

INFO COM

licence
master
doctorat

TIFFANY ANDRY
SUZANNE KIEFFER
FRANÇOIS LAMBOTTE



Les fondamentaux de la visualisation de données

> Les bases expliquées

> Les meilleures pratiques

> Des exemples concrets

+ EN LIGNE



Exercices, QCM, schémas
interactifs à compléter

deboeck **B**
SUPÉRIEUR

INFO & COM
licence
master
doctorat

**TIFFANY ANDRY
SUZANNE KIEFFER
FRANÇOIS LAMBOTTE**



Les fondamentaux de la visualisation de données

deboeck **B**
SUPÉRIEUR



constitue la bibliothèque de référence de l'étudiant des 1^{er} et 2^e cycles en information-communication.

Elle est dirigée par **Benoît Grevisse**, **Marc Lits** et **Jacques Walter**. Benoît Grevisse est professeur à l'École de journalisme de Louvain et membre actif de l'Observatoire de recherche sur les médias et le journalisme (UCL, Belgique). Marc Lits est professeur émérite de la Faculté des sciences économiques, sociales, politiques et de communication (UCL, Belgique). Jacques Walter est professeur, directeur du Centre de recherche sur les médiations (Université de Lorraine, France) et codirecteur de la revue *Questions de communication*. Ils sont entourés d'un **comité scientifique international**, garantie supplémentaire de la qualité de la collection et de ses proximités avec les programmes des différentes écoles de journalisme et de communication.

TITRES PARUS

- Andry T., Kieffer S. et Lambotte Fr., *Les fondamentaux de la visualisation de données*
- Antoine F., *Analyser la radio*
- Damian-Gaillard B., Montañola S. et Saitta E., *Genre et journalisme*
- Degand A. et Grevisse B., *Journalisme en ligne*
- Derèze G., *Méthodes empiriques de recherche en information et communication*. 2^e édition
- Derèze G., Diana J.-Fr, Standaert O., *Journalisme sportif*
- Frère M.-S., *Journalismes d'Afrique*
- Grevisse B., *Écritures journalistiques*. 2^e édition
- Grevisse B., *Déontologie du journalisme*. 2^e édition
- Guillot C. et Benmoyal Bouzaglo S., *Les fondamentaux de la communication*
- Jaspers J.-J., *Journalisme de télévision*
- Koutroubas Th. et Lits M., *Communication politique et lobbying*
- Lallemand A., *Journalisme narratif en pratique*
- Lits M., Desterbecq J., *Du récit au récit médiatique*. 2^e édition
- Marthoz J.-P., *Journalisme international*. 3^e édition
- Marthoz J.-P., *Couvrir les migrations*
- Pasquier M., *Communication des organisations publiques*. 2^e édition
- Pignard-Cheynel N. et van Dievoet L., *Journalisme mobile*
- Sepulchre S., *Décoder les séries télévisées*. 2^e édition
- Verhaegen P., *Signe et communication*

SOMMAIRE

COUVERTURE

INTRODUCTION

CHAPITRE 1. Les fondements de la visualisation d'information

CHAPITRE 2 . La perception visuelle

CHAPITRE 3. La représentation des données

CHAPITRE 4 . Principes de conception des visualisations de données

CONCLUSION

ENVIE DE TESTER VOS CONNAISSANCES ?

BIBLIOGRAPHIE GÉNÉRALE

TABLE DES FIGURES

INDEX

NOTES

RÉSUMÉ DU LIVRE

INTRODUCTION

Graphique sur l'évolution chiffrée de la pandémie de COVID-19, tableau de bord sur nos performances sportives quotidiennes, suivi d'une campagne de communication sur les réseaux sociaux, les visualisations de données sont devenues incontournables dans notre quotidien. Comment expliquer l'essor de ce mode d'information et de communication ?

Depuis 2010, la digitalisation de notre société s'est fortement accélérée. La production de contenus sur le web a connu une croissance exponentielle, rendue possible par la conjonction de plusieurs facteurs que nous prenons le temps de rappeler ici.

L'infrastructure du web (fibre optique, réseau mobile 4G/5G) s'est grandement améliorée avec une bande passante qui a permis un doublement du nombre de personnes ayant une connexion à haut débit depuis 2010 (Roser *et al.*, 2015). Ainsi, l'entreprise CISCO¹ projette que nous passerons de 3,9 milliards en 2018 à 5,9 milliards de personnes ayant un accès à Internet d'ici 2023. En parallèle, le nombre d'appareils connectés passera de 8 milliards à plus de 12 milliards. Chacun de ces appareils ayant des performances décuplées, en particulier les smartphones, nous sommes devenus, au quotidien, en capacité de produire autant de contenus qu'en une année il y a près de 100 ans. Ainsi, en 2018, on estime à 3 milliards

le nombre d'utilisateurs des réseaux sociaux². Des chiffres qui donnent le vertige.

Face à cette hyperconnectivité et cette production massive de contenus, l'un des plus grands défis est celui de l'économie de l'attention (Citton, 2013). Comment attirer le regard, l'intérêt du lecteur ou de la lectrice et l'inviter à consacrer de précieuses secondes à lire ou découvrir ce que vous lui proposez ? Dans cette bataille quotidienne, ce sont les images et les vidéos qui ont progressivement supplanté les contenus textuels sur le web. En 2021, ce sont les plateformes de contenus vidéos et images qui dominent le classement en matière de contenus les plus consommés sur plateforme³.

Avec l'accélération des innovations numériques, nous déléguons aux machines, outils de rétention tertiaire selon Stiegler (Rétention | Ars Industrialis, s. d.), le stockage et la rétention de tous nos échanges et productions numériques. On stocke, on trie, mais on ne jette rien (Van House & Churchill, 2008). Ce stockage massif de nos données a permis le développement de l'économie de la donnée. Du *big data* aux *data analytics* en passant par l'intelligence artificielle, l'exploitation des données est devenue le moteur des prises de décisions dans toutes les sphères de l'organisation : optimisation des performances de production, adaptation des capacités ou encore personnalisation de l'offre de contenus et de services passent aujourd'hui par l'analyse en temps réel de nos traces numériques.

Face à cette masse de données, il s'agit pour les producteurs de contenu qu'il ou elle soit journaliste, data analyste ou animatrice de communauté de produire de l'intelligibilité sur la base de données collectées et traitées grâce à un algorithme. En effet, l'exploitation des données, vu la quantité amassée, ne fait également plus depuis longtemps l'objet d'un monopole des humains (Lamb & Kling, 2003). Par la nature des traces et par l'évolution des modalités de capture,

de nouvelles formes d'exploitation sont à l'œuvre : les algorithmes, que Cardon (2015) décrit comme des objets dénués d'intelligence. Malgré cela, leur utilisation n'est pas pour autant sans conséquence. En effet, une fois collectées et stockées, les traces de nos échanges font l'objet de calculs, de mise en association par ces algorithmes qui se chargent ensuite de traduire ces résultats sous forme de visualisations de données. C'est ce que nous définissons comme une mise en intelligibilité des données.

Pour Bachimont (1996), la production de nouvelles connaissances par les algorithmes repose sur une technique d'inscription computationnelle qui dépasse et diffère de l'inscription graphique : « L'hypothèse que nous formulons est que l'informatique, sous la forme des systèmes formels automatiques, fournit précisément un nouveau type de support, les supports dynamiques, auquel doit correspondre un type spécifique de synthèse, et par conséquent une rationalité spécifique, que nous proposons de baptiser "raison computationnelle" » (Bachimont, 1996 pp.14-15). Cependant, même si cette raison computationnelle produit de nouvelles connaissances, celles-ci restent très abstraites (des 1 et des 0) si elles ne sont pas traduites d'un écrit-calcul dans un écrit-graphique accessible aux utilisateurs finaux (Collomb, 2016). Dans cet exercice, la visualisation d'information et de données a pris une place prépondérante de par leur efficacité communicationnelle.

Bien que l'influence des algorithmes sur notre quotidien domine les discours médiatiques, nous souhaitons, à travers cet ouvrage, conscientiser le lecteur et la lectrice quant à l'importance des visualisations de données comme le médium privilégié dans nos interactions avec ces calculs. Que ce soit un graphique sur l'évolution de la pandémie de COVID-19, un tableau de bord d'aide à la prise de décision ou encore dans le suivi de nos performances physiques

quotidiennes sur notre smartphone, ces mises en forme influencent grandement nos manières d'interpréter ces informations et d'agir en conséquence.

Cet ouvrage est un des fruits du projet FEDER USERMEDIA/MEDIAFACTORY UCL Mons, portant sur la mise en intelligibilité des données produites sur les réseaux sociaux numériques. Dès le départ, en 2013, nous avons pris conscience de la nécessité de travailler sur la conception des visualisations de données et leur compréhension par un public de non-initiés.

Ce public est au cœur des réflexions que nous proposons dans cet ouvrage, il vise d'une part à conscientiser les concepteurs et conceptrices de visualisation de données aux origines et à l'évolution des grands principes de conception, aux familles de visualisations de données les plus appropriées en fonction de la construction de l'histoire et du message souhaité et, d'autre part, de permettre à ce public de non-initiés de poser un regard critique sur la visualisation d'information et de données à laquelle il ou elle est exposé quotidiennement en tant que citoyen.ne dans toutes les facettes de notre vie.

Enfin, alors que la visualisation de données prend une place prépondérante dans notre quotidien, l'inscription de cette matière aux programmes de formation en sciences de l'information et de la communication et plus largement en sciences humaines et sociales n'est que très récente. Nous espérons équiper celles et ceux qui, dans le monde francophone, souhaitent bénéficier d'une base théorique solide pour construire leurs compétences dans ce domaine. De nombreuses illustrations accompagnent également les lecteurs et lectrices dans leur découverte des principes et familles de visualisations de données. Cet ouvrage ne prétend pas se substituer à une formation plus technique visant la maîtrise d'un logiciel de

conception de visualisation de données. Cependant, il représente un excellent guide pour celui ou celle qui, équipé d'un logiciel, devra poser des choix à chaque étape du processus de conception (types de données, types de croisement, type de message, type de public).

Pour vous permettre de saisir toutes les subtilités de la visualisation de données, cet ouvrage est organisé comme suit :

Le premier chapitre présente une perspective historique de la visualisation de données. Son objectif est de vous montrer que l'utilisation des graphiques de données est loin d'être un phénomène nouveau. Il initie également aux objectifs essentiels de la visualisation de données et à quelques éléments importants de terminologie.

Le second chapitre aborde la perception visuelle. Qu'est-ce que le sens de la vue ? Comment fonctionne l'œil ? D'où viennent les couleurs ? Ainsi, le lecteur est amené à comprendre que le pouvoir de la visualisation de données repose sur des phénomènes physiques concrets : par exemple, en colorant de rouge un numéro dans une série de chiffres, il est plus facile de l'identifier. C'est le traitement préattentif qui est stimulé, essentiel en visualisation de données. Après l'avoir découvert, vous serez sensibilisé à diverses lois d'illusion d'optique qui entrent en jeu dans la conception de représentation graphique. Comprendre la vision, c'est comprendre comment fonctionne la visualisation de données.

Le troisième chapitre entre dans le vif du sujet en présentant les données et leur représentation au lecteur. Effectivement, il existe différents types de données et de variables qui, une fois identifiés, permettent de choisir la représentation idéale. Néanmoins, les visualisations de données se classent en différentes familles les visualisations statistiques, les visualisations temporelles, les cartes et, pour terminer, les hiérarchies et réseaux. Après une présentation de

celles-ci, le lecteur a accès à un catalogue qui ne recense pas moins de 43 visualisations différentes dans une fiche détaillée : à quelles données et à quels types de variables correspondent ces visualisations, quelles sont leurs forces et leurs faiblesses, etc. Son objectif est d'aider le concepteur de visualisation de données débutant à choisir la bonne technique ou, autrement dit, le bon graphique.

Le quatrième chapitre rassemble toutes les règles et principes de conception des visualisations de données qui aident à réaliser une visualisation correctement perceptible et adaptée à son audience. Découvrez ainsi les *do's and don'ts* de la visualisation à travers ce dernier chapitre : de véritables trucs et astuces à mettre en place pour éviter tout faux pas dans la conception de sa représentation graphique.

Cet ouvrage se termine tout naturellement sur une conclusion qui vise à vous rappeler les éléments essentiels parcourus au sein des chapitres.

Comment lire cet ouvrage ?

Pour les uns, ce livre permettra une initiation dans le domaine de la visualisation de données, en proposant un cadrage en termes de culture générale. Les deux premiers chapitres rempliront parfaitement cet objectif, tandis que les deux derniers apportent des connaissances plus concrètes.

Pour les autres, il peut s'agir d'un guide, à partir du chapitre 3. En effet, lorsque l'on conçoit une visualisation de données, il y a des éléments importants à prendre en compte. Le type de données et le type de variables, ainsi que le message à communiquer, permettent de faire un choix entre plusieurs techniques de visualisation. Le lecteur

pourra s'il le souhaite s'aider du catalogue pour faire un choix entre 43 techniques de visualisation. Par la suite, il pourra s'aider du quatrième chapitre pour réaliser ses choix de conception graphique : quels principes respecter, quelles couleurs utiliser, etc.

En somme, cet ouvrage peut être lu traditionnellement, du premier au dernier chapitre, pour les débutants et les curieux. Pour ceux qui souhaitent mettre la main à la pâte et réaliser leurs propres visualisations de données, il peut être utilisé de manière sélective, tel un guide, en choisissant sa technique de visualisation dans le catalogue ou encore en s'assurant grâce aux principes de conception recensés que leur design est correctement conçu.

Même si beaucoup de règles régissent la création d'une visualisation de données, n'oubliez jamais qu'elle est une conséquence de choix de conception : c'est une interprétation, un message créé par un auteur sur base d'un jeu de données concret. À travers cet ouvrage, nous espérons développer, d'une part, une perspective critique chez le lecteur de visualisations de données et, d'autre part, d'aider les créateurs de visualisations de données à adopter de bons réflexes de conception.

Bibliographie indicative

Bachimont, B. (1996). Intelligence artificielle et écriture dynamique : De la raison graphique à la raison computationnelle. *Au nom du sens*, 290-319.

Collomb, C. (2016). *Un concept technologique de trace numérique*. Compiègne. <http://www.theses.fr/2016COMP2286>

Cardon, Dominique. (2015). *À quoi rêvent les algorithmes : Nos vies à l'heure des big data*. Seuil.

Citton, Y. (2013). The Attention Economy and the New Forms of Digital Exploitation. *Multitudes*, 54(3), 163-175.

Lamb, R., & Kling, R. (2003). Reconceptualizing Users as Social Actors in Information Systems Research. *Management Information Systems Quarterly*, 27(2). <http://aisel.aisnet.org/misq/vol27/iss2/4>

Rétention | Ars Industrialis. (s. d.). Consulté 4 janvier 2018, à l'adresse <http://arsindustrialis.org/r%C3%A9tention>

Roser, M., Ritchie, H., & Ortiz-Ospina, E. (2015). Internet. *Our World in Data*. <https://ourworldindata.org/internet>

Van House, N., & Churchill, E. F. (2008). Technologies of memory : Key issues and critical perspectives. *Memory Studies*, 1(3), 295-310.

CHAPITRE 1

LES FONDEMENTS DE LA VISUALISATION D'INFORMATION

1. Histoire de la visualisation d'information
2. L'utilité de la visualisation d'information : objectifs et définition

La visualisation de données est une discipline en plein essor aujourd'hui, grâce au développement du numérique ces dernières décennies. Désormais, visualiser les grands jeux de données issus de nos systèmes d'information et de communication est un enjeu sociétal important. Cette pratique n'est pourtant pas si neuve. Visualiser les données signifie les transcrire sous une forme visuelle à des fins de compréhension et de communication. Ne dit-on pas qu'une image vaut mieux qu'un long discours ? La vision et la compréhension semblent ainsi liées depuis longtemps... Dans ce chapitre, vous découvrirez une perspective historique de la visualisation de données afin de comprendre, dans un second temps, les raisons pour lesquelles on visualise les données. Le chapitre se terminera par la présentation de visualisations de données judicieuses, réalisées au XIX^e siècle.

1. Histoire de la visualisation d'information

Nombreux ont été les hommes avant nous qui ont souhaité mieux comprendre des informations qu'ils avaient à disposition. Cela remonte à la préhistoire. Pour « mieux penser », nos prédécesseurs ont eu besoin de réaliser des représentations physiques de l'information. Cela leur permettait ainsi de « sortir » l'information de leur tête pour mieux la « regarder » d'un point de vue extérieur (Kirsh, 2010). En vérité, il n'y a pas « une histoire » de la visualisation

de données. Elle trouve sa source dans plusieurs disciplines telles que les mathématiques, la géographie ou encore la sémiologie. Mais avant d'être un objet scientifique à proprement parler, la visualisation des données a émergé des besoins et des pratiques sociales des premiers êtres civilisés.

1.1 Compter à la préhistoire

Pour commencer, on a retrouvé un bâton de comptage au Congo, daté de plus de 20 000 ans. Cet « Os d'Ishango », sur lequel se présentent des entailles espacées de façon régulière, serait le premier système de comptage. Ainsi l'être humain eut besoin de compter, mais également de se remémorer l'information de manière externe. Ces bâtons étaient utilisés avant l'existence de l'écriture (Huylebrouck, 2008).

Plus tard, vers 3500 av. J.-C., les Sumériens utilisaient des jetons de comptage pour un tas de choses : compter le nombre de brebis dans un troupeau, relever les impôts, etc. À la différence, un jeton de comptage pouvait dès lors représenter plusieurs unités, identifiées grâce à un symbole. Ainsi naquit le nombre et, plus tard, l'écriture cunéiforme (Launay, 2016). Avec l'apparition du nombre, nous comprenons ainsi que l'homme a besoin d'un support externe à sa pensée pour pouvoir réfléchir. Le bâton et les jetons de comptage sont bien une façon de *visualiser* de l'information.



Figure 1 : LOs d'Ishango, exposé au Museum des Sciences Naturelles à Bruxelles (CC0)

1.2 L'influence de l'arbre

La religion, catholique notamment, a également donné un coup de pouce à la visualisation de données. Ainsi, en 1086, la première visualisation sous forme d'arbre vit le jour (figure 2). Il s'agit de « L'arbre de Jessé ». Celui-ci vise simplement à représenter de façon visuelle l'arbre généalogique de Jésus, en partant de Jessé, père du roi David. La racine se lit ainsi au bas, et Jésus se trouve en haut de l'arbre (Focant, 1989). Dès le Moyen Âge, il s'inscrit définitivement dans l'histoire de la visualisation de données grâce à la métaphore hiérarchique qu'il représente. Il permet ainsi de montrer des relations d'ordre dans la religion, dans la connaissance, en sciences, en généalogie et autres, en utilisant bien souvent des formes figuratives, comme dans la figure 3 (Lima, 2014).



Figure 2 : Arbre de Jessé dans Le livre de chasse de Phébus, France, fin du xv^e siècle (Domaine public)

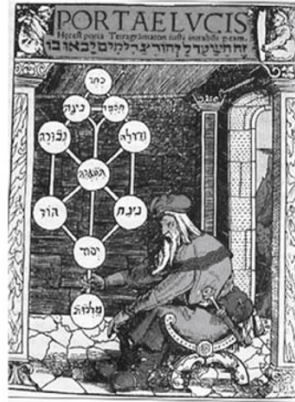


Figure 3 : Exemple médiéval d'un diagramme d'arbre.
 Illustration reproduite dans Lima (2014), Visual complexity,
 Princeton Architectural Press, p. 24 (Domaine public)

Raymond Lulle, durant la Renaissance, produit d'ailleurs plusieurs arbres de ce type, dont le fameux *arbor scientiae* (arbre de la science), dans lequel bourgeonnent différents autres arbres (*arbor elementalis*, *vegetalis*, *sensualis*, *imaginalis*, *moralis*, etc.) (Eco, 2010). La figure 4 illustre l'*arbor elementalis*, montrant ainsi les objets du monde composés des quatre éléments.

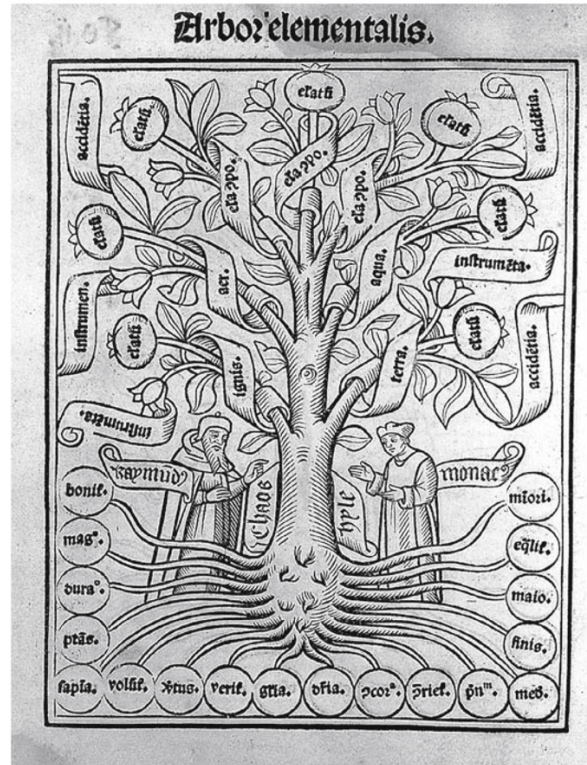


Figure 4 : Arbor Elementalium de Raymond Lulle (CC BY)

1.3 Le développement de la cartographie

Un autre domaine ayant aussi favorisé la visualisation de données est la géographie, et plus particulièrement la cartographie. Ainsi, les êtres humains eurent besoin de plans et de cartes pour se repérer correctement. Les premières cartes retrouvées ne montrent pas la Terre, mais le ciel. Les premières représentations célestes montrent la position des étoiles et sont datées de – 16 500 dans les grottes de Lascau. D'autres représentations spatiales ont existé au cours de l'histoire, mais les premières cartes du monde ont vu le jour dans l'Antiquité grecque. Cependant, Anaximandre et Hécathée de Milet (vers – 500 av. J.-C.) ou encore Ptolémée (vers – 150 av. J.-C. ; figure 5) ne sont pas les seuls à avoir essayé de cartographier le monde tel qu'ils le connaissaient (Jacob, 1992). Les Babyloniens et Chinois s'y sont également appliqués à l'époque. La cartographie

connaît ensuite une longue et florissante histoire : au Moyen Âge, Jérusalem se trouve au centre des cartes qui ne mentionnent que l'Europe, l'Afrique et l'Asie. Plus tard, la découverte de l'Amérique par Christophe Colomb modifie l'apparence des cartes. De nombreuses écoles de cartographie se développent au cours de l'histoire, et la carte se décline à plusieurs applications, comme à celle de la guerre. Un exemple en est la carte d'*État-Major*, demandée par Napoléon lui-même (figure 6). De nombreux cartographes sont devenus célèbres pour la finesse de leur art, comme Mercator (1512-1594) (Jacob, 1992). Par ailleurs, Jacques Bertin (1918 – 2010), père de la *Sémiologie Graphique* (1967) qui est fondamentale en visualisation d'information, était lui aussi cartographe (*infra*).



Figure 5 : Reconstitution de la carte du Monde de Ptolémée (xv^e siècle) (Domaine public) (CC0)

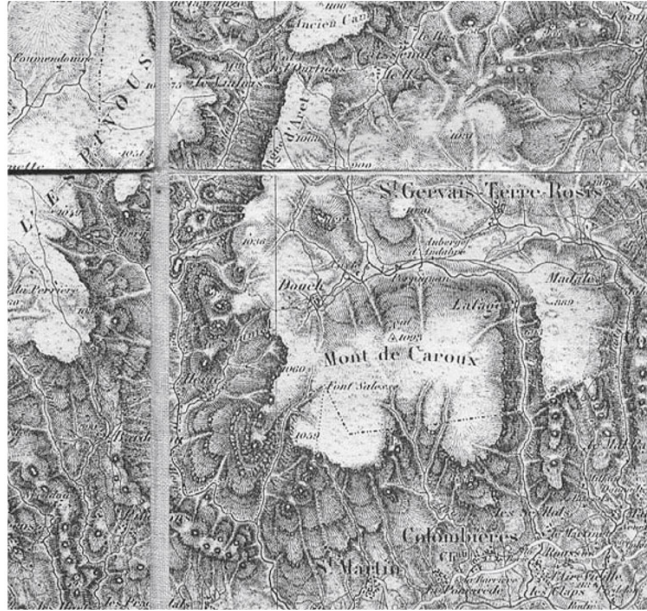


Figure 6 : Carte d'état-major, 1866 (Domaine public)

1.4 La représentation visuelle des statistiques

Il va sans dire que la visualisation d'information n'aurait pu se développer davantage sans l'émergence des statistiques. Évidemment, nous pourrions même remonter à l'émergence et à l'histoire des mathématiques, mais ce n'est pas l'objet de notre propos. En revanche, nous ne pouvons omettre de mentionner les statistiques, dont la visualisation est devenue fondamentale aujourd'hui. En démographie, en médecine, en collecte d'information sur les états, en physique... Les statistiques nécessitent depuis longtemps qu'on puisse voir les données afin de mieux les comprendre.

En 1370, Nicolas Oresme est le premier à représenter sur un graphique les différents mouvements physiques, grâce aux axes « *longitudo* » et « *latitudo* », qui correspondent aux abscisses et ordonnées que nous connaissons aujourd'hui. Il a ainsi utilisé des formes géométriques en relation les unes avec les autres pour illustrer des différences de grandeur. Il est aussi le premier à représenter une

série de nombres dans ce qui semble être un histogramme (Parocchia, 2017).

D'autres auteurs se sont essayés à la représentation graphique, mais c'est William Playfair qui, en 1786, met au point trois représentations graphiques qui ressemblent énormément à ce que nous utilisons encore aujourd'hui. Il crée ainsi le diagramme en ligne (ou courbes), le diagramme en barres et le diagramme circulaire (Costigan-Eaves, Macdonald-Ross, 1990). Bien que ses graphiques n'aient pas connu un succès fou de son vivant, ils sont devenus des outils communément utilisés en visualisation de données. Ce géologue – mathématicien – économiste – touche à tout publie ses premiers graphiques dans « *L'Atlas Commercial et Politique* » afin d'aider les hommes politiques à faire des choix rapidement, en saisissant l'information en un coup d'œil ou presque (Costigan-Eaves, Macdonald-Ross, 1990). Les thématiques abordées sont donc politiques, économiques et sociétales. Les figures 7, 8 et 9 représentent différents graphiques de l'économiste. C'est le début de l'utilisation de visualisation de données abstraites dans l'aide à la décision.

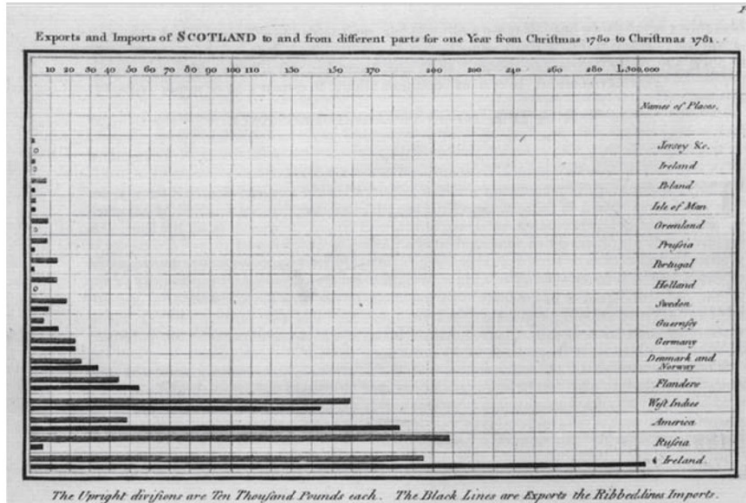


Figure 7 : Graphique en barres de William Playfair dans The Commercial and Political Atlas, 1786 (CC0)

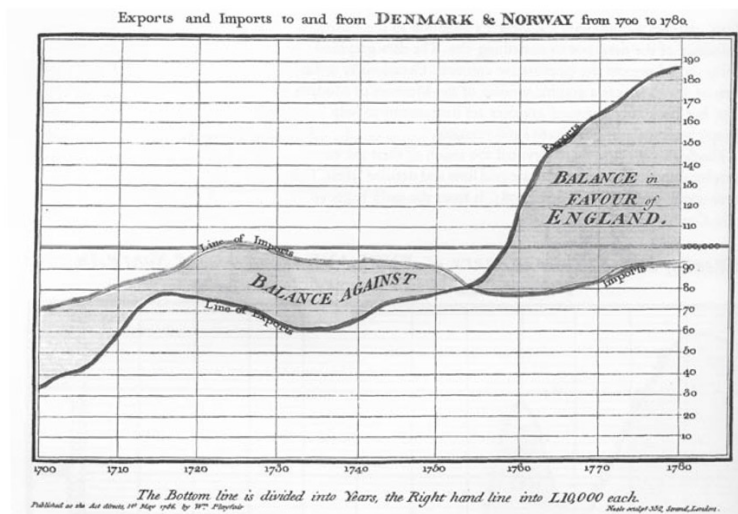


Figure 8 : Graphique en lignes de William Playfair dans The Commercial and Political Atlas, 1786 (CC0)

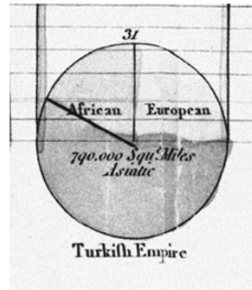


Figure 9 : Graphique en secteurs de William Playfair dans « Playfair's commercial and political atlas and statistical breviary », réédition de 2005 (CC0)

1.5 Les exemples marquants du XIX^e siècle

Après Playfair, le XIX^e siècle devient particulièrement florissant pour la visualisation d'information. Diverses représentations graphiques ont marqué l'histoire. Voici trois personnalités qui sont citées très souvent dans le domaine comme références : John Snow (1854), Florence Nightingale (1858) et Joseph Minard (1869).

En 1854, le docteur John Snow cherche à comprendre d'où vient une épidémie de choléra à Londres. Sur un plan du quartier de Soho, il recense alors le lieu d'habitation ou de travail des personnes infectées. Grâce à cela, il identifie que le nombre d'infectés augmente au fur et à mesure que l'on se rapproche d'une pompe de distribution d'eau publique à Broad Street et en conclut que l'eau de cette pompe est probablement empoisonnée (figure 10 – Snow, 1854).

En 1858, l'infirmière et statisticienne Florence Nightingale produit à titre privé des notes concernant l'impact de la guerre de Crimée sur l'armée britannique, à la demande du Secrétaire d'État à la Guerre. Cet ouvrage contient un graphique statistique (figure 11) qui montre que la maladie épidémique est responsable de plus de décès britanniques au cours de la guerre de Crimée que les blessures sur le champ de bataille. Sur son diagramme, il y a 12 secteurs,

correspondant chacun à un mois de l'année. La taille de leur aire est proportionnelle aux nombres de décès et les différentes couleurs employées dans ces secteurs en représentent la cause. La couleur grise représente la cause infectieuse, la couleur rose représente les décès dus aux blessures et la couleur noire représente les autres causes (Nara, 2015, p. 229). Le graphique que Nightingale a utilisé pour expliquer des statistiques complexes de manière simple, claire et convaincante, est devenu le Nightingale Rose Diagram, fortement utilisé aujourd'hui en météorologie.

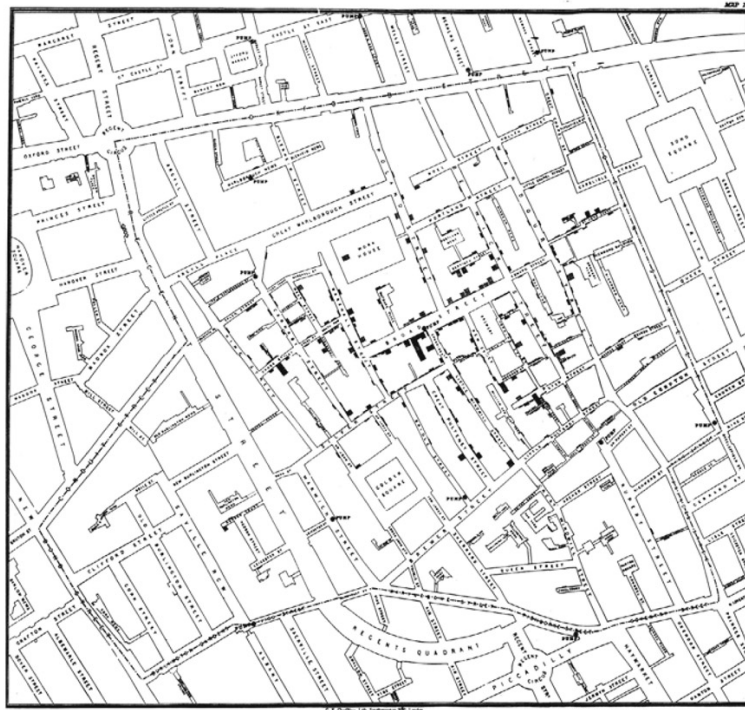


Figure 10 : Carte originale de John Snow, 1854.
Les cas de choléra sont représentés grâce aux barres noires
(Domaine public)

En 1869, Charles Joseph Minard, illustre pionnier en matière de design graphique, réalise sa représentation graphique la plus célèbre : la « Carte figurative des pertes successives en hommes de l'Armée française dans la campagne de Russie en 1812-1813 » (figure 12). De

nombreux professeurs en visualisation d'information la considèrent encore comme un exemple incroyable mêlant localisation, quantité, concision et complexité à la fois. La raison en est que sur une simple image, il réussit à illustrer plusieurs variables relatives à la désastreuse campagne de Napoléon. Ces variables sont

- la localisation et l'itinéraire de l'armée indiquant les points de séparation et de regroupement des unités et en indiquant l'aller et le retour des troupes ;
- le nombre de pertes de l'armée ;
- les variations de la température lors de la mission.

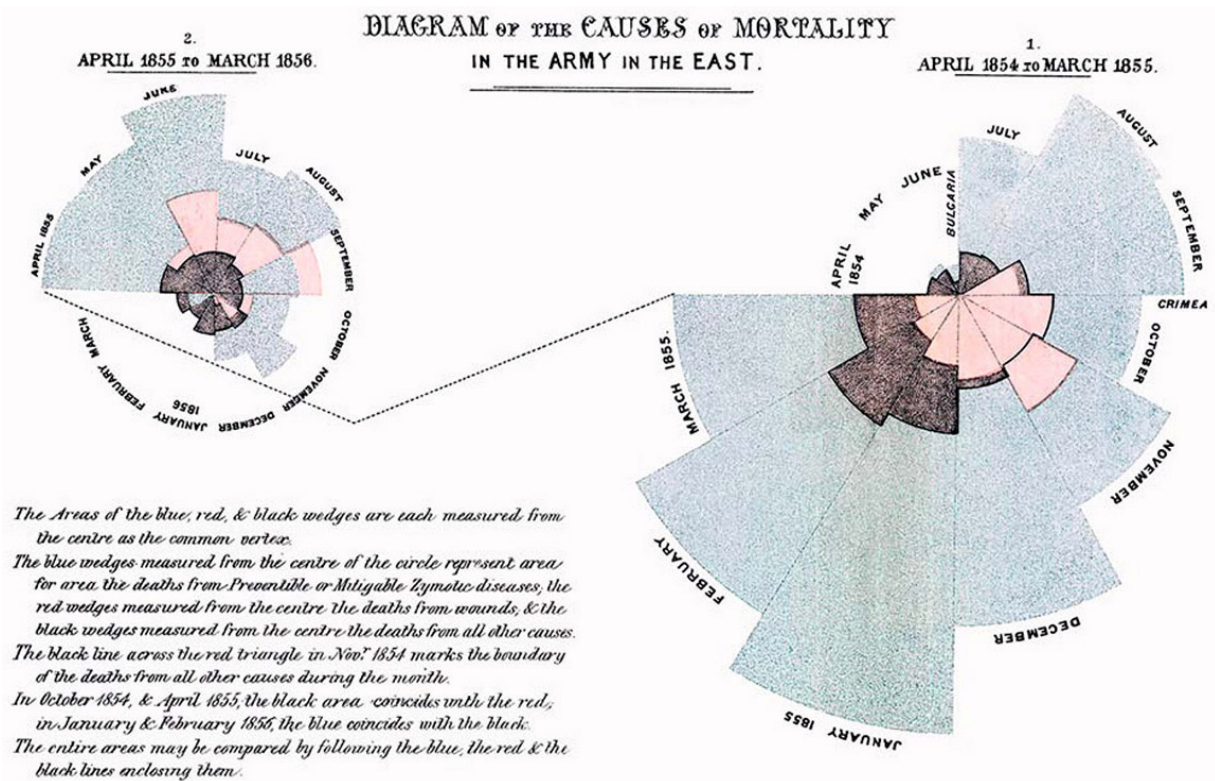


Figure 11 : Diagram of the causes of mortality in the army in the East par Florence Nightingale, 1858 (Domaine public)

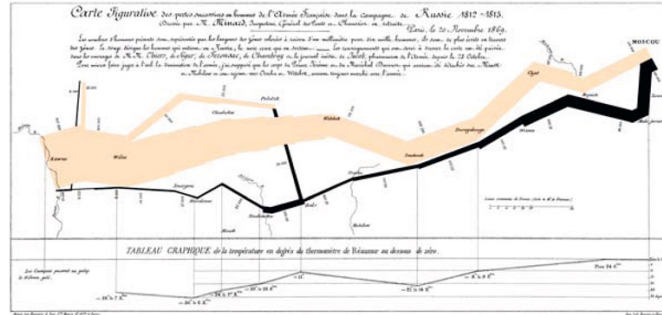


Figure 12 : Charles Joseph Minard, Carte figurative des pertes successives en hommes de l'Armée française dans la campagne de Russie en 1812-1813, 1869 (Domaine public)

1.6 Infovis et Scivis

La visualisation de données s'est également améliorée au fil du temps grâce aux innovations dans le domaine informatique. Il y a ainsi deux types de visualisations de données. Elles sont très différentes et il ne faut pas les confondre.

Premièrement, les **scivis**, aussi appelées visualisations de données scientifiques, s'appliquent, comme leur nom l'indique, aux données scientifiques, souvent dans le domaine de la physique (Card, 2012). Les **scivis** concernent les données continues (cf. [Chapitre 3](#)). Les données visualisées sont des phénomènes physiques, tangibles, invisibles, que la visualisation rend donc apparents (Telea, 2014). En **scivis**, la simplification des données est possible, malgré une perte d'information minimale qui laisse le message compréhensible. La simplification est assez naturelle. Par ailleurs, le codage des données sur la surface disponible est assez contraint. De fait, il est implicite, car il se réalise dans un espace géométrique qui fait sens ; il n'y a qu'une seule possibilité, qu'une seule manière de placer les points dans l'espace (Telea, 2014). De fait, les **scivis** ont trois attributs majeurs. Premièrement, les points de données ont toujours un emplacement clair, précis et localisé dans la réalité. Ensuite, elles

concernent des valeurs numériques continues (des quantités décimales). Pour terminer, le phénomène mesuré est tout naturellement continu : entre deux points de mesure sur l'image, les valeurs varient continuellement, ce qui est impossible avec des valeurs discrètes (telles qu'un nom, par exemple) (Telea, 2014). La figure 13 illustre un exemple de *scivis* : il s'agit de la représentation des fibres anatomiques qui constituent l'architecture de la matière blanche du cerveau humain. Elles sont visualisées selon un code de couleur en fonction de la direction du flux nerveux dans le cerveau (Horn *et al.*, 2014). L'information est visualisée sur un code couleur, et il a été possible de simplifier la représentation en ne visualisant pas les tissus humains et autres composantes du cerveau. De même, on comprend également que les données représentées sur cette *scivis* ne peuvent être visualisées ailleurs dans l'espace, parce qu'elles y représentent un phénomène physique qui n'existe tout simplement pas à un autre endroit.

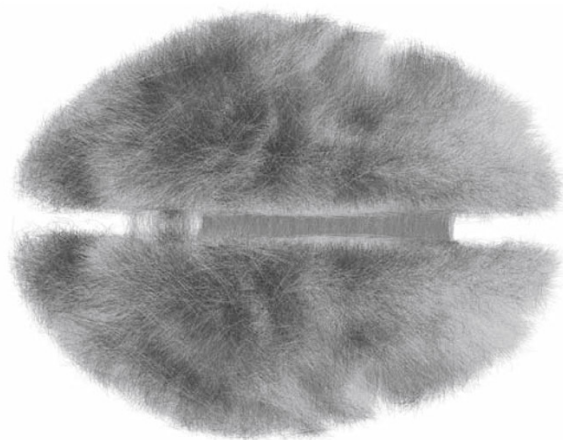


Figure 13 : Scivis : représentation des connectomes humains, par Horn *et al.*, 2014 (CC0)

Ensuite, l'*infovis*, aussi appelée visualisation d'information, mêle entités abstraites et représentation spatiale afin d'augmenter la cognition (Card, 2012). « Abstrait » signifie qu'il n'y a pas

d'emplacement physique de ces données dans la « réalité ». Une visualisation doit avoir un espace pour exister et cet espace est abstrait. Les *infovis* concernent donc des données de type abstrait et dès lors, il est très difficile, voire impossible, de supprimer les données de ces visualisations : cela reviendrait à perdre une part d'information trop importante. De fait, l'encodage des données sur la représentation graphique est explicite et même arbitraire. L'auteur de la représentation est, au contraire des *scivis*, totalement libre de trouver la manière de disposer les données dans l'image afin que son objectif soit servi au mieux (Telea, 2014). La figure 14 est un exemple d'*infovis* : tous les éléments présents sur ce graphique sont du fait de son créateur. Les barres représentées n'existent pas en tant que tel dans la réalité, même si, de façon abstraite, elles illustrent un phénomène. Le texte, les icônes et les barres pouvaient être disposés différemment tout en représentant un phénomène identique.

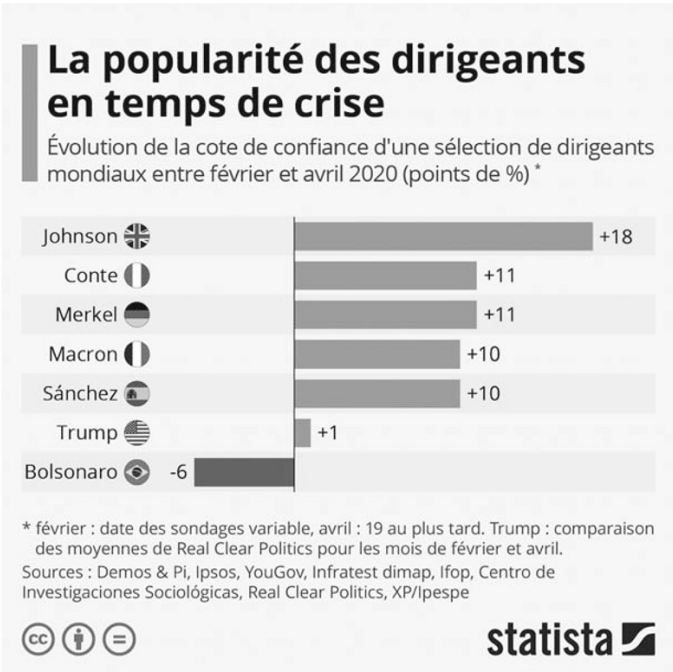


Figure 14 : Infovis : la popularité des dirigeants en temps de crise, par Statista, 2020 (CC0)

Les *scivis* et les *infovis* sont des catégories de visualisation très différentes, connaissant elles-mêmes des sous-catégories. Par exemple, les *scivis* se subdivisent en *flow visualisation*, visualisation géospatiale, visualisation médicale et autre, tandis que l'*infovis* se subdivise en visualisation en réseau (*network vis.*), *software visualisation* (relative aux programmes informatiques), visualisations relatives à la sécurité de l'état, etc. Comme on le voit, il est possible de catégoriser les visualisations de données en fonction de leur domaine d'application. Cependant, tout n'est pas forcément immuable. Ainsi, ces dernières années, certains chercheurs ont montré la plus-value d'un croisement entre *scivis* et *infovis*, pour apporter de nouvelles clés de compréhension au monde de la visualisation de données.

Dans cet ouvrage, nous aborderons les visualisations de données principalement sous l'angle de la visualisation d'information, dite *infovis*.

1.7 Conclusion

En conclusion, la visualisation d'information ne date pas d'hier et l'humain, depuis la Préhistoire, a eu besoin de visualiser pour améliorer sa réflexion. Il s'agit d'un besoin pour les êtres humains qui l'utilisent comme support à leur propre mémoire d'une part, et pour prendre du recul d'autre part : en « regardant les données », ils prennent des décisions (Cardon, 2012).

Alors que l'on retrouve des traces des premières visualisations de données il y a plus de 20 000 ans, c'est à la fin du XVIII^e siècle que la visualisation de données a connu un essor important, en s'inscrivant définitivement comme outil, grâce aux innovations graphiques de William Playfair. Quant à la cartographie, c'est une autre histoire : son évolution a été constante et a suivi les améliorations

technologiques et la découverte du monde de manière générale. Mais qu'en est-il aujourd'hui ?

À partir des années 2000, l'utilisation de l'ordinateur et d'Internet a définitivement changé les pratiques de représentation graphique qui, grâce aux outils accessibles, sont de plus en plus inventives, permettant notamment l'interaction avec la visualisation de données.

Depuis 2010, on constate un engouement pour la visualisation de données, non seulement parce que les évolutions technologiques la rendent accessible à tous ou presque, mais aussi parce que les capacités de capture et de stockage d'informations enregistrées dans des bases de données numériques ont connu une croissance exponentielle (Brasseur, 2015). Cette grande masse, appelée *big data*, nécessite une visualisation afin d'être exploitable (Cardon, 2012). Ce n'est donc pas la visualisation de données qui est nouvelle, mais plutôt le regard que nous portons sur les données... L'avènement du numérique contribue à écrire une nouvelle page de la visualisation d'information : avec de nouveaux outils naissent de nouvelles pratiques de visualisation.

Si vous souhaitez en apprendre plus sur l'avènement des différentes représentations graphiques au cours de l'histoire, vous pouvez explorer les ressources externes (qui sont en anglais). Il s'agit de deux lignes du temps montrant l'apparition de nouveaux graphiques.

2. L'utilité de la visualisation d'information : objectifs et définition

2.1 Une aide à la pensée

En tant qu'êtres humains, nos réflexions peuvent être nombreuses. Mais combien de fois, lorsque nous avons besoin d'effectuer un calcul

ou d'organiser un événement, nous saisissons-nous d'un crayon pour établir la clarté ? Cet acte nous permet ainsi « d'extérioriser » nos pensées pour ensuite les regarder, faire le point et tirer une conclusion. La visualisation de données suit le même principe. Effectivement, les visualisations d'information sont des représentations externes qui permettent de créer du sens supplémentaire, de réaliser des actions externes à notre système de pensée « interne ». En somme, on utilise un support à notre pensée pour l'augmenter. Il s'agit d'un véritable approfondissement de notre réflexion (Kirsh, 2010). Les visualisations d'information nous permettent de réaliser une réelle économie cognitive : elles diminuent notre charge cognitive en facilitant le travail effectué pour comprendre l'information et la traiter (Zhu, Chen, 2008).

Souvent, lorsque les analystes réalisent une visualisation de données, c'est parce qu'elle leur permet d'amplifier leur réflexion. Un tableau affichant de nombreuses données peut être compliqué à saisir ; comment en retirer les informations principales ? Comment détecter les tendances contenues dans ces données ? Comment les interpréter, comment prendre des décisions ? En offrant une vision claire du problème et des données, la visualisation est un moyen de répondre à toutes ces questions. Ensuite, lorsque l'on voudra communiquer ces données dans un article de presse ou un rapport d'entreprise, la visualisation de données sera un support permettant d'expliquer l'information, de la mettre en scène, sans que le destinataire de la visualisation ne doive rechercher les données dans un tableau ou dans un paragraphe de texte. Ainsi, une visualisation de données correctement réalisée permet d'étendre la cognition, mais aussi la mémorisation. Correctement réalisée, elle nous permet d'ancrer un souvenir dans la mémoire. Il s'agit d'une véritable « cristallisation de

la connaissance » ; c'est ce qu'on appelle le « *sensemaking visuel* » (Chi, Card, 1999).

L'aide à la prise de décision est un champ d'application central de la visualisation des données. Comment mettre en exergue les informations cruciales au moment du processus de prise de décision dans un environnement complexe (Ramly *et al.*, 2012 ; Payne, 1976) ? La visualisation de données, qui clarifie l'information, en l'organisant précisément et en révélant les liens entre les données, est un outil précieux dans ce genre de situations. En organisation, les outils d'aide à la décision sont souvent présentés sous la forme de *dashboard* (tableau de bord informatique) regroupant une série de visualisation de données agrégant les informations utiles à l'utilisateur. Ces tableaux de bord mettent en lumière les informations et les enjeux qui permettent aux décisionnaires d'une entreprise d'échafauder les scénarios dans l'exercice de prise de décision (Few, 2006).

2.2 Explorer et communiquer : les grandes finalités de la *dataviz*

La visualisation de données permet de poursuivre deux grands objectifs : explorer l'information et communiquer l'information. Les logiques de création de ces supports visuels sont alors tout à fait différentes.

2.2.1 Explorer

Dans ce cas, l'exploitation du graphique est plutôt individuelle. La visualisation de données aide l'utilisateur à trouver des informations et à générer des idées. C'est lui le créateur de la visualisation, c'est lui qui choisit tous les éléments qui la composent. Son objectif est d'explorer les données et de les comprendre. Les visualisations de

données produites dans un but exploratoire ont souvent une courte vie, car elles évoluent beaucoup au cours de leur utilisation. Le travail personnel de l'utilisateur sur cette visualisation est donc très important : certains choix de construction et d'utilisation de la représentation sont propres à la personne qui l'utilise (Friendly, 2008). Les personnes qui explorent les données sont souvent des analystes qui produisent eux-mêmes leurs visualisations.

2.2.2 Communiquer

La seconde raison d'utiliser une représentation graphique est d'avoir pour ambition de présenter des données à des fins de communication. Dans ce cas, il faut décider quelle information on souhaite mettre en lumière et ce, dans un dispositif approprié à l'audience. La façon dont la représentation sera perçue par l'utilisateur doit être dans les préoccupations principales du créateur de la visualisation (Friendly, 2008). Dès lors, il doit établir à quels besoins la visualisation de données doit répondre. Les questions auxquelles répondent les données exploitées dans une visualisation doivent être définies bien avant que la conception de la visualisation des données ne commence. Un même jeu de données peut en effet répondre à plusieurs questions à la fois. Le sujet de la visualisation de données ainsi que son audience sont des éléments déterminants dans sa présentation (Kirk, 2016). Les visualisations de données produites à des fins de communication sont très répandues aujourd'hui, dans la presse, sur des blogs ou encore dans des rapports d'activité en entreprise.

2.2.3 Communiquer et explorer : la mise à disposition d'Analytics pour les utilisateurs

Il est aussi commun que la visualisation de données combine ces deux objectifs à la fois. Ainsi, imaginons que vous êtes l'administrateur de

la page Facebook d'une association ou bien que vous teniez une boutique en ligne. Les réseaux sociaux et hébergeurs de sites web proposent aujourd'hui différents tableaux de bord (*dashboards*) présentant certaines statistiques afin que vous puissiez jauger la portée de vos actions passées et que vous en meniez d'autres en conséquence. Ainsi, ces outils vous fournissent des *analytics* (statistiques) sous forme de visualisations agrégées dans un *dashboard*. Les visualisations d'information les plus importantes et utiles à la réalisation d'un ou plusieurs objectifs sont consolidées et disposées sur un seul écran afin que les informations puissent être suivies en un coup d'œil (Few, 2006). Face à cet outil, vous prenez dès lors le rôle de l'explorateur qui sera amené à prendre des décisions. Attention toutefois à rester critique par rapport aux instruments proposés par des plateformes externes : que se cache-t-il derrière les mesures présentées ? Est-ce toujours transparent ? C'est dès lors à l'utilisateur de se poser les bonnes questions quant à la pertinence et à l'efficacité des outils proposés.

2.3 Visualisation d'information, de données : synonymes et faux-amis

2.3.1 Définition générale

La visualisation d'information (infovis) se définit comme l'utilisation de représentations visuelles interactives, assistées par ordinateur, mettant en scène des données pour amplifier la cognition

(Card et al., 1999)

Par « assistées par ordinateur », on entend que ces visualisations sont réalisées grâce à un ordinateur et la plupart du temps affichées sur un écran.

Par « interactives », on entend que ces visualisations permettent un flux d'information bidirectionnel entre un ordinateur et un utilisateur.

Les représentations sont des informations affichées sous une forme visuelle en utilisant des attributs tels que l'emplacement, la longueur, la forme, la couleur et la taille pour former une image et permettre aux utilisateurs de voir des tendances et des exceptions qui ne seraient pas visibles autrement.

Amplifier la cognition, c'est permettre une économie cognitive en augmentant la mémoire.

Aujourd'hui, la visualisation de données peut être définie comme une représentation visuelle assistée par ordinateur, soit statique, soit interactive et mouvante (Card, 2012), dont le principal objectif est « de traduire des données sous une forme visuelle pertinente, simple, didactique et pédagogique » (Brasseur, 2015), dans un souci de lisibilité et de mémorisation motivé par l'apparition de « corrélations visuelles ou des relations d'ordre entre les données » (Saulnier, Thièvre, Viaud, 2006, p. 57). Cette représentation est génératrice d'une « amplification cognitive en termes d'acquisition et d'utilisation de l'information » (Chauvin, 2005). On peut aussi noter que « la visualisation de données désigne tous types de représentations visuelles qui facilitent l'exploitation, l'analyse et la communication de données » (Fredriksson, 2015, p. 36).

