



**HAL**  
open science

# Analyses factorielles multivariées en Génétique et (Gén)omique. Du neuf avec du vieux.

Denis Laloë

► **To cite this version:**

Denis Laloë. Analyses factorielles multivariées en Génétique et (Gén)omique. Du neuf avec du vieux..  
École thématique. TransModGen, Toulouse (En ligne), France. 2023, pp.63. hal-04443620

**HAL Id: hal-04443620**

**<https://hal.inrae.fr/hal-04443620>**

Submitted on 7 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Analyses Factorielles multivariées en Génétique et (Gén)omique. Du neuf avec du vieux

7 Février 2023

Denis Laloë  
GABI - GiBBS

TransModGen, Toulouse, 7 février 2023

## Introduction

Ensemble de méthodes partageant +/- une approche et une formalisation commune. Remonte au début du 20ème siècle

### Quelques jalons

- ▶ ACP : Pearson K. 1901. On lines and planes of closest fit to systems of points in space. *Philos Mag A*. 6: 559-572
- ▶ ACP : Hotelling H. 1933. Analysis of a complex of statistical variables into principal components. *J Educ Psychol*. 25: 417-441
- ▶ SVD : Eckart C and Young G. 1936. The approximation of a matrix by another of a lower rank. *Psychometrika*. 1: 211-218.1
- ▶ Correspondances : Fisher, R.A. 1940. The precision of discriminant functions. *Annals of Eugenics*, 10, 422- 429.
- ▶ Correspondances : Benzécri J.P., 1965-66, *Leçons sur l'analyse factorielle et la reconnaissance des formes*, Rennes
- ▶ Rao, C. R. (1964). RDA : The use and interpretation of principal component analysis in applied research. *Sankhya Ser. A* 26 329?359
- ▶ Schéma de dualité : Escoufier, Y. 1977. Operators related to a data matrix. In *Recent Developments in Statistics* (J. Barra, F. Brodeau, G. Romier and B. van Cutsem, eds.) 125?131. Amsterdam. North Holland.
- ▶ Analyse Factorielle Multiple : Escoufier, B. 1984. L'analyse factorielle multiple. *Cahiers du Bureau universitaire de recherche opérationnelle Série Recherche*, 42, 3-68.
- ▶ Aspects historiques : Armatte, M. 2008. Histoire et Préhistoire de l'Analyse des données par JP Benzecri: un cas de généalogie rétrospective. *Journal électronique d'histoire des probabilités et de la statistique*, 4(2), 1-24.

## Introduction

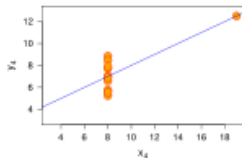
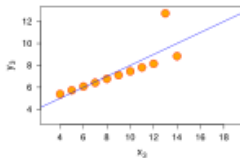
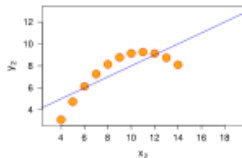
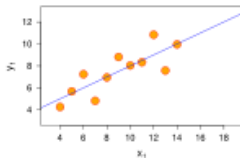
# Données

- ▶ *Le modèle doit suivre les données, non l'inverse, J P Benzécri*
- ▶ Observation vs Expérimentation
  - ▶ Donnée préexistante (Sciences sociales / Ecologie)
  - ▶ Pas de structure a priori : non supervisée ?
  - ▶ Synthèse (vision holistique / corrélations partielles / causalité) :
    - ▶ Approche de Durkheim : Pour dégager des relations causales, une relation binaire ne suffit pas, il faut intégrer plusieurs variables dans l'analyse et considérer leurs relations
    - ▶ Approche de Benzécri : C'est de la synthèse, sans a priori, que les causes émergent.

# Représentation des données

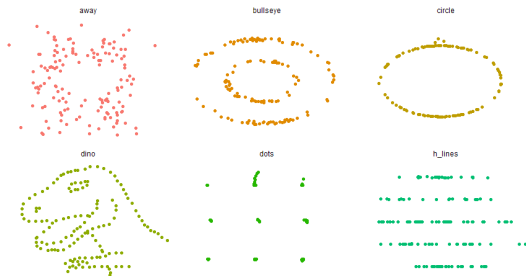
## Accent mis sur la visualisation

Approche géométrique : représentation de données sous forme de nuages de points (plutôt que des résumés quantitatifs) F J Anscombe, 1973



## Représentation des données

Approche géométrique : représentation de données sous forme de nuages de points (plutôt que des résumés quantitatifs)



## Analyse géométrique des données

- ▶ Approche géométrique : représentation de données sous forme de nuages de points (plutôt que des résumés quantitatifs)
- ▶ Efficience : synthèse par optimisation d'un critère (Inertie,...)
  - ▶ Critère existant pour analyses simples (Eckart-Young)
  - ▶ Peut dépendre de la question

## Pas d'a priori / Non supervisé ?

### Non supervisé : S'affranchir des idées a priori

- ▶ Optimisation d'un critère objectif (Inertie)
- ▶ Substitution des facteurs à *l'arbitraire échafaudage des idées a priori*

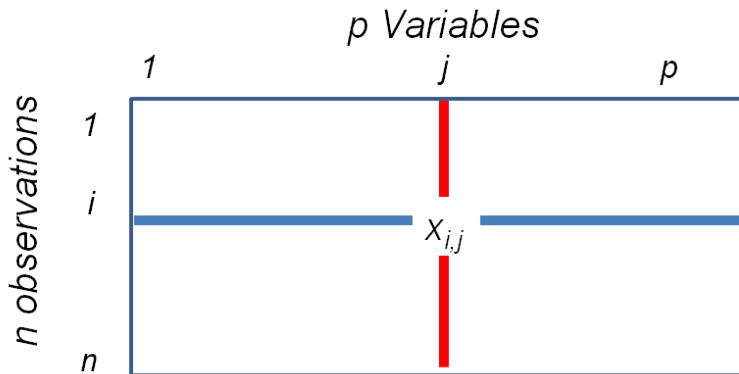
### Non supervisé : oui mais ?

