

# Chapitre 1

## Visualisation interactive d'information

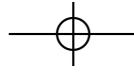
---

### 1.1 Introduction

La visualisation interactive d'information propose un cadre pour trouver des représentations graphiques à des données abstraites, qui permettent de les explorer et de les comprendre. Elle donne les clés pour trouver des représentations utilisant au mieux les capacités du système perceptuel humain et proposer des interactions utilisant au mieux les entrées de l'utilisateur.

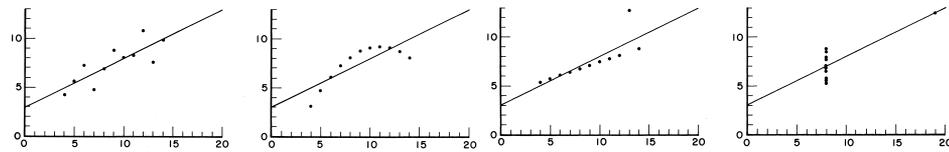
Dans le cadre de la science des données, la visualisation peut rendre des services à différents niveaux. Tout d'abord, elle permet d'explorer les données elles-mêmes, en amont de toute analyse par des moyens automatiques : c'est l'occasion par exemple de vérifier que des hypothèses posées sur les données (par exemple, la nature de leur distribution) sont effectives. Le statisticien [Anscombe \(1973\)](#) a ainsi illustré l'utilité de la visualisation en produisant quatre jeux de données indiscernables par leurs propriétés statistiques (moyennes, régressions linéaires, erreurs quadratiques, etc.)

Un simple affichage des nuages de points (et des régressions linéaires qu'ils déterminent) aide à comprendre que les phénomènes capturés par les données sont de natures différentes : lien linéaire mais bruité entre deux variables, lien quadratique mal capturé par la régression, liens linéaires très



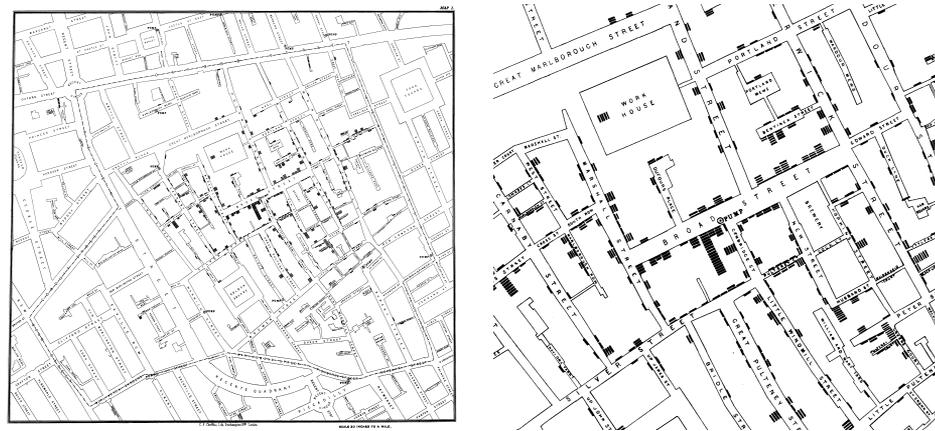
## 1CHANTIEREDUCATION VISUALISATION INTERACTIVE D'INFORMATION

forts mais parasités par un artefact (figure 1.1). Là où les statistiques ne permettent pas de discerner les jeux de données, un simple coup d'œil sur une visualisation élémentaire permet d'émettre des hypothèses sur la nature des liens entre les variables en présence.



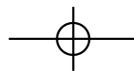
**Figure 1.1** - Le *quartet* d'Anscombe (1973) visualisé.

Cet aspect exploratoire de la visualisation d'information est illustré par la fameuse analyse de l'épidémie de choléra qui a frappé Londres en 1854. En figurant sur un fond de carte la localisation des 578 victimes de l'épidémie et des fontaines publiques (figure 1.2), le docteur J. Snow formula l'hypothèse que cette maladie se contractait du fait de la contamination d'un point d'eau et non d'une transmission par voie aérienne (Johnson 2006).

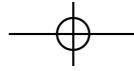


**Figure 1.2** - La carte (à gauche) de l'épidémie de choléra à Londres en 1854 par le docteur J. Snow. À droite est visualisé un détail centré sur la fontaine infectée.

La visualisation sert également à présenter des données afin de transmettre les informations découvertes suite à l'exploration pour convaincre de la réalité de phénomènes qui, enfermés dans des tableaux de chiffres,







## 1 CHAPITRE DONNÉES VISUALISER EN UNE VUE INTERACTIVE D'INFORMATION

vision du monde basée sur des faits, et non des a priori. La visualisation interactive d'information est un outil majeur pour parvenir à cette fin.

Ce chapitre propose donc de donner quelques clés pour choisir de bonnes représentations, que ce soit au cours de l'exploration ou pour exposer des résultats. Ben Shneiderman, un des fondateurs du domaine, a résumé les propriétés d'une bonne visualisation dans son célèbre *mantra* : « *overview first, zoom and filter, then details on demand* » (Shneiderman 1996) qu'on peut traduire par : « d'abord la vue d'ensemble ; ensuite, permettre de zoomer et de filtrer ; et enfin, fournir les détails à la demande ».

### 1.2 Des données au graphique

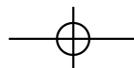
L'étape clé de la production d'une visualisation consiste à choisir l'encodage graphique (le *visual mapping* de Card et Mackinlay 1997) : quelle variable de l'image (la position, la couleur, etc.) va servir à encoder quel attribut des données (l'âge, la taille, le poids, etc. des individus listés dans notre jeu de données). Pour cela, il convient de connaître les différents types d'attributs qu'on peut avoir à traiter et ce qui les caractérise du point de vue des tâches attendues de comparaisons et d'agrégations ; mais également de connaître les variables de l'image qui sont disponibles pour les encoder et la manière dont elles sont perçues par l'être humain.

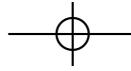
#### 1.2.1 Les données

Comme dans toute tâche d'analyse de données, celles-ci peuvent arriver sous diverses formes. Les problématiques de nettoyage et de mise en forme sont éludées ici et on considère que nos données sont formatées en tables. Nous appellerons *items* les lignes d'une telle table de données (d'autres termes se rencontrent, par exemple : *individu* ou *observation*). Chaque item est caractérisé par les valeurs de ses *attributs* (ou encore *dimensions*). Le nombre d'items et le nombre d'attributs les caractérisant dimensionnent un jeu de données et rendent le problème de la visualisation plus ou moins difficile.

#### Les natures des attributs

La nature de chaque attribut a des implications : des données dont les attributs sont des valeurs numériques (âge, taille) ne vont pas forcément





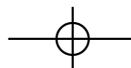
## CHAPITRE 1. VISUALISATION D'INTERACTIONS D'INFORMATIONS

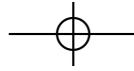
bénéficier d'une même visualisation que celles dont les attributs prennent leurs valeurs dans des catégories (espèce, modèle). Il importe donc de connaître les différentes natures d'attributs que l'on peut rencontrer. Card et Mackinlay (1997) les classent en trois grandes catégories, en réutilisant une classification classique des statistiques. Cette dernière est fonction de critères importants pour la visualisation comme pour les statistiques : quelle comparaison peut-on établir entre deux valeurs d'un attribut et quel résumé ou quelle agrégation peut-on faire d'un ensemble de valeurs d'un attribut.

Les attributs **nominaux** sont ceux qui ont le moins de structure. Ils prennent leurs valeurs parmi celles d'un ensemble (fini ou non), qui ont pour seule comparaison possible l'identité : deux individus sont-ils identiques ou différents selon cet attribut ? Un exemple d'attribut nominal est le nom d'une personne (nombre infini de valeurs possibles), ou son sexe (H/F). Il faut bien comprendre que la nature d'un attribut n'est pas forcément liée à son type numérique : des nombres peuvent être utilisés pour étiqueter des items sans avoir d'autre sens que l'identification. Par exemple, si on numérote des items, la seule comparaison entre l'item 2 et le 4 est qu'ils sont différents ; on ne peut conclure que l'un est plus petit que l'autre ou que le numéro 4 est deux fois plus grand que le 2.

Pour résumer un ensemble de valeurs d'un attribut nominal, on peut donner la distribution (combien d'items pour chaque valeur), les  $k$  éléments les plus fréquents, etc. Si l'ensemble des valeurs possibles est muni d'une structure (par exemple, un attribut nominal *ville* structuré hiérarchiquement par la géographie : *département, région, état*), cette dernière est utilisable pour grouper des valeurs avant de donner une distribution. Enfin, si l'on doit résumer un ensemble par une unique valeur, on ne peut que donner la plus fréquente, i.e. le *mode*.

