

Annexe V

Élaboration d'une carte thématique

A. Introduction

A5.1. La présente annexe propose un bref aperçu des éléments liés à la conception à considérer pour créer des cartes thématiques. Il est impossible de traiter de toutes les questions relatives au contenu informatif des cartes dans le cadre de cet aperçu; il convient dès lors de consulter un manuel au besoin. Les cartographes distinguent différents types de cartes. Les cartes générales servent de cadre de référence à des fins d'orientation. Elles présentent surtout les objets géographiques réels que l'on peut observer sur le terrain. Ces objets sont soit naturels — cours d'eau, montagnes, littoraux — soit artificiels, comme les routes ou les localités. Les cartes de référence montrent aussi des objets qui ne sont pas visibles sur le sol. Le meilleur exemple en est les frontières politiques et le réseau de référence montrant les latitudes et les longitudes. Les cartes topographiques entrent dans la catégorie des cartes générales ou des cartes de référence. Elles jouent un rôle important dans la cartographie des secteurs de recensement car elles fournissent des informations sur les objets que l'agent recenseur utilise pour s'orienter dans le secteur de travail qui lui est affecté.

A5.2. Les *cartes thématiques* présentent plus d'intérêt lorsqu'il s'agit d'exposer des résultats de recensement. Elles montrent la répartition géographique des phénomènes physiques et culturels qu'il n'est pas facile d'observer directement sur le sol. Elles peuvent être fondées sur des informations qualitatives ou quantitatives. On peut citer à titre d'exemple de cartes du premier type une carte montrant la répartition de la population par langue maternelle ou par religion. Les cartes thématiques quantitatives, appelées parfois cartes statistiques, fournissent par contre des informations sur les dimensions relatives des objets cartographiés. On peut citer à titre d'exemple une carte sur laquelle les symboles représentant les villes d'un pays sont gradués en fonction des dimensions de chaque ville. Un autre exemple est celui d'une carte sur laquelle des unités déclarantes telles que des districts sont ombrées en fonction de leur densité démographique. La plupart des cartes établies pour un atlas de recensement sont de ce type.

B. Principes de la conception des cartes

A5.3. Bien que fréquemment utilisées à des fins d'analyse, les cartes se prêtent mal à la présentation de données exactes. Sur une carte, les valeurs de données sont exprimées par des symboles. Le cartographe doit affecter les valeurs de données à des intervalles de classes pour obtenir un nombre raisonnable de catégories représentées sous forme de couleurs ou de symboles. Cela signifie qu'une partie des informations est perdue dans la présentation de la carte. Si les cartes permettent de présenter correctement des tendances, des grandeurs relatives et la répartition des valeurs d'indicateurs, on obtient de meilleurs résultats avec les tableaux de données ou les cartes numériques dont la base de données peut être interrogée si l'on est intéressé par des valeurs exactes.

A5.4. L'établissement de cartes de présentation est l'opération par laquelle le cartographe communique une idée ou un concept au lecteur (Monmonier, 1993). Elle est comparable à d'autres modes de communication d'informations qualitatives ou quantitatives sous forme graphique au moyen de diagrammes, d'images ou d'autres illustrations. En conséquence, les mêmes principes de conception qui président à la conception graphique s'appliquent aussi à la cartographie.

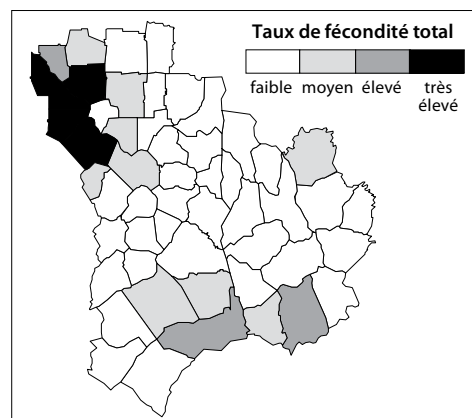
A5.5. La simplicité et la clarté sont les principes essentiels en matière de conception. De nombreuses cartes sont encombrées de détails parce que le cartographe a voulu présenter trop de choses sur un petit espace. Tufte (1983) offre un concept utile avec son rapport « maximisation des données/consommation d'encre » : appliqué à la cartographie, il signifie que l'encre utilisée doit servir à représenter des données géographiques plutôt que des informations non essentielles. Il faut donc renoncer à fournir des informations superflues. Les titres commençant par « carte de ... » ou « légende » sont inutiles, de même que de nombreux encadrés, lignes et souvent mais pas toujours, les flèches indiquant le nord et l'échelle. Comme la plupart des principes, celui-ci a naturellement ses limites. Certains éléments de la carte, comme la légende elle-même, un titre concis et des informations sur la source sont évidemment requis pour comprendre la carte.

A5.6. La simplicité signifie aussi qu'aucun espace ne doit être perdu. Aujourd'hui, on dispose presque partout d'imprimantes laser à haute résolution et par conséquent les cartes n'ont pas besoin d'être imprimées en très grand format pour montrer tous les détails. La carte peut être imprimée dans un format d'autant plus réduit qu'elle est mieux dessinée. Mais économiser l'espace signifie aussi éviter des polices, des symboles de légendes ou des encarts disproportionnés.

A5.7. La notion de hiérarchie visuelle est un autre concept important. Elle s'applique à tous les éléments figurant sur la carte elle-même ainsi qu'à la disposition de toutes ses composantes. Sur la carte elle-même, le choix des couleurs ou des symboles reflète l'ordre de valeurs des données. Sur une carte de mortalité infantile, par exemple, les unités déclarantes présentant les valeurs les plus élevées pourront être ombrées avec la couleur la plus brillante ou la nuance de gris la plus sombre. Ce sont les points chauds qui doivent retenir immédiatement l'attention du lecteur. Sur la figure A5.1, par exemple, les catégories de taux faibles à élevés sont volontairement ombrées en nuances de gris clair pour faire ressortir la catégorie des « taux très élevés ». C'est le contraste entre les couleurs sombres et les nuances plus délicates qui crée la hiérarchie visuelle. Une zone relativement claire entourée de teintes sombres ressortira tout autant. Le choix des couleurs est examiné ci-après plus en détail.

Figure A5.1

Établissement d'une hiérarchie visuelle par le choix des couleurs ou des nuances de gris



A5.8. Le cartographe peut aussi recourir à d'autres stratagèmes pour diriger l'attention du lecteur vers une zone particulière de la carte. Il peut entourer d'un contour précis les objets les plus importants sur la carte, par exemple, en les faisant ressortir par rapport à l'arrière-plan. On utilise parfois des annotations ou des flèches orientées vers certains objets, mais cela a souvent pour effet d'encombrer la carte.

A5.9. Les mêmes principes s'appliquent à la composition d'ensembles de la carte. Les éléments les plus importants de la carte sont l'information cartographique elle-même, le titre et la légende expliquant les symboles utilisés. Ils doivent être les éléments les plus apparents sur la page de la carte. On n'introduira d'autres éléments qu'avec prudence.

A5.10. L'utilisation de la couleur dans les cartes thématiques statistiques implique plusieurs considérations sous-jacentes (voir également par. A5.4 ci-dessus). Les cartographes doivent avoir conscience des corrélations négatives entre une présentation visuelle recourant à la couleur et les coûts supplémentaires liés à l'encre et à d'autres fournitures. Le daltonisme ne doit pas être perdu de vue. Les cartes devraient éviter l'utilisation du rouge et du vert pour des valeurs de données, puisque les daltoniens ne peuvent pas les distinguer normalement. Plus généralement, une carte couleur doit être reproductible sur une imprimante en noir et blanc sans perte du contenu informatif. Cela signifie qu'il faut respecter la densité de l'ombrage et remplacer les configurations de remplissage au besoin. Enfin, les cartographes doivent veiller à ne pas heurter une partie quelconque de la population par leurs choix de conception. Ils doivent être au fait des sensibilités des diverses régions ou groupes de la population. Certains symboles ou couleurs peuvent s'accompagner de connotations positives ou négatives pour différents groupes ethniques ou raciaux du pays. Lors de la conception d'une carte, les symboles associés à des stéréotypes concernant un sous-groupe quelconque de la population doivent être évités. Et surtout, les cartographes doivent être conscients que les couleurs doivent être utilisées judicieusement et non gratuitement.

1. Éléments d'une carte thématique

A5.11. Une carte thématique est constituée de plusieurs éléments. La carte elle-même consiste en une carte de base indiquant les limites des secteurs présentant de l'intérêt, par exemple, les frontières du pays et éventuellement des objets de référence tels que les principaux cours d'eau ou les grandes villes. Ils orientent le lecteur qui souhaite comparer la valeur d'une variable dans une partie du pays à cette même valeur dans une autre partie. Le deuxième élément principal est le recouvrement de la carte thématique qui présente la répartition géographique de la variable.

A5.12. Outre les données cartographiques réelles, une carte susceptible d'être publiée comprend d'autres éléments. Il peut s'agir des éléments suivants :

- Les **titres et les sous-titres** doivent être courts et très descriptifs.
- Les **sources de données, remerciements et date d'établissement de la carte** renseignent l'utilisateur sur la fiabilité et la crédibilité de la carte. Il convient par ailleurs d'ajouter les informations relatives aux droits d'auteur. Certaines institutions qui établissent régulièrement des cartes ajoutent aussi des numéros de référence et de version à usage interne. Il convient d'y ajouter des explications importantes pour permettre de comprendre le contenu de la carte. Dans le cas des cartes imprimées en grand format, les paramètres de la projection cartographique doivent aussi être indiqués.
- La **légende de la carte** décrit la façon dont les valeurs de la variable cartographiée ont été traduites en symboles cartographiques, par exemple, quelles

couleurs sont utilisées pour représenter un intervalle donné de valeurs de densité démographique. Il importe de toujours inclure les unités de mesure dans la légende, par exemple le « nombre d'habitants au kilomètre carré ».

- **L'échelle de la carte** permet au lecteur de mesurer les distances sur la carte. Pour une série de cartes thématiques comme celles d'un atlas de recensement, où toutes les cartes sont tracées à la même échelle, cette information n'a pas besoin de figurer sur chaque page. Il en va de même pour les cartes de dimensions relativement réduites de régions connues où il est peu probable que le lecteur veuille effectuer des mesures de distances. Il est en général préférable d'ajouter une échelle graphique plutôt que de spécifier une échelle numérique (par exemple 1/1 000 000). Si la carte est réduite ou agrandie pendant la photocopie, l'échelle graphique reste utilisable. Par contre, l'échelle cartographique nominale utilisée pour tracer la carte initiale sera erronée.
- Une **flèche indiquant le nord** n'est pas absolument nécessaire sur une carte de référence étant donné que toutes les cartes sont orientées vers le nord. C'est particulièrement vrai si la carte montre une zone géographique connue telle qu'un pays tout entier. Si l'on fait tourner la carte pour mieux l'adapter à la page, on devra dans tous les cas inclure une flèche indiquant le nord.
- Les **limites et bords intérieurs de cartes** servent à séparer différents éléments de la carte; l'utilisation de ces éléments graphiques est surtout une question de conception. On a une impression de fouillis si la carte comporte un trop grand nombre de lignes et d'encadrés. On ne devra donc utiliser des limites supplémentaires que si les éléments de la carte ne sont pas suffisamment séparés.
- Des **noms de lieux et des étiquettes** servant à identifier des objets géographiques ou de zones statistiques.
- Un **graticule**, c'est-à-dire le réseau des latitudes et des longitudes (parallèles et méridiens) qui facilite l'orientation sur la carte. Ces éléments doivent être inclus sur les cartes à petite échelle.
- Des **cartes de positionnement** sont utilisées pour indiquer l'emplacement de la zone couverte par la carte principale. Par exemple, une carte de densité démographique au niveau du district peut être accompagnée d'une petite carte montrant l'emplacement de ce district dans le pays ou la province.
- Les **cartons intérieurs** sont semblables aux cartes de positionnement. Mais au lieu de montrer l'emplacement de la zone couverte par la carte principale, ils montrent une portion réduite de la carte à une échelle cartographique supérieure. Par exemple, une carte au niveau d'une province peut être accompagnée d'un petit carton intérieur montrant la région de la capitale ou fournissant des informations plus détaillées sur un district moins étendu.
- Du **texte et des annotations** fournissent des informations ou des explications de base qui doivent être courtes et appropriées.
- Des **éléments graphiques supplémentaires** peuvent inclure un histogramme montrant la répartition géographique de la variable ou le symbole du bureau qui a établi la carte.

A5.13. Les figures A5.2 et A5.3 montrent deux exemples de cartes comprenant beaucoup d'éléments de la carte thématique. La figure A5.2. est une carte d'unités administratives de premier niveau dans le pays hypothétique de Poplandia. La carte est couverte d'un réseau de latitudes et de longitudes définissant les références géodésiques. La capitale nationale, les chefs-lieux administratifs de provinces et les principaux

cours d'eau sont ajoutés à titre de référence. Tous les objets sont étiquetés correctement en utilisant différentes polices pour divers types d'objets. La marge au-dessous de la carte comporte une échelle graphique, la légende indiquant les types d'objets thématiques représentés et l'origine de la carte. Si le bureau statistique dispose d'un logo, celui-ci peut également être ajouté sur chaque carte. Une flèche indiquant le nord a été omise pour deux raisons. La carte ne présente pas d'orientation particulière et les lignes représentant la longitude indiquent clairement que le nord se situe au haut de la carte. L'autre raison, moins évidente, est que la projection géodésique utilisée pour la carte fait converger les latitudes vers le nord. Cela sous-entend que le nord se trouve dans une direction légèrement différente aux différentes longitudes.

Figure A5.2

Échantillon de carte des unités administratives et des principaux centres urbains

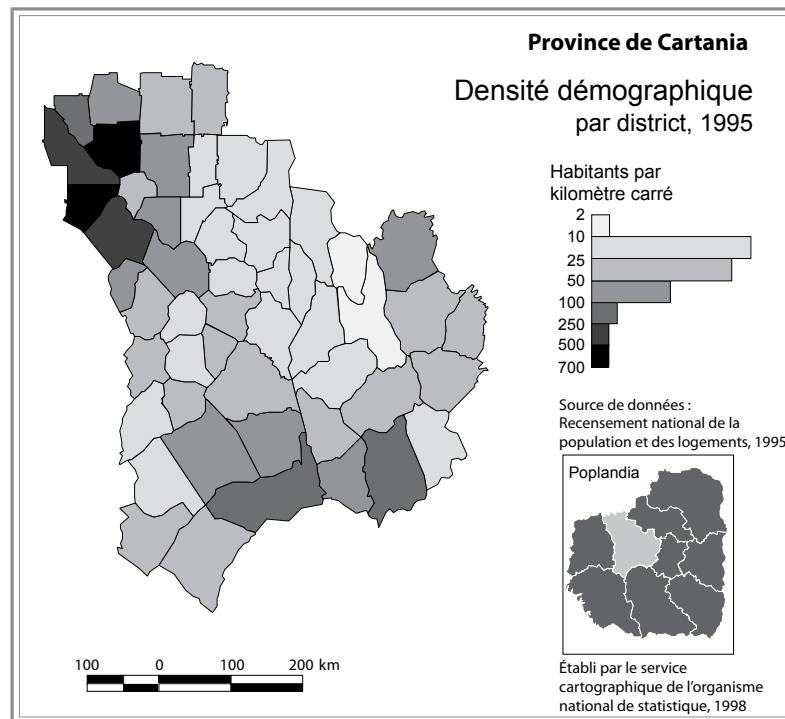


A5.14. La carte thématique représentée sur la figure A5.3. montre la densité démographique dans une des provinces de Poplandia. Une carte de ce type pourrait par exemple accompagner des tableaux montrant les caractéristiques démographiques par province dans une publication censitaire. Une telle carte pourrait par ailleurs comporter des graphiques ou d'autres représentations graphiques. Le tracé de la carte est très simple. Le titre décrit le thème de la carte et le sous-titre indique la zone géographique. Au lieu de la légende traditionnelle présentant des couleurs dans des cases de dimensions égales, la légende de cette carte montre les catégories de densité démographique sous la forme d'un histogramme. Celui-ci joue le rôle d'une légende traditionnelle — associant les valeurs à des nuances de couleurs — et il indique en outre la répartition

de fréquence des valeurs du district. Pour des cartes plus complexes, comportant plusieurs zones, on peut ajouter le nombre réel de districts relevant de chaque catégorie. Par souci de simplicité et de clarté, cela n'a pas été fait dans le cas de la présente carte. Au-dessous de la légende et de l'indication de la source, une petite carte de localisation montre l'emplacement de la province de Cartania dans le pays. Il est souvent inutile d'ajouter des annotations à une carte de localisation représentant le pays, parce que la forme du pays est facilement reconnaissable pour les lecteurs.

Figure A5.3

Exemple de carte thématique représentant la densité démographique



2. Niveaux de mesure et variables graphiques

a) Dimensions spatiales

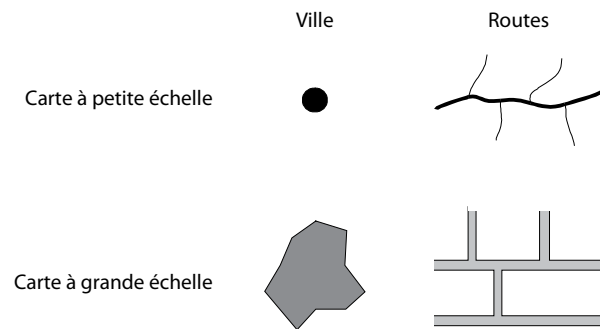
A5.15. Les cartes thématiques n'indiquent pas seulement l'emplacement d'un objet, elles fournissent aussi des informations concernant cet objet — la valeur de la variable en chaque point géographique. Une carte thématique est donc composée d'éléments géographiques et de certains attributs correspondant à ces éléments. Cela signifie qu'en dessinant une carte thématique, nous devons envisager la dimension spatiale des objets géographiques et nous devons avoir présent à l'esprit le niveau de mesure de chaque variable. Ces deux facteurs déterminent les options cartographiques disponibles pour produire une carte attrayante sur le plan visuel, précise et facile à interpréter.

