



HAL
open science

La visualisation de données, entre usages démonstratifs et heuristiques

Martin Grandjean

► **To cite this version:**

Martin Grandjean. La visualisation de données, entre usages démonstratifs et heuristiques. Bohet Baptiste; Pringuet Virginie. Les devenirs numériques des patrimoines, Editions de la Maison des Sciences de l'Homme, pp.199-217, 2022, 978-2-7351-2891-4. halshs-03845625

HAL Id: halshs-03845625

<https://shs.hal.science/halshs-03845625>

Submitted on 9 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La visualisation de données, entre usages démonstratifs et heuristiques

Martin Grandjean

Grandjean, Martin "La visualisation de données, entre usages démonstratifs et heuristiques", in Bohet, Baptiste et Pringuet, Virginie, *Les devenirs numériques des patrimoines*, Editions de la Maison des sciences de l'homme, Paris, 2022, pp. 199-217.

Enjeux de la visualisation de données

Pratique transdisciplinaire prenant sa source dans les méthodes mathématiques bien avant l'avènement de l'ordinateur, la visualisation de données est le versant visuel et coloré de l'analyse de données statistiques. Même si la visualisation de données fait partie d'un vaste continuum (Jessop 2008), elle s'est assez récemment constituée en discipline, celle des *visual studies*, avec ses précurseurs, dont Jacques Bertin (Bertin 1967), John Tukey (Tukey 1977) ou Edward Tufte (Tufte 1983), et un nouvel âge d'or il y a une dizaine d'années, incarné par exemple par Alberto Cairo (Cairo 2012), Nathan Yau (Yau 2011) ou Mike Bostock (Bostock *et al.* 2011). Il se publie désormais chaque année des dizaines de nouvelles monographies ou recueils d'exemples inspirants (voir par ex. Schwabish 2021 pour un des derniers exemples en date). Devenu un véritable objet de l'histoire de l'art, on observe aujourd'hui un retour vers des formes *vintage*, une redécouverte des graphiques fascinants et parfois excentriques de Charles-Joseph Minard, Florence Nightingale, John Snow ou encore W. E. B. Du Bois. Si elles sont souvent étudiées pour elles-mêmes, une analyse longitudinale de ces visualisations « historiques » montre que leur développement a connu des périodes plus ou moins fastes (Friendly 2007) et nous permet de relativiser le caractère innovant des productions actuelles.

Les formes visuelles que nous connaissons aujourd'hui sont donc largement inspirées des premières grandes études statistiques du XIX^e siècle ainsi que de l'utilisation quasi quotidienne d'outils produisant des résultats chiffrés ou des représentations visuelles à interpréter (sinusoïdale explicitant une fréquence, graphe d'une fonction en analyse mathématique, bande colorée du spectre électromagnétique d'un corps céleste, simulation de molécules ou de génomes, calcul de la trajectoire de particules élémentaires après collision, etc.), mais elles n'en sont pas moins le reflet graphique d'une abstraction supposée pure, vraie, objective : la donnée. Or, la donnée est rarement « donnée », mais construite par un appareillage destiné à la recueillir ou une procédure chargée de la compiler. Et si, comme on va le voir, l'inventaire des types de visualisations possibles montre une convergence des formes quel que soit le contexte scientifique, la façon dont on crée la donnée, elle, diffère

fortement d'une discipline à l'autre. La visualisation de données en sciences humaines et sociales, par exemple, aura beau utiliser des codes graphiques qui la rendront quasi identique à une visualisation statistique produite par des chercheurs en sciences de la vie, elle ne consiste pas en la représentation du résultat d'une manipulation expérimentale, d'une observation au moyen d'un appareillage ou d'un processus mathématique conceptuel. Au contraire, elle traduit visuellement un jeu de données préalablement modélisées, extraites et encodées par le chercheur. De l'enquête sociologique dont on tire des distributions statistiques à la compilation d'événements historiques récurrents sur une carte, en passant par les analyses économiques de l'évolution de la valeur de certaines marchandises ou encore le comptage d'occurrences de termes dans de grandes concordances littéraires ou religieuses, les méthodes quantitatives en sciences humaines et sociales nous ont habitués à la production de visualisations de données qui nécessitent un travail considérable pour passer de la source à la donnée, puis de la donnée à la représentation graphique, et impliquent un régime interprétatif particulier (Drucker 2020).

Plutôt qu'une histoire de la visualisation, ce que nous proposons ici est un questionnement sur le statut de ces images, leur rhétorique et leur inscription dans une démarche d'herméneutique de la donnée. Un panorama des usages, une typologie. Or, à quoi sert une typologie des usages, puisqu'il existe déjà comme on va le voir de nombreuses typologies des formes ? Au contraire des catalogues d'exemples, qui étendent l'horizon et développent la créativité, une typologie des usages sert à prendre conscience de ce que l'on fait (et clarifier pourquoi), de comprendre ce que l'on peut faire dans un contexte donné et de choisir ce que l'on va faire en connaissance de cause. En bref, c'est un outil qui permet une critique de l'acte de visualisation lui-même, entendu comme une pratique scientifique faisant intégralement partie de l'activité de recherche, puis de vulgarisation.

Effets de la démocratisation de la visualisation de données

Comme pour confirmer la transversalité de la visualisation, on observe depuis quelques décennies le développement d'outils et de méthodologies qui témoignent de la diffusion de cette pratique dans toutes les disciplines, même – et surtout – hors des sciences expérimentales : les chercheurs en sciences sociales, en histoire, en études littéraires ou artistiques, rejoints par les milieux du journalisme et du design graphique, se projettent désormais aussi dans des analyses et des représentations visuelles qui dépassent largement la simple présentation de résultats qu'ils pratiquaient jusqu'ici pour embrasser des démarches beaucoup plus exploratoires. Mais quel que soit le champ de recherche, si l'automatisation des processus et la massification des corpus sont des conséquences bienvenues de ce renouveau de la visualisation de données pour le rafraîchissement des approches et le développement de nouvelles manières de faire et de communiquer la recherche, cette évolution d'une visualisation illustrative à une visualisation faisant partie intégrante du déroulement de la recherche amène également son lot de questions.

