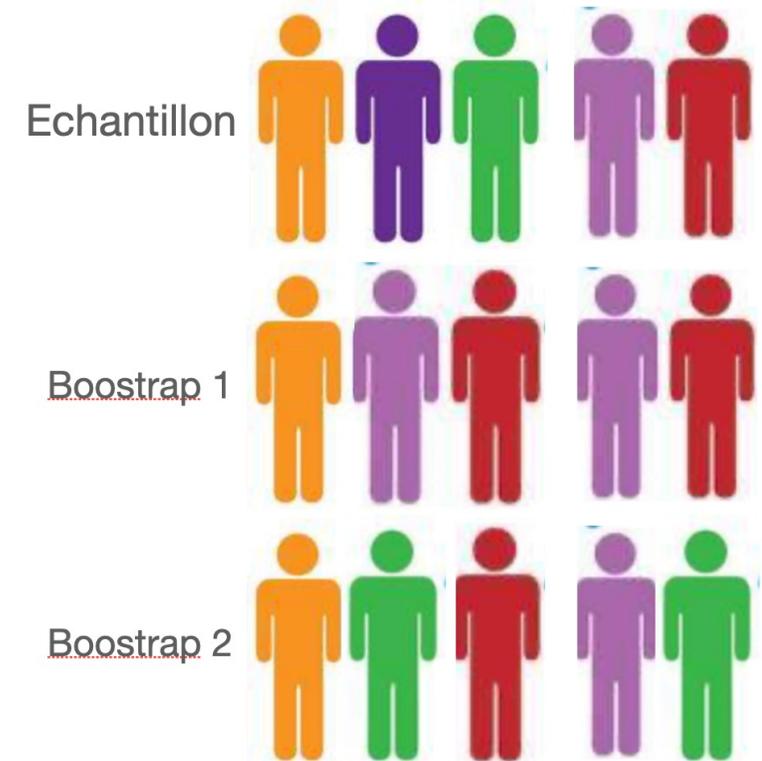
- le stress normalisé $S_n = \frac{\sum_{i < j} (\delta_{i,j} - \hat{\delta}_{i,j})^2}{\sum_{i < j} \delta_{i,j}^2}$