A5.16. Les objets géographiques sont représentés dans une base de données géographique par des primitives géométriques : des points, des lignes et des surfaces. D'autres catégories, moins souvent utilisées en cartographie, ajoutent une troisième et une quatrième dimensions : le volume et l'espace-temps. La forme géométrique utilisée pour un objet du monde réel dépend parfois de l'échelle spatiale de la carte ou

de l'ensemble de données. Par exemple, on pourra représenter un village ou une ville comme des surfaces sur des cartes à grande échelle mais ils seront représentés sous la forme d'un point sur des cartes à plus petite échelle au niveau de la province ou du pays (voir figure A5.4). Une route pourra être représentée par une ligne sur une carte de province, mais elle prendra la forme d'une double ligne — c'est-à-dire un objet surfacique — sur une carte de ville.

Figure A5.4

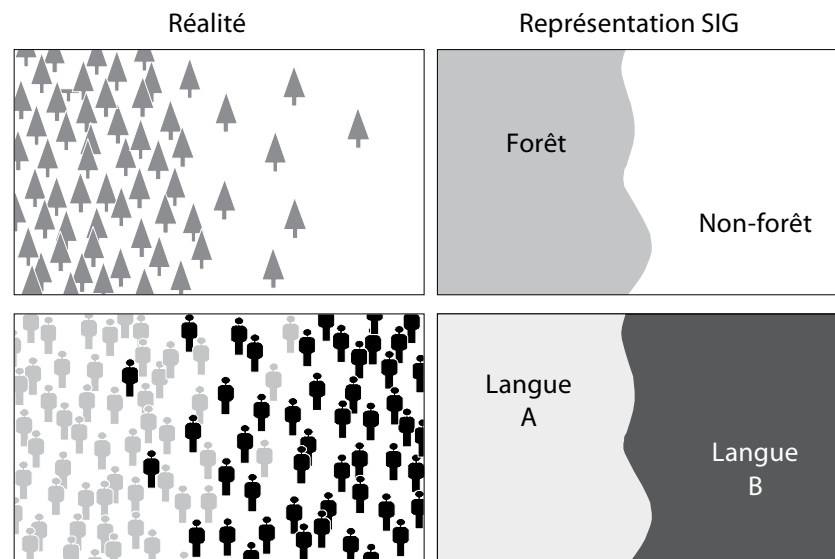
Effets de la généralisation sur l'affichage d'objets spatiaux



A5.17. Il importe de se souvenir que les limites et les emplacements ne sont pas toujours aussi clairement définis qu'ils le paraissent dans la représentation discontinue d'une carte ou d'une base de données géographique. Les objets complexes du monde réel ont souvent besoin d'être généralisés, simplifiés ou abstraits pour pouvoir être représentés dans la base de données d'un ordinateur. Par exemple, de nombreux objets du monde réel n'ont pas de limites précises. Il existe souvent une zone de transition entre la forêt et la non-forêt. Si la forêt est représentée sous la forme d'un objet surfacique (et non sous forme de points correspondant à chaque arbre), il y aura nécessairement une perte d'informations (voir figure A5.5).

Figure A5.5

La complexité du monde réel doit parfois être simplifiée pour faciliter sa représentation dans un SIG



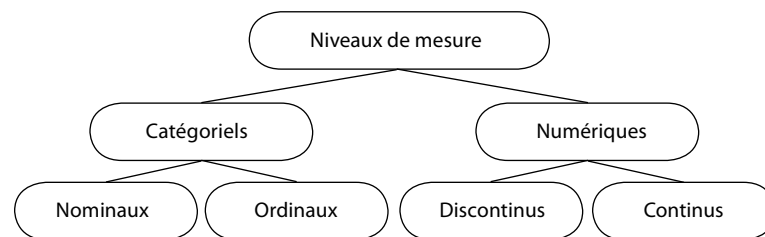
A5.18. Dans le domaine socioéconomique on a des exemples de limites mal définies avec la répartition des groupes ethniques ou linguistiques. Bien que le mode de répartition de ces groupes soit parfois très bien défini, il existera probablement à la périphérie de chaque région des zones où se mêlent les divers groupes ethniques ou linguistiques. Les cartographes utilisent parfois des lignes pointillées pour représenter ces limites mal définies, bien que cela ne résolve pas le problème de l'emplacement de la limite sur la carte.

b) Niveaux de mesure

A5.19. Il faut mentionner un autre facteur important : la façon dont est mesurée la variable dont on souhaite établir la carte. Il faut surtout distinguer les informations numériques de celles relatives à des catégories (voir figure A5.6). Les données relatives à des catégories se subdivisent à leur tour en données nominales ou ordinales. Les données nominales ou qualitatives décrivent simplement un type d'objet, sans qu'il y ait un ordre naturel entre ces catégories. On peut citer par exemple les différents types de maisons, par exemple les maisons en pierre ou en bois. Par contre, les données ordinales supposent une hiérarchisation des catégories, même si l'on ne connaît pas l'intervalle qui les sépare. Par exemple, à partir des réponses obtenues suite à une enquête, nous pourrions classer les ménages en fonction de leur niveau de bien-être, faible, moyen ou élevé. Mais nous ignorons si la différence entre un niveau faible et un niveau moyen est la même qu'entre un niveau moyen et un niveau élevé.

Figure A5.6

Mesure des variables



A5.20. Si nous pouvons quantifier la différence entre catégories, nous avons affaire à des données numériques. Les données discontinues correspondent à des dénombrements, par exemple le nombre de chambres par ménage, ainsi que la population totale. Les variations continues ou variables proportionnelles peuvent prendre toute valeur souhaitée. Elles peuvent être mesurées avec une grande précision. En ce qui concerne les données de recensement, les variables continues sont en général des indicateurs calculés pour des unités de recensement agrégées telles que la densité démographique, la proportion de population ayant accès à l'eau potable ou le taux de fécondité total.

c) Variables graphiques

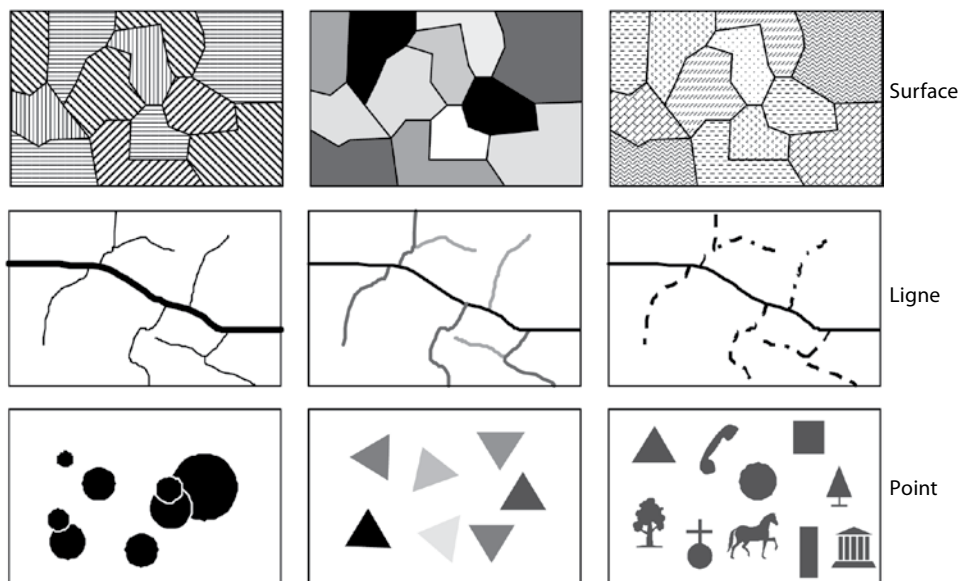
A5.21. Sur une carte thématique, les symboles graphiques indiquent au lecteur des différences de valeur ou de catégorie entre des objets géographiques. Les concepts de symbolisation appliqués à la cartographie sont semblables à ceux élaborés pour les applications infographiques par Bertin (1983; voir aussi MacEachren, 1995). Bertin distingue les variables graphiques suivantes :

- La **taille**, indicateur de différences ordinales ou numériques. La taille est très importante pour les objets ponctuels ou linéaires, pour indiquer par exemple la dimension des villes à l'aide de cercles gradués ou l'importance de la migration entre régions à l'aide de lignes ou de flèches d'épaisseur variable.
- L'**orientation** est utilisée, par exemple, pour les hachures d'objets surfaciques. De même, des objets ponctuels géométriques peuvent être présentés dans différentes orientations. L'orientation n'implique aucune différence concernant la grandeur d'une variable et elle est donc utile pour présenter des données nominales.
- La **texture** correspond à la densité d'un motif régulier qui varie suivant les zones. Elle peut servir à représenter des différences ordinales ou numériques. Ce peut être un raccourci utile lorsque les périphériques de sortie n'ont que des capacités limitées pour imprimer des couleurs ou des nuances de gris. La texture est aussi très utile pour représenter une information stratifiée dans laquelle deux variables sont figurées l'une au-dessus de l'autre. Mais il n'est pas facile de préserver la clarté de ces cartes qui sont donc utiles surtout pour des applications d'analyse exploratoire.
- La **forme** est très importante pour les objets ponctuels. Les ensembles de symboles et de polices présents dans les logiciels commerciaux de SIG et les logiciels de cartographie informatique comportent un grand nombre de symboles différents. Les symboles les plus connus en cartographie sont ceux qui représentent des bâtiments tels que des lieux de culte ou des hôpitaux.
- La **couleur** est tout à fait appropriée pour illustrer des différences numériques et dans une certaine mesure géographiques. Le choix de la couleur est l'une des questions les plus importantes en ce qui concerne la conception d'une carte et c'est pourquoi il est examiné plus loin de façon plus détaillée.

A5.22. En principe, chacune de ces dimensions graphiques peut servir à caractériser les divers types d'objets géographiques — à savoir les objets ponctuels, linéaires et surfaciques. Mais dans la plupart des cas, on utilise seulement un sous-ensemble de variables graphiques pour les différents types d'objets. La figure A5.7 en fournit

Figure A5.7

Variables graphiques correspondant à des polygones, des lignes et des points



quelques exemples. Les variables graphiques d'une carte thématique sont choisies de façon à correspondre au type de mesure utilisé pour l'indicateur faisant l'objet de la carte. Par exemple, les dimensions et la couleur sont très importantes pour représenter des valeurs numériques. Les formes des symboles ponctuels ou la texture des objets surfaciques représentent des valeurs nominales différentes.

3. Types de cartes thématiques

a) Représentation cartographique d'objets discontinus

A5.23. Les données de recensement compilées en vue de leur publication sont constituées de nombres agrégés par unité déclarante telle que le district ou le secteur de recensement. Le meilleur moyen de représenter ces données consiste dans l'utilisation de cartes choroplèthes. Le terme choroplèthe vient des mots grecs « choros » (lieu) et « plèthe » (valeur). Les cartes choroplèthes montrent des données correspondant à des unités déclarantes discontinues souvent définies indépendamment de la répartition spatiale effective des données (par exemple des limites administratives). Le symbole — c'est-à-dire la couleur ou la texture — utilisé pour ombrer chaque unité déclarante, est déterminé par la valeur. Les cartes choroplèthes diffèrent des cartes de catégories de zones pour lesquelles les unités déclarantes sont déterminées par les données. Par exemple, sur une carte représentant un couvert forestier, les unités déclarantes seront définies à partir des limites séparant les zones forestières et les zones non forestières.

A5.24. Un exemple de carte choroplèthe a déjà été présenté à la figure A5.3. Les cartes choroplèthes sont construites en subdivisant d'abord en un ensemble de catégories toute la série de valeurs de données correspondant aux unités déclarantes. On assigne ensuite une couleur ou un type d'ombrage à chaque catégorie. Étant donné que les données d'une énumération suivent un ordre naturel, le choix des couleurs ou des tons répond à une certaine logique, par exemple une progression des nuances de clair à sombre ou des motifs d'irrégulier à régulier. L'objectif consiste à donner au lecteur une idée intuitive de la grandeur de la valeur dans chaque unité déclarante. Il existe de nombreuses façons différentes de déterminer les symboles utilisés pour ombrer les cartes choroplèthes. Le choix dépend du type de variable, de l'intervalle des valeurs de données et aussi du périphérique de sortie utilisé pour présenter la carte. Le choix des symboles étant très important, il sera examiné en détail dans la prochaine section.

A5.25. Les cartes choroplèthes donnent de bons résultats lorsqu'il s'agit de représenter la répartition générale des valeurs de données sur une carte ou de comparer les répartitions entre différentes cartes. Il n'est en général pas possible d'obtenir la valeur exacte de chaque unité déclarante parce que les couleurs ou nuances représentent seulement des intervalles de valeurs semblables. Pour présenter les valeurs exactes, il est préférable d'utiliser des tableaux de données ou de procéder à une requête interactive dans un SIG.

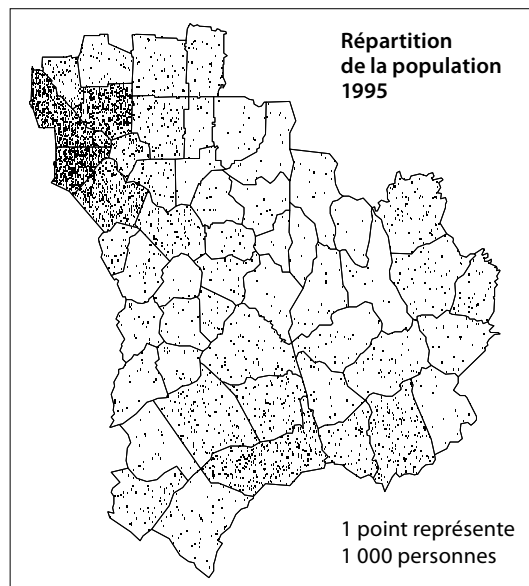
A5.26. Les valeurs utilisées pour produire des cartes choroplèthes sont presque toujours des coefficients, des proportions ou des densités. Il peut s'agir de coefficients géographiques, où une valeur de données telle que la population est divisée par la surface pour calculer la densité démographique. Ou il peut s'agir de coefficients généraux dans lesquels le dénominateur est une valeur autre qu'une surface, par exemple le taux de natalité brute exprimé par le nombre de naissances par 1 000 habitants. Le plus souvent, lorsqu'on établit une carte de variables socioéconomiques, la dimension des unités déclarantes n'est pas constante. Par exemple, il existe souvent de très grandes différences de dimensions et de population entre divers districts et provinces. Si nous voulions établir la carte d'une variable de dénombrement telle que la population totale

plutôt qu'un coefficient, les districts les plus grands seraient probablement ombrés avec les couleurs les plus sombres, même si leur population était peu nombreuse par rapport à leur superficie. Le recours aux cartes choroplèthes est donc inapproprié si l'on veut représenter des valeurs absolues.

A5.27. Une autre méthode utilisable pour afficher des données de dénombrement est celle des cartes de points. Ce type de carte a été utilisé pour la première fois en France en 1830 pour établir une carte montrant la répartition de la population. Sur des cartes à points, le symbole d'un point sert à représenter une ou plusieurs unités d'une variable cartographiée. Par exemple, chaque point peut représenter 10 000 personnes ou 1 000 ménages. La grandeur de la variable est alors représentée par la densité plus ou moins élevée des points dans l'unité déclarante. La figure A5.8 montre une carte à points typique, présentant une répartition de la population.

Figure A5.8

Carte de densité par points



A5.28. Pour placer les points sur la carte, deux méthodes sont applicables. Le cartographe, connaissant la répartition effective de la population dans chaque district, peut choisir l'emplacement des points à partir de celle-ci. Par exemple, il placera davantage de points dans les zones urbaines et autour de celles-ci que dans les régions rurales moins peuplées. Pour certaines applications, on a fait appel à des cartes d'utilisation des sols ou de couverture du sol pour aider à déterminer les densités de points dans chaque unité déclarante. On peut aussi recourir à des masques virtuels pour qu'aucun point ne soit placé dans des zones dont on sait qu'elles sont inhabitées, comme les nappes d'eau, les forêts très denses ou les réserves naturelles protégées.

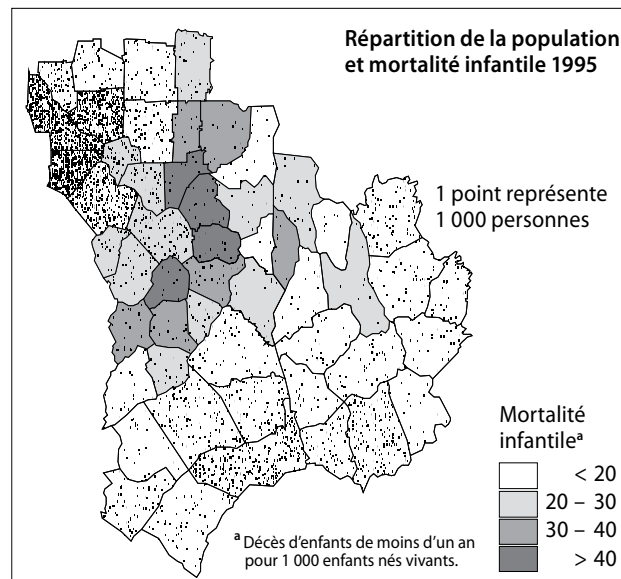
A5.29. La seconde possibilité consiste à placer les points aléatoirement dans chaque district. Dans ce cas, la densité des points reflète la densité d'ensemble de la valeur. Les logiciels de SIG et de cartographie sur ordinateur offrant des fonctions de représentation de la densité placent généralement les points de façon aléatoire. L'utilisateur ne peut choisir que la dimension de chaque point ou le symbole utilisé pour les points. Ce symbole peut être choisi de façon à représenter la variable faisant l'objet de la carte, mais un simple point offre généralement la représentation la plus claire.

