

Comment présenter nos données pour mieux communiquer ?

La datavisualisation : synthèse et simplicité



Christine Lagarenne*, Frédéric Minodier** et Odile Samson***

L'image est un formidable vecteur de transmission : les messages passent quasi instantanément de l'œil au cerveau pour devenir information. L'image a été mise au service de la statistique, en particulier au XIX^e siècle où cartes et tableaux ont été popularisés en France. Pour en augmenter l'impact, elle a été adossée à un récit qui en explicite les messages ; les technologies web ont complété l'éventail en facilitant l'accès à un très grand nombre de données et en les appréhendant à travers des supports dynamiques. La sémiologie graphique et le design sont complémentaires : la sémiologie améliore l'efficacité des illustrations pour la bonne compréhension et le design réduit l'effort de lecture. Dans la réalisation de la datavisualisation ou dataviz, les règles fondamentales du métier de statisticien doivent être respectées (métadonnées, rigueur et présentation du chiffre). Des solutions simples et génériques favorisent l'expérience utilisateur et permettent au statisticien de conserver la maîtrise technique dans un environnement hautement évolutif. Avec la dataviz comme vecteur de communication, la statistique publique continue à éclairer le débat public et à l'alimenter, avec toujours plus de données accessibles au plus grand nombre.

 *The image is a powerful vector of transmission: messages are passing quite instantly from the eye to the brain where they become information. That's why image has been used by statisticians, starting from 19th century when maps and charts were made popular (Rendgen, 2020). Storytelling was added to spell out the intended message, before web technologies completed the scope by enabling the general public to easily access a very large amount of data and by helping to comprehend them through dynamic media. Graphic semiology and design are complementary: semiology has improved the effectiveness of illustrations in terms of comprehension; design helps to reduce the reading effort. When it comes to producing a datavisualisation or dataviz, the statistician's basic rules must be kept in mind (metadata, rigour and presentation of figures). When implementing, simple and generic solutions should drive choices to enhance user experience and enable statisticians to keep the control for technology, especially in the highly evolving environment. Using dataviz, a mean of communication, official statistics continue to enlighten public debate and to keep it well informed with large amounts of highly accessible data.*

* Cheffe du département de l'offre éditoriale, DDAR, Insee.
christine.lagarenne@insee.fr

** À la date de la rédaction, chef de la division édition et diffusion internet, DDAR, Insee.
frederic.minodier@insee.fr

*** Cheffe de la section graphique, DDAR, Insee.
odile.samson@insee.fr

La datavisualisation ou dataviz regroupe l'ensemble des représentations visuelles des données, depuis les diagrammes en bâtons utilisés de longue date jusqu'aux infographies et tableaux de bord dynamiques. Quel que soit le support, elle est vecteur d'information ou de communication, au format papier ou de plus en plus numérique. Elle permet de rendre compte de façon synthétique d'un ensemble de données mais aussi d'accompagner l'utilisateur dans sa lecture afin qu'il s'approprie plus facilement les résultats d'une étude.

“ **La datavisualisation ou dataviz regroupe l'ensemble des représentations visuelles des données.** ”

Dans le contexte de développement massif des données, la datavisualisation est une opportunité pour la statistique publique de les exploiter encore davantage. Elle permet de diffuser, d'une part, les données de manière plus élaborée et,

d'autre part, les résultats des études statistiques et économiques de façon plus accessible. Elle participe à la stratégie de lutte contre les infox (*fake news*) en touchant un public large grâce à sa facilité de lecture.

L'évolution des différentes formes de dataviz est illustrée par quelques exemples de la littérature sur le sujet et des pratiques de l'Insee et de services statistiques ministériels. Les pré-requis pour la développer et rester à niveau tout en respectant les critères de diffusion de la statistique publique sont ensuite détaillés.

► L'image, un média très utile...

“ **« Une image vaut mille mots », a dit le philosophe Confucius.** ”

« Une image vaut mille mots », a dit le philosophe Confucius. Les données sont mieux perçues quand elles sont représentées sous un format visuel et non plus textuel. Non seulement le cerveau traite les images à très grande vitesse¹ mais en plus, le taux de reconnaissance d'une image après 3 jours est de 90 % (*Haber, 2013*). Il faut dire qu'en moyenne, on ne lit qu'une dizaine de caractères par fixation oculaire (*Brysbaert, 2019*). Les lecteurs ne lisent pas vraiment dans un premier temps. Ils butinent, ils scannent, ils ont une lecture sélective et associative.

Le cerveau cherche à mettre en forme, à donner une structure signifiante à ce qu'il perçoit, afin de le simplifier et de l'organiser devant la complexité de notre environnement. La théorie de la *Gestalt* (i.e. « forme » en allemand) inspirée notamment par Christian von Ehrenfels (*Dortier, 2012*) au début du XX^e siècle traite de la psychologie de la forme. Il existe six lois, dont trois vont être utiles à la dataviz. Tout d'abord, la loi de la bonne forme, loi principale dont les autres découlent : notre cerveau cherche à reconnaître en toute chose des formes simples et stables qui lui sont familières et l'utilisation directe de ces formes dans les représentations améliore notre rapidité de lecture et de compréhension. Ensuite, la loi de proximité regroupe les points les plus proches les uns des autres et est utilisée dans les histogrammes par exemple.

¹ <https://news.mit.edu/2014/in-the-blink-of-an-eye-0116>.



Notre cerveau cherche à reconnaître en toute chose des formes simples et stables qui lui sont familières.

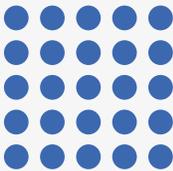


Enfin, la loi de similarité attribue un caractère commun (couleur, taille, etc.) à des éléments permettant de les regrouper visuellement même s'ils sont éparés (*figure 1*).

Pour que le plus grand nombre de personnes puisse s'appropriier les données, notamment statistiques, la forme graphique est un moyen essentiel, largement utilisé. Il permet de synthétiser mais aussi et surtout

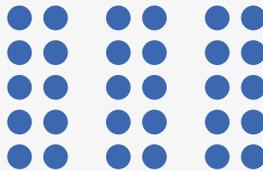
de distinguer les éléments les plus importants. La simplicité de la représentation (courbe, diagramme en bâtons) est primordiale. Des représentations plus techniques ont également fait leurs preuves pour accéder à l'information, en permettre une compréhension rapide et apporter un soutien à l'écrit². Elles font désormais partie de notre vie de tous les jours (*encadré*), comme les cartes géographiques.

► Figure 1 - Trois lois utiles à la dataviz



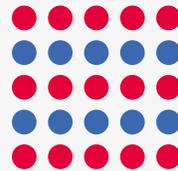
Bonne forme (prégnance)

Notre cerveau cherche à reconnaître des formes simples et stables qui lui sont familières.



Proximité

Les éléments proches sont spontanément associés pour créer un groupe.



Similarité

Les formes similaires (forme, couleur, orientation, etc.) sont regroupées spontanément entre elles.

0	1	2	3	2	6	9	4	5	5	7	6	1
2	0	3	0	6	4	1	0	6	3	0	4	2
1	5	6	3	0	2	5	6	9	7	4	1	0
3	6	5	1	8	0	0	6	7	0	9	1	5
0	1	2	4	8	8	1	4	1	3	6	9	4
1	5	0	2	1	0	8	7	0	6	9	5	4
1	3	5	9	7	3	2	6	9	4	5	5	7
6	1	2	0	3	7	6	4	1	0	3	9	0



0	1	2	3	2	6	9	4	5	5	7	6	1
2	0	3	0	6	4	1	0	6	3	0	4	2
1	5	6	3	0	2	5	6	9	7	4	1	0
3	6	5	1	8	0	0	6	7	0	9	1	5
0	1	2	4	8	8	1	4	1	3	6	9	4
1	5	0	2	1	0	8	7	0	6	9	5	4
1	3	5	9	7	3	2	6	9	4	5	5	7
6	1	2	0	3	7	6	4	1	0	3	9	0

Combien y a-t-il de chiffres 3 ?

Grace à la loi de similarité, juste en coloriant une donnée celle-ci est immédiatement lisible.

² Voir notamment le rapport « La culture statistique des Français : constats, enjeux et perspectives », INSEE N° 2023_14/ DG75-B001, IGÉSR N° 21-22 316A – février 2023, IGAC N° 2023-05. https://intranet.insee.fr/jcms/18876156_DBFileDocument/mi-2022-6-litteratie-statistique-rapport-ig-21-04-2023?details=true.

► Encadré 1. La représentation cartographique, au service des voyageurs

La représentation cartographique s'est appuyée sur une connaissance de plus en plus fine de notre planète au fil des siècles. Les photos aériennes, les satellites dont le fameux *Global Positioning System* (GPS) ont permis de préciser les mappemondes de la Renaissance et les relevés des différentes expéditions scientifiques du siècle des Lumières. Pour autant, la précision de la carte n'est pas toujours nécessaire. C'est ainsi qu'est né le plan du métro londonien dont l'archétype date de 1931. Ce type de plan est maintenant la norme.

