QuantiLille 2023

Module Multiniveau - Journée géographique

Clémentine Cottineau & Julie Vallée 30 Juin 2023

Icebreaker

Prénom

A quoi me sert le multiniveau?

Espace géographique préféré

Multiniveau & géographie - Partie théorique

Penser les niveaux comme des espaces géographiques

Le niveau d'analyse du phénomène étudié (niveau 1)

Un phénomène propre à des personnes

- réussite scolaire
- état de santé
- statut d'activité (employé ou au chômage)
- etc

Un phénomène propre à un espace géographique

- pollution
- prix du vin (qui y est produit)
- densité de population
- taux de chômage
- croissance de la population résidente
- etc

=> réflexion sur la maille géographique (en termes de forme et d'échelle) à considérer

Et quel niveau 2?

Question thématique:

A quelle(s) échelle(s) géographique(s) sont organisés les acteurs et les processus qui peuvent influencer les phénomènes?

Contrainte pragmatique:

Quelle est la maille géographique (en termes de forme et d'échelle) à laquelle les variables explicatives sont disponibles ?

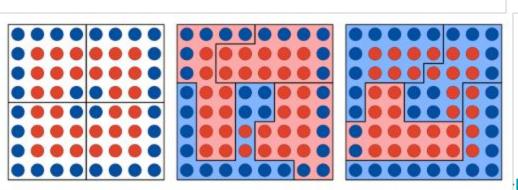
Considération statistique:

Quelle est l'influence de la maille géographique sur les résultats? Quels biais ?

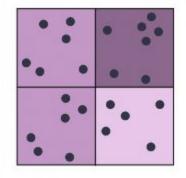
LE MAUP

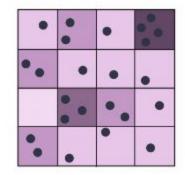
Les biais statistiques peuvent venir de:

- l'échelle d'agrégation
- la forme des mailles









Effets de contexte et échelle(s) géographique(s)

Galster, G. C. (2011). The mechanism (s) of neighbourhood effects: Theory, evidence, and policy implications. In *Neighbourhood effects research: New perspectives* (pp. 23-56). Dordrecht: Springer Netherlands.

Social-Interactive Mechanisms : Social contagion, collective socialization, Social Networks, Cohesion and control, Competition, Relative deprivation, parental mediation

Environmental Mechanisms: Exposure to violence, Physical surroundings, toxic exposure

Geographical Mechanisms: Spatial Mismatch, Public services

Institutional Mechanisms: Stigmatization, Local institutional resources, Local market actors

> Quelle(s) échelle(s) pertinente(s) pour saisir l'organisation des acteurs et des processus?

Modéliser les différences du phénomène au niveau 2

<u>Étape 1</u>: Établir l'existence des disparités entre les unités spatiales de niveau 2 par l'étude de la variance de niveau 2

=> Mesurer l'ampleur de la variation du phénomène observé entre les unités spatiales

Étape 2 : Voir dans quelle mesure ces disparités entre les unités spatiales de niveau 2 sont liées...

- à la répartition spatiale des caractéristiques de niveau 1 = l'effet de composition
- à l'influence des caractéristiques de niveau 2 sur les comportements des individus = l'effet du contexte géographique

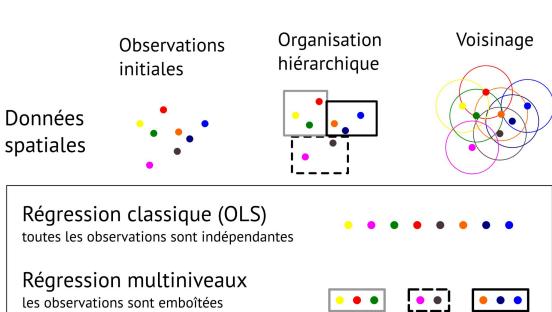
Ce qui change pour la modélisation multiniveau quand le niveau 2 est un espace

 On peut cartographier les résultats pour rechercher des explications (et variables) géographiques

- Les unités de niveau 2 (ou plus) ont des propriétés géographiques (contiguïté, proximité, etc.)