Pour résumer très simplement, il s'agit de la représentation et la présentation de données pour faciliter la compréhension » (Kirk, 2016, p. 19, notre traduction), et plus particulièrement de la

présentation visuelle de données abstraites (Sallaberry, 2011). Les données abstraites ; c'est ce dont nous parlions précédemment lorsque nous abordions *l'infovis*. Souvent, la visualisation d'information et la visualisation de données sont deux termes utilisés de façon substitutive, comme des synonymes. Mais est-ce vraiment le cas ?

2.3.2 Quelle est la différence entre visualisation d'information et visualisation de données ?

La visualisation d'information vise à transmettre des idées abstraites à un public de manière visuellement intuitive et efficace, afin de stimuler les utilisateurs pour qu'ils acquièrent de nouvelles connaissances. On peut la considérer comme l'étude de représentations visuelles interactives pour renforcer la cognition humaine. La visualisation de données, souvent considérée comme un sous-domaine de la visualisation d'information, est la science de la représentation visuelle des « données » comprises en tant que telles. Ainsi, les données sont de l'information... Mais de l'information, ce n'est pas nécessairement des données statistiques.

La visualisation de données vise également à faire comprendre la signification des données en les plaçant dans un environnement visuel tel que des graphiques statistiques et un affichage graphique afin que les utilisateurs puissent faire des comparaisons ou déterminer la causalité.

Il est évident que visualisation d'information et visualisation de données partagent une histoire et des objectifs communs. Certaines caractéristiques permettent néanmoins de les distinguer. Pour commencer, alors que la visualisation d'information se concentre sur des données abstraites, la visualisation de données se concentre sur l'affichage de nombres et de statistiques plus directement. Malgré une

similarité de progrès mathématiques, algorithmiques et informatiques dans les deux disciplines, la visualisation d'information se concentre davantage sur les perspectives analytiques de l'information, tandis que la visualisation de données se concentre plus techniquement sur une réduction de la dimensionnalité (essayer de réduire l'espace en y affichant toujours autant de données) (Kim *et al.*, 2016).

Au fond, pourquoi une telle distinction alors que l'objectif, au départ, est similaire ? C'est très simple : la visualisation d'information est à considérer au sens large. Il s'agit d'un domaine de recherche à part entière, qui s'intéresse aux propriétés cognitives, perceptuelles et psychologiques des représentations visuelles. C'est donc tout naturellement que la visualisation de données vient s'y inscrire. En somme, l'*infovis* englobe la *dataviz* ! La visualisation d'information, comme son nom l'indique, concerne la mise en scène d'information abstraite en son sens général. La visualisation de données, quant à elle, concerne la visualisation de statistiques d'un point de vue plus technique.

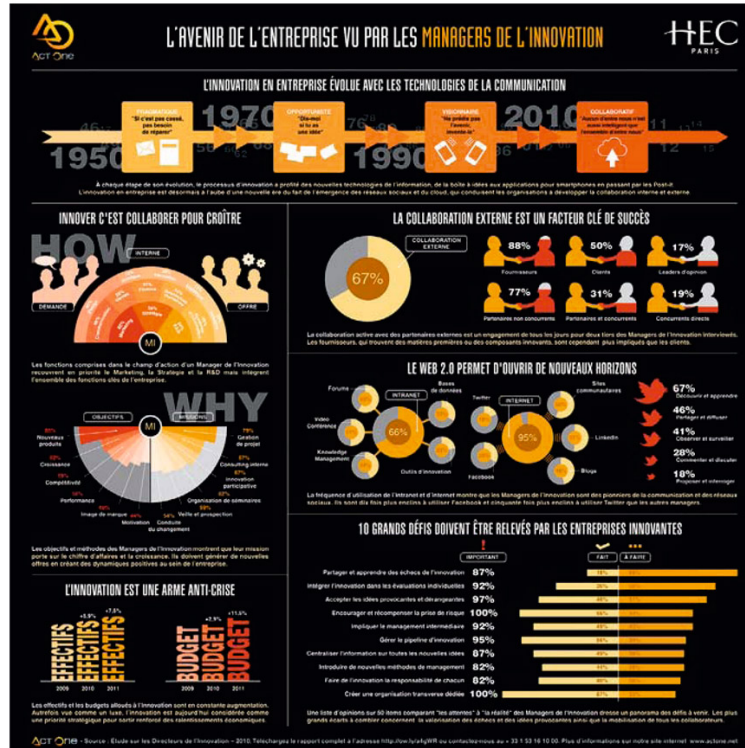


Figure 15 : Exemple d'infographie : Étude HEC Act One, par Laure Joy, 2010 (CC BY-SA)

Si la visualisation d'information a toujours existé, elle a réellement commencé à être théorisée à la fin des années soixante. Elle a constamment évolué, jusqu'à l'apparition de visualisations assistées par ordinateur. L'interaction est alors entrée en jeu et est devenue fondamentale en visualisation d'information. Avec ces mêmes évolutions informatiques s'est également développé le web dans les années 2000, jusqu'à devenir ce que nous connaissons aujourd'hui. L'utilisation de l'Internet et du web favorisent une production immense de données numériques. Ces données sont donc plus facilement comprises une fois visualisées. C'est tout l'enjeu de la visualisation de données : comment visualiser dans un espace restreint une quantité inimaginable de données ? La visualisation de données est également devenue une discipline à part entière, bien qu'intégrée dans la visualisation d'information. Ainsi, bien qu'une

visualisation de données soit une visualisation d'information, l'inverse n'est pas garanti. Un bel exemple est celui de l'infographie, qui visualise beaucoup d'information de façon judicieuse, sans être pour autant une visualisation de données, et ce, tout en étant une visualisation d'information. La figure 15 est un exemple de ce type de visuel. On entendra aussi davantage parler de visualisation d'information en recherche et de visualisation de données en business.

Résumé

Alors qu'on pourrait croire que l'utilisation des visualisations de données est toute neuve, elle ne date en réalité pas d'hier ! Les premières traces de visualisations de données datent de la préhistoire et les racines sociales de cette pratique sont variées. Au fil du temps, la visualisation des données comme objet scientifique s'est développée dans plusieurs disciplines. Il existe deux grandes catégories de visualisation de données : les visualisations scientifiques et les visualisations d'information. Nous nous concentrerons essentiellement sur la visualisation d'information (*infovis*) dans cet ouvrage. Enfin, il est essentiel d'éviter la confusion entre un ensemble de termes utilisés dans le domaine ! Les voici listés :

- ***Dataviz*** : simple contraction des termes « *Data visualization* » en anglais américain.
- ***Infovis*** : simple contraction des termes anglo-saxons « Information visualisation ».
- **Visualisation** : de façon générale, en *infovis* ou en *dataviz*, on peut appréhender le mot « visualisation » de deux façons en français. Il peut d'une part représenter l'acte de visualiser, de transcrire les informations sous une forme visuelle pertinente. C'est le mécanisme technique par lequel l'ordinateur génère l'image. D'autre part, cela peut aussi désigner l'objet final : le graphique, la visualisation. La visualisation, c'est ainsi l'acte de visualiser, tout autant que le résultat final : la représentation graphique. Tous n'expriment pas forcément l'idée du même point de vue.

Infographie : représentation visuelle de plusieurs informations sur un même support visuel, rassemblant du texte, des illustrations, des statistiques et des graphiques.

- ***Visual analytics*** : ils émanent de la visualisation de l'information en se concentrant sur le raisonnement analytique facilité par des interfaces visuelles interactives. Elles sont généralement affichées dans un *dashboard*.
- ***Dashboard*** : affichage des visualisations d'informations les plus importantes et utiles à la réalisation d'un ou plusieurs objectifs. Elles sont consolidées et disposées sur un seul écran afin que les informations puissent être suivies en un coup d'œil (Few, 2006). Il est souvent utilisé afin de prendre des décisions. Un exemple de *dashboard* sur l'évolution du nombre de contaminés lors de la pandémie de coronavirus en 2020 est disponible dans les ressources externes.
- **Graphique** : représentation visuelle mobilisant les dimensions du plan afin de communiquer des données. Se dit « chart » ou « graph » en anglais.
- **Graphe** : dessin mathématique représentant des relations (liens) entre des entités (nœuds). Se dit également « graph » en anglais. Attention aux faux-

amis !

- **Diagramme** : représentation visuelle mettant en relation des éléments dans le plan appartenant à minimum deux composantes.
- **Cartographie** : représentation visuelle illustrant des données de localisation.

Ressources externes et bibliographie indicative

Contenu interactif

1. Une ligne du temps interactive des infographies les plus emblématiques, par R. J. Andrews, 2017 : <https://history.infowetrust.com/>
2. Jalons dans l'histoire de la cartographie thématique, des graphiques statistiques et de la visualisation des données, par Michael Friendly et Daniel J. Denis, 2001 : <https://www.datavis.ca/milestones/>
3. Un exemple de *dashboard* très complet (en anglais) : L'évolution du COVID 19, par The Center for Systems Science and Engineering, Johns Hopkins University : <https://gisanddata.maps.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6>
4. « Quelle différence entre infographie et data visualisation », par Abilways (*Spécialiste de la formation professionnelle des entreprises privées et publiques et des institutions européennes*) : <https://www.youtube.com/watch?v=GgmSehds3tk>.

Bibliographie indicative

Bertin, J. (1967), *Sémiologie graphique*, Paris, Mouton.

Cardon, D. (2012). Regarder les données. *Multitudes*, 49 (2), 138, <https://doi.org/10.3917/mult.049.0138>

Lima, M. (2014). *The book of trees: visualizing branches of knowledge*. Princeton Architectural Press.

Tufte, E. R. (1983). *The visual display of quantitative information*, Cheshire, CT: Graphics Press.

CHAPITRE 2

LA PERCEPTION VISUELLE

1. L'œil
2. La couleur
3. Le traitement préattentif
4. Les lois d'illusion et de perception
5. Conclusion

La visualisation d'information repose sur les propriétés de la perception visuelle. C'est ce qui, d'ailleurs, en fait son succès, à travers l'adage « une image vaut mieux qu'un long discours ».



Figure 16 : Le vase de Rubin (CC BY-SA)

Mais lorsque l'on regarde une visualisation de données, que se passe-t-il ? Grâce à l'œil, nous faisons fonctionner la **vue**, qui est tout naturellement le sens nous permettant de voir. La **vision** est l'interprétation cognitive du sens de la vue.

Que voyez-vous sur la figure 16 ? Deux visages ? Un vase ? Les deux ?

Cette illusion d'optique, dite « ségrégation fond-forme », vous fait ainsi prendre conscience que la vue et la perception des formes ne sont pas les seules choses qui se produisent lorsque vous regardez une image. Votre cerveau produit une interprétation et crée du sens en un rien de temps. Il s'agit de la **vision**, qui n'est donc pas seulement un processus physique, mais aussi cognitif.

Vous le comprendrez aisément, l'usage grandissant des visualisations d'information n'est pas à l'abri de biais induit par la conception ou l'interprétation. Dans la plupart des cas, ces biais amènent des incompréhensions ou servent parfois à de la manipulation délibérée menant à des prises de décisions pouvant être lourdes de conséquences.

Dans ce chapitre sur le pouvoir de la vision, notre objectif est de comprendre comment la perception visuelle fonctionne. Nous aborderons également la couleur, le traitement préattentif et différentes lois de perception, comme les lois de la *Gestalt Theory*. Comprendre ces éléments est un prérequis afin de réaliser de bons choix de conception de visualisations de données.

1. L'œil

1.1 Les sens humains

Selon Aristote (384-322 av. J.-C), philosophe grec de l'Antiquité, la perception sensorielle des humains comprend cinq sens : le goût, l'odorat, l'ouïe, le toucher et la vue. Bien qu'il n'y ait pas de consensus en neurophysiologie concernant le nombre exact de sens humains, il semble désormais admis que la perception sensorielle des humains est bien plus vaste et comporterait au moins neuf sens, soit au moins quatre de plus que les cinq sens d'Aristote : le sens de l'équilibre, le sens du mouvement et de position du corps, le sens de la douleur, et le sens du chaud et du froid. Dans les ressources externes, vous trouverez en guise d'exemple une vidéo abordant les sens humains⁴.

L'équibrioception est un sens physiologique qui existe grâce aux yeux et à la perception visuelle, aux oreilles et à leur système vestibulaire, ainsi qu'aux autres sens spatiaux (proprioception) (Benamran, 2015).

La proprioception est aussi appelée kinesthésie. C'est la perception de la position de son propre corps, consciente ou inconsciente. Elle est rendue possible par l'ensemble des informations nerveuses transmises au cerveau par les différents récepteurs présents sur nos muscles (Benamran, 2015).

La nociception est le sens de la douleur. Elle est un sous-sens du toucher, car notre peau est recouverte de nocicepteurs, tout comme d'autres parties de notre corps (exemple ; une rage de dents) (Benamran, 2015).

La thermoception est le sens de la chaleur. La plupart des mammifères ressentent cette chaleur externe à leur propre corps grâce à des capteurs de chaleur pour le chaud et de capteurs d'absence de chaleur pour le froid (Benamran, 2015).

1.2 Fonctionnement de l'œil et de la vue

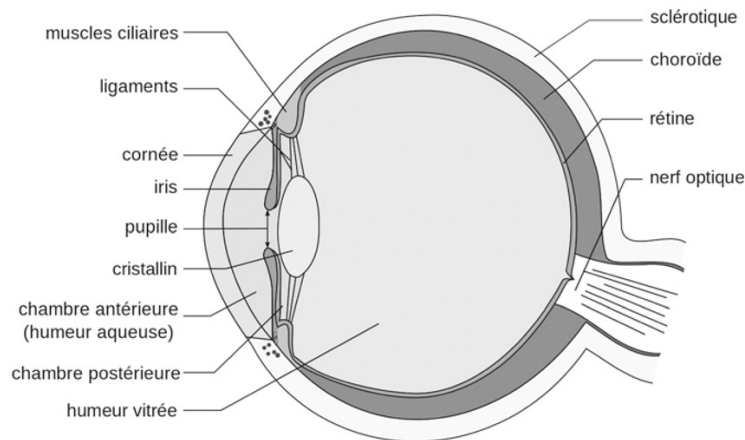


Figure 17 : Schéma d'une coupe longitudinale d'œil humain, par Talos, Jakov et Falcox, 2014, (CC BY-SA)

L'œil est l'organe récepteur de la vue, logé dans l'orbite, et dont les principales parties sont la pupille, le cristallin et la rétine. La vue est la capacité sensorielle (sens) qui nous permet de voir. La vision, ou perception visuelle, fait référence aux processus cognitifs et

physiologiques par lesquels la lumière émise par l'environnement détermine les détails des représentations sensorielles (formes, couleurs, textures, mouvement, distance et relief). Ces processus permettent au cerveau d'interpréter le sens de la vue.

La cornée est le premier élément de l'œil rencontré par la lumière. L'iris, par mouvements de fermeture et d'ouverture, permet de réguler la quantité de lumière perçue. C'est par la pupille qu'entre la lumière dans l'œil, afin d'atteindre le cristallin. Le cristallin peut faire varier la focalisation de la lumière afin de faire converger les rayons lumineux sur la rétine, au fond de l'œil. Ce phénomène s'appelle l'accommodation. La rétine est pourvue de récepteurs de lumière : les bâtonnets (sensibles à la luminosité) et les cônes (sensibles aux couleurs). L'accommodation permet d'assurer la netteté de l'information visuelle qui est envoyée au cerveau grâce au nerf optique. L'information visuelle est donc un message de type nerveux.

La rétine comporte de nombreux récepteurs sensibles à la lumière. Ils sont appelés cellules photoréceptrices ou photorécepteurs. Les photorécepteurs sont les cellules qui réagissent aux radiations lumineuses. Les cônes sont les cellules photoréceptrices localisées au centre de la rétine qui permettent la vision des couleurs. Les bâtonnets sont les cellules photoréceptrices localisées en périphérie de la rétine qui réagissent aux faibles intensités lumineuses. Les neurones (ganglionnaires et bipolaires) transfèrent les informations de la rétine au cerveau par l'intermédiaire du nerf optique.

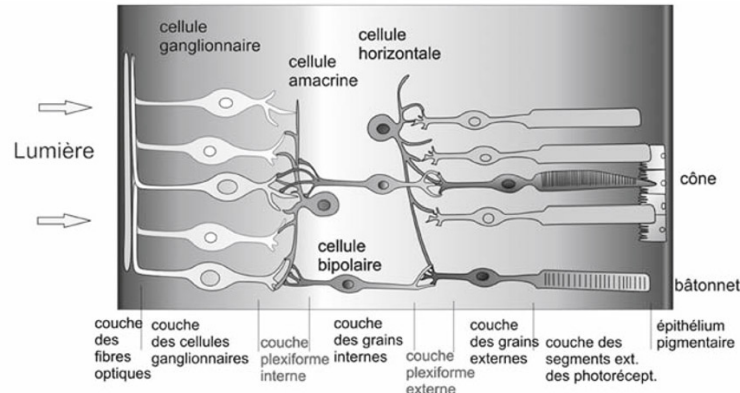


Figure 18 : Structure de la rétine, imitée de Purves *et als* Neurosciences, De Boeck, 2005, par Pancrat (CC BY-SA)

Les humains et quelques autres primates ont une vision trichromatique, c'est-à-dire impliquant trois types de cônes : les cônes S (pour « *short* » en anglais, longueur d'onde courte) associés aux couleurs bleuâtres avec un maximum dans les violet-bleus ; les cônes M (pour « *medium* » en anglais, longueur d'onde moyenne) associés aux couleurs vertes ; et les cônes L (pour « *long* » en anglais, longueur d'onde longue) associés aux couleurs vertes à rouges avec un maximum vert-jaune. Le daltonisme résulte d'une déficience d'un ou plusieurs des trois types de cônes de la rétine oculaire qui affecte la perception des couleurs (exemple : vert et rouge identiques). Les oiseaux et les poissons ont une vision quadrichromatique, c'est-à-dire impliquant quatre types de cônes. Cette caractéristique leur permet de distinguer des surfaces que les humains perçoivent comme identiques.

1.3 La vision et l'intensité lumineuse : vision diurne, nocturne et crépusculaire

Trois réseaux peuvent être mobilisés en fonction de l'intensité lumineuse : les cônes pour la vision diurne (forte intensité lumineuse), les bâtonnets pour la vision nocturne (faible intensité lumineuse).

lumineuse), et les bâtonnets périphériques et les cônes centraux pour la vision crépusculaire.

La vision **photopique** est la forme que prend la vision de jour, lorsqu'il y a beaucoup de lumière (ou, de manière plus générale, lorsque l'éclairage est important). En raison de la forte intensité lumineuse, ce sont les cônes qui jouent un rôle important.

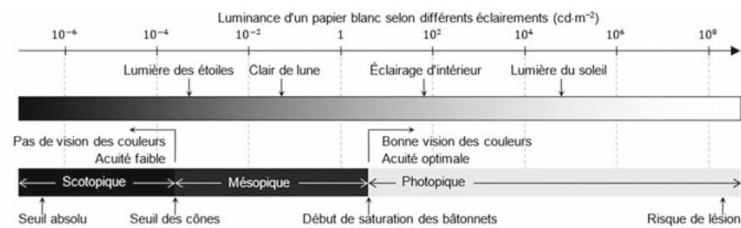


Figure 19 : Domaines de vision d'après Robert Sève, Science de la couleur, 2009, (ISBN 2-9519607-5-1) (CC BY-SA)

La vision **scotopique** est la forme que prend la vision de nuit, lorsqu'il y a peu de lumière (ou, de manière plus générale, lorsque les conditions d'éclairage sont basses). Les bâtonnets y jouent un rôle important en raison de la faible intensité lumineuse.

La vision **mésotopique** est la forme que prend la vision au crépuscule (ou, plus généralement, lorsque les conditions d'éclairage sont moyennes). Les bâtonnets périphériques et les cônes centraux sont mobilisés dans ce cas.

Découvrez l'aperçu de la vision de jour et de nuit sur la Figure 20.



Figure 20 : Le Bode Museum et la Tour de Télévision à Berlin, vue de jour et de nuit, [Pixabay.com](https://pixabay.com) (CC0)

2. La couleur

2.1 D'où viennent les couleurs ?

La visibilité d'un objet est rendue possible par la lumière. En réalité, les couleurs n'ont pas d'existence propre : l'image perçue par l'œil constitue un message nerveux qui est envoyé au cerveau via le nerf optique. Grâce au cerveau, nous percevons ces images colorées : il « construit » les couleurs grâce aux messages électriques en provenance de la rétine (Valeur, 2016). Souvenez-vous du paysage berlinois présenté précédemment (figure 20). Ce qui diffère sur ces photographies d'un paysage semblable, c'est l'intensité de l'éclairage au moment où les photos ont été capturées. L'éclairage de jour est élevé et stimule les cônes à travers la vision photopique tandis que l'éclairage de nuit est faible à modéré, ce qui stimule davantage les bâtonnets et la vision scotopique. Ainsi, le ciel n'est pas bleu en soi et les arbres ne sont pas verts en soi non plus. C'est le cerveau qui, à travers le niveau de lumière disponible, fait percevoir ces éléments comme tels (Valeur, 2016). En somme, l'œil voit très distinctement les couleurs de jour uniquement grâce aux cônes qui sont sensibles aux fortes intensités lumineuses. La couleur est visible grâce à la lumière ; si elle n'est pas présente ou en faible intensité, il est tout simplement impossible de voir la couleur.

2.2 Les conditions de visibilité d'un objet

Tout est donc une question de lumière. Il y a trois conditions à satisfaire afin de voir un objet :

1. L'objet est éclairé par une source lumineuse (solaire, lunaire ou artificielle).
2. L'objet diffuse des rayons lumineux. Il réfléchit la lumière extérieure ou diffuse sa propre lumière.

3. Les rayons lumineux diffusés par l'objet entrent directement dans l'œil.

La diffusion est le phénomène qui se produit lorsque la lumière frappe un corps qui la renvoie alors dans toutes les directions. Sur la figure 21, la lumière du Soleil réfléchiée par l'arbre entre directement et en ligne droite dans l'œil. Évidemment, sur ce schéma, le mur empêche la lumière d'entrer dans l'œil du bas, ce qui ne permet pas de voir l'arbre. L'interaction de l'œil avec la lumière permet la perception des couleurs.

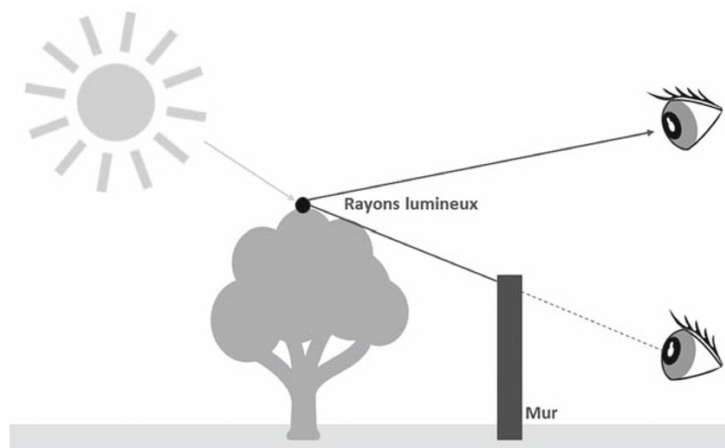


Figure 21 : Diffusion de la lumière, par les auteurs

2.3 La longueur d'onde

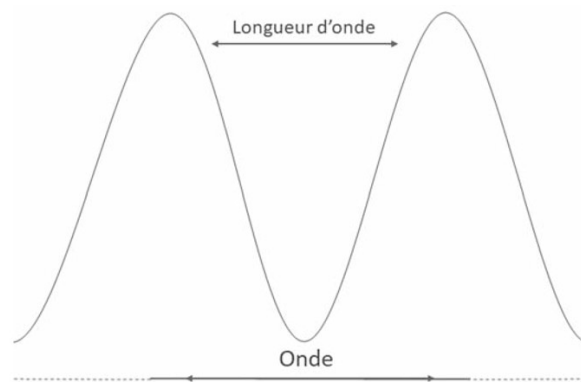


Figure 22 : La longueur d'onde, par les auteurs (CC0)

La lumière est composée d'ondes électromagnétiques. Ces ondes sont caractérisées par leur longueur. La longueur d'onde s'apparente à la distance entre deux phases ondulatoires, comme dans la figure 22. On dit alors que l'onde est courte lorsqu'il y a une faible distance entre deux crêtes successives. Si cette distance est grande, on parle « d'onde longue ». La lumière se présente à nous comme un ensemble d'ondes électromagnétiques de longueur différente. Les ondes frappent différents objets. Certaines d'entre elles sont absorbées, d'autres sont réfléchies par les objets. En réfléchissant ces ondes, et donc cette lumière, les objets font parvenir la lumière jusqu'à l'œil. Cela se rapporte d'ailleurs aux trois conditions de visibilité d'un objet abordées sur la page précédente (Sève, 2009).

2.4 L'interprétation des ondes par le cerveau

L'œil perçoit les couleurs en journée, lorsque la vision photopique fonctionne, stimulant ainsi les cônes présents sur la rétine. Il y a trois types de cônes, correspondant à des longueurs d'ondes différentes :

- Les cônes « S » (*short*), qui correspondent aux longueurs d'ondes courtes, jusqu'à 437 nanomètres. Ils correspondent ainsi aux nuances de bleu.
- Les cônes « M » (*medium*), qui correspondent aux longueurs d'ondes moyennes, jusqu'à 533 nanomètres. Ils correspondent ainsi aux nuances de vert.
- Les cônes « L » (*long*), qui correspondent aux longueurs d'ondes longues, jusqu'à 564 nanomètres. Ils correspondent ainsi aux nuances de rouge.

En vision nocturne (ou scotopique), les bâtonnets sont sensibles à une longueur d'onde allant jusqu'à 498 nanomètres. La figure 23

permet de voir à quoi correspondent les différents cônes et les bâtonnets.

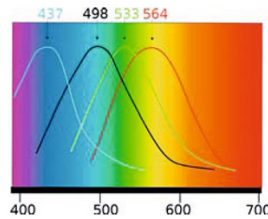


Figure 23 : Spectre visible d'absorption des cônes. La courbe noire représente les bâtonnets. Par Pancrat, 2011 (CC BY-SA)

Toute une gamme d'ondes entre simultanément dans l'œil qui voit plusieurs couleurs en même temps. Plusieurs cônes réagissent à différents degrés. Le cerveau interprète alors une information en mélangeant les sensations colorées : l'association du bleu et du rouge forment ainsi la sensation colorée « magenta », par exemple. Les différentes longueurs d'ondes interprétées par notre cerveau font donc ressentir différentes sensations colorées.

L'étendue des longueurs d'ondes perceptibles par l'œil humain forme ce que l'on appelle « le spectre visible ». Cela signifie qu'il existe d'autres longueurs d'ondes que nous ne percevons pas, comme les micro-ondes. Le spectre visible fait partie du spectre électromagnétique, rassemblant toutes les ondes électromagnétiques qui existent. Les longueurs d'onde décrites en nanomètres expriment les caractéristiques physiques des couleurs. Les teintes produites par le cerveau grâce aux messages nerveux encodés par les différents cônes et transmis par le nerf optique sont les caractéristiques psychologiques de la couleur.

Le bleu, le vert et le rouge sont des teintes. Il est possible de caractériser la couleur autrement encore. Deux autres éléments influencent la façon dont la couleur est perçue. Ce sont aussi des

interprétations psychologiques de la couleur provenant d'un phénomène physique. Il s'agit de la saturation et de la luminosité.

Teinte, saturation et luminosité

Ainsi, la couleur est l'association de la teinte, de la saturation et de la luminosité qui permettent au cerveau de transformer les caractéristiques physiques des ondes en une sensation psychologique, qui nous permet donc d'interpréter ce qui nous entoure :

- La teinte : plusieurs longueurs d'onde sont présentes et réfléchies simultanément. Il y a une longueur d'onde dominante.
- La saturation : plus il y a de longueurs d'onde différentes de la dominante, plus la saturation est faible. Moins il y a de longueurs d'onde différentes de la dominante, plus la saturation est élevée. Si toutes les ondes sont présentes en une même quantité, nous voyons une tonalité de gris.
- La luminosité : plus la quantité de longueurs d'onde augmente, plus la luminosité est forte. Plus la quantité diminue, plus la luminosité est faible. Si toutes les longueurs d'ondes sont présentes en quantité maximale, nous percevons du blanc. À l'inverse, si elles sont absentes, nous percevons du noir.

La couleur n'est donc pas seulement une question de teinte. Dans les logiciels de conception d'image par exemple, la notion de RGB (Red Green Blue) et de HSL (Hue Saturation Luminance) est souvent utilisée pour permettre à l'utilisateur de définir la couleur la plus précise possible. Dans les ressources externes, vous pourrez faire varier ces éléments sur une roue colorimétrique afin de les comprendre plus aisément.

2.5 La synthèse additive et la synthèse soustractive

Il est possible de faire apparaître la couleur en utilisant deux procédés : la synthèse additive ou la synthèse soustractive. La figure 24 en montre un exemple.



Figure 24 : Synthèse additive et soustractive, par les auteurs

Les couleurs primaires de la synthèse additive sont le rouge, le vert et le bleu. En synthèse additive, la somme de trois lumières de couleurs primaires de même intensité donne du blanc. En faisant varier la proportion de chacune des lumières, on peut obtenir la couleur de son choix. En synthèse additive, on parle de superposition de lumières colorées et non de couleurs. Le blanc est la somme de toutes les couleurs uniquement lorsque l'on additionne plusieurs sources lumineuses (Lejeune *et al.*, 2006). C'est le procédé utilisé par les écrans.

La synthèse soustractive consiste à supprimer une partie du spectre de la lumière blanche pour obtenir une couleur différente. Cela consiste à combiner les propriétés d'absorption de plusieurs corps pour obtenir une nouvelle couleur. Cela peut impliquer la superposition de filtres ou encore le mélange de plusieurs couleurs. La synthèse soustractive joue sur la capacité d'un pigment à ne diffuser que certaines teintes (Lejeune *et al.*, 2006). Lorsque l'on mélange toutes les couleurs d'une boîte de peinture, on obtiendra du marron ou du noir. De la même façon, en synthèse soustractive,

lorsque l'on dit que les couleurs primaires sont le cyan, le magenta et le jaune, ce n'est vrai que lorsque l'on parle de pigmentation. C'est le procédé utilisé par les imprimantes.

2.6 La couleur en contexte : contrastes et illusions

Une fois contextualisée, une couleur peut être perçue différemment en fonction de son environnement. Michel Eugène Chevreul s'en rend compte en 1824, lorsqu'il est nommé directeur des teintures à la Manufacture des Gobelins. Il était difficile à l'époque de reproduire à l'identique certaines couleurs sur le tissu. Chevreul s'est alors rendu compte que le problème ne venait pas de la chimie, mais de notre perception visuelle : des couleurs sont perçues différemment si elles sont juxtaposées, disposées les unes à côté des autres ou regardées seules (Valeur, 2016).

Les différents contrastes peuvent donc nous faire percevoir les couleurs différemment.

Vous découvrez ainsi :

- **Le contraste de luminosité** (Figure 25) : une même couleur est perçue plus foncée sur un fond clair que sur un fond sombre (Parsons, 1924). Sur la figure 31, le carré de couleur moutarde semble plus foncé lorsqu'il est contextualisé dans le carré gris plutôt que dans le carré noir. Autrement dit, le carré noir rend cette couleur plus saillante. L'œil s'adapte à l'intensité lumineuse moyenne d'une scène. D'ailleurs, de manière générale, on préconise en visualisation de données d'employer un fond blanc ou gris afin d'éviter les effets de contraste et de problèmes de visibilité.



Figure 25 : Contraste de luminosité, par les auteurs (CC0)

- **Le contraste de saturation** (Figure 26) : pour l'œil, la vivacité d'une couleur dépend de la scène environnante. Ainsi, une couleur semblera plus pâle lorsqu'elle sera contextualisée dans un ensemble plus soutenu. Elle semblera donc plus vive si elle est entourée d'une couleur fade (Parsons, 1924). Sur la figure 26, le plus petit carré, de couleur identique dans les deux contextes, semble plus vif dans un environnement gris, qui est fade.



Figure 26 : Contraste de saturation, par les auteurs (CC0)

- **Le contraste simultané** (Figure 27) : l'éclat de deux couleurs complémentaires est plus intense lorsqu'on les juxtapose. Lorsqu'il perçoit une couleur, l'œil suscite naturellement la perception de la couleur qui lui est complémentaire. Deux couleurs complémentaires mises l'une sur l'autre se renforcent donc mutuellement (Valeur, 2016). Par exemple, le cyan est complémentaire au rouge. C'est la raison pour laquelle les carrés rouges ressortent davantage sur le carré cyan de la figure 27. Vous en doutez ? C'est peut-être à cause de la proximité des carrés bleus et jaunes. Cachez simultanément le carré jaune puis bleu pour mieux constater la saillance des carrés rouges !



Figure 27 : Contraste simultané, par les auteurs (CC0)

- **Le contraste successif** (Figure 28) : il s'agit d'une illusion qui peut faire voir une couleur qui n'apparaît pas. Pour comprendre cela, une petite expérience est possible : sur la figure 28, fixez les cercles bleus pendant environ 20 secondes. Ensuite, maintenez votre regard sur le cercle à droite. Voyez-vous cette tache jaune qui apparaît ? La raison est physiologique et psychologique : lorsque l'on regarde une tache bleue, les neurones stimulés par les cônes « S » (cônes sensibles au bleu) saturent, d'une certaine façon, laissant ainsi la place aux cônes « M » et « L », qui concernent les couleurs complémentaires et les longueurs d'ondes plus grandes. Le temps que la situation des cônes « S » se rétablisse, une sensation de jaune apparaît, grâce aux deux autres types de cônes (Valeur, 2016).



Figure 28 : Contraste successif, par les auteurs (CC0)

2.7 Le daltonisme

Le daltonisme est une anomalie visuelle. Les personnes qui sont concernées par cette infirmité légère (daltoniens) sont atteintes d'une défaillance au niveau des cônes. Un ou plusieurs des trois cônes de la rétine dysfonctionnent. L'anomalie est génétique et est détectée chez l'enfant avant l'âge de 3 ans (Lanthony, 2001). Différentes déficiences existent.

Le dysfonctionnement le plus courant est la confusion du vert et du rouge. Cette déficience atteint principalement les hommes, car les gènes concernant les récepteurs de ces couleurs se concentrent sur le chromosome X. Les hommes ont un seul chromosome X (XY). S'il est atteint, l'individu sera daltonien. Les femmes (XX), en revanche, ne sont daltoniennes que si leurs deux chromosomes X sont atteints. Le daltonisme peut donc être transmis aux filles qui sont porteuses saines et qui peuvent ensuite le transmettre à leurs propres enfants. La déficience du bleu se répartit quant à elle équitablement entre hommes et femmes parce que le gène codant les récepteurs de la couleur bleue se trouve sur le chromosome 7 (Lanthony, 2001). Le daltonisme concerne 8 % de la population masculine et 0,5 % de la population féminine. En Belgique, il y a environ 458 000 daltoniens (Vision des couleurs, 2020).

La visualisation d'information et les daltoniens

En visualisation d'information, la couleur est très importante parce qu'elle permet de coder de l'information. Pour les daltoniens, cela peut vite devenir problématique. Par exemple, le vert et le rouge sont deux couleurs extrêmement utilisées en raison de leur forte symbolique culturelle : le vert est perçu comme « bon » et le rouge comme « mauvais », à l'image du feu de signalisation où le vert signifie « passez » et le rouge « arrêtez-vous ». Mais comment un daltonien atteint de déficience vert-rouge perçoit-il le feu de signalisation ? La figure 29 représente la complexité de saisir la différence de couleur pour ces personnes.

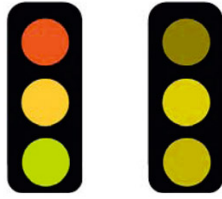


Figure 29 : Différence de perception des couleurs d'un feu de signalisation entre une personne à la vision normale (gauche) et une personne atteinte de daltonisme (déficience rouge-vert ; droite), par les auteurs. (CC0)

Si, concernant les feux de signalisation, il est possible de se remémorer l'ordre précis dans lequel les ampoules s'illuminent, il est évidemment moins aisé de comprendre une information codée dans les nuances de rouge et de vert pour un daltonien qui la voit pour la première fois. L'utilisation du bleu peut éventuellement contrer le problème.

Évidemment, les personnes atteintes de daltonisme sont minoritaires dans la population. Mais, dans les cas où le concepteur d'une visualisation de données sait qu'il y a un risque que son audience comporte des daltoniens, il peut adapter son visuel en conséquence. La préconisation majeure que l'on retrouve dans le domaine est de choisir d'autres éléments que la couleur pour communiquer l'information, afin d'éviter tout risque de confusion. Ainsi, au lieu d'utiliser une légende colorée pour communiquer de l'information, le concepteur choisira des formes ou des hachures, et ce, dans la mesure du possible et dans le respect des différents principes de conception de visualisations d'informations. Ces principes suivent dans les chapitres suivants.

À l'heure où les écrans sont énormément utilisés, la question du daltonisme suscite beaucoup d'intérêt, comme dans cette ressource externe⁵ abordant la conception d'interfaces pour daltoniens.



<https://colour-blindness.org/>

3. Le traitement préattentif

3.1 Définition

Le traitement préattentif est l'accumulation subconsciente d'informations provenant de l'environnement qui nous entoure. L'information disponible est traitée de manière préattentive. Cela signifie que le cerveau filtre et traite ce qu'il choisit comme important (cette action n'est pas consciente). Les informations qui ont la plus grande importance déterminée de façon subconsciente par le cerveau sont alors sélectionnées pour une analyse plus approfondie qui est, elle consciente et attentive (Van der Heijden *et al.*, 1996).

Le traitement préattentif touche tous les sens, mais nous nous concentrerons uniquement sur la vision. Il est davantage connu dans la littérature anglo-saxonne, dans laquelle il est appelé *preattentive processing*.

Sur la figure 30, il est possible de compter les boules rouges en une fraction de seconde. Dans le cas de cette image, la couleur est traitée de façon préattentive par le cerveau et permet de mettre ces boules rouges en évidence par rapport aux bleues. De plus, lorsqu'il y a moins de quatre éléments (donc moins de quatre boules rouges dans ce cas), il n'y a même pas besoin de compter. « La compréhension de ce qui est traité de manière préattentive est probablement la contribution la plus importante que la science visuelle puisse apporter à la visualisation des données » (Ware, 2004, p. 145, traduction libre).

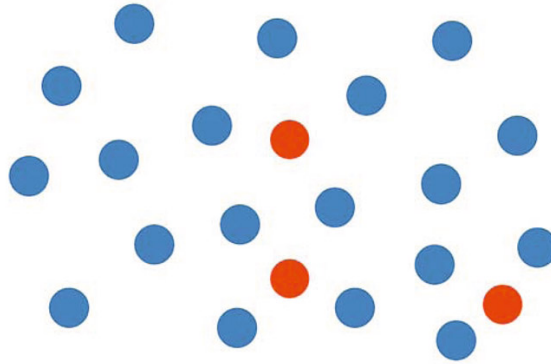


Figure 30 : Dispositions préattentives de la couleur, par les auteurs (CC0)

Le traitement préattentif se produit en moins de 500 millisecondes lorsqu'il faut effectuer une tâche et en moins de 10 millisecondes lorsque l'on observe un objet individuel. Il s'agit d'un traitement parallèle : on perçoit en un coup d'œil, tandis qu'en traitement attentif, le traitement est séquentiel. On réfléchit à une seule chose à la fois, ce qui demande un minimum de concentration.

Tableau 1 : Temps de traitement pour une tâche, un objet individuel et la temporalité

	Traitement préattentif	Traitement attentif
Tâche	≤ 500 ms	> 500 ms
Objet individuel	≤ 10 ms	> 10 ms
Temporalité	Traitement parallèle	Traitement séquentiel

3.2 Exemple

Soit la tâche « compter les occurrences du chiffre 3 » (Figure 31). L'exécution de la tâche sur l'affichage (a) requiert la lecture scrupuleuse et donc le balayage séquentiel de chaque chiffre. La durée d'exécution est alors de l'ordre de la dizaine de secondes.

Les caractéristiques préattentives déterminent les informations qui attirent notre attention. C'est important dans la visualisation d'information, car cela permet d'orienter l'attention du destinataire vers les informations les plus importantes du graphique. Dans la figure 32, vous pouvez voir clairement que les différentes caractéristiques préattentives attirent votre attention. C'est le cas de l'orientation, de la courbure, de la forme, de la taille, du nombre, de la couleur/valeur, de la « clôture », de la forme concave ou convexe ou encore de l'addition. Mais restez prudent ! Certains éléments peuvent donner l'impression d'être préattentifs sans l'être, comme la jonction ou le parallélisme.

Il est possible de détecter le stimulus visuel le plus saillant visuellement (i.e., ayant le plus de poids visuel) de manière préattentive dans toutes les images ci-dessus. Une astuce pour déterminer si un stimulus visuel est préattentif ou non consiste à se demander ce qu'il se passerait si l'on augmentait le nombre de distracteurs. Un distracteur est un stimulus visuel moins saillant visuellement.

Collin Ware (2004, p. 152) illustre ce phénomène avec l'exemple suivant. Il est facile de repérer un seul faucon dans un ciel plein de pigeons, mais si le ciel contient une plus grande variété d'oiseaux, le faucon sera plus difficile à voir. Selon l'auteur, deux facteurs sont importants pour déterminer si une cible se démarque de manière préattentive : le degré de différence de la cible par rapport aux distracteurs et le degré de différence des distracteurs les uns par rapport aux autres (Quinlan et Humphreys, 1987 ; Duncan et Humphreys, 1989) cités dans (Ware, 2004 ; p. 152).

Les caractéristiques préattentives peuvent être organisées en un certain nombre de catégories (Ware, 2004) :

- Forme d'une ligne : orientation, longueur, largeur.

- Forme d'un objet : taille, courbure, regroupement spatial, effet flouté (*blur*), marques ajoutées.
- Couleur : teinte, saturation, luminosité.
- Mouvement : scintillement, clignotement, direction du mouvement.
- Position spatiale : position 2D, profondeur stéréoscopique, forme convexe/concave de l'ombrage.

Ces caractéristiques sont très importantes en visualisation d'information, car elles entrent en compte dans les lois de perception visuelle, de la *Gestalt Theory* (chapitre suivant) ainsi que différents principes de conception de visualisations d'information.

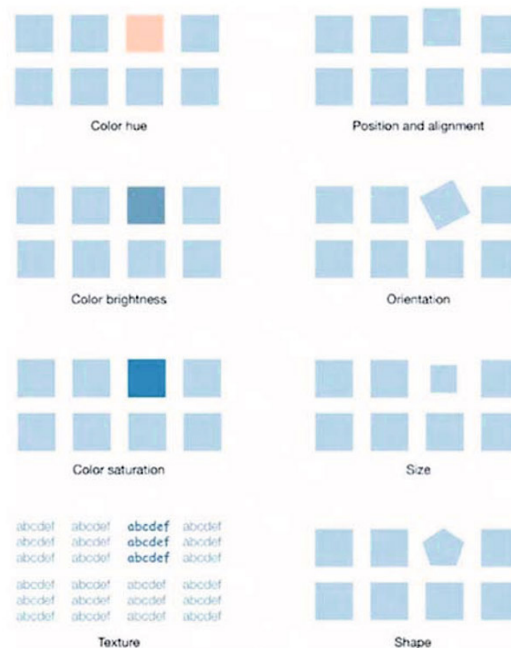


Figure 32 : Caractéristiques pré-attentives, par Meedanphotos (CC BY 2.0)

3.4 L'attention sélective

Malgré tout, la cognition est limitée. L'attention sélective nous permet de favoriser le traitement d'une caractéristique pertinente d'un

stimulus, tout en ignorant les éléments distracteurs présents dans le même environnement. Ainsi, l'attention visuelle est la capacité à maintenir les processus mentaux concentrés sur quelque chose que nous observons. Comme pour les autres types d'attention, elle peut être sélective (Mialet, 1999). Il se pourrait ainsi qu'en se concentrant sur un élément en particulier, on ne remarque pas les éléments autour. La meilleure expérience à ce sujet est celle du « Monkey Business », menée par Simon et Chabris (1999). Prenez-en connaissance grâce à la vidéo intitulée « Le test d'attention sélective de Simon et Chabris (1999) », répertoriée dans les ressources externes de la fin de ce chapitre.

3.5 La cécité au changement

La cécité au changement est un phénomène de la perception visuelle qui se produit lorsqu'un changement dans un stimulus visuel est introduit et que l'observateur ne le remarque pas. Il s'agit d'une limite fondamentale de l'attention humaine (Rensik *et al.*, 1997). Connaissez-vous le jeu des 7 différences ? Il est révélateur de ce phénomène puisque vous avez besoin de vous concentrer pour détecter ces différences.

De même, sur la figure 33, il vous aura peut-être fallu un moment pour remarquer que la branche d'arbre a disparu sur la seconde image, ou bien encore que des personnes ont disparu sur la gauche. Dans les ressources externes de ce chapitre, une vidéo intitulée « *The door study* » vous montre une expérience de cécité au changement.



Figure 33 : Globe and high court, 2009, Par Jiron (CC BY-SA)

4. Les lois d'illusion et de perception

4.1 La perception dépasse la simple sensation : vers la reconstruction visuelle

En psychologie, la théorie de *Gestalt*, ou théorie de la forme, permet de comprendre, si l'on retourne à ses prémices, pourquoi la perception ne se résout pas seulement au bon fonctionnement de nos sens. Elle est aussi une réaction qui consiste à répondre à un signal. Évidemment, les sens sont indispensables au bon fonctionnement de la perception. La vue est un sens précurseur, c'est-à-dire qui fonctionne « à distance ». Cela nécessite, pour chacun, le développement « d'un espace subjectif ». De cette manière, selon les gestaltistes :

- les sens ont d'abord induit des réactions à des signaux. Par apprentissage, ces signaux sont devenus, par procédés complexes, des « formes » ;
- du fait de l'apparition de ces formes, les significations se sont étendues ;
- le tout mène progressivement vers la représentation et la symbolisation, qui poussent à s'interroger sur les liens entre perception et cognition.

(Thinès, 2016, en ligne)

En réalité, la perception visuelle répond à un problème de taille : celui de faire correspondre un stimulus avec un objet physique, voire avec un objet abstrait (comme la droite géométrique) (Froment, Masnou, Morel, 1998, p. 1). Dès lors, c'est une « reconstruction visuelle » qui s'opère. Les principes de la reconstruction visuelle ont été mis au jour par les chercheurs de la *Gestalt Theory*. Il existe donc le « gestaltisme géométrique », selon lequel toute vision est une illusion. Il arrive que la reconstruction visuelle échoue : dès lors, « la correspondance entre objet perceptif et objet physique est erronée. Il s'agit donc d'illusions visuelles distorsives, encore qualifiées d'optico-géométriques » (Froment, Masnou et Morel, 1998, p. 1). Par ailleurs, d'autres types d'illusions visuelles existent.

Tous ces genres d'illusions sont générés par trois éléments :

- le phénomène physiologique, c'est-à-dire le fonctionnement physique de notre œil, sensible à la lumière, aux contrastes, aux contours, couleurs et autre ;
- nos capacités cognitives traduites par les facultés de notre cerveau à interpréter les images en fonction de ce que l'on connaît (notre culture, ce qui nous est étranger), à comprendre le mouvement, à établir des corrélations entre divers éléments de l'image, le tout avec ses limites qui lui sont propres ;
- les principes *Gestalt*, que nous verrons par la suite.

(Thinès, 2016 ; Hurter, 2010 ; Telea, 2014)

4.2 Les lois de la Gestalt Theory

Prendre connaissance des lois de la *Gestalt Theory* permet de comprendre quels mécanismes nous développons, visuellement et cognitivement, face aux images que nous voyons. La *Gestalt Theory* est très large. Dans ce chapitre sont seulement abordés les lois et principes concernés par le « domaine de la constitution immédiate

des objets visuels » (Froment, Masnou, Morel, 1998). Effectivement, la *Gestalt Theory* dépasse de loin l'analyse d'éléments visuels d'un point de vue cognitif : la linguistique, la philosophie, la psychologie, la phénoménologie et bien d'autres sont également concernés par cette théorie qui revendique un fonctionnement « holistique » du cerveau. Par ce terme, les gestaltistes veulent dire qu'ils s'intéressent au fonctionnement du cerveau dans sa globalité (Lima, 2013). De cette manière, il fonctionnerait de manière unitaire et, dans son processus cognitif, « la reconnaissance visuelle des formes ne découle plus d'une collection d'éléments – points et lignes – mais opère, quand ils sont vus en tant que motif reconnaissable dans son ensemble » (Lima, 2013). Plus tard mutée en « holisme » (Lima, 2013), la *Gestalt Theory* prône un principe qui dicte toute sa pensée :

Ce que nous percevons comme un tout est différent de ses parties
(Smith-Gratto & Fisher, 1998)

Les lois de la *Gestalt Theory*, fondamentales en perception visuelle, sont utiles pour la visualisation d'information ; correctement appliquées au texte et aux graphiques, elles peuvent « augmenter l'efficacité éducative » (Smith-Gratto & Fisher, 1998, p. 362). Voici les lois et principes les plus connus :

- La **loi de la bonne forme** : un ensemble d'éléments sera d'abord perçu comme une forme simple et complète (Crombez, 2014). C'est de ce principe que découlent les autres lois, puisque le tout est perçu différemment des parties ! Sur la figure 34, vous percevrez ainsi un carré avant de percevoir 16 cercles.

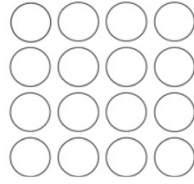


Figure 34 : Loi de la bonne forme, par les auteurs (CC0)

- Le **principe de proximité** : les éléments représentés proches les uns des autres sont perçus comme un groupe (Coren & Girgus, 1980). Ce regroupement perceptif a lieu quand une distance assez faible sépare les sujets en question (Froment, Masnou & Morel, 1998).



Figure 35 : Loi de proximité, par les auteurs (CC0)

- Le **principe de similitude** : de nouveau par regroupements perceptifs, les formes qui se ressemblent sont assemblées (Froment, Masnou & Morel, 1998). Les éléments similaires sont perçus comme groupés ensemble (Coren & Girgus, 1980), de telle sorte que l'attention soit susceptible de se focaliser sur une forme différente (Telea, 2014).

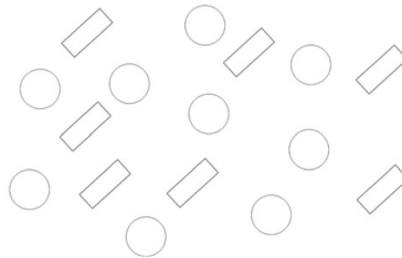


Figure 36 : Loi de similitude, par les auteurs (CC0)

- La **loi de continuité de direction** : l'œil suit les trajectoires naturellement présentes dans l'image (Coren & Girgus, 1980). Il

s'agit d'une « rigidification des objets visuels formés par des courbes régulières » (Froment, Masnou & Morel). Encore une fois, il s'agit d'un regroupement perceptif. Ce principe est lui-même composé...

- Du principe de fermeture (une courbe fermée). L'œil peut également compléter lui-même l'espace, par prolongement de la courbe ou si la forme lui est familière (Telea, 2014).



Figure 37 : Principe de fermeture, par les auteurs (CC0)

- De la complétion amodale : lorsque l'œil détecte une jonction en T, « notre perception l'interprète comme étant le résultat de l'occlusion d'un objet par un second objet placé avant » (Froment, Masnou, Morel, 1998, p. 6).



Figure 38 : Principe de complétion amodale, par les auteurs (CC0)

- Le principe de transparence/ombrage : « présence d'une jonction en X, les quatre régions concernées par la jonction se regroupent par paires de manière à reconstruire deux objets se recouvrant partiellement » (Froment, Masnou, Morel, 1998, p. 8).

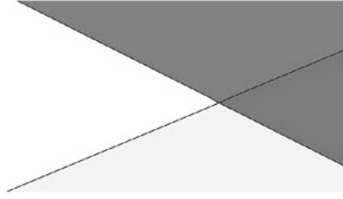


Figure 39 : Principe de transparence/ombrage, par les auteurs (CC0)

- Le principe de perspective. De la manière la plus simple qu'il soit, nous interprétons une scène en perspective quand l'image est perçue comme « la représentation plane d'une image tridimensionnelle » (Froment, Masnou & Morel, 1998 p. 8). De même, nous distinguons ainsi le fond de la figure au premier plan (Telea, 2014).

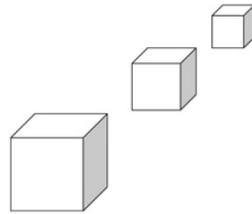


Figure 40 : Exemple de perspective, par les auteurs (CC0)

- La **loi de destin commun** : nous concevons comme une même forme des éléments perçus comme allant dans une direction similaire (Crombez, 2014).



Figure 41 : Loi de destin commun, par les auteurs (CC0)

4.3 La connectivité

La connectivité est plus puissante que la proximité ou la similarité pour le regroupement. Relier différents objets graphiques par des lignes est un moyen très puissant d'exprimer qu'il existe une relation entre eux. La connectivité est la méthode la plus courante pour représenter les relations entre entités.

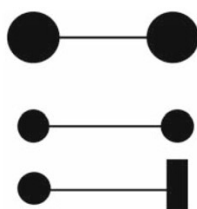


Figure 42 : Représentation de la connectivité, par les auteurs (CC0)

4.4 La ségrégation fond-forme

Une figure est un objet perçu comme étant au premier plan. Le fond est tout ce qui se trouve derrière la figure. La perception de la figure par opposition au fond peut être considérée comme l'acte perceptuel fondamental d'identification des objets. L'exemple ci-dessous montre la figure classique du vase de Rubin, dans laquelle il est possible de percevoir soit deux visages nez à nez, soit un vase noir centré sur l'écran. Le fait que les deux perceptions ont tendance à alterner suggère que des processus concurrents peuvent être impliqués dans la construction de figures.



