

# A new approach for analyzing a set of hierarchies

Cadoret M.\*, Lê S.\* and Pagès J.\*

\* Applied mathematics department  
Agrocampus Ouest, France

September 9, 2010

GfKI - CLADAG 2010



# Outline

- 1 Objectives
- 2 Data collection
- 3 Statistical analysis
- 4 Application
- 5 Conclusion

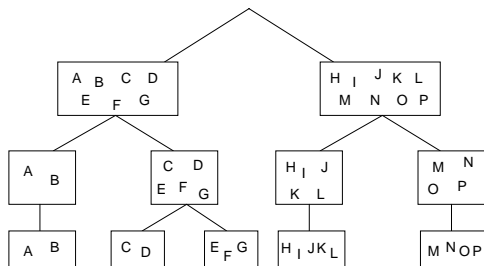
# Introduction

- Interested in graphical representation of several (non-indexed) hierarchies
- At least 2 possibilities:
  - As a hierarchy consensus (Adams, 1972)
  - As an Euclidean representation of the hierarchies (Critchley and Heiser, 1988): representation of the terminal nodes, etc.
- Interested in the representation of the diversity of hierarchies  
⇒ Euclidean representation

# Outline

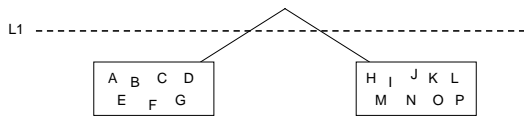
- 1 Objectives
- 2 Data collection
- 3 Statistical analysis
- 4 Application
- 5 Conclusion

# Data collection (1)



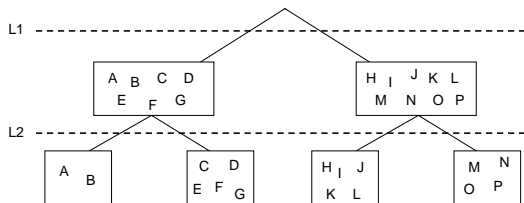
	L1	L2	L3
A			
B			
C			
D			
E			
F			
G			
H			
I			
J			
K			
L			
M			
N			
O			
P			

# Data collection (1)



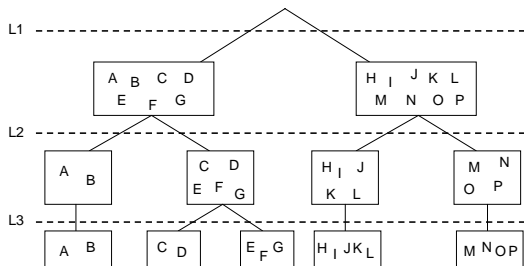
	L1	L2	L3
A	G1		
B	G1		
C	G1		
D	G1		
E	G1		
F	G1		
G	G1		
H	G2		
I	G2		
J	G2		
K	G2		
L	G2		
M	G2		
N	G2		
O	G2		
P	G2		

# Data collection (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	

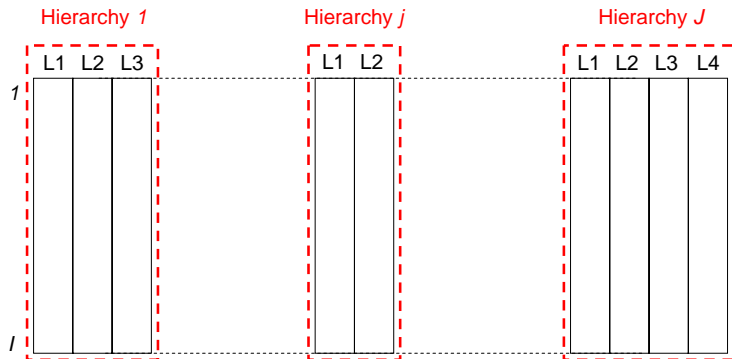
# Data collection (1)



	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 collection (2)



# Outline

- 1 Objectives
- 2 Data collection
- 3 Statistical analysis**
- 4 Application
- 5 Conclusion

# Objectives

What we have:

- A data table with a group structure on the variables

# Objectives

What we have:

- A data table with a group structure on the variables

What we want:

- Obtain graphical representations of objects, hierarchies, etc.
- Balance the influence of each hierarchy
- Keep the information inside each hierarchy

# Objectives

What we have:

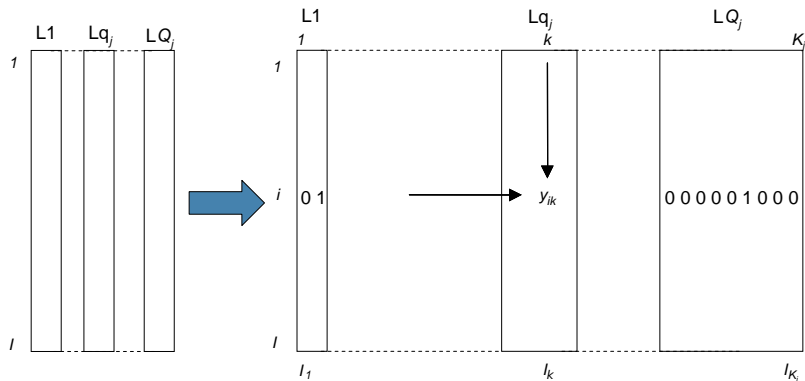
- A data table with a group structure on the variables

What we want:

- Obtain graphical representations of objects, hierarchies, etc.
- Balance the influence of each hierarchy
- Keep the information inside each hierarchy

⇒ Multiple Factor Analysis (MFA; Escofier and Pagès, 1998) in which 1 hierarchy corresponds to 1 group of variables

# Disjunctive data table associated with hierarchy $j$



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

# Multiple Factor Analysis

MFA is looking for:

- Objects oppositions provided by several hierarchies
- Objects oppositions provided at upper levels of these hierarchies

MFA provides different representations:

- An objects representation
- An hierarchies representation
- A levels representation

# Objects representation

Distance between 2 objects:

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

with:

- $Q_j$  the number of level of hierarchy  $j$
- $l$  the number of objects
- $l_k$  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



## Hierarchies representation

Coordinate of hierarchy  $j$  on the axis of rank  $s$ :

$$\sum_{k \in Q_j} \eta^2(z_s, L_k),$$

with:

- $Q_j$  the number of level of hierarchy  $j$
- $z_s$  the axis  $s$
- $L_k$  the level  $k$  of the hierarchy  $j$

⇒ A hierarchy will have an higher coordinate on the axis  $s$  than the objects opposition highlighted by axis  $s$  appears at an upper level of the hierarchy

# Levels representation

Coordinate of level  $k$  on the axis of rank  $s$ :

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

with:

- $z_s$  the axis  $s$
- $L_k$  the level  $k$

⇒ A level will have an higher coordinate on the axis  $s$  than the object opposition highlighted by axis  $s$  corresponds to the level

# Outline

- 1 Objectives
- 2 Data collection
- 3 Statistical analysis
- 4 Application**
- 5 Conclusion

# Data

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



# Binary hierarchical sorting

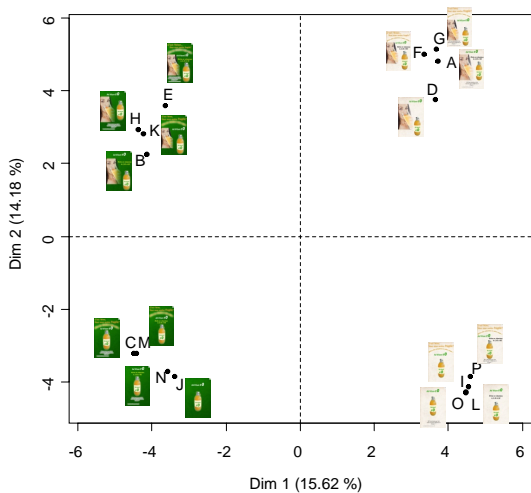
Each subject is asked:

- Firstly, to separate the objects (advertisements) into 2 groups
- Secondly, to separate these 2 groups into 2 sub-groups each, if she/he wants
- And so on, until the subject obtains homogenous groups

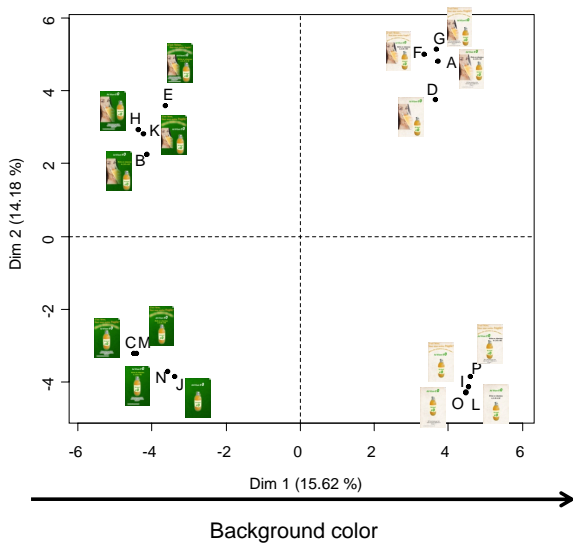
# Example of hierarchical sorting: subject number 3



# Advertisements representation

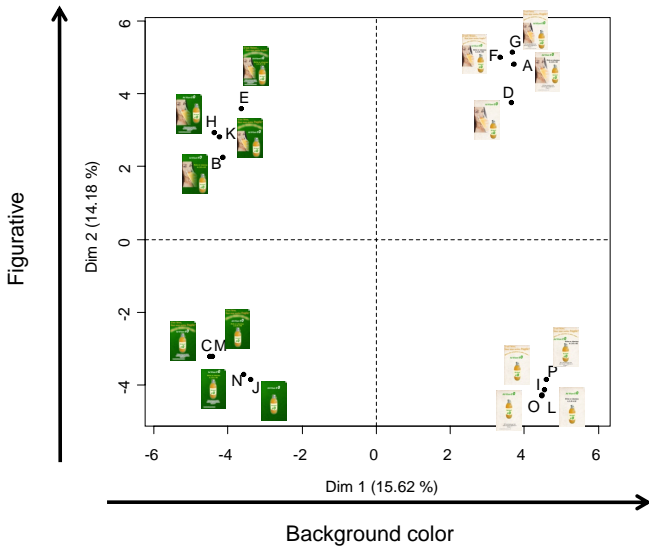


# Advertisements representation

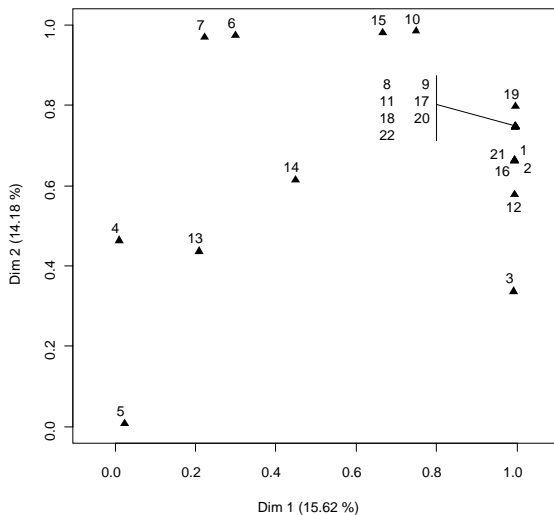




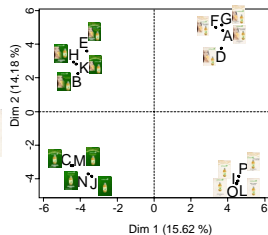
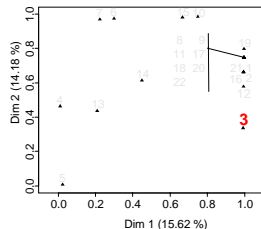
# Advertisements representation



