

MEMOIRE

*EN QUOI LA
REPRESENTATION DES
DONNEES NECESSITE-T-
ELLE UN APPRENTISSAGE
?*

CLAIRE
DARRIGADE

SEPTEMBRE 2018

Table des matières

Introduction.....	2
1. La visualisation des donnees : un enjeu historique.....	6
1.1 L'histoire des représentations graphiques	6
1.2 Etat de l'art	9
1.3 L'apprentissage : la litteratie en visualisation	12
2. Les mecanismes et pratiques de la visualisation des données	16
2.1 Les mecanismes : entre l'economie de l'attention et la psychologie cognitive... 16	
2.2 Le processus de creation d'une visualisation de données	21
3. Cas pratique : creation de visualisations de données.....	25
3.1 Le contexte du processus de creation	25
3.2 La creation de visualisations de données	29
3.2.1 L'étendue de la cote de confiance des Présidents de la République pendant leur mandat.....	32
3.2.2 Classement des Présidents de la République selon leur cote de confiance par année de mandat.....	34
3.2.3 L'évolution de cote de confiance mensuelle des présidents de la République depuis 1981.....	36
Conclusion.....	39
Références.....	42

INTRODUCTION

« LA PLUS GRANDE VALEUR D'UNE IMAGE, C'EST QUAND ELLE NOUS OBLIGE A REMARQUER CE QUE NOUS NE NOUS ATTENDIONS PAS A VOIR » JOHN TUCKEY, EXPLORATORY DATA ANALYSIS (1977)

Les systèmes de visualisation sont issus des civilisations ancestrales. Par exemple, l'écriture sumérienne est la langue écrite connue la plus ancienne. Ses premiers témoignages datent de -3300 avant JC. Il s'agit d'une époque très prospère pour la Mésopotamie et ce particulièrement dans la cité d'Uruk en Iraq - l'une des plus puissantes villes de la civilisation Sumer - où les échanges commerciaux sont florissants. L'écriture sumérienne est alors inventée par un groupe d'administrateurs qui travaillaient au sanctuaire principal de la ville et utilisaient l'écriture comme un outil de gestion des échanges.

Cette écriture se traduisait par des symboles logographiques - c'est-à-dire des dessins représentatifs d'une notion ou d'une suite phonique constituée par un mot¹. Un peu moins d'un millier de signes ont été énumérés, ils étaient gravés à l'aide d'un roseau - le calame - sur des tablettes d'argile humide puis séchées au soleil. Cette écriture a été qualifiée de cunéiforme puisque l'utilisation du calame générait ces formes en coin des différents signes.



¹ <http://www.cnrtl.fr/definition/logogramme>

Des éléments de vie ont donc été conceptualisés et retranscrits de manière visuelle : par exemple, le mouton était représenté par une croix entourée d'un cercle. Dans le contrat ci-contre², les comptables ont détaillé la valeur de la vente d'un champ et d'une maison. Ce type de contrat permettait de vérifier les affirmations de chacun en cas de litige.

L'écriture sumérienne peut dès lors être considérée comme la première retranscription visuelle connue de la réalité en signes conceptualisés. Elle facilitait la compréhension et officialisait la valeur des échanges en représentant les biens - et donc les données - de manière visuelle permettant ainsi de rendre tangibles ces informations. Elle illustre l'importance de la place du visuel comme moyen d'échanges et de transmission de savoir, ce depuis plus de 3000 ans avant JC.

Quelques millénaires plus tard, la thématique de la visualisation des données est plus que jamais omniprésente.

C'est avec l'arrivée du Big Data - expression datant de 1997 qui désigne l'ensemble des données numériques produites par l'utilisation des nouvelles technologies à des fins personnelles ou professionnelles³ - que l'intérêt porté au traitement de grands nombres de données augmente. Puisque l'ère du numérique ne signifie pas uniquement l'augmentation du volume des données mais aussi l'accroissement de leur variété et de la vitesse à laquelle elles sont produites, récoltées et analysées⁴, l'enjeu repose désormais sur l'exploitation de ces données afin de les rendre lisibles et d'en tirer des informations utiles notamment pour la prise de décisions. Aujourd'hui, seulement 1% des données produites dans le monde sont exploitées⁵.

Tandis que le principe de la société d'hyperconsommation (Lipovetsky, 2006) est de démultiplier la quantité de biens à consommer et que les économistes étudient ces ressources matérielles, la nouvelle économie de l'attention porte sur « le temps

²Contrat archaïque concernant la vente d'un champ et d'une maison Vers 2600 avant J.-C. Shuruppak. Musée du Louvre.

³<http://www.futura-sciences.com/tech/definitions/informatique-big-data-15028/>

⁴Laney D. (2001). 3D Data Management : Controlling Data Volume, Velocity, and Variety. *Meta Group*.

⁵Vincenot P. (2016). La dataviz : aide à la décision et à la compréhension. *I2D - Informations, données et documents*, volume 53, p. 54-55

d'attention qui constitue aujourd'hui l'objet d'appropriation central »⁶. Ce nouveau modèle économique est fondé sur les techniques de la rhétorique : comment accrocher l'écoute ou le regard ; comment captiver l'attention afin de marquer l'esprit pour qu'il reste attentif et que le message s'imprime en lui ? C'est au centre de ce type de problématiques qu'entre en jeu la visualisation des données qui a pour but de « faciliter la compréhension des données en les retranscrivant efficacement sous forme visuelle »⁷. En ce sens, « efficacement » signifie que la lecture de la représentation produite doit se faire sans effort, de manière presque automatique et ce en utilisant les réflexes cognitifs du système visuel. Certaines caractéristiques visuelles dites « pré-attentives » (Treisman, 1986), comme les couleurs ou les rayons d'une courbure, sont perçues par l'œil humain à une vitesse inférieure à 200 millisecondes, de manière presque inconsciente. Dans ce cas, il est donc possible de raisonner sur un nombre important de données. Au contraire, une représentation qui n'est pas pré-attentive augmente le temps d'acquisition mental qui devient alors proportionnel au nombre de données présentées.

Dans le processus de création d'une représentation visuelle, la détermination de la cible et donc du public est essentielle. En effet, à mesure que la visualisation des données se propage dans les médias et les entreprises, il semblerait que les lecteurs ne soient pas encore habitués à ce nouveau langage. Bien que s'appuyant sur le système réflexif visuel humain, l'utilisation des codes graphiques de la visualisation des données ne peut être interprétée que si elle est connue du lecteur. C'est pourquoi ce mémoire cherchera à répondre en trois temps à la question suivante : *en quoi la représentation des données nécessite-t-elle un apprentissage ?* En premier lieu, la réflexion présentera un état de l'art de la visualisation des données, en retraçant l'évolution historique des représentations graphiques, les différents courants de pensées concernant la visualisation des données et comment ce phénomène nécessite un temps d'apprivoisement de nouveaux codes et donc un apprentissage en littératie visuelle⁸. Une seconde partie exposera les pratiques de la visualisation de données en allant des mécanismes de l'économie de l'attention au processus de création d'une data visualisation. Enfin, cette recherche aboutira à un cas

⁶ Citton, Y. (2012). Traiter les données : entre économie de l'attention et mycélium de la signification. *Multitudes*, 49,(2), 143-149. doi:10.3917/mult.049.0143.

⁷ Lagnel J.M. (2017). Manuel de data visualisation.

⁸ Terme définie par l'International Visual Literacy Association comme un « *groupe de compétences acquises en vue de l'interprétation et de la production de messages visuels* ».

concret de création de visualisations de données illustrant l'utilité de cette pratique et la nécessité d'un apprentissage qu'elle engendre.

1. LA VISUALISATION DES DONNÉES : UN ENJEU HISTORIQUE

Cette première partie a pour but de d'ancrer le sujet dans son contexte historique. Puisqu'il n'existe pas à ce jour de définition officielle de la visualisation des données, il convient de présenter les évolutions des représentations graphiques mais aussi des différents courants théoriques qui accompagnent ce développement pour enfin, réaliser l'ampleur des enjeux de la visualisation des données.

1.1 L'HISTOIRE DES REPRÉSENTATIONS GRAPHIQUES

Cette partie retrace l'évolution de la visualisation des données et présente les grandes découvertes ayant mené à nos modes de représentations graphiques actuels.

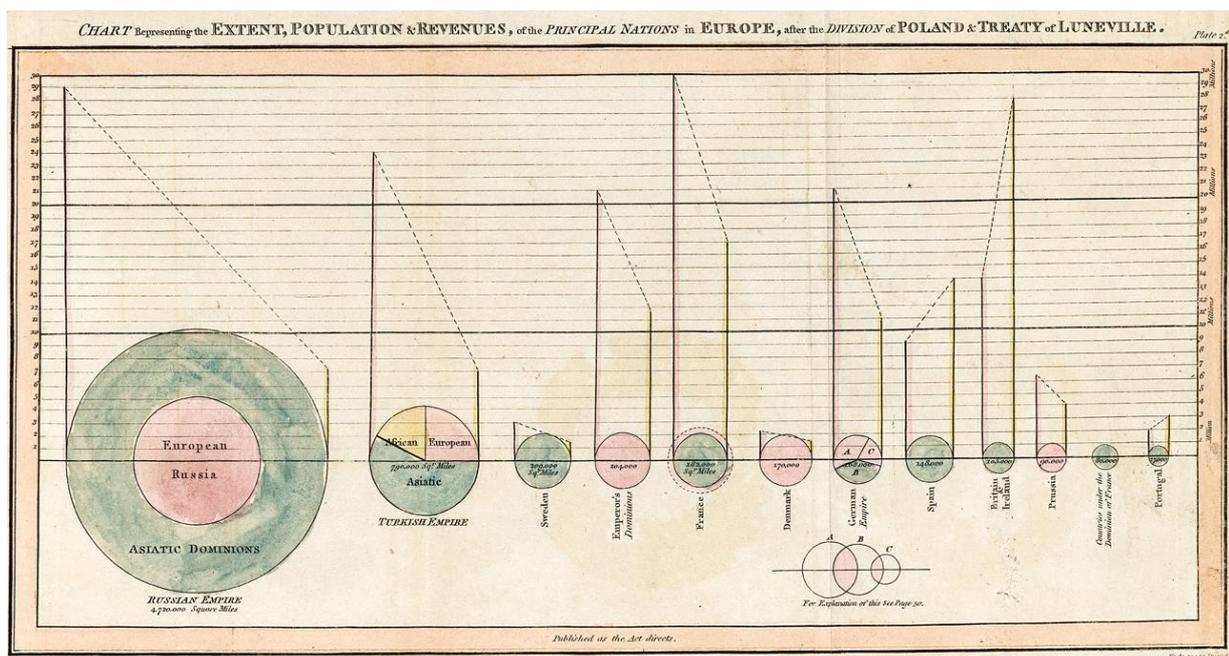
Si les systèmes de comptage existaient avant la naissance de l'écriture, le développement des représentations graphiques de données apparaît au Moyen Age, en 1370. L'un des premiers à représenter sous forme graphique le rapport entre deux variables est Nicole Oresme. Il utilise deux axes perpendiculaires, qu'il appelle *latitudo et longitudo*, pour représenter les variations de la chaleur d'un corps en fonction de l'endroit où on la mesure. Par la suite il développera les premières représentations des mouvements dans son *Tractatus de configuratione qualitatum et motuum*. Il y exprime dès lors l'intérêt de la représentation visuelle des données pour la compréhension de phénomènes divers :

« Les propriétés de cette qualité, en seront examinées plus clairement et plus facilement dès lors que quelque chose qui lui est semblable est dessiné en une figure plane, et que cette chose, rendue claire par exemple visible, est saisie rapidement et parfaitement par imagination ...car l'imagination des figures aide grandement à la connaissance des choses mêmes »⁹.



