Euclidean representations of a set of hierarchies using Multiple Factor Analysis

Cadoret M.*, Lê S.* and Pagès J.*

* Applied mathematics department
Agrocampus Ouest, France

9 February 2011

Correspondence Analysis and Related Methods 2011

Outline

Introduction

Interested in:

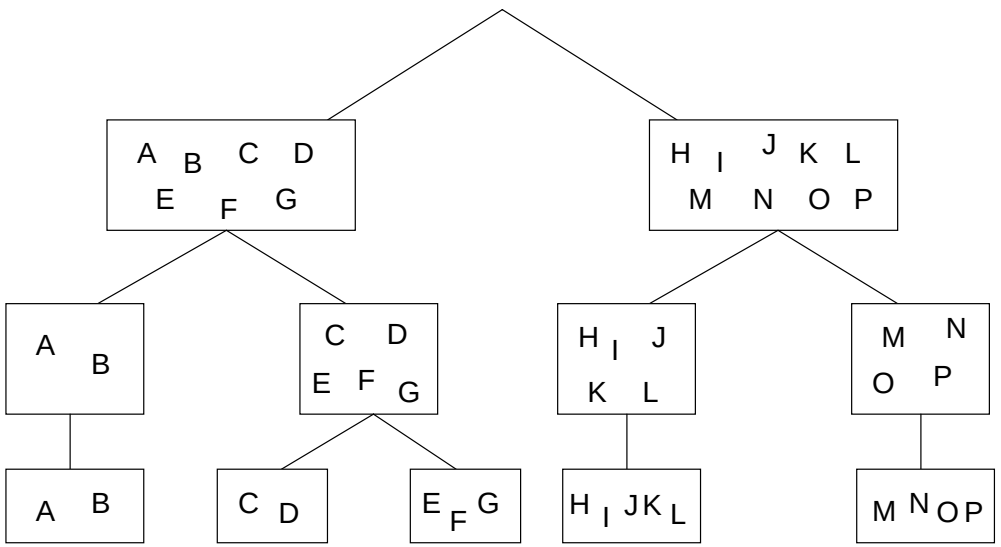
- Set of non-indexed hierarchies
- Synthetic graphical representations

At least 2 possible graphical representations:

- As a hierarchy consensus (Adams, 1972)
  - Same shape of the data
  - Consensus difficult to obtain when the number of hierarchies increases
- As an Euclidean representation of the hierarchies: representation of the terminal nodes, etc.

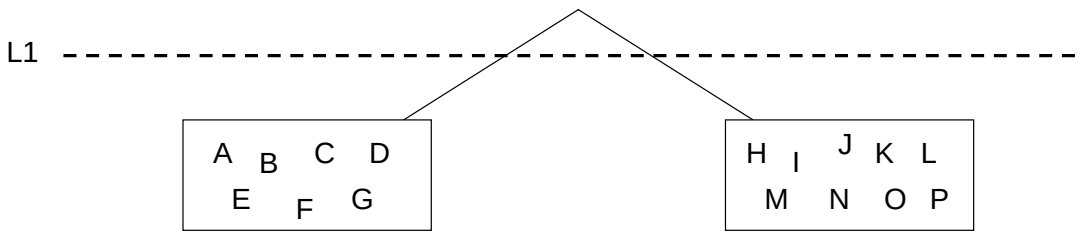
Outline

Data coding (1)



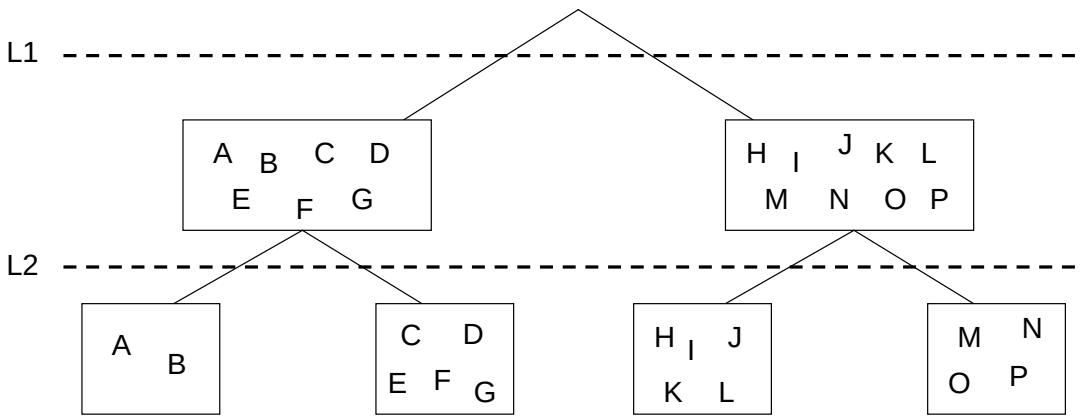
	L1	L2	L3
A			
B			
C			
D			
E			
F			
G			
H			
I			
J			
K			
L			
M			
N			
O			
P			

Data coding (1)



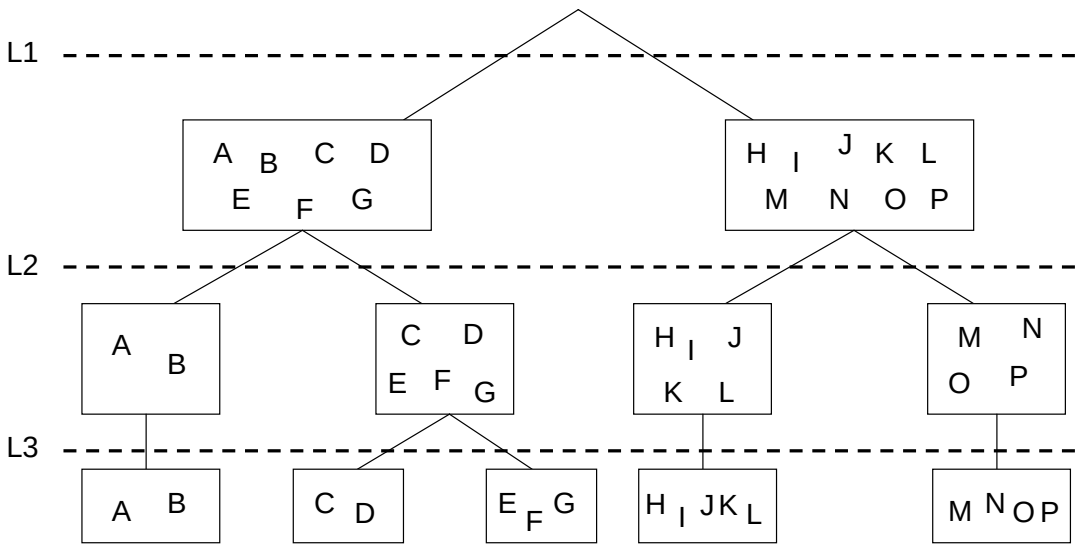
	L1	L2	L3
A	G1		
B	G1		
C	G1		
D	G1		
E	G1		
F	G1		
G	G1		
H	G2		
I	G2		
J	G2		
K	G2		
L	G2		
M	G2		
N	G2		
O	G2		
P	G2		

Data coding (1)