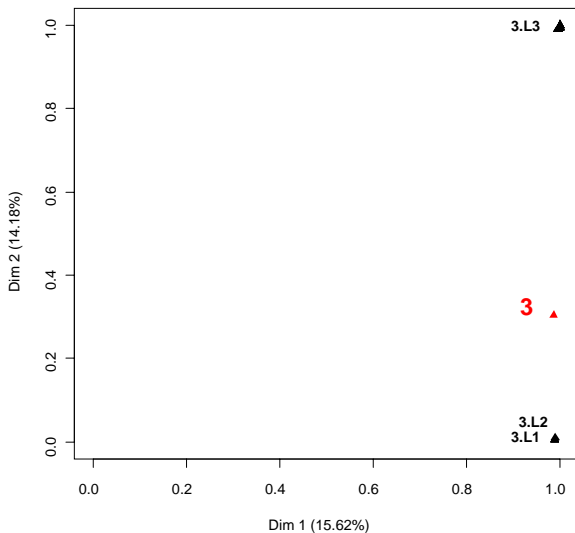
# Hierarchies representation



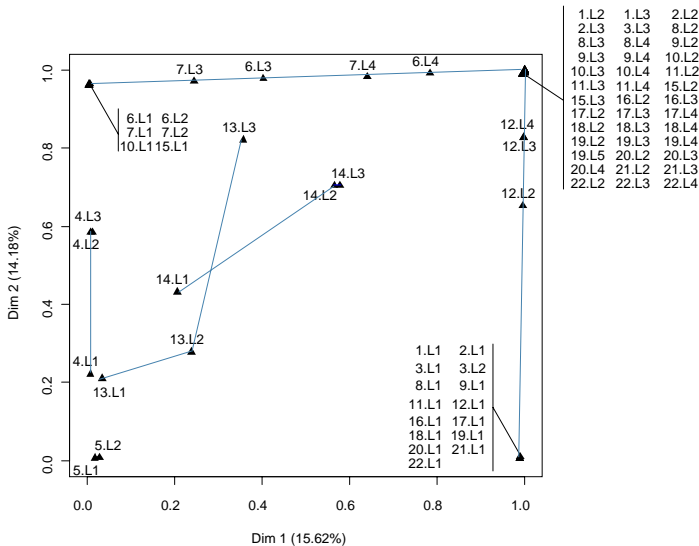
# Hierarchies representation: subject number 3



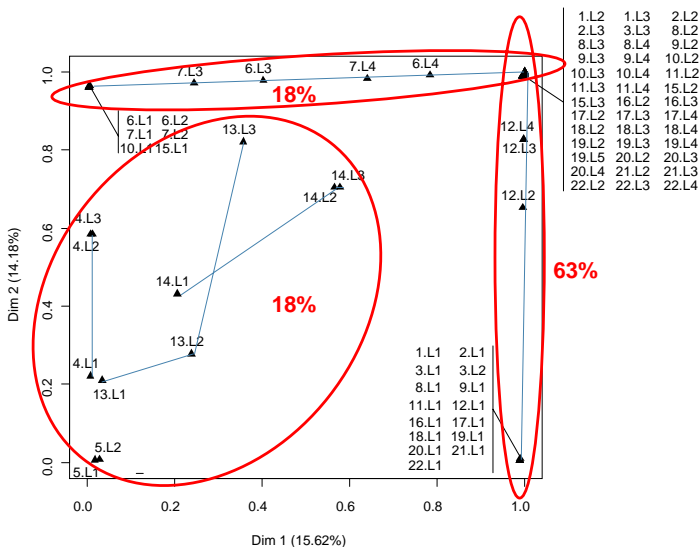
# Levels representation: subject number 3



# Levels representation: trajectories



# Levels representation: trajectories



# Outline

- 1 Objectives
- 2 Data collection
- 3 Statistical analysis
- 4 Application
- 5 Conclusion

# Conclusion

- Methodology providing rich and interpretable results:
  - Representation of objects, hierarchies, levels
  - Representations related to each other
- Can be used to obtain groups of hierarchies
- Allows the simultaneous taking into account of hierarchies and partitions in a same analysis