⁹ <https://www.lmno.cnrs.fr/archives/oresme/Oresme.html>

Enfin, le troisième scientifique ayant eu un impact particulier sur l'évolution des représentations graphiques est William Playfair. Il est le pionnier des graphiques statistiques et notamment de 3 grands types de visualisations classiques utilisés de manière très fréquente de nos jours : les courbes, les histogrammes et les graphiques sectoriels. En ce qui concerne les courbes, il invente plus précisément la série statistique chronologique sous forme de courbe. Playfair est un économiste écossais du XVIII^{ème} siècle (1759-1823), il rédige un premier ouvrage *l'Atlas commercial et politique* dans lequel 44 graphiques sont publiés. Cet ouvrage porte sur le commerce et les créances britanniques de 1700 à 1782, il décide d'illustrer ses dires par des graphiques expliquant l'activité économique au fil du temps. Dans un second ouvrage, *The statistical Breviary (1801)*, il publie des diagrammes statistiques regroupant un grand nombre de données. Comme par exemple ce diagramme montrant l'étendue de la population et les revenus des principaux pays européens en fonction de leur taille. Les barres verticales rouges représentent la population en millions d'habitants et les barres vertes le total des taxes collectées en millions de livres sterling. La superficie des pays est représentée par des cercles de tailles différentes. Enfin, la ligne en pointillés met en relation le nombre d'habitants et la valeur des taxes. Bien que la pente de cette droite ne soit pas significative, le fait que celle-ci soit orientée vers la droite ou bien vers la gauche permet une interprétation. Par exemple, la Grande Bretagne est le pays avec les taxes les plus élevées par rapport à sa population.



Playfair expliqua ses démarches de la sorte : « *De tous les sens, c'est la vue qui donne l'idée la plus exacte et la plus prompte de tout ce qui est susceptible de lui être représenté ; et quand il s'agit de reconnaître la proportion qui existe entre diverses quantités, ou diverses grandeurs, l'œil a une supériorité étonnante pour la saisir. Par l'habitude constante et presque involontaire de comparer divers objets, il acquiert une facilité étonnante à les juger.* »¹⁰

Les créations graphiques se sont donc développées depuis plusieurs siècles déjà et l'importance de la visualisation dans la compréhension, la transmission et l'apprentissage est une idée déjà intégrée dans nos sociétés. Cependant, les modèles graphiques classiques existants comportent certaines limites, notamment lorsqu'il s'agit de représenter des jeux de données complexes. C'est là tout l'enjeu de la visualisation des données.

1.2 ETAT DE L'ART

Les avancées réalisées en représentations graphiques ont amené plusieurs théoriciens à en officialiser les règles de bonnes pratiques. Il existe par ailleurs, plusieurs courants de pensées qui s'opposent autour de l'utilisation, abusive pour certains, de la visualisation des données. Cette partie présente donc un résumé des débats actuels autour de la visualisation des données.

Aujourd'hui, la visualisation des données est un sujet de débat perpétuel. Certains dénoncent des pratiques dénuées de sens, n'utilisant pas les bons codes et ne permettant pas de faciliter la compréhension d'un message – comme Bertin ou Tufte – qui centrent leurs propos sur l'utilisation de la représentation graphique comme moyen de valorisation des données uniquement. D'autres prônent la beauté des données avant tout et tendent à accentuer le côté artistique de la visualisation des données, comme David McCandless. Dans le cadre de cette recherche, la visualisation des données sera définie comme une représentation visuelle permettant de faciliter la compréhension d'un ensemble de données.

¹⁰ Playfair, W. (1802). *Eléments de statistique*.

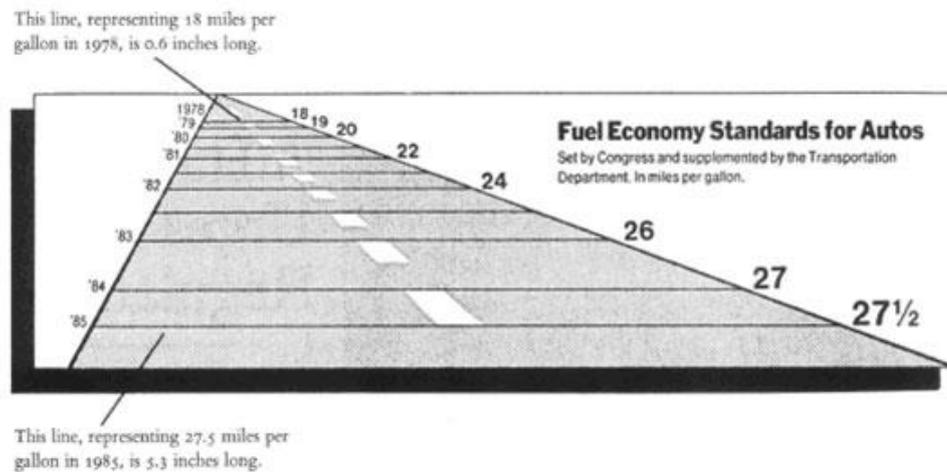
Jacques Bertin est un géographe français qui a théorisé la sémiologie graphique, qu'il définit comme « *l'ensemble des règles d'un système graphique de signes pour la transmission d'une information* » en 1967 dans son ouvrage *Sémiologie graphique*. Cette œuvre décrit les règles de bonnes pratiques à mettre en place pour réaliser des visualisations graphiques d'informations. Cet ouvrage est donc une référence des techniques et méthodes de représentation qui influe encore les spécialistes actuels. Il y explique que le choix des formes, des couleurs et de l'organisation des représentations graphiques sont essentiels et ne peuvent être réalisés de manière aléatoire. Chaque décision prise lors du processus de création d'un graphique doit avoir du sens. Il critique les représentations « dénuées de sens », qui selon lui ne prennent pas en compte le double attachement entre utilité et simplicité dans le graphique. Bertin fait partie de ceux qui pensent que les graphiques sont fréquemment utilisés de manière abusive et constate par ailleurs que « *les cartes et les diagrammes inutiles sont encore très nombreux* »¹¹.

Edward Tufte est l'une des principales figures de la visualisation des données et de cette lutte contre les dérives. Il théorise et conceptualise la visualisation des données en tant que telle. Tufte développe le principe de « *data-ink ratio* » qui consiste à montrer les données avant tout et de supprimer le superflu. C'est-à-dire que toute encre utilisée pour autre chose que l'écriture des données et des échelles n'étant pas essentielle à la compréhension du graphique doit être retirée. Selon lui, cette encre superflue pourrait détourner l'attention du lecteur et l'attirer vers des éléments non pertinents pour la transmission du message. Cet indicateur est développé dans son premier ouvrage - *The visualisation of quantitative data* – tout comme le concept du « *lie factor* ». C'est une valeur permettant d'évaluer la relation entre la taille d'un effet affiché dans un graphique et la taille de l'effet réel dans les données. Il part du principe que la représentation des nombres mesurée physiquement sur le graphique doit être proportionnelle aux quantités représentées. Par exemple, Tufte analyse le graphique ci-dessous paru dans le New York Times.

Il calcule le facteur de mensonge – « *lie factor* » - de la manière suivante :

¹¹ Bertin J. (1981). Théorie matricielle de la graphique. *Communication et langages*, vol. 48, n° 1, p. 62-74.

$$\text{Lie factor} = \frac{\text{size effect shown in graphic}}{\text{size effect in data}} = \frac{(5,3 - 0,6)/0,6}{(27,5 - 18)/18} = \frac{7,83}{0,53} = 14,8$$



Ainsi dans ce graphique, les standards d'économie d'essence pour les automobiles entre 1978 et 1985 ont augmenté de 53% tandis que le graphique représente via les lignes horizontales une augmentation de 783%. Le « facteur de mensonge » est évalué à 14,8. D'après Tufte, il est essentiel pour la bonne compréhension d'une visualisation graphique que ce facteur soit compris entre 0,95 et 1,05¹². Au-delà et en-deçà de ces valeurs, un graphique peut être considéré comme distordu et donc mensonger.

Pendant que certains prônent l'utilité et l'efficacité des représentations minimalistes, un courant plus récent lancé par David Mc Candless, auteur de l'ouvrage *Information is beautiful* publié en 2009, développe les concepts de design et de beauté de l'information dans les visualisations de données. Mc Candless explique que l'intérêt de la visualisation des données est de raconter des histoires avec les données et donc de représenter les relations entre les données afin de permettre des comparaisons de manière aisée. Ce courant considère que la beauté des représentations permet la séduction du public. C'est cette séduction qui amplifie alors l'attractivité de visualisation de données. Ce phénomène est appelé l'esthétique informationnelle. D'après les recherches menées par Norman (2002), lier le design et l'émotion à la visualisation de données permettrait d'augmenter l'efficacité de celle-ci en accentuant son attractivité. Cela engendrerait une meilleure mémorisation ainsi qu'une interprétation plus juste. Cependant le risque de rendre la visualisation inaccessible est important - comme l'explique Alain Garnier (2015) - lorsqu'une visualisation est belle, elle provoque une incitation à essayer de comprendre.

¹² <https://moz.com/blog/data-visualization-principles-lessons-from-tufte>

Mais le risque de cette "séduction par les sens" est qu'elle soit mal interprétée par l'esprit voire inintelligible. Ce qui détourne du but premier de la visualisation des données.

C'est dans ce cadre que l'apprentissage entre en jeu en permettant à tous lecteurs, qu'ils soient experts ou non, d'avoir les moyens de comprendre une visualisation de données.

1.3 L'APPRENTISSAGE : LA LITTÉRATIE EN VISUALISATION

La visualisation des données doit permettre une lecture rapide et efficace, qui améliore la mémorisation des informations transmises via l'utilisation de codes visuels communs. Pourtant, plusieurs études montrent qu'il n'est pas si facile de bouleverser les codes de la graphique¹³ « classique » et de rendre plus accessibles à tout un chacun les informations représentées dans une visualisation (Drot-Delange, 2016). Cette popularisation de la visualisation des données se doit d'être accompagnée et soutenue par l'apprentissage de ces nouveaux codes, c'est-à-dire de la littératie visuelle.

La première définition de la littératie visuelle est la suivante : « *ensemble de compétences visuelles qu'un être humain développe par le regard et d'autres expériences sensorielles* » (Debes, 1969). Celle-ci a évolué et depuis 2001, elle est définie par l'International Visual Literacy Association comme un « *groupe de compétences acquises en vue de l'interprétation et de la production de messages visuels* » (Brill, Kim et Branch, 2001). L'idée étant la création d'un référentiel commun permettant à tous l'appréhension d'un message visuel. Ces compétences peuvent être qualifiées de transversales puisqu'elles recourent différentes disciplines comme la statistique, le design et l'informatique.

Parce que la visualisation des données se propage dans les communications institutionnelles comme dans les médias, l'enjeu de la lecture et de la compréhension de ces représentations est essentiel. Cette propagation peut être facilement identifiée dans les médias, où le développement du journalisme de données - mouvement visant à exploiter les données statistiques pour les mettre à disposition du grand public - fait rage.

¹³ « La graphique utilise les propriétés du plan pour faire apparaître les relations de ressemblance, d'ordre ou de proportionnalité entre des ensembles donnés [...] la graphique désigne un système de signes ». (Bertin, 1970).

Par exemple, les journaux *Le Monde*¹⁴, *Libération* ou le *New York Times* ont lancé des rubriques spécialisées dans ce domaine. Une émission télévisée *#DataGueule* diffusée sur France 4 propose des vidéos décryptant l'actualité via des animations de données¹⁵. Les entreprises comme par exemple Air France utilisent aussi la visualisation des données en tant qu'outil de communication. Les visualisations de données permettent des communications simples publiables sur les réseaux sociaux et notamment Twitter. L'enjeu de la lecture et de la compréhension de ces représentations est donc bien essentiel. Pour s'assurer de la bonne transmission d'un message, il convient préalablement de le concevoir de manière objective et universelle de sorte qu'il ne souffre aucune mésinterprétation. Or la construction d'une visualisation de données s'apparente à un processus de création, à chaque étape duquel des choix subjectifs de représentation sont faits et qui peuvent conduire à compromettre la visée objective initiale. C'est à ce niveau que prend place l'apprentissage de la littératie en visualisation, les individus doivent être dotés des compétences permettant d'appréhender, d'interpréter mais aussi de porter un regard critique sur leur lecture. Comme disait Jacques Bertin en 1981 sur ses lecteurs : « *Mais connaissent-ils la graphique et font-ils la différence entre la graphique et le graphisme ?* »¹⁶. L'apprentissage de la littératie en visualisation permet aux individus d'être capables d'observer les différences entre déformations liées aux données et déformations liées au style du designer, ce qui est une des clés de compréhension de la visualisation des données. Et pour ce faire, chacun doit être doté de la capacité à recontextualiser les données d'un graphique.

