

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

Cyprien PIERRE 

2025-02-10

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 les éléments impliquant la mise en oeuvre de solution programmatique d'accessibilité et pouvant intégrer une logique globale de préparation de graphiques numériques tels que:

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

Ce rapport n'aborde pas les éléments nécessitant l'emploi de matériels spécifiques tels que :

- L'exploration en réalité mixte

- Les retours haptiques
- L'emploi d'écran à relief (braille)

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 utilisation.

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

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 leurs chargement vise à assurer la bonne expérience des utilisateurs.

Couleurs

Tables

Une table ou tableur permet d'exposer des données brutes organisées en lignes ou en colonnes.¹²

Sauf mention contraire, les recommandations sur les tables sont issues du livre "Show me the numbers" de Stephen Few.¹³ 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.

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

Composition

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 lignes doit être plus grand qu'entre deux colonnes,

Insérer une ligne vide toutes les 5 lignes pour faciliter le balayage visuel.¹⁴

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 et, 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.

Il convient de rappeler le nom de la catégorie en cas de changement de page et de maintenir la structure du tableur sur toutes les catégories et de rappeler les titres des colonnes en cas de changement de page. Les catégories doivent être ordonnées suivant un ordre logique (eg. chronologique, alphabétique, par classement, etc.)

Ne pas effectuer de rotation sur un tableur, son orientation doit respecter celle du texte du document.

Maintenir les alignements les titres avec celui des données en colonnes.

Chiffres

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.

Texte

Aligner les textes à gauche. Centrer les dates et utiliser une convention stricte d'écriture de ces données telles que "YYYY-MM-DD".¹⁵ 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.

Sommaire et agrégats

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 tableau, il convient de les afficher immédiatement à droite des colonnes de catégories ou immédiatement en dessous des entêtes de colonnes, suivant la nature du sommaire.

Les produits d'un calcul doivent être affichés dans la colonne immédiatement à droite de la colonne source de données.

Mise en évidence

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.

Méthode de sélection

Graphiques

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

Arc

Hémicycle, souvent utilisé pour visualiser la composition d'un effectif.

Arc temporel

Utilisé pour illustrer les connexions des catégories à travers le temps. L'ordre des données influence la perception de l'importance des connexions.¹⁶

Peut être utilisé pour illustrer les distances entre deux points. Cet usage met bien en évidence l'augmentation des distances.¹⁶

Bande de valeur

Affiche les points de données d'une catégorie.¹⁷ Les valeurs peuvent être affichés par des lignes ou des cercles pleins mais les cercles sont à privilégier lorsqu'il est nécessaire de déagglomérer les données.

Il est possible d'utiliser de la transparence pour illustrer les zones de valeurs compactes.¹⁷ Utiliser une palette de couleur divergente met en évidence la répartition des valeurs mesurées et souligner la séparation de celles-ci par rapport à un axe de référence.¹⁷

Les points peuvent être légèrement décalés les uns des autres par un offset pour éviter le masquage de données par surimposition. Ce mode de représentation est appelé "essaim d'abeille" (beeswarm).¹⁷

Histopoint : Wilkinson dot plot (Leland Wilkinson) Affichage des bandes de valeurs pour chaque agrégat d'un histogramme.¹⁷

Epis de blé : Wheat plot (SFew)

Barre divergente

Affiche correctement un ensemble de valeur quantitative triées par catégories lorsque des valeurs négatives sont utilisées.⁶ L'axe des abscisses doit être égal à 0 et toutes les barres doivent être projetées depuis cet axe.

Barres divergentes empilées

Utilisé pour illustrer les résultats d'enquêtes impliquant un sentiment.⁶ 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?)

Barre empilées

Représente le total de chaque catégorie et la proportion de leurs composantes. Les composantes doivent être homogènes. Si les proportions sont plus importantes que les totaux, considérer d'utiliser l'affichage empilée à 100%.¹²

Seule la catégorie ayant sa base à 0 sera facile à comparer pour le lecteur car les autres ne partagent pas la même base.¹⁸ Il est préférable de tracer chaque diagramme de barre dans un graphique distinct et des les disposer en [Matrice de graphique](#).¹⁸

Barre jumelée

Permet d'afficher plusieurs séries. Devient plus difficile à lire avec plus de deux séries.⁶ Les barres d'une même catégories sont collées les unes aux autres pour souligner leurs appartenance à une même catégorie (ratio 1:0). Espacer les jeux de barres jumelées d'une largeur égale à une barre (ratio n:1).

