

# Beyond numbers: Multiple correspondence analysis

## **Ctrl+R Seminar**

Thierry Rossier & Pierre Benz

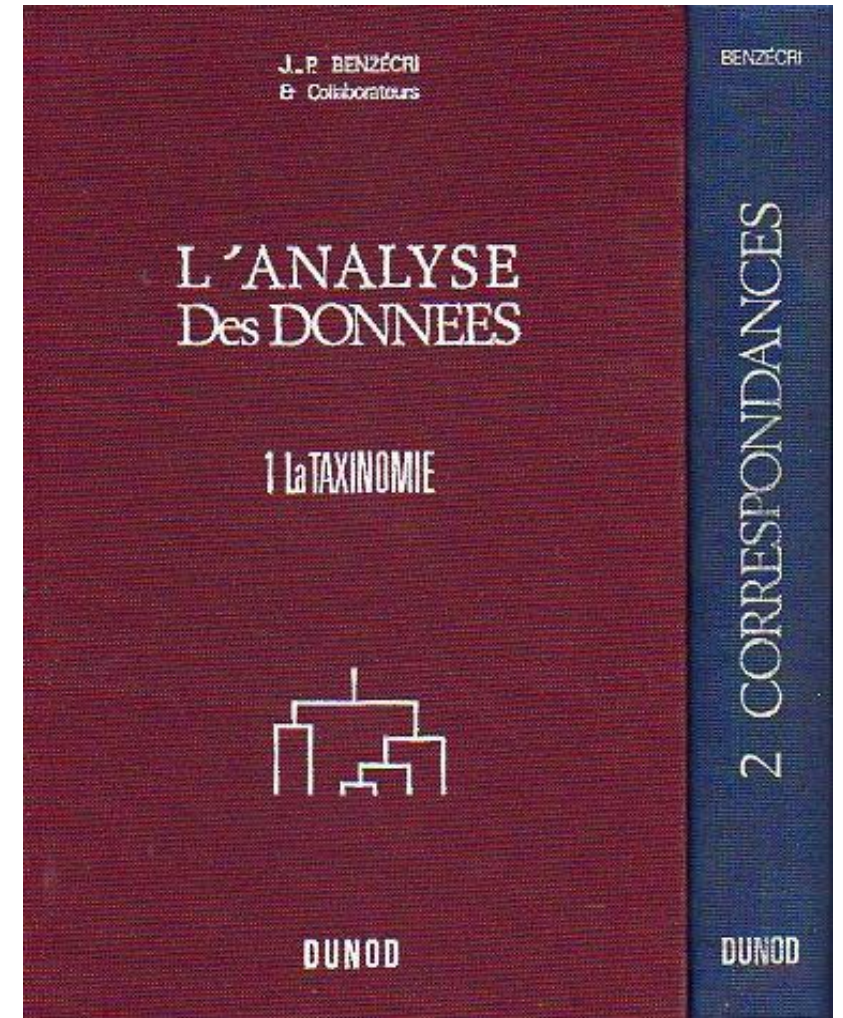
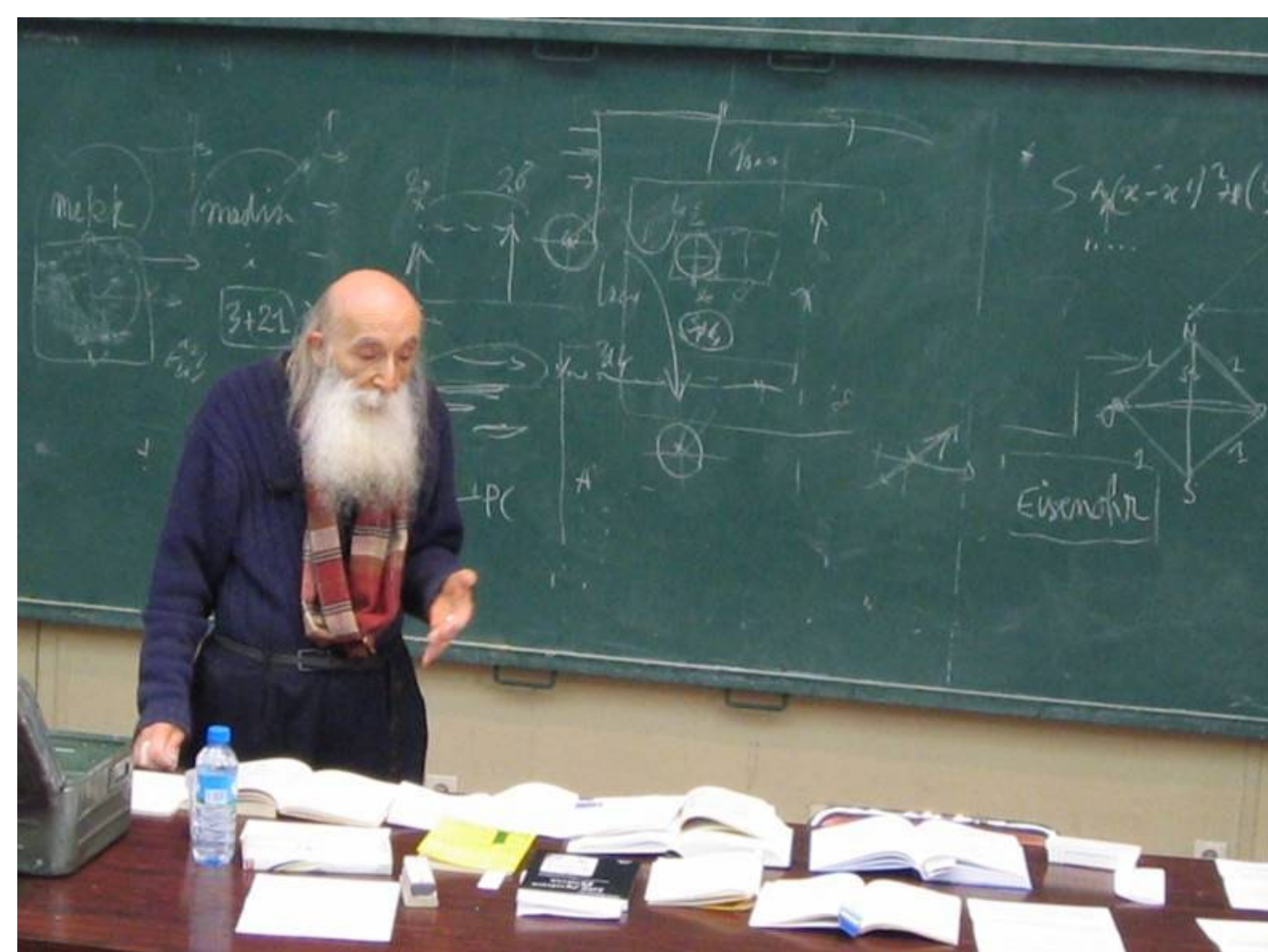
December 5th 2024

