



[Visualisation de données]

# 6 Visualisation de données

*La source principale de ce chapitre est « Effective Visual Communication for the Quantitative Scientist » (2019), par M. Vandemeulebroecke, M. Baillie, A. Margolskee et B. Magnusson. La seconde source est le magnifique ouvrage de Edward Tufte ↗, « The Visual Display of Quantitative Information ». Je vous recommande également « #Makeover Monday », d'Andy Kriebel et Eva Murray.*

La communication visuelle est une compétence essentielle pour un statisticien/scientifique des données. Elle est présente autant à l'écrit qu'à l'oral. Cette communication efficace sert à mieux comprendre les données, ce qui est la clef de notre travail. Le papier de Vandemeulebroecke propose trois lois de communication visuelle efficace : avoir un objectif clair, montrer les données clairement, et rendre le message évident.

6.1 Excellence graphique .....	34
6.2 Représenter les données les révèles .....	35
6.3 Exemples historiques et particuliers de représentations graphiques .....	39
6.4 Illustrations de ce qu'il ne faut pas faire .....	54
6.5 Loi 1 : avoir un objectif clair. ....	67

6.6 Loi 2 : montrer les données clairement. ....	71
6.7 Loi 3 : rendre le message évident. ....	79

## 6.1 Excellence graphique

L'excellence d'une représentation statistique est constitué d'idées complexes communiquées avec clarté, précision et efficacité. La représentation doit :

- montrer les données;
- inciter le spectateur à réfléchir à la substance plutôt qu'à la méthodologie, au design graphique ou à la méthode utilisée dans la production;
- éviter de déformer ce que disent les données;
- présenter de nombreux chiffres dans un espace réduit;
- rendre cohérents de grands ensembles de données;
- encourager l'œil à comparer différentes données;
- révéler les données à plusieurs niveaux de détail, de la vue d'ensemble à la structure fine;
- servir un objectif clair : description, exploration, tabulation, décoration;
- être étroitement intégrées à la description statistique et narrative d'un ensemble de données.

## 6.2 Représenter les données les révèles

Considérons cet exemple d'Anscombe (Am. Stat., 1973)<sup>1</sup>

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

<sup>1</sup> Francis J. Anscombe « Graphs in Statistical Analysis », The American Statistician 27 pp. 17-21 (1973) ↗

En R (copiez/collez/commentez):

```
x1<-c(10.0,8.0,13.0,9.0,11.0,14.0,6.0,4.0,12.0,7.0,5.0)
y1<-c(8.04,6.95,7.58,8.81,8.33,9.96,7.24,4.26,10.84,4.82,5.68)

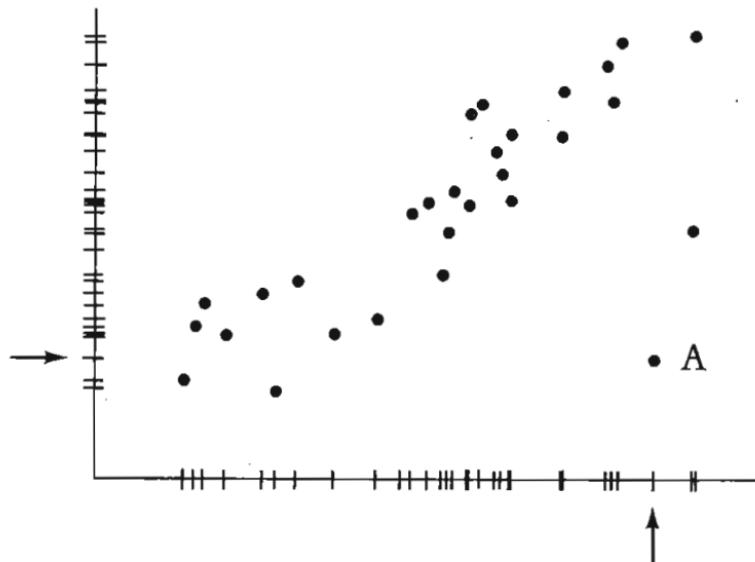
x2<-c(10.0,8.0,13.0,9.0,11.0,14.0,6.0,4.0,12.0,7.0,5.0)
y2<-c(9.14,8.14,8.74,8.77,9.26,8.10,6.13,3.10,9.13,7.26,4.74)

x3<-c(10.0,8.0,13.0,9.0,11.0,14.0,6.0,4.0,12.0,7.0,5.0)
y3<-c(7.46,6.77,12.74,7.11,7.81,8.84,6.08,5.39,8.15,6.42,5.73)

x4<-c(8.0,8.0,8.0,8.0,8.0,8.0,8.0,19.0,8.0,8.0,8.0)
y4<-c(6.58,5.76,7.71,8.84,8.47,7.04,5.25,12.50,5.56,7.91,6.89)

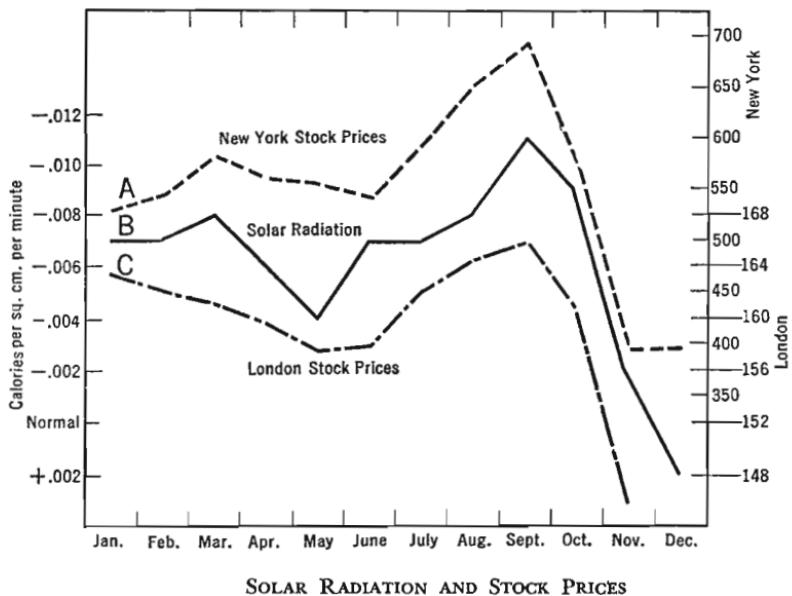
lm(y1~x1) ; lm(y2~x2) ; lm(y3~x3) ; lm(y4~x4)
```

Remarquez dans cet exemple<sup>2</sup> comment le point *A* ne se remarque pas dans les distributions marginales:



<sup>2</sup> Stephen S. Brier et Stephen E. Fienberg, « Recent Econometric modelling of Crime and Punishment: Support for the Deterrence Hypothesis », 1980.

Évidement, un « beau graphique bien fait » n'est pas un gage de qualité non plus:



SOLAR RADIATION AND STOCK PRICES

- A. New York stock prices (Barron's average). B. Solar Radiation, inverted, and C. London stock prices, all by months, 1929 (after Garcia-Mata and Shaffner).

## 6.3 Exemples historiques et particuliers de représentations graphiques

Cette section présente d'autres représentations graphiques, dont certaines sont célèbres de par leur histoire, d'autres de simples curiosités. Quelques hyperliens sont parfois présents si vous voulez en savoir plus.

L'élément central de cette section est d'établir qu'une image vaut vraiment mille mots, et permet parfois de comprendre bien mieux que de longs tableaux de chiffres.

