Ségrégation urbaine: un éclairage par les données de téléphonie mobile

Présentation Journées de Méthodologie Statistique

Lino Galiana

14 juin 2018

Enjeux

Comment croiser données administratives et mobiles?

Mesures de la ségrégation

Résultats

Enjeux

Les différentes dimensions de la ségrégation

- Ségrégation est un **phénomène multi-dimensionnel**:
  - Séparation spatiale des groupes sociaux (*ségrégation spatiale*)
  - Absence d'interactions entre ceux-ci (*ségrégation sociale*)
- Approche par **ségrégation résidentielle**:
  - Mesure séparation dans espace résidentiel
  - Pas de mesure interactions sociales ou rencontre des populations
  - Avantage données administratives: vision exhaustive sur la dimension revenu
- Approche par **enquête**:
  - Possibilité de suivre la mobilité
  - Contrainte taille échantillon

Données de téléphonie mobile: information sur comportement individuel sur dimensions manquantes des données administratives

Comment mesurer la ségrégation ?

- **Indices de ségrégation**:
  - Comparaisons spatiales (entre les villes et au sein de celles-ci)
  - Comparaisons temporelles
- Revenu: **variable continue** $\Rightarrow$ *Rank ordered index* (Reardon & O'Sullivan, 2004)
  - Indices classiques: variables discrètes (bi- ou multimodales)
  - Rangs qui sont utilisés dans la construction d'indices
  - Permet des comparaisons d'indices entre des villes dont dispersion revenu diffèrent
- **Deux dimensions ségrégation** identifiées par Reardon & O'Sullivan (2004):
  - *Exposition*: ampleur par laquelle les groupes sociaux se rencontrent
  - *Concentration*: ampleur par laquelle les groupes sociaux sont concentrés dans l'espace

Données de téléphonie mobile: mesure de l'exposition (dans espace physique et social) et possibilité mesure alternative de la concentration (par rapport approche résidentielle)

Choix de l'unité spatiale

- Problème d'agrégation spatiale (MAUP):
  - Distribution des revenus au sein des unités spatiales affectée par découpage géographique
  - Hétérogénéité des unités spatiales pose problème
- Privilégier une **échelle spatiale homogène**:
  - *Population fixe* ? (IRIS)
  - *Superficie fixe* ? (grille)
  - *Unités spatiales à partir données mobiles* ? (tesselation de Voronoï)

Choix d'une grille avec carreaux 500x500m pour avoir espaces homogènes ⇒ probabilisation des événements mesurés au niveau des antennes

Données utilisées

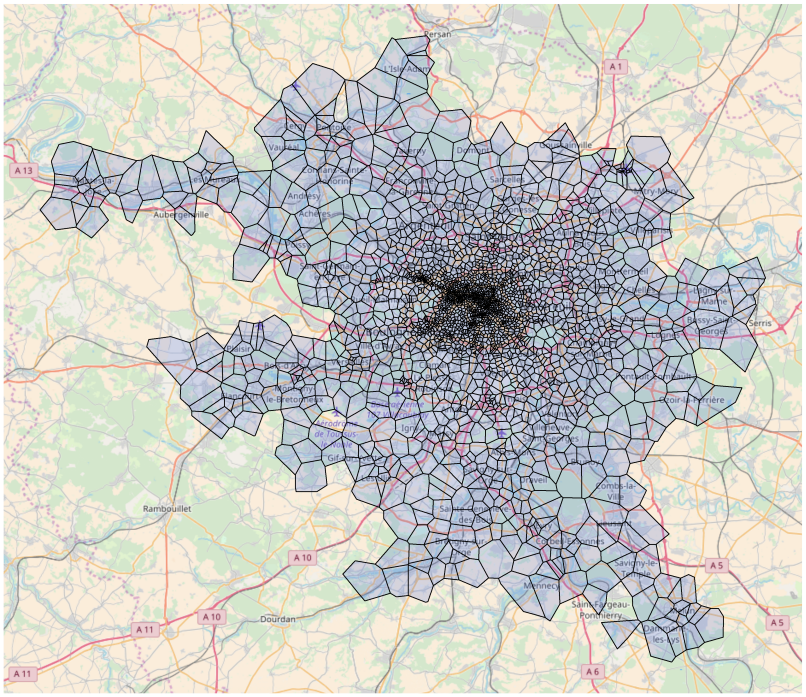
- Comptes rendus d'appels d'**Orange** (5 mois sur **année 2007**):
  - Données individuelles: 18 millions de carte SIM
  - $\approx$ 3.5 milliards d'interactions
  - Données actives: appels & sms ($\neq$ données passives)
- **Données géolocalisées**:
  - Interactions entre utilisateurs
  - Mobilité (avec horodatage)
- **Evénements mesurés au niveau d'une antenne**
  - Passage à une grille
- **Pas de caractéristique individuelle**
  - Données Filosofi: distribution spatiale du revenu
  - Utilisation dimension spatiale pour croiser avec données administratives

