

Tirage coordonné d'échantillons : une application à l'Échantillon-Maître Nautile et à l'enquête Emploi

Thomas Merly-Alpa, Ludovic Vincent, Laurent Costa,
Clément Guillo, Nicolas Paliod, Martin Chevalier, Thomas Deroyon

Séminaire de Méthodologie Statistique - Fidéli-Nautile : une nouvelle source d'informations et base de sondage sur les individus et les ménages

29 novembre 2018



Rappel : problématique et contexte

- Renouveler l'Échantillon-Maître (EM)
 - Une unité primaire est un regroupement de communes ;
 - C'est l'ensemble des logements dans lesquels se concentrera la collecte des enquêtes ménages ;
 - Le nouvel EM sera mis en place en 2020 : tirage spatialement équilibré.
- Renouveler l'échantillon de l'enquête Emploi en continu (EEC)
 - Obligation de renouvellement : fin de vie de l'échantillon actuel ;
 - Le nouvel échantillon de l'EEC débutera en juillet 2019.
- Coordination des deux échantillons

- 1 Le nouvel échantillon de l'enquête Emploi
Rappels sur l'enquête Emploi
Tirage de l'échantillon des secteurs

- 2 La coordination avec l'Échantillon-Maître Nautille
Pourquoi coordonner les tirages ?
Comment coordonner les tirages ?
Choix des allocations régionales

- 1 Le nouvel échantillon de l'enquête Emploi
- 2 La coordination avec l'Échantillon-Maître Navile

Quelques rappels sur l'Enquête Emploi en Continu

- L'enquête Emploi en Continu (EEC) est le pendant français de la *Labour Force Survey*.
- Elle permet de mesurer le taux de chômage, le taux d'activité ainsi que d'autres éléments sur l'emploi (temps partiel, ...).
- L'échantillon est « aréolaire » : il n'est pas issu directement d'un tirage de logements, mais d'un tirage de groupes de logements proches (60 000 logements interrogés par trimestre).

Quelques rappels sur l'Enquête Emploi en Continu

- Ces groupes, ou « grappes », sont des ensembles géographiquement compacts d'une vingtaine de logements.
- Ces grappes sont regroupées en secteurs comportant 6 (ou 7) grappes.
- La collecte se fait au sein d'un secteur pour assurer un effet de pseudo-panel sur la mesure de l'évolution des taux de chômage.

Constitution des secteurs Emploi

- On commence par construire les grappes et secteurs.
- **Champ** : logements des fichiers Fidéli qui correspondent "aux résidences principales avec des occupants" (définition proche de celle du recensement).
→ 28 282 687 logements
- En 2008, le champ était uniquement basé sur les fichiers de la taxe d'habitation.
- L'échantillon est construit selon une logique ascendante : construction de grappes, puis de secteurs de 6 ou 7 grappes.

Constitution des grappes Emploi

- **Contraintes** de construction :
 - Grappe de 17 à 24 logements en ayant pour objectif 20 logements.
 - Tous les logements d'un même étage doivent se retrouver dans la même grappe.
- En 2008, le tri selon les variables foncières dans la base de sondage avait permis de créer **un** chemin, le long duquel les grappes étaient construites.
- **Nouvelle idée** : pouvoir utiliser **plusieurs** chemins différents et construire ces chemins sur la proximité géographique (X,Y).

Bilan de la constitution

2008		2018
199 500	Nombre de secteurs	231 966
4%	Secteurs à 7 grappes	1,7%
1 204 838	Nombre de grappes	1 395 866
88,15%	Grappes de 17-24 logements	99,71%
59,36%	Grappes de 19-21 logements	80,15%
2,3 km	Distance moyenne des grappes	1,5 km
21,5 log	Taille moyenne des grappes	20,2 log

- Tirage spatialement équilibré des secteurs.
- Tirage stratifié car il existe des contraintes de précision régionale.
- Variables d'équilibrage issues de Fideli :
 - Variables socio-démographiques ;
 - Variables liées aux revenus : proxy de l'activité et du chômage.

- 1 Le nouvel échantillon de l'enquête Emploi
- 2 La coordination avec l'Échantillon-Maître Nautille**

Pourquoi coordonner les tirages ?

