



**Direction des Statistiques Démographiques et Sociales**

Unité Méthodes Statistiques

Division Méthodologie d'Elaboration et d'Analyse de Données

Timbre F410

**Statistique, Imagerie et Sciences Cognitives**

**Monique LE GUEN**

**Document n° F 9614**

**Juillet 1996**

**Série des documents de travail**

**de la Direction des Statistiques Démographiques et Sociales**

# Statistique, Imagerie et Sciences Cognitives

*Monique LE GUEN*

*CNRS-INSEE- Unité Méthodes Statistiques*

## Résumé

Depuis deux décennies l'enseignement et la pratique de la Statistique sont largement dominés par les procédures calculatoires et logiques des Mathématiques. Les Mathématiques ne sont pas à la portée de tous, alors que la "pensée" statistique doit être partagée par le plus grand nombre.

Comment faciliter l'apprentissage de la pensée statistique? "Je n'ai des idées que parce que j'ai des **images**" a dit EULER. Les Sciences Cognitives cherchent à démontrer que Euler avait raison. Le rôle de l'imagerie dans les processus cognitifs, les stratégies d'apprentissage développées par les apprenants sont actuellement objets de travaux de recherche et leurs résultats commencent à se répandre en dehors de leur communauté de recherche.

Les recherches actuelles en Sciences Cognitives peuvent nous apporter des aides pour concevoir des enseignements intelligemment assistés par ordinateur (EIAO).

## Mots-clés

Statistique, Visualisation Scientifique, Imagerie, Sciences Cognitives, Pédagogie, Enseignement Intelligemment assisté par Ordinateur (EIAO).

## Plan

1. Analyse de Données et Statistiques conventionnelles.....	3
2. Statistique et Visualisation : un nouveau poste de travail.....	3
3. Sciences Cognitives, Communication et Pédagogie .....	5
4. Vers la réconciliation.....	6
Annexe 1 Comment les chiffres peuvent masquer la réalité .....	10
Annexe 2 Introduction aux Sciences Cognitives.....	11

## 1. Analyse de Données et Statistiques conventionnelles

Pour comprendre, remontons 25 ans en arrière. Dans les années 1970-80, l'utilisation croissante de l'ordinateur a permis la réalisation de bibliothèques de programmes informatiques conduisant à étudier un vaste panel de méthodes mathématiques et statistiques.

Les méthodes d'analyse de données se sont substituées aux méthodes statistiques conventionnelles qui elles, faisaient largement appel aux tests et aux modèles nécessitant toujours des *hypothèses supplémentaires*. Ce furent l'analyse en composantes principales, l'analyse des correspondances, les classifications, l'analyse relationnelle<sup>1</sup> etc.

Ces méthodes de traitement de l'information regroupées en France sous l'appellation "Analyse de Données" ont la même préoccupation, *faire émerger d'un ensemble de données, une structure particulière qui restitue l'essentiel de l'information tout en réduisant la masse de données* (Marcotorchino, 1991). Ces méthodes ont également en commun d'avoir apporté un début de représentations visuelles. Cependant les sorties avec les aides à l'interprétation peuvent encore être difficiles à comprendre pour le non spécialiste de la méthode employée. A la complexité du sujet à analyser vient s'ajouter la complexité des méthodes de traitements, distances sous-jacentes et algorithmes.

## 2. Statistique et Visualisation : un nouveau poste de travail

Depuis l'introduction de l'interface graphique de Apple, reprise sur toutes les plates-formes par tous les grands constructeurs, les pratiques informatiques ont évolué. Les utilisateurs manipulent intuitivement leurs données et leurs fichiers souvent sans l'obligation de connaître à fond des techniques informatiques nécessitant un long apprentissage. Les icônes, les fenêtres, les menus déroulants et les souris, ont remplacé les commandes ésotériques.

