

REPRÉSENTATION CARTOGRAPHIQUE ET ÉLABORATION DE TYPOLOGIES

Philippe LOUCHART

*Institut d'Aménagement et d'Urbanisme de la Région Ile-de-France
(IAURIF)*

L'analyse des résultats du recensement de 1990 a donné lieu, en Ile-de-France, à une collaboration approfondie entre la direction régionale d'Ile-de-France de l'INSEE et l'Institut d'aménagement et d'urbanisme de la région Ile-de-France. Celle-ci s'est concrétisée par la publication des deux tomes de l'*Atlas des Franciliens*¹.

Le tome I, paru à l'automne 1991, s'attachait, à travers 32 articles et 94 cartes, à présenter les évolutions révélées par les premiers résultats du recensement de 1990, c'est-à-dire le dénombrement de la population et des logements.

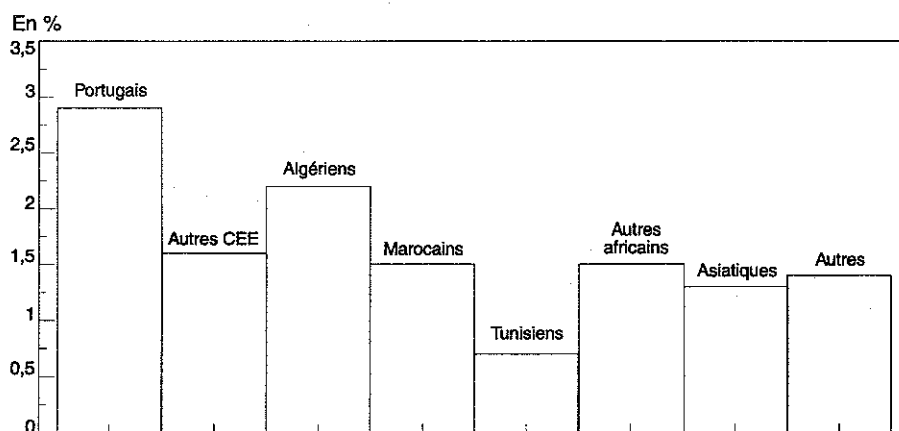
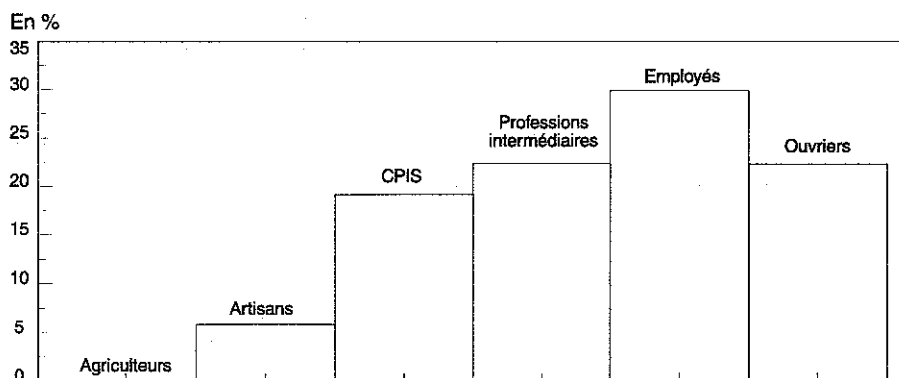
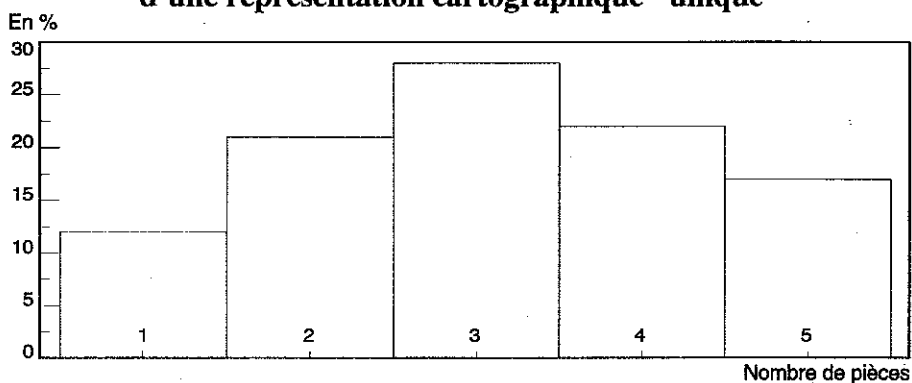
Le tome II, paru au début de l'année 1993, a permis de tracer un portrait détaillé de la région capitale à partir de l'analyse et de la cartographie des données communales issues du dépouillement exhaustif et du sondage au quart du recensement de 1990.

