

Méthode de sélection et de configuration des représentations graphiques des corpus d'informations

Cyprien PIERRE 

2025-02-04

Abstract

Abstract

Rédiger l'abstract ici **Mots clés** : Visualisation de données, Graphique, Méthode

Introduction

La visualisation de donnée est une discipline ancienne à la littérature riche et précise. Cependant, il n'existe, à ce jour pas de synthèse opérationnelle ni de méthode harmonisée dans la conception des représentations graphiques des informations. Ce rapport vise à fournir un point de départ cohérent à de futurs travaux en représentation visuelle d'informations. Il est basé sur une revue de littérature rigoureuse et les recommandations qu'il porte sont issues des lectures réalisées. Une tentative est réalisée d'objectiver les usages des diverses représentations.

L'origine de ces travaux prend place dans un besoin grandissant de solution de visualisation d'information dans le secteur de la construction. Les technologies de Big Data, BIM, VDC et les disciplines de la science des données ont fortement pénétré ce secteur sur l'ensemble du cycle de vie des ouvrages.¹ L'utilisation renforcée de l'IoT, l'émèrgence des BIS et BOS et l'intégration des solutions d'intelligences artificielles renforcent d'autant plus la nécessité pour ce secteur de se doter de moyens robuste en exploration et en communication des informations.

Le secteur de la construction est historiquement peu digitalisé en

comparaison avec les autres industries. Pour adresser ce retard, il est important de publier des études opérationnelle claires et industrialisables dont chaque acteur de l'industrie de la construction puisse bénéficier immédiatement. Ce rapport s'inscrit dans cette démarche.

Il existe de multiples manières d'interagir avec un graphique.^{2,3}

Ce rapport aborde :

- L'exploration visuelle
- La lecture des tables
- L'écoute des descriptions
- La sonification des données
- L'interactivité avec les commandes de clavier

Ces éléments impliquent la mise en oeuvre de solution programmatique d'accessibilité pouvant intégrer une logique globale de préparation de graphiques numériques.

Ce rapport n'aborde pas :

- L'exploration en réalité mixte
- Les retours haptiques
- L'emploi d'écran à relief (braille)

Ces éléments nécessitant l'emploi de matériels spécifiques.

Démarche

La conduite de cette revue s'est déroulée de la manière suivante : Nous avons commencé par dresser un panorama des types de graphiques existant en nous appuyant sur les travaux de From Data to Viz,⁴ The Graphic Continuum,⁵ Lexique Visuel,⁶ Insights for ArcGis.⁷ Cette première action nous a permis de pré-catégoriser les graphiques selon leurs utilisations.

Ensuite, nous avons étudié les livres de divers auteurs et en avons tirés un ensemble de règles générales quant à la préparation des graphiques. Cela concerne les aspects visuels (couleurs, polices...), des conseils sur la pertinence dans le choix des types graphiques suivant les objectifs attendus, quelques éléments d'accessibilités et des retours d'expériences. Sur cette base, nous avons affiné la catégorisation préétablies et commencé à regrouper les recommandations.

Pour compléter les dispositions propres à l'accessibilité, nous avons étudié le rapport Centering accessibility in data visualization.⁸ Nous en avons tiré un aperçu global des approches possibles pour améliorer l'accessibilité des graphiques ainsi que de nombreux conseils et bonnes pratiques.

En parallèle, nous avons étudié diverses chartes de publications issues de divers médias et institutions. Ces lectures nous ont permis de nourrir les recommandations en matière de conception visuelle des graphiques ainsi que de dresser une liste d'éléments standardisables et d'autres pouvant être laissés libres de personnalisation avec quelques conseils.

Pour finir, nous avons étudié les articles scientifiques publiés sur ces sujets avec une profondeur de recherche à 5 ans. Ce choix de profondeur résulte d'une évolution conséquente depuis le début de la décennie des sciences de l'information. Certains papiers étudiés sont plus anciens mais ont été très régulièrement cités par d'autres publications et nous avons choisis de les considérer. De cette étape, nous avons

affiné les recommandations en identifiant les résultats d'expérimentations.