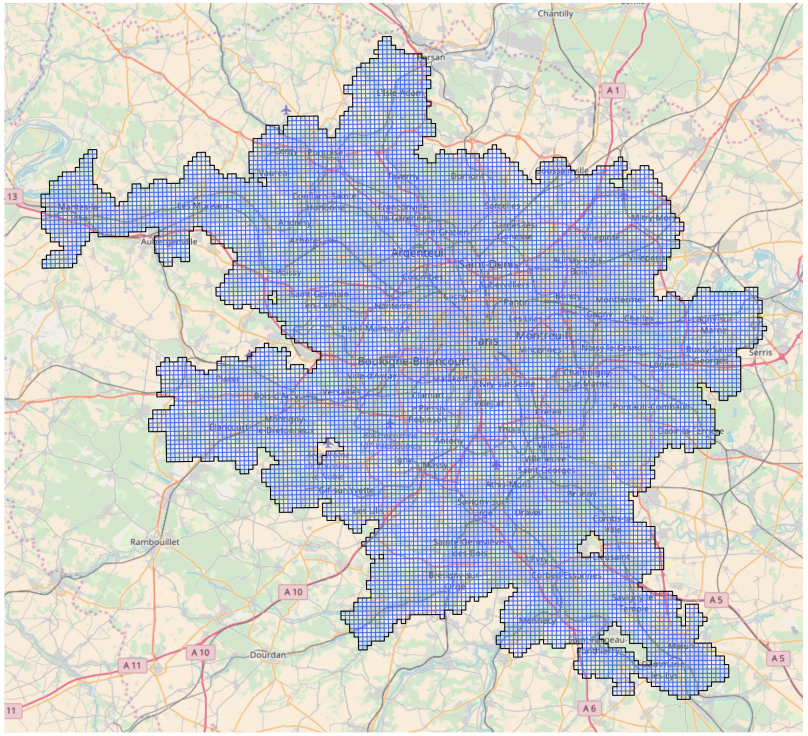
Comment croiser données administratives et mobiles?

Passage du voronoï à la grille

- **Tesselation de Voronoï**:
  - Modélisation de l'aire de couverture des antennes (en pratique, celle-ci est inconnue)
  - Espaces très hétérogènes: très petits dans les zones denses, très larges dans zones moins couvertes
- **Changer d'unité spatiale fondamentale**:
  - Probabiliser localisation des événements pour s'abstraire des problèmes d'approximation de l'aire de couverture
  - Grille plus homogène que les voronoï pour agrégation spatiale
  - Passage du voronoï au carreau 500m^2 par aires d'intersection $\mathbb{P}(c_i|v_j) = \frac{\mathcal{S}(c_i \cap v_j)}{\mathcal{S}(v_j)}$
- Analyse sur trois grandes **unités urbaines**:
  - Retient antennes dans limites des unités urbaines de Paris, Lyon, Marseille
  - Centre denses et périphéries

