



Les carroyages : la remise au carré de l'analyse spatiale

Enjeux et perspectives

Jean-Luc LIPATZ
INSEE/DG





Septembre 2009

Rapport du groupe de travail
**STATISTIQUES ET NOUVELLES TENDANCES
DE LOCALISATION DES POPULATIONS ET
DES ACTIVITES SUR LE TERRITOIRE**

Président : Francis CUILLIER

*“Troisième point de convergence : la nécessité de
**faire évoluer la géographie de production et de
diffusion des statistiques territoriales.**”*

**3.1.2 - Un fichier fournissant des estimations de
population et d'emploi** (et si possible de personnes
en formation...)

selon un maillage « **carreau** » de surface à définir et/ou
selon un maillage parcelle cadastrale avec la surface
(pour calculer des **densités**).



Les enjeux

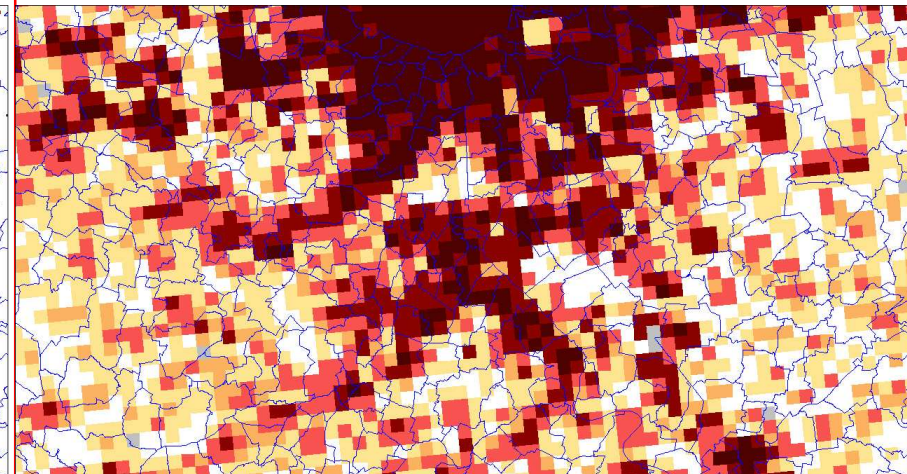
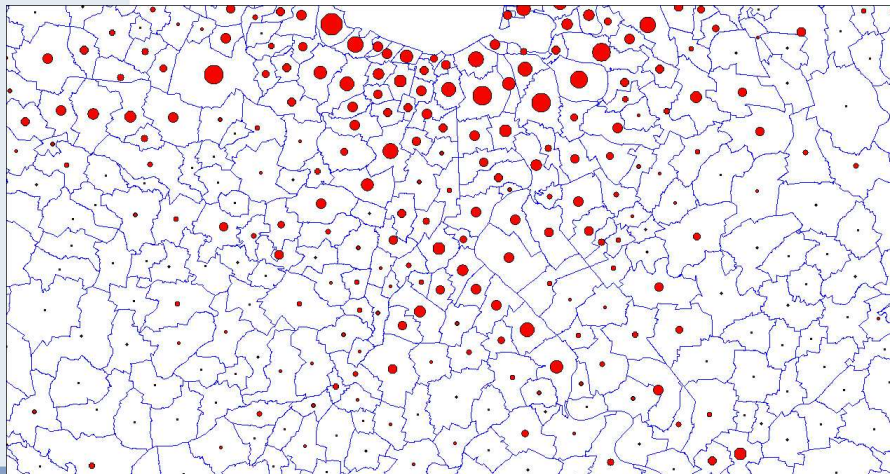
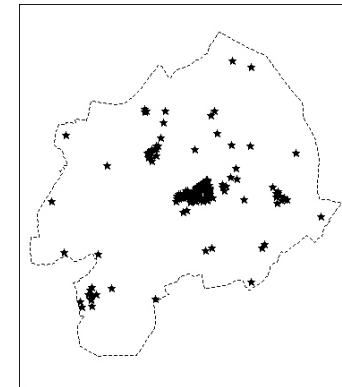
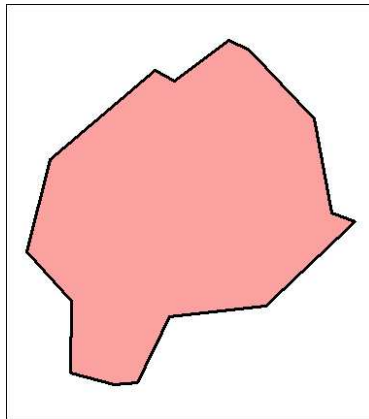
Le point de vue d'un statisticien spatial

Les travaux européens



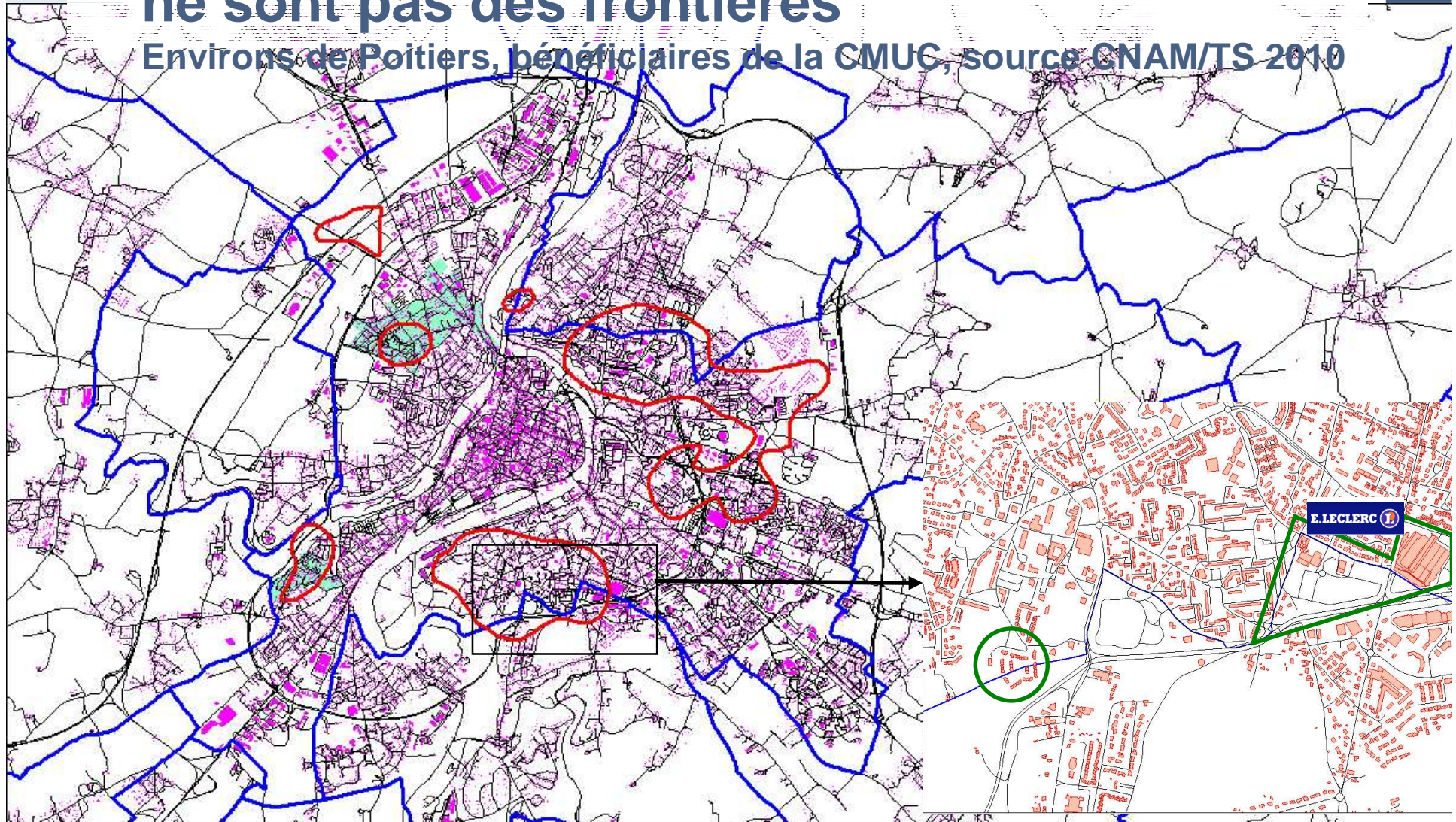


Un credo : Les phénomènes humains transcendent les découpages administratifs



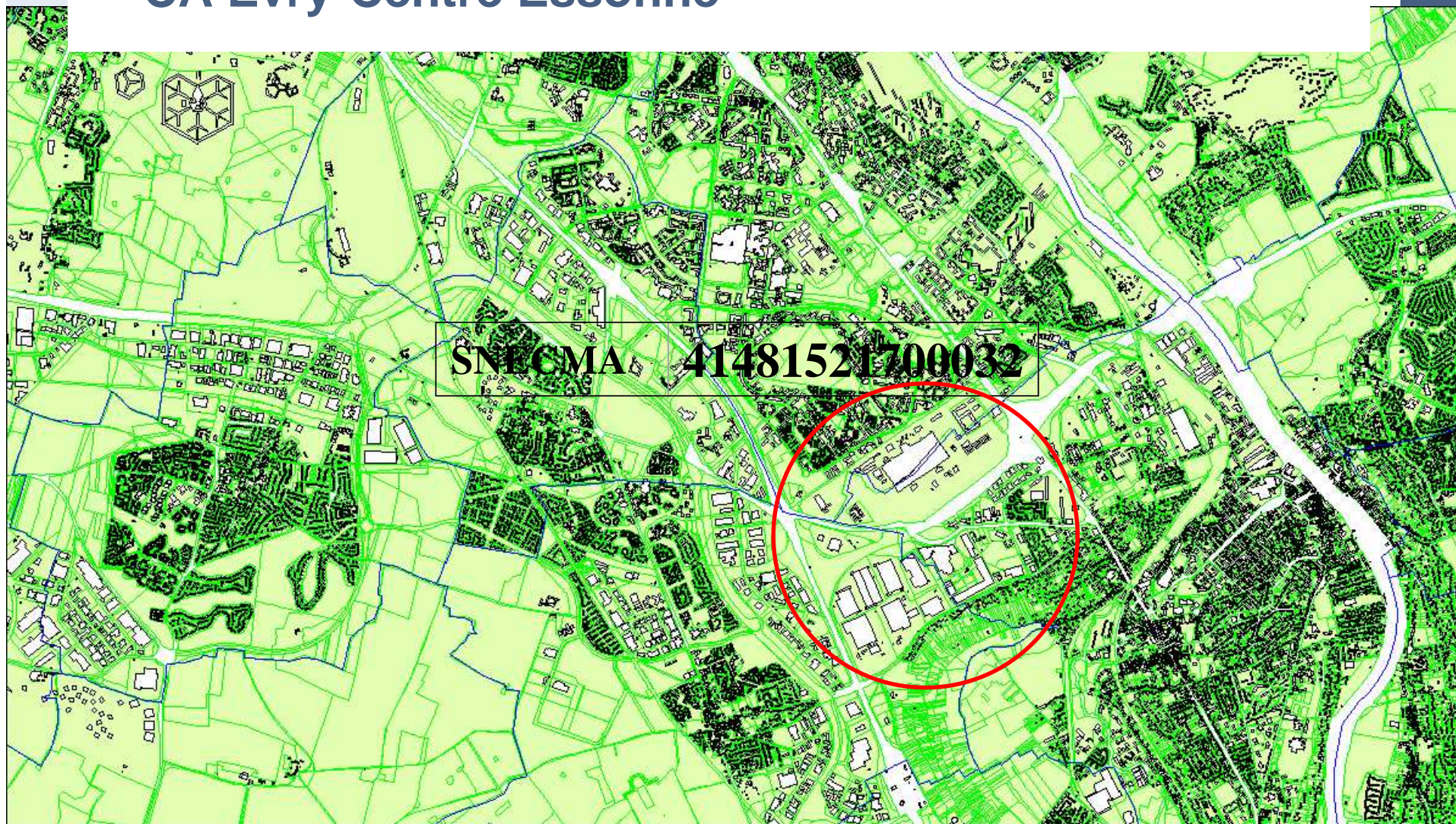
Les limites administratives ne sont pas des frontières

Environs de Poitiers, bénéficiaires de la CMUC, source CNAM/TS 2010



Les frontières administratives ne sont pas des limites

CA Evry-Centre Essonne





Les outils : la concurrence de deux mondes

- › Géocodage : affecter un code de zone (COG, ZUS, IRIS , NQP, (ilot),...)
 - + facile à utiliser
 - + supporte les localisations approximatives (adresses incomplètes)
 - la dimension géographique est faiblement présente
 - n'est faisable que sur des zones prédéfinies
 - susceptible de biaiser les analyses

La statistique classique

- › Géolocalisation : affecter une position géographique
 - nécessite des outils d'analyse
 - pas possible pour toutes les observations
 - + la dimension géographique est complètement intégrée
 - + flexibilité maximum
 - + impartial
 - + input idéal pour les techniques d'analyse spatiale

L'approche de l'INSEE
À l'échelle de la ville
(ou + si affinités)

- › Carroyages : réconcilier les deux approches

Page 7





Deux objectifs pas convergents

- › Permettre le suivi de territoires d'action
 - Données géocodées sur insee.fr
 - Extensions locales

Production « cadrée »
«Exactitude » du chiffre

- › Autoriser les analyses contextuelles : repérage de nouveaux territoires d'action, évaluation de territoires existants
 - Données carroyées
 - DIAF-RP
 - Prestations d'analyse

Approximations tolérées
Souplesse maximum

... quelle que soit la taille de la commune
et sans contrainte de frontières communales



Pourquoi faire si compliqué?

Nos faux amis les zonages

La deuxième dimension oubliée

1 Bénéficiaires de la couverture maladie universelle complémentaire (ZULU)

Ville de Niort par IRIS.
Source insee.fr

					Zone		Population totale		Population À bas revenu	
					AL	A	acmuc	CCmuc		
					Allocataires CNAM	Population couverte	Allocataires CNAM bénéficiaires de la CMUC	Population bénéficiaire de la CMUC		
8	Code région	Code unité urbaine (agglomération)	Code département / commune	Nom de la commune	Code IRIS	Nom de la zone	Allocataires CNAM	Population couverte	Allocataires CNAM bénéficiaires de la CMUC	Population bénéficiaire de la CMUC
9	54	79501	79191	Niort	0101	CENTRE	2296	2918	300	430
10	54	79501	79191	Niort	0102	GARE	1373	1770	153	242
11	54	79501	79191	Niort	0103	HOPITAL	911	1186	66	110
12	54	79501	79191	Niort	0201	NORD PARTIE INFÉRIEURE	1068	1429	55	90
13	54	79501	79191	Niort	0202	NORD PARTIE SUPÉRIEURE	1621	2192	129	272
14	54	79501	79191	Niort	0203	BATICOOP	1688	2180	57	99
15	54	79501	79191	Niort	0301	SUD RIBRAY	1886	2443	85	133
16	54	79501	79191	Niort	0302	SAINT-FLORENT	1602	2117	32	65
17	54	79501	79191	Niort	0303	BOUCLE SEVRE	2133	2712	87	143
18	54	79501	79191	Niort	0401	CLOU BOUCHET NORD	1240	1692	187	389
19	54	79501	79191	Niort	0402	VENISE VERTE	1241	1558	37	69
20	54	79501	79191	Niort	0403	FOUR CHAROT GAVACHERIE	1066	1455	182	339
21	54	79501	79191	Niort	0404	CLOU BOUCHET SUD	1249	1901	330	655
22	54	79501	79191	Niort	0501	SAINT-LIGUAIRE-NANTES	1414	1932	16	23
23	54	79501	79191	Niort	0502	COTEAU SAINT-HUBERT	1361	1880	35	79
24	54	79501	79191	Niort	0503	SAINTE-PEZENNE	1385	1878	39	67
25	54	79501	79191	Niort	0504	ZONE D'ACTIVITE DE SAINT-LIGUAIRE	91	116	1	1
26	54	79501	79191	Niort	0505	ZONE D'ACTIVITE TECHNOPOLE		95		
27	54	79501	79191	Niort	0601	SURIMEAU-ANTES CHOLETTE	1359	1979	50	99
28	54	79501	79191	Niort	0602	JUSTICES-BRIZEAUX	1549	2152	53	116
29	54	79501	79191	Niort	0603	MOULIN A VENT-SOUCHE	1311	1767	35	66
30	54	79501	79191	Niort	0604	CHAMPCLAIROT	1587	2121	67	114
31	54	79501	79191	Niort	0605	ZONE D'ACTIVITE DE SOUCHE	109	140	6	13
32	54	79501	79191	Niort	0701	CHAMPOMMER	2081	2791	95	136
33	54	79501	79191	Niort	0702	ES ST	1328	1794	35	78
34	54	79501	79191	Niort	0703	ZONE D'ACTIVITE DE SAINT-FLORENT		81		

Forte proportion de bas revenus

Faible proportion de bas revenus

Cela ne décrit pas les disparités socio spatiales!

