

Diplôme national de master

Domaine — Sciences humaines et sociales

Mention — Sciences de l'information et des bibliothèques

Parcours — Publication numérique

Mémoire — Août 2021

Mémoire

La littératie des datavisualisations à destination du grand public

Auteur — Stéphanie Louison

Sous la direction de M. Éric Guichard
Enseignant-chercheur à l'Essib et à l'ENS de Lyon

REMERCIEMENTS

Je remercie mon directeur de mémoire, M. Éric Guichard, pour ses conseils judicieux, et son travail qui m'a permis de percevoir tous les enjeux de la culture numérique et plus particulièrement des datavisualisations.

Merci beaucoup aux deux datajournalistes qui ont bien voulu m'accorder un peu de leur temps, pour nos échanges enrichissants.

Je remercie également toutes mes camarades de promotion, Sandrine Chenevez, Lina Davy, Lucie Delaunay, Myriam Hamla, Léa Mosnier, Emma Soulier et Marie Vial-Bonacci, pour leur soutien, leurs conseils, et l'excellente année que nous avons passée ensemble.

Un remerciement spécial à Johanna Lamour, qui a commencé à travailler sur la datavisualisation avec moi.

Enfin, un immense merci à mes deux merveilleuses relectrices pour leur soutien, leur attention et leurs conseils ; ainsi qu'à celui qui a eu la patience d'écouter mes analyses.

Résumé

Dans ce mémoire, nous cherchons à identifier une ou plusieurs formes de littératie qui seraient mobilisées par les datavisualisations médiatiques informatisées à destination du grand public. Pour cela, nous définissons de nombreux concepts, en particulier ceux de données, visualisation, datavisualisation, et littératie. Grâce à deux entretiens semi-directifs exploratoires ainsi que de la littérature scientifique, nous présentons l'écosystème du datajournalisme, ainsi que les biais inhérents à la réalisations de datavisualisations. Nous établissons théoriquement les contours d'une littératie des datavisualisations. Cependant, une analyse sémiologique d'un corpus de presse numérique de médias grand public tend à montrer que cette compétence spécifique n'est pas très mobilisée. Les médias américains, et dans leur foulée les français, semblent transférer ces compétences vers une maîtrise de la littératie informatique.

Descripteurs Datavisualisation, représentation graphique de données, datajournalisme, littératie, littératie des données, littératie informatique.

Abstract

In this thesis, we try to identify some kind of literacy that would be necessary to read data visualizations in digital mass media. In that purpose, we define many concepts, such as data, visualization, data visualization, and literacy. Thanks to two exploratory interviews and some scientific literature, we understand the ecosystem of data journalism, as well as the biases involved in the creative process of data visualization. We try to define what would be data visualization literacy. However, our semiological analysis of digital press from mainstream media tends to show that this specific skill is not so usefull in practice. The American media, who pave the way to the French ones, seem to transfer these skills to computer literacy.

Keywords data, visualization, data visualization, literacy, data literacy, computer literacy, data journalism

Droits d'auteur



Cette création est mise à disposition selon le contrat : « **Paternité—Pas d'Utilisation Commerciale—Pas de Modification 4.0 France** » disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/4.0/deed.fr> ou par courrier postal à Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

Introduction	1
Cadre théorique	5
1 Les concepts de donnée et de visualisation	5
1.1 Une définition de la donnée	5
1.2 Une définition de la visualisation	11
2 Le concept de datavisualisation	21
2.1 Les caractéristiques de la datavisualisation	21
2.2 Les représentations graphiques de données à travers l’histoire .	28
2.3 La datavisualisation : un concept, deux approches	36
3 Le concept de littératie	47
3.1 Une définition de la littératie	47
3.2 La littératie informatique	49
3.3 La littératie des données	50
3.4 Quelle littératie pour les datavisualisations ?	52
Analyse	59
4 Méthodologie	59
4.1 Analyse sémiologique	61
4.2 Approche sémiotique dans l’étude des médias	62
4.3 Choix du corpus	63
5 Contexte de production des datavisualisations médiatiques numériques	65
5.1 Compte-rendu des deux entretiens exploratoires	65
5.2 Le datajournalisme	67
5.3 Processus de réalisation des datavisualisations journalistiques .	71
5.4 Récit médiatique et rhétorique des datavisualisations	76

6 Analyse sémiologique d'un corpus de datavisualisations journalistiques	79
6.1 Quelle place pour les datavisualisations dans le média?	80
6.2 Quelle littératie est mobilisée?	88
Conclusion	91
Annexe	103

INTRODUCTION

Depuis le début de la pandémie de COVID-19, les représentations graphiques de données se multiplient dans les médias d'actualité à destination du grand public. La presse écrite, papier comme numérique, les sites institutionnels, ou encore les informations télévisuelles s'appuient chaque jour sur de nouvelles données quantitatives présentées sous la forme de datavisualisations ou d'infographies tantôt statiques, tantôt animées, parfois même interactives.

Pourquoi ce type de productions numériques s'impose-t-il aujourd'hui dans les cercles journalistiques ? Nous pouvons supposer que les médias s'adaptent aux attentes et habitudes de consommation de l'information de leur public. Dans ce cas, la question doit être recentrée sur les récepteurs de ces représentations visuelles. Les lecteurs et/ou spectateurs adhèrent-ils à ce nouveau produit ? Si oui, pour quelles raisons ? Comprennent-ils tous les datavisualisations dont ils sont abreuvés quotidiennement ? Sont-ils en mesure de porter un regard critique sur ces graphiques ?

Ce mémoire tentera d'apporter des réponses à ces questions en s'appuyant, tout d'abord, sur un état de l'art puisant dans la littérature scientifique. En effet, le phénomène décrit précédemment a déjà attiré l'attention de plusieurs chercheurs en sciences sociales, qui ont rédigé des articles répondant en partie à ces interrogations initiales.

Nous approfondirons ensuite une problématique non encore pleinement explorée par la recherche :

Quelle(s) littérature(s) sont mobilisée(s) par les datavisualisations médiatiques informatisées à destination du grand public ?

Cette exploration se basera sur une étude sémiotique de corpus, éclairée par les réponses obtenues lors de deux entretiens exploratoires auprès de datajournalistes.

Cadre théorique

CHAPITRE 1

LES CONCEPTS DE DONNÉE ET DE VISUALISATION

La notion de datavisualisation n'est évidente ni à définir, ni à délimiter. Étudions la construction de ce terme : cet anglicisme est composé du préfixe « *data* », que l'on traduit en français par « données », et du mot « visualisation ». Aussi pouvons-nous en déduire que la « datavisualisation » correspond à une façon de « visualiser » des « données ».

Mais que sont exactement des « données »? Et qu'est-ce que « visualiser »? C'est en définissant ces deux termes que nous parviendrons à construire le concept de datavisualisation.

1.1 Une définition de la donnée

On parle très souvent de « données » depuis l'avènement de l'informatique. Pourtant, selon DATACTIVIST¹ (2019), les données sont apparues en même temps que l'écriture. En effet, ces professionnels estiment que les nombres inscrits vers -3200 sur les tablettes mésopotamiennes (correspondant à des comptes) constituaient déjà des données.

Ainsi, la définition de la donnée n'est pas propre aux domaines de l'informatique et des nouvelles technologies. Nous devons donc trouver une définition large, commune à toutes les disciplines des sciences de l'information et de la communication (SIC).

1.1.1 La donnée en SIC

Dans ce champ de recherche, LELEU-MERVIEL et USEILLE (2008) ont réalisé un état de l'art concernant trois notions qui nous intéressent particulièrement ici : celles de l'information, de la connaissance et de la donnée.

Les deux co-auteurs définissent la donnée dans les termes suivants :

« Une donnée est en principe un fait objectif, le plus souvent quantifiable. »

1. Dataactivist est une société coopérative promouvant l'ouverture des données et leur réutilisation. Pour plus d'informations, voir : <https://dataactivist.coop/fr/>

Cette explication demeure bien vague pour identifier ce qu'est ou n'est pas une donnée.

Référons-nous alors à la citation originale qui a induit cette reformulation. Celle-ci est attribuée à MÉLÈSE² et datée de 1979.

« Une donnée est un enregistrement, dans un code convenu par un groupe social, de certains attributs d'un objet ou d'un événement. »

Cette définition laisse moins de place à l'ambiguïté, et permet d'établir les critères caractéristiques d'une donnée. Celle-ci correspond à un « enregistrement », c'est-à-dire à une consignation sur un support, « dans un code convenu par un groupe social », code correspondant à une forme d'écriture³, d'un contenu qualifié d'« attributs d'un objet ou d'un événement ».

Pour éclairer ce dernier point, il est nécessaire de s'intéresser à la génération de ces « attributs ». Pour cela, il faut lire le paragraphe du même article consacré à MUGUR-SCHÄCHTER⁴. Cette dernière parle plutôt de qualificatifs, qui consistent souvent en des « nombre[s] arbitrairement grand[s] mais fini[s] ». Ceux-ci doivent permettre de décrire et, éventuellement, quantifier une « capture de fragments de substance purement factuels, encore a-conceptuels, obtenus par une découpe volontaire dans la densité du réel ». Les éléments ainsi qualifiés sont donc des faits élémentaires arrachés à leur contexte. Ces faits décontextualisés perdent leur sens, et « sont traités comme une matière première pour des sémantisations progressives ».

Ainsi, nous pouvons résumer en disant qu'une donnée correspond à l'écriture (ou « enregistrement », les deux termes impliquant une conservation dans le temps), dans un code convenu par un groupe social, d'une valeur qualitative ou quantitative caractérisant un fait élémentaire décontextualisé.

Cette définition peut s'appliquer à des données :

- Notées sous forme numérique sur papier (par exemple, un relevé manuscrit de température toutes les heures dans une ville) ;
- Écrites alphabétiquement (une liste des noms de tous les étudiants d'une promotion serait une série de données) ;
- Enregistrées informatiquement dans un code binaire (comme les cookies retiennent nos visites sur des sites web).

LELEU-MERVIEL et USEILLE ajoutent un point important à leur définition : « une donnée est en principe (...) objecti[ve] ». En effet, les données sont censées être décontextualisées. De même, DMYTROVA (2018) estime que « les données n'engagent aucune signification *a priori* ». Cependant, il faut garder à l'esprit que les événements et objets auxquels elles se rapportent sont sélectionnés, de même que les critères et la méthode qui permettent de les qualifier ou encore l'échelle qui les quantifie. Ainsi,

2. Jaques MÉLÈSE. *Approches systémiques des organisations*. Editions Hommes et Techniques. Suresnes, 1979.

3. L'écriture peut être alphabétique ou pictographique, analogique ou numérique. Elle correspond à « un ensemble de signes [fixés sur un support qui peuvent être compris] par une autre personne que celle qui les traça », selon les mots de AVÉROUS VERCLYTTTE, reprenant la thèse de HERRENSCHMIDT. Nous reviendrons plus loin dans ce mémoire sur la notion d'écriture.

4. Mioara MUGUR-SCHÄCHTER. *Sur le tissage des connaissances*. Hermès. Paris, 2006.

les données ne sont pas neutres, mais dépendantes d'un contexte social, qui engage les êtres humains ou les algorithmes programmés pour cela à conserver certaines données plutôt que d'autres et à les enregistrer dans un code qui fait sens pour eux.

1.1.2 Le lien entre donnée, information et connaissance

Comme nous l'avons dit précédemment, les données s'appliquent à un fait décontextualisé. Aussi, il est difficile de donner du sens à une série de données. Si nous reprenons l'exemple du relevé de température, les chiffres notés ne sont d'aucune utilité s'ils ne sont pas rattachés à leur unité (°C ou °F), au lieu où ont été effectuées les mesures, au jour et aux heures correspondant à chaque valeur ou encore à la précision du thermomètre utilisé.

C'est tout l'enjeu de l'information : celle-ci consiste à donner du sens à des données « qui en constituent la matière première » (LELEU-MERVIEL et USEILLE 2008). Ainsi, la mise en relation des données permet de construire une information.

Cependant, celle-ci peut prendre une signification différente en fonction du récepteur, de ses expériences antérieures et de ses connaissances préalables. Par exemple, le relevé de température, une fois associé à toutes autres données qui permettent de le recontextualiser, présente une information concernant l'évolution de la température en un point précis et un jour particulier. Le sens qu'y donnera un météorologue (qui a des points de comparaison en fonction de la ville et peut s'en servir pour faire des prédictions...) est différent de celui que lui attribuera un citoyen lambda (qui se rappellera peut-être simplement qu'il a eu très chaud ce jour-là). Un activiste engagé dans la protection du climat, quant à lui, en extraira encore d'autres informations (par exemple, que ces températures sont néfastes pour telle espèce animale ou végétale).

Aussi, l'information issue de la mise en relation des données prend des sens différents selon le récepteur. Les connaissances construites par chaque personne à partir des mêmes informations sont donc singulières pour chacun (LELEU-MERVIEL et USEILLE 2008).

Cette différence de niveau sémantique entre la donnée, l'information et la connaissance sont souvent représentées sous la forme de la pyramide dite « data information knowledge wisdom », attribuée à Russell Ackoff (1989) : les données, considérées comme une matière première brute, sont traitées et assemblées pour permettre de les interpréter, et de les transformer ainsi en information. Cette dernière est intégrée au système de pensée de chaque personne qui lui attribue une signification en devenant une connaissance. Puis, en donnant du sens à des connaissances, on aboutit théoriquement, selon ce modèle, à la compétence ou à la sagesse (DATACTIVIST 2019).

Ainsi, des données brutes ne représente pas grand-chose sans traitement. Elles ne dévoilent leur potentiel que lorsqu'elles sont agencées pour former une information, qui, elle-même, n'a de signification que relativement aux expériences et connaissances préalables d'un récepteur.

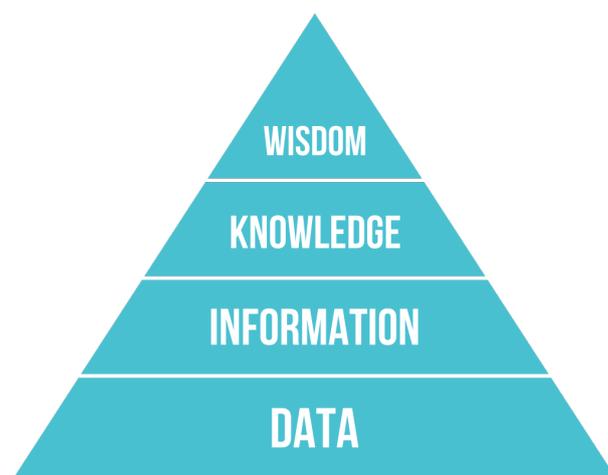


FIGURE 1.1 – Pyramide « data information knowledge wisdom » selon Russell Ackoff (1989)

1.1.3 Un contexte propice à la production et à l'exploitation de données

Puisque les données existent depuis l'invention de l'écriture et qu'elles nécessitent un traitement complexe pour délivrer une information, pourquoi en parle-t-on autant aujourd'hui ?

Le Big data

La réponse à cette question réside dans la masse de données sans précédent à traiter de nos jours. En effet, l'informatique tout entier repose sur des communications et enregistrements de données en code binaire. D'abord exploitées sur machines par des logiciels de type bases de données relationnelles et tableurs (inventés dans les années 1970), ces nouveaux outils se retrouvent dépassés avec l'avènement du web dans les années 1990–2000 (DATACTIVIST 2019).

En effet, chaque contenu publié, chaque consultation, chaque communication en ligne, chaque clic est susceptible de générer des données. Avec 4,5 millions d'internautes en 2020, plusieurs pétaoctets⁵ d'entre elles sont produits chaque année (ALI 2020). Le graphique en figure 1.2 illustre cette frénésie informatique génératrice de données massives.

Ces données massives (communément appelées « Big data ») ont donc commencé à poser problème aux moteurs de recherche dès les années 2000. Google a ainsi créé en 2001 une base de données spécialisée, Big Table, accompagnée dès 2004 d'un algorithme, MapReduce. À partir de ce modèle, le logiciel open source Hadoop voit le jour. Cette référence dans l'univers du Big data gère de gigantesques quantités de données non structurées (BRASSEUR 2015).

5. Un pétaoctet correspond à 10^{15} octets, soit un million de milliards d'octets.

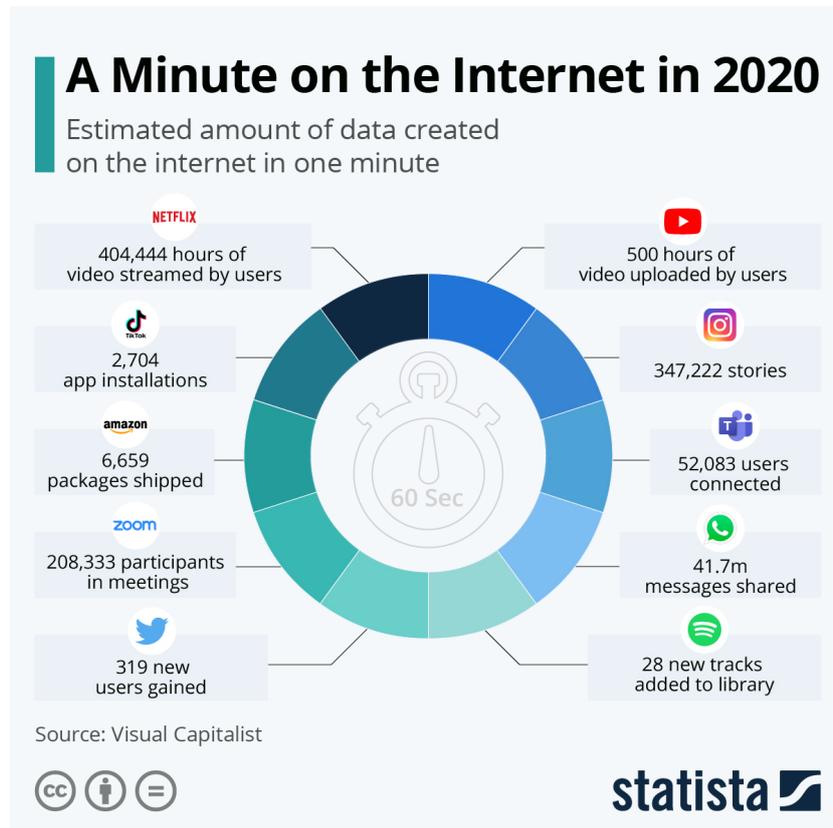


FIGURE 1.2 – Ce qu’il s’est passé en une minute sur internet en 2020
 Datavisualisation réalisée par Claire Jenik, à partir de données collectées par *Visual Capitalist*
 Issue de [https://www.statista.com/chart/17518/
 data-created-in-an-internet-minute/](https://www.statista.com/chart/17518/data-created-in-an-internet-minute/)

Ainsi, avec la démocratisation du web est né le Big data⁶, c’est-à-dire les immenses volumes de données générés sur internet en continu et en provenance de sources diverses. Des outils informatiques ont alors été mis en place pour gérer cet afflux massif et permettre d’en faire une analyse productrice d’information.

L’Open data

En parallèle de ce « déluge de données », des mouvements citoyens demandent à avoir accès à toutes ces sources potentielles d’information. Au nom de la transparence et du contrôle démocratique des institutions, ils réclament l’ouverture des données (J. DENIS et GOËTA 2013).

Si le mouvement de l’Open data s’est formalisé au cours de la rencontre de Sébastopol (en Californie) de 2007 qui a adopté une déclaration définissant les huit grands principes de l’Open Government Data pour faire pression sur les instances politiques (GOËTA 2019), il a des origines plus anciennes. En effet, dès les années 1960–1970, des courants démocratiques tendent

6. Bien que le terme « Big data » n’ait commencé à être employé que dans les années 2010, le phénomène prend bien racines dès le début des années 2000.

à favoriser l'accès aux données publiques (voté en France en 1978), et obligent même les administrations françaises à ouvrir leurs données depuis 1997. C'est finalement sous la présidence d'OBAMA que les États-Unis basculent vers une politique de transparence avec « l'Open government » (PAQUIENSÉGUY 2016). Cette décision engendre une approche similaire dans toutes les démocraties occidentales, avec notamment des directives européennes et des lois françaises en faveur de l'Open data, notamment la *Loi pour une République Numérique* de 2016 (CHAGNOUX 2020). Ainsi, de nombreux jeux de données sont aujourd'hui publiés en accès libre sur des sites tels que <https://www.data.gouv.fr/fr/>, <https://data.europa.eu/fr>, ou encore <https://www.data.gov/> (aux États-Unis) qui propose plus de 310 000 jeux de données.

Dans cette démarche, la figure du citoyen est centrale. C'est lui qui souhaite avoir accès aux données, à travers des communautés telles que Regards citoyens⁷ ou Démocratie ouverte⁸ (PAQUIENSÉGUY 2016). Le citoyen veut être « informé », et contrôler les données enregistrées par ses institutions, et ainsi l'information qui pourrait en découler. C'est ce que J. DENIS et GOËTA évoquent en parlant de « culture de l'audit » (J. DENIS et GOËTA 2013).

Il s'agit d'une façon pour la société civile de montrer son importance, car comme dit précédemment, les données peut générer de l'information, et cette dernière est source de pouvoir. Les citoyens veulent partager cette prérogative avec leur administration.

Pendant, il n'est pas dit que des individus lambdas puissent extraire la moindre information de jeux de données dites « brutes », décontextualisées, alors qu'ils ne savent ni comment elles ont été produites, ni dans quel but (2013). Comme le souligne PAQUIENSÉGUY (2016) :

« Certes le citoyen en sera informé mais sans pouvoir en tirer aucune conclusion pour autant. Dans la très grande majorité des [cas], faire faire sens aux jeux de données afin d'en tirer une information utile n'est pas chose simple pour le citoyen. »

Résumé — La notion de donnée

Nous avons vu que les données, aujourd'hui produites massivement, sont des valeurs codées de façon digitale (sous forme de bits) ou analogique (nombres, mots, lettres, symboles ou encore ondes électromagnétiques) et enregistrées sur un support quelconque (allant de la feuille de papier au disque dur). Ces valeurs qualifient ou quantifient un fait, un événement ou un objet. Cependant, des données isolées ne rendent pas compte du contexte dont elles sont issues et n'ont donc pas de sens en elles-mêmes. C'est par une mise en relation de plusieurs données que l'on arrive à voir émerger une information.

7. Pour accéder au site de Regards citoyens, cliquez sur l'url suivante : <https://www.regardscitoyens.org/qui-sommes-nous/>

8. Pour accéder au site de Démocratie ouverte, cliquez sur l'url suivante : <https://www.democratieouverte.org/nous-connaitre/>

La quantité astronomique de données générées quotidiennement aujourd'hui est donc potentiellement source de très nombreuses informations. C'est pourquoi, à l'ère du Big data, de nombreuses entreprises, sociétés, voire États, investissent dans la captation et le traitement de données, car les informations qu'elles peuvent révéler sont susceptibles de relever d'enjeux stratégiques majeurs. De même, les citoyens veulent avoir un droit de regard dessus afin de contrôler leurs administrations.

Pourtant, révéler le potentiel informationnel des données n'est pas facile. En effet, comment faire en sorte que des accumulations de valeurs révèlent un message lisible rapidement, clair, et sans ambiguïté ? Cette tâche est assurée par la visualisation de données.

1.2 Une définition de la visualisation

Pour faire apparaître le sens des jeux de données, on les représente souvent graphiquement. Le récepteur a alors accès à l'information *via* un visuel. Aussi, dans le cadre d'un questionnement sur la façon dont les usagers comprennent les représentations graphiques d'informations, il convient de s'interroger sur les particularités de la visualisation.

1.2.1 À la base de la visualisation : l'image

Comme REYES, nous considérerons la datavisualisation comme « un type d'image contemporaine » (REYES 2015). Mais qu'est-ce qu'une image ? Quels visuels peuvent être considérés comme tels ?

Délimiter le champ de l'image

METZ (1970) rappelle que l'image est souvent associée à la représentation, plus ou moins réaliste, d'objets ou d'événements. Cependant, il souligne que cette notion d'analogie, c'est-à-dire à le critère de ressemblance d'une image à son sujet, est réductrice. En effet, il existe des images non-figuratives, tels que certains tableaux d'art moderne qui ne représentent rien volontairement, mais donnent à voir malgré tout des couleurs, des formes, des traits, des textures... (METZ 1970). Nous admettons donc qu'une image ne se définit pas par son analogie à un modèle physique.

Nous pouvons néanmoins reprendre le début de la définition précédente, en l'enrichissant de quelques éléments techniques. Nous dirons donc qu'une image correspond à la représentation iconique, plus ou moins réaliste, d'objets ou d'événements en deux dimensions. En effet, nous pouvons remarquer que les images ne comportent jamais plus de deux dimensions, et ne sont donc pas en relief. Autrement, on parle d'objet, de sculpture, de réalité virtuelle... La notion de « représentation graphique » exclura cependant toutes les « séquence[s] de mots (ou de morphèmes) » composants ensuite des phrases (METZ 1970). Autrement dit, nous ne considérerons pas qu'une image puisse être composée d'une suite de mots, mais uniquement d'un ensemble d'icônes, de signes et/ou de formes, assemblés de façon non linéaire (contrairement à un texte). Cependant, nous n'oublions pas que l'image peut être en étroite relation avec du texte (ce qui est bien souvent le cas, comme nous le verrons par la suite).

Cette définition de l'image englobe donc tout autant les peintures que les dessins, les graphiques de données que les affiches publicitaires, les illustrations de livres que les cases de bandes dessinées, les photographies analogiques que celles visibles sur un écran. La technique utilisée pour produire l'image ne sera pas prise en considération, pas plus que le support d'apparition de l'image : celle-ci pourra être tracée à la main au crayon de papier sur une feuille aussi bien que tracée avec une bombe de peinture sur un mur, ou encore par un algorithme sur un écran. Seul l'affichage final nous intéressera dans la notion d'image.

Cette dernière remarque prend toute son importance quand nous constatons que les images sont extrêmement présentes sur l'interface que constitue nos écrans.

L'écran comme interface des images numériques

Les images numériques ont nécessairement certaines particularités liées au medium qui permet leur affichage. Elles sont confrontées à des contraintes de couleurs (liées à des mélanges rouge-vert-bleu qui ne permettent pas la même palette que celle d'un peintre) et de résolution (l'image s'affichant pixel par pixel, elle est dépendante du nombre de pixels par unité de longueur de l'écran). Cependant, notre but n'étant pas ici de porter un avis sur la qualité des images numériques ni de comprendre leurs aspects techniques, nous ne nous pencherons pas sur ces problématiques.

En revanche, afin de comprendre la réception des images (qui nous intéresse particulièrement dans ce mémoire), il nous faut aborder l'élément qui permet aux usagers d'entrer en contact avec l'image numérique : l'interface écranique.

Comme SEUX (2014), nous considérerons l'écran « non pas dans sa seule dimension d'objet (...) mais aussi dans sa qualité de médiateur, c'est-à-dire en tant qu'interface permettant d'accéder à des contenus et générant des échanges ». L'écran constitue donc une interface entre les encodages digitaux qui constituent le langage machine et la perception de l'être humain.

Une interface est un dispositif qui assure une médiation (GIROUD 2021). Cependant, contrairement à cet auteur, nous ne considérerons pas l'interface comme n'ayant « pas de consistance propre ». Au contraire, nous considérerons que l'objet technique qu'est l'écran, délimité par un cadre qui contraint la vision et centre le regard, en est une (SEUX 2014).

En tant qu'un objet technique, l'écran est donc un dispositif qui assure une médiation, au sens de NORMAN, qui a « mis sur le devant de la scène la notion de médiation comme corollaire (...) à celle de dispositif, signifiant hétérogénéité de technique, de naturel et d'humain qui s'interpénètrent et s'auto-constituent »⁹, selon l'état de l'art de PÉQUIGNOT. Ainsi, le dispositif écranique met une technologie au service de l'être humain, qui se l'approprie et le fait évoluer jusqu'à ce que l'objet offre des possibilités d'interactions intuitives (PÉQUIGNOT 2021). C'est ce phénomène que décrit SEUX en parlant de « cacher les contraintes électroniques et computationnelles dans la structure physique de l'objet de manière à faciliter l'accès [aux contenus médiés] » (SEUX 2014).

9. Donald NORMAN. *Raisons pratiques*. T. 4. Les artefacts cognitifs. 1993, p. 15-34.

L'écran devient donc un dispositif médiateur, donc une interface, pour accéder à de très nombreux contenus (cinématographiques, télévisuels, informatiques). En particulier, dans le cas d'un appareil informatique (tablette, ordinateur, smartphone, lecteur mp3, liseuse...), il est indispensable pour permettre à l'être humain d'interpréter les informations codées de façon binaire en les présentant sous une apparence plus signifiante pour l'être humain. Tel est le propos de GIROUD (2021) :

« Le numérique est de nature calculatoire, dans la mesure où il ne porte que sur des entités formelles dénuées de sens, réduites à leur simple manipulabilité, à l'instar des calculi (...), alors seule la médiation du symbole permet de le représenter. »

L'écran permet donc de « manipule[r] et [d']interprète[r] les calculi du numérique » (GIROUD 2021). En effet, c'est bien *via* l'écran que le système technique (l'ordinateur ou tout autre objet informatique) peut être manipulé indirectement, sans avoir à modifier directement les circuits électroniques du disque dur, ou de tout autre élément constitutif du *hardware*. En ce sens, l'interface écranique est « pragmatique dans la mesure où sa finalité est d'assurer une prise, et donc un contrôle sur le numérique » (GIROUD 2021).

Pour permettre à l'être humain d'agir sur le numérique, l'interface écranique s'adapte à lui pour devenir un objet ergonomique. Par exemple, elle reprend une « grammaire d'action » qui représente métaphorique[ment] nos relations quotidiennes au monde, telles que le « bureau », la « poubelle », les « dossiers »...

L'écran se transforme en une interface si intuitive qu'elle nous semble aujourd'hui quasiment invisible, comme le souligne GIROUD : « l'interface s'efface pour ne plus être au centre (...) de notre attention, mais au contraire à sa périphérie » (GIROUD 2021). C'est ce que GOULARD qualifie « d'interfaces ambiantes », en argumentant qu'ils « se fondent dans notre univers quotidien pour nous apporter de l'information en périphérie de notre attention » (GOULARD 2015). Cette situation amène SEUX à aborder la dimension sociale de l'écran :

« L'invisibilité des écrans, ce phénomène paradoxal qui consiste à ne plus voir les écrans alors qu'ils ne cessent de se multiplier, n'est donc pas seulement un effet de notre familiarité avec l'objet, c'est une transparence socialement construite. »

SEUX 2014

Cette dimension sociale, mise en exergue par SEUX s'accompagne de son corollaire culturel. En effet, l'écran, comme tout dispositif de médiation, « modifie notre perception » (GIROUD 2021), puisque des aspects technologiques contraignent les contenus affichés. Par exemple, cette interface ne permet pas l'affichage en trois dimensions, fait appel exclusivement à la vue (avec un ajout possible d'interface sonore, mais l'informatique ne convoque pas les cinq sens à ce jour), elle privilégie le défilement par *scrolling* vertical... Ainsi, cette interface homme/ordinateur modifie les pratiques des individus et donc leur culture (GIROUD 2021). En particulier, notons que le défilement vertical est devenu un standard aussi connu que le feuilletage de pages, que l'information est aujourd'hui très souvent consultée numériquement, que l'écrit, s'il est toujours très présent, privilégie les formats plus courts que les

livres (de type article ou même brève), et que l'image est mise en avant sur ces interfaces visuelles.

L'écran, qui se nourrit de la culture humaine tout en modifiant celle-ci, est devenu une « interface culturelle »(GIROUD 2021. Il permet notamment de donner à voir des images numériques dont l'outil informatique contraint l'apparence, et dont l'interface conditionne le format et les usages.

La place de l'image dans les sociétés de l'écran

L'image occupe une position centrale dans les sociétés modernes. En effet, de nos jours, de très nombreuses images nous entourent au quotidien, puisqu'elles peuvent être, comme nous l'avons mentionné précédemment, placardées dans les rues sous forme d'affiches, imprimées sur divers types de papier (en particulier les photographies), croisées au détour d'une page de livre ou de journal, affichées sur un mur en tableau ou tapisserie décorative... De plus, des images sont accessibles à tout moment sur les écrans qui envahissent notre environnement : les photographies personnelles des téléphones portables laissent la place aux images humoristiques vues sur le web ou illustrations d'articles de presse en ligne... Elles peuvent même être animées à la télévision, au cinéma ou sur internet. Comme le traduit FACCIOLI (2007) à partir d'un texte de GITLIN datant de 2003 :

« Les écrans (...) en une seule minute sont aptes à montrer plus d'images que celles qu'aurait pu contenir en plusieurs vies une riche maison flamande du XVII^e siècle. »¹⁰

Cette situation est liée aux facilités actuelles à produire des images en grandes quantités offertes par les nouvelles technologies développées depuis la fin du XX^e siècle (FACCIOLI 2007). En particulier, l'outil informatique a démocratisé cet usage de l'image par le plus grand nombre. Celle-ci devient même bien souvent le « canal privilégié de communication », selon FACCIOLI (2007). Par exemple, il est de notoriété publique que poster une image dans un tweet favorise l'intérêt que les autres utilisateurs y accordent, puisque celui-ci reçoit 18 % de clics supplémentaires et 150 % du nombre de retweets qu'il aurait obtenu sans. On peut voir ces statistiques sur l'histogramme présenté en figure 1.3, réalisé avec des données collectées par l'entreprise conseil en marketing digital Buffer¹¹. Une information est donc beaucoup plus relayée sur les médias sociaux si elle associe une image au texte.

