

Dealing with the change of administrative divisions over time with

Kim Antunez

 antuki13

Technical Interview with Shopify
13 February 2020

A data science project (1/4)

« A data science project you have worked on yourself independently »



- ~~Producing word embeddings with word2vec~~
- ~~Web app: interactive graphs about a French opinion survey~~
- **COGugaison and CARTElette : R packages in open-source for manipulating French spatial databases produced at different dates.**

« We want you to be able to explain the project inside and out. »

A data science project (2/4)

Available online [here](#)

Le Baromètre d'opinion de la DREES

Outil de visualisation

Documentation et données

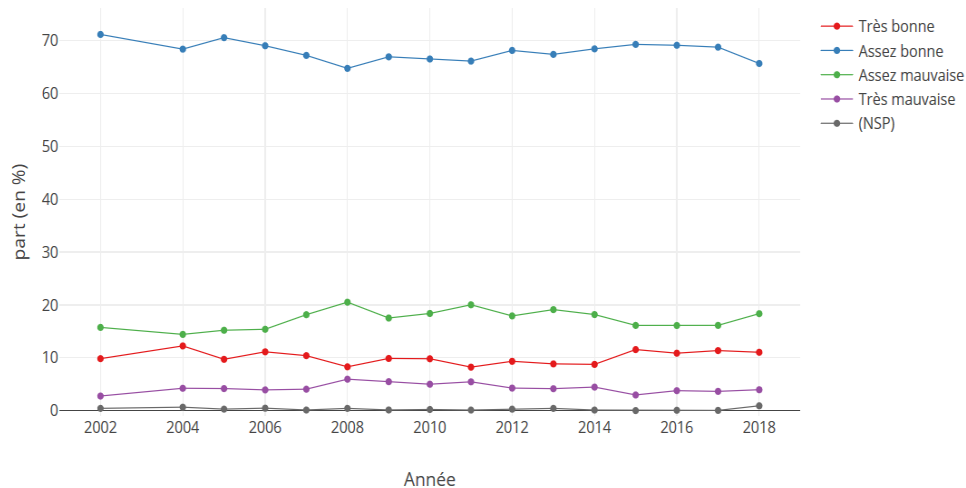
Mentions légales

Le Baromètre d'opinion de la DREES suit chaque année depuis 2000 l'évolution de l'opinion des Français à l'égard de la santé, de la protection sociale et des inégalités sociales.

Graphique

Tableau

OG1. Vous personnellement, comment qualifieriez-vous votre situation actuelle ?



Source : DREES, Baromètre d'opinion 2002-2018
Champ : Population de France métropolitaine âgée de 18 ans ou plus

Module

OPINION GÉNÉRALE

Question

OG1. Vous personnellement, comment qualifieriez-vous votre situation actuelle ?

Très bonne Assez bonne Assez mauvaise
Très mauvaise (NSP)

☐ Retirer du champ les « Ne se prononce pas »

Ventilation

Aucune

Options

☐ Se restreindre à un seul millésime

A data science project (3/4)

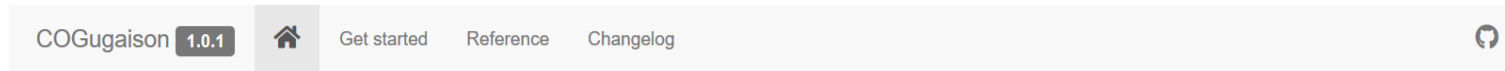


COGugaison is an R package for manipulating french spatial databases produced at different dates.

CARTElette is a complement to the COGugaison package. The goal is to create geographical layers that correspond to the situation of the division of the French territories (France and Overseas) on the first January of each year.

Packages on github: [antuki/COGugaison](https://github.com/antuki/COGugaison) and [antuki/CARTElette](https://github.com/antuki/CARTElette).