Tout d'abord, la place de plus en plus grande qu'occupe la computation dans la recherche ravive la tension entre approches qualitatives et quantitatives. Présentant le vaste éventail des possibilités offertes par les technologies numériques (dont la visualisation de données est souvent et logiquement l'un des éléments les plus visibles), la tentation de considérer celles-ci comme une « révolution » laisse souvent entendre en creux que les méthodes préexistantes sont ou seront à terme remplacées. Les facultés esthétiques et heuristiques de la visualisation renforcent d'ailleurs souvent volontairement ou involontairement cette impression : un projet

présenté au moyen d'un site web, d'un poster ou d'un article de presse tire généralement largement parti d'une iconographie attrayante. La visualisation, avec son pouvoir objectivant, totalisant, qui laisse croire qu'on peut saisir d'un seul coup tout un sujet, va donc y être privilégiée et donner ainsi l'impression que le projet en question se résume à ces aspects techniques. De fait, dans des disciplines dont la culture numérique n'est pas encore arrivée à un certain degré de maturité, un projet de recherche qui se concentrerait sur le développement d'un outil d'analyse et de visualisation *ad hoc*, activité jusqu'ici réservée à des acteurs spécifiques (un département d'informatique, une entreprise spécialisée dans le numérique, une agence de graphisme), risque de susciter la méfiance de collègues craignant un abandon du terrain aux *computer sciences*. Ces appréhensions ne sont pas toutes infondées, elles rappellent la nécessité de conserver à la visualisation de données sa fonction de serviteur de la recherche et de maintenir toujours explicitement le lien entre le matériau de recherche de base, les données qui en sont extraites et finalement leur représentation graphique pour que cette dernière ne se substitue pas à l'objet fondamental. Ce rappel est d'autant plus important que l'aspect innovant et extrêmement spécifique de certaines technologies oblige ceux qui les pratiquent à s'isoler dans des communautés qui les coupent parfois de leur discipline et rendent ainsi leurs résultats et leurs publications peu intelligibles pour leurs pairs. À ne fréquenter que des collègues qui travaillent au quotidien sur des questions de visualisation de données ou sur des méthodes computationnelles, on en oublie que ces nouvelles compétences s'éloignent à ce point des méthodes habituelles qu'on perd le langage commun, la *lingua franca* de la communauté disciplinaire d'origine.

L'apparente démocratisation de l'accès aux outils est une autre conséquence à double tranchant des récents développements de la visualisation de données. Le foisonnement de solutions clés en main peut en effet donner l'impression que l'analyse de données et leur représentation graphique constituent une démarche accessible à tout le monde. S'il est évidemment précieux d'éviter que les ressources logicielles ne soient accaparées par une caste de spécialistes, il faut se rendre à l'évidence et admettre que des compétences personnelles et un certain temps d'apprentissage sont nécessaires pour comprendre le fonctionnement d'un outil et en tirer une valeur ajoutée. Et si l'absence de compétences liées à des aspects techniques – comme la maîtrise de fonctionnalités avancées, la gestion d'une interface en ligne de commande ou le design graphique fait parfois grincer des dents les spécialistes de ces outils qui voient de nouveaux arrivants piétiner les logiciels dont ils sont les « gardiens du temple », le savoir-faire qui manque le plus cruellement est celui qui touche d'une part à la compréhension globale de ce que la visualisation peut apporter à un objet d'étude et d'autre part aux fondamentaux de la statistique. Par exemple le nombre d'articles scientifiques prenant des courbes de *n-gram* de l'outil de Google Books¹ comme sources sans avoir paramétré la langue, la sensibilité à la casse ou qui ne discutent pas l'opacité du corpus est considérable. L'outil est un démonstrateur très limité (Grandjean, 2014a) mais tout en effrayant une partie de la communauté des études littéraires, il fascine manifestement celle des sciences du *big data* pour le meilleur et pour le pire (Michel *et al.* 2011). Il n'est sans doute égalé que par le nombre des couvertures de livres sur lesquelles trône fièrement un *word cloud* des mots les plus utilisés, généralement sans reconnaissance d'expressions, de sélection de *stop words* à exclure et de réflexion autour de ce que signifie la fréquence d'apparition d'un terme dans un texte (sans parler des biais du corpus et de la difficulté de lecture d'une telle visualisation). Le monde du journalisme n'est d'ailleurs pas épargné puisque c'est dans les colonnes de nos médias – qui participent pourtant activement à la popularisation de la visualisation de données auprès du grand public – que l'on trouve le plus

¹ Google Books Ngram Viewer : <https://books.google.com/ngrams>

souvent des axes tronqués sans avertissement, des bulles dont la taille est calculée sur le diamètre plutôt que sur la surface et des cartes choroplèthes qui ne sont pas pondérées en fonction de la population des régions étudiées. Ces exemples très simples expliquent partiellement que le processus de normalisation de l'usage de la visualisation soit ralenti par une partie de la communauté universitaire qui, confrontée à des usages contre-productifs et conscients de l'effet d'autorité – voire de la performativité – des visualisations de données, refuse de considérer certaines méthodes comme valables dans ce contexte. Or, comme le fait remarquer Martyn Jessop (2008, 288), la crise de validité de la visualisation de données n'est pas sans précédent dans l'histoire des sciences : il faut du temps, des pionniers puis du consensus et de la transparence pour qu'une méthode ou un outil soient définitivement adoptés et reconnus valides par une communauté scientifique.

La représentation graphique de données statistiques, une pratique qui évolue

Ces constats sur les limites de la démocratisation de la visualisation de données rappellent que la représentation graphique n'est qu'une étape du traitement d'un jeu de données et, plus largement, de l'étude d'une question scientifique. Si c'est naturellement ce produit visuel qui permet de constater l'évolution des méthodes et des outils, une représentation graphique est rarement le point final d'une démonstration puisqu'elle procède d'une simplification – ou d'une « réduction » (Manovich, 2011, 38) – et ne peut donc se suffire à elle-même pour exprimer la richesse de l'objet qu'elle décrit. Or cette simplification, qui prouve son utilité dans de nombreuses situations (pensons au plan de métro dont les codes visuels facilitent la lecture en tordant la réalité physique), n'a rien d'une pratique récente et n'a bien sûr pas attendu le développement de l'informatique. À tel point que la représentation du territoire sous la forme d'une carte augmentée de marqueurs indiquant des éléments de géographie physique ou humaine est une forme de visualisation qui préexiste à l'écriture puisqu'on trouve par exemple des chaînes de montagnes sur des peintures rupestres. Suivra la cartographie symbolique et politique, l'objectivation de la « frontière », par exemple, qui fera du plan cadastral une visualisation performative. La cartographie céleste est d'ailleurs un exemple parfait de représentation d'informations simplifiées à but utilitaire : même si dès l'Antiquité certains atlas astronomiques cherchent une forme d'exhaustivité, la plupart sont des cartes simplifiées des astres les plus lumineux et des constellations majeures à destination des marins et des voyageurs. Dans le registre des représentations visuelles hautement codifiées, on trouve également les arbres généalogiques médiévaux : sur une ou plusieurs faces du cénotaphe d'un gisant, par exemple, la représentation en bas-relief des blasons organisés selon les mariages et les parentés, explicitant parfois même le tronc et les branches de l'arbre de Jessé dont nos arbres actuels reprennent la symbolique, permet au pouvoir en place de renforcer sa légitimité auprès des visiteurs qui embrassent d'un coup d'œil la généalogie du monarque défunt sans pour autant avoir besoin de savoir lire.

