

Direction de la Diffusion et de l'Action régionale

H 2018/03

Guide de sémiologie graphique

Département de l'action régionale

Document de travail



Institut National de la Statistique et des Études Économiques

SOMMAIRE

Introduction	2
Objectifs d'un graphique.....	3
1. Que veut-on montrer ?.....	3
2. Pourquoi choisir une représentation graphique ?.....	4
3. Raconter une histoire	4
Les limites de la perception visuelle et du contexte éditorial.....	5
1. Illusions graphiques et fausses interprétations, daltonisme.....	5
2. Erreurs d'appréciation des ordres de grandeur.....	5
Règles.....	7
1. Les variables visuelles.....	7
2. Principes de lisibilité d'un graphique.....	8
3. Un niveau de lecture autonome.....	8
4. Une lecture globale immédiate et un message simple.....	8
5. Un nombre limité de variables représentées.....	8
6. Un graphique doit être habillé.....	8
7. Quelques principes d'un bon graphique.....	9
8. Des codes graphiques classés selon leur capacité à transmettre fidèlement l'information.....	9
9. Bien choisir sa représentation	11
Les différents types de graphiques.....	15
1. Diagrammes en bâtons.....	18
2. Courbes.....	19
3. Graphiques en secteurs.....	20
4. Nuages de points.....	21
5. Box plot de Tukey ou boîte à moustaches.....	23
6. Aires empilées.....	24
7. Radars.....	24
8. Graphiques moins classiques.....	25
Quelques pièges à éviter.....	33
1. Des surfaces qui se chevauchent.....	33
2. Le mirage de la 3D.....	33
3. Des surfaces censées être comparables et qui ne le sont pas.....	34
4. L'effet spaghetti.....	34
5. Le mouvement pendulaire ou le graphique torticolis.....	35
6. Axes en déséquilibre.....	36
7. Trouver le bon ratio hauteur / largeur.....	36
8. Des graphiques "modernes".....	36
Pour en savoir plus.....	38

INTRODUCTION

Ce guide de sémiologie graphique vise à rappeler quelques règles de sémiologie et à promouvoir des bonnes pratiques.

Dans ses numéros 115, 120 et 1261, le *Courrier des Statistiques* soulignait l'importance du choix et de la qualité des figures. Ces sujets sont restés d'actualité. La généralisation de la couleur dans les publications a fait évoluer les pratiques. Par ailleurs, l'utilisation des graphiques et des cartes s'est beaucoup développée dans la presse et plus largement dans les publications. De nouvelles illustrations sont apparues, notamment avec l'arrivée de nouveaux logiciels ou langages sur le marché (R et QGIS par exemple).

La prise en compte des règles de lisibilité favorise la compréhension des messages de l'Insee par les acteurs publics, les enseignants, les journalistes... et plus généralement l'accès à la culture statistique pour tous.

La plupart des règles présentées dans ce guide sont communément admises. Mais il y a toujours des exceptions et les avis peuvent diverger sur l'opportunité ou non du choix de tel ou tel graphique.

Ce guide a été réalisé en 2015 dans le cadre d'un groupe de travail, piloté par Thomas Le Jeannic (département de l'action régionale). Y ont participé : Sylvain Duverne et Marianne Borzic (direction régionale Pays de la Loire) ; Gilles Zemis (direction régionale Bourgogne-Franche-Comté) ; Lionel Cacheux (direction régionale Grand-Est) ; Bernard Nozières (direction régionale Occitanie) ; Mélanie Bigard et Joachim Timoteo (direction régionale Provence-Alpes-Côte d'Azur) ; Hélène Valdelièvre et Alain Da Silva (direction générale, département de l'offre éditoriale).

Le Département de l'action régionale
DDAR – Insee

¹ *Courrier des statistiques* n° 2009_01 n° 126 - La perception graphique : mieux construire et interpréter les graphiques – O. Monso, I. de St Paul – Insee

Courrier des statistiques n° 2007_07 n° 120 - Savoir compter, savoir conter, épisode n° 6 : habiller un article, difficile mais indispensable
Courrier des statistiques n° 2005-09 n° 115 – Savoir compter, savoir conter, épisode n° 3 : des graphiques et des tableaux

OBJECTIFS D'UN GRAPHIQUE

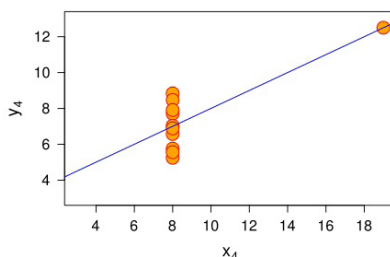
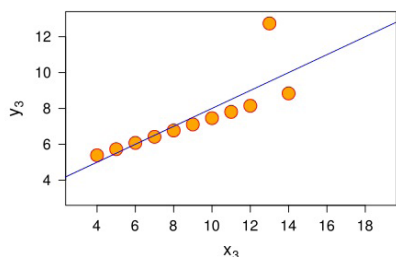
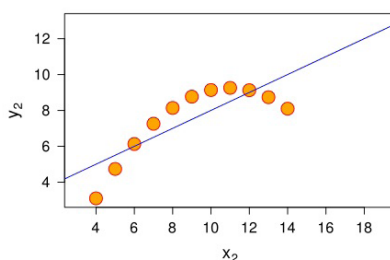
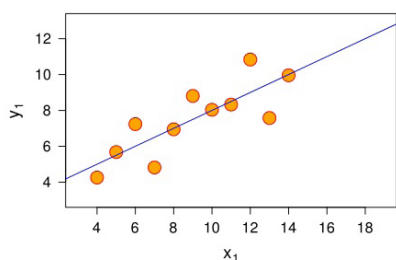
« Un bon croquis vaut mieux qu'un long discours » Napoléon

La vision est notre sens le plus puissant. Environ 70 % des récepteurs sensoriels du corps humain sont dans l'œil. La vision est ainsi la plus grande source d'informations pour le cerveau. Le système visuel obéit cependant à des règles. Quand les données sont bien présentées, les modèles ou configurations peuvent être facilement perçus. Mais si les règles ne sont pas respectées, les données peuvent alors être soit incompréhensibles, soit biaisées.

Les neurosciences ont montré que le traitement de l'information se faisait de façon complémentaire dans les deux hémisphères cérébraux. Plus analytique, l'hémisphère gauche est particulièrement apte à traiter l'information verbale, en procédant de façon linéaire et séquentielle. Plus synthétique, l'hémisphère droit est très efficace pour le traitement visuel et spatial, celui des images. Il recherche et construit des structures, en reconnaissant les relations entre éléments séparés. Sa façon de traiter l'information est beaucoup plus globale. Il serait le siège de l'intuition créatrice. Aucun des deux hémisphères ne peut fonctionner seul de manière efficace. Leur complémentarité donne à la pensée toute son efficacité.



1 Que veut-on montrer ?



L'exemple suivant d'Anscombe montre l'intérêt des graphiques par rapport à des statistiques ou des indicateurs sophistiqués. Les quatre distributions suivantes présentent des caractéristiques statistiques communes : même moyenne, même modèle de régression linéaire, même somme des carrés des résidus, même coefficient de corrélation, même R^2 . Toutefois, la répartition des données est très différente. Visualiser rapidement les données à l'aide d'un graphique adapté permet parfois d'aller plus vite à l'information. Il est alors nécessaire de choisir le type de représentation (tableau, graphique, carte, schéma...) en fonction du message que l'on souhaite délivrer.

En vue de choisir le meilleur graphique, le statisticien doit adopter le point de vue du lecteur et se demander :

- Quel est l'objectif de ce graphique ? présenter simplement des données, illustrer une relation...
- Quel type de comparaisons veut-on faire ? comparaison spatiale, temporelle...
- Quels éléments du graphique vont retenir l'attention du lecteur ? une couleur, une forme particulière...
- Ces éléments servent-ils à la compréhension du message ?
- À quel type de lecteurs s'adresse-t-on ? des journalistes, des étudiants, des chercheurs...
- Le graphique est-il approprié pour ses destinataires ?

② Pourquoi choisir une représentation graphique ?

Les graphiques sont des outils essentiels pour l'analyse et la représentation de données quantitatives. L'image, couplée à un raisonnement, facilite la compréhension, la mémorisation et diminue le temps d'apprentissage. Les figures sont les premiers éléments regardés par un lecteur avant même de commencer la lecture du texte. Par ailleurs, les figures illustrent un message compris dans le texte et le complètent également, en apportant notamment une information plus précise et plus exhaustive. La nouvelle charte graphique de l'Insee préconise de faire référence aux figures dans le texte et d'en placer systématiquement une en première page de publication.

La représentation graphique comporte des avantages par rapport à un tableau :

- Appréhension directe et globale de l'information : un graphique remplace avantageusement un long tableau ; il permet de « traiter » une information d'un seul coup d'œil. Un bon graphique exploite la capacité de l'œil à traiter simultanément beaucoup d'informations.
- Délivre efficacement un message : il permet notamment de raconter une histoire.
- Constituer un bon outil de communication : plus facilement repris par les médias ou dans une présentation publique.

Ainsi, concevoir une bonne représentation graphique est primordial pour la transmission de l'information.

Emmanuel Goldenweiser a proposé en 1916 une typologie des graphiques en quatre grandes catégories selon l'objectif qu'ils poursuivent : le graphique référentiel, illustratif, analytique et de recherche.

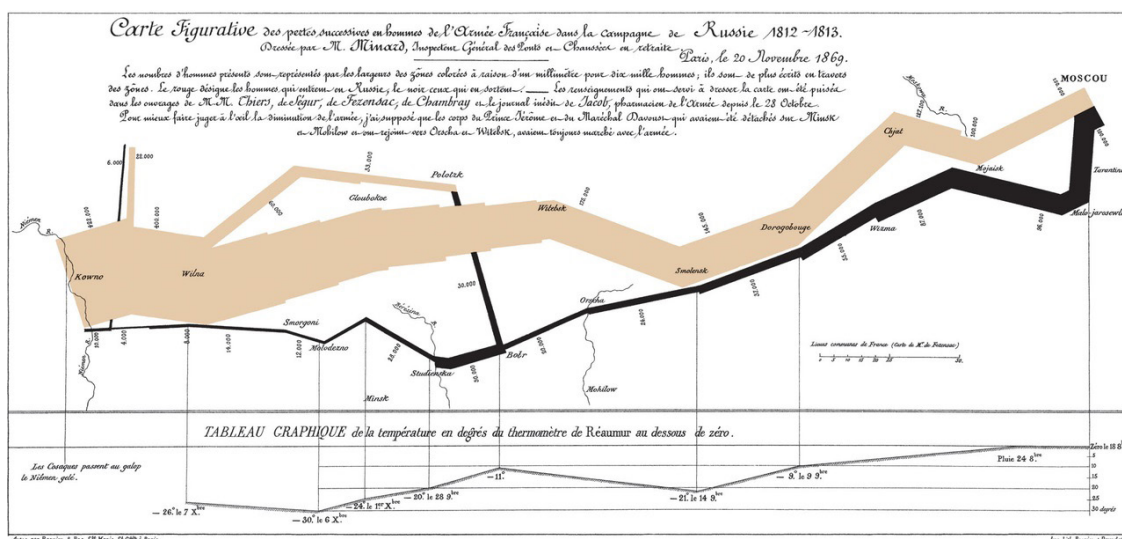
1. Le graphique référentiel tend à l'exhaustivité. Il doit être clair et simple car son seul rôle est d'exposer une information de manière plus compréhensible que ne le ferait un tableau de données.
2. Le graphique illustratif vise à fixer un fait important dans l'esprit du lecteur en s'appuyant sur sa mémoire visuelle. Il implique une sélection des informations à présenter.
3. Le graphique analytique montre visuellement une relation que l'auteur veut mettre en valeur, de façon plus aisée que ne l'aurait fait un texte écrit ou un tableau. Il peut par exemple mettre en évidence le lien existant entre deux variables.
4. Le graphique de recherche est un peu à part et reste le plus souvent au stade expérimental. Il a pour but d'aider à établir une relation inconnue, ou dont l'auteur n'est pas forcément certain qu'elle existe.

Le présent guide vise à présenter les graphiques de communication, qu'ils soient illustratifs ou analytiques.

③ Raconter une histoire

Napoléon et la campagne de Russie :

La carte figurative réalisée en 1869 par Charles Joseph Minard (1781-1870), ingénieur français, est un bon exemple de graphique qui à lui seul « raconte une histoire ». La déroute de l'armée française et les pertes humaines au cours du temps jusqu'à Moscou (diagramme beige représentant le nombre de soldats) puis durant la retraite (diagramme noir) y sont représentées de façon claire et détaillée. Il inclut des éléments sur la géographie des lieux et sur la rudesse des conditions climatiques (graphique du dessous), montrant ainsi le lien entre la défaite française et le froid enduré par les troupes. D'après Edward Tufte, il « pourrait bien être le meilleur graphique statistique jamais dessiné ».



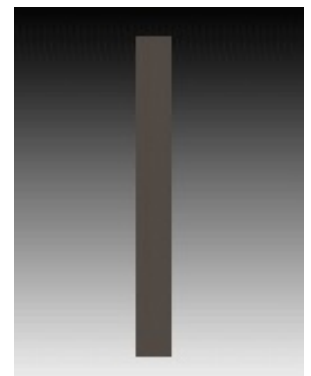
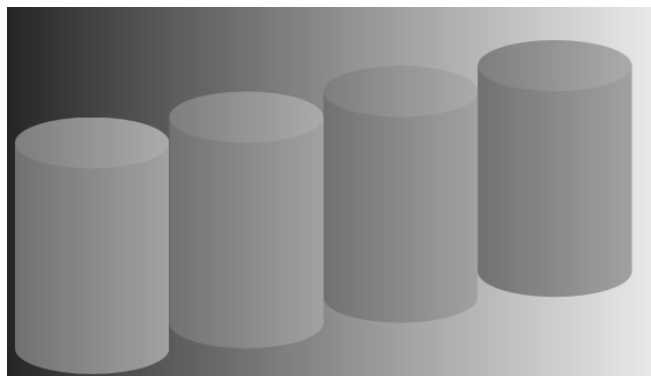
¹ Cette illustration sera d'ailleurs proposée en vignette pour les publications nationales de l'Insee qui apparaîtront dans la rubrique actualité du nouveau site Insee.fr en 2016.

LES LIMITES DE LA PERCEPTION VISUELLE ET DU CONTEXTE ÉDITORIAL

L'œil et le cerveau humain n'étant pas parfaits, des illusions optiques peuvent perturber l'analyse d'une figure.

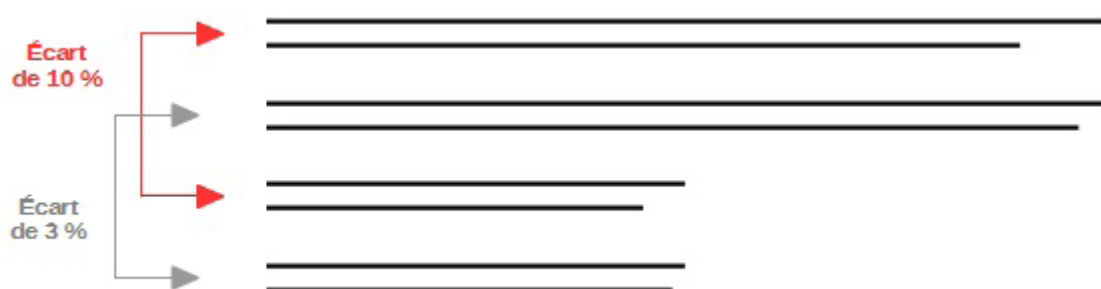
① Illusions graphiques et fausses interprétations, daltonisme

L'arrière-plan a une influence sur la perception des couleurs grisées des objets. Ainsi, dans les graphiques suivants, tous les gris sont identiques, contrairement à la perception que nous en avons.



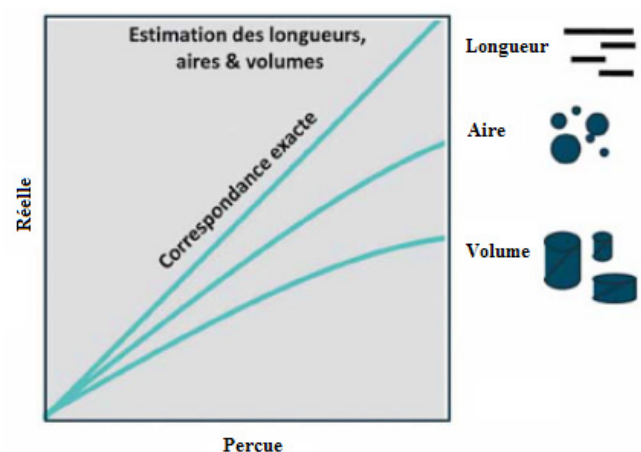
② Erreurs d'appréciation des ordres de grandeur

Dans le graphique suivant, nous percevons très bien une différence de 10 %, moins bien de 3 %.



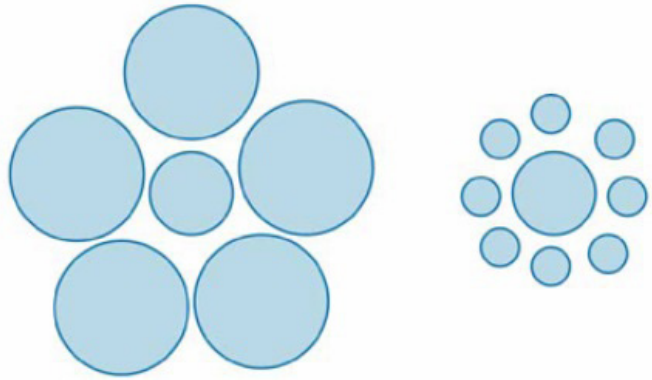
Les différences de longueurs sont mieux perçues que les différences de surfaces, et encore plus de volume :

Représentation de l'écart de perception en fonction de la nature du support et de l'intensité de la variation.



Dans l'exemple ci-contre, lequel des deux cercles centraux est le plus grand, celui de gauche ou de droite ?

En fait, ils sont identiques.



Sur la base d'une construction réelle

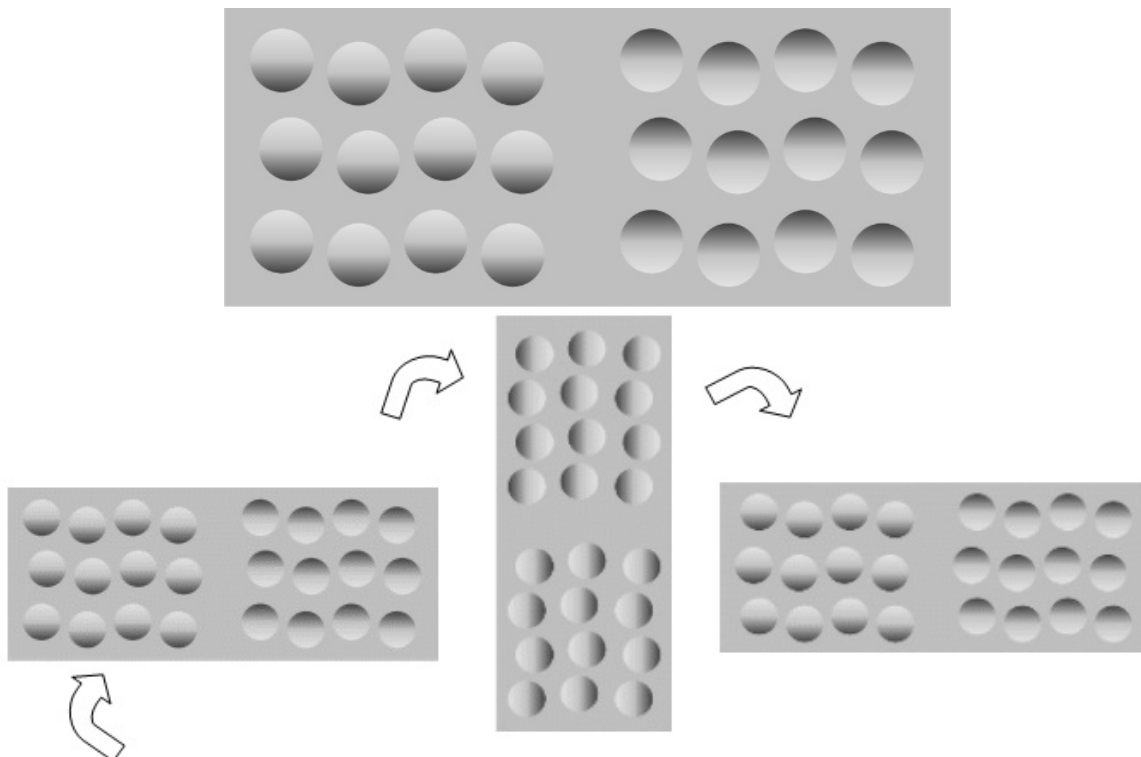


Sur la base d'une perception visuelle

Par ailleurs, l'œil sous-estime systématiquement la surface du cercle. Les graphiques ci-contre partent d'un même cercle de référence (de surface 100 unités). Le graphique de gauche représente, en respectant les proportions mathématiques, un cercle ayant une surface 3 fois plus grande, 6 fois plus grande et 12 fois plus grande que le cercle de référence. Le graphique de droite montre les cercles qui seraient représentés si on se fiait à notre seule perception visuelle. Ils sont sensiblement plus grands.