Et si la distribution géographique du phénomène ne correspond pas aux mailles?

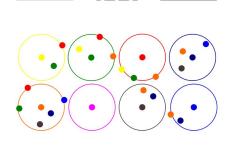
Les modèles GWR sont plus appropriés pour prendre en compte l'auto-corrélation spatiale



dans des entités indépendantes

Régression à pondération géographique (GWR)

les coefficients sont estimés localement en fonction du voisinage des observations



Multiniveau & géographie - Partie empirique :

une analyse pas à pas des résultats sous R

Cas d'étude: le vignoble bourguignon

Focus sur la Côte d'Or

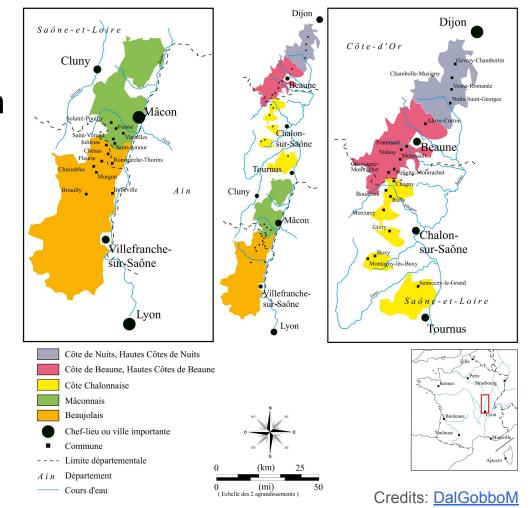
= Côte de Beaune + Côte de Nuits.

Des vins reconnus internationalement,

Vendus aux enchères plusieurs milliers d'euros,

Distingués par des AOC (territoire défini),

Nom de "Bourgogne" comme signal de qualité.



Les données Niveau 1 (vignobles)

On a dans un tableau les caractéristiques de 2391 vignobles en termes de:

- prix moyen du vin (millésimes 1998, 2002 et 2003)
- qualité pédologique et météo (radiations solaires, pluie, etc.)
- appellation d'origine contrôlée en 5 niveaux (Côteaux bourguignons < Bourgogne < Village < Premier cru < Grand cru.)
- surface

Sources:

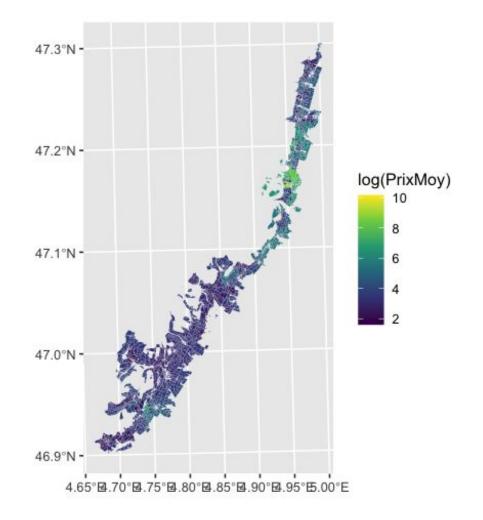
- Ay J.S., Hilal M., 2020, « Les déterminants naturels et politiques des AOC viticoles de Côte-d'Or », Cybergeo: European Journal of Geography, document 973, DOI: https://doi.org/10.4000/cybergeo.36443
- https://www.hachette-vins.com/guide-vins/actualite-vin/487/les-meilleurs-millesimes-des-vins-de-france/

Les données Niveau 1

Le prix:

Grande autocorrélation spatiale positive

=> Les vignobles proches géographiquement ont des prix moyens similaires.