A5.30. Certains programmes spécialisés écrits permettent de placer les points en recourant à d'autres couches de la carte. Le placement manuel des points compte tenu de la connaissance que le cartographe peut avoir de la variable est naturellement très fastidieux.

A5.31. Les cartes de points constituent un moyen efficace pour représenter les informations relatives à la densité, à condition que le placement des points soit guidé par la répartition géographique réelle de la variable représentée ou que la répartition dans chaque unité déclarante soit dans l'ensemble homogène. Cette méthode présente le grand avantage que les cartes de densité par points se prêtent bien à la reproduction, qu'elles soient photocopiées ou imprimées, car il s'agit essentiellement de cartes monochromes (en noir et blanc). Les cartes de densité par points peuvent être utilisées en combinaison avec des cartes choroplèthes pour illustrer simultanément deux variables — par exemple la carte de la figure A5.9 montre qu'il n'existe pas de relation entre les densités démographiques élevées et les taux élevés de mortalité infantile. Dans ce cas, la densité des points ne devra pas être très élevée pour que l'on puisse facilement déterminer les couleurs ou les nuances correspondant aux districts sous-jacents.

Figure A5.9

Combinaison de cartes de points et de cartes choroplèthes



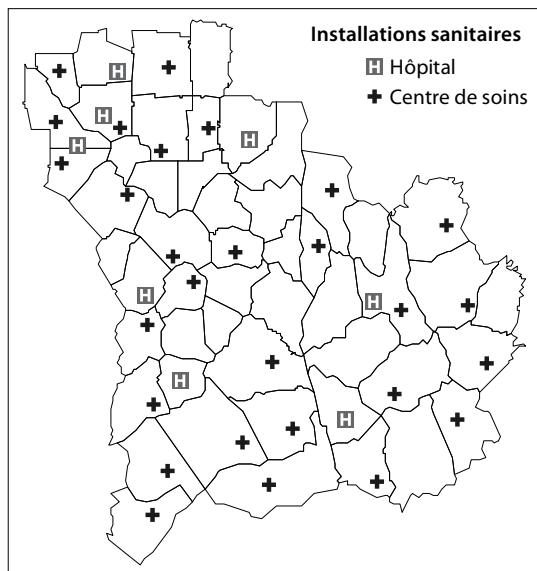
b) Données ponctuelles nominales

A5.32. Le type le plus simple de carte de points est celui où chaque point représente un élément distinct comme une exploitation agricole ou un hôpital. Ces données ponctuelles nominales représentent des catégories d'objets plutôt qu'un attribut compte ou dimension. Dans les cartes de symboles ponctuels simples, l'emplacement du point représente correctement l'emplacement de l'élément. La dimension, la couleur ou le symbole utilisé peuvent refléter divers types d'objets, tels que des centres de services de santé distincts des hôpitaux, comme sur la figure A5.10. On pourra utiliser des figures géométriques simples comme des cercles, des carrés et des triangles pour représenter divers types d'objets ponctuels. Inversement, les logiciels de cartographie numérique ou de SIG permettent à l'utilisateur de spécifier un symbole qui corresponde au type d'objet représenté. Par exemple, la carte figurant sur la figure A5.10 montre la réparti-

tion de deux types d'installations sanitaires au moyen de symboles faciles à interpréter. Les symboles utilisés sont généralement des caractères d'une police de caractères ou des dessins en mode point. La plupart des logiciels ont leurs propres polices qui offrent un grand nombre de symboles cartographiques classés par thème, comme les transports, les services collectifs ou établissements ouverts au public. Certains systèmes permettent aussi à l'utilisateur d'importer des symboles en mode point conçus par lui. Le mieux consiste toutefois à utiliser un jeu de symboles standard chaque fois que cela s'avère possible afin de réduire le besoin de spécialisation nécessaire pour la reproduction.

Figure A5.10

Représentation d'objets ponctuels distincts



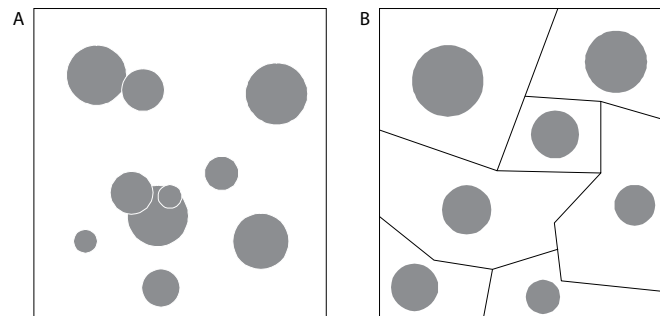
c) Symboles ponctuels proportionnels

A5.33 Les symboles ponctuels peuvent aussi servir à représenter une quantité en un certain lieu. Un type apprécié de carte de recensement, par exemple, montre l'emplacement et la dimension des grandes villes au moyen de cercles et de carrés gradués en fonction des valeurs numériques de chaque objet. Ces cartes sont qualifiées de proportionnelles ou de cartes de symboles gradués. Les cartes de symboles gradués sont utiles pour représenter la valeur absolue d'une variable. Elles le sont moins pour représenter une valeur relative telle qu'une densité ou un coefficient.

A5.34. Il existe deux types de cartes de symboles gradués. Dans un cas, les données se réfèrent à un objet ponctuel tel qu'une ville ou un ménage. Dans ce cas, l'emplacement du symbole correspond à l'emplacement de l'objet (voir figure A5.11A). Dans le second cas, les symboles sont utilisés pour représenter les valeurs d'objets surfaciques tels que des districts. Dans ce cas, on doit choisir un emplacement représentatif dans chaque unité déclarante (voir figure A5.11B). Il faut noter que la plupart des systèmes entourent chaque cercle d'un halo afin que l'on puisse distinguer des cercles très rapprochés les uns des autres. Le système trace les plus grands cercles en premier lieu pour éviter que les plus petits soient recouverts.

A5.35. Comme précédemment, un logiciel de cartographie numérique nous permettra de choisir un symbole correspondant au thème de la carte. Ces symboles figuratifs peuvent rendre la carte plus attrayante. Cependant, si les symboles sont trop complexes, ils risquent de détourner l'attention du lecteur de la principale information

Figure A5.11

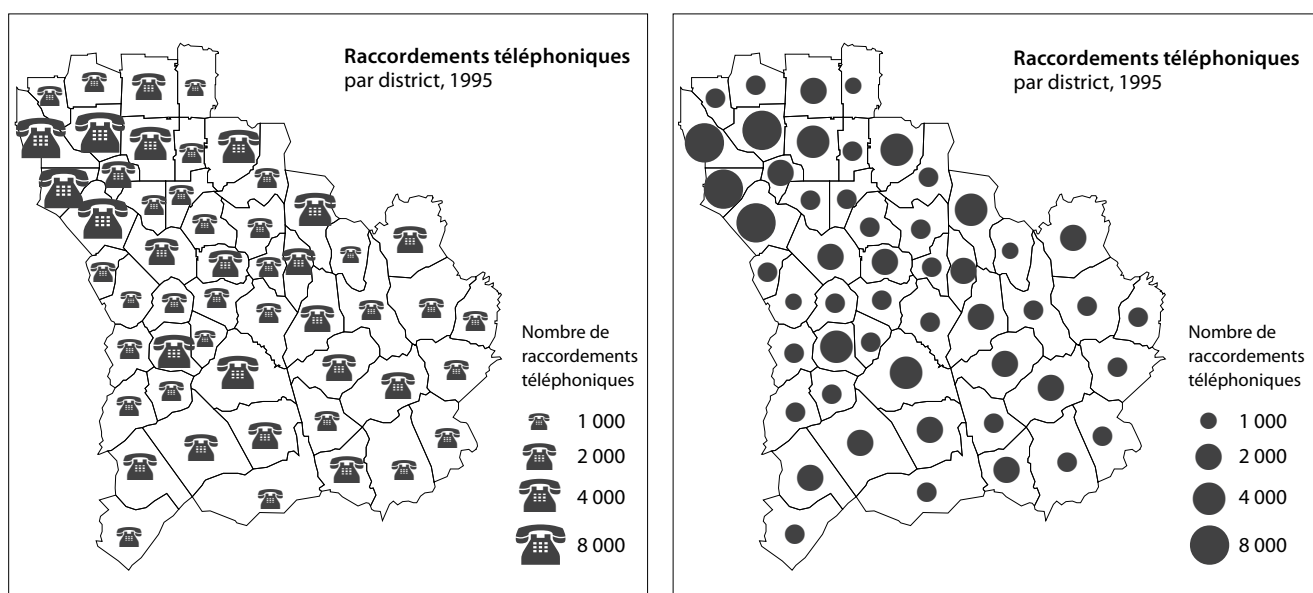
Symboles proportionnels correspondant à des objets ponctuels et surfaciques

qu'il s'agit de communiquer : la grandeur relative de la variable dans différentes régions. Comparons les deux versions d'une carte montrant le nombre de raccordements téléphoniques sur la figure A5.12. Bien que le symbole représentant l'appareil téléphonique soit très simple ; il est plus difficile d'apprécier la dimension de la variable sur la carte de gauche que sur la carte de droite plus simple. Le cartographe doit présenter l'information de façon claire et intelligible tout en s'efforçant de rendre la carte attrayante. Dans la plupart des cas, on obtiendra de meilleurs résultats avec des symboles simples qui ne détournent pas l'attention du lecteur de la grandeur relative de la variable étudiée.

A5.36. Des symboles proportionnels peuvent aussi être utilisés pour préserver deux variables en même temps. Par exemple, la dimension des cercles peut représenter le nombre de ménages dans une unité déclarante, tandis que la couleur ou la nuance de gris de chaque cercle indique le pourcentage de ménages ayant un raccordement au téléphone. Ici encore, le cartographe doit éviter de surcharger la carte d'informations. Si le nombre d'unités déclarantes est très élevé ou si les unités sont très petites, il peut être préférable de présenter les deux variables sur des cartes séparées.

A5.37. Outre les cercles, on utilise couramment des symboles géométriques tels que des carrés ou des triangles. En faisant varier l'orientation des triangles, on peut

Figure A5.12

Pictogrammes ou symboles graphiques simples

montrer la valeur de variables divergentes telles que l'immigration et l'émigration de chaque unité déclarante (voir figure A5.13). Des nuances de gris ou des couleurs différentes facilitent encore l'interprétation.

A5.38. D'autres cartes apparentées aux cartes de symboles gradués représentent les différences de valeurs à l'aide d'un certain nombre de symboles standardisés

Figure A5.13

Présentation de l'importance et de la direction des flux au moyen de symboles graphiques simples

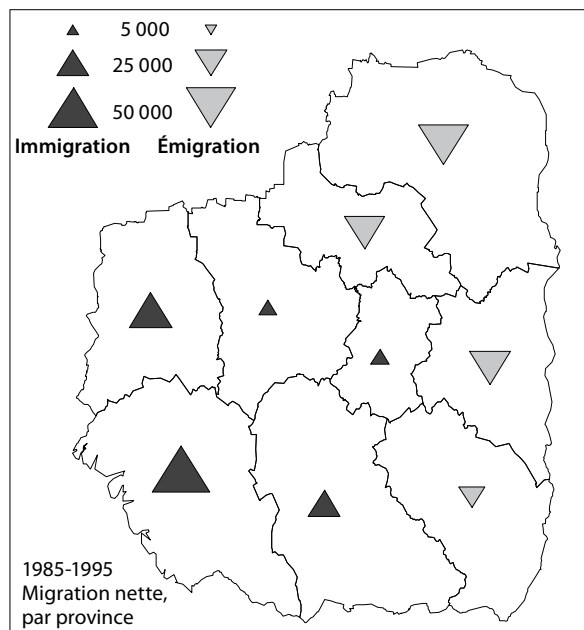
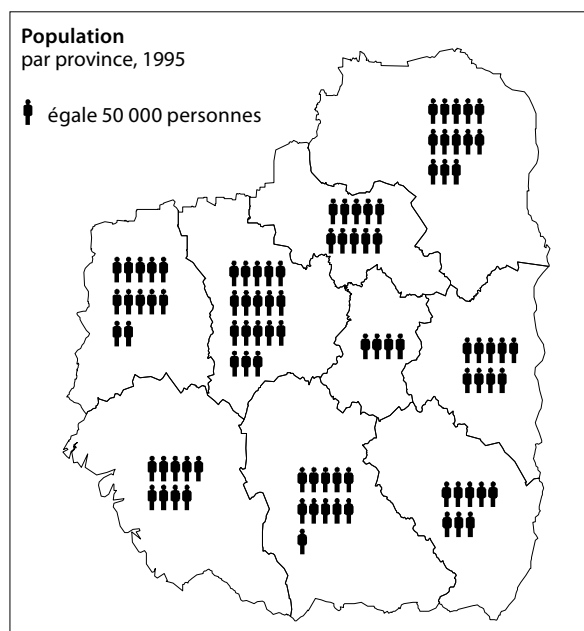


Figure A5.14

Représentation de valeurs de données faisant varier le nombre de symboles cartographiques pour chaque objet



tracés pour chaque unité géographique. Par exemple, la population totale peut être représentée comme sur la figure A5.14. Ce type de carte était très apprécié en cartographie thématique. Mais, de même que pour les symboles figuratifs, ces cartes sont facilement surchargées de détails et difficiles à interpréter. Il est préférable de représenter la grandeur des différentes valeurs à l'aide de symboles proportionnels.

d) Cartes-diagrammes

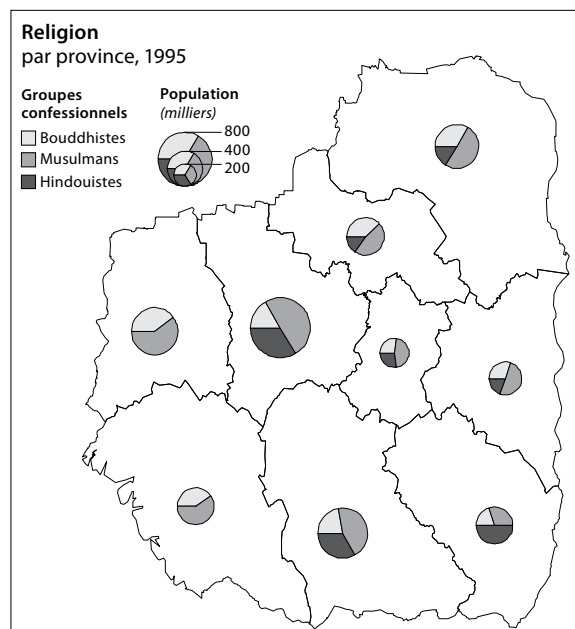
A5.39. Les cartes présentant des informations statistiques sous forme de diagramme pour chaque observation géographique rencontrent un grand succès du fait de leur disponibilité dans les logiciels commerciaux de SIG et de cartographie numérique. Comme plusieurs types de cartes examinés précédemment, les cartes-diagrammes sont facilement surchargées d'un trop grand nombre d'informations. Il existe malheureusement beaucoup d'exemples de cartes de ce type dont il est difficile ou impossible d'extraire des informations.

A5.40. Les types les plus courants de cartes-diagrammes utilisent des graphiques à secteurs, à barres ou à colonnes. Les diagrammes sont en général échelonnés de telle sorte que la dimension de chaque graphique à secteurs, par exemple, reflète l'importance du dénominateur. Par exemple, la figure A5.15 montre la répartition géographique de la proportion des principaux groupes confessionnels. Les diagrammes sont échelonnés en fonction de la population totale. Nous devons donc présenter deux types d'information dans la légende : la couleur qui se réfère à chaque groupe confessionnel et les totaux de population qui correspondent à une dimension donnée du graphique à secteurs.

A5.41. Les cartes-diagrammes donnent les meilleurs résultats lorsqu'il y a relativement peu d'observations géographiques et très peu de groupes à représenter. Par exemple, une carte-diagramme à secteurs qui comporte seulement deux catégories peut être très utile en combinaison avec une carte choroplèthe simple pour montrer plusieurs variables simultanément (voir figure A5.16) : La répartition spatiale des divers niveaux

Figure A5.15

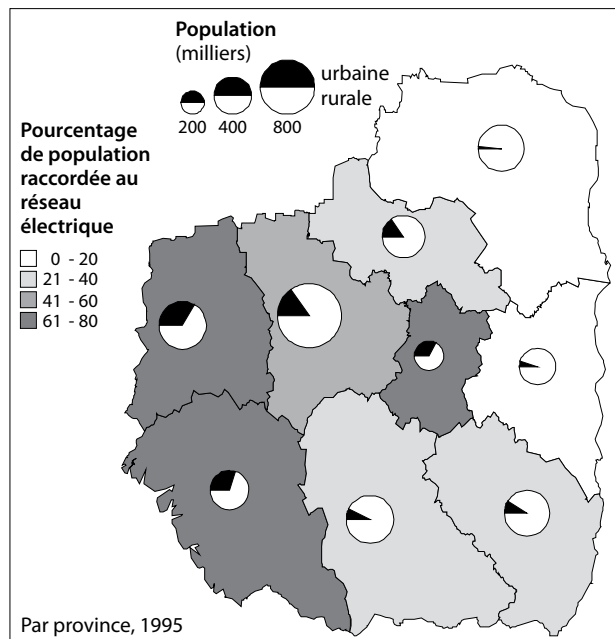
Carte-diagramme à secteurs



d'accès au réseau électrique, la population totale dans chaque province et la proportion de la population rurale ou urbaine. Sur cette carte, il semblerait que les provinces ayant une proportion élevée de population urbaine ont un coefficient plus élevé de raccordements au réseau électrique. Une carte bien conçue qui n'est pas surchargée de symboles, de couleurs et de nuances peut faciliter l'analyse multivariable de différentes variables. Cependant, la carte-diagramme à secteurs et les cartes semblables peuvent facilement devenir difficiles à interpréter et leur utilisation devrait être limitée aux cas dans lesquels le message cartographique n'est pas encombré de trop de symboles et de catégories.