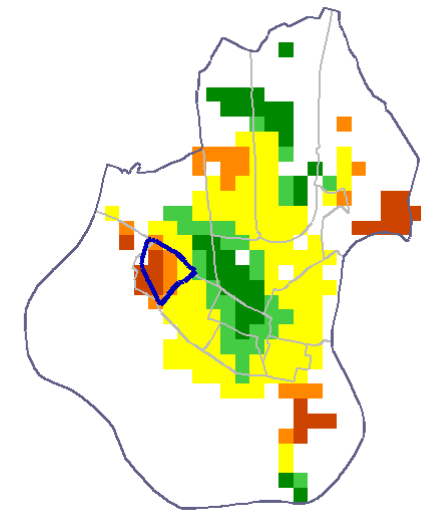


Nos faux amis les zonages

Le « Modifiable Area Unit Problem » - le piège

Code IRIS	Nom de la zone	Population couverte	Population bénéficiaire de la CMUC	
0101	CENTRE VILLE NORD	2315	435	19%
0102	CENTRE VILLE SUD	1909	414	22%
0201	CHAMPLAIN	2487	284	11%
0202	RENAN-PARAT	1251	93	7%
0203	LA PHILAUDRIE	1420	93	7%
0204	CHANTE-ALOUETTE	1496	116	8%
0301	CASSE AUX PRETRES	2116	252	12%
0302	LE GRAND BREUIL	1365	80	6%
0303	LIBERATION	1030	322	31%
0401	LA BEAUME-AVANT GARDE	1643	173	11%
0402	PETIT MARSEILLE-LA MAURATIERE	2171	406	19%
ZUS	PETIT MARSEILLE	954	355	37%

ZUS « Petit-Marseille » à Rochefort
*Une petite portion d'un IRIS hétérogène
(en rouge le parc social)*



rf12008 - Surreprésentation (occtyp='2') 0.28 << >> 2.30

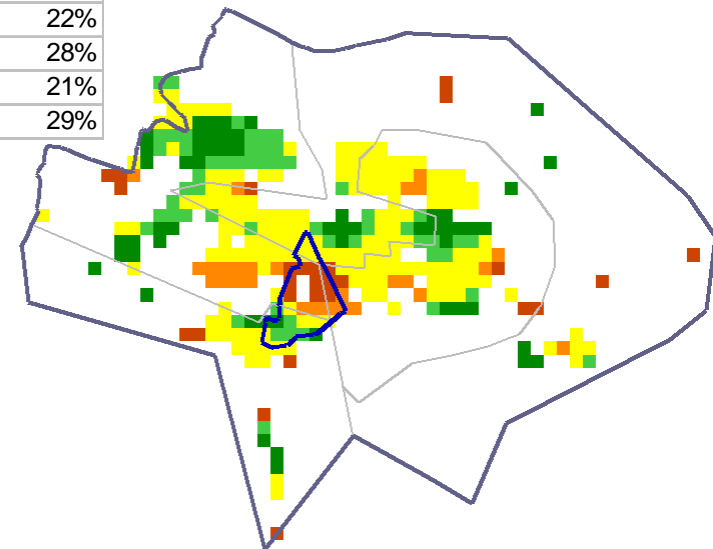




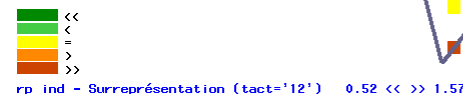
Nos faux amis les zonages

La nécessité d'une vision plus fine

Nom de la zone	Population couverte	Population bénéficiaire de la CMUC	
CROUIN	2072	492	24%
BELLEVUE	1192	381	32%
PETIT MARSEILLE	964	355	37%
OZON, LES RENARDIERES	5155	1792	35%
BASSEAU GRANDE GARENNE	3882	1254	32%
BEL AIR, GRAND FONT	1681	582	35%
MA CAMPAGNE	3141	681	22%
CHAMPS DE MANOEUVRE	3314	1188	36%
RESIDENCE PIERRE LOTI (RESIDENCE KARINE)	987	167	17%
MIREUIL-LALEU-LA PALLICE-LA ROSSIGNOLETTE	13963	1898	14%
VILLENEUVE LES SALINES	6309	1403	22%
CLOU BOUCHET-TOUR CHABOT-GAVACHERIE	5034	1427	28%
BEL AIR-POITIERS OUEST	2714	561	21%
BELLEJOUANNE-POITIERS SUD	1078	308	29%



ZUS « Pierre Loti » à Aytré
*Un territoire hétérogène à recibler
 (en rouge les chômeurs)*





Les carroyages

- + Découpage du territoire en mailles carrées de taille fixe : un zonage simple, donc potentiellement universel

 - Simple à construire

 - Simple à comprendre : précision géographique constante

- + Un zonage qui exprime naturellement les relations dans l'espace

- + Un outil d'analyse impartial, sans a priori sur les utilisations ultérieures

- + Une façon de gérer les approximations dans le positionnement

Et sous réserve d'un peu de concertation :

- + Un zonage unique sur tout « le » territoire

- + Un zonage unique dans le temps

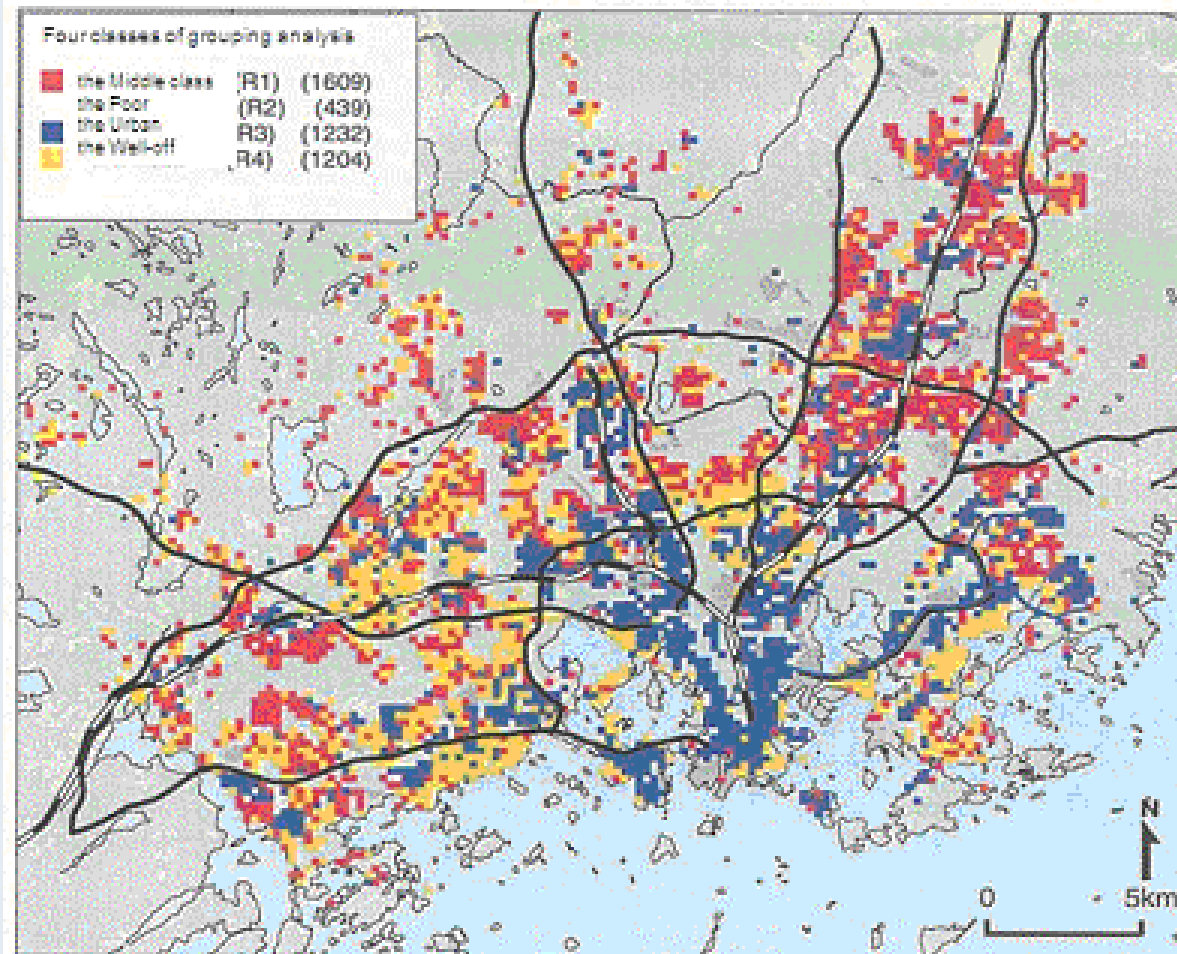
Et aussi (et surtout) :

- + Des possibilités de détail géographique bien au delà des zonages administratifs

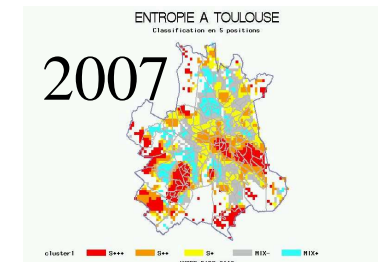


Les travaux européens et si on n'était pas seuls dans l'univers?

En Finlande



Data: Statistics Finland

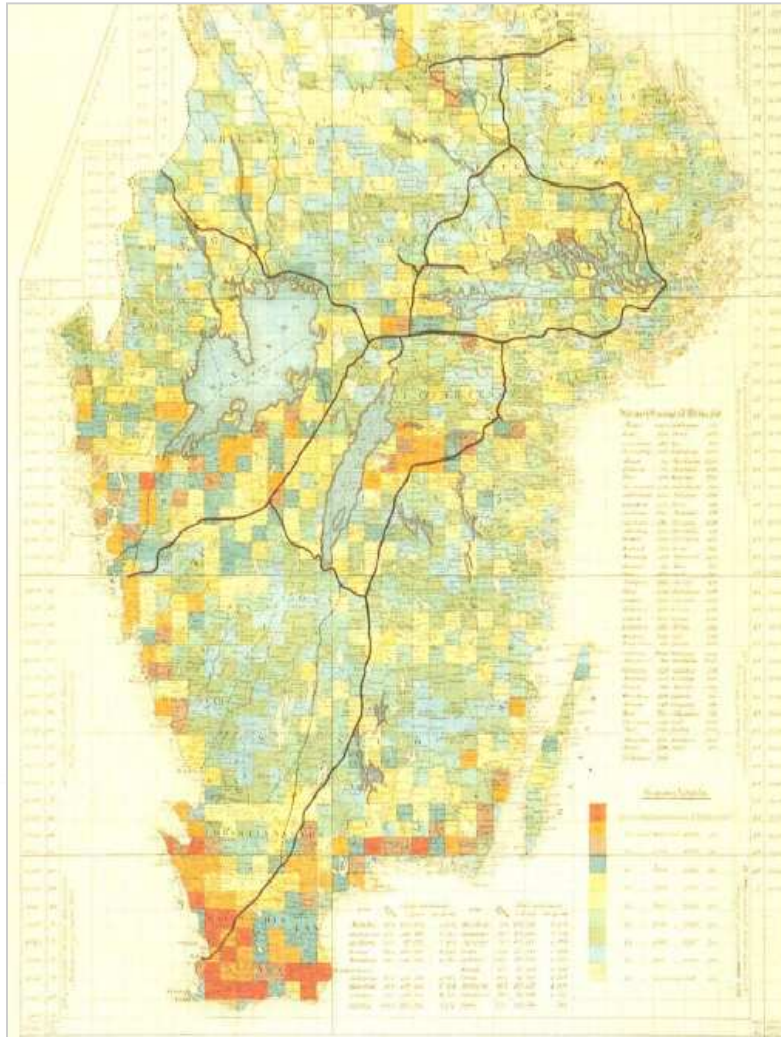


Ségrégation sociale au sein de l'aire métropolitaine d'Helsinki.

Classification automatique en 4 classes sur 20 variables (Vaattovaara 1998).



