

Savoir compter, savoir conter

Épisode n° 3 – Des graphiques et des tableaux

 Corinne Benveniste*

Les tableaux et graphiques¹ relèvent d'un autre niveau de lecture que le texte. Ils peuvent par exemple susciter la curiosité du lecteur et lui donner envie d'entrer dans le texte pour en savoir plus. De manière générale, un tableau, un graphique ne sont pas mis dans l'article seulement pour faire joli, mais pour répondre à des questions que le lecteur se pose. De plus, comme les intertitres, ils aèrent le texte et rendent la mise en page plus attractive.

La logique qui régit le texte s'applique aussi aux tableaux et graphiques : offrir l'essentiel, visible d'emblée. Ceci veut dire que l'auteur a fait un choix, afin de ne pas laisser le lecteur désarmé devant une multiplicité de chiffres dans un tableau ou une multitude de courbes dans un graphique : autrement dit, l'auteur a retravaillé ces chiffres et ces courbes pour les besoins de la publication.

L'illustration est autonome

Tableaux et graphiques sont des illustrations, des compléments ou des « zooms » sur un point précis. Dans tous les cas, une figure est autonome : on doit pouvoir la comprendre sans avoir recours au texte de l'article. Elle a un titre, et ce dernier donne une information. Elle ne présente aucune ambiguïté (d'unité, d'échelle, de date, etc.), quitte à ce qu'elle soit assortie d'une note de lecture.

Tableau ou graphique, comment choisir ?

Dans une publication, mieux vaut privilégier les graphiques. Certes, les chiffres précis sont attendus sur certains thèmes : ainsi la population de la France, le nombre de chômeurs par sexe, âge ou qualification, le PIB

(produit intérieur brut), le commerce extérieur sont des chiffres repères, souvent déclinés en plusieurs rubriques dans un tableau. Néanmoins l'utilisateur de chiffres peut, de plus en plus fréquemment, puiser de l'information gratuitement sur un site internet, dans des cédéroms ou des ouvrages spécialisés. Dans un article, le tableau doit seulement situer les chiffres les plus significatifs et les plus neufs issus de l'étude. Les graphiques donnent une image plus globale, mais dont le mouvement et les niveaux se saisissent d'un seul coup d'œil. Ainsi, on voit mal un tableau de chiffres – même lisible – restituer la richesse d'information de deux pyramides des âges superposées à deux dates, ou encore de trois courbes mensuelles comparant des prix sur une longue période.

Le tableau n'est pas exhaustif

Les statisticiens travaillent sur des tableaux. C'est leur matière première. Dans un article, l'auteur ne peut restituer toute l'information qu'il a traitée ; il choisit. De même, dans un tableau, il choisit de mettre en valeur des résultats. Ceci implique d'avoir d'abord réduit le tableau aux chiffres essentiels à donner : un lecteur ne va pas mémoriser un tableau entier à x lignes, y colonnes et z critères ! Le format d'une publication est limité, la taille du tableau doit l'être aussi. C'est donc à l'auteur de trier les informations les plus inédites, les plus inattendues ou les plus proches des préoccupations de ses lecteurs. Le chiffre important peut être une valeur, une répartition en pourcentage ou une évolution annuelle, tout dépend du thème.

Commencer par l'essentiel

Valable pour le corps de l'article, cette règle l'est aussi pour un tableau. Ainsi, dans une comparaison temporelle, il est préférable que les années les plus récentes figurent dans la partie gauche du tableau, de manière à ce qu'elles apparaissent en premier à la lecture, alors que dans le tableau d'origine (le tableau de travail), les années sont souvent par ordre chronologique. L'année la plus récente peut aussi être mise en gras pour attirer l'attention.

Dans le cas d'une variable qualitative non ordonnée, le tableau peut s'affranchir de l'ordre artificiel des modalités de cette variable figurant dans la nomenclature d'origine et trier autrement les données en les classant par exemple selon un ordre croissant ou décroissant des effectifs par modalité² : c'est ainsi que lors des résultats du recensement, on publie la population des villes ou des départements par ordre d'importance, et non par ordre numérique ou alphabétique de ceux-ci.

