

QUELQUES PRINCIPES DE CARTOGRAPHIE STATISTIQUE

INSTITUT GEOGRAPHIQUE NATIONAL

par Jean-Philippe GRELOT

L'information géographique est multiple et complexe. Son analyse est nécessaire pour le chercheur comme pour le responsable de la vie publique. L'étude d'un tableau de chiffres est longue et n'apporte pas l'étalement dans l'espace qu'est celui des données, elle ne fait pas ressortir la similarité et les particularismes régionaux. Seule une carte peut les rendre, à condition de respecter deux règles : la pertinence de la simplification statistique et celle du rendu graphique.

La simplification statistique est nécessaire, et c'est un reproche que les statisticiens adressent parfois à l'encontre de la cartographie. Le cartographe doit connaître les principaux types de distributions de probabilités et de statistiques. Il doit être apte à mener ou du moins à comprendre les méthodes d'analyse statistique employées pour déboucher sur la carte qu'il doit mettre en forme.

Si l'on peut lui pardonner de ne pas être un spécialiste en la matière, on ne lui trouve aucune excuse dans la partie qui lui est propre, la transcription graphique. Pourtant, certains textes relèvent des fautes graves en matière de conception ou d'utilisation de logiciels ; d'autres s'inquiètent de cet état de fait.

Malgré ces mises en garde, le nombre de cartes mal conçues ne fait que croître et prend des proportions catastrophiques avec la multiplication des micro-ordinateurs qui possèdent quelques possibilités graphiques, et par conséquent cartographiques. Que ces piètres cartes soient élaborées par des néophytes en la matière ne fait que traduire la place donnée à la formation graphique, sans même dire esthétique, dans l'enseignement.

Il faut prendre autant de soin à concevoir graphiquement une carte que l'on en a pris lors de l'analyse statistique; aucune des deux tâches n'est intellectuellement inférieure à l'autre. Après la présentation de quelques exemples d'erreurs et le rappel des principales règles, les raisons pour lesquelles les règles sont transgressées seront analysées et quelques remèdes seront proposés.

◆ 1 - Quelques erreurs courantes

Les cartes que l'on rencontre le plus fréquemment utilisent une représentation zonale, c'est-à-dire que la totalité de la zone à laquelle s'attache une valeur statistique est colorée. Plus précisément, il s'agit de cartes choroplètes, c'est-à-dire représentant un phénomène spatial sur un découpage du territoire en zones qui ne

relèvent pas de ce phénomène (contrairement à une carte d'occupation du sol ou à une carte géologique), mais résultent d'une autre logique, généralement une partition administrative. En toute rigueur, cette représentation ne peut s'appliquer qu'à des phénomènes de densité ; elle ne devrait s'appliquer à d'autres phénomènes (proportions, classifications) que si les unités géographiques étaient d'égale surface. Cela mis à part, les erreurs les plus courantes sont relatives à l'emploi de la couleur, de la valeur visuelle, et à la conception de la légende.

■ 1.1 - Couleur

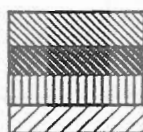
La couleur est souvent employée pour représenter une grandeur statistique ordonnée :

jaune	de 100 à 80
vert	de 80 à 60
rouge	de 60 à 40
bleu	de 40 à 20

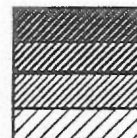
Le seul cas où cet emploi peut être accepté est celui de l'utilisation d'une progression passant par les couleurs dégradées et traduisant un phénomène du type hypsométrie pour lequel les unités géographiques sont des polygones parfaitement emboîtés et dont l'emboîtement suit scrupuleusement la totalité des classes de la légende. On peut traiter ainsi des cartes de température, d'ensoleillement ou de pluviosité, ou encore les distances ou les temps d'accès à un service ou à un équipement. Il faut cependant être là extrêmement prudent et veiller à l'esthétique du résultat.

■ 1.2 - Valeur

Que l'on utilise simplement le noir et blanc ou les nuances plus ou moins foncées d'une couleur, on doit éviter deux fautes :



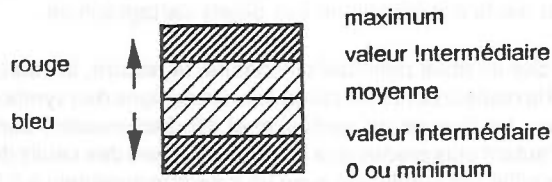
de 100 à 80
de 80 à 60
de 60 à 40
de 40 à 20



supérieur à 20
de 20 à 0
de 0 à -20
inférieur à -20

La première est une succession désordonnée des valeurs visuelles lorsque la variable statistique croît ; la seconde est la succession, cette fois ordonnée du clair au foncé, des valeurs visuelles, mais avec une variable statistique partant de valeurs négatives pour aller vers des valeurs positives, avec ou sans seuil pour la valeur 0 ; notons que pratiquement tous les phénomènes d'évolution présentent ce type de distribution numérique.

Le cas suivant se rencontre également :



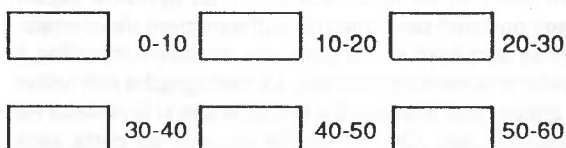
C'est à mon avis un mauvais emploi de la couleur, car la séparation visuelle entre les deux couleurs employées n'est jamais interprétée comme celle des unités situées respectivement au-dessus et au-dessous de la moyenne. On s'en tire habituellement en déclarant que l'on représente non pas le phénomène, mais l'écart à la moyenne: il vaut mieux chercher une représentation graphique plus efficace.

