

# Diffusion au carreau sur Insee.fr

Commission Territoire du CNIS



Mesurer pour comprendre

# Une méthodologie revue

---

- La source des données : Revenus Fiscaux Localisés 2010
  - ⇒ règle de diffusion des données sur les revenus fiscaux des ménages : au moins 11 ménages par unité géographique d'observation
- Abandon des procédures de « brouillage » utilisées lors de la première livraison
- Nouveau traitement de la confidentialité :
  - Regroupement des carreaux de trop faibles effectifs
  - Traitement spécifique appliqué aux variables « à risque »
- Données disponibles pour :
  - France métropolitaine
  - Réunion, Martinique

# Des carreaux aux rectangles

---

- Carreaux habités de moins de 11 ménages
  - Regroupés dans des mailles rectangulaires
    - ⇒ facilement appréhendables par leur forme
- Carreaux habités de 11 ménages ou plus
  - Diffusés comme tels ou inclus dans un rectangle pour un certain nombre d'entre eux (pour des raisons de construction des rectangles)
- Au total : 698 659 mailles (rectangles + carreaux seuls)

# Le traitement des variables « à risque »

---

- Les revenus sont « winsorisés »
  - Revenus par UC > 8<sup>ème</sup> décile (29 336 €) : descendus à ce seuil
  - Revenus par UC < 40% de la médiane (7 500 €) : remontés à ce seuil
- Autres variables « à risque »
  - **Objectif :**  
Informations diffusées ne doivent pas correspondre à une proportion > 80%
  - **Traitement :**  
Si > 80%, on diminue le total du plus petit nombre entier afin de passer à une proportion < 80%  
On complète la variable par une indicatrice précisant si le traitement a été opéré ou non

# Deux fichiers diffusés : carreaux et rectangles

---

- Les variables du fichier des carreaux habités
  - Identifiants géographiques du carreau
  - Nombre de carreaux habités du rectangle auquel appartient le carreau
  - Identifiants géographiques du rectangle auquel appartient le carreau
  - Nombre total d'individus du carreau

# Deux fichiers diffusés : carreaux et rectangles

---

## ➤ Variables non traitées du fichier des rectangles

- Nombre total d'individus
- Nombre total de ménages (i.e. de résidences principales)
- Nombre total d'individus par classes d'âge  
0-3 ans ; 4-5 ans ; 6-10 ans ; 11-14 ans ; 15-17 ans ; 25 ans et plus
- Nombre total de ménages de 5 personnes et plus
- Nombre total de ménages présents depuis 5 ans ou plus dans leur logement actuel
- Nombre total de ménages en logement individuel
- Nombre total de ménages en logement collectif
- Surface cumulée des résidences principales

# Deux fichiers diffusés : carreaux et rectangles

---

- Variables « à risque » et donc traitées, du fichier des rectangles
- Somme des revenus fiscaux par UC winsorisés des individus
  - Nombre de ménages dont le revenu fiscal par UC est  $<$  au seuil de bas revenus (60% de la médiane)
  - Nombre total d'individus de 65 ans ou plus
  - Nombre total d'individus de 75 ans ou plus
  - Nombre de ménages d'une personne
  - Nombre de ménages propriétaires

# Mode d'utilisation des fichiers

---

- Le fichier de rectangles : un fichier intermédiaire
  - ⇒ éviter de l'utiliser tel quel (en particulier pour faire des cartes) car :
    - superficies variables
    - possibilité de fortes disparités spatiales internes dans les grands rectangles
- Pour cartographier les effectifs d'individus :
  - utiliser le fichier de carreaux diffusés
- Pour cartographier les autres variables :
  - **Nécessité de construire un nouveau fichier de carreaux**  
Désagrégation de l'effectif du rectangle entre les carreaux qui le composent (au prorata de la population totale).

# Tests réalisés

---

- Travaux répartis entre deux équipes de la DG
  - Contrôle des programmes
  - Confrontation avec d'autres fichiers diffusés (risques de rupture de confidentialité par croisement de fichiers)
    - Décision de ne plus diffuser de revenu moyen à l'Iris et à la ZUS
  - Cohérence avec les données d'autres sources (RP)
  - Contrôle de conformité (absence de données sur des mailles de moins de 11 ménages)
  - Cartographie

# Consultation des utilisateurs : les modalités

---

- Consultation lancée fin juin auprès de 13 utilisateurs
  - Agences d'urbanisme, chercheurs, réseau équipement, collectivités territoriales, géomaticien
  - 9 réponses reçues
  
- Envoi de :
  - Fichiers tests (une région)
  - Note décrivant la nouvelle méthodologie
  - Mode opératoire sous Mapinfo
  
- 3 questions posées

# Consultation des utilisateurs : les messages

---

➤ Question 1 : Nouvelle méthodologie compréhensible ?  
Ajouts à faire dans la documentation ?

- Méthodologie bien comprise
- Quelques suggestions pour améliorer la documentation, compléter les métadonnées
- Quelques suggestions pour améliorer la méthode de constitution des rectangles (contrainte de distance...)

# Consultation des utilisateurs : les messages

---

➤ Question 2 : Avis sur le mode de diffusion (deux fichiers à fusionner + étape de désagrégation des données sur les rectangles)

- Manipulation un peu lourde
- Risque d'erreurs lors de l'étape de désagrégation
- Risque que des fichiers carreaux différents soient fabriqués, utilisés et diffusés
- Risque de sous-utilisation du produit
- Risque d'utilisation du fichier de rectangles tel quel (et donc de représentations cartographiques qui peuvent être erronées)

## Propositions :

Diffuser le fichier des carreaux estimés (au lieu de le faire fabriquer par les utilisateurs).

éventuellement : deux fichiers de carreaux à fusionner (un fichier de carreaux fiables > 10 ménages et un fichier de carreaux estimés)

# Consultation des utilisateurs : les messages

---

- Question 3 : Difficultés rencontrées pour l'opération de fusion-désagrégation ? Suggestions ?
  - Opération qui ne pose aucune difficulté pour des utilisateurs tous « experts » dans ce type de manipulation de fichiers.

# Consultation des utilisateurs : les messages

---

## ➤ Remarques diverses

### ■ Sur la liste des variables

Regret de ne plus voir figurer le revenu médian

Critique de la suppression de quelques variables par rapport à la liste initiale (nombre de ménages au carreau, résidences secondaires, en particulier).

Choix des classes d'âge mal compris (découpage très fin avant 15 ans, et très grossier au delà).

### ■ La modification de certaines variables par le seuil de 80%

Rupture des propriétés d'agrégation additive des données

### ■ Winsorisation du revenu

Critique du choix du 8<sup>ème</sup> décile (le 9<sup>ème</sup> aurait été préféré, car moins limitatif pour les analyses territoriales fines)

Critique aussi de la winsorisation en tant que telle, car « *introduisant des valeurs arbitrairement fausses sur l'un des sujets les plus importants pour la société* »

# Consultation des utilisateurs : les messages

---

## ➤ Une demande...

- Serait-il envisageable de mettre à disposition des fichiers moins restrictifs pour des « utilisateurs avertis travaillant pour l'aide à la décision publique » ?

---

**Merci pour votre attention !**