Grâce au développement et la diffusion hors du milieu recherche et universitaire des bibliothèques de programmes statistiques, il est possible d'obtenir sans grand effort des analyses statistiques appréhendées sous des angles et des points de vue différents. Ces analyses conjuguées à des représentations visuelles fixes pour le moment, mais animées dans un avenir proche, permettent de produire dans le domaine de la statistique deux classes d'images : des images d'analyse et des images de communication (Lucas, 1993).

Les *images d'analyse* conduisent à la prise de connaissance d'une situation donnée. C'est l'objet des graphiques comme les courbes, les histogrammes, les camemberts etc. qui sont enseignés et utilisés classiquement. La deuxième classe d'images regroupées sous le vocable d'*images de communication* sont destinées à mettre en valeur une idée, un résumé de données, à frapper l'imagination. Autant pour les images d'analyse des règles de conceptions doivent être connues, apprises et respectées (Bertin, 1974) autant pour les images de communication toute liberté est laissée à l'imagination et à la fantaisie de l'auteur pourvu qu'il communique son message.

---

<sup>1</sup> L'analyse relationnelle fut proposée en 1977 par F. MARCOTORCHINO et P. MICHAUD. Ils établirent une jonction entre les différents sujets de recherche permettant d'unifier les méthodes basées sur les relations de similarités et celles de préférences. Plus récemment F. Marcotorchino mit en évidence la liaison existant entre l'analyse relationnelle et l'analyse factorielle des données.

Cependant la frontière entre ces deux classes est encore floue du fait du peu de recherches réalisées sur la compréhension des images en liaison avec la psychologie cognitive (Denis, 1987) et plus généralement les sciences cognitives.

En 1989, SAS Institute a développé sur la plate-forme Mac-Apple, le logiciel JMP<sup>2</sup> qui associe la facilité d'utilisation du Macintosh à l'analyse exploratoire et visuelle des données. Cette nouvelle approche de la statistique nettement plus conviviale permet au novice une sensibilisation et un apprentissage beaucoup plus rapide à l'usage des statistiques, mais permet également au statisticien professionnel de gagner un temps précieux dans l'analyse et l'interprétation de ces résultats statistiques.

Ces nouveaux environnements informatiques doivent remettre en cause les méthodes de travail en statistique tout comme l'enseignement de cette discipline. Etudier la statistique et savoir l'utiliser dans de bonnes conditions est nécessaire dans presque tous les domaines de la connaissance, mais cette acquisition reste encore longue et laborieuse.

La cause est peut-être due à la forte influence des mathématiciens sur l'enseignement et la *pratique* statistique. En effet, les mathématiciens ont été très réticents à introduire les ordinateurs dans leurs recherches, et le débat existe encore dans leur microcosme pour savoir si les mathématiques peuvent être considérées comme une science d'expérience et d'observation. Selon John Huppard<sup>3</sup> l'usage des ordinateurs en mathématiques appliquées n'est pas contesté, mais en mathématiques pures cette pratique n'est pas encore passée dans les moeurs (Huppard, 1993).

Pourtant, pour comprendre les mathématiques tout comme la statistique, et en faire bon usage, sentir *concrètement* ce que cachent les chiffres, peut aider énormément; voir par exemple le graphique n°1 "Comment les chiffres peuvent masquer la réalité".

Huppard cite comme exemple l'engouement actuel pour les fractals, inventées en 1920 par Fatou et Julia et présentées dans deux mémoires primés à l'Académie des Sciences à cette même date. Leurs travaux sont restés sans suite jusqu'à ce que Mandelbrot les sorte de l'oubli en montrant sur des images spectaculaires, ce à quoi cela correspondait, et en donnant un nom à cette nouvelle branche des mathématiques "la Géométrie Fractale". Les images mentales que Fatou et Julia avaient, ils ne les ont pas mises sous une forme facile à communiquer<sup>4</sup>.