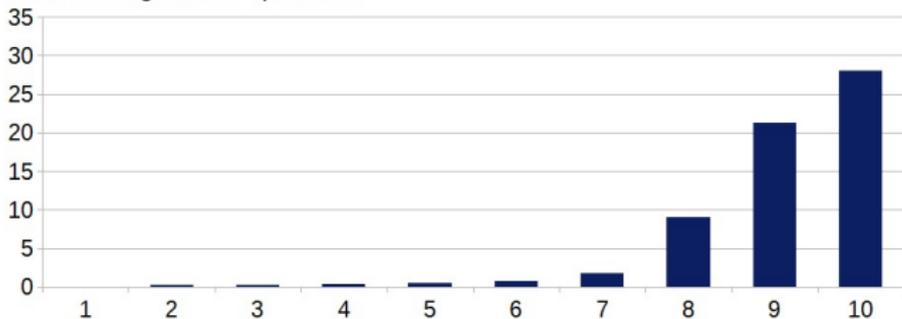
- Car les échantillons des enquêtes ménages et de l'enquête Emploi sont renouvelés à moins d'un an d'intervalle.
- Car cela limite les déplacements d'enquêteur dans des secteurs Emploi isolés.
- Car cela facilite les remplacements entre enquêteurs.

Tirer les secteurs dans les UP

- Tirer les secteur dans les UP ?
 - Certaines UP s'épuisent trop vite à cause de la non-réinterrogation des ménages ;

Épuisement des UP en cas de tirage de l'EEC dans les UP

Pourcentage d'UP épuisées

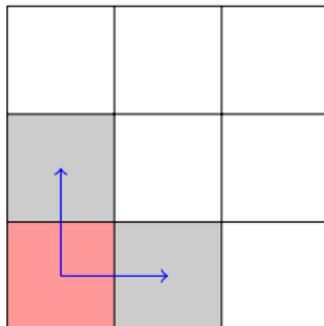


Année après la première utilisation de l'échantillon-maître

- L'effet grappe est trop marqué : perte de précision importante.

Quelle solution ?

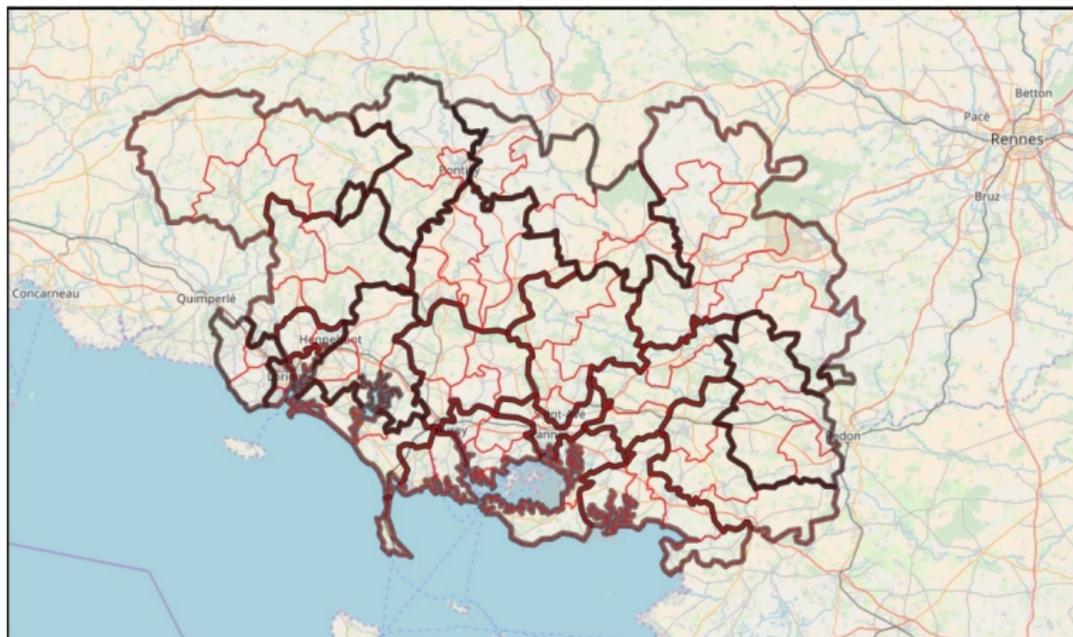
- Contrainte de la coordination : les secteurs emploi doivent être à proximité des unités primaires tirées.
- Tirer dans les unités primaires voisines :
 - Permet de garantir un tirage proche sans épuisement ni effet grappe trop fort ;
 - Problèmes : comment calculer les probabilités de sélection ? Comment gérer les frontières ?



Constituer de nouvelles zones

- Solution : dessiner de nouvelles zones (les unités de coordination - UC) en regroupant des unités primaires - UP
- Regrouper combien d'UP ? Combien de logements ?
- On décide d'un seuil minimal de 10 000 logements par UC :
 - Épuisement moindre des UP tirées ;
 - Le tirage de l'EEC est plus précis avec des UC plus grandes ;
 - Les UC restent peu étendues.