Choix du nombre de dimension du MDS

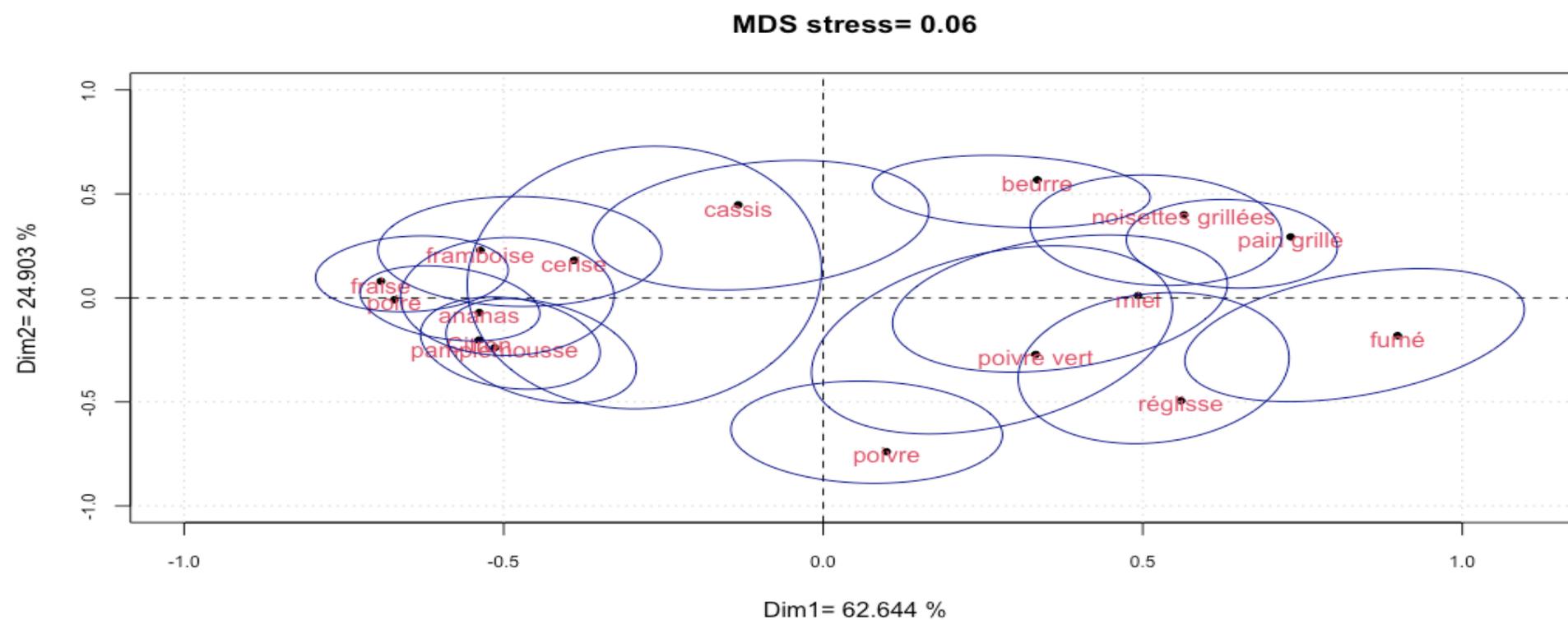


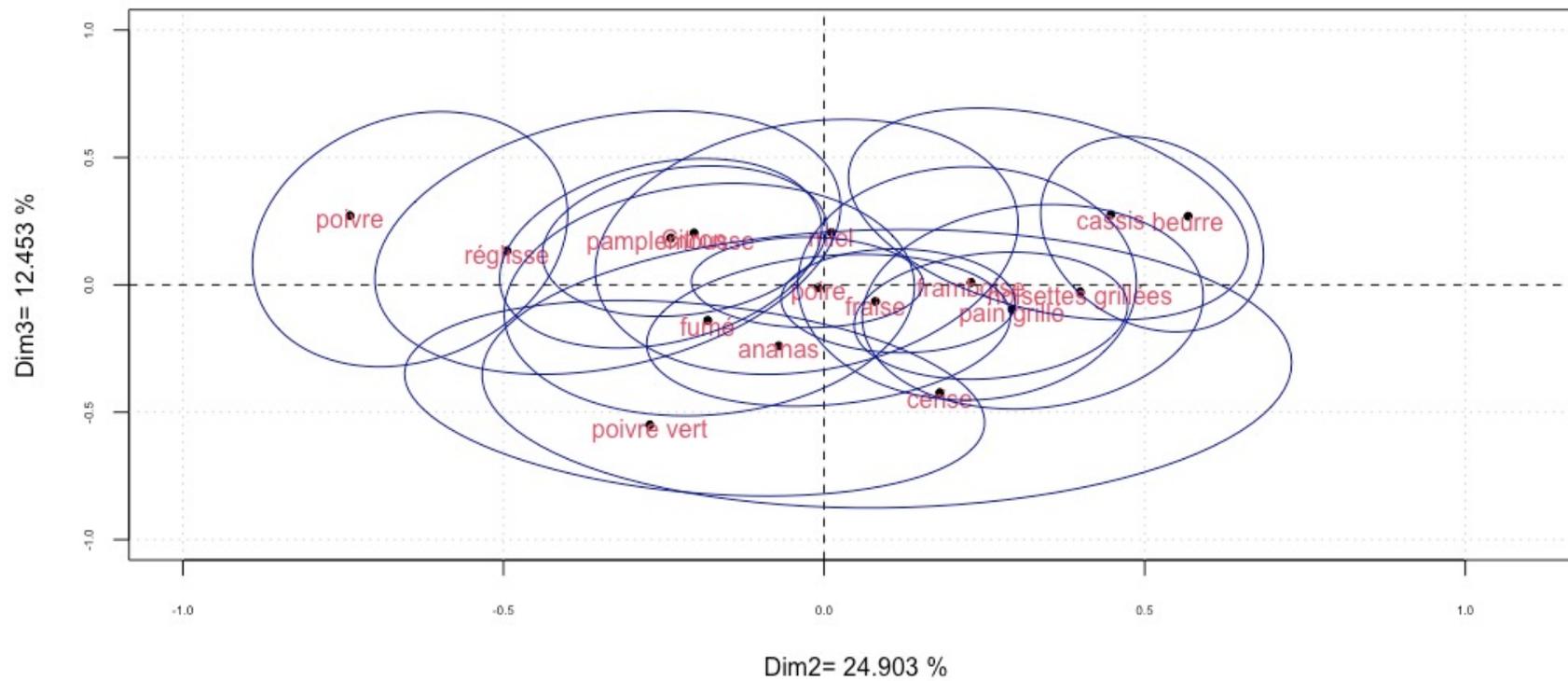
Ellipse de confiance

- On se pose la question de la stabilité de la configuration. D'autres individus proposeraient d'autres partitions \leadsto autre configuration.
- Pseudo-individus simulés à partir des individus de l'échantillon (Bootstrapping) : on tire des individus avec remise dans l'échantillon et on construit la configuration obtenue (en général $B = 500$ tirages).
- Construction d'ellipses de confiance à partir de ces B configurations.



