

# Construire l'information géographique : des données individuelles à l'analyse des territoires

Luc Guibard, Université de Paris et UMR 8504 Géographie-cités,  
luc\_guibard@orange.fr ;

Romain Leconte, Université de Paris, UMR 8504 Géographie-cités,  
romain.leconte@univ-paris-diderot.fr

<b>Type de Feuille</b>	Feuille de cours ( <b>TD</b> )
<b>Niveau</b>	Licence 2 ou 3
<b>Durée</b>	2 séances de 1h30 à 2h

## Objectifs

1. Décomposer les étapes de la construction de l'information géographique à partir de données individuelles : récolte de l'information auprès des individus, transformation en tableau élémentaire, en matrice d'information géographique, et en matrice de relation spatiale. Construire les objets géographiques, des individus aux territoires, par l'agrégation spatiale.
2. Identifier différents types de questions en analyse spatiale et mettre en forme des tableaux de données adaptée.
3. Sensibiliser aux enjeux éthiques des données individuelles, tant du point de vue de leur manipulation (anonymisation, secret statistique) que du point de vue de leur production par des tiers (organismes publics) et par les individus eux-mêmes.
4. Observer la répartition spatiale de la pauvreté à Paris avec les données des Caisses d'Allocations Familiales (CAF).

## Mots-clés

Information géographique, analyse spatiale, données individuelles, éthique des données, géographie sociale

## Remarques sur la réception auprès des étudiants

### Atouts de l'exercice

La mise en situation des étudiant·e·s dans une démarche administrative, qui leur est pour partie familière, permet à la fois une mise en activité rapide du groupe et suscite une certaine curiosité pour l'exploitation par la géographie de cette situation.

Pour des étudiants ayant déjà manipulé des tableaux de données, l'exercice permet de tisser des liens entre les différents enseignements de statistique, cartographie et SIG en détaillant la formalisation de ces tableaux de données.

### Difficultés

---

L'exercice manipule de nombreux termes dans des acceptions spécifiques à l'analyse de données (individu, variable, maille...) qui peuvent entraîner de la confusion pour les étudiant·e·s qui n'auraient pas de compétences préalables en statistiques spatialisées.

L'exercice implique un temps d'acculturation au fonctionnement de la CAF et des catégories qu'elle mobilise (foyer allocataire, seuil de bas revenu).

Les étudiant·e·s ont des niveaux de sensibilité et d'information très hétérogènes sur les enjeux éthiques et politiques des données personnelles. Les questions du TD peuvent soulever des échanges longs et contradictoires mais nécessaires.

---

*Cette feuille a bénéficié de l'expérience collective et cumulative des enseignants successifs d'analyse spatiale de l'Université Paris Diderot.*

## Table des matières

<b>A. Présentation.....</b>	<b>3</b>
<b>B. Support de cours à destination des étudiant·e·s.....</b>	<b>5</b>
<b>C. Proposition d'accompagnement pédagogique.....</b>	<b>22</b>
<b>Pour aller plus loin.....</b>	<b>31</b>
Bibliographie .....	32

## A. Présentation

### *Objectif général*

L'information géographique, par-delà la diversité de ses définitions et de ses formalisations, constitue une composante indispensable, car transversale de l'enseignement des méthodes et des outils de l'analyse géographique. La cartographie (Lambert et Zanin, 2016), la statistique spatiale (Dumolard et *al.*, 2003), l'analyse spatiale (Pumain et Saint-Julien, 2010) ou les SIG (Aschan-Leygonie et *al.*, 2019) ont pour socle commun l'analyse d'un matériau structuré sous forme d'information géographique, en trois composantes : sémantique, topologique et géométrique.

L'information géographique est une construction, issue d'un processus de sélection, de hiérarchisation et de formalisation. Cette étape de la démarche géographique présente une large dimension analytique. L'explicitation de cette démarche peut s'envisager comme un préalable aux exercices d'initiation méthodologique (Baron et Zanin, 2005a, 2005b, 2006 ; Commenges, 2017) mais également dans le cadre d'un retour réflexif sur les analyses produites.

L'exercice proposé consiste à confronter les étudiants à un matériau brut qui n'a pas été encore mis en forme pour l'analyse géographique. Il s'agit de lui faire subir les transformations nécessaires pour le faire correspondre aux différents formats que peut prendre l'information géographique : tableau élémentaire, matrice d'information géographique, matrice de relation spatiale.

### *Application*

Le cas présenté ici invite les étudiants à s'interroger sur les dimensions spatiales des inégalités sociales. Ces dimensions peuvent être la localisation (où se concentrent les plus pauvres ou les plus riches ?), la distance aux ressources (quelle accessibilité aux services, à l'emploi ?), et la territorialisation des politiques publiques (par exemple, la politique de la ville, le Réseau d'Éducation Prioritaire...). Dans le cadre de cette réflexion, on propose aux étudiants d'analyser l'inscription spatiale de la pauvreté à Paris en utilisant les données de la Caisse d'Allocations Familiales (CAF) de Paris. Plus précisément, c'est la localisation des foyers allocataires à bas revenus qui est étudiée, un foyer allocataire étant constitué d'un. e allocataire et de l'ensemble des personnes partageant son logement et ses ressources.

Cette application permet d'aborder les étapes de la construction de l'information géographique, de la collecte de renseignements au guichet de la CAF à la production d'un résultat cartographique mettant en évidence les espaces de pauvreté à Paris. À partir d'un formulaire de *Déclaration de situation pour les prestations familiales et les aides au logement* édité par les CAF, on interroge les étapes de sélection, de hiérarchisation et de formalisation qui ont conduit à la représentation cartographique. La représentation cartographique n'est mobilisée ici que comme l'une des possibilités d'analyse parmi d'autres qu'offre un formalisme général de l'information géographique.

### *Mise en pratique*

Les étudiants sont tout d'abord mis dans la situation de devoir effectuer une démarche administrative courante, en prenant connaissance d'un formulaire de la CAF. Si cette situation est banale pour nombre d'entre eux — 47,8 % de la population française était

couvert par la CAF en 2018<sup>1</sup> — on veillera à ne pas faire d'équivalence entre le fait d'être allocataire de la CAF et la situation de pauvreté. En effet, plus des deux tiers des allocataires de la CAF possèdent des revenus supérieurs au seuil de bas revenus. De plus, l'approche de la pauvreté par le revenu mobilisée dans cette application est restrictive. Il convient en effet d'insister sur le caractère multidimensionnel de la pauvreté. Dans un second temps, on formalisera l'information recueillie dans le formulaire en réfléchissant aux principaux choix qu'implique la sélection de variables et de mailles d'agrégation spatiale, ainsi qu'aux questions géographiques auxquelles peut répondre ce formalisme. Enfin, un moment est consacré à sensibiliser les étudiant·e·s aux implications éthiques de l'analyse des données individuelles à caractère personnel en géographie et des règles à respecter en termes d'anonymisation et de secret statistique.

---

1 Données au 31 décembre 2018, d'après l'*open data* de la CAF, <http://data.caf.fr/dataset>, consulté le 23.09.2020.

## B. Support à destination des étudiant·e·s

### *Liste des documents :*

- Doc. 1 : Formulaire — Déclaration de situation pour les prestations familiales et les aides au logement de la CAF
- Doc. 2 : Tableau — Extrait fictif de la base de données individuelles de la CAF Île-de-France sous forme d'un tableau élémentaire d'information
- Doc. 3 : Texte — La matrice d'information géographique d'après B. Berry, 1964
- Doc. 4 : Figure — La construction des objets géographiques, *in* Pumain, Saint-Julien, 2010, p.46
- Doc. 5 : Cartes – Densité des foyers allocataires à bas revenus à Paris selon trois maillages.
- Doc. 6 : Texte — L'approche spatiale des données d'après M. Goodchild et D. Janelle, 2004
- Doc. 7 : Figure — Typologie des matrices de relations spatiales, *in* Grasland, 1999
- Doc. 8 : Texte — Les enjeux des données individuelles selon R. Silberman, 2011.
- Doc. 9 : Figure — Un cas théorique d'application du secret statistique

### *Questions :*

**1. Le formulaire de déclaration de situation de la CAF (Doc. 1) est composé d'une série de questions sur l'identité du demandeur, sa situation familiale et ses revenus, afin de déterminer l'éligibilité du demandeur à une prestation sociale et son montant.**

- a. Vous souhaitez effectuer une demande d'aide auprès de la CAF de votre département. Après avoir lu le formulaire, identifiez :
- les **individus** décrits par le formulaire
  - leurs principaux **attributs**
  - les **modalités** des attributs

b. Quelles informations peuvent servir à l'analyse de la distribution spatiale des allocataires de la CAF à bas revenus à Paris ? Comment présenteriez-vous ces informations sous la forme d'un tableau élémentaire ?

### **2. À partir des documents 2 à 4 :**

a. Rappelez ce que sont les trois dimensions d'une matrice d'information géographique, en vous appuyant sur la lecture de l'extrait de l'article de Brian Berry (Doc. 3).

b. Identifiez les **unités spatiales** élémentaires renseignées dans la base CAF, en vous référant au Doc. 4. Quelles autres unités spatiales pertinentes pourriez-vous proposer pour conduire des analyses géographiques sur les bas revenus à Paris ?

c. Imaginez la structure possible d'une **matrice d'information géographique** construite à partir de ces données. À quelles questions géographiques une telle matrice permettrait-elle de répondre ? N'hésitez pas à repartir des différents sens de lecture de la matrice proposés par B. Berry (en ligne, en colonne...).

**3. À partir du Doc. 5 et de la lecture du Doc. 6, répondez aux questions suivantes :**

- a. Quel est l'intérêt d'une représentation cartographique de la matrice d'information géographique ? Qu'observez-vous sur les cartes du Doc. 5 que la matrice d'information géographique ne permettait pas d'observer ?
- b. Quel effet a le choix de la maille d'agrégation sur la représentation cartographique ? Quelle maille vous semble la plus adaptée, pourquoi ?

**4. À partir des Doc. 2, 3 et 7, répondez aux questions suivantes :**

- a. Quelles sont les limites de la matrice d'information géographique soulignées par B. Berry ? Sachant que ces déclarations annuelles peuvent décrire l'évolution de la situation d'un même individu dans le temps (grâce à l'identifiant individuel), quelle information du tableau élémentaire permettrait d'étudier une forme de relations entre unités spatiales ?
- b. À partir du Doc. 7, proposez une autre mise en forme de l'information géographique permettant d'en rendre compte. De quel type de matrice s'agit-il ?

**5. À partir des Doc. 8 et 9, répondez aux questions suivantes :**

- a. Identifiez les principaux enjeux de la production, de la manipulation et de l'analyse des données individuelles. En quoi l'utilisation des données de la CAF s'inscrit-elle dans ces enjeux ?
- b. Si vous deviez appliquer les principes du secret statistique au Doc. 9, quelles mailles ne pourraient pas être représentées ? Quel maillage permet de conserver le plus d'informations ?

## Dossier documentaire :

### Doc. 1 : Formulaire — Extraits de la Déclaration de situation pour les prestations familiales et les aides au logement de la CAF et de la Déclaration de ressources pour l'année 2013



santé  
famille  
retraite  
services

## Déclaration de situation pour les prestations familiales et les aides au logement



11423\*06

Livres V et VIII du code de la Sécurité sociale  
Livres III du code de la construction et de l'habitation

Une seule déclaration de situation suffit, même pour plusieurs demandes de prestations.  
Répondez à toutes les questions qui vous concernent, joignez toutes les pièces justificatives demandées, n'oubliez pas de dater et signer votre déclaration de situation.