- ▶ Popper: All observation is selective and theory-laden; there are no pure or theory-free observations.
- ▶ Brestoux:
  - ▶ Choix des métriques
  - ▶ Recueil/Construction des données (Echantillonnage, environnement);
  - ▶ Transformation des données (Centrage, standardisation)
  - ▶ De la pauvreté de l'a priori à la pauvreté de l'interprétation

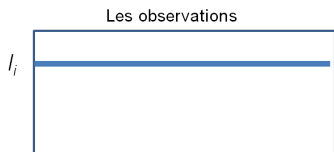
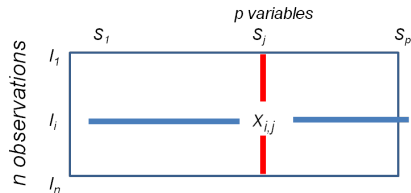


## L'ACP

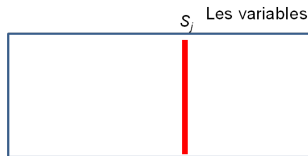
Un tableau de variables quantitatives



## Un tableau, deux points de vue



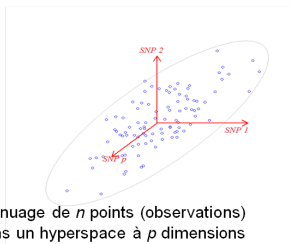
La  $i$ ème observation définie par  $p$  variables



La  $j$ ème variable définie par  $n$  observations

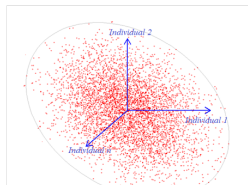
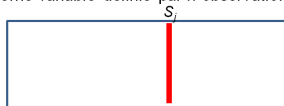
## Deux représentations géométriques

La  $i$ ème observation définie par  $p$  variables



Un nuage de  $n$  points (observations)  
dans un hyperspace à  $p$  dimensions  
(variables)

La  $j$ ème variable définie par  $n$  observations

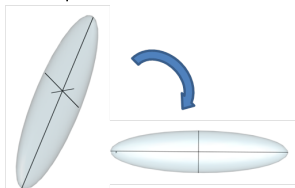


Un nuage de  $p$  points (variables)  
dans un hyperspace à  $n$  dimensions  
(observations)

## Transformation des nuages

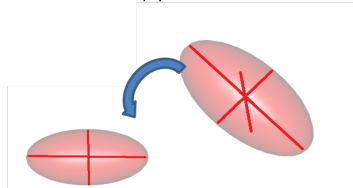
### Un même mécanisme : le schéma de dualité

Hyperspace des observations  
 $n$  points dans  $\mathbb{R}^p$



distance entre observations

Hyperspace des variables  
 $p$  points dans  $\mathbb{R}^n$



distance entre variables

- ▶ Décomposition en valeurs singulières
- ▶ Décomposition canonique

## Schéma de dualité

$$(\mathbf{X}, \mathbf{Q}, \mathbf{D}) \Leftrightarrow \begin{array}{ccc} \langle p \rangle & \xrightarrow{\mathbf{Q}} & \langle p \rangle \\ \mathbf{X}' \uparrow & & \downarrow \mathbf{X} \\ \langle n \rangle & \xleftarrow{\mathbf{D}} & \langle n \rangle \end{array}$$

- ▶  $\mathbf{X}_{(n, p)}$  : matrice de données (transformées)
- ▶  $\mathbf{Q}_{(p, p)}$  : matrice d'un produit scalaire, carrée symétrique
- ▶  $\mathbf{D}_{(n, n)}$  : matrice d'un produit scalaire, carrée symétrique

## Le schéma de dualité - ACP

Maximisation of the correlation between variables and components

$$V = X'X/n$$

$$VA = A\Delta$$

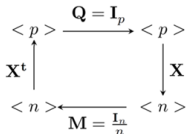
$$A'A = I$$

Principal axes

Variable scores

$$C = X'B$$

Best approximation (rank  $l$ )  
Eckart and Young



Diagonalisation

$$X'X \quad XX'$$

mêmes valeurs propres non nulles

$$\lambda_1 > \lambda_2 > \dots > \lambda_r > 0$$

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$$

Transition formulae

$$XA\Delta^{-0,5} = B$$

$$X'B\Delta^{0,5} = A$$

Singular value decomposition

$$X = B\Lambda^{0,5}A'$$

$$\hat{X}_l = \sum_{i=1, l} \sqrt{\lambda_i} \mathbf{b}_i \mathbf{a}_i'$$

Maximisation of the dispersion of individuals

Observations

$$W = XX' / n$$

$$WB = B\Delta$$

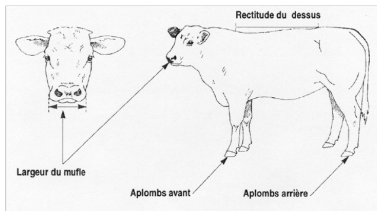
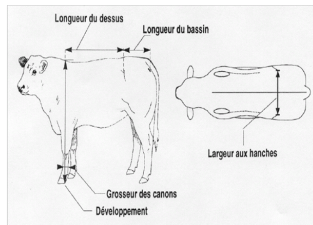
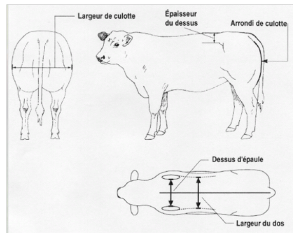
$$B'B = I$$

Principal components

Observation scores

$$L = XA$$

## Pointage des bovins allaitants



# Pointage des bovins allaitants

## Une session d'homogénéisation

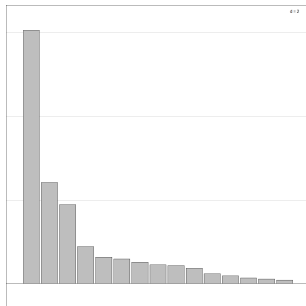
Homogénéiser la façon de pointer des techniciens pointeurs

- ▶ 43 animaux
- ▶ 4 techniciens
- ▶ 15 postes



## ACP sur le pointage des bovins allaitants

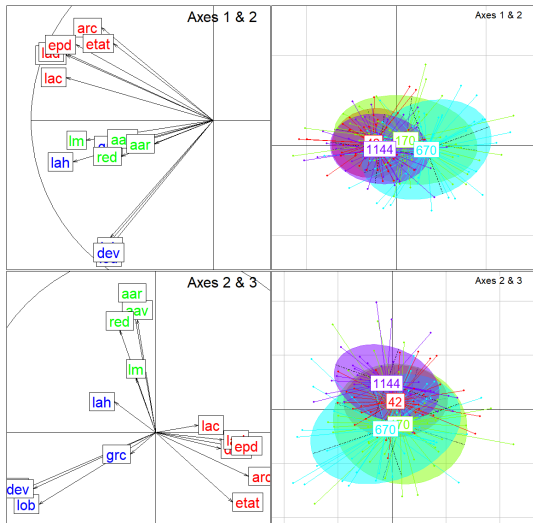
### Les inerties



Choix du nombre de composantes ?