Les données Niveau 2 (Communes)

Les 2391 vignobles sont regroupés dans 31 communes de Côte d'Or. Ces communes peuvent être caractérisées en termes de:

- superficie
- population
- côte d'appellation
- hiérarchie administrative

Sources:

- https://geo.data.gouv.fr/en/datasets/cac9f2c0de2d3a0209af2080854b6f6a7ee3d9f4

> Ouvrir un nouveau projet sur RStudio

- Créer un Dossier "Data"
- Télécharger les données sur edu.nl/jd6ej



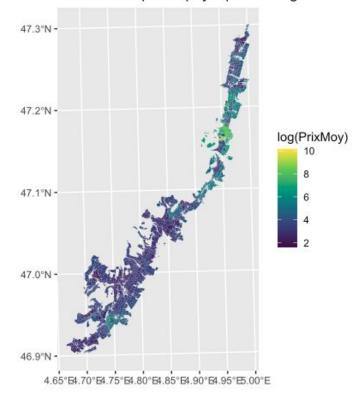
- Dé-zipper le dossier et déplacer son contenu dans le dossier "Data"
- Créer un script "Vin_Bourgogne.R" à la source du projet

 "Voir" s'il y a des disparités spatiales dans le phénomène étudié

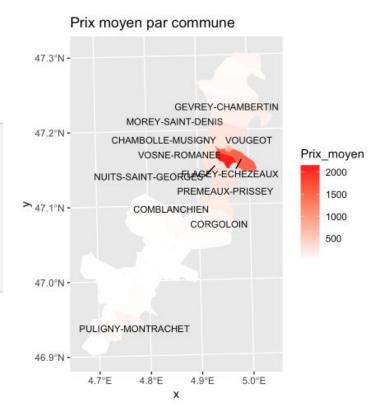
Avec ggplot, utilisation de geom_sf pour les objets spatiaux.

```
ggplot(data = vignobles) +
geom_sf(aes(fill = log(PrixMoy)), lwd=0.01, colour="white") +
scale_fill_viridis_c(option = "viridis")
```

Niveaux de qualité physique des vignobles



Agréger l'information au niveau 2

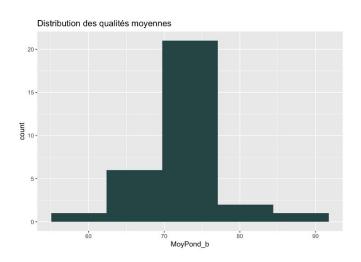


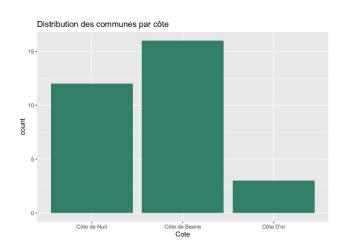
Modèle multiniveau vide, avec intercept aléatoire: Y ~ (1 | SpatialUnit)

```
mnnull <- lmer(LogPrix ~ 1 + (1 | LIBCOM), data=vignobles final cr, REML=F, na.action=na.omit)
summary(mnnull)
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
    method [lmerModLmerTest]
## Formula: LogPrix ~ 1 + (1 | LIBCOM)
     Data: vignobles final cr
                BIC logLik deviance df.resid
       AIC
    4487.1 4504.4 -2240.5 4481.1
## Scaled residuals:
      Min
               10 Median
## -5.4605 -0.3560 0.0610 0.4361 4.9734
## Random effects:
## Groups Name
                       Variance Std.Dev.
## LIBCOM (Intercept) 0.8638
                                0.9294
## Residual
                        0.3611 0.6009
## Number of obs: 2377, groups: LIBCOM, 31
## Fixed effects:
              Estimate Std. Error df t value Pr(>|t|)
## (Intercept) 0.1520
                           0.1677 30.9052 0.907
                                                    0.372
```

