

Problématisation

B. Lafon, 2016-2017

V. Visualiser, analyser, conclure

1. Du traitement à la présentation des données : les indicateurs

- a. Présentation des données en tableaux
- b. Différents types de graphiques :
histogrammes et secteurs
- c. Représentations cartographiques
- d. Graphiques divers et enrichis
- e. L'AFC et les « mappings »

a. Présentation des données en Tableaux

La présentation basique des données est le tableau croisé ou tableau de contingence, ventilant deux séries de variables. L'une est portée en lignes, l'autre en colonnes. Les cases sont alors le nombre d'occurrences.

Il peut aussi s'agir d'un tableau synthétique présentant une série de pourcentages.

Il peut aussi s'agir d'un tableau présentant des regroupements, ce qui techniquement se réalise par la fusion de cases (cellules).

Le but est de donner la présentation la plus claire possible, au besoin en utilisant certaines possibilités des logiciels utilisés

Exemples dans les diapos suivantes (commentés en cours).

Le sentiment que la France souffre des grands changements y compris de la construction européenne

La France :	Préférence partisane				
	<i>Ensemble</i>	<i>Extrême gauche</i>	<i>Gauche</i>	<i>Droite</i>	<i>Front national</i>
tire profit de la mondialisation des échanges économiques	25	16	25	32	20
en souffre	42	48	41	39	47
ni l'un, ni l'autre	33	36	34	29	33
	100%	100%	100%	100%	100%
tire profit de la construction européenne	27	18	28	33	21
en souffre	41	50	37	40	52
ni l'un, ni l'autre	32	32	35	27	27
	100%	100%	100%	100%	100%
tire profit de la mobilité croissante des travailleurs au sein de l'Union européenne	18	14	18	23	6
en souffre	41	46	41	36	48
ni l'un, ni l'autre	41	40	41	41	46
	100%	100%	100%	100%	100%

Tableau 3 : L'impact des grands changements du monde sur la France

Le parcours politique souhaité à gauche et à droite

Question : Pour vous, est-il indispensable, préférable mais sans plus ou pas vraiment important que le prochain président de la République ait été... ? (1)

	Ensemble	Sympathisants de gauche	Sympathisants UDF/UMP
Désigné par un parti politique			
- Indispensable	11	13	13
- Préférable mais sans plus	33	38	36
- Pas vraiment important	53	47	50
- Sans opinion	3	2	1
Ministre			
- Indispensable	9	6	14
- Préférable mais sans plus	41	39	49
- Pas vraiment important	49	55	36
- Sans opinion	1	0	1
Maire d'une grande ville			
- Indispensable	6	6	6
- Préférable mais sans plus	35	35	39
- Pas vraiment important	58	59	55
- Sans opinion	1	0	0
Député ou sénateur			
- Indispensable	6	6	8
- Préférable mais sans plus	38	41	39
- Pas vraiment important	54	52	51
- Sans opinion	2	1	2
Président ou patron d'un parti politique			



Attentes des apprenants vis-à-vis de la formation

Q3 APP. Parmi les raisons que je vais vous citer, pouvez-vous me dire pourquoi vous avez choisi une FOAD plutôt qu'un enseignement en présentiel ? (plusieurs réponses possibles)

Base : ensemble des apprenants		Ensemble	Formation initiale	Formation continue
AU TOTAL		%	%	%
Réponses supérieures à 2%	Pour gérer votre temps à votre convenance	64	69	60
	Par contrainte professionnelle	50	39	57
	Par contrainte géographique	43	39	48
	Par intérêt pour la technologie / le média utilisé	38	45	32
	Parce qu'il n'existe aucun équivalent dans les formations classiques	25	27	22
	Parce que vous possédiez une expérience antérieure de formation à distance	12	13	10
	Par contraintes familiales (non cité)	5	5	5
	Suite à un refus de l'université / inscription tardive (non cité)	4	9	1
	Vous n'avez pas eu le choix (non cité)	3	4	2
	Pour des raisons financières (non cité)	3	2	5
	Volonté liée aux professeurs / Fac (non cité)	3	2	5

b. Différents types de graphiques : histogrammes et secteurs

Les histogrammes ou graphiques en “bâtons” permettent de présenter des données discontinues, qu’il s’agisse de variables ou de classes (périodes, populations).

Généralement et de manière très concrète, la question se pose souvent de savoir s’il est plus pertinent de présenter ses données en histogramme ou en secteurs.

La finalité est différente :

- les histogrammes permettent de comparer des données (nombre d’occurrences ou pourcentages),
- les secteurs présentent les proportions de l’ensemble d’une population.

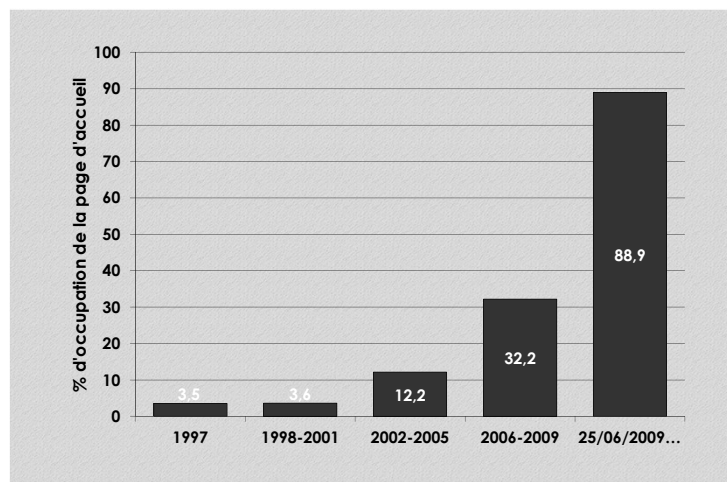
La présentation par secteurs est davantage “fermée” et met tous les éléments sur un pied d’égalité.

Pour des présentations complexes et plus souples
(personnalisables), on préférera l'histogramme.

Pour davantage de simplicité et un esprit plus synthétique, on
présentera ses données par secteur (plus rare)

Exemples dans les diapos suivantes (commentés en cours).