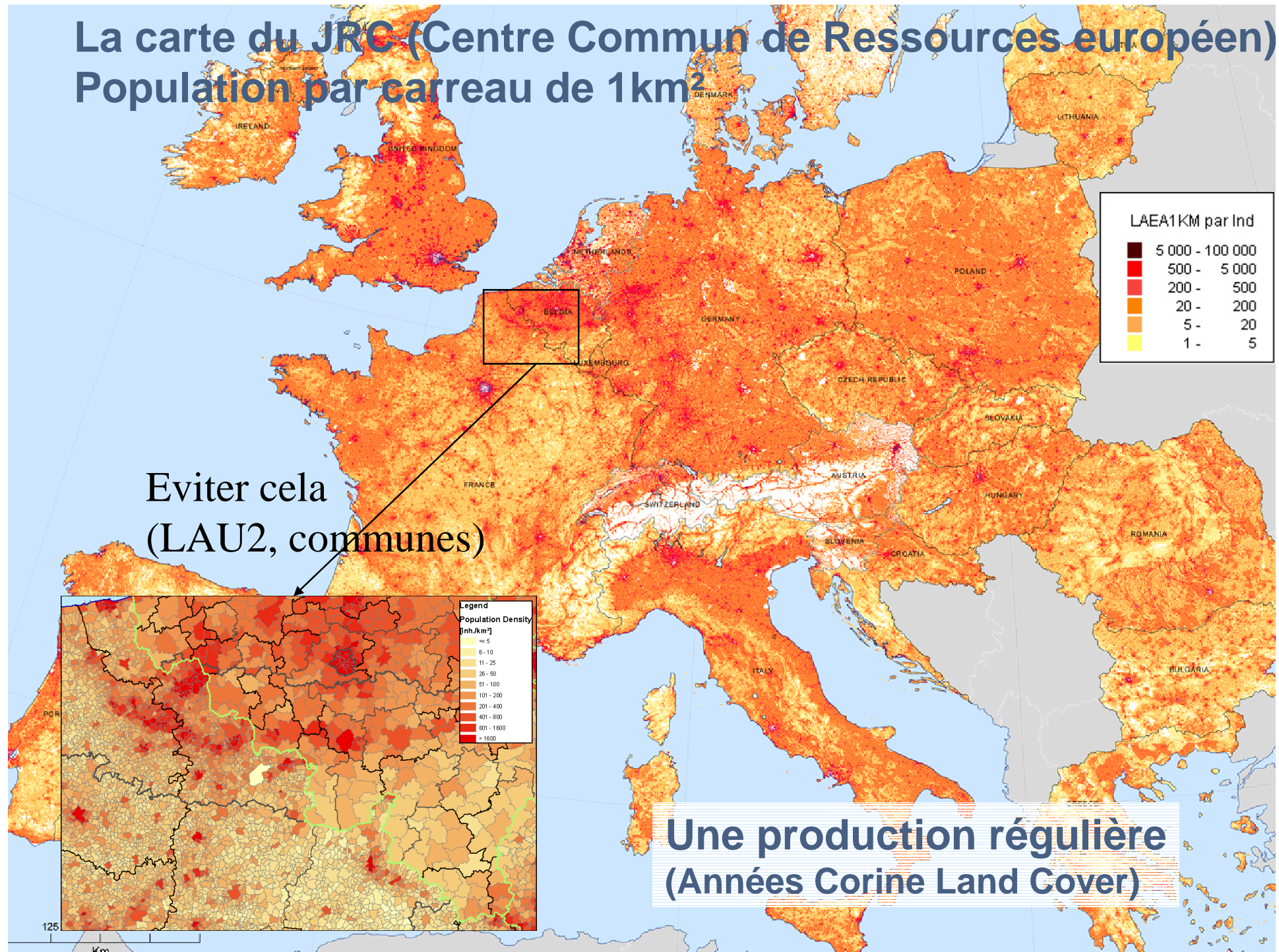
Rien de nouveau



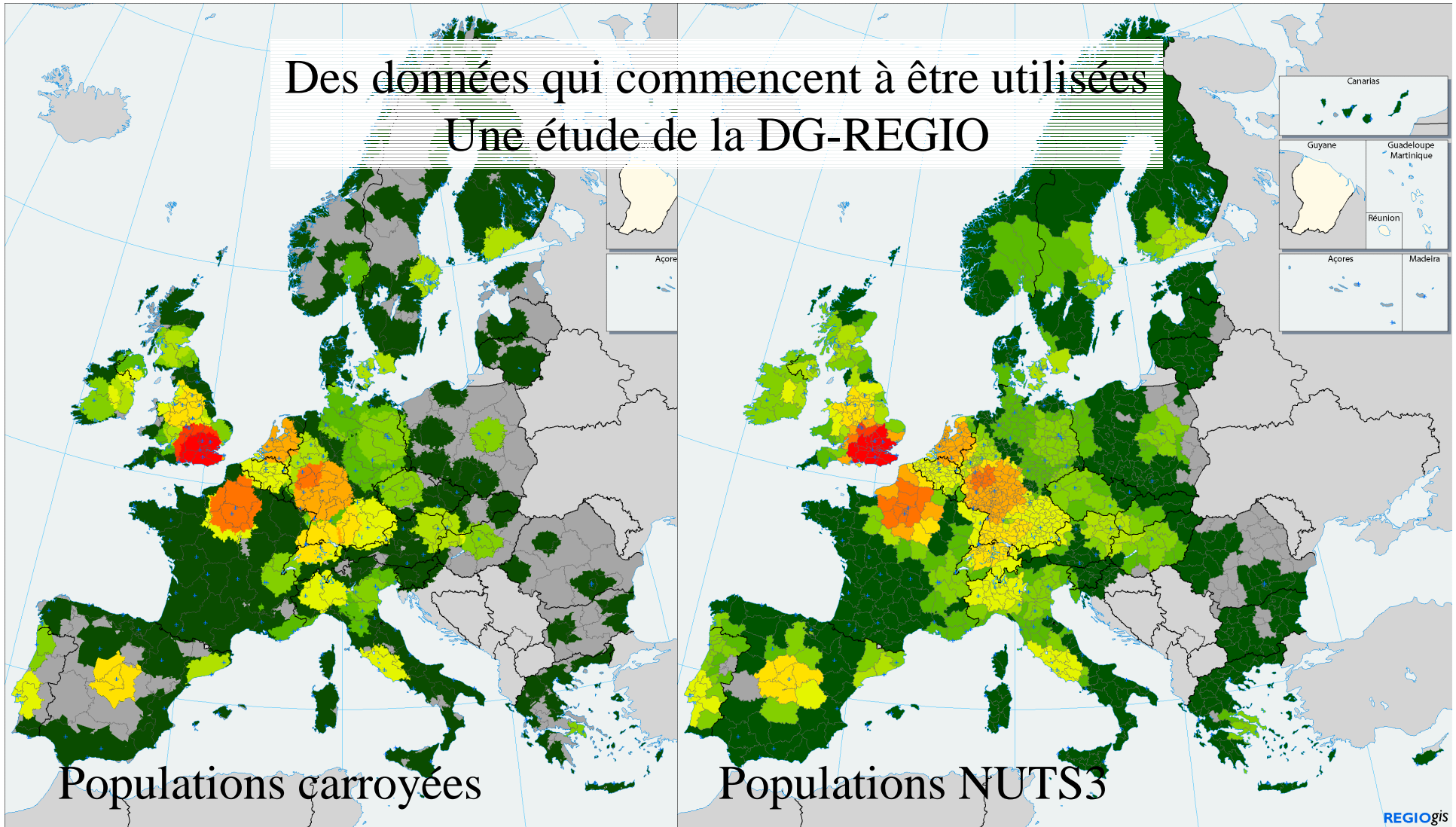
Sud de la Suède Densité de population 1857



La carte du JRC (Centre Commun de Ressources européen) Population par carreau de 1km²



Des données qui commencent à être utilisées Une étude de la DG-REGIO



Accessibility to long distance passenger flights at main airports, 2007

Cumulated daily number of passenger flights available within 90 minutes of travel by road to main airports

0	26 - 50	Travel time to nearest main airport > 90 min
1	51 - 100	No data
2 - 5	101 - 150	Airport
6 - 10	151 - 200	
11 - 25	201 - 250	

Long distance flights : > 5000 Km.
Sources: Eurostat, EuroGeographics, JRC, Statistics Finland, Statistics Sweden, REGIO-GIS.

0 1,000 Km

© EuroGeographics Association for the administrative boundaries

Accessibility to long distance passenger flights at main airports, 2007

Cumulated daily number of passenger flights available within 90 minutes of travel by road to main airports

0	26 - 50	Travel time to nearest main airport > 90 min
1	51 - 100	No data
2 - 5	101 - 150	Airport
6 - 10	151 - 200	
11 - 25	201 - 250	

Population weighted average at NUTS3 level.
Long distance flights : > 5000 Km.
Sources: Eurostat, EuroGeographics, JRC, Statistics Finland, Statistics Sweden, REGIO-GIS.

0 1,000 Km

© EuroGeographics Association for the administrative boundaries



Un autre exemple : Les villes selon la DG-REGIO

Des carreaux pour constituer des zonages

Audit urbain: **ville** = structure administrative gérant l'espace urbain. En France des EPCI.

- Cohérence avec la classification des NUTS?
- Où s'arrêter?

Méthode en cours d'expérimentation:

- présence d'une zone à haute densité de population (> 1500 habitants/km²)
 - Faisant abstraction de tout zonage : ensemble de carreaux contigus (hors diagonales)
données JRC à 1km (ou mieux)
- totalisant au moins 50 000 habitants

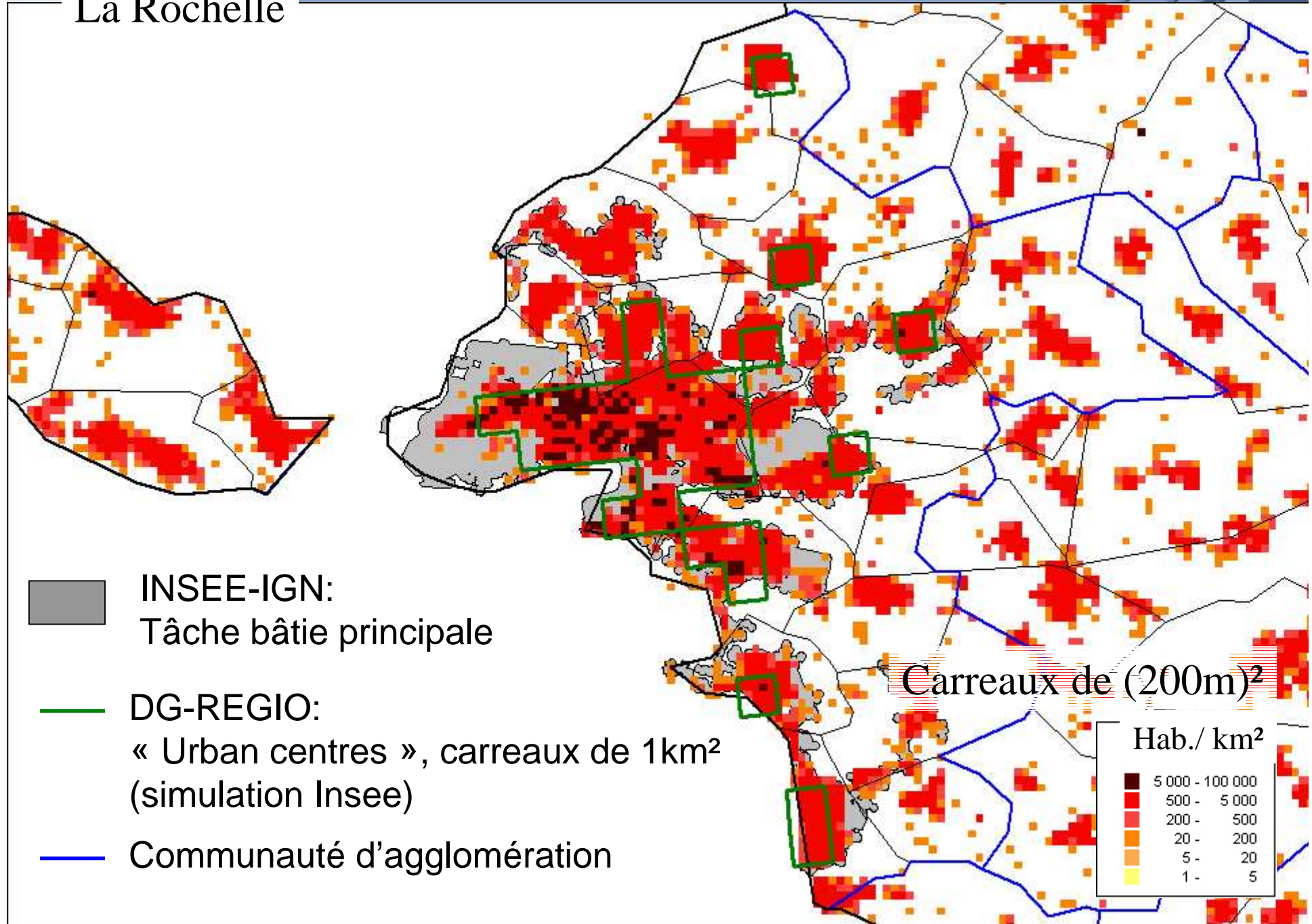
Extension en cours à une **classification des LAU2** (cf. nos unités urbaines)

- Type 1 : « urbain » majoritairement dans une zone à haute densité
- Type 2 : majoritairement dans une zone à densité intermédiaire (au moins 5000 habitants, entre 300 et 1500 habitants au km²)
- Type 3 : « rural »

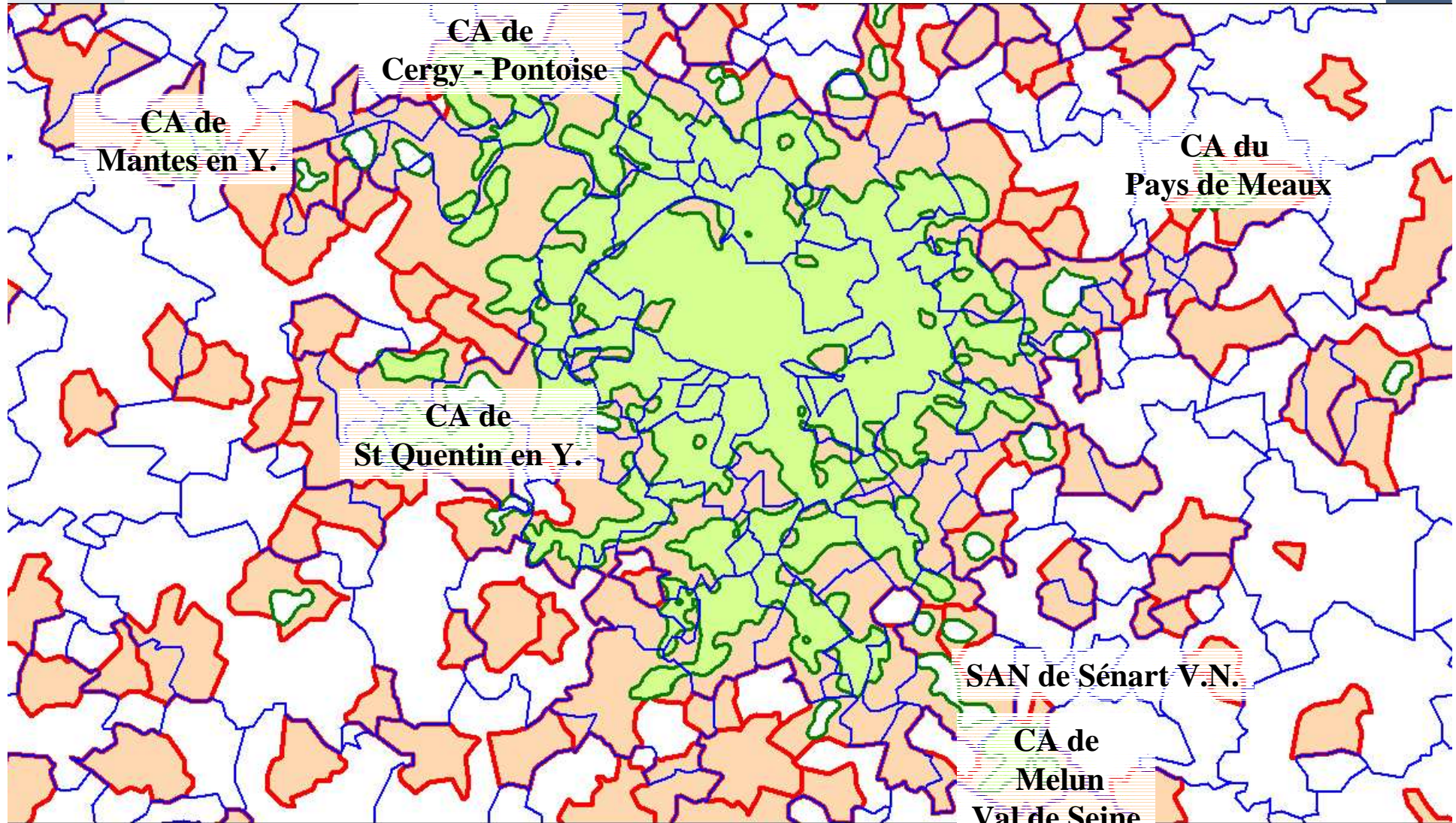
Page 20



La Rochelle



Paris : une structure à l'intérieur de la tache urbaine





L'ESSnet Geostat

Finalité

- ➔ Mettre l'ensemble des Etats Membres en mesure de diffuser des estimations de population par carreau de 1 km² à partir des recensements de 2010/2011.
- › Outputs attendus
 - Base de données (ou outil de diffusion compatible INSPIRE) avec indicateurs de qualité
 - Corpus méthodologique
 - Production : amélioration des méthodes de désagrégation
 - Diffusion : normalisation, traitement de la confidentialité
- › Geostat 1A : Travaux sur 2010-2011 avec financement d'Eurostat
<http://www.efgs.info/data/france/grided-data-from-the-census>
- › Geostat 1B : Prolongements en 2012-2013 : des partenaires supplémentaires, d'autres données, une meilleure résolution, une meilleure qualité