Des effets d'optique jouent également sur la vision des volumes. Dans ce dessin, lesquels parmi ces objets sont concaves, lesquels sont convexes ? A priori ceux de gauche sont convexes, ceux de droite sont concaves.

Retournons maintenant le dessin. Ceux qui étaient convexes sont devenus concaves et vice versa.



Ce phénomène s'explique par un effet de la lumière. L'œil s'attend à voir une lumière qui vient d'en haut, il interprète donc l'image en fonction de cela.

Pour en savoir plus sur les illusions d'optique :

<http://tpe610.free.fr/illusions.htm>

<http://tpeesperance2010.e-monsite.com/pages/iii-les-illusions-d-optique.html>

RÈGLES

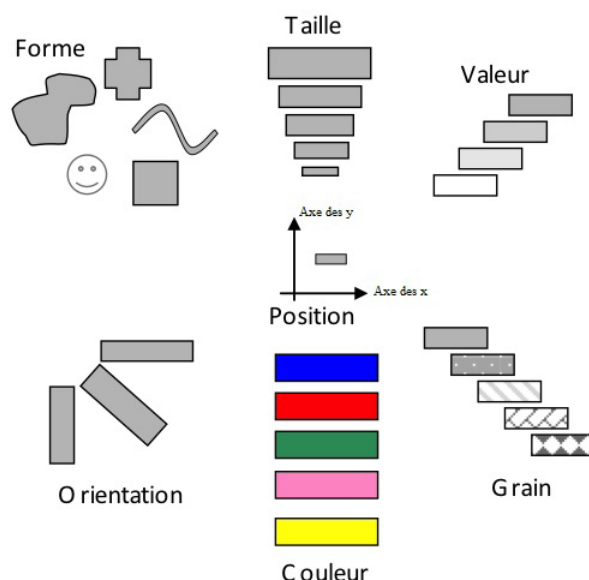
« It's better to be approximately right than exactly wrong » John W. Tukey

① Les variables visuelles

Les variables rétinienne de Bertin

Jacques Bertin est considéré comme le père fondateur de la discipline de la visualisation de l'information, y compris outre-Atlantique. Il recense six « variables rétinienne » qui viennent compléter les coordonnées de placement de chaque objet sur un plan doté des axes x et y. Chacune de ces variables peut être utilisée pour visualiser des informations. En fonction de la nature de la variable choisie, différents types de messages et d'informations peuvent être transmis.

À ces six variables, S.K. Card et J. Mackinlay ont ajouté deux variables supplémentaires, la connexion et l'inclusion.



Variable	Définition	Informations	Exemples d'utilisations
Forme	Dessin ou pictogramme servant à qualifier la nature de l'information	Indique la nature de l'information restituée seule ou qui peut être combinée avec d'autres variables (taille, couleur...)	Pour distinguer des types d'activité différents, des équipements différents
Taille	Surface/aire de la forme	Grandeur/taille en référence à une échelle de valeurs	Pour présenter des nombres absolus, des volumes
Valeur	Variation de couleur entre 2 extrêmes	Positionnement dans une échelle de valeurs avec un dégradé d'une même couleur entre deux extrêmes	Pour présenter des pourcentages, des taux, des ratios
Grain	Trame ou motif	Combinaison des valeurs ou d'informations	Moins utilisé aujourd'hui du fait de l'utilisation des couleurs
Couleur	Valeur au sein d'un nuancier de couleurs	Est utilisée soit pour traduire des informations de nature différente (une couleur par information) soit de nature ordonnée (froid/chaud)	Pour une typologie Par convention : • vert pour l'agriculture, gris pour l'industrie, marron pour le tertiaire • couleurs froides / chaudes = valeurs négatives / positives
Orientaion	Direction de l'objet	Traduit une évolution à la hausse ou à la baisse	Peu utilisée de nos jours
Coordonnées	Placement sur les deux axes x et y	Coordonnées géographiques ou par rapport à deux axes	Carte, courbe ou nuage de points
Connexion	Lien visuel entre différents objets	Causalité, hiérarchie, groupement	Carte de réseau de villes, liens entre des entreprises
Inclusion	Objet graphique inclus dans un autre	Hiérarchie, groupement	Ville-centre dans son agglomération, inclusion d'une sous-population

② Principe de lisibilité du graphique

Pour être lisible et facilement compréhensible, un graphique doit satisfaire aux critères suivants :

- Un niveau de lecture autonome
- Une lecture globale immédiate
- Un message simple
- Un nombre limité de variables représentées
- Un graphique doit être habillé

③ Un niveau de lecture autonome

Le graphique doit si possible être compréhensible sans légende, grâce au titre et au respect des règles de sémiologie (couleurs, symboles proportionnels...). Un graphique doit pouvoir se comprendre de façon autonome, sans référence au texte.

④ Une lecture globale immédiate et un message simple

« *The best displays are often the simplest* » Stephen Few

« *La perfection est atteinte, non pas lorsqu'il n'y a plus rien à ajouter, mais lorsqu'il n'y a plus rien à retirer* »
Antoine de Saint-Exupéry

Un graphique efficace permet d'apporter l'information d'un seul coup d'œil : lecture globale immédiate comportant un message simple. Un graphique répond à ces critères si, pour obtenir une réponse correcte et complète à une question donnée, il requiert un temps d'observation court.

Un bon test est de demander à un lecteur combien de temps il met pour comprendre le graphique et en dégager un message :

- moins de 5 secondes
- de 5 à 10 secondes
- de 10 à 30 secondes
- de 30 secondes à 1 minute
- de 1 à 5 minutes
- plus de 5 minutes
- jamais.

Au-delà de 30 secondes, il est conseillé de trouver un autre graphique.

⑤ Un nombre limité de variables représentées

Trop d'informations tue l'information. Un lecteur ne peut généralement pas mémoriser plus de dix objets différents. Il ne faut donc pas multiplier les objets ou les couleurs sur un même graphique.

⑥ Un graphique doit être habillé

Pour être efficace, un graphique doit comporter toutes les informations nécessaires à sa compréhension.

Les éléments indispensables à l'habillage d'un graphique sont les suivants :

→ En dehors de l'image :

- Un titre informatif : un message simple et efficace
- Un sous-titre explicatif
- Les sources (en italique). Chaque source doit être séparée par une virgule et chaque organisme par un point virgule.
- Éventuellement : note, lecture, champ (à positionner dans cet ordre avant les sources)

→ Dans l'image :

- Les unités
- Les libellés des axes
- Une légende
- Éventuellement, un quadrillage

De plus, les titres et les légendes doivent être clairs et précis afin d'éviter toute ambiguïté.

7 Quelques principes d'un bon graphique (D'après Claude Angers (1987))

• Un bon graphique **transmet le plus fidèlement possible l'information d'origine** : il ne déforme pas l'information contenue dans les données.

Pour Edward Tufte, l'excellence graphique nécessite de restituer fidèlement les données. Il s'agit de donner le plus grand nombre d'informations pour un temps d'attention le plus limité tout en ayant recours à la plus petite quantité d'encre et au plus petit espace de représentation. Il estime cette excellence au travers de trois critères : le ratio entre encre utilisée/informations restituées, le coefficient de distorsion par rapport à la réalité et l'absence d'éléments décoratifs n'apportant pas d'informations supplémentaires.

• Il **optimise le ratio quantité d'encre/quantité d'information** : pas d'éléments décoratifs sans information supplémentaires.

Plus le ratio quantité d'encre/quantité d'information est faible, plus le graphique est efficace. Il est inutile de rajouter des éléments de décoration qui n'apportent rien à la compréhension du graphique, d'autant plus si cette décoration biaise la perception comme les graphiques en 3D ou une couleur de fond. Le graphique ne doit pas servir à distraire l'attention du lecteur du véritable sens des données.

- Il est **de préférence multivarié**, en décrivant simultanément deux ou plusieurs variables interreliées.
- Lorsqu'on représente des valeurs à l'aide de **surfaces**, ces dernières doivent être **proportionnelles aux valeurs**.
- L'indicateur choisi doit être **en lien avec les messages** que l'on veut transmettre.

8 Des codes graphiques classés selon leur capacité à transmettre fidèlement l'information

Suite à des expériences auprès de groupes de personnes, William Cleveland et Robert McGill ont caractérisé dix codes graphiques élémentaires et les ont classés selon leur capacité à transmettre fidèlement l'information d'origine.

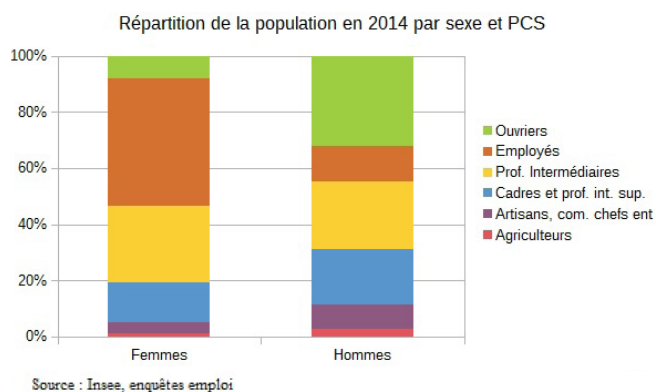
Les Codes graphiques élémentaires chez Cleveland et McGill (1984)

Code graphique élémentaire	Degré de précision des jugements
1. Position sur une échelle commune	1 ^{er}
2. Position sur des échelles différentes	2 ^e
3. Longueur	3 ^e
4. Direction	
5. Angle	
6. Aire	4 ^e
7. Volume	5 ^e
8. Courbure	
9. Densité de couleur	6 ^e
10. Intensité de couleur	

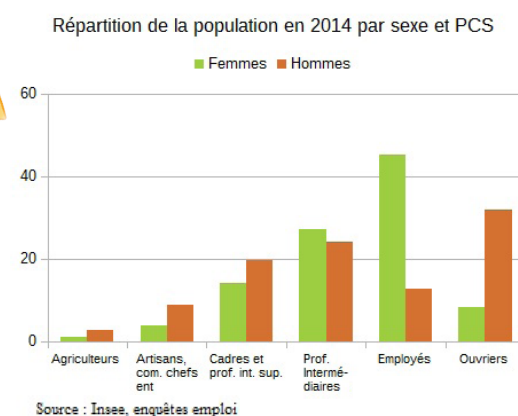
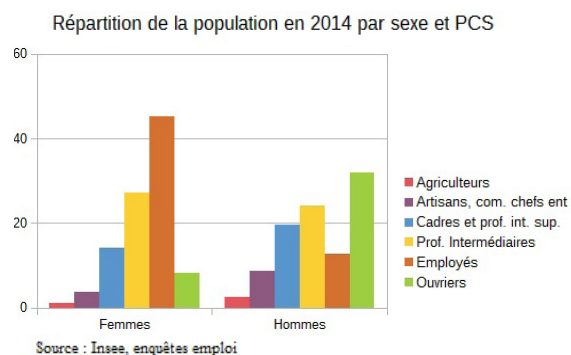
Les deux exemples suivants ont été proposés par Olivier Monso et Thibaut de Saint Pol dans le courrier des statistiques d'avril 2009. Les histogrammes empilés, bien souvent utilisés, font appel à un jugement sur les longueurs, lesquelles ne sont pas positionnées à partir du même endroit sauf les deux modalités extrêmes. Les deux autres histogrammes font également appel à un jugement sur les longueurs mais aussi à un jugement de position sur une échelle commune, réputé plus précis que le jugement de position sur des échelles différentes (cf. tableau ci-dessus).



Exemple



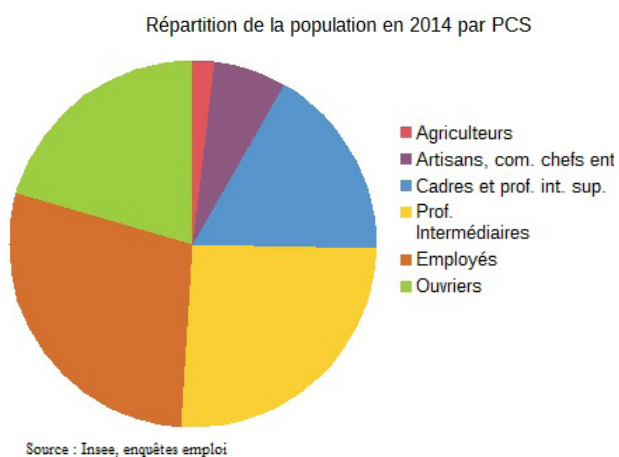
Proposition alternative



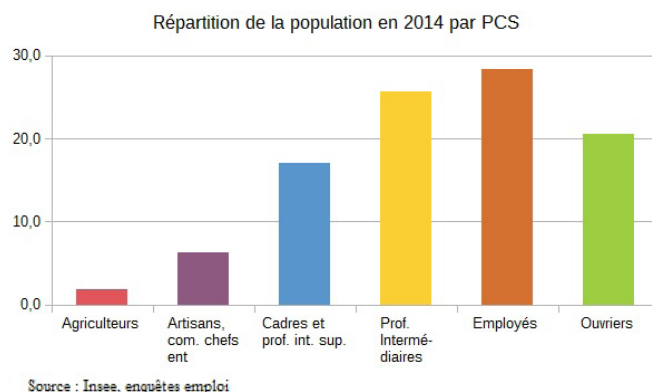
L'autre exemple proposé est le camembert, le jugement étant basé cette fois sur l'angle et la surface. Ce type de code graphique est également moins précis que la position sur une échelle commune.



Exemple



Proposition alternative



« Dans un graphique, il peut y avoir plusieurs niveaux de lecture, mais le premier niveau doit obligatoirement être lisible » Eugène Horber.

9 Bien choisir sa représentation

« Avec trois nombres vous faites une phrase, avec quatre à vingt nombres vous faites un tableau, au-delà de vingt nombres vous faites un graphique » E.Tufte.

→ Carte, graphique ou tableau ?

Le graphique est adapté pour la **comparaison de données non localisées** ou lorsque leur localisation est réduite et connue (les départements d'une région, par exemple). Pour des données par région française, un diagramme en bâtons peut avantageusement remplacer une carte, notamment quand plusieurs variables sont à représenter, et à la condition de trier ces variables par ordre croissant (ou décroissant).

Le graphique permet également une **bonne représentation des évolutions** et met en évidence les tendances.

→ Données quantitatives ou qualitatives ?

Les données à représenter peuvent être soit quantitatives, soit qualitatives.

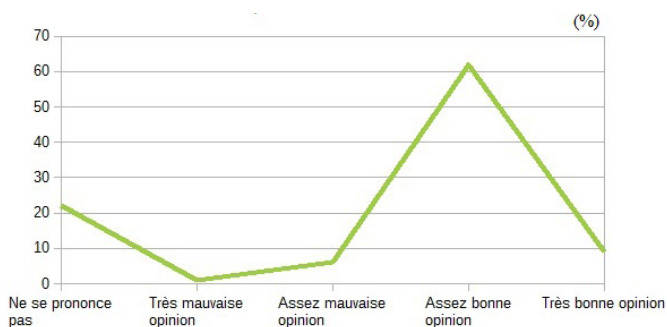
- Les variables quantitatives ont des nombres pour modalité. Les variables quantitatives continues peuvent prendre une infinité de valeurs. Les variables quantitatives discrètes ou discontinues ne peuvent prendre qu'un nombre limité de valeurs (par exemple, le nombre d'enfants).
- Les variables qualitatives ont pour modalité des mots. On peut distinguer les variables qualitatives ordonnées dont les modalités ont un ordre logique (par exemple : « tout à fait d'accord », « plutôt d'accord », « plutôt pas d'accord », « pas du tout d'accord »), et les variables qualitatives non ordonnées (profession, pays de naissance, etc.).

Une règle simple est de réserver les courbes aux données quantitatives continues. Les diagrammes en bâtons (ou les camemberts) sont plutôt réservés aux données quantitatives discrètes ou aux données qualitatives. Mais la frontière n'est pas toujours évidente. Ainsi, des années peuvent être comparées avec un diagramme en bâtons. En revanche, il est faux de représenter des grandeurs discontinues par des courbes, comme dans l'exemple suivant.



► Exemple

Opinion du public à l'égard de l'Insee

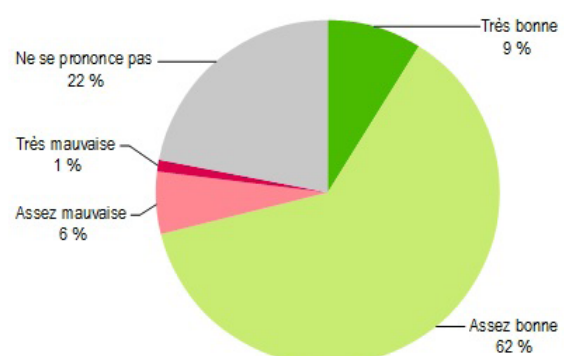


Source : baromètre "Image de l'Insee" mai 2015, CSA



► Proposition alternative

Opinion du public à l'égard de l'Insee



Lecture : 9 % des français ont une «très bonne» opinion de l'Insee.

L'enquête a été menée par téléphone, du 18 au 20 2015. L'échantillon était composé de 1004 personnes, constituant un échantillon national représentatif de la population française âgée de 18 ans ou plus. La méthode est celle des quotas : sexe, âge, profession de la personne de référence au sein du foyer, région et catégorie d'agglomération.

Source : CSA, baromètre « Image de l'Insee », mai 2015.

► Penser à trier ses données

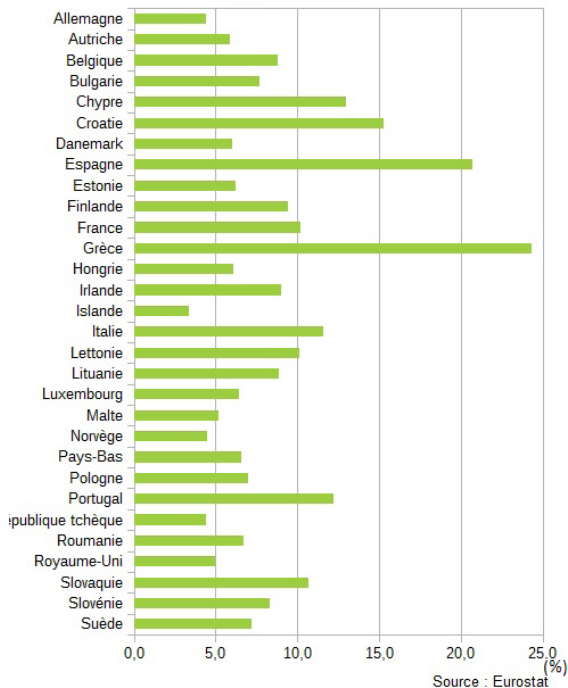
Il est important de trier les données pour les diagrammes en bâtons.

Dans l'exemple suivant, il est plus utile que les données soient triées par valeurs plutôt que par ordre alphabétique. Par ailleurs, il est conseillé de distinguer le territoire d'analyse (ici la France) par une couleur différente.



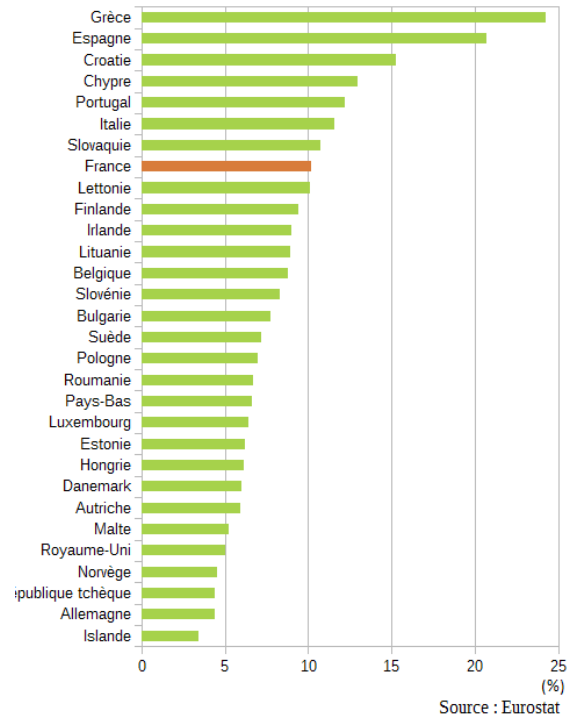
► Exemple

Taux de chômage dans l'Union Européenne
Décembre 2015



► Proposition alternative

Taux de chômage dans l'Union Européenne
Décembre 2015

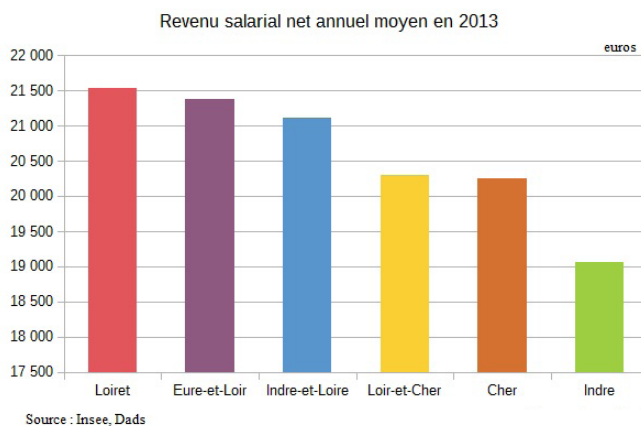


► User des couleurs sans en abuser

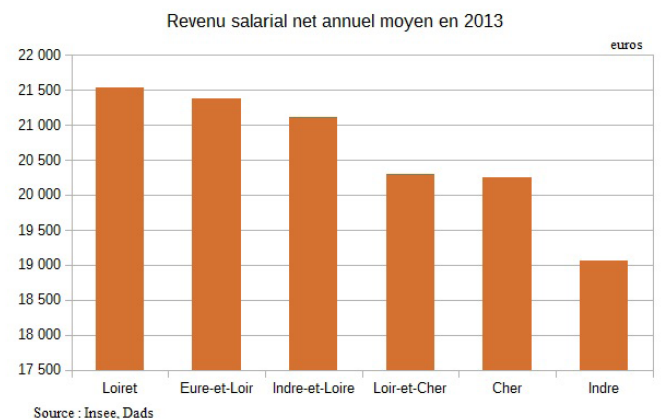
« Utiliser trop de couleurs diminue sa force » Stephen Few.



