Communes nouvelles : préparation des données

Auteur anonymisé

06 septembre 2021

# Mots-clés

La revue a créé un index des mots clés en français et en anglais : <https://journals-openedition-org.ezproxy.u-paris.fr/cybergeo/33>

# Objectifs généraux

L’objet de cette documentation consiste à décrire les données et géométries utilisées dans les analyses ainsi qu’à préparer les fichiers en amont pour faciliter leur usage ultérieur.

Ce projet utilise les **données INSEE et le fond géographique de référence aux géographies du 1er janvier 2011** (avant toute création de commune nouvelle).

Outre la documentation des sources utilisées, les objectifs de ce document consistent à créer :

* Les géométries de référence au 1er janvier 2011 et au 1er janvier de l’année la plus récente pour laquelle des données ont été publiées par l’INSEE (2021) avec le même niveau de généralisation cartographique. Ces objets seront appelés geom2011 et geom\_new.
* Les géométries des communes fusionnantes devenant des communes nouvelles : geomfus2011 avant les fusions (au 1er janvier 2011), et geomCN\_new après les fusions.
* Des données statistiques contextuelles pour l’ensemble des communes en 2011 et pour l’année la plus récente (fusion des données 2011). Ces données sont d’ordre socio-économique (emploi, structure par âge, etc.), relatives à des zonages (EPCI d’appartenance, catégories d’aires urbaines), et plus important, permettent de décrire la nature des fusions communales à l’oeuvre.

La partie 6 présente l’ensemble des données disponibles aux géographies du 1er janvier 2011 (df2011) et celles du 1er janvier de la dernière année disponible (df\_new).

Une simple jointure entre ces données et les géométries de référence est dès lors nécessaire pour procéder aux analyses.

# 1 - Géométries de référence

Le fond géographique [Geofla®Communes édition 2011 France Métropolitaine](https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html#geofla) est importé, c’est l’objet geom2011. Les arrondissements marseillais, lyonnais et parisiens sont agrégés au niveau communal pour un meilleur appariement avec les données statistiques de référence qui seront utilisées ultérieurement. Les communes de Corse sont aussi supprimées, étant donné qu’elles ne sont pas concernées par le processus de création de communes nouvelles. Cela améliorera aussi la résolution des représentations cartographiques à venir.

Ce fond communal est ensuite agrégé au niveau des départements, c’est l’objet dep.

On exporte finalement le code communal de ces géométries, qui servira de référence pour les jointures avec les données attributaires à venir.

# Géométries des communes françaises au 1er janvier 2011 (source : IGN)  
geom2011 <- st\_read(dsn = "data-raw/geom/COMMUNE.shp")

## Reading layer `COMMUNE' from data source   
## `E:\Projet\_CN\_2020\_Perso\data-raw\geom\COMMUNE.SHP' using driver `ESRI Shapefile'  
## Simple feature collection with 36610 features and 18 fields  
## Geometry type: MULTIPOLYGON  
## Dimension: XY  
## Bounding box: xmin: 99226 ymin: 6049647 xmax: 1242375 ymax: 7110524  
## Projected CRS: RGF93 / Lambert-93

geom2011 <- geom2011[geom2011$CODE\_REG != "94",] # Suppression de la Corse  
geom2011 <- st\_transform(geom2011, crs = 3035) # Transformation Lambert-93  
  
  
# Suppression et regroupement des arrondissements de Paris, Lyon et Marseille  
arrcom <- read\_excel("data-raw/geom/com\_arm.xlsx") # Table de passage des arrondissements  
com <- merge(geom2011, arrcom, by = "INSEE\_COM", all.y = TRUE) # Géométries des arrondissements  
com <- aggregate(x = com[,"CODGEO"], by = list(com$CODGEO), # On les aggrège  
 FUN = head, 1)  
  
com$NOM\_COMM <- c("MARSEILLE", "LYON", "PARIS")  
com$CODE\_DEPT <- c("13","69","75")  
  
geom2011 <- geom2011[!geom2011$INSEE\_COM %in% arrcom$INSEE\_COM, ] # Suppression des arrondissements  
  
colnames(geom2011)[3] <- "CODGEO"  
geom2011 <- rbind(geom2011[,c("CODGEO","NOM\_COMM", "CODE\_DEPT")],  
 com[,c("CODGEO","NOM\_COMM", "CODE\_DEPT")])  
  
  
# Agrégation en départements  
dep <- aggregate(x = geom2011[,"CODE\_DEPT"], by = list(geom2011$CODE\_DEPT), # Agrégation com > dépt  
 FUN = head, 1)  
  
# Codgeo de référence  
df <- st\_set\_geometry(geom2011[,c("CODGEO", "CODE\_DEPT")], NULL)  
  
# Forcer character  
df$CODGEO <- as.character(df$CODGEO)  
df$CODE\_DEPT <- as.character(df$CODE\_DEPT)

# 2 - Données contextuelles statistiques

## 2.1 - Fichiers sources et indicateurs cibles

Afin de caractériser les spécificités des communes nouvelles, plusieurs tables statistiques à l’échelon communal sont importées. Ces tables statistiques sont pour la plupart définies en géographie du 1er janvier 2011.

Voici une table qui résume les indicateurs ciblés pour cette collecte de données - constituant généralement des numérateurs et dénominateurs qui pourront, par la suite, être composés en ratios.

Ces données sont accessibles sur le site Web de l’INSEE à partir du nom du jeu de données Excel présent dans la table ci-dessous.

