

Avanies et organisation administrative française

THOMAS SOUBIRAN

CERAPS (UMR 8026 CNRS - Université de Lille)

<https://numa.hypotheses.org>

**École d'été méthodes
quantitatives en sciences sociales**

Lille, 28 juin 2021

- ▶ **prolégomènes** pour préparer le terrain avant de rentrer dans le vif du sujet
- ▶ remarques plus ou moins désordonnées sur **les données** couramment utilisées
- ▶ et **les surprises** qu'elles peuvent parfois réserver. . .
- ▶ . . . souvent liées **aux idiosyncrasies** de l'organisation administrative française
- ▶ façon de présenter à la fois :
 - ▶ l'objet et certaines données
 - ▶ et comment travailler avec elles
 - ▶ notamment les choses auxquelles il faut faire attention

- ▶ traiter **l'information géographique** peut nécessiter toucher à des domaines **très divers** :

- ▶ mathématiques
- ▶ informatique
- ▶ sémiologie
- ▶ ...

et ce, à des degrés tout aussi variables **en fonction des applications**

- ▶ mais aussi, de façon peut-être moins attendue, à

- ▶ **l'analyse des politiques publiques**
- ▶ et **de l'action publique**
- ▶ réglementations en vigueur, normes, nomenclatures, ...

- ▶ parce qu'on se retrouve souvent à devoir se caler sur **un découpage administratif** ou un autre
- ▶ et utiliser des données produites par **la puissance publique**

Exemple :

- ▶ la plus part des États —tous ?— **réglementent l'information géographique**
 - ▶ pour **coordonner** l'action publique
 - ▶ mais aussi parce que c'est un acte de **souveraineté** sur un territoire
- ▶ en définissant, p. ex., **une projection cartographique officielle**
- ▶ en France, **le système national de références géographiques** est défini par le décret **n° 2000-1276 du 26 décembre 2000**

portant application de l'article 89 de la loi n° 95-115 du 4 février 1995 modifiée d'orientation pour l'aménagement et le développement du territoire relatif aux conditions d'exécution et de publication des levés de plans entrepris par les services publics
- ▶ couramment appelé « **Lambert 93** »

Lambert, car elle utilise la projection conique conforme de J. Lambert, et 93 car la norme a été élaborée en 1993
- ▶ c'est cette norme que de nombreuses **entités publiques** —comme l'IGNF— utilisent
- ▶ et c'est donc dans cette projection que sont diffusés de nombreux **fonds de carte**
- ▶ **souvent** mais pas **toujours**

- ▶ **Note** : tout cela sera développé **demain**
- ▶ **géodésie** : \simeq mesure de la terre
- ▶ qui permet **de positionner des points**
- ▶ au moyen **de coordonnées géographiques**

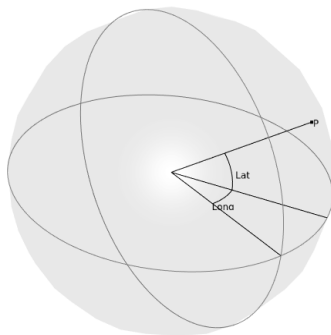
- ▶ **latitude** λ : angle **avec l'équateur**

axe nord-sud

- ▶ **longitude** ϕ : angle **avec un méridien de référence**

- ▶ comme le méridien de Greenwich
- ▶ axe est-ouest

Coordonnées géographiques



- ▶ **fonds de carte** : ensembles de primitives géométriques

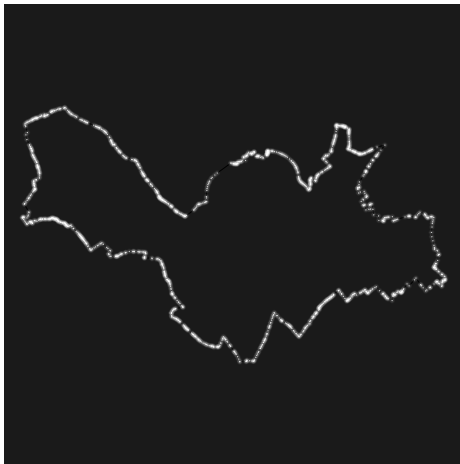
points, (poly)lignes, polygones

- ▶ qui représentent un objet géographique

Exemple :

- ▶ les adresses sont représentées par **des points**
 - ▶ coordonnées ϕ et λ —coordonnées polaires en degré—
 - ▶ ou x et y dérivés longitude, latitude —coordonnées cartésiennes en mètres—
 - ▶ et, éventuellement, le relief z
- ▶ les limites administratives —communes, cantons, . . .— sont représentées par **des polygones**
- ▶ qu'on peut définir comme des **~listes de points**

Exemple : carte des limites administratives de la commune de Lille



polygone composé de 1 587 points

- ▶ les mesures —géodésiques— des coordonnées sont réalisées sur **un ellipsoïde**
- ▶ soit une ellipse à **trois dimensions**
la terre n'est pas ronde. . .
- ▶ mais les cartes n'ont que **deux dimensions**
- ▶ le passage de l'un à l'autre s'appelle **une projection**
- ▶ qui induit **des déformations** plus moins marquées
- ▶ avec une dimension en moins, **on perd nécessairement quelque chose. . .**

► les projections conservent **de façon exclusive** :

► soit **les angles**

projections conformes

► soit **les aires**

projections équivalentes

► soit **les distances**

dans certaines direction

► le choix d'une projection **dépend aussi** :

► de **l'échelle**

globe, continent, pays, . . .