Figure 43 : Le vase de Rubin, par Bryan Derksen, 2007 (CC BY-SA)

4.5 Les illusions d'optique

Souvenez-vous : au début de ce chapitre, nous assimilions la perception visuelle à une « reconstruction visuelle ». D'ailleurs, selon le « gestaltisme géométrique », toute vision est une illusion. Il arrive que la reconstruction visuelle échoue et que « la correspondance entre objet perceptif et objet physique soit erronée. Il s'agit donc d'illusions visuelles distorsives, encore qualifiées d'optico-géométriques » (Froment, Masnou et Morel, 1998, p. 1). Par ailleurs, d'autres types d'illusions visuelles existent ! Aujourd'hui, les possibilités créatives sont multiples en visualisation d'information. Connaître les différentes illusions d'optique possibles peut guider vers de meilleurs choix de conception.

4.5.1 Les illusions visuelles distorsives

Ce sont les premières illusions qui ont été étudiées. Elles intéressent fortement les psychologues, dont les gestaltistes. D'une part, elles sont intéressantes pour « analyser l'origine de nos représentations de l'espace visuel » du côté des psychologues traditionnels, (Thiéry *et al.*, 1896, p. 496), de l'autre, pour découvrir quels principes guident la reconstruction visuelle pour les gestaltistes (Froment, Masnou, Morel, 1998), elles présentent des éléments qui semblent inégaux ou de tailles différentes alors qu'il n'en est rien. Voici trois illusions distorsives très connues.

L'illusion de Hering

Sur la figure 44, le dessin radial composé par les droites qui apparaissent en fond donne une illusion de perspective qui agit alors sur les droites parallèles qui sont au premier plan. Malgré leur parallélisme, ces droites semblent courbées.

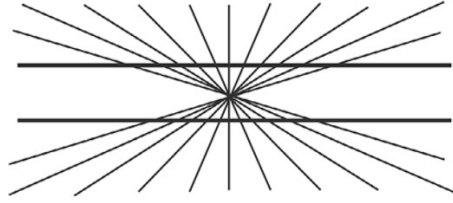


Figure 44 : Illusion de Hering, par Gwestheimer, 2011 (CC BY-SA)

L'illusion de Müller-Lyer

En contemplant la figure 45, vous pourriez avoir l'impression que les deux droites présentées sont de longueurs différentes. Pourtant, elles sont exactement de la même longueur. Elles subissent pourtant un effet de déformation à cause de la disposition de leurs terminaisons (les flèches) qui sont alors des éléments déformants.

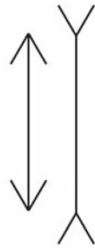


Figure 45 : Illusion de Müller-Lyer, par les auteurs, (CC0)

Illusion de Ponzo

Cette illusion suggère que l'esprit humain juge la taille d'un objet en fonction de son arrière-plan. Ainsi, sur la figure 46, bien que nous voyons deux lignes non parallèles, nous interprétons une perspective, à l'image d'une voie ferrée. Nous imaginons donc deux droites parallèles qui s'éloignent de nous. Ainsi, les segments de droite horizontaux, de taille pourtant identique, nous semblent différents.



Figure 46 : Illusion de Ponzo, par les auteurs, (CC0)

4.5.2 Les illusions ambiguës

Ces illusions révèlent la capacité du cerveau et de l'œil à inverser une position perceptible. Un objet peut-être perçu de deux façons, il y a une double perception (Pritchard, 1958). L'illusion ambiguë la plus connue est le Cube de Necker. En dessin, il existe les lois de la perspective, qui permettent de donner une illusion de profondeur se rapprochant du réel. Les lois de la perspective ne sont pas respectées dans le cube de Necker. Si ces lois étaient respectées, une seule perception serait possible (Orbach J. *et al.*, 1963).

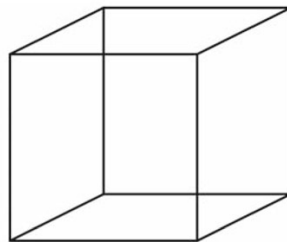


Figure 47 : Le cube de Necker, par BenFrantzDale, 2017, (CC BY-SA)

4.5.3 Les illusions impossibles

Il s'agit d'images en deux dimensions qui donnent l'illusion de percevoir une figure, un objet en trois dimensions. Cet effet peut induire des interprétations perceptuelles contradictoires (Penrose & Penrose, 1956 ; Figures 48 et 49). Finalement, considérée dans son ensemble, la structure semble impossible.

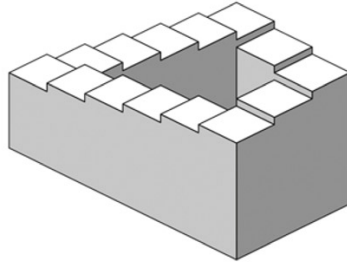


Figure 48 : L'escalier impossible de Penrose et Penrose, 1958
(Domaine Public)

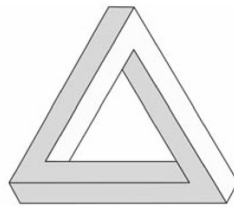


Figure 49 : Le triangle de Penrose et Penrose, 1958 (Domaine Public)

4.5.4 Les illusions de contraste

Différentes valeurs de gris apposées les unes à côté des autres et de façon graduelle peuvent amener un tremblement dans la perception, notamment au niveau de l'intensité. Une même teinte de gris aura l'air plus intense disposée à côté d'une autre. L'exemple le plus souvent utilisé pour illustrer ce phénomène s'appelle « les bandes de Mach » (figure 50).



Figure 50 : Bandes de Mach (Free Art License 1.3)

De même, la couleur en contexte, telle qu'elle a été présentée dans ce chapitre sous le titre « La couleur » fait également partie des illusions

d'optique, puisque dépendant des autres couleurs présentes dans son environnement, une couleur peut sembler plus ou moins vive. Rappelez-vous donc que la perception de la lumière réfléctée par les couleurs est différente en fonction du contexte.



Figure 51 : Contraste de saturation, par les auteurs (CC0)

4.5.5 Les illusions de mouvement

Les couleurs et formes, associées ensemble, parfois avec certains effets de contraste, peuvent générer une impression de mouvement. De plus, les mouvements de la tête et de l'œil tendent à renforcer cette même impression de mouvement. Mais encore, plus que la simple forme géométrique et couleur, la brillance (luminance) et l'orientation de ces formes parfois colorées induisent également des effets de mouvement (Pinna & Brelstaff, 2000). En exemple, la figure 52 qui semble bouger si l'on fixe le point noir en s'approchant, s'éloignant de l'image, ou si l'on hoche la tête.

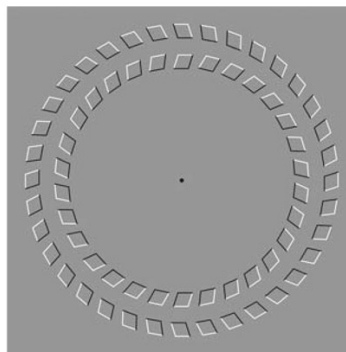


Figure 52 : Anneaux concentriques de Pinna et Brelstaff, 2000 (CC0)

4.5.6 La forme en contexte

Selon le contexte, certaines formes peuvent nous sembler différentes.

L'illusion d'Ebbinghaus

L'œil humain évalue difficilement les aires. Mais une fois celles-ci mises en contexte, c'est encore pire. Ainsi, sur la figure 53, vous expérimentez l'illusion d'Ebbinghaus : les cercles centraux ont la même taille. Vous parieriez pourtant le contraire.

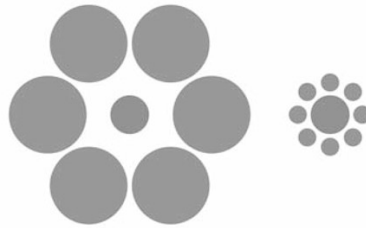


Figure 53 : Illusion d'Ebbinghaus, 1901 (Domaine public)

L'illusion de Poggendorf

Rappelez-vous de la loi de continuité de la *Gestalt Theory* : c'est exactement la même chose qui se produit dans la figure 54. Parfois, nous réalisons de « mauvaises » continuités, car, d'un point de vue géométrique, c'est bien la ligne inférieure qui est dans le prolongement de la droite, et non la ligne supérieure.

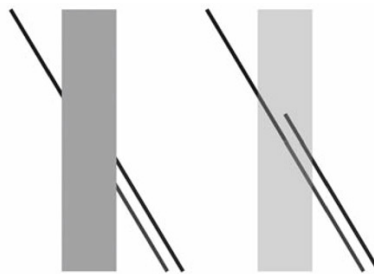


Figure 54 : Illusion de Poggendorf, par Fibonacci, 2007
(CC BY-SA)

5. Conclusion

Ces différentes lois sont à garder en tête lorsque l'on conçoit un graphique quelconque. Les différents principes de la *Gestalt Theory*

ainsi que les illusions d'optique peuvent influencer les choix de conception de visualisation d'information. Il s'agit de concepts de base en termes de perception visuelle qui, dès que l'on aborde la conception de visualisation d'information, deviennent des principes de design fondamentaux. Ces principes de design sont abordés dans le chapitre 4. Ils y sont abordés du point de vue du concepteur de visualisations d'information.

Résumé

Pour rappel, la vision est l'interprétation cognitive du sens de la vue. Elle fait référence aux processus cognitifs et physiologiques par lesquels la lumière émise par l'environnement détermine les détails des représentations sensorielles. Sans lumière, nous ne pouvons tout simplement pas voir ! C'est d'ailleurs la différence d'intensités lumineuses, transmises par les longueurs d'ondes présentes autour de nous, qui nous permet de percevoir les couleurs.

Le traitement préattentif est l'accumulation subconsciente d'informations provenant de l'environnement qui nous entoure : le cerveau « choisit » ce qu'il considère comme important ou non. En visualisation d'information, certains traits ont des caractéristiques plus préattentives que d'autres, c'est-à-dire qu'un élément sera saillant visuellement si on lui attribue cette caractéristique (par exemple : une couleur ou une orientation différente des autres éléments présents dans l'environnement).

Pour terminer, selon la *Gestalt Theory*, « le tout est différent de ses parties » : cela signifie que nos processus cognitifs nous font voir bien plus qu'un ensemble de traits lorsque nous regardons une image, la perception dépasse la sensation ; elle va jusque l'interprétation. C'est la raison pour laquelle il existe différentes illusions de perception qu'il est utile de connaître dans le domaine de la visualisation d'information.

Ressources externes et bibliographie indicative

Contenu interactif

1. Nous avons plus que 5 sens, par e-penser : <https://www.youtube.com/watch?v=6XQhUJ5fqMk>
2. Événement couleur : comprendre la perception et la relativité de la couleur, par Idiko Tv : <https://www.youtube.com/watch?v=RK2MDkbhqL0>
3. La roue colorimétrique de Galactic.Ink : <https://galactic.ink/sphere/>
4. Le projet Visionary, qui met en place des alternatives pour daltoniens en ce qui concerne la conception d'interfaces : <https://colour-blindness.org/>
5. Vidéo : le test d'attention sélective de Simon et Chabris (1999) : <https://www.youtube.com/watch?v=vJG698U2Mvo>
6. Vidéo : The Door Study, expérience de cécité au changement : <https://www.youtube.com/watch?v=FWSxSQspiQ>

Bibliographie indicative

Benamran, B. (2015), Prenez le temps d'y penser, Tome 1, Hachette Livres, Éditions Marabout, Paris

Crombez, N. (2014), *La Gestalt: Psychologie de la forme*, 50minutes.fr, Business. ISBN 280625695X

Lanthony, P. (2001), *Vision des couleurs et daltonisme*, Ediss

Sève, R. (2009). *Science de la couleur : Aspects physiques et perceptifs*, Challagam Éditions

CHAPITRE 3

LA REPRÉSENTATION DES DONNÉES

1. Types de données et types de variables
2. Les familles de visualisation
3. CATALOGUE

Visualiser les données équivaut à représenter un message. Connaître les types de données, les types de variables et les différentes techniques de visualisation permet de préparer le message à communiquer grâce aux différentes techniques de visualisation. Ces techniques se classent par famille et chaque famille permet de représenter un type de message. Dans ce chapitre, vous découvrirez comment choisir la visualisation adéquate, en rapport avec votre message.

1. Types de données et types de variables

1.1 L'acquisition des données

Avant de visualiser les données, la première étape est bien évidemment d'en faire l'acquisition. L'analyste explore donc les données avant de les visualiser dans un but communicationnel.

1.1.1 Le contexte d'acquisition des données (Kirk, 2016)

Afin de réaliser une analyse sur un sujet qui l'intéresse, l'analyste établit ses besoins en termes d'information. Il établit ainsi les critères auxquels les données doivent correspondre et décide quel type de données ne sera pas pertinent afin de répondre à sa question initiale.

Par exemple, imaginons que l'analyste se demande « quelle a été l'évolution des émissions de CO₂ en Belgique ces dix dernières

années ? ». N'ayant pas produit ces données lui-même, il va mener une recherche afin de trouver des sources pouvant lui apporter cette information. Il trouvera peut-être les données en ligne, téléchargeables sur un site web, ou peut-être ailleurs : faut-il contacter quelqu'un pour les recevoir ? Remplir un formulaire ? Le jeu de données (*data set* en anglais) identifié par l'analyste peut dépasser la question qu'il se pose. Il pourrait, par exemple, contenir les émissions de CO₂ de tous les pays d'Europe et ce, non pas annuellement, mais mensuellement. L'analyste doit donc examiner les données et les transformer (*infra*) pour qu'elles puissent répondre à sa question. Il peut aussi lui-même élargir ses perspectives : comparer les émissions de CO₂ de la Belgique avec celles des pays voisins au cours des 10 dernières années est-il enrichissant et intéressant ? A-t-il le temps et l'espace pour se le permettre ?

En vérité, les données portent assez mal leur nom. Elles ne sont pas tangibles, disponibles « dans la nature » et saisissables pour être analysées. Les données sont construites : les humains les font émerger de différentes activités afin d'analyser les phénomènes. Nous évoluons dans une société où une importance grandissante est accordée aux données et aux nombres (*data-driven society*). Les données donnent une idée objective et aseptisée de la réalité, mais il n'en est rien : les gens produisent les données. Il est donc important de toujours se demander **d'où**, **comment** et **par qui** les données représentées sont-elles acquises. Il y a effectivement différentes méthodes d'acquisition de données d'origine différente.

Les données que l'on produit soi-même :

- La collecte de données primaires : lorsque les données dont l'analyste a besoin n'existent pas ou lorsqu'il a besoin d'avoir le contrôle total sur les données, il récolte les données brutes (*raw data*). Les méthodes sont nombreuses : expérimentations,

questionnaires, relevé manuel de mesures précises, etc. Manuelle ou assistée par ordinateur, la collecte de données primaires reste coûteuse en temps (et parfois en argent), mais garantit le plus grand niveau de contrôle sur les données qui sont en adéquations avec les besoins de l'étude.

- La collecte manuelle et l'enrichissement des données : les données dont l'analyste a besoin existent, mais ont besoin d'être rassemblées et enrichies. Par exemple, cela peut concerner des passages de textes relevés sur du papier et qui n'existent pas au format numérique. L'analyste les encode numériquement afin de les analyser et les enrichit en fonction de ses besoins. Souvent, il s'agit de terminer la collecte de données commencées par d'autres en ajoutant des valeurs manquantes.

Les données produites par d'autres :

- Les données fournies à l'analyste : un partenaire, client ou collègue de l'analyste peut tout simplement lui transmettre un jeu de données prêt à l'emploi.
- Téléchargement web : certains jeux de données sont disponibles sur le web grâce à des plateformes permettant d'extraire le jeu de données souhaité.
- Rapports systèmes : en entreprise, il est commun d'utiliser les données liées à une activité afin d'en produire une analyse.
- Les services de tiers : l'analyste peut recourir à des tiers qui produiront les données pour lui dans le cas où cela est trop compliqué ou coûteux en temps pour lui.
- Les API : les *application programme interface*, disponibles gratuitement sur de nombreuses interfaces, permettent de créer des applications qui accèdent aux traces numériques liées à l'usage d'applications ou encore à la consommation et la

production de contenus sur un site web ou un réseau social numérique.

1.1.2 L'examen des données

Une fois les données acquises, l'analyste examine les propriétés des données : le type des données, la taille du jeu de données ainsi que la qualité des données. La taille du jeu de donnée concerne les attributs et formats des données : quelle est la plus haute valeur ? Quelle est la plus basse ? Combien de valeurs y a-t-il ? Examiner la qualité des données permet d'identifier les problèmes dans le jeu de données : y a-t-il des doublons ? Des erreurs d'encodage ? Des données manquantes ? Des changements d'échelles ? L'ajout ou la suppression de question ? etc. Cette étape mène souvent au nettoyage et à la transformation des données nécessaires à la visualisation (Kirk, 2016). En fonction de cet examen des données et des objectifs poursuivis, l'analyste optera pour une technique appropriée d'analyse des données.

Ce choix repose non seulement sur l'identification des variables et des types de données, mais aussi sur l'analyse statistique des données qui permet de formuler le message à véhiculer.

L'analyse statistique des données

Lors de l'analyse des données, il y a deux possibilités : produire des statistiques descriptives (on décrit seulement les données) ou inférentielles (on tente une prédiction avec un intervalle de confiance).

Statistiques descriptives

- Tendances : médiane, moyenne, mode
- Écart-type, variance, valeurs extrêmes ou aberrantes
- Distribution : maximum, minimum, quartiles, quantiles, médiane

Statistiques inférentielles

- Corrélation : interdépendance (coefficient de corrélation, Pearson, Chi-square)
- Causalité : relation causale entre deux événements (t-test, ANOVA)
- Régression : prédiction des changements dans les variables (régression simple/multiple)
- Tests non-paramétriques quand les autres tests ne s'appliquent pas (Wilcoxon, sign)

Besoin d'un coup de pouce en statistique ? Trouvez de l'aide dans les ressources externes⁶.

L'analyse statistique permet à l'analyste d'observer et d'explorer ses données. En fonction des variables disponibles et du type de données à disposition, il pourra formuler le message à représenter et choisir une technique de visualisation appartenant à une famille de visualisation.

1.2 Les types de variables

Lorsque l'analyste dispose enfin d'un jeu de données exploitable et qu'il mène son analyse, il peut envisager de visualiser les données.

La première étape pour choisir la technique de visualisation la plus adéquate possible à l'ensemble de données est l'analyse du type de variable. S'agit-il d'une variable **discrète** ou **continue** ? Une variable

est discrète lorsqu'elle peut prendre un nombre fini ou dénombrable de valeurs. Plutôt que de valeurs, on parle alors de modalités. Le passage d'une modalité à une autre est « brutal », c'est-à-dire sans continuité. C'est le cas par exemple des catégories (ex. genre d'une personne ou pays de naissance) et des niveaux (ex. niveau scolaire ou échelle de Richter). Une variable est continue lorsqu'elle peut prendre un nombre infini ou indénombrable de valeurs. Le passage entre deux valeurs comporte une infinité de possibilités, il est continu. Les variables continues ne peuvent prendre que des valeurs réelles (ex. $\pi = 3,141\ 59\dots$).

Ainsi, certaines techniques de visualisation sont plus adaptées à certains types de données. Par exemple, si l'on souhaite montrer l'évolution du nombre d'inscrits à l'université entre 2009 et 2019, on peut utiliser un diagramme à bâtons, car cette représentation est plus adéquate pour montrer une variable discrète. Si l'on souhaite montrer l'évolution des températures sur les dix dernières heures, on peut utiliser un histogramme, car cette représentation est plus adéquate pour montrer une variable continue.

Attention ! Il existe des « zones grises » entre variables discrètes et continues, au sein desquelles le designer peut faire des choix méthodologiques, pour autant que ces choix soient correctement justifiés. Considérons l'âge par exemple. L'âge est une durée que l'on peut arrondir à l'année près (ex. 22 ou 43 ans), alors qu'en réalité une personne n'a **exactement** 22 ans qu'un infime instant. Il est d'usage d'exprimer son âge à l'année près. Pourtant une durée peut être exprimée en une infinité de possibilités : minutes, secondes, millisecondes, microsecondes, nanosecondes, picosecondes et plus encore avec l'infiniment petit.

1.3 Les types de données

La deuxième étape pour choisir la technique de visualisation la plus adéquate possible à l'ensemble de données est l'analyse du type de données. S'agit-il de données qualitatives ou quantitatives ? Pour le savoir, une astuce consiste à se demander s'il est possible d'effectuer une opération mathématique simple (ex. soustraction ou division) sur les données. Si la réponse est non, alors il s'agit de données qualitatives. Si la réponse est oui, alors il s'agit de données quantitatives. Attention, les données numériques n'impliquent pas nécessairement des données quantitatives. Par exemple, le code postal, bien qu'exprimé de façon numérique, est une donnée qualitative, car soustraire ou diviser des codes postaux n'a aucun sens.

Typiquement, un ensemble de données ou *data set* en anglais est composé de plusieurs variables, et comprend donc des données qualitatives et quantitatives. C'est le message que le designer souhaite véhiculer à propos du *data set* qui influence le choix de la technique de visualisation. Par exemple, la carte et le diagramme à bâtons peuvent tous deux montrer le nombre d'habitants par pays en Europe. Mais le diagramme à bâtons semble plus adapté pour étudier le classement entre pays (angle arithmétique), la carte plus adaptée pour montrer similitudes et différences entre pays (angle géographique). Il est donc essentiel de comprendre à quel type de données on a à faire !

Comme illustré ci-dessous, il existe de nombreuses classifications des données. Les classifications de Bertin (1983), Mackinlay (1986) et Ware (2004) comportent trois classes de données : qualitatives, ordinales et quantitatives. Seule celle de Stevens (1946) comporte quatre classes de données, facilement mémorisables avec l'acronyme « NOIR » pour nominales, ordinales, intervalles et ratio. Cette classification est la plus précise.

Tableau 2 : Classification des types de données, par auteur
(Stevens, Bertin, Mackinglay et Ware)

Stevens (1946)	Nominal : couleur, pays, nom, genre	Ordinal : niveau d'alerte, expertise, satisfaction	Intervalle : date, distance, température	Ratio : âge, taille, poids, taille
Bertin (1983)	Catégories non- ordonnables : couleur, pays, nom, genre	Ordinal (organisation universelle de catégories) : temps (âge, génération, généalogie), perception (couleur, chaleur, taille), habitudes (grades, hiérarchies)	Métriques (mesurable) : date, taille, poids, température	
Mackinglay (1986)	Catégories	Ordinal	Quantitatif	
Ware (2004)	Qualitatif (catégories) : couleur, pays, nom, genre	Entiers (discret) : niveaux	Réels (continus) : âge, date, distance, taille, poids, température	

1.3.1 NOIR, selon Stevens (1946)

La classification NOIR, selon Stevens, est la plus complète, bien qu'elle soit la plus ancienne des quatre classifications présentées. C'est celle qui vous aidera à déterminer le message à communiquer de la façon la plus précise possible.

Les données nominales sont qualitatives. Ce type de données permet de catégoriser et de distinguer. Par exemple, une pomme peut être verte, rouge ou jaune. Ces données ne sont pas ordonnables entre elles, une catégorie ne vaut pas mieux qu'une autre : les relations entre catégories sont arbitraires.

Les données ordinales sont qualitatives. Elles sont aussi catégorielles et, comme leur nom l'indique, il y a une relation d'ordre entre elles. Exemple : la taille d'un vêtement (S, M, L) ou bien encore les réponses à une enquête, de type « pas du tout d'accord » à « complètement d'accord » basées sur une échelle.

Les données d'intervalle sont quantitatives et assez peu communes. Elles sont numériques et l'intervalle entre deux catégories a toujours la même valeur ; un zéro ne signifie pas qu'il y a absence de phénomène. Une opération arithmétique comme une multiplication ou division est impossible avec des données d'intervalle. L'exemple le plus simple est la température : la valeur zéro (0 °C) est arbitraire et ne signifie pas l'absence de température ! L'heure de départ et d'arrivée d'un transport aérien s'exprime aussi en données d'intervalle.

Les données de ratio sont les données quantitatives les plus souvent rencontrées. Numériques, elles expriment la pleine puissance expressive des nombres réels. Avec les données de ratio, on peut dire que : « L'objet A est deux fois plus grand que l'objet B ». La masse d'un objet ou encore l'argent se définissent sur des échelles de ratio.

Ne pas se laisser troubler par la nature des données

La classification du type de données de Stevens (1946), représentée par l'acronyme NOIR, permet de qualifier la qualité de l'attribut d'une entité dont parle Ware (2004). Il est compréhensible que la nature des données puisse être troublante. Ainsi, les données textuelles sont qualitatives. Elles peuvent être nominales et ordinales. En ce qui concerne les données de type ordinal, elles peuvent être représentées par des données textuelles ou numériques (dans le cadre d'une compétition : Or, Argent, Bronze ou 1, 2, 3). De même, le code postal est une catégorie : il serait étrange de le diviser, bien que sous forme numérique, c'est un type de données catégoriel (et donc nominal). Ainsi, la nature des données importe finalement peu : ce qui compte est bien ce qu'elles expriment en soi. Pour communiquer un message de façon claire et précise, il faut donc choisir une visualisation qui sera en phase avec ce même message, et donc en phase avec le type de variables et le type de données (NOIR) dont l'analyste dispose. Cela est abordé dans le chapitre suivant.

2. Les familles de visualisation

2.1 Le « Zoo de la visualisation » (Heer *et al.*, 2010)

2.1.1 Présentation du zoo

Dans « A tour through the visualization zoo », Heer et ses collègues (2010) présentent les grandes familles des techniques de visualisation : chronologies, cartes, visualisations statistiques, hiérarchies et réseaux. Le titre de l'article repose sur une métaphore entre les familles de techniques de visualisation et le regroupement d'espèces animales dans un zoo (faune Antarctique, félins d'Afrique, insectes, oiseaux, reptiles, etc.).

En général, les techniques de visualisation candidates pour véhiculer un message donné sont multiples. Néanmoins, certains choix peuvent s'avérer incongrus, voire aberrants. Par exemple, choisir le camembert pour montrer des quantités est une aberration, car cette

technique de visualisation n'est adaptée que pour montrer les parties d'un tout. En effet, le camembert montre la contribution relative de catégories à un tout. La contribution relative de chaque catégorie est exprimée en pourcentage, et la somme des contributions est nécessairement égale à 100 %. Ainsi, par exemple, pour montrer les quantités brutes de chaque fruit en stock, on utilisera un diagramme à barres. Pour montrer la contribution relative de chaque fruit au stock, on utilisera un camembert. Dans les deux cas, on triera les valeurs pour faciliter les comparaisons et le balayage visuel.



Figure 55 : Métaphore du « Zoo de la visualisation », par les auteurs (CC0)

Les familles de visualisations identifiées dans le « Zoo de la visualisation » (Heer *et al.*, 2010) permettent donc de choisir une technique de visualisation en adéquation avec le message à communiquer. Après avoir choisi une famille de technique de visualisation à employer (par exemple : vouloir représenter quelque chose qui s'est passé dans le temps et choisir la famille des séries chronologiques), il y a une multitude de techniques de visualisation parmi lesquelles choisir. Si l'objectif est, par exemple, de comparer les données entre elles, certaines techniques conviendront mieux que d'autres. La plupart des techniques se trouve dans le catalogue et sont catégorisées selon le « Zoo de la visualisation » qui est présenté dans le paragraphe suivant (avec quelques ajustements).

2.1.2 La classification de Heer et ses collègues (2010)

Heer et ses pairs (2010) ont catégorisé différentes représentations graphiques en tenant compte de leur famille de visualisation. Ils parlent de ces graphiques à titre d'exemple. Il en existe bien plus, mais ceux-ci permettent de se s'accoutumer aux familles de visualisation de façon intrigante et pratique.

1. Les séries chronologiques

Les séries chronologiques permettent de représenter des phénomènes temporels, et donc des valeurs qui changent au fil du temps. Dans ce type de visualisations, le temps est une variable explicative.

- **Index chart** : avec certaines formes de données chronologiques, les valeurs brutes sont moins importantes que des changements relatifs. Un graphique d'index est un graphique linéaire qui indique le changement en pourcentage pour un ensemble de données temporelles basées sur un point d'indice sélectionné. Par exemple, le graphique de la figure 56 indique la variation en pourcentage du taux de chômage (l'indice) à New Castle et en Nouvelle-Galles-du-Sud (Australie) d'août 1988 jusque décembre 2009 (période de temps).

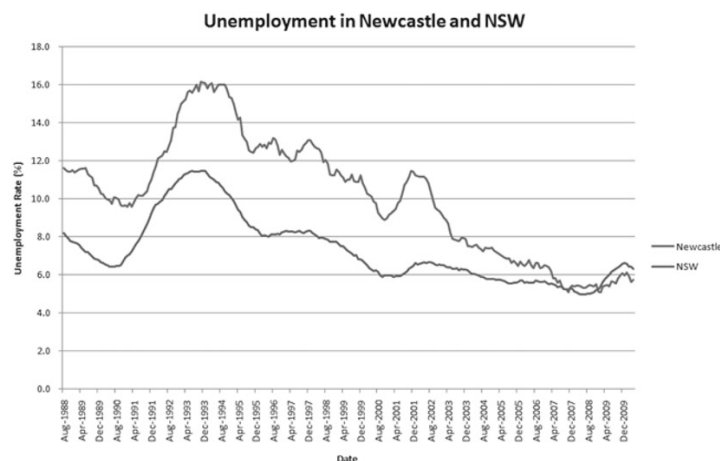


Figure 56 : Taux de chômage à New Castle et en Nouvelle-Galles-du-Sud au cours du temps (Domaine public)

- **Stacked graph** : cette forme empilée de séries chronologiques permet d'obtenir la somme visuelle des valeurs. Ce type de graphique décrit des modèles agrégés et permet d'effectuer des recherches dans un sous-ensemble de séries individuelles. Sur la figure 57, on voit clairement que les hectares consacrés à la culture du Sauvignon blanc ont clairement augmenté au cours du temps, tandis que cela n'a pas ou peu évolué en ce qui concerne la culture du Merlot. Ces graphiques présentent quelques limites, comme le fait de ne pas supporter les chiffres négatifs et d'être incompatible avec les données d'intervalle. En outre, l'empilement peut parfois rendre difficile l'interprétation précise des tendances qui se trouvent au sommet d'autres courbes. L'utilisation de l'interaction (recherche et filtre) peut facilement régler ce problème.

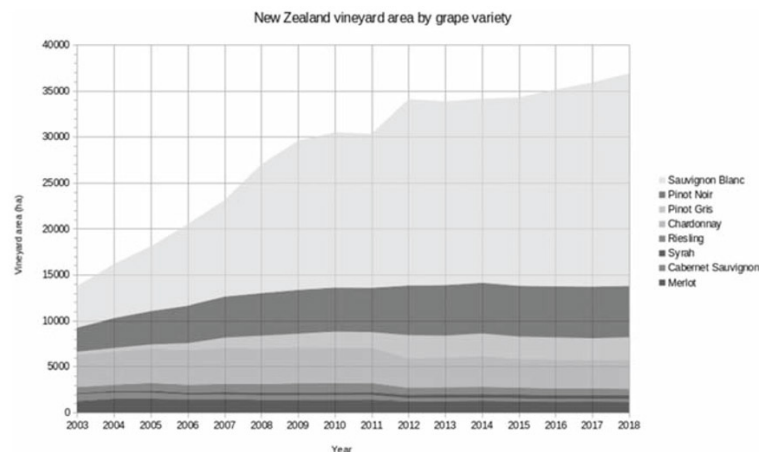


Figure 57 : Évolution de la surface des vignobles en fonction de la variété des raisons au cours du temps, 2019, par Jonathanischoice (CC BY-SA)

- **Small multiples** : parfois, tracer des séries chronologiques dans un même espace peut produire des chevauchements de courbes qui réduisent la lisibilité. Les *small multiples* permettent de contourner cela en montrant chaque série dans son propre

graphique. Ils peuvent être construits avec presque n'importe quel type de visualisation et pas seulement les séries chronologiques. Mais, dans ce cas en particulier, ils se révèlent très utiles en permettant une visualisation plus efficace que de s'obstiner à afficher toutes les données dans un même espace chargé et difficilement lisible.

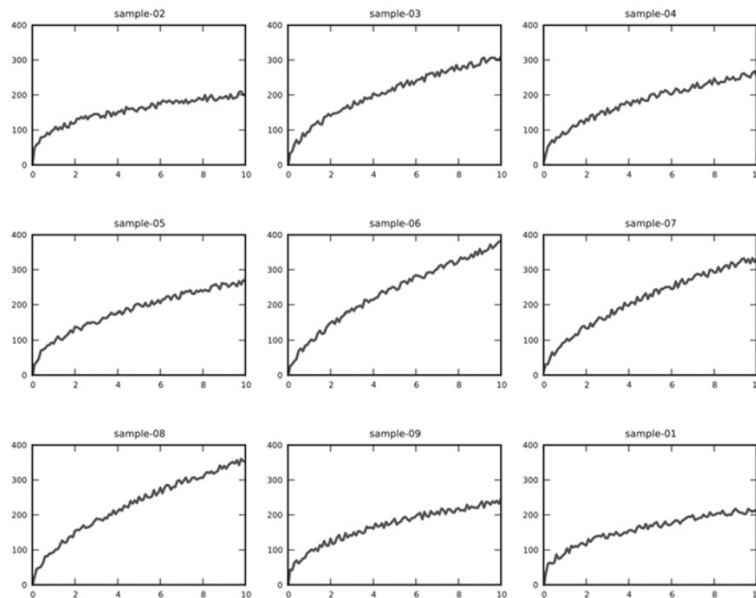


Figure 58 : Exemple de Small Multiples, par Dr Drang (2013)
(CC BY-SA)

- **Horizon graph** : le *Horizon graph* une technique permettant d'augmenter la densité des données temporelles tout en préservant la résolution et en réduisant drastiquement l'espace utilisé. À partir d'un graphique en ligne standard, on colore les valeurs positives en bleues et les négatives en rouge (une gradation est possible si on le souhaite). Puis, on « reflète » les valeurs négatives dans la même zone que les valeurs positives. On double encore ensuite la densité des données en divisant une fois de plus le graphique en bandes et en couches pour créer une forme imbriquée. Le résultat est un graphique qui

préserve la résolution des données, mais n'utilise qu'un quart de l'espace. Le *Horizon graph* prend un temps de compréhension conséquent, mais il est plus efficace qu'un graphique standard lorsque l'espace disponible est très petit : il utilise moins de 17 % de l'espace utilisé généralement par un graphique en ligne traditionnel.

2. Les distributions statistiques

Ces visualisations permettent de mieux comprendre les propriétés statistiques des données. Dans un but exploratoire, cela sert à tester des hypothèses ou à prédire des valeurs futures. Cela sert également à décider des transformations à apporter aux données ainsi que de leur modélisation. Les techniques les plus connues sont l'histogramme et le *box plot*, mais d'autres techniques permettent de visualiser des distributions statistiques.

- ***Stem-and-leaf plot*** : le *stem-and-leaf plot* permet d'évaluer une collection de nombres. Il regroupe généralement les nombres en fonction du premier chiffre significatif (*stem* ou tige), puis empile les valeurs dans chaque regroupement par le deuxième chiffre significatif (*leaf* ou feuille). Les données représentent elles-mêmes les distributions de fréquences (correspondant aux barres d'un histogramme). Imaginons que vous deviez représenter la distribution des pulsations cardiaques de 10 personnes venant de terminer un effort physique. La tige peut représenter des dizaines, des centaines, etc., à gauche de l'axe vertical. Les unités sont représentées sur les feuilles, à droite de l'axe vertical. Cette représentation permet d'avoir une idée claire des nombres faisant partie de la distribution.

Pulsations cardiaques des participants : 78, 79, 83, 91, 95, 98, 147, 163, 164, 201

7		8, 9
8		3
9		1, 5, 8
14		7
16		3, 4
20		1

Figure 59 : Exemple de stem-and-leaf plot, par les auteurs

- **Q- Q Plot** : le *Q-Q Plot* (ou diagramme quantile-quantile) compare deux distributions en représentant graphiquement les quantiles de la première distribution par rapport à ceux de la deuxième. Si les deux sont similaires, les valeurs tracées se situeront approximativement le long de la ligne centrale diagonale. Si les deux sont liées de manière linéaire, les valeurs se situeront à nouveau le long d'une ligne, mais avec une pente et une interception variable. On peut ainsi comparer des distributions. Sur la figure 60, on voit qu'elles sont relativement semblables.

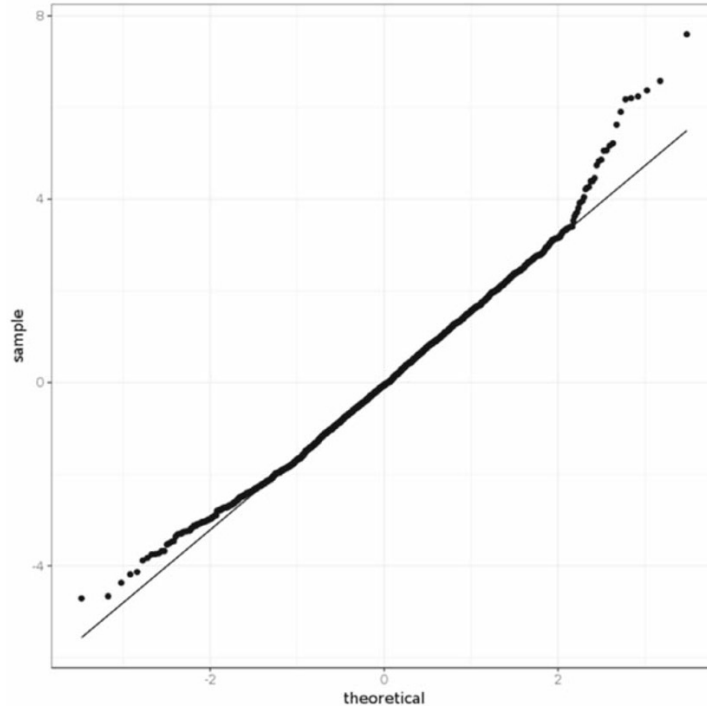


Figure 60 : Exemple de Q-Q Plot comparant les distributions de deux relevés de temps de lecture, par Groceryheist, 2018 (CC BY-SA)

- ***SPLOM (Scatter Plot Matrix)*** : le *SPLOM* représente les relations entre de multiples variables. Il arrive souvent que les données soient multivariées et donc difficiles à représenter, en partie à cause de la difficulté d’imaginer mentalement des données en plus de trois dimensions. Utiliser plusieurs petits diagrammes de dispersion montrant un ensemble de relations par paires permet de pallier le problème. Un *SPLOM* permet ainsi l’inspection visuelle de plusieurs corrélations. La figure 61 montre le *SPLOM* d’un jeu de données concernant les variétés de l’Iris. Le *data set* contient un ensemble de 150 enregistrements sous cinq attributs : longueur de sépale, largeur de sépale, longueur de pétale, largeur de pétale et espèce.

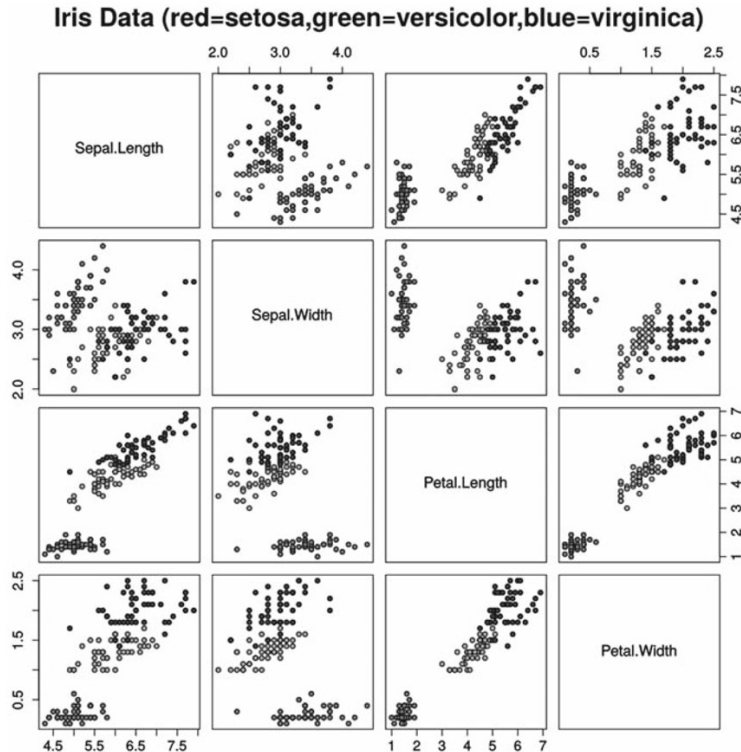


Figure 61 : The scatterplot of Iris flower data set, collected by Edgar Anderson and popularized in the Machine learning community by Ronald Fisher (CC BY-SA)

- **Coordonnées parallèles** : dans les coordonnées parallèles, les données sont tracées de façon répétitive sur des axes parallèles, tandis que les points correspondants sont reliés avec des lignes. Chaque ligne (ou polyligne) représente une seule ligne dans la base de données. Les croisements de lignes entre les colonnes parallèles indiquent souvent une corrélation inverse. Sur la figure 62, chaque ligne correspond à une voiture. Les couleurs permettent de les répertorier par marque. Pour 8 attributs, les similitudes et différences des voitures sont perceptibles sur cette représentation.

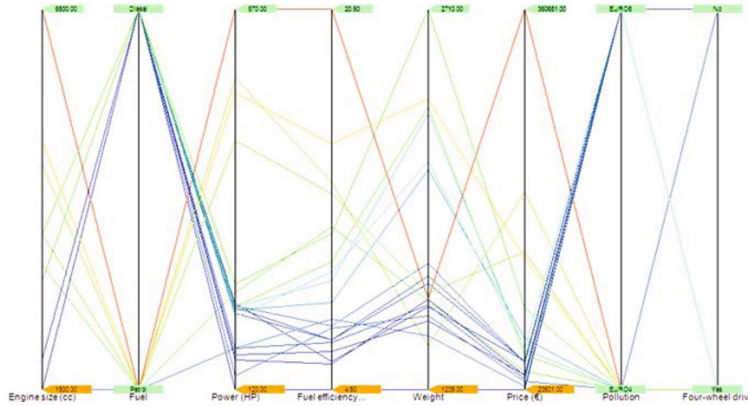


Figure 62 : Coordonnées parallèles décrivant certaines caractéristiques d'un ensemble de voitures. (Domaine public)

3. Cartes géographiques

Les cartes représentent des données géographiques, de localisation. Elles peuvent se charger d'information supplémentaire ou encore être déformées afin de raconter une histoire.

- **Flow map** : en traçant des lignes ou flèches sur une carte géographique, une carte de flux peut représenter le mouvement d'une quantité dans l'espace et (implicitement) dans le temps. Les points de cheminement, la direction, l'épaisseur des lignes et la couleur peuvent tous être utilisés pour présenter les dimensions de l'information. Comme toute visualisation, la *Flow Map* peut bénéficier de l'interaction ou d'une présentation dynamique. À titre d'exemple, cette vidéo⁷ représente l'exportation internationale d'armes par les États-Unis de 1950 à 2017 grâce à une *Flow Map*.



<https://vimeo.com/279923192>

- **Choropleth map** : une approche standard pour communiquer ces données consiste à utiliser un codage couleur de la zone

géographique (États, provinces, etc.) afin de représenter des données agrégées par zones. On appelle cela une carte choroplèthe (*choropleth map*). Dans la figure 63, on peut constater que la luminosité augmente graduellement, tout comme la densité de population dans les comtés concernés.

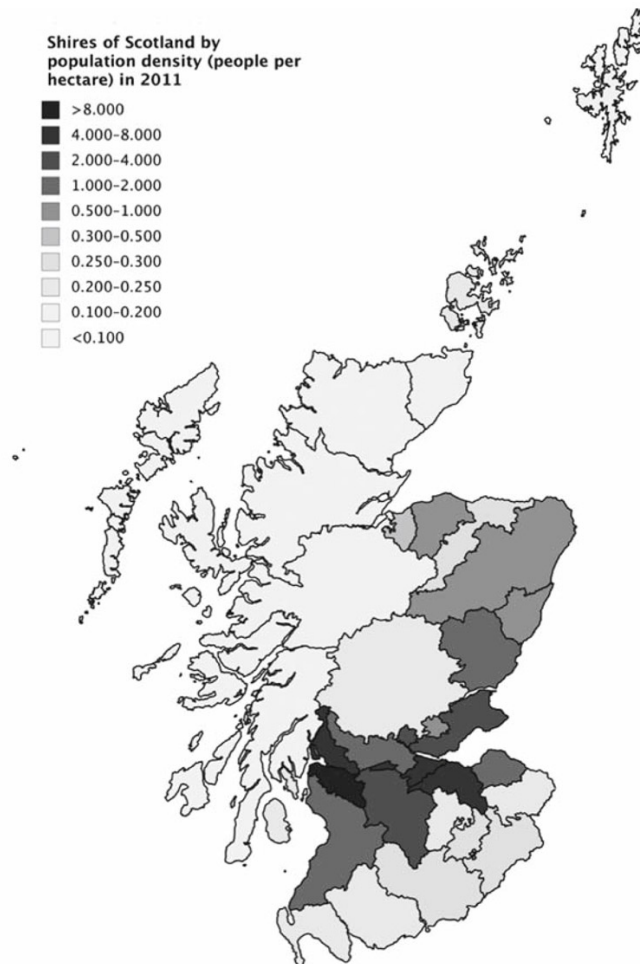


Figure 63 : Comtés d'Irlande par densité de population (personnes par hectares) en 2011 (CC BY-SA)

- **Graduated symbol map** : la *graduated symbol map* permet de placer différents symboles sur une carte sous-jacente. Cette approche évite de confondre la zone géographique avec les valeurs des données et permet de visualiser plus de dimensions (par exemple, la taille, la forme et la couleur des symboles). En

plus des formes simples telles que les cercles, les cartes à symboles gradués peuvent utiliser de simples graphiques comme les diagrammes circulaires.

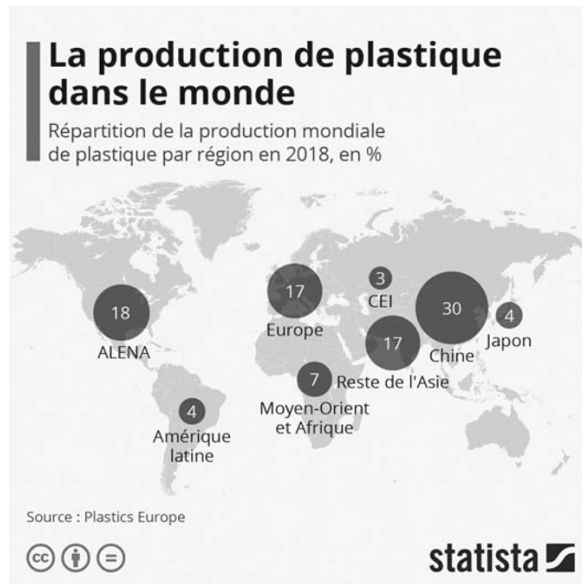


Figure 64 : La production de plastique dans le monde, par Tristan Gaudiaut, 2020 (CC0)

- **Cartograms** : un cartogramme déforme les régions géographiques de sorte que la région encode directement une variable de données. Un exemple courant consiste à redessiner chaque pays du monde en le dimensionnant proportionnellement à sa population ou à son produit intérieur brut.



Figure 65 : An alternative way of visualising the world's population, using a variant on the cartogram. Max Roser/Our World in Data (CC BY)

4. Hiérarchies et réseaux

Les réseaux montrent des relations entre les données (par exemple : qui est ami avec qui, quels sont les amis en commun). Les entités sont représentées par des nœuds (points) et les relations par des liens (arcs). Les hiérarchies montrent également des relations, avec une idée d'ordre entre ces relations. On parle alors d'un nœud racine et de nœuds parents, comme dans un organigramme.

Dans cette catégorie, Heer et ses pairs (2010) donnent différents exemples : le diagramme nœud-lien, les *adjacency diagrams*, les *enclosure diagrams*, le *force-directed layout*, l'*arc diagram* et la *matrix view*. Ci-dessous n'apparaissent que le diagramme nœud-lien, l'*arc diagram* et la *matrix view*. La raison est simple : toutes ces visualisations représentent des graphes. Un graphe est un calcul mathématique modélisant la relation entre des nœuds et des liens. Il y a de multitudes façons de dessiner un graphe, surtout lorsqu'il y a beaucoup de données. De nombreux algorithmes différents existent et permettent de disposer ces données dans l'espace. C'est ce qu'on appelle les *layouts*. Ils sont répertoriés dans le catalogue.

- **Diagramme nœud-lien** : les diagrammes nœud-lien sont les formes les plus simples qui modélisent les réseaux, en mettant en relation des nœuds et des liens, comme son nom l'indique.

Lorsque le réseau a une racine et qu'il est orienté, c'est une hiérarchie. L'organigramme, par exemple, est diagramme nœud-lien qui représente un arbre (*tree*). La figure 66 est un diagramme nœud-lien qui montre la visualisation du réseau LinkedIn de Florent Darrault. Il est central dans le réseau et on voit les différentes relations entre les nœuds (personnes) faisant partie de son réseau. Les couleurs servent à identifier différents clusters.

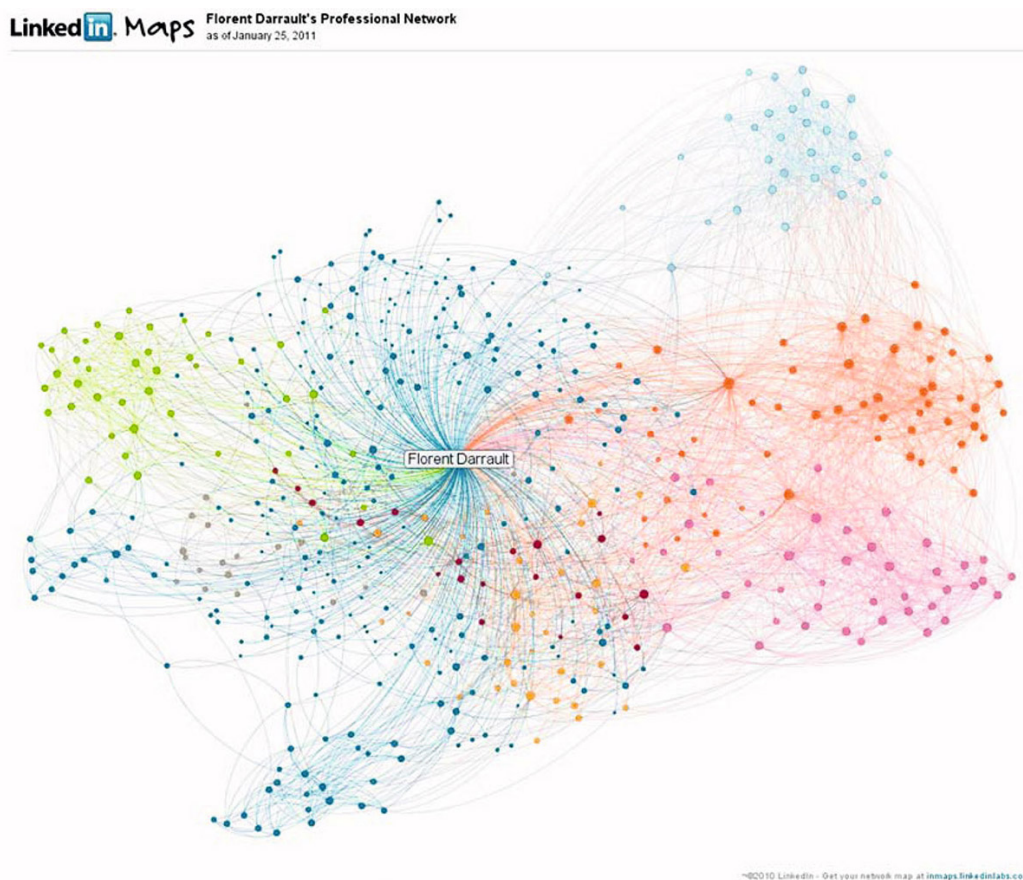


Figure 66 : InMaps pour LinkedIn : visualiser son réseau par Florent Darrault, 2011 (CC BY-SA)

- **Arc Diagram** : un diagramme en arc ou *arc diagram* utilise une disposition unidimensionnelle des nœuds, avec des arcs circulaires pour représenter les liens. Il s'agit à nouveau d'un

layout. Bien que cette représentation ne puisse transmettre la structure globale du graphique aussi efficacement que d'autres *layouts* en deux dimensions, avec un bon ordre des nœuds, il est facile d'identifier les ensembles et les ponts. La figure 67 montre un exemple de diagramme en arc : on voit quels points sont reliés entre eux. La disposition des points pourrait être différente d'un point de vue visuel.

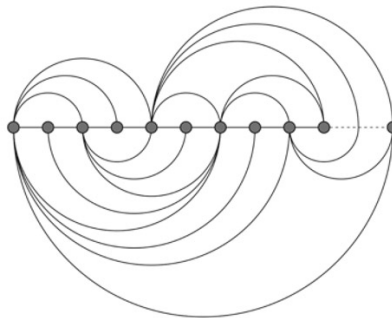


Figure 67 : Exemple de diagramme en arc, par David Eppstein, 2013 (CC0)

- **Matrix View** : la *matrix view* est souvent utilisée par les mathématiciens et les informaticiens. Chaque valeur de la ligne i et de la colonne j de la matrice correspond au lien existant entre le nœud i et le nœud j . Compte tenu de cette représentation, une visualisation naît alors et il suffit de montrer la matrice. L'utilisation de la couleur ou de la saturation au lieu du texte permet de percevoir plus rapidement les valeurs associées aux liens. Le problème de l'ordre des nœuds s'applique tout autant à la *matrix view* qu'au diagramme en arc, de sorte que l'ordre des lignes et des colonnes est important : c'est un algorithme de détection des communautés qui crée l'affichage.