Les barres sont colorées avec une palette de couleur de catégories. Les couleurs sont utilisées pour distinguer chaque barres dans une catégories, les couleurs doivent être homogènes pour chaque catégorie. Si les différentes valeurs affichées sont d'importances similaires, il convient d'utiliser

une palette de couleurs à faibles contrastes. N'utiliser qu'une seule couleur par ensemble de données apparentées.¹⁹

Barre ordonnée

Les catégories sont ordonnées par décroissance de leurs valeurs de quantité¹⁸ sauf si les catégories ont un autre ordre logique (eg. temporel). Dans ce second cas, cet ordre doit être utilisé pour la construction du graphique.²⁰

Note: si catégories temporelles, utiliser un histogramme?

Lorsqu'un ordre de classement est à afficher, l'utilisation d'un graphique horizontal est à privilégier.⁹

William Playfair (1786) pionnier de l'utilisation de ce type de graph ?

Les barres doivent être affichées à l'horizontal s'il y a beaucoup de valeurs à afficher ou si les labels des catégories s'entrecroisent.^{6,21,20,19,18}

Il convient de maintenir l'affichage vertical du graphique si les catégories affichées sont des unités temporelles.¹⁹

Ce graphique est plus lisible pour un nombre limité d'entrées.¹²

L'espace entre deux barres est égal à la moitié de la largeur d'une barre (ratio 1:0.5). Les barres ont toutes la même largeur.¹⁹

Utiliser une couleur homogène à l'ensemble des barres. Une couleur de contraste peut être utilisée pour mettre en évidence certaines catégories.

Ne pas utiliser de motifs de remplissage. Les barres doivent être remplies par un solide coloré.¹⁹

Ne pas utiliser de bordures de barres. Ces dernières doivent naturellement contraster avec l'arrière plan.

Si de grands écarts de valeurs empêchent la bonne visualisation des écarts sur une partie du graphique, il est recommandé d'ajouter un second graphique en barre ordonnée n'affichant pas les catégories aux grandes valeurs.¹⁸

Ne pas afficher de séparateurs au niveau

des catégories, les barres remplissent elles même ce rôle, associées aux espaces entres elles.¹⁸

Pour les petits graphiques statiques, il convient d'apposer directement un label à chaque barre contenant la valeur de celles-ci. Par cette action, il convient également de masquer l'axe des absyesses.¹⁸

Pour les graphiques dynamiques, interactifs ou de grande largeur, il convient d'afficher des lignes d'aide à la lecture en projection des repères de l'axe des valeurs. Ces lignes doivent être faiblement contrastées et être tracées en arrière plan des barres.¹⁸

Ne pas segmenter les barres.¹⁸

Ne pas ajouter d'iconographie aux barres.²²

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 leurs surface pour éviter d'exagérer les différences.¹⁸ Ce type de graphique utilise un double encodage de la quantité : la position et la taille.²³ 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.¹⁸

Boîte à moustache

Résume les distributions multiples en montrant la médiane et l'étendue des données.⁶

Ce graphique est plus efficace lorsqu'il y a de nombreuses mesures pour une entrée et que leur distribution doit être représentée (eg : mesure médiane, basse, haute, incertitude...).¹²

Les lignes s'étendant au dessus et en dessous sont appelées les moustaches.²⁴

Voir trx de son inventeur : John W. Tukey

Il est possible d'afficher la position d'un point précis du jeu de donnée sous-jacent. (couleurs mise en évidence ?)

Les boîtes à moustaches suivent les dispositions graphiques des barres ordonnées.¹⁹

Bosses

Affiche l'évolution d'un classement entre différents jalons.^{12,6} Ce graphique n'affiche pas les valeurs sous-jacente mais uniquement le positionnement de la catégorie dans le classement.²⁵ Si la valeur source est également à représenter, ce graphique peut se transformer en graphique en ruban, combinant les disposition du graphique en bosse avec celles du graphique de Sankey.²⁵

Camembert

Diagramme standard, difficile à lire.⁶ Limiter à 5 (4?) entrées maximum¹²

Ce graphique doit être limité à l'illustration d'un état donnée. Il est important de garder à l'esprit que ce graphique ne donne qu'une représentation et n'est pas adapté à l'affichage de valeurs précises.²⁶