Si l'on s'en tient toutefois à une définition plus étroite – et plus répandue – de la visualisation de données, c'est-à-dire de la représentation graphique de données statistiques, on cite généralement comme première occurrence de visualisation l'« Atlas politique et commercial » de William Playfair (1786). Inspiré par les *timelines* de Joseph Priestley (1765), qui visualisent vingt ans plus tôt pour la première fois la durée de vie de deux milliers de personnalités sur un axe du temps couvrant presque trois millénaires, Playfair propose de nombreuses séries chronologiques documentant la balance du commerce extérieur britannique et est dès lors considéré comme l'inventeur de l'histogramme, le graphique en

barres. Presque un siècle plus tard, et sortant du cadre purement statistique pour adopter des codes graphiques qui font désormais référence, la « Carte figurative des pertes successives en hommes de l'armée française dans la campagne de Russie 1812-1813 » réalisée par Charles Joseph Minard en 1869, décrite par Arthur Robinson (1967) mais surtout popularisée par Edward Tufte (1983) est un exemple de visualisation extrêmement parlante parce que simplifiée à l'extrême (Grandjean 2014b) qui inspire aujourd'hui encore beaucoup de graphistes d'information lors de la réalisation de cartographies de flux.

De l'essor des atlas statistiques et leurs fascinantes cartes thématiques dès la seconde moitié du XIX^e siècle aux représentations physiques de statistiques à destination du grand public au tournant du XX^e siècle (Debluë 2018), puis finalement à la « sémiologie graphique » de Jacques Bertin (1967), les codes évoluent et se stabilisent mais les pratiques demeurent relativement descriptives et les visualisations de données proprement heuristiques y sont encore rares.

Visualiser : quelles données et pour quels usages ?

Une typologie des usages de la visualisation de données

Dans le sillage de l'« analyse de données exploratoire » thématisée par John Tukey (1977), se développe une distinction entre l'approche illustrative et une approche plus expérimentale rendue possible par le traitement de données statistiques à plus grande échelle. Cette distinction entre la visualisation que nous qualifierons « de démonstration » et la visualisation « de recherche » est au cœur de la question des usages, de l'intégration de l'analyse et de la visualisation de données dans le processus de recherche². C'est pourquoi nous construisons ici une typologie générale de ces représentations graphiques³. Si la réflexion sur les types de représentations graphiques n'est de loin pas absente de la littérature scientifique, on constate que la majorité des travaux qui utilisent des visualisations de données en discutent leur production ainsi que leurs caractéristiques techniques ou visuelles mais n'en détaillent pas toujours les objectifs. Il nous semble donc important de clarifier les usages, le rôle de ces objets dans le processus de création et de diffusion de connaissance.

Cette typologie des usages rejoint donc évidemment d'autres classifications qui distinguent généralement les visualisations de données selon le type de données ou le type de forme visuelle. Parmi les typologies orientées autour du type de données visualisées, on trouve le très utile « Visual Vocabulary » du *Financial Time*⁴ qui indique 75 façons de traiter graphiquement des données selon qu'elles sont temporelles, spatiales, de classement, de corrélation, de déviation, de distribution, de magnitude, ou encore qu'elles indiquent des flux ou des pourcentages. Dans le champ des humanités numériques, Martyn Jessop (2008) distingue les visualisations selon qu'elles proviennent d'analyses spatiales, quantitatives, textuelles, temporelles ou tridimensionnelles.

² Nous préférons la notion de visualisation de « recherche » puisque le terme « exploratoire », proposé par Tukey (1977) désigne la démarche plus que la visualisation en elle-même et peut créer la confusion avec les visualisations interactives, cliquables ou manipulables permettant de naviguer dans une base de données en ligne, que ce soit un outil de recherche ou de simple présentation des données. Dans les paragraphes qui suivent, ces visualisations interactives, ou « interfaces » sont présentées comme une catégorie intermédiaire entre « démonstration » et « recherche » pure.

³ Cette typologie démonstration/recherche est esquissée dans Grandjean (2015) et Grandjean (2018).