2.2 La classification utilisée dans cet ouvrage

La classification de Heer et ses collègues (2010) n'est pas exhaustive. Pour retrouver un nombre plus étendu de visualisations classées par famille et pour une description plus complète des visualisations présentées dans ce chapitre, rendez-vous dans le catalogue.

La classification utilisée dans cet ouvrage – et qui par ailleurs structure aussi le catalogue – s'inspire du « Zoo de la visualisation ». Elle se structure comme suit :

- **Les visualisations statistiques** : cette catégorie ne se concentre pas uniquement sur les distributions statistiques et inclut les tableaux.
- **Les séries chronologiques** : tout comme dans le « Zoo », y sont répertoriées les visualisations dont le temps est une variable explicative.
- **Les cartes** : ne concerne pas seulement les cartes géographiques, mais la localisation de manière plus générale.
- **Les hiérarchies et réseaux** : concerne les mêmes visualisations que dans le « Zoo ».

3. CATALOGUE⁸

Les possibilités de représentation graphique sont nombreuses. Dans ce catalogue, découvrez ou redécouvrez-les. Chaque visualisation a une fonction précise et correspond à un nombre de variables, à un type de variables et à un type de données. Ces éléments vous sont renseignés pour chaque visualisation du catalogue.

De façon systématique, dans le catalogue, chaque visualisation est détaillée dans une fiche présentant ses caractéristiques, ses points forts et ses points faibles.

Visualisations de données détaillées dans le catalogue

Les visualisations statistiques
Visualisations temporelles
Cartes
Hiérarchies et réseaux

3.1 Les visualisations statistiques

Le diagramme à barres

Nom français	Diagramme à barres
Nom anglais	<i>Bar chart</i>
Fonction	Distribution statistique
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	intervalles ratio

Animaux trouvés morts dans les parcs de New-York (2004-2014)

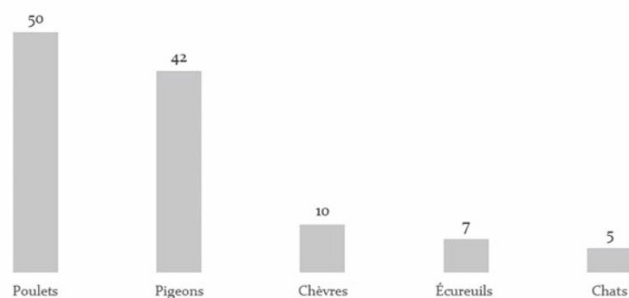


Figure 68 : Diagramme à barres, par les auteurs

Analyse :

Le diagramme à barres, ou diagramme en bâtons, permet de montrer la distribution d'une variable qualitative discrète. Typiquement, les modalités (catégories ou niveaux) sont placées en abscisse et forment

des barres équidistantes de largeur égale, et les fréquences (valeurs numériques) en ordonnée. Seule la hauteur des barres varie, représentant ainsi la valeur numérique associée à chaque catégorie (Archambault *et al.*, 2015). Il est également possible de regrouper plusieurs barres en différents groupes pour montrer les valeurs de plusieurs variables (1 à N variables). Enfin, le diagramme à barres existe aussi sous forme de barres empilées. S'y ajoute alors la fonction « partie d'un tout ».

Le diagramme à barres permet de comparer des catégories discrètes de données numériques (Kirk, 2016). L'exemple montre la distribution des animaux trouvés morts dans les parcs de New York entre 2004 et 2014. On remarque qu'on a trouvé dix fois plus de poulets (50) que de chats (5).

Forces :

- Représentation graphique très accessible et intuitivement compréhensible (Cleveland & McGill, 1984).
- Différences entre catégories rapidement perceptibles par l'œil humain.

Faiblesse :

- Représentation graphique aisément confondue avec l'histogramme (qui montre la distribution de variables continues).

Bibliographie indicative :

Archambault, S. G., Helouvry, J., Strohl, B., & Williams, G. (2015). Data visualization as a communication tool. *Library Hi Tech News*, 32(2), 19. <https://doi.org/10.1108/LHTN-10-2014-0098>

Cleveland, W. S., & McGill, R. (1984), "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical

Methods”, *Journal of the American Statistical Association*, 79(387), 531-554. <https://doi.org/10.1080/01621459.1984.10478080>Kirk, Andy (2016). *Data Visualization : A handbook for data driven design* (Sage). Thousands Oaks.

L’histogramme

Nom français	Histogramme
Nom anglais	<i>Histogram</i>
Fonction	Distribution statistique
Nombre de variables	1
Type(s) de variables	continue
Type(s) de données	intervalles ratio

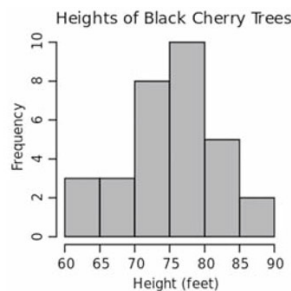


Figure 69 : Black cherry tree histogram (CC BY-SA)

Analyse :

L’histogramme est « un graphique formé de rectangles juxtaposés et tels que la surface de chaque rectangle et la surface totale des rectangles aient une signification » (Bertin, 1967, p. 210). L’histogramme permet de représenter la distribution de la variable. Il sert à montrer les valeurs des données en classes ou groupes. Par groupe, un rectangle est réalisé et correspond à la valeur du groupe. En somme, il s’agit d’un graphique en bâtonnets limité par des

contraintes. La somme des barres, des surfaces, doit être égale à 100 % (Vanderschick & Wautelet, 2003).

Ce qui différencie un histogramme d'un digramme en bâtons (bar chart) est la mesure de la fréquence, c'est-à-dire la proportion d'une valeur dans un groupe. Dans un histogramme, cette proportion est mesurée par la surface des colonnes. Dans un bar chart, elle est mesurée par la hauteur des barres.

Forces :

- Parfait pour estimer quelles sont les valeurs extrêmes et s'il y a des écarts ou des valeurs inhabituelles.
- Donne une bonne vue d'ensemble de la distribution des probabilités.
- Pas besoin de compétences statistiques poussées pour le comprendre.

Faiblesse :

- Peut être confondu avec le bar chart.

Bibliographie indicative :

Bertin, J. (1967). « Sémiologie graphique », Paris, Mouton.

Vandeschrick, C., & Wautelet, J-M. (2003), « De la statistique descriptive aux mesures des inégalités », Academia/Bruylant/L'Harmattan : Louvain-la-Neuve/Paris, <http://hdl.handle.net/2078.1/79654>

Le diagramme à barres empilées

Nom français	Graphique à barres empilées
Nom anglais	<i>Stacked bar chart</i>
Fonction	Distribution statistique
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	ordinales intervalles ratio

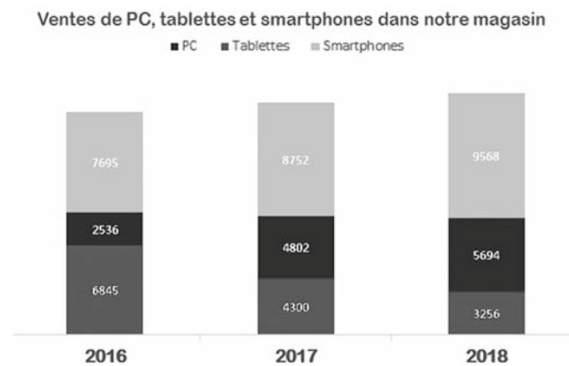


Figure 70 : Diagramme à barres empilées, par les auteurs

Analyse :

Le graphique à barres empilées permet de représenter une variable en catégories et sous-catégories, afin les comparer entre elles. Les catégories sont représentées par des barres horizontales ou verticales qui sont divisées en sous-catégories, chaque barre montrant la contribution relative des sous-catégories à la catégorie entière (Archambault *et al.*, 2015). En somme, il est utilisé pour montrer comment une catégorie est divisée en sous-catégories et quelle est la contribution de chaque sous-catégorie à la catégorie.

Selon Kirk (2016), ces graphiques peuvent soit :

- Superposer les groupes en fonction de leur valeur. La valeur totale de la barre est la somme de tous les groupes.

- Montrer le pourcentage de chaque groupe par rapport à la barre qui vaut 100 %. La valeur d'un groupe est donc un pourcentage relatif au total de la barre. Dans ce cas, il est facile d'identifier les parties (groupes) d'un tout (barres) dans ces graphiques.

Force :

- Peut être réalisé de façon horizontale ou verticale.

Faiblesses :

- Devient plus difficile à lire lorsque le nombre de groupes au sein d'une barre augmente.
- Difficile de comparer les sous-catégories entre elles, car elles ne sont pas alignées.
- Parfois confondu avec le diagramme de Gantt.

Bibliographie indicative :

Archambault, S. G., Helouvry, J., Strohl, B., & Williams, G. (2015). Data visualization as a communication tool. *Library Hi Tech News*, 32(2), 19. <https://doi.org/10.1108/LHTN-10-2014-0098>

Kirk, A. (2016). *Data visualisation: a handbook for data driven*

Le diagramme circulaire

Nom français	Diagramme circulaire (camembert)
Nom anglais	<i>Pie chart</i>
Fonction	Distribution statistique
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	intervalles ratio

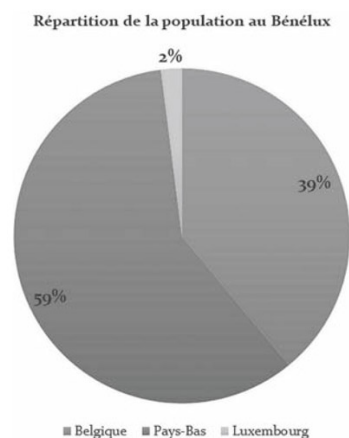


Figure 71 : Diagramme circulaire, par les auteurs

Analyse :

Le diagramme circulaire, aussi appelé diagramme à secteurs ou camembert, est un graphique statistique permettant de montrer les parties d'un tout (Kirk, 2016). Chaque partie, secteur ou part de camembert représente une valeur numérique qui correspond à une proportion, à un pourcentage. La surface, l'angle central et la longueur de l'arc de chaque partie sont proportionnels à la proportion représentée, et la somme des valeurs numériques représentées est toujours égale à 100 % (Easton et McColl, 1997).

Forces :

- Représentation graphique aisément accessible.
- Véhicule rapidement une vue d'ensemble des données.

Faiblesses :

- Illisible au-delà de cinq catégories.
- Totalement inefficace pour supporter la comparaison quand les parties sont de taille similaire (ex. 32, 33 et 35 %).
- Version 3D souvent utilisée pour des raisons esthétiques au détriment de l'intelligibilité des données, l'effet 3D augmentant

la charge cognitive et agissant ainsi comme un distracteur (Rangecroft, 2003).

Bibliographie indicative :

Easton J.V, McColl H.J. (1997), *Statistics Glossary*, [THML en ligne] (consulté le 22/02/17), <http://www.stats.gla.ac.uk/steps/glossary>

Kirk, Andy (2016.). *Data Visualization : A handbook for data driven design* (Sage). Thousands Oaks.

Rangecroft, M. (2003). As easy as pie. *Behaviour & Information Technology*, 22(6), 421-426. <https://doi.org/10.1080/01449290310001615437>

Le diagramme en anneau

Nom français	Diagramme en anneau
Nom anglais	<i>Donut Chart</i>
Fonction	Distribution statistique
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	intervalles ratio



Figure 72 : Diagramme en anneau, par les auteurs

Analyse :

Très similaire au diagramme circulaire, le diagramme en anneau part sur le même principe. L'ensemble des segments constitue un total de 100 % et représente différentes catégories que l'on peut comparer les unes par rapport aux autres. Il est creux au centre, laissant ainsi un espace vide pouvant être utilisé pour y insérer le titre ou un texte court. Il n'y a donc pas d'angles dans ce diagramme (Evergreen, 2019).

Forces :

- Comparaison aisée.
- Apprécié pour ses qualités esthétiques.
- Permet un gain de place en insérant du texte en son centre.

Faiblesses :

- Plus difficile à lire que le diagramme circulaire étant donné l'absence d'angles : la perception de taille entre les segments est moins aisée (Skau & Kosara, 2016).
- Illisible lorsqu'il y a trop de segments.

Bibliographie indicative :

Evergreen, S. D. H. (2019). *Effective Data Visualization: The Right Chart for the Right Data*. SAGE Publications.

Skau, D., & Kosara, R. (2016). Arcs, Angles, or Areas : Individual Data Encodings in Pie and Donut Charts. *Computer Graphics Forum*, 35 (3), 121 -130. <https://doi.org/10.1111/cgf.12888>

Les cercles imbriqués

Nom français	Cercles imbriqués
Nom anglais	<i>Nested/concentric circles</i>
Fonction	Distribution statistique
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	intervalles ratio

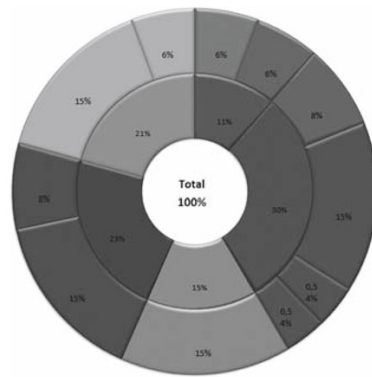


Figure 73 : Graphique en cercles imbriqués

Analyse :

La visualisation en cercles imbriqués apporte plus de précisions sur un diagramme circulaire (camembert). Elle fournit une vue d'ensemble notamment pour les données hiérarchiques. Elle permet ainsi de voir les regroupements et relations structurelles d'un tout (Weixin & *et al.*, 2006). Chaque partie extérieure représente une proportion des cercles plus centraux. La somme des valeurs numériques représentées est égale au tout central, soit à 100 % (Easton et McColl, 1997).

Forces :

- Permet d'avoir rapidement une vue d'ensemble des données.
- Fait ressortir les grandes hiérarchies ou composition d'un tout.

Faiblesses :

- Illisible au-delà de deux cercles et cinq catégories.
- Devient inutile lorsque les valeurs sont trop proches les unes des autres.

Bibliographie indicative :

Weixin, W., Hui, W., Guozhong, D., Hongan, W. (2006) Visualization of Large Hierarchical Data by Circle, CHI 2006 Proceedings, Visualization and Search Packing, Montréal, Canada. 517-520

Easton J.V, McColl H.J. (1997), *Statistics Glossary*, [HTML en ligne] (consulté le 10/04/20), <http://www.stats.gla.ac.uk/steps/glossary>

Le diagramme de Nightingale

Nom français	Diagramme de Nightingale
Nom anglais	<i>Nightingale rose chart</i>
Fonction	Distribution statistique
Nombre de variables	1
Type(s) de variables	discrète continue
Type(s) de données	ordinales intervalles ratio

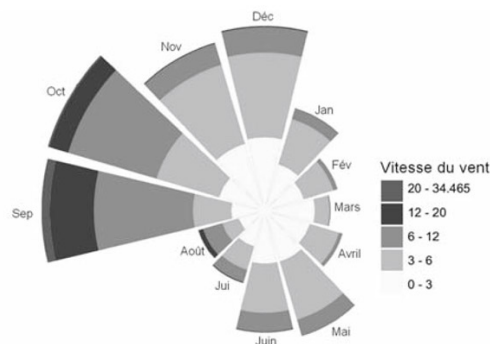


Figure 74 : Diagramme de Nightingale Rose représentant la vitesse du vent par mois en nœuds, par les auteurs

Analyse :

Inventé initialement par la réformatrice médicale Florence Nightingale pour communiquer les décès évitables de soldats pendant la guerre de Crimée, ce graphique est réalisé sur des coordonnées polaires (Brasseur, 2005). Chaque catégorie ou intervalle des données est divisé en segments égaux sur ce graphique radial. La valeur est représentée par la distance entre chaque segment et le centre du diagramme. Chaque anneau à partir du centre de la grille peut être utilisé comme échelle pour tracer la taille du segment et représenter une valeur supérieure. La valeur est représentée de manière zonale et non par le rayon des segments (Wells, 2000).

Forces :

- Permet de montrer un grand niveau de complexité dans les données.

Faiblesses :

- Segments extérieurs davantage mis en valeur en raison de leur plus grande taille de surface
- Distorsion visuelle importante qui peut engendrer des problèmes de compréhension.
- À manipuler avec une grande prudence au vu de la distorsion visuelle qui peut être source d'erreurs d'interprétation.
- Préférable d'utiliser cette représentation graphique avec un public averti.

Bibliographie indicative :

Brasseur, L. (2005). Florence Nightingale's Visual Rhetoric in the Rose Diagrams. *Technical Communication Quarterly*, 14(2), 161–182.

Wells N. A. (2000). Are There Better Alternatives to Standard Rose Diagrams?, *Journal of Sedimentary Research*, 70:1, 37-46

Le nuage de points

Nom français	Nuage de points
Nom anglais	<i>Scatter plot</i>
Fonction	Corrélation
Nombre de variables	2
Type(s) de variables	discrète continue
Type(s) de données	ordinales intervalles et ratio

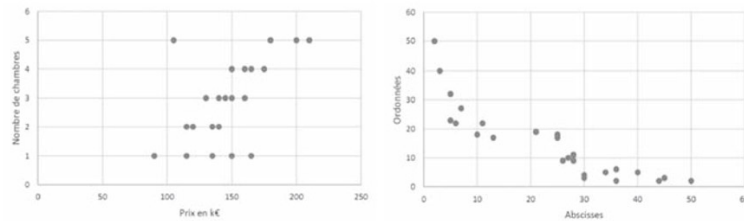


Figure 75 : Nuages de points, par les auteurs

Analyse :

Un nuage de points est un diagramme mathématique utilisant des coordonnées cartésiennes pour afficher les valeurs de deux variables. Les valeurs sont affichées sous la forme d'un ensemble discret de points (le nuage), chaque point ayant la valeur d'une variable déterminant la position sur l'axe horizontal et la valeur de l'autre variable déterminant la position sur l'axe vertical (Utts, 2014). Typiquement, la variable indépendante est affichée sur l'axe des abscisses, la variable dépendante sur l'axe des ordonnées. Dans le cas de deux variables indépendantes, les variables peuvent être affichées sur n'importe quel axe. Attention : bien que le nuage soit un ensemble discret de points, cette représentation graphique permet

d'afficher tant des variables discrètes que des variables continues. L'exemple du haut, inspiré de (Spence, 2014, p. 45), représente le prix de maisons en fonction du nombre de chambres et met en évidence les points suivants : plus il y a de chambres, plus la maison est chère (tendance globale) ; il existe des maisons à deux, trois ou quatre chambres moins chères que certaines maisons à une chambre (compromis local) ; une maison à cinq chambres ne coûte que 105 k€ (donnée aberrante). L'exemple du bas montre une corrélation négative.

Forces :

- Permet d'analyser rapidement et facilement les tendances, le classement et les intervalles de données, les corrélations (positive, négative, nulle, linéaire, exponentielle ou en U).

Faiblesse : N/A.

Bibliographie indicative :

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1), 17-21.

Spence, R. (2014). *Information Visualization-An Introduction* 3rd Edition, Springer.

Utts, J. M. (2014). *Seeing through statistics*. Cengage Learning.

La matrice de nuages de points

Nom français	Matrice de nuages de points
Nom anglais	<i>Scatter plot matrix (SPLOM)</i>
Fonction	Corrélation
Nombre de variables	N
Type(s) de variables	discrète continue
Type(s) de données	ordinales intervalles ratio

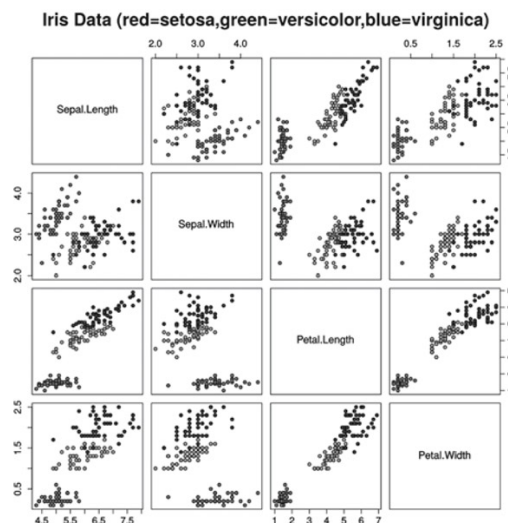


Figure 76 : The scatterplot of Iris flower data set, collected by Edgar Anderson and popularized in the Machine learning community by Ronald Fisher

Analyse :

La matrice de nuages de points, connue sous l'acronyme de *SPLOM* pour *Scatter Plot Matrix* en anglais, inventée par John Hartigan en 1975 (Plotly, 2018), est un outil graphique qui utilise plusieurs nuages de points pour déterminer la corrélation entre une série de variables. Il s'agit d'une grille (ou une matrice) de nuages de points utilisée pour visualiser les relations bivariées entre les combinaisons de variables.

Forces :

- Explorer de nombreuses relations dans un même graphique (Heer *et al.*, 2010, p. 62).
- Permet une visualisation des corrélations entre n'importe quelle paire de variables (Heer *et al.*, 2010, p. 62).
- Possibilité d'utiliser des techniques d'interaction telles que le *brushing* et le *linking* pour explorer les données (Heer *et al.*, 2010, p. 62).

Faiblesse : N/A.**Bibliographie indicative :**

Heer, J., Bostock, M. & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communication of the ACM*, 53, 59-67. Doi:10.1145/1743546.1743567

Plotly. (2018). *What is a SPLOM chart ? Making scatterplot matrices in Python*. Retrieved from <https://medium.com/plotly/what-is-a-splom-chart-make-scatterplot-matrices-in-python-8dc4998921c3>

Le graphique à bulles

Nom français	Graphique à bulles
Nom anglais	<i>Bubble chart</i>
Fonction	Corrélation
Nombre de variables	3
Type(s) de variables	discrète continue
Type(s) de données	ordinales intervalles ratio

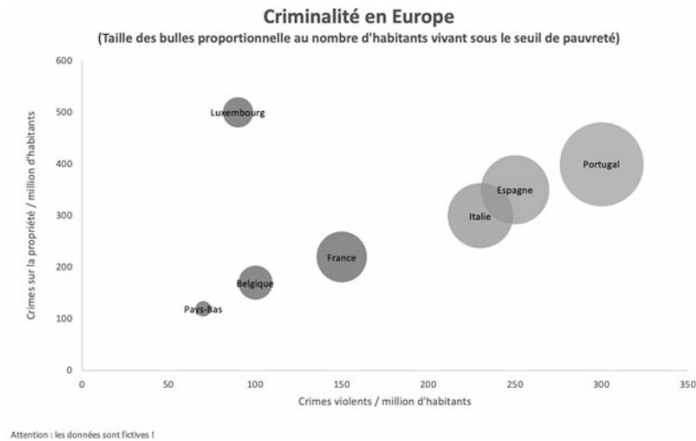


Figure 77 : Exemple de graphique à bulles, par les auteurs

Analyse :

Le *bubble chart* permet de comparer rapidement des proportions et des concentrations entre différents éléments, sur base de la taille de leur représentation visuelle et de leur position dans un repère cartésien (Hardin *et al.*, 2012). Les éléments sont représentés par des cercles dont la surface est proportionnelle à la valeur que l'on veut représenter (Viégas *et al.*, 2007). Les cercles sont ensuite positionnés sur les axes X et Y. Le *bubble chart* permet ainsi de représenter trois variables (Bock, s.d.) : les variables X et Y (grâce à l'emplacement du cercle sur le graphe), et une troisième variable grâce à la taille du cercle. Le *bubble chart* est à mi-chemin entre le *scatter plot* et le *proportional area chart* (The Data Visualisation catalogue, s.d.).

Dans l'exemple ci-dessus, les bulles permettent de visualiser les relations entre 3 variables, pour chacun des pays européens représentés : les crimes violents par million d'habitants (X), les crimes sur la propriété par million d'habitants (Y) et le nombre d'habitants vivant en dessous du seuil de pauvreté (taille de la bulle). Ce *bubble chart* permet de montrer que les crimes violents et les crimes sur la propriété se déroulent majoritairement dans les pays au sein desquels beaucoup d'habitants vivent sous le seuil de pauvreté (car les cercles

les plus grands sont situés dans la partie supérieure droite du graphe, tandis que les cercles les plus petits se trouvent dans la partie inférieure gauche). On peut en conclure que plus le pays compte beaucoup d'habitants vivant sous le seuil de pauvreté, plus le nombre de crimes (violents et sur la propriété) est important. Il y a tout de même un pays qui fait exception : le Luxembourg, dont le cercle est de petite taille (peu d'habitants vivant sous le seuil de pauvreté), mais qui se trouve pourtant plus haut dans le repère (beaucoup de crimes sur la propriété, mais peu de crimes violents).

Forces :

- Permet d'ajouter une 3^e dimension au *scatter plot*.
- Permet une visualisation rapide des proportions et des concentrations.

Faiblesses :

- Difficile à lire lorsqu'il y a un trop grand nombre de bulles ou qu'elles se chevauchent souvent.
- L'œil humain est mauvais pour évaluer les surfaces.

Bibliographie indicative :

Bock, T. (s.d.). What is a bubble chart. Displayr. Retrieved from <https://www.displayr.com/what-is-a-bubble-chart/?fbclid=IwAR0BAmBhwHUNHOa7masEJ5aK8v9ydOMcqs2X8z1tNe3jWZLLYrQu7Rlquka>

Bubble chart (s.d.). The data visualisation catalogue. Retrieved from https://datavizcatalogue.com/methods/bubble_chart.html

Hardin, M., Hom, D., Perez, R. & Williams, L. (2012). Which chart or graph is right for you? Tell impactful stories with data. Tableau. Retrieved from <http://eng260.tolearn.net/wp->

content/uploads/2015/08/which_chart.pdf?fbclid=IwAR0iiYzGaz1Tjb8NzraUw-uCy3BSFFVxICq1IBmxY9u6RZH2aYmRaCnOM70

Kirk, A. (2016). *Data Visualization : A handbook for data driven design*. London : Sage.

Viègas, F.B., Wattenberg, M., van Ham, F., Kriss, J. & McKeon, M. (2007). Many Eyes, A site for Visualization at internet scale. *IEEE Symposium on Visualization and Computer Graphics*, 13(9), 1121

Le graphique en bulles groupées

Nom français	Graphique en bulles groupées
Nom anglais	<i>Packed bubble chart, bubble chart, packed bubbles views, packed circle chart, bubble cloud, clustered bubble chart</i>
Fonction	Distribution statistique
Nombre de variables	2
Type(s) de variables	discrète
Type(s) de données	nominales ordinales intervalles ratio

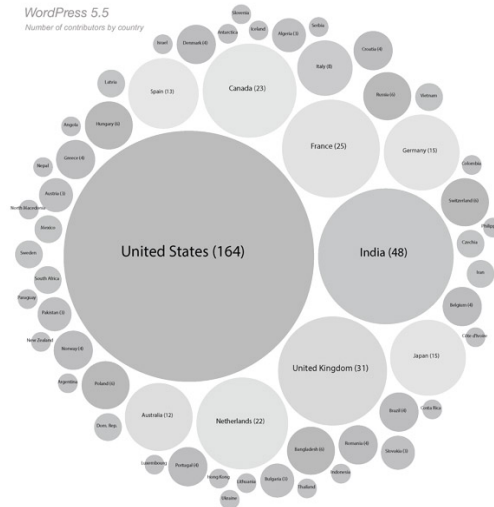


Figure 78 : Bubble chart showing the quantity of professions people, creating programming languages (CC BY-SA 4.0)

Analyse :

Les graphiques à bulles groupées ont une structure de données plus simple que les graphiques à bulles classiques. Un tableau plat à une dimension avec des volumes est suffisant. La position x/y de la bulle est calculée automatiquement à l'aide d'un algorithme qui rassemble les bulles dans un *cluster*. La configuration des points de données prend en charge la définition des couleurs et des valeurs indiquées. La fonction, glisser-déposer, a également été ajoutée pour permettre à l'utilisateur de déplacer rapidement une bulle entre les séries, puis de vérifier comment leurs relations vont changer.

Forces :

- Format compact pour la représentation d'un grand nombre de catégories de données.
- Esthétique agréable, il véhicule rapidement une vue d'ensemble des données.
- Particulièrement efficace pour la mise en avant de la position de différents éléments dans une structure hiérarchique.

Faiblesses :

- Limite du nombre de données pour une bonne lisibilité (Yi, 2019).
- Difficile de tirer des valeurs exactes par la simple taille des bulles (Yi, 2019).

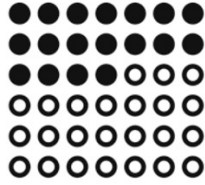
Bibliographie :

Highcharts. 2020. Packed Bubble. Retrieved from: <https://www.highcharts.com/docs/chart-and-series-types/packed-bubble>

Yi, Mike. 2019. A complete guide to Bubble charts. Retrieved from: <https://chartio.com/learn/charts/bubble-chart-complete-guide/>

Le diagramme à points

Nom français	Diagramme à points
Nom anglais	<i>Dot matrix chart</i>
Fonction	Distribution statistique
Nombre de variables	N
Type(s) de variables	discrète
Type(s) de données	nominales ordinales intervalles ratio



Created by Evangeline White
from Noun Project

Figure 79 : Exemple de dot matrix chart : les points noirs correspondent à un groupe ; les points blancs correspondent à un autre groupe. Par Evangeline White (The noun project), 2020 (CC BY)

Analyse

Les tableaux de matrice de points affichent des données discrètes en unités de points, chacune colorée pour représenter une catégorie particulière et regroupée dans une matrice. Ils sont utilisés pour donner un aperçu rapide de la distribution et des proportions de chaque catégorie dans un ensemble de données et également pour comparer la distribution et les proportions dans d'autres ensembles de données, afin de découvrir des modèles.

Lorsqu'une seule variable/catégorie est utilisée dans l'ensemble de données et que tous les points sont de la même couleur, un diagramme à matrice de points peut être utilisé pour montrer principalement les proportions (Ribecca, s.d.)

Force :

- Représentation graphique permettant de donner un aperçu rapide des proportions des catégories dans un ensemble de données (Insightwhale, 2020)

Faiblesse :

- Représentation graphique fonctionnant uniquement avec des nombres modestes ou très ronds où un point peut représenter, par exemple, 10 ou 100 (Shaw, 2017).

Bibliographie indicative :

Insightwhale (2020). How to visualize frequency distribution : *Best methods*

Ribbecca, S (s. d). The data visualisation catalogue. *Dot matrix chart*

Shaw, T. (2017). Good data visualization practices : *Dot plots & kin*

Le diagramme à pictogrammes

Nom français	Diagramme à pictogrammes
Nom anglais	<i>Pictogram chart</i>
Fonction	Distribution statistique
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	intervalles ratio



Figure 80 : Exemple de diagramme à pictogrammes, par les auteurs

Analyse :

Les diagrammes à pictogrammes utilisent des icônes pour représenter des petits ensembles de données discrètes. En général, les icônes représentent le sujet ou la catégorie de données. Il se présente de la même manière que le diagramme à barres, mais au lieu de barres, il utilise des colonnes d'images afin de comparer chaque catégorie entre elles (<https://infogram.com/>, s. d.). Chaque icône peut représenter une unité ou un nombre quelconque d'unités. Ce type de visualisation est couramment utilisé dans les communications commerciales ou les articles de presse pour comparer visuellement les données (<https://datavizcatalogue.com/>, s. d.).

Forces :

- Vue d'ensemble potentiellement plus attrayante que celle d'un diagramme à barres.
- L'utilisation d'icônes peut parfois aider à surmonter les différences de langue, de culture et d'éducation.
- Les icônes peuvent donner une vue plus représentative (explicite) des données (<https://infogram.com/>, s. d.).

Faiblesses :

- Inadapté pour les grands ensembles de données.
- Certaines icônes peuvent induire une confusion sur ce qu'elles représentent.
- Efficace uniquement avec des petits nombres entiers. (<https://infogram.com/>, s. d.).

Bibliographie :

Create a Pictorial Chart. (s. d.). Infogram. Consulté 29 avril 2020, à l'adresse <https://infogram.com//create/pictorial-chart>

Pictogram Chart—Learn about this chart and tools to create it. (s. d.).
Consulté 29 avril 2020, à
l'adresse <https://datavizcatalogue.com/methods/pictogram.html>

Le diagramme quantile-quantile

Nom français	Diagramme quantile-quantile
Nom anglais	<i>Quantile-Quantile plot (Q-Q plot)</i>
Fonction	Distribution statistique
Nombre de variables	1
Type(s) de variables	discrète continue
Type(s) de données	ordinales intervalles ratio

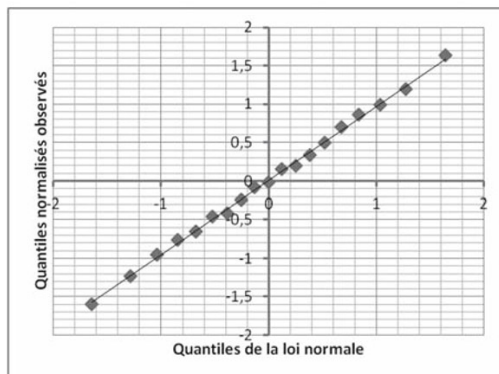


Figure 81 : Exemple de diagramme quantile-quantile, par les auteurs

Analyse :

Un diagramme Quantile-Quantile (*Q-Q plot*) permet d'analyser rapidement la fréquence de distribution d'une variable quantitative. Plus puissant que l'histogramme ou la boîte à moustaches, il permet de comparer deux distributions statistiques en affichant leurs quantiles. Si les distributions sont similaires, les valeurs tracées se situeront à peu près le long de la diagonale centrale (voir exemple).

Si elles sont linéairement dépendantes, les valeurs tracées se situeront à nouveau le long d'une droite, mais avec une pente et une intersection variables.

L'exemple sert à comparer une loi de distribution préalablement centrée et réduite avec une loi normale $N(0, 1)$. Pour un autre exemple d'utilisation du *Q-Q plot*, voir Figure 2 b dans (Heer *et al.*, 2010 ; p. 62) qui compare un ensemble de données à trois distributions statistiques : uniforme, normale et mélange ajusté de trois distributions normales. Le mélange ajusté de trois distributions normales fournit un meilleur ajustement avec l'ensemble de données ; ce qui suggère qu'un modèle statistique à trois composantes pourrait être plus approprié.

Forces :

- Représentation graphique permettant d'analyser efficacement la distribution statistique d'un ensemble de données (Kohler & Kreuter, 2005).
- Représentation graphique plus puissante que l'histogramme ou la boîte à moustaches.

Faiblesses :

- Représentation graphique nécessitant des connaissances avancées en statistiques : distribution statistique, quantile, lois de probabilité (Heer *et al.*, 2010).
- Représentation graphique incongrue pour montrer des données ordinales.

Bibliographie indicative :

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Commun. Acm*, 53 (6), 59-67.

Kohler, U., & Kreuter, F. (2005). *Data analysis using Stata*. Stata press.

La boîte à moustaches

Nom français	Boîte à moustaches
Nom anglais	<i>Box and whisker plot</i>
Fonction	Distribution statistique
Nombre de variables	1
Type(s) de variables	discrète continue
Type(s) de données	ordinales intervalles ratio

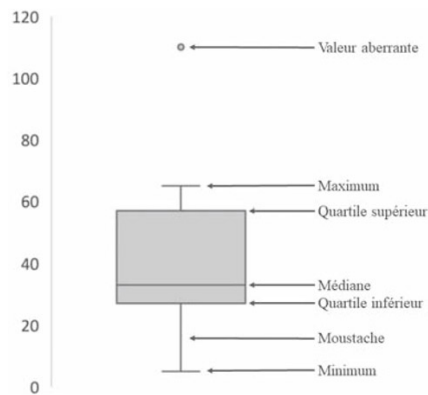


Figure 82 : Annotations d'un exemple de boîte à moustaches, par les auteurs

Analyse :

Une boîte à moustaches permet d'analyser rapidement la distribution d'une variable quantitative. Dans l'exemple ci-dessus, la boîte bleue contient 50 % des données mesurées. Les deux zones situées de part et d'autre de la médiane correspondent au quartile supérieur et au quartile inférieur, chaque quartile contenant 25 % des données. La moustache supérieure relie la boîte au maximum, la moustache inférieure au minimum. Comme elle ne se trouve pas à l'intérieur de

la boîte, la valeur 110 est suspecte et doit être examinée pour vérifier s'il s'agit d'une valeur aberrante ou d'une erreur de saisie. L'asymétrie de la boîte à moustaches (médiane située en bas de la boîte et moustaches de longueur différente) indique une asymétrie de la distribution des données mesurées. Donc, cette distribution ne suit pas une loi normale.

Forces :

- Représentation graphique permettant de comparer rapidement et facilement la distribution de N variables en juxtaposant N boîtes à moustaches (Potter *et al.*, 2006).
- Représentation graphique permettant de réaliser un gain de place important par rapport à l'histogramme pour montrer une ou plusieurs distributions.

Faiblesses :

- Représentation graphique nécessitant des connaissances en statistiques : distribution statistique, quartile, médiane ou valeur aberrante. Voir Singh (2019) pour une explication reposant sur une métaphore avec le visage humain.
- Représentation graphique incongrue pour montrer des données ordinales.

Bibliographie indicative :

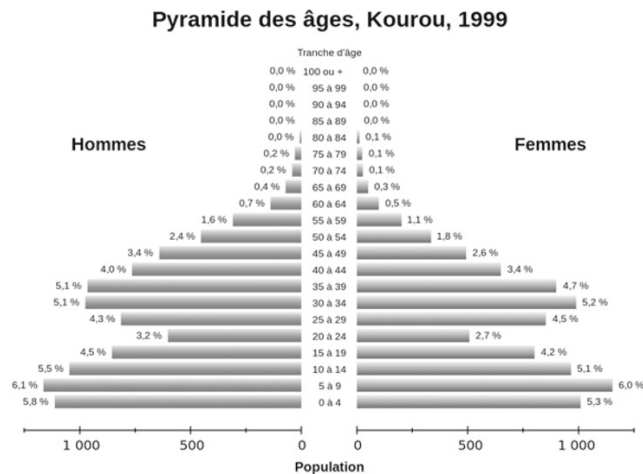
Potter, K., Hagen, H., Kerren, A., & Dannenmann, P. (2006). Methods for presenting statistical information: The box plot. *Visualization of large and unstructured data sets*, 4, 97-106.

Singh, S. (2019). Box Plot versus Human Face. *CHANCE*, 32 (2), 28-30.

Williamson, D. F., Parker, R. A., & Kendrick, J. S. (1989). The box plot: a simple visual method to interpret data. *Annals of internal medicine*, 110(11), 916-921

La pyramide des âges

Nom français	Pyramide des âges
Nom anglais	<i>Population pyramid, age-gender pyramid</i>
Fonction	Structure d'une population
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	nominales intervalles ratio



Source : INSEE, recensement de la population 1999, exploitation principale et complémentaire

Figure 83 : Pyramide des âges, Kourou, 1999

Analyse :

Une pyramide des âges, aussi appelée « diagramme de la structure des âges », est une paire d'histogrammes inversés et accolés qui représente la répartition par genre et par âge d'une population à un instant donné. Le côté gauche montre généralement le nombre

d'hommes et le côté droit, le nombre de femmes. Les tranches d'âges sont le plus souvent indiquées au centre et parfois sur le côté gauche, et les effectifs sont exprimées le plus souvent en chiffres ou en pourcentages. La longueur de chaque barre est mesurée par les valeurs (effectifs) disposées sur l'abscisse (axe X – horizontal), et les âges sont disposés verticalement sur l'ordonnée (axe Y – vertical) (Laplante, 2005).

Les sections rétrécies de la pyramide indiquent soit une baisse de la natalité (*baby-krach*) due à une période de difficultés économiques exceptionnelles, soit un taux de mortalité élevé dû à une circonstance importante (une guerre, par exemple). Utilisée le plus souvent pour analyser la situation démographique d'un pays, la pyramide des âges donne aux responsables politiques un outil visuel pour l'aide à la prise de décisions éclairées pour l'avenir. Ce type de graphique est aussi utilisé dans les institutions et les entreprises (Laplante, 2005).

Il existe différents profils des pyramides des âges :

- le profil « champignon » : la base est très étroite et le sommet large. Cela montre une population vieillissante ;
- le profil « pyramidal » : la base est large et le sommet est fin ;
- le profil « poire » : la base s'élargit avant de rétrécir vers la fin ;
- le profil « pagode » : la base est rétrécie et a des encoches, ce qui correspond à une période de natalité ou de mortalité importante ;
- le profil « sablier » : la base est large, puis elle se rétrécit au milieu avant de s'élargir au sommet ;
- le profil « ballon de rugby » : la base est étroite, puis elle s'élargit au milieu avant de se rétrécir au sommet.

Forces :

- Idéal pour détecter les changements ou les différences dans les schémas de population.
- Utile pour interpréter une population à un moment donné et spéculer sur l'évolution future de celle-ci.
- La construction d'une seule pyramide des âges peut être réalisée indifféremment, à partir des quantités absolues ou/et des quantités relatives (Bonin, 1991).
- Utile pour divers domaines : écologie, sociologie, économie, etc.

Faiblesses :

- L'évaluation de la quantité d'individus correspondant à une classe d'âge, pour l'un des deux sexes, reste très approximative. La recherche d'un nombre précis relève de la consultation d'un tableau de données (Bonin, 1991).
- Quel que soit le nombre de classes d'âge représentées, la comparaison des structures d'âge masculine et féminine est difficile et longue, car obligation de lire, à partir de l'axe vertical, vers la gauche pour les uns et vers la droite pour les autres. Seules des différences nettement marquées permettent d'apprécier les inégalités entre hommes et femmes au niveau de l'ensemble (Bonin, 1991).
- Peu lisible si le nombre ou le pourcentage n'apparaît pas à côté de chaque bâton.
- Problème d'études des différences et de comparaisons accentuées lorsque deux ou plusieurs pyramides, présentées en juxtaposition ou en superposition, sont mises en relation. La comparaison de deux ou plusieurs pyramides passe par l'utilisation des pourcentages, les pyramides établies à partir des

quantités absolues risquant de ne faire apparaître que les quantités totales de population (Bonin, 1991).

Bibliographie indicative :

Bonin, S. (1991). Représentation graphique des structures de population. *Espace Populations Société*, 3, 539-547. Retrieved from https://www.persee.fr/doc/espos_0755-7809_1991_num_9_3_1503

Laplante, A. (2005). *La Pyramide des âges*. Retrieved from <https://www.ebsi.umontreal.ca/jetrouve/illustre/barres3.htm> (Laplante se présente comme professeure à l'Université de Montréal).

Le diagramme en étoile

Nom français	Diagramme en étoile
Nom anglais	<i>Radar chart</i> (également appelé <i>spider chart</i>)
Fonction	Relations
Nombre de variables	N
Type(s) de variables	discrète
Type(s) de données	ordinales intervalles ratio

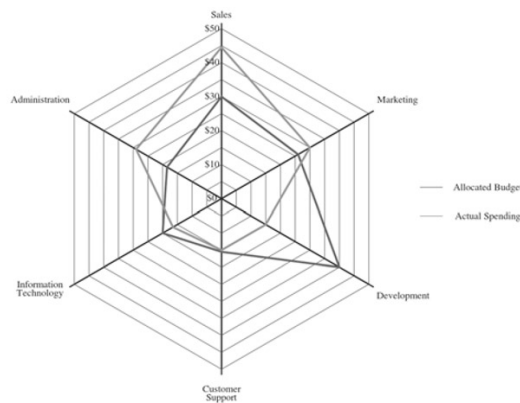


Figure 84 : Spider Chart, 2008, par Ordoon (Domaine public)

Analyse :

Le diagramme en étoile est idéal pour montrer les relations entre les données, car il permet d'afficher des données multivariées. Chaque rayon correspond à une variable (Kirk, 2016). Les rayons ont habituellement la même unité et la même échelle. Tout en gardant un bon niveau de visibilité, on pourrait afficher jusqu'à huit variables. Dans l'exemple ci-contre, il y en a six.

Le diagramme en étoile peut également être affiché grâce à des polygones. Cela pourrait néanmoins créer certains problèmes d'occlusion à maîtriser.

Force :

- Détection rapide des minima et maxima au sein de différents ensembles.

Faiblesses :

- Difficile de donner une vue d'ensemble lorsque plusieurs variables ont la même valeur sur un axe donné (Hongliang *et al.*, 2008)
- Rapidement illisible si le nombre de catégories est trop élevé (ne pas dépasser huit catégories).
- Occlusion possible dans le cas d'utilisation de polygones et/ou de grands jeux de données.

Bibliographie indicative :

Hongliang, L., Anxin, L., Bin, Z., Tiefu, Z., & Xin, Z. (2008). A Fuzzy Comprehensive Evaluation Method of Maintenance Quality Based on Improved Radar Chart. *2008 ISECS International Colloquium on Computing, Communication, Control, and Management*, 1, 638-642. <https://doi.org/10.1109/CCCM.2008.208>

Kirk, Andy (2016.). Data Visualization : A handbook for data driven design (Sage). Thousands Oaks.

Le tableau

Nom français	tableau
Nom anglais	<i>table</i>
Fonction	Valeurs
Nombre de variables	N
Type(s) de variables	discrète continue
Type(s) de données	nominales ordinales intervalles ratio

Rangs des appréciations récoltées

user	RBelle	RIntéré	RClaire	RCompris	RJaime	Dataink1	Embellisheddataink1
1	379.000	468.500	371.500	499.500	388.500	82.00%	0.00%
2	551.000	468.500	230.500	89.500	388.500	82.00%	0.00%
4	379.000	120.500	41.500	499.500	70.000	82.00%	0.00%
6	686.500	664.500	602.500	499.500	699.000	82.00%	0.00%
8	551.000	261.000	602.500	499.500	388.500	82.00%	0.00%
10	551.000	664.500	371.500	499.500	572.500	82.00%	0.00%
12	686.500	468.500	132.000	187.000	70.000	82.00%	0.00%
14	551.000	261.000	41.500	187.000	388.500	82.00%	0.00%
16	551.000	468.500	132.000	187.000	572.500	82.00%	0.00%
18	551.000	261.000	230.500	499.500	572.500	82.00%	0.00%
20	205.000	468.500	132.000	187.000	214.000	82.00%	0.00%
23	686.500	468.500	371.500	499.500	699.000	82.00%	0.00%
25	379.000	261.000	230.500	187.000	214.000	82.00%	0.00%
27	551.000	468.500	230.500	187.000	572.500	82.00%	0.00%
29	686.500	468.500	602.500	499.500	699.000	82.00%	0.00%
31	551.000	664.500	230.500	499.500	572.500	82.00%	0.00%
33	551.000	468.500	371.500	499.500	572.500	82.00%	0.00%
35	686.500	664.500	230.500	499.500	388.500	82.00%	0.00%
37	686.500	468.500	132.000	187.000	572.500	82.00%	0.00%
39	686.500	468.500	230.500	499.500	388.500	82.00%	0.00%
2	379.000	34.500	41.500	44.500	70.000	47.00%	0.00%
3	551.000	120.500	41.500	44.500	214.000	47.00%	0.00%
5	686.500	34.500	41.500	499.500	70.000	47.00%	0.00%
7	551.000	120.500	41.500	13.500	214.000	47.00%	0.00%
9	379.000	468.500	132.000	499.500	572.500	47.00%	0.00%
11	686.500	664.500	230.500	499.500	572.500	47.00%	0.00%

Figure 85 : Tableau de données représentant les rangs d’appréciations récoltées par questionnaires, dans (Andry, 2021) – par les auteurs

Analyse :

Un tableau est une série d'informations, de données, disposées de façon claire et systématique, permettant une consultation rapide et globale (Usito, s. d.). On distingue deux types de tableau : les tableaux à caractère qualitatif (utiliser pour organiser les données) et les tableaux à caractère quantitatif. Notons qu'il est possible d'avoir un troisième type qui est une combinaison des deux types précédents. On distingue deux sortes de tableaux à caractère quantitatif : à variable discrète ou à variable continue. Les données d'un tableau peuvent être les microdonnées (données brutes), ou des données agrégées (traitées au niveau statistique).

La structure d'un tableau est la suivante. Les lignes indiquent les unités statistiques (plus petit élément décrit par une enquête, exemple : étudiant). Les colonnes renseignent les variables (exemple : pays, sexe, âge, année). Les cellules, c'est-à-dire les intersections entre lignes et colonnes, contiennent les valeurs que prennent les variables pour chacune des unités statistiques (exemple : dans le tableau ci-dessus, on a un individu de nationalité indienne et de sexe masculin, avec comme situation familiale célibataire...).

Forces :

- Vision claire et rapide de l'information.
- Permet de lire des nombres exacts.
- Facilitent la comparaison de paires de valeurs connexes.

Faiblesses :

- Ne permet pas de comparer des ensembles entiers de nombres.
- Traitement uniquement verbal (texte).
- Ne permet pas un aperçu rapide de l'histoire contenue dans les chiffres.

Bibliographie indicative :

Usito. (s. d.). Usito. Usito. Consulté 14 avril 2020, à l'adresse <https://usito.usherbrooke.ca/définitions/table>

Les tableaux de données. (s. d.). Consulté 14 avril 2020, à l'adresse <https://commentprogresser.com/statistique-tableau.html>

3.2 Visualisations temporelles

Le graphique linéaire

Nom français	Graphique linéaire
Nom anglais	<i>Line graph</i>
Fonction	Chronologie
Nombre de variables	2
Type(s) de variables	continue
Type(s) de données	intervalles ratio

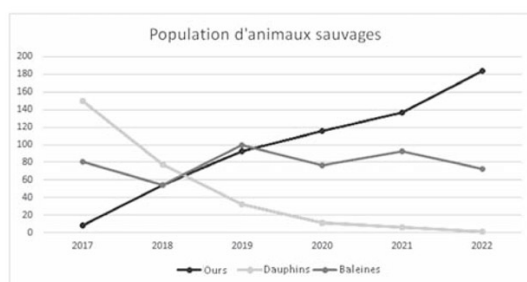


Figure 86 : Population d'animaux sauvages, données fictives, par les auteurs

Analyse :

Le graphique linéaire, *line chart* ou *line graph* en anglais, permet d'afficher des informations sous la forme d'une série de points de données reliées par de segments linéaires. Ce genre de graphique est généralement utilisé pour visualiser les tendances des données sur

des intervalles de temps. Ils sont donc utiles pour visualiser les données qui changent continuellement au fil du temps (Acharya & Chellappan, 2017, p. 592) et montrer la tendance dans l'ensemble des données.

Il existe de nombreuses formes ou variations de graphique linéaire dépendant des points de données à tracer tels que *Curve line chart*, *Step line chart*, etc. (Khan & Khan, 2011, pp. 5-6).

Forces

- Permet une analyse rapide des données : montrer les changements au fil du temps, même les plus petits
- Permet de voir plusieurs groupes et de comparer leur évolution dans le temps.
- Utilisé pour montrer les changements à court et à long terme.

Faiblesses

- Ce graphique n'est utile que si la notion du temps est une des deux variables.
- Ce graphique n'est généralement bon que s'il y a au moins 50 valeurs de données.
- Moins visuellement attrayant que d'autres graphiques.

Bibliographie indicative :

Acharya, S. & Chellappan, S. (2017), *Pro Tableau A step by step guide*, Berkeley, p. 592.

Bruce, O. (2015), *What are the advantages of using a line graph to represent data?*, eNotesEditorial, Retrieved from : <https://www.enotes.com/homework-help/what-advantages-using-line-graph-represent-data-514721>

Khan, S. & Khan, M. (2011), *Data and Information Visualization Methods, and Interdactive Mechanisms : A survey*, International Journal of computer applications, pp. 5-6.

Le graphique en aires empilées

Nom français	Graphe en aires empilées
Nom anglais	<i>Stacked area chart</i>
Fonction	Chronologie
Nombre de variables	N
Type(s) de variables	continue
Type(s) de données	intervalles ratio

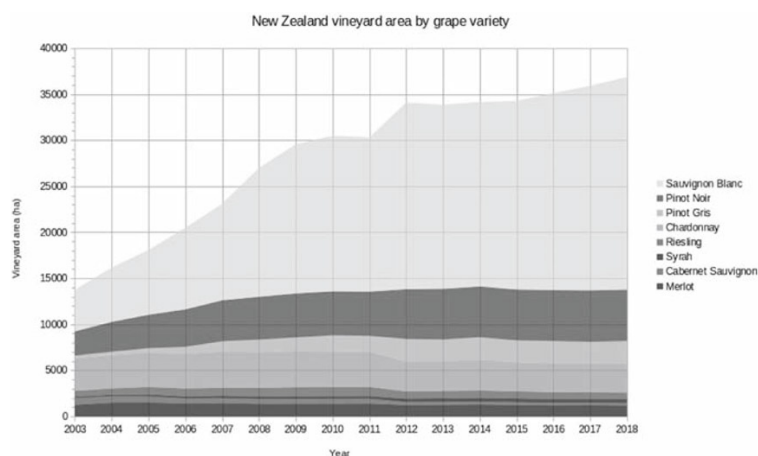


Figure 87 : New Zealand vineyard area by grape variety, by Jonathanischoice, https://commons.wikimedia.org/wiki/File:NZ_vineyard_area_by_grape_variety.png (CC BY-SA)

Analyse :

Yan Holtz (s.d.) et Edward Tufte (2010) définissent le graphique en aires empilées (*stacked area chart* en anglais) comme étant une variation du graphique en aire. Le graphique en aires empilées

permet de montrer l'évolution de N variables quantitatives, tandis que le graphique en aire montre l'évolution d'une seule variable quantitative. L'axe des abscisses correspond au temps où une variable quantitative ordonnée et l'axe des ordonnées représentent la valeur d'une autre variable.