Comme l'expliquent plusieurs articles du journal *Le Monde*, les médias et les entreprises publient beaucoup de graphiques faussés par exemple en utilisant des échelles distordues. Grace à l'apprentissage en littératie visuelle, les lecteurs pourraient avoir un regard critique face à ces graphiques et en comprendre le sens réel. Par exemple, le journal *Le Point* avait repris une publication trompeuse d'Air France publiée de 2015¹⁷. Sur la visualisation de données communiquée les écarts entre les deux barres de chacun

¹⁴ https://www.lemonde.fr/les-decodeurs/article/2014/12/31/des-chiffres-et-des-images-l-annee-2014-en-12-datavisualisations_4539835_4355770.html

¹⁵ <https://education.francetv.fr/matiere/education-aux-medias/premiere/programme/datagueule>

¹⁶ Bertin J. (1981). Théorie matricielle de la graphique. *Communication et langages*, vol. 48, n° 1, p. 62-74.

¹⁷ https://www.lemonde.fr/les-decodeurs/article/2015/10/26/pourquoi-il-faut-se-mefier-des-graphiques_4797002_4355770.html

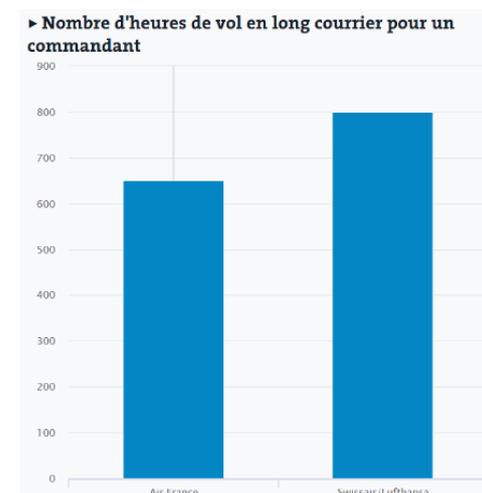
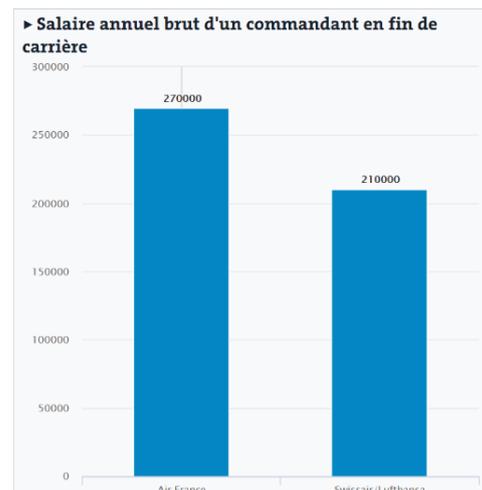
des graphiques sont accentués. *Le Point* a donc décidé de reproduire des graphiques « corrects » et de les présenter en comparaison de la publication d'Air France. Les différences sont très importantes.

Publication d'Air France

Air France, des chiffres qui parlent



Publication du Point



Sur le premier graphique représentant le salaire annuel brut d'un commandant en fin de carrière, l'écart entre les données des compagnies aériennes est de 28% mais il est représenté de manière équivalente à un écart de 200% sur la visualisation des données. De la même manière, le second graphique représentant le nombre d'heures de vol en long-courrier pour un commandant a été distordu. L'écart entre les données est de 23% mais le graphique présente un écart de 200%.

C'est dans ce contexte que s'ancre l'apprentissage en littératie visuelle. Les communications de données des entreprises, institutions et médias sont diffusées aisément de nos jours. Il s'agit le plus souvent de visualisations de données, permettant d'assurer l'efficacité des communications, publiées sur les réseaux sociaux ou bien sur des sites internet. L'enjeu de doter les lecteurs des compétences pour interpréter ces visualisations est donc de plus en plus présent et l'apprentissage est une clé d'amélioration de la situation actuelle.

2. LES MECANISMES ET PRATIQUES DE LA VISUALISATION DES DONNEES

2.1 LES MECANISMES : ENTRE L'ECONOMIE DE L'ATTENTION ET LA PSYCHOLOGIE COGNITIVE

L'une des plus anciennes disciplines connues – la rhétorique – se donnait pour objet la captation de l'attention du public. Les principes de la rhétorique recourent aujourd'hui les enjeux de la communication : comment accrocher l'écoute et le regard ? Comment présenter des données de manière simplifiée afin de frapper l'esprit du lecteur ? La genèse de la visualisation des données s'immisce dans les problématiques de l'économie de l'attention. C'est aujourd'hui le temps de captation de l'attention qui importe, que ce soit pour l'efficacité publicitaire ou pour le temps de lecture et de compréhension d'un graphique dans un journal. La démultiplication de l'utilisation des graphiques dans les médias accentue d'autant plus cet enjeu. Par exemple auprès des instituts de sondage : comment produire des rapports efficaces, plus pertinents que ceux des concurrents ? C'est du côté des théories de la psychologie cognitive qu'il faut s'appuyer pour construire de véritables outils de communication efficaces et donc pertinents. La pertinence/efficacité d'une représentation visuelle des données doit être mesurable suivant différents critères :

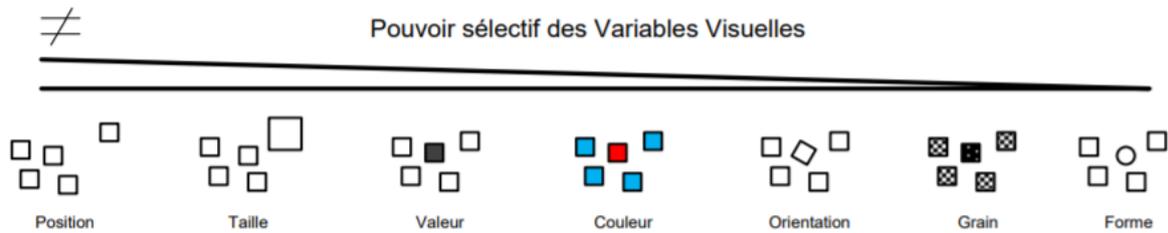
1. L'utilisation des caractéristiques pré-attentives est essentielle ainsi que le respect des règles et codes graphiques.

La notion de perception pré-attentive désigne la capacité du système perceptif humain à analyser certaines caractéristiques des représentations graphiques de manière très rapide, c'est-à-dire en moins de 250 millisecondes, sans effort et donc de façon presque automatique. Ces composantes graphiques pré-attentives telles que la position, la taille, la couleur, l'orientation ou la forme¹⁸ sont dites sélectives. Elles possèdent la capacité de

¹⁸ Hurter, C. (2010). *Caractérisation de visualisations et exploration interactive de grandes quantités de données multidimensionnelles* (Doctoral dissertation, Université Paul Sabatier-Toulouse III).

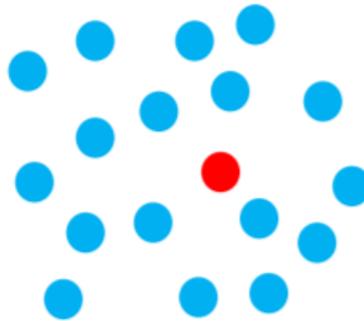
distinguer un objet parmi d'autres sans avoir à tous les tester. Voici sur la figure 1 le pouvoir sélectif des différentes variables visuelles de manière décroissante.

Figure 1 : Bertin, 1983



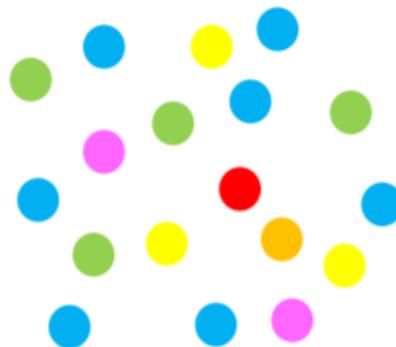
Par exemple, la perception d'un point rouge parmi un ensemble de points bleus est immédiate (voir Figure 2).

Figure 2



Cependant, cette perception pré-attentive de la couleur est limitée. La même expérience reproduite avec des cercles de différentes couleurs (voir Figure 3) prouve que la distinction du cercle rouge demande plus de temps. Il est nécessaire d'analyser chaque élément séquentiellement de manière non-automatique et contrôlée.

Figure 3



A l'opposé de la perception automatique, la perception contrôlée est beaucoup plus lente et mobilise entièrement le système cognitif. Il s'agit par exemple de la lecture d'un texte.

2. Le temps de lecture et de compréhension d'une visualisation de données est un critère d'efficacité

La puissance d'une image est incontestable : 70% des récepteurs sensoriels sont situés dans les yeux et 50% du cerveau est dédié au traitement de l'image¹⁹. Cela ne veut pas dire que toute chose vue est aisément comprise ou mémorisée. Dans son ouvrage *Le Chant du signe. Aventures et mésaventures de nos interprétations quotidiennes*, Lionel Naccache explique la puissance des signes visuels que l'on perçoit quotidiennement. Il explique : «*Les signes symboliques qui nous entourent ont été conçus pour éveiller en nous une interprétation absolument claire, univoque et indiscutable de leur signification : un feu vert, un feu rouge [...] la croix verte de la pharmacie, les petits rectangles verticaux de nos téléphones portables ... Aussitôt perçus, nous voyons bien ce qu'ils veulent dire* ». De la même manière, une visualisation de données cherche à éveiller ce type d'interprétation. L'attention humaine a récemment été mesurée comme égale à 8 secondes²⁰ - c'est-à-dire inférieure à celle d'un poisson rouge – d'où le défi de faire passer un message en un laps de temps très restreint. Les visualisations de données transforment les codes graphiques classiques et la perception de ce qui autrefois avait une signification « univoque » se voit désormais modifiée. Le temps de lecture et de compréhension d'une visualisation de données est donc bien un critère d'efficacité et un enjeu à relever.

3. Le lecteur ne doit pas se perdre, le message transmis doit être clair et simple. Cela implique un travail de mise en page et de valorisation de l'information²¹.

Effectivement, la lecture d'un support se fait de la manière suivante : l'entrée du support se fait de manière naturelle dans le coin supérieur gauche, puis poursuit en

¹⁹ Lagnel JM. (2017). *Manuel de data visualisation*.

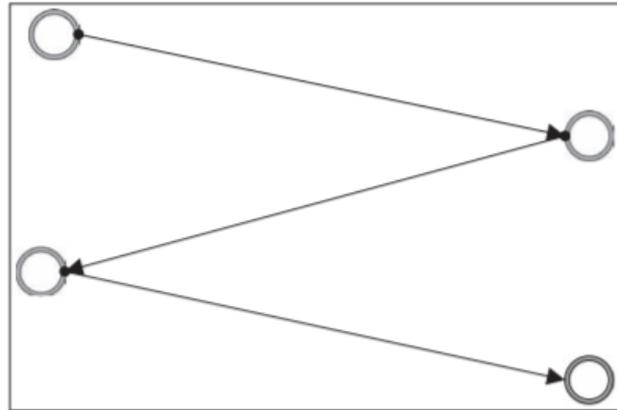
²⁰ https://www.francetvinfo.fr/replay-radio/nouveau-monde/moins-de-concentration-que-les-poissons-rouges-a-cause-des-ecrans_1778207.html

²¹ Lebellet, B. (2012). *Convaincre avec des graphiques efficaces: sous Excel™, PowerPoint™, Tableau™...* Editions Eyrolles.

diagonal jusqu'au premier tiers droit, revient en diagonal sur le second tiers gauche pour enfin finir dans le coin inférieur droit.