Plus proche de nos métiers, les horaires des transports en commun adoptent la forme d'un diagramme arborescent (*steam-and-leaf*). Pour concentrer l'information, l'axe des ordonnées représente les heures et celui des abscisses les minutes de l'horaire de départ du train.

Dans les deux cas, on prend de la distance par rapport à l'objet que l'on veut étudier : on est loin de l'exactitude géographique ou de la représentation linéaire du temps. « La carte n'est pas le territoire » (*Korzybski, 1933*).

► ... qui illustre la statistique depuis le XIX^e siècle

Le développement de la forme graphique s'est fait progressivement. À titre d'exemple, les évolutions temporelles sont représentées sous forme de courbe depuis le X^e siècle (*Andry et alii, 2022*), où elles servaient à décrire le mouvement des planètes. L'utilisation pour la statistique est plus récente.



Cette démarche cherche à valoriser les données chiffrées. L'usage de ces représentations graphiques simples est intentionnel pour permettre une compréhension au premier coup d'œil.



Ainsi, à la fin du XVIII^e siècle, William Playfair a pour la première fois utilisé les trois modes fondamentaux de représentation statistique, à savoir la courbe pour des séries temporelles, le diagramme en bâtons pour des effectifs ou des proportions, et le camembert pour des proportions. Cette démarche cherche à valoriser les données chiffrées. L'usage de ces représentations graphiques simples est intentionnel pour permettre une compréhension au premier coup d'œil. Au XIX^e siècle, l'utilisation des premiers graphiques permet d'identifier des corrélations, et de nouvelles représentations plus complexes apparaissent. Ainsi, en 1854,

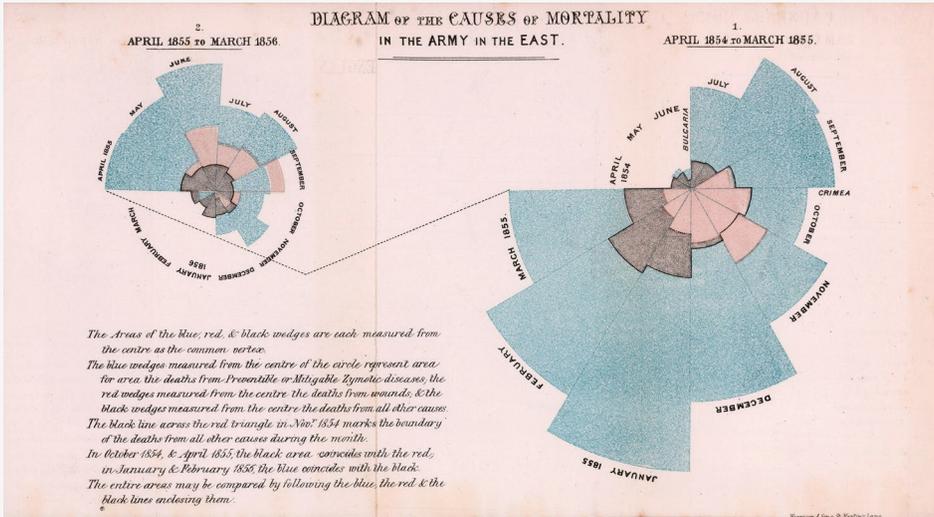
John Snow³ met en regard les décès et les points d'eau sur une carte, ce qui lui permet de trouver l'origine de l'épidémie de choléra à Londres : une pompe infectée.

Infirmière lors de la guerre de Crimée durant l'hiver 1854-1855, Florence Nightingale parvient à attirer l'attention du politique sur le fait qu'on ne meurt pas uniquement sur le champ de bataille, mais surtout des suites de ses blessures en raison de conditions sanitaires déplorables, grâce à un diagramme en crête de coq (*figure 2*). Il s'agit d'un diagramme polaire ; chaque secteur, correspondant à un mois donné est découpé en trois aires concentriques selon trois causes de décès (mort au combat, maladie infectieuse, autre cause).

Qu'il s'agisse de proportions ou d'évolutions temporelles, le statisticien dispose de plusieurs types de représentations graphiques adaptées aux données pour faire parler les chiffres et mettre en avant des résultats, pour établir les faits et convaincre. D'où l'importance du choix de la visualisation pour identifier et porter les messages.

³ "On the Mode of Communication of Cholera" by John Snow, originally published in 1854 by C.F. Cheffins, Lith, Southampton Buildings, London, England. <https://archive.org/details/b28985266/page/n57/mode/2up>.

► **Figure 2 - Diagramme des causes de mortalité dans l'armée**



Source : Harrison and sons, Saint Martin's Lane

► **Raconter des histoires : le storytelling**

Les objectifs d'une dataviz sont non seulement d'accompagner et de faire comprendre les données mais aussi de communiquer, d'illustrer des messages et de donner envie aux lecteurs de lire (accrocher le lecteur). Une datavisualisation est à la frontière entre diffusion et communication.

Pour intéresser le lecteur, raconter une histoire à l'aide de visuels permet de capter son attention : en anglais, on parle de *storytelling*. Il peut prendre des formes variées : la carte de la campagne de Russie réalisée par Charles Joseph Minard (1869), une des premières réalisations de datavisualisation multidimensionnelle, s'apparente au *storytelling* d'avant le numérique.



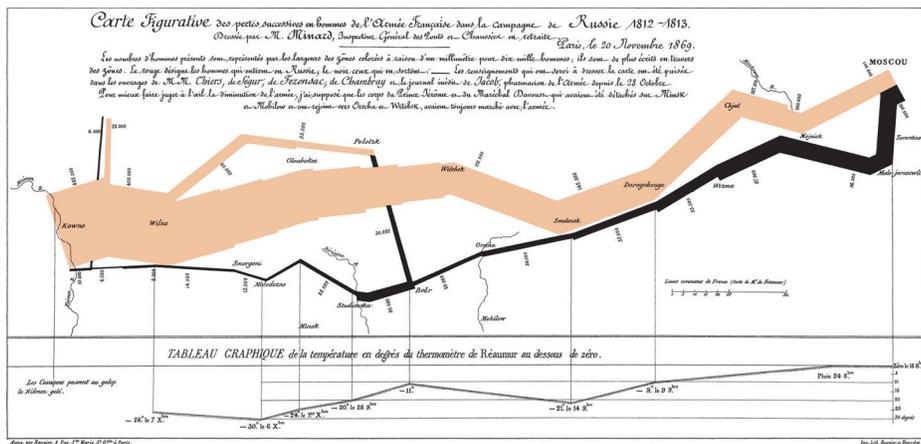
L'auteur réussit à raconter sous forme graphique l'évolution de l'armée française entre juin et décembre 1812.



L'auteur réussit à raconter sous forme graphique l'évolution de l'armée française entre juin et décembre 1812 (**figure 3**). Dans un même espace, il décrit simultanément son itinéraire, en indiquant topographiquement les lieux (cours d'eau, villes) et sa direction (en brun pour la campagne jusqu'à Moscou, en noir pour la retraite) ; la dimension statistique est ajoutée par la largeur de la coloration qui représente la taille de l'armée à

chaque moment, ainsi que l'évolution de la température de l'air pendant la retraite. Minard a développé une cartographie originale dans le but de « faire apprécier immédiatement par l'œil, autant que possible, les proportions des résultats numériques » (*Palsky, 1996*).

► Figure 3 - Carte figurative de la campagne de Russie 1812-1813



Source : Charles Joseph Minard, 1869.
 Reproduction provenant de l'École nationale des Ponts et Chaussées, Paris.

Plus simplement, faire ressortir une information sur un graphique complexe constitue déjà en soi du *data storytelling*. Par exemple, distinguer une courbe parmi plusieurs, isoler une part dans un camembert, fournit aux lecteurs une clé de lecture implicite, un message.

► Vers l'infographie et le défilement narratif (*scrollytelling*), deux formes de *storytelling*

Une infographie regroupe un graphique, un diagramme ou toute autre image visuelle, comme un pictogramme ; elle est destinée à présenter des informations complexes sous une forme facilement compréhensible⁴, adaptée à la diffusion sur support numérique ou papier. À titre d'exemple, une infographie sur l'étude *Femmes et Hommes : une lente décade des inégalités* met en lumière les orientations post-bac des femmes. Les comparaisons femmes-hommes sont illustrées par deux podiums (6 chiffres extraits de deux figures) (*figure 4*).