Les stratégies de communications digitales illustrent bien le constat de FACCIOLI (2007) :

« Alors que dans le passé il existait des modalités et des lieux de représentation visuelle socialement définis et alors que les sujets étaient des destinataires passifs [FACCIOLI pense notamment aux musées, où l'on pouvait voir des tableaux], aujourd'hui les observateurs s'approprient les images, les décontextualisent et les utilisent pour communiquer. »

Si les images sont utilisées dans un but de communication auprès du grand public, c'est en particulier parce qu'elles ont un impact sur ceux qui

10. Todd GITLIN. *Sommersi dai media*. it. Milan : Etas, 2003.

11. Pour connaître cette entreprise, rendez-vous sur : <https://buffer.com/about>.

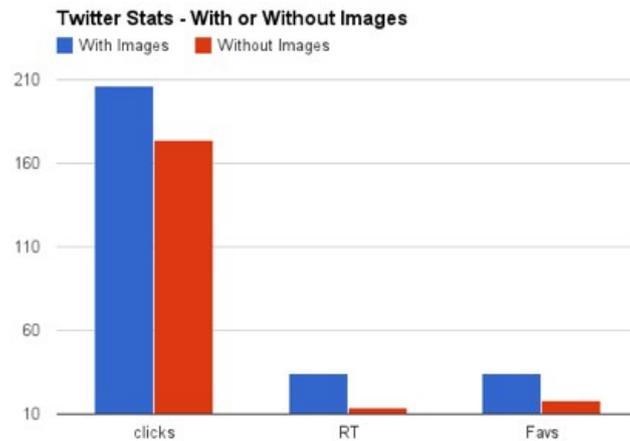


FIGURE 1.3 – Comparaison du nombre de clics, retweets et mise en favoris d'un tweet avec image et d'un autre sans image.

Datavisualisation réalisée par Buffer, à partir d'une moyenne sur 100 tweets.

Issue de <https://buffer.com/resources/>

[the-power-of-twitters-new-expanded-images-and-how-to-make-the-most-of-it/](https://buffer.com/resources/the-power-of-twitters-new-expanded-images-and-how-to-make-the-most-of-it/)

les regardent (FACCIOLI 2007). En effet, elles sont par nature polysémiques, comme l'a énoncé BARTHES. Chacun peut les interpréter à l'aune des expériences qu'il a vécu. Comme le dit FACCIOLI (2007) :

« Ce qui nous frappe (nous pique) [en regardant une image,] le punctum, comme l'a défini Roland BARTHES, c'est la piqûre, le petit trou, la petite coupure, c'est un détail qui creuse dans notre mémoire, dans notre passé remué, dans nos blessures et dans les choses qui sont importantes pour nous, et qui conduisent à découvrir nos émotions. »

Ainsi, regarder une image est un processus subjectif qui fait appel aux émotions. Ce phénomène est donc tout à fait approprié à la culture actuelle de consommation de l'information du grand public. En effet, nous sommes entrés dans une « économie de l'attention », c'est-à-dire que la plupart des gens, submergés par la masse d'informations qui les entourent, n'accordent que quelques instants à chaque contenu (CITTON 2014). Il faut donc, pour les créateurs desdits contenus, capter l'attention de l'utilisateur, qui s'arrêtera d'autant plus longtemps sur une page, un article, ou tout autre document, que celui-ci fera appel à ses émotions, et lui présentera une information rapidement décryptée. Nous sommes passés « d'une culture du discours, textuel et diachronique, à une culture synchronique de l'image à fort impact immédiat » (FACCIOLI 2007).

Les images, aujourd'hui omniprésentes dans notre vie, façonnent nécessairement notre expérience sensible du monde. La vision occupe une dimension centrale dans notre existence, et une culture visuelle s'instaure pour chacun. Celle-ci permet d'attribuer des significations aux éléments observés

(FACCIOLI 2007). L'étude de la réception des images nous permettra de comprendre l'impact que ce phénomène peut avoir sur les individus.

1.2.2 La réception des images

Lorsque l'on est spectateur d'une image, on ne fait pas qu'admirer les formes et couleurs que l'on a sous les yeux et que notre vue perçoit. On en retire également des émotions et un sens.

La perception

C'est la perception qui permet, en première instance, de voir l'image et de recevoir son message.

Bien que ce mot fasse partie du langage courant, le concept de « perception » est extrêmement complexe à définir. En effet, il fait appel à de nombreuses disciplines : si les sciences cognitives s'y intéressent particulièrement pour mettre en exergue les sensations et leur lien avec les pensées et actions, la psychologie se le ré-approprie aussi pour expliquer des observations, et la sociologie y fait parfois incursion. Les sensations elles-mêmes sont le domaine de la biologie ou de la physique. Enfin, en SIC, la perception joue un rôle pratique pour déterminer les modes de communication les plus efficaces en fonction de la réception d'un public cible.

Nous essayerons malgré tout de donner ici une définition générale de ce terme, en nous appuyant sur les commentaires de BARBARAS sur les cours de SIMONDON, philosophe spécialiste des questions d'information, ainsi que sur les écrits de VIAL, chercheur en sciences du design.

Voici la définition qu'en donne le premier :

La perception est « la modalité originaire de la relation d'un vivant à son milieu, la forme sous laquelle il entre d'abord en contact avec lui. »

BARBARAS 2013

Cette expression reste extrêmement large, et ne nous donne que peu de précisions. Nous pouvons néanmoins en retirer l'information suivante : la perception correspond au lien premier d'un être vivant avec son environnement. En effet, ne dit-on pas que l'on « perçoit » le monde qui nous entoure ?

Cela comporte deux dimensions. D'une part, il s'agit de ressentir : la perception correspond à « une fonction du corps » (VIAL 2013), une activité « biologique » (BARBARAS 2013) qui regroupe toutes les sensations d'un être vivant. Cela englobe tous les stimuli que l'on reçoit de son environnement, qu'ils soient olfactifs, auditifs, gustatifs, kinesthésiques ou visuels.

D'autre part, ces sensations entraînent des réactions neuronales, jusqu'au cerveau, où la perception devient « une fonction de la conscience » (VIAL 2013) à « portée informative » (BARBARAS 2013), c'est-à-dire qu'elle permet à l'être vivant qui ressent son environnement de se le représenter. Il ne s'agit pas seulement de faire appel aux sens pour « explor[er] » son milieu, mais également de « [l']élabor[er] », donc de se le représenter mentalement (BARBARAS 2013).

Ainsi, percevoir une image fait appel à la vue, qui permet de repérer des formes et couleurs, mais consiste également à identifier le contenu de l'image observée. Elle permet donc de recevoir le message que propose toute représentation iconique en première instance.

Cette approche initiale de l'image est donc personnelle, car liée au sens et à l'activité cérébrale de chaque individu. Cependant, ce dernier baigne dans une culture qui oriente sa perception. En effet, l'être humain acquiert des « habitudes ou aptitudes (...) en tant que membre d'une société » (VIAL 2013). Aussi VIAL élabore d'ailleurs le concept de « culture perceptive » : selon lui la perception « est aussi une fonction sociale — au sens où elle est conditionnée par des facteurs culturels ». Cette idée semble confirmée par des sociologues tels que BOURDIEU et DELSAUT, qui soulignent la différence entre la façon dont les tableaux de la Renaissance italienne étaient perçus par leurs contemporains et les analyses que l'on peut en faire aujourd'hui, avec une culture différente (BOURDIEU et DELSAUT 1981).

Notons également que BARBARAS décrit la perception comme un processus actif, et non contemplatif. En effet, nous pouvons ressentir des objets inanimés, mais surtout des événements, des éléments mouvants. Ainsi, nous pouvons percevoir « de la distance spatiale [possiblement évolutive] et (...) un écoulement temporel », qui s'inscrivent dans une situation non figée (BARBARAS 2013). Cette caractéristique nous intéresse puisqu'elle pourra s'appliquer à l'image animée ou interactive.

Le récepteur d'une iconographie commence donc par percevoir l'image qu'il a sous les yeux. Il comprend ensuite sa signification, ou plutôt un de ses multiples sens.

La polysémie

Les images portent toujours un sens en elles-mêmes. Selon BARTHES, elles en ont même plusieurs. Il explique que « toute image est polysémique », c'est-à-dire qu'elle peut être interprétée de plusieurs manières différentes. En effet, un ensemble iconique regroupe deux niveaux de sens :

- L'image représente objectivement des éléments, que notre œil perçoit (c'est-à-dire qu'il voit et identifie) : c'est « l'image dénotée » ;
- Elle offre également un message symbolique, qualifié de « connoté », qui provient de l'interprétation des signes que l'on voit objectivement dans l'image dénotée. On appelle ces signes des « signifiants », et leur interprétation le « signifié » (BARTHES 1964).

De plus, le message connoté lui-même peut recevoir plusieurs significations, dépendantes de l'observateur. En effet, chaque signifiant peut être associé à plusieurs signifiés en fonction du spectateur, de « [s]es états d'âme, [s]es vécus, [s]es expériences passées, [s]a mémoire, [etc.] » (FACCIOLI 2007). Les signifiés sont donc subjectifs, puisqu'ils dépendent de la personne qui les observe. Comme le résume FACCIOLI (2007) :

« Les usagers amènent leurs façons de regarder et, de cette manière, les significations sont renégociées. »

Bien que l'interprétation d'une image soit propre à chacun, sa connotation fait appel à des stéréotypes culturels (BARTHES 1964). Ainsi, des spectateurs

de même culture sont orientés, tout d'abord, dans leur façon de percevoir; puis ils attribuent des significations proches à une même iconographie. Citons FACCIOLI (2007) qui nous rappelle que :

« La vision d'une image a toujours lieu dans des contextes sociaux particuliers qui en conditionnent l'impact. »

Un dernier élément doit être pris en considération pour expliquer les sens pris par une image : le contexte de son observation. En effet, en fonction du lieu de son apparition, un effet d'autorité peut influencer le regard que l'on porte sur elle. Par exemple, voir une image dans un musée lui apporte une dimension qu'elle n'aurait pas si on la voyait dans son salon (FACCIOLI 2007).

L'image, qui est de plus en plus présente dans notre environnement quotidien, est donc perçue, puis prend une signification aux yeux du spectateur. Nous intéressant à la datavisualisation, il nous faut à présent nous pencher sur la place des visualisations au sein de la catégorie des images, et nous interroger sur les spécificités de leur réception.

1.2.3 La visualisation

FACCIOLI (2007) définit le concept de visualisation : selon elle, il s'agit de « rendre visible, montrer, produire une image » à partir d'une réalité (ou de ce que l'on imagine être une réalité). Ainsi, nous assimilerons toute image qui cherche à rendre compte d'un phénomène à une visualisation. Par exemple, un schéma qui essaie de montrer des relations ou des processus de façon iconographique est en une. De même, une carte qui cherche à rendre visible un territoire et ses dynamiques est une visualisation. Même une simulation informatique représentant la ville du futur serait une visualisation puisqu'elle représente ce que l'auteur pense être la réalité de demain. À l'inverse, nous excluons ici de ce concept les images purement artistiques qui ne cherchent pas à rendre compte d'un phénomène, les images associées à des œuvres fictionnelles telles que les illustrations de romans, d'albums ou de bandes dessinées, mais aussi les photographies (si elles ne sont ni modifiées ni insérées à un système iconographique d'une autre nature). En effet, les photographies sont des témoignages d'un instant, mais ne rendent pas perceptible un événement dans son ensemble.

On parlera ici de visualisation exclusivement pour désigner des images créées (soit par l'être humain, soit par un outil informatique) dans le but de rendre visible un phénomène¹². Notons que, selon FACCIOLI, les visualisations sont liées à une « construction sociale », ce qui revient à dire que la façon de visualiser des événements ou des objets est culturelle (FACCIOLI 2007).

Par ailleurs, cette dernière insiste sur la tendance actuelle à « visualiser des choses non visuelles à l'aide de supports technologiques ». Autrement dit, nous essayons aujourd'hui de rendre visibles des éléments qui ne le sont pas. Par exemple, lors d'une échographie, on voit sous ses yeux une image en noir

12. Par phénomène, nous entendons un ensemble de faits appartenant à la réalité sensible du monde et qui peuvent être théorisés. Il s'agit ici d'un concept en lien avec son étymologie grecque : étudier un phénomène (terme provenant de *φαινόμενον* : ce qui apparaît) consiste à rechercher la cause d'événements pour comprendre ceux-ci.

et blanc qui représente des densités de matières à l'intérieur d'un être vivant. Ce qui s'affiche sur l'échographe montre une réalité qui n'est pas accessible aux sens de l'être humain (pas même lors d'une dissection qui donnerait à voir des éléments en reliefs, et non des contrastes de densité des tissus). FACCIOLI souligne donc cette « tendance à visualiser l'existence : tout est visualisé, des modèles informatiques à la médecine, aux interfaces graphiques et aux téléphones, et jusqu'à la vie privée » (FACCIOLI 2007).

Une autre préoccupation de l'auteure est la tendance à créer des visualisations qui ne représentent même plus une réalité, mais des simulations de ce qui est, a été, ou pourrait être. Elle écrit :

« La technologie rend visible une réalité qui est seulement simulation, mais que nous pouvons vivre comme réelle. »

FACCIOLI 2007

On retrouve ici la notion de perception : le récepteur d'une visualisation la « vi[t] comme réelle » puisqu'il l'a sous les yeux, la voit et la perçoit. Même s'il ne s'agit que d'une simulation, il a l'impression que celle-ci a une certaine réalité, puisqu'il peut la ressentir. C'est pourquoi l'on peut dire que créer une visualisation « équivaut (...) à donner une définition de la réalité, (...) à produire une vision du monde » (FACCIOLI 2007).

Cela peut inciter à penser que celui qui réalise une visualisation détient une forme de pouvoir sur son récepteur, puisqu'il exprime sa propre vision d'un phénomène. FACCIOLI (2007) évoque d'ailleurs le « pouvoir de visualiser, et donc de véhiculer des significations ».

Les médias de masse, en particulier, emploient de très nombreuses visualisations, que ce soit à la télévision, dans les journaux papier ou en ligne. Si leur intention, est louable (rendre perceptibles pour leurs lecteurs et/ou spectateurs un phénomène souvent complexe), ils contribuent ce faisant à construire une réalité qui n'est pas accessible par l'expérience de vie quotidienne de leurs usagers. Ceux-ci peuvent donc soit adhérer à une vision des événements et, plus généralement, de leur monde véhiculée par le média, soit la rejeter. Nous savons d'ailleurs que ce phénomène de rejet des médias traditionnels est de plus en plus fréquent pour des individus qui, bien souvent, s'oppose au système dans son ensemble, et donc à toute instance qui lui donnerait une vision du monde qui n'est pas en adéquation avec la sienne. Pour expliciter cette idée, citons FACCIOLI (2007) :

« De nombreuses études récentes portant sur les effets des médias convergent vers la notion de "réalité médiatique", c'est-à-dire la production d'images du monde qui entrent dans les processus subjectifs de construction de la réalité sociale. »

Résumé — La notion de visualisation

Nous avons donc défini la visualisation comme une image créée (soit par l'être humain, soit par un outil informatique) pour rendre visible un phénomène. Cette définition a nécessité de se pencher sur le concept d'image, que nous avons associé à une représentation iconique à deux dimensions,

plus ou moins réaliste, d'objets ou d'événements, constituée d'un ensemble d'icônes, de signes et/ou de formes, assemblés sans suivre l'ordre du langage (par opposition à un texte écrit, qui suit l'ordre du langage oral).

Par ailleurs, nous avons pu noter la multiplication des images dans notre environnement quotidien (bien que celles-ci ne soient pas toujours des visualisations). En effet, la communication digitale, *via* l'interface que constitue l'écran, s'appuie très souvent sur des représentations visuelles afin de capter l'attention de ses récepteurs. Les visualisations sont d'ailleurs aujourd'hui très employées par les médias grand public.

Ce constat nous a obligé à interroger la réception des images. Nous avons établi que l'observateur commence par percevoir l'image, c'est-à-dire la voir et se la représenter mentalement, puis qu'il lui attribue un ou plusieurs sens parmi ceux que la représentation iconique peuvent prendre.

C'est là un point qu'il nous reste à éclaircir : si une image est, par nature, polysémique, comme l'affirme BARTHES, comment une visualisation peut-elle donner une seule et unique vision d'un phénomène ? Car son but est bien de montrer des phénomènes de façon explicite et univoque. Pour répondre à cette question, nous allons nous pencher sur les processus mis en place par la datavisualisation pour contourner la polysémie propre aux iconographies.

CHAPITRE 2

LE CONCEPT DE DATAVISUALISATION

La datavisualisation couple les deux concepts que nous avons défini précédemment : il s'agit de la visualisation de données. Selon nous, les datavisualisations seraient donc des images (c'est-à-dire de représentations en deux dimensions constituées d'un ensemble d'icônes, de signes et/ou de formes ¹) créées soit par l'être humain, soit par un outil informatique, à partir de données (valeurs codées de façon digitale ou analogique et enregistrées sur un support, qualifiant ou quantifiant un fait, un événement ou un objet décontextualisé ²). Ce procédé permet d'exploiter des données, de les mettre en relation graphiquement, et ainsi d'en faire émerger des informations afin de rendre visible un phénomène.

Afin de valider ou d'apporter des précisions à cette définition, nous allons nous appuyer sur quelques auteurs qui ont proposé un certain nombre de caractéristiques définissant, selon eux, les datavisualisations. Ensuite, nous étudierons l'évolution des représentations graphiques de données dans le monde occidental, afin de comprendre comment ce concept a émergé, et ce qui le constitue intrinsèquement. À la lumière de ces éclaircissements, nous pourrions ré-élaborer notre définition du concept. Enfin, nous étudierons les deux courants de pensées qui, selon nous, contribuent aux datavisualisations que nous trouvons actuellement dans les médias français : l'approche sémiologique de BERTIN, et celle qui relève du design centré utilisateur de TUFTE.

2.1 Les caractéristiques de la datavisualisation

La plupart des définitions de la datavisualisation que l'on trouve en ligne correspondent approximativement à celle que propose Éduscol. Le « Portail du numérique » de cette institution ministérielle définit la datavisualisation comme « l'utilisation de représentations visuelles interactives et informatisées de données pour simplifier la connaissance. Elle permet en effet de faciliter la lecture de données en les présentant sous formes d'images, de graphiques, de pictogrammes, de cartes... ». Et d'ajouter :

1. Voir la partie 1.2
2. Voir la partie 1.1

« Elle rend explicite les informations complexes ou noyées dans une grande quantité de paramètres. (...) Son objectif est d'éclaircir un ou des phénomènes en lui donnant un aspect plus ergonomique qu'un tableur rempli de chiffres. »

Certains éléments sont communs avec la définition que nous avons établi précédemment : l'exploitation de données sous une forme visuelle, dans le but de montrer des informations et d'expliquer des phénomènes. En revanche, plusieurs notions ont ici été ajoutées : l'ergonomie, le recours obligatoire à l'outil informatique et l'interactivité. De plus, elle évoque plusieurs forme que peut prendre la datavisualisation selon elle : « images, graphiques, pictogrammes, cartes... »

Reprenons ces caractéristiques une par une en nous référant à des experts français du domaine, qui se sont réunis pour décrire tous les aspects de la datavisualisation dans un numéro de la revue *I2D* en 2015.

2.1.1 La matière première : les données

Tous les auteurs consultés s'accordent pour dire que les données sont la matière première qui constitue les datavisualisations. Certains montrent que tout le processus de conception d'une représentation graphique tourne autour de ce concept, à l'image de HACHOUR qui cite toutes les étapes de collecte et de transformation des données (nettoyage) en vue de leur représentation visuelle. Il dit : « la visualisation de données se fonde sur une analyse préalable, qualitative et/ou quantitative, de données brutes ou structurées (*datamining*) » (HACHOUR 2015).

BRASSEUR (2015), quant à lui, fait explicitement le lien entre l'augmentation considérable du nombre de données aujourd'hui (le Big data) et l'engouement pour les datavisualisations aujourd'hui. ZAZA (2015) met en évidence le mouvement de l'Open data avec d'un portail de réutilisation visuelle de données mises à disposition par la ville de Paris. Elle s'interroge d'ailleurs sur la possibilité que ce mouvement conduise à des exploitations de données aidant à la « prise de décision [du] grand public ».

La question des données est donc centrale pour la datavisualisation. DRUCKER (2011) remet en cause le terme même de « donnée » : elle revendique l'utilisation du mot « capta », car celui-ci implique un processus actif de récolte et donc une construction souffrant de biais cognitifs, contrairement à une valeur « donnée » et objective (DRUCKER 2011).

2.1.2 La forme : une représentation visuelle

La datavisualisation « traduit des données sous une forme visuelle pertinente » (BRASSEUR 2015). Elle emploie un système de signes visuels pour exprimer des données. Selon REYES, elle « combine régulièrement des éléments graphiques (lignes, points, couleurs) organisés dans une forme (...) et s'enrichit d'éléments aidant sa communication et sa compréhension (textes, légendes, images, figures) ». REYES propose une classification des datavisualisations en « quatre modèles visuels typiques, à savoir les graphes cartésiens, les diagrammes de réseaux, la cartographie et les modèles expérimentaux » (REYES 2015). Cette classification reprend celle que

BERTIN³ en lui ajoutant les « modèles expérimentaux » qui correspondent aux formes les plus récentes. Ces derniers sont qualifiés de « visualisations avancées » par BRASSEUR, qui estime que leurs « courbes, formes et jeux de couleurs » rendent ces représentations « esthétiques » (BRASSEUR 2015).

Ainsi, bien que certains auteurs (tel MANOVICH) classe la cartographie en dehors du champ de la datavisualisation, la réalisation de cartes statistiques, qui positionnent des données sur une représentation schématique d'un territoire, est « l'une des disciplines fondatrices de la visualisation de données » (HACHOUR 2015). C'est cette discipline qui a posé les bases de la « transformation sémiotique » consistant à représenter « dans un ensemble de signes [une] réalité ». D'ailleurs, comme nous le verrons bientôt, l'un des théoriciens des graphiques de données les plus reconnus en France (textscBertin) était cartographe. Aussi nous admettons que les cartes statistiques font partie des datavisualisations, au même titre que « histogrammes, courbes, secteurs, camemberts, aires, bulles, etc., » (BRASSEUR 2015), c'est-à-dire toute forme de représentation visuelle de jeux de données.

2.1.3 L'objectif : mettre en évidence et transmettre des informations

Comme nous l'avons expliqué précédemment, le croisement de données permet de faire émerger des informations. L'objectif des représentations graphiques de données est alors de mettre en valeur ces informations. Citons quelques uns des auteurs consultés, qui corroborent cette idée :

« Les données sont devenues le moyen privilégié d'encapsuler l'information toujours plus complexe produite par nos sociétés. »

GOULARD 2015

Ou encore :

« Dans un contexte de génération massive de données, la datavisualisation devient un moyen privilégié d'accès à l'information. (...) Elle rend les mégadonnées (big data) commensurables à nos moyens perceptifs et cognitifs. »

AMATO 2015a

L'objectif premier de toute datavisualisation est donc de montrer des informations visuellement. Certains auteurs évoquent également l'importance de présenter l'information de façon « efficace [pour être] décodée le plus rapidement possible » (BRASSEUR 2015). La capacité à résumer une information en un visuel simple et vite décriptable est le point fort de la datavisualisation, ce qui en fait un « outil heuristique » puissant (AMATO 2015a). De plus, il est parfois noté que les datavisualisations peuvent « faciliter l'apprentissage d'éléments de connaissance sur un objet complexe » (HACHOUR 2015). Leur capacité à résumer visuellement et à faire apparaître des tendances pourrait en effet, selon certains, favoriser la mémorisation. Il s'agit probablement de l'un des objectifs de la datavisualisation, quoiqu'il ne soit pas toujours vérifié dans la pratique.

3. Nous retrouverons les recommandations de BERTIN dans la fin de cette partie.

Enfin, l'information portée par une représentation graphique de données est souvent destinée à être communiquée. En effet, selon FREDRIKSSON (2015), « la visualisation de données désigne tous types de représentations visuelles qui facilitent l'exploitation, l'analyse et la communication de données ». L'un des buts d'une datavisualisation est donc bien la communication des informations qu'elle met en évidence. C'est également l'opinion de BOY, qui parle même de « nouveau médium de communication » (FEKETE 2015).

Cet acte de communication a donc un récepteur. Celui-ci peut varier selon le contexte. Par exemple, de nombreux articles mentionnent la *Business Intelligence* qui consiste à présenter graphiquement des données métier aux membres d'une entreprise, souvent aux personnes en charge des décisions stratégiques. En effet, les datavisualisations sont depuis toujours (et demeurent) des outils d'aide à la décision, comme le souligne HACHOUR (2015) :

« Les représentations graphiques font sens en tant que messages participant à la décision et à l'action. »

BRASSEUR (2015), lui aussi, considère les datavisualisations comme un outil essentiel à la prise de décision des personnes dotées d'autorité :

« Ces graphiques permettent aux décideurs d'examiner rapidement les données et de réduire le temps de compréhension des informations essentielles. »

Pour d'autres, il s'agit d'apporter des informations au « grand public » (ZAZA 2015). En particulier, FREDRIKSSON et GOULARD pensent que la datavisualisation doit viser la « transmission [du] sens auprès de la société civile » (FREDRIKSSON 2015), afin de permettre aux individus de se réappropriier les données, de les comprendre malgré leur quantité, d'offrir l'opportunité de les manipuler, et d'exercer un droit de regard sur elles (GOULARD 2015). Les représentations graphiques de données doivent, selon ces deux auteurs, favoriser la « compréhension des données et de leurs processus de traitement ».

« En traduisant visuellement les données, [la datavisualisation] s'appuie sur la capacité de notre cerveau à s'approprier des quantités d'information bien plus importantes lorsque ces dernières lui sont présentées visuellement, grâce à la spatialisation de l'information et aux métaphores visuelles. »

GOULARD 2015

Il s'agit ainsi de faire de la datavisualisation un outil d'aide à la compréhension et à la décision citoyenne. FREDRIKSSON (2015) évoque cela explicitement :

« La visualisation de données [vise à] aider à la démocratisation de l'accès à une connaissance structurée issue de la complexité. »

2.1.4 Le design pour favoriser l'ergonomie et la perception

Comme nous l'avons dit, la datavisualisation est un outil heuristique qui facilite la perception d'informations grâce à une « organisation visuelle

de[s] données » qui « en révé[le] la structure » (FREDRIKSSON 2015). Si l'acte même de représenter graphiquement des informations constitue une aide à la cognition, certains professionnels estiment cela insuffisant, et pensent nécessaire de se tourner vers le design pour comprendre les besoins des utilisateurs.

Pour comprendre cette pratique, nous référerons à VIAL 2014 (2014). Celui-ci nous explique que le design est apparu aux États-Unis dans les années 1920, dans le domaine industriel. Son objectif est alors de favoriser la vente de produits de consommation « en leur donnant des formes aérodynamiques, lisses et arrondies » (VIAL 2014). En une trentaine d'années, ce nouveau concept donne de si bons résultats sur le plan commercial que les entreprises associent les services de design et de marketing, et en font ainsi une discipline communicationnelle.

« En effet, transformer les objets en signifiants consommationnels grâce au design consiste à en faire des canaux de communication qui transmettent un message (commercial, idéologique) entre l'émetteur (la marque) et le récepteur (le consommateur). »

VIAL 2014

Créer un design est donc avant tout un acte de communication (tout comme les datavisualisations). Cela est particulièrement vrai concernant le design d'information, qui fait intervenir toutes les modalités de communication moderne : le visuel avec le design graphique, les nouveaux outils informatiques avec le design interactif. Ce dernier « envisage l'interaction entre une interface numérique et un usager en termes de "synergie communicationnelle" » (VIAL 2014), c'est-à-dire que l'utilisateur doit en comprendre naturellement les fonctionnalités, tandis que l'interface évolue pour s'adapter à l'utilisateur. Cette idée s'inscrit dans le mouvement du design « centré-utilisateur », très en vogue depuis les années 1980 (VIAL 2014).

« Le design centré-utilisateur permet à un artefact bien conçu de communiquer de lui-même et par lui-même avec l'utilisateur, sans qu'il y ait besoin d'instructions d'utilisation. »

VIAL 2014

Il s'agit donc d'une approche « principalement ergonomique et cognitive » (VIAL 2014) : l'utilisation de l'objet du design doit être intuitive pour l'individu, qui n'a alors pas besoin d'explication pour l'utiliser.

C'est cette forme de design « centré[e] utilisateurs » (RINATO et al. 2015) qui est de plus en plus souvent utilisée par les datascientists, qui cherchent la meilleure façon de présenter leurs données et de permettre l'interaction avec elles afin de proposer une expérience intuitive et ergonomique au récepteur. Ce dernier, ne se heurtant pas aux difficultés de la prise en main d'une interface et à la lecture fastidieuse d'une légende, est alors sensé comprendre facilement ce qui lui est présenté puisqu'il s'agit alors d'un outil cognitif optimisé pour la compréhension. Comme le mentionne FREDRIKSSON, le design d'information doit mettre en valeur le sens du message porté par la visualisation. Elle décrit le « design d'information » comme un « équilibre entre des enjeux relevant de l'exactitude scientifique, de l'esthétique, de l'efficacité cognitive et de l'accessibilité, déterminés par un contexte » (FREDRIKSSON 2015). Cela

implique bien les principes du design interviennent dans le processus de conception d'une datavisualisation pour la rendre ergonomique et facilement accessible à la pensée des récepteurs. C'est également la position défendue par trois designers, RINATO et al. (2015), qui disent « [s']interrog[er] sur la manière de faire le lien entre l'intention et l'utilisateur », donc chercher la meilleure façon de représenter graphiquement les données pour que le récepteur puisse les découvrir, voire les explorer, aisément. Ils pensent que leur rôle en tant que créateurs de datavisualisations est de « défricher le chemin » et « d'extraire un sens lisible et simple à de comprendre » prêt à être consommé par l'utilisateur (RINATO et al. 2015).

2.1.5 L'outil informatique

L'informatique est maintenant considéré par beaucoup comme le mode de création privilégié des datavisualisations. En effet, beaucoup pensent que le traitement de données d'une part, et leur représentation graphique d'autre part, se sont « étendu[s] et perfectionné[s], notamment grâce à l'informatisation qui permet de traiter des quantités de données de plus en plus massives » (HACHOUR 2015). Cela est effectivement une réalité : les capacités de calcul des ordinateurs et leurs méthodes itératives permettent de travailler sur de grands volumes de données, qu'il aurait été difficile de traiter à la main. De plus, les interfaces graphiques s'étant perfectionnées au point que l'on ai aujourd'hui de nombreux logiciels WYSIWYG⁴ permettent de créer des visuels correspondant exactement aux souhaits des concepteurs. C'est ce qu'évoque CASILLI et ARRUABARRENA (2015) en disant que « la visualisation [est une] pratique [qui] consiste à se familiariser et à expérimenter avec des outils logiciels, des codes esthétiques ».

Cependant, peut-on pour autant écarter du champ de la datavisualisation toutes les représentations graphiques réalisées à la main ? Comme MANOVICH 2011, dans une perspective historique, nous considèrerons ici que les graphiques de données et cartes statistiques produits pour tout autre outil que l'informatique reste une datavisualisation à condition qu'il réponde aux autres critères évoqués.

Par ailleurs, il nous semble réducteur de considérer que les datavisualisations ne puissent être reçues que *via* une interface écranique. Bien que certains privilégient nettement ce mode de réception, à l'instar de BRASSEUR qui parle de « publication des résultats sur Internet ou intranet », nous considèrerons que les représentations visuelles de données peuvent tout aussi bien être vues sur un écran que sur tout autre support, allant du papier journal jusqu'à la peinture murale.