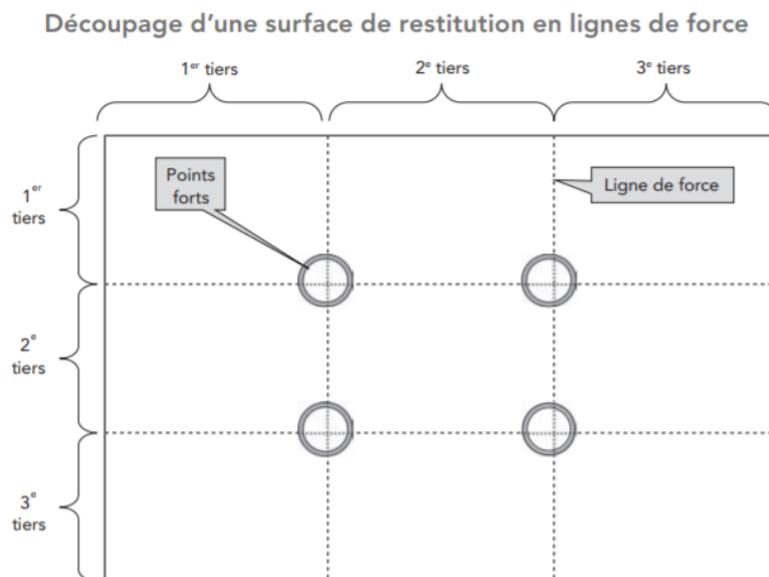
Figure 4

Parcours classique de l'œil



Parallèlement, la disposition des formes doit être organisée judicieusement de manière à améliorer la mémorisation et la compréhension du message. Ceci nécessite de s'appuyer sur les grands principes de décomposition d'une image. En découpant en tiers verticaux et tiers horizontaux, il est possible d'identifier les « lignes de forces » et « points de force » d'une page, endroits où les principaux axes de lecture doivent être placés.

Figure 5



Par ailleurs, lors de la création d'une visualisation, il semble important de se reporter à la théorie de la Gestalt ou la psychologie des formes, développée au XX^{ème} siècle en Allemagne. Ce système explique que le cerveau cherche automatiquement à structurer et organiser les éléments complexes pour les mettre en forme. La pensée saisirait donc d'abord une image globale de haut en bas, avant d'analyser la somme des parties. La structuration des formes ne se ferait pas de manière aléatoire mais serait régulée par différentes lois dites naturelles. Il en existe six différentes :

- la loi de la bonne forme ;
- la loi de la continuité ;
- la loi de la proximité ;
- la loi de la similitude ;
- la loi de destin commun ;
- la loi de clôture.

Parmi ces lois, celle de la proximité et de la similitude peuvent être particulièrement pertinentes pour la visualisation des données. Celles-ci expliquent que « *nous regroupons d'abord les points les plus proches les uns des autres* »²² - loi de la proximité – et que « *si la distance ne permet pas de regrouper les points, nous nous attachons ensuite à repérer les éléments les plus similaires entre eux pour percevoir une forme* » - loi de la similitude.

Figure 6 : Loi de la proximité

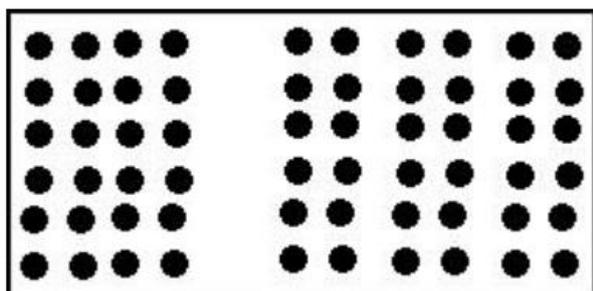
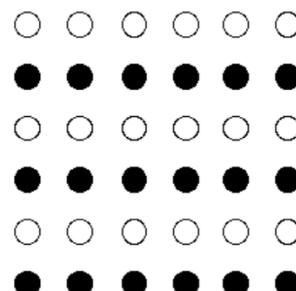


Figure 7 : Loi de la similitude



Les mécanismes de captation de l'attention d'un individu sont fortement liés à la psychologie cognitive de l'être humain. Il convient donc de se référer au fonctionnement de notre système cognitif et aux règles qui ont été établies afin de produire un message visuel percutant.

²² https://iris.univ-tlse2.fr/moodle-ent/pluginfile.php/180517/mod_resource/content/0/Gestalt.html

2.2 LE PROCESSUS DE CREATION D'UNE VISUALISATION DE DONNEES

Le traitement des données ne peut débuter que lorsqu'un angle d'analyse est défini. Cet angle est caractérisé par une problématique qui permet de guider tout le processus de création d'une visualisation de données. Il permet de savoir ce que l'on cherche à montrer. Par ailleurs, il est impératif de cerner son public : un même sujet peut être traité différemment en fonction de la cible. Savoir à qui la visualisation des données va s'adresser est donc impératif, cela permet d'adapter la conception au lecteur et d'éviter qu'il doive s'adapter à la visualisation. Plus le public est spécialisé, plus le rendu peut être complexe.

La première étape est de récolter et de nettoyer les données²³. Soit le client possède déjà une base de données – il est tout de même nécessaire d'en vérifier le contenu avant de se lancer dans la production - soit il est nécessaire d'en constituer une en rassemblant des informations. La seconde étape est de retravailler le fichier en changeant les libellés, mettre à jour certaines informations, s'imprégner de la base de données en identifiant le type de données des variables. Une fois que ce premier travail est réalisé, le processus de traitement de données consiste à chercher des liens, des relations entre les différentes variables de la base de données. Il faut nécessairement garder en tête, la problématique de départ, c'est-à-dire ce que l'on cherche à expliquer et à transmettre et à quel type de public.

Il convient par la suite d'analyser les données et d'aboutir à des interprétations qui seront choisies pour être par la suite représentées dans la visualisation de données. Le traitement des données consiste à chercher si les données peuvent être comparées par catégorie, avec des proportions ou bien s'il existe des relations entre elles. Des problèmes peuvent survenir pendant cette étape comme par exemple la présence de valeurs manquantes dans la base ou bien des formats des variables non adaptées qu'il faut alors nettoyer en tranchant sur la méthode à choisir. Lorsque le traitement des données porte sur un nombre important de variables, il est parfois intéressant d'utiliser des indicateurs synthétiques tels que la moyenne, la médiane ou encore l'écart-type.

²³ Arruabarrena, B. (2015). L'écosystème numérique de la datavisualisation. *I2D-Information, données & documents*, 52(2), 56-58.

Le choix d'utilisation d'un graphique dépend de ce qui a été mis en lumière pendant le traitement des données²⁴. Pour comparer des valeurs entre plusieurs catégories, les types variés de barres sont appropriés (barre multiples, empilées, verticales circulaires ou encore radiales).

Figure 8 : Barres verticales circulaires

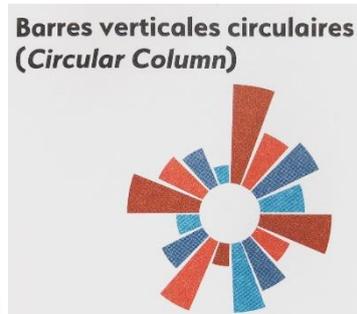
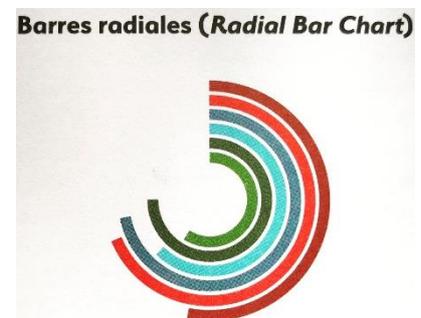


Figure 9 : Barres empilées



Figure 10 : Barres radiales



Pour représenter des évolutions dans le temps ou des tendances, les histogrammes, les courbes ou les streamgraphs sont les modèles les plus pertinents.

Figure 11 : Courbes

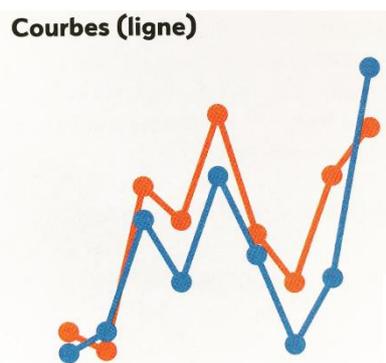


Figure 12 : Histogramme

Histogramme (Histogram)

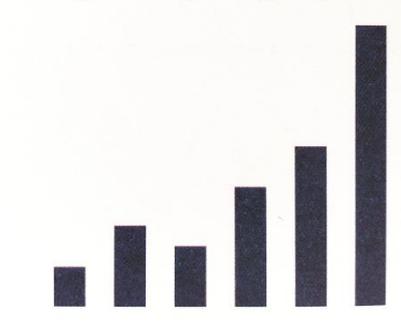
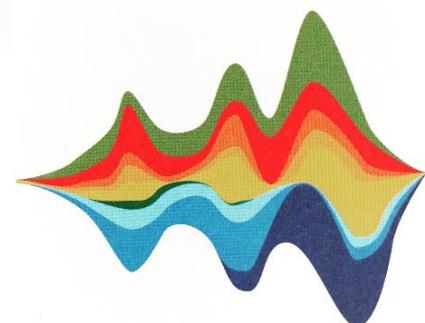


Figure 13 : Streamgraph

Streamgraph (Streamgraph)



²⁴ Lagnel JM. (2017), *Manuel de datavisualisation*.

La seconde étape consiste à choisir le support. Bien souvent, le client aura déjà exposé en amont ses choix concernant le support. Cette phase de la conception est particulièrement importante. Chaque support comporte des contraintes à prendre en compte. Une visualisation fixe (papier ou web) est ce qui est le plus rapidement concevable. Une visualisation destinée à être imprimée comporte des contraintes : d'une part il n'y aura pas d'interactivité ni d'animation et d'autre part, le risque de manquer de place rapidement est assez important. Dans ce cas, le volume du texte est souvent un élément problématique, il faut aller rapidement à l'essentiel et bien agencer l'espace. Il est parfois préférable de produire une visualisation interactive, notamment quand les données sont complexes et nombreuses. Par exemple, une visualisation avec un grand nombre de courbes peut rapidement devenir illisible sur papier tandis qu'une version interactive – où le passage de la souris sur la courbe indique la valeur de celle-ci - peut permettre une lecture facilitée. Pour réaliser ces visualisations il existe différentes outils et logiciels.

L'outil le plus communément utilisé pour les représentations graphiques reste Excel. Il permet de produire des graphiques simples qui peuvent être améliorés manuellement. L'un des avantages d'Excel est l'automatisation de la production de graphiques à l'aide de macros. Ce système permet de produire des graphiques très rapidement en mettant à jour uniquement les nouvelles données, ce qui est très pratique pour la production des baromètres en institut de sondage. Cependant, l'offre des types de graphiques est assez restreinte et le choix n'est pas aussi important que ce que proposent d'autres outils. Voici un aperçu des différents outils permettant de créer des visualisations de données.

