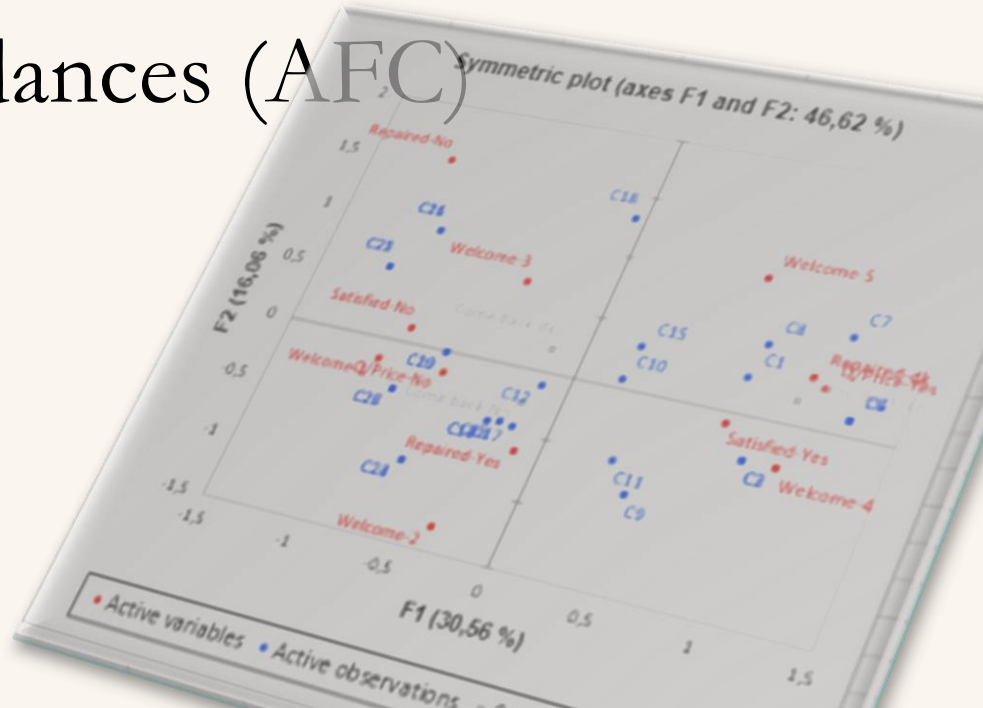
Statistique Inférentielle
Chapitre 23 : Test de corrélation

Chapitre 23 : Test de corrélation

- Régression simple
- Analyse en Composantes Principales (ACP)
- Analyse des Correspondances Multiples (ACM)
- Analyse Factorielle des Correspondances (AFC)