Test ANOVA pour comparer variances inter et intra unités spatiales

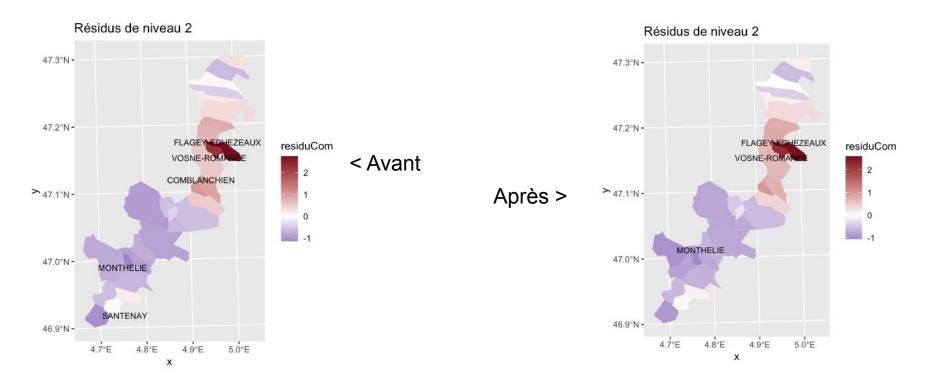
Déterminer ce qui dans les différences des prix entre *vignobles* (niveau 1) relève des *communes* (niveau 2), par exemple:





Variables explicatives qui diminuent la variance inter-unités spatiales: Quality + Surface + Source + AOC

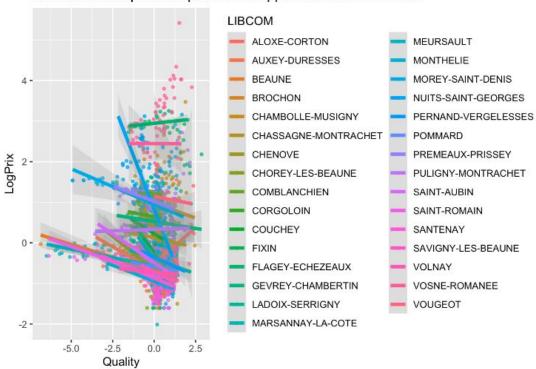
Variables explicatives qui diminuent la variance inter-unités spatiales: Quality + Surface + Source + AOC



Variables explicatives statistiquement associées à la « variable à expliquer » mais sans faire diminuer la variance inter unités spatiales. Comment l'interpréter?

Étape 3. Des interactions aux modèles avec "pentes aléatoires"

Relation entre qualité et prix selon l'appartenance communale



Étape 3. Des interactions aux modèles avec "pentes aléatoires"