► de **la latitude**

► le Lambert 93 ne concerne donc que **la Métropole**

► les projections les collectivités l'outre-mer utilisent l'UTM (Universal Transverse Mercator)

► ce **document** de l'IGNF liste les systèmes de référence en vigueur

- ▶ le Lambert 93 est une projection **conforme**
- ▶ qui combine deux **types de projections** :

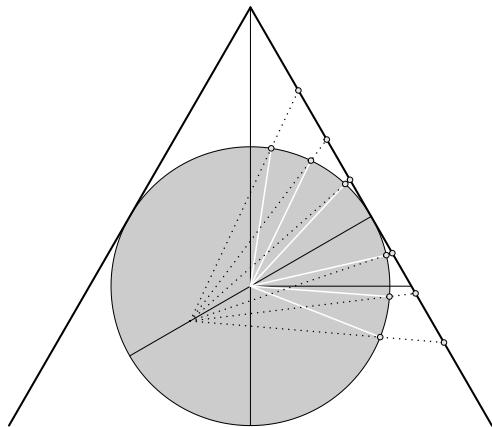
- ▶ projection conique

souvent utilisées pour les régions intermédiaires entre l'équateur et le pôle

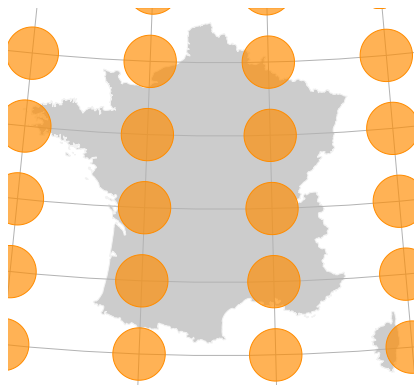
- ▶ projection sécante

qui permet de réduire les distortions

Projection conique



Carte de France projetée dans le Lambert 93



avec ses indicatrices de Tissot

- ▶ toujours s'assurer que :

- ▶ **les données sont projetées**

pas des coordonnées GPS

- ▶ et, le cas échéant, **dans quelle projection**

- ▶ on peut encore trouver

même si c'est devenu rare

- ▶ des coordonnées projetées avec **l'ancien système de projection** utilisé avant l'entrée en vigueur de la norme actuelle

qui utilise aussi une projection conforme de Lambert mais avec des paramètres différents

- ▶ il y a aussi **une variantes locale** de la norme qui divise le territoire métropolitaine **en 9 zones**

pour corriger les altérations linéaires pour les cartes à échelle plus réduite

- ▶ parfois les coordonnées sont projetées dans **un référentiel différent**

Exemple : la projection Web Mercator qui utilisée par OSM et d'autres services en ligne

- ▶ parfois aussi, les données n'ont tout simplement pas **de meta-données**

Production de l'information géographique

- ▶ utilisation **très précoce** de l'informatique **en géomatique**

nombre d'applications nécessitent en effet des calculs importants

- ▶ les développements logiciels ultérieurs et, plus particulièrement, l'apparition des **SIG** dans les années 1970

systèmes d'information géographiques : ~ environnements informatiques intégrés pour le traitement de l'information géographique

- ▶ ont facilité **la production** et **l'utilisation** de l'information géographique
- ▶ aussi lié au développement **d'infrastructures**

réseau GPS puis GLONASS —Russie—, Beidou —Chine—, Galileo —UE—

- ▶ **tout un chacun** peut produire de l'information géographique

ne serait-ce qu'à l'insu de son plein grès dès qu'un téléphone borne, *a fortiori* lorsqu'on utilise un « smart » phone ou le wifi

- ▶ utilisation d'un téléphone pour **faire** de la géolocalisation et pas seulement pour **être** géolocalisé

Exemple : enquête de Jacobo Grajales

Production de l'information géographique

- ▶ **multiplication** des producteurs et diffuseurs

- ▶ public —politiques d'ouverture des données—
- ▶ privé —plutôt associatifs. . .—
- ▶ particuliers, . . .