Le camembert permet d'accentuer l'appartenance d'une partie individuelle à un ensemble homogène et complet. Il illustre parfaitement des fractions simples.²⁷ Les quantités n'ont pas leurs place dans un camembert, ce sont les fractions (taux, pourcentages) qui y sont illustrés. Ainsi, l'ensemble des valeurs constituante sont strictement égale à 100% du total.²⁰ Il représente plus efficacement les parts dont les angles sont familiers aux lecteurs soit respectivement 90°, 180° et 240° pour les taux 25%, 50% et 75%.

Ce graphine n'est pas adapté à ces cas de figure :

- les parts sont de tailles similaires,²¹

- Il y a beaucoup de catégories à afficher pour atteindre l'objectif du message,
- il y a une comparaison entre deux graphiques à réaliser,²⁶
- il y a un changement entre plusieurs états.²⁷

Il convient de :

- Arranger les parts par ordre de valeur croissant en commençant à '12h'.²⁶
- Utiliser un espace vide entre chaque part pour les faire ressortir,²¹
- Utiliser une palette de couleur de mise en évidence pour faire ressortir une seule part ou,
- Utiliser une palette de couleur pour faire ressortir deux parts,
- Apposer les labels directement sur les parts au moyen d'étiquette.²¹

Les parties n'étant pas mises en avant peuvent être agrégées en une catégories "autres". Leurs informations ne servant pas à appuyer le message du graphique.²⁶

Dans le cas où une seule partie de l'ensemble est mise en avant, il est à considérer l'utilisation d'un **Marqueur**. Le camembert n'apportant aucune information particulière à votre propos.

Il n'est pas encore clairement établi si la comparaison des valeurs est réalisée par la perception des angles ou par la perception des longueurs d'arcs des différentes parties.²⁶

Ce diagramme possède un avantage majeur au regard des autres modes de représentation des parties d'un ensemble, il est très populaire et ainsi accessible à usager.²⁶ Il est également reconnu que les formes circulaires attirent l'intérêt des utilisateurs.²⁶

Carte de fréquentation

Excellente visualisation pour suivre la fréquence d'une activité. Favorise la quantité à la qualité (exemple : github).⁶

Cascade

Affiche les gains et pertes entre deux états d'un système.¹⁸ Utilisé principalement pour les analyses financières.⁶

Chandelier

Affiche des bilans quotidiens (eg. ouverture, fermeture, point haut, point bas).⁶ La couleur est utilisée pour représenter la direction de l'évolution de la valeur (augmente, descend).

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 clôture de la période.¹⁷ Il convient d'utiliser une palette de couleur binaire pour contraster les périodes aux soldes positifs des périodes aux soldes négatifs.

Chronologie circulaire

Pour l'affichage des valeurs discrètes de taille variable à travers de multiples catégories.⁶

Chronologie verticale

Affiche le temps sur l'axe Y. Particulièrement adapté à l'affichage sur mobile.⁶

Colonne vertébrale

Utilisé pour afficher la quantité de réponses négatives et positives à une question. Le spine divise une valeur quantitative unique en deux composants distincts contrastés par un déterminant binaire.⁶ Utilisé pour afficher la quantité de réponses négatives et positives à une question.

Coordonnées parallèles

Montre en simultané la valeur et la relation de catégories hétérogènes pour plusieurs systèmes.¹⁶ Permet de valoriser la valeur.⁶

Pratique pour observer les schémas et relations entre les systèmes.¹² Ce graphique est plus efficace pour explorer des jeux de données que pour les présenter à un audimat.²¹

Chaque axes de catégories peut avoir sa propre échelle. Cela représente mieux les valeurs de chaque catégorie mais cela implique également de bien labeliser les axes pour distinguer leurs échelles.¹⁶

Afficher un grand nombre de jeux de données sur un même graphe risque de le rendre illisible.¹⁶ Dans ce cas il convient d'utiliser une palette de couleur de mise en évidence et de sélectionner un ou deux jeux de données à mettre en avant.

Lors de l'exploration d'un jeu de donnée à travers une interface informatique, il convient de mettre en avant la ligne pointée par l'utilisateur et de labeliser chaque points de catégorie avec la valeur associé.²¹

Corde

Illustre les flux bidirectionnels et le gagnant net dans une matrice.⁶

Dans la mesure du possible, il convient de flécher le sens du flux.