► Exemple 1



► Proposition alternative

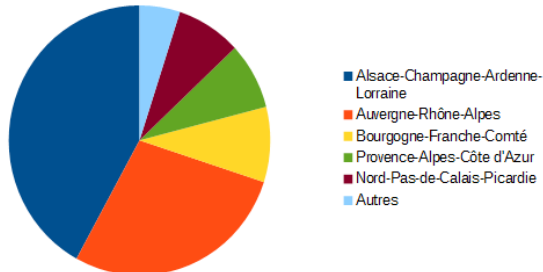


Entre les couleurs par défaut de certains logiciels et des couleurs saturées, il est préférable de choisir des couleurs plus douces.



Exemple 2

Répartition des actifs transfrontaliers selon leur lieu de résidence

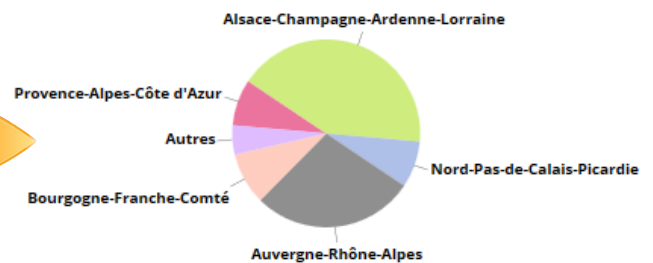


Source : Insee, recensement de population 2013



Proposition alternative

Répartition des actifs transfrontaliers selon leur lieu de résidence en %

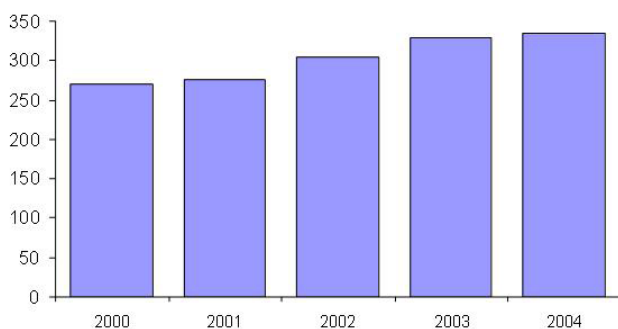


Champ : actifs transfrontaliers, France métropolitaine
Source : Insee, recensement de la population de 2013, exploitation complémentaire.

Du bon usage du temps

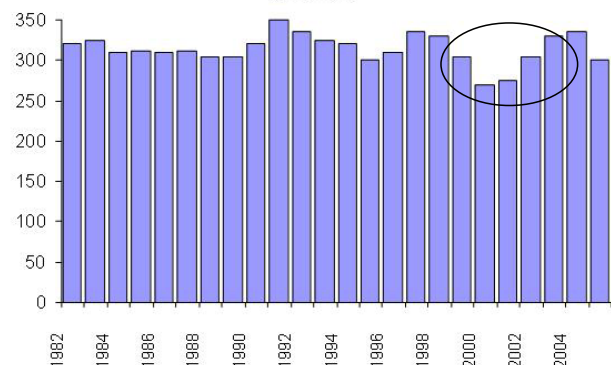
Pour la représentation de séries temporelles, il est important de se demander si la période choisie ne fausse pas le jugement. En effet, on peut facilement induire en erreur en sélectionnant une période particulière. Les données en soi ne sont pas fausses mais c'est leur sélection, plus ou moins volontaire, qui biaise l'information. Voici un exemple qui montre une tendance à la hausse entre 2000 et 2004, qui est toute relative dès lors que l'on regarde depuis 1982. Par ailleurs, il est recommandé de représenter le temps sur l'axe horizontal des x.

Nombre de délits enregistrés en Suisse en milliers



Source : Statistique policière de la criminalité

Nombre de délits enregistrés en Suisse en milliers

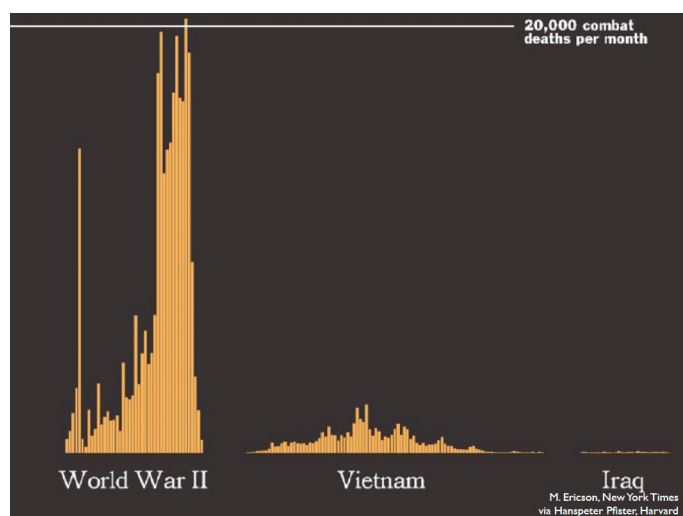
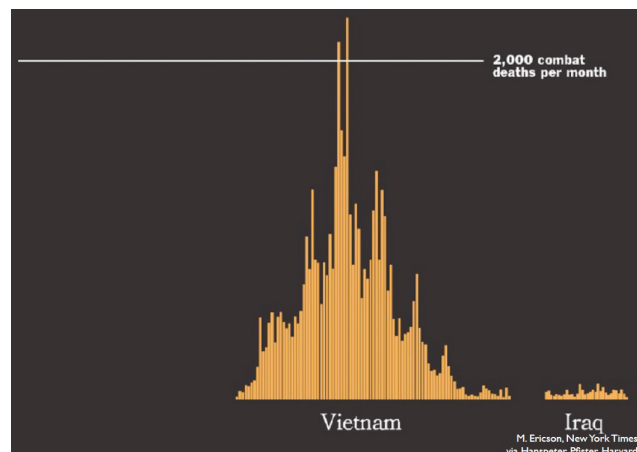


Source : Statistique policière de la criminalité

Le choix du référentiel de comparaisons

On doit toujours se demander à quoi on veut comparer les données.

Voici un exemple édifiant sur le nombre de soldats américains tués au combat durant trois grands conflits. Le constat diffère selon le choix des éléments de comparaison.

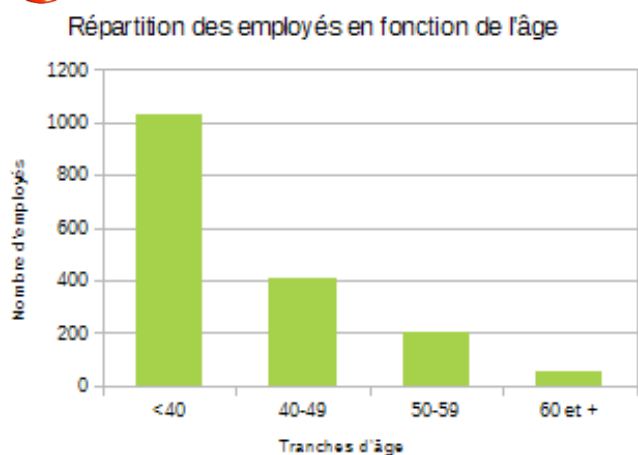


► Du bon découpage en classes

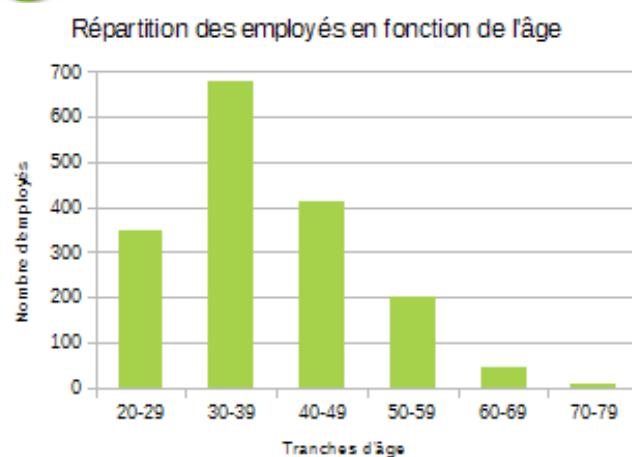
Le découpage en tranches dans le cas de variables quantitatives continues (par exemple l'âge) n'est pas neutre. Par ailleurs, pour pouvoir comparer les effectifs par tranche, il faut veiller à ce que l'étendue de ces tranches soit identique, comme dans le graphique suivant de droite. La règle d'égalité d'étendue de tranches peut parfois ne pas être respectée, comme dans l'exemple de droite où l'analyse ne porte que sur les 40-59 ans. ■



► Exemple



► Proposition alternative

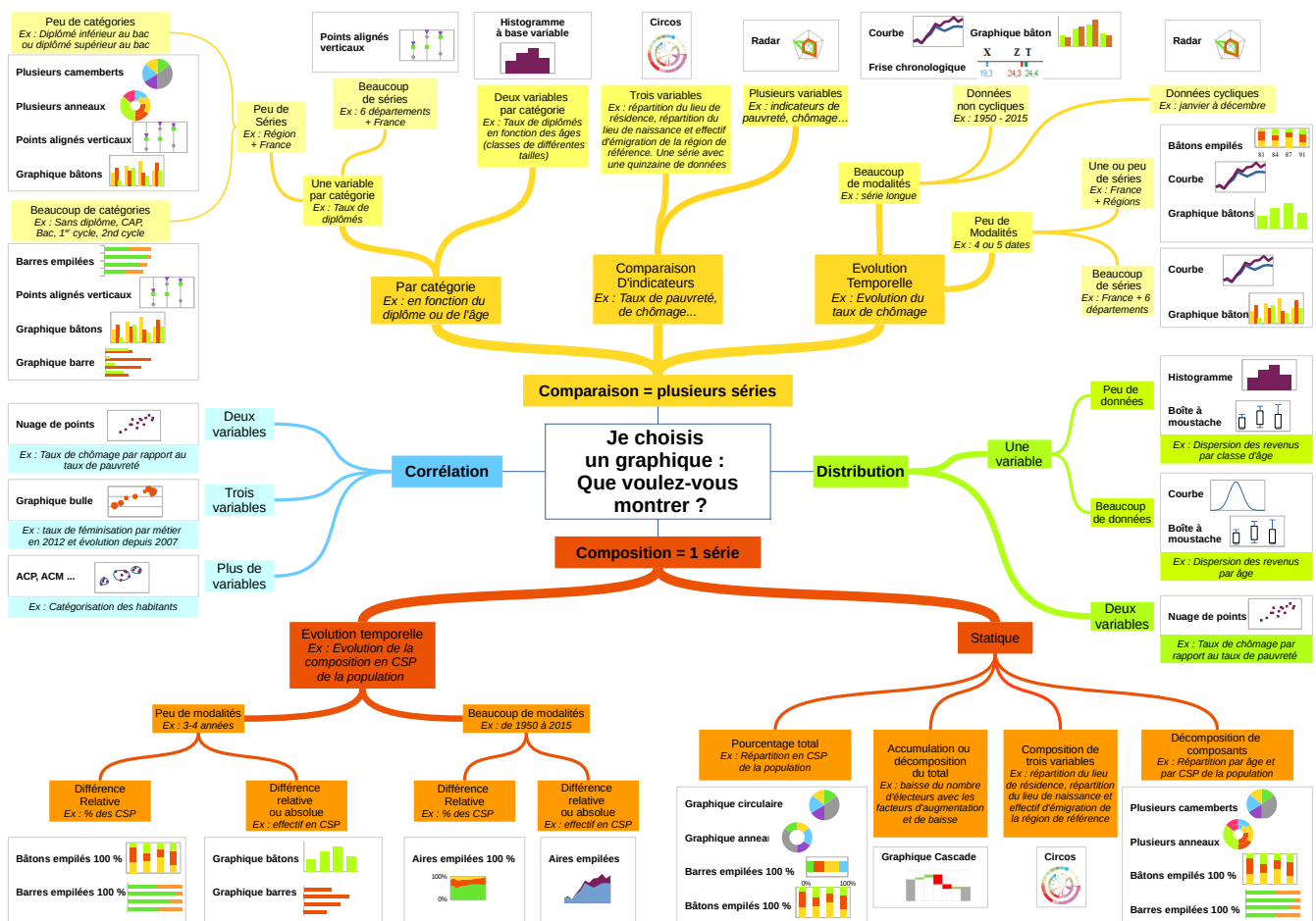


Source : Insee, Recensement de la population

LES DIFFÉRENTS TYPES DE GRAPHIQUES

Il existe des centaines de graphiques. Les logiciels comme R ou des compléments à Excel (Sparklines pour Excel® de Fabrice Rimlinger) en proposent de nombreux. Les pages Web offrent des quantités d'exemples de graphiques nouveaux (datavision, dataviz) qui peuvent alimenter notre inspiration. Cependant, la lisibilité, la clarté et la rigueur de la représentation doivent primer sur l'aspect purement décoratif et novateur.

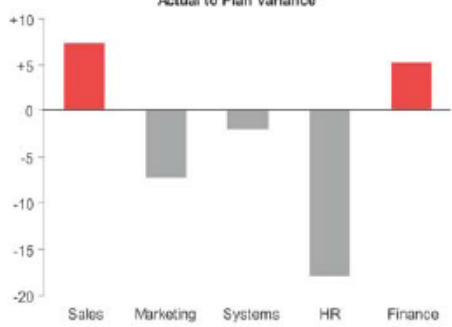
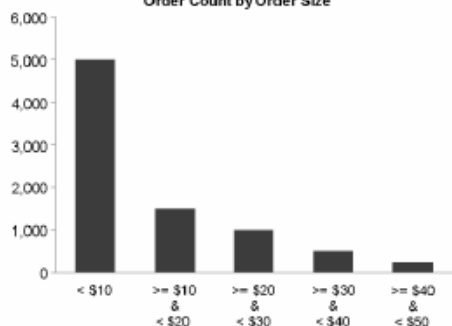
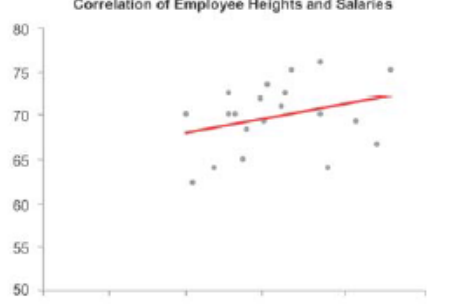
Voici un tableau synoptique pour choisir son graphique :



Stephen Few classe les graphiques en sept familles :

- Les comparaisons de catégories (Nominal Comparaison)
- Les séries temporelles (Time-Series)
- Les classements (Ranking)
- Les décompositions (Part-to-whole)
- Les variations (Deviation)
- Les distributions (Frequency Distribution)
- Les relations et corrélations (Correlation)

Type/Description	Encoding Methods	Example																																							
Nominal Comparison A simple comparison of the categorical subdivisions of one or more measures in no particular order	Bars only (horizontal or vertical)	<table border="1"><caption>Q1 2003 Calls by Region</caption><thead><tr><th>Region</th><th>Calls</th></tr></thead><tbody><tr><td>North</td><td>3,000</td></tr><tr><td>East</td><td>4,500</td></tr><tr><td>South</td><td>2,200</td></tr><tr><td>West</td><td>4,800</td></tr></tbody></table>	Region	Calls	North	3,000	East	4,500	South	2,200	West	4,800																													
Region	Calls																																								
North	3,000																																								
East	4,500																																								
South	2,200																																								
West	4,800																																								
Time Series Multiple instances of one or more measures taken at equidistant points in time	- Lines to emphasize overall pattern - Bars to emphasize individual values - Points connected by lines to slightly emphasize individual values while still highlighting the overall pattern - Always place time on the horizontal axis	<table border="1"><caption>2003 Sales</caption><thead><tr><th>Month</th><th>Category 1 (Grey)</th><th>Category 2 (Orange)</th></tr></thead><tbody><tr><td>Jan</td><td>2,200</td><td>1,000</td></tr><tr><td>Feb</td><td>2,500</td><td>1,000</td></tr><tr><td>Mar</td><td>2,800</td><td>1,000</td></tr><tr><td>Apr</td><td>2,500</td><td>1,000</td></tr><tr><td>May</td><td>3,000</td><td>1,000</td></tr><tr><td>Jun</td><td>2,800</td><td>1,000</td></tr><tr><td>Jul</td><td>2,500</td><td>1,000</td></tr><tr><td>Aug</td><td>2,800</td><td>500</td></tr><tr><td>Sep</td><td>3,200</td><td>1,000</td></tr><tr><td>Oct</td><td>3,000</td><td>1,000</td></tr><tr><td>Nov</td><td>3,500</td><td>1,000</td></tr><tr><td>Dec</td><td>3,800</td><td>1,000</td></tr></tbody></table>	Month	Category 1 (Grey)	Category 2 (Orange)	Jan	2,200	1,000	Feb	2,500	1,000	Mar	2,800	1,000	Apr	2,500	1,000	May	3,000	1,000	Jun	2,800	1,000	Jul	2,500	1,000	Aug	2,800	500	Sep	3,200	1,000	Oct	3,000	1,000	Nov	3,500	1,000	Dec	3,800	1,000
Month	Category 1 (Grey)	Category 2 (Orange)																																							
Jan	2,200	1,000																																							
Feb	2,500	1,000																																							
Mar	2,800	1,000																																							
Apr	2,500	1,000																																							
May	3,000	1,000																																							
Jun	2,800	1,000																																							
Jul	2,500	1,000																																							
Aug	2,800	500																																							
Sep	3,200	1,000																																							
Oct	3,000	1,000																																							
Nov	3,500	1,000																																							
Dec	3,800	1,000																																							
Ranking Categorical subdivisions of a measure ordered by size (either descending or ascending)	- Bars only (horizontal or vertical) - To highlight high values, sort in descending order - To highlight low values, sort in ascending order	<table border="1"><caption>Headcount</caption><thead><tr><th>Department</th><th>Headcount</th></tr></thead><tbody><tr><td>Manufacturing</td><td>230</td></tr><tr><td>Sales</td><td>160</td></tr><tr><td>Engineering</td><td>50</td></tr><tr><td>Operations</td><td>40</td></tr><tr><td>Finance</td><td>30</td></tr><tr><td>Info Systems</td><td>30</td></tr><tr><td>Legal</td><td>20</td></tr><tr><td>Marketing</td><td>10</td></tr></tbody></table>	Department	Headcount	Manufacturing	230	Sales	160	Engineering	50	Operations	40	Finance	30	Info Systems	30	Legal	20	Marketing	10																					
Department	Headcount																																								
Manufacturing	230																																								
Sales	160																																								
Engineering	50																																								
Operations	40																																								
Finance	30																																								
Info Systems	30																																								
Legal	20																																								
Marketing	10																																								
Part-to-Whole Measures of individual categorical subdivisions as ratios to the whole	- Bars only (horizontal or vertical) - Use stacked bars only when you must display measures of the whole as well as the parts	<table border="1"><caption>Regional % of Total Expenses</caption><thead><tr><th>Region</th><th>% of Total Expenses</th></tr></thead><tbody><tr><td>West</td><td>33%</td></tr><tr><td>East</td><td>30%</td></tr><tr><td>North</td><td>21%</td></tr><tr><td>South</td><td>15%</td></tr></tbody></table>	Region	% of Total Expenses	West	33%	East	30%	North	21%	South	15%																													
Region	% of Total Expenses																																								
West	33%																																								
East	30%																																								
North	21%																																								
South	15%																																								

Deviation Categorical subdivisions of a measure compared to a reference measure, expressed as the differences between them	• Lines to emphasize the overall pattern only when displaying deviation and time-series relationships together • Points connected by lines to slightly emphasize individual data points while also highlighting the overall pattern when displaying deviation and time-series relationships together • Bars to emphasize individual values, but limit to vertical bars when a time-series relationship is included • Always include a reference line to compare the measures of deviation against	Actual to Plan Variance  <table><tr><th>Department</th><th>Variance</th></tr><tr><td>Sales</td><td>+8</td></tr><tr><td>Marketing</td><td>-8</td></tr><tr><td>Systems</td><td>-2</td></tr><tr><td>HR</td><td>-18</td></tr><tr><td>Finance</td><td>+5</td></tr></table>	Department	Variance	Sales	+8	Marketing	-8	Systems	-2	HR	-18	Finance	+5
Department	Variance													
Sales	+8													
Marketing	-8													
Systems	-2													
HR	-18													
Finance	+5													
Frequency Distribution Counts of something per categorical subdivisions (intervals) of a quantitative range	• Vertical bars to emphasize individual values (called a <i>histogram</i>) • Lines to emphasize the overall pattern (called a <i>frequency polygon</i>)	Order Count by Order Size  <table><tr><th>Order Size Range</th><th>Order Count</th></tr><tr><td>< \$10</td><td>5,000</td></tr><tr><td>>= \$10 & < \$20</td><td>1,500</td></tr><tr><td>>= \$20 & < \$30</td><td>1,000</td></tr><tr><td>>= \$30 & < \$40</td><td>500</td></tr><tr><td>>= \$40 & < \$50</td><td>200</td></tr></table>	Order Size Range	Order Count	< \$10	5,000	>= \$10 & < \$20	1,500	>= \$20 & < \$30	1,000	>= \$30 & < \$40	500	>= \$40 & < \$50	200
Order Size Range	Order Count													
< \$10	5,000													
>= \$10 & < \$20	1,500													
>= \$20 & < \$30	1,000													
>= \$30 & < \$40	500													
>= \$40 & < \$50	200													
Correlation Comparisons of two paired sets of measures to determine if as one set goes up the other set goes either up or down in a corresponding manner, and if so, how strongly	• Points and a trend line in the form of a scatter plot • Bars may be used, arranged as a <i>paired bar graph</i> or a <i>correlation bar graph</i>, if scatter plots are unfamiliar • (Note: For descriptions of these graphs, see my book <i>Show Me the Numbers</i>.)	Correlation of Employee Heights and Salaries 												

Les graphiques classiques

- Diagrammes en bâtons, pyramides des âges
- Courbes
- Camemberts, secteurs
- Nuages de points
- Boîtes à moustaches (Box plot)
- Aires empilées
- Radars

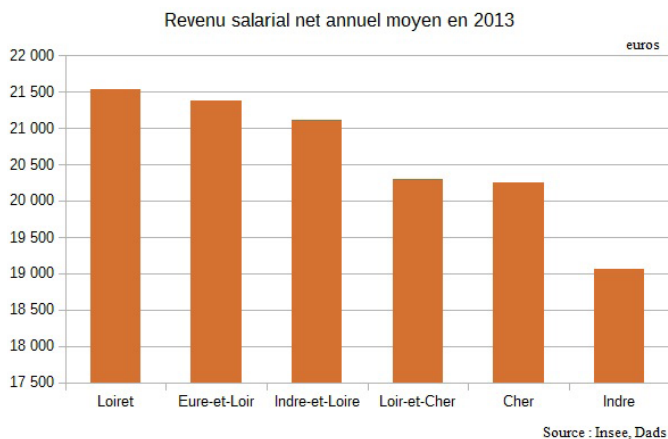
Les graphiques moins classiques

- Graphique en une ligne, chronogramme, histogramme en cascade, bulles temporelles, Treemap, histogramme avec relations entre les barres, Circos, matrice de points, tapis...