- ▶ mais, en France, **deux acteurs** publics historiques de poids :

- ▶ **l'IGNF** —Institut National de l'Information Géographique et Forestière—
anciennement IGN
- ▶ **l'INSEE** —Institut national de la statistique et des études économiques—

- ▶ de par, à la fois

- ▶ **le volume** de leur production
- ▶ mais aussi le caractère **incontournable** de leur production

- ▶ L'IGNF propose **de nombreuses ressources**
- ▶ dont une grande partie sont diffusées selon les termes de la **licence ouverte Etalab 2.0**
 - ▶ vecteur ou raster
 - ▶ cartes, ortho-images, géographie physique et administrative, . . .
- ▶ voir cette **page** pour plus d'information
- ▶ L'IGNF travaille **en partenariat**
- ▶ notamment avec la mission Etalab pour produire **la Base adresse nationale** (BAN)
 - qui renseigne les coordonnées géographiques de plus de 20 M d'adresses en France
- ▶ ou encore avec **l'INSEE**

- ▶ place particulière de l'INSEE dans **la statistique publique**
- ▶ l'INSEE a un rôle fondamental de par les **registres**

RNIPP, SIREN

- ▶ et **les codes** qu'il administre

y compris en matière **d'information géographique**

- ▶ parmi les codes, on peut notamment **citer** :

- ▶ la PCS
- ▶ la NAF —Nomenclature d'activités française—, . . .

- ▶ mais aussi le COG —**Code officiel géographique**—

- ▶ code crée **en 1943** par le Service national des statistiques sous la direction de R. Carmille

comme le NIR —Numéro d'inscription au registre—

- ▶ qui rassemble **les codes et les libellés** des différentes entités géographiques en France

communes, des cantons, des arrondissements, des départements, des régions, des collectivités d'outre-mer, des pays et territoires étrangers

- ▶ le COG est un ensemble **de fichiers**
- ▶ accessibles à partir de **cette page** sur le site de l'INSEE
- ▶ actualisés **touts les ans**
- ▶ voir aussi la page **Géographie administrative et d'étude**

Note : le site de l'INSEE recèle beaucoup de choses mais . . .

- ▶ depuis sa refonte, le site de l'INSEE n'est plus conçu pour être **navigable** mais **indexable**
comme en témoigne la faible profondeur du site
- ▶ qui provient **des difficultés à organiser** la masse grandissante d'information mise en ligne
- ▶ en conséquence de quoi, utilisez **un moteur de recherche web**
- ▶ avec persévérance

- ▶ le COG est constitué d'un ensemble de fichiers qui associent un libellé à un code, un type et, le cas échéant, des entités liées
- ▶ **Exemple** : le fichier `commune2021`

Treize champs :

TYPECOM	COM	REG	DEP	CTCD	ARR	TNCC
COM	01001	84	01	01D	012	5
COM	01002	84	01	01D	011	5
COM	01004	84	01	01D	011	1
COM	01005	84	01	01D	012	1

NCC	NCCENR	LIBELLE	CAN	COMPARENT
ABERGEMENT CLEMENCIAT	Abergement-Clémenciat	L'Abergement-Clémenciat	0108	
ABERGEMENT DE VAREY	Abergement-de-Varey	L'Abergement-de-Varey	0101	
AMBERIEU EN BUGEY	Ambérieu-en-Bugey	Ambérieu-en-Bugey	0101	
AMBERIEUX EN DOMBES	Ambérieux-en-Dombes	Ambérieux-en-Dombes	0122	

TYPECOM	Type de commune
COM	Code commune
REG	Code région
DEP	Code département
CTCD	Code de la collectivité territoriale ayant les compétences départementales
ARR	Code arrondissement
TNCC	Type de nom en clair
NCC	Nom en clair (majuscules)
NCCENR	Nom en clair (typographie riche)
LIBELLE	Nom en clair (typographie riche) avec article
CAN	Code canton. Pour les communes « multi-cantonaux » code déclinée 99 à 90 (pseudo-canton) ou de 89 à 80 (communes nouvelles)
COMPARENT	Code de la commune parente pour les arrondissements municipaux et les communes associées ou déléguées.

- ▶ l'essai **d'importation** du fichier csv `commune2021` dans R

```
1 commune2021 <- read.table(  
2     file = "commune2021.csv"  
3     , sep = ","  
4     , header = T  
5     , stringsAsFactors = F ## !!!IMPORTANT!!!  
6 )
```

lance une erreur : **la ligne 177 n'avait pas 12 éléments**

- ▶ pourtant, la ligne 177 **n'a rien de particulier**

```
COMD,01182,,,,,0,GROSLEE,Groslée,Groslée,,01338
```

