

posons  $H_0$  vraie dans nos populations d'intérêt. La statistique de test  $F$  suit une loi de Fisher à (4 ; 1671) degrés de liberté. La valeur observée de la statistique de test est :

$$F = \frac{\frac{\sigma_{inter}^2}{9-1}}{\frac{\sigma_{intra}^2}{n-9}} = 5,72.$$

la probabilité d'observer une valeur de  $F$  supérieure ou égale à  $\hat{F}$  est inférieure au seuil de 1%, donc on rejette l'hypothèse d'égalité des moyennes au 99,9%.

Il existe une différence entre les durées d'interviews moyennes des sans emploi (64,0 minutes), des CAP, BEP, BEPC (65,4 minutes), des diplômés du supérieur (67,8 minutes), des diplômés bac+2 (67,8 minutes) et des diplômés supérieurs (70,0 minutes). Cette différence est significative au seuil de 99,9% : les durées d'interviews sont de plus en plus longues lorsqu'on est plus diplômé.

### articles d'application conseillés

Régis Blazy, Bertrand Chopard, Agnès Finay, Jean-Daniel Guigou, « Entreprises et difficultés : l'arbitrage des tribunaux entre maintien de l'emploi et apurement du passif », *Économie et statistique*, n°443, 2011, p. 51-75.  
Viviane Le Hay, « Chapitre 3 : l'organisation infranationale des valeurs, opinions et attitudes en Europe : une étude par les dissimilarités régionales », « *Sous les plis des drapeaux* », la mosaïque territoriale ? Une analyse infranationale des valeurs, opinions et attitudes en Europe occidentale (1970-1999), Thèse, IEP de Paris, Sociologie, 2010, p. 189-259.  
Pierre-Jean Marescaux, Olivier Desrichard, « Permutation périodique des tâches au sein d'une équipe : motivation, fatigue ou norme sociale ? », *Le travail humain*, vol. 65, n°2, 2002, p. 127-158.  
David L. Strayer, William A. Johnston, « Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone », *Psychological science*, vol. 12, n°6, 2001, p. 462-466.

## Les représentations graphiques

Le chapitre 1, via la question de la représentation d'une distribution statistique, a introduit une première représentation graphique : l'histogramme. Ce chapitre 8 se propose de balayer les types de visualisations graphiques les plus couramment employés en sciences sociales<sup>1</sup>. Nous traiterons ici des représentations utilisées à des fins de lisibilité pédagogique, en vue de représenter synthétiquement et de mettre en évidence « d'un seul coup d'œil » les tendances générales d'un tableau de données<sup>2</sup>. Avant d'exposer les règles fondamentales de l'élaboration graphique, nous présenterons ses principes de construction, qui dépendent principalement du type et du nombre de variables à représenter. Nous aborderons enfin les cas particuliers des données spatio-temporelles, qui répondent à des enjeux spécifiques.

Un message nous semble important : s'il est nécessaire, par construction, de faire des choix (modalité que l'on représente, détermination de la graduation des axes par exemple), il faut avoir conscience qu'ils ne sont en aucun cas neutres et qu'ils engagent l'analyste, le technicien et le lecteur, dans une lecture de la réalité sociale dont il faut absolument cerner les éventuels écueils, biais ou autres pièges d'interprétation.

### 1. Quelle représentation graphique ?

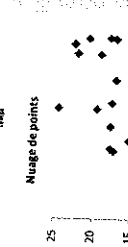
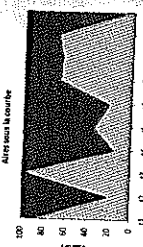
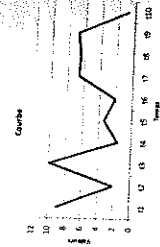
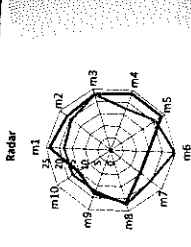
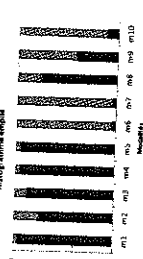
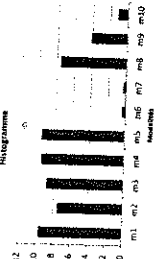
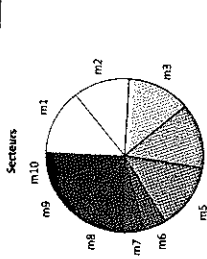
#### 1.1. Les questions à se poser pour trancher

Le choix du type même de représentation graphique ne se fait pas au hasard, selon son humeur, ou simplement pour des questions esthétiques, bien au contraire ! S'il fallait dresser une liste des représentations graphiques usuelles en sciences sociales, il s'agirait (l'encart 1 en propose une illustration) : de secteurs (connus également sous le nom de camemberts), d'histogrammes (et d'histogrammes empilés), de diagrammes en boîte, de radars, de courbes (et d'aires sous la courbe) et de nuages de points.

<sup>1</sup> Il ne s'agira pas ici d'aborder tous les cas particuliers possibles, en trop grand nombre.

<sup>2</sup> Ainsi, il ne sera pas abordé la question particulière de la représentation de plans factoriels par exemple.

# Encart 1 Les représentations graphiques usuelles



Nous reviendrons plus tard sur l'ensemble des éléments énumérés ci-dessus, et aborderons en fin de chapitre et à grands traits le cas particulier de la représentation de données géographiques par l'outil cartographique. Il existe bien entendu d'autres figures utiles à l'analyste pour des objectifs graphiques particuliers, mais il ne s'agit pas ici d'en exposer un catalogue exhaustif.

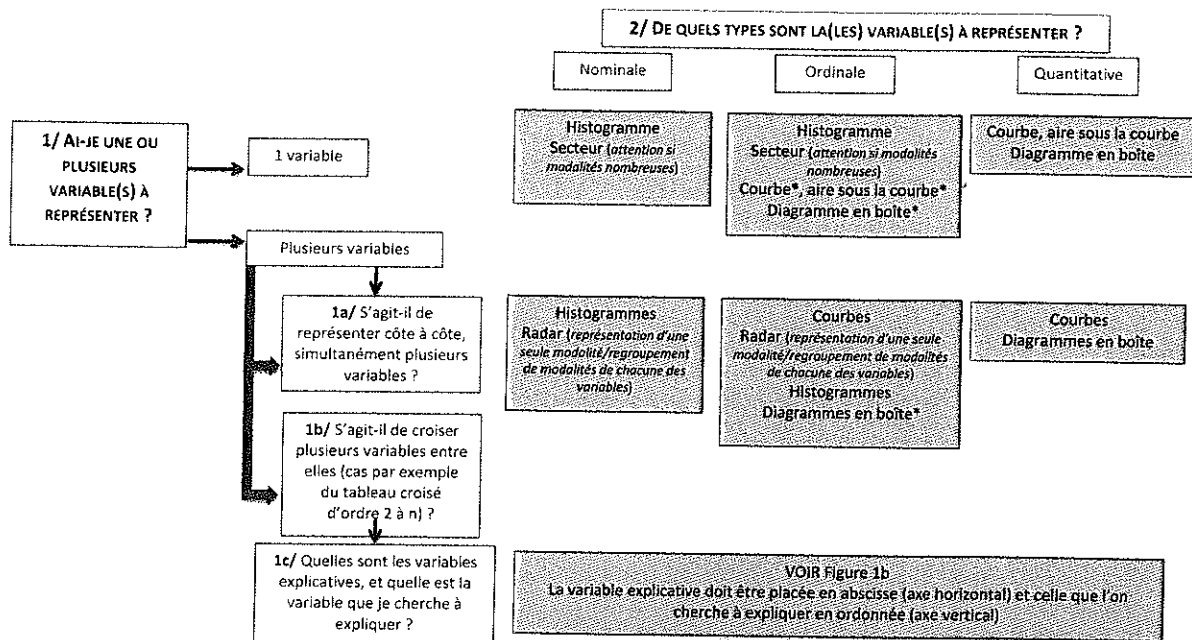
En revanche, ce qu'il est important de bien comprendre, c'est qu'à chacune de ces représentations coïncide un tableau de données bien spécifique, et pas un autre. Dès lors, pour décider du type de représentation à mettre en œuvre, il convient au préalable de répondre aux questions suivantes. Alors, et alors seulement, il deviendra possible de trancher.

- 1. Ai-je une ou plusieurs variables à représenter ? ;
- 2. Si j'ai plusieurs variables : s'agit-il de représenter ensemble, c'est-à-dire côte à côte, leur distribution, ou bien sont-elles croisées les unes avec les autres (cas par exemple du tableau croisé d'ordre 2 à  $n^1$ ) ? Dans ce dernier cas de figure, quelles sont les variables explicatives, et quelle est celle que je cherche à expliquer ? ;
- 3. Pour l'ensemble des variables en présence, quel(s) est(sont) leur(s) type(s) ?

