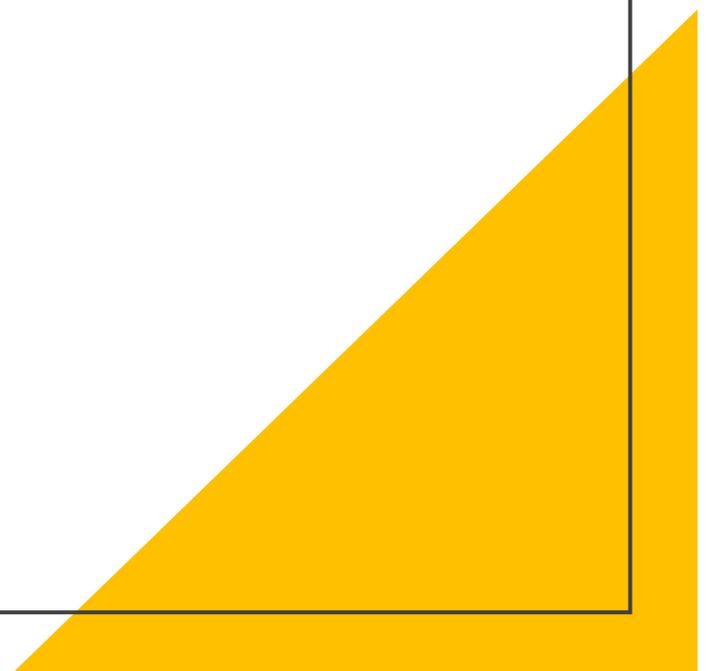


La statistique descriptive multidimensionnelle en histoire de la pensée économique

Julien Gradoz — Université de Lille



Ressources utilisées dans ce cours 1/2

- [Statistique descriptive multidimensionnelle](#) — Alain Baccini
- [Méthodes factorielles de représentation et de discrimination](#) — Lise Vaudor
- [Analyse de données avec R](#) — François Husson, Sébastien Lê, Jérôme Pagès