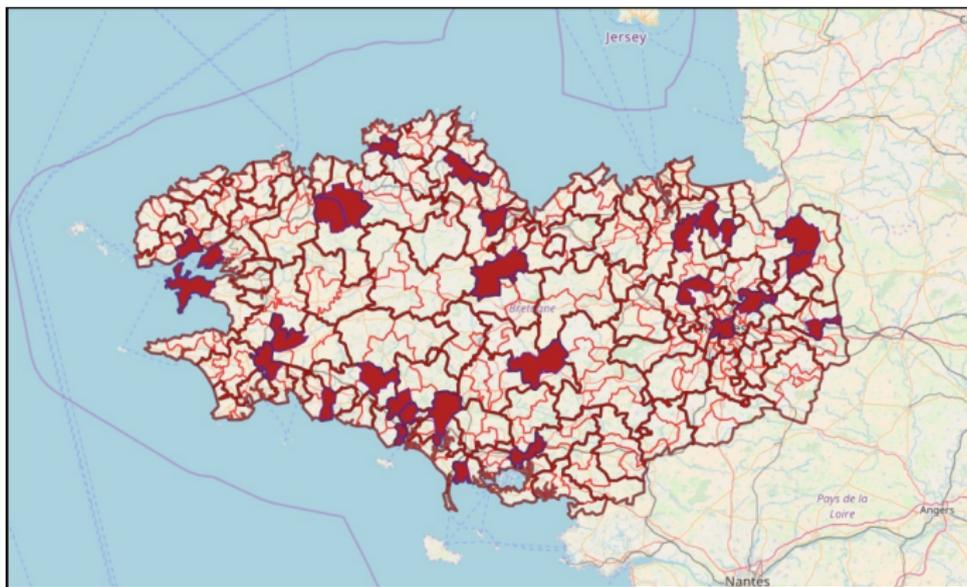
Exemple du Morbihan



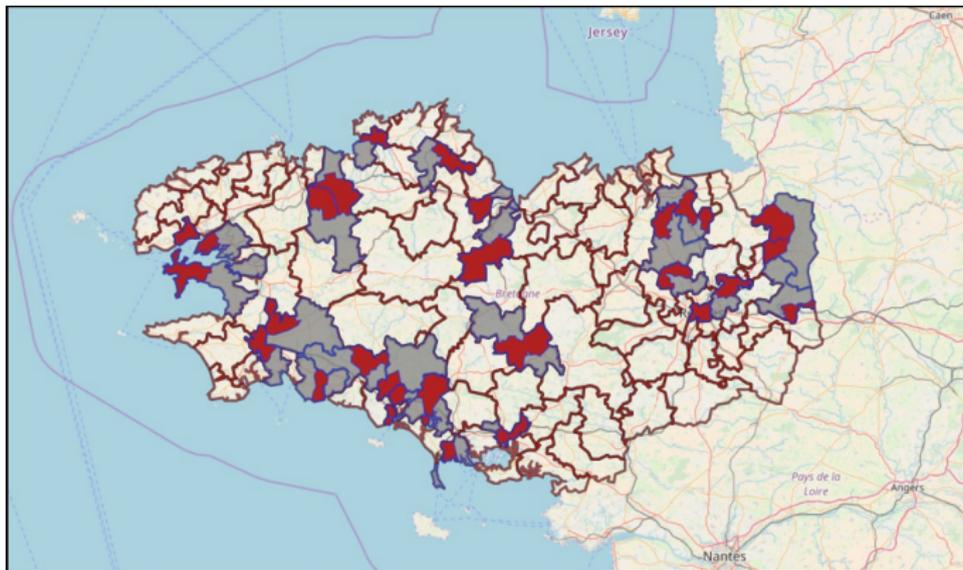
- Méthode intuitive : tirage à 2 degrés
 - Tirage des UC en premier lieu
 - Tirage des UP et des secteurs dans les UC tirées
- Tirer 1 UP par UC tirée entraîne une moindre qualité de l'échantillon-maître.
- Ne pas contraindre le tirage d'1 UP par 1 UC limite la coordination.
- Cette solution n'est pas retenue !

- Le problème du tirage direct des UC vient du trop grand nombre de contraintes sur le tirage des UP.
- Le tirage des secteurs pour l'EEC aurait fonctionné dans la méthode de tirage direct.
- Idée : Tirer les UP en premier pour limiter les contraintes sur les UC.
- Tirage indirect des UC :
 - Étape 1 : Tirage des UP ;
 - Étape 2 : Récupération des UC associées aux UP tirées ;
 - Étape 3 : Tirage des secteurs dans les UC récupérées.