2.1.6 Les fonctionnalités interactives

En lien avec le support informatique et l'expérience utilisateur offerte par le design, beaucoup de spécialistes actuels établissent une relation entre datavisualisation et interactivité. Il s'agit en effet d'une fonctionnalité nouvelle

4. WYSIWYG est l'acronyme de « what you see is what you get ». Les logiciels WYSIWYG permettent de visualiser sur l'écran le résultat que l'on pourra imprimer ou publier en ligne. Il s'oppose aux logiciels fonctionnant exclusivement par des lignes de commandes et nécessitant une compilation avant d'obtenir un aperçu.

permise par l’outil informatique, qui rendrait la représentation graphique qui en bénéficie encore plus performante cognitivement. Pour FREDRIKSSON 2015 (2015), l’interactivité « confér[e] à l’utilisateur un rôle actif au sein du processus, en lui permettant d’appréhender et de manipuler les données ». Aux yeux de GOULARD (2015), « en zoomant, filtrant, naviguant, basculant, déplaçant, masquant, etc., les utilisateurs sont plongés dans une posture active qui facilite leur appropriation de l’information ». Cette idée est également partagée par les chercheurs en informatique HASCOËT et BEAUDOUIN-LAFON :

« Pour être réellement efficace, la visualisation d’information doit être interactive. En effet, l’être humain est particulièrement habile à extraire des informations d’un environnement qu’il contrôle directement et activement par rapport à un environnement qu’il ne peut qu’observer de manière passive. Selon l’approche écologique de la perception due au psychologue J.J. GIBSON (1979), la perception est indissociable de l’action : il faut agir pour percevoir et il faut percevoir pour agir. On parle de couplage (ou boucle) action-perception. »

HASCOËT et BEAUDOUIN-LAFON 2001

C’est bien à l’amélioration de l’outil informatique que l’on doit cette opportunité nouvelle. En effet, HASCOËT et BEAUDOUIN-LAFON nous apprennent que « la visualisation interactive requiert des temps de réponse interactifs (de l’ordre du dixième de seconde) » afin d’avoir un affichage en temps réel. Cela n’est possible que depuis « l’avènement des postes de travail graphiques et l’augmentation de la puissance de calcul » (HASCOËT et BEAUDOUIN-LAFON 2001). De plus, aujourd’hui, les interfaces que sont les écrans peuvent même être tactiles, comme le mentionne GOULARD (2015), ce qui permet des mouvements plus naturels qu’avec une souris (qui constituerait un dispositif médiateur supplémentaire pour accéder à l’information).

Cependant, on peut se demander quelles actions peuvent être considérées comme des interactions. Cette question est légitime, étant donné que l’usage de « barres de défilement pour naviguer le contenu d’une fenêtre, ou [de] boîtes de dialogues pour spécifier des paramètres » consistent déjà en des interactions en l’être humain et la machine (HASCOËT et BEAUDOUIN-LAFON 2001). Nous ne nous intéresserons ici qu’aux interactions avec les données, et non avec l’ensemble de l’image affichée à l’écran.

Grâce aux fonctionnalités interactives, l’utilisateur n’est plus seulement observateur, mais peut manipuler les données sur la datavisualisation : en afficher certaines, en cacher d’autres, et se focaliser ainsi sur les informations qui l’intéresse. Selon GOULARD, cela autorise l’utilisateur à « découper l’information trop dense en fragments plus faciles à manipuler pour un humain » que chacun est libre d’explorer à sa guise. Il s’agirait d’un procédé « rédui[sant] la complexité au sein des données » (GOULARD 2015). D’après HASCOËT et BEAUDOUIN-LAFON, l’interactivité « exploi[t] les caractéristiques du système visuel humain pour faciliter la manipulation et l’interprétation de données informatiques variées », puisque « les travaux en perception visuelle ont montré que l’être humain a une perception d’abord globale (...) d’une scène, avant de porter son attention aux détails » HASCOËT et BEAUDOUIN-LAFON 2001. Aussi les effets de filtre, de zoom,

et les transformations sémantico-géométriques, qui permettent l’affichage de différents niveaux de détails, paraissent-ils particulièrement en adéquation avec les facultés de perception humaines.

Bien que l’interactivité soit mentionnée par de nombreux auteurs consultés, qui en soulignent d’ailleurs souvent les intérêts, la grande majorité s’accorde pour dire qu’il s’agit d’une fonctionnalité supplémentaire (non obligatoire pour définir une datavisualisation). De plus, à l’instar de l’utilisation de l’outil informatique, nous ne pouvons pas considérer l’interactivité comme une dimension *sine qua non* de la datavisualisation si l’on veut conserver une perspective historique. En effet, si les représentations graphiques de données intègrent aujourd’hui de plus en plus souvent des fonctionnalités interactives, elles ont également une histoire : les être humains n’ont pas attendu l’avènement de l’ordinateur pour se rendre compte du potentiel heuristique des représentations visuelles d’informations.

Ainsi, nous considérerons que les datavisualisation s’attachent à représenter visuellement des données afin de mettre en évidence, et parfois communiquer, des informations. Si tous soulignent l’importance de s’attacher à respecter la perception humaine, plusieurs experts français revendique leur utilisation du design centré utilisateur pour apporter plus encore d’ergonomie. Par ailleurs, nous avons mentionné que nous considérerions ici que la datavisualisation n’implique pas nécessairement le recours à l’informatique ni à l’interactivité, ce qui nous semblerait réducteur, en particulier quand les plus grand designers modernes (tel TUFTE) continuent à s’inspirer de représentations graphiques manuscrites des XVIII^e et XIX^e siècles.

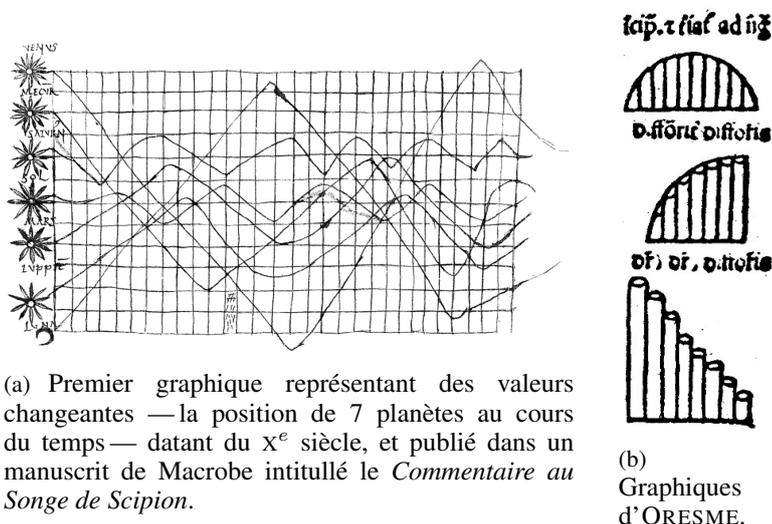
2.2 Les représentations graphiques de données à travers l’histoire

Par l’étude historique qui suit, nous souhaitons comprendre l’évolution des représentations graphiques de données au fil du temps et replacer la datavisualisation dans son contexte actuel. En effet, si cette pratique a une longue l’histoire, la situation actuelle associant un « déluge de données » et une multiplication d’images promet plus encore ce type de médiation.

2.2.1 Les débuts de la datavisualisation

Les toutes premières représentations graphiques du monde sont très anciennes : les premiers diagrammes représentant la position des astres et les cartes de voies navigables avec des systèmes de coordonnées remontent à l’Antiquité (FRIENDLY 2006). Selon BERTIN, « les plus anciennes représentations graphiques découvertes sont des cartes géographiques gravées sur argile, et qui datent vraisemblablement du 3^e millénaire avant J.-C » (BERTIN 1970). Cependant, ces représentations visuelles ne permettaient alors que l’affichage de positions « reprod[uisant] la nature visible », qu’elles ne modifient que par leur échelle (BERTIN 1970). Il faut attendre le Moyen-Âge « pour découvrir que les deux dimensions (...) pouvaient utilement représenter autre chose que l’espace visible » et ainsi arriver à créer des visualisations de données monosémiques grâce à un « “système de signes”, complet, indépendant, et possédant ses lois propres » (BERTIN 1970).

Ce n'est en effet qu'au X^e siècle de notre ère qu'un premier graphique représentant des données quantitatives dans un repère gradué représente l'inclinaison de sept corps célestes au cours du temps. Au XIV^e siècle, l'évêque ORESME a l'idée de représenter graphiquement des séries de valeurs contenues dans un tableau⁵ (FRIENDLY 2006).



(a) Premier graphique représentant des valeurs changeantes — la position de 7 planètes au cours du temps — datant du X^e siècle, et publié dans un manuscrit de Macrobe intitulé le *Commentaire au Songe de Scipion*.

(b) Graphiques d'ORESME.

FIGURE 2.1 – Premières datavisualisations européennes au Moyen-Âge
Source : FRIENDLY 2006.

Il s'agit cependant de cas isolés, les représentations graphiques de jeux de données étant loin d'être fréquentes au Moyen-Âge. Nous pouvons identifier plusieurs raisons techniques à cela. Tout d'abord, il y avait alors peu de données à représenter. S'il en existait d'ordre administratif, qui étaient consignées par écrit dans des registres, les mesures scientifiques n'avaient pas encore cours. En effet, la plupart des outils de mesures scientifiques précis n'avaient pas encore été inventés. Par ailleurs, ces visualisations manuscrites devaient être tracées sur un support. Or écrire était un luxe au Moyen-Âge : le papier n'arrivant en Europe qu'au XI^e siècle, et ne commençant à être répandu qu'au XIV^e, il fallait écrire sur des parchemins. Ceux-ci étaient extrêmement onéreux, et ne pouvaient en aucun cas être gâchés pour dessiner des graphiques qui auraient servi à comprendre soi-même un phénomène, avant d'envisager toute diffusion.

Au XVII^e siècle, des datavisualisations commencent à être tracées par les savants pour exploiter les données quantitatives issues de leurs observations, et ainsi construire des connaissances. À nouveau, des avancées technologiques en sont la cause : à l'époque de nouveaux outils de mesures et d'observation précis apparaissent, tels que la lunette astronomique. Dès lors, ils peuvent relever des données empiriques, comme l'a fait HALLEY en notant la pression atmosphérique en fonction de l'altitude en 1686 (à partir desquelles a été tracé le graphique présenté en figure 2.3(a)). De plus, les innovations en matière d'impression (la presse typographique existant déjà depuis plus d'un siècle) et de reproductions d'images (avec l'invention de la xylographie et

5. Voir les graphiques en figure 2.1

de la chalcographie) permettent de diffuser des ouvrages comprenant des illustrations, notamment sous forme de graphiques de données.

Les représentations visuelles deviennent un appui pour la pensée, comme en témoigne le schéma de l'astronome flamant VAN LANGREN réalisé en 1644 dans le but d'établir la distance réelle entre Rome et Tolède. Pour atteindre son objectif, l'astronome place sur un axe plusieurs mesures de cette distance évaluées précédemment par divers savants, et il estime que la distance réelle doit se trouver au milieu (FRIENDLY 2006). Il s'agit d'une façon intuitive d'envisager la notion de médiane grâce à des outils graphiques.

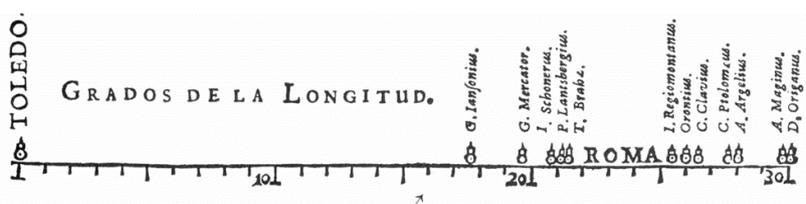


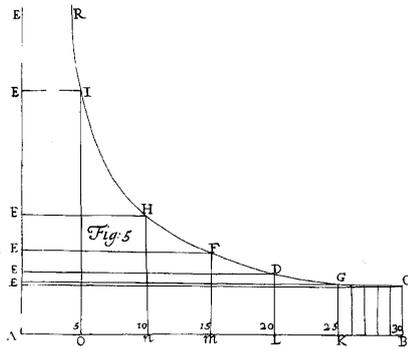
FIGURE 2.2 – Diagramme de VAN LANGREN

Source : http://www.fi.uu.nl/wiskrant/artikelen/hist_grafieken/begin/begin.html

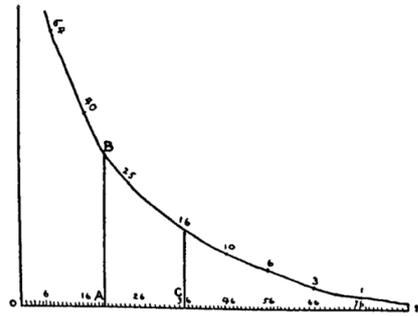
Les datavisualisations naissent donc pour se représenter le monde, donner à voir et à comprendre. Cette compréhension devient de plus en plus fine grâce aux développements majeurs qui touchent le champ des mathématiques. Ainsi, le XVII^e siècle se dote d'un arsenal théorique qui permet d'attribuer du sens aux visualisations de données. On commence alors à représenter graphiquement des relations mathématiques théoriques, telles que des équations, à l'initiative de FERMAT et DESCARTES (FRIENDLY 2006). De plus, les théories mathématiques des statistiques et des probabilités se formalisent grâce à PASCAL, FERMAT ou encore GAUNT. C'est ainsi que le premier graphique statistique est réalisé par HUYGENS en 1669. Il représente l'espérance de vie restant à un individu en fonction de l'âge de celui-ci. Il s'appuie sur des statistiques de l'âge de décès de la population de Londres établies par GAUNT (FRIENDLY 2006).

Les représentations visuelles de données deviennent durant ce siècle un outil pour la pensée des érudits, puisqu'elles permettent de comprendre des informations et d'établir des théories à partir de données issues d'observations et de mesures. Cependant, leurs destinataires demeurent peu nombreux : seul un cercle restreint, une élite lettrée et fortunée, peut accéder aux ouvrages qui représentent ces datavisualisations.

La plupart des formes de graphiques « classiques » que nous retrouvons encore aujourd'hui ont été élaborées au XVIII^e siècle. PLAYFAIR joue un rôle majeur dans ces inventions. Cet ingénieur écossais fait publier à Londres en 1786 l'*Atlas commercial et politique*, connu pour être le premier ouvrage majeur contenant des graphiques statistiques. L'objectif de PLAYFAIR est de représenter le plus simplement possible des informations précises et complexes (FRIENDLY 2006). S'il commence par réaliser des datavisualisations dans le seul but de « mettre de l'ordre dans ses propres idées », il réalise que le résultat va « au-delà de ce qu'il espérait » et permet de « rassembler dans un même



(a) Évolution de la pression atmosphérique en fonction de l'altitude, réalisée par Halley en 1686



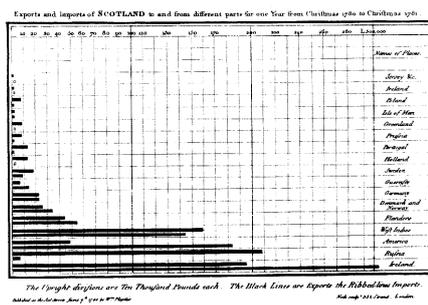
(b) Graphique de Huygens, 1669

FIGURE 2.3 – Datavisualisations du XVII^e siècle
Source : FRIENDLY 2006.

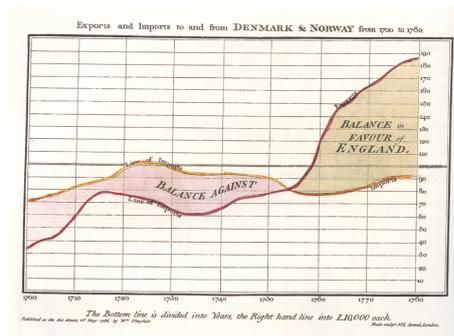
visuel des éléments dispersés (...) des faits ayant parfois un lien déjà établis entre eux, parfois non » et ainsi de faire apparaître des relations entre ces diverses données (PLAYFAIR 1805).

Partant de ce constat, il décide que ses datavisualisations doivent être publiées : l'*Atlas commercial et politique* contient 40 planches en couleurs montrant :

- Des courbes dans des repères cartésiens (qui existaient déjà et avaient été formalisées au siècle précédent) ;
- Des frises chronologiques (datavisualisations associant graphiquement des dates à des événements, inventées par PRIESTLEY en 1765) ;
- Un histogramme (le premier publié).



(a) Premier histogramme, présentant les imports/exports de l'Écosse en 1781



(b) Graphique montrant la balance économique entre le Royaume-Uni et le Danemark et la Norvège

FIGURE 2.4 – Deux graphiques réalisés par PLAYFAIR publiés en 1786 dans *The Commercial and Political Atlas*
Source : Wikipédia

Il imagine quinze ans plus tard des diagrammes circulaires, publiés en 1801

dans l'ouvrage *Éléments de statistiques*, à Londres ⁶ (FRIENDLY 2006).

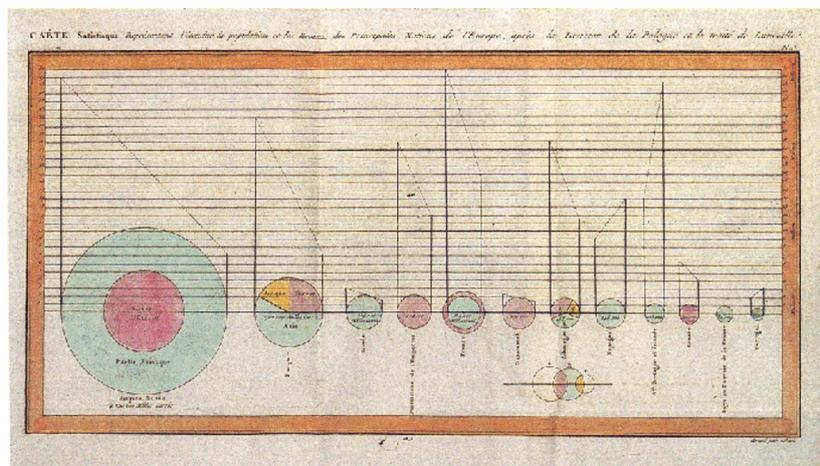


FIGURE 2.5 – Premiers diagrammes circulaires, réalisés par PLAYFAIR en 1801
Source : Wikipédia

Ces ouvrages ne sont pourtant pas encore destinés au grand public, mais adressés à une élite bourgeoise d'hommes d'affaire, qui prend de l'importance à cette époque. Les datavisualisations sont alors des outils d'aide à la décision, en présentant des informations stratégiques pour éclairer les choix de ces capitaines d'industrie (FRIENDLY 2006).

2.2.2 L'essor des datavisualisations au XIX^e siècle

Au XIX^e siècle, les datavisualisations prennent un essor significatif, en particulier grâce aux techniques d'impression qui facilitent la reproduction et la diffusion d'images. La lithographie, inventée en 1786 et perfectionnée jusqu'à l'apparition de la chromolithographie en 1837, simplifie la chaîne du livre illustré puisque ni graveur ni coloristes ne sont nécessaires : on dessine directement sur la pierre sans avoir à la graver. Les graphiques peuvent alors être directement exécutés par l'auteur et être reproduites sur de très nombreux exemplaires. Les délais et les coûts de l'impression en couleurs diminuent (FRIENDLY et D. DENIS 2020). De plus, le XIX^e siècle promeut la statistique en tant que discipline scientifique. Récolter des données et les rendre compréhensibles afin d'éclairer certains phénomènes devient une pratique récurrente et reconnue. Des activités de recensement de données (en particulier démographiques) amène une augmentation de la quantité de données disponibles (RENDGEN 2019).

De nombreuses datavisualisations (courbes, histogrammes, digrammes circulaires, cartes statistiques) présentent alors les informations que l'on peut tirer de ces données, et permettent d'en retirer des connaissances. Par exemple, la carte du docteur SNOW (en 1854), qui applique les données relatives à une épidémie de choléra sur la carte Londres, a permis de trouver l'origine de cette maladie : un puits contaminé sur Broad Street.

6. Cet ouvrage est consultable en ligne via le site *Gallica* à l'adresse <https://gallica.bnf.fr/ark:/12148/bpt6k81477j/f1.item>

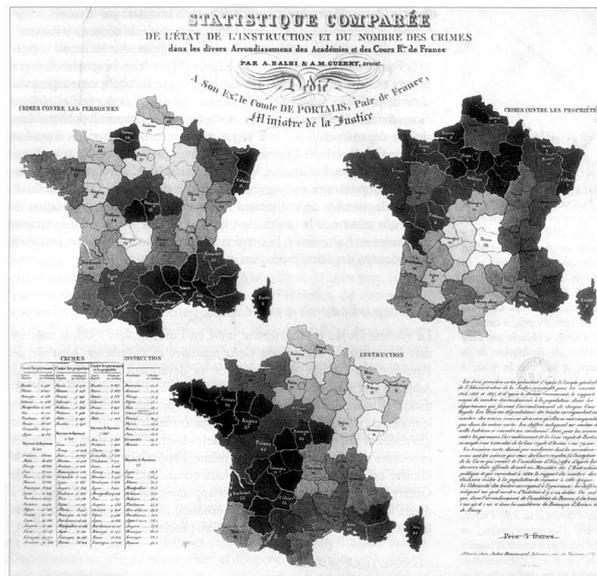


FIGURE 2.6 – Graphique de GUERRY comparant comparées l’instruction et le nombre de crimes en France
Source : Wikipédia

Ces visualisations expliquent donc toujours des phénomènes, mais qui sont désormais moins physiques que sociologiques. Parfois produites pour des instances politiques (comme la carte de corrélation entre criminalité et instruction dessinée par l’avocat GUERRY en 1825 à partir de données collectées par le Compte Général du Ministère de la Justice, présentée en figure 2.6), elles servent encore d’aide à la décision (FRIENDLY 2008).

Les datavisualisations ne sont donc désormais plus cantonnées aux publications scientifiques et aux atlas, elles ne sont plus publiées uniquement lors d’initiatives individuelles. Elles apparaissent dans des publications officielles pour le compte d’institutions reconnues. De plus, elles se démocratisent grâce à la presse (RENDGEN 2019). En effet, les journaux du XIX^e siècle, qui s’intéressent beaucoup à la politique, n’hésite pas à montrer des chiffres que les hommes politiques eux-mêmes utilisent. La quantité de graphiques et de cartes statistiques croît, et leurs formes se diversifient. Par exemple on voit l’apparition du stéréogramme en 1880 avec PEREZZO qui représente la population par âge et par année historique, selon trois axes. GALTON dessine des diagrammes de corrélations et met au point le système graphique qui devient la base de nos cartes météorologiques modernes. C’est aussi l’époque de MINARD, connu entre autres pour sa carte de flux figurant les pertes successives en hommes de l’armée française au cours de la campagne de Russie en 1812–1813, datant de 1869 (RENDGEN 2019).

Avec cette profusion de nouveauté et de publications, FRIENDLY estime que le XIX^e siècle correspond à un « âge d’enthousiasme », ou même un « âge d’or » pour les datavisualisations (FRIENDLY 2006).

des ordinateurs s'accompagnent d'améliorations dans le traitement de larges quantités de données statistiques (ou de tout autre jeu de données). À partir des années 1990, des environnements informatiques permettent même la génération de représentations graphiques de données interactives (FRIENDLY 2006).

2.2.4 Un contexte actuel propice à l'explosion du nombre de datavisualisations

Depuis une trentaine d'années, les outils informatiques permettant de créer des datavisualisations sont donc de plus en plus performants. Cela coïncide avec l'entrée des ordinateurs dans les foyers, ainsi qu'à la mise en réseau des informations *via* internet. Cette combinaison d'événements, qui permet à tout un chacun d'employer des outils puissants pour créer ses propres représentations graphiques de données et de les publier en ligne, a grandement contribué à la démocratisation des datavisualisations. Cela est d'autant plus vrai que, récemment, des logiciels extrêmement ergonomiques, tels que SAS ou Tableau public, sont disponibles aux novices en matière de datavisualisation afin que tout le monde puisse en réaliser (AMATO 2015b). Ainsi, alors que les représentations graphiques de données étaient l'apanage d'experts (savants, scientifiques ou statisticiens) par le passé, elles peuvent aujourd'hui être produites par n'importe quel individu.

De plus, étant bien souvent diffusées en ligne, tous les citoyens ont accès à ces visualisations *via* les écrans, qui sont très largement répandus dans la population. Le grand public peut donc lire ces graphiques, cartes et réseaux. Ils jouent toujours le même rôle d'aide à la décision que par le passé, mais ils s'adressent désormais à tous et à toutes, alors qu'ils étaient auparavant réservés à une élite lettrée, bourgeoise, dans le but de prendre des décisions politiques, commerciales ou industrielles majeures (AMATO 2015b).

Aujourd'hui, les citoyens obtiennent visuellement des informations présentées de façon efficace et rapide à appréhender, et peuvent ainsi faire des choix éclairés à l'heure où ils revendiquent de plus en plus de pouvoir, de contrôle de leurs institutions et d'intervention dans les processus démocratiques. Ce mode de médiation de l'information favorise les projets civiques et/ou de communautés citoyennes. En effet, les données sont produites en grande quantité, bien souvent par des services publics ou grâce à des renseignements personnels. Les citoyens estiment donc souvent qu'ils ont un droit de regard sur ces données et les informations que l'on peut en tirer. S'instaure alors un nouveau rapport aux données, qui semblent être omniprésentes, à portée de clic, et pourtant souvent inexploitable. Certains décident, de leur propre initiatives, de les réutiliser dans un but informationnel⁷ (FREDRIKSSON 2015).

Malgré cette tentative de prise en main des données par les citoyens, la plupart des datavisualisations restent réalisées par des professionnels de l'information et/ou de la communication. En particulier, en entreprise, les décisions s'appuient souvent sur des informations présentées sous une forme graphique. Cette dernière « rend accessibles des informations existantes

7. On peut en voir des exemples publiés sur le site <https://www.data.gouv.fr/fr/>

et, plus largement, améliore la supervision et la gouvernance » (AMATO 2015b). un domaine d'expertise a même été créé : celui de la *business intelligence*, destinée à fournir de l'information décisionnelle grâce à l'analyse de datavisualisations. AMATO (2015) affirme d'ailleurs que :

« Pour les organisations, le traitement visuel des données devient un moyen incontournable pour la prise de décision. Il permet de mettre au jour des données non exploitées, mais aussi d'établir une base de connaissances cumulatives pour anticiper des actions. »

Les médias journalistiques se sont également appropriés cette forme de médiation pour présenter certaines informations (FREDRIKSSON 2015). Dans ce cas, c'est avant tout l'aspect « synthétique » de l'image qui est recherché, et vient souvent en « complémen[t] du traditionnel régime textuel plus analytique » (AMATO 2015b). Cependant, la visualisation peut parfois se suffire à elle-même, comme le souligne BOY (2015) :

« Dans certains articles de presse en ligne (dont ceux du New York Times), [la datavisualisation] se substitue même au texte. »

Aujourd'hui, les datavisualisations, attrayantes visuellement, sont donc démocratisées. Elles interviennent dans de nombreux champs professionnels (médecine, intelligence économique, politique, sciences sociales, etc.) en tant qu'outil d'aide à la décision, mais servent aussi à la communication, quelle soit marketing ou journalistique, pour convaincre.

Comment ces « outils heuristiques » (ARRUABARRENA 2015b) réussissent-ils à mettre en lumière des informations qui aident à la décision ou convainquent à l'action ? Nous allons tenter de répondre à cette question en étudiant la façon dont ces outils sont conceptualisés par quelques grands noms du domaine.

2.3 La datavisualisation : un concept, deux approches

2.3.1 De la graphique en France...

L'article « La graphique », du cartographe BERTIN (1970) est une référence en matière de représentation graphique de données en France. En effet, la plupart des experts français du domaine y font référence, à l'image de FEKETE (2015) qui dit :

« C'est le cartographe Jacques BERTIN qui, au début des années 70, a posé les bases de la construction graphique de la visualisation. »

FREDRIKSSON précise que « la sémiologie graphique de Jacques BERTIN (...) constitu[e] l'ensemble des règles d'un système graphique de signes pour la transmission d'une information » (FREDRIKSSON 2015).

Aussi, nous pouvons estimer que, aux yeux de la plupart des spécialistes français de ce domaine, la datavisualisation est un anglicisme qui correspond au concept que BERTIN appelait « graphique ». En effet, « la graphique » a pour objectif de représenter visuellement « toute série homogène d'observations,

tout comptage » (BERTIN 1970), autrement dit des jeux de données. De plus, il s'agit bien d'une image monosémique (comme nous le verrons dans les paragraphes qui suivent) qui vise à expliquer des phénomènes. Nous retrouvons bien là toutes les caractéristiques fondamentales de la datavisualisation.

Une définition de la datavisualisation

BERTIN explique les mécanismes de « la graphique ». Il commence par la qualifier de « "langage" destiné à l'œil », autrement dit d'image, et l'oppose à tous les « langages sonores », qui « obéissent aux lois de la linéarité ». Ceux-ci sont, à son sens, destinés initialement à être perçus par l'oreille, et peuvent être verbalisés (lus ou chantés). Il cite en particulier les « écritures musicales, verbales et mathématiques », autrement dit les partitions, les textes et suites de formules et autres équations mathématiques. Ces formes d'expression nécessitent de prendre en compte une dimension temporelle, puisque l'on ne peut les lire et les entendre que linéairement, une ligne après l'autre. En effet, il précise que « les transcriptions scripturales de la musique, du verbe et des mathématiques ne sont que des formules de mémorisation de systèmes fondamentalement sonores, et que ces formules n'échappent pas au caractère linéaire et temporel de ces systèmes » (BERTIN 1970).

La datavisualisation est donc une forme de langage non dépendante de contraintes temporelles, selon BERTIN. Comme lui, nous excluons donc de son champ les « image[s] animée[s] », qui sont soumises à de telles contraintes. Les datavisualisations peuvent être qualifiées de langage car elles présentent un système de signes destinés à être perçus par l'être humain visuellement, comme toute image. Cependant, la graphique ne correspond pas à n'importe quelle image. En effet, BERTIN précise qu'il s'agit d'images « monosémiques », c'est-à-dire employant exclusivement des signes dont la signification est précisée et portée à la connaissance du récepteur en amont de l'observation de l'assemblage des signes. En effet, il explique que « un graphique ne se conçoit qu'une fois précisée, par la légende, l'unique signification de chaque signe ». Citons encore les effets de l'emploi d'un système de signes monosémiques :

« Lorsqu'on emploie un système monosémique, le domaine considéré est rigoureusement précisé et délimité, si grand soit-il. Toute interprétation *a priori*, toute discussion sur le mot est par définition réglée au préalable. »

BERTIN 1970

C'est ainsi que la graphique se distingue des autres types d'images (figuratives ou non) qui sont polysémiques⁸ dans lesquelles au contraire « l'assemblage des signes que [permet de] déduire la signification » (BERTIN 1970).

La graphique n'est pas seulement une image monosémique, mais également une visualisation en raison de sa fonction explicative. En effet, selon les mots de BERTIN, elle « fait partie des systèmes de signes fondamentaux que l'homme a construits pour retenir, comprendre et communiquer les

8. Voir partie 1.2

observations nécessaires à sa survie et à sa vie pensante ». Il ne s'agit pas d'une simple « illustration tribulaire des seules règles de l'esthétique », mais d'un « instrument de recherche » puisqu'elle permet de mettre en évidence les rouages de certaines situations et de certains faits, et donc d'expliquer des phénomènes. BERTIN souligne également son rôle de « mémoire artificielle », car en fixant signes signifiants sur un support, elle permet de les retrouver et de les relire de façon diachronique, comme toute forme d'écriture (BERTIN 1970). Les datavisualisations sont donc, comme le dit ARRUBARRENA (2015), un « outil heuristique » puissant.

La datavisualisation est donc, selon BERTIN, une image, utilisant un système de signes monosémique représentant des données, dont le but est de comprendre un phénomène et de garder une trace pour s'en souvenir et/ou communiquer. Comme cet auteur de référence, nous considérerons que la datavisualisation « recouvre l'univers des réseaux, celui des diagrammes et enfin l'univers des cartes qui s'échelonne de la reconstitution atomique à la transcription des galaxies » (BERTIN 1970). Ces trois catégories de datavisualisations se différencient par le nombre de « composantes » à croiser, une composante désignant un ensemble de données d'une même nature (ou une « série d'analyse » dans le langage des logiciels informatiques). Par exemple, des valeurs de distances (quelque soit leur nombre) forment une composante. De même, un nombre d'individus serait une autre composante. Mais si l'on souhaite croiser ces deux types de données (par exemple, pour établir un diagramme montrant le nombre d'individus vivant à différentes distances du centre-ville d'une métropole quelconque), le résultat est alors une datavisualisation à deux composantes.