B

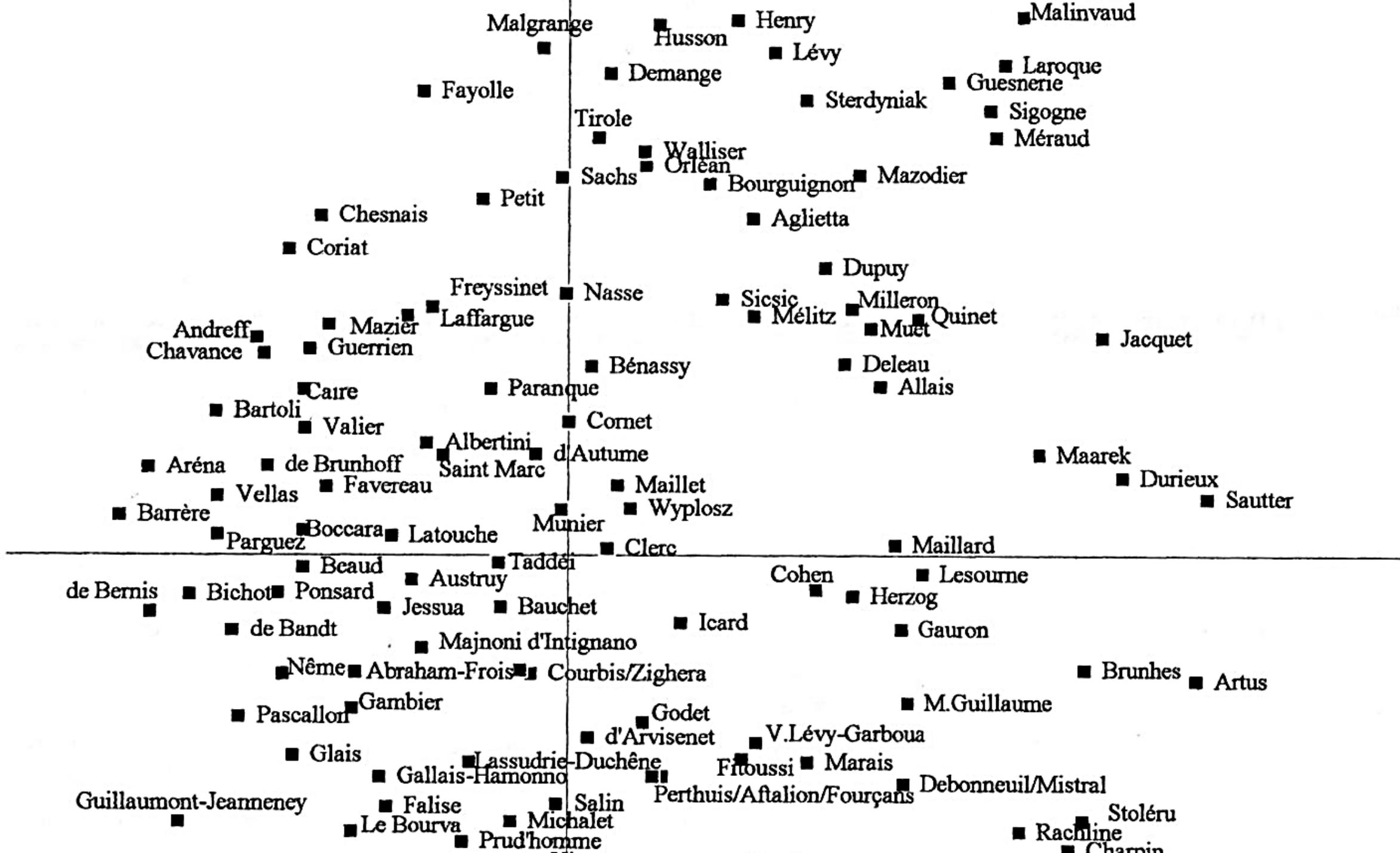
IV



Ressources utilisées dans ce cours 2/2

- [Analyse en Composantes Principales](#) — Marie Chavent
- [Analyses factorielles simples](#) — Xavier Bry





Appel pour sortir de la pensée unique



■ Père employé, ouvrier

■ Père Nsp

Economistes non-conformistes

Comité dir. AFSE

Carr. univ. seule

Autre univ. RP

Autre Paris

■ Conseiller scientifique

Titulaire/AFSE

■ Doctorat

Médaille CNRS

Dir., sous-dir./Prix

■ Autre administration académique internationale

■ Père agric., comm.

Père professions

Femme

■ Univ. Province

■ Province

■ Père cadre moyen

Paris X

■ SSCI=1 à 10

médicales

Paris I-Paris II

Institutions internationales

■ Père ingénieur

■ Auteur article Le Monde

■ Agrégation

Banque de France

■ Père enseignant

■ Etudes scientifiques autres

■ Professeur associé

■ CNU 1996

■ SSCI=0

■ Père haut fonc.

Paris bourgeois

■ DES et <

■ Père cad. sup. privé

■ Association financière

Diplôme Etats-Unis

■ Société d'économie politique

Etudes lettres-droit-sc. po.

Carr.

Who's who

■ Cab. gauche

■ Prix Nouvel Economiste

Elu local

politique (+univ/adm)

■ Père officier

■ Père fonc. sup.

Election, prix Institut

Bottin mondain

juridiques

■ Chroniqueur presse

■ Autre association internationale

■ Cadre dirigeant

IEP Paris

■ Père patron

■ Enseignant IEP

■ Elu national

Introduction 1/2

B

- Dans cette séance, nous allons présenter une technique de *statistique descriptive multidimensionnelle*, et montrer ses applications en histoire de la pensée économique.
- On désigne par *statistique descriptive multidimensionnelle* l'ensemble des méthodes de la statistique descriptive (ou exploratoire) permettant de traiter *simultanément* un nombre *quelconque* de variables.

Introduction 2/2

B

- Ces méthodes sont purement *descriptives*, c'est-à-dire qu'elles ne supposent, a priori, aucun modèle sous-jacent, de type probabiliste.
- Par exemple, lorsqu'on considère des variables quantitatives, il n'est pas nécessaire de supposer que ces variables sont distribuées selon des lois normales.

La statistique descriptive multidimensionnelle en histoire de la pensée économique 1/2

- De plus en plus de travaux en histoire de la pensée économique recourent à ces techniques. Quelques exemples :
 - Lebaron ([1997](#)).
 - Numéro spécial de la revue *Historical Social Research* ([2018](#)).
 - Cherrier et Saïdi ([2018](#)).
 - Tusset ([2021](#)).
 - Rossier et Benz ([2021](#)).

La statistique descriptive multidimensionnelle en histoire de la pensée économique 2/2

- Quelques exemples de problématiques auxquelles ces méthodes permettent de répondre :
 1. Quelles sont les spécificités des *business schools* dans la formation des économistes aux USA ?
 2. Quelles sont les principales divergences entre les économistes français vis-à-vis des principaux thèmes de politique publique ?
 3. Les experts à la Commission européenne : un groupe social homogène ?

L'enjeu de la visualisation 1/2

B

- Supposons que l'on s'intéresse à une base de données comprenant 2 ou 3 variables quantitatives.
- Dans ce cas, il est facile de visualiser le *nuage de points* de cette base de données.
- En revanche, dès que l'on s'intéresse à une base de données comprenant plus de trois variables, il est difficile de comprendre ce qu'il se passe, car il est impossible de « visualiser » le nuage de points.

L'enjeu de la visualisation 2/2

B

- *L'analyse en composantes principales (ACP)* permet de *rabattre* un nuage de points de dimension $N > 3$ dans un espace de dimension réduite (par exemple 2), en déformant le moins possible l'information initiale contenue dans le nuage de points.
 - Dit autrement, l'ACP est une méthode visant à obtenir le résumé le plus *pertinent* possible des données initiales, tout en permettant de les représenter dans un espace de dimension réduite.
- Encore faut-il définir un critère de pertinence !