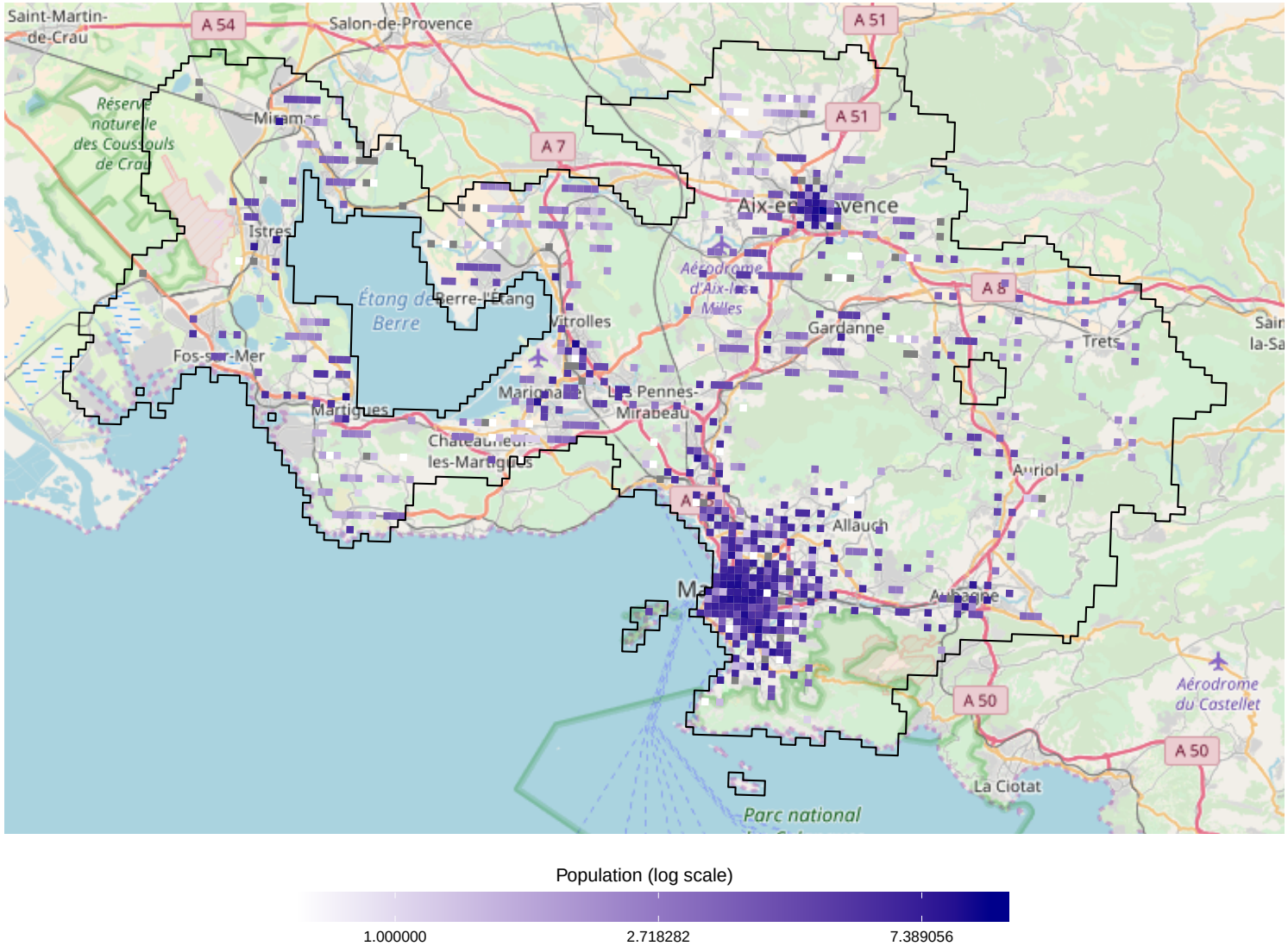
Passage du voronoï à la grille





Attribution d'un lieu de vie

Détection du domicile de l'utilisateur de téléphone à partir de ses appels en soirée



