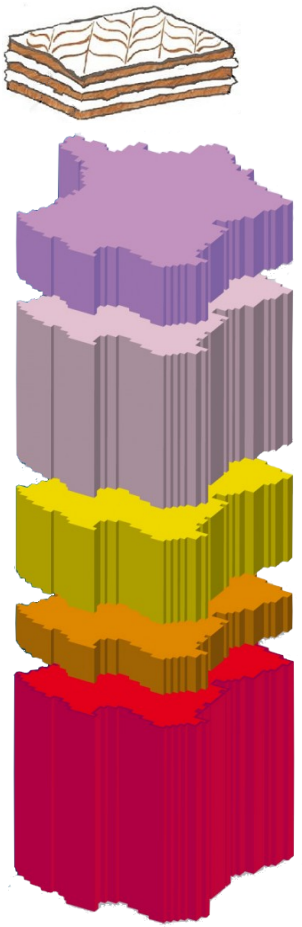
Dealing with the change of administrative divisions over time with R

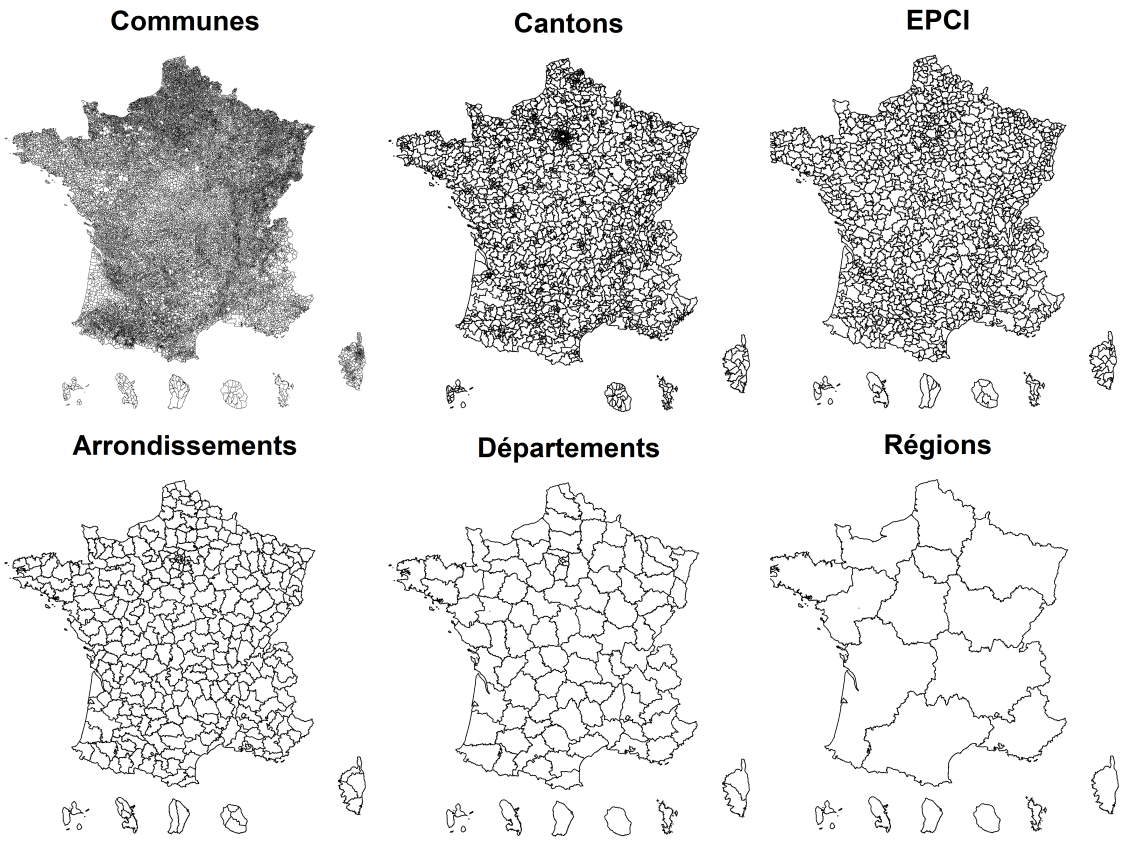