À l'ère du numérique, une autre forme de scénarisation se développe : le *scrollytelling* qui est une animation de datavis dans une page *web*. Ce terme provient de la contraction entre *storytelling* et *scroll*, le défilement haut-bas sur un écran. Elle incorpore tous les éléments multimédias (sons, textes, vidéos, infographies, animations, photos ou dessins) de façon fluide grâce à la parallaxe⁵. Elle est plus ou moins interactive et peut orienter l'utilisateur vers différents scénarios.

⁴ Selon le Dictionnaire Collins en langue anglaise : *infographics, a graph, diagram, or other visual image designed to present complex information in an easily understandable form*. Les dictionnaires Robert et Larousse donnent une définition plus générique de l'infographie (image via un ordinateur).

⁵ La parallaxe est un effet visuel lié à une vitesse de défilement différente entre le premier et l'arrière-plan lors du *scroll* (défilement).

► Figure 4 - Part des femmes dans les différentes formations d'enseignement supérieur en 2020-2021

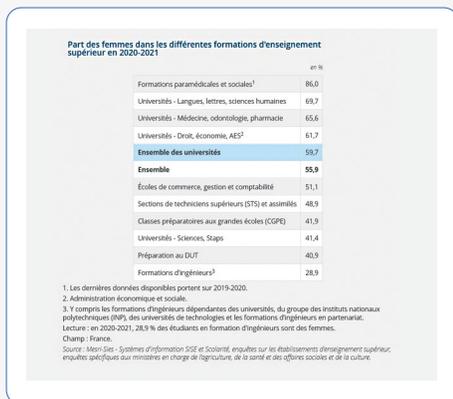
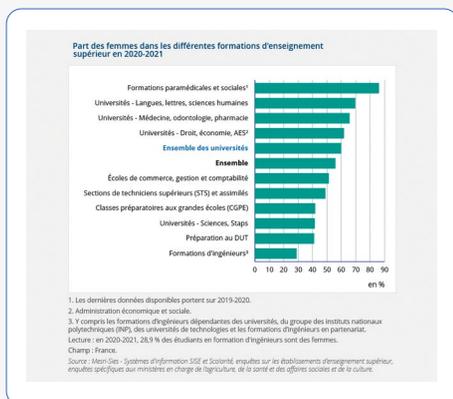


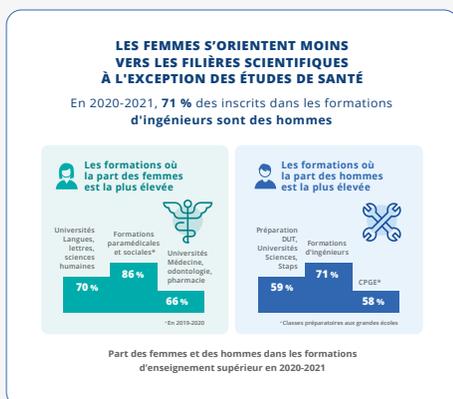
Tableau de données

- Les données sont difficiles à appréhender ;
- Pas d'éléments saillants ;
- Il faut lire, ce qui nécessite du temps.



Datavisualisation

- Mise en évidence des éléments faibles ou forts par la longueur des barres et le jeu des couleurs ;
- Lecture rapide des données ;
- S'accompagne souvent d'un texte explicatif.



Infographie

- Visualisation immédiate des éléments à retenir d'une étude : texte + chiffres ;
- Nécessite pour le producteur un tri de l'information ;
- Jeux de textes, graphiques, chiffres, pictos ;
- Permet une compréhension facile et une meilleure communication.

Collection : Insee Références
 Femmes et hommes, l'égalité en question (Édition 2022)

Le service statistique du ministère chargé de l'agriculture a présenté de cette manière les résultats du recensement agricole de 2020 (*Le Grand*, 2022) ; de même, l'office national de statistique du Royaume-Uni utilise cette façon de faire pour diffuser les résultats du recensement de la population (*How the population changed where you live: Census 2021*). Le *scrollytelling* est engageant pour le lecteur mais nécessite toutefois un grand effort de scénarisation, comme les infographies associées à des études. La démarche consiste à hiérarchiser les résultats, les présenter par ordre d'importance décroissante selon la technique de la pyramide inversée, technique rédactionnelle journalistique (*Angel*, 2009).

► Datavisualisation interactive

Avec le *scrollytelling*, l'utilisateur est actif, il fait défiler les pages, mais la datavisualisation interactive va plus loin dans sa relation avec les chiffres : elle autorise la libre exploration des données par tout un chacun. Elle offre la possibilité à l'utilisateur d'explorer les données, à l'image du statisticien qui les examine avant de les diffuser (**encadré 2**).

Le développement de l'interactivité et la multiplicité des représentations graphiques possibles autorise la prise en main, par le plus grand nombre, d'ensembles de données de grande dimension. Les pyramides des âges interactives ont été l'un des premiers exercices de ce type : l'outil permet par le biais d'une animation de faire défiler les résultats sur plus de 70 ans. Plus proche de nous, l'ouvrage de référence « Tableaux de l'économie française », a été entièrement refondu et repensé dans une optique utilisateur : si la version papier annuelle a été conservée, le produit est aujourd'hui modernisé, actualisé en continu avec les dernières informations, fournissant un meilleur service aux internautes avec différents niveaux de lecture, tout en préservant une simplicité d'alimentation pour les gestionnaires. La définition de cet ouvrage a nécessité des compromis dans le choix des indicateurs et leur mise en valeur, en évaluant leur caractère structurant, leur actualité, leur capacité d'appréhension par le public. La nomenclature des thèmes a été calquée sur celle d'« insee.fr » pour que l'internaute se retrouve facilement dans les rubriques. Dernier en date, l'outil interactif sur *les espérances de vie* rassemble dans une même enveloppe toutes les espérances de vie calculées par l'Insee depuis 1946, déclinées par sexe, âge, territoire, diplôme, niveau de vie et catégorie sociale. L'utilisateur peut s'approprier cette mine d'informations par une vidéo didactique et modifier les différents critères par le jeu de figures interactives.

Cependant, le parcours utilisateur doit être suffisamment réfléchi pour que l'internaute s'y retrouve. L'expérience utilisateur (UX, acronyme pour l'anglais *User eXperience*) doit être pensée grâce à des tests, notamment avec des utilisateurs externes.

► Diffuser massivement...

Le statisticien conserve son rôle d'accompagnement dans la mise à disposition des données. Tout d'abord, le choix des croisements doit être pertinent (une simple liste à la Prévert des variables ne suffit pas pour une diffusion large) tout comme la mise en forme des données. Les métadonnées (concepts, variables et modalités) sont essentielles pour permettre la bonne appréhension des données.



Les métadonnées (concepts, variables et modalités) sont essentielles pour permettre la bonne appréhension des données.



La statistique publique diffuse de plus en plus massivement des données structurées, nécessaires à la datavisualisation. Elle met ainsi à disposition des outils⁶ permettant à la fois de mieux découvrir (catalogue de jeux de données) et mieux explorer les données (tableaux dynamiques) ; les métadonnées sont essentielles pour ces usages. Les outils de gestion des données (*datastewardship* ou *datamanagement*⁷) permettent en effet d'élargir la réutilisation des données ; d'autres acteurs peuvent s'en emparer pour proposer d'autres valorisations des statistiques publiques.

► Encadré 2. La dataviz au service de la qualité

L'aspect exploratoire des données renvoie au travail du statisticien, dans son appropriation initiale des données collectées, avant leur diffusion. Ainsi, « l'analyse exploratoire des données, c'est donc, pêle-mêle, des instruments efficaces et simples de mise en œuvre, des logiciels puissants et ergonomiques, la réhabilitation des graphiques comme outils d'analyse, mais aussi et surtout une attitude du statisticien face à son problème et ses données. » (*Destandau et alii, 1999*).

Le statisticien mobilise différentes représentations des données pour les analyser, repérer des sous-populations particulières, identifier les erreurs ; sur le *web*, les possibilités sont variées, le recours aux outils de datavisualisation contribue à l'amélioration de la qualité des données en multipliant les possibilités de contrôles. À titre d'exemple, une carte de déplacements domicile-études a ainsi permis de

corriger des données aberrantes dans la diffusion du recensement de la population. La première datavisualisation interactive sur les salaires est née des outils de validation de ces données. Cet outil interactif (*Dataviz salaires*) permet une exploration en profondeur de la distribution des salaires, avec différentes dimensions d'analyse (âge, secteur d'activité, géographie, etc.). Ce type d'outil est un moyen d'analyse des données puissant : il permet de repérer, pour une sous-population particulière, certains résultats qui n'auraient pas été mis en évidence ou n'auraient pas trouvé leur place dans les publications à vocation large.

Mis à disposition sur le site de l'Insee, l'outil permet aux internautes d'accéder à des données extrêmement détaillées de façon simple et rapide (*user-friendly*). C'est un exemple d'application du concept de réutilisation dans l'acronyme FAIR^{*}.