1.3 - Conception de la légende

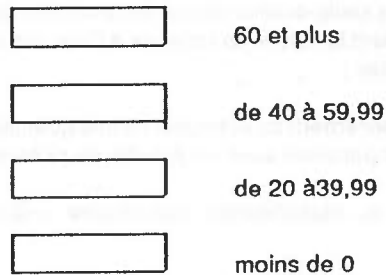
Deux aspects relèvent de la conception de la légende: le choix du nombre de classes et la disposition graphique de la légende.

Le nombre de classes de la légende est en rapport direct avec le nombre d'entités géographiques cartographiées. D'une manière générale, on doit éviter les classes à très faible effectif ; la légende d'une carte avec moins de 25 entités géographiques comportera moins de cinq classes.

Spécialement dans le cas d'une progression ordonnée, le parcours de lecture de la légende doit être logique (mouvement mathématiquement monotone). Doit être proscrite une légende telle que :



qui nécessite des allers-retours oculaires ; dans ce cas, la lecture naturelle se fait dans le sens vertical et non dans le sens horizontal, ce qui vient encore aggraver l'élément perturbateur. Ce qui est vrai pour les caissons l'est aussi pour les valeurs numériques ; le schéma suivant est déconseillé, qui nécessite un parcours de nombres en zig-zag :



2 - Rappel des règles graphiques

Le graphique ou la graphique est un langage de communication. Comme tout langage, il obéit à des règles qui doivent être connues et donc apprises, même si certains énoncent l'universalité et le caractère naturel du langage graphique.

Ces règles ont été longuement exposées, notamment dans l'ouvrage de Jacques BERTIN "Sémiologie graphique" publié en 1967, ou présentées avec des variantes mineures. Contentons-nous de les rappeler.

Les objets graphiques ont trois modes d'implantation: le point, la ligne, la zone.

Un symbole graphique mis en place est la combinaison de neuf variables visuelles

- position,
- forme,
- taille,
- orientation,
- texture ou structure, définie comme un motif graphique élémentaire, visuellement perceptible,
- grain, qui résulte de l'agrandissement "photographique" d'un motif,
- valeur,
- couleur,
- saturation ou pureté de la couleur.

Les six premières relèvent de la perception spatiale, les trois dernières de la perception spectrale.

La variable position est citée ici alors qu'elle est souvent passée sous silence. La cartographie générale l'utilise dans au moins deux cas : la mise en page de la carte et le positionnement des toponymes. Elle trouve des aspects en cartographie socio-économique lorsque l'on déforme sensiblement les unités géographiques tout en conservant une ressemblance morphologique et surtout en respectant les positions respectives : c'est le cas des cartogrammes construits à partir de formes géométriques simples, ou encore des réseaux déformés par un modèle de contraintes mécaniques.

La saturation d'une couleur, d'autant plus forte que sa gamme spectrale est plus étroite, est difficile à juger. Sa perception dépend fortement du contraste simultané engendré par la proximité des plages colorées, et semble perturber la lecture plutôt que d'ordonner. Nous n'en parlerons pas davantage.

Les cartes socio-économiques représentent des phénomènes dont la nature se rattache à l'une des catégories suivantes :

. Le **dénombrement** ou estimation d'une quantité absolue par comparaison avec un échelle de référence ;

. l'**ordre** ou classification quantitative relative des objets ;

. la **typologie**, ou classification qualitative des unités géographiques selon une taxonomie.

Leur lecture met en oeuvre des relations visuelles, parmi lesquelles on retrouve le dénombrement et l'ordre, ainsi que l'associativité et la sélectivité. L'associativité, ou équivalence, traduit une classification purement qualificative, par des signes qui peuvent être perçus comme visuellement semblables ; on lit généralement signe après signe, sans opérer de regroupement visuel : c'est typiquement la lecture ana-

lytique d'un inventaire. La sélectivité, ou discrimination, au contraire, conduit à opérer des regroupements en familles, à partir d'une légende où les symboles sont bien différenciés. La performance de l'association et de la sélectivité dépend du mode d'implantation graphique du phénomène étudié.

On dresse alors (voir ci-dessous) les deux tableaux des représentations permises par les variables visuelles avec une efficacité plus ou moins grande, et des variables visuelles que l'on peut employer en fonction du mode d'implantation des objets cartographiés.

Sur un objet ponctuel ou linéaire, la texture, la valeur et la couleur sont tributaires des dimensions des symboles ; les risques de confusion et d'indiscernabilité sont d'autant plus grands que l'on s'approchera des seuils de lisibilité : un cercle doit avoir un diamètre supérieur à 0,2 mm ; un carré doit avoir un côté supérieur à 0,4 mm ; la largeur minimale d'une ligne (ou d'un blanc entre deux lignes colorées) est de 0,2 mm.

	Phénomènes			Modes d'implantation		
	Dénombrement	Ordre	Typologie	Point	Ligne	Zone
Couleur			xx	x	x	xx
Valeur		xx		x	x	xx
Taille	xx	xx		xx	xx	xx
Grain		x	x		xx	xx
Texture			x	x	x	xx
Orientation			x	xx	x	xx
Forme			x	xx	x	

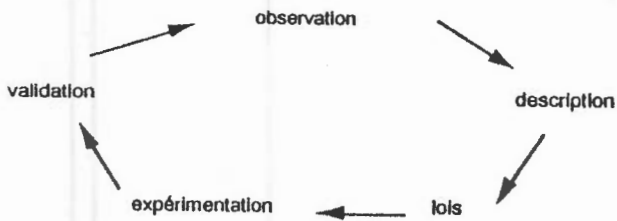
◆ 3- Pourquoi les règles sont-elles transgressées ?