A data science project (4/4)



COGugaison

Afin de mieux connaître les utilisateurs des packages COGugaison et CARTElette et de mieux répondre à vos besoins, merci de répondre à cette rapide enquête en ligne : https://antuki.github.io/2019/11/08/opinion_package/

Le découpage des territoires français, en particulier les communes, n'est pas un phénomène immuable. Chaque année certaines communes changent de codes, ou bien de nom, fusionnent ou encore se divisent. Certains périmètres supra-communaux changent également, comme celui des cantons qui a été récemment redéfini. C'est à l'Insee que revient le suivi de ces changements afin d'établir chaque année le code officiel géographique (COG).

Ce package R a alors pour objectif global de manipuler des données communales produites à différents millésimes et de les agréger à différents niveaux supra-communaux. Plus précisément, il permet actuellement de :

- détecter le millésime du code officiel géographique d'une table de données communales : fonction `COG_akinator` et appairer une liste de communes d'une base de données avec celle du COG d'un millésime donné avec `appairier_COG`
- visualiser les modifications communales (fusions, défusions, changements de codes ou de noms) qui ont eu lieu entre deux dates : `modifications_communes`
- visualiser sur un graphique interactif la trajectoire d'une commune donnée, c'est-à-dire toutes les modifications qui ont eu lieu depuis 1968 : `trajectoire_commune` et `trajectoire_commune_shiny`.
- transformer des tables de données numériques en géographie au premier janvier d'une année souhaitée : `changement_COG_varNum`.
- transformer des typologies de communes en géographie au premier janvier d'une année souhaitée en ayant le choix entre plusieurs hypothèses de classement en cas de fusion de communes de classes différentes (attribuer une unique classe à toutes les communes fusionnées, attribuer la classe qui contient le plus de population, définir une classe absorbante, une classe absorbée ou une classe spécifique aux regroupements de plusieurs communes de classes différentes) : `changement_COG_typo`. Il est également possible d'isoler dans une table les communes fusionnées appartenant à des classes différentes : `changement_COG_typo_details`



Links

Browse source code at
<https://github.com/antuki/COGugaison>

Report a bug at
<http://github.com/antuki/COGugaison/issues>

License

GPL (>= 2) | file [LICENSE](#)

Developers

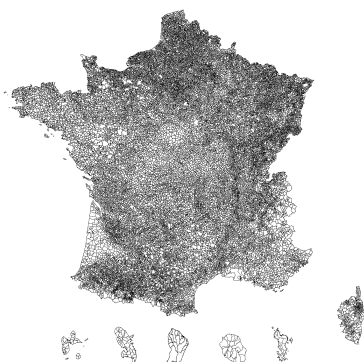
antuki
Maintainer, author

Administrative divisions...

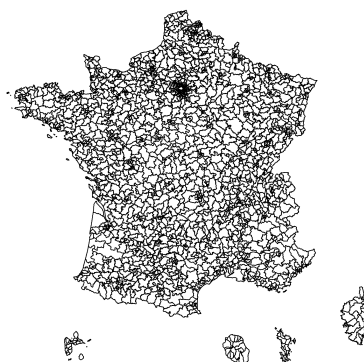


The French territorial « mille-feuille »