|  |  |  |  |
| --- | --- | --- | --- |
| INSEE\_CODE | DESCRIPTION | SOURCE | PRODUCTEUR |
| CODGEO | Code INSEE, découpage géographique au 01/01/2011 | Toutes tables | INSEE |
| LIBGEO | Nom de la commune, découpage géographique au 01/01/2011 | Toutes tables | INSEE |
| CODE\_DEPT | Code départemental | Toutes tables | INSEE |
| CATAEU2010 | Catégorie de communes dans le zonage en aires urbaines 2010 | table-appartenance-geo-communes-11.xls | INSEE |
| REG | Région | table-appartenance-geo-communes-11.xls | INSEE |
| AU2010 | Aire urbaine 2010 | table-appartenance-geo-communes-11.xls | INSEE |
| EPCI | EPCI | table-appartenance-geo-communes-11.xls | INSEE |
| P09\_ACT1564 | Actifs 15-64 ans en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| P09\_CHOM1564 | Chômeurs 15-64 ans en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| P09\_ETUD1564 | Elèv. Etud. Stag. non rémunérés 15-64 ans en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| P09\_RETR1564 | Retraités Préretraités 15-64 ans en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| C09\_ACT1564\_CS1 | Actifs 15-64 ans Agriculteurs exploitants en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| C09\_ACT1564\_CS2 | Actifs 15-64 ans Artisans, Comm., Chefs entr. en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| C09\_ACT1564\_CS3 | Actifs 15-64 ans Cadres, Prof. intel. sup. en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| C09\_ACT1564\_CS4 | Actifs 15-64 ans Prof. intermédiaires en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| C09\_ACT1564\_CS5 | Actifs 15-64 ans Employés en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| C09\_ACT1564\_CS6 | Actifs 15-64 ans Ouvriers en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| P09\_EMPLT | Emplois au LT en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| C09\_EMPLT\_AGRI | Emplois au LT Agriculture en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| C09\_EMPLT\_INDUS | Emplois au LT Industrie en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| C09\_EMPLT\_CONST | Emplois au LT Construction en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| C09\_EMPLT\_CTS | Emplois au LT Commerce, Transports, Services divers en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| C09\_EMPLT\_APESAS | Emplois au LT Adm publique, Enseignement, Santé, Act sociale en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| P09\_ACTOCC | Actifs occupés 15-64 ans en 2009 | base-cc-emploi-pop-active-2009.xls | INSEE |
| P09\_POP | Population en 2009 | base-cc-evol-struct-pop-2009.xls | INSEE |
| P09\_POP0014 | Pop 0-14 ans en 2009 | base-cc-evol-struct-pop-2009.xls | INSEE |
| P09\_POP1529 | Pop 15-29 ans en 2009 | base-cc-evol-struct-pop-2009.xls | INSEE |
| P09\_POP3044 | Pop 30-44 ans en 2009 | base-cc-evol-struct-pop-2009.xls | INSEE |
| P09\_POP4559 | Pop 45-59 ans en 2009 | base-cc-evol-struct-pop-2009.xls | INSEE |
| P09\_POP6074 | Pop 60-74 ans en 2009 | base-cc-evol-struct-pop-2009.xls | INSEE |
| P09\_POP75P | Pop 75 ans ou plus en 2009 | base-cc-evol-struct-pop-2009.xls | INSEE |
| C09\_ACTOCC15P\_ILT1 | Nombre d’actifs travaillant dans leur commune (ou arrondissement municipal) de résidence | BTX\_FM\_DTR\_2009.xls | INSEE |
| C09\_ACTOCC15P\_ILTAUT | Nombre d’actifs travaillant hors de leur commune (ou arrondissement municipal) de résidence | BTX\_FM\_DTR\_2009.xls | INSEE |
| C09\_ACTOCC15P | Total : nombre d’actifs de 15 ans ou plus ayant un emploi | BTX\_FM\_DTR\_2009.xls | INSEE |
| Potentiel.Financier | Potentiel financier des communes | Potentiel\_financier\_2011.xls | DGCL |
| dotation.globale.de.fonctionnement | Dotation globale de fonctionnement des communes en 2011 | city\_all.csv | DGCL |
| Nombre.de.foyers.fiscaux | Nombre de foyers fiscaux dans la commune en 2011 | Ircom\_2011\_revenus2010.xls | Ministère de l’économie |
| Revenu.fiscal.de.référence.des.foyers.fiscaux | Revenu fiscal de référence de l’ensemble des foyers imposés sur la commune | Ircom\_2011\_revenus2010.xls | Ministère de l’économie |
| Impôt.net..total. | Impôt net total sur la commune | Ircom\_2011\_revenus2010.xls | Ministère de l’économie |
| Nombre.de.foyers.fiscaux.imposables | Nombre de foyers fiscaux imposables sur la commune | Ircom\_2011\_revenus2010.xls | Ministère de l’économie |

## 2.2 - Préparation des données statistiques

Le code ci-dessous importe les indicateurs contenus dans ces tables de référence, réalise une jointure avec les codes communaux 2011 et filtre les indicateurs pour ne conserver que les indicateurs d’intérêt. Le tout est inclu dans un objet df2011 qui constituera l’objet incluant les données contextuelles d’intérêt pour cette étude, aux géographies du 1er janvier 2011.

# Importer les tableaux de données d'intérêt  
datatmp1 <- data.frame(read\_excel("data-raw/stats\_insee/base-cc-emploi-pop-active-2009.xls", sheet = "COM\_2009", skip = 5))  
datatmp2 <- data.frame(read\_excel("data-raw/stats\_insee/base-cc-evol-struct-pop-2009.xls", sheet = "COM\_2009", skip = 5))  
datatmp3 <- data.frame(read\_excel("data-raw/stats\_insee/BTX\_FM\_DTR\_2009.xls", sheet = "TOTAL", skip = 5))  
datatmp4 <- data.frame(read\_excel("data-raw/stats\_insee/Potentiel\_financier\_2011.xls", sheet = "Rapport1", skip = 2))  
datatmp5 <- data.frame(read\_excel("data-raw/stats\_insee/table-appartenance-geo-communes-11.xls", sheet = "Liste\_COM", skip = 5))  
datatmp6 <- read.csv("data-raw/stats\_insee/city\_all.csv", sep = ",", stringsAsFactors = FALSE, fileEncoding = "UTF8")  
datatmp6 <- subset(datatmp6, datatmp6$année == "2011")  
  
# Nécessité d'une procédure un peu plus longue pour les données concernant les revenus  
sheet\_names <- excel\_sheets("data-raw/stats\_insee/Ircom\_2011\_revenus2010.xls") # On liste les feuilles  
datatmp7 <- data.frame()  
for (i in 1:length(sheet\_names)){  
 feuille <- sheet\_names[i] # récupère le nom de la feuille  
 IRFeuille <- data.frame(read\_excel("data-raw/stats\_insee/Ircom\_2011\_revenus2010.xls", sheet = feuille, skip = 20))  
 datatmp7 <- rbind(datatmp7, IRFeuille, stringsAsFactors = FALSE)  
 rm(feuille, IRFeuille) # suppression des objet intermédiaires  
}  
# Renommer variables  
colnames(datatmp7)[10] <- "Traitements.et.salaires\_Nbr.foyers"  
colnames(datatmp7)[11] <- "Traitements.et.salaires\_Montant"  
colnames(datatmp7)[12] <- "Retraites.et.pensions\_Nbr.foyers"  
colnames(datatmp7)[13] <- "Retraites.et.pensions\_Montant"  
datatmp7 <- subset(datatmp7, datatmp7$Revenu.fiscal.de.référence.par.tranche..en.euros. == "Total") # Sélection des seules variables totales  
datatmp7[, c(5:length(datatmp7))]<- lapply(datatmp7[, c(5:length(datatmp7))], as.numeric) # Variables en numériques  
# Création d'un CODGEO  
datatmp7$CODGEO <- substr(datatmp7$DÉP, 1, 2)  
datatmp7$CODGEO <- paste(datatmp7$CODGEO, datatmp7$Commune, sep = "")  
  
# Jointure avec le champ CODGEO 2011 de référence (issu des géométries)  
df <- merge(df, datatmp1, by = "CODGEO", all.x = TRUE)  
df <- merge(df, datatmp2[,c(1, 5:length(datatmp2))], by = "CODGEO", all.x = TRUE)  
df <- merge(df, datatmp3[,c(1, 3:length(datatmp3))], by = "CODGEO", all.x = TRUE)  
df <- merge(df, datatmp4, by.x = "CODGEO", by.y = "Code.INSEE", all.x = TRUE)  
df <- merge(df, datatmp5[,c(1, 5:length(datatmp5))], by = "CODGEO", all.x = TRUE)  
df <- merge(df, datatmp6[,c(5, 12)], by.x = "CODGEO", by.y = "cog..code.officiel.géographique.", all.x = TRUE)  
df <- merge(df, datatmp7[,c(5:length(datatmp7))], by = "CODGEO", all.x = TRUE)  
  
# Ne conserver que les indicateurs d'intérêt   
selec <- target$INSEE\_CODE  
df2011 <- df[,selec]  
  
# Renommer le code des variables  
colnames(df2011) <- target$recoding

Les catégories du zonage en aires urbaines sont aussi regroupées.