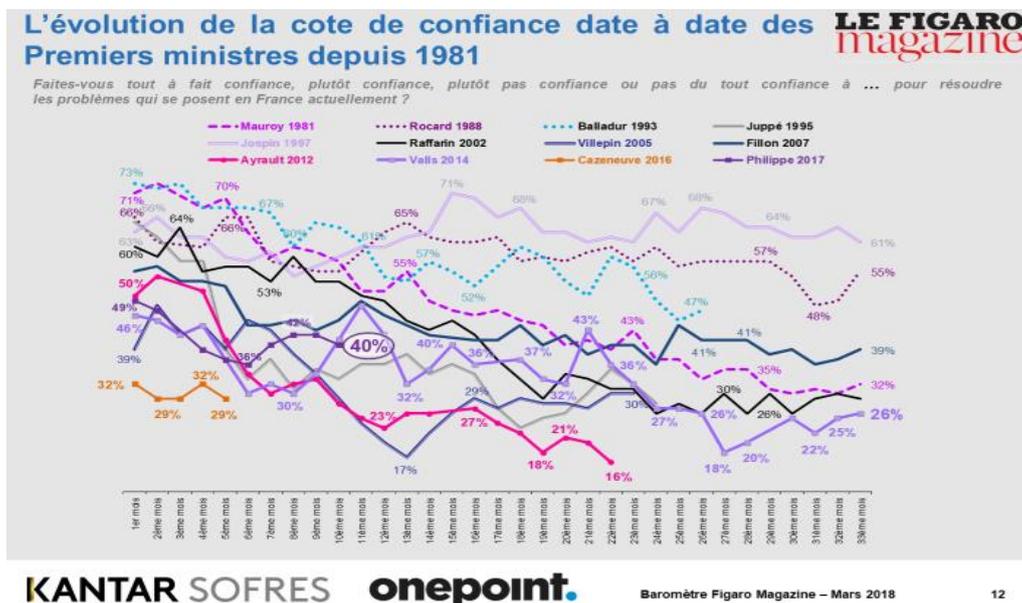
- Afin de visualiser rapidement des données de manière assez simple, l'utilisation de Google Chart Tools peut être utile dans un premier temps.
- L'outil Infogram, est un des outils web les plus utilisés. La prise en main est assez simple et propose une tablette de graphiques assez complète. Il existe une version payante professionnelle offrant plus d'options.
- Enfin, le logiciel Tableau Public est le plus connu des outils de visualisation de données. C'est un logiciel gratuit, assez puissant, qui permet de créer des

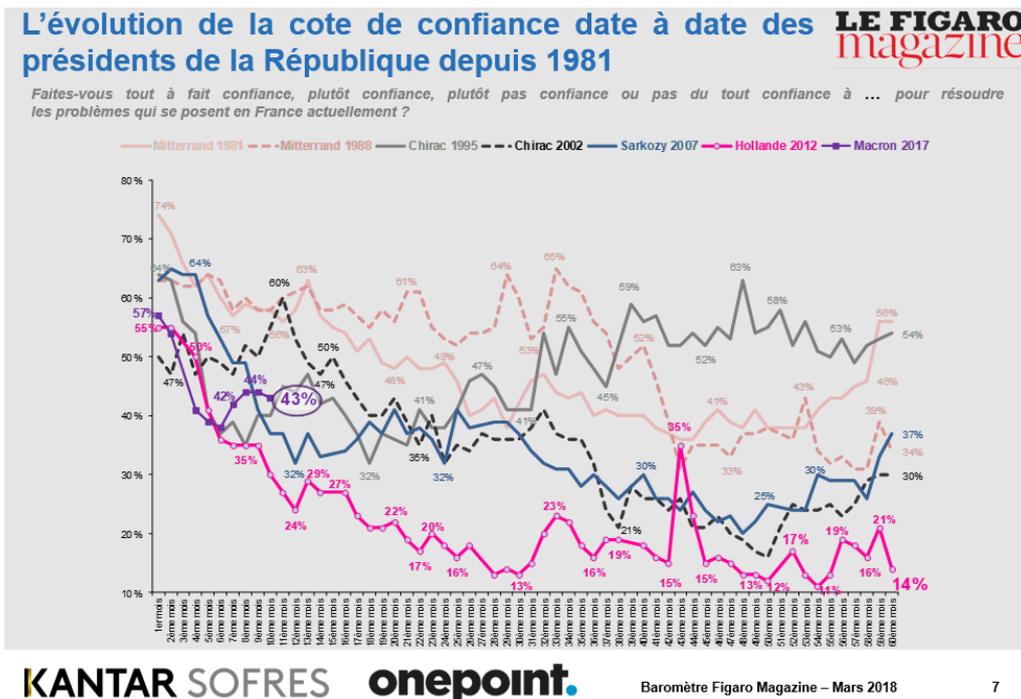
visualisations complexes et interactives. Ce logiciel est utilisé par beaucoup de professionnels de l'information comme Orange, BNP Paribas ou encore *The Guardian* et *le Wall Street Journal*. L'interface est plutôt complexe de prime abord et peu engageante, elle nécessite un certain temps de prise en main. Mais il existe beaucoup de tutoriels très complets pour aider à avancer dans la conception des visualisations de données.

3. CAS PRATIQUE : CREATION DE VISUALISATIONS DE DONNEES

3.1 LE CONTEXTE DU PROCESSUS DE CREATION

Le baromètre politique du figaro magazine est l'une des vitrines de Kantar Public. C'est une étude quantitative menée auprès d'un échantillon représentatif de la population de 1000 personnes. Depuis quelques années, le baromètre est réalisé tous les mois. Il mesure dans une première partie la confiance des Français dans le couple exécutif puis, dans une seconde partie il présente les cotes d'avenir des personnalités politiques et enfin dans la troisième partie il dévoile l'opinion sur les différents partis. Ce baromètre est publié dans les médias pour le grand public mais il est avant tout utilisé par les politologues. La demande de renouvellement du visuel du baromètre du *Figaro Magazine* provient d'une requête en interne, au sein de l'équipe Kantar Public France. La chargée d'étude qui produit ce baromètre cherche à moderniser le rapport, l'ensemble des graphiques, le template c'est-à-dire la trame de fond du rapport pour le rendre plus attractif avec un problème majeur concernant deux courbes rendues complètement illisibles par le surplus de données. L'une concerne l'évolution de la cote de confiance des différents Premiers ministres depuis 1981 et l'autre celle des Présidents. Les données sont issues de l'étude du mois de mars 2018. L'enjeu est ici de rendre ces courbes plus lisibles via une visualisation de données adaptée.





Le déroulement de l'expérience réalisée s'est fait en plusieurs étapes.

- La première étape a été la définition des objectifs du projet :

Le public : un même sujet est traité de manière différente suivant son public. Un des problèmes du baromètre Figmag est de viser deux types de publics différents : le grand public et les spécialistes/politologues. Tandis que pour ces derniers, les données publiées doivent être très précises et complexes (lisible au date à date), les lecteurs du grand public ont besoin d'une lecture simplifiée et donc synthétisée. L'enjeu est donc de trouver une alternative aux deux courbes présentées ci-dessus pouvant être à la fois interprétée par le grand public et satisfaisant les besoins des spécialistes.

Les contraintes : le livrable est un format Power Point. Les visualisations de données réalisées seront donc statiques et au format A4, imprimables. Par ailleurs, le rapport est produit de manière quasi-automatique à partir d'une maquette et d'une macro programmée sur support Office. En raison du manque de temps entre la récupération des données et le rendu du livrable, il est essentiel que la chargée d'études produisant le rapport n'ait rien à ajouter de façon manuelle dans le livrable. Tout doit être automatisé. Enfin, le baromètre est réalisé depuis une trentaine d'années, les modifications apportées ne peuvent pas être trop extrêmes et la question de la nécessité d'un accompagnement du public se pose.

L'objectif: transformer les courbes de cote de confiance pour les rendre plus lisibles sans perdre de l'information.

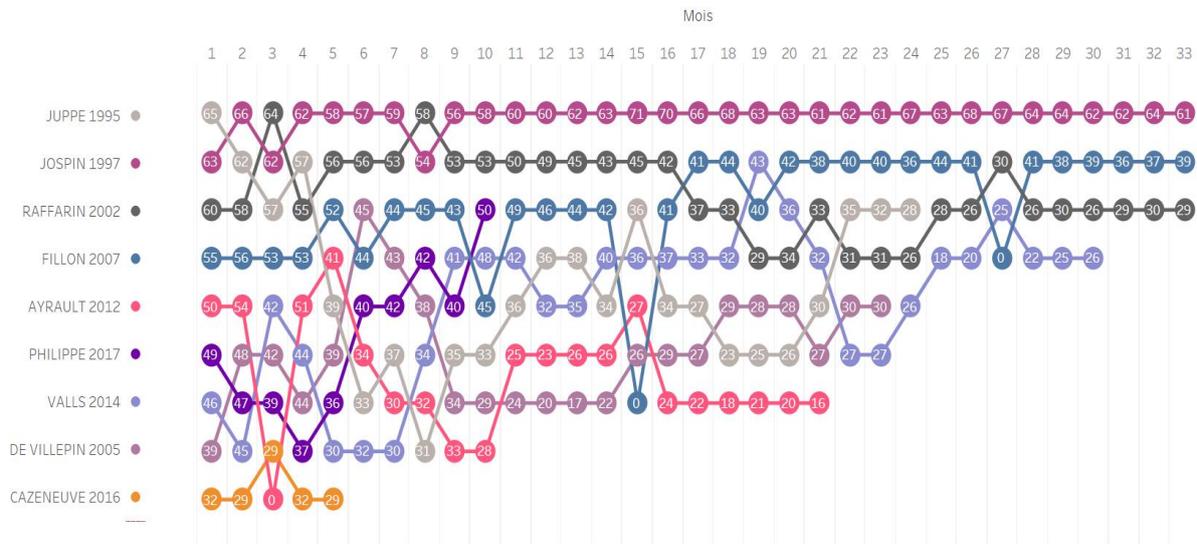
Les courbes représentant l'évolution de la cote de confiance des Premiers ministres ou des Présidents permettaient trois types d'analyse :

1. Le premier axe est celui de l'évolution dans le temps : une courbe permet de visualiser rapidement la tendance de la cote de confiance des différentes personnalités dans le temps ;
2. La seconde analyse des courbes est celle de la comparaison entre les différentes séries. Par exemple, la comparaison de la cote de confiance des différents Premiers Ministres ou Présidents permet d'apprécier la popularité de ces personnalités les unes par rapport aux autres ;
3. Enfin, le troisième axe d'analyse est celui de la fluctuation et donc de la dispersion des données d'une série qui représentent les variations temporelles de la popularité d'une personnalité. Il est toujours intéressant de repérer dans une courbe, et dans ce cas pour une personnalité donnée le point minimal et maximal de la cote de confiance.

- La seconde étape a été la présentation d'une piste de visualisation de données

Après avoir réfléchi sur le sujet, plusieurs éléments d'analyse ont été soulevés. L'idée était de comprendre quelles étaient les informations présentées sur les courbes, et qu'est-ce qu'il fallait essayer de traduire à l'aide d'une visualisation de données. Le premier axe d'analyse des courbes qui a été sujet à la recherche d'une visualisation adaptée a été celui de la comparaison de la cote de confiance des personnalités à une date donnée. Cela a abouti à l'envie de réaliser un classement des personnalités par date, suivant leur cote de confiance. En s'inspirant d'un graphique en pente – ou Slopegraph – inventé par Edward Tufte en 1983, le but était de montrer les variations entre les différentes situations dans le temps. Ce type de graphique permet de visualiser facilement un changement de position dans un classement dans le temps. Dans un premier temps, la recherche a porté sur la courbe de l'évolution de la cote de confiance des Premiers Ministres depuis 1995 sur les 30 premiers mois au pouvoir. Voici ce que donnait cette première visualisation de données :

Premiers Ministres depuis 1995



L'un des problèmes majeurs dans cette visualisation de données était la volonté de garder toute l'information, c'est-à-dire de vouloir d'une part faire un classement et d'autre part d'afficher la valeur de la cote de confiance. Cette première piste de visualisation de données a été présentée à la responsable de la production de cette étude.

Plusieurs difficultés ont été soulevées lors de cette réunion :

- L'affichage de la valeur de la cote de confiance peut provoquer des incompréhensions. Par exemple, la première ligne du classement est composée de valeurs de cote de confiance très différentes pouvant aller de 71% à 56%. Ainsi, l'affichage de la valeur de la cote de confiance perturbe la lecture du classement. D'autre part, le trop grand nombre d'informations représentées sur ce graphique entraîne un manque de clarté ;
- Le changement brutal d'un graphique produit depuis 30 ans n'est pas facile à accepter et nécessite un temps d'adaptation et d'acceptation pour tous les lecteurs, que ce soit en interne ou en externe. La crainte par ailleurs, d'une incompréhension du public doit être prise en compte. Il faut donc penser à réaliser un accompagnement permettant l'apprentissage des nouveaux codes graphiques utilisés.

