Beyond numbers:
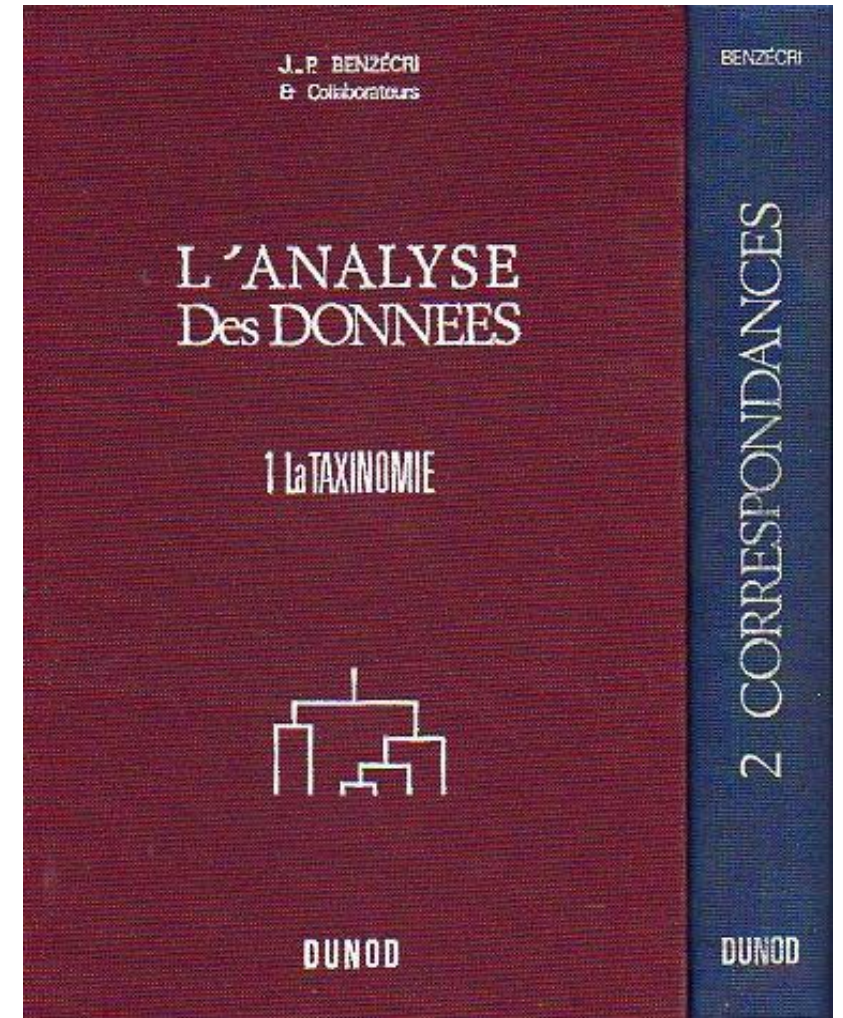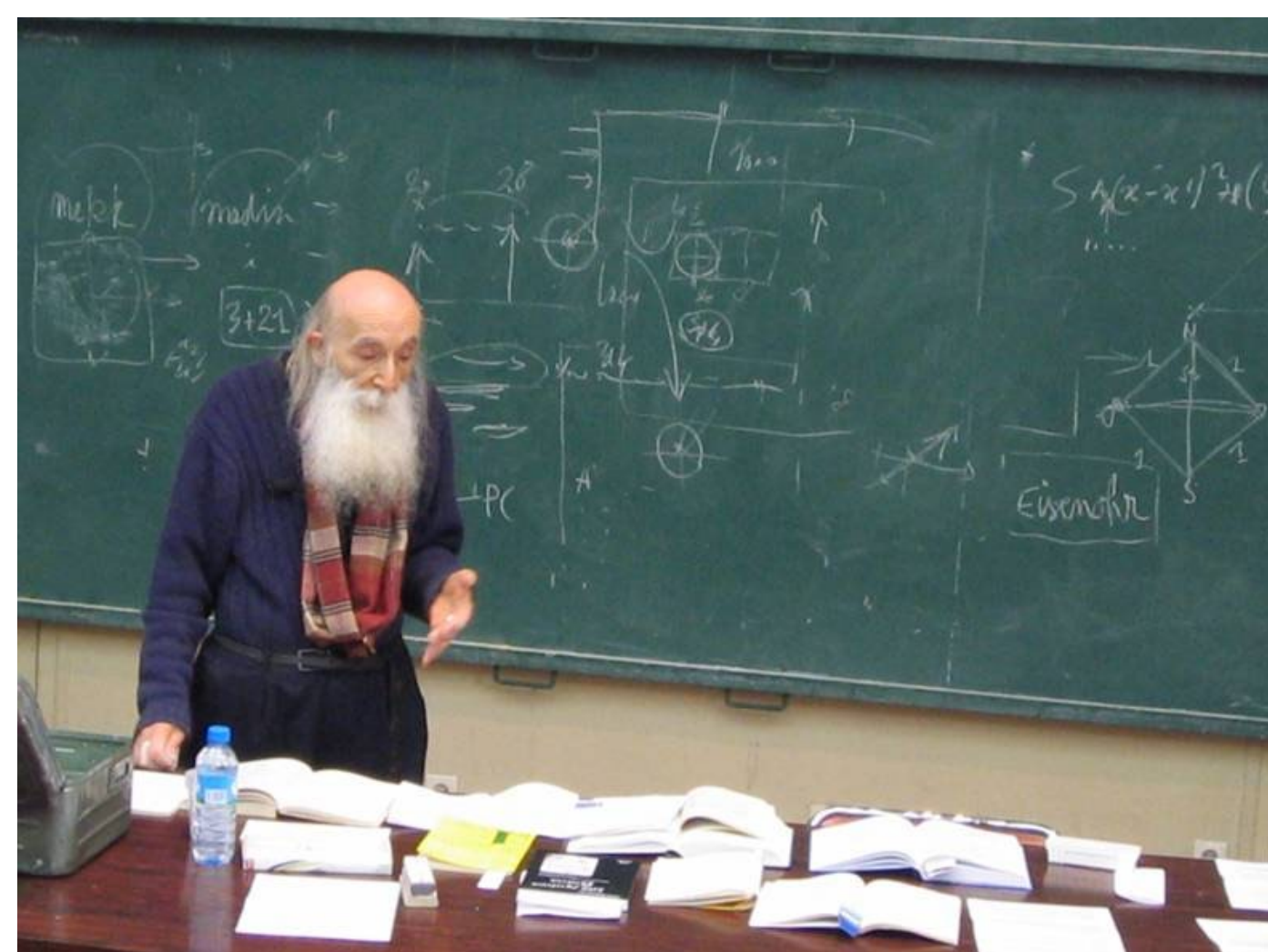