# Pour créer des catégories plus vastes à partir des   
# Utilisation du ZAU pour cibler les communes appartenant à un petit pôle (unité urbaine et couronne)  
  
# Utilisation du ZAU pour cibler les communes du rural profond  
df2011$ZAU\_POL [df2011$CATAEU2010 %in% c("221", "222")] <- "OUI"  
  
# Utilisation du ZAU pour cibler les communes du rural profond  
df2011$ZAU\_RUR [df2011$CATAEU2010 %in% c("300", "400")] <- "OUI"  
  
# Utilisation du ZAU pour cibler les communes du rural profond + couronnes petits pôles : Marges Sans Pôles  
df2011$ZAU\_MAR\_SP [df2011$CATAEU2010 %in% c("222", "300", "400")] <- "OUI"  
  
# Utilisation du ZAU pour cibler les communes du rural profond + petits pôles et couronnes  
df2011$ZAU\_MAR [df2011$CATAEU2010 %in% c("221", "222", "300", "400")] <- "OUI"  
  
# Utilisation du ZAU pour cibler les communes périurbaines et les autres  
df2011$ZAU\_PERI [df2011$CATAEU2010 %in% c("112", "120")] <- "OUI"  
  
# Utilisation du ZAU pour cibler les communes des Grandes aires urbaines, des aires urbaines et les autres  
df2011$ZAU\_AU [df2011$CATAEU2010 %in% c("111", "112", "120")] <- "GrdAirUrb"  
df2011$ZAU\_AU [df2011$CATAEU2010 %in% c("211", "212", "221", "222")] <- "AirUrb"  
df2011$ZAU\_AU [df2011$CATAEU2010 %in% c("300", "400")] <- "Autres"  
  
# Tout ce qui ne fait pas partie des catégories = NON   
selecV <- c("ZAU\_POL", "ZAU\_RUR", "ZAU\_MAR\_SP", "ZAU\_MAR", "ZAU\_PERI", "ZAU\_AU")  
df2011[selecV][is.na(df2011[selecV])] <- "NON"  
  
# # Transformation en facteur des catégories des aires urbaines  
# df2011$CATAEU2010 <- as.factor(df2011$CATAEU2010)  
# df2011$CODGEO <- as.character(df2011$CODGEO)

# 3 - Fusions communales et communes nouvelles

La [table de passage communale](https://www.insee.fr/fr/information/2028028) produite par l’INSEE est importée. Elle restitue l’ensemble des changements qui ont touché des communes en France depuis 2003. Cette table ne concerne pas exclusivement les communes nouvelles.

Pour cette raison, on associe à ce fichier d’origine d’autres tables mises à disposition par l’INSEE : le listing des communes nouvelles ( [un fichier par année depuis 2016](https://www.insee.fr/fr/information/2549968)). Pour récupérer les créations de communes nouvelles antérieures au 2 janvier 2015, un fichier est créé à partir de l’interface d’[historique des communes](https://www.insee.fr/fr/metadonnees/historique-commune?taille=100&debut=0&modification=MG&p-debut=2003&p-fin=2014) de l’INSEE. NB : Les tables de passages de l’INSEE sont du 2 janvier de l’année indiquée jusqu’au 1er janvier de l’année suivante, inclue (par exemple, la table de passage “2015”) comprend toutes les fusions du 2 janvier 2015 au 1er janvier 2016. Voici le récapitulatif des informations mises à disposition par l’INSEE :

|  |  |  |  |
| --- | --- | --- | --- |
| INSEE\_CODE | DESCRIPTION | SOURCE | PRODUCTEUR |
| annee\_modif | Année de modification | table\_passage\_geo2003\_geo2021.xlsx | INSEE |
| COM\_FIN | Code INSEE de la commune finale | table\_passage\_geo2003\_geo2021.xlsx | INSEE |
| COM\_INI | Code INSEE Commune initiale | table\_passage\_geo2003\_geo2021.xlsx | INSEE |
| LIB\_COM\_FIN | Libellé de la commune finale | table\_passage\_geo2003\_geo2021.xlsx | INSEE |
| LIB\_COM\_INI | Libellé de la commune initiale | table\_passage\_geo2003\_geo2021.xlsx | INSEE |
| DepComN | Code Insee de la commune nouvelle | communes\_nouvelles\_2012…2021.xls | INSEE (fichiers consolidés à partir de 2015) |
| NomCN | Nom de la commune nouvelle | communes\_nouvelles\_2012…2021.xls | INSEE (fichiers consolidés à partir de 2015) |
| DepComA | Code Insee de l’ancienne commune | communes\_nouvelles\_2012…2021.xls | INSEE (fichiers consolidés à partir de 2015) |
| NomCA | Nom de l’ancienne commune | communes\_nouvelles\_2012…2021.xls | INSEE (fichiers consolidés à partir de 2015) |
| ChefLieu | Statut de commune chef-lieu (O=OUI, N=NON) | communes\_nouvelles\_2012…2021.xls | INSEE (fichiers consolidés à partir de 2015) |
| ComDLG | Statut de Commune déléguée (O=OUI, N=NON) | communes\_nouvelles\_2012…2021.xls | INSEE (fichiers consolidés à partir de 2015) |
| Date1 | Date de l’arrêté préfectoral (JJ/MM/AAAA) | communes\_nouvelles\_2012…2021.xls | INSEE (fichiers consolidés à partir de 2015) |
| Date2 | Date d’effet (si différente du 1er janvier de l’année suivante) | communes\_nouvelles\_2012…2021.xls | INSEE (fichiers consolidés à partir de 2015) |
| Date3 | Date de publication au JO (renseignée si la parution au JO a eu lieu à la date de mise en ligne) | communes\_nouvelles\_2012…2021.xls | INSEE (fichiers consolidés à partir de 2015) |

On regroupe toutes ces informations au sein d’un seul dataframe. Si la date d’effet (Date2) est manquante au niveau des données fournies par l’INSEE, on applique la valeur au 1er janvier de l’année suivante à ce champ.

À noter que pour mettre à jour la table de passage avec d’éventuelles fusions ultérieures, il suffit de rajouter les fichier les plus récents. Travail pour l’intégration des données publiées chaque année : - Intégration nouvelle table fusions 2021 - intégration nouvelle table évolutions 2003-2021 - rechercher la dernière année connue pour la modifier dans le texte.