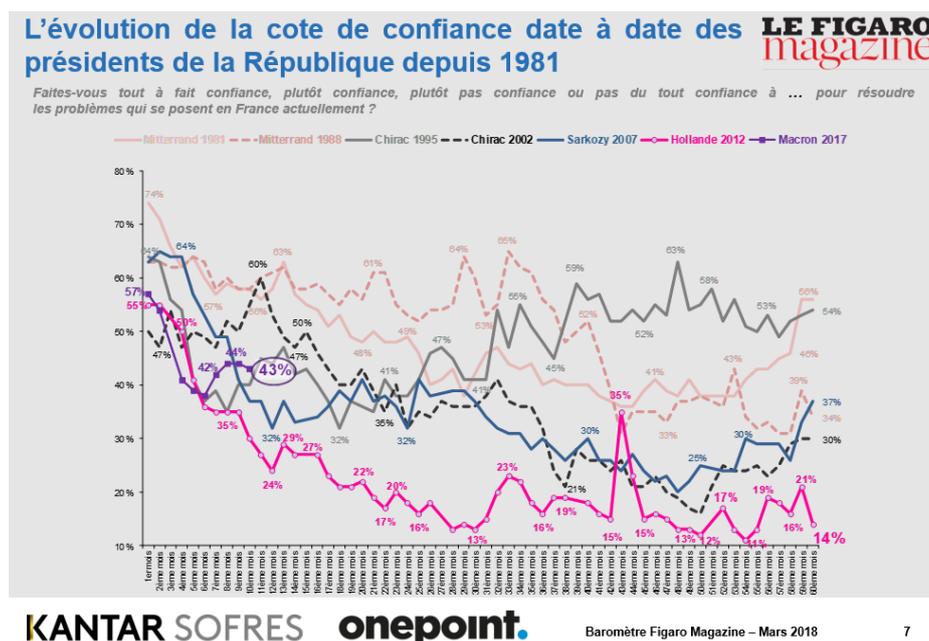
- Le logiciel utilisé pour la création de la représentation visuelle s'appelle Tableau Public. Il est gratuit mais ne permet pas de produire des visualisations de données de manière automatique. Il faut donc chercher des alternatives sur l'outil Excel.

- Troisième étape : les modifications du projet initial

A la suite de cette première réunion, il a été beaucoup plus difficile de dégager du temps pour échanger sur le projet. Par ailleurs, la restriction d'une création de visualisation de données à l'aide du logiciel Excel était assez contraignant et ne permettait pas de mettre en lumière la réflexion présentée dans ce mémoire. C'est pourquoi j'ai décidé de continuer ma réflexion sur cette problématique tout en utilisant l'outil Tableau Public, qui offre une palette de charte bien plus importante. Dès lors, le projet a pu prendre une autre direction. L'objectif est donc de rester sur le même contexte, avec l'enjeu d'un double public à satisfaire, des trois axes de lecture des courbes à clarifier et simplifier tout en mettant de côté la restriction des outils.

3.2 LA CREATION DE VISUALISATIONS DE DONNEES

Pour la création d'une visualisation de données adaptée à la problématique, les données utilisées sont celles de la cote de confiance des Présidents de la République depuis 1981. Ces données ont été produites au mois de mars 2018 par Kantar Public France. Il est possible de dupliquer les visualisations de données créées pour la courbe de l'évolution de la cote de confiance des Premiers Ministres depuis 1981.



Le principal problème de ce graphique est son manque de clarté. Les différentes courbes se confondent ce qui amplifie la difficulté de la lecture.

Il convient donc de rappeler les trois axes d'analyse de cette courbe qui ont été dégagés un peu plus tôt. Cette courbe présente 3 grands axes d'analyse ou de lecture :

1. L'évolution globale de la cote de confiance des personnalités dans le temps ;
2. La comparaison entre les différentes personnalités ;
3. Les fluctuations et la dispersion de la cote de confiance des personnalités.

La visualisation des données à produire comporte donc 3 thèmes à illustrer :

1. La tendance de la cote de confiance des différentes personnalités ;
2. Le classement des personnalités par popularité ;
3. Le résumé de la dispersion de la cote de confiance des différentes personnalités pendant leur mandat.

Dans chacune de ces visualisations des données le choix des couleurs a été fait en fonction des partis politiques des personnalités. Faire correspondre la gamme de couleur entre les trois représentations était nécessaire pour créer une sorte d'homogénéité de l'information. Dans la courbe initiale, les axes en noir étaient très visibles et obstruaient légèrement la lecture du graphique. Le choix d'utiliser une trame gris clair permet de valoriser les données représentées et de ne pas encombrer la visualisation tout en gardant une certaine structure classique. L'importance du nombre de séries statistiques a amené à l'utilisation d'indicateurs synthétiques tels que le minimum, le maximum et la moyenne pour l'une des visualisations de données. La quantité d'informations présente dans la courbe de l'évolution de la cote de confiance des Présidents de la République a engendré une segmentation des axes de lecture et donc la production de trois visualisations de données complémentaires illustrant de manière simplifiée les différentes entrées de lecture de la courbe.

Les choix qui ont été faits pour réaliser ces visualisations de données portent plus sur les besoins d'une cible grand public que ceux d'experts politologues. Effectivement, le grand public a besoin d'une information simplifiée et pas particulièrement précise tandis que les politologues recherchent une information précise et complexe. C'est pourquoi l'idée était d'essayer de combiner les besoins de ces deux cibles. Il a fallu tout d'abord

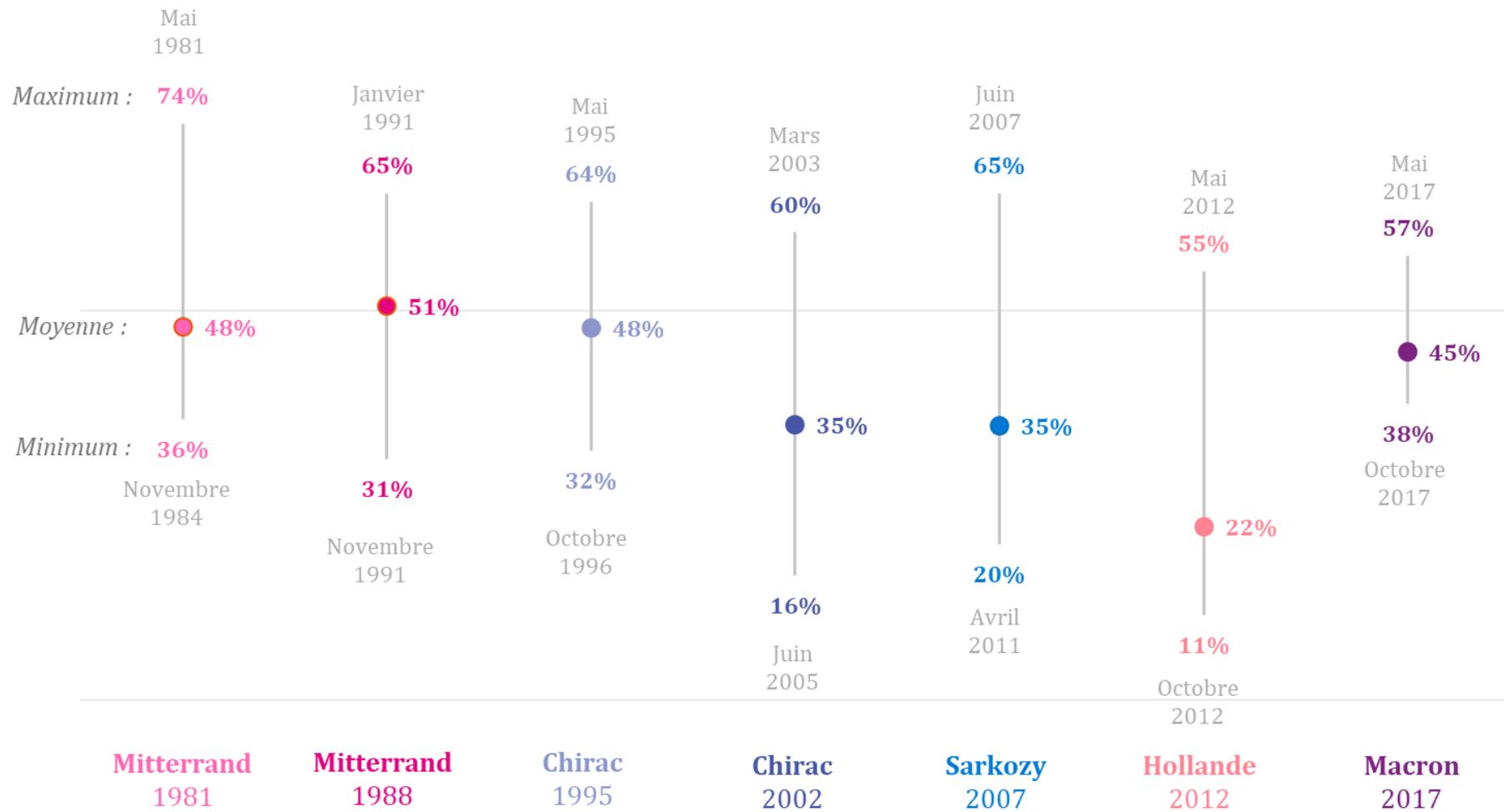
prendre du recul par rapport à la courbe initiale et chercher à faire apparaître les grandes lignes de cette information. Le grand public n'a pas besoin de faire la comparaison de la cote de confiance des différentes personnalités date à date. Ce qui est important c'est de pouvoir comparer la popularité des différents Présidents pendant leur mandat de manière globale. Ainsi, le parti-pris de cette création de visualisations de données est de répondre aux besoins du grand public avant tout et ensuite d'essayer de garder une certaine précision dans les données présentées.

3.2.1 L'étendue de la cote de confiance des Présidents de la République pendant leur mandat

Cette première visualisation de données a été réalisée sur la base d'indicateurs synthétiques : le minimum, le maximum et la moyenne. Il a semblé préférable d'illustrer la moyenne des séries car c'est une notion plus connue du grand public et plus abordable que la médiane par exemple. L'une des alternatives à cette visualisation était l'utilisation de boîtes à mouches pour représenter plus en détails la dispersion des séries, cependant ce type de graphique nécessite une explication plutôt approfondie. C'est ici le choix de la simplification qui a été fait. Les boîtes à moustaches pourraient être particulièrement adaptées à la cible des politologues car elles permettent de synthétiser de manière assez détaillée la cote de confiance des différents Présidents pendant leur mandat. Elles illustrent de manière approfondie la dispersion des séries statistiques.

Les différents Présidents de la République depuis 1981 sont placés sur un axe horizontal, en ordonnée est représentée la cote de confiance en pourcentage. Ainsi pour chaque personnalité, il est possible de distinguer la valeur de la cote minimale et maximale atteinte pendant un mandat mais aussi la moyenne de la cote de confiance sur ce mandat. Rappelons que le choix des couleurs suit celui des partis politiques des personnalités. Il semble nécessaire d'ajouter une grille d'interprétation pour le lecteur non habitué à ce type visualisation. Il s'agirait d'une petite phrase indiquant un exemple de lecture, placée sous la visualisation. La visualisation suivante a été conçue via le logiciel Excel. La comparaison de la popularité des Présidents est donc grandement simplifiée. L'ajout des dates auxquelles ces valeurs extrêmes ont été mesurées permet de se repérer et de pouvoir aller chercher des éléments externes pour mieux appréhender cette représentation.