- ▶ et —sous 'NIX—

```
1 awk -F',' '{print NF}' commune2021.csv | sort | uniq -c
```

retourne 37743 12 —toutes les lignes ont **le même nombre de champs**—

- ▶ la cause de l'erreur survient dans les faits **dès la première ligne** :
L'Abergement-Clémenciat
- ▶ et provient d'un paramètre par défaut de la fonction `read.table` : `quote = "\"\""`
qui fait de l'apostrophe **un délimiteur de chaîne de caractère**
- ▶ désactiver l'option avec `quote = ""`

```
1 commune2021 ← read.table(  
2   file = "commune2021.csv"  
3   , sep = ","  
4   , quote = ''  
5   , header = T  
6   , stringsAsFactors = F  
7 )
```

permet de supprimer l'erreur

- ▶ il s'agit **d'un problème courant** sur les fichiers csv contenant notamment du texte **en français** mais non échappé par, p. ex., "
- ▶ les messages d'erreur de R, c'est pas toujours ça...

- ▶ le COG n'est **pas utilisé** que par l'INSEE
- ▶ et d'un fichier à l'autre ou d'un producteur à l'autre
- ▶ on retrouve **les même champs**
- ▶ mais **pas toujours désignés de la même façon**

Exemple : si on utilise des données IGNF et INSEE, les fichiers IGNF utilisent souvent INSEE_COM et non COM

- ▶ donc, si vous travaillez à partir **de plusieurs sources**

ce qui est souvent le cas : même des cartes d'aspect simple peuvent nécessiter plusieurs couches

- ▶ ça peut valoir la peine de **normaliser** les champs communs

ou, au moins, ceux qui vont servir à **l'appariement** entre les deux

- ▶ peu importe la façon, tant c'est fait **de façon cohérente**

Le format du code commune

- ▶ quand on travaille avec le COG, il faut aussi toujours faire attention au format **des codes**

- ▶ **code commune :**

- ▶ concaténation du code du département et du n° de la commune dans le département
- ▶ les deux champs se trouvant parfois dans **deux colonnes distinctes**

- ▶ codés sur **cinq caractères** mais de façon différente pour la Métropole et l'Outre-mer :

- ▶ le département (ou collectivités pour certains territoires d'outre-mer :

- ▶ Métropole : **deux caractères**
- ▶ Outre-mer : **trois caractères**

préfixe : 97 ou 98

- ▶ la commune :

- ▶ Métropole : **trois caractères**

max. de 999 communes par départements

- ▶ Outre-mer : **deux caractères**

max. de 99 communes par départements

► **Exemple :**

- St-Denis dans la Seine-St-denis —93— : **93**066
- St-Denis dans l'île de La Réunion —974— : **974**11

- de plus, **complément en tête** de code en métropole pour les départements dont le numéro est **inférieur à dix**

Exemple : L'Abergement-Clémenciat : 01001

- généralement par **un zéro** mais aussi parfois par un blanc

mais aussi parfois rien. . .

- lors d'appariements utilisant le code commune —ou autre—, **vérifiez toujours le formatage** ou formatez le code de façon appropriée

- ▶ les identifiants du COG sont représentés par **des chaînes de caractères** et non par **des nombres**
- ▶ car les identifiants du COG sont **des codes** et non **des n°**
- ▶ le code département et donc tous les autres qui sont en sont dérivés
- ▶ sont en effet **alphanumériques** depuis 1976
- ▶ et **la division** du département de la Corse —20— en deux —**2A et 2B**—
suite à la loi du 15 mai 1975 portant réorganisation de la Corse
- ▶ il faut donc toujours traiter les codes du COG comme **des chaînes de caractère**
- ▶ l'oublier peut parfois jouer des mauvais tours

- ▶ mais, parfois, traiter la variable **comme numérique**
- ▶ pour créer indicatrice DOM-COM

```
1 dcom ← as.integer(commune2021$COM) > 97000
```

marche quand même

- ▶ par contre, l'inverse est plus délicat. p. ex.,

```
1 metro ← !dcom # -ie: !(as.integer(cog.commune2021$COM) > 97000)
```

n'est **pas suffisant** car les communes de la Corse valent NA et pas FALSE avant le complément et

code	as.integer(code)	as.integer(code) > 970	!as.integer(code) > 970
01	1	F	T
2A	NA	NA	NA
974	974	T	F

parce que

$\neg \text{NA} \rightarrow \text{NA}$

- ▶ voir `scripts.R` pour une solution possible

La pérennité des identifiants

- ▶ l'administration française n'a pas de notion **d'identifiants pérennes**
- ▶ **Exemple** : les départements

- ▶ depuis la création des départements sous la Constituante en 1790, **leur nombre n'a cessé de varier**
- ▶ d'abord **83**
- ▶ et **jusqu'à 130** sous l'Empire

86 à la Restauration

- ▶ XIX^e–1945 :

- ▶ 1848 : création de trois départements en Algérie
- ▶ 1860 : annexion définitive de la Savoie et de Nice
- ▶ allers-retours de l'Alsace et de la Moselle