Les « têtes » d'un tableau (titre d'une ou de plusieurs colonnes) doivent être simples, concises et lisibles. Les tableaux ont toujours un

* Corinne Benveniste est responsable éditoriale au sein de la Division rédaction des publications du Département de l'offre éditoriale de l'Insee. Elle est plus particulièrement responsable du domaine économique dans l'équipe éditoriale des Insee Première.

1. Cet article n'aborde pas le cas des graphiques dits « d'analyse », comme le *box plot* (également appelé « boîte à moustaches »). Voir à ce sujet le *Courrier des statistiques* n° 90 de juin 1999 et son dossier spécial sur l'analyse exploratoire des données (articles de Dominique Ladiray, Monique Le Guen et Sophie Destandau).

2. Ce point, fondamental, est abordé à trois reprises dans cet article, sous trois angles différents. Le lecteur n'en sera que mieux averti !

titre, une unité, une source, parfois une note de lecture.

Les graphiques

Un graphique présente l'avantage, par rapport à un texte ou à un tableau, de permettre une lecture globale, des comparaisons visuelles très rapides. Il est moins précis qu'un tableau, mais il attire l'œil et focalise sur un point précis, un zoom ou un mouvement dans le temps. Certes, un graphique peut se lire point par point, comme un tableau. Mais ce n'est pas son principal intérêt. Il permet surtout la perception directe par l'œil d'une information. Par exemple, une courbe d'évolution sur 40 ans du prix du pétrole, en euros et en dollars, vaut mieux que la série de chiffres qui a servi à la fabriquer.

Un graphique doit comporter un titre « énonciatif » (par exemple : « Les grandes surfaces en 2003 ») qui permet de savoir avec suffisamment de précision de quoi il s'agit. Ce titre, intégré au graphique, est éventuellement complété par un titre « informatif » (par exemple : « Les maxidiscomptes ont encore progressé ») ou par un petit commentaire, qui met en exergue l'information principale à retenir. Parfois, le titre « informatif » est suffisant. Le graphique doit comporter aussi une source, une unité et, éventuellement, une légende (qui reprend les couleurs ou valeurs du graphique avec leur intitulé).

3. On appelle variable *quantitative* une variable dont les modalités sont des nombres, fruits d'une mesure. On distingue les variables quantitatives *continues*, qui peuvent prendre une infinité (ou un très grand nombre) de valeurs – comme l'âge ou le salaire d'un individu –, et les variables quantitatives *discrètes*, ou *discontinues*, qui ne peuvent prendre qu'un nombre limité de valeurs (généralement entières) – comme le nombre d'enfants d'un ménage ou le nombre de pièces d'un logement.

4. On appelle variable *qualitative* une variable dont les modalités sont des mots. On distingue les variables qualitatives *ordonnées*, dont les modalités ont un ordre logique (comme les réponses « tout à fait d'accord », « plutôt d'accord », « plutôt pas d'accord », « pas du tout d'accord »), et les variables qualitatives *non ordonnées*, dont les modalités n'ont pas d'ordre logique (comme la profession, la couleur des yeux, le pays de naissance, etc.)

5. Dans tout ce qui suit, lorsque la source d'un graphique n'est pas précisée, c'est qu'il a été inventé spécialement pour illustrer cet article à titre pédagogique.

Un graphique est imprimé

L'auteur ne doit jamais perdre de vue que son graphique sera imprimé. Avant de se lancer dans de magnifiques représentations en couleur sur écran, il doit vérifier le nombre de couleurs disponibles dans la publication : noir, noir plus une couleur, ou quadrichromie. Dans ce dernier cas, l'auteur doit se souvenir que si le graphique présente un nombre trop important de couleurs, les lecteurs ayant la particularité d'être daltoniens ne pourront pas distinguer toutes les nuances. De manière générale, l'auteur doit penser au résultat imprimé sur papier. La limitation au noir et blanc oblige à s'en tenir à des graphiques simples (nombre réduit de courbes ou de trames). Mais simple ne signifie pas simpliste... L'auteur doit aussi penser à la taille réelle qu'aura son graphique une fois imprimé, taille qui ne sera jamais une pleine page d'écran.