Représentation obtenue





Partition consensuelle

Mesure d'accord entre deux partitions

On demande à 2 participants de partager en groupes 5 produits numérotés {1,2,3,4,5}.

On obtient les deux partitions suivantes : $P_1 = \{\{1,2,3\}, \{4,5\}\}$ et $P_2 = \{\{1,2\}, \{3,4,5\}\}$.

On va mesurer l'accord entre ces deux partitions en regardant si chaque paire de produits est :

groupé dans P_1 (Gr P1),

séparé dans P_1 (sep P1).

De même dans P_2 .

	Gr P2	Sep P2
Gr P1	4	2
Sep P1	2	2

Rand Index (RI)

$$RI(P_1, P_2) = \frac{a+d}{P(P-1)/2}$$

où $a + d$ est le nombre d'accord entre P_1, P_2 (ie lorsqu'elles regroupent ou qu'elles séparent les deux produits).

$$0 \leq RI(P_1, P_2) \leq 1$$

0 : désaccord total et 1 : accord total

Inconvénient : Il augmente en moyenne lorsque le nombre N de sujets augmente.

Adjusted Rand Index (ARI)

RI est remplacé par

$$ARI(P_1, P_2) = \frac{RI(P_1, P_2) - \overline{RI}}{1 - \overline{RI}}$$

où \overline{RI} est la moyenne du RI pour 2 partitions dont le consensus n'est dû qu'au hasard.

Conséquences

1. $ARI(P_1, P_2) = 0$ lorsque le consensus n'est dû qu'au hasard,
2. $ARI(P_1, P_2) = 1$ lorsque le consensus est parfait (ie les partitions sont identiques).