À noter également que les dates des fichiers Excel fournis par l’INSEE ne sont pas formatés de façon identique. Au format texte jusqu’à 2017, au format date à partir de 2018. Nous avons préalablement converti les champs date au format texte pour l’ensemble des fichiers. Cela signifie qu’en cas de mise à jour des données il faudra de nouveau convertir les dates au format texte, ou inversement, tout transformer au format date depuis le début dans les fichiers Excel.

# 3.1 - Identification des communes fusionnantes

On commence par identifier les communes fusionnantes à partir de la table de passage globale fournie par l’INSEE depuis la géométrie de 2003.

# comFus <- data.frame(read\_excel("data-raw/communes\_nouvelles/table\_passage\_geo2003\_geo2020.xlsx", sheet = "Liste des fusions", skip = 5))  
comFus <- data.frame(read\_excel("data-raw/communes\_nouvelles/table\_passage\_geo2003\_geo2021.xlsx", sheet = "Liste des fusions", skip = 5))  
  
# Pour éviter les doublons dans la table INSEE  
comFus <- aggregate(x = comFus, by = list(comFus$COM\_INI),   
 FUN = head, 1) # On garde le premier élément (car df trié par date)  
  
# Renommer avec la commune fusionnée la plus récente  
comFus <- comFus[order(comFus$annee\_modif, decreasing = TRUE),]  
comFus$LIB\_COM\_FIN <- ave(comFus$LIB\_COM\_FIN, comFus$COM\_FIN, FUN = function(x) x[1])  
comFus <- comFus[,c(2:length(comFus))]  
  
# Identification des communes fusionnantes dans la table d'entrée  
df2011 <- merge(df2011, comFus[,c("COM\_INI", "annee\_modif")], by.x = "CODGEO",  
 by.y = "COM\_INI", all.x = TRUE)  
df2011$annee\_modif[!is.na(df2011$annee\_modif)] <- "OUI"  
df2011$annee\_modif[is.na(df2011$annee\_modif)] <- "NON"  
colnames(df2011)[length(df2011)] <- "FUSION"

# 3.2 - Identification et caractérisation des communes nouvelles

Il s’agit ici de sélectionner, dans chaque table de passage fournie par l’INSEE, les communes fusionnantes et les communes nouvelles créées. Certains cas particuliers demandent un traitement spécifique : - Bettoncourt-le-Haut a fusionné en 1972 avec Épizon, sous le statut de commune-associée. Au 1er janvier 2013, avec la création de la commune-nouvelle d’Épizon, la commune de Bettoncourt-le-Haut devient commune déléguée. Comme on ne dispose pas des données 2011 pour Bettoncourt-le-Haut, on préfère ignorer cette commune en considérant que la commune nouvelle d’Épizon est le résultat de la fusion des seules Épizon et Pautaines-Augeville. Source : <https://www.insee.fr/fr/metadonnees/cog/commune/COM52046-bettoncourt-le-haut>. - On doit également traiter à la main le cas de la commune de L’Oudon, qui a changé de code communal : issue de la fusion-association de dix communes en 1973, avait jusqu’en 1990 le code de 14472. Du fait d’un transfert de chef-lieu, le code a été modifié en 14697. Après un nouveau transfert de chef-lieu en 2014, l’INSEE décide en 2016 de ré-attribuer le code de 14472. Cette commune a fusionné au 1er janvier 2017 au sein de la commune nouvelle de Saint-Pierre-en-Auge. Pour éviter les problèmes d’associations entre les données, nous avons laissé l’ancien code 14697, qui est celui utilisé dans les données INSEE que nous utilisons (millésime 2011).

# Import tables de référence INSEE  
tmp <- data.frame(read\_excel("data-raw/communes\_nouvelles/communes\_nouvelles\_2012\_2014.xls",   
 sheet = "Communes\_nouvelles\_2012\_2014"))  
# Traitement du cas particulier de Bettoncourt-le-Haut et Épizon  
tmp <- subset (tmp, DepComA != "52046")  
  
tmp1 <- data.frame(read\_excel("data-raw/communes\_nouvelles/communes\_nouvelles\_2015.xls",   
 sheet = "Communes nouvelles 2015"))  
tmp1$Date2[is.na(tmp1$Date2)] <- "01/01/2016"  
  
tmp2 <- data.frame(read\_excel("data-raw/communes\_nouvelles/communes\_nouvelles\_2016.xls",   
 sheet = "Communes nouvelles 2016"))  
tmp2$Date2[is.na(tmp2$Date2)] <- "01/01/2017"  
  
# Traitement du cas particulier de la commune de L'Oudon  
# tmp2$DepComA[tmp2$NomCA == "Oudon (L')"] <- "14697"  
tmp2$DepComA[tmp2$DepComA == "14472"] <- "14697"  
  
tmp3 <- data.frame(read\_excel("data-raw/communes\_nouvelles/communes\_nouvelles\_2017.xls",   
 sheet = "Communes nouvelles 2017"))  
tmp3$Date2[is.na(tmp3$Date2)] <- "01/01/2018"  
  
tmp4 <- data.frame(read\_excel("data-raw/communes\_nouvelles/communes\_nouvelles\_2018.xls",   
 sheet = "Communes\_nouvelles\_2018"))  
tmp4$Date2[is.na(tmp4$Date2)] <- "01/01/2019"  
  
tmp5 <- data.frame(read\_excel("data-raw/communes\_nouvelles/communes\_nouvelles\_2019.xls",   
 sheet = "Communes\_nouvelles\_2019"))  
  
tmp6 <- data.frame(read\_excel("data-raw/communes\_nouvelles/communes\_nouvelles\_2020.xlsx",   
 sheet = "Communes\_nouvelles\_2020"))  
tmp6$Date1 <- format(as.Date(tmp6$Date1, format = "%Y-%m-%d"),"%d/%m/%Y") # Pour mettre la date au même format que les fichiers précédents  
tmp6$Date2 <- format(as.Date(tmp6$Date2, format = "%Y-%m-%d"),"%d/%m/%Y") # Pour mettre la date au même format que les fichiers précédents  
  
# Regroupement des colonnes identiques des différents dataframe  
comNew <- data.frame()  
list\_df <- list(tmp, tmp1, tmp2, tmp3, tmp4, tmp5, tmp6)  
  
for(item in list\_df){   
 items <- item[, names(tmp)] # La référence sont les noms de colonne du premier fichier  
 comNew <- rbind(comNew, items)   
  
}  
  
# Supprimer les erreurs de saisie  
comNew <- subset(comNew, comNew$DepComN != "NA")  
  
# Reformatage année fusion  
comNew$FusAn <- as.Date(comNew$Date2, tryFormats = "%d/%m/%Y")  
  
# Découpage en phases temporelles de fusion  
comNew$FusPhas <- cut(comNew$FusAn,   
 breaks = as.Date(c("2011-1-1", "2015-1-2", "2016-1-2",  
 "2017-1-2","2018-1-2", "2020-1-2","2021-1-2")),   
 labels = c("Phase 1", "Phase 2", "Phase 3", "Phase 4", "Phase 5", "Phase 6"))  
  
# Pour fusions successives et à code DepComA identique, ne conserver que l'élément le plus ancien  
comNew <- aggregate(x = comNew, by = list(comNew$DepComA), # on évite les doublons  
 FUN = head, 1) # On ne garde que le premier élément (comNew trié par ordre chronologique)  
  