⁴ <http://ft.com/vocabulary>

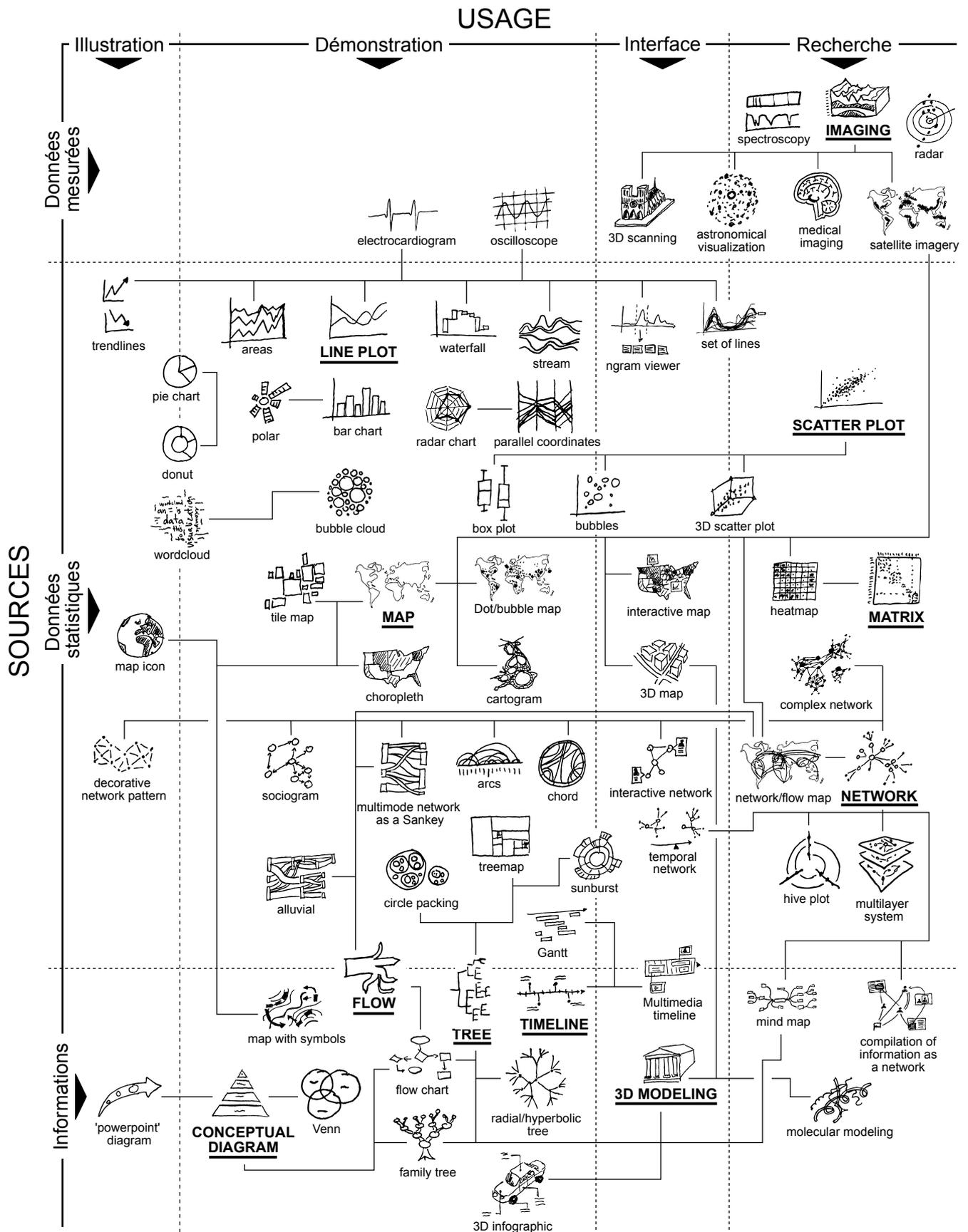


Figure 1. Typologie de la visualisation de données selon deux axes : le type de sources utilisées (vertical) et le type d'usage (horizontal) (Grandjean 2022).

Les typologies de formes, elles, se présentent souvent comme des « catalogues » destinés à inspirer les analystes et graphistes d'information. On trouve par exemple le *Data Visualization Catalog*⁵ (60 types de visualisations) ou le *Data Viz Project*⁶ (154 types), qui ont tous les deux par la suite également développé des classifications liées au type de données, ou des listes liées à des dépôts d'exemples et de tutoriels, comme *Flowingdata*⁷ (59 types). De leur côté, la plupart des outils et bibliothèques de visualisation disposent de leur propre typologie de formes, comme celle de *D3.js*⁸, par exemple (168 éléments, qui sont parfois des variations ou des animations sur la base des types répertoriés). Enfin, notons que l'on peut également classer les visualisations sur la base du traitement des données elles-mêmes, ce que l'on trouve chez Johanna Drucker (2010, 33) sous le nom de « méthodes d'interprétation algorithmique » : présentation descriptive, visualisation analytique, modélisation visuelle, visualisation procédurale et modèles évolutifs.

La typologie des usages que nous proposons ici se déploie selon deux axes principaux, détaillés plus bas et représenté sur la figure 1 comme un tableau à double entrée. Dans ce tableau, 70 exemples des visualisations le plus souvent utilisées, organisées par grandes familles, servent de points de repères. Sur l'axe vertical, les visualisations sont classées en fonction des sources, on y distingue (a) les représentations issues de la compilation d'informations, de (b) celles issues de la collecte de données ou (c) de la mesure expérimentale. L'axe horizontal propose un classement en fonction des usages, qui différencie les représentations destinées à (a) illustrer un propos, (b) être utilisées pour démontrer ou présenter une situation par les données, (c) servir d'interface, de point d'accès pour interagir avec les données et finalement (d) celles qui ont vocation à tenter d'apporter une nouvelle connaissance sur l'objet étudié dans le cadre d'un processus de recherche.

Ce tableau est destiné à alimenter la réflexion, à accompagner la réalisation d'une visualisation en remettant au centre la question du « pourquoi » : sommes-nous en train de choisir une forme pour ses attraits graphiques ou parce qu'elle se prête mieux à un usage heuristique ? Notre but est-il de comprendre un grand set de données complexes ou bien cherchons-nous à produire une visualisation intelligible pour expliquer simplement le sujet à nos lecteurs ? En tant qu'aide à la décision, ce tableau ne peut donc pas être exhaustif et couvrir tous les cas de figure (il ne cherche en tout cas pas à répertorier toutes les formes), puisqu'il est possible que certains types de visualisations appartiennent à plusieurs catégories ou permettent des usages différents selon les disciplines.

⁵ <https://datavizcatalogue.com/>

⁶ <https://datavizproject.com/>

⁷ <https://flowingdata.com/chart-types/>

⁸ <https://observablehq.com/@d3/gallery>

Visualiser de l'information ou des données

Typologie des sources		
Informations	Données statistiques	Données mesurées

La nuance entre ce que recouvrent les notions d'« infographie » et de « visualisation de données » est d'autant plus difficile à percevoir que l'une est souvent utilisée pour l'autre. Mais la différence est pourtant importante pour comprendre en quoi la démarche diverge même si son résultat est parfois d'apparence très proche : produire une représentation en se basant sur une compilation d'informations est un acte qui implique une « mise en page » graphique et souvent manuelle, alors que la visualisation d'un jeu de données est une opération qui peut être effectuée automatiquement par un logiciel (même si, dans la plupart des cas, une intervention est nécessaire pour en améliorer le rendu graphique). C'est donc précisément selon le type de source utilisée que l'on distingue l'« infographie » de la « visualisation de données » puisque leur produit est parfois visuellement très proche (une modélisation de bâtiment en 3D, par exemple, peut ressembler à s'y méprendre au scan tridimensionnel du même bâtiment, mais ce sont des objets diamétralement opposés dans leur constitution). Proches aussi sont les représentations graphiques produites par des logiciels de mesure, des visualisations de données que nous séparons ostensiblement dans le schéma de la fig. 1 (en haut) pour rendre plus claire la nature « captée » des données dont il est question ici. Sans pour autant oublier que la plupart des autres visualisations de données statistiques portent également sur des « data » qui sont avant tout des « capta » (Drucker 2011), la spécificité de ces outils de mesure est qu'ils forment un système intégré qui produit directement une visualisation sans passer par la création d'un jeu de données sous forme de tableur.

Si l'usage d'un logiciel n'est pas un critère suffisant pour discriminer à coup sûr une visualisation de données d'une infographie – par exemple l'écolier qui reporte l'emplacement de points sur du papier millimétré est bien en train de produire une visualisation de données, même s'il n'utilise qu'un crayon, et l'illustrateur qui dessine un schéma en coupe d'une automobile fait certainement appel à un logiciel –, se demander si la représentation graphique que l'on a devant les yeux peut être obtenue à l'aide d'un logiciel statistique est un moyen assez fiable pour distinguer la visualisation de données des représentations « dessinées » sur la base d'informations.

Alors que ce premier axe de notre typologie laisse penser que la visualisation de données est un art plus pur que celui du *design* d'information, cet entremêlement nous rappelle au contraire que ces deux types de représentations graphiques sont complémentaires puisqu'ils expriment des sources de nature différente.

