# Visualisation des données et analytique d'affaires

Shadi Shuraida

## Introduction

Dans le cadre de l’analytique d’affaires, qui est le processus d'interprétation des données afin d'améliorer la prise de décision au sein de l'organisation, la visualisation des données commence seulement à être appréciée à sa juste valeur pour les avantages qu'elle peut apporter aux organisations. La visualisation des données, c'est-à-dire l'utilisation d'images pour représenter des informations, est utilisée dans toutes les organisations par les employés qui ont besoin d'interagir avec les données et de les analyser pour y déceler des modèles significatifs, des tendances et des anomalies (Stodder, 2013). En effet, face au déluge de données, la visualisation des données est devenue essentielle pour la compréhension et la prise de décision au sein des organisations. Comme l'a déclaré Edward Tufte dans son ouvrage séminal, The Visual Display of Quantitative Information, « les graphiques révèlent les données » (Tufte, 2001, p.13). Cela dit, comme nous le verrons dans les sections suivantes, si de bonnes visualisations de données sont essentielles pour prendre de meilleures décisions, de mauvaises visualisations de données peuvent embrouiller ou induire en erreur les utilisateurs.

Le rôle des technologies de l'information (TI) dans ce domaine a été crucial pour faire progresser la visualisation des données. Ces technologies ont fourni aux utilisateurs des outils puissants qui leur permettent d'interpréter les données et de communiquer leurs connaissances à d'autres personnes.

Ce chapitre est structuré comme suit. La section suivante présente un bref historique de l'évolution des visualisations de données, suivi d'une discussion sur l'importance et les avantages de l'utilisation des visualisations de données. Ensuite, une vue d'ensemble de la littérature concernant ce qui rend les visualisations de données efficaces est présentée. Vient ensuite un résumé de l'utilisation des visualisations de données d'affaires, y compris les tableaux de bord. Enfin, ce chapitre se termine par une discussion et quelques démonstrations de visualisations trompeuses.

## Brève histoire de la visualisation de données

La visualisation des données est souvent considérée comme un développement relativement moderne dans le domaine de l'analyse. Mais en fait, la représentation graphique des informations est profondément enracinée dans les premières cartes, graphiques et lignes de temps (Sharda et al., 2018).

Plus précisément, l’histoire de la visualisation des données remonte à au moins 200 ans av. J.-C. Les visualisations de cette époque ont pris la forme des tableaux décrivant la position des étoiles et d'autres astres, ainsi que lors de l'élaboration de cartes destinées à faciliter la navigation et l'exploration (Few, 2007). L'idée de coordonnées a été utilisée par les géomètres égyptiens de l'Antiquité pour le tracé des villes, òu les positions terrestres et célestes ont été localisées par quelque chose qui s'apparente à la latitude et à la longitude, au moins jusqu'en 200 av. J.-C. Plus tard, le mathématicien et astronome Claude Ptolémée (a créé une projection cartographique d'une terre sphérique en latitude et longitude à Alexandrie qui a servi de référence pendant plus de mille ans (McCluskey, 1998).

Au 16e siècle, les techniques et les instruments d'observation et de mesure précises, tels que la triangulation pour cartographier les lieux avec exactitude, ont marqué le début du domaine de la visualisation des données (Friendly et Wainer, 2021). Avec l'expansion territoriale au XVIIe siècle, les questions de temps, de distance et d'espace sont devenues primordiales pour l'élaboration des cartes et la navigation. Ce siècle est également marqué par les progrès de la géométrie (par exemple les coordonnées cartésiennes), de la mesure et de la statistique (Friendly et Wainer, 2021). C’est ainsi qu’au 17e siècle que la représentation visuelle de données quantitatives par rapport à des échelles de coordonnées bidimensionnelles, la forme la plus courante de ce que nous appelons les graphiques est apparue. Mais la première visualisation de données statistiques peut être attribuée à l’astronome et cartographe Michael Florent van Langren, en 1644, qui a utilisé une seule dimension pour montrer les 12 estimations connues à l’époque de la différence de longitude entre Tolède et Rome présenté dans figure 6.1 (Friendly et al., 2017).

Une image contenant texte, Police, reçu

Description générée automatiquement

Figure 6.1. Le graphique de van Langren de 1644 sur les déterminations de la distance en longitude de Tolède à Rome. Source : Friendly et al. (2017)

Bien que ces visualisations aient été à l'origine de la présentation graphique d'informations quantitatives, la plupart des développements ont eu lieu au cours des deux derniers siècles et demi. Plus précisément, on attribue à William Playfair l’invention des graphiques modernes, y compris les diagrammes circulaires, et les graphiques à barres et à lignes. L’ingénieur et économiste politique écossais a déclaré la guerre aux tableaux. Le graphique présenté dans la figure 6.2 était publié dans son livre *Commercial and Political Atlas* en 1786, et montre la balance commerciale entre l’Angleterre d’un côté, et le Danemark et la Norvège de l’autre, entre 1700 et 1780.

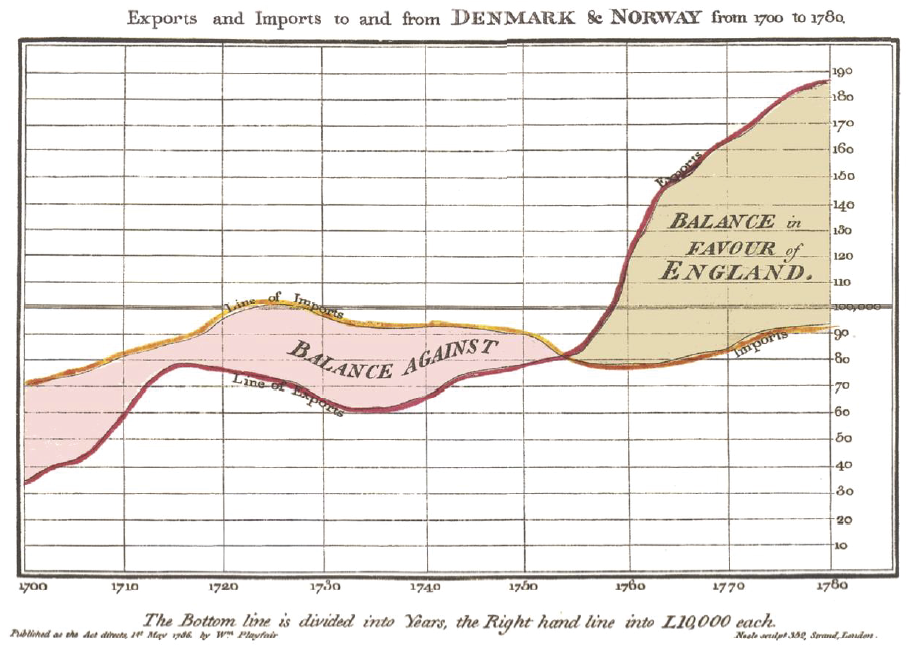


Figure 6.2. Balance commerciale entre l'Angleterre et le Danemark et la Norvège de 1700 à 1780. Source : <https://www.historyofinformation.com/detail.php?entryid=2929>

La seconde moitié du 19e siècle est considérée comme l'âge d'or des graphiques statistiques. Cette période a été marquée par des innovations et une croissance rapide des visualisations. Ces innovations ont également été rendues possibles par la création d'offices statistiques dans toute l'Europe, qui ont collecté des données quantitatives dans le but de la planification sociale, le commerce et les transports (Friendly et al., 2017).