**Espace réservé sur la page d'accueil Ina.fr aux
services de visionnage (post 25/06/09)**

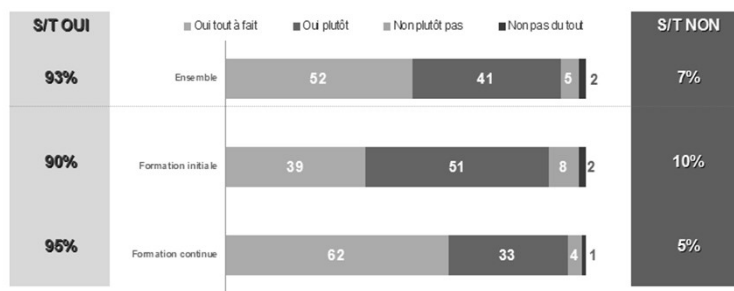


Ipsos **Campus numériques**

Adéquation entre la formation et les attentes

Q4 APP. Considérez-vous que la formation que vous suivez correspond à vos attentes et aux objectifs que vous vous étiez fixés ?

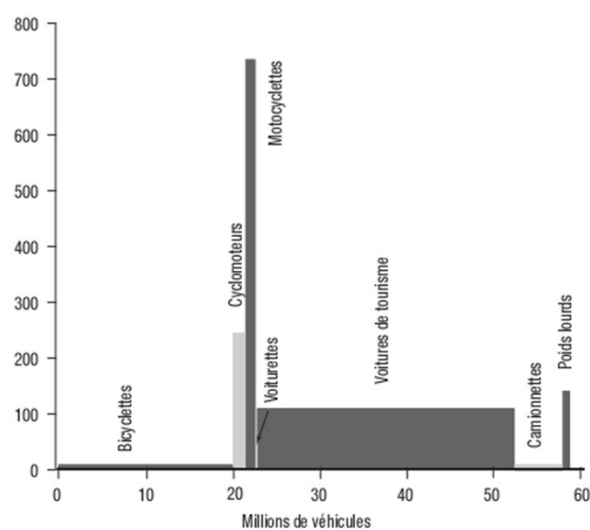
Base : ensemble des apprenants



Partie II : Efficacité des dispositifs- Motivations et satisfaction des apprenants

15

Tués par million de véhicules



Un autre type d'histogramme plus complexe existe et comporte 3 dimensions.

Un graphique en 3 dimensions permet d'ajouter un troisième axe et donc une série d'informations supplémentaires par rapport à un graphique à deux axes.

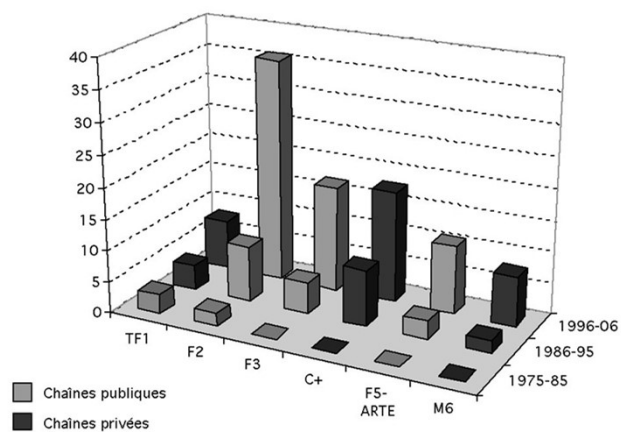
Ce type de graphique est très utile dans une optique comparative.

Les limites tiennent au grand nombre de données contenues et au fait que certaines en masquent d'autres.

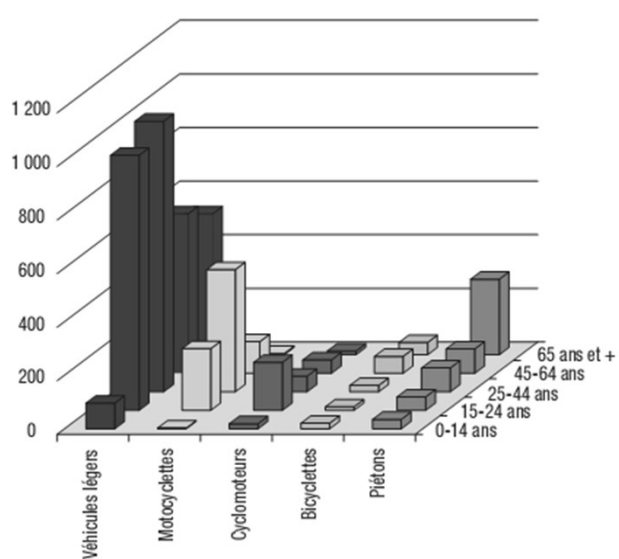
Il faut donc présenter ce type de graphique sous le meilleur angle, des logiciels tels que Excel permettant de faire des rotations dans l'espace pour parvenir à la meilleure lisibilité.

Exemples dans les diapos suivantes (commentés en cours).