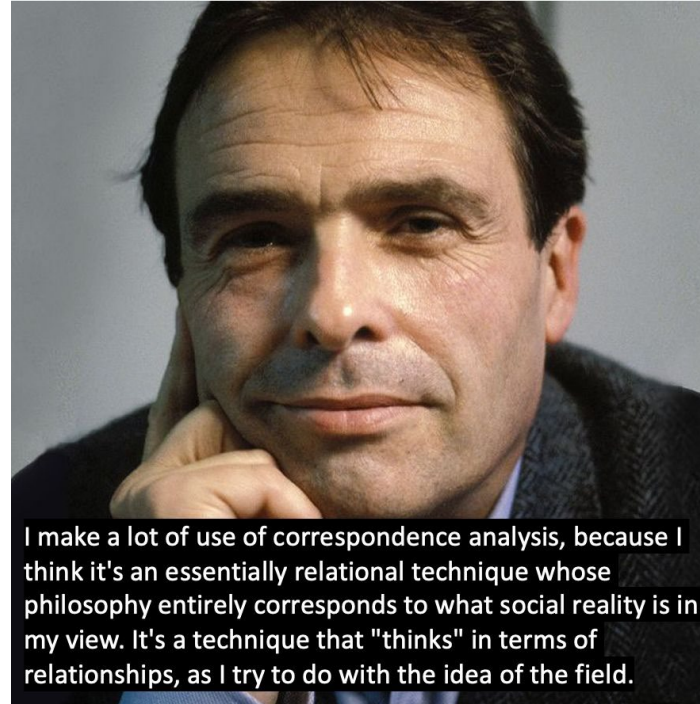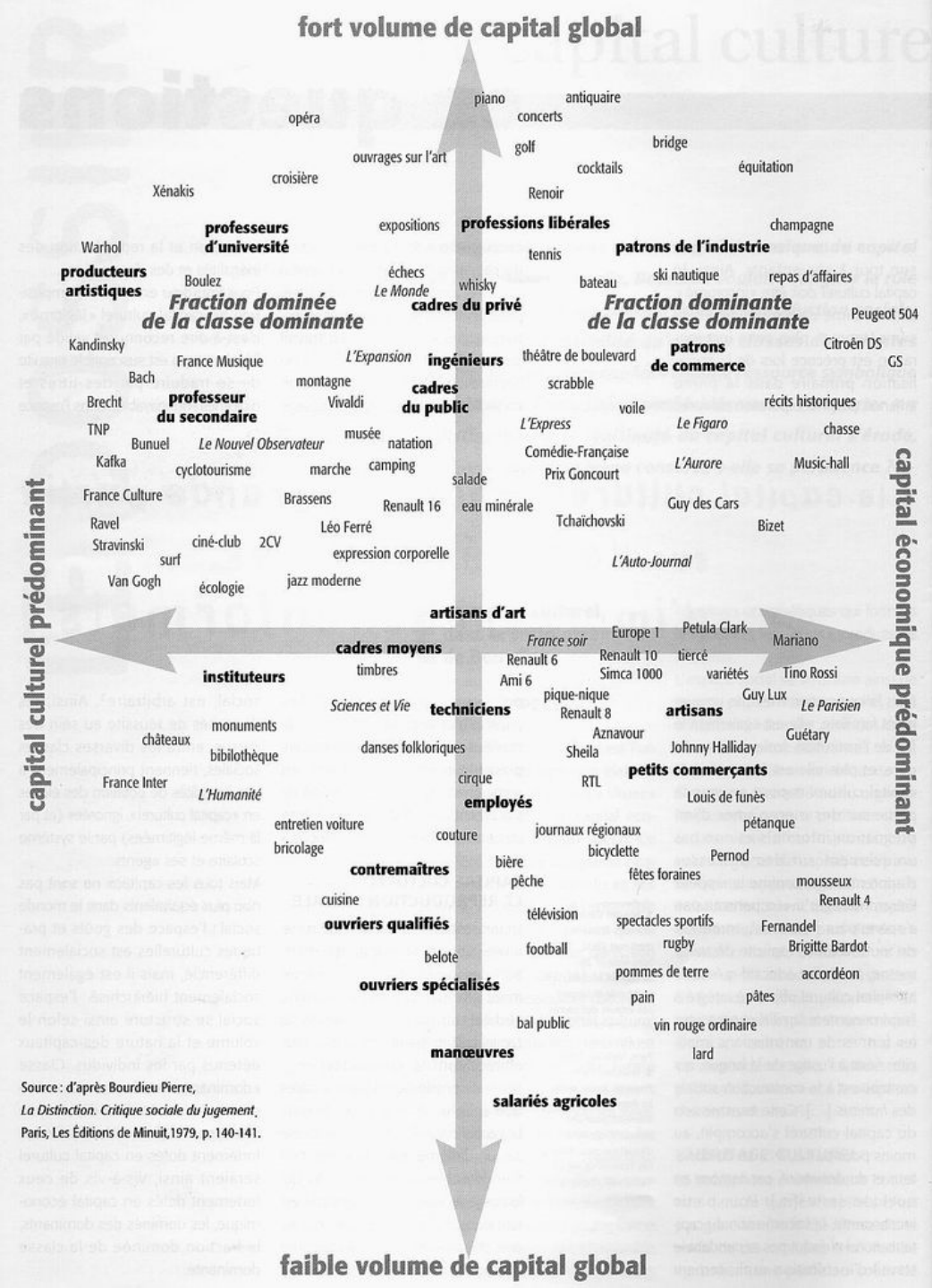# Multiple correspondence analysis