Les attributs **ordonnés** sont plus riches que les nominaux : ils disposent d'une relation d'ordre. Pour deux valeurs, on peut donc tester l'identité (ou l'égalité), mais aussi, à partir de la relation d'ordre, laquelle est la plus petite. Une échelle d'évaluation de type Likert, « pas du tout d'accord », « pas d'accord », « indifférent », « d'accord », « tout à fait d'accord » est un exemple d'attribut ordonné. Les tailles de vêtement XS, S, M, L, XL en sont un autre : on peut dire que  $XS \neq L$  et que  $XS \prec L$ . On ne peut par contre rien affirmer sur l'écart entre ces deux tailles (aucun calcul n'est envisageable sur les valeurs : on ne dispose pas de structure algébrique).





## 1 CHAPITRE DONNÉES QUALITATIVES EN QUELQUE INTERACTIVE D'INFORMATION

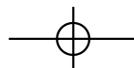
Pour résumer des valeurs ordonnées, il est possible, comme pour les attributs nominaux, de donner la distribution, les éléments les plus fréquents, etc. On peut en plus tirer parti de la relation d'ordre pour structurer l'ensemble des valeurs selon des classes (en choisissant des bornes et en regroupant les valeurs comprises entre deux bornes successives) et construire ainsi des histogrammes. Des indications sur la distribution des valeurs peuvent également être fournies grâce au minimum et au maximum, ainsi qu'à l'aide de quantiles. Enfin, si on doit résumer une distribution par une unique valeur, une alternative au mode, toujours disponible, est la *médiane* qui est la valeur centrale quand on place les valeurs dans l'ordre qui les structure.

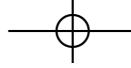
Enfin, les attributs **quantitatifs** ont encore plus de structure : on peut tester l'identité de deux valeurs, si une valeur est plus grande qu'une autre, mais surtout à quel point elle est plus grande. Pour les attributs quantitatifs, on parlera d'*échelle d'intervalle* si seule la différence entre deux valeurs fait sens. Par exemple, si un attribut est une date, on peut quantifier la différence entre deux dates par la durée qui les sépare. Par contre, le quotient de deux dates n'a pas de sens. Si le quotient de deux valeurs a un sens, on parlera d'*échelle de ratio* : c'est le cas des attributs pour lesquels la valeur nulle n'est pas arbitraire. Par exemple, si un attribut est une taille, on peut quantifier l'écart entre deux valeurs par leur différence (3 m est 1 m plus grand que 2 m), mais aussi par leur rapport (3 m est 1.5 fois plus grand que 2 m).

Pour résumer des valeurs quantitatives, on dispose des mêmes indicateurs que pour les valeurs ordonnées et d'autres sont calculables. En particulier, les moments statistiques résument la distribution : la *moyenne* en donne une valeur centrale, l'*écart-type* caractérise son étalement, etc.

### 1.2.2 L'image

La perception visuelle met en jeu des phénomènes complexes, dont les mécanismes sont étudiés depuis longtemps, mais pas pour autant encore tous compris. Nous présentons ici des modèles simplifiés, voire simplistes qui nous permettent, malgré leurs approximations, de raisonner sur les visualisations.





## CHAPITRE 1. VISUALISATION ET INTERACTIONS EN INFORAPHTIQUE

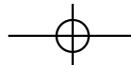
### Éléments sur la perception

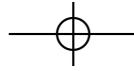
Le première chose dont il faut être conscient est qu'il y a une distinction à faire entre le *stimulus* visuel et sa *perception*. Le stimulus, pour nous une image caractérisée complètement par la couleur des pixels de l'écran, sera perçu après avoir traversé le système optique de l'œil, l'encodage par la rétine et l'interprétation du cerveau. Ce processus comble des manques de notre système visuel en ajoutant de l'information inférée à partir de diverses sources, notamment de notre expérience acquise. Ainsi, la tâche aveugle présente dans les champs visuels des deux yeux n'est pas perçue comme un trou dans la vision, pas plus que notre quasi-insensibilité à la couleur en vision périphérique induit une perception en noir et blanc. À plus haut niveau, des phénomènes d'invariance font que nous identifions des objets indépendamment du point de vue et de la distance.



**Figure 1.4** - Photo tirée du n° 58 de février 1958 de *Life Magazine* par Gregory (1970) pour illustrer l'émergence des formes selon la *Gestalt*. Ce motif est aussi présent dans le tableau *le Torero Hallucinogène* de Salvador Dalí (1970).

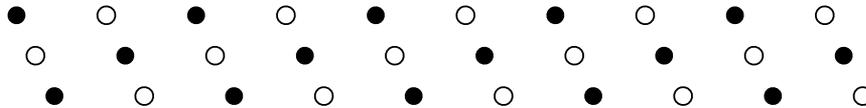
La théorie de la *Gestalt*, développée au début du XX<sup>e</sup> siècle, résume cette distinction entre stimulus et perception par la formule : « le tout est différent de la somme des parties ». Elle explique par exemple que notre perception des formes est *émergente* : notre cerveau injecte énormément d'information en plus du stimulus pour parvenir à une perception. La figure 1.4 illustre cet aspect : nous distinguons assez rapidement le sujet de la photo, un dalmatien de trois quarts arrière, malgré le caractère très épars de l'information. Encore plus étonnant : lorsqu'on demande où se situe cette scène, beaucoup de personnes répondent qu'elle a lieu dans un parc et, si on demande ce qui les conduit à cette affirmation, la présence d'un arbre est souvent évoquée. Et pourtant, cet arbre qui met sur la voie





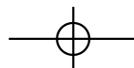
## 1 CHAÎNES DE DONNÉES VISUELLES EN QUÊRE D'INTERACTIVE D'INFORMATION

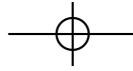
d'un parc se situe hors cadre ; seul son pied, et peut-être son ombre ou des feuilles tombées au sol, se devinent ! C'est d'ailleurs sûrement la présence du chien qui nous met sur la voie de cette interprétation. Le quart inférieur droit de l'image, avec le chien, ou le quart supérieur gauche, avec le pied de l'arbre, ne pourraient conduire, isolément, à l'interprétation du parc : cette perception est une construction. Outre l'émergence, la *Gestalt* liste des phénomènes qui affectent notre perception des formes : l'*invariance*, citée plus haut, la *réification*, qui nous fait percevoir des formes dans les interstices d'autres formes, ou enfin la *multistabilité*, qui exprime qu'un stimulus peut conduire à des perceptions multiples.



**Figure 1.5** - Des groupes émergent du fait de leur proximité ou de leur ressemblance.

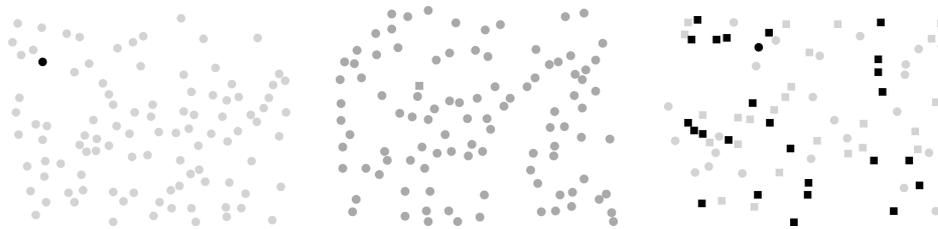
Au-delà de la perception des formes, la *Gestalt* explore ce qui conduit à la perception des groupes. Par exemple, dans l'image de la figure 1.5, les marques peuvent nous apparaître groupées trois par trois de deux manières différentes : soit des diagonales orientées vers le haut à droite comportant trois marques blanches ou trois marques noires, soit des diagonales orientées vers le haut à gauche comportant deux marques d'une couleur et une marque de l'autre. Ces regroupements sont induits par la similarité de certaines variables graphiques : la proximité des positions ou les couleurs identiques. Nous percevons donc spontanément des structures induites par le partage de certaines caractéristiques visuelles. Cette perception spontanée est appelée par les psychologues expérimentaux *perception pré-attentive* : la cognition n'entre pas en jeu et elle a donc lieu en temps constant quel que soit le nombre d'éléments dans le champs visuel (Treisman et Gelade 1980). Par exemple, trouver un carré (resp. une marque bleue) au milieu de cercles (resp. de marques rouges) sur la figure 1.6 ne dépend pas du nombre total de marques. Par contre, ces propriétés interfèrent : trouver le carré bleu au milieu de carrés rouges et des ronds bleus ou rouge prend un temps qui croît à nouveau avec le nombre de marques du fait de l'analyse consciente que nous devons faire des pro-





## CHAPITRE 1. VISUALISATION INTERACTIVES D'INFORMATIONS

priétés graphiques de chaque objet. Comme montré par [Vergheze \(2001\)](#), la réalité est plus complexe qu'une dichotomie entre traitement parallèle (pré-attentif) et séquentiel : il existe en fait d'un continuum entre ces deux extrêmes. Pour la visualisation d'information, rechercher au maximum les perceptions pré-attentives est une bonne chose : les détails encodés ainsi sauteront aux yeux plutôt que de nécessiter une recherche visuelle.