Ce diagramme incite les utilisateurs à l'explorer plus en détail.¹⁶

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.⁶

Courbe de densité

La densité représente la probabilité de distribution des variables.²¹ Pertinent lorsque le nombre de données est très grand, autrement il convient d'utiliser

un histogramme.²⁴ 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.²⁴

Donut

Dans un graphique en donut, les valeurs sont à comparer en fonction des longueurs d'arcs.

Comme le diagramme circulaire, le centre peut être utilisé pour afficher un total ou encore une icône.⁶

Il est, autrement, en tout point similaire aux [Camembert](#)

Etats parallèles

Montre les changements sur un flux entre un état et au moins un autre. Adapté au traçage d'un processus complexe.⁶ Cette représentation découpe un ensemble en quantité individuelles de plusieurs catégories. Les bandes sont tracées entre chaque catégories pour représenter leurs relations. La largeur des bandes indiquent la proportion des catégories corréllées les unes aux autres.²⁸ A utiliser à la place des mosaïques et treemaps lorsqu'il y a plus de deux variables de regroupement des données.²⁷

Eventail

Utilisé pour montrer l'incertitude des projections futures.⁶

Frise de Priestley

Parfait lorsque la date et la durée sont des éléments clés dans l'histoire de la donnée.⁶

Haltères

Utilisé pour comparer un minimum et un maximum de plusieurs catégories.⁶ La relation entre le maximum et le minimum

d'une catégorie est représenté par la barre reliant les deux valeurs.¹²

Peut être utilisé pour illustrer le changement d'une valeur entre deux état données. 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.

Histogramme

Utilisé pour afficher la fréquence d'une variable simple regroupée en intervalle de fréquence (agrégat).²¹ Il affiche ainsi une distribution statistique de donnée. Il n'y a pas d'espace entre deux barres (ratio 1:0) ce qui permet de favoriser la forme de la distribution.⁶

La nature du graphique (barre) implique un calcul d'agrégat pour chaque barre.²⁹ La largeur de l'agrégat est à évaluer par le concepteur au regard du niveau de détail souhaité.²⁹

L'histogramme n'est pas adapté aux cas suivant :

- l'affichage de l'ensemble des valeurs, pour cela il convient d'utiliser une [Courbe de densité](#).²⁹
- s'il y a peu de valeurs dans le jeu de données,³⁰ préférez l'utilisation d'une [Bande de valeur](#).
- l'affichage de plusieurs distributions,²⁹ utilisez une décomposition en [Matrice de graphique](#).

L'utilisation d'un histogramme permet d'identifier des schémas de distributions tels que :¹⁷

- **Incliné à gauche** : les données sont majoritairement sur la gauche du graphique

- **Incliné à droite** : les données sont majoritairement sur la droite du graphique
- **Bimodale** : lorsque deux pics de valeurs sont constatables
- **Multimodale** : lorsque plus de deux pics de valeurs sont constatables
- **Symétrique** : lorsqu'il y a un nombre approximativement égal de valeurs de chaque côté d'une valeur centrale
- **Uniforme** : lorsque toutes les valeurs sont distribuées de façon globalement homogène

Horizon

Permet de vérifier un équilibre entre deux borne par rapport à une base.⁶ 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

Jauge

Représentation métaphorique la vitesse d'un système.

Ligne

Inventé par Playfair en 1786³¹ Représentation graphique standard et familière pour la représentation des séries temporelles.⁶

Les lignes relient chaque point de mesure. Pour de grands jeux de données, il convient d'utiliser une ligne lissée sur l'ensemble du graphique plutôt qu'une ligne brisée par chaque point de donnée.²⁷

La base des ordonnées doit être égale ou inférieure à la plus faible valeur affichée.^{Wong2010}

Il n'existe pas de règle stricte quand à la limite de séries affichables sur un

graphique.²⁵ Il convient de s'intéresser à la perception du lecteur recherché par le concepteur pour arbitrer les choix.²⁵

- En cas d'impression, limiter le nombre de lignes simultanées à 4 au maximum,^{Wong2010}
- Si l'interaction est possible avec le graphique, il est possible de tracer toutes les lignes dans un même graphique à condition d'utiliser un système de mise en évidence,²⁵
- Ces cas échouant, séparez les lignes et tracez une [Matrice de graphique](#).