Page 23





Geostat

Données disponibles

› Geostat 1A

– Participants au projet : NO, FI, EE, AT, SL, PT, FR

– (PL)

– Contributions volontaires : UK(England+Wales)+SW+DK

Ensemble : 59 % du territoire continental et 44 % de la population;
résultat une carte très composite

– Contributeurs potentiels: CH, Northern Ireland, EI, ES, Lettonie,

› Geostat 1B

Objectif : harmoniser le processus, l'étendre à quelques variables
démographiques

– CZ

– (BG)

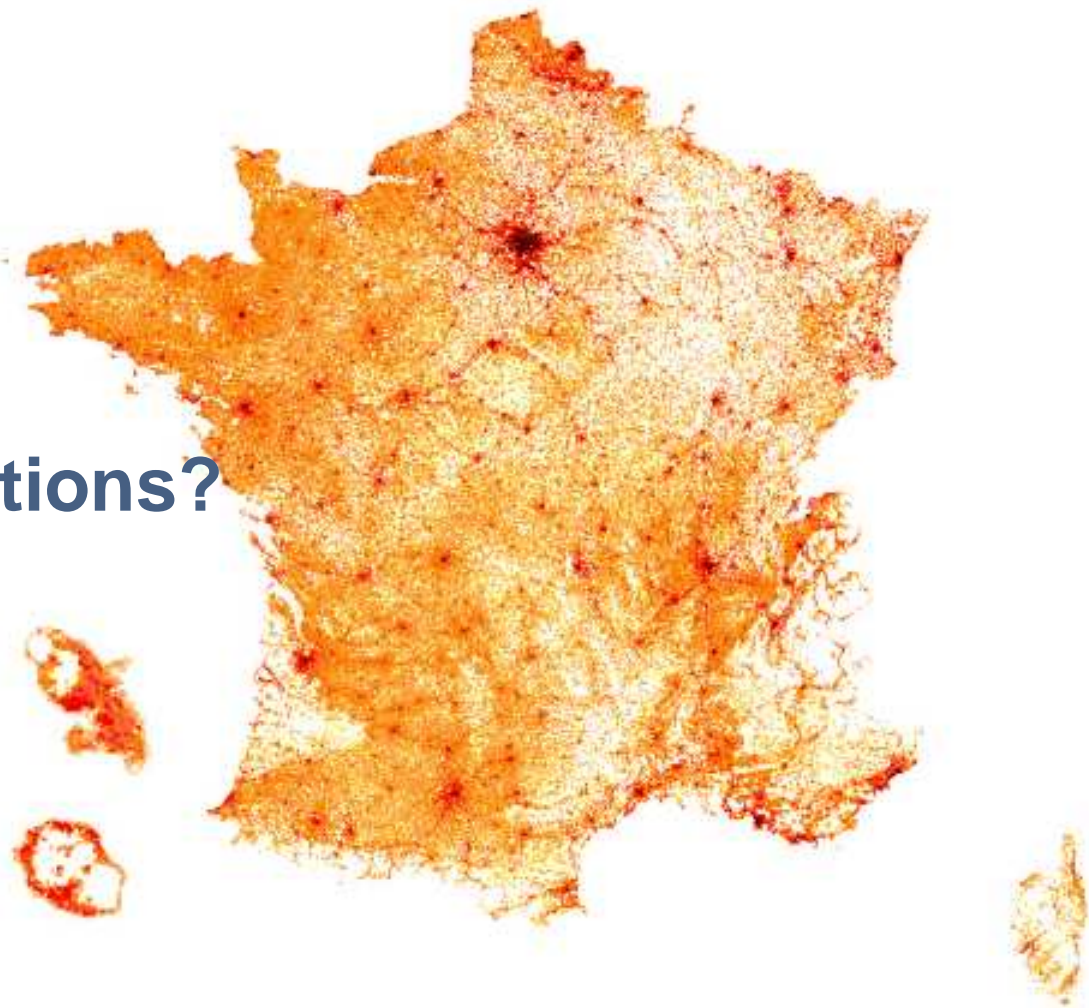
➔ Il reste une place pour les estimations

Page 24





Des questions?





Les données

Sources

Géoréférencement

Produits

Page 26





L'accès aux données

La contractualisation

- › Des points d'appui
 - CNIS : avis d'opportunité
 - Législation (loi Borloo), besoins de l'ONZUS
- › Un accord donnant donnant
 - Restitution aux partenaires de leur fichiers enrichis
 - Complément avec d'autres données, de l'assistance
- › Un cadre légal
 - Une convention entre les deux organismes
 - Une déclaration auprès de la CNIL
 - Un environnement de travail sécurisé



L'accès aux données

Nos partenaires

- › Impôts (DGFIP)
 - Impôt sur le revenu, taxe d'habitation, mais également cadastre et répertoire des ménages
- › Pôle emploi : chômage enregistré
- › Caisse nationale d'assurance maladie
 - Mutuelle complémentaire sous condition de revenus (CMUC)
 - Répertoire des médecins
- › Caisses d'allocations familiales : prestations sociales diverses
- › Education nationale : élèves
- › DARES : mesures d'aide à l'emploi
- › Direction des sports : licenciés et clubs
- ... Plus les ressources internes ailleurs à l'Insee
 - Répertoire des entreprises et des établissements
 - Base permanente des équipements
 - Déclaration des données sociales (sous convention) : emploi
 - Le recensement

Prestations
« gratuites »



Le spatial dans les données administratives

› Accès

- Convention avec l'organisme, généralement sur la transmission d'un fichier individuel sans les noms mais avec les variables d'intérêt et l'adresse postale

› Spatialisation

- Identification ou vérification/correction de la commune
- A partir des libellés d'adresse, commune par commune : transformation en références spatiales

- Application automatique
- Travaux manuels
- Redressements statistiques



Un chaîne à finalité statistique

- › Un environnement applicatif
 - Une application d'appariement automatique GeolocV2
 - Des bases adresse de référence

- › Des traitements manuels ciblés complémentaires
 - Sélectionnés en fonction des besoins

- › Des modules statistiques
 - De redressement des échecs de codage
 - De validation des résultats

- › ... et un réseau humain
 - Pilotage : pôle de localisation géographique (5 p. à Bordeaux)
 - Traitements manuels : équipes en directions régionales (20 ETP)
 - Traitements statistiques : pôle DSU (5 p. à Orléans)

GEOLOCV2 : Exemple de reprise (codages douteux)



APPLICATION DE GEOCODAGE - XP - - [GEOLOC - Menu reprise Voies (session=RP2)]

File View Tools Solutions Window Help

Commune Nb.Obs nîlot %ilot %iris %zon1 %zon2 %zon3 %adr VNapp
 97411- Saint-Denis 292 133 54.45 93.84 95.55 95.89 54.45 12

libellé de la voie sélectionnée = AV LECONTE DELISLE #97490#

Liste des libellés de voies

- DCV (12 obs)- AV LECONTE DELISLE #97490# // AV HUBERT DELISLE
- DCV (3 obs)- BD DE LA TRINITE #97400# // BD NOTRE DAME DE LA TRINITE
- DCV (2 obs)- R MONSEIGNEUR DE BEAUMONT #97400# // RUE MGR DE BEAUMONT
- DLV (2 obs)- R RUISSEAU BLANC #97417# // CHE DU RUISSEAU BLANC
- DCV (2 obs)- R JOSEPH BEDIER #97490# // RUE AMEDEE BEDIER
- DLV (1 obs)- ALL DES ECOLES #97400# // RLE DE L ECOLE
- DLV (1 obs)- R DE CAEN #97400# // RUE DE L AMIRAL DECAEN
- DCV (1 obs)- R GENERAL DE GAULLE #97400# // RUE DU GAL DE GAULLE
- DCV (1 obs)- R MGR MONDON #97400# // RUE MONSEIGNEUR MONDON
- DCV (1 obs)- R MONSEIGNEUR BEAUMONT #97400# // RUE MGR DE BEAUMONT
- DLV (1 obs)- RTE DE SAINT FRANCOIS #97400# // BD DE SAINT FRANÇOIS
- DLV (1 obs)- RTE DE ST FRANCOIS #97400# // BD DE SAINT FRANÇOIS

Traiter les adresses de la v

Nb Libellés	Rivoli	Idvoie	Libellé de voie CICN2
1	0415	(97411JLLR000943)	AV HUBERT DELISLE
	0432	(97411JLLR000758)	RUE DU CHER FRERE DENIS
Nb Obs Total	0010	(97411JLLR000151)	RUE DE L ABATTOIR
12	1107	(97411JLLR000267)	CHE DE LA PLAINE D AFFOUCHES
	1490	(97411JLLR000152)	RUE DE L AFRIQUE
Rechercher	0455	(97411JLLR000233)	ALL DE LA FLUTE SAINT ALEXIS
DELISLE	????	(97411JLLR000157)	RUE DE L AMITIE
	0094	(97411JLLR000159)	RUE DE L ANJOU
Rechercher FMZ	0110	(97411JLLR000161)	RUE DE L ARTILLERIE
	0117	(97411JLLR000163)	RUE DE L AVENIR
	????	(97411JLLR000136)	RUE DE BAILLY
Rechercher AMZ	0679	(97411JLLR000939)	ALL HONORE DE BAI ZAC

GEOLOCV3 : La nouvelle application

Affichage Colonne Recherche

	Depcom	Libellé Voie	Libellé Apparié	Validé	R1	nb obs
Non-Validé <input type="checkbox"/>	<input checked="" type="checkbox"/> 56260	R DOCTEUR JOSEPH AUDIC		N	XXX	224
	<input type="checkbox"/> 56260	R DU DOCTEUR JOSEPH AUDIC		N	XXX	5

Confirmer Invalider Associer Stand by Exclure

Echos du referentiel

	Depcom	Libellé Voie	Rivoli	Score
<input type="checkbox"/>	56260	ALL FRANCOIS JOSEPH BROUSSAIS	1156	9
<input type="checkbox"/>	56260	R CHARLES JOSEPH LE SUEUR	0656	9
<input type="checkbox"/>	56260	CAR GIRATOIRE DOCTEUR GROSSE	0899	6
<input type="checkbox"/>	56260	IMP JOSEPH LOTH	1661	6
<input type="checkbox"/>	56260	PL JOSEPH LE BRIX	1640	6

Valider Valider carreau A expertiser

Retour accueil

Rue : R DOCTEUR JOSEPH AUDIC

Recherche OpenLs

voie	commune	CCP	x	y	p	typ
R DOCTEUR JOSEPH AUDIC	Vannes	S6000	-2.774220	47.680446	100	voie

<input type="checkbox"/>	Depcom	Num	Libelle voie	Libelle voie Referentiel	Id Voie App	Complemen Adresse	Code Postal	Type	Id
<input type="checkbox"/>									

ajouter voie supprimer voie ajouter carreaux supprimer carreaux modifier carreau associer carreau

Problème : Voie inconnue

Recherche Géoportail, GoogleMaps, openstreetmap

Localisation carroyée

Enrichissement du référentiel

Terminé





La géolocalisation à tout prix

Réutiliser les outputs de GeolocV2

- › Les éléments de géocodage contiennent une part de géographie **approchée** (tous calculs hors GEOLOC)
 - Appariement réussi avec le RIL Natif : 85 %
 - Codage ilot et voie (une façade d'ilot) | Barycentre : 10 %
 - Codage ilot seul
 - Codage IRIS (ou ZUS) et voie
 - Codage IRIS (ou ZUS) seul
 - Codage voie seule
 - Pas mieux que la commune
- Imputation non déterministe sur un carreau : 5 %

Données carroyées : cumuls sur **arrondi à 100 m** des coordonnées Lambert II (en fait pas de zonage)





La question des bases adresses

Préférence aux produits maison

› Le RIL en majeur

Environ 900 communes

- Géolocalisation à la plaque adresse
- Géocodage sur quelques zonages de base (IRIS, ZUS, ZFU)
- autres zonages par superposition de fonds de carte

› des compléments disparates...

Environ 300 communes avec des référentiels réduits

- ne permettant que de faire du géocodage
- parfois limités aux seules ZUS
- sans mise à jour régulière

...qui sont remis en cause

par les nouveaux référentiels « G3 » issus de l'appariement cadastre-fichiers fiscaux sur 35 000 communes

- Géolocalisation à la parcelle
- pas de géocodage : à faire par superposition de fonds de carte



La situation mi Juillet 2011

- › Avenant à la convention avec la DGFIP, portant sur une **livraison annuelle** de l'ensemble du cadastre vectorisé ou non (image de ce qui est mis en ligne sur www.cadastre.gouv.fr .
 - ➔ 2ème livraison en mai-juin 2011
- › Mise à disposition gratuite de la BD parcellaire de l'IGN
 - ➔ téléchargée en mai 2011
- › Après transformation des fichiers, trois cas de figure :
 - Cadastre vectorisé : 22 000 communes
 - Cadastre images avec tables de passage coordonnées locales vers coordonnées géographiques : 14 000 communes
 - Pas de cadastre : Îles de Sein, de Molène, essentiel de la Guyane, Mayotte



L'utilisation du cadastre à fin de géolocalisation

- › Fichiers fiscaux : injection directe de coordonnées par chaînage de trois tables de passage

Cadastre : identifiant parcelle -> coordonnées

TH : identifiant logement -> identifiant parcelle

IRPP : identifiant contribuable -> identifiant logement

→ Seul le chaînage Cadastre – TH est imparfait

Millésimes différents : aujourd'hui, cadastre trop frais!

Erreurs de référence cadastrale dans la TH (0.1 % sur le test de Reims)

Références cadastrales pas encore présentes

0,4 %
(RFL 2009)

- › Autres fichiers administratifs : le même chaînage permet de construire un fichier de libellés d'adresses géolocalisées (source TH et IRPP).

– → Matière première pour GEOLOC modulo une transformation en référentiels « G3 » générés en juin 2011, sur **la plupart des communes**

– Le processus inclura l'ensemble des adresses bâties et non bâties de la DGFIP à partir de mi 2012.

Page 36

- › Le recensement par contagion...





Histoires d'AMZ

ilot iris zus

0.txt - Bloc-notes

Référentiel RIL

Format Affichage ?

1730000000000001R	ABBE GREGOIRE	1	VILLENEUVE LES SALINES	ES17	0902	HORSZONE
1730000000000001R	ABBE GREGOIRE	2	ZONE VILLENEUVE LES SALINE	ES16	0902	HORSZONE
1730000000000001R	ABBE GREGOIRE	2	FIEF DE VILLENEUVE	ES16	0902	HORSZONE
1730000000000001R	ABBE GREGOIRE	3		ES17	0902	HORSZONE
1730000000000001R	ABBE GREGOIRE	4		ES16	0902	HORSZONE
1730000000000001R	ABBE GREGOIRE	5	FIEF DE VILLENEUVE	ES17	0902	HORSZONE
1730000000000001R	ABBE GREGOIRE	6		ES16	0902	HORSZONE
1730000000000001R	ABBE GREGOIRE	7	ZAC DE VILLENEUVE	ES17	0902	HORSZONE
1730000000000001R	ABBE GREGOIRE	8	VILLENEUVE LES SALINES	ES16	0902	HORSZONE
1730000000000001R	ABBE GREGOIRE	9	ZA LES FIEFS DE VILLENEUVE	ES17	0902	HORSZONE
1730000000000001R	ABBE GREGOIRE	10		ES16	0902	HORSZONE
1730000000000002R	ABRAHAM DUQUESNE	2		ES17	0902	HORSZONE

Des zonages et des x,y

0.txt - Bloc-notes

Référentiel G3

Format Affichage ?