Sémantique

Éléments généraux

Polices

Le choix d'une police de texte a de multiples impacts sur la perception des graphiques. Il convient de sélectionner une police accessible et de préférence sans sérif pour un usage informatique (Arial, Calibri, Source Sans Pro, Verdana...).⁹ Les polices avec serif peuvent être considérées pour la production d'éléments imprimés. Il n'y a pas de consensus clair sur l'impact des sérifs sur l'accessibilité des polices.¹⁰ Il n'y a pas non plus de consensus clair sur l'efficacité des polices conçues pour adresser des problématiques d'accessibilités telles que la dyslexie. Il convient d'utiliser une police utilisant une hauteur fixe pour les chiffres.¹⁰

Il est conseillé de restreindre l'utilisation de l'italique car les textes affichés de la sorte sont plus difficiles à lire. Il est également conseillé de limiter l'utilisation de la graisse et du soulignement à des cas spécifiques pour ne pas surcharger les présentations.

La taille de la police joue un rôle majeur dans l'accessibilité du texte. Il est recommandé d'utiliser une hauteur de police de 12 points.⁹ Le nombre de tailles et de type de police doit être limité en nombre.⁹

Certaines polices peuvent être utilisées pour projeter des icônes (NerdFont, StateFace...).¹¹ L'intérêt des icônes est discuté plus tard dans ce rapport.

Il est important de prévoir le chargement de toutes polices non standard (eg. Source Sans Pro, NerdFont...) dans l'interface utilisateur si celles-ci sont utilisées puisqu'elles ne sont probablement pas installées dans le système d'exploitation de l'utilisateur. Prévoir leur chargement vise à assurer la bonne expérience des utilisateurs.

Tables

Une table ou tableur permet d'exposer des données brutes organisées en lignes ou en colonnes.

Sauf mention contraire, les recommandations sur les tables sont issues du livre *Show me the numbers* de Stephen Few.¹⁰ L'auteur y rentre très en détails sur chaque point de paramétrage. Il y indique notamment les orientations en matière de construction de tableur lorsqu'il s'agit du choix premier d'affichage de données. Ce rapport s'intéresse à la conception de graphique. Dans ce cadre, les tableurs sont des éléments complémentaires pouvant être affichés par l'utilisateur pour explorer plus précisément les données préalablement affichées.

Pour composer une table de donnée lisible, il est recommandé de :

- Séparer les entrées avec un espace vide,
 - Lorsque les données sont présentées en colonnes, l'espace entre deux colonnes doit être plus grand qu'entre deux lignes,
 - Lorsque les données sont présentées en ligne, l'espace entre deux ligne doit être plus grand qu'entre deux colonnes,
 - Insérer une ligne vide toutes les 5 lignes pour faciliter le balayage visuel¹²
- Utiliser une ligne horizontale pour séparer les entêtes de colonnes des données,
- Utiliser une ligne horizontale pour séparer les catégories lorsque les données sont triées par catégories,
 - Dans ce cas de figure, ne pas répéter le nom de la catégorie sur toutes les lignes,
 - Les noms des catégories doivent être dans la première colonne, les sous-catégorie dans la seconde colonne, etc. S'il y a plusieurs subdivisions, les noms des catégories doivent être apposées sur la même ligne,
- Rappeler le nom de la catégorie en cas de changement de page,
- Maintenir la structure du tableur sur toutes les catégories,
- Rappeler les titres des colones en cas de changement de page,
- Les catégories doivent être ordonnées suivant un ordre logique (eg. chronologique, alphabetique, par classement, etc.)
- Valeurs sommarisées
 - Utiliser une ligne verticale pour séparer les valeurs placées en colonne à droite de toutes les valeurs
 - Utiliser une ligne horizontale si ces valeurs sont placées en une ligne en bas du tableau,
 - Si ces valeurs sont le message important de votre tableur, il convient de les affichés immédiatement à droite des colonnes de catégories ou immédiatement en dessous des entêtes de colonnes, suivant la nature du sommaire.
- Ne pas effectuer de rotation sur un tableur, son orientation doit respecter celle du texte du document,
- Les produits d'un calcul doivent être affichés dans la colonne immédiatement à droite de la colonne source de données
- Uniformiser les alignnements
 - Les titres des colonnes suivent les alignnements des données
 - Aligner les chiffres à droite,
 - * Homogénéiser les décimales (généralement 2 ou 3 décimales suffisent suivant le contexte),
 - * Indiquer les valeurs négatives avec le symbole moins (-), ici la notion de valeur symbolique est importante. Certains choisissent d'identifier les valeurs négatives entre parenthèses, cependant

il ne s'agit pas d'une représentation naturelle répandue pour une telle identification.