Nombre de fictions sur le cancer diffusées par chaîne et par décennie



Tués par catégories d'usagers en fonction de l'âge



c. Représentations cartographiques

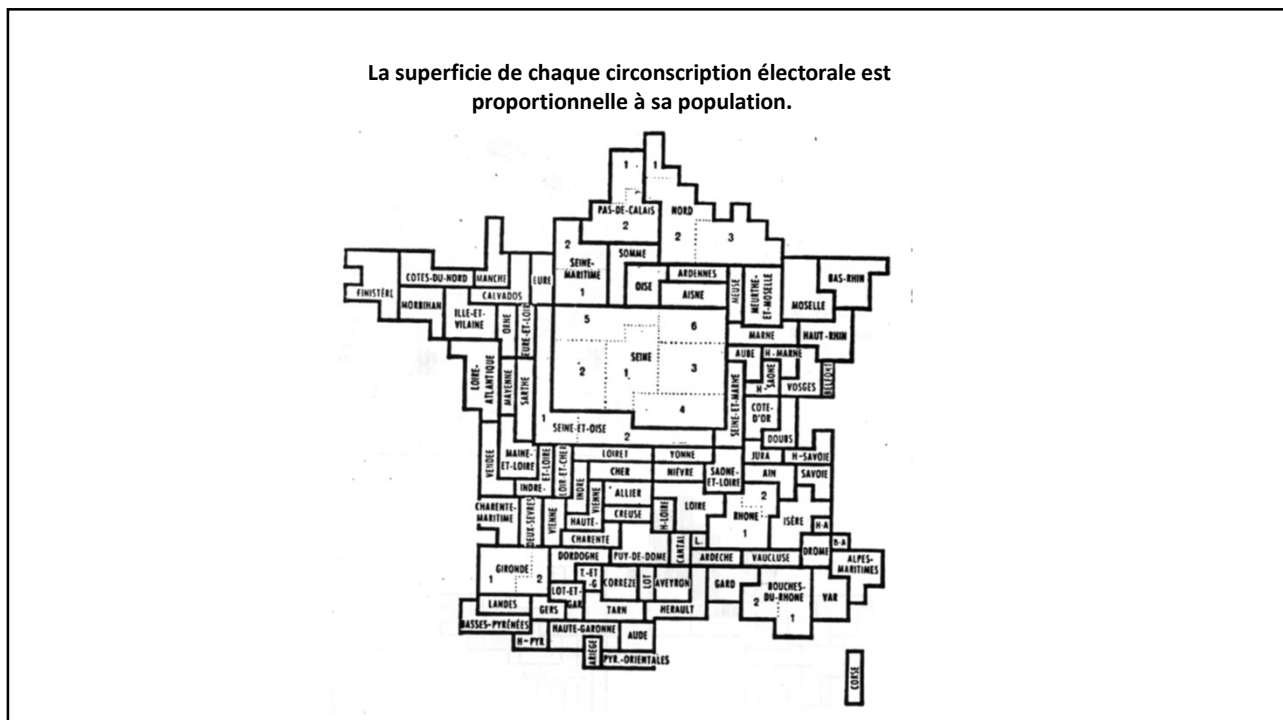
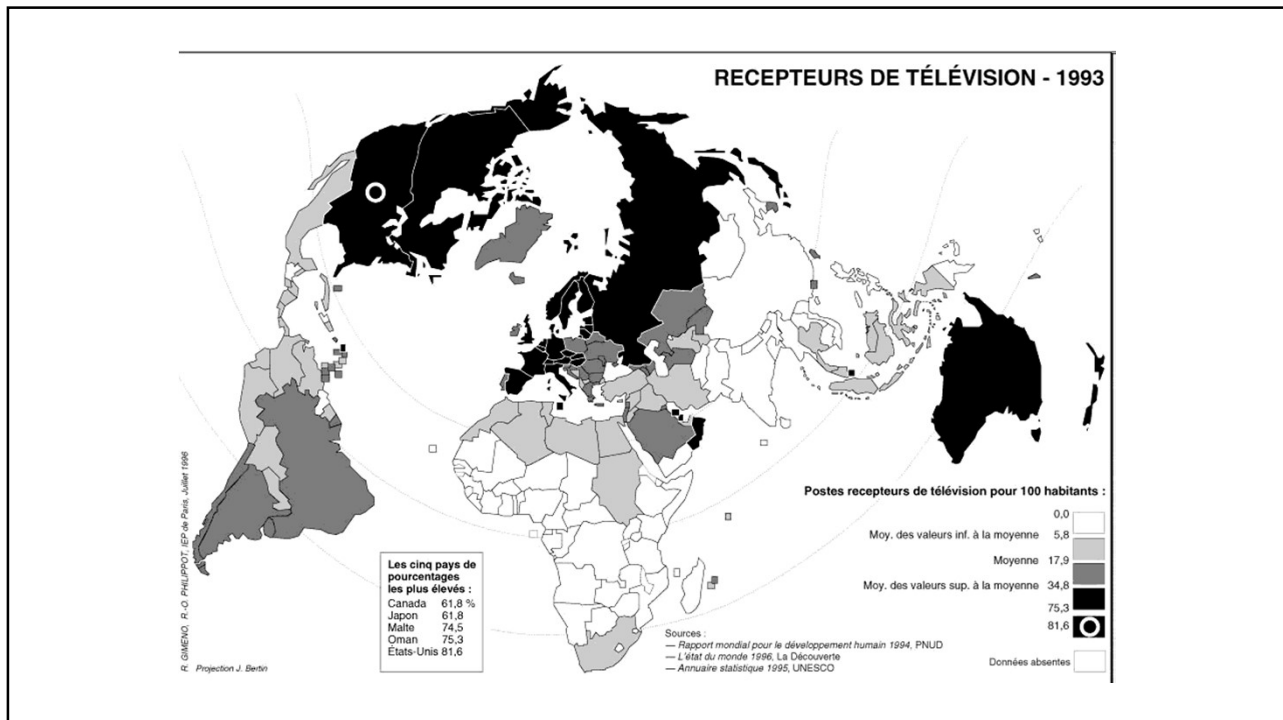
Les cartes permettent une représentation géographique des données : elles peuvent être utiles pour comparer différentes zones, et peuvent remplacer de nombreuses données.

Une carte peut être descriptive ou comporter des données quantitatives : dans ce cas, des classes seront symbolisées par différents aplats ou textures, ou encore par des surfaces variables.

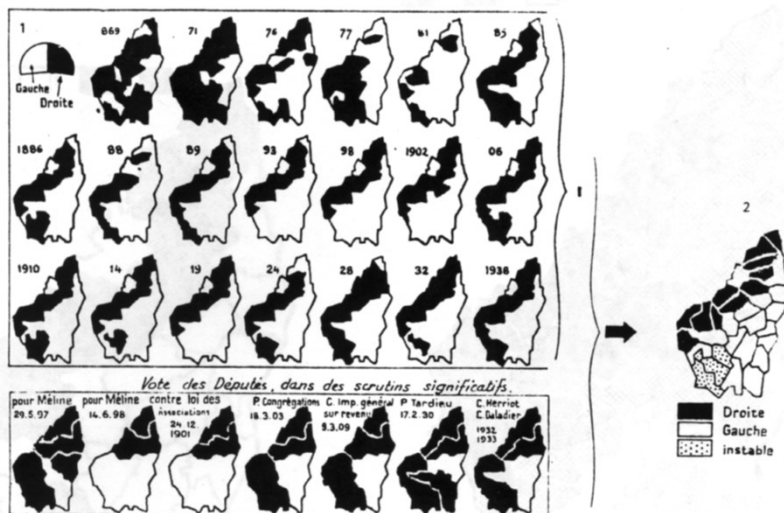
D'autres études ont utilisé la géographie et la comparaison de cartes comme méthode de recherche.