Imputation du revenu aux utilisateurs de téléphone

- Filosofi ⇒ **Calcul moments de la distribution du revenu**
  - Revenu par UC avec coordonnées (x, y)
  - Agrégation au niveau de la grille: médiane par grille
  - Respect secret statistique
- Utilisateur téléphone ⇒ **imputation du revenu**
  - Revenu médian dans le voisinage du lieu de vie (grille de résidence identifiée)
- Approfondissement ⇒ **attribution aléatoire du revenu**
  - Plus de quantiles dans le futur
  - Gérer la population différente selon les unités spatiales

Mesures de la ségrégation

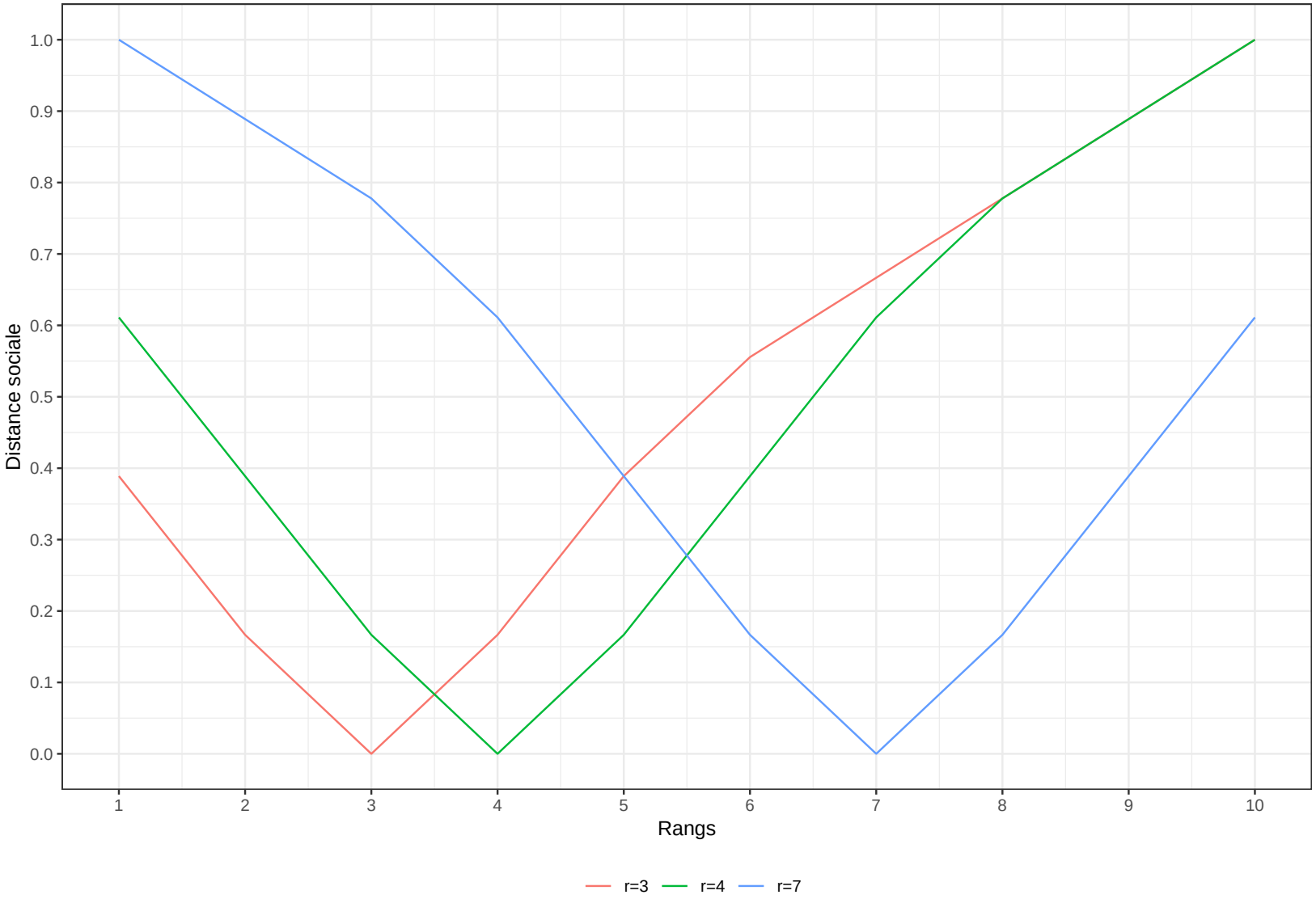
Mesure de distance sociale

- **Ordonne** les individus du plus pauvre au plus riche
- Création d'une **distance sociale**