Kim Antunez
antuki13

useR! 2019 - Toulouse, France
11 July 2019

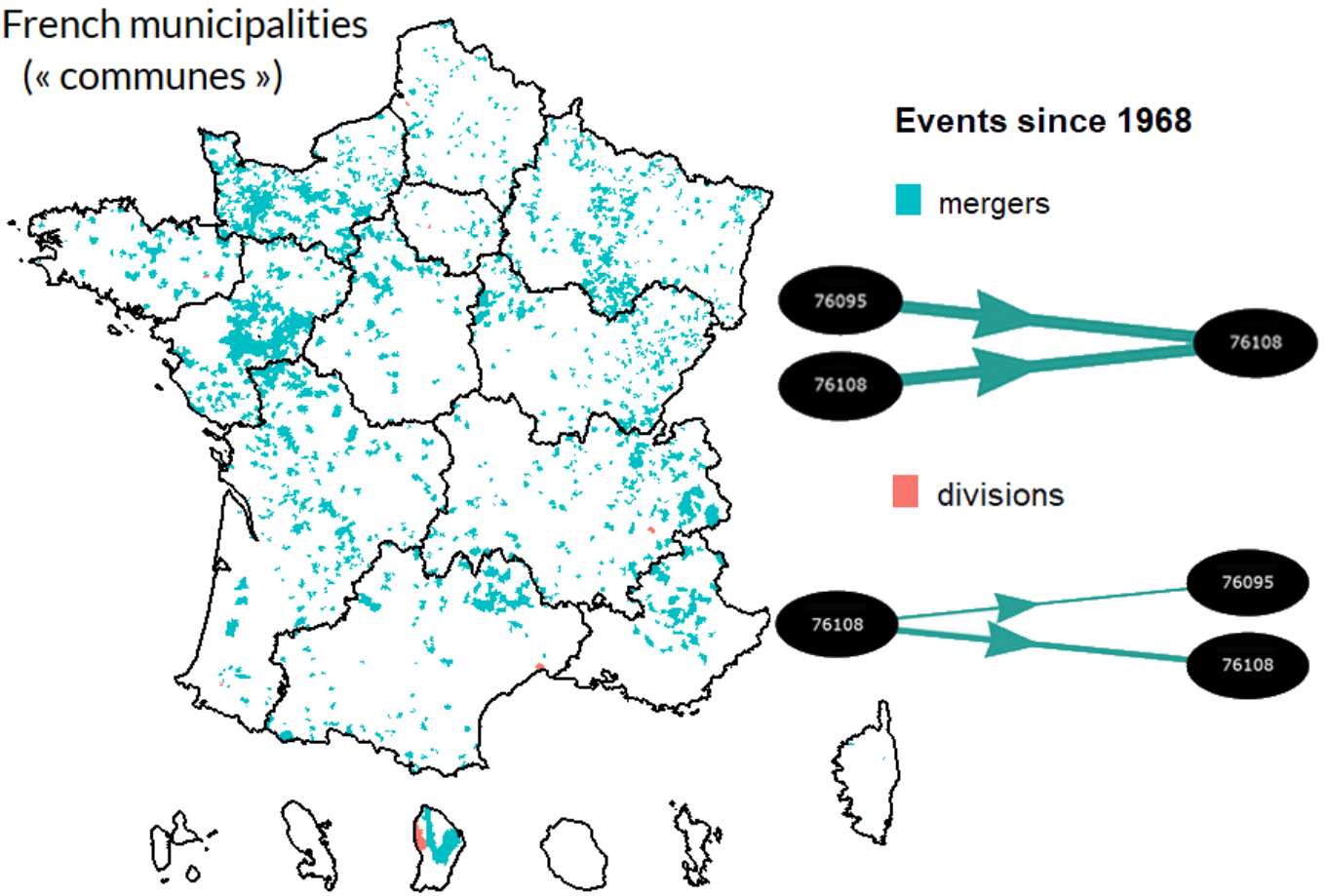
Administrative divisions...