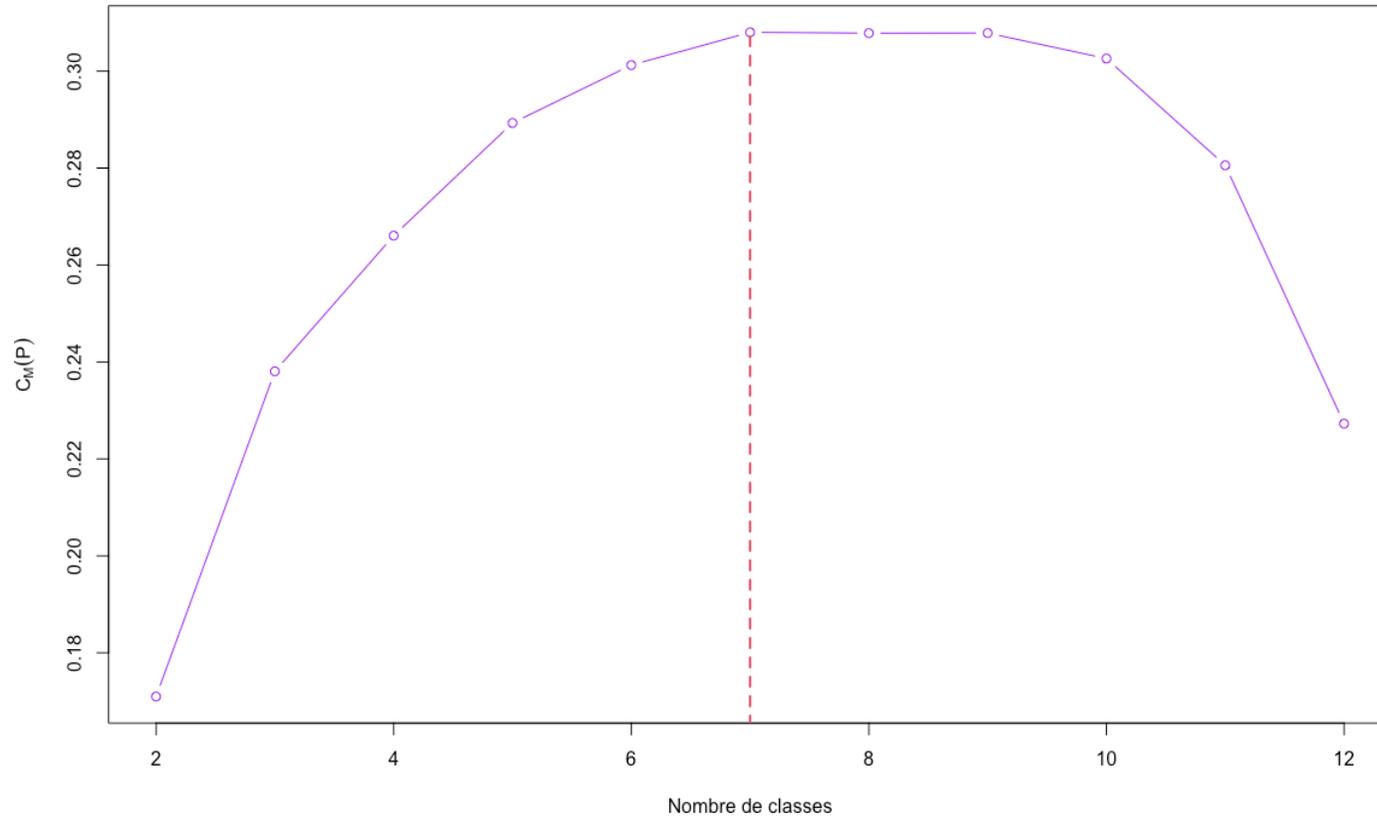
Mesure de consensus

- On cherche une partition P telle que la moyenne des ARI de chaque sujet avec cette partition est maximale c'est à dire

$$C_M(P) = \frac{1}{N} \sum_{n=1}^N ARI(P, P_n) \text{ maximale.}$$

- On procède de façon itérative en fixant le nombre K de classes dans la partition consensuelle. On choisit la valeur de K pour laquelle $C_M(P)$ est maximal.
- **Attention** : Cet algorithme est sensible au choix de la première partition.

Recherche du consensus

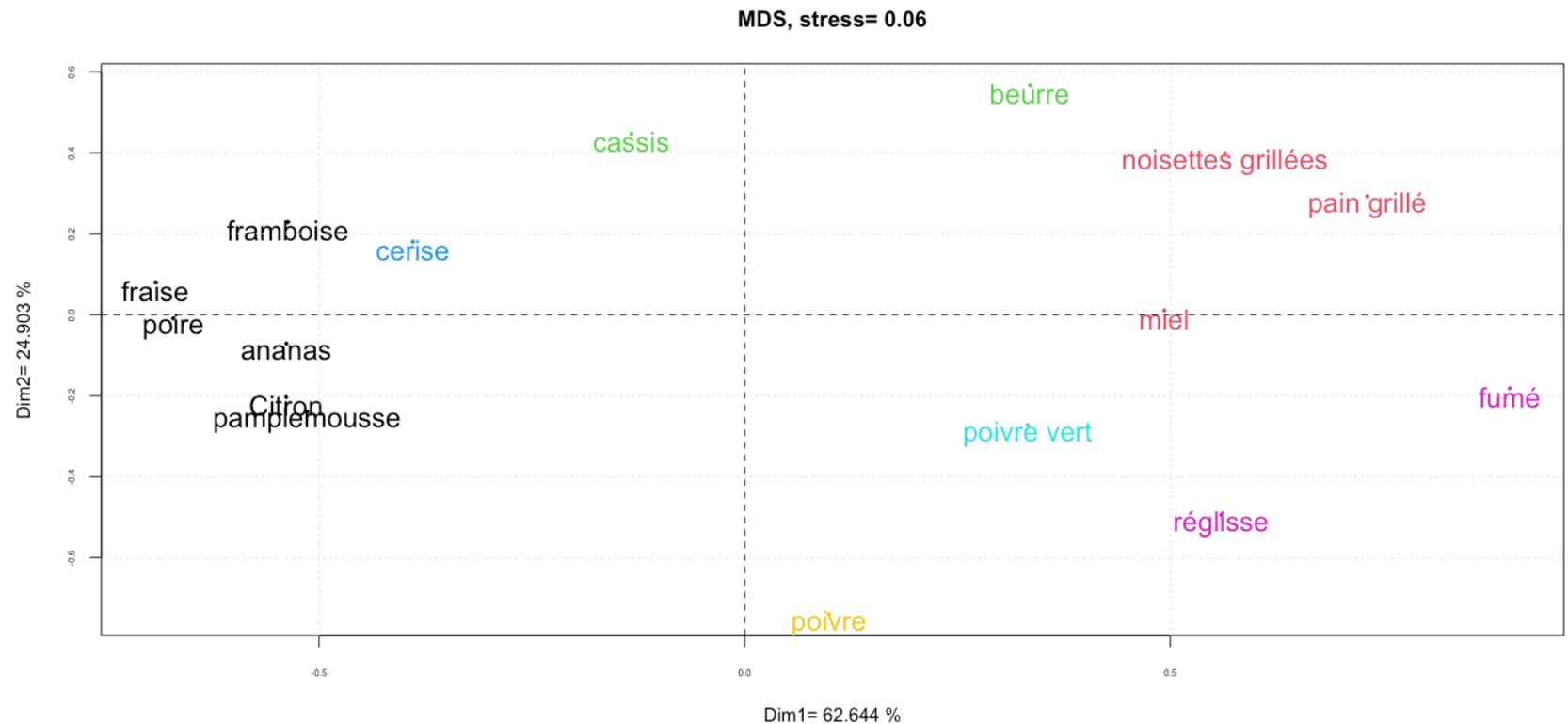


Résultat pour l'exemple

Partition en 7 classes avec
 $C_M(P) = 0.308$

arome	Consensus
Citron	1
pamplemousse	1
ananas	1
poire	1
miel	2
beurre	3
pain grillé	2
noisettes grillées	2
fraise	1
framboise	1
cerise	4
cassis	3
poivre vert	5
fumé	6
poivre	7
réglisse	6