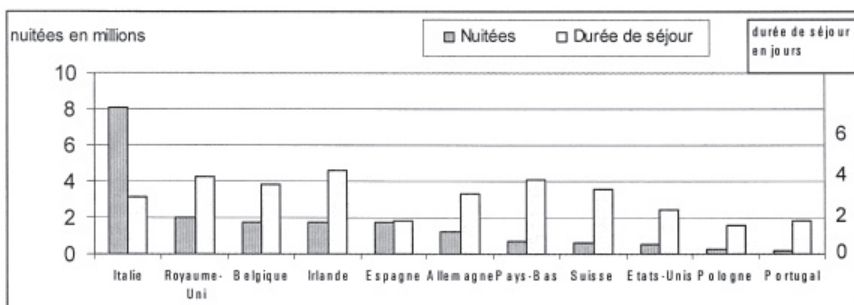
Des formules consacrées

Les lecteurs sont habitués à lire certaines formes de représentations graphiques. Mieux vaut les surprendre par le contenu (ce dont parle le graphique) que par le contenant

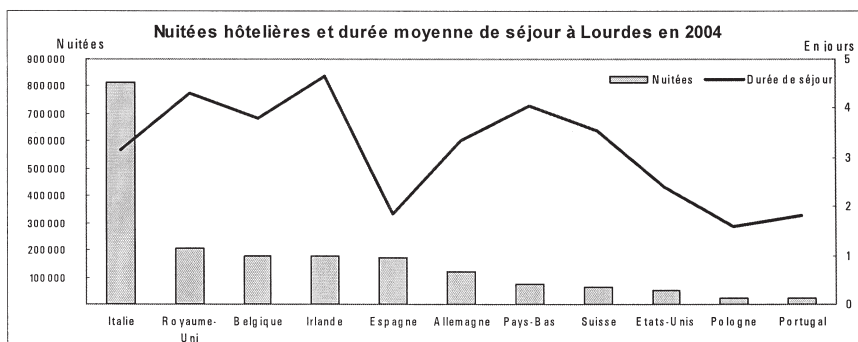
(la présentation du graphique). Certains types de graphiques sont des classiques : il en va ainsi, de la pyramide des âges déjà évoquée. Elle obéit à de bonnes pratiques de représentations visuelles et statistiques. Les courbes sont réservées aux variables quantitatives continues, et les bâtons aux variables quantitatives discrètes³ ou bien qualitatives⁴. La limite n'est pas toujours évidente, mais autant il n'est pas choquant de comparer des années présentées en bâtons, autant il est faux de mettre sur une courbe continue des grandeurs manifestement discontinues (graphiques 1 et 2, respectivement comme exemples à suivre et ne pas suivre⁵).

On voit bien (graphique 2) qu'on ne peut relier entre elles des durées de séjours de personnes provenant de différents pays, car la courbe ainsi obtenue donne une impression fautive de continuité. C'est donc la présentation des données sous forme de bâtons (graphique 1) qu'il faut privilégier. Hélas le tableur Excel ne permet pas de faire directement une représentation à deux échelles avec des bâtons (la double échelle du graphique 2 étant nécessaire,

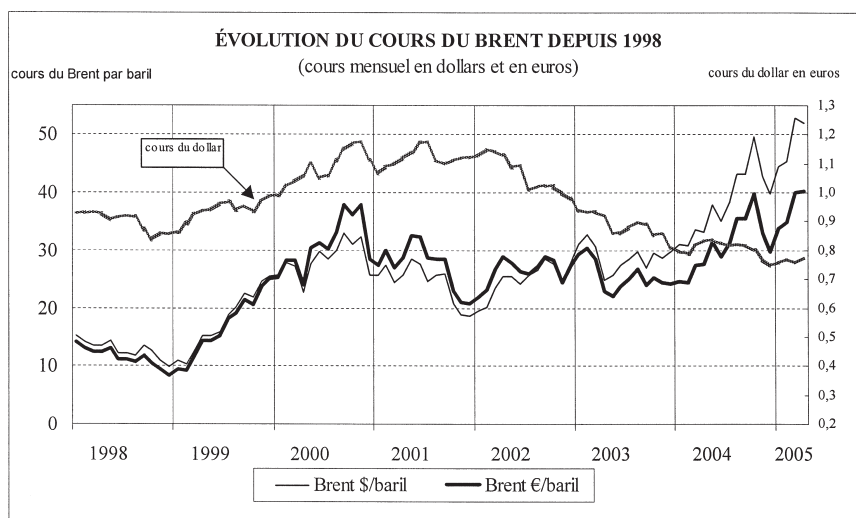
Graphique 1 : Représentation « satisfaisante » des nuitées et durées de séjour