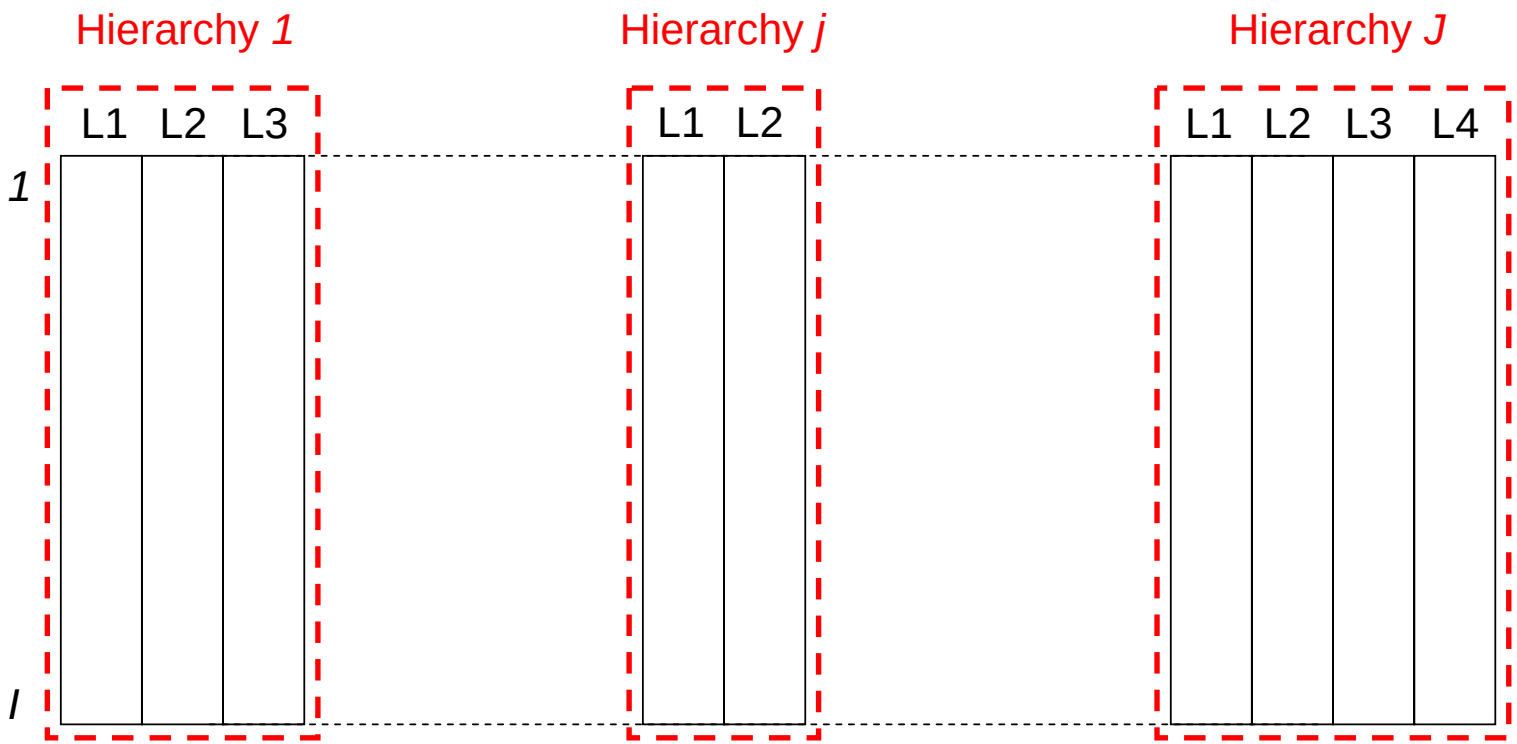
	L1	L2	L3
A	G1	G1	
B	G1	G1	
C	G1	G2	
D	G1	G2	
E	G1	G2	
F	G1	G2	
G	G1	G2	
H	G2	G3	
I	G2	G3	
J	G2	G3	
K	G2	G3	
L	G2	G3	
M	G2	G4	
N	G2	G4	
O	G2	G4	
P	G2	G4	

Data coding (1)



	L1	L2	L3
A	G1	G1	G1
B	G1	G1	G1
C	G1	G2	G2
D	G1	G2	G2
E	G1	G2	G3
F	G1	G2	G3
G	G1	G2	G3
H	G2	G3	G4
I	G2	G3	G4
J	G2	G3	G4
K	G2	G3	G4
L	G2	G3	G4
M	G2	G4	G5
N	G2	G4	G5
O	G2	G4	G5
P	G2	G4	G5

Data coding (2)



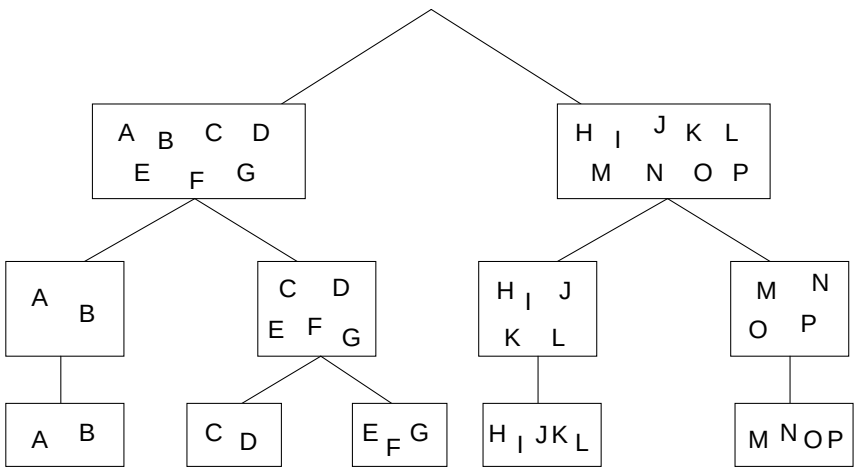
Outline

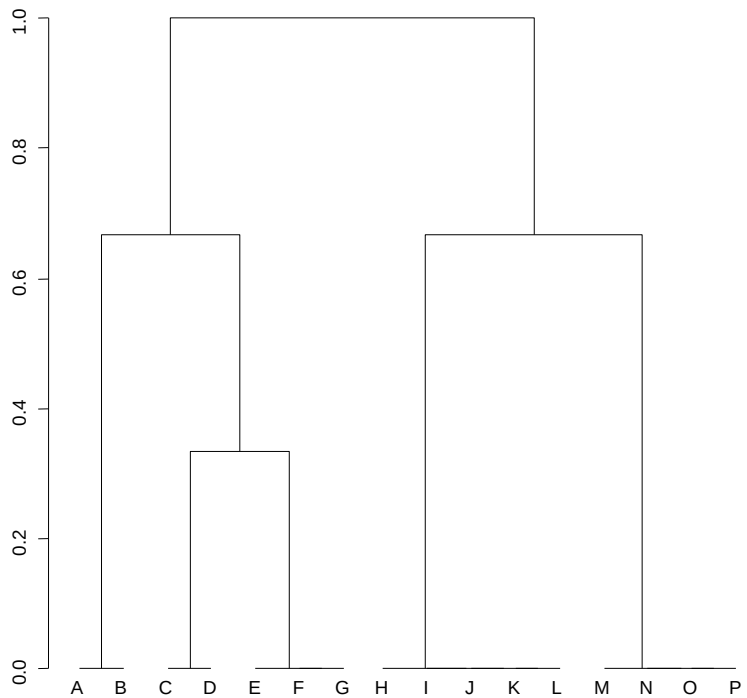
Check on data coding and analysis of 1 hierarchy

- Data table with qualitative variables
- Multiple Correspondence Analysis + Ascendant Hierarchical Classification on the dimensions