1 Diagrammes en bâtons

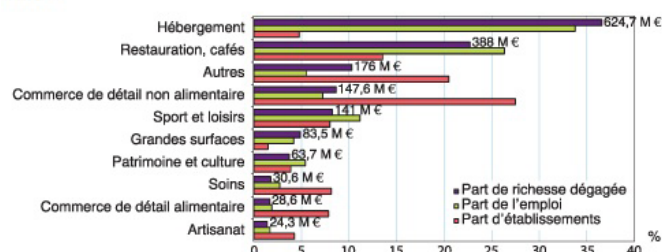
Les diagrammes en bâtons sont efficaces pour comparer des données entre elles. Ils sont par ailleurs faciles à lire. Il existe différents types de diagrammes en bâtons : en bâtons simples, en bâtons groupés ou en bâtons empilés.

► Les diagrammes en bâtons simples



► Les diagrammes en bâtons groupés

Répartition des établissements, des emplois et de la richesse dégagée touristique par secteur d'activité en Aquitaine en 2011

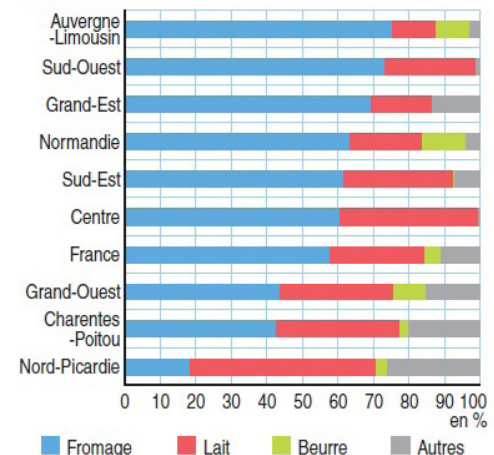


► Les diagrammes en bâtons empilés

L'utilisation de ces graphiques est délicate car ils ne sont pas toujours lisibles. Le cas suivant est un exemple réussi de diagramme en bâtons empilés. Il permet de comparer facilement la répartition des salariés selon l'activité. On peut ajouter les chiffres des effectifs pour ne pas perdre l'information en volume. Visuellement, on peut surtout comparer les différences entre les modalités des deux extrêmes. Pour les modalités intermédiaires, il est préférable d'avoir des valeurs assez différentes, ce qui est le cas ici. Dans le cas contraire, il est préférable de privilégier l'histogramme classique. Comme toujours, il faut choisir le mode de représentation qui fait le mieux passer le message que l'on veut transmettre.

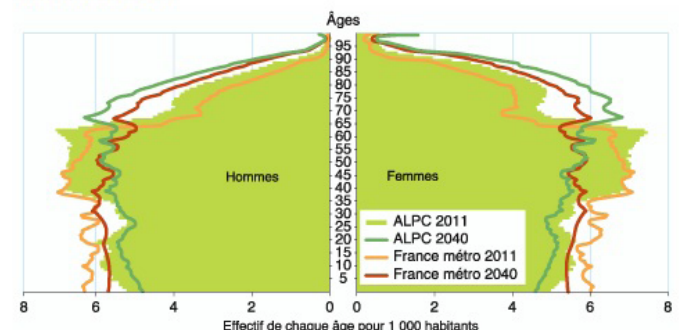
Le bassin Auvergne-Limousin très spécialisé dans la production de fromages

Répartition des salariés de l'industrie laitière selon l'activité en 2012

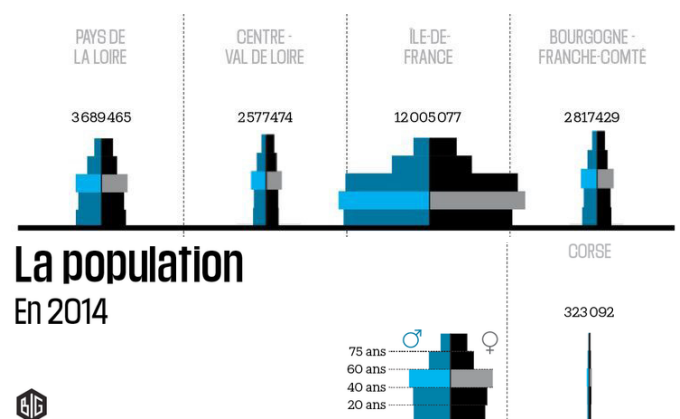


La pyramide des âges est un exemple particulier mais assez classique de diagramme en bâtons. Il est possible de superposer des courbes, ce qui permet de comparer une pyramide particulière avec une pyramide de référence. Dans ce cas-là, on représente des parts de population plutôt que des effectifs.

Pyramide des âges de l'ALPC et de la France métropolitaine en 2011 et en 2040



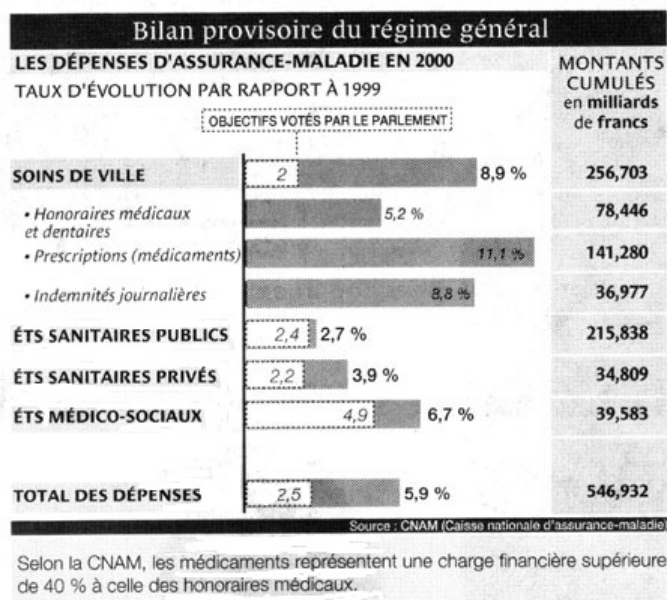
Il est aussi possible de représenter des effectifs en pyramide, ce qui permet de comparer à la fois les volumes et les profils de pyramides.



Un exemple de la revue Pénombre d'octobre 2001 montre comment mieux représenter plusieurs informations, avec à la fois des pourcentages d'évolution et des volumes. Le graphique de gauche vient de la presse et ne permet pas de rendre compte des masses en présence. Le graphique de droite, plus percutant, montre à la fois les évolutions et les volumes correspondants. Ce type de graphique n'est pas courant car il n'est pas disponible en standard dans les logiciels de bureautique.



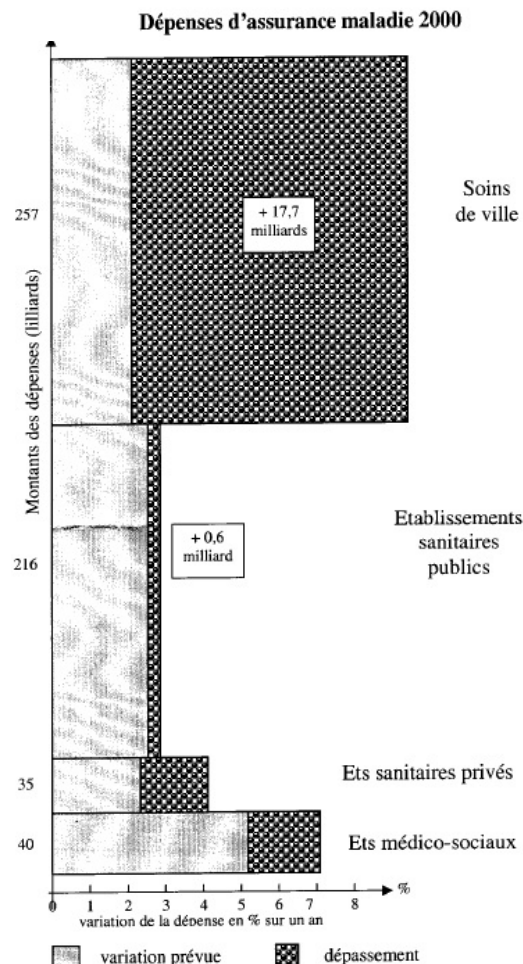
Exemple



Source : CNAM (Caisse nationale d'assurance-maladie)



Proposition alternative

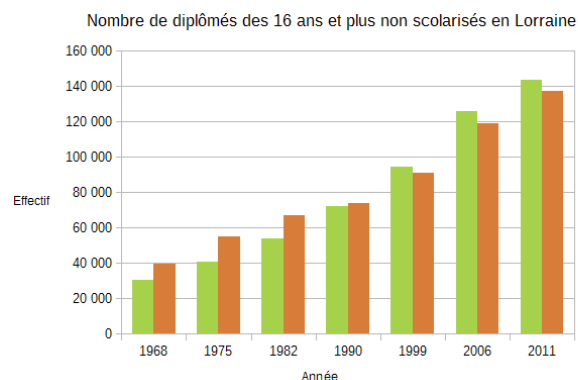


Source : CNAM

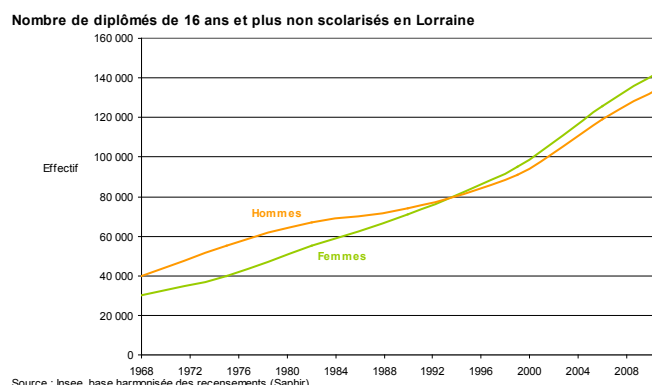
2 Courbes

Les courbes sont faciles à lire et permettent la représentation de nombreuses modalités. Elles sont bien adaptées aux données temporelles.

Dans l'exemple suivant, les mêmes données sont représentées en bâtons et en courbes. Les bâtons sont efficaces pour comparer des données entre elles. Mais les courbes montrent plus clairement les changements d'évolution.



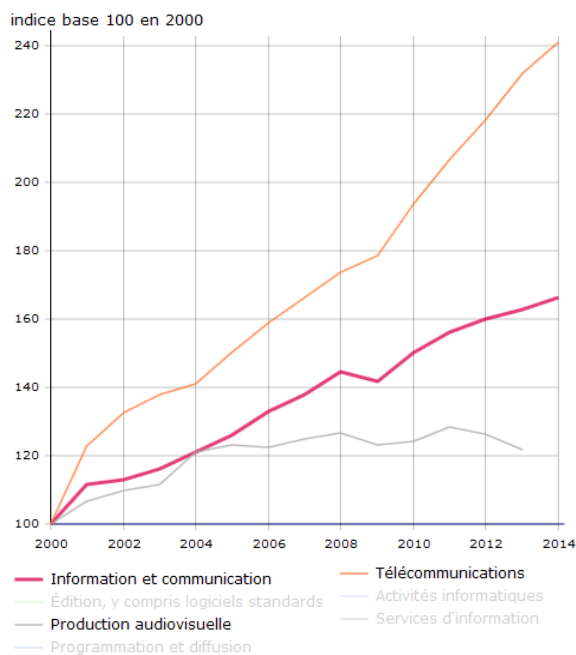
Source : Insee, base harmonisée des recensements (Saphir)



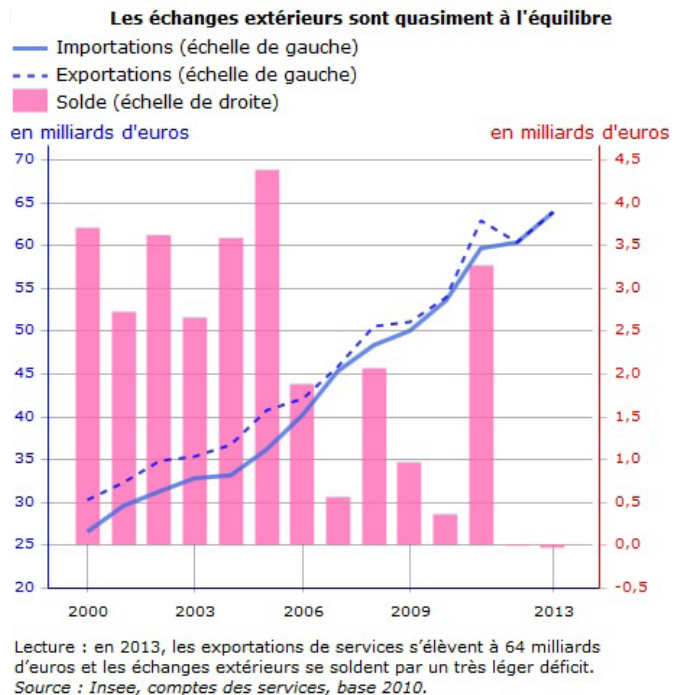
Source : Insee, base harmonisée des recensements (Saphir)

Dans le cas de plusieurs échelles, il est préférable de colorier différemment les échelles de la couleur de la courbe concernée et de faire référence à cette échelle dans l'intitulé de l'axe. Le graphique ci-contre (à droite) est un exemple de superposition d'un diagramme en bâtons et de courbes avec des échelles différentes. L'échelle de gauche de couleur bleu correspond aux courbes bleues « Importations » et « Exportations ». L'échelle de droite de couleur rouge correspond à l'histogramme rose « Solde » (la couleur rouge aurait été un meilleur choix). Les échelles sont également référencées dans la légende.

Évolution de la production par branche en volume



Champ : France, branches de l'information et de la communication.
Source : Insee, comptabilité nationale.



L'interactivité des graphiques au format SVG dynamique permet à l'internaute de ne retenir que quelques courbes (3 courbes sur 7 possibles dans le graphique ci-contre à gauche) ou d'afficher les valeurs en passant la souris dessus.

3 Graphiques en secteurs



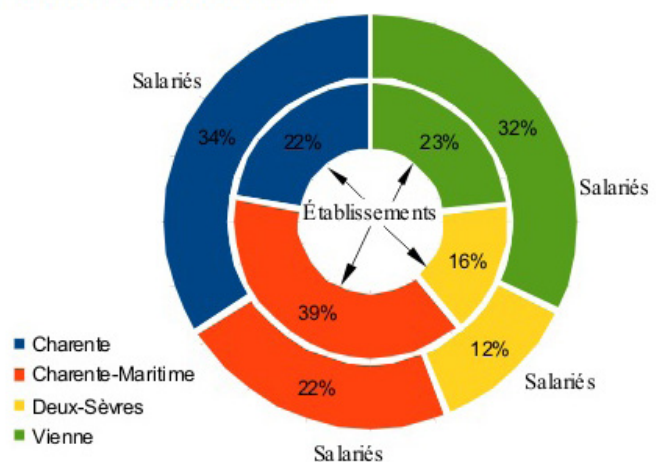
Exemple

Pour les graphiques en secteurs, une tendance récente consiste à représenter deux anneaux concentriques.

Or il est difficile de comparer les valeurs des deux anneaux. Dans l'exemple suivant, la valeur 22 % en rouge paraît aussi grande que la valeur 39 % et le 12 % en jaune paraît plus grand que le 16 %.

Lecture : la couronne centrale représente le poids des établissements employeurs dans la région par département. La couronne extérieure représente le poids des salariés (en ETP) des départements en Poitou-Charentes. À titre d'exemple, en Charente-Maritime, les salariés de la filière représentent 22 % des salariés picto-charentais de la filière alors que, dans ce même département, les établissements liés à la filière pèsent 39 % dans l'ensemble des établissements de la filière régionale.

Répartition des salariés et des établissements employeurs de la filière Image par département



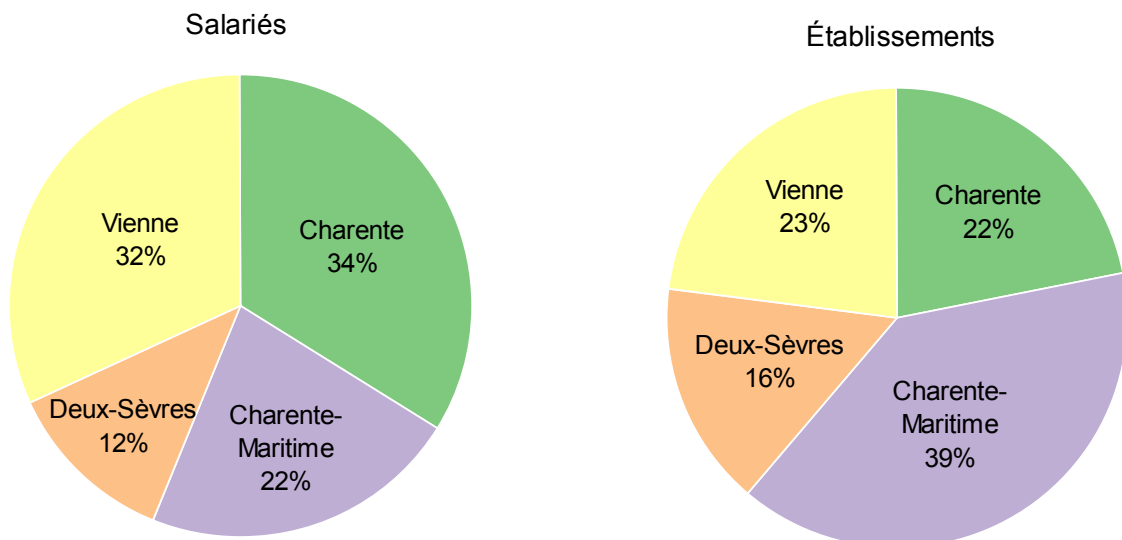
Source : Insee, Clap 2012



Proposition alternative

Une proposition alternative est de représenter deux graphiques en secteurs. Ils ont l'avantage d'être connus du grand public et de pouvoir représenter un grand nombre de modalités. En revanche, ils sont pauvres en information et en fiabilité de représentation car la perception visuelle des surfaces peut être biaisée. Enfin, ces diagrammes n'ont pas d'intérêt quand les profils ne sont pas marqués (surfaces voisines).

Répartition des salariés et des établissements employeurs de la filière par département



Source : Insee, Clap 2012.