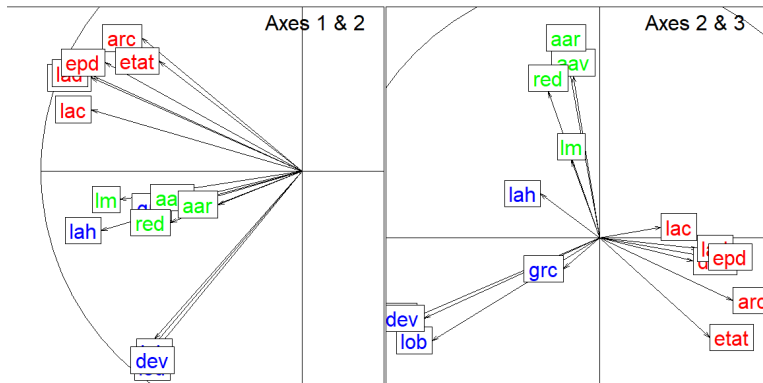
## ACP sur le pointage des bovins allaitants

Représentation conjointe: Cercle des corrélations; Carte factorielle



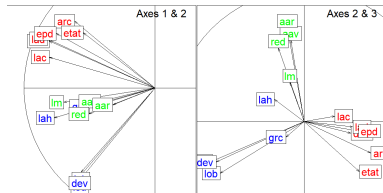
## ACP sur le pointage des bovins allaitants

### Le cercle des corrélations



## Taille, forme, valeurs propres

### Le cercle des corrélations



- ▶ 1ère composante : taille : un arbre qui cache la forêt ?
- ▶ Composantes suivantes: formes

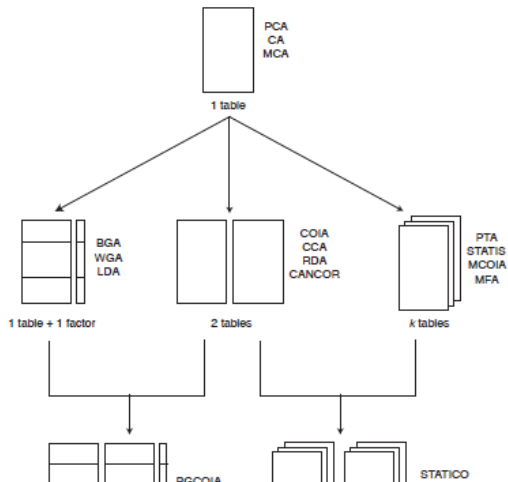
## Une représentation partagée

- ▶ Représentation compréhensible de relations entre variables
- ▶ Dialogue
  - ▶ Notion de taille générale (1er axe) peu perçue : on perçoit plus les oppositions entre variables (forme, axes suivants)
  - ▶ Dialogue possible car mise en évidence de la forme
  - ▶ Position particulière de la grosseur des canons
  - ▶ intérêt de diminuer le nombre de postes pointés

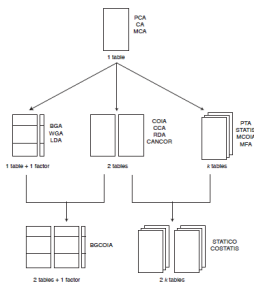
## Les extensions

Thioulouse, 2011. Simultaneous analysis of a sequence of paired ecological tables: A comparison of several methods. Ann.

Appl. Stat. 5 (4) 2300 - 2325



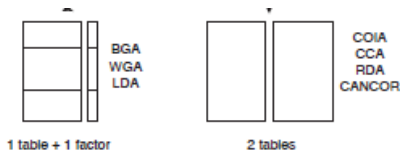
## Les extensions



- ▶ 2, k-tables
- ▶ Structures particulières
- ▶ Intégration de données
- ▶ Modélisation; prédiction

## Deux tableaux

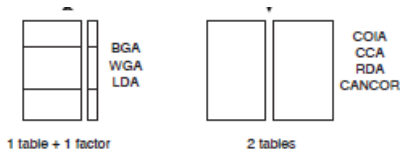
- ▶ Dissymétrie: un tableau explique l'autre: Analyse de redondance (RDA)
- ▶ Symétrie: Les deux tableaux jouent le même rôle: Analyse de coinertie (COIA) / Corrélation canonique (CANCOR)
- ▶ Séparation selon un facteur : Analyse discriminante (LDA) / Analyse entre groupes (BGA)





## Tableaux et dimensions

Etape d'inversion de matrices (prédiction,...)  $\Rightarrow$  contrainte sur les dimensions des tables



Relations entre deux tables	Inversion de matrice	Matrice	Conséquences
Coinertie COIA X vs Y	non	non	-
Corrélations canoniques CANCOR	oui	$(X'X) ; (Y'Y)$	$p_x < n ; p_y < n$
Analyse de redondance RDA $Y=f(X)$	oui	$(X'X)$	$p_x < n$
Analyse discriminante LDA $X=f(g)$	oui	variance intra-classe	$p_x < n$
Analyse entre groupes (BGA) $X=f(g)$	non	non	-

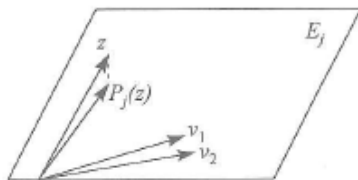
## Relations symétriques entre deux tableaux

- ▶ Corrélations Canoniques
- ▶ Co-inertie
- ▶ Coefficient  $R_v$

## Relations entre deux tableaux

### Corrélations canoniques

$$\begin{array}{ccc}
 \langle p \rangle & \xrightarrow{(X'DX)^{-1}} & \langle p \rangle \\
 X'DY \uparrow & & \downarrow Y'DX \\
 \langle m \rangle & \xleftarrow{(Y'DY)^{-1}} & \langle m \rangle
 \end{array}$$



**Figure 88.** Inadaptation du coefficient de corrélation multiple dans le cas de variables liées.

La variable  $z$  est presque orthogonale à chacune des variables  $v_1$  et  $v_2$  du groupe  $K_j$ . Or, son coefficient de corrélation multiple avec  $E_j$  vaut presque 1.