#### ► Allocataire

Personne qui demande les prestations familiales et qui est responsable du dossier

Monsieur  Madame

Nom de famille (de naissance) : \_\_\_\_\_

Nom d'usage : \_\_\_\_\_  
(facultatif et s'il y a lieu)

Prénoms (dans l'ordre de l'état civil) : \_\_\_\_\_

Date de naissance :

Lieu de naissance : ville \_\_\_\_\_ département

Ou pays en cas de naissance à l'étranger : \_\_\_\_\_

Nationalité :

Française  UE, EEE\* ou Suisse  autre

Numéro de Sécurité sociale :

Date d'entrée en France si vous résidiez à l'étranger :

\* Cf liste des pays en page 4

Si vous êtes inscrit à la Caf/MSA de votre département ou avez été inscrit, au cours des 2 dernières années, en France ou à l'étranger, auprès d'un organisme versant des prestations ou avantages familiaux, précisez :

Nom de l'organisme : \_\_\_\_\_

Pays : \_\_\_\_\_

Numéro d'allocataire ou numéro de dossier : \_\_\_\_\_

#### ► Conjoint(e), concubin(e) ou pacsé(e)

Monsieur  Madame

Nom de famille (de naissance) : \_\_\_\_\_

Nom d'usage : \_\_\_\_\_  
(facultatif et s'il y a lieu)

Prénoms (dans l'ordre de l'état civil) : \_\_\_\_\_

Date de naissance :

Lieu de naissance : ville \_\_\_\_\_ département

Ou pays en cas de naissance à l'étranger : \_\_\_\_\_

Nationalité :

Française  UE, EEE\* ou Suisse  autre

Numéro de Sécurité sociale :

Date d'entrée en France s'il résidait à l'étranger :

\* Cf liste des pays en page 4

S'il est inscrit à la Caf/MSA de son département ou a été inscrit, au cours des 2 dernières années, en France ou à l'étranger, auprès d'un organisme versant des prestations ou avantages familiaux, précisez :

Nom de l'organisme : \_\_\_\_\_

Pays : \_\_\_\_\_

Numéro d'allocataire ou numéro de dossier : \_\_\_\_\_

#### ► Votre adresse

Votre adresse complète : \_\_\_\_\_

Code postal :     Commune : \_\_\_\_\_ Pays : \_\_\_\_\_

Numéros de téléphone - Domicile :       Autre (travail ou portable) :

Adresse mél : \_\_\_\_\_ @ \_\_\_\_\_

Depuis quelle date résidez-vous à cette adresse ? :

Combien de personnes vivent à votre domicile (y compris vous et votre conjoint) ? : \_\_\_\_\_

• Si vous vivez en couple et que votre conjoint, concubin ou pacsé réside à l'étranger, précisez :

le pays : \_\_\_\_\_ depuis quelle date :

⚠ Pour bénéficier des prestations familiales, vous et vos enfants devez résider habituellement en France, sauf si votre situation relève des règlements communautaires ou d'accords internationaux.

S 7103 j - 10/2013

#### Emplacement réservé

Date de la demande

DSIT

Page 1/4

IDX W 1005004 U



## 2 Déclaration de situation pour les prestations familiales et les aides au logement

### ► Situation familiale

#### Vous vivez en couple

- Vous vivez en couple sans être marié ni pacsé depuis le .....
- Vous êtes marié depuis le .....
- Vous êtes pacsé depuis le .....
- Vous avez repris la vie commune depuis le .....

#### Vous vivez seul(e)

- Vous êtes célibataire
- Vous êtes divorcé depuis le .....
- Vous êtes séparé légalement (séparation sans divorce avec intervention du juge) depuis le .....
- Vous êtes séparé sans intervention du juge ou avez rompu votre pacs depuis le .....
- Vous êtes veuf ou veuve depuis le .....

#### Parents séparés

• Si l'autre parent d'un de vos enfants est dans une ou plusieurs des situations suivantes, cochez la(les) case(s) correspondante(s) :

- il réside dans un des Etats de l'UE, EEE\* (hors France) ou en Suisse
- il travaille pour un employeur dont le siège est situé dans l'un de ces pays
- il perçoit une pension (retraite, invalidité), ou une indemnisation (maladie, chômage) de l'un de ces pays
- Une pension alimentaire est-elle fixée et/ou versée pour vos enfants par l'autre parent ?  oui  non

\* Cf liste des pays en page 4

### ► Enfants à votre charge et autres personnes de votre foyer

Si besoin, vous pouvez déclarer d'autres enfants ou personnes sur papier libre à joindre à ce formulaire.

Nom et prénoms (dans l'ordre de l'état civil)	Date et lieu ou pays de naissance	Date d'arrivée au foyer et lien de parenté (Fille, fils, nièce, neveu, enfant recueilli, parent, aucun lien...)	Situation actuelle (Scolarité, apprentissage, activité professionnelle...*)	Enfant résidant à l'étranger	Enfant en résidence alternée**
1. ....	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. ....	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. ....	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. ....	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. ....	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>

\* Si l'enfant est salarié ou apprenti, précisez son dernier salaire net mensuel et la date de début de son activité, et date de fin de contrat si apprenti.

\*\* En cas de séparation et de résidence alternée d'un ou plusieurs de vos enfants entre votre domicile et celui de l'autre parent, vous pouvez demander le partage des allocations familiales. Téléchargez le formulaire «Enfant(s) en résidence alternée - Déclaration et choix des parents» sur [caf.fr](http://caf.fr) ou [msa.fr](http://msa.fr) ou demandez-le à votre Caf ou MSA.

### ► Si vous êtes âgé(e) de moins de 25 ans ou si votre conjoint(e), concubin(e) ou pacsé(e) est âgé(e) de moins de 25 ans

⚠ Si vous demandez des prestations, vos parents et/ou ceux de votre conjoint ne pourront plus bénéficier des prestations qu'ils perçoivent pour vous en tant qu'enfant à leur charge.

	■ Parents de l'allocataire	■ Parents du conjoint, concubin ou pacsé
L'un de vos parents perçoit-il pour vous des prestations familiales, Apl, Rsa... ?	<input type="checkbox"/> oui <input type="checkbox"/> non	<input type="checkbox"/> oui <input type="checkbox"/> non
Si <b>oui</b> , Nom du parent allocataire :	.....	.....
Prénoms :	.....	.....
Adresse :	.....	.....
Code postal et commune :	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Organisme de prestations familiales :	.....	.....
Numéro d'allocataire (s'il en possède un) :	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Numéro de sécurité sociale :	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>

#### Emplacement réservé



### 3 Déclaration de situation pour les prestations familiales et les aides au logement

#### ► Situation(s) professionnelle(s) actuelle(s)

- |   | ■ Allocataire  | ■ Conjoint(e), concubin(e) ou pacsé(e)                     |
|---|--|--|
| ■ Salarié (y compris en préretraite progressive)..... | <input type="checkbox"/> depuis le _____                   | <input type="checkbox"/> depuis le _____                   |
| Nature du contrat (Cdd, Cdi, intérim, etc.) .....     | _____  | _____  |
| ■ Apprenti .....                                      | <input type="checkbox"/> depuis le _____<br>fin le : _____ | <input type="checkbox"/> depuis le _____<br>fin le : _____ |
| ■ Stagiaire de la formation professionnelle .....     | <input type="checkbox"/> depuis le _____                   | <input type="checkbox"/> depuis le _____                   |
| ■ Travailleur indépendant ou employeur .....          | <input type="checkbox"/> depuis le _____                   | <input type="checkbox"/> depuis le _____                   |
| ■ Non salarié agricole .....                          | <input type="checkbox"/> depuis le _____                   | <input type="checkbox"/> depuis le _____                   |
| ■ Auto-entrepreneur .....                             | <input type="checkbox"/> depuis le _____                   | <input type="checkbox"/> depuis le _____                   |
| ■ Conjoint collaborateur .....                        | <input type="checkbox"/> depuis le _____                   | <input type="checkbox"/> depuis le _____                   |

#### Pour toutes ces situations, précisez :

S'il y a lieu, le nom de l'employeur ou de l'organisme

de formation et son adresse : .....

Vous ou votre employeur cotisez **en France** :

à l'Urssaf  à la MSA (régime agricole)

à l'Urssaf  à la MSA (régime agricole)

autre régime, lequel ? .....

autre régime, lequel ? .....

à l'étranger :

précisez le pays : .....

précisez le pays : .....

■ Chômeur (indemnisé ou non).....  depuis le \_\_\_\_\_  depuis le \_\_\_\_\_

Si indemnisé, précisez l'organisme .....

S'agit-il de chômage partiel ? .....

oui  non

oui  non

■ Retraité, pensionné (y compris en préretraite totale) .....

depuis le \_\_\_\_\_

depuis le \_\_\_\_\_

De quel(s) régime(s) percevez-vous une pension ?

■ En maladie .....

depuis le \_\_\_\_\_

depuis le \_\_\_\_\_

■ Autre cas (congé maternité, congé parental, hospitalisation, détention sauf régime de semi-liberté, longue maladie, etc.) : .....

Précisez la situation : .....

Précisez la situation : .....

Nom et adresse de l'établissement en cas

d'hospitalisation ou de détention : .....

**Si pour une de ces situations, les indemnités** Nom de l'organisme et pays : Nom de l'organisme et pays :

ou pensions sont versées par un pays étranger :

\_\_\_\_\_

\_\_\_\_\_

■ Etudiant .....

depuis le \_\_\_\_\_

depuis le \_\_\_\_\_

■ Sans activité professionnelle .....

depuis le \_\_\_\_\_

depuis le \_\_\_\_\_

depuis toujours

depuis toujours

#### ► Déclaration sur l'honneur

Je certifie sur l'honneur l'exactitude de cette déclaration et des documents joints.

Fait à : \_\_\_\_\_ Le : \_\_\_\_\_

Si le signataire est un représentant de l'allocataire, précisez ci-dessous

ses nom, prénom, qualité et adresse :

\_\_\_\_\_

Signature de l'allocataire ou de son représentant

**Vous avez l'obligation de signaler immédiatement tout changement modifiant cette déclaration.**

La Caf/MSA vérifie l'exactitude des déclarations (Article L.114-19 du code de la Sécurité sociale). La loi punit quiconque se rend coupable de fraude ou de fausse déclaration (Articles L.114-9 - dépôt de plainte de la Caisse pouvant aboutir à : travail d'intérêt général, amende ou peine de prison, L. 114-13 - amende, L. 114-17 du code de la Sécurité sociale - prononcé de pénalités).

La loi 78-17 du 06/01/1978 modifiée relative à l'informatique, aux fichiers et aux libertés s'applique aux réponses faites sur ce formulaire. Elle garantit un droit d'accès et de rectification pour les données vous concernant auprès de l'organisme qui a traité votre demande.

Emplacement réservé



# Ressources de l'année 2013

**Déclarez SANS LES CENTIMES tous les revenus perçus en France, hors de France ou versés par une organisation internationale** (articles R. 532-3, R. 831-6 et D. 542-10 du code de la Sécurité sociale et R. 351-5 du code de la Construction et de l'habitation).