Huppard fait la remarque suivante : "Kahane<sup>5</sup> qui a lui même contribué à l'évolution mathématique de la géométrie fractale, peut lire et comprendre maintenant les oeuvres de Fatou et Julia, ce qu'il n'aurait pas pu faire, s'il n'avait pas vu les images de B. Mandelbrot". Huppard

---

<sup>2</sup> JMP Software de SAS Institute a servi de prototype pour la réalisation de SAS/INSIGHT.

<sup>3</sup> John HUBBARD Professeur de Mathématique, Université de Cornell, spécialiste des représentations visuelles des fractales. Conférence du Séminaire d'Histoire des Sciences et des techniques Université Paris-Sud ORSAY: "Les Mathématiques et l'expérimentation amis ou ennemis?" 22 Février 1993 .

<sup>4</sup> La géométrie fractale a trouvé une application promise à un développement important dans la communication matérielle des images numérisées. La compression et la décompression fractales des images utilisent une technique basée sur le théorie des "fonctions fractales itératives". Cette méthode s'attache à rechercher des entités fractales au sein d'une image, c'est à dire des motifs pouvant être stockés de manière compacte à l'aide d'équations mathématiques.

<sup>5</sup> Jean Pierre KAHANE professeur de Mathématique Université Paris-Sud Orsay.

ajoute encore, "Je pense que bien souvent en Sciences Humaines et Sociales, les ordinateurs sont utilisés comme outil pour intimider l'auditoire, plutôt que comme outil de connaissance".

C'est également le même message que Benzecri essaye de faire passer depuis de nombreuses années dans les *Cahiers de l'Analyse de données* (Benzecri, 1976-1996) ou dans un ouvrage à accès plus aisé, *l'Encyclopédie Universalis* (Benzecri, 1973), lorsqu'il dit : « L'économiste, qui entreprend de lier par des équations différentielles quelques grandeurs dont les noms couvrent l'ensemble de son objet, a devant lui une belle carrière mathématique : les instruments de calcul les plus raffinés le serviront; et les conclusions numériques ultimes, issus de son modèle, peuvent, au moins en partie, se traduire dans le langage du sens commun en autant d'oracles qui, même s'ils sont tous inacceptables, seront indiscutables comme l'enchaînement rigoureux des formules.... il est a priori douteux que l'élaboration mêmes des données économiques conduise à une description adéquate des faits au moyen de variables se prêtant, comme celles de la physique classique, aux finesses du calcul infinitésimal. »

### 3. Sciences Cognitives, Communication et Pédagogie

S'il est passé dans les esprits que l'apprentissage de la statistique ne pouvait se faire sans micro-ordinateur, il n'est pas encore reconnu par la plupart des gens, que l'image lorsqu'elle est couplée à un raisonnement, facilite la compréhension, la mémorisation, et diminue le temps d'apprentissage, *même en mathématique*.

Une illustration de cette situation, est l'enseignement qu'effectue Marie Berrondo-Agrell (Berrondo-Agrell & Agrell, 1992; Berrondo-Agrell & Fourastié, 1995). Depuis 20 ans, M. Berrondo-Agrell enseigne le calcul des probabilités en utilisant comme support à la compréhension des règles de calcul des probabilités, la représentation visuelle développée par John Venn<sup>6</sup> et généralisée à l'ordre n, par A. Edwards (1989).

Les diagrammes de Venn dits aussi représentation d'Euler<sup>7</sup> sont utilisés classiquement dans l'éducation anglo-saxonne depuis longtemps, aussi bien pour enseigner les probabilités et les statistiques (Spiegel, 1975), que pour enseigner les tables de vérité, l'algèbre de Boole ou les lois de Morgan en informatique. Dans les livres des auteurs français on ne trouve pas souvent de telles analogies.

La réticence à utiliser des images concrètes comme aide à la compréhension en mathématique et statistique semble être plus française. Aux USA, des études sur la perception graphique des données qualitatives ou quantitatives sont menées très officiellement à AT&T Bell Laboratories (Cleveland & McGill, 1987). Ces recherches, pour concevoir des graphiques statistiques, associent statisticiens et psychologues, aussi bien au plan théorique qu'au plan expérimental. A notre connaissance, en France, les deux communautés ont peu de liens.