On a dit que le langage graphique était universel. Cette affirmation doit être nuancée, car tout symbolisme figuratif ou abstrait porte en lui des connotations culturelles variant donc dans l'espace et le temps. Le postulat d'universalité a un avatar pervers, l'innéité : puisque ce langage est universel, alors il fait partie de la conscience collective innée, il est naturel, et n'a pas besoin de faire l'objet d'apprentissage.

Cette dénégation de la formation doit être combattue. Les règles sont peu nombreuses, elles n'ont rien qui aille à l'encontre du bon sens, il est du devoir de tout cartographe de les apprendre. Elles seront peut-être enfin un jour inscrites au programme de l'enseignement élémentaire. En attendant, chacun peut s'exercer en critiquant des cartes existantes pour en dégager les imperfections, en se posant à chaque fois la question: "Comment peut-on traiter plus justement et plus efficacement cette représentation ?" sans perdre de vue les moyens techniques mis en oeuvre.

Il faut déjouer le piège de la facilité. Les options par défaut des logiciels sont généralement mauvaises ; ne le seraient-elles pas, elles correspondraient à une catégorie de représentations, donc une catégorie de cartes. Tant que l'on ne disposera pas d'un système expert assez puissant pour intégrer suffisamment de connaissances acquises en ce domaine, on devra travailler à améliorer le rendu graphique. Le cartographe doit rester un artiste, prêt à reprendre son ouvrage si le résultat ne le satisfait pas. Qu'il le veuille ou non, sa carte sera appréhendée comme une image et regardée selon des critères esthétiques. L'efficacité de la communication visuelle n'est atteinte que si le lecteur entre dans le jeu graphique à son insu ; ce n'est pas à lui de faire l'effort de compréhension, c'est au concepteur de faire l'effort de créativité, ce qui n'est somme toute pas plus difficile que de trouver la bonne méthode d'analyse statistique pour tirer la substantifique moëlle des chiffres.

Les contraintes matérielles sont parfois présentées comme une excuse. C'est un tort : à la limite, il vaut mieux ne rien montrer plutôt que diffuser une mauvaise carte. Malgré toute la bonne volonté qu'il aura pu y mettre, l'auteur d'une mauvaise carte a toute les chances de transmettre un message erroné. N'oublions pas que la démarche scientifique est un cycle hypothético-déductif :



Les données guident le choix de la représentation, leur distribution statistique celui du seuillage : c'est la phase de description, qui s'achève par le rendu de la distribution spatiale. Celle-ci est analysée et exprimée comme une loi, qu'il convient de vérifier ; l'expérimentation est, par exemple, l'analyse des écarts au modèle de distribution ; s'ils sont acceptables, la loi est validée, et le phénomène décrit vient enrichir le savoir pour des observations ultérieures. Le cycle est souvent parcouru plusieurs fois, ce qui correspond à la définition de la loi par approximations successives.

Dès lors, la démarche à suivre est tributaire des ressources. L'auteur doit se dire : je dispose de telles données, elles me révèlent tel ou tel aspect géographique ou socio-économique ; avec les moyens graphiques dont je dispose, je puis concevoir telle et telle représentation ; quel phénomène puis-je cartographier, avec laquelle de ces représentations ? Encore une fois, faisons le parallèle avec l'analyse des données : telles données, tel logiciel permettent de faire telle et telle chose ; si cela ne suffit pas, l'analyste ira chercher d'autres données ou d'autres méthodes d'analyse, il ne s'arrêtera pas à un état inachevé pour énoncer doctement une interprétation qu'il jugera honnêtement sujette à caution.

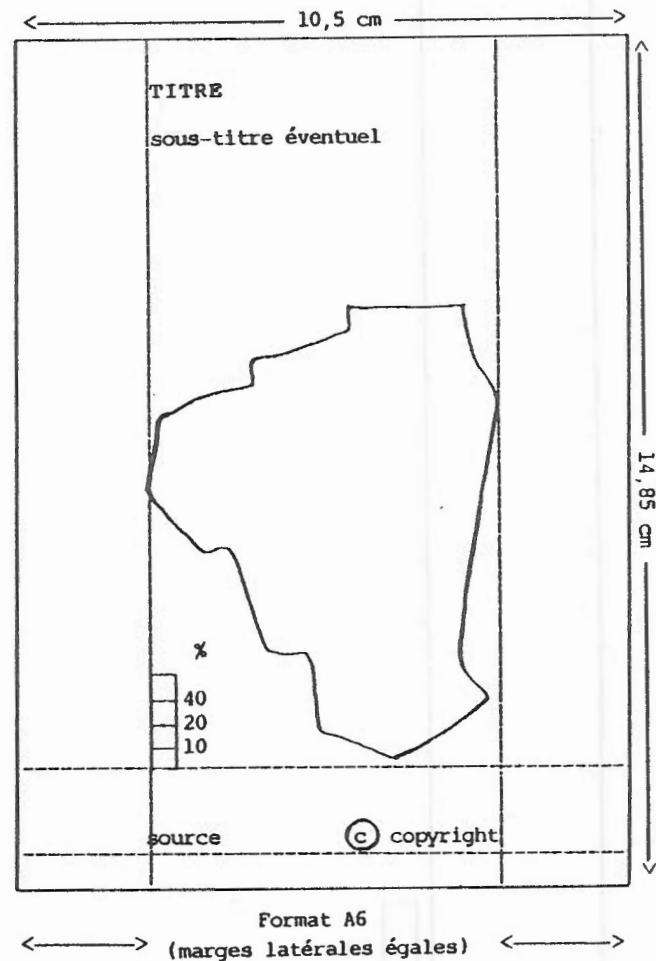
◆ 4 - Propositions

■ 4.1 - Mise en page

Le territoire cartographié s'inscrit dans un espace ; la page ou l'écran graphique. Le mise en page ne consiste pas à disposer les éléments comme ils se présentent, mais à construire cet espace. Les éléments à disposer ou habillage, sont :