Représentation de la partition consensuelle sur le plan latent :



Classification des sujets

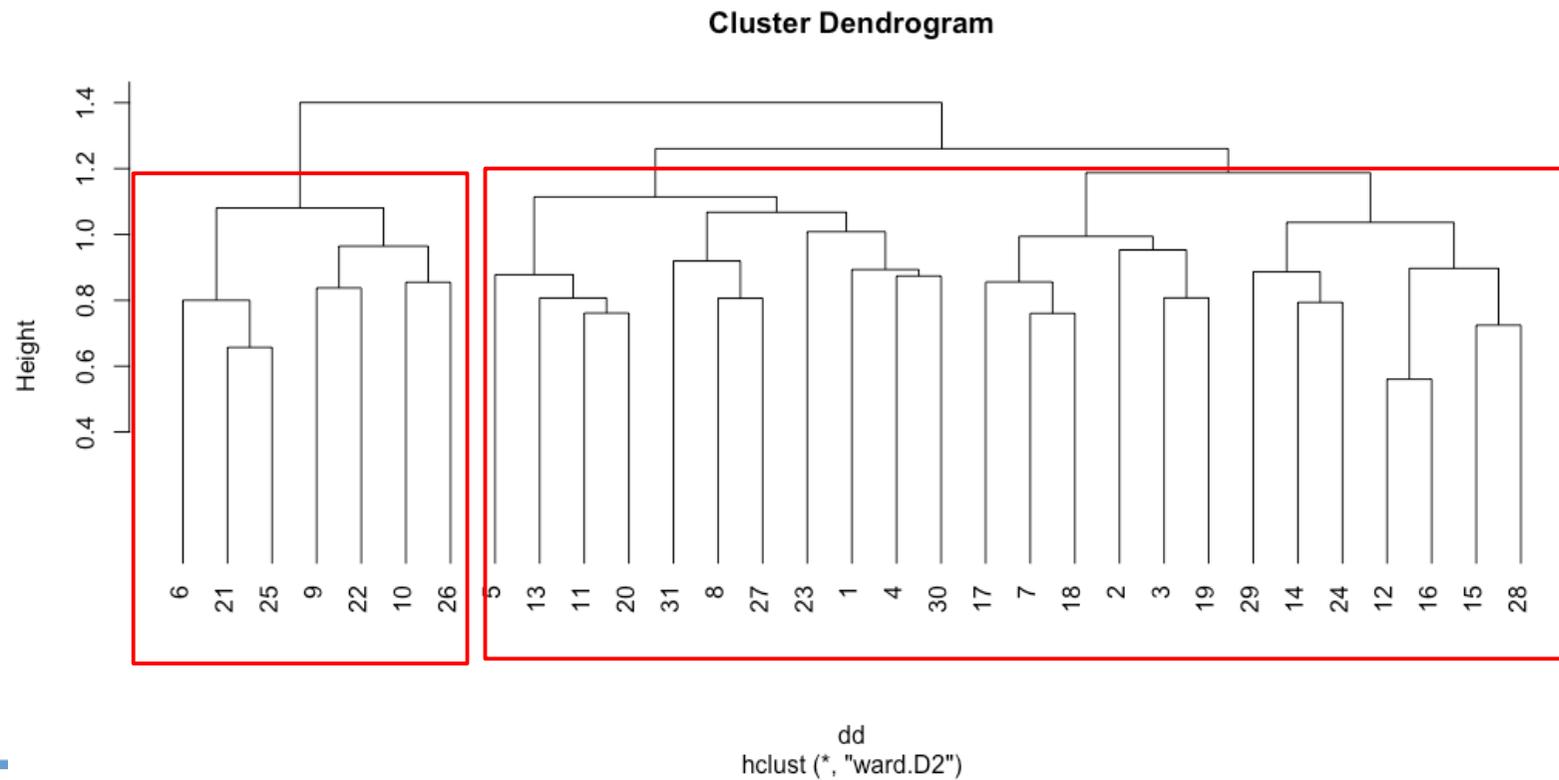
- L'*ARI* permet de définir une distance euclidienne entre les sujets,

$$d_{ARI}(S_1, S_2) = \sqrt{1 - ARI(P_1, P_2)}$$

où le sujet S_i propose la partition P_i .