Figure A5.16

Combinaison de cartes choroplèthes et de cartes-diagrammes à secteurs

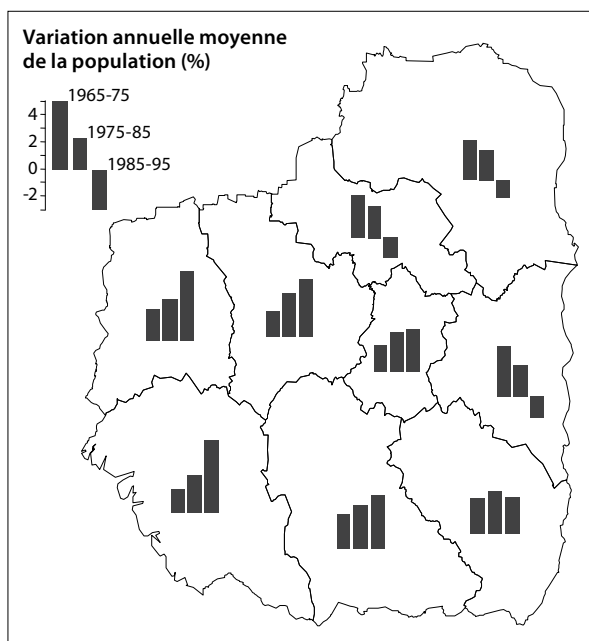


A5.42. Les cartes-diagrammes servent aussi à illustrer les tendances au cours du temps. La carte de la figure A5.17, par exemple, montre la variation annuelle moyenne en pourcentage de la population dans chaque province entre les trois derniers recensements. Les diagrammes à barres sont très simples, sans bordure et sans ligne de base, car pour ces données, il ne peut y avoir de doute quant aux barres correspondant à un accroissement ou à une diminution de la population. Comme précédemment, il s'agit en fait de montrer des changements relatifs dans le temps et non des valeurs exactes, qu'il est préférable de présenter sous forme de tableau.

A5.43. Un type de diagramme qui présente une grande importance en ce qui concerne les données de recensement de la population est naturellement la pyramide des âges. Les pyramides peuvent être combinées avec une carte de base des unités déclarantes pour montrer comment la répartition par âge et par sexe varie par région (voir figure A5.18). Les pyramides de population sont des diagrammes très complexes. Cela sous-entend qu'elles ne peuvent être représentées de façon satisfaisante que si le nombre de régions sur la carte est relativement faible. En général, cela signifie qu'elles ne seront présentées sur un atlas de recensement qu'au premier niveau subnational. Cependant, en pratique les logiciels commerciaux de SIG et de cartographie numérique ne produisent pas automatiquement des diagrammes-pyramides. Il faut donc les créer extérieurement, par exemple dans un tableur et les ajouter à une carte de base dans un logiciel graphique ou dans le module de mise en page d'un programme de cartographie numérique.

Figure A5.17

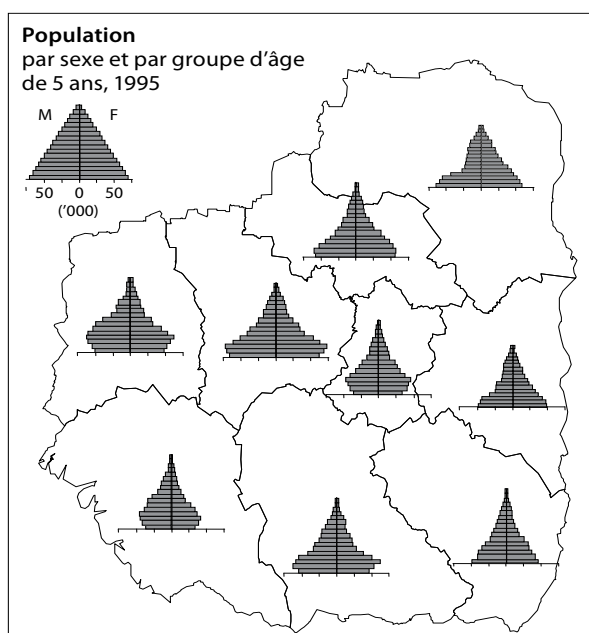
Carte montrant l'évolution dans le temps au moyen d'histogrammes



A5.44. Les pyramides de population présentées pour plusieurs régions sont significatives s'il existe des différences dans la forme des pyramides. Si les répartitions par âge et par sexe sont plutôt constantes dans tout le pays, les cartes résultantes ne fourniront guère d'indications intéressantes. Sur la figure A5.18, il semblerait que les provinces du sud-est aient connu une baisse de la fécondité au cours des 15 dernières années, tandis que cela n'a pas été le cas pour les provinces du nord. En outre, on

Figure A5.18

Combinaison de cartes et de pyramides de population

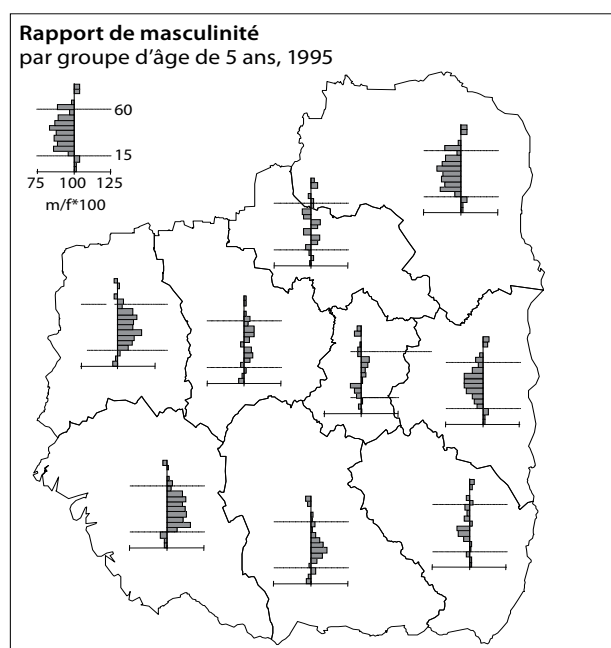


constate que les provinces du nord-est présentent une répartition dissymétrique du rapport de masculinité. Il semble qu'il y ait plus de femmes que d'hommes dans la population économiquement active. Au sud-ouest, la situation paraît être inverse.

A5.45. Les variations du rapport de masculinité peuvent être illustrées en utilisant un type différent de diagramme à barres, comme le montre la figure A5.19. Ces diagrammes montrent l'excédent ou le déficit d'hommes ou de femmes dans chaque province. La tendance qui apparaissait sur la carte des pyramides de population est ici beaucoup plus nette. Cependant, la carte est assez complexe et visuellement peu attrayante. On examinera un autre moyen de représenter les rapports de masculinité à la fin de la présente annexe.

Figure A5.19

Représentation des rapports de masculinité sur une carte



e) Cartes de flux

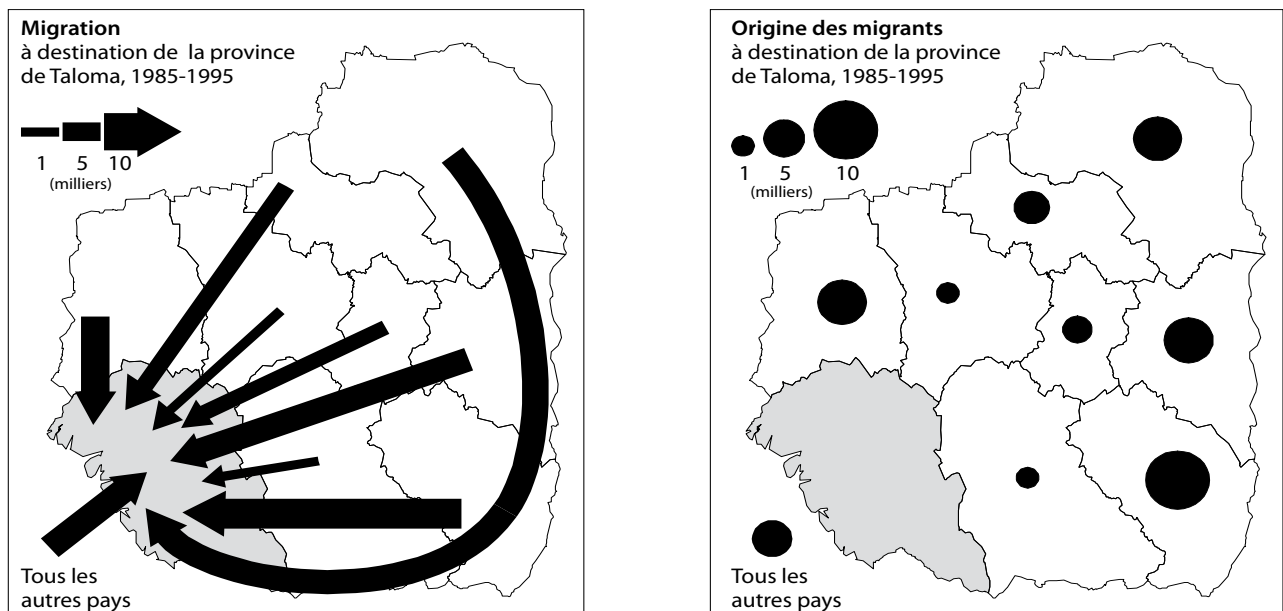
A5.46. La migration est une variable démographique qui représente les mouvements de population d'une partie du pays à une autre (migration interne) ou entre le pays et le reste du monde (migration internationale). La migration peut être représentée sur les cartes de diverses façons. Les taux de migration sont représentés au moyen de cartes choroplèthes des taux d'immigration, d'émigration ou de migration nette. Le volume de l'immigration ou de l'émigration peut être représenté en utilisant des cartes de symboles gradués (voir plus haut figure A5.13). Inversement, si l'on dispose d'informations complètes sur la migration, des cartes de flux — appelées également cartes de lignes flux — peuvent être utilisées. Ces cartes montrent divers aspects de la migration : l'itinéraire du flux de migration et la direction (de ... à), en utilisant une flèche comme symbole et l'importance du flux, en faisant varier l'épaisseur de la ligne.

A5.47. Les cartes de migration peuvent devenir rapidement très complexes. Même avec notre carte échantillon de provinces comportant seulement neuf unités déclarantes, il existe 72 flux éventuels — sans compter la migration internationale ou la migration à l'intérieur des provinces. C'est pourquoi on établit rarement des cartes complètes des flux montrant tous les itinéraires de migration possibles à l'intérieur

de la région du pays. Il existe plusieurs options possibles. L'une consiste à ne pas tenir compte des flux migratoires les moins importants et de représenter les plus importants et les plus significatifs. Une autre possibilité consiste à établir des cartes séparées pour chaque province montrant seulement l'immigration ou l'émigration de la province (voir figure A5.20). Pour nos provinces échantillons, cela donnerait une série de neuf paires de cartes. Même ces cartes plus simples peuvent facilement devenir surchargées. Le cartographe doit souvent dessiner des flèches courbes autour de la carte si les régions d'origine et de destination sont très éloignées.

Figure A5.20

Autres présentations des flux entre régions



A5.48. Dans les cartes de flux utilisant des flèches comme symboles, l'impression visuelle dépend de la longueur et de l'épaisseur de la flèche. Une flèche étroite et assez longue peut avoir un impact visuel plus grand qu'une flèche plus courte et plus large du fait de sa surface plus grande. Dans certains cas, le cartographe voudra peut-être utiliser ce fait pour attirer l'attention sur un flux migratoire intéressant en provenance d'une région éloignée, mais souvent le lecteur aura des difficultés pour évaluer l'importance relative des flux représentés par des flèches de longueur différente. Si l'on veut mettre l'accent sur le niveau absolu de migration en provenance de chaque région d'origine, d'autres présentations seront plus appropriées. Par exemple, au lieu de flèches, on pourra utiliser des symboles gradués pour montrer l'importance des flux migratoires par origine ou par destination (voir figure A5.20).

A5.49. En utilisant des types spéciaux de symboles gradués, chaque carte peut montrer à la fois la migration à destination de la province ou en provenant, comme le montre la figure A5.21. Ici, on utilise des demi-cercles de différentes couleurs ou nuances de gris pour distinguer l'immigration de l'émigration.

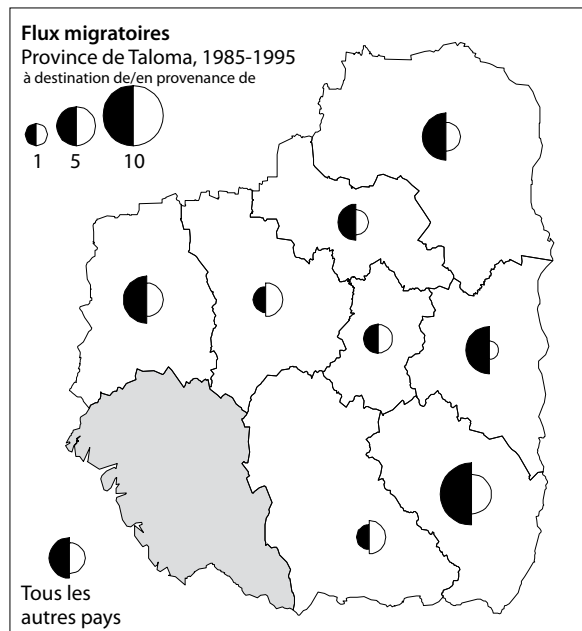
f) Représentation de phénomènes continus

A5.50. Les types de cartes présentés dans les précédentes sections conviennent pour des données concernant des objets géographiques distincts tels que des emplacements ponctuels ou des surfaces. Cependant, certains phénomènes géographiques

sont continus. La température ou l'altitude, par exemple, varient de façon continue dans l'espace. Mais il est aussi possible de considérer la répartition de la population comme une variable qui varie de façon plus ou moins continue. Les unités déclarantes sont définies assez arbitrairement et les valeurs agrégées présentées sous forme de tableau pour ces unités ne permettent pas de voir les variations spatiales à l'intérieur de chaque unité. Les atlas de population et, de plus en plus, les ensembles de données de SIG, présentent donc parfois la densité démographique et la répartition de la population comme des variables continues.

Figure A5.21

Représentation de l'immigration et de l'émigration

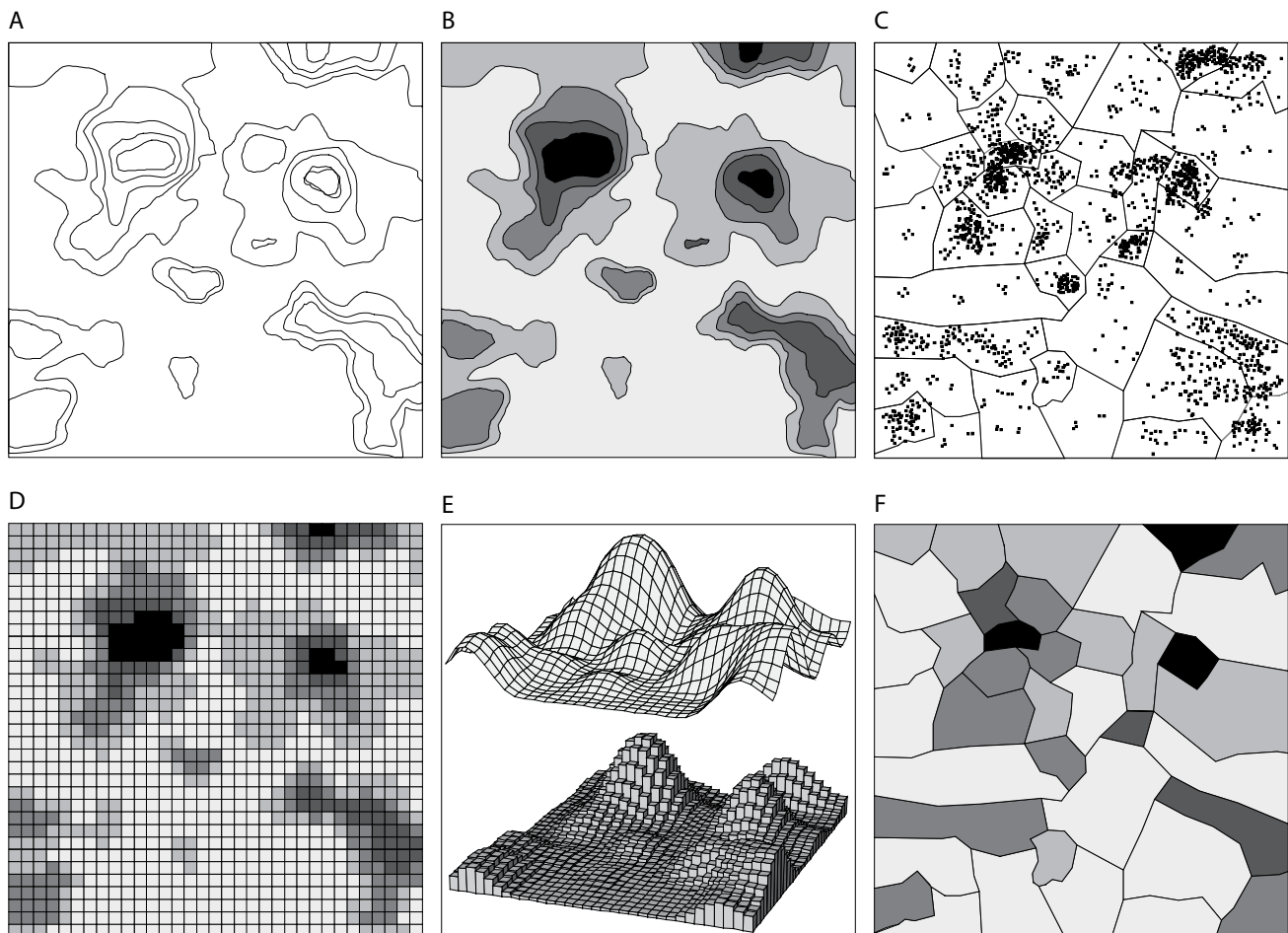


A5.51. La continuité effective ne peut pas être représentée facilement sur une carte imprimée ou dans une base de données d'ordinateur. Même si nous pouvions théoriquement établir une valeur différente pour chaque point exact du pays, il nous faudrait discrétiser les données d'une façon ou d'une autre aux fins de la cartographie. Plusieurs méthodes sont présentées sur la figure A5.22.