$$\text{Quand } x \leq \frac{N}{2}, \quad d_{x \to y} = \begin{cases} 0 \text{ si } j = x \\ \frac{|4x-4j|-1}{2(N-1)} \text{ si } j \leq 2x-1 \text{ et } j \neq x \\ \frac{j-1}{N-1} \text{ si } j > 2x-1 \end{cases}$$

$$\text{Quand } x > \frac{N}{2}, \quad d_{x \to y} = \begin{cases} 0 \text{ si } j = x \\ \frac{|4x-4j|-1}{2(N-1)} \text{ si } j \geq 2x-N \text{ et } j \neq x \\ \frac{N-j}{N-1} \text{ si } j < 2x-N \end{cases}$$

Mesure de distance sociale



Indice de ségrégation sociale

- Utilisation des **interactions téléphoniques**
- Mesure l'*exposition* aux interactions avec des individus "*éloignés socialement*"
- **Moyenne pondérée des distances sociales entre un individu et ses contacts par le nombre d'interactions**
- **Indice individuel**:
  - $\neq$ indices classiques
  - 0.5: profil de communication uniforme

- $s_{x \to y} = 1 - d_{x \to y}$ similarité entre x et y
- Individu x communique avec contacts $(y_1, ..., y_m)$ à la fréquence $(f_1, ..., f_m)$

$$SSI_x = \frac{\sum_{j=1}^{m} f_j s_{x \to y_j}}{\sum_{j=1}^{m} f_j}$$

Indice de ségrégation spatiale

- Utilisation des **déplacements individuels** et de la **dimension temporelle** des données mobiles
- Mesure l'*exposition* aux autres groupes sociaux dans l'espace et le temps
- **Moyenne des distances sociales entre un individu et les individus fréquentant le même lieu au même moment pondéré par la probabilité de coprésence dans la grille**
- **Indice individuel**

Indice de ségrégation spatiale

- Soit $\mathbb{P}_x(c_j, T)$ probabilité de se trouver dans le lieu c_j pendant la période de temps T
- Soit $\mathcal{U}(l, T)$ l'ensemble des individus ayant une probabilité non nulle d'être présent au même instant dans ce lieu
- Soit $\mathcal{C}_T$ l'ensemble des carreaux visités par x pendant la période T