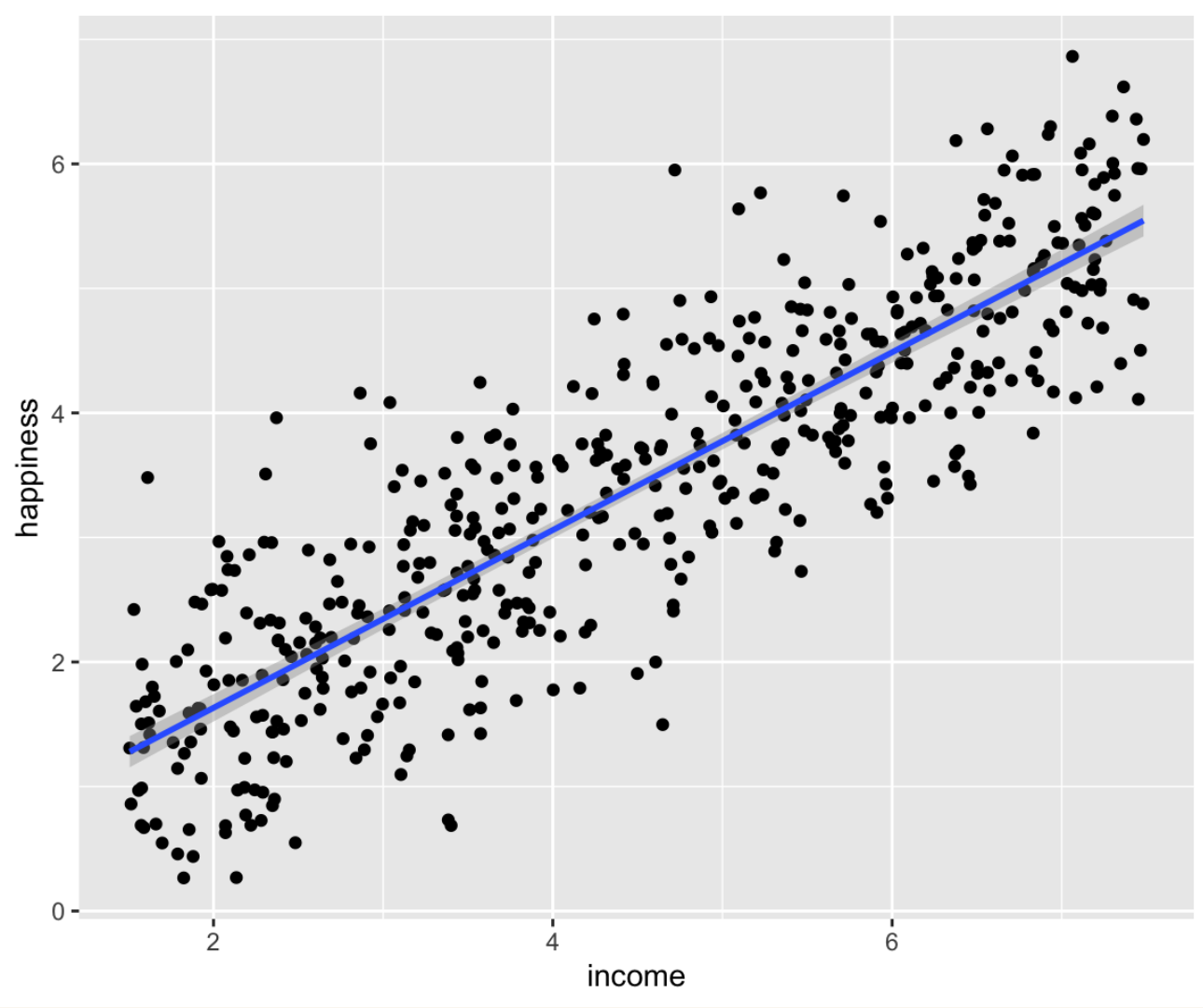
Régression simple

Rappel

cf chapitre 02 "statistiques bivariées"