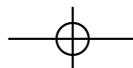
**Figure 1.6** - Trouver une marque foncée au milieu des claires (gauche) ou un carré au milieu de cercles (centre) demande un temps qui ne dépend pas du nombre de marques, alors que trouver le cercle foncé au milieu de cercles et de carrés clairs ou foncés (droite) prend un temps qui croît linéairement en fonction du nombre de marques.

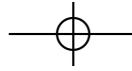
### Les variables de l'image

La visualisation consiste donc à encoder les attributs des données à l'aide de **marques** graphiques : points, lignes ou surfaces (si on ne considère pas les visualisations utilisant la 3<sup>e</sup> dimension de l'espace). Ces marques sont caractérisées par des **variables graphiques** : couleur, forme, etc. utilisables pour encoder les divers attributs des données.

[Bertin \(1967\)](#) a recensé de manière systématique ces variables graphiques exploitables pour différencier des marques : position, taille, valeur, texture, couleur, orientation et forme, définies ci-après. Bertin a caractérisé ces variables du point de vue des jugements visuels qu'elles autorisent :

- *Association* : est-ce que des marques partageant le même niveau de cette variable graphique nous apparaissent comme appartenant à un groupe ?
- *Sélection* : est-ce que deux marques ayant des niveaux différents nous apparaissent comme différentes (et si oui, combien de niveaux distincts peut-on percevoir) ?
- *Ordre* : peut-on ordonner des niveaux distincts ?





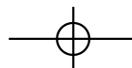
## 1 CHAPITRE DONNÉES QUALITATIVES EN VISUALISATION INTERACTIVE D'INFORMATION

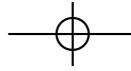
- *Quantité* : peut-on quantifier la différence entre deux niveaux distincts ?

La **position** est bidimensionnelle : elle permet d'encoder deux attributs. Ces deux dimensions du plan sont les variables de l'image les plus expressives : elles permettent l'association (des marques proches dans l'espace nous apparaissent constituant un groupe) et la sélection (une marque à distance d'un groupe apparaît comme une anomalie), de déterminer un ordre (de gauche à droite ou de haut en bas) et de quantifier la différence entre deux marques (une marque est deux fois plus loin de l'origine qu'une autre par exemple). La position est donc la variable graphique à exploiter en priorité pour construire une visualisation. Certaines visualisations utilisent également la troisième dimension de l'espace, la profondeur, comme une troisième variable graphique. La *Note sur l'utilisation de la 3D* explique pourquoi ce n'est pas forcément une bonne idée.

La **taille** des marques permet tous les jugements hormis l'association des points de même taille (au sein d'un nuage de points, ils ne nous sauteront pas aux yeux pour les isoler du reste). Par contre, [Bertin \(1967\)](#) nous indique que, si on utilise au plus 4 tailles différentes, la sélection fonctionne : nous percevrons ces 4 groupes. L'ordre, du plus petit au plus grand, est également perceptible sans effort. Enfin, on peut évaluer l'écart entre deux marques de tailles différentes. Il faut cependant prendre garde comme le rappellent [Cleveland et McGill \(1984\)](#) : si on sait estimer assez fidèlement le rapport entre deux longueurs, il n'en est pas de même pour les surfaces. En deux dimensions, on sous-estime les rapports de tailles : un cercle de rayon double d'un autre sera perçu comme ayant une surface moindre que quadruple du premier. Il faut donc veiller, si on souhaite que les jugements de quantité soient possibles, à ne faire varier la taille que selon une seule dimension à la fois. Après la position, la taille est tout de même la variable graphique la plus expressive.

La **valeur** des marques est leur niveau de gris, du clair au foncé, indépendamment de la teinte. Comme la taille, cette variable ne permet pas l'association, mais fonctionne pour la sélection si on utilise au plus 3 niveaux de gris pour les marques ponctuelles (4 pour les lignes et 5 pour les surfaces). L'ordre, du clair au foncé, est perçu sans effort. Par contre, les jugements sur la quantité ne sont pas possibles : on ne sait dire si un gris est deux fois plus clair qu'un autre. La valeur reste indiquée pour encoder tout type d'attribut à condition que seule l'information d'ordre soit importante à transmettre.





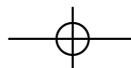
## CHAPITRE 1. VISUALISATION INTERACTIVES D'INFORMATIONS

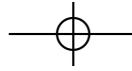
Les variables suivantes sont moins expressives et sont à considérer dans un second temps pour encoder des attributs moins importants, en veillant à ce que ces informations supplémentaires n'interfèrent pas avec les plus importantes.

Le **grain** est une variable graphique qui nécessite un remplissage des marques par un motif périodique (par exemple, hachures, carreaux, trame). Le grain est alors l'échelle de ce motif (par exemple, l'espacement entre les hachures ou la taille des carreaux). Toujours selon Bertin (1967), le grain autorise les mêmes jugements que la valeur, mais également l'association. Il faut cependant noter que, pour la perception visuelle, le grain interfère avec les variables de taille et de valeur. Comment, par exemple, interpréter des rayures de largeurs identiques sur des formes identiques mais dont la taille a doublé ? De même, des rayures plus ou moins denses peuvent être perçues comme plus ou moins foncées, interférant ainsi avec la variable de valeur. Le grain est donc une variable graphique à manier avec précaution.

La **couleur** de Bertin est en fait ce que les peintres appellent la teinte : c'est la couleur, mais indépendamment de sa valeur. Le terme n'a donc pas le sens qu'on lui donne dans le langage courant. Ainsi, le rose et le rouge sont des variantes d'une seule couleur : le rouge. De même, un bleu clair, un bleu foncé, ou un bleu nuit ont pour couleur le bleu. La couleur permet l'association et la sélection, dans la mesure d'un maximum de 7 niveaux distincts. Il n'y a par contre pas d'ordre directement perceptible sur les couleurs : même si le vert se situe entre le rouge et le violet dans l'arc-en-ciel, on ne peut pas dire que le vert est plus violet que ne l'est le rouge. L'arc-en-ciel n'est pas une échelle que chacun a en tête permettant de définir un ordre sur les couleurs. Sans ordre, il est enfin impossible de prononcer un jugement de quantité. La couleur doit donc être réservée pour encoder des catégories. On peut tout de même envisager d'autres usages, mais avec beaucoup de précautions (voir la *Note sur l'utilisation de la couleur* ci-après).

L'**orientation** est l'angle que fait une marque avec l'horizontale. Pour utiliser cette variable, il faut que la forme s'y prête et ne présente pas trop de symétries (avec des marques circulaires ou des points, par exemple, cette variable n'est pas du tout utilisable). Comme la couleur, elle permet cependant, lorsque son usage est possible, de faire des jugements d'asso-





## 1 CHAÎNES DONNÉES VISUALISABLES EN QUELQUE INTERACTIVE D'INFORMATION

ciation et de sélection, mais avec un nombre de classes plus restreint : 4 niveaux sont un maximum à utiliser.

Enfin, la **forme** est la variable graphique la moins expressive : elle n'autorise que le jugement d'association. Cependant, un choix judicieux des formes (par exemple, rond, carré, triangle) permet d'encoder efficacement un attribut nominal quand les autres variables graphiques sont déjà toutes utilisées.

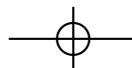
### Note sur l'utilisation de la couleur

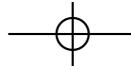
Il est possible d'utiliser la couleur de manière à percevoir un ordre moyennant quelques précautions et en se restreignant à une partie du spectre : par exemple un dégradé qui va du rouge au vert (en passant par le jaune) ou du jaune au bleu (en passant par le vert). Sur cette dernière échelle, on sait par exemple dire d'un vert s'il est plus jaune ou plus bleu qu'un autre vert.

Il faut cependant être très prudent avec cette utilisation : le daltonisme le plus fréquent (8% des hommes en France) altère la perception des couleurs du rouge au vert, ce qui devrait amener à proscrire l'usage de ce type de dégradé. Pour le dégradé du jaune au bleu, l'usage de la couleur interagit avec l'utilisation de la valeur : le jaune est plus clair que le bleu ! Il est donc difficile d'utiliser la couleur et la valeur comme deux variables ordonnées pour encoder deux attributs distincts des données.