- le titre, aussi concis que possible ;
- un sous-titre éventuel, plus technique, comportant la date de validité si elle ne figure pas dans le titre ;
- le territoire cartographié ;
- la légende ;
- les sources utilisées (fichiers géographiques, statistiques, logiciel...)
- le copyright ;
- le cas échéant, un commentaire explicatif de quelques phrases.

L'arrangement de ces éléments dépend de la forme du territoire cartographié. En s'appuyant sur ses extensions extrêmes, on disposera l'habillage de manière à obtenir un rectangle légèrement plus petit que la page (au format normalisé de préférence) ou l'écran.



■ 4.2 — Carte de situation

Il n'est pas souhaitable de figurer sur les cartes des données de localisation, autres que les limites des unités statistiques, qu'il s'agisse d'un réseau géographique ou de noms de villes. En effet, ces éléments occulteraient le thème représenté, ou partageraient des unités cartographiques. Pourtant, et surtout si le territoire est mal connu du lecteur, ces informations doivent être communiquées. On les représentera soit dans un cartouche de plus petit format pour les cartes isolées, soit sur une carte d'introduction pour un recueil de plusieurs cartes ; dans ce cas, l'expérience montre qu'une feuille volante de papier convient parfaitement car la lecture par comparaison de deux cartes à même échelle se fait aisément, alors qu'un calque altère les couleurs de la carte qu'il couvre, et qu'un plastique transparent est relativement onéreux.

14,85 cm

1990 - TITRE
sous-titre



10 km

commentaire bref

source

© copyright

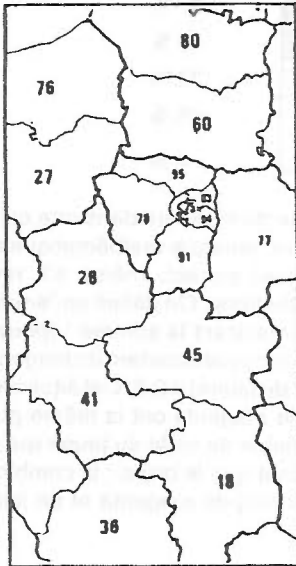
format A5

4.3 - Généralisation

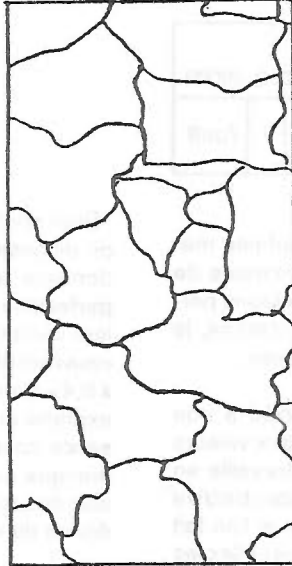
La carte n'est pas un fond que l'on colore ; c'est un tout, et le fond lui-même doit être adapté au phénomène représenté. Il arrive fréquemment que l'on extraie du fond initial les limites d'unités agrégées ; ainsi, les limites cantonales d'un fichier communal ; cette opération est en fait la première phase d'une généralisation, qui s'accompagne d'une simplification géométrique des limites. Il faut en effet veiller à supprimer tout détail qui nuirait à la lecture, ou qui donnerait aux limites cartographiées une illusion de précision en contradiction avec le cumul statistique auquel on a dû procéder.

Cette simplification géométrique doit faire apparaître une structure. Les deux extrêmes sont présentés dans la figure ci-après. On évitera un lissage mou (a), pour préférer un trait volontaire (b) qui donne du dynamisme à la carte.

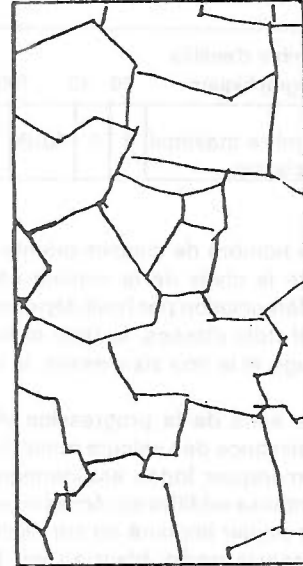
La généralisation offre un autre avantage ; elle réduit sensiblement le temps de calcul ou de dessin de la carte. Le degré de généralisation sera aussi choisi en relation directe avec la résolution graphique du support de tracé ; une carte sur écran sera plus généralisée qu'une carte sur imprimante à laser.



fond original



(a)



(b)

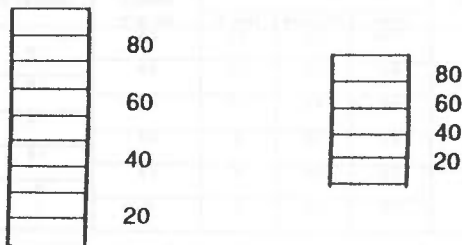
4.4 - Légende

La concision est de rigueur, qui sous-tend la logique. Les caissons sont disposés verticalement, allant de haut en bas des valeurs élevées aux valeurs faibles. L'indication des seuils face aux intervalles séparant les caissons évite les fastidieux : de 40 à 59,99 , de 60 à 79,99 ... à la précision souvent illusoire.