4 Nuages de points

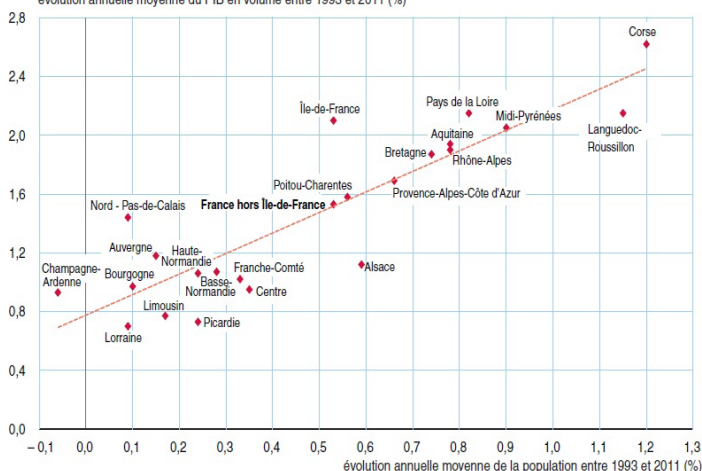
Les nuages de points sont des graphiques permettant de montrer des liens et des corrélations. Ces graphiques, davantage utilisés à l'heure actuelle, délivrent rapidement un message, à la condition qu'il soit simple.



Propositions

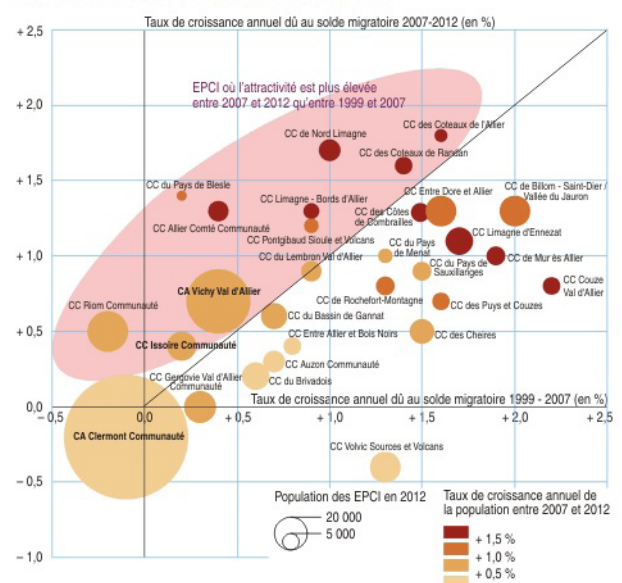
La croissance du PIB est corrélée à la croissance de la population

évolution annuelle moyenne du PIB en volume entre 1993 et 2011 (%)



On peut compléter ces nuages de points en faisant des ronds proportionnels aux effectifs ou en faisant des regroupements. Cependant, le risque est de rendre peu compréhensible le graphique, en accumulant trop d'information, comme dans le graphique de droite.

Nette progression de l'attractivité pour les EPCI situés entre les territoires de Vichy Val d'Allier et Clermont communauté - Population des EPCI sur l'axe Vichy-Brioude et taux de croissance dû au solde migratoire entre 1999-2007 et 2007-2012

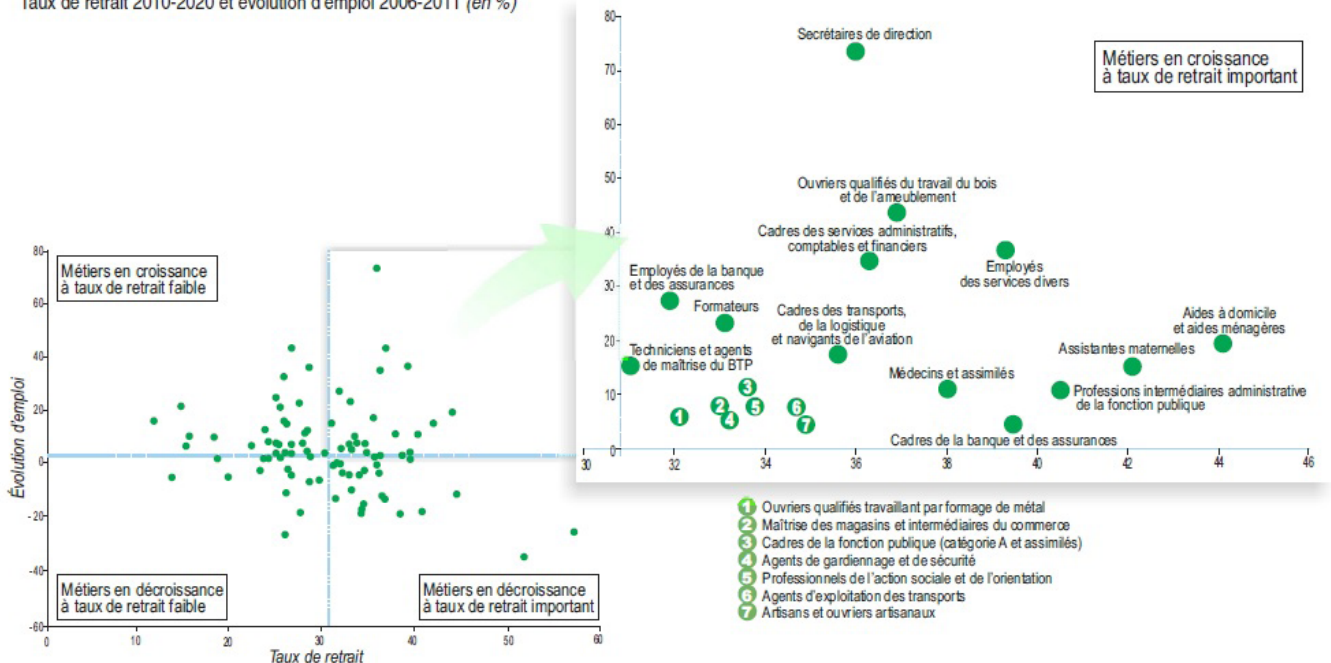


Sources : Insee, Recensements de la population 1999, 2007 et 2012.

On peut également zoomer sur une partie du nuage. Dans l'exemple suivant, on cible les métiers en croissance à taux de retraite important.

Un enjeu de renouvellement pour une vingtaine de métiers

Taux de retrait 2010-2020 et évolution d'emploi 2006-2011 (en %)



Source : Insee, recensements de la population et modèle de micro simulation

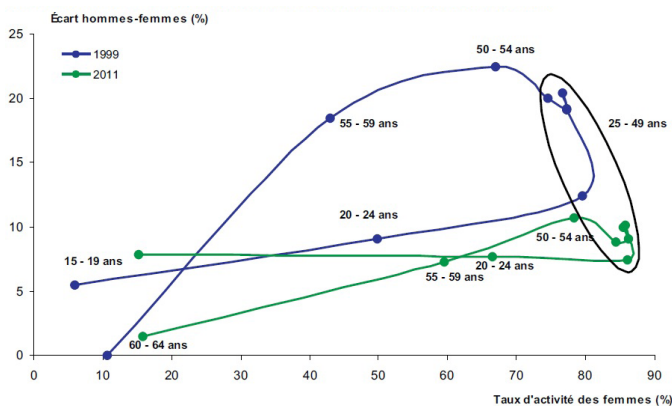
Parfois les points sont reliés entre eux. Dans l'exemple ci-dessous, les points sont reliés entre eux selon la tranche d'âge croissante, ce qui rend la lecture du graphique difficile.



► Exemple

Un écart de taux d'activité hommes-femmes en baisse

Taux d'activité des femmes et écart avec le taux d'activité des hommes par tranches d'âge

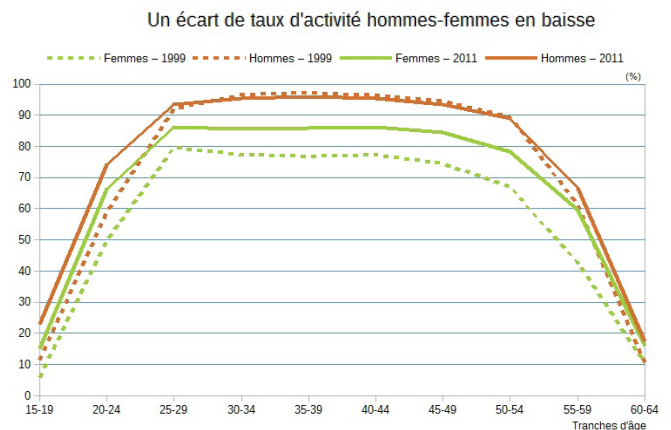


Source : Insee, recensements de la population 1999 et 2011



► Proposition alternative

Une proposition alternative est de représenter les taux d'activité par âge des hommes et des femmes. L'écart entre les hommes et les femmes est alors visualisé par les différences entre les deux courbes.



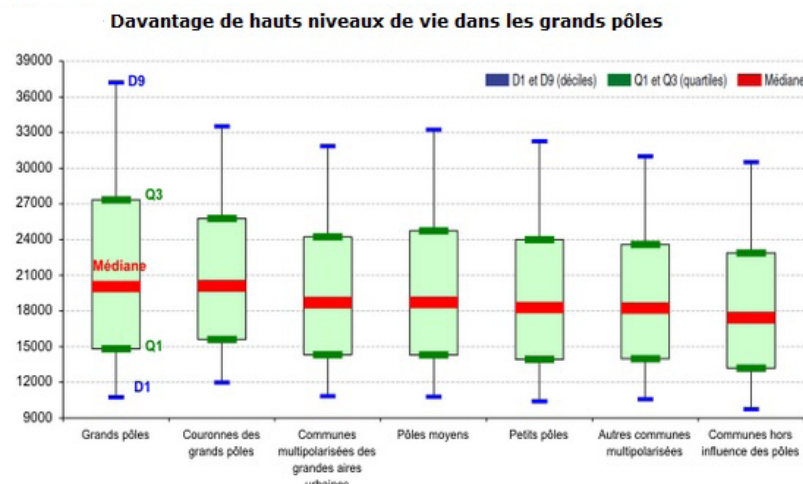
Source : Insee, recensements de la population 1999 et 2011

5 Box plot de Tukey ou boîte à moustaches



Exemple

Ces graphiques sont utiles à l'analyse mais constituent plus des graphiques d'études que des graphiques de diffusion. En effet, ils nécessitent quelques bases en statistiques pour pouvoir les interpréter ou une bonne note de lecture.



Fractiles du niveau de vie annuel en Aquitaine - Limousin - Poitou-Charentes (ALPC), par territoire du zonage en aires urbaines 2010, en euros

Lecture : 10 % des personnes des ménages fiscaux ordinaires des grands pôles de l'ALPC ont un niveau de vie annuel supérieur à 37200 euros (D9) ; 75 % un niveau de vie inférieur à 27350 euros (Q3).

Note : les couronnes des moyens et petits pôles, qui n'hébergent que 1,3 % de la population fiscale de l'ALPC, ne sont pas figurées ici.

Champ : population des ménages fiscaux ordinaires de l'ALPC

Source : Insee-DGFiP-Cnaf-Cnav-CCMSA, Fichier localisé social et fiscal 2012

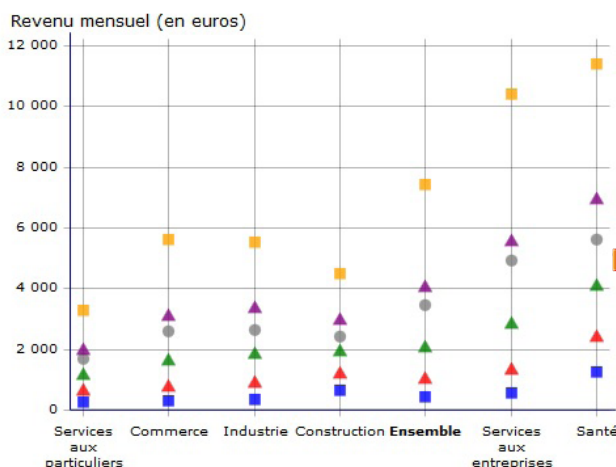
Plus facile à lire sous la forme ci-dessous ..mais c'est encore mieux en changeant les couleurs.



Exemple

Dispersion des revenus d'activité non salariaux en 2012 (hors auto-entrepreneurs) selon le secteur d'activité

■ D1 ▲ Q1 ▲ médiane
● moyenne ▲ Q3 ■ D9



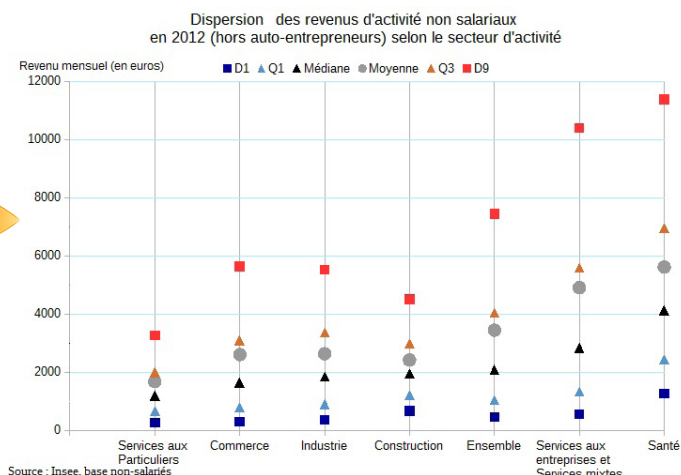
Lecture : dans la construction, le revenu moyen d'activité des non-salariés (hors auto-entrepreneurs et hors revenus nuls) s'élève en 2012 à 2 410 euros par mois. Un non-salarié de la construction sur dix perçoit moins de 660 euros mensuels (D1), un quart gagne moins de 1 190 euros (Q1) et la moitié moins de 1 940 euros (médiane) ; en haut de l'échelle, un quart perçoit plus de 2 970 euros (Q3) et un sur dix gagne plus de 4 500 euros (D9).

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2012, hors agriculture, hors auto-entrepreneurs, hors revenus nuls et hors taxés d'office (définitions).

Source : Insee, base non-salariés.



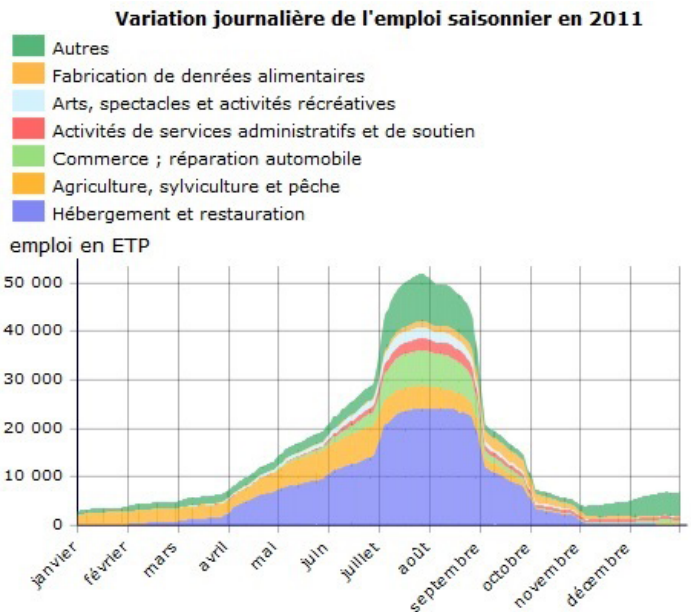
Proposition alternative



Source : Insee, base non-salariés

6 Aires empilées

Les graphiques en aires empilées sont bien adaptés pour des populations complémentaires et pour les dimensions temporelles. Cependant, il faut éviter « l'effet montagne » en représentant un trop grand nombre de variables. Le risque est de considérer que les courbes sont les unes derrière les autres et non un empilement de couches (effet cumulé).



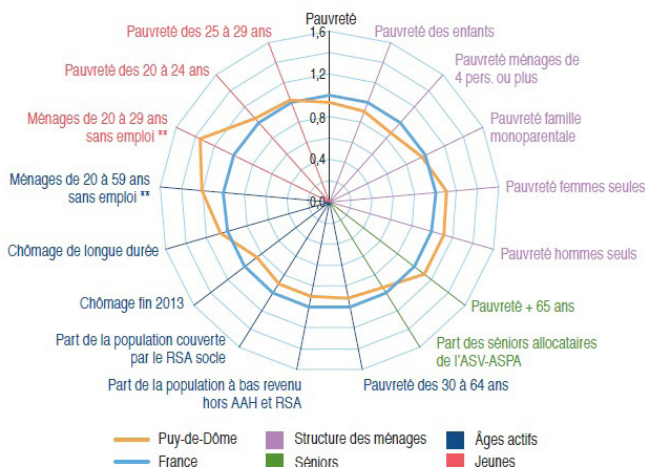
7 Radars



Exemple

Le graphique en radars permet facilement de comparer des profils différents sur plusieurs variables. Ils sont souvent utilisés à l'Insee pour représenter des statistiques multivariées, mais parfois avec quelques défauts : trop de modalités (axes) représentées, trop de courbes, trop de niveaux.

Une précarité des jeunes plus marquée que la moyenne métropolitaine
Position relative* du Puy-de-Dôme par rapport à la moyenne française



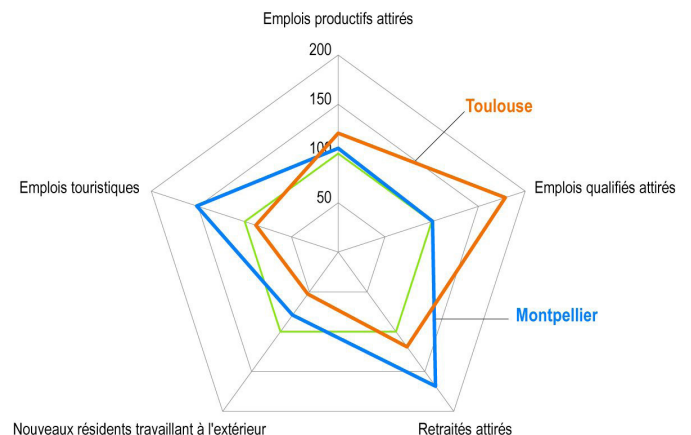
* Les taux sont plus marqués dans le département qu'en moyenne nationale pour les valeurs au-dessus de 1, et inversement pour les valeurs inférieures.
** Part des ménages dont aucun membre n'a d'emploi selon l'âge de la personne de référence.
Sources : Insee, RDL 2011, RP 2011, Taux de chômage localisés ; Pôle emploi ; CNAF ; MSA.



Proposition alternative

Pour que ce type de graphique puisse délivrer un message simple rapidement lisible, il faut que les profils présents sur le radar soient très différents. Certains choix de couleurs peuvent faciliter la lecture, comme dans le graphique ci-dessous.

Toulouse attire particulièrement les actifs qualifiés et les emplois productifs
Indicateurs mesurant la capacité des grandes zones d'emploi à attirer les ressources extérieures en 2013

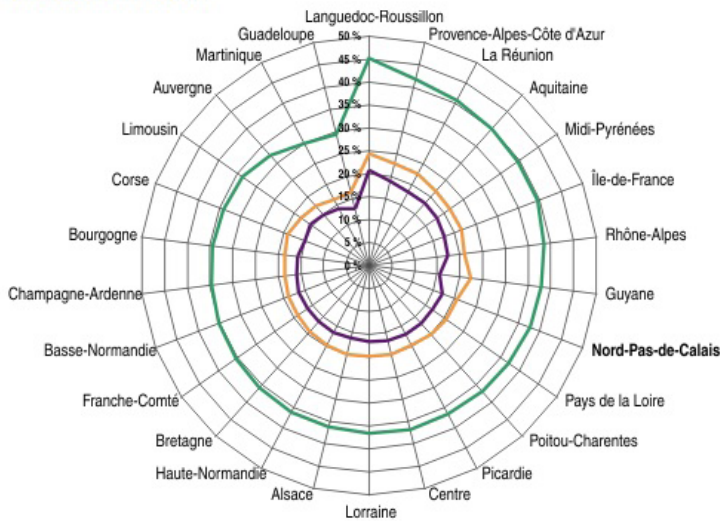


Note : les indicateurs sont normés à 100, le trait vert représente la moyenne des 16 zones d'emploi de plus de 230 000 emplois.
Lecture : le taux d'arrivées d'emplois qualifiés à Toulouse est à 179, supérieur de 79 % à la moyenne du groupe des grandes zones d'emploi.
Sources : Insee, recensements de la population 2008 et 2013, Sirene, Clap, Lifi 2011 à 2013, DADS, Acoess 2013.



Exemple

Plus d'un tiers du stock d'établissements se renouvelle chaque année dans la région



— Taux de sortie moyen — Taux d'entrée moyen — Taux de renouvellement moyen

Note : Les taux d'entrée et de sortie sont mesurés en ramenant les entrées et les sorties au nombre d'établissements au 1er janvier de l'année.