Ce second tome, plus volumineux (65 articles et 175 cartes), est surtout plus riche en informations. Cette richesse d'information nous a conduit à faire preuve d'un peu d'imagination pour cartographier celle-ci sans trop la simplifier, en particulier, lorsque pour certaines variables, il a été décidé de représenter l'ensemble de leurs modalités sur une seule carte : logements selon le nombre de pièces, ménages selon le nombre de personnes, actifs selon leur catégorie socioprofessionnelle, personnes seules dans l'ensemble des ménages selon l'état matrimonial, étrangers dans la population selon leur nationalité, ... (voir graphiques ci-après).

Pour ce faire, les techniques classiques de classification automatique ont été utilisées, mais d'une façon pas très classique. Il s'agissait en effet de produire une représentation cartographique des données analysées, avec les contraintes de lisibilité que cela implique, et non d'élaborer une véritable typologie des communes franciliennes. En outre, l'ensemble des 1300 communes d'Ile-de-France devait être représentées, et non les plus peuplées uniquement.

1. IAURIF-INSEE, Atlas des Franciliens, Tome I : Population et logements, 80 pages, 94 cartes couleurs, Paris 1991 ; Tome II : Age, emploi, modes de vie ..., 160 pages, 175 cartes couleurs, Paris, 1992.

Quelques exemples de variables ayant fait l'objet d'une représentation cartographique "unique"



Source : IAU RIF, décembre 1993

Après un bref rappel des principales méthodes de discrétisation utilisées pour cartographier des variables quantitatives, la démarche suivie et les règles adoptées pour élaborer ces cartes "en typologie" sont présentées, tout comme les cartes obtenues.

Le résultat est assez instructif : alors qu'aucune contrainte de contiguïté spatiale n'a été intégrée explicitement lors de l'élaboration de ces "typologies", la représentation cartographique montre une très forte spécialisation de l'espace, variable, mais pas toujours, selon la nature de l'information traitée.

1. Discrétisation de variables quantitatives : principes et limites

La principale difficulté que soulève la représentation cartographique d'une seule variable, quantitative et continue, comme la proportion d'actifs entre 20 et 24 ans, tient dans le choix des seuils.

De très nombreuses "formules" ont été proposées pour "découper" un ensemble de n valeurs en m classes, mais il n'existe pratiquement jamais de solution unique au problème du choix des seuils. Toutes donnent, en général, des résultats différents, et la partition retenue dépend, en fait, d'un nombre élevé de paramètres qui très souvent changent d'une carte à l'autre : l'allure générale de la distribution qu'il faut décrire, le but recherché par la représentation cartographique, le public visé, l'échelle retenue, la nécessité éventuelle de représenter, sur une même page, la même variable à différentes dates, etc ...²

Mais, une fois les seuils choisis, si la variable comporte un nombre élevé de modalités, et qu'une seule d'entre elles est cartographiée, le résultat obtenu risque d'être relativement "pauvre". La variable "catégorie socioprofessionnelle", par exemple, comprend un nombre élevé de modalités.

Cartographier les seuls ouvriers permettra, certes, de distinguer aisément les secteurs où la proportion d'ouvriers est élevée, de ceux où cette proportion est faible. En revanche, ceci ne permettra pas, lorsque cette proportion est faible, de savoir si c'est en raison d'une proportion élevée d'agriculteurs, d'employés, ou encore de cadres supérieurs. En ce sens, la représentation obtenue est "pauvre", puisque seule une partie de l'information a été cartographiée.

2. Sur ce sujet, voir, par exemple, la synthèse réalisée par : CAUVIN Colette, REYMOND Henri et Abdelaziz SERRADI, "Discrétisation et représentation cartographique", Montpellier, GIP RECLUS, coll. Reclus modes d'emploi, 1987, 116 pages.

Il est vrai que la visualisation des différentes modalités d'une variable sur une seule carte nuit à sa lisibilité. Pour cette raison, la solution retenue généralement consiste à présenter plusieurs cartes séparées : une pour chaque modalité. Leur juxtaposition est ensuite supposée restituer l'information complète. Ce n'est pas la solution qui a été retenue, pour différentes raisons, liées essentiellement à des problèmes d'échelle, de dimension de la page et, au final, de lisibilité. En deçà d'une certaine échelle, il devient en effet impossible de distinguer les 1.300 communes qui composent l'Ile-de-France.