On peut par contre encoder de manière redondante un unique attribut en utilisant les deux variables couleur et valeur. C'est par exemple ce que propose l'échelle de couleur *viridis* proposée par la bibliothèque *matplotlib*<sup>4</sup> : elle fait simultanément varier la couleur du jaune au bleu et la valeur du clair au foncé. *Cubehelix* (Green 2011) propose une échelle de couleurs « arc-en-ciel » où l'ordre peut être perçu grâce à la valeur (la couleur ne renforçant pas cette information puisque l'arc-en-ciel n'est pas un ordre perçu sur les couleurs). Avec ces deux échelles, la couleur peut être supprimée sans perte d'information puisqu'elle est redondante dans le premier cas et inutile dans le second. Sa présence est essentiellement esthétique. Il faut donc en réserver l'usage au cas où elle ne peut pas être mieux employée, c'est-à-dire si on n'a pas d'information de catégorie à encoder.

4. Choosing Colormaps – Matplotlib documentation, <[matplotlib.org/users/colormaps.html](http://matplotlib.org/users/colormaps.html)>.





## CHAPITRE 1. VISUALISATION DES DONNÉES GÉOGRAPHIQUES

Dans tous les cas, le choix des couleurs doit faire l'objet d'une attention particulière si on souhaite maximiser leur pouvoir discriminant. Plutôt que se fier à ses goûts personnels, toujours discutables, on se reposera pour ses choix sur l'expérience acquise au fil du temps, par exemple par les cartographes, que [Brewer \(2018\)](#) met à la disposition de tout un chacun dans un outil en ligne remarquable : *ColorBrewer*<sup>5</sup>.

### Note sur l'utilisation de la 3D

On remarque que, dans la liste des variables graphiques, la troisième dimension n'est pas mentionnée, alors que l'outil informatique permettrait de l'exploiter, soit directement à l'aide de dispositifs d'affichage stéréographiques, soit par projection sur des écrans classiques. Ajouter une variable d'espace est extrêmement tentant : on a vu que les dimensions de la position sont les variables les plus expressives. Pourquoi alors se priver d'une troisième variable aussi bonne ?

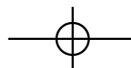
En fait, l'usage de la troisième dimension présente des inconvénients qui en limitent l'utilité à des cas particuliers (par exemple, lorsque les données sont intrinsèquement tridimensionnelles). Le premier problème vient de l'interaction de la troisième dimension avec la variable de taille. Notre perception de la profondeur repose en partie sur la perspective, qui nous sert à évaluer la distance d'objets connus en fonction de leur taille apparente. La perception de ces deux variables interfère et les utiliser simultanément pour encoder deux attributs conduit à un mélange d'information que notre système perceptif n'est pas capable de séparer.

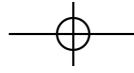
Pourquoi alors préférer la taille ? Encore une fois, notre perception de la troisième dimension passe par des projections (sur l'écran, sur nos rétines) et ces dernières impliquent de l'occultation : plusieurs marques risquent de se superposer et se masquer. La variable de taille n'a pas ce défaut, d'où le choix de l'utiliser en priorité si on n'a pas d'argument fort pour la troisième dimension.

### 1.2.3 Encodage visuel

Pour créer une visualisation, il reste à établir quels attributs des données vont être encodés par quelles variables visuelles. Faire les bons choix dans

5. *Color advice for cartography*, <[colorbrewer.org](http://colorbrewer.org)>.





## 1 CHAPITRE DONNÉES QUALITATIVES EN VISUALISATION INTERACTIVE D'INFORMATION

cette combinatoire n'est pas évident, mais on peut se référer à des recommandations faites par Bertin (1967) (en se basant sur une réflexion sur sa propre expertise) ou Mackinlay (1986) (en se basant sur des résultats de psychologie expérimentale).

La stratégie la plus simple est d'allouer les variables graphiques, peu nombreuses, en ordre d'expressivité décroissante. Si les données n'ont pas plus d'attributs que le nombre de variables, les attributs quantitatifs sont à encoder en priorité à l'aide de la position puis de la taille, etc. Les attributs ordonnés bénéficieront des variables libres suivantes. Enfin, les attributs nominaux seront encodés en utilisant les variables restantes.

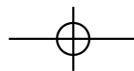
Ces règles simples conduisent déjà à concevoir des visualisations efficaces, et surtout évitent de commettre des impairs.

### **Exemple d'encodage : *BubbleChart***

La visualisation principale de Gapminder (figure 1.7) utilise un *BubbleChart*. Elle montre le lien entre espérance de vie et revenu par habitant pour tous les pays de la planète. Chaque pays est représenté par une marque (cercle) dont les variables graphiques encodent différents attributs :

- la position sur l'axe horizontal encode le revenu par habitant ;
- la position sur l'axe vertical encode l'espérance de vie ;
- la taille encode le nombre d'habitants ; et
- la couleur encode la zone géographique.

Trois des quatre attributs représentés sont quantitatifs : le revenu, l'espérance de vie et le nombre d'habitants. Hans Rosling a donc logiquement choisi les trois variables les plus expressives pour les représenter : les deux composantes de la position et la taille. Dans ce trio, la nature seule des attributs ne permet pas de décider de l'allocation des variables. Le choix se fait donc en fonction de l'intention de la visualisation. Ici, la corrélation entre revenu et espérance de vie voulait être mise en avant ; ces deux attributs sont donc encodés avec la variable la plus expressive : la position. La distribution des cercles dans l'espace nous montre que, plus les revenus sont bas, plus l'espérance de vie est faible. La taille des pays, encodée par la taille des marques, ne semble pas avoir d'influence sur ces deux autres attributs.



CHAPITRE 1. VISUALISATION INTERACTIVES D'INFORMATIONS

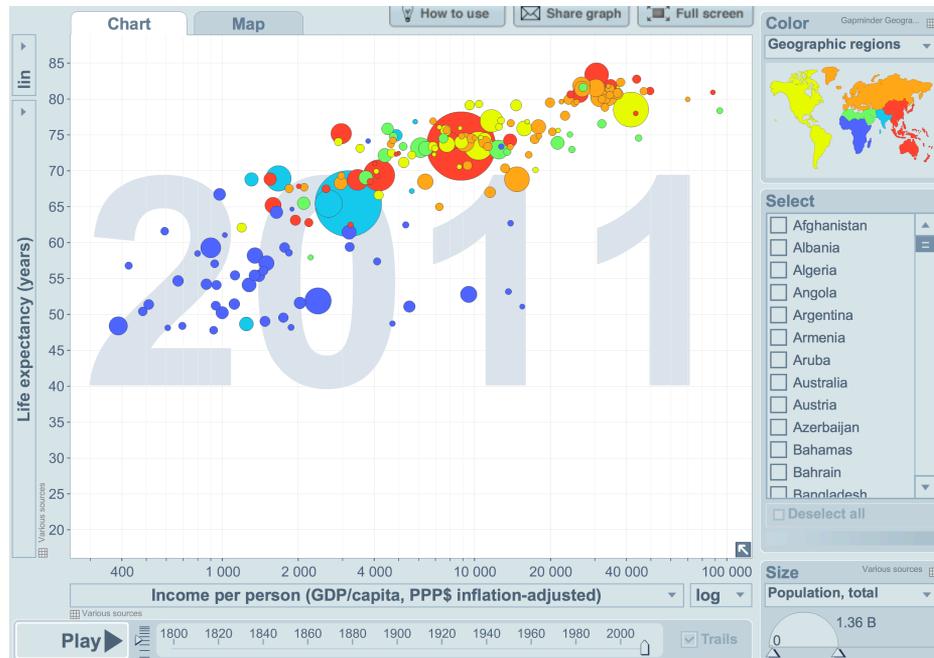


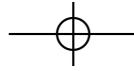
Figure 1.7 - Capture de l'interface de Gapminder.

L'attribut de région géographique (découpage en 6 zones montrées sur la carte en haut à droite, figure 1.7) est encodé par la couleur<sup>6</sup>. Ce choix se justifie car les variables de valeur ou de grain ne permettent pas d'encoder 6 niveaux sans risques de confusion, et surtout par le fait qu'elles risquent d'induire un ordre alors que les zones géographiques ne sont pas ordonnées, puisqu'il s'agit d'un attribut nominal.

L'association des variables de position et de couleur met des groupes en évidence : l'Afrique noire (en bleu foncé) se situe nettement à l'extrémité inférieure gauche du nuage de points. Viennent ensuite, avec des localisations moins marquées, le sous-continent indien (en bleu clair), l'Afrique sub-saharienne et l'Asie mineure (vert), l'Asie et l'Océanie (rouge), les Amériques (jaune) et l'Eurasie (orange).