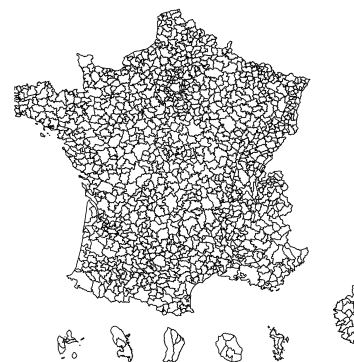
Communes



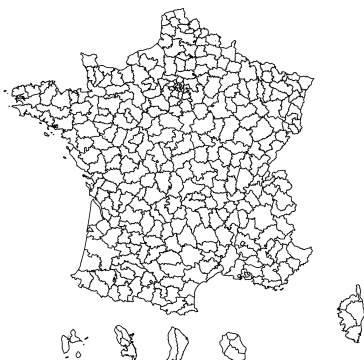
Cantons



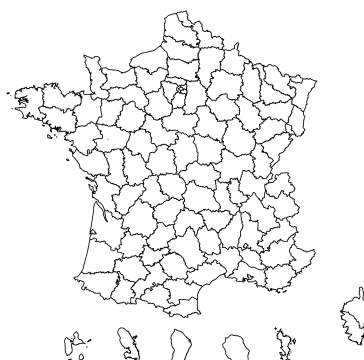
EPCI



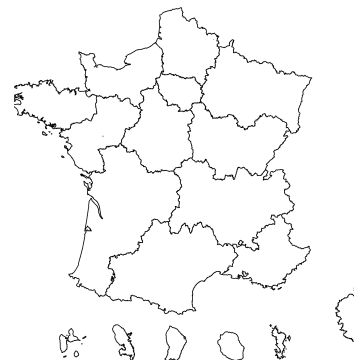
Arrondissements



Départements

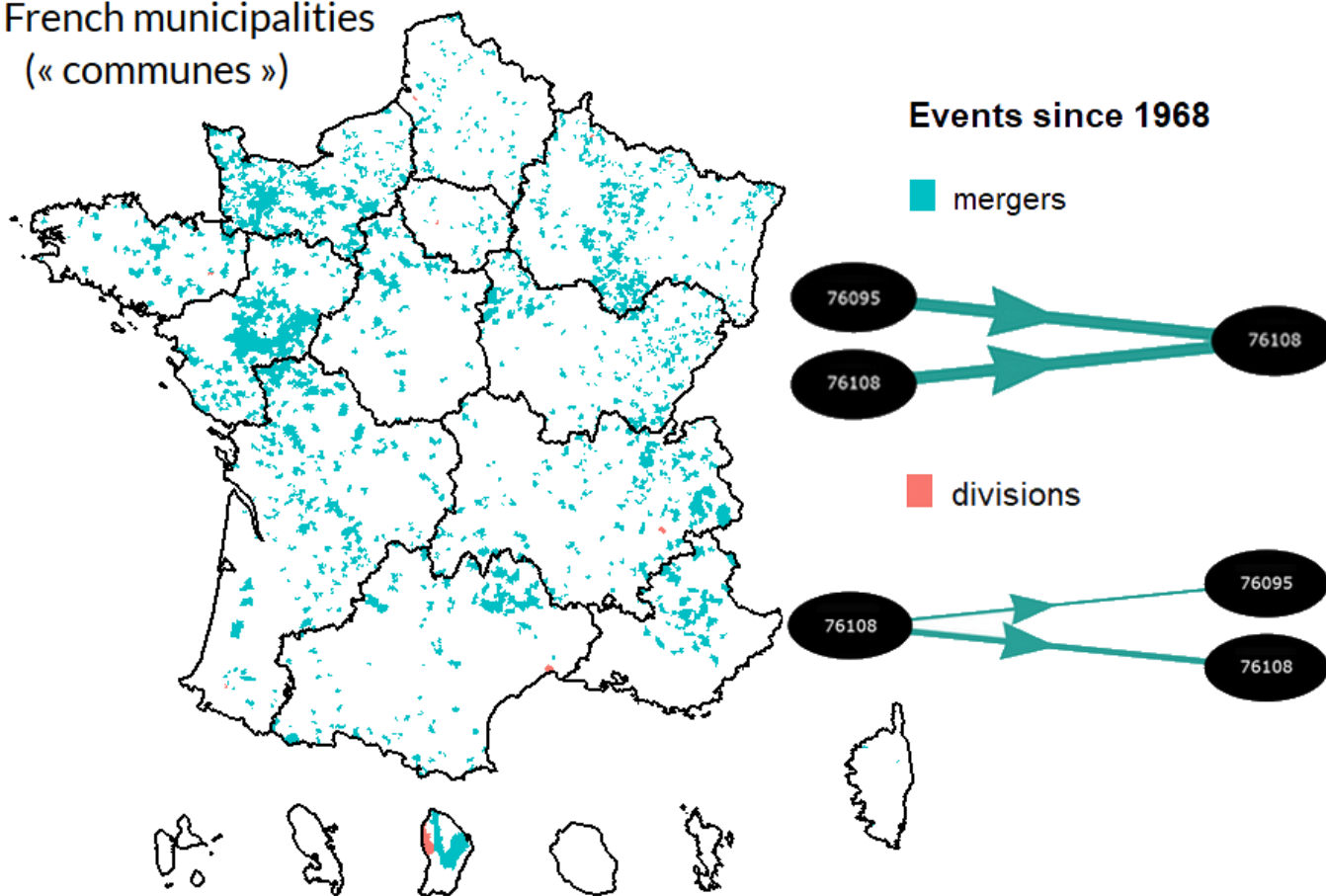


Régions



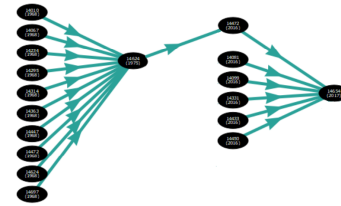
... change over time

French municipalities
(« communes »)



Source : Insee, COG 2019. Map layer antuki/CARTElette

See modifications over the years



⚠ All functions and parameters are translated from French for this presentation

```
municipality_evolution_graph(code = "76108", year = 2014) # uses {visNetwork}
```



```
evol <- municipalities_evolutions(begin_date="01-01-2011", end_date="01-01-2014")
```

```
grep("(76095)|(76108)", evol$mergers, value = TRUE)
```

```
## 2012-01-01: Bois-Guillaume-Bihorel (76108) is a merger of Bihorel (76095),  
Bois-Guillaume (76108).
```

```
grep("(76095)|(76108)", evol$divisions, value = TRUE)
```

```
## 2014-01-01: Bois-Guillaume (76108) divided into Bois-Guillaume (76108), Bihorel  
## (76095).
```


Guess the year of a database



| code | name |
|-------|----------------------|
| 76094 | Bierville |
| 76095 | Bihorel |
| 76096 | Biville-la-Baignarde |

```
head(db, 2)
```

```