C'est le cas des travaux de l'un des pionniers de la sociologie politique, André Siegfried.

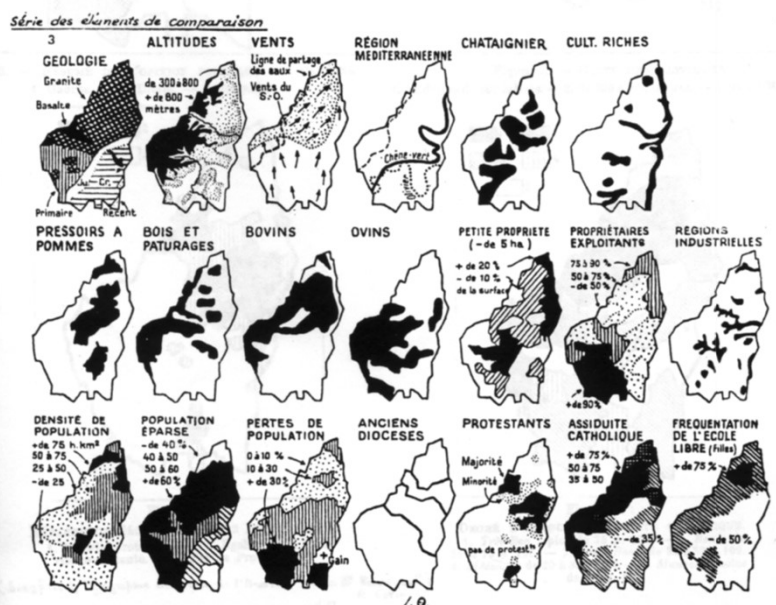
Exemples diapos suivantes.



André Siegfried,
 "Géographie électorale de l'Ardèche sous la IVème République"



André Siegfried,
 "Géographie électorale de l'Ardèche sous la IVème République"



d. Graphiques divers et enrichis

Lorsque les données à traiter sont complexes, on peut éprouver la nécessité d'opter pour des représentations plus originales.

1er cas : juxtaposition d'histogrammes présentant les mêmes caractéristiques dans un but synthétique et comparatif. Le but est alors de présenter des histogrammes très simples afin de pouvoir les comparer entre eux.

2ème cas : représentation graphique hybride entre le tableau croisé et l'histogramme.

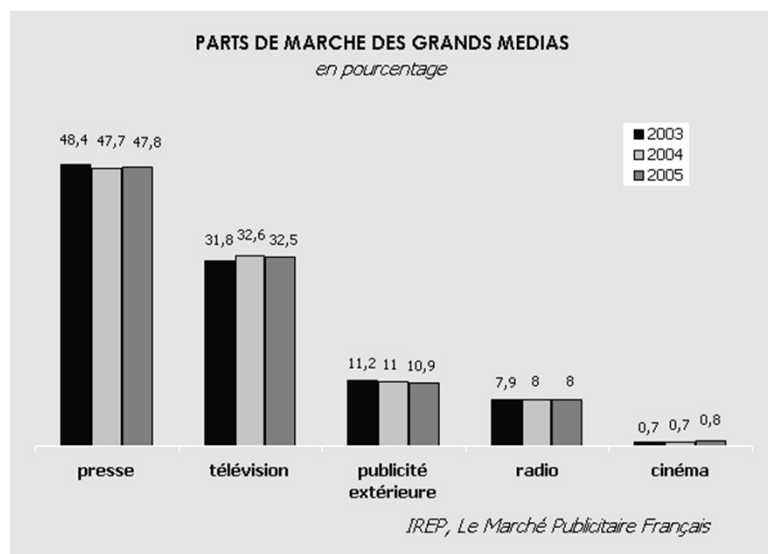
On peut ainsi mieux visualiser des fréquences ou la répartition des populations. L'avantage de ce type de représentation est de pouvoir du premier coup d'œil distinguer des éléments structurants.

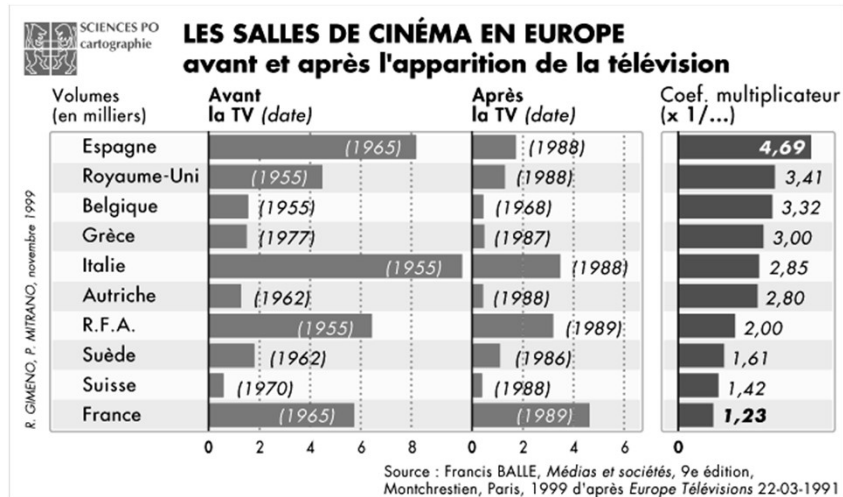
L'inconvénient est le manque de précision : il s'agit plutôt de présenter qualitativement des données quantitatives.

3ème cas : histogramme enrichi.

Il s'agit dans ce cas de retoucher le graphique et d'y adjoindre des renseignements supplémentaires dans le but de le rendre plus pertinent pour l'analyse.

On peut ainsi trier les données, coloriser les bâtons, ajouter des éléments tels que légendes supplémentaires, flèches et commentaires, etc.