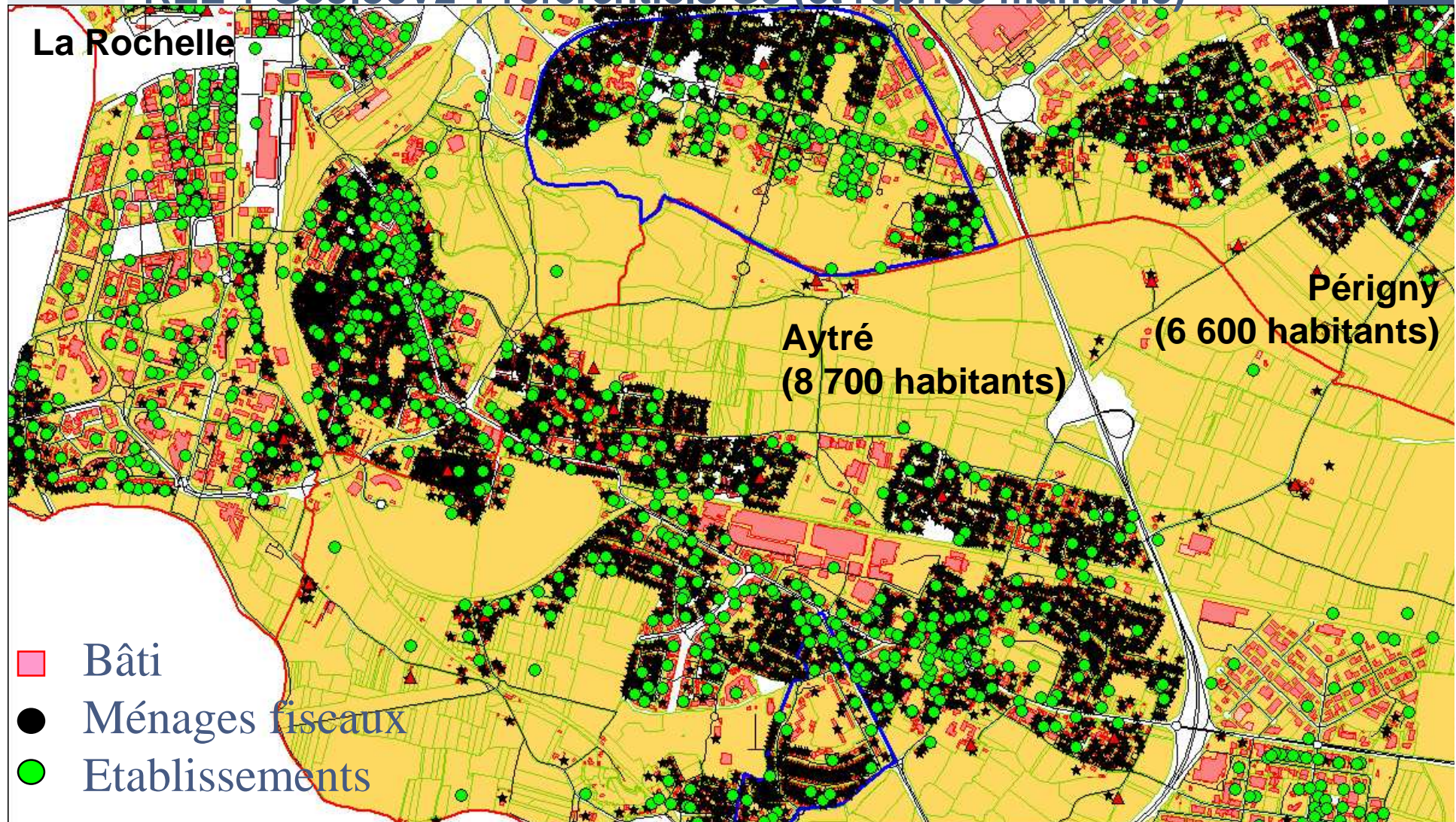
17300JLL0000000	VERSION	.	ADR2010 FT2009 G=CADASTRE P=L93	3.07.1
17300JLL0000001R	ABBE GREGOIRE	9999		101-028
17300JLL0000002R	ABRAHAM DUQUESNE	0002		030-039
17300JLL0000002R	ABRAHAM DUQUESNE	0004		030-039
17300JLL0000002R	ABRAHAM DUQUESNE	0006		030-039
17300JLL0000002R	ABRAHAM DUQUESNE	0007		030-041
17300JLL0000002R	ABRAHAM DUQUESNE	0009		030-041
17300JLL0000002R	ABRAHAM DUQUESNE	0011		030-041
17300JLL0000002R	ABRAHAM DUQUESNE	0013T		030-041
17300JLL0000002R	ABRAHAM DUQUESNE	0015		030-041
17300JLL0000002R	ABRAHAM DUQUESNE	0036		030-041
17300JLL0000003R	DE L ABREUVOIR	0001T		030-041

Des x,y, des carreaux et rien d'autre



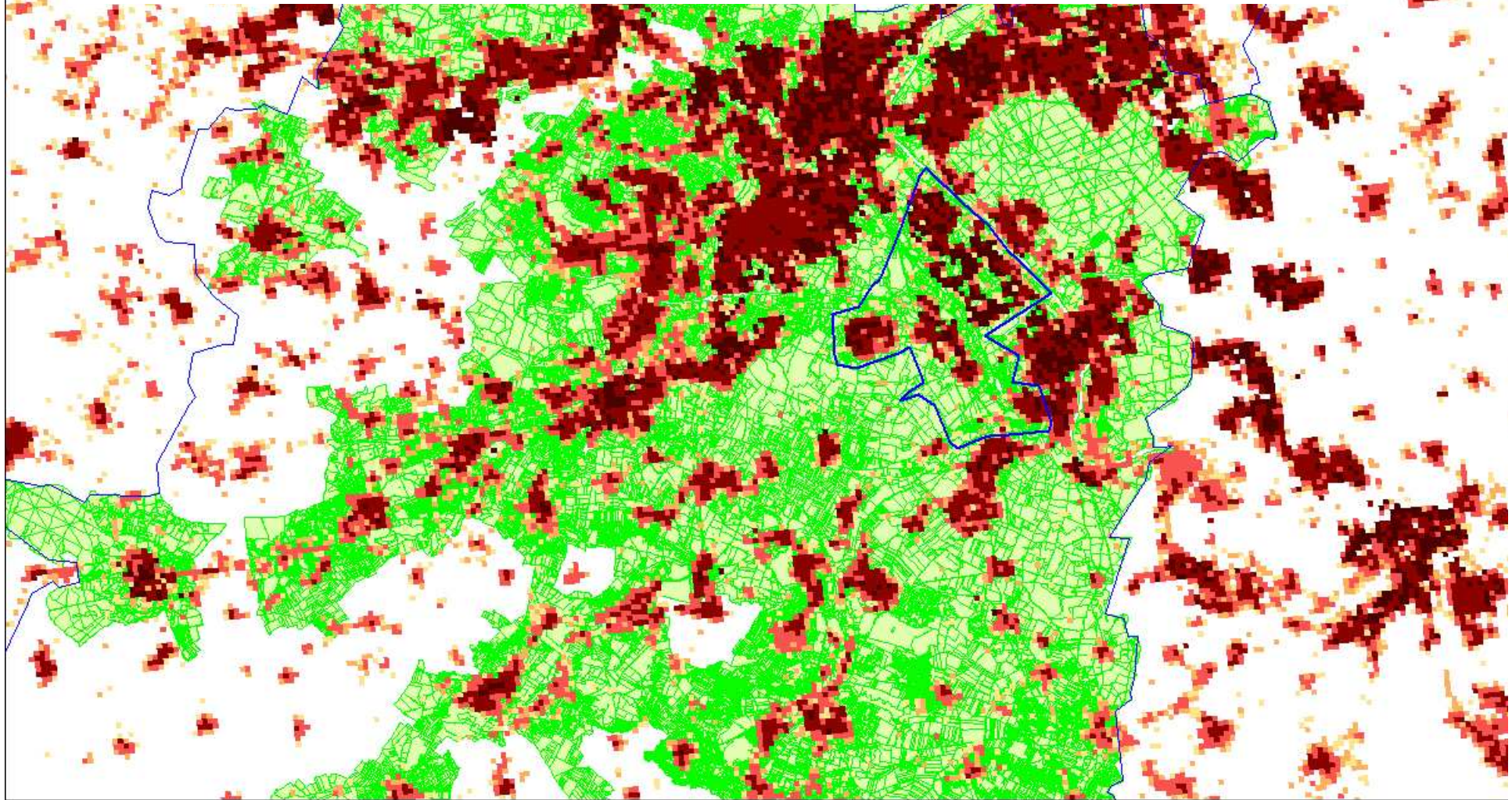
Le futur est déjà là

REE + GeolocV2 + référentiels G3 (et reprise manuelle)



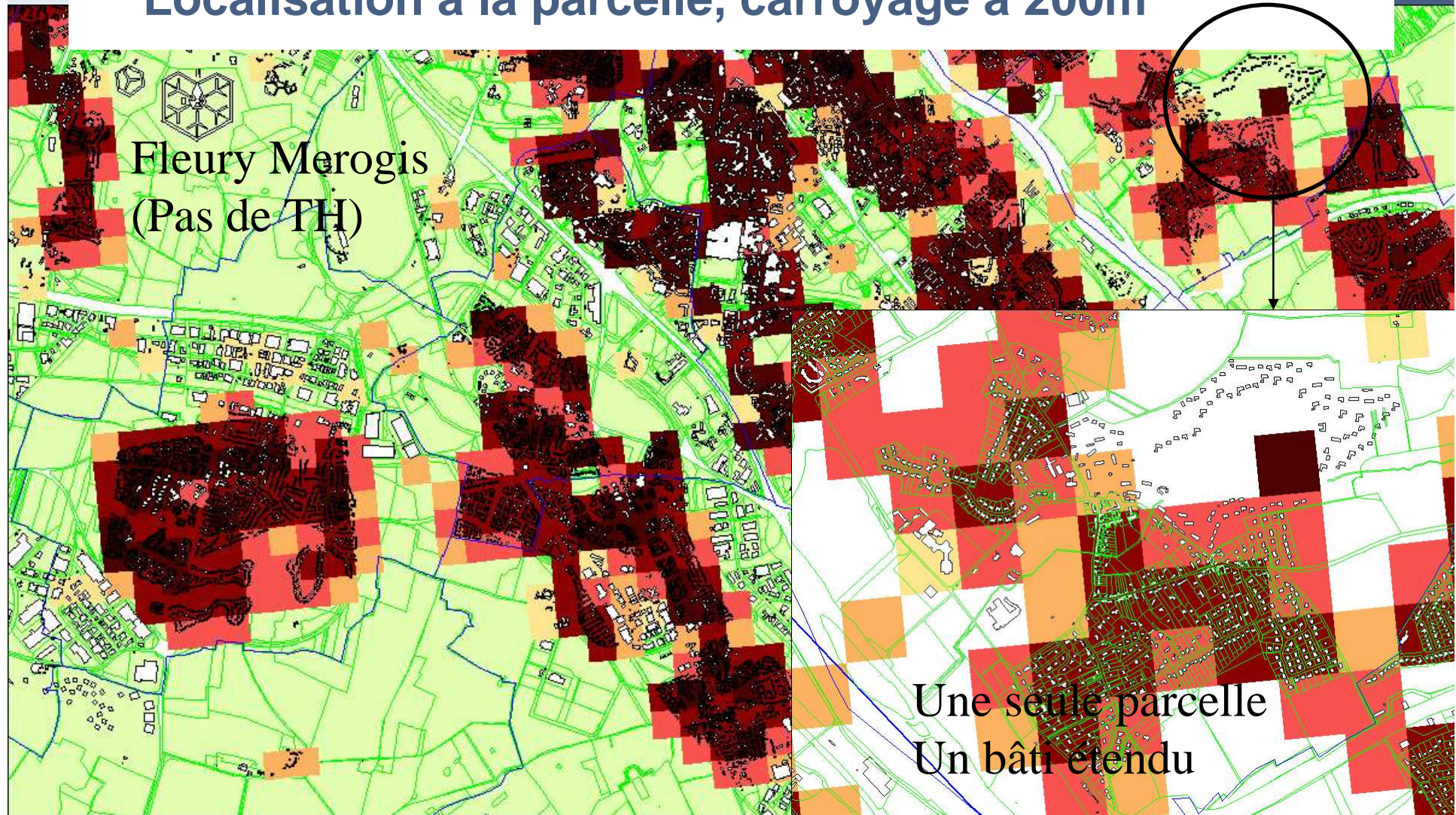
Département de l'Essonne

La non vectorisation n'est pas un obstacle



Populations fiscales : Les limites de l'exercice

Localisation à la parcelle, carroyage à 200m



Les limites de l'exercice : C'est quoi une adresse?

Saint-Michel sur Orge 91570

RIL, cadastre et carroyage à 200m





Conséquences sur l'utilisation des données

Un carroyage est plus une géolocalisation avec un positionnement approximatif qu'un géocodage sur un zonage particulier

Les données carreau par carreau n'ont pas forcément de signification précise

Les données doivent être interprétées via des procédures statistiques impliquant des ensembles spatiaux plus larges

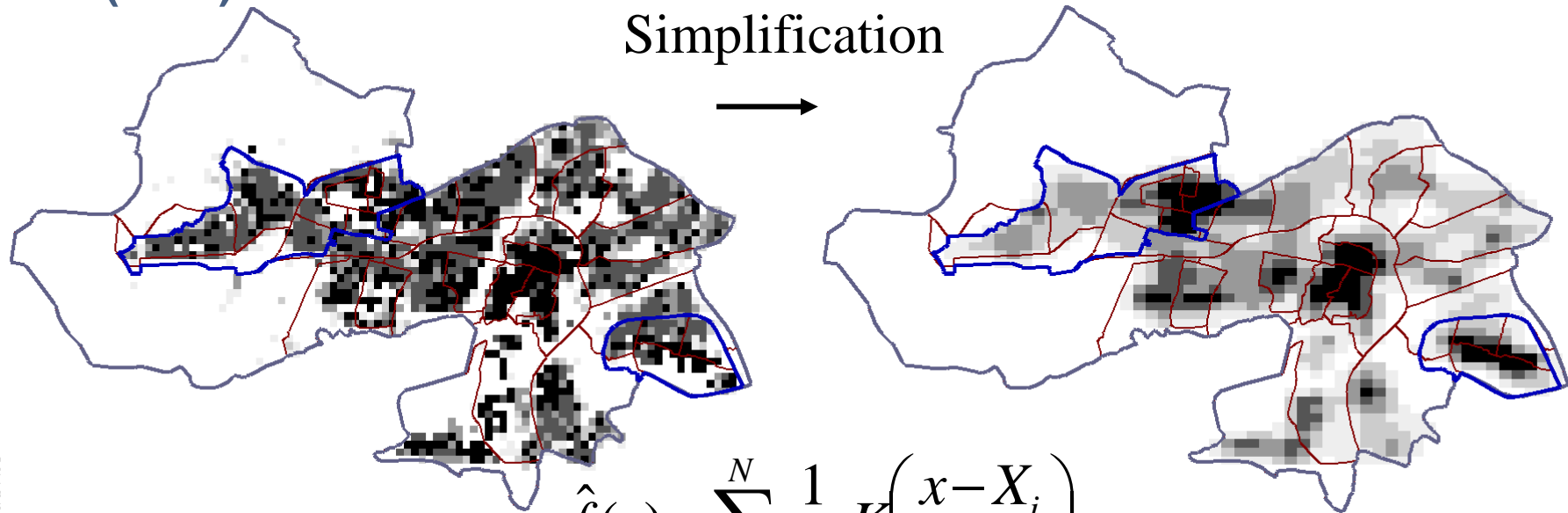
→ Une diffusion carroyée s'apparente à la diffusion d'une image : la couleur précise de chaque pixel n'est pas importante (cf. JPEG)