##      code      men   women  
## 1 01001 385.0000 380.000  
## 2 01002 142.6132 113.107
```

```
OGC_guess(db$code) # guesses the year of the Official Geographic Code (OGC)
```

```
## [1] "2016"
```

```
codes <- c("99086", db$code[-1]) # creates a vector of ID of municipalities  
merge_OGC(codes = codes, OGC = 2016)$not_in_db # ID not in the db
```

```
## 01001
```

```
merge_OGC(codes = codes, OGC = 2016)$not_in_OGC # ID not in the OGC
```

```
## 99086
```

Change the year of a database

| code | name |
|-------|----------------------|
| 76094 | Bierville |
| 76095 | Bihorel |
| 76096 | Biville-la-Baignarde |



- **qualitative variable** [character]

- 👉 See [change_OGC_typology](#)

- *divisions* ▶ easy: copy the lines
 - *mergers* ▶ several hypotheses: assign the class that contains the most population, define an absorbent or absorbed class...

```
nrow(db)
```

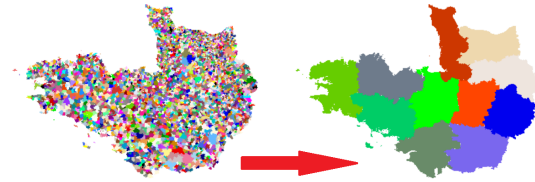
```
## [1] 35887
```

```
db_2019 <- db %>% # changes the year of a numeric variable (from 2016 to 2019)
  change_OGC_numeric(2016:2019)
```

```
str(db_2019)
```

```
## 'data.frame':   34972 obs. of  3 variables:
## $ code : chr  "01001" "01002" "01004" "01005" ...
## $ men : num  385 142.6 6778.5 819 54.5 ...
## $ women: num  380 113.1 7246.5 783.9 49.5 ...
```