A5.52. Le moyen le plus courant pour représenter des données continues consiste à utiliser des isolignes ou des maillages réguliers. Les isolignes — le mot grec « iso » signifie égal — sont des lignes de valeur constante et on les appelle aussi lignes de niveau (figure A5.22A). Elles sont utilisées sur les cartes topographiques qui indiquent l'altitude. Les cartes de niveaux peuvent aussi être ombrées et elles ressemblent alors davantage à des cartes choroplèthes (voir figure A5.22B). Les couleurs représentent des valeurs dans la série de données comprise entre deux intervalles de niveau. Les cartes à points peuvent aussi être utilisées pour fournir une description plus continue de la répartition de la population ou d'une variable semblable. Comme on l'a indiqué plus haut, la plupart des logiciels de SIG produisent des cartes de points en traçant aléatoirement des points dans chaque unité déclarante. Dans le cas présent, nous n'obtenons aucune information supplémentaire par rapport à une carte choroplèthe. Mais si les points sont placés en fonction de données supplémentaires sur la couverture terrestre ou des emplacements de villages, par exemple, on pourra obtenir une image plus continue de la répartition de la variable (voir figure A5.22C).

A5.53. Aux fins de modélisation et d'analyse dans un SIG, les données continues sont en général enregistrées dans des maillages réguliers (figure A5.22D). La dimension des mailles est choisie de façon à préserver la variabilité dans l'ensemble de données, mais un maillage serré donne des dimensions de fichiers très importantes. Enfin, les logiciels de cartographie numérique, de même que les logiciels d'infographie en général, offrent différents moyens pour présenter sous forme de surfaces des ensembles de données variant de façon continue. La figure A5.22E montre deux exemples : un modèle fil de fer et un diagramme à barres à deux dimensions. Ces techniques sont très utiles pour présenter des informations sur les terrains fondées sur un modèle numérique d'altitude. Parfois, ces cartes peuvent aussi très bien illustrer la répartition de la population. Sur ces cartes, des collines et des pics représentent des groupements de densité démographique très élevées, tandis que les vallées symbolisent des zones faiblement peuplées. Cependant, en ce qui concerne la population et d'autres informations socioéconomiques semblables, il est souvent difficile d'évaluer la véritable répartition spatiale sur des surfaces. Intuitivement, nous ne pouvons interpréter des altitudes, mais il est beaucoup plus difficile d'associer rapidement les hauteurs de surface d'autres variables à leurs valeurs respectives. Il est donc en général plus indiqué de recourir à des techniques cartographiques plus classiques. Par comparaison, la figure A5.22F montre une carte choroplèthe où les unités déclarantes ne sont pas déterminées par la répartition des données.

Figure A5.22

Différentes méthodes cartographiques utilisables pour présenter des données continues

C. Classification des données

A5.54. On a examiné dans les sections précédentes les outils dont dispose le cartographe pour présenter des informations thématiques sur les cartes. Le concepteur de la carte doit choisir les variables graphiques et les types de cartes thématiques les plus appropriés pour la variable à représenter. Dans certains cas, les types de symbole et les valeurs de variables seront mis en correspondance, un à un. Ce sera le cas si l'on représente un petit nombre de catégories nominales, par exemple, avec des symboles ponctuels de taille semblable, mais de forme différente. Toutefois, même avec des données catégoriques, il est souvent nécessaire de représenter plusieurs objets ayant des valeurs similaires par le même symbole graphique. Par exemple, les ménages unis ou plurifamiliaux peuvent être représentés par le même symbole ponctuel. Les données numériques doivent presque toujours être catégorisées avant d'être appariés à des dimensions ou des couleurs de symboles.

A5.55. Le processus par lequel des observations présentant des valeurs semblables sont regroupées pour être représentées par le même symbole graphique est appelé classification. Il est semblable aux méthodes de classification dans le domaine statistique, qui groupent les valeurs en catégories de façon à réduire au minimum la variance des observations dans la même catégorie et à maximiser la variance entre différentes catégories. Les logiciels de cartographie numérique proposent des méthodes par défaut pour affecter des symboles à des valeurs ou à des intervalles de valeurs. Même si les interfaces graphiques utilisateur se sont améliorées, ces valeurs par défaut peuvent ou non être appropriées pour la variable représentée — le plus souvent elles ne le sont pas. Des outils de classification inappropriés donnent souvent des tracés de cartes inappropriés ou même susceptibles d'induire en erreur. On examinera donc plus en détail les possibilités de classification aux paragraphes suivants.

A5.56. Les classes de données numériques sont en général des intervalles de valeurs contigus. Le nombre de classes est déterminé par plusieurs facteurs : la répartition des données (c'est-à-dire la variation des valeurs dans l'ensemble de données), l'exactitude recherchée de la représentation des données et — facteur non moins important — la capacité du périphérique de sortie de présenter de petites différences de couleurs et de texture. On n'améliore pas nécessairement une carte thématique en accroissant le nombre de classes, car il est de plus en plus difficile pour le lecteur de distinguer différentes classes. Il importe plutôt de déterminer les intervalles de classes de façon à refléter exactement les variations de l'ensemble de données.

A5.57. Le choix d'une technique de classification appropriée dépend de la répartition des données de la variable. Une méthode permettant d'obtenir une carte exacte et visuellement attrayante pour un ensemble de données uniformément réparti (c'est-à-dire qui comporte un nombre à peu près égal de valeurs élevées, moyennes et faibles) ne donnera pas nécessairement de bons résultats pour une répartition de données dissymétrique — c'est-à-dire une répartition comportant beaucoup de valeurs faibles et peu de valeurs élevées.

A5.58. Pour établir des cartes de qualité appropriée pour la publication, les données devront donc toujours être évaluées en utilisant des diagrammes statistiques. Malheureusement, les logiciels de SIG et de cartographie numérique n'ont que des capacités limitées en matière de tracé de diagrammes. Mais ils permettent d'exporter les données vers des tableurs ou des programmes statistiques qui possèdent des fonctions développées pour tracer des diagrammes.

A5.59. Le type le plus utile de diagramme pour déterminer les intervalles de classe est un classement par ordre de grandeur. Tous les points correspondant aux

données sont classés en fonction de leur valeur de bas en haut. Ils sont ensuite tracés l'un à côté de l'autre — l'axe des x indique le rang de chaque observation et l'axe des y les valeurs de données. Les intervalles verticaux ou *discontinuités naturelles* entre des points de données voisins sont de bons candidats pour tracer des limites de classes, mais souvent on aura plus ou moins d'intervalles que le nombre de classes souhaité.

A5.60. Les pages qui suivent présentent des exemples de méthodes de classification courantes pour trois variables comportant des répartitions différentes de données statistiques. La variable densité démographique présente une répartition dissymétrique. Il existe de nombreuses valeurs dans l'intervalle de 21 à environ 110 personnes au kilomètre carré, et seulement quelques valeurs très élevées. La valeur la plus élevée (791) est égale à environ deux fois et demi la deuxième valeur par ordre d'importance (320). Cela n'est pas rare en ce qui concerne la densité démographique. Le district présentant les valeurs les plus élevées, par exemple, peut inclure la capitale d'une province par ailleurs rurale. La seconde variable est le taux d'alphabétisation pour les districts. Les valeurs sont assez uniformément réparties, ce qu'indique la ligne presque droite correspondant aux observations dans le classement par ordre de grandeur. Il n'y a pas de valeurs extrêmes.

A5.61. La troisième variable citée est l'indice synthétique de fécondité. Le classement par ordre de grandeur montre un accroissement rapide des valeurs pour les observations inférieures, une section moyenne étendue avec une augmentation beaucoup moins rapide et de nouveau un accroissement plus rapide des valeurs pour les observations très élevées vers la droite. Cela correspond à une répartition dite normale, qui est caractérisée par un plus faible nombre de valeurs extrêmement faibles et élevées et de nombreuses observations dans les intervalles intermédiaires. Naturellement, les exemples cités ne sont là qu'à titre d'illustration. Les mêmes variables pour d'autres zones géographiques pourront présenter des répartitions très différentes.

A5.62. Les exemples montrent que l'apparence d'une carte dépend au plus haut point du choix de la méthode de classification, qui peut ou non être appropriée pour la répartition des données. Cela confirme que les méthodes de classification fournies dans les logiciels de SIG doivent être utilisées avec une certaine prudence.

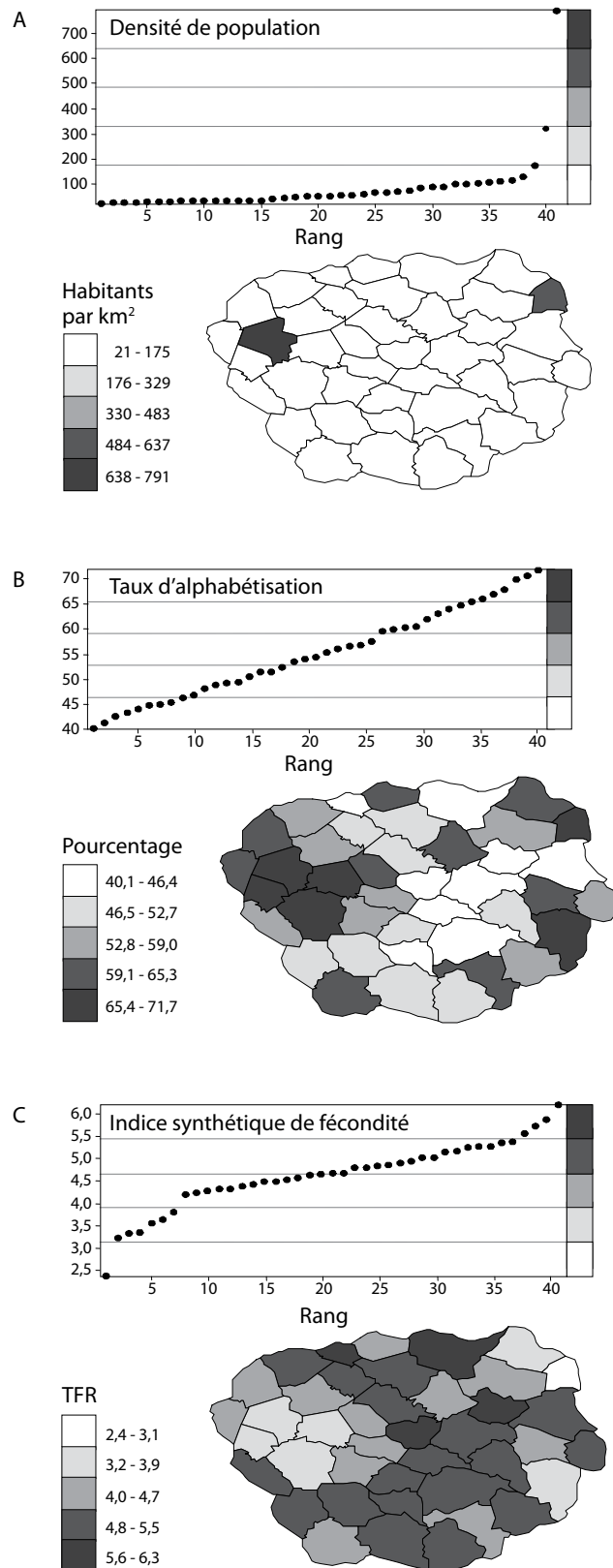
1. Classification de données en série

A5.63. L'une des méthodes de classification les plus simples consiste à diviser la gamme des valeurs de données en *intervalles égaux* (figure A5.23). Le cartographe détermine d'abord combien il utilisera de classes. L'intervalle de valeurs de données — la valeur la plus élevée moins la plus faible — est ensuite divisé par le nombre de classes pour obtenir l'incrément, appelée également raison de la progression. La première classe s'étend alors de la valeur la plus faible à cette même valeur augmentée de l'accroissement et les classes suivantes sont définies en ajoutant l'accroissement à la valeur supérieure de la classe précédente. Il peut être nécessaire d'arrondir si les nombres sont indiqués dans la légende avec un faible degré de précision.

A5.64. Pour la variable densité démographique, la valeur inférieure est égale à 21 et la plus élevée à 791. L'intervalle est donc de 770. Puisque nous voulons utiliser cinq catégories, la raison de la progression est donc égale à $770/5$ soit 154. Ainsi la première catégorie va de 21 à 175, la suivante de 176 à 329 et ainsi de suite.

A5.65. La carte de densité démographique montre pourquoi cela peut occasionner des difficultés. L'intervalle de valeurs est influencé par une valeur très élevée. La raison de la progression est donc si élevée que l'intervalle de la première catégorie comprend toutes les observations sauf deux. Il est évident que la carte obtenue n'est pas très instructive.

Figure A5.23
Intervalles égaux



A5.66. Cette méthode donne de biens meilleurs résultats avec le taux d'alphabétisation, qui est réparti de façon plus uniforme. L'ensemble de données est divisé en nombres à peu près égaux d'observations dans chaque catégorie et la carte résultante donne une bonne impression de la répartition de l'alphabétisation entre les districts.

A5.67. Enfin, la carte de l'indice synthétique de fécondité donne lieu à des difficultés comparables, mais moins marquées que celles rencontrées pour la carte de la densité démographique. On ne compte qu'une seule observation dans la catégorie inférieure et la carte est plutôt dominée par les valeurs des intervalles de la catégorie moyenne. Cependant, par une coïncidence, les intervalles de classes entre la deuxième et la troisième et entre la quatrième et la cinquième catégories correspondent assez bien aux discontinuités dans la répartition des données.

A5.68. Outre ce procédé des intervalles égaux, il existe d'autres solutions pour la classification des données en série. L'une d'entre elles consiste à utiliser une progression géométrique régulière telle que 0-2, 2-4, 4-8, 8-16, etc. Cela peut donner de bons résultats pour des répartitions dissymétriques de données comme dans le cas de la variable densité démographique.

2. Classification statistique

A5.69. Une méthode de classification consiste à avoir un nombre approximativement égal d'observations géographiques dans chaque catégorie. Ce résultat peut être obtenu en utilisant le concept statistique des quantiles, qui divisent l'ensemble de données en classes comportant le même nombre d'observations. Si l'on a quatre classes, on parle de quartiles, s'il y en a cinq, de quintiles, etc.

A5.70. Pour déterminer les quantiles, le nombre d'observations est divisé par le nombre de catégories souhaité et, si nécessaire, arrondi à l'entier le plus proche. Dans le classement par ordre de grandeur, les premières observations n sont affectées à la première catégorie, les n suivantes à la seconde, etc. Tout nombre impair est affecté à la première ou à la dernière catégorie.

A5.71. L'établissement de cartes par quantiles est assuré par de nombreux logiciels de cartographie numérique et c'est pourquoi cette méthode est devenue très courante pour l'établissement des cartes.

A5.72. Les trois échantillons de cartes ont tous un aspect attrayant. On a par définition une bonne répartition des observations entre les classes de telle sorte que toutes les cartes utilisent pleinement toute la gamme des gris.

A5.73. Si l'on considère la répartition des données, la classification concernant la variable taux d'alphabétisation paraît très appropriée. En fait, la carte ne paraît pas très différente de celle qui utilise des intervalles égaux.

A5.74. Cependant, pour les cartes de la densité démographique et de l'indice synthétique de fécondité, nous constatons que la méthode regroupe des valeurs semblables dans des catégories différentes. Pour l'indice synthétique de fécondité, par exemple, les deux observations présentant les valeurs les plus élevées dans l'intervalle de données inférieur (2,4-4,2) ressemblent beaucoup plus aux observations de la seconde catégorie qu'aux observations de la première catégorie. Bien plus, on a trois observations avec une valeur de 5,3, une qui est affectée à la quatrième classe et deux à la cinquième (certains logiciels de cartographie numérique assouplissent le critère du nombre égal d'observations pour éviter de tels cas).