<sup>1</sup> Cf. chapitre 6.

<sup>2</sup> Nous traiterons ici uniquement des représentations du croisement de deux variables et de la distribution côte à côte de variables similaires dans leur type et leur unité de mesure.

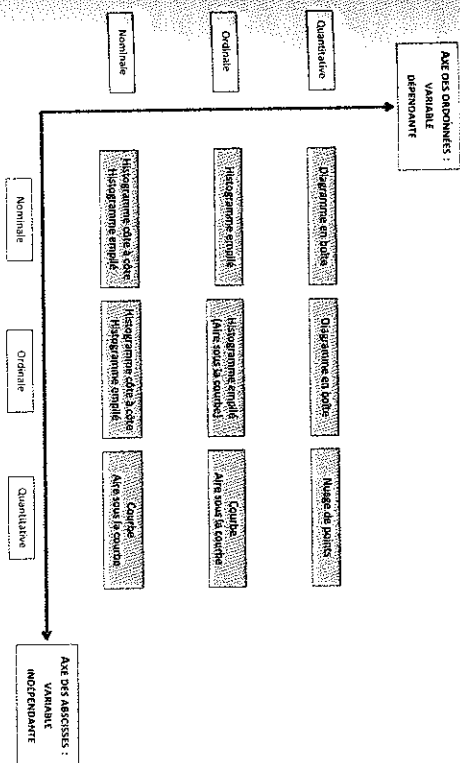
Figure 1a : Les questions à se poser pour déterminer la représentation graphique idéale



\* Cas limites pour une variable ordinale. Tout dépend en réalité du nombre de modalités de cette dernière (à partir d'un certain seuil – comme les échelles de 0 à 10 –, elle peut être considérée tout à la fois comme une variable ordinale et comme une variable continue discrète).

En fonction des réponses apportées à chacune des questions posées ci-dessus, et en suivant le cheminement indiqué dans la figure 1a, puis le cas échéant dans la figure 1b, il devient possible d'identifier la représentation graphique idéale. Notons toutefois qu'avec un peu d'expérience, cette étape finit par être invisible et s'opère intuitivement.

Figure 1b : Choix de la représentation graphique dans le cas du croisement de deux variables



## 1.2. Un exemple pas à pas à partir d'une variable à expliquer nominale : passage de l'analyse univariée à la multivariée

Prenons l'exemple de la question suivante, tirée de l'enquête Eurobaromètre 72.2<sup>1</sup> :

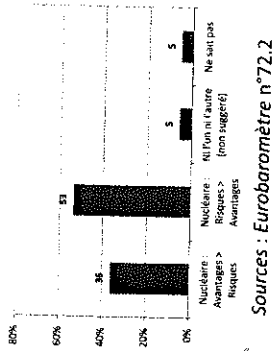
« Quand vous pensez à l'énergie nucléaire, qu'est-ce qui vous vient à l'esprit en premier ? »

- les avantages liés à l'énergie nucléaire sont plus importants que les risques ;
- les risques que représente l'énergie nucléaire sont plus importants que les avantages ;
- ni l'un ni l'autre (non suggéré) ;
- ne sait pas. »

Si l'on s'intéresse exclusivement au cas de la France et en suivant les indications fournies par la figure 1a, on pourrait représenter cette unique variable nominale par un histogramme (figure 2a). Notons qu'une représentation en secteurs est également envisageable.

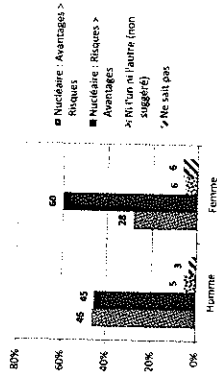
<sup>1</sup> Terrain réalisé en septembre-octobre 2009.

Figure 2a : L'opinion sur l'énergie nucléaire en France, %



Imaginons à présent que l'on souhaite évaluer la différence d'appréciation entre les hommes et les femmes français sur la question<sup>1</sup>. Il conviendra alors, comme l'indiquent la figure 1a puis la figure 1b, de privilégier un histogramme côte à côte (figure 2b)<sup>2</sup>. Rappelons que dès lors que l'on croise des variables, il faut toujours placer en abscisse (axe horizontal) la variable explicative et en ordonnée celle que l'on cherche à expliquer. À la lecture du graphique, on saisit d'une part combien l'opinion diffère entre les femmes et les hommes (jusqu'à 18 points d'écart sur la première modalité)<sup>3</sup>, d'autre part que les premières apparaissent davantage réticentes à l'égard de l'énergie nucléaire que les seconds.

Figure 2b : L'opinion sur l'énergie nucléaire en France en fonction du sexe, %

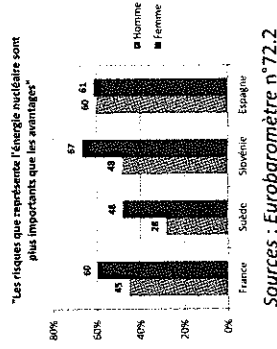


Dans la mesure où nous travaillons ici à partir des données de l'Eurobaromètre<sup>4</sup>, il s'agit en réalité de s'interroger sur d'éventuelles différences entre pays concernant la thématique proposée<sup>5</sup>. Mais que faire dans ce cas précis,

- 1 Il s'agit donc d'un croisement entre deux variables nominales. Celle à expliquer (ou dépendante) est l'opinion sur l'énergie nucléaire.
- 2 À ce stade, un graphique en secteurs n'est plus envisageable car ce type de figure ne permet pas de représenter le croisement de deux variables.
- 3 Il conviendrait bien entendu de réaliser quelques tests statistiques pour s'assurer de la robustesse de ce premier résultat descriptif, mais ce n'est pas l'objet de ce chapitre (voir chapitre 6).
- 4 Pour une présentation des Eurobaromètres, se reporter à Dominique Joye, « Chapitre 1 : Les grandes enquêtes internationales des sciences sociales », Alain Chenu, Laurent Lesnard (dir.), *La France dans les comparaisons internationales : guide d'accès aux grandes enquêtes statistiques en sciences sociales*, Paris, Presses de Sciences Po, 2013, p. 19-30.
- 5 Nous retenons ici pour l'exemple quatre pays : la France, la Suède (Europe du Nord), la Slovaquie (Europe de l'Est) et l'Espagne (Europe du Sud).

dès lors que l'on souhaite ajouter une dimension d'analyse supplémentaire ? En effet, intégrer au graphique précédent une variable supplémentaire rendrait rapidement l'information publiée par trop bavarde et trop dense<sup>1</sup>. Nous vous recommandons alors de simplifier l'information relative à la variable à expliquer, en ne conservant qu'une seule modalité de réponse (celle qui intéresse votre démonstration). C'est la démarche que nous avons entreprise dans la figure 2c en choisissant de représenter la 2<sup>e</sup> modalité de réponse (*Les risques que représente l'énergie nucléaire sont plus importants que les avantages*)<sup>2</sup>. D'une part, nous constatons par l'intermédiaire de ce graphique que les quatre pays conservés n'ont pas la même appréhension du risque lié à l'énergie nucléaire (la Suède apparaît par exemple comme le pays le moins réservé). D'autre part, l'écart enregistré dans le cas français n'est pas systématique (il semble inexistant en Espagne). En tout état de cause, retenons du point de vue technique que le plus souvent, la meilleure voie consiste, autant que faire se peut, à simplifier l'information disponible, en ne conservant qu'une seule modalité de la variable dépendante. Si une représentation à plat, en deux dimensions telle que nous l'avons proposée, peut apparaître à certains lecteurs peu vendeur ou inesthétique, nous attirons son attention sur les dangers d'interprétation occasionnés par le choix d'une reproduction en trois dimensions dans l'encart 2 de ce chapitre.

Figure 2c : L'opinion sur l'énergie nucléaire en fonction du pays et du sexe, %



- 1 De la figure 2b, qui contient déjà 8 bâtons, nous aboutirions à  $8 \times 4 = 32$  bâtons !
- 2 Attention : si dans notre exemple, le choix de la modalité est relativement indifférent, ce n'est pas toujours le cas. Il faut être en mesure d'argumenter sa sélection et de s'assurer que cela ne dissimule pas d'effets importants qu'il ne faudrait pourtant pas occulter.

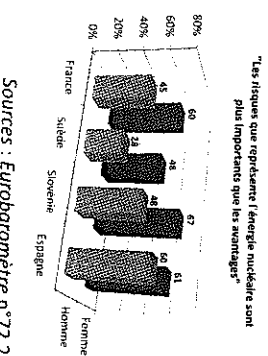
### Encart 2 : ATTENTION PRUDENCE ! La représentation graphique en trois dimensions (3D) : Attention aux illusions d'optique !