Visualiser pour (dé)montrer ou pour chercher

Typologie des usages			
Illustration	Démonstration	Interface	Recherche
À but <u>décoratif</u> , la visualisation n'a pas de sens particulier	Utile à la simplification et à la communication d'un résultat, la visualisation donne à <u>voir</u> les données	Objet interactif, la visualisation est un <u>point d'accès</u> aux données	Partie prenante du processus de recherche, la visualisation aide le chercheur à <u>comprendre</u> ses données

L'axe présenté plus haut distinguant les visualisations selon leur source, il propose une typologie triviale au chercheur qui sait précisément sur quel genre de matériel il travaille. Ce deuxième axe introduit une typologie souvent beaucoup moins évidente. En effet, si la différence entre une visualisation produite pour illustrer ou rendre visible un phénomène (observé dans les données) et celle qui est réalisée sur la base d'un jeu de données difficile à synthétiser ou trop large et complexe semble assez claire en théorie, elle ne l'est souvent plus au moment de la confrontation pratique. Suis-je en train de produire cette image pour moi-même ou pour mon lecteur ? M'apporte-t-elle réellement une connaissance ou des hypothèses nouvelles sur mon objet ou m'aide-t-elle surtout à exprimer, simplifier et résumer mon jeu de données pour que mon lecteur y voie plus clair en un coup d'œil (la simplification visuelle étant donc un biais introduit volontairement pour faciliter le cheminement interprétatif du lecteur) ? Dans les faits, il n'est pas rare que, partant du principe que la visualisation participe au processus herméneutique et va nous aider à comprendre nos données, on se lance dans une visualisation « de recherche » avant de se rendre compte que celle-ci n'apporte finalement rien de particulier. À l'inverse, il se peut également que la représentation graphique d'un jeu de données apparemment simple n'en soit pas moins épiphanique et qu'elle fasse apparaître une relation ou une tendance insoupçonnée qui nous donne matière à réflexion, ou que ce processus de visualisation permette de repérer des erreurs dans un fichier mal encodé. À mi-chemin entre ces deux approches, le développement récent de technologies permettant de produire des visualisations interactives brouille encore la distinction entre usage démonstratif et heuristique : si l'utilisateur a désormais en partie la main sur l'interface, on le place en position de chercheur en lui donnant accès à tout ou partie des données, ce qui ne signifie pas pour autant qu'il n'est pas guidé par les choix initiaux de la plateforme.

Derrière cette typologie des usages apparaît donc une gradation de complexité implicite : plus le jeu de données est complexe, plus l'utilisation d'un logiciel de traitement statistique ou de visualisation est susceptible d'axer la démarche dans une dimension exploratoire. Mais là encore, point de généralités puisque les usages heuristiques de la représentation graphique de données statistiques n'ont pas attendu la puissance de calcul de l'ordinateur. Les matrices en papier de Jacques Bertin en sont un excellent exemple (Perin *et al.* 2019) : une interface analogique, à proprement parler « bricolée » par le chercheur qui prend un rôle actif dans l'ordonnancement des données, préfigure largement le *big data* et ses algorithmes. Enfin, l'un

des paradoxes de la visualisation « de recherche » est qu'elle produit des représentations graphiques qui ne sont parfois lisibles que par leur auteur, qui s'est plongé suffisamment dans les données et a tourné et retourné l'objet assez longtemps pour le déchiffrer. La publication du résultat visuel n'est donc pas toujours souhaitable puisque trop complexe, peu esthétique et finalement inutile pour le lecteur si ce n'est pour lui permettre de comprendre le cheminement de l'auteur de l'étude.

Afin d'incarner cette typologie des usages, nous proposons ci-dessous de passer brièvement en revue quatre grandes catégories de visualisations.

Les visualisations statistiques : des formes simples pour des résultats visuels solides

Les représentations graphiques de données statistiques sont la catégorie reine de la visualisation de données, parce qu'elles préexistent à la plupart des autres formes (à l'exception notable de la cartographie) et qu'elles sont généralement dénuées d'éléments esthétisants qui pourraient faire obstacle entre la donnée et son interprétation. Leur usage est le plus souvent démonstratif : en effet, une série de lignes ou un histogramme sont basés sur des tableaux de valeurs qu'il est déjà possible de lire à l'œil nu, sans visualisation graphique particulière. Alors que certaines formes simplistes, comme des flèches ou des lignes de tendances montantes/descendantes, ou encore les décriés « camemberts » (*pie chart*) – considérés par Bertin comme « inutiles » (Perin *et al.* 2019, 3) – ne peuvent servir qu'à de l'illustration, d'autres se prêtent mieux à des usages plus exploratoires. C'est le cas des visualisations de *ngrams* lorsqu'elles servent d'interface pour naviguer dans des collections de textes, des grands sets linéaires interactifs où l'objectif est de permettre au lecteur de cliquer pour isoler une série dans son contexte (la pandémie du Covid-19 nous en a fourni de nombreux exemples) ou des graphiques en bulles lorsque ceux-ci évoluent dans le temps, à l'exemple du *Gapminder*⁹ de Hans Rosling.

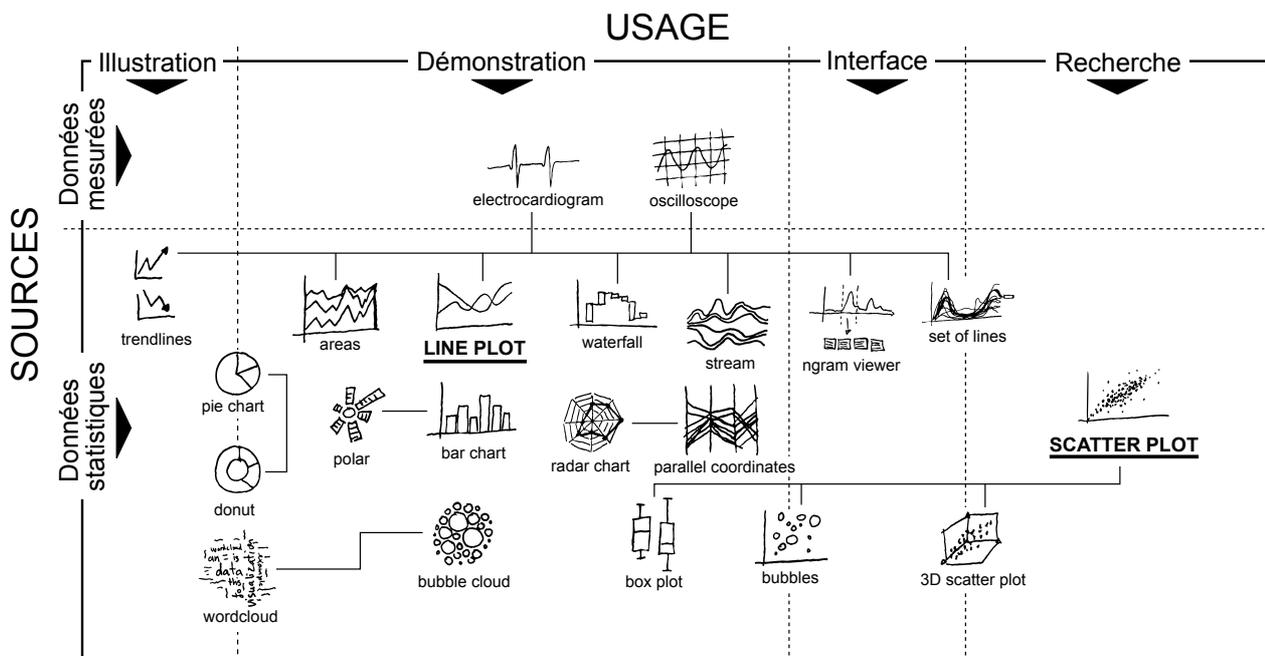


Figure 2. Typologie de la visualisation de données, avec focale autour des représentations statistiques (Grandjean 2022).