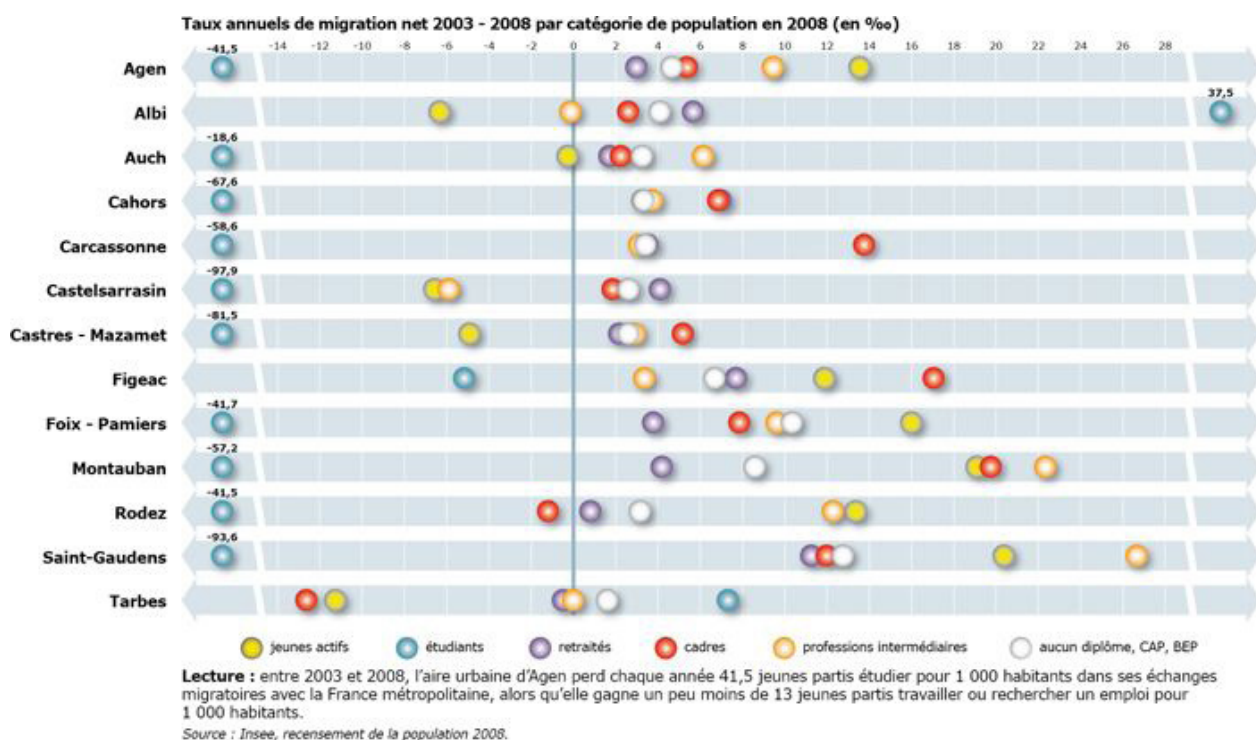
Source : Insee, REE et Clap.

Une tendance récente dans les publications régionales de l'Insee consiste à utiliser les radars pour représenter les régions françaises en les classant. L'intérêt du radar qui permet de voir rapidement des profils différents est alors perdu. Dans le graphique ci-contre, il est difficile de comparer les valeurs des régions non adjacentes sur la spirale.

8 Graphiques moins classiques

Graphiques sur une ligne

Souvent faciles à construire et à lire, les graphiques en ligne sont plutôt efficaces.

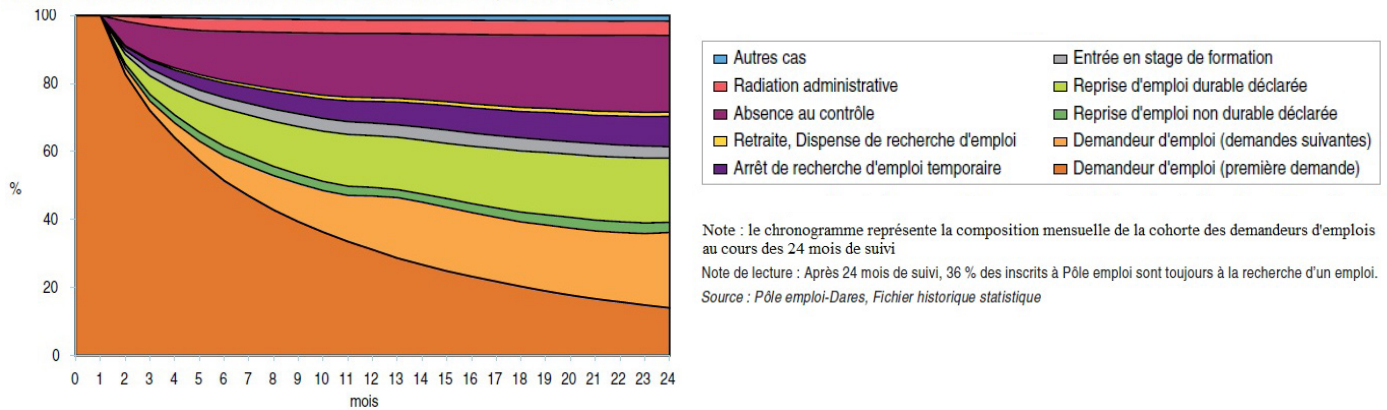


► Chronogramme

Ci-dessous la représentation graphique de l'évolution de la situation d'une cohorte de chômeurs pendant 24 mois.

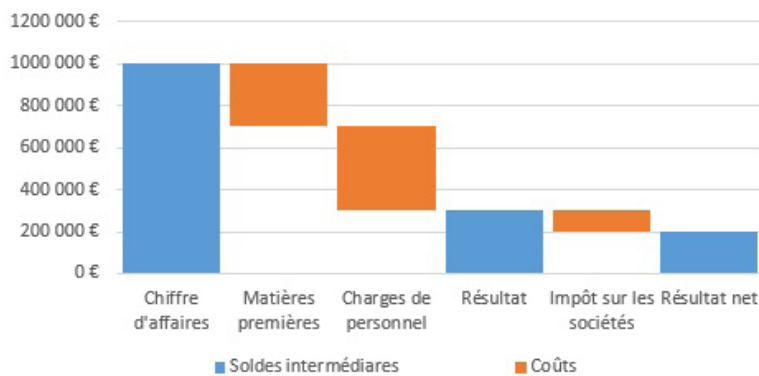
Au bout de 24 mois de suivi, 22 % des inscrits à Pôle emploi sont en situation de reprise d'emploi

Évolution de la situation des chômeurs dans les 24 mois suivant l'inscription à Pôle emploi



► Histogramme en cascade

Calcul du bénéfice net

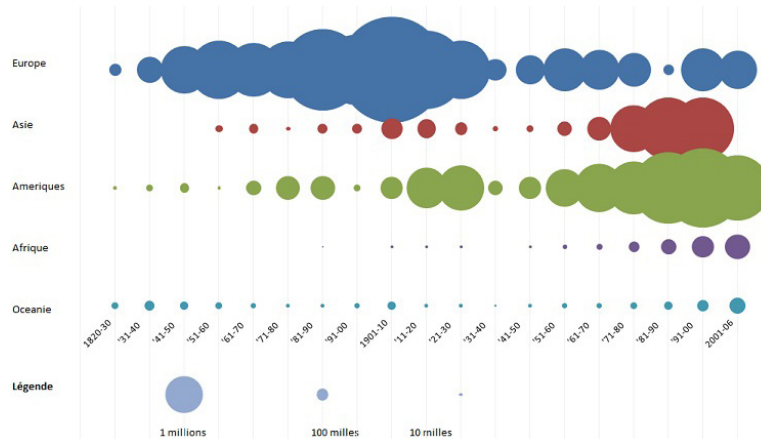


Source : Entreprise X

Ce graphique permet de montrer les étapes ou les différents facteurs permettant de passer d'une valeur à une autre.

► Bulles temporelles

L'immigration aux Etats-Unis au 19ème et 20ème siècle par continent d'origine

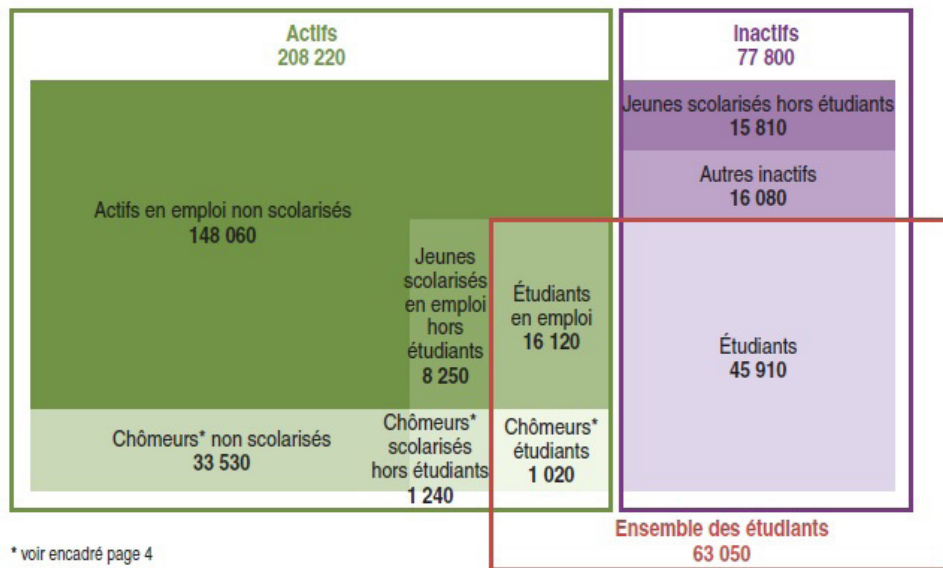


Source : Statistical Yearbooks of the Immigration and Naturalization Service (mise à jour depuis 2006 via INS Annual Reports)

Treemap

Sur ce Treemap, la surface des rectangles est proportionnelle aux effectifs des sous-populations.

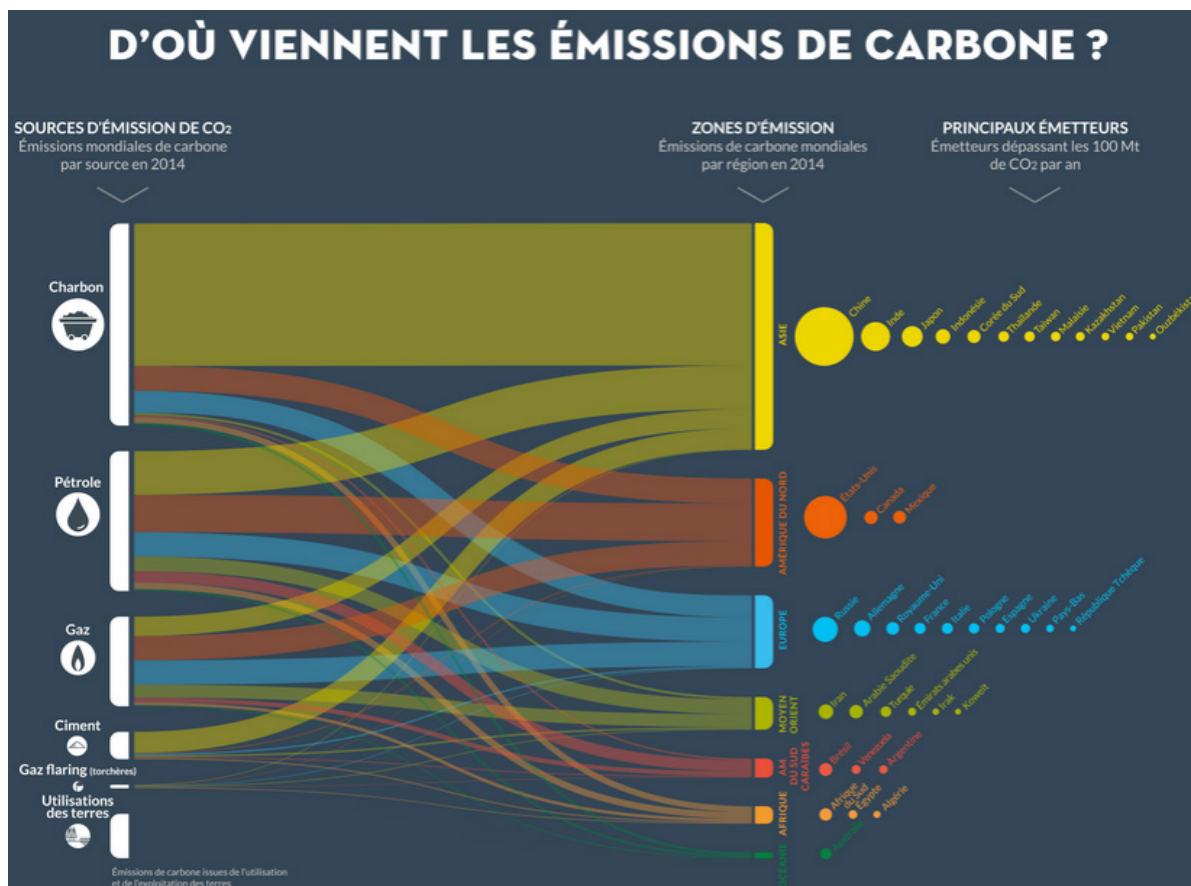
286 000 Alsaciens âgés de 18 à 29 ans



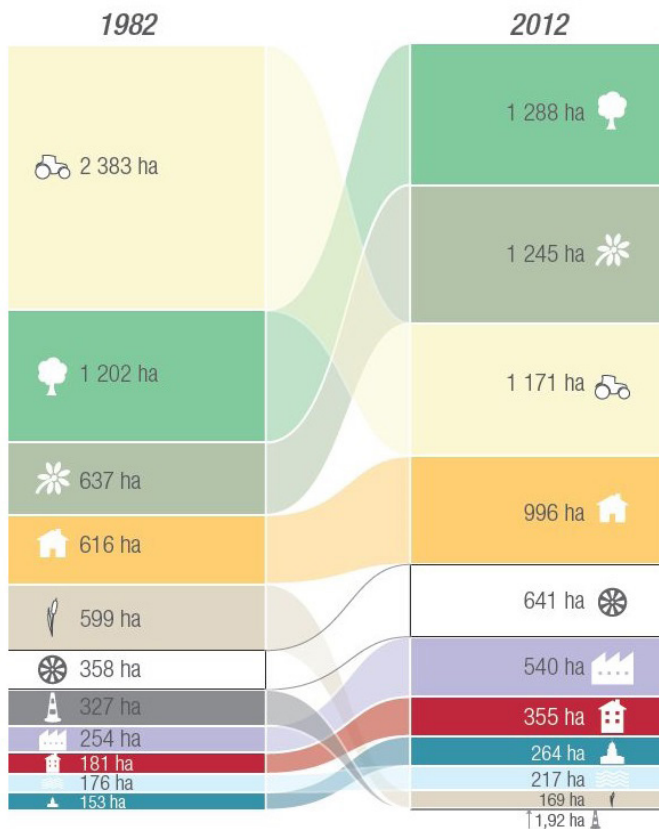
Source : Insee, RP2007, exploitation complémentaire au lieu de résidence

Histogramme avec relations entre les barres

Ce type de graphique permet de mettre en relation plusieurs variables et de raconter une histoire.

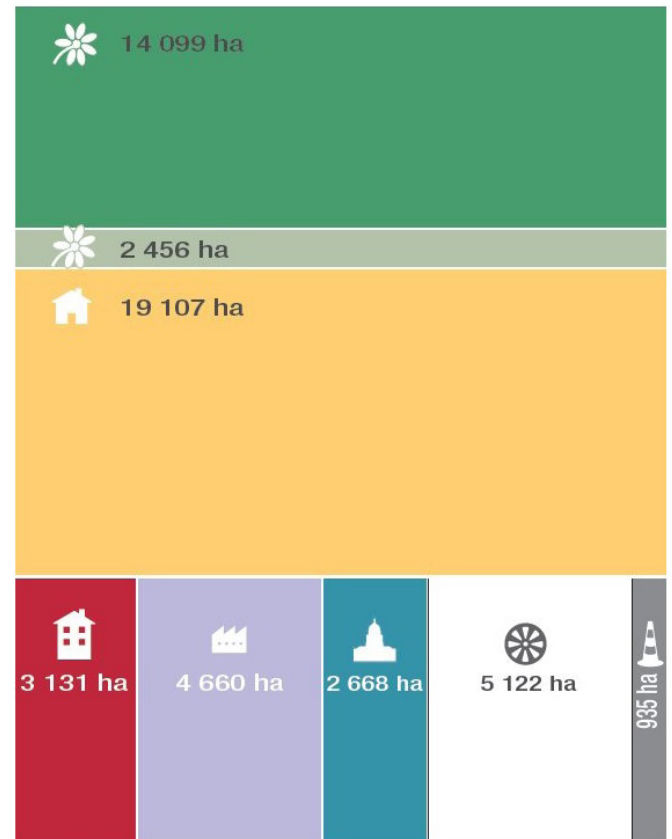


Composition du sol de Saint-Quentin-en-Yvelines entre 1982 et 2012



Source : Mos 1982 à 2012, IAU îdF

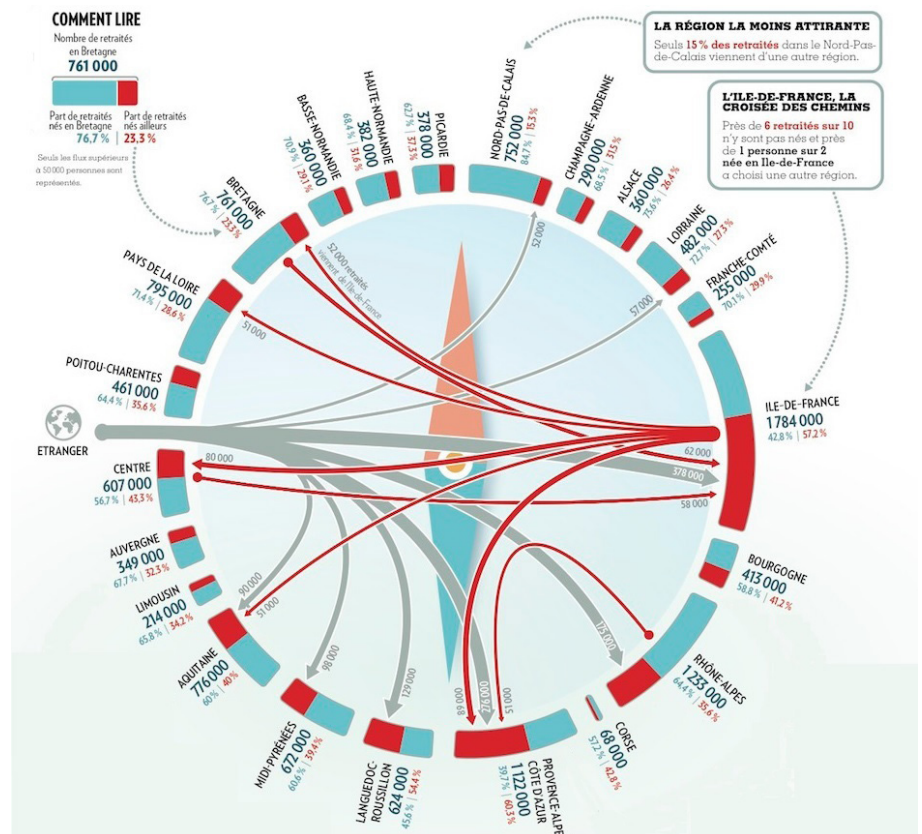
Composition des espaces urbains yvelinois en 2012



Source : Mos 1982 à 2012, IAU îdF

Chord diagram, Circos

Le Circos a été popularisé pour le séquençage du génome humain. Il permet de visualiser les relations entre différentes parties. Il peut être intéressant comme graphique d'études, à condition de ne pas faire plusieurs cercles concentriques et de limiter le nombre de modalités. Ce type de graphique gagne à être présenté en mode interactif (<http://www.global-migration.info/>).



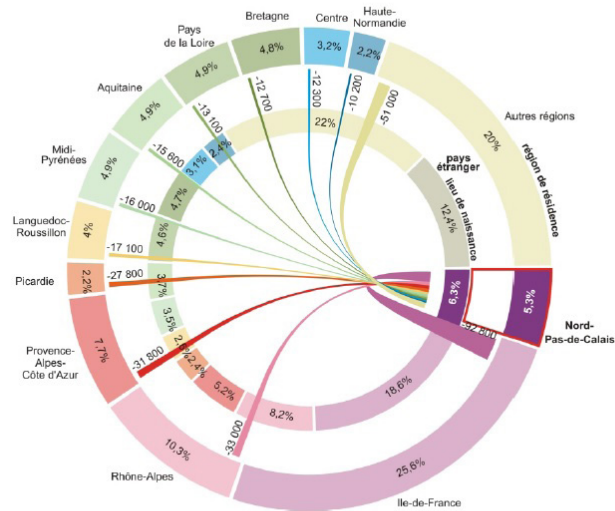


Exemples

Dans l'exemple suivant à gauche, le graphique est complexifié par deux cercles concentriques, de rayon et de hauteur différents, qui sont difficiles à comparer, et par un nombre important de couleurs.

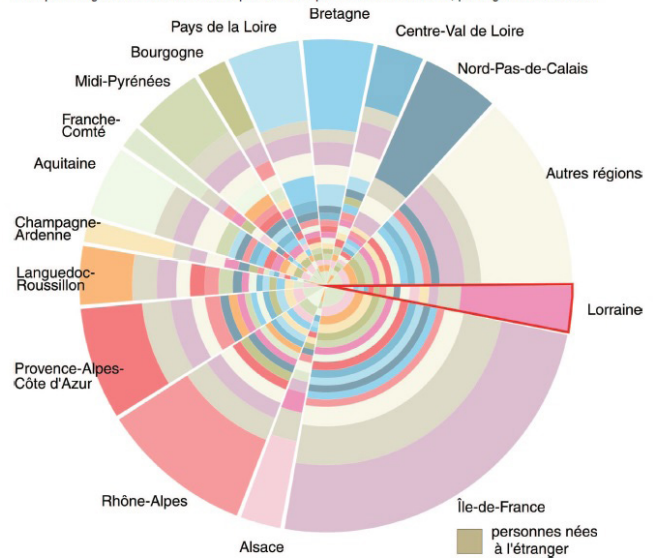
Le graphique de droite pose le problème de comparaison entre des formes et des tailles différentes. Les secteurs ne présentent à la fois ni le même angle ni la même taille. Or, il est difficile de comparer et hiérarchiser des valeurs en fonction de surfaces de formes différentes. De plus, trop d'informations sont présentes dans ce graphique, avec beaucoup de couleurs différentes.