-4.740e-02 7.855e-02 3.139e+01 -0.603 0.550536

3.694e-01 8.405e-02 3.036e+01

7.577e-01 2.616e-01 3.176e+01

1.298e+00 1.663e-01 3.004e+01

MeanQuality

CoteCôte D'or

CoteCôte de Nuit

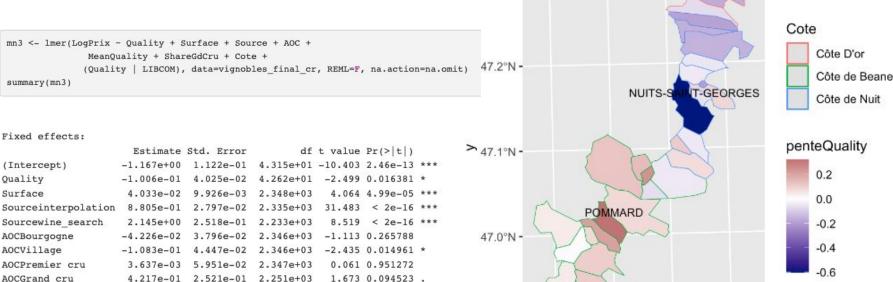
ShareGdCru

Distribution des effets aléatoires de la variable Quality selon les communes

4 9°F

5.0°F

4.8°E



4.395 0.000125 ***

2.896 0.006783 **

7.807 1.03e-08 ***

47.3°N -

46.9°N

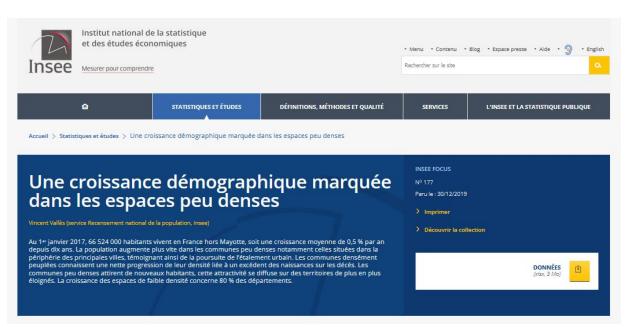
4 7°F

Attention!

Bien distinguer

- modèle avec intercept aléatoire Random-intercept model : Y ~ 1 + X + (1 | SpatialUnits)
 - => l'intercept (1) varie par (|) groupe (SpatialUnits)
- modèle avec intercept et slope *aléatoires* Random-intercept and slope model : Y ~ 1 + X + (1 + X | SpatialUnits)

TD



<u>Objectif</u>

Confirmer et enrichir certains résultats de cette fiche grâce à des régressions multiniveaux

https://www.insee.fr/fr/statistiques/4267787

MISE EN PLACE

Charger des packages

```
library(lme4) # pour la modélisation multiniveau
library(ImerTest) # pour l'évaluation des modèles
library(sf) # pour les objets spatiaux (cartes)
library(tidyverse) # pour la manipulation des données
library(readxl) # pour ouvrir les fichiers xlsx
library(languageR)
library (dplyr)
```

Télécharger données fiche Insee et créer dataframe pop_comm

https://www.insee.fr/fr/statistiques/fichier/4267787/if177.xlsx

Fonction read_excel (Choisir "Figure complémentaire 1")

Note : grille communale de densité (1: commune densément peuplée ; 2: commune de densité intermédiaire ; 3: commune peu dense ; 4: commune très peu dense)

A Figure complémentaire 1	B - Population 2017 de	C 7 des communes	et évolution entre 2007 et 2017	E	F	G	Н	Taux de variation annuel de la population dú au solde migratoire apparent 2007-2017 (en %)	Variation annuelle du nombre d'habitants par km² 2007-2017 (en%)
Région	Départemen	Code t commune Insee	Commune	Grille de densité	Population municipale légale 2017 (en%)	Taux de variation annuel de la population 2007 2017 (en %)	Taux de variation annuel de la population dú au solde naturel 2007-2017 (en %)		
Auvergne-Rhône-Alpes	01	01001	L'Abergement-Clémenciat	3- commune peu dense	776	-0,4	0,4	-0,8	-0,2
Auvergne-Rhône-Alpes	01	01002	L'Abergement-de-Varey	4- commune très peu dense	248	2,4	1,0	1,5	0,6
Auvergne-Rhône-Alpes	01	01004	Ambérieu-en-Bugey	2- commune de densité intermédiaire	14 035	1,0	0,8	0,2	5,3

Nettoyer & Vérifier

Simplifier la variable *densite* (par exemple "3 - commune peu dense" => 3)

Supprimer les communes avec NA pour le taux de variation (*txvar*)

Vérifier la population totale en 2017 (*P17_POP*) dans les données importées (/ fiche Insee)

Au 1er janvier 2017, 66 524 000 habitants vivent en France hors Mayotte (sources). Entre 2007 et 2017, la population a augmenté de 0,5 % par an en moyenne, soit 292 400 personnes supplémentaires chaque année.