Graphique 2 : Représentation « fautive » des nuitées et durées de séjour



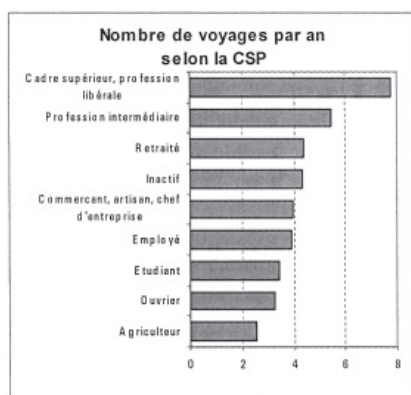
Graphique 3 : exemple d'utilisation des courbes



Source : Insee Première n° 1024.

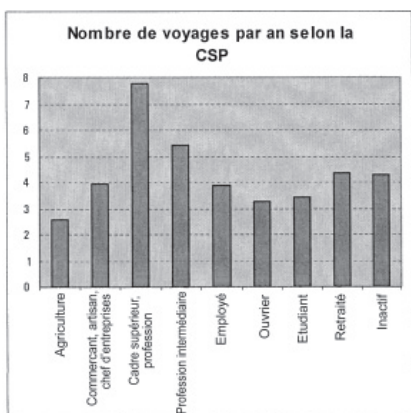
Graphiques 4 : exemple d'utilisation du diagramme en bâtons horizontaux

Ce graphique répond notamment à la question : quelle profession part le plus souvent en voyage ?



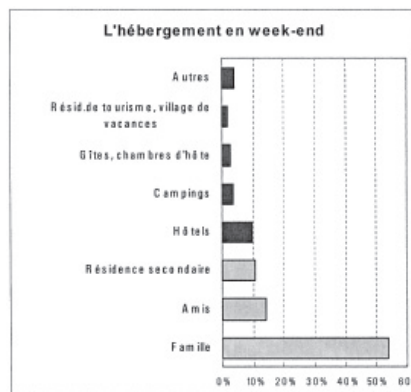
Graphiques 5 : exemple d'utilisation du diagramme en bâtons verticaux

Celui-ci ne répond à aucune question, et oblige à se tordre le cou !



Source : Le tourisme en France.

Graphique 6 : Deux composantes sur un même graphique



puisque les nuitées sont exprimées en millions alors que les durées de séjour le sont en jour). Dans ce cas, à moins d'utiliser des astuces qui s'apparentent à du « bricolage » et qu'il serait trop long d'expliquer ici, mieux vaut demander de l'aide aux maquetistes.

Les courbes, pour représenter le temps

Les axes x et y permettent de représenter des composantes « longues » où les points sont très nombreux. Ils permettent aussi une représentation « ordonnée ». Certaines composantes sont ordonnées par nature, par exemple le temps ou les âges, qu'il n'est pas question de bousculer.

La courbe de référence gagne à être renforcée, ici le prix du baril

de pétrole en euros puisque nous sommes en Europe (graphique 3).

À éviter : trop de courbes sur le même graphique, surtout lorsqu'elles se superposent ou se croisent.

Pour les variables discrètes ou qualitatives, utiliser les bâtons

C'est la représentation graphique la plus courante. Chaque modalité du caractère est représentée par un bâton dont la hauteur est proportionnelle à l'effectif. C'est le graphique le plus approprié pour représenter un caractère qualitatif ou un caractère quantitatif discret.

On l'a dit précédemment, les nomenclatures statistiques ne sont pas sacrées. Les composantes qui ne sont pas ordonnées par nature (cas des variables qualitatives non ordonnées) sont « ordonnables » de la modalité ayant la plus petite valeur à celle ayant la plus grande, ou inversement. Cela permet de guider l'œil pour une lecture directe et rapide de l'information. Pour les intitulés longs, il est plus judicieux de choisir les barres horizontales (graphiques 4 et 5).

Les deux couleurs ou valeurs⁶ (graphique 6) permettent de distinguer l'hébergement gratuit de l'hébergement payant. Dans le texte correspondant, il suffit alors à l'auteur de préciser que l'hébergement gratuit est très majoritaire.