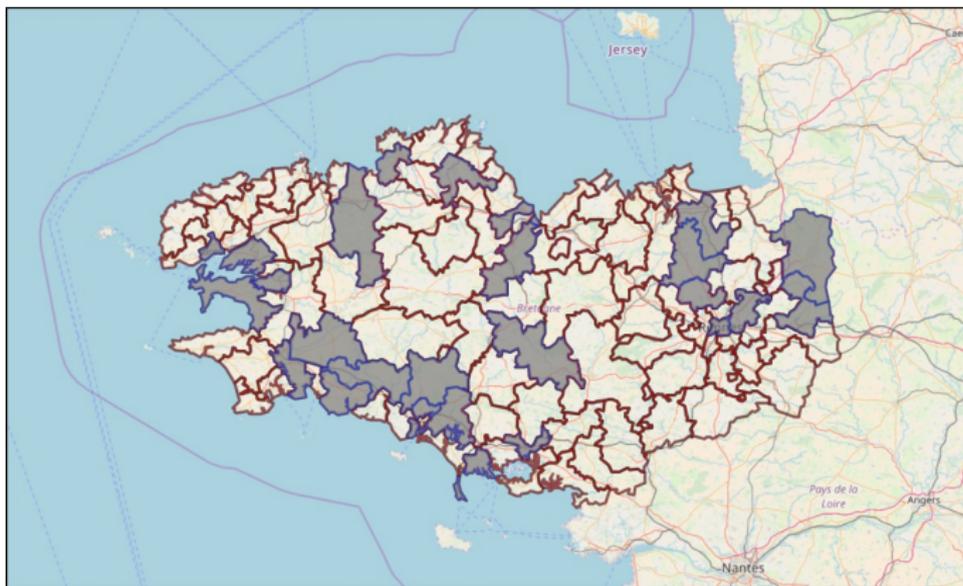
Étape 1 : Tirage des UP



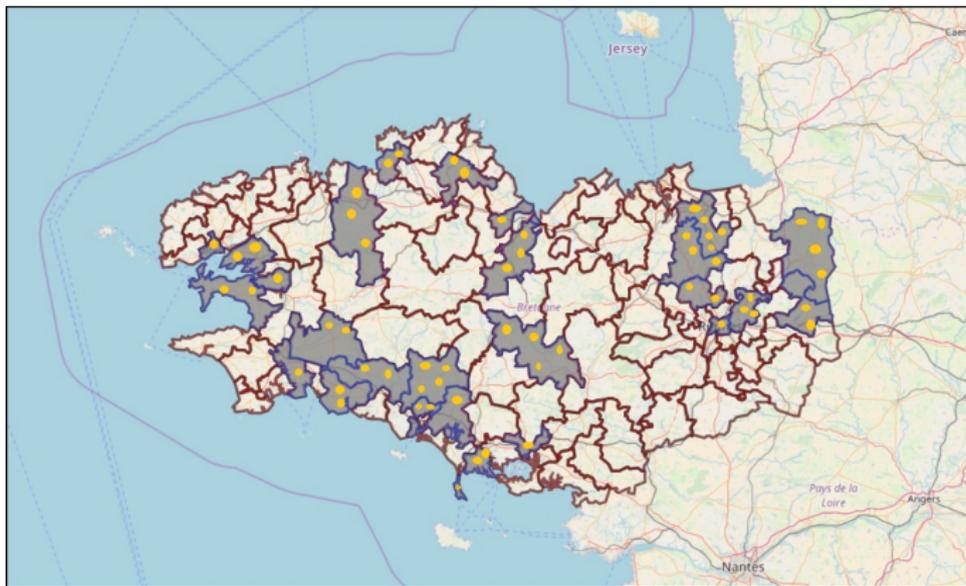
Étape 2 : Récupération des UC



Étape 2 : Récupération des UC



Étape 3 : Tirage des secteurs



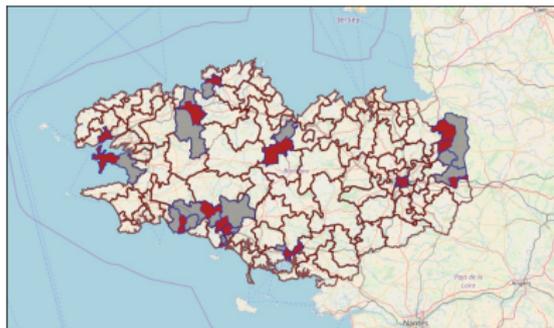
L'équilibrage indirect des UP

- Question restante : comment tirer les UP pour qu'elles soient précises pour l'EM et pour que les UC soient précises sur l'EEC ?
- Les UP sont équilibrées sur des variables d'intérêt des enquêtes ménages.
- Les UP déterminent indirectement les secteurs qui pourront être tirés.
- Nécessité d'intégrer des variables dans le tirage des UP **pour que les UC soient tirées de manière précise sur les variables de l'EEC.**