## Relations entre deux tableaux

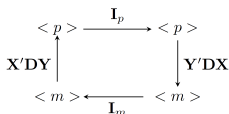
### Corrélations canoniques

$$\begin{array}{ccc}
 \langle p \rangle & \xrightarrow{(\mathbf{X}'\mathbf{D}\mathbf{X})^{-1}} & \langle p \rangle \\
 \mathbf{X}'\mathbf{D}\mathbf{Y} \uparrow & & \downarrow \mathbf{Y}'\mathbf{D}\mathbf{X} \\
 \langle m \rangle & \xleftarrow{(\mathbf{Y}'\mathbf{D}\mathbf{Y})^{-1}} & \langle m \rangle
 \end{array}$$

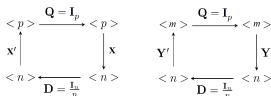
- ▶  $p \gg n \rightarrow (\mathbf{X}'\mathbf{D}\mathbf{X})$  non inversible
- ▶ Besoin d'une régularisation  $\rightarrow (\mathbf{X}'\mathbf{D}\mathbf{X} + \alpha_{\mathbf{X}}\mathbf{I})^{-1}$
- ▶ Choix de  $\alpha_{\mathbf{X}}$  par cross-validation

# Relations entre deux tableaux

## Coinertie - Rv



$$covv(\mathbf{X}, \mathbf{Y}) = \sum \mu_{\mathbf{X}, \mathbf{Y}}$$



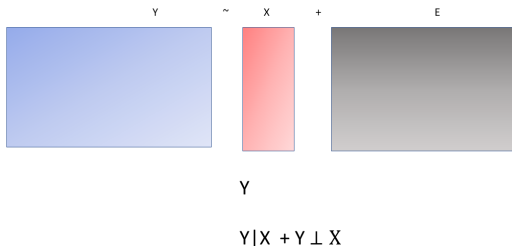
$$vav(\mathbf{X}) = \sum \lambda_{\mathbf{X}}^2$$

$$vav(\mathbf{Y}) = \sum \lambda_{\mathbf{Y}}^2$$

$$Rv_{\mathbf{X}, \mathbf{Y}} = \frac{covv(\mathbf{X}, \mathbf{Y})}{\sqrt{vav(\mathbf{X})vav(\mathbf{Y})}}$$

## Deux tableaux : Analyse de Redondance - RDA

RDA: Redundancy Analysis (aka PCAIV - Analyse sur variables instrumentales) Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. Sankhya: The Indian Journal of Statistics, Series A, 329-358.



Interprétation de la variation d'un ensemble de variables (Y) en fonction d'un autre ensemble de variables (X)

- ▶ Approche "modèle linéaire" dans un contexte multivarié
- ▶  $\hat{Y} = X[X^tX]^{-1}X^tY$
- ▶ Contraintes sur le nombre de variables de X / nombre de lignes des tableaux

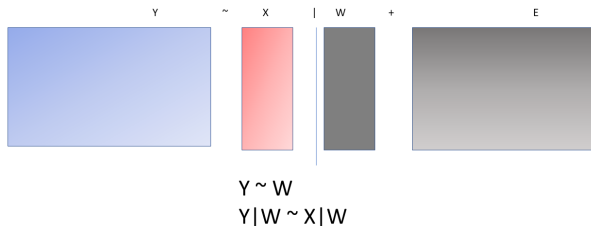
## L'analyse de redondance

- ▶ Approche "modèle linéaire" dans un contexte multivarié
- ▶ Tableau d'analyse de variance
- ▶ Tests de significativité (permutation) et Pourcentage d'inertie due
  - ▶ au modèle
  - ▶ aux axes
  - ▶ aux différents facteurs

## Deux (trois) tableaux : partial RDA

partial RDA: Redundancy Analysis with conditioning (partialling out)

Borcard, D., Legendre, P., and Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73(3), 1045-1055.



$$\mathbf{U} = \mathbf{Y}_{\text{res}|W}$$

$$\mathbf{V} = \mathbf{X}_{\text{res}|W}$$

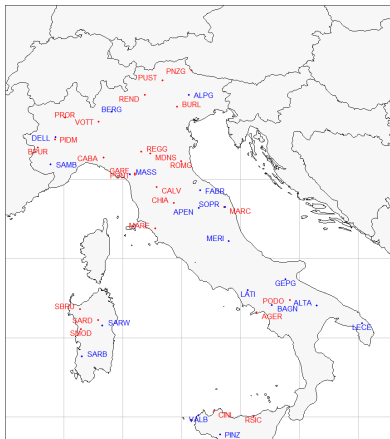
$$\hat{\mathbf{U}} = \mathbf{V}[\mathbf{V}^t\mathbf{V}]^{-1}\mathbf{V}^t\mathbf{U}$$



# RDA et génomique environnementale

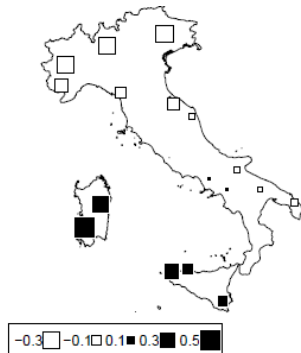
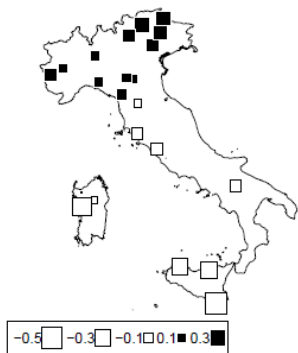
## Structuration géographo-climatique des ruminants en Italie

Senczuk et al, 2022. How Geography and Climate Shaped the Genomic Diversity of Italian Local Cattle and Sheep Breeds. *Animals*, 12(17), 2198.



## Structuration géographico-climatique des ruminants en Italie

Structuration de la diversité de bovins et ovins italiens en fonction de la géographie (latitude/longitude) et du climat (base de données Climond).  
 Diversité = f(Géographie)



## Structuration géographico-climatique des ruminants en Italie

Décomposition de l'inertie en fonction de la géographie, du climat conditionné par la géographie

Table 5. Decomposition (percentage and *p*-values) of the inertia according to geography and climate (partialled out for geography).