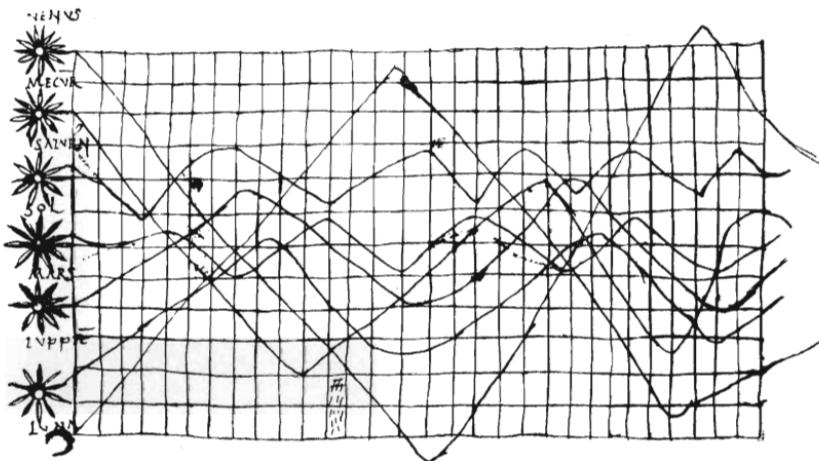


Chambers et al. : “There is no single statistical tool that is as powerful as a well-chosen graph.”<sup>3</sup>

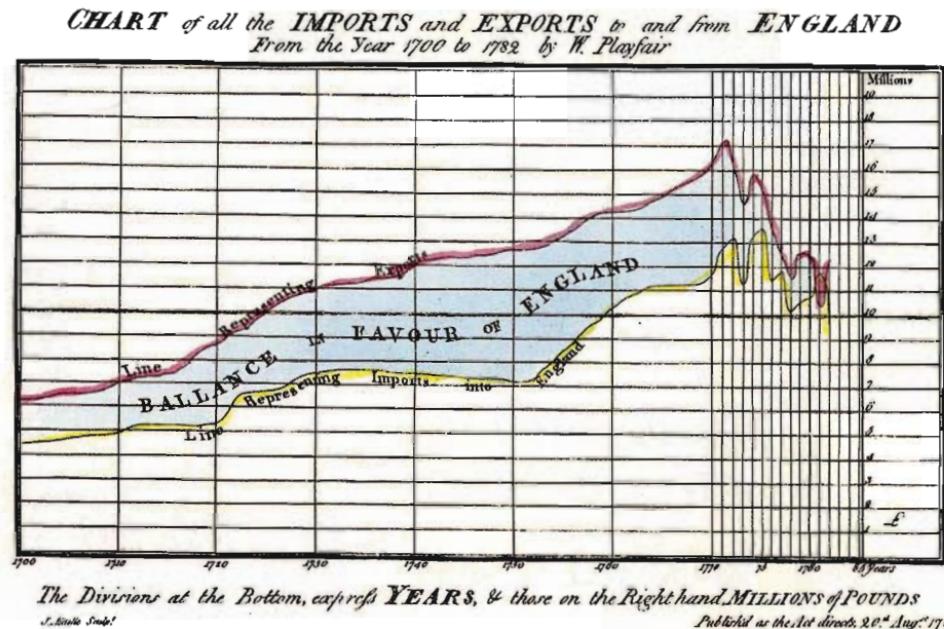
---

<sup>3</sup> Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A. Graphical Methods for Data Analysis (Wadsworth, Pacific Grove, CA, 1983).

L'illustration suivante du dixième (ou peut-être du onzième) siècle représente des inclinaisons des orbites planétaires en fonction du temps, et fait apparemment partie d'un texte destiné aux écoles des monastères, est le plus ancien exemple connu d'une tentative de représentation graphique de valeurs changeantes. Il apparaît comme une merveille mystérieuse et isolée dans l'histoire des graphiques de données, puisque le prochain graphique existant d'une série chronologique tracée apparaît quelque 800 ans plus tard.



William Playfair, un économiste, propose la première vrai représentation d'une série chronologique(1786) [Wiki](#) [Wiki](#).



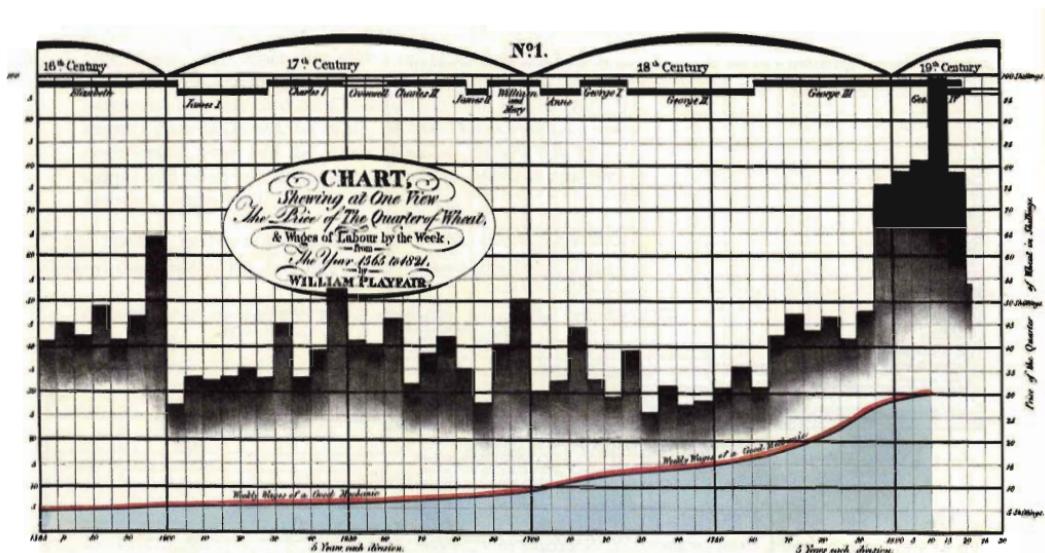
Extrait de « The Commercial and Political Atlas » (1786):

« Information, that is imperfectly acquired, is generally as imperfectly retained; and a man who has carefully investigated a printed table, finds, when done, that he has only a very faint and partial idea of what he has read; and that like a figure imprinted on sand, is soon totally erased and defaced. The amount of mercantile transactions in money, and of profit or loss, are capable of being as easily represented in drawing, as any part of space, or as the face of a country; though, till now, it has not been attempted. Upon that principle these Charts were made; and, while they give a simple and distinct idea, they are as near perfect accuracy as is any way useful. On inspecting any one of these Charts attentively, a sufficiently distinct impression will be made, to remain unimpaired for a considerable time, and the idea which does remain will be simple and complete, at once including the duration and the amount. »

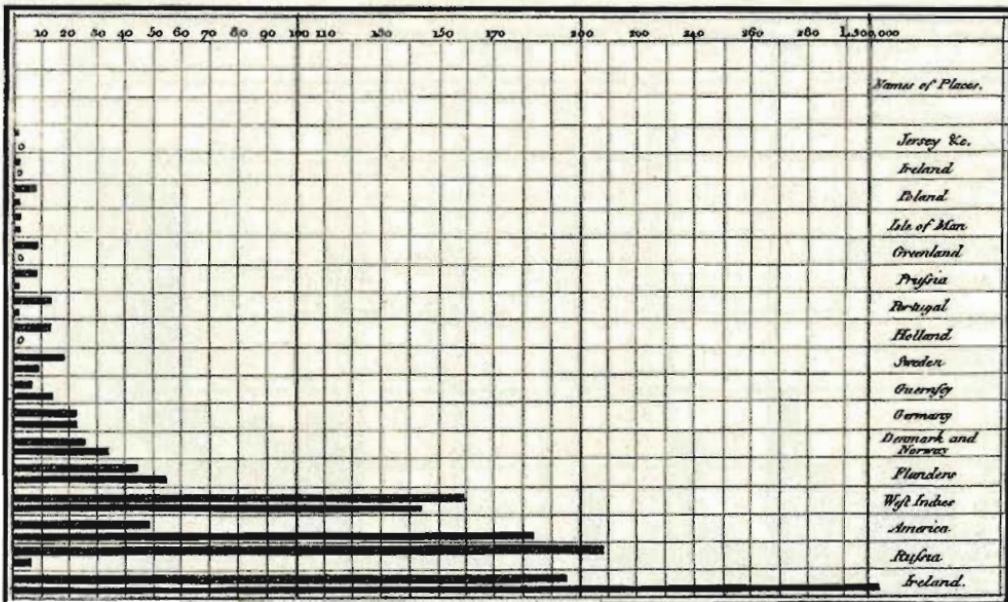
Dans la langue de Molière:

« L'information, qui est imparfaitement acquise, est généralement aussi imparfaitement retenue ; et un homme qui a examiné soigneusement un tableau imprimé, constate, une fois terminé, qu'il n'a qu'une idée très faible et partielle de ce qu'il a lu ; et que, comme une figure imprimée sur le sable, elle est bientôt totalement effacée. Le montant des transactions commerciales en argent, ainsi que les profits et les pertes, peuvent être aussi facilement représentés en dessin que n'importe quelle partie de l'espace ou que la surface d'un pays, bien que cela n'ait pas été tenté jusqu'à présent. C'est sur ce principe que ces cartes ont été réalisées et, bien qu'elles donnent une idée simple et distincte, elles sont aussi proches de la précision parfaite que possible. En examinant attentivement l'une ou l'autre de ces cartes, on en retire une impression suffisamment distincte pour qu'elle reste intacte pendant un temps considérable, et l'idée qui reste est simple et complète, comprenant à la fois la durée et la quantité. »

Ici, Playfair trace trois séries chronologiques parallèles : les prix, les salaires et les règnes des rois et reines britanniques:



Exports and Imports of SCOTLAND to and from different parts for one Year from Christmas 1780 to Christmas 1781.