Numéro d'allocataire : \_\_\_\_\_

N° de Sécurité sociale : \_\_\_\_\_  
(de l'allocataire)

Nom, prénom : \_\_\_\_\_

Adresse : \_\_\_\_\_

1 PERSONNES DONT LES RESSOURCES SONT À DÉCLARER	VOUS	VOTRE CONJOINT, CONCUBIN OU PACSÉ	ENFANT OU AUTRE PERSONNE
Nom	_____	_____	_____
Prénom	_____	_____	_____
Date de naissance	_____	_____	_____
<b>ABSENCE DE RESSOURCES EN 2013</b> cochez la case	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>2 SALAIRES</b> • traitements, salaires, heures supplémentaires et indemnités journalières de Sécurité sociale (maladie, maternité, paternité, fraction imposable d'accident du travail ou de maladie professionnelle) • frais réels déductibles	_____ € _____ €	_____ € _____ €	_____ € _____ €
<b>3 INDEMNITÉS JOURNALIÈRES D'ACCIDENT DU TRAVAIL OU DE MALADIE PROFESSIONNELLE</b> (fraction non imposable)	_____ €	_____ €	_____ €
<b>4 ALLOCATIONS DE CHÔMAGE ET PRÉRETRAITES</b>	_____ €	_____ €	_____ €
<b>5 REVENUS DES NON SALARIÉS (Bic - BNC - BA - MICRO BIC...)</b> • adhérent d'un centre de gestion agréé, ou "régime micro" ou <b>auto-entrepreneur</b> • non adhérent d'un centre de gestion agréé • forfait non fixé : cochez la case	_____ € _____ € <input type="checkbox"/>	_____ € _____ € <input type="checkbox"/>	_____ € _____ € <input type="checkbox"/>
<b>6 DÉFICITS DE L'ANNÉE 2013</b> • professionnels • fonciers	_____ € _____ €	_____ € _____ €	_____ € _____ €
<b>7 RETRAITES, PENSIONS ET RENTES IMPOSABLES</b> (les Allocations supplémentaires vieillesse ou invalidité et l'Allocation de solidarité aux personnes âgées ne sont pas à déclarer)	_____ €	_____ €	_____ €
<b>8 PENSIONS ALIMENTAIRES REÇUES</b>	_____ €	_____ €	_____ €
<b>9 AUTRES REVENUS</b> • revenus fonciers • contrat d'épargne-handicap • autres	_____ € _____ € _____ €	_____ € _____ € _____ €	_____ € _____ € _____ €
<b>10 CHARGES DÉDUCTIBLES</b> • pensions alimentaires versées • CSG déductible sur les revenus du patrimoine • épargne retraite et cotisations volontaires de Sécurité sociale	_____ € _____ € _____ €	_____ € _____ € _____ €	_____ € _____ € _____ €

Je certifie sur l'honneur l'exactitude de cette déclaration. Je prends connaissance que ma Caisse peut vérifier les montants déclarés. Je m'engage à signaler immédiatement à ma Caisse tout changement intervenant dans ma situation.

Signature obligatoire

Le : \_\_\_\_\_

La Caf vérifie l'exactitude des déclarations (Article L. 114-19 du code de la Sécurité sociale). La loi punit quiconque se rend coupable de fraude ou de fausse déclaration (Articles L. 114-9 - dépôt de plainte de la Caisse pouvant aboutir à : travail d'intérêt général, amende ou peine de prison, L. 114-17 du code de la Sécurité sociale - prononcé de pénalités -, articles 313-1 à 313-3, 441-1 et 441-6 du code Pénal).

La loi 78-17 du 06/01/1978 modifiée relative à l'informatique, aux fichiers et aux libertés s'applique aux réponses faites sur ce formulaire. Elle garantit un droit d'accès et de rectification pour les données vous concernant auprès de l'organisme qui a traité votre demande.

S 7123 i

Emplacement réservé

DR

Page 1/1

IDX P 3201015 Z

Doc. 2 : Table – Extrait fictif de la base de données individuelles de la CAF Île-de-France sous forme d'un tableau élémentaire d'information

Matricule	NIR	Date de naissance allocataire	Situation familiale	Enfants	Autres personnes à charge	Situation pro. allocataire	Situation pro.conjoint	RUC (euros)	Bas revenus	Adresse	Ancienne adresse
10001	1598094337404	2/6/1959	pacse	1	0	chomeur	chomeur	1390	0	44 r. de Lille 75007	9 bld. de Magenta 75010
10002	2750319174145	3/3/1975	marie	1	0	salarie	salarie	2231	0	18 r. Clisson 75013	20 av. du bel-air 75012
10003	1710885487946	11/8/1971	marie	0	0	salarie	chomeur	1008	1	104 r. Balard 75015	79 r. du Mont Cenis 75018
10004	2620562514194	17/5/1962	separe	2	1	autre	/	1037	1	1 r. Lepage 75016	31 r. Charcot 75013
10005	1570432134846	17/4/1957	veuf	0	0	salarie	/	1453	0	22 r. H. Moreau 75018	65 r. Boursault 75017
10006	1500334599522	12/3/1950	couple	0	0	retraite	salarie	/	/	110 bld. Exelmans 75016	9 r. de Trévise 75009
10007	1830883533690	18/8/1983	celibataire	3	0	auto-entrepreneur	/	2712	0	83 q. de la gare 75013	90 r. de la Jonquière 75017
10008	1881070525240	20/10/1988	celibataire	1	0	etudiant	/	/	/	1 r. Charles Friedel 75020	70 bld. de Rochechouart 75018
10009	2810879360589	12/8/1981	divorce	3	0	chomeur	/	922	1	5 r. Michal 75013	24 r. le Sueur 75016
10010	2540595198563	16/5/1954	marie	0	1	salarie	salarie	1843	0	80 r. de l'Amiral Roussin 75015	9 r. claudie Tillier 75012
10011	2461276302204	8/12/1946	couple	0	0	retraite	retraite	/	/	12 sq. de Clignancourt 75018	65 r. de Rivoli 75001
10012	1920832243105	1/8/1992	pacse	0	0	auto-entrepreneur	salarie	1598	0	116 r. des Pyrénées 75020	34 r. de la Duée 75020
10013	2921273150251	29/12/1992	couple	1	0	apprenti	stagiaire	820	1	42 r. d'Auteuil 75016	28 av. Wilson 75016
10014	2740395425339	29/3/1974	couple	1	0	maladie	salarie	1037	1	60 r. Letort 75018	6 r. Alibert 75010
10015	2930243209619	31/2/1993	celibataire	0	0	salarie	/	1453	0	11 r. Serpente 75006	7 r. Christiani 75018
10016	1471063490341	26/10/1947	marie	0	0	retraite	retraite	/	/	75 r. Vieille du Temple 75003	98 r. du chemin vert 75011
10017	2830650807705	16/6/1983	celibataire	0	0	sans-activite	auto-entrepreneur	1175	0	1 r. Joseph Bara 75006	22 r. Visconti 75006
10018	1900456403554	21/4/1990	pacse	1	0	autre	chomeur	1959	0	24 r. de Constantinople 75008	9 r. J. Hachette 75015
10019	1650930615432	20/9/1965	divorce	1	0	salarie	/	1822	0	95 r. de Rome 75017	45 r. Vineuse 75016
10020	1560517619158	5/5/1956	marie	1	0	retraite	autre	962	1	400 r. Saint-Honoré 75001	53 r. de Picpus 75012
10021	1910194285761	26/1/1991	celibataire	0	0	etudiant	/	/	/	22 r. de la Harpe 75005	23 av. René Fonck 75019
10022	1780678500508	28/6/1978	celibataire	4	0	independant	/	840	1	58 r. de Saussure 75017	36 r. de Richelieu 75001
10023	1520978573321	30/9/1952	marie	0	0	retraite	retraite	/	/	43 r. Charlot 75003	7 r. Liancourt 75014
10024	1771225251477	24/12/1977	separe	1	0	auto-entrepreneur	/	1114	0	19 r. de Tolbiac 75013	15 r. de la chapelle 75018
10025	1891089392469	19/10/1989	celibataire	0	0	salarie	/	1660	0	111 r. d'Alecia 75014	13 r. du plateau 75019

### Doc. 3 : Texte – La matrice d'information géographique d'après B. Berry (1964)

Réfléchissons un instant à la nature d'une observation élémentaire, exploitée dans une perspective géographique. Cette observation décrit un attribut particulier d'un lieu donné et peut être qualifiée de « fait géographique ». En général, ce fait géographique s'inscrit plutôt dans un ensemble d'observations, qui décrivent soit les caractéristiques d'un même attribut pour une série de lieux, soit une série d'attributs combinés en un même lieu. Ces deux séries méritent d'être examinées plus attentivement. Si les caractéristiques d'un même attribut sont enregistrées pour une série de lieux, on fera porter l'analyse sur les variations spatiales du phénomène, d'un lieu à l'autre. Ces variations seront probablement cartographiées, car de même que les données statistiques peuvent être réarrangées sous la forme de résumés de distributions statistiques, les géographes aiment réarranger leurs données sous la forme de distributions spatiales. L'étude des organisations spatiales qui ressortent de telles cartes est l'un des objets essentiels de la géographie. Si des attributs variés sont renseignés pour un même lieu, ils constitueront la base d'inventaires locaux et de l'étude géographique de lieux singuliers. C'est à l'appui de tels inventaires que les géographes étudient la manière dont différents phénomènes se combinent et interagissent en un même lieu.

Supposons maintenant que toute une série d'attributs ait été enregistrée pour toute une série de lieux. Nous pourrions très bien imaginer l'existence de tels « fichiers géographiques » exhaustifs du monde (même si ce rêve pourrait devenir un réel cauchemar, mais c'est un autre problème). Une manière efficace de structurer cette information consiste à construire un tableau (ou une matrice) rectangulaire. A quoi ressemble une telle « matrice géographique » ? Chaque caractéristique correspond à une colonne, et chaque lieu à une ligne (...)<sup>2</sup>. L'intersection de n'importe quelle colonne avec n'importe quelle ligne définit une cellule, et chaque cellule correspond à un fait géographique, pour tel attribut et tel lieu (...). Si l'on considère une matrice géographique telle que celle qui vient d'être décrite, combien d'approches d'analyse régionale sont-elles envisageables ? On peut considérer :

- (1) L'arrangement des cellules au sein de toute ou partie d'une colonne ; ou bien
- (2) L'arrangement des cellules au sein de toute ou partie d'une ligne.

Le premier sens de lecture conduit à l'étude de distributions spatiales et de cartes. Le deuxième conduit à l'étude de l'association de variables en un même lieu et à celle d'inventaires localisés. On conviendra sans hésiter que ces deux approches sont au fondement de toute géographie (...).

La définition d'un fait géographique que nous avons exposée jusqu'à présent est incomplète, car un certain attribut observé en un lieu donné est valable pour un instant particulier. A n'importe quel autre instant, il pourrait être différent ; la variation est tout autant temporelle que spatiale. Le temps pourrait aussi être subdivisé à l'infini, mais il est plus pratique d'imaginer que la matrice géographique est complétée par une troisième dimension (...), arrangée en une série de coupes ou de « tranches » temporelles extraites d'un temps continu, de la même manière que les colonnes étaient l'extraction d'une infinité de caractéristiques possibles et que les lignes étaient sélectionnées parmi un nombre infini de lieux. Chaque tranche résume ou saisit donc les variations d'attributs d'un lieu à un autre et pour une certaine période de temps (...).