Lorsque l'on cherche à résumer l'information contenue par au moins deux variables, on se retrouve très vite confronté à la difficulté de bien la représenter graphiquement. En particulier, comment procéder pour les résultats d'un tableau croisé ? Comment traiter l'information provenant du croisement de trois variables ? Une solution parfois adoptée est de la représenter en trois dimensions, une pour chacune des variables. Nous souhaitons par cet encart vous alerter sur cette pratique et vous recommander un usage parcimonieux de cette technique.

En effet, si une figure en trois dimensions respecte fort logiquement les règles du dessin en perspective, l'œil du lecteur, lui, a tendance à lire l'information diffusée comme si elle était représentée en deux dimensions, ce qui engendre un décalage problématique entre ce que montre effectivement le graphique, et ce qu'en voit le lecteur.

Reprenons l'exemple de la question relative à l'énergie nucléaire. Nous pourrions nous dire : « mais pourquoi ne pas représenter la figure 2c en trois dimensions ? Cela serait beaucoup plus élégant ! ». Certes. Procédons donc ainsi. Dans la figure 3, les deux axes horizontaux correspondent aux deux variables explicatives (pays et genre) et l'axe vertical indique la proportion de réponse pour la modalité « les risques que représente l'énergie nucléaire sont plus importants que les avantages ». Afin de cerner au premier coup d'œil la limite d'un tel choix graphique, nous avons intégré, pour chaque histogramme, le pourcentage qui lui est associé.

Figure 3 : L'opinion sur l'énergie nucléaire en fonction du pays et du sexe, %  
(représentation en « 3D »)



Que constatons-nous ? La lecture de cet histogramme en 3D est peu évidente si on n'y prend pas garde. Si on ne regarde pas le chiffre précisé sur le graphique (car bien souvent ces chiffres n'apparaissent pas sur les figures), on a tendance à lire, sans tenir compte de la

perspective, qu'un peu moins de 40% des hommes français ont opté pour cette modalité de réponse... Ils sont en réalité 45% ! De la même manière, on dirait que les femmes françaises sont un peu moins de 60% sur la même position... Alors qu'elles sont en réalité très exactement 60% !

Dès lors, si les proportions ne sont pas précisées sur le graphique, on peut être amené à effectuer de lourdes erreurs de lecture et d'interprétation. Dans ce cas, l'histogramme classique tel que nous l'avons proposé (figure 2c) ou même le tableau de données (tableau 1) apparaissent donc plus indiqués.

Tableau 1 : L'opinion sur l'énergie nucléaire en fonction du pays et du sexe, %

	Homme	Femme
France	45	60
Suède	28	48
Slovénie	48	67
Espagne	60	61

Sources : Eurobaromètre n°72.2

Une fois cette première étape relative au choix de la représentation graphique résolue, il ne faut en aucun cas considérer le travail achevé, loin s'en faut. Afin de faire concorder le type de représentation à son objectif de visualisation de l'information, quelques principes et précautions, faisant bien souvent appel au bon sens, s'imposent.

## II. Les règles fondamentales de la représentation graphique

### II.1. Des critères typographiques parfois imposés

Nous entendons par critères typographiques imposés, les normes de présentation prescrites par les éditeurs et autres commanditaires (thèses, revues, etc.), en particulier la possibilité d'utiliser la couleur ou non. Cela dit, si une variable nominale comprenant dix modalités de réponses est difficile à représenter directement par un histogramme en noir et blanc, l'emploi de la couleur ne résout pas tout. La conception du graphique dépend dès lors de contraintes typographiques certes, mais surtout de la préparation préalable des variables que l'on souhaite analyser et de leur recodage adapté en amont<sup>1</sup>.

Ensuite, au-delà de la possibilité offerte de représentations graphiques en couleur, il convient de jouer sur leur taille, le format paysage ou portrait, les polices de caractère, l'épaisseur des tracés et sur les pointillés, hachures et autres quadrillages.

Également de notre point de vue, la sobriété ainsi que la simplicité de la mise en forme graphique sont bien souvent les meilleures alliées pour mettre en valeur son tableau de données.

<sup>1</sup> Se reporter au chapitre 3 pour les notions concernant les méthodes de recodage des variables.

Pour finir, rappelons qu'un bon graphique suppose de façon triviale de lui associer :

- d'une part un titre précisant le cas échéant les unités de mesure utilisées ;
- d'autre part une légende précise et complète.

## II.2. La graduation des axes : un choix qui n'est pas neutre

Il faut avoir conscience qu'un choix inapproprié dans la construction des échelles des axes (en abscisse comme en ordonnée) est susceptible d'aboutir à des problèmes de surinterprétation des résultats représentés ou à des difficultés à visualiser le plus important de l'analyse. A tout le moins, il est clair que les décisions opérées sont tout sauf neutres. Elles conduisent à mettre en exergue un phénomène particulier, parfois au détriment d'autres, qui peuvent dès lors être minorés, voire occultés. Pour l'analyste soucieux de rigueur et de méthode, il s'agit de déterminer une échelle en fonction de la problématique élaborée, ce scientifiquement et en dehors de toute considération subjective ; mais d'autres critères peuvent amener à dériver vers une représentation graphique partielle. Un maître mot donc : vigilance !

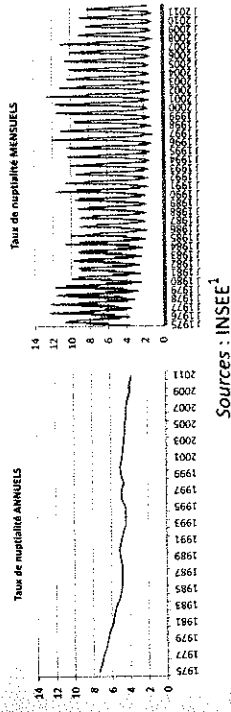
La représentation du taux de nuptialité<sup>1</sup> de 1975 à 2011 (figure 4) constitue un exemple typique de l'effet du choix de l'échelle des abscisses, **en particulier de son niveau de précision**. L'INSEE fournit en effet des données très détaillées sur la question : il nous est ainsi possible de représenter mensuellement les taux de nuptialité aussi bien qu'annuellement. Mais un niveau de précision mensuel a-t-il un sens ? En réalité, tout dépend de la question que l'on se pose ! La courbe de gauche relie des points annuels, tandis que celle de droite relie des points mensuels. Si la première souligne plutôt l'érosion tendancielle de cette pratique au cours du temps (division par deux de la proportion en 35 années)<sup>2</sup>, la seconde met davantage en exergue l'extrême saisonnalité du mariage, plus volontiers organisé au moment des beaux jours<sup>3</sup>. Aucun de ces deux graphiques ne peut être considéré comme fondamentalement faux ; dès lors, la décision se doit d'être adaptée à la question posée.

<sup>1</sup> Le taux de nuptialité annuel est défini comme suit sur le site Internet de l'INSEE : « Le taux de nuptialité est le rapport du nombre de mariages de l'année à la population totale moyenne de l'année » (consulté le 15 janvier 2014).

<sup>2</sup> Notons toutefois que le nombre des mariages est reparti à la hausse en 2012 (Insee Première, n° 1482).

<sup>3</sup> La question des avantages fiscaux à se marier en milieu d'année est également à prendre en compte. Cet exemple a été choisi à des fins illustratives. Nous mettons volontairement de côté la question de la primo-nuptialité et des remariages, qui n'est pas notre objet ici.

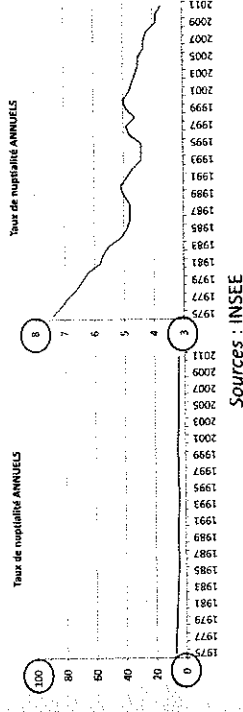
Figure 4 : Effet de l'unité de mesure (annuelle à gauche/mensuelle à droite) de l'échelle des abscisses sur la courbe du taux de nuptialité (France métropolitaine)



Sources : INSEE<sup>1</sup>

A partir de ce même exemple (mais figure 5 cette fois), examinons maintenant l'effet de l'amplitude de l'échelle de l'axe des ordonnées, **en particulier du choix de grossir ou non, par un effet de loupe, la fenêtre d'observation** (détermination des bornes minimales et maximales).