- Lorsque l'on cherche à établir des « relations entre les éléments d'une seule composante », la datavisualisation obtenue est un réseau. Par exemple, un arbre généalogique qui montre les liens entre des individus (données de même nature) en est un. Il en va de même pour les plans des lignes de métro, qui mettent en évidence les liens entre des stations (une seule composante) sans prendre en considération le terrain.
- Quand on souhaite établir des « relations entre [plusieurs] composantes », on forme un diagramme. Ainsi, une courbe montrant l'évolution des prix d'un produit (composante 1) au cours du temps (composante 2) relève de la catégorie des diagrammes selon BERTIN. De même, un histogramme comparant le nombre cumulé de personnes infectées (composante 1) et celui des morts (composantes 2) du coronavirus dans plusieurs pays (composante 3) est un diagramme.
- Lorsque l'on ordonne un réseau à l'organisation spatiale observée, on obtient une carte. Par exemple, lorsque le réseau des stations de métro parisien est appliqué à une représentation de l'espace parisien, on obtient une carte du métro (BERTIN 1970).

Un outil heuristique

La thèse de BERTIN (à laquelle nous adhérons) est qu'une datavisualisation, si elle utilise les propriétés naturelles de la perception visuelle, permet comprendre des phénomènes très rapidement. Comme le dit BERTIN (1970) :

« Il faudrait au moins 20 000 instants successifs de perception pour comparer deux tableaux de chiffres de 100 lignes sur 100 colonnes. Que les chiffres soient transcrits graphiquement et la comparaison est aisée, et peut même être instantanée. »

Cependant, il ne suffit pas de représenter les données visuellement pour que leur contenu informationnel apparaisse évident à l'observateur. Encore faut-il suivre les lois de la perception visuelle naturelle, qu'énonce BERTIN. Nous montrons les deux exemples que prend le cartographe pour démontrer cette idée en figures 2.8 et 2.9.

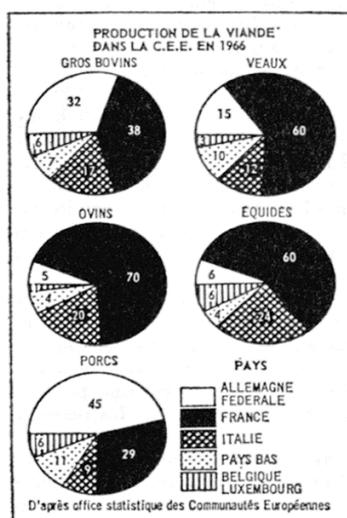


Fig. 4. Exemple de construction inefficace et inutile.

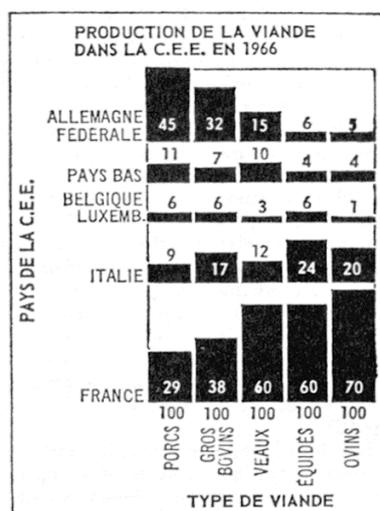


Fig. 5. Exemple de construction efficace et utile. Même information que fig. 4. Mais construction en x, y, z. En x, les types de viande; en y, les pays de la C.E.E.; en z, les quantités.

FIGURE 2.8 – Comparaison de l'efficacité de deux diagrammes par BERTIN

Source : BERTIN 1970

Dans chacun de ces cas, la première datavisualisation, bien qu'elle soit esthétiquement irréprochable et construite à partir de données bien exploitées, ne permet pas d'obtenir une vue d'ensemble de la situation, et donc de se représenter le phénomène dans sa globalité. L'information est bien présente, mais il faut la lire en se focalisant successivement sur chaque zone, chaque forme, et en se référant à la légende. Au contraire, la seconde visualisation offre une vision d'ensemble interprétable en quelques secondes : concernant la figure 2.8, nous voyons d'un coup d'œil que l'Allemagne fédérale était, en 1966, le plus grand producteur de porc de la CEE, alors que la France produisait beaucoup d'ovins. De plus, on visualise bien que la France produisait à cette époque plus de viande que tous les autres pays de la CEE. Ainsi, nous déduisons l'information efficacement et rapidement. De même, la seconde carte de la figure 2.9 permet d'identifier instantanément les zones de l'Est de la France les plus chères.

Comme le souligne BERTIN (1970) :

« Sans image d'ensemble significative, [l'observateur] ne peut dégager les relations essentielles que fournit cette information.

Le prix du terrain dans la France de l'Est.

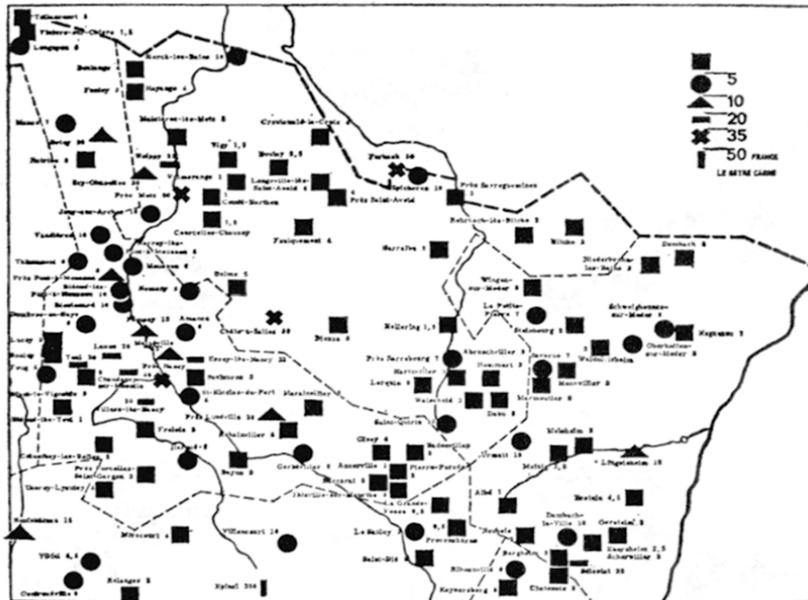


Fig. 6. Transcription des prix par une variation de forme.

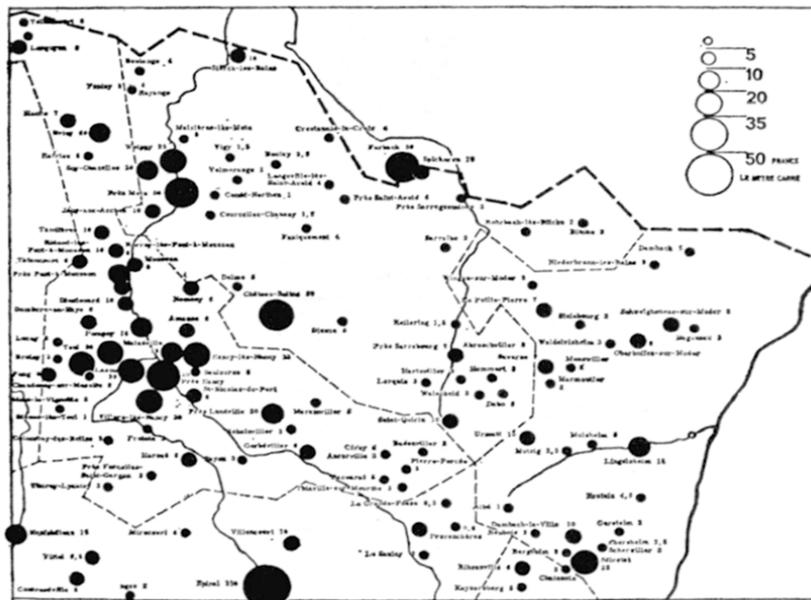


Fig. 7. Transcription des prix par une variation de taille.

FIGURE 2.9 – Comparaison de l'efficacité de deux cartes par BERTIN
Source : BERTIN 1970

Faute de mieux, il ne retient que quelques pourcentages. »

Ainsi, pour concevoir une datavisualisation efficace, il est nécessaire de connaître les caractéristiques de la perception visuelle. Selon BERTIN, celle-ci « dispose de trois variables sensibles : la variation des taches [élémentaires] et les deux dimensions du plan ». Autrement dit, l'œil perçoit naturellement que

chacune des deux dimensions du plan représente une composante différente, de même que les différences de signes (« variation des taches ») sont associées spontanément à d'autres composantes. C'est pourquoi les datavisualisations peuvent nous communiquer « dans un instant de perception (...) les relations entre trois variables » en se servant simplement de la « puissance considérable de la vision » (BERTIN 1970).

Les taches élémentaires (en forme de points, lignes ou surfaces) qui constituent les datavisualisations peuvent varier selon huit variables visuelles facilement perceptibles :

- Les deux dimensions du plan (DP);
- Leur taille (T);
- Leur valeur (V);
- Leur grain (G);
- Leur couleur (C);
- Leur orientation (OR);
- Et leur forme (F).

La figure 2.10 illustre ces huit variables visuelles.

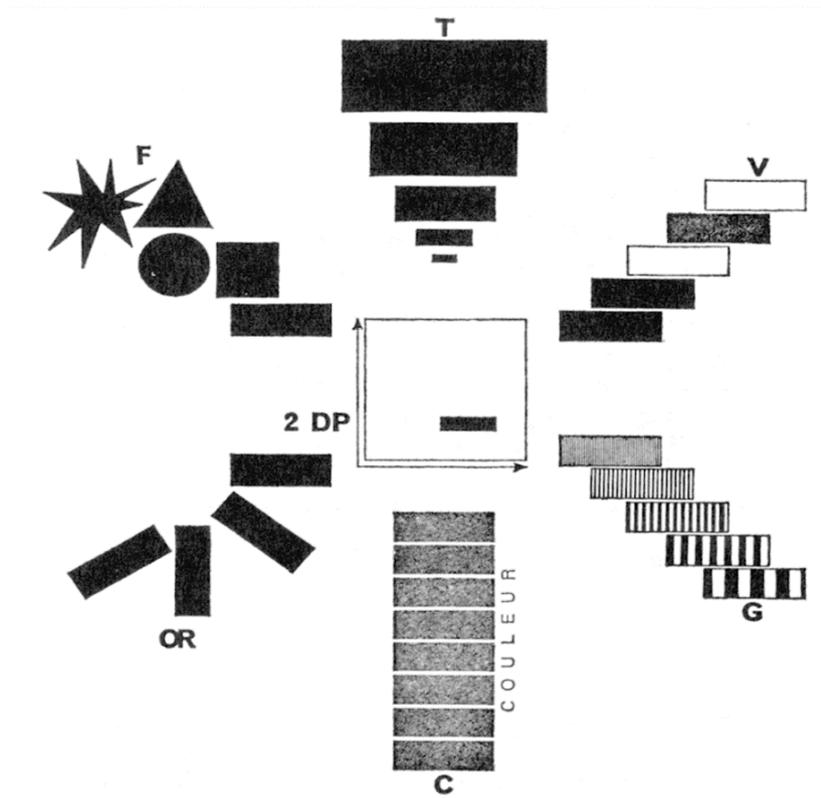


FIGURE 2.10 – Les huit variables visuelles selon BERTIN

Source : BERTIN 1970

Ces huit variables sont utilisées pour mettre en évidence des relations de ressemblance, de différence, d'ordre et de proportionnalité entre les éléments de chaque composante. Cependant, certaines caractéristiques visuelles permettent de représenter certains types de relation et pas d'autres. Le tableau ci-dessous reprend les recommandations de BERTIN sur ce sujet.

Variable visuelle	Ressemblance	Différence	Ordre	Proportionnalité
Forme	✓			
Orientation	✓	✓		
Couleur	✓	✓		
Grain	✓	✓	✓	
Valeur	✓	✓	✓	
Taille	✓	✓	✓	✓
Dimensions du plan	✓	✓	✓	✓

Par ailleurs, BERTIN souligne que la perception visuelle permet d'accéder à plusieurs niveaux d'informations : on peut percevoir l'ensemble de la figure (ce qui apporte une information générale), mais on peut également décrypter les données associées à une tâche élémentaire (ce qui permet d'obtenir une information de « niveau élémentaire »). En s'intéressant à des « groupement[s] de tâches », on peut même avoir des « niveaux moyens de lecture ». Aussi BERTIN considère-t-il (à juste titre) que les datavisualisations peuvent représenter énormément d'informations sur une seule image, en fonction du niveau de lecture de l'observateur (BERTIN 1970).

Enfin, toute datavisualisation doit être construite. Son créateur doit alors analyser l'information à traiter : le nombre de composantes, le nombre de données dans chaque composante, le type de relation à faire apparaître (ordonner, mettre en valeur des différences ou des similitudes, rendre visible des effets de proportionnalité...). De plus, dès qu'une datavisualisation doit représenter plus de trois composantes, il est nécessaire de faire des compromis entre exhaustivité de l'information et la simplicité de lecture. En effet, plus on ajoute de composantes, plus la représentation graphique se complexifie. Aussi BERTIN conseille-t-il de privilégier l'exhaustivité lorsque la datavisualisation « sert de mémoire artificielle », c'est-à-dire quand on peut se référer au document sans avoir à le mémoriser (par exemple, une carte de métro), mais de réduire la quantité d'informations de niveau élémentaire lorsqu'il le graphique est utilisé à des fins communicationnelles ou pédagogiques afin de ne mettre en valeur que les corrélations essentielles (BERTIN 1970).

Cette analyse permet alors de choisir la représentation qui prend le mieux en compte la perception visuelle afin d'avoir une compréhension optimisée en fonction du besoin. Comme le résume BERTIN (1970) :

« Construire une représentation graphique consiste donc à transcrire chaque composante (...) par une variable visuelle, de telle façon que la construction soit conforme à l'image naturelle [et] que la lecture n'exige que le nombre minimum d'instant de perception. »

2.3.2 ... au design originaire des États-Unis

Si BERTIN est une référence dans le domaine de la datavisualisation aux yeux des experts français, il n'en va pas de même Outre-Atlantique. Il est d'ailleurs surprenant de constater que les articles américains ne le mentionnent quasiment jamais, à l'exception de ceux qui retracent l'histoire du domaine et qui ont procédé à des recherches fouillées quand aux évolutions des représentations graphiques de données au XX^e siècle. Aux États-Unis, les figures qui font autorité en la matière sont TUKEY, statisticien qui a promu la

datavisualisation au rang de science, et le « gourou des graphiques », TUFTE (COOKSON 2013).

Les enseignements de TUFTE

Les approches américaine et française sont sur deux lignes idéologiques très différentes. Bien que l'objectif demeure inchangé (mettre en évidence des phénomènes de façon claire et univoque en faisant apparaître graphiquement les relations entre des données) et que les moyens employés soient sensiblement les mêmes (un système de signes monosémique dessiné, bien souvent, par l'outil informatique), les philosophies qui les sous-tendent et permettent de les théoriser sont bien distinctes. En effet, là où « la graphique » de BERTIN hérite des codes de la cartographie et insiste sur la présence d'une légende et d'un système de signes conventionnel, TUFTE et les autres spécialistes américains s'orientent plutôt vers le design, l'interaction utilisateur, et le mélange entre texte, image et visualisation.

Les datavisualisations américaines revendiquent en particulier leur design centré utilisateur et leur ergonomie. TUFTE a d'ailleurs été professeur de statistiques et de design d'information à l'université de Yale (MATEY 1998). Lorsqu'il décrit lui-même les visualisations qui l'ont marqué, il ne cite que peu de datavisualisations conventionnelles (histogrammes, courbes, cartes...). Il évoque plutôt des travaux atypiques et expérimentaux, comme la carte de MINARD sur la campagne de Napoléon de 1812–1813 (très belle et efficace mais qui demeure un exemple unique, jamais reproduit avec succès); des cartes du relief des Alpes de IMHOF, sur lesquelles les reliefs sont colorisés (et pas seulement représentées par des courbes de niveaux, comme c'est traditionnellement le cas); ou encore des innovations de BOSTOCK, le créateur de la librairie JavaScript D3.js, qui permet de réaliser des visuels extrêmement divers et novateurs (dont on peut voir des exemples sur <https://d3js.org/>). De plus, en évoquant son propre travail, TUFTE ne fait pas référence exclusivement à des datavisualisations, mais également à des œuvres d'art (des sculptures mégalithiques, par exemple). Les titres de ses livres sont également évocateurs, comme *Beautiful Evidence* (2006). On sent bien que son approche de la datavisualisation est orientée vers une recherche esthétique. Il s'agit, bien entendu, de faire passer un message clair, mais au moyen de représentations ergonomiques et élégantes.

Nous ne retrouvons pas ici l'exigence principale de BERTIN, à savoir l'emploi d'un système de code monosémique, ni les catégorisations des différentes formes graphiques (diagrammes, cartes et réseaux). L'idée de TUFTE est plutôt de présenter un maximum de données expliquant un phénomène, en les rendant explicites grâce aux enseignements du design informationnel. Il conseille en particulier de limiter les figures présentées au minimum pour se concentrer sur l'information centrale, de ne pas hésiter à expliciter le contexte directement grâce à des indications écrites sur le graphique, de ne représenter que le minimum de données d'un même jeu et de supprimer celles qui n'apportent pas d'information supplémentaire de façon à augmenter la lisibilité... Il suggère même de modifier les axes des graphiques pour les borner aux valeurs maximales et minimales des données représentées, ou de ne pas les tracer et d'afficher simplement des suites de valeurs (*Tufte on Design and Data* 2021).

Selon ce statisticien, six principes fondamentaux doivent structurer les datavisualisations :

- Montrer des comparaisons ;
- Mettre en évidence les liens de causalité ;
- Utiliser de nombreuses variables (plus de trois « composantes » au sens de BERTIN) ;
- Intégrer texte, images et chiffres dans un même visuel ;
- Se focaliser sur le contenu informationnel, et limiter à la « pollution graphique » (ou « *chartjunk* »), c'est-à-dire toute écriture ou dessin inutile à la compréhension ;
- Afficher les sources de données afin d'établir sa crédibilité (*Tufte on Design and Data* 2021).

Ainsi, contrairement à BERTIN, TUFTE ne propose pas de méthode, ni de code à maîtriser et à appliquer pour créer une datavisualisation. Il semble plutôt attiré par les innovations, et promeut la lisibilité et les mises en évidence de liens de causalité pour comprendre l'information rapidement. Il s'oppose à BERTIN, non pas sur l'objectif (qui est toujours de construire un outil efficace cognitivement et favorisant la compréhension), mais dans la forme envisagée pour les représentations graphiques de données : là où BERTIN prône d'emploi d'un code fort pour transmettre l'information, TUFTE préfère l'étude au cas par cas et la mise en valeur de l'information par le design.

Un exemple d'experts en datavisualisations américains : MANOVICH

Dans la lignée du statisticien, des experts de la visualisation d'information (en particulier américains) affirment clairement le lien entre datavisualisation et design en SIC. La différence entre ces deux champs est explicitement définie par MANOVICH. Selon lui, la datavisualisation a pour objectif de mettre en évidence la structure d'un jeu de données. Le design d'information prend ensuite la relève pour trouver la façon la plus adaptée de montrer cette structure visuellement (MANOVICH 2011). Les deux concepts sont donc intimement liés dans la culture américaine.

La datavisualisation est définie par MANOVICH comme une cartographie visuelle de données discrètes. Il ne donne aucune indication sur la forme que peuvent prendre ces représentations graphiques. En effet, il explique que, à ses yeux, les représentations graphiques classiques (comme celles que recommandent BERTIN) s'appuient trop sur des caractéristiques spatiales pour représenter les informations importantes⁹, et pas assez sur des distinctions de couleurs¹⁰. Il émet l'hypothèse que le peu de couleurs différentes sur les datavisualisations antérieures aux années 1990–2000 serait lié à l'absence d'outil adéquat pour reproduire facilement, mais que l'outil informatique est en train de modifier les pratiques. Hors, si cela est vrai pour certains designers tels que MCCANDLESS, il n'en est rien pour la plupart, qui estiment contre productif de se diriger vers une débauche de couleurs au détriment de la

9. Nous verrons plus loin que les deux dimensions du plan permettent de représenter tous les types de relations qu'une datavisualisation peut montrer, cela correspond à une capacité de la perception humaine.

10. La couleur ne symbolise cognitivement que la ressemblance ou la différence, comme nous le verrons plus loin avec la « sémiologie graphique » de BERTIN

focalisation de l'attention sur certains points particuliers ¹¹.

De plus, MANOVICH considère que les données ne sont pas, comme nous l'avons défini, des valeurs qualifiant ou quantifiant un élément, mais cet élément lui-même. Selon lui, il vaudrait ainsi mieux ne pas créer un histogramme ou une courbe représentant la colorimétrie d'une image, mais proposer au spectateur l'image elle-même, quitte à la décomposer pour en faire ressortir les différentes couleurs. L'encodage des données dans un système de signes univoque est ici complètement remis en cause. MANOVICH considère, au contraire, que ce procédé « réduit » les données puisqu'il est impossible de montrer tous les aspects d'un élément sur un graphique. Il estime donc préférable de montrer directement l'élément. Il appelle cela la « visualisation directe », dont il donne des exemples ¹².

Mais dans ce cas, a-t-on encore à faire à une datavisualisation ? Puisque nous venons de dire que cette proposition ne coïncide pas avec notre définition initiale de la donnée (que nous maintiendrons tout au long de ce mémoire), ces représentations ne sont pas, à nos yeux, des datavisualisations. On pourra les considérer comme des visualisations artistiques, mais elles ne répondent pas aux critères établis précédemment par BERTIN et qui nous semblent mieux correspondre aux concepts de donnée et de visualisation que nous avons donné en amont.

Aussi, bien que le design informationnel prenne de plus en plus de place dans le domaine de la datavisualisation actuellement, en particulier sous l'impulsion de designers américains, une bonne connaissance de la perception humaine et la sémiologie nous semblent plus pertinentes pour créer des datavisualisations efficaces. C'est donc BERTIN qui nous servira de référence tout au long de ce mémoire, car il semble s'appuyer sur des modèles corroborés par les sciences cognitives. De plus, bien que le design de données prôné par TUFTE propose des productions intéressantes et inspirantes, elles correspondent toutes à des prototypes, puisque leur principe fondamental est d'être adapté à chaque situation, à chaque jeu de données à visualiser et à chaque information à communiquer. Il nous semble donc difficile de prendre pour référentiel des modèles uniques par essence.

Résumé — La notion de datavisualisation

Comme BERTIN, nous considérerons donc que les datavisualisations sont des images monosémiques représentant les relations entre une ou plusieurs composantes de données. Les frises chronologiques, diagrammes circulaires, histogrammes, courbes ou nuages de points dans des repères cartésiens, cartes, etc. en font partie du moment qu'ils présentent une légende ne laissant planer aucune ambiguïté quant à l'interprétation des données représentées.

11. Voir pour cela la critique de FEW à l'url : <http://www.perceptualedge.com/blog/?p=935>.

12. Voici quelques exemples de « visualisation directe » selon MANOVICH :

- <https://brendandawes.com/projects/cinemaredux>;
- <https://fathom.info/traces/>;
- <http://lab.softwarestudies.com/p/cultural-analytics.html>.

L'objectif de ces images monosémiques est avant tout de mettre en évidence des informations issues du croisement de données. Pour cela, la perception humaine naturelle doit être respectée, que ce soit en suivant les règles établies par BERTIN ou en se basant sur le design centré utilisateur.

Par ailleurs, nous avons établi que les datavisualisations ont une longue histoire, et ne sont pas nées à l'ère de l'informatique. Leur intérêt heuristique (c'est-à-dire leur capacité à s'appuyer sur des processus cognitifs pour faciliter la compréhension) leur permet depuis longtemps de jouer un rôle d'aide à la décision. Ces outils visuels ont évolués avec les techniques de leur époque, jusqu'à proposer aujourd'hui parfois des fonctionnalités interactives. Cependant, selon nous, il peut y avoir datavisualisation sur un support non numérique et sans interactivité.

Enfin, nous avons souligné que les datavisualisations ont longtemps été réservées à des experts, mais qu'elles tendent à se démocratiser depuis l'avènement de l'informatique. Cette situation nous invite à nous interroger sur la réception de ce type de message visuel : tout le monde possède-t-il aujourd'hui la littératie nécessaire pour comprendre les datavisualisations ?

CHAPITRE 3

LE CONCEPT DE LITTÉRATIE

À l'instar de BERTIN, nous considérons les datavisualisations comme une forme de langage. Il s'agit d'une forme particulière de transmission de message par un code visuel. Suivant ces considérations, nous pouvons appliquer aux représentations graphiques de données le concept de littératie, que nous allons définir ci-après.

3.1 Une définition de la littératie

Concept apparu dans les années 1960, la littératie concerne les pratiques d'écritures et de lectures. En particulier, l'anthropologue GOODY et LEJOSNE (2006) a beaucoup étudié ce sujet, et est aujourd'hui une référence. En effet, l'œuvre de sa vie a consisté à expliquer les différences entre les sociétés « littératiennes », pratiquant l'écriture, et « non-littératiennes ».

3.1.1 Une définition de l'écriture

Puisque la littératie se définit par rapport à l'écriture, nous devons établir de façon univoque ce que regroupe ce dernier terme.

GOODY parle de l'écriture comme d'un « medium de communication entre les hommes », consistant en un « ensemble de signes visibles » tracés sur un support. Cette matérialité de l'écriture lui permet d'être transmise « dans l'espace et (...) dans le temps ». Il considère l'écriture comme une « technologie de l'intellect » performante heuristiquement puisqu'elle permet de développer des « procédures analytiques et critiques propres aux nouveaux modes de la pensée littératiennne » (GOODY et LEJOSNE 2006). Pour mieux comprendre cette idée de lien entre écriture et pensée logique, citons une de ses phrases :

« Le fait d'écrire, en donnant aux mots une forme physique et en les mettant à disposition en même temps que leurs signifiés pour une analyse bien plus approfondie et minutieuse que l'oral ne l'autorise, encourage la réflexion privée. »

GOODY, WATT et LEJOSNE 2006

L'écriture selon GOODY ne concerne pas uniquement la prose : il parle également de listes, de tableaux et d'organigrammes. Ces derniers peuvent être considérés comme des datavisualisations (des réseaux au sens de BERTIN. Aussi peut-on admettre que les datavisualisations font partie du vaste ensemble nommé « écriture ».

Si l'écriture est une part importante de la littératie, il ne faut pas oublier sa réception. En effet, s'il s'agit de générer des signes visibles du côté du producteur, le récepteur de la communication doit pouvoir en tirer un sens. C'est ce qu'évoquent BAZERMAN, DONAHUE et LEJOSNE (2006) dans leur article « Écriture, organisation sociale et cognitions : repenser GOODY » :

« La littératie active des mécanismes psychologiques par lesquels nous créons du sens et nous nous ajustons sur ce que les autres cherchent à nous communiquer. »

3.1.2 Une définition de la littératie

Maintenant que nous avons explicité le terme d'écriture, nous pouvons donner une définition précise de la littératie. Pour cela, nous nous appuyons sur la définition de JAFFRÉ (2005), cité par DROT-DELANGE (2014). Selon lui, la littératie « désigne l'ensemble des activités humaines qui impliquent l'usage de l'écriture, en réception et en production. Elle met un ensemble de compétences de base, linguistiques et graphiques, au service de pratiques, qu'elles soient techniques, cognitives, sociales ou culturelles. Son contexte fonctionnel peut varier d'un pays à l'autre, d'une culture à l'autre, et aussi dans le temps ».

Il s'agit bien là d'une définition en adéquation avec celle que propose GOODY, puisqu'elle évoque « l'écriture, en réception et en production », ainsi que « son contexte » en fonction du pays, de la culture et de l'époque. En revanche, si nous choisissons cette définition comme base, c'est parce qu'elle énonce clairement que la littératie mobilise « un ensemble de compétences de base, linguistiques et graphiques », ce qui implique la nécessité d'un apprentissage.

La notion de littératie est aujourd'hui très présente au sein des sciences de l'éducation : elle y désigne au sens strict la maîtrise de la lecture et de l'écriture; mais peut aussi être employé dans une vision plus large, notamment au Canada, pour désigner un ensemble de compétences minimales par les élèves, englobant en plus des aptitudes au calcul, des connaissances scientifiques de base, ou encore la capacité à décoder les messages visuels transmis par les médias, et la compréhension de l'informatique (GRIIPTIC 2019).

Si, à l'origine, le concept de littératie n'incorporait pas les écritures informatiques (trop récentes à l'époque), il est légitime aujourd'hui, alors que nous lisons continuellement des inscriptions en lignes et que nous décodons des pictogrammes appelés « *smiley* », de s'interroger sur la littératie informatique. Bien entendu, tout texte, en français ou en anglais par exemple, apparaissant sur un écran d'ordinateur relèvera de la littératie générale. Mais qu'en est-il des programmes informatiques, des algorithmes, et des jeux de données ?

3.2 La littératie informatique

3.2.1 Précisions sur les termes

Le concept de « littératie informatique » (ou *computer literacy* en anglais) provient de la culture anglo-saxonne. En effet, DROT-DELANGE (2014), qui a réalisé un état de l'art sur cette notion, constate que « l'expression n'est pas utilisée dans les travaux francophones, à de rares exceptions près, plutôt récentes ». Une autre expression peut-être identifiée dans les textes : la « littératie numérique » (*digital literacy*). Cette dernière est plus récente (apparue à la fin des années 1990 en réponse à la diffusion de nombreux objets numériques, alors que le concept de *computer literacy* remonte aux années 1980). Bien que les mots soient différents car adaptés aux dénominations en vogue à leur époque, les deux expressions renvoient souvent à la même notion (DROT-DELANGE 2014).

Il faut noter que ce concept ne fait pas l'unanimité : il est, en effet, contesté par certains, selon lesquels il aurait été inventé non pas pour « former des citoyens mais des consommateurs à l'aise avec l'informatique », qui achèteraient des objets numériques d'autant plus qu'ils en maîtriseraient les codes et usages (DROT-DELANGE 2014).

3.2.2 La pensée informatique

Si la notion véhicule des controverses et est perçue différemment selon les auteurs, le point départ de toute réflexion sur la « littératie informatique » est l'idée que le numérique apporte une nouvelle façon de raisonner et de résoudre des problèmes, au même titre que l'écriture. On parle alors de « pensée informatique » (*computational thinking*). DROT-DELANGE cite le rapport du CSTA (2011) pour donner une définition de cette nouvelle façon de penser :

« La pensée informatique y est définie comme une “approche pour résoudre des problèmes que l'on peut mettre en œuvre avec un ordinateur. Elle implique l'usage de concepts, tel que l'abstraction, la récursivité, l'itération, le traitement et l'analyse de données et la création d'artefacts réels et virtuels”. »

La pensée informatique serait « un savoir fondamental pour tous, véhiculé par les technologies numériques, tout comme l'imprimerie aurait permis la diffusion des savoirs comme lire–écrire–compter ». C'est d'ailleurs ce que pensent de hautes instances de l'enseignement, à l'image de la *Royal Society* britannique, qui dit qu'une « éducation de haute qualité en informatique permet aux élèves de comprendre et de changer le monde grâce à la pensée informatique. Elle développe et exige une pensée logique et de précision. Elle combine créativité et rigueur ». La littératie informatique devient alors capitale puisqu'elle permet de développer « les moyens intellectuels de penser et de se représenter un problème, de formaliser sa pensée, d'étudier et de critiquer celle des autres » (DROT-DELANGE 2014).

3.2.3 Une définition de la littératie informatique

Dans les citations précédentes, on perçoit le lien avec la définition de la littératie générale, basée sur l'écriture. D'ailleurs, certains vont jusqu'à

considérer les langages de programmation informatique comme une forme d'écriture. En effet, il s'agit bien d'un ensemble de signes consignés sur un support (disque dur, serveurs...) dans un objectif de communication à travers le temps et l'espace, puisqu'ils « transmettra[en]t des intentions humaines (...) à une machine, mais aussi à d'autres personnes ». Les tenants de cette position considèrent alors que la maîtrise de la littératie informatique passe par une connaissance de l'algorithmique et une maîtrise de la programmation (DROT-DELANGE 2014).

D'autres objectent que cette posture est trop techniciste, et que le but de la littératie est avant tout « l'acquisition de compétences critiques nécessitant une connaissance générale des ordinateurs et de leurs usages, ainsi que des enjeux sociétaux de l'informatique » (DROT-DELANGE 2014). En arrière-plan de cette dernière assertion, il y a la notion de citoyen, qui doit être assez éclairé pour se rendre compte de « rôle et [de] l'impact de l'informatique dans la société » (DROT-DELANGE 2014).

Cette dernière vision est plus proche de celle qui nous intéresse : dans le contexte où le numérique est présent partout, il devient capital d'avoir la capacité de poser un regard critique sur les diverses productions qu'il nous propose. La littératie informatique sera donc considérée ici comme un ensemble de connaissances et compétences à acquérir pour développer cette approche critique de l'informatique et de ses contenus, sans nécessité de maîtriser la programmation.