- * Séparer les digit d'un nombre par un espace tous les 3 caractères. Dans le cas de grands nombres, il convient de les arrondir à la précision utile (dixaine, centaine...). Par défaut, la précision affichée doit correspondre à celle de la source d'information.
 - * Si une valeur numérique réfère à une information de catégorie elle doit être traitée comme un texte.
- Aligner les textes à gauche,
 - Centrer les dates et utiliser une convention stricte d'écriture de ces données telles que YYYY-MM-DD.¹³ La composition de la date doit être indiqué à l'utilisateur. Le choix du format doit respecter le niveau de précision associée à la mesure. Implicitement, le choix d'un formatage de plus haut niveau que la précision de la mesure induit une agrégation des valeurs.
 - Centrer les données dont la largeur de caractère est fixe

Si des valeurs spécifiques doivent être mises en avant, il est possible d'utiliser l'une ou l'autre de ces solutions :

- mettre le texte en gras,
- remplir la cellule d'une couleur.

Il est recommandé de limiter cette opération à un nombre réduit de valeur. Si cela n'est pas possible, il convient de sélectionner un autre mode de visualisation.

Des prescriptions spécifiques à la préparation de tableaux pour certains graphiques sont apportés le cas échéant dans la suite de ce rapport.

Graphiques

Marqueur

Affiche une valeur simple, plus pertinent lorsqu'associé à une tendance sur une période donnée.

Jauge

Représentation métaphorique la vitesse d'un système.

Tiret

Présentation d'une valeur mesurée ou calculée au regard d'une valeur cible et Pertinent lors de la réalisation d'un benchmark.

Barres divergentes

Barres divergentes empilées

Utilisé pour illustrer les résultats d'enquêtes impliquant un sentiment. Les barres doivent être calculées en pourcentage et alignées à 100% comme pour une partie d'un ensemble. Les réponses sont affichées de façon ordonnée et du plus négatif au plus positif. Il convient d'utiliser une palette de couleur divergente. Lors de la préparation d'une enquête impliquant des sentiment, il est conseillé de proposer un choix de valeur paires pour éviter les votes neutres sauf si la représentation de l'indifférence est une volonté de l'étude. De plus, **expliquer la limite de nombre de zones facile à visualiser** (max 6?)

Spine

Utilisé pour afficher la quantité de réponses négatives et positives à une question. Le spine divise une valeur quantitative unique en deux composants distinct contrasté par un déterminant binaire. Utilisé pour afficher la quantité de réponses négatives et positives à une question.

Horizon

Permet de vérifier un équilibre entre deux borne par rapport à une base. Utilisé pour suivre l'évolution d'une variable dans le temps et devant respecter un écart-type (eg. tension à 230V +/- 5%). L'information la plus importante est le dépassement de l'écart-type. Pour correctement mettre en évidence ces phénomènes, il est recommandé d'utiliser une palette de couleur de

Cascade

Affiche les gains et pertes entres deux états d'un système.¹⁴

Bougies

Illustre les évolutions d'un stock sur une temporalité donnée. La ligne indique l'état le plus haut et l'état le plus bas de la journée. La barre indique l'état de l'ouverture et de la cloture de la période.¹⁵ Il convient d'utiliser une palette de couleur binaire pour contraster les périodes aux soldes positifs des périodes aux soldes négatifs.

Nuage de point

Représente la relation entres deux variables continues. La couleur est utilisée pour distinguer les catégories. Si des milliers de point de données sont à afficher, il est préférable d'utiliser soit :

- des cercles ouverts
- des petits cercles
- des cercles avec une transparence
Cela permet de limiter le masquage des points de données par surimposition.⁹

Nuage de bulle

Il s'agit d'un nuage de point dont la taille des bules est utilisé pour représenter la valeur quantitative.