Figure 5 : Effet du choix de l'amplitude de l'échelle des ordonnées (à gauche : [0 ; 100] / à droite : [3 ; 8]) sur la courbe du taux de nuptialité (France métropolitaine)



Sources : INSEE

Souhaitant donc représenter graphiquement l'évolution du taux de nuptialité de 1975 à 2011, deux possibilités extrêmes pourraient être envisageables (cf. figure 5) :

- soit (cas du graphique à gauche) choisir de montrer l'ensemble des valeurs théoriquement possibles : un taux, par définition, pouvant évoluer entre 0 et 100, on fait varier alors l'axe des ordonnées de 0 à 100 ;
- soit (cas du graphique à droite), opter pour une représentation au plus près de la distribution : on fait alors varier l'échelle des ordonnées du minimum (inclus) de la distribution à son maximum (inclus). Dans notre exemple, l'axe varie donc de 3 à 8 et non plus de 0 à 100.

Il est manifeste que selon le choix opéré, l'effet visuel auquel on aboutit est radicalement différent et l'interprétation effectuée peut en être affectée. Si le premier graphe donne l'impression d'une relative stabilité du taux de nuptialité au cours du temps, le second semble au contraire montrer une baisse rapide de ce dernier. La représentation idoine se situe entre les deux.

<sup>1</sup> <<http://www.bdm.insee.fr/bdm2/choixCriteres.action?codeGroupe=62>>.

15 janvier 2014.



Nous souhaitons ici sensibiliser le lecteur : soyez systématiquement attentif aux critères d'élaboration d'une représentation graphique ! Cela est vrai quand vous-même, bien entendu, êtes amené à en réaliser une. Mais cela l'est peut-être encore davantage quand vous consultez un document illustré par ce type d'outil visuel. Au risque de nous répéter, un maître mot : vigilance !

Mais alors, que faire dans notre exemple ? Quels choix effectuer ? Là encore, pas de réponse univoque, mais sans doute une représentation intermédiaire, telle qu'on la trouve en figure 4 (à gauche), est-elle le bon chemin.

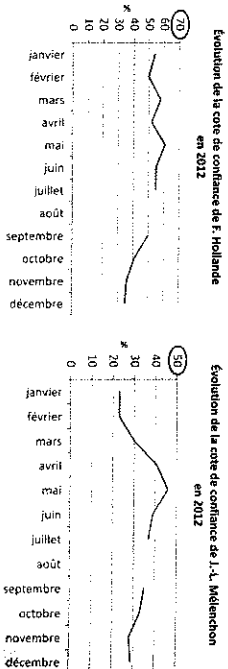
En tout état de cause, une règle essentielle doit être respectée dans le cas du recours à plusieurs graphiques sur un même thème (dans un même article scientifique par exemple) : il faut impérativement s'assurer de la **reproduction à l'identique** des axes d'une représentation à l'autre. C'est l'objet de notre section à venir.

### II.3. L'homogénéité des axes pour une comparaison rigoureuse

Lorsqu'il s'agit de confronter plusieurs graphiques, ce qui est bien souvent le cas d'un chapitre d'ouvrage ou d'un article scientifique, il est essentiel de garantir leur comparabilité.

Un tel objectif est atteint à la seule condition que les bornes minimales et maximales ainsi que les graduations des axes soient rigoureusement identiques pour l'ensemble des graphiques susceptibles d'être comparés entre eux.

Figure 6a : Représentations biaisées (axes verticaux distincts) – Évolution des cotes de confiance mensuelles de F. Hollande et de J.-L. Mélenchon (janvier-décembre 2012)

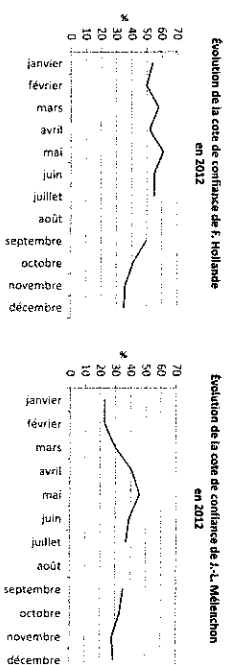


Sources : Baromètre politique mensuel, TNS Sofres, Figaro Magazine, janvier 2012-décembre 2012. Il n'y a pas d'interrogation au mois d'août de chaque année.

La figure 6a constitue un exemple de cet écueil : les deux graphiques, qui représentent l'évolution respective de la cote de confiance de François Hollande et de Jean-Luc Mélenchon, sont dotés de la même échelle des abscisses. Une lecture hâtive, qui ne prêterait pas attention aux axes verticaux, nous mènerait aux constats suivants : les deux personnalités politiques évoluent de manière distincte mais avec des amplitudes globalement comparables, et même un maximum supérieur dans le cas de Jean-Luc Mélenchon. Or si les logiques d'évolution des cotes de confiance sont en effet propres à chacun, la suite de

l'interprétation est faussée ! Dans l'optique de comparer plusieurs représentations, il faut toujours **déterminer une échelle commune**, c'est-à-dire une graduation, un minimum et un maximum communs, qui minimise dans le même temps un risque d'effet loupe abusif (paragraphe précédent). Ici, un axe allant de 0% à 70% semble le meilleur compromis.

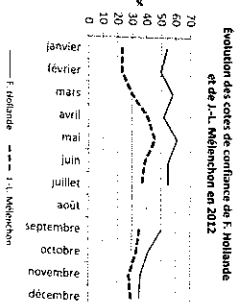
Figure 6b : Représentations comparables (axes verticaux identiques) – Évolution des cotes de confiance mensuelles de F. Hollande et de J.-L. Mélenchon (janvier-décembre 2012)



Sources : Baromètre politique mensuel, TNS Sofres, Figaro Magazine, janvier 2012-décembre 2012

Les deux représentations graphiques ci-dessus (figure 6b) sont comparables et donc satisfaisantes. Notons qu'une dernière solution dans ce cas précis pourrait également consister à représenter les deux courbes sur le même graphique (figure 6c).

Figure 6c : Représentations simultanées – Évolution des cotes de confiance mensuelles de F. Hollande et de J.-L. Mélenchon (janvier-décembre 2012)



Sources : Baromètre politique mensuel, TNS Sofres, Figaro Magazine, janvier 2012-décembre 2012

### Encart 3 : ATTENTION PRUDENCE !

*La graduation des axes : le cas particulier du croisement de deux variables à la même unité de mesure ou le syndrome « il ne doit pas y avoir une tête qui dépasse ! »*

Quelle est la principale précaution à prendre lorsque l'on souhaite croiser deux variables comportant la même unité de mesure ? Indiscutablement et toujours afin de ne pas créer artificiellement d'illusion d'optique, il faut s'assurer de l'homogénéité effective des deux axes du point de vue de leur graduation ainsi que de leurs bornes minimale et maximale.

Prenons un exemple tiré de l'enquête *European Social Survey*<sup>1</sup>. À l'occasion de sa troisième vague, elle pose la question suivante aux personnes interrogées :

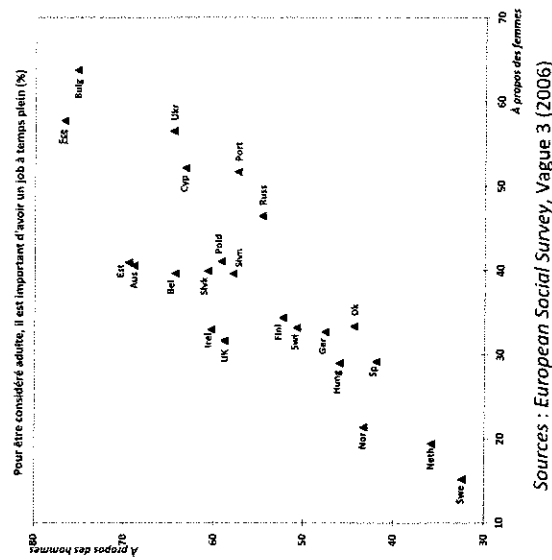
« Quelques questions maintenant sur le moment où [une femme / un homme]<sup>2</sup> peut être considéré(e) comme un(e) adulte. Pour pouvoir être considéré(e) comme un(e) adulte, dans quelle mesure est-il important pour [une femme / un homme]...  
... d'avoir une activité professionnelle à plein temps ? »

Nous souhaitons représenter graphiquement dans quelle mesure les représentations que les Européens ont de l'adulte sont susceptibles de varier, en pourcentages, selon que l'on parle des hommes ou bien des femmes et selon les pays. Il s'agit, en d'autres termes, de confronter les niveaux nationaux des réponses (cumul des réponses « plutôt important » et « très important ») selon que l'on parle des hommes ou des femmes. Un nuage de points s'avère tout à fait indiqué dans un tel cas de figure<sup>3</sup> comme on peut le découvrir en figures 7a et 7b (la question relative aux hommes est placée en ordonnée ; celle relative aux femmes en abscisse).