Le modèle que nous venons de développer embrasse la plupart des approches imaginées et mises en œuvre par les géographes, mais ne les englobe pas toutes ; ce modèle est en soi limité. Il y a en effet d'importantes questions que la matrice que nous venons de présenter ne permet pas de traiter.

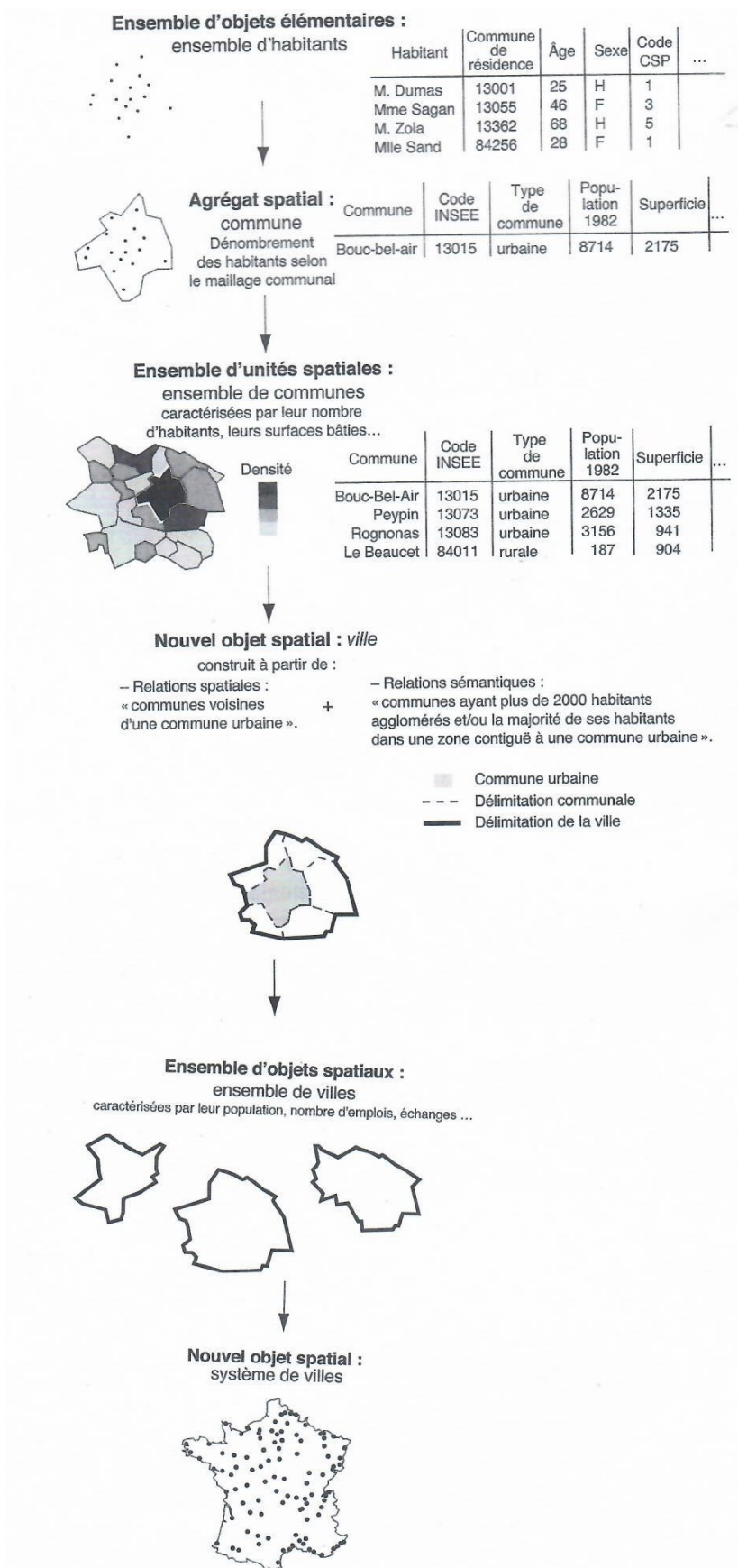
---

2 Dans le texte original de Brian Berry, la présentation de la matrice est inversée (attributs en ligne, lieux en colonne) par rapport aux matrices que les géographes ont l'habitude d'utiliser. Dans ce document, on a modifié le texte original et repris la présentation usuelle du tableau d'information géographique (lieux en ligne, attributs en colonne) pour éviter toute ambiguïté.

Cette discussion a été initialement développée dans le cadre d'une réflexion sur la Théorie Générale des Systèmes. Cette théorie nous dit que tout système, y compris « l'écosystème mondial dont les hommes sont une composante dominante », peut être analysé à partir de différents niveaux, les trois premiers niveaux correspondant aux structures statiques, à la connectivité des parties du système (organisation fonctionnelle), et aux processus dynamiques (...). On peut considérer que la partie du système qui a le plus intéressé les géographes jusqu'à présent correspond au premier de ces trois niveaux, celui des structures statiques — configurations analysées dans l'espace et dans le temps. Cela ne dit rien du deuxième niveau, celui des interconnexions, des flux et des interactions entre les lieux, et cela laisse aussi de côté le troisième niveau, celui de la compréhension de processus dynamiques et interdépendants.

*D'après Berry B.J.L., 1964, « Approaches to Regional Analysis: A synthesis », Annals of the Association of American Geographers, 54 (1), p.2-11.*

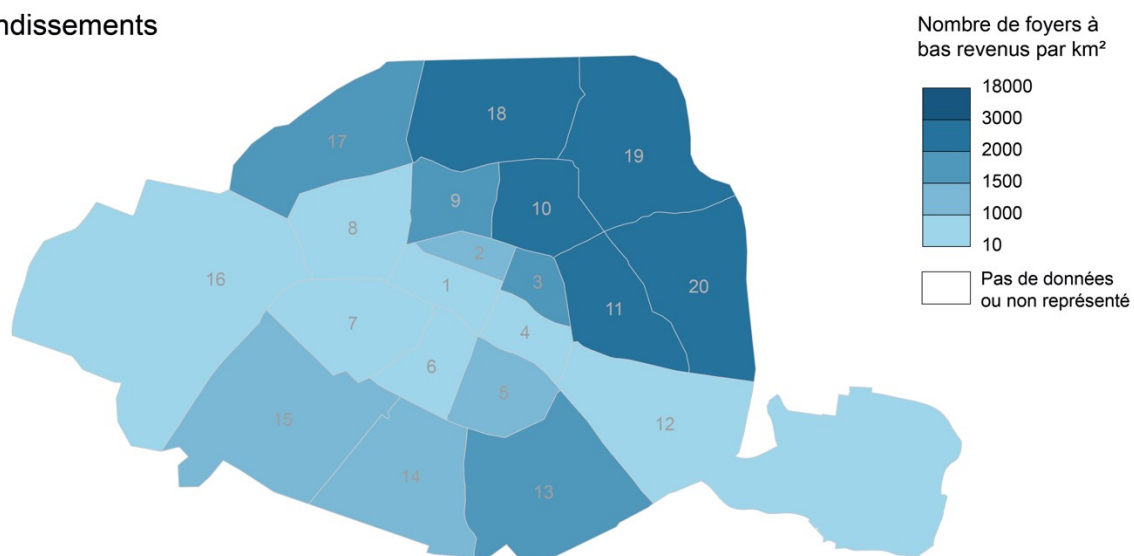
Doc. 4 : Figure – La construction des objets géographiques, *in* Pumain, Saint-Julien, 2010, p.46





## Doc. 5 : Cartes – Densité des foyers allocataires à bas revenus à Paris selon trois maillages.

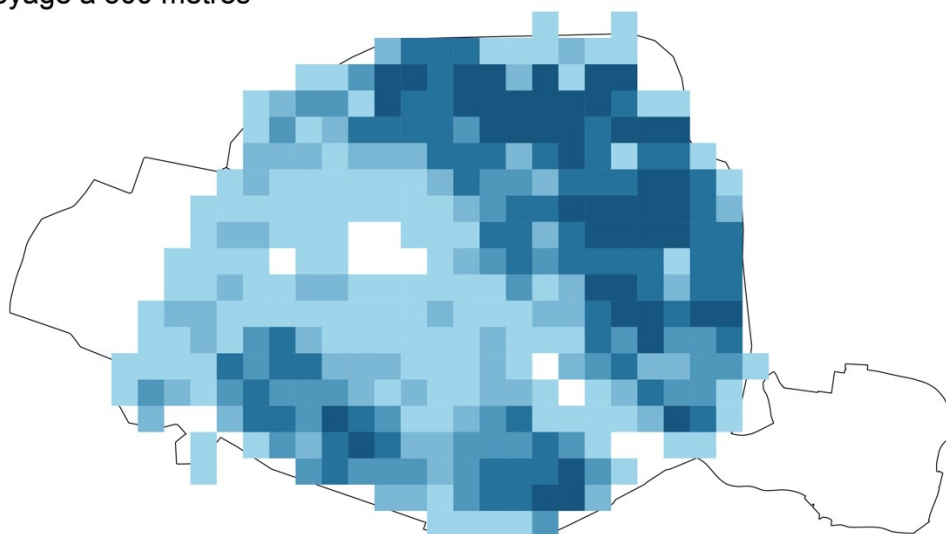
### Arrondissements



### IRIS



### Carroyage à 500 mètres



Données : Caisses d'allocations familiales d'Ile-de-France (2016)

Auteurs: Luc Guibard (Univ. Paris Diderot), Romain Leconte (Univ. Paris Diderot) et Bruno Fayard (CTRAD), 2018

## La pensée spatiale en sciences sociales

Les tableaux de dénombrement ou statistiques sont un moyen si commun d'organiser l'information que nous ne pensons rarement voire jamais à leur signification. Les tableaux sont utilisés pour organiser l'information issue des recensements décennaux de chaque État<sup>3</sup>, pour rapporter la performance comparée des économies nationales, pour résumer la performance des écoles d'un État, et pour un ensemble d'autres objectifs. Dans la plupart des cas, une simple convention est appliquée dans la préparation du tableau (Nicol et Pexman, 1999). Les lignes du tableau représentent les entités, les objets ou les cas dont on reporte les attributs, et tous les objets appartiennent à un type homogène, comme les comtés, les États ou les pays. Chaque colonne du tableau représente un des attributs, permettant au lecteur de faire des comparaisons en lisant une colonne de bas en haut ou de haut en bas, ou en opérant des calculs statistiques comme les moyennes des colonnes. Toutes les entités citées ici s'avèrent être géographiques, dans le sens où elles occupent des localisations distinctes à la surface de la Terre ; dans le cas du recensement ces objets vont des régions aux États, comtés, municipalités, mailles statistiques ou aux quartiers.

Les tableaux permettent d'organiser l'information pour faciliter la comparaison et pour l'analyse avec des tableurs ou des logiciels de statistique, mais ils ignorent largement une propriété spécifique de chaque objet : sa localisation géographique. Un tableau des États des États-Unis trié par ordre alphabétique place, par exemple, l'Alabama en haut et le Wyoming en bas, et dans seulement deux cas (Floride et Géorgie, Illinois et Indiana), les États qui sont géographiquement contigus apparaissent côte à côte dans des lignes du tableau. En effet, les tableaux détruisent toute l'importance qu'on peut accorder à la localisation géographique : est-ce que les États qui sont géographiquement proches les uns des autres ont tendance à partager les mêmes propriétés, et est-ce que ces observations permettent de mieux éclairer les processus à l'œuvre dans le paysage ; y a-t-il des entités dont les attributs sont radicalement divergents de ceux des États voisins ? [...] Clairement, il est impossible de trier un tableau des États dans un ordre qui reflète complètement la proximité géographique relative ou le contexte géographique d'un État (Abel et Mark, 1990).