Il est encore plus difficile de repérer la même commune sur plusieurs cartes de petit format. Il est, en outre, pratiquement impossible de reconstituer, mentalement, la répartition des actifs d'une commune donnée par catégorie socioprofessionnelle, par exemple, à partir de plusieurs cartes juxtaposées : l'une sur les agriculteurs, l'autre sur les commerçants ou artisans, une autre encore sur les ouvriers, etc.

Pour ces différentes raisons, la solution retenue a été de réaliser une seule carte, de "grand format", en recourant à des techniques d'analyse un peu plus complexes (techniques de classification automatique), mais dont le but est toujours le même : simplifier la réalité pour en produire une représentation cartographique lisible qui reste la plus informative possible.

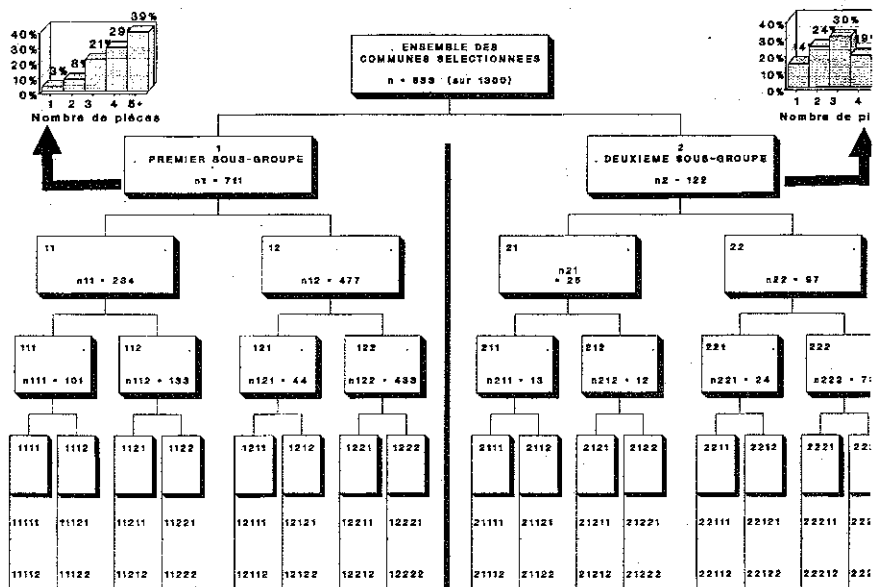
L'objectif poursuivi est de "typer" les communes d'Ile-de-France, en les regroupant dès qu'elles présentent des similitudes. Il s'agit, par exemple, de regrouper les communes qui ont une répartition identique, ou voisine, des logements selon le nombre de pièces, des ménages selon le nombre de personnes qui les composent, des actifs selon leur catégorie socioprofessionnelle, etc.

2. Classification automatique et cartographie

Les techniques de classification automatique ont ici été utilisées pour produire une représentation cartographique des données analysées, et non pour élaborer une véritable typologie des communes franciliennes. Cet objectif a introduit une contrainte forte : celle de limiter, *a priori*, le nombre de groupes de communes à une quinzaine, au maximum, pour que les cartes restent "lisibles". Cet objectif explique aussi la démarche en deux étapes qui, systématiquement, a été adoptée pour élaborer les cartes en typologie présentées dans l'Atlas des Franciliens.

Lors de la **première étape**, les communes de moins de 2.000 habitants ou de moins de 200 résidences principales, dans lesquelles vivent moins de 5% des Franciliens, ont été exclues de l'analyse, soit entre 470 et 790 communes sur les 1.300 que comptent l'Ile-de-France. Ont aussi été exclues, parmi les communes restantes, toutes celles qui comptaient des effectifs trop faibles, compte tenu de la variable étudiée, pour faire l'objet d'une analyse très détaillée.

PROCESSUS DE SEGMENTATION DU FICHIER



EX : LOGEMENTS SELON LE NOMBRE DE PIÈCES

En fait, c'est sur un sous-ensemble de 200 à 800 communes, parmi les plus peuplées, qu'ont été appliquées, de façon systématique, les techniques d'analyse factorielle et de classification hiérarchique ascendante³, selon une démarche itérative qui s'inspire de celle utilisée pour réaliser la "typologie socioprofessionnelle de l'Ile-de-France"⁴.