3.3 La littératie des données

3.3.1 Une définition de la littératie des données

La littératie des données (ou *data literacy* en anglais, aussi parfois orthographié *data literacy*) se rapproche de la littératie informatique en ce sens qu'elle ne consiste pas à comprendre les données et à savoir en tirer des informations, mais aussi à prendre du recul quant à la production, la fiabilité et la qualité des données. C'est ce qu'exprime J. DENIS (2018) lors de son retour sur la troisième conférence *Data Literacy* :

« Développer la *data literacy*, ça n'est pas seulement apprendre à tout le monde à lire correctement des données, c'est aussi interroger la fabrique même des données, les moments et les acteurs de leur écriture. Et c'est, dans le meilleur des cas, outiller et former à cette écriture. »

Cette qualification rejoint certaines définitions proposées par ARRUBARRENA, KEMBELLEC et CHARTRON (2019), qui expliquent :

« Le concept de data littératie est la capacité des individus "d'accéder, d'interpréter, d'évaluer de manière critique, de gérer, de manipuler et d'utiliser éthiquement les données". En ce sens, l'enjeu d'une data littératie est le développement "d'une culture des données", i.e. la capacité à évaluer de manière éthique et critique les données. »

D'autres définitions — rassemblées par ARRUBARRENA, KEMBELLEC et CHARTRON (2019) — prennent moins en compte cette analyse critique des mécanismes de créations de données, et se concentrent plutôt sur l'utilisation des données (comme, par exemple, MANDINACH et GUMMER (2013), qui mentionnent « la capacité à comprendre et à utiliser les données de manière effective pour la prise de décision ») ou au contraire sur des « compétences mathématiques, informatiques et statistiques ». Cependant, ces deux approches seraient réductrices, en se focalisant soit sur le traitement, soit sur l'utilisation des données.

Ainsi, selon les deux définitions retenues, la littératie des données englobe toute une culture de la donnée : il est nécessaire d'interroger la fabrique des données, d'être capable de juger de leur qualité, de savoir se positionner quant à l'éthique en matière de collecte et de réutilisation de données, de comprendre les mécanismes du *Big data*, ou encore de connaître le cadre juridique (RGPD en France) pour prétendre maîtriser la littératie des données.

3.3.2 Lire et comprendre des données

Si les données peuvent être étudiées sous l'angle de la littératie, c'est parce qu'elles « sont une affaire de langage. Et plus encore, elles sont affaire d'écrit ». En effet, des données sont des signes inscrits sur des supports (souvent informatiques, mais pas nécessairement). De plus, elles servent à communiquer des informations (dont elles constituent les briques essentielles) entre les êtres humains, et « sont ancrées » dans un contexte. Enfin, « les données font partie des choses qui nous aident à penser le monde, à l'appréhender » (J. DENIS 2018). Les données correspondent donc bien à tous les critères que GOODY établit pour décrire une écriture, c'est-à-dire qu'il s'agit d'un système de signes, sur un support, sous-tendus par une activité intellectuelle, dans un contexte social.

Manipuler des données consiste donc soit à écrire, soit à lire. Mais « cette écriture et cette lecture ne vont pas de soi. Elles passent par un apprentissage » puisqu'elles emploient un langage différent de ceux auxquels nous sommes habitués. En effet, les données ne sont pas un « langage universel, transparent, univoque » (J. DENIS 2018). Citons ARRUBARRENA, KEMBELLEC et CHARTRON (2019) qui soulignent que « la dissociation des données de leur contexte peut rendre l'analyse difficile, et la réutilisation impossible (...) avec le risque d'être mal interprétées ». On peut illustrer cette affirmation en évoquant les « jeux de données dites "métier" [produites au sein d'entreprises ou d'administrations] qui (...) sont manipulés sans aucun problème par celles et ceux qui les utilisent quotidiennement, [mais] restent hermétiques et difficiles à appréhender » pour toute personne extérieure.

Dans le cas des datavisualisations, il s'agit (comme dit précédemment) de représenter des données par un système de signes graphiques. Aussi, la question du recueillement des données se pose-t-elle nécessairement dès l'origine de toute réalisation de datavisualisation. Or, comme le dit J. DENIS :

« les données ne sont jamais à disposition dans les systèmes d'information, prêtes à l'emploi, attendant sagement d'être ramassées et rassemblées pour que leur puissance potentielle soit enfin libérée ».

Plusieurs spécialistes du sujet proposent d'ailleurs d'abandonner le terme de « données » pour le remplacer par celui « d'obtenues »¹.

Les méthodes de production et de rassemblement des données doivent alors être comprises pour maîtriser la chaîne de la donnée, compétence capitale de la data littératie. Cela est d'autant plus centrale aujourd'hui que les générations de données sont le plus « invisibilisées » dans les entreprises et les administrations.

Ainsi, la littératie des données consiste à être capable d'interpréter des données (donc à savoir les lire et leur attribuer un sens) mais aussi de les manipuler (savoir les représenter graphiquement basiquement ou faire des calculs simples pour en retirer des informations pertinentes). Ces compétences s'apprennent, à l'instar de l'écriture, qui consiste à assembler des mots en des phrases qui font sens. La littératie des données implique aussi de conserver un regard critique sur les données utilisées, le contexte de leur production, ainsi que les choix effectués lors de leur représentation visuelle.

3.4 Quelle littératie pour les datavisualisations ?

Lire des datavisualisations impose de maîtriser la littératie générale de GOODY, puisqu'il y a des mots, chiffres, etc. à décoder et à comprendre. Si la datavisualisation est issues d'un processus de production informatisé, il faut également posséder la littératie informatique suffisante pour être capable de porter un regard critique sur l'image. De plus, si la datavisualisation est présentée sur un écran, il est nécessaire de savoir utiliser le dispositif médiateur, ce qui relève également de la littératie informatique. Enfin, il faut maîtriser absolument la littératie des données pour comprendre leur origine, et savoir ce qu'elles peuvent représenter.

Cependant, on peut se demander si elles mobilisent une littératie spécifique, propre à la représentation graphique en elle-même.

3.4.1 Une littératie spécifique pour les datavisualisations ?

Cette question est légitime, car les représentations graphiques de données imposant leurs propres codes (puisque nous avons établi qu'elle correspondent à une forme d'écriture), elles pourraient imposer la maîtrise de compétences spécifiques dédiées à la lecture et à l'écriture de ce code de signes monosémiques forts (au sens de BERTIN). BOY (juill. 2015) explicite d'ailleurs bien cette interrogation :

« Une visualisation n'est pas une image "gratuite" ou purement illustrative : elle utilise des codes graphiques spécifiques qu'un lecteur doit savoir interpréter. »

Nous pourrions donc estimer, comme ARRUBARRENA (juill. 2015), que la datavisualisation est une « compétence info-communicationnelle [à part entière], c'est-à-dire une littératie ».

Pour BOY, la littératie appliquée au concept de datavisualisation se définirait comme : « la capacité qu'a un individu de traduire des questions

1. Voir la partie 1.1 où l'on expose la revendication du terme *capta* par DRUCKER

posées au niveau visuel en questions relatives aux données, et de traduire des questions relatives aux données en “requêtes” visuelles » (BOY 2015). Il s’agirait alors d’une « compétence passerelle » qui consiste à repérer des sèmes (c’est-à-dire des signes signifiants au sens de BARTHES) permettant de répondre à une problématique, à les mettre en relation, puis à interpréter l’ensemble pour lui donner sens, avec la possibilité d’en « faire l’analyse ou la critique ».

Pour illustrer la notion de littératie des datavisualisations, BOY prend l’exemple suivant :

« Alors que détecter un point rouge au milieu d’un groupe de points bleus peut être très rapide, une visualisation n’est réellement efficace que si le lecteur comprend à quoi fait référence ce point rouge au niveau des données. À défaut, l’information que présente la visualisation ne sera pas communiquée. »

Cette définition de la littératie des datavisualisations est calquée sur la définition de la littératie générale : si la datavisualisation emploie un système de code monosémique (ce qui en fait une forme d’écriture), elle nécessite forcément des compétences spécifiques d’encodage et de décodage de l’information (des compétences de lecture et d’écriture) et relève donc d’une littératie.

Cependant, cette définition fait disparaître les littératies informatique et de données. DESFRICHES-DORIA (2015) module cette définition en lui ajoutant l’expression « la capacité à utiliser en confiance une datavisualisation donnée ». Cette notion de « confiance » fait échos à la capacité à porter un regard critique que nous avons attribuée aux littératies informatique et de données. D’ailleurs, DESFRICHES-DORIA exprime elle-même cette idée en argumentant que la littératie des datavisualisations « pourrait (...) inclure la compréhension du fonctionnement, même schématique, des algorithmes qui transforment et traduisent les données en visualisation » (DESFRICHES-DORIA 2015). Il nous semble donc important de prendre en compte cette capacité d’analyse réflexive des datavisualisations dans notre définition de leur littératie.

Dans cette optique, il faudrait s’intéresser à l’influence humaine sur les algorithmes. En effet, ces derniers sont régis par des êtres humains, qui peuvent influencer la perception des datavisualisations par leurs récepteurs par certains choix implémentés dans les algorithmes.

3.4.2 La capacité à porter un regard critique

La littératie des datavisualisations implique donc non seulement d’être capable de lire celle-ci, de comprendre ce dont il est question sur la représentation graphique, de faire le lien entre légende et éléments représentés ; mais elle suppose également une capacité à porter un regard critique sur toute datavisualisation. En faisant un parallèle avec la littératie classique (au sens de savoir lire et écrire), peut-on estimer que quelqu’un maîtrise la lecture s’il sait décoder les signes, les assembler, leur donner du sens, mais sans avoir la capacité d’analyse suffisante pour identifier un discours trompeur ou fallacieux ?

En effet, toute datavisualisation est une construction générée par des individus. Elle ne peut pas donner à voir la réalité, puisqu'elle n'en est qu'une représentation. Elle tend seulement et modestement à représenter un aspect d'une réalité, selon un angle d'approche.

Pour expliciter cet aspect, nous pouvons nous remporter à BARTHES. Ce dernier estime que tout dessin, c'est-à-dire à toute image tracée par la volonté de l'être humain, « oblige à un ensemble de transpositions réglées; (...) l'opération du dessin (...) oblige tout de suite à un certain partage entre le signifiant et l'insignifiant : le dessin ne reproduit pas tout, et souvent même fort peu de choses, sans cesser cependant d'être un message fort ». Donc, selon BARTHES, toute représentation graphique implique des choix. C'est le cas, *a fortiori*, de la datavisualisation. En effet, celle-ci est par essence incomplète, puisque des choix sont nécessaires pour la réaliser. Ces choix dépendent du contexte.

BESSE (2006) explique très précisément ce phénomène dans le cas de la cartographie (qui, selon notre définition basée sur les travaux BERTIN, appartient au domaine de la datavisualisation). En posant la question de l'exactitude des cartes, il constate :

« La sélection des informations et la schématisation des représentations sont intrinsèques à la réalisation de la carte (...) Toute carte propose une version ou une interprétation de la réalité territoriale à laquelle elle réfère, en fonction des intentions qui sont déployées vis-à-vis de cette réalité. »

Il en conclut que, pour étudier une carte, il est nécessaire de la replacer dans son contexte, puisque « la carte porte l'empreinte d'une intention, d'un intérêt, d'origine économique, politique, idéologique, voire même scientifique, qui vont orienter la sélection des informations puis leur présentation cartographique ». En effet, la carte est influencée par la culture dans laquelle elle est créée, et dont « elle traduit (...) les valeurs » par ses choix des éléments qu'elle présente, mais aussi par ses prises de position graphiques : « les couleurs, les symboles, les lignes, les noms, les systèmes de projection, mais aussi le cadre, l'échelle, le format, ou les éléments décoratifs » ne sont pas neutres, et véhiculent des valeurs (BESSE 2006). Par exemple, il y a bien une différence dans la façon de concevoir le monde entre une projection Mercator du monde (qui place l'Europe au centre), et une projection Cassini (dans laquelle les continents sont représentés selon leur position autour des pôles).

Mais si la carte est dépendante d'un contexte, elle est aussi un objet de pouvoir dans le sens où elle est susceptible « comme beaucoup d'autres images, d'orienter les décisions de la société et de la culture, tout simplement parce qu'elle met en forme la vision du monde sur laquelle ces décisions prennent appui ». La carte devient un outil de communication et d'influence (BESSE 2006).

Bien entendu, toutes ces remarques qui sont vraies pour les cartes le sont aussi pour les autres types de datavisualisations (réseaux et diagrammes, selon BERTIN), qui nécessitent tous une sélection des données puis des choix de représentation. Si la *Sémiologie graphique* de BERTIN nous donne une méthode pour réaliser des datavisualisations rigoureusement objectives car

obéissant à une série de règles si stricte qu'elles laissent peu de marge à d'autres formes d'expressions culturelles, les choix du designer désireux de faire passer un message vont être influencés par des codes sociaux, voire vont être pensés pour influencer la perception des récepteurs.

Citons, par exemple, la courbe des ventes cumulées d'iPhone qu'Apple présente à la presse en 2014. Les données sont sans doute exactes, mais leur choix n'est pas neutre : au lieu de représenter les ventes par année, Apple a choisi de représenter les ventes *cumulées*, ce qui implique nécessairement une courbe croissante, puisque chaque année présente les ventes des années précédentes ajoutées à celle de l'année en cours. Par ses choix stratégiques de représentation, Apple cherche à donner l'image d'une entreprise en pleine croissance. Cet exemple est emblématique de l'emploi de nombreuses datavisualisations pour faire du *storytelling*. Les récepteurs de ces datavisualisations doivent donc être en mesure de conserver une vision critique (CANOPÉ 2019).

Il ne faut cependant pas généraliser ces pratiques : les datavisualisations à visée scientifique et celles de la presse conservent un idéal d'objectivité, qui ne peut cependant jamais être atteint car, comme dit précédemment, les choix de données comme de représentation graphique ne sont pas neutres. Pour autant, bien que les datavisualisations présentent seulement une vision de la réalité, elles n'en demeurent pas moins efficaces heuristiquement tant qu'elles sont « susceptible[e] de rendre service dans la perspective du but que l'on s'est fixé ». Autrement dit, même si le contexte de la production d'une représentation graphique doit être conservé à l'esprit, celle-ci peut être intéressante (voire indispensable) pour comprendre certains phénomènes.

Résumé — La notion de littératie appliquée aux datavisualisations

Nous avons établi que l'écriture est un « medium de communication entre les hommes », consistant en un « ensemble de signes visibles » tracés sur un support qui permet une transmission de l'information « dans l'espace et (...) dans le temps » (GOODY et LEJOSNE 2006). Aussi, nous pouvons admettre que les datavisualisations, si elles reposent sur un code monosémique (comme le pense BERTIN), peuvent être assimilées à une forme d'écriture.

Elles relèvent donc de la littératie générale, définie comme « l'ensemble des activités humaines qui impliquent l'usage de l'écriture, en réception et en production » (DROT-DELANGE 2014). Autrement dit, toute personne qui possède les compétences nécessaires pour lire et écrire maîtrise la littératie générale.

Nous avons cependant été amenés à ajouter une autre dimension à la littératie des datavisualisations : une capacité à porter un regard critique sur l'image produite. Pour ajouter cette idée à notre concept, nous avons dû recourir aux notions de littératie informatique (un ensemble de connaissances et compétences à acquérir pour développer cette approche critique de l'informatique et de ses contenus, sans nécessité de maîtriser la programmation), et de littératie des données (« capacité des individus [à] accéder, d'interpréter, d'évaluer de manière critique, de gérer, de manipuler et d'utiliser éthiquement les données » selon ARRUBARRENA, KEMBELLEC et CHARTRON).

Avec tous ces éléments, nous en venons à placer la littératie des datavisualisation au croisement de trois autres littératies :

- La littératie générale, dans la mesure où lire et tracer une datavisualisation nécessite des phases d’encodage et de décodage pour donner du sens aux données, que l’on transforme alors en informations significatives ;
- La littératie informatique, qui consiste à savoir utiliser l’interface de lecture, mais également à maîtriser la pensée informatique et donc à comprendre le cheminement des algorithmes (même si cette compréhension reste basique, schématique) ;
- La littératie des données, qui nécessite une culture de la donnée, composée de connaissance sur la fabrication des données, leur extraction, leur parcours jalonné de transformations et leur écosystème général (acteurs, réglementations, éthique, etc.).

Au terme de cet état de l’art, nous avons donc repéré des concepts littératiens qui pourraient être appliqués aux datavisualisations. Nous pourrions donc estimer avoir répondu à notre problématique initiale :

Quelle(s) littératie(s) sont mobilisée(s) par les datavisualisations médiatiques à destination du grand public ?

Cependant, nous avons formulé une hypothèse lorsque nous avons établi le lien entre littératie et datavisualisations : que les datavisualisations soient réalisées sur le modèle proposé par BERTIN. Il nous reste donc à savoir si les datavisualisations médiatiques à destination du grand public sont construites sur cette base.

Analyse

CHAPITRE 4

MÉTHODOLOGIE

Afin de lier les concepts de datavisualisations et de littératie, nous avons décidé de considérer les datavisualisations à destination du grand public. En effet, étudier la littératie d'un groupe d'experts n'aurait aucun sens : il n'était donc pas possible d'interroger, par exemple, des chercheurs, ni des étudiants en master ou en doctorat dans le supérieur, puisqu'ils ont probablement reçu une formation spécifique concernant les représentations graphiques de données, ne serait-ce que pour présenter les résultats de leurs recherches. Il est donc impossible de se limiter à un groupe de personnes homogène, d'autant plus si celui-ci s'est déjà intéressé (même de loin) au sujet de l'enquête. Cela pourrait biaiser les résultats. Dès lors que nous avons établi que les datavisualisations qui nous intéressent ne sont pas destinées à des experts, il faut nous tourner vers celles qui sont communiquées au grand public.

Nous employons l'expression « grand public » pour signifier que nous cherchons à étudier un public varié et « sans spécificité particulière » (SPIES 2021), s'opposant au « public spécifique » (LECOLLE 2021). Émerge ici la notion de « public », qu'il nous faut définir. Pour cela, nous ferons appel à la définition qui en est donnée par CITTON dans le « Lexique » de son ouvrage *Lire, interpréter, actualiser. Pourquoi les études littéraires ?* : un « public » correspond à une « collection d'individus apparemment autonomes et indépendants, qui ne se connaissent et ne se voient pas (contrairement à une foule), mais qui tendent malgré cette séparation spatiale à penser et à agir de la même façon, parce qu'ils “se retrouvent” circulairement dans les médias qui informent leur sensibilité et leur idéologie, selon une logique qui relève des lois du marché davantage que d'un contrôle politique »¹. Bien que cette définition soit très pertinente dans le contexte de l'économie de l'attention étudiée par CITTON, nous la nuancerons pour l'adapter à notre étude. En effet, si nous présumons que l'ensemble d'un public « pens[e] et (...) agi[t] de la même façon », alors nous pouvons penser qu'ils comprennent les graphiques de la même façon, et que la littératie des datavisualisations est homogène chez tous les destinataires privilégiés de chaque média. Nous

1. Référence citée par WIKIPEDIA. *Public*. fr. Page Version ID : 182695594. Mai 2021. URL : <https://fr.wikipedia.org/w/index.php?title=Public&oldid=182695594> (visité le 22/08/2021)

verrons si cela est vrai ou non par la suite, mais il s'agit plutôt d'une conclusion issue des observations de CITTON que d'une définition générale. Aussi nous utiliserons le *Publictionnaire* du Centre de recherche sur les médiations pour reformuler ce concept de façon plus neutre : nous dirons alors qu'un public est une « collection d'individus apparemment autonomes et indépendants, qui ne se connaissent et ne se voient pas (contrairement à une foule) » (CITTON 2007) « rassemblé[e] conjonctuellement pour un but ou par un goût commun » (LECOLLE 2021).

Le point de rassemblement d'un public est souvent la consultation d'un média commun, comme l'évoque CITTON. Dans ce cas, le « grand public » désigne l'audience d'un média qui s'adresse à toute catégorie d'individus, sans distinction de classe sociale, opinion politique, etc. (SPIES 2021). Il ne s'agit pas d'un public populaire (au sens de « issu de catégories populaires »), mais d'une façon de désigner les destinataires de médias qui se veulent universels, adressés à tous, sans exiger du récepteur la moindre spécificité (SPIES 2021). SPIES précise que, bien que cette notion soit « paradoxale » puisque le grand public est une « forme invisible », elle demeure « nécessaire pour penser les pratiques culturelles ».

Ainsi, dans le cadre d'une étude sur la littératie, nous nous intéressons au « grand public ». Mais avec quelle méthode ? Pour pouvoir appréhender la littératie des datavisualisations chez le grand public, il faudrait interroger de nombreuses personnes aux profils variés, sans attrait particulier pour la datavisualisation. Hors il aurait été difficile d'obtenir des réponses de personnes non concernées par le sujet. Dans le cas d'une étude quantitative, l'effectif de répondants aurait été trop faible pour donner des résultats concluants. De plus, il aurait fallu constituer un panel comprenant des individus de diverses catégories sociales, générations, professions, etc. L'analyse quantitative ne paraissait donc pas réalisable avec les contraintes temporelles imposées. Quant à une étude qualitative, cette méthode n'aurait pas permis d'interroger suffisamment de gens pour tirer la moindre conclusion. Le choix des personnes à interviewer aurait également posé question : qui choisir pour être représentatif du grand public ? Par ailleurs, il n'aurait pas été très éthique de questionner des gens concernant leur compréhension de graphiques : ils auraient probablement vécu cela comme une évaluation.

En raison de toutes ces considérations, nous avons donc opté pour une méthode d'étude indirecte : en faisant une analyse sémiotique de datavisualisations journalistiques auxquelles a accès le grand public, nous verrons en détails quels éléments des représentations graphiques de données nécessitent une forme quelconque de littératie. Nous pourrions ainsi répondre à la problématique suivante :

Quelle(s) littératie(s) sont mobilisée(s) par les datavisualisations médiatiques informatisées à destination du grand public ?

Puisque nous nous focaliserons sur les datavisualisations journalistiques, nous devons dans un premier temps nous renseigner sur leur contexte de production, et ainsi développer notre propre littératie des données en suivant leur parcours au sein des rédactions de presse. Pour cela, nous nous appuierons sur des articles scientifiques ainsi que sur deux entretiens exploratoires

menés avec des datajournalistes français, un du *Parisien* et l'autre du *Monde diplomatique*.

Une fois ce contexte établi, nous étudierons un corpus de datavisualisations journalistiques. Cette partie constituant le point central de notre analyse, nous devons maîtriser la sémiotique et l'analyse d'un corpus de presse numérique. Nous allons à présent décrire les méthodologies que nous utiliserons dans la suite de ce mémoire.

4.1 Analyse sémiologique

Pour commencer, définissons les mots « sémiotique », qui « désigne l'élaboration de concepts portant sur la production du sens » tandis que le terme « sémiologie » « renvoie à l'analyse concrète d'objets particuliers » (JEANNERET 2019). Notre analyse de corpus consistera donc à identifier des « signes [que nous interpréterons en tant que] phénomènes » communicationnels qui influencent notre « perception » (JEANNERET 2019).

Concernant les théories sur lesquelles nous nous appuyerons, nous nous placerons dans le cadre établi par notre état de l'art. En effet, nous y avons déjà abordé la question de l'image et de sa polysémie, de l'interprétation des signes (avec la distinction entre le signifiant, c'est-à-dire le signe que l'on voit, et le signifié, ou l'interprétation que l'on retire du signifiant). Nous avons évoqué les messages portés par les images, celles-ci associant une « image littérale dénotée » (correspondant à l'identification de la scène présentée, sans interprétation) et une « image symbolique connotée » (c'est-à-dire la perception que l'on retire de l'image par des associations d'idées culturellement ancrées) (BARTHES 1964). Nous nous référerons à la partie 1.2 pour plus de détails.

D'autre part, étudiant des datavisualisations, nous devons nous référer à d'autres sources qui abordent la question des images monosémiques. Pour cela, nous nous baserons principalement sur BERTIN (dont nous retrouverons les idées en partie 2). En effet, bien qu'il soit moins connu que BARTHES et moins souvent cité dans des analyses sémiotiques, il est le fondateur du courant de la sémiologie graphique (courant qui a d'ailleurs pris le nom de son propre livre).

De plus, comme l'explique FEKETE (2015), les enseignements de BERTIN (qui reposait à l'époque de leur formalisation sur des intuitions et des observations empiriques) ont été depuis corroborées par la psychologie cognitive. En effet, il évoque la psychologue américaine TRIESMAN qui, en 1985, aurait « mis en évidence la raison d'une visualisation efficace. » En particulier, cette scientifique a identifié des variables visuelles « pré-attentives », c'est-à-dire « perçues par un œil humain très rapidement [soit en moins de 0,2 secondes] inconsciemment et sans effort », ce qui permet de « raisonner sur une grande quantité de données très rapidement et pratiquement indépendamment du nombre de ces données » (FEKETE 2015). Ces caractéristiques pré-attentives correspondent à plusieurs des variables visuelles identifiées par BERTIN, ainsi qu'à leur impact sur la perception. Par exemple, la couleur permet de marquer la ressemblance, mais surtout la différence puisque : « en regardant des points dessinés sur une feuille de papier ou un écran, il est très facile, lorsque la plupart des points sont bleus, de voir un ou deux points rouges. Que l'on regarde 10, 100 ou 10 000 points

bleus, la présence d'un point rouge "saute aux yeux" ». De même, la forme (assimilée par la psychologue à son rayon de courbure) permet de symboliser la ressemblance entre des éléments (FEKETE 2015). Il a également été montré que seules un certains nombres de caractéristiques pré-attentives peuvent être gérées par la perception humaine concomitamment (sept au maximum). Nous n'entretons pas dans les détails de l'étude de la cognition humaine (qui dépasse nos compétences), mais ces éléments tendent à montrer que la sémiologie graphique de BERTIN est un outil valable pour une analyse.

4.2 Approche sémiotique dans l'étude des médias

Étudiant un corpus de datavisualisations de presse numérique, il nous faut savoir comment appliquer les principes de l'étude sémiotique aux médias. Nous nous appuyons pour cela sur l'ouvrage *Médias et médiatisation : Analyser les médias imprimés, audiovisuels, numériques* publié en 2019.

Commençons par délimiter le champ que recouvre les « médias ». En effet, ce mot est très employé en SIC, et peut désigner plusieurs éléments. Comme LAFON, nous estimerons que « [l]es médias (ou dispositifs médiatiques) (...) compren[nent] les médias institués de manière multiséculaire comme la presse, de manière séculaire comme la radio et la télévision, ou de manière récente comme les médias natifs du Web (presse numérique, médias audiovisuels en ligne tel Netflix) ou les médias sociaux (médias diffusant *via* Facebook par exemple). »

Aussi le corpus d'articles de presse numérique incluant des datavisualisations est-il bien inscrit dans le champ des médias. Cependant, JEANNERET (2019) précise qu'en sémiotique, le concept de média doit être re-problématisé. En effet, pour pouvoir étudier les signes, signifiants et signifiés, il faut comprendre que ceux-ci s'inscrivent dans un dispositif de communication². Ce que nous nommerons « médias » dans la suite de ce mémoire correspondra à des « dispositifs médiatiques », incluant des personnes, des aspects techniques incluant des algorithmes et des logiciels, des pratiques ainsi qu'une histoire et des spécificités culturelles propres au milieu du journalisme. Nous reviendrons sur tous ces aspects en partie 5.

Dans notre analyse, nous essaierons de donner du sens aux « création[s] des hommes » (JEANNERET 2019) que sont les datavisualisations qui composeront notre corpus. Nous nous appuyons sur des considérations culturelles, principalement, que nous détaillerons au cas par cas. Comme le conseille JEANNERET (2019), nous n'avons pas réalisé de grille d'analyse (trop cloisonnante), mais nous avons suivi un protocole qui a consisté à :

- Observer généralement le corpus (comme le suggère JEANNERET) pour identifier des points communs et des éléments de divergences entre les différents médias. En particulier, nous avons prêté attention dans cette phase aux « formes globales », sans nous appesantir sur les

2. La notion de dispositif étant entendue ici au sens de FOUCAULT : « un ensemble résolument hétérogène comportant des discours, des institutions, des aménagements architecturaux, des décisions réglementaires, des lois, des mesures administratives, des énoncés scientifiques, des propositions philosophiques, morales, philanthropiques ». Ce concept sera adapté aux médias en tant « [qu']artefact industriellement produit qui configure les conditions matérielles de la communication »

détails (JEANNERET 2019). Afin d'identifier des questionnements qui puissent justifier nos observations, nous nous sommes appuyés sur la connaissance de l'écosystème des datavisualisations journalistiques, qui nous a permis de repérer des points clés dans l'écosystème du journalisme.

- Consulter d'autres datavisualisations de chacun des médias afin de corroborer ou non ces premières impressions.
- Développer chacune des thématiques observées sur chacun des médias en « inventoriant (...) systématiquement certaines catégories de signes » (JEANNERET 2019).
- Établir un lien entre la problématique et les observations effectuées.

Après avoir suivi les deux premières étapes, nous avons décidé de nous interroger sur la place des datavisualisations dans les médias informatisés, leur rôle dans l'économie de l'attention, la place de la narration et de l'interaction dans l'énonciation, la question de l'objectivation d'un article de presse par des données chiffrées, ainsi que l'origine desdites données. Nous tenterons d'identifier le positionnement laissé au récepteur par le journaliste (ou le datajournaliste), son degré de liberté dans l'analyse, la perception qu'il peut avoir des datavisualisations journalistiques étudiées, pour finir par nous demander quelle forme de littérature est convoquée par le corpus choisi.

4.3 Choix du corpus

Notre corpus a été constitué en choisissant des journaux à destination du grand public, qui ont tous sur leur site internet une rubrique « infographie », « datavisualisation » ou tout autre terme montrant qu'ils utilisent cette pratique régulièrement. Afin de pouvoir comparer les articles présentant des datavisualisations entre eux, nous avons choisi d'étudier des tableaux de bord de l'évolution de la pandémie de COVID-19, qui ont toutes été construites à partir de jeux de données similaires. Afin d'éclairer certains points qui ne seraient pas assez saillants sur ces tableaux de bord assez standardisés et soumis à des contraintes temporelles fortes (ils ont dû être créés en toute hâte au début de la pandémie), nous avons décidé de nous référer également à d'autres datavisualisations produites par les mêmes journaux, mais sur le plus long terme et donc plus fidèles aux visions des datajournalistes.

Considérant, comme GUICHARD (2019) que les nouvelles pratiques culturelles d'écritures proviennent prioritairement des États-Unis, et ayant remarqué lors de nos recherches d'informations que la plupart des articles scientifiques américains étaient beaucoup plus nombreux que les articles francophones sur la datavisualisation, nous étudierons deux médias des États-Unis (le *Washington Post* et le *New York Times*). Nous partirons du principe qu'ils évoluent dans un milieu culturel qui est déjà plus en pointe sur la datavisualisation. En effet, lorsque l'on entre la requête « data visualisation » sur Google scholar, on obtient 17 résultats en langue française sur 1 183 en langue anglaise. De plus, parmi ces résultats, seuls 7 articles français sont antérieurs à 2015, contre 668 déjà à cette période en anglais. Bien que cette estimation puisse être contestée (l'orthographe du terme peut varier, le mot employé en français pour désigner le même concept n'était peut-être pas le même il y a dix ans, des articles anglophones peuvent provenir

d'autres pays que les États-Unis, etc.), la proportion est si écrasante que nous émettrons l'hypothèse que les médias américains sont plus en pointe que les médias français sur ce sujet, et que l'avenir des datavisualisations correspondra probablement aux visions que l'on en a Outre-Atlantique.

Tout au long de notre analyse, nous garderons à l'esprit que nous ne faisons que « propose[r] des interprétations », et que nous ne pouvons prétendre à la découverte d'une vérité absolue (JEANNERET 2019). une

CHAPITRE 5

CONTEXTE DE PRODUCTION DES DATAVISUALISATIONS MÉDIATIQUES NUMÉRIQUES

Pour pouvoir réaliser une étude sémiotique de datavisualisations journalistiques, il faut avoir quelques renseignements sur le contexte de production de ces objets. En effet, il nous faut maîtriser la littérature des données du journalisme pour pouvoir ensuite les commenter. De plus, il nous faut des informations concernant la situation d'énonciation propre aux médias. Pour cela, nous allons nous appuyer sur des articles scientifiques ainsi que sur les réponses que nous avons obtenu lors de deux entretiens exploratoires avec des datajournalistes dans des rédactions françaises : *Le Parisien* et *Le Monde diplomatique*. La grille ayant servi à ces entretiens semi-directifs sont présentées en annexe 6.2.2.

5.1 Compte-rendu des deux entretiens exploratoires

Les entretiens ont duré 50 minutes pour le premier et 1 h 05 pour le second. Il en est ressorti que les deux personnes interrogées avaient des profils et des pratiques très différents.