# Prise en compte du dernier nom de commune nouvelle  
comNew <- comNew[order(comNew$FusAn, decreasing = TRUE),]  
comNew$NomCN <- ave(comNew$NomCN, comNew$DepComN, FUN = function(x) x[1])  
comNew <- comNew[,c(2:length(comNew))] # Ne garder que les colonnes utiles  
  
# Jointure avec le fichier d'origine  
df2011 <- merge(df2011, comNew[,c("DepComA","ChefLieu","ComDLG", "FusAn","FusPhas")],  
 by.x = "CODGEO", by.y = "DepComA", all.x = TRUE, all.y = TRUE)  
  
# Identification communes nouvelles  
df2011$COM\_NOUV[!is.na(df2011$FusPhas)] <- "OUI"  
df2011$COM\_NOUV[is.na(df2011$FusPhas)] <- "NON"  
df2011$FusPhas <- as.character(df2011$FusPhas)

# 3.3 - Gestion des scissions communales

Dans la table de passage, des scissions communales sont présentes. On fait en sorte ici que ces dernières ne soient pas considérées comme des fusions ou des communes nouvelles.

Il faut, là encore, noter un cas particulier : en 2012, la commune de Robert-Magny-Laneuville-à-Rémy (52427) se scinde en Robert-Magny (52427) et Laneuville-à-Rémy (52266). Au 1er janvier 2016, Robert-Magny fusionne avec Montier-en-Der (52331) pour former La Porte du Der (52331). Comme on ne dispose pas des données 2011 pour Laneuville-à-Rémy, on préfère ignorer cette scission. Source : <https://www.insee.fr/fr/metadonnees/cog/commune/COM52427-robert-magny>.

# Tout ce qui est scission n'est pas considéré comme commune nouvelle ou fusion  
# comSci <- data.frame(read\_excel("data-raw/communes\_nouvelles/table\_passage\_geo2003\_geo2020.xlsx", sheet = "Liste des scissions", skip = 5))  
comSci <- data.frame(read\_excel("data-raw/communes\_nouvelles/table\_passage\_geo2003\_geo2021.xlsx", sheet = "Liste des scissions", skip = 5))  
  
  
# Cas particulier de la commune de Robert-Magny-Laneuville-à-Rémy (52427).  
comSci <- subset(comSci, COM\_INI != "52427")  
comSci <- as.vector(unique(comSci$COM\_FIN))  
  
df2011$FUSION [df2011$CODGEO %in% comSci] <- "NON"  
df2011$COM\_NOUV [df2011$CODGEO %in% comSci] <- "NON"  
df2011$ChefLieu [df2011$CODGEO %in% comSci] <- NA  
df2011$ComDLG [df2011$CODGEO %in% comSci] <- NA  
df2011$FusAn [df2011$CODGEO %in% comSci] <- NA  
df2011$FusPhas [df2011$CODGEO %in% comSci] <- NA  
  
df2011$FusAn <- as.character(df2011$FusAn)

# 4 - Agrégation des indicateurs en version la plus récente et calcul des ratios

Les données d’intérêt sont fusionnées dans les géographies du 1er janvier de l’année la plus récente grâce à la table de passage 2003-20XX fournie par l’INSEE et la fonction comix qui permet d’agréger les données selon leurs spécificités statistiques (données textuelles, de stock, de ratio)

On importe préalablement les zonages de référence pour l’année la plus récente et on calcule les typologies dérivées du zonage en aires urbaines comme plus haut. Les variables de stock et les informations relatives aux communes nouvelles sont ensuite agrégées puis jointe à cette table de référence. Les codes INSEE aux géographies du 1er janvier de l’année la plus récente sont aussi joints à la table de données du 1er janvier 2011.

# Import de la table de zonage de l'année la plus récente  
# df\_new <- data.frame(read\_excel("data-raw/stats\_insee/table-appartenance-geo-communes-20.xlsx", sheet = "COM", skip = 5))  
df\_new <- data.frame(read\_excel("data-raw/stats\_insee/table-appartenance-geo-communes-21.xlsx", sheet = "COM", skip = 5))  
  
df\_new <- df\_new[,c("CODGEO","LIBGEO","DEP","REG")] # NB : certains éléments ne sont plus dans les données à partir de 2021 : "CATAEU2010","AU2010", "EPCI"... Ont donc été supprimées. Il faudra songer à intégrer le nouveau zonage lorsqu'il sera disponible. Ci-dessous, commenté, le code qui servait à l'intégration des données jusqu'en 2020  
  
# df\_new <- df\_new[,c("CODGEO","LIBGEO","DEP","REG","CATAEU2010","AU2010","EPCI")]  
# colnames(df\_new)[1:3] <- c("CODGEO\_new","LIBGEO\_new","CODE\_DEPT")   
# # Création catégories  
# df\_new$ZAU\_POL [df\_new$CATAEU2010 %in% c("221", "222")] <- "OUI"  
# df\_new$ZAU\_RUR [df\_new$CATAEU2010 %in% c("300", "400")] <- "OUI"  
# df\_new$ZAU\_MAR\_SP [df\_new$CATAEU2010 %in% c("222", "300", "400")] <- "OUI"  
# df\_new$ZAU\_MAR [df\_new$CATAEU2010 %in% c("221", "222", "300", "400")] <- "OUI"  
# df\_new$ZAU\_PERI [df\_new$CATAEU2010 %in% c("112", "120")] <- "OUI"  
# df\_new$ZAU\_AU [df\_new$CATAEU2010 %in% c("111", "112", "120")] <- "GrdAirUrb"  
# df\_new$ZAU\_AU [df\_new$CATAEU2010 %in% c("211", "212", "221", "222")] <- "AirUrb"  
# df\_new$ZAU\_AU [df\_new$CATAEU2010 %in% c("300", "400")] <- "Autres"  
# # Tout ce qui ne fait pas partie des catégories = NON   
# selecV <- c("ZAU\_POL", "ZAU\_RUR", "ZAU\_MAR\_SP", "ZAU\_MAR", "ZAU\_PERI", "ZAU\_AU")  
# df\_new[selecV][is.na(df\_new[selecV])] <- "NON"  
  
  
colnames(df\_new)[1:3] <- c("CODGEO\_new","LIBGEO\_new","CODE\_DEPT")   
  
  
  
# Agrégation des données de 2011 vers l'année la plus récente (stock + texte)  
# Import de la table de passage   
# modif <- as.data.frame(read\_xlsx("data-raw/communes\_nouvelles/table\_passage\_geo2003\_geo2020.xlsx", col\_types = "text", sheet = "Table de passage", skip = 5))  
modif <- as.data.frame(read\_xlsx("data-raw/communes\_nouvelles/table\_passage\_geo2003\_geo2021.xlsx", col\_types = "text", sheet = "Table de passage", skip = 5))  
colnames(modif) <- c("CODGEO\_INI", "CODGEO\_new", "LIBGEO\_new" )  
  