---

<sup>6</sup>John VENN, mathématicien et logicien au Collège Gonville et Caius de Cambridge a décrit dans son livre *Symbolic Logic* en 1881, une symbolisation visuelle des propositions logiques dans l'esprit de Lewis CARROLL.

<sup>7</sup>EULER a écrit: "Il existe un rapport profond entre le raisonnement logique et l'espace", et on lui prête la remarque "Je n'ai des idées que parce que j'ai des images".

L'émergence des sciences cognitives qui sont des sciences multidisciplinaires, cherchent à redonner de l'importance à l'imagerie et à l'observation par le biais des modèles mentaux et des représentations mentales. De nombreux travaux scientifiques ont étudié le rôle de l'imagerie dans les processus cognitifs et commencent à se répandre en dehors de leur communauté de recherche (Gyselink & Tardieu, 1993).

Selon les psychologues cognitivistes, pour comprendre un texte, qu'il soit littéraire ou scientifique, le lecteur doit développer une stratégie de représentation mentale pour construire un *modèle mental* (*ibid*). Leurs expériences révèlent que cette construction est facilitée par des illustrations lorsque celles-ci sont proposées en *fin de lecture*.

Des expériences menées en psychologie cognitive font apparaître une supériorité de la reconnaissance visuelle et une meilleure mémorisation pour des textes ou des mots à forte valeur d'imagerie (Thon et al, 1993). Ces auteurs, Thon, Marquié et Maury, aux vues de leurs résultats pensent que l'encodage en mémoire et la reconnaissance d'une image ne passent pas nécessairement par un recodage verbal de l'information perceptive. La reconnaissance d'une image peut donc demander moins de temps que l'utilisation d'un encodage de type verbal.

D'autres psychologues supposent que les dessins sont mieux retenus car ils font l'objet d'un double encodage : un encodage verbal et un encodage imagé. Le premier se situerait dans l'hémisphère gauche du cerveau, le second dans l'hémisphère droit. Ce double ancrage expliquerait l'efficacité de l'imagerie mentale dans la résolution de problèmes et peut contribuer à un explication de la généralisation des *icônes* et des interfaces graphiques dans le milieu informatique.

Si l'illustration peut remplir une fonction *facilitatrice*, elle peut aussi devenir *inhibitrice*. Ces effets inhibiteurs sont sans doute plus difficiles à mesurer et la littérature sur le sujet est moins prolifique.

#### 4. Vers la réconciliation

« Tant que l'Algèbre et la Géométrie ont été séparées, leurs progrès ont été lents et leur usages bornés; mais lorsque ces deux sciences se sont réunies, elles se sont prêtées des forces mutuelles et ont marché ensemble d'un pas rapide vers la perfection. C'est à Descartes qu'on doit l'application de l'Algèbre à la Géométrie, application qui est devenue la clef des plus grandes découvertes dans toutes les branches des Mathématiques. » Lagrange, 1795.

Le XIX<sup>ème</sup> siècle, en Mathématique, fut marqué par l'importance de la Géométrie. L'intuition géométrique, l'art de raisonner sur des figures était une exigence de l'époque. Au XX<sup>ème</sup> siècle, l'introduction des mathématiques modernes a changé radicalement les modes de raisonnement. Du presque tout *géométrie* on est passé au tout *algèbre*. On a formé d'excellents calculateurs formels mais on a laissé sur le bord du chemin ceux qui ne pouvaient pas suivre. Ce fut « une erreur scientifique et pédagogique..., replacé dans l'époque, c'était la vision des structures - Piaget et Bourbaki s'épaulant mutuellement- mise en oeuvre dans l'enseignement » nous dit maintenant J.P. Kahane (1995).