Matrice de température

Représentation graphique d'une table de valeur homogène.¹⁶ Intéressante pour afficher une tendance lorsqu'un très grand nombre d'information est impliqué.

En utilisant une palette de couleur séquentielle, un lecteur aura tendance à considérer une couleur sombre comme une valeur haute. En utilisant une palette de couleur divergente (eg. basée sur la température de la couleur), ce phénomène s'estompe.

L'organisation des catégories changeant la forme de la matrice de température, il convient de retenir une organisation par valeur croissante ou décroissante à une colonne donnée. Il convient de signaler cette configuration au lecteur.¹⁷

Matrice de bulle

Ce graphique est une version agrégée du nuage de bulle et le la matrice de température.

Il simplifie la visualisation des relations entre un couple de variable. A réserver pour un usage d'exploration de données.¹⁶ La taille des bules représente la force de la corrélation. Cette force est la représentation du coefficient de Pearson.¹⁸

La couleur peut être utilisée pour distinguer les catégories (à vérifier). Elle peut être également utilisée comme double encodage de la corrélation (à vérifier).

Le tableur associée est une matrice de corrélation organisant les valeurs de coefficient de Pearson.

Pareto

Ce type de graphique montre la relation entre les valeurs quantitatives de catégories organisées en colonne ordonnée et la valeur cumulée affichée en ligne brisée. Le premier point de la ligne vérifiant ($\tau \geq 80\%Total$) est affichée. Est utilisé pour illustrer les éléments important dans un système qualité.

Arbre de mot

Représente les liens sémantiques. Utilisé pour classer les réponses textuelles d'une enquête.(voir Paige Jarreau) Utilisé pour illustrer l'usage des mots dans un texte.(voir Martin Wattenberg & Fernanda Viegas 2007) La taille des mots représente la fréquence de leurs usages.¹¹

Histogramme

Utilisé pour afficher une seule distribution de donnée. Il n'y a pas d'espace entre deux barres ce qui permet de favoriser la forme de la distribution. La nature du graphique (barre) implique un calcul d'agrégat pour chaque barre. La largeur de l'agrégat est à évaluer au regard du niveau de détail souhaité. L'histogramme n'est pas adapté à l'affichage de l'ensemble des valeurs, pour cela il convient d'utiliser une courbe de densité.

Polygone de fréquence

Utilisé pour afficher simultanément plusieurs distributions de données. Fonctionne comme un histogramme. Chaque agrégat est représenté par un point et ils sont reliés entre eux par un segment droit. Pour faciliter la lecture, il convient de limiter le nombre de distributions affichées à 4 ce qui correspondant notamment au nombre de symboles aisément discriminables.

Courbe de densité

La densité représente la probabilité de distribution des variables.¹⁶ Pertinent lorsque le nombre de données est très grand, autrement il convient d'utiliser un histogramme.¹⁹ En cas d'affichage de plusieurs courbe, utiliser des remplissages transparents pour faciliter la distinctions. L'air de la courbe est usuellement égal à 1 et de ce fait, l'échelle de l'axe des ordonnées dépend de l'axe des abscisses.¹⁹

Courbe cumulative

Utilisé pour représenter une distribution inégale. L'axe des ordonnées est toujours la fréquence cumulative et l'axe des abscisses est toujours une mesure.

Bande de valeur (strip plot)

Affiche les points de données d'une catégorie. Les points sont légèrement décalés les uns des autres par un offset pour éviter le masquage de données par surimposition. Ce type de graphique peut être associé à un graphique en violon ou en lignes de crêtes lorsque l'on veut exposer graphiquement les données sources des tracés.¹⁹

Boite à moustache

Valeur lettrée

Représente chaque découpage statistique ($x\sigma$) par une boite distincte. Convient en remplacement des boites à moustache lorsqu'il y a un très grand nombre de données disponibles. Permet de faire des estimations fiables.^{hofmannLettervaluePlotsBoxplots2017}

Violon

Représente fidèlement les distributions d'un très grand nombre de données. Utilisé pour faire des comparaisons entre des catégories. Très pertinent pour représenter les distributions bimodales.¹⁹ Le tracé du demi-violon est identique au tracé d'une courbe de densité. Celui-ci commence à la valeur la plus petite et termine à la valeur la plus grande.

