

---

## **Conférences spéciales**

---



# VISUALISATION GRAPHIQUE ET SEMI-GRAPHIQUE

Eugène Horber

## Résumé de l'intervention

... Depuis un certain temps le terme "visualisation de données" est à la mode. Pour beaucoup il signifie techniques graphiques hautement sophistiquées, pour d'autres ce n'est qu'une expression pour parler de graphiques. La visualisation est en réalité d'abord une approche alternative, une façon particulière de travailler avec les données.

L'objectif de cette présentation est de montrer qu'à côté du regard habituel du statisticien (distributions, paramètres, modèles, etc.) il existe un autre regard plus visuel qui tire sa force de l'interaction dynamique entre les outils, les capacités visuelles du cerveau humain et les moyens informatiques. Cette contribution vise dans un premier temps à caractériser la visualisation pour ensuite proposer un tour d'horizon des différentes familles d'outils graphiques et visuels. L'exposé s'appuiera sur une série d'illustrations et d'exemples ...

## Plan structurel de la présentation

### *Les fondements de la visualisation*

Afin de comprendre le potentiel, les limites et les difficultés de la visualisation, il est indispensable de s'intéresser à la perception visuelle (aspects psychologiques et conceptuels), à la communication graphique et visuelle (propriétés, contraintes, difficultés), ainsi qu'aux préférences et habitudes visuelles. La visualisation des données comme *approche analytique* doit repenser le rôle plus actif et plus responsable de l'analyste (statisticien, utilisateur, expert) responsable face à la démarche, au choix des outils et à l'interprétation.

## ***Visualisation semi-graphique et typographique***

Déjà sur papier, la présentation visuelle de tableaux et de listes mériterait plus d'attention, car, à part faciliter la lecture, une bonne présentation peut mettre en évidence des structures et orienter l'analyse statistique. Utilisés de façon interactive et itérative, de simples tableaux ou listes se transforment en outils analytiques puissants, à condition que l'utilisateur soit capable de donner une direction et une cohérence à l'analyse en adaptant ces outils aux besoins concrets d'une analyse (tris, sélections, codages, colorations, etc.)

## ***Graphiques simples, graphiques statistiques***

La démarche interactive transforme de simples graphiques tels que les histogrammes et les graphiques de dispersion en outils analytiques puissants. Les boîtes à pattes (*boxplots*), branchages (*stemleaf*) et d'autres outils graphiques et semi-graphiques introduits par l'analyse exploratoire des données illustrent l'efficacité des graphiques pour caractériser les variables et pour poser un diagnostic statistique. Pour répondre aux multiples besoins de la pratique, un grand nombre de graphiques statistiques est à disposition : les *plots* de diagnostic (par exemple le *Q'Q plot*, les graphiques de symétrie ou de linéarité), les graphiques de résidus, les aides à la transformation ou encore les méthodes de lissage (par exemple LOWESS) et d'autres méthodes qui visent à faciliter la lecture des graphiques ou des tableaux d'un point de vue analytique.

## ***Visualisations multivariées***

Pour tenter d'appréhender plusieurs dimensions, les statisticiens ont déployé des trésors d'imagination. Des graphiques qui tentent de broser le portait d'observations ou de variables (par exemple les faces de Chernov<sup>1</sup> ou les graphiques en étoile) aux méthodes dimensionnelles (par exemple composantes principales), la boîte à outils du statisticien déborde de propositions qui gagnent encore en puissance, si elles sont combinées dans un environnement véritablement interactif.

1. Le représentant le plus connu de cette famille de graphiques, mais peut-être également le plus critiquable.

## *Graphiques dynamiques*

La manipulation directe d'objets graphiques sur un écran rajoute une toute nouvelle dimension à l'analyse graphique. Mais comment décrire verbalement l'expérience visuelle du choix<sup>1</sup> d'un groupe d'observations isolées sur un graphique de dispersion qui est instantanément repéré par une couleur distincte sur tous les autres graphiques affichés sur l'écran (histogrammes, graphiques de dispersion ou encore arbres de classification) ?

## *Conclusion*

Les techniques visuelles et l'analyse statistique véritablement interactive constituent sans doute des atouts majeurs de la pratique statistique, mais exigent une redéfinition du rôle de l'analyste et requièrent d'autres attitudes et aptitudes. ... et expérience que la qualité scientifique et la cohérence d'un résultat ne peut être garantie que par une démarche explicite, responsable et consciente et par un public (scientifique ou non) attentif et critique dissipant les illusions que l'on aurait pu avoir à l'égard des outils statistiques.

1. À l'aide d'une souris ou d'un autre instrument de pointage.

---

## BIBLIOGRAPHIE

---

- BERTIN, Jacques (1973), *Sémiologie graphique. Les diagrammes, les réseaux, les cartes*. Paris : MONTON/GAUTHIER VILLARS.
- CHAMBERS, J.M. et al. (1983), *Graphical Methodes for Data Analysis*, Boston : Duxbury Press.
- CLEVELAND, WILLIAM S. (1993), *Visualizing Data*, Murray Hill : AT & T Bell Laboratoires.
- CLEVELAND, WILLIAM S. (1986), *The Collected Works of John Tukey, Volume V : Graphics 1965-1985*. MONTEREY : WADSWORTH & BROOKS/Cole.
- SPENCE, Ian, LEWANDOWSKY, STEPHAN (1990), *Graphical Perceptions*, In : FOX, JOHN ; LONG SCOTT J. *Modern Methods of Data Analysis*, Newbury Park : Sage.
- TIERNEY, Luke (1990), *LISP STAT : An Object-Oriented Environment for Statistical Computing and Dynamic Graphics*, New York : Wiley.
- TUFTE, Edward. R. (1983), *The Visual Display of Quantitative Information*, Cheshire : Graphics Press.
- TUFTE, Edward, R. (1990), *Envisioning Information*, Cheshire : Graphics Press.
- TUKEY, J.W. (1977), *Exploratory Data Analysis*, Reading : Addison & Wesley.