334 000 natifs du Nord-Pas-de-Calais aujourd'hui dans d'autres régions
Régions de naissance et de résidence pour les diplômés du supérieur et les adultes en cours d'études



Guide de lecture : le cercle interne représente la décomposition en fonction de la région de naissance, y compris les pays étrangers. Le cercle externe représente la décomposition en fonction de la région actuelle de résidence, uniquement en France. Les arcs symbolisent les mobilités qui ont pris place depuis le Nord-Pas-de-Calais comme région de naissance vers d'autres territoires pris comme régions de résidence. Par exemple, 33 000 natifs du Nord-Pas-de-Calais, diplômés du supérieur ou en cours d'étude, résident aujourd'hui en région Rhône-Alpes.
Source : Insee, recensement de la population 2011.

Sept résidents lorrains diplômés sur dix sont nés en Lorraine
Principales régions de naissance des diplômés du supérieur et des étudiants, par région de résidence



Lecture : chaque secteur de disque symbolise une région de résidence et se trouve décomposé en bande de couleurs, selon la répartition de la population en fonction des lieux de naissance. La surface des bandes de couleur est proportionnelle au poids démographique de chaque région de naissance dans la région de résidence actuelle.

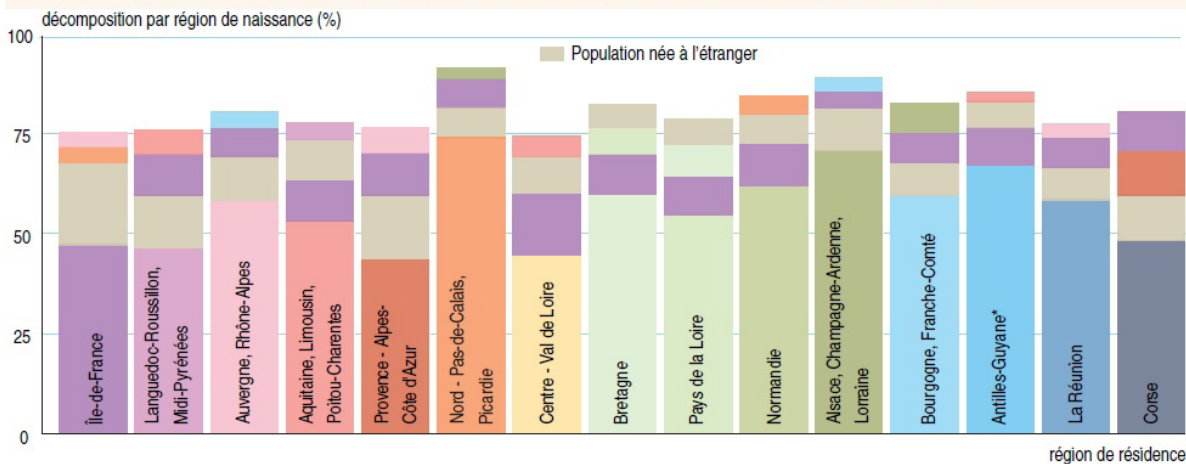
Source : Insee, recensement de la population 2012



Proposition alternative

La présence de pourcentages permettrait de pallier en partie ce problème. Un diagramme empilé avec moins d'informations serait peut-être plus efficace (voir ci-dessous).

Les quatre principales régions de naissance des diplômés du supérieur et des étudiants, par région de résidence



Lecture : chaque barre de l'histogramme retrace la décomposition par région de naissance pour une région de résidence donnée. À chaque région, de naissance ou de résidence, correspond une couleur spécifique. Seules les quatre principales régions de naissance figurent pour chaque région de résidence. Par exemple, l'Île-de-France, en violet, est la deuxième région d'origine des résidents en Bretagne.

* Pour assurer la lisibilité de l'illustration, les régions de Guadeloupe, Martinique et Guyane ont été regroupées sous l'appellation « Antilles - Guyane ».

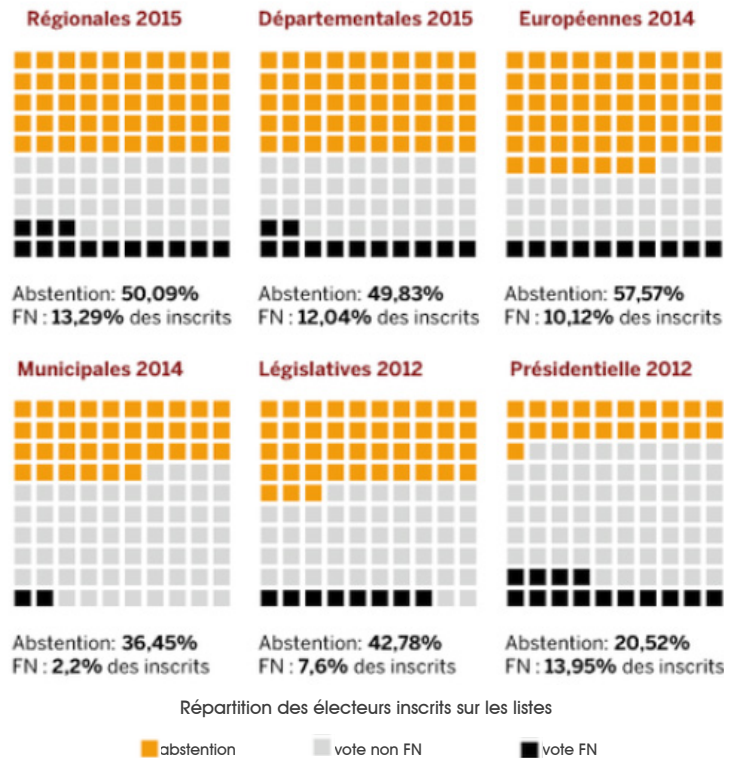
Source : Insee, recensement de la population 2012

► Matrice de points

FN et abstention, réponse en 15 élections

Ces graphiques montrent le score du Front national au premier tour de toutes les élections depuis 2007, non pas en part des voix exprimées, mais en part des inscrits (électeurs inscrits sur les listes), ce qui permet de mesurer leur score indépendamment de l'abstention, qui figure en jaune sur les graphiques

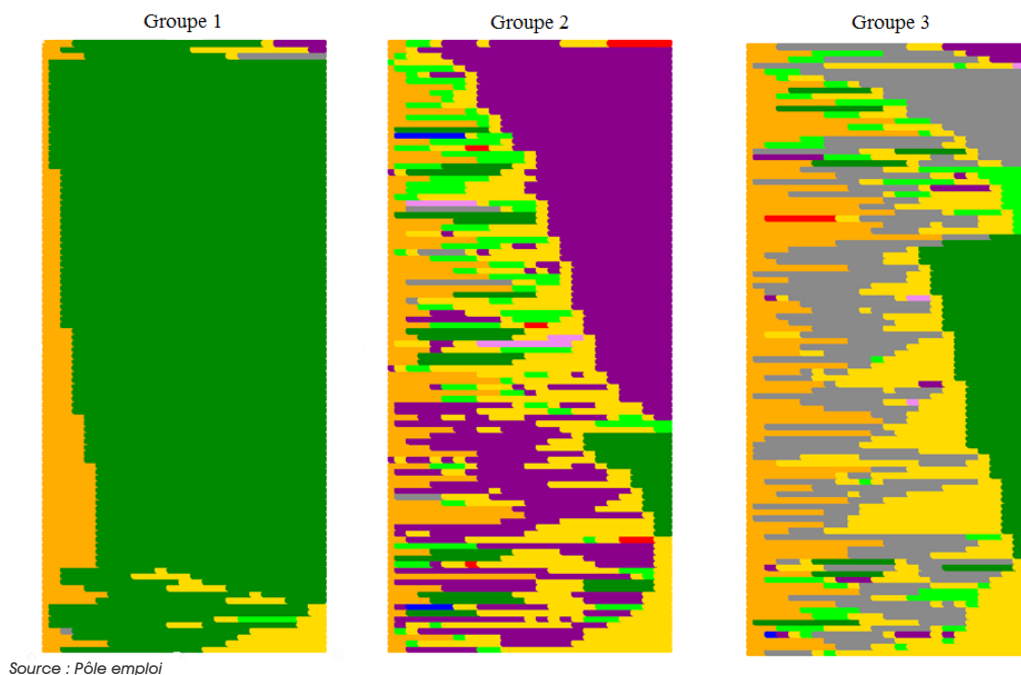
La matrice de points pourrait être une alternative à un histogramme classique. Il est en effet aisé de comparer rapidement la surface des différentes couleurs. De plus, les pourcentages sont faciles à reconstituer, chaque petit carré représentant 1 %.



► Tapis (ou index plots)

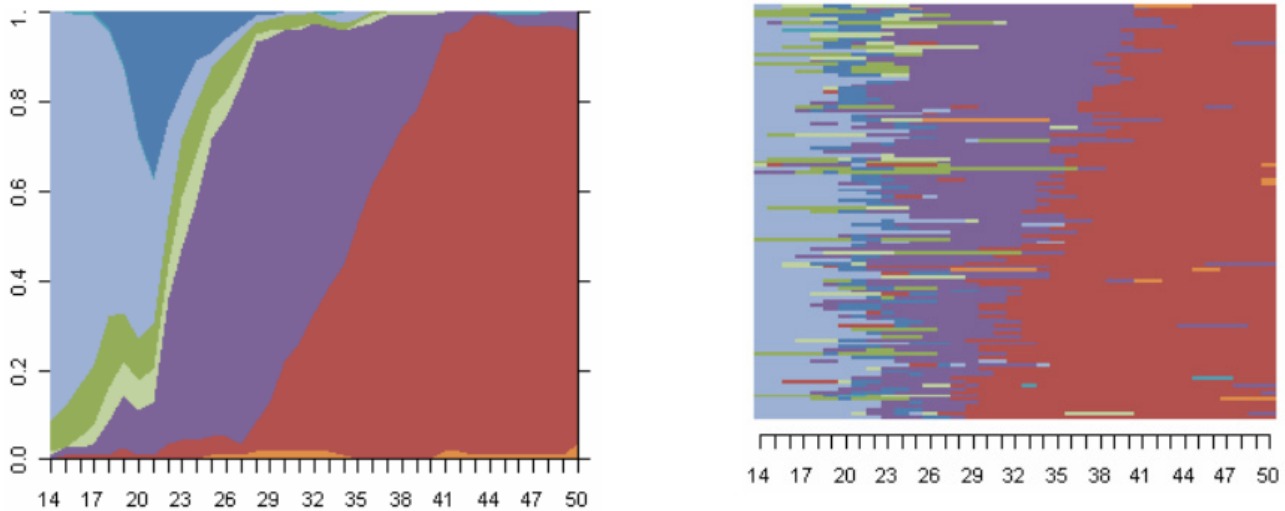
Le tapis est plus un graphique de recherche qu'un graphique de communication. Il permet de présenter des parcours individuels. Les exemples suivants sont réalisés à partir de 3 échantillons de 100 parcours individuels.

Parcours de demandeurs d'emploi



Chaque ligne représente le parcours d'un demandeur d'emploi au cours du temps en fonction de son activité représentée par une couleur (chômage, formation, stage, retour à l'emploi, ...)

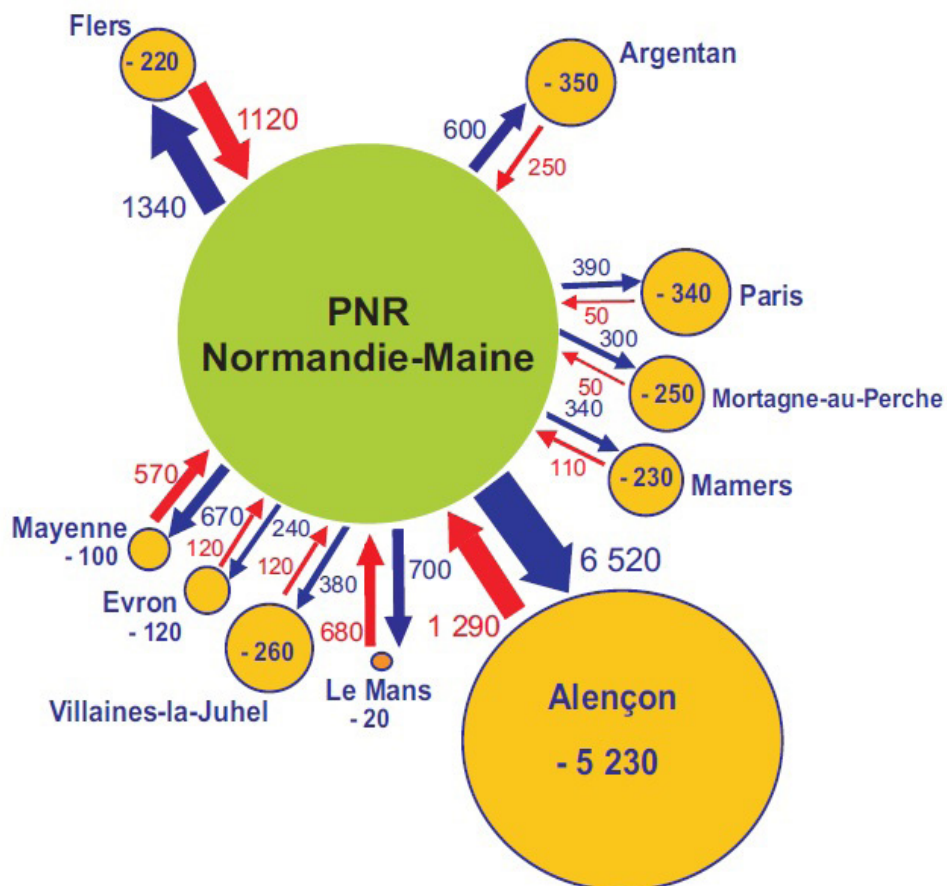
Dans l'exemple ci-dessous, on voit la différence entre un tapis où chaque individu est représenté (graphique de droite) et un graphique en aires cumulées où les individus ont été agrégés (graphique de gauche).



► Des graphiques à la place de cartes

Dans cet exemple, une carte est schématisée avec des flux et des bilans migratoires entre zones géographiques. Cela permet d'aller à l'essentiel de l'information sans souci d'exactitude spatiale.

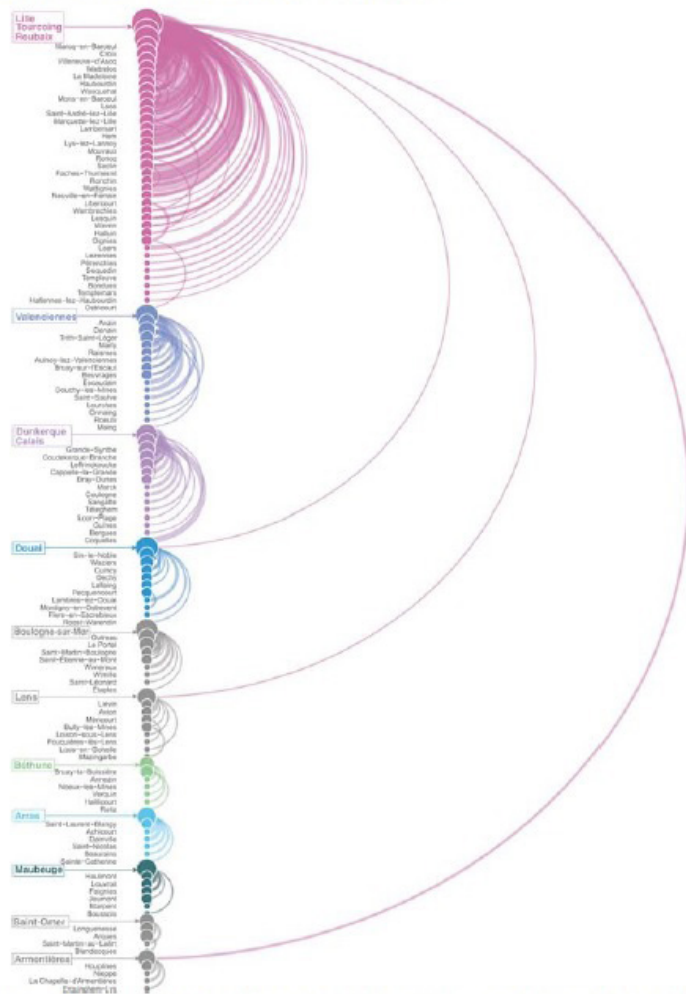
Intensité des échanges domicile-travail avec les principales aires urbaines



Source : Insee, recensement de la population 2010, exploitation principale

Arc Diagrams

Des réseaux de villes nombreux, aux côtés du réseau lillois...
Principaux réseaux de villes du Nord-Pas-de-Calais en 1975



Dans cet autre exemple, il n'y a plus de références spatiales.

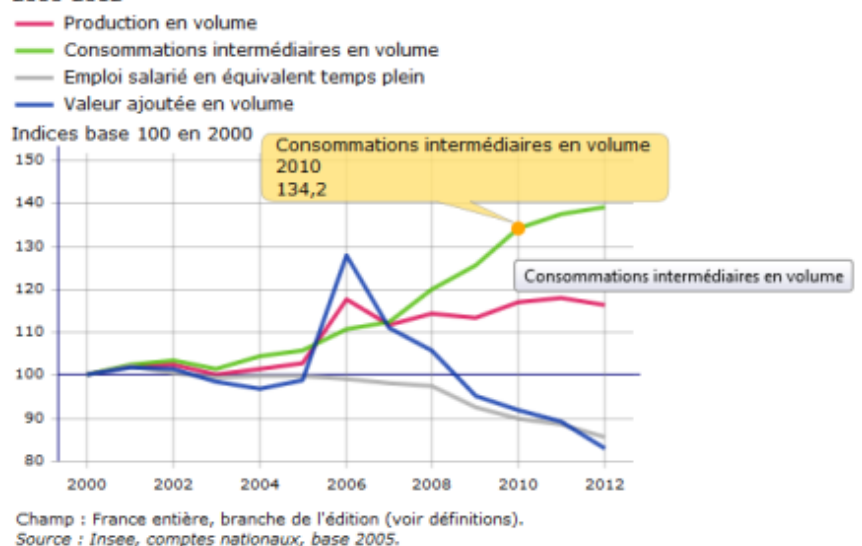
Note : Chaque couleur correspond à un réseau distinct. Les arcs de cercle représentent les liaisons d'au moins 400 navetteurs. Les communes sont symbolisées par des bulles dont la taille représente le nombre de connexions impliquant la ville dans un réseau.
Source : Insee, recensement de la population 1975

Graphiques dynamiques

La nouvelle tendance sur les sites internet est le graphique dynamique. Les mêmes règles sémiologiques doivent être respectées. Ces graphiques permettent de montrer plus d'informations. Ainsi, un graphique animé peut ajouter plus facilement une troisième dimension comme le temps à un graphique en deux dimensions.

Le graphique dynamique peut également jouer le rôle d'une mini-base de données. L'internaute promène le curseur et dévoile telle ou telle méta-information, ou telle ou telle variable supplémentaire et contextuelle. ■

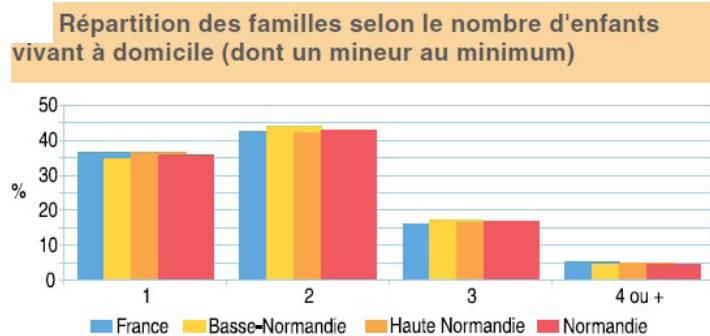
Évolution de la production, des consommations intermédiaires et des effectifs de la branche de l'édition - 2000-2012



QUELQUES PIÈGES À ÉVITER

① Des surfaces qui se chevauchent

Dans le diagramme en bâtons, on compare des surfaces. Ces surfaces doivent apparaître dans leur intégralité de façon à éviter de fausser la perception que l'on en a.



Source : Insee, enquête "La famille et les logements" 2011

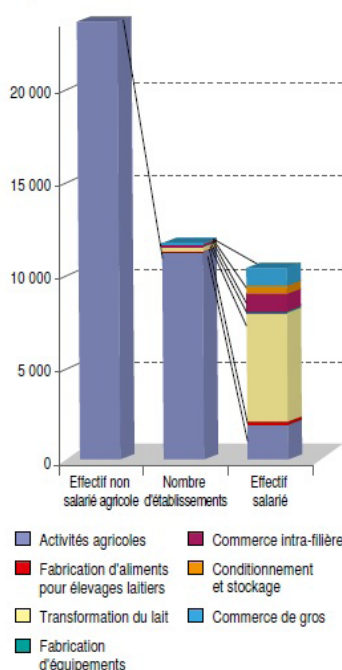
② Le mirage de la 3D

Les effets tridimensionnels faussent la perception visuelle. Selon l'angle de vue, certaines parties d'un camembert en 3D paraissent plus importantes que d'autres en raison de l'effet de volume. Dans le diagramme à barre, ce même effet volume avantage les barres de droites.