A l'aube du XXI<sup>ème</sup> siècle, une réconciliation entre les deux parties semble en marche. Les facilités d'usage des nouveaux moyens informatiques : -matériels, logiciels, ateliers de génie logiciels, multi-média etc,- vont permettre de repenser l'enseignement en concevant de véritables "Enseignements Intelligemment Assistés par Ordinateur" (EIAO).

Nous savons tous intuitivement que la communication entre des individus qui n'ont pas les mêmes connaissances, et donc pas les mêmes raisonnements, passe plus facilement par un vecteur graphique, un dessin ou une analogie. Mais ce que nous apportent les résultats des sciences cognitives, ce sont des explications, des validations et des procédures à mettre en oeuvre. Les connaissances acquises sur les processus d'apprentissage développés par les apprenants, ont fortement progressé ces vingt dernières années. Les retombées de ces résultats, en communication et surtout en pédagogie devraient apporter des gains de productivité dans l'apprentissage de nouvelles connaissances.

C'est un changement de mentalité profond dans notre culture occidentale. Les Français notamment, ont une vieille tradition d'enseignement par le raisonnement logique et abstrait, qui s'oppose sur ce plan à la conception plus pragmatique des Américains. Pour étayer nos convictions citons Jean Largeault (1992) analysant le livre de René Thom (1991) "*Prédire n'est pas expliquer*".

« Ce qui est en cause, c'est une conception des rapports entre la pensée et la réalité. L'épistémologie française a longtemps enseigné que la pensée est d'autant plus noble qu'elle est plus abstraite, plus déductive et plus éloignée de l'empirisme immédiat. A ces leçons de l'intellectualisme brunshvicgien ou bachelardien s'ajoute aujourd'hui le rationalisme des règles combinatoires et des algorithmes. On ne semble pas s'apercevoir que plus on calcule moins on comprend. »

Orsay le 20 Juin 1995

## REFERENCES

- ANDLER D. (1992) -sous la direction de- "Introduction aux Sciences Cognitives" Folio, essais, Editions Gallimard, 514 pages.
- BENZECRI J.P. (1973) "La place de l'a priori" Encyclopédia Universalis, Vol 17, Organum pp11-24
- BENZECRI J.P. (1976-1977) "Histoire et Préhistoire de l'Analyse de Données", les Cahiers de l'analyse de Données, 1976 n°1 pp9-32, n°2 pp101-120, n°3 pp221-241, n°4 pp 434-346, 1977 n°1 pp9-40
- BENZECRI J.P. (1988) "D'Aristote à l'analyse multidimensionnelle : logique et ontologie dans l'induction statistique", Les Cahiers de l'Analyse de Données Vol XIII 1988 n°3 pp 325-338
- BERRONDO-AGRELL M. et AGRELL P. (1992) "Vers une syntaxe des diagrammes de Venn : lutte contre un mythe" SSP tome 133 n°1, 1992, pp140-150.
- BERRONDO-AGRELL M. et FOURASTIE J. (1995) "Pour comprendre les probabilités", Hachette Collection les Fondamentaux, 160 pages.
- BERTIN (1974) "La graphique et le traitement graphique de l'information", Flammarion.
- BRETON T. (1991) "La dimension invisible, le défi du temps et de l'information", Editions Odile Jacob.

BRISSARD F. (1989) "Développez l'intelligence de votre enfant par la méthode de la Garanderie", Editions Le Rocher.

CHANGEUX J.P. et CONNES A. (1989) "Matière à pensée", Editions O. Jacob.

CLEVELAND W. S. et MCGILL R. (1987) "Graphical Perception : The visual decoding of quantitative information on graphical displays of data", J.R. Statist. Social A 1987, 150 Part 3, pp 192-229.

COURRIER du CNRS (1987) n° 66-67-68 Spécial Imagerie Scientifique.

COURRIER du CNRS (1992) n° 79, Dossiers Scientifiques , Sciences Cognitives.