L'étendue de la cote de confiance des Présidents de la République pendant leur mandat



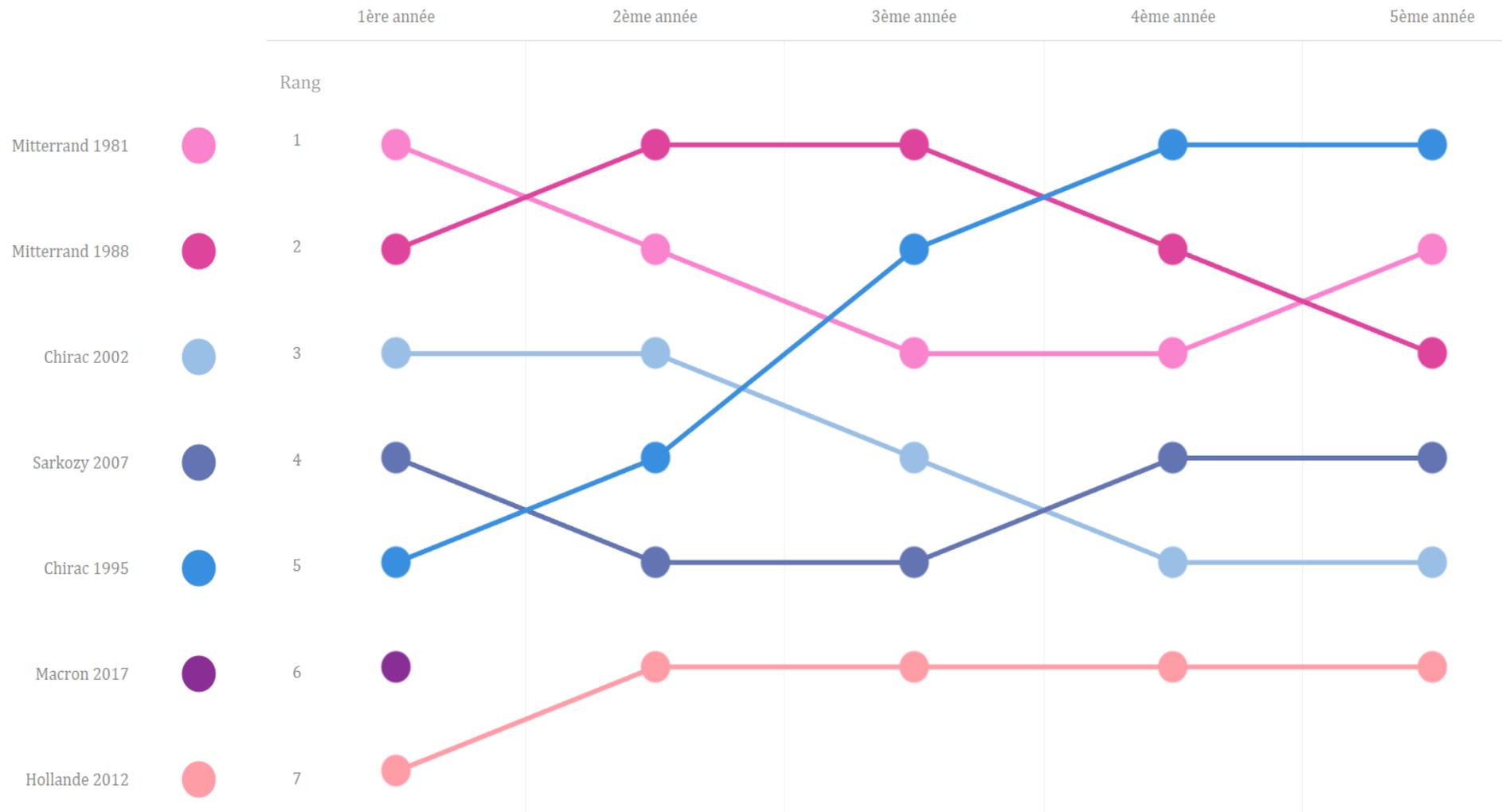
3.2.2 Classement des Présidents de la République selon leur cote de confiance par année de mandat

Cette visualisation de données est basée sur la première piste élaborée au début de la réflexion sur le sujet. L'un des problèmes de cette piste était le manque de clarté lié au surplus d'informations. C'est pourquoi il a semblé judicieux de calculer une cote de confiance moyenne par année de mandat, afin de diminuer la quantité d'information tout en représentant plusieurs états de variation. Par ailleurs, l'affichage des valeurs de la cote de confiance obstruait la lecture de la représentation, la valeur moyenne de la cote de confiance par année de mandat n'est donc pas apparente. Une fois que cette nouvelle variable a été calculée il a fallu l'ordonner c'est-à-dire calculer un rang allant de 1 à 7 ; 1 désignant la valeur la plus élevée et 7 la moins élevée. Pour créer cette visualisation de données, l'outil utilisé était Tableau Public.

Cette visualisation permet de comparer rapidement les positionnements des personnalités les unes par rapport aux autres en fonction des années passées au pouvoir. Pour les deux mandats de Mitterrand, ces valeurs ont été calculées uniquement sur les cinq premières années et non par sur le mandat complet afin de faciliter la comparaison. Sont ainsi présentés en abscisse les années de mandat et en ordonnée le rang des personnalités avec en haut le premier et en bas le septième. Le code couleur est le même que pour la visualisation précédente, il est indiqué sur la gauche dans l'ordre du classement de la première année de mandat. Il semblerait que la publication de cette visualisation de données nécessite, du moins pour la première fois, une note explicative de la même manière que la visualisation précédente. Pour un public n'étant pas habitué à ce genre de représentations, le bouleversement des codes peut être un peu brutal. Mais une note explicative, permet d'aider à la compréhension, à l'appropriation et donc à l'apprentissage ces nouveaux modes de représentations.

Dans cette visualisation, les cercles représentent les rangs des personnalités suivant les années de mandat. Les traits reliant ces rangs ne correspondent à aucune valeur, ils sont présents afin de faciliter la visualisation des changements de position d'une année à l'autre d'un mandat.

Classement des Présidents selon leur cote de confiance par année de mandat

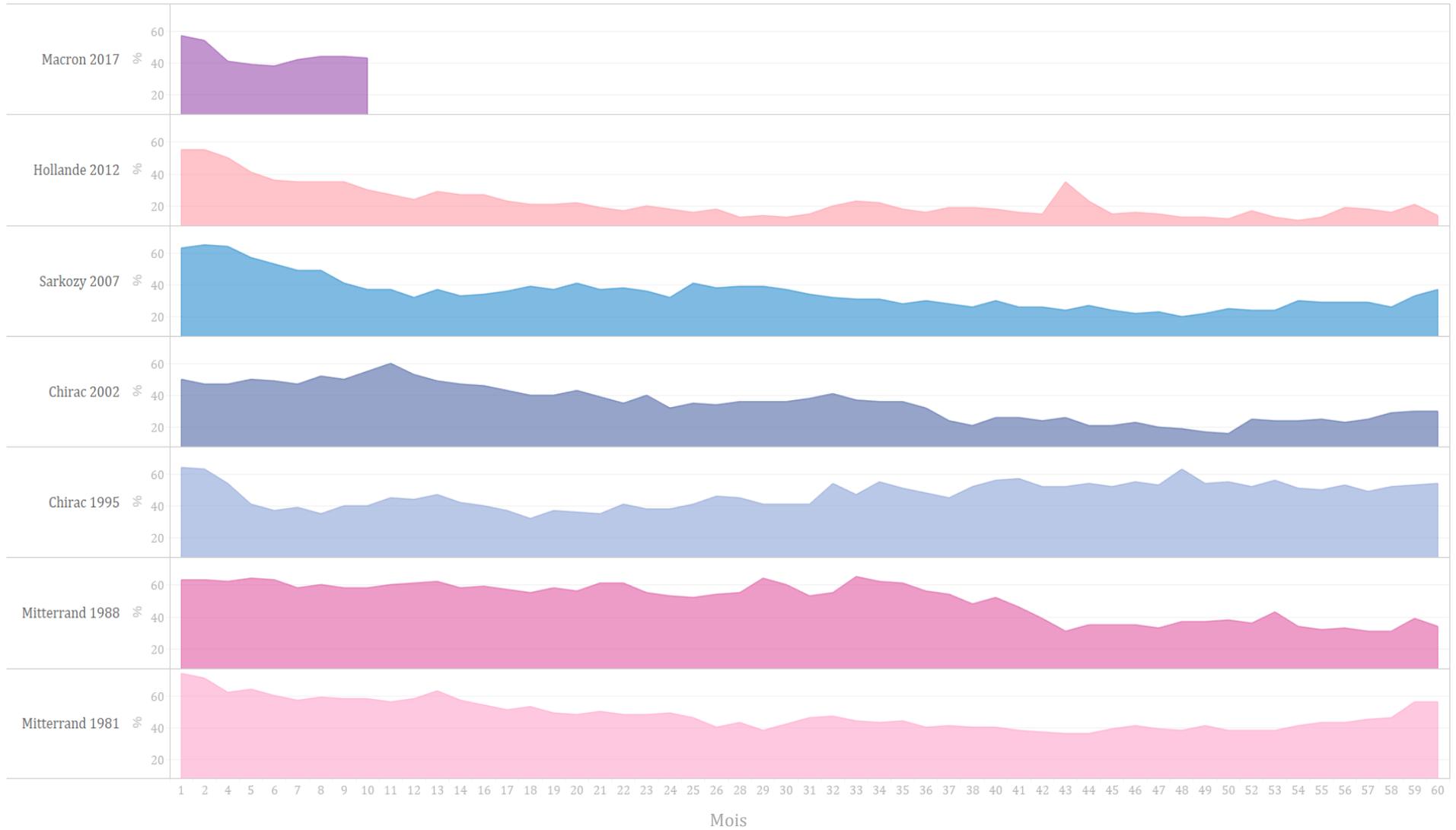


3.2.3 L'évolution de cote de confiance mensuelle des présidents de la République depuis 1981

Dans les deux représentations précédentes, l'information a été simplifiée si bien qu'il n'est pas possible de retrouver la valeur de la cote de confiance des différentes personnalités, date à date. Cet élément est particulièrement important pour le public des politologues mais peut tout autant être utile pour le grand public. C'est la raison pour laquelle il paraissait nécessaire de produire une visualisation de données de type courbe. Cependant le nombre trop important de séries représentées sur la courbe initiale rendait la lecture très confuse voire impossible. C'est en partant de ce point de réflexion qu'est venue l'idée de démultiplier l'axe des ordonnées (de la cote de confiance) par sept pour obtenir des courbes dissociées. De cette manière, l'axe des abscisses reste le même que celui de la courbe initiale : les mois passés au pouvoir. Et l'axe des ordonnées est toujours la cote de confiance en pourcentage. Cependant, la visualisation est divisée en sept lignes horizontales. Dans chacune de ces lignes est représentée par une aire l'évolution de la cote de confiance d'un Président de la République dans le temps. L'aire permettait de mieux percevoir les variations des séries plutôt qu'une simple courbe. Chaque ligne est divisée par trois sous-ligne qui sont des repères de valeur de la cote de confiance à 20%, 40% et 60%. L'ordre vertical des Présidents de la République a été fait de manière chronologique, avec le Président au pouvoir actuellement en haut et le Président le plus ancien en bas. Le code couleur est toujours organisé de la même manière que dans les visualisations de données précédentes. Enfin, certains mois ont été masqués pour cause de données manquantes.

Cette dernière visualisation de données semble être la plus facilement abordable par le grand public. Elle ne nécessite par de commentaire particulier sur la façon de l'interpréter.

Evolution de la cote de confiance par mois des présidents de la République pendant leur mandat



Après présentation de cette visualisation à différentes personnes, il apparaît que le but premier fixé étant la comparaison date à date des cotes de confiance des différents Présidents n'est pas atteint. Il est possible de voir de manière générale l'allure de l'évolution de la cote de confiance et par exemple celles des personnalités lors des premiers ou derniers mois de mandat. Cette visualisation permet aussi d'observer la densité des séries.

Dans cette situation, il serait particulièrement intéressant de créer cette même visualisation sous forme interactive sur internet. Le logiciel Tableau Public propose de publier en ligne des visualisations de données interactives. Ainsi, en passant le curseur sur un point d'une courbe, une barre verticale pourrait apparaître et afficher les valeurs des différentes séries pour ce mois donné. Dans ce cas, l'objectif de la comparaison date à date de ces sept séries serait atteignable.

