

QUAND LE STATISTICIEN JOUE SUR LA FASCINATION DE LA CARTE

La saga cartographique de l'Insee

par *Christophe Terrier*

www.christophe-terrier.com

L'Insee a pour mission d'établir des statistiques mais également de les analyser. Parmi les statistiques produites par l'Insee figurent des données portant sur des territoires. Il est donc naturel que l'Insee produise des cartographies à base de statistiques. La réalisation de cartes statistiques avec des techniques diverses a soulevé et soulève encore des problèmes relevant tantôt de la technique statistique, tantôt de la sémiologie graphique mais aussi parfois des deux à la fois. Cet espace de rencontre entre la statistique et la cartographie offre un champ d'étude que l'on ne finit pas d'explorer. On évoquera les techniques de lissage, employées pour répondre à la question : comment cartographier correctement des données établies sur des maillages hétérogènes ? On évoquera bien sûr les cartes interactives qui se sont fortement développées. On n'oubliera pas les cartes qui déforment la géographie sous la pression des données statistiques (anamorphoses selon différents modèles) et parfois même du temps (anamorphoses animées). Et puis les films cartographiques (évolution de la densité de population sur 100 ans), et puis les cartographies des flux, et puis... Sans oublier le dernier pavé lancé par l'Insee dans la mare des statistico-cartographes avec la mise à disposition de données carroyées. On se centrera ici sur quelques cartes produites par le « monde statistique ». En examinant leurs aspects statistiques et leurs aspects graphiques, on s'interrogera sur les enseignements que l'on peut en tirer en matière de sémiologie statistico-cartographique. Et peut-être pourrons nous, avancer des éléments de réponse aux questions fondamentales : une carte statistique doit-elle être belle, voire fascinante, pour être parlante et utile ? Une belle carte peut-elle véhiculer une information que contredirait une analyse statistique bien faite ?

Introduction

En quarante ans d'Insee, essentiellement consacrés aux statistiques locales – donc à la géographie –, j'ai beaucoup utilisé la cartographie sous toutes ses formes. Ayant été responsable national des études régionales à l'Insee, puis détaché à la Datar et ensuite au ministère du Tourisme, j'ai contribué à la réalisation et à l'édition de nombreuses cartes statistiques dont certaines ont pu avoir un impact non négligeable dans le débat public. C'est à partir de cette expérience que je vais essayer ici d'aborder la question de la relation entre cartes et chiffres en évoquant quelques étapes choisies de cette saga cartographique.

Le point de vue adopté ici est celui du statisticien – pas obligatoirement inculte en matière de géographie mais en tous cas pas spécialement formé à cette discipline – à la conquête de la cartographie.

Au commencement était le chiffre

La mission de l'Insee, c'est de faire du chiffre, on appelle ça des statistiques. Établir des statistiques, c'est faire l'état de la situation, faire l'état des lieux. D'abord on définit très précisément ce que l'on mesure et dans quel cadre, puis on dénombre et on compte. En matière

géographique, l'Insee est investi d'une double mission : celle de donner le cadre de référence et celle de mesurer ce qui se passe dans les compartiments définis par ce cadre.

Pour le cadre, l'Insee est chargé de la gestion du « Code officiel géographique » (Insee, 2016 et Lang, 2015) qui définit l'existence et la composition des communes et leur appartenance aux niveaux géographiques supérieurs : canton, arrondissement, département, région. Chaque territoire se voit attribuer un code ; pour la commune ce code est composé de 5 chiffres dont les 2 premiers pour le département et les 3 suivants l'ordre de la commune dans le département. Pour sa part, l'IGN se charge de définir la topographie de ces mailles territoriales.

Pour la mesure, l'Insee établit, notamment par le recensement, les statistiques – démographiques, économiques, sociales - se rapportant à chaque territoire.

Du chiffre à la carte

Les statistiques une fois produites peuvent être éditées sous une forme numérique (tableaux) ou sous une forme graphique (courbes, cartes...). C'est à la forme cartographique que l'on s'intéressera ici mais les autres formes graphiques de la représentation statistique

forment un corpus plus large dont la sémiologie n'est évidemment pas absente.

Donc l'Insee fait des cartes à partir des statistiques. Le principe habituel est d'appliquer à la surface attribuée à un territoire une forme graphique correspondant à la valeur de la variable telle que déterminée par l'opération statistique. Le cadre, c'est-à-dire la surface, la forme et l'emplacement dévolus au territoire, est donné par le « fond de carte », fourni en général par un organisme tiers, souvent l'IGN. Les formes graphiques utilisées sont diverses, du simple à-plat de couleur aux formes les plus variées. Il arrive aussi que la mesure statistique porte non pas sur le territoire lui-même mais sur la relation qu'il entretient avec d'autres territoires : appartenance à une zone ou flux entre territoires. Il faudra dans ce cas utiliser d'autres formes graphiques. Dans tous les cas, il s'agit, à partir d'une mesure statistique, de donner à voir sous une forme cartographique l'information recueillie par l'opération statistique.