RDA Component		Cattle (Fst = 0.177)		Sheep (Fst = 0.144)	
		% Inertia	<i>p</i> -Value	%	<i>p</i> -Value
Geog (Lat*Long)	RDA1	7.5	0.03	11.5	< 0.01
	RDA2	4.6	0.66	8.1	0.02
	RDA3	2.6	0.97	6.7	0.14
	Total	14.7		26.3	
Clim   Geog	RDA1	6	0.09	7.5	0.18
	RDA2	4.1	0.56	4.8	0.74
	Total	10.1		12.3	

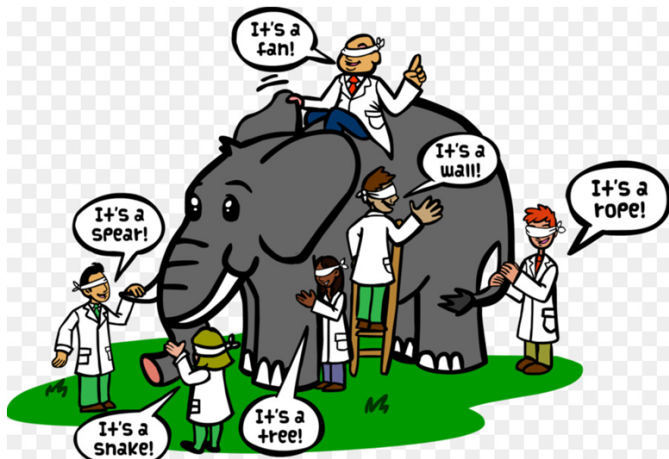
- ▶ Bovins : 14,7 %
- ▶ Ovins : 26,3 %
- ▶ Prise en compte un peu simpliste de la géographie (Latitude:Longitude).
- ▶ Spatial PCA et graphes de voisinage. Laloë et al. 2010. Spatial trends of genetic variation of domestic ruminants in Europe. *Diversity*, 2(6), 932-945.

## Retour sur la RDA

- ▶ Méthode puissante et robuste ( BCA / analyses discriminantes)
- ▶ Attention aux artefacts quand  $n \ll p$ . Nécessité de faire des tests...
- ▶ Thioulouse, J. et al ,2021. Overcoming the spurious groups problem in between-group PCA. *Evolutionary Biology*, 48(4), 458-471.

## Analyses multitable

- ▶ Appréhension d'un phénomène selon plusieurs points de vue
- ▶ Les variables sont structurées en plusieurs groupes



## Analyses multitables

Ensemble d'individus décrits par plusieurs groupes de variables

- ▶ Relations entre tableaux (interstructure) - *Coefficient de corrélation*
- ▶ Recherche d'un compromis - *Droite de régression*
- ▶ Intrastructure (Ecart au compromis) - *Ecart à la droite de régression*

## Analyses multitable

Plusieurs stratégies Pas d'analogie au théorème d'Eckart-Young

- ▶ Relations entre tableaux
  - ▶ Mesure de proximité entre tableaux :  $Rv$
  - ▶ ACP sur les  $Rv$  : Statis
- ▶ Recherche explicite d'un compromis : Analyse de co-inertie multiple
- ▶ Recherche d'une synthèse : Analyse Factorielle Multiple

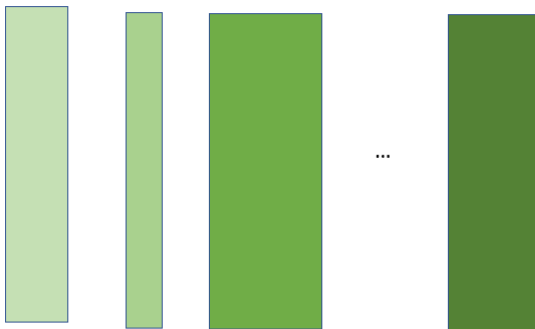
## L'Analyse Factorielle Multiple (AFM, Escofier et Pagès, 1997)

- ▶ L'AFM traite simultanément des tableaux croisant les mêmes individus
- ▶ Représentation superposée des ACP partielles
- ▶ Pondération équilibrée des différents groupes de variables



## L'Analyse Factorielle Multiple

Mêmes observations,  $K$  tables de données



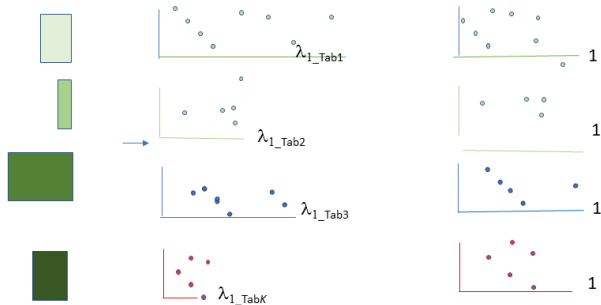
$K$  blocs de données

# L'Analyse Factorielle Multiple

## Une ACP pondérée

Etape de pondération par  $1/\lambda_1$ :

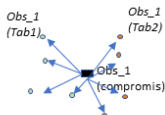
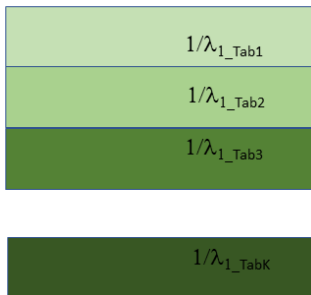
Standardisation par l'échelle du premier axe de l'ACP de chaque bloc de données



## L'Analyse Factorielle Multiple

### Une représentation simultanée du compromis et des analyses partielles

ACP sur la combinaison  
des blocs pondérés



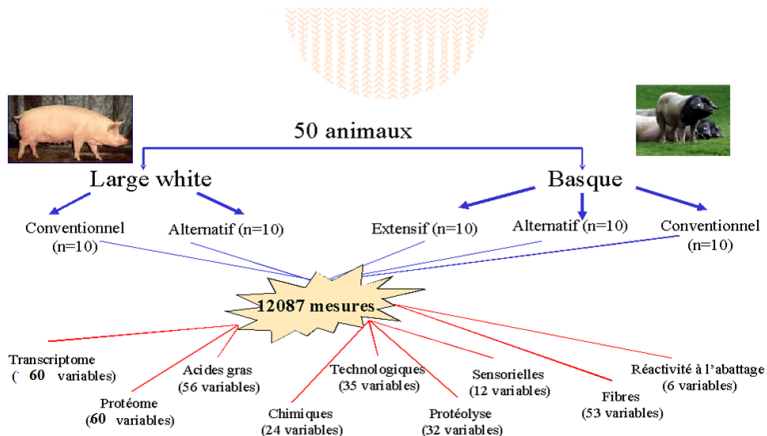
ACP :