The Upright divisions are Ten Thousand Pounds each. The Black Lines are Exports the Ribbed lines Imports.

Published at the Author's Office June 7<sup>th</sup> 1782 by W<sup>m</sup> Playfair

No 100, 1<sup>st</sup> Strand, London.

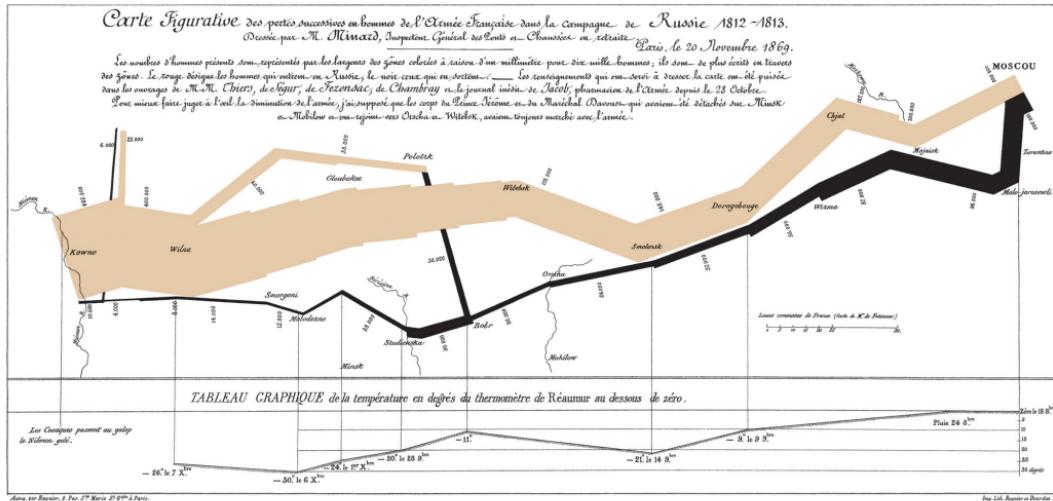


« This Chart is different from the others in principle, as it does not comprehend any portion of time, and it is much inferior in utility to those that do; for though it gives the extent of the different branches of trade, it does not compare the same branch of commerce with itself at different periods; nor does it imprint upon the mind that distinct idea, in doing which, the chief advantage of Charts consists: for as it wants the dimension that is formed by duration, there is no shape given to the quantities. »

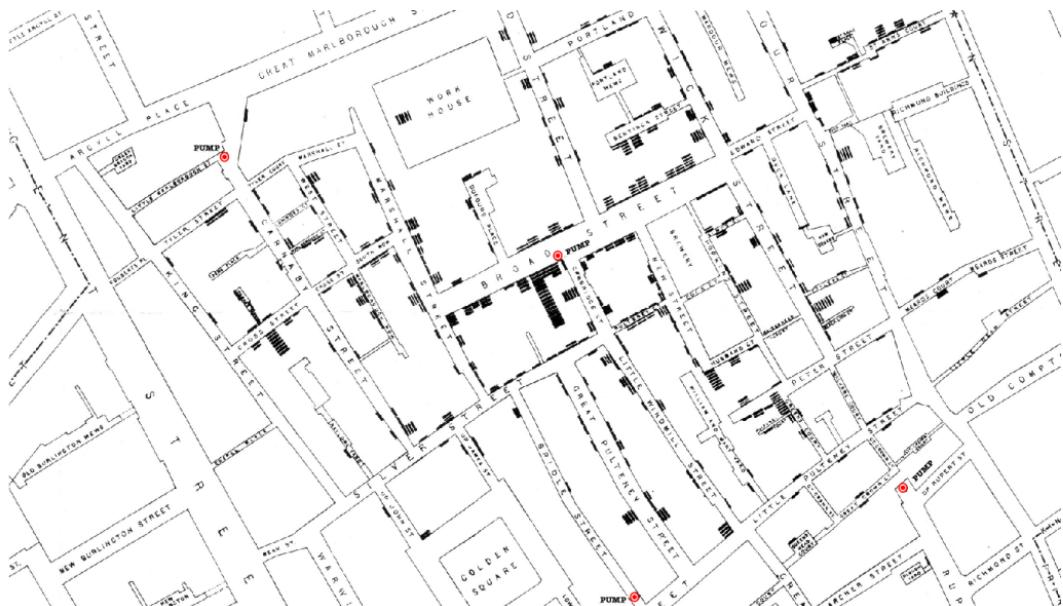


« Ce tableau diffère des autres par son principe, car il ne comprend aucune portion de temps, et son utilité est bien inférieure à celle des autres ; car s'il donne l'étendue des différentes branches du commerce, il ne compare pas la même branche du commerce avec elle-même à différentes époques ; il n'imprime pas non plus dans l'esprit cette idée distincte, en quoi consiste le principal avantage des tableaux : car comme il manque la dimension formée par la durée, il ne donne pas de forme aux quantités. »

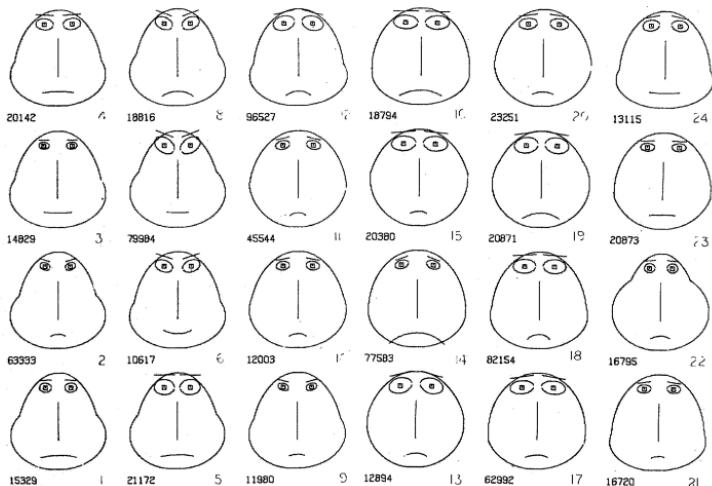
Charles Joseph Minard est resté à l'histoire comme le concepteur de représentations graphiques originales, dont la « Carte figurative des pertes successives en hommes de l'armée française dans la campagne de Russie 1812-1813 » [Wiki](#)



En 1854 se produit à Londres une épidémie de Cholera. John Snow, un médecin, à l'idée de faire une représentation graphique qui permettra d'identifier la source de l'épidémie ([Wiki](#), [Wiki](#)).

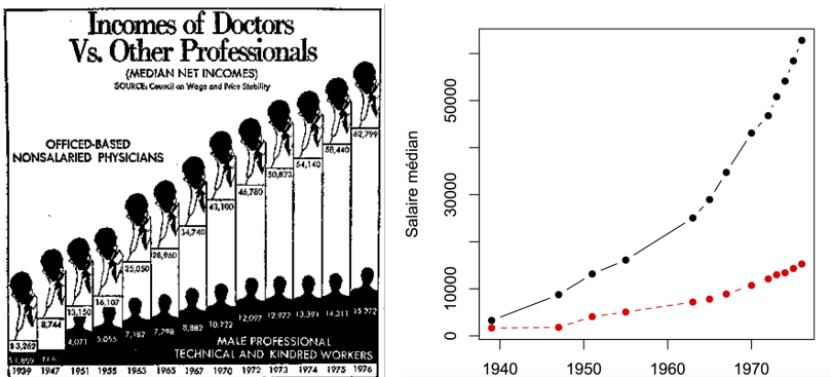


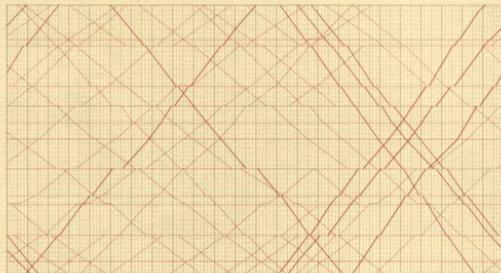
Dans un article de 1973, Herman Chernoff propose une représentation graphique multidimensionnelle de données en utilisant des visages, puisque nous sommes habitués à différencier des visages. Ainsi, chaque caractéristique du visage est une représentation d'une variable. Voici un exemple tiré de l'article. Notez que la représentation, bien qu'intéressante, est peu utilisée ( Wiki ↗, Wiki ↗, Journal of the American Statistical Association).