- ▶ après 1945 :

- ▶ Outre-mer : St-Pierre-et-Miquelon est aujourd'hui une collectivité d'outre-mer mais a d'abord été un département, à l'inverse, Mayotte est un département depuis 2011
- ▶ suppression des départements de **la Seine** et de **la Seine-et-Oise**

- ▶ la loi **du 10 juillet 1964 portant réorganisation de la région parisienne** prévoyait **le remplacement** de

- ▶ la Seine (75) par les départements de Paris (75) Hauts-de-Seine (92), de la Seine-Saint-Denis (93), et du Val-de-Marne (94)
- ▶ et de la Seine et Oise (78) par les départements de l'Essonne (91), des Yvelines (78), et du Val-d'Oise (95)

- ▶ Paris et les Yvelines ont donc **récupéré** les n° de la Seine et de la Seine-et-Oise respectivement
- ▶ les autres n° reprennent les n° utilisés **en Algérie** jusqu'en 1955 :

- ▶ 91 : département d'Alger
- ▶ 92 : département d'Oran
- ▶ 93 : département de Constantine
- ▶ 94 : Territoires du Sud

Note : le n° 99 du département de Bône, crée en 1955, a ensuite été utilisé pour indiquer la naissance à l'étranger minterligne

- ▶ de plus, le territoire de l'Algérie a été **redécoupé** entre 1956 et 1959 avec l'attribution de n° alphanumériques

9A, 9B, . . .

- ▶ **cas plus récent** : les collectivités d'outre-mer
- ▶ depuis les années 1980, les codes de certaines collectivités **ont été modifiés** et ce, **parfois plusieurs fois**
 - ▶ Mayotte : 98402 de 1943 à 1985 puis 985 de 1985 à 2003 et 976 depuis
 - ▶ ou encore pour Saint-Barthélemy ou Saint-Martin
- ▶ ce qui peut poser problème, p. ex., avec données électorales depuis le début de la V^e République
- ▶ qui utilisent le code du département pour identifier les circonscriptions
- ▶ le recyclage de codes peut donc poser problèmes avec des données **à long terme** mais aussi **à moyen terme**

- ▶ **cas particulier** : les trois communes à arrondissements
- ▶ pour lesquels on a un code **pour la commune** et des codes **pour les arrondissements**
 - ▶ **Paris** : 75056 pour la commune et 75101 pour le 1^{er} arrondissement, 75102 pour le deuxième, . . . jusqu'à 75120
 - ▶ **Lyon** : 69123 pour la commune et 69381, . . . , 69389 pour les neuf arrondissements
 - ▶ **Marseille** : 13055 pour la commune et 13201, . . . , 13216 pour les seize arrondissements
- ▶ donc, pour les données sur la France entière, toujours faire attention si les données **sont agrégées à la commune ou à l'arrondissement**

Exemple : l'INSEE diffuse parfois un fichier principal agrégé à la commune population et un fichier à part pour les arrondissements

- ▶ le code commune COG n'a donc **rien à voir** avec le code postal
- ▶ codes communes et codes postaux sont distincts car attribués par **deux administrations différentes** pour des **finalités différentes**
- ▶ le COG —et de nombreuses bases— ne contiennent pas de code postal
et attention lorsqu'ils en contiennent —voir plus loin
- ▶ pour géoréférencer des adresses, il faut utiliser une mappe
- ▶ comme, p. ex., celle proposée par La poste
- ▶ mais cette mappe n'est **pas injective**
 - ▶ un même code postal peut correspondre à deux communes—ou plus—
 - ▶ et donc à plusieurs codes communes

- ▶ là où ça se complique, c'est qu'une commune peut aussi avoir **plusieurs codes communes**
qui ne correspondent évidemment pas aux codes postaux
- ▶ deux autres cas de **communes particulières** :
 - ▶ les communes **associées**
 - ▶ et **déléguées**
- ▶ pour expliquer, petit détour et retour en arrière
- ▶ et un peu d'histoire **des politiques publiques**

- ▶ la **Nomenclature des unités territoriales statistiques** —NUTS— est une nomenclature d'EUROSTAT
- ▶ qui vise à rapprocher les divisions administratives des États membres
- ▶ **trois niveaux** actuellement
- ▶ **cinq précédemment**, le niveau 5 correspondant aux communes
- ▶ la France, avec ses 35 000+ communes représentant alors $\frac{1}{3}$ des NUTS 5

- ▶ la part prise par la France s'explique notamment par **des réformes** réalisées par d'autres pays
- ▶ qui ont drastiquement **réduit le nombre** de \sim « communes »
- ▶ soit **progressivement** comme en Suède, au Royaume-Uni, ou en Allemagne
- ▶ ou **d'un coup** comme en Belgique où une loi de 1975 a fait passer le nombre de municipalité de 2 359 à 589

Note : mais les anciens nom peuvent rester en usage et, plus généralement, personne ne connaît le COG. . .