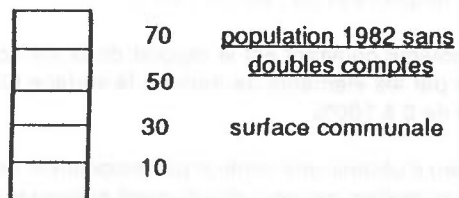
L'unité dans laquelle les seuils sont exprimés est écrite une seule fois, au-dessus et à l'aplomb des seuils : il ne sert à rien de la répéter après chaque seuil, cela ne fait que surcharger sans apporter d'information supplémentaire. En revanche, il peut être utile d'expliciter cette unité par une formule donnant la définition précise des statistiques utilisées. Par exemple, l'unité "habitants/km²" sera complétée par la mention:

population 1982 sans doubles comptes

surface communale



habitants/km²



Si la légende ne peut être disposée en colonne, par exemple sur un écran graphique, elle sera disposée sur une seule ligne selon le même principe ; la croissance se fait cette fois de gauche à droite, les seuils étant situés au-dessus des intervalles séparant les caissons.



Le nombre de classes doit être cohérent avec le nombre d'entités géographiques. En dehors de cas particuliers, on se tiendra dans les limites suivantes :

Nombre d'entités géographiques		20	40	100	300	1000	4000	20000	
Nombre maximal de classes		3	4	4ou5	5	5ou6	6	7	7ou8

Le nombre de classes oriente dans une certaine mesure le choix de la couleur car les performances de différenciation par l'oeil dépendent d'elle. Le jaune permet trois classes, le bleu et le violet cinq classes, le rouge et le noir six classes, le vert sept classes.

Le sens de la progression visuelle associée à une croissance des valeurs numériques faibles aux valeurs numériques fortes est identique, que l'on travaille en synthèse additive sur écran ou en synthèse soustractive sur papier imprimé ou sur cliché ; il le reste si l'on fait passer le fond du blanc au noir. Les valeurs visuelles les plus claires sont associées aux grandeurs numériques faibles, les plus foncées aux grandeurs élevées.

■ 4.5 - Principes d'obtention des couleurs

L'impression des couleurs se fait au moyen de textures très fines, appelées trames, que l'on combine entre elles. Chaque élément de trame reçoit ensuite des pigments colorés, et les éléments de trame sont suffisamment fins pour que l'oeil opère la synthèse des couleurs. Les caractéristiques d'une trame sont la linéature, l'orientation et le pouvoir couvrant.

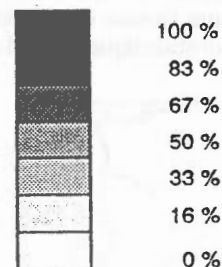
La linéature exprime la finesse de la trame ; on retiendra que l'impression sur papier journal se fait avec des trames à 80 lignes au pouce alors que l'impression ordinaire se fait avec des trames à 130 lignes au pouce.

L'orientation joue pour éviter les phénomènes de moirage. Ainsi, on combine les couleurs primaires cyan, magenta, jaune et le noir en leur attribuant les orientations respectives 30°, 60°, 0°, 45°.

Le pouvoir couvrant est le rapport de la surface couverte par les éléments de trame à la surface totale, et varie de 0 à 100%.

Au lieu d'obtenir une couleur par association des couleurs primaires, on peut directement mélanger les encres. On se réfère alors au nuancier appelé "gamme Pantone", et on travaille en ce cas sur une couleur que l'on peut dégrader en modulant le pouvoir couvrant.

Toutefois, on maîtrise mieux le rendu des couleurs en utilisant les combinaisons des trois couleurs primaires (cyan, magenta, jaune). La référence est donnée par une charte, où chaque couleur primaire est désignée par son pouvoir couvrant. Ainsi, dans une carte où chaque couleur primaire est décomposée en sept niveaux allant du blanc (pouvoir couvrant nul) à l'aplats (pouvoir couvrant 100%), les pouvoirs couvrants successifs seront, en graduation continue : 0%, 16%, 33%, 50%, 67%, 83%, 100%.



Pour choisir une gamme de couleurs dans une carte, on dispose d'une méthode presque mathématique qui donnera un résultat visuel correct, même s'il reste perfectible sur le plan esthétique. On définit un "équivalent niveau de gris" comme étant la somme : (pouvoir couvrant du cyan) x 0,4 + (pouvoir couvrant du magenta) x 0,4 + (pouvoir couvrant du jaune) x 0,2. Cet équivalent exprime que le cyan et le magenta ont la même puissance colorimétrique, double de celle du jaune qui est presque aussi réfléchissant que le blanc ; la combinaison des trois aplats de cyan, de magenta et de jaune donne du noir.