* L'acronyme FAIR est utilisé pour désigner les propriétés de l'open data : Findability, Accessibility, Interoperability, and Reuse ou Facilité de recherche, Accessibilité, Interopérabilité et Réutilisation.

► ... en gardant la rigueur du statisticien

Pour viser le plus grand nombre, il est conseillé d'utiliser des mots simples⁸ sans renoncer à la rigueur scientifique ni au recours aux concepts adéquats. Par exemple, il faudra veiller à ce que les définitions soient aisément accessibles sur le site de diffusion.

6 Le projet Melodi développe ces outils pour l'Insee. Les services statistiques des ministères en charge de l'enseignement supérieur, de la santé et de la culture ont déjà des solutions opérationnelles.

7 *Datastewardship* et *datamanagement* sont deux termes utilisés pour désigner les processus, outils et techniques de gestion des données stockées dans une entreprise, dans un triple objectif de cohérence, qualité et sécurité.

<https://www.oracle.com/fr/database/definition-data-steward/>

<https://www.lebigdata.fr/data-management>

8 Pour rendre accessibles à un public élargi certains de ses résultats d'études, la DREES (Direction de la Recherche, des Études, de l'Évaluation et des Statistiques), service statistique du ministère chargé de la santé et des solidarités, publie des transcriptions selon la méthode Facile à lire et à comprendre (FALC).

<https://www.culture.gouv.fr/Thematiques/Developpement-culturel/Culture-et-handicap/Facile-a-lire-et-a-comprendre-FALC-une-methode-utile>

► Encadré 3. La théorie autour de la sémiologie graphique

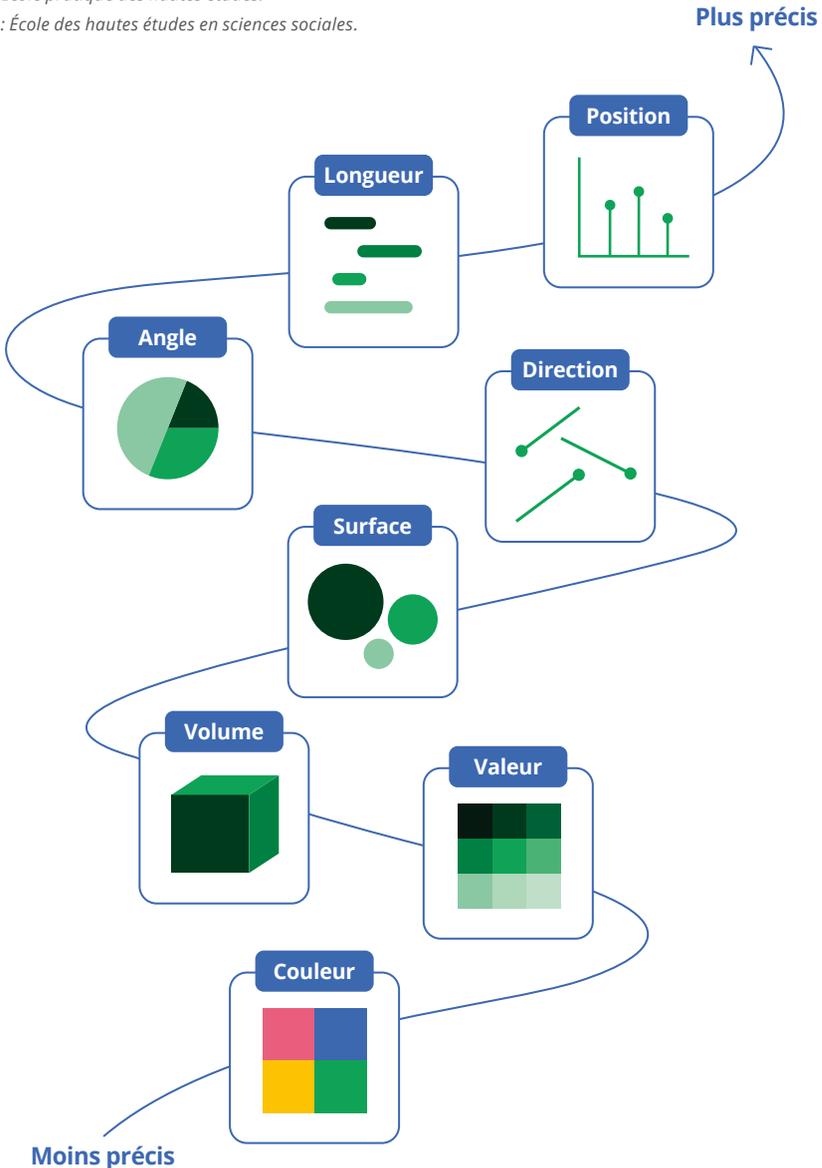
Les variables visuelles clés d'un graphique sont la taille, la valeur, le grain, la couleur, l'orientation et la forme (Bertin, 1967). Selon la nature des caractéristiques des données à représenter, Jacques Bertin a déterminé la combinaison de variables visuelles la plus adaptée : «... si pour obtenir une réponse correcte et complète à une question donnée, [...] une construction requiert un temps d'observation plus court qu'une autre construction, on dira qu'elle est plus efficace pour cette question ». Ce corpus théorique a été construit grâce à l'expérience qu'il

a accumulée au cours de sa carrière au laboratoire de cartographie de l'EPHE* (aujourd'hui EHESS**).

Quelques années plus tard, les fondements d'une théorie scientifique ont été posés pour étayer les principes sémiologiques (Cleveland et Mac Gill, 1984), en s'appuyant sur des tests de perception des valeurs par les différents éléments d'un graphique. Le système précédent de variables visuelles a été raffiné. In fine, ajouter une échelle (ou plusieurs) est préférable aux autres façons d'indiquer une valeur, l'ombre et la couleur n'arrivant qu'en dernier.

* EPHE : École pratique des hautes études.

** EHESS : École des hautes études en sciences sociales.



Plus généralement, pour garder une diffusion cohérente, il est essentiel de respecter une charte graphique et des règles éditoriales, en particulier les règles de la statistique publique sur les outils de dataviz : unité statistique, échelles de données clairement affichées, titre, date, champ pour chaque illustration, affichage des sources, données accessibles en téléchargement, respect du secret statistique (*Darriau, 2020*). Cela participe à l'image de marque de l'institut de statistique.

Comme pour une publication de quatre pages ou plus, la réalisation d'une datavisualisation et notamment d'une infographie nécessite une sélection drastique de l'information à présenter, une focalisation sur les résultats les plus significatifs. Il faut renoncer à tout dire, ce qui peut être difficile pour le chargé d'études. Gardons toutefois en tête que, dans l'univers statistique, la dataviz reste le plus souvent complémentaire au contenu détaillé sous forme d'études, de documents de travail, etc. Elle ne remplace ni les données ni les publications et la méthodologie associée. Elle joue un rôle de produit d'appel, d'incitation à lire l'étude.

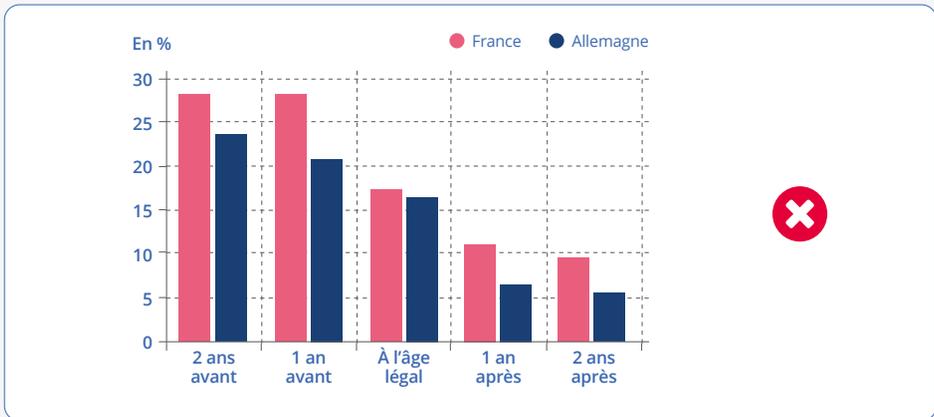
► Avec des représentations graphiques soignées... ---

Une datavisualisation a 30 fois plus de chances d'être vue et lue qu'un simple texte ou tableau. Mais comment faire pour que la dataviz transcrive au mieux les données ? La sémiologie répond à cette question (**encadré 3**) : c'est « l'ensemble des règles d'un système graphique de signes pour la transmission d'une information » (*Bertin, 1967*) qui vise l'efficacité. L'efficacité est renforcée par l'idée du minimalisme (*Tufte, 1983*) : on vise alors « l'excellence graphique ». Cette dernière est atteinte lorsque la quantité d'informations est transmise au lecteur dans un temps minimal et avec le moins d'encre possible. La notion de taux d'encre (*data-ink ratio*) est introduite (**figure 5**) : il s'agit du rapport entre la partie – minimale et nécessaire – du graphique représentant des données qu'on ne peut pas effacer sans réduire les informations diffusées, et le total de l'encre imprimée du graphique. Ainsi, sur un quadrillage, on évitera de mettre une grille apparente qui relègue les données au second plan ou *a minima* gêne la lecture. L'idée est d'éliminer autant que faire se peut tout ce que l'on peut enlever à un graphique sans perdre de sens pour la visualisation (*chart junk*).