A la première étape, les communes retenues sont réparties en deux groupes, de taille variable, mais très "opposées" au regard des variables étudiées. Dans l'exemple reproduit dans le schéma ci-dessus, la première segmentation oppose 711 communes où dominent la proportion de grands logements (68% des logements comptent 4 pièces ou plus), à 122 communes où ce sont les petits logements qui sont la dominante (68% des logements comptent 1, 2 ou 3 pièces). A la seconde étape, les mêmes techniques sont de nouveau appliquées à chacun de ces deux sous-ensembles de communes,

3. Les programmes suivants du logiciel d'analyse des données ADDAD ont été utilisés en chaîne, leurs fichiers paramètres étant générés automatiquement par des programmes FORTRAN ad hoc :

ANCORR (analyse factorielle des correspondances)

CAHVOR (classification ascendante hiérarchique selon la maximisation du moment centré d'ordre deux d'une partition, utilisant la méthode des voisins réductibles).

IMPCAH (Représentation d'une classification hiérarchique)

CLACAH (Stockage de partitions à partir des paramètres d'une hiérarchie)

4. BESSY, N. TABAY : Typologie socioprofessionnelle de l'Ile-de-France", Insee, Direction régionale de Paris, 1990, 81 pages.

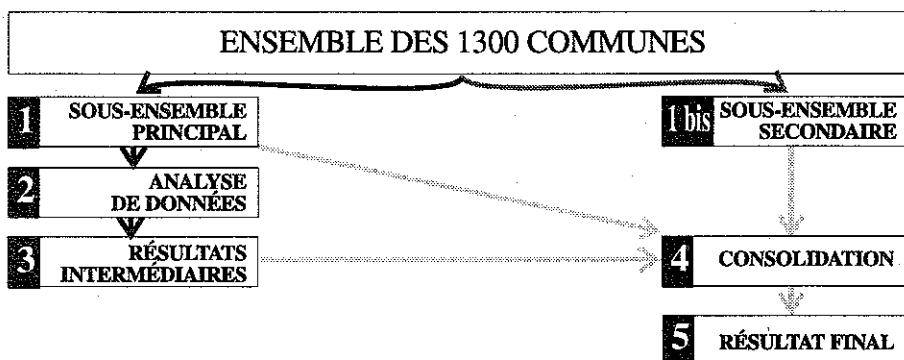
conduisant à une partition en quatre groupes de l'ensemble initial de communes, et ainsi de suite, jusqu'à obtenir 32 groupes différents.

Ainsi, parmi les 122 communes où les petits logements dominent, par exemple, la seconde étape permet de faire un tri entre 25 communes où la proportion de studios et de deux pièces avoisine 60% et 97 communes où le parc de logements est un peu plus grand (22% de deux pièces, 31% de 3 pièces et 22% de 4 pièces). Quant à celles où la proportion de grands logements est élevée, la seconde étape entraîne, par exemple, une distinction entre 234 communes qui comptent à peu près autant de 3 pièces, de 4 pièces que de 5 pièces ou plus (entre 26% et 29%), et 477 communes où près de la moitié des logements comptent 5 pièces ou plus ; et ainsi de suite.

A chaque étape, de nouvelles oppositions, de moins en moins prononcées, sont mises en évidence. A l'issue du processus, les 200 à 800 communes de départements sont ventilées en un maximum de 32 groupes différents qui, bien sûr, ne peuvent être représentés sur une seule carte.

Souvent, le nombre de groupes réellement différents est inférieur à 32 car, très tôt dans le processus de segmentation, certains sous-ensembles ne comportent plus qu'un nombre très restreint de communes ou ne diffèrent plus que de façon marginale. Il arrive aussi qu'un ou plusieurs de ces 32 groupes comportent un nombre très élevé de communes. Ce ou ces groupes sont alors de nouveau segmentés jusqu'à ce que les différences mises en évidence deviennent minimales.

A l'issue de cette phase, le nombre de groupes différents obtenus, s'il est inférieur à 32 est, en revanche, toujours supérieur à la quinzaine, nombre maximum de groupes qui peuvent être représentés, et identifiés, sur une même carte, fût-elle de grand format. Il faut donc opérer des choix.



Pour ce faire, trois principaux critères ont été utilisés : le nombre de communes dans chaque groupe, la population concernée et, bien sûr, la plus ou moins grande proximité entre chaque groupe, mesurée par la distance du chi-deux.

Une fois ces choix effectués, la **seconde étape** consiste à affecter chaque commune d'Ile-de-France, y compris celles qui ont été exclues de l'analyse initiale, à un groupe, et un seul, parmi les dix à quinze qui ont été retenus. Pour ce faire, on calcule, pour chaque commune, la "distance" entre le profil de la commune et le profil moyen de chacun des groupes retenus. Le même calcul est aussi effectué par rapport au profil moyen des communes exclues lors de la première étape, celles-ci pouvant présenter des caractéristiques différentes de celles mises en évidence à l'issue de la première étape.