A un autre niveau, on peut se demander si les États sont les unités appropriées pour rassembler des statistiques. Bien sûr, on peut utiliser des unités plus grandes comme des agrégations d'États comme les régions ou les pays, ou des unités plus petites comme les comtés, qui eux-mêmes s'agrègent en États, mais la question plus générale est de savoir si aucune de ces unités standardisées n'est optimale, et si une manière un peu différente de découper les États-Unis ne serait pas plus éclairante. Par exemple, les régions métropolitaines dépassent souvent les limites des États qui ont été dessinées presque arbitrairement il y a bien longtemps, mais elles sont toujours moins arbitraires que les limites issues de l'agrégation de leurs États respectifs (par exemple, les régions métropolitaines de New-York ou de Kansas City). Les États varient aussi grandement en taille, de plus de 1,5 millions de km<sup>2</sup> — pour l'Alaska à 3 144 km<sup>2</sup> — pour Rhode Island, ruinant tout effort qui chercherait à mener une analyse à l'échelle des États des processus opérant dans le paysage sur des échelles de distances particulières. Peut-être l'espace géographique doit être considéré comme un espace continu sur lequel on a imposé des limites de mailles statistiques quelques peu arbitraires. [...]

## L'analyse spatiale

Le terme d'analyse spatiale ou d'analyse des données spatiales suggère une insistance sur la localisation dans le déroulé de l'analyse, ou plus formellement, suggère que les résultats des

3 Il s'agit ici des États fédérés américains.



analyses utilisant une des techniques de ce type sont en partie dépendants de la localisation des objets qui sont analysés — si les localisations changent, les résultats aussi. L'analyse spatiale traite les données de manière *transversale*, par opposition aux analyses *longitudinales* ou aux analyses de séries temporelles. Les recensements décennaux sont un exemple marquant de données transversales, parce que chaque recensement fournit une image d'une population d'un pays à un temps *t*. La succession des recensements peut, en principe, être assemblée pour fournir une série longitudinale, mais les longs intervalles entre les dates de recensement et les changements de définitions des variables ou des mailles statistiques en font un défi. [...]

Les techniques d'analyse spatiale répondent à la question qui ouvre ce chapitre : la perte du contexte spatial quand des informations transversales sont présentées sous la forme de tableau. Supposons, par exemple, que cette même information sur les performances économiques des pays, ou le pourcentage de propriétaires par État étaient présentés sous la forme d'une carte plutôt que d'un tableau. Dans cette forme, il serait facile de comparer la part de propriétaires de chaque État à celle de chacun des quartiers, d'identifier les anomalies quand les États ont des niveaux très dissemblables de ceux de leurs voisins, et de comparer les régularités observées sur la carte avec la connaissance préalable que l'on a des caractéristiques de chaque État et de ses voisins. Fondamentalement, la création d'une carte est un moyen simple et intuitif d'analyse spatiale, ou peut-être plus exactement, un moyen d'armer le cerveau et l'œil humain pour faire de l'analyse spatiale. Comme toute forme de représentation de l'information, elle est sujette à manipulation par le cartographe, par des choix mal appropriés de couleurs ou de discrétisation (voir Monmonnier, 1991) ; et par la prévalence accordée à la précision planimétrique, les cartes tendent à sur-représenter les grands États et peuvent rendre les petits États presque invisibles (ainsi, les cartogrammes fournissent un moyen intéressant d'ajustement de la représentation visuelle (Dorling, 1995)).

*D'après Goodchild M. F., Janelle D. G., 2004, Spatially Integrated Social Science, Oxford University Press, pp. 3-6*



**Doc. 8 : Texte — Les enjeux des données personnelles selon R. Silberman (2011)**

La recherche en sciences sociales utilise très largement des données individuelles que ce soit des données sur les personnes ou sur des entités comme des entreprises ou des collectivités territoriales. Qu'il s'agisse d'autoriser le recueil de ces données ou de les utiliser pour des traitements statistiques, leur protection est devenue avec le développement de l'informatique une préoccupation centrale pour les citoyens.

La question centrale est en réalité celle de l'identification possible, dans des bases de données, de personnes physiques ou morales, identification qui pourrait de ce fait révéler certaines de leurs caractéristiques et comportements. [...] Le nom, qu'il s'agisse de celui de la personne physique, de l'entreprise ou de toute autre personne morale, permet bien entendu l'identification encore que les cas d'homonymie puissent laisser une zone d'incertitude. L'absence dans un fichier du nom, mais également de l'adresse, le plus souvent considérée comme un élément d'identification directe car singulier (même s'il s'agit de l'adresse d'un immeuble), ne suffit cependant pas à considérer qu'un tel fichier est anonymisé. La notion d'*indirectement nominatif* est désormais au cœur du débat sur l'identification possible des personnes dans une base de données. C'est la notion qui figure désormais dans la loi française dite « Informatique et Libertés » comme dans la directive européenne de 1995 en la matière. Si le terme est plutôt utilisé à propos de l'identification des personnes physiques dans ces lois, la même notion prévaut pour les entreprises ou les personnes morales. Depuis les années 1990, on tend de plus en plus à considérer que le croisement de plusieurs variables peut permettre de ré-identifier des personnes physiques ou morales. Disposer par exemple de la date de naissance, de l'état matrimonial, du sexe et de la profession d'une personne avec une variable de localisation, même sur une zone relativement large, pourrait ainsi conduire à identifier une personne dans le cas par exemple où il s'agirait d'une profession rare : ainsi, un notaire divorcé et père de quatre enfants dans une ville de province.

L'idée initiale de ces régimes de protection porte en effet sur la question du préjudice : il s'agit de protéger des données qui, sauf lorsqu'elles sont déjà dans le domaine public, portent sur des faits, des actions, des opinions, etc., qui, s'ils étaient divulgués, porteraient préjudice aux intérêts ou à la réputation de la personne concernée. Dans le cas des personnes physiques, on parle de vie privée ou d'intimité. La présence de variables dites *sensibles* au sens de la loi conduit naturellement à encore plus de précautions. Dans le cas d'entreprises, il s'agit d'une atteinte à ce qu'on appelle le secret des affaires qui peut porter préjudice, s'il est rompu, aux intérêts de l'entreprise, qu'il s'agisse d'informations portées à la connaissance des services publics (le fisc notamment) ou d'autres entreprises (entorse au principe de la concurrence).

La loi française « Informatique et Libertés » distingue quatre catégories [de données sensibles] : l'origine, les opinions politiques et religieuses, la santé et la sexualité.

Les travaux des chercheurs recourant à l'analyse spatiale ou utilisateurs de par leur domaine de recherche (santé, immigration par exemple) de données sensibles ont été de leur côté particulièrement impactés par l'élévation des niveaux d'anonymisation qui est d'autant plus forte que sont en jeu des variables spatiales et des variables sensibles. [...] La diffusion du recensement français de 1999 a marqué un point d'inflexion sensible : le niveau géographique disponible dans les fichiers détails a ainsi considérablement augmenté (restriction de la diffusion de résultats du recensement de la population à des zones comportant au moins 2 000 habitants), rendant particulièrement difficiles les travaux des géographes mais aussi ceux de la sociologie urbaine sur les questions de ségrégation spatiale qui nécessitent l'accès à des données localisées fines pour construire des agrégats pertinents. [...] Poussée à son terme, la question du seuil d'anonymisation

va bien au-delà de ces deux types de variables particulièrement sensibles : dès lors que le croisement de plusieurs variables ferait apparaître des configurations singulières avec trop peu d'individus par case, le risque d'identification existerait.

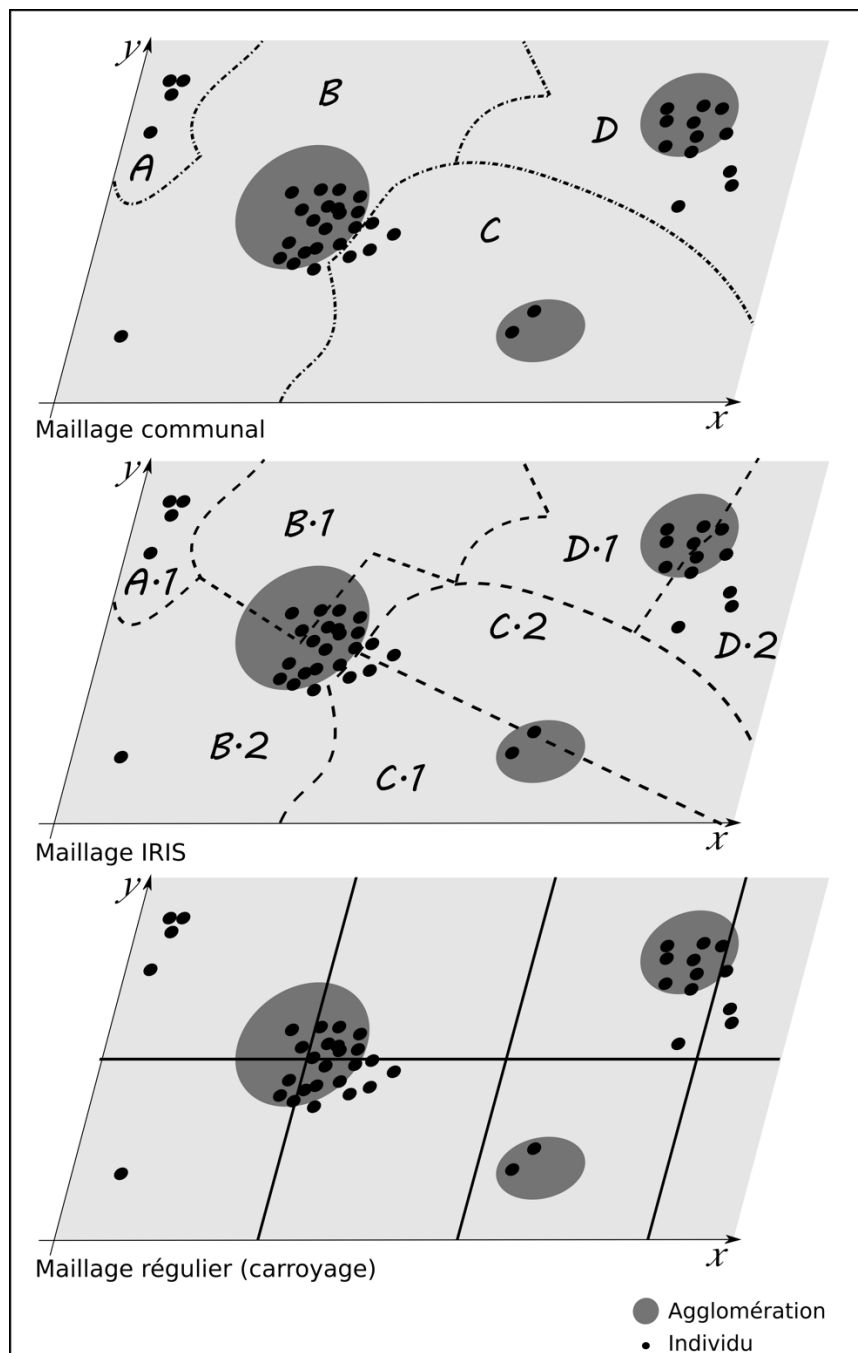
*Extraits de Silberman, R. (2011). Chapitre 12. La protection des données individuelles en France et la recherche en sciences sociales. Dans : Alain Chenu éd., La France dans les comparaisons internationales : guide d'accès aux grandes enquêtes statistiques en sciences sociales (pp. 183-204). Paris, Presses de Sciences Po.*

## Doc. 9 : Figure — Un cas théorique d'application du secret statistique

### Les principes du secret statistique

Conformément aux dispositions relatives au secret statistique et à la confidentialité contenues dans la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques, et de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, les données et la documentation diffusées :

- Ne doivent pas permettre d'identifier, directement ou indirectement, des ménages, des individus ou une personne par qui que ce soit. En tous les cas, les données mises à disposition ne doivent contenir aucun code, numéro de matricule ou de moyens de décodage possible.
- Doivent être complètement anonymisées. Elles ne comprennent pas d'informations nominatives, à caractère personnel ou de renseignements individuels ou sensibles.
- Aucun dénombrement inférieur à cinq n'est communiqué à des tiers.