Représentation hybride

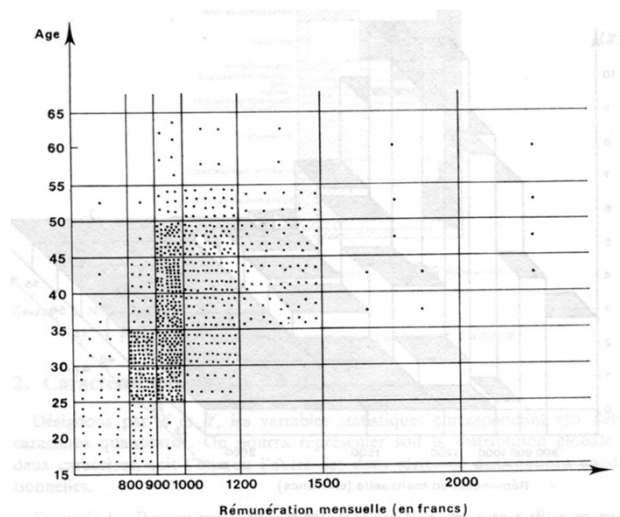


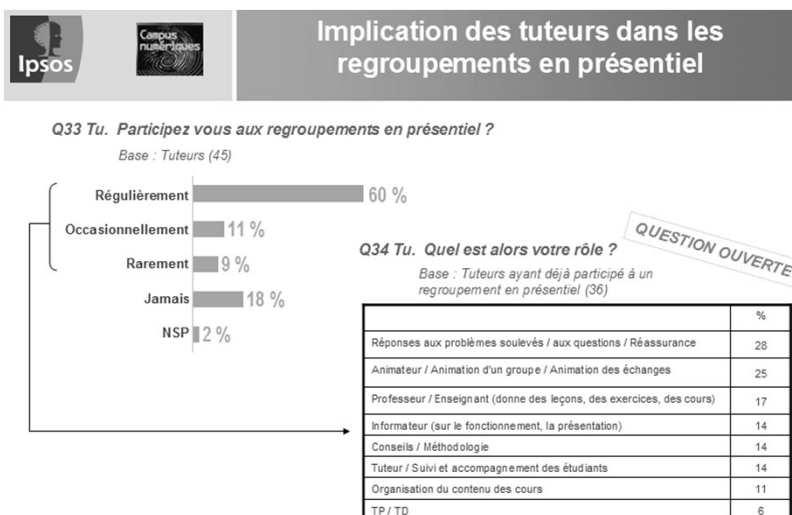
FIG. III.5 Représentation graphique de la distribution des fréquences des ouvriers selon l'âge et la rémunération mensuelle.

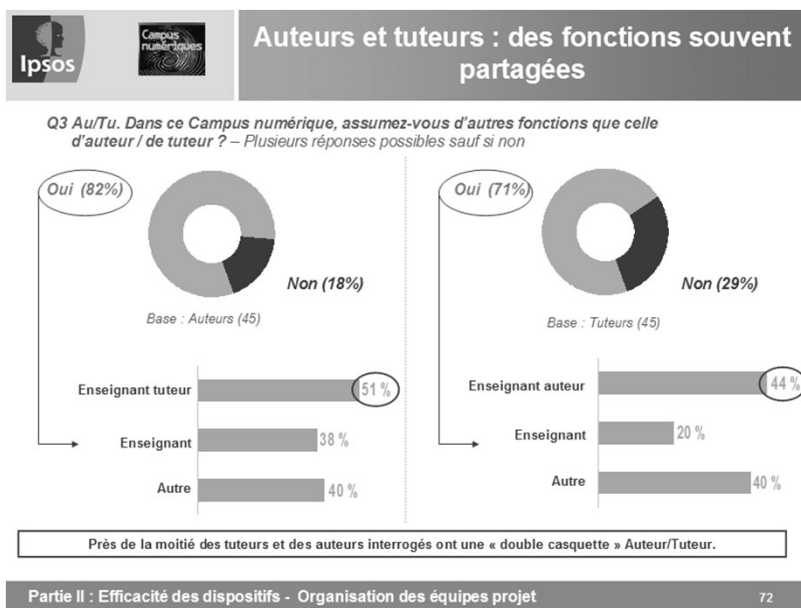
Représentation hybride

Figure 1 : domaines d'intervention médiatique des acteurs politiques par période trisannuelle (en nombre de reportages)

	63	66	69	72	75	78	81	84	87	90	93
	-	-	-	-	-	-	-	-	-	-	-
	65	68	71	74	77	80	83	86	89	92	95
Jeu politique (904)											
Traditions républicaines (813)											
Agriculture (562)											
Politique extérieure (453)											
Education / Recherche (451)											

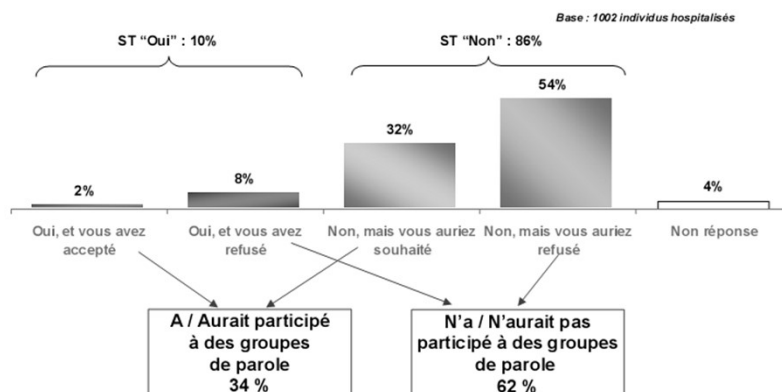
Histogramme enrichi





Participation à des groupes de parole

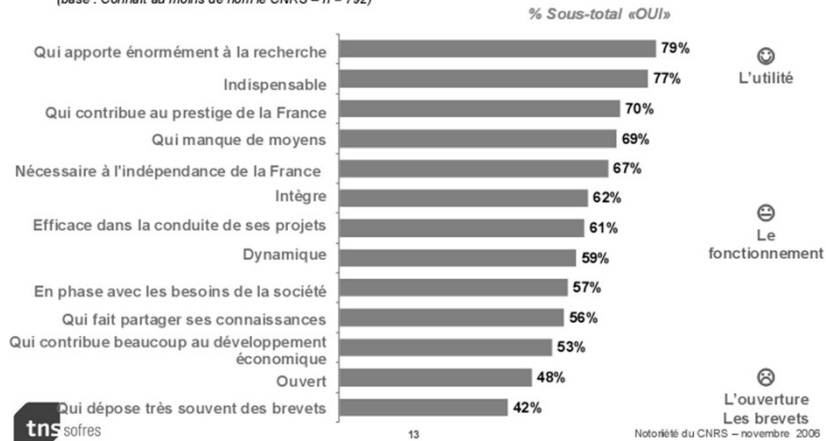
Question 46 : Vous a-t-on proposé de participer à des groupes de parole de patients pour vous aider à surmonter votre maladie ?