The French territorial « mille-feuille »

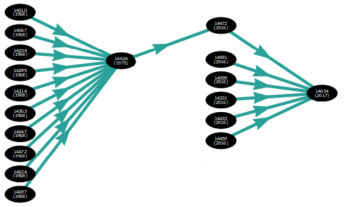


… change over time



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")
```

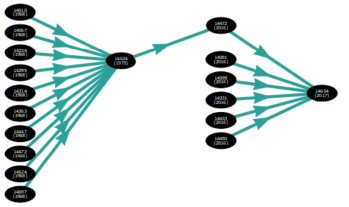
```
grep("(76095)|(76108)", evol$mergers, value = TRUE)
```

```
## 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).
```

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...

- **quantitative variable** [numeric]
  👉 See `change_OGC_numeric`
  - *mergers* ▸ easy: sum the lines
  - *divisions* ▸ divide lines proportionally to population

```
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 ...
```

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...

- **quantitative variable** [numeric]
  👉 See `change_OGC_numeric`
  - *mergers* ▸ easy: sum the lines
  - *divisions* ▸ divide lines proportionally to population

```
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 ...
```

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...

- **quantitative variable** [numeric]
  👉 See `change_OGC_numeric`
  - *mergers* ▸ easy: sum the lines
  - *divisions* ▸ divide lines proportionally to population

```
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 ...
```

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...

- **quantitative variable** [numeric]
  👉 See `change_OGC_numeric`
  - *mergers* ▸ easy: sum the lines
  - *divisions* ▸ divide lines proportionally to population

```
nrow(db)
```

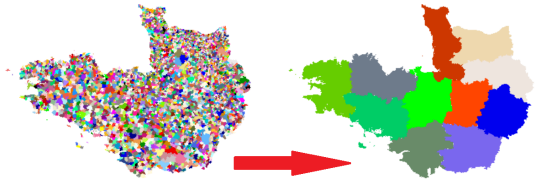
```
## [1] 35887
```

```
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

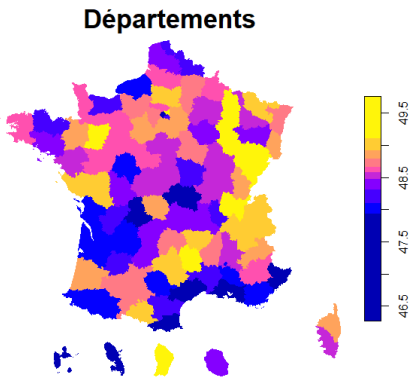


```
db_dep <- db_2019 %>% # aggregates the database (municipalities -> départements)
  aggregate_OGC(OGC = 2019, administrative_division = "DEP")
```

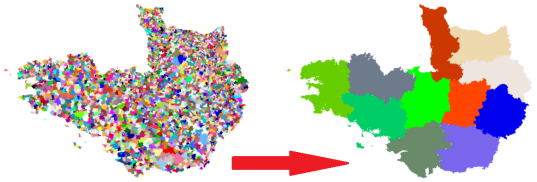
```
str(db_dep)
```

```
## '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))
```



Aggregate a database

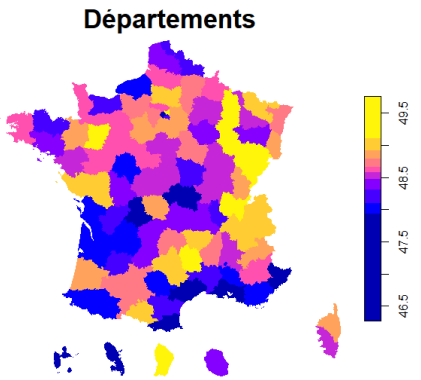


```
db_dep <- db_2019 %>% # aggregates the database (municipalities -> départements)
  aggregate_OGC(OGC = 2019, administrative_division = "DEP")
```

```
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))
```



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 functionnalities VS local (French!) specificities
- **Reach non R-users**
  - Shiny Apps
  - API

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 functionnalities VS local (French!) specificities
- **Reach non R-users**
  - Shiny Apps
  - API

@antuki13 antuki antuki.github.io

Packages on github: antuki/COGugaison and antuki/CARTElette.

Slides created with R package **xaringan** with the R-Ladies theme.

Also with remark.js, knitr, and R Markdown.