(NB: on est exactement dans la même configuration avec les données RP à l'IRIS, mais pour des raisons d'approximation statistique, non géographique)



Des données fines à un zonage

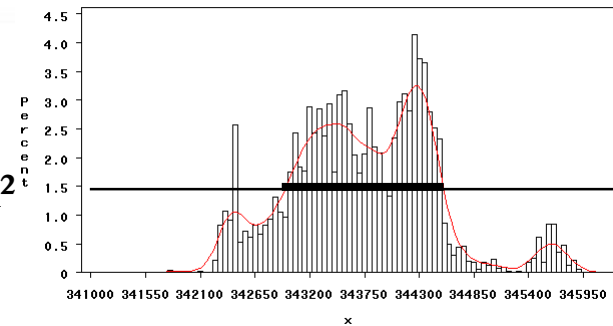
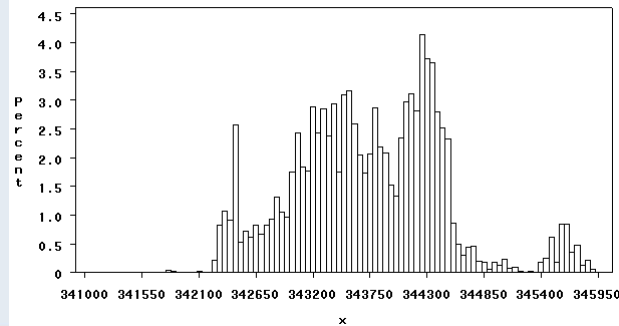
L'estimation non paramétrique de densité de probabilité (KDE)



$$\hat{f}(x) = \sum_{i=1}^N \frac{1}{Nh} K\left(\frac{x - X_i}{h}\right)$$

$h=200\text{m}$

1500 hab/km^2



Analyser une sous population

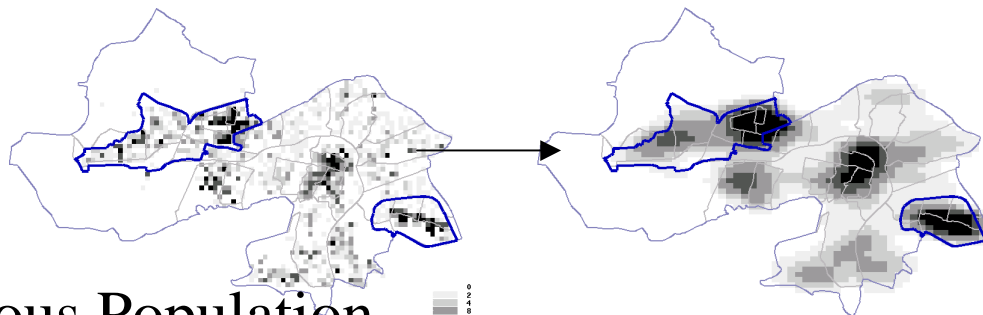
Zones à enjeu

Principe de cartographie

100m

200m

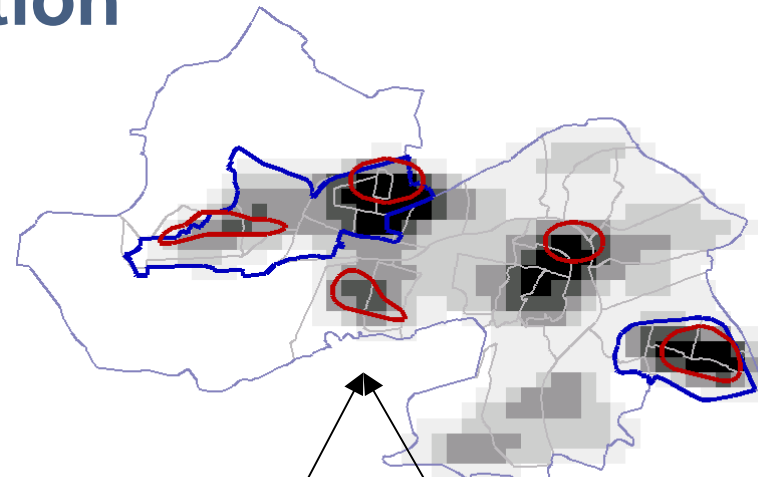
1) *Simplification* : f_{SP} , f_{PT}



Sous Population

« en difficulté »

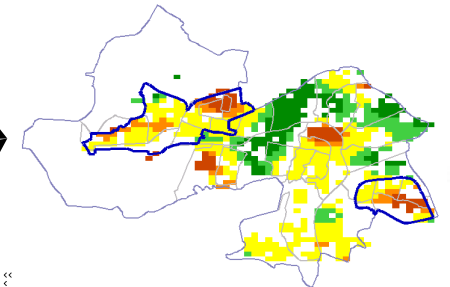
Population Totale



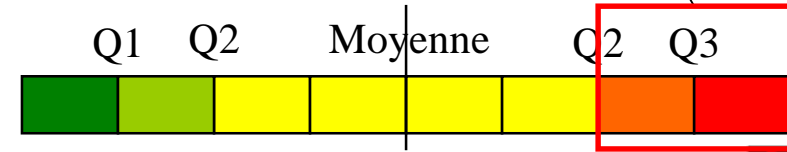
3) *Extraction des contours*
Superposition

f2003_17300_xy - Surreprésentation br (MEDIAN=1.14, MIN=100)
Carroyage 200x200m. (c) Insee 2006

2) f_{SP}/f_{PT}



ca12003_17300_xy - Surreprésentation br/atlac 0.67 << >> 1.22



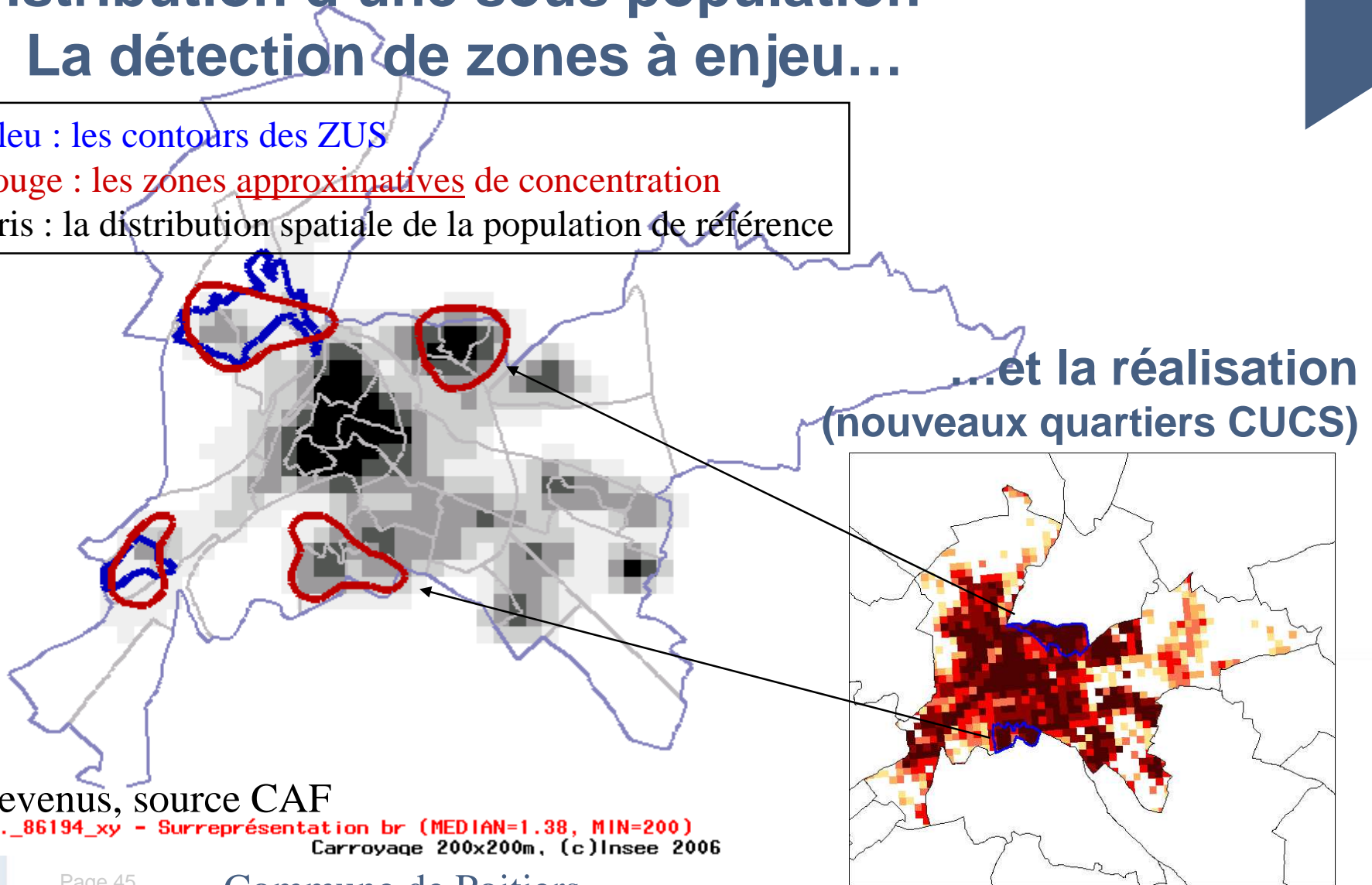
Distribution d'une sous population

La détection de zones à enjeu...

En bleu : les contours des ZUS

En rouge : les zones approximatives de concentration

En gris : la distribution spatiale de la population de référence



Bas revenus, source CAF

caf2003._86194_xy - Surreprésentation br (MEDIAN=1.38, MIN=200)
Carroyage 200x200m, (c)Insee 2006

Page 45

Commune de Poitiers





Le recensement

Page 46



Jean-Luc LIPATZ

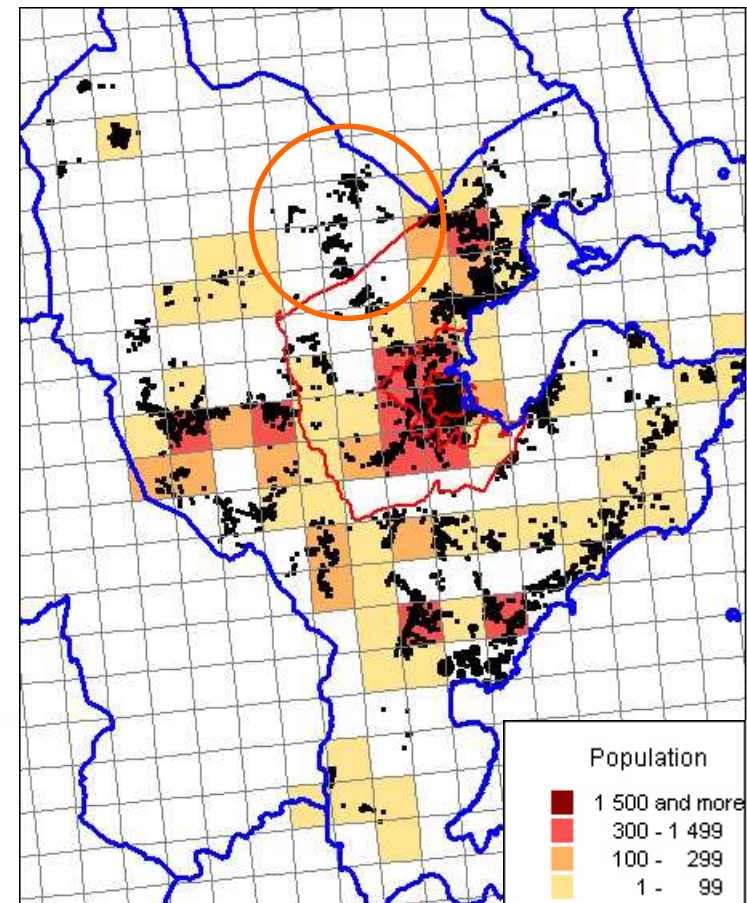
CRIS Ile de France

15 mars 2012



Acte I – les plus de 10 000 habitants

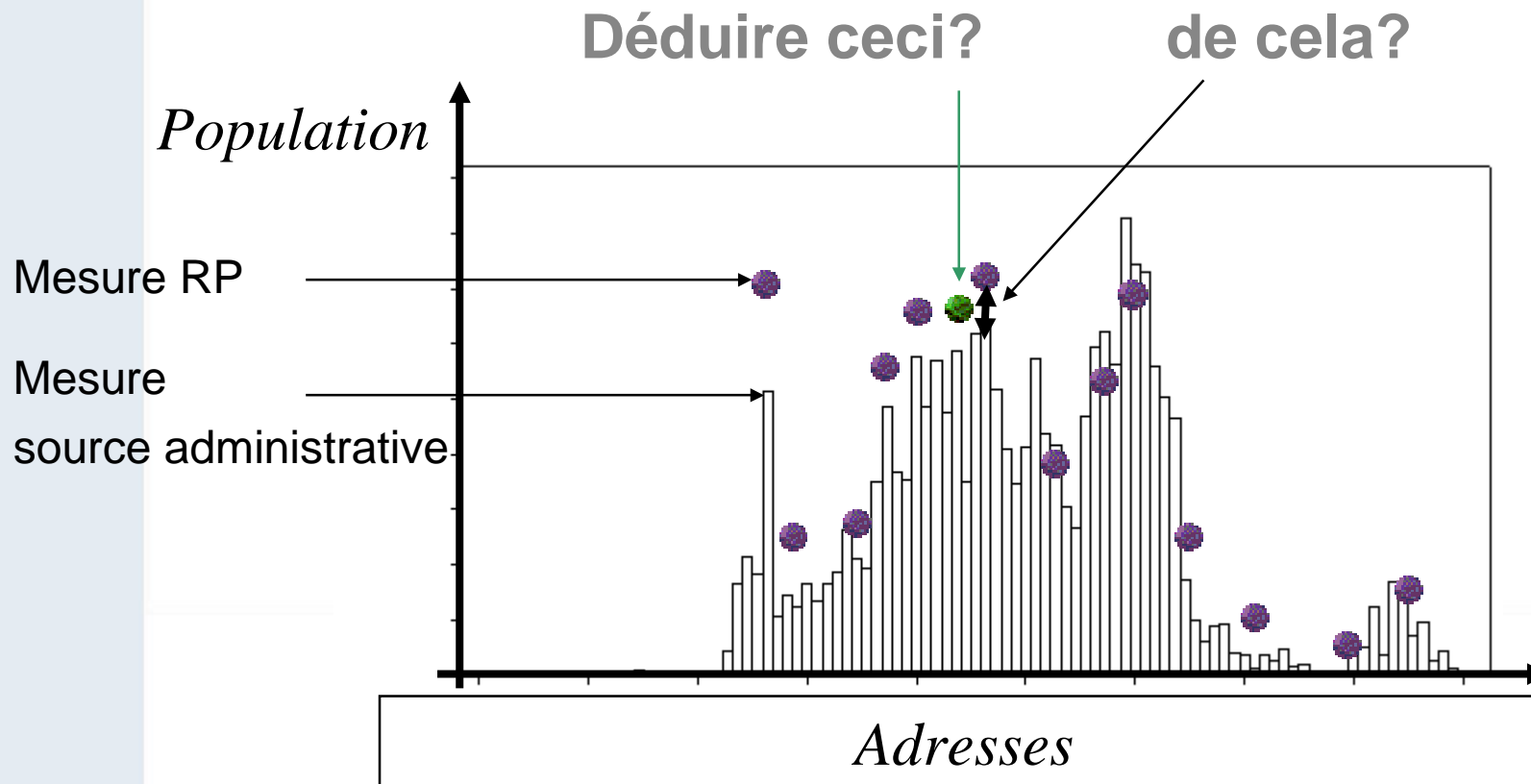
- › L'échantillon ne garantit pas une couverture spatiale
 - ➔ IRIS et zones DIAF-RP (pas toujours)
- › Mais on dispose de beaucoup d'information auxiliaire géolocalisée
 - RIL
 - Données administratives