Le CNRS apparaît comme une institution indispensable, mais peu ouverte ...



Q9 D'après vous, le CNRS est-il un organisme ... ?
(base : Connaît au moins de nom le CNRS – n = 792)



e. L'AFC (analyse factorielle des correspondances) et les « mappings »

L'AFC est requise si l'on veut éviter d'en rester au niveau d'un jeu de causes simples et élémentaires.

Il faut traiter le complexe en tant que tel.

Cette complexité se reflète dans des dossiers très volumineux, qu'il s'agisse de relevés, d'enquêtes ou d'archives.

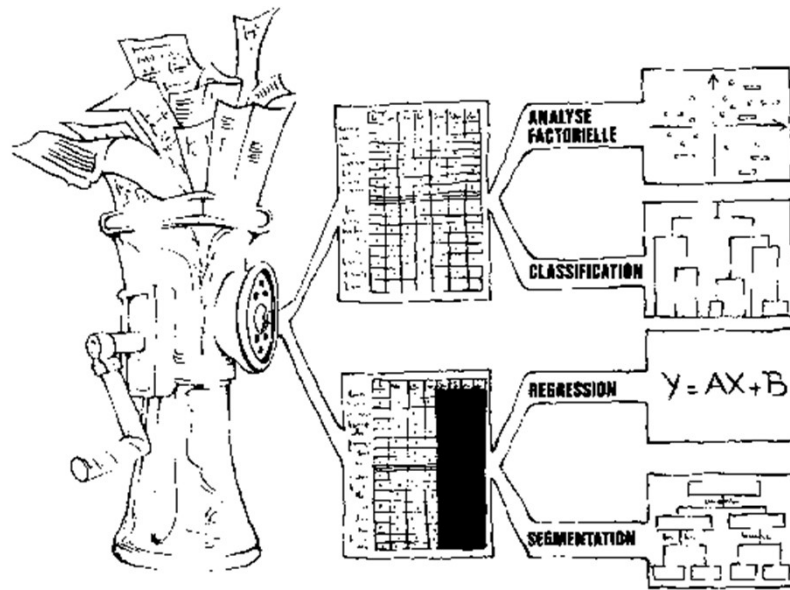
Il faut transformer ce dossier en un *tableau* qui en représente les éléments chiffrés.

Cette opération de codage est une opération fondamentale qui conditionne la qualité des résultats obtenus.

→ construction d'un tableau croisé, étape essentielle de l'AFC.

On classe les différentes techniques de codage des données en deux grandes catégories la seconde permettant une analyse factorielle :

- codage où certaines variables, c'est-à-dire certaines colonnes, jouent un rôle particulier (régression et segmentation).
- codage où toutes les variables, c'est-à-dire toutes les colonnes du tableau, jouent des rôles identiques (analyse factorielle ou classification)



Les différentes techniques

Exemple de segmentation

TAUX DE RÉUSSITE EN PREMIER CYCLE DE L'UNIVERSITÉ SELON :

		Taux	Écart
la CSP du père	populaire	42,7	8,4
	moyenne et supérieure	48,9	
l'exercice d'une activité salariée	oui	25,9	37,5
	non	58,4	
l'âge à la première inscription en fac	< 20 ans	61,3	29,8
	≥ 20 ans	31,5	
l'obtention d'une mention au bac	oui	62,2	20,1
	non	42,1	
la section du secondaire	classique	63	23,1
	moderne	39,9	

Exemples introductifs : application de l'analyse factorielle à divers domaines

1/ Le marketing

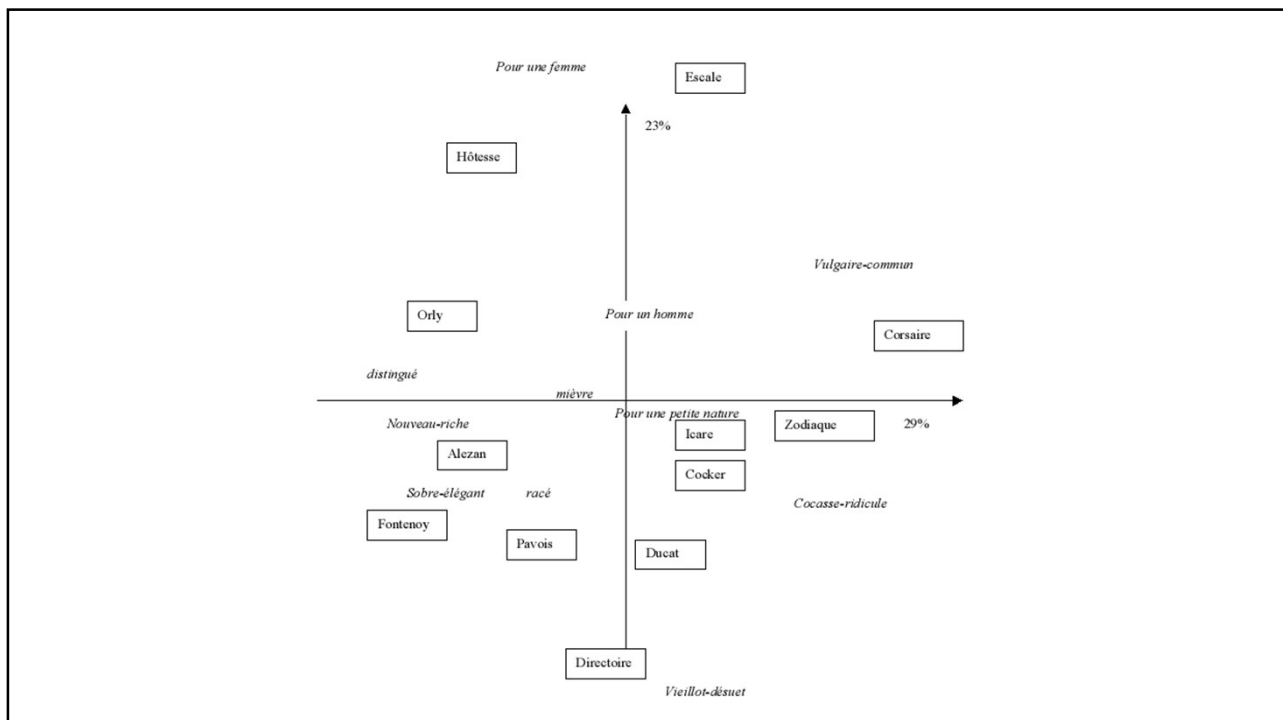
Par exemple les services commerciaux d'une entreprise qui s'attachent à définir les besoins des consommateurs, ou à « positionner » un nouveau produit.