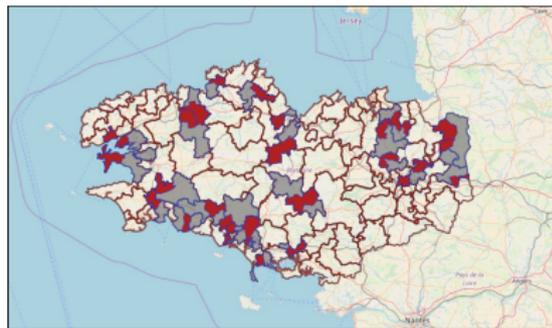
Les allocations régionales concernent deux éléments :

- le nombre d'UP tirées par région ;
- le nombre de secteurs Emploi tirés par région.

Tirer plus d'UP ou de secteurs ?



Tirage de 15 UP



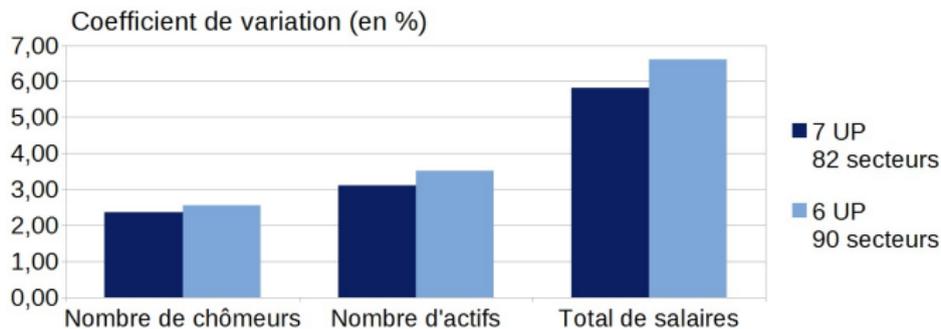
Tirage de 29 UP

	+ d'UP	+ de secteurs
précision EM	+	neutre
précision EEC	+	+

Pour la précision de l'EEC, quel arbitrage entre UP et secteurs ?

Tirer plus d'UP ou de secteurs ?

Influence des paramètres sur la précision de l'enquête Emploi dans le Limousin



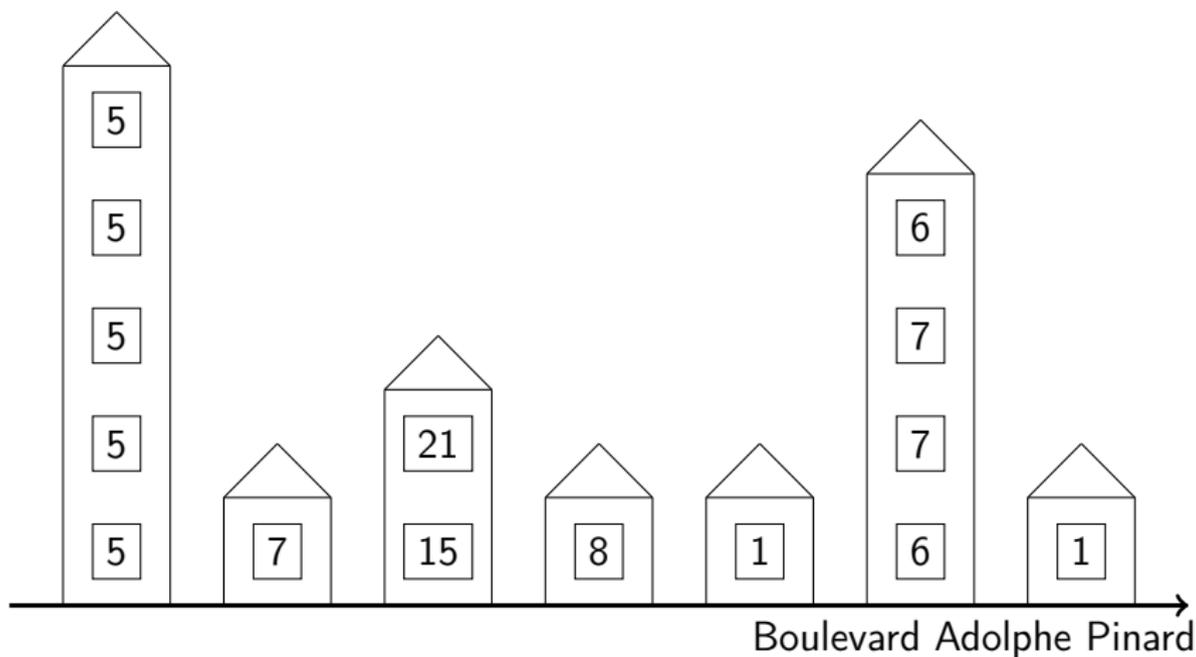
- Faible effet d'une augmentation de 10% de secteurs ;
- Fort effet d'une diminution de 10% d'UP.