De la carte faite à la main aux automatismes

On n'a pas attendu l'informatique pour faire des cartes à partir de statistiques. On peut se référer à l'excellent ouvrage de Gilles Palsky « Des chiffres et des cartes » (Palsky, 1996). Je reste personnellement fasciné par un atlas réalisé sous l'égide de la Datar par l'IGN et une équipe universitaire parmi laquelle figurait Jacques Bertin (Datar-IGN, 1967). Cet atlas, qui a été réalisé dans les années 1950, donc avant l'apparition des méthodes de cartographie automatique, comporte en particulier une superbe carte de la densité de population par commune. Savoir que les 36 000 communes ont été peintes à la main, une par une, à l'aquarelle, avec une gamme de 12 couleurs correspondant aux seuils retenus pour les classes de densité me remplit d'admiration pour l'artiste qui l'a réalisée. Il mérite bien le titre administratif à l'époque en vigueur à l'IGN : « artiste cartographe en chef ».

La carte statistique, summum du mensonge

On a beaucoup glosé sur les statistiques qui seraient la forme la plus évoluée du mensonge : il y a les petits mensonges, les gros mensonges et puis les statistiques... Souvent les politiques utilisent les statistiques pour leur faire dire n'importe quoi et parfois le contraire de ce qu'elles signifient. Quand on a, par ailleurs, lu l'ouvrage « Comment faire mentir les cartes » (Monmonier, 1993), on peut se demander si le cocktail « statistiques » plus « cartographie » ne risque pas de nous emmener vers des sommets en matière de mensonge.

En quarante ans d'Insee, essentiellement consacrés aux statistiques locales – donc à la géographie –, j'ai beaucoup utilisé la cartographie sous toutes ses formes. À la période où j'étais responsable national des études régionales à l'Insee, puis détaché à la Datar, je disposais d'un jeu de cartes statistiques me permettant de faire des exposés sur à peu près tous les sujets couverts par la statistique territoriale. Je me souviens en particulier d'un débat organisé à Rouen par les collègues de l'Insee avec des universitaires locaux. Après les exposés nous avons eu un débat assez vif avec nos interlocuteurs, débat au cours duquel j'avais projeté (c'était l'époque des rétroprojecteurs) de nombreuses cartes à l'appui de mes arguments. Je ne me souviens plus du sujet de la controverse mais je me souviens très bien que notre « adversaire » avait fini par jeter l'éponge tout en nous lançant sur un ton accusateur : « vous jouez sur la fascination de la carte ».

Depuis ce temps je m'interroge : dois-je être fier d'avoir fait ou utilisé des cartes suffisamment parlantes pour transmettre clairement une information ou dois-je être honteux d'utiliser des armes déloyales avec de belles images fascinantes qui jettent de la poudre aux yeux ?

Du dénombrement statistique à sa communication

Il y a plusieurs façons de communiquer l'information produite par une collecte de statistiques. On peut se contenter de donner les chiffres bruts de collecte ; ce serait un peu comme donner un morceau de minerai de fer à qui a besoin d'un couteau. On peut effectuer des redressements ou des calculs pour fournir une information plus « juste », « toutes choses égales par ailleurs », ou plus « informative » en calculant des rapports, des pourcentages, des évolutions...

Ensuite, on peut aussi présenter ces résultats sous forme de tableau ou sous une forme graphique qui fera mieux apparaître l'information apportée par ces statistiques. Au-delà des courbes et des diagrammes souvent utilisés par les statisticiens, la forme cartographique offre un vecteur particulièrement intéressant pour « faire parler » les chiffres lorsqu'ils portent sur des territoires.

Chacune de ces étapes nécessite d'opérer des choix et qui dit choix dit « expertise » et donc risque de subjectivité, voire d'erreur ou de mensonge si l'on est incompetent ou malintentionné.

Le premier niveau de choix, primordial, est celui de la nomenclature : qu'est-ce que l'on mesure exactement et dans quelles catégories classe-t-on ce qui fait l'objet

de la mesure ? En matière d'information localisée, les choix sont souvent multiples. Par exemple, pour un recensement de la population, on choisit la maille territoriale de collecte : on cumule dans la même case tous les logements et les individus dénombrés à l'intérieur du périmètre de cette maille. On sera ensuite obligé d'établir des conventions pour déterminer qui « habite » dans les logements recensés : ce n'est pas parce que l'enquêteur ne trouve personne dans le logement à un moment donné que personne n'y habite ; on peut être parti travailler pour la journée ou la semaine, ou parti en voyage ou autre. Et que faire avec ceux qui habitent régulièrement à plusieurs endroits (élèves pensionnaires, étudiants, travailleurs lointains...). Et les enfants en garde alternée, et les retraités multi-résidents ?