droite de régression …

coefficient de correlation …

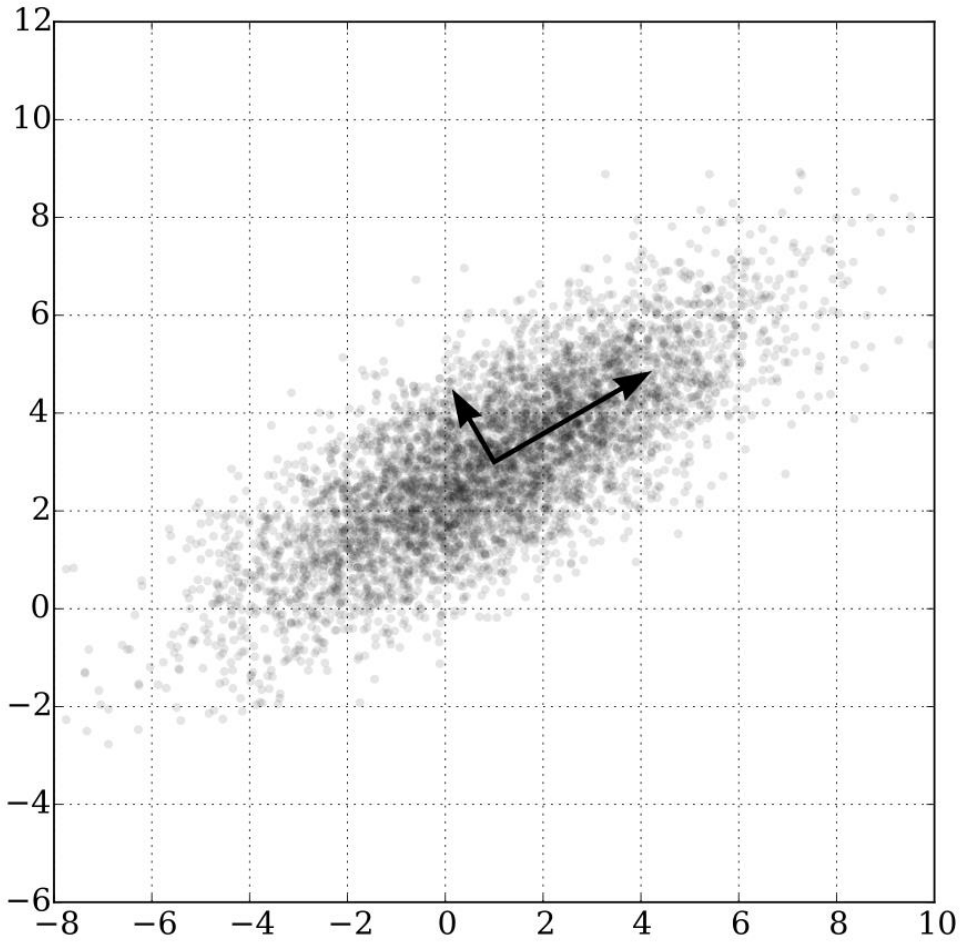


Analyse en Composantes Principales

ACP

ACP : Principe

- Transformer des variables corrélées en nouvelles variables décorrélées les unes des autres.
- Ces nouvelles variables sont nommées "composantes principales". Elle résument l'information en réduisant le nombre de variables.



ACP : Exemple 1 (simplifié)

base	math	phys	fran	angl
jean	6,0	6,0	5,0	5,5
alan	8,0	8,0	8,0	8,0
anni	6,0	7,0	11,0	9,5
moni	14,5	14,5	15,5	15,0
didi	14,0	14,0	12,0	12,5
andr	11,0	10,0	5,5	7,0
pier	5,5	7,0	14,0	11,5
brig	13,0	12,5	8,5	9,5
evel	9,0	9,5	12,5	12,0

analyse bivariée	math	phys	fran	angl
math		0,98	0,23	0,51
phys			0,40	0,65
fran				0,95
angl				

Conclusions intuitives :