DAMASIO A. R. (1994) "L'erreur de Descartes, la raison des émotions", Editions O. Jacob, 368 pages.

DENIS M. (1987) "Image mentale et pensée scientifique", Le Courrier du CNRS Juin 1987 n°67-68 pp 111-112.

DIEUDONNE J. (1987) Pour l'honneur de l'esprit humain", Les mathématiques d'aujourd'hui Hachette.

DUPUY J. P. (1994) "Aux origines des Sciences Cognitives", Editions la Découverte, 188 pages.

EDWARDS A. (1989) "Venn diagrams for many sets" Newscientist 7 january 1989 N° 1646 pp 51-56.

de La GARANDERIE A. (1982) "Pédagogie des moyens d'apprendre", Le Centurion.

GARFIELD J.(1995) "How students learn statistics", International Statistics Review 1995,63,1, pp25-34

GENINET A. (1993) "La gestion mentale en Mathématiques, Application de la 6ème à la 2de", Editions Pédagogie RETZ.

GYSELINK V. et TARDIEU H. (1993), "Modèles mentaux et Illustrations dans la Compréhension de textes". Actes du Colloque Images et langages, Multimodalité et Modélisation Cognitive, CNRS 1er et 2 Avril 1993 pp199-203.

IMBERT M. (1992) "Neurosciences et sciences cognitives" p49-76, in ANDLER D. (1992) "Introduction aux sciences cognitives", Folio essais, Editions Gallimard .

ISRAEL G. (1996) "La mathématisation du réel, essai sur la modélisation mathématique", Editions du Seuil, 368 pages.

JMP<sup>®</sup> et SAS<sup>®</sup>-INSIGHT de SAS Institute, Logiciels de Statistiques et Visualisation.

JUAN de MENDOZA J.L. "Deux images pour un mot : le rôle de l'hémisphère droit", Actes du Colloque Images et langages, Multimodalité et Modélisation Cognitive, CNRS 1er et 2 Avril 1993 pp41-48.

KAHANE J.-P. "Mathématiques et Formation", La pensée 302; Avril-Mai-Juin 1995, pp89-98.

LE GUEN M. et JAFFEUX C.(1989) "La conjonction Analyse de Données et Statistique inférentielle pour conduire à une meilleure perception visuelle", R.S.A vol XXXVII n°3 1989.

LE GUEN M. (1994) "Séries chronologiques et Analyse de données, des images de synthèse pour répondre à la question : ex-post, la classification COB des SICAV est-elle pertinente? ", JSSP, tome 135, n°1, 1994.

LUCAS M. (1993) "La communication graphique" Le Courrier du CNRS, Février1993 n°80 p102.



MARCOTORCHINO F. (1991) "La classification automatique aujourd'hui", Publications Scientifiques et Techniques IBM France, Novembre 1991 pp35-91.

MOLES A.A. (1990) "Les sciences de l'imprécis", Seuil 1990.

ROSENFELD I. (1994) "L'invention de la mémoire", Champ, Flammarion

SHEPARD R. N. (1990) "l'œil qui pense, visions, illusions, perceptions", Seuil.

R. SPIEGEL M. (1975) "Probabilités et Statistique cours et problèmes", Série Schaum 3ième tirage.

STEWART I. (1989) "les dentelures de l'esprit : quand les diagrammes de Venn deviennent fractals", Pour la Science n°138 Avril 1989.

STUART M. (1995) "Changing the teaching of statistics", The Statistician (1995), 44, n°1, pp 45-54.

THOM R. (1991) "Prédire n'est pas expliquer", Editions ESHEL.

THON B. MARQUIE J.C. et MAURY P. (1993), "Le texte, l'Image et leurs Traitements Cognitifs". Actes du Colloque Images et langages, Multimodalité et Modélisation Cognitive, CNRS 1-2 Avril 1993, pp29-37.

VARELA F.J. (1988) "Connaître les Sciences Cognitives", Tendances et perspectives, Seuil, 123 pages

WEIL-BARAIS A. (1993) "L'Homme Cognitif", PUF 570 pages.