La sélection fait ressortir les anomalies : par exemple, la position de certains marques les fait sortir de la tendance générale. Nous avons des pays

6. Pour une version en couleur la figure, se référer à <[iihm.imag.fr/blanch/infovis/](http://iihm.imag.fr/blanch/infovis/)>.



## 1 CHAPITRE DONNÉES VISUALISER EN QUÊRE ACTIVE D'INFORMATION

avec une espérance de vie faible (en bas) qui ont pourtant un revenu élevé (plus à droite que ceux à espérance de vie comparable). Si on regarde dans le détail, il s'agit de pays africains riches en ressources naturelles (pétrole, etc.) mais dont la population ne profite manifestement pas de cette richesse.

### **Exemple d'encodage : la campagne de Russie**

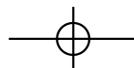
La visualisation des pertes de l'armée française lors de la campagne de Russie par C. J. Minard (figure 1.3) encode l'effectif et le trajet de l'armée, mais aussi la température à diverses étapes de ce trajet. Ces deux informations sont données par deux marques (lignes) juxtaposées sur deux graphes. Celui du haut nous communique la trajectoire de l'armée en encodant ses attributs par :

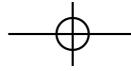
- la position sur l'axe horizontal pour la longitude ;
- la position sur l'axe vertical pour la latitude ;
- la taille de la ligne (son épaisseur) pour l'effectif de l'armée ; et
- la couleur la ligne pour encoder le sens du trajet (aller en beige, retour en noir).

La visualisation du bas ajoute les informations temporelles et de température pour la partie retour du trajet. Elle partage l'axe horizontal avec la première visualisation, ce qui relie les données entre elles : on peut retrouver la date et la température des grandes étapes de la retraite en suivant les lignes verticales qui appartiennent les points des deux graphiques. Cette seconde visualisation réutilise la position sur l'axe vertical pour encoder la température. La date n'est pas encodée visuellement : elle est indiquée à l'aide de texte. Le texte est également utilisé pour donner les valeurs précises de l'effectif de l'armée et pour la température. La visualisation de C. J. Minard respecte donc ainsi, avec plus d'un siècle d'avance, les préconisations de B. Shneiderman : le graphique fournit une vue d'ensemble de cette campagne de Russie et le texte, par la lecture, le plus précis niveau de détail disponible sur les données.

### **Pour aller plus loin**

[Card et Mackinlay \(1997\)](#) ont proposé une notation synthétique pour l'encodage visuel. Ils utilisent une table à double entrée présentant en lignes





## CHAPITRE 1. VISUALISATION INTERACTIVE D'ENCODAGES PERMANENTS

les attributs des données et en colonnes les variables graphiques. Cette notation sert à caractériser des visualisations existantes, mais c'est également un outil pour explorer l'espace de conception en envisageant toutes les combinaisons possibles d'encodage sans se limiter à des idées préconçues. On lira donc utilement cet article classique, dans lequel ils présentent la notation et montrent son utilité sur des cas concrets de visualisations.

### 1.3 Encodages avancés

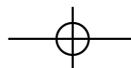
On a vu avec l'encodage de la campagne de Russie que la correspondance « une variable graphique pour un attribut des données » n'est pas toujours respectée (l'axe vertical encode à la fois la latitude et la température dans les deux parties de la visualisation). Dans la pratique, on a souvent besoin de réutiliser des variables graphiques ou de les utiliser de manière plus avancée, notamment pour montrer des relations entre individus présentés par la visualisation. Dans cette partie, nous montrons à l'aide d'exemples que les principes généraux vus précédemment ne sont qu'une base sur laquelle il faut construire pour s'adapter aux particularités de chaque jeu de données et des questions que l'on se pose sur ces données.

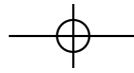
#### 1.3.1 Utilisation multiple des variables graphiques

Dans la visualisation de la campagne de Russie, l'axe vertical est utilisé deux fois. Le réemploi d'une variable graphique est souvent rendu nécessaire par le fait que les données sont susceptibles d'être décrites par plus d'attributs que nous n'avons de variables graphiques à notre disposition. Inversement, on a parfois intérêt à encoder un attribut à l'aide de plusieurs variables graphiques.

##### Réutilisation des variables graphiques

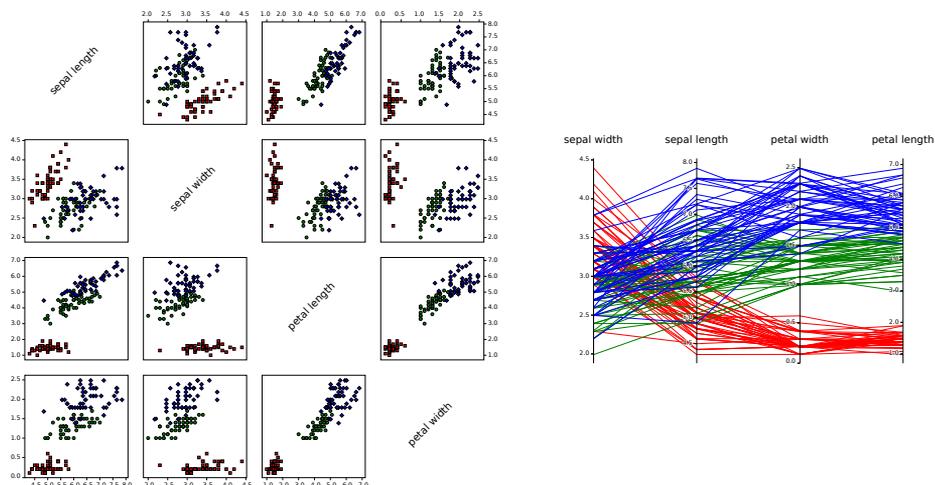
Si un jeu de données comporte de nombreux attributs quantitatifs, d'égale importance a priori, on aimerait pouvoir utiliser les variables de position, les plus expressives, pour tous ces attributs. Or, on ne dispose que de deux variables de position. On peut cependant construire pour chaque couple d'attributs un nuage de points. La figure 1.8 montre à gauche de tels nuages de points pour un jeu de données à quatre dimensions quanti-





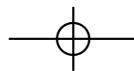
## 1 CHANGEMENTS DANS LA VISUALISATION INTERACTIVE D'INFORMATION

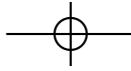
tatives (les longueurs et largeurs des pétales et sépales de diverses espèces d'iris). Les nuages de points sont organisés en table comportant une ligne et une colonne pour chaque attribut. Tous les couples sont donc représentés et les axes horizontaux et verticaux sont réutilisés autant de fois qu'il y a d'attributs. Les relations (corrélation ou autre) entre attributs apparaissent visuellement du fait des groupements ou des alignements qui sont immédiatement perçus dans certains nuages de points. Un individu est représenté par une marque dans chacun des nuages de points et seule une version interactive de cette visualisation permet de faire le lien entre les points : si l'utilisateur désigne un individu dans une des vues, il peut être mis en évidence dans les autres par le programme.



**Figure 1.8** - Le jeu de données *Iris*, à quatre dimensions quantitatives, représenté par une matrice de nuages de points (gauche) et des coordonnées parallèles (droite).

La figure 1.8 montre à droite une autre possibilité pour représenter le même jeu de données. Il s'agit cette fois d'utiliser autant d'axes verticaux que d'attributs quantitatifs. Cette fois, les individus ne sont plus représentés par des points, mais par des lignes qui relient les différentes coordonnées le long de ces axes. Cette visualisation, appelée *coordonnées parallèles* (Inselberg 1985), permet avec un peu d'habitude de repérer visuellement grâce aux motifs que forment les lignes, par exemples horizontaux (entre les deux dernières colonnes) ou en forme de croix (entre





## CHAPITRE 1. VISUALISATION INTERACTIVE EN CODAGE PERMANENT

les deux premières colonnes), les corrélations positives ou négatives entre attributs successifs.

Pour comprendre des jeux de données plus complexes avec ce type de visualisations, l'interaction est primordiale. Le fait de pouvoir sélectionner un sous-ensemble des individus sur un axe et d'observer comment se répartit ce sous-ensemble sur les autres axes, ou encore de pouvoir réordonner les axes pour juger de la corrélation entre deux attributs particuliers, sont des actions qu'on ne peut conduire avec une simple image. Pour réaliser l'importance des techniques d'interaction associées aux visualisations, on pourra utilement consulter les vidéos de présentation des techniques de visualisation *ScatterDice*<sup>7</sup> (Elmqvist *et al.* 2008) et *ImAxes*<sup>8</sup> (Cordeil *et al.* 2017) qui enrichissent ces deux visualisations pour autoriser l'exploration de jeux de données bien plus complexes.