Créer variable P07_POP à partir de txvar et P17_POP

Quelle formule?

```
txvar <- ((pop2017 / pop2007)^(1/n) - 1)*100 (avec n=10 car 10 ans entre les deux valeurs)

txvar <- (pop2017 / pop2007)^0.1 - 1)*100)

=> tvar/100 +1 = pop2017/pop2007)^0.1

=> (tvar/100 + 1)^10 = pop2017/pop2007

=> pop2007 = pop2017 / (tvar/100 + 1)^10
```

pop_comm\$P07_POP <-pop_comm\$P17_POP / (((pop_comm\$txvar/100) + 1)^10)

Regarder la base de données

*	REG \$	DEP =	CODGEO	Commune	densite *	P17_POP	txvar =	P07_POP
1	Auvergne-Rhône-Alpes	01	01001	L'Abergement-Clémenciat	3	776	-0.4	808
2	Auvergne-Rhône-Alpes	01	01002	L'Abergement-de-Vare y	4	248	2.4	196
3	Auvergne-Rhône-Alpes	01	01004	Ambérieu-en-Bugey	2	14035	1.0	12706
4	Auvergne-Rhône-Alpes	01	01005	Ambérieux-en-Dombes	3	1689	0.9	1544
5	Auvergne-Rhône-Alpes	01	01006	Ambléon	4	111	-1.2	125
6	Auvergne-Rhône-Alpes	01	01007	Ambronay	3	2726	1.9	2258
7	Auvergne-Rhône-Alpes	01	01008	Ambutrix	2	752	1.5	648
8	Auvergne-Rhône-Alpes	01	01009	Andert-et-Condon	3	330	0.2	323
9	Auvergne-Rhône-Alpes	01	01010	Angle fort	3	1115	2.0	915
10	Auvergne-Rhône-Alpes	01	01011	Apremont	3	376	1.2	334
11	Auvergne-Rhône-Alpes	01	01012	Aranc	4	326	1.3	286
12	Auvergne-Rhône-Alpes	01	01013	Arandas	4	144	-0.9	158
13	Auvergne-Rhône-Alpes	01	01014	Arbent	2	3367	-0.2	3435
14	Auvergne-Rhône-Alpes	01	01015	Arboys en Bugey	4	644	0.9	589

Importer la couche géographique des départements

URL: https://geoservices.ign.fr/adminexpress

Télécharger le dossier "ADMIN-EXPRESS-COG-CARTO" édition 2021 par territoire France Métropolitaine

<a href="https://wxs.ign.fr/x02uy2aiwjo9bm8ce5plwqmr/telechargement/prepackage/ADMINEXPRESS-COG-CARTO_SHP_TERRITOIRES_PACK_2021-05-19\$ADMIN-EXPRESS-COG-CARTO_3-0_SHP_LAMB93_FXX_2021-05-19/file/ADMIN-EXPRESS-COG-CARTO_3-0_SHP_LAMB93_FXX_2021-05-19.7z

Le dézipper

Importer la couche DEPARTEMENT.shp dans R avec la fonction st_read

```
dep_fr <- st_read
     ('Data/ADMIN-EXPRESS-COG-CARTO_3-0__SHP_LAMB93_FXX_2021-05-19/ADMIN-EXPRESS-COG-CARTO/1_
     DONNEES_LIVRAISON_2021-05-19/ADECOGC_3-0_SHP_LAMB93_FXX/DEPARTEMENT.shp',
     quiet = TRUE)</pre>
```

Faire une carte "vide" des départements

```
ggplot(data = dep_fr) +
    geom_sf() +
    ggtitle("Les départements français")
```