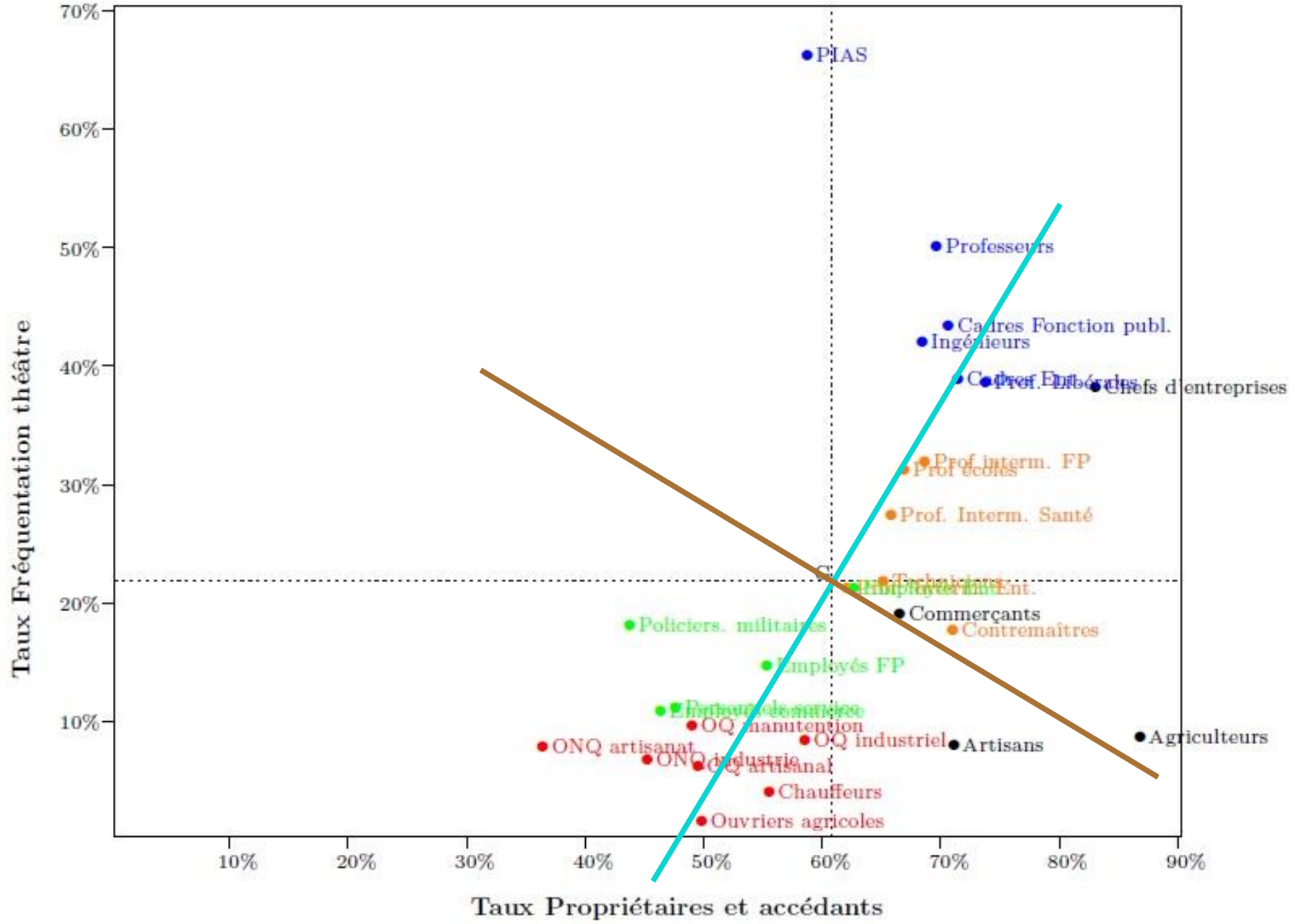
1. Il y a 2 composants principaux
   - math-phys & fran-angl
   - matières scientifiques vs littéraires
2. Il y a des composants secondaires
   - phys-angl, math-angl, …

ACP : Exemple 2

- Catégories socio-professionnelles avec
  - le taux de propriétaires
  - la proportion de membres allant au théâtre
- Indicateurs statistiques
  - moyenne
  - variance
  - variance plus élevée dans la 2e colonne
- Sources
  - INSEE
  - https://www.politika.io/fr/article/analyse-correspondances-multiples