- représentation des gènes
  - \* 1 point compromis
  - \* K points partiels
- Représentation des échantillons

## Une application. Qualité de viande chez le porc

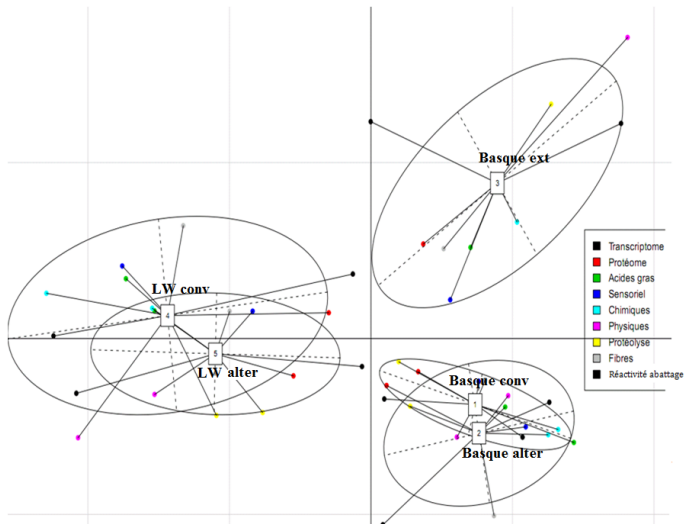
Thèse de B Salmi / C Larzul

Salmi B. et al (2010). Multivariate analysis to compare pig meat quality traits according to breed and rearing system , Proceedings of the 9th WCGALP, Leipzig, August 1-6, 2010, 442



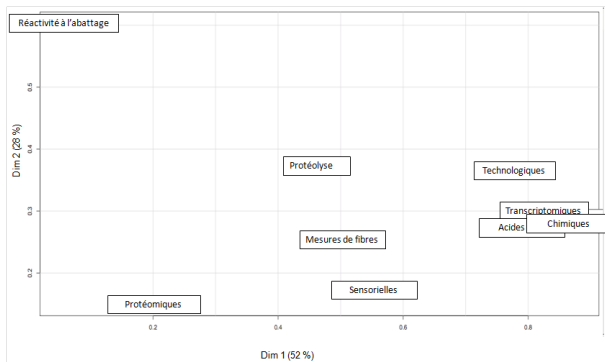
## AFM entre races\* systèmes d'élevage

### Consensus et nuages partiels



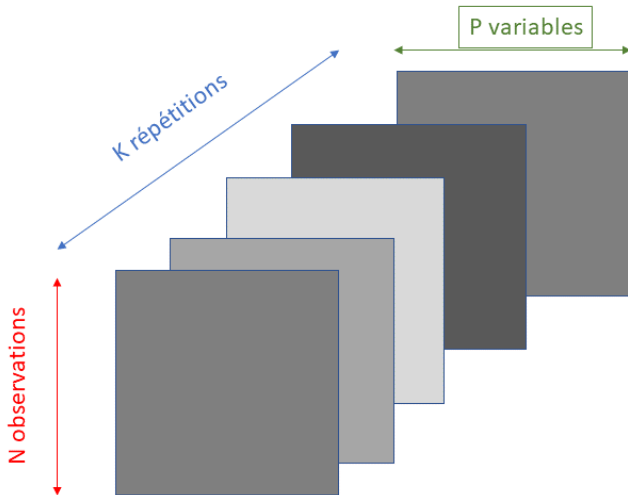
## AFM entre races\* systèmes d'élevage

Lien des groupes de variables et des composantes

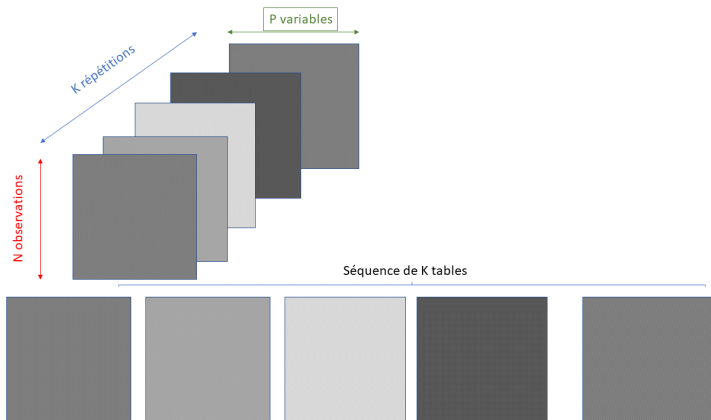


Lien du groupe  $k$  sur la composante  $i$  : inertie projetée du nuage  $k$  sur la composante  $i$

## Structure en cube



## D'une structure en cube à une séquence de K tables L'Analyse Triadique Partielle



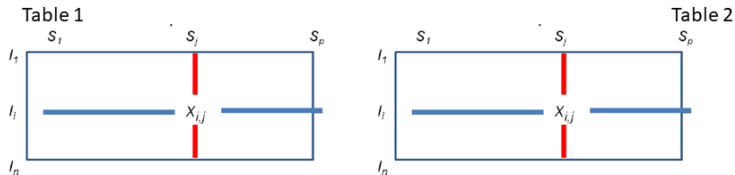


## L'analyse triadique partielle en résumé

- ▶ Calcul d'une corrélation moyenne entre tables
- ▶ Création d'une table compromis - moyenne pondérée des tables
- ▶ Coordonnée moyenne +  $K$  coordonnées partielles (une par table)
  - ▶ pour les lignes (observations)
  - ▶ pour les colonnes (variables)
- ▶ Si les coordonnées sont +/- les mêmes, la variable s'exprime semblablement quel que soit la table
- ▶ Sinon, la variable s'exprime différemment selon la table

## PTA : Détail des calculs.1

### Quantification des relations entre les tables

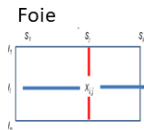
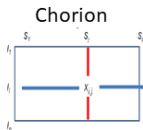


1. Comment quantifier la ressemblance entre deux tables ?

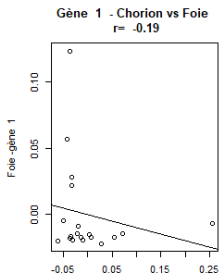
## PTA : Détail des calculs.2

Corrélation entre les expressions d'un même gène dans deux tables.