# Slides and material

[https://github.com/BenzPierre/ctrl-R\\_MCA](https://github.com/BenzPierre/ctrl-R_MCA)

Jean-Paul Benzécri (1932-2019)

J-P Benzécri: *L'analyse des données*. Dunod, 1973







# MCA

- Part of a larger framework: geometric data analysis (GDA)
- Multivariate, multidimensional and descriptive methods
- Dimensionality reduction
- Examples: principal component analysis (PCA); multidimensional scaling; multiple factor analysis etc.
- Often represented geometrically
- See [https://fr.wikipedia.org/wiki/Analyse\\_des\\_donn%C3%A9es](https://fr.wikipedia.org/wiki/Analyse_des_donn%C3%A9es)

# MCA in a nutshell

- Factor analysis for categorical data
- Meaning that instead of forcing categorical data into numeric scales to do a PCA (like many quant scholars wrongly do, I'm afraid)
- You can just respect statistical laws and do an MCA instead

# The key ideas of MCA (Le Roux and Lebaron, 2015)

1. A geometric model
2. A formal approach
3. An «inductive» and «descriptive» procedure

# 1) A Geometric Model

- A data table is transformed into a cloud of categories and a cloud of individuals
- Two phases:
  1. Construction of clouds: define metrics (Euclidian distances)
  2. Reduction of dimensionality: looking for the best dimensional adjustment



## 2) A Formal Approach

- Based on mathematical structures
- Multidimensional and Euclidian geometry
  - Alignment
  - Barycentre
  - Distances and angles

### 3) An Inductive and Descriptive Procedure

- Benzécri: «The model is supposed to follow the data, not the data the model»
- No technical hypotheses: such as multinormality, homoscedaticity, etc.)
- Inductive does not mean only exploratory. Constructing the space also depends on thorough reflection and conceptualisation of the used variables and modalities
- We can have a look at the variations and interactions between different groups of variables
- Can be combined with other techniques (regression analyses etc.)

# How does an MCA work?

- An MCA is a kind of a gigantic, multiple cross table between all the variables (and their categories) \* individuals
- This multiple cross table can be imagined as a multidimensional cloud where the categories of the modalities are more or less close according to their co-occurrence
- This cloud is then projected on a two-dimensional plane in a way that maximises the explained variance
- Categories which are close to each other, are frequently shared by the same individuals; categories which are distant from each other are, are hardly shared by the same individuals