Check on data coding and analysis of 1 hierarchy

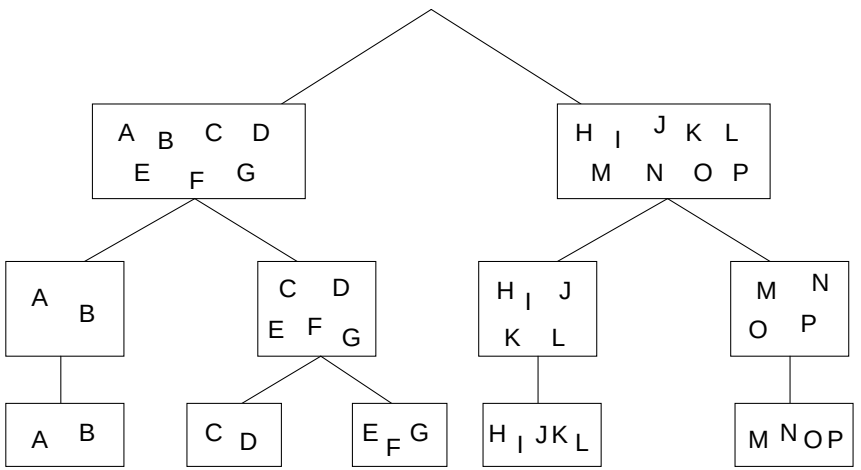
- Data table with qualitative variables
- Multiple Correspondence Analysis + Ascendant Hierarchical Classification on the dimensions

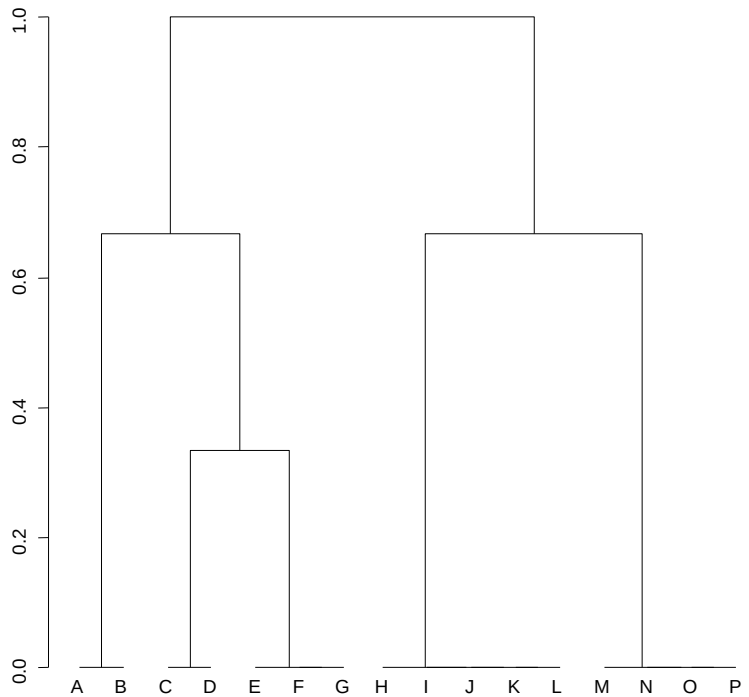




Check on data coding and analysis of 1 hierarchy

- Data table with qualitative variables
- Multiple Correspondence Analysis + Ascendant Hierarchical Classification on the dimensions





$\Rightarrow$ We found the initial hierarchy

Objectives

From a data table with a group structure on the variables, we want to perform a global factorial analysis such as:

Objectives

From a data table with a group structure on the variables, we want to perform a global factorial analysis such as:

- it provides graphical representations of objects, hierarchies and levels of hierarchy
- the influence of each hierarchy is balanced

Objectives

From a data table with a group structure on the variables, we want to perform a global factorial analysis such as:

- it provides graphical representations of objects, hierarchies and levels of hierarchy
- the influence of each hierarchy is balanced

$\Rightarrow$ Multiple Factor Analysis (MFA; Escofier and Pagès (1982)) in which 1 hierarchy corresponds to 1 group of variables

Multiple Correspondence Analysis (MCA)

MCA is looking for dimensions z_s that maximize:

$$\frac{1}{Q}\sum_{q}^{Q}\eta^2(z_s, L_q),$$

with:

- Q the number of qualitative variables
- z_s the axis s
- L_q the qualitative variable q

Multiple Correspondence Analysis (MCA)

MCA is looking for dimensions z_s that maximize:

$$\frac{1}{Q}\sum_{q}^{Q}\eta^2(z_s, L_q),$$

with:

- Q the number of qualitative variables
- z_s the axis s
- L_q the qualitative variable q

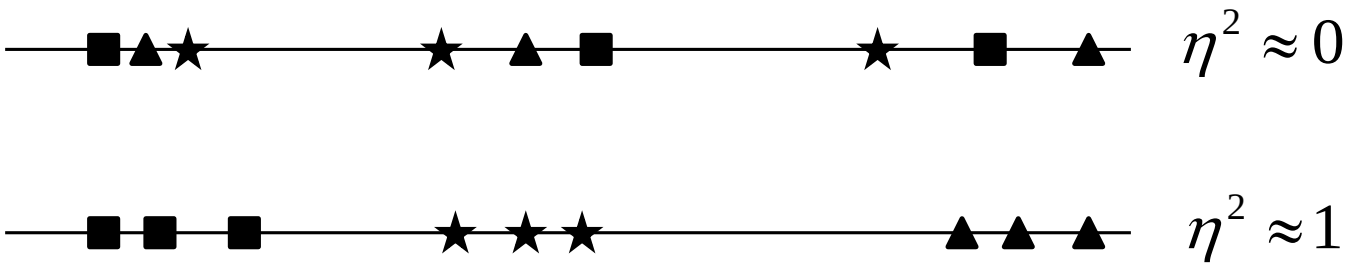
Multiple Correspondence Analysis (MCA)

MCA is looking for dimensions z_s that maximize:

$$\frac{1}{Q}\sum_{q}^{Q}\eta^2(z_s, L_q),$$

with:

- Q the number of qualitative variables
- z_s the axis s
- L_q the qualitative variable q



Multiple Factor Analysis (MFA)

MFA is looking for dimensions z_s that maximize the following criterion:

$$\sum_{j}^{J} \frac{1}{Q_j} \sum_{q}^{Q_j} \eta^2(z_s, L_q),$$

with:

- Q_j the number of level of hierarchy j
- z_s the axis s
- L_q the level q of the hierarchy j

Multiple Factor Analysis (MFA)

MFA is looking for dimensions z_s that maximize the following criterion:

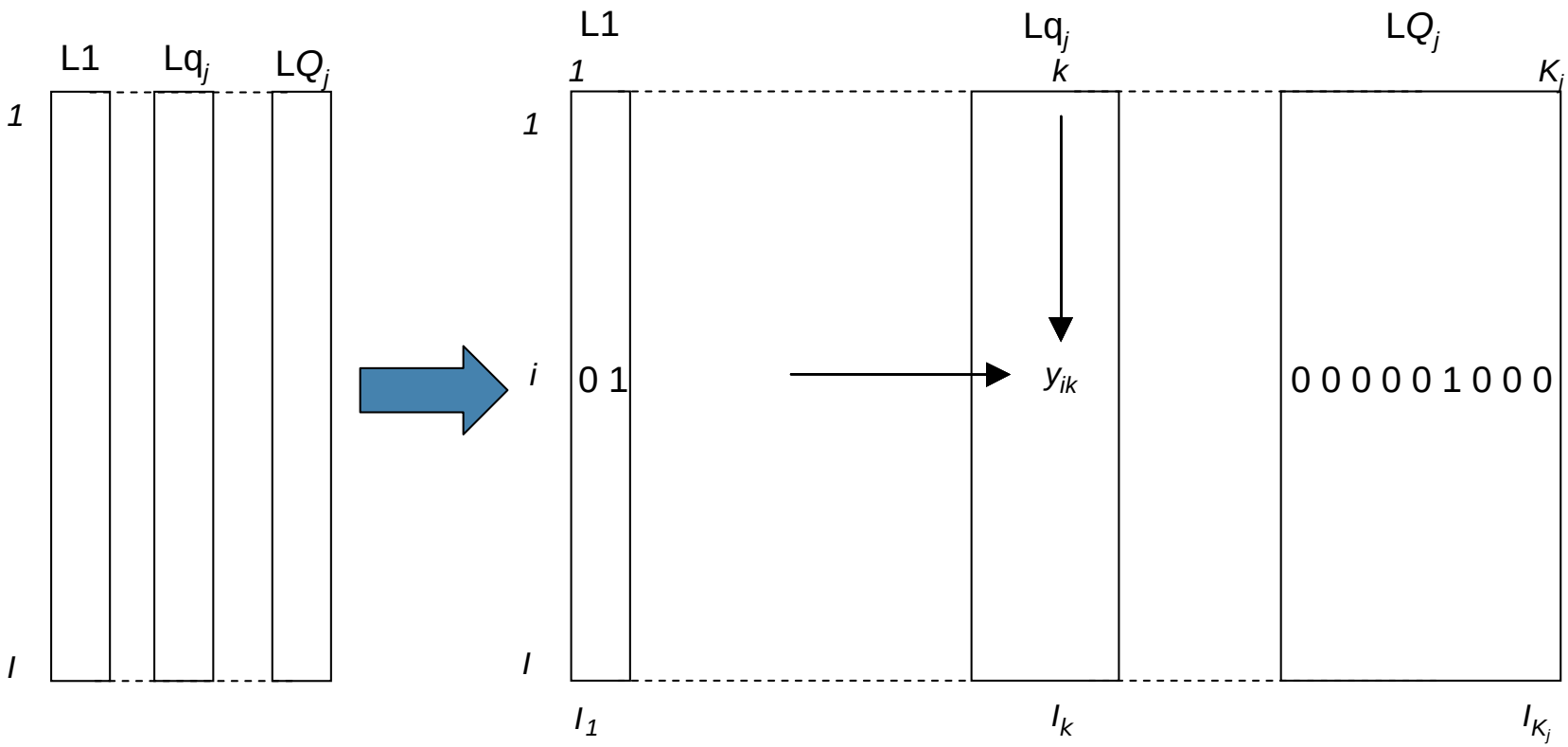
$$\sum_{j}^{J} \frac{1}{Q_j} \sum_{q}^{Q_j} \eta^2(z_s, L_q),$$

with:

- Q_j the number of level of hierarchy j
- z_s the axis s
- L_q the level q of the hierarchy j

$\Rightarrow$ In this particular case: criterion maximized by MFA $\Leftrightarrow$ sum of criteria maximized by MCA

Disjunctive data table associated with one hierarchy j



Each level (associated with a hierarchy) is represented by a set of dummy variables

Object representation

Distance between 2 objects:

$$d^2(i,l) = \sum_j \frac{1}{Q_j} \sum_{k \in K_j} \frac{I}{I_k}(y_{ik} - y_{lk})^2 = \sum_j d^2_{MCA_j}(i,l),$$

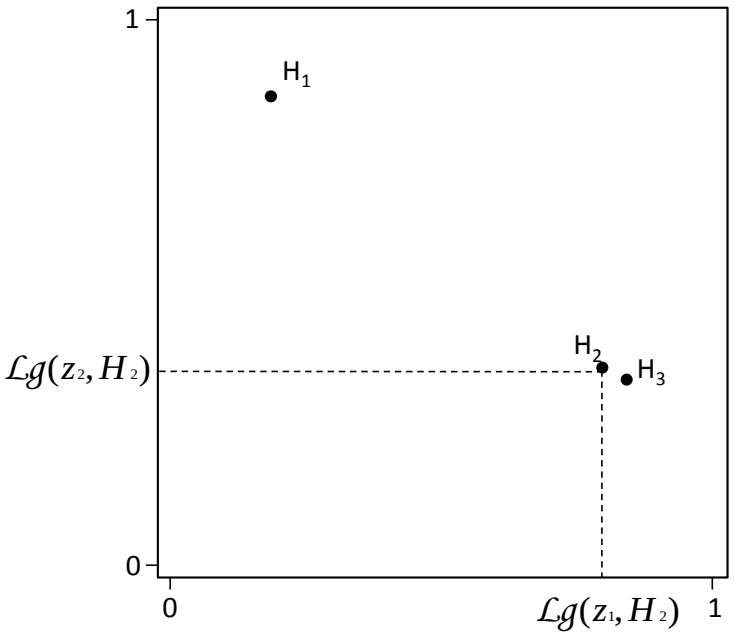
with:

- Q_j the number of level of hierarchy j
- I the number of objects
- I_k the number of objects into the group k
- y_{ik} the element of the disjunctive data table which is equal to 1 if the object i belong to group k and 0 in the opposite case

In this particular case: sum of usual distance in MCA

$\Rightarrow$ 2 objects will be closer than they belong to the same group for a lot of hierarchies

Global hierarchy representation



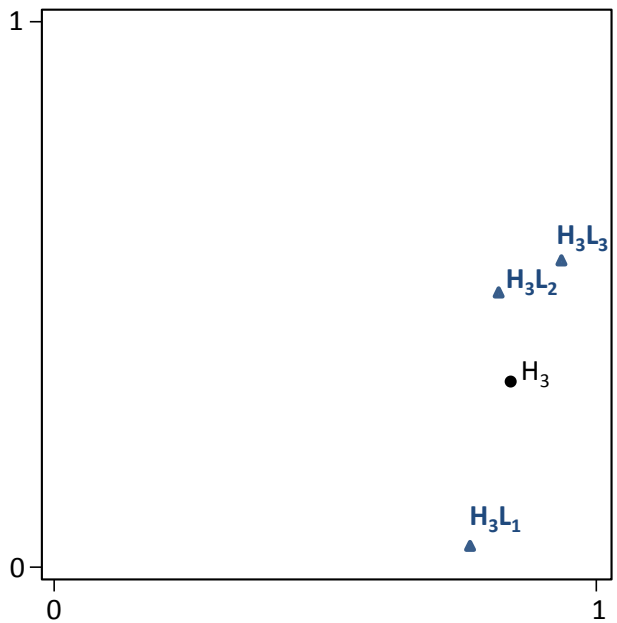
Coordinate of hierarchy j on axis s:

$$\frac{1}{Q_j} \sum_{q \in Q_j} \eta^2(z_s, L_q),$$

with:

- Q_j the number of level of hierarchy j
- z_s the axis s
- L_q the level q of the hierarchy j

Level representation



Coordinate of level q on axis s:

$$\eta^2(z_s, L_q),$$

with:

- z_s the axis s
- L_q the level q

2 consequences:

- Levels ordered along each axis
- Hierarchy = barycenter of its levels

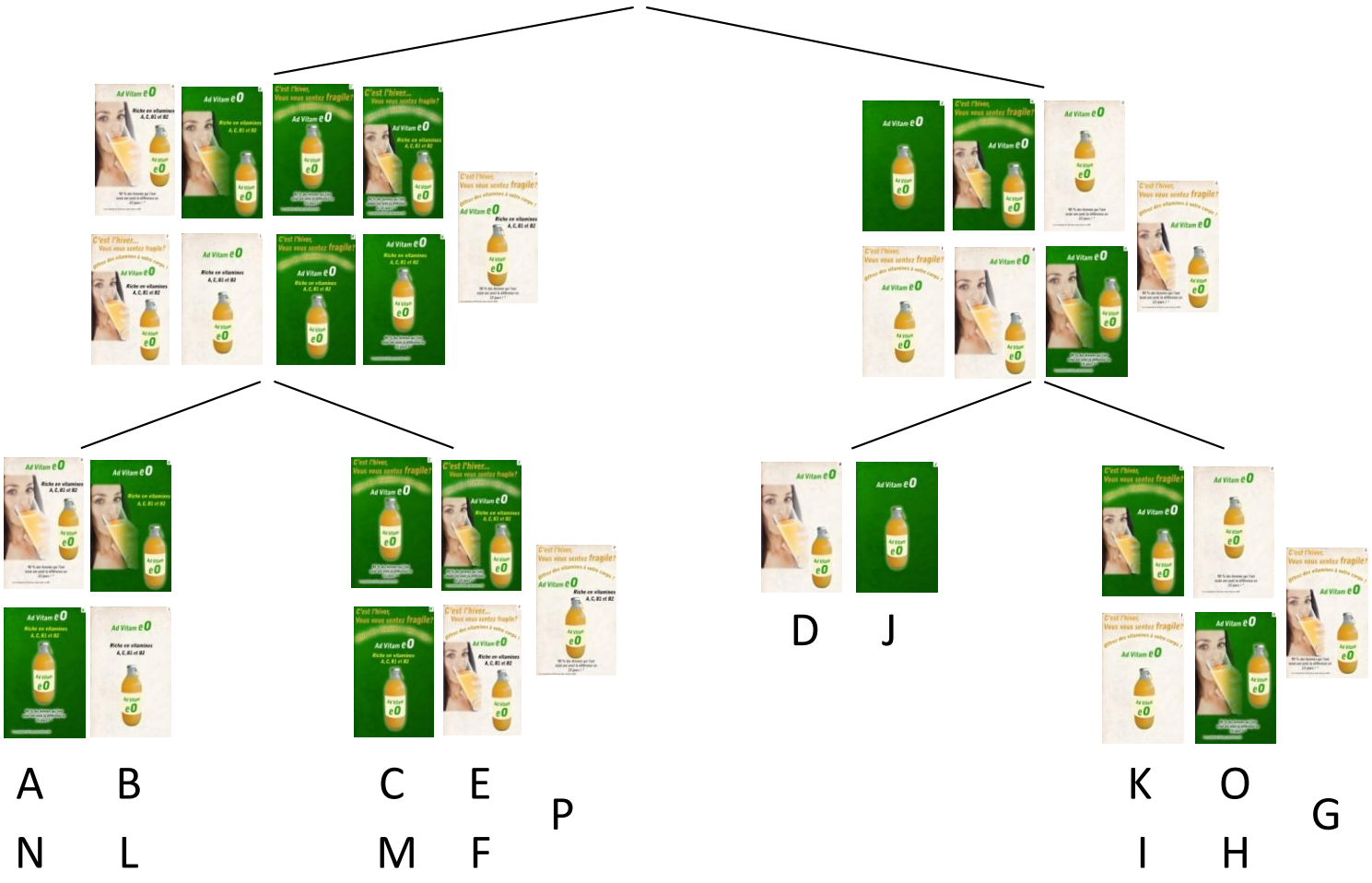
Outline

Data

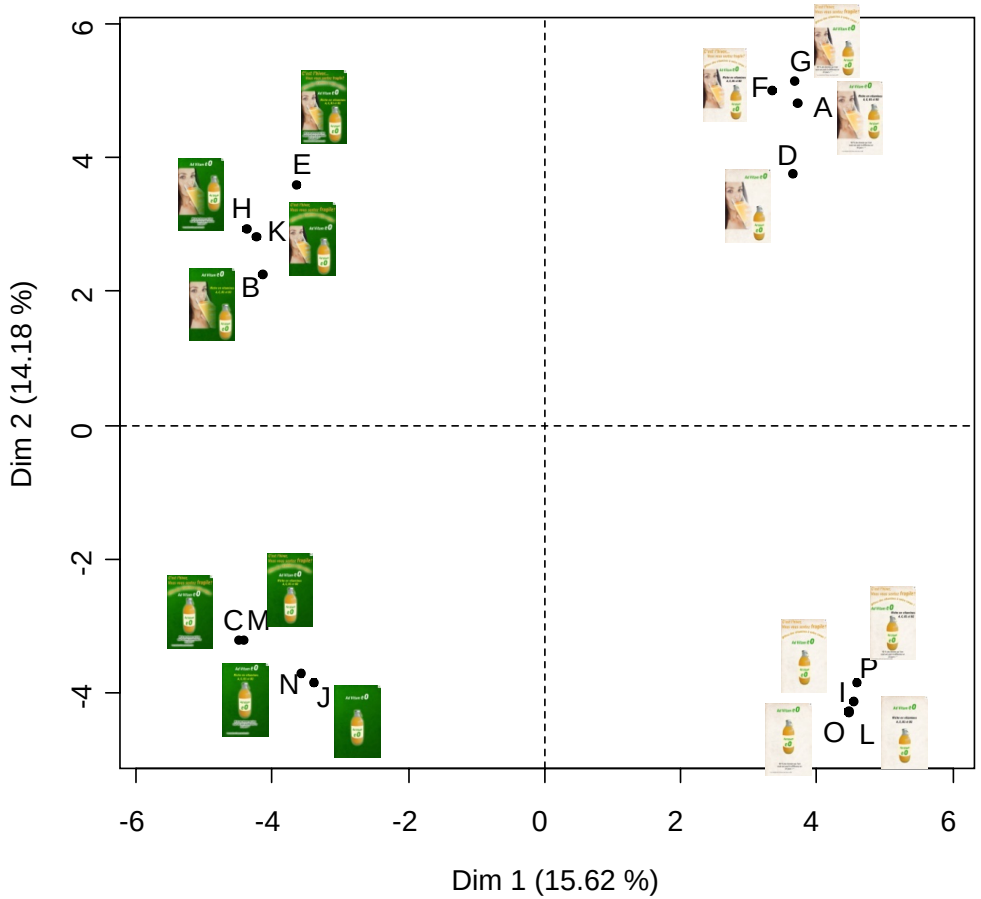
- 16 advertisements concerning an orange juice
- Advertisements built according to a 2^{5-1} fractional factorial design
- 22 subjects
- Hierarchical sorting