L'exemple ci-dessus montre l'évolution des parcelles de vignobles en Nouvelle-Zélande, en fonction de la variété de raisin. Nous pouvons vite constater que le Sauvignon Blanc est la variété la plus répandue en Nouvelle-Zélande depuis 2003.

Audrey Laplante (2009) identifie cinq types de graphique en aires empilées qui dépendent de la façon dont les aires sont agencées : aires empilées avec effet 3D, deux aires empilées, aires multiples empilées où aucune aire n'en cache une autre, aires empilées avec projection, et le graphique en aires empilées à 100 % qui couvre tout l'espace.

Forces :

- Permet de montrer des variations dans le temps et également l'amplitude du changement.
- Utilise tout l'espace disponible et rend les données mieux visibles.
- Idéal pour représenter une série d'éléments qui forment un tout et qui évoluent dans le temps.

Faiblesses :

- L'espace du graphe réduit le nombre d'éléments à observer, au plus il y en a, plus illisible le graphique deviendra.
- L'empilement peut rendre difficile l'interprétation précise des tendances. La recherche et le filtrage sont souvent utiles pour compenser ce problème.

- Peut être remplacé, pertinemment, par de multiples graphes en aires, voire des graphes en lignes.
- Heer et ses collègues (2010) précisent que ce type de graphique ne prend pas en compte les nombres négatifs et n'ont aucun sens pour les données qui ne doivent pas être additionnées comme la température.

Bibliographie indicative :

Heer, J. *et al.* (2010). A tour through the visualization zoo. Queue, 8,5.

Holz, Y. (s.d.).The Python Graph Gallery. Site web : <https://python-graph-gallery.com/stacked-area-plot/>. Consulté le 28 avril 2020.

Holz, Y. (s.d.). From data to viz. Site web : <https://www.data-to-viz.com/>. Consulté le 15 avril 2020.

Laplante A. (2009). Biblio-Internet – Université de Montréal. Site web : <https://www.ebsi.umontreal.ca/jetrouve/>. Consulté en ligne le 18 avril 2020.

Tufte E., *The Visual Display of Quantitative Information*, Cheshire, CT, Graphics Press, 2001, 2^e éd.

Le graphique en grille

Nom français	Graphique en grille
Nom anglais	<i>Small multiples</i>
Fonction	Chronologie
Nombre de variables	2
Type(s) de variables	discrète continue
Type(s) de données	intervalles ratio

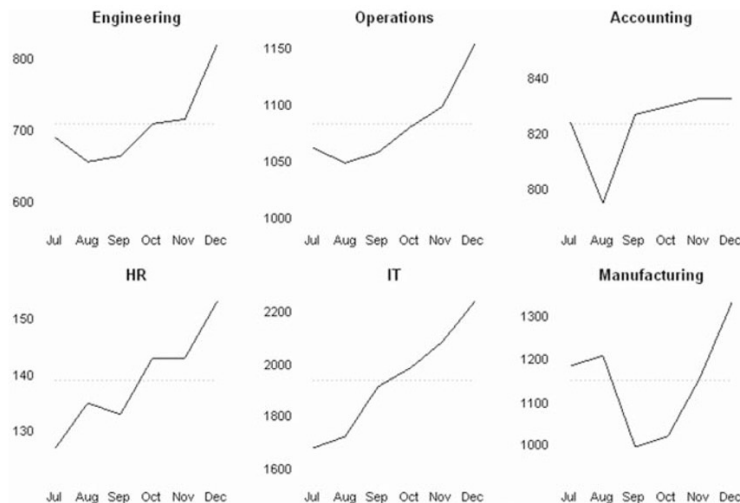


Figure 88 : Departmental salary expenses (Domaine public)

Analyse :

Les *Small multiples* sont un type de visualisation proposé pour la première fois par Edward Tufte (1983). L'objectif est de comparer l'ensemble des données changeantes et évolutives dans le temps (Brock, 2018). Ils rassemblent plusieurs graphiques dans une grille et peuvent porter plusieurs noms : graphique en treillis, graphique en grille ou graphique en panneaux. Toutes ces appellations font référence à une « série organisée de graphiques similaires, affichant une vue différente du même ensemble de données » (Brock, 2015).

Forces :

- Les *small multiples* peuvent être une force dans votre travail de visualisation lorsque vous possédez des données présentant un « overplotting » : un ensemble complexe de données. En effet, dès qu'on se trouve devant ce genre de visualisation, il est difficile de comparer et de contraster les données. Ils permettent d'avoir une vision d'ensemble. (Brock, 2018)
- Les *small multiples* permettent très facilement une comparaison de toutes les données.

Faiblesses :

- Ils ne sont pas la meilleure visualisation pour les problématiques impliquant une seule variable. Par exemple, si l'on veut comparer des parkings payants uniquement en fonction de prix, il est conseillé d'utiliser un autre type de visualisation. Il ne faut donc pas penser que le graphique en grille soit la solution à tout (Brock, 2018).
- Les *small multiples* peuvent prendre une place considérable dans le tableau de bord final d'une recherche. Bien que chaque graphique soit petit, on se retrouve avec une visualisation très grande qui ne s'adapte pas forcément à tous les supports et navigateurs. On peut prendre l'exemple des navigateurs sur téléphone mobile qui ne présentent pas de manière optimale les *small multiples* (Brock, 2018)

Bibliographie indicative :

- Brock, T. (2015, 11 août). *An introduction to small multiples*. Infragistics, https://www.infragistics.com/community/blogs/b/tim_brock/posts/an-introduction-to-small-multiples
- Brock, T. (2018, 26 août). *Whare are small multiples*. Display, <https://www.displayr.com/what-are-small-multiples/>
- Tufte, E. (1983). *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT 1983.

Le graphique en spirale

Nom français	Graphique en spirale
Nom anglais	<i>Spiral plot</i>
Fonction	Chronologie
Nombre de variables	1
Type(s) de variables	continue
Type(s) de données	nominales ordinales

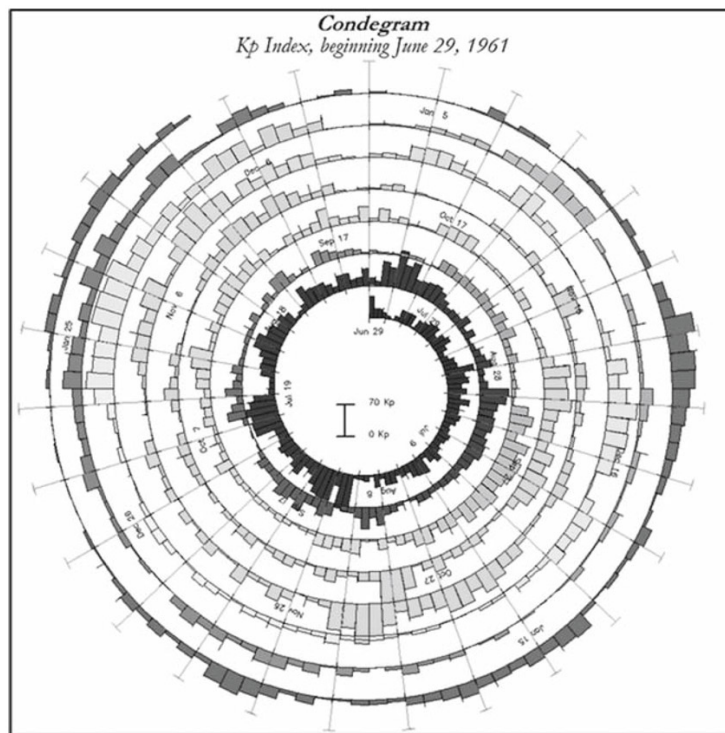


Figure 89 : Condegram Spiral Plot, 2011, FigN44 (Domaine Public)

Analyse :

Selon le catalogue de *Data Visualisation* développé par Severino Ribecca (s.d.), le graphique en spirale – aussi appelé spirale de séries temporelles – permet de tracer des données temporelles le long d’une spirale d’Archimède. Il peut contenir des barres, des lignes ou des points le long du chemin de la spirale. Le chemin commence au

centre et évolue vers l'extérieur. Selon Ribecca, les diagrammes en spirale sont efficaces pour montrer des tendances sur une grande période de temps. Selon Marc Weber, Marc Alexa et Wolfgang Muller (2001), le graphique en spirale permet de visualiser de grands ensembles de données et d'identifier de manière intuitive des structures périodiques dans les données. Il fonctionne aussi bien pour la visualisation de données nominales que pour celle de données quantitatives en utilisant une métaphore de visualisation similaire. Dans leur étude, les auteurs inscrivent les graphiques en spirale comme une amélioration dans l'identification des modèles périodiques. Ils permettent de compléter d'autres techniques de visualisation concernant les séries chronologiques. Par ailleurs, cette visualisation convient à une démonstration en 3D en mobilisant les concepts de *zoom*, de focalisation et de liaison des données. Par exemple, si la spirale représentée sur la figure 97 devait représenter une année, nous pourrions afficher les données en attribuant une couleur pour chaque mois de l'année. De cette manière, on pourrait décomposer les périodes et les comparer les unes avec les autres.

Forces :

- Le graphique à spirale combine les avantages des graphiques classiques – à barres, à points ou à lignes – avec ceux des graphiques circulaires. Les premiers sont efficaces pour l'analyse des séries de données et les graphiques circulaires sont idéals pour comparer les comportements périodiques dans de plus petits ensembles de données dans une série chronologique.
- Il permet une lecture comparative des ensembles de données.
- Il peut facilement amener aux graphiques à spirale multiples permettant la comparaison entre plusieurs ensembles de données.

- Il permet à l'œil humain de détecter facilement les structures dans la visualisation.
- Les cycles sur le graphique sont facilement visibles et permettent la comparaison des ensembles de données périodiques.
- La forme de la spirale permet une bonne visualisation de la continuité des données.
- Force prédominante par rapport aux autres visualisations : l'identification des modèles périodiques.

Faiblesses :

- La spirale peut se poursuivre à l'infini.
- Le graphique en spirale n'est pas optimal pour les petits ensembles de données.

Bibliographie indicative :

Butters, T. (2014). Spiral Chart Visualisation : an interactive data exploration tool. Sabisu. Retrieved from <http://www.sabisu.co/wp-content/uploads/2015/05/spiral-chart-visualisation.pdf>

Makhabel, D., Mishra, P., Danneman, N. & Heimann, R. (2017). R : Mining spatial, Text, Web, and Social Media Data, 69-70. Numbai, Inde : Packt Publishing.

Ribbecca, S. (s. d). The data visualisation catalogue. Consulté le 10 mai 2020, à l'adresse <https://datavizcatalogue.com/index.html>

Weber, M., Alexa, M. & Muller, W. (2001). Visualizing Time-Series on Spirals. DBLP. doi : 10.1109/INFVIS.2001.963273

Le diagramme de Gantt

Nom français	Diagramme de Gantt
Nom anglais	<i>Gantt chart</i>
Fonction	Ordonnancement de tâches
Nombre de variables	N
Type(s) de variables	discrète
Type(s) de données	nominales ordinales intervalles

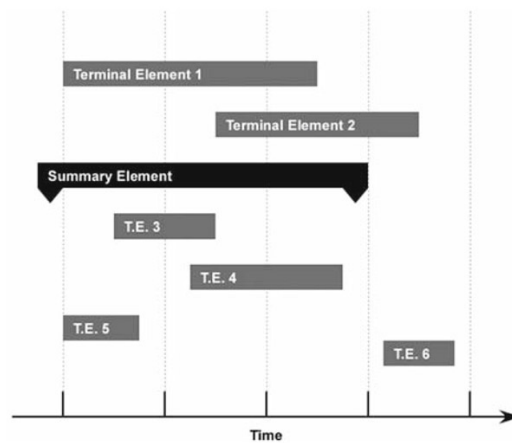


Figure 90 : Gantt Chart, 2006, par Maksim (CC BY-SA 3.0)

Analyse :

Le diagramme de Gantt est un outil d'organisation qui facilite la gestion de projet. Il permet d'afficher une liste d'éléments (tâches, activités, événements, etc.) et leur durée dans le temps. Il est ainsi possible de voir quand un élément commence, termine et s'il chevauche un autre élément (Mayor, 2001).

Les éléments sont organisés, placés verticalement, et le temps est renseigné sur l'axe des abscisses. Souvent exprimée en mois (unité de temps), la durée des éléments est proportionnelle la longueur de la barre horizontale. La couleur permet d'encoder des catégories d'éléments.

Forces :

- Parfait pour planifier et estimer la durée totale d'un projet.
- Permet d'identifier les conflits de dates et les goulots d'étranglement entre les éléments.

Faiblesses :

- Difficulté de comparer la taille des activités entre elles.

Bibliographie indicative :

Maylor, H. (2001). Beyond the Gantt chart : Project management moving on. European management journal, 19 (1), 92-100.

Le graphique de flux

Nom français	Graphique de flux
Nom anglais	<i>Stream graph</i> ou <i>horizon graph</i>
Fonction	Chronologie
Nombre de variables	2
Type(s) de variables	continue
Type(s) de données	intervalles ratio

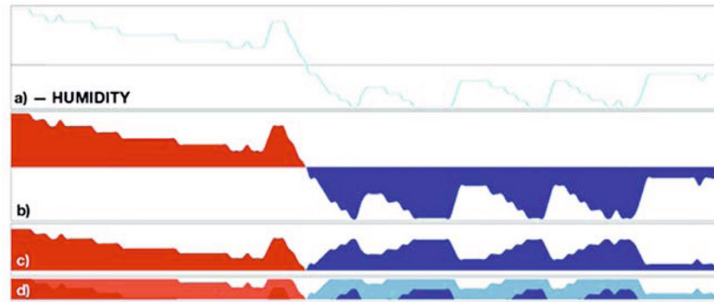


Figure 91 : Horizon Chart ; Figure 6 in Akram Hassan, K., Liu, Y., Besançon, L., Johansson, J., & Rönnerberg, N. (2019). A study on visual representations for active plant wall data analysis. *Data*, 4 (2), 74, (CC BY), DOI:10.3390/data4020074,

Analyse :

L'*horizon graph* permet de représenter une variable selon N modalités ou N catégories de façon chronologique et de présenter les valeurs positives ou négatives des modalités sur la même abscisse (Aquaclusters, 2015 ; Perin *et al.*, 2012).

L'origine de ce graphique est le graphique en aires empilées. En divisant ce graphique initial en couches consécutives, les graphiques de flux présentent les données de séries temporelles dans un espace compact tout en préservant la visibilité des données. Il s'agit d'une option plus efficace en termes d'espace pour afficher une quantité dense de données. Les couleurs basiques telles que le rouge et le bleu ou le rouge et le vert sont souvent utilisées pour encoder les valeurs positives ou négatives, en fonction des sujets abordés.

Forces :

- Permet de représenter des données négatives et positives sur le même axe d'abscisse.
- Permet la visualisation d'un gros ensemble de données sur un même graphique.

Faiblesses :

- Ne convient pas pour analyser des données individuelles et discrètes.

Bibliographie indicative :

Aquacluster,

<https://www.aquaclusters.com/app/home/project/public/aquadatastudio/wikibook/New-Features-17.0/page/Horizon-Charts/Horizon-Charts>, (En ligne), consulté le 9 avril 2020.

Flowingdata, <https://flowingdata.com/2015/07/02/changing-price-of-food-items-and-horizon-graphs/>, (En ligne), consulté le 9 avril 2020.

Perin C., Vernier F., Fekete J.-D. « Progressive Horizon Graphs: Improving Small Multiples Visualization of Time Series », in *IEEE Conference on Information Visualization (INFOVIS)*, IEEE, Oct 2012, Seattle, Washington, USA, United States. p. hal-00734497 (En ligne), consulté le 13 avril 2020.

3.3 Cartes

La carte à bulles

Nom français	Carte à bulles
Nom anglais	<i>Bubble map</i>
Fonction	Distribution géographique
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	ordinales intervalles ratio

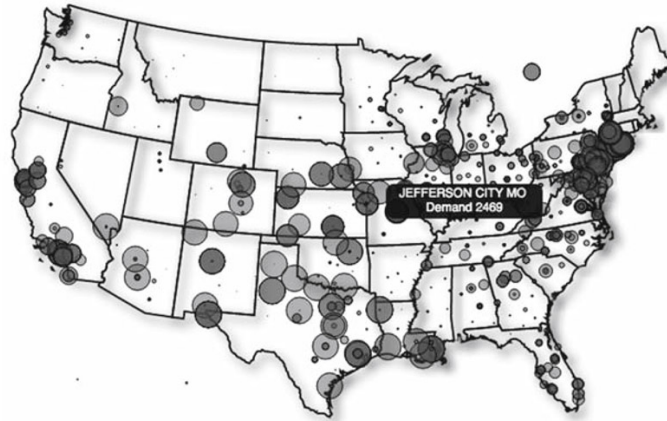


Figure 92 : Bubble Map, 2015 (CC BY-SA)

Analyse :

Sur ce type de cartes, les valeurs quantitatives d'une variable statistique sont associées à leur localisation géographique. Les données sont représentées par des aires (surfaces) proportionnelles aux valeurs dans l'ensemble de données. La couleur peut être utilisée pour distinguer des catégories (Kirk, 2016).

Forces :

- Permet de comparer des valeurs quantitatives sur des régions géographiques en évitant les problèmes de distorsion entraînés par les anamorphoses.

Faiblesses :

- L'œil humain est mauvais élève lorsqu'il s'agit d'évaluer et de comparer des surfaces et des volumes.
- Risque de chevauchement et d'occlusion entre des bulles trop grosses.

Bibliographie indicative :

Kirk, Andy (2016). Data Visualization : A handbook for data driven design (Sage). Thousands Oaks.

La carte choroplèthe

Nom français	Carte choroplèthe
Nom anglais	<i>Choropleth map</i>
Fonction	Distribution géographique
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	ordinales intervalles ratio

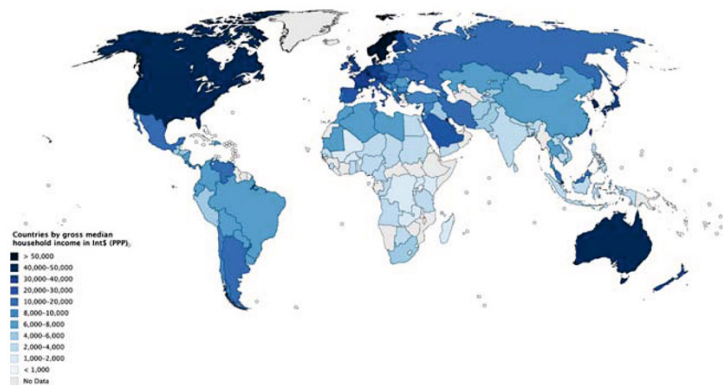


Figure 93 : Countries by gross median household income in Int\$ (PPP), 2018 (CC BY-SA)

Analyse :

Les cartes Choropleth affichent les valeurs quantitatives d'une variable quantitative discrète de manière géolocalisée, c'est-à-dire sur une carte. Chaque région géographique est caractérisée par une zone polygonale, représentant traditionnellement la région illustrée. Ces régions sont colorées, ombragées ou structurées par rapport à la variable quantitative. Les valeurs sont donc encodées par la couleur, ce qui permet de visualiser les valeurs sur une zone géographique (Kirk, 2016). Typiquement, les couleurs correspondent à un schéma

de couleurs séquentiel : la saturation augmente sur une seule et unique teinte (Tennekes, 2018).

Notre exemple montre le revenu médian brut des ménages à parité de pouvoir d'achat. Les pays où la valeur de cet indicateur est la plus élevée sont l'Australie, le Canada, la Corée du Sud, les États-Unis d'Amérique, le Luxembourg, la Norvège, Singapour et la Suède (bleu très foncé). Les pays où elle est la plus faible se situent en Afrique et en Asie du Sud Est (bleu très clair).

Force :

- Permet de comparer des régions géographiques entre elles en évitant les problèmes de distorsion entraînés par les anamorphoses.

Faiblesses :

- Cartes de densité erronées si encodage de données brutes au lieu de valeurs normalisées (Heer *et al.*, 2010).
- Perception de certaines zones affectées par la couleur des leurs voisins (Heer *et al.*, 2010).
- Manque de précision lors de la comparaison de zones très distantes.

Bibliographie indicative :

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, 53(6), 59-67.

Kirk, A. (2016). *Data Visualization : A handbook for data driven design* (Sage). Thousands Oaks.

Tennekes, M. (2018). tmap: Thematic Maps in R. *Journal of Statistical Software*, 84(6), 1-39.

L'anamorphose géographique

Nom français	Anamorphose géographique
Nom anglais	<i>Area cartogram</i>
Fonction	Distribution géographique
Nombre de variables	1
Type(s) de variables	discrète continue
Type(s) de données	nominales ordinales intervalles ratio

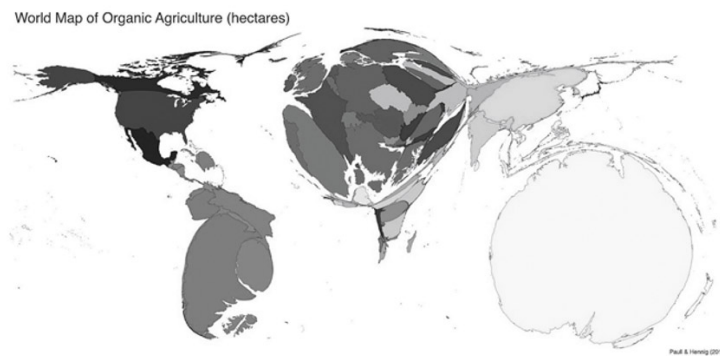


Figure 94 : World Map of Organic Agriculture, 2016, par Paull & Hennig, 2016 (CC BY-SA)

Analyse :

Les anamorphoses géographiques changent la signification des espaces visibles sur les cartes (Bertin, 1967). Il s'agit du codage d'une variable non géographique par la déformation du plan géographique. En d'autres termes, l'objectif n'est pas de représenter la forme géographique réelle, mais de montrer l'ampleur des données issues du terrain en augmentant ou diminuant la taille de cet espace géographique (ce qui n'est dès lors plus représentatif de la réalité géographique) (Dent, 1975).

Sur notre exemple, on pourrait imaginer que la différence de taille soit utilisée pour signifier le nombre d'hectares de culture organique.

La couleur peut être utilisée pour distinguer les pays ou pour encoder une troisième variable.

Force :

- Facile de repérer le pays montrant la plus haute valeur.

Faiblesses :

- Nécessite que le public connaisse très bien les zones géographiques illustrées.
- Cela ne permet pas de lire ou comparer avec précision les valeurs de la carte.
- La précision géographique est perdue.

Bibliographie indicative :

Bertin, J. (1967). « Sémiologie graphique », Paris, Mouton.

Dent, B. D. (1975). Communication aspects of value-by-area cartograms. *The American Cartographer*, 2(2), 154-168.

La carte à symboles gradués

Nom français	Carte à symboles gradués
Nom anglais	<i>Graduated symbol map</i>
Fonction	Distribution géographique
Nombre de variables	2 3
Type(s) de variables	discrète
Type(s) de données	nominales ordinales intervalles ratio



Figure 95 : US incarceration rate graduated symbol map, 2015
(CC BY-SA)

Analyse :

Les cartes à symboles gradués affichent des valeurs quantitatives de manière géolocalisée avec des symboles de tailles différentes, les symboles les plus grands représentant les valeurs supérieures. Ces cartes permettent donc de montrer une différence quantitative entre les différentes zones géographiques représentées. L'une des spécificités de ce type de cartes est qu'il est parfois possible de représenter une variable supplémentaire, pouvant être qualitative, en utilisant une couleur différente sur les symboles représentés (Bertini *et al.*, 2019). Il est ainsi possible de représenter un parti politique, par exemple.

Pour créer une carte à symboles gradués, les données sont d'abord réparties en différentes classes, toutes représentées par une même forme. Il n'est pas nécessaire d'utiliser des cercles ou des carrés, mais ils sont généralement plus simples à comparer (GIS Geography, 2018). Une taille spécifique est ensuite attribuée à chacune de ces classes avant de placer les symboles sur la carte.

Dans l'exemple en figure 95, la légende présente quatre classes différentes, chacune étant associée à un intervalle spécifique qui

représente le taux d'incarcération aux États-Unis. Ces symboles sont repris tels quels sur la carte à symboles gradués.

Forces :

- La taille de la zone géographique n'induit pas la perception de la valeur en erreur (GIS Geography, 2018).
- Plus facile de distinguer les valeurs les plus faibles et les plus élevées que sur une carte avec différentes intensités de couleurs (Bertini *et al.*, 2019).
- Possibilité de représenter une autre variable avec une couleur différente sur les symboles, par exemple un parti politique (Bertini *et al.*, 2019).

Faiblesses :

- Mauvaises interprétations possibles : un symbole deux fois plus grand qu'un autre ne signifie pas que la valeur est deux fois plus grande, car ce sont des classes (Boyes, 2019).
- Difficile de distinguer les valeurs s'il y a trop de classes, et de très grandes valeurs peuvent cacher les autres sur la carte (Boyes, 2019).
- Carte monotone si les valeurs ont toutes quasiment la même taille (Boyes, 2019).

Bibliographie indicative :

Bertini E. and Felix C. (2019), *Information Visualization : Graduated Symbol Maps*. Consulté le 20 avril 2020, à l'adresse <https://fr.coursera.org/lecture/information-visualization-advanced-techniques/graduated-symbol-maps-cihJ2>

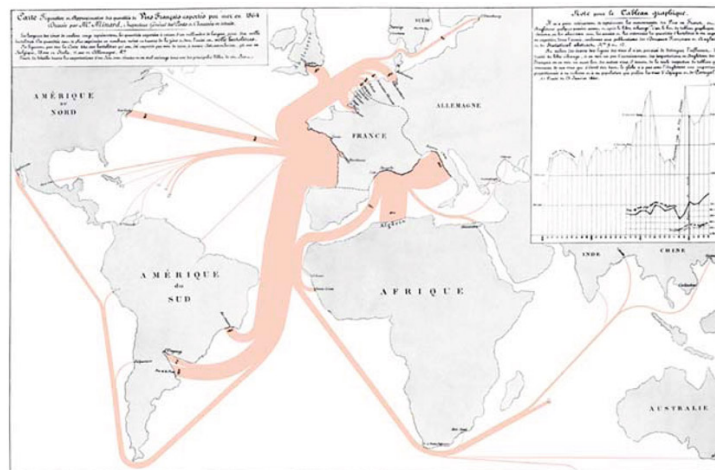
Boyes D. (2019), *GIS Data Acquisition and Map Design : Graduated symbol maps*. Consulté le 20 avril 2020, à

l'adresse <https://www.coursera.org/lecture/gis-data-acquisition-map-design/graduated-symbol-maps-FvB1d>

GIS Geography. (2018). *Dot Distribution vs Graduated Symbols vs Proportional Symbol Maps*. Consulté le 20 avril 2020, à l'adresse <https://gisgeography.com/dot-distribution-graduated-symbols-proportional-symbol-maps/>

La carte de flux

Nom français	Carte des flux
Nom anglais	<i>Flow map</i>
Fonction	Mouvements géolocalisés
Nombre de variables	1
Type(s) de variables	discrète continue
Type(s) de données	nominales ordinales intervalles ratio



Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard, 1845-1860*, a portfolio of his work held by the Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris.

Figure 96 : Minard, 1865, Carte figurative et approximative des quantités de vin français exportés par mer en 1864, (Domaine public)

Analyse :

La carte des flux montre les mouvements d'une variable entre des lieux représentés sur une carte. Des liens indiquent l'itinéraire du flux entre une origine et une destination. La largeur des liens permet d'indiquer une valeur numérique relative au flux (par exemple, le volume ou la fréquence du flux). La couleur des liens permet d'encoder une variable supplémentaire, typiquement une catégorie de flux (Kirk, 2016). L'exemple ci-dessus illustre les exportations de vin français en 1864. On observe des flux importants de la côte atlantique vers l'Amérique du Sud et l'Europe du Nord, ainsi que du bassin méditerranéen vers l'Algérie et l'Italie.

Dans un souci d'intelligibilité, une carte des flux ne devrait représenter qu'une seule variable à la fois (ex. trajet domicile-travail, flux migratoire de population, flux de marchandises). La direction du flux entre son origine et sa destination est indiquée soit par une flèche (représentations statiques), soit par le mouvement (représentations dynamiques). Il est possible de représenter soit des variables discrètes, soit des variables continues.

Forces :

- Fusionner les lignes de flux pour éviter les croisements permet d'éviter l'occlusion et l'encombrement visuel.
- Très intuitif.

Faiblesse :

- Parfois difficile à situer dans l'espace si le concepteur n'affiche pas la carte géographique en arrière-plan ou en filigrane.

Bibliographie indicative :

Kirk, A (2016). Data Visualization : A handbook for data driven design, Thousands Oaks, Sage.

La carte à points

Nom français	Carte à points
Nom anglais	<i>Dot map</i>
Fonction	Distribution géographique
Nombre de variables	1 2
Type(s) de variables	discrète
Type(s) de données	ordinales intervalles ratio

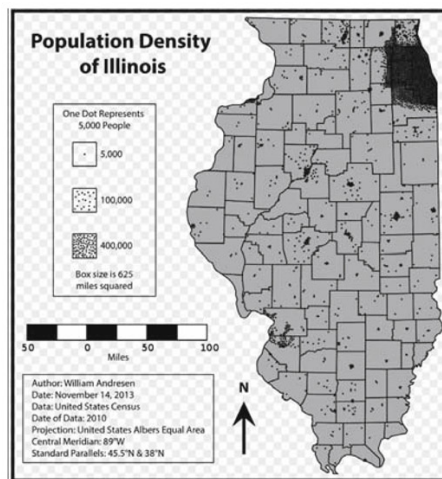


Figure 97 : Dot Density Map of the Population of Illinois, 2013, (CC BY-SA)

Analyse :

Les cartes à points représentent une valeur quantitative pour un espace géographique donné en le remplissant par des points. Tous les points ont une taille identique et correspondent à une même valeur donnée. Il convient au préalable de choisir une donnée unique qui sera donc représentée par ce point (GIS Geography, 2018). Par exemple, un point peut représenter 5 000 habitants. En faisant varier le nombre de points présents dans une même zone géographique, nous pouvons montrer les différences de densité dans les différentes

zones géographiques (Bertin. 1983). Il n'est pas obligé d'utiliser des points : d'autres formes géométriques peuvent également être utilisées telles que des carrés ou des triangles.

Forces :

- Visibilité quasi instantanée des différences entre les zones géographiques.
- Possibilité d'ajouter une variable par l'utilisation de bulles.

Faiblesses :

- Difficultés d'évaluer précisément les quantités dans une même zone géographique.

Bibliographie indicative :

Bertin, J. (1983). *Semiology of graphics; diagrams networks maps* (No. 04 ; QA90, B7.)

GIS Geography. (2018). *Dot Distribution vs Graduated Symbols vs Proportional Symbol Maps*. Consulté le 22 avril 2020, à l'adresse <https://gisgeography.com/dot-distribution-graduated-symbols-proportional-symbol-maps/>

Le cartogramme de Dorling

Nom français	Cartogramme de Dorling
Nom anglais	<i>Dorling Cartogram</i>
Fonction	Distribution géographique
Nombre de variables	3
Type(s) de variables	discrète
Type(s) de données	nominales intervalles ratio



Figure 98 : Obesity rate, 2020, by Bob Rudis • CC-BY-SA-4.0

Analyse :

Le cartogramme de Dorling appartient à la famille des cartes. Il consiste à remplacer les objets géographiques observés par des cercles en évitant les recouvrements et en préservant les relations (Institut National de la Statistique et des Études Économiques, 2018). L'objectif d'une telle méthode est de représenter de manière plus fidèle qu'une carte choroplèthe la réalité d'une situation observée. En effet, en s'affranchissant de la forme réelle des objets et en les remplaçant par des cercles, l'aire de ceux-ci peut incarner directement une variable supplémentaire (Rajerison, 2013).

Un cartogramme de Dorling est une représentation abstraite d'une configuration spatiale, car la forme, la typologie et le centroïde de la figure de départ ne sont pas forcément respectés. Cependant, dans un souci de clarté et pour se rapprocher au maximum de la configuration géographique réelle, ce type de visualisation tente de limiter au maximum la distance entre la position finale de chaque objet et sa position d'origine (Dorling Cartogram, s.d.).

Forces :

- Représentation plus fidèle et efficace d'une situation par rapport à une carte choroplèthe.

- Permet de visualiser aisément des proportions tout en conservant les relations entre objets.
- Permet d'incarner une 3^e (voir une 4^e) variable.

Faiblesse :

- Ne conserve pas la forme initiale de la zone géographique observée, ce qui peut causer une difficulté à identifier celle-ci.

Bibliographie indicative :

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A Tour Through the Visualization Zoo. *Graphics*, 8(5), 59-67.

Institut National de la Statistique et des Études Économiques. (2018). Guide de sémiologie cartographique. Consulté à l'adresse https://www.union-habitat.org/sites/default/files/articles/pdf/2018-11/guide_insee.pdf

Rajerison, M. (2013, juillet 28). Un petit peu de « geohacking » et de scraping, da (o) rling?. Consulté à l'adresse <https://datagistips.hypotheses.org/62>

La carte de connexions

Nom français	Carte de connexions
Nom anglais	<i>Connection map</i>
Fonction	Relations géolocalisées
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	nominales ordinales



Figure 99 : International route map of Philippine Airlines as of October 2013. Based on the route map in the Mabuhay magazine, 2013 (CC BY-SA)

Analyse :

Une carte de connexions, ou *connection map* en anglais, montre les connexions entre plusieurs points sur une carte. Le lien entre deux localisations se fait généralement avec une ligne droite ou courbée (Data Visualisation Catalogue, s.d.). On l'utilise le plus souvent pour afficher les itinéraires de vols (Data visualization & Human rights, s.d.). En comparaison à la carte des flux qui montre le mouvement d'une variable d'un lieu à un autre, les liens de la carte des connexions n'est représenté ni par des flèches directionnelles ni par le changement d'épaisseur variable des connecteurs (Data visualization & Human rights, s.d.). Au lieu de montrer les mouvements, elle met simplement en évidence les liens qui se font d'un endroit à l'autre (Data Visualisation Catalogue, s.d.). Il est toutefois important de souligner qu'à ce jour, aucune littérature scientifique ne renseigne à ce sujet. Les cartes de connexions sont le plus souvent utilisées dans les publications journalistiques.

Forces :

- Utile pour identifier des tendances concernant la répartition des connexions géographique. (« Connection map », n. d.).

- Utile pour identifier une concentration à travers les connexions.

Faiblesses :

- Difficultés de déterminer les connexions quand un trop grand nombre se chevauche.
- Possible confusion comme étant un type de carte de flux.
- Aucune documentation officielle à son sujet.

Bibliographie indicative :

Data Visualisation Catalogue. (s.d.). *Connection map*. Retrieved from https://datavizcatalogue.com/methods/connection_map.html

Data visualization & Human rights. (s.d.). *Connection map*. Retrieved from <http://visualizingrights.org/kit/charts/connection-map.html>

La carte de chaleur

Nom français	Carte de chaleur
Nom anglais	<i>Heat map</i>
Fonction	Distribution géographique
Nombre de variables	1
Type(s) de variables	discrète continue
Type(s) de données	ordinales intervalles ratio



Figure 100 : Heatmap, 2010, par City University Interaction Lab (CC BY-SA)

Analyse :

La carte de chaleur est une représentation de données statistiques qui utilise un spectre coloré faisant correspondre l'intensité d'une grandeur variable à une gamme de tons ou à des nuances de couleurs sur une jauge opposant deux tons (Köpp *et al.*, 2014). Les valeurs statistiques sont donc représentées par ce spectre qui peut chevaucher une carte de type géographique ou tout autre support visuel sur lequel des données de localisation ont été prélevées (photo, site web, etc.).

Notre exemple montre une carte de chaleur représentant un site web sur lequel des données de regard ont été prélevées. Plus le spectre est rouge, plus cette zone de l'écran a été observée. On constate ainsi une densité plus forte à cet endroit, grâce à une légende de couleurs partant du bleu-vert au rouge.

Forces :

- Les zones où l'activité est importante facile à repérer.

Faiblesses :

- Ne fonctionne que sur des dispositifs statiques.

Bibliographie indicative :

Köpp, C., von Mettenheim, H.-J., & Breitner, M. H. (2014). Decision Analytics with Heatmap Visualization for Multi-step Ensemble Data : An Application of Uncertainty Modeling to Historical Consistent Neural Network and Other Forecasts. *Business & Information Systems Engineering*, 6(3), 131-140. <https://doi.org/10.1007/s12599-014-0326-4>

3.4 Hiérarchies et réseaux

L'arbre

Nom français	Arbre
Nom anglais	Tree
Fonction	Hiérarchie
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	nominales ordinales intervalles ratio

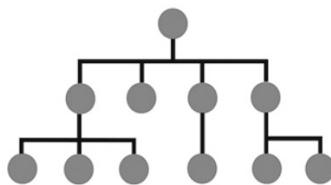


Figure 101 : Exemple d'arbre, par les auteurs

Analyse

Un arbre est un réseau dans lequel deux nœuds sont connectés par exactement un chemin. L'arbre est « enraciné » : il y a une « racine » qui est à la tête de la hiérarchie. On parle de nœud « parent » lorsqu'un nœud en précède un autre plus haut dans la hiérarchie, de nœud « enfant » lorsqu'un nœud succède à un autre plus bas dans la hiérarchie (Wilkinson, 2008). Les extrémités de l'arbre sont appelées les feuilles. Une feuille n'a pas de nœud « enfant ».

L'exemple ci-dessus est un dessin classique d'arbre, aussi appelé « hiérarchie ». Il y a de nombreuses manières de visualiser les hiérarchies. C'est pourquoi de nombreux affichages ont été conçus (Heer *et al.*, 2010).

Forces :

- Vision claire et précise de la hiérarchie.
- Sa forme la plus classique, l'organigramme, est bien connue du grand public.

Faiblesse : N/A.

Bibliographie indicative :

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, 53(6), 59. <https://doi.org/10.1145/1743546.1743567>

Wilkinson, L. (2008), « Graph-theoretic graphics », *Handbook of Data Visualization*, Springer Berlin Heidelberg, pp. 121-150

L'arbre hyperbolique

Nom français	Arbre hyperbolique (layout)
Nom anglais	<i>Hyperbolic tree (layout)</i>
Fonction	Hiérarchie
Nombre de variables	1
Type(s) de variables	discrète continue
Type(s) de données	nominales ordinales intervalles ratio

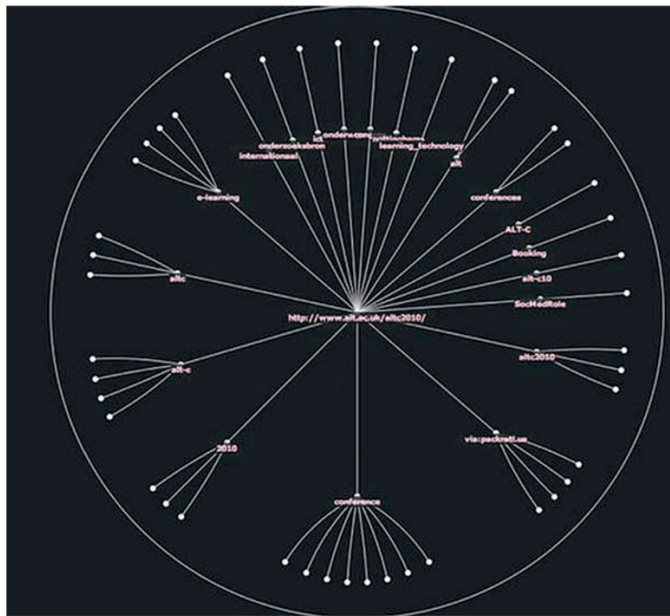


Figure 102 : Hyperbolic Tree, 2010, by AJ Hirst (CC BY 3.0)

Analyse :

Comme son nom l'indique, l'arbre hyperbolique est un arbre (ou hiérarchie), à savoir un réseau dans lequel deux nœuds sont connectés par exactement un chemin. La racine de l'arbre est située au centre du diagramme circulaire. Les nœuds « enfants » sont disposés en cercle de façon exponentielle à mesure que la distance par rapport à la racine augmente. Un arbre hyperbolique peut contenir beaucoup plus d'informations dans la même quantité

d'espace qu'un arbre hiérarchique standard (Urribarri *et al.*, 2013). Par rapport à un arbre classique, rien ne change donc, si ce n'est la disposition des nœuds.

Forces :

- Possibilité de visualiser de très grands arbres.
- Gain d'espace considérable.

Faiblesse :

- Difficilement saisissable en un coup d'œil.

Bibliographie indicative :

Urribarri, D. K., Castro, S. M., & Martig, S. (2013). Gyrolayout: A Hyperbolic Level-of-Detail Tree Layout. *J. UCS*, 19 (1), 132-156.

Le sunburst

Nom français	Sunburst
Nom anglais	<i>Sunburst Diagram</i>
Fonction	Hiérarchie
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	nominales ordinales intervalles ratio



Figure 103 : Exemple de sunburst, par les auteurs

Analyse :

Le sunburst est une des nombreuses façons de visualiser un arbre. Pour rappel, un arbre est un réseau dans lequel deux nœuds sont connectés par exactement un chemin. La hiérarchie est ici montrée à travers une série d'anneaux qui sont subdivisés pour illustrer les nœuds. Chaque anneau correspond à un niveau de la hiérarchie, le cercle central représentant le nœud racine et la hiérarchie s'en éloignant. La taille de chaque « tranche » est ainsi proportionnelle à la taille du nœud et à sa valeur. Les anneaux sont divisés en fonction de leur relation hiérarchique avec la tranche « parent ». Différents groupes ou catégories peuvent être codés sur les variables visuelles grâce à la couleur des tranches (Kirk, 2016).

Forces :

- Attrait visuel.
- Catégories clairement différenciables.

Faiblesses :

- Confusion autour de la perception des proportions même pour un utilisateur averti.
- Peu pertinent pour afficher un grand ensemble de données, car la hiérarchie n'est pas perceptible en un seul coup d'œil

(*overview*), ce qui oblige l'audience à explorer chaque tranche pour comprendre hiérarchie et proportions.

- Malgré son aspect simple, il arrive facilement qu'on oublie qu'il s'agit bel et bien d'un arbre (hiérarchie), et non d'un diagramme classique.

Bibliographie indicative :

Kirk, Andy (2016.). *Data Visualization : A handbook for data driven design* (Sage). Thousands Oaks.

Liu, C., & Wang, P. (2015). A Sunburst-based hierarchical information visualization method and its application in public opinion analysis. 2015 8th International Conference on Biomedical Engineering and Informatics (BMEI), 832-836.

<https://doi.org/10.1109/BMEI.2015.7401618>

Le treemap

Nom français	Treemap
Nom anglais	<i>Treemap</i>
Fonction	Hiérarchie
Nombre de variables	2 3
Type(s) de variables	discrète continue
Type(s) de données	ordinales intervalles ratio



Figure 104 : Gradient grouped treemap, 2010 (Domaine public)

Analyse :

Le treemap est une manière de visualiser un arbre, en y incluant une idée de quantité. Il s'agit d'une série de rectangles disposée dans un espace défini. La taille des rectangles est significative des dimensions d'intérêt représentées. Chaque catégorie se voit attribuer une zone rectangulaire avec leurs sous-catégories rectangles imbriqués à l'intérieur (Turo & Johnson, 1992). De plus, la superficie de la catégorie mère est le total de ses sous-catégories. La couleur des rectangles peut également supporter une variable. Il permet de disposer des données hiérarchiques en utilisant des rectangles pour représenter les branches d'un arbre. Il s'agit donc bel et bien d'un graphe hiérarchique – un arbre – dont la hiérarchie entre les nœuds est représentée par la taille des cases employées (Day & Dickinson, 2014). Cette méthode permet également d'inclure des cases dans un même rectangle, pour signifier des groupes. Elle a été initialement imaginée pour visualiser l'occupation de l'espace sur les disques durs des ordinateurs, mais peut être utilisée à d'autres fins sans problème.

Force :

- Très utile pour comparer les proportions entre les catégories d'un arbre en fonction de la taille de leur case.

Faiblesse :

- N'affiche pas les niveaux hiérarchiques aussi clairement que d'autres visualisations d'arbres.

Bibliographie indicative :

Day L. & Dickinson R. (2014), « Treemap Presentation as a Corporate Dashboard », AD07.Day, [PDF en ligne] (consulté le 22/02/17), <https://www.cs.umd.edu/hcil/treemap-history/BNSF-AD07.Day.pdf>

Turo, D., & Johnson, B. (1992, October). Improving the visualization of hierarchies with treemaps: design issues and experimentation. In Proceedings Visualization'92 (pp. 124-131). IEEE.

Le diagramme de nœuds-liens radial

Nom français	Diagramme de nœuds-liens radial
Nom anglais	<i>Radial node-link diagram</i>
Fonction	Hiérarchie
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	nominales ordinales intervalles ratio

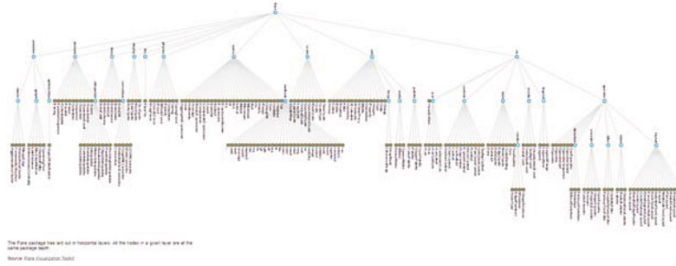


Figure 105 : Exemple dans Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, 53(6), p. 64 : Figure 4a. Radial node-link diagram of the Flare package hierarchy.

Analyse :

Le diagramme nœud-lien radial est une visualisation se basant sur le modèle classique de représentation sous forme d'arbre. L'exemple ci-dessus issu du *Flare Visualization Toolkit* montre un diagramme nœuds-liens radial, où les différentes données composant l'outil de visualisation Flare (nommées dans l'exemple « Flare Package ») sont réparties hiérarchiquement.

Ce diagramme présente un réseau dans lequel plusieurs nœuds sont connectés par un seul chemin entre eux, qui peuvent prendre la forme d'une courbe. La « racine » de l'arbre se caractérise par le point situé le plus haut sur la représentation (en suivant ici dans notre exemple l'axe vertical), il s'agit de la tête de la hiérarchie et c'est donc de ce point que partent les liens.

Forces :

- Peut permettre de visualiser de très grandes hiérarchies de données.
- Donne une bonne vision d'ensemble de la hiérarchie des données.

Faiblesses :

- Les premiers niveaux de hiérarchie sont forts espacés et lisibles, tandis que les niveaux les plus bas sont très resserrés et bien moins lisibles.
- Peut rapidement devenir illisible et il est alors difficile d'identifier chaque donnée individuellement sans passer par le *zoom*.

Bibliographie indicative :

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, 53(6), 64.

Le réseau

Nom français	Réseau
Nom anglais	<i>Cartesian node-link diagram</i>
Fonction	Relations
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	nominales ordinales intervalles ratio

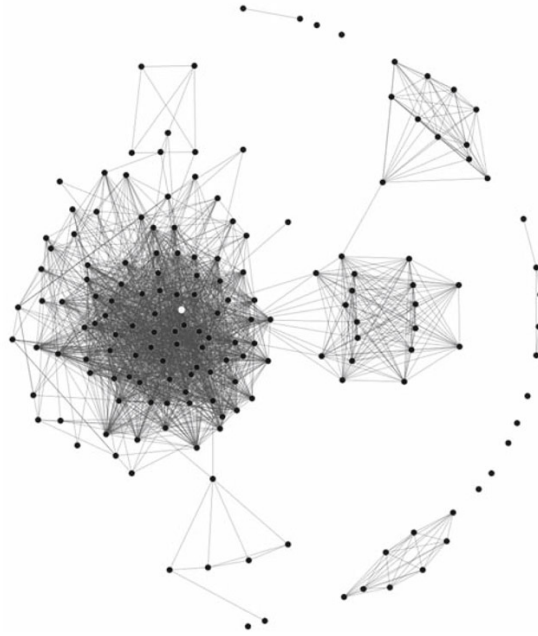


Figure 106 : Social Network Diagram, 2014 (CC BY-SA)

Analyse

Les réseaux sont des ensembles de points (sommets) et de lignes (arcs) qui lient certains couples de points. Un ensemble de nœuds connectés par des liens identifie la position des nœuds de manière générale dans l'espace. Le réseau peut être orienté à l'aide de flèches (une direction est indiquée sur les lignes) ou non. Un réseau non orienté est dit symétrique, car les nœuds se distribuent uniformément dans l'espace (Beauguitte, 2010). Un réseau est donc tout simplement un ensemble de points reliés entre eux par des lignes. Pour que des groupes (*clusters*) puissent être identifiés, c'est sur leur aspect visuel qu'il faut donc jouer. Pour cela, des *layouts* existent. Il s'agit d'algorithmes qui permettent de déterminer le tracé du graphe. Ainsi, on choisit un *layout* dans un programme de visualisation de réseaux.

Forces :

- Possibilité de visualiser de grands jeux de données.
- Identification claire de *clusters*.

- Attrait visuel.

Faiblesse :

- Pour les très grands jeux de données, de l'occlusion peut apparaître entre les liens et les rendre moins visibles.

Bibliographie indicative :

Beauguitte, L. (2010). Graphes, réseaux, réseaux sociaux : Vocabulaire et notation. <https://halshs.archives-ouvertes.fr/halshs-00541898/document>

Le diagramme à cordes

Nom français	Diagramme à cordes
Nom anglais	<i>Chord diagram</i>
Fonction	Relations
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	nominales ordinales intervalles ratio

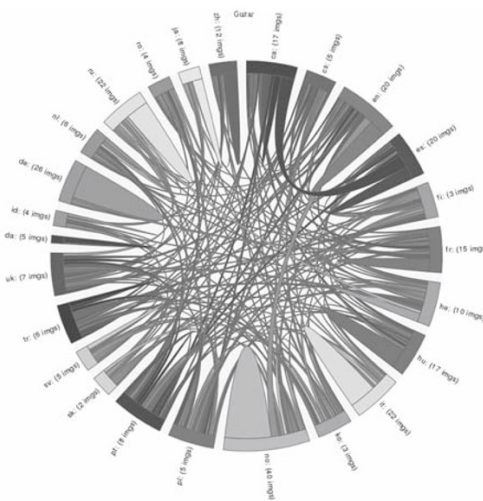


Figure 107 : Image Dive guitar chord diagram, 2018 (CC BY-SA)

Analyse

Le diagramme à cordes est un réseau qui affiche les relations (liens) entre les catégories ou modalités (nœuds) d'une variable. Pour rappel, un réseau est un ensemble de nœuds (sommets) et de liens (arcs) entre deux nœuds. Le diagramme à cordes est de forme circulaire. Les nœuds représentent des catégories et sont affichés sur la circonférence du cercle, par « tranches ». La taille de ces tranches est proportionnelle à la taille des nœuds. Les relations entre les catégories sont signifiées grâce à des arcs (Iturbe & al., 2016).

Les relations entre catégories sont utilisées pour montrer qu'elles ont quelque chose en commun. Le diagramme à cordes permet de comparer les similitudes au sein d'un ensemble de données ou entre différents groupes de données. La taille des arcs est également porteuse d'information, comme la fréquence de la relation (Kirk, 2016).

L'exemple sur la figure 107 montre l'utilisation de l'image à travers les éditions linguistiques pour l'article anglais sur « *Guitar* ».

Force :

- Idéal pour comparer les points communs au sein d'un ensemble de données ou entre différents groupes de données.

Faiblesse :

- Lisibilité du diagramme influencée par le nombre de données (nœuds et liens) : un trop grand nombre de liens encombre le diagramme (l'interactivité peut néanmoins aider à contourner ce problème).

Bibliographie indicative :

Iturbe, M., Garitano, I., Zurutuza, U., & Uribeetxeberria, R. (2016, February). Visualizing Network Flows and Related Anomalies in Industrial Networks using Chord Diagrams and Whitelisting. In VISIGRAPP (2: IVAPP) (pp. 101-108).

Kirk, Andy (2016.). Data Visualization : A handbook for data driven design (Sage). Thousands Oaks.

L'algorithme par modèle de forces

Nom français	Algorithme par modèle de forces (Système masse-ressort)
Nom anglais	<i>Forced-directed layout</i>
Fonction	Relations
Nombre de variables	1
Type(s) de variables	discrète
Type(s) de données	nominales ordinales

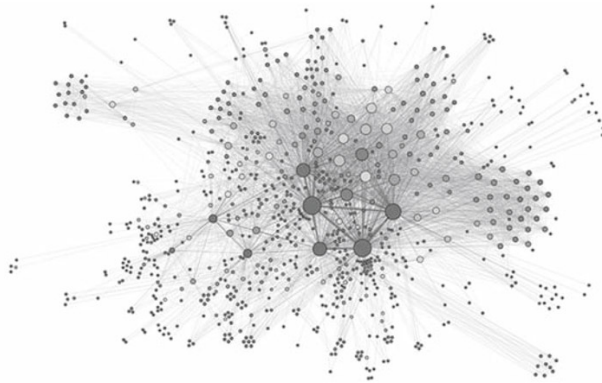


Figure 108 : Social network visualization. Graph containing 800 vertices and 10 000 edges, 2016 (CC BY-SA)

Analyse :

L'algorithme par modèle de forces permet de visualiser des réseaux exprimés par des sommets reliés entre eux par des arêtes. Ce type de graphe est utilisé notamment pour analyser des réseaux sociaux, pour faire de la cartographie, etc. (Wikipedia, 2020). Les algorithmes par modèle de force emploient un système masse-ressort. Ce système s'exprime par des actions exercées sur les sommets et les arêtes. Ces actions sont identifiées par des forces d'attraction qui associent un ressort à chaque arête et des forces de répulsion entre chaque paire de sommets. Il se peut que des forces mineures interviennent également. Ces forces mineures sont exprimées par exemple sous la forme d'une gravité où tous les sommets sont attirés vers le centre du graphe (Nadal & Melançon, 2012).