- ▶ en France, **politiques similaires** ont aussi été menées et ce, depuis longtemps
- ▶ mais avec des résultats **un peu moins spectaculaires**

- ▶ les politiques de réduction du nombre de communes ont été mises en œuvre quasiment **dès leur création**
 - ▶ la loi du 14 décembre 1789 conduit à la création > 40 000 communes
 - ▶ les communes de moins de 250 habitants étant invitées à se réunir dès la loi du 20 août 1790
- ▶ les différents régimes qui se succèdent continuèrent l'effort
- ▶ avec un succès **plus que mitigé**
- ▶ après quelques essais infructueux, la III^e République finit, elle, par **entériner la situation**
- ▶ et crée **les syndicats intercommunaux** en 1890
- ▶ **d'autres établissements** de coopération intercommunale seront plus tard créés
- ▶ qui déboucheront sur la création **des établissements publics de coopération intercommunale** (EPCI)

métropoles, communautés urbaines, communautés d'agglomération et communautés de communes

100 fois remettez votre ouvrage

- ▶ quelques velléités commencent à se faire sentir sous la **IV^e République**
- ▶ mais les choses reprennent seulement sous la **V^e**
- ▶ avec loi n° 71-588 du 16 juillet 1971 sur les fusions et regroupements de communes dite « **loi Marcellin** »
 - du nom du ministre de l'intérieur de l'époque R. Marcellin
- ▶ qui prévoit **deux modes de rapprochement** entre les communes :
 - ▶ soit une fusion simple
 - ▶ soit une fusion comportant la création d'une ou plusieurs communes associées
- ▶ avec, là aussi, **un résultat mitigé**
 - quelques centaines de fusions au bout de dix ans et aussi quelques dé-fusions
- ▶ conduisant par la n° 2010-1563 du 16 décembre 2010 de réforme des collectivités territoriales
- ▶ à créer un nouveau régime de regroupement des communes, les « **communes nouvelles** »
- ▶ mais qui maintient **les communes associées** pour les communes déjà fusionnées
 - sous la forme de communes déléguées
- ▶ pour les communes ne souhaitant pas devenir une commune nouvelles

- ▶ la première conséquence de tout ça est que le nombre de communes **change presque tous les ans**

le nombre de communes pouvant aussi varier du fait qu'il y a plusieurs façons de définir les communes et donc autant de façons de les compter. . .



- ▶ donc attention si vous voulez faire **des comparaisons dans le temps**

ce qui illustre plus généralement la difficulté de constituer des séries longues à partir des catégories de l'entendement administratif

- ▶ retour au point de départ :

- ▶ tout cela a aussi des conséquences sur les opérations de **géoréférencement**
- ▶ **cas pratique** : Lille

- ▶ Lille a pour code postal **59000** et pour code commune **59350**
- ▶ Lille a **deux communes** associées/délégées

- ▶ Hellemmes-Lille depuis 1977 —code commune : 59298—

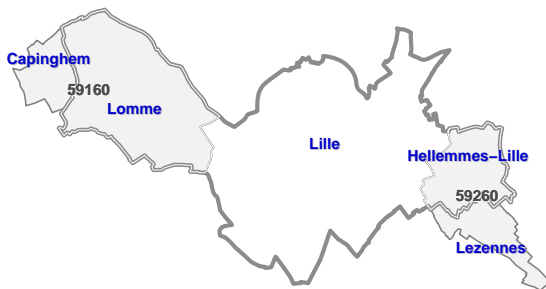
que personne n'appelle « Hellemmes-Lille »

- ▶ Lomme depuis 2000 —code commune : 59355—

- ▶ d'autre part,

- ▶ Hellemmes a **le même code postal** que Lezennes —59260—
- ▶ et Lomme a **le même code postal** que Capinghem —59160—

Carte des communes associées de Lille et de leur codes postaux



Petits détails complémentaires :

- ▶ le nombre non négligeable de noms **de voies communes à plusieurs communes** associées ou déléguées
- ▶ ou toutes :
 - en effet, les communes du Nord se caractérisent souvent par :
 - ▶ une avenue Jean Jaurès
 - ▶ une école primaire Roger Salengro
 - ▶ un beffroi
- ▶ de plus, les communes associées-déléguées ne sont souvent **pas distinguées** dans les bases
 - Lomme et Hellemmes ont généralement 59350 pour code commune, c-à-d celui de Lille