Le rôle de l'inertie dans l'ACP 1/2

IV

- Le critère de pertinence à partir duquel construire l'ACP est *l'inertie restituée* du nuage de points.
- L'inertie d'un nuage de point correspond à la somme des distances au carré de chacun des points du nuage par rapport au centre du nuage. Elle s'interprète donc comme « l'étalement » du nuage de points.

Le rôle de l'inertie dans l'ACP 2/2

IV

L'objectif fondamental de l'ACP est de faire passer des « axes » dans le nuage de point initial, permettant de réaliser des projections de ce nuage de points dans un espace de dimension réduite, tout en sachant que ces axes sont construits de telle sorte à ce que la projection obtenue restitue le maximum d'inertie initiale du nuage de points.

Métaphore de la « prise de vue » 1/4



- Pour comprendre le fonctionnement de l'ACP, nous pouvons utiliser la métaphore de la *photographie d'un fruit*.
- En effet, une photographie revient fondamentalement à projeter dans un espace de dimension 2 (la photographie) un objet initialement en dimension 3.

Métaphore de la « prise de vue » 2/4



- Sur l'image ci-dessous, des fruits ont tout d'abord été photographiés du dessus, puis de côté.

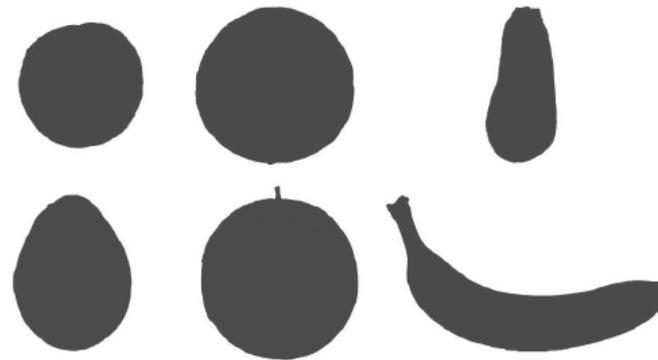


FIGURE 1.5

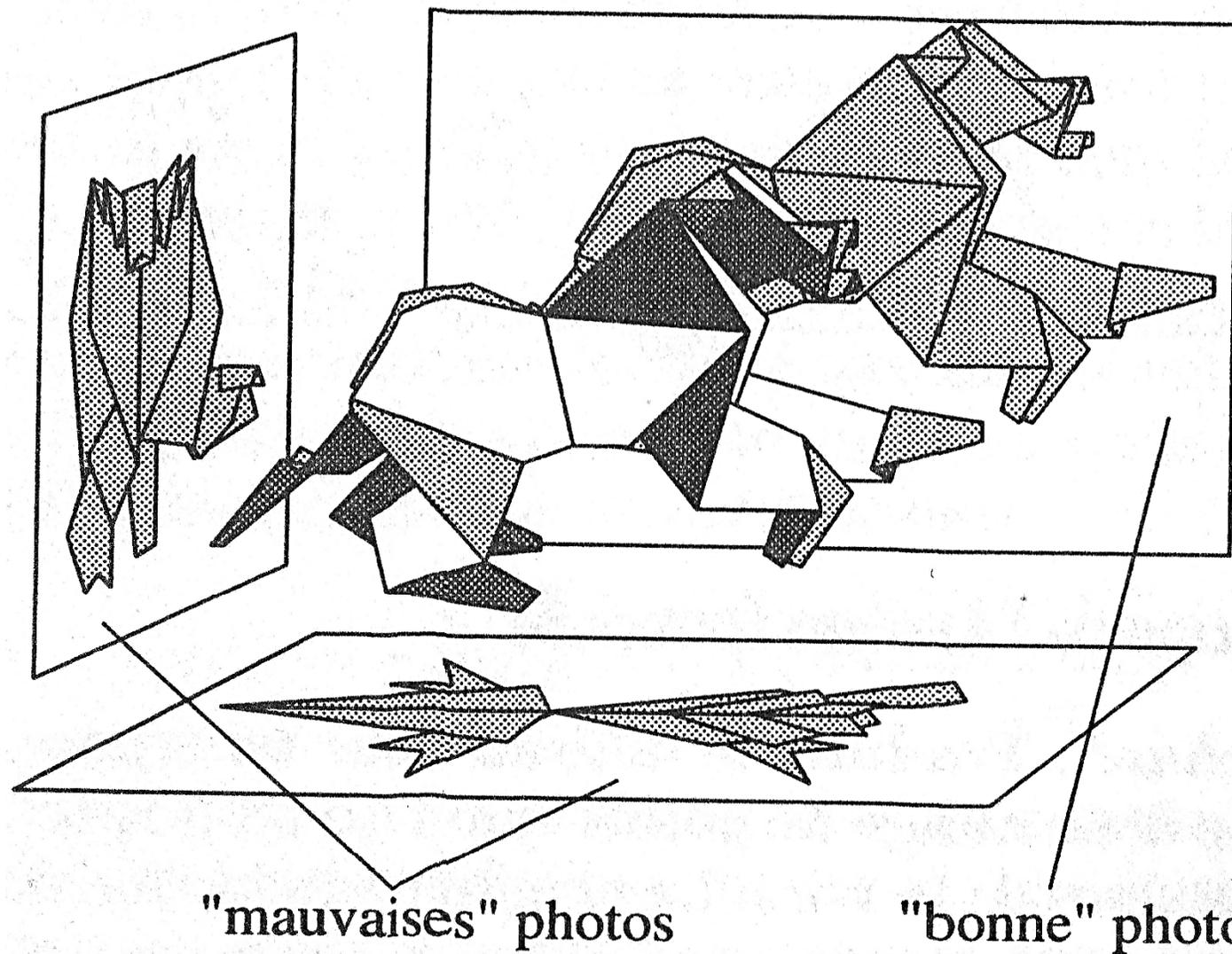
Two-dimensional representations of fruits: from left to right, an avocado, a melon, and a banana; each row corresponds to a different representation.