Carreaux de 1km²
Estimation mode DIAF-RP



Remplir les trous de collecte corriger les biais administratifs et temporels





Des régressions locales

L'espace considéré comme une variable explicative

$$Pop_{RP} = \alpha(s).Pop_{Adm} + \beta(s).Logements + \varepsilon$$

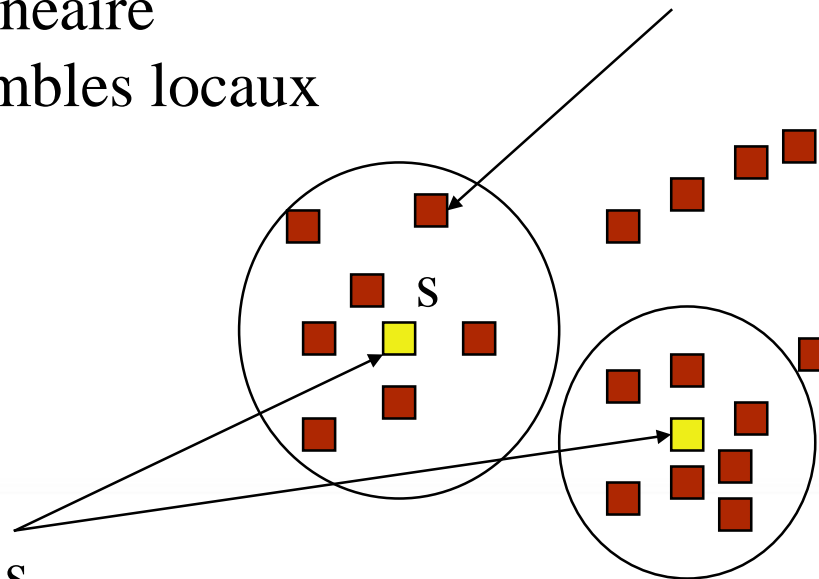
Régressions linéaire
sur sous-ensembles locaux

$\alpha(s), \beta(s)$

Estimations

$$\hat{Pop}_{RP}(s) = \alpha(s).Pop_{Adm}(s) + \beta(s).Logements(s)$$

Pondérations décroissantes
avec la distance



Pop_{Adm} : CNAM/TS ou RFL



Acte II – les moins de 10 000 habitants

Des données fiscales aux données RP

Deux sources parlant de la même réalité

Avec des interprétations / descriptions différentes :

- Qualité (RP : valeurs manquantes, fisc : défauts d'enregistrement)
- Etudiants (RP : au lieu de résidence, fisc : dans leur famille)
- Enfants (RP : données complètes, fisc : seulement l'âge)
- Résidence secondaires
- Notion d'immeuble
- Personnes hors ménages ordinaires

Mais 95 % des enregistrements parlent des mêmes personnes

→ Importer la localisation depuis les données fiscales



Le processus

- 1) Appariement flou des deux fichiers individuels en utilisant :
 - La commune de résidence
 - Le sexe
 - La date de naissance
 - La commune (éventuellement seulement le département) de naissance

➔ Appariements multiples
- 2) Calcul d'un score individuel
 - Qualité de la reconnaissance du lieu de naissance?
 - Enregistrement administratif correspondant à une résidence secondaire, à autre chose qu'une maison ou un appartement?
- 3) Calcul par agrégation d'un score au niveau immeuble RP
- 4) Choix de l'appariement le plus probable, invalidation des appariements improbables
 - Le recensement est collecté par grappes
- 5) Interpolation des appariements invalidés et des échecs d'appariement
 - Entre deux enregistrements RP voisins dans le fichier, lorsqu'ils ont été identifiés dans la même rue



Les résultats

Rien n'est parfait

- Le lieu de naissance n'est parfaitement connu d'aucune des deux sources : intégré au calcul du score.
- La date de naissance des personnes âgées est parfois mal saisie
- Les données manquantes du recensement sont comblées par clonage : impossible à corriger, les ménages correspondants sont placés au barycentre du district de collecte pour..

→ Mais le résultat est plus que satisfaisant

Département de la Vienne, communes de moins de 10 000 habitants, population des ménages ordinaires

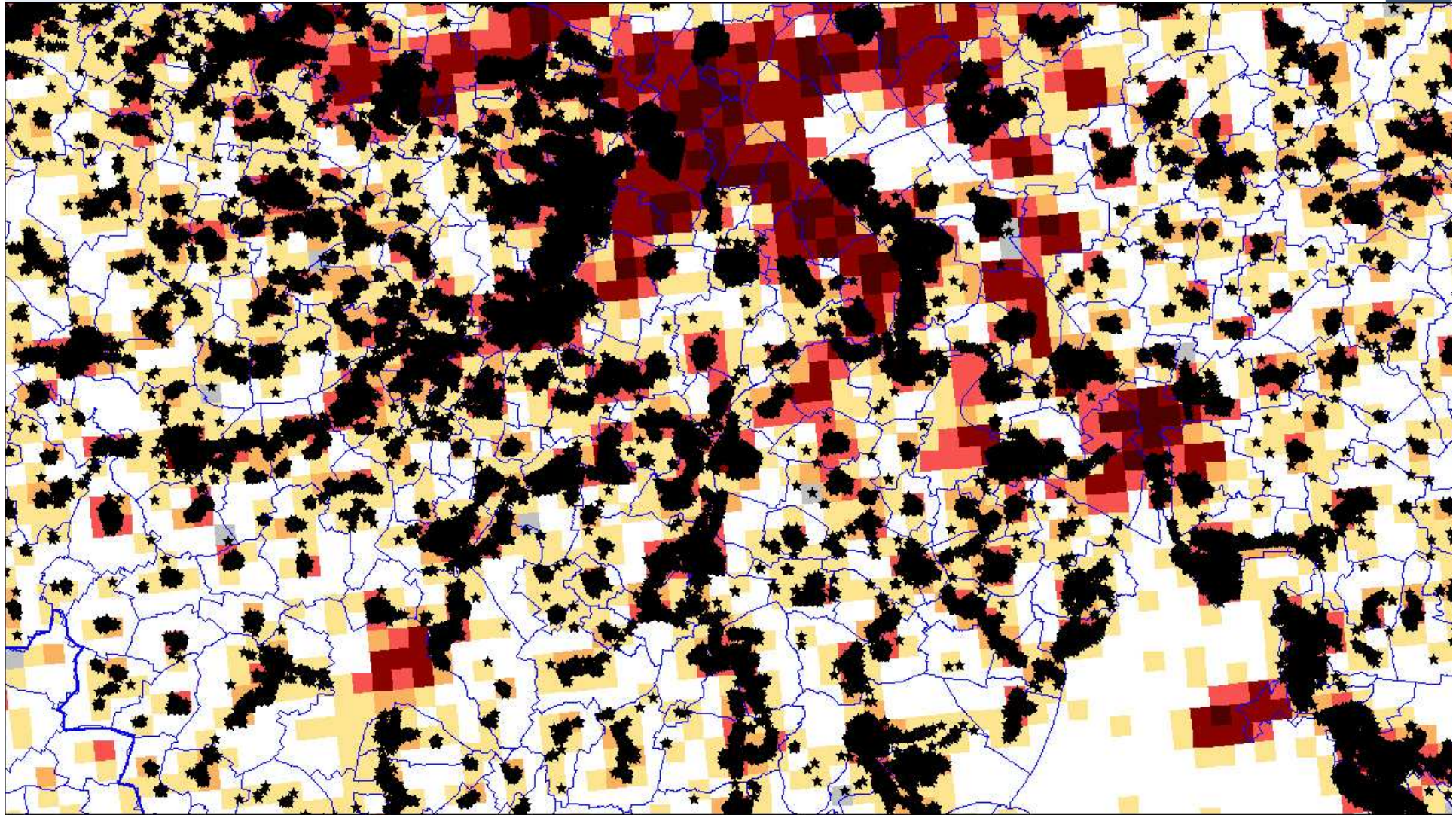
Positionné au barycentre	0.6 %
Résultant du repérage de l'immeuble	97.4 %
Interpolé	2.0 %

Page 52



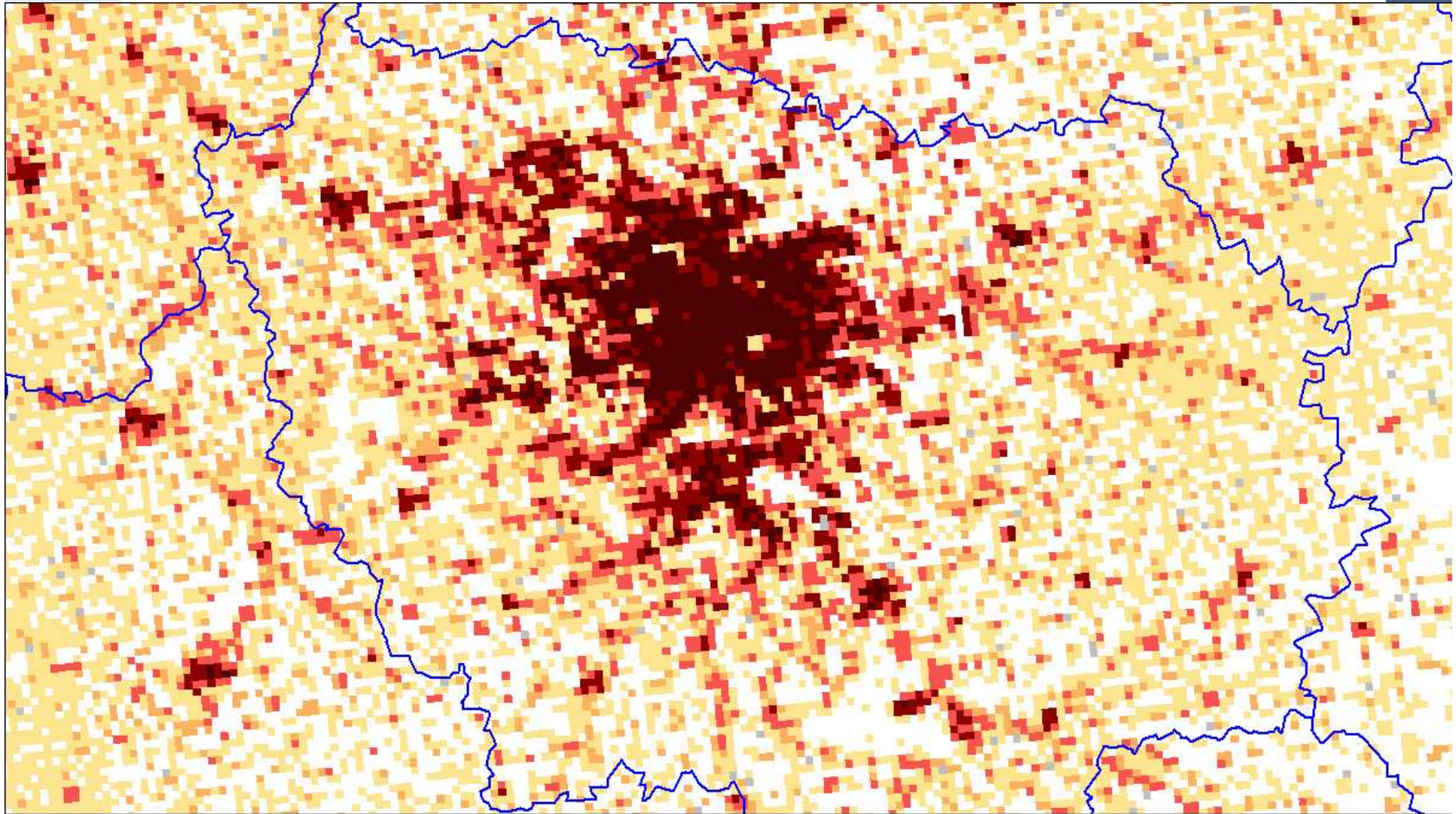


Des carreaux pour réconcilier les approches





...à consommer sans limites





Au final

Un dispositif à deux vitesses

- › Du standard : champ géographique cible = France entière, données géolocalisées, recours à Geoloc limité
 - RFL | opérationnel
 - RP |
 - *SIRENE* |
 - *BPE* | travaux en cours
 - *CLAP* |

- › Du spécifique : champ géographique restreint, pas forcément de géolocalisation, recours à Geoloc obligatoire
 - *CAF* (en attendant le rattachement à RFL)
 - DADS accent migrations domicile – travail
 - CNAM, DEPP pour les besoins politique de la ville
 - DARES, Sports dans le cadre de prestations
 - Naissances



La diffusion de données carroyées

Préconisation 6 : mettre en place les conditions qui permettront aux utilisateurs de rassembler des **informations « clés »** issues de sources diverses selon une même maille géographique, le « carreau » ; ceci **afin de les agréger et de les analyser** sur leurs propres terrains infracommunaux d'observation et d'action.