⁹ <https://www.gapminder.org/tools/>

De son côté, le nuage de points est la preuve que la visualisation « de recherche » n’aboutit pas forcément à un résultat difficile à lire. Souvent, même s’il est le produit d’un jeu de données volumineux – et donc impossible à « comprendre » pour le chercheur confronté à un fichier comptant des milliers de lignes – sa visualisation est facilement lisible et se prête bien à la mise en évidence de tendances, de distributions, ou le repérage de positions sortant de la moyenne.

La cartographie : une grande variété d’usages

Fascinante en ce qu’elle permet de superposer une couche de données à ce qui nous semble être une réalité tangible (même si « la carte n’est pas le territoire ») – il faut d’ailleurs se méfier de son pouvoir de « séduction » (Bahoken *et al.* 2020) –, la cartographie se prête à des visualisations d’une très grande richesse. Mais une carte est un objet bien différent si elle est le support d’un dessin thématique (comme le serait une carte militaire historique), d’une visualisation de données statistiques (une carte politique par circonscription électorale) ou du produit d’une captation (une carte satellite des émissions de chaleur).

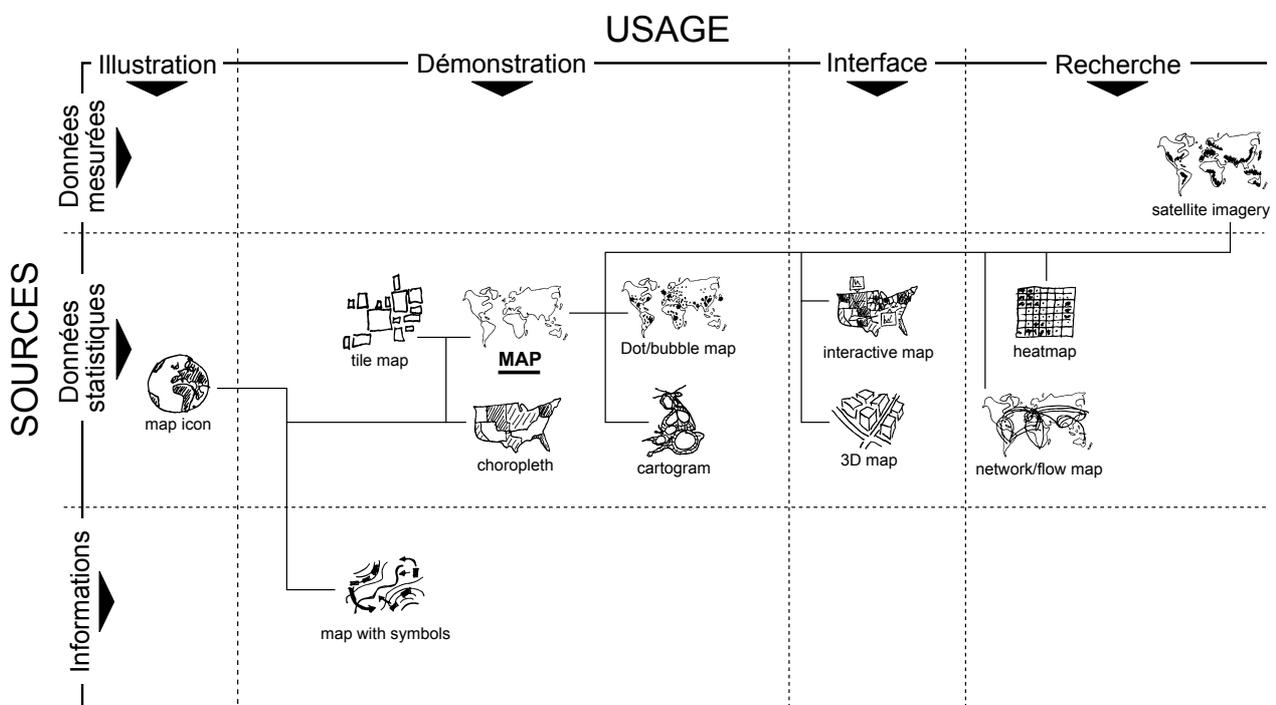


Figure 3. Typologie de la visualisation de données, avec focale autour des représentations cartographiques (Grandjean 2022).

La carte est souvent un objet destiné à être montré, mais, décrivant l’évolution de la cartographie géographique, Éric Guichard (2006, 51) met très bien en évidence en quoi le quotidien du chercheur est transformé par le fait de penser la représentation graphique comme faisant partie du processus de recherche :

Cette « démocratisation » de la vitesse a changé la nature de la cartographie en lui donnant une dimension méthodologique et heuristique qu’elle n’avait pas quand la carte était l’aboutissement de mois de travail. [...] Cette démarche rappelle une attitude scientifique qui s’appuie sur l’expérimentation [...]. Parfois ces allers-retours entre sources et cartes produites invitent à reformuler la problématique à l’origine de la démarche cartographique.