Métaphore de la « prise de vue » 3/4



- Nous constatons que certaines « prises de vue » déforment moins la forme initiale du fruit que d'autres, et qu'elles permettent donc de mieux restituer la forme initiale du fruit en trois dimensions.
- Avec l'ACP, c'est exactement cette idée : il s'agit de trouver la « prise de vue » qui permettra de restituer au mieux l'inertie initiale du nuage de points.

Métaphore de la « prise de vue » 4/4

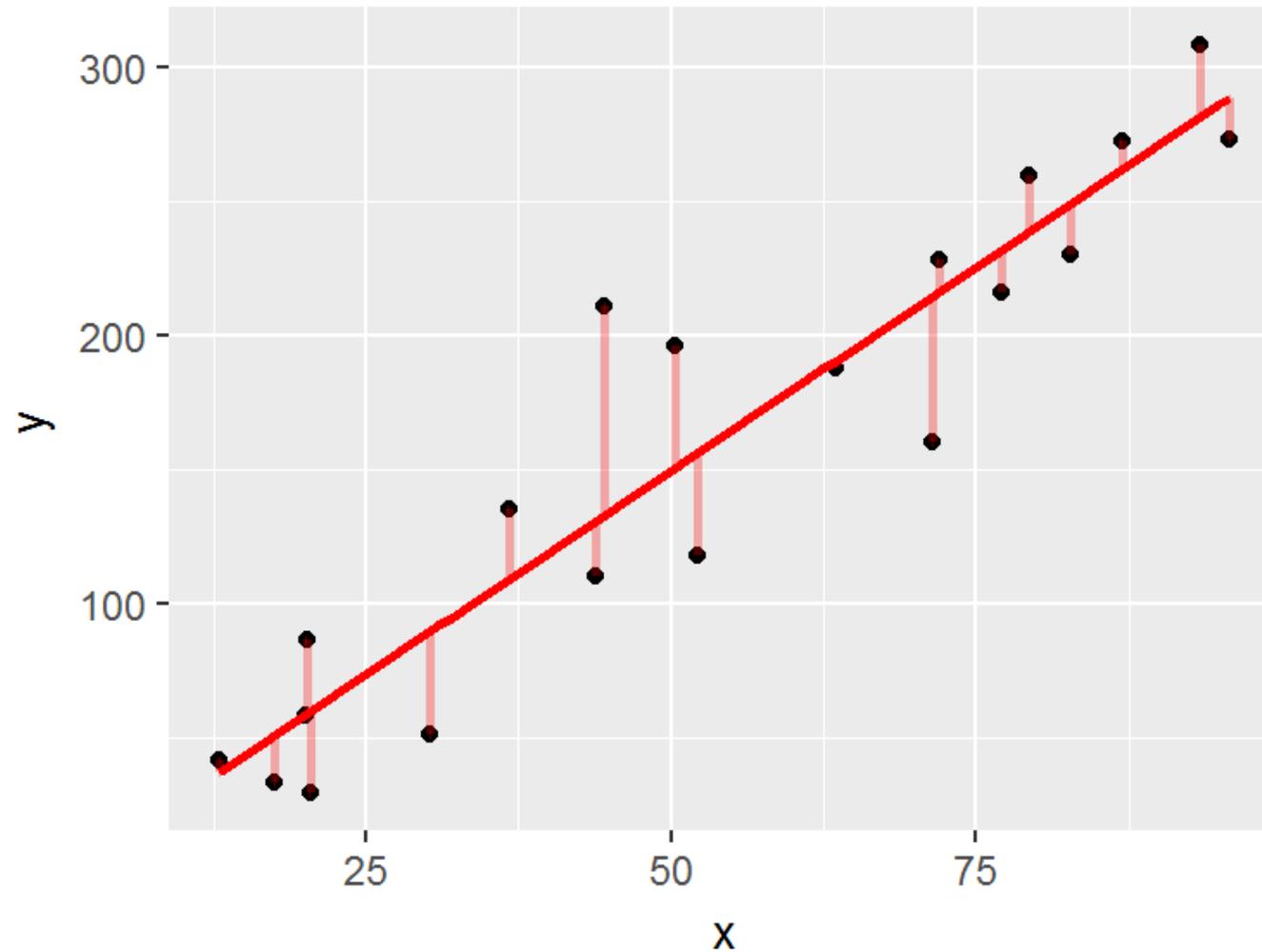


Métaphore de la régression linéaire 1/3

- On peut aussi comprendre le fonctionnement de l'ACP en s'appuyant sur la *droite de régression linéaire*.
- En effet, une droite de régression linéaire cherche à restituer le mieux possible la « tendance » d'un nuage de points en deux dimensions.
- Pour ce faire, la droite de régression linéaire va *projeter* un nuage de points en deux dimensions le long d'une droite de dimension 1.