→ Il est pertinent d'échantillonner plus d'UP.

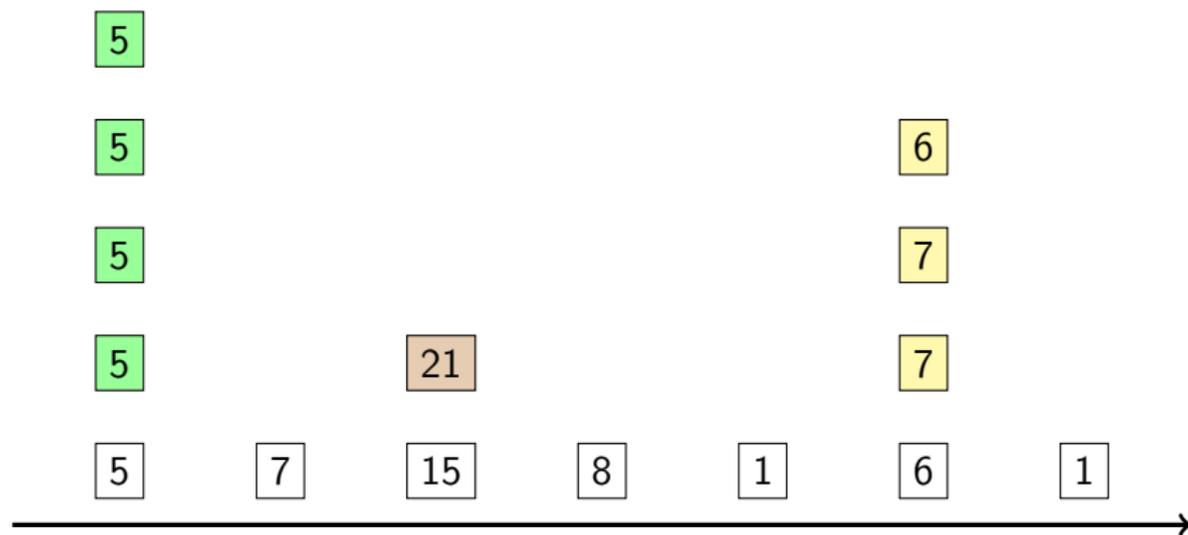
Pour la première fois, **EM et échantillon de l'EEC sont coordonnés** :

- Facilite la collecte
 - Optimisation de l'étendue des grappes et secteurs de l'EEC, et des unités primaires de l'EM
 - Constitution des unités de coordination (UC) regroupant les UP tirées et les secteurs
- En conservant une bonne précision
 - Abandon du recensement de la population pour l'EM
 - Amélioration des variables d'équilibrage mobilisées
 - Apport de l'équilibrage spatial
 - Équilibrage indirect des UC via l'équilibrage des UP