La création de trois visualisations de données complémentaires permet d'approfondir la lecture des informations présentées la courbe initiale. Le surplus du nombre de séries représentées sur une seule courbe gênait la lecture. Il a donc fallu extraire l'ensemble des données représentées dans cette courbe sous forme d'axes d'analyse, dans le but de ne pas supprimer certaines informations présentes initialement de la réflexion. Le fil conducteur de ces visualisations de données a été la lisibilité et la simplification de l'information. Cependant, le changement de codes visuels et graphiques doit être accompagné de note de lecture, aidant le public à comprendre la visualisation. Bien que la simplicité et en quelques sortes la vulgarisation soit l'objectif principal d'une visualisation de donnée, celle-ci ne peut se permettre de bouleverser les habitudes de lecture sans être accompagné d'une notice ou d'une aide. C'est dans cette optique que l'idée d'apprentissage paraît incontournable dans le domaine de la visualisation de données.

CONCLUSION

Les représentations visuelles sont depuis toujours un outil de communication, de transmission et d'éducation. Elles étaient utilisées dans les civilisations anciennes, en tant qu'écriture, pour produire des traités commerciaux et permettre de retracer la valeur des biens vendus. Petit à petit, les retranscriptions visuelles de données chiffrées ont été utilisées comme outil, comme appui des scientifiques ou des grands penseurs pour mieux comprendre les phénomènes qui les entouraient. Ce fut par exemple le cas de Nicole Oresme, qui étudiait les variations de chaleur suivant les différentes parties du corps humain. Il utilisa alors un repère cartésien pour y placer ces données et ainsi tracer des courbes. De nouvelles révolutions en termes de représentations graphiques ont été réalisées par la suite et ont amené à la conception des codes graphiques actuels et universels.

Dans un environnement où la quantité de données produites augmente de manière exponentielle et où seul 1% des données mondiales produites est réellement traité, de nouveaux types de représentations visuelles se développent. L'intérêt de ces visualisations de données est de simplifier l'information qui peut parfois être très complexe. Ainsi, de nouveaux types de représentations ont été développées pour que le lecteur soit en capacité d'interpréter les données qu'il observe et de les mémoriser. Cependant, le bouleversement des codes graphiques engendré par l'observation d'une visualisation de données est à l'encontre de l'objectif principal de la visualisation. L'objectif d'une visualisation de données est bien de simplifier un message visuel. Mais lorsque les repères habituels d'un graphique ne sont plus présents, il faut apprendre à se repérer d'une manière différente. Ainsi, la visualisation des données nécessite un temps d'adaptation et parfois même d'acceptation. C'est dans ce contexte qu'entre en jeu l'apprentissage, notamment en littératie visuelle, concept défini de la sorte « *groupe de compétences acquises en vue de l'interprétation et de la production de messages visuels* » (Brill, Kim et Branch, 2001).

La nouvelle économie de l'attention porte aujourd'hui sur le temps d'attention qui constitue l'objet principal d'appropriation. Les communications s'orientent donc désormais vers la captation de l'attention du public. Et cette attention visuelle est plus que

jamais précieuse. Différents mécanismes relevant de la psychologie cognitive ont été identifiés comme efficaces pour provoquer le déclenchement de cette attention visuelle du lecteur. Les entreprises ont bien intégré ces mécanismes. Elles communiquent désormais via des visualisations de données parfois trompeuses, comme dans l'exemple de la communication d'Air France reprise par le journal *Le Point*. L'utilisation d'une visualisation de données permet par exemple de ne pas afficher les axes d'abscisse et d'ordonnée - comme l'aurait fait un graphique classique - et donc de jouer plus facilement sur des distorsions d'échelles pour faire passer un message. C'est aussi en ce sens, que l'apprentissage en littératie visuelle est utile. Il s'agit non seulement d'acquérir les capacités d'interprétation et de production de messages visuels mais aussi des compétences permettant de porter un regard critique sur une visualisation de données.

Dans le cas pratique présenté, ce sont les deux facettes de la littératie visuelle qui sont utilisées. La première est celle de la capacité d'interprétation d'un message visuel. La production de la courbe de l'évolution de la cote de confiance des Présidents de la République date à date depuis 1981 utilise des codes classiques, c'est-à-dire interprétable par un lecteur lambda. En revanche, la composition des trois data visualisations complémentaires nécessite une note de lecture ou une aide, du moins pour une première lecture. L'objectif de cette production est bien de faciliter la compréhension d'une courbe peu lisible et d'en simplifier la lecture mais le changement de codes graphiques implique le besoin d'un accompagnement pour le lecteur. Ce phénomène est par ailleurs amplifié par le fait que la courbe publiée dans le baromètre politique de Kantar Public est produite depuis 30 ans, le lecteur est donc habitué à déchiffrer cette courbe et à y repérer les évolutions mensuelles. Et il aurait été tout à fait intéressant de réaliser une étude qualitative auprès des lecteurs du magazine en amont de la production des visualisations de données afin d'aligner cette recherche sur les attentes exprimées des lecteurs. Le changement de représentations est doublement accompagné du phénomène d'adaptation et de temps d'appropriation des informations. Cependant, une fois l'apprentissage réalisé la lecture devrait être simplifiée. La seconde facette de la littératie visuelle utilisée est celle de la production d'un message visuel. Se confronter à la production d'une visualisation de données tout en ayant un objectif fixé en amont, permet effectivement de développer des compétences en littératie visuelle. Le manque de lisibilité de la courbe a amené à avoir un regard critique sur celle-ci, permettant d'identifier les forces et les faiblesses d'une représentation en ligne. Ces points forts et ces points faibles ont par la

suite été réétudiés pour aboutir à la création de visualisations de données présentant la même information de façon simplifiée.

REFERENCES

- Amato E.A. (2015). Enjeux et opportunité de la datavisualisation : interagir avec les données.
- Arruabarrena, B. (2015). L'écosystème numérique de la datavisualisation. *I2D-Information, données & documents*, 52(2), 56-58.
- Arruabarrena B. (2015). Datavisualisation : principes, enjeux et perspectives pour des utilisateurs non-experts.
- Bertin, J. (1970). La graphique. *Communications*, 15(1), 169-185.
- Bertin J. (1981). Théorie matricielle de la graphique. *Communication et langages*, vol. 48, n° 1, p. 62-74.
- Börner, K. et Polley, D. E. (2014). Visual insights: a practical guide to making sense of data. Cambridge (Mass.) : The MIT Press
- Cardon, D. (2012). Regarder les données. *Multitudes*, (2), 138-142.
- Casilli, A. (2015). Voir pour comprendre et comprendre pour voir. *I2D - Information, données & documents*, volume 52,(2), 40-41. <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-40.htm>
- Citton, Y. (2012). Traiter les données : entre économie de l'attention et mycélium de la signification. *Multitudes*, 49,(2), 143-149. doi:10.3917/mult.049.0143.
- Conversy, S. (2015, October). De la pré-attentivité des conjonctions visuelles. In *27ème conférence francophone sur l'Interaction Homme-Machine*. (p. alt2).
- Debes, J. L. (1969). The loom of visual literacy--An overview. *Audiovisual Instr.*
- Desfriches Doria, O. (2015). Quels dispositifs numériques pour appréhender la datavisualisation ?. *I2D - Information, données & documents*, volume 52,(2), 54-56. <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-54.htm>.
- Branch R., Brill J., & Kim D. (2001). Visual literacy defined: The results of a delphi study. Can IVLA (operationally) define visual literacy? Selected Readings of the International

Visual Literacy Association 2001–Exploring the Visual Future: Art Design, Science & Technology. Robert E. Griffin, Jung Lee & Vicki S. Williams (eds.), 9-15.

Drot-Delange, B. (2016, September). Visualisation interactive de données: les apports d'une étude de cas menée auprès d'étudiants de licence. In *Quand le Big Data transforme l'éducation la formation et les apprentissages*.

Fekete, J. D., & Boy, J. (2015). Recherche en visualisation d'information ou Dataviz: pourquoi et comment?. *I2D–Information, données & documents*, 52(2), 32-33.

Guerin, C., Chandesris, M. & Remy, A. (2016). Vers des algorithmes à dess(e)ins : intégration du design dans la conception d'une datavisualisation. *Sciences du Design*, 4,(2), 30-50. <https://www.cairn.info/revue-sciences-du-design-2016-2-page-30.htm>

Hachour, H. (2015). De la fouille à la visualisation de données : un processus interprétatif. *I2D – Information, données & documents*, volume 52,(2), 42-43. <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-42.htm>.

Hurter, C. (2010). *Caractérisation de visualisations et exploration interactive de grandes quantités de données multidimensionnelles* (Doctoral dissertation, Université Paul Sabatier-Toulouse III).

Lagnel J.M. (2017). *Manuel de data visualisation*.

Lebelle, B. (2012). *Convaincre avec des graphiques efficaces: sous Excel™, PowerPoint™, Tableau™...* Editions Eyrolles.

Lebrun M. (2015). La littératie visuelle : Génèse, défense et illustration. *Revue de Recherches en LMM (r2lmm.ca)*, vol. 2.

Lipovetsky, G. (2006). *Le bonheur paradoxal: essai sur la société d'hyperconsommation* (Vol. 377). Paris: Gallimard.

Lloveria, V. (2015). Data design-moi un mouton. *Communication & Organisation*, (46), 99-112.

McCandless, D. (2012). *Information is beautiful* (pp. 978-0007294664). London: Collins.

Naccache, L. (2017) *Le Chant du signe. Aventures et mésaventures de nos interprétations quotidiennes*, Odile Jacob

Norman D. (2002), « Emotion & Design : Attractive Things Work Better », *Interactions*, vol. 9, n° 4, juillet, p. 36-42.
DOI : [10.1145/543434.543435](https://doi.org/10.1145/543434.543435)

Shapley D. (1998). The da Vinci of Data. *The New York Times*

Sheps, A. (1999). Joseph Priestley's time charts: the use and teaching of history by rational dissent in late eighteenth-century England. *Lumen: Selected Proceedings from the Canadian Society for Eighteenth-Century Studies/Lumen: Travaux choisis de la Société canadienne d'étude du dix-huitième siècle*, 18, 135-154.

Tufte, E. R. (1985). The visual display of quantitative information. *Journal for Healthcare Quality*, 7(3), 15.

Treisman, A. (1986). Features and objects in visual processing. *Scientific American*, 255(5), 114B-125.

Vaisman C. (2015). La visualisation, un langage sans parole.

Vincenot, P. (2016). La dataviz: aide à la décision et à la compréhension. *I2D-Information, données & documents*, 53(4), 54-55.

Womack, Ryan (2014). **Data Visualization and Information Literacy**. *IASSIST Quarterly*, 38(1), 12-17. Retrieved from <https://doi.org/doi:10.7282/T3X92CZF>

Wright, R. D. (Ed.). (1998). *Visual attention* (No. 8). Oxford University Press.

« Visualisation de données, les leçons de l'histoire », Gaëtan Gaborit, mai 2012, slideshare.net, <http://fr.slideshare.net/gaetangab/visualisation-de-donnees-les-leons-de-l-histoire-13139160>

<https://washingtonmonthly.com/magazine/mayjune-2011/the-information-sage/#.WvS4TVIJZig.twitter>

<https://cursus.edu/articles/33818/visualisation-de-donnees-dessiner-pour-informer-et-comprendre#.W2iQXSgzY2w>

http://archives.lesechos.fr/archives/cercle/2014/03/06/cercle_92883.htm

<https://education.francetv.fr/matiere/education-aux-medias/premiere/programme/datagueule>

<https://moz.com/blog/data-visualization-principles-lessons-from-tufte>

<http://www.etopia.be/spip.php?article569>

https://www.francetvinfo.fr/replay-radio/nouveau-monde/moins-de-concentration-que-les-poissons-rouges-a-cause-des-ecrans_1778207.html

https://www.lemonde.fr/les-decodeurs/article/2015/10/26/pourquoi-il-faut-se-mefier-des-graphiques_4797002_4355770.html

https://www.lemonde.fr/les-decodeurs/article/2014/12/31/des-chiffres-et-des-images-l-annee-2014-en-12-datavisualisations_4539835_4355770.html