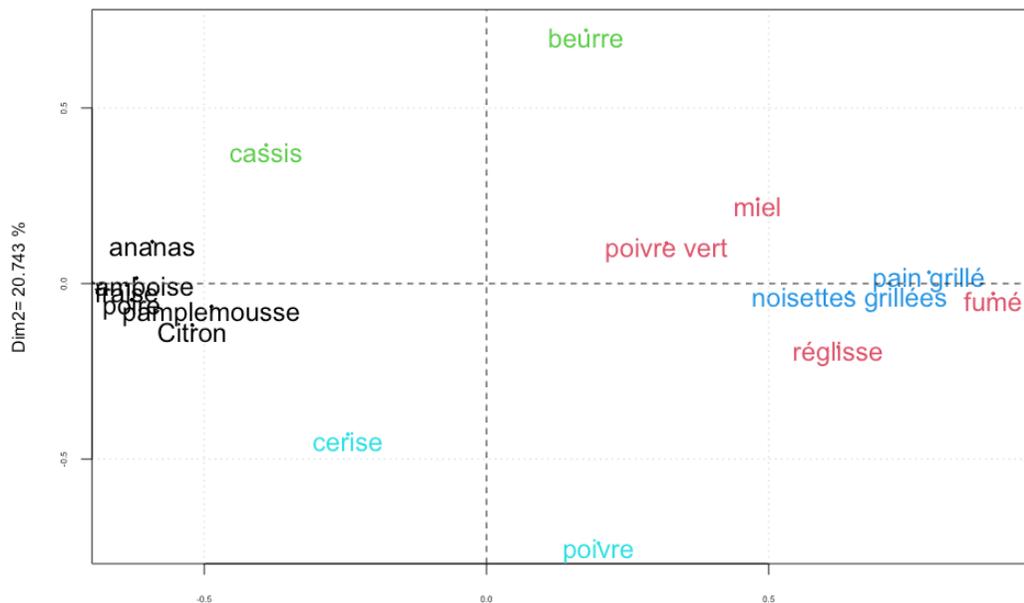
- On réalise ensuite une Classification ascendante hiérarchique avec cette distance. On établit le consensus sur les groupes obtenus.

CAH avec d_ARI



Consensus dans les deux classes

MDS, stress= 0.068

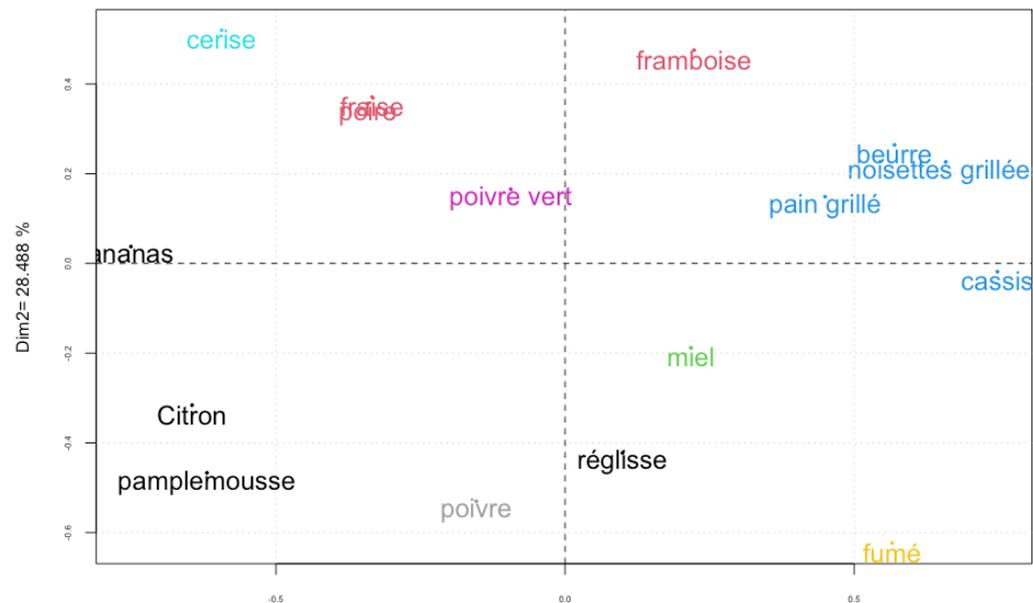


7 sujets

5 groupes

$$C_M(P) = 0.32$$

MDS, stress= 0.032



24 sujets

9 groupes

$$C_M(P) = 0.45$$

Informations additionnelles

Informations additionnelles (1)