Le premier interviewé (un homme d'une quarantaine d'années) a un parcours professionnel déjà long : il a commencé par travailler à l'exploitation graphique de données en 2006 dans le cadre d'une mission pour l'ONU, puis a continué son parcours professionnel en mettant à profit ses compétences techniques (informatique, traitement statistique de données) jusqu'à être embauché dans la presse. Il se considère comme un journaliste de données, sur le même plan que les autres professionnels de son journal, avec lesquels il travaille beaucoup pour produire conjointement des articles. Il maîtrise parfaitement la chaîne de la donnée, connaît de nombreux outils de datavisualisation, et se tient informé dans son domaine d'activité par la lecture de livres scientifiques (notamment *How charts lie* de CAIRO). Il considère qu'un graphique « classique » est bien plus évident à comprendre pour ses lecteurs (il précise qu'il veut s'adresser à tous, au grand public) et qu'il vaut mieux ne pas se lancer dans des innovations graphiques que les usagers ne comprennent pas. Il stipule qu'une étude a d'ailleurs été menée par son

quotidien, auprès d'un panel représentatif de son lectorat, qui a montré que ses récepteurs apprécient de trouver l'information présentée sur un graphique simple pour sa rapidité de lecture, mais sont découragés par des formats atypiques. Il considère qu'il faut communiquer en utilisant des codes communs à tout son public. Il prend l'exemple d'une expérimentation réalisée par son équipe : réaliser une datavisualisation de réseau (pour mettre en évidence les liens entre des personnes impliquées dans des attentat). Le projet est vite devenu tentaculaire, et ne parlait pas à au public de son journal. Aussi l'équipe a-t-elle décidé de modifier la structure du projet pour s'orienter vers un format « réseaux sociaux », avec des profils accessibles par recommandation. Il valorise la clarté, la simplicité, et les codes graphiques classiques. Il revendique de faire des choix dans ses représentations de données afin de ne pas surcharger celles-ci.

La seconde personne interrogée est une femme d'une vingtaine d'années (entre 20 et 30). Après des études de lettres puis une licence d'urbanisme, elle est en reconversion en étudiant le développement interactif. Elle en est alternance dans un média pour lequel elle réalise des cartes interactives. Contrairement au premier interviewé, elle aborde très vite la notion de design. Elle privilégie les formats esthétiques et souvent novateurs. Elle suit d'ailleurs des médias qui proposent ce genre de contenus sur les réseaux sociaux et *via* des newsletter. Elle maîtrise de nombreux logiciels graphiques (suite Adobe par exemple) et a appris en autodidacte à utiliser des bibliothèques javascript pour améliorer ses datavisualisations. Elle valorise l'exhaustivité de l'information, quoiqu'elle reconnaisse le besoin supérieur de lisibilité. Elle conseille l'interactivité pour solutionner ce problème, et notamment les filtres. Elle ne maîtrise pas encore pleinement la chaîne de la donnée, car elle reste en apprentissage, et mentionne le fait de n'avoir jamais « nettoyé » de jeu de données. Concernant l'interaction, elle espère que son design permet à un individu qui possède les codes informatiques saura explorer. Elle pense néanmoins également à apporter quelques indications concernant les interactions essentielles sur la visualisation. Malgré toutes ses compétences informatiques, elle estime tout autant les datavisualisations en ligne que papier.

Nous sommes donc confrontés à deux profils très différents. Le premier privilégie les graphiques de données classiques, avec un code fort, et suit ainsi la voie de BERTIN ; tandis que la seconde s'oriente plutôt vers une approche designée des datavisualisations. Dans les deux cas, quelques points communs émergent :

- Ils considèrent tous les deux que les datavisualisations peuvent être papier ou numérique, et ne sont pas nécessairement interactives.
- Ils voient leurs représentations graphiques de données comme destinées à tous (au grand public) et pensent que l'image attire le lecteur.
- Ils sont peu loquaces sur les sources de leurs données. Leur travail semble commencer une fois le jeu de données entre les mains. Il restent flous quand aux problèmes de production des données, quoiqu'ils semblent se diriger plus volontiers vers des sources institutionnelles autant que possible.

5.2 Le datajournalisme

5.2.1 Une évolution du métier de journaliste

Revenons aux fondations du journalisme : tout journal a pour objectif principal de diffuser des informations auprès d'un public. C'est déjà sur ce principe qu'est fondé le tout premier journal français : *La Gazette* de RENAUDOT créé en 1631 (HARMAND 2020).

Par la suite, la question de l'objectivité de cette information fait débat : en effet, au XIX^e siècle, la presse est ouvertement politisée et revendique ses contenus orientés (HARMAND 2020). Cette tendance trouve ses limites lorsqu'en 1826, une nouvelle taxe oblige les responsables éditoriaux à repenser leur modèle économique. Ils sont alors financés par la publicité, ce qui leur impose de conquérir un public fidèle et nombreux pour attirer les annonceurs, et également de satisfaire ces derniers (HARMAND 2020). Ils doivent donc suivre une ligne éditoriale non polémique.

Comme l'explique HARMAND, « deux modèles économiques et visions du journalisme » coexistent alors : quand certains (comme le journal *La Presse* de GIRARDIN) « prône[nt] une information rigoureusement neutre — sans se cacher d'assurer des débouchés commerciaux à l'industrie », d'autres au contraire promeuvent leurs opinions politiques sans publicité et suscitent le débat, à l'instar du *National* de THIERS et CARREL. Dans le premier cas, la presse « s'efface devant "l'opinion populaire" en donnant aux foules ce qu'elles demandent pour ne pas perdre leur lectorat », alors que dans le second, les journaux sont « direct[eurs] d'opinion » (HARMAND 2020). Cette dichotomie entre les deux modèles nés au XIX^e siècle se retrouve dans la suite toujours bien vivace dans la culture journalistique.

Au XX^e siècle, les journalistes se professionnalisent sous l'impulsion de LIPPMANN, et constituent une élite médiatique encouragée à « exposer le monde tel qu'il est ». Ils aspirent alors à l'objectivité. Pourtant, malgré tous les efforts de la profession pour atteindre l'impartialité, ils ne parviennent pas à convaincre les lecteurs qui exigent encore plus d'objectivité. La profession « atteint la fin du XX^e siècle en pleine "crise de confiance et de légitimité" » (HARMAND 2020).

Dans ce contexte, l'analyse de données quantitatives est une façon d'objectiver l'information. LIPPMANN recommande d'ailleurs cette pratique pour aider à la décision les « hommes d'action », entendus comme des membres d'une élite politique ou industrielle (HARMAND 2020). À partir des années 1960, certains journalistes américains (en particulier MEYER) emploient des méthodes quantitatives et proposent des schémas graphiques (THEVIOT 2020).

En effet, l'exploitation de valeurs chiffrées tend à faire paraître les pratiques journalistiques comme impartiales. THEVIOT (2020) exprime cela :

« Le journalisme de données permettrait *a priori* de se détacher de la critique récurrente du manque d'objectivité ou de neutralité du journaliste. »

Cependant, l'auteur nuance cette impression : « ce présupposé d'une donnée "pure", "objective" » est un mythe, puisque « les données sont,

elles-aussi travaillées, choisies, par les institutions qui les communiquent » (THEVIOT 2020).

Outre cette objectivation du travail journalistique par les chiffres, l'utilisation de bases de données massives est censée « maximiser les sources pour redonner confiance dans le travail des journalistes » (THEVIOT 2020). En effet, la quantité de données devrait être plus représentative d'une population que des enquêtes journalistiques ne permettant d'interroger que peu de personnes faute de temps. « La profusion de données disponibles [actuellement], liées notamment au mouvement d'ouverture des données » (THEVIOT 2020) favorise cette « évolution du métier de journaliste » (HARMAND 2020).

5.2.2 La situation actuelle

Aujourd'hui, l'utilisation de données chiffrées est devenu une pratique courante dans environ la moitié des rédactions journalistiques. Pour vérifier cela, nous nous appuyons sur une étude réalisée en 2017 pour le compte du Google news lab, reposant sur des entretiens qualitatifs avec 56 journalistes américains, français, britanniques, et allemands, complétés par une enquête quantitative auprès de 900 employés du secteur médiatique dans ces quatre pays. Les médias enquêtés sont aussi bien de la presse papier que des chaînes d'information télévisuelles, des journaux *pure player* (c'est-à-dire exclusivement en ligne) ou encore des journaux cumulant plusieurs de ces modalités. D'après cette étude, la moitié des équipes de rédactions enquêtées utilisent régulièrement des données dans le cadre de leur travail (BOWERS, SCHWABISH et ROGERS 2017). L'enquête lie ce phénomène de journalisme des données avec le Big data (puisque de grandes quantités de données sont aujourd'hui produites) et le mouvement de l'Open data (qui rend ces jeux de données accessibles). De plus, les progrès des technologies informatiques offrent également des outils puissants pour analyser ces données, puis pour les communiquer (BOWERS, SCHWABISH et ROGERS 2017).

Pourtant, ces nouveaux terrains d'investigation ne séduisent pas tous les journalistes (CHAGNOUX 2020). THEVIOT (2020) et CHAGNOUX (2020) note d'ailleurs qu'en France, cette pratique est implantée dans la presse nationale et parisienne, mais que les quotidiens locaux peinent à se laisser convaincre. Si les journaux provinciaux ont « conscience de l'importance du datajournalisme pour le futur de la presse locale », des difficultés de mise en œuvre apparaissent (CHAGNOUX 2020). En effet, les reporters habitués à des enquêtes de terrain qualitatives, basées principalement sur des entretiens, envisagent cette « numérisation du journalisme » plutôt comme une contrainte chronophage les obligeant à maîtriser des technologies informatiques poussées (THEVIOT 2020).

L'étude de données implique des manipulations techniques complexes qui nécessitent « une solide formation en (...) en statistiques et informatique, en particulier en programmation » (CHAGNOUX 2020). CHAGNOUX estime que ces bases théoriques ne sont pas assez enseignées « dans les formations initiales des journalistes et reste difficilement accessibles en autoformation à des profils non scientifiques » qui constituent la grande majorité de la profession (CHAGNOUX 2020). Ainsi, environ la moitié des répondants à l'enquête de BOWERS, SCHWABISH et ROGERS envisagent l'analyse de

données comme une compétence spécialisée, et non commune à tous les journalistes. La mise en forme graphique de ces données pour en faire la matière première d'un article serait une spécialisation à part entière selon quatre enquêtés sur cinq (BOWERS, SCHWABISH et ROGERS 2017). Enfin, même en maîtrisant les compétences techniques indispensables, les enquêtes de BOWERS, SCHWABISH et ROGERS et de CHAGNOUX suggèrent que les journalistes ne disposent pas d'assez de temps pour produire des résultats satisfaisants à leurs yeux : une journée seulement est consacrée à la création de la majorité des datavisualisations de presse (BOWERS, SCHWABISH et ROGERS 2017).

Cette nouvelle méthode pour faire émerger et transmettre l'information peut donc être vécue par certains comme une contrainte. Par ailleurs, son intérêt n'est pas évident aux yeux de tous les journalistes. En particulier, les employés de longue date de journaux locaux défendent leurs méthodes qui leur semblent plus nuancées et qualitatives, moins soumises à des impératifs de communication et de rentabilité. Ils voient parfois le journalisme de données comme déshumanisant, faisant intervenir des robots, et amenant à faire des raccourcis et à produire des visualisations simplificatrices (THEVIOT 2020). Cette opinion assimilant l'avenir du journalisme à des productions robotisées sans médiateur humain est étayée par les récentes « expériences d'algorithmes d'écriture automatique agencant des données en formats narratifs [tel que] le “Quakebot” du *Los Angeles Times* [qui] rédige des alertes aux tremblements de terre depuis 2014 » (HARMAND 2020).

« Le journalisme à l'ancienne, “artisanal”, se voit ainsi opposé au journalisme dit moderne, ou “augmenté”. »

THEVIOT 2020

Aussi le journalisme de données est-il une pratique émergente non encore adoptée par tous les journalistes. Nous remarquons d'ailleurs que méthodes de production, l'organisation interne des rédactions sur ces questions, et même les canaux de diffusions sont variables d'un journal à un autre.

Même la définition du terme « datajournalisme » fait débat. S'agit-il d'enrichir un article classique en s'appuyant sur des chiffres pour prouver ses dires (pratique nommée « enrichissement par les données » par BOWERS, SCHWABISH et ROGERS)? Ou bien parle-t-on de recherche des informations directement issues des données (ce qui correspondrait, selon les termes de BOWERS, SCHWABISH et ROGERS, à du « datajournalisme d'investigation »)? Il existe même une troisième voie : illustrer des données chiffrées (que l'on accepte alors comme un constat, et qui ne font alors pas débat) par des enquêtes journalistiques de terrain, pour raconter l'histoire vécue derrière les chiffres. Il s'agit alors d'offrir plus de sens aux lecteurs que de simples valeurs numériques ne peuvent en avoir (BOWERS, SCHWABISH et ROGERS 2017).

Le journalisme de données n'est donc pas encore une pratique uniforme. S'il est bien implanté dans certains médias à portée nationale, son intégration dans les rédactions locales pose plus de problèmes. En effet, le défaut de compétences techniques demeure un obstacle. Dès lors, il semble que cette pratique devienne une spécialisation dédiée à des personnes spécifiques : les datajournalistes. Ce nouveau métier peut cependant être décliné de différentes manières, consistant parfois à extraire et travailler des données pour illustrer un article, ou au contraire se servant de données comme base d'une information.

Dans le cadre de ce mémoire, nous nous focaliserons exclusivement sur les pratiques journalistiques impliquant la production de datavisualisations.

5.2.3 Les datavisualisations journalistiques

Si certains continuent à contester le journalisme de données, en particulier quand il consiste à produire une visualisation et non un texte, l'intérêt d'utiliser des données pour enrichir et objectiver les articles de presse semble aujourd'hui pris en compte dans de plus en plus de rédactions de journaux. De plus, CHAGNOUX note que l'interactivité de certaines datavisualisations apporte une apparence d'objectivité en faisant croire au lecteur qu'il peut manipuler les données pour en tirer ses propres conclusions (CHAGNOUX 2020).

Outre cette objectivation de l'information par les chiffres, les datavisualisations journalistiques, en tant qu'images, présentent un attrait pour les lecteurs/observateurs dans le cadre de « l'économie l'attention » (DYMYROVA 2018). En effet, aujourd'hui, « les formats web et (...) l'image joue[nt] un rôle prépondérant dans le partage d'informations » en particulier « sur les réseaux sociaux ». Ces images sont censées être « engageante[s] » pour le récepteur (CHAGNOUX 2020). La communication avec le grand public est donc favorisée par ce type de production, qui capte l'attention. C'est pourquoi les journaux investissent dans la création de datavisualisations : « ret[enir] l'attention (...) fait vendre » (THEVIOT 2020).

L'information communiquée est censée être plus facile à appréhender pour le grand public grâce aux représentations graphiques. En effet, selon CHAGNOUX (2020) la datavisualisation a une « dimension pédagogique » qui correspond à l'ambition du journaliste à être un médiateur, « un passeur d'informations » selon THEVIOT (2020). Cette dernière souligne d'ailleurs que le *Guide du datajournalisme*¹ professe explicitement cette volonté de « vulgarisation » :

« En utilisant des données, il est possible de transformer une abstraction en quelque chose qui parle à tout le monde. »

THEVIOT résume tout cela :

« La mission pédagogique du journaliste se transforme alors en mission de communiquant qui vise à vulgariser pour mieux faire passer une information au plus grand nombre afin de capter un plus large public. »

THEVIOT 2020

Ainsi, la datavisualisation convainc de plus en plus de journaux (en particulier pour un usage en ligne) parce qu'elle permet, en principe, un « gain d'intelligibilité » : comme toute autre forme de journalisme, elle tend à « informer utilement le public ». De plus, elle présente une « esthétique [attractive et] légitimante » en bénéficiant de l'apparence de l'objectivité conférée par les données quantitatives (HARMAND 2020).

1. Jonathan GRAY et al. *Guide du datajournalisme. Collecter, analyser et visualiser les données*. Eyrolles, 2013. 220 p. ISBN : 9782212136852. URL : <https://jplusplus.github.io/guide-du-datajournalisme/index.html>.

5.3 Processus de réalisation des datavisualisations journalistiques

Comprendre comment sont créées les datavisualisations journalistiques nous permettra de mettre en évidence les éventuels biais de ce processus. En effet, on sait déjà que le journalisme induit des biais dans les informations qu’il communique « imputés tant à la nature humaine du journaliste qu’à son environnement complexe » (HARMAND 2020). L’information est sélectionnée en fonction de sa disponibilité, de préconceptions sociales, de contraintes budgétaires, de logiques de divertissement parfois... L’information est donc choisie, construite et présentée d’une façon déterminée : en ce sens, on peut dire qu’elle est designée (HARMAND 2020). Par ailleurs, comme le montre admirablement CAIRO (2020), les datavisualisations sont aussi des constructions que l’on peut faire mentir. En effet, il est toujours possible de représenter des données de façon à exprimer exactement le contraire de la réalité. En étant conscients de tous ces possibles biais dans la chaîne de la donnée journalistique, il est important de connaître son parcours pour comprendre ce qui relève du construit dans les datavisualisations de presse.

En nous appuyant sur ARRUBARRENA (2015) et HARMAND (2020), nous considérerons que ce processus peut être décomposé en trois grandes étapes, que nous allons à présent étudier.

5.3.1 La collecte des données

La première étape consiste, évidemment, à récolter des données, matière première de toute datavisualisation. Deux possibilités sont alors offertes : soit les créer à partir de sa propre enquête (ce qui arrive parfois dans la presse, mais rarement, en faisant appel à une société spécialisée dans les sondages par exemple), soit (et c’est le cas le plus courant) les extraire bases de données « brutes » (ARRUBARRENA 2015b). Or, comme le font remarquer J. DENIS et GOËTA (2013), les « données brutes » sont un mythe : en effet, tout jeu de données est « fabriqué » en fonction d’un contexte social qui encourage les personnes ou algorithmes² à sélectionner certains événements ou objets à qualifier ou à quantifier, à choisir une méthode d’obtention des données, à mesurer en fonction d’une échelle connotée culturellement, etc. (voir la partie 1.1). HARMAND (2020) suggère même de « questionner en amont la source même de ces données, susceptible d’édulcorer la matière fournie de ses “arrière-pensées politiques” ». Sans aller jusqu’à imaginer des données intentionnellement falsifiées, nous pouvons admettre que la production de celles-ci est soumise à des biais cognitifs. Aussi, avant même le début du processus de création d’une datavisualisation journalistique, les données ne peuvent être considérées comme neutres, représentations fidèles de la réalité.

Le datajournaliste choisit donc un ou plusieurs de ces jeux de données. Il va souvent les chercher dans les bases de données ouvertes accessibles en ligne. Il peut également utiliser des programmes destinés récupérer automatiquement certaines données, tels que les *crawlers* (ARRUBARRENA 2015b). Les jeux de données extraits sont alors étudiés par le datajournaliste, qui évalue s’il

2. Nous admettons que les algorithmes étant programmés par des êtres humains, ils reproduisent les biais cognitifs de ceux qui établissent leurs instructions.

peut ou non en tirer des informations, voire une datavisualisation efficace. Pour cela, il devrait commencer par vérifier la méthode de collecte (afin de s'assurer que les biais évoqués précédemment sont minimales). Il s'assure également que les données soient « suffisantes et/ou comparables entre elles », comme le dit HARMAND (2020) :

« Un jeu de données peut s'avérer fragmentaire ou obsolète ; les données peuvent être difficiles à comparer, comme l'a montré la disparité de comptage des décès entre les différents pays durant l'épidémie de Covid-19 en 2020. »

HARMAND 2020

Cette étape de sélection des données qui seront représentées est, bien entendue, exécutée avec professionnalisme par les journalistes, sans volonté de tromper leur public. Il n'empêche qu'elle n'est pas exempte de biais : certaines sources sont privilégiées par rapport à d'autres, certains jeux de données sont exclus de l'analyse si leurs résultats ne paraissent pas en cohérence avec les enquêtes de terrain, etc. On se trouve ici dans un processus de captation des éléments de base, qui justifie l'emploi du mot « capta » que revendique DRUCKER (2011) en lieu et place des « data » qui seraient incontestables et objectives.

5.3.2 La transformation des données

Les données sélectionnées sont ensuite nettoyées grâce à des logiciels spécialisés (tel Open Refine). Cette opération consiste à supprimer les doublons, effacer les données aberrantes, ajouter des métadonnées qui explicitent un contexte, effectuer des tris, convertir les données spécifiques pour leur utilisation ultérieure dans certains logiciels de visualisation, etc. (ARRUABARRENA 2015b).

« Ce travail de la donnée est indispensable pour leur consolidation dans des tables exploitables par les outils visuels. Parfois fastidieuse, cette étape n'en reste pas moins un travail préalable essentiel pour garantir la qualité des données. »

ARRUABARRENA 2015b

Les données ainsi nettoyées sont alors prêtes à être analysées. Cette analyse implique la génération de représentations graphiques sur lesquelles le journaliste espère voir « ressortir [d]es *patterns* » (ARRUABARRENA 2015b), c'est-à-dire une structure interprétable par des méthodes statistiques. Pour structurer visuellement les données afin d'identifier ces *patterns*, les journalistes peuvent créer leurs propres programmes ou, le plus souvent, utiliser des algorithmes déjà prêts à l'emploi (par exemple Force Atlas). Puis ils doivent spécifier des paramètres graphiques « permettant d'attribuer des formes et des couleurs » (ARRUABARRENA 2015b). De nombreux outils clés en main existent à présent et sont particulièrement prisés par les journalistes non experts en datavisualisation (CHAGNOUX 2020) : « Gephi pour l'analyse visuelle de réseau, CartoDb pour la cartographie géographique, R pour l'analyse mathématique et statistique », etc. (ARRUABARRENA 2015b).

D'après HARMAND (2020), « le temps de l'analyse comporte aussi son lot de biais ». En effet, il est nécessaire d'être capable d'une « formalisation

théorique solide » pour savoir tirer des conclusions pertinentes du résultat visuel obtenu. Il faut éviter « [d']établir des [relations] arbitraires » entre les composantes du graphique (HARMAND 2020).

Il y a alors deux modalités de réflexion :

- Soit le journaliste a une hypothèse à vérifier, et il peut infirmer ou illustrer son idée de départ grâce à la visualisation ;
- Soit il « découv[re de façon] inattendue de nouvelles structures » « sur le mode de la sérenpidité » (ARRUABARRENA 2015b).

Les deux méthodologies sont soumises à des biais cognitifs. Dans la première configuration, le résultat que l'on cherche absolument à démontrer peut nous aveugler et nous empêcher de porter un regard lucide sur le graphe obtenu. De plus, quand la méthodologie fait défaut, nous encourageés à déformer et recalculer à partir des données jusqu'à voir apparaître ce que l'on imagine vrai, car il existe toujours une configuration des données qui permettra cela. HARMAND cite d'ailleurs TUFTE qui dit :

« Si vous torturez les données assez longtemps, elles vous révéleront n'importe quoi. »

Avec la deuxième méthode (chercher sans savoir ce que l'on va découvrir avec sérenpidité), il est tout autant possible de commettre des erreurs. En effet, comme le souligne HARMAND (2020) :

« Trouver des corrélations sans cause est d'ailleurs une caractéristique des algorithmes statistiques. »

On peut alors théoriser des phénomènes sur la base d'une simple corrélation (un *pattern* ayant été identifié) qui n'établit en fait aucun lien de causalité. Ces corrélations trompeuses sont illustrées par CAIRO (2020) dans son fameux *How charts lie*, avec un graphique qui représente l'espérance de vie des individus de plus de 14 ans dans chaque pays en fonction de leur consommation annuelle de cigarettes. Cette datavisualisation est présentée en figure 5.1.

Avec cette représentation graphique, il est possible de tracer une courbe de tendance croissante, qui montrerait que plus l'on fume, plus l'espérance de vie augmente. Or ce résultat est biaisé par la non prise en compte d'un facteur primordial : le niveau de développement des pays représentés. Si ce facteur est également pris en considération (comme c'est le cas en figure 5.2), il devient plus net que le niveau de vie a une influence sur l'espérance de vie. Ce graphique présentant de très nombreuses informations, une décomposition peut être nécessaire (montrée en figure 5.3) pour que l'on voit clairement apparaître trois groupes à l'espérance de vie distincte correspondant à trois niveaux de vie différents. On remarque d'ailleurs que les pays les moins riches ont la plus faible espérance de vie et la plus faible consommation de cigarette, ce qui contredit absolument la corrélation initialement observée avec le graphique biaisé.

Ainsi, comme le dit HARMAND (2020), le datajournalisme doit utiliser des données, mais ne doit pas se laisser piloter « aveuglément (...) par la technique ». Contrairement à ce que certains croient, il ne s'agit pas d'entrer des données dans un programme, qui calcul et fait apparaître un graphe tout prêt. Il faut une analyse fine et une bonne connaissance des méthodes de

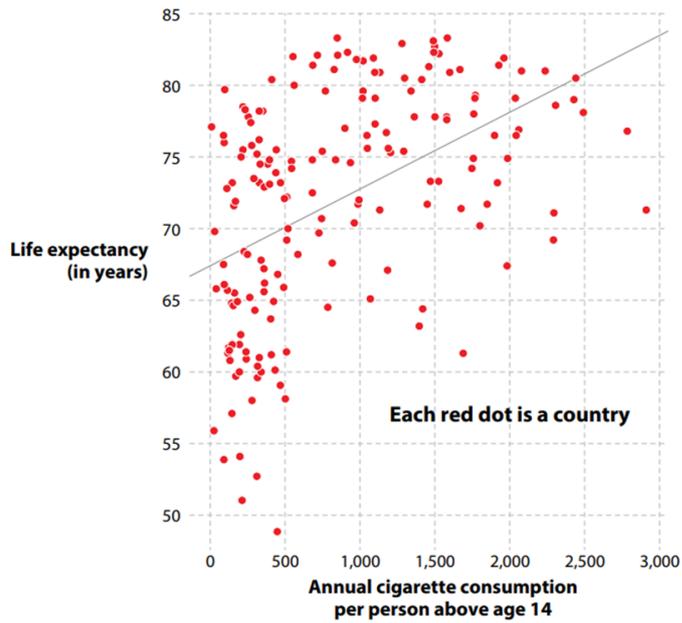


FIGURE 5.1 – Espérance de vie des individus de plus de 14 ans dans plusieurs pays en fonction de leur consommation annuelle de cigarettes

Source : CAIRO 2020

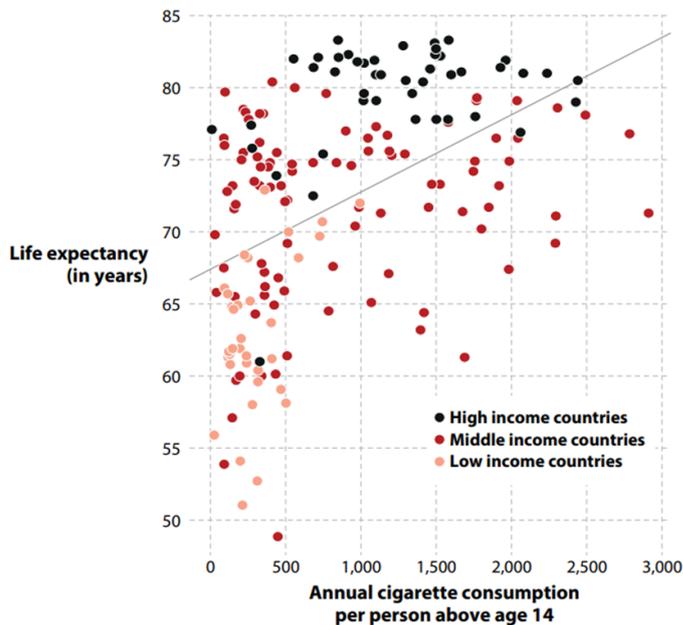


FIGURE 5.2 – Espérance de vie des individus de plus de 14 ans dans plusieurs pays de niveaux de vie différents en fonction de leur consommation annuelle de cigarettes

Source : CAIRO 2020

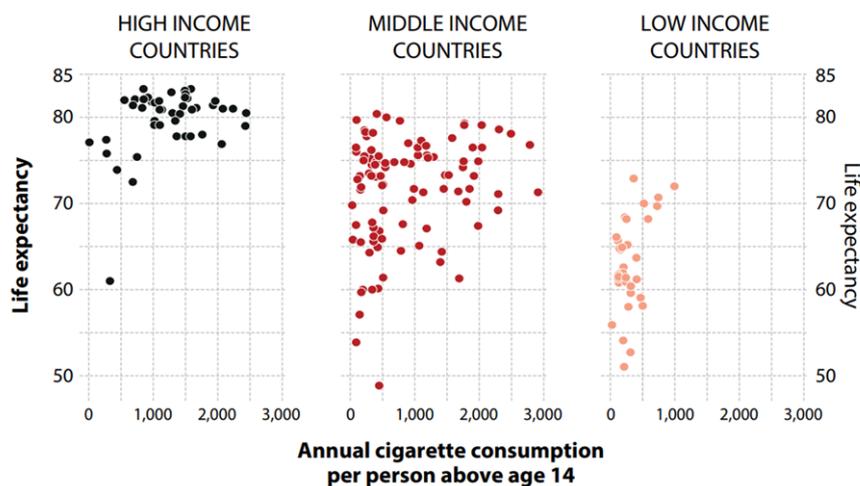


FIGURE 5.3 – Espérance de vie des individus de plus de 14 ans dans plusieurs pays en fonction de leur consommation annuelle de cigarettes et de leur niveau de vie

Source : CAIRO 2020

collecte des données, mais aussi des méthodes statistiques et appliquer avec rigueur le raisonnement scientifique (souvent dans une logique hypothético-déductive).

5.3.3 Restitution des données sous forme de datavisualisation

Cette dernière étape, qui consiste à créer le visuel final pour la communication, est le point clé pour les réalisateurs de datavisualisations journalistiques. En effet, toutes les étapes précédentes n'ont été réalisées que pour obtenir la datavisualisation finale, un produit qui doit permettre rendre le plus explicite possible le message auprès du grand public (comme nous l'avons mentionné en partie 5.2).

Là encore, le journaliste de données a le choix : soit il utilise des outils clé en main qui lui affichent automatiquement une datavisualisation classique, soit il crée son propre visuel avec des outils graphiques de la suite Adobe, par exemple, ou avec de la programmation informatique en HTML/CSS, avec des libraires javascript spécifiques (comme D3.js) pour créer de l'interactivité, etc. (ARRUABARRENA 2015b). Grâce à cette profusion d'outils et de méthodes, les datavisualisations prennent aujourd'hui des formes innombrables, selon HARMAND (2020). L'objectif des créateurs de datavisualisations doit alors être de « restituer une information intelligible » tout en faisant appel aux « préceptes du design d'information » le plus souvent (HARMAND 2020).

Le design d'information est utilisé régulièrement dans le milieu du journalisme. Il permet faciliter la communication avec le récepteur, et aide à la mise en récit des articles.

5.4 Récit médiatique et rhétorique des datavisualisations

Avec le mythe de l'objectivité journalistique, nous pourrions imaginer que les informations soient présentées à travers un discours descriptif, neutre. Or d'après LITS (2009), il n'est en rien. En effet, il explique que deux types de discours cohabitent dans les médias : le récit, « tel que défini par la linguistique textuelle, à savoir que le type textuel dominant y est bien le type narratif », et l'argumentation. Il souligne que le modèle narratif est dominant, ce qui « install[e] l'utilisateur dans une posture de consommation de type narratif » (LITS 2009). Ainsi, l'énonciation journalistique consisterait à mettre en récit des faits, afin de répondre à une demande des destinataires. Cette situation dans laquelle émetteur et récepteur communiquent *via* une « médiation particulière qui nous permet de donner du sens dans le chaos du monde, de l'organiser et de se construire une identité narrative, tant individuelle que collective » fait dire à LITS que « le médiatique serait (...) intrinsèquement narrativisé ». Il se demande même si l'audience des médias grand public n'aurait pas « une exigence d'histoire irréductible » (LITS 2009). Bien que LITS se concentre sur des notions textuelles, il n'en note pas moins que les médiations visuelles sont de plus en plus employées, et que les procédés de mise en récit y sont les mêmes.

Il note cependant qu'un autre courant promeut l'argumentation dans le discours médiatique. Il faudrait passer du narratif à l'interprétatif. Ce mouvement ne s'est cependant pas imposé, même si on rencontre un modèle qui allie narration et argumentation, et dans lequel « la narration [n']est qu'une étape dans un dispositif plus large où l'argumentation devient un élément central » (LITS 2009). La mise en récit est alors destinée à mieux servir une argumentation, à étayer l'opinion de l'émetteur, ou à l'illustrer pour que le récepteur lui affecte plus de sens.