Forces :

- Identifier des nœuds importants dans un tout.
- Traite beaucoup de données d'une variable.
- Esthétiquement beau et facile à comprendre.

Faiblesses :

- Graphe dépendant d'un algorithme de calcul.

- Perd en visibilité si le nombre de données est trop élevé.

Bibliographie indicative :

Nadal, M. ; Melançon, G. ; Dessin de graphe assisté par un algorithme génétique, 9^e édition de la conférence MAnifestation des JEunes Chercheurs en Sciences et Technologies de l'Information et de la Communication – MajecSTIC 2012 (2012), Nicolas Gouvy, Oct 2012, Villeneuve d'Ascq, France. hal-00780201

La carte de chaleur en classe hiérarchique²

Nom français	Carte de chaleur en classe hiérarchique
Nom anglais	<i>Cluster heat map</i>
Fonction	Hiérarchie
Nombre de variables	1 2
Type(s) de variables	discrète continue
Type(s) de données	ordinales intervalles ratio

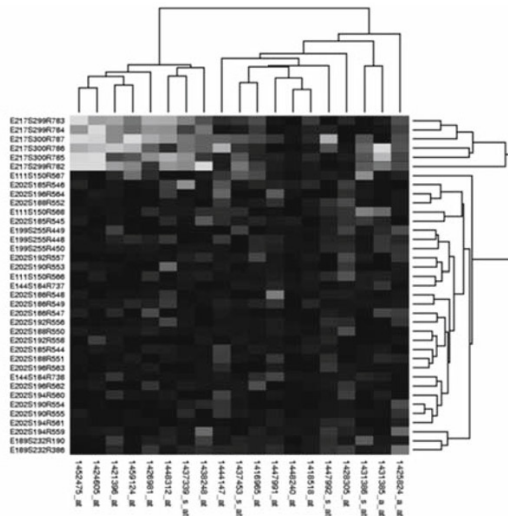


Figure 109 : Cluster heatmap, 2006 (Domaine public)

Analyse :

L'exemple ci-dessus est une carte de chaleur de microréseau (séquence ADN), les lignes représentent des gènes et les colonnes représentent des échantillons. Chaque cellule est colorisée en fonction du niveau d'expression de ce gène dans cet échantillon (Wilkinson & Friendly, 2009).

La carte de chaleur en classe hiérarchique est une représentation de données statistiques qui utilise une échelle de couleur pour représenter les valeurs des données. Cette carte permet d'identifier les valeurs similaires, car ces dernières sont affichées sous la forme de zones qui ont des couleurs similaires.

Les variables dans les colonnes sont numériques, ce sont celles qu'on aura tendance à regrouper. Ces variables doivent contenir des valeurs numériques à partir d'une échelle linéaire. Les exemples incluent le score du test, le nombre, la taille, le poids, l'âge, la température, etc. Tandis que les variables dans les lignes sont souvent des étiquettes pour faciliter l'interprétation (NCSS Statistical Software, s.d.).

Force :

- Facile de repérer les groupes de valeurs similaires.

Faiblesse : N/A**Bibliographie indicative :**

NCSS Statistical Software. Clustered Heat Maps. Consulté le 13 mai 2020, à l'adresse https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Clustered_Heat_Maps-Double_Dendrograms.pdf

Wilkinson, L., & Friendly, M. (2009). The history of the cluster heat map. *The American Statistician*, 63(2), 179-184.

Les coordonnées parallèles

Nom français	Coordonnées Parallèles
Nom anglais	<i>Parallel coordinate plot</i>
Fonction	Relations
Nombre de variables	N
Type(s) de variables	discrète continue
Type(s) de données	ordinales intervalles ratio

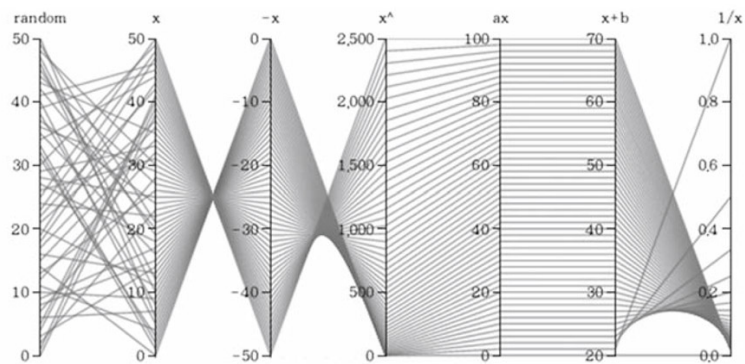


Figure 110 : Representative sample for parallel coordinates, 2015 (CC BY-SA)

Analyse :

Les coordonnées parallèles sont idéales pour montrer les relations entre les données, car elles permettent d'afficher des données multivariées. Chaque axe correspond à une variable. Les axes sont placés parallèlement les uns aux autres et n'ont pas forcément la même échelle puisque chaque variable peut fonctionner sur une unité de mesure différente. Effectivement, chaque élément multivarié « correspond à une polyligne traversant les axes » (Tuor *et al.*, 2016, p. 300).

Les données sont représentées par une série de lignes qui se connectent sur tous les axes. Les axes peuvent être placés à

l'horizontale (comme dans cet exemple) ou de manière verticale. Grâce aux coordonnées parallèles, il est possible de repérer visuellement des schémas ou modèles (patterns), ce qui est un point fort en analyse exploratoire (Tuor *et al.*, 2016, p. 300). Les coordonnées parallèles permettent de voir les liens et relations entre les différentes données multivariées représentées. Chaque segment de ligne représente l'enregistrement de la donnée dans des catégories différentes. Au plus des données continues seront ajoutées, au plus une « forme » schématique apparaîtra (Kirk, 2016).

Forces :

- Possibilité d'afficher un très grand nombre de données.
- Clarté dans l'affichage des relations et des schémas émergents.
- C'est l'une des seules visualisations de données qui permet d'afficher différentes échelles de manière aussi précise et sans confusion.

Faiblesses :

- Lorsque le nombre de polygones est trop grand, les segments s'enchevêtrent et se superposent. Les tendances sont alors difficilement identifiables dans la visualisation (Tuor *et al.*, 2016).
- Bien que cela soit d'usage, représenter des données discrètes engendre une perturbation visuelle, puisque ce type de visualisation n'est pas réellement adapté aux dimensions catégorielles (Tuor *et al.*, 2016).
- Attention, le terme « *plot* » en anglais pour « *parallel coordinates plot* » est un faux-ami, car ce graphique ne prend pas la forme d'un graphique de relations traditionnel avec deux axes (x, y).
- Nécessite un public averti.

Bibliographie indicative :

Kirk, A. (2016), *Data Visualization : A handbook for data driven design*, Sage, Thousands Oaks.

Tuor R., Evéquoz F., Lalanne D. (2016), *Parallel Bubbles : Visualisation de données catégorielles dans des Coordonnées Parallèles*. Actes de la 28^e conférence francophone sur l'Interaction Homme-Machine, Fribourg, Suisse. pp. 299-306, 10.1145/3004107.3004142.

Le diagramme de Venn

Nom français	Diagramme de Venn
Nom anglais	<i>Venn Diagram</i>
Fonction	Relations
Nombre de variables	2 3
Type(s) de variables	discrète
Type(s) de données	nominales ordinales intervalles ratio

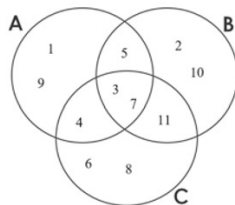


Figure 111 : Exemple de diagramme de Venn, par les auteurs

Analyse :

Le diagramme de Venn permet de montrer les relations entre plusieurs variables, typiquement entre trois variables. Chaque variable est représentée par un cercle qui contient une collection d'entités qui ont hypothétiquement quelque chose en commun. Ce

sont les intersections entre les cercles, i.e., les zones où les cercles se chevauchent, qui permettent de vérifier si oui ou non il existe des relations entre les variables. Les entités qui apparaissent aux intersections correspondent aux éléments que les variables ont en commun (Coumet, 1965). Il est aussi possible de réaliser un diagramme de Venn pour seulement deux variables.

L'exemple sur la figure 111 montre un diagramme de Venn à trois variables. On observe que les variables A, B et C ont en commun les chiffres 3 et 7, les variables A et B le chiffre 5.

Force :

- Peut afficher des données numériques et des données textuelles.

Faiblesse :

- Contrairement à ce qui est préconisé, la lecture du diagramme de Venn devient difficile au-delà de trois variables/

Bibliographie indicative :

Coumet, E. (1965). Les diagrammes de Venn. *Mathématiques et Sciences humaines*, 10, 31-46.

Résumé

Pour résumer, avant de se lancer dans la conception d'une visualisation de données, il est indispensable de réfléchir au type de variables et au type de données dont on dispose. En effet, il est fondamental de se pencher sur la nature de l'information à communiquer plutôt que sur sa potentielle apparence.

Une variable peut être discrète, c'est-à-dire prendre un nombre fini ou dénombrable de valeurs, ou bien elle peut être continue. Dans ce cas, elle peut prendre un nombre infini ou indénombrable de valeurs.

Les données peuvent être nominales, ordinales, d'intervalle ou de ratio. L'acronyme NOIR de Stevens (1946) permet de s'en souvenir facilement.

Lorsque l'on connaît bien les variables et les données dont on dispose et que l'on est en mesure d'imaginer un message à communiquer, il est temps de choisir une technique de visualisation. Il existe de nombreuses techniques de visualisation. Nous en avons répertorié 43 dans le Catalogue présent dans ce chapitre.

Les techniques de visualisation peuvent être classées en familles, comme ça l'est présenté dans le « Zoo de la Visualisation » selon Heer et ses collègues (2010). Dans notre catalogue, les techniques de visualisation se répartissent entre différentes familles que sont les visualisations statistiques, les visualisations temporelles, les cartes, les hiérarchies et réseaux.

Après avoir choisi une technique de visualisation, le concepteur peut se préoccuper d'en imaginer le design. Il est régi par des règles et principes qui font l'objet du chapitre suivant.

Ressources externes et bibliographie indicative

Ressources externes

Statistiques : Ebook gratuit en ligne accompagné de vidéos explicatives :
How2stats, par Gilles Gignac : <http://www.how2statsbook.com/>.

Bibliographie indicative

Heer, J., Bostock, M., & Ogievetsky, V. (2010), A tour through the visualization zoo, *Communications of the ACM*, 53(6), 59.
<https://doi.org/10.1145/1743546.1743567>

Kirk, A. (2016), *Data visualisation: a handbook for data driven design*. Sage

OHowell, D., Bestgen, Y., Yzerbyt, V., & Rogier, M. (2008), *Méthodes statistiques en sciences humaines*, De Boeck, Bruxelles.

Ware, C. (2004). *Information visualization: perception for design*. San Francisco, CA: Morgan Kaufman

CHAPITRE 4

PRINCIPES DE CONCEPTION DES VISUALISATIONS DE DONNÉES

1. Efficacité et excellence graphique
2. Les variables visuelles
3. L'intégrité graphique
4. Visualiser en simplicité
5. Évitez le *chartjunk*
6. La reconnaissance de patterns
7. La couleur

Évitez les faux pas de la dataviz !

Enfin, ça y est : après avoir analysé vos données, vous avez décidé d'un message à communiquer et vous avez choisi une technique de visualisation. Vous allez pouvoir vous attarder aux détails visuels : légendes, étiquettes de données, couleurs, etc. Cet « enrobage » visuel est une étape souvent plaisante. Malheureusement, dans l'enthousiasme, les créateurs de visualisation de données en font parfois trop. Comment savoir si les choix de conception sont les bons ? Comment faire pour que la visualisation soit le plus compréhensible possible ? Les différents principes de conception présentés dans ce chapitre répondront à vos questions.

Pendant longtemps, les innovations graphiques ont manqué d'un vocabulaire, de règles communes, jusqu'à ce que Jacques Bertin, cartographe français du xx^e siècle, rende accessible cette grammaire graphique en 1967, grâce à son manifeste *Sémiologie Graphique*. Il y aborde les différents types de construction graphique, leur grammaire ainsi que les variables visuelles permettant de coder de l'information sur une image en deux dimensions. À partir des années 80, le Professeur américain Edward Tufte publie lui aussi de nombreux principes de design graphique et de communication visuelle, visant à rendre la visualisation d'information claire, compréhensible, élégante

et surtout fidèle aux données (Tufté, 1983). D'autres chercheurs contemporains comme Ben Shneidermann ou encore Stephen Few ont également enrichi le domaine qui constitue désormais une discipline à part entière.

1. Efficacité et excellence graphique

1.1 L'efficacité selon Jacques Bertin (1967)

Jacques Bertin est un cartographe français. Son apport aux principes de conception des visualisations de données est considérable. En 1967, il écrit la *Sémiologie graphique*, un manifeste décrit comme une étude des bases de la représentation graphique. À l'époque, les visualisations de données s'imprimaient encore, étant alors des images de communication. À la lecture, l'ouvrage peut sembler daté. Il regorge pourtant de bons conseils en termes de conception graphique.

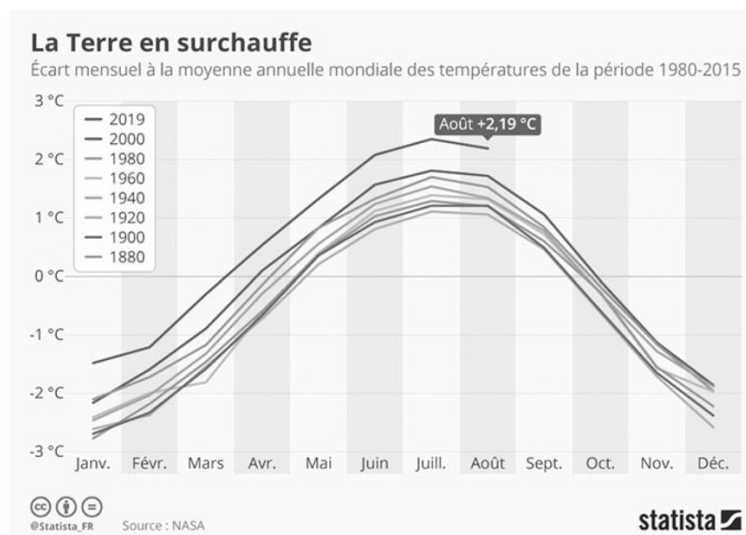


Figure 112 : La Terre en surchauffe, par Tristan Gaudiaut, 2019 (CC0)

Pour Bertin, il était important qu'un graphique soit compréhensible et perceptible dans un instant minimum de vision. Pour lui, lorsqu'un individu lit un graphique, il concrétise d'abord dans sa pensée les variables du graphique et l'invariant, puis il passe à travers différentes étapes du processus de lecture :

1. L'identification externe : repose sur les habitudes acquises comme la reconnaissance de mots, de formes et de couleurs.
2. L'identification interne : le lecteur reconnaît comment sont encodées les différentes variables.
3. La perception des correspondances originales : le lecteur comprend le croisement des variables, ce qui l'amène à formuler des questions consciemment ou inconsciemment. Par exemple : « quelle était la température moyenne au cours du mois de février en 2000 ? » (figure 112).

Les questions rendues possibles par la perception des correspondances originales sont en nombre fini : on peut donc les définir :

1. Les questions induites par un seul élément de la variable : « en mars, quelle était la température moyenne en 2000 ? » C'est le niveau élémentaire de lecture : le lecteur en tire des informations élémentaires.
2. Les questions induites par un groupe d'éléments ou de catégories : « durant les trois premiers mois de l'année, quelle était l'évolution de la température ? » Ce sont les niveaux moyens de lecture.
3. Les questions induites par l'ensemble du graphique : « durant toute cette période, quelle a été l'évolution de la température ? » C'est le niveau supérieur, la lecture d'ensemble,

qui réduit toute l'information en une relation d'ordre entre les variables.

En ciblant toutes les différentes questions auxquelles une visualisation de données peut répondre, on peut définir les objectifs que l'information vise.

C'est par rapport aux questions que l'on peut se poser lors de la lecture d'un graphique – et surtout par rapport aux réponses que l'on trouve – que Bertin définit **l'efficacité graphique**.

Si pour obtenir une réponse correcte et complète à une question donnée, et toutes choses égales, une construction requiert un temps d'observation plus court qu'une autre construction, on dira qu'elle est plus efficace pour cette question.

(Bertin, 1967, p. 139)

En cela, Bertin applique la notion de coût mental (Zipf, 1935) à la perception visuelle. La différence entre une construction efficace et inefficace pour Bertin est alors très nette : elle se mesure par le temps dont l'individu a besoin pour répondre aux questions qu'il se pose (consciemment ou inconsciemment) et atteindre la compréhension. Il lie ainsi l'efficacité aux opérations mentales nécessaires lors de la lecture d'un dessin.

Ainsi, « les constructions les plus efficaces sont celles dans lesquelles toute question, quel qu'en soit le type et le niveau, obtient une réponse dans un seul instant de perception, une réponse perceptible en une seule image » (Bertin, 1967, p. 146).

C'est l'utilisation du système graphique de signes propre à la visualisation de données qui rend les représentations graphiques efficace. Ces signes sont pour la plupart arbitraires et permettent à l'homme d'intégrer des relations complexes entre des variables en quelques instants.

Évitez de confondre !

Attention ! Si vous lisez la sémiologie graphique aujourd'hui, vous pourriez être confus : Bertin parle d'invariant, de composantes et des variables (visuelles ou rétinienne). Avec le temps, le langage a évolué et ces mots ne sont plus ceux que l'on emploie aujourd'hui.

L'invariant est la partie de l'information qui reste fixe sur toute l'étendue du graphique, c'est « la définition complète et commune à toutes les données » (Bertin, 1967, p. 16)

Les composantes sont les concepts de variation. Aujourd'hui, on les appelle tout simplement « variables ».

Les variables rétinienne (plus tard appelées variables visuelles) correspondent selon Bertin à ce que nous considérons aujourd'hui comme des techniques d'encodage visuel.

Nous veillerons donc à employer les termes actuels... Autant ne pas s'emmêler les pinces !

1.2 L'excellence graphique

Edward Tufte, professeur en design d'information à l'université de Yale, a publié plusieurs ouvrages dans lesquels il aborde le design d'information. Dans celui qui est sans doute sa plus grande contribution, *The visual display of quantitative information*, il définit ce qu'il appelle « l'excellence graphique ». En se résumant en trois principes, l'excellence graphique...

1. est la présentation bien conçue de données intéressantes. C'est une question de fond, de statistiques et de design.
2. consiste en la communication d'idées complexes menée avec clarté, précision et efficacité.
3. consiste à donner au lecteur la communication du **plus grand nombre d'informations, dans le temps le plus court, dans l'espace le plus petit, et avec le moins d'encre possible.**

(Tufte, 2001, p. 51, traduction libre)

Quant à Tufte, il présente **l'excellence graphique** comme atteignant un haut niveau de qualité, parce que la représentation doit absolument révéler les données. Le graphique en lui-même doit être d'une telle efficacité que le récepteur pense d'abord à sa substance (ses données) sans s'inquiéter de la méthodologie déployée pour mettre le graphique lui-même en place (Tufte, 1983).

L'excellence graphique nécessite de dire la vérité à propos des données.

(Tufte, 2001, p. 51)

Pour résumer, l'excellence graphique, c'est

- Montrer les données de manière claire, dans leur contexte et sans les déformer ;
- Montrer une grande quantité de données de manière concise tout en restant cohérent ;
- Stimuler l'exploration et l'analyse des données en incitant l'audience à réfléchir au fond plutôt qu'à la forme. On peut vulgariser ce dernier point comme suit : si l'audience ne se concentre pas sur les données, mais essaye de comprendre les intentions du designer, c'est raté...

1.3 Efficacité, excellence... Une vision fonctionnaliste ?

L'efficacité est plutôt traduite chez Bertin en termes de fonctionnement : un graphique va mieux « fonctionner » qu'un autre. Bertin est, dans sa vision du terme, très tourné vers le récepteur qui cherche à répondre à une question. Il y a donc un certain aspect « utilisateur », mais pourtant, il ne se concentre pas sur les caractéristiques personnelles du lecteur. Bertin cherche à agir sur le graphique dans l'idée que tout lecteur saura se poser suffisamment de questions et y répondre, sans forcément que d'autres facteurs (contexte, caractéristiques personnelles, compétences) n'entrent en compte dans cette interaction.

L'idée d'excellence graphique de Tufte peut être rapprochée à l'idée d'efficacité de Bertin : « le critère de l'efficacité que Bertin nous présente en mettant l'accent sur la dimension temporelle de l'observation n'est pas loin de l'excellence graphique promulguée par le gourou américain du design graphique » (Sung-Do Kim, 2008, p. 33). En exprimant que l'excellence graphique doit communiquer le plus grand nombre d'information dans le plus petit espace et avec le moins d'encre possible, c'est aussi sur la capacité à « fonctionner » du graphique que Tufte se concentre.

2. Les variables visuelles

2.1 Le plan



Figure 113 : L'espoir d'une reprise économique en V, par Tristan Gaudiaut, 2020 (CC0)

Pour Bertin, l'image se construit sur le plan géométrique. C'est en utilisant les dimensions du plan que l'on construit une représentation graphique. Évidemment, au vu de son époque, il ne parle que des représentations statiques. Ainsi, sur le plan, on peut réaliser l'image selon deux ou trois dimensions. Sans accès au dynamisme ou à

l'interaction, dépasser la troisième dimension est impossible. Une représentation statique peut ainsi comporter jusqu'à 3 dimensions au maximum, parfois 4 dans de rares cas. S'il y a plus de dimensions à transcrire, il faut penser à rédiger son information en deux représentations visuelles distinctes. En effet, « la perception visuelle n'admet qu'un nombre réduit de variables » (Bertin, 1967, p. 9).

C'est grâce aux variables visuelles (ou rétiniennes) que l'on code une information sur la troisième dimension de l'image. Sur la figure 113, qui est un graphique en ligne (série chronologique), la variable des années est transcrite sur la première dimension du plan (axe x) tandis que le taux de croissance du PIB réel figure sur la deuxième dimension du plan (axe y). Le croisement de ces deux dimensions permet de déceler des correspondances. Mais il existe une troisième information (le pays) qui est codée grâce à la couleur sur une troisième dimension. La couleur est une variable visuelle.

2.2 Les variables visuelles selon Bertin (1967)

Lorsqu'il est nécessaire d'afficher une troisième variable (ou une deuxième, dans le cadre de la cartographie), il faut faire appel aux variables visuelles (rétiniennes) qui permettent de coder cette information sur une troisième dimension. Elles sont au nombre de six :

- la taille ;
- la valeur ;
- le grain ;
- la couleur ;
- l'orientation ;
- la forme.

Les variables visuelles s'organisent selon différents niveaux de perceptions. Cela signifie que les variables visuelles n'ont pas toutes les mêmes propriétés perceptives. Bertin les a donc classées en fonction de ces propriétés. C'est en quelque sorte une application des principes de la *Gestalt theory* : l'utilisation de ces variables nécessite à peine la lecture de la légende tant elles semblent intuitives et permettent une compréhension quasi-spontanée. Les niveaux de perception sont les suivants :

Tableau 3 : Niveaux de perception selon Jacques Bertin (1967)

<p>La perception associative</p>	<p>≡</p>	<p>Permet d'égaliser une variation, de regrouper des correspondances. L'œil peut spontanément reconstruire l'homogénéité de l'image malgré une variation visuelle. En somme, c'est ce qui « permet à une catégorie nominale d'être perçue malgré l'influence des autres variables dans une même image » (Hurter, 2010).</p>	<p>Forme, orientation, couleur, grain</p>
<p>La perception sélective</p>	<p>≠</p>	<p>Permet de distinguer différentes catégories et familles. Elle concerne les catégories nominales. Elles sont perçues directement, comme mises en avant dans l'image (Hurter, 2010).</p>	<p>Taille, valeur, grain, couleur (et parfois orientation)</p>
<p>La perception ordonnée</p>	<p>○</p>	<p>Donne une idée d'ordre entre différentes catégories. La perception ordonnée est stimulée lorsqu'au premier regard, le classement visuel des catégories apparaît (Hurter, 2010).</p>	<p>Taille, valeur, grain. Certains auteurs assurent que la couleur permet la perception ordonnée (en jouant sur la luminosité d'une même teinte).</p>
<p>La perception quantitative</p>	<p>Q</p>	<p>Donne une idée d'ordre entre différentes catégories. La perception ordonnée est stimulée lorsqu'au premier regard, le classement visuel des catégories apparaît (Hurter, 2010).</p>	<p>Taille</p>

De cette manière, la **taille** permet d'agrandir ou de diminuer la surface d'un signe. Une figure, une forme, une ligne ou les éléments contenus dans une zone peuvent tous varier de taille (Bertin, 1967).

La **valeur** consiste en des catégories qui varient entre le noir et le blanc. Bertin définit la valeur comme « le rapport entre les quantités

totales de noir et de blanc perçues dans une surface donnée » (Bertin, 1967, p. 73). Tufte considère que la valeur comporte une forme hiérarchique toute naturelle et facilement perceptible (Tufte, 1983). La raison en est que la variation de valeur est ordonnée.

Le **grain** est « la quantité de taches séparables contenues dans une surface unitaire » (Bertin, 1967, p. 79). Bertin parle du grain nul, composé d'un tel nombre d'éléments qu'ils ne sont plus distinguables. Pour lui, « le grain trop fin ne se réduit pas et disparaît de toute reproduction microfilmée » (Bertin, 1967, p. 79). Le grain peut varier en longueur et en taille, à noter que « plus les taches sont grandes, plus le nombre des paliers séparables est grand » (Bertin, 1967, p. 79). Bertin dénonce également le possible effet vibratoire du grain. Effectivement, associé à une valeur de noir et de blanc, le grain peut provoquer une sensation visuelle désagréable que Bertin explique par un effet physiologique (une résonance est créée par l'individu) et par un effet psychologique (l'individu hésite à placer son intention sur le fond ou la forme).

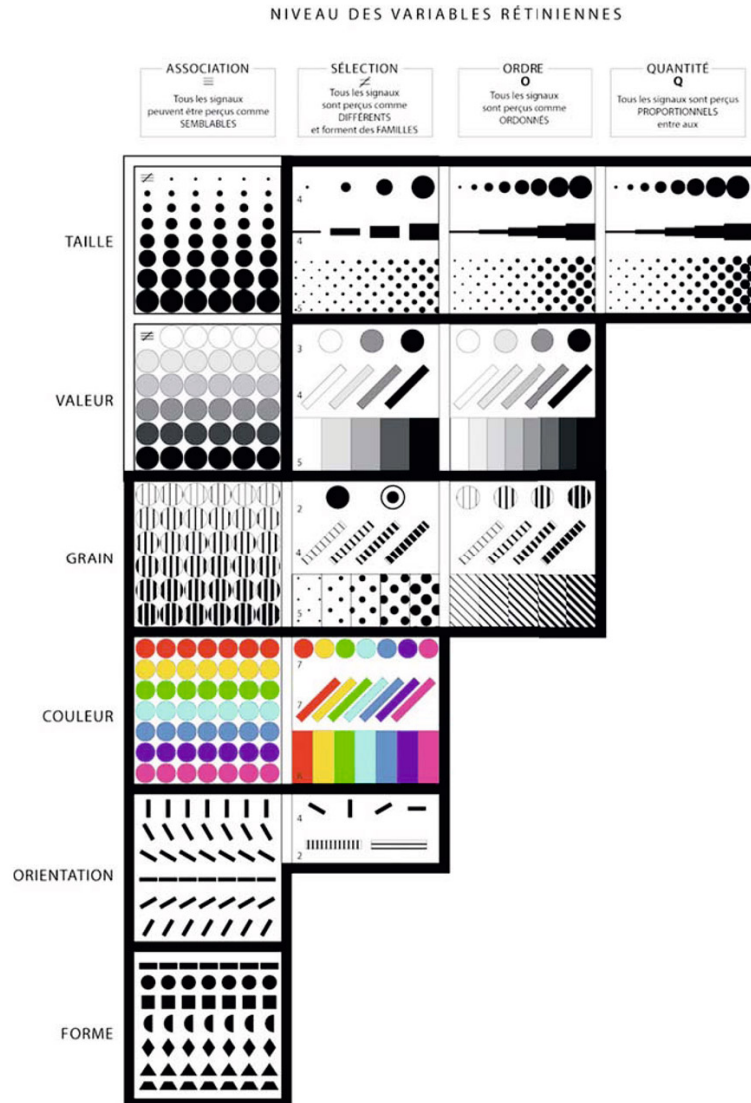


Figure 114 : Tableau des propriétés des variables rétiniennees (Bertin, 1967, p. 96)

L'**orientation** concerne les signes qui sont placés dans des sens différents, « c'est la différence d'angle entre des champs créés par plusieurs signes parallèles » (Bertin, 1967, p. 93). En implantation linéaire, il faut se limiter à deux orientations.

La **forme** recouvre des possibilités infinies. « Une tache de surface constante peut prendre une infinité de formes différentes » (Bertin, 1967, p. 95). Ce qui est intéressant dans la forme est le fait qu'on

puisse la reconnaître lorsqu'elle est semblable à plusieurs reprises dans la même représentation (Bertin, 1967).

La **couleur** permet d'associer et de différencier. Pour Bertin, les couleurs ne peuvent être utilisées pour ordonner ; elles ont plutôt une intention sélective. Il est pourtant possible de composer des cartes de couleur qui induisent un ordre naturel et intuitif (Telea, 2014). En effet, la variation de valeur (dans ce cas, faire varier la luminosité d'une même couleur) peut être induite dans la variation de couleurs. Ainsi, ce n'est pas la variation de couleurs (teinte) qui est utile à la perception ordonnée, mais la variation de valeur (luminosité) induite dans la couleur, si l'on parle dans les termes de Bertin. Tufte considère aussi la couleur comme un quantifieur naturel (Tufte, 1990). Un chapitre entier est consacré aux principes de conception relatifs à la couleur.

Les différentes variables visuelles sont résumées dans la figure 114, en fonction de leurs propriétés perceptives. Les variables visuelles peuvent être combinées entre elles au sein d'une même représentation graphique.

2.3 Les variables visuelles aujourd'hui

Avec le temps, de nombreuses études ont été menées sur les variables visuelles. Plusieurs auteurs comme Cleveland et McGill (1984), Mac Eachren (1995) et ses collègues (2012) ou encore Roth (2017) ont évalué leur pouvoir d'encodage et les ont élargies en fonction des évolutions informatiques et technologiques. En 1967, à l'époque de Bertin, les représentations graphiques étaient majoritairement réalisées à la main et imprimées. Aujourd'hui, les évolutions technologiques permettent l'accessibilité de la représentation graphique grâce à de nouveaux supports de lecture, mais aussi grâce

à l'apparition d'outils rendant aisée la création de visualisations de données.

Roth (2017) présente ainsi non pas 6, mais 12 variables visuelles, présentées dans la figure 115. Initialement destinées à la géographie, ces variables visuelles s'appliquent également à la visualisation de données en général. On y retrouve les variables visuelles traditionnelles de Bertin : taille, forme, orientation, couleur, valeur et grain auxquelles sont ajoutées la localisation, la teinte (couleur), la saturation (couleur), l'arrangement, la netteté, la résolution et la transparence. Sur ce schéma, on voit pour chaque variable visuelle :

- si la variable visuelle permet une perception associative ou sélective (oui ou non) ;
- dans quelle mesure (bon, marginal ou pauvre) elle permet d'encoder une variable nominale, ordinale ou quantitative.

Les 7 variables visuelles supplémentaires se définissent comme ceci :

- La **localisation** décrit la position d'un symbole par rapport à un cadre de coordonnées.
- La **teinte** de la couleur décrit la longueur d'onde dominante de l'information sur la partie visible du spectre électromagnétique (par exemple, bleu, vert, rouge) et est l'une des trois variables visuelles associées à la perception de « couleur ».
- La **valeur** de la couleur décrit la quantité relative d'énergie émise ou réfléchi par l'information. La variation de la valeur des couleurs se traduit par la perception d'ombres, ou de zones relativement claires (forte émission ou réflexion de la lumière) et sombres (faible émission ou réflexion de la lumière). En conséquence, la valeur de la couleur est parfois appelée « *lightness* » dans la théorie des couleurs. Ainsi, comme la couleur permet bien de coder une information ordonnée.

- La **saturation** des couleurs décrit la crête spectrale de l'information sur le spectre visible. C'est la troisième des trois variables visuelles associées à la perception des couleurs. Les couleurs vives ou saturées émettent ou réfléchissent la lumière dans une bande très concentrée du spectre visible, tandis que les couleurs pastel ou désaturées émettent ou réfléchissent la lumière de manière uniforme dans le spectre visible.
- L'**arrangement** décrit la disposition des marques graphiques constituant un symbole cartographique. L'arrangement consiste à modifier un modèle régulier. Il est particulièrement utilisé dans les cartes.
- La **netteté** décrit l'acuité de la limite des symboles utilisés. Elle est également appelée « profondeur de champ » et « flou » (MacEachren *et al.* 2012) ont constaté que la netteté était la variable visuelle la plus efficace pour représenter l'incertitude dans le contexte de la symbolisation des points.
- La **résolution** décrit la précision spatiale où le symbole est affiché. Plutôt utilisée en cartographie, elle sert à généraliser en réalisant la suppression significative de détails, car la complexité du monde réel est abstraite. Cela sert ainsi à s'adapter à l'échelle réduite de la carte.
- La **transparence** décrit la quantité de mélange graphique entre un symbole et le fond de la représentation (ou les symboles) sous-jacents.

(Roth, 2017)

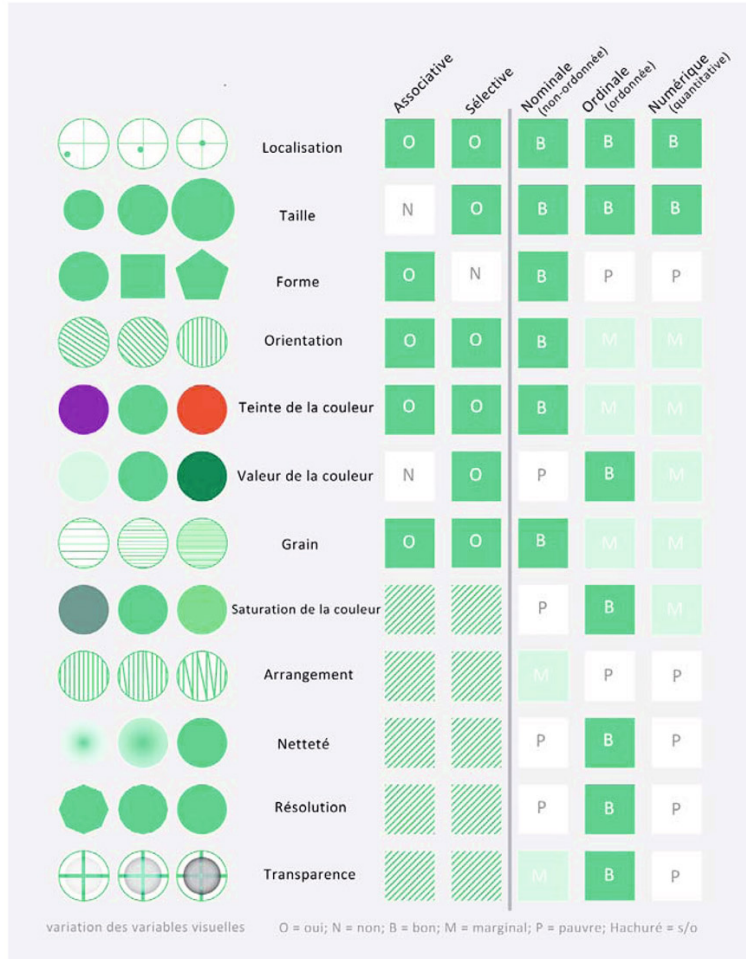


Figure 115 : Traduction de « Visual variables and their syntactics. Figure derived from Bertin (1983), MacEachren (1995), and MacEachren *et al.* (2012). », dans Roth, R. E. (2017). Visual Variables. In D. Richardson, N. Castree, M. F. Goodchild, A. Kobayashi, W. Liu, & R. A. Marston (Éds), International Encyclopedia of Geography : People, the Earth, Environment and Technology (p. 1-11). John Wiley & Sons, Ltd.

2.4 L'interaction

L'interaction ne sera pas abordée en détail dans cet ouvrage, mais il convient d'en dire quelques mots. Effectivement, les évolutions technologiques et informatiques permettent aujourd'hui de pallier certains problèmes – comme l'impossibilité, par exemple, de réaliser

un graphique classique dans lequel apparaissent quatre dimensions différentes, ou encore d'éviter le problème de lisibilité des données lorsqu'elles sont nombreuses sur la représentation graphique. Là où les variables visuelles sont parfois insuffisantes par rapport à l'information à représenter, l'interaction s'avère utile parce qu'elle laisse aussi au lecteur davantage de libertés, comme celle d'explorer ou encore de sélectionner l'information dont il a besoin. La visualisation est alors conçue tout en permettant à l'utilisateur de réaliser certaines tâches comme :

- *Overview* : est une vue d'ensemble des données.
- *Zoom* : l'utilisateur peut zoomer sur des éléments d'intérêt.
- *Filter* : grâce à une liste déroulante par exemple, l'utilisateur peut filtrer les éléments (in) intéressants.
- *Details-on-demand* : le lecteur peut sélectionner un élément ou un groupe et obtenir des détails supplémentaires.
- *Relate* : permet de voir les relations entre les éléments.
- *History* : permet au lecteur de conserver un historique des actions (*undo*, *replay*) et vise l'affinement progressif de l'information.
- *Extract* : permet à l'utilisateur d'extraire des sous-collections de données et des paramètres de requête.

(Shneidermann, 1996, p. 337)

En guise d'exemple, nous présentons la visualisation interactive⁹ suivante (QR code). Elle présente la corrélation entre l'espérance de vie et le revenu moyen à un niveau mondial. La couleur représente le continent, tandis que la taille de la bulle représente la population totale des pays représentés. Une bulle représente un pays ; une étiquette apparaît lorsque la souris survole une bulle. De nombreuses libertés sont laissées à l'utilisateur : filtrer les pays, zoomer et même

avoir une vision temporelle. En cliquant sur « Play », la vue temporelle et animée se déclenche seule. À vous : explorez !



[https://www.gapminder.org/tools/#\\$chart-type=bubbles](https://www.gapminder.org/tools/#$chart-type=bubbles)

3. L'intégrité graphique

3.1 *Distorsion et lie factor*

Les chiffres donnent une idée aseptisée de la réalité. Ils ont un air objectif, neutre. Et pourtant, la conception d'un message est loin d'être objective. Certains choix de conception d'une visualisation de donnée peuvent parfois donner l'illusion que les chiffres représentent un phénomène plus grand ou plus petit qu'en réalité. Par exemple, le diagramme en bâtonnet de la figure 116 ne présente pas une information correcte. Regardez l'axe des ordonnées (Y). Rien ne vous choque ?

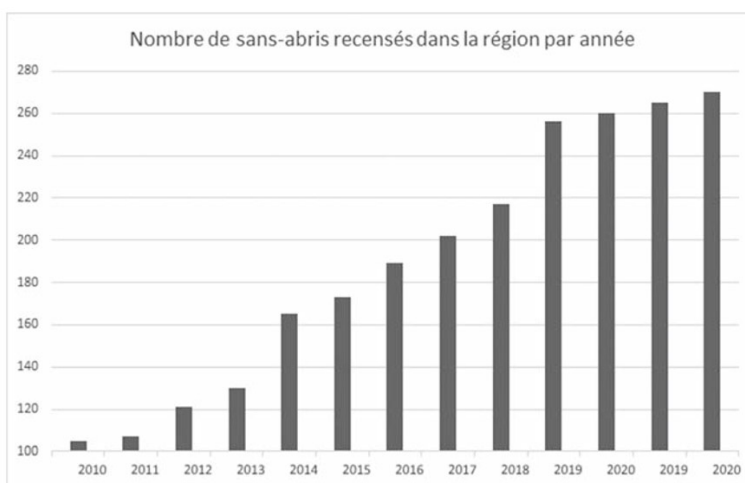


Figure 116 : Diagramme en bâtonnets représentant « Le nombre de sans-abris recensés dans la région par année » avec distorsion (chiffres fictifs), par les auteurs (CC0)

Vous pouvez constater sur l'axe Y que l'échelle ne commence pas à 0, mais à 100. Ainsi, vous avez l'impression qu'il y a peu de sans-abris en 2010 et 2011... Or, il y en a plus de cent ! Ce choix de conception a tendance à amplifier le phénomène, à donner une impression exagérée de la réalité au lecteur. Ainsi, la figure 117 montre la représentation exacte de ces nombres. On constate dès lors que l'augmentation est moins drastique qu'il n'y paraît, même si elle est importante. Les figures 116 et 117 montrent exactement les mêmes données.

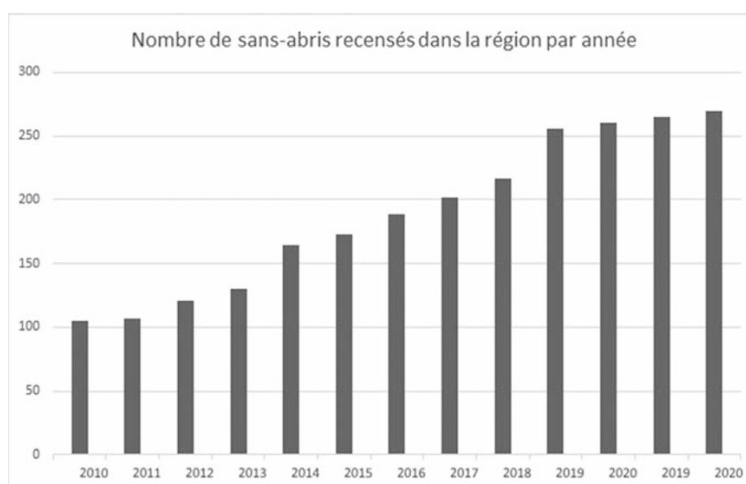


Figure 117 : Diagramme en bâtonnets représentant « Le nombre de sans-abris recensés dans la région par année » avec distorsion (chiffres fictifs), par les auteurs (CC0)

Il arrive que des distorsions volontaires ou involontaires se glissent dans les représentations graphiques. Les distorsions concernent l'effet visuel perçu. Un graphique ne présente pas de distorsion si la représentation visuelle des données est cohérente avec les valeurs numériques.

Il est donc important de veiller à ce que la représentation des nombres, mesurée physiquement sur la surface du graphique en lui-même, soit directement proportionnelle aux quantités numériques

représentées (Tufte, 2001, p. 56). Un étiquetage clair, détaillé et complet permet d'éviter l'ambiguïté dans les données. Étiqueter les événements importants dans les données aide le lecteur.

Le *lie factor* est un ratio permettant de vérifier si l'effet montré par le graphique est réellement celui contenu dans les données (Tufte, 2001, p. 57).

$$\text{Lie factor} = \frac{\text{taille de l'effet montré dans le graphique}}{\text{taille de l'effet montré dans les données}}$$

La taille de l'effet se calcule comme ceci :

$$\text{Taille de l'effet} = \frac{(\text{la dernière valeur} - \text{la première valeur})}{\text{la première valeur}}$$

Si la valeur obtenue par le calcul du *lie factor* se révèle inférieure à 0,95 ou supérieure à 1,05, le graphique présente de la distorsion. Le calcul est assez simple à réaliser sur papier : il suffit de mesurer la représentation visuelle des données en cm. Le calcul fonctionne également très bien avec des pixels et est donc possible sur ordinateur.

Voici le calcul des *lie factors* pour les diagrammes en bâtonnets des figures 116 et 117 :

1. Diagramme avec distorsion (figure 116)

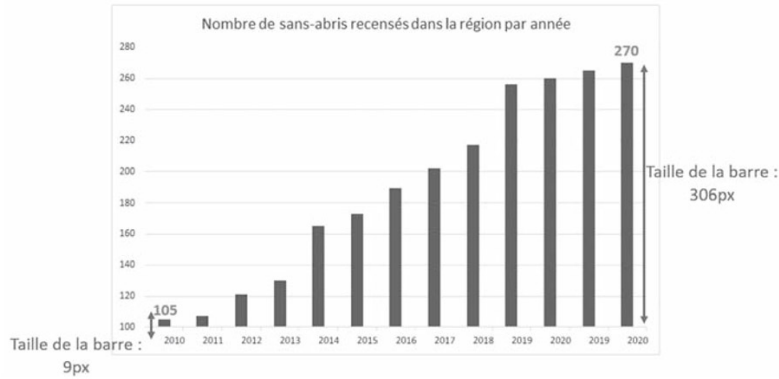


Figure 118 : Mesure du lie factor sur la figure 116

Le calcul :

$$\frac{\frac{(306 - 9)}{9}}{\frac{(270 - 105)}{105}} = \frac{33}{1,57} = 21,01$$

Le *lie factor* est de 21,01. Ce nombre est bien supérieur à 1,05 : il indique une distorsion énorme dans le graphique !

2. Diagramme sans distorsion (figure 117)

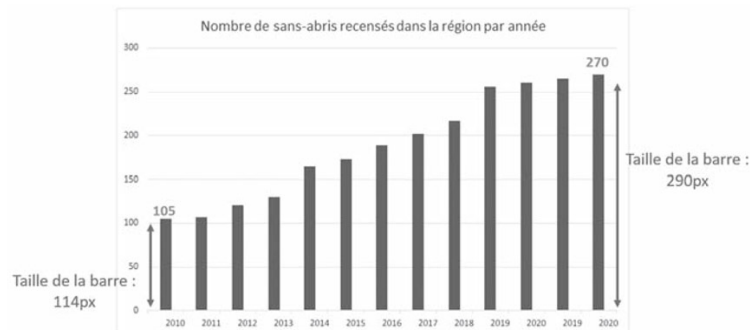


Figure 119 : Mesure du lie factor sur la figure 117

Le calcul :

$$\frac{\frac{(290 - 114)}{114}}{\frac{(270 - 105)}{105}} = \frac{1,54}{1,57} = 0,98$$

Le *lie factor* est de 0,98. Il est donc supérieur à 0,95 et montre bien qu'il n'y a aucune distorsion dans le graphique.

3.2 Les six principes de l'intégrité graphique

L'intégrité graphique vise donc à réaliser des visualisations de données « honnêtes » qui révèlent les données, « la vérité contenue dans les données » (Tufte, 2001). Il existe ainsi six principes de l'intégrité graphique.

3.2.1 Proportionnalité visuelle et numérique

La représentation des nombres physiquement mesurée sur le graphique en lui-même doit être directement proportionnelle aux quantités contenues dans l'ensemble de données (2001, p. 77). Il s'agit, tout simplement, de respecter le *lie factor* et d'éviter la distorsion.

3.2.2 Privilégier la variation des données à l'esthétique

Montrez la variation des données et non la variation du design, faire passer le design d'une visualisation avant la communication adéquate des données peut générer un affichage corrompu. En somme, pensez au message avant de penser à son esthétique. Sur la figure 120, on voit bien que l'esthétique a primé dans le choix de conception. Il est très difficile de comprendre à quoi les différentes étiquettes font référence et de déduire une information correcte.

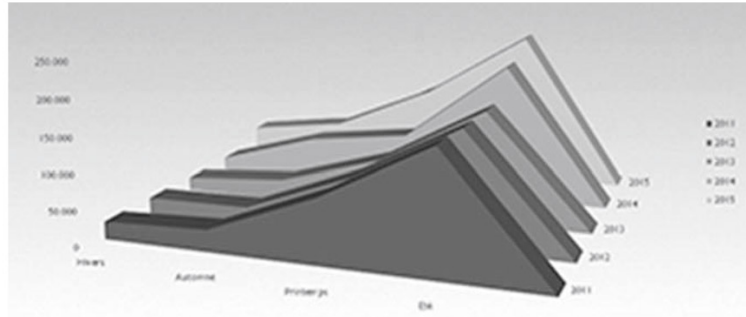


Figure 120 : Vente des maillots de bain au cours des saisons de 2011 à 2015, chiffres fictifs, par les auteurs (CC0)

3.2.3 Les dimensions visuelles égales aux dimensions contenues dans les données



Figure 121 : Les bistrotts en voie de disparition ?, par Tristan Gaudiaut, 2020 (CC0)

Le nombre de dimensions porteuses d'information (variables) représentées dans un graphique ne doit pas dépasser le nombre de dimensions qui existent dans les données. Par exemple, l'utilisation de deux dimensions pour montrer des données unidimensionnelles est une technique faible et inefficace, capable de ne traiter que de très petits ensembles de données, avec souvent des erreurs de conception et une ambiguïté de perception. Cela concerne principalement les aires visuelles et les mesures quantitatives. Utiliser

des aires pour montrer des données unidimensionnelles peut donner une illusion erronée de ce que représentent réellement les données. La figure 121 en montre un exemple.

Sur cette visualisation, la troisième icône, indiquant le débit de boisson en 2016, est un peu trop grande : elle devrait être plus petite pour illustrer correctement la représentation des nombres. Le rapport de taille entre les icônes n'est pas proportionnel au rapport numérique qui existe entre les données, comme on le voit dans le tableau suivant.

Tableau 4 : Différence entre le rapport numérique et le rapport de taille dans les éléments visuels de la figure 121

Rapport numérique entre les nombres		Rapport de taille dans le visuel entre les icônes dans le visuel	
Entre 500 000 et 200 000	2,5	Entre la première et la deuxième icône	2,5
Entre 200 000 et 38 800	5,15	Entre la deuxième et la troisième icône	4
Entre 500 000 et 38 800	12,88	Entre la première et la troisième icône	10,53

La troisième icône devrait donc être 12,88 fois plus petite que la première, et non 10,53 fois. Dans cet exemple, la différence n'est pas flagrante, mais elle peut être dramatique dans d'autres cas.

L'œil humain n'excelle pas dans la comparaison des aires (Cleveland & McGill, 1984). En utilisant un symbole à deux dimensions pour exprimer des données à une dimension (comme le débit de boisson), on perd en précision. La hauteur de barres dans un diagramme en bâtonnets ou la distance entre l'axe X et les sommets d'un diagramme en ligne permettent d'afficher les données avec plus

d'exactitudes, d'autant plus que les propriétés perceptives y sont meilleures.

3.2.4 Contextualiser les données

Les graphiques ne doivent pas citer les données hors contexte. Si rien n'est expliqué à propos des données, pas même l'échelle ou l'unité de mesure, comment comprendre l'information ? En découle le principe suivant.

3.2.5 Ne pas lésiner sur l'étiquetage

Étiqueter les événements/explications importants dans les données sur le graphique lui-même : l'étiquetage direct doit être utilisé pour éviter toute distorsion et ambiguïté. Les étiquettes de données sont importantes : elles permettent de contextualiser ou d'exprimer un phénomène important révélé par les données. Il convient également de prendre soin de rédiger correctement les titres, échelles, unités, légendes, etc.

En effet, une visualisation de données est titrée dans la plupart des cas. Le titre permet de faire comprendre au lecteur quelle relation est mise en avant entre les différentes composantes du graphique (Bertin, 1967). Il ne faut pas le confondre avec la légende. La légende intervient dans tout graphique à partir du moment où un codage supplémentaire est ajouté pour comprendre une information. La légende, le titre et les différentes annotations sur le graphique constituent l'étiquetage. Un étiquetage précis, complet et de qualité diminue le risque d'ambiguïté sur la représentation graphique (Tufte, 1983). Par ailleurs, sans étiquetage du tout, il est impossible de comprendre une visualisation.

3.2.6 Les mesures monétaires

Dans les séries chronologiques, les unités de mesure monétaires déflatées et normalisées sont presque toujours meilleures que les unités nominales.

4. Visualiser en simplicité

4.1 Le *data-ink ratio* (Tufté, 2001)

Tufté a élaboré une théorie graphique relative à la conception de visualisations de données. Le *data-ink* y prend une place importante. Littéralement traduit par « encre des données » en français, c'est bien à l'encre présente sur le graphique (ou pixels) que Tufté fait allusion.

En effet, la majeure partie de l'encre utilisée dans un graphique devrait être consacrée aux données. Le *data-ink* correspond justement à l'encre non effaçable et non redondante consacrée aux données. Si l'on supprime le *data-ink*, le graphique perd son contenu, sa substance. À l'inverse, le *non-data-ink* ne véhicule pas de données, mais est utilisé pour représenter d'autres informations comme les axes, les échelles, les étiquettes et les arêtes.

Le *data-ink ratio* désigne la proportion d'encre qui est consacrée aux données par rapport à la quantité totale d'encre utilisée dans un graphique. Il se calcule comme de cette façon :

$$\text{Data-ink ratio} = \frac{\text{Encre consacrée aux données}}{\text{Total d'encre utilisée dans le graphique}}$$

Si le *data-ink ratio* est proche de 1, l'encre utilisée est essentiellement consacrée aux données. S'il est proche de 0, une majeure partie de l'encre utilisée peut être effacée sans que le graphique ne perde de son contenu, de sa substance.