L'apposition de bâtons « origine 0 vers la droite et vers la gauche », autrement dit le fait de mettre sur un même graphique des bâtons horizontaux partant vers la gauche et des bâtons partant vers la droite, est utile pour représenter les pyramides des âges, ou encore les oppositions entre exportations et importations

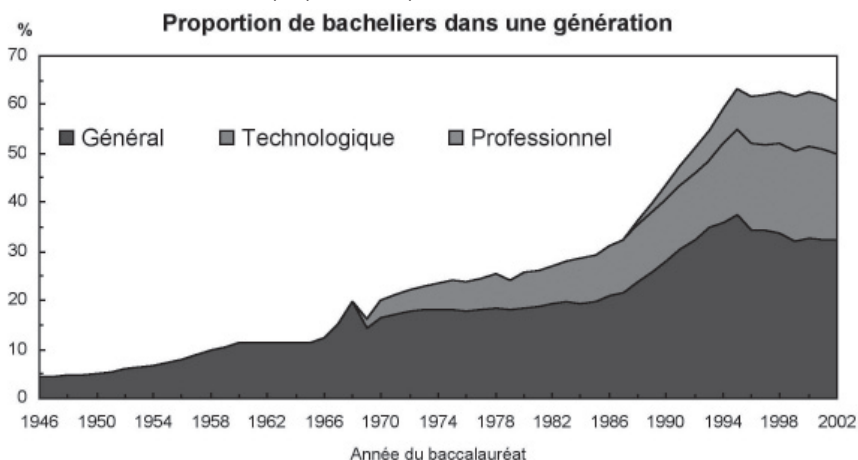
6. On appelle *valeur* le degré de clarté d'une couleur, du sombre au clair. Valeur et couleur sont des caractéristiques souvent indépendantes l'une de l'autre : une couleur (le bleu par exemple) peut avoir une valeur faible ou forte selon qu'il s'agit d'un bleu foncé ou clair ; à l'inverse, deux couleurs différentes (le rouge et le bleu par exemple) peuvent avoir la même valeur si elles sont aussi claires (ou foncées) l'une que l'autre. Dans ce dernier cas, l'impression en noir et blanc d'un document contenant ces deux couleurs, donnera exactement le même gris pour les deux.

de produits. On peut aussi proposer une figure comparant les bâtons deux à deux ou décomposer chaque bâton en deux ou trois parties. Il faut cependant limiter les découpages pour garder une bonne lisibilité.

Pour comparer des évolutions : des aires

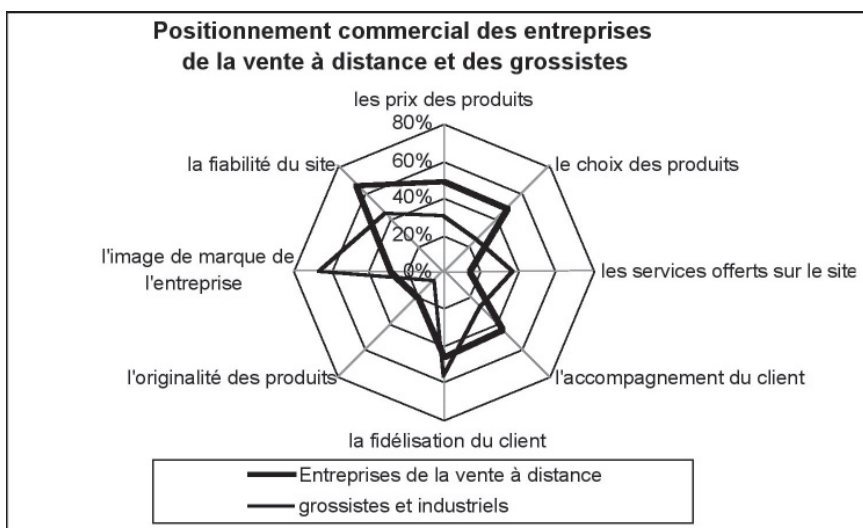
Cetypedereprésentation(graphique 7) met en valeur l'évolution d'une variable comportant deux ou trois modalités. Ici on voit bien, en lecture globale, l'augmentation de la proportion de bacheliers au cours du temps ; en lecture plus décomposée, l'apparition de nouveaux bacs et la répartition des types de bacheliers. Au-delà de trois composantes graphiques, la lecture devient difficile.