## C. Proposition d'accompagnement pédagogique

### *Contextualisation*

Les caisses d'allocations familiales (CAF) constituent la Branche Famille du système de Sécurité Sociale français. Elles sont en charge des prestations familiales (allocations familiales, Prime à l'accueil du jeune enfant), de l'action sociale (définie selon les besoins départementaux) ainsi que du versement des prestations sociales concernant le logement (Aide Personnalisée au Logement, Allocation de Logement Sociale et Allocation de Logement Familiale) et de la lutte contre la précarité (le Revenu de Solidarité Active notamment). Pour bénéficier d'une de ces prestations, il faut transmettre à la CAF de son département un dossier de demande (en ligne ou au guichet). Celui-ci doit contenir toutes les informations qui permettront de déterminer les droits du demandeur ainsi que le montant des prestations (Doc. 1). Les renseignements collectés sont centralisés par les CAF départementales dans des bases de données à partir desquelles elles assurent le suivi des dossiers et établissent le bilan de leur activité (nombre de bénéficiaires de chaque prestation, coûts engagés, *etc.*).

Ces bases de données constituent une source d'information très riche pour l'analyse géographique. En effet, relativement à d'autres sources de données habituellement mobilisées — notamment celles proposées par l'Institut national de la statistique et des études économiques (Insee) à partir du recensement de la population — les données des allocataires de la CAF sont actualisées tous les mois et elles contiennent des renseignements précis sur les lieux de résidence (adresse), la situation familiale et les ressources des ménages (données de la Direction Générale des Finances Publiques). Elles permettent des analyses longitudinales, car il est possible de suivre les foyers allocataires dans le temps. Bien qu'elles ne soient pas représentatives de l'ensemble de la population française, elles couvrent près de la moitié de celle-ci (32 millions de personnes au 31 décembre 2018, soit 47,8 % de la population)<sup>4</sup>.

### *Question 1*

a. Pour formaliser l'information contenue dans les déclarations de situation, il faut identifier les individus décrits par le formulaire, leurs attributs et les modalités de ces attributs. Il s'agit de mettre en correspondance ces renseignements avec les notions élémentaires de l'analyse statistique en géographie.

En statistique, un individu est « un des objets constitutifs de l'ensemble observé » (Grasland, 1998). Chaque individu est unique et possède un identificateur propre (son nom, un code). Il peut également être désigné par sa position dans la population étudiée (le sixième dossier reçu ce mois par exemple). L'expression « individu » est donc à prendre au sens large. Il s'agit d'une unité statistique. Selon le sujet d'étude, celle-ci peut être une personne physique, un arbre, un article de journal, une commune...

Dans le Doc. 1, plusieurs personnes physiques sont décrites : le futur allocataire qui effectue la demande de prestation, son conjoint éventuel, ses enfants et toute autre personne à sa charge. Les caractéristiques de chacun et les relations qu'ils ont entre eux sont prises en compte pour déterminer les prestations auxquelles ils ont droit. Cependant, du point de vue

---

<sup>4</sup> D'après l'*open data* de la CAF, <http://data.caf.fr/dataset>., consulté le 23.09.2020.

de la CAF, toutes ces personnes constituent un seul dossier : le « foyer allocataire »<sup>5</sup> qui est l'**individu** statistique élémentaire. Chaque foyer allocataire, composé d'une ou plusieurs personnes, est représenté par une personne de référence, sur laquelle on aura le plus d'informations. L'ensemble des membres du foyer constitue la « population couverte » par les CAF (Chemineau et Bergeron, 2016). Il ne faut pas la confondre avec la population dans son sens statistique qui désigne l'ensemble des individus statistiques sur lesquels porte l'analyse.

Un **attribut** (aussi appelé variable ou caractère) permet de décrire un individu statistique. Ses modalités sont l'ensemble des valeurs qu'il peut prendre (sans que celles-ci soient nécessairement observées). Les **modalités** d'un attribut sont incompatibles (l'attribut n'a qu'une modalité par individu) et exhaustives (tout individu doit pouvoir avoir une modalité) (Grasland, 1998). Le tableau d'information géographique présenté dans le Doc. 2 reprend les principaux attributs qu'il est possible de construire à partir du formulaire. Ces derniers portent l'identité des membres du foyer (nom, âge), sur leur situation professionnelle et familiale ainsi que sur leurs ressources. Certains sont quantitatifs, les salaires par exemple, et d'autres catégoriels, comme la situation professionnelle.

b. La pauvreté peut être appréhendée de plusieurs manières. On peut ainsi distinguer les approches multidimensionnelles — qui considèrent que la pauvreté implique à la fois des aspects économiques, sociaux, physiques et psychologiques (Anand et Sen, 1997) — d'une approche fondée uniquement sur un critère économique, le niveau de vie des individus. Cette deuxième approche est celle qui est généralement utilisée par les Instituts nationaux de statistique lorsqu'il s'agit de déterminer le taux de pauvreté d'un pays. Les individus sont considérés comme pauvres lorsque leur niveau de vie est sous un certain seuil, qui peut être relatif (60 % du revenu médian en France ou pour Eurostat) ou absolu (aux États-Unis, ce seuil est fixé à 10 830 \$ par an pour une personne seule par exemple).

- Mesurer la pauvreté

Au vu des données à notre disposition, la deuxième approche semble à privilégier dans la mesure où aucune information n'est disponible quant à la santé des individus, leur niveau d'éducation, ou leur capacité d'accès à certaines ressources, différents aspects essentiels à une prise en compte multidimensionnelle de la pauvreté. Le Doc. 1 permet par contre d'avoir des informations très précises sur l'âge et les relations des membres du foyer ainsi que sur leurs revenus. Ces informations permettent le calcul d'un indicateur classique du niveau de vie : le Revenu par Unité de Consommation (RUC)<sup>6</sup>. Le RUC est égal au rapport entre le revenu mensuel disponible avant impôts et le nombre d'unités de

---

<sup>5</sup> Le terme de « foyer allocataire » n'est pas synonyme des notions de « ménage » ou de « foyer fiscal » utilisées par l'Insee ou l'administration fiscale. En effet, selon l'Insee, un ménage désigne généralement « l'ensemble des occupants d'un même logement sans que ces personnes soient nécessairement unies par des liens de parenté ». Deux personnes en colocation qui perçoivent des Aides aux logements forment un même ménage, mais deux foyers allocataires. De même, un foyer fiscal « désigne l'ensemble des personnes inscrites sur une même déclaration de revenus : par exemple, un couple non marié où chacun remplit sa propre déclaration de revenus compte pour deux foyers fiscaux », mais pour un seul foyer allocataire.

<sup>6</sup> Le RUC n'est calculé que pour les foyers pour lesquels les CAF ont des informations précises sur les ressources. Les personnes de plus de 65 ans et les étudiants ne percevant qu'une aide au logement ne sont notamment pas pris en compte.

consommation du foyer<sup>7</sup>. Procéder à cette division permet de rendre les revenus de chaque foyer comparables en les rapportant à la situation d'une personne seule et sans enfant.

Le RUC de chaque foyer est ensuite comparé au « seuil de bas revenus » qui correspond à 60 % du RUC médian (calculé sur l'ensemble de la population à partir de l'Enquête sur les revenus fiscaux et sociaux)<sup>8</sup>. Au 31 décembre 2016, le seuil de bas revenus était de 1 045 euros par mois et par unité de consommation. 4,4 millions de foyers allocataires vivaient sous ce seuil, soit une population de 9,8 millions de personnes (d'après *l'open data* de la CAF).

- Localiser la pauvreté

Afin d'étudier la dimension spatiale de la pauvreté, la variable bas revenus doit être croisée avec des données de localisation. C'est de ce croisement qu'émerge une information géographique qui pourra, par la suite, faire l'objet d'opérations d'agrégations spatiales. D'après Jean Denègre et François Salgé (2004), l'information géographique peut être définie comme un « ensemble reliant une information relative à un objet ou un phénomène du monde terrestre, décrit par sa nature, son aspect et ses attributs (...) et sa localisation sur la surface terrestre, décrite dans un système de référence explicite » (p. 12). Dans le Doc. 1, les données de localisation disponibles sont les adresses des foyers allocataires. Ces adresses peuvent être transformées en données de localisation absolue grâce à un géocodage. Cette opération permet d'obtenir les coordonnées des foyers allocataires dans un système de référence géographique.

Une fois que l'ensemble des informations nécessaires à l'analyse a été identifié, il est finalement possible de rassembler ces renseignements au sein d'un tableau élémentaire d'informations (Doc. 2). Ce tableau permet de décrire les individus qui composent la population étudiée (en lignes) à l'aide des attributs descriptifs sélectionnés (en colonnes). Il constitue la première étape de la formalisation de l'information géographique.

## ***Question 2***

La transformation du tableau élémentaire d'informations en matrice d'information géographique procède au regroupement des individus selon leur localisation. Cette procédure d'agrégation spatiale revient à construire des individus géographiques (*cf.* Doc. 4) impliquant deux choix analytiques : la définition d'une maille spatiale pertinente et le choix de méthodes pour résumer les variables.

---

7 Comme l'Insee, l'échelle utilisée est celle de « l'OCDE modifiée » (1 point pour le premier adulte, 0.5 pour toute autre personne de plus de 14 ans et 0.3 par enfant de moins de 14 ans). Mais à la différence de l'Insee la CAF rajoute 0.2 points pour les familles monoparentales.

8 Bien que le seuil de pauvreté corresponde également à 60% du RUC médian, le seuil de pauvreté et le seuil de bas revenus ne sont pas tout à fait équivalents. Comme évoqué précédemment, ils ne sont pas calculés avec la même échelle d'équivalence pour les UC, ils portent par ailleurs sur des champs différents (foyers fiscaux contre foyers allocataires) et ne tiennent pas compte de la même façon de la fiscalité. En 2017 le seuil de pauvreté défini par l'Insee était de 1041 euros par mois contre 1052 euros par mois pour le seuil de bas revenus défini par la CAF.