Exemple d'une étude qui avait pour objet de choisir un nom pour une nouvelle cigarette. Cette carte résume l'information recueillie en associant 11 attributs à 12 marques possibles de cigarettes.

Le tableau de données est tiré de « Optimisation en Analyse ordinale des données » J-F MARCOTORCHINO et Pierre MICHAUD, Masson 1979.

Le cas est réel, le nom finalement choisi par la SEITA a été « Fontenoy »

	Viellot-désuet	Nouveau-riche	Sobre-élégant	Cocasse(ridicule)	Racé	Miévre	Distingué	Vulgaire-commun	Pour un homme	Pour une femme	Pour une petite nature
Orly	1	20	9	1	4	3	11	4	9	9	7
Alezan	2	9	23	3	33	9	9	4	12	3	5
Corsaire	14	1	1	15	7	1	1	32	23	9	2
Directoire	38	11	15	15	8	7	17	2	4	8	7
Ducat	18	10	7	6	3	7	4	6	7	4	11
Fontenoy	10	9	11	5	6	5	21	0	13	2	2
Icare	9	1	6	12	6	12	6	9	5	6	6
Zodiaque	5	1	2	18	4	9	1	7	5	8	11
Pavois	9	20	7	4	5	6	5	3	10	1	9
Cocker	4	9	12	25	15	9	4	10	5	6	24
Escale	0	7	3	2	5	6	5	12	13	23	10
Hôtesse	1	12	17	2	3	13	27	7	9	33	5



Le premier axe s'interprète comme une opposition sémantique (vulgaire/distingué), le second axe représente l'idée de mode (vieillot-désuet/pour une femme), le troisième voit racé et Alezan s'opposer à désuet et Directoire: c'est le facteur de l'originalité, alors que le quatrième est un facteur de virilité où «pour une petite nature» s'oppose à «pour un homme».

La cible visée était un public «distingué et masculin»: Fontenoy et Alezan convenaient tous deux a priori.

Cet exemple offre peu d'intérêt statistique, mais il introduit bien à la forme des résultats.

2/ En Sociologie

Sentiment d'appartenance religieuse , une Analyse Factorielle

Source « *La Croix* » 17 nov 1998 . Tableau des profils-colonnes en %.

	D	B	Sp	F	Ir	It	N	U-K
Catholique	45	65	85	58	93	80	30	9
Protestant	43	1	0	1	2	1	17	47
Autre religion	1	3	1	3	1	1	4	2
Sans religion	11	32	14	38	4	18	49	42