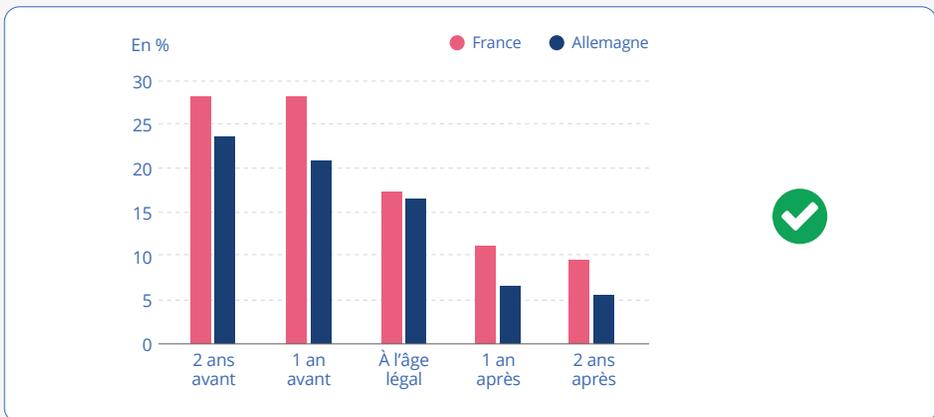
Les couleurs participent également à une meilleure compréhension des graphiques. Mais trop de couleurs rendent un graphique illisible ! Elles doivent être les plus épurées possible avec une palette restreinte et harmonieuse (**figure 6**). Par exemple, il faut attribuer une seule couleur à un seul et même type de données. Enfin, tous les graphiques doivent suivre les normes d'accessibilité par rapport aux différentes déficiences visuelles (daltoniens, malvoyants, etc.).

► Figure 5 - Les taux d'encrage

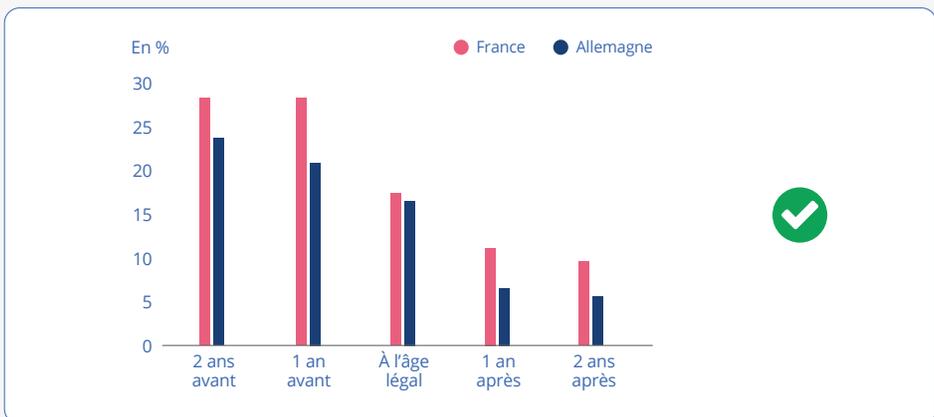
Taux d'encrage important



Taux d'encrage moyen

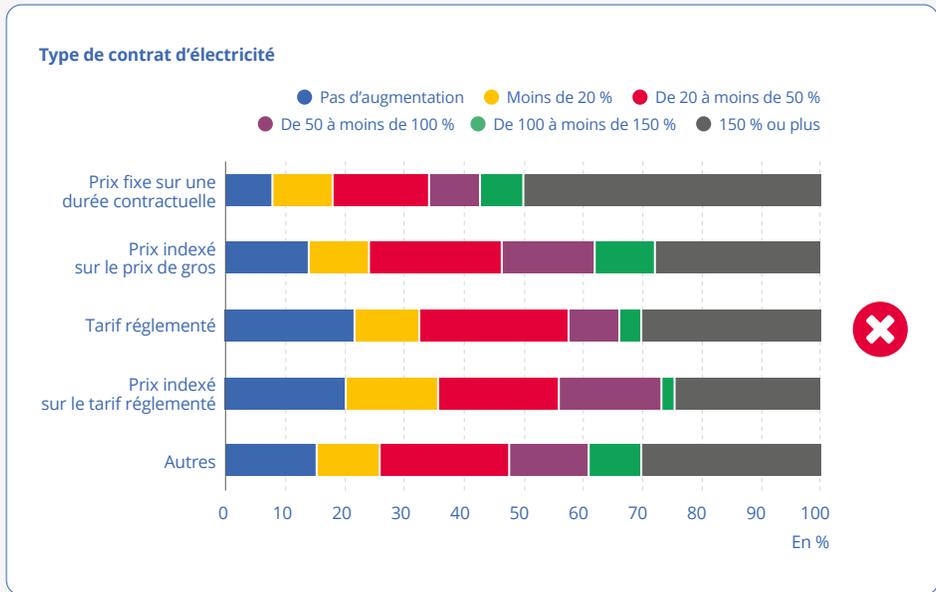


Taux d'encrage bas

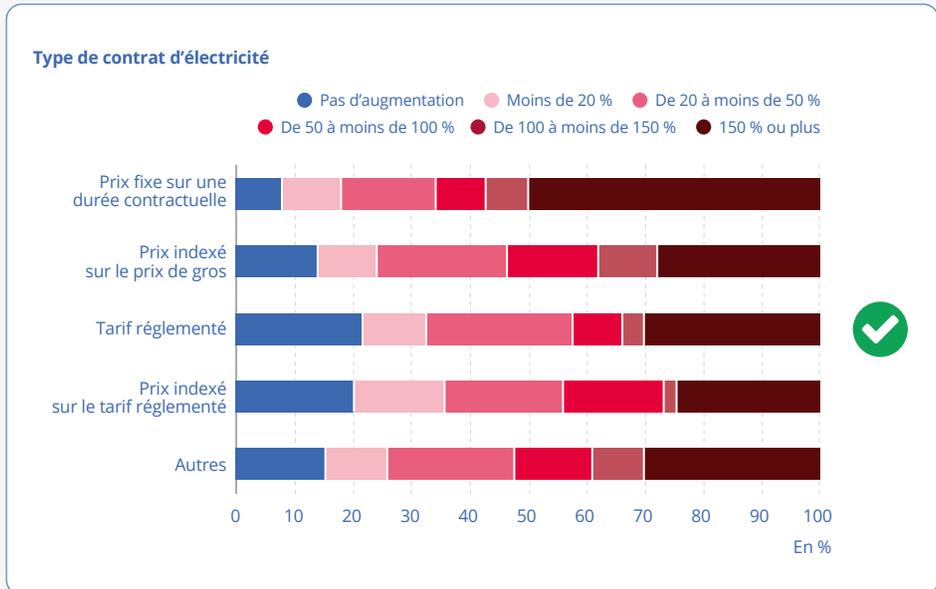


► **Figure 6 - Exemple : Prix de l'électricité**

Trop de couleurs



Palette de couleurs restreinte



► ... et des illustrations parlantes



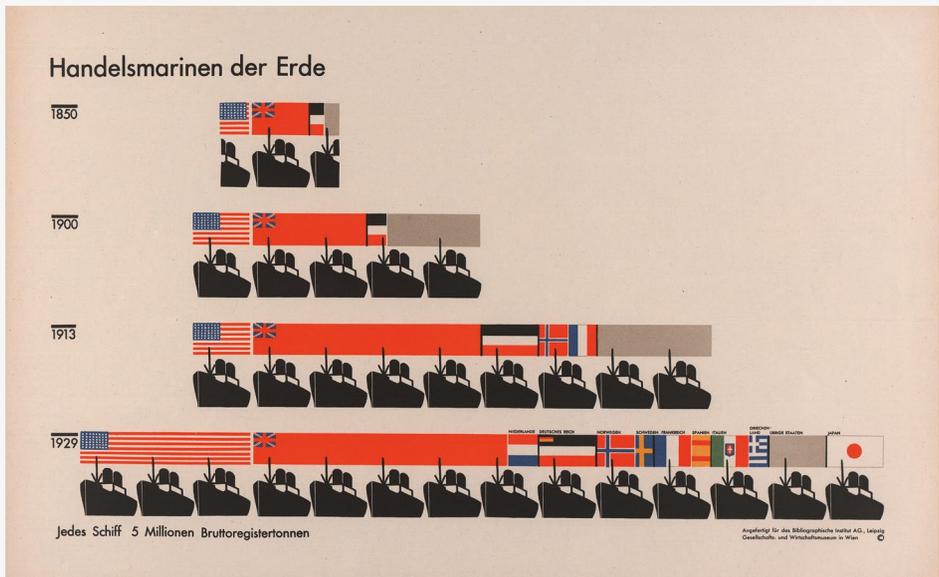
**Une meilleure
accessibilité à
l'information pour la
population la plus large.**



Comme pour le texte, un recours au langage courant facilite la transmission du message à l'exemple de la convention d'usage : rouge pour le négatif et vert pour le positif. Ainsi, en Allemagne, les isotypes (*International System Of Typographic Picture Education*) sont créés (*Neurath et alii, 1925*) formulant un langage visuel, simple et universel influencé par l'esthétique du *Bauhaus*⁹ et permettant une meilleure accessibilité à l'information pour la population la plus large. Ce

sont les ancêtres de nos pictogrammes (*figure 7*) que l'on retrouve dans la dataviz, et surtout dans les infographies.