# Agrégation  
tmp <- comix(x = df2011, xid = "CODGEO", app = modif,  
 app.init = "CODGEO\_INI", app.target = "CODGEO\_new",   
 var = c("CODGEO", colnames(df2011)[c(8:40)], "FUSION", "FusAn",  
 "FusPhas", "COM\_NOUV", "CATAEU2010"),  
 var.type = c("text", rep("stock",33), rep("text",5)))  
  
# Gestion des "OUI-NON" > transformé en "OUI"  
selec <- c("FUSION", "COM\_NOUV")  
tmp[,selec][tmp[,selec] == "OUI-NON"] <- "OUI"  
tmp[,selec][tmp[,selec] == "NON-OUI"] <- "OUI"  
  
# Jointure avec le fichier de référence le plus récent  
df\_new <- merge(df\_new, tmp, by.x = "CODGEO\_new", by.y = "CODGEO\_new",  
 all.x = TRUE)  
  
# Suppression des territoires corses, des DOM et des communes associées (pas de geom)  
out <- c("94","01","02","03","04","06")  
df\_new <- df\_new[!df\_new$REG %in% out,]   
df\_new <- df\_new[!is.na(df\_new$CODGEO),]  
  
  
# Jointure des codes géographiques récents pour df2011  
df2011 <- merge(df2011, modif, by.x = "CODGEO", by.y = "CODGEO\_INI", all.x = TRUE)  
  
# Gestion des chefs-lieux dans les fusions successives  
# Suppression des chefs lieux des communes ayant fusionné a posteriori  
df2011$ChefLieu <- ifelse(df2011$CODGEO == df2011$CODGEO\_new, "O", "N")  
  
# Transformation en factor d'un champ  
#df\_new$CATAEU2010 <- as.factor(df\_new$CATAEU2010) # Ne sont plus dans les données à partir de 2021  
df2011$CATAEU2010 <- as.factor(df2011$CATAEU2010)

Pour les jeux de données aux géographies 2011 et pour l’année la plus récente, des ratios d’intérêt sont compilés à partir des stocks préalablement chargés et agrégés. Ces ratios sont définis sur les bases de ce qui est renseigné dans la feuille “ratios” du fichier meta.xls

ratio <- as.data.frame(read\_excel("data-raw/meta.xlsx", sheet = "ratios"))  
  
# Compilation des ratios pour les communes 2011  
for (i in 1:nrow(ratio)){  
 df2011[paste0(ratio[i,"CODE"])] <- df2011[[ratio[i, "Numerator\_Code"]]] /   
 df2011[[ratio[i, "Denominator\_Code"]]] \* ratio[i,"Coeff"]  
}  
  
# Compilation des ratios pour les communes à la dernière géométrie connue  
for (i in 1:nrow(ratio)){  
 df\_new[paste0(ratio[i,"CODE"])] <- df\_new[[ratio[i, "Numerator\_Code"]]] /   
 df\_new[[ratio[i, "Denominator\_Code"]]] \* ratio[i,"Coeff"]  
}

# 5 - Résumé des indicateurs disponibles par fichier de données de référence

Deux jeux de données ont ainsi été produits :

* Le premier, df2011, contient l’ensemble des données communales françaises désirées aux géographies du 1er janvier 2011.
* Le second, df\_new, contient les informations aux géographies du 1er janvier de la dernière année pour laquelle les données ont été publiées par l’INSEE. Les données 2011 ont été regroupées en fonction du fichier des fusions délivré par l’INSEE et la fonction comix.

Voici les codes respectifs que ces fichiers contiennent, ainsi qu’un descriptif de leur contenu.