Ainsi, une variable statistique exprimant une grandeur relative induit une progression de valeurs qui, comme nous l'avons vu, se fait dans le même sens que la progression de la variable. Au sens strict, une gradation de valeurs suffit. Mais, soit que l'on travaille en trichromie, soit que l'on module la couleur pour des raisons esthétiques, on recourt très simplement aux pouvoirs couvrants et à l'équivalent niveau de gris pour opérer un choix raisonnable. D'une nuance claire à la nuance immédiatement plus foncée, le pouvoir couvrant de chaque couleur primaire croît, ou reste égal ; les différences tabulaires des équivalents niveaux de gris croissent lorsque l'on va des nuances claires aux nuances foncées ; par exemple :

Variable	Pouvoir couvrant (en %)			Equivalent niveau de gris	Différence tabulaire
	cyan	magenta	jaune		
80	100	80	20	76	18
60	80	60	10	58	16
40	60	40	10	42	14
20	40	30	0	28	12
10	20	20	0	16	8
	10	10	0	8	

A noter que le blanc total n'est jamais la nuance la plus claire d'une gradation, il signifie absence de valeur.

Pour une typologie, on cherchera les nuances qui ont le même équivalent niveau de gris ; par exemple, on aura :

Pouvoir couvrant (en %)			Equivalent niveau de gris
cyan	magenta	jaune	
60	0	80	40
50	20	60	40
40	30	60	40
30	40	60	40
20	60	40	40
0	80	40	40

Là encore, il faut comprendre que les principes calculatoires fournissent un guide, et les considérations esthétiques permettent de moduler.

Les périphériques graphiques (traceurs, écrans...) sont classés en trois catégories selon leur aptitude à manier les deux variables visuelles valeur et couleur :

- les périphériques monochromes tels que les traceurs électrostatiques noir et blanc ou les écrans cathodiques dont les particules de phosphore sont identiquement colorées en blanc, ou en vert, ou en orange, etc...
- les périphériques polychromes "tout ou rien" qui typiquement codent la couleur sur trois bits dont chacun correspond à une couleur de base (cyan-magenta-jaune ou rouge-vert-bleu) ; certains ordinateurs personnels fonctionnent selon ce principe ;
- les périphériques polychromes à large éventail de nuances, tels que les traceurs électrostatiques couleur ou les écrans graphiques, qui constituaient voilà peu le haut de gamme, et qui tendent aujourd'hui à devenir la norme.

Les aptitudes de ces différents périphériques sont données dans le tableau ci-dessous :

Type de périphérique	Valeur visuelle	
	valeur	couleur
monochrome	oui	non
polychrome "tout ou rien"	non	oui
polychrome à nuances	oui	oui

Le seul emploi de la variable valeur avec les périphériques polychromes "tout ou rien" est celui où l'on passe du blanc au noir en intercalant une couleur primaire et une couleur binaire qui la contient. On obtient ainsi les progressions blanc, cyan, bleu, noir d'une part ; blanc, magenta, rouge, noir d'autre part. Les autres combinaisons (cyan-vert, magenta-bleu, jaune-rouge, jaune-vert) ne peuvent pas être considérées comme des progressions de valeur. Toutefois, avec ce type de périphérique, l'usage des trames d'une bibliothèque graphique (qui sont en fait des textures), permet d'obtenir des dégradés des couleurs disponibles.

Il y a deux façons d'obtenir les documents tramés qui vont servir à l'impression :

- soit on effectue, pour chaque couleur finale à obtenir, la décomposition en trames correspondant aux trois couleurs primaires ; c'est la rédaction par couleurs séparées, qui conduit directement au résultat ;
- soit on produit une planche pour chaque couleur composite à obtenir ; sur chaque planche, les surfaces occupées par la couleur figurent en aplat (pouvoir couvrant 100 %), et le photogaveur se charge de la traduction en trames et du regroupement ; c'est la rédaction par thèmes séparés, que tout traceur graphique permet d'effectuer.

4.6 - Choix de la représentation

Une grandeur absolue se représente toujours en implantation ponctuelle ou linéaire par des symboles dont on fait varier la taille. La comparaison visuelle se fait bien lorsque la forme est simple ; cercle, carré. La variable statistique peut être seuillée, c'est-à-dire que l'on utilisera le même symbole pour tous les individus groupés dans une même classe par seuillage. Si le phénomène est à implantation ponctuelle, la surface S du symbole est proportionnelle à la variable $V: S = kV$; le coefficient K est choisi de manière que la surface totale occupée par les symboles soit de l'ordre de 1/7 de la surface du territoire cartographié : il s'agit d'une approximation, d'autant que l'on ne tient pas compte du recouvrement entre symboles voisins ; lorsqu'il y a recouvrement, les petits symboles entourés d'une couronne évidée, viennent sur les plus gros. Pour un phénomène à implantation linéaire, tel que flux ou échange, la largeur du symbole est proportionnelle à la variable et, là encore, la surface des symboles est de l'ordre de 1/7 de celle du territoire ; mais la gestion des superpositions se fait de manière plus souple : en général, les flux sont d'autant plus petits que le point d'arrivée est plus éloigné du point de départ, et la lecture peut se faire par prolongement, même si les symboles les plus larges effacent les plus étroits. Le diamètre minimal d'un cercle est de 0,2mm, le côté minimal d'un carré est 0,4 mm, la largeur minimale d'une ligne 0,2 mm.

Une grandeur relative est représentée par une variation de valeur en implantation zonale ; il suffit d'aller du clair au foncé dans la même couleur, mais on y ajoute souvent une variation de la couleur elle-même. Par grandeur relative, on entend ici soit une densité (phénomène rapporté à la surface sur laquelle il s'applique) auquel cas la représentation est légitime, soit une proportion ; la représentation souffre alors de l'éventuelle hétérogénéité des surfaces des unités cartographiées. Il faut, en particulier, y veiller dans les zones étendues où les deux termes de la proportion ont des valeurs faibles, mais où leur rapport est grand sans réellement être significatif ; il est, dans ce cas, préférable d'indiquer, par un symbole approprié, que la donnée n'est pas significative.