- ▶ selon l'observation que les « avanies » ne **s'additionnent** pas mais **se multiplient**
- ▶ en combinant tout cela, on obtient un mélange...**détonant**
- ▶ p. ex. lors **de géoréférencement d'adresses**
- ▶ si on n'y prend pas garde, résultats **défiant la géographie** garantis
- ▶ et pas toujours facilement détectables, surtout si on a beaucoup d'adresses
- ▶ **toujours vérifier le résultat** :
 - ▶ idéalement, **en automatisant** sur l'ensemble par des tests unitaires
pas seulement manuellement sur une petite partie
 - ▶ vous vous direz merci quand vous modifierez votre script des mois après

- ▶ l'INSEE diffuse l'information à **différents niveaux d'agrégation**
région, départements, communes, . . .
- ▶ mais aussi à un niveau **infra-communal** : IRIS
Îlots regroupés pour l'information statistique
- ▶ **découpage du territoire** de certaines communes selon différents critères
 - ▶ > 15 000 IRIS qui concernent principalement les communes de > 5 000 habitants
la liste est disponible à cette [page](#)
 - ▶ seulement **les départements**
sauf Mayotte entre 2011 et 2020 pour des résultats pas avant fin 2025
- ▶ diffusion des données principalement issues du **recensement de population**
- ▶ comme la population : <https://www.insee.fr/fr/statistiques/4228434>
- ▶ plus de données sont disponibles à **l'ADISP**

Le recensement et sa « rénovation »

- ▶ le **Recensement général de population** RGP a connu en France
- ▶ une profonde **transformation** —« rénovation »—
- ▶ au début des années 2000
- ▶ en effet, le recensement est aujourd'hui. . .
- ▶ . . .une enquête **par sondage probabiliste**

La dérive des recensements

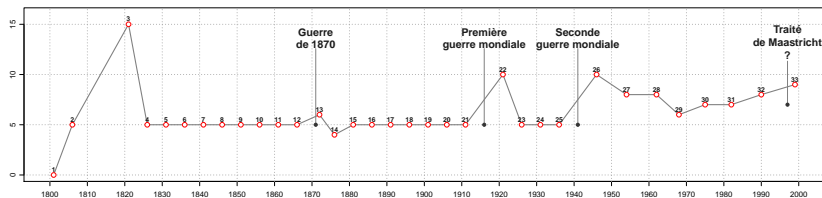


FIGURE 1 – Intervalles inter-censitaires 1801–1999

- ▶ depuis les 1801, les recensements se sont déroulés en France **tous les cinq ans**
à quelques exceptions près
- ▶ à partir de la Seconde guerre mondiale, l'intervalle inter-censitaire a toutefois commencé **à dériver**
- ▶ pour atteindre **neuf ans** pour le 33^e recensement (1999), soit deux ans après l'année prévue initialement
le précédent datant de 1990

Le recensement et sa « rénovation »

- ▶ cette dérive s'explique en partie par **le coût et la complexité** croissant de l'opération
- ▶ dans les années 1990, l'INSEE a complètement revu **le dispositif**
- ▶ le chantier de la « rénovation » du RGP a d'abord abouti à l'inscription du « recensement » **dans la loi**

titre V —« Des opérations de recensement »— de la loi n° 2002-276 du 27 février 2002 relative à la démocratie de proximité

- ▶ jusque-là, les recensements étaient organisés par décret
- ▶ la mise en place effective **à partir de 2004**
- ▶ selon des modalités **drastiquement différentes**

Le recensement et sa « rénovation »

- ▶ jusque-là, le recensement était réalisé en France métropolitaine à **intervalles (ir)réguliers** sur une période de plusieurs mois
- ▶ où les communes étaient divisées **en îlots**
- ▶ et où **~tous les ménages** de chaque îlot étaient visités par des agents recenseurs
- ▶ le recensement est désormais réalisé **« en continu »**
- ▶ selon une méthode conçue par J.-C. Deville et M. Jacod
 - inspirée par le statisticien L. Kish
- ▶ souvent présenté comme **une spécificité française**
 - mais le *Census Bureau* réalise aux EU l'enquête ACS —*American Community Survey*— suit la même inspiration pour compléter le recensement qui a lieu tous les dix ans
- ▶ dans un contexte international **de réflexion sur les recensements**
 - qui a parfois conduit à leur suppression pure et simple

L'enquête de recensement

- ▶ la collecte se déroule sur **cinq ans** par rotation

2004–2008, 2009–2013, 2014–2018, . . .

- ▶ en distinguant les communes selon **un critère de taille**

- ▶ communes de moins de 10 000 habitants : la commune est enquêtée intégralement **une fois** au cours de la période
- ▶ communes de plus de 10 000 habitants : **une fraction des adresses** de la commune est recensée chaque année

10 000 habitants \simeq médiane de la population française

- ▶ **tirage aléatoire** dans les deux cas

- ▶ communes de moins de 10 000 habitants : tirage de communes qui sont réparties en cinq groupes
- ▶ communes de plus de 10 000 habitants : tirages de 8% des **d'adresses**

base de sondage le Répertoire d'immeubles localisés —RIL—

soit 40% des adresses au bout de cinq ans

► tirage aléatoire :

- communes de moins de 10 000 habitants : tirage équilibré sur un ensemble de variables

le nombre de logements, tranches d'âges, sexe, . . .