Métaphore de la régression linéaire 2/3

IV



Métaphore de la régression linéaire 3/3

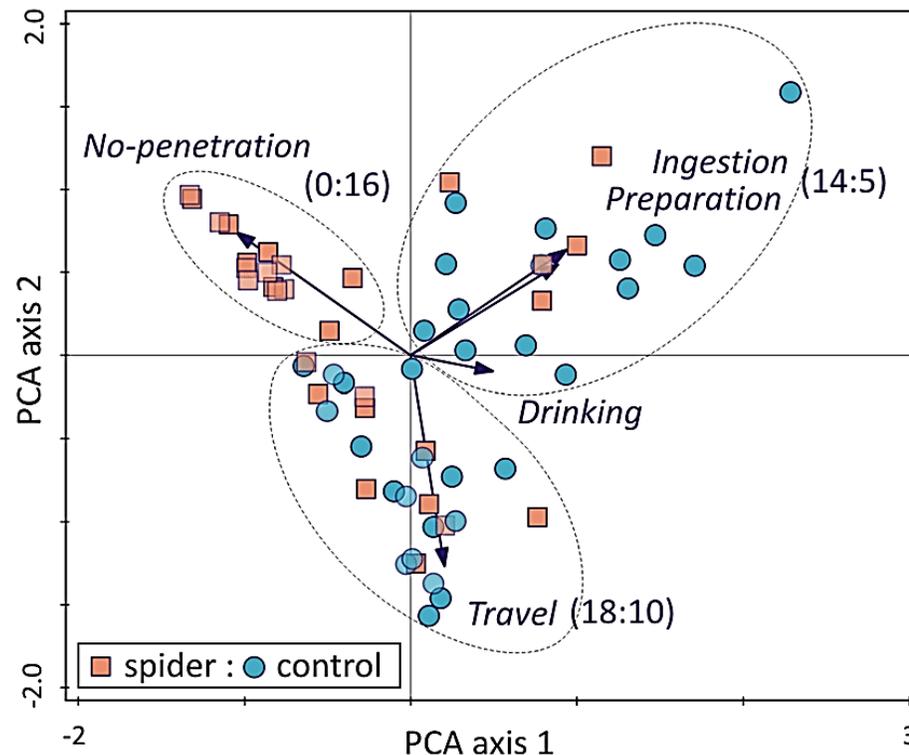
- La projection des points le long de la droite de régression linéaire respecte le critère des *moindres carrés ordinaires*.



- Si on réalisait une ACP dans un nuage de points de dimension 2, nous pourrions également obtenir une projection de nos points le long d'une droite. Seulement, le positionnement des points respecterait désormais le critère de *maximisation de l'inertie restituée*.

Le nombre d'axes d'une ACP 1/2

- Quand on réalise une ACP, on projette généralement le nuage de points initial dans un plan en deux dimensions, le *plan factoriel*.