### Composition des variables graphiques

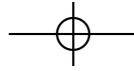
Dans les exemples précédents, la réutilisation d'une variable graphique concerne exclusivement les variables de position, et ce n'est pas un hasard : cela permet la juxtaposition de visualisations. Cette dernière a cependant un inconvénient : l'espace alloué à chaque sous-partie est réduit. Pour faire passer le plus possible d'information au travers d'une surface réduite, on peut alors envisager d'utiliser conjointement plusieurs variables graphiques pour représenter un seul attribut des données.

C'est ce que proposent les *Horizon Graphs* (Reijner 2008), qui généralisent les échelles bicolores de Saito *et al.* (2005). La figure 1.9 donne un exemple de cette technique de visualisation grâce à laquelle on représente un grand nombre de séries temporelles dans un espace limité. Dans cet exemple, qui présente l'évolution du taux de chômage dans divers états des États-Unis d'Amérique, l'axe horizontal encode le temps et l'axe vertical est réutilisé pour encoder l'écart à la moyenne du taux de chômage pour chacun des états. Si seule la position sur l'axe vertical était utilisée, les lignes seraient difficiles à comparer et les variations seraient difficiles à percevoir car comprimées verticalement. Les *Horizon Graphs* proposent donc d'« enrouler » l'axe vertical sur lui-même en encodant l'écart du taux de

7. *Multidimensional visual exploration using scatterplot matrix navigation*, [<youtu.be/2bYIRcO-gwg>](https://youtu.be/2bYIRcO-gwg).

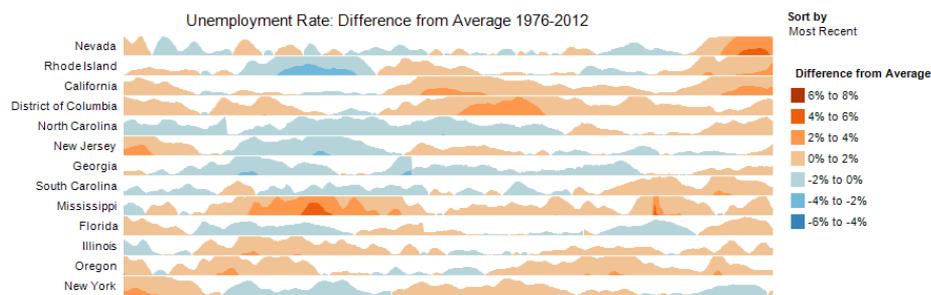
8. *Immersive multivariate data visualisation*, [<youtu.be/hxqJJ934Reg>](https://youtu.be/hxqJJ934Reg).





## 1.3.1 L'ENCODAGE DES VALEURS EN SAISON INTERACTIVE D'INFORMATION

chômage modulo une constante (ici 2%). La valeur (du plus clair au plus foncé) est utilisée pour encoder le nombre d'enroulements et on peut ainsi juger au premier coup d'œil à quel point deux états diffèrent. Enfin, la couleur<sup>9</sup> (teintes de rouges ou de bleus) encode si la variation est positive ou négative. L'écart à la moyenne du taux de chômage est donc séparé en trois parties (signe, quotient et reste de la valeur absolue divisée par une constante), communiquées via trois variables graphiques (couleur, valeur et hauteur).



**Figure 1.9** - Évolution des écarts à la moyenne des taux de chômage de divers états des États-Unis d'Amérique visualisée grâce à des *Horizon Graphs*.

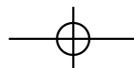
### 1.3.2 Encodage des liens entre individus

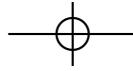
Beaucoup de jeux de données comportent des informations de relation entre individus. Si pour chaque individu on dispose d'un lien vers un parent, on parle de hiérarchies (par exemple une taxonomie des espèces animales), alors que si plusieurs liens sont possibles, on parle de réseaux<sup>10</sup>. Parfois, les liens eux-mêmes peuvent comporter des attributs (nature et longueur d'un axe routier reliant deux villes, etc.)

Si nous sommes habitués aux représentations graphiques des hiérarchies et des réseaux sous forme d'arbres et de graphes « nœuds-liens » dans lesquels les individus sont représentés par des marques ponctuelles et les liens par des segments les reliant, des représentations alternatives existent qui peuvent se montrer très efficaces.

9. Pour une version en couleur la figure, se référer à <[iihm.imag.fr/blanch/infovis/](http://iihm.imag.fr/blanch/infovis/)>.

10. On préfère le terme « réseau » à celui de « graphe » car ce dernier, bien que plus usuel en informatique, a le défaut d'être polysémique dans notre contexte car il désigne également une représentation graphique.

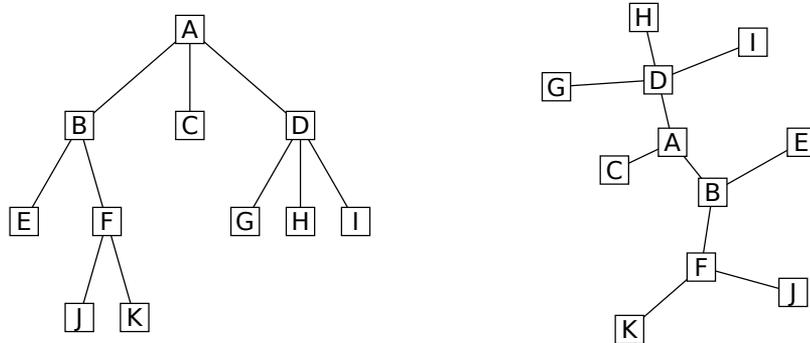




## CHAPITRE 1. VISUALISATION INTERACTIVE D'ENCODAGES MANOÏÉS

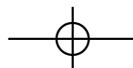
### Hiérarchies

Les représentations usuelles des hiérarchies sont de type « nœuds-liens ». Du fait que ces représentations soient très courantes, on oublie souvent la diversité des algorithmes qui existent pour placer les nœuds de ces arbres dans le plan. La figure 1.10 donne, pour un même arbre, deux exemples de placements de ses nœuds. Les divers algorithmes qui existent offrent des compromis différents entre place nécessaire et lisibilité; et ils s'adaptent à des types d'arbres particuliers (étendus en largeur ou en profondeur, équilibrés ou non, etc.) Un examen systématique de ces placements et des métriques quantifiant leurs propriétés est proposé par [McGuffin et Robert \(2010\)](#).



**Figure 1.10** - Deux représentations « nœuds-liens » d'un même arbre avec différents algorithmes de placement.

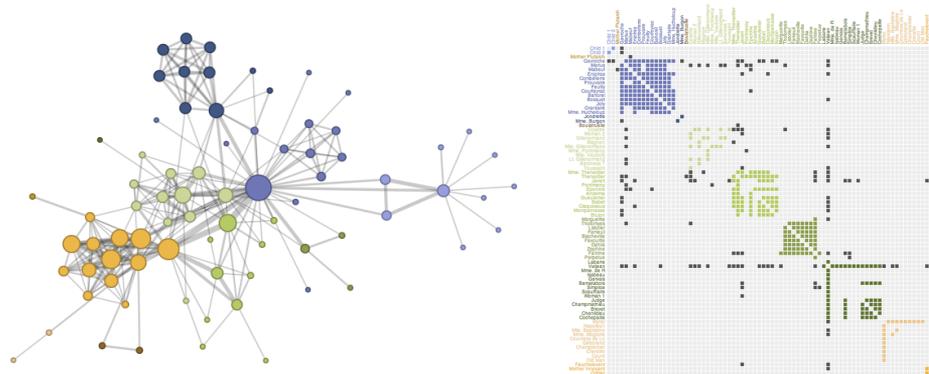
On connaît moins les représentations d'arbres qui utilisent l'emboîtement pour figurer la hiérarchie. Cette idée, déjà présente chez [Bertin \(1967\)](#), a été popularisée par [Shneiderman \(1992\)](#) sous le nom de *Tree-maps*. Elle consiste à utiliser toute la surface disponible pour figurer la racine de l'arbre, puis à partager cette surface entre les enfants de la racine et de procéder ainsi récursivement jusqu'à atteindre les feuilles de l'arbre. De nombreuses variantes de cette idée existent et proposent des stratégies de partition offrant là encore des compromis variés en termes de lisibilité ou de stabilité du résultat si on modifie partiellement la structure de l'arbre. Si les *Tree-maps* originels utilisent des rectangles, d'innombrables variations existent qui adoptent d'autres géométries de nœuds : des cercles ou encore





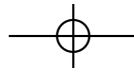
## CHAPITRE 1. VISUALISATION INTERACTIVE D'UN CODAGE PERMANENT

d'occultation : les individus sont représentés par les lignes et les colonnes de la matrice et les liens deviennent des marques ponctuelles, présentes ou non, à l'intersection de ces lignes et colonnes. La figure 1.12 présente à droite un réseau visualisé sous cette forme. Des motifs nous apparaissent alors, qu'on peut apprendre à interpréter en termes de théorie des graphes : les carrés remplis le long de la diagonale sont des cliques (des ensembles d'individus complètement interconnectés entre eux), les motifs en forme de croix sont des individus centraux (plus connectés que les autres et assurant la connexité du réseau). Pour que ces motifs apparaissent, l'ordre dans lequel les individus sont placés en lignes et colonnes est bien sûr primordial. Il s'agit encore d'une problématique de placement, mais cette fois en une dimension. De nombreux algorithmes aux propriétés variées existent et leur différents mérites pour la visualisation ont été étudiés par [Behrisch et al. \(2016\)](#).