A5.75. Les cartes par quantiles doivent donc être utilisées avec prudence. Très souvent, des valeurs semblables se retrouvent affectées à des catégories différentes et

Figure A5.24

Établissement de cartes par quantiles (fréquence égale)

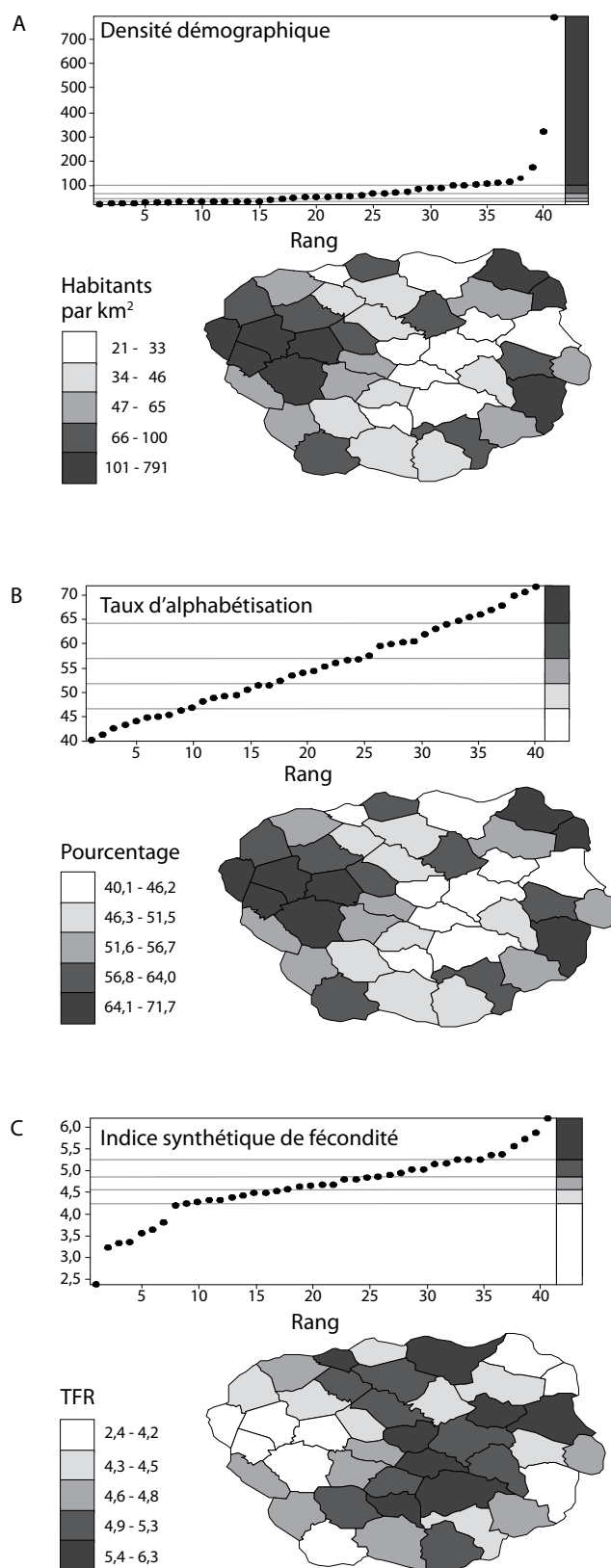
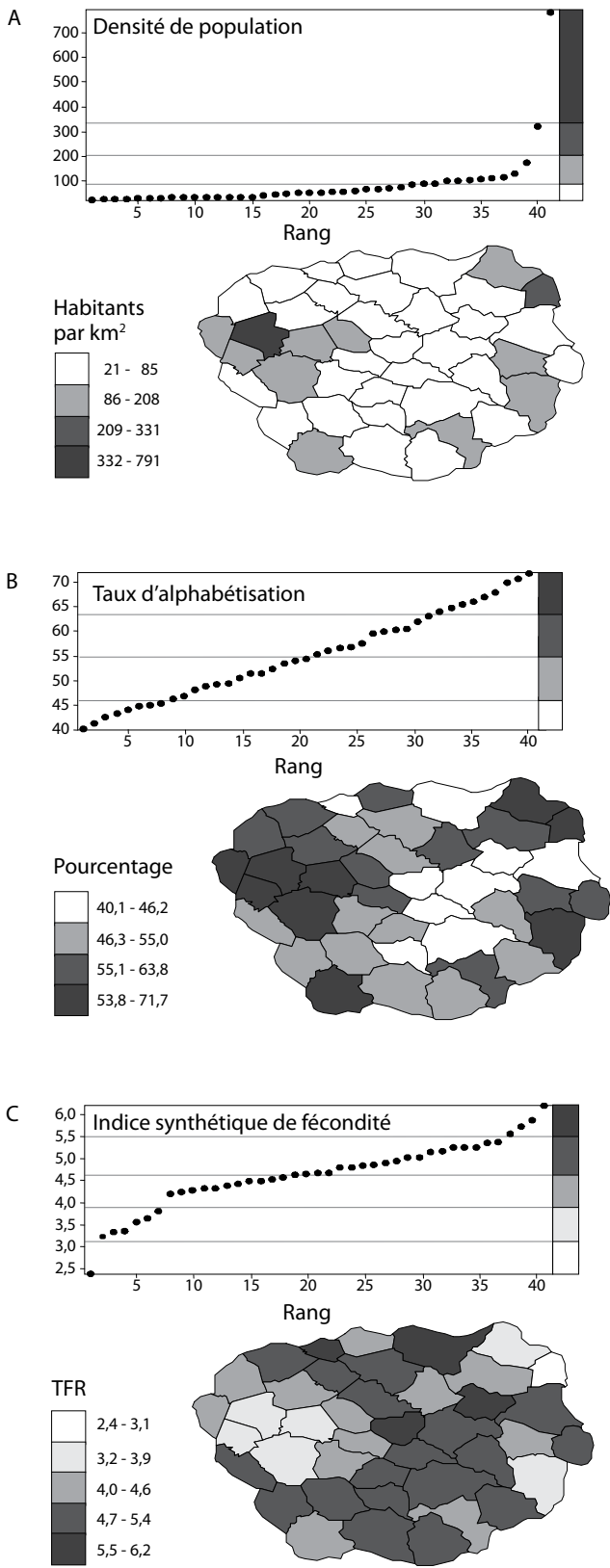


Figure A5.25
Écart-type



des valeurs différentes sont affectées à la même classe. Bien que les cartes obtenues soient visuellement attrayantes, cette impression risque d'induire en erreur.

A5.76. Il existe une autre technique de classification statistique fondée sur des évaluations sommaires de la répartition des données. Une méthode consiste à déterminer des intervalles de classes en utilisant l'écart-type de la répartition de la variable. L'écart-type est égal à la racine carrée de la variance. La variance est calculée en tant que moyenne des carrés des écarts entre les valeurs de données et la valeur moyenne globale. Par exemple, pour le taux d'alphabétisation, l'écart-type est égal à 8,9.

A5.77. Les catégories de cartes fondées sur les écarts-types montrent ainsi comment des observations individuelles — concernant par exemple des districts — se comparent à la valeur moyenne d'une province ou du pays tout entier.

A5.78. Les catégories sont obtenues en soustrayant l'écart-type de la moyenne (55 pour le taux d'alphabétisation) ou en le lui ajoutant. Les intervalles des catégories sont donc constants, comme avec la méthode des intervalles égaux.

A5.79. En ce qui concerne le taux d'alphabétisation, le premier intervalle de données (40,1-46,2) correspond à des valeurs supérieures à un écart-type, mais inférieures ou égales à deux écarts-types au-dessous de la moyenne. La répartition des données étant assez dense, toutes les valeurs sont comprises entre +/- deux écarts-types et quatre catégories seulement sont nécessaires. Comme nous le voyons sur la figure A5.25B, cette méthode divise les valeurs du taux d'alphabétisation en nombres d'observations à peu près égaux dans chaque catégorie, ce qui donne une carte avec un bon contraste visuel.

A5.80. Cependant, cette approche convient beaucoup moins bien pour la variable densité démographique. Parce qu'il y a beaucoup de valeurs peu élevées, la densité démographique moyenne est plutôt faible (85,4) et l'écart-type plutôt élevé (124,8). La première catégorie — correspondant à des valeurs situées à moins d'un écart-type de la moyenne — devrait donc être en fait comprise entre -39,5 et 85,4. D'autre part, la valeur la plus élevée (791) est à plus de cinq écarts-types de la moyenne. Nous devrions donc utiliser beaucoup plus de catégories, dont la plupart ne contiendraient aucune observation. Au contraire, la catégorie la plus importante sur la carte présentée ici inclut toutes les valeurs à plus d'un écart-type de la moyenne. Il est évident que les écarts-types ne constituent pas un bon choix pour cette variable.

A5.81. Les écarts-types donnent des résultats un peu meilleurs pour l'indice synthétique de fécondité, avec une moyenne de 4,6 et un écart-type de 0,8. Cependant, seule la valeur très faible de 2,4 se situe dans la catégorie la plus basse, soit à plus de deux écarts-types au-dessous de la moyenne.

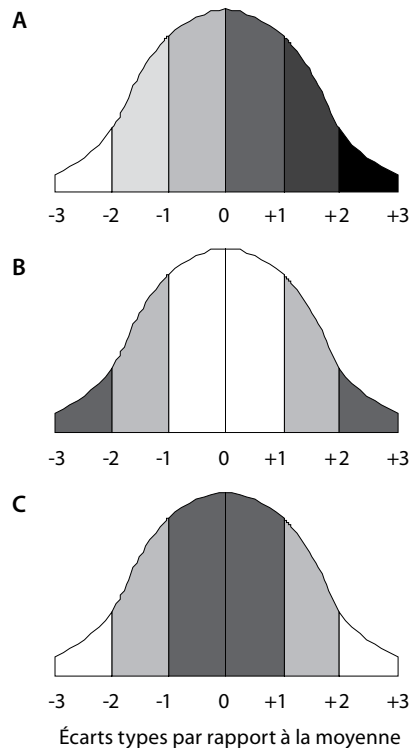
A5.82. La méthode de classification par les écarts-types est intéressante parce qu'intuitive, étant étroitement associée aux techniques statistiques descriptives. Elle donne de bons résultats si les données sont normalement réparties, avec une variance relativement faible, de telle sorte qu'au plus six catégories suffisent pour inclure toutes les valeurs.

A5.83. Les écarts-types peuvent être utilisés pour représenter différents types de tendances dans un ensemble de données [voir figure A5.26; voir également Dent (1999)]. Dans les exemples fournis sur la figure A5.25, on utilise une échelle de gris allant du clair au sombre. Les cartes illustrent la progression des valeurs faibles aux valeurs élevées de densité démographique, de taux d'alphabétisation et d'indice synthétique de fécondité correspondant au classement par catégories présenté sur la figure A5.26A. C'est en fait l'application la moins commune de la classification fondée sur les écarts-types.

A5.84. Cette méthode est plus couramment utilisée pour illustrer des tendances divergentes. Par exemple, pour montrer des niveaux de revenu, nous souhaiterions peut-être faire ressortir quels sont les districts les plus pauvres et le plus riches. Dans ce cas, nous affecterons des couleurs vives ou une texture marquée aux districts dont les valeurs se situent à plus d'un et deux écarts-types de la moyenne et des couleurs relativement pâles aux districts situés au centre de la répartition de données (figure A5.26B).

Figure A5.26

Affectation de couleurs aux catégories déterminées par les écarts-types



A5.85. Si l'on s'intéresse seulement à la distance par rapport à la moyenne — sans se préoccuper de savoir si les valeurs sont supérieures ou inférieures à la moyenne — on pourra utiliser les mêmes couleurs de part et d'autre. S'il s'agit aussi de savoir si les valeurs sont supérieures ou inférieures à la moyenne, on devra utiliser différentes couleurs ou textures pour chaque côté. Par exemple, sur une carte imprimée en couleurs, on pourra affecter des teintes rouges allant du clair au sombre aux catégories inférieure à la moyenne et des teintes bleues correspondantes aux catégories supérieures à la moyenne.

A5.86. Dans d'autres cas, nous souhaiterions peut-être faire ressortir les intervalles moyens (figure A5.26C). MacEachren (1994), par exemple, examine une carte de l'Irlande du Nord (Royaume-Uni) publiée par Fothergill et Vincent (1985) qui montre la proportion de Protestants et de Catholiques. Sur cette carte, les valeurs de l'ordre de 50 %, indiquant un nombre à peu près égal de Protestants et de Catholiques, sont mises en lumière en affectant aux catégories moyennes une couleur vive (jaune). Les zones où les Catholiques ou les Protestants sont en nette majorité sont présentées avec des couleurs plus pâles (vert et orange, respectivement).

3. Points d'inflexion naturels

A5.87. Comme nous l'avons vu avec les précédents exemples, la plupart des méthodes donnent des cartes de nature à induire en erreur pour des variables qui ne présentent pas une répartition très uniforme. Il arrive souvent que des valeurs semblables soient affectées à des catégories différentes ou que des valeurs très différentes soient regroupées ensemble. Une méthode logique pour classer les données en cartographie consiste à trouver un mode de regroupement des données qui optimise l'opération en minimisant les différences entre valeurs dans chaque catégorie et en maximisant la variation entre groupes.

A5.88. Cet objectif peut être atteint en procédant à une inspection visuelle de la répartition des données et ensuite au choix des inflexions entre catégories. Des exemples de cette méthode sont présentés sur la figure A5.27. Pour la variable indice synthétique de fécondité, cela est très simple, car elle présente plusieurs inflexions distinctes dans la répartition.

A5.89. Cela s'avère plus difficile pour les deux autres variables. Pour la densité démographique, une stricte application de la méthode pourrait consister à affecter toutes les valeurs inférieures à la même catégorie et les valeurs supérieures à un certain nombre de catégories séparées. Mais cela doit être mis en balance avec le souci de préserver une variation peu marquée dans les intervalles de valeurs inférieures.

A5.90. De même, pour la variable taux d'alphabétisation uniformément répartie, les points d'inflexion ne sont pas très nets car la différence de valeur entre les observations ne varie pas beaucoup.

A5.91. Malgré ces difficultés, la classification à partir des inflexions naturelles, prenant explicitement en considération la répartition des données, permet en général d'obtenir des représentations cartographiques précises des données avec un bon contraste visuel.

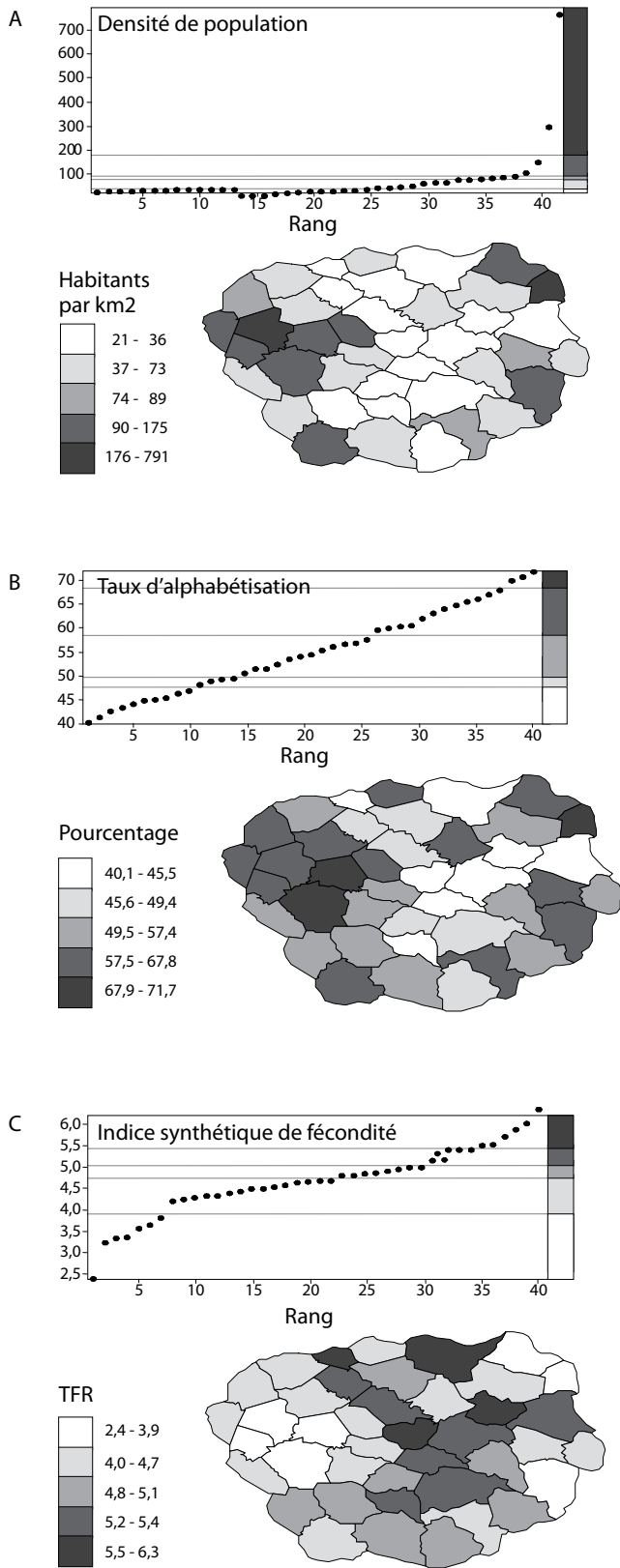
A5.92. Plutôt que de se fier à une appréciation quelque peu subjective, on peut aussi laisser l'ordinateur déterminer les inflexions naturelles ou optimales. De nombreux logiciels de SIG et de cartographie numérique possèdent des fonctions permettant de définir des points d'inflexion naturels fondés sur une évaluation automatisée de la répartition des données (méthode de classification optimale de Jenk). On peut aussi utiliser des fonctions de classification ou de regroupement dans des logiciels statistiques.

4. Cartes choroplèthes sans intervalles de catégories

A5.93. Avec les cartes choroplèthes dites sans catégories, le cartographe n'a pas à choisir de méthode de classification. Grâce à une technologie améliorée d'affichage et d'impression, les écrans d'ordinateurs et les imprimantes peuvent obtenir une large gamme de couleurs différentes ou de nuances de gris. Pour une carte sans catégorie ou de catégorie n , les valeurs de données détermineront directement, par exemple, le pourcentage de gris. Par exemple, pour une variable exprimée en pourcentage, nous pouvons choisir le niveau de gris sur une échelle allant de 0 % de gris (blanc) à 100 % de gris (noir) qui correspond à chaque valeur de l'observation. Mais si la méthode de reproduction peut donner un nombre suffisant de nuances distinctes, il est recommandé d'éviter le blanc comme couleur, car c'est aussi généralement la couleur du fond de page.