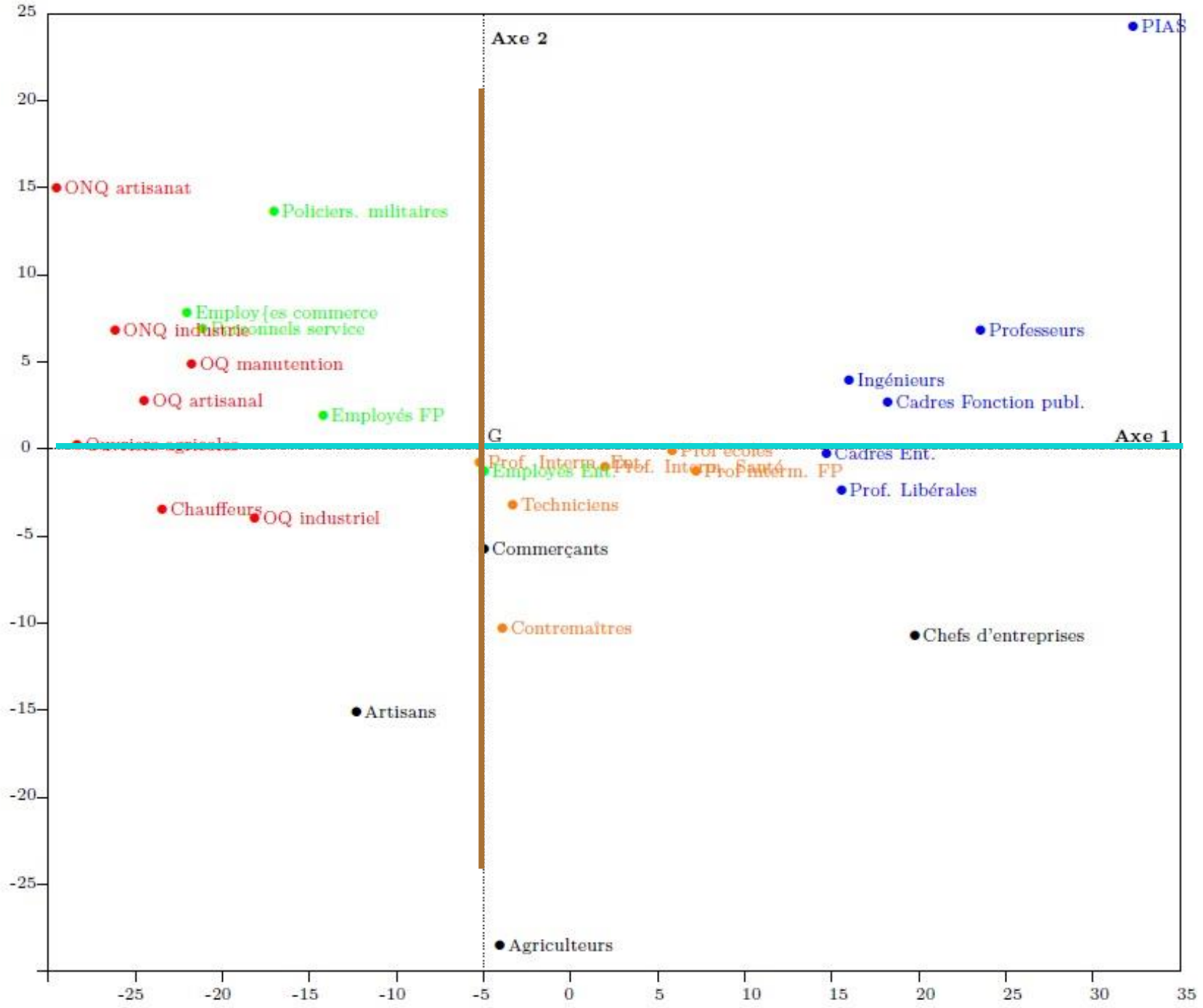
	Taux de propriétaires ou accédants (%)	Fréquentation théâtre (%)
Agriculteurs	85,9	7,8
Artisans	70,2	7,2
Commerçants et assimilés	65,6	18,2
Chefs d'entreprise de 10 salariés ou plus	82,1	37,3
Professions libérales	72,8	37,7
Cadres de la fonction publique	69,7	42,5
Professeurs, professions scientifiques	68,7	49,2
Professions de l'information, des arts et des spectacles	57,8	65,4
Cadres administratifs et commerciaux d'entreprise	70,5	38,0
Ingénieurs et cadres techniques d'entreprise	67,5	41,2
Professeurs des écoles, instituteurs et assimilés	66,0	30,4
Professions intermédiaires de la santé et du travail social	64,9	26,6
Professions intermédiaires administratives de la fonction publique	67,7	31,0
Professions intermédiaires administratives et commerciales	61,2	20,4
Techniciens	64,2	20,9
Contremaîtres, agents de maîtrise	70,1	16,9
Employés civils et agents de service de la fonction publique	54,4	13,9
Policiers et militaires	42,8	17,2
Employés administratifs d'entreprise	61,7	20,4
Employés de commerce	45,4	10,0
Personnels des services directs aux particuliers	46,7	10,3
Ouvriers qualifiés de type industriel	57,6	7,6
Ouvriers qualifiés de type artisanal	48,6	5,4
Chauffeurs	54,6	3,2
Ouvriers qualifiés de la manutention, du magasinage	48,1	8,8
Ouvriers non qualifiés de type industriel	44,3	5,9
Ouvriers non qualifiés de type artisanal	35,5	7,0
Ouvriers agricoles	48,9	0,8
Moyenne	*60,5*	*21,5*
Variance	*143,9*	*253,5*

ACP : Exemple 2



ACP : Exemple 2

Les repères ont été transformés :
Rotation du nuage autour du point G pour le rendre horizontal.