Ces exemples, non exhaustifs, ont pour seul but d'attirer l'attention sur le fait que l'expertise, et donc le risque de subjectivité, existe avant même que l'opération de collecte statistique soit entreprise. Et encore parle-t-on ici d'objets statistiques définissables sans trop d'ambiguïté : logements et individus. La situation se complique lorsqu'on traite de sujets plus difficiles à cerner comme les emplois et les entreprises, objets statistiques ô combien ambigus dont la définition relève de conventions multiples.

Le deuxième niveau de choix correspond au traitement statistique après la collecte. Il y a bien sûr toutes les opérations qui font le métier de l'expert statisticien, en particulier le redressement de la non-réponse. Il y a ensuite les calculs de pourcentages et autres opérations permettant, à partir du matériau brut de collecte, de tirer de l'information utilisable pour l'analyse.

Le troisième niveau consiste à présenter l'information sous forme graphique. On se centrera ici sur les formes cartographiques de l'information ; il faut cependant noter que les distinctions classiques entre tableau statistique, graphique statistique et carte statistique s'effacent largement avec les outils interactifs actuellement disponibles qui permettent de combiner allègrement toutes ces formes de représentation de l'information.

Pour les statisticiens, la cartographie a longtemps été considérée comme un simple procédé de présentation d'une information élaborée par des procédés statistiques. À l'usage, on s'est rendu compte – surtout avec l'avènement des méthodes cartographiques complexes – que la carte pouvait aussi être un outil d'analyse statistique à part entière, donnant à voir une information qui n'aurait pas été révélée par les méthodes statistiques habituelles.

Des cartes d'affichage

Un grand nombre de cartes réalisées à partir de statistiques ont pour principal intérêt d'offrir un affichage géographique de l'information. Dans le cas, par exemple, de cartes de France par région, vous irez regarder l'information sur l'Alsace quelque part en haut à droite de la carte au lieu d'aller la chercher dans un tableau classé par code de région ou par ordre alphabétique (ce qui, dans le cas des régions, n'est pas le même ordre ; et on ne parle pas du désordre introduit par le nouveau regroupement régional). En plus, la carte offre l'avantage de montrer les positions géographiques respectives et donc de permettre une comparaison facile entre les régions voisines. Et puis, un éventuel effet géographique, s'il existe, sera visible sur la carte et non sur le tableau statistique.

Ces cartes, que l'on qualifiera de « cartes d'affichage » forment l'essentiel de la production statistico-cartographique. Leur sémiologie graphique a déjà été largement étudiée et on ne s'y étendra pas ici. Tout au plus fera-t-on remarquer que les outils grand public de cartographie automatique ont fait perdre à certains toute prudence statistique dans le traitement de l'information. En particulier, les systèmes automatiques de répartition en tranches, utilisés sans une bonne analyse statistique préalable, amènent parfois à faire n'importe quoi ; je l'ai vu faire, et pas seulement chez les « non-statisticiens ».

Au passage, on pourrait s'intéresser aux différentes étiquettes accolées aux métiers de « géographe » ou de « statisticien » qui se sont fortement rapprochés. En matière de traitement des données localisées, le niveau de qualification statistique est souvent très élevé chez ceux qui sont étiquetés « géographes », au point même que l'Insee – institution au sein de laquelle l'approche territoriale n'est pas la culture dominante – a fait appel à ces « géographes » pour former à la statistique spatiale ses « statisticiens » chargés d'études territoriales.

Quelques cartes particulières

On tournera ici autour de quelques cartes particulières réalisées et publiées par les statisticiens de l'Insee ou des services de la statistique publique et qui ont pu soulever des questions spécifiques en matière de sémiologie statistico-cartographique.

La cartographie au petit point

Il faut sans doute avoir connu la mécanographie ou les débuts de l'informatique – celle où la carte perforée était reine et où les écrans n'existaient pas – pour comprendre la fascination qu'exerçait l'image sur ces informaticiens qui en étaient privés. Quelle

fascination de voir apparaître sur des imprimantes qui ne connaissaient que les caractères (lettres et chiffres) des images reproduites par décomposition en points (de nos jours on parlerait de pixellisation). Les fichiers (sous forme de cartes perforées et plus tard de bandes magnétiques) circulaient sous le manteau permettant de représenter la Vénus de Milo, Bambi ou autres...