1. Comment quantifier la ressemblance entre deux tables ?



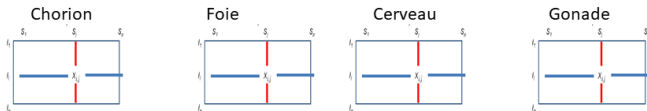
1 gène :  
Corrélation entre  
expression dans le  
chorion et expression  
dans le foie  $r = -0,19$



## PTA : Détail des calculs.3

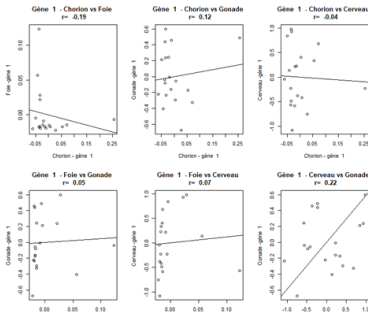
Corrélation entre les expressions d'un même gène pour toutes les tables 2 à 2.

1. Comment quantifier la ressemblance entre deux tables ?



Gène 1

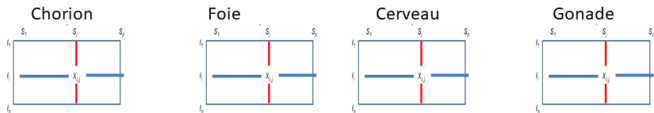
	chorion	foie	gonade	cerveau
chorion	1	-0,19	0,12	-0,04
foie	-0,19	1	0,05	0,07
gonade	0,12	0,05	1	0,22
cerveau	-0,04	0,07	0,22	1



## PTA : Détail des calculs.4

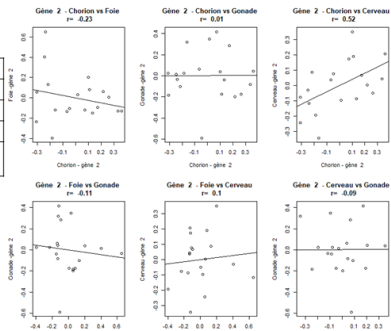
Corrélation entre les expressions d'un même gène pour toutes les tables 2 à 2.

1. Comment quantifier la ressemblance entre deux tables ?



Gène 2

	chorion	foie	gonade	cerveau
chorion	1	-0,23	0,01	0,52
foie	-0,23	1	-0,11	0,1
gonade	0,01	-0,11	1	-0,09
cerveau	0,52	0,1	-0,09	1



## PTA : Détail des calculs.5

### Moyenne des corrélations.

#### 1. Comment quantifier la ressemblance entre deux tables ?

Gène 1

	chorion	foie	gonade	cerveau
chorion	1	-0,19	0,12	-0,04
foie	-0,19	1	0,05	0,07
gonade	0,12	0,05	1	0,22
cerveau	-0,04	0,07	0,22	1

Gène 2

	chorion	foie	gonade	cerveau
chorion	1	-0,23	0,01	0,52
foie	-0,23	1	-0,11	0,1
gonade	0,01	-0,11	1	-0,09
cerveau	0,52	0,1	-0,09	1

Gène 16602

	chorion	foie	gonade	cerveau
chorion	1	0,52	0,18	0,03
foie	0,52	1	0,01	0,17
gonade	0,18	0,01	1	0,3
cerveau	0,03	0,17	0,3	1

Corrélations moyennes entre tables  
(RV)

	cerveau	chorion	foie	gonade
cerveau	<b>1.00</b>	<b>0.09</b>	<b>0.12</b>	<b>0.14</b>
chorion	<b>0.09</b>	<b>1.00</b>	<b>0.12</b>	<b>0.06</b>
foie	<b>0.12</b>	<b>0.12</b>	<b>1.00</b>	<b>0.06</b>
gonade	<b>0.14</b>	<b>0.06</b>	<b>0.06</b>	<b>1.00</b>

## PTA : Détail des calculs.6

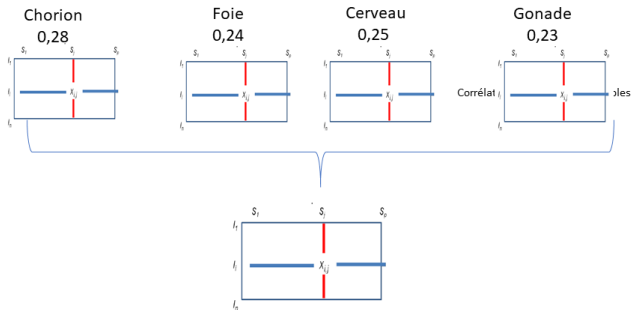
Compromis (moyenne pondérée des tables)

### Analyse Triadique Partielle

Etablissement d'un compromis: table moyenne

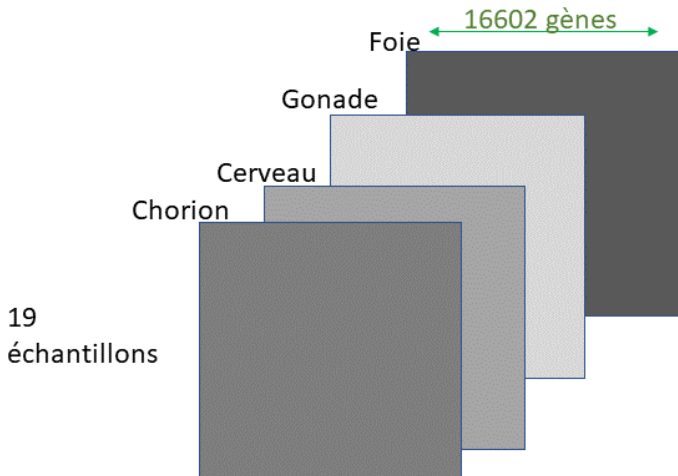
	cerveau	chorion	foie	gonade
cerveau	1.00	0.09	0.12	0.14
chorion	0.09	1.00	0.12	0.06
foie	0.12	0.12	1.00	0.06
gonade	0.14	0.06	0.06	1.00