À partir du Doc. 2, on peut construire la matrice d'information géographique suivante :

Arrondissement	Nb de foyers	RUC moyen	Nb enfants moyen	Hommes*	Femmes	Mariés	Célibataires	...
1	1	962	1	1	0	1	0	
3	2	/	0	2	0	1	0	
5	1	/	0	1	0	0	0	
6	2	1 314	1	0	2	1	1	
7	1	1 390	0	1	0	0	0	
8	1	1 959	0	1	0	0	1	
13	4	1 744	1,75	2	2	1	1	
14	1	1 660	0	1	0	0	1	
15	2	1 425	2,5	1	1	2	0	
16	3	928	2,67	2	1	0	0	
17	2	1 331	3	2	0	0	0	
18	3	1 245	0,67	2	1	0	0	
20	2	1 598	0,5	2	0	0	1	

\* les variables de sexe peuvent être obtenues à partir du premier chiffre du NIR (numéro de sécurité sociale) présent dans le formulaire : le numéro des hommes commence par un 1, celui des femmes par un 2.

La maille d'agrégation choisie est celle de l'arrondissement parisien. On utilise le code postal de la variable « adresse » pour procéder au regroupement des individus. Les variables quantitatives peuvent être résumées par des indicateurs de position dans la distribution statistique (minimum, maximum, moyenne, médiane). Ici, la variable « RUC » devient une variable décrivant le revenu par unité de consommation moyen des foyers. Les variables qualitatives se transforment par un renversement où les modalités deviennent variables : la variable « Situation familiale » devient une série de 7 variables (célibataire, couple, pacsé, marié, divorcé, séparé, veuf) pour lesquelles on renseigne le nombre de foyers.

La matrice d'information géographique possède plusieurs sens de lecture qui permettent de traiter les données selon différentes perspectives :

- La première direction de lecture oppose une entrée par les **individus** (ici présentés en ligne) et une entrée les **variables** (en colonne) de la matrice. Si la lecture de la matrice se fait en suivant les valeurs que prennent les différents attributs des individus (de gauche à droite), on s'inscrira alors dans une **analyse verticale**, méthode d'observation des spécificités et des associations attributaires des individus. Si la lecture de la matrice se fait en suivant les valeurs que prend un attribut pour chaque individu (de haut en bas), on s'inscrira alors dans une **analyse horizontale**, méthode d'observation des individus par rapport à leur position dans la population. La lecture verticale de la matrice ne renvoie pas nécessairement à une approche spatiale des données : elle cherche à caractériser des individus, qu'ils soient localisés ou non, par leurs propriétés attributaires. L'analyse horizontale de l'information géographique a pour spécificité de positionner spatialement les individus les uns par rapport aux autres, grâce à la variable de localisation. Dans une perspective de géographie régionale, elle cherche à mettre en évidence des structures dans la

distribution spatiale du phénomène observé. Ces deux sens de lecture ne sont pas exclusifs et l'analyse horizontale est bien souvent précédée d'une analyse verticale des individus.

À partir de ces deux sens de lecture, on peut poser deux séries de questions pour interroger la matrice d'information géographique. **Une approche verticale permet de caractériser les arrondissements parisiens** : Quels sont les arrondissements où la majorité des foyers sont pauvres ? ; quels sont les quartiers où les foyers pauvres sont composés de parents célibataires ? ... **Une approche horizontale permet de qualifier la distribution spatiale de la pauvreté** : La part des foyers pauvres dans les arrondissements augmente-t-elle avec la distance au centre ? ; les arrondissements de l'Ouest parisien sont-ils beaucoup plus riches que les arrondissements de l'Est ? ; les arrondissements concentrant les foyers pauvres sont-ils proches les uns des autres ou éloignés ? ...

- La seconde direction de lecture concerne les plans temporels de la matrice : lit-on les différents plans de la matrice ou un seul ? Si un plan de la matrice est pris isolément, on s'inscrira dans une analyse *transversale*, méthode d'observation d'une population dans sa globalité à un instant *t*. Si les plans de la matrice sont pris ensemble, on s'inscrira dans une analyse *longitudinale*, méthode d'observation suivie d'une population au cours d'une période de temps.

La lecture longitudinale de la matrice permet de complexifier les questions posées en cherchant à **qualifier l'évolution des caractères** : la part de foyers pauvres a-t-elle augmenté ou diminué dans les arrondissements ? ; quels arrondissements ont changé de profil sociodémographique ? ... Cette lecture permet aussi de **qualifier l'évolution des distributions spatiales** en répondant aux questions suivantes : les arrondissements les plus pauvres sont-ils de plus en plus loin du centre ? ; la différence entre les arrondissements de l'ouest et de l'Est s'est-elle accrue ou a-t-elle diminué ? ; les arrondissements pauvres sont-ils plus proches ou plus éloignés les uns des autres ?

### **Question 3**

Au sein de la matrice d'information géographique, la localisation des unités spatiales n'est qu'une variable parmi d'autres. C'est ce qui conduit M. Goodchild à parler de « perte du contexte spatial » (Doc. 3). En effet, les unités statistiques (en ligne) peuvent être rapprochées dans la matrice d'information géographique si elles se ressemblent ou si elles sont proches. Mais cette démarche ne permet pas de rendre compte de l'ensemble des relations spatiales qu'elles ont entre elles (contiguïté, distance...).

La carte, à l'inverse, constitue un mode de représentation de l'information géographique qui rend compte du contexte spatial. La lecture d'une carte permet de comparer les modalités d'un attribut pour des unités spatiales voisines, de repérer des discontinuités spatiales, des logiques d'autocorrélation, des phénomènes de concentration ou encore des gradients. Elle facilite aussi la confrontation des distributions observées aux « connaissances qu'on a des contextes locaux et des caractéristiques des lieux qui composent l'espace d'étude » (Doc. 3). Elle permet l'élaboration d'hypothèses pour expliquer des structures spatiales observées.

Le Doc. 5 est composé de trois cartes. Celles-ci représentent la densité des foyers allocataires à bas revenus selon trois mailles géographiques : les arrondissements, les Îlots Regroupés pour l'Information Statistique (IRIS) et un carroyage de 500 mètres. Les IRIS sont un maillage créé par l'Insee à partir de critères démographiques (environ 2000 habitant·e·s par

maille) et morphologiques (homogénéité du bâti). Proches de l'îlot en milieu urbain dense, ils ont longtemps constitué le niveau le plus fin pour la diffusion des données infracommunales issues du recensement de la population<sup>9</sup>.

Les adresses des foyers allocataires étant géocodées, il a été possible d'agréger les données selon chacun de ces maillages. La lecture de ces cartes permet à la fois de repérer des structures spatiales dans la localisation des foyers allocataires à Paris et de réfléchir aux effets des choix d'agrégation sur la représentation de l'information.

Tout comme la matrice d'information géographique, une carte peut être construite et lue selon les différentes perspectives d'analyses évoquées en question 2. Pour rester lisible, elle doit représenter la distribution d'un nombre limité d'attributs. Sachant que les données cartographiées ici portent seulement sur l'année 2016 et que la pauvreté est approchée à partir de seuil de bas revenus, parmi les sens d'analyse évoqués en question 2, les cartes du Doc. 5 offrent des perspectives transversales d'analyse de la **distribution spatiale** des foyers pauvres à Paris. La représentation cartographique n'est donc qu'une possibilité parmi d'autres d'analyse spatiale des données.

Sur la carte par arrondissement il est possible de constater que la densité des foyers se trouvant sous le seuil de bas revenus s'organise selon un gradient centre/périphérie (plus faible au centre), et deux discontinuités principales, Est/ouest et nord/sud. Néanmoins, les arrondissements sont peu nombreux et ils ont une surface importante. De ce fait l'information est particulièrement agrégée et les contrastes sont atténués. La carte est facilement lisible, mais elle est peu précise. Il n'est pas possible d'en tirer des conclusions quant aux structures locales de la pauvreté à Paris.

La comparaison de cette carte avec la carte par IRIS résume bien l'arbitrage entre précision et lisibilité qui doit être réalisé au moment du choix du niveau d'agrégation. La carte par IRIS contient en effet plus d'informations, mais elle s'avère plus difficile à lire et à interpréter. Par ailleurs, étant donné leur petite taille, certains IRIS se trouvent sous le seuil de confidentialité (moins de 100 allocataires y résident, tous revenus confondus). Il n'a pas été possible d'y représenter la densité des foyers à bas revenus. Sur cette carte, il est possible d'observer l'hétérogénéité interne aux arrondissements. Par exemple, dans le 13<sup>e</sup>, il y a de fortes différences entre la rive de Seine et le quartier de Porte d'Italie au sud.

Finalement, le cas des bois de Vincennes et de Boulogne révèle les limites de l'agrégation par arrondissement. Ces derniers appartiennent au 12<sup>e</sup> et au 16<sup>e</sup> arrondissement de Paris et de ce fait leur superficie a été prise en compte au moment du calcul des densités. Pourtant sur la carte par IRIS on constate qu'ils ne sont pas habités par des allocataires. La prise en compte des bois a donc artificiellement fait baisser les mesures de densité pour ces arrondissements.

Ces deux cartes illustrent en fait le MAUP (*Modifiable Areal Unit Problem*) en géographie (Openshaw, 1984), aussi évoqué par M. Goodchild dans le Doc. 3. Le MAUP peut être défini comme « l'influence du découpage spatial sur les résultats de traitements statistiques ou de modélisation » (Salima et De Bellefon, 2018, p. 59). Cette influence renvoie à deux problèmes interdépendants : un effet de taille et un effet de forme. Avec des mailles de grande taille, les résultats sont moins contrastés (arrondissements) alors qu'une forte hétérogénéité est visible avec les IRIS. Par ailleurs la lecture du phénomène est influencée par la morphologie et le positionnement des unités spatiales : la structure en escargot des arrondissements renforce l'effet de gradient centre-périphérie, par rapport à un découpage en tranches de pizza qui mélangerait au sein d'une même unité des quartiers centraux et périphériques. Le MAUP peut donc constituer une limite à la comparabilité des résultats, en particulier lorsque des maillages administratifs irréguliers sont utilisés.

---

9 Depuis 2010, les données du recensement de la population sont aussi disponibles selon un carroyage à 200m du territoire français : <https://www.insee.fr/fr/statistiques/2520034?sommaire=4176305>

L'usage d'un carroyage permet de neutraliser en grande partie le MAUP. En effet en agrégeant les données selon une grille régulière, il est plus facile de comparer les résultats obtenus, car tous les carreaux ont la même taille et la même forme. Il reste néanmoins à déterminer la dimension et le positionnement de la grille les plus adaptés à l'analyse. Ici la densité des allocataires à bas revenus a été représentée selon un carroyage à 500 mètres. Cette échelle constitue un bon compromis entre précision et lisibilité. En effet, la carte permet de repérer des phénomènes de concentration dans la localisation des foyers allocataires à bas revenus à Paris (certains quartiers comme les Batignolles, la Goutte d'Or, Belleville ou les Olympiades ressortent) ainsi que des structures spatiales plus générales. La densité des allocataires à bas revenus décroît depuis ces quartiers révélant des gradients locaux.

#### *Question 4*

La matrice de relation spatiale est une table où les lignes et les colonnes correspondent à des unités spatiales. À chaque couple d'unité spatiale  $(i, j)$  on associe une information décrivant un type de relations entre les deux unités. Le Doc. 7 permet d'identifier quatre types de relations spatiales :

- **Relations de proximité spatiale** : basées sur les distances, la connexité, ou la contiguïté entre les unités spatiales.
- **Relations d'appartenance territoriale** : basées sur l'appartenance commune des unités à des entités territoriales de niveau supérieur.
- **Relations de similarité** : basées sur la ressemblance ou de la dissemblance entre deux unités.
- **Relations d'échange** : basée sur la mesure des flux entre deux unités.