D : Allemagne

B : Belgique

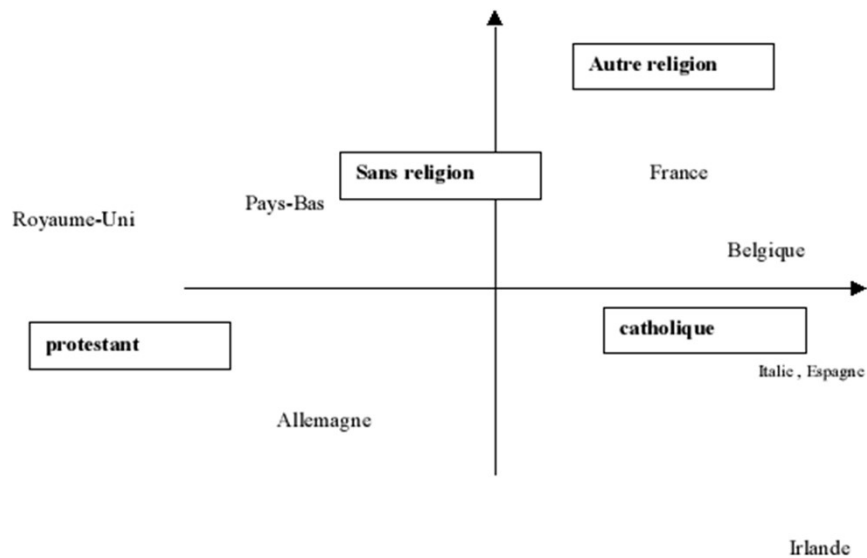
Sp : Espagne

Ir : Irlande

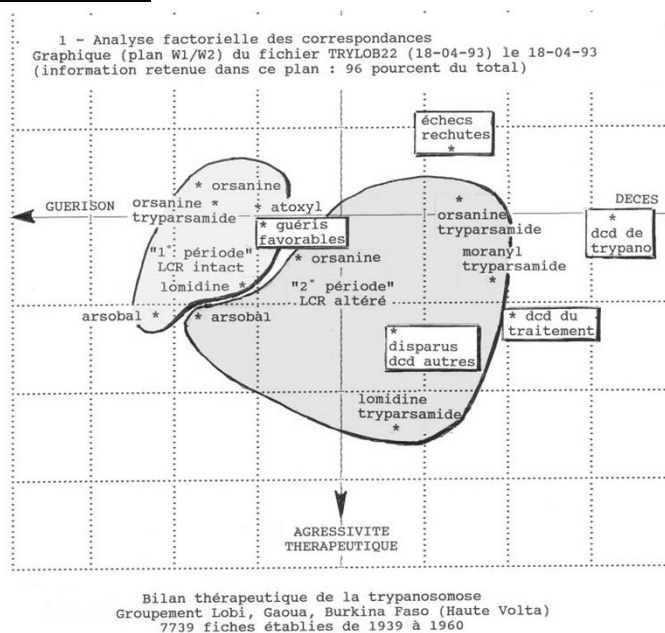
It : Italie

N : Pays-Bas

U-K : Royaume-Uni



3/ En médecine



4/ En agronomie

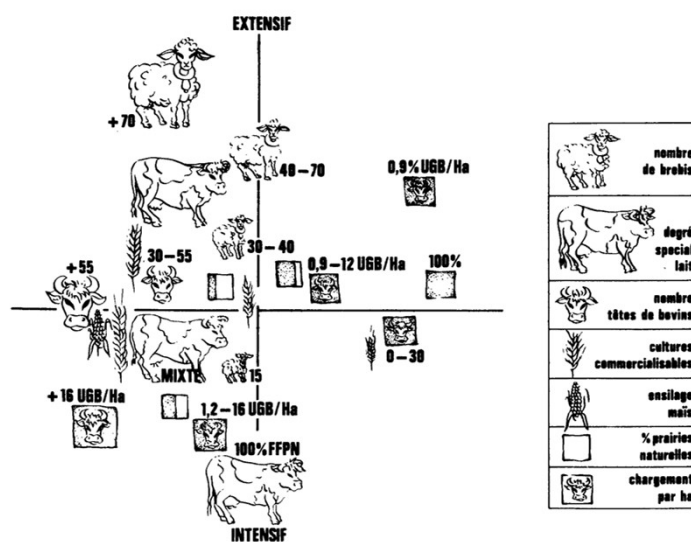


Figure 5 : Economie Agricole

étude réalisée en déc. 1975 pour la Chambre d'Agriculture de Maine-et-Loire (échantillon de 202 agriculteurs).

Axe 1 horizontal : sur cet axe on trouve à l'extrême droite les agriculteurs ayant des bas revenus avec un système traditionnel peu rentable et à l'extrême gauche les agriculteurs ayant de très hauts revenus

Axe 3 vertical : il met en évidence l'opposition des niveaux d'intensification; au fur et à mesure qu'on se déplace vers le bas, on passe des systèmes extensifs vers des systèmes très spécialisés

L'AFC – Principes de base

Exemple fondé sur les notes obtenues dans diverses matières par les élèves d'une même classe.

L'AFC permet de repérer les élèves en fonction du profil de leurs notes dans plusieurs matières.

Le dossier des élèves se présente comme un tableau où chaque élève représente une ligne et chaque matière une colonne.

Tableau croisé à soumettre à l'analyse

	Français	Math.	Chimie	Musique	Gym.	Dessin
Jean	3	9	7	2	10	1
Pierre	7	1	3	8	2	9
Gérard	4	6	5	4	7	4
Henry	6	4	4	6	5	7
Nathalie	3	7	7	3	8	2
Roger	4	8	6	4	9	5
Véronique	8	5	2	9	3	9

Figure 4 : Tableau des notes

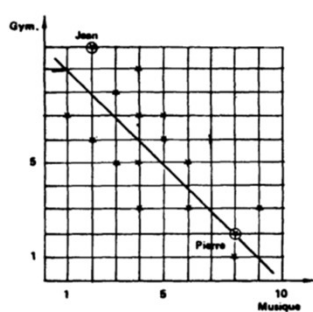


Figure 5 : Représentation géométrique de deux notes

	Musique	Gym.
Jean	2	10
Pierre	8	2
Gérard	4	7
Henry	6	5
Nathalie	3	8
Roger	4	9
Véronique	9	3

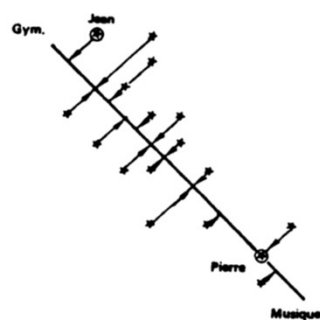
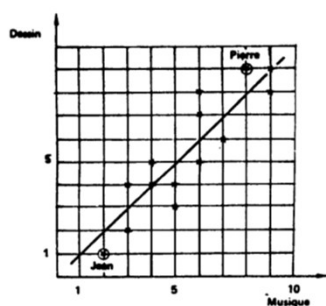


Figure 6 : Résumé du nuage « Musique-Gymnastique »



	Musique	Dessin
Jean	2	1
Pierre	8	9
Gérard	4	4
Henry	6	7
Nathalie	3	2
Roger	4	5
Véronique	9	9

Figure 7 : Nuage « Musique-Dessin »

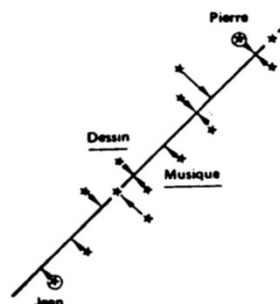


Figure 8 : Synthèse « Musique-Dessin »

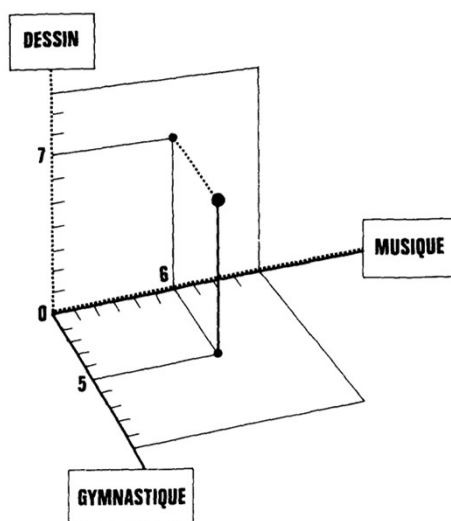


Figure 9 : Situation d'un élève dans trois dimensions



2. « Des idées aux faits, des faits aux idées »

Une fois le traitement des données établi :

- Commenter l'aspect des visualisations
- Commenter les indicateurs produits

⇒ En tirer des conclusions : « montée en généralité »

⇒ Retour sur la problématique : affiner la problématique