► Figure 7 - Les marines marchandes du monde



Lecture : Chaque navire représente 5 millions de tonnes brutes de marchandises.

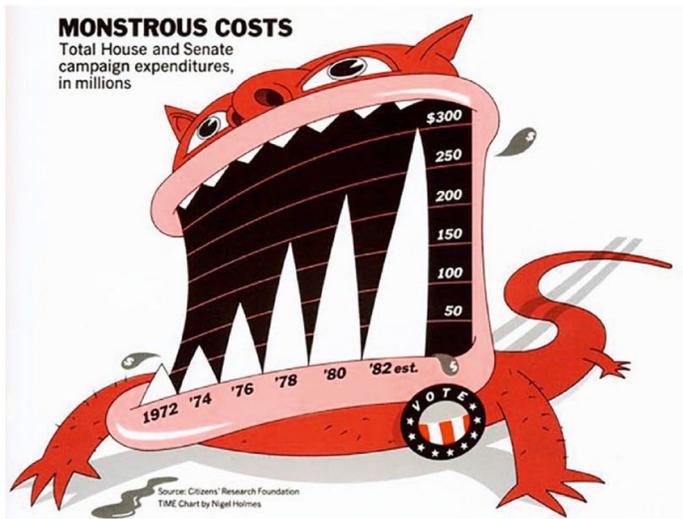
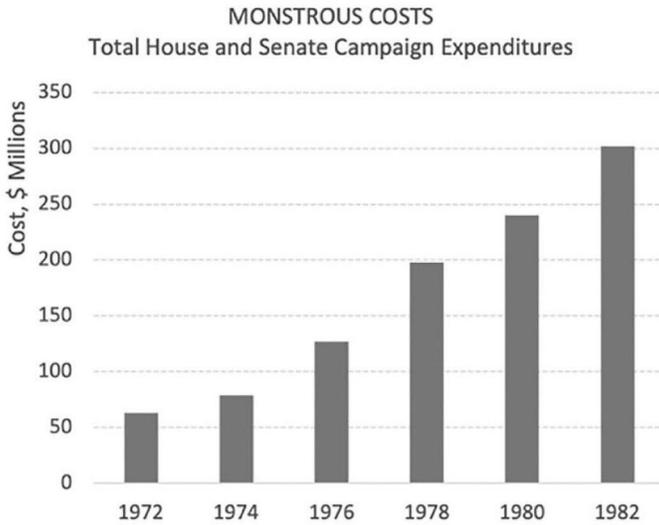
Source : 'Gesellschaft und Wirtschaft' page 55, Verlag des Bibliographischen Instituts AG, Leipzig, 1930.

Les journaux, à grand tirage notamment, utilisent fréquemment des illustrations qui vont au-delà de l'approche minimaliste pour mieux communiquer, comme l'illustre la représentation de coûts jugés monstrueux par un diagramme dans la gueule d'un monstre (*figure 8*). Des « embellissements » peuvent faciliter la transmission d'un message et sa mémorisation (*Bateman et alii, 2010*).

⁹ Courant artistique né en Allemagne à Weimar, 1919, concernant notamment l'architecture et le design.

Ces auteurs montrent, au travers d'une expérience sur deux séries de graphiques (avec et sans embellissements), que la lecture des graphiques n'est pas déformée par les ajouts, que la mémorisation à court terme (5 minutes) n'est pas significativement différente, mais qu'après deux ou trois semaines, les graphiques embellis sont mieux décrits que les autres.

► **Figure 8 - Coûts monstrueux**



Source : Holmes, N. *Designer's Guide to Creating Charts and Diagrams*, Watson-Guptill Publications, 1984.

On parle alors de design émotionnel (ou science cognitive, *Norman, 2012*) selon lequel les images attrayantes se mémorisent mieux parce qu'elles stimulent un plus grand plaisir d'utilisation. Les définitions de l'efficacité – uniquement centrées sur des attributs visuels et fonctionnels des visualisations de données – doivent être élargies, car elles ne prennent pas en compte l'individu et les caractéristiques qui lui sont propres : les émotions (*Kirk, 2016*).

De quoi faire réfléchir un statisticien, non ? Finalement, un juste équilibre doit être trouvé et, s'il est intéressant qu'un graphique marque son lecteur et lui permette de mémoriser le message d'un article, il ne doit pas sombrer dans une approche uniquement séduisante : le message et la rigueur doivent conserver la primauté.

L'auteur veillera à rester neutre, à identifier le vrai message et à ne pas faire mentir les statistiques, comme lorsque le choix de l'échelle masque le phénomène réel ; la dataviz véhicule alors un message significativement différent (*Huff, 1954*). Ce risque de « manipuler » les statistiques n'est pas nouveau, ni propre à leur diffusion de façon visuelle, mais il demeure voire est amplifié. Et ce d'autant plus que les nuances sont difficiles à transcrire ; la dataviz n'est ainsi pas forcément adaptée à tout type de données.

Les principes ci-dessus doivent s'appliquer quels que soient les supports de diffusion et de communication, dont la variété s'est considérablement étendue ces dernières années.

► Une variété de supports formant un écosystème

Au-delà du domaine des publications papier ou *web*, supports classiques, la statistique publique, notamment européenne au travers du programme DIGICOM¹⁰, s'est ouverte dans les années 2010 aux nouveaux supports et canaux de communication : la vidéo avec animation graphique (*motion design*), les réseaux sociaux, notamment sur le smartphone, particulièrement consommateurs d'images. La datavisualisation est donc adaptée à cet écosystème. La vidéo renforce la scénarisation des résultats statistiques, elle se rapproche fortement du *datastorytelling* puisqu'elle intègre un flot de narration (**Encadré 4**).

► Encadré 4. Webographie

Parce que la dataviz n'est pas un ensemble de règles figées mais un corpus en perpétuelle évolution avec une forte part d'innovation et qu'une bonne partie se développe sur internet, les sites ci-dessous permettront de découvrir différentes réalisations. On y trouve à la fois des sites généralistes et des exemples d'applications interactives spécifiques, à consommer sans modération.*

<https://informationisbeautiful.net/>.

<https://www.awwards.com/>.

<https://www.visualcapitalist.com/>.

<https://www.dataviz-inspiration.com/>.

<https://visualisingdata.com/>.

<https://www.data-to-art.com/>.

Exemples de *datascrolling* :

<https://www.spiegel.de/wissenschaft/zirkel-der-genies-a-90c50289-30ac-4a4b-bc49-348676ce6687>.

<https://vizagreste.agriculture.gouv.fr/age-et-devenir-des-exploitations-agricoles.html>.

* Sites en ligne [consultés le 1^{er} décembre 2023].

¹⁰ DIGICOM est un programme européen sur la communication numérique, actif entre 2018 et 2022.

Par ailleurs, les images favorisent la viralité de l'information *via* les médias ou les réseaux sociaux. À titre d'exemple, un tweet sur une publication aura d'autant plus d'impact qu'il sera associé à une image (graphique de la publication ou infographie spécifique). Les principaux résultats chiffrés de l'étude toucheront ainsi un grand nombre de lecteurs qui les garderont en mémoire.

Le recours à ces nouveaux médias engendre des étapes supplémentaires à la diffusion de l'étude ; pour la vidéo, la gestion (à la seconde près) du temps, le choix entre la captation de vue réelle ou le *motion design* doivent être réfléchis et pour les réseaux sociaux, un tweet de 280 caractères issu du chapô doit être rédigé.