Les réseaux et les flux : des représentations complexes

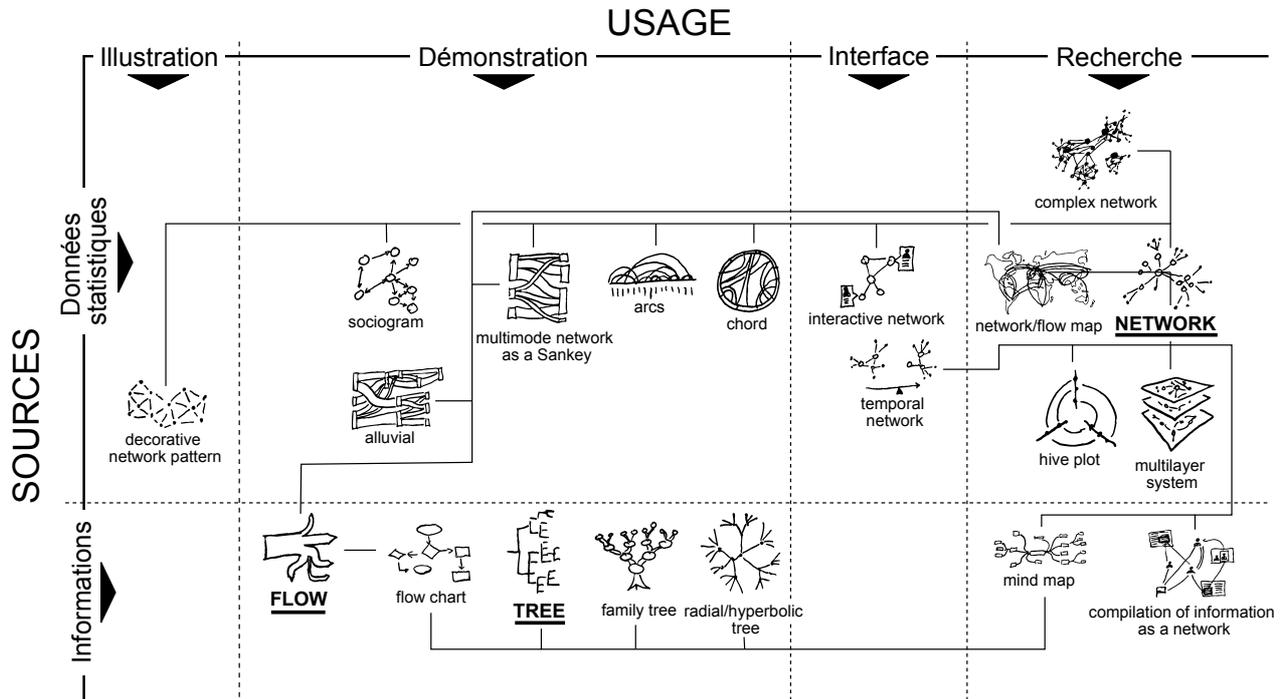


Figure 4. Typologie de la visualisation de données, avec focale autour des représentations de réseaux et de flux (Grandjean 2022).

Destinés à représenter des données relationnelles, c'est-à-dire fondamentalement des matrices (et donc des jeux de données difficilement « lisibles » dans un tableur), les visualisations de réseaux sont souvent des objets complexes. En conséquence, on les lit la plupart du temps comme des cartes (Grandjean et Jacomy 2019), avec une certaine distance, une perspective topologique qui met en évidence des *patterns*, des *clusters* ou des éléments centraux/périphériques (alors que d'autres ne les visualisent tout simplement pas pour en analyser uniquement les caractéristiques statistiques). Si les sociogrammes sont en général toutefois prévus pour une lecture plus proche, la plupart des autres usages de démonstration utilisent des formes simplifiées, empruntant aux représentations de flux. Dans le cas des réseaux, il est intéressant de constater que la typologie des usages (fig. 4, horizontal) est parfois difficile à discerner en raison de leur caractère subtil et des niveaux de lecture variables. De son côté, la typologie des sources (fig. 4, vertical) est particulièrement intéressante : il n'est pas rare d'emprunter la forme visuelle du réseau pour des visualisations qui ne se fondent pas sur des données à proprement parler : des *flow charts* aux *mind maps* en passant par les schémas compilant des informations hétérogènes (images, événements, personnes, textes, etc.), les représentations relationnelles ne sont pas toutes des visualisations de données et les formes les plus simples ont parfois des usages qui s'apparentent plus à de la recherche qu'à de la démonstration (on traduit bien *mind map* par « carte heuristique »).

Les visualisations d'information, entre dessin et mise en ordre des données

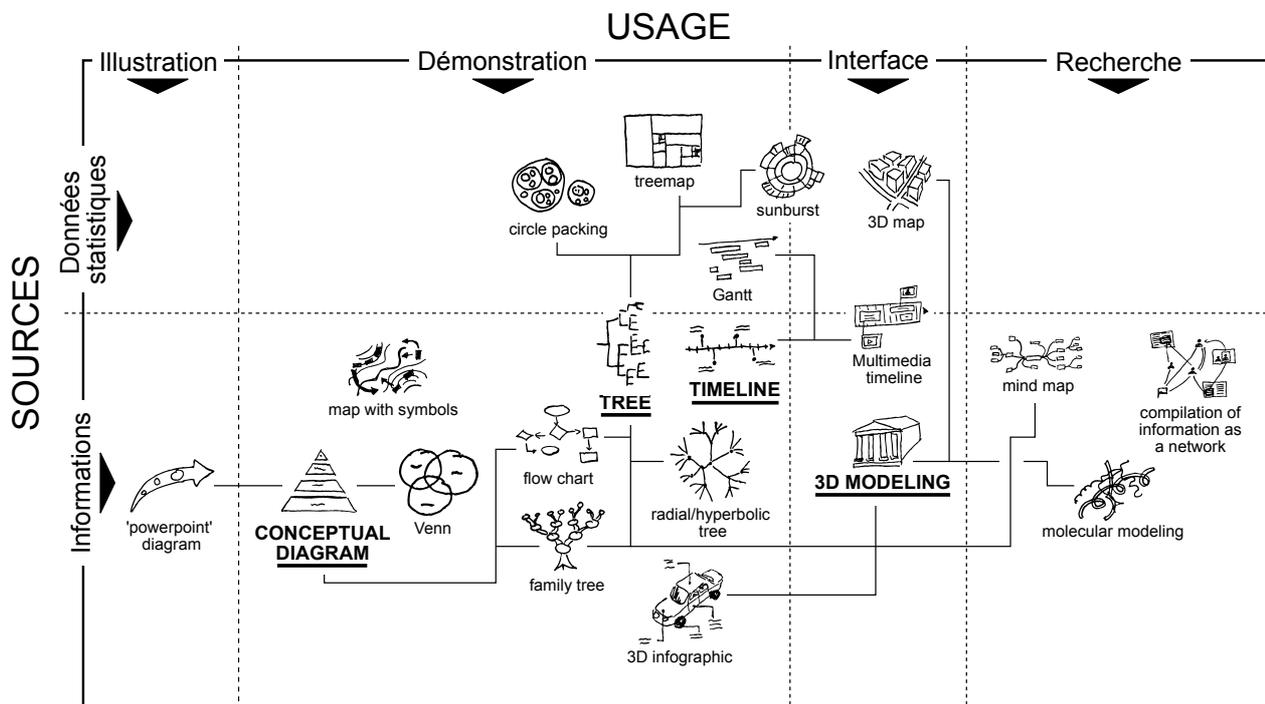


Figure 5. Typologie de la visualisation de données, avec focale autour des représentations d'information (Grandjean 2022).