Aggregate a database

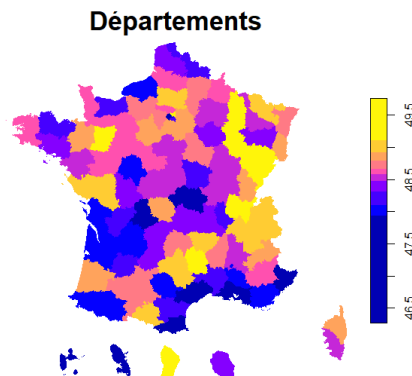


```
db_dep <- db_2019 %>% # aggregates the database (municipalities -> départements)
  aggregate_OGC(OGC = 2019, administrative_division = "DEP")
```

```
str(db_dep)
```

```
## 'data.frame':   100 obs. of  4 variables:
##  $ DEP      : chr  "01" "02" "03" "04" ...
##  $ LIBGEO: chr  "Ain" "Aisne" "Allier" "Alpes-de-Haute-Provence" ...
##  $ men      : num  309198 262693 163759 78600 68309 ...
##  $ women    : num  316839 276817 179006 83124 71588 ...
```

```
DEP_sf <- left_join(DEP_sf, db_dep, by=c("DEP"="DEP")) %>%
  mutate(prop=100*men/(men+women))
plot(DEP_sf %>% select(prop))
```



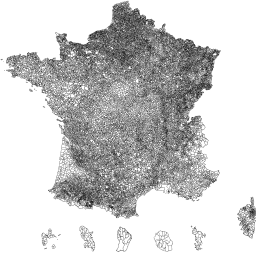
Use map layers adapted to the data


https://github.com/antuki/CARTElette/tree/master/ 80 %

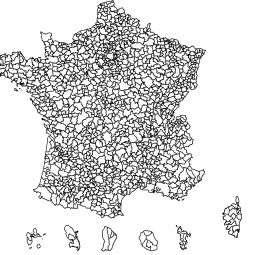
Branch: master ▾ CARTElette / couches_carto / IGN / COG2019 /

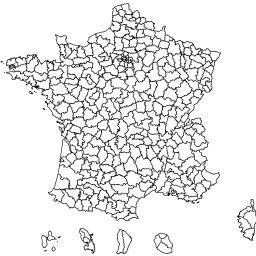
antuki MâJ COG2019 + param enlever_PLM

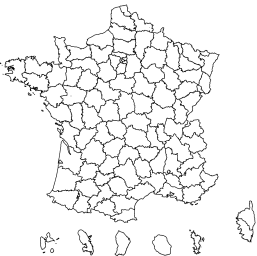
| | |
|-------------------------|---------------------------------|
| ARR_2019_CARTElette.dbf | MâJ COG2019 + param enlever_PLM |
| ARR_2019_CARTElette.prj | MâJ COG2019 + param enlever_PLM |
| ARR_2019_CARTElette.shp | MâJ COG2019 + param enlever_PLM |
| ARR_2019_CARTElette.shx | MâJ COG2019 + param enlever_PLM |
| COM_2019_CARTElette.dbf | MâJ COG2019 + param enlever_PLM |
| COM_2019_CARTElette.prj | MâJ COG2019 + param enlever_PLM |

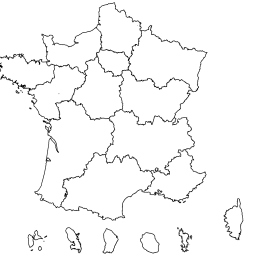
Communes


Cantons


EPCI


Arrondissements


Départements


Régions


Future improvements?

- **Create new functions**
 - add your own geographical levels
 - add your own distribution keys
- **Expand to other countries**
 - in Europe : Nomenclature of Territorial Units for Statistics (NUTS)
 - identify common fonctionnalités VS local (French!) specificities
- **Reach non R-users**
 - Shiny Apps
 - API



 [@antuki13](https://twitter.com/antuki13)  [antuki](https://github.com/antuki)  antuki.github.io

Packages on github: [antuki/COGugaison](https://github.com/antuki/COGugaison) and [antuki/CARTElette](https://github.com/antuki/CARTElette).

Slides created with R package [xaringan](https://github.com/antuki/xaringan) with the [R-Ladies theme](https://github.com/antuki/R-Ladies-theme).

Also with [remark.js](https://github.com/antuki/remark.js), [knitr](https://github.com/antuki/knitr), and [R Markdown](https://github.com/antuki/R-Markdown).