|  |  |
| --- | --- |
| codes | label |
| CODGEO | Code INSEE, découpage géographique au 01/01/2011 |
| LIBGEO | Nom de la commune, découpage géographique au 01/01/2011 |
| CODE\_DEPT | Code départemental |
| CATAEU2010 | Catégorie de communes dans le zonage en aires urbaines 2010 |
| REG | Région |
| AU2010 | Aire urbaine 2010 |
| EPCI | EPCI |
| P09\_ACT1564 | Actifs 15-64 ans en 2009 |
| P09\_CHOM1564 | Chômeurs 15-64 ans en 2009 |
| P09\_ETUD1564 | Elèv. Etud. Stag. non rémunérés 15-64 ans en 2009 |
| P09\_RETR1564 | Retraités Préretraités 15-64 ans en 2009 |
| C09\_ACT1564\_Agr | Actifs 15-64 ans Agriculteurs exploitants en 2009 |
| C09\_ACT1564\_ArtCom | Actifs 15-64 ans Artisans, Comm., Chefs entr. en 2009 |
| C09\_ACT1564\_Cadr | Actifs 15-64 ans Cadres, Prof. intel. sup. en 2009 |
| C09\_ACT1564\_ProfInt | Actifs 15-64 ans Prof. intermédiaires en 2009 |
| C09\_ACT1564\_Empl | Actifs 15-64 ans Employés en 2009 |
| C09\_ACT1564\_Ouvr | Actifs 15-64 ans Ouvriers en 2009 |
| P09\_EMPLT | Emplois au LT en 2009 |
| C09\_EMPLT\_AGRI | Emplois au LT Agriculture en 2009 |
| C09\_EMPLT\_INDUS | Emplois au LT Industrie en 2009 |
| C09\_EMPLT\_CONST | Emplois au LT Construction en 2009 |
| C09\_EMPLT\_CTS | Emplois au LT Commerce, Transports, Services divers en 2009 |
| C09\_EMPLT\_APESAS | Emplois au LT Adm publique, Enseignement, Santé, Act sociale en 2009 |
| P09\_ACTOCC | Actifs occupés 15-64 ans en 2009 |
| P09\_POP | Population en 2009 |
| P09\_POP0014 | Pop 0-14 ans en 2009 |
| P09\_POP1529 | Pop 15-29 ans en 2009 |
| P09\_POP3044 | Pop 30-44 ans en 2009 |
| P09\_POP4559 | Pop 45-59 ans en 2009 |
| P09\_POP6074 | Pop 60-74 ans en 2009 |
| P09\_POP75P | Pop 75 ans ou plus en 2009 |
| C09\_ACTOCC\_IN | Nombre d’actifs travaillant dans leur commune (ou arrondissement municipal) de résidence |
| C09\_ACTOCC\_OUT | Nombre d’actifs travaillant hors de leur commune (ou arrondissement municipal) de résidence |
| C09\_ACTOCC | Total : nombre d’actifs de 15 ans ou plus ayant un emploi |
| P11\_POT\_FIN | Potentiel financier des communes |
| P11\_DGF | Dotation globale de fonctionnement des communes en 2011 |
| P11\_FoyFisc | Nombre de foyers fiscaux dans la commune en 2011 |
| P11\_Rev\_Fisc | Revenu fiscal de référence de l’ensemble des foyers imposés sur la commune |
| P11\_IMP\_NET | Impôt net total sur la commune |
| P11\_FoyFisc\_Imp | Nombre de foyers fiscaux imposables sur la commune |
| ZAU\_POL | Petit pôle et couronne de petits pôles (catégories 221 et 222 du zonage en aires urbaines (OUI/NON) |
| ZAU\_RUR | Communes rurales et autres (catégories 300 et 400 du zonage en aires urbaines (OUI/NON) |
| ZAU\_MAR\_SP | Marges sans pôles (catégories 222, 300 ey 400 du zonage en aires urbaines (OUI/NON) |
| ZAU\_MAR | Rural profond + petits pôles et couronnes (catégories 221, 222, 300 et 400 du zonage en aires urbaines (OUI/NON) |
| ZAU\_PERI | Périurbain de grands pôles (catégories 112 et 120 du zonage en aires urbaines (OUI/NON) |
| ZAU\_AU | Typologie regroupée basées sur les aires urbaines : GrdAirUrb (catégories 111, 112, 120), AirUrb (211, 212, 221, 222) et Autres (300, 400) |
| FUSION | Code INSEE, découpage géographique le plus récent |
| ChefLieu | Nom de la commune, découpage géographique le plus récent |
| ComDLG | Est-ce que la commune a fusionné entre 2011 et aujourd’hui (OUI/NON) |
| FusAn | Commune nouvelle : Statut de commune chef-lieu (O=OUI, N=NON) |
| FusPhas | Commune nouvelle : Statut de Commune déléguée (O=OUI, N=NON) |
| COM\_NOUV | Date d’effet de la création de commune nouvelle |
| CODGEO\_new | Phase dans le processus de fusion (phase 1 = 2011-2015, 2 = 2015-2016, 3 = 2016-2017, 4 = 2017-2018, 5 = 2018-2020, 6 = 2020-2021 |
| LIBGEO\_new | Est que la commune est devenue une commune nouvelle (OUI/NON) |
| P09\_CHOM1564\_RT | Taux de chômage des 15-64 ans (%) |
| P09\_ETUD1564\_RT | Part des étudiants, stagiaires, non rémunérés dans les actifs de 15-64 ans (%) |
| P09\_RETR1564\_RT | Part des retraités et pré-retraités dans les actifs de 15-64 ans (%) |
| C09\_ACT1564\_Agr\_RT | Part des agriculteurs dans les actifs de 15-64 ans (%) |
| C09\_ACT1564\_ArtCom\_RT | Part des artisans, comm., chefs entr. dans les actifs de 15-64 ans (%) |
| C09\_ACT1564\_Cadr\_RT | Part des cadres, prof. intel. sup. dans les actifs de 15-64 ans (%) |
| C09\_ACT1564\_ProfInt\_RT | Part des prof. intermédiaires dans les actifs de 15-64 ans (%) |
| C09\_ACT1564\_Empl\_RT | Part des employés dans les actifs de 15-64 ans (%) |
| C09\_ACT1564\_Ouvr\_RT | Part des ouvriers dans les actifs de 15-64 ans (%) |
| C09\_EMPLT\_AGRI\_RT | Part de l’agriculture dans l’emploi au lieu de travail (%) |
| C09\_EMPLT\_INDUS\_RT | Part de l’industrie dans l’emploi au lieu de travail (%) |
| C09\_EMPLT\_CONST\_RT | Part de la construction dans l’emploi au lieu de travail (%) |
| C09\_EMPLT\_CTS\_RT | Part du commerce, transports, services divers dans l’emploi au lieu de travail (%) |
| C09\_EMPLT\_APESAS\_RT | Part de l’adm publique, enseignement, santé, act sociale dans l’emploi au lieu de travail (%) |
| P09\_POP0014Y\_RT | Part des 0-14 ans dans la population totale (%) |
| P09\_POP1529Y\_RT | Part des 15-29 ans dans la population totale (%) |
| P09\_POP3044Y\_RT | Part des 30-44 ans dans la population totale (%) |
| P09\_POP4559Y\_RT | Part des 45-59 ans dans la population totale (%) |
| P09\_POP6074Y\_RT | Part des 60-74 ans dans la population totale (%) |
| P09\_POP75PY\_RT | Part des 75 ans et plus dans la population totale (%) |
| C09\_ACTOCC\_IN\_RT | Part des actifs occupés travaillant dans leur commune de résidence (%) |
| C09\_ACTOCC\_OUT\_RT | Part des actifs occupés travaillant hors de leur commune de résidence (%) |
| C09\_EMP\_CONC\_RT | Concentration d’emploi au lieu de travail (nombre d’actifs occupés = 100) |
| P11\_POT\_FIN\_RT | Potentiel financier par habitant (euros) |
| P11\_DGF\_RT | Dotation globale de fonctionnement par habitant (euros) |
| P11\_Rev\_Fisc\_RT | Revenu fiscal de référence par habitant (euros) |
| P11\_IMP\_NET\_RT | Impôt net par habitant (euros) |
| P11\_FoyFisc\_Imp\_RT | Pourcentage de foyers fiscaux imposables |

Il s’agit exactement des mêmes codes utilisés pour df\_new, à la différence que l’identifiant géographique des communes correspond à CODGEO\_new et son libellé LIBGEO\_new. Dans df\_new, CODGEO correspond à la concaténation des communes constituant la commune fusionnée ou la commune nouvelle. Les variables dédiées aux chefs-lieux et aux communes déléguées (codes ChefLieu et ComDLG) ne sont pas inclues dans le jeu de données aux géographies récentes, car non adaptées à ce niveau de granularité géographique.

Les jeux de données de référence sont exportés dans le dossier data au format RData.

try(save(df2011, df\_new, file = "data/refdata.Rdata"))

# 6 - Agrégation des dernières géométries connues (2021), extraction des communes nouvelles et export des résultats

On agrège à présent les géométries aux géographies du 1er janvier 2011 vers les géométries les plus récentes pour lesquelles des données ont été publiées par l’INSEE (1er janvier 2021). Les communes nouvelles sont ensuite extraites de ces géométries de référence.

# Communes françaises au 1er janvier 2011  
geom2011 <- geom2011[,c("CODGEO")]  
  
# Communes françaises à la géométrie la plus récente  
geom\_new <- merge(geom2011, df2011[,c("CODGEO","CODGEO\_new")])  
geom\_new <- aggregate(x = geom\_new[,"CODGEO\_new"],   
 by = list(geom\_new$CODGEO\_new), FUN = head, 1)  
geom\_new <- geom\_new[,c("CODGEO\_new")]  
  
  
# Communes nouvelles avant et après fusion  
geomfus2011 <- merge(geom2011, df2011[,c("CODGEO", "COM\_NOUV")])  
geomCN\_new <- merge(geom\_new, df\_new[,c("CODGEO\_new", "COM\_NOUV")])  
  
geomfus2011 <- geomfus2011[geomfus2011$COM\_NOUV == "OUI",]  
geomfus2011 <- geomfus2011[,c("CODGEO")]  
geomCN\_new <- geomCN\_new[geomCN\_new$COM\_NOUV == "OUI",]  
geomCN\_new <- geomCN\_new[,c("CODGEO\_new")]

5 couches géographiques sont exportées dans le fichier geom.gpkg :

* Les géométries communales au 1er janvier 2011 (geom2011).
* Les géométries communales au 1er janvier de l’année la plus récente pour laquelle des données ont été publiées par l’INSEE (2021) (geom\_new).
* Les géométries des communes concernées par une création de communes nouvelles, avant fusion (geomfus2011).
* Les géométries des communes nouvelles, après fusion (geomCN\_new).
* Une couche d’habillage pour les représentations cartographiques : les départements (dep).