- communes de plus de 10 000 habitants : stratification

- immeubles de grande taille : tous enquêtés au bout de cinq ans

adresses d'au moins 60 logements chacune et cumulant au maximum 10% des logements de la commune

- autres immeubles : tirage équilibré

- les échantillons successifs sont appelés **groupes de rotation**

- les tirages ont lieu tous les ans avec

- des règles pour ajouter les nouvelles adresses
 - des règles lorsqu'une commune dépasse le seuil des 10 000 hab. en cours de cycle

- ▶ de par les modalités de la collecte, les statistiques d'intérêt comme la population légale doivent donc être **estimées**

estimations pour chaque année

- ▶ avec des estimateurs qui varient en fonction de **la date** dans le cycle de collecte, **du seuil** de $+/-10\ 000$ et **en Métropole ou dans les DOM**

Exemples

- ▶ pour **les communes $>10\ 000$ hab. en métropole** : la population des ménages pour une année n est estimée en multipliant le nombre de logements au 1^{er} janvier de l'année de référence par le nombre moyen de personnes par logement estimé lors des cinq dernières collectes par sondage (collectes $n-2$ à $n+2$)

pour 2019, on fait la moyenne de 2017 à 2021

- ▶ pour **les communes $<10\ 000$ hab.** : on retient le résultat de l'enquête de recensement n si elle a eut lieu cette année-là. Pour les autres années, les estimateurs varient si on veut calculer pour les années $n+1$ ou $n+2$ d'une part et, d'autre part, les années $n-1$ ou $n-2$.
Ces estimations mobilisent, elles, le fichier de la taxe d'habitation.

- ▶ ne soyez donc pas surpris si les edonnées issues du recensement comportent **des nombres flottants** et non **des entiers**

- ▶ du fait que les calculs pour une année peuvent nécessiter
- ▶ les années **suivant** ou —non—exclusif— les années **précédant** l'année considérée
- ▶ plusieurs estimateurs ne peuvent être calculés de façon instantanée mais pour **l'année médiane d'un cycle**
 - ▶ ainsi, une partie des populations légales publiées annuellement par l'INSEE
 - ▶ sont en fait les population estimées en $n-3$
 - ▶ mais projetée dans les limites administratives du millésime
 - c-à-d en intégrant les regroupements de communes intervenus l'année précédente
- ▶ attention donc **au millésime des fichiers**

C'était mieux avant ?

- ▶ les nouvelles modalités du recensement ont suscité **certaines réticences**
- ▶ d'origines diverses

élus, géographes et même au sein de l'INSEE

- ▶ et leur mise en œuvre, un —peu de— contentieux via des recours au Conseil d'État
- ▶ sans rentrer dans les détails de la polémique, on peut toutefois **noter** que :

- ▶ l'ancien recensement n'était **pas exhaustif**

~1% d'erreur

- ▶ surtout, le recensement fait l'objet de **deux types d'exploitations**

- ▶ **une exploitation principale** : toutes les réponses de tous les bulletins sont saisies

évolution et structure de la population, logements et migrations résidentielles, . . .

- ▶ **une exploitation complémentaire** : seule une fraction des réponses sont saisies
—25% en 1999—

la structure familiale des ménages, la profession et la catégorie socioprofessionnelle des personnes, . . .

- ▶ et c'est encore le cas aujourd'hui

- ▶ à la suite de A. Desrosières, il est courant **de distinguer**
 - ▶ statistique administrative
 - ▶ statistique d'enquête
- ▶ mais, dans les faits, la finalité des collectes de la statistique publique est **toujours « administrative »** quel que soit le mode de collecte
- ▶ car ces collectes sont issues de **politiques publiques** et sont un des moyens de **l'action publique**
 - ▶ la première finalité du RGP, c'est de **déterminer la population légale** et donc la dotation d'État des communes
 - ▶ plus généralement, le RGP est nécessaire pour l'application de **~350 articles législatifs** répartis dans 28 codes différents
 - nombre de conseillers municipaux, détermination du mode de scrutin, découpage des circonscriptions électorales, implantations d'officines de pharmacie et des débits de tabac, les barèmes de certaines taxes, . . .
- ▶ et cela vaut aussi pour **les autres collectes** de l'INSEE même si elles ne sont pas directement la mise en œuvre de dispositions législatives
- ▶ d'où l'importance de **se familiariser** avec les normes, les nomenclatures, les modalités, . . . à l'œuvre dans la production de ces données pour les utiliser **à bon escient**

Merci pour votre attention