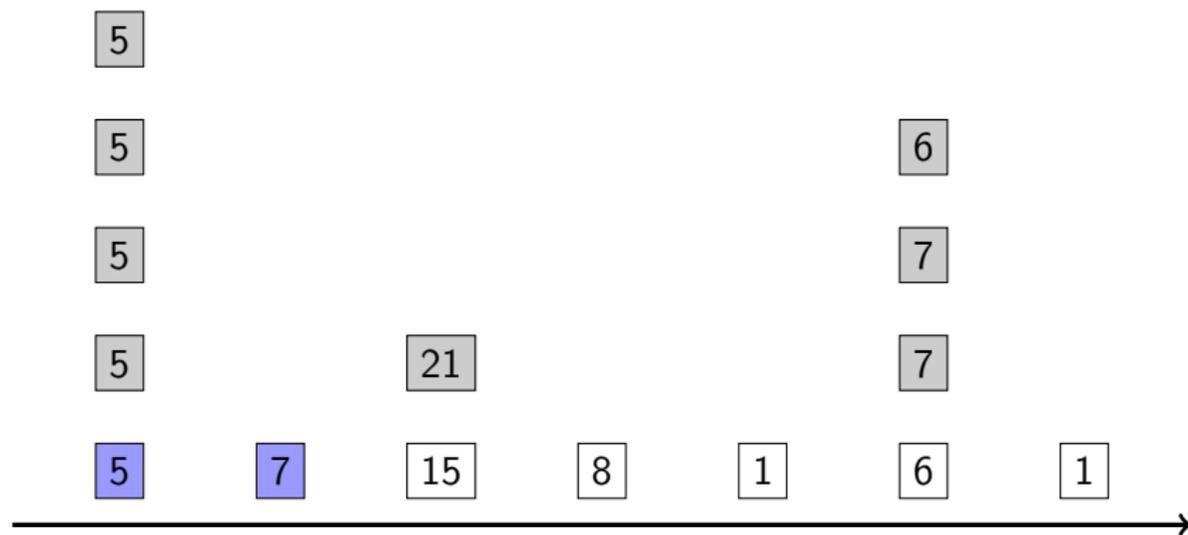
Merci pour votre attention !



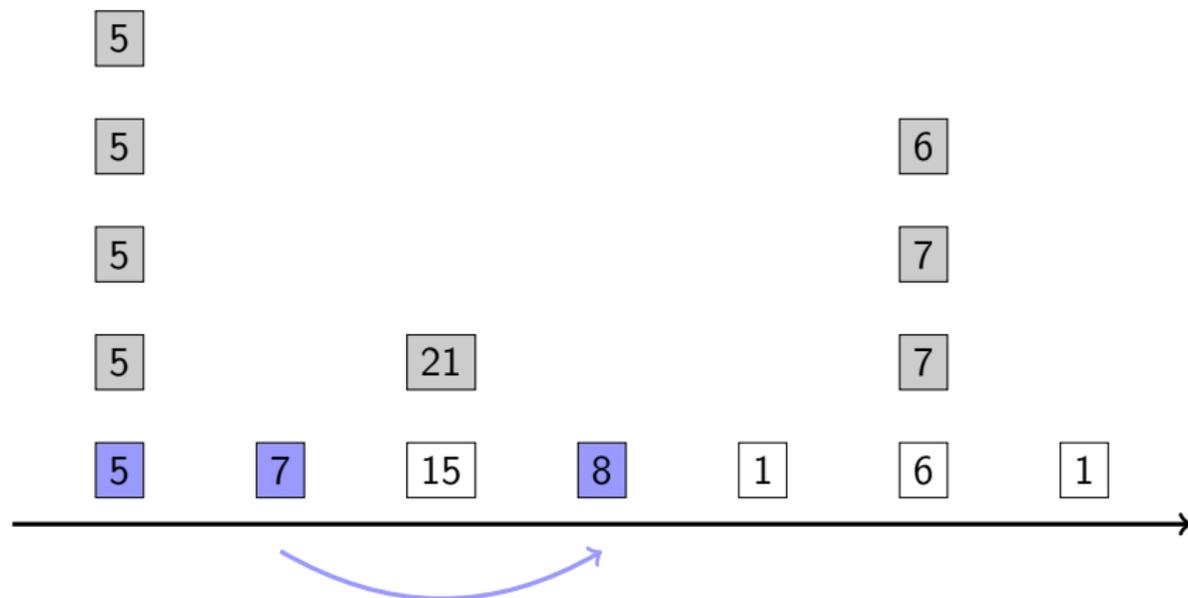
Étape 1 : Regroupement dans les immeubles