Un excellent exemple de visualisations de cette ère ayant conduit à des découvertes qui n'auraient peut-être pas été faites facilement autrement est le graphique populaire créé par le médecin John Snow dans les années 1850. La visualisation cartographique présentée dans la figure 6.3 montre l'emplacement des décès dus au choléra dans le centre de Londres en septembre 1854. Dans cette visualisation, les décès sont marqués par des points et les onze pompes à eau de la région sont marquées par des croix (Tufte, 2001).

Une image contenant diagramme, Plan, schématique, Dessin technique

Description générée automatiquement

Figure 6.3. Les flambées de choléra lors de l'épidémie londonienne de 1854. Source : Tufte (2001, p. 24)

À l’époque, les germes n’étaient pas encore compris et les gens croyaient que le choléra se propageait par ce que l’on appelle les miasmes dans l’air. L’épidémie soudaine et grave de choléra dans le Soho de Londres était donc un mystère (Rogers, 2013). Cette cartographie a permis d’observer que les cas étaient regroupés autour de la pompe à eau de rue Broad (aujourd’hui Broadwick), ce qui a amené John Snow à penser que le choléra était dû à l’eau contaminée plutôt qu’à l’air (Rogers, 2013). À la suite de cette observation, Snow a fait retirer la poignée de la pompe, mettant ainsi fin à la pandémie qui a coûté la vie à plus de 500 personnes (Tufte, 2001).

Mais l'un des graphiques multidimensionnels les plus populaires de cette époque est peut-être celui créé par Charles Joseph Minard en 1869 (voir figure 6.4). Dans ce graphique, Minard décrit la décimation de l'armée napoléonienne lors de la campagne de Russie de 1812. Ce qui est impressionnant dans ce graphique, c'est qu'il présente six types de données en deux dimensions : le nombre de soldats de Napoléon, la distance parcourue, la température, la latitude et la longitude, la direction et la position de l’armée par rapport à des dates spécifiques.

A chart of Napoleon’s 1812 march to Russia shows the number of soldiers during the march and retreat, geography, and temperature and data of the retreat. Text in the chart, translated from the map reads, Figurative Map of the successive losses in men of the French army in the Russian campaign 1812 through 1813. Drawn by Mr. Minard, inspector general of bridges and roads in retirement. Paris, 20 November 1869. The numbers of men present are represented by the widths of the colored zones in a rate of one millimeter for ten thousand men; these are also written beside the zones. Red designates men moving into Russia, black those on retreat. The information used for drawing the map were taken from the works of Messieurs Chiers, de Ségur, de Fezensac, de Chambray and the unpublished diary of Jacob, pharmacist of the Army since 28 October. In order to facilitate the judgment of the eye regarding the diminution of the army, I supposed that the troops under Prince Jèrôme and Marshal Davoust, who were sent to Minsk and Mobilow and who rejoined near Orscha and Witebsk, had always marched with the army. The chart has a temperature scale that reads, after translation from French, Graphic table of the temperature in degrees of Réaumur thermometer. The march started after crossing Niemen river into Kowno with 422,000 men. 22,000 men went north, while 400,000 continued the march to Vilna. 60,000 crossed Vilna moved northeast towards Glubokoe, and 33,000 marched towards Polotsk. The remaining 175,000 marched through Vitebsk. 145,000 men crossed Smolensk. They passed Doroboy and Gjat, where 187,000 men crossed the Moskowa River. 100,000 men reached Moscow. The retreat began on October 18 towards Tarantino and Malo-jarosewli. The temperature was 0 degrees. 96,000 men crossed Malo-jarosewli and moved towards Mojaisk. On October 24, the temperature was 0 degrees, and the weather was rainy. 87,000 men crossed Mojaisk into Wirma, and 55,000 men crossed Wirma. The temperature on November 9 was minus 9 degrees, and 37,000 men were alive. On November 15, the temperature was minus 21 degrees, and 24,000 men were alive. 20,000 men crossed Orscha into Botr, and the temperature was minus 11 degrees. 30,000 men from Polotsk joined the troop, and 50,000 men moved towards Studienska. The temperature was minus 20 degrees on November 28. 28,000 men crossed Studienska. The temperature on December 1 was minus 24 degrees. The temperature on December 6 was minus 30 degrees, and 12,000 men crossed Molodzeno into Smorgoni. The troops moved towards Neimen River. 12,000 men increased to 14,000, and then diminished to 8,000 men on December 7, when the temperature was minus 26 degrees. The number of men reduced to 4,000, and were joined by men from north of Kowno. 10,000 men crossed the Niemen River.

Figure 6.4. Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813. Source : <https://en.wikipedia.org/wiki/Charles_Joseph_Minard#The_map_of_Napoleon's_Russian_campaign>, consulté le 7 décembre 2023.

Contrairement aux 50 années qui l'ont précédée, la première moitié du 20e siècle a connu peu d'innovations en matière de visualisation de données. Tandis qu'un certain nombre de découvertes scientifiques en astronomie, physique, biologie et autres résultent de la popularisation des graphiques au cours de cette période, une plus grande importance a été accordée à la quantification et aux modèles statistiques formels dans les sciences sociales (Friendly and Denis, 2005).

La renaissance de la visualisation des données est apparue au milieu des années 1960 avec la confluence du développement des méthodes d'analyse statistique des données, et de l'avènement de la technologie de l’information et des appareils d'affichage (Friendly et al., 2017). En 1977, John Tukey, professeur de statistiques à l'université de Princeton, a créé une approche graphique simple et efficace pour explorer les données et leur donner un sens dans ce qu'il a appelé l'analyse exploratoire des données (Few, 2007).

Avant l'avènement des ordinateurs personnels, la création de graphiques visuels était un processus à forte charge de travail qui nécessitait des compétences et des matériaux spéciaux. Mais les progrès réalisés depuis dans le domaine du hardware, ainsi que l'accessibilité et l'omniprésence des logiciels, ont entraîné une explosion des méthodes de visualisation des données et des types de données. À l'aide d'applications informatiques avancées, les organisations aujourd’hui peuvent traiter de grandes quantités de données, qui leur fournissent des visualisations qui les guident dans leur prise de décision.

Cependant, cet intérêt soulève des questions essentielles : pourquoi un tel intérêt pour les visualisations de données (importance) et comment peuvent-elles être utiles à la prise de décision au sein de l'organisation (forme) ?