**Figure 1.12** - Un réseau visualisé sous forme de diagramme « nœuds-liens » (gauche) et d'une matrice d'adjacence (droite).

Les représentations « nœuds-liens » et par matrice d'adjacence ne sont pas à opposer : elles ont chacune leurs points forts et leurs points faibles comme montré par [Ghoniem et al. \(2004\)](#). Sur une visualisation « nœuds-liens », on peut par exemple trouver plus facilement un chemin entre deux individus que sur une matrice d'adjacence. Par contre, dès que le réseau est grand et que sa connectivité est importante, la matrice d'adjacence est plus adaptée à la plupart des autres tâches qu'on peut rencontrer dans l'exploration d'un graphe.



## CHAPITRE ALLIANCE INTERACTIVE D'INFORMATION

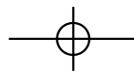
Pour tirer parti des points forts de ces deux types de visualisation, il est possible de les combiner en une visualisation hybride, en encodant les sous-parties denses d'un réseau à l'aide de matrices et la structure moins dense par des liens qui connectent ces parties denses. C'est ce que fait la technique *NodeTrix* (Henry *et al.* 2007), qui enrichit encore l'ensemble en proposant des techniques d'interaction pour que les utilisateurs structurent eux-mêmes cette visualisation, comme le montre la vidéo de démonstration<sup>12</sup>. Ce type de visualisation hybride existe aussi pour les arbres avec les *Elastic hierarchies* de Zhao *et al.* (2005) ou pour des mélanges d'arbre et de réseau avec les *Dendrogramix* de Blanch *et al.* (2015).

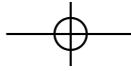
### 1.4 Pour aller plus loin

Ce chapitre n'a donné qu'une brève introduction à la visualisation interactive d'information. Certains aspects, bien que très importants, n'ont été qu'à peine effleurés. En premier lieu, l'importance de l'interaction est délicate à illustrer dans un ouvrage imprimé. Les techniques de visualisation présentées ici ne produisent pas simplement une image à partir d'un jeu de données. Elles sont souvent accompagnées d'un ensemble de techniques d'interaction qui permettent, au travers de la visualisation, de manipuler et d'explorer les données. Cette manipulation est capitale pour s'approprier et comprendre les données en leur donnant corps. Plus la manipulation sera directe, plus elles sortiront du domaine abstrait des tables de chiffres pour devenir des entités concrètes accessibles à notre compréhension. Depuis une dizaine d'années, les publications présentant une nouvelle technique de visualisation sont pour la plupart accompagnées d'une vidéo de démonstration et, parfois, du code pour la tester. Il ne faut donc pas hésiter à visionner ou à essayer ces démonstrations.

Un autre aspect passé sous silence ici est la question de l'évaluation : comment valide-t-on l'intérêt d'une nouvelle technique de visualisation ? S'il existe des critères familiers aux informaticiens (complexité des algorithmes mis en œuvre, optimisation d'un critère quantifiable comme le nombre de croisements des liens dans le placement d'un réseau, etc.), l'appréciation des utilisateurs est également à prendre en compte. On peut pour cela mener des expérimentations en se donnant des métriques (temps pour at-

<sup>12</sup>. *NodeTrix: a hybrid visualization of social networks*, [<youtu.be/7G3MxyOCHKQ>](https://youtu.be/7G3MxyOCHKQ).





## CHAPITRE 1. VISUALISATION INTERACTIVE D'INFORMATIONS

teindre un objectif, préférences des utilisateurs, etc.) et comparer ainsi designs alternatifs ou nouvelles techniques à l'état de l'art. Sur ces questions, on pourra se référer utilement au chapitre 4 du livre de [Munzner \(2014\)](#). Cet ouvrage dans son ensemble aidera par ailleurs le lecteur intéressé à approfondir les différents thèmes que nous n'avons pu qu'effleurer dans cette introduction au vaste domaine de la visualisation interactive d'information.

Enfin, avant d'inventer ses propres visualisations, ou pour rechercher de l'inspiration, on consultera utilement l'état de l'art. La littérature scientifique dans le domaine est foisonnante, mais heureusement, depuis quelques années, des articles recensant les techniques pour des types de données spécifiques voient le jour régulièrement et sont généralement accompagnés d'un site web offrant un accès aisé et structuré à cet état de l'art. Par exemple, si on s'intéresse aux visualisations de hiérarchies, on consultera le pionnier dans ce type de catalogue : <[treevis.net](#)>, site compagnon d'un état de l'art extensif par [Schulz \(2011\)](#) qui recense plus de 300 techniques de visualisation d'arbres. Dans la même veine, le site <[survey.timeviz.net](#)> est quant à lui associé au livre de [Aigner et al. \(2011\)](#) qui recense plus de 100 manières de représenter des séries temporelles. Une liste à jour de ce type d'états de l'art est maintenue par l'auteur sur le site dédié à ce chapitre : <[iihm.imag.fr/blanch/infos/](#)>.

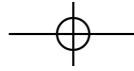
### 1.5 Exercices

#### Exercice 10.1

- Donner un exemple d'attribut pour chacun des types de données suivant : nominal, ordinal, quantitatif (intervalle et ratio).
- Expliquer, pour chacun de ces types de données, quelles sont les opérations de comparaison réalisables entre deux valeurs.
- Expliquer, pour chacun de ces types de données, quelles sont les opérations d'agrégation applicables pour résumer tout un ensemble de valeurs.

#### Exercice 10.2



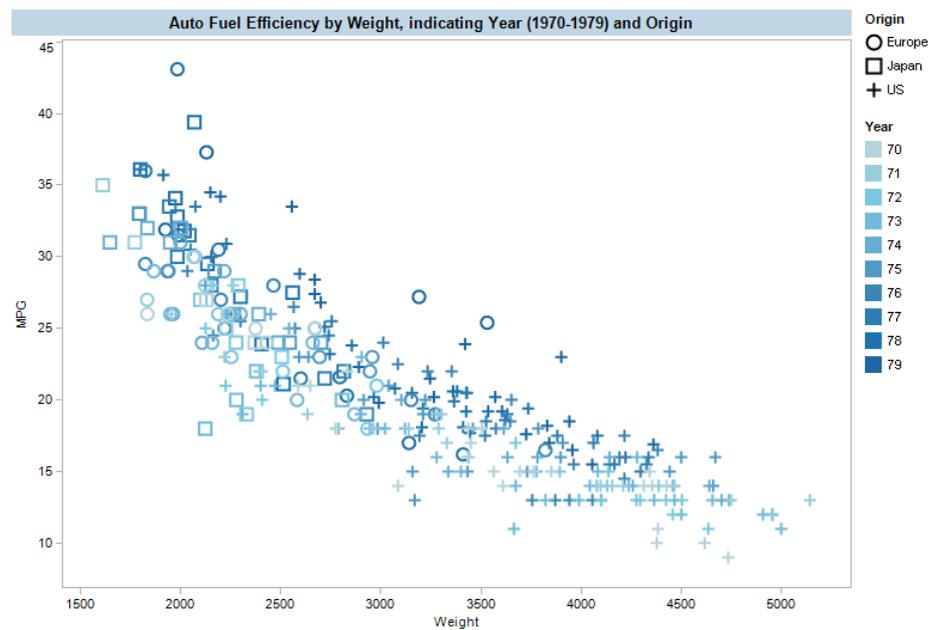


## 10 CHAPITRE 10 VISUALISATION INTERACTIVE D'INFORMATION

- Pour chaque variable graphique de Bertin (position, taille, forme, valeur du foncé au clair, couleur, orientation et texture), indiquer le type de jugement que l'on peut porter.
- Expliquez, pour chacune de ces variables, quels sont les types de données qu'elles encodent convenablement.