Graphique 7 : Exemple d'utilisation des aires



Source : TEF 2003-2004, page 61.

Graphique 8 : Exemple d'utilisation des radars



Source : Insee, Enquête sur le commerce électronique (2004) Insee première n° 999.

Des radars pour s'orienter

Popularisé par les comparaisons de performances des appareils du son et de l'image, ce type de figure repère les points forts (ici les préoccupations premières des entreprises, graphique 8) selon une échelle de 0 à 100 sur plusieurs axes. On voit ainsi que les entreprises de vente à distance (sur catalogue ou sur internet) ont pour priorité la fiabilité de leur site, le choix des produits et l'accompagnement de la clientèle. Leur image de marque est déjà établie. À l'inverse, les grossistes sur internet ont comme préoccupation première de se faire connaître et de proposer des services en ligne. Ils ne font pas d'effort particulier sur leurs prix, ils sont déjà très bien placés du simple fait de la suppression des intermédiaires. La lecture des radars, on le voit, est donc aisée.

Le diagramme en secteurs ou « camembert »

Très utilisé, ce type de graphique est familier au lecteur. Il a un côté affectif lié au « partage du gâteau ». À ce titre, il fait partie des graphiques de communication. Il est pourtant assez pauvre en information. En effet, bien que la figure soit en deux dimensions, on utilise non pas deux mais une seule dimension pour représenter les composantes (c'est-à-dire les parts du camembert), car c'est la surface des secteurs qui permet de les quantifier, or celle-ci est proportionnelle à l'angle du secteur, mais pas à son rayon, qui lui est constant.

De plus, la lecture de ce type de graphique présente un risque d'interprétation biaisée car l'œil distingue moins bien les différences entre secteurs (d'un camembert) qu'entre hauteurs (d'un histogramme). Si l'on tient tout de même à utiliser cette représentation, les secteurs ne doivent pas être trop nombreux : sept ou huit restent un maximum.

Les cartes

On n'a rien trouvé de mieux que la cartographie pour représenter les variables géographiques. Une carte bien construite parle mieux qu'un long discours. Mais l'étendue de ces territoires n'est pas forcément en rapport avec les phénomènes à cartographier. Ainsi l'Île-de-France, 2 % du territoire, 19 % de la population, doit-elle toujours faire l'objet d'un « cartouche », c'est-à-dire un zoom représenté à part, sur les cartes de France.

En outre, le découpage de la carte reprend souvent des découpages administratifs ; or, ces derniers n'ont pas nécessairement de réalité objective par rapport au phénomène étudié et peuvent alors induire le lecteur en erreur.

Par exemple, pour représenter des parts ou des variations entre deux dates selon un découpage administratif (commune, département), on peut choisir d'associer à chaque unité du découpage (la commune, le département) une nuance de gris plus ou moins foncée, ou bien une trame⁷ plus ou moins serrée. Mais

cette représentation a ses limites, car elle peut donner une importance visuelle exagérée (et donc fausse) à la trame appliquée à une grande étendue. C'est ce qui peut arriver par exemple avec Fontainebleau, plus grande commune d'Île-de-France en superficie, mais dont la zone habitée est en réalité très petite : dans ce cas, un découpage infra-communal, distinguant les zones habitées des forêts, permettrait d'éviter le problème. De façon générale, l'idéal est de choisir une échelle suffisamment petite pour que l'effet visuel soit juste : des trames appliquées à une carte de France découpée par commune ne se repèreront pas une par une, mais donneront l'effet (juste) d'une continuité de zones homogènes.

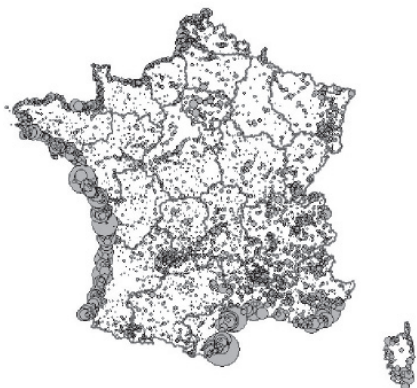
Autre possibilité, la représentation par points, dont la surface est proportionnelle à la grandeur considérée : c'est un outil classique pour montrer où se concentrent les populations. Ce type de carte est d'autant plus lisible que les points sont petits et nombreux (graphique 9).