On propose aux 31 sujets une liste de 36 descripteurs pour les 16 produits considérés. On obtient la table de contingence suivante :

	Acide	fumé	capiteux	agrumes	citron	fruit	naturel	épicé	sucré	chaud	agréable	bonbon
Citron	3	0	1	7	11	11	0	1	5	2	3	5
pamplemousse	5	0	1	9	9	13	0	1	4	2	2	3
ananas	5	0	1	5	4	13	0	1	9	0	2	8
poire	2	0	1	4	5	15	1	0	10	0	2	13
miel	1	2	1	1	3	1	1	1	4	3	1	1
beurre	2	1	0	0	0	2	0	1	3	1	1	2
pain grillé	0	5	1	0	0	0	0	1	1	2	0	0
noisettes grillées	0	3	1	1	0	1	0	0	2	3	1	0
fraise	3	0	1	3	2	14	0	0	13	0	2	14
framboise	1	1	1	1	3	12	1	0	10	0	2	9
cerise	1	1	1	1	1	4	0	0	5	2	3	5
cassis	1	1	1	1	1	8	1	0	4	3	1	2
poivre vert	3	3	0	1	1	1	1	0	2	2	1	0
fumé	0	19	0	0	0	0	0	1	0	2	0	1
poivre	3	1	0	1	3	1	1	4	1	7	1	1
régliasse	1	5	1	0	1	1	1	2	0	4	0	1

Caractérisation des axes latents

On effectue la régression de chacun des descripteurs sur les composantes de la MDS. Soit Y_h l'un des descripteurs proposés on écrit

$$Y_h = \beta_{0,h} + \beta_{1,h}C_1 + \beta_{2,h}C_2 + \beta_{3,h}C_3 + \varepsilon_h$$

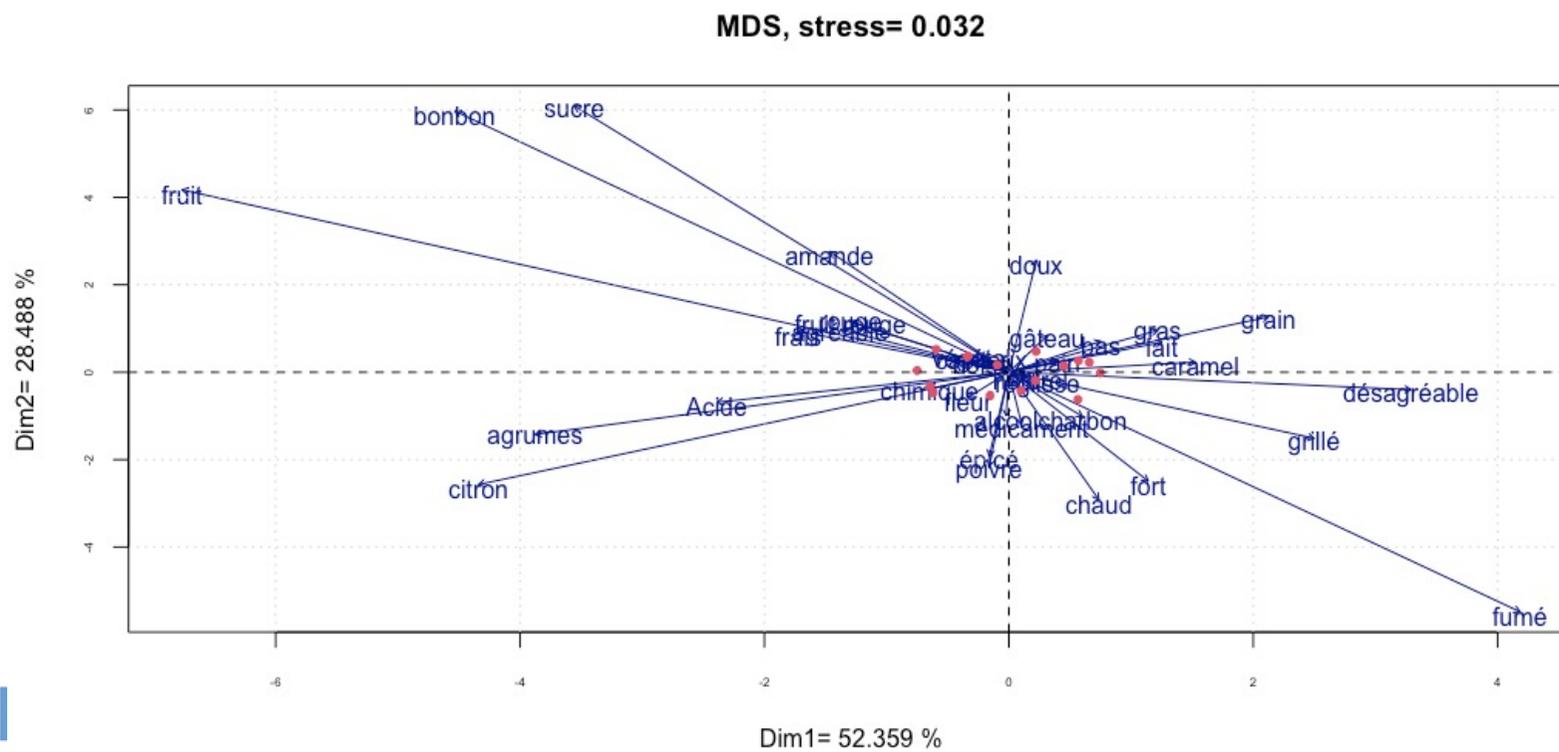
Les coefficients $\beta_{1,h}, \beta_{2,h}, \beta_{3,h}$ seront les coordonnées de Y_h dans les plans factoriels.