Mini-graphe (spark line)

Affiche la tendance d'évolution d'une variable ou d'une catégorie. Est généralement intégré dans un tableur. N'affiche aucun autre élément que la courbe et un point symbolisant l'état actuel. Cette représentation est très macroscopique et

visé à susciter un sentiment plutôt que d'exposer une valeur précise.¹⁵

Bulle proportionnelle ordonnée

Utilisé pour illustrer les grandes variations de quantité pour différentes catégories et lorsque la distinction des nuances n'est pas primordiale. Les bulles doivent être dimensionnées par leur surface pour éviter d'exagérer les différences.¹⁴ Ce type de graphique utilise un double encodage de la quantité : la position et la taille.²¹ Bien que ce type de représentation soit considéré comme engageant, il convient d'utiliser un graphique en barre ordonnée à la place lorsque l'évaluation des différences entre catégories est importante.¹⁴

Barre ordonnée

Lollipop

Utilisé pour afficher la position dans une liste ordonnée lorsque l'affichage de la valeur est moins important que la position. Également très efficace comparé aux barres ordonnées lorsqu'il y a un grand nombre de catégories dans le classement. Si une ligne connecte le point de la valeur à l'axe des ordonnées, l'axe des abscisses doit commencer à 0, autrement cela n'est pas obligatoire.

Haltères

Utilisé pour comparer un minimum et un maximum de plusieurs catégories. La relation entre le maximum et le minimum d'une catégorie est représentée par la barre reliant les deux valeurs.

Peut être utilisé pour illustrer le changement d'une valeur entre deux états donnés. Dans ce cas, il convient de flécher le sens du changement de l'état antérieur vers l'état final.

Dans ces deux cas de figures, les catégories doivent être ordonnées de la

plus grande valeur à la plus petite.

Les haltères sont toujours affichés à l'horizontal.

Pente

Affiche la variation des valeurs ou ratios des catégories entre deux états donnés.

Bosses

Affiche l'évolution d'un classement entre différents jalons. Ce graphique n'affiche pas les valeurs sous-jacentes mais uniquement le positionnement de la catégorie dans le classement.²² Si la valeur source est également à représenter, ce graphique peut se transformer en graphique en ruban, combinant la disposition du graphique en bosse avec celles du graphique de Sankey.²²

Méthode de sélection

Études futures

Des études complémentaires pourraient être menées sur les sujets suivants :

- La conception de bibliothèques de graphiques accessibles fonctionnant aussi bien dans un usage web que dans les PDF,

Évolution

La table de données sous-jacente doit être organisée par colonne de valeur temporelle croissante.

Classement

La table de données sous-jacente doit être organisée par ligne de classement croissante (du premier au dernier en partant de la première ligne).