## Pourquoi visualiser les données ?

Les avantages des visualisations de données sont soulignés dans de nombreuses publications et sont souvent présentés comme l'un des principaux chapitres des livres d'introduction à l'analyse des données (par exemple, Sharda et al., 2018, Jaggia et al., 2023). Cependant, l'importance des visualisations est présentée de manière remarquable par l'exemple du quartet d'Anscombe (1973) présenté dans la figure 6.5 (Tufte, 2001). À l'époque où le statisticien anglais a publié son article en 1973, les gens débattaient encore de la nécessité des graphiques de données (Schneider et Dineen, 2013). Alors il a créé un quartet d’ensemble de données fictives. Si on examine les statistiques sommaires des données présentées dans la figure 6.5, ces quatre ensembles de données semblent identiques : les quatre ensembles de données sont décrits par exactement le même modèle linéaire. Toutefois, ces ensembles de données sont très différents. Voyez-vous la différence dans les données présentées dans la figure 6.5 ?

Une image contenant texte, Police, capture d’écran, nombre

Description générée automatiquement

Figure 6.5. Quartet d’Anscombe. Source : Tufte (2001, p. 13).

Lorsque les quatre ensembles de données visualisés (figure 6.6), elles révèlent des schémas de relations remarquablement différents entre les dimensions x et y. L'idée du quartet d'Anscombe est simple et puissante : les différences peuvent être clairement révélées par la représentation graphique des données, et les statistiques sommaires traditionnelles peuvent être trompeuses.

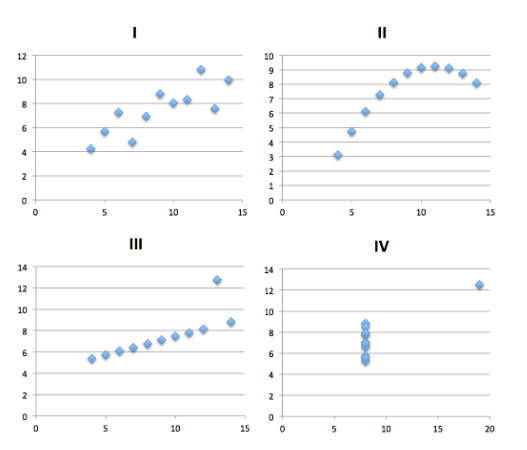


Figure 6.6. Graphique du quartet d'Anscombe (I - IV) ensembles de données.

Avec cet exemple, Anscombe (1973) a pu contrer l'idée de l'époque selon laquelle « les calculs numériques sont exacts, mais les graphiques sont approximatifs » et que « faire des calculs complexes est vertueuse, alors que regarder réellement les données est une tricherie » (p. 17). Mais outre le fait que la visualisation des données constitue une partie importante de l'analyse des données, cet exercice démontre également d'autres raisons pour lesquelles la visualisation des données est précieuse : ils révèlent les données (Tufte, 2001). En d'autres termes, les données présentées dans la figure 6.6 sont exactement celles présentées dans la figure 6.5. Pourtant, la figure 6.6 nous a aidés à mieux comprendre. Comment expliquer cela ?

Les spécialistes des sciences cognitives et les professionnels s'accordent à dire que la présentation graphique des données augmente la cognition (Hegarty, 2018) en soutenant la compréhension et le jugement de plusieurs manières (Zheng, 2017 ; Dudycz, 2010).

La première est que les visualisations organisent l'information dans une perspective spatiale. Les visualisations relationnelles organisent souvent l'information de manière que les représentations des entités liées soient proches, ce qui facilite la recherche et l'intégration de l'information (Chen et al., 2009), c'est-à-dire la compréhension. Les graphiques organisent les entités en les plaçant dans un espace défini par les axes x et y. Par conséquent, les entités similaires sont visualisées comme étant proches les unes des autres (Hegarty, 2018). Par exemple, le diagramme de dispersion en figure 6.7 démontre la relation entre l’espérance de la vie et le PIB par habitant de 177 pays.

Figure 6.7. Relation entre PIB par habitant et l’espérance de la vie à la naissance. Source de données : Données des Nations Unies sur l'espérance de vie, de https://population.un.org/wpp/Download/Standard/Mortality/, et les données de la Banque mondiale sur le PIB par pays, https://data.worldbank.org/indicator/NY.GDP.MKTP.CD, récupérées le 13 décembre 2023.

Dans ce graphique, on peut voir que des éléments similaires sont regroupés, par exemple le groupe des pays dont le PIB est inférieur à 20 000$ ou ceux dont le PIB est supérieur à 60 000$. Il faudrait une incroyable capacité cognitive humaine pour identifier cette organisation à partir des données textuelles des 177 pays.

Ainsi, l'organisation visuelle des informations réduit la charge cognitive nécessaire au traitement des informations et facilite la mémorisation, le rappel et la compréhension des informations (Borkin et al., 2013 ; Zheng, 2017). Cette réduction de la charge cognitive libère des ressources cognitives pour d'autres aspects de la réflexion, par exemple l'analyse de la relation entre les variables (Hegarty, 2018). Si nous examinons par exemple la relation entre l'espérance de vie et le PIB dans la figure 6.7, nous pouvons ensuite voir comment des éléments similaires tels que l'espérance de vie changent en fonction des différents niveaux du PIB (faible, moyen et élevé). C'est ainsi que les visualisations efficientes donnent un aperçu des structures, des relations et des tendances dans les données, ce qui nous permet d'interpréter et de comprendre facilement des ensembles de données potentiellement complexes (Zheng, 2017). Les visualisations nous permettent également d'identifier les zones de différence telles que les anomalies qui peuvent être obscures dans les textes ou les tableaux (Dudycz, 2010). Ceci est illustré dans la figure 6.7, ou l’on peut facilement déterminer si un pays dispose d’un PIB élevé, mais d’une espérance de vie plus faible, par exemple.

En tant que telle, une représentation graphique efficace des données simplifie l'analyse et la prise de décision dans le domaine de l'analyse des affaires. Cela dit, la question se pose de savoir ce qui constitue une visualisation efficace (Borkin et al., 2013; Alhadad, 2018; Hegarty, 2018).

## Qu'est-ce qui rend une visualisation efficace ?

Une visualisation de données efficace est une visualisation qui améliore la compréhension de l'information par l'individu (Hegarty, 2018 ; Alhadad, 2018). Ainsi, pour comprendre ce qui fait une visualisation efficace, on se réfère aux modèles cognitifs qui permettent de comprendre les visualisations d'informations. En ce sens, deux conclusions importantes et solides de la recherche sur la psychologie cognitive et la compréhension des graphiques peuvent nous aider à identifier ce qui constitue des visualisations de données efficaces.

1. La première est que l'efficacité de la visualisation d'un ensemble de données doit tenir compte de la tâche à effectuer avec ce visuel et de l'utilisateur de ce visuel (Hegarty, 2018; Alhadad, 2018).
2. La seconde, qui concerne l'utilisateur du visuel, est que la capacité humaine de traitement de l'information est limitée (Alhadad, 2018).