La diffusion de données carroyées

Trois questions

- 1) Le choix de la **géographie** : il y aurait du y avoir concertation, mais...
 - ... l'Europe pose le LAEA comme standard de projection
 - ... l'INSEE est parti sur 200m de coté, standard de fait (AUs)
- 2) Le choix des **indicateurs** : il pourrait y avoir concertation mais...
 - ... l'objectif n'est que de diffuser quelques indicateurs de base, pour une lecture sur un territoire plus étendu
 - ... la production à grande échelle ne fait que démarrer : la liste des indicateurs sera étendue en fonction des possibilités et besoins

Grande échelle = France entière, imposé par la philosophie des carroyages
- 3) La **validation par la CNIL** se fera source par source
On commence par les données de la filière fiscale (TH et RFL)

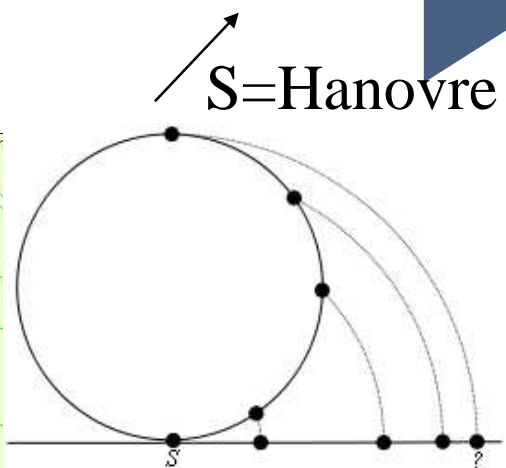
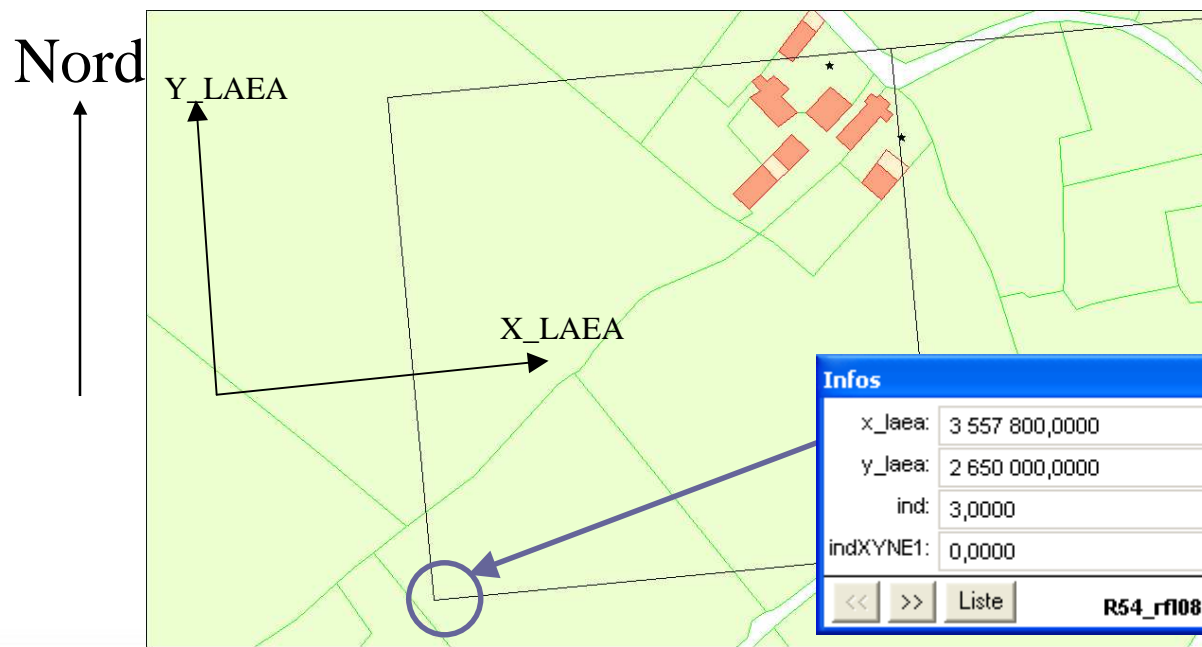
Page 57





1- La géographie

Le Lambert Azimutal Equal Area (LAEA) Des carreaux pas carrés



Infos

x_laea:	3 557 800,0000
y_laea:	2 650 000,0000
ind:	3,0000
indXYNE1:	0,0000

<< >> Liste R54_r108_LAEA200

Extrait de la base MIF/MID de diffusion, résolution 200m

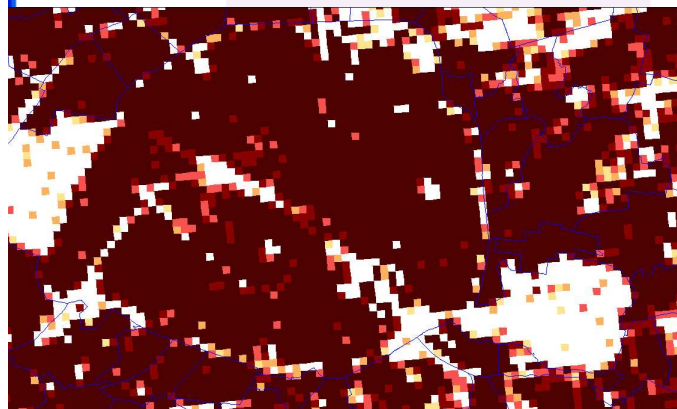
Coordonnées LAEA du coin Sud-Ouest

Population fiscale

Population fiscale non localisée à la parcelle

→ Environnement

☑ Travail - Emploi



:: Téléchargement des données

Chaque fichier compacté (format zip) contient les données sous deux formes :

- Un fichier en format dBase (DBF),
- Un fond de carte en format de transport MapInfo (MIF/MID).

Dans les deux cas, les carreaux décrits dans les fichiers sont uniquement ceux portant une population non nulle.

 [Télécharger les données](#) (format zip : 2 412 Ko)

:: Téléchargement les fonds de carte départementaux complets

Chaque fichier zippé (*.zip) contient le fonds de carte départemental complet en format de transport MapInfo (MIF/MID).

Les données de population sont proposées dans le fichier régional ci-dessus.

 [Télécharger la documentation](#)

 [Télécharger le fonds de carte département 75](#) (format zip : 64,5 Ko)

 [le département 77](#) (format zip : 4,3 Mo)

 [le département 78](#) (format zip : 1,5 Mo)

 [le département 91](#) (format zip : 1,1 Mo)

 [le département 92](#) (format zip : 112 Ko)

 [le département 93](#) (format zip : 147,5 Ko)

 [le département 94](#) (format zip : 151 Ko)

 [le département 95](#) (format zip : 748 Ko)

Infos

x_jaea: 3 755 000,0000

y_jaea: 2 887 200,0000

COG1: 75056

COG2: 92040

COG3:

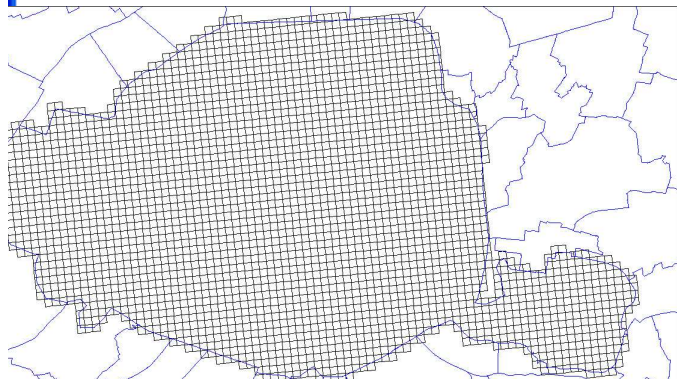
COG4:

COG5:

COG6:

id: LAEA200M_N14436E18775

<< >> Liste dep_75





2- Les indicateurs

› Des indicateurs ménages (à affiner : sources et détail)

- Population par tranches d'âge
- Ménages selon quelques caractéristiques simples
- Revenu des ménages
- Logement (source TH)
- ...

➔ Proposition opérationnelle en CNIS/Commission Territoires le 30/5

› Entreprises

- Le concept de carreau n'est ni le bon concept ni ce qui est demandé par le rapport Cuiller
- L'objectif à court terme (deux ans) est de mettre à disposition des bases d'établissements et d'équipement géolocalisés
 - La géolocalisation à grande échelle a démarré début 2012
- Pour l'emploi, pas à court terme : il faut encore progresser sur CLAP pour une exploitation fine. Le recours aux carreaux est souhaitable (établissements étendus).

Page 60





3- Le traitement de la confidentialité

C'est l'image globale qui est importante

- › Objectif : garantir qu'aucune donnée individuelle ne sera accessible ni directement ni par approximation fiable.

- › Définition de **populations à risque** (à faire par variable diffusée), exemple sur le revenu médian:
 - 1) Moins de 11 ménages
 - 2) Maille isolée ou ensemble de mailles isolées de moins de 80 ménages
 - 3) Ménages des mailles appartenant à un zonage standard (IRIS ou ZUS) de moins de 200 ménages
 - 4) Ménages des mailles où le taux de bas revenus est extrême par rapport à la situation communale (« outliers »)

- › Traitement : **Permutation** des ménages avec des ménages de mailles voisines, si les caractéristiques sont différentes, **imputation** d'un revenu moyen sinon
Exhaustivement sur les risques 1 et 2, un sur quatre sur les risques 3 et 4

Page 61





Des données [garanties] fausses...

Données originales :

id	n	SEXE_1	SEXE_2	AGE_1	AGE_2	AGE_3
00134	1	1	1	0	0	1
00135	1	1	0	0	1	0
00222	20	8	12	3	13	4
00308	11	8	3	3	8	0
00309	12	7	5	0	11	1
00394	3	2	1	0	1	2
00395	3	1	2	1	0	2
00473	1	0	1	0	1	0
00474	3	1	2	0	3	0

Brouillées

→ Pas de changement : hors sujet confidentialité

Données publiables :

id	n	SEXE_1	SEXE_2	AGE_1	AGE_2	AGE_3
00134	1	0	1	0	0	1
00135	1	0	1	0	1	0
00222	20	11	9	4	12	4
00308	11	6	5	2	9	0
00309	12	6	6	0	11	1
00394	3	3	0	0	1	2
00395	3	1	2	0	1	2
00473	1	0	1	0	1	0
00474	3	2	1	0	3	0

→ Pas de changement (mais qui le saura?)

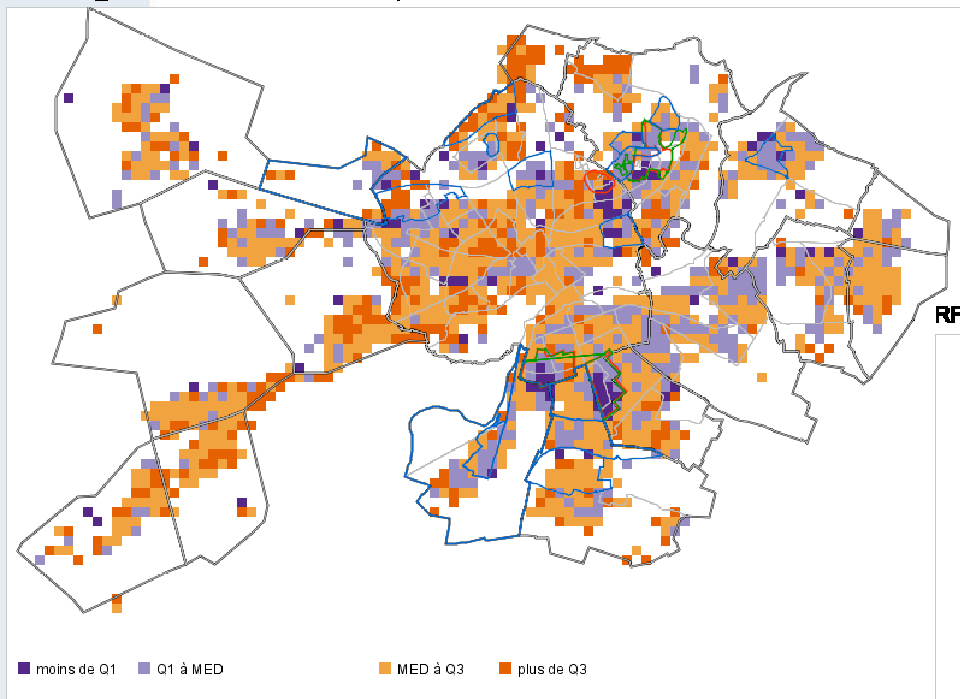




...mais dont l'utilisation reste possible

Revenu médian par carreau, agglomération de Caen

RFL2009_Caen - RFL - Revenu médian par carreau



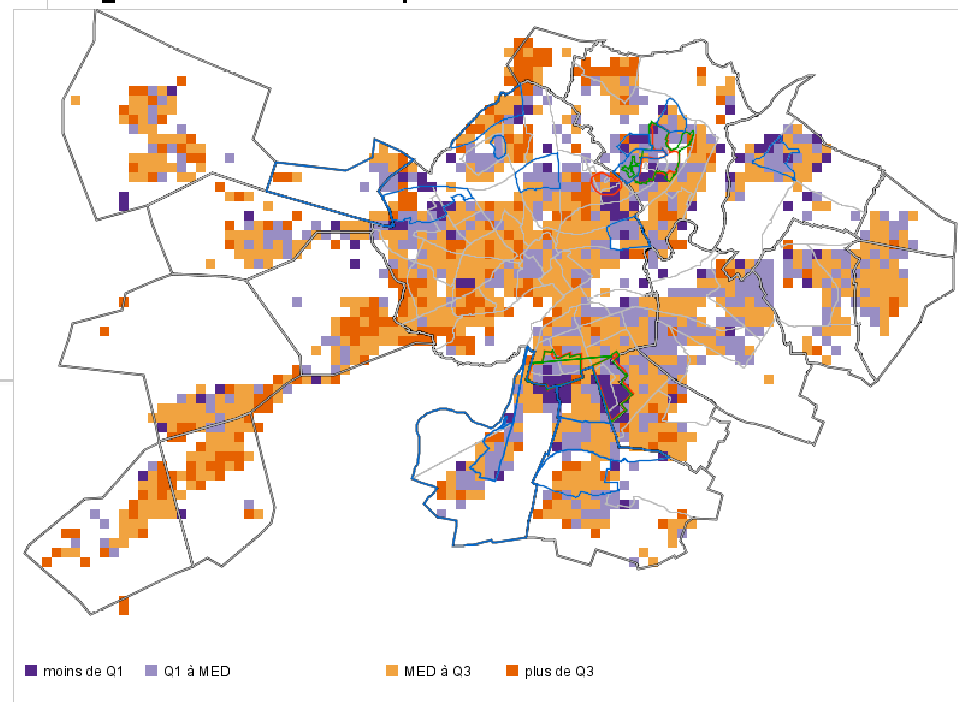
Données originales

Entre les deux :
10 000 ménages/80 000 brouillés

Page 63

Données brouillées

RFL2009_Caen - RFL - Revenu médian par carreau - Données traitées





L'Insee et la statistique publique

Thèmes

Bases de données

Publications et services

Régions

Définitions et méthodes

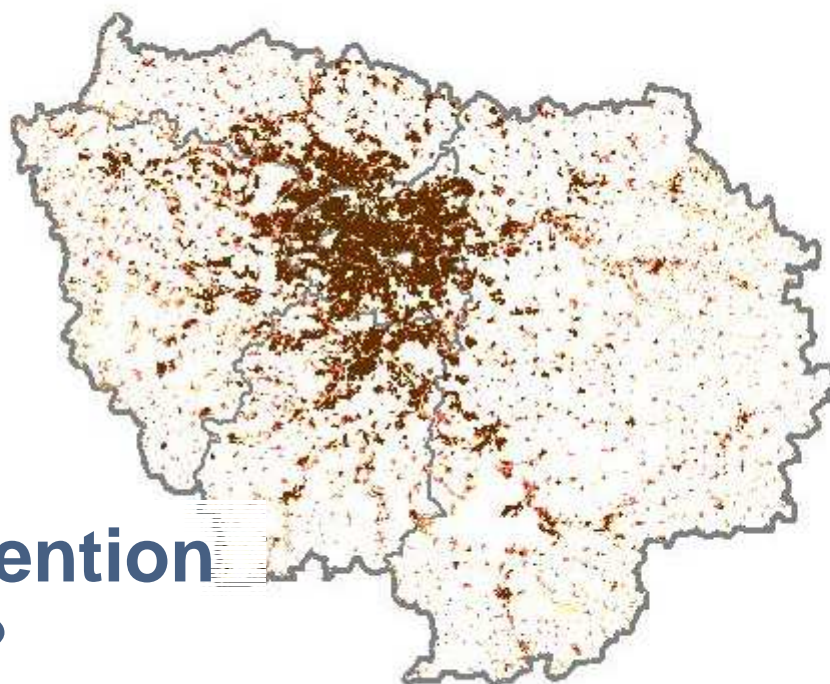
Accès public

Thèmes

- ⊖ Agriculture
- ⊖ Commerce
- ⊖ Comptes nationaux - Finances publiques
- ⊖ Conditions de vie - Société
- ⊖ Conjoncture
- ⊖ Économie
- ⊖ Enseignement - Éducation
- ⊖ Entreprises
- ⊖ Industrie - IAA - Construction
- ⊖ Population
- ⊖ Revenus - Salaires

⊖ Accueil ⊖ Thèmes ⊖ Territoire ⊖ Régions, dé... ⊖ Données car... ⊖ Données à 2... ⊖ Ile de France

■ Ile de France



:: Légende

Densité de population
Nombre d'habitants estimés
par carreaux de 200 m

- 0 - 4
- 5 - 9
- 10 - 19
- 20 - 59
- >= 60

:: Pour en savoir plus

- ⊖ Documentation sur données carroyées

:: Liens

- ⊖ Changer de région

Merci de votre attention
Des questions?

⊖ Régions, départements et