References

1. ASIA UNIVERSITY, TAICHUNG, TAIWAN; CHEN, Chiung-Hui. Research on the Application of Function-Technology-Aesthetics Framework in the Design Knowledge Modelling of Data Visualization. *Journal of Advances in Information Technology* [online]. 2020, pp. 10–14 [visited on 2025-01-22]. ISSN 17982340. Available from DOI: [10.12720/jait.11.1.10-14](https://doi.org/10.12720/jait.11.1.10-14).
2. 1. Centering Accessibility in Data Visualization. In: *Centering Accessibility in Data Visualization*. Urban Institute, 2022, pp. 4–8. Do Not Harm Guide, no. 3.
3. FRANK ELAVSKY. 2. The Right Tools for the Job: Learning and Building for Data Visualization and Accessibility. In: *Centering Accessibility in Data Visualization*. Urban Institute, 2022, pp. 10–18. Do Not Harm Guide, no. 3.
4. YAN HOLTZ; CONOR HEALY. *From Data to Viz*. 2018. Available also from: data-to-viz.com.
5. JONATHAN SCHWABISH; SEVERINO RIBECCA. *The Graphic Continuum* [online]. 2014. [visited on 2024-07-18]. Available from: <https://www.informationisbeautifulawards.com/showcase/611-the-graphic-continuum>.
6. *Lexique Visuel*. [N.d.]. Available also from: <https://github.com/Financial-Times/chart-doctor/tree/main>.
7. *Insights for ArcGIS* [online]. 2017. [visited on 2024-07-18]. Available from: <https://www.esri.com/content/dam/esrisites/fr-fr/media/pdf/data-visualization-options-in-insights-3.0.pdf>.
8. *Centering Accessibility in Data Visualization*. Urban Institute, 2022. Do Not Harm Guide, no. 3.
9. KRAUSE, Andreas; RENNIE, Nicola; TARRAN, Brian. Best Practices for Data Visualisation [online]. 2024, pp. 1–72 [visited on 2024-12-30]. Available from DOI: [10.5281/ZENODO.10600718](https://doi.org/10.5281/ZENODO.10600718).
10. 8. Table Design. In: *Show Me the Numbers - Designing Tables and Graphs to Enlighten*. Seconde Edition. Burlingame, California: Analytics Press, 2012, pp. 155–190. ISBN 978-0-9706019-7-1.
11. 10. Qualitative. In: *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks* [online]. New York: Columbia University Press, 2021, pp. 311–326 [visited on 2025-01-09]. ISBN 978-0-231-55015-4. Available from DOI: [10.7312/schw19310](https://doi.org/10.7312/schw19310).
12. *Ergonomie de l'interaction Homme-Système - Partie 125 : Recommandations Relatives à La Présentation Visuelle d'informations*. AFNOR, 2017. No. NF EN ISO 9241-125.
13. *Date et Heure - Représentations Pour l'échange d'information – Partie 1: Règles de Base*. 2019. No. ISO 8601-1.
14. 4. Comparing Categories. In: *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks* [online]. New York: Columbia University Press, 2021, pp. 68–132 [visited on 2025-01-09]. ISBN 978-0-231-55015-4. Available from DOI: [10.7312/schw19310](https://doi.org/10.7312/schw19310).

15. 6. Distribution. In: *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks* [online]. New York: Columbia University Press, 2021, pp. 179–216 [visited on 2025-01-09]. ISBN 978-0-231-55015-4. Available from doi: [10.7312/schw19310](https://doi.org/10.7312/schw19310).
16. III. The Graphics. In: *Data Visualization Made Simple: Insights into Becoming Visual*. New York: Routledge, Taylor & Francis Group, 2019, pp. 43–70. ISBN 978-1-138-50387-8 978-1-138-50391-5.
17. 6. Visualizing Amounts. In: *Fundamentals of Data Visualization* [online]. 1st. O'Reilly Media, 2019, pp. 45–58 [visited on 2024-12-30]. ISBN 978-1-4920-3108-6. Available from: <https://clauswilke.com/dataviz/>.
18. 8. Relationship. In: *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks* [online]. New York: Columbia University Press, 2021, pp. 249–288 [visited on 2025-01-09]. ISBN 978-0-231-55015-4. Available from doi: [10.7312/schw19310](https://doi.org/10.7312/schw19310).
19. 9. Visualizing Many Distributions at Once. In: *Fundamentals of Data Visualization* [online]. 1st. O'Reilly Media, 2019, pp. 81–92 [visited on 2024-12-30]. ISBN 978-1-4920-3108-6. Available from: <https://clauswilke.com/dataviz/>.
20. HOFMANN, Heike; WICKHAM, Hadley; KAFADAR, Karen. Letter-Value Plots: Box-plots for Large Data. *Journal of Computational and Graphical Statistics* [online]. 2017, vol. 26, no. 3, pp. 469–477 [visited on 2025-02-04]. ISSN 1061-8600, ISSN 1537-2715. Available from doi: [10.1080/10618600.2017.1305277](https://doi.org/10.1080/10618600.2017.1305277).
21. 12. Visualizing Associations among Two or More Quantitative Variables. In: *Fundamentals of Data Visualization* [online]. 1st. O'Reilly Media, 2019, pp. 117–130 [visited on 2024-12-30]. ISBN 978-1-4920-3108-6. Available from: <https://clauswilke.com/dataviz/>.
22. 5. Time. In: *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks* [online]. New York: Columbia University Press, 2021, pp. 133–178 [visited on 2025-01-09]. ISBN 978-0-231-55015-4. Available from doi: [10.7312/schw19310](https://doi.org/10.7312/schw19310).