À partir du tableau élémentaire d'information, on peut mesurer les déménagements de foyers allocataires à bas revenus entre arrondissements. On s'inscrit alors dans une matrice de relations d'échange permettant d'analyser la mobilité résidentielle.

La construction de cette matrice qui rend compte des déménagements entre arrondissements permet de réfléchir à nouveau aux implications du choix d'agrégation spatiale. La diagonale d'une matrice de relation spatiale indique en ligne et en colonne la même unité spatiale. Deux possibilités s'offrent alors : considérer que la diagonale ne fait pas sens et sera laissée vide, ou considérer que la diagonale permet de décrire des phénomènes internes à l'unité spatiale considérée. Dans notre cas, il faudra se demander si l'information sur le nombre de déménagements à l'intérieur du même arrondissement est pertinente. On sait qu'environ 40 % des mobilités résidentielles se font sans changer de commune de résidence (Baccaïni et Gamier, 2018), compléter la diagonale de la matrice permet alors de saisir cette réalité.

Dans ce cas, les échanges entre unités spatiales sont orientés, on obtient alors une matrice de relation dissymétrique : la relation de A vers B n'a pas la même valeur que la relation de B vers A. Au contraire, les matrices de relation non orientées ont pour propriété d'être symétriques : par exemple, la distance euclidienne de A à B est identique à celle de B à A.

Arr.	vers																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
de	1																1	1		
	2																			
	3																			
	4																			
	5																			
	6																			
	7																			
	8																			
	9																			
	10																			
	11																			
	12																			
	13																			
	14																			
	15																			
	16																			
	17																			
	18																			
	19																			
	20																			

### Question 5

Le respect de la vie privée et des libertés des individus constitue l'enjeu principal des données personnelles (cf. Doc. 8). Plus précisément, il ne doit pas être possible pour un tiers d'identifier directement ou indirectement (à partir du croisement de plusieurs attributs) une personne à partir des informations collectées par l'organisme. La diffusion non consentie d'informations personnelles peut nuire aux individus : discriminations, fichage, ciblage. En conséquence, l'accès aux données désagrégées doit être restreint et sécurisé, mais les analyses rendues publiques doivent également respecter des seuils de confidentialité garantissant l'anonymat des personnes.

Pour garantir ce principe, la collecte et le traitement des données individuelles doivent obéir à certains critères définis par la Loi (Loi « Informatique et Libertés » et Règlement Général sur la Protection des Données) :

- La collecte, la manipulation et la conservation de ces données doivent répondre à des besoins explicites.

- Les personnes fournissant des renseignements doivent être informées du type de données collectées, de leur utilisation et ont le droit d’y accéder, de les modifier ou de demander leur effacement.

À l’exception de la statistique publique, les données ne sont généralement pas collectées pour l’analyse géographique. Mobiliser ces données en dehors de leur cadre d’utilisation habituel peut paraître contradictoire avec les points précédents : la finalité première des données de la CAF concerne l’attribution des prestations et non un usage en sciences sociales. À travers une convention de mise à disposition des données garantissant la protection de la vie privée des allocataires, il est néanmoins possible pour des chercheurs extérieurs à la CAF d’analyser ces données. Dans ce cas, respecter les normes de confidentialité de ces données ne peut suffire. L’enjeu posé par ces données relève de la définition d’une éthique de la recherche en Sciences sociales :

« Du côté de la géographie humaine, la réflexion [éthique] s’organise autour de deux questions, abordées précocement par deux figures bien différentes. La première concerne l’articulation des valeurs morales avec la pratique de la géographie (Buttimer, 1974, question reprise notamment par Smith, 1998 et 2000). La seconde interroge l’utilité du savoir géographique produit, et invite les géographes à s’engager en réponse auprès des plus démunis et assujettis, et à mettre leurs savoirs et savoir-faire au service de ces derniers (William Bunge et les Expéditions géographiques de Détroit de 1969, voir Merrifield, [1995] 2001). La géographie serait engagée, ou ne serait pas. » (Collignon, 2010, p.70)

## **Conclusion : de l’individu au territoire, des enjeux renouvelés pour la géographie**

En mettant les étudiant·e·s au contact d’un matériau brut produit par une administration et en décomposant les étapes possibles d’une mise en forme pour l’analyse géographique, l’application proposée peut conduire à interroger plus largement le passage complexe de l’observation individuelle à l’analyse du territoire. Les individus statistiques de cet exercice sont des individus sociaux (personnes) ou spatiaux (lieux). L’articulation spécifique de ces individus propre à la géographie pose de longue date plusieurs défis à l’analyse agrégée des territoires et des groupes sociaux.

Le regroupement des individus selon leur position dans l’espace géographique et leurs attributs a conduit à formuler brièvement deux problèmes liés, le *Modifiable Area Unit Problem* (MAUP) et l’erreur écologique. Ces enjeux méthodologiques de l’agrégation spatiale et sociale constituent une piste de prolongement de la séquence. A partir des analyses menées par S. Openshaw sur les cas de leucémie infantile autour de la centrale nucléaire de Sellafield, P. Gould (1992) montre comment le regroupement inductif, débarrassé des mailles a priori, et itératif permet d’identifier des concentrations spatiales significatives d’un phénomène observé au niveau individuel. Cette démarche met la méthode d’agrégation spatiale au cœur de l’analyse pour rendre compte du rôle de l’espace et de ses échelles dans la structuration des phénomènes sociaux (Eckert et al., 2013).

Pour poursuivre l’analyse thématique de la pauvreté, le débat sur les « effets de quartier » (Bacqué, Fol, 2007) peut amener à approfondir la réflexion sur la relation individu-territoire en comprenant le risque d’erreur écologique (Grasland, 1994) qui pèse sur l’interprétation de relations observées au niveau des agrégats. Cette feuille peut également engager la réflexion sur l’intérêt que porte la discipline à de nouvelles données désagrégées (*open data* administratif, traces numériques...) pour lesquelles il est nécessaire de construire

les agrégats de l'analyse. Comme pour les données de la CAF, la manipulation de ces données nouvelles interroge la possibilité d'inférer des propriétés des territoires à partir de populations aux contours flous et implique de considérer les enjeux éthiques de la manipulation de données personnelles désagrégées.

### **Pour aller plus loin**

Fayard B., Chemineau D., 2018, « Données statistiques allocataires relatives à la pauvreté et la précarité en Île-de-France fin 2017 », *Recueil de données sociales des CAF en Ile-de-France*, n° 14, 16 pages.

Chemineau D., Flamand J., 2014, « La pauvreté en Ile-de-France en 2013 : 1,6 million de Franciliens vivent dans des foyers allocataires à bas revenus dont 739 000 jeunes de moins de 25 ans », *Bulletin d'information des CAF en Ile-de-France*, n° 19, 12 pages.

Bayardin V., Herviant J., Jabot D., Martinez C., Chemineau D., Glachant E., Guérin G., 2017, « En Île-de-France, la pauvreté s'est intensifiée dans les territoires déjà les plus exposés », *Insee Analyse*, n° 76, 4 pages.

## Bibliographie

Aschan-Leygonie C., Cauty C, Davoine P.A., 2019, *Les systèmes d'information géographique*, Armand Colin.

Bacqué M.-H., Fol S., 2007, « 14. Effets de quartier : enjeux scientifiques et politiques de l'importation d'une controverse », in Authier J.-Y. (dir.), *Le quartier*. La Découverte, pp. 181-193.

Baron M., Zanin C., 2005 a, « Un tableau de données : pour quoi faire ? », *Feuilles de Géographie*, I-2005, Feuille n° 56.

Baron M., et, Zanin, C., 2005b, "L'analyse de données : pour quoi faire ? Familiarisation au commentaire et à l'interprétation des résultats d'une Analyse en Composantes Principales normées (ACP normée)", *Feuilles de Géographie*, Feuille n° 57.

Baron M., Zanin C., 2006. " 'Le monde peut-il suffire ? ' Pour une ACP — Familiarisation à la réalisation et à l'interprétation des résultats d'une Analyse en Composantes Principales normées", *Feuilles de Géographie*, Feuille n° 58.

Bavoux J.J., Chapelon L. (dir), 2014, *Dictionnaire d'analyse spatiale*, Armand-Colin.

Chemineau D., Bergeron T., 2016, « La mobilité résidentielle des allocataires parisiens », *Dossier d'études des CAF en Ile-de-France n° 10*.

Commenges H., 2017, « Analyse de données et cartographie — Devoir sur table. Étude de la ségrégation urbaine au Cap (Afrique du Sud) », *Feuilles de géographie*, Feuille 2017-2.

Denègre J., Salgé F., 2004, « L'information géographique ». In Denègre J., *Les systèmes d'information géographique*, Presses Universitaires de France, pp. 12-37.

Dumolard P., Dubus, N. Charleux, L., 2003, *Les statistiques en géographie*. Editions Belin.

Eckert D., Baron M., Jégou L., 2013, « Les villes et la science: apports de la spatialisation des données bibliométriques mondiales », *Mappemonde*, n°110

Goodchild M. F., Janelle D. G. (Eds.), 2004, *Spatially integrated social science*. Oxford University Press.

Gould P., 1992, « Épidémiologie et maladie », in Bailly A., Pumain D., Ferras R. (dir.), *Encyclopédie de géographie*, Economica, pp. 947-967

Grasland C., 1994, « L'erreur écologique en géographie », *Feuilles de Géographie*, I-1994, Feuille n° 1.

Grasland C., 1998, « Chapitre 1 : Introduction générale – Définitions », In Grasland C., *Initiation aux méthodes statistiques en sciences sociales*, consulté le 25 septembre 2019, [http://grasland.script.univ-paris-diderot.fr/STAT98/stat98\\_1/stat98\\_1.htm](http://grasland.script.univ-paris-diderot.fr/STAT98/stat98_1/stat98_1.htm)

Lambert, N., Zanin C., 2016, *Manuel de cartographie : principes, méthodes, applications*. Armand Colin.

Le Goix R., 2018 « Analyse spatiale (chapitre 2) ». In Ciattoni A., Veyret Y, *Les fondamentaux de la géographie*, Armand Colin, pp. 44-67.



Morange M., Schmoll C., 2016, *Les outils qualitatifs en géographie : Méthodes et applications*. Armand Colin.

Openshaw S., 1984, *The modifiable areal unit problem*, T. CATMOG 38, GeoBooks, Norwich.

Pumain D., Saint-Julien T., 2010, *L'analyse spatiale (1) : Localisations dans l'espace*, 2<sup>e</sup> édition, Cursus-géographie, Armand-Colin.

Pumain D., Saint-Julien T., 2010, *L'analyse spatiale (2) : Les interactions spatiales*, 2<sup>e</sup> édition, Cursus-géographie, Armand-Colin.

Salima B.A., De Bellefon M.-P., 2018, « Indices d'autocorrélation spatiale ». In Loonis, V. (dir.), 2018, *Manuel d'analyse spatiale*, Insee Méthodes n° 131.

Sen A., Anand S., 1997, Concepts of Human Development and Poverty: A Multidimensional Perspective. In: *Poverty and Human Development: Human Development Papers 1997*. New York: United Nations Development Programme, pp. 1-20.