► Utiliser des composants les plus génériques possibles... —

Encore plus complexes techniquement, les outils interactifs se sont progressivement développés sur insee.fr. Le premier outil a été le simulateur d'inflation, puis les premières pyramides des âges, adaptés tous deux d'une expérience de l'office allemand de la statistique (Destatis). Il s'agit de véritables applications informatiques conçues le plus souvent autour d'une thématique ciblée.

L'expérience d'Eurostat en la matière est intéressante : dans le cadre du projet DIGICOM, un « *work package* » avait pour objectif de préparer une publication interactive sur l'ensemble du territoire européen, intégrant la traduction dans les différentes langues. Le 20 octobre



Il suffisait de développer un nouveau type de visualisation pour qu'il puisse être utilisé dans d'autres publications.



2017, jour de la statistique européenne, le projet a abouti à la publication « La vie des femmes et des hommes en Europe », reprise sur le site de l'Insee¹¹. Cet exercice a été reconduit sur plusieurs thématiques, dont plusieurs ont été traduites et diffusées sur insee.fr. En effet, l'architecture technique définie, à base de briques graphiques imbriquées avec du texte, était suffisamment souple pour s'adapter à d'autres thématiques et permettait de faire évoluer les fonctionnalités du produit : il suffisait de développer un nouveau type

de visualisation pour qu'il puisse être utilisé dans d'autres publications. Cette évolutivité a permis à Eurostat de mettre en ligne une dizaine d'outils en quelques années, en travaillant essentiellement sur les aspects éditoriaux et non plus sur les briques techniques.

► ... simples et robustes dans le temps —

À l'Insee, le choix privilégié a été un développement interne du site *web* : c'est une volonté liée à la place de la diffusion dans le cœur de métier de l'Institut. La bibliothèque graphique du site insee.fr a été conservée dans le cadre du projet Web4G (2014-2016) : conçue par des statisticiens pour les statisticiens, elle remplissait les fonctions attendues et utilisait une grammaire maîtrisée par les différentes équipes d'alimentation du site. Néanmoins, pour pallier son obsolescence, l'intégration de nouvelles bibliothèques est en cours, permettant de proposer 2 500 nouveaux produits chaque année, à l'état de l'art, tout en conservant la profondeur historique (dix ans) qui fait la richesse du site.

¹¹ <https://www.insee.fr/fr/outil-interactif/3142332/index.html>.

Pour les nouveaux outils interactifs, le retour sur investissement a été jugé meilleur pour les produits pérennes (dont les données sont mises à jour), comme le tableau de bord de l'économie française, que pour d'autres produits *ad hoc*. Cet outil a été conçu pour être alimenté en continu par des équipes internes, en intégrant d'emblée des capacités d'évolution pour le volet territorial. Cette façon de procéder permet de capitaliser en l'adaptant à différentes thématiques et de le maintenir dans la durée.

En effet, il est délicat, dans le cadre de la statistique publique qui porte sur du long terme, de développer des outils au cycle de vie court, au coût de maintenance élevé, avec des données non actualisées.

On est donc loin des solutions sophistiquées, même si de nos jours, avec les dernières innovations, il n'est plus nécessaire d'être un informaticien chevronné pour créer une application et la mettre en ligne sur le *web*.

► Maîtriser un environnement évolutif..

Ceci est rendu possible grâce à l'évolution technologique de ces vingt dernières années. Deux grands axes de transformation ont été mis en œuvre, avec d'une part l'apparition d'un écosystème mondial de la donnée et d'autre part une démocratisation du traitement de la donnée. On constate :

- l'explosion du volume de données diffusées dans les années 2000, avec le stockage sur CD puis DVD avant la généralisation de l'accès en ligne, avec les formats de données XML, CSV, JSON (*Dondon et alii, 2023*) ;
- le développement de l'*open data*, avec la directive Inspire¹² qui ouvre la voie en 2007, l'ouverture de data.gov en 2010 outre-Manche puis le changement de paradigme juridique avec la loi pour une République numérique et le service public de la donnée, également déployé au niveau européen avec la mise en place des jeux de données à forte valeur (*High Value Datasets*) : la mise à disposition gratuite de la donnée, hier encouragée, est aujourd'hui la norme¹³ ;
- le développement des outils de traitement (R v1 en 2000 et apparition de Rstudio en 2011) et notamment de datavisualisation (création de D3.js en 2011) (**Encadré 5**) ;
- le développement de l'accès aux données *via* internet avec notamment les API REST, qui permettent d'échanger facilement les données, et seulement les données. Les services *web*, comme ceux offerts sur le portail api.insee.fr, sont devenus la porte d'entrée vers les données, permettant de filtrer le champ d'intérêt ou les variables utiles.



Des mouvements à la fois rapides et profonds, qui défient le statisticien public.



Ces évolutions s'inscrivent dans des mouvements à la fois rapides et profonds, qui défient le statisticien public. Elles sont en effet sources de contraintes et d'opportunités. L'accès par API permet une mise à jour indépendante de la donnée et de sa mise en forme : une simple mise à jour des données et la nouvelle version de l'outil est prête ; ou encore, un simple changement d'interface et l'ensemble

¹² La directive européenne du 14 mars 2007, dite directive Inspire, vise à établir une infrastructure d'information géographique pour favoriser la protection de l'environnement.

¹³ Loi pour une République numérique du 7 octobre 2016 (article 1).

► Encadré 5. Un éventail d'outils logiciels sans cesse renouvelé pour la datavisualisation

Dans le prolongement de l'informatique décisionnelle, la promesse de l'outil « Tableau »* est de créer un flot de narration. Comme toute solution propriétaire pour laquelle le passage à l'échelle demande un investissement financier conséquent, « Tableau » n'est aujourd'hui pas répandu dans le service statistique public (SSP). En France, le logiciel est utilisé notamment par la DGFIP**. D'autres solutions peuvent être citées : *Qlik*, un acteur relativement ancien, *Microsoft PowerBI* et des solutions utilisées dans le monde journalistique comme *Datawrapper* ou *Infogram*. La logique de ces produits est de définir un produit de datavisualisation comme on le ferait avec un outil de bureautique ; la reproductibilité n'est pas en soi un objectif.

Derrière ces produits clé en main, d'autres outils accessibles nécessitent une prise en main plus complexe mais offrent des services complémentaires utiles pour le statisticien. Des bibliothèques (comme *Gapminder* et *Highcharts*) offrent des ensembles de graphiques simples pour les pages *web*. La présentation d'Hans Rosling*** sur l'évolution sur 200 ans de l'espérance de vie et du revenu de chaque pays est une référence en matière de datavisualisation : elle montre en particulier que, si la réalisation technique est importante, la scénarisation l'est encore plus pour engager le spectateur.

* <https://www.tableau.com/fr-fr>.

** DGFIP : La direction générale des Finances publiques est une direction de l'administration publique centrale française qui dépend du ministère chargé de l'économie.

*** <https://www.presentationzen.com/presentationzen/2010/07/hans-rosling-tips-on-presenting-data.html>.

**** <https://observablehq.com/>.

Le spectre est complété par la bibliothèque Javascript « D3.js », dont la richesse fonctionnelle est illustrée par la page d'accueil de son site. Cette richesse a pour contrepartie une complexité d'utilisation bien plus forte. L'importation et le traitement des données, leur affichage et leur présentation requiert un paramétrage confinant à l'informatique.

Pour le statisticien, les choses sont encore facilitées : nombre de bibliothèques « Javascript » sont portées sous R. Un simple « Enregistrer sous... » avec le package *htmlwidgets* permet alors de disposer d'un code directement lisible par un navigateur.

Le service statistique public a expérimenté différentes voies. La DREES a ainsi développé des datavisualisations sous *Rshiny*, mises sur internet par le biais du site shinyapps.io.

Dans la lignée des *notebooks Jupyter* utilisés pour générer des séquences reproductibles à partir des données, la technologie *Observable*****, développée par Mike Bostock, créateur de D3.js, permet de présenter l'ensemble d'une étude sous la forme d'une suite de morceaux (*chunks*) de texte, code et résultats. Chaque utilisateur peut personnaliser les résultats en les modifiant pour s'attacher à un axe d'analyse particulier ou/et en les exécutant pour dérouler le reste de l'analyse.

de l'historique du jeu de données est accessible. Mais attention, la pression pour une mise à disposition toujours plus rapide risque de nuire à la qualité des données. D'où l'importance d'une automatisation des contrôles sur les données initiales.

Le mouvement *open data* oblige également à une diffusion de plus en plus large, butant potentiellement sur les contraintes de confidentialité et demandant une analyse en profondeur de l'arbitrage entre utilité et protection de la donnée diffusée ; inversement, il donne ou *a minima* facilite l'accès à de nouvelles sources de données, permettant ainsi des croisements plus nombreux, apportant plus d'informations, qu'il faut alors expertiser puis mettre en valeur avec la datavisualisation.

► ... au bénéfice (mesuré) de l'utilisateur ?