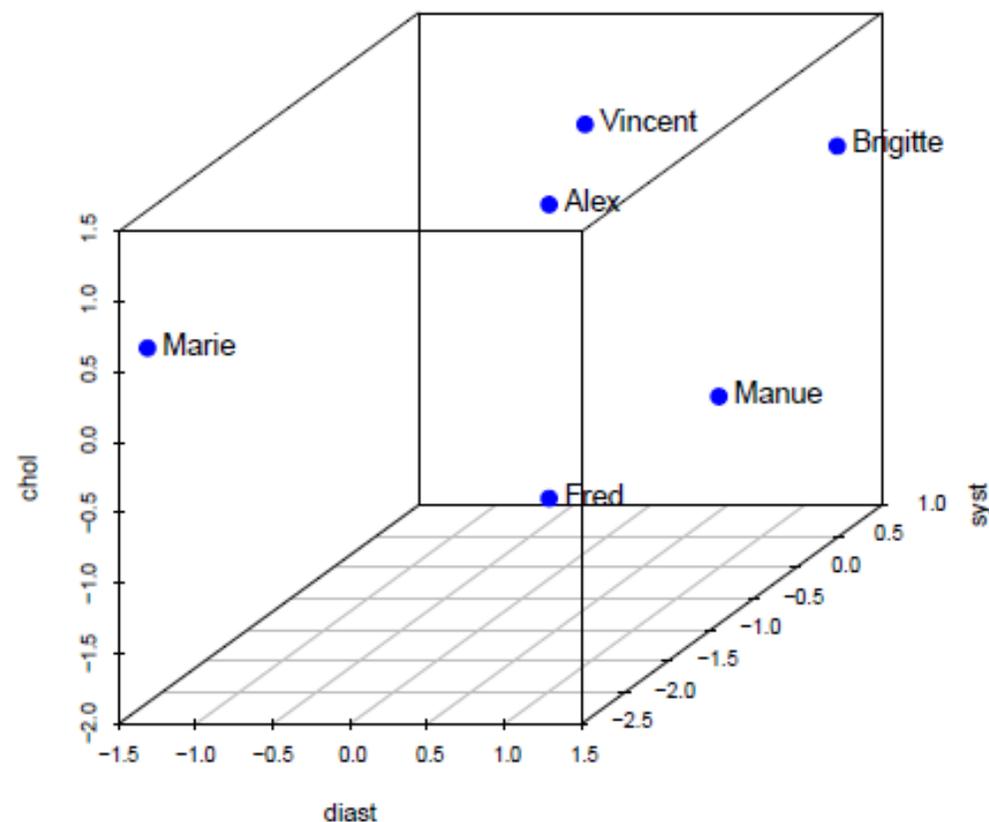
Le nombre d'axes d'une ACP 2/2

- Le graphique précédent est la projection d'un nuage de points de dimension $N > 3$ dans un espace de dimension 2.
- Le « plan factoriel » est constitué de deux axes factoriels, et ces deux axes factoriels sont construits de sorte à restituer le maximum d'inertie initiale du nuage de points de dimension $N > 3$.

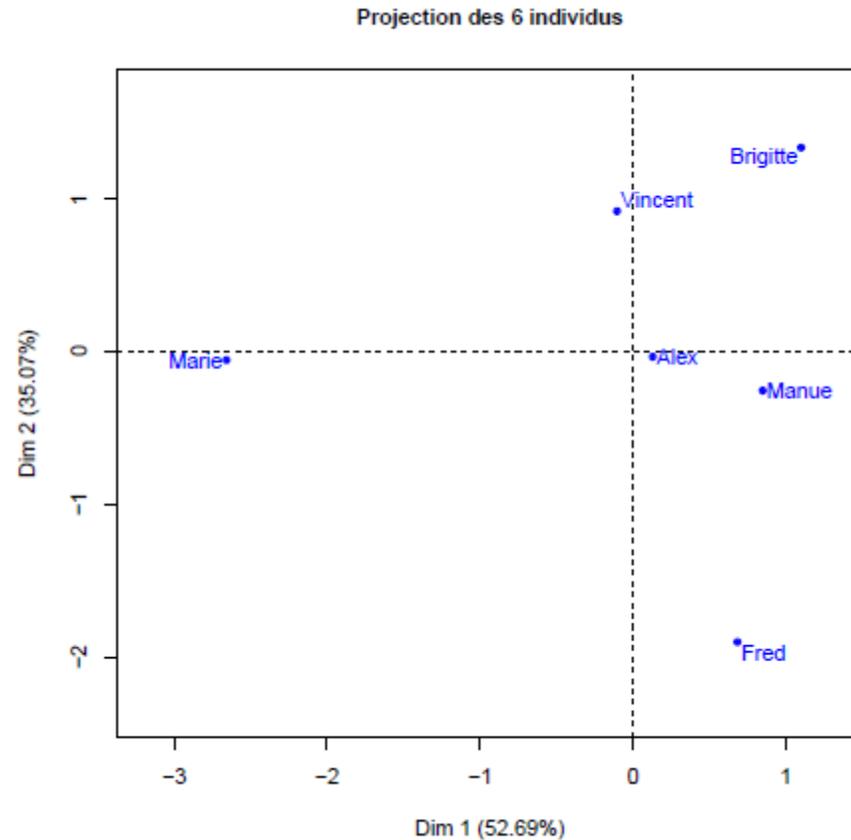
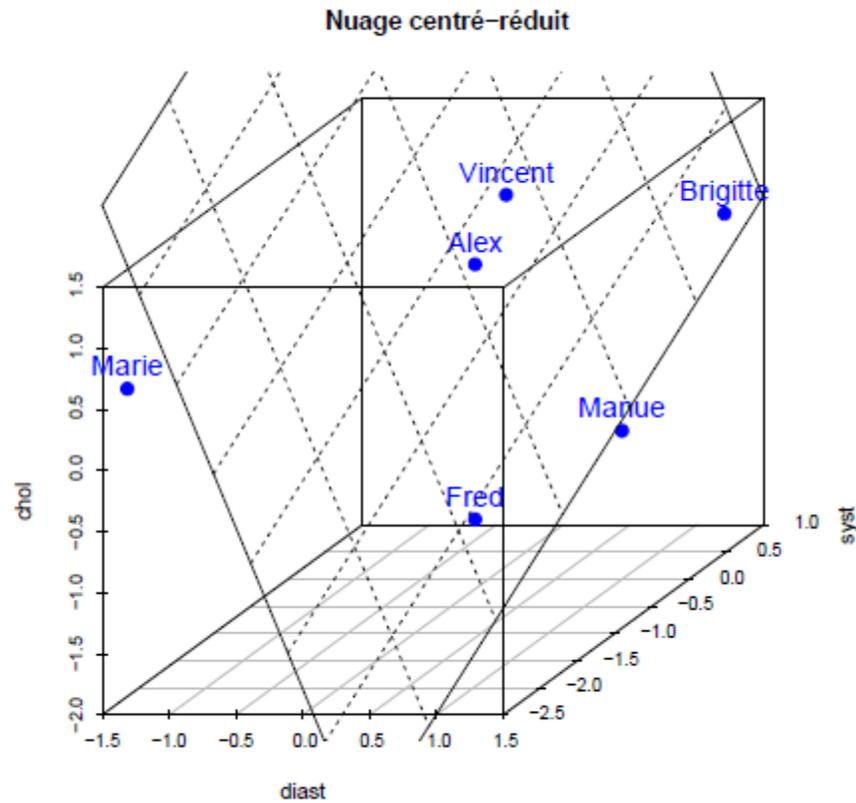
Exemple : si on voulait passer de la dimension 3 à la dimension 2