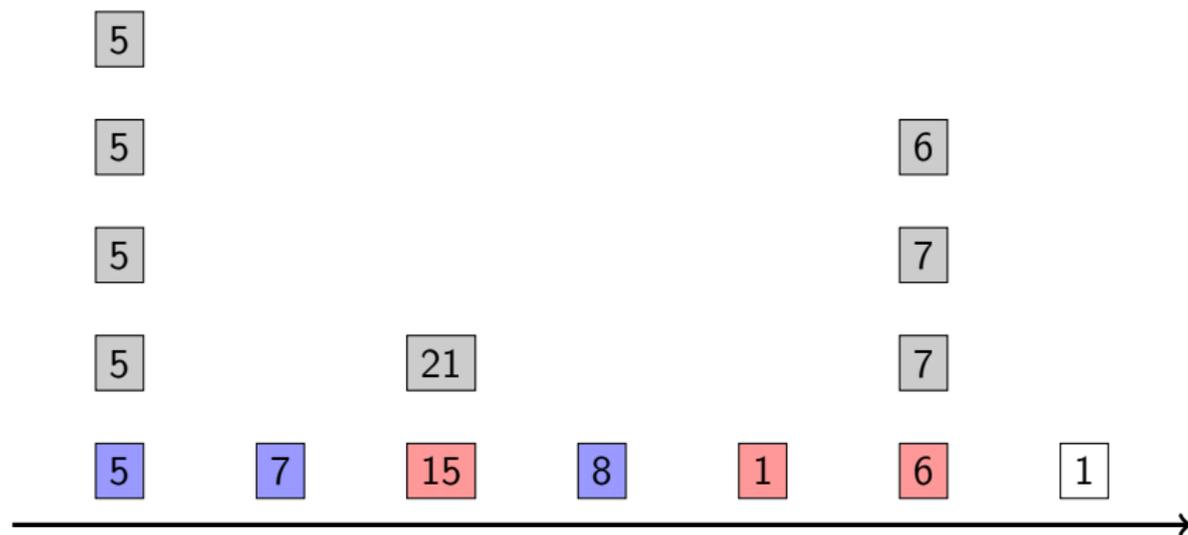
Étape 2 : Regroupement entre immeubles



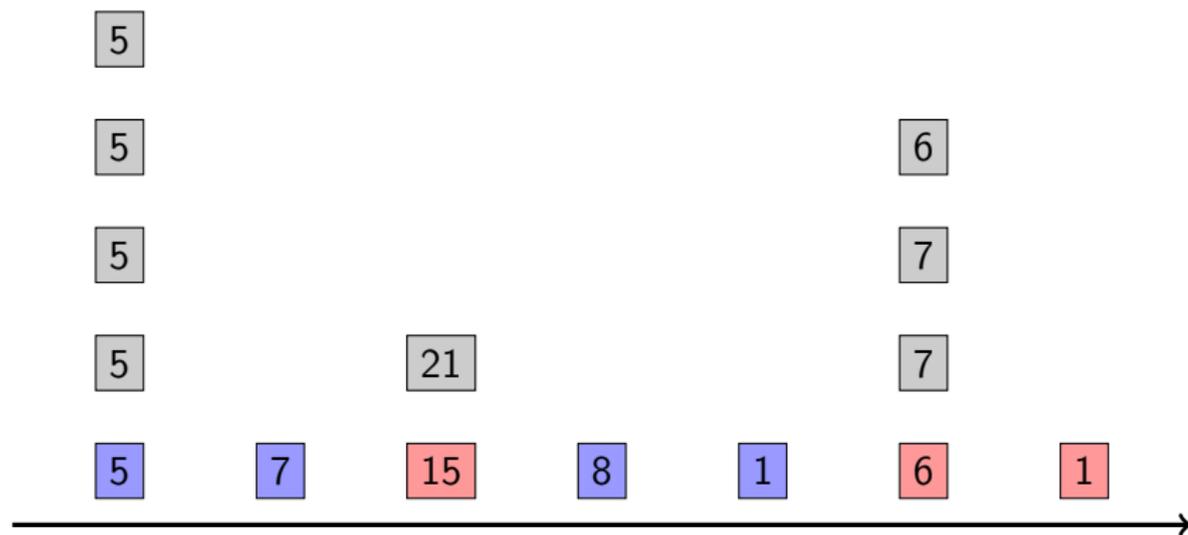
Étape 2 : Regroupement entre immeubles



Étape 2 : Regroupement entre immeubles



Étape 3 : Consolidation



Les variables transformées

- Création de variables transformées pour intégrer dans l'équilibrage des UP des variables de l'EEC qui permettront aux UC tirées d'être précises sur ces variables.
- On se place dans le contexte de la méthode généralisée de partage des poids.
- Les UC $j(i)$ sont tirées indirectement par leurs UP i .
- On pondère le lien $\theta_{i,j(i)}$ par le nombre de logements, proportionnel à la probabilité de sélection.

- Pour chaque variable d'équilibrage X au niveau UC, on construit les variables associées \tilde{X}_i au niveau UP :

$$\tilde{X}_i = \theta_{i,j(i)} X_{j(i)}$$

- Alors, si l'échantillon d'UP est équilibré sur \tilde{X} , on a un échantillon d'UC équilibré sur X :

$$\sum_{j \in S_{UC}} w_j^{UC} X_j = \sum_{j \in S_{UC}} \sum_{i | j(i)=j} \theta_{i,j} w_i^{UP} X_j = \sum_{i \in S_{UP}} w_i^{UP} \tilde{X}_i$$