Discrétisation

L'information géographique peut revêtir plusieurs formes. Parmi ces dernières, la «représentation» cartographique présente l'avantage de la lecture instantanée d'une image, une lecture qui doit aller du général au particulier. L'avantage que constitue cette lecture ne peut exister qu'au prix d'un investissement préalable fait lors de la conception et de la réalisation de la carte. Cet investissement concerne aussi bien la sélection pertinente de l'information que le traitement de cette information. Le traitement préalable de l'information dépend du type de caractère statistique que l'on veut cartographier. En fonction de ce caractère les choix des méthodes de discrétisation sont plus ou moins nombreux et les résultats cartographiques peuvent donner des images très variées.

La discrétisation est l'opération qui permet de découper en classes une série de variables qualitatives ou de variables quantitatives. Cette opération simplifie l'information en regroupant les objets géographiques présentant les mêmes caractéristiques en classes distinctes.

Une discrétisation est satisfaisante lorsqu'elle permet la création de classes homogènes et distinctes entre elles : les objets géographiques d'une même classe doivent se ressembler plus entre eux qu'ils ne ressemblent aux objets des autres classes. Discrétiser une série statistique constitue souvent l'ultime étape de la réduction, de l'organisation et de la hiérarchisation de l'information avant de construire une carte qui rend compte de la répartition géographique de cette même série statistique.

L'opération de discrétisation doit satisfaire à la fois aux exigences de la représentation cartographiques et à celles des principes statistiques. Elle doit conserver les caractéristiques essentielles présentées par les données, perdre le moins d'information possible, mais aussi respecter les règles de la perception visuelle afin de transmettre une information géographique efficace et de qualité.

Le nombre de classes optimum à réaliser dans une partition est toujours fonction du nombre d'individus observés (unités spatiales). Il existe un indice permettant de connaître le nombre de classes idéales pour une distribution, il faut le considérer uniquement comme une aide indicative. Il s'agit de l'indice de Huntsberger :

$$N(cl) = 1 + 3{,}3 \log_{10}(N)$$

N = nombre d'observations
N(cl) = nombre de classes

Les principes de la discrétisation diffèrent avec la nature de l'information :
-Si l'information est qualitative :

-ordinale : on cherche à conserver la «hiérarchie» des informations ;

-nominale : on cherche à définir un critère commun de regroupement pour aboutir à la construction d'une typologie.

Il n'existe pas de méthode spécifique pour discrétiser l'information qualitative : chaque regroupement d'objets géographiques est propre aux objectifs de la simplification choisie des données.
-Si l'information est quantitative :

de stock ou de rapport : on cherche à conserver la forme de la distribution ou à mettre en avant des valeurs particulières.

Le choix d'une méthode de discrétisation dépend des propriétés de la distribution et des objectifs que l'on s'est fixé quant à l'information à communiquer.
Trois informations sur les caractéristiques de la distribution doivent être préservées :

HYPERGEO

-l'ordre de grandeur, mesuré par les valeurs centrales de la distribution : mode, moyenne ou médiane. Ces valeurs peuvent être choisies comme limite de classes et faire apparaître les valeurs exceptionnelles qui informent sur les particularités de la distribution ;

-la dispersion, est une «mesure» de l'inégalité des valeurs, caractérisée par l'écart-type, l'intervalle inter-quantiles ou le coefficient de variation ; ces valeurs peuvent être choisies comme amplitude des classes. Cette mesure prend en compte la variance : elle permet de minimiser les différences entre les objets d'une même classe (variance intra-classe) et de maximiser les différences entre les classes (variance inter-classe) ;

-la forme, elle peut être de type normal, symétrique ou dissymétrique. Si la série statistique observée est petite, la réalisation du diagramme de distribution permet l'évaluation de la forme de la distribution. Si la série est grande, la comparaison des valeurs centrales permet de la déterminer.

Les principes de la discrétisation diffèrent avec les objectifs à atteindre :

-s'il s'agit de construire une carte unique, la meilleure méthode est celle qui rend le mieux compte de la structure interne des données : les « seuils naturels » observés sur le diagramme de distribution ou tout autre méthode qui respecte la forme de la distribution ;

-s'il s'agit d'établir des comparaisons sur le même ensemble d'objets géographiques (par exemple pour une collection de cartes), il faut absolument utiliser des méthodes de discrétisation basées sur le calcul de seuils statistiquement signifiants. Là encore tout dépend de l'objectif de la comparaison :

-Pour comparer les positions des mêmes «lieux» selon plusieurs caractères : les unités de mesure sont différentes il faut employer une méthode permettant d'utiliser les mêmes paramètres statistiques de chaque série (par exemple la moyenne et l'écart-type) ;

-Pour comparer les dispersions de plusieurs séries statistiques avec les mêmes lieux : cette comparaison n'est valable que si les données sont de même nature avec les unités de mesure identiques et l'on utilisera la méthode des amplitudes ou effectifs égaux en jouant sur la gamme des gris ou des couleurs.

-Pour comparer une même série à des dates différentes : il faut établir la discrétisation sur l'ensemble des valeurs observées de la série statistique pour l'ensemble des dates.

Les méthodes de discrétisation :
Il existe un grand nombre de méthode de discrétisation et de variantes. Le choix d'une méthode dépend à la fois des propriétés de la distribution, des possibilités de la représentation cartographique et des objectifs que l'on se fixe (message, public, support).
Les quatre méthodes les plus faciles à mettre en œuvre et les plus couramment utilisées en géographie sont les suivantes : la méthode des seuils naturels, la méthode des amplitudes égales, la méthode des effectifs égaux et la méthode de la progression géométrique.

HYPERGEO

[gallery link="file" ids="1132"]

Bibliographie

Bibliographie :
-Cauvin (C), Reymond (H), Serradj (A), Discrétisation et représentation cartographique, Montpellier, GIP RECLUS (Col. Reclus modes d'emploi), 1987.
-Palsky G., Des chiffres et des cartes. Naissance et développement de la cartographie quantitative au XIXe siècle, Paris, CTHS, 1996.
-M. Beguin et D. Pumain, La représentation des données géographiques, coll. Cursus Armand Colin, 2e édition 2000
-E. Blin et J.P. Bord, Initiation géo-graphique ou comment visualiser son information, Sedes, 2e édition 1995
-C. Zanin et M.L. Trémélo, Savoir faire une carte, coll. Sup Géographie Belin, 2003

HYPERGEO