Ces géométries de référence sont exportées dans le dossier “data.”

# Export des fichiers consolidés  
st\_write(obj = geom2011, dsn = "data/geom.gpkg", layer = "geom2011", delete\_layer = TRUE, quiet = TRUE)  
st\_write(obj = geom\_new, dsn = "data/geom.gpkg", layer = "geom\_new", delete\_layer = TRUE, quiet = TRUE)  
st\_write(obj = geomfus2011, dsn = "data/geom.gpkg", layer = "geomfus2011", delete\_layer = TRUE, quiet = TRUE)  
st\_write(obj = geomCN\_new, dsn = "data/geom.gpkg", layer = "geomCN\_new", delete\_layer = TRUE, quiet = TRUE)  
st\_write(obj = dep, dsn = "data/geom.gpkg", layer = "dep",   
 delete\_layer = TRUE, quiet = TRUE)

# [Annexes] Données additionnelles non intégrées au flux de préparation des données

Si on ne conserve pas ces informations, il faudra les supprimer du dossier raw-data pour éviter d’avoir à stocker des informations inutiles.

## Appartenances à un syndicat intercommunal

A voir si on le place dans la préparation des données > créer un nouvel objet “data\_epci.”

La carte des communes nouvelles ne ressemble pas à beaucoup de cartes connues (densités, zonages urbains, caractéristiques socio-économiques, profils démographiques…). On peut néanmoins voir une similitude avec la carte montrant la baisse du nombre de syndicats intercommunaux entre 2010 et 2019 (carte 2 p. 3 du *Bulletin d’information statistique de la DGCL*, n°134, avril 2019 : <https://www.collectivites-locales.gouv.fr/files/files/dgcl_v2/DESL/bis_134_interco_2019.pdf>) Les données concernant les appartenances aux intercommunalités sont archivées ici : <https://www.banatic.interieur.gouv.fr/V5/archivage/archivage.php>. Nous avons utilisé les données au 1er janvier 2011, disponibles ici : <https://www.banatic.interieur.gouv.fr/V5/fichiers-en-telechargement/fichiers-telech.php?arch=01/01/2011> Pour accéder aux appartenances à toutes les structures intercommunales, il faut télécharger les fichiers région par région, ce qui a été fait dans le dossier data-raw/Groupements\_intercommunaux\_2011. [Edit] Après test de chi-2, on peut rejeter l’hypothèse d’indépendance des variables “Fusion” et “Dernière intercommunalité,” mais sans que les Syndicats intercommunaux sans fiscalité propre soient extrêmement plus représentés chez les communes fusionnantes que chez les autres. Donc, a priori, la similarité entre les deux cartes n’est pas forcément lié à ce point. Ou alors, nécessiterait d’avoir les données plus récentes ? Répartition des communes fusionnantes ou non en fonction des dernières intercommunalités en 2009 : CA CC CU SIVOM SIVU SMF SMO NON 0.1 0.5 0.1 14.1 17.6 32.6 35.0 OUI 0.0 1.3 0.1 11.3 18.9 33.7 34.7 C’est, en revanche, davantage visible si on ne considère que les régions “Ouest” (Normandies, Bretagne, Pays-de-la-Loire), pour lesquelles le résultat du test est : X-squared = 103.52, df = 6, p-value < 2.2e-16

CA CC CU SIVOM SIVU SMF SMO

NON 0.2 0.4 0.1 7.0 36.4 35.0 20.9 OUI 0.0 1.1 0.2 4.6 23.2 43.6 27.3

# list.files("data-raw/Groupements\_intercommunaux\_2011")  
  
Int\_1 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Alsace.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_2 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Aquitaine.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_3 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Auvergne.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_4 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Basse-Normandie.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_5 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Bourgogne.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_6 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Bretagne.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_7 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Centre.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_8 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Champagne-Ardenne.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_9 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Corse.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_10 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Franche-Comté.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_11 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Haute-Normandie.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_12 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Île-de-France.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_13 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Languedoc-Roussillon.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_14 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Limousin.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_15 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Lorraine.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_16 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Midi-Pyrénées.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_17 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Nord-Pas-de-Calais.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_18 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Pays de la Loire.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_19 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Picardie.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_20 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Poitou-Charentes.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_21 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Provence-Alpes-Côte d'Azur.csv", stringsAsFactors=TRUE, sep = "\t")  
Int\_22 <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Rhône-Alpes.csv", stringsAsFactors=TRUE, sep = "\t")  
#Compliation des fichiers  
Interco\_Compil <- rbind (Int\_1, Int\_2, Int\_3, Int\_4, Int\_5, Int\_6, Int\_7,  
 Int\_8, Int\_9, Int\_10, Int\_11, Int\_12, Int\_13, Int\_14,  
 Int\_15, Int\_16, Int\_17, Int\_18, Int\_19, Int\_20, Int\_21, Int\_22)  
# Export des fichiers pour réutilisation ultérieure  
write.table(Interco\_Compil, "data-raw/Groupements\_intercommunaux\_2011/Compilation\_Interco.csv",  
 sep="\t", row.names=FALSE)  
# Import du fichier compilé  
Interco <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Compilation\_Interco.csv",  
 stringsAsFactors=TRUE, sep = "\t")

IntercoTout <- read.csv2("data-raw/Groupements\_intercommunaux\_2011/Compilation\_Interco.csv",  
 stringsAsFactors=TRUE, sep = "\t")  
IntercoTout$DateEffet <- as.Date(IntercoTout$Date.d.effet, tryFormats = "%Y-%m-%d")  
# summary(IntercoTout$DateEffet)  
  
# Sélection, pour chaque commune, de la ligne correspondant à la date la plus récente  
summary(IntercoTout$Type)  
IntercoSelect <- subset (IntercoTout, Type == "Commune")  
# Pour voir les occurrences  
# IntercoSelect <- unique(IntercoTout$Nom.membre)  
# IntercoSelect [1:4]  
# On identifie les doublons (quand une commune est présente une seule fois) en commençant par les plus anciens  
doublons <- which(duplicated(IntercoSelect[,"Nom.membre"], fromLast=TRUE))  
doublons[1:10]  
# On supprime les doublons  
DernierInterco <- IntercoSelect[-doublons,]  
# On ne garde que les donées qui nous intéressent  
DernierInterco <- DernierInterco[,c("Nom.membre","DateEffet","Nom.du.groupement", "Nature.juridique")]  
rm(IntercoTout)  
rm(IntercoSelect)  
rm(doublons)  
  
colnames(DernierInterco) [1] <- "Nom.commune"  
colnames(DernierInterco) [2] <- "DateDernierInterco"  
colnames(DernierInterco) [3] <- "NomDernierInterco"  
colnames(DernierInterco) [4] <- "NatureDernierInterco"