Exemples

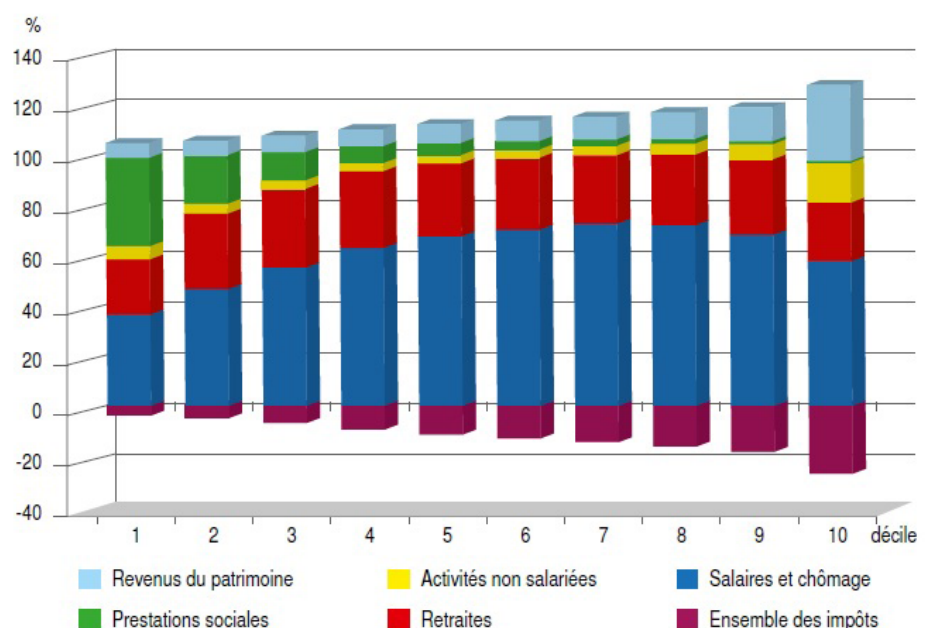
En marge du cœur de filière, peu d'établissements et peu d'effectifs salariés
Répartition des établissements et des effectifs en 2010 pour les activités agricoles et au 31 décembre 2011 pour les autres segments d'activité



Source : SSP (ministère chargé de l'agriculture)

Les prestations sociales constituent le tiers des revenus des plus pauvres

Décomposition du revenu disponible avant impôts par déciles de niveau de vie en 2012

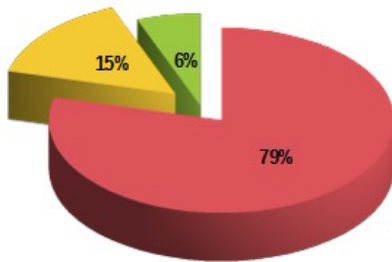


Sources : Insee-DGFiP-Cnaf-Cnav-CCMSA, Fichier localisé social et fiscal 2012



Exemple de camembert en 3D

Fréquence d'utilisation d'internet au cours des 3 derniers mois



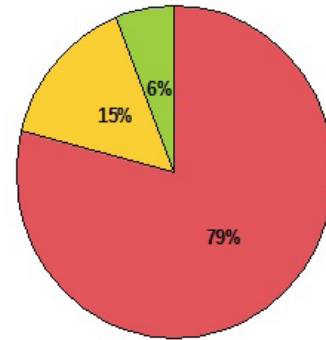
■ Tous les jours ou presque
■ au moins une fois par semaine
■ moins d'une fois par semaine

Source : Insee



Proposition alternative

Fréquence d'utilisation d'internet au cours des 3 derniers mois



■ Tous les jours ou presque
■ au moins une fois par semaine
■ moins d'une fois par semaine

Source : Insee

3 Des surfaces censées être comparables et qui ne le sont pas

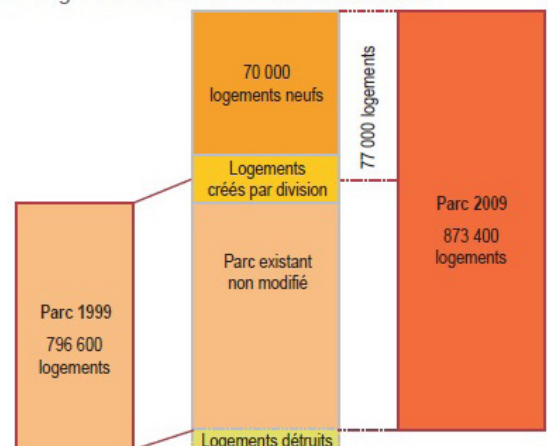


Exemple

Dans ce graphique, le parc de logements 2009 semble au moins deux fois plus gros que le parc 1999.

Le parc de logements évolue dans sa structure

Évolution du parc de logements entre 1999 et 2009 en Picardie



Source : Dreal, Filocom

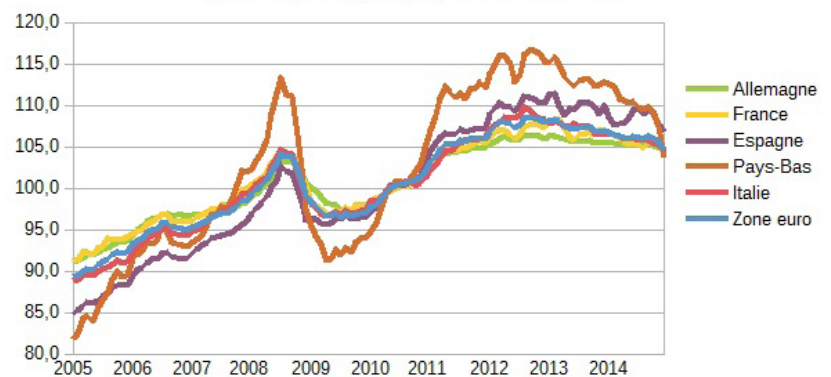
4 L'effet spaghetti



Exemple

Plusieurs petits graphiques valent parfois mieux qu'un seul gros graphique avec trop d'informations.

Évolution mensuelle des prix de production de l'industrie pour les principaux pays de la zone euro

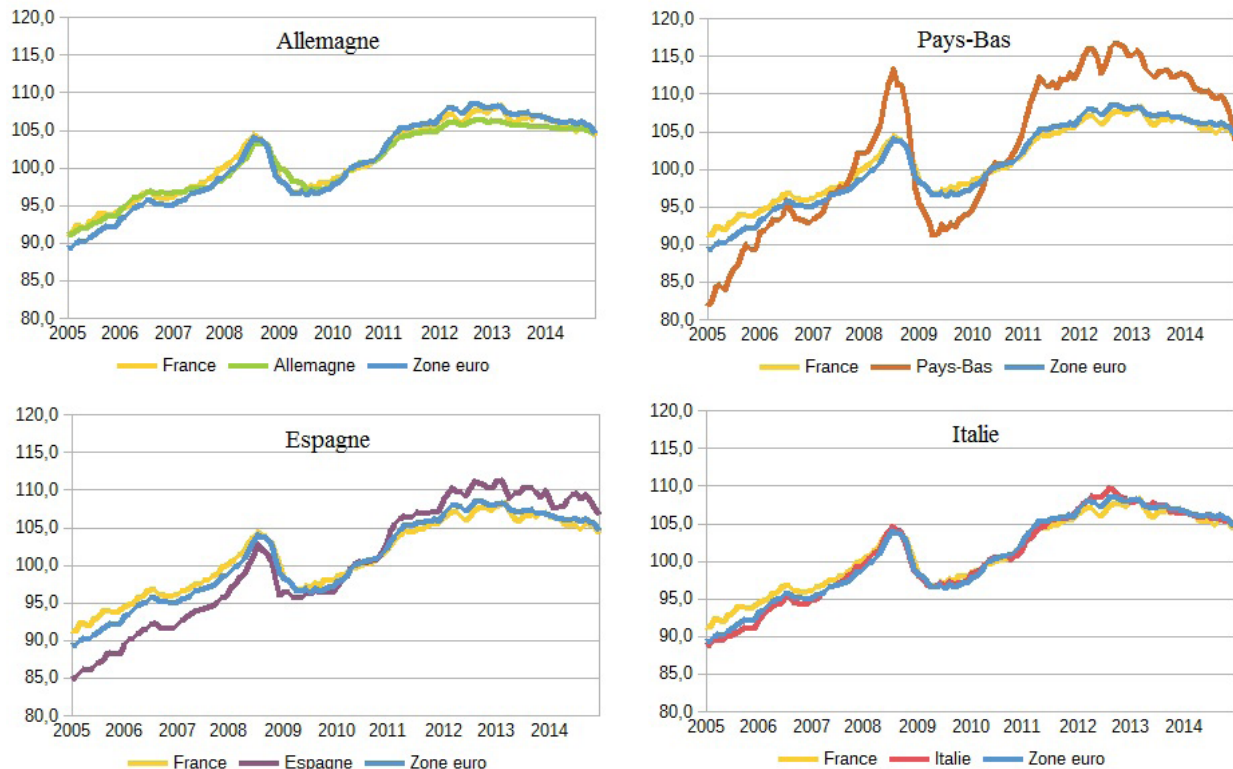


Source : Eurostat



Proposition alternative

Évolution mensuelle des prix de production de l'industrie pour les principaux pays de la zone euro



5 Le mouvement pendulaire ou le graphique torticollis

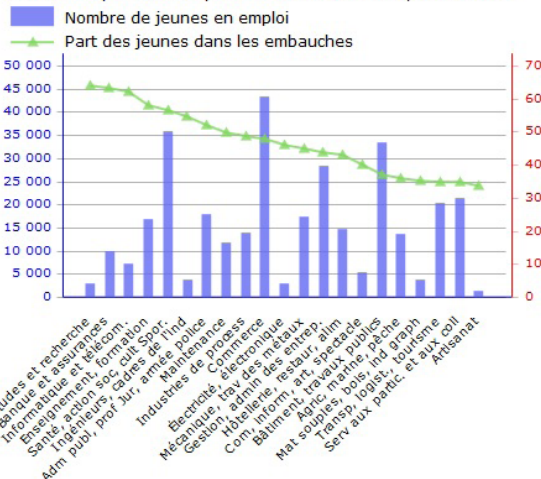
Un graphique efficace doit réduire au maximum la charge musculaire sur l'œil en lui évitant de faire le yo-yo entre diagramme, légende, libellés. Certains libellés de graphique obligent à se pencher pour les lire.



Exemple

Fortes parts d'embauche des jeunes dans les emplois qualifiés des services

Nombre de jeunes en emploi et part des jeunes (en %) dans les embauches par domaine professionnel dans les Pays de la Loire



Lecture : dans les Pays de la Loire, 43 300 jeunes ont un emploi dans le commerce. Dans ce domaine professionnel, les jeunes représentent 48 % des embauches.

Source : Insee, Recensement de la population (RP) 2009 - Enquête Emploi en Continu (EEC) 2006-2012.



Proposition alternative

Peu de reconversions possible dans les mêmes métiers



Note de lecture : l'échelle de gauche présente le nombre d'emplois du secteur des IAA en 2011. Celle de droite indique les pertes d'emplois observées entre 2006 et 2011, tant dans les IAA que dans les autres secteurs d'activité.

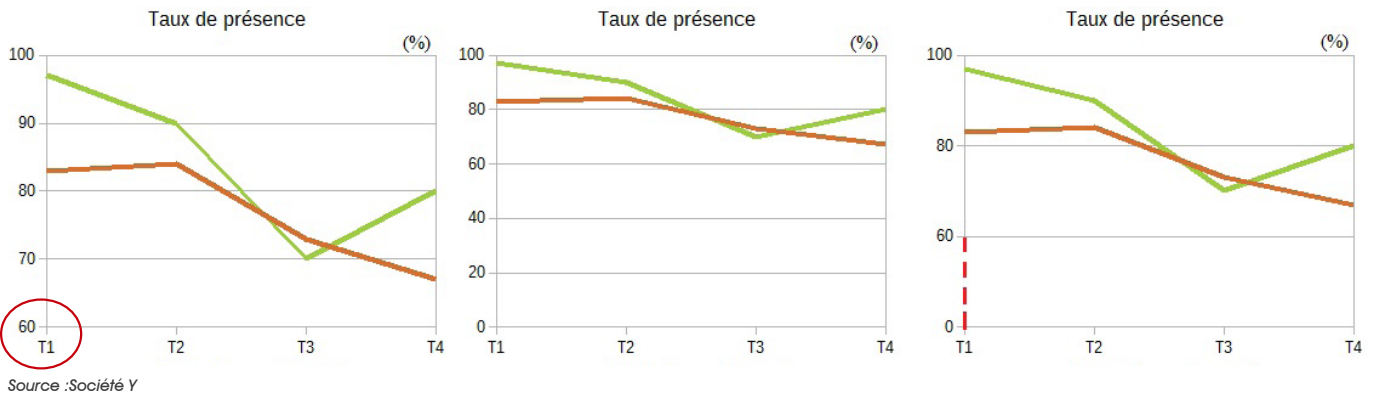
Source : Insee, recensements de la population 2006 et 2011

6 Axes en déséquilibre



Exemple

Il est préférable de ne pas occulter le point zéro dans l'axe (1^{er} graphique). Cependant, ce choix est parfois fait pour montrer les différences ou les variations de valeurs élevées, avec le risque d'amplifier des différences négligeables. Pour contourner le problème, il est possible d'indiquer par des petits points que l'axe vertical est incomplet et a été tronqué (3^e graphique).

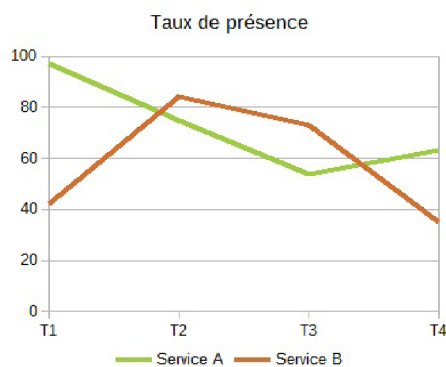


7 Trouver le bon ratio hauteur/largeur

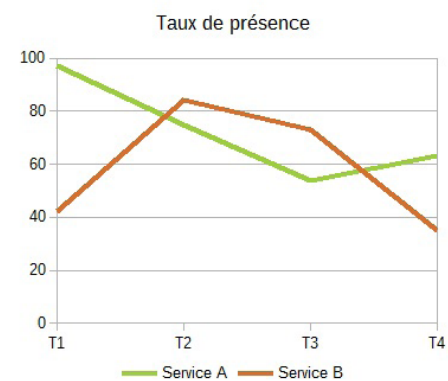


Exemple

Il est déconseillé d'étirer en longueur ou en largeur les graphiques pour éviter de les déformer, comme dans le graphique suivant.



Proposition alternative

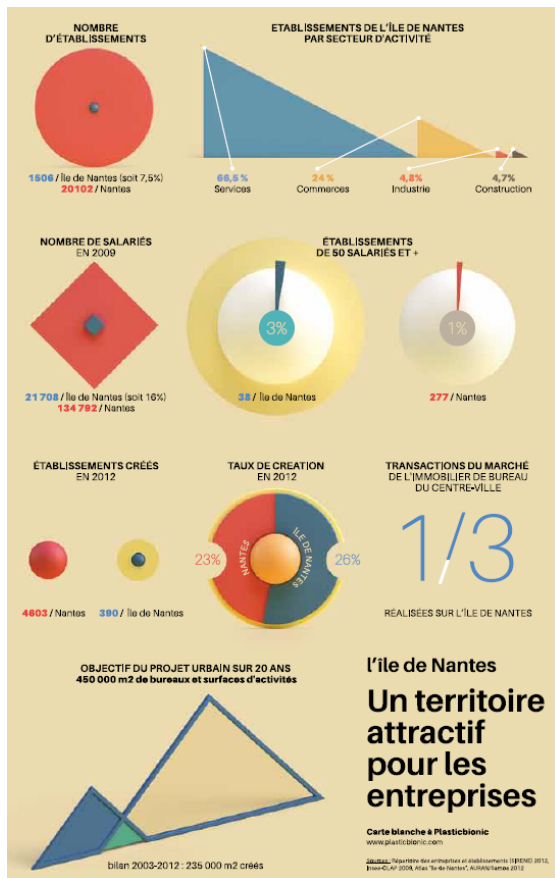


8 Des graphiques " modernes "

Une tendance récente (en particulier dans la presse) tente d'allier graphisme et esthétique. Le résultat laisse perplexe... ■



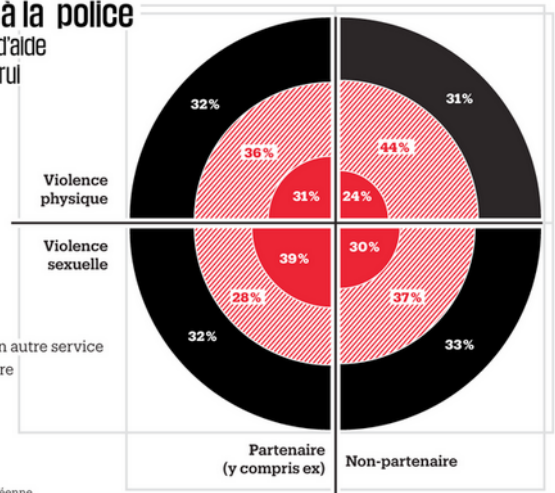
Exemples



Peu de recours à la police

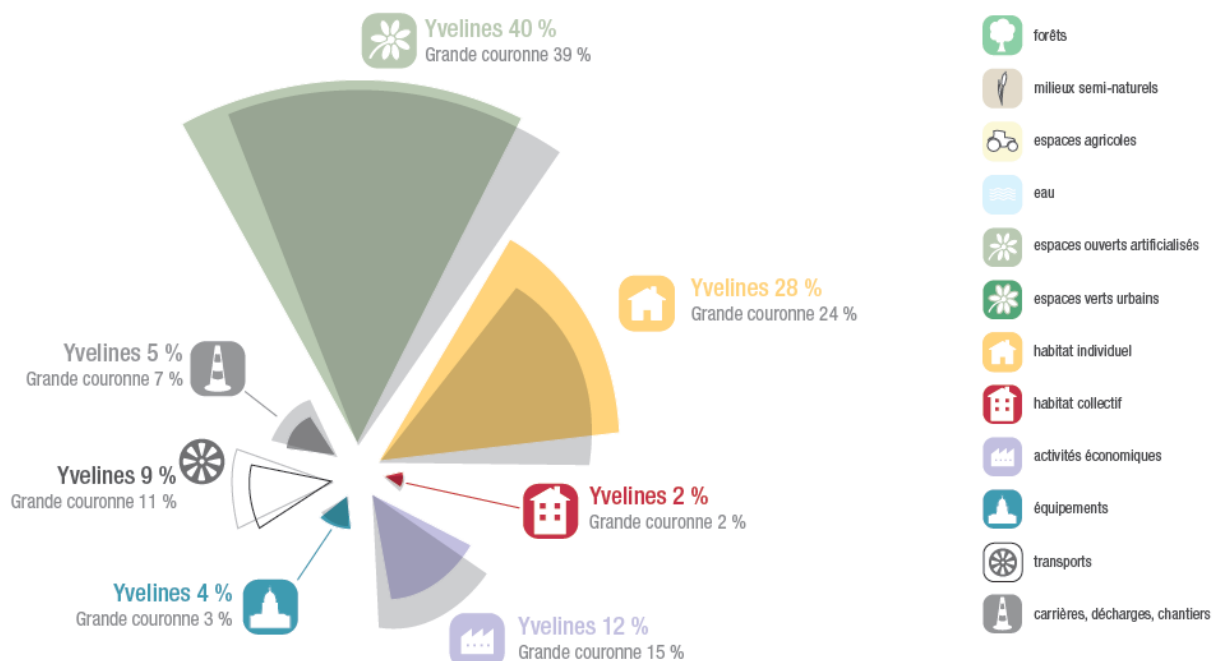
Recours aux services d'aide et discussion avec autrui du fait de violence le plus grave subi, depuis l'âge de 15 ans Par type de violence et d'auteur, en % dans l'Union européenne

- A contacté la police ou un autre service
- A parlé à quelqu'un d'autre
- N'en a parlé à personne



Source : Agence des droits fondamentaux de l'Union européenne

Infographie 1 : que sont devenus les espaces ruraux des Yvelines et de grande couronne consommés entre 1982 et 2012 ?
(en % de l'ensemble des espaces ruraux consommés dans les Yvelines et en grande couronne)



Source : Mos 1982 à 2012, IAU idF

POUR EN SAVOIR PLUS

(BIBLIOGRAPHIE)

► Angers Claude, Graphiques statistiques : forces et faiblesses, *Bulletin AMQ*, mai 1987

► Benveniste Corinne, Savoir compter, savoir conter, épisode n° 3 – des graphiques et des tableaux, *Courrier des statistiques* n° 115, septembre 2005

► Bertin Jacques, La graphique et le traitement graphique de l'information, 1977, Éd. Flammarion.

► Few Stephen, Telling Compelling Stories with Numbers, novembre 2008

► Lebel Bernard, Convaincre avec des graphiques efficaces, 2012

► Magaud Jacques, Pénombre n° 27, octobre 2001

► Monso Olivier et de Saint Pol Thibaut, La perception graphique : mieux construire et interpréter les graphiques, *Courrier des statistiques* n° 126, janvier-avril 2009. ■