A5.94. En pratique, cela ne donnera peut-être pas des résultats optimaux. Cela tient notamment au fait que de nombreuses variables ne vont pas de 0 à 100, mais ont des valeurs concentrées sur un intervalle plus restreint. La carte peut ainsi ne présenter en définitive que des nuances de gris très claires ou très sombres. Nous pouvons éviter

Figure A5.27
Points d'inflexion naturels



ce problème en « étirant » la répartition des données : l'utilisation de la couleur la plus claire pour la valeur la plus basse et de la plus sombre pour la valeur la plus élevée donnera des cartes plus faciles à interpréter par le lecteur.

A5.95. Mais en général, il y a une limite au nombre de nuances de gris et de couleurs qui peuvent être facilement distinguées. Si un système d'ombrage continu est utile à des fins analytiques, le classement des valeurs de données entre un nombre limité de catégories est généralement préférable pour des cartes établies à des fins de présentation.

5. Classification des données extérieures

A5.96. Dans certains cas, le système de classification est fourni par un organisme extérieur. Par exemple, pour établir une carte de la pauvreté par district, le cartographe utilise une valeur seuil de donnée du revenu moyen — appelée seuil de pauvreté — au-dessous de laquelle un district est considéré comme pauvre. Autre exemple de système de classification fourni extérieurement, celui dans lequel on procède à des comparaisons avec les cartes imprimées existantes pour lesquelles les données initiales ne sont pas disponibles. Pour permettre une comparaison exacte entre les cartes, par exemple pour le taux de fécondité dans les différentes provinces d'un pays, la classification doit être identique.

6. Remarques générales

A5.97. L'aperçu qui précède montre qu'il existe de nombreuses méthodes pour affecter des valeurs de données à des catégories. La plupart des logiciels de SIG et de cartographie numérique prennent en charge les intervalles égaux, les quantiles, les écarts-types et les points d'inflexion naturels. De plus, tous les logiciels permettent à l'utilisateur de définir une classification de données personnalisée.

A5.98. Chaque méthode a ses avantages et ses inconvénients, qui sont présentés sur le tableau A5.1. Le choix de la méthode la plus appropriée dépend de la répartition des données et des objectifs de la carte. En général, la répartition des données devrait toujours être évaluée en utilisant un diagramme statistique tel que le tableau d'ordre de grandeurs présenté plus haut. De tels outils sont proposés dans certains logiciels de SIG. Le nombre optimal de catégories et les meilleurs points d'inflexion paraîtront alors souvent tout à fait évidents.

Tableau A5.1

Évaluation des différentes techniques de classification

Méthode de classification	Avantages	Inconvénients
Intervalles égaux	Facile à mettre en œuvre. Méthode appropriée pour des données uniformément réparties.	Aucun rapport entre le système de classification et la répartition des données. Les intervalles de catégories étant fixes, des valeurs semblables peuvent être affectées à différentes catégories et des valeurs différentes à une même catégorie. Ne convient pas pour des répartitions dissymétriques de données ou des ensembles de données avec des valeurs hors norme.
Progression géométrique	Facile à mettre en œuvre. Appropriée pour des données avec une répartition très dissymétrique (beaucoup de valeurs faibles et peu de valeurs très élevées).	La progression géométrique appropriée doit être déterminée par l'utilisateur. Les intervalles de catégories étant fixes, des valeurs semblables peuvent être affectées à différentes catégories et des valeurs différentes à une même catégorie.

Méthode de classification	Avantages	Inconvénients
Quantiles (fréquence égale)	Un bon contraste visuel est garanti. Appropriée pour des données très uniformément réparties.	Des valeurs semblables ou identiques peuvent se retrouver dans des catégories différentes.
Écarts-types	Convient pour représenter des tendances divergentes centrées autour de la valeur moyenne. Établit un rapport entre des catégories individuelles et la valeur moyenne globale. Convient pour des données normalement réparties.	Des répartitions dissymétriques ou des ensembles de données contenant des valeurs hors norme (peu de valeurs très élevées ou très faibles) nécessiteront un grand nombre de catégories (plusieurs écarts-types au-dessus ou au-dessous de la moyenne)
Inflexions naturelles	Des valeurs semblables seront affectées à la même catégorie. Le nombre de catégories est souvent suggéré par le nombre d'inflexions.	Les intervalles de catégories obtenus peuvent être très inégaux. Exige une appréciation subjective (détermination visuelle). Ne permet pas la comparaison des cartes dans le temps.
Cartes choroplèthes sans classification	Il n'est pas nécessaire de définir des interruptions entre catégories. La nuance de gris ou la couleur est déterminée directement par la valeur de la donnée. Fait ressortir une répartition continue des valeurs dans l'ensemble de données.	La plupart des périphériques de sortie n'acceptent qu'un nombre limité de nuances de gris ou de couleurs distinctes. Les cartes présentant de faibles différences de gris ou de couleur sont difficiles à reproduire (par exemple à photocopier). Difficile à mettre en œuvre dans la plupart des logiciels de SIG et de cartographie.

A5.99. Il convient de noter que les points d'inflexion naturels ne sont pas appropriés si plusieurs cartes sont présentées ensemble à des fins de comparaison, par exemple, une série chronologique des rapports de masculinité par district ou des cartes d'accès à l'eau potable pour deux provinces du pays. Dans ces cas, les limites de classes doivent être choisies constantes. Un système de classification défini par l'utilisateur et fondé sur une évaluation de toutes les séries chronologiques doit être choisi à cette fin. Les cartes par quantiles peuvent aussi être utilisées, si l'objectif consiste seulement à comparer le classement des différentes observations dans le temps ou l'espace plutôt que les valeurs de données effectives. Deux cartes par quartiles pourraient par exemple présenter les 25 % des districts possédant les taux d'alphabétisation les plus élevés tels qu'ils apparaissent à la suite du dernier recensement et du recensement précédent.

D. Choix de la couleur

A5.100. Les exemples de cartes examinées dans la présente annexe utilisent tous des échelles de gris pour la symbolisation. Les publications en noir et blanc sont moins coûteuses à produire et les cartes utilisant des échelles de gris demeurent lisibles lorsqu'on les reproduit sur une photocopieuse en noir et blanc. Par contre, l'utilisation de la couleur offre au cartographe beaucoup plus de possibilités dans la conception de la carte. Le prix des imprimantes couleurs continue de baisser. D'autre part, dans un proche avenir, beaucoup plus de cartes seront présentées électroniquement sur des sites Web ou des publications électroniques. Dans ces cas, la couleur pourra être largement utilisée pour établir les cartes, même s'il conviendra de l'utiliser judicieusement.

A5.101. Il est utile de savoir comment un ordinateur interprète les couleurs lorsqu'on doit définir un système de couleurs pour une carte choroplèthe. Les couleurs sont définies dans un ordinateur en utilisant un modèle parmi plusieurs modèles de

couleurs. Deux des plus courants sont le modèle teinte-saturation-intensité (TSI) et le modèle rouge, vert et bleu (RVB). Le terme « teinte » désigne ce que l'on comprend généralement par la couleur, comme le « rouge » ou le « bleu ». Physiquement, la teinte se réfère à l'intervalle spectral de la lumière réfléchi et va du violet, avec une faible longueur d'onde, au bleu, vert, jaune, orange et rouge qui a la longueur d'onde la plus élevée dans le spectre visible. L'intensité est appelée parfois luminosité [teinte-saturation-luminosité (TSL)]. Elle détermine la différence entre, par exemple, un rose pâle et un rouge sombre, qui ont tous deux la même teinte. Enfin, la « saturation » est la mesure de la brillance ou de l'intensité. Une couleur ayant une saturation faible apparaît plus pâle ou grise tandis qu'une couleur à haute saturation paraît plus pure.

A5.102. Le modèle RVB est un modèle dans lequel les nouvelles couleurs sont définies par addition en combinant différents niveaux de rouge, de vert et de bleu. Les écrans d'ordinateur ou de téléviseur utilisent la méthode RVB. Des niveaux égaux des mêmes trois couleurs donnent des nuances de gris. Les niveaux inférieurs de rouge, vert et bleu combinés donnent du noir et les valeurs les plus élevées du blanc.

A5.103. Le choix de la couleur dépend du niveau de mesure de la variable, du type de carte utilisé et du message que le cartographe souhaite communiquer. L'œil humain distingue facilement différentes teintes, ce qui est tout à fait approprié pour distinguer des catégories distinctes. Par exemple, des cercles bleus peuvent être opposés à des cercles rouges pour montrer différents types d'écoles. Cependant, il faut tenir compte du daltonisme lorsqu'on choisit une couleur uniquement pour distinguer différents symboles cartographiques. Les daltoniens risquent de ne pas pouvoir distinguer le rouge et le vert — forme la plus courante du daltonisme touchant 1 % des hommes — ou le bleu et le jaune. Certains ne peuvent pas voir la partie verte du spectre des couleurs. En général, il est préférable de ne pas s'en remettre aux différences entre le rouge et le vert dans la composition des cartes.

A5.104. Pour présenter des variables mesurées en continu comme la population, le revenu ou des coefficients et des pourcentages, on utilise des variables graphiques qui comportent un classement distinct. Les différences de niveau de couleur (c'est-à-dire les nuances claires à sombres d'une même couleur) sont facilement associées aux valeurs de la variable, tandis que les nuances sombres sont en général associées aux valeurs les plus élevées de la donnée. Par exemple, les niveaux de densité démographique sont souvent représentés par des teintes rouges, allant des rouges très pâles pour une faible densité démographique aux rouges sombres pour les zones de forte densité démographique. Pour les répartitions de données dissymétriques, les niveaux de couleur ne sont naturellement pas liés en proportion directe aux valeurs des catégories de données. Sinon, dans l'exemple de la variable densité démographique utilisé dans la précédente section, par exemple, nous utiliserions de nombreuses nuances très pâles, difficiles à distinguer, de rouge clair pour les nombreuses valeurs peu élevées et une nuance très sombre pour un petit nombre de valeurs élevées. On utilise au contraire des degrés égaux de niveau de couleur pour représenter les catégories d'une progression géométrique ou similaire.

A5.105. Si une classification comporte de nombreuses catégories, nous pouvons nous retrouver avec un plus grand nombre de catégories que ce que l'on peut distinguer clairement sur une page imprimée. Dans ce cas, les nuances de couleurs proches les unes des autres peuvent être combinées — ce que l'on appelle une gamme de couleurs d'un spectre partiel. Pour utiliser à nouveau l'exemple de la densité démographique, nous pouvons commencer par des nuances de jaune clair et continuer jusqu'aux rouges sombres en passant par les oranges. Il importe de considérer qu'il doit y avoir une nette progression des couleurs moins dominantes aux couleurs plus

dominantes. Les cartes utilisant plusieurs teintes brillantes dominantes pour des valeurs faibles ou levées d'un intervalle continu ou ordinal de catégories ne communiquent pas un message clair et sont une source de confusion pour le lecteur.

A5.106. Il paraît approprié d'utiliser des couleurs différentes pour représenter un intervalle continu de données dans une application telle qu'une échelle de données divergente. Par exemple, une carte de l'immigration nette par unité administrative comportera un éventail de catégories partant de nombres négatifs élevés pour une forte émigration et atteignant des valeurs positives élevées correspondant à une forte immigration en passant par zéro. Pour faire ressortir les valeurs négatives et positives élevées — où la migration influe le plus sur la dynamique de la population — nous pouvons utiliser un système de couleurs allant par exemple du rouge intense au rouge pâle ou au rose, puis au blanc pour les taux de migration nette proches de zéro et du bleu pâle au bleu intense pour les taux d'immigration les plus élevés.

A5.107. Une dernière observation concernera la catégorie multivariable, avec laquelle deux variables sont présentées en combinaison. Par exemple, une carte pourra associer différents niveaux d'alphabétisation et de fécondité en utilisant une légende constituée essentiellement d'une matrice de combinaisons possibles des catégories alphabétisation et fécondité. Le cartographe doit trouver un système de couleurs approprié dans lequel, par exemple, les différences de niveaux d'alphabétisation sont représentées par des teintes voisines dans un système spectral partiel et les différences de taux de fécondité sont représentées par des variations de l'intensité de la couleur. Malheureusement, ces cartes sont difficiles à interpréter. Le lecteur doit constamment consulter la légende pour assortir les couleurs aux valeurs de données concernant les deux variables. En général, on devra éviter de recourir à des cartes multivariables. La section A5.6 présentera quelques approches alternatives à la présentation géographique d'informations multivariables.

A5.108. Revenant aux types de niveaux de mesure examinés plus haut, le tableau A5.2 présente, sous forme résumée, les directives concernant l'utilisation des nuances de gris et des couleurs [voir également Brewer (1994)].

Tableau A5.2
Choix des nuances de gris et des couleurs

Niveaux de mesure		Exemple	Cartes en noir et blanc	Cartes en couleurs
Nominaux	Binaires	Accès à l'eau potable	Blanc ou noir, gris pâle ou gris sombre.	Nuances fortement contrastées de différentes couleurs comme le bleu et le rouge ou le jaune et le vert.
	Catégoriels	Langue dominante (anglais, français, espagnol, etc.)	Variations de structure avec une dominance visuelle semblable.	Différentes couleurs avec des niveaux semblables de valeur et de saturation qui ne supposent aucun classement, par exemple bleu, vert, jaune, violet.
Ordinaux		Niveau d'éducation (école primaire, école secondaire, etc.)	Nuances de gris échelonnées avec des différences relativement importantes entre niveaux de gris. Les différences de texture font ressortir encore mieux le caractère ordinal des données.	Même nuance ou intervalle spectral partiel de couleurs, avec des différences relativement élevées entre catégories. Par exemple, jaune pâle, orange, rouge moyen, rouge sombre.
Discrets		Dimension des ménages (1, 2, 3... personnes) et non-dimension moyenne des ménages	Même situation que pour les données ordinales, mais des différences plus faibles entre nuances de gris sont acceptables.	Même situation que pour les données ordinales, mais des différences plus faibles entre nuances de gris sont acceptables.

Niveaux de mesure		Exemple	Cartes en noir et blanc	Cartes en couleurs
Continus	Séquentiels	Taux d'alphabétisation (toute valeur comprise entre 0 et 100 %).	Gamme continue de nuances de gris. Le niveau de gris peut ou non être proportionnel aux valeurs de données. Des différences faibles mais observables des niveaux de gris sont acceptables.	Gamme continue de la même couleur ou dans un intervalle spectral partiel. Des variations faibles du niveau de couleur sont acceptables.
	Divergents	Taux de masculinité (inférieur à un = plus de femmes que d'hommes ; supérieur à un = plus d'hommes que de femmes)	On doit recourir à des différences de texture/motif. Des ombrages pleins d'une part et des différences de texture d'autre part peuvent donner de bons résultats.	Couleur neutre (blanc ou gris) au centre avec une gamme continue de deux couleurs différentes de part et d'autre. Par exemple de l'orange clair à l'orange foncé pour les valeurs inférieures à un et du vert clair au vert foncé pour les valeurs supérieures à un.

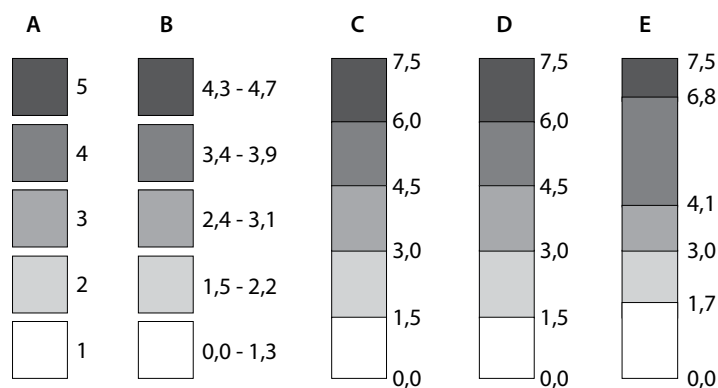
E. Conception de la légende cartographique

A5.109. Le niveau de la mesure peut être reflété dans la conception de la légende, qui établit la relation entre les valeurs de données ou les intervalles de valeurs et les symboles graphiques utilisés. Les logiciels de SIG et les logiciels de cartographie numérique comportent une fonction intégrée de conception des légendes qui convient pour la plupart des applications. Pour une conception cartographique plus soignée, cependant, nous pouvons modifier la légende par défaut, soit dans le module de tracé des logiciels cartographiques, soit dans un logiciel graphique extérieur.

A5.110. La figure A5.28 en fournit quelques exemples. Pour les données catégorielles, les divers cadres de légendes doivent être séparés les uns des autres (voir figure A5.28A). De même, on peut faire ressortir ainsi des séries de catégories non contigües — par exemple, il existe un intervalle entre la limite supérieure d'une catégorie et la limite inférieure de la suivante (voir figure A5.28B). En général, cependant, on devra éviter d'utiliser des légendes de ce type. Les cadres de légendes contigus font ressortir le caractère continu de variables telles que des coefficients ou des densités (voir figure A5.28C). La continuité des valeurs de données ressort encore plus si les cadres individuels de catégories ne sont pas entourés d'un contour (voir figure A5.28D). Enfin, la figure A5.28E montre l'exemple d'une légende pour la classification d'une variable continue avec des séries de catégories irrégulières.