Pour mesurer l'impact de la datavisualisation dans la statistique publique, on suit les consultations sur internet des produits concernés, le nombre de vues des vidéos, le nombre de tweets, etc. Le résultat de nos actions devient mesurable et est mesuré en continu.

► Figure 9 - Climate stripes



Lecture : L'artiste local Ian Rolls a créé une nouvelle fresque murale montrant l'augmentation moyenne des températures de l'air à Jersey, afin d'attirer l'attention sur le changement climatique.

Source : <https://www.bailiwickexpress.com/jsy/news/126-reasons-be-green/>. © Copyright Bailiwick Publishing 2023



La datavisualisation doit rester une opportunité pour mieux communiquer et porter une image de marque forte de la statistique publique en accord avec son temps.



L'objectif est non seulement de capter l'attention de l'internaute et de maximiser son temps de lecture mais également de susciter un retour de l'utilisateur (*feedback*) pour vérifier que l'on va bien au-devant de tous les publics et que les chiffres parlent (d'eux-mêmes). Une bonne dataviz peut être largement relayée, augmentant ainsi son impact dans le débat public. À titre d'exemple, l'illustration de l'évolution de la température mondiale annuelle depuis 1850 est partagée et désormais connue (*figure 9*).

La déontologie joue alors un rôle majeur, pour rester neutre et conserver le recul nécessaire à une prise en main objective du chiffre. La datavisualisation doit rester une opportunité pour mieux communiquer et porter une image de marque forte de la statistique publique en accord avec son temps.

► Aller plus loin dans la dataviz pour communiquer

En définitive, la statistique publique a pris le tournant de la datavisualisation : celle-ci a été diffusée à la fois dans une version statique, permettant une meilleure appréhension des publications par l'image et la scénarisation, et dans une version dynamique, offrant à voir des morceaux de données. Les problématiques sur ce dernier volet sont importantes et renvoient à l'identité même du statisticien : jusqu'où doit-on donner à voir les données et quel est le niveau de qualité nécessaire pour la diffusion ?

Des ressources spécifiques sont le plus souvent nécessaires (design, montage) : le besoin d'un « designer talentueux » est identifié (*Bateman et alii, 2010*). Le mariage des compétences du statisticien et du designer graphique est nécessaire pour réaliser des produits esthétiques, engageants mais aussi rigoureux et informatifs.

La large mise à disposition de données permet à d'autres acteurs de s'emparer de celles-ci et de réaliser leurs propres datavisualisations. Les *data*-journalistes relayent ainsi l'information diffusée par la statistique publique de plus en plus souvent sous forme de dataviz, ce qui redonne de la valeur aux données produites.

De manière plus légère, la statistique publique doit-elle s'engager dans le *data-art* qui pousse l'esthétique jusqu'à l'œuvre d'art dans la représentation des données ? L'engagement du « lecteur » est une question essentielle et le chiffre doit marquer les esprits. Lors de l'exposition en partenariat avec la SNCF pour les 75 ans de l'Institut, nos travaux ont été présentés sous la forme de grands panneaux pédagogiques. Au-delà de la qualité de nos données, la présence de l'Insee hors de ses murs est essentielle pour porter ses messages.

**Un moyen
extrêmement efficace
pour partager et
communiquer autour
des statistiques.**

À l'ère du *web*, avec toujours plus de données et de moins en moins de temps pour les lire, la dataviz s'impose comme un moyen extrêmement efficace pour partager et communiquer autour des statistiques. Elle se développe de plus en plus et de façon de plus en plus sophistiquée et la tendance ne semble pas près de s'arrêter. Elle a dès lors un vrai rôle à jouer dans le débat public, et un rôle à part entière dans la communication.

► Bibliographie

- ANDRY, Tiffany, KIEFFER Suzanne et LAMBOTTE François, 2022. De Boeck Supérieur. ISBN 978 2807341579
- ANGEL, Jean-William, 2009. Plan d'un article : inversez la pyramide ! In : *Courrier des statistiques*. Insee Hors série, décembre 2009, pp. 21-24. [en ligne]. [Consulté le 9 août 2023]. Disponible à l'adresse : <https://www.bnsp.insee.fr/ark:/12148/bc6p06xt18x.pdf>.
- BATEMAN, Scott, MANDRYK, Regan L., GUTWIN, Carl, GENEST Aaron, Mc DINE, David et BROOKS, Christopher, 2010. *Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts*. [en ligne]. [Consulté le 9 août 2023]. Disponible à l'adresse : <http://www.stat.columbia.edu/~gelman/communication/Bateman2010.pdf>.
- BERTIN, Jacques, 1967. *Sémiologie graphique : les diagrammes, les réseaux, les cartes*. Gauthier-Villars.
- BERTIN, Jacques, 1977. *La graphique et le traitement graphique de l'information*. Flammarion.
- BRYLSBAERT, Marc, 2019. *How many words do we read per minute? A review and meta-analysis of reading rate*. Journal of Memory and Language, Volume 109, Décembre 2019.
- CLEVELAND, William S. and MCGILL, Robert, 1984. *Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods*. Journal of the American Statistical Association, Vol. 79, No. 387. Septembre 1984, pp. 531-554.
- CLEVELAND, William S. *How William Cleveland Turned Data Visualization Into a Science*. [en ligne]. [Consulté le 9 août 2023]. Disponible à l'adresse : <https://priceconomics.com/how-william-cleveland-turned-data-visualization/>.
- DARRIAU, Valérie, 2020. Les données carroyées, des outils et méthodes innovants. Pour percevoir la réalité des territoires. In : *Courrier des statistiques*. [en ligne]. 31 décembre 2020. N° N5, pp. 53-73. [Consulté le 9 août 2023]. Disponible à l'adresse : <https://www.insee.fr/fr/information/5008701?sommaire=5008710>.
- DESTANDAU, Sophie, LADIRAY Dominique, LE GUEN Monique, 1999. In : *Courrier des statistiques*. [en ligne]. Juin 1999. Insee. N°90. [Consulté le 9 août 2023]. Disponible à l'adresse : <https://www.bnsp.insee.fr/ark:/12148/bc6p06xt287/f1.pdf>.
- DONDON, Alexis et LAMARCHE, Pierre, 2023. Quels formats pour quelles données ? In : *Courrier des statistiques*. [en ligne]. 30 juin 2023. Insee. N°N9, pp 86-103. [Consulté le 1^{er} décembre 2023]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7635827?sommaire=7635842>.
- DORTIER, Jean-François, 2012. *Une histoire des sciences humaines*. pp.150-153. ISBN 978 -2361061678
- HUFF, Darrel, 1954. *How to Lie With Statistics*. Norton, New York, ISBN 0-393-31072-8.
- KIRK, Andy, 2016. *Data visualisation: A handbook for data driven design*. Londres, Royaume-Uni : Sage Publications.

- KORZYBSKI, Alfred, 1933. *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*. p. 58.
- LE GRAND, Hervé, 2022. Le recensement agricole de 2020, cinq innovations qui feront date. In : *Courrier des statistiques*. [en ligne]. 20 janvier 2022. Insee. N° N7, pp. 48-67. [Consulté le 1^{er} décembre 2023]. Disponible à l'adresse : <https://www.insee.fr/fr/information/6035938?sommaire=6035950>.
- NEURATH, Otto, 1939. *Modern Man in the Making*. Lars Muller Publishers. ISBN 978-3037786765.
- NIGHTINGALE, Florence, 1858. "Diagram of the causes of mortality in the army in the East" Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army.
- NORMAN, Donald A., 2012. *Design émotionnel : pourquoi aimons-nous (ou détestons-nous) les objets qui nous entourent ?* Bruxelles, Belgique : De Boeck.
- PALSKY, Gilles, 1996. *Des chiffres et des cartes. Naissance et développement de la cartographie quantitative au XIX^e siècle*. Collection : Mémoires de la section de géographie physique et humaine - n° 19.
- PLAYFAIR, William, 1805. *A Statistical Account of the United States of America by D. F. Donnant*. London: J. Whiting. William Playfair
- RENDGEN, Sandra, 2020. *Le Système Minard - Anthologie des représentations statistiques de Charles-Joseph Minard* - Collection de l'École nationale des ponts et chaussées. 20 novembre 2020. Éditions B42. ISBN 978-2-490077-45-8.
- RODIGHIERO Dario, 2021. *Mapping Affinities: Democratizing Data Visualizations*. Métis Presses.
- STANDING, Lionel, CONEZIO, Jerry et HABER, Ralph, 2013. *Perception and memory for pictures: Single-trial learning of 2500 visual stimuli*. [en ligne]. [Consulté le 9 août 2023]. Disponible à l'adresse : <https://link.springer.com/article/10.3758/BF03337426>.
- TUFTS, Edward, 2001. *The Visual Display of Quantitative Information*. Graphics Press.