Exemple

Descripteur Acide :

Comp1	Comp2	Comp3
-2.39	-0.69	-0.69

Représentation graphique



Informations additionnelles (2): Matrice de Liking

Ces données sont issues d'un cours de sensométrie de François Husson <https://husson.github.io/data.html>

59 sujets ont réalisé un tri libre de 12 jus d'orange.

48 sujets ont réalisé un liking de ces 12 jus d'orange.



Consignes

4) Jugement d'ensemble

Appréciation hédonique (globale)

Très mauvais

1 2 3 4 5 6 7 8 9 10

Excellent

Commentaires :

Les données de Liking

	Juge.1	Juge.2	Juge.3	Juge.45	Juge.46	Juge.47	Juge.48
Jaf_SP_F	5	6	2	1	3	2	7
Jaf	2	3	4	3	3	3	8
Trop_AP_F	9	8	5	10	3	4	3
Bio	3	2	3	1	2	3	1
Jock	3	5	8	1	2	2	5
And_F	7	5	8	1	5	2	5
Trop_SP_F	7	7	5	6	4	4	7
Jaf_AP_F	3	4	4	7	3	2	3
Trop_SP	1	8	3	1	2	3	6
Trop_AP	3	7	4	2	3	1	4
Trop_pulipissimo	7	2	6	4	3	1	3
Jaf_israel	4	2	3	1	7	4	2

Projection des notes sur le plan factoriel

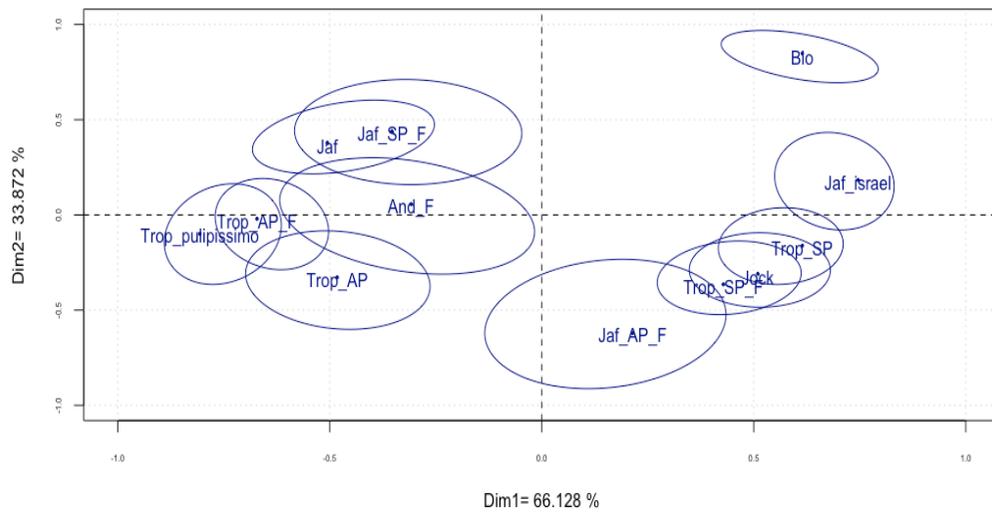
Plan factoriel à partir de la matrice de dissimilarités obtenue à l'issue du tri libre.

Soit $Juge_h$ les notes attribuées par le juge h pour les 12 jus d'oranges proposés

$$Juge_h = \beta_{0,h} + \beta_{1,h}C_1 + \beta_{2,h}C_2 + \varepsilon_h$$

Les coefficients de régression seront les coordonnées des notes dans le plan factoriel.

MDS, stress= 0.086



MDS, stress= 0.086