Ctrl+R Seminar

Thierry Rossier & Pierre Benz

December 5th 2024

Jean-Paul Benzécri (1932-2019)

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

fort volume de capital global

piano antiquaire
opéra concerts
bridge
ouvrages sur l'art golf cocktails équitation
croisière Renoir
Xénakis expositions professions libérales champagne
professeurs d'université tennis patrons de l'industrie
Warhol échecs ski nautique repas d'affaires
producteurs artistiques Boulez Le Monde whisky bateau
Fraction dominée de la classe dominante cadres du privé *Fraction dominante de la classe dominante* Peugeot 504
Kandinsky Citroën DS
France Musique L'Expansion ingénieurs théâtre de boulevard patrons de commerce GS
Bach montagne scrabble
Brecht professeur du secondaire Vivaldi cadres du public récits historiques
TNP L'Express voile Le Figaro chasse
Kafka Bunuel Le Nouvel Observateur musée natation Comédie-Française L'Aurore Music-hall
France Culture cyclotourisme marche camping Prix Goncourt
Ravel Brassens Renault 16 salade eau minérale Guy des Cars
Stravinski ciné-club 2CV Léo Ferré Tchaïchovski Bizet
surf expression corporelle L'Auto-Journal
Van Gogh écologie jazz moderne

capital culturel prédominant ← artisans d'art → capital économique prédominant

Europe 1 Petula Clark
cadres moyens *France soir* Renault 10 tiercé Mariano
instituteurs timbres Renault 6 Simca 1000 variétés Tino Rossi
Ami 6 pique-nique Guy Lux
Sciences et Vie techniciens Renault 8 artisans Le Parisien
Aznavour Guétary
danses folkloriques Sheila Johnny Halliday
cirque RTL petits commerçants
employés Louis de funès
entretien voiture journaux régionaux pétanque
bricolage couture bicyclette Pernod
contremaîtres bière fêtes foraines
cuisine pêche mousseux Renault 4
ouvriers qualifiés télévision spectacles sportifs Fernandel
belote football rugby Brigitte Bardot
pommes de terre accordéon
ouvriers spécialisés pain pâtes
bal public vin rouge ordinaire
manœuvres lard

salariés agricoles

faible volume de capital global

Source: d'après Bourdieu Pierre, *La Distinction. Critique sociale du jugement*, Paris, Les Éditions de Minuit, 1979, p. 140-141.

I make a lot of use of correspondence analysis, because I think it's an essentially relational technique whose philosophy entirely corresponds to what social reality is in my view. It's a technique that "thinks" in terms of relationships, as I try to do with the idea of the field.

"What is his evidence? Incomprehensible diagrams (…)".

Stanley Lieberson, ASA presidential address 1991 (about Bourdieu's use of MCA)

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

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