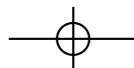
### Exercice 10.3

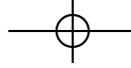
- En utilisant la taxonomie de [Card et Mackinlay \(1997\)](#), caractériser l'encodage graphique utilisé par la visualisation de la figure 1.13, qui présente un jeu de données dont chaque ligne caractérise une voiture par son autonomie (MPG pour *miles per gallon*, l'inverse de notre litre aux 100 km), son poids, son année de sortie et la région où elle est fabriquée (Europe, Japon, États-Unis d'Amérique).



**Figure 1.13** - Visualisation d'un jeu de données de voitures.

Expliquer en quoi ce choix est pertinent en vous référant aux recommandations de Bertin.





## CHAPITRE 1. VISUALISATION INTERACTIVE D'INFORMATIONS

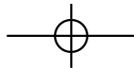
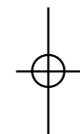
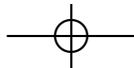
- b) En supposant qu'une question posée sur ce jeu de données soit de comprendre comment les consommations des véhicules ont évolué dans le temps, quelle représentation choisir ?  
Décrire cette représentation à l'aide de la taxonomie proposées par [Card et Mackinlay \(1997\)](#).  
Justifier cet encodage à la lumière des recommandations de Bertin.
- c) Pour chaque interaction disponible autour de la visualisation de Gapminder (figure 1.7), indiquer quelle étape du processus de transformation des données en visualisation décrit par [Chi et Riedl \(1998\)](#) est manipulée.

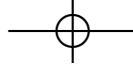
### Travaux pratiques

Le site [iihm.imag.fr/blanch/infovis/](http://iihm.imag.fr/blanch/infovis/) pointe vers un ensemble de ressources pour s'initier à l'usage de la bibliothèque D3.js<sup>13</sup>. Il propose des sujets de travaux pratiques pour s'entraîner à l'utiliser.

---

13. *D3.js Data-driven documents*, [d3js.org](http://d3js.org).





# Bibliographie

---

Wolfgang AIGNER, Silvia MIKSCH, Heidrun SCHUMANN et Christian TOMINSKI : *Visualization of Time-Oriented Data*. Springer, 2011.

Raymond J. ANDREWS : An interactive timeline of the most iconic infographics, 2017. URL <http://infowetrust.com/scroll/>.

Francis J. ANSCOMBE : Graphs in statistical analysis. *American Statistician*, 27(1) :17–21, 1973.

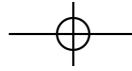
Michael BEHRISCH, Benjamin BACH, Nathalie HENRY RICHE, Tobias SCHRECK et Jean-Daniel FEKETE : Matrix reordering methods for table and network visualization. *Computer Graphics Forum*, 35 :24, 2016.

Jacques BERTIN : *Sémiologie graphique*. Mouton/Gauthier-Villars, 1967.

Renaud BLANCH, Rémy DAUTRICHE et Gilles BISSON : Dendrogramix : a hybrid tree-matrix visualization technique to support interactive exploration of dendrograms. In *Proc. PacificVis 2015*, pages 31–38, 2015.

Renaud BLANCH et Éric LECOLINET : Browsing zoomable treemaps : Structure-aware multi-scale navigation techniques. *IEEE Trans. on Vis. and Comp. Graph. (Proc. InfoVis 2007)*, 13(6) :1248–1253, 2007. URL <http://iihm.imag.fr/blanch/projects/ztm/>.

Cynthia BREWER : ColorBrewer : color advice for cartography, 2018. URL <http://www.ColorBrewer.org>.



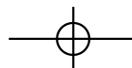

---

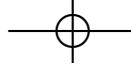
 BIBLIOGRAPHIE

---

 BIBLIOGRAPHIE

- Stuart K. CARD et Jock MACKINLAY : The structure of the information visualization design space. *In proc. InfoVis'97*, pages 92–99, 1997.
- Ed Huai-hsin CHI et John T. RIEDL : An operator interaction framework for visualization systems. *In proc. InfoVis'98*, pages 63–70, 1998.
- William S. CLEVELAND et Robert MCGILL : Graphical perception : Theory, experimentation, and application to the development of graphical methods. *J. Am. Stat. Assoc.*, 79(387) :531–554, 1984.
- Maxime CORDEIL, Andrew CUNNINGHAM, Tim DWYER, Bruce H. THOMAS et Kim MARRIOTT : Imaxes : Immersive axes as embodied affordances for interactive multivariate data visualisation. *In Proc. ACM UIST*, pages 71–83, 2017.
- Niklas ELMQVIST, Pierre DRAGICEVIC et Jean-Daniel FEKETE : Rolling the dice : Multidimensional visual exploration using scatterplot matrix navigation. *IEEE TVCG (Proc. InfoVis 2008)*, 14(6) :1141–1148, 2008.
- Mohammad GHONIEM, Jean-Daniel FEKETE et Philippe CASTAGLIOLA : A comparison of the readability of graphs using node-link and matrix-based representations. *In Proc. IEEE InfoVis 2004*, pages 17–24, 2004.
- Dave A. GREEN : A colour scheme for the display of astronomical intensity images. *Bull. Astr. Soc. India*, 39 :289–295, 2011.
- Richard L. GREGORY : *The Intelligent Eye*. Weidenfeld and Nicolson, 1970.
- Nathalie HENRY, Jean-Daniel FEKETE et Michael J. MCGUFFIN : Node-Trix : Hybrid representation for analyzing social networks. *IEEE Trans. on Vis. and Comp. Graph. (Proc. InfoVis 2007)*, 13(6) :1302–9, 2007.
- Alfred INSELBERG : The plane with parallel coordinates. *The Visual Computer*, 1(2) :69–91, 1985.
- Steven JOHNSON : *The Ghost Map : the Story of London's Most Terrifying Epidemic, and How it Changed Science, Cities and the Modern World*. Riverhead, 2006.



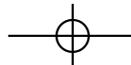


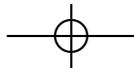
---

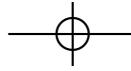
*BIBLIOGRAPHIE**BIBLIOGRAPHIE*

---

- Jock Mackinlay MACKINLAY : Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, 5(2) :110–141, 1986.
- Michael J. MCGUFFIN et Jean-Marc ROBERT : Quantifying the space-efficiency of 2D graphical representations of trees. *Information Visualization (IVS)*, 9(2) :115–140, 2010.
- Tamara MUNZNER : *Visualization Analysis and Design*. AK Peters Visualization Series, 2014.
- Hannes REIJNER : The development of the horizon graph. *In proc. Vis08 Workshop From Theory to Practice : Design, Vision and Visualization*, 2008.
- Takafumi SAITO, Hiroko Nakamura MIYAMURA, Mitsuyoshi YAMAMOTO, Hiroki SAITO, Yuka HOSHIYA et Takumi KASEDA : Two-tone pseudo coloring : Compact visualization for one-dimensional data. *In proc. InfoVis'05*, pages 173–180, 2005.
- Hans-Jörg SCHULZ : Treevis.net : a tree visualization reference. *IEEE Computer Graphics and Applications*, 31(6) :11–15, 2011.
- Ben SHNEIDERMAN : Tree visualization with tree-maps : 2-d space-filling approach. *ACM Trans. Graph.*, 11(1) :92–99, 1992.
- Ben SHNEIDERMAN : The eyes have it : A task by data type taxonomy for information visualizations. *In proc. Visual Languages*, pages 336–343, 1996.
- Anne TREISMAN et Garry GELADE : A feature-integration theory of attention. *Cog. Psycho.*, 12 :97–136, 1980.
- Preeti VERGHESE : Visual search and attention : a signal detection theory approach. *Neuron*, 31 :523–535, 2001.
- Shengdong ZHAO, Michael J. MCGUFFIN et Mark H. CHIGNELL : Elastic hierarchies : Combining treemaps and node-link diagrams. *In Proc. IEEE InfoVis 2005*, pages 57–64, October 2005.







# Index

---

attributs des données  
agrégation, 5  
attribut des données, 4  
comparaison, 5  
nominaux, 5, 14  
ordonnés, 5, 14  
quantitatifs, 6, 14  
échelle d'intervalle, 6  
échelle de ratio, 6

perception  
Gestalt, 7  
groupes, 8  
invariance, 7  
multistabilité, 8  
perception, 7  
pré-attentive, 8  
réification, 8  
stimulus, 7  
émergence, 8

variable graphique  
couleur, 11, 15  
daltonisme, 12  
dégradé, 12  
encodage graphique, 4  
forme, 11, 12  
grain, 11, 15  
marque, 9  
orientation, 11  
position, 10, 14  
taille, 10, 13, 14

troisième dimension, 13  
valeur, 10, 15  
variable de l'image, 4  
échelle de couleur, 12  
visualisation interactive d'information,  
1