Mais ces figures montrent combien l'uniformité des axes est primordiale à une lecture non biaisée du phénomène : si la figure 7a (représentation biaisée, qui ajuste au plus près de chaque série de

données les bornes minimale et maximale des axes) montre bien des conceptions différenciées d'être adulte au sein des pays européens, elle ne laisse pas transparaître visuellement un déséquilibre flagrant des représentations à propos des hommes et des femmes. Pourtant à bien y regarder, ce que permet en revanche la figure 7b (qui propose deux axes identiques pour les deux sexes), l'accès à une activité professionnelle rémunérée, et par ce biais à l'autonomie financière, constitue bel et bien, au-delà de la grande variabilité des représentations liées à l'âge adulte en Europe, un rite de passage vers l'âge adulte qui semble plus décisif dans tous les pays concernant les hommes que pour les femmes.

Figure 7a : Représentation biaisée (axes différents) – Positions nationales de l'importance d'avoir une activité professionnelle à temps plein pour être considéré adulte (à propos des hommes et à propos des femmes)<sup>1</sup>



Sources : *European Social Survey*, Vague 3 (2006)

Dès lors, nous voyons combien l'effort de rigueur dès le choix de ses axes constitue un élément déterminant de la qualité d'une représentation graphique<sup>2</sup>. Et l'uniformité des axes doit se prolonger jusqu'au bout de la démarche : si un centimètre représente un écart

<sup>1</sup> Aus : Autriche ; Bel : Belgique ; Bulg : Bulgarie ; Cyp : Chypre ; Dk : Danemark ; Est : Estonie ; Finl : Finlande ; Fce : France ; Ger : Allemagne ; Hung : Hongrie ; Irel : Irlande ; Neth : Pays-Bas ; Nor : Norvège ; Pold : Pologne ; Port : Portugal ; Russ : Russie ; Slvk : Slovaquie ; Slvn : Slovénie ; Sp : Espagne ; Swe : Suède ; Swi : Suisse ; UK : Royaume-Uni ; Ukr : Ukraine.

<sup>2</sup> Dans notre exemple, une autre solution aurait été de conserver la figure 7a, mais d'y ajouter la bissectrice, en précisant qu'un pays positionné au-dessus de la droite signifie que l'item apparaît plus souvent important pour un homme que pour une femme.

<sup>1</sup> Cet exemple est tiré de : Flora Chanvrlil, Viviane Le Hay, « Que signifie "être adulte" pour les Européens ? Des conceptions différenciées concernant les hommes et les femmes ? », *Note de recherche*, Sciences Po, 3 décembre 2008, <<http://ess.sciencespo.com/Que-signifie-etre-adulte-pour-les.html>>, consulté le 19 mai 2014.

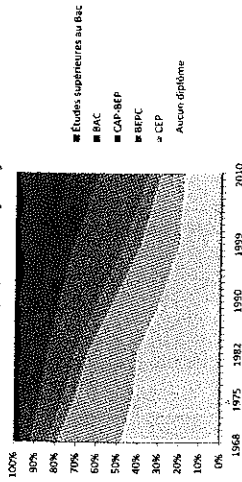
<sup>2</sup> Une expérimentation a été réalisée : une moitié de la population sélectionnée au hasard dans l'échantillon a été consultée concernant les femmes, et l'autre moitié concernant les hommes. Dans la mesure où les questions à propos des hommes et à propos des femmes sont posées à des sous-échantillons distincts, il n'y a pas de risque que les réponses sur un groupe influencent les réponses sur l'autre (effet combiné de halo et de désirabilité sociale : si l'on questionnait une personne successivement à propos des hommes puis à propos des femmes, nous pourrions raisonnablement faire l'hypothèse qu'une partie non négligeable de la population interrogée aurait une forte propension à répondre de façon quasiment identique à propos des hommes et à propos des femmes).

<sup>3</sup> Deux variables quantitatives (en pourcentages) sont croisées (voir figure 1b de ce chapitre).





Figure 9 : Évolution annuelle de la répartition des niveaux de diplôme dans la population française



Sources : INSEE, données des recensements de la population, France métropolitaine, hommes et femmes de 25 ans et plus, construction du graphique Vincent Tiberj<sup>1</sup>

#### Encart 4 - ATTENTION PRUDENCE ! La graduation d'un axe temporel : s'accommoder d'une série temporelle non linéaire ou incomplète

La difficulté que l'on rencontre bien souvent quand on a affaire à une évolution temporelle consiste à faire fi de l'absence de données pour certaines dates, ou encore d'informations présentes à des dates irrégulières dans le temps (c'est le cas par exemple des dates des élections présidentielles en France pour raison d'élection anticipée et suite au passage du septennat au quinquennat).

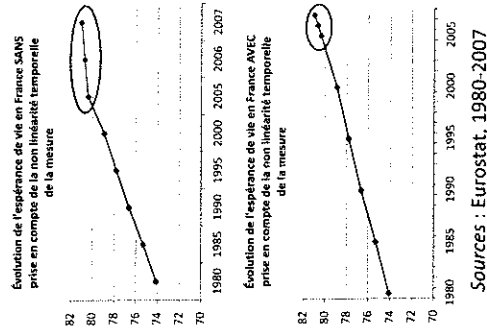
Pour le premier cas de figure (absence de données), soit l'information n'existe pas, c'est-à-dire qu'il n'y a pas de mesure effectuée à une date donnée et pourtant attendue, soit elle existe mais est inconnue. Les figures 6 de ce chapitre fournissent un excellent exemple de mesure qui pourrait être attendue mais qui n'a pas été réalisée : alors qu'on dispose d'une information mensuelle, celle-ci n'a pas été accomplie au moment de la pause estivale, au mois d'août 2012. Dans cet exemple, mieux vaut laisser un « blanc » dans la courbe. En revanche, si la mesure n'est pas connue mais devrait l'être en théorie (cas d'un fichier de données mal renseigné par exemple), on peut décider de relier les deux points entourant la valeur absente, la courbe passant ainsi par la moyenne de ces deux points. Il est alors vivement conseillé d'indiquer cette carence dans les sources.

Pour le cas d'une non-linéarité de la mesure, il est fondamental, toujours afin d'éviter des effets d'optique malheureux, d'adapter sa graduation au déroulement temporel effectif. La figure 10 constitue un exemple frappant de cet effet d'optique potentiel<sup>2</sup> : on dispose de

<sup>1</sup> <[http://www.insee.fr/fr/themes/detail.asp?reg\\_id=99&ref\\_id=pop-17ans-dipl](http://www.insee.fr/fr/themes/detail.asp?reg_id=99&ref_id=pop-17ans-dipl)>, 5 février 2014.  
<sup>2</sup> Les données sont adaptées aux besoins de la démonstration.

l'évolution de l'espérance de vie tous les 5 ans entre 1980 et 2005, puis de cette même mesure pour 2006 et 2007.

Figure 10 : Évolution annuelle de l'espérance de vie sans, puis avec prise en compte de la non-linéarité temporelle de la mesure



Sources : Eurostat, 1980-2007

La non-prise en compte de la non-linéarité de la mesure dans la représentation graphique du haut provoque un effet de seuil artificiel entre 2005 et 2007 qui disparaît lorsque l'échelle des abscisses est corrigée en conséquence (graphique du bas).

### III.2. La représentation des données géographiques

Les données géographiques peuvent être représentées à l'aide de cartes<sup>1</sup>. L'outil cartographique, s'il est bien utilisé, est plus adapté à ce type de données qu'une représentation graphique classique ou qu'une présentation des données par un tableau puisqu'il permet de mettre en évidence visuellement les logiques proprement spatiales d'un phénomène social (effet de proximité ou de distance, mécanismes de diffusion, etc.). Nous n'avons pas la prétention ici de détailler l'ensemble des règles inhérentes à la représentation cartographique, il s'agit d'un ouvrage à part entière. Nous allons néanmoins présenter quelques règles de base utiles dans le cadre de représentations cartographiques relativement simples, c'est-à-dire ne traitant que d'une seule variable<sup>2</sup>, cette dernière étant quantitative (afin d'aborder ainsi la question de son regroupement en classes)<sup>3</sup>.

<sup>1</sup> Pour aller plus loin, l'introduction présente quelques sites Internet utiles en termes de ressources cartographiques.

<sup>2</sup> Il est bien entendu tout à fait possible de représenter simultanément plusieurs variables sur une même carte. Par exemple, une première variable à l'aide de plages de couleurs et une seconde variable à l'aide de symboles.

<sup>3</sup> Le cas d'une variable nominale ou ordinale ne présente en effet pas de difficulté technique particulière.