Une représentation graphique est si puissante qu'il peut être tentant de tourner parfois les coins ronds... Voici un exemple tiré du livre d'Edward Tufte.

L'exemple est paru dans le Washington Post. Observez bien comme l'augmentation du salaire des médecins paraît linéaire dans la parution originale (à gauche), alors qu'elle semble exponentielle à droite (représentation juste); pourquoi ?

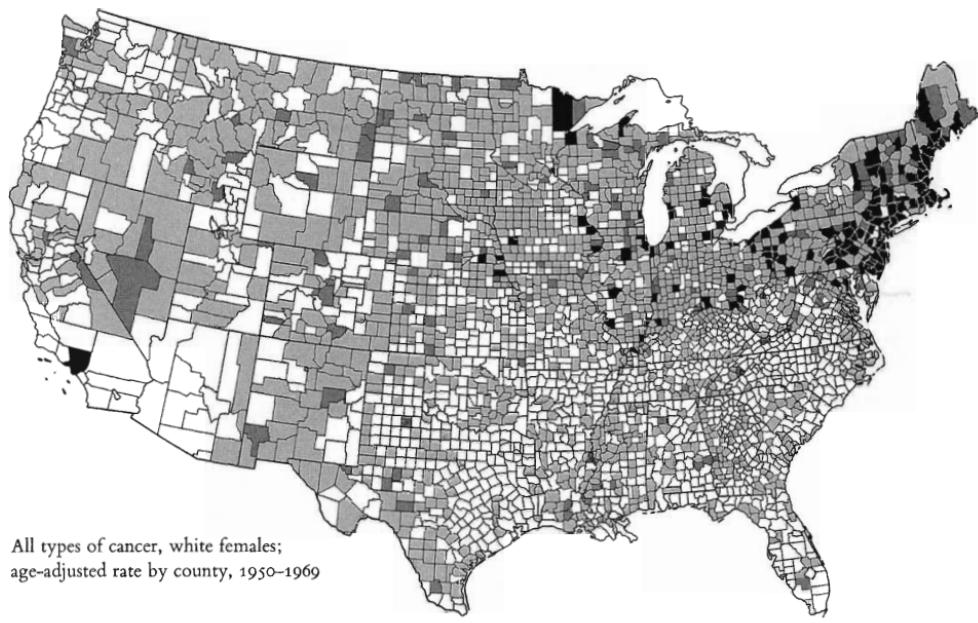


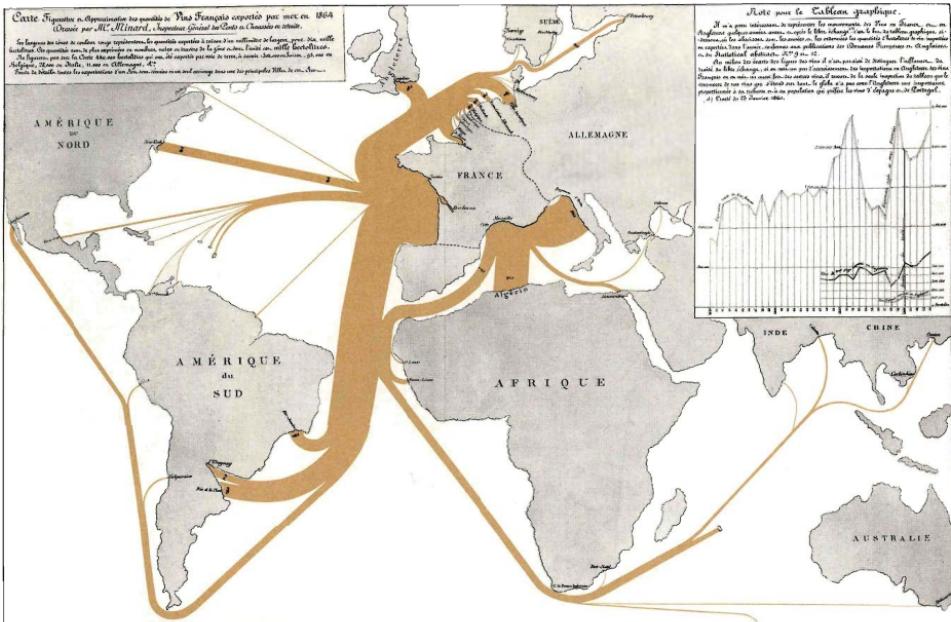


The Visual Display  
of Quantitative Information

EDWARD R. TUFTE

Illustration de la puissance d'une représentation spatiale : représentation de taux de mortalité ajusté pour l'âge pour différents cancers pour 3056 contés des États-Unis.





Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard*, 1845-1869, a portfolio of his work held by the Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris.

## 6.4 Illustrations de ce qu'il ne faut pas faire

Le diagramme en trois dimensions présente un défaut majeur: il ne représente pas la même information dépendant de l'axe selon lequel il est présenté. De plus, la tranche du diagramme fausse visuellement l'image globale : dans les diagrammes de la figure, la proportion « fr. et an. » est de 53% à gauche, en 2 dimensions, et de 66.8% à droite en 3 dimensions avec la tranche.

Bien que les diagrammes circulaires soient souvent utilisés dans les médias et dans le monde des affaires, ce sont de très mauvaises représentations; en effet, il a été montré qu'il est difficile de comparer les différentes sections d'un diagramme circulaire, et (presque) impossible de comparer des sections à travers différents graphiques. Ceci est illustré sur la figure 6.2, où l'on voit aisément que les diagrammes en bâtons sont plus utiles que les diagrammes circulaires si on veut comparer les catégories d'une même variable: pour les trois répartitions considérées dans la figure, les diagrammes circulaires se ressemblent tous, alors que les diagrammes en bâtons montrent clairement les différences. Bref, tout ceci explique pourquoi vous ne trouverez que rarement ce type de graphique dans le monde scientifique.

Pour conclure voici une unique figure avec les quatre représentations (fig. 6.1).

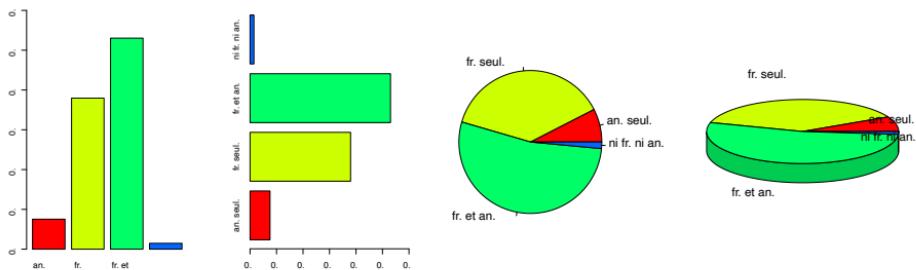


Figure 6.1 Les différentes représentations possibles.