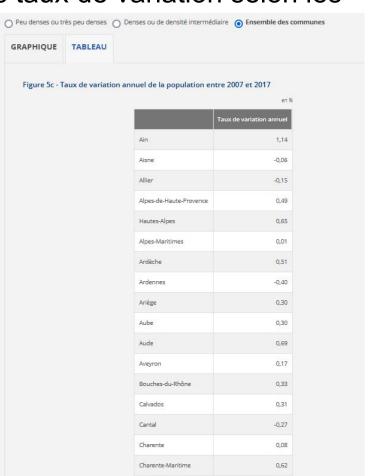


Question A. Visualiser les différences des taux de variation selon les

départements

=> Faire un <u>tableau</u> des taux de variation par département

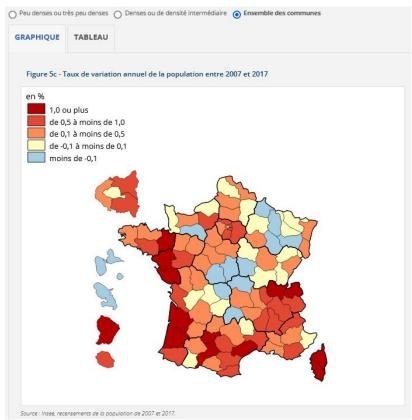
cf. Tableau 5c de la Fiche Insee



Question A. Visualiser les différences des taux de variation selon les départements

=> Faire une <u>carte</u> des taux de variation par département

cf. Carte 5c de la Fiche Insee



Question A. Visualiser les différences des taux de variation selon les départements

- => Faire un modèle multiniveau (vide) des taux de variation des communes avec :
 - les communes en niveau 1
 - les départements en niveau 2

###

- Calculer et interpréter les variances de niveau 1 et de niveau 2 ?
- Faire une carte des résidus de niveau 2 (département) pour le modèle vide. Que peut-on en dire ?

Question B. Est-ce que le taux de variation des communes est associé à leur densité ?

=> Faire un <u>tableau</u> des taux de variation selon la densité des communes

cf Figure 2



Question B. Est-ce que le taux de variation des communes est associé à leur densité ?

=> Faire un <u>modèle multiniveau</u> avec la densité des communes comme variable explicative (effet **fixe**)

###

 Observer les coefficients associés à la densité des communes. Que peut-on en conclure ?

Question C. Si on classe les communes en deux catégories selon leur densité, quelle classification est la plus pertinente à intégrer au modèle multiniveau ?

- => Si on fait comme dans la fiche Insee?
- 1: communes densément peuplées & de densité intermédiaire
- 2: communes peu denses & très peu denses
- => Si on isole les communes peu denses ?
- 0: communes densément peuplées & de densité intermédiaire & très peu denses
- 1: communes peu denses

###

- Observer les coefficients associés à la densité des communes en deux catégories. Que peut-on en conclure ?
- Si on faisait une cartographie par département des coefficients liés à la densité, que devrait-on observer ?
- Faire une cartographie par département des résidus de niveau 2 (département). Qu'observe-t-on ?
 Quelle(s) autre(s) variable(s) explicative(s) pourrait-on ajouter au modèle ?

Question D. Est-ce que le taux de variation des communes varie aussi selon la présence d'une grande ville dans le département ?

Créer variable au niveau du département (présence/absence d'une commune > 150 000 hb.) et la rapatrier au niveau de la table communes

Intégrer cette variable explicative de niveau 2 dans le modèle multiniveau

###

- Observer le coefficient associé à la présence/absence dans le département d'une commune > 150 000 hb. Que peut-on en conclure ?
- Faire une cartographie par département des résidus de niveau 2 (département).
 Qu'observe-t-on ?

<u>Question E.</u> Est-ce que la différence inter-départementale dans les taux de variation s'explique par....

... la composition des communes des départements en termes de densité (effets de composition) ?

... et par la présence d'une grande ville dans le département (effet de contexte) ?

###

 Comparer l'évolution des variances de niveaux 1 et 2 entre les différents modèles <u>Question F.</u> L'effet de la densité des communes varie-t-il selon les départements ?

Faire un modèle multiniveau avec la densité des communes comme variable explicative (effet **aléatoire**).

###

- Faire une cartographie par département des coefficients liés à la densité. Qu'observe-t-on ?
- Comparer l'évolution des variances de niveaux 1 et 2 entre les différents modèles