Figure A5.28

Différents types de légendes pour des cartes ombrées

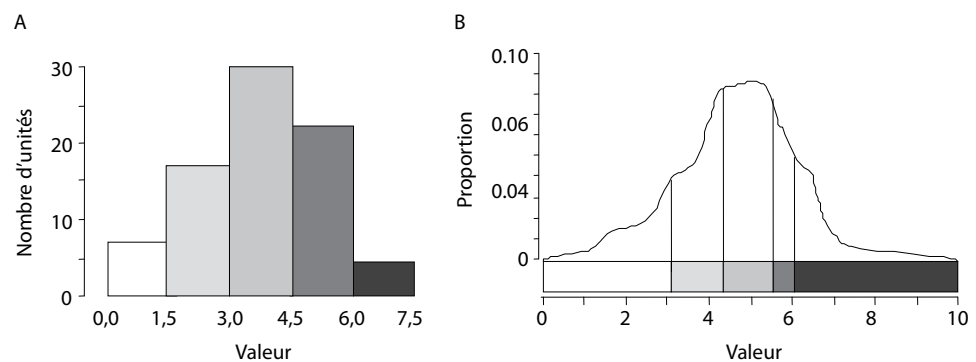


A5.111. Les légendes des figures A5.28C, D et E comportent des points d'inflexion plutôt que des séries de données distinctes. Lorsqu'on utilise des séries de données pour avoir une répartition continue, on se heurte à un problème qui consiste à présenter une même valeur de donnée dans deux catégories, par exemple, 0-10, 10-20, 30-40. Ce problème peut être résolu en utilisant le symbole inférieur à pour affecter chaque valeur à une seule catégorie : par exemple 0-<10, 10-< 20, 20-30. Avec des catégories ouvertes, on peut utiliser le symbole plus grand que ou égal : <10, 10-< 20, >20.

A5.112. La légende peut aussi être intégrée à un diagramme statistique qui résume la répartition des données de la variable. Des histogrammes dans lesquels les couleurs des barres correspondent à la nuance de couleur sont souvent utilisés à cette fin (voir figure A5.29A). Si les intervalles de catégories ne sont pas constants, les barres de l'histogramme peuvent être tracées avec une largeur variable. Si le logiciel cartographique ne prend pas en charge les histogrammes, il est possible d'en tracer dans un programme graphique ou de les importer d'un tableur ou d'un logiciel statistique. Il existe deux possibilités pour déterminer la hauteur de chaque barre. L'approche la plus conventionnelle consiste à utiliser le nombre d'unités géographiques dont les valeurs entrent dans chaque catégorie. Certains logiciels de cartographie numérique affichent le nombre d'unités entrant dans chaque catégorie de la légende. Cela comporte cependant une difficulté : les unités — par exemple des districts — peuvent avoir des tailles de population très différentes. Au lieu d'utiliser le nombre d'unités, on pourra donc déterminer la hauteur des barres de l'histogramme à partir de la taille de la population sous-jacente. Pour une carte de la densité démographique, par exemple, on utilisera le nombre d'habitants présents dans chaque intervalle de densité. Naturellement, la forme de l'histogramme sera très différente et le procédé utilisé devra être indiqué clairement sur la carte ou dans le texte qui l'accompagne.

Figure A5.29

Légendes montrant la répartition des données statistiques



A5.113. Les logiciels statistiques permettent aussi de calculer des schémas de densité montrant la répartition des données sous une forme plus continue qu'avec un histogramme (voir figure A5.29B). La surface située au-dessous de la courbe de densité est égale ou total à un, de telle sorte que la fréquence approchée d'une valeur de donnée individuelle quelconque peut être lue sur le graphique. Des légendes de ce type ont été utilisées, par exemple, dans l'*Atlas of United States Mortality* (*Atlas de la mortalité aux États-Unis*) [National Center for Health Statistics (NCHS), 1997].

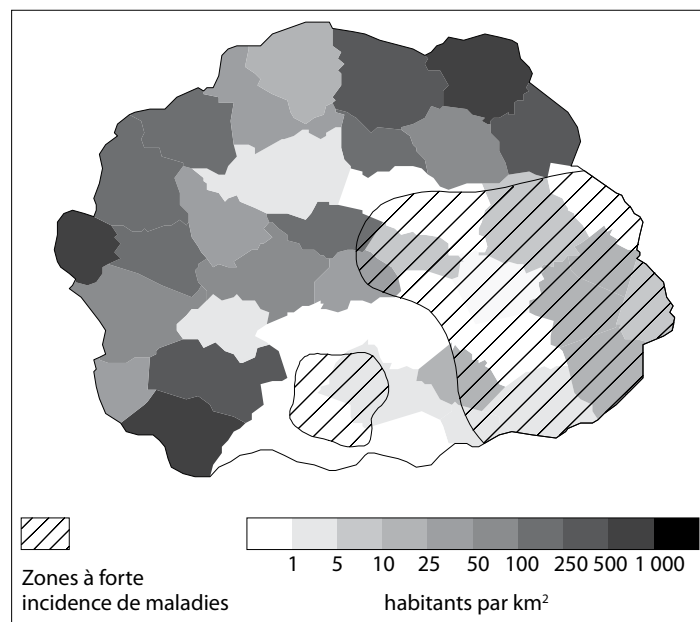
F. Cartes qui racontent des histoires

1. Cartes multivariées

A5.114. À quelques exceptions près, les exemples précédents n'ont montré qu'une variable à chaque fois. C'est le type le plus courant d'affichage utilisé dans un atlas de recensement. À des fins analytiques et pour illustrer les rapports entre variables, nous souhaiterions parfois afficher plus d'une variable à la fois. Dans la section consacrée au choix des couleurs, nous avons fait observer que les cartes multivariées qui utilisent un système de couleurs complexe pour montrer deux variables sur la même carte sont en général difficiles à comprendre. Une autre solution, comme nous l'avons mentionné plus haut, consiste à utiliser un tracé comportant une couleur de fond transparente sur une carte choroplèthe ombrée. Cela donne de bons résultats si la variable superposée ne comporte que quelques catégories ou s'il s'agit d'une variable binaire (présence ou absence) (voir figure A5.30).

Figure A5.30

Combinaison de symboles ombrés pleins et hachurés pour présenter deux variables sur la même carte



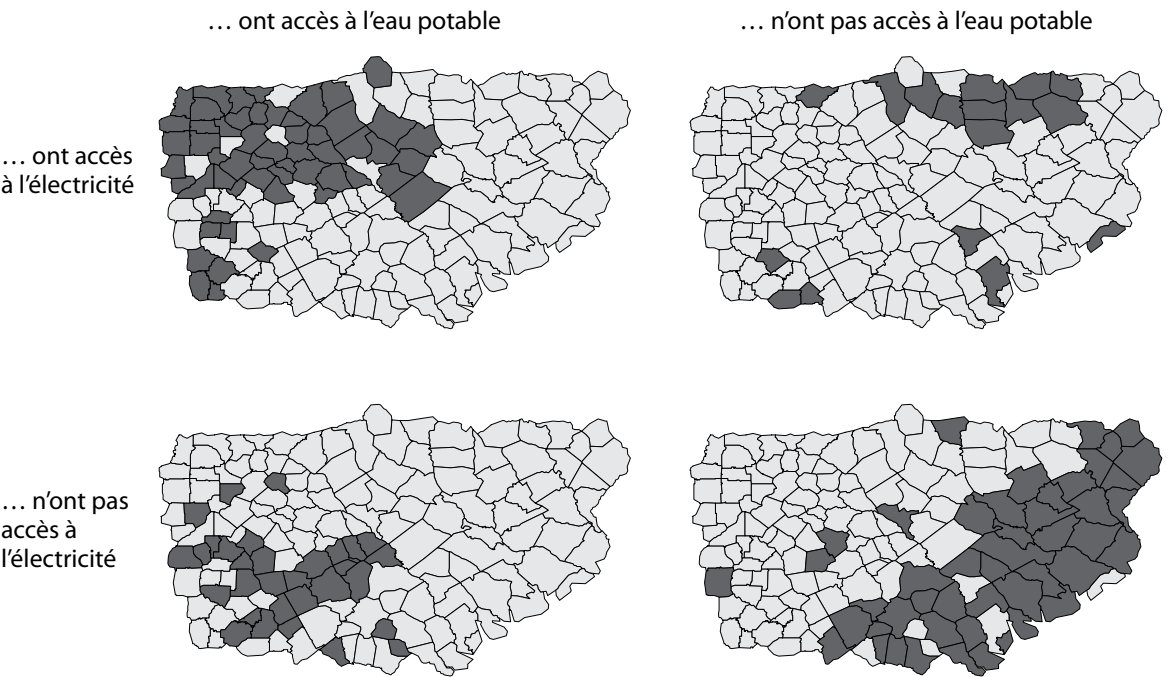
A5.115. Dans l'analyse de données statistiques, deux variables catégorielles qui ne prennent qu'un petit nombre de valeurs sont analysées en utilisant des tabulations croisées. Ces tableaux sont aussi appelés tableaux de contingence. Les lignes et les colonnes d'un tableau à double entrée montrent les catégories des deux variables et les cases montrent le nombre d'observations qui prennent les valeurs correspondantes pour chaque variable. Cet arrangement permet d'évaluer rapidement leurs relations. Par exemple, nous pourrions avoir converti deux variables d'un recensement d'habitations — le pourcentage de ménages ayant accès à l'eau potable et le pourcentage de ménages ayant accès à l'électricité — en deux variables binaires indiquant si la majorité des ménages du district a ou non accès à ces services publics. La tabulation croisée peut ressembler à ce qui suit :

Majorité des ménages ayant...			
	... accès à l'eau potable	... n'ayant pas accès à l'eau potable	Total
...accès à l'électricité	55	17	72
...n'ayant pas accès à l'électricité	31	48	79
TOTAL	86	65	151

A5.116. Si nous voulons présenter ces informations sous forme géographique, nous pouvons établir une carte comportant quatre catégories, une pour chaque case de la tabulation croisée. Cependant, étant donné que ces quatre catégories ne comportent aucun classement naturel, il sera difficile pour le lecteur de découvrir des tendances sur une telle carte. Une meilleure méthode consistera à traduire le concept d'un tableau à double entrée directement en langage cartographique. La figure A5.31. montre une carte équivalant à un tableau à double entrée. Chaque carte indique les districts corrélatifs à la case correspondante dans le tableau à double entrée. Les cartes n'ont pas besoin d'être assorties d'une légende détaillée, car la nuance sombre fait nettement ressortir les districts présentant de l'intérêt.

Figure A5.31
Carte équivalant à un tableau à double entrée

La majorité des habitants ...



A5.117. Des tendances apparaissent immédiatement même sur une petite carte qui ne couvre qu'environ un tiers de page. La plupart des districts du nord-ouest ont accès à la fois à l'eau potable et à l'électricité, tandis que la majorité des ménages dans les districts du sud-est n'ont accès ni à l'un ni à l'autre. Dans les tabulations croisées, les cases hors diagonale sont souvent les plus intéressantes. Dans certains districts du nord-est la plupart des ménages n'ont pas accès à l'eau potable, mais ont accès à l'électricité. Dans un groupe de districts au sud-ouest, la situation est inversée.

A5.118. Cette approche pourrait être étendue à des tableaux plus complexes, par exemple lorsqu'une variable présente trois valeurs (à savoir faible, moyenne et éle-

vée) et une autre comporte deux catégories. Il n'est pas nécessaire de tracer de grandes cartes. Même avec de nombreuses unités géographiques — dans le présent exemple 151 districts — de petites cartes suffisent, car seules deux couleurs ou nuances de gris contrastées sont requises.

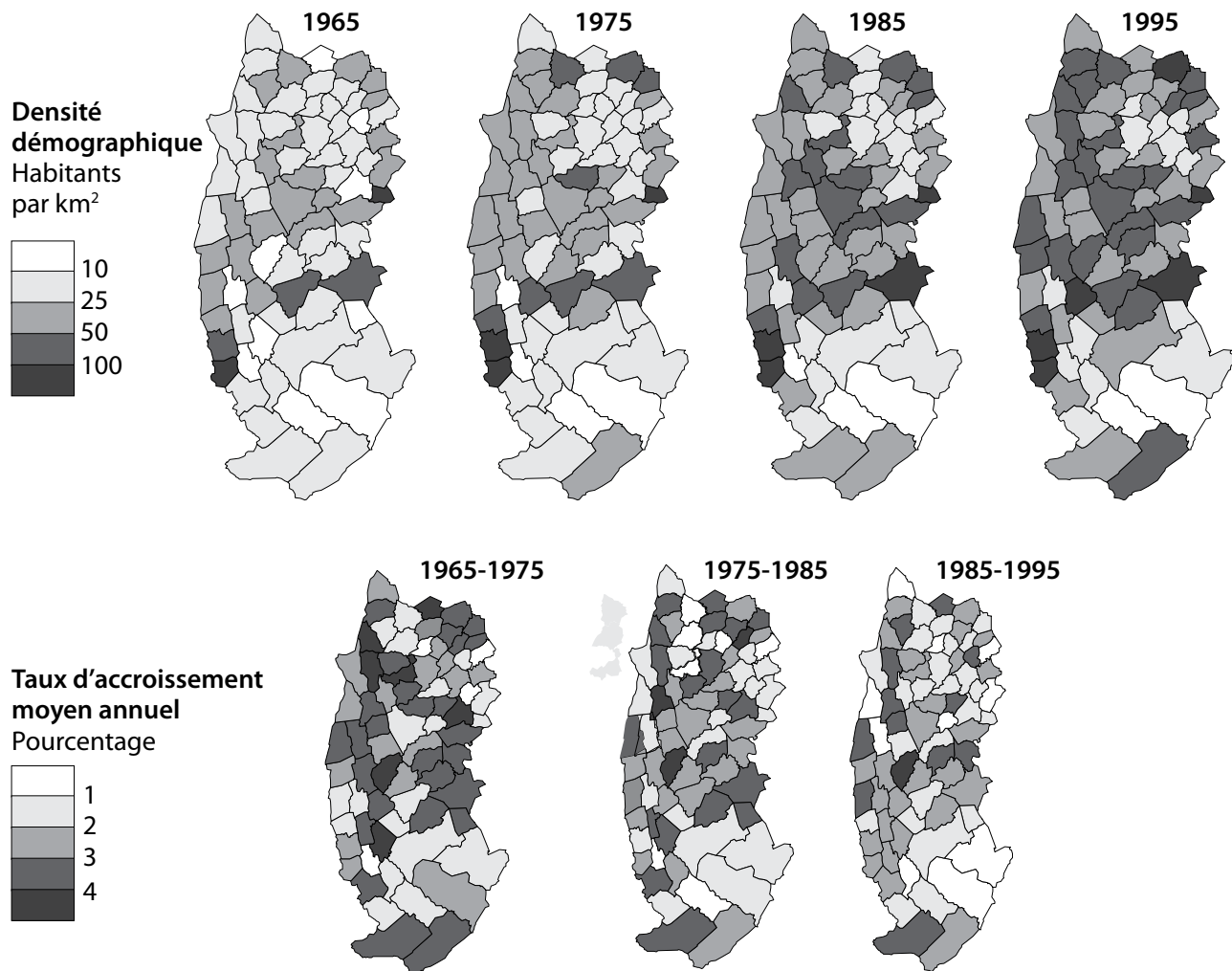
2. Petites cartes multiples

A5.119. On peut aussi présenter efficacement des informations évolutives en disposant les données sur plusieurs cartes. La figure A5.32 montre l'accroissement démographique dans le temps à partir des chiffres provenant de quatre recensements successifs. Les cartes de densité démographique montrent où l'accroissement démographique a été le plus élevé. Pour permettre des comparaisons dans le temps, les limites de catégories doivent être les mêmes sur toutes les cartes. Cela signifie que les systèmes de classification fondés sur la répartition des données (par exemple les inflexions naturelles) ne sont pas appropriés. Les cartes de densité sont complétées

Figure A5.32

Petites cartes multiples — représentation de l'évolution dans le temps

Dynamique de la population, 1965-1995



par trois cartes plus petites montrant les taux d'accroissement démographique moyen annuel entre les recensements.

A5.120. Des présentations du type présenté sur la figure A5.32 sont appelées petites cartes multiples [Bertin (1983) et Tufte (1983)]. Le même tracé de carte est répété pour chaque année ou chaque sous-groupe de population. Le tracé étant le même pour toutes les cartes, elles s'avèrent faciles à comprendre pour le lecteur. Cela permet au concepteur de la carte de présenter une plus grande densité d'informations que ce qui serait autrement possible. Les relations multivariées sont souvent plus évidentes avec des présentations comportant plusieurs cartes qu'avec des cartes composites accompagnées d'une légende éventuellement compliquée.

A5.121. Un document des Nations Unies, *Geographical Information Systems for Population Statistics* [Nations Unies (1997); voir figure 4.8] offre un autre exemple de carte utilisant le concept des petites cartes multiples. Il montre les taux de masculinité concernant des tranches d'âges de cinq ans dans 75 districts au Népal. Le graphique montre 17 petites cartes, avec une classification divergente centrée autour d'un taux de masculinité équilibré. La version en couleurs de cette carte montre un excédent de femmes dans différentes nuances de rouge et un excédent d'hommes sous forme de bleus. La version en noir et blanc utilise des nuances de gris pleines pour un excédent de femmes et une combinaison de points de densité variable pour un excédent d'hommes. À n'en pas douter, la couleur renforce le message communiqué par ces cartes. Bien qu'elles contiennent une grande quantité d'informations, les cartes peuvent être interprétées très facilement, car les groupements de valeurs semblables sont très apparents. Il est évident qu'un tableau de 1 275 valeurs (17 fois 75) serait beaucoup plus difficile à interpréter que les mêmes informations présentées sous forme de cartes géographiques. En fait, le graphique correspondant au Népal fait ressortir clairement certaines tendances qui peuvent être attribuées à la migration pendant le cycle de vie des hommes dans un certain nombre de districts.