Enfin, d'autres techniques plus complexes, dites de « lissage statistique », donnent des résultats intéressants. Elles ne seront pas abordées ici. Elles supposent généralement, pour un bon résultat, de disposer d'une impression en quadrichromie.

Les notes de lecture

Une note de lecture est parfois utile. Attention : si elle est longue, votre illustration est sans doute trop complexe.

Graphique 9 : exemple d'utilisation d'une représentation par points (concentration spatiale des campings)



Source : Insee, direction du Tourisme (à paraître dans le Tourisme en France)

Le tableur Excel

Les tableurs permettent de faire facilement des graphiques à partir d'un tableau. Mais ils ne donnent pas de résultat miracle. Excel est actuellement le tableur le plus répandu. Très performant pour les calculs, il l'est moins pour les représentations graphiques. Toutes les représentations ci-dessus y sont prévues, mais les options par défaut ne sont pas les meilleures, elles gagnent à être changées.

Excel présente aussi le défaut majeur de ne pas permettre la représentation graphique des histogrammes, utilisés dans le cas des variables quantitatives continues, et dont les rectangles successifs, contrairement à ceux des diagrammes en bâtons, peuvent avoir des largeurs proportionnelles le long de l'axe des abscisses.

- Ne pas user du relief

Une des options offertes mérite d'être totalement évitée pour la présentation de résultats statistiques, c'est l'option relief ou 3D. Elle peut paraître séduisante, mais fausse les perspectives et donc la lecture correcte des échelles. De même, les nombreuses fioritures de motifs proposées pour les bâtons sont inutiles, voire nuisibles à la lecture.

- User de la mise en ordre des données

Par définition, les variables quantitatives sont déjà ordonnées : tranche de revenus, de salaire, d'âge. Le graphique utilisera bien entendu le classement en question. Dans d'autres cas (variables qualitatives non ordonnées), la présentation initiale du tableau de chiffres qui a servi à construire le graphique est tributaire d'une nomenclature d'activités ou de professions. Cette nomenclature, classée selon une logique qui lui est propre (ordre alphabétique par exemple), n'est pas forcément adaptée à une représentation particulière. L'exemple du graphique 4 montrait ainsi comment un tri judicieux pouvait donner davantage de sens à un graphique, permettant de répondre à la question : qui part le plus souvent en voyage ?

- Élargir et opposer les barres

Des options permettent de donner plus de surface aux barres ou bâtons

en les élargissant. On facilite alors les comparaisons visuelles, de même qu'en réduisant l'espace vide entre deux barres. L'option « superposer les barres à 100 » permet de mettre en vis-à-vis des valeurs négatives et positives ou de construire des pyramides.

La couleur la plus visible (ou la valeur noire dans le cas d'un graphique en noir et blanc) doit être réservée au sujet traité. Si des données sont systématiquement comparées à une moyenne, ce sont ces données qui, tout au long de l'article, doivent rester les plus visibles dans les graphiques, et traitées de manière semblable d'un graphique à l'autre.

Enfin, si l'auteur publie son étude sur un support en noir et blanc, il gagnera toujours du temps et en fera gagner beaucoup à l'équipe de réalisation en testant son graphique au préalable sur une imprimante noir et blanc.

Pour résumer, quelques conseils pour publier une bonne illustration

Qu'il s'agisse d'un tableau ou d'un graphique, choisissez les éléments les plus importants dans la batterie de données dont vous disposez.

Pour mettre en valeur l'information essentielle, pensez à trier vos données (dans les graphiques et les tableaux) et à mettre en premier les plus récentes (tableaux).

Pour les graphiques, privilégiez des représentations classiques et faciles à lire.

Ne vous faites pas piéger par les couleurs : un bon graphique doit être pertinent en noir et blanc.

Un tableau et un graphique doivent impérativement comporter : un titre, si possible informatif ; une unité ; une source.

Le logiciel ne fait pas tout, tout seul : prenez conseil auprès des spécialistes.

7. Une trame est un « hachage » (ou quadrillage) plus ou moins serré, donc perçu visuellement comme étant plus ou moins foncé. On l'utilise, de même que les nuances uniformes de gris, pour représenter, sur une carte géographique, les modalités prises par une variable, l'idée (perçue intuitivement par l'œil) étant que plus c'est foncé, plus la valeur prise par la modalité est élevée.