Nuage centré-réduit des 6 individus



Exemple : si on voulait passer de la dimension 3 à la dimension 2



On construit le plan (*axe factoriel 1, axe factoriel 2*) qui restitue le maximum d'inertie initiale du nuage de points en trois dimensions

Le plan factoriel et interprétation 1/3

- Dans le graphique précédent, nous avons une base de données avec trois variables quantitatives.
- Nous cherchons à faire passer un « plan factoriel » dans le nuage de points, constitué de deux axes orthogonaux, et qui restitue le maximum d'inertie initiale de ce nuage. Dans la figure de droite, nous obtenons une « projection du nuage de points dans le plan factoriel ».

Le plan factoriel et interprétation 2/3

- La représentation du nuage de points dans un plan factoriel est la démarche traditionnelle de l'ACP.
- Par ce procédé, nous obtenons un *nouveau nuage de points*, différent du premier. En effet, ce nuage de points ne se situe plus dans le plan (x,y,z) , mais dans le plan (*axe factoriel 1, axe factoriel 2*).

Le plan factoriel et interprétation 3/3

- Par exemple, nous passons d'un nuage de points dans le repère (*poids, taille, âge*) à un nuage de points dans le repère (*axe factoriel 1, axe factoriel 2*).
- Dès lors, la question fondamentale est : ***comment interpréter ce nouveau nuage de points ?***

L'ACM 1/6

- Avant de répondre à cette question, nous allons présenter la méthode sur laquelle nous allons travailler durant cette séance, *l'analyse des correspondances multiples (ACM)*.
- Cette méthode est un cas particulier de l'ACP, où les données initiales sont des variables qualitatives.

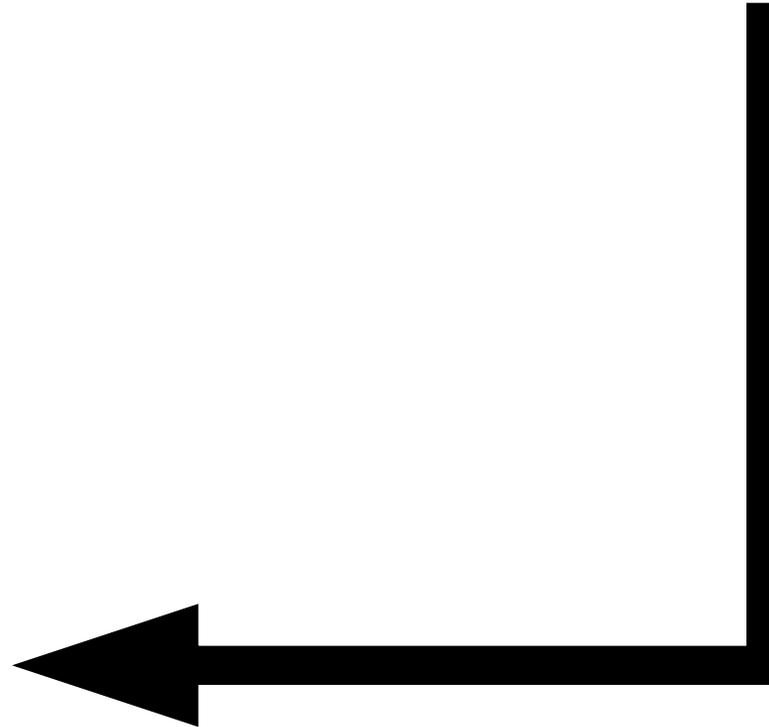
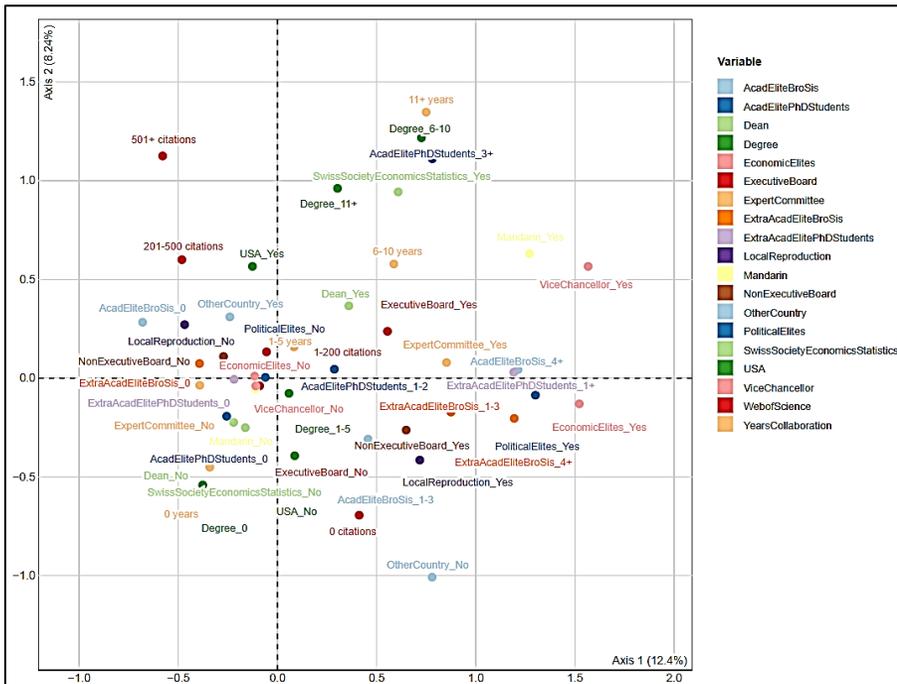
L'ACM 2/6