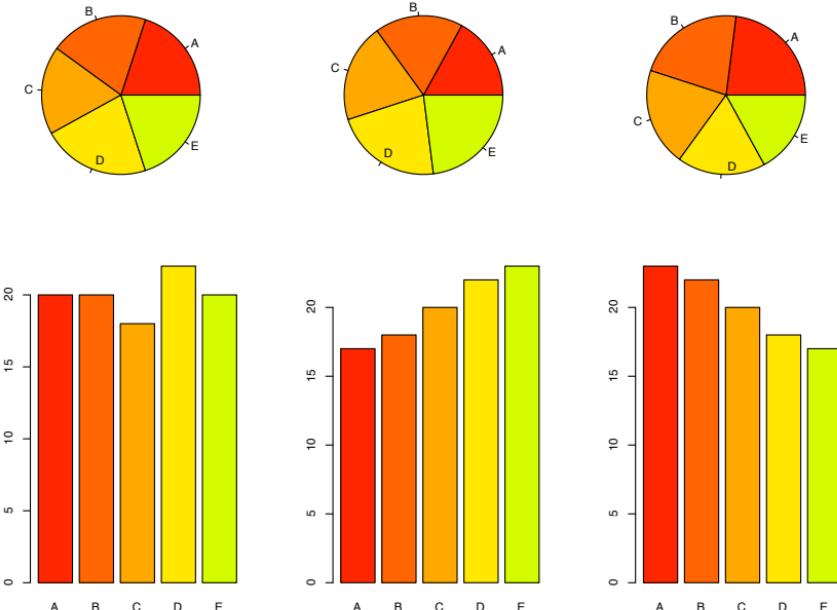
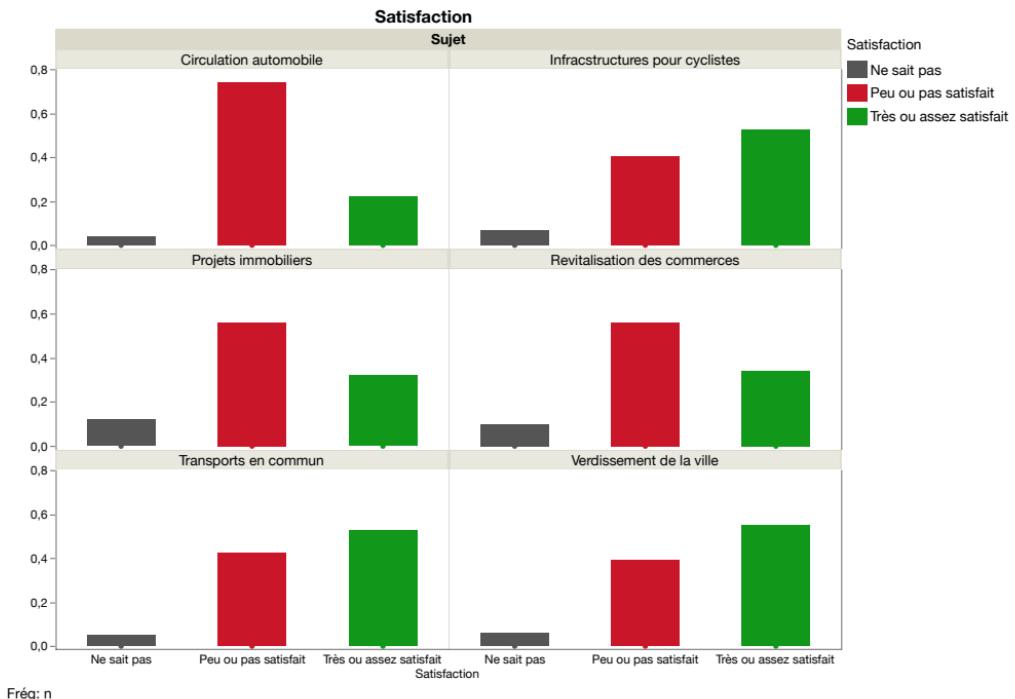
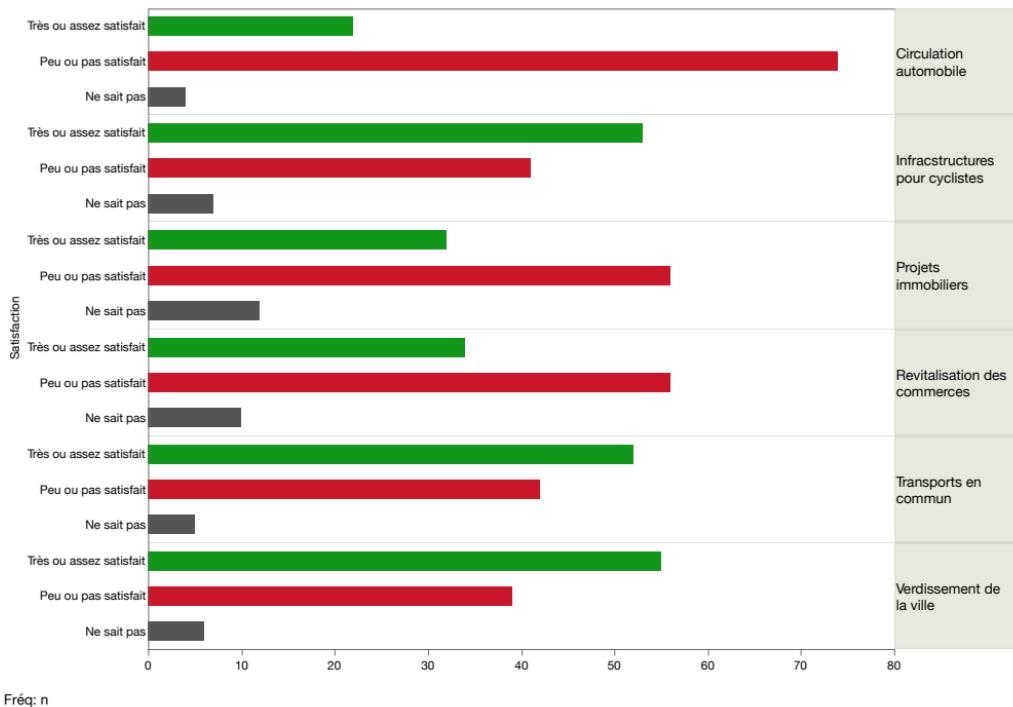


Figure 6.2 Exemples de diagrammes en bâtons avec diagrammes circulaires associés sur 3 ensembles (fictifs) de données.

## La Presse (Aut. 2020):







La Presse (Février 2024):

The screenshot shows a news article from La Presse+ about the results of a vote by the Fédération autonome de l'enseignement (FAE). The article title is "L'entente de principe adoptée à l'arraché". It features a photo of a protest or rally with red flags. The main content is a chart titled "RÉSULTATS DES VOTES DE LA FAE" comparing "ONT VOTÉ POUR" (yellow) and "ONT VOTÉ CONTRE" (red) for various teacher unions.

Syndicat	Ont voté Pour (%)	Nombre de membres	Ont voté Contre (%)	Nombre de membres
Syndicat de l'enseignement de la Pointe-de-l'Île (SEPI)	58 %	4384 MEMBRES	68 %	6346 MEMBRES
Alliance des professeures et professeurs de Montréal	52 %	11 343 MEMBRES	72 %	5259 MEMBRES
Syndicat des enseignants de l'Outaouais (SEO)	57 %	6582 MEMBRES	58 %	700 MEMBRES
Syndicat de l'enseignement de l'Ouest de Montréal (SEOM)	51 %	7266 MEMBRES	60 %	6400 MEMBRES
Syndicat des enseignants de la Haute-Yamaska (SEHY)	50.58 %	2001 MEMBRES		

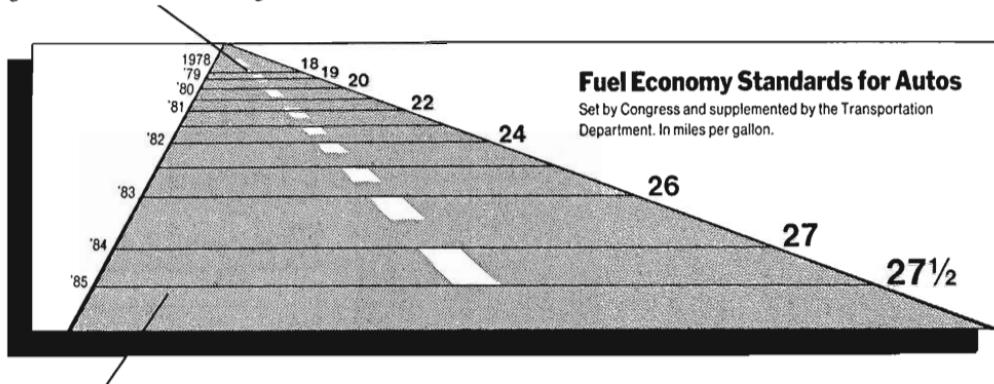
A note at the bottom right asks: "QU'EST-CE QUE LA « DOUBLE MAJORITÉ » ?"