Il s'agit aussi de sensibiliser le lecteur et d'attirer son attention sur les grands pièges d'une illustration cartographique abusive ou biaisée. Encore une fois : vigilance oblige ! Nous l'avons déjà vu dans le cas d'une représentation graphique usuelle, le choix de la graduation des axes est loin d'être anodin ou innocent. De la même façon concernant une carte, deux dimensions ne sont pas neutres :

- le choix de l'échelle (du maillage) géographique d'une part ;
- la manière de discrétiser la variable quantitative, c'est-à-dire d'opérer son regroupement en classes.

En premier lieu, il faut prendre conscience que le choix de partitionnement de l'espace produit nécessairement des effets de lecture spécifiques. Faut-il alors, si on en a la possibilité, privilégier un maillage fin (le canton par exemple si on raisonne dans le cas français) ou apparemment plus grossier (comme le département) ? Il n'y a bien sûr pas de réponse univoque à une telle question. Bien souvent, chaque grille de lecture apporte des éclairages différents mais complémentaires. Notons néanmoins que, contrairement à l'intuition, le maillage le plus fin n'est pas nécessairement le plus performant. Jean-Paul Terrenoire montre en 2003, en comparant les résultats relatifs à la pratique religieuse en France aux niveaux cantonal et départemental, que le niveau géographique le moins fin est celui qui met en évidence avec le plus d'acuité les phénomènes sociaux autour de la pratique religieuse : « les relations statistiques se montraient plus étroites et les coefficients de corrélation beaucoup plus élevés quand les calculs restaient dans le cadre départemental. »<sup>1</sup> Nous ne développerons pas plus avant ce point, mais retenons simplement que si « regrouper des unités spatiales entraîne une perte d'information, [...] la trop forte parcellisation de l'espace peut entraîner des phénomènes aléatoires dommageables. »<sup>2</sup> Il convient donc de déterminer sa partition spatiale avec soin.

En second lieu, lorsque l'on souhaite représenter des données géographiques quantitatives, par exemple la densité de population dans les communes d'un département donné, ou les résultats électoraux pour un candidat donné à une élection donnée dans les départements français, il est nécessaire de choisir une façon de discrétiser l'information : cela signifie qu'il faut déterminer un regroupement en classes de la variable initiale.<sup>3</sup> Plusieurs choix se présentent alors :

- puis-je utiliser des couleurs ou suis-je limité par une représentation en niveaux de gris ? ;

- faut-il choisir un regroupement en quantiles<sup>4</sup> ? selon la méthode des moyennes emboîtées<sup>5</sup> ? ou encore un regroupement raisonné<sup>6</sup> ? ;
- lorsque l'on souhaite représenter l'évolution temporelle d'un phénomène géographique, c'est-à-dire représenter un même phénomène avec plusieurs cartes pour différentes dates, faut-il choisir un regroupement en classes commun à l'ensemble des points temporels ou plutôt privilégier un regroupement propre à chaque date (cette question se décline d'ailleurs quelle que soit la comparaison de cartes que l'on souhaite effectuer) ?

Si la première question est dépendante de critères éditoriaux, les deux suivantes peuvent être aisément illustrées.

La distribution des densités de population par département français constitue un bon exemple de l'effet de la méthode de discrétisation sur la visualisation finale. La figure 11 compare de gauche à droite et de bas en haut : une carte proposant un découpage en quantiles des densités de population, une représentation privilégiant un regroupement selon la méthode des moyennes emboîtées et enfin une carte présentant un regroupement raisonné. Si la carte selon la méthode des moyennes emboîtées permet de visualiser facilement les *outliers* (individus statistiques aberrants) de la région parisienne et en particulier les trois départements les plus peuplés (Paris, les Hauts-de-Seine et la Seine-Saint Denis), elle écrase les départements à faible densité de population.<sup>7</sup> En effet la première catégorie (15-102 hab/km<sup>2</sup>) comprend 60 départements sur les 96, soit près des deux-tiers. La carte utilisant le regroupement en quantiles n'est plus sensible aux *outliers* et donc plus équilibrée, mais mélange dans la dernière catégorie des départements à très haute densité de population comme Paris (21060 hab/km<sup>2</sup>) à d'autres plus « moyens » comme la Haute-Savoie (11 hab/km<sup>2</sup>). Elle ne permet donc pas de distinguer les départements très urbanisés qui sortent du lot. La solution intermédiaire présentée en bas de cette figure permet de conserver les avantages des deux méthodes détaillées ci-dessus : s'agit d'une méthode raisonnée, basée sur le regroupement en quantiles, mais qui subdivise le dernier quartile en trois catégories distinctes, mettant ainsi en valeur les départements à haute (386-935 hab/km<sup>2</sup>) et très haute densité de population (5351-21060 hab/km<sup>2</sup>). Cette solution raisonnée permet de ne pas écraser les départements à faible densité de population dans une grande et unique catégorie tout en mettant en valeur les *outliers* à forte densité de population (et en quel sorte en isolant dans une classe à part les cas spécifiques et excentrés, du reste

1 Jean-Paul Terrenoire, « Cartographie et analyse écologique quantitative de la pratique religieuse rurale et urbaine en France », *Sociétés contemporaines*, vol. 1-2, n°49-50, 2003, p. 63-84.

2 Marie-Alexandre Laurent, Isabelle Thomas, « Modèle d'interaction spatiale et aggrégation des lieux : l'exemple des données criminelles », *L'Espace géographique*, n°3, 1997, p. 269-279.

3 Se reporter au chapitre 3 pour d'autres exemples de discrétisation.

4 Se reporter au chapitre 2 pour la définition des quantiles.

5 Il s'agit ici dans un premier temps de séparer les valeurs inférieures ou égales à la moyenne des valeurs supérieures, puis à l'intérieur de chacune des deux classes ainsi créées de calculer la moyenne de la classe et de recouper en deux celle-ci, jusqu'à obtenir le nombre de classes souhaitées.

6 Se reporter au chapitre 3 pour la définition d'un regroupement raisonné ou hypothèse déductif.

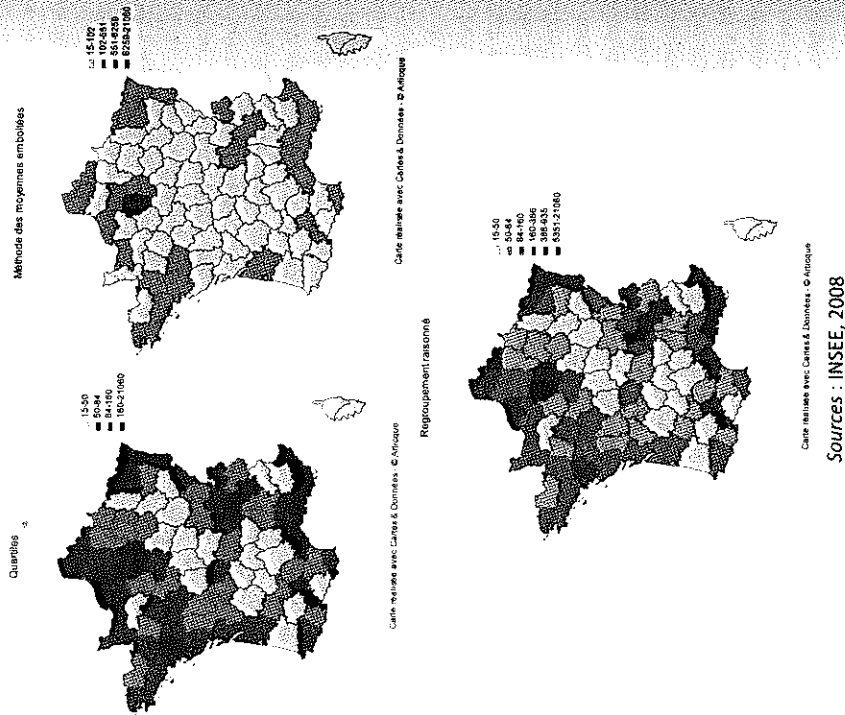
7 Il s'agit de trois méthodes couramment utilisées en sciences sociales.

8 Se reporter au chapitre 2 pour revenir sur ce principal piège de la moyenne.

9 Il n'y a aucun département ayant une densité de population comprise entre 935 et 1000 hab/km<sup>2</sup>, d'où la cassure nette entre ces deux classes.

la population statistique étudiée). En définitive, il est nécessaire de tenir compte de l'homogénéité de la distribution dans le choix de son mode de discrétisation, la présence de points aberrants pouvant venir contrarier des choix de regroupement pertinents par ailleurs.