ACP : Exemple 3 : énoncé

- Données
  - températures moyennes mensuelles de différentes villes d'un pays
- On veut résumer le climat du pays.
- Les villes les plus chaudes en janvier le sont aussi en juillet. Les variables sont donc corrélées positivement entre elles.
- L'ACP mettra en évidence un premier axe opposant les villes chaudes aux villes froides.
  Mais …
- Il existe des différences entre les villes dans l'écart entre les températures hivernales et estivales.
  - La présence de la montagne peut accentuer les écarts entre l'été et l'hiver.
  - La proximité de la mer peut adoucir ces écarts.
- L'ACP dégagera alors un second axe opposant les villes selon l'amplitude des températures.
  - climat océanique ou continental

ACP : Exemple 3 : traitement

- `acp-complément.xlsx`
  - certains traitements et graphiques réalisés sous tableur
- `acp-input.csv`
  - fichier de données pour R
  - 23 villes françaises, 12 mois
- `acp-main.R`
  - traitement automatisé complet, avec graphique
  - adapter seulement les paramètres
- `acp-output.png`
  - graphique généré par R

Analyse des Correspondances Multiples

ACM

ACM : Principe

- Méthode destinée à étudier l'association entre au moins deux variables qualitatives.
- L'ACM est aux variables qualitatives
  ce que l'ACP est aux variables quantitatives.
- Résultat : des cartes de représentation sur lesquelles on peut visuellement observer les proximités entre les catégories des variables qualitatives et les observations.

Analyse Factorielle des Correspondances

AFC

Analyse par tableau de contingence

- Méthode d'analyse quantitative des données
- Extension du test χ^2 d'indépendance
- Très fréquent en études de marché
- Cette technique d'étude de marché permet elle aussi d'adapter son marketing et sa communication en fonction des relations observées entre les données et les variables.

coul vs pays	Belgique	France	Suisse	Pérou	**Total / coul**
Rouge	48	69	210	11	**338**
Jaune	27	79	141	16	**263**
Vert	5	28	48	5	**86**
Total / pays	**80**	**176**	**399**	**32**	**687**

AFC : Principe

- Principe
  - hiérarchiser les informations contenues dans un tableau croisé
  - deux variables qualitatives
- Histoire :
  - Années 1960 : mise au point par Jean-Paul Benzécri, Faculté des sciences de Rennes, et par Jussieu, laboratoire de statistique multidimensionnelle à Paris.
- Exemple
  - Lors d'une élection, un panel d'électeurs indiquent leur commune et leur liste préférentielle.
  - L'AFC déterminera les "attirances" et les "répulsions" entre les communes et les listes.

AFC : Exemple

- On veut déterminer s'il existe un rapport entre l'âge et la fréquence d'achat en ligne.
- On crée un tableau avec
  - En colonne : la fréquence des achats en ligne
    - 1 fois par semaine, 1 fois par mois, 1 fois par an, …
  - En ligne : les tranches d'âge
    - 20 à 35 ans, 36 à 50 ans, 51 à 65 ans, 66 ans ou plus
- Les résultats de l'enquête sont entrés dans le tableau créé au préalable.

AFC : Exemples : énoncés

- le choix d'études
  - Le choix d'études des enfants est-il corrélé au milieu socio-professionnel des parents ?
  - Si oui, dans quel sens ? Attractions et répulsions ?
- les pommes
  - La couleur des pommes mises sur le marché est-elle corrélée au pays d'origine ?
  - Si oui, dans quel sens ? Attractions et répulsions ?

AFC : Exemples : traitement

- `afc-choixdetudes.csv, afc-pomme.csv`
  - fichier de données pour R
- `afc-choixdetudes.png, afc-pomme.png`
  - graphique généré par R
- `afc-complément.xlsx`
  - certains traitements et graphiques réalisés sous tableur
- `afc-main.R`
  - traitement automatisé complet, avec graphique
  - adapter seulement les paramètres

AFC : Exercices

- la mixité sociale …
  - exercice complet, partant de données brutes réelles