Ne pas utiliser ce graphique pour essayer d'illustrer une différence entre deux courbes.²⁵

Les axes doivent être labelisé (obligatoirement le titre de la série, optionnellement la valeur finale) directement en fin de ligne, à droite du graphique.^{9,21}

Il est possible de mettre en évidence un point spécifique en utilisant un symbole (de préférence un cercle plein) et en y apposant une étiquette affichant $f(x) = y$.

La distinction des séries est plus efficace en utilisant des teintes différentes. Dans ce cas, il convient également d'utiliser des tons différents pour assurer le maintiens des distinctions si le graphique est imprimé en nuance de gris.¹⁹ Il convient également de n'utiliser que des lignes continues pour les tracés.¹⁹ L'utilisation de symboles comme éléments différenciants entre les séries n'est pas pertinent.¹⁹

Ligne de crête

Ce graphique représente des courbes de densité séparées par un offset et est utilisé pour comparer la distribution d'une variable entre des catégories.²⁴ Parfait lorsque les schémas de distributions sont bien distincts entre les groupes.¹² Cela évite que les courbes de densités des catégories s'entrevêchent.¹⁷

Ce graphique se met très bien à l'échelle de grands jeux de données.²⁴

Il est adapté à :

- la comparaison des changements de distributions à travers le temps²⁷ associé à un gradient de couleur pour diriger le regard vers la distribution la plus récente,¹⁷
- la comparaison des évolutions de deux tendances à travers le temps.²⁴ associé à une palette de couleur binaire.

L'axe des absyesses affiche la valeur de réponse et l'axe des ordonnées affiche les catégories.²⁴

Il convient de classer les catégories suivant un ordre logique.¹⁷ Voir (insérer renvoi ici vers l'explication de la logique de classement de catégories)

Il est possible d'afficher les barres d'erreurs dans les lignes de crêtes. Ce mode de représentation s'appel un "demi-oeil".³²

Il est possible d'afficher les nuages de points dont l'estimation de la courbe de densité est issue. Ceux-ci sont affichés en dessous des courbes de densité. Cette forme de représentation est appelé "Nuage de pluie". (ref: Micah Allan ou Schwabish2021p212) Cette représentation offre à la fois une représentation générique et une visualisation précise des données.¹⁷ Bien que cette représentation soit très efficace, elle n'est pas très répandue.¹⁷ Il faut être vigilant à bien expliquer son fonctionnement lors de son utilisation dans un rapport. Dans le cas d'un nuage de pluie, il faut faire attention à ce que les points de données n'entre pas en surimposition avec les courbes de densités tracées en dessous.

Lollipop

Utilisé pour afficher la position dans une liste ordonnée lorsque l'affichage de la valeur est moins important que la position.⁶ Egalement très efficace comparé aux barres hordonnées lorsqu'il y a un grand nombre de catégorie dans le classement.¹² Si une ligne connecte le point de la valeur à laxe des ordonnées, l'axe des absisse doit commencer à 0, autrement cela n'est pas obligatoire.

Marqueur

Affiche une valeur simple, plus pertinent lorsqu'associé à une tendance sur une période donnée.¹²

Marimeko

Visualisation de la taille et des proportions d'un ensemble homogène de données.⁶ Exemple : répartition des parts de marché de 3 entreprises sur un axe et taux dans 3 catégories de marché pour chacune

La couleur peut être utilisée pour mettre en évidence une valeur.¹⁸ Attention au biais de corrélation lorsque ce graphique est utilisé pour représenter plusieurs variables.¹⁸

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.²¹ La taille des bulles représente la force de la corrélation. Cette force est la représentation du coefficient de Pearson.¹⁶

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.

Matrice de graphique

Origine Est utilisé pour comparer de multiples graphiques homogènes organisées en matrices. Chaque entité affiche le graphique d'une catégorie. Les graphiques sont tous à la même échelle.²¹

Conseils : Limitez les détails dans chacun des graphique.²¹ Regroupez les axes pour n'afficher qu'un seul axe Y par ligne et un seul axe X par colonne.