La représentation des nombres, telle que mesurée sur la surface d'un graphique, doit être directement proportionnelle aux quantités numériques représentées. Tuft propose le facteur mensonge (*The Lie Factor*) pour mesurer les distorsions :

$$\text{Facteur mensonge} = \frac{\text{Taille de l'effet illustrée dans le graphique}}{\text{Taille de l'effet dans les données}}.$$

Un exemple de calcul du « Facteur Mensonge » :

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

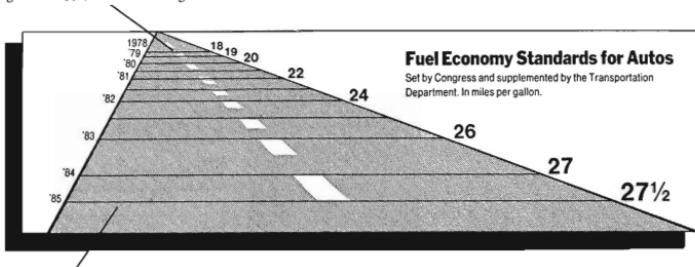
$$\text{Effet réel: } \frac{27.5 - 18}{18} \times 100 = 53\%,$$

$$\text{effet représenté: } \frac{5.3 - 0.6}{0.6} \times 100 = 783\%$$

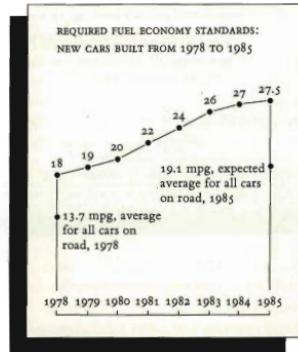
$$\text{Facteur mensonge} = \frac{\text{Taille de l'effet illustrée dans le graphique}}{\text{Taille de l'effet dans les données}} = \frac{783}{53} = 14.8.$$

## Version corrigée !

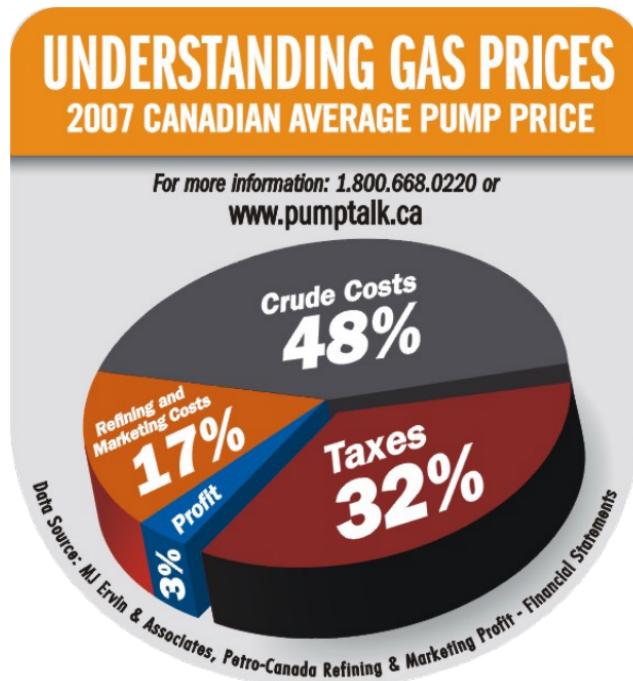
This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.



Voici une représentation qui aurait un score élevé !



Pour conclure, voici les principes de l'intégrité graphique proposés par Tufte :

- La représentation des nombres, telle que mesurée physiquement sur la surface du graphique lui-même, doit être directement proportionnelle aux quantités numériques représentées.
- Un étiquetage clair, détaillé et complet doit être utilisé pour éviter toute distorsion ou ambiguïté du graphique. Écrivez les explications des données sur le graphique lui-même. Indiquez les événements importants dans les données.
- Montrez la variation des données, pas celle de la conception.
- Dans les séries chronologiques sur la monnaie, les unités de mesure monétaire déflatées et normalisées sont presque toujours préférables aux unités nominales.
- Le nombre de dimensions porteuses d'information (variables) représentées ne doit pas dépasser le nombre de dimensions dans les données.
- Les graphiques ne doivent pas citer des données hors de leur contexte.

Ainsi, rappelons nous de la proposition de Vandemeulebroecke :

*Proposition: les trois lois de la communication visuelle.*

- *avoir un objectif clair (loi 1),*
- *montrer les données clairement,*
- *rendre le message évident.*

## 6.5 Loi 1 : avoir un objectif clair.

Quel est le but de cette présentation ou de cette communication ? Selon Doumont, il s'agit de la "loi zéro" de la communication professionnelle, "un principe si évident qu'il a longtemps été négligé". Soyez clair et explicite sur l'objectif que vous souhaitez atteindre. S'agit-il d'explorer certaines données, de transmettre une analyse déductive, de délivrer un message, de convaincre un public ou de soutenir une décision ? Il peut s'agir d'un mélange de ces objectifs - par exemple, même des graphiques exploratoires apparemment simples doivent servir à prendre une décision (peut-être implicite) (par exemple, sur la manière de poursuivre l'exploration). Chaque graphique, et plus généralement chaque communication, doit être adapté à son objectif spécifique.

Il est utile de définir la question scientifique à laquelle vous essayez de répondre, idéalement en discutant avec vos partenaires, et de la mettre par écrit de manière explicite. Essayez de ne pas regarder de données avant d'avoir formulé votre question. C'est le concept des "visualisations basées sur les questions" : laisser la question scientifique déterminer quelles données afficher et comment. (Par exemple, combinez des données provenant de différents domaines si cela permet de répondre efficacement à la question. Ne produisez pas uniquement des sorties standard par domaine de données - l'affichage doit être déterminé par la question à laquelle il répond, et non par la manière dont les données sont organisées). Comme le dit Diggle, nous "analysons les problèmes, pas les données". Cela ne signifie pas que la question ne peut pas être affinée après avoir vu les données. Nous pouvons très bien itérer sur l'espace du problème et l'espace de la solution, pour autant que nous le fassions consciemment. Senn illustre de nombreux exemples de questions de recherche mal formulées. L'une d'entre elles consiste souvent à se concentrer sur la mauvaise comparaison, par exemple en comparant une valeur post-traitement à la valeur de référence correspondante plutôt qu'à la valeur sous un traitement de contrôle. La plupart des graphiques quantitatifs présentent des comparaisons, et il est toujours utile de se demander "comparé à quoi ?". Si la comparaison n'est pas claire pour l'auteur, elle ne le sera pas non plus pour le lecteur.

Une partie de cette première loi (et de la troisième, voir ci-dessous) consiste également à bien connaître son public. Ensuite, il faut s'adapter à ce public. Ne supposez pas qu'il s'adaptera à vous. Vous ne pouvez pas contrôler votre public, mais vous pouvez contrôler les informations et les messages que vous lui transmettez, et comment. Votre public, c'est vous (qui essayez de voir les tendances dans les données), c'est vous dans quelques années (qui essayez de vous souvenir de ce que vous avez fait), ce sont des experts quantitatifs tels que vos pairs (intéressés par vos méthodes), des experts en la matière (avides de votre message principal), des décideurs (uniquement les titres), ou un mélange de tout cela ? Votre communication visuelle devra être différente en conséquence. Votre communication (intrigue, présentation et rapport) est destinée au public, pas à vous.

La clarté de l'objectif et de la question scientifique d'intérêt permettra de choisir les méthodes quantitatives appropriées pour y répondre. Ceci, ajouté à la clarté de votre public, permettra de définir les messages clés et la manière de les transmettre. (Sur l'aspect de la diffusion, voir également la loi 3 ci-dessous).

Il convient de noter que cette première loi est si importante qu'elle peut parfois défier d'autres bons principes. Si votre objectif principal est d'attirer l'attention, vous pouvez choisir une représentation graphique emblématique qui le fait bien, même si elle va à l'encontre de certaines des recommandations formulées plus loin. Cependant, vous ne devez jamais déformer les données.

## 6.6 Loi 2 : montrer les données clairement.

Montrer les données, en choisissant la représentation la plus simple. Choisir les designs simples plutôt que trop riches. Albert Einstein : Make your plot « as simple as it can be, but not simpler ». Ne vous trompez pas : c'est difficile de faire simple ! Maximisez le rapport données-encre<sup>4</sup>: enlevez le bruit, tout ce qui est inutile.



Figure 6.3 Exemples de représentations graphiques.

<sup>4</sup> data-to-ink ratio (also “data density index”)

Loi 2 : montrer les données clairement. Les diagrammes circulaires et en beignet des panneaux (a) et (b) ne permettent pas de voir l'ordre de grandeur de certains des segments. L'œil doit comparer des surfaces, des longueurs courbes (du contour) ou des angles, des attributs graphiques qui ne sont pas faciles à décoder. Le graphique en forme de beignet omet même les angles. Le graphique en mosaïque du panneau (c) ne repose que sur les surfaces ; là encore, il est difficile de déterminer l'ordre de grandeur. Il est préférable d'utiliser des longueurs avec une ligne de base commune ou des positions sur une échelle commune, comme dans un diagramme en barres ou un graphique en points. Le diagramme en barres du panneau (d), cependant, introduit une fausse dimension, qui n'est pas nécessaire et rend difficile le décodage des valeurs numériques à partir de la hauteur des barres. Les panneaux (e) et (f) sont suffisamment simples et montrent clairement les données. Ils classent également les données par ordre de grandeur pour faciliter encore plus la comparaison. Le graphique à points du panneau (f) utilise une quantité minimale d'encre et attire l'attention sur la position des points ; c'est la façon la plus efficace d'afficher ces données.