Sur la figure 122, deux exemples totalement opposés sont présentés. Sur le graphique de gauche, le *data-ink ratio* est assez bas et tend

vers 0 ($\sim 0,30$) : beaucoup d'encre présente sur le graphique a été utilisée sans pour autant servir les données. Sur le graphique de gauche, le *data-ink ratio* est assez haut et tend vers 1 ($\sim 0,90$). Il s'agit du même graphique : le fond coloré, le quadrillage inutile, l'encadrement des barres ainsi que l'ombrage des barres ont été enlevés afin qu'un maximum d'encre présente sur le graphique soit consacré aux données. Au final, le bruit graphique a été réduit, proposant ainsi un visuel plus épuré et plus clair. Selon Tufte : « **Less is more !** »

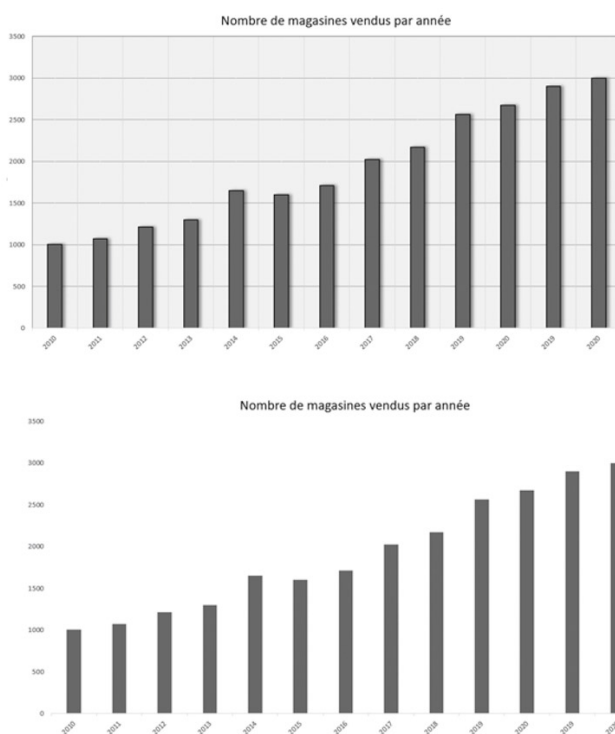


Figure 122 : Nombre de magazines vendus par année, chiffres fictifs. Comparaison de différents *data-ink ratios*, par les auteurs (CC0)

Cinq principes de la théorie des graphiques de données :

1. Montrez les données avant tout

Il faut penser à montrer les données avant de penser à leur esthétique. Est-ce que vos choix mettent correctement les données à l'avance ?

2. Maximisez le *data-ink ratio*

Un graphique minimaliste et épuré engendrera moins de bruit graphique et permettra de communiquer l'information de façon plus ciblée, permettant de se concentrer sur les données, sur le fond de l'information et non sur son esthétique.

3. Effacez le *non-data-ink*

En effaçant tout ce qui n'est pas du *data-ink*, le graphique minimaliste apparaît, révélant ainsi les données et seulement les données. Ce principe va de pair avec le précédent.

4. Effacez le *data-ink* qui est redondant

Si des informations se répètent sans que cela ne soit utile et si la redondance du *data-ink* ne sert pas à écarter une possible ambiguïté, mieux vaut enlever cette redondance afin de tendre vers un graphique épuré qui sera plus accessible.

5. Éditez, raffinez

Cette étape consiste à modifier le graphique et à l'alléger des éléments qui peuvent engendrer des effets involontaires et non nécessaires. Par exemple, dans la figure 123, le dégradé de couleurs, la grille et la 3D utilisée aident-ils réellement à comprendre l'information ? Ces éléments desservent plutôt l'information.

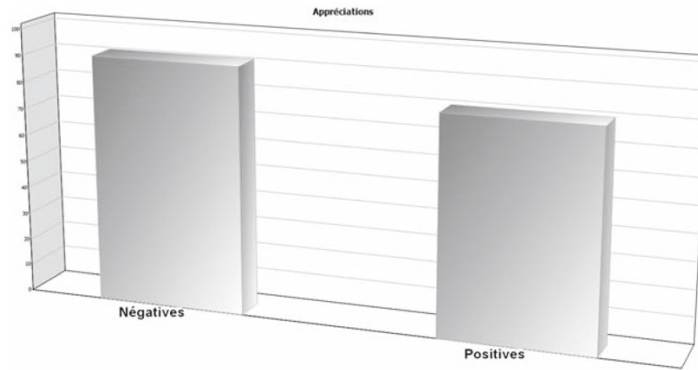


Figure 123 : Nombre d'appréciations, chiffres fictifs, par les auteurs (CC0)

4.2 La simplification, révélatrice de sens

Pour Bertin aussi, la simplification est un élément fondamental de la construction graphique. Lorsque l'on souhaite communiquer une information complexe (et d'autant plus en représentation graphique), la simplification est inévitable : elle est « une obligation de l'acte de communication », ayant pour objet de « faire comprendre » l'information au lecteur (1967, p. 166).

Donc, puisque la représentation graphique vue comme un objet communicationnel a pour objectif la compréhension du lecteur, sa simplification est obligatoire : « la simplification est créatrice » (Bertin, 1967, p. 166). Bertin appelle le traitement de l'information « simplification logique », car, tout simplement, « le traitement graphique de l'information s'opère par la simplification de l'image ». Contrairement aux images complexes qui rendent de nombreux niveaux de lecture possibles, les images simples, en réduisant l'hésitation du lecteur, sont bénéfiques à la mémorisation et à la communication. C'est là qu'interviennent trois règles de construction graphique.

« Construire une information en une image » (Bertin, 1967, p. 171). Cela consiste à rendre la visualisation de données compréhensible en

un minimum d'instant de perception, c'est-à-dire le plus rapidement possible. Les fioritures retardent donc cette rapidité de compréhension puisqu'elles ajoutent une charge à ces instants de perception. En découle le deuxième principe.

En ce qui concerne les représentations à dimensions ordonnables, simplifier l'image au maximum (1967, p. 171), en prenant garde à ne pas réduire le nombre de correspondances. Ce principe est très proche de la maximisation du *data-ink ratio* selon Tufte : on garde l'essentiel de la représentation des données, en simplifiant au maximum les éléments apparaissant autour de cette information.

Réduire l'image pour la simplifier lorsqu'elle comporte des dimensions ordonnées (1967, p. 177). Cette idée de réduction de l'image à des fins de simplification se retrouve aussi dans les travaux de Tufte. Il l'appelle le *shrink principle*.

Bertin et Tufte, deux véritables précurseurs en visualisation de données, s'accordent donc sur un point : la visualisation de données simple, minimaliste et épurée, sera la plus facilement compréhensible en comparaison aux représentations chargées d'éléments visuels qui ne se consacrent pas aux données, mais plutôt à la forme qu'au fond.

5. Évitez le *chartjunk*

Un des principes de visualisation de données les plus connus est celui-ci : « Évitez le *chartjunk* ». Mais au fond, de quoi s'agit-il, et pourquoi certains chercheurs et professionnels remettent-ils ce principe en question ?

5.1 Du bruit graphique aux ornements extrêmes

Pour de nombreuses personnes, ajouter une touche décorative aux visualisations de données est une grande tentation. Pourtant, Tufte décourage tout ajout esthétique qui n'est pas du *data-ink*. Il dit d'ailleurs que « la décoration intérieure des graphiques génère beaucoup d'encre qui ne dit rien de neuf au lecteur » (Tufte, 2001, p. 107, traduction libre). Penser à la variation d'esthétique avant de penser à montrer la variation des données engendre donc beaucoup de *non-data-ink* ou de *data-ink* redondant, ce qui tend ainsi vers ce que Tufte appelle le « *chartjunk* ».

Le *chartjunk* concerne tous les éléments superflus, voire indésirables, sur un graphique. Il inclut les illusions, les grilles, les effets 3D ou encore les effets moirés. Tufte plaide ainsi en faveur de la maximisation du *data-ink ratio*, avec comme objectif la réduction de la charge cognitive lors de l'exploration et l'analyse des données. La présence de décorations « artistiques » est aussi du *chartjunk*, n'ajoutant aucune valeur à l'information selon Tufte. Il distingue trois types de *chartjunk* :

- l'art optique involontaire ;
- les grilles ;
- les canards (*self promoting duck*).

(Tufte, 2001, p. 108)

Par « art optique involontaire », Tufte suggère qu'une conception graphique peut engendrer un tremblement physiologique dans l'œil du lecteur et être désagréable, comme dans les figures 124 et 125, où un mouvement de vibration apparaît.

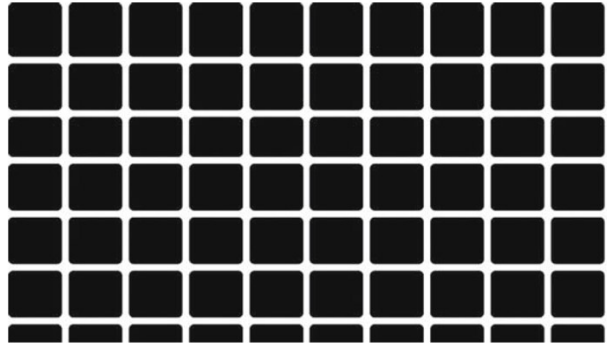


Figure 124 : Illusion visuelle, Pixabay, 2020 (Domaine public)

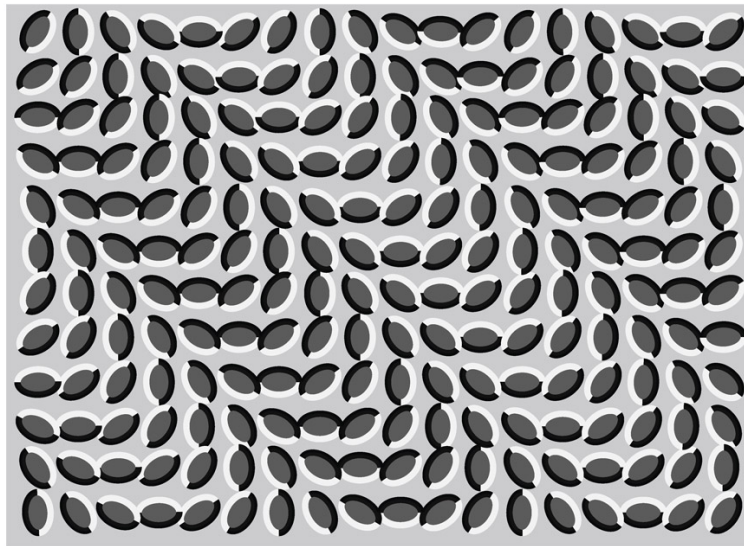
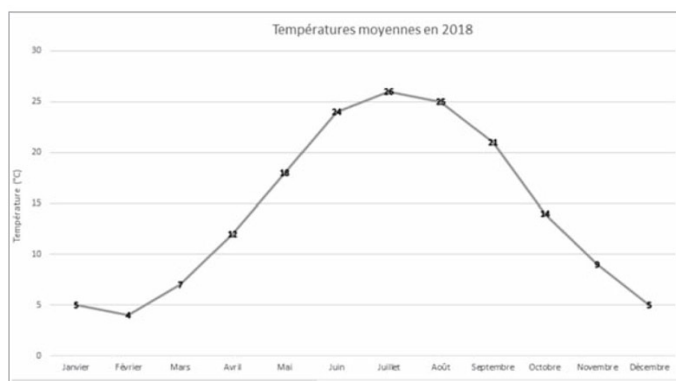
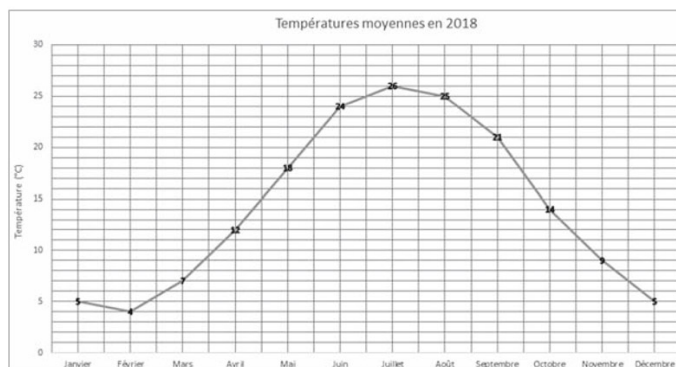


Figure 125 : Grains de café, Openclipart (Domaine public)

Les grilles, quant à elles, sont un élément qui devrait être supprimé lorsqu'il s'avère inutile. À de nombreuses reprises, ce n'est pas le cas. La présence d'une grille entre en « compétition » avec les données présentées (Tufte, 2001, p. 112). En revanche, si une grille est essentielle et aide à lire, il est primordial de la faire apparaître (Tufte, 2001, p. 116). La figure 126 est un exemple : une grille trop dense est gênante, tandis qu'une grille légère peut aider la lecture de l'information. Si l'étiquetage est minutieux, la grille peut complètement disparaître.

Le canard (*self-promoting duck*) est un style de graphique réalisé pour sa forme décorative et pour le design avant tout, et non en fonction des formes qui mettraient les données le plus en valeur. Les mesures de données deviennent alors tout simplement de purs éléments de design et le style prévaut sur l'information quantitative (Tufté, 2001, p. 116). Aussi, il consiste souvent en l'ajout d'une fausse perspective sur la structure des données (3D), ce qui est assez trompeur (Tufté, 2001, p. 118). Pour l'anecdote, Tufté a appelé ce type de *chartjunk* « *the duck* » en référence à un bâtiment architectural (The big duck) construit dans les années 1930 à Flanders dans l'État de New York (figure 127). Grand exemple de publicité littérale, ce bâtiment destiné à la vente de produits de canard a surtout été conçu pour son esthétique et non pour son ergonomie. L'idée est similaire avec le *chartjunk*, comme en témoigne la figure 128.



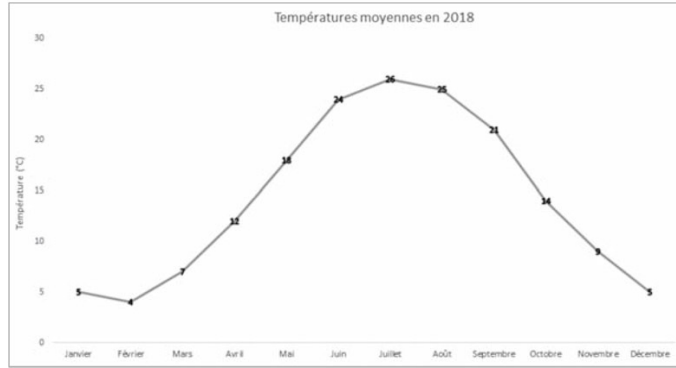


Figure 126 : Disposition de la grille sur « Températures moyennes en 2018 » (chiffres fictifs), par les auteurs (CC0)

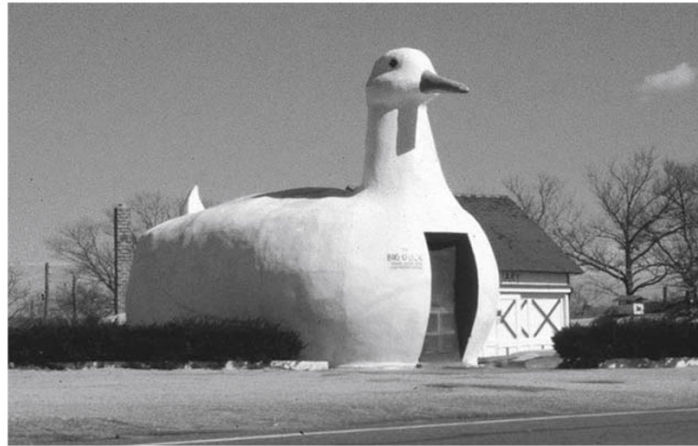


Figure 127 : The Big Duck (Domaine public)

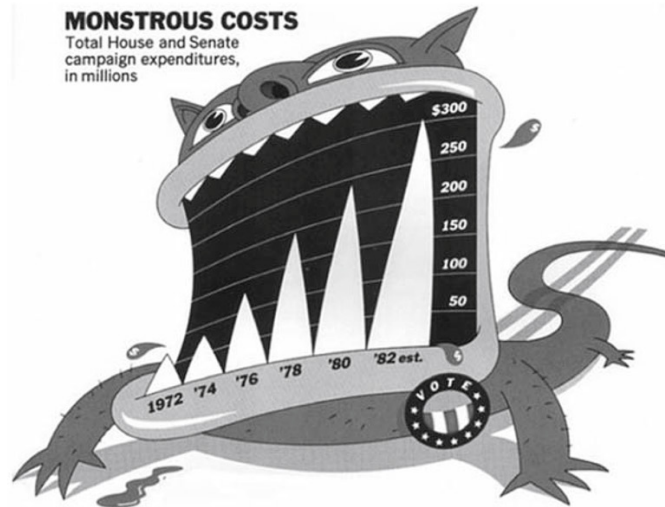


Figure 128 : « Monstrous costs » par Nigel Holmes dans Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A., McDine, D., & Brooks, C. (2010, April). Useful junk? The effects of visual embellishment on comprehension and memorability of charts. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 2573-2582). (CC BY-SA)

Selon Tufte, le *chartjunk* n'atteint pas le but de ses propagateurs. Le graphique ne deviendrait pas plus intéressant et plus attractif parce que des éléments ornementaux y seraient placés. Il peut ainsi transformer les données et le faire tourner au désastre. Le meilleur design provoque l'intrigue et la curiosité et révèle la vraie teneur des données, parfois grâce à un sens narratif, au sens du détail et parfois par une présentation simple et élégante de données intéressantes. Aucune information, découverte ou toute autre substance n'est produite par le *chartjunk* (Tufte, 2001). Tufte est on ne peut plus clair : « oubliez le *chartjunk*, en ce compris les effets de vibration, la grille et le canard » (Tufte, 2001, p. 121). Il inclut également dans le *chartjunk* les ornements et ajouts esthétiques, plus ludiques. Il pense que les promoteurs du *chartjunk* imaginent les données comme ennuyeuses, et qu'ils ont besoin d'ajouter de la décoration « cosmétique » afin de briser l'ennui (Tufte, 1990, p. 34).

5.2 Évitez la 3D à tout prix !

Comme tout bon *chartjunk* qui se respecte, l'utilisation de la 3D abusive engendre des problèmes de perception de l'information. Bien que souvent proposée en option dans les programmes qui permettent de concevoir des graphiques, elle est une véritable erreur de design. Imaginons que vous ayez deux dimensions à encoder dans un diagramme en bâtonnets ; la taille moyenne en fonction du sexe. Est-il logique d'y ajouter une troisième dimension visuelle (3D), alors que vous souhaitez seulement encoder deux dimensions sur le plan ? Voyez par vous-même sur la figure 129. Entre le graphique en bâtonnets en 2D et celui en 3D, vous constaterez que le graphique en 2D vous donne une perception bien plus claire de la différence entre les données présentées.

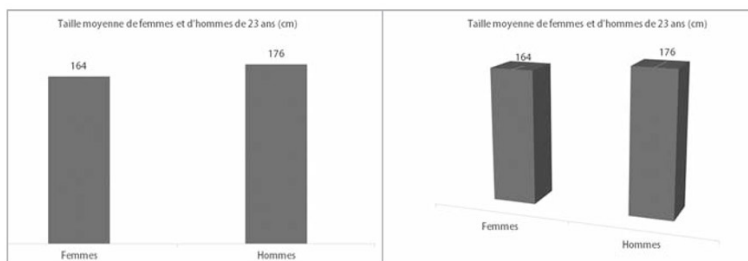


Figure 129 : Comparaison d'un diagramme en bâtonnets en 2D vs. 3D : Taille moyenne de femmes et d'hommes de 23 ans, chiffres fictifs, par les auteurs (CC0)

La 3D est aussi fréquemment utilisée dans les diagrammes circulaires. Cela n'a pas de sens non plus : ajouter de la 3D à ce type de diagrammes transforme le cercle en une ellipse : le cercle est en biais et la perception des angles est perturbée. Le lecteur ne perçoit donc pas clairement la variation de données. Une distorsion apparaît, comme sur la figure 130, utilisée comme exemple. Le même jeu de données est utilisé, mais on a l'impression que la part de 54 % est plus grande sur le graphique en 3D.

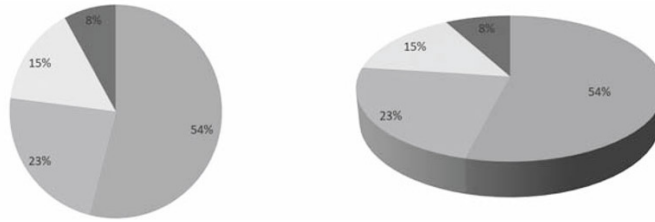


Figure 130 : Comparaison de diagrammes circulaires :
2D vs. 3D, par les auteurs (CC0)

5.3 L'ornementation utile ?

La notion de *chartjunk* pousse fortement au débat. Presque 20 ans après la première apparition du mot, une étude intitulée « Useful Junk ? The Effects of Visual Embellishment on Comprehension and Memorability of Charts » (Bateman *et al.*, 2010) remet le concept en question. Cette étude teste l'influence du *chartjunk* et du graphique minimaliste sur la compréhension et sur la mémorisation du lecteur. Les chercheurs ont comparé l'appropriation de graphiques minimalistes (figure 131) et de *chartjunk* de type « *duck* ». La figure 128 est issue de cette étude ; la figure 131 en est la version minimaliste.

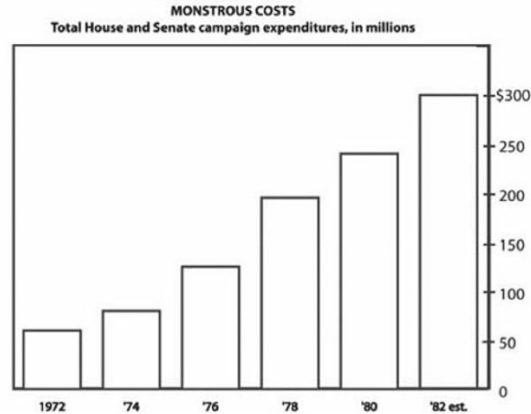


Figure 131 : « Monstrous costs », version minimaliste dans Bateman, *et al.* (2010, April). Useful junk? The effects of visual embellishment on comprehension and memorability of charts. In Proceedings of the SIGCHI conference on human factors in computing systems, p. 25

À propos des mêmes données, le *chartjunk* n'est pas forcément moins bien compris qu'un graphique minimaliste. Mais surtout, après 2 à 3 semaines, le *chartjunk* permet une mémorisation de l'histoire racontée par les données de manière nettement significative (Bateman *et al.*, 2010). L'ornementation peut donc avoir un effet positif. De plus, les participants comprenaient la valeur du message plus souvent dans les *chartjunk* que dans les graphiques minimalistes et surtout, ils ont souligné le plaisir de les utiliser, ce qui rendait la mémorisation plus facile (Bateman *et al.*, 2010). Few se permet néanmoins une remarque : « actuellement, les exemples dans cette étude ont consisté seulement en quelques graphiques embellis, créés par une seule personne – Nigel Holmes – et ils n'ont été un succès que dans un but extrêmement précis et limité » (Few S., 2011, p. 4, traduction libre). Few établit une critique méthodologique forte à l'encontre de l'article. Par exemple, que donnerait ce genre d'étude dans des conditions écologiques, en réalité ? Et en n'utilisant pas forcément des cas extrêmes de *chartjunk* ? (Few, 2011). Y aurait-il

donc différentes catégories ou niveaux de *chartjunk* ? Ces critiques le laissent penser.

D'ailleurs, les années suivantes, davantage d'études ont émergé. Ainsi, Borgo et ses collègues ont démontré que l'ajout de pictogrammes, d'images et de symboles aux graphiques minimalistes renforce la mémorisation, mais aussi la vitesse de rappel de l'information, bien que le temps de traitement au moment de la lecture soit plus long pour le *chartjunk*. La complexité des éléments figuratifs (images, pictogrammes) a un impact sur la performance du lecteur (Borgo *et al.* 2012). Haroz et ses pairs ont montré en 2015 que l'utilisation du pictogramme en visualisation de données impacte la mémoire, la vitesse de recherche de l'information, mais aussi les préférences des utilisateurs et leur engagement auprès de l'information (Haroz *et al.*, 2015). Selon l'équipe de Borkin (2016), les visualisations les plus mémorables en un coup d'œil sont celles qui reviennent rapidement en mémoire. Le titre et le texte sont des éléments essentiels de la visualisation, mais les éléments rapidement reconnaissables par l'œil humain (images et pictogrammes) tendent vers une reconnaissance et mémorisation rapide.

L'utilisation de couleurs encourage également la mémorisation. Le chapitre sur la couleur (4.8.) explique les principes de conception relatifs à la couleur.

5.4 L'esthétique des données : efficacité et plaisir

L'avis des différentes études qui prennent le contre-pied du *chartjunk* est souvent nuancé, mais elles tendent toutes généralement vers un constat semblable : le lecteur mémorise mieux un graphique qui comporte un élément figuratif ou des couleurs. Entre Tufte qui milite contre le *chartjunk* et les auteurs qui ont tendance à le dédramatiser, qui croire, que penser et quel comportement adopter ?

Plus loin que le *chartjunk* : l'efficacité et l'engagement

En vérité, la réflexion ne concerne pas simplement le *chartjunk*, mais l'efficacité de la visualisation de données en général. L'excellence graphique de Tufte et l'efficacité de Bertin tel qu'ils les définissent ont un point en commun : la représentation doit être comprise le plus rapidement possible, dans un minimum d'instant de perception et en utilisant le moins d'encre possible. Dans ce cas, bien sûr, il est logique de militer contre le *chartjunk*. Bertin est lui aussi réticent par rapport à l'ajout d'information cosmétique sur les représentations graphiques : au plus on ajoute de l'encre, et d'autant plus dans une intention décorative ou artistique, au plus le lecteur mettrait du temps à lire l'information, sans compter la distraction à laquelle il pourrait être sujet. Tufte et Bertin supposent que les lecteurs perçoivent tous l'information de la même manière ou, du moins, ils maximisent leurs efforts pour que ce soit le cas et que le message soit transmis de la manière la plus efficace possible.

Pourtant, tout pousse à penser aujourd'hui qu'il faudrait proposer une autre définition de l'efficacité des visualisations de données : tous les individus ne perçoivent pas les données de la même façon. Des éléments propres à leur expérience, à leurs préférences et caractéristiques vont forcément influencer leur lecture et leur engagement auprès de la visualisation de données. Ainsi, Kennedy et Hill estiment que différents facteurs humains et émotionnels entrent en jeu lorsque quelqu'un lit une visualisation de données. Ils sont répertoriés dans les tableaux suivants.

Tableau 5 : Facteurs humains, sociaux et propres à la visualisation, selon Kennedy & Hill, 2017

Le sujet de la visualisation	Une visualisation de données ne peut être présentée indépendamment de son sujet. Si les intérêts du lecteur sont touchés par la visualisation, il aura envie de s'engager. Sinon, il fera moins d'efforts.
La source d'où provient la visualisation	Si la visualisation présentée à un individu provient d'un blog en ligne qu'il adore, il y a de fortes chances qu'il s'y intéresse plus. <i>A contrario</i> , si elle vient d'un journal qu'il déteste, il aura des réticences avant même de commencer.
Croyances et opinions	L'individu s'engagera plus envers la visualisation si elle convient à ses croyances et opinions. La visualisation peut aussi élargir ses opinions, lui faire découvrir de nouvelles choses ou provoquer des remises en question.
Le temps	Avoir le temps suffisant pour traiter une visualisation est primordial.
La confiance en soi et ses compétences	Certaines personnes sont plus confiantes que d'autres en ce qui concerne leurs propres capacités à comprendre une visualisation de données. Le doute peut donc donner des motifs de réticence aux personnes qui n'ont pas cette confiance en leurs compétences de lecture et de compréhension de visualisations.
Les compétences langagières	Il est plus facile pour une personne de saisir une visualisation si elle en comprend la langue.
Les compétences mathématiques ou statistiques	Certains graphiques nécessitent ce type de compétences afin de comprendre des notions comme « moyenne » ou « échelle ». Les individus qui ont peu d'affinités avec ces compétences peuvent parfois se montrer réticents.
Les compétences de littératie visuelle	Elles aident l'individu à comprendre la signification attachée au visuel.
Les compétences informatiques	Elles sont utiles à l'individu quand il s'agit d'interagir avec la visualisation.

Les compétences de recul critique	Elles permettent à l'individu de se demander si le graphique est correct, si un élément y a davantage été mis en avant qu'un autre, de questionner l'information, etc.
--	--

Tableau 6 : Facteurs de réaction émotionnelle selon Kennedy & Hill, 2017

La visualisation en tant que visualisation	Les individus peuvent tout simplement réagir émotionnellement à une visualisation parce qu'elle en est une. Quelqu'un qui, par nature, n'aime pas les graphiques pourra se montrer agacé à sa vue par exemple. D'autres individus pourraient tout simplement réagir à la visualisation et exprimer une réaction émotionnelle parce que le visuel est plaisant.
Les données et le sujet traité	Imaginons qu'une visualisation traite du nombre de victimes du terrorisme sur une année et que le lecteur soit sensible à cette thématique. Il pourrait montrer des émotions comme de l'empathie, par exemple. De même, si un individu déteste une personnalité, sa réaction émotionnelle pourrait être de l'ordre de l'énervement. Son intérêt envers la visualisation pourrait être détourné, ou bien renforcé, mais accompagné d'une émotion négative.
La source ou le média	Certains médias peuvent sembler plus légitimes que d'autres à un individu, tout comme il peut aimer (ou pas) certaines sources.

Malgré tous les efforts du concepteur, la réception de la visualisation de données peut être différente. Bertin et Tufte misent sur la compréhension du message brut. Mais ne faudrait-il pas désormais impliquer l'engagement dans la question de l'efficacité ? Car les différents facteurs humains et émotionnels présentés ici favorisent ou défavorisent cet engagement. Est-ce bien utile de produire une visualisation claire et compréhensible si le public visé ne s'engage pas auprès d'elle ? La définition de l'efficacité gagnerait à être élargie. Le

débat est ouvert et laisse un peu de place à ce qu'on appellera « ornementation » plutôt que *chartjunk*.

Entre « *duck* » et ornementation épurée

Tous les *chartjunks* ne sont pas à mettre dans le même panier : de la distorsion du *data-ink* au simple ajout d'un pictogramme, il est possible de nuancer l'utilisation de l'ornementation. Plusieurs études ont donc montré l'intérêt de l'utilisation de la couleur ou d'un élément figuratif (pictogramme, image) : une mémorisation plus grande sur le long terme, un attrait visuel, la curiosité attisée, le plaisir d'utilisation, l'engagement et parfois même l'impression d'un effort minimisé (Andry, 2020). Néanmoins, si l'ornementation, qui est considérée comme du *chartjunk* par Tufte, n'est pas à bannir, elle est à utiliser avec prudence. Beaucoup de découvertes restent encore à réaliser sur le sujet, mais on sait aujourd'hui que :

- L'ornementation est fortement appréciée du lecteur lorsqu'elle est simple et épurée.
- L'ornementation qui est très chargée (beaucoup de bruit graphique) perturbe effectivement la réception.
- Les objets reconnaissables (images, pictogrammes) ont de nombreux effets positifs.

Bien souvent, le problème n'est pas seulement l'utilisation d'ornements, mais la mauvaise conception graphique de la visualisation. C'est la parfaite contre-application du principe de Tufte : « pensez aux variations de données avant de penser au design ». Ainsi, il est à supposer qu'en concevant la visualisation de données selon les principes de conception graphique et en y apportant de l'ornementation *a posteriori* et de façon simple et épurée, l'objectif de compréhension visé par l'excellence graphique et les bénéfices de l'ornementation seraient parfaitement associés. Le

résultat serait une visualisation compréhensible en un coup d'œil, élégante et qui pousse à l'engagement. De toute façon, il restera dans une certaine mesure toujours une inconnue, à savoir les caractéristiques personnelles du lecteur sur lesquelles on a finalement peu de contrôle.

De plus, Tufte parle également d'élégance graphique. Il dit :

L'élégance graphique se trouve souvent dans la simplicité du design et la complexité des données.

(Tufte, 2001, p. 177)

Ornementer son design n'en empêche pas la simplicité. Ainsi, la fameuse maxime de Tufte : « *Less is more* », s'applique aussi à l'ornementation : les visualisations présentant des ornements simples et épurés plaisent énormément aux lecteurs. Néanmoins, remplacer le corps-même du graphique par un ornement pourrait parfois perturber la compréhension (exemple : dans un diagramme en bâtonnet, dessiner des crayons à la place des bâtonnets).

Donc, utiliser des ornements ne doit pas :

- perturber la compréhension du corps du graphique.
- prendre la plus grande part d'espace visuel sur l'image.
- remplacer le *data-ink*.

Cela peut sembler contradictoire avec le principe de maximisation du *data-ink ratio*. En réalité, ça ne l'est pas tant : l'ornementation est un choix visant à rendre une visualisation attrayante et ludique. Son utilisation dépend du contexte de communication (*infra*) : elle est souvent inutile et obsolète. Néanmoins, lorsque son utilisation se justifie, il convient de l'employer avec parcimonie et en gardant en tête la simplicité préconisée par Tufte.

Pour terminer, avant même de se demander si l'ornementation pose problème, il est adéquat de vérifier si le choix de la technique de

visualisation est le bon et s'il sert au mieux le message. Plusieurs exemples sont critiqués ci-dessous.

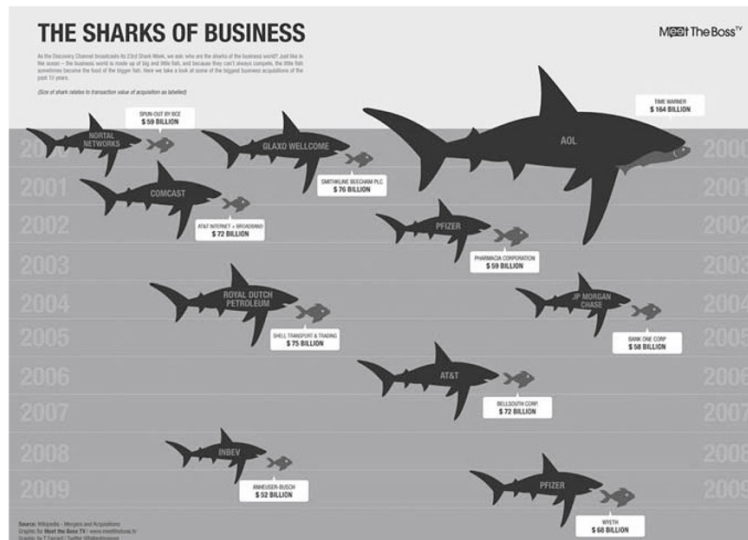


Figure 132 : The Sharks of Business, par T. Farant, 2010 (CC BY)

Cette visualisation se construit sur une métaphore : tout comme dans l’océan, le monde des affaires est constitué d’une hiérarchie de grands et de petits poissons, et que les petits poissons deviennent parfois la nourriture des plus gros. Le graphique est néanmoins chargé de couleurs, d’icônes et de nombreuses étiquettes de données. Le choix de conception nécessite l’exploration et que l’on s’attarde sur la visualisation. Il y a trois choses que les ornements ne permettent pas de savoir :

- La taille des requins est-elle la taille avant ou après avoir mangé le plus petit poisson ?
- Quelle est la taille des requins au final ?
- Les rapports de taille entre requins et poissons sont-ils respectés ? Difficile à dire à l’œil nu...

De plus, la dimension temporelle est ici anecdotique et donne l'illusion que l'information suit une évolution dans le temps. Or, le temps n'est pas une variable explicative.

Cette visualisation ne permet pas de saisir l'information en un coup d'œil. La métaphore visuelle passe ici avant l'illustration réelle des données.



Figure 133 : Cat Population, par les auteurs



Figure 134 : Fibre optique : quels pays sont les mieux équipés ?, par Tristan Gaudiaut, 2020 (CC0)

Très à la mode, le « pictogramme *chart* » peut causer un dérangement visuel. De plus, bien que précis, il alourdit l'information. Dans le cas de la figure 133 en particulier, l'addition des pictogrammes dans le visuel peut causer un grattement dans l'œil. L'étiquetage pourrait aussi être plus précis. Comme on le voit, le lecteur pourrait vite se perdre dans le décodage d'une information qui est pourtant très simple.

Les icônes représentant des drapeaux sont généralement appréciées du public. Dans la figure 134, l'idée de câbles qui sortent des bâtonnets du diagramme est bonne. Mais sont-ils porteurs d'information ou pas ? Est-ce du *data-ink* ou non ? Cela pourrait troubler le lecteur.

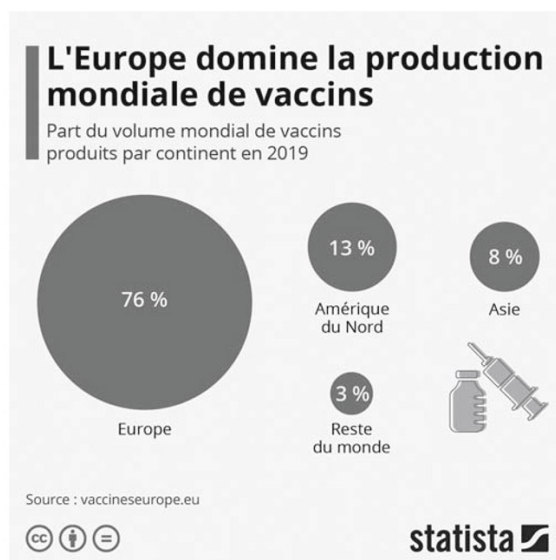


Figure 135 : L'Europe domine la production mondiale de vaccins, par Tristan Gaudiaut, 2020 (CC0)

Dans la figure 135, l'ornement est simple, épuré, et se situe à côté du *data-ink*. Il rappelle le sujet, il est cohérent et « non-invasif ». En revanche, les bulles ne sont pas du tout la meilleure façon de représenter l'information. La technique de visualisation a mal été choisie : pour montrer les différentes parties d'un tout (donc les parts

du volume mondial de vaccins produits par continent), un diagramme circulaire aurait permis de comparer en un coup d'œil les différentes parts au sein d'un même tout. Adopter la bonne technique de visualisation tout en gardant cette esthétique n'aurait pourtant pas enlaidi cette visualisation.



Figure 136 : La représentation des femmes dans les municipalités, par Tristan Gaudiaut, 2020 (CC0)

Un titre, une bonne légende, un étiquetage correct et un ornement minimaliste : autant dire que ce qui « décore » cette visualisation d'information dans la figure 136 n'est clairement pas dérangeant, tout en y apportant une certaine esthétique. La grille pourrait être enlevée si des étiquettes de données précisaient le chiffre exact entre les croisements de l'axe x et y. Néanmoins, l'auteur a ici fait le choix d'attirer l'attention du lecteur sur le chiffre final.

Quel objectif de communication ?

Évidemment, l'ornementation d'un graphique dépend de son contexte d'utilisation. Les pictogrammes peuvent dynamiser une présentation dans le monde professionnel, mais on imagine mal un directeur


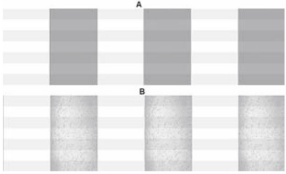
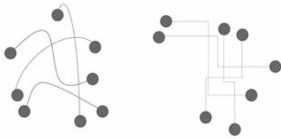
financier présenter son information à l'aide de petits requins. Ainsi, à chaque visualisation son contexte d'utilisation. Dans un journal, un magazine, à des fins d'information, l'ornementation pourra parfois se justifier. La visualisation de données est parfois même utilisée dans des dispositifs publicitaires, où cela se justifie pleinement. Cependant, il y a certains contextes informationnels où l'ornementation n'a pas forcément sa place. En somme, il est toujours bon de se questionner à propos du public qui va lire la visualisation de données et dans quel contexte.

6. La reconnaissance de patterns

6.1 L'application des principes Gestalt

Dans l'exploration des données, percevoir un *pattern* peut souvent mener à une idée clé : c'est la plus grande force de la visualisation de données. La compréhension de la perception des formes (patterns) fournit des règles de conception abstraites qui peuvent en apprendre beaucoup sur la façon dont il faut organiser les données de manière à ce que les structures importantes soient perçues (Ware, 2004). Dans le chapitre abordant la perception visuelle, les différentes lois de la Gestalt Theory sont présentées. Ces différentes lois permettent de comprendre différents principes de design relatifs à la perception. Elles ont été complétées et commentées par Ware (2004).

Tableau 7 : Application des principes Gestalt au design de visualisations de données

Description du principe	Application en termes de design
<p>Loi de proximité</p>  <p>Figure 137 : Loi de proximité, par les auteurs (CC0)</p>	<p>Les éléments représentés proches les uns des autres sont perçus comme un groupe. Le moyen le plus simple et le plus puissant afin de mettre en évidence les relations entre différentes entités de données est de les placer à proximité dans un écran (Ware, 2004, p. 190).</p>
<p>La loi de similitude</p>  <p>Figure 138 : Application de la loi de similitude par utilisation de la couleur et du grain, par les auteurs (CC0)</p> <p>Les éléments similaires sont perçus comme groupés ensemble.</p>	<p>La figure 138 montre deux façons différentes de séparer visuellement les informations des lignes et des colonnes. En A, le codage intégral en couleur et en niveaux de gris est utilisé. En B, la couleur jaune est utilisée pour délimiter les lignes et le grain est utilisé pour délimiter les colonnes. La couleur et le grain permettent de différencier. On voit alors un <i>pattern</i> qui peut être visuellement segmenté soit par des lignes ou par colonnes. Le lecteur peut ainsi facilement suivre l'un ou l'autre modèle (Ware, 2004, p. 191).</p>
<p>Loi de continuité</p>  <p>Figure 139 : Application de la loi de continuité : trajectoires lisses (gauche) vs. brusques changements de direction (droite), par les auteurs (CC0)</p>	<p>L'œil suit les trajectoires naturellement présentes dans l'image et il est plus susceptible de construire des entités visuelles à partir d'éléments visuels lisses et continus plutôt que de ceux qui contiennent des changements de direction brusques.</p> <p>Le principe de la bonne continuité peut être appliqué aux diagrammes nœud-lien. Il devrait être plus facile d'identifier les sources et les destinations des lignes de connexion si elles sont lisses et continues (Ware, 2004, p. 192).</p>

Loi de fermeture

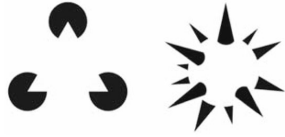


Figure 140 : Gestalt 7 : Closure (Domaine public)

Le contour fermé a tendance à être vu comme un objet. L'œil peut aussi compléter lui-même l'espace par prolongement de la courbe ou si la forme lui est familière.

Le diagramme de Venn est une parfaite application de la loi de fermeture : nous percevons à la fois les deux cercles et l'ensemble commun qu'ils créent grâce à la fermeture causée par les deux formes (cercles qui se croisent). Le lecteur interprète la région à l'intérieur d'un contour fermé comme définissant un ensemble d'éléments. Les principes les plus importants pour ce diagramme sont la fermeture et la continuité (Ware, 2004, p. 194).

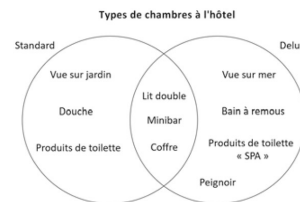


Figure 141 : Diagramme de Venn, par les auteurs (CC0)

Connectivité

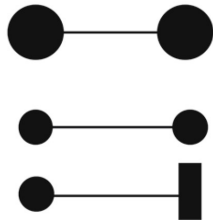


Figure 142 : La connectivité, par les auteurs (CC0)

Plus puissante que la proximité ou la similarité, la connectivité relie différents objets graphiques au moyen de lignes.

Relier différents objets graphiques par des lignes est un moyen très puissant d'exprimer qu'il existe une certaine relation entre eux. En effet, cela est fondamental pour le diagramme de nœuds et de liens, l'une des méthodes les plus courantes pour représenter les relations entre les concepts (Ware, 2004, p. 191).

Symétrie

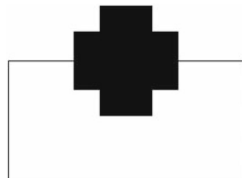
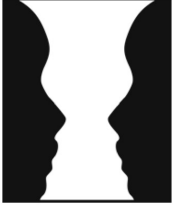


Figure 143 : Croix symétrique et rectangle, par les auteurs

Une application possible de la symétrie est dans les tâches où l'on recherche des similitudes entre deux ensembles différents de données temporelles. Il peut être plus facile de percevoir des similitudes si ces séries chronologiques sont disposées en utilisant la symétrie verticale (Ware, 2004, p. 193-194).

Permet la régularité entre les différentes parties d'un objet.	
<p>Ségrégation fond-forme</p>  <p>Figure 144 : Le vase de Rubin, par Bryan Derksen, 2007 (CC BY-SA)</p> <p>Une figure est un objet perçu comme étant au premier plan. Le fond est tout ce qui est perçu comme se trouvant derrière la figure.</p>	En conception graphique, il faut donc s'assurer que les motifs seront bien perçus comme des figures.

En résumé :

- S'assurer que les motifs soient bien perçus comme une figure et qu'il n'y ait pas de confusion → ségrégation fond-forme.
- Placer les symboles représentant les informations connexes à proximité les uns des autres → proximité et similarité.
- Utiliser la fermeture pour alléger le design.
- Éviter les changements de direction brusque lorsque l'on relie deux objets par une ligne → continuité et connectivité.

6.2 Les 10 tâches perceptuelles élémentaires (Cleveland & McGill, 1984)

En 1984, Cleveland et McGill ont conçu une liste de 10 tâches perceptuelles élémentaires, chacune étant une méthode pour représenter des données. Ils ont classé ces tâches en fonction de la précision avec laquelle le cerveau humain peut détecter les différences et faire des comparaisons. La figure 145 montre les tâches perceptuelles élémentaires, de la plus haute à la plus basse précision. Autrement dit, il s'agit du jugement de la précision des données

quantitatives, du moins précis au plus précis. Les tâches sont regroupées en fonction de la manière dont vous pouvez percevoir les différences dans les données en les utilisant.

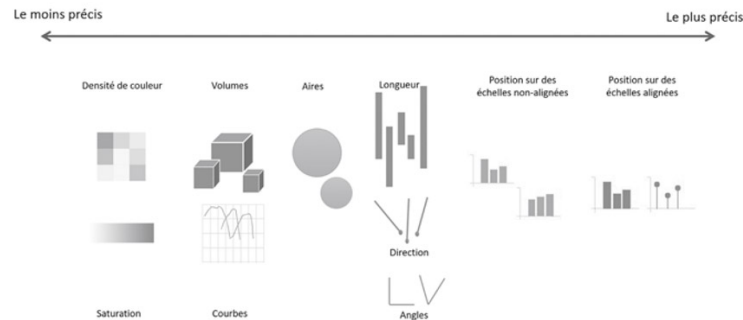


Figure 145 : Précision du jugement des données quantitatives.
Traduction de la figure originale de Cleveland et McGill
(1984), par les auteurs (CC0)

Les tâches perceptuelles sont les suivantes :

- position sur une échelle commune ;
- position sur des échelles non alignées ;
- longueur, direction et angle ;
- aires ;
- volume, courbes ;
- densité et saturation des couleurs.

Une forme graphique qui implique des tâches perceptuelles élémentaires conduisant à des jugements plus précis qu'une autre forme graphique (avec les mêmes informations quantitatives) aboutira à une meilleure organisation et augmentera les chances d'une meilleure perception des modèles. Un diagramme en bâtonnet est toujours supérieur à un diagramme à bulles si le but du graphique est de faciliter des comparaisons précises (Cairo, 2012).

Ainsi, plus on se déplace vers la droite sur l'échelle de Cleveland et de McGill (1984, figure 145), plus les lecteurs pourront porter des

jugements précis. Mais il arrive que l'objectif ne soit pas de permettre des comparaisons précises ou de classer des valeurs, mais de faciliter la perception de modèles plus larges. On ne le dira jamais assez : la représentation graphique doit correspondre à un message clair et précis. Le classement des tâches perceptuelles de Cleveland et de McGill est un excellent outil pour concevoir des visualisations de données de façon raisonnée plutôt que sur le seul goût esthétique. Il existe cependant des nuances et des exceptions en fonction du contexte et des circonstances (Cairo, 2012).

7. La couleur

7.1 Les dangers de la couleur : exemple

Les couleurs sont peut-être la propriété visuelle que les gens utilisent le plus souvent à mauvais escient dans la visualisation de données sans même s'en rendre compte. Par exemple, les variations des couleurs de l'arc-en-ciel dans les visualisations de données sont très populaires, mais en même temps les plus problématiques et les plus trompeuses pour les lecteurs. La figure 146 en montre un exemple.

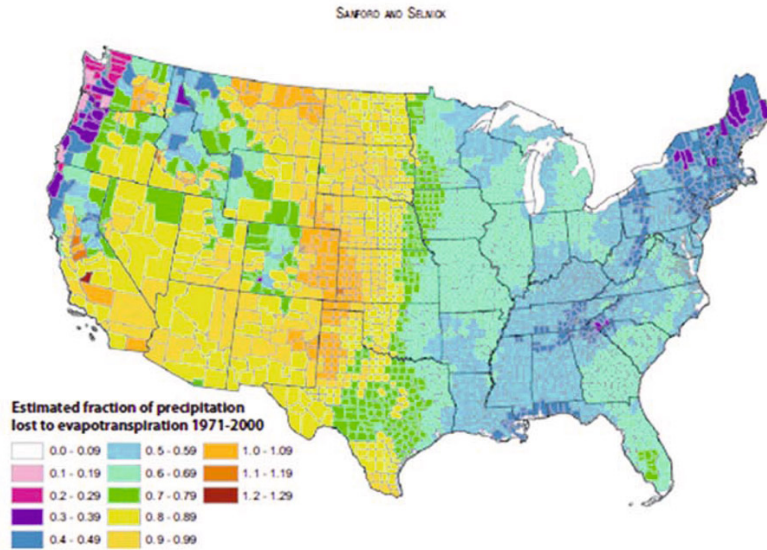


Figure 146 : Estimation du rapport annuel moyen entre l'évapotranspiration réelle (ET) et les précipitations (P) pour les États-Unis contigus pour la période 1971-2000.

Crédit image : Sanford & Selnick,
<https://doi.org/10.1111/jawr.12010>

En effet, sur cette figure, la valeur maximale est difficile à trouver. Le choix de la couleur ne la rend pas très saillante. De plus, la couleur rouge foncé ressemble quelque peu à la couleur mauve utilisée. Ces deux couleurs semblent similaires et encodent en réalité des valeurs très différentes. Mais encore, on remarque au milieu de la carte une scission claire entre deux nuances de vert. Une « frontière » semble dessiner clairement une différence forte entre deux valeurs. Pourtant, la légende montre que la différence entre les valeurs codées par ces couleurs n'est pas énorme. Utiliser des couleurs de teinte similaire et dont la luminosité évolue aurait limité cet effet. Attention toutefois : la différence perçue entre deux couleurs n'est pas toujours proportionnelle à la distance réelle sur le spectre visible des couleurs. Pour rappel, la perception des couleurs est abordée dans le chapitre sur la perception visuelle. Dans les ressources externes se trouve un

outil d'aide de choix de couleurs (*ColorBrewer*) dans le domaine de la cartographie.

Bien qu'attrayante et jouant un réel rôle d'engagement auprès du lecteur, la couleur n'est pas toujours nécessaire à la transmission de l'information : il n'y a pas forcément besoin de couleur lorsque d'autres variables visuelles aux mêmes propriétés perceptives peuvent être utilisées.

7.2 La couleur en contexte

Comme cela a été présenté dans le chapitre sur la perception visuelle, le contexte dans lequel la couleur est utilisée peut en influencer la perception. Voici de nouveau trois exemples :

1. Il est difficile de lire le texte lorsque **la luminosité est égale entre le texte et le fond**, malgré une grande différence chromatique (Ware, 2004, p. 115). La figure 147 l'illustre bien.

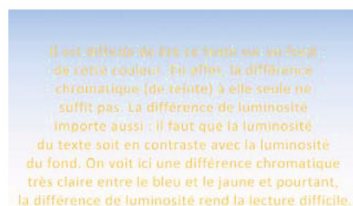


Figure 147 : Lecture de texte entre deux couleurs différentes d'un point de vue chromatique, avec variation de luminosité, par les auteurs, inspiré de (Ware, 2004, p. 115) (CC0)

2. En fonction du **contexte** dans lequel se trouve la couleur, c'est-à-dire des autres couleurs environnantes, une couleur pourra être perçue **plus vive ou moins vive**. Sur la figure 148, le carré central de droite semble ainsi plus lumineux.



Figure 148 : Contraste de saturation, par les auteurs. (CC0)

3. Le contraste de luminosité simultané s'applique également aux nuances de gris, comme sur la figure 149.

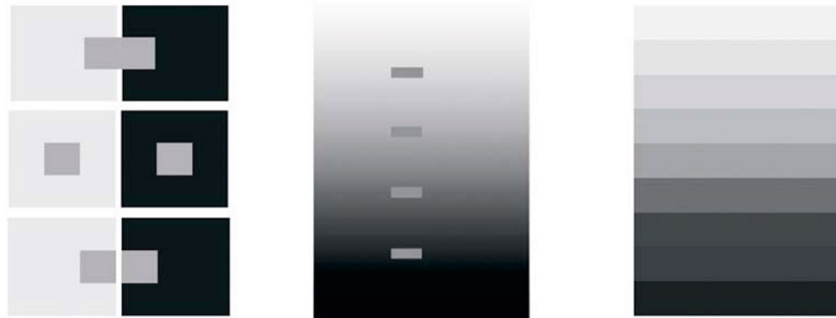


Figure 149 : Les nuances de gris et les contrastes, par les auteurs (CC0)

Ainsi, le contraste simultané avec les couleurs d'arrière-plan peut modifier considérablement l'apparence des couleurs, faisant ressembler une couleur à une autre. C'est une autre raison pour laquelle il est conseillé de n'avoir qu'un petit ensemble de codes de couleur. Une méthode pour réduire les effets de contraste consiste à placer une fine bordure blanche ou noire autour de l'objet codé en couleur. Ce dispositif est couramment utilisé avec les feux de signalisation ; par exemple, les signaux des trains sont affichés sur de grands disques à fond noir. En outre, nous ne devrions jamais afficher de codes utilisant des différences purement chromatiques avec le fond. Il devrait y avoir une différence de luminosité significative en plus de la différence de couleur.