Example of hierarchical sorting: subject number 3

Example of hierarchical sorting: subject number 5

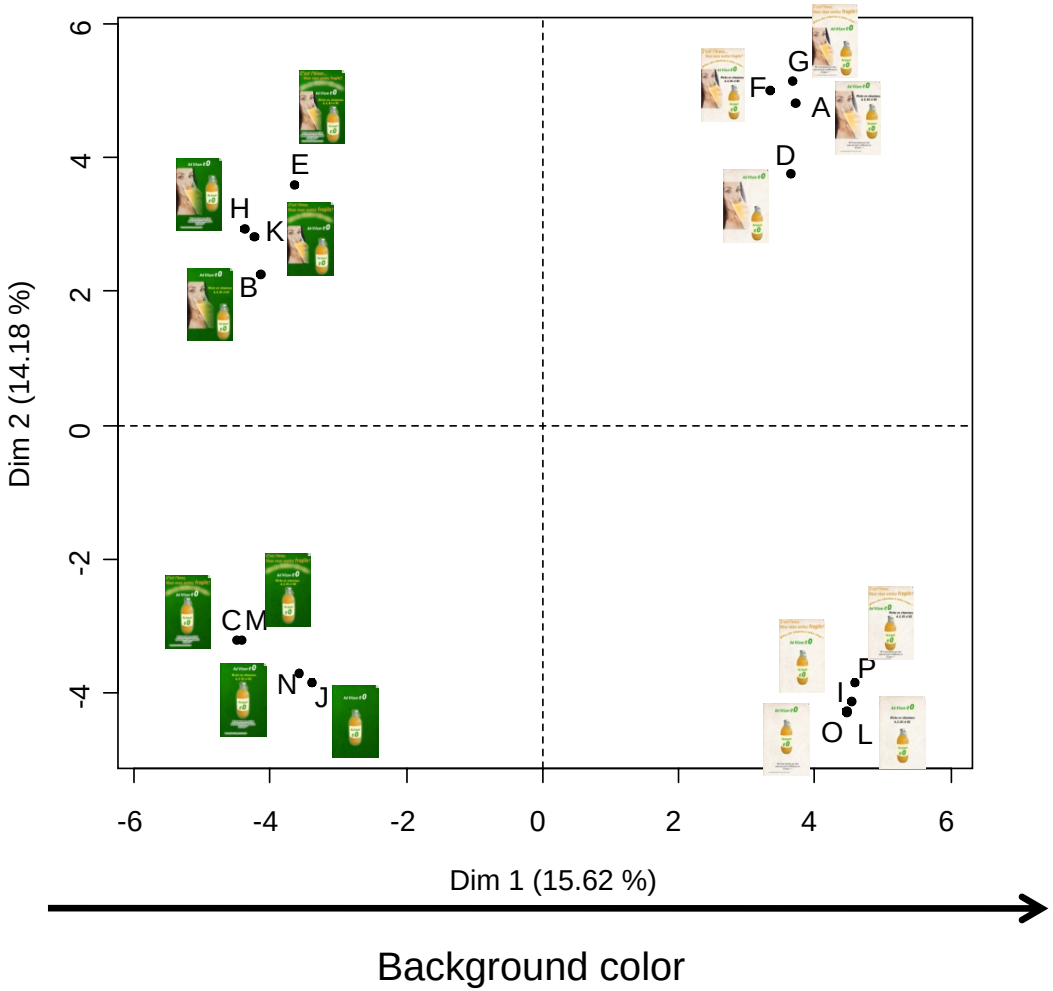


Advertisement representation



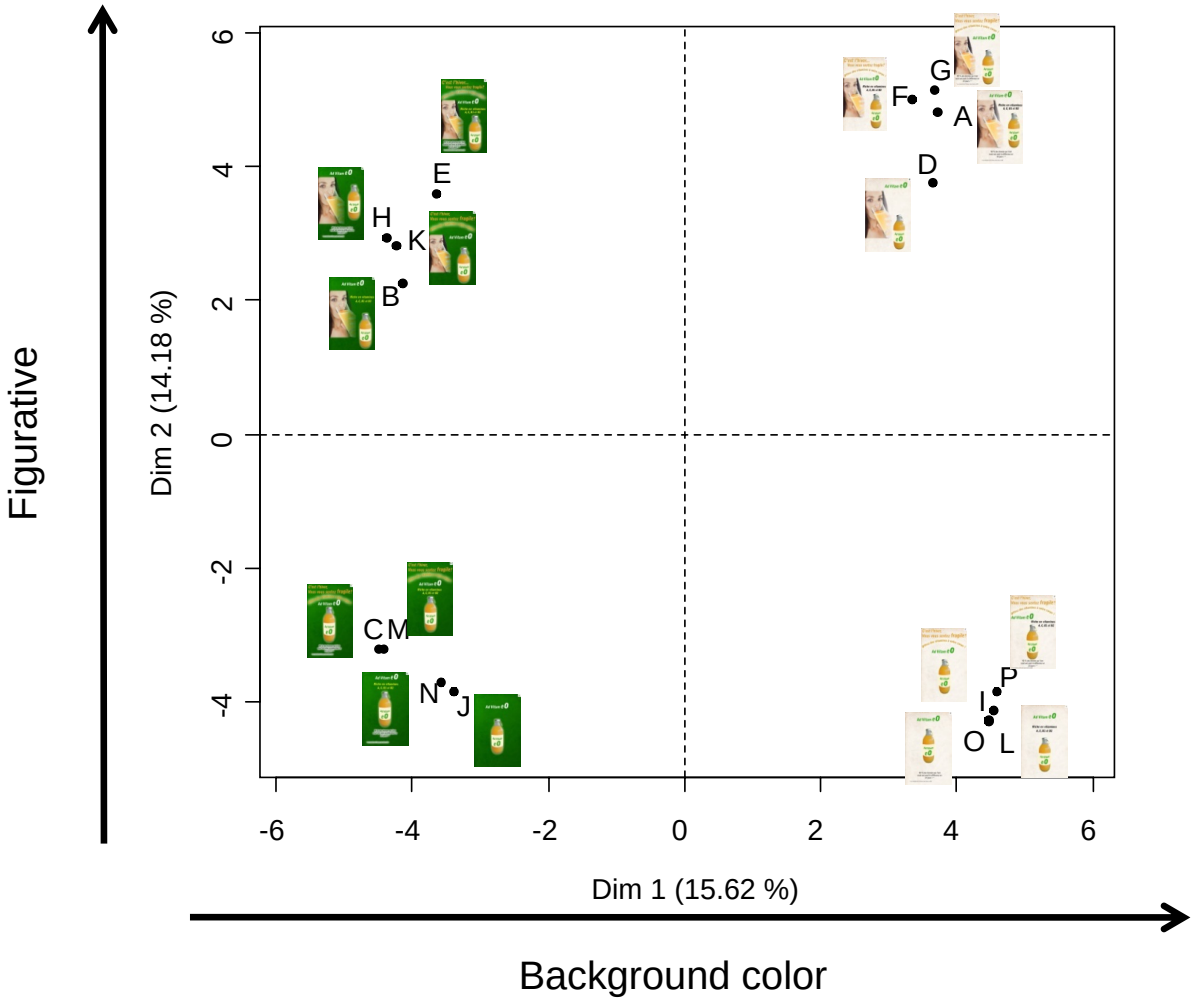
$$\lambda_1 = 16.55$$

Advertisement representation



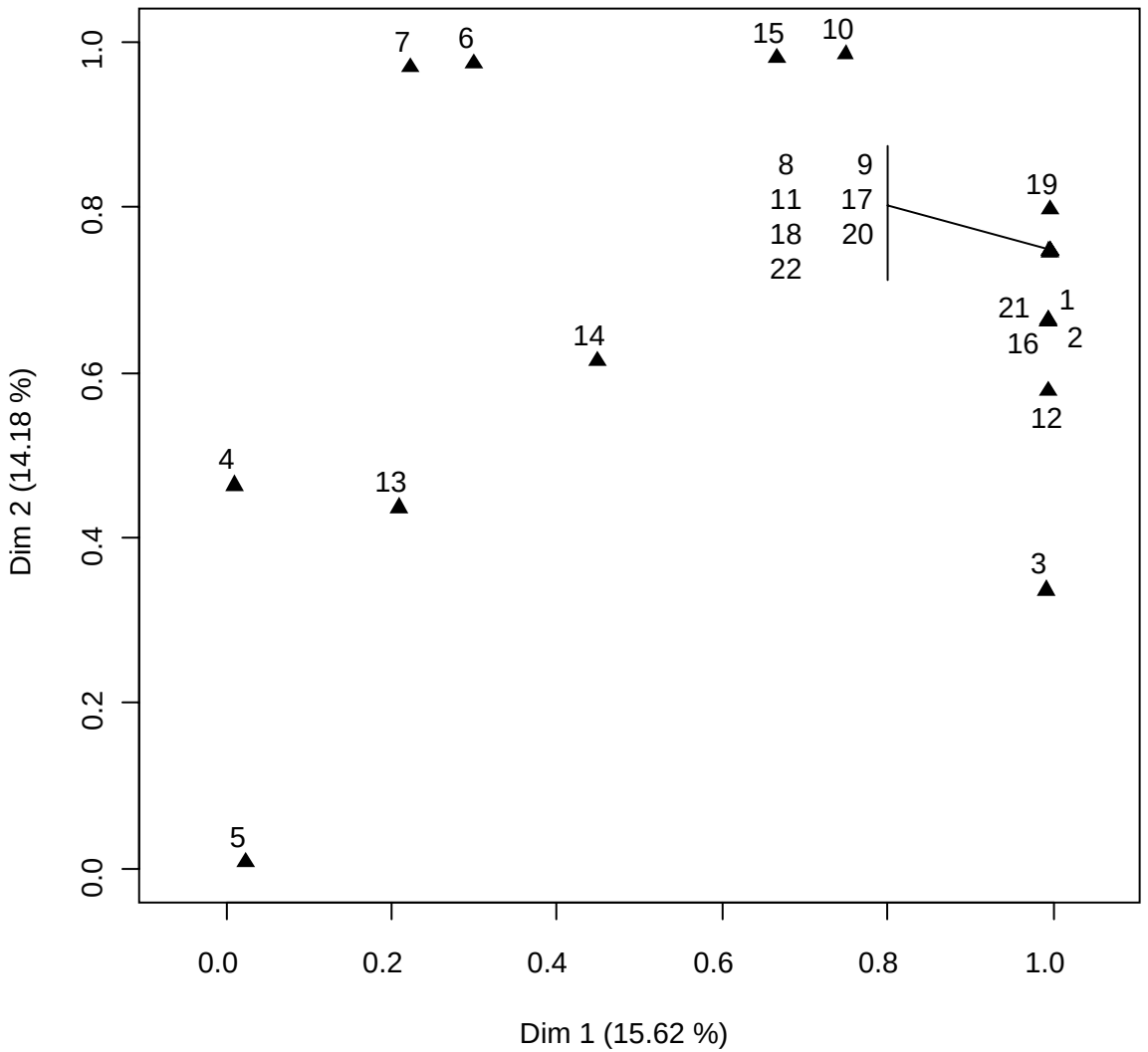
$$\lambda_1 = 16.55$$

Advertisement representation

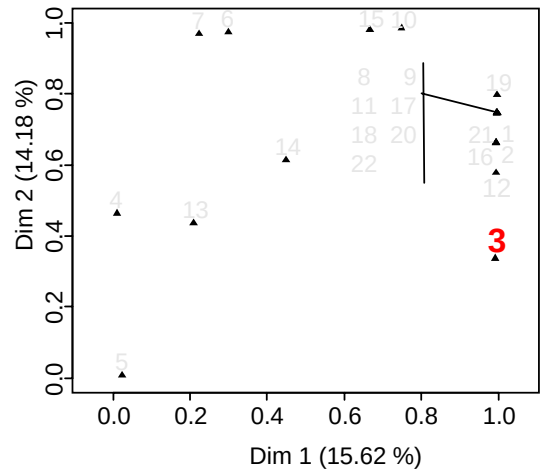


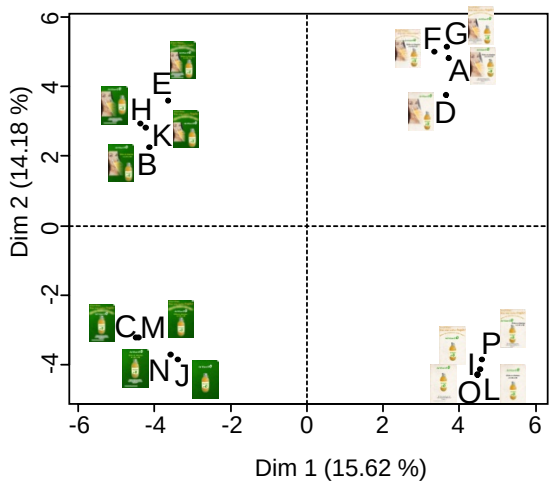
$$\lambda_1 = 16.55$$

