

Précarité énergétique - Rapport 06/05/2019 (SEMAINE 1)

1. Tâches effectuées

Tout d'abord, nous avons lu la documentation qui nous a été mise à disposition et à l'aide de celle-ci nous avons sélectionné les variables pertinentes parmi l'enquête EU-SILC réalisée afin de visualiser les données concernant le thème de la précarité énergétique.

=> tableau Excel (en annexes).

Nous pourrions au travers de statistiques descriptives représenter ces données.

- **Proposition :**

Nous avons fait le choix de réaliser notre application de visualisation des données avec le langage R et le package shiny :

- Réalisation de graphiques traditionnels (boxplot, barplot, histogramme, nuage de point, série chronologique, ...)
- Réalisation de plusieurs cartes dynamiques.
 - Une carte de l'Europe à l'échelle des pays
 - Une carte nationale pour chaque pays à l'échelle des régions (NUTS2 au mieux)

Ainsi, pour cela, nous avons commencé à s'intéresser aux données

- Mise en place d'une importation des données de manière automatique et flexible. Nous avons 4 fichiers .csv pour chaque pays et chaque année.
 - Household Register (D),
 - Personal Register (R),
 - Household Data (H),
 - Personal Data (P).

Le formatage et la structuration des fichiers est harmonisée de façon à ce qu'on puisse importer l'ensemble des données (tous les pays et toutes les années) sans avoir à dézipper les archives.

- Nous nous sommes limités aux variables principales nous permettant de mesurer "la précarité énergétique". Il en ressort 3 :
 - ability to keep adequately warm (HH050)
 - arrears on utility bills (HS021)
 - Leaking roof, damp walls/floors/foundation or rot in window frames ... (HH040)

Nous nous sommes également limités à l'année 2013 pour l'instant.

Nous avons eu pour objectif de visualiser l'étendue de ces variables sur des cartes à l'aide du package leaflet :

- Moyenne d'une variable pour chaque région NUTS2 pour tester dans un premier temps.

Codage des variables :

Après avoir importé les données et après avoir choisi des variables d'étude, nous avons créé un tableau associant des données à projeter pour chaque région.

En ce qui concerne les variables d'études, certaines étaient binaires comme :

- HH050 : Ability to keep home adequately warm (YES = 1 / NO = 2).
- HH040 : Leaking roof, damp walls/floors/foundation, or rot in window frames or floor (YES = 1 / NO = 2).

Ces variables sont recodées de façon à avoir (YES = 1 et NO = 0).

D'autres, en revanche, qualitatives non binaires.

- HS021 : Arrears on utility bills.
 - o 1 = Yes, once
 - o 2 = Yes, twice or more
 - o 3 = No

Nous avons choisi de transformer cette variable multi-modale en variable binaire en regroupant les modalités "Yes, once" et "Yes, twice or more".

- (YES = 1 / NO = 0)

Pour chaque région, nous avons effectué la moyenne de ces variables afin d'avoir les proportions des ménages ayant répondu YES ou NO pour chaque question.

Selon la tournure de la question on choisira d'effectuer $1 - \text{MOYENNE}()$ ou juste $\text{MOYENNE}()$.

Partie Cartographie :

Après diverses recherches sur internet, nous avons décidé d'utiliser les packages leaflet et geojsonio pour faire nos cartes :

- Données GIS : les données GPS des régions ont été récupérées sur le site d'Eurostat, avec la nomenclature NUTS2 en vigueur en 2013 (la nomenclature en vigueur en 2016 n'est plus la même !).
 - o *Remarque 1 : différentes précisions des tracés des frontières sont proposées, on utilise, pour l'instant, celle la moins précise pour alléger les temps de calculs.*
 - o *Remarque 2 : l'ordre des régions diffère selon le choix de la précision, bien faire attention à bien relancer l'ensemble du code lié au traitement des données sinon des régions seront affectées de valeurs étranges.*

Nous avons remarqué qu'il y avait des normes différentes en ce qui concerne la représentation des régions. En effet, à partir de 2016, les labels des régions ont changé. Les études ayant été réalisées utilisent l'ancienne norme NUTS (celle en vigueur en 2013), elle contient les anciens labels. Nous avons donc utilisé les données cartographiques de 2013.



Figure 1. Rendu-test de la carte à l'échelle des régions.

2. Problèmes

Après étude des fichiers excel, nous avons remarqué que les données fournies issues de l'enquête EU-SILC contiennent de nombreuses données manquantes qui vont nous gêner quant à la représentation de celle-ci :

1. Nous n'avons pas de données pour les pays suivants : Pologne, Portugal, République Tchèque et la Suède.
2. Lors du questionnaire, certaines régions ont été enregistrées sous la norme NUTS1 alors que la carte est en NUTS2 (Autriche et Hongrie).
3. En Allemagne, Slovaquie et Pays-Bas : aucune région n'a été précisée par les ménages (la variable identifiant la région du ménage existe mais n'est pas remplie du tout).

3. Solutions envisagées :

1. Nous ne pouvons rien faire.
2. La valeur prise par la région NUTS1 est injectée dans toutes les sous-régions NUTS2 (problème traité comme indiqué).
3. Nous pourrions représenter ces pays à l'échelle internationale (sur la carte mondiale). Toutefois, nous ne pourrions pas représenter ces pays à l'échelle nationale avec les valeurs de chaque variable pour chaque région.

4. Poursuite

Nous envisageons d'effectuer durant la SEMAINE 2 de :

- traiter les données après importation selon la décision qui sera prise suite à la réunion.
- réaliser une maquette de ce à quoi pourrait ressembler l'application.
- commencer les statistiques descriptives.