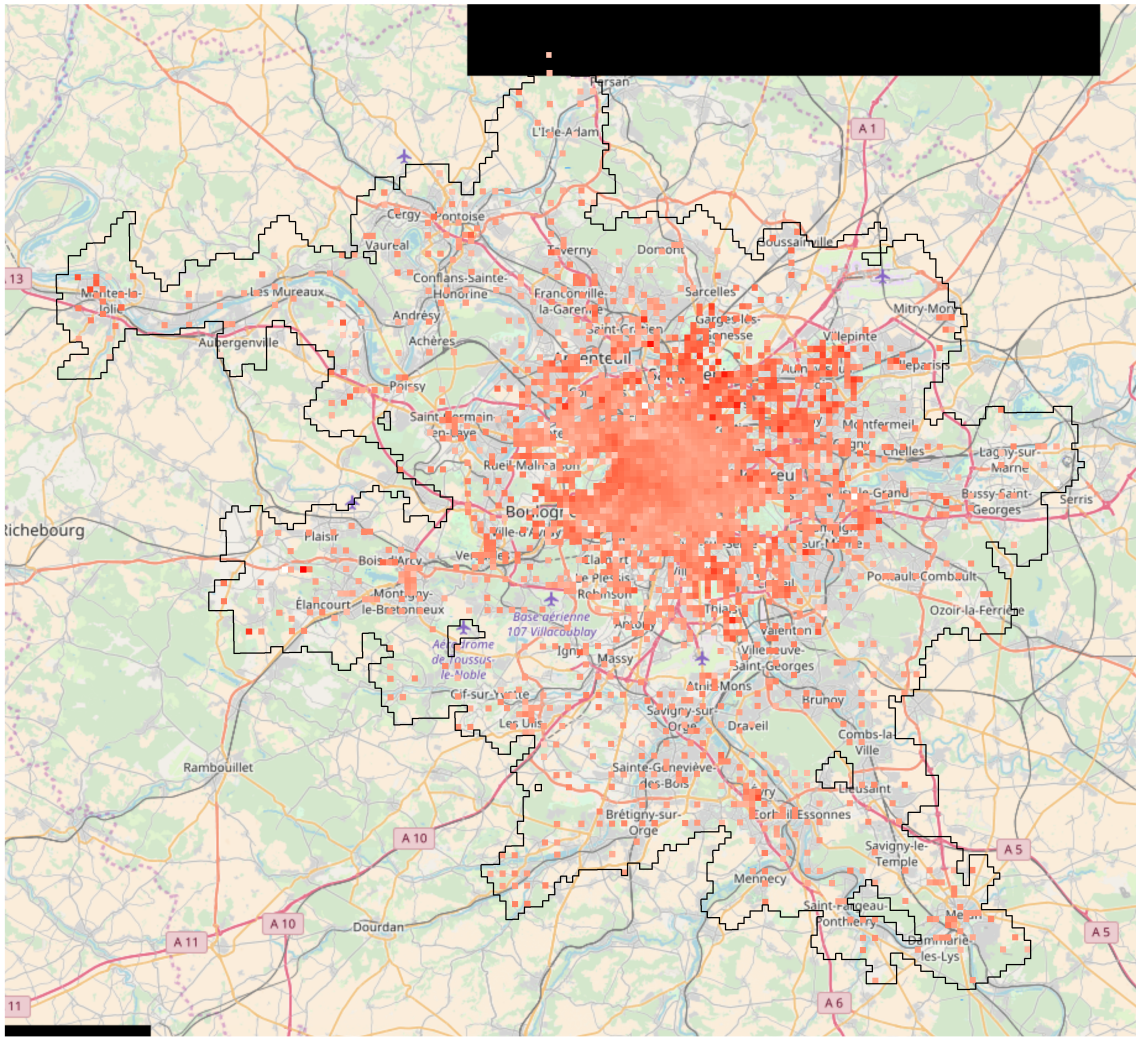
$$\begin{aligned} PSI_x(c_j, T) &= \frac{\sum_{y \in U(c_j,T)} \mathbb{P}_x(c_j, T)\mathbb{P}_y(c_j, T)s_{x \to y}}{\sum_{y \in U(l,T)} \mathbb{P}_x(c_j, T)\mathbb{P}_y(c_j, T)} \\ &= \frac{\sum_{y \in U(c_j,T)} \mathbb{P}_y(c_j, T)s_{x \to y}}{\sum_{y \in U(c_j,T)} \mathbb{P}_y(c_j, T)} \end{aligned}$$