Hierarchy representation

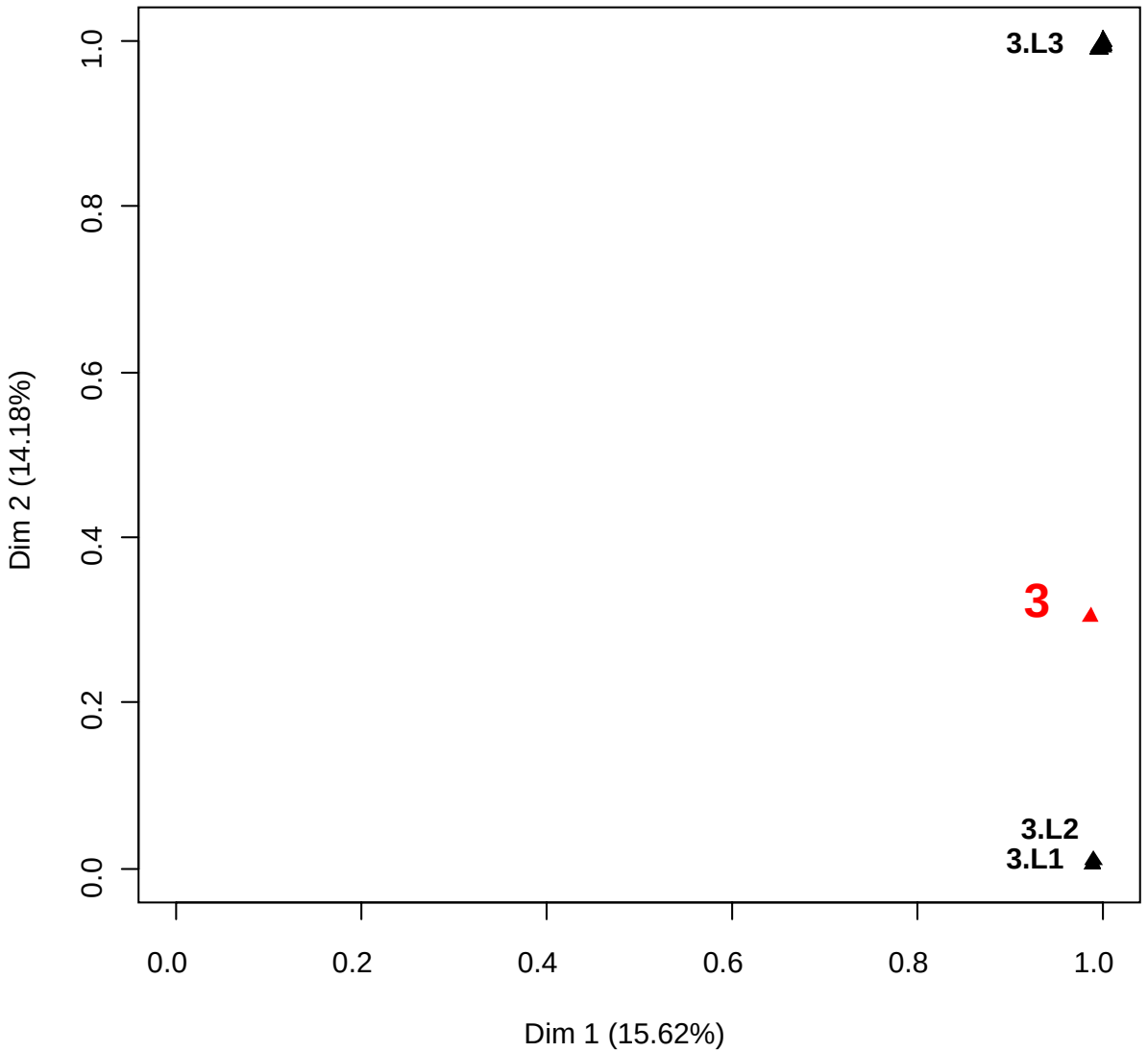


Hierarchy representation: subject number 3

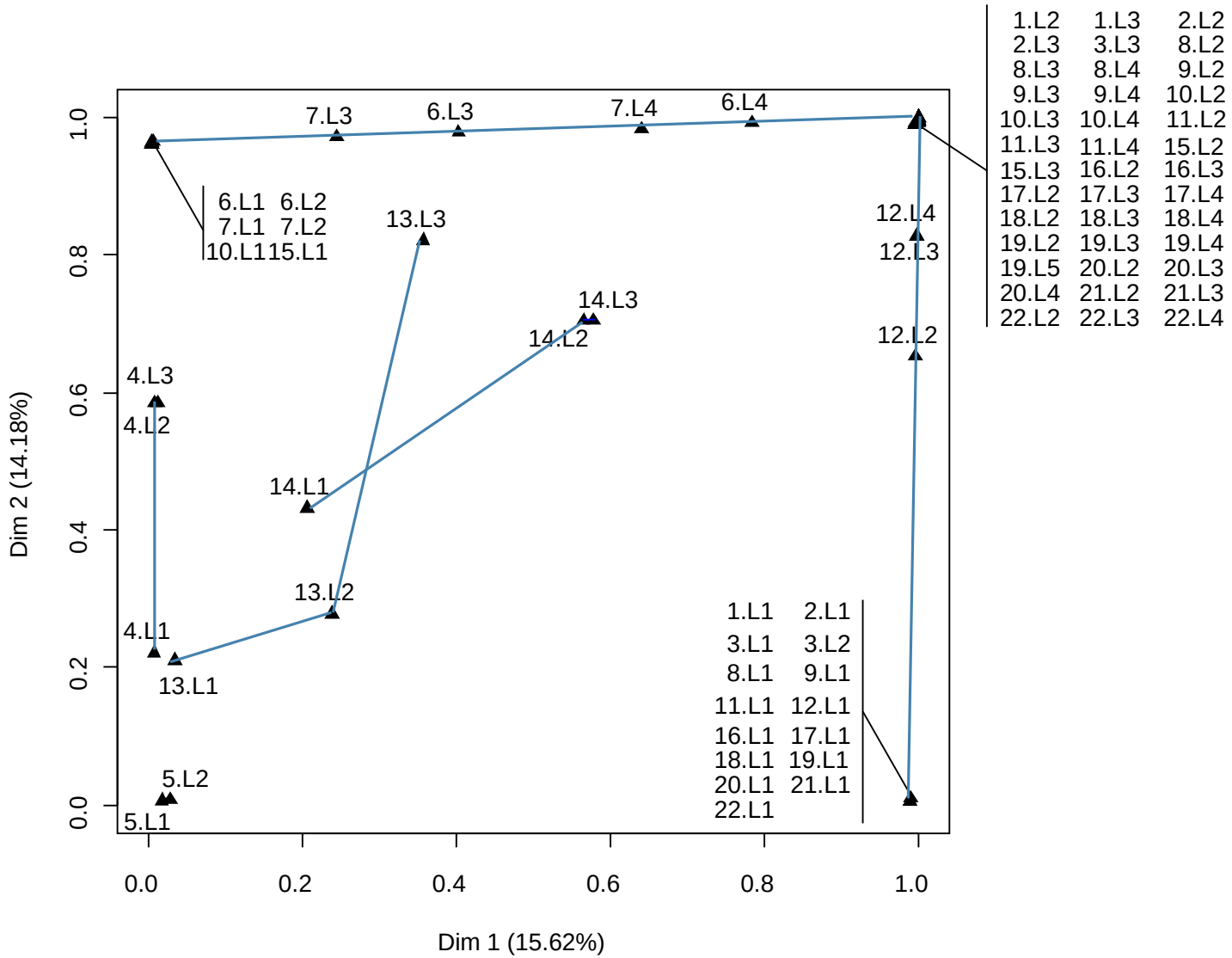




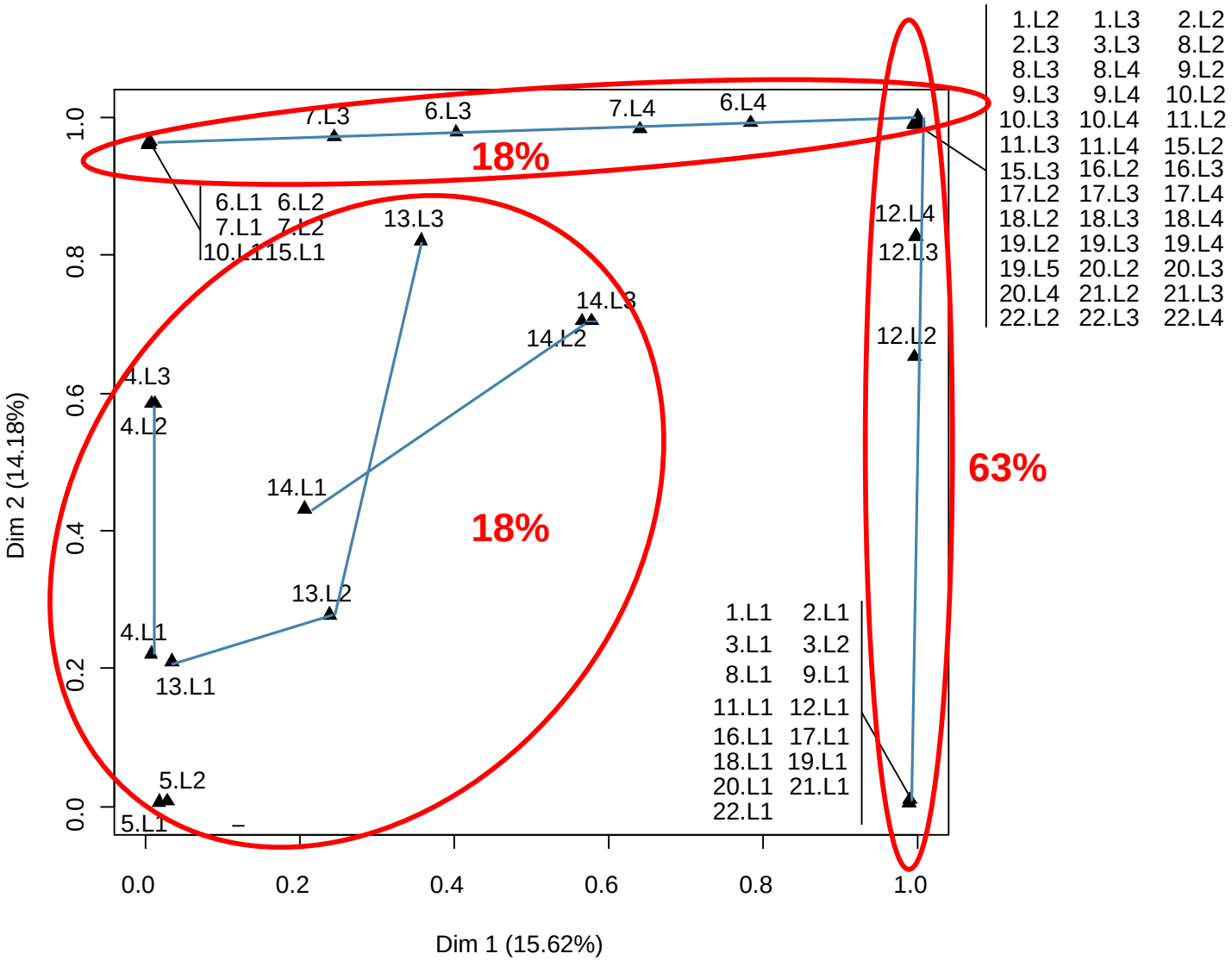
Level representation: subject number 3



Level representation: trajectories



Level representation: trajectories



Outline

Conclusion

- Methodology providing:
  - Representation of objects, hierarchies, levels
  - Representations related to each other
  - Representations interpretable according to simple rules
- In the example, suggests groups of hierarchies
- Allows the simultaneous taking into account of hierarchies and partitions in a same analysis
- Program available in the SensoMineR package

References

Adams, E. I. (1972). Consensus techniques and the comparison of taxonomic trees. *Systematic Zoology*, 21:390–397.

Escofier, B. and Pagès, J. (1982). Comparaison de groupes de variables définies sur le même ensemble d'individus. *Rapport de recherche INRIA*, 149.