Matrice de température

Représentation graphique d'une table de valeur homogène.²¹ 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.²⁰

Mini-graphe

Affiche la tendance d'évolution d'une variable ou d'une catégorie. Est généralement intégré dans un tableur. Un mini-graphe ne possède pas d'axes ni d'autres détails et est composé uniquement d'une courbe et d'un point indiquant la dernière valeur du jeu de donnée.²¹ Cette représentation est très macroscopique et vise à susciter un sentiment plutôt que d'exposer une valeur précise.¹⁷

Un mini-graphe ne possède pas d'axes ni d'autres détails.

Nuage de bulle

Il s'agit d'un nuage de point dont la taille des bulles est utilisé pour représenter une troisième valeur quantitative.⁶

Nuage de mots

Affiche des nuages de mots organisés en groupes sémantiques. La taille des mots indique leur fréquence d'utilisation dans le corpus étudié. La couleur des groupes peut représenter un choix de données qualitative (catégorisation).²¹ Cette visualisation est

une méthode d'exploration de texte utile pour réaliser une synthèse macroscopique. Il convient de filtrer les mots vides dit "stopwords" pour nettoyer la composition des nuages.¹¹ Des listes de mots vides peuvent être trouvées à l'adresse <https://github.com/stopwords-iso/>.

Nuage de point

Représente la relation entre deux variables continues.⁶ La couleur est utilisée pour distinguer les catégories. Si des milliers de points 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 point connecté

Montre l'évolution des données pour deux variables pertinentes si le modèle de progression est clair.⁶

Pour souligner la direction de l'évolution entre deux points il est possible de :

- Flécher la direction de l'évolution entre deux points si les points sont bien espacés et qu'un seul jeu de donnée est affiché,¹⁸
- Utiliser un gradient de couleur du plus clair vers le plus foncé.

Ce graphique sert à synthétiser des variations de valeurs, il est préférable de l'utiliser pour afficher des jeux de données dont la variation n'est pas mesurée régulièrement pour éviter d'avoir trop de points à relier et donc d'éviter de créer un graphique trop encombré.¹⁸

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. La zone précédent ce point est mise en évidence. Est utilisé pour illustrer les éléments important dans un système qualité.

Pente

Affiche la variation des valeurs ou ratios des catégories entre deux états donnés.⁶

Pictographe

Utilisation hautement contextuelle limiter strictement à l'affichage d'entiers naturels.⁶ Les pictographes peuvent augmenter significativement la précision d'acquisition d'un écart de ratio ainsi que la rapidité de la visualisation d'un minimum ou d'un maximum entre des catégories.²² A la manière des symboles groupées, si la transmission de valeurs précise est importante, il convient d'utiliser un graphique en barre ordonnée à la place.²²

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 4⁶ ce qui correspondant notamment au nombre de symboles aisément discriminables.

Quadrillage

Adapté à l'affichage des informations en pourcentages. Fonctionne mieux avec des entiers naturels.⁶ Généralement en grille de 10 x 10 carrés (pour un total de 100). Chaque

carré représente 1% d'un tout. Les carrés sont colorés sur la base de la taille d'une catégorie dans un groupe.¹²

Radar

Représentation compacte de la valeur de variables hétérogènes pour un seul système.⁶ Les radars sont utilisés pour identifier les variables ayant des valeurs similaires et les exceptions.²¹

La représentation des valeurs se fait à travers la forme globale du polygone. Il convient de coloriser l'intérieur de celui-ci pour bien faire ressortir sa forme.¹⁶

Utiliser une disposition en small-multiples plutôt que de surimposer plusieurs polygones de valeurs.¹⁶

La représentation des variables est fortement corrélée à la forme du polygone plutôt qu'aux valeurs des sommets.²¹ De ce fait, pour améliorer la forme du polygone et offrir une représentation avec le moins de distortions possibles, il convient d'homogénéiser les échelles des axes.