Une typologie est représentée par une variation de couleur ou de texture ; les couleurs sont choisies de manière à avoir le même impact visuel.

Certaines combinaisons de représentations sont possibles. Ainsi, lorsqu'une variable prend des valeurs négatives d'une part, positives d'autre part, on établit une distinction entre ces deux catégories en attribuant à chacune une couleur (couleur froide pour les valeurs négatives, chaude pour les valeurs positives), puis on module la valeur dans chaque catégorie.

4.7 - Discrétisation des données

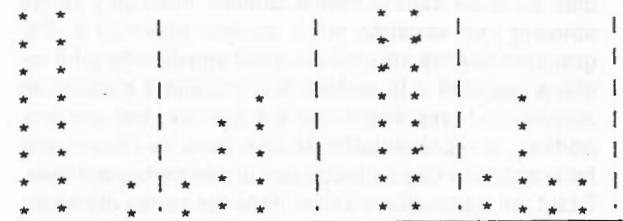
Pour tirer partie de la communication par l'image, en acceptant ses contraintes, et dégager les niveaux d'information correspondant aux niveaux de lecture, on regroupe les valeurs prises par la variable statistique en classes. Cette opération de discrétisation n'est pas aisée, car il ne faut pas oublier qu'elle influe sinon conditionne la lecture de la carte, et donc son interprétation. Il n'y a pas de clé universelle, il y a simplement des outils pour éviter les erreurs manifestes.

Le cas favorable est celui où un seul seuil existe "naturellement", en quelque sorte de manière exogène à l'acte cartographique ; il en est ainsi du résultat des élections dans un scrutin majoritaire, ou du taux de renouvellement de la population pour l'étude du taux de fécondité.

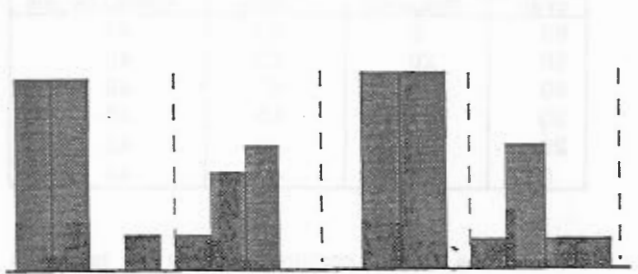
Les méthodes de discrétisation reposent sur un principe de bon sens ; un seuil doit être choisi de manière que, si on le fait varier un peu, le résultat cartographique ne change pas. C'est-à-dire que l'on cherche une certaine objectivité dans la fixation des seuils. On utilise pour cela des méthodes visuelles ou des méthodes mathématiques.

Les méthodes visuelles consistent à faire un graphique de la distribution, puis à placer les seuils selon la forme du graphique :

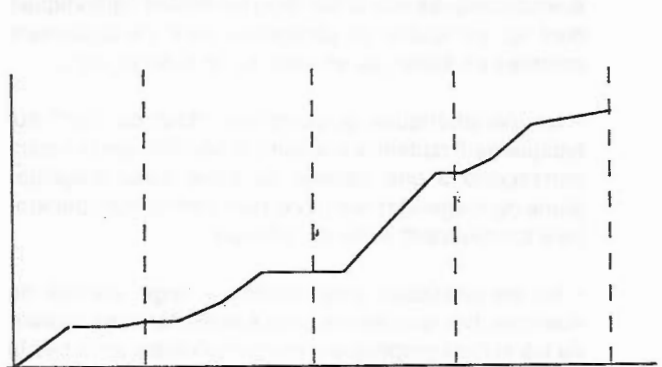
à l'état brut, on se contente de mettre, sur l'axe des abscisses, un point pour chaque valeur prise par la variable ; on place les seuils dans les interstices les plus longs.



si l'on a un très grand nombre d'individus prenant les mêmes valeurs, on trace un histogramme de fréquences, et on place les seuils dans les creux de l'histogramme ; les creux apparaissent d'autant plus nettement que la distribution présente plusieurs modalités ; toutefois, les classes de l'histogramme ne doivent pas être trop étendues, car elles prédestineraient alors en grande partie les seuils finaux :



des données brutes ou de l'histogramme, on dérive la courbe cumulative, dont les replats sont évidemment les creux de l'histogramme ou les interstices du premier diagramme ; les seuils sont donc placés aux valeurs correspondant aux replats ;



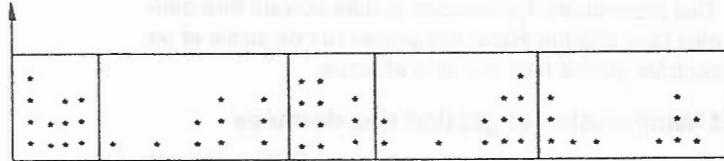
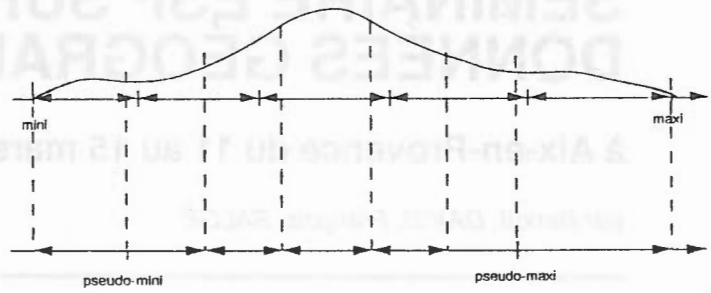
Les méthodes mathématiques déterminent des seuils que l'on adapte ensuite légèrement pour leur faire respecter le principe qui a servi de base aux méthodes visuelles.