WOODFIELD A. (1992) "Un modèle à deux étapes de la formation des concepts" p274-290, in Andler D. (1992) "Introduction aux sciences cognitives", Folio essais, Editions Gallimard.

#### **FILM :**

"Un violon dans la tête", de Claude EDELMANN, Festival des films scientifiques, Palaiseau 1992.

#### **Nota**

Cet article est paru dans le Bulletin de Méthodologie Sociologique, rubrique : Recherche en cours, avec une bibliographie plus succincte.

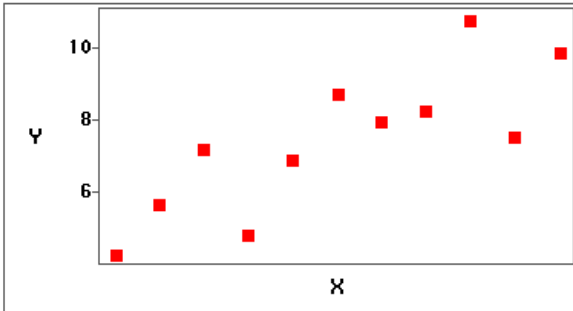
LE GUEN M. (1995) "*Statistique, Imagerie et Sciences Cognitives*", BMS Décembre 1995, n°49, pp90-100.

## Annexe 1 Comment les chiffres peuvent masquer la réalité

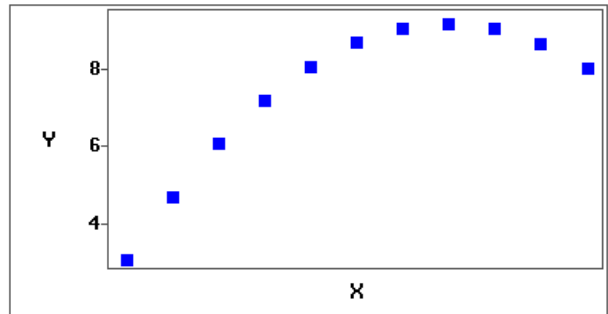
Pour chacun des groupes (Y,X) de données

le coefficient de corrélation est égal à 0.81

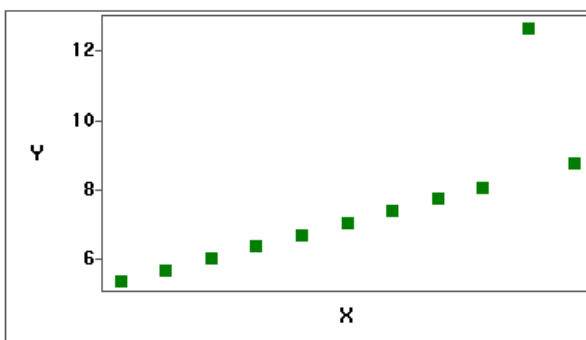
GRUPE = 1



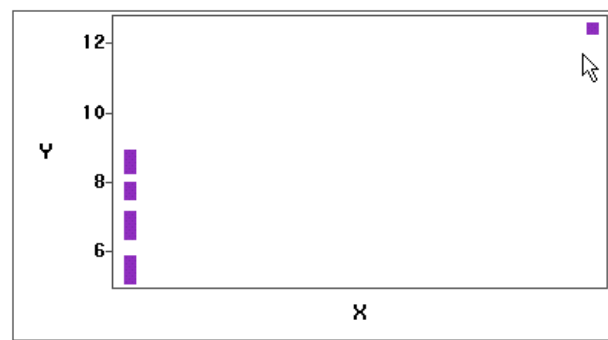
GRUPE = 2



GRUPE = 3



GRUPE = 4



### Source des données

MOLES A.A. (1990) « *Les sciences de l'imprécis* », Seuil.

## Annexe 2 Introduction aux Sciences Cognitives

*une schématisation*