Chaque commune est ensuite affectée au groupe dont elle est la plus proche, au sens de la distance du chi-deux. Lorsque toutes les communes ont été classées, le "profil-type" de chaque groupe est recalculé en faisant la moyenne des profils communaux, autrement dit en accordant la même importance à chaque commune, qu'elle compte 1.000 habitants ou, au contraire, 250.000 (l'ensemble de la démarche est résumé dans le schéma ci-contre).

Ces profils font l'objet d'une représentation graphique sous chaque carte en typologie, et servent de légende. La population totale des communes qui font partie de chaque groupe figure aussi sur ces graphiques, ceci afin de pondérer l'impression visuelle, parfois trompeuse, que peut laisser une lecture rapide des cartes.

Parfois, par souci de simplification et de fiabilité statistique, toutes les communes, dont l'effectif analysé était inférieur à un seuil donné, ont été regroupées ensemble ; par exemple, toutes les communes qui comptent moins de 20 étrangers, ou encore toutes celles dont la proportion de ménages d'une personne est inférieure à deux fois la moyenne régionale. L'objectif poursuivi est toujours le même : rendre compte de la diversité des situations communales, d'autant plus finement que les populations concernées sont nombreuses et les écarts observés, significatifs.

Bien sûr, le nombre de groupes ayant été, a priori, limité à une quinzaine, l'écart est, certes minimum, entre le profil d'une commune donnée et le profil du groupe auquel elle a été rattachée, mais cet écart peut être assez élevé. Qui plus est, l'écart avec un autre groupe peut s'avérer à peine plus fort. Faute de place, il n'a pas été possible d'indiquer, pour chaque commune, l'écart moyen avec chacun des groupes. Les lecteurs, qui ne s'intéresseront qu'à une ou deux des 1.300 communes franciliennes cartographiées, sont donc invités à vérifier la plus ou moins grande proximité de ces communes avec le profil moyen du groupe auquel elles ont été rattachées.

**Annexe 3 : probabilités de transition à partir du chômage selon les états précédemment occupés
(en pourcentage)**

Jeunes détenteurs d'un CAP ou d'un BEP

État précédent	Première entrée	TUC		Autre PEP		SIVP		CDD		CQ, CA.App		CDI	
		0	3	8	3	8	3	8	8	14	8	14	
Durée potentielle d'éligibilité (en mois)		0	3	8	3	8	3	8	8	14	8	14	
Probabilité de transiter vers													
CDI	(1)	15,9	3,9	10,1	12,1	17,6	21,1	24,7	28,1	21,4	24,0	25,8	28,0
	(2)	11,4	3,9	13,7	18,7	18,9	24,9	18,7	25,1	20,7	24,4	34,4	39,2
	(3)	11,2	3,6	14,7	19,7	18,8	24,5	18,7	25,1	20,8	24,1	34,3	39,1
CDD	(1)	41,4	17,1	53,1	54,8	33,4	34,7	52,1	52,7	21,5	22,3	50,3	50,7
	(2)	45,4	13,0	42,2	40,1	29,5	27,4	61,9	58,0	21,4	20,7	39,5	37,2
	(3)	44,5	11,7	39,9	37,8	26,8	24,9	62,5	58,7	19,2	18,6	39,7	37,4
PEP	(1)	30,7	70,3	30,1	28,1	42,6	39,6	14,2	12,7	43,7	42,5	15,7	14,7
	(2)	30,1	74,0	35,8	35,1	45,0	43,1	12,0	11,6	45,3	44,6	17,3	16,6
	(3)	29,5	74,5	38,7	37,8	48,8	46,7	11,4	11,0	50,0	49,2	17,2	16,6
Probabilité d'une durée au chômage supérieure à un an	(1)	16,8	7,0	9,6	8,9	7,7	7,0	13,7	11,3	29,0	28,8	15,7	14,1
	(2)	14,6	6,4	10,7	12,0	6,8	7,2	8,6	9,4	29,1	30,4	18,7	18,4
	(3)	18,8	8,5	9,1	15,3	6,6	11,4	11,2	17,7	34,3	43,2	30,2	36,9

Notes

- (1) Modèle avec hétérogénéité non observée et corrigé du biais d'échantillonnage dans le stock.
- (2) Modèle sans hétérogénéité non observée mais corrigé du biais d'échantillonnage dans le stock.
- (3) Modèle sans hétérogénéité non observée et sans correction du biais d'échantillonnage dans le stock.