Lorsque la visualisation a pour objectif principal de représenter une information ou l'organisation de données dans un espace en deux ou trois dimensions, le processus fait souvent plus appel à la créativité qu'au formalisme. Si les diagrammes conceptuels sont par nature des objets destinés à communiquer un principe, d'autres formes de dessins d'information sont à la limite de la visualisation de données statistiques : c'est par exemple le cas dans la catégorie des arborescences où l'organisation d'un dendrogramme/d'une *treemap* est bien un agencement visuel mais la taille de ses classes l'expression d'un jeu de données. Il en va de même pour une ligne du temps interactive : elle est visualisation de données pure si elle est construite sur des intervalles temporels, comme les biographies de Priestley (1765) mais se rapproche de l'infographie interactive dès lors qu'elle sert d'écran à des contenus multimédias.

Conclusion

Cet exercice de typologie montre que la démocratisation de la visualisation de données doit être accompagnée d'une éducation à l'usage et à la lecture de tels objets visuels. Même s'il est désormais courant de rencontrer des représentations graphiques de données statistiques dans les publications scientifiques, il s'agit la plupart du temps de formes très simples et relativement intuitives : histogrammes, courbes ou cartes géographiques, par exemple. Souvent, cette iconographie n'est d'ailleurs là que pour accompagner un propos, elle est subordonnée au texte et n'a donc qu'un rôle d'illustration, au mieux permet-elle de décrire une situation en quelques formes colorées, en espérant donner raison à l'adage « une image vaut mille mots ». En conséquence, la présentation d'une visualisation « de recherche » issue

d'un traitement de données complexes, comme une analyse des correspondances multiples ou une visualisation de réseau massive peut susciter des réactions mitigées parce qu'elle est difficilement lisible par un lecteur non averti. Alors que certains refusent ces visualisations comme moyen de preuve parce qu'ils soupçonnent – parfois à juste titre – qu'elles ne sont qu'une façon de chercher à impressionner, à pallier une faiblesse argumentative ou à cacher l'absence d'hypothèse de recherche, d'autres risquent au contraire de se laisser emporter par l'esthétique séduisante et l'impression de totalité que dégage un tel objet visuel, et d'en accepter l'interprétation sans remettre en question les choix de modélisation, l'analyse et les données sur lesquelles il se fonde. Ces deux réactions sont le produit d'une même cause, une forme d'« analphabétisme visuel », une incapacité à lire ces graphiques et à comprendre les enjeux de visualisation qui les sous-tendent. Il est donc d'autant plus important d'éviter les malentendus et, tout en utilisant la visualisation de données à son plein potentiel, d'être bien clair sur le statut de la représentation graphique dans le processus de recherche et son contexte rhétorique. Ainsi, même si le futur de l'analyse de données semble être dans la massification et l'automatisation (Andrienko *et al.* 2020), le défi réside probablement dans notre capacité à développer des stratégies permettant de maintenir l'intelligibilité et la simplicité de la visualisation tout en ne sacrifiant pas sa fonction heuristique.

Bibliographie

- ANDRIENKO, Gennady *et al.*, 2020. « Big Data Visualization and Analytics: Future Research Challenges and Emerging Applications », *BigVis 2020*.
- BAHOKEN, Françoise *et al.*, 2020. « La séduction des cartes du geoweb. Le cas des flux de migrants internationaux », *Cybergeog : European Journal of Geography*.
- BERTIN, Jacques, 1967. *Sémiologie graphique. Les diagrammes. Les réseaux. Les cartes*, Paris, Mouton et Gauthier-Villars.
- BOSTOCK, Michael *et al.*, 2011. « D3: Data-Driven Documents », *IEEE Trans. Visualization & Comp. Graphics*.
- CAIRO, Alberto, 2012. *The Functional Art: an introduction to information graphics and visualization*, Indianapolis, New Riders.
- DEBLUË, Claire-Lise, 2018. « De l'image numérique à la photographie : exposer les savoirs statistiques au tournant du 20^e siècle », *Transbordeur : photographie, histoire, société*, 2, 54-65.
- DRUCKER, Johanna, 2020. *Visualization and Interpretation. Humanistic Approaches to Display*, Cambridge, MIT Press.
- DRUCKER, Johanna, 2011. « Humanities Approaches to Graphical Display », *Digital Humanities Quarterly*, 5, 1.
- DRUCKER, Johanna, 2010. « Graphesis: Visual Knowledge Production and Representation », *Poetess Archive Journal*, 2, 1.
- FRIENDLY, Michael, 2007. « A Brief History of Data Visualization », in CHEN C. *et alii* (dir.), *Handbook of Computational Statistics: Data Visualization*, Heidelberg, Springer, 1-34.
- GRANDJEAN, Martin, 2014a. « Big Data littéraire. Et si Google n'avait pas compris son propre outil de 'Ngram' ? », <http://www.martingrandjean.ch/bigdata-litteraire-et-si-google-navait-pas-compris-le-ngram/>
- GRANDJEAN, Martin, 2014b. « Historical Data Visualization: Minard's map vectorized and revisited », martingrandjean.ch/historical-data-visualization-minard-map.
- GRANDJEAN, Martin, 2015. « Introduction à la visualisation de données : l'analyse de réseau en histoire », *Histoire et Informatique*, 18, 107-126.
- GRANDJEAN, Martin, 2018. *Les réseaux de la coopération intellectuelle. La Société des Nations comme actrice des échanges scientifiques et culturels dans l'entre-deux-guerres*, Lausanne, Université de Lausanne.

- GRANDJEAN, Martin ; JACOMY, Mathieu, 2019. « Translating Networks: Assessing Correspondence Between Network Visualisation and Analytics », *Digital Humanities 2019*, Utrecht.
- GUICHARD, Eric, 2006. « L'internet : retrouvailles de l'écriture et de la cartographie », *Revue de la Bibliothèque nationale de France*, 24, 51-55.
- JESSOP, Martyn, 2008. « Digital visualization as a scholarly activity », *Literary and Linguistic Computing*, 23, 3, 281-293.
- MANOVICH, Lev, 2011. « What is visualisation? », *Visual Studies*, 26, 1, 36-49.
- MICHEL Jean-Baptiste *et al.*, 2011. « Quantitative Analysis of Culture Using Millions of Digitized Books », *Science*, 331, 6014, 176-182.
- PERIN, Charles *et al.*, 2019. « Jacques Bertin's legacy in information visualization and the reorderable matrix », *Cartography and Geographic Information Science*, 46, 2, 176-181.
- PLAYFAIR, William, 1786. *Commercial and Political Atlas*, London.
- PRIESTLEY, Joseph, 1765. *A chart of biography*, London.
- ROBINSON, Arthur H., 1967. « The thematic maps of Charles Joseph Minard », *Imago Mundi*, 21, 1, 95-108.
- SCHWABISH, Jonathan, 2021. *Better Data Visualizations*, Columbia University Press.
- TUFTE, Edward R., 1983. *The Visual Display of Quantitative Information*, Cheshire, Graphics Press.
- TUKEY, John W., 1977. *Exploratory Data Analysis*, Reading, Pearson.
- YAU, Nathan, 2011. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*, Indianapolis, Wiley.