7.3 Taille des zones colorées

Il faut éviter d'encoder la couleur sur de petits objets ou de petites zones : cela la rend difficilement perceptible. Plus la zone est grande, plus il est facile de distinguer les couleurs. Les petits objets sur lesquels la couleur est codée doivent avoir des couleurs fortes et très saturées pour une discrimination maximale, c'est-à-dire pour qu'ils soient bien perceptibles.

Lorsque de grandes zones de codage couleur sont utilisées, par exemple, avec des régions cartographiques, les couleurs doivent être peu saturées et ne différer que légèrement les unes des autres. Cela permet de percevoir de petites cibles de couleur vive par rapport aux régions de fond.

Lorsque les couleurs sont utilisées pour mettre en évidence des régions de texte noir, elles doivent être claires (contraste de luminance minimum avec le papier blanc) et également de faible saturation. Cela permet de réduire les interférences avec le texte.

La figure 150 montre bien qu'il est plus difficile de percevoir la couleur sur une plus petite zone et que cela dépend du fond.

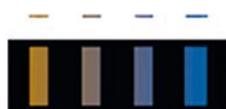


Figure 150 : Color field, inspirée de Ware (2004),
par les auteurs

7.4 L'inversibilité des couleurs

Le choix des couleurs devrait être facilement inversible. Il faut donc éviter les choix de couleurs :

- Dont la teinte, la saturation et la valeur se ressemblent.
- Qui sont perçues comme similaires (attention au daltonisme) → les couleurs qui semblent éloignées ou très différentes dans le

spectre visible des couleurs ne sont peut-être pas forcément perçues comme telles.

- Qui sont tout simplement difficiles à percevoir (par exemple, les couleurs sombres ou fortement désaturées).

En somme, il faut veiller à utiliser pleinement le spectre visible des couleurs (et éviter de se concentrer uniquement sur la teinte). Les différences de teinte, de saturation et de luminosité (valeurs) doivent être élevées plutôt que faibles. Le jaune et le violet, ou encore le vert et le rouge sont des couleurs opposées et facilement distinguables (inversibles). La figure 151 évoque l'évolution du spectre en fonction de la luminosité (valeur), de la teinte et de la saturation. Attention toutefois au message que vous souhaitez communiquer ! Si vous souhaitez illustrer des données ordinales, la variation de luminosité (valeur) sur une même teinte permettra mieux d'illustrer les variations de données (*infra*).

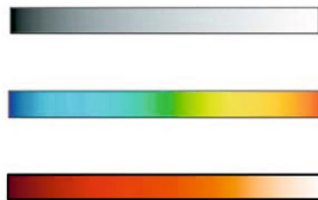


Figure 151 : Évolution du spectre en fonction de la luminosité (valeur), de la teinte et de la saturation, par les auteurs (CC0)

7.5 Représenter le type de données

Les différentes propriétés des couleurs permettent d'encoder les données de sorte que leur compréhension soit intuitive. Les choix de couleurs seront donc différents pour représenter des données nominales, ordinales, d'intervalle ou de ratio :

- La teinte convient pour représenter des données nominales.

- La saturation convient pour représenter des données ordinales, d'intervalle et de ratio
- La luminosité convient pour représenter des données ordinales, d'intervalle et de ratio.

La teinte, la saturation et la luminosité sont représentées sur la figure 152.

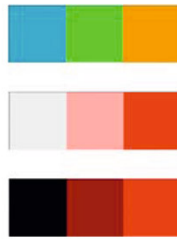


Figure 152 : HSV, par les auteurs (CC0)

Plus précisément, trois principes sont à garder en tête.

1. **Les palettes de couleurs séquentielles supposent un ordre** (figure 153).

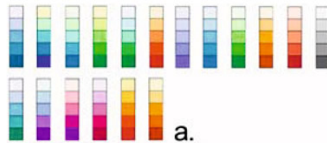


Figure 153 : Palettes de couleurs séquentielles choisies grâce à colorbrewer.org, par les auteurs (CC0)

Ces palettes conviennent pour représenter des données allant de valeurs faibles à élevées, soit sur une échelle ordinale (par exemple, froid, chaud, chaud), soit sur une échelle numérique (par exemple, classes d'âge de 0 à 9, 10 à 19 ans, etc.). Les différents stades de luminosité dominant l'aspect de ces palettes, généralement avec des couleurs claires pour les valeurs faibles et des couleurs sombres pour les valeurs élevées. Utiliser les couleurs sombres pour les plus hautes valeurs est une convention cartographique standard (« *Dark equals*

more »). Les palettes séquentielles peuvent être soit monochromes (par exemple, le même bleu, avec différents niveaux de clarté et de saturation), soit multicolores (par exemple, du jaune clair au vert foncé).

2. Les palettes de couleurs divergentes supposent un point de rupture évident (figure 154).

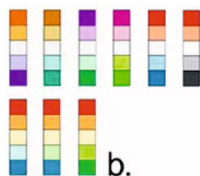


Figure 154 : Palettes de couleurs divergentes choisies grâce à colorbrewer.org, par les auteurs (CC0)

En jouant sur la teinte et la luminosité, ces palettes conviennent pour mettre en évidence un changement dans les classes de données critiques¹⁰ (ou un point de rupture). Il s'agit toujours des séquences de plusieurs teintes. Attention cependant, en raison de la façon dont la luminosité varie, ne fait pas de bonnes photocopies en noir et blanc.

3. Les palettes de couleurs qualitatives supposent l'apparition de catégories (figure 155).

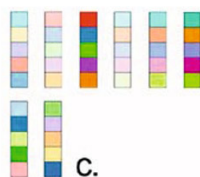


Figure 155 : Palettes de couleurs qualitatives choisies grâce à colorbrewer.org, par les auteurs (CC0)

Ces palettes s'appuient principalement sur les différences de teinte, tandis que la saturation et la luminosité sont maintenues constantes. Comme il n'y a pas de classement conceptuel dans les données

nominales, il est inapproprié d'utiliser ces palettes de couleur pour représenter les données nominales (par exemple, en utilisant une séquence de teintes uniques claires à foncées). Par contre ; évitez d'utiliser des couleurs « néon » fortes (c'est-à-dire à forte saturation) et des couleurs pastel (c'est-à-dire à faible saturation) dans le même schéma de couleurs qualitatif, car ces variations de saturation pourraient impliquer un ordre également.

Exemple :

Dans le cadre des élections présidentielles de 2004 aux États-Unis, cette carte choroplèthe présentée en figure 156 montre une utilisation pertinente des couleurs :

- Le rouge et le bleu sont opposés et supposent l'apparition de catégories : le bleu pour la victoire de Johnn Kerry et le rouge pour la victoire de George Bush dans les différents états.
- La variation de saturation pour illustrer le pourcentage de voix.

**Vote percentage difference in 2004 US Presidential Election as
Related to Voter Turnout**

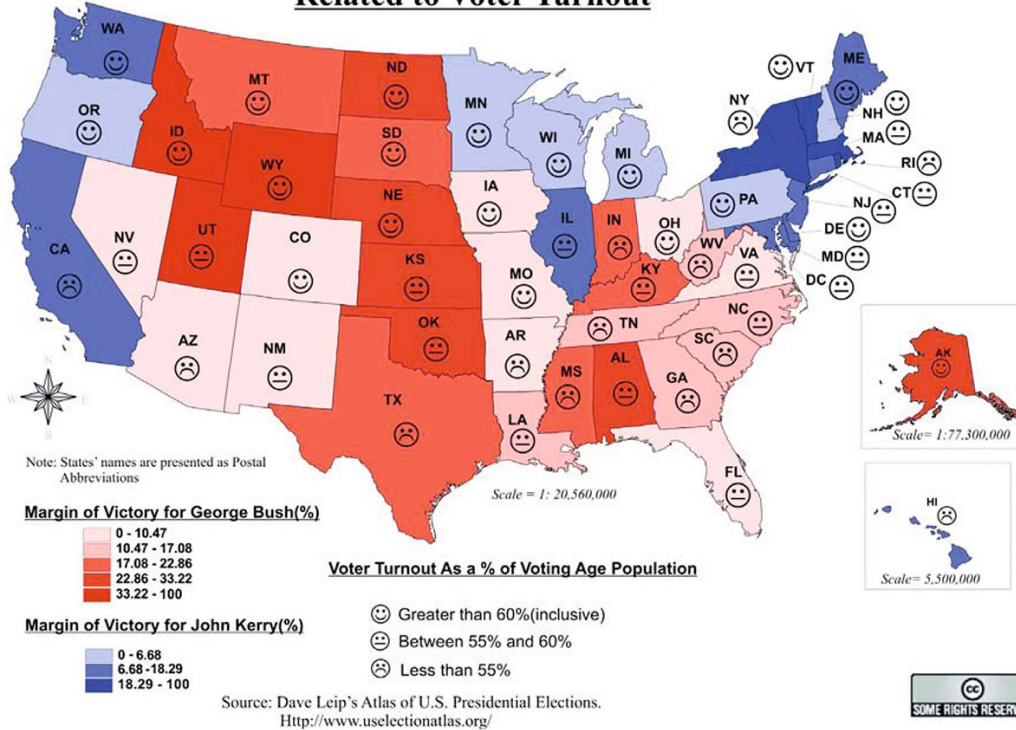


Figure 156 : Pourcentage des voix aux élections présidentielles de 2004 aux USA, Dave Leip, 2005 (CC BY-SA)

7.6 Différencier et explorer

Utiliser la couleur permet d'attirer l'attention sur un point d'intérêt, dans une optique plutôt explicative qu'exploratoire. On constate ainsi que la figure 158 oriente directement le lecteur vers le résultat le plus haut, contrairement à la figure 157. Le design doit être réalisé en fonction de ce que vous devez mettre en valeur, comme :

- des fourchettes de valeurs spécifiques ;
- des valeurs spécifiques ;
- le taux de variation de la valeur ;
- ou tout autre élément qui est jugé important.

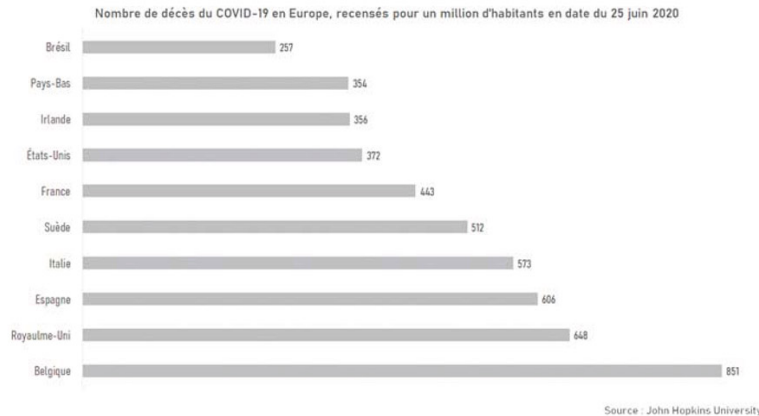


Figure 157 : Nombre de décès du COVID-19 en Europe, recensés pour un million d'habitants en date du 25 juin 2020, par les auteurs (CC0)

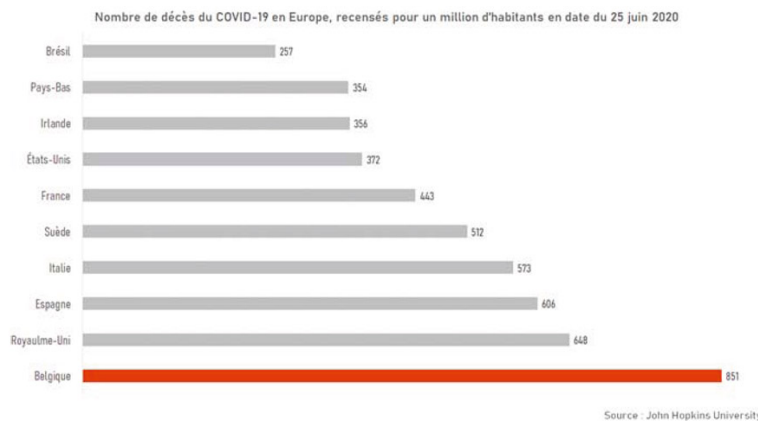


Figure 158 : Nombre de décès du COVID-19 en Europe, recensés pour un million d'habitants en date du 25 juin 2020, résultat belge mis en évidence, par les auteurs (CC0)

7.7 Pour conclure...

Finalement, il n'existe pas de règles universelles de design concernant la couleur, mais certaines lignes directrices assez techniques sont à garde en tête.

7.7.1 Quelles couleurs et combien ?

Le rouge, le vert, le jaune et le bleu reconnus par le cerveau comme primaires. S'il est nécessaire de se souvenir d'un code de couleur, ces

couleurs sont les premières à prendre en compte. Il est préférable de n'utiliser que quelques couleurs s'il s'agit de codes bien distincts. Le rouge, le vert, le jaune et le bleu, ainsi que le noir et le blanc offrent des choix naturels lorsqu'un petit ensemble de codes de couleur est nécessaire. Bien que le codage couleur soit un excellent moyen d'afficher des informations sur les catégories, seul un petit nombre de codes peut être rapidement perçu. Les estimations varient entre environ cinq et dix codes (Ware, 2004 ; p. 125). Si la couleur du fond est variée, essayer d'utiliser plus de 12 couleurs comme codes risque de se solder par un échec. Il est également préférable d'utiliser une correspondance individuelle entre les valeurs et les couleurs.

7.7.2 Le contraste

Pour montrer les détails dans une visualisation, il faut toujours avoir un contraste de luminance considérable entre les informations d'avant-plan et d'arrière-plan. Faire uniquement la différence par la variation chromatique est une erreur. Cela devrait être évident dans le cas d'un texte, bien que de nombreuses présentations PowerPoint enfreignent encore cette règle. Des bordures noires ou blanches autour des symboles colorés peuvent contribuer à les rendre distincts en assurant une rupture de contraste de luminance avec les couleurs environnantes. Le contraste des couleurs peut entraîner de grandes erreurs dans la représentation de la quantité. Les erreurs de contraste peuvent être réduites par des bordures autour de zones sélectionnées ou par l'utilisation d'arrière-plans sourds et relativement uniformes.

7.7.3 Premier plan et arrière-plan

Il est pertinent d'utiliser la saturation pour faire apparaître certains éléments et d'utiliser des couleurs plus saturées pour les petits symboles, les lignes fines ou les petits objets/zones ainsi que des couleurs moins saturées (c'est-à-dire des couleurs sourdes) pour les

grandes zones. La luminance peut modifier considérablement l'apparence des couleurs, mais il faut veiller à éviter le contraste de luminance simultané avec les couleurs de fond. Il est aussi important d'utiliser un niveau minimum de différence de luminance pour soutenir le contraste entre le premier plan et l'arrière-plan et pour montrer les détails dans la visualisation.

7.7.4 Palette de couleurs

Choisissez une palette de couleurs qui correspond au message à transmettre ainsi qu'au type de données utilisé.

Résumé

**Créer une visualisation respectueuse
des principes de conception : *checklist***

Jusqu'ici, vous avez pris connaissance de la plupart de règles qui existent en termes de conception de visualisation de données. Pour résumer, différentes questions, proposées sous forme de *checklist*, peuvent vous permettre de vérifier si votre représentation graphique est améliorable ou non.

1. La technique de visualisation

En relation avec le chapitre 3

- De quelles données disposez-vous ?
- Quel est le message que vous souhaitez communiquer ?
- Quelle est la technique de visualisation la plus adéquate pour représenter de telles données ?

2. Principes de base

Avez-vous vérifié l'intégrité de votre graphique ?

- Citer les données hors contexte : interdit.
- *Lie factor* à vérifier : la proportionnalité entre les nombres réels et les nombres représentés doit être respectée.
- Diagrammes à axes (x, y) : votre axe y commence-t-il bien à 0 ?
- Étiquetage clair et précis : votre graphique est-il correctement détaillé, fourni en informations ?
- Les données avant le design.
- Déflater et standardiser les unités monétaires.

Avez-vous pensé au *data-ink ratio* ?

- Est-il maximisé ?
 - *Sinon pourquoi, quelle en est la raison ?*
- Avez-vous utilisé de l'ornementation ?
 - *Si oui : se justifie-t-elle ? Quel est votre objectif de communication ?*

3. Variables visuelles

- Avez-vous encodé vos informations en utilisant les meilleures variables visuelles ?
- Souhaitez-vous stimuler la perception associative, sélective, ordonnée ou quantitative ? Avez-vous utilisé les variables visuelles qui y correspondent ?

4. Reconnaissance de *patterns*

- Avez-vous vérifié que la compréhension spatiale de votre design soit préattentive ?

- *Exemple : s'il y a des liens dans votre représentation, sont-ils représentés sous forme de courbe ou présentent-ils des cassures nettes ?*
- Les formes qui doivent être perçues comme étant semblables sont-elles représentées ensemble, proches les unes des autres ?
- Votre représentation graphique engendre-t-elle des illusions d'optique et des dérangements visuels ?
- Les formes que vous avez employées permettent-elles une grande précision dans le jugement des données quantitatives ?

5. La couleur

- La couleur utilisée encode-t-elle de l'information ?
- Si vous souhaitiez mettre une information en évidence par rapport aux autres, l'avez-vous fait grâce à la couleur ? Était-ce le moyen le plus approprié de le faire ?
- La palette de couleurs choisie correspond-elle à la perception que vous souhaitez stimuler (ordinaire ou catégorielle) ?
- Les couleurs utilisées engendrent-elles des dérangements visuels (couleurs éclatantes, dérangements visuels) ?
- Avez-vous vérifié que la couleur utilisée dans son contexte final soit bien celle que vous vouliez représenter initialement ?

Si vous savez répondre correctement à toutes ces questions et justifier vos réponses, votre conception graphique aura de grandes chances d'être de très bonne qualité.

Ressources externes et bibliographie indicative

Ressource interactive

Color Brewer 2.0 : Conseils en couleur pour la cartographie ¹¹

Bibliographie indicative

Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A., McDine, D., & Brooks, C. (2010, April). Useful junk? The effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2573-2582)

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387), 531-554.

Metz, C. (1971, August). Réflexions sur la « Sémiologie graphique » de Jacques Bertin. In *Annales. Histoire, Sciences Sociales* (Vol. 26, No. 3-4, pp. 741-767). Cambridge University Press.

Tufte, E. R. (2001), *The visual display of quantitative information* (Vol. 2), Cheshire, CT: Graphics press.

Ware, C. (2004). *Information visualization: perception for design*. Second Edition. Morgan Kaufmann

CONCLUSION

Depuis de nombreuses années, les usages socionumériques et l'utilisation de technologies numériques sont si foisonnants qu'on s'attend à une explosion exponentielle des données au cours du temps. Toutes ces données permettront des avancées technologiques et sociales dans bien des domaines. Mais, en premier lieu, il s'agit d'information brute. Pour comprendre le message contenu dans les données numériques, il est très commun de les visualiser. Dans de nombreuses sphères (entreprises, médias, politique, etc.), visualiser les données permet de les interpréter et de les comprendre. Cela permet également de les communiquer à une audience qui pourra comprendre une information claire et condensée : la visualisation de données est le médium principal par lequel les gens ont accès aux données (Kennedy et Hill, 2018). En outre, la visualisation de données acquiert également une place grandissante dans les espaces médiatiques numériques. Dans la presse en ligne, sur les blogs et les réseaux sociaux, elle gagne du terrain, notamment grâce à la puissance visuelle des infographies qui permettent de communiquer l'information rapidement dans un espace numérique très saturé. La visualisation de données se développe donc dans un contexte d'évolution intense. Notre environnement actuel est imprégné d'infobésité, comprise comme une surcharge d'information (Sauvajol-Rialland, 2014). Les graphiques ou visualisations d'information sont

un bon moyen de communiquer quelque chose rapidement et d'intéresser les gens. Ils sont à la fois un gain de temps et d'effort cognitif. Dans ce contexte de surcharge d'information, ils sont souvent utilisés. L'infobésité frappe également les organisations et les entreprises, où la quantité de données produite crée une course à l'information qui est devenue vitale pour rester efficace. C'est pourquoi les tableaux de bord d'analyse d'activité, rassemblant diverses visualisations de données, sont très utilisés aujourd'hui pour établir des rapports, prendre des décisions et comprendre globalement cette grande masse de données.

Aujourd'hui, nous avons donc tendance à mettre les choses sous forme de chiffres, de données, puis à les visualiser pour les comprendre, ce qui peut donner une impression de neutralité ou, comme le disent Kennedy et Hill (2018), une vision aseptisée de la réalité. Les données, sous leur forme visualisée, semblent objectives ; il semblerait en effet que « les chiffres parlent d'eux-mêmes ». Leur utilisation médiatique s'explique donc également par une dynamique d'intérêt, de rapidité et de conviction. Pourtant, la visualisation de données qui se veut totalement objective n'existe pas parce qu'elle est le résultat d'étapes qui visent à construire l'interprétation d'un phénomène : les données sont une construction, puisqu'elles sont extraites d'une activité. Elles sont ensuite nettoyées et exploitées dans un but précis : être représentées dans un graphique pour expliquer une histoire, pour communiquer une information. La visualisation de données exprime un message, pas un fait en soi. Elle est donc le résultat de choix visant à communiquer un message.

Dès lors, dans différents domaines et contextes, comment les concepteurs des visualisations de données font-ils pour que leurs visualisations soient les plus efficaces possibles ? En vérité, ils font assez peu appel à des référentiels ou à des guides qui existent,

principalement en langue anglaise. Ils s'y réfèrent peu parce que d'une part, le temps manque et que, d'autre part, ils ne s'y retrouvent pas forcément. Pourtant, de nombreux principes de conception des visualisations de données existent, notamment parce que la visualisation de données est loin d'être un phénomène neuf. Depuis longtemps, l'homme utilise la schématisation pour organiser sa pensée et la communiquer. William Playfair, en 1786, concevait pour la première fois et de manière aboutie les trois graphiques les plus connus au monde : le diagramme en ligne, le diagramme en barres et le diagramme circulaire. De nombreuses pratiques en visualisation de données ont ensuite foisonné. Ce n'est donc pas la visualisation de données qui est nouvelle, mais plutôt le regard que nous portons sur les données puisque nous les visualisons afin de les interpréter (Cardon, 2012). Transcrire des données statistiques en une forme visuellement compréhensible est un enjeu stimulé par la diversité d'outils, de techniques, de pratiques et de tendances.

À travers cet ouvrage, nous avons souhaité combler ce manque de référentiels en français, et ce, pour les débutants en visualisation de données et en communication, mais aussi pour les professionnels aguerris qui ont besoin d'une piqure de rappel dans le cadre de leur pratique.

Avoir conscience que la visualisation de données est une pratique sociale millénaire et comprendre comment fonctionne la perception visuelle, en ayant notamment connaissance de différentes notions comme le traitement préattentif ou encore les différentes lois de la *Gestalt Theory*, prédispose à réaliser de bons choix de conception en visualisation de données.

Selon Tufte (1983), la plus grosse erreur que l'on puisse commettre serait de penser au design et au look de sa visualisation avant même de penser au message qu'elle communique. En lui donnant raison,

nous avons proposé, dans le troisième chapitre de cet ouvrage, de se poser un temps pour réfléchir aux types de variables et aux types de données qui entrent en jeu dans la conception d'une visualisation de données. Il s'agit là des matériaux fondamentaux qui aident à choisir une technique de visualisation, à savoir *le bon graphique* qui arrivera parfaitement à transcrire le message que vous souhaitez délivrer. Au cours de votre pratique, revenez fréquemment consulter le catalogue, et voyez si vos objectifs de communication peuvent être atteints grâce à la technique de visualisation que vous comptez utiliser.

Après ces étapes fondamentales vient l'heure du design : quel enrobage ajouter à cette fameuse technique de visualisation ? C'est le moment de penser aux titres, aux légendes, aux échelles, à l'utilisation des couleurs, et peut-être même à quelques éléments d'ornementation du rendu final. Attention néanmoins, de nombreux principes de conception de visualisations de données existent. Ils permettent de cadrer la conception de la visualisation : les surfaces représentées sur votre graphique sont-elles proportionnelles aux données représentées ? Le message principal de votre représentation transparait-il clairement, dans un environnement épuré ? N'abusez-vous pas du *chartjunk* ? En contrôlant toutes ces notions comme le *lie factor*, le *data ink ratio*, les variables visuelles, *le chartjunk* ou encore les tâches perceptuelles élémentaires, vous vous assurez de réaliser une visualisation qui sera compréhensible pour votre audience.

Pour terminer, rappelons-nous que le produit final qu'est la visualisation de données réalisée à des fins de communication résulte d'un ensemble de choix humains et informatiques. Afin d'arriver à produire la visualisation, les données ne sont pas seulement extraites et traduites sous une forme visuelle. De fait, elles subissent de grandes transformations à plusieurs reprises. Ces transformations ont évidemment des implications importantes sur les données. La

visualisation de données est souvent présentée comme un accès transparent aux données alors qu'elle est le point final d'un processus complet. À travers cet ouvrage, vous avez connaissance des étapes de ce processus relevant de choix humains, sans pour autant vous consacrer à ce qui relève du domaine informatique et technique. Sans doute n'existe-t-il pas de visualisation de données parfaite : c'est la raison pour laquelle, en tant que lecteur ou concepteur de représentations graphiques, il est possible de développer un large sens critique en ce qui concerne la conception de ces fascinants objets de communication.

ENVIE DE TESTER VOS CONNAISSANCES ?

Rendez-vous sur
www.socialmedialab.be/dataviz



www.socialmedialab.be/dataviz

Découvrez des vidéos interactives, des exercices et plus d'exemples

BIBLIOGRAPHIE GÉNÉRALE

Andry, T. (2021). Beauté des données: quelle est l'influence de l'ornementation des visualisations de données sur la construction de sens de l'utilisateur ? (Doctoral dissertation, Université Catholique de Louvain).

Andry, T. (2020). Visualisation de données et design émotionnel peuvent-ils se conjuguer?. *Communiquer. Revue de communication sociale et publique*, (28), 53-71.

Bachimont, B. (1996). Intelligence artificielle et écriture dynamique : De la raison graphique à la raison computationnelle. *Au nom du sens*, 290-319.

Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A., McDine, D., & Brooks, C. (2010, April). Useful junk? The effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2573-2582)

Bertin (1967), *Sémiologie Graphique*, Paris, Mouton.

Borgo, R., Abdul-Rahman, A., Mohamed, F., Grant, P. W., Reppa, I., Floridi, L., & Chen, M. (2012), "An empirical study on using visual embellishments in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2759–2768

Borkin, Michelle A., Bylinskii, Zoya, Kim, Nam Wook, *et al.* Beyond Memorability: Visualization Recognition and Recall. Visualization and Computer Graphics”, *IEEE Transactions on*, 2016, vol. 22, no 1, p. 519-528

Brasseur, C. (2015). Usages visuels des données & Big data. *I2D – Information, données & documents*, me 52 (2), 44-46. Card, S. (2012), *Information Visualization in Human Computer Interaction Handbook : Fundamentals, Evolving Technologies, and Emerging Applications*, Third Edition. CRC Press, p. 516-548.

Cairo, A. *Functional Art, The: An Introduction to Information Graphics and Visualization*, New Riders, 2012. *San Francisco, US.*

Card, S. (2012), *Information Visualization in Human Computer Interaction Handbook : Fundamentals, Evolving Technologies, and Emerging Applications*, Third Edition. CRC Press, p. 516-548.

Cardon, D. (2012). Regarder les données. *Multitudes*, 49(2), 138. <https://doi.org/10.3917/mult.049.0138>

Cardon, Dominique. (2015). *À quoi rêvent les algorithmes : Nos vies à l'heure des big data*. Seuil.

Chauvin, S. (2005), *Visualisations heuristiques pour la recherche et l'exploration de données dynamiques: l'art informationnel en tant que révélateur de sens* (Doctoral dissertation, Université Paris VIII Vincennes-Saint Denis).

Chi E. H. H., & Card S. K. (1999), “Sensemaking of evolving web sites using visualization spreadsheets”, In *Information Visualization, 1999. (Info Vis' 99) Proceedings. 1999 IEEE Symposium on*, IEEE, p. 18-25.

Citton, Y. (2013). The Attention Economy and the New Forms of Digital Exploitation. *Multitudes*, 54(3), 163175.

- Cleveland, W. S., & McGill, R. (1984), Graphical Perception : Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, 79(387), 531-554. <https://doi.org/10.1080/01621459.1984.10478080>
- Collomb, C. (2016). *Un concept technologique de trace numérique*. Compiègne. <http://www.theses.fr/2016COMP2286>
- Costigan-Eaves, P., & Macdonald-Ross, M. (1990). William Playfair (1759-1823). *Statistical Science*, 318-326.
- Eco, U. (2010). *De l'arbre au labyrinthe*. Grasset.
- Few S. (2006), *Information dashboard design*. Chapter 1 : Clarifying the vision, O'Reilly, Newton.
- Few, S., & Edge, P. (2011). The chartjunk debate. *Visual Business Intelligence Newsletter*, no. June, 1-11.
- Focant, C. (1989). Hans-Ruedi Weber, Voici Jésus l'Emmanuel. La venue de Jésus dans l'art et la Bible. 1988. *Revue Théologique de Louvain*, 20(4), 481-481
- Fredriksson S. (2015), « Du design d'information à la visualisation de données : un enjeu de transmission de sens auprès de la société civile », *I2D – Information, données & documents 2015/2* (Volume 52), p. 36-36
- Friendly, M. (2008), « A brief history of data visualization. In *Handbook of data visualization* », Springer Berlin Heidelberg, pp. 15-56.
- Garlan, T. (2012). Deux siècles de cartographie des sédiments marins. *CFC-Cartes et Géomatique*, (211), 115-139. Horn, A., Ostwald, D., Reisert, M., & Blankenburg, F. (2014). The structural–functional connectome and the default mode network of the human brain. *Neuroimage*, 102, 142-151.

Haroz, S., Kosara, R., & Franconeri, S. L. (2015, April). Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 1191-1200).

Heer, J., Bostock, M., & Ogievetsky, V. (2010), A tour through the visualization zoo, *Communications of the ACM*, 53(6), 59.

<https://doi.org/10.1145/1743546.1743567>Bertin, J. (1967), *Sémiologie graphique: les diagrammes*, Paris, Mouton.

Hurter, C. (2010). *Caractérisation de visualisations et exploration interactive de grandes quantités de données multidimensionnelles* [Phdthesis, Université Paul Sabatier – Toulouse III].

<https://tel.archives-ouvertes.fr/tel-00610623/document>, p. 19-21

Huylebrouck D. (2008), *Afrique mathématiques*. Traduction française de Ferdinand Bourton. Brussel : VUB Press.

Jacob, C. (1992). *L'empire des cartes: approche théorique de la cartographie à travers l'histoire*. Albin Michel.

Kennedy, H., & Hill, R. L. (2018). The feeling of numbers: Emotions in everyday engagements with data and their visualisation. *Sociology*, 52(4), 830-848.

Kim, S. D. (2008), *La raison graphique de Saussure*, Cahiers Ferdinand de Saussure, 23-42.

Kim, M. C., Zhu, Y., & Chen, C. (2016), How are they different? A quantitative domain comparison of information visualization and data visualization (2000–2014), *Scientometrics*, 107(1), 123-165.

<https://doi.org/10.1007/s11192-015-1830-0>.

Kirk, A. (2016), *Data visualisation: a handbook for data driven design*, Sage.

- Kirsh, D. (2010). Thinking with external representations. *AI & SOCIETY*, 25(4), 441-454. <https://doi.org/10.1007/s00146-010-0272-8>
- Lamb, R., & Kling, R. (2003). Reconceptualizing Users as Social Actors in Information Systems Research. *Management Information Systems Quarterly*, 27(2). <http://aisel.aisnet.org/misq/vol27/iss2/4>
- Lanthony P. (2001), *Vision des couleurs et daltonisme*, Ediss
- Launay, M. (2016). Le grand roman des maths. De la préhistoire à nos jours. Flammarion.
- Lejeune, S., & Blin-Barrois, B. (Eds.). (2006). *Parlons couleur ! : langage, codes, création* (Vol. 4). Édisud
- Lima, M. (2014). The book of trees: visualizing branches of knowledge. Princeton Architectural Press.
- MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D., & Gahegan, M. (2012). Visual semiotics & uncertainty visualization : An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2496–2505.
- Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)*, 5(2), 110-141.
- Mialet J-P, *L'attention*, Paris, PUF, coll. « Que sais-je ? », 1999
- Nara, A. (2015). Visual analytics of movement, by Gennady Andrienko, Natalia Andrienko, Peter Bak, Daniel Keim and Stefan Wrobel, Berlin Heidelberg, Springer-Verlag, p. 229.
- Parrochia, D. (2017). Mathématiques et philosophie: des irrationnels aux motifs. ISTE Group.
- Parsons, J. H. (1924). *An introduction to the study of colour vision*. Cambridge University Press

Payne, J. W. (1976), « Task complexity and contingent processing in decision making: An information search and protocol analysis », *Organizational behavior and human performance*, 16(2), p. 366-387.

Ramly, N. N., Nor, F. M., Ahmad, N. H., & Aziz, M. H. (2012), « Comparative Analysis on Data Visualization for Operations Dashboard », *International Journal of Information and Education Technology*, 2(4).

Rensink R., O'Regan J., Clark J. (1997)., To See or not to See: The Need for Attention to Perceive Changes in Scenes, *In Psychological Science*. 8 (5): 368–373, doi:10.1111/j.1467-9280.1997.tb00427.x.

Rétention | *Ars Industrialis*. (s. d.). Consulté 4 janvier 2018, à l'adresse <http://arsindustrialis.org/r%C3%A9tention>

Roser, M., Ritchie, H., & Ortiz-Ospina, E. (2015). Internet. *Our World in Data*. <https://ourworldindata.org/internet>

Roth, R. E. (2017). Visual Variables. In D. Richardson, N. Castree, M. F. Goodchild, A. Kobayashi, W. Liu, & R. A. Marston (Éds.), *International Encyclopedia of Geography : People, the Earth, Environment and Technology* (p. 1-11). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118786352.wbieg0761>

Sallaberry, A. (2011), *Visualisation d'information : De la théorie sémiotique à des exemples pratiques basés sur la représentation de graphes et d'hypergraphes*, (Doctoral dissertation, Université Sciences et Technologies – Bordeaux I), <https://tel.archives-ouvertes.fr/tel-00646397/document>

Saulnier A., Thievre, J., & Viaud, M. L. (2006), « La perception du mouvement dans la visualisation : le cas des graphes », *Revue d'Interaction Homme-Machine* Vol, 7(2).

Sauvajol-Rialland, C. (2014). Infobésité, gros risques et vrais remèdes. *L'Expansion Management Review*, (1), 110-118.

- Sève, R. (2009). Science de la couleur: Aspects physiques et perceptifs, Challagam Éditions
- Shneiderman, B. (1996). The Eyes Have It : A Task by Data Type Taxonomy for Information Visualizations. Proceedings of the 1996 IEEE Symposium on Visual Languages, 336 –. <http://dl.acm.org/citation.cfm?id=832277.834354>
- Snow, J. (1854). The cholera near Golden-square, and at Deptford. Medical Times and Gazette, 9, 321-322.
- Stevens, S. S. (1946). On the theory of scales of measurement Science. New series, 103(2684).
- Telea, A. C. (2014). Data visualization: principles and practice. CRC Press.
- Tufte, E. R. (1983). Envisioning information. Cheshire, CT: Graphics press.
- Tufte, E. R. (2001). The visual display of quantitative information, Cheshire, CT: Graphics press.
- Valeur B. (2016), *Une belle histoire de la lumière et des couleurs*, Flammarion, Paris
- Van der Heijden, A. H. C. (1996). Perception for selection, selection for action, and action for perception. Visual Cognition, 3(4), 357-361.
- Van House, N., & Churchill, E. F. (2008). Technologies of memory : Key issues and critical perspectives. *Memory Studies*, 1(3), 295-310.
- Vision des couleurs (2020), *Statistiques*, repéré à <https://www.visiondescouleurs.com/statistiques/> en mai 2020.
- Ware, C. (2004). Information visualization: perception for design. Second Edition. Morgan Kaufmann.
- Zhu, B., & Chen, H. (2008). Information visualization for decision support. In Handbook on Decision Support Systems 2, Springer Berlin

Heidelberg, p. 699-722.

Zipf, G. K. (1935), *The psycho-biology of language*, Houghton, Mifflin.

TABLE DES FIGURES

Figure 1 : LOs d'Ishango, exposé au Museum des Sciences Naturelles à Bruxelles

Figure 2 : Arbre de Jessé dans Le livre de chasse de Phébus, France, fin du xv^e siècle

Figure 3 : Exemple médiéval d'un diagramme d'arbre

Figure 4 : Arbor Elementialis de Raymond Lulle

Figure 5 : Reconstitution de la carte du Monde de Ptolémée (xv^e siècle)

Figure 6 : Carte d'état-major, 1866

Figure 7 : Graphique en barres de William Playfair

Figure 8 : Graphique en lignes de William Playfair

Figure 9 : Graphique en secteurs de William Playfair

Figure 10 : Carte originale de John Snow, 1854

Figure 11 : Diagram of the causes of mortality in the army in the East par Florence Nightingale, 1858

Figure 12 : Charles Joseph Minard, Carte figurative des pertes successives en hommes de l'Armée française dans la campagne de Russie en 1812-1813, 1869

Figure 13 : Scivis : représentation des connectomes humains

Figure 14 : Infovis : la popularité des dirigeants en temps de crise

Figure 15 : Exemple d'infographie : Étude HEC Act One

Figure 16 : Le vase de Rubin

Figure 17 : Schéma d'une coupe longitudinale d'œil humain

Figure 18 : Structure de la rétine

Figure 19 : Domaines de vision

Figure 20 : Le Bode Museum et la Tour de Télévision à Berlin, vue de jour et de nuit

Figure 21 : Diffusion de la lumière

Figure 22 : La longueur d'onde

Figure 23 : Spectre visible d'absorption des cônes

Figure 24 : Synthèse additive et soustractive

Figure 25 : Contraste de luminosité

Figure 26 : Contraste de saturation

Figure 27 : Contraste simultané

Figure 28 : Contraste successif

Figure 29 : Différence de perception des couleurs d'un feu de signalisation entre une personne à la vision normale et une personne atteinte de daltonisme

Figure 30 : Dispositions préattentives de la couleur

Figure 31 : Saillance visuelle des chiffres

Figure 32 : Caractéristiques pré-attentives

Figure 33 : Globe and high court

Figure 34 : Loi de la bonne forme

Figure 35 : Loi de proximité

Figure 36 : Loi de similitude

Figure 37 : Principe de fermeture

Figure 38 : Principe de complétion amodale

Figure 39 : Principe de transparence/ombrage

Figure 40 : Exemple de perspective

Figure 41 : Loi de destin commun

Figure 42 : Représentation de la connectivité

Figure 43 : Le vase de Rubin

Figure 44 : Illusion de Hering

Figure 45 : Illusion de Müller-Lyer

Figure 46 : Illusion de Ponzo

Figure 47 : Le cube de Necker

Figure 48 : L'escalier impossible de Penrose et Penrose, 1958

Figure 49 : Le triangle de Penrose et Penrose, 1958

Figure 50 : Bandes de Mach

Figure 51 : Contraste de saturation

Figure 52 : Anneaux concentriques de Pinna et Brelstaff

Figure 53 : Illusion d'Ebbinghaus, 1901

Figure 54 : Illusion de Poggendorf

Figure 55 : Métaphore du « Zoo de la visualisation »

Figure 56 : Taux de chômage à New Castle et en Nouvelle-Galles-du-Sud au cours du temps

Figure 57 : Évolution de la surface des vignobles en fonction de la variété des raisons au cours du temps, 2019

Figure 58 : Exemple de Small Multiples

Figure 59 : Exemple de stem-and-leaf plot

Figure 60 : Exemple de Q-Q Plot comparant les distributions de deux relevés de temps de lecture

Figure 61 : The scatterplot of Iris flower data set

Figure 62 : Coordonnées parallèles décrivant certaines caractéristiques d'un ensemble de voitures

Figure 63 : Comtés d'Irlande par densité de population (personnes par hectares) en 2011

Figure 64 : La production de plastique dans le monde

Figure 65 : An alternative way of visualising the world's population, using a variant on the cartogram

Figure 66 : InMaps pour LinkedIn : visualiser son réseau

Figure 67 : Exemple de diagramme en arc

Figure 68 : Diagramme à barres

Figure 69 : Black cherry tree histogram

Figure 70 : Diagramme à barres empilées

Figure 71 : Diagramme circulaire

Figure 72 : Diagramme en anneau

Figure 73 : Graphique en cercles imbriqués

Figure 74 : Diagramme de Nightingale Rose représentant la vitesse du vent par mois en nœuds

Figure 75 : Nuages de points

Figure 76 : The scatterplot of Iris flower data set

Figure 77 : Exemple de graphique à bulles

Figure 78 : Bubble chart showing the quantity of professions people, creating programming languages

Figure 79 : Exemple de dot matrix chart

Figure 80 : Exemple de diagramme à pictogrammes

Figure 81 : Exemple de diagramme quantile-quantile

Figure 82 : Annotations d'un exemple de boîte à moustaches

Figure 83 : Pyramide des âges

Figure 84 : Spider Chart, 2008

Figure 85 : Tableau de données représentant les rangs d'appréciations récoltées par questionnaires

Figure 86 : Population d'animaux sauvages, données fictives

Figure 87 : New Zealand vineyard area by grape variety

Figure 88 : Departmental salary expenses

Figure 89 : Condogram Spiral Plot, 2011

Figure 90 : Gantt Chart, 2006

Figure 91 : Horizon Chart

Figure 92 : Bubble Map, 2015

Figure 93 : Countries by gross median household income in Int\$ (PPP), 2018

Figure 94 : World Map of Organic Agriculture, 2016

Figure 95 : US incarceration rate graduated symbol map, 2015

Figure 96 : Minard, 1865, Carte figurative et approximative des quantités de vin français exportés par mer en 1864

Figure 97 : Dot Density Map of the Population of Illinois, 2013

Figure 98 : Obesity rate, 2020

Figure 99 : International route map of Philippine Airlines as of October 2013

Figure 100 : Heatmap, 2010

Figure 101 : Exemple d'arbre

Figure 102 : Hyperbolic Tree, 2010

Figure 103 : Exemple de sunburst

Figure 104 : Gradient grouped treemap, 2010

Figure 105 : Radial node-link diagram of the Flare package hierarchy.

Figure 106 : Social Network Diagram, 2014

Figure 107 : Image Dive guitar chord diagram, 2018

Figure 108 : Social network visualization

Figure 109 : Cluster heatmap, 2006

Figure 110 : Representative sample for parallel coordinates, 2015

Figure 111 : Exemple de diagramme de Venn

Figure 112 : La Terre en surchauffe

Figure 113 : L'espoir d'une reprise économique en V

Figure 114 : Tableau des propriétés des variables rétinienne

Figure 115 : Traduction de « Visual variables and their syntactics »

Figure 116 : Diagramme en bâtonnets représentant « Le nombre de sans-abris recensés dans la région par année » avec distorsion (chiffres fictifs)

Figure 117 : Diagramme en bâtonnets représentant « Le nombre de sans-abris recensés dans la région par année » avec distorsion (chiffres fictifs)

Figure 118 : Mesure du lie factor sur la figure 116

Figure 119 : Mesure du lie factor sur la figure 117

Figure 120 : Vente des maillots de bain au cours des saisons de 2011 à 2015 (chiffres fictifs)

Figure 121 : Les bistrots en voie de disparition ?

Figure 122 : Nombre de magasins vendus par année, chiffres fictifs. Comparaison de différents *data-ink ratios*

Figure 123 : Nombre d'appréciations, chiffres fictifs

Figure 124 : Illusion visuelle, Pixabay, 2020

Figure 125 : Grains de café, Openclipart

Figure 126 : Disposition de la grille sur « Températures moyennes en 2018 » (chiffres fictifs)

Figure 127 : The Big Duck

Figure 128 : « Monstrous costs »

Figure 129 : Comparaison d'un diagramme en bâtonnets en 2D vs. 3D : Taille moyenne de femmes et d'hommes de 23 ans, chiffres fictifs

Figure 130 : Comparaison de diagrammes circulaires : 2D vs. 3D

Figure 131 : « Monstrous costs », version minimaliste

Figure 132 : The Sharks of Business

Figure 133 : Cat Population

Figure 134 : Fibre optique : quels pays sont les mieux équipés ?

Figure 135 : L'Europe domine la production mondiale de vaccins

Figure 136 : La représentation des femmes dans les municipalités

Figure 137 : Loi de proximité

Figure 138 : Application de la loi de similitude par utilisation de la couleur et du grain

Figure 139 : Application de la loi de continuité : trajectoires lisses vs. brusques changements de direction

Figure 140 : Gestalt 7 : Closure

Figure 141 : Diagramme de Venn

Figure 142 : La connectivité

Figure 143 : Croix symétrique et rectangle

Figure 144 : Le vase de Rubin

Figure 145 : Précision du jugement des données quantitatives

Figure 146 : Estimation du rapport annuel moyen entre l'évapotranspiration réelle (ET) et les précipitations (P) pour les États-Unis contigus pour la période 1971-2000

Figure 147 : Lecture de texte entre deux couleurs différentes d'un point de vue chromatique, avec variation de luminosité

Figure 148 : Contraste de saturation

Figure 149 : Les nuances de gris et les contrastes

Figure 150 : Color field, inspirée de Ware (2004)

Figure 151 : Évolution du spectre en fonction de la luminosité (valeur), de la teinte et de la saturation

Figure 152 : HSV

Figure 153 : Palettes de couleurs séquentielles choisies grâce à colorbrewer.org

Figure 154 : Palettes de couleurs divergentes choisies grâce à colorbrewer.org

Figure 155 : Palettes de couleurs qualitatives choisies grâce à colorbrewer.org

Figure 156 : Pourcentage des voix aux élections présidentielles de 2004 aux USA

Figure 157 : Nombre de décès du COVID-19 en Europe, recensés pour un million d'habitants en date du 25 juin 2020

Figure 158 : Nombre de décès du COVID-19 en Europe, recensés pour un million d'habitants en date du 25 juin 2020, résultat

belge mis en évidence

Licences utilisées pour la présentation des images

Domaine Public

CC 0 : <https://creativecommons.org/publicdomain/zero/1.0/deed.fr>

CC BY : <https://creativecommons.org/licenses/by-sa/2.0/be/>

CC BY-SA : <https://creativecommons.org/licenses/by-sa/2.0/be/>

Free Art License 1.3 : <https://artlibre.org/licence/lal/en/>

INDEX

C

Cartographie [1](#)

chartjunk [1](#)

contraste [1](#)

D

daltoniens [1](#)

dashboard [1](#)

diagramme [1](#)

données [1](#)

E

esthétique [1](#)

F

Facteurs de réaction émotionnelle [1](#)

Facteurs humains [1](#)

G

Gestalt Theory [1](#), [2](#)

graphe [1](#)

graphique [1](#)

I

Infographie [1](#)

infovis [1](#), [2](#)

J

John Snow [1](#)

L

la vue [1](#)

M

Minard [1](#)

N

NOIR [1](#)

P

Playfair [1](#)

S

scivis [1](#)

statistique [1](#)

V

variable [1](#)

Visual analytics [1](#)

visualisation de données [1](#)

visualisation d'information [1](#)

Licences utilisées pour la présentation des images

Domaine Public

CC 0 : <https://creativecommons.org/publicdomain/zero/1.0/deed.fr>

CC BY : <https://creativecommons.org/licenses/by-sa/2.0/be/>

CC BY-SA : <https://creativecommons.org/licenses/by-sa/2.0/be/>

Free Art License 1.3 : <https://artlibre.org/licence/lal/en/>

Ressources numériques

- Exercices interactifs
- Illustrations en couleur

Repérez les ressources numériques dans votre livre



lienmini.fr/ressourcesnum-dbs

Accédez directement à votre ressource :

Flashez le code avec votre
téléphone ou votre tablette



OU

Tapez l'URL
dans votre navigateur



Pour toute information sur notre fonds et les nouveautés dans votre domaine de spécialisation, consultez notre site web : www.deboecksuperieur.com

Couverture et maquette intérieure : cerise.be

© De Boeck Supérieur s.a., 2022
Rue du Bosquet, 7 – B-1348 Louvain-la-Neuve

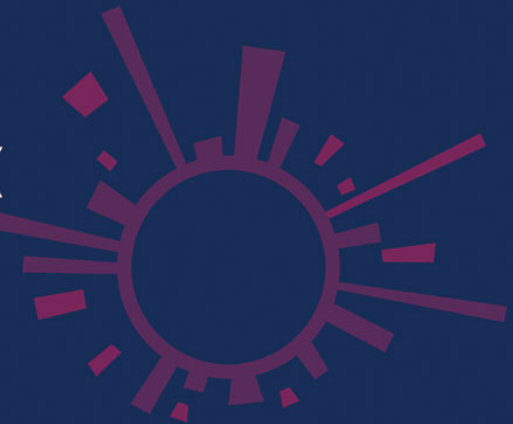
Tous droits réservés pour tous pays.

« Cette œuvre est protégée par le droit d'auteur et strictement réservée à l'usage privé du client. Toute reproduction ou diffusion au profit de tiers, à titre gratuit ou onéreux, de tout ou partie de cette œuvre, est strictement interdite et constitue une contrefaçon prévue par les articles L 335-2 et

suivants du Code de la Propriété Intellectuelle. L'éditeur se réserve le droit de poursuivre toute atteinte à ses droits de propriété intellectuelle devant les juridictions civiles ou pénales. »

Ce document numérique a été réalisé par [PCA](#)

Les fondamentaux de la visualisation de données



Graphiques dans les médias (presse, télévision, réseaux sociaux, plateformes web, etc.), outils d'aide à la décision en entreprise ou encore balises de suivi de nos activités physiques sur smartphone : la visualisation de données est aujourd'hui omniprésente.

Avec l'explosion du numérique, présenter de grands ensembles de données de manière intelligible est devenu un enjeu sociétal. Cet ouvrage montre le rôle joué par la visualisation en tant que médium privilégié dans nos interactions avec les données.

Il propose, d'une part, une perspective historique de la discipline et met en évidence l'importance de la perception visuelle dans l'interaction avec les données. Il offre, d'autre part, un guide à la conception de visualisations de données composé d'un catalogue répertoriant les grands classiques de la visualisation et d'une série de principes à respecter pour concevoir des visualisations claires, efficaces et fidèles aux données.

Cet ouvrage s'adresse aux étudiant·es et enseignant·es en sciences humaines.



Tiffany Andry est docteure en sciences de l'information et de la communication. Ses travaux de recherche s'intéressent à l'appropriation des visualisations de données.

Suzanne Kieffer

est docteure en informatique, professeure à l'école de communication de l'UCLouvain et membre du groupe de recherche en médiation des savoirs (GReMS).



François Lambotte est docteur en sciences de gestion et professeur en communication organisationnelle et numérique à l'UCLouvain. Il coordonne un programme de recherche dédié aux enjeux d'intelligibilité des données : production, extraction, traitement, visualisation et appropriation (www.socialmedialab.be).

Dans le cadre du nouveau Système Européen de Transfert de Crédits (E.C.T.S.), ce manuel couvre en France les niveaux : Licence 1-2-3, Master 1-2. En Belgique : Bachelier 1-2-3, Master 1-2

L 1-2-3

M 1-2

DEBOECK
SUPÉRIEUR

www.deboecksuperieur.com

En Suisse : Bachelor, Master
Au Canada : Baccalauréat, Maîtrise



NOTES

1. <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>
2. Source: <https://cdn.statcdn.com/Infographic/images/normal/13175.jpeg>
3. Source: <https://fr.statista.com/infographie/17351/applications-generant-la-plus-grande-part-du-traffic-internet-mobile/>
4. <https://www.youtube.com/watch?v=6XQhUJ5fqMk>
5. Le projet Visionary, qui met en place des alternatives pour daltoniens en ce qui concerne la conception d'interfaces : <https://colour-blindness.org/>
6. Statistiques : Ebook gratuit en ligne accompagné de vidéos explicatives : How2stats, par Gilles Gignac, <http://www.how2statsbook.com/p/chapters.html>
7. Vidéo représentant l'exportation internationale d'armes par les États-Unis de 1950 à 2017. *The United States of Arms from Will Geary on Vimeo*, <https://vimeo.com/279923192>
8. Contributeurs du catalogue : Benhammou, Rizlène ; Buttice, Sarah ; Dart, Marine ; De Cooman, Florence ; Debray, Denis ; Eeckhout, Thomas ; Engels, Aurélie ; Guillet de la Brosse, Tania ; Henreaux, Emilie ; Jaspers, Julien ; Kimus, Estelle ; Lebeau, Alizée ; Mataisse, Charlotte ; Michel, Solène ; Ndayiragije, Rénovat ; Noutcha, Michel ; Phan-Ngoc, Tina ; Pierre, Laura ; Selassi, Nabil ; Vaessen, Théo ; Vervaet, Julie ; Voituren, Lydia
9. Consultez cette visualisation en ligne : *How Does Income Relate to Life Expectancy?*, par Gap Minder Répéré à [https://www.gapminder.org/tools/#\\$chart-type=bubbles](https://www.gapminder.org/tools/#$chart-type=bubbles)
10. Moyenne, médiane ou zéro. Le nombre de couleurs est égal de chaque côté du point de rupture.
11. <https://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3>