Figure 11 : Représentation cartographique de la densité de population par département, France métropolitaine (hab/km<sup>2</sup>, en haut à gauche en quartiles, en haut à droite selon la méthode des moyennes emboîtées, en bas par un regroupement raisonné)



Les figures 12a et 12b montrent respectivement le choix d'une échelle propre à chaque point temporel dans le premier cas et celui d'une échelle commune dans le second. La première de ces deux figures permet de visualiser clairement les zones de force et de faiblesse du candidat FN à chacun des trois scrutins présentés : les zones de force se situent globalement « à l'est d'une ligne Le Havre-Perpignan »<sup>1</sup>, quelle que soit l'année considérée. Cependant les sextiles ne sont pas bornés de la même manière et la seconde figure permet de visualiser l'évolution du rapport de force entre 2002 et 2012 : on constate que le scrutin de 2012 est équivalent à celui de 2002 en termes de niveaux, avec un phénomène de diffusion autour des zones de force, tandis que le scrutin de 2007 est très en dessous des deux autres en termes de niveaux, avec seulement les quatre premiers sextiles communs représentés. Dans cet exemple, nous avons choisi une représentation en sextiles et non par les moyennes emboîtées. Il aurait été tout à fait possible d'utiliser une représentation à l'aide de la seconde méthode : chacune présente en réalité des avantages et des inconvénients. En revanche, à partir du moment où l'on doit représenter plusieurs points temporels, une discrétisation non systématique, basée sur un regroupement raisonné, s'avère plus délicate à mettre en œuvre. En effet, un regroupement raisonné commun à différents points temporels peut rapidement mener à une surinterprétation des résultats s'il n'est pas manié avec précaution.

<sup>1</sup> Pascal Perrineau, *La France au Front*, Paris, Fayard, 2014, p. 47.

**Figure 12a : Vote FN au premier tour de l'élection présidentielle de 2002 à 2012, France entière (% aux exprimés, sextiles propres à chaque scrutin)**



Sources : ministère de l'Intérieur, 2002-2012

**Figure 12b : Vote FN au premier tour de l'élection présidentielle de 2002 à 2012, France entière (% aux exprimés, sextiles communs aux trois scrutins)**



Sources : ministère de l'Intérieur, 2002-2012

# Encart 5 : RÉCAPITULONS UN PEU !

Pour réaliser un graphique présentant tous les signes des bonnes pratiques, il faut :

- rechercher la simplicité et la sobriété. À cet égard, les représentations en trois dimensions sont loin de constituer la panacée... ;
- être précis dans la présentation des données (titre, légende, source, etc.) ;
- identifier les types des variables en présence et s'y adapter ;
- tenir compte de la quantité d'information que l'on souhaite représenter (nombre de variables et de modalités), et le cas échéant la simplifier. Si les variables à expliquer sont nominales ou ordinales, il convient parfois de sélectionner une modalité. Attention ici à ne pas dissimuler d'informations importantes, cachées dans les modalités non représentées. Pour ce qui concerne les variables à expliquer quantitatives, il faut choisir un indicateur synthétique approprié (voir chapitre 3) qui constitue un résumé fidèle de la distribution (pour une représentation cartographique : déterminer un mode de discrétisation idoine) ;
- prendre le temps de réfléchir à son unité de mesure : est-elle la plus pertinente pour la question que je me pose (voir l'exemple du taux de nuptialité) ? Ce point doit être considéré s'agissant du choix du maillage géographique pour le cas d'une représentation cartographique ;
- déterminer soi-même les principales caractéristiques de ses axes des ordonnées et des abscisses : bornes minimales et maximales, graduation, etc. Surtout ne pas laisser décider le logiciel à sa place. Celui-ci est programmé pour des mesures par défaut, pas toujours appropriées aux phénomènes étudiés ! Ne pas fabriquer d'effet loupe abusif et s'assurer de la linéarité de sa graduation (le danger est réel dans le cas d'une échelle temporelle) ;
- s'assurer de l'homogénéité de son raisonnement : si une démonstration, un article de revue ou un chapitre d'ouvrage mobilise plusieurs représentations graphiques, les échelles des axes doivent impérativement être comparables tout au long du raisonnement. Un choix différent peut néanmoins être effectué mais ce de façon volontaire et justifié, en particulier concernant une série de représentations cartographiques.

## IV. Les exercices d'application

### IV.1. Le souhait de devenir propriétaire

À partir du tableau 2 suivant, trouver une représentation graphique appropriée sans utiliser d'histogramme.

Tableau 2 : Le souhait de devenir propriétaire de son logement, %

Souhaitez-vous dans l'avenir devenir propriétaire ?	
Non, pas du tout	32
Non, pas vraiment	11
Oui, plutôt	22
Oui, tout à fait	33
NSP	2

Sources : Les Français et le logement, TNS Sofres pour les États généraux du logement avril-mai 2011

### IV.2. Les qualités et défauts des candidats

À partir du tableau 3 suivant, trouver une représentation graphique permettant la comparaison entre les deux candidats au second tour de l'élection présidentielle française de 2012.

Tableau 3 : Qualitatifs des candidats au second tour de l'élection présidentielle de 2012, %<sup>1</sup>

	François Hollande	Nicolas Sarkozy
Il est honnête	59	37
Il est sincère	55	38
Il est sympathique	59	39
Il est dynamique	48	77
Il est compétent	52	56
Il vous inquiète	36	45
Il a la stature présidentielle	46	67
Il a des convictions	68	67
Il tiendra ses engagements	46	38
Il comprend les problèmes des gens comme vous	52	34

Sources : Prédiscopie 2012, Vague 10 post premier tour présidentiel, Cevipof/Monde/Fondapol/Fondation Jean Jaurès/ Ipsos/Logica Business Consulting, avril 2

<sup>1</sup> La question posée était la suivante : « Pour chacune des phrases suivantes, diriez-vous s'applique très bien, assez bien, assez mal ou très mal à chacune des personnalités suivantes ? Le tableau se lit ainsi : 59% des personnes interrogées pensent que la phrase « Il est honnête » s'applique très ou assez bien à F. Hollande. Pour plus d'informations sur cette enquête : <<http://www.cevipof.com/fr/2012/recherche/panel/>>, consulté le 19 mai 2014.



1. La densité de population et le type de logement

À partir du tableau 4 suivant, trouver une représentation graphique mettant de visualiser simultanément la répartition des deux variables densité de population » et « part de maisons dans les logements ».

Tableau 4 : Densité de population et part de maisons dans les logements par département, France métropolitaine

Département	Densité (hab./km²)	Maisons/Logements (%)	Département	Densité (hab./km²)	Maisons/Logements (%)
Ain	101	64,8	Indre-et-Loire	96	63,5
Aisne	73	76,4	Isère	160	49,7
Allier	47	71,4	Jura	52	63,8
Alpes-de-Haute-Provence	23	60,7	Landes	40	71,0
Alpes-Maritimes	252	74,0	Loire	155	50,0
Ardeches	54	24,4	Corse du Sud	184	63,3
Ardennes	56	69,6	Corse du Nord	96	68,1
Ariège	31	80,1	Corse du Sud	51	79,0
Aube	50	66,3	Corse du Nord	33	82,5
Avall	57	71,2	Corse du Sud	61	79,5
Avall	32	72,9	Corse du Nord	15	76,9
Bas-Rhin	229	47,9	Corse du Sud	108	70,5
Bas-Rhône	386	38,6	Corse du Nord	84	78,4
Calvados	122	60,3	Corse du Sud	60	78,4
Canal	26	74,2	Corse du Nord	59	79,9
Canal	50	82,1	Corse du Sud	31	84,2
Canal	80	76,0	Corse du Nord	104	73,7
Canal	43	78,5	Corse du Sud	168	55,2
Canal	41	76,5	Corse du Nord	72	76,7
Canal	35	51,3	Corse du Sud	447	68,6
Canal	22	55,0	Corse du Nord	136	67,8
Canal	60	82,1	Corse du Sud	48	78,6
Canal	85	85,7	Corse du Nord	21060	0,9
Canal	61	85,4	Corse du Sud	219	77,7
Canal	45	81,8	Corse du Nord	79	63,5
Canal	100	51,0	Corse du Sud	85	52,6
Canal	73	63,8	Corse du Nord	107	56,2
Canal	668	48,5	Corse du Sud	520	30,0
Canal	96	76,8	Corse du Nord	65	68,4
Canal	72	74,0	Corse du Sud	90	74,5
Canal	132	73,0	Corse du Nord	220	60,6
Canal	119	62,1	Corse du Sud	199	56,1
Canal	30	81,4	Corse du Nord	6388	26,1
Canal	142	64,3	Corse du Sud	92	75,2
Canal	35	52,2	Corse du Nord	65	78,7
Canal	193	77,1	Corse du Sud	0,3	78,1
Canal	45	72,9	Corse du Nord	213	46,5
Canal	30	73,3	Corse du Sud	535	32,5
Canal	24	42,7	Corse du Nord	932	47,4
Canal	45	75,7	Corse du Sud	168	53,1
Canal	63	41,9	Corse du Nord	151	51,9
Canal	51	80,4	Corse du Sud	92	52,9
Canal	68	65,7	Corse du Nord	81	73,0
Canal	213	52,3	Corse du Sud	65	63,9
Canal	8605	12,2	Corse du Nord	46	77,5
Canal	167	48,5	Corse du Sud	616	81,8
Canal	143	61,6	Corse du Nord	741	43,7
Canal	34	81,8	Corse du Sud	741	43,7

Sources : INSEE, 2008

IV.4. Pour aller plus loin : le taux de croissance du PIB

À partir du tableau 5 suivant, trouver une représentation graphique adaptée.