La figure 6.8 présente un modèle de compréhension de la visualisation d'informations adapté de Hegarty (2018). Conformément aux modèles antérieurs (par exemple, Carpenter et Shah, 1998), la compréhension des graphes implique l'interaction entre les processus de visualisation et de connaissance individuelle.

Une image contenant noir, obscurité

Description générée automatiquement

Figure 6.8. Modèle des différentes représentations (indiquées par des cases) et processus (indiqués par des flèches) impliqués dans la compréhension d'une visualisation d'information. Source : Hegarty (2018, p. 21, traduction libre).

Plus précisément, la capacité limitée de la personne à traiter l'information fait que seule une petite quantité d'informations visuelles peut être traitée à un moment donné (Alhadad, 2018). En conséquence, lorsqu'il est confronté à des entrées multiples (visuels dans ce cas), le système cognitif donne la priorité à certains éléments pour la suite du traitement, tout en en ignorant d'autres. Comme le montre la figure 6.8, l'attention est soit focalisée volontairement par un mécanisme descendant dirigé par les objectifs de l'observateur (processus descendant), soit par ce qui est visuellement saillant dans le graphique (processus ascendant) (Kahneman, 2003).

Par conséquent, l'information que les individus reconnaissent lorsqu'ils font passer du visuel à leur système de mémoire (encodage) dépend 1) de l'attention qu'ils portent à certains éléments du graphique (Carpenter et Shah, 1998), et 2) de leurs connaissances préalables des conventions du schéma du graphique (Alhadad, 2018; Hegarty, 2018). Ceci est important pour la conception graphique, car la première étape pour comprendre les informations importantes et pertinentes d'un visuel est liée à leur saillance dans le graphique et à la connaissance par l'utilisateur des conventions graphiques. Dans ce cas, des informations très saillantes, mais non pertinentes pour la tâche, telles que des images, peuvent détourner l'attention de l'utilisateur de l'encodage des informations pertinentes pour la tâche. Selon Tufte (2001), c'est le principe fondamental de la conception des graphiques : « avant tout, montrez les données » (p. 92, traduction libre de la citation).

De même, un utilisateur peut ne pas encoder les informations importantes d'un graphique s'il ne connaît pas la convention graphique présentée. Une convention consiste par exemple à représenter la variable indépendante sur l'axe des x et la variable dépendante sur l'axe des y, ou à considérer qu'une ligne droite représente une relation linéaire. Afin de comprendre pleinement la signification du visuel, l'utilisateur de l'information doit également faire d'autres déductions basées sur la connaissance du domaine. Dans la figure 6.7, par exemple, un décideur politique du gouvernement du Québec peut avoir besoin de connaître des informations non présentées dans le graphique, comme l’éventail moyen des valeurs du PIB et de l'espérance de vie au Canada. En effet, certaines études ont montré que les spectateurs commettaient des erreurs systématiques dans l'interprétation des graphiques lorsque leurs connaissances préalables étaient incompatibles avec les informations présentées dans le graphique (Shah et Hoeffner, 2002).

En substance, ce modèle montre que les caractéristiques visuelles de la visualisation des données influencent la perception, l'attention et l'encodage, tandis que l'expérience et les connaissances guident l'attention et la compréhension visuelles, et ce de manière concertée (Alhadad, 2018; Hegarty, 2018). Alors que les experts ou les novices en visualisation peuvent choisir intuitivement certaines caractéristiques de visualisation pour représenter efficacement les données, des études cognitives humaines suggèrent que ces caractéristiques sont incompatibles avec les connaissances préalables des utilisateurs et peuvent donc détourner l'attention des utilisateurs et les empêchent de comprendre les informations importantes (Zacks et al, 1998; Shah and Hoeffner, 2002).

### Principes de conception de visualisations efficaces

Sur la base des théories du traitement de l'information et des recherches antérieures sur la compréhension des visualisations, un ensemble de principes pour la conception de visualisations efficaces a été proposé (Tufte, 2001; Kosslyn, 2006; Alhadad, 2018; Hegarty, 2018). Ces principes visent essentiellement à réduire la charge cognitive (surcharge d'informations) et à permettre l'allocation de processus cognitifs pour comprendre les données et en tirer des conclusions. Il convient ici de souligner qu'il n'existe pas de meilleur visuel, sans prise en compte de la tâche et de l'utilisateur de la visualisation (Hegarty, 2018). Voici un résumé de certains principes importants tirés de cette littérature :

1. Principe de pertinence et saillance. La communication est plus efficace lorsque l'on présente suffisamment d'informations et pas plus (Tufte, 2001; Kosslyn, 2006; Hegarty, 2018). Les visualisations efficaces présentent à l'utilisateur les informations pertinentes nécessaires à l'objectif visé, sans le submerger d'encombrements visuels et de détails inutiles. En supprimant les éléments non pertinents tels que les détails excessifs, les embellissements visuels ou la densité de la mise en page, l'utilisateur alloue davantage de capacités cognitives au traitement des informations essentielles (Alhadad, 2018). Tufte (2001) considère qu'il s'agit là du principe fondamental des bons graphiques de données : « avant tout, montrer les données » (p. 92, traduction libre de la citation). Cela dit, le type d'informations contenues nécessaires à la compréhension dépend du graphique et de la tâche. Néanmoins, de ce principe découlent les conseils généraux de fournir des informations contextuelles importantes à partir desquelles les utilisateurs peuvent interpréter les données, par exemple des titres clairs et précis pour les graphiques (Kosslyin, 2006) et des étiquettes d'axes (Alhadad, 2018).

En outre, en présentant les informations nécessaires, le principe de saillance stipule que les visuels doivent être conçus pour signaler les informations les plus importantes. Ce principe est important pour attirer l'attention des utilisateurs et les aider à comprendre ce qui est important (Kosslyn, 2006).

2. Principe d'organisation perceptuelle (regroupement). Étant donné les capacités limitées des humains en matière de traitement de l'information et de mémoire, l'un des mécanismes permettant d'améliorer la mémoire et la compréhension est le regroupement perceptif (Gobet et al., 2001). Il consiste à regrouper des éléments en unités plus grandes ou plus larges en fonction de leur signification ou des associations apprises (Alhadad, 2018). Un exemple de regroupement est la façon dont nous nous souvenons de nos indicatifs téléphoniques en Amérique du Nord ; l'indicatif à trois chiffres est mémorisé comme une seule unité, plutôt que comme trois nombres distincts. Les stratégies de regroupement perceptif comprennent l'utilisation de paramètres visuels communs tels que la couleur, la forme ou l'emplacement pour faciliter le "regroupement" des informations dans des groupes compatibles avec les objectifs de la tâche. Par exemple, le regroupement de données basé sur la similitude (par l'utilisation de couleurs ou de symboles communs), la proximité (distance) ou la continuité (traçage de lignes entre les données) permet à l'utilisateur d'associer différents points de données en tant que groupes partagés. La littérature suggère que la redondance dans la combinaison de ces caractéristiques (comme la combinaison de la couleur et de la forme ; par exemple, cercles bleu vis-à-vis carrés orange) peut renforcer la segmentation pour un traitement plus efficace (Alhadad, 2018).