- L'ACM permet de transformer la base de données qualitative afin de la rendre compatible avec la création d'une ACP.
- L'ACM permet donc d'associer un nuage de points à une base de données qualitative, mais également de résumer l'information contenue dans cette base de données, qui est multidimensionnelle.

	PoliticalElites	ExpertCommittee	EconomicElites	ViceChancellor	Dean	Mandarin	SwissSocietyEconomicsStatistics	NonExecutiveBoard
1	No	No	No	No	Yes	No	No	No
2	No	Yes	No	No	No	No	No	No
3	No	No	No	No	Yes	No	No	No
4	Yes	No	Yes	No	Yes	No	No	Yes
5	No	No	No	No	No	No	No	No
6	No	No	No	No	No	No	No	No
7	No	No	No	No	No	No	No	No
8	No	No	No	No	No	No	Yes	No
9	No	No	No	No	No	No	No	No
10	Yes	Yes	Yes	No	No	No	No	Yes
11	No	No	No	No	Yes	No	No	No
12	No	No	No	No	Yes	No	No	No
13	No	No	No	No	Yes	No	No	No
14	No	No	No	No	No	No	No	Yes
15	No	No	No	No	No	No	No	No
16	No	Yes	No	No	No	No	No	No
17	No	No	No	No	Yes	No	No	No
18	No	No	No	No	No	Yes	No	Yes
19	No	No	No	No	Yes	Yes	No	Yes
20	Yes	Yes	Yes	No	No	No	Yes	Yes
21	No	No	No	No	No	No	No	Yes



Transformation adéquate des données (*tableau disjonctif complet*)



Réalisation d'une ACP permettant de projeter cette base de données qualitative dans un plan factoriel

L'ACM 3/6

- Les modalités proches signifient que des individus prennent souvent simultanément ces modalités dans la base de données, tandis que les modalités opposées sont rarement prises simultanément par des individus.
- Il est donc possible d'interpréter les axes à partir des oppositions et des rapprochements entre les modalités de la base de données.

L'ACM 4/6

- Notons qu'il est également possible de projeter des modalités qui ne participent pas à la construction de l'ACM, mais qui permettent néanmoins d'apporter des informations complémentaires.
- Cette approche est très utilisée en sociologie, afin de distinguer les variables qui concernent une thématique particulière et les variables qui concernent les caractéristiques sociodémographiques des individus.

L'ACM 5/6

- Avec cette méthode, il est possible de générer une ACM sur une thématique particulière, qui s'interprète alors comme un « champ » : *le champ des loisirs, le champ du cinéma, le champ des économistes...*
- Une fois le champ généré, on peut lui superposer un « espace social » généré à partir des caractéristiques sociodémographiques des individus. Dès lors, la position dans le « champ » peut être mise en rapport avec la position dans l'espace social des individus.

L'ACM 6/6

- Enfin, au-delà de représenter les modalités dans le plan factoriel, il est également possible de représenter les individus, qui sont proches des modalités qu'ils possèdent, et éloignés des modalités qu'ils ne possèdent pas.
- Pour comprendre cette idée, nous allons reproduire l'ACM d'un chapitre d'ouvrage de Rossier et Benz (2021).