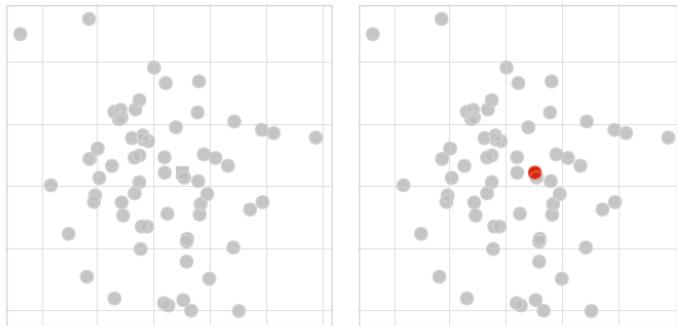
C'est la maxime de Tufte : "Avant tout, montrez les données". Montrez-les avec précision et clarté. Cette loi a plusieurs visages :

Simplifier ! "Simplifier pour clarifier". La tâche principale des scientifiques quantitatifs est de rendre simple ce qui est complexe : révéler la structure des données grâce à des modèles, faire des déductions grâce à des analyses, et distiller et transmettre les conclusions par le biais de la communication (visuelle). Choisissez le type de graphique approprié le plus simple ; préférez les modèles familiers aux modèles fantaisistes (voir également le classement de l'efficacité de Cleveland-McGill dans la loi 3 ci-dessous). Évitez les fausses dimensions et les camemberts. Faites en sorte que votre graphique soit "aussi simple que possible, mais pas plus simple" (attribué à Albert Einstein ; également "rasoir d'Occam" ou loi de parcimonie). "Comprenez, modifiez et simplifiez l'information et la conception en pensant à vos lecteurs.

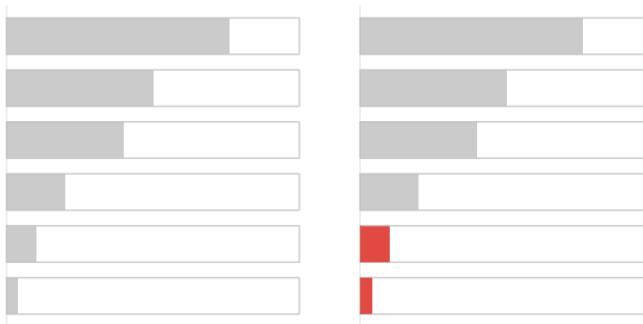
Maximiser le rapport données/encre (ou "indice de densité des données ") dans les limites du raisonnable. Maximiser le signal par rapport au bruit en supprimant le bruit : supprimer tout ce qui détourne l'attention de l'objectif du graphique. Rien n'est neutre : le choix des symboles ou des couleurs, l'arrière-plan, les polices de caractères, le style des lignes et les annotations. Ces éléments sont du bruit s'ils ne servent pas un objectif clair. Choisissez-les avec sagesse et parcimonie ; faites ressortir les données. Ne vous fiez pas aux valeurs par défaut des logiciels graphiques. Souvent, une utilisation intelligente de l'espace blanc peut structurer un affichage mieux que beaucoup d'encre. (Il en va de même pour les tableaux : ils sont souvent structurés de manière plus efficace en réservant les lignes noires au sens horizontal et en utilisant un alignement simple dans le sens vertical). N'encombrez jamais votre graphique avec de la "camelote " ("chart junk").

Afficher directement les données pertinentes. Dans un flux de travail quantitatif, cela signifie souvent qu'il faut examiner les données brutes et ne pas se contenter de statistiques sommaires. Cabanski et al. l'illustrent avec neuf ensembles de données qui présentent des schémas complètement différents malgré des moyennes marginales, des écarts types et des coefficients de corrélation identiques (voir également Anscombe et Matejka et Fitzmaurice). Demandez-vous quelle est la meilleure façon de résumer les caractéristiques pertinentes des données ; il se peut que ce ne soit pas la moyenne +/- l'écart-type. Lors de l'ajustement d'un modèle aux données pour en tirer des conclusions ou faire des prédictions, les quantités dérivées du modèle peuvent être les "données pertinentes" à afficher. Dans ce cas, une représentation graphique des données brutes peut être trompeuse si elle ne tient pas compte de covariables importantes. Dans les simulations (par exemple, pour la planification de l'étude), cette deuxième loi peut être appliquée aux données simulées et/ou aux quantités dérivées, selon le cas. Dans une communication finale, présentez de manière concise ce qui soutient le mieux votre message (voir également la loi 3 ci-dessous).

## La puissance du traitement préattentif.

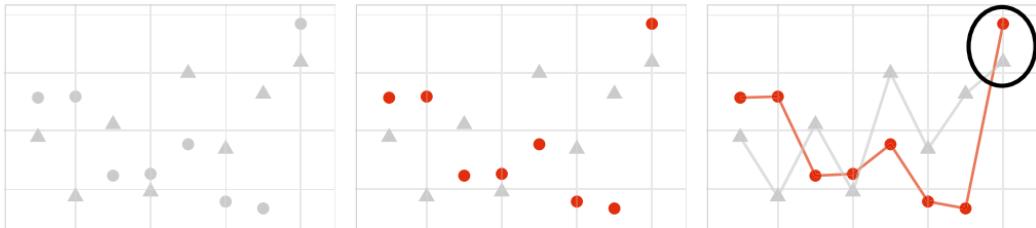


Un point de données est marqué par un symbole différent (à gauche) ou une couleur différente (à droite). Discerner le symbole différent requiert de l'attention, alors que la couleur différente "saute aux yeux" de manière préattentive.



Panneau de gauche, nous sommes attirés par une comparaison de longueurs. Dans le panneau de droite, nous introduisons la couleur pour attirer l'œil sur les deux barres inférieures en premier. Bien que la longueur soit déjà un attribut préattentif important, la couleur l'est encore plus. Nous pouvons l'utiliser pour guider l'attention des spectateurs à travers une intrigue et leur permettre de suivre l'histoire que nous voulons transmettre.

### Les principes du regroupement visuel:



Les principes du regroupement visuel : clôture, connexion, proximité et similarité. Le premier panneau montre deux groupes de points, identifiés par la similitude (des symboles de tracé). Ici, la proximité est une conséquence de l'affichage des données et ne peut être choisie délibérément. Le deuxième panneau introduit de la couleur pour renforcer le regroupement par similarité (en utilisant une redondance efficace et un traitement préattentif). Le troisième panneau regroupe les points de manière encore plus efficace grâce à des lignes de connexion (en plus des symboles et des couleurs du graphique). L'attention est particulièrement attirée sur une paire de points par une ellipse qui les entoure.

## 6.7 Loi 3 : rendre le message évident.

Si la deuxième loi était axée sur les données (dans le but de minimiser le bruit), la troisième concerne le message (et l'amplification du signal). Cela suppose que vous ayez un message à transmettre et que ce message soit clair, au moins pour vous-même. En cas de doute, revenez à la première loi.

La troisième loi impose de rendre son message aussi évident que possible. Citant Krzywinski et Cairo, "il est risqué d'inviter les lecteurs à tirer leurs propres conclusions". Ne vous contentez pas de rendre votre message facile à obtenir. Faites en sorte qu'il soit impossible de le manquer. Cela va au-delà des éléments graphiques et concerne tous les aspects de la communication.

La clarté de votre public, déjà mentionnée dans la première loi, est également une condition préalable à la troisième. Elle est nécessaire pour définir le message à transmettre (loi 1) et pour adapter son mode de diffusion (loi 3).

Voici quelques exemples concrets de la troisième loi :

- Choisissez judicieusement la manière de coder les données que vous affichez. La couleur et la surface permettent d'attirer l'attention, mais un observateur peut décoder les positions sur une échelle commune beaucoup plus facilement et avec plus de précision. Considérez le classement de l'efficacité des attributs graphiques pour le codage des valeurs numériques, tel que proposé par Cleveland, et Cleveland et McGill.
- Exploiter autant que possible le traitement préattentif.
- Certaines caractéristiques graphiques "sautent aux yeux", tandis que d'autres nécessitent une inspection minutieuse. Tenez-en compte dans votre choix de codage des données (classement d'efficacité Cleveland-McGill ; voir ci-dessus) et dans votre choix de symboles, de couleurs, de types de lignes, etc.
- Évitez le calcul mental. Si les différences ou les ratios sont le principal intérêt, montrez-les directement. Si les valeurs brutes et les différences sont toutes deux intéressantes, envisagez de les montrer toutes les deux. Utiliser des plages d'axes identiques sur plusieurs graphiques ou panneaux, le cas échéant.