Dans la figure 6.9a ci-dessous par exemple, des données fictives sont fournies concernant l'effet de trois stratégies différentes sur quatre trimestres. Cette figure utilise le regroupement des stratégies en utilisant des couleurs et des symboles différents pour chaque stratégie (similarité). Dans ce cas, en supposant que l'utilisateur veuille décider d'une stratégie en fonction de son impact dans le temps, il faut distinguer deux éléments, la stratégie et son effet à travers le temps. L'organisation de l'information peut donc être améliorée si nous regroupons également les stratégies en utilisant la continuation (ligne regroupant chaque stratégie) comme dans la figure 6.9b.

Une image contenant capture d’écran, texte, affichage, logiciel

Description générée automatiquement Une image contenant texte, ligne, capture d’écran, diagramme

Description générée automatiquement

Figures 6.9a et 6.9b. Comparaison des caractéristiques visuelles d'un diagramme à points avec codage redondant (forme/couleur) entre les catégories, avec des lignes en pointillés pour faciliter le suivi au sein de chaque stratégie.

3. Principe de compatibilité tâche-affichage. En rapport avec ce qui précède, la réflexion sur la manière dont les éléments d'un affichage visuel peuvent être regroupés en entités psychologiquement significatives peut soutenir l'attention et la cognition (Pinker, 1990) d'une manière compatible avec la tâche. Par exemple, le graphique linéaire (figure 9.6b) est mieux adapté à l'intégration d'informations sur la relation entre la stratégie et le temps qu'un diagramme à barres ou un simple diagramme à points.

Ce principe s'étend également à l'utilisation de variables visuelles qui véhiculent une signification cohérente avec des suppositions spatiales et culturelles communes, telles que le haut est bon, le bas est mauvais et les éléments graphiques plus grands représentent une plus grande quantité de quelque chose. D'autres suppositions communes incluent que les lignes indiquent des connexions et que la dimension horizontale est naturellement associée au temps (Kosslyn, 2006). Par exemple, la figure 6.10a ci-dessous va à l'encontre de trois conceptions courantes qui sont corrigées dans la figure 6.10b : la première est qu'un axe horizontal doit être utilisé pour indiquer le temps (les quarts), la deuxième est que les valeurs ou la hauteur sur l'axe vertical indiquent la quantité, et la troisième est que la couleur rouge est utilisée pour indiquer les valeurs négatives.

Une image contenant texte, capture d’écran, diagramme, ligne

Description générée automatiquement Une image contenant texte, capture d’écran, nombre, Police

Description générée automatiquement

Figures 6.10a et 6.10b. Comparaison de deux diagrammes à barres qui diffèrent dans l'affichage des caractéristiques de temps, de couleur et de valeur (positive ou négative).

4. Principe de la connaissance appropriée. Toute conception de visualisation efficace doit tenir compte des connaissances des utilisateurs. En fait, tous les principes ci-dessus ne sont pas utiles si l'utilisateur de la visualisation ne dispose pas des connaissances nécessaires pour interpréter le graphique présenté. Ces connaissances comprennent les conventions d'affichage visuel et les connaissances du domaine (Kosslyn, 2006; Hegarty, 2018). Par exemple, si la personne qui utilise le visuel n'est pas familiarisée avec le type de représentation utilisé (un diagramme en boîte à moustaches, par exemple), elle ne peut pas comprendre le contenu de l'information. De même, des connaissances dans un domaine tel que les statistiques sont nécessaires pour interpréter les lignes de régression linéaire, ou des connaissances en météorologie pour comprendre les cartes météorologiques. En tant que telle, bien que de bons visuels présentent de nouvelles informations, les concepts utilisés doivent être familiers aux utilisateurs et le format d'affichage doit être compréhensible pour eux.

## Visualisation des données d’affaires

Le terme « visualisation des données d'affaires » a une signification plus large que les seules activités commerciales, et se réfère généralement à de nombreuses activités humaines et organisationnelles qui assurent le fonctionnement d'un système. Il peut s'agir d'activités commerciales, non lucratives, éducatives, sportives, de divertissement, gouvernementales et bien d'autres encore. Dans ces activités, des données sont produites et enregistrées pour refléter tous les aspects des activités organisationnelles, puis elles sont analysées et font l'objet de rapports à différents niveaux. Les données d’affaires sont généralement abstraites, quantitatives, structurées ou semi-structurées et multidimensionnelles (Dudycz, 2010 ; Zheng, 2017). Ainsi, ces données sont transformées en informations significatives et utiles qui sont consommées par les humains. La transformation de ces données sous forme de graphiques est la visualisation des données.

Les objectifs de la visualisation des données d’affaires se concentrent principalement sur l'exploration des données et la prise de décision (Stodder, 2013; Zheng, 2017). L'exploration des données permet aux utilisateurs de l'organisation de comprendre les données en révélant des tendances, en localisant les valeurs aberrantes et en montrant les relations entre différents facteurs. Supposons, par exemple, qu'un responsable des opérations de transport public ait pour objectif d'explorer les facteurs ou les tendances susceptibles d'entraîner une augmentation de la fréquentation. Les observations faites par le gestionnaire à partir des visualisations de données peuvent par exemple conduire à des idées intéressantes sur les tendances, les valeurs aberrantes ou les relations dans le temps, les données démographiques ou les lieux géographiques. Celles-ci peuvent à leur tour conduire à une exploration plus poussée et à d'autres objectifs.

De même, la visualisation des données fournit aux responsables des informations qui les aident à prendre des décisions au sein de l'organisation. Par exemple, les visualisations de données peuvent contribuer à améliorer l'affectation des ressources de sécurité publique en fonction des concentrations de criminalité (Nepomuceno & Cabral Seixas Costa, 2019) ou aider les médecins à évaluer l'état de santé de leurs patients et à agir en conséquence (Wanderer, Nelson, Ehrenfeld, Monahan, & Park, 2016).

En tant que telles, les visualisations dans les organisations servent des buts et des objectifs multiples. Une fois que vous avez déterminé les données dont vous avez besoin, la question est de savoir quel format de visualisation utiliser ?

### Choisir la bonne visualisation pour la tâche

Comme indiqué ci-dessus, il n'existe pas de graphique idéal à utiliser. Le choix du graphique dépend de l'information transmise, de l'objectif de la visualisation et du public visé. Mais une fois que vous avez identifié les informations que vous souhaitez transmettre et que vous avez réfléchi à votre public, vous devez choisir un format. Ainsi, les graphiques les plus familiers et les plus utilisés sont notés dans le tableau 6.1 (Kosslyn, 2006 ; Stodder, 2013; Jaggia et al., 2023), avec des lignes directrices du contexte de leur utilisation en termes d'objectif et de type de données (voir Kosslyn, 2006, et Jaggia et al., 2023, et Sharda et al., 2018 pour des suggestions de design plus détaillées). Il convient de noter ici que les cas d’utilisations conseillés ci-dessous sont normatifs et ne s'appliquent pas à tous les instances et les contextes.