$$PSI_x(T) = \sum_{c_j \in \mathcal{C}_T} \mathbb{P}_x(c_j, T)$$

Résultats

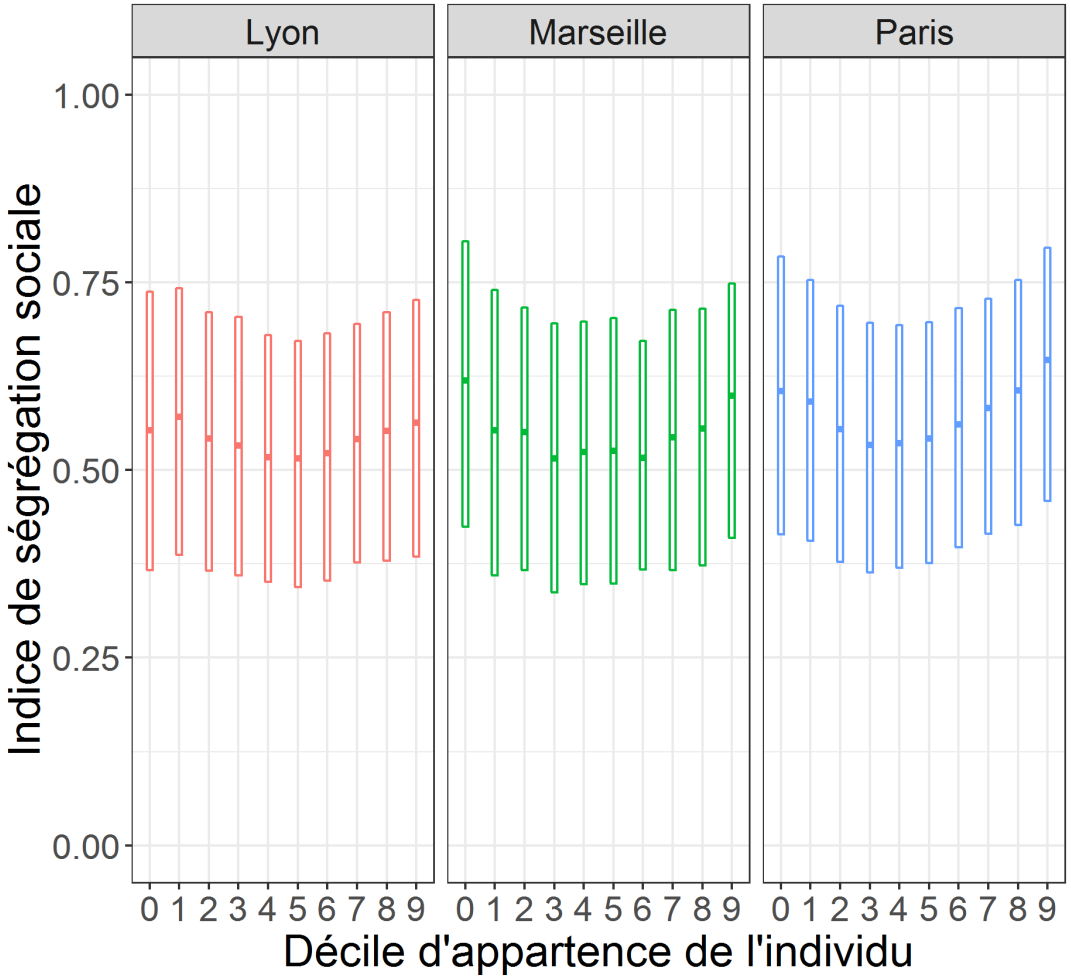
Ségrégation sociale à Paris

- Ségrégation élevée:
  - ouest (quartiers riches)
  - nord-est (quartiers pauvres)



Profil social

- Habitants des quartiers les plus riches et les plus pauvres sont plus ségréguées
- Paris plus ségréguée que Marseille et Lyon



Limites et approfondissements

- **Limites**:
  - Biais si choix de l'opérateur lié au revenu, avec écarts selon le lieu de résidence
  - Biais du choix du mode d'interactions sociales
  - Reproductibilité de l'étude: s'adapter au caractère changeant des communications numériques (*whatsapp*...)
- **Approfondissements**:
  - Ségrégation spatiale pour étudier dimensions complémentaires à l'approche résidentielle
  - Tirer aléatoirement un revenu plutôt qu'affecter le revenu médian
  - Construire des tests formels d'adéquation entre Filosofi et simulation de revenu
  - Analyser l'autocorrélation spatiale des indices de ségrégation
  - Comparer à d'autres indicateurs de ségrégation

Conclusion

- **Intérêt des données mobiles**:
  - Mesurer ségrégation sociale avec données de téléphonie
  - Evolution temporelle ségrégation pour un espace donné
- **Enjeux**:
  - Etudier un comportement individuel sans connaissance a priori des caractéristiques individuelles
  - Croiser données administratives et *big data*
- **Premiers résultats intéressants**
  - Ségrégation plus marquée à Paris
  - Profil de ségrégation plus fort aux extrêmes de la distribution
  - Des quartiers se distinguent particulièrement à Paris: nord-est, sud-ouest