C'est en utilisant ce procédé que l'on a pu réaliser les premières cartes statistiques automatiquement générées par l'ordinateur à partir de tableaux statistiques. L'Insee a publié sous cette forme de nombreuses cartes à partir du système SIRF (Statistiques et indicateurs des régions françaises).

Un logiciel maison, baptisé Aubusson, a même été développé en interne pour tirer le meilleur parti de cette technique (Terrier, 1979). L'Insee Lorraine a utilisé ce logiciel pour la réalisation d'un Atlas (Insee-Lorraine, 1980). Pour obtenir un rendu plus fin, les cartes étaient imprimées en plusieurs « listings » qui étaient ensuite recollés entre eux à la manière du papier peint pour former des panneaux larges d'environ un mètre qui étaient ensuite photographiés et réduits au format A4 pour l'édition. C'était un bel exploit technique mais, vu de notre fenêtre actuelle, ces cartes ne sont pas « belles » et n'offrent pas d'autre intérêt que d'être un marqueur des débuts de la cartographie assistée par ordinateur.

Une perception différenciée des cartes en relief

Le premier matériel « moderne » dont ont été dotés les services de l'Insee pour la réalisation de cartes était constitué de tables traçantes à quatre couleurs. Elles étaient utilisées via le logiciel SasGraph, extension du logiciel SAS (Statistical Analysis System) qui avait été introduit à l'Insee quelques années plus tôt. Avec ce type de matériel, il était plus facile de tirer des traits que de remplir des plages de couleur, d'où peut-être la multiplication des cartes en « oursins » pour représenter les liens établis entre des territoires par les déplacements domicile-travail ou par l'attraction des équipements.

Le logiciel comprenait également des fonctions permettant de déformer le relief de la carte en fonction des valeurs prises par la variable. L'exploration de cette technique s'est avérée, pour les apprentis-sorciers que nous étions, à la fois riche de promesses et déroutante. Riche parce que cette technique permettait de « mettre en relief » l'information statistique et donc de la faire apparaître de façon plus évidente. Déroutante parce que, en jouant sur les paramètres du système, l'image obtenue – et donc le message transmis – pouvait changer de façon significative (Terrier, 1989).

Je garde en mémoire une carte représentant, en relief, l'emploi par commune en Lorraine. Les collègues du service diffusion de l'Insee, baptisé à l'époque « Observatoire économique de Lorraine », avaient jugé que cette carte était trop « techno » et ne devait pas être publiée dans la revue « Économie Lorraine » destinée au public. Cette carte avait cependant été affichée, parmi d'autres, sur un mur du hall d'entrée du Conseil régional. Au sortir de la délibération de l'assemblée plénière du Conseil régional, le président de la région a eu son œil accroché par cette carte, en a compris immédiatement le message et a demandé à voir les cartes équivalentes d'autres régions pour pouvoir se comparer.

La première conclusion que j'en ai tiré, et que j'ai eu l'occasion de vérifier de nombreuses fois par la suite, c'est que tout le monde n'a pas la même perception visuelle. Une même carte peut être très parlante pour certains et rester hermétique pour d'autres. À partir de là, je me suis souvent efforcé, pour représenter la même statistique, de montrer plusieurs cartes utilisant des modes de représentations différentes. Certaines personnes sont rebutées par l'usage du relief, d'autres – parfois les mêmes – par les techniques de lissage. J'ai même rencontré un préfet pour qui l'absence des limites départementales sur une carte de France était une transgression telle qu'elle occultait totalement l'information portée par la carte.

La deuxième conclusion que j'en ai tiré est que, lorsqu'une carte est réellement comprise par un dirigeant, il en demande aussitôt une autre !

Quand la statistique et la cartographie ne s'accordent pas

Tout avait pourtant bien commencé. D'abord les organismes statistiques des différents pays d'Europe, fédérés par Eurostat, s'étaient accordés sur des nomenclatures d'unités territoriales statistiques : NUTS 1, 2, 3, 4 et 5. Pour la France on avait choisi pour NUTS1 les ZEAT (Zones d'étude et d'aménagement du territoire), pour NUTS2 les régions, pour NUTS3 les départements, rien pour NUTS4 et pour NUTS5 les communes. Tout le monde savait que les nomenclatures n'étaient pas totalement cohérentes d'un pays à l'autre. Bien sûr, les 36 000 communes françaises étaient très nombreuses comparées aux NUTS5 des autres pays européens qui avaient procédé – presque partout mais pas partout – à un regroupement de leurs communes. Mais, comme souvent dans les instances européennes, on était content d'avoir établi un compromis qui faisait consensus.