Tableau 6.1. Les graphiques les plus familiers et leurs cas d’utilisation conseillés.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Graphique** | **Cas d’utilisation conseillés (type de données)** | | | **Cas d’utilisation conseillés (objectifs)** |
| *Variables nominales* | *Données quantitatives et données de rangs* | *Données en pourcentage et en proportion* |
| Graphique en barres ou en colonne |  |  |  | Pour montrer les valeurs relatives des points, et particulièrement utiles pour visualiser les changements dans les données. |
| Graphique linéaire |  |  |  | Lorsque les conventions définissent des tendances significatives. Par exemple, pour suivre les changements ou les tendances dans le temps. |
| Graphique circulaire |  |  |  | Pour indiquer les quantités relatives approximatives. |
| Graphique à barres empilées |  |  |  | Pour donner une impression précise des parties d'un ensemble. |
| Diagramme de dispersion |  |  |  | Pour donner une impression générale de la relation entre deux variables. Particulièrement utile pour afficher la relation entre deux grands ensembles de données. |
| Carte thermique (heat map) |  |  |  | Pour identifier les combinaisons de variables nominales qui ont une signification économique. |

Il existe également de nombreuses variantes et extensions de ces graphiques plus familiers, comme le diagramme à bulles (extension du diagramme de dispersion), et l'histogramme (variante du diagramme à barres). En outre, les logiciels spécialisés et les nouveaux besoins ont donné lieu à une explosion du nombre de représentations visuelles disponibles utilisées par les organisations. Par exemple, les visualisations de nuages de mots sont souvent utilisées par les organisations pour analyser les messages des employés et des clients sur les médias sociaux ou les flux en ligne, ou des graphiques de réseau (ou nœuds-liens) pour analyser les relations dans les réseaux sociaux. Cela dit, bien qu’il existe de nombreux types de représentations de données utilisées dans les entreprises, Kosslyn (2006) suggère que l'utilisation et la survie des graphiques familiers depuis leur invention par William Playfair en 1786 jusqu'à aujourd'hui témoignent de leur utilité.

### Tableaux de bord en analytique d’affaires

Les tableaux de bord sont devenus des éléments courants des efforts analytiques de la plupart des organisations (Few, 2007; Sharda et al., 2018). Leur conception vise à imiter celle d’un tableau de bord de voiture qui fournit au conducteur des informations sur la vitesse actuelle du véhicule, le niveau de carburant et la température du moteur, afin qu’il puisse évaluer les conditions de fonctionnement actuelles et prendre les décisions appropriées. De même, un tableau de bord de données fournit aux gestionnaires les informations les plus importantes sur un seul écran de qui leur permet d’évaluer les performances de leur organisation afin de prendre des mesures efficaces (Dudycz, 2010).

Souvent, les gestionnaires sont confrontés au défi d'être inondés d'informations, de ne pas avoir assez d'informations ou de ne pas avoir les informations dont ils ont besoin au bon moment. Le but des tableaux de bord est de fournir la bonne information, au bon moment, à la bonne personne qui peut prendre une décision informée. En tant que tels, les tableaux de bord efficaces offrent aux organisations trois fonctionnalités principales: surveiller, analyser et gérer (Eckerson, 2010).

La première est qu'il permet aux responsables d'entreprise de surveiller les processus et activités critiques de l'entreprise à l'aide d'indicateurs de performance qui déclenchent des alertes en cas de problèmes potentiels. À l'aide d'indicateurs visuels (tableaux, graphiques, cadrans, etc.), les tableaux de bord doivent présenter les mesures de performance pertinentes pour l'utilisateur. Par exemple, un tableau de bord destiné à un directeur des opérations doit fournir des informations relatives à ses décisions, telles que le stock disponible et l'utilisation des capacités, tandis qu'un tableau de bord destiné à un directeur financier doit fournir des informations sur la situation financière de l'entreprise, telle que les liquidités disponibles, les obligations courantes liées à la dette, etc.

Deuxièmement, des tableaux de bord efficaces permettent à l'utilisateur d'analyser la cause profonde des problèmes mis en évidence par la fonction de surveillance, en explorant des informations pertinentes et opportunes à partir de perspectives multiples et à différents niveaux de détail. Enfin, des tableaux de bord efficaces permettent aux utilisateurs de gérer les personnes et les processus afin d'améliorer les décisions, d'optimiser les performances et d'orienter l'organisation dans la bonne direction. Plus précisément, les tableaux de bord traduisent la stratégie de l'organisation en mesures et en objectifs concrets, adaptés aux utilisateurs visés. Cela facilite la communication avec les différents groupes au sujet des objectifs de l'organisation et de ce qu'ils doivent faire dans leurs domaines respectifs pour atteindre les buts fixés. D’ailleurs, grâce à la possibilité de surveiller les performances, les tableaux de bord permettent aux utilisateurs d'évaluer leurs actions afin d'affiner la stratégie ou d'aligner leurs actions avec l'orientation stratégique de l'organisation.

En fournissant ces trois fonctionnalités, le défi de créer un tableau de bord efficace est d'afficher toutes les informations requises sur un seul écran (Sharda et al.,2018). En tant que tel, si le respect des principes de conception susmentionnés est nécessaire, il n'est pas suffisant. Compte tenu de la grande quantité d'informations fournies, il peut être fastidieux pour un décideur d'évaluer si un graphique présente des valeurs ou des tendances acceptables. Pour faciliter le traitement de l'information, des objets ou attributs visuels spécialisés sont parfois utilisés pour fournir le contexte d'évaluation (c'est-à-dire bon, mauvais, acceptable, etc..). Comme le montre l’exemple ci-dessous d’un tableau de bord des ressources humaines (figure 6.11), certains de ces attributs comprennent l'utilisation de molettes, de feux de signalisation ou de codes de couleur.