Réseau

Utilisé pour montrer la force et l'interconnectivité des relations de types différents.⁶ La position des nœuds n'a pas de fonction logique. Elle vise à rendre le diagramme le plus lisible possible.¹²

Ruisseau

Vérifier la métaphore liée au nom

A privilégier lorsque la visualisation des changements en proportions au fil du temps est plus importante que les valeurs individuelles.⁶

Sankey

Créé par : Matthew Henry Phineas Riall Sankey (1898) L'épaisseur représente la part d'un événement. Chaque flèche représente

l'entrée ou la sortie d'élément sur le système étudié.¹²

Les changements sont liés à une évolution temporelle et à l'association de catégories. La force de ce graphique est qu'aucun de ces deux aspects ne prévaut sur l'autre.¹⁸

Très utile pour représenter les flux entre des catégories ou entre des états.¹⁸

Il convient néanmoins de limiter le nombre de catégories affichées pour maintenir un bon niveau de lisibilité.¹⁸

Sismographe

Pour montrer des séries à grandes variations dans les données.⁶

Symbole groupé

Utile lorsqu'il est utile de compter ou de mettre en évidence des éléments individuels organisés en barre ou en colonne.⁶ Cependant, ce type de graphique n'est pas efficace dans cette action car il incite le lecteur à compter chaque point, ce qui peut s'avérer décourageant.³³ Un certain usage peut être imaginé à des fins décoratives visant à illustrer un élément déjà explicité dans un texte. Un graphique en barre ordonné est à privilégier aux symboles groupés.

Tiret

Présentation d'une valeur mesurée ou calculée au regard d'une valeur cible.⁶ Pertinent lors de la réalisation d'un benchmark.¹²

Tunnel

Utilisé pour suivre l'engagement d'une population (taille) à chaque étape d'un processus (barre).¹²

Treemap

Utilisé pour les relations hiérarchiques d'une partie vers l'ensemble. Difficile à lire s'il

y a trop d'éléments affichés.⁶ Le découpage des boîtes n'a pas besoin d'être ordonné, il peut y avoir plus de 2 niveaux de hiérarchie.¹²

Valeur lettrée

Représente chaque découpage statistique ($x\sigma$) par une boîte distincte. Convient en remplacement des boîtes à moustache lorsqu'il y a un très grand nombre de données disponibles. Permet de faire des estimations fiables.^{34,12}

Violon

Représente fidèlement les distributions d'un très grand nombre de données.⁶ Utilisé pour faire des comparaisons entre des catégories.¹² Très pertinent pour représenter les distributions bimodales.²⁴ 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.

Voronoi

Semblable à un diagramme de bulle mais transformées en zones dont l'ensemble des points de chacune est plus proche du centre de sa zone que de tout autre centroïde.⁶

Il est très pertinent pour illustrer les parties d'un ensemble lorsque de nombreuses sous-catégories sont à afficher.²⁶ En ce sens, il est préféré aux camemberts.

Etudes futures

Des études complémentaires pourraient être menées sur les sujets suivants :

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

Evolution

Date	Changements
2025-02-10	Rédaction du document

Glossaire

Médiane définition

Acronymes

ML Machine Learning

Références

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. Centering Accessibility in Data Visualization. In: SCHWABISH, Jonathan; POPKIN, Sue; FENG, Alice. *Centering Accessibility in Data Visualization*. Urban Institute, 2022, pp. 4–8. Do Not Harm Guide, no. 3.
3. FRANK ELAVSKY. 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. Table Design. In: STEPHEN FEW. *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. Qualitative. In: JONATHAN SCHWABISH. *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. MIKE YI. *How to Choose the Right Data Visualization*. Chartio, 2020. Available also from: <https://res.cloudinary.com/dun3slcfg/image/upload/v1711993110/cloud-files/How-to-Choose-the-right-data-visualization.pdf>.
13. STEPHEN FEW. *Show Me the Numbers - Designing Tables and Graphs to Enlighten*. Seconde Edition. Burlingame, California: Analytics Press, 2012. ISBN 978-0-9706019-7-1.
14. *Ergonomie de l'interaction Homme-Système - Partie 125 : Recommandations Relatives à La Présentation Visuelle d'informations*. AFNOR, 2017. Version de 2018-03-P. No. NF EN ISO 9241-125.
15. *Date et Heure - Représentations Pour l'échange d'information - Partie 1: Règles de Base*. 2019. No. ISO 8601-1.
16. Relationship. In: JONATHAN SCHWABISH. *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).