Moyenne pondérée : plus les tables se ressemblent, plus leur poids dans la moyenne est importante



## Un exemple: Transcriptome sur 4 tissus

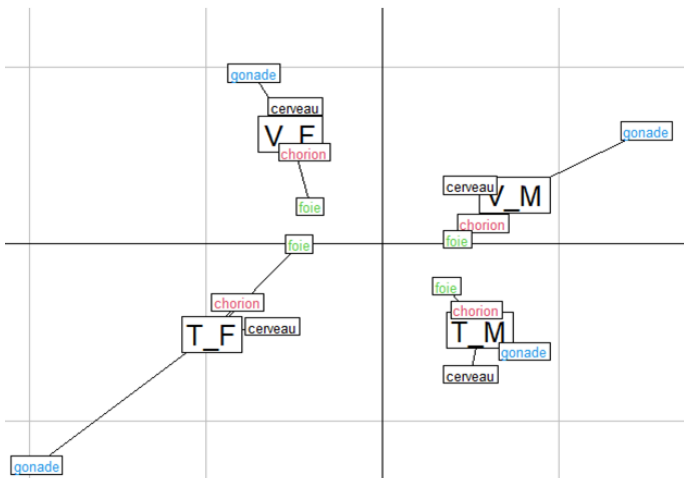
Projet BoSexDim (ANR/Apis-Gene) - Transcriptomes de 19 embryons bovins (Sexe / type de fécondation)





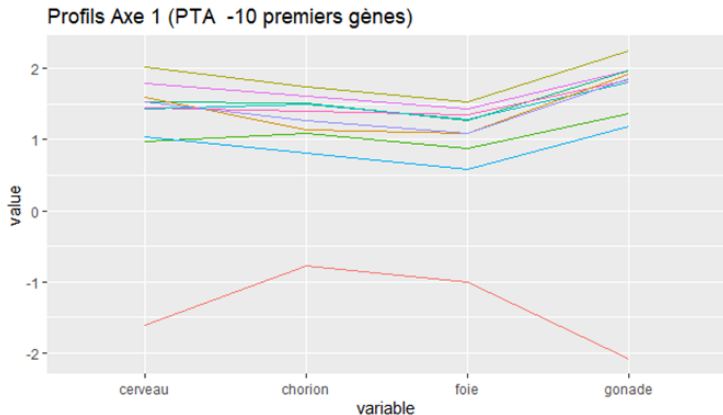
## Un exemple: Transcriptome sur 4 tissus

Compromis et points partiels des combinaisons Sexe\*Type de fécondation



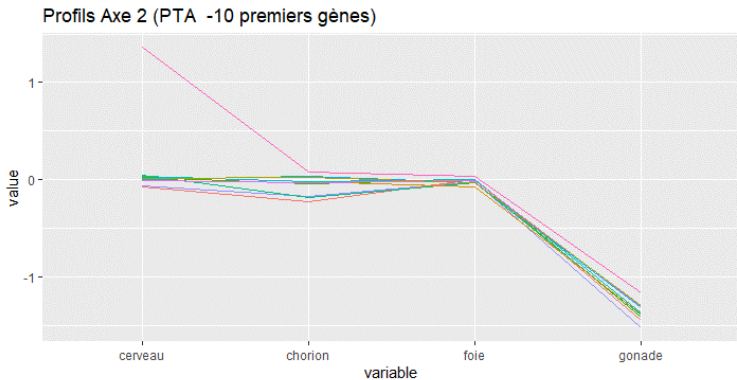
## Un exemple: Transcriptome sur 4 tissus

Expression des gènes "importants" sur l'axe 1 (Sexe)



## Un exemple: Transcriptome sur 4 tissus

Expression des gènes "importants" sur l'axe 2 (Type de fécondation)



## Retour sur la PTA

- ▶ Corrélations négatives entre tables ? Quid du compromis ?
- ▶ "Partiel": Choix de la dimension répétée
  - ▶ Il existe des méthodes 3-blocs (PARAFAC, tenseurs, ... )
  - ▶ Centrage
- ▶ Structure contrainte: pas de données manquantes
- ▶ Alternative : RDA avec un facteur "répétition"

## Conclusion : Approches statistiques Descriptive vs Inférentielle

### Description

- ▶ Sans modélisation / non supervisée
- ▶ Symétrie
- ▶ Pas de différence entre expliquée et explicative
- ▶ Pas d'inversion de matrice
- ▶ Pas de contraintes sur le nombre de variables

### Inférence

- ▶ Modélisation / Supervisée
- ▶ Asymétrie
- ▶ variables à expliquer vs variables explicatives
- ▶ Inversion de matrices
- ▶ Nombre d'observations  $\geq$  nombre de variables
- ▶ Etapes de régularisation

## Modernité de l'approche

- ▶ Structures variées
- ▶ Variables peu ou pas caractérisées
- ▶  $p \gg n$
- ▶ Pas ou peu d'hypothèses de distribution
- ▶ Possibilité de mixer quantitatif et qualitatif
- ▶ Accent sur la visualisation
- ▶ Big Data / Informatique

## Perspectives ?

### Modélisation de dépendances entre variables

- ▶ Dépendances spatiales entre variables
- ▶ Dépendances spatiales (ATAC-Seq / RNA-Seq) entre variables de blocs différents
- ▶ Dépendances temporelles : AFM ou PTA temporelle; autocorrélations

## Références additionnelles / Extensions R

- ▶ Dray, S. et Dufour, A-B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4).
- ▶ Lebart, L., Piron, M, Morineau, A. (2006). Statistique exploratoire multidimensionnelle. *Dunod*.
- ▶ Legendre P., Legendre L. (2012). Numerical ecology. *Elsevier*.
- ▶ Le Roux B., Rouanet, H, 2004. Geometric Data Analysis. *Kluwer*
  
- ▶ ade4. <http://pbil.univ-lyon1.fr/ade4/>
- ▶ FactomineR <http://factominer.free.fr/>
- ▶ vegan <http://cran.r-project.org/web/packages/vegan/>