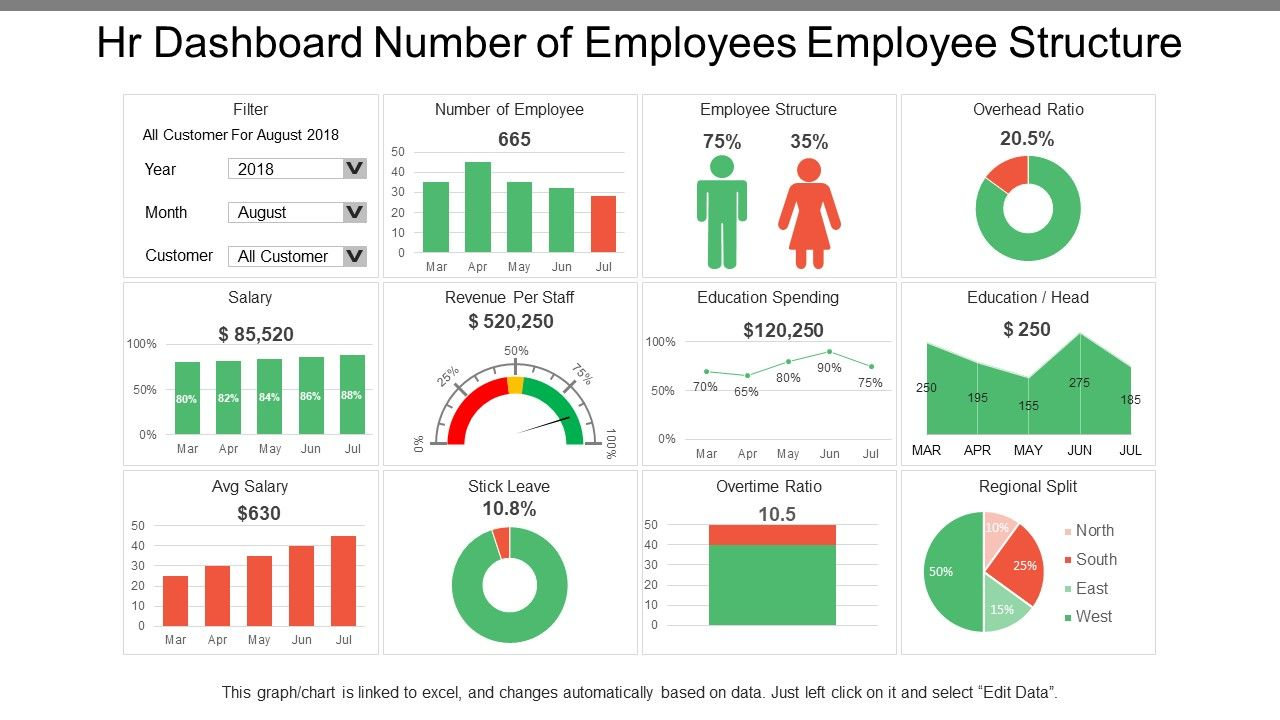


Figure 6.11. Exemple de tableau de bord des ressources humaines.

Outre le fait qu'un tableau de bord soit bien conçu, un aspect très important est son utilité. Les données doivent être fiables, actualisées et utiles pour que l'utilisateur puisse prendre des décisions informées.

## Visualisations trompeuses et conclusion

Même si les visualisations peuvent rendre l'information plus accessible, plus compréhensible et plus convaincante, elles peuvent aussi être facilement mal utilisées et mal comprises, même par leurs concepteurs. La distorsion de l'information via les visualisations de données est un problème depuis longtemps (Tufte, 2001), et sa prévalence a augmenté avec la popularité accrue des visualisations de données dans les médias populaires et sur Internet (Pandey, 2015).

Les concepteurs et les communicateurs qui utilisent des visuels peuvent, intentionnellement ou non, présenter de l'information déceptive. Si certaines visualisations trompeuses peuvent être le résultat de l'inattention ou de l'ignorance de leurs créateurs quant aux bonnes pratiques en matière de création visuelle, d'autres sont délibérément conçues pour induire en erreur et influencer le public cible (Pandey et al., 2015). Toutes les visualisations sont sujettes à la tromperie par le biais de diverses techniques qui peuvent inclure la couleur (voir figures 6.10 ci-dessus), des axes tronqués ou disproportionnés, ou des étiquettes manquantes. Pandey et al. (2015) ont démontré que les visualisations trompeuses ont un effet important sur la compréhension qu'ont les gens d'une question particulière, et ils identifient deux catégories principales dans lesquelles les visualisations peuvent être trompeuses.

La première est l'exagération ou la sous-estimation du message. Ce type de déception se produit lorsque l'information n'est pas déformée, mais lorsque l'étendue de l'information présentée est modifiée, c'est-à-dire que les données sont exagérées ou sous-estimées. L'axe des ordonnées tronqué en est un exemple. Prenons l'exemple de la figure 6.12a ci-dessous, qui compare le nombre de personnes bénéficiant de l'aide sociale à celui des personnes ayant un emploi à temps plein. Ce graphique déforme la proportionnalité, laissant entendre à première vue qu'il y a presque cinq fois plus de personnes bénéficiant de l'aide sociale que de personnes ayant un emploi à temps plein. Les chiffres indiqués sont exacts, mais lorsqu'ils sont présentés sur un axe des ordonnées complet qui commence à 0 dans la figure 6.12b, ils sont moins spectaculaires.

Une image contenant texte, capture d’écran, Police

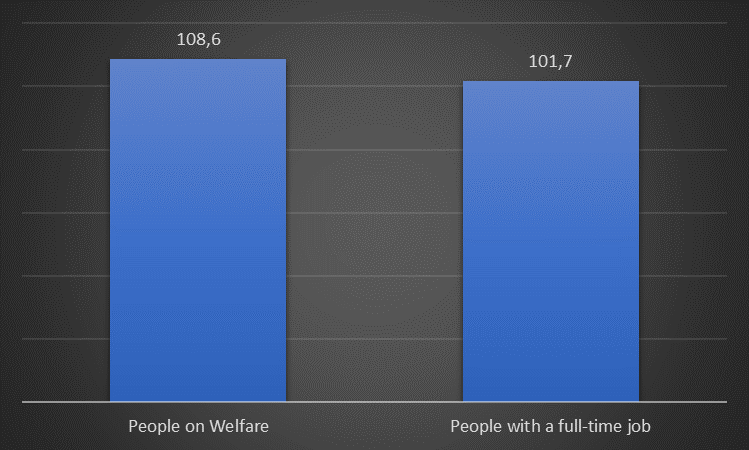
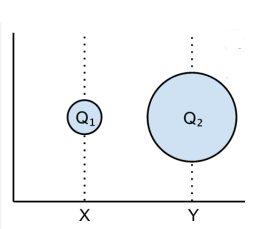
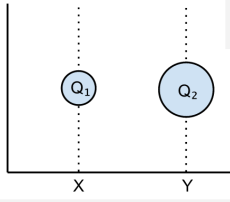
Description générée automatiquement 

Figure 6.12a et 6.12b. Exemples de visualisation trompeuse en utilisant l’axe des ordonnées tronqué. Source : https://www.mediamatters.org/fox-friends/dishonest-fox-chart-overstates-comparison-welfare-full-time-work-500-percent, consulté le 9 janvier 2024.

La comparaison des quantités dans les visualisations peut également être trompeuse en cartographiant les données dans une zone graphique. Au lieu de faire figurer la valeur des données dans la zone graphique (figure 6.13b), certaines visualisations trompeuses peuvent utiliser des variables autres que la surface, telles que le rayon (figure 6.13a). Cela fausse considérablement la représentation graphique des données.



**Quantité = rayon**



**Quantité = surface**

Figure 6.13a et 6.13b. Exemples de visualisation trompeuse montrant que la surface d’une visualisation pourrait être une distorsion de la quantité. Source : Pandey et al. (2015, p. 1472).