Si la mise en récit semble imposer la médiation par un texte, nous trouvons aussi la notion « récit de chiffres » chez KOREN (2009). Il n'est alors pas question de datavisualisations, mais l'effet objectivant induits par l'emploi de données chiffrées qui jouent un « rôle narratif et argumentatif déterminant ». L'auteur explique d'ailleurs que les données sont passées du statut de « “fait brut” ou information technique » à celui « d'événement saillant qui joue un rôle central dans la mise en intrigue » (KOREN 2009). On retrouve ici l'idée du datajournalisme d'investigation qui, à partir des données, cherche à reconstituer les causes, de ses observations, voire de la pratique consistant à illustrer des données chiffrées par une enquête de terrain, pour raconter l'histoire vécue derrière les chiffres.

La datavisualisation est particulièrement propice à une mise en récit des chiffres : elle présente ceux-ci graphiquement, et peut utiliser la mise en image pour en raconter sa propre version, son interprétation. Comme le dit CAETANO (2020), ce procédé « fai[t] émerger [des] façons de voir et de montrer la réalité, pour convertir les perceptions en opérations discursives capables d'attirer l'attention et la sensibilité des lecteurs, en agissant sur le paraître du monde » (CAETANO 2020).

De plus, CHATENET et CARDOSO (2020) souligne la dimension « objectivante » des données, et qui est utilisées par les médias pour « montrer la vraisemblance du raisonnement [du journaliste] plutôt que de l'objet, en

mobilisant l’imaginaire de la rigueur scientifique » (CHATENET et CARDOSO 2020). De même, HARMAND (2020) évoque le « poids rhétorique » de la technique employée pour générer des datavisualisations, à partir « [d’]une matière pure (source) (...) travaillée par un processus implacable (traitement algorithmique) produisant une visualisation péremptoire (restitution) » (HARMAND 2020). En d’autres termes, tous ces auteurs soulignent la dimension objectivante des données, de la technique et de l’apparence scientifique des datavisualisations.

Cependant, cette rhétorique n’est utilisée que pour mieux argumenter, et prouver un point de vue :

« [Les représentations graphiques de données] utilisent la puissance persuasive de l’image pour offrir ses arguments à l’appropriation du lecteur par l’intermédiaire de figures, d’effets plastiques, semi-symboliques ou de références culturelles qui véhiculent un message (...) et suscitent l’adhésion. »

CHATENET et CARDOSO 2020

Résumé — Le contexte de production des datavisualisations médiatiques

Ainsi, nous avons pu constater que le datajournalisme est un milieu encore émergent et en construction. Les pratiques, les outils et les productions y sont divers et probablement non encore stabilisés, de même que les processus de créations et jusqu’aux théories qui sous-tendent la pratique. Mes deux entretiens exploratoires ont très bien illustré cette diversité : les deux datajournalistes interrogés avaient des conceptions de leur métier très différentes.

De plus, nous avons mis en évidence que les datavisualisations journalistiques étaient soumises à de nombreux biais tout au long de leur processus de création, depuis le choix des données jusqu’aux choix graphiques de leur restitution, en passant par leur traitement informatisé. Aussi, bien que les datavisualisations tendent à objectiver le travail des journalistes, il est important de garder à l’esprit qu’il s’agit de produits médiatiques, soumis comme tout autre à des impératifs de communication et de mise en récit pour intéresser les récepteurs.

CHAPITRE 6

ANALYSE SÉMIOLOGIQUE D'UN CORPUS DE DATAVISUALISATIONS JOURNALISTIQUES

Dans le cadre de ce mémoire, nous allons étudier un corpus de datavisualisations en forme de tableaux de bord présentant l'état des lieux de la pandémie de COVID-19. En sélectionnant ce corpus, l'objectif est de comparer des datavisualisations construites à partir de jeux de données similaires. Afin d'éclairer certains points qui ne seraient pas assez saillants sur ces graphiques, nous nous référerons parfois à d'autres datavisualisations en ligne produites par le même journal, mais sur le plus long terme et donc plus fidèles aux visions des datajournalistes.

Les journaux français étudiés sont :

- *Les Échos* : « L'évolution du coronavirus en France », appuyé par « Poker pollueur » et « Next 40 » ;
- *Le Parisien* : « Covid-19: l'évolution de l'épidémie en France » avec un regard ponctuel sur « Coronavirus: le nombre de cyclistes a chuté de 85% à Paris depuis le début du confinement », « Couvre-feu, confinement et déconfinement: quel impact sur le trafic parisien? », et « Coronavirus: le nombre d'admissions aux urgences a fondu » ;
- *Le Figaro* : « Covid-19: suivez l'évolution de l'épidémie, département par département » appuyé par l'observation de « Le confinement a-t-il permis d'endiguer le nombre des décès? » ;
- *Le Monde* : « Covid-19: le tableau de bord de l'épidémie » en nous aidant de « Moins de bruit, plus de télé, pas d'avions ou presque: la "France à l'arrêt" en douze graphiques ».

Les journaux des États-Unis sont :

- *The Washington Post* : « More than ... people have died from coronavirus in the U.S. », avec l'éclairage de « At least ... people have received one or both doses of the vaccine in the U.S. »
- *The New York Times* : « Coronavirus in the U.S.: latest map and case count » avec le recours ponctuel à « How the virus won »

Il est possible d'accéder à toutes ces datavisualisations en cliquant sur les titres de chaque article. On pourra également les retrouver *via* le titre de l'article ou du journal au gré de leurs mentions dans l'analyse qui suit.

6.1 Quelle place pour les datavisualisations dans le média ?

Dans un premier temps, nous ferons une remarque commune à tous les tableaux de bord étudiés : ceux-ci présentent non pas une, mais plusieurs datavisualisations juxtaposées sur la même page. De plus, toutes comportent des paragraphes de texte. Aucun des journaux étudiés n'a donc fait le choix de présenter une image unique, sans le moindre texte. Cependant, les proportions entre ces éléments diffèrent d'un journal à l'autre.

Cette structure alliant texte et images peut être étudiée au prisme des enseignements sémiologiques de BARTHES (1964), qui note le rapport étroit que peuvent entretenir message iconique et message linguistique.

6.1.1 Taille et disposition des différents éléments

Les datavisualisations composant le tableau de bord du *New York Times* se démarquent par l'espace qu'elles occupent sur la page. En effet, elles recouvrent toute la largeur de l'écran, sur un ordinateur comme en format responsive sur un écran de smartphone ou de tablette¹. Les datavisualisations sont complètement assumées en tant que produit journalistique de première importance. Le texte, relégué à droite de l'écran, apparaît au contraire comme secondaire. Nous pouvons également constater que la seconde visualisation du *New York Times* étudiée, « How the virus won », occupe également tout l'espace de l'écran, quoiqu'elle laisse plus de place au texte qui vient se superposer devant l'image. Pourtant, c'est bien la visualisation qui demeure au centre de l'attention : elle reste sur l'écran tout au long du *scrolling* vertical bien qu'elle évolue, alors que les explications textuelles ne font que passer devant. C'est bien la carte statistique qui est le fil conducteur de l'article, tout le texte semble construit autour d'elle, apparaissant tantôt à droite, tantôt à gauche, au gré des zones libres que laissent l'absence de signe représentant des données sur le fond de carte.

En comparaison, les autres journaux laissent de larges marges autour de leurs graphiques. Une exception est à noter : *Le Parisien* semble considérer que son tableau de bord est constitué exclusivement de chiffres et de graphiques, qui occupent dès lors tout l'espace disponible. Cependant, contrairement au *New York Times*, aucune représentation graphique n'occupe à elle seule toute la largeur de la page. C'est la juxtaposition de deux éléments qui le fait. Nous pouvons en déduire que, si *Le Parisien* assume un article construit autour de données et de leurs visualisations, aucune de celles-ci n'est privilégiée. Pour l'observateur, tous les graphiques occupent le même espace et sont composés des mêmes couleurs. Il ne semble pas y avoir d'ordre précis de lecture, qui, par défaut, se fera de gauche à droite, et du haut vers le bas, comme dans un journal papier. À l'inverse, le *New York Times* met en valeur la carte interactive des « Hot spots », immense et colorée en teintes de rouge-orangé (couleurs perçues avec beaucoup d'intensité par l'œil humain) qui saute immédiatement aux yeux.

Nous devons cependant remarquer que, si *Le Parisien* assume son

1. Cette vue responsive a été obtenue grâce aux outils de développeur (raccourci Fn + F12) permettant d'afficher les différents formats.

tableau de bord entièrement dédié à des données quantitatives, c'est loin d'être une généralité pour ce quotidien régional. En effet, les trois autres articles identifiés (« Coronavirus: le nombre de cyclistes a chuté de 85% à Paris depuis le début du confinement », « Couvre-feu, confinement et déconfinement: quel impact sur le trafic parisien? », et « Coronavirus: le nombre d'admissions aux urgences a fondu ») ont une structure complètement différente, avec des datavisualisations insérées au milieu d'une colonne de texte, qui semblent venir uniquement illustrer le propos écrit. D'ailleurs, les chiffres importants sont repris dans les paragraphes rédigés. Le tableau de bord du <https://www.leparisien.fr/coronavirus-france/Parisien>, qui relègue le texte explicatif sans aucune analyse des données en bas de page, est donc une exception pour ce journal.

Les autres articles sont construits sur le même modèle : une colonne centrale alternant texte et datavisualisations. La quantité de texte demeure faible pour tous les articles français. Il ne s'agit pas d'analyse des chiffres, mais uniquement d'explications des données que les graphiques représentent. Cependant, un regard aux autres datavisualisations proposées par ces journaux nous montre que, comme pour *Le Parisien*, le tableau de bord journalier représente une exception. En effet, tous les autres articles analysent les chiffres dans des textes (qui peuvent prendre plus ou moins d'ampleur selon le journal). Ce n'est pas le cas pour le *Washington Post* qui présente une analyse rédigée détaillée, allant même jusqu'à intégrer les chiffres du jour dans son titre. Cette différence entre journaux français et américains peut être le signe d'une impréparation des médias français à faire face à des traitements de masses de données quotidiennement, qu'il faut mettre à jour et qu'il faudrait en plus commenter. Les quotidiens américains tels que le *New York Times* et le *Washington Post* semblent, pour leur part, avoir été prêts à gérer cette situation, ce qui montre leur maîtrise plus avancée du datajournalisme. Cependant, il ne faut pas généraliser cela à tous les médias américains : le *New York Times* et le *Washington Post* sont deux des journaux qui ont investi le plus tôt sur les datavisualisations journalistiques. Les autres quotidiens américains n'ont sans doute pas encore atteint leur niveau de maîtrise du processus.

Nous remarquons que les datavisualisations du *Washington Post* concernant l'expansion de la pandémie à travers les États-Unis présente des graphiques plus larges qu'ils ne le sont dans les autres articles de même nature de ce journal (comme « At least ... people have received one or both doses of the vaccine in the U.S. »). Le texte, en revanche, reste bien normé, délimité par les mêmes marges qu'à l'accoutumée. Ce contraste met en valeur subtilement l'expansion irrépressible du virus, qui déborde les limites habituellement autorisées. La carte interactive, en particulier, est marquante par son encombrement spatial. Étant donné la brièveté de la légende et la possibilité d'interactions avec l'utilisateur (les valeurs s'affichent quand on passe la souris au dessus d'un État), nous ne pouvons pas penser que ce choix de taille soit lié à un impératif de lisibilité. Il s'agit donc bien d'un effet de style, qui renforce l'impression de submersion par le virus.

De ces observations générales concernant la disposition des différents éléments (textuels et graphiques) sur la page, nous pouvons conclure que les deux journaux américains étudiés, le *Washington Post* et le *New York Times* semblent affirmer d'avantage leurs datavisualisations comme centre

de leur article. En effet, dans les deux cas, les représentations graphiques sont très volumineuses par rapport au texte et attirent le regard. De plus, le texte est construit autour des données : il reprend et explicite les valeurs clés présentées sur les graphiques, et propose des liens hypertextuels vers d'autres pages qui peuvent compléter l'article par d'autres informations (on peut donc parler d'enrichissements présents dans le texte). On peut donc penser que ces deux pionniers en matière de datavisualisations journalistiques pratiquent le datajournalisme d'investigation (BOWERS, SCHWABISH et ROGERS 2017).

Par opposition, les médias français semblent plus timides dans leur usage des datavisualisations, qui demeurent parfois illustratives (comme c'est souvent le cas au *Parisien*). Il faut cependant noter que l'abondance de chiffres liées à la pandémie de COVID-19 a l'air de remettre en question ces pratiques de simple enrichissement par les données (BOWERS, SCHWABISH et ROGERS 2017). En effet, les tableaux de bord montrant l'état de la pandémie se compose principalement de datavisualisations, et de peu de textes. Cependant, contrairement aux tableaux de bord américains, les paragraphes textuels n'analysent pas les données pour faire ressortir les chiffres clés : on laisse les lecteurs regarder par eux-mêmes les graphiques et les interpréter.

On peut supposer que les pratiques de datajournalisme des médias français vont tendre de plus en plus vers celles (plus stabilisées) des médias américains. Aussi observera-t-on peut-être une homogénéisation de ces pratiques dans les années à venir.

6.1.2 Économie de l'attention

Si tous les médias étudiés proposent des datavisualisations, c'est en raison de l'impact fort que les images ont sur l'attention des récepteurs. En effet, ceux-ci ont tendance à aller plus volontiers vers un médias présentant une image rapidement comprise, sans nécessité de passer du temps à extraire l'information d'un texte dense. De plus, les couleurs et la nouveauté attirent les nouveaux publics.

C'est là une situation bien comprise par le *New York Times*. En effet, ce journal propose des datavisualisations novatrices, comme sa carte mobile des États-Unis et ses représentations de virus mouvants dans l'article « How the virus won ». Même dans le cadre du tableau de bord, qui ne permet que peu de fantaisie, tout est mis en œuvre pour attirer le lecteur et le maintenir dans le microcosme du journal. En particulier, les couleurs employées (très vives, avec une dominante de rouge et de orange marquant le danger représenté par le virus) sont repérées facilement par le spectateur, qui va probablement explorer l'article des yeux.

En promenant simplement sa souris sur les graphiques, il va s'apercevoir que toutes les datavisualisations qui lui sont proposées sont interactives : des étiquettes apparaissent en effet pour indiquer les valeurs en chaque point des courbes présentées. De plus, il a la possibilité d'utiliser des filtres sur certains graphiques ou cartes pour afficher des données qui l'intéressent. Par exemple, la carte interactive affichée par défaut présente les *hot spots*, mais l'utilisateur peut également choisir d'afficher la carte des pourcentages de personnes vaccinées.

Enfin, la carte interactive offre plusieurs niveaux de lecture : une vision globale du phénomène à l'échelle nationale, mais également comté par comté. Aussi le récepteur a-t-il la possibilité de personnaliser l'information. Par

exemple, s'il vit dans le Minnesota et qu'il souhaite partir en vacances en Floride, il peut consulter la carte des niveaux de risques qui lui indiquera (chiffres à l'appui) que le virus circule beaucoup là où il espère se rendre. Le consommateur de l'information est donc actif, il est invité à découvrir. Le journal ne lui offre pas seulement des nouvelles, il lui permet de les personnaliser et en fait une expérience.

De plus, de nombreux liens sont proposés, notamment à la fin de l'article, pour accéder à d'autres pages du *New York Times* et continuer à « traquer le coronavirus ». Cette rhétorique de la traque renvoie presque à l'univers du jeu vidéo : l'utilisateur peut poursuivre le COVID-19 comme il le ferait face à une menace imaginaire dans un jeu fictif. Ce procédé de liens hypertextuels permet de garder l'utilisateur sur le site du média, qui capitalise alors sur l'attention du consommateur.

Ainsi, l'esthétique de l'image, l'ergonomie et l'interactivité, c'est-à-dire en un mot le design du tableau de bord du *New York Times* joue le jeu de l'économie de l'attention.

Le *Washington Post* met en place des procédés similaires, quoique plus subtiles. En effet, les couleurs sont moins vives, les images n'apparaissent pas au premier abord, et les liens hypertextes sont moins mis en valeur. Cependant, d'autres choix éditoriaux encourage le lecteur à demeurer sur la page. En premier lieu, le titre écrit en très grands caractères, avec le nombre de morts défilant sous les yeux du nouvel arrivant sur la page frappe les esprits. Il encourage à *scroller* vers le bas pour découvrir les datavisualisations interactives qui suivent. Là encore, deux procédés d'interactivité sont employés : des filtres et les étiquettes apparaissant au passage de la souris. De plus, la granularité de l'information peut être personnalisée. La carte interactive permet, en effet, une vision globale de la situation du pays tout entier, avec une possibilité de zoom sur un comté choisi par le récepteur. Cette fois, ce n'est pas l'exploration qui guide le lecteur, mais le texte, qu'il est invité à consulter puisqu'il forme une continuité avec les datavisualisations.

Le Monde, premier quotidien français à avoir investi le domaine des datavisualisations à peu près à la même période que ces homologues américains, reprend les mêmes procédés d'interactivité. Cependant, il n'a pas de titre marquant ni de couleurs interpellantes au premier abord. On est ici moins dans l'économie de l'attention que dans l'aide à la lecture des graphiques (avec les étiquettes de données interactives en particulier). De plus, les filtres proposés, s'ils permettent d'afficher d'autres jeux de données pour des experts, n'offre aucune expérience au simple citoyen qui ne peut guère personnaliser l'information. L'idée des cartes américaines présentant les données territoire par territoire est reprise, mais sans que le fond de carte de la France ne soit tracé. Comme l'intégralité de la visualisation ne s'affiche pas sur un seul écran (sauf sur iPhone), cette représentation graphique originale qui permettrait de consulter les données au cas par cas demeure probablement incomprise par certains (voir beaucoup), et apparaît comme une liste de graphes dans laquelle on ne se retrouve pas.

Ni *Le Figaro*, ni *Le Parisien* n'ont tenté d'attirer le regard du lecteur par la couleur. Ils ont misé sur un code simple et sans ambiguïté : l'association

rouge / vert / bleu

- Le rouge met en valeur les points les plus négatifs, comme le nombre de morts ou le nombre de personnes en réanimation ;
- Au contraire, le vert représente les valeurs positives, comme le nombre de retours à domicile, ou encore les variables en cours d'amélioration (comme le taux de positivité, en baisse) ;
- Le bleu présente les données avec neutralités. Il contraste bien avec le rouge (comme le soulignait BERTIN et permet donc de distinguer aisément les deux courbes superposées du *Figaro* par exemple (sans poser de problème aux daltoniens).

Les Échos ajoutent à ce panel de couleur le noir pour les décès. Ces trois médias français restent dans la sobriété, commençant par des résumés en chiffres pour *Les Échos* et *Le Parisien* (quoique les termes employés par ce dernier soient moins explicites de même que les tendances d'évolutions). Tous font le choix de limiter le nombre de graphiques de données :

- *Le Parisien* en proposant de tracer les graphiques en fonction d'un filtre ;
- *Les Échos* grâce à plusieurs filtres (dont un temporel et un moteur de recherche permettant d'afficher des courbes par département) ;
- *Le Figaro* en choisissant de ne représenter que deux jeux de données (le nombre de personnes hospitalisés et celui des individus en réanimation), reproduit pour chaque département.

Ces trois journaux français ne se sont pas alignés avec les trois précédents, plus axés sur le design et l'esthétique attractifs pour le consommateur (quoique le tableau de bord des *Les Échos* soit très élégant). Leur mot d'ordre semble au contraire être la sobriété, en termes de couleurs mais aussi de graphiques, qui ne sont pas tous affichés. Les fonctionnalités interactives servent à afficher des jeux de données différents dans un même espace et à afficher les valeurs des données dans des étiquettes. *Le Figaro* ne propose d'ailleurs aucune forme d'interactivité avec les données. Cependant, ce journal ne semble pas opposé à l'interactivité en général, puisque nous observons que certaines datavisualisations de l'article « Le confinement a-t-il permis d'endiguer le nombre des décès? » bénéficient du système d'étiquette de données quand on passe la souris du la représentation graphique. Pour *Le Figaro* comme pour *Le Parisien*, l'interactivité n'est pas incorporée systématiquement à toutes les datavisualisations réalisées. Elle demeure accessoire.

Si les choix éditoriaux de ces journalistes permettent à des experts de trouver toutes les informations dont ils ont besoin et au citoyen lambda de suivre l'évolution globale de la situation, l'expérience utilisateur est, pour sa part, amoindrie. L'utilisateur a du mal à personifier l'information : il peut, au mieux, voir quelques petites courbes concernant un département. Il a peut souvent accès à des cartes interactives (sauf pour *Les Échos*, mais la carte ne porte aucun point de repère).

Ainsi, nous pouvons constater que les médias américains, en particulier le *New York Times*, qui ont parfaitement intégrées les datavisualisations à leurs productions traditionnelles, utilisent ces dernières pour attirer l'attention de leurs usagers et les maintenir le plus longtemps possible sur leur site internet. Pour cela, ils utilisent le design qui offre une expérience utilisateur (parfois même personnalisable) avec de l'interactivité, mais aussi une grande importance

accordée à l'esthétique (notamment aux couleurs), et à l'ergonomie générale (l'utilisateur n'a pas d'indications d'utilisation des datavisualisations, il est censé tester les interactions possibles).

Si *Le Monde* semble vouloir s'inscrire dans ce même mouvement (quoiqu'il ne possède pas encore la maîtrise du *New York Times*), les trois autres médias français étudiés privilégient la simplicité et la sobriété, d'après nos observations. Ils ne paraissent pas utiliser les datavisualisations comme un argument commercial dans le contexte de l'économie de l'attention.

Nous notons également que *Le Figaro* et *Le Parisien* optent plus volontiers pour des graphiques de données classiques, qui respectent les enseignements et les codes de BERTIN, sans interactivité nécessaire. Ils ne s'attachent pas spécialement au design de l'information.

6.1.3 La narration par les données

Les observations précédentes ont un corollaire : si le *Washington Post* et le *New York Times* utilisent de nombreux procédés de design d'information pour attirer l'attention, ils mettent également leur données en récit.

Le *Washington Post* utilise un procédé classique du journalisme en guidant la lecture des datavisualisations par son texte, qui met prioritairement l'accent sur le nombre de morts. Les codes couleurs choisis pour les chiffres renforcent cette sensation de danger, avec le nombre de morts affiché en prune dans le titre (couleur de deuil, entre noir et rouge), et les représentations de données sur graphique ou sur carte en orange (couleur proche du rouge, qui éveille l'attention, et marque le risque).

Au contraire, dans l'article « At least ... people have received one or both doses of the vaccine in the U.S. », le *Washington Post* utilise des nuances de vert, comme pour montrer les bienfaits de la vaccination. Ceux qui sont vaccinés sont approuvés par l'énonciateur, c'est-à-dire le journaliste, le concepteur de datavisualisation, et plus généralement la rédaction du journal. Cet ensemble de personnes qui s'appuient sur des données quantitatives objectivantes et qui bénéficient de l'aura de leur prestigieux quotidien apparaissent au récepteur comme une élite bien informée, qui conseille fortement la vaccination pour faire opposition au danger que représente le coronavirus. Les codes couleurs et les choix graphiques n'ont alors rien d'anodin : ils sont désignés pour influencer la perception du récepteur. Il ne s'agit pas de couleurs facilitant la lectures de visualisations, mais interprétant déjà ces dernières.

De même, la carte interactive des États-Unis choisit de représenter le nombre de nouveaux cas ou de morts localement par des cercles dont la surface est proportionnelle à la valeur pour cent mille habitants. Si ce procédé est légitime en soi (il ne donne pas de fausse indication), il peut prêter à confusion : au lieu de considérer l'information comme étant relative à la zone au centre du cercle, un observateur peut y voir une expansion spatiale du virus dans son environnement. Cette sensation est renforcés par un habile jeu de transparence, qui permet aux cercles devenus trop larges pour demeurer dans les limites du comté qu'ils représentent de se recouvrir les uns les autres et de former des motifs plus ou moins intensément colorés. Cette situation peut être interprétés comme un « pattern » surlignant les zones les plus à risque, alors qu'il correspond en réalité à un effet de style. Bien sûr, les zones les

plus foncées sont au croisement de plusieurs comtés dont les indicateurs de cas pour cent mille habitants sont inquiétants. Cependant, le zone la plus foncée est située au croisement de deux valeurs. On ne peut donc pas tirer de conclusion quant au lieu précis qui fait l'objet de cette plus grande intensité.

Le *New York Times* utilise les mêmes procédés de narration par les couleurs (avec du rouge et de l'orange vifs pour marquer le danger, du noir pour les décès, du vert pour la vaccination) et les choix graphiques (des cercles de surface proportionnelles au nombre de décès ou de cas sur certaines représentations). L'interprétation des données sur la carte même est encore plus évidente sur la datavisualisation des niveaux de risque (*risk levels*) pour les personnes non vaccinées. Si l'information présentée s'appuie sur des chiffres (la moyenne du nombre de nouveaux cas par comté pour cent mille habitants), ce ne sont pas les données qui sont représentées mais leur interprétation qualitative par le datajournaliste, qui décide de placer ces valeurs sur une échelle qui peut paraître arbitraire car non explicitée. Les données numériques sont accessibles en passant le souris sur chaque comté, mais elle n'arrive qu'en dernière position, comme une simple justification factuelle objectivant le commentaire qui la précède : le niveau de risque estimé, suivi le plus souvent du conseil de porter le masque pour tous, personnes vaccinées ou non.

Bien que le *New York Times* n'utilise pas, dans ce cas, le texte pour faire le récit des données, il crée une narration de celles-ci directement par la représentation visuelle. Les chiffres sont alors objectivants, de même que l'impression de liberté laissée à l'individu d'explorer des données, alors que celles-ci sont déjà interprétées pour lui.

Il ne s'agit cependant pas des seuls procédés de mise en récit des données utilisés par le *New York Times*, qui n'hésite pas à recourir à l'intégration texte/datavisualisation comme dans l'article « How the virus won ». L'interaction y est limitée (il s'agit seulement d'interaction avec la fenêtre, grâce à l'ascenseur vertical que l'on utilise pour *scroller* vers le bas. Les données utilisées par le datajournaliste sont affichées sur un fond de carte, défilant au gré de la lecture du texte, mais sans légende chiffrée (tout au plus voit-on à quoi correspondent les points qui apparaissent) et déjà interprétées grâce au message textuel et aux codes graphiques. On se trouve là dans une scénarisation de l'information, qui pourrait constituer une vidéo. Le choix de ne pas s'orienter vers ce type de produit est probablement lié à deux facteurs :

- Le *New York Times* est un journal papier à l'origine, qui se diversifie en format numérique pour rencontrer un public qui s'informe sur internet. Il ne souhaite peut-être pas aller jusqu'à produire des vidéos, qui pourraient le faire passer dans une autre catégorie de médias, peut-être considérée comme moins sérieuse.
- Si la vidéo impose son vitesse et la voix d'un narrateur, la combinaison datavisualisation/texte en *scrolling* vertical laisse plus de latitude au lecteur, qui peut suivre le déroulement de la scénographie designée à son attention à son rythme, en ayant l'impression de découvrir les informations de lui-même par sa lecture. Le récepteur est plus impliqué que devant une animation qui le laisserait passif.

Le dernier procédé décrit de mise en récit des données alliant texte et datavisualisations dans un *scrolling* vertical semble être devenu une spécialité des *Échos*, bien que le journal ne l'ait pas employé sur son tableau de bord de

l'épidémie. Il est utilisé dans la plupart de ses « infographies ». Celles-ci ne revendiquent d'ailleurs pas le nom de « datavisualisation » malgré la présence de diagrammes, cartes et réseaux comme base même de la construction du récit. Ainsi, par exemple, les articles « Poker pollueur » et « Next 40 » sont désignés avec un défilement de texte et de datavisualisations, qui s'alimentent mutuellement : le message textuel met en évidence certaines informations portées par les graphiques, mais ceux-ci en portent plus encore. L'utilisateur peut les explorer interactivement, les lire intégralement grâce aux légendes complètes (en comparaison de celles du *New York Times*) et se forger une opinion sur l'interprétation journalistique portée par le texte.

Ainsi, bien que les données représentées et l'analyse qui en est faite par le média soient anglées² puisque les données sont sélectionnées pour corroborer l'interprétation, le lecteur a le loisir d'explorer les représentations avec un regard critique, comme il le ferait pour un article de presse classique. Contrairement au *New York Times*, il privilégie un code couleur moins connoté, qui ne paraît pas destiné à influencer la perception du récepteur mais plutôt à correspondre à une charte graphique interne au journal (dominantes de noirs/blancs/gris, avec des datavisualisations faisant intervenir des teintes précises de orange, jaune et violet).

D'autres médias français s'essaient à cette narration des données, avec moins de succès que *Les Échos* et que le *New York Times* (quoique nous ayons établi des différences fondamentales entre leurs deux approches). En particulier, *Le Monde* s'y essaie dans l'article « Moins de bruit, plus de télé, pas d'avions ou presque: la "France à l'arrêt" en douze graphiques ». Cependant, la mise en forme ne met pas en valeur les visualisations, avec le texte passant devant les graphiques et masquant ceux-ci, qui sont pourtant interactifs et ont une légende complète. On voit que l'enquête se base sur les données, que celles-ci ont donné lieu à des représentations graphiques, et que de ces graphes ont émergé des informations que le journaliste nous restitue textuellement. Cependant, il est difficile de trouver une position de l'ascenseur vertical qui permette la lecture intégrale de chaque datavisualisation, et donc de porter un regard critique sur les analyses journalistiques qui en résultent.

À l'instar du *New York Times*, les datavisualisations semblent seulement venir légitimer l'information, chiffres à l'appui. Cependant, contrairement au média américain, les couleurs et autres effets graphiques ne semblent pas construits pour influencer la perception du lecteur et corroborer les dires du journaliste. Les graphiques sont en effet classiques, avec des couleurs qui semblent aléatoires quoique toutes bien discernables pour gagner en lisibilité.

Nous trouvons une autre tentative de ce genre de mélange datavisualisation/interprétation textuelle dans l'article « Le confinement a-t-il permis d'endiguer le nombre des décès? » du *Figaro*. Peut-on alors parler de mise en récit des données? Selon nous, il s'agit plutôt de présenter une analyse prête à l'emploi directement liée à l'image qui attire le regard.

Cela interroge sur la perception que les datajournalistes ont de leurs récepteurs : pensent-ils qu'ils ne réussiraient pas à lire les datavisualisations présentées autrement? En effet, nous avons vu que le seul média qui, lorsqu'il

2. Nous faisons référence ici à l'angle journalistique d'attaque d'un sujet : la rédaction du journal assume un point de vue sur une thématique.

a recourt à ce procédé de mélange texte / datavisualisation, laisse la possibilité de lire l'intégralité des graphiques de données est *Les Échos*. Tous les autres imposent leur interprétation sur leurs datavisualisations, de façon explicite dans le cas des médias français, et en influençant la perception des usagers dans le cas des médias américains étudiés.

La question de la littératie impliquée dans les lectures de ces datavisualisations de presse numérique s'impose donc : en tant que récepteurs, sommes-nous censés les lire, les déchiffrer, les interpréter et porter dessus un regard critique ; ou devons-nous simplement utiliser nos capacités de lecture de texte pour accéder directement à l'analyse proposée par le média ? Finalement, toutes les observations que nous avons faites sur les datavisualisations de notre corpus de presse nous interrogent sur l'emploi par les médias d'une littératie propre aux datavisualisations, qui pourrait bien n'être finalement qu'un recours aux compétences classiquement mobilisées par la presse écrite (quelle soit papier ou numérique) : savoir lire et comprendre des textes.

6.2 Quelle littératie est mobilisée ?

La question de la littératie que nous venons de poser nécessitera que nous étudions en détail chacune des formes de littératie que nous avons mentionné en partie 3.

6.2.1 La littératie des données

Nous notons que les tableaux de bord étudiés mentionnent tous :

- La date de dernière mise à jour (ou la fréquence de mise à jour) ;
- La provenance des données exploitées.

Cela étant dit, des différences apparaissent quand au statut de cette dernière information. D'une part, la source des données ne bénéficie pas toujours d'un lien hypertexte pour pouvoir retrouver le jeu de données initial. C'est le cas du *Figaro*, du *Parisien* et du *Monde*.

Ensuite, certains mettent ces informations en exergue, à l'image du *Figaro* avec son encadré « sources et méthodologie » ; alors que d'autres cachent l'information, comme le *New York Times* qui impose de cliquer sur les mentions « *about this data* » spécifiées en tous petits caractères sous chaque graphique.

Enfin, la méthodologie de traitement des données est tantôt spécifiée, tantôt non. Il semble n'y avoir aucune uniformité en matière d'affichage de sources et de méthodologie, si ce n'est de mentionner la source. La précision de la provenance des données est conseillée par TUFTE afin de légitimer le graphique et d'instaurer une relation de confiance entre l'émetteur et le récepteur.

Cependant, cette pratique ne permet en aucun cas de se faire une idée critique quand au traitement des données, ni même concernant leur contexte de production. Bien qu'il faille maîtriser la littératie des données pour savoir que des biais méthodologiques peuvent exister, et qu'il faut par conséquent nuancer les informations présentées sur ces datavisualisations, les datajournalistes ne donnent aucun clé qui permette l'analyse critique de leur méthode.