17. Distribution. In: JONATHAN SCHWABISH. *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).
18. Comparing Categories. In: JONATHAN SCHWABISH. *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).
19. Component-Level Graph Design. In: STEPHEN FEW. *Show Me the Numbers - Designing Tables and Graphs to Enlighten*. Second Edition. Burlingame, California: Analytics Press, 2012, pp. 205–256. ISBN 978-0-9706019-7-1.
20. Visualizing Amounts. In: WILKE, Claus O. *Fundamentals of Data Visualization* [online]. 1st ed. O'Reilly Media, 2019, pp. 45–58 [visited on 2024-12-30]. ISBN 978-1-4920-3108-6. Available from: <https://clauswilke.com/dataviz/>.
21. The Graphics. In: SOSULSKI, Kristen. *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.
22. TRAN, Tien; LEE, Hae-Na; PARK, Ji Hwan. Discovering Accessible Data Visualizations for People with ADHD. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems* [online]. Honolulu HI USA: ACM, 2024, pp. 1–19 [visited on 2025-01-22]. ISBN 979-8-4007-0330-0. Available from DOI: [10 . 1145 / 3613904.3642112](https://doi.org/10.1145/3613904.3642112).
23. Visualizing Associations among Two or More Quantitative Variables. In: WILKE, Claus O. *Fundamentals of Data Visualization* [online]. 1st ed. O'Reilly Media, 2019, pp. 117–130 [visited on 2024-12-30]. ISBN 978-1-4920-3108-6. Available from: [https : / / clauswilke . com / dataviz/](https://clauswilke.com/dataviz/).
24. Visualizing Many Distributions at Once. In: WILKE, Claus O. *Fundamentals of Data Visualization* [online]. 1st ed. O'Reilly Media, 2019, pp. 81–92 [visited on 2024-12-30]. ISBN 978-1-4920-3108-6. Available from: [https : / / clauswilke . com/dataviz/](https://clauswilke.com/dataviz/).
25. Time. In: JONATHAN SCHWABISH. *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).
26. Part-to-Whole. In: JONATHAN SCHWABISH. *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks* [online]. New York: Columbia University Press, 2021, pp. 289–310 [visited on 2025-01-09]. ISBN 978-0-231-55015-4. Available from DOI: [10 . 7312 / schw19310](https://doi.org/10.7312/schw19310).
27. Directory of Visualizations. In: WILKE, Claus O. *Fundamentals of Data Visualization* [online]. 1st ed. O'Reilly Media, 2019, pp. 37–44 [visited on 2024-12-30]. ISBN 978-1-4920-3108-6. Available from: <https://clauswilke.com/dataviz/>.
28. Visualizing Nested Proportions. In: WILKE, Claus O. *Fundamentals of Data Visualization* [online]. 1st ed. O'Reilly Media, 2019, pp. 105–116 [visited on 2024-12-30]. ISBN 978-1-4920-3108-6. Available from: [https : / / clauswilke . com/dataviz/](https://clauswilke.com/dataviz/).
29. Visualizing Distributions: Histograms and Density Plots. In: WILKE, Claus O. *Fundamentals of Data Visualization* [online]. 1st ed. O'Reilly Media, 2019, pp. 59–70 [visited on 2024-12-30]. ISBN 978-1-4920-3108-6. Available from: <https://clauswilke.com/dataviz/>.

30. WEISSGERBER, Tracey L.; MILIC, Natasa M.; WINHAM, Stacey J.; GAROVIC, Vesna D. Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLOS Biology* [online]. 2015, vol. 13, no. 4, pp. 1–10 [visited on 2025-01-22]. ISSN 1545-7885. Available from DOI: [10.1371/journal.pbio.1002128](https://doi.org/10.1371/journal.pbio.1002128).
31. Becoming Visual. In: SOSULSKI, Kristen. *Data Visualization Made Simple: Insights into Becoming Visual*. New York: Routledge, Taylor & Francis Group, 2019, pp. 1–24. ISBN 978-1-138-50387-8 978-1-138-50391-5.
32. Visualizing Uncertainty. In: WILKE, Claus O. *Fundamentals of Data Visualization* [online]. 1st ed. O'Reilly Media, 2019, pp. 181–206 [visited on 2024-12-30]. ISBN 978-1-4920-3108-6. Available from: <https://clauswilke.com/dataviz/>.
33. Silly Graphs That Are Best Forsaken. In: STEPHEN FEW. *Show Me the Numbers - Designing Tables and Graphs to Enlighten*. Seconde Edition. Burlingame, California: Analytics Press, 2012, pp. 271–294. ISBN 978-0-9706019-7-1.
34. 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).