La deuxième catégorie proposée par Pandey et al. (2015) est l'inversion du message. Comme mentionné dans le principe 3 ci-dessus, les gens associent les directions à des tendances telles que l'augmentation à la hausse et la diminution à la baisse. Cette interprétation humaine fait de l'axe inversé l'une des techniques de distorsion les plus courantes susceptibles d'induire les individus en erreur. Comme le montre la figure 6.14 ci-dessous, le lecteur pourrait penser que les décès par arme à feu en Floride ont diminué (comme le suggère la tendance) à la suite de la mise en œuvre de la loi "stand your ground". Cependant, en regardant de plus près le graphique, on constate que l'axe des ordonnées est inversé et que les chiffres augmentent vers le bas.

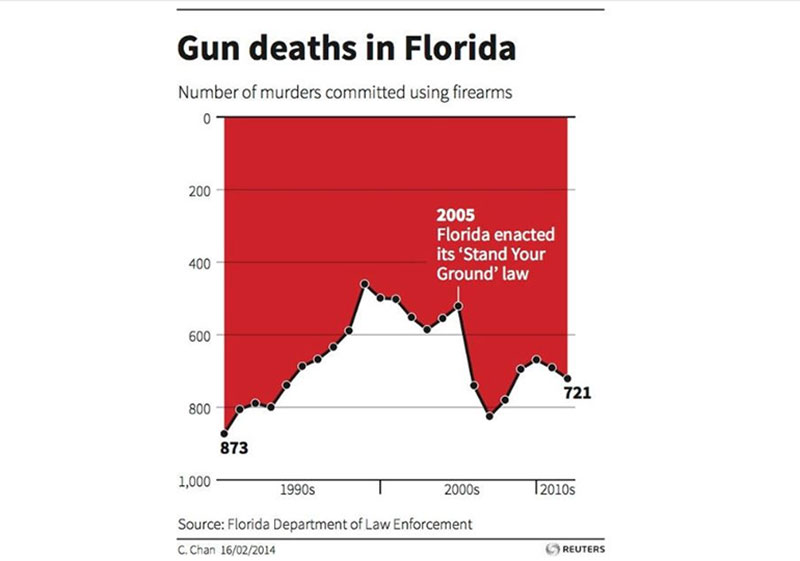


Figure 6.14. Exemple de visualisation trompeuse en utilisant l’inversion du message. Source : https://policyviz.com/2023/02/07/10-ways-to-mislead-with-data-visualization/, consulté le 9 janvier 2024.

En conclusion, bien que ces techniques soient les principales utilisées dans la visualisation trompeuse, il en existe beaucoup d'autres, moins courantes. Il incombe donc aux concepteurs des visualisations de respecter les principes et normes de conception des graphiques afin de transmettre des informations exactes. Néanmoins, les études en sciences cognitives suggèrent que tous les problèmes actuels d'interprétation des visualisations ne peuvent pas être résolus uniquement par une bonne conception visuelle. Ces études suggèrent également que les parties prenantes doivent être mieux informées sur les conventions et principes visuels, afin d'être de meilleurs consommateurs de visualisations de données.

## Bibliographie

Alhadad, S. S. (2018). Visualizing data to support judgement, inference, and decision making in learning analytics: Insights from cognitive psychology and visualization science. *Journal of Learning Analytics, 5*(2), 60-85.

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician, 27*(1), 17-21.

Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics, 19*(12), 2306-2315.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*(2), 75.

Chen, M., Ebert, D., Hagen, H., Laramee, R. S., van Liere, R., Ma, K. L., Ribarsky, W., Scheuermann, G. & Silver, D. (2009). Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications, 29*(1). 12-19.

Dudycz, H. (2010). Visualization methods in Business Intelligence systems–an overview. Business Informatics (16). *Data Mining and Business Intelligence, 104*, 9-24.

Eckerson, W. W. (2010). Performance dashboards: measuring, monitoring, and managing your business. John Wiley & Sons.

Few, S. (2007). Data visualization: past, present, and future. *IBM Cognos Innovation Center*.

Friendly, M., & Denis, D. (2005). The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences, 4*1(2), 103-130.

Friendly, M., Sigal, M., & Harnanansingh, D. (2017). The milestones project: a database for the history of data visualization. Dans *Visible Numbers* (pp. 219-234). Routledge.

Friendly, M., & Wainer, H. (2021). A history of data visualization and graphic communication. Cambridge, MA: Harvard University Press.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*(6), 236-243.

Hegarty, M. (2018). Advances in cognitive science and information visualization. Dans Score reporting research and applications, ed. D. Zapata-Rivera. Routledge.

Jaggia, S., Kelly, A., Lertwachara, K., & Chen, L. (2023). Business analytics: Communicating with numbers. 2e ed. McGraw Hill LLC.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review, 93*(5), 1449–1475.

Kosslyn, S. M. (2006). Graph design for the eye and mind. New York, NY: Oxford University Press

McCluskey, S. C. (1998). Astronomies and cultures in early medieval Europe. Cambridge University Press.

Nepomuceno, T. C. C., & Costa, A. P. C. S. (2019). Spatial visualization on patterns of disaggregate robberies. *Operational Research, 19*, 857-886.

Pandey, A. V., Rall, K., Satterthwaite, M. L., Nov, O., & Bertini, E. (2015, April). How deceptive are deceptive visualizations? An empirical analysis of common distortion techniques. *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 1469-1478.

Pinker, S. (1990). A theory of graph comprehension. Dans *Artificial intelligence and the future of testing*, R. Freedle (Ed.), 73–126. Hillsdale, NJ: Erlbaum.

Rogers, S. (2013, 15 mars). John Snow's data journalism: the cholera map that changed the world. *The Guardian*. https://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map

Schneider, E., & Dineen, C. (2013). Adding a dimension to Anscombe's quartet: Open source, 3‐D data visualization. *Proceedings of the American Society for Information Science and Technology*, *50*(1), 1-3.

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, *14*, 47-69.

Sharda, R., Delen, D., & Turban, E. (2018). Business intelligence, analytics, and data science: a managerial perspective. Pearson.

Stodder, D. (2013). Data visualization and discovery for better business decisions. *TDWI Research*.

Tufte, E.R. (2001). The visual display of quantitative information. 2e éd. Graphics Press

Wanderer, J. P., Nelson, S. E., Ehrenfeld, J. M., Monahan, S., & Park, S. (2016). Evolving health informatics semantic framework and metadata-driven architectures. *Journal of Medical Systems*, *40*(12), 1-9.

Zacks, J., Levy, E., Tversky, B., & Schiano, D. J. (1998). Reading bar graphs: Effects of extraneous depth cues and graphical context. *Journal of Experimental Psychology: Applied*, *4*(2), 119.

Zheng, J. G. (2017). Data visualization for business intelligence. Dans *Global business intelligence* (ed. J.M. Munoz). Routledge. 67-82.