Cette situation est paradoxale : alors que l'ouverture des données est censée assurer une transparence, et que leur réutilisation apparaît comme une médiation vers le citoyen, celui-ci n'a pas le loisir de comprendre comment les données ont été traitées.

Ainsi, la littératie des données est nécessaire pour avoir un regard averti sur les datavisualisations médiatiques à destination du grand public. Cependant, ne pas disposer de cette compétence n'empêche en rien de comprendre l'information communiquée par la représentation graphique. Les médias, dans leur grand majorité (à l'exception du *Figaro*) semblent donc ne pas s'intéresser à cette forme de littératie, et tentent de la rendre optionnelle.

6.2.2 Littératie des datavisualisations ou littératie informatique ?

La véritable question qui demeure encore posée est la suivante : a-t-on besoin de savoir lire un graphique ou une carte pour comprendre les datavisualisations journalistiques étudiées ? En effet, la lecture de cartes et de graphiques est une compétence qui s'enseigne au collège et au lycée, et qui consiste à acquérir les codes de la projections sur les axes pour les diagrammes, de la représentation du territoire pour les cartes, etc. C'est d'ailleurs ce qui justifie notre emploi du terme « littératie » appliqué aux datavisualisations.

Or, comme nous l'avons déjà évoqué, nous n'avons pas besoin de maîtriser ces codes pour percevoir le message global des articles très connotés des médias américains, qui au moyen de codes couleur et d'autres artefacts graphiques induisent un sentiment de menace et de submersion par le virus.

Cependant, qu'en est-il si l'on souhaite lire l'information plus en détail ? A-t-on besoin de maîtriser des codes pour connaître le nombre de nouveaux cas aux États-Unis survenus le 7 mai, par exemple ? La réponse nous semble négative : en effet, ce n'est pas en lisant la valeur sur les axes que l'on identifie la réponse, mais grâce à l'étiquette interactive qui apparaît lorsque l'on passe la souris dessus. Avec toutes les fonctionnalités interactives (étiquettes, filtres) et les textes mentionnant les valeurs importantes, les données sont lues pour nous. Grâce au design d'interaction, la littératie des datavisualisations n'est plus nécessaire.

Si cela est clair sur les articles américains étudiés, il n'en va pas de même pour les médias français. *Les Échos* et *Le Monde* proposent les mêmes procédés interactifs et des textes présentant les valeurs à retenir. Ils s'orientent donc dans la même voie que les médias américains. Cependant, leurs légendes complètes et rendent leurs datavisualisations lisibles malgré tout, ce qui n'est pas le cas, par exemple, de celle du *New York Times* : « How the virus won ».

Les datavisualisations du *Figaro* et du *Parisien* restent classiques et nécessitent une lecture. Elles reprennent les variables visuelles de BERTIN sans prendre en compte le design d'interaction. Il faut donc encore une littératie des données pour les lire.

Il reste un point à souligner : si les datavisualisations américaines n'exigent pas de littératie particulière des datavisualisations, elles nécessitent en revanche de maîtriser l'outil informatique, et de connaître les codes de l'interactivité (par exemple, le passage de la souris sur un point pour en afficher la valeur). Dans ce cas, ne pourrait-on dire que les datavisualisations médiatiques informatisées à destination du grand public remplacent la littératie

des datavisualisations par la littératie informatique, supposant sans doute qu'une plus grande part de leur lectorat maîtrise cette dernière ?

CONCLUSION

Dans ce mémoire, nous avons pu définir de nombreux concepts : ceux de données, visualisation, datavisualisation, littératie, etc. Toutes les informations issues de nos recherches ainsi que deux entretiens semi-directifs exploratoires nous ont permis de comprendre l'écosystème du datajournalisme, et de percevoir les biais inhérents à la réalisations de datavisualisations.

Grâce à notre état de l'art, nous avons pu établir les contours d'une littératie des datavisualisations. Pourtant, au terme de cette étude, cette définition nous paraît bien inutile puisqu'il semble, d'après notre analyse sémiologique, que les médias grand public emploient de multiples procédés pour minimiser au maximum les besoins de cette compétence spécifique. Les médias américains, et dans leur foulée les français, semblent transférer ces compétences vers une maîtrise des codes informatiques.

Rappelons notre problématique de départ :

Quelle(s) littératie(s) sont mobilisée(s) par les datavisualisations médiatiques informatisées à destination du grand public ?

Nous pouvons y répondre que les médias à destination du grand public tendent à minimiser l'usage de littératie autre que générale et informatique.

Cependant, nous nuancerons en remarquant que certains médias français ne se sont pas encore engagés dans cette voie, et qu'il est toujours nécessaire de maîtriser la lecture graphique pour les comprendre. De plus, nous ajouterons qu'à l'heure du Big data et de l'ouverture des données, la littératie des données (entendue comme la capacité à accéder aux jeux de données, et à les interpréter et manipuler de façon critique) est un enjeu de plus en plus important, qui reste encore trop peu exploré dans les formations.

BIBLIOGRAPHIE

- [Ali20] Aran ALI. *Here's What Happens Every Minute on the Internet in 2020*. en-US. Sept. 2020. URL : <https://www.visualcapitalist.com/every-minute-internet-2020/> (visité le 06/08/2021).
- [Ama15a] Étienne Armand AMATO. « Enjeux et opportunités de la datavisualisation : interagir avec les données ». fr. In : *I2D — Information, données documents* Volume 52.2 (juill. 2015), p. 34-35. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-34.htm> (visité le 19/08/2021).
- [Ama15b] Étienne Armand AMATO. « Formalisation et accessibilité des datas : de la complexité à la simplicité dans un monde numérique ». fr. In : *I2D — Information, données documents* Volume 52.2 (juill. 2015), p. 30-31. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-30.htm> (visité le 19/08/2021).
- [Arr15a] Béa ARRUBARRENA. « Datavisualisation : des données à la connaissance ». fr. In : *I2D — Information, données documents* Volume 52.2 (juill. 2015), p. 28-29. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-28.htm> (visité le 19/08/2021).
- [Arr15b] Béa ARRUBARRENA. « L'écosystème numérique de la datavisualisation ». fr. In : *I2D — Information, données documents* Volume 52.2 (juill. 2015), p. 56-58. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-56.htm> (visité le 19/08/2021).
- [AKC19] Béa ARRUBARRENA, Gérald KEMBELLEC et Ghislaine CHARTRON. « Data littératie et SHS : développer des compétences pour l'analyse des données ». fre. In : (2019). URL : https://www.researchgate.net/publication/331235075_Data_litteratie_SHS_developper_

- des _ competences _ pour _ l % 27analyse _ des _ donnees.
- [Avé08] Valérie AVÉROUS VERCLYTTE. « Clarisse Herrenschmidt, Les trois écritures. Langue, nombre, code ». fr. In : *Mots. Les langages du politique* 86 (mars 2008), p. 122-126. ISSN : 0243-6450. URL : <http://journals.openedition.org/mots/13872> (visité le 06/08/2021).
- [Bar13] Renaud BARBARAS. « Préface ». fr. In : Gilbert SIMONDON. *Cours sur la perception (1964–1965)*. Paris : Presses Universitaires de France, 2013, IX-XVI. ISBN : 978-2-13-062529-2. URL : <https://www.cairn.info/cours-sur-la-perception-1964-1965--9782130625292-page-IX.htm>.
- [Bar64] Roland BARTHES. « Rhétorique de l'image ». fr. In : *Communications* 4.1 (1964), p. 40-51. DOI : 10.3406/comm.1964.1027. URL : https://www.persee.fr/doc/comm_0588-8018_1964_num_4_1_1027 (visité le 22/02/2021).
- [BDL06] Charles BAZERMAN, Christiane DONAHUE et Jean-Claude LEJOSNE. « Écriture, organisation sociale et cognitions : repenser Goody ». fr. In : *Pratiques* 131.1 (2006), p. 95-115. DOI : 10.3406/prati.2006.2120. URL : https://www.persee.fr/doc/prati_0338-2389_2006_num_131_1_2120 (visité le 22/02/2021).
- [Ber70] Jacques BERTIN. « La graphique ». fr. In : *Communications* 15.1 (1970), p. 169-185. DOI : 10.3406/comm.1970.1221. URL : https://www.persee.fr/doc/comm_0588-8018_1970_num_15_1_1221 (visité le 22/02/2021).
- [Bes06] Jean-Marc BESSE. *Cartographie et pensée visuelle : réflexions sur la schématisation graphique*. fr. Juin 2006. DOI : 10/document. URL : <https://halshs.archives-ouvertes.fr/halshs-00256710> (visité le 22/02/2021).
- [BD81] Pierre BOURDIEU et Yvette DELSAUT. « Pour une sociologie de la perception ». fr. In : *Actes de la Recherche en Sciences Sociales* 40.1 (1981), p. 3-9. DOI : 10.3406/arss.1981.2131. URL : https://www.persee.fr/doc/arss_0335-5322_1981_num_40_1_2131 (visité le 10/08/2021).
- [BSR17] Danielle BOWERS, Jonathan SCHWABISH et Simon ROGERS. *Data journalism in 2017 : the current stage and challenges facing the field today*. En. Rapport de recherche. Google news lab, 2017, p. 24. URL : <https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf>.
- [Boy15] Jeremy BOY. « Littératie en visualisation de données ». fr. In : *I2D — Information, données documents* Volume 52.2 (juill. 2015), p. 32-33. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-32.htm> (visité le 19/08/2021).

- [Bra15] Christophe BRASSEUR. « Usages visuels des données & Big data ». fr. In : *I2D — Information, donnees documents* Volume 52.2 (juill. 2015), p. 44-46. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-44.htm> (visité le 06/08/2021).
- [Cae20] Kati CAETANO. « Des données au sens dans le discours de l'information. La rhétoricité de l'infographie ». fr. In : *Interfaces numériques* 9.3 (déc. 2020). ISSN : 2259-1001. DOI : 10.25965/interfaces-numeriques.4399. URL : <https://www.unilim.fr/interfaces-numeriques/index.php?id=4399> (visité le 20/08/2021).
- [Cai20] Alberto CAIRO. *Georgia international conference on information literacy. How charts lie : what you design is not what people see.* en. Archives de conférence. 2020. URL : https://digitalcommons.georgiasouthern.edu/gaintlit/2020/2020/79/?utm_source=digitalcommons.georgiasouthern.edu%2Fgaintlit%2F2020%2F2020%2F79&utm_medium=PDF&utm_campaign=PDFCoverPages.
- [Can19] CANOPÉ. *L'usage de la data visualisation dans les médias.* en. 2019. URL : <https://www.clemi.fr/index.php?id=626> (visité le 22/02/2021).
- [CA15] Antonio CASILLI et Béa ARRUBARRENA. « Voir pour comprendre et comprendre pour voir ». fr. In : *I2D — Information, donnees documents* Volume 52.2 (juill. 2015), p. 40-41. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-40.htm> (visité le 19/08/2021).
- [Cha20] Marie CHAGNOUX. « La datavisualisation, double point d'entrée du datajournalisme dans la PQR ». fr. In : *Interfaces numériques* 9.3 (déc. 2020). ISSN : 2259-1001. DOI : 10.25965/interfaces-numeriques.4480. URL : <https://www.unilim.fr/interfaces-numeriques/index.php?id=4480> (visité le 20/08/2021).
- [CC20] Ludovic CHATENET et Stéphanie CARDOSO. « Du graphique à l'infographie. De l'art de faire parler les images ». fr. In : *Interfaces numériques* 9.3 (déc. 2020). ISSN : 2259-1001. DOI : 10.25965/interfaces-numeriques.4408. URL : <https://www.unilim.fr/interfaces-numeriques/index.php?id=4408> (visité le 20/08/2021).
- [Cit14] Yves CITTON. « Introduction ». fr. In : *L'économie de l'attention : Nouvel horizon du capitalisme?* Google-Books-ID : hGuuAwAAQBAJ. La Découverte, juin 2014. ISBN : 9782707182968.

- [Coo13] Clive COOKSON. « Edward Tufte ». In : *Financial Times* (juill. 2013). URL : <https://www.ft.com/content/dda1cb5c-f4c0-11e2-a62e-00144feabdc0> (visité le 17/08/2021).
- [Dat19] DATAACTIVIST. *Répertoire commun des modules de cours pour les IA et 5A de Sciences Po Saint-Germain-en-Laye*. fr. 2019. URL : <https://github.com/dataactivist/SPoSGL> (visité le 05/08/2021).
- [Den18] Jérôme DENIS. *Une littératie des données ?* Autopublication pdf. Centre de sociologie de l'innovation— Mines ParisTech Retours sur la troisième conférence «Data Literacy» 5 et 6 octobre 2018, Aix-en-Provence. 2018.
- [DG13] Jérôme DENIS et Samuel GOËTA. « La fabrique des données brutes. Le travail en coulisses de l'open data ». fr. In : fév. 2013. URL : <https://halshs.archives-ouvertes.fr/halshs-00990771> (visité le 07/08/2021).
- [Des15] Orélie DESFRICHES-DORIA. « Quels dispositifs numériques pour appréhender la datavisualisation ? » fr. In : *I2D — Information, donnees documents* Volume 52.2 (juill. 2015), p. 54-56. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-54.htm> (visité le 19/08/2021).
- [Dro14] Béatrice DROT-DELANGE. « Littératie informatique : quels ancrages théoriques pour quels apprentissages ? » fre. In : *Spirale - Revue de recherches en éducation* 53.1 (2014), p. 121-132. DOI : 10.3406/spira.2014.1054. URL : https://www.persee.fr/doc/spira_0994-3722_2014_num_53_1_1054 (visité le 22/02/2021).
- [Dru11] Johanna DRUCKER. « Humanities Approaches to Graphical Display ». In : *Digital Humanities Quarterly* 005.1 (mars 2011). ISSN : 1938-4122.
- [Dym18] Valentyna DYMYTROVA. « Data journalisme, entre pratique créative innovante et nouvelle médiation experte ? Une analyse conjointe des discours et des productions journalistiques ». In : *XXI Congrès de la SFSIC Création, créativité et médiations*. Paris, France : MSH Paris Nord, juin 2018. URL : <https://hal.archives-ouvertes.fr/hal-01817597> (visité le 20/08/2021).
- [Fac07] Patrizia FACCIOLI. « La sociologie dans la société de l'image ». fr. In : *Societes* no 95.1 (mars 2007), p. 9-18. ISSN : 0765-3697. URL : <https://www.cairn.info/revue-societes-2007-1-page-9.htm> (visité le 07/08/2021).
- [Fek15] Jean-Daniel FEKETE. « Recherche en visualisation d'information ou Dataviz : pourquoi et comment ? » fr. In : *I2D — Information, donnees documents* Volume 52.2 (juill. 2015), p. 32-33. ISSN : 2428-2111. URL : <https://www.cairn.info/revue->

- i2d-information-donnees-et-documents-2015-2-page-32.htm (visité le 19/08/2021).
- [Fre15] Sylvia FREDRIKSSON. « Du design d'information à la visualisation de données : un enjeu de transmission de sens auprès de la société civile ». fr. In : *I2D — Information, donnees documents* Volume 52.2 (juill. 2015), p. 36-36. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-36.htm> (visité le 19/08/2021).
- [Fri06] Michael FRIENDLY. « A Brief History of Data Visualization ». In : *Handbook of Computational Statistics : Data Visualization*. Sous la dir. de C. CHEN, W. HÄRDLE et A UNWIN. T. III. Heidelberg : Springer-Verlag, 2006.
- [Fri08] Michael FRIENDLY. « The Golden Age of Statistical Graphics ». In : *Statistical Science* 23.4 (2008). Sous la dir. d'Institute of MATHEMATICAL STATISTICS, p. 502-535. DOI : 10.1214/08-STS268. URL : https://arxiv.org/PS_cache/arxiv/pdf/0906/0906.3979v1.pdf.
- [FD20] Michael FRIENDLY et Daniel DENIS. *Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization. An illustrated chronology of innovations by Michael Friendly and Daniel J. Denis*. 2020. URL : <http://datavis.ca/milestones/index.php>.
- [Gir21] Guillaume GIROUD. « L'interface de la représentation, représentations de l'interface ». fr. In : *Interfaces numériques* 10.1 (avr. 2021). ISSN : 2259-1001. DOI : 10.25965/interfaces-numeriques.4542. URL : <https://www.unilim.fr/interfaces-numeriques/index.php?id=4542> (visité le 11/08/2021).
- [Git03] Todd GITLIN. *Sommersi dai media*. it. Milan : Etas, 2003.
- [Goë19] Samuel GOËTA. *Rapport sur l'open data*. fr. Rapp. tech. Dataactivist, pour le Conseil départemental de l'Ardèche, 2019. URL : <https://dataactivist.coop/ardeche/rapport/partiel.html#aux-origines-de-louverture-des-donn%C3%A9es-retours-sur-quelques-grandes-dates-fondatrices> (visité le 07/08/2021).
- [GL06] Jack GOODY et Jean-Claude LEJOSNE. « La technologie de l'intellect ». fr. In : *Pratiques* 131.1 (2006), p. 7-30. DOI : 10.3406/prati.2006.2114. URL : https://www.persee.fr/doc/prati_0338-2389_2006_num_131_1_2114 (visité le 22/02/2021).
- [GWL06] Jack GOODY, Ian WATT et Jean-Claude LEJOSNE. « Les conséquences de la littératie. Jack Goody et Ian Watt ». fr. In : *Pratiques* 131.1 (2006), p. 31-68. DOI : 10.3406/prati.2006.2115. URL : <https://www.persee.fr/doc/>

- prati_0338-2389_2006_num_131_1_2115 (visité le 22/02/2021).
- [Gou15] Caroline GOULARD. « Les interactions hommes-données mettent l'Humain au cœur de l'industrie des données ». fr. In : *I2D — Information, données documents* Volume 52.2 (juill. 2015), p. 62-63. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-62.htm> (visité le 19/08/2021).
- [Gra+13] Jonathan GRAY et al. *Guide du datajournalisme. Collecter, analyser et visualiser les données*. Eyrolles, 2013. 220 p. ISBN : 9782212136852. URL : <https://jplusplus.github.io/guide-du-datajournalisme/index.html>.
- [GRI19] Groupe de recherche interuniversitaire sur l'intégration pédagogique des technologies de l'information et de la communication GRIIPTIC. *Principales réalisations*. 2019. URL : <https://griiptic.ca/realisations.html> (visité le 23/08/2021).
- [Gui19] Éric GUICHARD. « Les nouveaux maîtres de l'écriture du monde ». In : *Contemporary French and Francophone Studies. The Google Era ? / L'Ère Google ?* 23.4 (2019), p. 490-501. DOI : 10.1080/17409292.2020.1723247. URL : <https://hal.archives-ouvertes.fr/hal-02310616> (visité le 23/08/2021).
- [Hac15] Hakim HACHOUR. « De la fouille à la visualisation de données : un processus interprétatif ». fr. In : *I2D — Information, données documents* Volume 52.2 (juill. 2015), p. 42-43. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-42.htm> (visité le 19/08/2021).
- [Har20] Florian HARMAND. « Ce que font les données au journalisme : Un point de vue du design ». fr. In : *Interfaces numériques* 9.3 (déc. 2020). ISSN : 2259-1001. DOI : 10.25965/interfaces-numeriques.4435. URL : <https://www.unilim.fr/interfaces-numeriques/index.php?id=4435> (visité le 20/08/2021).
- [HB01] Mountaz HASCOËT et Michel BEAUDOUIN-LAFON. « Visualisation interactive d'information ». fr. In : *Revue-i3* 1 (2001), p. 77-108. URL : <http://www.lirmm.fr/~mountaz/Publi/Upload/revueI3-hascoet.pdf>.
- [Her07] Clarisse HERRENSCHMIDT. *Les trois écritures : Langue, nombre, code*. Bibliothèque des sciences humaines. Paris, France : Gallimard, 2007. ISBN : 9782070760251.
- [Jea19] Yves JEANNERET. « Chapitre 4. Recourir à la démarche sémiocommunicationnelle dans l'analyse des médias ». fr. In : *Médias et médiatisation*. Presses universitaires de Grenoble, 2019. ISBN : 9782706142802. URL : <https://www.cairn.info/>

- medias - et - mediatisation -- 9782706142802 - page-105.htm (visité le 22/08/2021).
- [Kor09] Roselyne KOREN. « Le récit de chiffres : enjeux argumentatifs de la « narrativisation » des chiffres dans un corpus de presse écrite contemporain ». fr. In : *A contrario* n° 12.2 (déc. 2009), p. 66-84. ISSN : 1660-7880. URL : <https://www.cairn.info/revue-a-contrario-2009-2-page-66.htm> (visité le 23/08/2021).
- [Laf19] Benoit LAFON. « Présentation ». fr. In : *Médias et médiatisation*. Presses universitaires de Grenoble, 2019. ISBN : 9782706142802. URL : <https://www.cairn.info/medias-et-mediatisation--9782706142802-page-17.htm> (visité le 22/08/2021).
- [Lec21] Michelle LECOLLE. *Public (lexique)*. fr-FR. 2021. URL : <http://publictionnaire.huma-num.fr/notice/public-lexique/> (visité le 22/08/2021).
- [LU08] Sylvie LELEU-MERVIEL et Philippe USEILLE. « Quelques révisions du concept d'information ». In : *Problématiques émergentes dans les sciences de l'information*. Sous la dir. d'HERMÈS. Traité des sciences et techniques de l'information. Lavoisier, 2008, p. 25-56. URL : <https://hal.archives-ouvertes.fr/hal-00695777> (visité le 22/02/2021).
- [Lit09] Marc LITS. « La médiatisation du politique ou le passage d'un espace public délibératif à un espace public symbolique narratif ». fr. In : *A contrario* n° 12.2 (déc. 2009), p. 85-100. ISSN : 1660-7880. URL : <https://www.cairn.info/revue-a-contrario-2009-2-page-85.htm> (visité le 23/08/2021).
- [Man11] Lev MANOVICH. « What Is Visualisation ? » In : *Visual Studies* 26 (mars 2011), p. 36-49. DOI : 10.1080/1472586X.2011.548488.
- [Mat98] James R. MATEY. « Visual Explanations : Images and Quantities, Evidence and Narrative ». In : *Computers in Physics* 12.2 (mars 1998), p. 146-148. ISSN : 0894-1866. DOI : 10.1063/1.168637. URL : <https://aip.scitation.org/doi/abs/10.1063/1.168637> (visité le 20/08/2021).
- [Mé179] Jaques MÉLÈSE. *Approches systémiques des organisations*. Editions Hommes et Techniques. Suresnes, 1979.
- [Met70] Christian METZ. « Au-delà de l'analogie, l'image ». fr. In : *Communications* 15.1 (1970), p. 1-10. DOI : 10.3406/comm.1970.1212. URL : https://www.persee.fr/doc/comm_0588-8018_1970_num_15_1_1212 (visité le 08/08/2021).
- [Mug06] Mioara MUGUR-SCHÄCHTER. *Sur le tissage des connaissances*. Hermès. Paris, 2006.
- [Nor93] Donald NORMAN. *Raisons pratiques*. T. 4. Les artefacts cognitifs. 1993, p. 15-34.

- [Paq16] Françoise PAQUIENSÉGUY. « La datavisualisation des portails open data : une forme de médiation du et par le chiffre? » In : *Datalogie : Formes et imaginaires du numérique*. Sous la dir. d'Avenati Olaf & Chardel PIERRE-ANTOINE. Editions LOCO, mars 2016, p. 126-136. URL : <https://hal.archives-ouvertes.fr/hal-01552803> (visité le 22/02/2021).
- [Péq21] Julien PÉQUIGNOT. « Le numérique comme objet intellectuel ». fr. In : *Interfaces numériques* 10.1 (avr. 2021). ISSN : 2259-1001. DOI : 10.25965/interfaces-numeriques.4510. URL : <https://www.unilim.fr/interfaces-numeriques/index.php?id=4510> (visité le 11/08/2021).
- [Pla05] William PLAYFAIR. *An Inquiry Into the Permanent Causes of the Decline and Fall of Powerful and Wealthy Nations : Designed to Shew how the Prosperity of the British Empire May be Prolonged*. en. Google-Books-ID : MLvIAAAAMAAJ. Greenland et Norris, 1805.
- [Ren19] Sandra RENDGEN. *History of Information Graphics*. Sous la dir. de Julius WIEDMANN. Köln : Taschen, 2019. 462 p. ISBN : 978-3-8365-6767-1.
- [Rey15] Everardo REYES. « La datavisualisation comme image-interface ». fr. In : *I2D — Information, données documents* Volume 52.2 (juill. 2015), p. 38-39. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-38.htm> (visité le 19/08/2021).
- [Rin+15] Yves RINATO et al. « De l'usage à la forme : le design pour créer du sens ». fr. In : *I2D — Information, données documents* Volume 52.2 (juill. 2015), p. 66-67. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-66.htm> (visité le 19/08/2021).
- [Seu14] Christine SEUX. « Écran(s) ». fr. In : *Le Telemaque* n° 45.1 (juin 2014), p. 15-25. ISSN : 1263-588X. URL : <https://www.cairn.info/revue-le-telemaque-2014-1-page-15.htm> (visité le 11/08/2021).
- [Spi21] Virginie SPIES. *Grand public*. fr-FR. 2021. URL : <http://publicationnaire.huma-num.fr/notice/grand-public/> (visité le 22/08/2021).
- [The20] Anaïs THEVIOT. « Data-journalisme et pratiques infographiques : entre mise en concurrence et résistances des journalistes politiques locaux ». fr. In : *Interfaces numériques* 9.3 (déc. 2020). ISSN : 2259-1001. DOI : 10.25965/interfaces-numeriques.4469. URL : <https://www.unilim.fr/interfaces-numeriques/index.php?id=4469> (visité le 20/08/2021).
- [] *Tufte on Design and Data*. URL : <https://sites.google.com/site/tufteondesign/> (visité le 17/08/2021).

- [Vial3] Stéphane VIAL. *L'être et l'écran*. fr. Presses Universitaires de France, 2013. ISBN : 9782130621706. DOI : 10.3917/puf.vials.2013.01. URL : <https://www.cairn.info/l-etre-et-l-ecran--9782130621706.htm> (visité le 10/08/2021).
- [Vial4] Stéphane VIAL. « Le design, un acte de communication ? » fr. In : *Hermès, La Revue*. 70^e sér. 3 (2014), p. 174-180. URL : <https://www.cairn.info/revue-hermes-la-revue-2014-3-page-174.htm>.
- [Wik21] WIKIPEDIA. *Public*. fr. Page Version ID : 182695594. Mai 2021. URL : <https://fr.wikipedia.org/w/index.php?title=Public&oldid=182695594> (visité le 22/08/2021).
- [Zaz15] Ornella ZAZA. « Vers un Open data visuel : le portail Open Data Paris ». fr. In : *I2D — Information, données documents* Volume 52.2 (juill. 2015), p. 53-53. ISSN : 2428-2111. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-53.htm> (visité le 19/08/2021).

ENTRETIENS

PRESENTATION

Tout d'abord, merci beaucoup de m'accorder cet entretien.

1. Me présenter

Je me présente : je suis **Stéphanie Louison**, et je suis étudiante à **PENSSIB** (école d'information-documentation) en **Master 2 de Publication Numérique**.

2. Présenter l'objectif de l'étude

Dans le cadre de mon **mémoire** de fin d'études, je mine actuellement une enquête sur les **datavisualisations**, et en particulier sur leur **réception**.

3. Expliquer à la personne comment, pourquoi et par quel intermédiaire on l'a contactée

C'est pour cela que je t'ai demandé de m'accorder cet entretien : en tant que **datajournaliste**, tu es un **spécialiste des datavisualisations**, mais aussi de leur **transmission à un public**.

4. Préciser les modalités de l'enquête

Toutes les réponses que tu me donneras seront utilisées dans un **cadre exclusivement universitaire**. Es-tu d'accord pour poursuivre l'entretien sachant qu'il sera **enregistré** ? Il est néanmoins possible, si tu préfères, que cet entretien soit **anonymisé**.

- Comment s'organise votre **travail au quotidien** ? Avez-vous des **contraintes temporelles** (une datavisualisation par jour par exemple) ? Des réunions d'équipes ? Des brainstormings ?
- **Qui choisit le thème** qu'aborderont vos datavisualisations : faites-vous des propositions, ou est-ce l'équipe éditoriale qui vous pose une commande ?
- D'ailleurs, quels sont les **thèmes les plus généralement abordés** par vos datavisualisations ?

PRATIQUES DE REALISATIONS DE DATAVISUALISATIONS

- Pourrais-tu me décrire le **processus pour réaliser une datavisualisation** ? Que faites-vous concrètement ?
- Si nous reprenons toute la chaîne de la donnée (que j'aimerais mieux comprendre) :
 - **Qui sélectionne les données** ? Pourriez-vous me citer des sources de données que vous avez déjà utilisées ?
 - **Comment traitez-vous les données** ? Avec quels logiciels, quelles méthodes (statistiques...) ?
 - Comment choisissez-vous comment les **représenter graphiquement** (designer, sans artistique de chacun,

TALON SOCIOLOGIQUE

Je te laisse te présenter.

- **Depuis quand travailles-tu au Parisien ?**
- Comment en es-tu **venu à travailler ici** ? Comment t'es-tu **formé** à la réalisation de représentations graphiques de données ? As-tu travaillé dans **d'autres secteurs précédemment** ?
- Peux-tu me présenter ton **poste** et ton rôle au Parisien ? Quelles sont tes **missions** ? Comment décrirais-tu ton **métier** ?
- D'ailleurs, comment appelles-tu ce que tu fais : de la **datavisualisation**, du **design d'information**, de l'**infographie de presse** ?

CADRE PROFESSIONNEL ET ECOSYSTEME

- Je crois que tu travailles avec une **équipe**. Faites-vous tous le même travail, ou avec chacun une spécialité ?
- Avec quels **autres corps de métier** travaillez-vous ? (rédaction du journal, service web, marketing, développeurs externes, designers...)

brainstorming d'équipe, s'adapter à un modèle dans un logiciel, suivre des recommandations éditos... ?

- Privilégiez-vous des **modèles de graphiques « classiques »** (courbes, histogrammes...) ou préférez-vous **innover** avec des **formats originaux (lesquels ?)** ?
- Quelle importance accordez-vous à l'**interactivité** ? En quoi est-ce important ou non (lecteur plus libre de découvrir, plus de données, plus attractif, plus nouveau...) ?
- Quel est le **lien entre le texte d'un article et la datavisualisation** ? Penses-tu qu'il faille un texte autour de la datavisualisation, ou une bonne datavisualisation doit-elle se suffire à elle-même ? Au contraire, le texte « polie »-t-il la datavisualisation ? SI ! y a un texte, qui le rédige ? Un journaliste ou un développeur de datavis ?

RETOUR SUR CES PRATIQUES

- **À qui s'adressent** les datavisualisations ?
 - A tous ? A un certain public ? Si oui, lequel ?
 - Penses-tu que tes datavisualisations soient **compréhensibles pour tous** ? Ou penses-tu que certaines personnes y soient plus sensibles que d'autres ?
 - Penses-tu que **certaines compétences (en lectures de graphiques, de cartes...)** sont nécessaires à vos récepteurs pour comprendre les informations que vous

voulez transmettre ? Ou pense-tu, au contraire, que **sans aucune formation à la lecture de graphique**, n'importe quel enfant puisse lire tes datavisualisations ?

- Sont-elles créées justement dans ce but, avec une dimension « **vulgarisatrice** » ?
- Qu'est-ce qui rend tes datavisualisations accessibles à tous à ton avis ?
- As-tu parfois des **retours de lecteurs / d'usagers** sur tes productions ? Si oui, qu'en disent-ils ? Ont-ils l'impression de mieux comprendre certains points ? Ces remarques sont-elles semblables les mêmes que pour des articles écrits ?
 - Penses-tu qu'il existe une **demande** pour les datavisualisations ? Les utilisateurs attendent-ils cela, ou se contenteraient-ils plus volontiers d'articles bien rédigés ?
 - Pourquoi avoir créé un **compte Twitter** spécifiquement pour vos infographies ? Qu'a-t'en occupé, d'ailleurs ?
- Je crois savoir que la notion d'**objectivité** est une notion floue des pratiques journalistiques : penses-tu que les datavisualisations soient **plus objectives** que les articles **journalistiques** ? Ou qu'elles soient **moins** ataquables ou remises en cause par les lecteurs ?
- Quelle part occupe le **design** dans ton approche de la datavisualisation ? Est-il important ? Si oui, sous quel aspect (attraitif, facile d'utilisation, soigné...) ?

- Pour conclure : **Qu'est-ce qu'une bonne datavisualisation selon toi ?**

Merci infiniment pour cet échange qui m'a fait beaucoup avancer sur ma connaissance et ma compréhension de ce sujet complexe.