- Exploiter les principes de regroupement visuel. Les entités graphiques sont regroupées le plus efficacement en fonction de la clôture, de la connexion, de la proximité et de la similarité (dans cet ordre). En d'autres termes, les objets similaires sont perçus comme appartenant au même groupe, tout comme les objets proches les uns des autres, reliés par des lignes ou enfermés dans un sous-espace commun. Voir la figure 3 pour une illustration : ces mécanismes peuvent fournir des informations contextuelles à un graphique d'une manière simple mais puissante.
- Réduire au minimum le mouvement des yeux de l'observateur. Placer les éléments à comparer à proximité les uns des autres. Préférer l'étiquetage direct à la légende.
- Attirer l'attention du lecteur sur les points principaux. Utiliser des éléments graphiques appropriés (par exemple, surlignage en gras ou en couleur, lignes de référence, encerclement, etc.) Faire suivre d'un étiquetage explicite (par exemple, "le traitement A est plus performant que le traitement B de X %").
- Ajouter des informations significatives à un graphique pour raconter l'histoire dans son ensemble (par exemple, inclure des lignes de référence, des effets de référence, des inférences, etc.)

- Utilisez une redondance efficace. Transmettez le même message par plusieurs canaux, afin de l'amplifier et de donner au public une deuxième chance de l'obtenir. Utilisez les mots et les images à l'unisson. Par exemple, en plus de montrer les données, envisagez d'annoter la direction de l'axe "bon" ou "mauvais", et indiquez ce que vous voyez en termes simples. Ne pas confondre redondance (encombrement inutile du graphique) et redondance efficace (transmission d'un message par plusieurs canaux complémentaires).
- Laissez chaque graphique se suffire à lui-même. Utilisez des étiquettes et des légendes informatives et expliquez les abréviations. N'obligez pas le lecteur à chercher dans le texte pour comprendre une figure.
- Ajoutez toujours un titre à votre graphique. Envisagez de le formuler comme une conclusion et non comme une description (par exemple, "la concentration plasmatique dépend du poids corporel" plutôt que "graphique de la concentration plasmatique en fonction du poids corporel"). Pensez à votre public lorsque vous faites ce choix.

Sélectionner le bon graphe :

### Selecting the right base graph

Consider if a standard graph can be used by identifying suitable designs based on the:

(i) **purpose** (i.e. message to be conveyed or question to answer) and (ii) **data** (i.e. variables to display).

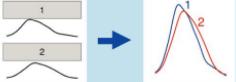
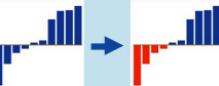
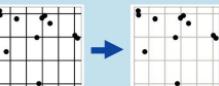
Example plots categorized by purpose

Deviation	Correlation	Ranking	Distribution	Evolution	Part-to-whole	Magnitude
Chg. from baseline 	Scatter plot 	Horizontal bar chart 	Boxplot 	Kaplan Meier 	Stacked bar chart 	Vertical bar chart 
Waterfall 	Heat map 	Dotplot 	Histogram 	Line plot 	Tree map 	Forest plot 

## Classement de l'efficacité :

Effectiveness Ranking								
Volume	Color hue	Depth: 3d position	Color intensity	Area	Slope or Angle	Length	Position on unaligned scale	Position on common scale
Least accurate					Most accurate			
volume charts	poorly designed heat maps	multivariate density plots	heat maps	bubble charts, mosaic charts	line graphs, pie charts	stacked bar charts, waterfall chart	small multiple plots	dot plots, bar charts, parallel coordinate plots

Faciliter les comparaisons, et utiliser la couleur pour accentuer ou distinguer

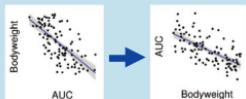
Facilitating Comparisons	Color for emphasis or distinction
<h3 data-bbox="260 221 615 256">Proximity improves association</h3>	<p data-bbox="717 221 1141 256">Restrained use of color is highly effective in organizing a narrative and calling attention to certain elements.</p>
<p data-bbox="219 277 383 338">Place labels next to data instead of using legends</p> 	<p data-bbox="717 277 1093 316">Think carefully before introducing additional color. Do you really need it?</p>
<p data-bbox="219 375 356 437">Group together elements to be compared directly</p> 	<p data-bbox="717 375 881 424">Do not use color to differentiate between categories of the same variable</p> 
<h3 data-bbox="308 472 581 501">Ease visual inspection</h3>	
<p data-bbox="219 523 397 584">Order values to help compare across many categories</p> 	<p data-bbox="717 441 902 543">Use colors or shades to represent meaningful differences such as positive/negative values, treatments or doses</p> 
<p data-bbox="219 620 397 681">Judgments are easier to make on a common vertical scale</p> 	<p data-bbox="717 564 929 646">Be consistent, use the same color to mean the same thing in a series of graphs (e.g. treatment, dose)</p> 
<h3 data-bbox="294 716 595 746">Reduce mental arithmetic</h3>	
<p data-bbox="219 768 410 829">Plot the final comparison e.g. mean difference not two means Exception: if comparator is of interest in itself</p> 	<p data-bbox="717 697 929 758">Use a bold, saturated or contrasting color to emphasize important details.</p> 
<p data-bbox="219 865 410 904">Use reference lines and other visual anchors.</p>	<p data-bbox="717 789 929 871">Emphasize the data by minimizing unnecessary ink, e.g. soften gridlines with a light color</p> 
	<p data-bbox="717 891 1113 932">Utilize existing resources for selection of appropriate palettes such as Color brewer or Munsell</p>



## Considérations sur la mise en œuvre :

### Implementation Considerations

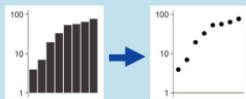
Plot cause on the x-axis and effect on the y-axis. Use this standard convention in order to avoid misinterpretation.



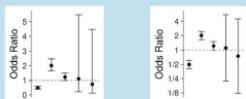
Aspect ratio can influence interpretation. Aim for a 45 degree angle of change to avoid over-interpretation of slope.



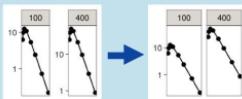
Use position for comparisons rather than length (i.e. dots instead of bars), especially for non-linear scales (e.g. log scale or % change)



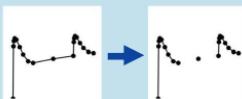
Do not plot log-normally distributed variables on a linear scale (e.g. hazard ratio, AUC, CL)



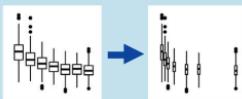
When displaying data measured on the same scale, also plot them on the same scale for easy comparison.



Connected data imply continuity. Do not connect data across a disconnected or uneven time scale.



Visits displayed close together are perceived to be closer in time. Space the visits proportional to the time between each in order to avoid confusion.  
Exception: baseline or pre-dose



Plot data and inferences to support stories about models.

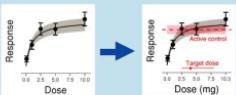


## Lisibilité et clarté :

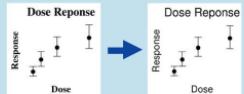
### Legibility and Clarity

Effective graphs stand alone. They use titles, annotations, labels, shapes, colors, and textures to deliver important information.

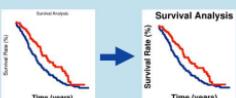
Label axes with clear measurement units and provide annotations that support the message.



Keep the font style simple, sans serif is easier to read.



Use font size to create hierarchy (e.g. set titles 2pt larger than all other labels to make them more prominent)



Display text with enough contrast to be visible. Favor the use of dark on light instead of light on dark whenever possible.



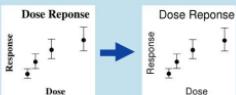
Do not type too small or too condensed. Break long titles into two lines. Shift or adjust size of labels that overlap.



Bold or italics should only be used for layering or emphasis. Emphasizing everything means nothing gets emphasized.



Keep the font style simple, sans serif is easier to read.



Try not to set text at an angle, as this decreases readability. Think of alternative solutions such as transposing the graph.



Exemple : analyse exploratoire de la relation exposition-réponse pour un médicament inhalé qui agit localement dans le poumon.  
 (a) Un nuage de points de la réponse en fonction de l'exposition est amélioré par (b) l'application de bons principes graphiques, et fondamentalement modifié par (c) la révision de la question d'intérêt. AUC<sub>0-24</sub> , aire sous la courbe de concentration en fonction du temps de 0 à 24 heures ; VEMS, volume expiratoire forcé en 1 seconde.