Tableau 5 : Taux de croissance du PIB en 2012 dans les pays de l'UE27

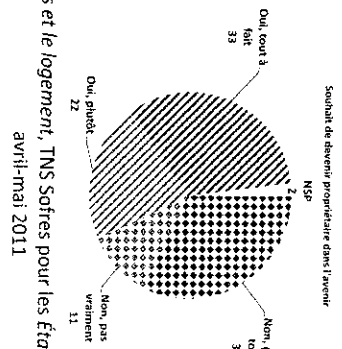
Pays	Taux de croissance du PIB en 2012 (%)	Appartenance à la zone euro en 2012
Allemagne	0,7	zone euro
Autriche	0,8	zone euro
Belgique	0,2	zone euro
Bulgarie	0,8	zone euro
Chypre	-2,4	zone euro
Danemark	-0,6	zone euro
Espagne	-1,4	zone euro
France	0,2	zone euro
Grèce	-6,4	zone euro
Hongrie	-1,7	zone euro
Irlande	0,9	zone euro
Italie	-2,4	zone euro
Lettonie	5,6	zone euro
Lituanie	3,6	zone euro
Luxembourg	0,2	zone euro
Malte	0,8	zone euro
Pays-Bas	1	zone euro
Pologne	2	zone euro
Portugal	-3,2	zone euro
République tchèque	-1,3	zone euro
Roumanie	0,3	zone euro
Royaume-Uni	0,3	zone euro
Slovaquie	2	zone euro
Slovenie	-2,3	zone euro
Suède	0,8	zone euro

Sources : Eurostat, 2012

IV.5. Les corrigés

[IV.1.] Le graphique en secteurs, aussi appelé « camembert » pour sa forme circulaire, est tout à fait adapté pour la représentation de la distribution d'une seule variable nominale. Il peut être amélioré en choisissant un mode de remplissage cohérent pour les réponses positives, les réponses négatives et les sans réponses.

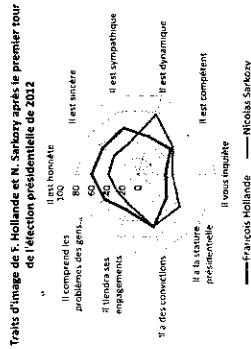
Figure 13 : Le souhait de devenir propriétaire, %



Sources : Les Français et le logement, TNS Sofres pour les États généraux du logement, avril-mai 2011

2.] S'il est possible de représenter ces données sous forme d'histogramme en dix dimensions, une représentation graphique en radar est particulièrement appropriée à la comparaison des deux candidats sur chaque qualificatif. Ce type graphique permet de visualiser directement les qualificatifs sur lesquels tel ou candidat est meilleur que l'autre.

Figure 14a : Qualificatifs des candidats au second tour de l'élection présidentielle de 2012, %



Sources : Présidescopie 2012, Vague 10 post premier tour présidentiel, Cevipof/Le Monde/Fondapol/Fondation Jean Jaurès/Ipsos/Logica Business Consulting, avril 2012

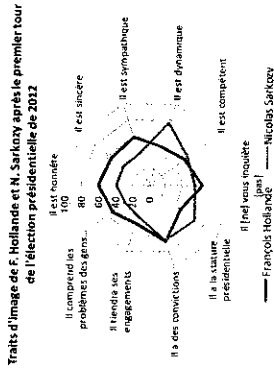
Plus généralement, ce type de graphique permet de comparer :

- différents critères entre eux (ici les qualificatifs) et leur niveau selon différentes modalités (ici les candidats) ;
- une situation observée à une situation idéale.

Si cette seconde utilisation est peu utilisée en sociologie, elle peut être utile dans certains cas, par exemple comparer les distributions sociodémographiques servées dans son échantillon aux vraies valeurs dans la population d'intérêt.

Pour aller plus loin : l'item « Il vous inquiète » est négatif contrairement aux autres. Il serait donc judicieux de représenter l'opinion complémentaire<sup>1</sup> pour cet item afin de n'observer que des opinions positives sur le graphique en radar et en illustrer ainsi la lecture (figure 14b).

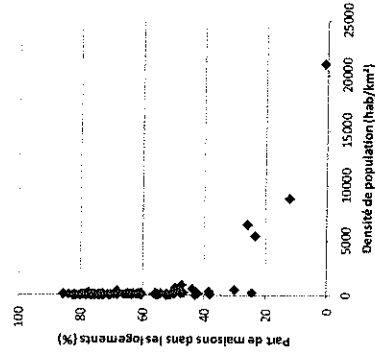
Figure 14b : Qualificatifs des candidats au second tour de l'élection présidentielle de 2012, %, item « Il vous inquiète » inversé



Sources : Présidescopie 2012, Vague 10 post premier tour présidentiel, Cevipof/Le Monde/Fondapol/Fondation Jean Jaurès/Ipsos/Logica Business Consulting, avril 2012

[IV.3.] Le nuage de points est adapté pour représenter la distribution de variables quantitatives.

Figure 15a : Densité de population et part de maisons dans les logements par département, France métropolitaine

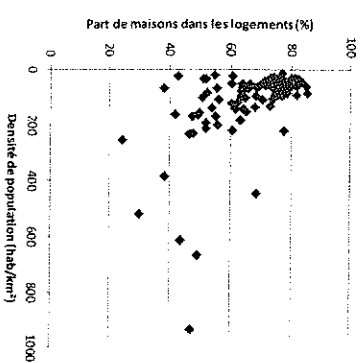


Sources : INSEE, 2008

Sous cette forme, le graphique fait clairement apparaître plusieurs outliers, correspondant à Paris et aux trois départements de la petite couronne francilienne. Ces quatre départements spécifiques doivent donc être étudiés à part car ils rendent la représentation graphique illisible.

<sup>1</sup> Soit les réponses correspondant à « s'applique assez ou très mal » concernant cet item pour chacun des 2 candidats (64% pour F. Hollande et 55% pour N. Sarkozy contre respectivement 36% et 45%).

Figure 15b : Densité de population et part de maisons dans les logements par département, France métropolitaine, hors-Paris et petite couronne

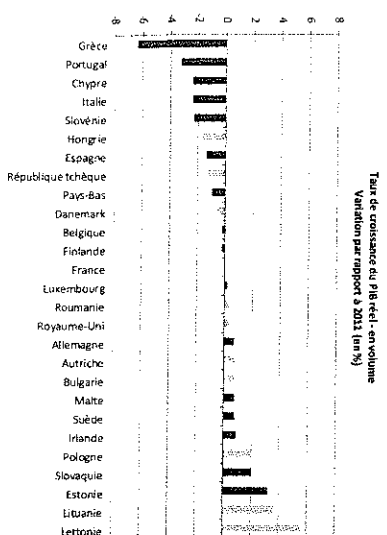


Sources : INSEE, 2008

Le coefficient de corrélation (cf. chapitre 5) constitue un complément indispensable pour l'interprétation du lien entre ces deux variables.

[IV.4.] Il existe plusieurs solutions en fonction de l'information que l'on souhaite mettre en valeur. S'il s'agit de séparer les pays à croissance négative de ceux à croissance positive, il faut trier son fichier de données selon les valeurs du taux de croissance du PIB. La visualisation de l'appartenance ou non à la zone euro, secondaire dans ce cas, peut se faire en coloriant différemment les barres de l'histogramme selon que le pays correspondant appartient ou non à la zone euro.

Figure 16 : Taux de croissance du PIB en 2012 dans les pays de l'UE27



Sources : Eurostat, 2012

Dans le graphique ci-dessus, les pays appartenant à la zone euro sont représentés en noir plein, tandis que ceux n'appartenant pas à la zone euro sont représentés en gris clair.

## V. Les articles d'application conseillés

- Michel Bussi, Jérôme Fourquet, Céline Colange, « Analyse et compréhension du vote lors des élections présidentielles de 2012 : l'apport de la géographie électorale », *Revue française de science politique*, vol. 62, n°5, 2012, p. 941-963.
- François Héran, « Les animaux domestiques », *Données sociales 1987*, INSEE, p. 417-423.
- Olivier Monso, Thibault de Saint Pol, « La perception graphique : mieux construire et interpréter les graphiques », *Courrier des statistiques*, n°126, 2009, p. 67-74.