Si l'on a montré par un test adéquat (χ^2 par exemple) que la variable elle-même, ou la transformée de la variable (on dira variable recodée), suit une loi normale, on prendra comme seuils, en désignant par m la moyenne et σ l'écart-type :

- si la valeur moyenne est en soi significative, $m \pm p \sigma$, p prenant des valeurs entières positives à partir de 0 ;

- si la valeur moyenne n'est pas significative, c'est-à-dire si l'on considère qu'une faible variation autour de la moyenne n'est pas de nature à modifier l'interprétation que l'on doit faire, $m + (2p + 1) \frac{\sigma}{2}$, p prenant des valeurs entières positives à partir de 0.

Lorsque cette méthode n'est pas applicable, on effectue un seuillage par quantiles, en pratiquant l'équirépartition soit des individus eux-mêmes, avec un nombre d'individus par classe égal au quotient de la population totale par le nombre de classes, soit des surfaces correspondantes si les unités statistiques ont des superficies très hétérogènes. De la même façon que pour la méthode issue de la normalisation, on adapte les seuils ainsi calculés pour les faire tomber sur des replats de la courbe cumulative. On veille dans tous les cas à ce que deux individus de même occurrence tombent dans la même classe.



Si la distribution n'est pas trop dispersée, on calcule des classes d'égale étendue. Si n , $mini$ et $maxi$ désignent respectivement le nombre de classe, la valeur minimale et la valeur maximale de la distribution, l'étendue de chaque classe sera :

$$\frac{maxi-mini}{n} \text{ et les seuils seront}$$

$$mini, mini + \frac{maxi-mini}{n}, mini + 2 \frac{maxi-mini}{n}, \dots$$

$$mini + (n-1) \frac{maxi-mini}{n}, maxi$$

Dans le cas où la distribution est dispersée, avec des valeurs extrêmes ayant un faible nombre d'occurrence, on adapte la méthode des classes d'égale étendue, afin d'éviter de dilater les classes centrales qui recevraient la majeure partie des individus. A cet effet, on élimine de la distribution les individus prenant les 5% de valeurs les plus faibles et les 5% de valeurs les plus fortes. On désignera respectivement par "pseudo-mini" et "pseudo-maxi" les valeurs minimale et maximale de cette distribution dérivée, sur laquelle on applique la méthode des classes d'égale étendue. On repasse ensuite à la distribution initiale en reclassant les individus éliminés dans les classes extrêmes. Si n est le nombre de classe, l'étendue des classes a été transformée

de la valeur $\frac{maxi-mini}{n}$ en une valeur plus faible

$$\frac{pseudo-maxi - pseudo-mini}{n}$$

et les seuils $mini + p \frac{maxi-mini}{n}$ (pour p prenant les valeurs entières de 1 à $n-1$)

sont devenus $pseudo-mini + p$

$$\frac{pseudo-maxi - pseudo-mini}{n} ;$$

les valeurs $pseudo-mini$ et $pseudo-maxi$ ne sont pas des seuils (voir figure ci-après).

La partie concentrée de la distribution apparaît ainsi mieux ventilée dans les différentes classes.

4.8 - Séries de cartes

Lorsque l'on doit représenter non une carte unique, mais une série de cartes, le choix de discrétisation doit se faire entre un seuillage unique pour l'ensemble des cartes et un seuillage pour chaque carte.

Le nombre de classes doit être choisi en fonction du nombre d'unités géographiques dans une carte, ou du nombre moyen lorsque la série porte sur des entités géographiques différentes : par exemple, représentation de la densité de population par commune pour chaque département d'une région administrative.

Lorsque la série porte ainsi sur plusieurs entités :

- des seuils identiques donnent des équivalences d'ensemble, une indication de l'homogénéité de la distribution sur la totalité de l'espace : lorsque les entités sont contiguës, on peut d'ailleurs les rassembler en une carte unique ;
- des seuils adaptés pour chaque carte permettent la comparaison des structures spatiales globales, se déroulant sur l'ensemble de l'entité, ou locales (typiquement autour des agglomérations) : ce procédé gomme l'écart entre les valeurs moyennes de chaque entité, et éventuellement les divergences de concentration des distributions statistiques, mais représente la variabilité locale de la distribution.

Lorsque, sur une même entité géographique, on dispose d'une série multi-temporelle :

- des seuils identiques donneront la propagation d'ensemble, la tendance générale, autour de laquelle les particularismes locaux et leur évolution apparaîtront en second plan ;
- des seuils adaptés à chaque carte, c'est-à-dire à chaque période, montreront le degré de stabilité de la distribution géographique, des groupes géographiques dominants, une fois gommée la tendance générale de l'évolution.

◆ CONCLUSION

La banalisation de l'outil cartographique avec la diffusion des micro-ordinateurs est un facteur favorable à la connaissance géographique des phénomènes socio-économiques. Les fastidieuses heures de dessin ne sont plus d'actualité, et la carte est produite très rapidement : que son concepteur veuille bien consacrer quelques minutes à "penser" sa représentation, afin que la communication qu'il veut établir avec son lecteur soit efficace. Alors, la carte ne sera pas un gadget mais un message.